AN ADAPTIVE BROADBAND ESTIMATOR OF THE FRACTIONAL DIFFERENCING COEFFICIENT

Clifford Hurvich†, *Eric Moulines*‡ *and Philippe Soulier*⋆

† New York University, New York, USA
‡ Ecole Nationale Suprieure des Tlcommunications, Paris, France
⋆ Universit d'Evry Val d'Essonne, Evry, France
churvich@stern.nyu.edu, moulines@tsi.enst.fr, soulier@maths.univ-evry.fr

ABSTRACT

We consider semiparametric fractional exponential (FEXP) estimators of the memory parameter d for a potentially nonstationary linear long-memory time series with smooth additive trend. We use differencing to annihilate the trend, followed by tapering to handle the potential non-invertibility of the differenced series. We propose a method of pooling the tapered periodogram which leads to more efficient estimators of d than existing pooled, tapered estimators. We establish asymptotic normality of the estimator. Finally, we consider minimax rate-optimality and feasible nearly rate-optimal estimators. Some simulations are presented to illustrate our findings. Applications to measure the Hurst coefficient of network traffic data will be presented at the time of the conference.

1. INTRODUCTION

In contrast with wavelet based techniques [1], most Fourier domain estimators known to date fail in the presence of additive trends. This feature is very annoying when analyzing real-data, because the presence of additive trends is the rule rather than the exception. A possible solution to combat these trends was proposed by [3], who recommended differencing the data p times to eliminate the trend, followed by tapering with a new class of tapers that are invariant to the mean and which afford sufficient leakage control to offset the potential strong noninvertibility of the differenced series. These estimators somehow mimic the behavior of the wavelet estimators which also implicitly use a combination of differentiation (the order of differentiation being linked with the number of vanishing moment of the wavelet) and tapering.

In this paper, we consider FEXP introduced in [5, 7] estimators of d for a potentially nonstationary linear series with smooth additive trend. We use differencing and tapering as described above, with a taper in the class proposed by [3]. We propose a method of pooling the tapered periodogram which leads to more efficient estimators of d than existing pooled, tapered estimators. The techniques to robustify the estimators can be applied as well to any periodogram based estimators, including, *e.g.* the GPH estimator.

2. MAIN ASSUMPTIONS AND NOTATIONS

In this section, we make precise some notations on the observed process and on the main tool of analysis, *i.e.* the periodogram. Let $X = (X_t)_{t\in\mathbb{Z}}$ be a process whose p-th difference is stationary for some non negative integer p, and define $Y = (I - B)^p X$ where B is the backshift operator. It will be assumed in the sequel that Y admits a spectral density f that can be expressed as

$$f(x) = |1 - e^{ix}|^{-2d} f^*(x), \tag{1}$$

where f^* is an even, positive, continuous function on $[-\pi, \pi]$ and $-p - 1/2 < d < 1/2$. This assumption is equivalent to the existence of a stationary process Y^* with spectral density f^* such that

$$(I - B)^d Y = Y^*, \tag{2}$$

where $(I - B)^d$ is defined by the infinite series expansion

$$(1 - B)^d = \sum_{j=0}^{\infty} \frac{\Gamma(-d + j)}{\Gamma(-d)\Gamma(j + 1)} B^j .$$

Note that, because of the differentiation, the process Y might be non invertible. Define

$$d_n^Y(x) = (2\pi n)^{-1/2} \sum_{t=1}^{n} Y_t e^{itx}, \tag{3}$$

$$I_n^Y(x) = |d_n^Y(x)|^2 = (2\pi n)^{-1} \Big|\sum_{t=1}^{n} Y_t e^{itx}\Big|^2. \tag{4}$$

In order to deal with possible overdifferencing, the use of a taper is necessary. In a novel family of data taper is considered. Let

$$h_{t,n} = 1 - e^{2i\pi t/n}$$

and, for any integer $s \geq 0$, define the tapered DFT of order s as

$$d_{s,n}^Y(x) = (2\pi n a_s)^{-1/2} \sum_{t=1}^{n} h_{t,n}^s Y_t e^{itx}, \;\; I_{s,n}^Y(x) = |d_{s,n}^Y(x)|^2 \tag{5}$$

where s is the taper order and $a_s := \binom{2s}{s}$ is the normalization constant. It is easily seen that the tapered Fourier transform is invariant to shift in the mean and that the decay rate of the kernel in the frequency domain increases with the kernel order. This property means that high-order kernels are more effective in controlling leakage, which is important when dealing with possibly non-invertible series.

