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 using a local Whittle criterion function. We allow the optional inclusion of an additional term to account for possible 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 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. Applications of GPH in the context of volatility have been presented in Andersen and Bollerslev (1997a,b), Ray and Tsay (2000), and Wright (2002), among others.

[*]New York University : 44 W. 4'th Street, New York NY 10012, USA
[†]École Nationale des Télécommunications, 46 rue Barrault, 75005 Paris, France
[‡]Université Paris X, 200 avenue de la République, 92001 Nanterre cedex, France

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 with variance σ_η^2, 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). Bollerslev and Mikkelsen (1999) study FIEGARCH forecasts of volatility, while Baillie, Cecen and Han (2000) study high frequency data using FIEGARCH. 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 mutually independent.

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.

Another popular estimator of the memory parameter is the Gaussian semiparametric estimator (GSE), introduced by Künsch (1987), and later studied by Robinson (1995b) for processes which are linear in a Martingale difference sequence. For the LMSV model, results analogous to those of Deo and Hurvich (2001) were obtained by Arteche (2003) for the GSE estimator, based once again on log squared returns. The use of GSE instead of GPH allows the assumption that $\{Y_t\}$ in (1.1) is Gaussian to be weakened to linearity in a Martingale difference sequence. Arteche (2003) requires the same restriction on m as in Deo and Hurvich (2001).

Sun and Phillips (2003) proposed a nonlinear log-periodogram regression estimator $\hat{d}_{\mathrm{NLP}}$ of d^*, using Fourier frequencies $1, \ldots, m$. They partially account for the noise term $\{\eta_t\}$ in (1.1), through a first-order Taylor expansion about zero of the spectral density of the observations. They establish the asymptotic normality of $m^{1/2}(\hat{d}_{\mathrm{NLP}} - d^*)$ under assumptions including $n^{-4d^*} m^{4d^*+1/2} \to$ Const. Thus, $\hat{d}_{\mathrm{NLP}}$, with a variance of order $n^{-4d^*/(4d^*+1/2)}$, converges faster than the GPH estimator, but still arbitrarily slowly if d^* is sufficiently close to zero. Sun and Phillips (2003) also assumed that the noise and signal are Gaussian. This rules out most LMSV models, since $\log e_t^2$ is typically non-Gaussian.

Recently, Hurvich and Ray (2003) have proposed a local Whittle estimator of d^*, based on log squared returns in the LMSV model. The local Whittle estimator, defined precisely in Section 2.1, may be viewed as a generalized version of the GSE estimator. Hurvich and Ray (2003) included an additional term in the Whittle criterion function to account for the contribution of the noise term $\{\eta_t\}$ in (1.1) to the low frequency behavior of the spectral density of $\{X_t\}$. The estimator is obtained from numerical optimization of the criterion function. It was found in the simulation study of Hurvich and Ray (2003) that the local Whittle estimator can strongly outperform GPH, especially in terms of bias when m is large.

We assume that the observed process $\{X_t\}$ is the sum of a long-memory signal $\{Y_t\}$ which is linear in a Martingale difference sequence $\{Z_t\}$, and a white noise $\{\eta_t\}$ which is potentially contemporaneously correlated with $\{Z_t\}$. Our signal plus noise model, made precise in Section 2 below, includes both the LMSV and FIEGARCH models as special cases, by allowing a contemporaneous correlation between the shocks in the signal and noise processes.

Many empirical studies have found estimates of the memory parameter in the log-squared returns, d^*, which are close to or even greater than $1/2$, indicating possible nonstationarity of volatility. For example, Hurvich and Ray (2003) 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 analyzing a similar data set with a parametric LMSV model, Harvey (1998), who explicitly allowed for the nonstationary case in his definition of the model, obtained an estimated memory parameter of 0.868. In view of these empirical findings, we allow in this paper for the possibility that d^* exceeds $1/2$. Specifically, we assume here that $d^* \in (0, 1)$.

In the context of our general signal plus noise model, allowing all of the generalizations described above, we will show that under suitable conditions our local Whittle estimator $\hat{d}_n$ based on the first m Fourier frequencies is consistent. Then, we will establish the $\sqrt{m}$-consistency and asymptotic normality of $\hat{d}_n$ 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 twice differentiable, 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 and GSE estimators of persistence in volatility and over the NLP regression estimator of Sun and Phillips (2003).

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 practitioners working with stochastic volatility models find the assumption to be overly restrictive.

The remainder of this paper is organized as follows. In Section 2.1, we define the local Whittle estimator $\hat{d}_n$. Section 3 presents results demonstrating the consistency of the local Whittle estimator of both d^* and of the auxiliary parameter θ^*. Section 4 gives a central limit theorem for $\hat{d}_n$. The estimates of the parameters (d^*, θ^*) converge at different rates, and in the case of the estimates of θ^* the rates 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 $\sqrt{m}(\hat{d}_n - d^*)$. In Section 5, 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. In Section 6 we report the results of a simulation study on the properties of the local Whittle estimator.

2 Definitions and notations

We generalize the model (1.1) to a potentially nonstationary signal plus noise model, in which the observed process is either

$$X_t = \begin{cases} \mu + Y_t + \eta_t, & \text{(stationary case)} \\ \mu + \sum_{s=1}^t Y_s + \eta_t, & \text{(nonstationary case)}, \end{cases} \tag{2.1}$$

$\{Y_t\}$ is a weakly stationary zero mean process and $\{\eta_t\}$ is a zero mean white noise with variance σ_η^2. Our main concern in this paper is the memory parameter of $\{X_t\}$, denoted by d^*. The stationary case corresponds to $d^* \in (0, 1/2)$ and the nonstationary case corresponds to $d^* \in [1/2, 1)$.

In the stationary case, we lose no generality in assuming that $\{Y_t\}$ 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 $\{Y_t\}$ has mean zero ensures that $\{X_t\}$ is free of linear trends. This does entail some loss of generality, but our estimator, which makes no use of differencing or tapering, is not invariant to such trends, and would presumably be adversely affected by them. In any case, deterministic trends in volatility are perhaps somewhat artificial from an economic standpoint.

We now present precise assumptions on the signal process $\{Y_t\}$. We assume first that the weakly stationary process $\{Y_t\}$ admits an infinite order moving average representation with respect to a zero mean, unit variance white noise (*i.e.* an uncorrelated second order stationary sequence) $\{Z_t\}$:

$$Y_t = \sum_{j \in \mathbb{Z}} a_j Z_{t-j}, \tag{2.2}$$

with $\sum_{j\in\mathbb{Z}} a_j^2 < \infty$. In order to guarantee that the returns are a Martingale difference sequence, one could assume that $a_j = 0$ $(j \leq 0)$. This assumption would imply that $\{r_t\}$ is adapted to the natural filtration $\{\mathcal{F}_t\}$ of $\{e_t, Z_t\}$, Y is predictable with respect to this filtration and

$$\mathbb{E}[r_t \mid \mathcal{F}_{t-1}] = \exp(Y_t/2)\mathbb{E}[e_t] = 0.$$

We do not make such an assumption here, in order to consider the problem in its fullest generality. Thus, we do not require the returns to be a Martingale difference sequence. Additional assumptions on $\{Z_t\}$ will be specified as needed.

We define $a(x) = \sum_{j\in\mathbb{Z}} a_j \mathrm{e}^{\mathrm{i}jx}$ and assume that it can be expressed as

$$a(x) = (1 - \mathrm{e}^{\mathrm{i}x})^{-d_Y} a^*(x), \quad x \in [-\pi, \pi] \setminus \{0\},$$

where $d_Y \in [-1/2, 1/2)$, a^* is a function that is continuous at 0, and $a^*(0) \neq 0$. The quantity d_Y is the memory parameter of the time series $\{Y_t\}$. The stationary case corresponds to $d_Y \in (0, 1/2)$, and the nonstationary case corresponds to $d_Y \in [-1/2, 0)$. The case $d_Y = 0$, which corresponds to short memory in volatility, will be addressed separately in Section 5.

The spectral density of $\{Y_t\}$ is given by $f_Y(x) = |a(x)|^2/(2\pi)$, and can be expressed as

$$f_Y(x) = |1 - \mathrm{e}^{\mathrm{i}x}|^{-2d_Y} f_Y^*(x), \tag{2.3}$$

with $f_Y^*(x) = |a^*(x)|^2/(2\pi)$.

The concept of pseudo spectral density has been defined for nonstationary processes. See, e.g., Solo (1992), Hurvich and Ray (1995), Velasco (1999). To generalize this concept so that it applies to our signal plus noise process $\{Y_t\}$, we first state additional assumptions on the second-order dependence structure of the bivariate sequence $\{Z_t, \eta_t\}$. Specifically, we assume that:

$$\forall t \in \mathbb{Z},\ \mathbb{E}[\eta_t Z_t] = \rho_\eta\, \sigma_\eta \quad \text{and} \quad \forall s \neq t,\ \mathbb{E}[\eta_s Z_t] = 0. \tag{2.4}$$

The parameter ρ_η accounts for the possible contemporaneous correlation between Z_t and η_t, assumed 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 $\{e_t\}$ is i.i.d. $\mathcal{N}(0,1)$, and (2.4) is in force. 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 $\{X_t\}$ defined in (2.1) is then

$$f_X(x) = \begin{cases} f_Y(x) + \frac{2\rho_\eta\sigma_\eta}{2\pi}\mathrm{Re}(a(x)) + \frac{\sigma_\eta^2}{2\pi}, & \text{(stationary case)}, \\ |1 - \mathrm{e}^{\mathrm{i}x}|^{-2} f_Y(x) + \frac{2\rho_\eta\sigma_\eta}{2\pi}\mathrm{Re}((1 - \mathrm{e}^{\mathrm{i}x})^{-1}a(x)) + \frac{\sigma_\eta^2}{2\pi}, & \text{(nonstationary case)}. \end{cases} \tag{2.5}$$

In both cases, under additional smoothness assumption on the behavior of a^* about 0 (that will be made precise in the next section), f_X admits the following expansion at 0:

$$f_X(x) \sim x^{-2d^*} f_Y^*(0) + \mathrm{Re}\left((1-\mathrm{e}^{\mathrm{i}x})^{-d^*}\right) \frac{2\,\rho_\eta \sigma_\eta \sqrt{f_Y^*(0)}}{\sqrt{2\pi}} + \frac{\sigma_\eta^2}{2\pi}, \tag{2.6}$$

with

$$d^* = \begin{cases} d_Y \in (0,1/2), & \text{(stationary case)}, \\ d_Y + 1 \in [1/2,1), & \text{(nonstationary case)}, \end{cases} \tag{2.7}$$

where the symbol $\sim$ indicates that the ratio of the left hand side to the right hand side of the above formula tends to 1 as $x \to 0^+$. Thus, in the stationary case, $\{X_t\}$ has the same memory parameter as $\{Y_t\}$, namely d_Y, while in the nonstationary case $\{X_t\}$ has the same memory parameter as the partial sum of $\{Y_t\}$, namely $d_Y + 1$.

Remark 2.1. In the stationary case where the returns are $r_t = \mathrm{e}^{Y_t/2} e_t$, and $Y_t = \sum_{j=1}^{\infty} a_j Z_{t-j}$, Surgailis and Viano (2002) have proved that under the additional assumptions that $\mathbb{E}[\mathrm{e}^{u|Z_1|}] < \infty$ for all $u > 0$ and that $\{Z_t\}$ and $\{e_t\}$ are i.i.d. sequences, the memory parameter of the series $\{|r_t|^u\}$ is the same as the memory parameter of $\{Y_t\}$. Thus, for both the LMSV and FIEGARCH models, under the above mentioned restrictions, the squared returns and the log-squared returns have the same memory parameter. In the nonstationary case, the relationship between these two memory parameters remains an open question.

2.1 The Local Whittle Estimator

Consider a covariance stationary process $\{X_t\}$ with spectral density

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

where $d^* \in (-1/2, 1/2)$ and f_X^* is a positive function which is 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. The parametric model used in GSE replaces f_X^* by a constant. This method yields a consistent and asymptotically normal estimator of $d^* \in (-1/2, 1/2)$, under mild assumptions both on f_X^* and the process $\{X_t\}$. These results were later extended to the nonstationary case $d^* \in [1/2, 1)$ by Velasco (1999) who proved the consistency for d^* in this range and asymptotic normality for $d^* \in [1/2, 3/4)$.

In some situations however, the local-to-zero parameterization of $f_X^*(x)$ by a constant may be inefficient. An example is the situation of a long-memory process observed in an additive noise. In order to improve the efficiency, one can try to fit a more complex local parametric model for f_X^*. In the local Whittle estimator, defined below in a general setting, $f_X^*(x)$ is replaced by $G(1 + h(d, \theta, x))$, where G is a positive constant and h is a function tailored to the problem at hand. The additional parameter θ can be seen as a nuisance parameter which is included to allow some flexibility in the modelling of f_X^* about 0.

The discrete Fourier transform and the periodogram ordinates of any process $\{V_t\}$ evaluated at the Fourier frequencies $x_j = 2j\pi/n$, $j = 1, \dots, n$, are respectively denoted by

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

The local Whittle contrast function, based on the observations $X_1, \dots, X_n$, is defined as

$$\hat{W}_m(d, G, \theta) = \sum_{k=1}^{m} \left\{ \log \left(G x_k^{-2d} (1 + h(d, \theta, x_k) \right) + \frac{I_{X,k}}{G x_k^{-2d} (1 + h(d, \theta, x_k))} \right\} \tag{2.8}$$

where $m < n/2$ is a bandwidth parameter (the dependence on n is implicit). Concentrating G out of $\hat{W}_m$ yields the following profile likelihood

$$\begin{aligned} \hat{J}_m(d, \theta) &= \log \left(\frac{1}{m} \sum_{k=1}^{m} \frac{x_k^{2d} I_{X,k}}{1 + h(d, \theta, x_k)} \right) + m^{-1} \sum_{k=1}^{m} \log\{x_k^{-2d} (1 + h(d, \theta, x_k))\} \\ &= \log \left(\frac{1}{m} \sum_{k=1}^{m} \frac{k^{2d} I_{X,k}}{1 + h(d, \theta, x_k)} \right) + m^{-1} \sum_{k=1}^{m} \log\{k^{-2d} (1 + h(d, \theta, x_k))\}. \end{aligned} \tag{2.9}$$

The local Whittle estimator is any minimand of the empirical contrast function $\hat{J}_m$ over the admissible set $\mathcal{D}_n \times \Theta_n$ (which may depend on the sample size n):

$$(\hat{d}_n, \hat{\theta}_n) = \arg \min_{(d,\theta) \in \mathcal{D}_n \times \Theta_n} \hat{J}_m(d, \theta). \tag{2.10}$$

Note that $(\hat{d}_n, \hat{\theta}_n)$ depends on h, $\mathcal{D}_n$ and Θ_n.

We now specify three different parameterizations that we will use for estimation of the memory parameter in the model (2.1).

(P0)

$$h(d, x) \equiv 0, \quad \mathcal{D}_n = [-1/2, 1]. \tag{2.11}$$

Here, there is no parameter θ and the definition of Θ_n is thus irrelevant. This parameterization is used for the GSE estimator.

