Analysis of a Hurst Parameter Estimator Based on the Modified Allan Variance

Alessandra Bianchi*, Stefano Bregni†, Irene Crimaldi‡, Marco Ferrari§

*University of Padua, Department of Mathematics, Padova, Italy, Email: bianchi@math.unipd.it

†Politecnico di Milano, Dep. of Electronics e Information, Milan, Italy, Email: bregni@elet.polimi.it

‡ IMT, Institute for Advanced Studies, Lucca, Italy, Email: irene.crimaldi@imtlucca.it

§IEIIT, National Research Council (CNR), Milan, Italy, Email: marco.ferrari@ieiit.cnr.it

***Abstract*—In order to estimate the Hurst 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 an other method of common use based on wavelet analysis. Here we link it to the wavelets setting and stress why a different analysis for the two approaches is required. We then focus on the asymptotic analysis of the MAVAR log-regression estimator and provide new formulas for the related confidence intervals. By numerical evaluation, we analyze these formulas and make a comparison between three suitable choices on the regression weights, also optimizing over different choices on the data progression.**

I. Introduction

Internet traffic, as well as many different kinds of real data (Hydrology, Economics, Biology), has been demonstrated to exhibit self-similarity and long-range dependence (LRD) on various time scales [1], [2], [3]. In a self-similar random process, a dilated portion of a realization, by the scaling Hurst parameter H, has the same statistical characterization than the whole. On the other hand, the LRD is commonly equated to an asymptotic power-law behaviour of the spectral density of a related stationary random process, and it is thus characterized by the exponent α of such a power-law.

Though a self-similar process can not be stationary (and thus not even LRD), these two proprieties 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 of the increments, providing an asymptotic power-law with exponent $\alpha = 2H - 1$. The most paradigmatic example of this connection is given by the fractional Brownian motion and by its increment process, the fractional Gaussian noise [4].

Among the different techniques introduced in the literature in order to estimate the Hurst parameter H, here we focus on a method based on the log-regression of the Modified Allan Variance (MAVAR). The MAVAR is a well known time-domain quantity generalizing the classic Allan variance [5], [6], [7]. It has been proposed for the first time as a traffic analysis tool in [8] and then its performance in estimating LRD has been evaluated by simulation [8], [9].

Among other examples, it has been successfully applied in estimating the LRD of real IP traffic [10] and of GSM call arrivals [11], while from the theoretical point of view, its good behavior have been confirmed by some recent results where, under the assumption that the signal process is a fractional Brownian motion, the asymptotic normality of the estimator has been shown [12].

These works have pointed out the high accuracy of the method in estimating the parameter H, in particular when compared with the well-established log-diagram based on Daubechies wavelets [8], [9].

The aim of this work is to analyze theoretically and numerically the performance of the MAVAR log-regression estimator for different choices on the regression weights. Here we focus on three different weights: The simple linear regression (SLR)-weights, used in the implementation of the method in [10], [11]; the Abry-Veitch (AV)-weights [13]; the FMRT-weights, introduced in a paper by Fay et al. to analyze the performance of a (Daubechies) wavelet-based estimator [19].

We first show that though a wavelet representation of the MAVAR estimator can be given, in analogy to the well known connection between Allan variance and Haar-wavelets family [13], the two approaches are intrinsically different and a different analysis is required.

Following the asymptotic analysis performed in [12], that shows that the MAVAR estimator is consistent and asymptotically normal distributed, we provide new explicit formulas for the related confidence intervals of the Hurst parameter. By numerical evaluations of such formulas, we show that the asymptotic variance decreases with the rate predicted by their analytical asymptotes, independently of the choice on the weights.

We then perform a comparison between the results obtained for the different regression weights, and optimize the results over different possible choices on the data progression.

The results of the numerical analysis are organized in tables, to be used as a reference, displaying the behavior of the confidence intervals as the size of traffic series and the value of the parameter H vary.

II. Self-Similarity and Long-Range Dependence

According to [3], we consider a centered self-similar real-valued stochastic process $X = \{X(t),\, t \in \mathbb{R}\}$, with $X(0) = 0$, that can be interpreted as the signal process. By *self-similarity* of X we refer to the existence of a parameter

978-1-4673-0921-9/12/$31.00 ©2012 IEEE

$H \in (0,1)$, called *Hurst index or Hurst parameter* of the process, such that, for all $a > 0$, it holds

$$\{X(t)\,, t \in \mathbb{R}\} \stackrel{d}{=} \{a^{-H}X(at),\, t \in \mathbb{R}\}\,. \qquad (1)$$