When considering log-transform of the periodogram, it is appropriate to use pooling, which consists in computing averages of periodogram values along blocks of size m, prior to logging. The following pooling scheme is used. Define $K := K(m,n,s) = [(n-m-s)/2(m+s)]$ and for $k = 1, \cdots, K$,

$$\bar{I}^Y_{m,p,s,n,k} = \sum_{j=(m+s)(k-1)+1}^{(m+s)(k-1)+m} I^Y_{q,n}(x_j).$$

Periodogram ordinates at m consecutive ordinates are aggregated, and s frequencies are dropped out of every $m+s$. This will lead to a loss of efficiency of order $m/(m+s)$, but considerably simplify the derivations. This loss is limited provided the pooling index $m \gg s$ (which is in general a safe practice; the pooling size should not however be chosen to large in order to avoid too much bias).

As emphasized in the introduction, the principle of the FEXP estimator is to fit to the data a finite dimensional FEXP model of order q, where q is a non decreasing sequence of integers that depends on the sample size n. Recall that a stationary process with spectral density f satisfying (1) is a FEXP model of order q if

$$f^*(x) = \exp\{\sum_{j=0}^{q-1} \theta_j h_j(x)\},$$

where $h_0 = 1/\sqrt{2\pi}$ and $h_j(x) = \cos(jx)/\sqrt{\pi}$, $j \geq 1$. For such a model, it is natural to estimate simultaneously d and the coefficients θ_j by log-periodogram regression. In the case of interest here, the log-periodogram regression will be performed on the log-pooled-tapered periodogram of the p-times differenced process $Y = (I-B)^p X$ with a data taper of order $s \geq p$ (it is of utmost importance for all the derivations that follow to match the order of differencing and the taper order). Denote $\gamma_{m,p,s} = \mathbb{E}_0[\log(\bar{I}^Y_{m,p,n,k})]$, where $\mathbb{E}_0$ stands for the expectation when Y is a standard white Gaussian noise. Let $Y_{n,k} = \log(\bar{I}^Y_{m,p,s,n,k}) - \gamma_{m,p}$, $y_k = (2k-1)\pi/2K$ and $g(x) = -2\log|1-e^{ix}|$ and

$$(\hat{d}_{m,p,s}(q), \hat{\theta}_0, \cdots, \hat{\theta}_{q-1}) = \arg\min_{\bar{d},\bar{\theta}_0,\cdots,\bar{\theta}_{q-1}} \sum_{k=1}^{K} \Big(Y_{n,k} - \bar{d}g(y_k) - \sum_{j=0}^{q-1} \bar{\theta}_j h_j(y_k)\Big)^2. \quad (6)$$

When d is the only parameter of interest, the centering term $\gamma_{m,p,s}$ is irrelevant and $\hat{d}_{m,p,s}(q)$ can be obtained directly without computing $\hat{\theta}_0, \cdots \hat{\theta}_q$. To give the explicit expression of $\hat{d}_{m,p,s}(q)$, we need some additional notation. Define the following norm on $\mathbb{R}^K$: $\|u\|_n = 2\pi K^{-1} \sum_{k=1}^K u_k^2$, let $< .,. >_n$ denote the associated scalar product and identify any function φ with the vector $(\varphi(y_1), \cdots, \varphi(y_K))$. Let $H_{q,n}$ be the orthogonal projector on the q-dimensional linear subspace of $\mathbb{R}^K$ spanned by the vectors $[h_0, \cdots, h_{q-1}]$. Note that these vectors are orthonormal w.r.t the scalar product $< .,. >_n$. Denote

$$\tilde{g}^*_q := g - H_{q,n} g = \tilde{g}^*_q = g - \sum_{j=0}^{q-1} < g, h_j >_n h_j, \; \tilde{\gamma}_q = \|\tilde{g}^*_q\|^2_n. \quad (7)$$

