Estimating long memory in volatility

Clifford M. Hurvich[*] Eric Moulines[†] Philippe Soulier[†‡]

Abstract

We consider semiparametric estimation of the memory parameter in a model which includes as special cases both the long-memory stochastic volatility (LMSV) and fractionally integrated exponential GARCH (FIEGARCH) models. Under our general model the logarithms of the squared returns can be decomposed into the sum of a long-memory signal and a white noise. We consider periodogram-based estimators which explicitly account for the noise term in a local Whittle criterion function. We allow the optional inclusion of an additional term to allow for a correlation between the signal and noise processes, as would occur in the FIEGARCH model. We also allow for potential nonstationarity in volatility, by allowing the signal process to have a memory parameter $d^* \geq 1/2$. We show that the local Whittle estimator is consistent for $d^* \in (0,1)$. We also show that a modified version of the local Whittle estimator is asymptotically normal for $d^* \in (0,3/4)$, and essentially recovers the optimal semiparametric rate of convergence for this problem. In particular if the spectral density of the short memory component of the signal is sufficiently smooth, a convergence rate of $n^{2/5-\delta}$ for $d^* \in (0,3/4)$ can be attained, where n is the sample size and $\delta > 0$ is arbitrarily small. This represents a strong improvement over the performance of existing semiparametric estimators of persistence in volatility. We also prove that the standard Gaussian semiparametric estimator is asymptotically normal if $d^* = 0$. This yields a test for long memory in volatility.

1 Introduction

There has been considerable recent interest in the semiparametric estimation of long memory in volatility. Perhaps the most widely used method for this purpose is the estimator (GPH) of Geweke and Porter-Hudak (1983). The GPH estimator of persistence in volatility is based on an ordinary linear regression of the log periodogram of a series that serves as a proxy for volatility, such as absolute returns, squared returns, or log squared returns of a financial time series. The single explanatory variable in the regression is log frequency, for Fourier frequencies in a neighborhood which degenerates towards zero frequency as the sample size n increases.

[*]New York University : 44 W. 4'th Street, New York NY 10012, USA
[†]École Nationale des Télécommunications, 46 rue Barrault, 75005 Paris, France
[‡]Département of Mathématiques, Université d'Evry Val d'Essonne, 91025 Evry cedex, France

Applications of GPH in the context of volatility have been presented in Andersen and Bollerslev (1997a,b), Ray and Tsay (2000), and Wright (2000), among others.

To derive theoretical results for semiparametric estimates of long memory in volatility, such as GPH, it is necessary to have a model for the series which incorporates some form of stochastic volatility. One particular such model is the long-memory stochastic volatility (LMSV) model of Harvey (1998) and Breidt, Crato and de Lima (1998). The LMSV model for a weakly stationary series of returns $\{r_t\}$ takes the form $r_t = \exp(Y_t/2)e_t$ where $\{e_t\}$ is a series of i.i.d. shocks with zero mean, and $\{Y_t\}$ is a weakly stationary linear long-memory process, independent of $\{e_t\}$, with memory parameter $d^* \in (0, 1/2)$. Under the LMSV model, the logarithms of the squared returns, $\{X_t\} = \{\log r_t^2\}$, may be expressed as

$$X_t = \mu + Y_t + \eta_t, \tag{1.1}$$

where $\mu = \mathbb{E}[\log e_t^2]$ and $\{\eta_t\} = \{\log e_t^2 - \mathbb{E}[\log e_t^2]\}$ is an i.i.d. process, independent of $\{Y_t\}$.

Another model for long memory in volatility is the fractionally integrated exponential GARCH (FIEGARCH) model of Bollerslev and Mikkelsen (1996). This model builds on the exponential GARCH (EGARCH) model of Nelson (1991). The weakly stationary FIEGARCH model takes the form $r_t = \sigma_t e_t$, where the $\{e_t\}$ are i.i.d. with zero mean and a symmetric distribution, and

$$\log \sigma_t^2 = \omega + \sum_{j=1}^{\infty} a_j g(e_{t-j}) \tag{1.2}$$

with $g(x) = \theta x + \gamma(|x| - \mathbb{E}|e_t|)$, $\omega > 0$, $\theta \in \mathbb{R}$, $\gamma \in \mathbb{R}$, and real constants a_j such that the process $\log \sigma_t^2$ has long memory with memory parameter $d^* \in (0, 1/2)$. If θ is nonzero, the model allows for a so-called leverage effect, whereby the sign of the current return may have some bearing on the future volatility. As was the case for the LMSV model, here we can once again express the log squared returns as in (1.1) with $\mu = \mathbb{E}[\log e_t^2] + \omega$, $\eta_t = \log e_t^2 - \mathbb{E}[\log e_t^2]$, and $Y_t = \log \sigma_t^2 - \omega$. Here, however, the processes $\{Y_t\}$ and $\{\eta_t\}$ are not independent of each other. In view of our goal of semiparametric estimation of d^*, we allow more generality in our specification of the weights a_j than Bollerslev and Mikkelsen (1996), who used weights corresponding to a fractional ARIMA model. As far as we are aware, no theoretical justification of any semiparametric estimator of d^* has heretofore been presented for the FIEGARCH model.

Assuming that the volatility series $\{Y_t\}$ is Gaussian, Deo and Hurvich (2001) derived asymptotic theory for the GPH estimator based on log squared returns in the LMSV model. This provides some justification for the use of GPH for estimating long memory in volatility. Nevertheless, it can also be seen from Theorem 1 of Deo and Hurvich (2001) that the presence of the noise term $\{\eta_t\}$ induces a negative bias in the GPH estimator, which in turn limits the number m of Fourier frequencies which can be used in the estimator while still guaranteeing $\sqrt{m}$-consistency and asymptotic normality. This upper bound, $m = o[n^{4d^*/(4d^*+1)}]$, becomes increasingly stringent as d^* approaches zero.

Recently, Hurvich and Ray (2001) have proposed a local Whittle estimator of d^*, once again based on log squared returns in the LMSV model. This estimator explicitly accounts for the noise term $\{\eta_t\}$ in (1.1). It was found in the simulation study of Hurvich and Ray (2001) that the local Whittle estimator can strongly outperform GPH, especially in terms of bias when m is large.

The local Whittle estimator, defined precisely in Section 1.1, may be viewed as a generalized version of the Gaussian semiparametric estimator (GSE) of Künsch (1987), which was studied by Robinson (1995b) under the assumption that the series of observations is linear in Martingale differences. We assume instead that we observe log squared returns $\{X_t\}$ which are the sum of a long-memory signal and a white noise. Our signal plus noise model, made precise in Section 1.1 below, includes both the LMSV and FIEGARCH models as special cases.

In the local Whittle estimator as originally proposed by Hurvich and Ray (2001), an additional term was included in the Whittle criterion function to account for the contribution of the noise term to the low frequency behavior of the spectral density of $\{X_t\}$. We will generalize this idea further by allowing the inclusion of one more term, as described below. The estimator is obtained from numerical optimization of the criterion function.

Many empirical studies have found estimates of the memory parameter in volatility, d^*, which are close to or even greater than 1/2, indicating possible nonstationarity of volatility. For example, Hurvich and Ray (2001) obtained a value of the local Whittle estimator $\hat{d}_n = 0.556$ for the log squared returns of a series of Deutsche Mark / US Dollar exchange rates with $n = 3485$ and $m = n^{0.8}$. In view of these empirical findings, we allow in this paper for the possibility that d^* exceeds 1/2. Specifically, we assume here that $d^* \in (0, 1)$.

As mentioned above, in the case of the FIEGARCH model the signal and noise processes will not be independent of each other. We allow (optionally) the addition of a term to the Whittle criterion to account for a contemporaneous correlation between the shocks in the signal and noise processes. This allows the FIEGARCH model to fit within our general framework.

In the context of our general signal plus noise model, allowing all of the generalizations described above, we will show that our local Whittle estimator $\hat{d}_n$ based on the first m Fourier frequencies is $\log^5(m)$-consistent. Using this result together with a modification to semiparametric Whittle-type estimators originally suggested for linear long-memory processes by Andrews and Sun (2001), we will establish the $\sqrt{m}$-consistency and asymptotic normality of a correspondingly-modified local Whittle estimator, $\hat{d}_n^*$, for $d^* \in (0, 3/4)$.

As long as the spectral density of the volatility (signal) process is sufficiently regular, our asymptotic results are free of upper restrictions on m arising from the presence of the noise term. In particular, if the spectral density of the short memory component of the signal is C^2, then we obtain asymptotic normality of $\sqrt{m}(\hat{d}_n^* - d^*)$ if $m = [n^\zeta]$ with $0 < \zeta < 4/5$. This represents a strong improvement over the GPH estimator of persistence in volatility.

Since we use the Whittle likelihood function we are able to avoid the assumption that the signal is Gaussian. This assumption was required by Deo and Hurvich (2001), but many practioners working with stochastic volatility models find the assumption to be overly restrictive.

The remainder of this paper is organized as follows. In Section 1.1, we define the local

Whittle estimator $\hat{d}_n$. Section 2 presents a theorem on the $\log^5(m)$ consistency of $\hat{d}_n$. Section 3 gives a central limit theorem for the modified estimator, $\hat{d}_n^*$. The estimates of the parameters (d^*, θ^*) converge at different rates, and in the case of the estimates of θ^* the rates may depend on d^*. Fortunately, however, the limiting covariance matrix of a suitably normalized vector of parameter estimates does not depend on θ^*. We present an expression, in terms of d^*, for the variance of the asymptotic distribution of $\sqrt{m}(\hat{d}_n^* - d^*)$. This expression takes a simple form when the signal and noise processes are known to be uncorrelated with each other. In Section 3.1, we prove that the standard GSE, without any of the additional terms considered in our local Whittle estimator, is asymptotically normal if $d^* = 0$. This yields a test for long memory in volatility.

1.1 The Local Whittle Estimator

Let X be a process with spectral density f_X that can be expressed as

$$f_X(x) = |1 - \mathrm{e}^{\mathrm{i}x}|^{-2d^*} f_X^*(x),$$

$d^* \in (0, 1/2)$, where f_X^* is a positive function, which is moreover smooth in a neighborhood of the origin. The GSE estimator of d^* consists in locally fitting a parametric model for f_X^* by minimizing the Whittle contrast function. Originally, the parametric model fitted replaces f_X^* by a constant. This method yields a consistent and asymptotically normal estimator of d^*, under mild assumptions both on f_X^* and the process X. Its rate of convergence is also known to be optimal under certain assumptions. In some situations however, this parameterization might be inefficient. An example is the situation of a long-memory process observed in an additive noise, in which case the rate of convergence of the GSE depends on d^* and is not optimal. In order to improve this rate of convergence, one can try to fit a more complex parametric model. Instead of replacing f_X^* by a constant in a neighborhood of 0, it is replaced by $G(1 + h(d, \theta, x))$, where (d, θ) belongs to the set of admissible parameters $\mathcal{D}_n \times \Theta_n$ which might depend on the sample size n. To be more precise, we introduce some notation. The discrete Fourier transform and the periodogram ordinates of any process U evaluated at the Fourier frequencies $x_j = 2j\pi/n$, $j = 1, \ldots, n$, are respectively denoted by

$$d_{U,j} = (2\pi n)^{-1/2} \sum_{t=1}^{n} U_t \mathrm{e}^{-\mathrm{i}tx_j}, \quad \text{and} \quad I_{U,j} = |d_{U,j}|^2.$$

The local Whittle contrast function is defined as

$$W_m(d, G, \theta) = \sum_{k=1}^{m} \left\{ \log\left(G x_k^{-2d}(1 + h(d, \theta, x_k) \right) + \frac{I_{X,k}}{G x_k^{-2d}(1 + h(d, \theta, x_k))} \right\} \tag{1.3}$$

where $m < n/2$ is a bandwidth parameter (the dependence on n is implicit). Concentrating G out of W_m yields the profile likelihood

$$\hat{J}_m(d, \theta) = \log\left(\frac{1}{m} \sum_{k=1}^{m} \frac{x_k^{2d} I_{X,k}}{1 + h(d, \theta, x_k)} \right) + m^{-1} \sum_{k=1}^{m} \log\{x_k^{-2d}(1 + h(d, \theta, x_k))\}. \tag{1.4}$$

The local Whittle estimates of d^* and θ^* are any minimand of the empirical contrast function $\hat{J}_m$ over a compact set:

$$(\hat{d}_n, \hat{\theta}_n) = \arg \min_{(d,\theta)\in\mathcal{D}_n\times\Theta_n} \hat{J}_m(d,\theta).$$

We generalize the model (1.1) to a signal plus noise situation, where the signal process Y exhibits long memory with memory parameter $d_Y \in (-1/2, 1/2)$, but the observed process X is either

$$X_t = \mu + Y_t + \eta_t, \tag{1.5}$$

$$\text{or } X_t = \mu + \sum_{s=1}^{t} Y_s + \eta_t, \tag{1.6}$$

according to whether X is stationary or nonstationary, where $(\eta_t)_{t\in\mathbb{Z}}$ is a zero mean white noise with variance σ_η^2. We assume moreover that Y admits an infinite order moving average representation with respect to a zero mean, unit variance white noise Z:

$$Y_t = \sum_{j\in\mathbb{Z}} a_j Z_{t-j}, \tag{1.7}$$

with $\sum_{j\in\mathbb{Z}} a_j^2 < \infty$, and for each t, η_t is independent of $\{Z_s, s \neq t\}$. We lose no generality in assuming that Y has zero mean, since the estimators considered in this paper are all functions of the periodogram at nonzero Fourier frequencies. In the nonstationary case, the assumption that Y has mean zero ensures that X is free of linear trends.

