

RECENT ADVANCES ON THE SEMI-PARAMETRIC ESTIMATION OF THE LONG-RANGE DEPENDENCE COEFFICIENT

JEAN-MARC BARDET, ERIC MOULINES, AND PHILIPPE SOULIER

ABSTRACT. Semi-parametric estimation of the long-range dependence parameter was a subject of major interest in recent years. The purpose of this paper is to put in a common framework several recent contributions of that topic. We will focus in particular on *spectral methods*, which consist in estimating the exponent of the singularity of the spectral density function at zero frequency and on *time-domain* methods, which are based on the estimation on the rate of convergence to zero of the autocovariance coefficients and/or on the scaling of the variance of processes deduced from the original process by generalized aggregation.

RÉSUMÉ. L'estimation du coefficient de longue portée dans un contexte semi-paramétrique est un sujet dont l'importance s'est considérablement développé au cours des dernières années. L'objet de cet article est de synthétiser un certain nombre de développement récents dans ce domaine. Nous nous concentrons en particulier sur les méthodes spectrales, qui consistent à estimer l'exposant de la singularité de la densité spectrale à la fréquence nulle, et les méthodes temporelles, basées sur l'exposant de décroissance de la fonction d'autocorrélation à l'infini, ou de façon similaire, sur la loi d'échelle de la variance de processus déduits du processus initial par aggrégation (généralisée).

1. INTRODUCTION

An ARMA process $\{X_t\}_{t \in \mathbb{Z}}$ is often referred to as a short memory process since the auto-covariance coefficient (or dependence) between X_k and $X_{k'}$ decreases "rapidly" as $|k - k'| \rightarrow \infty$. In fact, the autocorrelation function is geometrically bounded, i.e. $|\rho(k)| \leq Cr^{|k|}$, with $C < \infty$ and $0 < r < 1$. A long memory process is a stationary process for which $\rho(k) = L(k)k^{2d-1}$ where $L(k)$ is a slowly varying function as $k \rightarrow \infty$, where $0 < d < 1/2$ is the memory parameter (LRD). There is evidence that long memory processes occur quite frequently in fields as diverse as hydrology, economy, computer networks.

The series of the autocorrelation coefficients being square summable, the spectral measure of $\{X_t\}$ is absolutely continuous w.r. to the Lebesgue measure on $[-\pi, \pi]$. Denote $f(x)$ the spectral density function. Since $\sum_k k^{2d-1} = \infty$, the spectral density is infinite at the zero frequency. More precisely, it may be shown that $f(x) = L(x)x^{-2d}$, where $L(x)$ is a slowly varying function as $x \rightarrow 0^+$.

The aim of this contribution is to summarize some recent advances on the estimation of the LRD parameter, using spectral-domain techniques and time-domain

© Société de Mathématiques Appliquées et Industrielles. Typeset by L^AT_EX.

J.M. Bardet: Université Paul Sabatier Toulouse III, Laboratoire de Statistiques et de Probabilités, 118 route de Narbonne, 31004 Toulouse Cedex, E-mail: bardet@cict.fr.

E. Moulines: ENST/TSI & CNRS, URA 820, GdR ISIS, 46, rue Barrault, 75 634 Paris Cedex 13, France. E-mail: moulines@tsi.enst.fr.

Ph. Soulier: Université d'Evry Val d'Essonne, Laboratoire de Mathématiques et d'Informatique, 91025 Evry Cedex, E-mail: soulier@lami.univ-evry.fr.

E. Moulines was partially supported by France Telecom-CNET, CTI n0 PE95-7633.

techniques. Spectral techniques aim at estimating the exponent of the power-law singularity of the spectral density at zero-frequency (see section 2). Depending on the kind of assumptions one is willing to do on the behavior of the spectral density $f(x)$ for $x \neq 0$ (smoothness), different estimators can be constructed. When only weak statements can be done on the regularity of $f(x)$ outside a "small" neighborhood of the zero-frequency, one has to resort to "local" methods, whose convergence and asymptotic normality can be established under rather weak assumptions on the spectral density (see section 2.1). If the spectral density is known to be smooth for $x \neq 0$, alternative estimates can be constructed, showing generally better convergence rate and, in practice, much better behavior (see section 2.2).

Instead of estimating the singularity of the spectral density at zero frequency, it is possible to estimate the rate of convergence to zero of the autocovariance coefficients of the process for large lags (section 3). Since covariance are "time-domain" quantities (as opposed to the spectral density !), the techniques which are either directly or implicitly based on the scaling law of the autocovariance coefficients for large lags are often referred to as under the catch-up name of time-domain methods, though the difference between time-domain and frequency-domain methods is somewhat messier than this clear-cut definition may indicate. Time-domain and frequency-domain methods are "conceptually" equivalent, because the singularity of the spectral density function at zero frequency translates (through Tauberian Theorems) into a rate of convergence of its Fourier coefficients (namely, the autocovariance coefficients) for large lags. Of course, from a statistical standpoint, these two approaches for estimating the memory parameter are markedly different. The rate of convergence of the autocovariance coefficients to zero can also (and almost equivalently) be assessed by studying the scaling of the variance of aggregated processes (the simplest example being $X_k^{(N)} = N^{-1} \sum_{t=1+Nk}^{N(k+1)} X_t$) with respect to the aggregation size (see section 3.1). Generalizations of this approach, based on the concept of generalized quadratic variations are considered (section 3.2).

2. SPECTRAL-DOMAIN TECHNIQUES

Let $\{X_t\}_{t \in \mathbb{Z}}$ be a covariance stationary with spectral density $f(x)$. It is assumed that the spectral density of $\{X_t\}$ writes $f(x) = \sin^{-2d}(x/2)f^*(x)$ where $f^*(x)$ is a "smooth" function (in a sense precisely defined below) that governs the short-memory component of the spectral density. When $d = 0$, $\{X_t\}$ is short-range dependent, while for $d \neq 0$ the spectral density is singular at zero. The process is long-range dependent for $0 < d < 1/2$ (the autocovariance function is not absolutely summable). When $-1/2 < d < 0$, the spectral density of the process is zero at zero frequency, but the process $\{X_t\}$ is still invertible; such a situation occurs for example when modeling the first differences of a process which is not stationary but less so than a unit root process (a situation which is frequent in econometric applications or when studying environmental time-series).