Assuming further that the process X has *weakly stationary* increments, we get the following expression for the autocovariance function

$$\mathrm{Cov}(X(s), X(t)) = \frac{\sigma_H^2}{2}\left(|t|^{2H} - |t-s|^{2H} + |s|^{2H}\right)\,, \qquad (2)$$

with $\sigma_H^2 := \mathbb{E}[X^2(1)]$. Denoting by Y_τ the τ-increment process of X, defined as $Y_\tau(t) = \frac{X(t+\tau)-X(t)}{\tau}$, it also turns out that the autocovariance function of Y_τ, given by $R_{Y_\tau}(t) = \mathrm{Cov}(Y_\tau(s), Y_\tau(s+t))$, satisfies asymptotically the following power law [1]

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

In particular, if $H \in (\frac{1}{2}, 1)$, the process Y_τ displays *long-range dependence*, in the sense that there exists $\alpha = 2H-1 \in (0,1)$ such that the spectral density of the process Y_τ, $f_{Y_\tau}(\lambda)$, satisfies the condition

$$f_{Y_\tau}(\lambda) \sim c|\lambda|^{-\alpha} \quad \text{as } \lambda \to 0\,,$$

for some finite constant $c \neq 0$. The parameter α is often referred as *memory parameter* of the process Y_τ [14], [15], [16]. Thus, under the assumption that X is a self-similar process with weakly stationary increments, we embrace the two main empirical properties of a wide collection of real data.

A basic example of the connection between these two properties is provided by the *fractional Brownian motion* $B_H = \{B_H(t),\, t \in \mathbb{R}\}$ [4], that is a centered Gaussian process with autocovariance function given by (2) with

$$\sigma_H^2 = \frac{1}{\Gamma(2H+1)\sin(\pi H)}\,. \qquad (3)$$

It can be shown 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 $G_{\tau,H}(t) = \frac{B_H(t+\tau)-B_H(t)}{\tau}$, called *fractional Gaussian noise*, turns out to be a weakly stationary Gaussian process [4], [17], displaying long memory for $H > \frac{1}{2}$.

III. The Modified Allan Variance

In this section we introduce and recall the main properties of the Modified Allan variance (MAVAR) [6], [5], and of the log-regression estimator of the Hurst parameter based on it [8], [9], [10].

A. Definition of MAVAR and related estimator

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 $t_i - t_{i-1} = \tau_0$, i.e. $t_i = t_1 + \tau_0(i-1)$. For any integer $p \geq 1$, we set $\tau = \tau_0 p$ and define the modified Allan variance (MAVAR) as

$$\sigma^2_{p,\tau_\bullet} := \frac{1}{2\tau^2}\mathbb{E}\left(\frac{1}{p}\sum_{i=1}^{p} X(t_i+2\tau) - 2X(t_i+\tau) + X(t_i)\right)^2 \qquad (4)$$

where $\mathbb{E}$ denotes the space-average, that is the average over the set of possible values taken by the signal process X [6]. 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 $t_1, \ldots, t_n$, with constant sampling period τ_0. In other words we set $X_i = X(t_i)$ for $i = 1, \ldots, n$. For $k \in \mathbb{Z}$, let us define

$$d_{p,\tau_\bullet,k} := \frac{1}{\sqrt{2}\,\tau p}\sum_{i=1}^{p}(X_{k+i+2p} - 2X_{k+i+p} + X_{k+i})\,, \qquad (5)$$

and notice that, from the hypotheses on X of Sec. II, $\{d_{p,\tau_\bullet,k}\}_k$ is weakly stationary. Moreover, by definitions (4) and (5), $\sigma^2_{p,\tau_\bullet} = \mathbb{E}\left[d^2_{p,\tau_\bullet,k}\right]$.

The ITU-T standard estimator [18] for the modified Allan variance (MAVAR estimator), also used in [7][8][9][10], except for the different notation is given by

$$\widehat{\sigma}^2_{p,\tau_0}(n) := \frac{1}{n_p}\sum_{k=0}^{n_p-1} d^2_{p,\tau_0,k}\,, \qquad (6)$$

for $p \in \{1, \ldots, \lfloor n/3 \rfloor\}$ and $n_p := n - 3p + 1$, where the space-average $\mathbb{E}[\cdot]$ is replaced by the empirical average over the observations sample.

B. MAVAR and wavelet estimators

Consider the generalized process $Y = \{Y(t),\, t \in \mathbb{R}\}$ defined through the set of identities

$$\int_{t_1}^{t_2} Y(t)dt = X(t_2) - X(t_1)\,, \qquad \forall t_1, t_2 \in \mathbb{R}\,. \qquad (7)$$

In short, we write $Y = \dot{X}$. With this definition we can rewrite the MAVAR and its related estimator as functions of the process Y. In particular we can write

$$d_{p,\tau_\bullet,k} = \frac{1}{\sqrt{2p^2}\tau_0}\sum_{i=1}^{p}\left(\int_{t_{i+k+p}}^{t_{i+k+2p}} Y(t)dt - \int_{t_{i+k}}^{t_{i+k+p}} Y(t)dt\right)\,. \qquad (8)$$