(P1)

$$h(d, \theta, x) = \theta x^{2d}, \ \mathcal{D}_n = [\epsilon_n, 1], \quad \Theta_n = [0, \epsilon_n^{-2}], \tag{2.12}$$

where $\{\epsilon_n\}$ is a sequence that tends to zero as n tends to infinity at a rate that will be specified in the sequel. This parameterization is used for the local Whittle estimator in the LMSV model when ρ_η is known to be zero, as in Hurvich and Ray (2003). Our parameterization conforms with this model: indeed, the expansion (2.6) of the spectral (or pseudo spectral) density f_X at 0 when $\rho_\eta = 0$ can be expressed as $f_X(x) \sim x^{-2d^*} f_Y^*(0)(1 + h(d^*, \theta^*, x)$, with h as in (2.12) and

$$\theta^* = \frac{\sigma_\eta^2}{2\pi f_Y^*(0)}. \tag{2.13}$$

Note that if $d^* \in (0,1)$, the definition of $\mathcal{D}_n$ and Θ_n implies that for all sufficiently large n, we will have $d^* \in \mathcal{D}_n$ and $\theta^* \in \Theta_n$.

(P2)

$$\begin{aligned} &h(d,\theta,x) = \theta_1 x^{2d} \mathrm{Re}\left((1-\mathrm{e}^{\mathrm{i}x})^{-d}\right) + \theta_2 x^{2d}, \\ &\mathcal{D}_n = [\epsilon_n, 1], \quad \Theta_n = [-2\epsilon_n^{-1}, 2\epsilon_n^{-1}] \times [0, \epsilon_n^{-2}], \end{aligned} \tag{2.14}$$

where $\{\epsilon_n\}$ is as described above. This parameterization is used for the local Whittle estimator when ρ_η is not required to be zero, as in the FIEGARCH model and the LMSV model with contemporaneous correlation between $\{Z_t\}$ and $\{\eta_t\}$. Here again, the expansion (2.6) can be expressed as $f_X(x) \sim x^{-2d^*} f_Y^*(0)(1 + h(d^*, \theta^*, x))$, with h as in (2.14) and

$$\theta^* = (\theta_1^*, \theta_2^*) \ \text{ with } \ \theta_1^* = \frac{2\rho_\eta \sigma_\eta}{\sqrt{2\pi f_Y^*(0)}} \quad \text{and} \quad \theta_2^* = \frac{\sigma_\eta^2}{2\pi f_Y^*(0)}. \tag{2.15}$$

We denote the local Whittle estimators associated with the parameterizations (P0), (P1) and (P2) by $(\hat{d}_n^{(i)}, \hat{\theta}_n^{(i)})$, $i = 0, 1, 2$, respectively. Note that $\hat{d}_n^{(0)}$ is simply the GSE estimator, based on a parameterization which does not involve the noise term. In some of our discussions, as should be clear from the context, we reserve the term "local Whittle estimator" to refer only to the parameterizations (P1) and (P2) but not (P0).

Remark 2.2. The presence of an ϵ_n sequence tending to zero in parametrizations (P1) and (P2) allows the admissible parameter space to depend on n and to become larger as n increases. This in turn will allow us to state and prove our main theoretical results without making arbitrary restrictions on the true parameters, as is done in much of the current literature (see, e.g., Robinson 1995b). Nevertheless, if we took ϵ_n to be fixed and positive, then our main results would continue to hold as long as the true parameters lie in the corresponding admissible parameter space.

Remark 2.3. We explain here the (perhaps) surprising form of the parameterization (P2). For $x \in (0, \pi]$ and $d \in (0,1)$, it is well known that $|1 - \mathrm{e}^{\mathrm{i}x}|^{-2d} = x^{-2d}(1 + O(x^2))$, but it should be noted that

$$\mathrm{Re}\left\{(1-\mathrm{e}^{\mathrm{i}x})^{-d}\right\} = \{2\sin(x/2)\}^{-d} \cos\{d(x/2 + \pi/2)\} = x^{-d}\cos(\pi d/2)\{1 + O(x)\},$$

where the term $O(x)$ cannot be improved. Replacing $\mathrm{Re}\left\{(1-\mathrm{e}^{\mathrm{i}x})^{-d}\right\}$ with x^{-d} in (P1) would not only change the value of the parameter θ_1, but also create a bias term that would result in a slower rate of convergence for $\hat{d}_n^{(2)}$ (moreover depending on d^*) than the rate we will be able to establish below.

Remark 2.4. The correction term $h(d, \theta, x)$ in parameterizations (P1) and (P2) is the key element which allows us to attain a better rate of convergence for the local Whittle estimator, in comparison to the ordinary GSE and GPH estimators. Indeed, the use of $h(d, \theta)$ frees the optimal rate of convergence of the local Whittle estimator from an undesirable dependence on d^*, a problem faced by the ordinary GPH estimator considered in Deo and Hurvich (2001).

3 Consistency of the local Whittle estimator

In order to prove consistency of the local Whittle estimator $\hat{d}_n$, we consider the following assumptions.

(H1) $\{Z_t\}$ is a zero mean unit variance white noise such that

$$\frac{1}{n}\sum_{t=1}^{n}(Z_t^2 - 1) \xrightarrow{P} 0 \tag{3.1}$$

and for any $(s, t, u, v) \in \mathbb{N}^4$ such that $s < t$ and $u < v$, $\mathbb{E}[|Z_u Z_v Z_s Z_t|] < \infty$ and

$$\mathbb{E}[Z_u Z_v Z_s Z_t] = \begin{cases} 1 & \text{if } u = s \text{ and } t = v \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

Remark 3.1. This assumption is the weakest one under which we were able to construct our proof of consistency, and is satisfied under a variety of conditions. For instance, it is implied by assumption A3 of Robinson (1995b) which states that $\{Z_t\}$ is a martingale difference sequence satisfying $\mathbb{E}[Z_t^2|\sigma(Z_s, s < t)] = 1$ a.s. (which implies (3.2)) and strongly uniformly integrable (which implies (3.1)). Note that (3.1) holds when $\{Z_t^2\}$ is ergodic. Finally, note that (3.2) rules out the case $s = t$ since it assumes that $s < t$ and $u < v$, and therefore (3.2) does not imply that $\mathbb{E}[Z_t^4] = 1$ (which would be impossible except for a Rademacher random variable).

For reference, we recall the assumption on $\{\eta_t\}$.

(H2) $\{\eta_t\}$ is a zero mean white noise with variance σ_η^2 such that for each $s \neq t$, $\mathbb{E}[\eta_s Z_t] = 0$ and for each t, $\mathbb{E}[\eta_t Z_t] = \rho_\eta \sigma_\eta$.

Note that ρ_η is the correlation between Z_t and η_t, which is assumed to be constant.

(H3) $\{Y_t\}$ admits the linear representation (2.2) and the function $a(x) = \sum_{j\in\mathbb{Z}} a_j \mathrm{e}^{\mathrm{i}jx}$ can be expressed for $x \in [-\pi, \pi] \setminus \{0\}$ as $a(x) = (1 - \mathrm{e}^{\mathrm{i}x})^{-d_Y} a^*(x)$, where $d_Y \in (-1/2, 1/2)$, a^* is integrable over $[-\pi, \pi]$, $a^*(-x) = \overline{a^*(x)}$ for all $x \in [-\pi, \pi]$ and there exist $\vartheta \in (0, \pi]$, $\beta \in (0, 2]$ and $\mu > 0$ such that a^* is differentiable at 0 if $\beta > 1$ and for all $x \in [-\vartheta, \vartheta]$,

$$|a^*(x) - a^*(0)| \le \mu |a^*(0)||x|^\beta, \quad \beta \in (0, 1] \tag{3.3}$$

$$|a^*(x) - a^*(0) - x a^{*\prime}(0)| \le \mu |a^*(0)||x|^\beta, \quad \beta \in (1, 2]. \tag{3.4}$$

Remark 3.2. The function $(1-\mathrm{e}^{ix})^{-d}$ is defined for $d \in (-1/2, 1/2) \setminus \{0\}$ and $x \in [-\pi, \pi] \setminus \{0\}$ by

$$(1-\mathrm{e}^{\mathrm{i}x})^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(d+j)}{\Gamma(d)\, j!} \mathrm{e}^{\mathrm{i}jx}.$$

This series is absolutely convergent if $d < 0$ and converges in the mean square if $d > 0$. For $d = 0$, we set $(1-\mathrm{e}^{\mathrm{i}x})^0 \equiv 1$. Since by assumption $a^*(-x) = \overline{a^*(x)}$, thus $|a^*|^2$ is an even function. If moreover a^* is differentiable at 0, then $a^{*\prime}(0) = -\overline{a^{*\prime}(0)}$ and for all $\beta \in (0, 2]$, there exists a constant C such that for all $x \in [-\vartheta, \vartheta]$, it holds that

$$\big||a^*(x)|^2 - |a^*(0)|^2\big| \le C|x|^\beta. \tag{3.5}$$

Remark 3.3. In the related literature (Robinson (1995b), Velasco (1999), Andrews and Sun (2001)), it is usually assumed moreover that the function a is differentiable in a neighborhood of zero, except at zero, with $|xa'(x)|/|a(x)|$ bounded on this neighborhood. Hence our assumptions are weaker than those of the above references.

Theorem 3.1. *Assume* **(H1)**, **(H2)** *and* **(H3)** *and* $d^* \in [0, 1)$. *Let m be a non-decreasing sequence such that*

$$\lim_{n\to\infty} \big(m^{-1} + m/n\big) = 0, \tag{3.6}$$

and set $\epsilon_n = (\log(n/m))^{-1/2}$. Then, the local Whittle estimators $\hat{d}_n^{(i)}$, $i = 0, 1, 2$ are consistent.

Remark 3.4. Arteche (2003), Theorem 1, proved the consistency $\hat{d}_n^{(0)}$ (the GSE) for a long-memory process observed in independent noise (with no contemporaneous correlation) in the stationary case *i.e.* for $d^* \in (0, 1/2)$. Theorem 3.1 extends this result to the nonstationary case $d^* \in (1/2, 1)$ and to the case where the noise $\{\eta_t\}$ is (possibly) contemporaneously correlated with $\{Y_t\}$, thus covering the FIEGARCH model. Furthermore, Theorem 3.1 implies that $\hat{d}_n^{(1)}$ is consistent even if ρ_η is nonzero.

Remark 3.5. The local Whittle estimator $\hat{d}_n^{(i)}$, $i = 1, 2$ is consistent if $d^* = 0$, but by construction its rate of convergence is at most ϵ_n. The necessity of introducing the sequence ϵ_n comes from certain technicalities in the proof of Theorem 1. We do not know if it is possible to define $\mathcal{D}_n = [0, 1]$ under (P1) and (P2). In any case, if $d^* = 0$, the parameter θ^* need not be identifiable. Thus, if a small value of $\hat{d}_n^{(i)}$ is obtained, it should be better to test for $d^* = 0$ by using the standard GSE. We will establish the validity of this procedure in section 5.

Theorem 3.1 provides no information about the behavior of $\hat{\theta}_n^{(i)}$, $i = 1, 2$. This is because, as $n \to \infty$, the objective function becomes flat as a function of θ. Thus a special technique is needed to prove the consistency of $\hat{\theta}_n^{(i)}$, $i = 1, 2$ which requires strengthened assumptions. This technique was first used in a similar context by Sun and Phillips (2003). We now introduce these assumptions.

(H4) $\{Z_t\}$ is a martingale difference sequence such that for all t, $\mathbb{E}[Z_t^4] = \mu_4 < \infty$ and $\mathbb{E}[Z_t^2 \mid \sigma(Z_s, s < t)] = 1$ a.s.

Remark 3.6. **(H4)** implies **(H1)**. More precisely, it implies that $\{Z_t^2 - 1\}$ is a square integrable martingale difference sequence and that $n^{-1}\sum_{t=1}^{n}(Z_t^2 - 1) = O_P(n^{-1/2})$.

(H5) $\{\eta_t\}$ is a zero mean white noise with variance σ_η^2 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 $u < v$,

$$\mathbb{E}[\eta_u\eta_v\eta_s\eta_t] = \begin{cases} \sigma_\eta^4 & \text{if } u = s \text{ and } t = v \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

$$\operatorname{cum}(Z_{t_1}, Z_{t_2}, \eta_{t_3}, \eta_{t_4}) = \begin{cases} \kappa & \text{if } t_1 = t_2 = t_3 = t_4, \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

Theorem 3.2. *Assume* **(H2)**, **(H3)**, **(H4)** *and* **(H5)** *and* $d^* \in (0,1)$. *Let* m *be a non-decreasing sequence of integers such that*

$$\lim_{n\to\infty}\left(m^{-4d^*-1+\delta}n^{4d^*} + n^{-2\beta}m^{2\beta+1}\log^2(m)\right) = 0 \tag{3.9}$$

for some arbitrarily small $\delta > 0$ *and set* $\epsilon_n = (\log(n/m))^{-1/2}$. *Then* $\hat{d}_n^{(i)} - d^* = O_P((m/n)^{2d^*})$ *and* $\hat{\theta}_n^{(i)} - \theta^* = o_P(1)$, $i = 1,2$.

Remark 3.7. The first term in (3.9) imposes a lower bound on the allowable value of m, requiring that m tend to ∞ faster than $n^{4d^*/(4d^*+1)}$. This condition can be fulfilled only when $\beta > 2d^*$. Note that $\beta > 2d^*$ always holds if $\beta = 2$, 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 $n^{4d^*/(4d^*+1)}$. When $\beta \le 2d^*$, then it is no longer possible to prove that $\hat{\theta}_n$ is a consistent estimator of θ^*, and the proposed local Whittle estimator will not perform better than the standard GSE.

4 Asymptotic normality of the local Whittle estimator

We focus here on $\hat{d}_n^{(i)}$ and $\hat{\theta}_n^{(i)}$, $i = 1,2$. Corresponding results for the GSE estimator will be presented in Section 5. The main result of this section is Theorem 4.3, which is a central limit theorem for the vector of estimated parameters. Before presenting this main result, we outline the main steps in the proof, which are not completely standard due to the asymptotic flatness of the objective function.

For ease of notation here, in the discussion below, we omit the superscript in $\hat{d}_n^{(i)}$. Contrary to standard statistical theory of minimum contrast estimators, the rates of convergence of $\hat{d}_n - d^*$ and of $\hat{\theta}_n - \theta^*$ are different, where d^* is defined in (2.7) and θ^* is defined in (2.13) in the LMSV case and (2.15) in the FIEGARCH case. To account for the difference in these rates, we prove that $D_n^*(\hat{d}_n - d^*, \hat{\theta}_n - \theta^*)$ is asymptotically normal with zero mean, where D_n^* is a deterministic diagonal matrix whose diagonal entries tend to ∞ at different rates, as defined below. Our proof starts with a second order Taylor expansion of the contrast function. The gradient of the contrast function evaluated at the estimates vanishes, since they are consistent and converge to an interior point of the parameter set.

