Semi-parametric estimation of the long-range dependence parameter : A survey

Jean-Marc Bardet

Gabriel Lang

Georges Oppenheim,

Anne Philippe

Stilian Stoev

Murad S. Taqqu

ABSTRACT In semi-parametric models, the long-range dependence parameter is estimated without assuming that the short-range dependence structure of the covariance function is known. We review some of the estimation methods and present new ones which have not been tested before on simulated data.

1 Introduction

We present methods of estimation for the long-range dependence parameter. The statistician may make different hypotheses on the underlying population model. He has three basic choices. He may consider a parametric model, where the parameter H of long-range dependence is one of several parameters that determine the model. In this case, classical methods of estimation apply; for instance, Fox and Taqqu [11] and Avram [3] proved that the Whittle contrast estimator has a $\sqrt{N}$ rate of convergence where N is the number of data, when the observed process is Gaussian. Giraitis and Surgailis [13] proved the same kind of results for a class of linear processes. The Whittle estimator is asymptotically efficient and fast to compute. Some Monte-Carlo experiments indicate that these methods work well in practi-

AMS Subject classification: Primary 60G18, 62G05; secondary 62E25.

Keywords and phrases: fractional Gaussian noise, self-similarity, time series, FARIMA, estimation

cal situations (see Taqqu and Teverovsky [26]). In the second instance, the statistician, may want to use a purely graphical technique, assuming only asymptotic self-similarity (see [25]). The estimation methods typically involve regressing the logarithm of various sampling statistics, and often in practice, the region where the regression is performed is chosen after looking at related plots of the data. Some are more effective than others. The often used Rescaled Range method (R/S), for instance, is particularly poor. These graphical methods were studied and compared in Taqqu *et al.* [27] and Taqqu and Teverovsky [26]. We focus here on a third intermediate situation, by considering semi-parametric models, where long-range dependence is described by the parameter H but the structure of the covariance function at short ranges is not specified. In this case, a theory can be developed, although in practice, the estimation often involves regression as well. The goal is to estimate H, either with or without estimating the short-range dependence structure of the covariance function. Some of these methods, which are also described in Taqqu et al. [27, 26], are recalled here and used for comparison; new additional methods which have not been tested before on simulated data sets are also included.

In Section 2, we describe each method, list the main references, and indicate briefly known theoretical results, the order of the complexity of the algorithm and the semi-parametric models that are covered. In Section 3, we present and comment on Monte-Carlo results. We use the same data sets as in Taqqu and Teverovsky [26], to facilitate the comparison.

Our basic notation is as follows: $X = \{X_k, k \in \mathbb{N}\}$ denotes a long memory stationary centered time series with an autocovariance function $r(k)$, such that $\sum |r(k)| = \infty$. The spectral density is denoted by g, N is the number of discrete observations of X, and the Fourier frequencies we consider are $\lambda_j = 2\pi j/N,\ j = 1, 2, \ldots, N$.

2 Semi-parametric estimators

We assume that the stationary process X has spectral density which satisfies the following condition:

$$g(\lambda) = g_*(\lambda)|1 - \exp(-i\lambda)|^{1-2H} \tag{2.1}$$

with $1/2 < H < 1$. The function g_* is continuous and not vanishing at 0. When necessary, additional non-parametric assumptions on g_* will be made, for instance about its regularity. The process X is long-range dependent because $g(\lambda) \sim g_*(0)|\lambda|^{1-2H}$, as $\lambda \to 0$. The goal is to estimate H.

2.1 Time domain methods

We do not present here several methods which are covered in Taqqu and Teverovsky [26]. We also do not present methods which involve counting the number of times the sample path crosses a fixed level. Theoretical results on their convergence are not available and the simulations we performed indicate that they possess a large bias as well as a large variance.

Wavelet analysis

The wavelet transformation is used to detect features of a signal at different scales and locations. Thus it is a natural tool in the study of processes with scaling properties (for example, self-similar or long-range dependent stochastic processes, which are asymptotically self similar). The analysis of long-range dependent processes was initiated by Flandrin [10] and developed further by Abry et al. [2] and Veitch and Abry [28]. A survey of the above papers is presented in Abry et al. [1].

We shall briefly review some recent asymptotic results concerning the wavelet estimators of Gaussian long-range dependent processes (for more details, see Bardet et al. [6]). Let $X = \{X_t\}_{t\in\mathbb{R}}$ be a stationary $\mathbb{L}^2$-process. The wavelet coefficient of X for scale $n > 0$ and location $k \in \mathbb{N}$ is defined by

$$c_X(n,k) = \frac{1}{\sqrt{n}} \int_{\mathbb{R}} X_t \psi\left(\frac{t}{n} - k\right) dt, \tag{2.2}$$

where the integration is considered in the $\mathbb{L}^2$-sense. The function ψ is called the mother wavelet and it is assumed to have $Q (\geq 1)$ zero moments, that is,

$$\int_{\mathbb{R}} t^m \psi(t) dt = 0 \quad \text{for} \quad m = 0, 1, \dots, Q-1 \quad \text{and} \quad \int_{\mathbb{R}} t^Q \psi(t) dt \neq 0.$$

In applications one typically uses wavelets ψ with compact support or with fast decay (e.g. Daubechies, Mexican Hat or Meyer wavelets).

The wavelet coefficients $c_X(n,k)$ have the following properties:

1. If $\{X_t\}_{t\in\mathbb{R}}$ is stationary, so is $\{c_X(n,k)\}_{k\in\mathbb{N}}$ for any $n > 0$.
2. If $\{X_t\}_{t\in\mathbb{R}}$ is Gaussian, so is $\{c_X(n,k),\ n > 0,\ k \in \mathbb{N}\}$.
3. If $\{X_t\}_{t\in\mathbb{R}}$ is a long-range dependent process which satisfies (2.1), then the zero moment property of the wavelet ψ implies that $\{c_X(n,k)\}_{k\in\mathbb{N}}$ is a short-range dependent process, for any scale $n > 0$.

4. If $\{X_t\}_{t\in\mathbb{R}}$ is a Gaussian process which satisfies (2.1) then under additional regularity conditions on $g_*(\lambda)$ one has

$$\mathbb{E}c_X(n,k)^2 \sim C_{\psi,H,g_*} n^{2H-1}, \quad \text{as} \quad n \to \infty \quad \text{for all} \quad k \in \mathbb{N}, \tag{2.3}$$

where $C_{\psi,H,g_*} > 0$ and where $f_1(t) \sim f_2(t)$, as $t \to \infty$ means that $f_1(t)/f_2(t) \to 1$, as $t \to \infty$ (one of the regularity conditions on g_* is $g_*(\lambda) - g_*(0) \sim \mathrm{const}|\lambda|^\beta$, as $\lambda \to 0$, with $\beta > 0$; see Bardet et al. [6] for details).

