Semiparametric spectral estimation for fractional processes

Eric Moulines
Ecole nationale supérieure
des Télécommunications
46, rue Barrault,
75013 Paris, France

Philippe Soulier
Université d'Evry Val d'Essonne
Département de Mathématiques
91025 Evry Cedex, France

January 25, 2000

Abstract

In this contribution, recent methodological and theoretical developments in the field of semiparametric estimation for fractional processes are presented. Fractional processes are a special case of long range dependent processes. The main subject is the semiparametric estimation of the fractional differencing coefficient or memory parameter. Emphasis is put on two directions : non stationary processes and minimax and adaptive estimation.

Contents

1 Introduction : spectral methods for fractional processes

This papers is concerned with the presentation of new methods and results in the field of semiparametric estimation of long range dependent processes. Robinson's seminal papers of 1994 and 1995 [44, 46, 47] and his review paper [45] presented the first rigourous treatment of the asymptotic theory of semiparametric estimators of the memory parmeter, among which the famous GPH estimator which had been proposed as early as 1983. The theoretical, methodological and technical ideas contained in these papers were so powerful that they induced developments of the theory in many directions. First, new estimators of the long memory parameters were introduced and their asymptotic theory was either direct applications or refinements of some of Robinson's results.

An important new direction of investigation was the extension of the class of fractional processes to non incertible or non stationary processes. This was first investigated by Hurvich and Ray (1995) [31] and later by Velasco in a series of papers [52, 54]. The key ingredient to study such processes is tapering. The classical tapers such as the cosine bell taper, Kolmogorov tapers were investigated, and recently, a new class of complex tapers, very well suited to the study of fractional processes was introduced by Hurvich and Chen (1999) [27].

Models with a fractional pole outside zero were introduced very naturally as early as 1989 by Gray Zhang and Woodward (1989) [19] and further extensions followed. Suprisingly, this direction was not investigated from a statistical point of view until very recently, and no papers have been published yet on the problems of estimation of the fractional pole and of the fractional differencing coefficient when the location of the fractional pole is unknown.

Giraitis, Robinson and Samarov (1997,2000) [15, 16] initiated the investigation of minimax and adaptive estimation of the fractional differencing coefficient. research in this direction is very important since all the semiparametric estimators of the fractional coefficients involve parameters which depend on some kind of smoothness of the spectral density and these parameters must be selected automatically from the data. The adaptive estimation theory, still relatively new provides very simple and powerful tools to select these parameters, which will probably be widely prefered to classical selection procedures.

The distribution of the processes for which a satisfactory theory is established is rather restircted. The theory, if not the method, of adaptive estimation is restricted to Gaussian processes. The weak convergence of some estimators has been recently extended to the class of linear non-Gaussian processes. Very recently, non linear models exhibiting long range dependence have been introduced (cf. Giraitis, Robinson and Surgailis (1999) [17]), but there is as yet no statistical theory for these processes. Fractionally integrated processes with infinte variance have been introduced by Kokoszka and Taqqu (1995) [35], but semiparametric estimation has not yet been considered for these processes.

This contribution aims at presenting an overview of the problems described above, with an emphasis on two directions : non invertible and non stationary processes and minimax and adaptive estimation. It will not be an exhaustive review of all the existing literature, and the last two problems metioned (infinite variance and non linear processes) will be omitted.

The rest of this paper is organized as follows. Section 2 presents the fractional models under consideration. Section 3 presents the specific tools and methodology of spectral estimation for fractional processes. Sections 4 and 5 respectively present the two families of semiparametric estimators of the fractional differencing coefficient d, namely the local and global estimators in the case where the fractional pole is known to be zero. Section 6 presents the asymptotic theory of these estimators. Emphasis is put on minimax estimation theory and on the relatively new theory of adaptive estimation. Section 9 presents some of the main probability tools and technical results that are used in the derivations of section 6

2 Fractional processes

Let $Y = \{Y_t\}_{t\in\mathbb{Z}}$ be a covariance stationary covariance process, with mean μ, covariance $\gamma(\tau) := \mathrm{cov}(Y_{t+\tau}, Y_t)$ and spectral density $f^*(x)$. For $d < 1/2$, let $X = \{X_t\}_{t\in\mathbb{Z}}$ be the covariance stationary process defined by

$$(1-B)^d X_t = Y_t \tag{1}$$

where B is the backshift operator,

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

and Γ denotes the Gamma function. The process $\{X_t\}_{t\in\mathbb{Z}}$ is said to be a fractionally integrated process with fractional differencing parameter d. This class of processes was introduced by Granger and Joyeux (1980) [18] and Hosking (1981) [23]. Allowing d to take fractional values produces a fundamental change in the correlation structure that a fractional process can have when compared with the correlation structure of a "standard" time-series model, such as an ARMA(p,q) process. The covariance coefficients $\rho(\tau)$ of a fractional process decline at a hyperbolic rate (see Brockwell and Davis, (1991) [9], Theorem 13.2.2)

$$\rho(\tau) = O(\tau^{-1+2d})$$

while the autocovariance of a stationary ARMA process decays exponentially. For $0 < d < 1/2$, a fractional process is said to have long-memory or long-range dependence and its covariance sequence $\rho(j)$ is not summable. For $d < 0$, the process is said to have *intermediate memory*, in the sense that the covariance coefficients $\rho(j)$ are summable, though declining at an hyperbolic rate. The spectral density of the process X defined by (1) is given by

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

When $0 < d < 1/2$, the spectral density is unbounded at zero, whereas the spectral density of an ARMA process is bounded.

A popular class of fractional processes, introduced by Granger and Joyeux (1980) [18] is the class of ARFIMA model (standing for AutoRegressive Fractionaly Integrated Moving Average), in which Y is a causal invertible ARMA(p,q) process. Another class of fractional process of interest is the class of FEXP models (standing for fractionaly integrated exponential models), in which the spectral density f^* of Y is modeled as the exponential of a finite order trigonometric polynomial, *i.e.* there exist coefficients θ_j, $1 \le j \le p$ such that

$$f^*(x) = \exp\{\sum_{j=0}^{q-1} \theta_j \cos(jx). \tag{3}$$

A fractional process with spectral density f satisfying (2) with f^* as in (3) will be referred to in the sequel as a FEXP(q) process. Such processes, which generalize the so-called Bloomfield exponential models (cf. Bloomfield (1973) [7]), have been proposed as an alternative to ARFIMA processes by Beran (1993) [4] and Robinson (1994) [45].

The model defined by (1) has been recently generalized to fractional models with a fractional pole outside the origin. Gray, Zhang and Woodward (1989) [19] and Viano, Deniau and Oppenheim (1995) [55] among others have considered general fractional models which allow complex as well as real roots on the unit circle. The generic form is

$$\lambda(B)X_t = Y_t \tag{4}$$

where, as above, Y_t is a covariance stationary process with positive and finite spectral density and $\lambda(B)$ is given by

$$\lambda(B) = (1 - 2\cos(\lambda)B + B^2)^d,$$

where $d < 1$ if $\lambda \in (0, \pi)$ specifies the frequency at which the singularity occurs. If $0 < d < 1/2$, the process X is stationary and long-memory. When Y is a white noise, X is referred to as a Gegenbauer process. Gray, Zhang and Woodward (1989) [19] have shown that, for $d < 1/2$ and smooth enough, $f^*(x)$ the autocovariance function and the spectral density of Gegenbauer processes are given by

$$f(x) = [4\{\cos(x) - \cos(\lambda)\}^2]^{-d} f^*(x), \tag{5}$$

$$\rho(\tau) = c\tau^{2d-1}\cos(2\pi\lambda\tau)(1 + o(1)). \tag{6}$$

When $0 < d < 1/2$, Gegenbauer processes are particularly appropriate for data with slowly damping autocovariances which also have cyclic patterns. Gray, Woodward and Zhang (1989) [19] further developped the model through the inclusion of more than one Gengenbauer factor.

An important generalization of the previous models is to allow values of the fractional differencing parameter d larger than $1/2$ (see Hurvich and Ray (1995) [31], Beran, Bhansali and Ocker (1998) [5], Velasco (1999) [52, 54]). Note however that in such case the definition (1) cannot be readily applied. We follow here the construction presented Hurvich and Ray (1995) [31]. A process X is said to be a fractional process of order $d \in]-1/2, p + 1/2[$, where $p \in \mathbb{N}$, if the p-th order difference of this process $\tilde{X}_t = (1 - B)^p X_t$ satisfies equation (1) with $\tilde{d} := d - p < 1/2$. Note that this class include processes with deterministic polynomial trend of order $(p - 1)$. Differencing arbitrarily p times a process X in this class yields a stationary but possibly non-invertible process. Hence, it is important in practice to derive procedures which are able to deal with possibly non-invertible processes of order $-p - 1/2 < d < 1/2$.

3 Spectral estimation

In classical time series analysis, it is customary to distinguish time domain and spectral domain approaches. Roughly speaking, time domain methods are based on empirical estimates autocovariance function (or other time-domain quantities, such as running mean empirical variance) while spectral methods are based on the discrete Fourier transform coefficients of the observed data. In the context of fractional processes, spectral methods have been shown very convenient and powerful. We will describe in this section the various problems, specific tools and methodology of spectral estimation.

3.1 The periodogram

The oldest and most natural tool of spectral estimation is the periodogram. Given an observation $X_1, \cdots, X_n$, the ordinary discrete Fourier transform (DFT) and the periodogram are respectively defined as

$$d_n^X(x) = (2\pi n)^{-1/2} \sum_{t=1}^{n} X_t e^{itx}, \tag{7}$$

$$I_n^X(x) = |d_n^X(x)|^2 = (2\pi n)^{-1} \Big|\sum_{t=1}^{n} X_t e^{itx}\Big|^2. \tag{8}$$

It is well known that the periodogram ordinates of a standard Gaussian white noise evaluated at Fourier frequencies $x_k = 2k\pi/n$, $1 \le k \le \tilde{n} := [(n-1)/2]$ form a sequence of i.i.d. r.v. distributed as standard exponentials. This property no longer holds when the process X is not a Gaussian white noise. However, under miscellaneous "weak dependence" conditions, it holds that

- the periodogram is an asymptotically unbiased estimate of the spectral density, *i.e.* $\mathbb{E}[I_n^X(x_k)] = f(x_k) + O(n^{-1})$, $1 \le k \le [(n-1)/2]$, where the $O(n^{-1})$ term is uniform in k,
- the periodogram ordinates are asymptotically uncorrelated, $\mathrm{cov}(I_n^X(x_k), I_n^X(x_l)) = O(n^{-1})$, for $1 \le k \ne l \le [n/2]$, where the $O(n^{-1})$ is uniform in k, l,
- For any given u, and any u-tuple of distinct integers $(k_1, k_2, \cdots, k_u)$,

  $$I_n(x_{k_1})/f(x_{k_1}), \cdots, I_n(x_{k_u})/f(x_{k_u}$$

  are asymptotically independent standard exponentials.

Note that the latter property is no longer true when considering an increasing number of Fourier frequencies.

For fractional long-memory processes ($0 < d < 1/2$, $x_0 = 0$), Künsch (1986) [36] and later Hurvich and Beltrao (1993) [24] proved that none of the above mentioned properties remain valid. Indeed for a fractional process with spectral density verifying (2), for fixed $1 \le k < j \le \tilde{n}$, it holds that

$$\lim_{n\to\infty} |\mathbb{E}[I_n(x_k)/f(x_k)] - 1| \ne 0, \tag{9}$$

$$\lim_{n\to\infty} |\mathrm{cov}(I_n(x_k)/f(x_k), I_n(x_j)/f(x_j))| \ne 0. \tag{10}$$

See Hurvich and Beltrao (1993) [24], Lemma xx for an expression of these limits. Note that because the spectral density can be either infinite ($d > 0$) or zero ($d < 0$) at zero, it is more appropriate to consider the "normalized" periodogram, *i.e.* the raw periodogram normalized by the inverse of the spectral density. (9) shows that the bias (generally) does not vanish for large n. Nevertheless, under appropriate conditions, one may show (see section 9) that there exists a sequence $r(f;k)$ verifying for all n, and all $1 \le k \le \tilde{n}$,

$$|\mathbb{E}[I_n(x_k)/f(x_k)] - 1| \le r(f;k)$$

and such that $\lim_{k\to\infty} r(f;k) = 0$, which means that the bias is small for frequencies sufficiently far away from zero. Similarly, (10) shows that the normalized periodogram ordinates are not asymptotically uncorrelated. However, as shown *e.g.* by Robinson (1995) [47] or Moulines and Soulier (1999) [41], under appropriate technical conditions, there exists a sequence $r(f;k,j)$ such that, for all n and all $1 \le k < j \le \tilde{n}$,

$$|\mathrm{cov}(I_n(x_k)/f(x_k), I_n(x_j)/f(x_j))| \le r(f;k,j).$$

The fact that this bound does not vanish for large n makes the derivations more intricate, but what really matters is that, one can still obtains bounds such as

$$\sum_{1\le k<j\le \tilde{n}} r(f;k,j) = o(n^{\epsilon}),$$

for some relevant $\epsilon > 0$.

Finally, for any given u, and any u-tuples $(k_1, k_2, \cdots, k_u)$, the r.v. $[I_n(x_{k_1})/f(x_{k_1}), \cdots, I_n(x_{k_u})/f(x_{k_u})]$ converges in distribution to

$$(Z_1^2 + Z_2^2)/2, (Z_3^2 + Z_4^2)/2, \cdots, (Z_{2u-1}^2 + Z_{2u}^2)/2$$

where $Z := (Z_1, Z_2, \cdots, Z_{2u})$ is a zero-mean multivariate Gaussian vector with non-singular (non diagonal) covariance matrix $\Gamma(k_1, \cdots, k_u)$ (see Deo (1997) [11], corollary 3 for an expression of this matrix).

3.1.1 Tapering

In some situations, it is required to have a better control of the bias term or the covariance term: a classical way to achieve this objective in spectral analysis is to use a data taper, i.e. to apply a taper on the observed data prior to computing the discrete Fourier transform. Another motivation to use a taper is to deal with (possibly) over-differentiated processes, *i.e.* time-series for which d can be less that $-1/2$: such values of the fractional differencing coefficients can be observed as a consequence of a pre-processing consisting in differencing the original time-series in order to combat either non-stationarity in the data (*e.g.* fractionally integrated time-series) or the presence of polynomial trends. Such an approach has already been suggested by Hurvich and Ray (1995) [31], Velasco (1999) [52, 53, 54] and Hurvich and Chen (1999) [27] (see also Deo and Hurvich (1999) [28] who show that tapering can be helpful for estimating the mean of a potentially over-differenced long-memory time-series) [1].

Let $(h_{t,n})_{1\le t\le n}$ be a triangular array of complex numbers. The tapered discrete Fourier transform and the tapered periodogram of $(X_1, \cdots, X_n)$ are respectively defined as

$$d_{h,n}^X(x) := (2\pi \sum_{t=1}^{n} |h_{t,n}|^2)^{-1/2} \sum_{t=1}^{n} h_{t,n} X_t e^{itx} \text{ and } I_{h,n}^X(x) := |\omega_n(x)|^2. \tag{11}$$

Velasco (1999) [52, 53, 54] has considered several tapering schemes such as the cosine bell taper and the Kolmogorov tapers. Hurvich and Chen (1999) [27] have considered a novel family of data taper depending on a single control parameter p, referred to as the taper order. This family of tapers is specified as follows. Define

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

[1] When the G-frequency $x_0 \neq 0$, this type of behavior may occur as a consequence of filtering the observed data with a filter having a zero at $e^{\pm j x_0}$

and, for any integer $p \geq 0$, the tapered DFT of order p

$$d^X_{p,n}(x) := (2\pi n a_p)^{-1/2} \sum_{t=1}^{n} h^p_{t,n} X_t e^{itx}, \quad I^X_{p,n}(x) := |d^X_{p,n}(x)|^2 \tag{12}$$

where the subscript p (with a slight abuse of notation) denotes the taper order and $a_p := n^{-1} \sum_{t=1}^n |h_{t,n}|^{2p} = \binom{2p}{p}$ is the normalization constant. Note that, contrary to most commonly used data tapers, this taper is complex-valued. For $p = 0$, the ordinary periodogram (8) is obtained. Define for $p \in \mathbb{N}$, $D_{p,n}(x)$ the normalized kernel function $D_{p,n}(x) := (2\pi n a_p)^{-1/2} \sum_{t=1}^n h_{t,n} \exp(itx)$. It follows from elementary calculations that

$$D_{p,n}(x) = (2\pi n a_p)^{-1/2} \sum_{k=0}^{p} \binom{p}{k} (-1)^k \exp(-ik\pi/n) D_n(x + x_k). \tag{13}$$

where $D_n(x) := \sum_{t=1}^n e^{-ixt}$ denotes the Dirichlet kernel. The latter relation implies that $D_{p,n}(x_k) = 0$ for $k \in \{1, \cdots, [(n-2p-1)/2]\}$, so that the tapered Fourier transform is invariant to shift in the mean. This invariance is achieved without the need to restrict attention to a coarse grid of Fourier frequencies, as is necessary for the Kolmogorov taper applied to the original (non-differentiated) time-series (see Velasco (1999) [52]). As shown in Hurvich and Chen (1999) [27], the "decay rate" of the kernel in the frequency domain increases with the kernel order. This property means that high-order kernel are more effective to control leakage, which is of utmost importance when dealing with non-invertible (over-differentiated) series. For $p = 0$, DFT ordinates of white noise at any two Fourier frequencies are uncorrelated. This property is unfortunately lost by tapering. For $p \geq 1$, the correlation of the DFT ordinates of a white noise sequence $X_1, \cdots, X_n$ at Fourier frequencies x_k, x_j, $1 \leq k < j \leq [n/2]$ does not vanish for $(j-k) \leq p$.

