

DOCUMENTOS DE TRABAJO

BILTOKI

D.T. 2010.04

Semiparametric inference in correlated long memory signal plus
noise models.

Josu Arteche

eman ta zabal zazu



Universidad Euskal Herriko
del País Vasco Unibertsitatea

Facultad de Ciencias Económicas.
Avda. Lehendakari Aguirre, 83
48015 BILBAO.

Documento de Trabajo BILTOKI DT2010.04

Editado por el Departamento de Economía Aplicada III (Econometría y Estadística)
de la Universidad del País Vasco.

Depósito Legal No.: BI-1145-2010

ISSN: 1134-8984

URN: RePEc:ehu:biltok:201004

SEMIPARAMETRIC INFERENCE IN CORRELATED LONG MEMORY SIGNAL PLUS NOISE MODELS

J. Arteche *
University of the Basque Country

20th April 2010

Abstract

This paper proposes an extension of the log periodogram regression in perturbed long memory series that accounts for the added noise, also allowing for correlation between signal and noise, which represents a common situation in many economic and financial series. Consistency (for $d < 1$) and asymptotic normality (for $d < 3/4$) are shown with the same bandwidth restriction as required for the original log periodogram regression in a fully observable series, with the corresponding gain in asymptotic efficiency and faster convergence over competitors. Local Wald, Lagrange Multiplier and Hausman type tests of the hypothesis of no correlation between the latent signal and noise are also proposed.

JEL classification: C22; C13

Keywords: long memory; signal plus noise; log periodogram regression; semiparametric inference.

* *Corresponding address:* Departamento de Economía Aplicada III (Econometría y Estadística); University of the Basque Country (UPV-EHU); Bilbao 48015, Spain; Tl: (34) 94 601 3852; Fax: (34) 94 601 3754; Email: josu.arteche@ehu.es.

1 INTRODUCTION

The analysis of economic and financial time series often needs to deal with situations in which a variable is not directly observable because it suffers contamination from some noise. The variable of interest in this case is a latent signal and the contaminating noise is usually incorporated in an additive form (perhaps after a logarithmic transformation) such that the observed series is of the form

$$z_t = y_t + u_t \tag{1}$$

where y_t is the latent signal and u_t is the perturbing noise, usually considered to be weak dependent or even white noise. It is assumed without loss of generality that z_t has a zero mean since the results described hereafter would not be altered by the addition of a non zero constant.

This situation is quite common in economic time series where the latent variable often shows strong persistence (Granger, 1966). For example economic mechanisms where the short run and long run behaviour of the series are affected by different factors may give rise to a series such as (1) where y_t and u_t represent the short and long run effects respectively. A second example is the measurement error which is concomitant to many economic variables and makes the variable of interest into a latent signal. A similar situation arises also in rational expectation models where the ex ante variable y_t may exhibit long range dependence (e.g. Sun and Phillips, 2004, for the analysis of the long run Fisher equation).

In financial time series the noise may emerge in stock price series as a result of price discreteness or microstructure effects. In this case the noise represents the discrepancy between transaction prices and implicit efficient prices (Hasbrouck, 1993) and causes the weak autocorrelation empirically observed in many financial return series. This weak linear dependence of the returns is far from strong persistent. However, the noise in prices brings about a noisy series of squared returns, often used as proxies of the volatility, that masks the strong persistence in the (latent) volatility dynamics (Andersen and Bollerslev, 1998). The realized volatility (RV), built upon a summation of high frequency squared returns over a specified period, inherits the effects of this microstructure noise and can be considered to conform the specification in (1) with y_t the latent volatility and u_t emerging because of the microstructure noise. In fact a strong persistence has been often found in RV series,

so that ARFIMA models are usually employed (see Andersen et al, 2003, Deo et al, 2006, or Lieberman and Phillips, 2008). Alternatively, the strong persistence empirically found in the volatility of financial series can be modeled by means of stochastic volatility models. The Long Memory in Stochastic Volatility (LMSV) introduced independently by Harvey (1998) and Breidt et al. (1998) characterizes the returns as $r_t = \exp(y_t/2)\varepsilon_t$ with y_t the long memory component and ε_t an independent white noise. The logs of the squared returns are of the form in (1) with a noise $u_t = \log \varepsilon_t^2$. In this context long and short-lived shocks are jointly considered in y_t . However, different types of news may affect volatility in different ways such that short and long run effects could be separately incorporated in the volatility specification as $r_t = \exp((a_1 y_t + a_2 g_t)/2)\varepsilon_t$ for some weak dependent g_t and constants a_1 and a_2 (Bollerslev and Jubinski, 1999, Veiga 2006). In this case $\log r_t^2$ is the sum of a long memory process and a weak dependent noise $g_t + u_t$, rather than a white noise perturbing variable. Other log ARCH type models are also of this form with a latent signal corresponding to the persistent underlying volatility component.

Independence between signal and noise is usually assumed in the analysis of these series. However, this hypothesis is in many cases hard to sustain. The factors that affect the short run behaviour of a series might have also some effect on its long run behaviour and viceversa. Additionally, the potential correlation between a true economic variable and a measurement error has been demonstrated in a number of papers. For example, Bound et al. (1994) mention that the assumption of uncorrelation between the latent variable and the measurement error in labour market data “reflects convenience rather than conviction”. Similarly, De Jong et al. (1998) show that transaction costs and lagged adjustment to information give rise to correlation between the underlying price and noise in stock price series that may transfer to the RV such that correlation between noise and latent volatility is expected in RV series. If the volatility is considered by means of stochastic volatility models, the leverage effect typically found in financial time series may introduce correlation between the latent volatility and the added noise. Additionally, if short and long-lived shocks to volatility are modeled separately their independence is at least debatable because there may exist shocks with both short and long run effects.

Independence of signal and noise is at least a debatable issue and the specification in (1) with correlated y_t and u_t can be considered as appropriate for many economic and financial

series. Ignoring either the presence of the noise or the correlation between signal and noise in the estimation procedure may mask the strong persistence of the latent signal. Considering both the noise and the correlation, the spectral density of z_t in (1) can be expressed in terms of the spectral and cross-spectral densities of y_t and u_t such that

$$f_z(\lambda) = f_y(\lambda) + f_u(\lambda) + 2\text{Re}f_{yu}(\lambda) \quad (2)$$

where $\text{Re}(a)$ denotes the real part of a . The cross spectral density only arises in the case of non null correlation of signal and noise.

The long memory of y_t determines the behaviour of the spectral density at the origin such that $f_y(\lambda) \sim C_y \lambda^{-2d}$ as $\lambda \rightarrow 0$ for a positive constant C_y . If the added noise does not show persistence or has less memory than y_t , and under a reasonable correlation structure between y_t and u_t (see the assumptions below), the long memory property of y_t transmits to z_t and the spectral density of the observable z_t shares the divergency of $f_y(\lambda)$ at the origin with the same memory parameter $d > 0$. This spectral property entitles the estimation of the memory parameter of the latent signal using semiparametric or local techniques originally proposed for fully observable long memory series, which only consider spectral behaviour around frequency zero. However, the added noise affects the properties of these estimators, inducing a large bias which limits the efficiency by compelling the use of frequencies very close to the origin. This effect has been analyzed by Deo and Hurvich (2001) and Arteche (2004) for the log-periodogram regression and the local Whittle estimators respectively. To reduce this bias Sun and Phillips (2003), Hurvich et al. (2005) and Arteche (2006) propose modifications of both estimators that include the added noise in the estimation procedures. Sun and Phillips (2003) and Arteche (2006) consider only the case of independent signal and noise in a log periodogram regression context. Hurvich et al. (2005) extend the local Whittle estimator by including explicitly in the estimation procedure both the added white noise and the potential correlation between signal and noise by incorporating terms that account for the spectral density of u_t and the non null cross spectral density of y_t and u_t . However, their proposal does not account fully for the correlation of signal and noise, which limits the asymptotic efficiency and rate of convergence further than claimed (see the corrigendum Hurvich et al., 2008).

Here we extend the log periodogram regression in analogous directions but account for more general correlation structures. As in other similar extensions, the local Whittle type

estimators dominate those based on a log periodogram regression in an asymptotic mean squared error sense. However, the local Whittle extension of Hurvich et al. (2005) is here improved in three directions. First, we allow for a more general specification of the correlation of signal and noise, covering more realistic situations. Second, the possibility of non contemporaneous correlation is also considered. Finally, weak dependence of the noise is allowed, which is particularly relevant not only in economic series but even in some extensions of the LMSV such as those in Bollerslev and Jubinski (1999) and Veiga (2006).

We also consider the nonstationary case as in Velasco (1999) and Hurvich et al. (2005). For that purpose we define the nonstationary y_t as $y_t = y_0 + \sum_{t=1}^t v_t$ where y_0 is a random variable not depending on t and v_t is weakly stationary with memory parameter $d - 1 \in [-1/2, 0)$. The pseudo spectral density function of y_t is then

$$f_y(\lambda) = |1 - \exp(i\lambda)|^{-2} f_v(\lambda) \sim C_v \lambda^{-2d} \quad \text{as } \lambda \rightarrow 0$$

for $d \in [1/2, 1)$. The pseudo spectral density function of z_t can then be written in terms of the spectral density function of v_t and the cross spectral density of v_t and u_t as

$$f_z(\lambda) = |1 - \exp(i\lambda)|^{-2} f_v(\lambda) + f_u(\lambda) + 2\text{Re} \{ (1 - \exp(i\lambda))^{-1} f_{vu}(\lambda) \}$$

or in terms of pseudo spectral and cross spectral densities as in (2) with $f_{yu}(\lambda) = (1 - \exp(i\lambda))^{-1} f_{vu}(\lambda)$.

The rest of the paper is organized as follows. Section 2 describes the assumptions needed in our analysis and illustrates them with some primary examples. Section 3 introduces the proposed estimator and Section 4 shows its asymptotic properties, particularly consistency and asymptotic normality. Section 5 proposes Wald, Lagrange Multiplier and Hausman type tests of the hypothesis of no correlation between signal and noise. The finite sample performance of the proposed estimators and testing procedures are examined in Section 6 and in Section 7 they are applied to a series of daily RV of S&P500 futures index. The technical details are relegated to the Appendices.

2 BASIC ASSUMPTIONS AND EXAMPLES

Assumption 1: The signal y_t is a Gaussian process with a (pseudo) spectral density

satisfying

$$\begin{aligned} \text{a) } f_y(\lambda) &= C_y \lambda^{-2d_0} (1 + O(\lambda^{\beta_1})) \\ \text{b) } f_y(\lambda) &= C_y \lambda^{-2d_0} (1 + G_y \lambda^{\beta_1} + O(\lambda^{\beta_1 + \iota})) \end{aligned}$$

as $\lambda \rightarrow 0^+$, for some $\iota > 0$, finite positive C_y , finite G_y , $0 < d_0 < 1$, $\beta_1 > 1 + d_0$ and in a neighbourhood of the origin $f_y(\lambda)$ is differentiable with first derivative $O(\lambda^{-1-2d_0})$.

Assumption 2: The added noise u_t is Gaussian with a spectral density satisfying

$$f_u(\lambda) = f_u(0)(1 + O(\lambda^{\beta_2}))$$

as $\lambda \rightarrow 0^+$, $\beta_2 > \beta_1 - 2d_0$, for a positive finite $f_u(0)$ and in a neighbourhood of the origin $f_u(\lambda)$ is differentiable with first derivative $O(\lambda^{-1})$.

Assumption 3: As $\lambda \rightarrow 0^+$, the (pseudo) cross spectral density function of y_t and u_t satisfies

$$\text{Re}(f_{yu}(\lambda)) = \lambda^{-d_0} \left(C_{yu} \cos \left[d_0 \left(\frac{\lambda}{2} - \frac{\pi}{2} \right) \right] + G_{yu} \lambda \sin \left[d_0 \left(\frac{\lambda}{2} - \frac{\pi}{2} \right) \right] + O(\lambda^{\beta_3}) \right)$$

for finite constants C_{yu} and G_{yu} , $\beta_3 > \beta_1 - d_0$ and in a neighbourhood of the origin $f_{yu}(\lambda)$ is differentiable with first derivative $O(\lambda^{-1-d_0})$.