Denote finally $\mathbf{Y}_n = (Y_{n,1}, \cdots, Y_{n,K})'$. It is then easily seen that

$$\hat{d}_{m,p,s}(q) = \frac{< \tilde{g}^*_q, \mathbf{Y}_n >_n}{\tilde{\gamma}_q} = \frac{2\pi}{K\tilde{\gamma}_q} \sum_{k=1}^{K} \tilde{g}^*_q(y_k) Y_{n,k}. \quad (8)$$

When the true model is not of a finite order, these estimators will be biased by the remainder of the infinite series expansion of $\log(f^*)$. Assuming that f^* is positive over $[-\pi, \pi]$, define

$$\theta_j = \int_{-\pi}^{\pi} h_j(x) \log(f^*(x)) dx, \quad (9)$$

$$l^*_q = \log(f^*) - \sum_{j=0}^{q-1} \theta_j h_j. \quad (10)$$

Define

$$\epsilon_{n,k} = \log(\bar{I}^Y_{m,p,s,n,k}/f(y_k)) - \gamma_{m,p,s}. \quad (11)$$

Then, we can write

$$Y_{n,k} = dg(y_k) + \sum_{j=0}^{q-1} \theta_j h_j + \epsilon_{n,k} + l^*_q(y_k), \quad (12)$$

Using these notations, the FEXP estimator can be decomposed as

$$\hat{d}_{m,p,s}(q) = d + \xi_{m,p}(q) + b_q(f^*), \quad (13)$$

with

$$\xi_{m,p,s}(q) = \frac{2\pi}{K\tilde{\gamma}_q} \sum_{k=1}^{K} \tilde{g}^*_q(y_k)\epsilon_k,$$

$$b_q(f^*) = \frac{2\pi}{K\tilde{\gamma}_q} \sum_{k=1}^{K} \tilde{g}^*_q(y_k) l^*_q(y_k).$$

The term $\xi_{m,p,s}(q)$ is a stochastic fluctuation term and the term $b_{m,p,s}(f^*, q)$ is a bias term which depends on the short-memory component f^* and on the truncation index q, but not on d, and indirectly on m, p and s through the choice of the frequencies y_k. If $\sum_{j=0}^{\infty} |\theta_j| < \infty$, then $l^*_q = \sum_{j=q}^{\infty} \theta_j h_j$, the bias term $b_{m,p,s}(f^*, q)$ is controlled by the rate of decay of the coefficients θ_j. More precisely,

Lemma 1 *There exists a finite constant ς which does not depend on f^* such that for all $q \leq K/2$,*

$$2\sqrt{\pi}|b_{m,p,s}(f^*, q)| \leq (1 + \varsigma q/K) \sum_{j=q}^{\infty} |\theta_j|.$$

3. THEORETICAL RESULTS

In the following sections, it is assumed that $p = s$ and that the trend is polynomial. The results below hold with obvious modifications whenever $s \geq p$ (note however that tapering will in general introduce some losses in efficiency). Extensions to smooth non-polynomial trends (which can be modelled as $\tilde{X}_t = s(t/n) + X_t$ where s is a smooth function) are considered in a forthcoming paper. If the differenced process $Y = (I-B)^p X$ is stationary and satisfies equation (1), then it admits the spectral density

$$f(x) = |1-e^{ix}|^{-2d} f^*(x) = |1-e^{ix}|^{-2d} (2\pi)^{-1} |a^*(x)|^2,$$

To derive asymptotic results, assumptions on the distribution of the driving noise Z and on a^* are required.

