Consistent estimation of the memory parameter for nonlinear time series

Violetta Dalla, Liudas Giraitis and Javier Hidalgo
London School of Economics, University of York, London School of Economics

18 August, 2004

Abstract

For linear processes, semiparametric estimation of the memory parameter, based on the log-periodogram and local Whittle estimators, has been exhaustively examined and their properties are well established. However, except for some specific cases, little is known about the estimation of the memory parameter for nonlinear processes. The purpose of this paper is to provide general conditions under which the local Whittle estimator of the memory parameter of a stationary process is consistent and to examine its rate of convergence. We show that these conditions are satisfied for linear processes and a wide class of nonlinear models, among others, signal plus noise processes, nonlinear transforms of a Gaussian process ξ_t and EGARCH models. Special cases where the estimator satisfies the central limit theorem are discussed. The finite sample performance of the estimator is investigated in a small Monte-Carlo study.

JEL classification. C14, C22.
Short title: Consistent estimation of the memory parameter
Keywords and phrases: Long memory, semiparametric estimation, local Whittle estimator.

Corresponding Author: L. Giraitis, Department of Mathematics, University of York, Heslington, York YO10 5DD, U.K.; lg8@york.ac.uk

1 Introduction

Consider a weakly stationary process $(X_t)_{t\in Z}$ (abbreviated as (X_t) in what follows), which is observed at times $t = 1, 2, ..., n$, with an unknown mean μ, variance σ^2 and spectral density $f(\lambda)$, such that

$$f(\lambda) = |\lambda|^{-\alpha_0} g(\lambda), \quad |\lambda| \le \pi, \tag{1.1}$$

where

$$g(\lambda) \to b_0, \quad \text{as } |\lambda| \to 0,$$

$|\alpha_0| < 1$ and $0 < b_0 < \infty$. When $\alpha_0 = 0$, we say that (X_t) has short memory. If $0 < \alpha_0 < 1$, we say that the process has long memory, whereas when $-1 < \alpha_0 < 0$, it is said that the process is antipersistent.

When the spectral density $f(\lambda)$ in (1.1) is correctly specified by a finite dimensional parameter, say $g(\lambda) \equiv g(\lambda; \theta_0)$, then under some additional regularity assumptions, the parameters α_0 and θ_0 can be consistently estimated by the parametric Whittle estimator. Hannan (1973) proved consistency of this estimator for a wide class of short memory linear and nonlinear time series (X_t). In the case of Gaussian and linear processes, the Whittle estimator is known to be $n^{1/2}$-consistent and asymptotically normal. For $\alpha_0 \ge 0$, this was shown by Fox and Taqqu (1986), Dahlhaus (1989) and Giraitis and Surgailis (1990). The case when α_0 can be negative has been recently examined by Velasco and Robinson (2000).

Semiparametric estimation of the memory parameter α_0 requires less *a priory* known information about the spectral density $f(\lambda)$. Besides (1.1), it imposes no additional parametric specification or restrictions on $f(\lambda)$ (or $g(\lambda)$) outside the frequency $\lambda = 0$. A number of semiparametric estimators of α_0 has been developed for Gaussian and linear processes. Among others, we can mention the well-known log-periodogram and local Whittle estimators introduced by Geweke and Porter-Hudak (1983) and Künsch (1987), respectively, and explored by Robinson (1995a, b). See also Moulines and Soulier (1999) and Hurvich and Brodsky's (2001) broad-band estimators and exact local Whittle estimation method by Shimotsu and Phillips (2005), the latter being also valid for non-stationary time series. For a recent review on semiparametric estimators of the memory parameter and their statistical properties see Moulines and Soulier (2003).

As a rule, semiparametric estimators have a slower rate of convergence than parametric ones and are pivotal, i.e. their asymptotic distribution does not depend on unknown parameters. For example, if (X_t) is a fourth order stationary linear sequence with spectral density (1.1) such that $g(\lambda) = b_0 + O(\lambda^2)$, as $\lambda \to 0$, then Robinson's (1995b) results imply that, the local Whittle estimator $\widehat{\alpha}$, defined by (2.1) below, has an asymptotic standard normal distribution:

$$\sqrt{m}(\widehat{\alpha} - \alpha_0) \xrightarrow{d} N(0, 1), \quad \text{as } n \to \infty, \tag{1.2}$$

where $m = o(n^{4/5} \log^{-2} n)$. The existing semiparametric estimation theory is based on the assumption that (X_t) is a linear process. However, in some empirical applications, e.g. financial econometrics, nonlinear models are rather common, and hence the practitioner faces the question of to which extend results like (1.2) are still valid. For nonlinear models some results for the log-periodogram estimator for stochastic volatility models were obtained by Deo and Hurvich (2001), Sun and Phillips (2003) and for the local Whittle estimator by Hurvich, Moulines and Soulier (2005) and Arteche (2004).

The main purpose of this paper is to derive, for a wide class of time series models, general and easy to check conditions, under which the local Whittle estimator is a consistent estimator of the memory parameter α_0 and to investigate its rate of convergence. In particular, we show that our results are valid for nonlinear transformations $G(\xi_t)$ of a Gaussian process (ξ_t), the EGARCH process and for a signal plus noise type process $X_t = Y_t + Z_t$ when the memory parameter of the noise (Z_t) is smaller than the memory parameter of the signal (Y_t). The latter model extends the so-called stochastic volatility models, see Deo and Hurvich (2001) and Hurvich, Moulines and Soulier (2005).

Furthermore, from our results we can draw two main conclusions. Firstly, for nonlinear time series the rate of convergence of the local Whittle estimator $\widehat{\alpha}$ to α_0 is typically slower than in linear or Gaussian models, and hence to achieve the same level of accuracy, a larger sample is required. Secondly, the central limit theorem (1.2) with $m = o(n^{4/5})$ might no longer hold. Thus, estimation and testing procedures designed for linear processes that are based on (1.2) might not be appropriate for nonlinear ones.

The remainder of the paper is as follows. Section 2 presents the main results of the paper, whereas in Sections 3 and 4 we discuss various applications and examples. Sections 5 and 6 contain the proofs. Finally, a Monte-Carlo study in Section 7 examines the finite sample performance of the estimator.

2 Consistency of the local Whittle estimator

2.1 Consistent estimation

To estimate α_0 we shall use the local Whittle estimator $\widehat{\alpha}$, see Künsch (1987) and Robinson (1995b), defined as the minimizer

$$\widehat{\alpha} \equiv \widehat{\alpha}_n = \mathrm{argmin}_{[-1,1]} U_n(\alpha), \tag{2.1}$$

of the local objective function

$$\begin{aligned} U_n(\alpha) &= \log\left(\frac{1}{m}\sum_{j=1}^m \lambda_j^\alpha I_n(\lambda_j)\right) - \frac{\alpha}{m}\sum_{j=1}^m \log\lambda_j \\ &= \log\left(\frac{1}{m}\sum_{j=1}^m j^\alpha I_n(\lambda_j)\right) - \frac{\alpha}{m}\sum_{j=1}^m \log j. \end{aligned}$$

Here $\lambda_j = 2\pi j/n$, $j = 1, \ldots, m$, are the Fourier frequencies,

$$I_n(\lambda_j) = (2\pi n)^{-1}\left|\sum_{t=1}^n X_t e^{it\lambda_j}\right|^2$$

is the periodogram of the variables X_t, $t = 1, .., n$ and $m = m_n$ is an integer bandwidth parameter such that

$$m \to \infty, \quad m = o(n), \quad \text{as } n \to \infty.$$

Note that in semiparametric models, the spectral density function has property (1.1) and is only locally "parameterized" around $\lambda = 0$ by the parameters α_0 and b_0. Therefore, contrary to the parametric Whittle estimation, which employs the full spectrum of frequencies, the local Whittle estimator uses only the first m Fourier frequencies.

The main aim of this subsection is to derive a semiparametric analog to Hannan's (1973) result, who showed that if the process (X_t) is ergodic and has parametrically specified spectral density $f(\lambda)$, then, under mild assumptions on $f(\lambda)$, the unknown parameters can be consistently estimated by the parametric Whittle estimator.

As in the case of the parametric Whittle estimator, the local Whittle estimator is based on the whitening principle of the normalized periodogram at the Fourier frequencies. Roughly speaking, it means that the sequence

$$\eta_j = \frac{I_n(\lambda_j)}{f(\lambda_j)}, \quad 1 \le j \le m,$$

behaves as if η_j were independent and identically distributed (i.i.d.) random variables with unit mean, which is well-known when (X_t) is an i.i.d. Gaussian sequence. However, when (X_t) is a non-Gaussian or a sequence of dependent random variables, then the η_j's are neither independent nor uncorrelated random variables. In spite of that, under assumption (1.1), Lemma 6.3 below implies that

$$E\eta_j = 1 + o(1), \quad \text{as } j \to \infty,\, n \to \infty, \tag{2.2}$$

uniformly in $1 \le j \le m$, and under some additional regularity assumptions,

$$\mathrm{Cov}(\eta_j, \eta_k) \to 0 \tag{2.3}$$

when $j \ne k$ and $j, k \to \infty$. See Lahiri (2003), for sufficient and necessary conditions for the validity of (2.3).

One of the main consequences of (2.2) and (2.3) is that (η_j) satisfies a weak law of large numbers (WLLN):

$$m^{-1} \sum_{j=1}^{m} \eta_j \xrightarrow{P} 1, \quad \text{as } n \to \infty. \tag{2.4}$$

Note that (2.4) indicates that (η_j) behaves as an "ergodic" sequence with mean 1.

Next, setting

$$\eta_j^* = \frac{I_n(\lambda_j)}{b_0 \lambda_j^{-\alpha_0}}, \quad 1 \le j \le m, \tag{2.5}$$

it follows from (1.1) that

$$\eta_j^* = \eta_j b_0^{-1} g(\lambda_j) = \eta_j (1 + o(1)), \quad \text{as } n \to \infty,$$

uniformly in $1 \le j \le m$. Therefore, (2.4) is equivalent to the WLLN property of (η_j^*):

$$m^{-1} \sum_{j=1}^{m} \eta_j^* \xrightarrow{P} 1, \quad \text{as } n \to \infty. \tag{2.6}$$

Indeed, by (2.2),

$$m^{-1}E\Big|\sum_{j=1}^{m}(\eta_j-\eta_j^*)\Big| = o(m^{-1})\sum_{j=1}^{m}E\eta_j = o(1),$$

so that convergence (2.4) implies (2.6). On the other hand, it is also clear that (2.6) implies (2.4). In addition, (2.2) shows that

$$E\eta_j^* \le C, \tag{2.7}$$

uniformly in $1 \le j \le m$, where C is a finite constant. Note that the variables η_j and η_j^* are invariant with respect to the mean of (X_t) since the periodogram $I_n(\lambda_j)$ is self-centering at the Fourier frequencies λ_j, $1 \le j \le n-1$.

To derive the consistency of the estimator $\widehat{\alpha}$ we introduce the following assumptions:

Assumption A. (X_t) is a weakly stationary sequence with spectral density $f(\lambda)$ satisfying (1.1).

Assumption B. The renormalised periodograms η_j^*, $1 \le j \le m$ satisfy the WLLN property (2.6), for any sequence $m = m_n \to \infty$ such that $m = o(n)$.

The following theorem shows that if $\eta_j^*, 1 \le j \le m$, satisfies Assumption B then $\widehat{\alpha}$ is a consistent estimate of α_0.

THEOREM 2.1 *Suppose that (X_t) satisfies Assumptions A and B. Then, as $n \to \infty$,*

$$\widehat{\alpha} \xrightarrow{P} \alpha_0. \tag{2.8}$$

Moreover,

$$\widehat{\alpha} - \alpha_0 = -Q_m(1 + o_P(1)) + O_P(m^{-1}\log m), \tag{2.9}$$

where

$$Q_m = m^{-1}\sum_{j=1}^{m}(\log(j/m)+1)b_0^{-1}\lambda_j^{\alpha_0}I_n(\lambda_j). \tag{2.10}$$

The proof of Theorem 2.1 is based on (2.6) and (2.7). In fact, a closer look at the proof shows that for the consistency of $\widehat{\alpha}$ the requirement of stationarity of (X_t) is not needed, as the following proposition indicates.

PROPOSITION 2.1 *Assume that $X_1, X_2, ..., X_n$, $n \ge 1$, is a sequence of random variables. If there exist $\alpha_0 \in (-1,1)$ and $b_0 > 0$ such that $\eta_j^*, 1 \le j \le m$ satisfy assumptions (2.6) and (2.7), then Theorem 2.1 holds.*

Proof of Theorem 2.1 is based only on properties (2.6) and (2.7) of (η_j^*), and therefore the proof of Proposition 2.1 is a standard extension of that of Theorem 2.1.

It is worth observing that Proposition 2.1 indicates that for consistency (2.8) only the asymptotic stationarity of (X_t) is required. The classical example of non-stationary process (X_t) satisfying (2.6) and (2.7) is a model generated by

$$X_t = (1-L)^{-\alpha_0/2}\, u_t \mathcal{I}\,(t>0)\,,$$

where (u_t) is a weakly dependent process.

The following theorem provides an expansion for $\widehat{\alpha} - \alpha_0$ which is helpful for analysing the rate of convergence and deriving the asymptotic distribution of the estimator $\widehat{\alpha}$. We first introduce

Assumption $\mathrm{T}(\alpha_0, \beta)$. There exist $\alpha_0 \in (-1, 1)$, $\beta \in (0, 2]$, finite $b_0 > 0$ and $b_1 \neq 0$ such that the spectral density f has property

$$f(\lambda) = |\lambda|^{-\alpha_0}(b_0 + b_1|\lambda|^{\beta} + o(|\lambda|^{\beta})), \quad \text{as } \lambda \to 0. \tag{2.11}$$

The parameter β characterizes the smoothness of the function $g(\lambda)$ in (1.1). For example, for Autoregressive Fractionally Integrated Moving Average (ARFIMA$(p, \alpha_0/2, q)$) models, (2.11) holds with $\beta = 2$.

Theorem 2.2 *Suppose that (X_t) satisfies Assumption B. Then under Assumption $T(\alpha_0, \beta)$, we have that*

$$\widehat{\alpha} - \alpha_0 = -(m/n)^{\beta}(b_1/b_0)B_{\beta} - (Q_m - EQ_m)(1 + o_P(1)) \tag{2.12}$$

$$+ o_P(m^{-1/2} + (m/n)^{\beta}),$$

where $B_{\beta} = (2\pi)^{\beta}\beta/(\beta + 1)^2$.