Assumption 1 imposes a particular spectral behaviour of y_t around zero, slightly relaxing Assumption 2 in Sun and Phillips (2003) and allowing for a non stationary y_t . The local specification in a) is required for consistency and b) is needed for the asymptotic normality. As in Henry and Robinson (1996) this local specification permits us to obtain the leading part of the asymptotic bias of local estimators of d_0 in terms of G_y . Only positive values of d_0 are considered. The condition $d_0 > 0$ guarantees that the long memory of y_t is transferred to z_t since by Assumption 2 u_t is weak dependent (e.g. stationary ARMA), which in view of the economic and financial examples given in the introduction is the empirically most interesting case. For $d_0 \leq 0$ the persistence of z_t would be that of u_t with a zero memory parameter. Finally, Gaussianity is required for the sake of simplicity of the proofs as in Sun and Phillips (2003). Gaussianity of the noise precludes the possibility of LMSV where the noise is not Gaussian. However, Gaussianity of RV series has been supported by Andersen et al. (2003) and Lieberman and Phillips (2008) among others. LMSV has been allowed by Deo and Hurvich (2001) and Hurvich and Soulier (2002) for the original log periodogram

regression when the u_t is a white noise (note that we allow for a weak dependent u_t) that is not accounted for in the estimation procedure. Gaussianity of u_t is more difficult to relax if the noise is included in the estimation procedure since in that case a non linear transformation of its periodogram has to be considered for the asymptotics of the estimator. Gaussianity of the signal is even more difficult to avoid and only Velasco (2000) for the original log periodogram regression in a fully observable series has replaced that assumption by restrictions in higher order moments, but even in that simplest context the asymptotic properties are only obtained if tapering is previously applied, with the consequent loss of efficiency.

Assumption 3 imposes a local behaviour of the (pseudo) cross spectral density of signal and noise. We call C_{yu} and G_{yu} the *low* and *high frequency correlation parameters* respectively since the latter is multiplied by the frequency and is thus negligible with respect to the low frequency correlation for frequencies close to zero. It is based on the extended use of the fractional difference operator $(1 - L)^d$ such that in its phasor form

$$\begin{aligned} (1 - e^{\pm i\lambda})^d &= (2 \sin \frac{\lambda}{2})^d \exp \left\{ \pm id \left(\frac{\lambda}{2} - \frac{\pi}{2} \right) \right\} \\ &= \lambda^d (1 + O(\lambda^2)) \exp \left\{ \pm id \left(\frac{\lambda}{2} - \frac{\pi}{2} \right) \right\} \end{aligned} \quad (3)$$

Hurvich et al. (2005) consider only the cosine term, ignoring the sine imaginary part so that only the low frequency correlation is accounted for. In some situations this omission generates a bias that would limit the number of frequencies or bandwidth used in their estimation further than claimed (see Remarks 3 and 4 below). We found it necessary also to include the sine imaginary part to fully account for the correlation of signal and noise.

Assumption 3 implies that the phase at zero frequency is fixed to be $d_0\pi/2$, which is related to the use of the one sided fractional filter $(1 - L)^d$, but other semiparametric structures give also rise to such a phase (see Robinson 2008 and the examples below). We found this assumption necessary for identifiability of the low and high frequency correlation parameters. Consider instead an unknown phase parameter γ_0 and modify correspondingly Assumption 3 as in Robinson (2008) such that, as $\lambda \rightarrow 0^+$

$$\text{Re}(f_{yu}(\lambda)) = \lambda^{-d_0} \left(C_{yu} \cos \gamma_0 + G_{yu} \lambda \sin \gamma_0 + O(\lambda^{\beta_3}) \right). \quad (4)$$

In this framework C_{yu} and γ_0 are not jointly identifiable (the effect is similar to an unknown memory of the added noise discussed in Remark 10 below) and knowledge of one of them is

needed to allow for the estimation of the other since the information on both parameters is concentrated in a single term $C_{yu} \cos \gamma_0$. Assumption 3 imposes a particular behaviour of the phase not only at zero frequency but also at neighbouring frequencies. Another possibility is to specify the real part of the (pseudo) cross spectral density as in (4) with $d_0\pi/2$ instead of γ_0 but in this case G_{yu} loses its correlation interpretation because it includes a term depending on d_0 and generated by the approximation of the cosine.

Different spectral smoothness parameters are permitted in the (pseudo) spectral and cross spectral densities of y_t and u_t . The restrictions $\beta_1 > 1 + d_0$, $\beta_2 > \beta_1 - 2d_0$ and $\beta_3 > \beta_1 - d_0$ guarantee that the asymptotic bias of the proposed estimator is $O(\lambda^{\beta_1})$ and also ensure identifiability of all the parameters to be estimated. For that $\beta_1 > 1 + d_0$, $\beta_2 > 1 - d_0$ and $\beta_3 > 1$ because otherwise the remainder in our regression model would be of an order of magnitude larger than $O(\lambda^{1+d_0})$ and at least G_{yu} would not be identifiable. These conditions are not very restrictive because in a wide range of situations, such as in ARMA and ARFIMA models, $\beta_1 = \beta_2 = \beta_3 = 2$.

For the purpose of illustration consider the two following cases (non stationary signals can be defined as explained in the Introduction).

Case 1: Let

$$\begin{aligned} y_t &= (1 - L)^{-d_0} \alpha(L) w_t \quad \text{for } w_t \sim NID(0, \sigma_w^2) \\ u_t &= \beta(L) \varepsilon_t \quad \text{for } \varepsilon_t \sim NID(0, \sigma_\varepsilon^2) \\ E(w_t \varepsilon_t) &= \rho \quad \text{and} \quad E(w_t \varepsilon_s) = 0 \quad t \neq s \end{aligned}$$

where L is the lag operator, $\alpha(L) = \sum_{j=0}^{\infty} \alpha_j L^j$ and $\beta(L) = \sum_{j=0}^{\infty} \beta_j L^j$ such that $\alpha(x)$ is twice differentiable with the second derivative satisfying a Lipschitz condition of order ι at one (implying $\beta_1 = 2$) and $\beta(x)$ is twice differentiable with a bounded second derivative in the neighbourhood of one (such that $\beta_2 = 2$). For example u_t can be a weak dependent ARMA and y_t a stationary fractional ARIMA, but other models such as the Bloomfield exponential and fractionally integrated Bloomfield exponential are also covered. Here $f_u(0) = \sigma_\varepsilon^2 \beta(1)^2 / 2\pi$, $C_y = \sigma_w^2 \alpha(1)^2 / 2\pi$, $G_y = d_0/12 + \{\alpha'(1)^2 - [\alpha''(1) + \alpha'(1)]\alpha(1)\} / \alpha(1)^2$ and the cross spectral density is

$$f_{yu}(\lambda) = \frac{\rho}{2\pi} (1 - e^{-i\lambda})^{-d_0} \alpha(e^{-i\lambda}) \overline{\beta(e^{-i\lambda})}$$

where the overline denotes complex conjugation. Using (3) we have that $f_{yu}(\lambda)$ satisfies Assumption 3 with $\beta_3 = 2$ and

$$C_{yu} = \frac{\rho}{2\pi}\alpha(1)\beta(1) \quad , \quad G_{yu} = \frac{\rho}{2\pi}[\alpha'(1)\beta(1) - \beta'(1)\alpha(1)]$$

A particular situation arises when y_t is a fractional noise and u_t is white noise such that $\alpha(L) = \beta(L) = 1$ and Assumptions 1 and 2 hold with $f_u(0) = \sigma_\varepsilon^2/2\pi$, $C_y = \sigma_w^2/2\pi$ and $G_y = d_0/12$. Regarding the cross spectral density, Assumption 3 holds with $C_{yu} = \rho/2\pi$ and $G_{yu} = 0$. More generally, $G_{yu} = 0$ whenever $\alpha'(1)\beta(1) - \beta'(1)\alpha(1) = 0$ and the imaginary part of $(1 - e^{i\lambda})^{-d}$ plays no role in the local specification of the cross spectral density such that the high frequency correlation in Assumption 3 can be discarded. This particular case represents the context covered by the extension of the local Whittle estimator in Hurvich et al. (2005).

Case 2: It is sometimes of interest to permit non contemporaneous correlation of signal and noise innovations such that

$$E(w_t \varepsilon_{t-s}) = \rho \quad \text{and} \quad E(w_t \varepsilon_{t-k}) = 0 \quad k \neq s$$

for some $s \neq 0$, to allow for example for lagged adjustment to information. In this case the cross spectral density function is

$$f_{yu}(\lambda) = \frac{\rho}{2\pi} e^{-i\lambda s} (1 - e^{-i\lambda})^{-d_0} \alpha(e^{-i\lambda}) \overline{\beta(e^{-i\lambda})}$$

which satisfies

$$\text{Re}(f_{yu}(\lambda)) = \lambda^{-d_0} \left\{ C_{yu}^* \cos \left[d_0 \left(\frac{\lambda}{2} - \frac{\pi}{2} \right) + \lambda s \right] + G_{yu}^* \lambda \sin \left[d_0 \left(\frac{\lambda}{2} - \frac{\pi}{2} \right) + \lambda s \right] + O(\lambda^2) \right\}$$

where

$$C_{yu}^* = \frac{\rho}{2\pi} \alpha(1)\beta(1) \quad , \quad G_{yu}^* = \frac{\rho}{2\pi} (\alpha'(1)\beta(1) - \beta'(1)\alpha(1))$$

Since

$$\begin{aligned} \cos \left[d_0 \left(\frac{\lambda}{2} - \frac{\pi}{2} \right) + \lambda s \right] &= \cos \left[d_0 \left(\frac{\lambda}{2} - \frac{\pi}{2} \right) \right] - \lambda s \sin \left[d_0 \left(\frac{\lambda}{2} - \frac{\pi}{2} \right) \right] + O(\lambda^2) \\ \sin \left[d_0 \left(\frac{\lambda}{2} - \frac{\pi}{2} \right) + \lambda s \right] &= \sin \left[d_0 \left(\frac{\lambda}{2} - \frac{\pi}{2} \right) \right] + \lambda s \cos \left[d_0 \left(\frac{\lambda}{2} - \frac{\pi}{2} \right) \right] + O(\lambda^2) \end{aligned}$$

$f_{yu}(\lambda)$ satisfies Assumption 3 with

$$C_{yu} = C_{yu}^* \quad \text{and} \quad G_{yu} = G_{yu}^* - s C_{yu}$$

The intertemporal correlation precludes the possibility of a null G_{yu} even when y_t is a fractional noise and u_t is white noise.

3 ESTIMATION UNDER CORRELATION OF SIGNAL AND NOISE

Under Assumptions 1, 2 and 3, the (pseudo) spectral density of z_t at frequency λ_j satisfies

$$f_z(\lambda_j) = C_y \lambda_j^{-2d_0} (1 + \theta'_0 X_j(d_0) + \varsigma_j) \quad (5)$$

where $\theta_0 = (\theta_{10}, \theta_{20}, \theta_{30})'$, $X_j(d) = (A_j(d)\lambda_j^d, \lambda_j^{2d}, B_j(d)\lambda_j^{1+d})'$, $A_j(d) = \cos\left[d\left(\frac{\lambda_j}{2} - \frac{\pi}{2}\right)\right]$, $B_j(d) = \sin\left[d\left(\frac{\lambda_j}{2} - \frac{\pi}{2}\right)\right]$, $\varsigma_j = O(\lambda_j^{\beta_1})$ under Assumption 1a) and $\varsigma_j = G_y \lambda_j^{\beta_1} + O(\lambda_j^{\beta_1+l})$ for $l = \min(\iota, d_0)$ under Assumption 1b) and

$$\theta_{10} = \frac{2C_{yu}}{C_y}, \quad \theta_{20} = \frac{f_u(0)}{C_y}, \quad \theta_{30} = \frac{2G_{yu}}{C_y}$$

Taking logarithms of (5) and considering only Fourier frequencies $\lambda_j = 2\pi j/n$, $j = 1, 2, \dots, [n/2]$ for n the sample size, we have

$$\log I_{zj} = a_0 + d_0(-2 \log \lambda_j) + \log(1 + \theta'_0 X_j(d_0)) + \varsigma_j + U_{zj} \quad (6)$$

where $a_0 = \log C_y - c$, $c = 0,577216$ is Euler's constant, $U_{zj} = \log(I_{zj} f_z^{-1}(\lambda_j)) + c$ and I_{zj} is the periodogram of z_t , $t = 1, 2, \dots, n$, at frequency λ_j

$$I_{zj} = I_z(\lambda_j) = |w_{zj}|^2 \quad \text{for} \quad w_{zj} = w_z(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n z_t \exp(-i\lambda_j t)$$

The ς_j in (6) is different from that in (5) but we use the same notation because they are asymptotically equivalent in the sense that they coincide up to an $o(\lambda_j^{\beta_1})$ term.