In the *parametric* approach, a finite-dimensional parametric model is assumed to hold for $f^*(x)$. A key example is the ARFIMA(p, d, q) (standing from Auto Regressive Fractionnally Integrated Moving Average) model (Granger and Joyeux (1980)) in which $f^*(x)$ is assumed to be a rational function, $f^*(x) = |P(e^{ix})|^2/|Q(e^{ix})|^2$, where $P(z)$ and $Q(z)$ are polynomial of degrees p and q , respectively. Another example is the FEXP(p) (standing from Fractionnal Exponential) model (Robinson (1990), Beran (1993)), where $f^*(x)$ is an exponential of a finite order trigonometric polynomial, $f^*(x) = \exp(\sum_{k=0}^p \theta_k \cos(kx))$. The parameters of $f(x)$, including d , may then be estimated using log-periodogram regression (Moulines and Soulier, 1998), Gaussian maximum likelihood (see Dahlhaus (1989)) or the Gaussian-Whittle contrast function approach (see Fox and Taqqu, 1986, Giraitis and Surgailis, 1990,

Robinson, 1995b). In all cases, the estimators may be shown to be asymptotically normal (under some additional assumptions - linearity, existence of moments) with the usual rate of convergence \sqrt{n} , where n is the sample size, provided that the *parametric model is correctly specified*. However, the estimator may be inconsistent if the model is *misspecified*, e.g. when the model order is under- or over-estimated. This drawback provides motivation for *semi-parametric* estimation of d , for which the short-memory component $f^*(x)$ is allowed to belong to a wider class of functions. There are currently two classes of semi-parametric frequency-domain methods: the *local methods* which fit the spectral density only in a neighborhood of zero-frequency and the *global methods* which attends to model the spectral density on the whole frequency range.

2.1. LOCAL METHODS

The local methods aim at constructing estimators that are consistent without any restrictions on $f^*(x)$ away from zero frequency (apart from integrability on $(-\pi, +\pi]$, a property which is implied by the underlying assumption of covariance stationarity), *i.e.* it is only assumed that

$$f(x) \simeq Cx^{-2d} \text{ as } x \rightarrow 0+ \quad (2.1)$$

where \simeq means that the ratio of the left- and right-hand sides tend to 1, C is positive, and $-1/2 < d < 1/2$. In applications, the singularity exponent d and the constant C are most often unknown. Local methods aim at constructing estimates of C and d by a *local* estimation technique, *i.e.* methods which are based only on the low-frequencies ordinates of the periodogram. The idea of using local techniques has been initiated in the early work of Geweke and Porter-Hudak (1983). It has been then refined and expanded by several authors, and in particular by Kunsch (1987). Theoretical foundations of the local methods were laid down in a series of paper by Robinson (1995-1997). Data-driven and adaptive methods were considered by Hurvich and his co-authors. There are basically three *local* estimators, which are detailed below.

2.1.1. THE GPH ESTIMATOR. Let m be a fixed integer, and for all n , let $n_m = 2m[n/2m]$ and $K_n = [n/2m]$. Define the discrete Fourier transform and the periodogram of $\{X_1, \dots, X_{n_m}\}$ as:

$$\omega_n(x) = (2\pi n_m)^{-1/2} \sum_{t=1}^{n_m} X_t e^{itx} \quad I_n(x) = |\omega_n(x)|^2.$$

These quantities are evaluated at the Fourier frequencies $x_s = 2\pi s/n$, $1 \leq s \leq n$ (note that the correction for the unknown mean is not necessary). Following the procedure proposed by Robinson (1994), the frequency axis is divided in non-overlapping segments of size m , and the periodogram is averaged over each segments. More precisely, for $k = 1, \dots, K_n$, denote $J_k = \{m(k-1) + 1, \dots, mk\}$ and

$$Y_{n,k} = Y_{n,2K_n-k+1} = \log \left(2e^{-\tau_m} \sum_{i \in J_k} I_n(x_i) \right),$$

where $\tau_m = \psi(m)$, and $\sigma_m^2 = m\psi'(m)$, where $\psi(z)$ is the digamma function $\psi(z) = \Gamma'(z)/\Gamma(z)$, $\Gamma(z)$ being the gamma function. Denote $g(x) = -2\log|1 - e^{ix}|$ and $y_k = (2k-1)\pi/2K_n$, $1 \leq k \leq K_n$. For $0 < L < M \leq K_n$ some appropriately chosen trimming numbers L and M (denoting respectively the minimal and the

maximal indexes of periodogram ordinates taken into consideration), define

$$\hat{d}(L, M) \triangleq \frac{\sum_{j=L+1}^M \left(g(y_j) - (M-L)^{-1} \sum_{j=L+1}^M g(y_j) \right) Y_{n,j}}{\sum_{j=L+1}^M \left(g(y_j) - (M-L)^{-1} \sum_{j=L+1}^M g(y_j) \right)^2}, \quad (2.2)$$

which is the least-square estimator of d in the *linear regression model* given by the identity

$$Y_{n,j} = c + dg(y_j) + \eta_j, \quad l+1 \leq j \leq m \quad (2.3)$$

The estimator $\bar{d}(0, M)$ (the lowest trimming number L is set to zero) was introduced in the seminal contribution of Geweke and Porter-Hudak (1983) (GPH). Since $f(x)$ can be approximated by $C \sin^{-2d}(x/2)$ in only a neighborhood of zero frequency, an asymptotic theory requires that M tend to infinity slower than n . The motivation for the GPH estimator stems from the fact that, for *short-memory time series*, the periodogram ordinates at Fourier frequencies may be asymptotically represented as $I_n(x_i) \simeq f(x_i)Y_{i,n}/2$, where $Y_{i,n}$, $1 \leq i \leq n$ is a set of independent random variables distributed as central chi-square with two degrees of freedom. It was shown that this decomposition do not longer holds for long-memory series, which casts for almost ten years a serious doubt on the validity of such approaches. A rigorous asymptotic analysis of the GPH family of estimator has been undertaken by Robinson (1995), under the following assumptions

- **(A1)** [Smoothness at zero-frequency] There exists $0 < C < \infty$, $-1/2 < d < 1/2$ and $0 < \alpha < 2$ such that $f(x) = Cx^{-2d}(1 + O(x^\alpha))$; in addition, in a neighborhood $(0, \epsilon)$ of the origin, $f(x)$ is differentiable and $|f'(x)| = O(x^{1-2d})$,
- **(A2)** [Gaussianity] $\{X_t\}$ is a Gaussian process,
- **(A3)** [Choice of trimming number] as $n \rightarrow \infty$,

$$M^{1/2} \log(M)/L + L \log^2(n)/M + M^{1+1/(2\alpha)}/n \rightarrow 0.$$

It holds that (Robinson (1995), Theorem 3)