3.1.2 Pooling

When considering non linear transformations of the periodogram (such as the log-periodogram), it is appropriate to consider the pooled periodogram, as introduced by Hannan (1970) [20] and later used by Robinson (1995) [47]. Pooling consists in computing running averages of periodogram values along blocks of size m, prior to applying the non-linear transformation. In order to guarantee independence of the tapered periodogram ordinates, for a data taper of order p, we will pool $m > p$ successive values of the periodogram, and we drop, at the end of the block, p discrete Fourier transform coefficients. More precisely, define $K(m,n,p) = [(n/2 - p)/(m + p - 1)]$ and for $k \in \{1, \cdots, K(m,n,p)\}$, $\mathcal{J}_{m,p,k} = \{(m+p-1)(k-1)+1, \cdots, (m+p-1)(k-1)+m\}$. The pooled periodogram is defined as

$$\bar{I}_{m,p,n,k} = \sum_{j \in \mathcal{J}_{m,p,k}} I^X_{p,n}(x_j).$$

For $p \geq 1$ and $m \geq 2$, the distribution of $\bar{I}_{m,p,n,k}$ is not a central chi-square, even when $(X_1, \cdots, X_n)$ is a white Gaussian noise. In such case however, it is possible to determine the distribution of $\bar{I}_{m,p,n,k}$ exactly. It follows from (13) that

$$d^X_{p,n}(x_k) = a_p^{-1/2} \sum_{j=0}^{p} \binom{p}{j} (-1)^j \exp(-ij\pi/n) U_{n,k+j}.$$

where $U_{n,l} := d^X_n(x_l)$. The r.v.'s $U_{n,l}$ are complex Gaussian with zero mean and, for $0 \leq j, l \leq \tilde{n}$, $\mathbb{E}[U_{n,j}\bar{U}_{n,l}] = (2\pi)^{-1}\delta_{j,l}$ and $\mathbb{E}[U_{n,j}U_{n,l}] = 0$, where δ is the Kronecker symbol. Hence, we have, for

$1 \leq l, j \leq \tilde{n}$, $\mathbb{E}[d^X_{p,n}(x_l) d^X_{p,n}(x_j)] = 0$, $\mathbb{E}[d^X_{p,n}(x_j) \bar{d}^X_{p,n}(x_l)] = 0$, if $|l-j| > p$ and

$$\begin{aligned}\mathbb{E}[d^X_{p,n}(x_j) \bar{d}^X_{p,n}(x_l)] &= (2\pi)^{-1} a_p^{-1} (-1)^{l-j} \sum_{u=0}^{p-l+j} \binom{p}{j-l+u} \binom{p}{u} \exp(-i((l-j)+2u)\pi/n) \\ &=: \varsigma_{p,n}(l-j) \end{aligned} \tag{14}$$

if $|l-j| \leq p$. When $X := (X_k)_{k\in\mathbb{Z}}$ is a white Gaussian noise, $\bar{I}_{m,p,n,k}$ is distributed as $\|W_{m,p,n}\|^2/2$ where $W_{m,p,n} = [W^{(1)}_{m,p,n}, W^{(2)}_{m,p,n}, \cdots, W^{(2m)}_{m,p,n}]$ is a $2m$-dimensional zero-mean Gaussian vector with covariance matrix $\Sigma_{m,p,n}$, with entries $[\Sigma_{n,m,p}]_{u,v}$, $1 \leq u, v \leq 2m$ defined as

$$[\Sigma_{m,p,n}]_{2u-1,2v-1} = [\Sigma_{m,p,n}]_{2u,2v} = \varsigma_{p,n}(u-v)/2, \tag{15}$$

$$[\Sigma_{n,m,p}]_{2u-1,2v} = [\Sigma_{m,p,n}]_{2u,2v-1} = 0. \tag{16}$$

where we have set $\varsigma_{p,n}(t) = 0$ for $t \geq p$. It is worthwhile to note that this covariance matrix does not depend on the frequency index k and that, as n goes to infinity, $\Sigma_{n,m,p}$ converges to $\Sigma_{m,p}$ defined as in (15) with

$$\varsigma_p(v) = \lim_{n\to\infty} \varsigma_{n,p}(v) = (2\pi)^{-1} a_p^{-1} (-1)^v \sum_{u=0}^{p-v} \binom{p}{v+u} \binom{p}{u} = (2\pi)^{-1} a_p^{-1} (-1)^v \binom{2p}{p+v}.$$

Hence, the r.v. $W_{m,p,n}$ converges in distribution to $W_{m,p}$, where $W_{m,p}$ is a zero-mean multivariate Gaussian r.v. with covariance $\Sigma_{m,p}$ and, by the continuous mapping theorem, $\bar{I}_{m,p,n,k}$ converges in distribution to $\|W_{m,p}\|^2/2$. We will use the following notations. Let ϕ be a function such that $\mathbb{E}|\phi(\|W_{m,p}\|^2/2)|^2 < \infty$. Define

$$\gamma_{m,p,n}(\phi) = \mathbb{E}[\phi(\|W_{m,p,n}\|^2/2)], \quad \sigma^2_{m,p,n}(\phi) = \mathrm{var}[\phi(\|W_{m,p,n}\|^2/2)], \tag{17}$$

$$\gamma_{m,p}(\phi) = \mathbb{E}[\phi(\|W_{m,p}\|^2/2)], \quad \sigma^2_{m,p}(\phi) = \mathrm{var}[\phi(\|W_{m,p}\|^2/2)]. \tag{18}$$

For $p = 0$, and $\phi(x) = \log(x)$ it is well known that $\gamma_{m,0}(\phi) = \psi(m)$ and $\sigma^2_{m,0}(\phi) = \psi'(m)$ where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function (see for instance Johnson and Kotz (1970) [33]). In the sequel, since we will mainly state results for the logarithm, we denote $\gamma_{m,p} := \gamma_{m,p}(\log)$ and $\sigma^2_{m,p} := \sigma^2_{m,p}(\log)$ (resp. $\gamma_{m,p,n}$ and $\sigma_{m,p,n}$).

3.2 Semiparametric estimation

In the semi-parametric setting, a full parametric model is not specified for the spectral density f of the observed process X, but the parameter of interest is finite dimensional. In the context of fractional processes, the most important semiparametric problem is the estimation of the fractional differencing parameter d and of the G-frequency when unknown. f^* is then considered as an infinite dimensional nuisance parameter.

A very classical way to obtain a semiparametric estimator is to consider an increasing family of parametric models Θ_p, $p \in \mathbb{N}^*$, and for a given sample size n, to fit a parametric model of size p_n, where p_n is an increasing sequence of integers. This type of method yields simultaneously an estimator of the spectral density (and of any functional thereof) over the whole frequency range $[-\pi, \pi]$. These methods are therefore referred to as *global methods*. The difficulty of such methods is to find a criterion to choose p_n

from the data, a problem known as model selection. In the previous section, several parametric models were descirbed. There are many ways to estimate efficiently (or with a reasonable loss of efficiency) parameters of a parametric model. It is well known however that there are specific contrast functions which are particularly well suited to a parametric model (see *e.g.* Taniguchi (1987) [50]). The parametric models of fractional processes described above include in particular the fractional autoregressive (FAR), which is a particular case of the ARFIMA model, and the fractional exponential model (FEXP). For these models, the natural contrast functions are respectively Whittle's contrast and the logarithmic contrast. These contrast functions define minimum contrast estimators (or M-estimators) by minimizing an empirical contrast function :

$$\theta \to \mathcal{K}_n(\theta) = n^{-1} \sum_{i=1}^{[(n-1)/2]} k(f_\theta(x_i), I_n(x_i)), \tag{19}$$

where f_θ is the spectral density corresponding to the value of the parameter θ, and $k(u,v) = \log(u) + v/u$ in the case of Whittle's contrast and $k(u,v) = (\log(u) - \log(v))^2$ in the case of the logarithmic contrast. Whittle's contrast is well suited to finite order non fractional AR models when the parameters are the coefficients of the regression. In the case of a FAR process, the minimization procedure is not so simple, but still reasonably complex, see section 5.1 below. The logartihmic contrast yields an estimator which is extremely simple for a finite order FEXP process when the parameters to estimate are d and the coefficients of the trigonometric polynomial defining the spectral density (cf. (3)), since it has a linear explicit expression with respect to the log-periodogram ordinates, see section 5.2 below.

When the only parameter of interest is d, the drawback of global methods is that they may seem to imply unnecessary assumptions on the spectral density. Since, for instance, in the case of a known fractional pole at zero, d determines the behaviour of f at zero, it seems natural to try to obtain a consistant estimator of d without any prior knowledge about f outside an arbitrarily small neighborhood of zero. These methods are therefore referred to as *local methods*. Whittle's contrast yields the Gaussian semiparametric estimator (GSE) proposed by Künsch (1987) [37] and the logarithmic contrast yields the GPH estimator of Geweke and Porter-Hudak (1983) [14]. (see sections 4.1 and 4.2 below).

4 Semi-parametric spectral estimation of d : local methods

The so-called local methods aim at constructing estimators that are consistent without any restrictions on $f^*(x)$ away from zero, apart from integrability on $[-\pi, +\pi]$. The range of applications of local methods will thus by nature be limited to the estimation of the fractional differencing coefficient d and, incidentally, to the estimation of $f^*(0)$. Since the behavior of the spectral density is specified only in a neighborhood of the zero frequency, local estimators use only periodogram ordinates belonging to this neighborhood. The idea of using local techniques has been initiated in the early work of Geweke and Porter-Hudak (1983) [14], who proposed to estimate d by regressing $-2d\log(x) + C$ on the first M_n log-periodogram ordinates (where M_n is a non-decreasing sequence), leading to the so-called GPH estimator, which is most famous estimator of the fractional differencing coefficient. Künsch (1987) [37] later introduced the so-called Gaussian semi-parametric estimator (GSE), which is a local version of the discrete form of the Whittle likelihood. Other, but less popular local methods include the averaged periodogram, proposed by Robinson (1994) [44], (see also, Lobato, 1997 [39]), the smoothed periodogram estimate, introduced by Reisen (1994) [43], and the log-spectral density regression estimators (see Azaïs and Lang, 1999, [38]). We will describe in detail the GSE and the GPH estimators, and will briefly review other methods proposed in an already vast literature.

4.1 The Gaussian semi-parametric estimator

The discrete local Whittle contrast function is defined as

$$Q_M(C,d) = M^{-1}\sum_{k=1}^{M}\left\{\log(C|1-e^{ix_k}|^{-2d}) + \frac{I_{p,n}^X(x_k)}{C|1-e^{ix_j}|^{-2d}}\right\} \tag{20}$$

where $M := M_n$, the trimming number, is a non-decreasing sequence such that $\lim_{n\to\infty}(M_n^{-1}+M_n/n) = 0$. Note that, since $I_{p,n}^X(x_k)$, for $1 \leq j \leq [(n-1)/2]$ is invariant to shift in the mean, the Whittle contrast function also is invariant, which is a significant advantage in the long-memory context. The integral form of the Whittle likelihood, which is not mean-shift invariant, is thus not used. Compared to the exact form Whittle' contrast fucntion (19), the two main differences are **(i)** the range of the sum is $\{1,\cdots,M\}$, *i.e.* only the Fourier frequencies in a neighborhood of the zero frequency are taken into account, and **(ii)** the spectral density is replaced by its equivalent at zero. Note that pooling is irrelevant since Whittle's contrast is based on a linear functional of the periodogram.

Replacing in (20) the unknown constant C by its estimate $\hat{C}_M(d)$ (obtained for a given value of d),

$$\hat{C}_M(d) = M^{-1}\sum_{j=1}^{M} I_{p,n}(x_j)|1-e^{ix_j}|^{2d}$$

yields the following profile likelihood

$$R(M;d) := Q_M(\hat{C}_M(d),d) = \log\left(M^{-1}\sum_{j=1}^{M}|1-e^{ix_j}|^{2d} I_{p,n}(x_j)\right) + dM^{-1}\sum_{j=1}^{M} g(x_j). \tag{21}$$

Let $[\Delta_1,\Delta_2] \subset]-p-1/2, 1/2[$ be a closed interval of admissible estimates of d (note that the lower bound depends on the kernel order p; see below). Δ_1 and Δ_2 can be chosen arbitrarily close to $-p-1/2$ and $1/2$ or can reflect a priori knowledge on d. For instance, $\Delta_1 > -1/2$ precludes non-invertibility. The local Whittle estimator of d is defined as the value of $d \in [\Delta_1,\Delta_2]$ which minimizes (21).

$$\hat{d}_p^{\mathrm{GSE}}(M) := \arg\min_{\bar{d}\in[\Delta_1,\Delta_2]} R(M,\bar{d}). \tag{22}$$

This can be done by searching this minimum over a grid, or by using any form of one-dimensional minimization algorithms.

4.2 The GPH estimator

Consider the pooled periodogram $\bar{I}_{m,p,n,k}$ (see section 3.1.2), and define

$$Y_k = \log(\bar{I}_{m,p,n,k}) - \gamma_{m,p,n}, \tag{23}$$

Given a trimming number M, the local log-periodogram regression estimator is defined as the minimum of the local least-squares criterion,

$$(\hat{d}_M^{GPH}, \hat{c}_M)) = \arg\min_{\bar{d},\bar{c}} \sum_{k=1}^{M} (Y_k - \bar{d}g(y_k) - \bar{c})^2, \tag{24}$$

where

$$y_k := (2k-1)\pi/2K(m,p,n), \quad 1 \leq k \leq K(m,p,n), \tag{25}$$

$$g(x) = -2\log|1-e^{ix}| = 2\sum_{j=1}^{\infty} \frac{\cos(jx)}{j}. \tag{26}$$

Contrary to the GSE estimator, (24) can be solved in close form

$$\hat{d}_{m,p}^{\mathrm{GPH}}(M) = \sum_{k=1}^{M} \beta_k(M) Y_k, \tag{27}$$

$$\text{with} \quad \beta_k(M) = \frac{(g(y_k) - \bar{g}_M)}{\sum_{k=1}^{M}(g(y_k) - \bar{g}_M)^2}, \quad \text{and} \quad \bar{g}_M = \sum_{k=1}^{M} g(y_k). \tag{28}$$

Note that this estimator is also mean-shift invariant. This estimator was first introduced (with $p = 0$, $m = 1$) by Geweke and Porter-Hudak (GPH) (1983) [14].

4.3 Other local estimators

4.3.1 The average periodogram

The averaged periodogram method, introduced by Robinson (1994a) [44] (see also Lobato (1997) [39] for multi-dimensional extensions) also is based upon the idea of regressing the low-frequency periodogram ordinates. The approach is however slightly different because it is based on an estimate of a statistics referred to as the averaged periodogram,

$$\hat{F}_{p,n}(x) = \frac{2\pi}{n} \sum_{j=1}^{[nx/2\pi]} I_{p,n}^X(x_j) \tag{29}$$

Note again that $F_{p,n}(x)$ is mean-shift invariant. There are many available statistical results on the averaged periodogram, including pointwise converge (for fixed x), and functional weak limit theorem of Donsker's type (on $[0,\pi]$). Most of these results assume that X_t is weakly dependent (e.g. having spectral density which is smooth and bounded). Robinson (1994a) [44] shows that (with $p = 0$, $0 < d < 1/2$) provided that

$$f(x) = L(1/x)x^{-2d}, \; x \to 0^+ \tag{30}$$

where L is a slowly-varying function at infinity, and additional conditions on the dependence and the heterogeneity of the white noise sequence in the Wold's representation of the process X_t, that

$$\hat{F}_{0,n}(x_M)/F(x_M) \to_P 1 \tag{31}$$

where

$$F(x) := \int_0^x f(u)du$$

and M is a sequence such that $M^{-1} + Mn^{-1} \to 0$ as $n \to \infty$. It may be conjectured that this result still holds when $p \geq 1$. Eq. (30) implies that

$$F(x) \simeq L(1/x)x^{1-2d}1 - 2d, x \to 0^+$$

Thus, $L(x_M^{-1})x_M^{1-2d}/(1-2d)$ can be substituted for the denominator in (31), whenever needed. Observe that the previous relation implies, for any $q > 0$,

$$\frac{F(qx)}{F(x)} \simeq q^{1-2d} L(q^{-1}x^{-1})/L(x^{-1}) \simeq q^{1-2d}, \ x \to 0^+$$

The form of the slowly-varying function L need not be specified, which is an advantage when compared to both the local Whittle and the GPH estimators. This suggests the following estimator

$$\hat{d}_{p,n} = 0.5 - \frac{\log(\hat{F}_{p,n}(qx_M)/\hat{F}_{p,n}(x_M))}{2\log(q)}$$

Note that this estimator is location and scale invariant. It is shown in Robinson (1994a) [44] (Theorem 3, pp. 532) that this estimator is consistent for $0 < d < 1/2$ (with $p = 0$). Lobato and Robinson (1996) [40] have shown that this estimator is asymptotically normal for $0 < d < 1/4$, but has non-Gaussian limit for $1/4 < d < 1/2$, with a rate convergence depending on the unknown parameter d. This is a serious drawback in practice.

4.3.2 Smoothed log-periodogram regression

Reisen [43] and Azaïs et Lang (1993,1999) [38] suggested to replace in the GPH estimator the periodogram by a "smoothed" estimate of the spectral density. Reisen proposed to use a Blackman-Tukey type estimate of the spectral density

$$\hat{f}_M(x) = \frac{1}{2\pi} \sum_{s=-M}^{M} k(s/M)\hat{\rho}(s)\cos(sx) \tag{32}$$

where $k(u)$ is the so-called lag-window generator, a fixed continuous even function in the range $-1 \leq u \leq 1$, with $k(0) = 1$ and $k(-u) = -k(u)$ and $\hat{\rho}(s)$ are the sample (biased) autocovariance coefficients, defined as

$$\hat{\rho}(s) = n^{-1} \sum_{t=1}^{n-|\tau|} (X_t - \bar{X})(X_{t+\tau} - \bar{X})$$

with $\bar{X} = n^{-1}\sum_{t=1}^{n} X_t$. The parameter M (usually referred to as the truncation point or the window bandwidth) is a function of n (the sample size), chosen such that, as $n \to \infty$, $m/n \to 0$. Reisen (1994) [43] suggested to use a Parzen lag window generator (which has the property to produce positive estimates of the spectral density).

Azaïs and Lang (1993,1999) suggested, instead of using (32), to use a Daniell type estimate of the spectral density,

$$\hat{f}_M(x) := \sum_k K((x - x_k)/M) I_{p,n}(x_k) \tag{33}$$

where $K(u)$ is the (frequency-domain) lag-window generator and M is the window bandwidth, also chosen in such a way that $M^{-1} + M/n \to 0$. Using either (32) or (33), estimators of the spectral density can be obtained at frequencies, say $0 < z_{1,n} \le z_{2,n} \le \cdots \le z_{L,n}$. The value L and the frequencies $z_{1,n}, \cdots, z_{L,n}$ must verify: $L \ge 2$ and $\lim_{n\to\infty} z_{L,n} = 0$, but can otherwise be chosen arbitrarily.

Reisen (1994) [43] suggested to set $z_{i,n} = x_i$, and L such that $L^{-1} + L/n \to \infty$ (much like for the GPH estimator). In a limited Monte-Carlo experiment, Reisen (1994) showed that the smoothed estimator compares favorably with the raw GPH estimator ($m = 1$), for a set of zero-mean ARFIMA$(1, d, 0)$ and ARFIMA$(0, d, 1)$ processes. Reisen (1994) discusses the consistency and the asymptotic normality of this estimator, but his arguments are mainly heuristics and do not appear to amount to a proof.

Azaïs and Lang (1993,1999) [2, 38] recommended to use $M = 2$ and proved the consistency and asymptotic normality of their semi-parametric estimator in the Gaussian case. It must be noted that even though this estimator might be of little practical use, it was the first for which an asymptotic theory was rigorously established. We will not discuss these estimators further.