Our main interest is to estimate the memory parameter d_0 of the latent signal y_t , although the rest of parameters θ_{10} , θ_{20} and θ_{30} may also play an important role and their joint estimation not only reduces the bias of the estimates of d_0 but is also of interest in itself. The parameter θ_{20} is the long run noise to signal ratio and θ_{10} and θ_{30} correspond to the low frequency and high frequency correlations respectively between signal and noise. Any correlation between innovations of signal and noise entails $\theta_{10} \neq 0$. It also generally implies $\theta_{30} \neq 0$ but there are some particular cases where $\theta_{30} = 0$, as indicated in the previous section. Non contemporaneous correlation, however, always implies $\theta_{30} \neq 0$.

In order to avoid possible inconsistencies caused by spectral misspecifications at frequencies far from the origin, we focus only on the m Fourier frequencies closest to zero, as is usual

in other semiparametric or local estimators of d_0 . The Augmented Log-Periodogram regression Estimator (ALPE) is obtained by applying least squares to the non linear regression model

$$\log I_{zj} = a + d(-2 \log \lambda_j) + \log(1 + \theta' X_j(d)) + U_{zj} \quad j = 1, 2, \dots, m, \quad (7)$$

where the ζ_j term that is omitted will lead the bias of the estimates. This is an extension of the ALPE of Arteche (2006) to account for possible correlation between y_t and u_t covering also non stationary values of d_0 . Approximating $\log(1 + \theta' X_j(d))$ locally by $\theta' X_j(d)$ as in Sun and Phillips (2003) results in a regression model that is linear on θ but still non linear on d , giving rise to an extra bias term of order $O(\lambda_m^{2d_0})$ (Arteche, 2006) and making θ_{20} and θ_{30} unidentifiable if signal and noise are correlated. Also, ignoring the high frequency correlation as in Hurvich et al. (2005) would create an extra bias term of order $O(\lambda_m^{1+d_0})$ - except in those particular cases where $\theta_{30} = 0$ mentioned in the previous section- that would limit the size of m , requiring a more restrictive bandwidth than that allowed in Assumption 4 below. In consequence the assumption in equation (3.9) in Hurvich et al. (2005) seems insufficient to make the effect of this omitted term asymptotically negligible and a more restrictive assumption on the evolution of m seems necessary as mentioned in Remark 1 below and acknowledged by the authors in their corrigendum (Hurvich et al., 2008).

The ALPE is formally defined as

$$(\hat{d}_{ALP}, \hat{\theta}_{ALP}) = \arg \min_{\Delta \times \Theta} Q(d, \theta) \quad (8)$$

where $\Delta = [\Delta_1, \Delta_2]$, $0 < \Delta_1 < \Delta_2 < 1$, $\Theta = \Theta_1 \times \Theta_2 \times \Theta_3$ for $\Theta_1 = [\Theta_{11}, \Theta_{12}]$, $-\infty < \Theta_{11} < \Theta_{12} < \infty$, $\Theta_2 = [0, \Theta_{22}]$, $0 < \Theta_{22} < \infty$, $\Theta_3 = [\Theta_{31}, \Theta_{32}]$, $-\infty < \Theta_{31} < \Theta_{32} < \infty$, and

$$Q(d, \theta) = \frac{1}{m} \sum_{j=1}^m \left\{ \log^\dagger I_{zj} + d(2 \log \lambda_j)^\dagger - \log^\dagger(1 + \theta' X_j(d)) \right\}^2$$

where for a general ξ_j we use the notation $\xi_j^\dagger = \xi_j - \bar{\xi}$ where $\bar{\xi} = \sum \xi_j / m$.

4 ASYMPTOTIC PROPERTIES OF THE ALPE

Theorem 1 *Under Assumptions 1a)-3, $\hat{d}_{ALP} - d_0 = o_p(1)$ if $m^{-1} + mn^{-1} \rightarrow 0$ as $n \rightarrow \infty$ and $\hat{d}_{ALP} - d_0 = O_p(\lambda_m^{1+d_0})$, $\hat{\theta}_{1ALP} - \theta_{10} = O_p(\lambda_m)$, $\hat{\theta}_{2ALP} - \theta_{20} = o_p(\lambda_m^{1-d_0})$ and $\hat{\theta}_{3ALP} - \theta_{30} = o_p(1)$ if $mn^{-1} + n^{2(1+d_0)(1+\delta)} m^{-2(1+d_0)(1+\delta)-1} \rightarrow 0$ as $n \rightarrow \infty$ for some arbitrarily small $\delta > 0$ and $d_0 < 3/4$.*

Theorem 1 shows the consistency of \hat{d}_{ALP} as long as $0 < d_0 < 1$. For the consistency of $\hat{\theta}_{ALP}$ we need a more refined rate of convergence of \hat{d}_{ALP} to avoid the flatness of $Q(d, \theta)$ as a function of θ . This is only achieved for $d_0 < 3/4$, when the m^{2d_0-2} terms in the bounds in Corollary 2 are dominated by the $m^{-1/2}$ terms. With that bounds we get consistency (at different rates) of the estimators of all the parameters.

For the asymptotic normality of $(\hat{d}_{ALP}, \hat{\theta}_{ALP})$ a more restrictive assumption on the rate of increase of the bandwidth is required.

Assumption 4(K): For $\delta > 0$ arbitrarily small and $0 < d_0 < 3/4$, as $n \rightarrow \infty$,

$$\frac{n^{2(d_0+1)(1+\delta)}}{m^{1+2(d_0+1)(1+\delta)}} \rightarrow 0 \quad \text{and} \quad \frac{m^{\beta_1+1/2}}{n^{\beta_1}} \rightarrow K$$

for a finite constant K .

The first condition in Assumption 4(K) imposes a lower bound on the growth rate of m , ensuring the consistency of the ALPE. A larger bandwidth is required as the value of d_0 increases to guarantee consistency of all the estimators of the elements in θ . It ensures that all the elements in the diagonal of the normalizing matrix D_n of the gradient and Hessian defined in Theorem 2 go to infinity. The upper bound is imposed by the second condition and is the conventional $O(n^{2\beta_1/(2\beta_1+1)})$ rate in the log periodogram regression when applied to a fully observable long memory series. These two restrictions are always compatible because $\beta_1 > d_0 + 1$ and δ is arbitrarily small.

The asymptotic distribution depends on the location of θ_0 in the parameter space. We first consider the case of existence of added noise such that all the parameters to be estimated are in the interior of the parameter set.

Assumption 5: $(d_0, \theta_0) = (d_0, \theta_{10}, \theta_{20}, \theta_{30})$ is an interior point of the parameter space, $(d_0, \theta_{10}, \theta_{20}, \theta_{30}) \in (\Delta_1, \Delta_2) \times (0, \Theta_{11}) \times (\Theta_{21}, \Theta_{22}) \times (\Theta_{31}, \Theta_{32})$.

Theorem 2 *Under Assumptions 1b)-4(K) and 5*

$$D_n \begin{pmatrix} \hat{d}_{ALPE} - d_0 \\ \hat{\theta}_{ALP} - \theta_0 \end{pmatrix} \xrightarrow{d} N \left(\Omega^{-1}b, \frac{\pi^2}{6} \Omega^{-1} \right)$$

for $D_n = D_n(d_0)$ and $D_n(d) = \sqrt{m} \text{diag}(1, \lambda_m^d \cos(\pi d/2), \lambda_m^{2d}, \lambda_m^{1+d} \sin(\pi d/2))$, $b = b(d_0)$

with

$$b(d) = \begin{pmatrix} -\frac{2\beta_1}{(\beta_1+1)^2} \\ \frac{\beta_1 d}{(\beta_1+1)(d+\beta_1+1)(d+1)} \\ \frac{2d\beta_1}{(\beta_1+1)(2d+\beta_1+1)(2d+1)} \\ -\frac{\beta_1(1+d)}{(\beta_1+1)(d+\beta_1+2)(d+2)} \end{pmatrix} K(2\pi)^{\beta_1} G_y$$

and $\Omega = \Omega(d_0)$ for

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

Remark 1: Theorems 1 and 2 show that the ALPE is consistent for $0 < d_0 < 1$ and asymptotically normal as long as $d_0 < 3/4$. Lemmas 3 and 6 in Appendix B can similarly be used to show consistency (for $d_0 < 1$) and asymptotic normality (for $d_0 < 3/4$) of the estimators proposed by Sun and Phillips (2003) and Arteche (2006) when signal and noise are independent, extending the work by Velasco (1999) for nonstationary series. Note that Velasco required to trim out l frequencies close to the origin. However, bounding the contributions of the low frequencies as advocated by Hurvich et al. (1998) and used in the proof of Theorem 2 results in that this trimming is not necessary for the classical log periodogram regression nor the extensions here considered.

Remark 2: The inclusion of regressors in the estimation procedure inflates the asymptotic variance of the estimator of d_0 by a multiplicative constant that is higher the lower d_0 is. Table 1 shows the asymptotic variances of $\sqrt{m}\hat{d}_{ALP}$, $\sqrt{m}\hat{d}_1$ and $\sqrt{m}\hat{d}_2$ for $d_0 = 0.2, 0.36$ and 0.48 , where \hat{d}_1 is obtained by nonlinear log-periodogram regression with no correlation between signal and noise (that is omitting θ_1 and θ_3 in the estimation procedure, see Sun and Phillips, 2003 and Arteche, 2006) and \hat{d}_2 ignores the high frequency correlation (omits θ_3) as in Hurvich et al. (2005). The inclusion of these regressors reduces the bias and allows a broader bandwidth such that, although the variance increases significantly, the asymptotic efficiency can be improved by using a larger m .

Remark 3: If $\theta_{30} = 0$ the asymptotic covariance matrix of the correspondingly restricted ALPE is $\pi^2/6$ times the inverse of the (3×3) left upper submatrix of Ω . This is the covariance matrix of the local Whittle estimator of Hurvich et al. (2005), denoted by Γ^* in their Proposition 4.1¹, apart from the multiplicative constant, which is $\pi^2/6$ here

¹Proposition 4.1 of Hurvich et al. (2005) contains two typos. The element (2,2) of Γ^* should be divided

Table 1: Asymptotic variances of $\sqrt{m}\hat{d}_1$, $\sqrt{m}\hat{d}_2$ and $\sqrt{m}\hat{d}_{ALP}$

	\hat{d}_1	\hat{d}_2	\hat{d}_{ALP}
$d_0 = 0.2$	5.038	181.354	609.551
$d_0 = 0.36$	2.347	33.493	100.855
$d_0 = 0.48$	1.714	16.297	45.759

due to the different estimation procedure. This is the typical discrepancy between other local Whittle and log-periodogram regression based estimators. However, if $\theta_{30} \neq 0$, the omission of θ_3 in the restricted estimation generates an extra bias due to the $O(\lambda_j^{1+d_0})$ high frequency correlation component in the (pseudo) cross spectral density, which is not explicitly considered in the estimation. A more restrictive assumption on the growth rate of m should then be imposed for asymptotic normality. Precisely $m^{2(1+d_0)+1}n^{-2(d_0+1)} = O(1)$ as $n \rightarrow \infty$ should hold instead of the second part of Assumption 4. A similar condition seems also to be necessary in Hurvich et al. (2005) (see the corrigendum by Hurvich et al. 2008).

Remark 4: The analogue local Whittle type estimator of our ALPE is the Modified Gaussian Semiparametric Estimator (MGSE) defined as

$$(\tilde{d}, \tilde{\theta}) = \arg \min_{\Delta \times \Theta} \left\{ \log \left(\frac{1}{m} \sum_{j=1}^m \frac{\lambda_j^{2d} I_{zj}}{1 + \theta' X_j(d)} \right) + \frac{1}{m} \sum_{j=1}^m \log \{ \lambda_j^{-2d} (1 + \theta' X_j(d)) \} \right\} \quad (9)$$

which corresponds to the estimator of Hurvich et al. (2005) but including the high frequency correlation. This case represents a new parametrization (P3) in Hurvich et al. (2005) and under a similar set of assumptions their results cover also this possibility such that consistency and asymptotic normality are expected to hold with an asymptotic covariance matrix as that in Theorem 2 with $\pi^2/6$ replaced by one. Comparing their assumptions with those needed here for the ALPE, the main advantage of the local Whittle extension is that Gaussianity of signal and noise are not required and instead it is assumed that the stationary part of the signal admits an infinite moving order representation with respect to a martingale difference sequence with bounded fourth moment and u_t is a zero mean white noise with finite fourth moment. The assumptions required for the short memory part of the signal are implied by our Assumption 1. Note also that the ALPE covers a more general situation because it allows for weak dependence in the added noise and

by two and $\cos(d_0\pi/2)$ should be multiplying in their normalizing matrix as in our D_n .

non contemporaneous correlation. Finally the upper bound in the rate of increase of the bandwidth for the asymptotic normality of the MGSE is slightly more restrictive due to a $\log m$ term that appears in formula (3.9) of Hurvich et al (2005) and is avoided in our Assumption 4.