Now we claim that, for p fixed, this random variable 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 set $\tau_0 = 2^j$ and $t_1 = 2^j$, so that $t_i = 2^j i$, 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

$$\begin{aligned} d_{k,j} &:= d(2^j, p, k) = \langle Y; \psi_{k,j}\rangle \\ \text{with} \quad \psi_{k,j}(s) &:= 2^{-j}\psi(2^{-j}s - k)\,. \end{aligned} \qquad (9)$$

An easy check shows that the function $\psi(s) := \frac{1}{p\sqrt{p}}\sum_{i=1}^{p}\psi^i(s)$, where

$$\psi^i(s) := \frac{1}{\sqrt{2p}}\left(\mathrm{I}_{[i+p,i+2p]}(s) - \mathrm{I}_{[i,i+p]}(s)\right)\,, \qquad (10)$$

satisfies Eq. (9). Notice also that the components ψ^i, $i = 1, \ldots p$, of ψ are suitably translated and re-normalized 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 already pointed out in [13].

Although the MAVAR can be related to the above Haar-type function family, we will show that the MAVAR and wavelets log-regression estimators do not match, as the regression runs on different parameters. For the wavelet-based estimators p is fixed and the regression parameter is j (related to τ_0), while for the MAVAR estimator (see Eq. (14)) the regression is on

p with τ_0 fixed. Because of this difference, it is not possible to apply the results available in the wavelets framework [14], [15], [16].

C. The MAVAR log-regression estimator

As proven in [12], applying the covariance formula (2) we get, for $H \in (1/2, 1)$,

$$\left|\sigma^2_{p,\tau_\bullet} - \sigma^2_H \tau^{2H-2} K(H)\right| \le \sigma^2_H \tau^{2H-2} O_H(p^{-1}), \qquad (11)$$

where

$$K(H) := \frac{2^{2H+4} + 2^{2H+3} - 3^{2H+2} - 15}{2(2H+1)(2H+2)}. \qquad (12)$$

This asymptotic relation suggests the following estimation method for the parameter H.

Let n be the sample size, i.e. the number of the observations, choose $\bar{p}, \bar{\ell} \in \mathbb{N}$ and an increasing sequence $\{a_\ell\}_{\ell \in \mathbb{N}}$ such that $1 \le \bar{p} a_{\bar{\ell}} \le p_{\max}(n) = \lfloor n/3 \rfloor$. Let $\underline{w} = (w_0, \ldots, w_{\bar{\ell}})$ be a vector of weights satisfying the conditions

$$\sum_{\ell=0}^{\bar{\ell}} w_\ell = 0 \quad \text{and} \quad \sum_{\ell=0}^{\bar{\ell}} w_\ell \log(a_\ell) = 1. \qquad (13)$$

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

$$\widehat{\mu}_n := \sum_{\ell=0}^{\bar{\ell}} w_\ell \log\left(\widehat{\sigma}^2_{a_\ell \bar{p}, \tau_\bullet}(n)\right). \qquad (14)$$

Roughly speaking, the idea behind this definition is to use the approximation $\widehat{\sigma}^2_{a_\ell \bar{p}, \tau_\bullet}(n) \simeq \sigma^2_{a_\ell \bar{p}, \tau_\bullet}(n)$ in order to get, by (11) and (13),

$$\begin{aligned}\widehat{\mu}_n &\simeq \sum_{\ell=0}^{\bar{\ell}} w_\ell \log(\sigma^2_{a_\ell \bar{p}}) \\ &\simeq \sum_{\ell=0}^{\bar{\ell}} w_\ell \log\left(\sigma^2_H \left(\tau_0 a_\ell \bar{p}\right)^\mu K(H)\right) = \mu,\end{aligned}$$

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

- compute the modified Allan variance by (6) for integer values $a_\ell \bar{p}$, with $1 \le a_\ell \bar{p} \le p_{\max}(n) = \lfloor n/3 \rfloor$;
- compute the weighted MAVAR log-regression estimator by (14) in order to get an estimate $\widehat{\mu}$ of μ;
- estimate H by $\widehat{H} = (\widehat{\mu} + 2)/2$.

Apart from the general notation, that gives freedom in choosing the increasing sequence $\{a_\ell\}_{\ell \in \mathbb{N}}$ and the weights vector $\bar{w}$, the above procedure corresponds to that proposed in [8], [9] and is analogous to others based on log-regression estimations.

IV. Asymptotic Normality of the Estimator

In [12], under the assumption that X is a *fractional Brownian motion with Hurst index* $H \in (1/2, 1)$, two convergence results are proven in order to justify the above approximations and to get the rate of convergence of $\widehat{\mu}_n$ toward $\mu = 2H - 2$.