Define $a(x) = \sum_{j\in\mathbb{Z}} a_j \mathrm{e}^{\mathrm{i}jx}$. Having fractional differentiation in mind, we assume that a can be expressed for $x > 0$ as

$$a(x) = (1 - \mathrm{e}^{\mathrm{i}x})^{-d_Y} a^*(x),$$

with $d_Y \in (-1/2, 1/2)$ and for some function a^*, smooth in a neighborhood of 0. The spectral density of the process Y is then $f_Y = |a|^2/(2\pi)$, and it can be expressed as

$$f_Y(x) = |1 - \mathrm{e}^{\mathrm{i}x}|^{-2d_Y} f_Y^*(x), \tag{1.8}$$

with $f_Y^* = |a^*|^2/(2\pi)$. Define $U_t = \sum_{s=1}^{t} Y_s$ and $f_U(x) = |1 - e^{ix}|^{-2} f_Y(x)$. The function f_U is referred to as a pseudo spectral density of U. See, e.g., Solo (1992), Hurvich and Ray (1995), Velasco (1999).

We do not rule out the possibility that for each t, Z_t and η_t are correlated. More precisely we define

$$\rho_\eta = \mathbb{E}[\eta_t Z_t]/\sigma_\eta, \tag{1.9}$$

the correlation of Z and η and we assume that it is constant. One such example is the FIEGARCH model with standard Normal multiplying shocks, for which $\eta_t = \log(e_t^2) - \mathbb{E}[\log(e_t^2)]$, $Z_t = \theta e_t + \gamma(|e_t| - \sqrt{2/\pi})$, and $(e_t)_{t\in\mathbb{Z}}$ is i.i.d. $\mathcal{N}(0,1)$. Since we assume $\mathbb{E}[Z_t^2] = 1$, θ and γ are linked by the relation $\theta^2 + \gamma^2(1 - 2/\pi) = 1$. In that case, $\rho_\eta = \gamma \mathrm{cov}(|e_0|, \log(e_0^2))/\sigma_\eta$, where $\sigma_\eta^2 = \pi^2/2$.

In general, the spectral density or pseudo spectral density of the process X defined in (1.5) or (1.6) is then

$$f_X(x) = \begin{cases} f_Y(x) + \frac{2\rho_\eta\sigma_\eta}{2\pi}\mathrm{Re}(a(x)) + \frac{\sigma_\eta^2}{2\pi}, & \text{(stationary case)}, \\ f_U(x) + \frac{2\rho_\eta\sigma_\eta}{2\pi}\mathrm{Re}((1-\mathrm{e}^{\mathrm{i}x})^{-1}a(x)) + \frac{\sigma_\eta^2}{2\pi}, & \text{(non stationary case)}. \end{cases} \tag{1.10}$$

In both cases, f_X admits the following expansion at 0:

$$f_X(x) \sim x^{-2d^*} f_Y^*(0) + \mathrm{Re}\left((1-\mathrm{e}^{\mathrm{i}x})^{-d^*}\right) \frac{2\,\rho_\eta\sigma_\eta\sqrt{f_Y^*(0)}}{\sqrt{2\pi}} + \frac{\sigma_\eta^2}{2\pi},$$

with $d^* = d_Y$ in the stationary case and $d^* = d_Y + 1$ in the non stationary case.

In order to guarantee that the returns are a Martingale difference sequence, it is helpful to assume that $a_j = 0$ $(j \le 0)$ in the case where ρ_η is assumed to be nonzero. We do not make such an assumption here, in order to consider the problem in its fullest generality.

The local Whittle estimator, including a term accounting for ρ_η is obtained by taking $\mathcal{D}_n = [\epsilon_n, 1]$, where ϵ_n is a sequence that tends to zero as n tends to infinity and $\Theta = [-2T, 2T] \times [0, T^2]$ and

$$h(d,\theta,x) = \theta_1 x^{2d}\mathrm{Re}\left((1-\mathrm{e}^{\mathrm{i}x})^{-d}\right) + \theta_2 x^{2d}. \tag{1.11}$$

The "true values" of the parameters are then

$$\begin{gathered} d^* = d_Y \;\; \text{(stationary case)}, \;\; d^* = d_Y + 1 \text{ (non stationary case)} \\ \text{and } \theta^* = (\theta_1, \theta_2) \;\; \text{with} \;\; \theta_1^* = \frac{2\rho_\eta\sigma_\eta}{\sqrt{2\pi f_Y^*(0)}} \quad \text{and} \quad \theta_2^* = \frac{\sigma_\eta^2}{2\pi f_Y^*(0)}. \end{gathered}$$

Note that $d^* \in (0,1)$ implies that $d_Y \in (0, 1/2)$ in the stationary case and $d_Y \in (-1/2, 0)$ in the non stationary case. In the case where ρ_η is known to be zero, we would use simply

$$h(d,\theta,x) = \theta x^{2d}, \tag{1.12}$$

where the "true" values of the parameters are d^*, $\theta^* = \theta_2^*$ as given above, and $\Theta = [0, T]$.

2 Consistency of the local Whittle estimator

Our results will be derived under regularity conditions on the function $a(x) = \sum_{j\in\mathbb{Z}} a_j \mathrm{e}^{\mathrm{i}jx}$. We introduce the following functional class:

Definition 1. *For $\alpha \in (0, \pi]$, $\beta > 0$ and $0 < \mu < \infty$, $\mathcal{F}(\alpha, \beta, \mu)$ is the set of functions g defined on $[-\pi, \pi]$ satisfying $\int_{-\pi}^{\pi} |g(x)| dx \le \mu$ and for all $x \in [-\alpha, \alpha]$,*

$$|g(x)| \le \mu |x|^\beta. \tag{2.1}$$

We now introduce our assumptions.

(A1) $Z = (Z_t)_{t\in\mathbb{Z}}$ is a zero mean unit variance white noise such that

$$\frac{1}{n}\sum_{t=1}^{n}(Z_t^2 - 1) \xrightarrow{P} 0 \tag{2.2}$$

and for any $(s,t,u,v) \in \mathbb{N}^4$ such that $s < t$ and $u < v$, $Z_u Z_v Z_s Z_t$ is integrable and

$$\mathbb{E}[Z_u Z_v Z_s Z_t] = \begin{cases} 1 & \text{if } u = s \text{ and } t = v \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

Remark 2.1. This assumption is implied by assumption A3 of Robinson (1995b) which states that Z is a martingale difference sequence satisfying moreover $\mathbb{E}[Z_t^2|\sigma(Z_s, s < t)] = 1$ a.s. (which implies (2.3)) and strongly uniformly integrable (which implies (2.2)).

(A2) There exists a real number $\varsigma > 0$ such that

$$\frac{1}{n}\sum_{t=1}^{n}(Z_t^2 - 1) = O_P(n^{-\varsigma}),$$

For reference, we recall the assumption on η.

(A3) $\eta = (\eta_t)_{t\in\mathbb{Z}}$ is a white noise with variance σ_η such that for each t, η_t is independent of $\{Z_s, s \neq t\}$ and for each t, we define $\rho_\eta = \mathbb{E}[\eta_t Z_t]/\sigma_\eta$.

(A4) Y admits the linear representation (1.7) and the function $a(x) = \sum_{j\in\mathbb{Z}} a_j \mathrm{e}^{\mathrm{i}jx}$ can be expressed for $x > 0$ as $a(x) = (1 - \mathrm{e}^{\mathrm{i}x})^{-d_Y} a^*(x)$ $(x > 0)$, where $\big(a^*(0)^{-1}a^* - 1\big) \in \mathcal{F}(\alpha, \beta, \mu)$ for some $\alpha \in (0, \pi]$, $\beta \in (0, 2]$ and $\mu > 0$.

Theorem 2.1. *Assume* **(A1)**, **(A3)** *and* **(A4)**. *Let m be a non decreasing sequence such that*

$$\lim_{n\to\infty} \big(m^{-1} + m/n\big) = 0. \tag{2.4}$$

Define $\epsilon_n = (\log(n/m))^{-1/2}$. Then $\hat{d}_n$ is a consistent estimate of d^*. If moreover **(A2)** *holds and the sequence m satisfies*

$$\lim_{n\to\infty} \log^{2s}(m)\mathrm{e}^{-\sqrt{\log(n/m)}} = 0, \tag{2.5}$$

for some positive integer s, then $\hat{d}_n - d^* = o_P(\log^{-s}(m))$.

Remark 2.2. It must be noted that we only prove consistency for $\hat{d}_n$ but not of $\hat{\theta}_n$. In order to prove asymptotic normality of $\hat{d}_n$, we will need to modify the definition of the estimator of d^*.

Remark 2.3. Assumption (2.5) holds as long as $m \leq n^\delta$ for some $\delta < 1$.

Proof of Theorem 2.1. Define $\mathcal{D}_1 = (-\infty, d^* - 1/2 + \epsilon) \cap \mathcal{D}_n$ and $\mathcal{D}_2 = [d^* - 1/2 + \epsilon, +\infty) \cap \mathcal{D}_n$, for some small positive real ϵ to be set later. As originally done in Robinson (1995b), we separately prove that $\lim_{n\to\infty} \mathbb{P}(\hat{d}_n \in \mathcal{D}_1) = 0$ and that $(\hat{d}_n - d^*)\mathbf{1}_{\mathcal{D}_2}(\hat{d}_n)$ tend to zero in probability. Note that $\mathcal{D}_1$ is empty if it is assumed that $d^* \in (0, 1/2)$ and ϵ is chosen small enough. In case $\mathcal{D}_1$ is not empty, the proof of $\lim_{n\to\infty} \mathbb{P}(\hat{d}_n \in \mathcal{D}_1) = 0$ is a straightforward adaptation of the proof of Robinson (1995b, pp. 1638-1639). For the sake of completeness, we provide a proof in the Appendix. We now prove that $\lim_{n\to\infty} \mathbb{P}(\hat{d}_n \to d^*, \hat{d}_n \in \mathcal{D}_2) = 1$. Denote

$$
\begin{aligned}
&\alpha_k(d,\theta) = \frac{1 + h(d^*, \theta^*, x_k)}{1 + h(d, \theta, x_k)}, \quad K_m(s) = \log\left(\frac{1}{m}\sum_{k=1}^m k^{2s}\right) - \frac{2s}{m}\sum_{k=1}^m \log(k), \\
&J_m(d,\theta) = \log\left(\frac{1}{m}\sum_{k=1}^m x_k^{2d-2d^*}\alpha_k(d,\theta)\right) - \frac{1}{m}\sum_{k=1}^m \log\left(x_k^{-2d}\{1 + h(d,\theta,x_k)\}\right), \\
&R_m(d,\theta) = J_m(d,\theta) - J_m(d^*,\theta^*) - K_m(d - d^*) \text{ and} \\
&E_n(d,\theta) = \frac{\sum_{k=1}^m k^{2d-2d^*}\alpha_k(d,\theta)\{x_k^{2d^*} I_{X,k}/(f_X^*(0)(1 + h(d^*,\theta^*,x_k)) - 1\}}{\sum_{j=1}^m j^{2d-2d^*}\alpha_j(d,\theta)}.
\end{aligned}
$$