THEOREM 2.1. *Assume (A1–A3). Then, as $n \rightarrow \infty$,*

$$2M^{1/2}/\log(n)(\hat{d}(L, M) - d) \rightarrow_{\mathcal{L}} \mathcal{N}(0, m\psi'(m));$$

Assumption **(A3)** implies that the weakest possible upper bound on the maximum number of periodogram ordinates M relative to the sample size n is $M^5/n^4 \rightarrow 0$ (an upper bound which is reached when $\alpha = 2$, which implies the existence and the continuity of the second order derivative of $f^*(x)$ at zero). Note that, for $0 < \alpha < 2$ "acceptable" sequences for the trimming numbers L and M always exist. It should be stressed that, *no matter how smooth $f^*(x)$ is outside a neighborhood of zero*, the rate of convergence of the GPH estimator is $n^{-\alpha}$, with $\alpha < 2/5$.

REMARK 2.2. Note that, for $m = 1$, $\sigma_m^2 = \pi^2/6$. The recurrence relation $\psi'(m+1) = \psi'(m) - m^{-2}$ indicates that the function $m\psi'(m)$ *decreases* in m , taking values 1.289 at $m = 2$ and 1.185 at $m = 3$, and tending to 1 as $m \rightarrow \infty$. There is thus always some advantages to "pool" the periodogram ordinates prior to compute logged variables.

REMARK 2.3. As shown by Hurvich, Deo and Brodski (1997), consistency and asymptotic normality of the estimator hold without trimming the L lowest frequencies, and that finite-sample performance is better without trimming. A similar result may be inferred from Theorem 3 in Moulines and Soulier (1997).

REMARK 2.4. In a recent paper Hurvich, Deo and Brodski (1997) have computed the bias and the variance of the original GPH estimator (setting the lowest trimming number to $L = 0$). It is shown in this contribution that if the spectral density is $f(x) = |1 - e^{ix}|^{-2d} f^*(x)$ (where $f^*(x)$ is an even positive function on $[-\pi, \pi]$,

with some additional regularity properties), then the bias of the periodogram behaves as $C_m f^{*''}(0)/f^*(0)M^2/n^2$, where C_m is some constants that depend only upon m . The same authors also show that the variance of the GPH estimator is C'_m/M and thus obtain a theoretical expression for the mean-square error (MSE) of the estimator. It is shown that the optimal rate for the trimming number M^* is proportional to $n^{4/5}$ (which is markedly higher than the rate $n^{1/2}$ originally suggested by Geweke and Porter-Hudak), but the proportionality constant depends on $f^*(x)$ through the value of its second-order derivative at zero. This formula can be used to derive plug-in estimate of the optimal bandwidth.

2.1.2. LOCAL WHITTLE ESTIMATOR. Instead of computing a log-periodogram regression, one may, as originally suggested by Kuensh (1987), a local Gaussian-Whittle minimum contrast estimator. Consider the objective function

$$Q(C, d) \triangleq M^{-1} \sum_{j=1}^M \{ \log(Cx_j^{-2d}) + x_j^{2d} I_n(x_j)/C \} \quad (2.4)$$

Define Θ the closed interval of admissible estimates of d , $\Theta = [\Delta_1, \Delta_2]$, where $-1/2 < \Delta_1 < \Delta_2 < 1/2$. We can choose Δ_1 and Δ_2 arbitrarily close to $-1/2$ and $1/2$ or we can choose them to reflect a prior knowledge on d (e.g. $\Delta_1 = 0$ if we know that $f(0) > 0$). Define

$$(\hat{d}, \hat{C}) = \operatorname{argmin}_{0 < C < \infty, d \in \Theta} Q(C, d) \quad (2.5)$$

The asymptotic theory of this estimator is studied in Robinson (1995), under the assumption that M tends to infinity to 0 more slowly than n , so that the proportion of the frequency band involved in the estimation degenerates (relatively slowly) to zero as n increases. The following assumptions are introduced

- **(W1)** as $x \rightarrow 0^+$, $f(x) \simeq C_0 x^{-2d_0} (1 + O(x^\beta))$, where $0 < C_0 < \infty$, $d_0 \in [\Delta_1, \Delta_2]$, $\beta \in (0, 2)$. In addition, in a neighborhood $(0, \epsilon)$ of the origin, $f(x)$ is differentiable and $df^{1/2}(x)/dx = O(f^{1/2}(x)x^{-1})$ as $x \rightarrow 0^+$.
- **(W2)** X_t is a linear process,

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad (2.6)$$

where Z_t is a martingale increment $E(Z_t | \mathcal{F}_{t-1}) = 0$ w.p.1 and

$$E(Z_t^2 | \mathcal{F}_{t-1}) = 1, \quad E(Z_t^3 | \mathcal{F}_{t-1}) = \mu_3, \quad E(Z_t^4) = \mu_4 \quad (2.7)$$

where $\mathcal{F}_t = \sigma(Z_s, s \leq t)$, the σ -algebra of events generated by $Z_s, s \leq t$.

- **(W3)** as $n \rightarrow \infty$, $1/M + M^{1+2\beta}(\log(M))^2/n^{2\beta} \rightarrow 0$.

Assumption (W1) is a regularity condition, analogous imposed in the parametric case. Assumption (W2) takes the innovation in the Wold representation to be a fourth-order stationary martingale difference sequence. Assumption (W3) is once again essential, because M must tend to infinity for consistency, while it must do so more slowly than n , because (A1) specifies the form of the spectral density only for $x \rightarrow 0^+$. Note that (W3) imposes an upper bound on the rate of increase of M with n , the weakest version arising when $\beta = 2$.

THEOREM 2.5. *Assume that (W1-W3) hold. Then, as $n \rightarrow \infty$,*

$$M^{1/2}(\hat{d} - d) \rightarrow_{\mathcal{L}} \mathcal{N}(0, 1/4) \quad (2.8)$$

Note that, similar to the log-periodogram regression technique, the optimal rate of convergence is $n^{-\alpha}$, with $\alpha < 2/5$. The variance in the limiting distribution is constant over d_0 and indeed completely free of unknown parameters, so Eq. (2.8) is simple to use in approximate rules of inference.

eqresulWHI indicates that \hat{d} is asymptotically more efficient (for the same M sequence) than the GPH estimate (though the loss of efficiency is minimal if the trimming number M is large).