(A1) The sequence $(Z_t)_{t\in\mathbb{Z}}$ is i.i.d., $\mathbb{E}[Z_1] = 0$, $\mathbb{E}[Z_1^2] = 1$ and

$$\int_{-\infty}^{\infty} |\mathbb{E}[e^{itZ_0}]|^r dt < \infty, r > 1 \quad (14)$$

Regularity conditions for a^* are conveniently stated in terms of a functional class.

Definition 1 *For $\mu > 1$ and $\rho > 1$, let $\mathcal{L}^*(\mu,\rho)$ be the class of positive continuously differentiable functions u on $[-\pi,\pi]$ such that for all $0 < |x|, |y| \le \pi$,*

$$\frac{\max_{0\le z\le\pi}|u(z)|}{\min_{0\le z\le\pi}|u(z)|} \le \mu, \tag{15}$$

$$\frac{|u(x)-u(y)-(x-y)u'(x)|}{\min_{0\le z\le\pi}|u(z)|} \le \mu\frac{|y-x|^\rho}{(|x|\wedge|y|)^\rho}. \tag{16}$$

Theorem 1 *Let p and m be integers such that $p \ge 1$ and $m \ge 4$. Assume* **(A1)** *and that $\mathbb{E}|Z_0|^{2m+1} < \infty$. Assume that $a^* \in \mathcal{L}^*(\mu,\rho)$ for some $\rho > 1$, and $-p-1/2 < d < 1/2$. If the Fourier coefficients of $l^* = \log(f^*)$ are absolutely summable and if q is a non-decreasing sequence of integers such that $\lim_{n\to\infty}(q^{-1} + q\log^2(n)n^{-1}) = 0$ and*

$$\lim_{n\to\infty}\sqrt{n/q}\sum_{k=q}^{\infty}|\theta_j| = 0, \tag{17}$$

then $\sqrt{n/q}(\hat{d}_{m,p}(q)-d)$ is asymptotically zero-mean Gaussian with variance $(m+p)\sigma^2_{m,p}$.

Here, $\sigma^2_{m,p} = \mathbb{E}_0[Y^2_{n,1}] = \mathbb{E}_0[Y^2_{n,k}]$ for $1 \le k \le \tilde{n}$. It may be shown that, for any given p, by increasing the blocksize m, the efficiency loss can be made arbitrarily close to the corresponding loss incurred by tapering in the Gaussian semiparametric estimator.

4. MINIMAX ESTIMATION

For simplicity, it is assumed in this section that Z_t is Gaussian; it is likely that extensions to more general distributions are possible (and perhaps to linear process w.r.t to a martingale innovations), but the whole theory in not yet available. The following proposition gives an expression mean square error of the FEXP estimator.

Proposition 1 *Let $\mu > 1$, let p be a non negative integer and $\delta_1 \in (0, p+1/2)$, $\delta_2 \in (0,1/2)$. Let q be a non decreasing sequence such that $\lim_{n\to\infty}(q^{-1} + q\log^r(n)n^{-1}) = 0$ with $r = 6$ in the non tapered case (*i.e. $p = 0$*) and $r = 4$ in the tapered case (*i.e. $p \ge 1$*). Then*

$$\lim_{n\to\infty}\sup_{-\delta_1\le d\le\delta_2}\sup_{f^*\in\mathcal{L}^*(\mu)} \left|\frac{n}{q}\mathbb{E}_{d,f^*}[(\hat{d}_{m,p}-d)^2] - (m+p)\sigma^2_{m,p}\right| = 0.$$

Lemma 1 and Proposition 1 yield a bound for the mean square error of $\hat{d}_{m,p}$. Let w be a decreasing sequence and define

$$\mathcal{G}^*(w) = \{f^* : f^*(x) = \exp\{\textstyle\sum_{j=0}^{\infty}\theta_j h_j(x)\}, \quad \forall q \ge 0,\ \displaystyle\sum_{j=q}^{\infty}|\theta_j| \le w(q).\}. \tag{18}$$

Two examples of sequences w are of interest. For $\beta > 0$ and $L > 0$, let $w_{\beta,L}(q) = L(1+q)^{-\beta}$ and define $\mathcal{S}^*(\beta,L) = \mathcal{G}^*(w_{\beta,L})$. Define also $v_{\beta,L}(q) = Le^{-\beta q}$ and $\mathcal{A}^*(\beta,L) = \mathcal{G}^*(v_{\beta,L})$. The class $\mathcal{A}^*(\beta,L)$ is of special interest since the spectral density of any causal stable and invertible ARMA processes is contained in such a class.