With this notation, we get

$$
\hat{J}_m(d,\theta) = \log(1 + E_n(d,\theta)) + J_m(d,\theta) + \log(f_X^*(0)). \tag{2.6}
$$

Due to the strict concavity of the log function, (d^*, θ^*) minimizes J_m and, by definition, $(\hat{d}_n, \hat{\theta}_n)$ minimizes $\hat{J}_m$. Hence we have

$$
\begin{aligned}
0 &\le J_m(\hat{d}_n, \hat{\theta}_n) - J_m(d^*, \theta^*) \\
&= J_m(\hat{d}_n, \hat{\theta}_n) - \hat{J}_m(\hat{d}_n, \hat{\theta}_n) + \hat{J}_m(\hat{d}_n, \hat{\theta}_n) - \hat{J}_m(d^*, \theta^*) + \hat{J}_m(d^*, \theta^*) - J_m(d^*, \theta^*) \\
&= \log(1 + E_n(d^*, \theta^*)) - \log(1 + E_n(\hat{d}_n, \hat{\theta}_n)) + \hat{J}_m(\hat{d}_n, \hat{\theta}_n) - \hat{J}_m(d^*, \theta^*) \\
&\le \log(1 + E_n(d^*, \theta^*)) - \log(1 + E_n(\hat{d}_n, \hat{\theta}_n)) \\
&\le 2 \sup_{(d,\theta)\in\mathcal{D}_2\times\Theta} |\log(1 + E_n(d,\theta))|.
\end{aligned}
$$

Proposition 2.1 below states that E_n converges in probability to zero, uniformly with respect to $(d,\theta) \in \mathcal{D}_2 \times \Theta$. Thus we obtain that $J_m(\hat{d}_n, \hat{\theta}_n) - J_m(d^*, \theta^*)$ converges in probability to 0. Note now that K_m converges uniformly on compact sets of $(-1, +\infty)$ to the function $K(s) = 2s - \log(1 + 2s)$. Hence it can be bounded below uniformly with respect to m: there exists a constant $c > 0$ which depends only on $\mathcal{D}$ such that for all $m \ge 2$ and $d \in \mathcal{D}$,

$$
K_m(d - d^*) \ge c(d - d^*)^2. \tag{2.7}
$$

Hence

$$
0 \le c(\hat{d}_n - d^*)^2 \le K_m(\hat{d}_n - d^*) \le J_m(\hat{d}_n, \hat{\theta}_n) - J_m(d^*, \theta^*) - R_m(\hat{d}_n, \hat{\theta}_n).
$$

To conclude the proof of the consistency of $\hat{d}_n$, we need only prove that $R_m(d,\theta)$ converges to zero in probability uniformly with respect to $(d,\theta) \in \mathcal{D} \times \Theta)$. By definition of h and $\mathcal{D}_n$, we first obtain a bound for $\alpha_k(d,\theta) - 1$:

$$\sup_{d \in \mathcal{D}_n, \theta \in \Theta} |\alpha_k(d,\theta) - 1| \le C \mathrm{e}^{-\sqrt{\log(n/m)}}.$$

To bound R_m, note that it can be expressed as

$$R_m(d,\theta) = \log\left(\frac{\sum_{k=1}^m k^{2d-2d^*} \alpha_k(d,\theta)}{\sum_{j=1}^m j^{2d-2d^*}}\right) - \frac{1}{m}\sum_{k=1}^m \log\left(\alpha_k(d,\theta)\right).$$

Hence we obtain

$$\sup_{(d,\theta) \in \mathcal{D}_n \times \Theta} |R_m(d,\theta)| \le C \mathrm{e}^{-\sqrt{\log(n/m)}}. \tag{2.8}$$

We now prove the second part of Theorem 2.1. For any positive real A and any positive integer m, define $\mathcal{D}_{A,m} = \{d \in \mathcal{D} : \ 2\log^5(m)|d - d^*| > A\}$. We want to prove that

$$\limsup_{n\to\infty} \mathbb{P}(\hat{d}_n \in \mathcal{D}_{A,m}) = 0.$$

Since for large enough m, $\mathcal{D}_1 \subset \mathcal{D}_{A,m}$, and since we already know that $\lim \mathbb{P}(\hat{d}_n \in \mathcal{D}_1) = 0$, we can restrict our attention to $\tilde{\mathcal{D}}_{A,m} = \mathcal{D}_{A,m} \cap \mathcal{D}_2$.

Since $(\hat{d}_n, \hat{\theta}_n)$ minimizes $\hat{J}_m$, it holds that

$$\begin{aligned}
\mathbb{P}(\hat{d}_n \in \tilde{\mathcal{D}}_{A,m}) &\le \mathbb{P}(\inf_{\theta\in\Theta} \inf_{d \in \tilde{\mathcal{D}}_{A,m,s}} \{\hat{J}_m(d,\theta) - \hat{J}_m(d^*,\theta^*)\} \le 0) \\
&= \mathbb{P}(\inf_{\theta\in\Theta} \inf_{d \in \tilde{\mathcal{D}}_{A,m}} \{J_m(d,\theta) - J_m(d^*,\theta^*) + \log(1+E_n(d,\theta)) - \log(1+E_n(d^*,\theta^*))\} \le 0) \\
&\le \mathbb{P}(\inf_{\theta\in\Theta} \inf_{d \in \tilde{\mathcal{D}}_{A,m}} \{K_m(d-d^*) + R_m(d,\theta) + \log(1+E_n(d,\theta)) - \log(1+E_n(d^*,\theta^*))\} \le 0).
\end{aligned}$$

Since $x \to K_m(x)$ is strictly convex, applying (2.7), for large enough m, yields

$$\inf_{d \in \mathcal{D}_{A,m}} K_m(2(d-d^*)) = K_m(A\log^{-s}(m)) \wedge K_m(-A\log^{-s}(m)) \ge cA^2\log^{-10}(m).$$

Hence

$$\mathbb{P}(\hat{d}_n \in \tilde{\mathcal{D}}_{A,m}) \le \mathbb{P}(\sup_{\theta\in\Theta} \sup_{d \in \tilde{\mathcal{D}}_{A,m}} (|\log(1+E_n(d,\theta))| + |R_m(d,\theta)|) \ge cA^2\log^{-10}(m)).$$

The proof of Theorem 2.1 is concluded by applying (2.5), (2.8) and Proposition 2.1. □

Proposition 2.1. *Assume* **(A1)**, **(A3)** *and* **(A4)**. *Then* $\sup_{(d,\theta)\in\mathcal{D}_2\times\Theta} |E_n(d,\theta)| = o_P(1)$. *If moreover* **(A2)** *holds, then, there exists a positive real number* η *such that*

$$\sup_{(d,\theta)\in\mathcal{D}_2\times\Theta} |E_n(d,\theta)| = O_P\left((m/n)^\eta + m^{-\eta}\right),$$

Proof. Denote

$$\gamma_k(d,\theta) = \frac{\alpha_k(d,\theta)k^{2d-2d^*}}{\sum_{j=1}^m \alpha_j(d,\theta)j^{2d-2d^*}},$$

$$r_k = \frac{I_{X,k}}{x_k^{-2d^*} f_Y^*(0)(1+h(d^*,\theta^*,x_k))} - 2\pi I_{Z,k}.$$

Then

$$E_n(d,\theta) = \sum_{k=1}^m \gamma_k(d,\theta) r_k + \sum_{k=1}^m \gamma_k(d,\theta)(2\pi I_{Z,k} - 1) =: E_{1,n}(d,\theta) + E_{2,n}(d,\theta).$$

Since h is uniformly bounded on $(0,1)\times\Theta\times[0,2\pi m/n]$, we obtain

$$|E_{1,n}(d,\theta)| \le \frac{C\sum_{k=1}^m k^{2d-2d^*}|r_k|}{\sum_{j=1}^m j^{2d-2d^*}}.$$

Applying (4.21) in Theorem 4.1, we obtain, for some $\gamma > 0$ and $C > 0$:

$$\mathbb{E}[|r_k|] \le C(k^{-\gamma} + (k/n)^\gamma).$$

If $d \in \mathcal{D}_2$, there exists a constant $c(\epsilon)$ such that

$$\sum_{j=1}^m j^{2d-2d^*} \ge c(\epsilon) m^{2d-2d^*+1}.$$

Without loss of generality, we assume that $2\epsilon < \gamma$. Thus, we obtain:

$$\begin{aligned}
\mathbb{E}\left[\sup_{(d,\theta)\in\mathcal{D}_2\times\Theta} |E_{1,n}(d,\theta)|\right] &\le c\sum_{k=1}^{m-1} \sup_{d\in\mathcal{D}_2} \frac{|k^{2d-2d^*} - (k+1)^{2d-2d^*}|}{m^{2d-2d^*+1}} \sum_{j=1}^k \mathbb{E}[|r_j|] + cm^{-1}\sum_{j=1}^m \mathbb{E}[|r_j|] \\
&\le c\sum_{k=1}^{m-1} \sup_{d\in\mathcal{D}_2} \frac{k^{2d-2d^*-1}}{m^{2d-2d^*+1}} \sum_{j=1}^k \mathbb{E}[|r_j|] + cm^{-1}\sum_{j=1}^m \mathbb{E}[|r_j|] \\
&\le c\sum_{k=1}^{m-1} (k/m)^{2\epsilon} k^{-2} \sum_{j=1}^k (j^{-\gamma} + (j/n)^\gamma) + cm^{-1}\sum_{j=1}^m (j^{-\gamma} + (j/n)^\gamma) \\
&\le cm^{-2\epsilon}\sum_{k=1}^{m-1} k^{-2+2\epsilon}\left(k^{1-\gamma} + n^{-\gamma}k^{\gamma+1}\right) + cm^{-\gamma} + c(m/n)^\gamma \\
&\le c\left(m^{-2\epsilon} + (m/n)^{2\epsilon}\right).
\end{aligned}$$

Write now $2\pi I_{Z,k} - 1 = n^{-1}\sum_{t=1}^n (Z_t^2 - 1) + 2n^{-1}\sum_{1\le s<t\le n} \cos\{(s-t)x_k\} Z_s Z_t$ and

$$\begin{aligned}
E_{2,n}(d,\theta) &= n^{-1}\sum_{t=1}^n (Z_t^2 - 1) + 2n^{-1}\sum_{k=1}^{m-1} \gamma_k(d,\theta) \sum_{1\le s<t\le n} \cos\{(s-t)x_k\} Z_s Z_t \\
&=: E_{2,1,n} + E_{2,2,n}(d,\theta).
\end{aligned}$$

Under assumption **(A1)**, $E_{2,1,n} = o_P(1)$ and under **(A2)**, $E_{2,1,n} = O_p(n^{-\varsigma}) = o_P(m^{-\varsigma})$. Consider now $E_{2,2,n}$. Applying Robinson (1995b) Eq. (3.20), we have

$$\mathbb{E}\left[\left(n^{-1}\sum_{1\le s<t\le n}\sum_{j=1}^{k}\cos\{(s-t)x_j\}Z_sZ_t\right)^2\right] \le k.$$

Hence, applying again summation by parts, we get

$$\begin{aligned}\mathbb{E}[\sup_{d\in\mathcal{D}_2\times\Theta}|E_{2,2,n}|] &\le c\sum_{k=1}^{m-1}\sup_{(d,\theta)\in\mathcal{D}_2\times\Theta}\frac{\left|(k+1)^{2d-2d^*}\alpha_{k+1}(d,\theta)-k^{2d-2d^*}\alpha_k(d,\theta)\right|}{m^{2d-2d^*+1}}\sqrt{k}\\ &\le c\sum_{k=1}^{m-1}\sup_{d\in\mathcal{D}_2\times}\frac{k^{2d-2d^*-1}}{m^{2d-2d^*+1}}\sqrt{k}\le cm^{-2\epsilon}.\end{aligned}$$

Hence Proposition 2.1 holds with $\eta = \varsigma\wedge(2\epsilon)$. □

3 Asymptotic normality of the modified local Whittle estimator

The usual method for proving asymptotic normality of a consistent minimum contrast estimate is to make a second order Taylor expansion of the contrast function and to say that the gradient of the contrast function evaluated at the estimates vanishes, since it is consistent and the true value is assumed to be an interior point of the parameter set. In the present context, we have only proved the consistency of $\hat{d}_n$, but not that of $\hat{\theta}_n$. Hence we cannot use this argument. Instead we will modify the definition of the estimator, following Andrews and Sun (2001).