5 Semi-parametric spectral estimation of d : global methods

The local Whittle and the GPH estimators have global counterparts. Instead of estimating d and $f^*(0)$ on a vanishing neighborhood of zero frequency, global estimators jointly estimate d and f^* over $[-\pi, \pi]$. Examples of global methods include the FAR estimator introduced by Bhansali and Kokozska (1999) [34], and the FEXP estimator, proposed by Robinson (1994) [45] and investigated independently by Moulines and Soulier (1999) [41] and Hurvich and Brodsky (1997) [26]). These estimators share the common feature, in contrast with local estimators, that they provide simultaneously a non-parametric estimator of the spectral density f.

5.1 The FAR estimator

The FAR estimator is obtained by fitting a finite order fractional autoregressive process (FAR(q)) and letting the order increase with the sample size. Any estimator of the parameters of an autoregressive process could be used in principle, but the most natural one is Whittle's estimator. We follow here the construction of Kokoszka and Bhansali who consider stationary invertible linear process and thus do not make use of tapers, and consider the integral form of Whittle's contrast. To state the results more precisely, let

$$\mathbf{F} = \{\boldsymbol{\beta} = (d, b_0, b_1, \cdots) : |d| < 1/2, \sum_{j=0}^{\infty} |b_j| < \infty\}. \tag{34}$$

Define

$$f(\boldsymbol{\beta}, x) = |1 - e^{ix}|^{-2d} \left| \sum_{j=0}^{\infty} b_j e^{ix} \right|^{-2}, \tag{35}$$

$$G(\boldsymbol{\beta}) = \int_{-\pi}^{\pi} \frac{I_n(x)}{f_{\boldsymbol{\beta}}(x)} dx. \tag{36}$$

$f_{\boldsymbol{\beta}}$ is the spectral density of a process X which admits the following representation

$$(I - B)^d X = Y, \;\; Y_t = \sum_{j=0}^{\infty} b_j Y_{t-j} + Z_t, \; t \in \mathbb{Z},$$

where Z is a white noise (Z is the innovation sequence of Y, provided that $b(z) := 1 - \sum_{i=1}^{\infty} b_i z^i \neq 0$ for $|z| \leq 1$). Denote σ^2 the innovation variance. The FAR(p) estimator of the differencing coefficient (and, as a by-product, of the spectral density) is then defined by minimizing Whittle's contrast over the finite FAR models of order p. For $\beta = (d, b_0, b_1, \cdots) \in \mathbf{F}$, denote $\beta(q) = (d, b_0, b_1, \cdots, b_q, 0, 0, \cdots)$, the truncation of β of order q, and let $\mathbf{E}$ be a compact subset of $\mathbf{F}$ such that the true parameter β^0 is an interior point of $\mathbf{E}$. Define

$$\hat{\beta}(q) = \arg\min\{G(\beta(q)); \; \beta \in \mathbf{E}\},$$
$$\hat{\sigma}^2(q) = G(\hat{\beta}(q)).$$

If q is fixed, then this estimator is not consistent except if the true model has a finite order smaller than q. Now, letting the truncation order $q = q_n$ increase to infinity, yet slowly enough so that consistent estimates of the AR coefficients can be obtained, one can expect to construct a consistent sequence of estimators of d and of the spectral density. Here, the choice of the trimming number is replaced by the selection of a truncation point (*i.e.* by model selection).

Note that the criterion $G(\beta(q))$ is non-convex. However, for a fixed value of d,

$$\bar{b}(q) := (\bar{b}_0, \cdots, \bar{b}_q) \to G(d, \bar{b}_0, \cdots, \bar{b}_q)$$

is convex and, using any standard convex optimization procedure, one can obtain

$$\hat{b}(d, q) = (\hat{b}_0(d), \cdots, \hat{b}_q(d)) = \arg \min_{\bar{b}_0, \cdots, \bar{b}_q} G(d, \bar{b}_0, \cdots, \bar{b}_q)$$

Plugging this estimate into $G(d, \bar{b}(q))$ yields the following profile likelihood

$$d \to G(d, \hat{b}(d, q))$$

which can now be minimized over a one-dimensional grid.

5.2 The FEXP estimator

Like the GPH estimator, the FEXP estimator is based on log-periodogram regression. The principle of the FEXP estimator is to estimate simultaneously d and the coefficients of a truncated expansion of $\log f^*$ on the cosine basis. These coefficients are often referred to as the *cepstrum coefficient* in the time-series literature. We use the following conventions. Define $h_0 = 1/\sqrt{2\pi}$ and $h_j(x) = \cos(jx)/\sqrt{\pi}$, $j \geq 1$ and, assuming that f^* is positive over $[-\pi, \pi]$,

$$\theta_j = \int_{-\pi}^{\pi} h_j(x) \log(f^*(x))dx. \tag{37}$$

Write now $l^* : \log(f^*) = \sum_{j=0}^{q-1} \theta_j h_j + l_q^*$, with

$$l_q^* = \log(f^*) - \sum_{j=0}^{q-1} \theta_j h_j = \sum_{j=q}^{\infty} \theta_j h_j. \tag{38}$$

The infinite series in the rhs of (38) is well defined if $\sum_{j=0}^{\infty} |\theta_j| < \infty$. The log-periodogram regression estimates of d and of $\theta_0, \cdots, \theta_q$ are given by

$$(\hat{d}_{m,p}^{\text{FEXP}}(q), \hat{\theta}_0, \cdots, \hat{\theta}_{q-1}) = \arg \min_{\bar{d}, \bar{\theta}_0, \cdots, \bar{\theta}_{q-1}} \sum_{k=1}^{K} \Big(Y_k - \bar{d} g(y_k) - \sum_{j=0}^{q-1} \bar{\theta}_j h_j(y_k) \Big)^2. \tag{39}$$

Compared with the FAR estimator, which requires the minimization of a non-convex function, the FEXP estimator is substantially simpler, because the estimators in (39) can be expressed in close form as a linear combination of the "observations" Y_k. To give the explicit expression of $\hat{d}_{m,p}^{\text{FEXP}}(q)$, some additional notations are needed. Define the following norm on $\mathbb{R}^K(m,p,n)$ (see section 3.1.2) : $\|u\|_n = 2\pi K(m,p,n)^{-1} \sum_{k=1}^{K} (m,p,n) u_k^2$, let $< .,. >_n$ denote the associated scalar product and identify any function ϕ with the vector $(\phi(y_1), \cdots, \phi(y_{K(m,p,n)}))$. 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

$$\tilde{g}_q^* := g - H_{q,n} g = \tilde{g}_q^* = g - \sum_{j=0}^{q-1} < g, h_j >_n h_j, \quad \text{and} \quad \tilde{\gamma}_q = \|\tilde{g}_q^*\|_n^2. \tag{40}$$

Denote finally $\mathbf{Y}_n = (Y_1, \cdots, Y_{K(m,p,n)})^T$. It is then easily seen that

$$\hat{d}_{m,p}^{\text{FEXP}}(q) = \frac{< \tilde{g}_q^*, \mathbf{Y}_n >_n}{\tilde{\gamma}_q} = \frac{2\pi}{K \tilde{\gamma}_q} \sum_{k=1}^{K(m,p,n)} \tilde{g}_q^*(y_k) Y_k. \tag{41}$$

A noteworthy computational adavantage of the FEXP estimator over the FAR estimator is that when d is the sole parameter of interest, it is not necessary to evaluate $\hat{\theta}_0, \cdots, \hat{\theta}_{q-1}$ to determine $\hat{d}_{m,p}^{\text{FEXP}}(q)$ (the complexity of this algorithm is thus considerably smaller than that of the FAR estimator).

6 Asymptotic theory in the case of a single pole at zero

6.1 Assumptions

In the sequel, it is assumed that $X = (X_t)_{t \in \mathbb{Z}}$ admits a linear representation with respect to a weak white noise sequence $Z = (Z_t)_{t \in \mathbb{Z}}$ such that $\mathbb{E}[Z_0] = 0$ and $\mathbb{E}[Z_0^2] = 1$, *i.e.* there exists a sequence of real numbers $(a_j)_{j\in\mathbb{Z}}$ such that $\sum_{j\in\mathbb{Z}} a_j^2 < \infty$ and

$$X_t = \mu + \sum_{j=-\infty}^{\infty} a_j Z_{t-j}. \tag{42}$$

Note that X is a Gaussian process if and only if Z is a Gaussian white noise. Eq. (42) is equivalent to the existence of the spectral density of X which then reads

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

where $a(x) := \sum_{j\in\mathbb{Z}} a_j e^{ijx}$ is defined in $L^2([-\pi,\pi], dx)$. This class of processes is still too large and must be restricted to three subclasses. Gaussian processes, linear processes in the strict sense *i.e.* with respect to an i.i.d. sequence and linear processes with respect to a fourth order stationary martingale increment sequence.

(A1) The processus X (or equivalently Z) is Gaussian.

(A2) Z is a sequence of i.i.d. random variables with finite moments up to the fourth-order and there exists a real $r \geq 1$ such that

$$\int_{-\infty}^{\infty} |\mathbb{E}[e^{itZ_0}]|^r dt < \infty. \tag{43}$$

This integrability condition is a strengthening of the so-called Cramer condition (see, e.g. Bhattacharya and Rao [6]). This condition ensures in particular that, for $n \geq r$, $d_n^Z(x_k)$ has a density w.r.t the Lebesgue measure, and thus that the periodogram $I_n^Z(x_k)$ is, for $n \geq r$ and all k, almost surely positive.

(A3) Z is a martingale increment sequence, *i.e.* $E[Z_k|\mathcal{F}_{k-1}] = 0$ where $\mathcal{F}_k = \sigma(Z_l, l \leq k)$ the filtration of the process Z. Moreover,

$$\mathbb{E}[Z_k^2|\mathcal{F}_{k-1}] = 1, \text{ a.e.} \tag{44}$$

$$\mathbb{E}[Z_k^3|\mathcal{F}_{k-1}] := \mu_3, \text{ a.e.} \tag{45}$$

$$\mathbb{E}[Z_k^4] := \kappa_4 + 3, \tag{46}$$

where κ_4 denotes the 4th-order cumulant of Z. We now state assumptions allowing to derive moment bounds for the approximation of the normalized periodogram $I_{p,n}^X(x_k)/f(x_k)$ by the periodogram of the white noise sequence $I_{p,n}^Z(x_k) := (2\pi a_p)^{-1/2} \sum_{t=1}^n h_{t,n}^p Z_t \exp(-itx_k)$ and moment bounds for normalized DFT coefficients $d_{p,n}^X(x_k)/\sqrt{f(x_k)}$. These assumptions are conveniently stated in terms of the function a^* defined by

$$a^*(x) = |1 - e^{ix}|^d a(x). \tag{47}$$

The following rather weak assumption has been originally proposed in Robinson (1995) [47] and has since been used in many contributions. It is conveniently expressed in terms of a functional class.

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

$$\frac{\max_{0 \le z \le \alpha} |a^*(z)|}{\min_{0 \le z \le \alpha} |a^*(z)|} \le \mu, \tag{48}$$

$$\frac{|a^*(x) - a^*(y)|}{\min_{0 \le z \le \alpha} |a^*(z)|} \le \mu \frac{|y - x|}{|x| \wedge |y|}. \tag{49}$$

Definition 2 *For $\mu > 1$, $\rho > 1$ and $0 < \alpha \le \pi$, let $\mathcal{L}^*(\mu, \rho, \alpha)$ be the class of square integrable functions a^* on $[-\pi, \pi]$, positive and continuously differentiable on $[-\alpha, \alpha]$ and such that (48) holds and for all $0 < |x|, |y| \le \alpha$,*

$$\frac{|a^*(x) - a^*(y) - (x - y) a^{*\prime}(x)|}{\min_{0 \le z \le \alpha} |a^*(z)|} \le \mu \frac{|y - x|^\rho}{(|x| \wedge |y|)^\rho} \tag{50}$$

Finally, an assumption of a different nature is needed to obtain bounds on the bias of the estimators of d. This assumption is naturally written in terms of the function f^* defined in (1), which can be expressed here as $f^* = |a^*|^2$, since the bounds for the bias depend on the distribution of the process only through f^*. It is convenient to define functional classes on which uniform bounds for the bias can be obtained. We consider two different classes. The first one is tailored to obtain uniform bounds for local estimators. For $\beta > 0$, $C_1 > 1$, $C_2 > 0$ and $\alpha \in (0, \pi]$, define

$$\mathcal{F}^*(\beta, C_1, C_2, \alpha) = \left\{ f^* : C_1^{-1} \le |f^*(0)| \le C_1, \ \forall x \in [0, \alpha], \ |f^*(x) - f^*(0)| \le C_2 |x|^\beta \right\}. \tag{51}$$

The bias of the FEXP estimator is naturally controlled by the rate of decay of the Fourier coefficients of $\log(f^*)$. Let w be a decreasing sequence and define the following class :

$$\mathcal{G}^*(w) = \left\{ f^* : f^*(x) = \exp\{\textstyle\sum_{j=0}^\infty \theta_j h_j(x)\}, \ \forall q \ge 0, \ \displaystyle\sum_{j=q}^\infty |\theta_j| \le w(q). \right\}. \tag{52}$$

Two examples of sequences w are of interest. For $\beta > 0$ and $L > 0$, let $w_{\beta,L}(p) = L(1+p)^\beta$. It can be easily seen that $\mathcal{G}^*(w_{\beta,L}) \subset \mathcal{F}^*(\beta, C_1, C_2, \pi))$ with $C_1 = e^{4\gamma/\beta}$ and $C_2 = 8\gamma e^{4\gamma/\beta}/\beta$. Define now $v_{\gamma,L}(p) = Le^{-\beta}$. for $\gamma > 0$ and $L > 0$. The class $\mathcal{G}^*(v_{\gamma,L})$ is of special interest since the spectral density of any causal stable and invertible ARMA processes is contained in such class. More precisely, the class $\mathcal{G}^*(w_{\gamma,L})$ contains the spectral densities of all ARMA processes with spectral density $f^*(x) = |P(e^{ix})/Q(e^{ix})|^2$ where P and Q are polynomials with no common root and no root inside the disk with radius e^γ. Note that for a suitable value of μ, the following inculsion hold.

$$\forall \beta > 1, \forall L > 0, \quad \mathcal{G}^*(w_{\beta,L}) \subset \mathcal{L}^*(\mu, \pi), \tag{53}$$

$$\forall \beta > 3/2, \forall L > 0, \quad \mathcal{G}^*(w_{\beta,L}) \subset \mathcal{L}^*(\mu, 2, \pi), \tag{54}$$

$$\forall \gamma > 0, \forall L > 0, \quad \mathcal{G}^*(v_{\gamma,L}) \subset \mathcal{L}^*(\mu, 2, \pi). \tag{55}$$

6.2 Weak convergence

In this section, we review the limiting distribution of some of the estimators presented in the previous section. We will restrain the study of the local estimators to the GSE and the GPH estimator.

6.2.1 Gaussian semi-parametric estimator

The consistency and asymptotic normality of the Gaussian semi-parametric estimator (GSE) have been established by Robinson (1995b) [46], for stationary invertible process $-1/2 < d < 1/2$ with no data taper ($p = 0$). Velasco (1999) [52] established asymptotic normality of the non-tapered version of the GSE estimator for $d \in [0.7, 0.75)$, and explained why the theory breaks down when d exceeds $3/4$.

This served to motivate either

- tapering, and more specifically the use of specific data tapers in the Kolmogorov class, following the suggestion by Velasco (1999) [52]; the problem with this approach stems from the fact that the efficiency loss incurred from using these tapers may be quite substantial.
- differencing and tapering, as suggested by Hurvich and Ray (1995) [31] and Hurvich and Chen (1999) [27]. This solution leads to a certain inflation in the variance, yet this increase is moderate. This is the approach considered in this contribution. The problem is equivalent to estimate the differencing coefficient of a potentially over-differentiated, and thus not invertible, time-series. Hurvich and Chen, 1999 shows consistency and asymptotic normality with $p = 1$ for $-3/2 < d < 1/2$ (or $-1/2 < \tilde{d} < 3/2$ prior to differencing).

Note again that since the GSE is based on linear functionals of the periodogram, pooling is irrelevant.

Theorem 1 *Assume that X is a process such that* **(A3)**, *$a^* \in \mathcal{L}^*(\mu, \alpha)$ for some $\mu > 1$ and $\alpha \in (0, \pi]$ and $-p - 1/2 < d < 1/2$. Assume that*

$$\lim_{n \to \infty} (M^{-1} + M^{2\beta/(2\beta+1)} \log^{1/\beta}(M) n^{-1}) = 0. \tag{56}$$

Then, $\sqrt{M}(\hat{d}_p^{\mathrm{GSE}}(M) - d)$ is asymptotically zero-mean Gaussian with variance $\Phi(p)/4$.

The values of $\Phi(p)$ are given in Hurvich and Chen (1999) [27].

6.2.2 GPH estimator

We now study the GPH estimator. $\hat{d}_{m,p}^{\mathrm{GPH}}(M) - d$ is naturally decomposed between a stochastic term and a bias term as follows.

$$\begin{aligned} \hat{d}_{m,p}^{\mathrm{GPH}}(M) - d &= \sum_{k=1}^{M} \beta_k(M) \epsilon_k + \sum_{k=1}^{M} \beta_k(M) \log(f^*(y_k)/f^*(0)) \\ &:= \xi_{m,p}(M) + b_M(f^*), \end{aligned} \tag{57}$$

where β_k are defined in (28) and

$$\epsilon_k := \log(\bar{I}_{m,p,n,k}) - \gamma_{m,p,n}. \tag{58}$$

$\xi_{m,p}(M)$ is a stochastic fluctuation term, whereas $b_M(f^*)$ is the bias caused by the approximation of f^* by a constant. The asymptotic theory for the GPH estimator thus follows from weak convergence of the stochastic term $\xi_{m,p}(M)$ and from a bound for the bias $b_M(f^*)$. The former requires assumptions on the distribution of the process X while the latter requires an assumption on f^*.

Theorem 2 *Let p be an integer. Assume* **(A2)**, $f^* \in \mathcal{F}^*(\beta, C_1, C_2, \alpha)$, $-p-1/2 < d < 1/2$ *and*

$$\lim_{n\to\infty}(M^{-1} + M^{2\beta+1}n^{-2\beta}) = 0. \tag{59}$$

If X is non Gaussian, assume in addition that $m \geq 4$. Then $\sqrt{M}(\hat{d}^{\mathrm{GPH}}_{m,p}(M) - d)$ is asymptotically zero-mean Gaussian with variance $\sigma^2_{m,p}/4$.