In particular, it is shown that if $\bar{p} = \bar{p}(n)$ is a sequence of integers such that $\bar{p}(n) \to +\infty$, $n\bar{p}(n)^{-1} \to +\infty$ and $n\bar{p}(n)^{-3} \to 0$ as $n \to +\infty$, then (for a fixed $\bar{\ell}$)

$$\rho_n(\underline{w}, H)^{-1} \left(\widehat{\mu}_n - \mu\right) \xrightarrow[n \to +\infty]{d} \mathcal{N}(0, 1) \qquad (15)$$

with

$$\rho_n(\underline{w}, H) \sim c(\underline{w}, H) \sqrt{\frac{\bar{p}}{n}} \xrightarrow[n \to +\infty]{} 0,$$

where $c(\underline{w}, H)$ is a suitable constant depending on $\underline{w}$ and H. Two important consequences of (15) are the following:

1) The MAVAR log-regression estimator is *consistent*, i.e. the bias $(\widehat{\mu}_n - \mu)$ converges in probability to zero.
2) Given an estimate of H, say $\widehat{H}$, and the corresponding estimate of the normalizing coefficients $\rho_n(\underline{w}, \widehat{H})$, we get the *asymptotic confidence interval* for the parameter H:

$$\widehat{H} - q_{1-\beta/2} \frac{\rho_n(\underline{w}, \widehat{H})}{2} \le H \le \widehat{H} + q_{1-\beta/2} \frac{\rho_n(\underline{w}, \widehat{H})}{2} \qquad (16)$$

where $q_{1-\beta/2}$ is the $(1 - \beta/2)$-quantile of the standard normal distribution. The length of the confidence interval is $q_{1-\beta/2} \rho_n(\underline{w}, \widehat{H})$.

The coefficient $\rho^2_n(\underline{w}, H)$ can be approximated by the following quantity (see [12])

$$\begin{aligned}&\frac{1}{K(H)^2} \sum_{\ell=0}^{\bar{\ell}} \sum_{\ell'=0}^{\bar{\ell}} \left(\frac{a_{\ell \vee \ell'}}{a_{\ell \wedge \ell'}}\right)^{2H+2} \frac{w_\ell w_{\ell'}}{n_\ell n_{\ell'}} \times \\ &\times \sum_{k=0}^{n_{\ell \vee \ell'} - 1} \sum_{k'=0}^{n_{\ell \wedge \ell'} - 1} \Phi_H\left(\frac{k - k'}{\bar{p}\, a_{\ell \vee \ell'}}, \frac{|a_\ell - a_{\ell'}|}{a_{\ell \vee \ell'}}\right),\end{aligned} \qquad (17)$$

with $K(H)$ given in (12), $\ell \vee \ell' = \max\{\ell, \ell'\}$, $\ell \wedge \ell' = \min\{\ell, \ell'\}$, $n_\ell := n - 3a_\ell \bar{p} + 1$ and

$$\begin{aligned}\Phi_H(x, y) := \int_{\mathbb{R}} &\left[\gamma_H(x, 0, r)\gamma_H(0, y, r) \times \right. \\ &\left. \times \int_{-\infty}^{r} \gamma_H(x, 0, v)\gamma_H(0, y, v)\, dv\right] dr.\end{aligned}$$

with

$$\begin{aligned}&\gamma_H(x, y, v) := \left[(H + 1/2)\Gamma(H + 1/2)\sigma_H\right]^{-1} \\ &\left\{\left[(x + 3(1 - y) - v)^+\right]^{H+1/2} - 3\left[(x + 2(1 - y) - v)^+\right]^{H+1/2} + \right. \\ &\left. +3\left[(x + (1 - y) - v)^+\right]^{H+1/2} - \left[(x - v)^+\right]^{H+1/2}\right\}\end{aligned}$$

V. Weights

The explicit expression of the weights clearly depends on the sequence $\{a_\ell\}$. In our investigation we considered the linear progression, $a_\ell = 1 + \ell$, and the geometrical progression, $a_\ell = r^\ell$ with $r > 1$. The latter has provided the better numerical results which are then presented in the next section. Here we focus on the geometrical progression sequence and we give explicit formulas for three particular weights vectors, with increasing complexity, as proposed in the literature for the log-regression procedure.

- The *simple linear regression (SLR) weights* are defined as

$$w_\ell^{SLR} := \frac{\left(\ell - m(\bar{\ell})\right)}{\log(r) \sum_{\ell=0}^{\bar{\ell}} \left(\ell - m(\bar{\ell})\right)^2}$$

with

$$m(\bar{\ell}) := (\bar{\ell}+1)^{-1} \sum_{\ell=0}^{\bar{\ell}} \ell = \frac{\bar{\ell}}{2}$$

- Following Abry and Veitch [13], we can define the *AV weights*

$$w_\ell^{AV} := \frac{\big(\ell - m(\bar{\ell})\big) r^{-\ell}}{\log(r) \sum_{\ell=0}^{\bar{\ell}} \big(\ell - m(\bar{\ell})\big)^2 r^{-\ell}}$$

where

$$m(\bar{\ell}) := \frac{\sum_{\ell=0}^{\bar{\ell}} \ell\, r^{-\ell}}{\sum_{\ell=0}^{\bar{\ell}} r^{-\ell}}.$$