From an empirical perspective, the main advantage of the standard LPE over the local Whittle lies in its simple implementation as a linear regression. This property however is lost in the ALPE as well as in other log periodogram regression based estimators that account for the added noise (Sun and Phillips, 2003, Arteche, 2006), which require nonlinear optimization. Nevertheless the ALPE seems to be superior in finite samples to the MGSE, at least in the cases analyzed in the Monte Carlo in Section 7. This can be partly explained by the fact that the empirical implementation of both estimators need to restrict $1 + \theta' X_j(d)$ to be positive and this expression appears twice in the contrast function of the MGSE and only once in the ALPE.

Remark 5: If the (pseudo) cross spectral density is not considered explicitly in the estimation procedure when in fact $\theta_{10} \neq 0$, the wrongly restricted estimator of d_0 remains consistent with an appropriate bandwidth choice but θ_{20} is not identifiable because the remainder is of order $O(\lambda_m^{d_0})$, i.e. higher than the order of the regressor corresponding to θ_{20} . In consequence θ_{20} can not be estimated consistently. Based on this characteristic, a Hausman type test for correlation between signal and noise is introduced in Section 5.

Remark 6: If the high frequency correlation is not included in the estimation when actually $\theta_{30} \neq 0$, the restricted estimators of d_0 , θ_{10} and θ_{20} are consistent if $n^{4d_0(1+\delta)} m^{-1-4d_0(1+\delta)} \rightarrow 0$ for some small $\delta > 0$. The asymptotic bias of the restricted ALPE of d_0 can be approximated in this case by $\theta_{30} \lambda_m^{1+d_0} \sin(\pi d_0/2) H(d_0)$ where $H(d_0) > 0$ is a constant function of d_0 . The same expression approximates the asymptotic bias of the local Whittle estimator of Hurvich et al. (2005).

Remark 7: The asymptotic bias of \hat{d}_{ALP} can be approximated by

$$Abias(\hat{d}_{ALP}) = \left(\frac{m}{n}\right)^{\beta_1} \tilde{\Omega}_1 b_K$$

where $\tilde{\Omega}_1$ is the first row of Ω^{-1} and $b_K = b/K$. The asymptotic variance is

$$Avar(\hat{d}_{ALP}) = \frac{\pi^2}{6m} \tilde{\Omega}_{11}$$

and consequently the asymptotic mean squared error can be approximated by

$$AMSE(\hat{d}_{ALP}) = \frac{\pi^2}{6m} \tilde{\Omega}_{11} + \left(\frac{m}{n}\right)^{2\beta_1} (\tilde{\Omega}_1 b_K)^2$$

The “optimal” bandwidth that minimizes $AMSE(\hat{d}_{ALP})$ is

$$m_{ALPE}^{opt} = n^{2\beta_1/(2\beta_1+1)} \left[\frac{\pi^2 \tilde{\Omega}_{11}}{24(\tilde{\Omega}_1 b_K)^2} \right]^{1/(2\beta_1+1)}$$

such that $AMSE(\hat{d}_{ALP}) = O(n^{-2\beta_1/(2\beta_1+1)})$ if $m = m_{ALPE}^{opt}$. The ALPE then achieves the same rate of mean square error convergence as the standard LPE applied to a fully observable long memory series.

Remark 8: The bandwidths that minimize the AMSE of $\hat{\theta}_{i,ALPE}$ are $O(n^{2\beta_1/(2\beta_1+1)})$ for all $i = 1, 2, 3$, although with different multiplicative constants in each case. The optimal rates of convergence of $\hat{\theta}_{i,ALPE}$ in a mean squared error sense are then $O(n^{-(\beta_1-d_0)/(2\beta_1+1)})$, $O(n^{-(\beta_1-2d_0)/(2\beta_1+1)})$ and $O(n^{-(\beta_1-1-d_0)/(2\beta_1+1)})$ for $i = 1, 2, 3$ respectively, which can be quite slow for large d_0 .

Remark 9: We do not consider the possibility of $d_0 = 0$, so that the asymptotic distribution in Theorem 2 can not be used to test the hypothesis of short memory. However if $d_0 = 0$ there is not need to extend the original LPE and GSE since such an extension does not imply a bias reduction. The classical LPE and GSE can then be used for that purpose as suggested by Hurvich and Soulier (2002) and Hurvich et al. (2005) since both maintain the asymptotic properties as if no added noise were present. In particular both are asymptotically normal and $n^{\beta_1/(2\beta_1+1)}$ -consistent with an optimal bandwidth choice. Note also that if $d_0 = 0$ the vector of parameters θ need not be identifiable. Thus, if we suspect that the value of d_0 could be zero, we should test that possibility by means of the standard LPE or GSE. If we find evidence of a positive d_0 then some extension accounting for the possible existence of noise should be applied. The results in Theorem 2 can be used however to test the stationarity of the series by means of the null hypothesis $d_0 = 1/2$ against $d_0 < 1/2$.

Remark 10: The literature on perturbed long memory has focused on a weak dependent added noise, which is the main case of interest in economics and finance, as illustrated in the introduction. Consider however that, instead of Assumption 2, u_t is a long memory process with a spectral density satisfying

$$f_u(\lambda) = C_u \lambda^{-2d_{u0}} (1 + O(\lambda^{\beta_2}))$$

as $\lambda \rightarrow 0^+$, for a positive finite constant C_u and $-0.5 < d_{u0} < d_0$ where d_{u0} is a new parameter to be estimated. In this context the (pseudo) cross spectral density function of y_t and u_t satisfies as $\lambda \rightarrow 0^+$,

$$\operatorname{Re}(f_{yu}(\lambda)) = \lambda^{-d_0-d_{u0}} \left(C_{yu} \cos \left[(d_0 - d_{u0}) \left(\frac{\lambda}{2} - \frac{\pi}{2} \right) \right] + G_{yu} \lambda \sin \left[(d_0 - d_{u0}) \left(\frac{\lambda}{2} - \frac{\pi}{2} \right) \right] + O(\lambda^{\beta_3}) \right)$$

for finite constants C_{yu} and G_{yu} . The two examples given in Section 2 with the added noise replaced by $u_t = (1 - L)^{-d_{u0}} \beta(L) \varepsilon_t$ satisfy this specification. The (pseudo) spectral density function of z_t is then

$$f_z(\lambda_j) = C_y \lambda_j^{-2d_0} (1 + \theta'_0 X_j(d_0 - d_{u0}) + \varsigma_j)$$

and the regression model is

$$\log I_{zj} = a + d(-2 \log \lambda_j) + \log (1 + \theta' X_j(d - d_u)) + U_{zj} \quad j = 1, 2, \dots, m \quad (10)$$

There is here a problem of asymptotic identification. Denote by $\psi_0 = (d_0, \theta'_0, d_{u0})'$ the parameters to be estimated and let

$$q_j(\psi) = d(-2 \log \lambda_j)^\dagger + \log (1 + \theta' X_j(d - d_u))^\dagger$$

Denote by X_0 the $m \times 5$ matrix with elements $[X_0]_{ji} = \partial q_j(\psi_0) / \partial \psi_i$ and $D_n^\perp = D_n^\perp(d_0 - d_{u0})$ where $D_n^\perp(d) = \sqrt{m} \operatorname{diag}(1, \lambda_m^d \cos(\pi d/2), \lambda_m^{2d}, \lambda_m^{1+d} \sin(\pi d/2), \lambda_m^d \cos(\pi d/2) \log \lambda_m)$. Approximating sums by integrals it can be shown that

$$\lim_{n \rightarrow \infty} D_n^{\perp-1} X_0' X_0 D_n^{\perp-1} = \Omega^\perp$$

where $\Omega^\perp = \Omega^\perp(d_0 - d_{u0})$ such that $\Omega^\perp(d)$ is a 5×5 symmetric matrix with the first submatrix of order 4 equal to $\Omega(d)$ in Theorem 2 and the fifth column (row) equal to the vector

$$\theta_{10} \begin{pmatrix} \frac{2d}{(1+d)^2} \\ \frac{d^2}{(2d+1)(1+d)^2} \\ -\frac{2d^2}{(2d+1)(1+d)(3d+1)} \\ \frac{d}{2(d+1)(2+d)} \\ \frac{d^2 \theta_{10}}{(2d+1)(1+d)^2} \end{pmatrix}$$

The matrix Ω^\perp is also the limit of the properly normalized Hessian matrix and is singular since the last column is just $-\theta_{10}$ times the second one, so the strong asymptotic identifiability condition is not satisfied (Davidson and MacKinnon, 2004, Chapter 6). Note that the

information of the regression model on d_u is asymptotically dominated by $\theta_1 A_j (d - d_u) \lambda_j^{d-d_u}$ such that both d_u and θ_1 cannot be identified.

A long memory added noise precludes identifiability in the estimation techniques that account for it, whereas it only affects the rate of convergence of the basic local Whittle and LPE estimators that ignore the noise by constraining the rate of increase of the bandwidth (Arteche, 2004). For identifiability d_{u0} should be known but this is not realistic unless $d_{u0} = 0$ as assumed before, where such an imposition relies on the characteristics of the noise in the different situations mentioned in the introduction.

In practice we are unlikely to be able to discern a priori whether the series is perturbed by an added noise or not. It is then interesting to analyze also the case of no added noise such that $\theta_{20} = 0$ lies on the boundary of the parameter space, which affects the limiting distribution of the estimators. Note also that $\theta_{20} = 0$ precludes the possibility of correlation since it obviously implies $\theta_{10} = \theta_{30} = 0$.

Theorem 3 *Let Assumption 1b)-4(K) hold and let $d_0 \in (\Delta_1, \Delta_2)$ and $\theta_0 = 0$. Then*

$$\begin{aligned} \sqrt{m}(\hat{d}_{ALP} - d_0) &\xrightarrow{d} -\tilde{\Omega}_1 \eta \{\tilde{\Omega}_3 \eta \leq 0\} - \Omega_{11}^{-1} \eta_1 \{\tilde{\Omega}_3 \eta > 0\} \\ D_n^{**}(\hat{\theta}_{ALP} - \theta_0) &\xrightarrow{d} -\tilde{\Omega}^{**} \eta \{\tilde{\Omega}_3 \eta \leq 0\} \end{aligned}$$

where $\tilde{\Omega}_i$ is the i -th row of the matrix $\tilde{\Omega} = \Omega^{-1}$, D_n^{**} is the 3×3 low right submatrix of D_n , $\tilde{\Omega}^{**}$ is the low 3×4 submatrix of Ω^{-1} and $\eta = (\eta_1, \eta_2, \eta_3, \eta_4)' \sim N(-b, \pi^2 \Omega / 6)$.

The proof of Theorem 3 is a straightforward extension of that in Theorem 4 in Sun and Phillips (2003) and is thus omitted.

5 TESTS FOR CORRELATION BETWEEN SIGNAL AND NOISE

The asymptotic distribution of the ALPE makes easy implementation of standard asymptotic inference conceivable not only on d_0 but also on the components of the vector θ_0 . In particular, it is of special interest to test the hypothesis of no correlation between signal and noise for two reasons. First, it is of interest per se because the existence of correlation

influences subsequent analysis, for example for forecasting RVs or for estimating economic mechanisms involving series with a measurement error correlated with the latent variable. Second, it is of technical interest because it can be used as a tool for prior selection of a suitable estimation strategy. If no evidence of correlation is found the ALPE and MGSE should be adapted to this information because in that case introducing terms to account for the correlation in the estimation procedure inflates unnecessarily the variance with the consequent loss of efficiency.