Remarks

- Theorem 2 improves on Robinson (1995) [47] Velasco (1999) [53] in the following way.
  1. The regularity condition on f^* is minimal. Robinson (1995) [47] Theorem 3 requires that $f^* \in \mathcal{L}^*(\mu, \alpha)$ for some $\alpha > 0$ to obtain bias and covariance bounds which yield the central limit theorem. However, it is possible to avoid this assumption by adapting the proof using bounds similar to Eqs. (3.5)-(3.6) in Giraitis, Robinson and Samarov (1997) [15] and Lemmas 3.1 and 3.2 in Giraitis, Robinson and Samarov (2000) [16].
  2. Theorem 3 in Robinson (1995) [47] is stated with a lower trimming number, which must go to infinity at a certain rate. This was suggested by Künsch (1986) [36], in view of the bias and the correlation of the periodogram ordinates at low-frequencies. It has now been shown that this lower trimming is not necessary. Hurvich, Deo and Brodsky (1998) [29], Theorem 2, also do not make use of the lower trimming number, but at the cost of the unnecessary assumption that f^* is twice differentiable at zero.
  3. The use of Hurvich's taper allows to consider non invertible (or equivalently, non stationary) processes, whereas Theorem 3 in Robinson (1995) [47] is stated for stationary invertible processes.
  4. We also improve on Velasco (1999) [53] since we require only a finite number of moments on the innovation sequence Z.
- Note that (59) implies that $M \to \infty$ as $n \to \infty$. It is noteworthy that asymptotic normality holds for $\hat{d}^{\mathrm{GPH}}_{m,p}(M)$ no matter how slowly M tends to infinity. The condition $M^{1+2\beta}n^{-2\beta} \to 0$ guarantees that the squared bias is negligible compared to the variance of the fluctuation term.
- The variance in the limiting distribution does not depend on d or any other unknown parameters, so Theorem 2 is simple to use to construct asymptotic confidence intervals.

6.2.3 The FAR estimator

The asymptotic theory for the FAR estimator is still incomplete. Kokoszka (1999) and Bhansali [34] have proved the consistency and conjectured the asymptotic normality of the FAR estimator when the coefficients of the AR representation of the short-memory component of the spectral density, $(b_j)_{j\in\mathbb{N}}$ decay exponentially fast, *i.e.* $|b_j| \leq C\rho^j$, with $\rho < 1$.

Theorem 3 (Kokoszka (1999) and Bhansali [34] Theorem 3.1) *Let* $\mathbf{E}$ *be a compact subset of* $\mathbf{F}$ *and assume that the true parameter* β^0 *is an interior point of* $\mathbf{E}$. *Assume also that, there exists* $C < \infty$ *and* $\rho < 1$ *such that* $|a_j| \leq C\rho^j$. *Let* q *be an non-decreasing sequence such that* $\lim_{n\to\infty}(q^{-1} + n^{-1}q) = 0$. *Then the* $\hat{d}^{\mathrm{FAR}}$ *is consistent sequence of estimator of* d.

6.2.4 The FEXP estimator

The FEXP estimtor $\hat{d}_{m,p}^{\mathrm{FEXP}}(q)$ admits a similar decomposition as the GPH estimator.

$$\hat{d}_{m,p}^{\mathrm{FEXP}}(q) = d + \xi_{m,p}(q) + b_q(f^*), \tag{60}$$

with (see (41) and (58) for notations)

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

The FEXP estimator has been studied over a much wider class of spectral densities than the FAR estimator. The minimal condition to obtain a bound for the bias of the FEXP estimator is the summability of the cepstrum coefficients $(\theta_j)_{j\in\mathbb{N}}$.

Lemma 1 (Iouditsky, Moulines and Soulier (1999), Proposition 1) *Define*

$$\theta_q^* = \sum_{j=q}^{\infty}|\theta_j|. \tag{61}$$

There exists a constant $\varsigma < \infty$ *such that for all* $q \leq K/2$,

$$|b_q(f^*)| \leq (1 + \varsigma q/K)\theta_q^*/(2\sqrt{\pi}).$$

Here, we state two theorems, because the assumptions for Gaussian and non-Gaussian processes are different. More precisely, a weaker assumption on a^* is sufficient when the process is Gaussian.

Theorem 4 *Assume that* X *is a Gaussian process with spectral density given by (1)* $f^* \in \mathcal{L}^*(\mu,\pi)$ *and* $-p - 1/2 < d < 1/2$. *If the Fourier coefficient of* $l^* = \log(f^*)$ *are absolutely summable and if* q *is a non-decreasing sequence of integers such that*

$$\lim_{n\to\infty}(q^{-1} + q\log^5(n)n^{-1}) = 0, \tag{62}$$
$$\lim_{n\to\infty}\sqrt{n/q}\sum_{k=q}^{\infty}|\theta_j| = 0, \tag{63}$$

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

Theorem 5 *Assume* **(A2)**, $a^* \in \mathcal{L}^*(\mu, \rho, \pi)$ *for some* $\rho > 1$, *and* $-p - 1/2 < d < 1/2$. *If the Fourier coefficient of* $l^* = \log(f^*)$ *are absolutely summable and if* q *is a non-decreasing sequence of integers such that (62) and (63) hold, then* $\sqrt{n/q}(\hat{d}^{\mathrm{FEXP}}_{m,p}(q) - d)$ *is asymptotically zero-mean Gaussian with variance* $m\sigma^2_{m,p}$.

6.3 Minimax Estimation

From a theoretical point of view, a usual way to assess the quality of a semi- or non-parametric estimator is through its minimax properties. In the minimax approach, the performance of an estimator are evaluated w.r.t to a *minimax lower bound.* When dealing with the estimation of the fractional differencing coefficient, such lower bounds are defined as follow. Let R be an admissible risk function, *i.e.*

$$R_X(\hat{d}_n, d) = \mathbb{E}_X\left[c(\hat{d}_n - d)\right]$$

where c is a positive bowl-shaped loss function, $\mathbb{E}_X$ is the expectation under the law of process X, and $\hat{d}_n$ is any estimator based on $X_1, \cdots, X_n$. Let $\mathcal{P}$ be a given class of processes *i.e.* $\mathcal{P}$ is a set of laws defined on a probability space. L_n is a lower bound for $(R, \mathcal{F})$ if

$$L_n \leq \inf_{\hat{d}_n} \sup_{X \in \mathcal{P}} R_X(\hat{d}_n, d).$$

where the infimum $\inf_{\hat{d}_n}$ is taken over all possible estimators $\hat{d}_n$ of d based on $\{X_1, \cdots, X_n\}$ and the supremum is evaluated over all processes in the class $\mathcal{P}$. An estimator $\hat{d}_n$ which attains this bound up to a constant,

$$\limsup_{n\to\infty} \frac{\sup_{X\in\mathcal{P}} R_X(\hat{d}_n, d)}{L_n} < \infty$$

is said to be *minimax rate-optimal.* It is called *minimax efficient* if it attains the exact bound,

$$\limsup_{n\to\infty} \frac{\sup_{X\in\mathcal{F}} R_X(\hat{d}_n, d)}{L_n} = 1.$$

For the quadratic loss function, such minimax lower bounds have been found for functional classes of spectral densities related to the local methods by Giraitis, Robinson and Samarov (1997) [15] and for functional classes related to the FEXP estimator by Iouditsky, Moulines and Soulier (1999) [32]. In both case, the tapered log-periodogram regression estimator ($\hat{d}^{\mathrm{GPH}}_{m,p}$ and $\hat{d}^{\mathrm{FEXP}}_{m,p}$) has been proved to be asymptotically minimax rate optimal for stationary possibly non-invertible processes ($-p-1/2 < d < 1/2$) Gaussian processes. Soulier (1999) [49] has proved minimax rates of convergence for the quadratic risk of the spectral density.

6.3.1 Minimax lower bounds

Theorem 6 (Giraitis Robinson and Samarov (1997), Theorem 1) *Let* $\delta, \Delta > 0$, $C_1 \geq 1$, $C_2 > 0$ *and* $\beta \geq 0$, $\alpha > 0$. *There exists a constant* $c > 0$ *such that,*

$$\liminf_n \inf_{\hat{d}_n} \sup_{-\Delta \leq d \leq \delta} \sup_{f^* \in \mathcal{F}^*(\beta, C_1, C_2, \alpha)} \mathbb{P}_{d,f^*}\left(n^{\beta/(2\beta+1)}|\hat{d}_n - d| \geq c\right) > 0, \tag{64}$$

where the infimum $\inf_{\hat{d}_n}$ *is taken over all possible estimators* d *based on* $\{X_1, \cdots, X_n\}$ *of a covariance stationary process* $\{X_t\}_{t\in\mathbb{Z}}$ *with spectral density* $f = e^{dg} f^*$.

Remarks

- Having in mind an optimality result for the GPH estimator, it is very important that the rate $n^{2\beta/(2\beta+1}$ depends neither on δ nor Δ since the GPH estimator can be shown to be rate optimal only if d is bounded from below and bounded away from $1/2$.
- (64) is equivalent to the following minimax lower bound for the quadratic of any estimator of d.

$$\liminf_n \inf_{\hat{d}_n} \sup_{-\Delta\le d\le\delta} \sup_{f\in\mathcal{F}^*(\beta,C_1,C_2,\alpha)} n^{-2\beta/(2\beta+1)} \mathbb{E}_{d,f^*}[(\hat{d}_n - d(f))^2] > 0. \tag{65}$$

We now give the minimax lower bounds for the classes $\mathcal{G}^*(w_{\beta,L})$ and $\mathcal{G}^*(v_{\gamma,L})$. Bounds for more general classes are given in Iouditsky, Moulines and Soulier (1999).

Theorem 7 (Iouditsky, Moulines and Soulier (1999) [32], Theorem 1) *Let* $\beta > 0$, $\gamma > 0$, $L > 0$, $\delta > 0$ *and* $\Delta > 0$. *There exists a positive constant* c *such that*

$$\liminf_n \inf_{\hat{d}_n} \sup_{-\Delta\le d\le\delta} \sup_{f^*\in\mathcal{G}^*(w_{\beta,L})} n^{2\beta/(2\beta+1)} \mathbb{E}_{d,f^*}[(\hat{d}_n - d)^2] \ge c, \tag{66}$$

$$\liminf_n \inf_{\hat{d}_n} \sup_{-\Delta\le d\le\delta} \sup_{f^*\in\mathcal{G}^*(v_{\gamma,L})} n \log^{-1}(n) \mathbb{E}_{d,f^*}[(\hat{d}_n - d)^2] \ge 1/2\gamma, \tag{67}$$

where the infimum $\inf_{\hat{d}_n}$ *is taken over all possible estimators of* d *based on* $\{X_1, \cdots, X_n\}$ *of a covariance stationary process* $\{X_t\}_{t\in\mathbb{Z}}$ *with spectral density* $f = e^{dg} f^*$.

Remark The inclusion $\mathcal{G}^*(w_{\beta,L}) \subset \mathcal{F}(\beta, C_1, C_2, \pi)$ (for appropriate values of C_1, C_2) shows that the bound (64) is actually implied by (66).

We now consider the estimation of the spectral density f. In a context of weak dependence, *i.e.* $d = 0$, it is pertinent to assess the performance of an estimator $\hat{f}_n$ by its L^2 risk $\mathbb{E}_f[\|f - \hat{f}_n\|^2_{L^2}]$. Minimax bounds in that framework have been given by Efroimovich and Pinsker (1982) [12]. In the context of fractional processes, the spectral density is not necessarily square integrable, but its logarithm is, so a natural choice for the risk of an estimator $\hat{l}_n$ of $l = \log(f)$ is $\|l - \hat{l}_n\|^2_n$, where the norm $\|.\|_n$ is defined in section 5.2. We must introduce another functional class. Define

$$\mathcal{S}^*(\beta, L) = \{f^* : f^* \exp\{\sum_{j=0}^{\infty} \theta_j h_j\},\ \sum_{j=0}^{\infty} (1+j)^{2\beta} \theta_j^2 \le L^2\}.$$

Theorem 8 (Soulier (1999) [49], Theorem 1) *Let* $\beta > 0$, $\gamma > 0$, $L > 0$, $\delta > 0$ *and* $\Delta > 0$. *There exists a positive constant* c *such that*

$$\liminf_n \inf_{\hat{l}_n} \sup_{-\Delta\le d\le\delta} \sup_{f^*\in\mathcal{S}^*(\beta,L)} n^{2\beta/(2\beta+1)} \mathbb{E}_{d,f^*}[\|\hat{l}_n - l\|^2_n] \ge c, \tag{68}$$

$$\liminf_n \inf_{\hat{d}_n} \sup_{-\Delta\le d\le\delta} \sup_{f^*\in\mathcal{G}^*(v_{\gamma,L})} n \log^{-1}(n) \mathbb{E}_{d,f^*}[\|\hat{l}_n - l\|^2_n] \ge 2\pi/\gamma, \tag{69}$$

where the infimum $\inf_{\hat{l}_n}$ *is taken over all possible estimators of* $l = \log(f)$ *based on* $\{X_1, \cdots, X_n\}$ *of a covariance stationary process* $\{X_t\}_{t\in\mathbb{Z}}$ *with spectral density* $f = e^{dg} f^*$.

6.3.2 Upper bounds

The GPH estimator and the FEXP estimator can both be expressed in the following way :

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

where q is the upper trimming number (denoted M above) in the case of the GPH estimator, and the truncation number in the case of the FEXP estimator. When f^* belongs to one of the functional classes introduced above, it has been shown that the leading term in the variance of the stochastic term is the same as in the case of a white noise, *i.e.*

$$\sigma_q^2 = \sigma_{m,p}^2 \sum_{k=1}^{K} \beta_{n,k}^2(q),$$

(cf. Giraitis, Robinson and Samarov (1997,2000) [16, 16], Moulines and Soulier (1999) [41] and Iouditsky, Moulines and Soulier (1999) [32]). It is easily seen that $q\sigma_q^2$ is uniformly (with respect to n) bounded and that for each n, σ_q^2 is strictly decreasing in the case of the GPH estimator and strictly increasing in the case of the FEXP estimator. In the minimax framework discussed above, it is assumed that f^* belongs to some functional class, say $\mathcal{F}^*$. If a uniform bound, say v_q, is available for the bias term $b_q(f^*)$, *i.e.* for all $f^* \in \mathcal{F}^*$, $|b_q(f^*)| \leq v_q$ and if moreover v_q is either non decreasing or non increasing with respect to q (for a given n, the dependence in n being implicit), then theoretical best choice for q is thus the one that balances the variance σ_q^2 and the squared bias v_q^2 : the sequence q_n must be chosen such that $c^{-1} \leq \liminf_n \sigma_{q_n} v_{q_n} \leq \limsup_n \sigma_{q_n}^{-1} v_{q_n} \leq c$ for some real $c > 1$. Such a choice yields the following uniform bound for the mean square error of $\hat{d}(q_n)$.

$$\limsup_n \sum_{-\Delta\leq d\leq\delta} \sup_{f^*\in\mathcal{F}^*} \sigma_{q_n}^{-1} \mathbb{E}_{d,f^*}[(\hat{d}(q_n) - d)^2] \leq C.$$

We now illustrate this method for the GPH and the FEXP estimators.

Rate optimality of the GPH estimator First we state a theorem that generalizes Theorem 2 of Giraitis, Robinson and Samarov (1997) [15] to the non invertible case.

Theorem 9 *Let* $\beta > 0$ *and* $p \geq 0$, $\alpha \in (0, \pi]$. *Let* $\Delta \in (0, p + 1/2)$ *and* $\delta \in (0, 1/2)$. *Consider the GPH estimator* $\hat{d}_{m,p}^{\mathrm{GPH}}(q_n)$ *with* $M_n = [n^{2\beta/(2\beta+1)}]$.

$$\limsup_n \sup_{-\Delta\leq d\leq\delta} \sup_{f^*\in\mathcal{F}^*(\beta,C_1,C_2,\alpha)} n^{2\beta/(2\beta+1)} \mathbb{E}_{d,f^*}[(\hat{d}_{m,p}^{\mathrm{GPH}}(M_n) - d(f))^2] \leq \bar{C}(\beta, C_1, C_2, \delta, \Delta).$$

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

Rate optimality of the FEXP estimator

Theorem 10 (Iouditsky, Moulines and Soulier (1999) [32]) *Let $\beta > 1$, $\gamma > 0$, $L > 0$, $\Delta > 0$ and $0 < \delta < 1/2$. Define $q_n(\beta, L) = [(Ln)^{1/(1+2\beta)}]$ and $q_n(\gamma) = [\log(n)/2\gamma]$.*

$$\limsup_n \sup_{-\Delta<d<\delta} \sup_{f^* \in \mathcal{G}^*(w_{\beta,L})} n^{2\beta/(2\beta+1)} \mathbb{E}_{d,f^*}[(\hat{d}^{\mathrm{FEXP}}_{m,p}(q_n(\beta,L)) - d)^2] \le \bar{C}(\beta,\delta,\Delta) L^{1/(2\beta+1)}, \tag{70}$$

$$\lim_{n\to\infty} \sup_{-\Delta<d<\delta} \sup_{f^* \in \mathcal{G}^*(\nu_{\gamma,L})} \log^{-1}(n) \mathbb{E}_f[(\hat{d}^{\mathrm{FEXP}}_{m,p}(q_n(\gamma)) - d)^2] = m\sigma^2_{m,p}/2\gamma, \tag{71}$$

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

Remarks

- Note that when $p = 0$, $\sigma^2_{m,0} = \psi'(m)$. Since $m\psi'(m)$ tends to 1 as m tends to infinity, (71) implies that the lim inf in (67) is actually a limit and is equal to $1/2\gamma$.
- (70) can be extended to the case $\beta < 1$ if the supremum is restricted to functions $f^* \in \mathcal{L}^*(\mu, \pi)$. The constant on the right-hand side of (70) in that case also depends on μ.
- Here again, it is possible to deal with over-differenced time-series, by using tapers. The resulting estimators are still rate optimal but the loss of efficiency $m\sigma^2_{m,p}$ in the analytic case does not tend to 1 as m tends to infinity for fixed $p \ge 1$.

The FEXP estimator provides also an estimator of the spectral density. More precisely, define

$$\hat{l}^{\mathrm{FEXP}}_{m,p}(q) = \hat{d}^{\mathrm{FEXP}}_{m,p}(q) + \sum_{j=0}^{q-1} \hat{\theta}_j h_j. \tag{72}$$

The next theorem states its rate optimality (cf. Moulines and Soulier (2000) [42] and Soulier (1999) [49]).

Theorem 11 *Let $\beta > 1$, $\gamma > 0$, $L > 0$, $\Delta > 0$ and $0 < \delta < 1/2$. Define $q_n(\beta, L) = [(Ln)^{1/(1+2\beta)}]$ and $q_n(\gamma) = [\log(n)/2\gamma]$.*

$$\limsup_n \sup_{-\Delta<d<\delta} \sup_{f^* \in \mathcal{S}^*(\beta,L)} n^{2\beta/(2\beta+1)} \mathbb{E}_{d,f^*}[\|l - \hat{l}^{\mathrm{FEXP}}_{m,p}(q)\|_n^2] \le \bar{C}(\beta,\delta,\Delta) L^{1/(2\beta+1)}, \tag{73}$$

$$\lim_{n\to\infty} \sup_{-\Delta<d<\delta} \sup_{f^* \in \mathcal{G}^*(\nu_{\gamma,L})} \log^{-1}(n) \mathbb{E}_{d,f^*}[\|l - \hat{l}^{\mathrm{FEXP}}_{m,p}(q)\|_n^2] = 2\pi m\sigma^2_{m,p}/\gamma, \tag{74}$$

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

Remarks

- As for the estimation of d, a consequence of theorems 8 and 11 is that in the case $p = 0$, the liminf in (68) is actually a limit, and the FEXP estimator is quasi efficient.