- Following Fay, Moulines, Roueff and Taqqu [19], we can compute a preliminary estimate, say $\widehat{H}^{(1)}$, of H (for instance, applying the first or the second method) and then use it in order to define the ***FMRT weights***

$$\underline{w}^{FMRT} := \frac{1}{\log(r)} D^{-1} B (B^T D^{-1} B)^{-1} \underline{b}$$

where

$$\underline{b} := \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad B := \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & \bar{\ell} \end{pmatrix}^T,$$

and $D = D(n, \widehat{H}^{(1)})$ is the symmetric matrix with entries (for a generic H)

$$[D(n,H)]_{\ell',\ell} := \frac{r^{(\ell\vee\ell')4H}}{r^{(\ell\wedge\ell')4}} \frac{1}{n_\ell n_{\ell'}} \times \\ \times \sum_{k=0}^{n_{\ell\vee\ell'}-1} \sum_{k'=0}^{n_{\ell\wedge\ell'}-1} \Phi_H\left(\frac{k-k'}{r^{\ell\vee\ell'}\bar{p}}, \frac{|r^\ell - r^{\ell'}|}{r^{\ell\vee\ell'}} \right).$$

VI. Numerical Results

In this section we present some numerical results that provide the (approximated) variance of the MAVAR estimator, $\rho_n^2(\underline{w}, H)$, and the relative confidence intervals for the three different weights listed above (SLR, AV, FMRT). The numerical evaluations have been realized for different choices on the parameters, and the most interesting results are presented, with comments, in the next figures and tables.

The variance of the MAVAR estimator is almost unchanged with H as shown in Fig. 1, thus for our analysis we used a fixed $H = 0.7$ to reduce the parameters space. We have first investigated the behavior of $\rho_n^2(\underline{w}, H)$ as a function of n with p following a geometrical growth $\bar{p}r^\ell$, for $0 \le \ell \le \bar{\ell}$. In order to satisfy the hypotheses which are behind convergence (15), the value of $\bar{p}$ has been chosen as $\bar{p} = \bar{p}(n) = \lfloor n^\delta \rfloor$ with $\delta = 0.35$ (formally, any value of $\delta \in (1/3, 1)$ is admissible, but the best results are obtained for δ close to $1/3$). We used two distinct values for the parameter r of the geometrical progression, $r = 1.1$ and $r = 2$, and for each of them we fixed a value $\bar{\ell}$ with the only restriction that $\bar{p}r^{\bar{\ell}} \le \lfloor n/3 \rfloor$.

In Fig. 2 we plot the results for $r = 1.1$ and $\bar{\ell} = 30$. Each marker is associated to one of the three weights (SLR, AV and FMRT) of the previous section as listed in the legend. The lines with markers show the results obtained by numerical evaluation of (17), while the dashed lines represent the corresponding theoretical asymptotes $n^{\delta-1} = n^{-0.65}$ (see Eq. (15)).

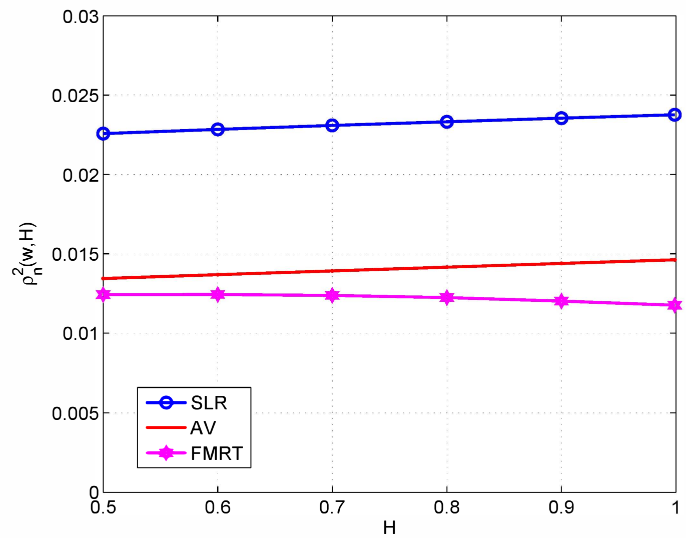

Figure 1. Trend of $\rho_n^2(\underline{w}, H)$ as a function of the Hurst parameter H, with SLR, AV and FMRT weights and geometrical progression $p = \bar{p}r^\ell$, $\bar{p} = \lfloor n^\delta \rfloor$ with $\delta = 0.35$, $r = 2$ and $n = 4096$.