We shall suppose that X satisfies the assumptions listed in properties $1-4$. These properties suggest using $c_X(n,k)$ to estimate the parameter H, as follows. Assume that the process X_t is observed for all continuous times $t \in [0,N]$. Consider a collection of scales $n_j > 0$, $j = 1,\ldots,J$, and let $S_N(n_j) = \frac{1}{N_j}\sum_{k=1}^{N_j} c_X(n_j,k)^2$ be the empirical mean of $c_X(n_j,\cdot)^2$, where $N_j = [N/n_j]$ is the number of available* coefficients on scale n_j. Then $S_N(n_j)$ is a consistent estimator of $\mathbb{E}c_X(n_j,\cdot)^2$, for fixed n_j as $N \to \infty$ (see Bardet et al [6]), and using (2.3) one obtains the linear model

$$\log(S_N(n_j)) = (2H-1)\log n_j + K_{\psi,H,g_*} + \epsilon(n_j,N),$$

where $K_{\psi,H,g_*} \in \mathbb{R}$ and $\epsilon(n_j,N) \xrightarrow{P} 0$, as both n_j and N/n_j converge to infinity. Then a linear regression of $(\log(S_N(n_j,k))_{j=1}^J$ against $(\log(n_j))_{j=1}^J$ provides an estimator $\widehat{H}(N)$ of H. This estimator is consistent in the sense that $\widehat{H}(N) \xrightarrow{P} H$, as both the scales n_j and the number of wavelet coefficients $[N/n_j]$ on scale n_j tend to infinity.

Remarks.

1. The greater Q, the "closer" the process $\{c_X(n,k),\ k \in \mathbb{N}\},\ n > 0$ to white noise and (theoretically) the better the estimators $\widehat{H}(N)$. In fact,
$$\mathbb{E}c_X(n,k)c_X(n,k') = \mathcal{O}(|k-k'|^{2(H-Q)-2}).$$
(A precise expansion of $\mathbb{E}c_X(n,k)c_X(n',k')$ can be found in Bardet et al. [6].) However, empirical studies have shown that the asymptotic variance of the estimator $\widehat{H}(N)$ seems to increase as Q increases.

2. In practice one uses Mallat's fast pyramidal algorithm to compute the wavelet coefficients $c_X(n_j,k)$. This imposes further restrictions on the mother wavelet ψ (i.e. orthogonality) and on the scales n_j (i.e.

---

*In practice, if the support of the wavelet ψ is (essentially) compact one can compute roughly $N_j = [N/n_j]$ coefficients $c_X(n_j,k)$ on the scale n_j (see (2.2). Thus it is natural to assume that $N_j = [N/n_j]$.

they should be dyadic), however theoretically, only conditions on the regularity and the number of zero moments of ψ are required in the above estimation framework.

• **Convergence rate and level of the asymptotic:** To obtain consistent and asymptotically normal estimators of H (see (2.3)) one needs to take scales n_j which grow to infinity together with the number N_j of coefficients $c_X(n_j, k)$. More precisely, as N tends to infinity, we keep J fixed and let n_j converge to infinity at the same rate for all $j = 1, \ldots, J$ (i.e. if $n_1 < n_2 < \ldots < n_J$, then n_J/n_1 bounded). Then for all $j = 1, \ldots, J$, we have $S_N(n_j)/\mathbb{E}c_X(n_j, 0)^2 \xrightarrow{P} 1$, as $N_j = [N/n_j] \to \infty$. Moreover, if $n_1 = n_1(N)$ satisfies the conditions $N_1 = [N/n_1] \to \infty$ and $Nn_1^{-(1+2\beta)} \to 0$, as $N \to \infty$, one obtains a central limit theorem for $\widehat{H}(N)$, with a convergence rate which depends on the choice of $n_1(N)$:

$$\sqrt{\frac{N}{n_1(N)}}(\widehat{H}(N) - H) \xrightarrow[N \to \infty]{d} \mathcal{N}(0, \sigma^2).$$

The asymptotic variance depends on H and the wavelet ψ (see Bardet et al. [6]). Asymptotic confidence intervals can be obtained, however they involve the unknown parameter H.

• **Complexity of the algorithm:** The method involves a linear regression on a finite number of terms J. The computational complexity of the wavelet estimators is $\mathcal{O}(N)$, provided that the integration in (2.2) has a complexity $\mathcal{O}(1)$.

• **Restrictions on the model and robustness:** The theoretical results presented here have been developed only for Gaussian processes, satisfying the assumptions listed in properties $1-4$ above (see Bardet et al. [6]). Corresponding results for linear fractional stable motion, a self-similar process with infinite variance, have been obtained by Pipiras, Taqqu and Abry [21].

Quadratic variations

In the previous method of estimation it was assumed that the continuous-time process $X = \{X_t\}_{t\in\mathbb{R}}$ is available, so that the wavelet coefficients $c_X(n, k)$ can be computed exactly. In this subsection we shall describe another method of estimation of H, which is based on the discrete-time observations $X_1, X_2, \ldots, X_N$ of the process X. Let n_j, $j = 1, \ldots, J$ be a collection of integer scales. First, one may consider approximated wavelet coefficients:

$$d_X(n_j, k) = \frac{1}{\sqrt{n_j}} \sum_{i=1}^{N} \psi\left(\frac{i}{n_j} - k\right) X_i. \tag{2.4}$$

This expression is a natural approximation of $c_X(n_j,k)$ by a Riemann sum with a spacing $\Delta_i = 1$. Note, however, that ψ may not have zero moments on the grid $i/n_j - k,\ i = 1,\dots,N$ and hence the coefficients $\{d_X(n_j,k),\ k \in \mathbb{N}\}$ may not therefore be short-range dependent. However, they will become so as n_j (and N) go to infinity.

One may also consider coefficients defined as follows:

$$d_X(n_j,k) = \sum_{i=1}^{p_j} a_i^j X_{n_j(k+i)}, \tag{2.5}$$

where $a^j = \{a_i^j\}_{i=1}^{p_j},\ p_j \in \mathbb{N},\ j = 1,\dots,J$, is a collection of *discrete filters* which have Q zero moments (that is, $\sum_{i=1}^{p_j} i^q a_i^j = 0$, for all $q = 0,\dots,Q-1$ and $\sum_{i=1}^{p_j} i^Q a_i^j \neq 0$). Due to the zero moment property of the filters $a^j,\ j = 1,\dots,J$, the coefficients $d_X(n_j,k)$ and the wavelet coefficients $c_X(n_j,k)$ of the continuous-time process X have similar properties. In fact $\{d_X(n_j,k),\ k \in \mathbb{N}\}$ in (2.5) will be short-range dependent if $Q \geq 1$.

One can replace $c_X(n_j,k)$ by $d_X(n_j,k)$ and use the same regression-based estimation procedure (with the same choice of the scales $n_j,\ j = 1,\dots,J$) as for the wavelet method. The resulting quantities $S_N(n_j) = \frac{1}{N_j}\sum_{k=1}^{N_j} d_X(n_j,k)^2$ are called *generalized quadratic variations* of the process $\{X_k,\ k \in \mathbb{N}\}$ and the corresponding estimators are called generalized quadratic variations estimators. This method (as well as the wavelet method) can be applied to processes with stationary increments such as fractional Brownian motion. The generalized quadratic variations have been introduced in Istas and Lang [17] to estimate the regularity of a Gaussian process with stationary increments and applied also by Bardet [4] to obtain a test for a Gaussian time series to be fractional Brownian motion.