- It is remarkable that in the analytic class $\mathcal{G}^*(\nu_{\gamma,L})$, the minimax rate of convergence is the same for the estimation of d and of $\log(f)$, whereas for $\beta > 1/2$, the inclusion $\mathcal{S}^*(\beta, L) \subset \mathcal{G}^*(w_{\beta-1/2,L})$ shows that the minimax rate of convergence for the estimation of d is slower than the minimax rate of convergence for the estimation of $\log(f)$. This situtation is similar to the problem of pointwise or global estimation of a density.

7 Model selection for the estimation of d

The problem of the minimax approach of the previous section is that it provides unfeasible estimators. Thus automatic selection procedures are needed for the choice of the trimming number in local methods or for the truncation number in the FEXP estimator or the FAR estimator. Based on simulations, Geweke and Porter-Hudak (1983) [14] suggested that $M = K^{1/2}$ be used as an upper trimming number for the GPH estimator, and this choice has been since widely adopted in the literature. As can be seen from theorem 9, this choice is not asymptotically optimal, and even practitioners may prefer a theoretically founded choice. Traditional approaches have been proposed for the selection of the upper trimming number of local estimators : plug-in methods (Henry and Robinson (1996) [21], Hurvich and Deo (1999) [28]), frequency-domain cross-validation (Hurvich and Beltrao (1994) [25]). For the estimation of $\log(f)$, Moulines and Soulier (2000) [42] have proved a certain type of asymptotic optimality of a selection procedure based on Mallows' C_p statistics for the FEXP estimator. These traditional approaches, even when not purely heuristically justified lack a rigorous measure of accuracy for the estimators they are used to built. For instance, the plug-in procedure of Hurvich and Deo (1999) provides a choice of an "optimal" trimming number $\hat{M}_n$ such that if M_n^{opt} asymptotically minimizes the mean square error of the GPH estimator, then $\lim_{n\to\infty}(M_n^{opt}/\hat{M}_n) = 1$, but this result could be said irrelevant since the relevant objective should be for instance to bound $\mathbb{E}[(\hat{d}_{m,p}^{\mathrm{GPH}}(\hat{M}) - d)^2]$. In sharp contrast, the relatively recent theory of adaptive estimation, initiated among others by Lepski (1990) proposes a very simple method for automatically selecting from the data the parameters of a semi- or non-parametric estimator (such as the trimming number M in the GPH estimator or the truncation order q in the FEXP estimator). In the context of the estimation d, this method yields

- a bound for the mean square error of $\hat{d}(\hat{q})$ for a given spectral density that is equal to the optimal mean square error for this spectral density, possibly up to a logarithmic factor; this property is referred to as adaptivity to the target function;
- a uniform bound the mean square error of $\hat{d}(\hat{q})$ over classes of spectral densities which is usually the minimax rate for these classes, up to a logarithmic factor; this logarithmic loss can usually be proved unavoidable in absence of prior knowledge on the functional class. This property is referred to as adaptivity in the minimax sense.

We will mainly focus on the description of this method in the next section and briefly present the traditional plug-in methods in sections 7.2 and 7.3 below.

7.1 Adaptive estimation of the fractional differencing coefficient

Consider an estimator of d such that $\hat{d}(q)$ can be decomposed in a stochastic fluctuation term $\xi(q)$ and a bias term $b_q(f^*)$. This situation is generic, but is more specifically adapted to the case of the GPH and FEXP estimators. Assume that $\mathbb{E}[\xi^2(q)] = \sigma_q^2(1+o(1))$ where the term $o(1)$ is uniform with respect to q (*i.e.* tends to zero uniformly wrt q as the sample size n increases), and σ_q^2 is either non increasing or non decreasing. In order to handle the GPH and the FEXP estimators at the same time, we will consider that in both case, the variance σ_q^2 is decreasing. This amounts to a reparametrization of the FEXP estimator. Let κ be a positive real number (the value of which will be given later). A truncation index q is said *admissible* if

$$\forall q' < q, |\hat{d}(q') - \hat{d}(q)| \leq \kappa\sqrt{\log(n)}\sigma_{q'}.$$

$\hat{q}$ is then defined as the largest admissible integer, and the adaptive estimator is $\hat{d}(\hat{q})$.

In order to assess the performance of the adaptive estimator, we introduce a deterministic sequence of integers $q_n^*(f^*)$ (depending upon the spectral density under consideration) that satisfies the following properties.

$$\forall q \leq q_n^*(f^*), \quad |b_{n,q}(f^*)| \leq \sqrt{\log(n)}\sigma_q. \tag{75}$$

(75) means that it is not required that the sequence $q_n^*(f^*)$ balances the squared bias and variance. Rather, the bias term is allowed to exceed the variance by a logarithmic factor. The aim of this method is to prove that under certain technical conditions, $\hat{q}$ is greater than $q_n^*(f^*)$ with a high probability. We now state precisely these technical conditions and the theorem they allow to prove on the mean square error of the adaptive estimator.

(T1) Let $\delta \in (0, 1/2)$ and let $\mathcal{F}^*$ be a functional class. For all $p \geq 1$, there exists a constant $C(\delta, \mathcal{F}^*, p)$ such that for all $|d| \leq \delta$, all $f^* \in \mathcal{F}^*$ and all $q \leq q_n^*(f^*)$,

$$\mathbb{E}_{d,f^*}[\xi_{n,q}^{2p}] \leq C^p(\delta, \mathcal{F}^*, p)\sigma_q^p.$$

(T2) $\mathcal{F}^*$ is a functional class such that there exists a constant $C(\mathcal{F}^*)$ such that

$$\forall f^* \in \mathcal{F}^*, \quad \forall q \leq q_n^*(f^*), \quad |b_{n,q}(f^*)| \leq C(\mathcal{F}^*).$$

(T3) Let $\delta \in (0, 1/2)$, $\rho > 0$ and let $\mathcal{F}^*$ be a functional class. There exists a constant $C(\delta, \mathcal{F}^*)$ and an integer $N(\delta, \mathcal{F}^*, \rho)$ such that for all $n \leq N(\delta, \mathcal{F}^*, \rho)$, $q \leq q_n^*(f^*)$, and for all $|d| \leq \delta$ and $f \in \mathcal{F}^*$),

$$\mathbb{P}_{d,f^*}\left(|\xi_{n,q}| > (\kappa/2 - 1)\sqrt{\log(n)}\sigma_q\right) \leq C(\delta, \mathcal{F}^*)n^{-(\kappa/2-1)^2/2(1+\rho)}.$$

The functional class is introduced to allow for both kinds of adaptivity described above. If adaptivity to the target function is seeked, 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 12 *Let $\delta \in (0,1/2)$. 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 (75),* **(T1)**, **(T2)** *and* **(T3)** *hold. Then if $\kappa > 6$, there exists a constant $C(\delta, \mathcal{F}^*, \kappa)$ such that for large enough n (depending on δ, $\mathcal{F}^*$ and κ), and for all $f \in \mathcal{F}^*$,*

$$\mathbb{E}_{d,f^*}[(\hat{d}(\hat{q}) - d)^2] \leq (1+\kappa)^2 \sigma^2_{q_n^*(f^*)} \log(n) + C(\delta, \mathcal{F}^*, \kappa) n^{-1}.$$

Remarks

- It must be stressed that the sequence $q_n^*(f^*)$ is only used to assess the performance of the adaptive estimator $\hat{d}(\hat{q})$. The dependence in f^* is also stressed, since it means that $\hat{d}(\hat{q})$ is *adaptive to the target function*. The bound for the mean square error depends only on f^*, but not on any prior assumption on this function, except the minimal ones that allow the derivation of **(T1)**-**(T3)**.
- Since σ_q is non increasing, then $q_n^*(f^*)$ should be chosen as the greatest integer such that **(T2)** and **(T3)** hold. Whenever possible, $q_n^*(f^*)$ must be chosen such that $\sigma^2_{q_n^*(f^*)}$ balances $b^2_{n,q_n^*(f^*)}(f^*)\log(n)$ and the adaptive estimator then balances the squared bias and variance for a given spectral density, up to a logarithmic factor. This is usually the best that can be achieved by an adaptive estimator over a functional class, as is illustrated in the next subsections.
- The exponential inequality **(T3)** is obviously the hardest of the three tools to obtain. In the context of log-periodogram regression, such an inequality has been proved only for Gaussian processes and using a taper of order at least one, under both intermediate ($d \leq 0$) and long memory ($d > 0$) conditions. Thus the range of application of Theorem 12 is implicitly restricted to Gaussian processes. The technique to obtain such an inequality was first presented in Giraitis, Robinson and Samarov (2000) and adapted to the form **(T3)**, by Iouditsky, Moulines and Soulier (1999).
- The value of κ is optimal given the technique of proof. It could not be improved even if the regression noise was actually i.i.d. zero-mean Gaussian with known variance. The difference would be that κ could be chosen exactly equal to 6 instead of strictly greater and that the bound would not depend on the class $\mathcal{F}^*$ and would be valid for all n. If an upper bound is assumed for $q_n^*(f^*)$, as is the case in Giraitis *et al.* then κ can be chosen smaller. This is not advisable however, if one wants the estimator to be really adaptive, that is with the least possible prior knowledge about f^*.

This result is illustrated in the next sections for the GPH and the FEXP estimator.

7.1.1 Adaptive GPH estimator

If f^* belongs to a class $\mathcal{F}^*(\beta, C_1, C_2, \alpha)$ for some unknown $\beta > 0$, Lemma 3.1 and 3.2 in Giraitis, Robinson and Samarov (2000) [16] can be adapted to prove that (75), **(T1)**, **(T2)** and **(T3)** hold with $M_n^* = M_n^*(\beta) = [K(\beta, C_1, C_2) n^{2\beta/(2\beta+1)}]$, for some constant $K(\beta, C_1, C_2)$. This yields the following corollary.

Corollary 1 *Let p be a positive integer, $\Delta \in [0, p+1/2)$, $\delta \in (0,1/2)$, $\beta > 0$, $C_1 > 1$, $C_2 > 0$ and $\alpha \in (0,\pi]$. There exists a constant $\bar{C} := \bar{C}(\beta_*, \beta^*, \delta, \Delta, \alpha)$ suxh that*

$$\limsup_n \sup_{\beta_* \leq \beta \leq \beta^*} \frac{n^{2\beta/(2\beta+1)}}{\log(n)} \sup_{-\Delta \leq d \leq \delta} \sup_{f^* \in \mathcal{F}^*(\beta, C_1, C_2, \alpha)} \mathbb{E}_{d,f^*}[(\hat{d}^{\mathrm{GPH}}_{m,p}(\hat{M}) - d)^2] \leq \bar{C},$$

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

In Giraitis, Robinson and Samarov (2000) [16], another adaptive version of the GPH estimator has been proposed, specifically adapted to this family of functional classes. Instead of comparing the estimators based on arbitrary values of m, without any prior assumption on the type of functional class to which f^* belongs as done here, it is assumed that f^* does belong to some $\mathcal{F}^*(\beta, C_1, C_2, \alpha)$. Estimators based on values $M(\gamma) = n^{2\gamma/(2\gamma+1)}$ where γ belongs to a grid of width $1/\log(n)$ are compared to obtain a value $\hat{\beta}$ which can be interpreted as an estimator of the true smoothness. This adaptive estimator achieves the rate $(\log^2(n)/n)^{2\beta/(2\beta+1)}$ if f^* actually belongs to the class $\mathcal{F}^*(\beta, C_1, C_2, \alpha)$. Thus the estimator of Giraitis, Robinson and Samarov (2000) [16] is better than the one presented here when $\beta < 1/2$, but worse when $\beta > 1/2$. The advantage of one construction over the other for this specific family of functional classes is thus not decisive. What could be considered as a decisive advantage of the construction presented here is that it is intrinsic and is independent of any prior knowledge about the smoothness of f^*, and even if it is assumed that f^* belongs to some class $\mathcal{F}^*(\beta, C_1, C_2, \alpha)$, then the estimator does not depend on prior bounds β_* and β^* for β, whereas the construction in Giraitis, Robinson and Samarov (2000) [16] explicitly depends on β_* and β^*. For instance, if the true f^* belongs to $\mathcal{F}^*(\beta, C_1, C_2, \alpha)$ for some $\beta > \beta^*$, the estimator presented here will achieve the rate $\log(n)/n^{2\beta/(2\beta+1)}$, whereas the rate of convergence of the GRS estimator would be limited at $(\log^2(n)/n)^{2\beta^*/(2\beta^*+1)}$. However, both estimators fail to achieve the following lower bound, proved in Giraitis, Robinson and Samarov (2000) [16].

Theorem 13 (Giraitis, Robinson and Samarov (2000), Theorem 2.1) *For any $\alpha > 0$, $\beta^* > \beta_* > 0$, $C_1 > 1$, $C_2 > 0$, $\Delta > 0$, and $\delta \le 1/2$, there exists a positive constant $\underline{C} := \underline{C}(\beta_*, \beta^*, C_1, C_2, \delta, \Delta)$ such that*

$$\liminf_n \inf_{\check{d}_n} \sup_d \sup_{\beta_* \le \beta \le \beta^*} (n/\log(n))^{2\beta/(2\beta+1)} \sup_{f^* \in \mathcal{F}^*(\beta, C_1, C_2, \alpha)} \mathbb{E}_{d,f^*}[(\check{d}_n - d)^2] \ge \underline{C},$$

where the infimum $\inf_{\check{d}_n}$ is taken over all possible estimators $\check{d}_n$ of d based on n observations $\{X_1, \cdots, X_n\}$ of a covariance stationary process $(X_t)_{t\in\mathbb{Z}}$ with spectral density $f = e^{dg} f^*$ with $-\Delta \le d \le \delta$ and $f^* \in \mathcal{F}^*(\beta, C_1, C_2, \epsilon)$.

Thus the conjecture remains open to prove that this bound is actually the adaptive minimax rate of convergence. To prove that this conjecture holds true, one should exhibit an adaptive estimator that achieves the rate $(\log(n)/n)^{2\beta/(2\beta+1)}$ when f^* actually belongs to the class $\mathcal{F}^*(\beta, C_1, C_2, \alpha)$.

It seems plausible that the rate $(\log(n)/n)^{2\beta/(2\beta+1)}$ is the exact adaptive rate of convergence, but Lepski's method, as it is implemented here, cannot yield an adaptive estimator that attains this rate for the following heuristic reason. As mentioned earlier, the choice of $\hat{M}$ by Lepski's method does not result in the balance of the variance and squared bias, since the bias is allowed to exceed the variance by a logarithmic factor. The problem here is that the covariance bounds for the noise sequence $\epsilon_{n,k}$ obtained uniformly over the class $\mathcal{F}^*(\beta, C_1, C_2, \alpha)$ (which seems difficult to improve) include a term which is of the same order of magnitude as the bias term. This means that any increase in the bias results in the same increase in the variance, and thus the goal of Lepski's method cannot be achieved for this class. From a technical point of view, the rate $(\log(n)/n)^{2\beta/(2\beta+1)}$ could be achieved if the exponential inequality **(T3)** did hold with $m_n^* \approx n^{2\beta/(2\beta+1)}(\log(n))^{1/(2\beta+1)}$. For the same technical reasons mentioned above, that can be clearly understood by looking at the proof of Lemma 3.2 in Giraitis, Robinson and Samarov (2000) [16], this cannot be achieved uniformly over the class $\mathcal{F}^*(\beta, C_1, C_2, \epsilon)$.

We now introduce a restricted functional class on which the rate $(\log(n)/n)^{2\beta/(2\beta+1)}$ can be achieved. For $\beta > 0$, $C_1 > 1$ and $C_2 > 0$, define

$$\mathcal{H}^*(\beta, C_1, C_2, \alpha) = \{f^* : C_1^{-1} \le f^*(0) \le C_1, \forall x, y \in [-\alpha, \alpha], \ |\frac{f^*(x) - f^*(y)}{f^*(0)}| \le C_2 |x - y|^\beta\}.$$

This Lipschitz condition of order β allows to improve on Lemma 3.1 of Giraitis, Robinson and Samarov (2000) [16] so that (75), **(T1)**, **(T2)** and **(T3)**, hold for $\mathcal{H}^*$ and $m_n^* = n^{2\beta/(2\beta+1)}(\log(n))^{1/(2\beta+1)}$. Thus we have the following upper bound.

Corollary 2 *Let p be a positive integer, $\Delta \in [0, p + 1/2)$, $\delta \in (0, 1/2)$, $\beta > 0$, $C_1 > 1$, $C_2 > 0$ and $\alpha \in (0, \pi]$. There exists a constant $\bar{C} := \bar{C}(\beta_*, \beta^*, \delta, \Delta, \alpha)$ suxh that*

$$\limsup_n \sup_{\beta_* \le \beta \le \beta^*} (n/\log(n))^{2\beta/(2\beta+1)} \sup_{-\Delta \le d \le \delta} \sup_{f^* \in \mathcal{H}^*(\beta, C_1, C_2, \alpha)} \mathbb{E}_{d,f^*}[(\hat{d}_{m,p}^{\mathrm{GPH}}(\hat{M}) - d)^2] \le \bar{C},$$

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

Even though the strict inclusion $\mathcal{H}^*(\beta, C_1, C_2, \alpha) \subset \mathcal{F}^*(\beta, C_1, C_2, \alpha)$ holds, the lower bound of Theorem 2.1 of Giraitis, Robinson and Samarov (2000) [16] is still valid.

Theorem 14 *For all $\beta^* > \beta_* > 0$, $\delta \in (0, 1/2)$, $\Delta > 0$, $C_1 > 1$ and $C_2 > 0$, there exists a constant $\underline{C}$ such that*

$$\liminf_n \sup_{\beta_* \le \beta \le \beta^*} \inf_{\check{d}} (n/\log(n))^{2\beta/(2\beta+1)} \sup_{-\Delta \le d \le \delta} \sup_{f^* \in \mathcal{H}^*(\beta, C_1, C_2, \alpha)} \mathbb{E}_{d,f^*}[(\check{d}_n - d)^2] \ge \underline{C}.$$

Thus the adaptive GPH estimator is adaptive rate optimal over the classes $\mathcal{H}^*(\beta, C_1, C_2, \alpha)$.