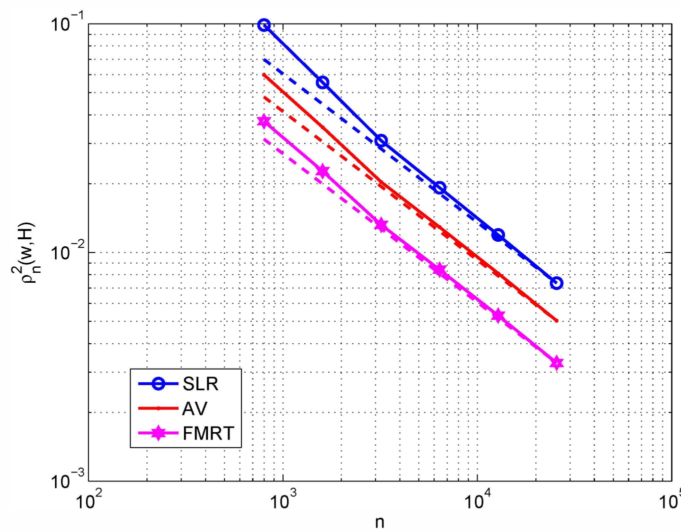

Figure 2. Trend of $\rho_n^2(\underline{w}, H)$ as a function of n, with SLR, AV and FMRT weights and geometrical progression $p = \bar{p}r^\ell$, $\bar{p} = \lfloor n^\delta \rfloor$ with $\delta = 0.35$, $r = 1.1$, $\bar{\ell} = 30$ and $H = 0.7$.

Using the same notation, in Fig. 3 we plot the results for $r = 2$ and taking $\bar{\ell} = 4$.

The two figures show that the approximation formula (17) for the variance $\rho_n^2(\underline{w}, H)$, that we have used for the numerical evaluation, provides results which are in quite good agreement with the theoretical behavior (dashed line), independently on the choice on the weights and on the other parameters. In particular, for n sufficiently large, we get very small values of $\rho_n^2(\underline{w}, H)$ and thus small confidence intervals.

In tables I and II, we list the value of the confidence intervals related to Fig. 2, namely for $r = 1.1$ and, respectively, $\bar{\ell} = 30$

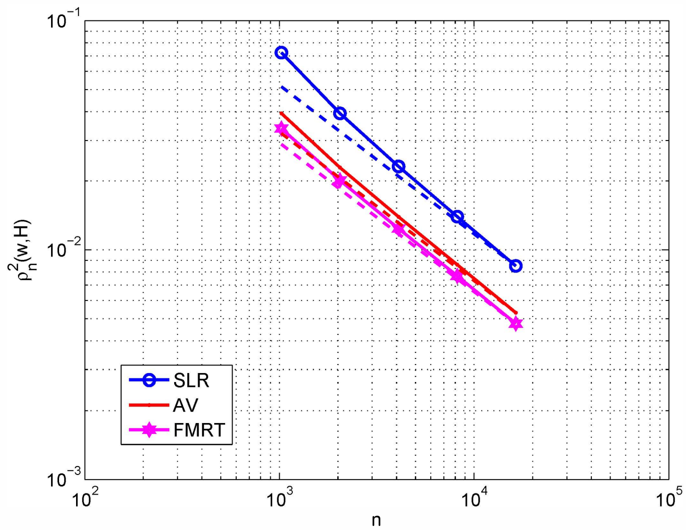

Figure 3. Trend of $\rho_n^2(\underline{w}, H)$ as a function of n, with SLR, AV and FMRT weights and geometrical progression $p = \bar{p}r^\ell$, $\bar{p} = \lfloor n^\delta \rfloor$ with $\delta = 0.35$, $r = 2$, $\bar{\ell} = 4$ and $H = 0.7$.

Table I
LENGTH OF THE CONFIDENCE INTERVAL AS A FUNCTION OF n, WITH SLR, AV AND FMRT WEIGHTS AND GEOMETRICAL PROGRESSION $p = \bar{p}r^\ell$, $\bar{p} = \lfloor n^\delta \rfloor$ WITH $\delta = 0.35$, $r = 1.1$, $\bar{\ell} = 30$ AND $H = 0.7$.

n	3200	4096	6400	12800	25600
$\bar{p}$	16	18	21	27	34
$1.96\rho_n(w_{SLR}, H)$	0.3442	0.3197	0.2716	0.2141	0.1680
$1.96\rho_n(w_{AV}, H)$	0.2796	0.2605	0.2227	0.1765	0.1391
$1.96\rho_n(w_{FMRT}, H)$	0.2253	0.2101	0.1798	0.1427	0.1124

PSfrag replacements
$\bar{\ell}$

and $\bar{\ell} = 45$. In Table III we list the value of the confidence intervals related to Fig. 3, namely for $r = 2$ and $\bar{\ell} = 4$.

Comparing the results displayed in tables I-III, as r and $\bar{\ell}$ vary, we can deduce that

- the best (smallest) value of $\rho_n^2(\underline{w}, H)$ is obtained at $r = 2$ for the SLR and AV weights, and at $r = 1.1$ for the FMRT weight;
- the value of $\rho_n^2(\underline{w}, H)$ is also sensitive of $\bar{\ell}$ and in particular, for $r = 1.1$, the value $\bar{\ell} = 45$ provides better results for the AV and FMRT weights, while $\bar{\ell} = 30$ provides better results for the SLR weights.