However, testing such a hypothesis is not a trivial issue since it involves assessing the correlation between two unobservable series whose spectral behaviour is only locally restricted. Uncorrelation between signal and noise corresponds in our local setup to the null hypothesis

$$H_0 : \theta_{10} = \theta_{30} = 0 \quad (11)$$

The Wald test statistic for this hypothesis is

$$W = \frac{6}{\pi^2} \hat{\theta}_{ALP}' D_n^*(\hat{d}_{ALP}) \Omega^*(\hat{d}_{ALP}) D_n^*(\hat{d}_{ALP}) \hat{\theta}_{ALP}^*$$

where $\hat{\theta}_{ALP}^* = (\hat{\theta}_{1,ALP}, \hat{\theta}_{3,ALP})'$, $D_n^*(d) = \sqrt{m}(\lambda_m^d \cos(\pi d/2), \lambda_m^{1+d} \sin(\pi d/2))$ and $\Omega^*(d) = \Omega_1(d) - \Omega_3(d)\Omega_2^{-1}(d)\Omega_3(d)'$ where Ω_1 , Ω_2 and Ω_3 are 2×2 matrices with elements (by row) the (2,2), (2,4), (4,2) and (4,4) (for Ω_1), (2,1), (2,3), (4,1) and (4,3) (for Ω_3) and (1,1), (1,3), (3,1) and (3,3) (for Ω_2) of the matrix $\Omega(d)$ in Theorem 2. The following corollary establishes the typical properties of the Wald type testing procedure based on W . The proof is straightforward and is omitted.

Corollary 1 *Let Assumptions 1b)-3, 4(0) and 5 hold. Under H_0 in (11), $W \xrightarrow{d} \chi_2^2$. By contrast $W \xrightarrow{p} \infty$ if $\theta_{10} \neq$ and/or $\theta_{30} \neq 0$. Also under the local alternative $H_1 : (\theta_{10}, \theta_{30})' = D_n^{*-1}(d_0)\delta$ for a non null vector $\delta = (\delta_1, \delta_3)'$, W has a non central chi-squared asymptotic distribution, $\chi_2^2(6\delta'(\pi^2\tilde{\Omega}^*(d_0))^{-1}\delta)$.*

Note that in the examples described in Section 2 $\theta_{30} \neq 0$ implies $\theta_{10} \neq 0$ but $\theta_{30} = 0$ and $\theta_{10} \neq 0$ is also a possibility under the alternative. The slow convergence of the estimators of θ_{10} and θ_{30} affects the finite sample performance of this test. In fact, we have found through simulations (not reported but available upon request) that the finite sample performance is rather poor. In order to avoid estimation of θ_{10} and θ_{30} an asymptotically equivalent

Lagrange Multiplier (LM) type test can be used. For the null hypothesis in (11) the LM test statistic has the form

$$LM = S^*(\hat{d}_{ALP}^R, \hat{\theta}_{ALP}^R)' D_n^*(\hat{d}_{ALP}^R)^{-1} \left[\frac{\pi^2}{6} \Omega^*(\hat{d}_{ALP}^R) \right]^{-1} D_n^*(\hat{d}_{ALP}^R)^{-1} S^*(\hat{d}_{ALP}^R, \hat{\theta}_{ALP}^R) \quad (12)$$

where $S^*(d, \theta) = (S_2(d, \theta), S_4(d, \theta))$ is the vector of the second and fourth elements of the score $S(d, \theta)$ in the proof of Theorem 2 in Appendix A and $\hat{d}_{ALP}^R, \hat{\theta}_{ALP}^R$ are the ALPE under the restriction of the null. These estimators correspond to those analyzed in Arteché (2006) under independence of signal and noise. The asymptotic properties of the testing procedure based on this statistic are those of the Wald type test in Corollary 1. The matrix $\Omega^*(\hat{d}_{ALP}^R)$ can be replaced by consistent estimates. In particular we can form similar matrices with the elements of the Hessian estimates of Ω in the proof of Theorem 2, $D_n(\hat{d}_{ALP}^R)^{-1} H(\hat{d}_{ALP}^R, \hat{\theta}_{ALP}^R) D_n(\hat{d}_{ALP}^R)^{-1}$ or $D_n(\hat{d}_{ALP}^R)^{-1} J(\hat{d}_{ALP}^R, \hat{\theta}_{ALP}^R) D_n(\hat{d}_{ALP}^R)^{-1}$. Using these alternatives the LM statistic has a simpler expression since there is no need to use the normalizing matrix D_n whose elements cancel out in (12), so no information is needed on the non standard rates of convergence of the estimators of the different parameters. The form of the statistic in this case is

$$LM = S^*(\hat{d}_{ALP}^R, \hat{\theta}_{ALP}^R)' \left[\frac{\pi^2}{6} \Xi^*(\hat{d}_{ALP}^R, \hat{\theta}_{ALP}^R) \right]^{-1} S^*(\hat{d}_{ALP}^R, \hat{\theta}_{ALP}^R) \quad (13)$$

with Ξ^* defined similarly to Ω^* with respect to $\Xi = H$ or J .

Considering that under correlation between signal and noise θ_{20} cannot be identified unless the correlation is explicitly considered in the estimation procedure, a Hausman type test for correlation can be easily designed based on the difference between $\hat{\theta}_{2,ALPE}$, consistent under null and alternative hypotheses but less efficient than $\hat{\theta}_{2,ALPE}^R$ if no correlation exists, and $\hat{\theta}_{2,ALPE}^R$ which is not consistent under the alternative of correlated signal and noise.

Theorem 4 *Under Assumptions 1b)-3, 4(0), 5 and if H_0 in (11) holds, then as $n \rightarrow \infty$*

$$H = \Upsilon^{-1} m \lambda_m^{A \hat{d}_{ALP}} (\hat{\theta}_{2,ALPE} - \hat{\theta}_{2,ALPE}^R)^2 \xrightarrow{d} \chi_1^2$$

where

$$\Upsilon = \frac{\pi^2}{6} [L_1 \Omega(\hat{d}_{ALP})^{-1} L_1' - L_2 \Omega_2(\hat{d}_{ALP})^{-1} L_2']$$

for $L_1 = (0, 0, 1, 0)$ and $L_2 = (0, 1)$.

The proof of this theorem relies on the consistency of \hat{d}_{ALP} and the fact that $\sqrt{m}\lambda^{2d_0}(\hat{\theta}_{2,ALPE} - \hat{\theta}_{2,ALPE}^R) \xrightarrow{d} N(0, \Upsilon)$, which can be shown using the details in the proofs of Theorem 2 (for $\hat{\theta}_{2,ALPE}^R$) and of Theorem 3 in Arteche (2006) (for $\hat{\theta}_{2,ALPE}^R$) on the convergence of the respective Hessians and scores, extended to the nonstationary case (see the proof in Appendix A for details). This implies that the asymptotic covariance between $\hat{\theta}_{2,ALPE}^R$ and $\hat{\theta}_{2,ALPE}$ is equal to the variance of $\hat{\theta}_{2,ALPE}^R$, which is efficient in the class of augmented log periodogram regression estimators under uncorrelation of signal and noise. This structure is similar to other Hausman type tests.

The restricted \hat{d}_{ALP}^R could have been used instead of \hat{d}_{ALP} in the normalizing factor since it is also consistent under the null, but we use \hat{d}_{ALP} due to its consistency under the null and the alternative. As before, $m^{-1}\lambda^{-4\hat{d}_{ALP}}L_1\Omega(\hat{d}_{ALP})^{-1}L_1'$ and $m^{-1}\lambda^{-4\hat{d}_{ALP}}L_2\Omega_2(\hat{d}_{ALP})^{-1}L_2'$ can be replaced by finite sample Hessian based approximations $L_1\Xi(\hat{d}_{ALP}, \hat{\theta}_{ALP})^{-1}L_1'$ and $L_2\Xi_2(\hat{d}_{ALP}, \hat{\theta}_{ALP})^{-1}L_2'$. In this case there is no need for the normalizing factor $m^{-1}\lambda^{-4d_0}$ in the construction of the H statistic.

6 FINITE SAMPLE PERFORMANCE

The finite sample performance of the proposed ALPE, which is denoted here by ALPEsin, is compared with the following estimators:

- The modified Gaussian semiparametric or local Whittle estimator of Hurvich et al. (2005) in their (P1) specification, i.e. ignoring the correlation between signal and noise. This estimator is denoted by MGSE.
- The modified Gaussian semiparametric or local Whittle estimator of Hurvich et al. (2005) in their (P2) specification, i.e. accounting only for the low frequency correlation and ignoring the high frequency correlation as if $\theta_{30} = 0$. This estimator is denoted by MGSEcos.
- The modified Gaussian semiparametric estimator accounting for both the low and high frequency correlation as in formula (9). We denote this estimator by MGSEsin.
- The ALPE ignoring the correlation as suggested by Arteche (2006).
- The ALPE accounting only for the low frequency correlation as the MGSEcos (ALPEcos).

Under independence of signal and noise Arteche (2006) shows that the MGSE and the ALPE outperforms the estimators that do not account for the added noise (see also Hurvich and Ray, 2003). No results exist however on the correlated signal and noise case. Although this situation is partially covered in Hurvich et al. (2005) with the MGSEcos estimator, their Monte Carlo only considers MGSE in the independent case. The ALPEsin and MGSEsin are then new proposals and are expected to have a lower bias under general correlation structures. This assertion is confirmed by analyzing different situations. We first consider

$$z_t = \sigma_y y_t + u_t \quad (14)$$

where $(1 - L)^d y_t = w_t$ and σ_y is chosen such that the long run noise to signal ratio is $nsr = 2\pi^2$, (similar results, available upon request, are obtained for $nsr = \pi^2$), which is close to the ratios considered by Arteche (2006), Deo and Hurvich (2001), Hurvich and Ray (2003) and Sun and Phillips (2003). We only show the results for $d = 0.4$. Qualitatively similar conclusions derive for other values of the memory parameter. We have in particular analyzed also $d_0 = 0.2$ and $d_0 = 0.7$ (results available upon request) and the only difference is that, as expected, the larger d_0 the larger the bandwidth allowed in every estimator but the effects here discussed for $d = 0.4$ remain unaltered. With this definition of z_t two different scenarios are explored

Model 1: $u_t = \varepsilon_t$ and

Model 2: $u_t = \varepsilon_t + 0.8\varepsilon_{t-1}$

for

$$\begin{pmatrix} \varepsilon_t \\ w_t \end{pmatrix} \sim NID \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

and $\rho = 0, -0.8$.

Finally a non contemporaneous correlation example corresponding to an LMSV model is discussed. In this case the Gaussianity assumption of the added noise does not hold but we consider it relevant to analyze the applicability of the ALPE also in this context for its empirical interest. Its performance is compared with the local Whittle extensions whose asymptotic properties do not rely on the Gaussianity neither of the signal nor the noise.

Model 3: $z_t = \sigma_y y_{t-1} + u_t$ with y_t defined as in Model 1 and $u_t = \log \varepsilon_t^2$ for ε_t standard normal. For the sake of brevity we only show the results for the case of a correlation between

w_t and u_t equal to -0.8 . The results with null correlation (available upon request) are similar to those in Model 1. Note that correlation between w_t and ε_t is possible maintaining the null correlation between w_t and u_t .

For Models 1, 2 and 3, $\sigma_y^2 = nsr^{-1}$, $(1 + 0.8)^2 nsr^{-1}$ and $n sr^{-1} \pi^2 / 2$ respectively such that the signal to noise ratio is the same in all three models. In Model 1 $\theta_{30} = 0$ and the MGSEcos and ALPEcos are expected to perform better than the ALPEsin. However, in Models 2 and 3 with $\rho \neq 0$ there is a non null θ_{30} such that its omission increases the bias for large bandwidths and the ALPEsin and MGSEsin are expected to have a lower bias. The ALPE and MGSE are explicitly designed for $\rho = 0$ and their behaviour when $\rho = -0.8$ is expected to be worse than the other estimators, at least in terms of bias.

A negative correlation raises a practical complication in the application of the different correlation corrected estimators. Whereas $1 + \theta'_0 X_j(d_0)$ is always positive for a large enough sample size, in finite samples it can be negative, even when evaluated at the true set of parameter values, which prevents the logarithms from being taken in the objective functions. To circumvent this problem we truncate the argument in the logarithms by considering instead $\max(1 + \theta'_0 X_j(d), 10^{-200})$, which is asymptotically equivalent to $1 + \theta'_0 X_j(d)$ for a large enough n . Note also that a negative ρ implies that the real part of the cross spectral density diverges to $-\infty$ as λ approaches the origin, greatly affecting the spectral behaviour of z_t and the estimation of the parameters. We have also performed a similar analysis with a positive ρ and the results (not reported but available upon request) show that the benefits of the correlation correction are not as evident as with a negative correlation; the bias decreases significantly but the variance inflation often gives rise to a higher mean square error.