• **Convergence rate and level of the asymptotics:** No proper reference exists on the asymptotics of these estimators in a general *semi-parametric* context, that is under the assumption (2.1). To obtain consistency and asymptotic results one may follow two approaches. First, one can consider $d_X(n_j,k)$, as defined in (2.4), where $X_1, X_2, \dots, X_N$ are discrete-time observations of a continuous-time process $\{X_t,\ t \in \mathbb{R}\}$. This approximation involves a bias which should be precisely evaluated and, for an appropriate asymptotic choice of the scales $n_j = n_j(N)$, it may be negligible. The consistency and asymptotic normality of the wavelet estimators, would then imply the consistency and asymptotic normality of these estimators, as well. A second approach is to use $d_X(n_j,k)$ as defined in (2.5) and consider X as a discrete-time stationary process $\{X_k,\ k \in \mathbb{N}\}$. One can rewrite all conditions of Bardet et al. [6] for the spectral density of X, which will now be a function on $[-\pi,\pi]$ rather than on $\mathbb{R}$. The proofs in Bardet et al. [6], which involve only manipulations with the spectral density could, in principle, be adapted to this case. Thus, consistency and asymptotic normality results of the generalized quadratic variations estimators could be obtained. The asymptotic results in the case of generalized quadratic variations should be exactly the same as the ones in the case of wavelet estimators.

The *parametric* case of fractional Gaussian noise was studied in Bardet [4] (with generalized quadratic variations). The rate of convergence of the estimators in this case is $\sqrt{N}$. Coeurjolly [8] has studied empirically the effect of Q on the asymptotic variance in the special case when X is a fractional Gaussian noise. The conclusion in this case is that $Q=1$ and $Q=0$ is asymptotically optimal when $H \geq 0.65$ and $H < 0.65$, respectively.

• **Complexity of the algorithm :** The complexity of the computations is the same as that of the wavelet estimators, that is $\mathcal{O}(N)$.

•**Restrictions on the model and robustness:** Theoretical results have been developed for fractional Brownian motion (Bardet [4, 5]) and linear fractional stable motion (Stoev, Pipiras and Taqqu [24]).

2.2 Spectral methods

For more details on this section, we refer to the general presentation of spectral methods in Moulines and Soulier [20].

Local Whittle estimator

Under the hypothesis (2.1), the Whittle estimation involves minimizing the contrast:

$$U_N(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \log g(\lambda) + \frac{I_N(\lambda)}{g(\lambda)} d\lambda.$$

where $I_N(\lambda)$ is the periodogram of $(X_1, .., X_N)$:

$$I_N(\lambda) = \frac{1}{2\pi N} \sum_{j=1}^{N} \left| X_j e^{-ij\lambda} \right|^2. \tag{2.6}$$

After renormalizing the spectral density $g(\lambda)$ so that $\int_{-\pi}^{\pi} \log g(\lambda) d\lambda = 0$, the estimator is defined by $(\hat{H}(N), \hat{g}_*(N)) = \mathrm{Argmin}_{(H,g_*)} U_N(\theta)$. Since the estimation of the function g_* is not of interest here, the Whittle contrast will be computed in the neighborhood of 0, where the function g_* may be considered as the constant $g_*(0)$. One then obtains the *local Whittle estimator* propoed by Robinson [23] and detailed by Beran [7]. The integral of the Whittle contrast is replaced by a discrete sum, which is easily computed from the discrete observations. This sum is computed over a range with a maximal frequency that tends to zero as N tends to infinity. Let m be an integer depending on N and denote the limit of $g(\lambda)\lambda^{2H-1}$, as $\lambda \to 0$, by the constant $G = g_*(0)$. The local contrast is

$$U_N(H,G) = \frac{1}{m}\sum_{j=1}^{m} \log G\lambda_j^{1-2H} + \frac{I_N(\lambda_j)}{G\lambda_j^{1-2H}},$$

and the estimator is defined by $\hat{\theta}(N) = \mathrm{Argmin}_{(H,G)\in\Theta} U_N(H,G)$, where $\Theta = [H_1, H_2] \times (0, +\infty)$.

• **Complexity of the algorithm :** The computation of U_n is $\mathcal{O}(m \log N)$. But the minimization step is of unknown complexity.

• **Convergence rate and level of the asymptotic :** The estimator is consistent if g_* is differentiable in an interval $(0, \delta)$, its derivative is $O(\lambda^{-1})$ in the neighborhood of 0, m tends to infinity with N and is negligible with respect to N. Under a stronger assumption on the differentiability of g_* in the neighborhood of 0, and a corresponding restriction on the rate of m, the estimator of H is asymptotically normal with rate $\sqrt{m}$ and asymptotic variance $1/4$ (for details, refer to Robinson [23]).

• **Restrictions on the model and robustness :** The model is Gaussian, or linear with "innovations" which are martingale differences satisfying a condition of uniform integrability.

Local log-periodogram estimators

Geweke and Porter-Hudak [12] propose to estimate the exponent of the spectral density g satisfying (2.1) by a log-regression on the periodogram computed at the first $m(N)$ Fourier frequencies (see Hurvich *et al.* [15]):

$$\begin{aligned} \log(I_N(\lambda_j)) &= \log(g_*(0)) + (1-2H)\log(|1-e^{-i\lambda_j}|) \\ &\quad + \log(g_*(\lambda_j)/g_*(0)) + \log(I_N(\lambda_j)/g(\lambda_j)) \qquad (2.7) \end{aligned}$$

• **Convergence rate and level of the asymptotic:** The estimator converges if the parameter m grows like N^α, with $0 < \alpha < 1$. In that case, the Fourier frequencies tend to 0 and so does the third term in (2.7). But the fourth term, which plays the role of the error term in the regression, is not i.i.d. and its expectation depends on j. Geweke and Porter Hudak's proof [12] of asymptotic normality of the estimator, using an approximation by i.i.d. variables, was not convincing. Based on an empirical study, they also indicated that the MSE is minimal for $m = \sqrt{N}$. Hurvich *et al.* [15] show that if g_* is three times differentiable at 0 and if $\alpha < 4/5$, then there is asymptotic normality with rate $\sqrt{m}$ and the asymptotic variance equals $\pi^2/24$. Assuming that $g_*''(0) \neq 0$ at 0, the optimal MSE is $\mathcal{O}(N^{2/5})$, obtained for $m = CN^{4/5}$, where C is an explicit constant, but it depends on the derivatives of g_* at 0, which are unknown in the semi-parametric context. Robinson [22] suggests a modified version of this estimator by discarding the first l Fourier frequencies and averaging the periodogram in blocks of fixed size, or equivalently – Lang and Azaïs [18] – by integrating the periodogram with a kernel vanishing at 0 before applying the log function. Similar results can then be obtained under weaker differentiability conditions on g_*. The optimal MSE in this case depends on the second term of the expansion of g_* at 0.

• **Complexity of the algorithm :** The complexity is that of the ordinary regression problem of size m, say $\mathcal{O}(m)$.

• **Restrictions on the model and robustness :** The models of interest are usually Gaussian, with g_* differentiable at 0 and bounded outside a neighborhood of 0.

Global log-periodogram estimator

The preceding methods where called local, since they use Fourier frequencies in a bandwidth converging to 0. The present method of Moulines and Soulier [19] will be called global, because it uses the whole set of observable Fourier frequencies. The downside is that it is necessary to estimate not only the parameter H of interest but also the "short memory part" that corresponds to the function $\log(g_*)$. This function is expanded in a $\mathbb{L}^2$ basis:

$$\log(g_*) = \sum_{j=0}^{\infty} \theta_j h_j. \tag{2.8}$$

In the FEXP (frequency-exponential) estimation method, the chosen basis is generated by the cosine functions $h_j(\lambda_j) = \cos(j\lambda_j)/\sqrt{\pi}$. The FEXP method involves fixing the order m in the approximation $\sum_{j=0}^{m} \theta_j h_j$ of $\log(g_*)$, and computing the linear regression of the log periodogram with the explanatory variables $\log|1 - e^{i\lambda_j}|$ and $h_j(\lambda_j)$. More precisely, the set of the indices $(1, \dots, N)$ is divided into K blocks of fixed size m, the last indices being dropped. The Fourier frequencies are divided into corresponding blocks. For each of these blocks, the periodogram is averaged (pooling) with a weight pattern (data tapering). The resulting variables are rescaled to have an expectation equal to 1 and the log of the result defines the variables $(Y_{N,k})_{(k=1,\dots,K)}$. An ordinary least square regression is then performed:

$$Y_{N,k} = R(1 - e^{-ix_k})(1 - 2\hat{H}) + \sum_{j=0}^{m} \hat{\theta}_j h_j(x_k) + \epsilon_{N,k}, \ k = 1, \dots, K,$$

where $x_k = (2k+1)\pi/N$. Note that the error terms are neither independent nor centered, and if m is kept fixed, the bias is not asymptotically zero.