Denote $H_n(d,\theta) = \int_0^1 \nabla^2 \hat{J}_m(d^* + s(d-d^*), \theta^* + s(\theta-\theta^*))ds$. With this notation, a first order Taylor expansion yields

$$0 = m{D_n^*}^{-1}\nabla\hat{J}_m(\hat{d}_n,\hat{\theta}_n) = \\ m{D_n^*}^{-1}\nabla\hat{J}_m(d^*,\theta^*) + m{D_n^*}^{-1}H_n(\hat{d}_n,\hat{\theta}_n){D_n^*}^{-1}D_n^*\left((\hat{d}_n,\hat{\theta}_n)-(d^*,\theta^*)\right). \quad (4.1)$$

The next step is to prove that $m{D_n^*}^{-1}\nabla\hat{J}_m(d^*,\theta^*)$ converges in distribution to a non-degenerate Gaussian random variable with zero mean and $m{D_n^*}^{-1}H_n(\hat{d}_n,\hat{\theta}_n){D_n^*}^{-1}$ converges in probability to a non-singular matrix. This is stated in the following two propositions.

Proposition 4.1. *Assume* **(H2)**, **(H3)**, **(H4)** *and* **(H5)**. *If $d^* \in (0,3/4)$, $\beta > 2d^*$ and m is a non-decreasing sequence of integers such that* (3.9) *holds, then $m{D_n^*}^{-1}\nabla\hat{J}_m(d^*,\theta^*)$ converges to the Gaussian distribution with zero mean and variance Γ^* with*
(i) $D_n^* = m^{1/2}\mathrm{Diag}\left(1, x_m^{2d^*}\right)$ and

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

under (P1), assuming ρ_η is known to be 0;
(ii) $D_n^* = m^{1/2}\mathrm{Diag}\left(1, x_m^{d^*}/\cos(\pi d^*/2), x_m^{2d^*}\right)$ and

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

under (P2).

Proposition 4.2. *Assume* **(H2)**, **(H3)** *and* **(H4)**. *If $d^* \in (0,3/4)$, $\beta > 2d^*$ and m is a non-decreasing sequence of integers that satisfies* (3.9), *then $m{D_n^*}^{-1}H_n(d,\theta){D_n^*}^{-1}$ converges in probability to Γ^*, uniformly with respect to $(d,\theta) \in \{d : |d-d^*| \le \log^{-5}(m)\} \times \Theta$, with D_n^* and Γ^* defined as in Proposition 4.1.*

Remark 4.1. An important feature is that Γ^* does not depend on the parameter θ^*. This was already noticed by Andrews and Sun (2001) in the context of local polynomial approximation.

Since the matrix Γ^* is invertible, the matrix ${D_n^*}^{-1}H_n(\hat{d}_n,\hat{\theta}_n){D_n^*}^{-1}$ is invertible, with probability tending to one. Hence (4.1) yields:

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

This and Propositions 4.1 and 4.2 yield our main result. Recall that our estimators are defined in section 2.1 with $\epsilon_n = (\log(n/m))^{-1/2}$.

Theorem 4.3. *Assume* **(H2)**, **(H3)**, **(H4)**, **(H5)**, $d^* \in (0, 3/4)$, $\beta > 2d^*$ *and let* m *be a non-decreasing sequence of integers that satisfies* (3.9)*. Then, under (P1), assuming* ρ_η *is known to be zero,* $D_n^* \left((\hat{d}_n^{(1)}, \hat{\theta}_n^{(1)}) - (d^*, \theta^*) \right)$ *is asymptotically Gaussian with zero mean and covariance matrix*

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

Under (P2), $D_n^* \left((\hat{d}_n^{(2)}, \hat{\theta}_n^{(2)}) - (d^*, \theta^*) \right)$ *is asymptotically Gaussian with zero mean and covariance matrix*

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

Corollary 4.4. *Under the assumptions of Theorem 4.3,* $m^{1/2}(\hat{d}_n^{(1)} - d^*)$ *is asymptotically Gaussian with zero mean and variance*

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

if ρ_η *is known to be 0; and* $m^{1/2}(\hat{d}_n^{(2)} - d^*)$ *is asymptotically Gaussian with zero mean and variance*

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

Remark 4.2. It is seen that the local Whittle estimator $\hat{d}_n^{(i)}$, $i = 1, 2$ is able to attain the same rate of convergence under the signal plus noise model (1.1) as that attained by the standard GSE $d_n^{(0)}$ in the case of no noise, as long as $\beta > 2d^*$.

Remark 4.3. The asymptotic variance of $\hat{d}_n^{(i)}$ increases when d^* is small, but this loss is compensated by the gain in the rate of convergence with respect to the standard GSE (and GPH).

5 Asymptotic normality of the standard GSE

Theorem 3.1 states that the GSE is consistent if $d^* \in (0, 1)$. We now state that it is asymptotically normal if $d^* \in (0, 3/4)$ but with a rate of convergence slower than the local Whittle estimator considered above.

Theorem 5.1. *Under the assumptions of Theorem 4.3, if* $d^* \in (0, 3/4)$ *and* m *satisfies*

$$\lim_{n \to \infty} \left(m^{-1} + m^{2\gamma^*+1} n^{-2\gamma^*} \right) = 0, \tag{5.1}$$

with $\gamma^* = d^*$ *if* $\rho_\eta \neq 0$ *and* $\gamma^* = 2d^*$ *if* $\rho_\eta = 0$*; then* $m^{1/2}(\hat{d}_n^{(0)} - d^*)$ *is asymptotically normal with variance* $1/4$.

Remark 5.1. If ρ_η is zero, then (5.1) requires that $m = o(n^{4d^*/(4d^*+1)})$, so the upper bound on m to ensure $\sqrt{m}$-consistency of GSE is essentially the same as that required for GPH by Deo and Hurvich (2001). If, however, $\rho_\eta \neq 0$, then the upper bound on m for GSE becomes more stringent, $m = o(n^{2d^*/(2d^*+1)})$, since the nonzero value of ρ_η increases the asymptotic bias of the GSE. Similar restrictions would presumably apply in this situation for GPH.

When $d^* = 0$, the theory of Robinson (1995b) cannot be directly applied to prove consistency and asymptotic normality of $\hat{d}_n^{(0)}$, since the process $X_t = Y_t + \eta_t$ is not necessarily linear with respect to a martingale difference sequence. Nevertheless, if we strengthen the assumptions on the noise $\{\eta_t\}$, we can prove the consistency and asymptotic normality of $\hat{d}_n^{(0)}$ when $d_Y = d^* = 0$.

Theorem 5.2. *Assume that* $d^* = d_Y = 0$, **(H2)**, **(H3)**, **(H4)**, **(H5)** *and that* η *is a martingale difference sequence that satisfies*

$$\operatorname{cum}(Z_u, Z_v, Z_s, \eta_t) = \gamma \quad \text{if } s = t = u = v \text{ and } 0 \text{ otherwise.} \tag{5.2}$$

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* $m^{1/2}\hat{d}_n^{(0)}$ *is asymptotically Gaussian with zero mean and variance 1/4.*

These results yield a test for long memory in volatility based on the standard GSE estimator, since Theorem 5.2 gives the asymptotic distribution of $m^{1/2}\hat{d}_n^{(0)}$ under the null hypothesis $d^* = 0$ and Theorem 5.1 shows that $m^{1/2}\hat{d}_n^{(0)} \to \infty$ if $d^* > 0$. Another test for long memory in volatility, based on the ordinary GPH estimator, was justified by Hurvich and Soulier (2002). 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.

6 Simulations

We present here some simulation results on the performance of the proposed local Whittle estimator, denoted here by LW. A comprehensive simulation study on the LW estimator was performed by Hurvich and Ray (2003), who included a proposal for constructing accurate finite-sample standard errors for LW. The concise set of results we present here was generated in the preparation of Hurvich and Ray (2003), but not reported there due to lack of space.

For each of three sample sizes ($n = 1000$, $n = 5000$, $n = 10000$), and for each of two values ($nsr = 5$, $nsr = 10$) of the noise to signal ratio $nsr = \sigma_\eta^2/(2\pi f_Y^*(0))$, 1000 realizations were generated from an LMSV model with standard Gaussian shocks e, and signal process $\{Y_t\}$ given by the ARFIMA$(1, d^*, 0)$ model $(1-B)^{d^*}(1-\phi B)Y_t = Z_t$, where $\phi = 0.8$ and $d^* = 0.4$. The innovations Z_t were *iid* Gaussian with mean zero, and variance chosen such that the specified value of nsr was obtained. Since the e_t are standard Gaussian, we have $\sigma_\eta^2 = \pi^2/2$. The profile likelihood (2.9) with parameterization (2.12) was minimized, for $d \in [.01, .75]$ and $\theta \in [e^{-8}, e^{20}]$. Note that the admissible parameter space here does not depend on n, so the ϵ_n sequence is fixed. See Remark 2.2.

Table 1 reports the bias, standard error (SE) and root mean squared error (RMSE) for LW, as well as the GPH estimator of Geweke and Porter Hudak (1983), and the bias reduced local polynomial log periodogram regression estimator of Andrews and Guggenberger (2003), denoted by AG. It was shown in Andrews and Guggenberger (2003) that the AG estimator has improved bias properties compared to GPH for Gaussian processes if the spectral density of the observations is sufficiently smooth. We used the simplest version of AG, in which a single additional term x_j^2 is included in the log periodogram regression. All three estimates of the memory parameter were constructed from the simulated log squared return series, $\{X_t\}_t$. For each realization and each estimator, three different bandwidths were considered ($m = [n^{0.6}]$, $m = [n^{0.7}]$, $m = [n^{0.8}]$).

Both GPH and AG suffer from negative bias, which worsens significantly as m or nsr is increased, presumably due to the noise term η that neither of these estimators was designed to explicitly account for. On the other hand, the bias of LW is stable with respect to nsr, and increases only modestly in m, due to the autoregressive component in the model. In most cases, LW is the best estimator in terms of RMSE, though LW has a higher standard error than GPH and AG. Overall, these results are consistent with existing theory.

APPENDIX: Proofs

The estimators introduced in section 2.1 are minimum contrast estimators. Empirical processes are the main tools in the study of such estimators. Since the Whittle contrast is based on the spectral density of a second order stationary time series, the empirical process involved is often referred to as the empirical spectral process. See for instance Dahlhaus and Polonik (2002) or Soulier (2002). In the first section of this appendix, we state two Propositions which provide the tools to derive the asymptotic properties of minimum contrast estimators: a uniform weak law of large numbers and a central limit theorem for the spectral empirical process. Their proof is very technical and is postponed to Appendix F. Using these tools, we prove our main results in the following sections. Appendix B and C deal with the main statistical issues of this paper, namely the consistency of the estimators of d^* and θ^*. The proof of the consistency of $\hat{d}_n$, in Appendix B, is essentially the same as the original proof of Robinson (1995b), but is more concise here thanks to the use of Proposition A.1. The proof of the consistency of $\hat{\theta}_n$, in Appendix C is rather involved. We have tried to make it clear, though concise. It is the longest and more difficult part of this proof. Appendices D and E contain the proof of the asymptotic normality results, which are quite standard and made very short by again referring to Propositions A.1 and A.2.

A Results for the empirical spectral process

Define

$$f_{X,k} = x_k^{-2d^*} f_Y^*(0)\{1 + h(d^*, \theta^*, x_k)\} \tag{A.1}$$

where the function h is defined either in (P0), (P1) or (P2), and for a positive integer m and $c = (c_1, \dots, c_m) \in \mathbb{R}^m$,

$$\mathcal{Z}_m(c) = \sum_{k=1}^{m} c_k \{f_{X,k}^{-1} I_{X,k} - 1\}. \tag{A.2}$$

For $\epsilon \in (0, 1]$ and $K > 0$, let $\mathcal{C}_m(\epsilon, K)$ be the subset of vectors $c \in \mathbb{R}^m$ such that

$$\text{for all } k \in \{1, \dots, m-1\}, \quad |c_k - c_{k+1}| \le K m^{-\epsilon} k^{\epsilon - 2}, \quad |c_m| \le K m^{-1}. \tag{A.3}$$

Proposition A.1 (Uniform weak law of large numbers).

1. *Assume* **(H1)**, **(H2)** *and* **(H3)**. *Let m be a non-decreasing sequence of integers such that $\lim_{n\to\infty}\{m\,n^{-1} + m^{-1}\} = 0$. Then, for any $\epsilon \in (0, 1)$, any constant $K < \infty$ and any $d^* \in (0, 1)$,*

$$\sup_{c \in \mathcal{C}_m(\epsilon, K)} \mathcal{Z}_m(c) = o_P(1).$$

2. *Assume moreover that* **(H4)**, **(H5)** *and one of the following assumptions hold.*

   (2.i) h *is given by (P1),* $\rho_\eta = 0$, $d^* \in (0, 3/4)$ *and* m *satisfies*

$$\lim_{n\to\infty} \left(m^{-1} + m^{2\beta+1}\log^2(n)n^{-2\beta}\right) = 0; \tag{A.4}$$

   (2.ii) h *is given by (P2),* $d^* \in (0, 3/4)$ *and* m *satisfies (A.4);*

   (2.iii) $h \equiv 0$, $d^* \in (0, 3/4)$ *and* m *satisfies*

$$\lim_{n\to\infty} \left(m^{-1} + m^{2\gamma^*+1}n^{-2\gamma^*}\right) = 0, \tag{A.5}$$

   with $\gamma^* = d^*$ *if* $\rho_\eta \neq 0$ *and* $\gamma^* = 2d^*$ *if* $\rho_\eta = 0$*;*

   (2.iv) $d^* = 0$, $h \equiv 2\rho_\eta\sigma_\eta/\sqrt{f_Y^*(0)/(2\pi)} + \sigma_\eta^2/(2\pi f_Y(0))$, η *satisfies the assumptions of Theorem 5.2 and* m *satisfies* (A.4).