2.2. GLOBAL METHODS

In the local approaches, only the behavior of the spectral density at zero frequency is specified (see assumption (A1) or (W1)). If one may assume (as it is often the case) that the spectral density is smooth over the *whole* frequency range, then better rate of convergences can be obtained. Provided that $f^*(x)$ is smooth enough (see below), then one may construct estimators that may have mean squared error as small as $O(\log(n)/n)$ (i.e. equal to the parametric rate up to a $\log(n)$ factor).

2.2.1. LOG-PERIODOGRAM REGRESSION. Denote $h_0(x) = 1/\sqrt{2\pi}$ and $h_j(x) = 1/\sqrt{\pi} \cos(jx)$, for $j > 0$. In the sequel, it is assumed that

- **(G1)** $f^*(x)$ is bounded away from zero on $[-\pi, \pi]$, differentiable with bounded derivative. In addition, the coefficients $\{\theta_j\}_{j \geq 0}$ of the Fourier series expansion of $\log f^*(x) = \sum_{j=0}^{\infty} \theta_j h_j(x)$ are such that $\sum_{j=0}^{\infty} j^\beta |\theta_j| < \infty$.

Note that, if $\beta \geq 1$, the assumption $\sum j^\beta |\theta_j| < \infty$ implies that $f^*(x)$ is continuously differentiable. The assumption that $f^*(x)$ is differentiable implies the convergence of the series $\sum j^\beta |\theta_j|$ for $0 \leq \beta < 1/2$. This assumption on the rate of decrease of the Fourier coefficients seems unavoidable.

A semi-parametric estimate of the long-range dependence parameter d , \hat{d}_p can be obtained by truncating the (infinite) Fourier expansion of $l^*(x)$ to its first p coefficients, and letting $p \rightarrow \infty$ at a certain rate. This kind of approach is often used in non-parametric functional estimation, and referred to as *projection estimates*.

In the log-periodogram regression framework (see the notations above), a semi-parametric estimator of d can be obtained as

$$\hat{d}_{p,n} \triangleq \left(\sum_{k=1}^{2K_n} \tilde{g}_{p,n}^2(y_k) \right)^{-1} \left(\sum_{k=1}^{2K_n} \tilde{g}_{p,n}(y_k) Y_{n,k} \right), \quad (2.9)$$

where $\tilde{g} = g - \sum_{j=0}^p \tilde{\alpha}_j h_j(x)$, and $\tilde{\alpha}_j = \pi/K_n \sum_{k=1}^{2K_n} g(y_k) h_j(y_k)$ (note that $\sum_{j=0}^p \tilde{\alpha}_j h_j(\underline{y})$ where $\underline{y} = (y_1, \dots, y_{K_n})^T$ is the orthogonal projection of $g(\underline{y})$ on the linear space spanned by $[h_0(\underline{y}), \dots, h_p(\underline{y})]$). We can now state a central limit theorem for the log-periodogram regression estimator of d , under the additional assumptions

- **(G2)** $\{X_t\}$ is a Gaussian process,
- **(G3)** As $n \rightarrow \infty$, $p^3 \log^2(n)/n^2 + n/p^{(1+2\beta)} \rightarrow 0$.

The case perhaps most widely considered in practice is where the θ_j decay exponentially fast to zero so that $\log f^*(\cdot)$ has an infinite number of continuous derivatives. In this case one may set $p = \log(n)$. It follows from Moulines and Soulier (1998) that

THEOREM 2.6. *Assume (G1-G3). Then, $(n/p)^{1/2}(\hat{d}_n - d)$ is asymptotically normal*

$$\sqrt{\frac{n}{p}}(\hat{d}_n - d) \rightarrow_d \mathcal{N}(0, m\sigma_m^2),$$

$$\lim_{n \rightarrow \infty} np_n^{-1} E(\hat{d}_n - d)^2 = m\sigma_m^2.$$

In addition Moulines and Soulier (1998) gives an expression of the bias and the variance of the estimates. Under **(G1)**, the bias may be shown to be bounded by $\log(n)/p^\beta$ and the optimal rate of convergence is $O(n^{-\beta/(2\beta+1)})$. Provided that θ_j decreases to zero exponentially, it follows from theorem 2.6 that if we take p proportional to $\log(n)$ with a suitably chosen proportionality constant then $E(\hat{d}_{p,n} -$

$d)^2 = O(\log(n)/n)$. Furthermore, if we take p proportional to $\log(n)^{1+\epsilon}$ for any fixed $\epsilon > 0$ and using any positive proportionality constant, then the MSE of $\hat{d}_{p,n}$ is $O(\log(n)^{1+\epsilon}/n)$. On the other hand, the fastest rate of convergence for "local" regression methods is $O(n^{-4/5})$. This is why, in many circumstances, the "global" semi-parametric estimator strongly outperforms the "local" one.

REMARK 2.7. Recently, Bhansali and Kokoszka (1997) have proposed a semi-parametric estimator of d based on fitting a model which is fully parametric but not assumed to be correctly specified. They assume that the true spectral density obeys and ARFIMA($\infty, d, 0$) model, and then show that estimates of d based on fitting potentially incorrect ARFIMA($p, d, 0$) (i.e. fractional autoregressive or FAR model) models by least-squares will be consistent for d , assuming that $p \rightarrow \infty$ as $n \rightarrow \infty$. The proposed FAR estimator may be viewed also as providing a parametric approach to semi-parametric estimation. Bhansali and Kokoszka (1997) did not obtain expressions for the asymptotic bias and variance of their estimators. Nevertheless, in simulations from various ARFIMA(p, d, q) models, the bias and the variance of the FAR estimator were found to compare favourably to those of the GPH estimator.

REMARK 2.8. A similar result is known to hold in a somewhat analogous context of non parametric spectral estimation. There, it has been shown by Berk (1974) that if the true process obeys an AR(∞) model and an AR(p) model is fitted to the data by least squares, with $p \rightarrow \infty$ at a suitable rate, then the resulting spectral estimate is consistent for the true spectral density in the sense that the mean integrated square error of the estimate tends to zero. Furthermore, if the true spectral density is infinitely differentiable, so that the autoregressive coefficients go to zero at an exponential rate, then taking p proportional to $\log(n)$ with a suitable proportionality constant yields a mean integrated squared error of $O(\log(n)/n)$, while the corresponding smoothed periodogram estimator can achieve a mean integrated square error no better than $O(n^{-4/5})$, no matter how smooth the spectral density f^* is.