Next, we present a simple sufficient condition which implies Assumption B. Denote

$$\Delta_m = \max_{1 \le k \le m} E\Big|\sum_{j=1}^{k}(\eta_j^* - E\eta_j^*)\Big|.$$

First note that (2.7) implies $\Delta_m \le Cm$, where C denotes a generic constant in what follows. The next proposition shows that $\Delta_m = o(m)$ implies Assumption B, which together with (1.1), as Theorem 2.1 indicates, is a sufficient condition for the consistency of the estimator $\widehat{\alpha}$.

Proposition 2.2 *Suppose that (X_t) satisfies Assumption A and that $\Delta_m = o(m)$. Then (X_t) satisfies Assumption B, and*

$$\widehat{\alpha} \xrightarrow{P} \alpha_0, \quad \textit{as } n \to \infty.$$

2.2 Consistency rate

In this section we examine the rate of convergence of the estimator $\widehat{\alpha}$.

Proposition 2.3 *Suppose that the spectral density function of (X_t) satisfies Assumption* $\mathrm{T}(\alpha_0, \beta)$ *and $\Delta_m = o(m/\log^2 n)$. Then*

$$\widehat{\alpha} - \alpha_0 = O_P\Big(\Delta_m m^{-1}\log m + m^{-1/2} + (m/n)^{\beta}\Big), \tag{2.13}$$

and $Q_m - EQ_m$ in (2.12) can be written as

$$Q_m - EQ_m = V_m + o_P((m/n)^\beta), \tag{2.14}$$

where

$$V_m = m^{-1} \sum_{j=1}^{m} (\log(j/m) + 1)(\eta_j - E\eta_j).$$

The last proposition shows that the rate of convergence of $\widehat{\alpha}$ is determined by the stochastic order of magnitude of $Q_m - EQ_m$ which can be controlled by the order of magnitude of Δ_m.

Our next step is to find simple sufficient conditions in terms of (X_t), which imply that $\Delta_m = o(m)$. To that end, let (X_t) be a 4-th order stationary sequence. Denote the 4-th order cumulant of the variables $X_{t_1}, X_{t_2}, X_{t_3}, X_{t_4}$ by $c_X(t_1, ..., t_4) := Cum(X_{t_1}, X_{t_2}, X_{t_3}, X_{t_4})$, defined by

$$c_X(t_1, ..., t_4) = E[X_{t_1} X_{t_2} X_{t_3} X_{t_4}] - E[X_{t_1} X_{t_2}] E[X_{t_3} X_{t_4}] - E[X_{t_1} X_{t_3}] E[X_{t_2} X_{t_4}]$$
$$-E[X_{t_1} X_{t_4}] E[X_{t_2} X_{t_3}].$$

Recall that without loss of generality we can assume that $EX_t = 0$, and, by 4-th order stationarity, $c_X(t_1, ..., t_4) = c_X(t_1 - t_4, t_2 - t_4, t_3 - t_4, 0)$.

Denote

$$D_n^* = \sum_{t_1, t_2, t_3 = -n}^{n} |c_X(t_1, t_2, t_3, 0)|,$$

$$D_n^{**} = \max_{|t_1|, |t_2| \le n} \sum_{u=-n}^{n} |c_X(t_1, t_2 + u, u, 0)|.$$

Note that for a wide class of 4-th order stationary short memory sequences (X_t), the 4-th order cumulant satisfies the condition

$$D_\infty^* = \sum_{t_1, t_2, t_3 = -\infty}^{\infty} |c_X(t_1, t_2, t_3, 0)| < \infty. \tag{2.15}$$

For example, (2.15) holds for stationary invertible $ARMA(p, q)$ models. Observe also that $EX_t^4 \le C$ implies that

$$D_n^{**} \le Cn.$$

We shall use D_n^* and D_n^{**} to estimate Δ_m and in particular to derive conditions which imply $\Delta_m = o(m)$.

LEMMA 2.1 *Suppose that (X_t) is a 4-th order stationary sequence whose spectral density function satisfies (1.1) with $\alpha_0 \in (-1, 1)$. Then, as $n \to \infty$,*

$$m^{-1} \Delta_m = O\Big(m^{-1/2} + \Big(\frac{D_n^*}{n}\Big)^{1/2} \Big(\frac{m}{n}\Big)^{\alpha_0}\Big). \tag{2.16}$$

Moreover, if $0 \le \alpha_0 < 1$, then

$$m^{-1} \Delta_m = O\Big(m^{-1/2} + \Big(\frac{D_n^{**}}{n}\Big)^{1/2} \Big(\frac{n}{m}\Big)^{1-\alpha_0} \log n\Big). \tag{2.17}$$

Combining Lemma 2.1 and Proposition 2.3 we obtain the following corollary.

COROLLARY 2.1 *Suppose that (X_t) is a 4-th order stationary sequence whose spectral density f satisfies Assumption* $\mathrm{T}(\alpha_0, \beta)$. *Then, as $n \to \infty$,*

$$\widehat{\alpha} - \alpha_0 = O_P\Big(m^{-1/2}\log m + (m/n)^{\beta} + r_n\Big),$$

where

(i) $r_n = 0$, *if* $\alpha_0 = 0$ *and* $D^*_\infty < \infty$,

(ii) $r_n = n^{1/2}m^{-1}\log m$, *if* $\alpha_0 \in (-1, 0)$ *and* $D^*_\infty < \infty$,

(iii) $r_n = \left(\frac{D^*_n}{n}\right)^{1/2}\left(\frac{m}{n}\right)^{\alpha_0}\log^2 n$, *if* $\alpha_0 \in (0, 1)$,

(iv) $r_n = \left(\frac{D^{**}_n}{n}\right)^{1/2}\left(\frac{n}{m}\right)^{1-\alpha_0}\log^3 n$, *if* $\alpha_0 \in (0, 1)$,

assuming that in (ii)-(iv), $m = m_n$, D^*_n and D^{**}_n are such that

$$r_n \to 0, \quad \text{as } n \to \infty.$$

REMARK 2.1 In Corollary 2.1, $r_n \to 0$ in case (ii) if m is such that $n^{1/2}m^{-1}\log m = o(1)$, and in case (iii) if $D^*_n = o(n/\log^4 n)$. In case (iv), $r_n \to 0$ holds if $D^{**}_n = O(n^{\gamma}/\log^6 n)$ for some $0 \le \gamma < 1$ and m is such that $n^{(1+\gamma)/2}m^{-1} = O(1)$.

If (X_t) has short memory, i.e. $\alpha_0 = 0$ and $D^*_\infty < \infty$, then (i) shows that $\widehat{\alpha} = O_P\Big(m^{-1/2}\log m + (m/n)^{\beta}\Big)$, whereas (iv) indicates, that in long memory case $\alpha_0 \in (0, 1)$, $r_n \to 0$ and the estimator $\widehat{\alpha}$ is consistent provided that $D^{**}_n = O(n^{\gamma})$ $(0 \le \gamma < 1)$ and m is chosen large enough.

Proofs of Theorems 2.1, 2.2, Propositions 2.2, 2.3 and Lemma 2.1 are given in Section 5. We finish the section showing how our general results hold for linear processes.

2.3 Linear process: an example

Semiparametric estimation of the memory parameter of a linear sequence (X_t) has been well investigated, see Robinson (1995a, b). It is nevertheless of interest to show that our general results in the previous subsection hold true for a linear sequence. It is said that (X_t) is a *linear sequence* if

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} a_j^2 < \infty, \tag{2.18}$$

where the ε_j are i.i.d. random variables with zero mean and unit variance.

Under Assumption A, it follows from Robinson (1995b, Theorem 1) that $\widehat{\alpha} \xrightarrow{P} \alpha_0$. We now show that under these conditions, the consistency of $\widehat{\alpha}$ is a consequence of our Theorem 2.1.

PROPOSITION 2.4 *Suppose that a linear sequence (X_t), given by (2.18), satisfies Assumption A, and $m = o(n)$. Then (X_t) satisfies Assumption B, so that*

$$\widehat{\alpha} \xrightarrow{P} \alpha_0, \quad \text{as } n \to \infty.$$

Proof of Proposition 2.4. To show that (X_t) satisfies Assumption B, it suffices to examine (2.4), since under Assumption A, (2.4) implies (2.6), see Subsection 2.1. Write

$$\sum_{j=1}^{k}\eta_j = \sum_{j=1}^{k}2\pi I_\epsilon(\lambda_j)+\sum_{j=1}^{k}(\eta_j-2\pi I_\epsilon(\lambda_j)) =: p_{n,1}(k)+p_{n,2}(k), \quad 1\le k\le m, \tag{2.19}$$

where

$$I_\varepsilon(\lambda_j)=(2\pi n)^{-1}\Big|\sum_{t=1}^{n}\varepsilon_t e^{it\lambda_j}\Big|^2.$$

Then (2.4) follows if, as $k\to\infty$,

$$k^{-1}p_{n,1}(k)\xrightarrow{P}1, \quad k^{-1}p_{n,2}(k)\xrightarrow{P}0. \tag{2.20}$$

Under Assumption A, Robinson (1995b, Relation (3.17)) derived the bound

$$E|\eta_j-2\pi I_\epsilon(\lambda_j)|\le C(j^{-1}\log j)^{1/2},$$

which holds uniformly in $1\le j\le m$. Thus

$$E|k^{-1}p_{n,2}(k)|\le Ck^{-1/2}(\log k)^{1/2}\to 0, \tag{2.21}$$

which implies that $k^{-1}p_{n,2}(k)\xrightarrow{P}0$ by Markov inequality. Similarly, proceeding as in Robinson (1995b, pp. 1637-1638), it can be shown that $k^{-1}p_{n,1}(k)\xrightarrow{P}1$, which completes the proof of (2.20). Since (X_t) satisfies Assumptions A and B, Theorem 2.1 implies that $\widehat{\alpha}\xrightarrow{P}\alpha_0$. □

Next proposition provides bounds for D_n^{**} and Δ_m.

Proposition 2.5 *Let (X_t) be a linear sequence (2.18) and $E\varepsilon_0^4<\infty$. Then*

$$D_n^{**}\le C, \quad n\ge 1. \tag{2.22}$$

In addition, if (X_t) satisfies Assumption $\mathrm{T}(\alpha_0,\beta)$ *and (2.25) below holds, then*

$$m^{-1}\Delta_m = O\Big(m^{-1/2}\log^{1/2}m+(m/n)^{\beta}\Big), \quad \textit{as } n\to\infty. \tag{2.23}$$

Proof of Proposition 2.5. First we show that $D_n^{**}\le C$. To that end, set $a_v=0$ for $v<0$. Then, using the equality

$$c_X(0,t_1,t_2,t_3)=(E\varepsilon_0^4-3)\sum_{v=-\infty}^{\infty}a_v a_{v+t_1}a_{v+t_2}a_{v+t_3},$$

we conclude that

$$\begin{aligned}|D_n^{**}| &\le C\max_{|t_1|,|t_2|\le n}\sum_{u=-n}^{n}\sum_{v=-\infty}^{\infty}|a_v a_{v+t_1}a_{v+t_2+u}a_{v+u}| \\ &\le C\max_{|t_1|\le n}\sum_{v=-\infty}^{\infty}|a_v a_{v+t_1}|\sum_{u=-\infty}^{\infty}a_u^2\le C\Big(\sum_{u=-\infty}^{\infty}a_u^2\Big)^2<\infty,\end{aligned}$$

which proves (2.22).

It remains to show (2.23). Using (2.19) we can write

$$p_n(k) \equiv \sum_{j=1}^{k} \eta_j^* = p_{n,1}(k) + p_{n,2}(k) + R_n(k), \quad 1 \le k \le m,$$

where $R_n(k) = \sum_{j=1}^{k}(\eta_j^* - \eta_j)$. Under Assumption A,

$$E|\eta_j^* - \eta_j| \le |1 - b_0^{-1} g(\lambda_j)| E\eta_j \le C|1 - b_0^{-1} g(\lambda_j)| \le C(m/n)^{\beta} \tag{2.24}$$

which implies that

$$E|R_n(k) - ER_n(k)| \le 2E|R_n(k)| \le Cm(m/n)^{\beta}.$$

Next, proceeding as in the proof of Robinson's (1995b) Theorem 2, it is easily seen that, uniformly in $1 \le k \le m$,

$$E|p_{n,1}(k) - Ep_{n,1}(k)| \le Cm^{1/2},$$

whereas (2.21) implies that

$$E|p_{n,2}(k) - Ep_{n,2}(k)| \le 2E|p_{n,2}(k)| \le Cm^{1/2} \log^{1/2} m.$$

The last three estimates imply that

$$\Delta_m \le \max_{1 \le k \le m} E|p_n(k) - Ep_n(k)| \le C(m^{1/2} \log^{1/2} m + m(m/n)^{\beta}),$$

to prove (2.23). $\square$

Robinson (1995b) showed that if a linear sequence (X_t) satisfies Assumption T(α_0, β) with $0 < \beta \le 2$, $E\varepsilon_0^4 < \infty$ and

$$(d/d\lambda)\alpha(\lambda) = O(|\alpha(\lambda)|/\lambda), \quad \text{as } \lambda \to 0+, \tag{2.25}$$

where $\alpha(\lambda) = \sum_{j=0}^{\infty} a_j e^{ij\lambda}$, then

$$m^{1/2}(\widehat{\alpha} - \alpha_0) \xrightarrow{d} N(0,1), \quad \text{as } n \to \infty, \tag{2.26}$$

for $m \to \infty$ such that $m = o(n^{2\beta/(1+2\beta)} \log^{-2} n)$.

Proposition 2.5 implies that if $m = o(n/\log^{2/\beta} n)$ then $\Delta_m = o(m/\log^2 m)$, and therefore our Theorem 2.2 together with (2.14) yield the expansion

$$\widehat{\alpha} - \alpha_0 = -V_m - (m/n)^{\beta}(b_1/b_0)B_{\beta} + o_P(m^{-1/2} + (m/n)^{\beta}), \tag{2.27}$$

which is valid when $m = o(n/\log^{2/\beta} n)$. Moreover, $EV_m = 0$ and by Robinson's (1995b) Theorem 2,

$$m^{1/2} V_m \xrightarrow{d} N(0,1), \quad \text{as } n \to \infty. \tag{2.28}$$

Relations (2.27) and (2.28) imply the convergence (2.26) if $m = o(n^{2\beta/(1+2\beta)})$.