7.1.2 Adaptive FEXP estimator

For clarity, we first redefine the selection procedure using the appropriate notations. Let $\epsilon_K = \log^4(K)$. An integer $q < \epsilon_K K$ is admissible if

$$\text{for all } q < r \le \epsilon_K K, |\hat{d}_{m,p}^{\mathrm{FEXP}}(r) - \hat{d}_{m,p}^{\mathrm{FEXP}}(q)| \le \kappa \sqrt{\log(K)} \sigma_r,$$

where here $\sigma_q^2 = 4\pi m \sigma_{m,q}^2 / n\tilde{\gamma}_q$. Let $\hat{q}$ be the smallest admissible integer. The adaptive estimator is thus $\hat{d}_{m,p}^{\mathrm{FEXP}}(\hat{q})$. It has been proved in Iouditsky, Moulines and Soulier that (75), **(T1)**, **(T2)** and **(T3)** hold for $f^* \in \mathcal{L}^*(\mu, \pi)$ and $q_n^*(f^*)$ defined as follows :

$$q_n^*(f) = \max\{q : 1 \le q \le K, \ \sigma_q \sqrt{\log(K)} \le \theta_q^*\} + 1, \tag{76}$$

where θ_q^* is defined in (61). In this context, Theorem 12 yields

Corollary 3 *Let p be a positive integer, let $\delta \in (0,1/2)$, $\Delta \in [0,p+1/2)$ and $0 < \beta_* < \beta^* < \infty$. There exists a constant $\bar{C} := \bar{C}(\beta_*, \beta^*, L^*, \delta, \Delta)$ such that*

$$\limsup_n \sup_{-\Delta \le d \le \delta} \sup_{\beta_* \le \beta \le \beta^*} \sup_{0<L<L^*} \sup_{f^* \in \mathcal{G}(w_{\beta,L})} (n/\log(n))^{2\beta/(2\beta+1)} \mathbb{E}_{d,f^*}[(\hat{d}_{m,p}^{\mathrm{FEXP}}(\hat{q}) - d)^2] \le \bar{C},$$

$$\limsup_n \sup_{-\Delta \le d \le \delta} \sup_{\beta_* \le \beta \le \beta^*} \sup_{0<L<L^*} \sup_{f \in \mathcal{G}^*(v_{\beta,L})} \frac{n}{\log^2(n)} \mathbb{E}_{d,f^*}[(\hat{d}_{m,p}^{\mathrm{FEXP}}(\hat{q}) - d)^2] \le \bar{C},$$

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

[42])

7.2 Plug-in method for the GPH estimator

We briefly present in this section the rsults of Hurvich, Deo and Brodsky (1998) [29] and Deo and Hurvich (1999) [28]. The pooled, non tapered periodogram is considered, so the results only apply to the range $-1/2 < d < 1/2$. The plug-in method is based on an expansion of the mean square error of the GPH estimator under the assumption that f^* is three times differentiable in a neighborhood of 0. Under this assumption, if $\lim_{n\to\infty}(M^{-1} + M\log(M)n^{-1}) = 0$, then

$$\mathbb{E}[(\hat{d}_{m,0}^{\mathrm{GPH}}(M) - d)^2] = \frac{4\pi^4}{81} \left\{ \frac{f^{*\prime\prime}(0)}{f^*(0)} \right\}^2 \frac{M^4}{K^4} + \frac{m\psi'(m)}{4M} + O\Big(\frac{M\log^3(M)}{n}\Big) + O\Big(\frac{M^4}{n^4}\Big) + o\Big(\frac{1}{m}\Big). \tag{77}$$

Neglecting the remainder term in the mean squared error, assuming that $f^{*\prime\prime}(0) \neq 0$, and minimizing with respect to M yields the asymptotically optimal choice for M,

$$M^{\mathrm{opt}} = CK^{4/5}, \; C := \Big(\frac{81 m\psi'(m)}{64\pi^4}\Big)^{1/5} \left\{ \frac{f^{*\prime\prime}(0)}{f^*(0)} \right\}^{2/5} \tag{78}$$

Since M^{opt} depends on the behavior of the unknown function f^* at zero, the plug-in method consists in estimating $\hat{C}$ of C. Since f, hence f^* is even, a Taylor expansion of $\log(f^*)$ at zero yields

$$\log f^*(y_k) = \log f^*(0) + (\Delta/2)y_k^2 + y_k^3 R(y_k) \tag{79}$$

where $\Delta := -f^{*\prime\prime}(0)/f^*(0)$ and $R(y_k)$ is uniformly bounded in the neighborhood of the origin. This suggests to estimate Δ as the third coefficient in an ordinary linear regression of $Y_{n,k}$ on a regression matrix with columns $(1, g(y_k), y_k^2/2)$, $1 \le k \le L\}$, where $L = \min(AK^\delta, K)$, for some arbitrary constant A and $0 < \delta < 1$. From lemma 1 in Hurvich and Deo [28], it can be infered that δ should be set equal to $6/7$, yielding a consistent estimator $\hat{\Delta}$ of Δ. Using this value, a consistent estimator of C is $\hat{C} = \big(\frac{81 m\psi'(m)}{64\pi^4}\big)^{1/5}\hat{\Delta}$. This estimator of $\hat{C}$ can then be used to construct a regression estimator $\hat{d}_{\hat{M}}^{GPH}$, with $\hat{M} = \hat{C}K^{4/5}$, but there is no theoretical result (such as an evaluation of the mean square error) to justify this choice.

7.3 Plug-in estimator for Gaussian semi-parametric estimator

In the case of the Gaussian semiparametric estimator (and for similar reasons in the case of the FAR estimator), there are as yet no satisfactory method of choosing the trimming number M, the main reason being that the GSE is implicitly defined. The only selection method presented in the litterature is the plug-in method of Henry and Robinson (1996) [21] which mimicks the construction of the plug-in bandwidth estimate for the GPH. We briefly resume here the heuristics of this approach. It is suggested to approximate the bias of the GSE by $\frac{16\pi^2}{9}\frac{f^{*\prime\prime}(0)}{f^*(0)}\frac{M^2}{n^2}$ and the variance by $4/M$. Balancin the approximate square bias and variance yields the following tentative value of the optimal bandwidth,

$$M^{\rm opt} = Cn^{4/5}, \ C = (\frac{3}{4\pi})^{4/5}(\frac{f^*(0)}{f^{*\prime\prime}(0)})^{2/5}$$

This optimal bandwidth is similar to the one derived for the GPH estimator, and a consistent estimator of C can be obtained along the same lines as in section 7.2.

8 Poles with unknown location

In this section, we briefly describe the problem and the existing partial answers of semi-parametric estimation of the fractional differencing parameter and of the G-frequency when the G-frequency is unknown. More precisely, let X be a stationary process with spectral density f that writes

$$f(x) = |1 - e^{i(x-\omega_0)}|^{-d(\omega_0)}|1 - e^{i(x+\omega_0)}|^{-d(\omega_0)}f^*(x), \tag{80}$$

where $0 < d(\omega_0) < 1$ if $\omega_0 \in (0,\pi)$ and $0 < d(\omega_0) < 1/2$ if $\omega_0 \in \{0,\pi\}$. and f^* is an even positive continuous 2π-periodic function on $[-\pi,\pi]$. The notation $d(\omega_0)$ stresses the fact that the G-frequency ω_0 is now considered as unknown. In this context, the first problem is to find a consistent estimator of the G-frequency ω_0 and preferably, to have one with the best possible rate of convergence. The second, and perhaps less important in practice, problem is to derive the asymptotic distribution for such an estimator.

The question of the best attainable rate of convergence (in a minimax sense) for such an estimator remains open. If the memory parameter is not bounded away from zero, then it is easily seen that it is impossible to consistently estimate ω_0, since when $d(\omega_0)$ tends to zero, the very definition ω_0 becomes meaningless. If $d(\omega_0)$ is positive and bounded away from zero it has been conjectured that the best possible rate might be n. In a semi-parametric context, Yajima (1995) and Hidalgo (1999) have proposed estimators that attain the rate n^α for any $\alpha < 1$, for a given $d(\omega_0) > 0$. Yajima (1995) failed to derive an asymptotic distribution for his estimator, while Hidalgo (1999) proved asymptotic normality of his estimator except at zero and π where it is respectively asymptotically distributed as the positive and negative part of a zero-mean Gaussian variable. In any case, the asymptotic variance of Hidalgo's estimator is proportional to $d(\omega_0)^{-2}$ and thus blows up as $d(\omega_0)$ goes to zero, in accordance with the remarks above.

Most of the estimators $\hat{d}(0)$ of the differencing parameter introduced in the previous sections (for $\omega_0 = 0$) can be more or less easily adapted to yield an estimator $\hat{d}(\omega_0)$ of $d(\omega_0)$ if $\omega_0 \neq 0$ is assumed to be a priori known, and the properties of these estimators will remain essentialy unchanged. When ω_0 is unknown and is estimated using $\hat{\omega}_0$, it is more involved to prove that $\hat{d}(\hat{\omega}_0)$ yields a consistent estimator of $d(\omega_0)$. Another challenging problem is to show that the best possible rate of convergence of such an estimator is the same as when ω_0 is known.

8.1 Yajima (1995)

Yajima (1995) [57] has studied what might be considered as the simplest estimator of the G-frequency ω_0, *i.e.* the G-frequency is the frequency at which maximizes the periodogram

$$\hat{\omega}_n = \arg\max_{x \in [0,\pi]} I_n(x)$$

This estimator has long been used to estimate the frequency of a sinewave in white noise.

Theorem 15 (Yajima, 1995 [57] Theorem 1) *Assume that X is a Gaussian process with spectral density that satisfies (80) with f^* differentiable on $[0,\pi] \setminus \{\omega_0\}$ and such that*

$$\forall x \in [0,\pi],\ |f^{*\prime}(x)/f^*(x)| \leq C|x - \omega_0|^{-1}.$$

Then for all positive real $\alpha < 1$, $n^\alpha(\hat{\omega}_n - \omega_0)$ tends to zero in probability.

Remarks

- This theorem can be readily adapted to prove that the same result holds if the periodogram is maximized over a grid of Fourier frequencies. Define $\check{k}_n = \arg\max_{1 \leq k \leq \tilde{n}} I_n(x_k)$ and $\check{\omega}_n = 2\pi\check{k}_n$. Then under the same assumptions on f, it also holds that $n^\alpha(\hat{\omega}_n - \omega_0)$ tends to zero in probability for any $\alpha < 1$.
- The assumptions of Gaussianity is probably not necessary. and, using recent results on the periodogram of an i.i.d. sequence by Davis and Mikosch (1999), it is most likely that the rate of convergence can be improved to any increasing sequence $v_n = o(n/\log(n))$.
- Yajima (1995) has conjectured that $n(\hat{\omega}_n - \omega_0)$ converges in distribution to a heavy-tailed non-normal distribution, probably depending on d.
- It is possible to use this estimator to obtain a semi-parametric estimator of d in (80) by plugging the value $\hat{\omega}_0$ in the estimator presented in the previous section. This would probably yield consistent estimators of d. Since the rate of convergence of this estimator is faster than the best rate of convergence for the GPH estimator or the FEXP estimator, it can be conjectured that, these estimators achieve the same rate of convergence when the G-frequency is known.

8.2 Hidalgo (1999)

Hidalgo (1999) has proposed a more complex method to estimate the location of the pole in (80), which has the advantage of providing an asymptotic distribution to the estimator. We outline the method. Let k_n be an increasing sequence of integers. Let $1 \leq q \leq n$ be an integer. Let $(\gamma_{n,p})_{1 \leq p \leq k_n}$, be non negative real numbers such that $\sum_{p=1}^{k_n} \gamma_{n,p} = 1$. For each p le now $\tilde{\gamma}_{p,l}$, $1 \leq l \leq p$ be real numbers such that

$\sum_{l=1}^{p} \tilde{\gamma}_{p,l} = 0$. Let m_n be a sequence of integers and define $\tilde{I}_{n,j} = \sum_{h=-m_n}^{m_n} I_n(x_h)$. Define

$$\tilde{d}_p(x_q) = \sum_{j=1}^{p} \tilde{\gamma}_{p,j} \log(\tilde{I}_{n,j+q}), \tag{81}$$

$$\hat{d}(x_q) = \sum_{p=1}^{k_n} \gamma_{n,p} \tilde{d}_p(x_q). \tag{82}$$

This somewhat complex smoothing scheme has two motivations. The most important one is to avoid the non Gaussianity of the averaged periodogram estimator for $d > 1/2$. The second one is to reduce the asymptotic variance of the estimator. Define now

$$\hat{q}_n = \arg \max_{0 \le q \le [n/2]} \hat{d}(x_q), \quad \hat{\omega}_n = 2\pi \hat{q}_n / n.$$

(A4) X is a causal linear process with innovation process Z, *i.e.*

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty, \ \phi_0 = 1.$$

The innovation Z is a homoscedastic martingale increment sequence, *i.e.*

$$\mathbb{E}[Z_t | \mathcal{F}_{t-1}] = 0, \ \mathbb{E}[Z_t^2 | \mathcal{F}_{t-1}] = 1, \ \mathbb{E}[Z_t^s | \mathcal{F}_{t-1}] = \mu_s, 1 \le s \le 2r,$$

where $\mathcal{F}_t$ is the σ-field generated by Z_s, $0 \le s \le t$, and $r \le 2$ is an integer.

The function $\psi(x) = \sum_{j=0}^{\infty} \psi(e^{ix})$ is differentiable except at ω_0 and

$$|\psi'(x)| \le C \frac{|\psi(x)|}{|x - \omega_0|}.$$

Theorem 16 *Assume that Assumption* **(A4)** *holds for some* $r \ge 4$. *Let* $\epsilon < 3/(r+1)$. *Then for* $k_n = [n^\epsilon]$, *one can choose a sequence* m_n *and weights* $\gamma_{n,p}$ *and* $\tilde{\gamma}_{p,j}$ *such that*

- *if* $\omega_0 \in (0, \pi)$, *then* $\frac{n}{\pi k_n^{1/2}}(\hat{\omega}_n - \omega_0)$ *is asymptotically centered Gaussian with variance* σ^2/d^2 *where* σ^2 *depends only on the weights ;*
- *if* $\omega_0 = 0$ *(resp.* $\omega_0 = \pi$*), then* $\frac{n}{\pi k_n^{1/2}}(\hat{\omega}_n - \omega_0)$ *is asymptotically distributed as* $(\sigma/\alpha)\xi^+$ *(resp.* ξ^-*) where* ξ *is a* $\mathcal{N}(0,1)$ *r.v. ;*
- $k_n^{1/2}(\hat{d}(\hat{\omega}_n) - d)$ *is asymptotically centred Gaussian with variance* τ^2 *independent of* ω_0 *and* d.

The interest of this method is that it yields an estimator which converges at a rate arbitrarily close to n, (provided the innovation process Z has enough finite moments) and which is moreover asymptotically Gaussian, except at zero and π. Its drawback is that it cannot be used to obtain simultaneously an estimator of the pole and of the memory parmeter since the rate of convergence of one of the estimators can be improved only at the cost of a loss in the rate of convergence of the other estimator. The key ingredient in the derivation of theorem 16 is an invariance principle for the estimator $\hat{d}(x)$ Define $\xi_n(\tau) = k_n(\hat{d}(x_{q_n + [k_n^{1/2}]}) - \hat{d}(x_{q_n})$, where $2\pi q_n/n$ is the Fourier frequency closest to ω_0.

Proposition 1 *ξ_n converges in the space $\mathcal{D}$ of left-limited right-continuous (càdlàg) functions endowed with the topology of uniform convergence on compact sets to the Gaussian process ξ defined by*

$$\xi(\tau) = c_1 d\tau^2 + c_2 \tau \zeta,$$

where ζ is standard Gaussian.

8.3 Giraitis, Hidalgo and Robinson (1999)

Even though it was developed in a parametric framework, the method of Giraitis, Hidalgo and Robinson (1999) [22] is worth mentioning since it might be adapted to a semi-parametric context. Assume that X is a weakly stationary process with spectral density $f(\omega_0, \theta_0, .)$ belonging to a parametric model $f(\omega, \theta, .)$, $\omega \in [0, \pi]$, $\theta \in \Theta$, satisfying the following regularity assumptions.

(A5) $f(\omega,\theta,x) = \frac{\sigma^2}{2\pi} k(\omega,\theta,x)$, $k(\omega,\theta,x) = |4\sin((x+\omega)/2)\sin((x+\omega)/2)|^{-2d} h(\theta,x)$, where $0 < d < 1$ if $0 < \omega < \pi$ and $0 < d < 1/2$ if $\omega \in \{0,1\}$, h is a C^2 function and even with respect to x, and for all $\omega \in [0,\pi]$ and all $\theta \in \Theta$, $\int_{-\pi}^{\pi} \log(k(\omega,\theta,x))dx = 0$.

In the parametric framework considered here, an identifiability condition is also necessary.

(A6) $\inf_{\omega,\theta) \in [0,\pi]\times\Theta} \int_{-\pi}^{\pi} \frac{k(\omega_0,\theta_0,x)}{k(\omega,\theta,x)} dx = 2\pi$.
$\int_{-\pi}^{\pi} \{k(\omega_0,\theta_0,x) - k(\omega,\theta,x)\}^2 dx > 0$ for all $(\omega,\theta) \neq (\omega_0,\theta_0)$.
The matrix $\Omega_0 = \frac{1}{4\pi} \int_{-\pi}^{\pi} \nabla_\theta \log k(\omega_0,\theta_0,x) \left(\nabla_\theta \log k(\omega_0,\theta_0,x)\right)^T dx$ is positive definite.

The estimator of (ω_0, θ_0) is obtained by minimizing a discretized Whittle contrast over a discrete grid of Fourier frequencies. Define

$$S(\omega,\theta) = [n/2]^{-1} \sum_{j=0}^{[n/2]} \frac{I_n(x_j)}{k(x_j,\omega,\theta)}$$
$$(\hat{\omega},\hat{\theta}) = \arg \min_{0 \le j \le [n/2], \theta\in\Theta} S(x_j,\theta), \quad \hat{\sigma}^2 = S(\hat{\omega},\hat{\theta}).$$

Theorem 17 *Assume that assumption* **(A4)** *holds with $r = 2$ and that moreover there exits a real $\eta > 0$ such that $\sup_t \mathbb{E}[|Z_t|^{4+\eta}] < \infty$. Assume also that* **(A5)** *and* **(A6)** *hold. Then $n(\hat{\omega}_n - \omega_0) = O_P(1)$ and $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically zero-mean Gaussian with variance Ω_0^{-1}.*

It has not been proved even in the parametric context that the rate n is the best rate for the estimation of ω_0.

9 Theoretical and technical tools for the periodogram

The study of the estimators outlined above involve linear and non-linear functionals of the normalized periodogram of the form

$$S_{m,p,n}(X,\phi) := \sum_{k=1}^{K(m,p,n)} \beta_{n,k}\left(\phi(\bar{I}^X_{m,p,n,k}/f(y_k) - \gamma_{m,p,n}(\phi)\right) \tag{83}$$

where $(\beta_{n,k})$ is a triangular array of numbers and ϕ is a (possibly non-linear) function verifying $\mathbb{E}[|\phi(|W_{m,p,n}|^2/2)|^2] < \infty$, where $W_{m,p,n}$ is a Gaussian vector with covariance given by (15). We will mainly focus on the following two cases: $\phi(x) = x$ (GSE and FAR estimators) and $\phi(x) = \log(x)$ (GPH and FEXP estimators). It is possible to consider more general non-linear functions, at the expense of some additional technicalities.