REMARK 2.9. It is likely that similar result can be derived when Whittle approximate likelihood is used instead of a plain log-periodogram regression. Up to our best knowledge, there are no precise results in this direction.

2.2.2. ORDER SELECTION. The choice of the truncation number p is often ambiguous. It is customary in the non-parametric functional estimation setting to select the truncation number p using some kind of data-driven criterion. Such a procedure is documented below. A semi-parametric estimator of the log-spectral density may be defined as

$$\hat{l}_{p,n} = \hat{d}_p g + \sum_{j=0}^p \hat{\theta}_j h_j. \quad (2.10)$$

In this contribution, the quality of the semi-parametric estimator of the spectral density is assessed over all the frequency range. As usual in non-parametric regression problems, we use the *mean average squared error* (MASE), defined as

$$R_{p,n} = \frac{\pi}{K_n} \sum_{k=1}^{2K_n} \mathbb{E}(l(y_k) - \hat{l}_{p,n}(y_k))^2. \quad (2.11)$$

The optimal truncation order w.r.t to $R_{p,n}$ is $p_n = \operatorname{argmin}_{p \leq K_n} R_{p,n}$. Of course, $R_{p,n}$ is unknown and p_n must be estimated from the data. In this contribution, we consider the following estimate of p_n

$$\hat{p}_n = \operatorname{argmin}_{1 \leq p \leq K_n} \left(S_{p,n} + 4\pi \frac{p}{K_n} \sigma_m^2 \right). \quad (2.12)$$

where $S_{p,n} = \pi K_n^{-1} \sum_{k=1}^{2K_n} (Y_{n,k} - \hat{l}_{p,n}(y_k))^2$ is the averaged square residual (ASR). Asymptotic optimality of the C_p statistics for the orthogonal series estimation of regression has been established, under restrictive assumption on the noise sequence, by Li (1987) and Polyak and Tsybakov (1989). The optimality of the C_p statistics was obtained for the log-periodogram regression in Moulines and Soulier (1998b)

THEOREM 2.10. *Assume (G1-G2). Then,*

$$\mathbb{P} - \lim_{n \rightarrow \infty} \left(R_{\hat{p}_{n,n}} / \min_{1 \leq p \leq K_n} R_{p,n} \right) = 1. \quad (2.13)$$

where $\mathbb{P} - \lim$ denotes the convergence in probability. In addition $\lim_{n \rightarrow \infty} \hat{p}_n = \infty$ with probability one.

Moulines and Soulier (1998b) (see also Hurvich and Brodsky (1998)) show, on numerous examples, that the data-driven global estimate outperforms the "local" estimators.

3. TIME-DOMAIN METHODS

The power-type singularity of the spectral density function at zero-frequency may equivalently be expressed in time-domain as a condition on the Fourier coefficients of f , namely the auto-covariance coefficient $\gamma(h) = \text{cov}(X_i, X_{i+h})$; more specifically, it may be shown that (see section A) the covariance of a LRD process ($0 < d < 1/2$) may be expressed as $\gamma(h) = L(h) |h|^{2d-1}$ where $L(\cdot)$ is slowly varying function for $|h| \rightarrow \infty$. In the parametric case (FGN, FARIMA), the expression of L is known up to a finite number of parameters. In the semi-parametric context, only the long-range dependence of the time-series is assumed and the exact form of L is not known. There are different techniques for estimating d directly from *time-domain* statistics (e.g. auto-covariance, variance of aggregated processes, etc.). These techniques are generally regarded as more robust (in a somewhat loose sense) than spectral domain techniques, and are most often computationally simple. Early examples of time-domain estimation techniques (not covered in this survey) include

- **R/S statistics** the R/S statistic has been introduced by Hurst and later popularized by Mandelbrot and his coworkers (see for instance Mandelbrot and Taquq (1979)),
- **Variance time-plot** the variance time-plot (see Beran (1994), which consists in regressing the variance of aggregated process with respect to the aggregation scale. For a long-range dependence process (with finite variance), it is easily seen that the variance of $v_t(m) = m^{-1} \sum_{j=1}^m X_{tm+j}$ is asymptotically proportional to m^{2d-1} , for large m . Using this scaling property, an estimate of the LRD parameter d can be constructed by regressing the variance of the sample aggregated process (with different aggregation horizon) as a function of the aggregation horizon (more on this later).
- **Covariance regression** this technique consists in regressing the estimated auto-covariance sequence $\hat{\gamma}(h)$ for large lags, exploiting the fact that $\gamma(h) = L(h)h^{2d-1}$. This technique shares some similarity with the variance time-plot (see Hall et al. (1998) and the references therein).

There are some practical shortcomings with the three methods mentioned above, the more annoying for a practitioner being the choice of the regression interval, which is a very subtle and intricate issue. Also, the limiting distribution of these estimators is not straightforward : the rate of convergence and the form of the limiting distribution generally depends on the value of the LRD parameter, which makes it difficult to construct confidence intervals or test statistics.

In this section, we present an alternative procedure, the *log-variogram method*, which is based on the regression of generalized quadratic variations of $\{X_t\}$. This

method is straightforward to implement and, the rate of convergence of the estimator is \sqrt{n} , whatever the value of d is, provided that $0 \leq d < 1/2$.

3.1. LOG-VARIOGRAM

The classical *log-variogram method*, is deduced from theoretical results on quadratic variations of a process (see for instance Guyon and Leon (1989)). The results are more naturally stated using continuous-time processes observed as discrete time-instants.

(3.1)

It has been shown by Guyon and Leon (1989) that

PROPOSITION 3.1. *Assume that $X \triangleq \{X_t\}_{t \in \mathbb{R}^+}$. Then, for any $n \in \mathbb{N}$:*

1. $ES_n(k) = \gamma(k)$,
2. For $n \rightarrow +\infty$, $S_n(k) \rightarrow \gamma(k)$ in quadratic mean (q.m) and almost-surely (a.s.).