Define $(\hat{d}_n^*, \hat{\theta}_n^*)$ as a solution in $\mathcal{D}_n\times\Theta$ of $\nabla\hat{J}_m(d,\theta)=0$, (where ∇ denotes the differentiation with respect to d and θ), if there exists one, and if there are multiple solutions, choose the one closest (in the sense of any norm) to $(\hat{d}_n,\hat{\theta}_n)$. If there are no solutions, set $(\hat{d}_n^*,\hat{\theta}_n^*) = (\hat{d}_n,\hat{\theta}_n)$.

The first step in establishing the consistency and asymptotic normality of $(\hat{d}_n^*,\hat{\theta}_n^*)$ remains as usual to study the behavior of the gradient and Hessian of the contrast function $\hat{J}_m$. For this we must strengthen the assumptions on the noise sequences Z and η.

(A5) $(Z_t)_{t\in\mathbb{Z}}$ is a martingale difference sequence such that for all t, $\mathbb{E}[Z_t^4] := \mu_4 < \infty$ and $\mathbb{E}[Z_t^2 \mid Z_s, s<t] = 1$ a.s. (*i.e.* Z_t^2-1 is a square integrable martingale sequence).

Remark 3.1. **(A5)** implies **(A1)** and **(A2)** with $\varsigma = 1/2$.

(A6) $(\eta_t)_{t\in\mathbb{N}}$ is a zero mean white noise such that $\sup_{t\in\mathbb{N}}\mathbb{E}[\eta_t^4]<\infty$, a.s. and for all $(s,t,u,v)\in\mathbb{N}^4$ such that $s<t$ and $u<v$,

$$\mathbb{E}[\eta_u\eta_v\eta_s\eta_t] = \begin{cases}\sigma_\eta^2 & \text{if } u=s \text{ and } t=v\\ 0 & \text{otherwise.}\end{cases} \tag{3.1}$$

$$\mathrm{cum}(Z_{t_1},Z_{t_2},\eta_{t_3},\eta_{t_4}) = \begin{cases}\kappa & \text{if } t_1=t_2=t_3=t_4,\\ 0 & \text{otherwise}\end{cases} \tag{3.2}$$

Denote $\hat{J}_m^* = \hat{J}_m(d^*, \theta^*)$ and $H_n(d,\theta) = \int_0^1 \nabla^2 \hat{J}_m(d^* + s(d-d^*), \theta^* + s(\theta - \theta^*))ds$.

Proposition 3.1. *Assume* **(A3)**, **(A4)**, **(A5)** *and* **(A6)**. *If $d^* \in (0, 3/4)$, $\beta > 2d^*$ and m is a non decreasing sequence of integers such that*

$$\lim_{n\to\infty} \left(m^{-4d^*-1} n^{4d^*} + n^{-2\beta} m^{2\beta+1} \log^2(m) \right) = 0, \tag{3.3}$$

then $m D_n^{*-1} \nabla \hat{J}_m^*$ converges to the Gaussian distribution with zero mean and variance Γ^* with
(i) $D_n^* = m^{1/2} \mathrm{Diag}\left(1, (m/n)^{2d^*}\right)$ and

$$\Gamma^* = \begin{pmatrix} 4 & -\frac{4d^*(2\pi)^{2d^*}}{(1+2d^*)^2} \\ -\frac{4d^*(2\pi)^{2d^*}}{(1+2d^*)^2} & \frac{4d^{*2}(2\pi)^{4d^*}}{(1+2d^*)^2(1+4d^*)} \end{pmatrix}$$

if ρ_η is known to be 0;
(ii) $D_n^* = m^{1/2} \mathrm{Diag}\left(1, (m/n)^{d^*}, (m/n)^{2d^*}\right)$ and

$$\Gamma^* = \begin{pmatrix} 4 & -\frac{2d^*(2\pi)^{d^*}}{(1+d^*)^2} & -\frac{4d^*(2\pi)^{2d^*}}{(1+2d^*)^2} \\ -\frac{2d^*(2\pi)^{d^*}}{(1+d^*)^2} & \frac{2d^{*2}(2\pi)^{2d^*}}{(1+d^*)^2(1+2d^*)} & \frac{2d^{*2}(2\pi)^{3d^*}}{(1+d^*)(1+2d^*)(1+3d^*)} \\ -\frac{4d^*(2\pi)^{2d^*}}{(1+2d^*)^2} & \frac{2d^{*2}(2\pi)^{3d^*}}{(1+d^*)(1+2d^*)(1+3d^*)} & \frac{4d^{*2}(2\pi)^{4d^*}}{(1+2d^*)^2(1+4d^*)} \end{pmatrix}$$

if ρ_η is not known.

Proposition 3.2. *Assume* **(A3)**, **(A4)** *and* **(A5)**. *If $d^* \in (0, 3/4)$, $\beta > 2d^*$ and m is a non decreasing sequence of integers that satisfies (3.3), then $m D_n^{-1*} H_n(d,\theta) D_n^{-1*}$ converges in probability to Γ^*, uniformly with respect to $(d, \theta) \in \{d;\ |d - d^*| \le C \log^{-5}(m)\} \times \Theta$, with D_n^* and Γ^* defined as in Proposition 3.1.*

Remark 3.2. The first term in (3.3) imposes a lower bound on the allowable value of m, requiring that m tend to ∞ faster than $n^{4d^*/(4d^*+1)}$. This lower bound is necessary to ensure that all the elements of the matrix D_n^* tend to infinity. This condition can be fulfilled since by assumption $\beta > 2d^*$. In the case $\beta \le 2d^*$, the standard GSE will achieve the optimal rate of convergence and the present construction is then useless. Note that $\beta > 2d^*$ holds for $\beta = 2$, which is the most commonly accepted value for β. It is interesting that Deo and Hurvich (2001), assuming $\beta = 2$, found that for $m^{1/2}(\hat{d}_{GPH} - d^*)$ to be asymptotically normal with mean zero, where $\hat{d}_{GPH}$ is the GPH estimator, the bandwidth m must tend to ∞ at a rate *slower* than $n^{4d^*/(4d^*+1)}$.

Remark 3.3. Note that Proposition 3.2 holds under a weaker assumption on the noise η than Proposition 3.1: it is only assumed that the second moment of η is finite.

Remark 3.4. The assumption $a^*(0)^{-1} a^* - 1 \in \mathcal{F}(\alpha, \beta, \mu)$ is used to validate the Bartlett approximation. In the related literature (Robinson (1995b), Velasco (1999), Andrews and Sun (2001)), it is usually assumed moreoover that $a(x) = x^{-d} a^*(x)$ is differentiable in a neighborhood of zero, except at zero, with $x a'(x)$ bounded. Hence our assumptions are weaker than those of the above references.

Remark 3.5. An important feature is that Γ^* does not depend on the true value of the parameter θ^*. This was already noticed by Andrews and Sun (2001) in the context of local polynomial approximation.

Let now $(\tilde{d}_n, \tilde{\theta}_n)$ be a sequence of solutions of $\nabla \hat{J}_m(d, \theta) = 0$. A first order Taylor expansion yields

$$0 = {D_n^*}^{-1} \nabla \hat{J}_m(\tilde{d}_n, \tilde{\theta}_n) = {D_n^*}^{-1} \nabla \hat{J}_m(d^*, \theta^*) + {D_n^*}^{-1} H_n(\tilde{d}_n, \tilde{\theta}_n) {D_n^*}^{-1} D_n^* \left((\tilde{d}_n, \tilde{\theta}_n) - (d^*, \theta^*) \right).$$

As a consequence of Propositions 3.1 and 3.2, we trivially obtain the following corollary.

Corollary 3.1. *Under the assumptions of Proposition 3.1, if $(\tilde{d}_n, \tilde{\theta}_n)$ is a sequence of solutions of $\nabla \hat{J}_m(d, \theta) = 0$ such that $\tilde{d}_n$ is $\log^5(m)$ consistent, then $D_n^* \left((\tilde{d}_n, \tilde{\theta}_n) - (d^*, \theta^*) \right)$ is asymptotically Gaussian with zero mean and covariance matrix ${\Gamma^*}^{-1}$.*

We give explicit expressions of ${\Gamma^*}^{-1}$. If ρ_η is assumed to be zero, we obtain

$${\Gamma^*}^{-1} = \frac{(1+2d^*)^2}{16{d^*}^2} \begin{pmatrix} 1 & \frac{1+4d^*}{d^*(2\pi)^{2d^*}} \\ \frac{1+4d^*}{d^*(2\pi)^{2d^*}} & \frac{(1+2d^*)^2(1+4d^*)}{{d^*}^2(2\pi)^{4d^*}} \end{pmatrix}.$$

If ρ_η is not assumed to be zero, we obtain

$$\Gamma^{-1} = \frac{1}{16{d^*}^4} \begin{pmatrix} -1 & 0 & 0 \\ 0 & \frac{2(1+d^*)}{d^*(2\pi)^{d^*}} & 0 \\ 0 & 0 & \frac{1+2d^*}{2d^*(2\pi)^{2d^*}} \end{pmatrix}$$
$$\times \begin{pmatrix} (1+d^*)^2(1+2d^*)^2 & -2(1+d^*)(1+2d^*)^2(1+3d^*) & (1+d^*)(1+2d^*)(1+3d^*)(1+4d^*) \\ -2(1+d^*)(1+2d^*)^2(1+3d^*) & 4(1+d^*)^2(1+2d^*)(1+3d^*)^2 & -2(1+d^*)(1+2d^*)^2(1+3d^*)(1+4d^*) \\ (1+d^*)(1+2d^*)(1+3d^*)(1+4d^*) & -2(1+d^*)(1+2d^*)^2(1+3d^*)(1+4d^*) & (1+2d^*)^2(1+3d^*)^2(1+4d^*) \end{pmatrix}$$
$$\times \begin{pmatrix} -1 & 0 & 0 \\ 0 & \frac{2(1+d^*)}{d^*(2\pi)^{d^*}} & 0 \\ 0 & 0 & \frac{1+2d^*}{2d^*(2\pi)^{2d^*}} \end{pmatrix}.$$

We are now in a position to prove that $\nabla \hat{J}_m(\hat{d}_n^*, \hat{\theta}_n^*) = 0$, and $\hat{d}_n^*$ is $\log^5(m)$ consistent.

Proposition 3.3. *Under the assumptions of Proposition 3.1, $\nabla \hat{J}_m(\hat{d}_n^*, \hat{\theta}_n^*) = 0$ with probability tending to one and $\hat{d}_n^*$ is $\log^5(m)$ consistent.*

Proof. Applying Lemma 1 of Andrews and Sun (2001), (with, in their notation, $L_n = m\hat{J}_m$, $B_n = D_n^*$ and $K_n = m^{1/2}\log^5(m)$) we know that there exists a sequence $(\tilde{d}_n, \tilde{\theta}_n)$ such that $\mathbb{P}\left(\nabla \hat{J}_m(\tilde{d}_n, \tilde{\theta}_n) = 0\right) \to 1$ and $\tilde{d}_n$ is $\log^5(m)$ consistent. This implies that $(\hat{d}_n^*, \hat{\theta}_n^*)$ also shares these properties. Indeed, by definition, since there exists a solution of $\nabla \hat{J}_m(d, \theta) = 0$, with probability tending to one, $\nabla \hat{J}_m(\hat{d}_n^*, \hat{\theta}_n^*) = 0$. Since we know from section 2 that $\hat{d}_n$ is $\log^5(m)$ consistent, and since by definition $(\hat{d}_n^*, \hat{\theta}_n^*)$ is the closest solution to $(\hat{d}_n, \hat{\theta}_n)$ of $\nabla \hat{J}_m(d, \theta) = 0$, then $\hat{d}_n^*$ must also be $\log^5(m)$ consistent. $\square$

Proposition 3.3 and Corollary 3.1 yield the asymptotic normality of $\hat{d}_n^*$.

Theorem 3.1. *Assume* **(A3)**, **(A4)**, **(A5)** *and* **(A6)**. *If $d^* \in (0, 3/4)$, $\beta > 2d^*$ and m is a non decreasing sequence of integers that satisfies (3.3), then $m^{1/2}(\hat{d}_n^* - d^*)$ is asymptotically Gaussian with zero mean and variance*

$$\frac{(1+d^*)^2(1+2d^*)^2}{16{d^*}^4}.$$

If ρ_η is known to be 0, then $m^{1/2}(\hat{d}_n^* - d^*)$ is asymptotically Gaussian with zero mean and variance $(1+2d^*)^2/(16{d^*}^2)$.