• **Convergence rate and level of the asymptotic :** The FEXP estimator is consistent and asymptotically normal if m is of order N^γ, with $1/(2\alpha) < \gamma < 2/3$ (here α is the Hölder index of g_*). The problem of finding the m that yields a minimal MSE is equivalent to the choice of the parameter m in local methods: it can be expressed as a function of the non-observable g_*, defined by an adaptive procedure (see Iouditsky *et al.* [16]), but with a corresponding logarithmic loss in the minimal MSE.

• **Complexity of the algorithm :** The estimator is given in a closed form, as a sum of K terms involving the periodogram; the complexity is the same as the computation of the values of the periodogram, i.e. $\mathcal{O}(N\log(N))$.

• **Restrictions on the model and robustness :** The model is Gaussian, g_* is bounded away from 0 and α-Hölder for an $\alpha > 1$.

3 Simulations and comments

3.1 Description of the simulations

Data sets

The data sets are the same as in Taqqu and Teverovski [26]. The simulated processes are

- fractional Brownian motion, with $H = 0.5,\ 0.6,\ 0.7,\ 0.8$ and 0.9.
- the $(0, d, 0)$-FARIMA process, with $d = 0,\ 0.1,\ 0.2$ and 0.3, with an infinite variance innovation (Pareto distribution with $\alpha = 1.5$).
- two $(1, d, 1)$-FARIMA process, with $d = 0,\ 0.1,\ 0.2,\ 0.3$ and 0.4, with Gaussian innovations. For the first process, $\phi = 0.3$, $\theta = 0.7$ where ϕ denotes the AR coefficient and θ denotes the MA coefficient. For the second process, $\phi = -0.3$, $\theta = -0.7$.

There are 50 independent replications of each time series. The lengths of the series are 500 and 10000.

Estimation of the parameters

In the semi-parametric setting, the function g_* describing the short-range dependence structure of the process is unknown to the practitioner. Most of the estimation methods we have presented introduce a bandwidth or a scale parameter m, that should depend on N. The theoretical results which assume that m is a power of N and specify the exponent, are asymptotic in nature and do not indicate how to best choose the constant of proportionality. Since the constant depends on the function g_*, it cannot be computed in practice, and hence arbitrary values must be given to m. These values are specified below. Note that the global log-periodogram regression is an exception, because its authors developed an adaptive version which computes a 'good' value for m from the data.

- Quadratic variations: The parameter J (number of scales) is equal to 8 for $N = 500$ and 64 for $N = 10,000$. The scales are defined as $n_j = jn_1$, for $j = 1, \ldots, J$ and $n_1 = 3$ for $N = 500$ and $n_1 = 4$ for $N = 10,000$. The function ψ used to compute the coefficients $d_X(n_j, k)$ (see definition (2.4)) is the Haar wavelet (with one zero moment), i.e. $\psi(t) = 1$ for $t \in (0, 1/2]$ and $\psi(t) = -1$, for $t \in (1/2, 1]$.

- Wavelets. The method and chosen parameters are described in [1]. We take: $j_1 = 3$ and $j_2 = 5$ for for $N = 500$, $j_1 = 5$ and $j_2 = 10$ for $N = 10000$. The wavelet is a Daubechies wavelet with 2 vanishing moments of $N = 500$ and 4 vanishing moments of $N = 10000$ (this choice of Q is recommended in [1]).

- Local Whittle: The trimming parameter $m = 20$ for $N = 500$, $m = 300$ for $N = 10000$. This is the recommended value for the $(0, d, 0)$-FARIMA model.

- Local log-periodogram: The trimming parameter is chosen equal to the Geweke-Porter Hudak specification: $m = 100$.

- Global log-periodogram. All parameters are chosen automatically.

3.2 Results

TO THE EDITOR: PLACE FIGURES 1-4 AND TABLES 1.1-1.4 HERE

3.3 Comments

Data set length

A sample path of length 500 is too short. Every estimator has a large standard error (larger than 0.1 when the parameter itself is defined in a range of 0.5). The different methods give equivalent results in the case of the FGN and the FARIMA model with a 'small' short range dependence part (negative ARMA parameters). But in the case of the two FARIMA$(1, d, 1)$ models, the best methods are the local Whittle and the Global log-periodogram. In the following we consider the data sets of length 10000.

Comparisons on long data sets

Because the FGN process and the FARIMA$(0, d, 0)$ correspond to an ideal situation, they serve as reference models for the various methods. The self-similarity of fractional Brownian motion implies that all the methods should work particularly well for them. The local Whittle and Global log-periodogram method appear slightly more effective for all the models considered here.

Robustness

We only investigate how the performance of the estimators is affected by the following variations in the model:

- Internal variations: we consider FARIMA$(1, d, 1)$ with Gaussian innovations, but where the short-range dependence plays a stronger role than in the FGN or FARIMA$(0, d, 0)$ models. The range of short-range dependence for FARIMA$(1, d, 1)$ with negative coefficients is larger than for FARIMA$(1, d, 1)$ with positive coefficients.

- External variation: the FARIMA$(0, d, 0)$ model is non-Gaussian (infinite variance innovations). No theoretical results are available for these methods. The FARIMA models with $d = 0$ are not long-range dependent.

The influence of the short-range dependence structure is greater for the FARIMA$(1, d, 1)$ model with positive coefficients and this could produce a bias in the first three methods. This bias can be alleviated by choosing j_1 large enough (in the quadratic variations and wavelet methods). In our setting $j_1 = 5$ works well when the data size is 10,000. Correspondingly, it is the parameter m that should be adequately chosen in the local Whittle and local log-periodogram methods. The Global log-periodogram method was designed to find a compromise between bias and variance by adapting the parameter m of the regression on the data.

Since replacing the Gaussian innovations by ones that are non-Gaussian with finite variance does not really affect the results (see [26, 1]), we have not done that here. When the innovations are Pareto and hence have infinite variance, the methods appear to continue to work but the variability of the estimator is much greater.

Conclusions

All the methods considered here are quite effective with a slight edge for the Global log-periodogram and local Whittle. The advantage of the Global log-periodogram is that it is adaptive, and does not require the presence of a trimming parameter m. The wavelet and quadratic variations estimators are fast and easy to program and they have good enough performance (except for the wavelet estimator with non-Gaussian series). The local log-periodogram is the worst for the variability, but the trimming parameter has been chosen equal to $\sqrt{N}$ as in Geweke and Porter-Hudak [12]. This is not a good choice and the performance could have improved, had one chosen a better value, as it was done for the local Whittle estimator. But it may not be worth doing so, because the Global log-periodogram estimator does that automatically.