To develop a useful limiting theory, one needs study the limiting distribution theorems of vectors $\mathbf{S}_n(k_{1:m}) \triangleq (S_n(k_1), \dots, S_n(k_m))$, with $k_{1:m} \triangleq (k_1, k_2, \dots, k_m)$ distinct integer numbers referred to as *scales*. The limiting distribution of the vector as well as the rate of convergence depends upon the value of H ; for $H \in (0, 3/4)$, a central limit theorem holds with a rate of \sqrt{n} . For $H \in (3/4, 1)$, we have a non-central limit theorem, with a rate that depends on H . By adapting the results in Leon and Guyon (1989), we have

PROPOSITION 3.2. *Assume that $X \triangleq \{X_t\}_{t \in \mathbb{R}^+}$ is a FBM.*

1. if $H \in (0, 3/4)$, then

$$\sqrt{n}(\mathbf{S}_n(k_{1:m}) - \gamma(k_{1:m})) \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \mathcal{N}_m(0, \mathbf{F}(k_1, \dots, k_m)),$$

where $\gamma(k_{1:m}) \triangleq (\gamma(k_1), \dots, \gamma(k_m))^T$ and $\mathbf{F}(k_{1:m}) = (F_{ij}(k_{1:m}))_{1 \leq i, j \leq m}$ where

$$F_{ij}(k_{1:m}) = \gamma(1)^2 (k_i k_j)^{2H} G_{i,j}(k_{1:m})$$

$$G_{i,j}(k_{1:m}) = 2d_{ij}(k_i k_j)^{-2H} \sum_{k=-\infty}^{+\infty} \left(|kd_{ij} - k_j|^{2H} + |kd_{ij} + k_i|^{2H} - |kd_{ij} + k_i - k_j|^{2H} - |kd_{ij}|^{2H} \right)^2,$$

where d_{ij} is the greatest common divisor (GCD) of k_i and k_j .

2. if $H \in (3/4, 1)$, then,

$$n^{2-2H}(\mathbf{S}_n(k_{1:m}) - \gamma(k_{1:m})) \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} H(2H-1)\gamma(1)^t (k_1^2 \mathcal{Z}_1, \dots, k_m^2 \mathcal{Z}_m),$$

where $\mathcal{Z}_1, \dots, \mathcal{Z}_m$ are identically distributed centered random variables called Rosenblatt variables.

Proposition 3.2 can be extended if X is disturbed by an additional white noise, a result which is of some interest when the series is subject to measurements errors. As mentioned above, the Hurst parameter can be estimated by estimating the slope of the intercept of $L_n(k_1, \dots, k_m) = (\log S_n(k_1), \dots, \log S_n(k_m))^T$ on $(\log(k_1), \dots, \log(k_m))$. Denote $V(k_{1:m}) = (\log(k_1) - m^{-1} \sum_{i=1}^m \log(k_i), \dots, \log(k_m) - m^{-1} \sum_{i=1}^m \log(k_i))^T$ and $\hat{H}_{1,N}$ the ordinary least square estimate of the slope, i.e.

$$\hat{H}_{1,N} = \frac{V(k_{1:m})^t L_n(k_{1:m})}{2\|V(k_{1:m})\|^2}$$

The sampling property of this estimator may be deduced in a straightforward manner from Proposition 3.2. More precisely,

PROPOSITION 3.3. *Assume that $X \triangleq \{X_t\}_{t \in \mathbb{R}^+}$ is a FBM.*

1. if $H \in (0, 3/4)$, then

$$\sqrt{n}(\widehat{H}_{1,n} - H) \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0; \sigma_1(k_1, \dots, k_m))^2$$

where $\sigma_1^2(k_{1:m}) = V(k_{1:m})^T \mathbf{G}(k_{1:m}) V(k_{1:m}) / (4\|V(k_{1:m})\|^4)$ and matrix $\mathbf{G}(k_{1:m})$ is defined in Proposition 3.2.

2. if $H \in (3/4, 1)$, with $(\mathcal{Z}_1, \dots, \mathcal{Z}_m)$ the process defined in proposition 3.2:

$$n^{2-2H}(\widehat{H}_{1,n} - H) \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \frac{H(2H-1)}{2} \frac{V^T(k_{1:m})}{\|V(k_1, \dots, k_m)\|^2} (n_1^{2H-2} \mathcal{Z}_1, \dots, n_m^{2H-2} \mathcal{Z}_m).$$

When $0 < H < 3/4$, it is possible to estimate $\mathbf{G}(k_{1:m})$ using the estimate $\widehat{H}_{1,n}$. Using this preliminary estimator of the covariance, it is possible to further reduce the variance of $\widehat{H}_{1,n}$ by computing a generalized least square estimate of the slope of the intercept. Some developments along these lines may be found in Bardet (1997) for Gaussian processes or in Poggi and Viano (1997) for linear processes.

Simulations show that the GLS estimate of the slope compares favorably with the maximum likelihood (MLE) estimate, for $H < 3/4$. This result is consistent with the limiting distribution theory. But, for $H > 3/4$, the regression estimator loose much efficiency (the rate of convergence is no longer optimal). To avoid this dichotomy, we have to define an alternate statistics, referred to as the *Generalized quadratic variation*.

3.2. GENERALIZED QUADRATIC VARIATIONS

Assume again that $\{X_t\}_{t \in \mathbb{R}}$ is a FBM observed at finitely many regularly sampled points $\{X_0, X_1, \dots, X_n\}$. For $p > 1$ and $Q \geq 2$ two positive integers, we define the set $M_p(Q)$ as

DEFINITION 3.4. : A p -uplet $\mathbf{u} = (u_1, u_2, \dots, u_p)$ of p real numbers belongs to $M_p(Q)$

if the Q first moments of \mathbf{u} vanish, i.e. $\sum_{i=1}^p i^q u_i = 0$ for $q \in \{0, 1, \dots, Q-1\}$ and

$$\sum_{i=1}^p i^Q u_i \neq 0.$$

For $\mathbf{u} \in M_p(Q)$, the \mathbf{u} -quadratic variations of $\{X_t\}_{t \in \mathbb{R}}$ at scale $k \in \mathbb{N}^*$, is the process $U_k(j)$ defined as

$$U_n(j) = \left(\sum_{i=1}^p u_i X_{k(i+j)} \right)^2 \text{ for } j = 0, 1, \dots, [n/k] - p.$$

These \mathbf{u} -quadratic variations are particular instances of generalized quadratic variations introduced by Istas and Lang (1997) for extending certain results obtained with ordinary quadratic variations (see Guyon and Leon (1989)). An interesting example \mathbf{u} -quadratic variation is given by $\mathbf{u} = (1, -2, 1)$ ($Q = 3$) which corresponds to increments of increments. The ordinary quadratic variations corresponds to $\mathbf{u} = (-1, 1)$ (associated with the increments). It may be shown that (see Istas and Lang (1997))

PROPOSITION 3.5. *Assume that $\{X_t\}$ is FBM and let $\mathbf{u} \in M_p(Q)$, with $Q \geq 2$. Then, for $k \in \mathbb{N}^*$, and $\forall j \in \{0, 1, \dots, [N/n] - p\}$:*

$$EU_k(j) = \sigma^2 C_u(H) k^{2H}$$

where $C_u(H) = -\frac{1}{2} \sum_{k=1}^p \sum_{k'=1}^p u_k u_{k'} |k - k'|^{2H}$.