The Monte Carlo consists of 1000 replications of series composed of 4096 observations. We choose such a large sample size to minimize the effect of the truncation of $1 + \theta'_0 X_j(d)$ and because it is similar to the sample sizes of many of the financial time series which have formed the basis of several empirical applications on perturbed long memory as that analyzed in the next section. We analyze three different bandwidths, $m = n^{0.4}, n^{0.6}, n^{0.8}$. The latter increases at the same rate as the unfeasible ALPE optimal bandwidth but it can actually be far from this quantity due to the unknown multiplicative constant in m_{ALPE}^{opt} . There does not exist however a feasible version of m_{ALPE}^{opt} and we do not pursue the issue here. Plug-in versions cannot be justified because they need an appropriate estimate of G_y , which is so

far not available. Criteria based on the minimization of an objective function may be a better choice but their performance is often unsatisfactory (Arteche, 2004). Other adaptive procedures such as that in Giraitis et al. (2000), which adapt to the spectral smoothness of z_t could also be used, but any other bandwidth computed as the adaptive bandwidth times a constant would be equally (asymptotically) efficient. Recent results (Arteche and Orbe, 2009) using the bootstrap seem promising and can be easily extended to the ALPE based estimators but further research is required.

Table 2 shows the Monte Carlo biases and mean squared errors (MSE) in each of the different situations considered. For the minimization of the objective functions we have used the option `nlminb` in R, with the following restrictions: $0.01 < d < 0.9$, $-\exp(6) < \theta_1, \theta_3 < \exp(6)$ and $\exp(-20) < \theta_2 < \exp(6)$.

When $\rho = 0$ MGSE and ALPE tend to perform better and the inclusion of the terms accounting for correlation inevitably inflates the variance. The bias is however reduced if the noise is an MA(1) because the regressors for the correlation account indirectly for the weak dependence of the noise when a large bandwidth is used. When the signal and noise are correlated the ALPE and MGSE have lower MSE if a small bandwidth is used but the bias in both cases is quite large and tends to increase with m . The bias significantly decreases with the correlation correction such that for $m = n^{0.8}$ the ALPEsin tends to be the best option, not only in terms of bias but also in MSE.

The MGSE based estimators tend to perform poorly when correlation exists and a large bandwidth is used, even if the correlation is accounted for. This may be caused by the truncation of $1 + \theta' X_j(d)$ since it is at high frequencies that it is expected to have more impact and it plays its role in two terms of the contrast function in the MGSEcos and MGSEsin but only in one in the corresponding functions of the ALPEsin and ALPEcos.²

Finally, we analyze the performance of the different testing procedures for the hypothesis of no correlation between signal and noise. In the context considered in this paper the possible correlation is quite hard to detect because both series are unobservable. Moreover only the local spectral behaviour around the origin is restricted, thus permitting a great deal of flexibility. Table 3 shows the rejection frequencies of the Lagrange multiplier and

²A limited Monte Carlo, not reported but available upon request, confirm that the performance of the MGSE improves significantly in those cases where truncation is not needed and the parameter space is correspondingly adjusted, e.g. positive correlation and the elements in θ restricted to be positive.

Table 2: Bias and MSE with $nsr = 2\pi^2$

	ALPE	ALPE _{cos}	ALPE _{sin}	MGSE	MGSE _{cos}	MGSE _{sin}
u_t white noise						
$\rho = 0$						
$\mathbf{m} = \mathbf{n}^{0.4}$						
Bias	0.050	-0.127	-0.152	0.000	-0.130	-0.121
MSE	0.042	0.093	0.094	0.036	0.095	0.091
$\mathbf{m} = \mathbf{n}^{0.6}$						
Bias	0.019	-0.076	-0.128	-0.013	-0.072	-0.076
MSE	0.038	0.080	0.083	0.030	0.070	0.073
$\mathbf{m} = \mathbf{n}^{0.8}$						
Bias	0.021	-0.026	-0.062	-0.001	-0.023	-0.037
MSE	0.024	0.057	0.074	0.016	0.033	0.042
$\rho = -0.8$						
$\mathbf{m} = \mathbf{n}^{0.4}$						
Bias	0.136	-0.073	-0.100	0.098	-0.041	-0.038
MSE	0.054	0.088	0.093	0.041	0.090	0.092
$\mathbf{m} = \mathbf{n}^{0.6}$						
Bias	0.165	0.008	-0.070	0.166	0.174	0.162
MSE	0.056	0.071	0.078	0.049	0.068	0.070
$\mathbf{m} = \mathbf{n}^{0.8}$						
Bias	0.135	-0.107	0.064	0.113	0.186	0.182
MSE	0.097	0.091	0.059	0.067	0.091	0.088
$\mathbf{u}_t \sim \text{MA}(1)$						
$\rho = 0$						
$\mathbf{m} = \mathbf{n}^{0.4}$						
Bias	0.047	-0.128	-0.153	0.003	-0.149	-0.147
MSE	0.043	0.093	0.094	0.036	0.098	0.096
$\mathbf{m} = \mathbf{n}^{0.6}$						
Bias	0.017	-0.076	-0.135	-0.015	-0.092	-0.112
MSE	0.036	0.080	0.084	0.031	0.076	0.082
$\mathbf{m} = \mathbf{n}^{0.8}$						
Bias	-0.149	-0.050	-0.065	-0.166	-0.053	-0.015
MSE	0.038	0.047	0.066	0.038	0.037	0.043
$\rho = -0.8$						
$\mathbf{m} = \mathbf{n}^{0.4}$						
Bias	0.129	-0.069	-0.103	0.096	-0.051	-0.051
MSE	0.053	0.088	0.091	0.042	0.092	0.093
$\mathbf{m} = \mathbf{n}^{0.6}$						
Bias	0.160	0.020	-0.060	0.162	0.157	0.130
MSE	0.056	0.070	0.074	0.049	0.070	0.073
$\mathbf{m} = \mathbf{n}^{0.8}$						
Bias	0.125	0.152	-0.016	0.130	0.148	0.158
MSE	0.058	0.064	0.055	0.051	0.054	0.063
$\mathbf{z}_t = \sigma_y \mathbf{y}_{t-1} + \log \chi_1^2$						
$\rho = -0.8$						
$\mathbf{m} = \mathbf{n}^{0.4}$						
Bias	0.123	-0.068	-0.100	0.093	-0.080	-0.081
MSE	0.051	0.089	0.093	0.043	0.092	0.094
$\mathbf{m} = \mathbf{n}^{0.6}$						
Bias	0.163	0.005	-0.068	0.167	0.160	0.120
MSE	0.054	0.071	0.078	0.047	0.071	0.074
$\mathbf{m} = \mathbf{n}^{0.8}$						
Bias	-0.358	-0.369	0.116	-0.164	-0.124	-0.098
MSE	0.148	0.146	0.046	0.112	0.131	0.114

Hausman test statistics for the null hypothesis of no correlation in (11). The nominal significance level is 5% (compared with the critical values of χ_2^2 and χ_1^2 distributions) and a bandwidth $m = n^{0.8}$ is used (we have found that this large bandwidth gives better results). Wald type tests are not considered because their performance is rather poor as mentioned in the previous section. The Lagrange multiplier has two important advantages over the Wald and Hausman type tests. First, no estimation of θ_{10} and θ_{30} is required. Second, since everything is calculated under the null of no correlation no truncation of the arguments of the logarithms is needed in the contrast functions.

To construct the test statistics a feasible approximation of $\Omega(d_0)$ and $\Omega_2(d_0)$ is needed. Two options can be considered: First a plug-in version replacing d_0 by the corresponding consistent estimate has the advantage of requiring only estimation of d_0 , which gives rise to greater stability. Second we can use finite sample Hessian based approximations, as suggested by Sun and Phillips (2003) and Hurvich and Ray (2003), in two different forms: $J(d, \theta) + (H(d, \theta) - J(d, \theta))I(H(d, \theta) > 0)$ or just $J(d, \theta)$ as defined in the proof of Theorem 2 in Appendix A, with $d, \theta = \hat{d}_{ALP}^R, \hat{\theta}_{ALP}^R$ for the LM test and $d, \theta = \hat{d}_{ALP}, \hat{\theta}_{ALP}$ for the Hausman test statistic. In this case we need to estimate not only d_0 but also θ_0 (only θ_{20} in the LM test), which makes the approximation less stable. However, when using either of them in the construction of the test statistics, the normalizing matrix D_n^* in the LM and $m\lambda_m^{Ad_0}$ in the Hausman statistic are not needed due to the normalized convergence of both approximations to $\Omega(d_0)$ and $\Omega_2(d_0)$. Moreover we have found a better finite sample performance if either of these Hessian based approximation is used instead of the plug-in version and the rejection frequencies in Table 3 are obtained with both of them. The LM performs quite well if the added noise is white noise, particularly in the non-contemporaneous case. However, if the added noise shows some weak dependence, the LM testing procedure is not able to discriminate between the weak dependence of the noise and the correlation of signal and noise, since both arise as extra terms in the spectral density function. This situation could be corrected by approximating the spectral densities of the weak dependent innovations of signal and noise by local polynomials of finite orders, instead of constants. This would imply extending the nonlinear log periodogram regression with further elements to account for the weak dependence. By contrast, the Hausman test seems to be quite robust to weak dependence of the noise but is more conservative than the LM with low power in every situation.

Table 3: Rejection frequencies of $H_0 : \theta_{10} = \theta_{30} = 0$ (5 % sig. level) with $nsr = 2\pi^2$

ρ	u_t white noise	$u_t \sim MA(1)$	$z_t = \sigma_y y_{t-1} + \log \chi_1^2$
LM			
0	0.095 (0.096)	0.340 (0.341)	0.076 (0.077)
-0.8	0.889 (0.889)	0.554 (0.554)	0.999 (0.999)
Hausman			
0	0.090 (0.069)	0.017 (0.020)	0.109 (0.078)
-0.8	0.339 (0.266)	0.330 (0.261)	0.387 (0.253)

Rejection frequencies with $m = n^{0.8}$ using Hessian approximations (using J between brackets).

7 EMPIRICAL EXAMPLE: S&P500 REALIZED VOLATILITY

We analyze the persistence of the daily realized volatility (RV) for the S&P500 future index. We construct the RV series using intraday transaction prices of futures contracts as traded on the Chicago Mercantile Exchange (CME) from 8:30AM to 3:15PM. As in Martens et al. (2009) the series is computed as the sum of squared intraday returns plus the squared overnight return between closing price and opening price next day, with a five minute sampling frequency. The sample period is from January 3, 1994 until May 29, 2009, extending the series analyzed by Martens et al. (2009), which runs until December 29, 2006. We also omit incomplete days and the large negative return for the period September 11-17, 2001. The series comprises a total of 3837 days. As shown by De Jong et al. (1998) transaction costs and lagged adjustment to information give rise to correlation between the underlying price and noise in stock price series that can transfer to the RV. We analyze here if such a correlation between the latent volatility and the noise persists in the S&P500 RV and its effects on the estimation of the memory of the series.

Martens et al. (2009) use parametric techniques to get an estimate of the memory parameter around 0.5. Table 4 shows the estimates for bandwidths $m = 100, 200$ and 300 using the six semiparametric estimation techniques considered in the Monte Carlo, together with the standard errors calculated by means of the matrix $J(d, \theta)$ as explained in the previous section.³ There is a large discrepancy between the estimates accounting for the correlation and those ignoring it. Whereas the ALPE and the MGSE lies on the nonstationary region, the estimates accounting for the correlation shed some doubt even on

³Note that there is no theoretical justification for the standard error of MGSEsin since Hurvich et al. (2005) only covers MGSE and MGSEcos. The approximation for the standard error of MGSEsin is used appealing to the comments in Remark 4.

the strong persistence of the series. As shown in the Monte Carlo the correlation between signal and noise may induce a positive bias in the ALPE and MGSE and also in the ALPEcos and MGSEcos if a large bandwidth is used. We observe such a behaviour in Table 4 where estimates that account for correlation are similarly low with $m = 100$ but the ALPEcos and MGSEcos increase significantly with the bandwidth whereas the ALPEsin and MGSEsin remain close to zero. The possible existence of correlation between signal and noise is corroborated by the LM and Hausman type tests statistics, which often give values larger than 10^5 , clearly supporting the existence of such correlation.