   Then for all $\epsilon \in (0, 1]$ *there exists a constant* C *such that, for all* $K > 0$

$$\mathbb{E}\left[\sup_{c\in\mathcal{C}_m(\epsilon,K)} |\mathcal{Z}_m(c)|\right] \leq CKm^{-(1/2\wedge\epsilon)}\log^\delta(m), \tag{A.6}$$

   with $\delta = 1$ *if* $\epsilon = 1/2$ *and* $\delta = 0$ *otherwise.*

Proposition A.2. *Assume* **(H2)**, **(H3)**, **(H4)** *and* **(H5)**. *Let* m *be a non-decreasing sequence of integers and let* $(c_{m,k})_{1\leq k\leq m}$ *be a triangular array of real numbers that satisfy*

$$\sum_{k=1}^m c_{m,k} = 0 \quad \text{and} \quad \sum_{k=1}^m c_{m,k}^2 = 1 \tag{A.7}$$

$$\lim_{n\to\infty} \left\{\sum_{k=1}^m |c_{m,k} - c_{m,k+1}| + |c_{m,n}|\right\}^2 \log(n) = 0. \tag{A.8}$$

Assume either (2.i), (2.ii), (2.iii) or (2.iv) of Proposition A.1. Then $\sum_{k=1}^m c_{m,k} f_{X,k}^{-1} I_{X,k}$ *is asymptotically standard Gaussian.*

B Proof of Theorem 3.1 and of the consistency part of Theorem 5.2

In this section, we prove Theorem 3.1 and the consistency part of Theorem 5.2. This proof only uses the first part of Proposition A.1, and is valid for each of the four cases considered. The only difference between them is the remainder term $R_m(d, \theta)$ (defined below) which is identically zero in the case of the standard GSE, and which converges uniformly to zero over $\mathcal{D}_n \times \Theta_n$ in the case of the local Whittle estimator. Therefore, we omit the superscript in the notation of the estimators. Define

$$\mathcal{D}_{1,n} = (-\infty, d^* - 1/2 + \epsilon) \cap \mathcal{D}_n, \tag{B.1}$$
$$\mathcal{D}_{2,n} = [d^* - 1/2 + \epsilon, +\infty) \cap \mathcal{D}_n \tag{B.2}$$

where $\epsilon < 1/4$ is a positive real number to be set later and $\mathcal{D}_n$ is defined in (2.11), (2.12) or (2.14). As originally done in Robinson (1995b), we separately prove that $\lim_{n\to\infty} \mathbb{P}(\hat{d}_n \in \mathcal{D}_{1,n}) = 0$ and that $(\hat{d}_n - d^*)\mathbf{1}_{\mathcal{D}_{2,n}}(\hat{d}_n)$ tends to zero in probability. Note that $\mathcal{D}_{1,n}$ is empty if it is assumed that $d^* \in (0, 1/2)$ and ϵ is chosen small enough. We first prove that $(\hat{d}_n - d^*)\mathbf{1}_{\mathcal{D}_{2,n}}(\hat{d}_n)$ tends to zero in probability. Denote

$$\alpha_k(d,\theta) = \frac{1 + h(d^*, \theta^*, x_k)}{1 + h(d, \theta, x_k)},$$
$$\gamma_{m,k}(d,\theta) = \frac{k^{2d-2d^*}\alpha_k(d,\theta)}{\sum_{j=1}^m j^{2d-2d^*}\alpha_j(d,\theta)}, \gamma_m(d,\theta) = (\gamma_{m,k}(d,\theta))_{1\le k\le m},$$
$$J_m(d,\theta) = \log\left(\frac{1}{m}\sum_{k=1}^m x_k^{2d-2d^*}\alpha_k(d,\theta)\right) + \frac{1}{m}\sum_{k=1}^m \log\left(x_k^{-2d}\{1 + h(d,\theta,x_k)\}\right).$$

With this notation and the notation introduced in Appendix A, we get:

$$\hat{J}_m(d,\theta) = \log(1 + \mathcal{Z}_m(\gamma_m(d,\theta))) + J_m(d,\theta) + \log(f_Y^*(0)). \tag{B.3}$$

Due to the strict concavity of the log function, for any positive integer m and positive real numbers $a_1, \dots, a_m$, it holds that

$$\log\left(\frac{1}{m}\sum_{k=1}^m a_k\right) \ge \frac{1}{m}\sum_{k=1}^m \log(a_k).$$

Thus, (d^*, θ^*) minimizes J_m. Moreover, by definition, $(\hat{d}_n, \hat{\theta}_n)$ minimizes $\hat{J}_m$. Hence on the event $\{\hat{d}_n \in \mathcal{D}_{2,n}\}$,

$$\begin{aligned}
0 &\le J_m(\hat{d}_n, \hat{\theta}_n) - J_m(d^*, \theta^*) \\
&= J_m(\hat{d}_n, \hat{\theta}_n) - \hat{J}_m(\hat{d}_n, \hat{\theta}_n) + \hat{J}_m(\hat{d}_n, \hat{\theta}_n) - \hat{J}_m(d^*, \theta^*) + \hat{J}_m(d^*, \theta^*) - J_m(d^*, \theta^*) \\
&\le \log\{1 + \mathcal{Z}_m(\gamma_m(d^*, \theta^*))\} - \log\{1 + \mathcal{Z}_m(\gamma_m(\hat{d}_n, \hat{\theta}_n))\} \qquad \text{(B.4)} \\
&\le 2 \sup_{(d,\theta)\in\mathcal{D}_{2,n}\times\Theta_n} |\log\{1 + \mathcal{Z}_m(\gamma_m(d,\theta))\}|. \qquad \text{(B.5)}
\end{aligned}$$

Define

$$K_m(s) = \log\left(\frac{1}{m}\sum_{k=1}^m k^{2s}\right) - \frac{2s}{m}\sum_{k=1}^m \log(k).$$

The function K_m is twice differentiable on $(-1,\infty)$, $K_m'(0)=0$ and $s \mapsto K_m''(s)$ is bounded away from zero on compact subsets of $(-1,\infty)$. Thus, there exists a constant $c>0$ such that for all $m \geq 2$ and $d \in \mathcal{D}_{2,n}$,

$$K_m(d-d^*) \geq c(d-d^*)^2. \tag{B.6}$$

Hence, defining $R_m(d,\theta) = J_m(d,\theta) - J_m(d^*,\theta^*) - K_m(d-d^*)$, we obtain:

$$\begin{aligned} 0 \leq (\hat{d}_n - d^*)^2 \mathbf{1}_{\mathcal{D}_{2,n}}(\hat{d}_n) &\leq c^{-1} K_m(\hat{d}_n - d^*) \leq c^{-1} J_m(\hat{d}_n,\hat{\theta}_n) - c^{-1} J_m(d^*,\theta^*) - c^{-1} R_m(\hat{d}_n,\hat{\theta}_n) \\ &\leq 2c^{-1} \sup_{(d,\theta)\in\mathcal{D}_{2,n}\times\Theta_n} |\log\{1+\mathcal{Z}_m(\gamma_m(d,\theta))\}| - c^{-1} R_m(\hat{d}_n,\hat{\theta}_n). \end{aligned} \tag{B.7}$$

To bound R_m, note that it can be expressed as

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

Under (P0), $h(d,\theta,x) \equiv 0$ which implies $\alpha_k(d,\theta) \equiv 1$; thus, $R_m \equiv 0$. Under (P1) or (P2), there exist constants c, C such that

$$\sup_{k\in\{1,\dots,m\}} \sup_{(d,\theta)\in\mathcal{D}_n\times\Theta_n} |\alpha_k(d,\theta)-1| \leq C\log^2(n/m)\,\mathrm{e}^{-\sqrt{c\log(n/m)}}.$$

Hence we obtain

$$\begin{aligned} \sup_{(d,\theta)\in\mathcal{D}_n\times\Theta_n} |R_m(d,\theta)| &\leq C \sup_{(d,\theta)\in\mathcal{D}_n\times\Theta_n} \sup_{k=1,\dots,m} |\alpha_k(d,\theta)-1| \\ &\leq C\log^2(n/m)\,\mathrm{e}^{-\sqrt{c\log(n/m)}} = o(1). \end{aligned} \tag{B.8}$$

Note that this last bound is valid even when $d^*=0$, but that we cannot bound conveniently $R_m(d,\theta)$ if d is not bonded away from zero by ϵ_n, because the convergence of $R_m(d,\theta)$ to zero is not uniform on $[0,1]\times\Theta_n$, even if Θ_n were bounded. To conclude, we now show that there exists a constant K such that, for all $(d,\theta)\in\mathcal{D}_{2,n}\times\Theta_n$, the sequence $\gamma_m(d,\theta)\in\mathcal{C}_m(2\epsilon,K)$. The argument is the same as implicitly used in the proof of Theorem 1 in Robinson (1995b). Since we will reuse this argument later, we give a more detailed proof than needed at present. Note first that there exists a constant C such that, for all $(d,\theta)\in\mathcal{D}_{2,n}\times\Theta_n$,

$$\sum_{k=1}^m k^{2d-2d^*}\alpha_k(d,\theta) \geq Cm^{2d-2d^*+1} \tag{B.9}$$

$$\begin{aligned} &\left|k^{2d-2d^*}\alpha_k(d,\theta) - (k+1)^{2d-2d^*}\alpha_{k+1}(d,\theta)\right| \\ &\quad\leq \left|k^{2d-2d^*} - (k+1)^{2d-2d^*}\right|\alpha_k(d,\theta) + (k+1)^{2d-2d^*}\,|\alpha_k(d,\theta)-\alpha_{k+1}(d,\theta)| \\ &\quad\leq Ck^{2d-2d^*-1}\left\{|d-d^*| + x_m^{\gamma^*}\log(n/m)\right\}, \end{aligned} \tag{B.10}$$

with $\gamma^* = 2d^*$ under (P1) if $\rho_\eta = 0$ and $\gamma^* = d^*$ under (P2). Gathering (B.9) and (B.10) yields

$$\sup_{(d,\theta)\in\mathcal{D}_{2,n}\times\Theta_n} |\gamma_{m,k+1}(d,\theta) - \gamma_{m,k}(d,\theta)| \le Ck^{2\epsilon-2}m^{-2\epsilon}.$$

It is also easily seen that $\gamma_{m,m}(d,\theta) \ge Cm^{-1}$, uniformly over $(d,\theta) \in \mathcal{D}_{2,n} \times \Theta_n$. Thus there exists a constant K such that, for all $(d,\theta) \in \mathcal{D}_{2,n} \times \Theta_n$, the sequence $\gamma_m(d,\theta)$ is in the class $\mathcal{C}_m(2\epsilon, K)$, and applying Proposition A.1, we obtain that $(\hat{d}_n - d^*)\mathbf{1}_{\mathcal{D}_{2,n}}(\hat{d}_n) = o_P(1)$.

We now prove that $\lim_{n\to\infty} \mathbb{P}(\hat{d}_n \in \mathcal{D}_{1,n}) = 0$. Define $p_m = (m!)^{1/m}$. For $d \in \mathcal{D}_{1,n}$, if $1 \le j \le p_m$, then $(j/p_m)^{2d-2d^*} \ge (j/p_m)^{-1+2\epsilon}$ and if $p_m < j \le m$, then $(j/p_m)^{2d-2d^*} \ge (j/p_m)^{2\epsilon_n-2d^*}$. Define then $a_{m,j} = m^{-1}(j/p_m)^{-1+2\epsilon}$ if $1 \le j \le p_m$, $a_{m,j} = m^{-1}(j/p_m)^{2\epsilon-2d^*}$ otherwise and $a_m = (a_{m,j})_{1\le j\le m}$. As shown in Robinson (1995b, Eq. 3.22), if $\epsilon < 1/(4\mathrm{e})$, then for large enough n, $\sum_{j=1}^m a_{m,j} \ge 2$. Define $\mathcal{E}_j = f_{X,j}^{-1} I_{X,j}$ and $\zeta_n = \mathrm{e}^{-\sqrt{\log(n/m)}}$. We obtain:

$$\begin{aligned}
&\hat{J}_m(d,\theta) - \hat{J}_m(d^*,\theta^*) \\
&\qquad = \log\Big\{\frac{1}{m}\sum_{j=1}^m \Big(\frac{j}{p_m}\Big)^{2d-2d^*} \alpha_j(d,\theta)\mathcal{E}_j\Big\} - \log\Big\{\frac{1}{m}\sum_{j=1}^m \mathcal{E}_j\Big\} - m^{-1}\sum_{k=1}^m \log(\alpha_k(d,\theta)) \\
&\qquad \ge \log\Big\{\sum_{j=1}^m a_{m,j}\mathcal{E}_j\Big\} - \log\Big\{\frac{1}{m}\sum_{j=1}^m \mathcal{E}_j\Big\} + \log(1 - C\zeta_n) - \log(1 + C\zeta_n) \\
&\qquad \ge \log\Big\{2 + \mathcal{Z}_m(a_m)\Big\} - \log\Big\{1 + \mathcal{Z}_m(u_m)\Big\} + 2\log(1 - C\zeta_n),
\end{aligned}$$

where we have defined $u_m = (m^{-1}, \ldots, m^{-1}) \in \mathbb{R}^m$. Hence

$$\begin{aligned}
\mathbb{P}(\hat{d}_n \in \mathcal{D}_{1,n}) &\le \mathbb{P}\left(\inf_{(d,\theta)\in\mathcal{D}_{1,n}\times\Theta_n} \hat{J}_m(d,\theta) - \hat{J}_m(d^*,\theta^*) \le 0\right) \\
&\le \mathbb{P}\left(\log\Big\{2 + \mathcal{Z}_m(a_m)\Big\} - \log\Big\{1 + \mathcal{Z}_m(u_m)\Big\} + 2\log(1 - \zeta_n)) \le 0\right).
\end{aligned}$$

The sequences a_m and u_m belong to $\mathcal{C}_m(2\epsilon, K)$ for some constant K, hence, applying Proposition A.1, we obtain that $\lim_{n\to\infty} \mathbb{P}(\hat{d}_n \in \mathcal{D}_{1,n}) = 0$, which concludes the proof.

C Proof of Theorem 3.2

Throughout this section, the assumptions of Theorem 3.2 are in force. Recall that we only consider parameterizations (P1) and (P2). For notational clarity, we omit the superscript in

$\hat{d}_n^{(i)}$. For $\alpha, \beta \geq 0$, define

$$\rho_m(\alpha,\beta) = \frac{1}{m}\sum_{k=1}^m x_k^\alpha x_k^\beta - \frac{1}{m}\sum_{k=1}^m x_k^\alpha \times \frac{1}{m}\sum_{k=1}^m x_k^\beta,$$

$$\rho'_m(\alpha,\beta) = \frac{1}{m}\sum_{k=1}^m x_k^\alpha \mathrm{Re}\left\{(1-\mathrm{e}^{\mathrm{i}x_k})^\beta\right\} - \frac{1}{m}\sum_{k=1}^m x_k^\alpha \times \frac{1}{m}\sum_{k=1}^m \mathrm{Re}\left\{(1-\mathrm{e}^{\mathrm{i}x_k})^\beta\right\},$$

$$\begin{aligned}\rho''_m(\alpha,\beta) &= \frac{1}{m}\sum_{k=1}^m \mathrm{Re}\left\{(1-\mathrm{e}^{\mathrm{i}x_k})^\alpha\right\}\mathrm{Re}\left\{(1-\mathrm{e}^{\mathrm{i}x_k})^\beta\right\} \\ &\quad - \frac{1}{m}\sum_{k=1}^m \mathrm{Re}\left\{(1-\mathrm{e}^{\mathrm{i}x_k})^\alpha\right\} \times \frac{1}{m}\sum_{k=1}^m \mathrm{Re}\left\{(1-\mathrm{e}^{\mathrm{i}x_k})^\beta\right\},\end{aligned}$$

with the convention that if $\alpha = 0$, x_k^α is replaced by $\log(k)$. These coefficients can be viewed as empirical covariances, so that for any $0 \leq \alpha_1 < \cdots < \alpha_{k+q}$, the symmetric matrix M with entries $M_{i,j} = \rho_m(\alpha_i, \alpha_j)$ if $1 \leq i,j \leq k$, $M_{i,j} = M_{j,i} = \rho'_m(\alpha_i,\alpha_j)$ if $1 \leq i \leq k$ and $k+1 \leq j \leq k+q$ and $M_{i,j} = \rho''_m(\alpha_i,\alpha_j)$ if $k+1 \leq i,j \leq k+q$, is positive definite.