As before, $\log E(U_k(j))$ is linear w.r.t the logarithm of the scale $\log k$, and the slope of the intercept is proportional to H . Using the results of Iatas and Lang (1997), it may be shown that the empirical mean $S_n(k)$ of the \mathbf{u} -quadratic variation, $S_n(k) \triangleq ([n/k] - p + 1)^{-1} \sum_{j=0}^{[n/k]-p} u_k u_{k'}$ is a natural estimator of $EU_k(j)$ in the sense that $S_n(k)$ converges to $EU_k(j)$ when $n \rightarrow +\infty$ and obeys a central limit theorem, for all $0 < H < 1$.

PROPOSITION 3.6. *Assume that $\{X_t\}_{t \in \mathbb{R}}$ is a FBM. Let $\mathbf{u} \in M_p(Q)$ ($Q \geq 2$) be such that $C_u(H) \neq 0$ for any $H \in]0, 1[$. Then:*

$$\sqrt{n} (\log S_n(k_i) - 2H \log k_i - \log \sigma^2 C_u(H))_{1 \leq i \leq m} \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \mathcal{N}_m(0, \mathbf{F}(k_{1:m})),$$

where $\mathbf{F}(k_{1:m}) = (F_{ij}(k_{1:m}))_{1 \leq i, j \leq m}$ is defined as

$$F_{ij}(k_{1:m}) = \frac{d_{ij}}{2k_i^2 k_j^2 C_u^2(H)} \sum_{l=-\infty}^{+\infty} \left(\sum_{k=1}^p \sum_{k'=1}^p u_k u_{k'} |k_i k - k_j k' + l d_{ij}|^{2H} \right)^2.$$

A similar result holds, as previously, when the samples X_i are contaminated by an additive observation noise.

REMARK 3.7. When $|k - k'| \rightarrow +\infty$, we have $\text{cov}(U_{k_i}(k), U_{k_j}(k')) = \mathcal{O}(|k_i k - k_j k'|^{4H-4Q})$, for $\mathbf{u} \in M_p(Q)$. When $Q \geq 2$, the covariance coefficient $\text{cov}(U_{k_i}(k), U_{k_j}(k'))$ is square integrable, whatever the value of H (provided that $0 < H < 1$, which is a pre-requisite for $U_{k_i}(k)$ to obey a central limit theorem. For $Q = 1$ (ordinary quadratic variations), the covariance coefficient are square summable only if $0 < H < 3/4$. Since the covariance coefficient of $U_{k_i}(k)$ and $U_{k_j}(k')$ for large $|k - k'|$ is a decreasing function of Q , it seems at a first sight attractive to increase Q (to force an increased decorrelation). Simulations show however that $Q = 2$ is, in general, the best choice

Using the previous results, it is possible, as before, to derive the limiting distribution of the OLS and the GLS estimate of H . More precisely (using the same notations as before)

PROPOSITION 3.8. *Under the same assumptions than in Proposition 3.5, the ordinary least squares estimator of H , $\widehat{H}_{1,n}$, is asymptotically normal, i.e.*

$$\sqrt{n} (\widehat{H}_{1,n} - H) \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \sigma_1^2(k_{1:m}))$$

where $\sigma_1^2(k_{1:m}) = \frac{V(k_{1:m})^T F(k_{1:m}) V(k_{1:m})}{4 \|V(k_{1:m})\|^4}$.

A similar result holds for the GLS estimator. Simulations show that the GLS estimator compares favourably with the MLE estimator, while being much more faster to compute.

APPENDIX A. SOME CLASSICAL MODELS OF LONG-RANGE DEPENDENT TIME SERIES: DEFINITION AND TERMINOLOGY

We describe in this section the Fractional Gaussian Noise (FGN) and the FARIMA series : these are the simplest models that display long-range dependence. The best way to introduce FGN is through its "parent" *fractional Brownian motion* (FBM) $\{B_H(t), t \geq 0\}$. FBM is a Gaussian process with mean 0, stationary increments, variance $EB_H^2(t) = t^{2H}$ and covariance $EB_H(s)B_H(t) = (s^{2H} + t^{2H} - |s - t|^{2H})/2$. The FGN series $\{X_i, i \geq 1\}$ is the increment of fractional Brownian motion, namely $X_i = B_H(i + 1) - B_H(i)$, $i \geq 1$. It is a zero-mean, stationary, Gaussian time-series whose auto-covariance function $\gamma(h) = EX_i X_{i+h}$ is given by $\gamma(h) = ((h + 1)^{2H} - 2h^{2H} + |h - 1|^{2H})/2$, $h \geq 0$. An important point about γ is

that it satisfies $\gamma(h) \simeq H(2H-1)h^{2H-2}$, as $h \rightarrow \infty$, when $H \neq 1/2$. The X_i 's are positively correlated when $1/2 < H < 1$, and we say that they display *long range dependence*. The index H is the *Hurst coefficient*, measures the *intensity* of long-range dependence. The spectral density is

$$f(x) = C_H (2 \sin(x/2))^2 \sum_{k=-\infty}^{\infty} |x + 2\pi k|^{-(2H+1)} \simeq C_H x^{1-2H} \quad x \rightarrow 0^+$$