Table 4: Memory parameter estimates of S&P500 RV and tests of correlation

m	ALPE	ALPEcos	ALPEsin	MGSE	MGSEcos	MGSEsin	LM	H
100	0.703	0.027	0.027	0.695	0.023	0.023	$> 10^5$ ($> 10^5$)	$> 10^5$ ($> 10^5$)
s.e.	0.129	0.295	1.237	0.101	0.272	1.116		
200	0.626	0.289	0.107	0.609	0.281	0.082	3933.9 (3933.9)	$> 10^5$ ($> 10^5$)
s.e.	0.091	0.184	0.162	0.072	0.153	0.179		
300	0.700	0.299	0.082	0.705	0.308	0.092	21.087 (21.087)	$> 10^5$ ($> 10^5$)
s.e.	0.085	0.136	0.218	0.064	0.116	0.211		

LM and H are Lagrange Multiplier and Hausman test statistics for the null hypothesis of no correlation using Hessian approximations (using J between brackets).

APPENDIX A: PROOFS

Proof of Theorem 1: The method of proof in Sun and Phillips (2003) is used to avoid the flatness of $Q(d, \theta)$ as a function of θ . Here we need the lemmas in Appendix B to correctly account for the correlation between signal and noise and also to avoid the linearization of the logarithm term in the contrast function, which, as explained in the text, would introduce a higher order bias. Write

$$Q(d, \theta) - Q(d_0, \theta_0) = \frac{1}{m} \sum_{j=1}^m (V_j^\dagger)^2 + \frac{2}{m} \sum_{j=1}^m V_j^\dagger (U_{zj} + \varsigma_j) \quad (15)$$

where

$$V_j = V_j(d, \theta) = 2(d - d_0) \log \lambda_j + \log(1 + C_j^0) - \log(1 + C_j)$$

$$C_j = C_j(d, \theta) = \theta' X_j(d)$$

$$C_j^0 = C_j(d_0, \theta_0)$$

The consistency of \hat{d}_{ALP} is established first. Since $\frac{1}{m} \sum_{j=1}^m (V_j^\dagger)^2 = 4(d - d_0)^2(1 + o(1))$, where $o()$ holds uniformly in $\Delta \times \Theta$, we have to show that $\frac{2}{m} \sum_{j=1}^m V_j^\dagger (U_{zj} + \varsigma_j) = o_p(1)$

uniformly. First, by Lemmas 3 and 4 below

$$\sup_{\Delta \times \Theta} \left| \sum_{j=1}^m V_j^\dagger U_{zj} \right| = O_p \left[\left(\frac{1}{\sqrt{m}} + \frac{\log^2 m}{m^{2(1-d_0)}} \right) [\sup_{\Delta} |d - d_0| + \sup_{\Delta} \lambda_m^d] \right] = o_p(1)$$

Also by the definition of ς_j and by summation by parts

$$\sup_{\Delta \times \Theta} \frac{1}{m} \left| \sum_{j=1}^m V_j^\dagger \varsigma_j \right| = O(\lambda_m^{\beta_1}) = o(1)$$

which, together with the previous results, implies $\hat{d}_{ALP} - d_0 = o_p(1)$.

Next, consider $d_0 < 3/4$. Since $Q(\hat{d}_{ALP}, \hat{\theta}_{ALP}) - Q(d_0, \theta_0) \leq 0$ then

$$\frac{1}{m} \sum \widehat{V}_j^{\dagger 2} \leq -\frac{2}{m} \sum \widehat{V}_j^\dagger (U_{zj} + \varsigma_j) \quad (16)$$

where $\widehat{V}_j^\dagger = V_j^\dagger(\hat{d}_{ALP}, \hat{\theta}_{ALP})$. By Lemmas 3 and 4 and summation by parts the right hand side is bounded in probability by

$$O_p \left(\frac{1}{\sqrt{m}} + \lambda_m^{\beta_1} \right) = o_p(\lambda_m^{2\Delta_1})$$

where the second bound comes from the assumption $n^{2(d_0+1)(1+\delta)} m^{-2(d_0+1)(1+\delta)-1} = o(1)$. Since $m^{-1} \sum \widehat{V}_j^{\dagger 2} = 4(\hat{d}_{ALP} - d_0)^2(1+o(1)) + O(\lambda_m^{2\Delta_1})$ we have that $\hat{d}_{ALP} - d_0 = O_p(\lambda_m^{\Delta_1}) = o_p(\lambda_m^{\Delta_1/2})$. Consider now $d \in \Delta_n^1 = \{d : |d - d_0| < \kappa \lambda_m^{\Delta_1/2}\}$ for some generic constant $\kappa > 0$ and $(d, \theta) \in \Delta_n^1 \times \Theta$. Since by summation by parts $\frac{1}{m} \sum_{j=1}^m V_j^\dagger \varsigma_j$ equals

$$\begin{aligned} & \frac{2(d-d_0)}{m} \sum \left(\log j - \frac{1}{m} \sum \log k \right) \varsigma_j + \frac{1}{m} \sum \log^\dagger \left(\frac{1+C_j^0}{1+C_j} \right) \varsigma_j \\ &= O \left(|d-d_0| \lambda_m^2 + |\theta_{10} - \theta_1| \lambda_m^{2+d_0} + |\theta_{20} - \theta_2| \lambda_m^{2+2d_0} + |\theta_{30} - \theta_3| \lambda_m^{3+d_0} \right) \end{aligned} \quad (17)$$

uniformly over $(d, \theta) \in \Delta_n^1 \times \Theta$, we have that by (16), (17) and Corollary 2

$$\frac{1}{m} \sum_{j=1}^m (\widehat{V}_j^\dagger)^2 \leq O_p \left(\frac{1}{\sqrt{m}} \lambda_m^{\Delta_1/2} \right) + O \left(\lambda_m^{\Delta_1/2+\beta_1} \right) = o_p(\lambda_m^{a(1+\delta)})$$

for $a = 1 + d_0$, because $n^{2a(1+\delta)} m^{-2a(1+\delta)-1} = o(1)$. Then, using Lemma 5, $\hat{d}_{ALP} - d_0 = o_p(\lambda_m^{a(1+\delta)/2})$, $\hat{\theta}_{1,ALP} - \theta_{10} = o_p(\lambda_m^{a(1+\delta)/2-d_0})$ and $\hat{\theta}_{2,ALP} - \theta_{20} = O_p(\lambda_m^b)$ for $b = \max(0, a(1+\delta)/2 - 2d_0)$. The rest of the proof is made sequentially as in the proof of Theorem 2 in Sun and Phillips (2003) noting Corollary 2 and Lemma 5 below. \square

Proof of Theorem 2: The first order conditions are

$$S(\hat{d}_{ALP}, \hat{\theta}_{ALP}) = 0$$

where, omitting the dependence on (d, θ) and n for easy of notation, and denoting the vector $x_j = (x_{0j}, x_{1j}, x_{2j}, x_{3j})'$

$$S(d, \theta) = \sum_{j=1}^m x_j^\dagger W_j$$

with

$$\begin{aligned} x_{0j} = x_{0j}(d, \theta) &= \left(2 \log \lambda_j - \frac{C_j^d}{1 + C_j} \right), \\ x_{ij} = x_{ij}(d, \theta) &= -\frac{C_j^{\theta_i}}{1 + C_j}, \quad i = 1, 2, 3 \\ W_j = W_j(d, \theta) &= \log I_{zj} + d(2 \log \lambda_j) - \log(1 + C_j) \end{aligned}$$

for

$$\begin{aligned} C_j^d &= \frac{\partial C_j(d, \theta)}{\partial d} = \theta_1 \lambda_j^d [A_j(d) \log \lambda_j + A_j^d(d)] + 2\theta_2 \lambda_j^{2d} \log \lambda_j + \theta_3 \lambda_j^{1+d} [B_j(d) \log \lambda_j + B_j^d(d)] \\ C_j^{\theta_1} &= \frac{\partial C_j(d, \theta)}{\partial \theta_1} = A_j(d) \lambda_j^d \\ C_j^{\theta_2} &= \frac{\partial C_j(d, \theta)}{\partial \theta_2} = \lambda_j^{2d} \\ C_j^{\theta_3} &= \frac{\partial C_j(d, \theta)}{\partial \theta_3} = B_j(d) \lambda_j^{1+d} \end{aligned}$$

with $A_j^d(d) = -(\lambda_j - \pi)B_j(d)/2$ and $B_j^d(d) = (\lambda_j - \pi)A_j(d)/2$.

The elements of the Hessian matrix, $H = H(d, \theta)$ are:

$$\begin{aligned} H_{1,1} &= \sum_{j=1}^m (x_{0j}^\dagger)^2 - \sum_{j=1}^m \left(\frac{C_j^{dd}}{1 + C_j} - \frac{(C_j^d)^2}{(1 + C_j)^2} \right)^\dagger W_j \\ H_{1,(i+1)} &= H_{(i+1),1} = \sum_{j=1}^m x_{0j}^\dagger x_{ij}^\dagger - \sum_{j=1}^m \left(\frac{C_j^{d\theta_i}}{1 + C_j} - \frac{C_j^d C_j^{\theta_i}}{(1 + C_j)^2} \right)^\dagger W_j \quad \text{for } i = 1, 2, 3 \\ H_{(i+1),(k+1)} &= \sum_{j=1}^m x_{ij}^\dagger x_{kj}^\dagger + \sum_{j=1}^m \left(\frac{C_j^{\theta_i} C_j^{\theta_k}}{(1 + C_j)^2} \right)^\dagger W_j \quad \text{for } i, k = 1, 2, 3 \end{aligned}$$

where

$$\begin{aligned} C_j^{dd} &= \theta_1 A_j(d) \lambda_j^d \log^2 \lambda_j + 2\theta_1 A_j^d(d) \lambda_j^d \log \lambda_j + \theta_1 A_j^{dd}(d) \lambda_j^d + 4\theta_2 \lambda_j^{2d} \log^2 \lambda_j \\ &\quad + \theta_3 B_j(d) \lambda_j^{1+d} \log^2 \lambda_j + 2\theta_3 B_j^d(d) \lambda_j^{1+d} \log \lambda_j + \theta_3 B_j^{dd}(d) \lambda_j^{1+d} \\ C_j^{d\theta_1} &= A_j(d) \lambda_j^d \log \lambda_j + \lambda_j^d A_j^d(d) \\ C_j^{d\theta_2} &= 2\lambda_j^{2d} \log \lambda_j \\ C_j^{d\theta_3} &= B_j(d) \lambda_j^{1+d} \log \lambda_j + \lambda_j^{1+d} B_j^d(d) \end{aligned}$$

for $A_j^{dd}(d) = -(\lambda_j - \pi)^2 A_j(d)/4$ and $B_j^{dd}(d) = -(\lambda_j - \pi)^2 B_j(d)/4$. Then

$$(\hat{d}_{ALPE}, \hat{\theta}'_{ALP})' - (d_0, \theta_0)' = -H^{-1}(\bar{d}, \bar{\theta})S(d_0, \theta_0)$$

for $|(\bar{d}, \bar{\theta})' - (d_0, \theta_0)'| \leq |(\hat{d}_{ALPE}, \hat{\theta}'_{ALP})' - (d_0, \theta_0)'|$. Considering the parameter set $\Delta_n \times \Theta_n = \{(d, \theta) : |\lambda_m^{-(1+d_0)/2}(d - d_0)| < \epsilon, |\lambda_m^{-1/2}(\theta_1 - \theta_{10})| < \epsilon, |\lambda_m^{-(1-d_0)/2}(\theta_2 - \theta_{20})| < \epsilon \text{ and } |\theta_3 - \theta_{30}| < \epsilon\}$ for an arbitrary small $\epsilon > 0$, the asymptotic normality is proved by showing that:

- a) $\sup_{(d, \theta) \in \Delta_n \times \Theta_n} \|D_n^{-1}[H(d, \theta) - J(d, \theta)]D_n^{-1}\| = o_p(1),$
- b) $\sup_{(d, \theta) \in \Delta_n \times \Theta_n} \|D_n^{-1}[J(d, \theta) - J(d_0, \theta_0)]D_n^{-1}\| = o_p(1),$
- c) $D_n^{-1}J(d_0, \theta_0)D_n^{-1} \rightarrow \Omega$
- d) $D_n^{-1}S(d_0, \theta_0) \xrightarrow{d} N\left(-b, \frac{\pi^2}{6}\Omega\right)$

where $J(d, \theta)$ is a 4×4 matrix with elements $[J(d, \theta)]_{(i+1), (k+1)} = \sum_{j=1}^m x_{ij}^\dagger x_{kj}^\dagger$, for $i, k = 0, 1, 2, 3$. The (1,1) element of the left hand side of a) is

$$-\sup_{\Delta_n \times \Theta_n} \sum_{j=1}^m \left(\frac{C_j^{dd}}{1 + C_j} - \frac{(C_j^d)^2}{(1 + C_j)^2} \right)^\dagger (V_j + U_{zj} + \varsigma_j)$$

The term involving V_j is bounded by

$$O\left(\frac{1}{m} \sup_{\Delta_n \times \Theta_n} \sum_{j=1}^m \lambda_j^d \log^2 n [(d - d_0) \log n + \lambda_j^{d_0}]\right) = (\log^3 n \lambda_m^{2d_0}) = o(1)$$

under Assumption 4. Similarly the summand with ς_j is also $o(1)$ since $\varsigma_j = O(\lambda_j^{\beta_1})$. Finally the term involving U_{zj} is $o_p(1)$ using Lemma 3. The other elements are proved to be $o_p(1)$ similarly. b) and c) are easily proved noting the restrictions in the parameter space $\Delta_n \times \Theta_n$ and approximating sums by integrals.