Theorem 2 *Let p be a non negative integer, let $\beta > 1$, $\gamma > 0$, $L > 0$, $0 < \delta_1 < p+1/2$ and $0 < \delta_2 < 1/2$. Define $q_n(\beta,L) = [(Ln)^{1/(1+2\beta)}]$ and $q_n(\gamma) = [\log(n)/2\gamma]$. There exists a constant $\bar{C}$ which depends only on β, δ_1 and δ_2 such that*

$$\limsup_n \sup_{-\delta_1\le d\le\delta_2}\sup_{f^*\in\mathcal{S}^*(\beta,L)} n^{2\beta/(2\beta+1)}\mathbb{E}_{d,f^*}[(\hat{d}_{m,p}(q_n(\beta,L))-d)^2] \le \bar{C}L^{1/(2\beta+1)}, \tag{19}$$

$$\lim_{n\to\infty}\sup_{-\delta_1\le d\le\delta_2}\sup_{f^*\in\mathcal{A}^*(\gamma,L)} \frac{n}{\log(n)}\mathbb{E}_{d,f^*}[(\hat{d}_{m,p}(q_n(\gamma))-d)^2] = (m+p)\sigma^2_{m,p}/2\gamma, \tag{20}$$

where $\mathbb{E}_{d,f^*}$ denotes the expectation with respect to the distribution of a Gaussian process with spectral density $e^{dg}f^*$. If $0 < \beta < 1$ and $\mu > 1$, there exists a constant $\bar{C}$ which depends only on β, δ_1, δ_2 and μ such that

$$\limsup_n \sup_{-\delta_1\le d\le\delta_2}\sup_{f^*\in\mathcal{S}^*(\beta,L)\bigcap\mathcal{L}^*(\mu)} n^{2\beta/(2\beta+1)}\mathbb{E}_{d,f^*}[(\hat{d}_{m,p}(q_n(\beta,L))-d)^2] \le \bar{C}L^{1/(2\beta+1)}. \tag{21}$$

Corresponding lower bounds are discussed in [4].

5. ADAPTIVE ESTIMATION

In practice, automatic selection procedure is required to select q. We adapt here a technique initially proposed by [6], to select the bandwidth in kernel estimator. Let ϵ_n a non-increasing sequence such that $\epsilon_n \to 0$. A truncation index q is said *admissible* if, for all $q < q' \le \epsilon_n K$

$$|\hat{d}_{m,p}(q') - \hat{d}_{m,p}(q)| \le \kappa\sqrt{\log(n)}\sigma_{m,p}(q'),$$

where κ is a constant (whose value is given below) and $\sigma^2_{m,p}(q) := \mathbb{E}_0[d^2_{m,p}(q)]$. Let $\hat{q}$ be the smallest admissible integer. The adaptive estimator is then defined as $\hat{d}_{m,p}(\hat{q})$. In order to assess the performance of the adaptive estimator, denote $q^*_n(f^*)$

$$\forall q : q^*_n(f^*) \le q \le \epsilon_n K, \quad |b_q(f^*)| \le \sqrt{\log(n)}\sigma_{m,p}(q). \tag{22}$$