On the other hand, if $n^{2\beta/(2\beta+1)}/m \to 0$ and $m = o(n/\log^{2/\beta} n)$, they yield convergence to a deterministic limit,

$$(n/m)^{\beta}(\widehat{\alpha} - \alpha_0) \xrightarrow{P} -(b_1/b_0)B_{\beta}, \tag{2.29}$$

whereas, if $m = n^{2\beta/(2\beta+1)}$, we then have

$$m^{1/2}(\widehat{\alpha} - \alpha_0) \xrightarrow{d} N\Big(-(b_1/b_0)B_{\beta}, 1\Big).$$

3 Signal plus noise process

In this section we discuss estimation of the memory parameter of a stationary process when it is observed with noise. More specifically, let $X_t = Y_t + Z_t$, $t \in Z$ where (Y_t) denotes the signal and (Z_t) stands for the noise. This type of models (with an i.i.d. noise) has drawn much attention as they arise after taking the logarithmic transformation of the stochastic volatility model, introduced by Taylor (1994) and explored by Harvey, Ruiz and Shephard (1994). We shall show that the local Whittle estimator of the memory parameter of the signal, denoted by α_Y, remains consistent in the presence of a noise whose memory parameter is smaller than that of the signal. However, the noise can significantly increase the finite-sample bias of the estimator of α_Y, so that larger samples are required to achieve the same precision as in estimation without the noise (Z_t). In case of Gaussian or linear signal similar observations were made by Hurvich, Moulines and Soulier (2005), Arteche (2004), and by Deo and Hurvich (2001) and Sun and Phillips (2003) for the log-periodogram estimator, assuming that the signal and the noise are independent processes.

Our approach does not assume that the signal is a Gaussian or linear process and the noise is an i.i.d. sequence as well as it does not require that the signal is independent of the noise, which are common assumptions in the literature.

3.1 The sum of a Gaussian sequence and an i.i.d. Gaussian noise.

We first start looking at the rather simple model of a Gaussian signal (Y_t) and i.i.d. noise which corresponds to the model examined by Deo and Hurvich (2001). This rather simple example illustrates that adding a Gaussian i.i.d. noise to a Gaussian stationary sequence (Y_t) can significantly increase the bias of the local Whittle estimate.

Consider the sequence (X_t),

$$X_t = Y_t + Z_t, \tag{3.1}$$

where (Y_t) is a Gaussian sequence satisfying Assumption $\mathrm{T}(\alpha_Y, 2)$ and (Z_t) is a sequence of i.i.d. Gaussian random variables with zero mean and unit variance uncorrelated with (Y_t). Denote by $f_X(\lambda)$, $f_Y(\lambda)$ and $f_Z(\lambda)$, the spectral density functions of the sequences (X_t), (Y_t) and (Z_t), respectively.

Assume that $\alpha_Y > 0$. Then, because $f_Y(\lambda) = |\lambda|^{-\alpha_Y}(b_0 + b_1|\lambda|^2 + o(|\lambda|^2))$, $\alpha_Y \in (-1, 1)$ and $f_Z(\lambda) = 1/(2\pi)$, we obtain that

$$f_X(\lambda) = f_Y(\lambda) + f_Z(\lambda) = |\lambda|^{-\alpha_Y}(b_0 + b_1'|\lambda|^{\alpha_Y} + O(|\lambda|^2)), \quad \text{as } \lambda \to 0,$$

where $b_1' = 1/2\pi$. Thus (X_t) satisfies Assumption $\mathrm{T}(\alpha_X, \beta)$ with memory parameter $\alpha_X = \alpha_Y$ and smoothness parameter $\beta = \alpha_Y$. Moreover, the Gaussian sequence (X_t) with spectral density f_X can be written as a linear sequence (2.18), due to a well-known result by Cramér. Results (2.28) and (2.29), obtained for a linear process, imply that the central limit theorem

$$m^{1/2}(\widehat{\alpha}_X - \alpha_X) \xrightarrow{d} N(0,1) \tag{3.2}$$

holds when $m = o(n^{2\alpha_Y/(1+2\alpha_Y)})$ which requires m to be small when α_Y is close to 0, leading to wider confidence intervals, whereas if $n^{2\alpha_Y/(1+2\alpha_Y)}/m \to 0$ and $m = o(n/\log^{2/\alpha_Y} n)$, then the bias terms dominates and

$$(n/m)^{\alpha_Y}\,(\widehat{\alpha}_X - \alpha_X) \xrightarrow{P} -(b_1'/b_0)B_{\alpha_Y}.$$

On the other hand, if $\alpha_Y < 0$, then (Z_t) is the signal and we can write

$$f_X(\lambda) = b_1' + b_0 \, |\lambda|^{-\alpha_Y} + o(|\lambda|^{-\alpha_Y}),$$

which shows that (X_t) satisfies Assumption $\mathrm{T}(\alpha_X, \beta)$ with $\alpha_X = 0$ and $\beta = -\alpha_Y$, and expansion (2.27) implies that $\widehat{\alpha}_X - \alpha_X = O_P\left(m^{-1/2} + (m/n)^{-\alpha_Y}\right)$. The previous relations show that in case of $\alpha_Y \neq 0$,

$$\widehat{\alpha}_X - \alpha_X = O_P\left(m^{-1/2} + (m/n)^{|\alpha_Y|}\right).$$

Therefore, for m such that $n^\epsilon \leq m \leq n^{1-\epsilon}$ for some $0 < \epsilon < 1$, there exists $\epsilon' > 0$ such that $\widehat{\alpha}_X - \alpha_X = O_P(n^{-\epsilon'})$.

3.2 Estimating the memory parameter of a signal plus noise process

In this subsection we extend the results of the previous subsection to a more general situation when the signal may be correlated with the noise which can be a stationary short or long memory sequence.

More specifically, the following theorem shows that the local Whittle estimate is $n^{-\epsilon}$ consistent for some $\epsilon > 0$ for a wide class of signal plus noise processes.

Theorem 3.1 *Suppose that*

$$X_t = Y_t + Z_t, \quad t \in Z,$$

where (Y_t) and (Z_t) are covariance stationary processes. Assume that (Y_t) satisfies Assumptions A and B with parameters $b_0 = c_Y$ and $\alpha_0 = \alpha_Y$, and the spectral densities f_Y and f_Z of (Y_t) and (Z_t) satisfy

$$f_Y(\lambda) = c_Y |\lambda|^{-\alpha_Y} + o(|\lambda|^{-\alpha_Y}), \quad f_Z(\lambda) \leq c_Z |\lambda|^{-\alpha_Z}, \quad \text{as } \lambda \to 0,$$

with $-1 < \alpha_Z < \alpha_Y < 1$.

Then, as $n \to \infty$,

(i)

$$\widehat{\alpha}_X \xrightarrow{P} \alpha_Y. \tag{3.3}$$

Moreover,

$$\widehat{\alpha}_X - \alpha_Y = (\widehat{\alpha}_Y - \alpha_Y)(1 + o_P(1)) + O_P\Big((m/n)^{(\alpha_Y - \alpha_Z)/2} + m^{-1}\log m\Big) \tag{3.4}$$

where $\widehat{\alpha}_Y$ denotes the local Whittle estimator of (Y_t) if the sequence (Y_t) were observed.

(ii) If (Y_t) satisfies Assumption $\mathrm{T}(\alpha_Y, \beta)$ and $\Delta_m \leq C m^\gamma$ for some $0 < \gamma < 1$, then

$$\widehat{\alpha}_X - \alpha_Y = O_P\Big(m^{\gamma-1}\log m + m^{-1/2} + (m/n)^\beta + (m/n)^{(\alpha_Y - \alpha_Z)/2}\Big). \tag{3.5}$$

(iii) If (Y_t) is a linear process satisfying Assumption $\mathrm{T}(\alpha_Y, 2)$ and $m = o(n/\log n)$, then

$$\widehat{\alpha}_X - \alpha_Y = O_P(m^{-1/2} + (m/n)^{(\alpha_Y - \alpha_Z)/2}). \tag{3.6}$$

In addition,

$$m^{1/2}(\widehat{\alpha}_X - \alpha_Y) \xrightarrow{d} N(0,1) \tag{3.7}$$

if (2.25) holds and $m = o(n^{2r/(2r+1)})$, where $r = (\alpha_Y - \alpha_Z)/2$.

Remark 3.1 Theorem 3.1 shows that if (X_t) can be decomposed into a signal plus noise process $X_t = Y_t + Z_t$ where the signal (Y_t) satisfies Assumption B and has larger memory parameter than the noise (Z_t), then under unrestricted assumptions on the noise (Z_t), $\widehat{\alpha}_X$ is a consistent estimator of the memory parameter α_Y of the signal. Recall that Assumption B is satisfied by linear and Gaussian processes, see Proposition 2.4. Theorem 3.1 does not impose any restrictions on the dependence between the signal (Y_t) and the noise (Z_t). In particular, if (Y_t) and (Z_t) are uncorrelated, then the spectral density function of (X_t) can be written as $f_X = f_Y + f_Z$ which implies that the memory parameter α_X of (X_t) equals α_Y.

Proof of Theorem 3.1. (i) By Proposition 2.1, (3.3) is shown if (X_t) satisfies relations (2.6) and (2.7) with parameters $b_0 = c_Y$ and $\alpha_0 = \alpha_Y$. Denote $w_X(j) = (2\pi n)^{-1/2} \sum_{t=1}^n X_t e^{it\lambda_j}$, $I_X(\lambda_j) = |w_X(j)|^2$ and write

$$I_X(\lambda_j) = |w_Y(j) + w_Z(j)|^2 = I_Y(\lambda_j) + v_j, \tag{3.8}$$

where

$$|v_j| \leq I_Z(\lambda_j) + 2|w_Y(j)|\,|w_Z(j)|.$$

Then, we can write

$$m^{-1} \sum_{j=1}^m c_Y^{-1} \lambda_j^{\alpha_Y} I_X(\lambda_j) := S_{n,1} + S_{n,2},$$

where

$$S_{n,1} = m^{-1} \sum_{j=1}^m c_Y^{-1} \lambda_j^{\alpha_Y} I_Y(\lambda_j), \quad S_{n,2} = m^{-1} \sum_{j=1}^m c_Y^{-1} \lambda_j^{\alpha_Y} v_j.$$

Since (Y_t) satisfies Assumption B, then (2.6) implies that $S_{n,1} \xrightarrow{P} 1$. On the other hand, by Lemma 6.3 and the assumptions imposed on $f_Z(\lambda)$ and $f_Y(\lambda)$,

$$\begin{aligned} E|v_j| &\leq C\left(EI_Z(\lambda_j) + 2\left[EI_Y(\lambda_j)\, EI_Z(\lambda_j)\right]^{1/2}\right) \\ &\leq C\left(f_Z(\lambda_j) + f_Y^{1/2}(\lambda_j)\, f_Z^{1/2}(\lambda_j)\right) \leq C\lambda_j^{-(\alpha_Z+\alpha_Y)/2}, \end{aligned} \tag{3.9}$$

uniformly in $1 \leq j \leq m$, because $\alpha_Z < \alpha_Y$. Therefore

$$E|S_{n,2}| \leq Cm^{-1} \sum_{j=1}^m \lambda_j^{\alpha_Y} E|v_j| \leq C(m/n)^{(\alpha_Y - \alpha_Z)/2} \to 0,$$

as $n \to \infty$, because $\alpha_Z < \alpha_Y$ and $m = o(n)$. Thus, by Markov inequality, $S_{n,2} \xrightarrow{P} 0$ and hence $S_{n,1} + S_{n,2} \xrightarrow{P} 1$, which shows that (X_t) satisfies (2.6) of Assumption B.

Since (Y_t) satisfies Assumption A, relations (3.8), (3.9) and (2.7) yield that

$$\lambda_j^{\alpha_Y} EI_X(\lambda_j) \leq C$$

for all $1 \leq j \leq m$, and hence (X_t) satisfies condition (2.7). This completes the proof of (3.3).

Next we show (3.4). Since (X_t) satisfies the assumptions (2.6) and (2.7) with parameters c_Y and α_Y, then Proposition 2.1 holds true and (2.9) implies that

$$\widehat{\alpha}_X - \alpha_Y = -Q_m(1 + o_P(1)) + O_P(m^{-1} \log m), \tag{3.10}$$

where

$$\begin{aligned} Q_m &= m^{-1} \sum_{j=1}^{m} (\log(j/m) + 1) c_Y^{-1} \lambda_j^{\alpha_Y} I_X(\lambda_j) \\ &= m^{-1} \sum_{j=1}^{m} (\log(j/m) + 1) c_Y^{-1} \lambda_j^{\alpha_Y} I_Y(\lambda_j) + O_P((m/n)^{(\alpha_Y - \alpha_Z)/2}), \end{aligned}$$

in view of (3.8) and (3.9). Applying relation (2.9) of Theorem 2.1 to the first term of the displayed equality, it follows that

$$Q_m = -(\widehat{\alpha}_Y - \alpha_Y)(1 + o_P(1)) + O_P\Big((m/n)^{(\alpha_Y - \alpha_Z)/2} + m^{-1} \log m\Big),$$

where $\widehat{\alpha}_Y$ denotes the local Whittle estimator as if the sequence (Y_t) were observed. This and (3.10) prove (3.4).

(ii) Since $\Delta_m = O(m^\gamma)$ and (Y_t) satisfies Assumption $\mathrm{T}(\alpha_Y, \beta)$, then $\widehat{\alpha}_Y - \alpha_Y = O_P(m^{\gamma-1} \log m + m^{-1/2} + (m/n)^\beta)$ by (2.13) of Proposition 2.3 which together with (3.4) implies (3.5).

(iii) If (Y_t) is a linear sequence satisfying Assumption $T(\alpha_Y, 2)$, then in view of (2.27), it follows that $\widehat{\alpha}_Y - \alpha_Y = O_P(m^{-1/2} + (m/n)^2)$ when $m = o(n/\log n)$, which together with (3.4) yields (3.6), whereas (3.7) follows applying (3.2) in (3.4). □

4 Applications

In this section we discuss estimation of the memory parameter of nonlinear transformations of a stationary Gaussian sequence and of some stochastic volatility models. We show that the latter processes can be decomposed into a signal plus noise process, so that the results of Section 3 apply.

4.1 Nonlinear functions of a stationary Gaussian sequence

Suppose that

$$X_t = G(\xi_t), \quad t \in Z, \tag{4.1}$$

where (ξ_t) is a stationary Gaussian sequence with zero mean and variance 1, and $G : R \to R$ is a measurable function such that $EG(\xi_t)^2 < \infty$ and $EG(\xi_t) = 0$. Then, X_t can be written as the sum

$$X_t = \sum_{k=k_0}^{\infty} \frac{c_k}{k!} H_k(\xi_t), \tag{4.2}$$

where $H_k(\cdot)$ is the k-th Hermite polynomial and $c_k = E[G(\xi_t) H_k(\xi_t)]$, see Dobrushin and Major (1979) and Taqqu (1979). The minimal integer $k_0 \geq 1$ such that $c_{k_0} \neq 0$ is called the Hermite rank of G. We assume that the Gaussian sequence (ξ_t) has spectral density f_ξ and denote $r_\xi(t) = E[\xi_t \xi_0]$.