Acknowledgment. Murad S. Taqqu was partially supported by the NSF Grant ANI-9805623 at Boston University.

References

[1] Abry P., Flandrin P., Taqqu M.S. and Veitch D. Self-similarity and long-range dependence through the wavelet lens. In P. Doukhan, G. Oppenheim, and M.S. Taqqu, editors, *Long-range Dependence: Theory and Applications*. Birkhäuser, 2001. Appears in this volume.

[2] Abry P., Veitch D. and Flandrin, P. Long-range dependence: revisiting aggregation with wavelets. *J. Time Ser. Anal.*, 19 (1998), no. 3, 253–266.

[3] Avram F. On bilinear forms in Gaussian random variables and Toeplitz matrices. *Probab. Theory Relat. Fields*, 79(1):37–45, 1988.

[4] Bardet J.-M. Testing for the presence of self-similarity of Gaussian time series having stationary increments. *J. Time Ser. Anal.*, 21:497–516, 2000.

[5] Bardet J.-M. Statistical study of the wavelet analysis of fractional Brownian motion. *Preprint*, 2000.

[6] Bardet J.-M., Lang G., Moulines E. and Soulier Ph. Wavelet estimator of long-range dependent processes. *Statistical Inference for Stochastic Processes*, 3:85–99, 2000

[7] Beran J. Statistical methods for data with long-range dependence. *Stat. Sci.*, 7:404–427, 1992.

[8] Coeurjolly J.-F. Identification du mouvement brownien fractionnaire par variations discrètes. *Preprint LMC*, RR 1016-M:1–33, 1999.

[9] Coeurjolly J.-F. Inférence statistique pour les mouvements browniens fractionnaires et multifractionnaires. *PhD thesis*, Université Joseph Fourier Grenoble 1. Direction A. Antoniadis, 2000.

[10] Flandrin P. On the spectrum of fractional Brownian motions, *IEEE Transactions on Information Theory*, 1989, Vol. IT-35, No. 1, pp. 197-199

[11] Fox R. and Taqqu M.S. Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, 14(2):517–532, 1986.

[12] Geweke J. and Porter-Hudak S. The estimation and application of long-memory time series models. *J. Time Ser. Anal.*, 4:221–238, 1983.

[13] Giraitis L. and Surgailis D. A central limit theorem for quadratic forms in strongly dependent linear variables and its applications to the asymptotic normality of Whittle estimate. *Prob. Th. and Rel. Fields.*, 86:87–104, 1990.

[14] Guyon X. and Leòn J. Convergence en loi des H-variations d'un processus Gaussien stationnaire. *Ann. Inst. Poincaré*, 25:265–282, 1989.

[15] Hurvich C.M., Deo R. and Brodsky J. . The mean square error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series. *J. Time Ser. Anal.*, 19(1):19–46, 1998.

[16] Iouditsky A., Moulines E. and Soulier Ph. Adaptive estimation of the fractional differencing coefficient. *Preprint*, 1998.

[17] Istas J. and Lang G. Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Ann. Inst. Poincaré*, 33:407–436, 1994.

[18] Lang G. and Azaïs J-M. Non parametric estimation of the long range dependence exponent for Gaussian processes. *J.Stat. Pl. Inf.*, 80(1-2):59–80, 1999.

[19] Moulines E. and Soulier Ph. Broadband log-periodogram regression of time series with long-range dependence. *J. Time Ser. Anal.*, 27(4):1415–1439, 1998.

[20] Moulines E. and Soulier Ph. Semiparametric spectral estimation for fractional processes. In P. Doukhan, G. Oppenheim, and M.S. Taqqu, editors, *Long-range Dependence: Theory and Applications*. Birkhäuser, 2001. Appears in this volume.

[21] Pipiras V., Taqqu M.S. and Abry P. On the central limit theorem for some nonlinear functions of symmetric stable moving averages with applications to wavelet-based estimators. *Preprint*, 2001.

[22] Robinson P.M. Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23:1048–1072, 1995.

[23] Robinson P.M. Gaussian semiparametric estimation of long-range dependence. *The Annals of Statistics*, 23(5):1630–1661, 1995.

[24] Stoev S., Pipiras V. and Taqqu M.S. Estimation of the self-similarity parameter in linear fractional stable motion. *Preprint*, 2001.

[25] Taqqu M.S. Fractional Brownian motion and long-range dependence. In P. Doukhan, G. Oppenheim, and M. S. Taqqu, editors, *Long-range Dependence: Theory and Applications*. Birkhäuser, 2001. Appears in this volume.

[26] Taqqu M.S. and Teverovsky V. On estimating the intensity of long-range dependence in finite and infinite variance time series. In R. E. Feldman R. J. Adler and M. S. Taqqu, editors, *A Practical Guide to Heavy Tails*, pages 177–217. Birkhäuser Boston, 1998.

[27] Taqqu M.S., Teverovski V. and Willinger W. Estimators for long-range dependence: an empirical study. *Fractals*, 3(2):785–798, 1995.

[28] D. Veitch and P. Abry, A wavelet-based joint estimator of the parameters of long-range dependence, *IEEE Transactions on Information Theory*, 1999, Vol. 45(3), pp. 878-897

Jean-Marc Bardet, Laboratoire de Statistique et Probabilités, 118 route de Narbonne, 31062 Toulouse, FRANCE, e-mail: `bardet@cict.fr`

Gabriel Lang, Laboratoire GRESE, Ecole Nationale du Genie Rural, des Eaux et des Forets, 19 avenue du Maine, 75732 Paris Cedex 15, FRANCE, e-mail: `lang@engref.fr`

Anne Philippe, Laboratoire de Probabilités et Statistique, bat M2, UFR de mathematiques, Université des Sciences et Technologies, 59655 Villeneuve d'Ascq Cedex, FRANCE, e-mail: `Anne.Philippe@univ-lille1.fr`

Georges Oppenheim, Laboratoire de Modélisation Stochastique et Statistique, Bât 425, Université de Paris Sud, 91 405 Orsay Cedex, FRANCE, e-mail: `Georges.Oppenheim@math.u-psud.fr`

Stilian Stoev, Department of Mathematics, 111 Cummington St., Boston University, Boston, MA 02459, U.S.A., e-mail: `sstoev@math.bu.edu`

Murad S. Taqqu, Department of Mathematics, 111 Cummington St., Boston University, Boston, MA 02459, U.S.A., e-mail: `murad@math.bu.edu`