The purpose of this section is to give a sketch of the techniques involved to study such triangular arrays, which will in turn justify some of the technical assumptions introduced to prove the consistency and the asymptotic normality of the estimators.

9.1 Methodology

The usual technique to prove a CLT for a functional of the periodogram of a non-Gaussian linear process makes use of the so-called Bartlett's decomposition. This decomposition amounts to decompose the periodogram as follows

$$I^X_{m,p,n,k} = (2\pi)f(y_k)I^Z_{m,p,n,k} + R_{m,p,n,k}, \tag{84}$$

where f is the spectral density of the process X and $R_{m,p,n,k}$ is a remainder term. This decomposition of the periodogram was suggested in Bartlett (1955) and later thoroughly investigated by many authors (see, *e.g.* Walker (1965) [56], Chen and Hannan (1980) [10], Brockwell and Davis (1991) [9] for short range dependent processes; see Robinson (1995b) [46] for long-range dependent processes).

The leading term in the decomposition (84), $2\pi f(y_k)I^Z_{m,p,n,k}$, is sometimes referred to as the *pseudo-periodogram*. The Bartlett's decomposition suggests to prove a central limit theorem for $S_{m,p,n}(X,\phi)$ in two steps

- Prove a CLT for $S_{m,p,n}(Z,\phi)$,
- Show that $S_{m,p,n}(X,\phi) - S_{m,p,n}(Z,\phi)$ is asymptotically negligible (*i.e.* converges to zero in probability).

For linear functionals of the periodogram, the proof for these two steps are easily carried out and lead to a very satisfactory result, see Theorem 19 below. In the case of non linear functionals of the periodogram, these steps are both technically involved and lead to partial results under rather stringent assumptions on the innovation sequence Z and the function ϕ. Some of these assumptions can be weakened by assuming that X is a Gaussian process: the technique of proof in such case does not make use of the Bartlett's decomposition, and stronger results can be obtained.

9.2 Linear functionals

As already mentioned, for linear functionals, pooling is irrelevant, thus we assume in this section that $m = 1$. The Bartlett's decomposition then reads

$$S_{p,n}(X) = \sum_{k=1}^{K} \beta_{n,k}(I_{p,n}^X(x_k)/f(x_k) - 1) = \sum \beta_{n,k}(2\pi I_{p,n}^Z(x_k) - 1)$$
$$+ \sum_{k=1}^{K} \beta_{n,k}(I_{p,n}^X(x_k)/f(x_k) - 2\pi I_n^Z(x_k)) =: S_{p,n}(Z) + R_{p,n}$$

and it must be proved that $S_{p,n}(Z)$ is asymptotically Gaussian and that $R_{p,n}$ is asymptotically negligible. The proof of the central limit theorem for $S_{p,n}(Z)$ makes use of standard arguments on martingale increments sequences. The weight sequence must satisfy a strengthened Lindeberg-Lyapunov assumption.

(A7) $(\beta_{n,k})_{1 \le k \le K}$ is a triangular array of real numbers such that

$$\sum_{k=1}^{K} \beta_{n,k}^2 = 1, \tag{85}$$

$$\lim_{n\to\infty} \max_{1\le k\le K} |\beta_{n,k}| = 0, \tag{86}$$

$$\lim_{n\to\infty} n^{-1} \sum_{j\neq k=1}^{K} \beta_{n,j}\beta_{n,k} = \tau. \tag{87}$$

where $\varsigma_{p,n}$ is defined in (14).

(A8) $(\beta_{n,k})_{1 \le k \le K}$ is a triangular array of real numbers such that

$$\lim_{n\to\infty} \sum_{j,k=1}^{K} \beta_{n,k}\beta_{n,j}|\varsigma_{p,n}(j-k)|^2 = L_{2,p}. \tag{88}$$

(85) and (86) are the usual smallness assumptions to prove a CLT for a triangular array of i.i.d. r.v.'s. To see that (87) and (88) is necessary if Z is a non Gaussian white noise, note that, for $1 \le j,k \le K(m,p,n)$

$$4\pi^2\mathbb{E}[(I_{p,n}^Z(x_k) - 1)(I_{p,n}^Z(x_j) - 1)] = |\varsigma_{p,n}(j-k)|^2 + \kappa_4 n^{-1} a_p^{-2} n^{-1} \sum_{t=1}^{n} |h_{t,n}|^4.$$

and that $n^{-1}\sum_{t=1}^{n} |h_{t,n}|^4 = a_{2p}$ exists and is finite. Hence

$$\mathbb{E}\left\{\sum_{k=1}^{K} \beta_{n,k}(2\pi I_{p,n}(x_k) - 1)\right\}^2 = (2\pi)^2 \sum_{1\le k,j\le K} \beta_{n,k}\beta_{n,j}|\sigma_{p,n}(j-l)|^2 + \kappa_4(a_{2p}/a_p^2)\left\{n^{-1}\sum_{k\neq j=1}^{K} \beta_{n,k}\beta_{n,j}\right\}$$

Thus it is necessary that the limits (87) and (88) hold to prove the convergence of the variance of $S_{p,n}(Z)$. In this section and the following, the assumptions on the spectral density will be given in terms of the function a^* defined in(47).

Theorem 18 *Let $\mu > 1$, $\alpha \in (0, \pi]$ and $p \in \mathbb{N}$. Assume that X is a process such that* **(A3)** *holds and such that $-p - 1/2 < d < 1/2$ and $a^* \in \mathcal{L}^*(\mu, \alpha)$. Assume moreover that* **(A7)** *and* **(A8)** *hold. Then $S_{p,n}(Z)$ converges in distribution to a centred Gaussian law with variance $L_{2,p} + \kappa_4(a_{2p}/a_p^2)\tau$.*

Note that (85) and (87) imply that $\tau = \lim_n \left\{n^{-1/2} \sum_{k=1} \beta_{n,k}\right\}^2$. Hence, $\tau = 0$ when $\sum_{k=1}^n \beta_{n,k} = 0$. This condition holds for the GSE estimate, explaining why the fourth-order cumulant does not appear in the expression of the asymptotic variance of this estimator.

Lemma 2 *Under the assumptions of theorem 18, $R_{p,n} = o_P(1)$. If in addition $\lim_{n\to\infty} \beta_n^* \log^2(n) = 0$, with $\beta_n^* := \max_{1\le k\le K} |\beta_{n,k}|$, then $\mathbb{E}[R_{p,n}^2] = o(1)$.*

From this Lemma and Theorem 18, the asymptotic normality of $S_{p,n}(X)$ follows.

Theorem 19 *Under the assumptions of theorem 18,*

$$\sum_{k=1}^{K} \beta_{n,k}(I_n^X(x_k)/f(x_k) - 1) \to_d \mathcal{N}(0, L_{2,p} + \kappa_4(a_{2p}/a_p^2)\tau).$$

If moreover $\lim_{n\to\infty} (\beta_n^*)^2 \log^3(n) = 0$, then

$$\lim_{n\to\infty} \mathbb{E}\left[\left\{\sum_{k=1}^{K} \beta_{n,k}(I_n^X(x_k)/f(x_k) - 1)\right\}^2\right] = L_{2,p} + \kappa_4(a_{2p}/a_p^2)\tau.$$

9.3 Non linear functionals, linear non-Gaussian case

The case of non linear functionals is more involved and additional technical assumptions are needed on the distribution of the sequence Z and on the class of functions for which a CLT can be proved ϕ is restricted. For the sake of simplicity and since the applications presented here mainly concern the case $\log(x) = \phi(x)$, and we will only present the results under this assumption. The proof for non-linear functionals closely follows the proof for linear functionals. Define

$$U_{n,k} := \frac{f^{-1}(y_k)\bar{I}^X_{m,p,n,k} - 2\pi \bar{I}^Z_{m,p,n,k}}{2\pi \bar{I}^Z_{m,p,n,k}}$$

and $R_{n,k} = \log(1 + U_{n,k})$. Then

$$\begin{aligned}\sum_{k=1}^{K} \beta_{n,k}(\log(\bar{I}^X_{m,p,n,k}) - \gamma_{m,p,n}) &= \sum_{k=1}^{K} \beta_{n,k}(\log(2\pi \bar{I}^Z_{m,p,n,k}) - \gamma_{m,p,n}) + \sum_{k=1}^{K} \beta_{n,k} R_{n,k} \\ &=: S_{p,n}(Z, \log) + R_{p,n}.\end{aligned}$$

The proof then amounts to prove that $S_{p,n}(Z, \log)$ is asymptotically normal and that $R_n = o_P(1)$. A Central Limit Theorem for triangular arrays of non-linear functions of the periodogram of i.i.d r.v.'s has recently been obtained by Fay and Soulier (1999) [13]. We specialize their result to the case $\phi(x) = \log(x)$. An additional assumption is also needed on the weight sequence.

(A9) $(\beta_{n,k})_{1\le k\le K}$ is a triangular array of real numbers such that for all $\epsilon > 0$,

$$\max_{1\le k\le K} |\beta_{n,k}| = O(\mu_n^{-1/2+\epsilon}), \tag{89}$$

where $\mu_n := \#\{k : 1 \le k \le K, \beta_{n,k} \neq 0\}$.

This assumption implies that $\mu_n(\max_{1\le k\le K} |\beta_{n,k}|)^2$ is bounded by a slowly varying function of μ_n, the cardinal of the support of the weights $\beta_{n,k}$. This condition obviously holds for the GPH estimator. For the FEXP estimator, this assumption does not hold in general, because $\mu_n = K$ and $\max_{1\le k\le K} |\beta_{n,k}| \asymp \sqrt{q/n}\log(n)$. Nevertheless, this difficulty can be alleviated by truncating the sum $S_{p,n}(Z, \log)$ (see Hurvich, Moulines and Soulier (1999) [30]). The proof of this CLT is based on the Fréchet-Sohat moment technique, which amounts to show that the moments of $S_{p,n}(Z, \phi)$ converge to the moments of a Gaussian r.v. This technique requires the evaluation of moments of the form

$$\mathbb{E}[\log^{r_1}(\bar{I}^Z_{m,p,n,k_1}) \cdots \log^{r_u}(\bar{I}^Z_{m,p,n,k_u})]$$

which is done by using Edgeworth expansion techniques (see Battacharya and Rao, 1976). Since $\log(x)$ is singular at zero frequency, it is required to obtain an expansion of the p.d.f of $(d_n^{(Z,h)}(x_{n,k_1}), \cdots, d_n^{(Z,h)}(x_{n,k_p})$ for any $p \ge 1$ and any p-uplets of pairwise distinct integers $(k_1, \cdots, k_p)$. The validity of this expansion relies upon (43), which is a strengthening of the usual Cramer's conditions. The original idea to use Edgeworth expansions in this context is due to Chen and Hannan (1980) [10] and Velasco (1999) [53] obtained a central limit theorem for the GPH estimator for a non-Gaussian process.

Theorem 20 *Let $\mu > 1$, $p \in \mathbb{N}$ and $m \ge 4$. Assume that the process X satisfies* **(A2)** *with $\mathbb{E}|Z|^{2m+2} < \infty$, $a^* \in \mathcal{L}^*(\mu, \pi)$ and $-p - 1/2 < d < 1/2$. Assume moreover that* **(A7)** *and* **(A9)** *hold. Then,*

$$\sum_{k=1}^{K} \beta_{n,k}\{\log(2\pi \bar{I}^Z_{m,n,p,k}) - \gamma_{m,p,n}\} \to_d \mathcal{N}(0, \sigma^2_{p,m} + \kappa_4 \tau C_{m,p})$$

where $C_{m,p}$ is a positive constant.

A closed-form expression of $C_{m,p}$ can be found in Fay and Soulier (1999). Its expression is not relevant here since $\tau = 0$ for the application presented here (the GPH and the FEXP estimators). The treatment of the remainder term needs a strengthening of the assumption on f^*.

Lemma 3 *Let $\mu > 1$, $\rho > 1$, $p \in \mathbb{N}$ and $m \ge 4$. Assume that the process X satisfies* **(A2)** *with $\mathbb{E}|Z|^8 < \infty$, $a^* \in \mathcal{L}^*(\mu, \rho, \pi)$ and $-p - 1/2 < d < 1/2$. Assume moreover that $\sum_{k=1}^K \beta^2_{n,k} = 1$. Then $\sum_{k=1}^K \beta_{n,k} R_{n,k}$ converges in probability to zero.*

Remark The proof of this lemma also makes use of Edgeworth's expansions and does not even guarantee that $\mathbb{E}|\sum_{k=1}^K \beta_{n,k} R_{n,k}|$ exists. Note also that the control of the remainder term critically depends upon the fact that $f^* \in \mathcal{L}^*(\mu, \rho, \pi)$ that $p \ge 1$, making use of the "bias reduction" in the Bartlett approximation afforded by the taper. We can now state

Theorem 21 *Let $\mu > 1$, $\rho > 1$, $p \in \mathbb{N}$ and $m \geq 4$. Assume that the process X satisfies* **(A2)** *with $\mathbb{E}|Z|^{2m+2} < \infty$, $a^* \in \mathcal{L}^*(\mu, \rho, \pi)$ and $-p - 1/2 < d < 1/2$. Assume moreover that* **(A7)** *and* **(A9)** *hold. Then,*

$$\sum_{k=1}^{K} \beta_{n,k} \{\log(2\pi \bar{I}^Z_{m,n,p,k}) - \gamma_{m,p,n}\} \to_d \mathcal{N}(0, \sigma^2_{p,m} + \kappa_4 \tau C_{m,p}).$$

9.4 Non linear functionals, Gaussian case

As far as convergence in distribution is concerned, the Gaussian case can be viewed as a subcase of the linear case, except for a mild restriction in Theorem 4. The assumption of Gaussianty is used at its full strength in the derivation of the minimax and adaptive results which proved (up to now) only under the assumption that the process X is Gaussian. In this section we present the specific tools that are used to derive these results.

Recall that the periodogram, wether raw or tapered or pooled can always be expressed as the square modulus of a Gaussian vector, and the the log-periodogram for instance can be viewed as a function of a Gaussian vector. When computing the variance of a weighted sum of log-periodogram ordinates, it is thus necessary to obtain bounds for quantities such as $\mathrm{cov}(\phi(\xi_1), \phi(\xi_2))$ where ξ_1, ξ_2 are Gaussian vetors. It is well known that these bounds can be expressed in terms of the covariances between the components of ξ_1 and ξ_2 and of the Hermite rank of the function ϕ. Such bounds were first established by Taqqu (1977) [51] in the case of Gaussian *variables* and a covariance bound for functions of Gaussian *vectors* was first proved by Arcones (1994) [1]. In order to derive central limit theorems for weighted sum of log-periodogram ordinates, it is necessary to resort to the so-called method of moments. This method requires computation of moments of product of functions of Gaussian vectors such as $\mathbb{E}[\phi_1(\xi_1) \cdots \phi_u(\xi_u)]$. The needed bounds were obtained in Taqqu (1977) for Gaussian variables and generalized by Soulier (1998) [48] for functions of Gaussian vectors. Finally, as described in section 7.1, the key tool to prove the adaptivity properties of estimators is an exponetial inequality. We now introduce some notations and definitions. Let $\boldsymbol{\xi}$ denote a standard a dimensional Gaussian vector. For a measurable function $\phi : \mathbb{R}^a \to \mathbb{R}$, denote $\|\phi\| = \left\{\mathbb{E}[\phi^2(\boldsymbol{\xi})]\right\}^{1/2} = \left\{(2\pi)^{-a/2} \int_{\mathbb{R}^a} \phi^2(x) e^{-\frac{1}{2}x^T x} dx\right\}^{1/2}$. Let H_k denote the k-th Hermite polynomial. Any function ϕ such that $\|\phi\| < \infty$ can be expanded as

$$\phi = \sum_{k_1, \cdots, k_a \geq 0} \frac{c(k_1, \cdots, k_a)}{k_1! \cdots k_a!} H_{k_1} \cdots H_{k_a},$$

where $c(k_1, \cdots, k_a) = \mathbb{E}^0[(\phi(X) H_{k_1}(X_1) \cdots H_{k_a}(X_a)]$ is well defined if $\|\phi\| < \infty$. The Hermite rank of ϕ is the smallest integer τ such that there exist integers $k_1, \cdots, k_a$ which satisfy $k_1 + \cdots + k_a = \tau$ and $c(k_1, \cdots, k_a) \neq 0$. The Hermite rank of ϕ can be defined equivalently as the smallest integer τ such that there exists a polynomial P of degree τ with $\mathbb{E}[P(\boldsymbol{\xi})\phi(\boldsymbol{\xi})] \neq 0$. ϕ has Hermite rank 0 if and only if $\mathbb{E}^0[\phi(\boldsymbol{\xi})] \neq 0$. The Hermite rank of a function is defined here only with respect to the standard a-dimensional Gaussian distribution.

Let $u, d_1, \cdots, d_u$ be positive integers and let $d = d_1 + \cdots + d_u$. Let X be a d-dimensional centered Gaussian vector with covariance matrix Γ. Assume that X can be written as $X = (X_1^T, \cdots, X_u^T)^T \in \mathbb{R}^d$, where $X_i \in \mathbb{R}^{d_i}$ is a d_i dimensional standard Gaussian vector. Consider functions $\phi_i : \mathbb{R}^{d_i} \to \mathbb{R}$ with Hermite rank τ_i, $1 \leq i \leq u$.