Using the well-known properties of Hermite polynomials,

$$E[H_k(\xi_t)H_k(\xi_s)] = k! r_\xi^k(t-s) \tag{4.3}$$

$$E[H_k(\xi_t)H_m(\xi_s)] = 0 \quad \text{if } k \neq m, \tag{4.4}$$

we have that

$$r_X(t) := Cov(X_t, X_0) = \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} r_\xi^k(t), \quad EX_0^2 = \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} < \infty. \tag{4.5}$$

Therefore, because $r_\xi(t) \to 0$ $(t \to \infty)$, we conclude that the covariance function of (X_t) satisfies

$$r_X(t) = r_\xi^{k_0}(t) \left(\frac{c_{k_0}^2}{k_0!} + o(1) \right), \quad \text{as } t \to \infty. \tag{4.6}$$

When $\sum_{t \in Z} |r_\xi(t)|^{k_0} < \infty$, the last displayed equality implies that $\sum_{t \in Z} |r_X(t)| < \infty$, so that (X_t) behaves as a short memory process, whereas if $\sum_{t \in Z} |r_\xi(t)|^{k_0} = \infty$, then $\sum_{t \in Z} |r_X(t)| = \infty$ and (X_t) behaves as a long memory process.

Assumption SM. Under short memory, (X_t) has an absolutely summable autocovariance function:

$$\sum_{t \in Z} |r_X(t)| < \infty. \tag{4.7}$$

Assumption LM. Under long memory, the spectral density $f_X(\lambda)$ of (X_t) satisfies

$$f_X(\lambda) = b_0 |\lambda|^{-\alpha_X}(1 + o(1)), \quad \text{as } \lambda \to 0, \tag{4.8}$$

with $0 < \alpha_X < 1$ and $b_0 > 0$. In addition, we assume that the spectral density f_ξ of the Gaussian sequence (ξ_t) has property

$$f_\xi(\lambda) = |\lambda|^{-\alpha_\xi} g_\xi(\lambda) = |\lambda|^{-\alpha_\xi}(b_{0,\xi} + b_{1,\xi}\lambda^2 + o(\lambda^2)), \quad \text{as } \lambda \to 0, \tag{4.9}$$

where $0 < \alpha_\xi < 1$ and $b_{0,\xi} \neq 0$, and the covariance function of (ξ_t) satisfies

$$r_\xi(t) \sim c_1 t^{-1+\alpha_\xi}, \quad \text{as } t \to \infty, \quad (c_1 \neq 0). \tag{4.10}$$

It is well-known that if $g_\xi(\lambda)$ is a sufficiently smooth function, then (4.9) implies (4.10), see e.g. Lemma 4 in Fox and Taqqu (1986) and Yong (1974), whereas, (4.9) implies (4.8), see discussion below. Here "$a_n \sim b_n$" means that $a_n/b_n \to 1$, as $n \to \infty$.

Observing that (4.3) implies that

$$E[H_k(\xi_t)H_k(\xi_0)] = k! r_\xi^k(t) = k! \int_{-\pi}^{\pi} e^{i\lambda t} f^{(*k)}(\lambda) d\lambda,$$

where

$$f^{(*k)}(\lambda) = \int_{-\pi}^{\pi} ... \int_{-\pi}^{\pi} f_\xi(\lambda - x_1 - ... - x_{k-1}) f_\xi(x_1) ... f_\xi(x_{k-1}) dx_1 ... dx_{k-1}$$

is the k-th order convolution of $f_\xi(\lambda)$, we obtain that under both SM and LM assumptions, the spectral density f_X of (X_t) can be written as

$$f_X(\lambda) = \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} f^{(*k)}(\lambda), \tag{4.11}$$

(we assume that f_ξ is periodically extended to the real line R).

Under Assumption SM, the spectral density $f_X(\lambda)$ of (X_t) is a continuous bounded function and

$$f_X(\lambda) \to b_0 = f_X(0), \quad \text{as } \lambda \to 0, \tag{4.12}$$

so that $f_X(\lambda)$ satisfies assumption (1.1) with $\alpha_X = 0$ and $b_0 = f_X(0)$.

However, if f_ξ satisfies condition (4.9) and $0 < k(1-\alpha_\xi) < 1$, then it can be shown that

$$f^{(*k)}(\lambda) = k! s_k |\lambda|^{-1+k(1-\alpha_\xi)}(1+o(1)), \quad \text{as } \lambda \to 0, \tag{4.13}$$

for some $s_k > 0$, whereas if $k(1-\alpha_\xi) = 1$, then $f^{(*k)}(\lambda) = k! s_k |\log|\lambda||^{-1}(1+o(1))$.

Then (4.9), (4.11) and (4.13) imply that (X_t) satisfies Assumption LM and

$$f_X(\lambda) = \frac{c_{k_0}^2}{k_0!} f^{(*k_0)}(\lambda)(1+o(1)) = c_{k_0}^2 s_{k_0} |\lambda|^{-\alpha_X} + o(|\lambda|^{-\alpha_X}), \quad \text{as } \lambda \to 0, \tag{4.14}$$

where $\alpha_X = 1 - k_0(1-\alpha_\xi) > 0$, indicating the relationship between the long memory parameters α_X and α_ξ and the Hermite rank k_0. Note that $0 < k_0(1-\alpha_\xi) < 1$.

Theorem 4.1 *Suppose that a sequence (X_t) is defined by* (4.1).

(i) If (X_t) satisfies Assumption SM and the bandwidth parameter $m \to \infty$ is such that $m = o(n)$, then

$$\widehat{\alpha}_X \xrightarrow{P} \alpha_X = 0, \quad \text{as } n \to \infty.$$

(ii) If (X_t) satisfies Assumption LM, with memory parameter $0 < \alpha_X < 1$, and m is such that

$$n^\gamma \le m = o(n) \tag{4.15}$$

for some $1 - k_0^{-1} < \gamma < 1$ where $k_0 \ge 1$ is the Hermite rank of G, then

$$\widehat{\alpha}_X \xrightarrow{P} \alpha_X, \quad \text{as } n \to \infty. \tag{4.16}$$

(iii) If in case (ii) $k_0 = 1$ and (4.15) holds, then $\alpha_X = \alpha_\xi$ and

$$\widehat{\alpha}_X - \alpha_X = O_P\Big(m^{-1/2} + (m/n)^r\Big) \tag{4.17}$$

with $0 < r < \min(\alpha_\xi/2, (1-\alpha_\xi)/2)$. In addition,

$$m^{1/2}(\widehat{\alpha}_X - \alpha_X) \xrightarrow{d} N(0,1), \quad \text{as } n \to \infty, \tag{4.18}$$

if (2.25) holds and $m = o(n^{2r/(2r+1)})$.

Remark 4.1 We conjecture, that in part (ii) of Theorem 4.1, it can be shown that $\widehat{\alpha}_X - \alpha_X = O_P(n^{-\epsilon})$ for some $\epsilon > 0$.

Proof of Theorem 4.1 is based on the following proposition. In view of (4.2), write X_t as a signal plus noise process

$$X_t = \sum_{k=k_0}^{M} \frac{c_k}{k!} H_k(\xi_t) + \sum_{k=M+1}^{\infty} \frac{c_k}{k!} H_k(\xi_t) =: Y_t + Z_t, \tag{4.19}$$

where $M > k_0$ will be chosen later.

Proposition 4.1 *In case (i) and (ii) of Theorem 4.1, (X_t) satisfies Assumptions A and B with parameters $b_0 \neq 0$ and $\alpha_0 = \alpha_X$.*

In case (ii), the process $Y_t = (c_{k_0}/k_0!)H_{k_0}(\xi_t)$, corresponding to $M = k_0$ in (4.19), has memory parameter $\alpha_Y = \alpha_X$ and satisfies Assumptions A and B with parameters $b_0 \neq 0$ and $\alpha_0 = \alpha_Y$.

Proof of Proposition 4.1. (i) In this case (X_t) is a short memory sequence and $\alpha_X = 0$. Then Assumption A follows by noticing that by (4.12), $f_X(\lambda)$ satisfies (1.1) with $\alpha_X = 0$ and $b_0 = f_X(0)$. Next, in view of Proposition 2.2, (X_t) satisfies Assumption B, if

$$\Delta_m = \max_{1 \le k \le m} E \left| \sum_{j=1}^{k} b_0^{-1}(I_X(\lambda) - EI_X(\lambda)) \right| = o(m). \tag{4.20}$$

To show (4.20), we shall use decomposition (4.19) with large $M > k_0$. Using equalities (4.3) and (4.4), we obtain that

$$r_Z(t) = Cov(Z_t, Z_0) = \sum_{k=M+1}^{\infty} \frac{c_k^2}{k!} r_\xi^k(t),$$

so that

$$\sum_{t=1}^{\infty} |r_Z(t)| \le \left(\sum_{t=1}^{\infty} |r_\xi(t)|^M \right) \sum_{k=M+1}^{\infty} \frac{c_k^2}{k!} = \left(\sum_{t=1}^{\infty} |r_\xi(t)|^M \right) \epsilon_M \le C\epsilon_M$$

by (4.7), (4.6) and (4.5), where $\epsilon_M \to 0$ as $M \to \infty$ by the summability of $c_k^2/k!$. Therefore, the spectral density of (Z_t) satisfies the bound

$$\sup_{\lambda \in [0,\pi]} f_Z(\lambda) \le (2\pi)^{-1} \sum_{t=-\infty}^{\infty} |r_Z(t)| \le C\epsilon_M. \tag{4.21}$$

On the other hand, the same argument shows that

$$\sup_{\lambda \in [0,\pi]} f_Y(\lambda) \le (2\pi)^{-1} \sum_{t=-\infty}^{\infty} |r_Y(t)| \le C \tag{4.22}$$

where C does not depend on n.

Writing $I_X(\lambda_j)$ as $I_X(\lambda_j) = I_Y(\lambda_j) + v_j$, similarly as in (3.8), we obtain that

$$E\left|\sum_{j=1}^{k}(I_X(\lambda_j) - EI_X(\lambda_j))\right| \le E|S_k| + E|R_k|,$$

where

$$S_k = \sum_{j=1}^{k}(I_Y(\lambda_j) - EI_Y(\lambda_j)), \quad R_k = \sum_{j=1}^{k}(v_j - Ev_j).$$

Hence, to show (4.20) it suffices to estimate $E|R_k|$ and $E|S_k|$, $k = 1, ..., m$. Relations (3.9), (4.21) and (4.22) imply that $E|v_j| \le C\epsilon_M^{1/2}$, where the constant C is independent of M and n. The latter inequality implies the bound $E|R_k| \le Ck\epsilon_M^{1/2} = o\,(k)$, as $M \to \infty$, uniformly in $k = 1, ..., m$, so that (4.20) follows if we show that for any fixed M, $E|S_k| = o(m)$ uniformly in $k = 1, ..., m$, as $n \to \infty$. Applying the estimate (2.16) of Lemma 2.1 to the sequence (Y_t) and recalling that $\alpha_Y = 0$, we obtain that, uniformly in $1 \le k \le m$ and $n \ge 1$,

$$m^{-1}E|S_k| = O(m^{-1/2} + n^{-1/2}{D^*_{n,Y}}^{1/2}) = o\,(1)$$

after observing that Giraitis and Surgailis (1985, Relation (2.9)) implies that

$$\begin{aligned} D^*_{n,Y} &= \sum_{t_1,...,t_3=-n}^{n} |Cum(Y_{t_1}, Y_{t_2}, Y_{t_3}, Y_0)| \\ &\le n^{-1} \sum_{t_1,...,t_3,t_4=1}^{3n} |Cum(Y_{t_1}, Y_{t_2}, Y_{t_3}, Y_{t_4})| = o(n). \end{aligned}$$

(ii) Next, we consider the case when $\alpha_X > 0$. First, (X_t) satisfies Assumption A by (4.14). To show that (X_t) satisfies Assumption B, write $X_t = Y_t + Z_t$ as a signal plus noise process (4.19) with $M = k_0$. Then $Y_t = (c_{k_0}/k_0!)H_{k_0}(\xi_t)$. Note that (4.14) implies that the spectral density f_Y of (Y_t) has property

$$f_Y(\lambda) = b_{0,Y}|\lambda|^{-\alpha_Y} + o(|\lambda|^{-\alpha_Y}), \quad \text{as } \lambda \to 0,$$

where $\alpha_Y = \alpha_X = 1 - k_0(1 - \alpha_\xi) > 0$, whereas the spectral density f_Z of (Z_t) can be bounded by $f_Z(\lambda) \le C|\lambda|^{-\alpha_Z}$, as $\lambda \to 0$, where $\alpha_Z = \alpha_Y - \epsilon$ for some $\epsilon > 0$. We show below that the sequence (Y_t) satisfies relation (2.6) of Assumption B. Then the same argument as in the proof of Theorem 3.1 (i) yields that (X_t) satisfies Assumption B.

To prove (2.6), in view of Proposition 2.2, it suffices to show that

$$\Delta_{m,Y} = \max_{1 \le k \le m} b_{0,Y}^{-1} E\left|\sum_{j=1}^{k} \lambda_j^{\alpha_Y}(I_Y(\lambda_j) - EI_Y(\lambda_j))\right| = o(m). \tag{4.23}$$

Applying the estimate (2.17) of Lemma 2.1 to the sequence (Y_t), we obtain that

$$m^{-1}\Delta_{m,Y} = O\Big(m^{-1/2} + m^{-1}n^{1/2}D_n^{**1/2}(m/n)^{\alpha_Y}\log n\Big),$$

where $\alpha_Y = 1 - k_0(1-\alpha_\xi)$. It remains to estimate D_n^{**}. In case $k_0 = 1$, (Y_t) is a Gaussian sequence, $Cum(Y_{t_1}, Y_{t_2}, Y_{t_3}, Y_{t_4}) = 0$ and $D_n^{**} = 0$, so that (4.23) holds. Let $k_0 \geq 2$. Using cumulant formula (2.10) of Giraitis and Surgailis (1985), we have that

$$\Big|Cum(Y_{t_1}, Y_{t_2}, Y_{t_3}, Y_{t_4})| = C\Big|Cum(H_{k_0}(\xi_{t_1}), H_{k_0}(\xi_{t_2}), H_{k_0}(\xi_{t_3}), H_{k_0}(\xi_{t_4}))\Big|$$

$$\leq C\Big(r_\xi^2(t_1-t_3) + r_\xi^2(t_1-t_4) + r_\xi^2(t_2-t_3) + r_\xi^2(t_2-t_4)\Big).$$

By (4.10), $|r_\xi(t)|^2 \leq C|t|^{-2(1-\alpha_\xi)}$, where $2(1-\alpha_\xi) < 1$ because $k_0 \geq 2$ and $\alpha_X = 1 - k_0(1-\alpha_\xi) > 0$. Therefore,

$$D_n^{**} \leq C\sum_{t=1}^n r_\xi^2(t) \leq C\sum_{t=1}^n t^{-2(1-\alpha_\xi)} \leq Cn^{2\alpha_\xi - 1}$$

and hence

$$\begin{aligned} m^{-1}\Delta_{m,Y} &\leq C(m^{-1}n^{\alpha_\xi}(m/n)^{\alpha_Y}\log n + m^{-1/2}) \\ &\leq C\Big([(n/m)^{k_0}n^{-1}]^{1-\alpha_\xi}\log n + m^{-1/2}\Big) \to 0, \end{aligned}$$

as $n \to \infty$, because assumption (4.15) assures that $(n/m)^{k_0}n^{-1} \leq Cn^{-\epsilon}$ for some $\epsilon > 0$ which yields (4.23). □

Proof of Theorem 4.1. We showed in Proposition 4.1 that in case (i), and (ii), (X_t) satisfies Assumptions A and B which imply $\widehat{\alpha} \xrightarrow{P} \alpha_X$ by Theorem 2.1.