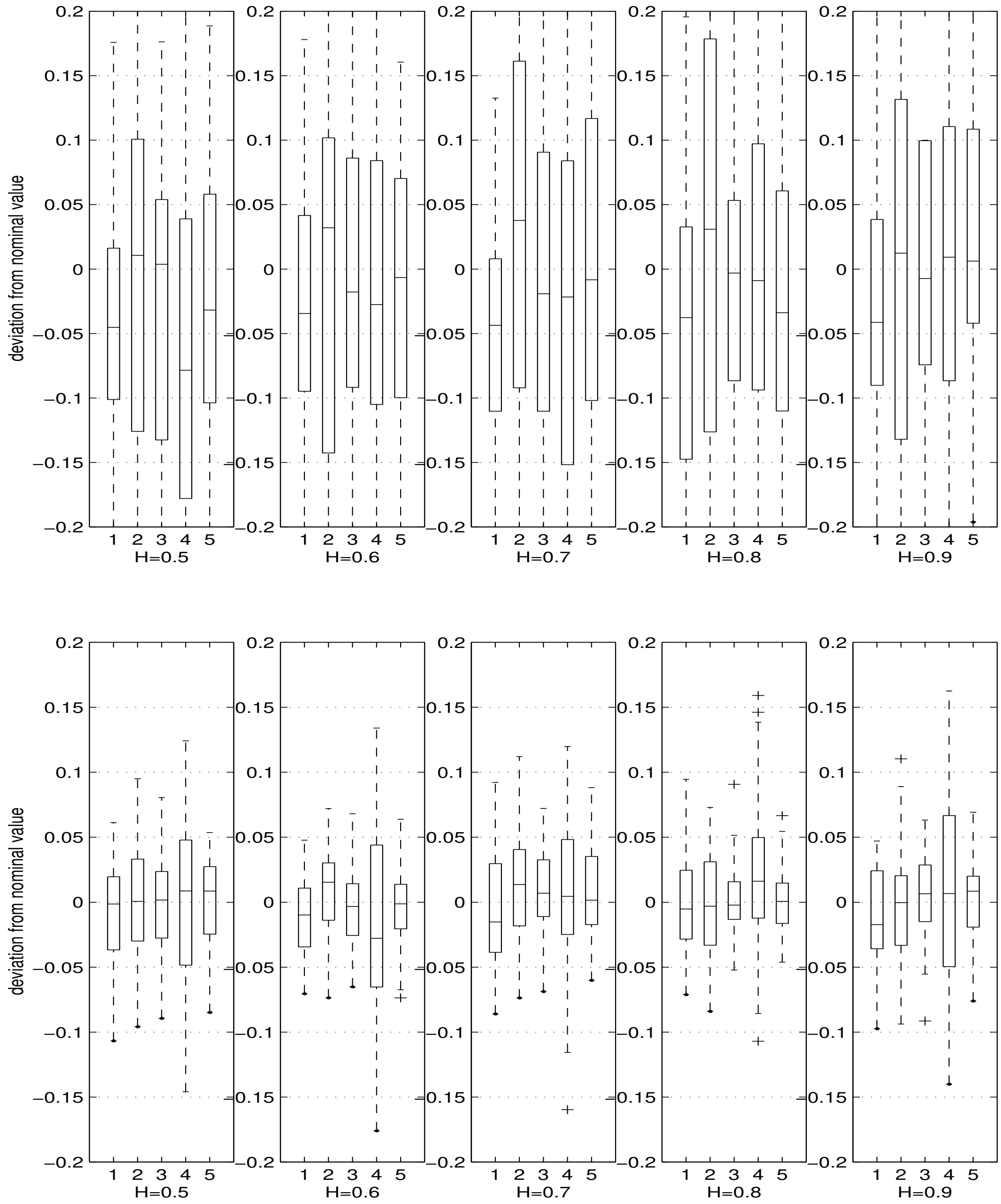

FIGURE 1. Boxplots for Fractional Gaussian noise series of length 500 (up) and 10000 (down). [Estimators] 1: Quadratic variations, 2: Wavelets, 3: Local Whittle, 4: Local log-periodogram, 5: Global log-periodogram

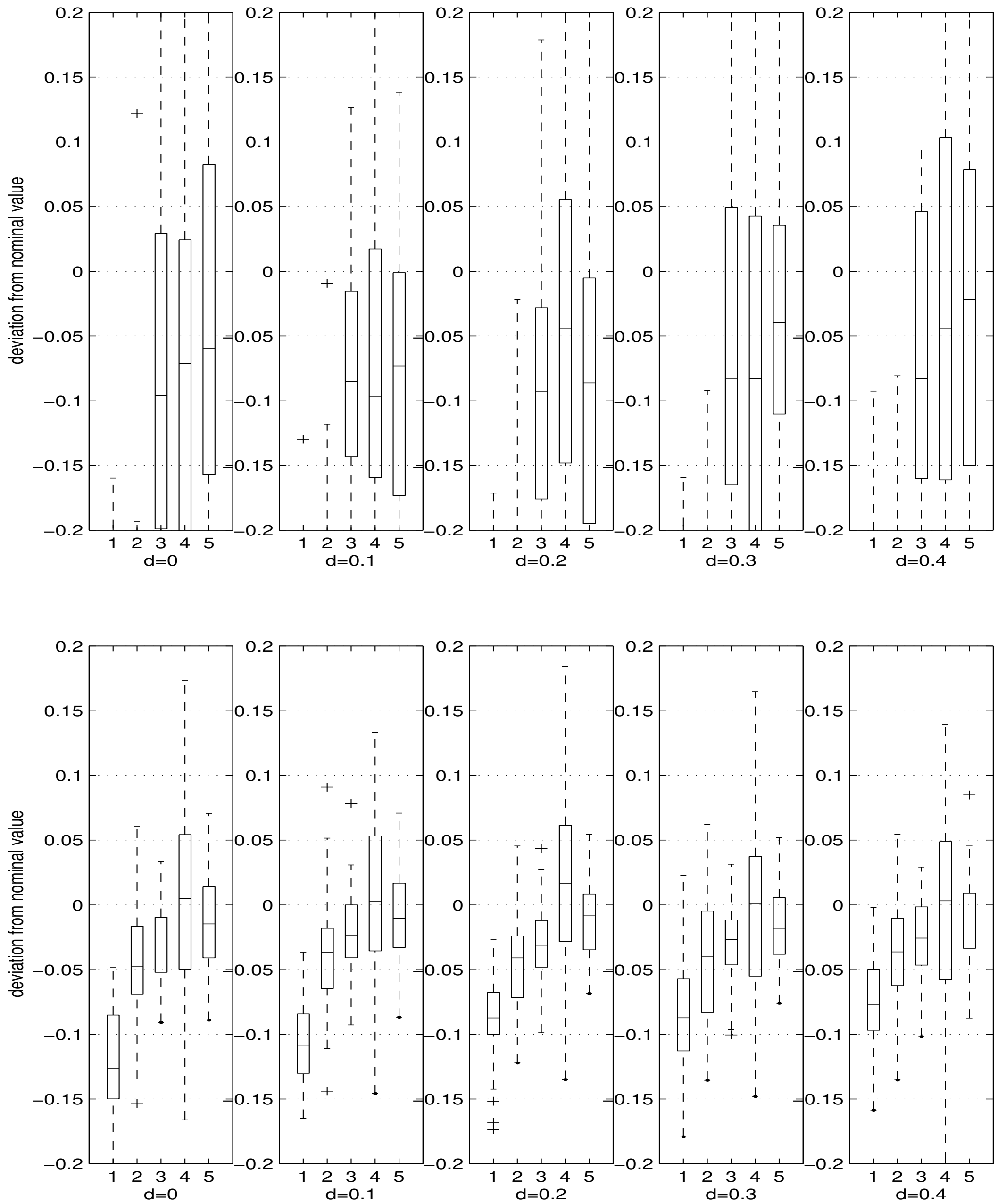

FIGURE 2. Boxplots for FARIMA$(1, d, 1)$ series of length 500 (up) and 10000 (down) with Gaussian innovations : $\phi = 0.3$, $\theta = 0.7$ with ϕ AR coeff. and θ MA coeff. [Estimators] 1: Quadratic variations, 2: Wavelets, 3: Local Whittle, 4: Local log-periodogram, 5: Global log-periodogram