Consider the following assumptions

- **(T1)** Let $\delta_1 \in (0, p+1/2)$ and $\delta_2 \in (0,1/2)$ and let $\mathcal{F}^*$ be a functional class. For all $r \ge 1$, there exists a constant $C(\delta_1,\delta_2,\mathcal{F}^*,r)$ such that for all $\delta_1 \le d \le \delta_2$, all $f^* \in \mathcal{F}^*$ and all q such that $q^*_n(f^*) \le q \le \epsilon_n K$,
$$\mathbb{E}[\xi_{m,p}(q)^{2r}] \le C^r(\delta_1,\delta_2,\mathcal{F}^*,r)\sigma^r_{m,p}(q).$$
- **(T2)** $\mathcal{F}^*$ is a functional class such that there exists a constant $C(\mathcal{F}^*)$ such that, for all $f^* \in \mathcal{F}^*$
$$\forall q : q^*_n(f^*) \le q \le \epsilon_n K, \ \ |b_{n,q}(f^*)| \le C(\mathcal{F}^*).$$
- **(T3)** Let $\delta_1 \in (0, p+1/2)$, $\delta_2 \in (0,1/2)$, $\rho > 0$ and let $\mathcal{F}^*$ be a functional class. There exists a constant $C(\delta_1,\delta_2,\mathcal{F}^*)$ and an integer $N := N(\delta_1,\delta_2,\mathcal{F}^*,\rho)$ such that for all $n \le N$, for all q such that $q^*_n(f^*) \le q \le \epsilon_n K$, and for all $d \in [\delta_1,\delta_2]$ and $f \in \mathcal{F}^*$,
$$\mathbb{P}\left(|\xi_{m,p}(q)| > (\kappa/2-1)\sqrt{\log(n)}\sigma_{m,p}(q)\right) \le C(\delta_1,\delta_2,\mathcal{F}^*)n^{-(\kappa/2-1)^2/2(1+\rho)}.$$

If adaptivity to the target function is desired, then the class $\mathcal{F}^*$ can be restricted to a single function f^*. If adaptivity in the minimax sense is the objective, then the class $\mathcal{F}^*$ should be, on the contrary, chosen as large as possible.

Theorem 3 *Let $\delta_1 \in (0, p+1/2)$ and $\delta_2 \in (0, 1/2)$. Let $\mathcal{F}^*$ be a functional class and for each $f^* \in \mathcal{F}^*$, let $q_n^*(f^*)$ be a sequence of integers such that (22),* **(T1)**, **(T2)** *and* **(T3)** *hold. If $\kappa > 6$, there exists a constant $C(\delta_1, \delta_2, \mathcal{F}^*, \kappa)$ such that for all $f \in \mathcal{F}^*$, for all $d \in [-\delta_1, \delta_2]$ and for large enough n (depending on δ, $\mathcal{F}^*$ and κ),*

$$\mathbb{E}[(\hat{d}(\hat{q}) - d)^2] \leq (1+\kappa)^2 \sigma_{m,p}(q_n^*(f^*))^2 \log(n) + C(\delta_1, \delta_2, \mathcal{F}^*, \kappa) n^{-1}.$$

We now illustrate the minimax adaptivity properties of the adaptive FEXP estimator in the functional classes introduced above. This amounts to finding a sequence q_n^* such that (22) holds with $q_n^*(f^*) = q_n^*$ for all functions f^* in a given functional class.

Corollary 1 *Let p be a positive integer, let $\delta_1 \in (0, p+1/2)$, $\delta_2 \in (0, 1/2)$ and $0 < \beta_* < \beta^* < \infty$. There exists a constant $\bar{C} := \bar{C}(\beta_*, \beta^*, L^*, \delta_1, \delta_2)$ such that*

$$\limsup_n \sup_{-\delta_1 \leq d \leq \delta_2} \sup_{\beta_* \leq \beta \leq \beta^*} \sup_{0 < L < L^*} \sup_{f^* \in \mathcal{S}^*(\beta, L)} (n/\log(n))^{2\beta/(2\beta+1)} \mathbb{E}_{d,f^*}[(\hat{d}_{m,p}(\hat{q}) - d)^2] \leq \bar{C},$$