Remark 3.6. The rate of convergence $o(n^{2\beta/(2\beta+1)})$ of the standard GSE in the case of no noise has been recovered. This rate of convergence is obviously optimal since the case of no noise is included in the noisy case. The asymptotic variance of $\hat{d}_n$ dramatically increases when d^* is small. Hence the gain in the rate of convergence with respect to the standard GSE is balanced by the loss in the asymptotic variance. Nevertheless, the simulations in Hurvich and Ray (2001) indicate that it is better to estimate the variance of the noise.

Remark 3.7. If $d = 3/4$, then it can be shown that $m^{1/2}(\hat{d}_n^* - d)$ converges to a non Gaussian distribution and if $d \in (3/4, 1)$ then the rate of convergence and the asymptotic distribution of $\hat{d}_n^* - d$ both depend on d^*. See Velasco (1999) in the standard case.

3.1 Asymptotic normality of the standard GSE when $d^* = 0$

The modified local Whittle estimator $\hat{d}_n^*$ is consistent in the case $d^* = 0$, but it would be difficult to obtain an asymptotic distribution for it. Instead, it is possible to test the hypothesis $d^* = 0$ using the standard GSE. By standard GSE, we mean

$$\hat{d}_n^{st} = \arg \min_{d \in [-\epsilon, \epsilon]} \left\{ \log \left(\sum_{k=1}^{m} k^{2d} I_{X,k} \right) - \frac{2d}{m} \sum_{k=1}^{m} \log(k) \right\},$$

for some arbitrary $\epsilon \in (0, 1/2)$. The theory of Robinson (1995b) cannot be directly applied in the present context to prove consistency and asymptotic normality of $\hat{d}_n^{st}$, since the process $X = Y + \eta$ is not necessarily linear with respect to a martingale difference sequence. Nevertheless, if Z and η satisfy assumptions **(A3)** and **(A5)**, we can define a martingale difference sequence ξ, which also satisfies **(A5)**. Note that in the present context, the spectral density of X has the same degree of smoothness at zero as f_Y. More precisely, we have $f_X^*(0) = f_Y^*(0) + 2\sqrt{f_Y^*(0)/(2\pi)}\rho_\eta\sigma_\eta + \sigma_\eta^2/(2\pi)$. Define then:

$$\xi_k = \frac{\sqrt{2\pi f_Y^*(0)} Z_k + \eta_k}{\sqrt{2\pi f_X^*(0)}}.$$

X does not admit a linear representation with respect to ξ, but we can adapt Lemmas 4.1, 4.2 and 4.3 to the present context.

Proposition 3.4. *Assume* **(A3)**, **(A4)**, **(A5)**, **(A6)** *and*

$$\mathrm{cum}(Z_u, Z_v, Z_s, \eta_t) = \gamma \quad \textit{if } s = t = u = v \textit{ and 0 otherwise.} \tag{3.4}$$

Assume moreover that η is a martinglale difference sequence. If m is a non decreasing sequence of integers that satisfies $\lim_{n\to\infty}(m^{-1}+n^{-2\beta}m^{2\beta+1}\log^2(m))=0$, then $m^{1/2}\hat{d}_n^{st}$ is asymptotically Gaussian with zero mean and variance 1/4.

This result yields a test for long memory in volatility based on the standard GSE estimator. Another test for long memory in volatility, based on the ordinary GPH estimator, was justified by Hurvich and Soulier (2000). Since the ratio of the asymptotic variances of the GPH and GSE estimators is $\pi^2/6$, the test based on the GSE estimator should have higher local power than the one based on GPH.

3.2 Proof of Propositions 3.1, 3.2 and 3.4

Proof of Proposition 3.1. Define

$$
\begin{gathered}
\mathcal{E}_k = \frac{x_k^{2d^*} I_{X,k}}{f_X^*(0)(1+h(d^*,\theta^*,x_k))} \\
S_m(d,\theta) = \frac{1}{m}\sum_{k=1}^m \alpha_k(d,\theta)k^{2d-2d^*}\mathcal{E}_k, \\
U_m(d,\theta) = mS_m(d,\theta)\nabla\hat{J}_m(d,\theta), \\
\delta_{0,k}(d,\theta) = 2\log(k) - 2m^{-1}\sum_{j=1}^m \log(j) - \frac{\partial_d h(d,\theta,x_k)}{1+h(d,\theta,x_k)} + m^{-1}\sum_{j=1}^m \frac{\partial_d h(d,\theta,x_j)}{1+h(d,\theta,x_j)}, \\
\delta_{i,k}(d,\theta) = \frac{\partial_{\theta_i} h(d,\theta,x_k)}{1+h(d,\theta,x_k)} - m^{-1}\sum_{\ell=1}^m \frac{\partial_{\theta_i} h(d,\theta,x_\ell)}{1+h(d,\theta,x_\ell)},\ i=1,2, \\
N_k(d,\theta) = (\delta_{0,k},\delta_{1,k},\delta_{2,k}), \\
N_k^* = N_k(d^*,\theta^*),\ S_m^* = S_m(d^*,\theta^*),\ U_m^* = U_m(d^*,\theta^*).
\end{gathered}
$$

With these notations, $mD_n^{*\,-1}\nabla\hat{J}_m(d^*,\theta^*) = (S_m^*)^{-1}D_n^{*\,-1}U_m^*$ and $U_m^* = \sum_{k=1}^m N_k^*\mathcal{E}_k$. We will prove that S_m^* tends to 1 in probability and that $D_n^{*\,-1}U_m^*$ is asymptotically Gaussian with covariance matrix Γ^*.

The proof of the asymptotic normality of $D_n^{-1}U_m^*$ is classically based on the so-called Wold device. We must prove that for any $x\in\mathbb{R}^3$, $x^T D_n^{*\,-1}U_m^*$ converges in distribution to a Gaussian random variable with mean zero and variance $x^T\Gamma^*x$. Define

$$
t_n^2(x) = \sum_{k=1}^m (x^T D_n^{*\,-1}N_k^*)^2,\quad c_{n,k}(x) = t_n^{-1}(x)x^T D_n^{*\,-1}N_k^*,\quad \text{and}\quad T_n = \sum_{k=1}^m c_{n,k}(x)\mathcal{E}_k.
$$

Using this notation, we have $x^T D_n^{*\,-1}U_m^* = t_n(x)T_n$ and it suffices to prove that T_n is asymptotically Gaussian with zero mean and unit variance and that $\lim_{n\to\infty} t_n(x)^2 = x^T\Gamma^*x$. This last property is obtained by elementary calculus (approximating sums by integrals) and its proof is omitted. To prove the asymptotic normality of T_n, observe that

$$\max_{1 \leq k \leq m} |c_{n,k}(x)| = O(\log(m) m^{-1/2}) \text{ and } |c_{n,k}(x) - c_{n,k+1}(x)| = O(k^{-1} m^{-1/2}).$$

Hence (4.3) holds and we can apply Theorem 4.1.

We conclude the proof by checking that S_m^* tends to 1 in probability. In view of the proof of Proposition 3.2, we will actually prove that $S_m(d,\theta)$ converges to 1 in probability uniformly with respect to $(d,\theta) \in \mathcal{D}_m \times \Theta$ where $\mathcal{D}_m := \{d;\ |d - d^*| \leq C \log^{-s}(m)\}$. Using the notations of section 2, we can write

$$S_m(d,\theta) = \frac{1}{m} \sum_{j=1}^{m} \alpha_j(d,\theta) j^{2d-2d^*}\ \{1 + E_n(d,\theta)\}.$$

By proposition 2.1, $E_n(d,\theta)$ converges in probability to 0 uniformly with respect to $(d,\theta) \in \mathcal{D}_m \times \Theta\}$. Moreover, on this set, it is easily seen that $\frac{1}{m}\sum_{j=1}^{m} \alpha_j(d,\theta) j^{2d-2d^*}$ converges uniformly to 1, and this concludes the proof. □

Proof of Proposition 3.2. We must prove that $m D_n^{*\,-1} \nabla^2 \hat{J}_m(d,\theta) D_n^{*\,-1}$ converges to Γ^* uniformly with respect to $(d,\theta) \in \mathcal{D}_m \times \Theta$. Using the notations introduced above, we have

$$m \nabla \hat{J}_m(d,\theta) = S_m^{-1} \sum_{k=1}^{m} N_k(d,\theta) \alpha_k(d,\theta) k^{2d-2d^*} \mathcal{E}_k.$$

Hence

$$\begin{aligned}
m \nabla^2 \hat{J}_m(d,\theta) = {} & S_m^{-1}(d,\theta) \sum_{k=1}^{m} N_k(d,\theta) \{\nabla(\alpha_k(d,\theta) k^{2d-2d^*})\}^T \mathcal{E}_k \\
& + S_m^{-1}(d,\theta) \sum_{k=1}^{m} \nabla N_k(d,\theta) \alpha_k(d,\theta) k^{2d-2d^*} \mathcal{E}_k \\
& - S_m^{-2}(d,\theta) \sum_{k=1}^{m} N_k(d,\theta) \alpha_k(d,\theta) k^{2d-2d^*} \mathcal{E}_k (\nabla S_m(d,\theta))^T \\
& =: S_m^{-1}(d,\theta) M_{1,n}(d,\theta) + S_m^{-1}(d,\theta) M_{2,n}(d,\theta) + S_m^{-2}(d,\theta) M_{3,n}(d,\theta).
\end{aligned}$$

Since we already know that $S_m^{-1}(d,\theta)$ converges uniformly to 1, we only need to prove that $D_n^{*\,-1} M_{1,n} D_n^{*\,-1}$ converges in probability to Γ^* uniformly with respect to $(d,\theta) \in \mathcal{D}_m \times \Theta$ and that $D_n^{*\,-1} M_{2,n} D_n^{*\,-1}$ and $D_n^{*\,-1} M_{3,n} D_n^{*\,-1}$ converge to 0. We will prove only the first fact, the other being routine applications of the same techniques.

Denote $M_{1,n}(d,\theta) = (M_{1,n}^{(i,j)}(d,\theta))_{0 \leq i,j \leq 2}$. For $i = 0,1,2$, let $D_{i,n}^*$ be the i-th diagonal element of the matrix D_n^*. For $j = 1,2$, we have:

$$\partial_{\theta_j} \alpha_k(d,\theta) = -\frac{\partial_{\theta_j} h(d,\theta,x_k)}{1 + h(d,\theta,x_k)} \alpha_k(d,\theta).$$

Hence for $i = 0, \ldots, u$ and $j = 1, \ldots, u$, we have

$$M_{1,n}^{(i,j)}(d,\theta) = -\sum_{k=1}^{m} \delta_{i,k}(d,\theta) \frac{\partial_{\theta_j} h(d,\theta,x_k)}{1+h(d,\theta,x_k)} \alpha_k(d,\theta) k^{2d-2d^*} \mathcal{E}_k.$$

Since $\sum_{k=1}^{m} \delta_{i,k} = 0$, we obtain:

$$D_{i,n}^{-1} D_{j,n}^{-1} M_{1,n}^{(i,j)}(d,\theta) = -D_{i,n}^{-1} D_{j,n}^{-1} \sum_{k=1}^{m} \delta_{i,k}(d,\theta) \delta_{j,k}(d,\theta) \tag{3.5}$$

$$- D_{i,n}^{-1} D_{j,n}^{-1} \sum_{k=1}^{m} \delta_{i,k}(d,\theta) \frac{\partial_{\theta_j} h(d,\theta,x_k)}{1+h(d,\theta,x_k)} \left(k^{2d-2d^*} \alpha_k(d,\theta) - 1\right) \tag{3.6}$$

$$- D_{i,n}^{-1} D_{j,n}^{-1} \sum_{k=1}^{m} \delta_{i,k}(d,\theta) \frac{\partial_{\theta_j} h(d,\theta,x_k)}{1+h(d,\theta,x_k)} (d,\theta) k^{2d-2d^*} \alpha_k(d,\theta)(\mathcal{E}_k - 1). \tag{3.7}$$

It is easily seen that the term on the right hand side of (3.5) converges to the expected limit. Since $d \in \mathcal{D}_m$ and $|D_{i,n}^{-1} \delta_{i,k}| \le C \log(n) m^{-1/2}$, we easily obtain that the term (3.6) is $O(\log^{2-s}(n))$. The term (3.7) can be expressed as

$$\sum_{k=1}^{m} c_{n,k}^{(i,j)}(d,\theta) \left(\mathcal{E}_k - 2\pi I_{Z,k}\right) + \sum_{k=1}^{m} c_{n,k}(d,\theta) \left(2\pi I_{Z,k} - 1\right),$$

where the coefficients $c_{n,k}^{(i,j)}(d,\theta)$ satisfy

$$\sup_{(d,\theta) \in \mathcal{D}_m \times \Theta} \max_{1 \le k \le m} |c_{n,k}^{(i,j)}(d,\theta)| = O(\log^2(m) m^{-1}).$$

Applying this bound and (4.21) in Theorem 4.1, we obtain, for some $\gamma > 0$:

$$\sum_{k=1}^{m} \sup_{(d,\theta) \in \mathcal{D}_m \times \Theta} |c_{n,k}(d,\theta)| \mathbb{E}[|\mathcal{E}_k - 2\pi I_{Z,k}|] \le \log^2(m) \left(m^{-\gamma} + (m/n)^{\gamma}\right).$$

To prove that $\sup_{(d,\theta) \in \mathcal{D}_m \times \Theta} |\sum_{k=1}^{m} c_{n,k}(d,\theta) \left(2\pi I_{Z,k} - 1\right)|$ converges in probability to 0, we use summation by parts as in the last part of the proof of Proposition 2.1. It can be shown that $\sup_{(d,\theta) \in \mathcal{D}_m \times \Theta} |c_{n,k}^{(i,j)} - c_{n,k+1}^{(i,j)}| \le C \log^2(m) m^{-1} k^{-1}$, and this suffices to prove the required result.