In case (iii), to derive (4.17) we shall use part (iii) of Theorem 3.1. Because $k_0 = 1$, we can write X_t as a signal plus noise process $X_t = Y_t + Z_t$, (4.19), with $Y_t = c_1H_1(\xi_t) = c_1\xi_t$ and $Z_t = \sum_{k=2}^\infty \frac{c_k}{k!}H_k(\xi_t)$. Let α_X, α_Y and α_Z be the memory parameters of (X_t), (Y_t) and (Z_t), respectively. Then $\alpha_X = \alpha_Y = \alpha_\xi$ since (ξ_t) is uncorrelated with (Z_t), in view of (4.4). We show below that $f_Z(\lambda) \leq C|\lambda|^{-\alpha_Z}$, as $\lambda \to 0$, with $\alpha_Z \geq 0$ such that

$$\alpha_Y > \alpha_Z = \begin{cases} 0, & \text{if } 2\alpha_\xi < 1 \\ 2\alpha_\xi - 1, & \text{if } 2\alpha_\xi > 1 \\ \epsilon, & \text{if } 2\alpha_\xi = 1 \end{cases} \tag{4.24}$$

for any $\epsilon \in (0, a_Y)$.

Indeed, if $2\alpha_\xi < 1$, then (4.6) applied to the sequence (Z_t) together with (4.10) imply that

$$\sum_{t=1}^\infty |r_Z(t)| \leq C\sum_{t=1}^\infty r_\xi^2(t) \leq C\sum_{t=1}^\infty t^{-2(1-\alpha_\xi)} < \infty.$$

Therefore the spectral density $f_Z(\lambda)$ of (Z_t) is a continuous function and $\alpha_Z = 0 < \alpha_\xi = \alpha_Y$.

If $2\alpha_\xi > 1$, then from equality $f_Z(\lambda) = \sum_{k=2}^\infty \frac{c_k^2}{k!}f^{(*k)}(\lambda)$, (4.14) and (4.9) it follows that that

$$f_Z(\lambda) \leq Cf^{(*2)}(\lambda) = c\lambda^{-\alpha_Z} + o(\lambda^{-\alpha_Z}), \quad \text{as } \lambda \to 0,$$

where $\alpha_Z = 1 - 2(1-\alpha_\xi) = 2\alpha_\xi - 1 > 0$ and $c > 0$. If $2\alpha_\xi = 1$ then

$$f_Z(\lambda) \le C f^{(*2)}(\lambda) \le C|\log(\lambda)|^{-1} \le C|\lambda|^{-\epsilon}$$

for any $\epsilon > 0$. This proves (4.24), since $\alpha_Z < \alpha_\xi = \alpha_Y$.

Thus (4.17) follows from (3.6) of Theorem 3.1, because Gaussian process $Y_t = c_1\xi_t$ can be represented as a linear sequence (2.18), the spectral density f_ξ satisfies condition $\mathrm{T}(\alpha_\xi, 2)$ and the conditions on m assures that $m = o(n/\log n)$, whereas (4.18) follows applying (3.2) in (3.4) of Theorem 3.1. $\square$

4.2 Estimation of the long memory parameter of a stochastic volatility model

In this section we consider the stochastic volatility model

$$r_t = \varepsilon_t \sigma_t, \quad t \in Z,$$

where (ε_t) is an i.i.d. noise with zero mean and finite variance and (σ_t) is a stationary volatility process independent of (ε_t). We shall analyse the long memory properties of the process

$$X_t = |r_t|^u = |\varepsilon_t|^u |\sigma_t|^u, \quad t \in Z \tag{4.25}$$

with some $u > 0$. Assume that the process $|\sigma_t|^u$ has long memory and satisfies (1.1) with parameters $b_0 > 0$ and $\alpha_Y > 0$. Then we can decompose X_t into a signal plus noise process

$$X_t = a|\sigma_t|^u + (|\varepsilon_t|^u - a)|\sigma_t|^u =: Y_t + Z_t, \tag{4.26}$$

where $a = E|\varepsilon_t|^u$. Since (Z_t) are uncorrelated variables, the spectral density function of (Z_t) is a constant and $\alpha_Z = 0$, whereas the signal (Y_t) has long memory. If the process of the u-th power $|\sigma_t|^u$ of the volatility σ_t satisfies Assumptions A and B with some $b_{0,Y} > 0$ and $\alpha_0 = \alpha_Y$, then $\alpha_X = \alpha_Y$ and by Theorem 2.1,

$$\hat{\alpha}_X \xrightarrow{P} \alpha_Y = \alpha_X.$$

Example: EGARCH process. Assume that the volatility $\sigma_t = f(\xi_t) > 0$ is a function of a stationary process (ξ_t) which is independent of (ε_t). Robinson (2001) showed that a wide class of stochastic volatility models with Gaussian (ξ_t) allow long memory behaviour in volatility. This type of models includes Exponential Generalized ARCH (EGARCH) model, suggested by Nelson (1991). A special case $f(\xi_t) = \exp(\xi_t)$, where (ξ_t) is a linear sequence was discussed in Breidt et al. (1998), Harvey (1998) and Surgailis and Viano (2002). Harvey (1998) examined the long memory properties of the process $X_t = |\varepsilon_t|^u \exp(u\xi_t)$, when (ξ_t) is a Gaussian sequence and showed, that for any $u > 0$,

$$r_X(t) = \mathrm{Cov}(X_t, X_0) \;=\; (EX_0)^2(e^{u^2 r_\xi(t)} - 1) \sim (uEX_0)^2 r_\xi(t), \quad \text{as } t \to \infty, \tag{4.27}$$

where $r_\xi(t) = \mathrm{Cov}(\xi_t, \xi_0)$ is the autocovariance function of (ξ_t). Relation (4.27) implies that the autocovariances $r_X(t)$ and $r_\xi(t)$ have the same rate of convergence to zero, as $t \to \infty$. Surgailis and Viano (2002) obtained similar result when (ξ_t) is a linear sequence.

We assume below that $\sigma_t = \exp(\xi_t)$ where (ξ_t) is a long memory Gaussian sequence with slowly decaying autocovariance

$$r_\xi(t) \sim c|t|^{-1+\alpha_\xi}, \quad \text{as } t \to \infty,$$

where $0 < \alpha_\xi < 1$ and the spectral density f_ξ of (ξ_t) satisfies (4.9).

In that case the sequence $Y_t = a\exp(u\xi_t)$, $t \in Z$ is a nonlinear transform of a Gaussian sequence (ξ_t) and has Hermite expansion $Y_t - EY_t = c_1 H_1(\xi_t) + ...$ with Hermite rank $k_0 = 1$ (since $c_1 \neq 0$). Therefore, by (4.14) of Section 4.1, the spectral density f_Y of (Y_t) has property

$$f_Y(\lambda) = b_{0,Y}|\lambda|^{-\alpha_\xi}(1 + o(1)), \quad \text{as } \lambda \to 0, \tag{4.28}$$

which implies that

$$f_X(\lambda) = f_Y(\lambda) + f_Z(\lambda) = b_{0,Y}|\lambda|^{-\alpha_\xi}(1 + o(1)), \quad \text{as } \lambda \to 0.$$

Hence, the sequences (X_t), (Y_t) and (ξ_t) have the same memory parameter

$$\alpha_X = \alpha_Y = \alpha_\xi > 0.$$

The next theorem shows that the local Whittle estimate $\widehat{\alpha}_X$ is a consistent estimate of the long memory parameter α_X of an EGARCH sequence (X_t) and satisfies the central limit theorem.

THEOREM 4.2 *Assume that $r_t = \varepsilon_t \exp(\xi_t)$ is an EGARCH model, (X_t) follows (4.25), $0 < \alpha_\xi < 1$, and m satisfies $n^\gamma \le m = o(n/\log n)$ for some $0 < \gamma < 1$. Then, as $n \to \infty$,*

$$\widehat{\alpha}_X - \alpha_X = O\Big(m^{-1/2} + (m/n)^r\Big) \tag{4.29}$$

for any $0 < r < \min(\alpha_\xi, 1 - \alpha_\xi)/2$. Moreover,

$$m^{1/2}(\widehat{\alpha}_X - \alpha_X) \xrightarrow{d} N(0,1), \tag{4.30}$$

if (ξ_t) satisfies (2.25) and $m = o(n^{2r/(2r+1)})$.

PROOF OF THEOREM 4.2. In the decomposition $X_t = Y_t + Z_t$ given in (4.26), the memory parameters of the sequences (Y_t) and (Z_t) have property $\alpha_Y > \alpha_Z = 0$. Proposition 4.1 and (4.28) imply that (Y_t) satisfies Assumptions A and B with parameters $b_{0,Y}$ and $\alpha_Y = \alpha_\xi$. Therefore by (3.4) of Theorem 3.1,

$$\widehat{\alpha}_X - \alpha_X = (\widehat{\alpha}_Y - \alpha_Y)(1 + o_P(1)) + O_P((m/n)^{\alpha_Y/2} + m^{-1}\log m) \tag{4.31}$$

where $\widehat{\alpha}_Y$ denotes the local Whittle estimator of (Y_t) as if the Y_t's were observed. By (4.17), $\widehat{\alpha}_Y - \alpha_Y = O_P(m^{-1/2} + (m/n)^r)$, which implies (4.29). Convergence (4.30) follows from (4.31) and (4.18). □

REMARK 4.2 In the short memory case, proof of consistency seems to be more technically involved. We conjecture that if (X_t) has short memory, then using similar techniques it can be show that $\widehat{\alpha}_X \xrightarrow{P} \alpha_X = 0$. Simulations support this result, see Tables 5 in Section 7.

Remark 4.3 Theorem 4.2 shows that the local Whittle estimator allows to estimate the long memory parameter of the powers $|r_t|^u$ of an EGARCH model r_t. On the other hand, the logarithms $X_{\log}(t) = \log |r_t|^u$ can be written as a signal plus noise process

$$X_{\log}(t) = Y_t + \eta_t, \tag{4.32}$$

where $Y_t = u\xi_t$ and $\eta_t = u \log |\varepsilon_t|$ are i.i.d. shocks, so that the memory parameter of the sequence (ξ_t) can be estimated applying the local Whittle estimator $\widehat{\alpha}_{\log}$ to $X_{\log}(t)$. In case of a linear process (ξ_t), consistency and asymptotic distribution of the local Whittle estimator $\widehat{\alpha}_{\log}$ were analysed in Hurvich, Moulines and Soulier (2005) and Arteche (2004).

Note that the model (4.32) is a particular case of a signal plus noise process discussed in our Theorem 3.1 which allows unrestricted mutual dependence of (ξ_t) and the noise (ε_t). Theorem 3.1 (iii) shows that if (ξ_t) is a linear process with the spectral density $f_\xi(\lambda) = |\lambda|^{-\alpha_\xi}(b_{0,\xi} + b_{1,\xi}\lambda^2 + o(\lambda^2))$ and $0 < \alpha_\xi < 1$, then the estimator $\widehat{\alpha}_{X_{\log}}$ satisfies

$$\widehat{\alpha}_{X_{\log}} - \alpha_\xi = O_P(m^{-1/2} + (m/n)^{\alpha_\xi/2}). \tag{4.33}$$

If (ξ_t) is a Gaussian sequence, the processes $\log(r_t^u)$, (ξ_t) and (r_t^u) have the same long memory parameter α_ξ which in view of (4.33) and Theorem 4.2 can be consistently estimated by the local Whittle estimator applied to $(|r_t|^u)$ or $(\log |r_t|^u)$.

5 Proofs of Theorems 2.1, 2.2, Propositions 2.2, 2.3 and Lemma 2.1

Proof of Theorem 2.1. 1. *Proof of consistency* (2.8). It suffices to show that for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$P\Big\{\inf_{\alpha\in[-1,1],|\alpha-\alpha_0|\geq\epsilon}(U_n(\alpha) - U_n(\alpha_0)) \geq \delta\Big\} \to 1, \quad \text{as } n \to \infty, \tag{5.1}$$

where

$$U_n(\alpha) = \log\left(\frac{1}{m}\sum_{j=1}^m (j/m)^\alpha I_n(\lambda_j)\right) - \frac{\alpha}{m}\sum_{j=1}^m \log(j/m).$$

Because $m^{-1}\sum_{j=1}^m \log(j/m) = -1 + o(1)$, we have that

$$U_n(\alpha) - U_n(\alpha_0) = \log L_n(\alpha) - \log L_n(\alpha_0) + \alpha - \alpha_0 + o(1),$$

where

$$L_n(\alpha) = m^{-1}\sum_{j=1}^m (j/m)^{\alpha-\alpha_0}\eta_j^*$$

and η_j^* is defined in (2.5).