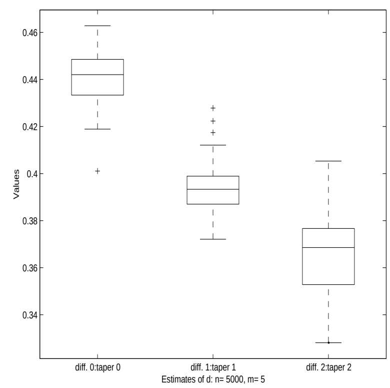
$$\limsup_n \sup_{-\delta_1 \leq d \leq \delta_2} \sup_{\beta_* \leq \beta \leq \beta^*} \sup_{0 < L < L^*} \sup_{f^* \in \mathcal{A}^*(\beta, L)} \frac{n}{\log^2(n)} \mathbb{E}_{d,f^*}[(\hat{d}_{m,p}(\hat{q}) - d)^2] \leq \bar{C},$$

where here again $\mathbb{E}_{d,f^*}$ denotes expectation with respect to the distribution of a Gaussian *process with spectral density $e^{dg} f^*$.*

A logarithmic loss appears here with respect to the minimax rates of convergence. For the classes $\mathcal{S}^*$, it has been shown in [4] that this loss is an unavoidable price to pay for adaptation.

6. SIMULATIONS

In this section, a small-scale Monte-Carlo experiment is presented. Due to lack of space, only a very limited set of results are actually presented. Matlab codes and extensive simulation results can be found at *http://www.tsi.enst.fr/ moulines*. In a first example, a fractional Gaussian noise (FARIMA(0,d,0)) ϵ_k is observed in presence of an additive perturbation, $Y_k = f(k/n) + \epsilon_k$, where f is a smooth function, here $f(t) = 2t^{0.3} + t$. Trends of the form t^α are long been known to induce "erroneous" long-memory. In this simulation, the fractional differencing coefficient is set to $d = 0.4$ and the number of observations is 5000. The choice of the truncation order q is based on the data driven procedure described above with $kappa = 3$. Three different settings are considered: the original estimator (no differencing and no tapering), and two versions of the proposed estimator (differencing once and tapering with a taper of order 1, differencing twice and using a taper of order 2). In all cases, the pooling size is set to $m = 5$. It is easily seen on the boxplot [1] below that the combination of tapering and differencing effectively leads to a significative reduction of the bias, at the expense of an increase in the variance (due to tapering). In this example, differentiation of order 1 appears to be sufficient for the purpose of bias reduction. Increasing the order of differentiation increase by a significant amount the estimation variance (recall that, with $m = 5$ and $p = 2$, the effective number of regression points is almost reduced by a factor 2).



[1] The box has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. Outliers are data with values beyond the ends of the whiskers.

7. REFERENCES

[1] P. Abry, P. Flandrin, M. Taqqu and D; Veitch, *Self-similarity and long-range dependence through the wavelet lens* To appear in Doukhan, Oppenheim, Taqqu, Eds.

[2] W. W. Chen and C. M. Hurvich, *Estimating fractional cointegration in the presence of polynomial trends*, Preprint, 1999.

[3] C. M. Hurvich and W. W. Chen, *An efficient taper for potentially overdifferenced long-memory time series*, J. of Time Series Analysis **21(2)** (2000), 155–180.

[4] A. Iouditsky, E. Moulines, and Ph. Soulier, *Adaptive estimation of the fractional differencing coefficient*, Prépublication de l'université d'Evry val d'Essonne, 1999, Accepted for publication in Bernouilli.

[5] J. Janacek, *Determining the degree of differencing for time series via the log spectrum*, J. of Time Series Analysis **3(5)** (1997), 177–183.

[6] O. Lepski, *On a problem of adaptive estimation in gaussian white noise.*, Theory Probab. Appl. **35(3)** (1990), 454–466.

[7] E. Moulines and Ph. Soulier, *Log-periodogram regression of time series with long-range dependence*, Annals of Statistics **27** (1999), no. 4, 1415–1439.