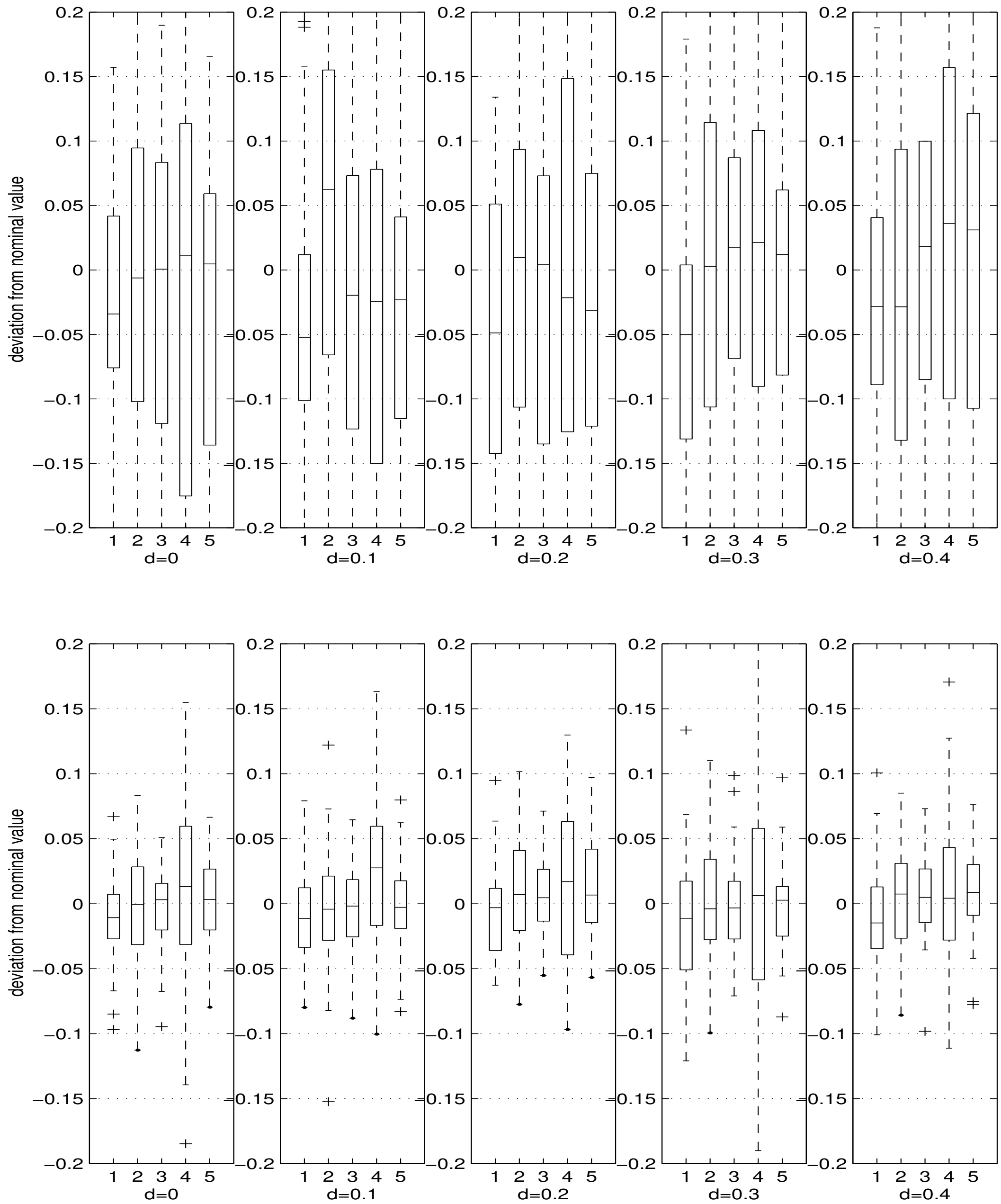

FIGURE 3. Boxplots for FARIMA$(1, d, 1)$ series of length 500 (up) and 10000 (down) with Gaussian innovations : $\phi = -0.3$, $\theta = -0.7$ with ϕ AR coeff. and θ MA coeff.. [Estimators] 1: Quadratic variations, 2: Wavelets, 3: Local Whittle, 4: Local log-periodogram, 5: Global log-periodogram

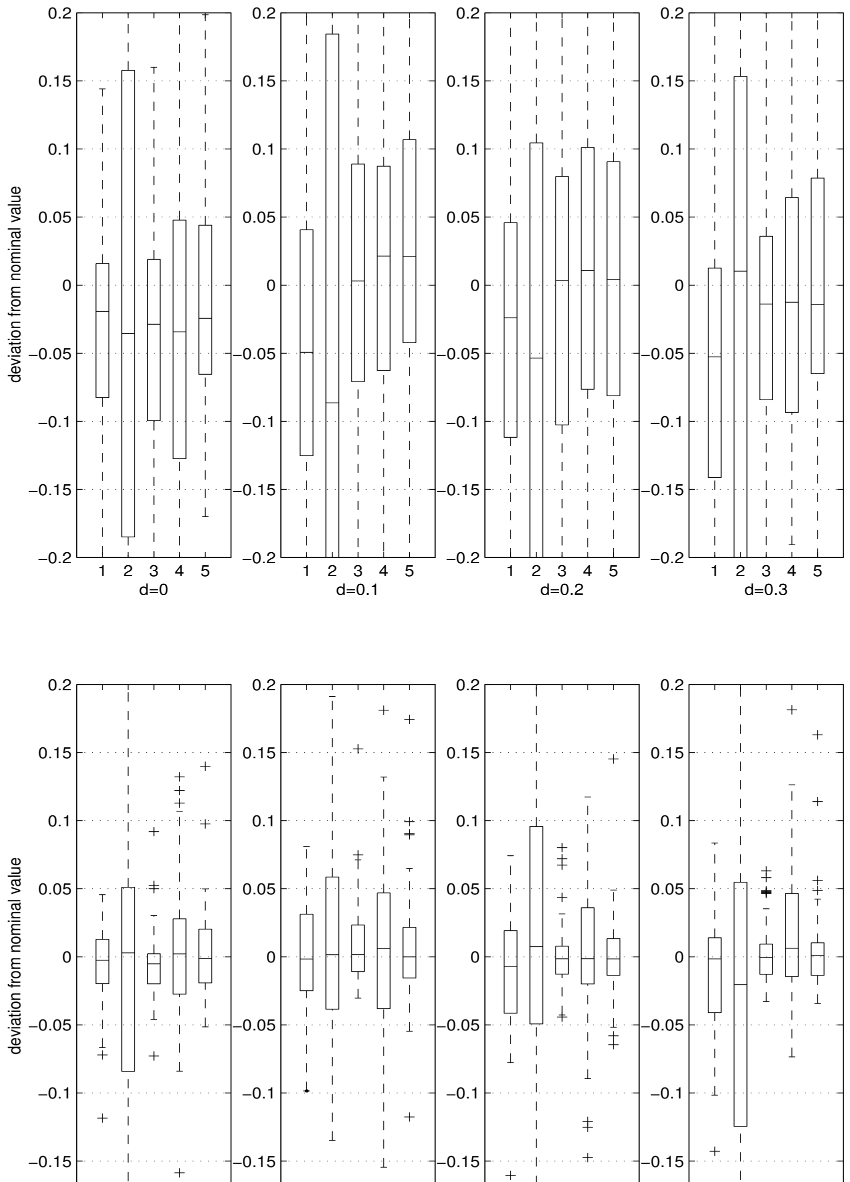

FIGURE 4. Boxplots for FARIMA$(0, d, 0)$ series of length 500 (up) and 10000 (down) with Pareto (infinite variance) innovations (alpha=1.5). [Estimators] 1: Quadratic variations, 2: Wavelets, 3: Local Whittle, 4: Local log-periodogram, 5: Global log-periodogram