We now consider the derivatives with respect to d: $\partial_d(\alpha_k(d,\theta) k^{2d-2d^*}) = \mu_k \alpha_k(d,\theta) k^{2d-2d^*}$ with $\mu_k = 2\log(k) - \frac{\partial_d h(d,\theta,x_k)}{1+h(d,\theta,x_k)}$. Hence,

$$D_{i,n}^{-1} D_{0,n}^{-1} M_{1,n}^{(i,0)}(d,\theta) = D_{i,n}^{-1} D_{0,n}^{-1} \sum_{k=1}^{m} \delta_{i,k}(d,\theta) \delta_{0,k}(d,\theta) \tag{3.8}$$

$$+ D_{i,n}^{-1} D_{0,n}^{-1} \sum_{k=1}^{m} \delta_{i,k}(d,\theta) \mu_k \left(k^{2d-2d^*} \alpha_k(d,\theta) - 1\right)$$

$$+ D_{i,n}^{-1} D_{0,n}^{-1} \sum_{k=1}^{m} \delta_{i,k}(d,\theta) \mu_k \alpha_k(d,\theta) k^{2d-2d^*} \alpha_k(d,\theta)(\mathcal{E}_k - 1).$$

As previously, the first term on the right hand side of (3.8) converges to the desired limit and the other terms tend to 0, uniformly with respect to $(d, \theta) \in \mathcal{D}_m \times \Theta$. $\square$

Proof of Proposition 3.4. To prove the consistency of the estimator, in view of Proposition 2.1, we only need to check that there exists a positive real η such that $\mathbb{E}[|I_{X,k}/\tilde{f}_{X,k} - 2\pi I_{\xi,k}|] \leq C(k^{-\eta} + (k/n)^{\eta})$. Here, we have defined $\tilde{f}_{X,k} = f_X(0)$ for all k. With this notation, we obtain:

$$\begin{aligned} I_{X,k} &= |d_{Y,k} + d_{\eta,k}|^2 \\ &= \left| d_{Y,k} - \sqrt{2\pi f_Y^*(0)} d_{Z,k} + \sqrt{2\pi f_X^*(0)} d_{\xi,k} \right|^2 = \left| d_{Y,k} - \sqrt{2\pi f_Y^*(0)} d_{Z,k} \right|^2 \\ &\quad + 2\sqrt{2\pi f_X^*(0)} \operatorname{Re}\left(\bar{d}_{\xi,k} \left(d_{Y,k} - \sqrt{2\pi f_Y^*(0)} d_{Z,k} \right) \right) + 2\pi f_X^*(0) I_{\xi,k}. \end{aligned} \tag{3.9}$$

Applying (4.1) and the Cauchy-Schwarz inequality, we obtain

$$\mathbb{E}\left[\left| d_{Y,k}/\sqrt{f_Y^*(0)} - \sqrt{2\pi} d_{Z,k} \right|^2 \right] \leq C(\log(k) k^{-1} + (k/n)^{\beta}),$$

$$\mathbb{E}\left[\left| \bar{d}_{\xi,k} \left(d_{Y,k}/\sqrt{f_Y^*(0)} - \sqrt{2\pi} d_{Z,k} \right) \right| \right] \leq C(\log^{1/2}(k) k^{-1/2} + (k/n)^{\beta/2}).$$

Hence we obtain $\mathbb{E}[|I_{X,k}/\tilde{f}_{X,k} - 2\pi I_{\xi,k}|] \leq C(k^{-\eta} + (k/n)^{\eta})$ for any $\eta < (1 \wedge \beta)/2$.

To prove the central limit theorem, note that ξ satisfies **(A3)**, with $\operatorname{cov}(Z_k, \xi_k) = 1 + \rho_\eta \sigma_\eta / (2\pi f_Y^*(0))$ and (3.4) implies that (3.2) holds with

$$\operatorname{cum}(Z_u, Z_v, \xi_s, \xi_t) = \operatorname{cum}(Z_0, Z_0, Z_0, Z_0) + 2\gamma/\sqrt{2\pi f_Y^*(0)} + \kappa/(2\pi f_Y^*(0)),$$

if $s = t = u = v$ and 0 otherwise. Hence we can apply Lemma 4.1 and we obtain that $\sum_{k=1}^m c_{n,k} I_{X,k}/\tilde{f}_{X,k} - 2\pi I_{\xi,k} = o_P(1)$ and $2\pi \sum_{k=1}^m c_{n,k} I_{\xi,k}$ is asymptotically standard Gaussian. $\square$

4 Technical results

We start by stating the results we use on the DFT and periodogram ordinates of a stationary long memory process Y satisfying assumption **(A4)**. Such results can be found in many references, starting with Robinson (1995b). We prefer to refer to Soulier (2002) which better suits our purpose. We first introduce some more notation. Define $\tilde{a}_k = \sqrt{2\pi f_Y^*(0)}(1 - \mathrm{e}^{\mathrm{i}x_k})^{-d_Y}$ and $\tilde{f}_{Y,k} = x_k^{-2d_Y} f_Y^*(0)$. With these definitions, we have, for some numerical constant C,

$$\left| \frac{\tilde{f}_{Y,k} - |\tilde{a}_k|^2/(2\pi)}{\tilde{f}_{Y,k}} \right| \leq C x_k^2.$$

The following Lemma gathers Lemmas 6.1, 6.2 and Theorem 6.1 of Soulier (2002), in the particular case of the non tapered periodogram (*i.e.* with $q = 0$ in the notations of that paper).

Lemma 4.1. *Assume* **(A1)** *and* **(A4)***. Then*

$$\mathbb{E}[|d_{Y,k}/\tilde{a}_k - d_{Z,k}|^2] \le C(\log(k)k^{-1} + (k/n)^{\beta}), \tag{4.1}$$

$$\mathbb{E}[|I_{Y,k}/\tilde{f}_{Y,k} - 2\pi I_{Z,k}|] \le C\left(\log^{1/2}(k)(k)^{-1/2} + (k/n)^{\beta/2}\right). \tag{4.2}$$

Assume moreover **(A5)** *and* **(A6)***. Let m be a non decreasing sequence of integers that satisfies (3.3) and let $(c_{n,k})_{1\le k\le m}$ be a triangular array of real numbers such that*

$$\sum_{k=1}^{m} c_{n,k} = 0, \qquad \sum_{k=1}^{m} c_{n,k}^2 = 1 \tag{4.3}$$

$$\lim_{n\to\infty}\left\{\sum_{k=1}^{m} |c_{n,k} - c_{n,k+1}| + |c_{n,\tilde{n}}|\right\}^2 \log(n) = 0, \tag{4.4}$$

Then

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\sum_{k=1}^{m} c_{n,k}\left(\frac{I_{Y,k}}{\tilde{f}_{Y,k}} - 2\pi I_{Z,k}\right)\right|\right] = 0, \tag{4.5}$$

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\sum_{k=1}^{m} c_{n,k}\bar{d}_{\eta,k}(d_{Y,k}/\tilde{a}_k - d_{Z,k})\right|\right] = 0, \tag{4.6}$$

and $2\pi\sum_{k=1}^{m} c_{n,k} I_{Z,k}$ and $\sum_{k=1}^{m} c_{n,k}\frac{I_{Y,k}}{\tilde{f}_{Y,k}}$ are asymptotically standard Gaussian.

We now deal with the approximation of the periodogram of the signal plus noise by the periodogram of the signal. Define $\tilde{f}_{X,k} = x_k^{-2d^*} f_Y^*(0)(1 + h(d^*, \theta^*, x_k))$.

Lemma 4.2. *Assume* **(A1)***,* **(A3)** *and* **(A4)***. If $d^* \in (0,1)$, then there exist $\gamma^* > 0$ and $C > 0$ such that*

$$\mathbb{E}\left[\left|\frac{I_{X,k}}{\tilde{f}_{X,k}} - \frac{I_{Y,k}}{\tilde{f}_{Y,k}}\right|\right] \le C\left(k^{-\gamma} + (k/n)^{\gamma}\right). \tag{4.7}$$

Proof. We first prove (4.7) in the stationary case $d^* = d_Y \in (0, 1/2)$. Write:

$$\begin{aligned}\frac{I_{X,k}}{\tilde{f}_{X,k}} - \frac{I_{Y,k}}{\tilde{f}_{Y,k}} &= \frac{I_{Y,k}}{\tilde{f}_{X,k}} - \frac{I_{Y,k}}{\tilde{f}_{Y,k}} + \frac{2\mathrm{Re}\left(d_{Y,k}\bar{d}_{\eta,k}\right)}{\tilde{f}_{X,k}} + \frac{I_{\eta,k}}{\tilde{f}_{X,k}} \\ &= \frac{\tilde{f}_{Y,k} - \tilde{f}_{X,k}}{\tilde{f}_{X,k}}\frac{I_{Y,k}}{\tilde{f}_{Y,k}} + \frac{2\sqrt{\tilde{f}_{Y,k}}}{\tilde{f}_{X,k}}\mathrm{Re}\left(\frac{d_{Y,k}}{\sqrt{\tilde{f}_{Y_k}}}\bar{d}_{\eta,k}\right) + \frac{I_{\eta,k}}{\tilde{f}_{X,k}}.\end{aligned}$$