where C_H is a constant. Besides FGN, we also consider FARIMA(0, d , 0). It is defined formally as $X_i = \Delta^{-d}\epsilon_i$, $i \geq 1$, where ϵ_i are i.i.d with mean 0 and finite variance, and where Δ is the differencing operator. A way to interpret $X_i = \Delta^{-d}\epsilon_i$ with a fractional value of d is as a moving average $X_i = \sum_{j=0}^{\infty} c_j \epsilon_{i-j}$, where $c_j = \Gamma(j+d)/(\Gamma(d)\Gamma(j+1))$, $j \geq 1$. Note that $c_j \simeq \Gamma(d)^{-1}j^{d-1}$ as $j \rightarrow \infty$ and that the autocovariance function γ satisfies $\gamma(h) \simeq C_d h^{2d-1}$, for $0 < d < 1/2$, where $C_d = \pi^{-1}\Gamma(1-2d)\sin(\pi d)$. For large h , the auto-covariance of a FARIMA(0, d , 0) has the same power decay as the auto-covariance of an FGN. Relating the two exponents give $d = H - 1/2$. The spectral density $f(x)$ of FARIMA(0, d , 0) also blows up at the origin like a power function $f(x) \simeq x^{-2d}/2\pi$, as $x! \rightarrow 0^+$. A FARIMA(0, d , 0) is a particular case of FARIMA(p , d , q) processes, a versatile family of models. FARIMA(p , d , q) is defined through the equations

$$X_i - \phi_1 X_{i-1} - \dots - \phi_p X_{i-p} = \Delta^{-d}\epsilon_i - \theta_1 \Delta^{-d}\epsilon_{i-1} - \dots - \theta_q \Delta^{-d}\epsilon_{i-q}$$

The FARIMA(p , d , q) series follow the same type of asymptotic relations for their auto-covariance and spectral density as do the FARIMA(0, d , 0) processes, although the actual functions are more complicated, and have additional short-term components. See Brockwell and Davis (1991), Beran (1994) and Samorodnitskyi and Taqqu (1994).

REFERENCES

- [1] P. ABRY, F. SELLAN (1996) *The wavelet-based synthesis for the fractional Brownian motion proposed by F. Sellan and Y. Meyer: remarks and fast implementation* Applied and Computational Harmonic Analysis Vol 3 pp. 377–383 1996.
- [2] J.M. BARDET (1997) *Testing self-similarity of processes with stationary increments*. Preprint. Orsay.
- [3] J.M. BARDET (1998) *Un test d'autosimilarité pour les processus gaussiens à accroissements stationnaires*. To appear.
- [4] J. BERAN, (1993) *Fitting long memory models by generalized regression*, Biometrika, 80, pp. 817–822.
- [5] J. BERAN, J. 1994. *Statistics for long memory processes*. Monographs on Stat. and Appli. Prob. 61. Chapman & Hall, 315 p.
- [6] R. DAHLHAUS (1989) *Efficient parameter estimation for self-similar processes*, The Annals of Statistics, 17, pp. 1749–1766.
- [7] J. GEWEKE AND S. PORTER-HUDAK (1983), *The estimation and application of long-memory time-series models*, Journal of Time Series Analysis, 4, pp. 221–238.
- [8] L.GIRAÏTIS, D. SURGAILIS (1990) *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. Field. 86, p. 87–104.
- [9] C. GRANGER, R. JOYEUX R (1980) *An introduction to long-memory series and fractional differencing*, J. of Time Series Analysis, pp. 15–30
- [10] P. HALL, H. KOUL, B. TURLACH (1998) *On estimating the self-similarity index for a long-range dependent process*. To appear.
- [11] C. HURVICH AND K. BELTRAO (1993), *Asymptotic for the low-frequency ordinates of the periodogram of a long-memory time-series*, J. Time Series Analysis, 14, pp. 455–472
- [12] C. HURVICH AND K. BELTRAO (1994), *Automatic semi-parametric estimation of the long-memory parameter of a long-memory time-series*, J. Time Series Analysis, 15, pp. 285–302
- [13] C. HURVICH, R. DEO, J. BRODSKI (1997), *The mean-square error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time-series*, J. Time Series Analysis.

- [14] X. GUYON, J. LEON (1989) *Convergence en loi des H -variations d'un processus gaussien stationnaire*. *Ann. Inst. Poincaré* 25, 265-282.
- [15] J. ISTAS, G. LANG (1997) *Quadratic variations and estimation of the local Hölder index of a Gaussian process*. *Ann. Inst. Poincaré* 33, 407-436.
- [16] B. MANDELBROT, M. TAQQU (1979) *Robust R/S analysis of long run serial correlation*. Proc. 42nd Session of the ISI, Manila. 2, p. 69-100.
- [17] J.M. POGGI, M. VIANO (1997) *An estimate of the fractal index using multiscale aggregates* J. of Time Ser. Anal.
- [18] K. LI (1997), *Asymptotic optimality for c_p , c_l , cross-validation and generalized cross-validation: discrete time index*, The Annals of Statistics, 15, pp. 958–975.
- [19] E. MOULINES AND P. SOULIER, (1997) *Log-Periodogram of time-series with long-range dependence*, Tech. Rep. Université d'Evry-Val d'Essonne, 1997, Accepted for publication in the *Annals of Statistics*
- [20] ———(1998), *Data driven order selection for projection estimator of the spectral density of time series with long-range dependence*, Tech. Rep. Université d'Evry Val d'Essonne, Accepted for publication in the *J. of Time Series Analysis*.
- [21] P. ROBINSON (1991), *Nonparametric and semi-parametric methods in econometrics and statistics*, in Proceedings of the Fifth International symposium in economic theory and econometrics, W. Barnett, J. Powell, and G. Tauchen, eds., 1991, pp. 437–457.
- [22] ———(1994a), *Time series with strong dependence*, vol. 1, Cambridge University Press, 1994, ch. 2 of *Advances in Econometrics*, Sixth World Congress, pp. 47–95.
- [23] ———(1994b), *Semiparametric analysis of long-memory time series*, The Annals of Statistics, 22 (1994), pp. 515–539.
- [24] ———(1995a), *Gaussian semi-parametric estimation of long-range dependence*, The Annals of Statistics, 23, pp. 1630–1661.
- [25] ———(1995b), *Log-periodogram regression of time-series with long-range dependence*, The Annals of Statistics, 23, pp. 1630–1661.
- [26] A. TSYBAKOV AND B. POLYAK (1989), *Asymptotic optimality of the c_p test for the orthogonal series estimation of regression*, Theory of Probability and Their Applications, 35 (1989), pp. 293–306.
- [27] G. SAMORODNITSKY, M. TAQQU (1994), *Stable non-Gaussian random processes*. Stochastic modeling. Chapman & Hall, 636 p.
- [28] M. TAQQU, V. TEVEROVSKY, W. WILLINGER (1995), *Estimators for long-range dependence: an empirical study*. *Fractals* 3, 785-798.