Table II
LENGTH OF THE CONFIDENCE INTERVAL AS A FUNCTION OF n, WITH SLR, AV AND FMRT WEIGHTS AND GEOMETRICAL PROGRESSION $p = \bar{p}r^\ell$, $\bar{p} = \lfloor n^\delta \rfloor$ WITH $\delta = 0.35$, $r = 1.1$, $\bar{\ell} = 45$ AND $H = 0.7$.

n	4096	8192	16384
$\bar{p}$	18	23	29
$1.96\rho_n(w_{SLR}, H)$	0.4601	0.3263	0.2425
$1.96\rho_n(w_{AV}, H)$	0.2562	0.1955	0.1509
$1.96\rho_n(w_{FMRT}, H)$	0.2005	0.1564	0.1220

Table III
LENGTH OF THE CONFIDENCE INTERVAL AS A FUNCTION OF n, WITH SLR, AV AND FMRT WEIGHTS AND GEOMETRICAL PROGRESSION $p = \bar{p}r^\ell$, $\bar{p} = \lfloor n^\delta \rfloor$ WITH $\delta = 0.35$, $r = 2$, $\bar{\ell} = 4$ AND $H = 0.7$.

n	4096	8192	16384
$\bar{p}$	18	23	29
$1.96\rho_n(w_{SLR}, H)$	0.2978	0.2315	0.1807
$1.96\rho_n(w_{AV}, H)$	0.2312	0.1818	0.1429
$1.96\rho_n(w_{FMRT}, H)$	0.2180	0.1718	0.1352

This last point suggests us to investigate the effect of an increase of $\bar{\ell}$ over $\rho_n^2(\underline{w}, H)$. We thus evaluate the variance as a function of $\bar{\ell}$, taking fixed $n = 4096$, $H = 0.7$, and $r = 1.1$. The trend is shown in Fig. 4 and Table IV lists the related confidence intervals. As one can see, with SLR weights there exists an optimal choice of $\bar{\ell}$ (approximately 20 in this setting), with the AV weights $\rho_n^2(\underline{w}, H)$ decreases until $\bar{\ell} = 40$ and then slightly increases, while with the FMRT weights it keeps decreasing with $\bar{\ell}$.

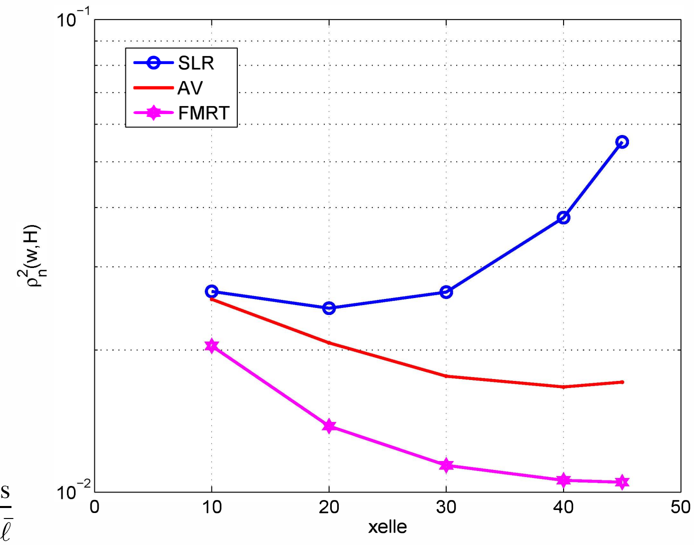

Figure 4. Trend of $\rho_n^2(\underline{w}, H)$ as a function of $\bar{\ell}$, with SLR, AV and FMRT weights and geometrical progression $p = \bar{p}r^\ell$. Here $n = 4096$, $\bar{p} = 18$, $r = 1.1$ and $H = 0.7$.

Table IV
LENGTH OF THE CONFIDENCE INTERVAL AS A FUNCTION OF $\bar{\ell}$, WITH SLR, AV AND FMRT WEIGHTS AND GEOMETRICAL PROGRESSION $p = \bar{p}r^\ell$, $\bar{p} = 18$, $n = 4096$, $r = 1.1$ AND $H = 0.7$.

$\bar{\ell}$	10	20	30	40	45
$1.96\rho_n(w_{SLR}, H)$	0.3197	0.3068	0.3191	0.3826	0.4605
$1.96\rho_n(w_{AV}, H)$	0.3136	0.2820	0.2600	0.2533	0.2568
$1.96\rho_n(w_{FMRT}, H)$	0.2799	0.2302	0.2093	0.2018	0.2011

VII. CONCLUSION

In this paper, we analyzed the behavior of the Modified Allan Variance (MAVAR) in estimating the Hurst parameter H of LRD traffic series. We have first provided a new representation of the MAVAR log-regression estimator that allows to put it in connection with the wavelet log-regression estimator, and to stress the analogies and the differences between the two.