Lemma C.1. *Let m be a non-decreasing sequence such that $\lim_{n\to\infty} m = \infty$. Then, for $\alpha > 0$, we have the following limits:*

$$\begin{aligned}\lim_{n\to\infty} x_m^{-\alpha-\beta}\rho_m(\alpha,\beta) &= \lim_{n\to\infty} x_m^{-\alpha-\beta}\rho'_m(\alpha,\beta)/\cos(\pi\beta/2) \\ &= \lim_{n\to\infty} x_m^{-\alpha-\beta}\rho''_m(\alpha,\beta)/\{\cos(\pi\alpha/2)\cos(\pi\beta/2)\} = \frac{\alpha\beta}{(1+\alpha)(1+\beta)(1+\alpha+\beta)}.\end{aligned}$$

Before proceeding, note now that under Assumption 3.2, the sequences $\log(n)$, $\log(m)$ and $\log(n/m)$ are of the same order of magnitude, in the sense that the ratio of any two of them is bounded. Therefore, whenever one of these sequences is involved, we will freely use the most convenient way to denote it. The first step in the proof of Theorem 3.1 is to prove a logarithmic rate of convergence for $\hat{d}_n$.

Lemma C.2.

$$\hat{d}_n - d^* = O_P\left(\log^{-5}(n)\right).$$

Proof. Theorem 3.1 implies that $\lim_{n\to\infty}\mathbb{P}(\hat{d}_n \in \mathcal{D}_{1,n}) = 0$, where $\mathcal{D}_{1,n}$ is defined in (B.1). We only need to prove that, for any constant $A > 0$, $\lim_{n\to\infty}\mathbb{P}\left\{|\hat{d}_n - d^*|\mathbf{1}_{\{\hat{d}_n\in\mathcal{D}_{2,n}\}} \leq A\log^{-5}(n)\right\} = 1$, where $\mathcal{D}_{2,n}$ is defined in (B.2).

Applying (B.7), (B.8) and Proposition A.1, we obtain:

$$\begin{aligned}0 &\leq (\hat{d}_n - d^*)^2\mathbf{1}_{\{\hat{d}_n\in\mathcal{D}_{2,n}\}} \leq C\sup_{(d,\theta)\in\mathcal{D}_{2,n}\times\Theta_n}|\log\{1+\mathcal{Z}_m(\gamma_m(d,\theta))\}| + C\mathrm{e}^{-c\sqrt{\log(n/m)}} \\ &= O_P(m^{-(1/2\wedge\epsilon)}\log(m)) + C\mathrm{e}^{-\sqrt{\log(n/m)}}.\end{aligned}$$

If m satisfies (3.9), then we obtain that $(\hat{d}_n - d^*)^2 \mathbf{1}_{\{\hat{d}_n \in \mathcal{D}_{2,n}\}} = O_P(\mathrm{e}^{-\sqrt{\log(n/m)}}) = o_P(\log^{-s}(n))$, for any positive integer s. □

At this point, for the sake of clarity, we treat the parameterizations (P1) and (P2) separately. We will give a detailed proof in the former case and a sketchier one in the latter case.

Proof of Theorem 3.1 under P(1)

Define $L_m(d,\theta) = J_m(d,\theta) - J_m(d^*,\theta^*)$ and

$$\mathcal{D}'_n = \{d \in \mathcal{D}_n : |d - d^*| \le \log^{-5}(n)\}. \tag{C.1}$$

Lemma C.3. *Let v_m be a deterministic sequence such that $\lim_{m\to\infty} \log^5(m) v_m = 0$. If $\hat{d}_n - d^* = o_P(v_m)$, then, under (P1) with $\rho_\eta = 0$,*

$$L_m(\hat{d}_n, \hat{\theta}_n) = O_P(m^{-1/2} \log(n)\{v_m + x_m^{2d^*}\})). \tag{C.2}$$

Proof. Define $\mathcal{D}_n(v_m) = \{d \in \mathcal{D}_n : |d - d^*| \le v_m\}$. Note that the assumption on v_m implies that $\mathcal{D}_n(v_m) \subset \mathcal{D}'_n$. Applying (B.4), we have

$$0 \le L_m(\hat{d}_n, \hat{\theta}_n) \le \log\{1 + \mathcal{Z}_m(\gamma_m(d^*,\theta^*))\} - \log\{1 + \mathcal{Z}_m(\gamma_m(\hat{d}_n, \hat{\theta}_n))\}.$$

Since we already know by the proof of Theorem 3.1 that $\sup_{(d,\theta)\in\mathcal{D}'_n\times\Theta} \mathcal{Z}_m(\gamma_m(d_n,\theta))\} = o_P(1)$, it suffices to prove that

$$\sup_{(d,\theta)\in\mathcal{D}_n(v_m)\times\Theta_n} |\mathcal{Z}_m(\gamma_m(d^*,\theta^*)) - \mathcal{Z}_m(\gamma_m(d,\theta))| = O_P(m^{-1/2}\log(n)\{v_m + x_m^{2d^*}\}).$$

Since $\mathcal{Z}_m$ is linear in its argument, this is equivalent to

$$\sup_{(d,\theta)\in\mathcal{D}_n(v_m)\times\Theta_n} |\mathcal{Z}_m\{\gamma_m(d^*,\theta^*) - \gamma_m(d,\theta)\}| = O_P(m^{-1/2}\log(n)\{v_m + x_m^{2d^*}\}). \tag{C.3}$$

By Proposition A.1, (A.6), we only have to check that if $(d,\theta) \in \mathcal{D}_n(v_m) \times \Theta_n$, then the sequence $\gamma_m(d,\theta) - \gamma_m(d^*,\theta^*)$ belongs to the class $\mathcal{C}(1, v_m \log(m) + \log^2(n/m)x_m^2)$. To check this, note first that $\gamma_{m,k}(d^*,\theta^*) \equiv 1/m$. Since $\mathcal{D}(v_m) \subset \mathcal{D}'_n$, by (B.9) and (B.10), there exists a constant C such that,

$$\sup_{(d,\theta)\in\mathcal{D}(v_m)\times\Theta_n} \sup_{k=1,\dots,m-1} |\gamma_{m,k}(d,\theta) - \gamma_{m,k+1}(d,\theta)| \le C k^{-1} m^{-1} \left\{v_m + x_m^{2d^*}\log(n/m)\right\}.$$

There only remains to bound $\gamma_{m,m}(d,\theta) - 1/m$. It is easily checked that

$$\sup_{(d,\theta)\in\mathcal{D}_n(v_m)\times\Theta_n} |\gamma_{m,m}(d,\theta) - 1/m| \le C m^{-1}\log(n)\left\{v_m + x_m^{2d^*}\right\}.$$

Thus, for all $(d,\theta) \in \mathcal{D}(v_m) \times \Theta_n$, $\gamma_m(d,\theta) \in \mathcal{C}_m(1, \log(n)\{v_m + x_m^{2d^*}\})$.

□

Lemma C.4. *Under parameterization (P1), if $\rho_\eta = 0$, there exists a constant C such that for all $(d, \theta) \in \mathcal{D}'_n \times \Theta_n$,*

$$\left| J_m(d,\theta) - J_m(d^*,\theta^*) - \frac{1}{2}(2d-2d^*, \theta-\theta^*) H_m^* (2d-2d^*, \theta-\theta^*)' \right| \\ \leq C \log^3(n) \left(x_m^{6d^*} + |d-d^*|^3 \right), \quad \text{(C.4)}$$

where the positive definite matrix H_m^* is defined by

$$H_m^* = \begin{pmatrix} \rho_m(0,0) & \rho_m(0,2d^*) \\ \rho_m(0,2d^*) & \rho_m(2d^*,2d^*) \end{pmatrix}.$$

Proof. For brevity, we introduce more notation. Denote $\Delta = 2d - 2d^*$ and

$$\nu_m = m^{-1} \sum_{k=1}^m \log(k), \qquad \nu_m^{(2)} = m^{-1} \sum_{k=1}^m \log^2(k), \quad \text{(C.5)}$$
$$A_m(d,\theta) = \frac{1}{m} \sum_{k=1}^m (\alpha_k(d,\theta) - 1), \qquad B_m(d,\theta) = \frac{1}{m} \sum_{k=1}^m (k^\Delta \alpha_k(d,\theta) - 1),$$
$$C_m(d,\theta) = \frac{1}{m} \sum_{k=1}^m (\alpha_k(d,\theta) - 1)^2, \qquad D_m(d,\theta) = \frac{1}{m} \sum_{k=1}^m \log(k)(\alpha_k(d,\theta) - 1).$$

Since

$$\sup_{k \in \{1,\ldots,m\}} \sup_{(d,\theta) \in \mathcal{D}'_n \times \Theta_n} |\alpha_k(d,\theta) - 1| = O(\log(n/m) x_m^{2d^*})$$

we have $\sup_{(d,\theta)\in\mathcal{D}'_n\times\Theta_n} A_m(d,\theta) = O(\log^2(n/m)\, x_m^{2d^*})$ and

$$\sup_{(d,\theta)\in\mathcal{D}'_n\times\Theta_n} \left| \frac{1}{m} \sum_{k=1}^m \log(\alpha_k(d,\theta)) - A_m(d,\theta) + \frac{1}{2} C_m(d,\theta) \right| = O(\log^3(n/m)\, x_m^{6d^*}).$$

In addition, there exists a constant C such that, for $\Delta \in \mathcal{D}'_n$, we have

$$\max_{k \in \{1,\ldots,m\}} \left| k^\Delta - 1 - \Delta \log(k) - \frac{1}{2} \Delta^2 \log^2(k) \right| \leq C \Delta^3 \log^3(m).$$

Using the previous bounds and the inequality, for all $a, b > 0$, $a^2 b \leq (2a^3 + b^3)/3$, we also obtain:

$$\left| \frac{1}{m} \sum_{k=1}^m (k^\Delta - 1)(\alpha_k(d,\theta) - 1) - D_m(d,\theta) \right| \\ \leq C \Delta^2 \log^2(m) \log(n/m) x_m^{2d^*} \leq C \log^3(n) \left(\Delta^3 + x_m^{6d^*} \right).$$

Writing now

$$B_m(d,\theta) = \frac{1}{m}\sum_{k=1}^{m}(k^{\Delta}-1) + \frac{1}{m}\sum_{k=1}^{m}(\alpha_k(d,\theta)-1) + \frac{1}{m}\sum_{k=1}^{m}(k^{\Delta}-1)(\alpha_k(d,\theta)-1)$$

we obtain that there exists a constant C such that, for all $(d,\theta) \in \mathcal{D}'_n \times \Theta_n$,

$$\left|B_m(d,\theta) - \Delta\nu_m - \frac{1}{2}\nu_m^{(2)}\Delta^2 - A_m(d,\theta) - \Delta D_m(d,\theta)\right| \le C\log^3(n)\left(\Delta^3 + x_m^{6d^*}\right),$$
$$\left|B_m^2(d,\theta) - \Delta^2\nu_m^2 - 2\Delta\nu_m A_m(d,\theta) - A_m^2(d,\theta)\right| \le C\log^3(n)(\Delta^3 + x_m^{6d^*}).$$

Thus, there exists a constant C such that, for all $(d,\theta) \in \mathcal{D}'_n \times \Theta_n$,

$$\left|\log\{1 + B_m(d,\theta)\} - B_m(d,\theta) + \frac{1}{2}B_m^2(d,\theta)\right| \le C\log^3(n)(\Delta^3 + x_m^{6d^*}).$$

Since $L_m(d,\theta) = \log\{1 + B_m(d,\theta)\} - \nu_m\Delta - \frac{1}{m}\sum_{k=1}^{m}\log\{\alpha_k(d,\theta)\}$ and $\nu_m^{(2)} - \nu_m^2 = \rho_m(0,0)$, we obtain, for all $(d,\theta) \in \mathcal{D}'_n \times \Theta_n$:

$$\Big|L_m(d,\theta) - \frac{1}{2}\rho_m(0,0)\Delta^2 - \Delta(D_m(d,\theta) - \nu_m A_m(d,\theta))$$
$$- \frac{1}{2}\{C_m(d,\theta) - A_m^2(d,\theta)\}\Big| \le C\log^3(n)(\Delta^3 + x_m^{6d^*}).$$

The proof is concluded by applying the following bounds, which are uniform over $\mathcal{D}'_n \times \Theta_n$:

$$\left|C_m(d,\theta) - A_m^2(d,\theta) - \rho_m(2d^*,2d^*)(\theta-\theta^*)^2\right| \le C\log^3(n)\{\Delta^3 + x_m^{6d^*}\},$$
$$|D_m(d,\theta) - \nu_m A_m(d,\theta) - \rho_m(0,2d^*)(\theta^*-\theta)| \le C\log^3(n)\{\Delta^3 + x_m^{6d^*}\}.$$

□

Proof of Theorem 3.2 under (P1). For brevity, denote $\tau_m^2 = \rho_m(0,0)$ and $\delta_m = \tau_m/\rho_m(0,2d^*)$. Applying Lemmas C.2 and C.4, we obtain that

$$L_m(\hat{d}_n,\hat{\theta}_n) = \frac{1}{2}\{2\tau_m(\hat{d}_n - d^*) + \delta_m(\hat{\theta}_m - \theta^*)\}^2 + O_P(\log^4(n/m)x_m^{4d^*}). \tag{C.6}$$

By Lemmas C.2 and C.3, we know that $L_m(\hat{d}_n,\hat{\theta}_n) = O_P(m^{-1/2})$. Assumption (3.9) implies that $m^{-1/2} = o(x_m^{2d^*})$. Thus

$$\lim_{n\to\infty} x_m^{-2d^*}\{2\tau_m(\hat{d}_n - d^*) + \delta_m(\hat{\theta}_m - \theta^*)\}^2 = 0.$$

Hence $2\tau_m(\hat{d}_n - d^*) + \delta_m(\hat{\theta}_m - \theta^*) = o_P(x_m^{d^*})$. By Lemma C.1 $\lim_{m\to\infty}\tau_m^2 = 1$ and $\delta_m = O(x_m^{2d^*})$, thus we obtain that $\hat{d}_n - d^* = o_P(x_m^{d^*})$.

Applying again Lemma C.3, we now obtain that $L_m(\hat{d}_n, \hat{\theta}_n) = O_P(m^{-1/2} \log(n) x_m^{d^*}) = o_P(x_m^{3d^*})$ under (3.9).