We assume that the variables η_j^* satisfy (2.6). Observe that $E\eta_j^* \leq C$ for all $1 \leq j \leq m$ by (1.1) and (6.10) of Lemma 6.3 below. Thus, by Lemma 6.1 below, for any $\epsilon' > 0$,

$$\sup_{\alpha\in[-1,1],\alpha\geq\alpha_0-1+\epsilon'} |L_n(\alpha) - L(\alpha)| \xrightarrow{P} 0,$$

where

$$L(\alpha) = \int_0^1 x^{\alpha-\alpha_0} dx = (1+\alpha-\alpha_0)^{-1}.$$

Hence, as $n \to \infty$, with probability tending to 1, uniformly in $\alpha \in [\alpha_0 - 1 + \epsilon', 1]$, $|\alpha - \alpha_0| \geq \epsilon$,

$$\begin{aligned} U_n(\alpha) - U_n(\alpha_0) &= \log L(\alpha) - \log L(\alpha_0) + \alpha - \alpha_0 + o(1) \\ &= -\log(1+\alpha-\alpha_0) + \alpha - \alpha_0 + o(1) \geq \delta_\epsilon > 0, \end{aligned} \tag{5.2}$$

since $-\log(1+x) + x > 0$ for $x > -1$.

On the other hand, uniformly in $\alpha \in [-1, \alpha_0 - 1 + \epsilon']$,

$$\begin{aligned} U_n(\alpha) - U_n(\alpha_0) &\geq \log L_n(\alpha_0 - 1 + \epsilon') - \log L_n(\alpha_0) + \alpha - \alpha_0 + o(1) \\ &\geq -\log(\epsilon') - 1 - \alpha_0 + o(1) \geq 1 \end{aligned} \tag{5.3}$$

when $\epsilon' > 0$ is small. Estimates (5.2) and (5.3) imply (5.1). □

2. *Proof of (2.9).* Suppose that $0 < \epsilon < \min(1-\alpha_0, \alpha_0 + 1)$. Since by (2.8) $\widehat{\alpha} \xrightarrow{P} \alpha_0$, then, as $n \to \infty$, $\mathbf{1}(|\widehat{\alpha} - \alpha_0| \leq \epsilon) = 1 + o_P(1)$ and $\mathbf{1}(\widehat{\alpha} - \alpha_0| > \epsilon) = o_P(1)$, where $\mathbf{1}(A)$ is the indicator function. We shall show below that

$$(\widehat{\alpha} - \alpha_0)\mathbf{1}(|\widehat{\alpha} - \alpha_0| \leq \epsilon) = -Q_m(1 + o_P(1)) + O_P(m^{-1}\log m). \tag{5.4}$$

Assuming, that (5.4) holds true, we conclude that

$$\begin{aligned} \widehat{\alpha} - \alpha_0 &= -Q_m(1 + o_P(1)) + O_P(m^{-1}\log m) + (\widehat{\alpha} - \alpha_0)\mathbf{1}(|\widehat{\alpha} - \alpha_0| > \epsilon) \\ &= -Q_m(1 + o_P(1)) + O_P(m^{-1}\log m) + (\widehat{\alpha} - \alpha_0) o_P(1), \end{aligned}$$

which implies that

$$\widehat{\alpha} - \alpha_0 = -Q_m(1 + o_P(1)) + O_P(m^{-1}\log m), \tag{5.5}$$

to prove (2.9).

We now show (5.4). First, we notice that

$$\frac{\partial}{\partial \alpha} U_n(\alpha) = \frac{T_n(\alpha)}{V_n(\alpha)},$$

where

$$T_n(\alpha) = \frac{1}{m}\sum_{j=1}^m (j/m)^{\alpha-\alpha_0} \nu_j \eta_j^*, \quad V_n(\alpha) = \frac{1}{m}\sum_{j=1}^m (j/m)^{\alpha-\alpha_0} \eta_j^* \geq 0, \tag{5.6}$$

with $\nu_j = \log j - m^{-1}\sum_{k=1}^m \log k$ and η_j^* given by (2.5). From Lemma 6.1 below, it follows that

$$V_n(\widehat{\alpha}) \geq \frac{1}{m}\sum_{j=1}^m (j/m)^\epsilon \eta_j^* \xrightarrow{P} \int_0^1 x^\epsilon dx > 0. \tag{5.7}$$

Observe that assumption $|\widehat{\alpha} - \alpha_0| \leq \epsilon$ implies that $\widehat{\alpha} \in (-1, 1)$. Therefore $\frac{\partial}{\partial\alpha} U_n(\widehat{\alpha}) = 0$ which yields that $T_n(\widehat{\alpha}) = 0$. By the mean value theorem,

$$T_n(\widehat{\alpha}) - T_n(\alpha_0) = \frac{\partial}{\partial \alpha} T_n(\alpha^*)(\widehat{\alpha} - \alpha_0), \tag{5.8}$$

where α^* is an intermediate point between $\widehat{\alpha}$ and α_0. To complete the proof of (5.4), it remains to show that

$$T_n(\alpha_0) = m^{-1}\sum_{j=1}^{m}(\log(j/m)+1)\eta_j^* + O_P(m^{-1}\log m) = Q_m + O_P(m^{-1}\log m) \qquad (5.9)$$

and

$$\frac{\partial}{\partial\alpha}T_n(\alpha^*) \xrightarrow{P} 1, \qquad (5.10)$$

which together with (5.8) imply (5.4).

Using relation

$$\nu_j = \log(j/m) + 1 + O(m^{-1}\log m),$$

see Robinson (1995b, Lemma 2), which holds uniformly in $1 \leq j \leq m$, we can write $T_n(\alpha_0)$ as

$$T_n(\alpha_0) = m^{-1}\sum_{j=1}^{m}(\log(j/m)+1)\eta_j^* + R_m = Q_m + R_m,$$

where, in view of (2.7),

$$E|R_m| = O(m^{-1}\log m)m^{-1}\sum_{j=1}^{m}E\eta_j^* = O(m^{-1}\log m),$$

which implies that $R_m = O_P(m^{-1}\log m)$ and proves (5.9).

It remains to show (5.10). To that end, write

$$\begin{aligned}\frac{\partial}{\partial\alpha}T_n(\alpha^*) &= m^{-1}\sum_{j=1}^{m}\nu_j(j/m)^{\alpha^*-\alpha_0}\log(j/m)\eta_j^* \\ &= m^{-1}\sum_{j=1}^{m}\Big(\psi(j/m;\alpha) + r_m(j/m;\alpha)\Big)\eta_j^*\end{aligned}$$

where $\psi(j/m;\alpha) = \log(j/m)(\log(j/m)+1)$ and

$$r_m(j/m;\alpha) = \nu_j(j/m)^{\alpha^*-\alpha_0}\log(j/m) - \psi(j/m;\alpha).$$

Note that

$$|r_m(j/m;\alpha)| = \Big|\Big(\log(j/m) + 1 + O(m^{-1}\log m)\Big)(j/m)^{\alpha^*-\alpha_0}\log(j/m) - \psi(j/m;\alpha)\Big|$$

$$\leq |\psi(j/m;\alpha)||(j/m)^{\alpha^*-\alpha_0} - 1| + |(j/m)^{\alpha^*-\alpha_0}\log(j/m)O(m^{-1}\log m) + O(m^{-1}\log^3 m)|.$$

We now show that the function $r_m(j/m;\alpha)$ satisfies assumptions (6.7)-(6.8) of Lemma 6.2. For any $0 < \gamma < 1$, uniformly in $\gamma \leq j/m \leq 1$, it holds that

$$|r_m(j/m;\alpha)| \leq C\Big|\gamma^{-|\alpha^*-\alpha_0|} - 1\Big| + o_P(1) = o_P(1),$$

because $\alpha^* \xrightarrow{P} \alpha_0$ which implies (6.7). On the other hand, uniformly in $0 < j/m \leq \gamma$,

$$|r_m(j/m;\alpha)| \leq C(j/m)^{-\epsilon}|\log(j/m)|^2,$$

so that $r_m(j/m;\alpha)$ satisfies (6.8). Then, by Lemma 6.2 below, we conclude that

$$\frac{\partial}{\partial\alpha}T(\alpha^*) \xrightarrow{P} \int_0^1 \log x(\log x + 1)dx = 1,$$

to prove (5.10). □

Proof of Theorem 2.2. Expansion (2.12) follows from relations (2.9)-(2.10) using the following expansion of EQ_m which is derived applying (6.10) of Lemma 6.3 below:

$$\begin{aligned} EQ_m &= m^{-1}\sum_{j=1}^{m}(\log(j/m)+1)b_0^{-1}g(\lambda_j)E\eta_j \\ &= m^{-1}\sum_{j=1}^{m}(\log(j/m)+1)\Big(1+(b_1/b_0)\lambda_j^\beta + o(\lambda_j^\beta)\Big)\Big(1+O(j^{-1}\log j)\Big) \\ &= (m/n)^\beta(b_1/b_0)(2\pi)^\beta\int_0^1(\log x+1)x^\beta dx + o\Big(m^{-1/2}+(m/n)^\beta\Big) \\ &= (m/n)^\beta(b_1/b_0)B_\beta + o\Big(m^{-1/2}+(m/n)^\beta\Big). \end{aligned}$$

□

Proof of Proposition 2.2. It suffices to show that (X_t) satisfies Assumption B. Since (X_t) satisfies assumption (1.1), then the convergence $\widehat{\alpha} \xrightarrow{P} \alpha_0$ follows from Theorem 2.1. Note that from (1.1) and Lemma 6.3 below, it follows that $E\eta_j^* = 1 + o(1)$, uniformly in $1 \leq j \leq m$, as $n \to \infty$, which implies that

$$m^{-1}\sum_{j=1}^{m}E\eta_j^* \to 1. \tag{5.11}$$

On the other hand, in view of assumption $\Delta_m = o(m)$,

$$E\Big|m^{-1}\sum_{j=1}^{m}(\eta_j^* - E\eta_j^*)\Big| \leq Cm^{-1}\Delta_m \to 0,$$

which together with (5.11) proves (2.6). □

Proof of Proposition 2.3. To show (2.13), set $S_k = \sum_{j=1}^{k}(\eta_j^* - E\eta_j^*)$. Summation by parts implies

$$Q_m - EQ_m = m^{-1}\sum_{j=1}^{m-1}\Big(\log(j/m) - \log((j+1)/m)\Big)S_j + m^{-1}S_m.$$

Thus,

$$E|Q_m - EQ_m| \le m^{-1} \sum_{j=1}^{m-1} j^{-1} E|S_j| + Cm^{-1} E|S_m| \le C\Delta_m m^{-1} \log m, \qquad (5.12)$$

because $E|S_k| \le \Delta_m$, which together with expansion (2.12) implies (2.13).

To prove (2.14), it suffices to show that

$$E|Q_m - EQ_m - V_m| = o((m/n)^\beta). \qquad (5.13)$$

Note that

$$Q_m - EQ_m - V_m = m^{-1} \sum_{j=1}^{m} (\log(j/m) + 1)(b_0^{-1} - g(\lambda_j)^{-1})(\eta_j^* - E\eta_j^*) b_0,$$

where, by Assumption $\mathrm{T}(\alpha_0, \beta)$ and (1.1),

$$b_0^{-1} - g(\lambda_j)^{-1} = (b_1/b_0^2)\lambda_j^\beta + o(\lambda_j^\beta),$$

uniformly in $1 \le j \le m$. Thus, by triangle inequality,

$$\begin{aligned} E|Q_m - EQ_m - V_m| &\le E\Big| m^{-1} \sum_{j=1}^{m} (\log(j/m) + 1)(b_1/b_0^2)\lambda_j^\beta (\eta_j^* - E\eta_j^*)\Big| \\ &+ m^{-1} \sum_{j=1}^{m} |\log(j/m) + 1| o(\lambda_j^\beta) E\eta_j^* =: R_1 + R_2. \end{aligned}$$

Write $p_j = (\log(j/m) + 1)(j/n)^\beta$. Then,

$$\begin{aligned} |p_j - p_{j+1}| &\le |\log j - \log(j+1)|(j/n)^\beta \\ &+ |\log((j+1)/m) + 1||(j/n)^\beta - ((j+1)/n)^\beta| \le C(m/n)^\beta j^{-1} \log m \end{aligned}$$

and summation by parts yields

$$\begin{aligned} R_1 &= CE\Big| m^{-1} \sum_{j=1}^{m-1} (p_j - p_{j+1}) S_j + m^{-1} p_m S_m \Big| \\ &\le C(m/n)^\beta \Big(m^{-1} \log m \sum_{j=1}^{m-1} j^{-1} E|S_j| + m^{-1} E|S_m| \Big) \\ &\le Cm^{-1} \Delta_m (m/n)^\beta \log^2 m = o((m/n)^\beta), \end{aligned}$$

because $\Delta_m = o(m/\log^2 m)$. On the other hand, using (2.7) we obtain that

$$R_2 = o((m/n)^\beta) m^{-1} \sum_{j=1}^{m} |\log(j/m) + 1| = o((m/n)^\beta),$$

to prove (5.13). $\square$

Proof of Lemma 2.1. Denote

$$v_X(\lambda_j) = (b_0\lambda_j^{-\alpha_0})^{-1/2}(2\pi n)^{-1/2}\sum_{t=1}^{n} X_t e^{it\lambda_j}. \tag{5.14}$$

Then $\eta_j^* = |v_X(j)|^2$ and we can write

$$Cov(\eta_j^*, \eta_p^*) = Cov(v_X(\lambda_j), v_X(\lambda_p))Cov(\overline{v_X}(\lambda_j), \overline{v_X}(\lambda_p))$$
$$+Cov(v_X(\lambda_j), \overline{v_X}(\lambda_p))Cov(\overline{v_X}(\lambda_j), v_X(\lambda_p)) + Cum\Big(v_X(\lambda_j), \overline{v_X}(\lambda_j), v_X(\lambda_p), \overline{v_X}(\lambda_p)\Big).$$

Thus,

$$\begin{aligned} E\Big(\sum_{j=1}^{k}(\eta_j^* \quad - \quad E\eta_j^*)\Big)^2 &= \sum_{j,p=1}^{k} Cov(\eta_j^*, \eta_p^*) \\ &\le \quad \sum_{j,p=1}^{k}\Big(|Cov(v_X(\lambda_j), v_X(\lambda_p))|^2 + |Cov(v_X(\lambda_j), \overline{v_X}(\lambda_p))|^2\Big) \\ &+ \quad \Big|\sum_{j,p=1}^{k} Cum(v_X(\lambda_j), \overline{v_X}(\lambda_j), v_X(\lambda_p), \overline{v_X}(\lambda_p))\Big| =: i_{n,1}(k) + i_{n,2}(k). \end{aligned}$$

Therefore,

$$\Delta_m \le \max_{1\le k\le m}\Big(i_{n,1}(k) + i_{n,2}(k)\Big)^{1/2} \le \max_{1\le k\le m}(i_{n,1}(k)^{1/2} + i_{n,2}(k)^{1/2}). \tag{5.15}$$

Now, by (6.12) of Lemma 6.3,

$$i_{n,1}(k) \le C\sum_{1\le j\le p\le m}(j^{-|\alpha_0|}p^{-2+|\alpha_0|}\log^2 m + 1_{\{j=p\}}) \le C(\log^3 m + m) \le Cm. \tag{5.16}$$

On the other hand, uniformly in $1 \le k \le m$,

$$\begin{aligned} i_{n,2}(k) \quad &\le \quad \sum_{j,p=1}^{k}\lambda_j^{\alpha_0}\lambda_p^{\alpha_0}b_0^{-2}(2\pi n)^{-2}\Big|\sum_{t_1,\ldots,t_4=1}^{n} e^{i(t_1-t_2)\lambda_j}e^{i(t_3-t_4)\lambda_p}Cum(X_{t_1}, X_{t_2}, X_{t_3}, X_{t_4})\Big| \\ &\le \quad Cn^{-2}(\sum_{j=1}^{k}\lambda_j^{\alpha_0})^2\sum_{t_1,\ldots,t_4=1}^{n}|c_X(t_1,t_2,t_3,t_4)| \\ &\le \quad Cn^{-1}(m/n)^{2\alpha_0}m^2\sum_{u_1,\ldots,u_3=-n}^{n}|c_X(u_1,u_2,u_3,0)| \\ &\le \quad Cm^2(m/n)^{2\alpha_0}n^{-1}D_n^*, \end{aligned}$$

which together with (5.16) and (5.15) imply the bound (2.16).