Under the assumption that the signal process is a fractional Brownian motion, the asymptotic analysis given in [12] applies and the MAVAR log-regression estimator turns out to be consistent and asymptotically normal distributed. Here we have provided, under the same hypotheses, new explicit expressions for the normalizing coefficients and detected their asymptotic behavior. These expressions have been used to obtain an explicit formula for the confidence intervals of the estimator.

All these formulas have been computed numerically taking into account different values parameters, such as the size of the traffic series, the value of the Hurst parameter, as well as the weight coefficients of the regression procedure. In particular we have considered three different regression weights commonly proposed in the literature, and compared the related estimator performance as the other parameters varies.

The numerical results show, on one hand, that the predicted asymptotic behavior provides an accurate approximation of the behavior of the estimator at finite sample. On the other hand, they provide a reference for the confidence intervals of the MAVAR log-regression estimator, as the length of the time series under consideration varies.

References

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

[2] V. Paxson and S. Floyd, "Wide-area traffic: the failure of Poisson modeling," *IEEE/ACM Trans. Networking*, Vol. 3, No. 6, pp. 226-244, June 1995.

[3] K. Park, and W. Willinger "Self-similar network traffic: an overview," *Park, K., Willinger, W. (Eds.), Self-Similar Network Traffic and Performance Evaluation*, 1–38, *Wiley (Interscience Division), New York*, 2000.

[4] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Rev.*, Vol. 10, pp. 422-437, Month? 1968.

[5] D. W. Allan, and J. A. Barnes, "A Modified Allan Variance with Increased Oscillator Characterization Ability," *Proc. 35th Annual Frequency Control Symposium*, Philadelphia, PA, 1981.

[6] L. G. Bernier, "Theoretical Analysis of the Modified Allan Variance," *Proc. 41st Annual Frequency Control Symposium*, Philadelphia, PA, 1987.

[7] S. Bregni, "Characterization and Modelling of Clocks," in *Synchronization of Digital Telecommunications Networks*, John Wiley & Sons, Chichester (UK), 2002.

[8] S. Bregni, and L. Primerano, "The modified Allan variance as time-domain analysis tool for estimating the Hurst parameter of long-range dependent traffic," *Proc. IEEE GLOBECOM*, Dallas, TX, Nov. 2004.

[9] S. Bregni, L. Jmoda, "Accurate Estimation of the Hurst Parameter of Long-Range Dependent Traffic Using Modified Allan and Hadamard Variances," *IEEE Trans. on Commun.*, Vol. 56, No. 11, 1900–1906, Nov. 2008.

[10] S. Bregni, and W. Erangoli, "Fractional noise in experimental measurements of IP traffic in a metropolitan area network," *Proc. IEEE GLOBECOM*, St. Louis, MO, Nov. 2005.

[11] S. Bregni, R. Cioffi, and M. Decina, "An Empirical Study on Time-Correlation of GSM Telephone Traffic," *IEEE Trans. on Wireless Commun.*, Vol. 7, No. 9, Sept. 2008, pp. 3428-3435.

[12] A. Bianchi, M. Campanino and I. Crimaldi, "Asymptotic normality of a Hurst parameter estimator based on the modified Allan variance." Submitted to *International Journal of Stochastic Analysis*, focus issue on *Fractional Brownian Motion.* Electronic preprint (2011) is available at http://amsacta.cib.unibo.it/3140/

[13] P. Abry, D. Veitch, "Wavelet analysis of long-range-dependent traffic," *IEEE Trans. Inf. Theory*, Vol. 44, No. 1, pp. 2–15, Jan. 1998.

[14] 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, Apr. 2007.

[15] 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," *J. Time Ser. Anal.*, Vol. 28, No. 2, pp. 155–187, Feb. 2007.

[16] E. Moulines, F. Roueff and M. S. Taqqu, "A wavelet Whittle estimator of the memory parameter of a nonstationary Gaussian time series," *Ann. Statist.*, Vol. 36, No. 4, pp. 1925–1956, Apr. 2008.

[17] A. M. Yaglom, "Correlation theory of stationary and related random functions. Vol. I. Basic results," *Springer Series in Statistics*. Springer-Verlag, New York, 1987.

[18] ITU T Rec. G.810 *Definitions and Terminology for Synchronisation Networks*, Rec. G.811 *Timing Characteristics of Primary Reference Clocks*, Rec. G.812 *Timing Requirements of Slave Clocks Suitable for Use as Node Clocks in Synchronization Networks*, Rec. G.813 *Timing Characteristics of SDH Equipment Slave Clocks (SEC)*, Geneva 1996-2003.

[19] G. Faÿ, E. Moulines, F. Roueff and M. S. Taqqu, "Estimators of long-memory: Fourier versus wavelets," *J. Econometrics*, Vol. 151, No. 2, pp. 159–177, Feb. 2009.