	FGN Nominal H				
	0.5	0.6	0.7	0.8	0.9
	QUADRATIC VARIATIONS				
BIAS	-0.0070	-0.0119	-0.0064	-0.0037	-0.0132
$\hat{\sigma}$	0.0360	0.0315	0.0405	0.0357	0.0381
$\sqrt{MSE}$	0.0363	0.0334	0.0406	0.0355	0.0400
	WAVELETS				
BIAS	0.0029	0.0095	0.0103	-0.0031	-0.0007
$\hat{\sigma}$	0.0454	0.0329	0.0445	0.0390	0.0434
$\sqrt{MSE}$	0.0450	0.0339	0.0452	0.0387	0.0429
	LOCAL WHITTLE				
BIAS	-0.0033	-0.0051	0.0072	0.0007	0.0048
$\hat{\sigma}$	0.0358	0.0280	0.0318	0.0261	0.0308
$\sqrt{MSE}$	0.0356	0.0282	0.0323	0.0259	0.0308
	LOCAL LOG-PERIODOGRAM				
BIAS	-0.0054	-0.0175	0.0015	0.0210	0.0088
$\hat{\sigma}$	0.0690	0.0703	0.0589	0.0604	0.0766
$\sqrt{MSE}$	0.0685	0.0718	0.0583	0.0634	0.0763
	GLOBAL LOG-PERIODOGRAM				
BIAS	0.0016	-0.0045	0.0077	0.0025	0.0009
$\hat{\sigma}$	0.0353	0.0307	0.0343	0.0240	0.0309
$\sqrt{MSE}$	0.0350	0.0307	0.0348	0.0239	0.0306

TABLE 1.1. Estimation results for H using 50 independent copies of time series of length 10000.

	FARIMA $(1, d, 1)$				
	Nominal d				
	0	0.1	0.2	0.3	0.4
	$\phi = 0.3$ and $\theta = 0.7$				
	QUADRATIC VARIATIONS				
BIAS	-0.1226	-0.1072	-0.0879	-0.0864	-0.0775
$\hat{\sigma}$	0.0397	0.0359	0.0334	0.0421	0.0381
$\sqrt{MSE}$	0.1288	0.1129	0.0940	0.0959	0.0861
	WAVELETS				
BIAS	-0.0446	-0.0360	-0.0422	-0.0400	-0.0380
$\hat{\sigma}$	0.0458	0.0437	0.0359	0.0457	0.0450
$\sqrt{MSE}$	0.0636	0.0563	0.0552	0.0604	0.0585
	LOCAL WHITTLE				
BIAS	-0.0298	-0.0224	-0.0304	-0.0282	-0.0286
$\hat{\sigma}$	0.0313	0.0328	0.0302	0.0311	0.0340
$\sqrt{MSE}$	0.0430	0.0395	0.0427	0.0417	0.0442
	LOCAL LOG-PERIODOGRAM				
BIAS	0.0012	0.0004	0.0098	-0.0017	-0.0083
$\hat{\sigma}$	0.0779	0.0676	0.0705	0.0656	0.0759
$\sqrt{MSE}$	0.0771	0.0669	0.0705	0.0650	0.0756
	GLOBAL LOG-PERIODOGRAM				
BIAS	-0.0126	-0.0091	-0.0101	-0.0153	-0.0120
$\hat{\sigma}$	0.0391	0.0348	0.0305	0.0329	0.0354
$\sqrt{MSE}$	0.0407	0.0357	0.0319	0.0360	0.0371

TABLE 1.2. Estimation results for $d = H - 0.5$ using 50 independent copies of time series of length 10000. The innovations are Gaussian.

	FARIMA $(1, d, 1)$				
	Nominal d				
	0	0.1	0.2	0.3	0.4
	$\phi = -0.3$ and $\theta = -0.7$				
	QUADRATIC VARIATIONS				
BIAS	-0.0114	-0.0087	-0.0066	-0.0123	-0.0079
$\hat{\sigma}$	0.0325	0.0323	0.0342	0.0443	0.0381
$\sqrt{MSE}$	0.0342	0.0331	0.0345	0.0456	0.0386
	WAVELETS				
BIAS	-0.0063	-0.0054	0.0102	-0.0006	0.0061
$\hat{\sigma}$	0.0421	0.0462	0.0391	0.0487	0.0405
$\sqrt{MSE}$	0.0421	0.0461	0.0400	0.0482	0.0406
	LOCAL WHITTLE				
BIAS	-0.0023	-0.0010	0.0066	-0.0017	0.0063
$\hat{\sigma}$	0.0307	0.0304	0.0283	0.0349	0.0302
$\sqrt{MSE}$	0.0304	0.0301	0.0288	0.0346	0.0306
	LOCAL LOG-PERIODOGRAM				
BIAS	0.0075	0.0248	0.0163	-0.0013	0.0127
$\hat{\sigma}$	0.0791	0.0607	0.0615	0.0818	0.0586
$\sqrt{MSE}$	0.0787	0.0650	0.0631	0.0810	0.0594
	GLOBAL LOG-PERIODOGRAM				
BIAS	0.0015	-0.0014	0.0107	-0.0007	0.0092
$\hat{\sigma}$	0.0340	0.0337	0.0349	0.0337	0.0319
$\sqrt{MSE}$	0.0337	0.0333	0.0362	0.0333	0.0329

TABLE 1.3. Estimation results for $d = H - 0.5$ using 50 independent copies of time series of length 10000. The innovations are Gaussian.

	FARIMA $(0, d, 0)$ with Pareto innovations			
	Nominal d			
	0	0.1	0.2	0.3
	QUADRATIC VARIATIONS			
BIAS	-0.0056	-0.0009	-0.0104	-0.0115
$\hat{\sigma}$	0.0309	0.0381	0.0427	0.0429
$\sqrt{MSE}$	0.0311	0.0378	0.0435	0.0440
	WAVELETS			
BIAS	-0.0163	-0.0072	0.0071	-0.0310
$\hat{\sigma}$	0.1092	0.1608	0.1433	0.1379
$\sqrt{MSE}$	0.1093	0.1593	0.1421	0.1400
	LOCAL WHITTLE			
BIAS	-0.0046	0.0090	0.0021	0.0037
$\hat{\sigma}$	0.0262	0.0312	0.0251	0.023
$\sqrt{MSE}$	0.0263	0.0322	0.0249	0.0236
	LOCAL LOG-PERIODOGRAM			
BIAS	0.0024	0.0049	-0.0022	0.0199
$\hat{\sigma}$	0.0622	0.0685	0.0619	0.0519
$\sqrt{MSE}$	0.0616	0.0680	0.0613	0.0551
	GLOBAL LOG-PERIODOGRAM			
BIAS	0.0026	0.0123	0.0034	0.0063
$\hat{\sigma}$	0.0341	0.0528	0.0434	0.0341
$\sqrt{MSE}$	0.0339	0.0537	0.0431	0.0344

TABLE 1.4. Estimation results for d using 50 independent copies of time series of length 10000. The innovations are Pareto and skewed Stable with $\alpha = 1.5$ and thus $H = d + 2/3$