To show (2.17), note that $i_{n,2}(k)$ can be written as

$$i_{n,2}(k) = \Big|\sum_{t_1,\ldots,t_4=1}^{n} B_k(t_1-t_2)B_k(t_3-t_4)Cum(X_{t_1}, X_{t_2}, X_{t_3}, X_{t_4})\Big|,$$

where

$$B_k(t) = (2\pi n)^{-1} \sum_{j=1}^{k} b_0^{-1} \lambda_j^{\alpha_0} e^{it\lambda_j}.$$

Then,

$$\begin{aligned}
i_{n,2} &= \Big| \sum_{t_1,\dots,t_4=1}^{n} B_k(t_1-t_2) B_k(t_3-t_4) c_X(t_1-t_2, 0, t_3-t_2, t_4-t_2) \Big| \\
&\le \Big| n \sum_{u_1,u_2,u_3=-n}^{n} B_k(u_1) B_k(u_2) c_X(u_1, 0, u_2+u_3, u_3) \Big| \\
&\le n \Big(\sum_{u=-n}^{n} |B_k(u)| \Big)^2 D_n^{**}.
\end{aligned}$$

We show below that in the case $\alpha_0 \ge 0$,

$$|B_k(t)| \le C(m/n)^{\alpha_0} |t|_+^{-1}, \tag{5.17}$$

where $|t|_+ = \max(|t|, 1)$, which implies that

$$i_{n,2} \le Cn(m/n)^{2\alpha_0} \log^2 n \, D_n^{**}$$

and together with (5.15) and (5.16) proves (2.17).

To check (5.17), set $s_p = \sum_{j=1}^{p} e^{it\lambda_j}$. Summation by parts yields that

$$B_k(t) = Cn^{-1-\alpha_0} \Big(\sum_{j=1}^{k-1} (j^{\alpha_0} - (j+1)^{\alpha_0}) s_j + k^{\alpha_0} s_k \Big).$$

Because

$$|s_p| = |e^{i\lambda_t}(1 - e^{ip\lambda_t})/(1 - e^{i\lambda_t})| \le 2/|1 - e^{i\lambda_t}| \le Cn/|t|_+,$$

we obtain that, for $\alpha_0 \ge 0$,

$$|B_k(t)| \le C|t|_+^{-1} n^{-\alpha_0} \Big(\sum_{j=1}^{k-1} |j^{\alpha_0} - (j+1)^{\alpha_0}| + k^{\alpha_0} \Big) \le C|t|_+^{-1} (m/n)^{\alpha_0},$$

to prove (5.17). □

6 Lemmas

LEMMA 6.1 *Assume that a triangular array of random variables $y_j \equiv y_{j,m}$, $1 \le j \le m$ is such that*

$$E|y_j| \le C \tag{6.1}$$

holds for all $1 \le j \le m$, $m \ge 1$, and for any $0 < \tau \le 1$,

$$[\tau m]^{-1} \sum_{j=1}^{[\tau m]} y_j \xrightarrow{P} 1, \quad \text{as } m \to \infty. \tag{6.2}$$

Suppose that a function $w(x;\alpha)$, $0 \le x \le 1$, $\alpha \in [a_1, a_2] \subset R$ has the following properties: for any $0 < b < 1$,

$$\sup_{b \le x \le 1} \sup_{\alpha \in [a_1,a_2]} |(\partial/\partial x) w(x;\alpha)| \le C < \infty \tag{6.3}$$

and there exists $0 < \gamma < 1$ and $c > 0$ such that

$$\sup_{\alpha \in [a_1,a_2]} |w(x,\alpha)| \le c x^{-\gamma}, \quad \text{as } x \to 0. \tag{6.4}$$

Then, as $m \to \infty$,

$$\sup_{\alpha \in [a_1,a_2]} \Big| m^{-1} \sum_{j=1}^{m} w(j/m;\alpha) y_j - \int_0^1 w(x;\alpha) dx \Big| \xrightarrow{P} 0. \tag{6.5}$$

Proof of Lemma 6.1. Let $0 < b \le 1$. Under assumption (6.3), the function $w(x,\alpha)$, $x \in [b,1]$ can be approximated by a step function $w^{\Delta}(x;\alpha)$ in $x \in [b,1]$ uniformly in $\alpha \in [a_1, a_2]$. Since the convergence (6.2) implies that for any $0 \le \tau_1 < \tau_2 \le 1$,

$$([\tau_2 m] - [\tau_1 m])^{-1} \sum_{j=[\tau_1 m]+1}^{[\tau_2 m]} y_j \xrightarrow{P} 1, \quad \text{as } m \to \infty,$$

a straightforward argument shows that

$$m^{-1} \sum_{j=[bm]+1}^{m} w(j/m;\alpha) y_j \xrightarrow{P} \int_b^1 w(x;\alpha) dx, \tag{6.6}$$

uniformly in α. Setting

$$J_m(\alpha) = m^{-1} \sum_{j=1}^{[bm]} w(j/m;\alpha) y_j,$$

from (6.4) and (6.1), it follows that uniformly in $m \ge 1$,

$$\begin{aligned} E \sup_{\alpha \in [a_1,a_2]} |J_m(\alpha)| &\le c m^{-1} \sum_{j=1}^{[bm]} (j/m)^{-\gamma} E|y_j| \le C m^{-1} \sum_{j=1}^{[bm]} (j/m)^{-\gamma} \\ &\le C \int_0^b x^{-\gamma} dx \to 0, \quad \text{as} \quad b \to 0. \end{aligned}$$

Hence, as $m \to \infty$ and $b \to 0$,

$$\sup_{\alpha \in [a_1,a_2]} |J_m(\alpha)| \xrightarrow{P} 0,$$

which together with (6.6) completes the proof of (6.5). □

LEMMA 6.2 *Assume that the random variables $y_j \equiv y_{j,m}, 1 \le j \le m$, and a function $w(x;\alpha)$ satisfy assumptions of Lemma 6.1. Suppose that the random variables $r_m(x,\alpha), m \ge 1$ are such that for any $0 < b < 1$, as $m \to \infty$,*

$$\sup_{b \le x \le 1} \sup_{\alpha \in [a_1,a_2]} |r_m(x,\alpha)| = o_P(1) \tag{6.7}$$

and there exist $0 < \gamma' < 1$ *and* $c > 0$ *such that*

$$\sup_{\alpha \in [a_1, a_2]} |r_m(x, \alpha)| = O_P(x^{-\gamma'}), \quad \text{as } x \to 0. \tag{6.8}$$

Then, as $m \to \infty$,

$$\sup_{\alpha \in [a_1, a_2]} \left| m^{-1} \sum_{j=1}^{m} \Big(w(j/m; \alpha) + r_m(j/m; \alpha) \Big) y_j - \int_0^1 w(x; \alpha) dx \right| \xrightarrow{P} 0. \tag{6.9}$$

Lemma 6.2 is a straightforward generalization of Lemma 6.1.

The next lemma deals with properties of the discrete renormalized Fourier transforms $v_X(\lambda_j)$, defined by (5.14).

LEMMA 6.3 *(Robinson (1995a)). Let assumption (1.1) be satisfied. Then uniformly in* $1 \le k < j = o(n)$, *as* $n \to \infty$,

$$E(I_n(\lambda_j)/f(\lambda_j)) = 1 + O(j^{-1} \log j), \tag{6.10}$$

$$E v_X(\lambda_j) v_X(\lambda_j) = O(j^{-1} \log j), \tag{6.11}$$

$$|E v_X(\lambda_j) \overline{v_X(\lambda_k)}| + |E v_X(\lambda_j) v_X(\lambda_k)| = O(k^{-|\alpha_0|/2} |j|^{-1+|\alpha_0|/2} \log j). \tag{6.12}$$

This result was derived by Robinson (1995a), but in the actual statement of his Theorem 2, (c) was replaced by the weaker bound $k^{-|\alpha_0|/2} |j|^{-1+|\alpha_0|/2} \log j \le k^{-1} \log j$.

7 Monte-Carlo experiment

To investigate the performance of the local Whittle estimator $\widehat{\alpha}_X$ in finite samples, we have conducted a set of Monte-Carlo experiments employing 10000 replications with sample sizes $n = 1024$ and 2048, and bandwidth parameters $m = [n^{0.5}], [n^{0.6}], [n^{0.7}]$ and $[n^{0.8}]$.

In Table 7 we report the bias and, in parenthesis, the mean squared error M.S.E. of the local Whittle estimator when (X_t) follows a Gaussian ARFIMA$(0, \alpha_0/2, 0)$ process with memory parameter $\alpha_0 = -0.8, -0.4, 0, 0.4, 0.8$, and generated by Davies and Harte (1987) algorithm. The results are similar to those reported in Robinson (1995b). The estimator $\widehat{\alpha}_X$ seems to have negative bias when the process has short or long memory, whereas in case of antipersistence the bias tends to be positive. For a given n and m, M.S.E. does not depend on α_0. The optimal bandwidth m minimizing the M.S.E. is of order $[n^{0.8}]$ which confirms the findings by Henry and Robinson (1996).

Table 1 gives the bias and M.S.E. of $\widehat{\alpha}_X$ for the signal plus noise process $X_t = Y_t + Z_t$, where (Y_t) and (Z_t) are Gaussian ARFIMA$(0, \alpha_Y/2, 0)$ and ARFIMA$(0, \alpha_Z/2, 0)$ processes with memory parameters $\alpha_Y = 0, 0.4, 0.8$ and $\alpha_Z = -0.8, -0.4, 0, 0.4$, respectively, and such that $\alpha_Y > \alpha_Z$. The signal (Y_t) and the noise (Z_t) are independent and have unit variance. Table 1 and Table 7 show, that, as the theory predicts, the noise significantly increases the bias of the estimator. The bias tends to decrease when the difference $\alpha_Y - \alpha_Z$ increases, and it remains always negative when the signal and the noise are independent. For a fixed

n and m, the M.S.E. varies across α_X, and the bandwidth minimizing M.S.E. depends on α_Y and α_Z. Overall, it appears that the bandwidth parameter $m = [n^{0.6}]$ results to the lowest M.S.E.

Tables 2 and 3 summarize the performance of $\widehat{\alpha}_X$ when $X_t = G(\xi_t)$ with $G(\xi_t) = \exp(\xi_t)$ and $G(\xi_t) = \xi_t^2$, where (ξ_t) is a Gaussian ARFIMA$(0, \alpha_\xi/2, 0)$ process with $\alpha_\xi = -0.4, 0, 0.4, 0.8$. The bias is again bigger than in the linear case except when $\alpha_\xi = 0$. The estimator performs better in case $X_t = \xi_t^2$, while in most cases the bias tends to be negative. For a fixed n and m, the M.S.E. varies across α_ξ, indicating that the optimal bandwidth parameter depends on α_ξ. Overall, the tables suggest that $m = [n^{0.7}] - [n^{0.8}]$ gives the lowest M.S.E..

Tables 4 and 5 contain the estimation results for r_t^2 and $\log(r_t^2)$, where r_t follows the EGARCH model $r_t = \varepsilon_t \exp(\xi_t)$, generated by an i.i.d. Gaussian sequence (ε_t) and an ARFIMA$(0, \alpha_\xi/2, 0)$ Gaussian process (ξ_t) with memory parameter $\alpha_\xi = 0, 0.4, 0.8$. Moreover, (ε_t) and (ξ_t) are independent sequences with unit variance. The tables indicate that the estimation of the memory parameter is more accurate when it is based on the sequence $(\log(r_t^2))$ than that based on (r_t^2), especially when the process (ξ_t) has long memory. Notice that the processes $\left(\log(r_t^2)\right)$ and $\left(r_t^2\right)$ can be written as signal plus noise model. In both cases, the signal is uncorrelated with the noise, which induces a negative bias, as the results of Table 1 would suggest. In addition, in the signal plus noise decomposition of r_t^2, the nonlinear signal $E[\varepsilon_t^2]\exp(2\xi_t)$ adds further negative bias, see Table 2. Tables 4 and 5 show that the estimator performs considerably better under short memory dependence and that the M.S.E. is not uniform across α_ξ. The tables suggest that the bandwidth parameters $m = [n^{0.6}] - [n^{0.7}]$ give the best finite sample performance.

Unsurprisingly, in all cases, the bias and the standard deviation (not reported here) decrease as n increases. Simulations show that for a fixed sample size n, as the bandwidth m increases, the standard deviation decreases, but overall the bias tends to increase which is in line with the theoretical results. In general, the standard deviation of the local Whittle estimator is on a similar level in linear and signal plus noise models but varies across α in ξ_t^2, $\exp(\xi_t)$ and EGARCH models. Various bias reduction methods for linear and signal plus noise models were discussed by Sun and Phillips (2003), Hurvich, Moulines and Soulier (2005) and Andrews and Sun (2004).

To conclude, both theoretical results and simulations suggest that the local Whittle estimator remains consistent also for nonlinear time series. However, the presence of a noise or nonlinearity worsens the behaviour of the estimator in finite samples and a larger sample size is needed to achieve a satisfactory accuracy. Although the choice of the optimal bandwidth parameter remains an open problem, for practical applications the simulation results suggest the use of $m = [n^{0.6}]$ for signal plus noise models, $m = [n^{0.7}] - [n^{0.8}]$ for nonlinear processes and $m = [n^{0.6}] - [n^{0.7}]$ for the EGARCH model.