Since $\mathbb{E}[I_{Y,k}/\tilde{f}_{Y,k}]$ is uniformly bounded over the class $\mathcal{F}(\alpha,\beta,\mu)$ and

$$\frac{\tilde{f}_{Y,k} - \tilde{f}_{X,k}}{\tilde{f}_{X,k}} + \frac{\sqrt{\tilde{f}_{Y,k}}}{\tilde{f}_{X,k}} \le C\left((k/n)^{d_Y} + (k/n)^{\beta}\right),$$

we obtain (4.7) with $\gamma = d_Y \wedge \beta$ in the stationary case. In the non stationary case, extra terms appear. Recall that $U_t = \sum_{s=1}^t Y_s$. Then

$$\begin{aligned}
d_{U,k} &= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n \sum_{s=1}^t Y_s \mathrm{e}^{\mathrm{i}tx_k} = \frac{1}{\sqrt{2\pi n}} \sum_{s=1}^n Y_s \sum_{t=s}^n \mathrm{e}^{\mathrm{i}tx_k} = \frac{d_{Y,k}}{1-\mathrm{e}^{\mathrm{i}x_k}} - \frac{\mathrm{e}^{\mathrm{i}x_k} \sum_{s=1}^n Y_s}{\sqrt{2\pi n}(1-\mathrm{e}^{\mathrm{i}x_k})}, \\
I_{U,k} &= \frac{I_{Y,k}}{|1-\mathrm{e}^{\mathrm{i}x_k}|^2} - \frac{2\mathrm{Re}(\mathrm{e}^{\mathrm{i}x_k} d_{Y,k}) \sum_{s=1}^n Y_s}{\sqrt{2\pi n}|1-\mathrm{e}^{\mathrm{i}x_k}|^2} + \frac{\left(\sum_{s=1}^n Y_s\right)^2}{2\pi n |1-\mathrm{e}^{\mathrm{i}x_k}|^2}, \\
I_{X,k} &= I_{U,k} + 2\mathrm{Re}(d_{U,k} \bar{d}_{\eta,k}) + I_{\eta,k} \\
&= \frac{I_{Y,k}}{|1-\mathrm{e}^{\mathrm{i}x_k}|^2} - \frac{2\mathrm{Re}(\mathrm{e}^{\mathrm{i}x_k} d_{Y,k}) \sum_{s=1}^n Y_s}{\sqrt{2\pi n}|1-\mathrm{e}^{\mathrm{i}x_k}|^2} + \frac{\left(\sum_{s=1}^n Y_s\right)^2}{2\pi n |1-\mathrm{e}^{\mathrm{i}x_k}|^2} \\
&\quad + 2\mathrm{Re}\left(\frac{d_{Y,k}}{1-\mathrm{e}^{\mathrm{i}x_k}} \bar{d}_{\eta,k}\right) - 2\mathrm{Re}\left(\frac{\mathrm{e}^{\mathrm{i}x_k} \sum_{s=1}^n Y_s}{\sqrt{2\pi n}(1-\mathrm{e}^{\mathrm{i}x_k})} \bar{d}_{\eta,k}\right) + I_{\eta,k}
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{I_{X,k}}{\tilde{f}_{X,k}} - \frac{I_{Y,k}}{\tilde{f}_{Y,k}} &= \frac{I_{Y,k}}{\tilde{f}_{Y,k}} \left(\frac{\tilde{f}_{Y,k}}{|1-\mathrm{e}^{\mathrm{i}x_k}|^2 \tilde{f}_{X,k}} - 1 \right) - \frac{2\mathrm{Re}(\mathrm{e}^{\mathrm{i}x_k} d_{Y,k}) \sum_{s=1}^n Y_s}{\sqrt{2\pi n}|1-\mathrm{e}^{\mathrm{i}x_k}|^2 \tilde{f}_{X,k}} + \frac{\left(\sum_{s=1}^n Y_s\right)^2}{2\pi n |1-\mathrm{e}^{\mathrm{i}x_k}|^2 \tilde{f}_{X,k}} \\
&\quad + \mathrm{Re}\left(\frac{2 d_{Y,k}}{(1-\mathrm{e}^{\mathrm{i}x_k}) \tilde{f}_{X,k}} \, \bar{d}_{\eta,k} \right) - \mathrm{Re}\left(\frac{2\mathrm{e}^{\mathrm{i}x_k} \sum_{s=1}^n Y_s}{\sqrt{2\pi n}(1-\mathrm{e}^{\mathrm{i}x_k}) \tilde{f}_{X,k}} \, \bar{d}_{\eta,k} \right) + \frac{I_{\eta,k}}{\tilde{f}_{X,k}}. \qquad (4.8)
\end{aligned}$$

Straightforward variance computations yield, for $d_Y \in (-1/2, 0)$, that

$$\mathbb{E}\left[\left(\sum_{s=1}^n Y_s \right)^2 \right] \leq C n^{2d_Y+1}. \qquad (4.9)$$

Thus

$$\begin{aligned}
\mathbb{E}\left[\left| \frac{I_{X,k}}{\tilde{f}_{X,k}} - \frac{I_{Y,k}}{\tilde{f}_{Y,k}} \right| \right] &\leq C \Big((k/n)^\beta + (k/n)^{2+2d_Y} + k^{d_Y} + k^{2d_Y} + (k/n)^{2+d_Y} \\
&\quad + n^{d_Y} (k/n)^{1+2d_Y} + (k/n)^{2+2d_Y} \Big) \leq C \left(k^{d_Y} + (k/n)^\beta + (k/n)^{2+2d_Y} \right).
\end{aligned}$$

This proves (4.7) in the non stationary case with $\gamma = (-d_Y) \wedge \beta$. $\square$

Lemma 4.3. *Assume* **(A3)**, **(A4)**, **(A5)** *and* **(A6)**. *Let m be a sequence of integers such that*

$$\lim_{n\to\infty} (m^{-1} + m^{2\beta+1} n^{-2\beta}) = 0. \qquad (4.10)$$

Let $(c_{n,k})_{1\leq k\leq m}$ be a triangular array of real numbers that satisfy (4.3). If $d^* \in (0, 3/4)$, then

$$\sum_{k=1}^m c_{n,k} \left\{ \frac{I_{X,k}}{\tilde{f}_{X,k}} - \frac{I_{Y,k}}{\tilde{f}_{Y,k}} \right\} = o_P(1). \qquad (4.11)$$

Proof. We first prove (4.11) in the stationary case.

$$\frac{I_{X,k}}{\tilde f_{X,k}} - \frac{I_{Y,k}}{\tilde f_{Y,k}} = \frac{\tilde f_{Y,k} - \tilde f_{X,k}}{\tilde f_{X,k}} \left(\frac{I_{Y,k}}{\tilde f_{Y,k}} - 2\pi I_{Z,k} \right) + \frac{2}{\tilde f_{X,k}} \mathrm{Re} \left(\tilde a_k \left(\frac{d_{Y,k}}{\tilde a_k} - d_{Z,k} \right) \bar d_{\eta,k} \right) \tag{4.12}$$

$$+ \frac{\tilde f_{Y,k} - \tilde f_{X,k}}{\tilde f_{X,k}} (2\pi I_{Z,k} - 1) + \frac{I_{\eta_k} - \sigma_\eta^2/(2\pi)}{\tilde f_{X,k}} + \frac{2}{\tilde f_{X,k}} \mathrm{Re} \left(\tilde a_k \left\{ d_{Z,k} \bar d_{\eta,k} - \frac{\rho_\eta \sigma_\eta}{2\pi} \right\} \right). \tag{4.13}$$

The terms in (4.13) can be easily bounded. Since Z and η satisfy assumptions **(A3)**, **(A5)** and (3.2), straightforward computations yield:

$$\mathbb{E}\Big[\Big(\sum_{k=1}^{m} \frac{c_{n,k}}{\tilde f_{X,k}} \Big\{ I_{\eta,k} - \sigma_\eta^2/(2\pi) + \left(\tilde f_{Y,k} - \tilde f_{X,k} \right) (2\pi I_{Z,k} - 1)$$

$$+ 2\,\mathrm{Re} \left(\tilde a_k \left\{ d_{Z,k} \bar d_{\eta,k} - \frac{\rho_\eta \sigma_\eta}{2\pi} \right\} \right) \Big\} \Big)^2 \Big] = O((m/n)^{2d^*}). \tag{4.14}$$

The terms in (4.12) are bounded by (4.5) and (4.6).

We now consider the non stationary case. Starting from (4.8), we write:

$$\frac{I_{X,k}}{\tilde f_{X,k}} - \frac{I_{Y,k}}{\tilde f_{Y,k}} = \left(\frac{\tilde f_{Y,k}}{|1 - \mathrm{e}^{\mathrm{i} x_k}|^2 \tilde f_{X,k}} - 1 \right) \left(\frac{I_{Y,k}}{\tilde f_{Y,k}} - 2\pi I_{Z,k} \right) \tag{4.15}$$

$$+ \mathrm{Re} \left(\frac{2\tilde a_k \left(\frac{d_{Y,k}}{\tilde a_k} - d_{Z,k} \right) \bar d_{\eta,k}}{(1 - \mathrm{e}^{\mathrm{i} x_k}) \tilde f_{X,k}} \right) + \mathrm{Re} \left(\frac{2\tilde a_k \left\{ d_{Z,k} \bar d_{\eta,k} - \frac{\rho_\eta \sigma_\eta}{2\pi} \right\}}{(1 - \mathrm{e}^{\mathrm{i} x_k}) \tilde f_{X,k}} \right) \tag{4.16}$$

$$+ \left(\frac{\tilde f_{Y,k}}{|1 - \mathrm{e}^{\mathrm{i} x_k}|^2 \tilde f_{X,k}} - 1 \right) (2\pi I_{Z,k} - 1) + \frac{I_{\eta,k} - \sigma_\eta^2/(2\pi)}{\tilde f_{X,k}} \tag{4.17}$$

$$- \frac{2\mathrm{Re}(\mathrm{e}^{\mathrm{i} x_k} d_{Y,k}) \sum_{s=1}^n Y_s}{\sqrt{2\pi n} |1 - \mathrm{e}^{\mathrm{i} x_k}|^2 \tilde f_{X,k}} \tag{4.18}$$

$$+ \frac{\left(\sum_{s=1}^n Y_s \right)^2}{2\pi n |1 - \mathrm{e}^{\mathrm{i} x_k}|^2 \tilde f_{X,k}} - \mathrm{Re} \left(\frac{2\mathrm{e}^{\mathrm{i} x_k} \sum_{s=1}^n Y_s}{\sqrt{2\pi n} (1 - \mathrm{e}^{\mathrm{i} x_k}) \tilde f_{X,k}} \bar d_{\eta,k} \right) \tag{4.19}$$

$$+ \frac{2\rho_\eta \sigma_\eta \mathrm{Re} \left(\tilde a_k (1 - \mathrm{e}^{\mathrm{i} x_k})^{-1} \right) + \sigma_\eta^2}{2\pi \tilde f_{X,k}} + \frac{\tilde f_{Y,k}}{|1 - \mathrm{e}^{\mathrm{i} x_k}|^2 \tilde f_{X,k}} - 1. \tag{4.20}$$

The terms in (4.15), (4.16) and (4.17) are similar to the terms that appear in the stationary case. We only consider the terms appearing in (4.18), (4.19) and (4.20). To deal with (4.20), note that

$$\tilde f_{X,k} = x_k^{-2} \tilde f_{Y,k} + \{ 2\rho_\eta \sigma_\eta \mathrm{Re}(\tilde a_k (1 - \mathrm{e}^{\mathrm{i} x_k})^{-1}) + \sigma_\eta^2 \}/(2\pi).$$

Hence, denoting $r_{n,k}$ the sum of the terms in (4.20), we have

$$r_{n,k} = \frac{\tilde f_{X_k} - x_k^{-2} \tilde f_{Y,k}}{\tilde f_{X,k}} + \frac{\tilde f_{Y,k}}{|1 - \mathrm{e}^{\mathrm{i} x_k}|^2 \tilde f_{X,k}} - 1 = \frac{\tilde f_{Y,k}}{x_k^2 \tilde f_{X,k}} \left(\frac{x_k^2}{|1 - \mathrm{e}^{\mathrm{i} x_k}|^2} - 1 \right).$$

Since $\frac{\tilde{f}_{Y,k}}{x_k^2 \tilde{f}_{X,k}}$ is bounded and $x_k^2 |1 - \mathrm{e}^{\mathrm{i}x_k}|^{-2} - 1 = O(x_k^2)$ (uniformly with respect to k and n), we obtain $\sum_{k=1}^m c_{n,k} r_{n,k} = O(m^{5/2} n^{-2}) = o(1)$ under condition (4.10).