Thus, by (C.6) and (3.9), we obtain that $\lim_{n\to\infty} x_m^{-3d^*} \{2\tau_m(\hat{d}_n - d^*) + \delta_m(\hat{\theta}_m - \theta^*)\}^2 = 0$, hence $\hat{d}_n - d^* = o_P(x_m^{3d^*/2})$. This in its turn implies, by Lemma C.3, that $L_m(\hat{d}_n, \hat{\theta}_n) = O_P(m^{-1/2} \log(n) x_m^{3d^*/2})$.

Iterating this procedure, we obtain that for all $k \geq 1$, $\hat{d}_n - d^* = o_P(x_m^{2d^*(1-2^{-k})})$ and $L_m(\hat{d}_n, \hat{\theta}_n) = O_P(m^{-1/2} \log(n) x_m^{2d^*(1-2^{-k})})$.

Under assumption (3.9), there exists an integer k^* such that $m^{-1/2} \log(n) x_m^{-2d^*(1+2^{-k^*})} = o(1)$. For this k^*, Lemma C.3 implies that $L_m(\hat{d}_n, \hat{\theta}_n) = O_P(x_m^{2d^*(1-2^{-k^*})} \log(n) m^{-1/2}) = o_P(x_m^{4d^*})$. Define $\kappa_m^2 = \tau_m^2 - \delta_m^2$. By Cauchy-Schwarz inequality, $\kappa_m > 2 > 0$ for $m \geq 2$. Thus, applying Lemma C.4 we finally obtain:

$$\{2\tau_m(\hat{d}_n - d^*) + \delta_m(\hat{\theta}_m - \theta^*)\}^2 + \kappa_m^2 (\hat{\theta}_n - \theta)^2 = o_P(x_m^{4d^*}).$$

And we can conclude that $\hat{d}_n - d^* = o_P(x_m^{2d^*})$ and $\hat{\theta}_n - \theta^* = o_P(1)$. □

Proof of Theorem 3.2 under (P2)

The scheme of the proof is the same as previously, but there is one more step because of the extra parameter involved, and because of bias terms of order x_m^* which appear now.

Lemma C.5. *Let v_m and w_m be deterministic sequences such that $\lim_{m\to\infty} \log^5(m) v_m = 0$ and $\lim_{m\to\infty} w_m = 0$. If $\hat{d}_n - d^* = o_P(v_m)$ and $\hat{\theta}_{1,n} - \theta_1^* = O_P(w_m)$, then, under (P2),*

$$L_m(\hat{d}_n, \hat{\theta}_n) = O_P(m^{-1/2} \log(n) \{v_m + w_m x_m^{d^*} + x_m^{2d^*}\})). \tag{C.7}$$

Proof. The proof is the same as the proof of Lemma C.3, with a more precise bound for $\alpha_k(d,\theta) - \alpha_{k+1}(d,\theta)$ that is a refinement of (B.10). More precisely, it holds that:

$$|\alpha_k(d,\theta) - \alpha_{k+1}(d,\theta)| \leq Ck^{-1}\{|d - d^*| + |\theta_1 - \theta_1^*| x_k^{d^*} + x_k^{2d^*}\}.$$

□

Lemma C.6. *Under (P2), there exists a constant C such that for all $(d,\theta) \in \mathcal{D}'_n \times \Theta_n$,*

$$\left| J_m(d,\theta) - J_m(d^*,\theta^*) - \frac{1}{2}(2(d-d^*), \theta_1 - \theta_1^*) \tilde{H}_m^* (2(d-d^*), \theta_1 - \theta_1^*)' \right| \\ \leq C \log^3(n) \{|d - d^*|^3 + x_m^{3d^*}\} \tag{C.8}$$

where the positive definite matrix $\tilde{H}_m^*$ is defined by

$$\tilde{H}_m^* = \begin{pmatrix} \rho_m(0,0) & \rho'_m(0,d^*) \\ \rho_m(0,d^*) & \rho''_m(d^*,d^*) \end{pmatrix}.$$

Define $\mathcal{D}''_n = \{d \in \mathcal{D}_n : |d - d^*| \leq x_m^{3d^*/2}\}$ *and* $\Theta'_n = \{\theta \in \Theta_n \mid |\theta - \theta^*| \leq x_m^{d^*/2}\}$*. Then, under (P2),*

$$\sup_{(d,\theta)\in\mathcal{D}''_n\times\Theta'_n} \Big| J_m(d,\theta) - J_m(d^*,\theta^*) \\ - \frac{1}{2}(2(d-d^*), \theta_1 - \theta_1^*, \theta_2 - \theta_2^*) K_m^* (2(d-d^*), \theta_1 - \theta_1^*, \theta_2 - \theta_2^*)' \Big| = o(x_m^{4d^*}), \quad \text{(C.9)}$$

where the positive definite matrix K_m^* *is defined by*

$$K_m^* = \begin{pmatrix} \rho_m(0,0) & \rho'_m(0,d^*) & \rho_m(0,2d^*) \\ \rho'_m(0,d^*) & \rho''_m(d^*,d^*) & \rho'_m(d^*,2d^*) \\ \rho_m(0,2d^*) & \rho'_m(d^*,2d^*) & \rho_m(2d^*,2d^*) \end{pmatrix}.$$

Proof. Define $\phi_k^* = \mathrm{Re}\left\{(1 - \mathrm{e}^{\mathrm{i}x_k})^{-d^*}\right\}$. Then, uniformly with respect to $(d,\theta) \in \mathcal{D}'_n \times \Theta_n$,

$$\alpha_k(d,\theta) - 1 = \left\{\theta_1^* - \theta_1 + O\left([|d - d^*| + x_k^{d^*}] \log(n)\right)\right\} \phi_k^* + (\theta_2^* - \theta_2) x_k^{2d^*} + O(x_k^{3d^*}). \quad \text{(C.10)}$$

To a first approximation, we obtain (C.8), which can be expressed in the following more convenient form:

$$L_m(d,\theta) = \frac{1}{2}\{\tau_m(d-d^*) + \delta'_m(\theta_1 - \theta_1^*)\}^2 + \frac{1}{2}\tilde{\kappa}_m^2(\theta_1 - \theta_1^*)^2 \\ + O\left(\{|d-d^*|^3 + x_m^{3d^*}\}\log^3(n)\right), \quad \text{(C.11)}$$

uniformly with respect to $(d,\theta) \in \mathcal{D}'_n \times \Theta_n$, and where we have defined $\delta'_m = \rho'_m(0,d^*)/\tau_m$ and ${\kappa'_m}^2 = \rho''_m(d^*,d^*) - {\delta'_m}^2$. Using (C.10) again, we can improve on the previous expansion to obtain (C.9), which can also be conveniently expressed as

$$L_m(d,\theta) = \frac{1}{2}\{\tau_m(d-d^*) + \delta'_m(\theta_1 - \theta_1^*) + \zeta_m(\theta_2 - \theta_2^*)\}^2 \\ + \frac{1}{2}\{\mu_m(\theta_1 - \theta_1^*) + \psi_m(\theta_2 - \theta_2^*)\}^2 + \frac{1}{2}\chi_m^2(\theta_2 - \theta_2^*)^2 + o\left(x_m^{4d^*}\right), \quad \text{(C.12)}$$

uniformly with respect to $(d,\theta) \in \mathcal{D}''_n \times \Theta'_n$, where μ_m is of order $x_m^{d^*}$, and ζ_m, ψ_m and χ_m are of order $x_m^{2d^*}$ (and an exact expression of these coefficients would not be helpful). □

Proof of Theorem 3.2 under (P2). As previously, the first step is to note that Lemmas C.2 and C.5 imply that $L_m(\hat{d}_n, \hat{\theta}_n) = O_P(m^{-1/2}) = o_P(x_m^{-2d^*})$ under (3.9). This, Lemma C.1 and (C.11) imply that $\hat{d}_n - d^* = o_P(x_m^{d^*})$ and $\hat{\theta}_{1,n} - \theta_1^* = o_P(1)$. This implies that the last term in (C.11) is actually $o_P(x_m^{3d^*}\log^3(n))$. This and Lemma C.5 imply that $L_m(\hat{d}_n, \hat{\theta}_n) = O_P(m^{-1/2}\log(n) x_m^{d^*}) = o_P(x_m^{3d^*})$ under (3.9). Hence, by considering again (C.11), we obtain that $\hat{d}_n - d^* = o_P(x_m^{3d^*/2})$ and $\hat{\theta}_{1,n} - \theta_1^* = o_P(x_m^{d^*/2})$. Knowing this, we can now use (C.12) and proceed iteratively as previously to conclude the proof of Theorem 3.2 under (P2). □

D Proof of Propositions 4.1 and 4.2

We outline the proof of these propositions under (P2), the proof under (P1) being exactly the same with one less parameter.

Proof of Proposition 4.1. Define

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

$$\delta_{0,k}(d,\theta) = 2\log(k) - 2m^{-1}\sum_{j=1}^{m}\log(j) - \frac{\partial_d h(d,\theta,x_k)}{1+h(d,\theta,x_k)} + m^{-1}\sum_{j=1}^{m}\frac{\partial_d h(d,\theta,x_j)}{1+h(d,\theta,x_j)},$$

$$\delta_{i,k}(d,\theta) = \frac{\partial_{\theta_i} h(d,\theta,x_k)}{1+h(d,\theta,x_k)} - m^{-1}\sum_{\ell=1}^{m}\frac{\partial_{\theta_i} h(d,\theta,x_\ell)}{1+h(d,\theta,x_\ell)}, \; i = 1,2,$$

$$N_k(d,\theta) = (\delta_{0,k}, \delta_{1,k}, \delta_{2,k}),$$

$$N_k^* = N_k(d^*,\theta^*), \; S_m^* = S_m(d^*,\theta^*), \; U_m^* = U_m(d^*,\theta^*).$$

With these notations, $m D_n^{*-1} \nabla \hat{J}_m(d^*,\theta^*) = (S_m^*)^{-1} D_n^{*-1} U_m^*$ 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

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

Hence (F.17) and (A.8) hold and we can apply Proposition A.2 to prove that T_n is asymptotically standard Gaussian.

We conclude the proof by checking that S_m^* tends to 1 in probability. In view of the proof of Proposition 4.2, we will actually prove that $S_m(d,\theta)$ converges to 1 in probability uniformly with respect to $(d,\theta) \in \mathcal{D}_n' \times \Theta_n$ where $\mathcal{D}_n'$ is defined in (C.1). Using the notations of section 3, we can write

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

By proposition A.1, $\mathcal{Z}_m(\gamma_m(d,\theta))$ converges in probability to 0 uniformly with respect to $(d,\theta) \in \mathcal{D}'_n \times \Theta_n$. 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 4.2. 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 $(d,\theta) \in \mathcal{D}'_n \times \Theta_n$. 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 $(d,\theta) \in \mathcal{D}'_n \times \Theta_n$ 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 $M_{1,n}(d,\theta) = (M_{1,n}^{(i,j)}(d,\theta))_{0\le i,j\le 2}$. 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 $i = 0,\dots,u$ and $j = 1,\dots,u$, 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:

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

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

$$- D_{i,n}^{-1} D_{j,n}^{-1} \sum_{k=1}^m \delta_{i,k}(d,\theta) \frac{\partial_{\theta_j} h(d,\theta,x_k)}{1 + h(d,\theta,x_k)} (d,\theta) k^{2d-2d^*} \alpha_k(d,\theta) (\mathcal{E}_k - 1). \tag{D.3}$$

It is easily seen that the term on the right hand side of (D.1) converges to the (i,j) entry of the asymptotic covariance matrix Γ^*. Since $d \in \mathcal{D}'_n$ and $|D_{i,n}^{-1}\delta_{i,k}| \leq C\log(n)m^{-1/2}$, we easily obtain that the term (D.2) is $O(\log^{2-s}(n))$. The term (D.3) can be expressed as $-\mathcal{Z}_m(c_m(d,\theta))$ with

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

It can be checked that for all $(d,\theta) \in \mathcal{D}'_n \times \Theta_n$, $y|c_{m,k}(d,\theta) - c_{m,k+1}(d,\theta)| \leq C\log^2(m)m^{-1}k^{-1}$ and $|c_{m,k}(d,\theta)| \leq Cm^{-1}$. Thus, for all $(d,\theta) \in \mathcal{D}'_n \times \Theta_n$, the sequence $c_m(d,\theta)$ belongs to the class $\mathcal{C}_m(1,K)$ for some constant K and we can conclude by applying Proposition A.1 that $\sup_{(d,\theta)\in\mathcal{D}'_n\times\Theta_n}\mathcal{Z}_m(c_m(d,\theta)) = o_P(1)$.

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,

$$\begin{aligned} D_{i,n}^{-1}D_{0,n}^{-1}M_{1,n}^{(i,0)}(d,\theta) &= D_{i,n}^{-1}D_{0,n}^{-1}\sum_{k=1}^{m}\delta_{i,k}(d,\theta)\delta_{0,k}(d,\theta) \\ &+ D_{i,n}^{-1}D_{0,n}^{-1}\sum_{k=1}^{m}\delta_{i,k}(d,\theta)\mu_k\left(k^{2d-2d^*}\alpha_k(d,\theta)-1\right) \\ &+ D_{i,n}^{-1}D_{0,n}^{-1}\sum_{k=1}^{m}\delta_{i,k}(d,\theta)\mu_k\alpha_k(d,\theta)k^{2d-2d^*}\alpha_k(d,\theta)(\mathcal{E}_k-1). \end{aligned} \tag{D.4}$$

As previously, the first term on the right hand side of (D.4) converges to the $(0,i)$ entry of Γ^* and the other terms tend to 0, uniformly with respect to $(d,\theta) \in \mathcal{D}'_n \times \Theta_n$. □

E Proof of Theorems 5.1 and 5.2

We already know that the standard GSE in consistent if $d^* \in [0,1)$. In order to prove the central limit theorem, we must first strengthen this result by proving a rate of convergence, as originally shown by Robinson (1995b). Hereafter, we omit the superscript in $\hat{d}_n^{(0)}$. Under the assumptions of Theorems 5.1 and 5.2, we can apply the second part of Proposition A.1, (2.iii) or (2.iv). Thus, noting that in the case under consideration here the remainder term $R_{m,n}$ is identically zero, (B.7) becomes:

$$0 \leq (\hat{d}_n - d^*)^2\mathbf{1}_{\mathcal{D}_{2,n}}(\hat{d}_n) \leq 2c^{-1}\sup_{(d,\theta)\in\mathcal{D}_{2,n}\times\Theta}|\log\{1+\mathcal{Z}_m(\gamma_m(d,\theta))\}| = O_P(m^{-1/2}).$$

Thus, $\hat{d}_n - d^* = O_P(m^{-1/4})$. We now briefly recall the way to prove the central limit theorem, since it is very standard and only uses a Taylor expansion and Propositions A.1 and A.2. Since $\hat{d}_n$ is consistent, with probability tending to one, it satisfies

$$0 = \frac{\partial\hat{J}_m(\hat{d}_n)}{\partial d} = \frac{2\sum_{k=1}^m k^{2\hat{d}_n}\log(k)I_{X,k}}{\sum_{k=1}^m k^{2\hat{d}_n}I_{X,k}} - 2\nu_m,$$

where ν_m is defined in (C.5). This implies, by a Taylor expansion

$$0 = \sum_{k=1}^{m} k^{2d^*} (\log(k) - \nu_m) I_{X,k} + 2(\hat{d}_n - d^*) \sum_{k=1}^{m} k^{2\tilde{d}_n} \log(k)(\log(k) - \nu_m) I_{X,k},$$

where $\tilde{d}_n$ lies between $\hat{d}_n$ and d^*. Define $c_m = (c_{m,k})_{1\le k\le m}$, with $c_{m,k} = \{\log(k) - \nu_m\}/\tau_m$, $\tau_m^2 = m^{-1}\sum_{k=1}^m \{\log(k) - \nu_m\}^2$ and

$$T_m = \tau_m^{-1} \sum_{k=1}^{m} k^{2\tilde{d}_n - 2d^*} \log(k)\, c_{m,k} \frac{I_{X,k}}{f_{X,k}}.$$

Then $\tau_m(\hat{d}_n - d^*) = -\frac{1}{2} T_m^{-1} \mathcal{Z}_m(c_m)$. It is easily seen that the sequence of weights c_m satisfies assumptions (A.7) and (A.8), so that $\mathcal{Z}_m(c_m)$ converges weakly to the standard Gaussian distribution. Because of the $m^{-1/4}$ consistency of $\hat{d}_n$, applying Proposition A.1, we obtain that T_m converges in probability to 1. Finally, $\lim_{m\to\infty} m^{-1}\tau_m^2 = 1$, which concludes the proof of Theorems 5.1 and 5.2.