Theorem 22 *Let δ denote the spectral radius of the matrix $\Gamma - I_d$ and let $\tau^* = \tau_1 + \cdots \tau_u$. For all real $\epsilon > 0$, there exists a constant $c_1(\epsilon, d, \tau^*)$ which depends only on ϵ, d and τ^* such that*

$$\left|\mathbb{E}\left(\prod_{i=1}^{u} \phi_i(X_i)\right)\right| \leq c_1(\epsilon, d, \tau^*) \prod_{i=1}^{u} \|\phi_i\|_2 \; \delta^{\tau^*/2}. \tag{90}$$

Remarks

- The condition $1-\delta \leq 1-\epsilon$ ensures that Γ ise invertible. this is obviously not superfluous. Consider for instance the case $X_1 = X_2 = X_3 = X$ and let ϕ be a function such that $\mathbb{E}[\phi^2(X)] < \infty$ and $\mathbb{E}[|\phi(X)|^3] = \infty$. Then (90) does not hold. It is however not clear wether the condition that the spectral radius of $\Gamma - I_d$ is strictly less than one, which implies the invertibility of Γ, is necessary.
- Let ρ^* be the maximum correlation between coordinates of any two vectors X_i and X_j for $i \neq j$, *i.e.* $\rho^* = \max_{1 \leq i < j \leq u} \max_{1 \leq k \leq d_i, 1 \leq l \leq d_j} |\mathbb{E}[X_i^{(k)} X_j^{(l)}]|$. A sufficient condition for $\delta < 1 - \epsilon$ is $\rho^* \leq (1-\epsilon)/(d-1)$. In the case of one-dimensional Gaussian variables, (90) can be obtained by combining Lemma 3.1, 3.2, 3.3 and Proposition 4.2 of Taqqu (1977), under the assumption $\rho^* \leq \epsilon$ for some $\epsilon < 1/(u-1)$. Thus theorem 22 is a bit more general.
- In order to prove a central limit theorem, the following weaker form of Theorem 22 is sufficient. If it is only assumed that s among the functions ϕ_i have Hermite rank $\tau_i \geq \tau$ for some integers $s \leq u$ and τ, if $\rho^* \leq (1-\epsilon)/(d-1)$ for some $\epsilon > 0$, then there exists a constant $c_2(\epsilon, d, \tau)$
$$\left|\mathbb{E}\left[\prod_{i=1}^{u} \phi_i(X_i)\right]\right| \leq c_2(\epsilon, d, \tau) \prod_{i=1}^{u} \|\phi_i\|_2 \; {\rho^*}^{s\tau/2}. \tag{91}$$
- For $s = u = 2$, (90) is Lemma 1 in Arcones (1993), and the constant is actually 1, *i.e.*, it can be written as
$$|\mathrm{cov}(\phi_1(X_1), \phi_2(X_2))| \leq \|\phi_1\|_2 \|\phi_2\|_2 \delta^{\tau}. \tag{92}$$
- Theorem 22 is based on an expansion in powers, rather than in Hermite polynomials as in Taqqu (1977) or Arcones (1994), of the density of the multivariate Gaussian distribution. This technique has been used by Robinson (1995) [47] to prove the asymptotic normality of the GPH estimator.

This theorem can be used to prove a central limit theorem by means of the method of moments, since it allows an evaluation of moments of a sum of functions of jointly Gaussian vectors. Combined with the bounds in Lemma 6 it yields central limit theorem for weighted sums of log-periodogram ordinates (cf. Moulines and Soulier (1999) [41] Theorem in the non tapered case).

Another important consequence of the Gaussian assumption is the expoenential inequality used to prove the adaptive properties of the GPH and FEXP estimators (see section 7.1). It is a consequence of the following simple inequality.

Proposition 2 *Let X be a d-dimensional centered Gaussian vector with covariance matrix Γ. Let $\psi : \mathbb{R}^d \to \mathbb{R}$ be a function such that $\|\psi\| < \infty$. Let $0 < \epsilon < 1$. There exists a constant $c_3(\epsilon)$ (which depends only on ϵ) such that if* $\mathrm{tr}\{(\Gamma - I_d)^2\} \leq 1 - \epsilon$,

$$|\mathbb{E}[\psi(X)]| \leq c_3(\epsilon) \|\psi\|_2. \tag{93}$$

The bound (93) is obviously weakest than (90), but its interest is that it is independent of the dimension of the Gaussian vector X. The main ingredient of its proof is simply Hölder inequality. It can be used to prove an exponential inequality for sums of functions of jointly Gaussian vectors. Let $X_1, \cdots, X_n$ be jointly d-dimensional Gaussian vectors such that the covariance matrix Γ_n of $X = (X_1, \cdots, X_n)$ satisfies the assumption of Proposition 2, *i.e.* $\mathrm{tr}\{(\Gamma - I_d)^2\} \le 1 - \epsilon$ for some $\epsilon > 0$ independent of n. Let now ϕ be a function and $\lambda_1, \lambda_2 \in [-\infty, +\infty$ such that for all $\lambda \in (\lambda_1, \lambda_2)$,

$$\mathbb{E}[e^{\lambda\phi(X_i)}] \le C_1 e^{C_\phi \lambda^2/2}.$$

Let now β_k, $1 \le k \le n$, be real numbers and define $\beta_n^* = \max_{1 \le k \le n} |\beta_k|$. Then, for all $\alpha \le \inf\{|\lambda_1|, |\lambda_2|\}/\beta_n^*$ and all $\rho > 0$,

$$\mathbb{E}[\exp\{\alpha \sum_{k=1}^{n} \beta_{n,k}\phi(X_i)] \le C(\epsilon, \rho) e^{C_\phi \alpha^2/2(1+\rho)}.$$

This inequality can be applied when the vectors X_i are the tapered DFT's of a Gaussian process with some adaptation (see Giraitis, Robinson and Samarov (2000) [16] and Iouditsky, Moulines and Soulier (1999) [32]). It can be understood from the bounds in lemma 6 that the use of taper is (technically) necessary to obtain this exponential inequality. Without a taper, the assumption of proposition 2 cannot be fulfilled.

9.5 Moment bounds for the tapered DFT's

The theorems in the previous sections are based on the probability tools already mentioned and on technical tools which are mainly bounds for moments of the approximation errors $d^X_{p,n}(x_k)/a(\tilde{x}_{p,k}) - d^Z_{p,n}(x_k)$ where $\tilde{x}_{p,k} := (p+1)^{-1}(x_k + \cdots + x_{k+p})$. The derivation of these bounds usually involve lengthy and tedious computations. The first such bounds were obtained by Robinson (1994,1995) [44, 46, 47] and later different bounds were obtained using the same basic ideas but different assumptions by Giraitis, Robinson and Samarov (1997) [15], Hurvich Deo and Brodsky (1998) [29] and Moulines and Soulier (1999). Similar bounds for the tapered DFT's were then investigated by Velasco (1999) [52, 53, 54], Giraitis, Robinson and Samarov (2000) [16], Hurviuch and Chen (1999) [27] and Hurvich, Moulines and Soulier (2000) [30]. From these bounds, bounds for the moments of the approximation errors of the pooled tapered periodogram by the pseudo-periodogram, *i.e.* $\bar{I}^X_{m,p,n,k}/f(y_k) - (2\pi)\bar{I}^Z_{m,p,n,k}$ are derived. We present here some of the bounds needed to prove the results of the previous sections and for a reference for future research in this field. It must be noted that all of the following bounds are valid under the rather weak assumption **(A3)**, with some strenghtening of the moment conditions when moments higher than 2 are computed. Define

$$u_{p,n,k} = \frac{1}{\sqrt{2\pi n a_p}} \sum_{t=1}^{n} h^p_{t,n} X_t \exp(-itx_k),$$
$$v_{p,n,k} = \frac{1}{\sqrt{2\pi n a_p}} \sum_{t=1}^{n} h^p_{t,n} Z_t \exp(-itx_k)$$

General assumptions In all the following lemmas, unless otherwise stated, p is a nonnegative integer, $0 \le \Delta < p + 1/2$, $0 < \delta < 1/2$, $\mu > 1$, $\rho > 1$, $\alpha \in (0, \pi]$ and C denotes a generic constant that depends only on these parameters. The bounds presented in these lemmas hold uniformly for any process X such that **(A3)** holds with spectral density $|1 - e^{ix}|^{-2d}|a^*(x)|^2$, where $-\Delta \le d \le \delta$ and $a^* \in \mathcal{L}^*(\mu, \alpha)$.

Lemma 4 *For all $n \geq 1$, for all $1 \leq k \leq j-p \leq [(n-2p-1)/2]\alpha/\pi$,*

$$|\mathbb{E}[|u_{p,n,j}|^2]-1| \leq r_p(j), \ |\mathbb{E}[u_{p,n,j}\overline{v_{p,n,j}}]-1]| \leq Cr_p(j), \tag{94}$$
$$|\mathbb{E}[u^2_{p,n,j}]| \leq Cr'_p(j), \ |\mathbb{E}[u_{p,n,j}v_{p,n,j}]| \leq Cr'_p(j) \tag{95}$$

where

$$r_p(j) = \begin{cases} \log(1+j)/j & p=0 \\ 1/j & p \geq 1 \end{cases}$$

and

$$r'_p(j) = \begin{cases} \log(1+j)/j & p=0 \\ 1/j^p & p \geq 1 \end{cases}$$

If in addition $p \geq 1$ and $a^* \in \mathcal{L}^*(\mu,\rho,\pi)$, then

$$|\mathbb{E}[|u_{p,n,j}|^2]-1| + |\mathbb{E}[u_{p,n,j}\overline{v_{p,n,j}}]-1]| \leq Cj^{-\rho}. \tag{96}$$

Lemma 5 *Assume that all the moments of the sequence Z are the same as in the case of an i.i.d. sequence with finite moments up to the order $2s$ for some positive integer. For all $n \geq 2p$ and for all $1 \leq k \leq [(n-2p-1)/2]\alpha/\pi$,*

$$\mathbb{E}\left[\left(\bar{I}^X_{m,p,n,k}/f(y_k) - (2\pi)\bar{I}^Z_{m,p,n,k}\right)^{2s}\right] \leq Cr^s_p(k). \tag{97}$$

If moreover $a^* \in \mathcal{L}^*(\mu,\rho,\alpha)$, then

$$\mathbb{E}\left[\left(\bar{I}^X_{m,p,n,k}/f(y_k) - (2\pi)\bar{I}^Z_{m,p,n,k}\right)^{2s}\right] \leq Ck^{-\rho s}. \tag{98}$$

Lemma 6 *Assume that $\alpha = \pi$, $n \geq 2p$ and $0 \leq k \leq j-p \leq [(n-2p-1)/2]$. Then*

$$|\mathbb{E}[u_{p,n,k}u_{p,n,j}]| + |\mathbb{E}[u_{p,n,k}\overline{u_{p,n,j}}]| \leq Cr_p(d;k,j), \tag{99}$$
$$|\mathbb{E}[u_{p,n,k}\overline{v_{p,n,j}}]| + |\mathbb{E}[u_{p,n,k}v_{p,n,j}]| + |\mathbb{E}[v_{p,n,k}\overline{u_{p,n,j}}]| + |\mathbb{E}[v_{p,n,k}u_{p,n,j}]| \leq Cr_p(d;k,j). \tag{100}$$

where,

$$r_p(k,j) := \begin{cases} \log(j)k^{-|d|}j^{1-|d|} & (p=0) \\ k^{-p}(j-k)^{-p}(k(j/k)^d \quad + j(k/j)^d) + (j-k)^{1-p}(k^{-1}(j/k)^d & \\ \quad + j^{-1}(k/j)^d + j^{1-2p}(k/j)^d) & (p \geq 1) \end{cases} \tag{101}$$

If $n \geq 2p$ and $0 \leq j-k \leq p$, then

$$|\mathbb{E}[u_{p,n,k}u_{p,n,j}] - \frac{1}{2\pi a_p}(-1)^{j-k}e^{i(j-k)\pi/n}\binom{2p}{p+j-k-1}| \leq Cr_p(d;k,j) \tag{102}$$

Remark As shown by the previous results, tapering, which is necessry to study non-invertible (or non-stationary) processes, has also the very important effect decrease the correlation between DFT's and bias of the approximation of the periodogram by the pseudo-periodogram (in the latter case, only if the spectral density is sufficiently smooth). Similar effect had been observed long ago for short-memory time series (see, *e.g.* Brillinger (1981) [8] for early references).

References

[1] M. A. Arcones. Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Annals of Probability*, 15(4):2243–2274, 1994.

[2] J. M. Azaïs and G. Lang. Estimation de l'exposant de longue dependance dans un cadre semi-parametrique. *C. R. Acad. Sci., Paris, Sér. I*, 316(6):611–614, 1993.

[3] M. S. Bartlett. *An introduction to stochastic processes.* Cambridge University Press, 1955.

[4] J. Beran. Fitting long memory models by generalized regression. *Biometrika*, 80:817–822, 1993.

[5] J. Beran, R. J. Bhansali, and D. Ocker. On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. *Biometrika*, 85:921–934, 1998.

[6] R. N. Bhattacharya and R. R. Rao. *Normal approximation and asymptotic expansions.* Wiley, 1st edition, 1976.

[7] P. Bloomfield. An exponential model for the spectrum of a scalar time series. *Biometrika*, 60:217–226, 1973.

[8] D. R. Brillinger. *Time Series. Data analysis and theory.* Holden-Day, 1981.

[9] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods.* Springer Series in Statistics. New York, Springer-Verlag, 1991.

[10] Z.-G. Chen and E. J. Hannan. The distribution of periodogram ordinates. *J. of Time Series Analysis*, 1:73–82, 1980.

[11] R. Deo. Asymptotic theory for certain redression models with long memory errors. *J. of Time Series Analysis*, 18(4):385–394, 1997.

[12] S. Yu. Efroimovich and M. S. Pinsker. Estimation of square integrable density on the basis of a sequence of observations. *Problems of information transimission*, 17(3):182–196, 1982.

[13] G. Fay and Ph. Soulier. The periodogram of an i.i.d. sequence. Prépublication de l'université d'Evry val d'Essonne, 1999.

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

[15] L. Giraitis, P. M. Robinson, and A. Samarov. Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence. *J. of Time Series Analysis*, 18:49–61, 1997.

[16] L. Giraitis, P. M. Robinson, and A. Samarov. Adaptive rate optimal estimation of the long memory parameter. *J. Multivariate Analysis*, 00:00–00, 2000.

[17] L. Giraitis, P. M. Robinson, and D. Surgailis. A model for long memory conditional heteroskedasticity. Preprint, 1999.

[18] C. W. J. Granger and R. Joyeux. An introduction to long memory time series and fractional differencing. *J. of Time Series Analysis*, 1:15–30, 1980.

[19] H. L. Gray, N. F. Zhang, and W. A. Woodward. On generalized fractional processes. *J. of Time Series Analysis*, 10:233–257, 1989.

[20] E. J. Hannan. *Multiple time series.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1970.

[21] M. Henry and P. M. Robinson. Bandwith choice in Gaussian semiparametric estimation of long range dependence. In *Athens conference on applied probability and time series. Eds. P. M. Robinson and M. Rosenblatt*, 1996.

[22] J. Hidalgo, L. Giraitis, and P. M. Robinson. Gaussian estimation of parametric spectral density with unknown pole. Preprint, 1999.

[23] J. R. M. Hosking. Fractional differencing. *Biometrika*, 60:165–176, 1991.

[24] C. M. Hurvich and K. Beltrao. Asymptotics of the low-frequency ordinates of the periodogram of a long-memory time series. *J. of Time Series Analysis*, 14:455–472, 1993.

[25] C. M. Hurvich and K. Beltrao. Automatic semiparametric estimation of the memory parameter of a long-memory time series. *J. of Time Series Analysis*, 15(3):285–302, 1994.

[26] C. M. Hurvich and J. Brodsky. Broadband semiparametric estimation of the memory parameter of a long memory time series using fractional exponential models. Technical report, New York University Leonard Stern School of Business, SOR-97-2, 1997.

[27] C. M. Hurvich and W. W. Chen. An efficient taper for potentially overdifferenced long-memory time series. *forthcoming in J. Time Series Analysis*, 1999.

[28] C. M. Hurvich and R. Deo. Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long-memory time series. *J. of Time Series Analysis*, 20(3):331–341, 1999.

[29] C. M. Hurvich, R. Deo, and J. Brodsky. The mean squared error of geweke and porter-hudak's estimator of a long-memory time-series,. *J. of Time Series Analysis*, 19:19–46, 1998.

[30] C. M. Hurvich, E. Moulines, and Ph. Soulier. The FEXP estimator for non Gaussian, potentially no stationary processes. Preprint, 2000.

[31] C. M. Hurvich and B. K. Ray. Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *J. of Time Series Analysis*, 16(1):17–41, 1995.

[32] A. Iouditsky, E. Moulines, and Ph. Soulier. Adaptive estimation of the fractional differencing coefficient. Prépublication de l'université d'Evry val d'Essonne, 1999.

[33] N. L. Johnson and S. Kotz. *Continuous univariate distributions I.* New York, Wiley, 1970.

[34] P. S. Kokoszka and R. J. Bhansali. Estimation of the long memory parameter by fitting fractionally differenced autoregressive models. Preprint, 1999.

[35] P. S. Kokoszka and M. S. Taqqu. Fractional ARIMA with stabel innovations. *Stochastic processes and their applications*, 73:79–99, 1995.

[36] H. R. Künsch. Discrimination between monotonic trends and long-range dependence. *J. Applied Probability*, 23:1025–1030, 1986.

[37] H. R. Künsch. Statistical aspects of self-similar processes. In *Probability theory and applications, Proc. World Congr. Bernoulli Soc.*, 1987.

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

[39] I. N. Lobato. Consistency of the averaged cross-periodogram in long memory series. *J. of Time Series Analysis*, 18(2):137–155, 1997.

[40] I. N. Lobato and P. M. Robinson. Average periodogram estimation of long memory. *J. Econom*, 73(1):303–324, 1996.

[41] E. Moulines and Ph. Soulier. Log-periodogram regression of time series with long-range dependence. *Annals of Statistics*, 27(4), 1999.

[42] E. Moulines and Ph. Soulier. Data driven order selection for projection estimator of the spectral density of time series with long range dependence. J. of Time Series Analysis, 2000.

[43] V. A. Reisen. Estimation of the fractional difference parameter in the ARIMA(p, d, q) model using the smoothed periodogram. *J. of Time Series Analysis*, 15(3):335–350, 1994.

[44] P. M. Robinson. Semiparametric analysis of long-memory time series. *Annals of Statistics*, 22:515–539, 1994.

[45] P. M. Robinson. Time series with long range dependence. In *Advances in econometrics. Proceedings of the sixth world congress*, 1994.

[46] P. M. Robinson. Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, 24:1630–1661, 1995.

[47] P. M. Robinson. Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, 23:1043–1072, 1995.

[48] Ph. Soulier. Some new bounds and a central limit theorem for functions of Gaussian vectors. Prépublication de l'université d'Evry val d'Essonne, 1998.

[49] Ph. Soulier. Adaptive estimation of the spectral density. Prépublication de l'université d'Evry val d'Essonne, 1999.

[50] M. Taniguchi. Minimum constrast estimation for spectral densities of stationary processes. *J. R. Statist. Soc. B*, 49:315–325, 1987.

[51] M. S. Taqqu. Law of the iterated logarithm for sums of nonlinear functions of Gaussian variables that exhibit long range dependence. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 40:203–238, 1977.

[52] C. Velasco. Gaussian semiparametric estimation of non-stationary time series. *J. of Time Series Analysis*, 20(1):87–127, 1999.

[53] C. Velasco. Non-Gaussian log-periodogram regression. *Econometric Theory, forthcoming*, 1999.

[54] C. Velasco. Non-Stationary log-periodogram regression. *Journal of Econometrics*, 91(2):325–371, 1999.

[55] M. C. Viano, C. Deniau, and G. Oppenheim. Long-range dependence and mixing for discrete time fractional processes. *J. of Time Series Analysis*, 16:323–328, 1995.

[56] A. M. Walker. Some asymptotic results for the periodogram of a stationary time series. *J. Aust. Math. Soc.*, 5:107–128, 1965.

[57] Y. Yajima. Estimation of the frequency of unbounded spectral density. Discussion paper series, Faculty of Economics, University of Tokyo.