Consider now the term (4.18), say R_n. Define $\tilde{c}_{n,k} = n^{d_Y} c_{n,k} \mathrm{e}^{\mathrm{i}x_k} \tilde{a}_k / (2\pi |1 - \mathrm{e}^{\mathrm{i}x_k}|^2 \tilde{f}_{X,k})$, $R_{n,1} = \sum_{k=1}^m \tilde{c}_{n,k} \left(\sqrt{2\pi} d_{Y,k} / \tilde{a}_k - \sqrt{2\pi} d_{Z,k} \right)$ and $R_{n,2} = \sum_{k=1}^m \tilde{c}_{n,k} \sqrt{2\pi} d_{Z,k}$. Then

$$R_n = n^{-1/2 - d_Y} \sum_{s=1}^n Y_s \ (R_{n,1} + R_{n,2}) \,.$$

Applying (4.9) and the Hölder inequality, we obtain

$$\mathbb{E}[|R_n|] \leq C \left(\mathbb{E}^{1/2}[R_{n,1}^2] + \mathbb{E}^{1/2}[R_{n,2}^2] \right) .$$

Since Z satisfies assumption **(A5)** and $|\tilde{c}_{n,k}| \leq C \, |c_{n,k}| k^{d_Y}$, it is easily seen that:

$$\mathbb{E}[R_{n,2}^2] \leq C \sum_{k=1}^m c_{n,k}^2 k^{2d_Y} = o(1).$$

The last equality follows straightforwardly from (4.10) and the assumption $d^* \in [1/2, 3/4)$ which implies that $d_Y \in [-1/2, -1/4)$. Applying (4.1) and twice the Cauchy-Schwarz inequality, we now bound $R_{1,n}$:

$$\mathbb{E}[R_{n,1}^2] \leq C \left(\sum_{k=1}^m |c_{n,k}| k^{2d_Y} \right)^{1/2} \left(\sum_{k=1}^m |c_{n,k}| \left\{ k^{-1} + (k/n)^\beta \right\} \right)^{1/2}$$

If $(t_k)_{k \geq 1}$ is a square summable sequence, then under condition (4.3), $\sum_{k=1}^m |c_{n,k} t_k| = o(1)$. For our purpose, we can even restrict ourselves to non increasing sequences. Split the sum at some $\ell \leq m$ to be fixed later and apply the Hölder inequality to the sum over $k \geq \ell$:

$$\sum_{k=1}^m |c_{n,k}| t_k \leq \ell t_1 \max_{1 \leq k \leq \ell} |c_{n,k}| + \left(\sum_{k \geq \ell} t_k^2 \right)^{1/2} .$$

These last two terms are simultaneously $o(1)$ as soon as the sequence $\ell = \ell(m)$ tends to infinity not too fast, that is in such a way that $\lim \left(\ell \max_{1 \leq k \leq m} |c_{n,k}| \right) = 0$, which is possible under (4.3). Hence, if $d_Y < -1/4$, then $\sum_{k=1}^m |c_{n,k}| k^{2d_Y} = o(1)$. Similarly, $\sum_{k=1}^m |c_{n,k}| k^{-1} = o(1)$. Moreover, $\sum_{k=1}^m |c_{n,k}| (k/n)^\beta = O(m^{\beta+1/2} n^{-\beta}) = o(1)$ under (4.10). Finally, $\mathbb{E}[R_{n,1}^2] = o(1)$.

Both terms in (4.19) can be dealt with straightforwardly. Applying the bound (4.9), we get

$$\sum_{k=1}^m |c_{n,k}| \mathbb{E} \left[\frac{\left(\sum_{s=1}^n Y_s \right)^2}{2\pi n |1 - \mathrm{e}^{\mathrm{i}x_k}|^2 \tilde{f}_{X,k}} \right] \leq C \sum_{k=1}^m |c_{n,k}| k^{2d_Y} = o(1)$$

by the same arguments as above. Since η satisfies **(A5)**, applying (4.9) and the Hölder inequality, we bound the last term:

$$\mathbb{E}\left[\left|\sum_{k=1}^m c_{n,k}\mathrm{Re}\left(\frac{2\mathrm{e}^{\mathrm{i}x_k}\sum_{s=1}^n Y_s}{\sqrt{2\pi n}(1-\mathrm{e}^{\mathrm{i}x_k})\tilde{f}_{X,k}}\,\bar{d}_{\eta,k}\right)\right|^2\right] \\ \leq C n^{2d_Y}\mathbb{E}\left[\left|\sum_{k=1}^m \frac{c_{n,k}\mathrm{e}^{\mathrm{i}x_k}}{(1-\mathrm{e}^{\mathrm{i}x_k})\tilde{f}_{X,k}}\,\bar{d}_{\eta,k}\right|^2\right] \leq C n^{2d_Y}(m/n)^{2+4d_Y} = o(1).$$

□

Gathering the previous Lemmas, we obtain the needed results for the periodogram of the signal plus noise.

Theorem 4.1. *Assume* **(A1)**, **(A3)** *and* **(A4)**. *Then there exist $\gamma > 0$ and $C > 0$ such that*

$$\mathbb{E}[|I_{X,k}/\tilde{f}_{X,k} - 2\pi I_{Z,k}|] \leq C\left(k^{-\gamma} + (k/n)^{\gamma}\right). \tag{4.21}$$

Assume moreover **(A5)** *and* **(A6)**. *Let m be a non decreasing sequence of integers that satisfies (3.3) and let $(c_{n,k})_{1\leq k\leq m}$ be a triangular array of real numbers that satisfies (4.3) Then*

$$\lim_{n\to\infty}\mathbb{E}\left[\left|\sum_{k=1}^m c_{n,k}\left(\frac{I_{X,k}}{\tilde{f}_{X,k}} - 2\pi I_{Z,k}\right)\right|\right] = 0. \tag{4.22}$$

If moreover $\sum_{k=1}^m c_{n,k} = 0$ and (4.4) holds, then $\sum_{k=1}^m c_{n,k}\frac{I_{X,k}}{\tilde{f}_{X,k}}$ is asymptotically standard Gaussian.

Appendix

By definition, $\mathbb{P}(\hat{d}_n \in \mathcal{D}_1) \leq \mathbb{P}(\inf_{(d,\theta)\in\mathcal{D}_1\times\theta_n}\hat{J}_m(d,\theta) - \hat{J}_m(d^*,\theta^*) \leq 0)$. Define $p_m = (m!)^{1/m}$. For $d \in \mathcal{D}_1$, if $1 \leq j \leq p_m$, then $(j/p_m)^{2d-2d^*} \geq (j/p_m)^{-1+2\epsilon}$ and if $p_m < j \leq m$, then $(j/p_m)^{2d-2d^*} \geq (j/p_m)^{2\epsilon_n-2d^*}$. Define then $a_j = (j/p_m)^{-1+2\epsilon}$ if $1 \leq j \leq p_m$ and $a_j = (j/p_m)^{2\epsilon_n-2d^*}$ otherwise. As shown in Robinson (1995b, Eq. 3.22), if $\epsilon < 1/(4\mathrm{e})$, then for large enough n, $\sum_{j=1}^m a_j \geq 2$. Moreover, if $d \geq \epsilon_n$, then, for large enough n, $\alpha_k(d,\theta) \geq 1 - C\mathrm{e}^{-\epsilon_n\log(n/m)} \geq 0$ for some constant C depending on d^* and Θ. Define $\zeta_n = C\mathrm{e}^{-\epsilon_n\log(n/m)}$.

We obtain:

$$
\begin{aligned}
&\hat{J}_m(d,\theta) - \hat{J}_m(d^*,\theta^*) \\
&\qquad = \log\Big\{\frac{1}{m}\sum_{j=1}^{m}\Big(\frac{j}{p_m}\Big)^{2d-2d^*}\alpha_k(d,\theta)\mathcal{E}_j\Big\} - \log\Big\{\frac{1}{m}\sum_{j=1}^{m}\mathcal{E}_j\Big\} + m^{-1}\sum_{k=1}^{m}\log(\alpha_k(d,\theta)) \\
&\qquad \geq \log\Big\{\frac{1}{m}\sum_{j=1}^{m}a_j\mathcal{E}_j\Big\} - \log\Big\{\frac{1}{m}\sum_{j=1}^{m}\mathcal{E}_j\Big\} + 2\log(1-\zeta_n) \\
&\qquad = \log\Big\{\frac{1}{m}\sum_{j=1}^{m}a_j + \frac{1}{m}\sum_{j=1}^{m}a_j(\mathcal{E}_j-1)\Big\} - \log\Big\{1+\frac{1}{m}\sum_{j=1}^{m}(\mathcal{E}_j-1)\Big\} + 2\log(1-\zeta_n) \\
&\qquad \geq \log\Big\{2+\frac{1}{m}\sum_{j=1}^{m}a_j(\mathcal{E}_j-1)\Big\} - \log\Big\{1+\frac{1}{m}\sum_{j=1}^{m}(\mathcal{E}_j-1)\Big\} + 2\log(1-\zeta_n).
\end{aligned}
$$

Hence

$$
\mathbb{P}(\hat{d}_n \in \mathcal{D}_1) \leq \mathbb{P}\left(\log\Big\{2+\frac{1}{m}\sum_{j=1}^{m}a_j(\mathcal{E}_j-1)\Big\} - \log\Big\{1+\frac{1}{m}\sum_{j=1}^{m}(\mathcal{E}_j-1)\Big\} + 2\log(1-\zeta_n) \leq 0\right).
$$

This last probability tends to zero as soon as $m^{-1}\sum_{j=1}^{m}(\mathcal{E}_j-1) = o_P(1)$ and $m^{-1}\sum_{j=1}^{m}a_j(\mathcal{E}_j - 1) = o_P(1)$. Since $E_n(d^*,\theta^*) = m^{-1}\sum_{j=1}^{m}(\mathcal{E}_j-1)$, Proposition 2.1 implies that this term is $o_P(1)$. As in Robinson (1995b, p. 1639), it is easily checked that $\sum_{j=1}^{m}a_j = O(m)$, $\sum_{1\leq j\leq p_m}a_j^2 = O(m^{2-4\epsilon})$ and $\sum_{p_m<j\leq m}a_j^2 = O(m)$. Thus we can apply Theorem 4.1, Eq. (4.22), to obtain that $m^{-1}\sum_{j=1}^{m}a_j(\mathcal{E}_j - 2\pi I_{Z,j}) = o_P(1)$. There only remains to check that under assumption **(A1)**, $m^{-1}\sum_{j=1}^{m}a_j(2\pi I_{Z,j}-1) = o_P(1)$. Expanding this sum as the term $E_{2,n}$ in the proof of Proposition 2.1, we obtain:

$$
\frac{1}{m}\sum_{j=1}^{m}a_j(2\pi I_{n,Z}-1) = \frac{1}{m}\sum_{j=1}^{m}a_j \times \frac{1}{n}\sum_{t=1}^{n}(Z_t^2-1) + \frac{2}{n}\sum_{1\leq s<t\leq n}\frac{1}{m}\sum_{j=1}^{m}a_j\cos((t-s)x_j)Z_sZ_t
$$

Since $\sum_{j=1}^{m}a_j = O(m)$ and $\sum_{1\leq j\leq p_m}a_j^2 = O(m^{2-4\epsilon})$, under assumption **(A1)**, both these terms are $o_P(1)$. The proof is concluded.

References

[1] Andersen, T. and Bollerslev, T. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *The Journal of Finance,* **LII,** (1997a), 975-1005.

[2] Andersen, T. and Bollerslev, T. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance,* **4,** (1997b), 115-158.

[3] Andrews, D.W.K. and Y. Sun. Local polynomial Whittle approximation of long-range dependence. Cowles Foundation discussion paper 1293, (2001).

[4] Bollerslev, T. and H.O. Mikkelsen. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics* **73** (1996), 151–184.

[5] Breidt, F. J., Crato, N., and de Lima, P. The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics,* **83,** (1998), 325-348.

[6] Brillinger, D.R. *Time Series. Data Analysis and Theory.* Holden-Day Series in Time Series Analysis, Holden-Day, San Frasncisco, 1981.

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

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

[9] Harvey, A. C. Long memory in stochastic volatility. In: Knight, J., Satchell, S. (Eds.), Forecasting volatility in financial markets. Butterworth-Heinemann, London, 1998.

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

[11] Hurvich, C.M. and B.K. Ray. The local Whittle estimator of long memory stochastic volatility. Preprint (2001).

[12] Hurvich, C.M. and Ph. Soulier. Testing for long memory in volatility. To appear in *Econometric Theory.*

[13] Künsch, H. R. Statistical aspects of self-similar processes. Proceedings of the World Congress of the Bernoulli Society, Tashkent, Vol. 1, (1987), 67-74.

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

[15] Ray, B.K. and Tsay, R. Long-range dependence in daily stock volatilities. *Journal of Business and Economic Statistics* **18** (2000), 254–262.

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

[17] Robinson, P.M. Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* **24** (1995b), 1630–1661.

[18] Solo, V. Intrinsic random functions and the paradox of 1/f noise. *SIAM J. Appl. Math.* **52** (1992), 270-291.

[19] Soulier, Ph. A uniform central limit theorem for linear combinations of tapered periodogram ordinates. Application to the estimation of the long-memory parameter. Available on http://www.tsi.enst.fr/∼soulier/paper.html (2002).

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

[21] Wright, J.H. Log periodogram estimation of long memory volatility dependencies with conditionally heavy tailed returns. *Board of Governors of the Federal Reserve System, International Finance Discussion Paper Number 685*, 2000.