F Proof of Propositions A.1 and A.2

We start with a simple lemma which we often use to prove that certain sums are o(1). In the sequel c, C denote numerical constants whose values may change upon each appearance.

Lemma F.1. *Let $(t_k)_{k\ge1}$ be a square summable sequence. Let $(c_{m,k})_{1\le k\le m}$ be a triangular array such that*

$$\sum_{k=1}^{m} c_{m,k}^2 = 1 \quad \textit{and} \quad \lim_{m\to\infty} \max_{k\in\{1,\dots,m\}} |c_{m,k}| = 0. \tag{F.1}$$

Then,

$$\lim_{m\to\infty} \sum_{k=1}^{m} c_{m,k} t_k = 0.$$

Proof. Split the sum at some $\ell \le m$ to be fixed later and apply the Cauchy-Schwarz inequality to the sum extending over $k \ge \ell$:

$$\sum_{k=1}^{m} |c_{m,k}||t_k| \le \ell \max_k |t_k| \max_{1\le k\le \ell} |c_{m,k}| + \left(\sum_{k\ge \ell} t_k^2 \right)^{1/2}.$$

These last two terms are simultaneously $o(1)$ as soon as the sequence $\ell = \ell(m)$ tends to infinity in such a way that $\lim\left(\ell \max_{1\le k\le m} |c_{m,k}|\right) = 0$. This is possible under (F.1). □

We now state without proof some results about the approximation of the DFT ordinates of the linear process $\{Y_t\}$, renormalised by a proxy for the spectral density, by the DFT ordinates of the white noise $\{Z_t\}$. These results are more or less straightforward adaptations of existing proofs for similar results. See, for instance Robinson (1995b), Velasco (1999b), Hurvich and Chen (2000).

Lemma F.2. *Assume* **(H1)**, **(H2)** *and* **(H3)**. *Define* $a_k = \sqrt{2\pi f_Y^*(0)}(1-\mathrm{e}^{\mathrm{i}x_k})^{-d_Y}$. *There exists a constant* C *such that for all* $k, j \le \vartheta n/\pi$,

$$\mathbb{E}\left[|d_{Y,k}/a_k - d_{Z,k}|^2\right] \le C\{\log(k)k^{-1} + (k/n)^\beta\}. \tag{F.2}$$

Assume moreover **(H4)** *and* **(H5)**. *Let* m *be a sequence of integers that satisfies* (A.4) *and let* $(c_{m,k})_{1\le k\le m}$ *is a triangular array of real numbers such that*

$$\lim_{m\to\infty}\left(\sum_{k=1}^m c_{m,k}^2\right)^{-1/2} \max_{1\le k\le m}|c_{m,k}| = 0, \tag{F.3}$$

Then

$$\mathbb{E}\left[\left|\sum_{k=1}^m c_{m,k}\left(\frac{I_{Y,k}}{f_{Y,k}} - 1\right)\right|\right] = o\left(\left\{\sum_{k=1}^m c_{m,k}^2\right\}^{1/2}\right), \tag{F.4}$$

$$\mathbb{E}\left[\left|\sum_{k=1}^m c_{m,k}\{\bar{d}_{\eta,k}d_{Y,k}/a_k - \rho_\eta\sigma_\eta/(2\pi)\}\right|\right] = o\left(\left\{\sum_{k=1}^m c_{m,k}^2\right\}^{1/2}\right). \tag{F.5}$$

We now deal with the terms involving the white noise sequence $\{\eta_t\}$. Recall that we have defined $f_{X,k} = x_k^{-2d^*}\{1 + h(d^*, \theta^*, x_k)\}$, with h as in (P0), (P1) or (P2).

Lemma F.3. *Assume* **(H1)**, **(H2)** *and* **(H3)** *and* $d^* \in (0,1)$. *Define* $f_{Y,k} = x_k^{-2d_Y} f_Y^*(0)$. *Then there exist constants* C_1 *and* C_2 *such that*

$$\mathbb{E}\left[\left|\frac{I_{X,k}}{f_{X,k}} - \frac{I_{Y,k}}{f_{Y,k}}\right|\right] \le C_1 k^{d_Y} + C_2(k/n)^{d^*}, \tag{F.6}$$

with $C_1 = 0$ *in the stationary case.*

Proof of (F.6), *stationary case.* $d^* = d_Y \in (0, 1/2)$. Write:

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

Since $\mathbb{E}[I_{Y,k}/f_{Y,k}]$ is uniformly bounded and $|f_{X,k}/f_{Y,k} - 1| \le Cx_k^{d^*}$ in the stationary case, under the three parameterizations, we obtain (F.6) in the stationary case. □

Proof of (F.6), *nonstationary case.* In the nonstationary case, extra terms appear. Recall that $U_t = \sum_{s=1}^{t} Y_s$. Then

$$d_{U,k} = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \sum_{s=1}^{t} Y_s \mathrm{e}^{\mathrm{i}tx_k} = \frac{1}{\sqrt{2\pi n}} \sum_{s=1}^{n} Y_s \sum_{t=s}^{n} \mathrm{e}^{\mathrm{i}tx_k} = \frac{d_{Y,k}}{1-\mathrm{e}^{\mathrm{i}x_k}} - \frac{\mathrm{e}^{\mathrm{i}x_k} \sum_{s=1}^{n} Y_s}{\sqrt{2\pi n}(1-\mathrm{e}^{\mathrm{i}x_k})},$$

$$I_{U,k} = \frac{I_{Y,k}}{|1-\mathrm{e}^{\mathrm{i}x_k}|^2} - \frac{2\mathrm{Re}(\mathrm{e}^{\mathrm{i}x_k} d_{Y,k}) \sum_{s=1}^{n} Y_s}{\sqrt{2\pi n}|1-\mathrm{e}^{\mathrm{i}x_k}|^2} + \frac{\left(\sum_{s=1}^{n} Y_s\right)^2}{2\pi n |1-\mathrm{e}^{\mathrm{i}x_k}|^2},$$

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

Hence,

$$\begin{aligned}
\frac{I_{X,k}}{f_{X,k}} - \frac{I_{Y,k}}{f_{Y,k}} &= \frac{I_{Y,k}}{f_{Y,k}}\left(\frac{f_{Y,k}}{|1-\mathrm{e}^{\mathrm{i}x_k}|^2 f_{X,k}} - 1\right) - \frac{2\mathrm{Re}(\mathrm{e}^{\mathrm{i}x_k} d_{Y,k}) \sum_{s=1}^{n} Y_s}{\sqrt{2\pi n}|1-\mathrm{e}^{\mathrm{i}x_k}|^2 f_{X,k}} + \frac{\left(\sum_{s=1}^{n} Y_s\right)^2}{2\pi n |1-\mathrm{e}^{\mathrm{i}x_k}|^2 f_{X,k}} \\
&\quad + \mathrm{Re}\left(\frac{2d_{Y,k}}{(1-\mathrm{e}^{\mathrm{i}x_k})f_{X,k}}\;\bar{d}_{\eta,k}\right) - \mathrm{Re}\left(\frac{2\mathrm{e}^{\mathrm{i}x_k}\sum_{s=1}^{n} Y_s}{\sqrt{2\pi n}(1-\mathrm{e}^{\mathrm{i}x_k})f_{X,k}}\;\bar{d}_{\eta,k}\right) + \frac{I_{\eta,k}}{f_{X,k}}. \qquad \text{(F.7)}
\end{aligned}$$

Straightforward variance computations yield (cf. Taqqu (2003), Proposition 4.1), for $d_Y \in (-1/2, 0)$, that

$$\mathbb{E}\left[\left(\sum_{s=1}^{n} Y_s\right)^2\right] \le C n^{2d_Y+1}. \qquad \text{(F.8)}$$

Also, in the nonstationary case, $f_{X,k} = x_k^{-2} f_{Y,k}(1 + O(x_k^{d^*}))$ under the three parameterizations, thus

$$\begin{aligned}
\mathbb{E}\left[\left|\frac{I_{X,k}}{f_{X,k}} - \frac{I_{Y,k}}{f_{Y,k}}\right|\right] &\le C\Big((k/n)^{1+d_Y} + k^{d_Y} + k^{2d_Y} + (k/n)^{1+d_Y} \\
&\qquad + n^{d_Y}(k/n)^{1+2d_Y} + (k/n)^{2+2d_Y}\Big) \le C\left(k^{d_Y} + (k/n)^{1+d_Y}\right).
\end{aligned}$$

This proves (F.6) in the nonstationary case. □

Lemma F.4. *Under the assumptions of part 2 of Proposition A.1,*

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\sum_{k=1}^{m} c_{m,k}\left\{\frac{I_{X,k}}{f_{X,k}} - \frac{I_{Y,k}}{f_{Y,k}}\right\}\right|\right] = 0. \qquad \text{(F.9)}$$

Proof of (F.9) *in the case* $d^* \in (0, 1/2)$.

$$\frac{I_{X,k}}{f_{X,k}} - \frac{I_{Y,k}}{f_{Y,k}} = \frac{f_{Y,k} - f_{X,k}}{f_{X,k}} \left(\frac{I_{Y,k}}{f_{Y,k}} - 1 \right) + \frac{2}{f_{X,k}} \mathrm{Re} \left(a_k \left\{ \frac{d_{Y,k}}{a_k} \bar{d}_{\eta,k} - \rho_\eta \sigma_\eta \right\} \right) \tag{F.10}$$

$$+ \frac{I_{\eta,k} - \sigma_\eta^2/(2\pi)}{f_{X,k}} + \frac{f_{Y,k} + 2\mathrm{Re}(a_k)\rho_\eta\sigma_\eta/(2\pi) + \sigma_\eta^2 - f_{X,k}}{f_{X,k}}. \tag{F.11}$$

Note that either $f_{X,k} - f_{Y,k} = 0$ under (P0) or $f_{X,k} - f_{Y,k} = O(x_k^{d^*})$ under (P1) or (P2). Hence the terms in (F.10) and the first term in (F.11) are bounded by applying Lemma F.2 and (F.4) with $I_{Z,k}$ instead of $I_{Y,k}/f_{Y,k}$. Consider now the last term in (F.11). Under (P1) with $\rho_\eta = 0$ or (P2), it is actually zero. Under (P0), it is of order $x_m^{d^*}$ if $\rho_\eta \neq 0$ and $x_m^{2d^*}$ if $\rho_\eta = 0$. To conclude the proof, note that, by the Cauchy Schwarz inequality, $\sum_{k=1}^m c_{m,k} x_k^\gamma = O(m^{\gamma+1/2} n^{-\gamma})$.

□

Proof of (F.9) *in the case* $d^* \in [1/2, 3/4)$. Starting from (F.7), we write:

$$\frac{I_{X,k}}{f_{X,k}} - \frac{I_{Y,k}}{f_{Y,k}}$$

$$= \left(\frac{f_{Y,k}}{|1 - \mathrm{e}^{\mathrm{i}x_k}|^2 f_{X,k}} - 1 \right) \left(\frac{I_{Y,k}}{f_{Y,k}} - 1 \right) + \mathrm{Re} \left\{ \frac{2a_k \left(\frac{d_{Y,k}}{a_k} \bar{d}_{\eta,k} - \frac{\rho_\eta \sigma_\eta}{2\pi} \right)}{(1 - \mathrm{e}^{\mathrm{i}x_k}) f_{X,k}} \right\} + \frac{I_{\eta,k} - \frac{\sigma_\eta^2}{2\pi}}{f_{X,k}} \tag{F.12}$$

$$- \frac{2\mathrm{Re}(\mathrm{e}^{\mathrm{i}x_k} d_{Y,k}) \sum_{s=1}^n Y_s}{\sqrt{2\pi n} |1 - \mathrm{e}^{\mathrm{i}x_k}|^2 f_{X,k}} + \frac{\left(\sum_{s=1}^n Y_s \right)^2}{2\pi n |1 - \mathrm{e}^{\mathrm{i}x_k}|^2 f_{X,k}} - \mathrm{Re} \left(\frac{2\mathrm{e}^{\mathrm{i}x_k} \sum_{s=1}^n Y_s}{\sqrt{2\pi n} (1 - \mathrm{e}^{\mathrm{i}x_k}) f_{X,k}} \bar{d}_{\eta,k} \right) \tag{F.13}$$

$$+ \frac{|1 - \mathrm{e}^{\mathrm{i}x_k}|^{-2} f_{Y,k} + 2\rho_\eta \sigma_\eta \mathrm{Re} \left(a_k (1 - \mathrm{e}^{\mathrm{i}x_k})^{-1} \right) / (2\pi) + \sigma_\eta^2/(2\pi) - f_{X,k}}{f_{X,k}}. \tag{F.14}$$

Under all three parameterizations, $|1 - \mathrm{e}^{\mathrm{i}x_k}|^{-2} f_{Y,k}/f_{X,k} - 1 = O(x_k^{d^*})$ at worst. Hence the terms in (F.12) are dealt with using Lemma F.2 as in the stationary case. We only consider the terms appearing in (F.13) and (F.14).