Bias and M.S.E. of $\hat{\alpha}$, X_t is ARFIMA$(0, \alpha_0/2, 0)$

α_0	$n=1024$ $m=[n^{0.5}]$	$[n^{0.6}]$	$[n^{0.7}]$	$[n^{0.8}]$	$n=2048$ $[n^{0.5}]$	$[n^{0.6}]$	$[n^{0.7}]$	$[n^{0.8}]$
-0.8	0.041	0.021	0.015	0.026	0.031	0.017	0.013	0.021
	(0.038)	(0.019)	(0.009)	(0.005)	(0.027)	(0.012)	(0.006)	(0.003)
-0.4	-0.011	-0.006	-0.002	0.008	-0.007	-0.003	0.000	0.007
	(0.049)	(0.021)	(0.009)	(0.004)	(0.032)	(0.013)	(0.005)	(0.002)
0	-0.019	-0.012	-0.008	-0.004	-0.014	-0.007	-0.004	-0.002
	(0.050)	(0.021)	(0.009)	(0.004)	(0.032)	(0.012)	(0.005)	(0.002)
0.4	-0.016	-0.009	-0.008	-0.013	-0.011	-0.005	-0.004	-0.009
	(0.049)	(0.021)	(0.009)	(0.004)	(0.032)	(0.012)	(0.005)	(0.002)
0.8	-0.023	-0.005	-0.004	-0.017	-0.010	0.001	0.001	-0.012
	(0.038)	(0.018)	(0.009)	(0.005)	(0.027)	(0.012)	(0.005)	(0.003)

α_Y	α_Z	$n=1024$ $m=[n^{0.5}]$	$[n^{0.6}]$	$[n^{0.7}]$	$[n^{0.8}]$	$n=2048$ $[n^{0.5}]$	$[n^{0.6}]$	$[n^{0.7}]$	$[n^{0.8}]$
0	-0.8	-0.082	-0.105	-0.146	-0.197	-0.060	-0.083	-0.124	-0.179
		(0.056)	(0.032)	(0.031)	(0.043)	(0.035)	(0.019)	(0.021)	(0.034)
0	-0.4	-0.107	-0.114	-0.129	-0.143	-0.090	-0.102	-0.118	-0.137
		(0.060)	(0.034)	(0.026)	(0.025)	(0.039)	(0.022)	(0.019)	(0.021)
0.4	-0.8	-0.056	-0.086	-0.152	-0.222	-0.035	-0.061	-0.119	-0.222
		(0.053)	(0.028)	(0.032)	(0.068)	(0.033)	(0.016)	(0.019)	(0.051)
0.4	-0.4	-0.092	-0.121	-0.171	-0.234	-0.067	-0.097	-0.145	-0.213
		(0.057)	(0.035)	(0.038)	(0.059)	(0.036)	(0.022)	(0.026)	(0.047)
0.4	0	-0.115	-0.124	-0.142	-0.165	-0.097	-0.111	-0.130	-0.156
		(0.062)	(0.036)	(0.029)	(0.031)	(0.041)	(0.024)	(0.022)	(0.026)
0.8	-0.4	-0.092	-0.144	-0.251	-0.389	-0.055	-0.105	-0.204	-0.347
		(0.051)	(0.042)	(0.073)	(0.156)	(0.032)	(0.024)	(0.047)	(0.123)
0.8	0	-0.143	-0.188	-0.259	-0.341	-0.104	-0.155	-0.226	-0.315
		(0.065)	(0.056)	(0.076)	(0.121)	(0.041)	(0.037)	(0.057)	(0.101)
0.8	0.4	-0.152	-0.163	-0.187	-0.219	-0.130	-0.149	-0.175	-0.208
		(0.068)	(0.047)	(0.044)	(0.052)	(0.047)	(0.035)	(0.036)	(0.045)

Table 1: Bias and M.S.E. of $\hat{\alpha}_X$, Signal plus Noise Process $X_t = Y_t + Z_t$, Y_t is Gaussian ARFIMA$(0, \alpha_Y/2, 0)$, Z_t is Gaussian ARFIMA$(0, \alpha_Z/2, 0)$

		$n = 1024$				$n = 2048$			
α_X	α_ξ	$m = [n^{0.5}]$	$[n^{0.6}]$	$[n^{0.7}]$	$[n^{0.8}]$	$[n^{0.5}]$	$[n^{0.6}]$	$[n^{0.7}]$	$[n^{0.8}]$
0	-0.4	-0.134	-0.145	-0.157	-0.170	-0.115	-0.128	-0.144	-0.161
		(0.063)	(0.041)	(0.033)	(0.032)	(0.043)	(0.028)	(0.026)	(0.028)
0	0	-0.020	-0.013	-0.006	-0.004	-0.014	-0.007	-0.003	-0.002
		(0.047)	(0.020)	(0.008)	(0.004)	(0.031)	(0.012)	(0.005)	(0.002)
0.4	0.4	-0.096	-0.100	-0.106	-0.118	-0.082	-0.089	-0.098	-0.112
		(0.059)	(0.032)	(0.022)	(0.020)	(0.039)	(0.021)	(0.016)	(0.016)
0.8	0.8	-0.112	-0.108	-0.112	-0.131	-0.095	-0.095	-0.104	-0.122
		(0.056)	(0.036)	(0.026)	(0.025)	(0.041)	(0.025)	(0.019)	(0.019)

Table 2: Bias and M.S.E. of $\hat{\alpha}_X$, $X_t = \exp(\xi_t)$, ξ_t is Gaussian ARFIMA$(0, \alpha_\xi/2, 0)$

		$n = 1024$				$n = 2048$			
α_X	α_ξ	$m = [n^{0.5}]$	$[n^{0.6}]$	$[n^{0.7}]$	$[n^{0.8}]$	$[n^{0.5}]$	$[n^{0.6}]$	$[n^{0.7}]$	$[n^{0.8}]$
0	-0.4	-0.020	-0.010	0.002	0.017	-0.015	-0.006	0.002	0.015
		(0.048)	(0.020)	(0.009)	(0.004)	(0.032)	(0.012)	(0.005)	(0.002)
0	0	-0.022	-0.013	-0.006	-0.003	-0.015	-0.008	-0.004	-0.001
		(0.048)	(0.020)	(0.009)	(0.004)	(0.032)	(0.012)	(0.005)	(0.002)
0	0.4	0.030	0.049	0.069	0.085	0.032	0.051	0.068	0.085
		(0.051)	(0.025)	(0.016)	(0.013)	(0.033)	(0.016)	(0.011)	(0.010)
0.6	0.8	-0.133	-0.097	-0.070	-0.056	-0.115	-0.080	-0.055	-0.042
		(0.102)	(0.059)	(0.034)	(0.021)	(0.080)	(0.043)	(0.024)	(0.014)

Table 3: Bias and M.S.E. of $\hat{\alpha}_X$, $X_t = \xi_t^2$, ξ_t is Gaussian ARFIMA$(0, \alpha_\xi/2, 0)$

		$n = 1024$				$n = 2048$			
α_{r^2}	α_ξ	$m = [n^{0.5}]$	$[n^{0.6}]$	$[n^{0.7}]$	$[n^{0.8}]$	$[n^{0.5}]$	$[n^{0.6}]$	$[n^{0.7}]$	$[n^{0.8}]$
0	0	-0.011	-0.007	-0.003	-0.002	-0.012	-0.006	-0.003	-0.001
		(0.032)	(0.013)	(0.006)	(0.002)	(0.021)	(0.008)	(0.003)	(0.001)
0.4	0.4	-0.317	-0.322	-0.328	-0.336	-0.314	-0.321	-0.330	-0.338
		(0.139)	(0.122)	(0.118)	(0.119)	(0.126)	(0.115)	(0.115)	(0.118)
		-0.454	-0.489	-0.525	-0.567	-0.428	-0.469	-0.513	-0.558
0.8	0.8	(0.262)	(0.273)	(0.299)	(0.336)	(0.229)	(0.247)	(0.281)	(0.323)

Table 4: Bias and M.S.E. of $\hat{\alpha}_{r^2}$, $r_t = \varepsilon_t e^{\xi_t}$, ξ_t is Gaussian ARFIMA$(0, \alpha_\xi/2, 0)$, ε_t is Gaussian i.i.d.$(0, 1)$

$\alpha_{\log r^2}$	α_ξ	$n=1024$ $m=[n^{0.5}]$	$[n^{0.6}]$	$[n^{0.7}]$	$[n^{0.8}]$	$n=2048$ $[n^{0.5}]$	$[n^{0.6}]$	$[n^{0.7}]$	$[n^{0.8}]$
0	0	-0.018	-0.009	-0.006	-0.003	-0.016	-0.006	-0.003	-0.002
		(0.049)	(0.021)	(0.009)	(0.004)	(0.031)	(0.012)	(0.005)	(0.002)
0.4	0.4	-0.131	-0.143	-0.161	-0.184	-0.116	-0.128	-0.148	-0.175
		(0.066)	(0.042)	(0.035)	(0.038)	(0.045)	(0.029)	(0.027)	(0.033)
0.8	0.8	-0.165	-0.218	-0.293	-0.376	-0.126	-0.181	-0.258	-0.349
		(0.074)	(0.070)	(0.095)	(0.146)	(0.047)	(0.045)	(0.072)	(0.124)

Table 5: Bias and M.S.E. of $\hat{\alpha}_{\log r^2}$, $r_t = \varepsilon_t e^{\xi_t}$, ξ_t is Gaussian ARFIMA$(0, \alpha_\xi/2, 0)$, ε_t is Gaussian i.i.d.$(0,1)$

Acknowledgements

The research of Violetta Dalla and Javier Hidalgo was supported by the Economic and Social Research Council grant R000239936. The research of Liudas Giraitis was supported by the Economic and Social Research Council grant R000239538.

REFERENCES

Andrews, D.W.K. and Sun, Y. (2004) Adaptive polynomial Whittle estimation of long-range dependence. *Econometrica* **72,** 569-614.

Arteche, J. (2004) Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models. *Journal of Econometrics* **119**, 131-154.

Breidt, F. J., Crato, N. and de Lima, P. (1998) On the detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* **83,** 325–348.

Dahlhaus, R. (1989) Efficient parameter estimation for self similar processes. *Annals of Statistics* **17**(4), 1749–1766.

Davies, R.B. and Harte, D.S. (1987) Tests for Hurst effect. *Biometrika* **74**, 95–102.

Deo, R. and Hurvich, C. (2001) On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. *Econometric Theory* **17**, 686-710.

Dobrushin, R. L. and Major, P. (1979) Non-central limit theorems for non-linear functions of Gaussian fields. *Zeitschrift für Wahrscheinlichkeit Theorie und verwandte Gebiete* **50**, 27-52.

FOX, R. AND TAQQU, M.S. (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics* **14**, 517-532.

GEWEKE, J. AND PORTER-HUDAK, S. (1983) The estimation and application of long-memory time series models. *J. of Time Series Analysis* **4**, 221-238.

GIRAITIS, L. AND SURGAILIS, D. (1985) CLT and other limit theorems for functionals of Gaussian process. *Zeitschrift für Wahrscheinlichkeitstheorie and verwandte Gebiete* **70**, 191-212.

GIRAITIS, L. AND SURGAILIS, D. (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittle's estimate. *Probability Theory and Related Fields* **86**, 87–104.

HANNAN, E.J. (1973) The asymptotic theory of linear time series models. *J. Appl. Probab.* **10**, 130-145.

HARVEY, A.C. (1998) Long memory in stochastic volatility. In: *Forecasting Volatility in the Financial Markets*, *Oxford* (eds. J. Knight and S. Satchell), pp. 307–320. Butterworth & Heineman.

HARVEY, A.C., RUIZ, E. AND SHEPHARD, N. (1994) Multivariate stochastic variance models. *Review of Economic Studies* **61**, 247-264.

HENRY, M. AND ROBINSON P.M (1996) Bandwidth choice in Gaussian semiparametric estimation of long range dependence, in Robinson, P.M. and Rosenblatt, M. (eds.), *Athens Conference on Applied Probability and Time Series*, pp. 220-232, Springer Verlag, New York.

HURVICH, C.M. AND BRODSKY, J. (2001) Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential model. *J. Time Ser. Anal.* **22**, 221-249.

HURVICH, C.M., MOULINES, E. AND SOULIER, P. (2005) Estimation of long memory in volatility. *Econometrica*, **73**, 1283-1328.

KÜNSCH, H. (1987) Statistical aspects of self-similar processes, in Yu. A. Prohorov and V. V. Sazonov (eds.), *Proceedings of the 1st World Congress of the Bernoulli Society*, Vol.1, Science Press, Utrecht, 67–74.

LAHIRI, S.N. (2003) A necessary and sufficient condition for asymptotic independence of discrete Fourier transforms under short- and long-range dependence. *Annals of Statistics* **31**, 6132-641.

MOULINES, E. AND SOULIER, P. (1999) Broad band log-periodogram regression of time series with long range dependence. *Annals of Statistics* **27**, 1415-1439.

MOULINES, E. AND SOULIER, P. (2003) Semiparametric spectral estimation for fractional processes. In: *Theory and applications of long-range dependence*, (eds. P. Doukhan, G. Oppenheim and M.S. Taqqu), pp. 251–301. Birkhäuser, Boston.

Nelson, D. B. (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**, 347–370.

Robinson, P.M. (1995a) Log-periodogram regression of time series with long range dependence. *Annals of Statistics* **23**, 1048-1072.

Robinson, P.M. (1995b) Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* **23**, 1630-1661.

Robinson, P. M. (2001) The memory of stochastic volatility models. *Journal of Econometrics* **101**, 195–218.

Shimotsu, K. and Phillips, P. C. B. (2005) Exact local Whittle estimation of fractional integration. *Annals of Statistics* **32**, 656 -692.

Sun, Y. and Phillips, P.C.B. (2003) Nonlinear log-periodogram regression for perturbed fractional processes. *Journal of Econometrics* **115,** 355-389.

Surgailis, D. and Viano, M.-C. (2002) Long memory properties and covariance structure of the *EGARCH* model. *ESAIM: Probability and Statistics* **6**, 311–329.

Taqqu, M.S. (1979) Convergence of integrated processes of arbitrary Hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **50**, 53-83.

Taylor, S.J. (1994) Modelling stochastic volatility: A review and a comparative study. *Mathematical Finance* **4**, 183-204.

Velasco, C. and Robinson, P.M.(2000) Whittle pseudo-maximum likelihood estimates for non-stationary time series. *Journal of the American Statistical Association* **95**, 1229-1243.

Yong, C.H. (1974) *Asymptotic behaviour of Trigonometric Series*. The Chinese University of Hong Kong, Hong Kong.