Consider the first term in (F.13), say R_n. Define $\tilde{c}_{m,k} = n^{d_Y} c_{m,k} \mathrm{e}^{\mathrm{i}x_k} a_k / (2\pi |1 - \mathrm{e}^{\mathrm{i}x_k}|^2 f_{X,k})$, $R_{n,1} = \sum_{k=1}^m \tilde{c}_{m,k} \left(\sqrt{2\pi} d_{Y,k}/a_k - \sqrt{2\pi} d_{Z,k} \right)$ and $R_{n,2} = \sum_{k=1}^m \tilde{c}_{m,k} \sqrt{2\pi} d_{Z,k}$. Then

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

Applying (F.8) 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 **(H4)** and $|\tilde{c}_{m,k}| \leq C |c_{m,k}| k^{d_Y}$, then by Lemma F.1, we obtain:

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

Applying (F.2) and the Cauchy-Schwarz inequality, we have:

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

Since $d^* \in [1/2, 3/4)$, thus $d_Y \in [-1/2, -1/4)$ and the series k^{2d_Y} is square summable. Hence, by Lemma F.1, $\sum_{k=1}^{m} |c_{m,k}| k^{2d_Y} = o(1)$ and $\sum_{k=1}^{m} |c_{m,k}| k^{-1} = o(1)$. Moreover, under either (A.4) or (A.5), $\sum_{k=1}^{m} |c_{m,k}| (k/n)^{\beta} = O(m^{\beta+1/2} n^{-\beta}) = o(1)$. Thus, $\mathbb{E}[R_{n,1}^2] = o(1)$.

The other terms in (F.13) can be dealt with straightforwardly. Applying the bound (F.8), we get

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

by Lemma F.1. Since η satisfies **(H4)**, applying (F.8) and the Hölder inequality, we bound the last term:

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

Finally, consider the term in (F.14). Since $\left| x_k^2 |1 - \mathrm{e}^{\mathrm{i}x_k}|^{-2} - 1 \right| \le C x_k^2$, it is under (P0) of order $x_k^{d^*}$ if $\rho_\eta \ne 0$ and of order $x_k^{2d^*}$ if $\rho_\eta = 0$; and under (P1) with $\rho_\eta = 0$ or (P2) of order x_k^2. Thus we obtain $\sum_{k=1}^{m} c_{m,k} r_{n,k} = o(1)$ under condition (A.5) or (A.4) respectively. □

We gather some of the previous results in the following corollary.

Corollary F.5. *Assume* **(H1)**, **(H2)** *and* **(H3)** *and* $d^* \in (0,1)$. *Then*

$$\mathbb{E} \left[\left| \sum_{j=1}^{k} \Big(\frac{I_{X,j}}{f_{X,j}} - 2\pi I_{Z,j} \Big) \right| \right] \le C \log(k) \left\{ \log(k) \mathbf{1}_{\{d^* < 1/2\}} + k^{d^*} \mathbf{1}_{\{d^* \ge 1/2\}} + (k/n)^{\beta \wedge d^*} \, k \right\}. \quad \text{(F.15)}$$

Assume moreover **(H4) (H5)** *and either (2.i), (2.ii) or (2.iii) of Proposition A.1, then, for all* $k \le m$,

$$\mathbb{E} \left[\left| \sum_{j=1}^{k} (f_{X,j}^{-1} I_{X,j} - 1) \right| \right] \le C k^{1/2} \quad \text{(F.16)}$$

We now consider the case $d^* = 0$.

Lemma F.6. *Under the assumptions of Theorem 5.2, denote*

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

with $f_X^*(0) = f_Y^*(0) + 2\sqrt{f_Y^*(0)/(2\pi)}\rho_\eta\sigma_\eta + \sigma_\eta^2/(2\pi)$. *Let* m *be a sequence of integers that satisfies* (A.4) *and let* $(c_{m,k})_{1\le k\le m}$ *be a triangular array of real numbers such that*

$$\sum_{k=1}^{m} c_{m,k}^2 = 1, \tag{F.17}$$

$$\lim_{m\to\infty} \max_{1\le k\le m} |c_{m,k}| = 0. \tag{F.18}$$

Then, for all $k \le m$,

$$\mathbb{E}\left[\left|\sum_{j=1}^{k} \{I_{X,j}/f_X^*(0) - 1\}\right|\right] \le C k^{1/2}, \tag{F.19}$$

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\sum_{k=1}^{m} c_{m,k}\{I_{X,k}/f_X^*(0) - 2\pi I_{\xi,k}\}\right|\right] = 0. \tag{F.20}$$

Proof. Note first that ξ satisfies **(H2)**, with $\mathrm{cov}(Z_k, \xi_k) = 1 + \rho_\eta\sigma_\eta/(2\pi f_Y^*(0))$ and (5.2) implies that (3.8) holds with

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

if $s = t = u = v$ and 0 otherwise. Write now:

$$\begin{aligned} I_{X,k} - 2\pi f_X^*(0) I_{\xi,k} &= |d_{Y,k} + d_{\eta,k}|^2 - 2\pi f_X^*(0) I_{\xi,k} \\ &= \left|d_{Y,k} - \sqrt{2\pi f_Y^*(0)} d_{Z,k} + \sqrt{2\pi f_X^*(0)} d_{\xi,k}\right|^2 - 2\pi f_X^*(0) I_{\xi,k} \\ &= \left|d_{Y,k} - \sqrt{2\pi f_Y^*(0)} d_{Z,k}\right|^2 + 2\sqrt{2\pi f_X^*(0)}\,\mathrm{Re}\left(\bar{d}_{\xi,k}\left\{d_{Y,k} - \sqrt{2\pi f_Y^*(0)} d_{Z,k}\right\}\right). \end{aligned} \tag{F.21}$$

The assumptions on $\{\eta_t\}$ ensure that $\{\xi_t\}$ has the same properties and the same relation to $\{Z_t\}$ as $\{\eta_t\}$, with possibly different values of the constants. Thus, the same arguments as above can be applied. □

We conclude the appendix by proving Propositions A.1 and A.2.

Proof of Proposition A.1. The idea of the proof is adapted from Robinson (1995b). It is based on summation by parts. Define $r_k = I_{X,k}/f_{X,k} - 2\pi I_{Z,k}$ and $s_k = n^{-1}\sum_{1\le s\ne t\le n} \mathrm{e}^{\mathrm{i}(t-s)x_k} Z_s Z_t$.

Then for any $c \in \mathbb{R}^m$,

$$\mathcal{Z}_m(c) = \sum_{k=1}^m c_k \frac{1}{n} \sum_{t=1}^n (Z_t^2 - 1) + \sum_{k=1}^m c_k (r_k + s_k) := \mathcal{Z}_{1,m}(c) + \mathcal{Z}_{2,m}(c).$$

By (A.3), for all $\epsilon \in (0,1)$, there exists C such that $\sup_m \sup_{c \in \mathcal{C}(\epsilon,K)} \sum_{k=1}^m |c_k| \le CK$. Thus $\sup_{c \in \mathcal{C}_m(\epsilon,K)} |\mathcal{Z}_{1,m}(c)| = o_P(1)$ under **(H1)**.

By (3.2), there exists a constant C such that for all $k = 1, \dots, m$, $\mathbb{E}\big[(\sum_{j=1}^k s_j)^2\big] \le Ck$. Thus, by summation by parts, (F.15) and the definition of the class $\mathcal{C}_m(\epsilon, K)$, we obtain

$$\begin{aligned}
\mathbb{E}\left[\sup_{c \in \mathcal{C}_m(\epsilon,K)} |\mathcal{Z}_{2,m}(c)|\right] &\le K m^{-\epsilon} \sum_{k=1}^{m-1} k^{\epsilon-2} \mathbb{E}\left[\left|\sum_{j=1}^k (r_j + s_j)\right|\right] + K m^{-1} \mathbb{E}\left[\left|\sum_{j=1}^m (r_j + s_j)\right|\right] \\
&\le CK m^{-\epsilon} \sum_{k=1}^{m-1} k^{\epsilon-2} \log(k) \left\{k^{d^* \vee \frac{1}{2}} + (k/n)^{\beta \wedge d^*} k\right\} + CK \log(m) \left\{m^{(d^* \vee \frac{1}{2}) - 1} + (m/n)^{\beta \wedge d^*}\right\} \\
&\le CK \log(m) \left\{m^{-\epsilon \wedge (1 - d^* \vee \frac{1}{2})} \log^\delta(m) + (m/n)^{\beta \wedge d^*}\right\} = o(1),
\end{aligned}$$

where $\delta = 1$ if $\epsilon + (d^* \vee \frac{1}{2}) = 1$ and zero otherwise. This proves the first part of Proposition A.1.

Under the assumptions of part 2 of Proposition A.1, we have $\mathbb{E}\left[\left|\sum_{j=1}^k (f_{X,j}^{-1} I_{X,j} - 1)\right|\right] \le C\sqrt{k}$, by Corollary F.5 in the case $d^* > 0$ or by Lemma F.6 in the case $d^* = 0$. Thus, applying summation by parts as above now yields, in both cases,

$$\begin{aligned}
\mathbb{E}\left[\sup_{c \in \mathcal{C}_m(\epsilon,K)} |\mathcal{Z}_m(c)|\right] &\le K m^{-\epsilon} \sum_{k=1}^{m-1} k^{\epsilon-2} \mathbb{E}\left[\left|\sum_{j=1}^k \Big(\frac{I_{X,j}}{f_{X,j}} - 1\Big)\right|\right] + K m^{-1} \mathbb{E}\left[\left|\sum_{j=1}^m \Big(\frac{I_{X,j}}{f_{X,j}} - 1\Big)\right|\right] \\
&\le CK m^{-\epsilon} \sum_{k=1}^{m-1} k^{\epsilon - 3/2} + CK m^{-1/2} \le CK m^{\epsilon \wedge 1/2} \log^\delta(m),
\end{aligned}$$

where $\delta = 1$ if $\epsilon = 1/2$ and zero otherwise. □

Proof of Proposition A.2. Since (A.8) implies (F.18), we can apply Lemmas F.2 and F.6, and we obtain that $\sum_{k=1}^m c_{m,k}\{f_{X,k}^{-1} I_{X,k} - 2\pi I_{\zeta,k}\} = o_P(1)$, with $\zeta = Z$ or $\zeta = \xi$. The proof of the asymptotic normality of $\sum_{k=1}^m c_{m,k} I_{\zeta,k}$ can be done along the lines of the proof of Theorem 2 in Robinson (1995b). The proof there is done for the special case $c_{m,k} = m^{-1/2}\{\log(k) - m^{-1} \sum_{j=1}^m \log(j)\}$, but only uses **(H4)** and condition (A.8). □

Table 1: Bias, standard error (SE), and root-mean-squared error (RMSE) for semi-parametric estimators of d^* in the LMSV-ARFIMA(1,0.40,0) model with $\phi = 0.8$

$n = 1000$

		$m = [n^{.6}]$			$m = [n^{.7}]$			$m = [n^{.8}]$		
		GPH	*AG*	*LW*	*GPH*	*AG*	*LW*	*GPH*	*AG*	*LW*
$nsr = 5$	Bias	-0.114	-0.071	0.023	-0.173	-0.076	0.036	-0.234	-0.115	0.079
	SE	0.097	0.210	0.177	0.066	0.135	0.166	0.051	0.091	0.148
	RMSE	0.149	0.222	0.179	0.185	0.155	0.170	0.240	0.147	0.168
$nsr = 10$	Bias	-0.195	-0.135	-0.003	-0.241	-0.158	0.009	-0.289	-0.194	0.046
	SE	0.093	0.200	0.221	0.065	0.132	0.206	0.050	0.092	0.190
	RMSE	0.216	0.241	0.221	0.249	0.206	0.206	0.293	0.214	0.195

$n = 5000$

		$m = [n^{.6}]$			$m = [n^{.7}]$			$m = [n^{.8}]$		
		GPH	*AG*	*LW*	*GPH*	*AG*	*LW*	*GPH*	*AG*	*LW*
$nsr = 5$	Bias	-0.079	-0.049	-0.007	-0.122	-0.064	-0.004	-0.197	-0.087	0.049
	SE	0.053	0.107	0.098	0.037	0.068	0.078	0.031	0.044	0.062
	RMSE	0.095	0.118	0.098	0.127	0.093	0.078	0.199	0.098	0.079
$nsr = 10$	Bias	-0.148	-0.085	-0.008	-0.200	-0.124	-0.011	-0.262	-0.168	0.031
	SE	0.054	0.108	0.128	0.036	0.067	0.103	0.032	0.043	0.079
	RMSE	0.158	0.138	0.128	0.203	0.141	0.104	0.264	0.174	0.084

$n = 10000$

		$m = [n^{.6}]$			$m = [n^{.7}]$			$m = [n^{.8}]$		
		GPH	*AG*	*LW*	*GPH*	*AG*	*LW*	*GPH*	*AG*	*LW*
$nsr = 5$	Bias	-0.071	-0.044	-0.009	-0.106	-0.061	-0.017	-0.182	-0.080	0.037
	SE	0.042	0.082	0.076	0.027	0.052	0.059	0.025	0.032	0.044
	RMSE	0.083	0.093	0.076	0.109	0.080	0.061	0.184	0.086	0.058
$nsr = 10$	Bias	-0.133	-0.073	-0.007	-0.184	-0.115	-0.016	-0.251	-0.158	0.022
	SE	0.043	0.086	0.102	0.029	0.051	0.074	0.027	0.033	0.056
	RMSE	0.140	0.113	0.103	0.186	0.126	0.076	0.252	0.161	0.060

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 Guggenberger, P. A bias-reduced log-periodogram regression estimator for the long-memory parameter. *Econometrica* **71** (2003), 675-712.

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

[5] Baillie, R.T. and A.A. Cecen and Y.-W. Han. High Frequency Deutsche Mark-US Dollar Returns: FIGARCH Representations and Non-Linearities. Multinational Finance Journal, **4**, (2000), 247-267.

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

[7] Bollerslev, T. and Mikkelsen, H.O. Long-term equity anticipation securities and stock market volatility dynamics, *Journal of Econometrics*, **93** (1999), 75-99

[8] 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.

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

[10] Dahlhaus, R. and Polonik, W. Empirical spectral processes and nonparametric maximum likelihood estimation for time series. In H. Dehling, T. Mikosch and M. Sørensen, editors, *Empirical process techniques for dependent data*, Birkhäuser, Boston, (2002).

[11] 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.

[12] 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.

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

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

[15] 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.

[16] Hurvich, C.M. and B.K. Ray. The local Whittle estimator of long-memory stochastic volatility. *Journal of Financial Econometrics*, **1,** (2003), 445–470.

[17] Hurvich, C.M. and Ph. Soulier. Testing for long memory in volatility. *Econometric Theory* **18** (2002), 1291–1308.

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

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

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

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

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

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

[24] Soulier, Ph. Empirical processes techniques for the spectral estimation of fractional processes. In H. Dehling, T. Mikosch and M. Sørensen, editors, *Empirical process techniques for dependent data*, Birkhäuser, Boston, (2002).

[25] Surgailis, D. and M.C. Viano. Long memory properties and covariance structure of the EGARCH model. *ESAIM* P&S, **6**, (2002), 311-329.

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

[27] Taqqu, M.S. Fractional Brownian motion and long-range dependence. In P. Doukhan, G. Oppenheim and M.S. Taqqu, editors, *Theory and Applications of Long-Range Dependence* Boston: Birkhäuser, (2003).

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

[29] Wright, J.H. Log periodogram estimation of long memory volatility dependencies with conditionally heavy tailed returns. *Econometric Reviews* **21** (2002), 397-417.