Finally in order to prove d) write $D_n^{-1}S(d_0, \theta_0) = m^{-1/2} \sum N_j(U_{zj} + \varsigma_j)$ for

$$N_j = \left(x_{0j}^\dagger, \frac{\lambda_m^{-d_0}}{\cos(d_0\pi/2)} x_{1j}^\dagger, \lambda_m^{-2d_0} x_{2j}^\dagger, \frac{\lambda_m^{-(1+d_0)}}{\sin(d_0\pi/2)} x_{3j}^\dagger \right)'$$

It is easily shown that $m^{-1/2} \sum N_j \varsigma_j \rightarrow -b$ approximating sums by integrals. The proof is completed by showing that for any vector $v = (v_1, v_2, v_3, v_4)'$, $m^{-1/2} \sum v' N_j U_{zj} \xrightarrow{d} N(0, v' \Omega v \pi^2 / 6)$. Divide now the sum in three parts

$$\frac{1}{\sqrt{m}} \sum_{j=1}^m v' N_j U_{zj} = T_1 + T_2 + T_3$$

where, for $\beta = \max(2, (1 - d_0)^{-1})$ and $1 > \alpha > \max\{1/2, 1/4(1 - d_0)\}$

$$T_1 = \frac{1}{\sqrt{m}} \sum_{j=1}^{\lfloor \log^\beta m \rfloor} v' N_j U_{zj}, \quad T_2 = \frac{1}{\sqrt{m}} \sum_{j=1+\lfloor \log^\beta m \rfloor}^{\lfloor m^\alpha \rfloor} v' N_j U_{zj}, \quad T_3 = \frac{1}{\sqrt{m}} \sum_{j=1+\lfloor m^\alpha \rfloor}^m v' N_j U_{zj}$$

Using Lemma 2 we can proceed as in Hurvich et al (1998) to show that T_1 and T_2 are both $o_p(1)$. Finally, since the elements in the vector N_j satisfy the assumptions in Lemma 6 and $m^{-1} \sum_{j=1+\lfloor m^\alpha \rfloor}^m (v' N_j)^2 = v' \Omega v (1 + o(1))$ we get the desired result. \square

Proof of Theorem 4: Note that

$$\begin{aligned} \hat{\theta}_{2,ALP} - \theta_{20} &= -L_1 H^{-1}(\bar{d}, \bar{\theta}) S(d_0, \theta_0) \\ \hat{\theta}_{2,ALP}^R - \theta_{20} &= -L_2 G^{-1}(\bar{d}, \bar{\theta}) R(d_0, \theta_{20}) \end{aligned}$$

where $G(\cdot)$ and $R(\cdot)$ are the Hessian matrix and the score of the restricted ALPE in Arteche (2006, page 2124) respectively. Defining $D_{n1} = D_{n1}(d_0) = \sqrt{m} \text{diag}(1, \lambda_m^{2d_0})$ we have that

$$\sqrt{m} \lambda_m^{2d_0} (\hat{\theta}_{2,ALP} - \hat{\theta}_{2,ALP}^R) = A_n B_n$$

where $A_n = (-L_1 D_n H^{-1}(\bar{d}, \bar{\theta}) D_n, L_2 D_{n1} G^{-1}(\bar{d}, \bar{\theta}) D_{n1})$ and $B_n = (D_n^{-1} S(d_0, \theta_0), D_{n1}^{-1} R(d_0, \theta_{20}))'$.

The desired result follows from the following convergences under the null hypothesis

$$\begin{aligned} A_n &\xrightarrow{p} (-L_1 \Omega(d_0)^{-1}, L_2 \Omega_2(d_0)^{-1}) \\ B_n &\xrightarrow{d} N\left(0, \frac{\pi^2}{6} \Omega_4(d_0)\right) \end{aligned}$$

for

$$\Omega_4(d_0) = \begin{pmatrix} \Omega(d_0) & \Omega_5(d_0) \\ \Omega_5(d_0)' & \Omega_2(d_0) \end{pmatrix}$$

and $\Omega_5(d_0)$ containing the first and third columns of $\Omega(d_0)$. The convergence of A_n is shown in a), b) and c) in the proof of Theorem 2 for the part related with H and as in formula (A.5) in Arteche (2006) for the terms concerning G . The weak convergence of B_n is shown in d) in the proof of Theorem 2 and as the last formula in the Appendix of Arteche (2006).

Finally

$$(-L_1 \Omega(d_0)^{-1}, L_2 \Omega_2(d_0)^{-1}) \Omega_4(d_0) \begin{pmatrix} -\Omega(d_0)^{-1} L_1' \\ \Omega_2(d_0)^{-1} L_2' \end{pmatrix} = L_1 \Omega(d_0)^{-1} L_1' - L_2 \Omega_2(d_0)^{-1} L_2'$$

for the form of the matrix $\Omega_5(d_0)$. \square

APPENDIX B: TECHNICAL LEMMAS

Lemma 1 is a variant of Lemma 1 in Sun and Phillips (2003) for the stationary case and Theorem 1 in Velasco (1999) for $d_0 \in [1/2, 1)$ and the proof is thus omitted. Lemma

2 extends Lemma 2 in Sun and Phillips (2003) to the nonstationary case and the proof is similar noting the bounds in Lemma 1 to uniformly control for the errors in the approximation of the covariances between normalized discrete Fourier transforms. Lemma 3 is a more general version of Lemma 3 in Sun and Phillips (2003), which we found more convenient for the proofs of the theorems. Its proof is similar using Lemma 2 and it is thus omitted.

Lemma 1 *Let $v_j = w_{zj} = w_{zj}/f_z^{1/2}(\lambda_j)$. Under assumptions 1-3, for any sequences of positive integers j and k such that $1 \leq k < j \leq m$ for $m/n \rightarrow 0$ as $n \rightarrow \infty$,*

$$a) E(v_j \bar{v}_j) = 1 + O(j^{-1} \log j + j^{2d_0-2} \log j),$$

$$b) E(v_j v_j) = O(j^{-1} \log j + j^{2d_0-2} \log j),$$

$$c) E(v_j \bar{v}_k) = O(k^{-1} \log j + (jk)^{d_0-1} \log k),$$

$$d) E(v_j v_k) = O(k^{-1} \log j + (jk)^{d_0-1} \log k).$$

Lemma 2 *Under assumptions 1-3 for $m/n \rightarrow 0$ and $\beta = \max(2, (1 - d_0)^{-1})$,*

$$a) Cov(U_{zj}, U_{zk}) = O(k^{-2} \log^2 j + (jk)^{2d_0-2} \log^2 k), \text{ uniformly for } \log^\beta m \leq k < j \leq m,$$

$$b) \lim_n \sup_{1 \leq j \leq m} EU_{zj}^2 < \infty,$$

$$c) EU_{zj} = O(j^{-1} \log j + j^{2d_0-2} \log j), \text{ uniformly for } \log^\beta m \leq k < j \leq m,$$

$$d) Var(U_{zj}) = \pi^2/6 + O(j^{-1} \log j + j^{2d_0-2} \log j), \text{ uniformly for } \log^\beta m \leq k < j \leq m,$$

Lemma 3 *Let $\{c_j(d, \theta')\}_{j=1}^m$ be a sequence of functions such that for some finite $b > 0$*

$$\begin{aligned} \sup_{\Delta \times \Theta} |c_j - c_{j-1}| &= O(k_{1m} j^{-1}) \text{ uniformly for } 2 \leq j \leq m \\ \sup_{\Delta \times \Theta} |c_m| &= O(k_{2m}), \quad \sup_{\Delta \times \Theta} |c_j| = O\left(\max(k_{1m}, k_{2m}) \frac{\sqrt{m}}{\log^b m}\right) \end{aligned}$$

uniformly for $1 \leq j \leq m$. Then

$$\sup_{\Delta \times \Theta} \left| \frac{1}{m} \sum_{j=1}^m c_j U_{zj} \right| = O_p \left[\max(k_{1m}, k_{2m}) \left(\frac{1}{\sqrt{m}} + \frac{\log^2 m}{m^{2(1-d_0)}} \right) \right]$$

Lemma 4 *For $(d, \theta) \in (\Delta \times \Theta)$ and V_j defined in the proof of Theorem 1,*

$$a) |V_j - V_{j-1}| = O(j^{-1}[|d - d_0| + \lambda_j^d + \lambda_j^{d_0}]), \text{ uniformly for } 2 \leq j \leq m,$$

$$b) |V_j^\dagger| = O(|d - d_0| \log m + \lambda_m^d + \lambda_m^{d_0}), \text{ uniformly for } 1 \leq j < m,$$

$$c) |V_m^\dagger| = O(|d - d_0| + \lambda_m^d + \lambda_m^{d_0}),$$

uniformly in $\Delta \times \Theta$ and if $d \in \Delta'_n = \{d : |d - d_0| < \kappa \lambda_m^v\}$ for some finite constant κ and $v > 0$ arbitrary small,

$$d) |V_j - V_{j-1}| = O(j^{-1}[|d - d_0| + |\theta_1 - \theta_{10}| \lambda_m^{d_0} + |\theta_2 - \theta_{20}| \lambda_m^{2d_0} + |\theta_3 - \theta_{30}| \lambda_m^{1+d_0}]), \text{ uniformly for } 2 \leq j \leq m,$$

$$e) |V_j^\dagger| = O(|d - d_0| \log m + |\theta_1 - \theta_{10}| \lambda_m^{d_0} + |\theta_2 - \theta_{20}| \lambda_m^{2d_0} + |\theta_3 - \theta_{30}| \lambda_m^{1+d_0}), \text{ uniformly for } 1 \leq j < m,$$

$$f) |V_m^\dagger| = O(|d - d_0| + |\theta_1 - \theta_{10}| \lambda_m^{d_0} + |\theta_2 - \theta_{20}| \lambda_m^{2d_0} + |\theta_3 - \theta_{30}| \lambda_m^{1+d_0}),$$

uniformly in $\Delta'_n \times \Theta$.

Proof: a) By the definition of V_j

$$|V_j - V_{j-1}| = \left| 2(d - d_0) \log \left(\frac{j}{j-1} \right) + \log \left(\frac{1 + C_j^0}{1 + C_{j-1}^0} \right) - \log \left(\frac{1 + C_j}{1 + C_{j-1}} \right) \right| \quad (18)$$

The first term of the right hand side of (18) is $O(|d - d_0| j^{-1})$ where hereafter the $O()$ are uniform over $2 \leq j < m$ and $\Delta \times \Theta$ or $\Delta'_n \times \Theta$ where appropriate. Now

$$\log \left(\frac{1 + C_j}{1 + C_{j-1}} \right) = \log \left(1 + \frac{C_j - C_{j-1}}{1 + C_{j-1}} \right) = O(|C_j - C_{j-1}|) = O\left(\frac{1}{j} \lambda_j^d\right)$$

using the following bounds

$$\begin{aligned} |\lambda_j^\alpha - \lambda_{j-1}^\alpha| &= \lambda_j^\alpha \left| 1 - \left(\frac{j-1}{j} \right)^\alpha \right| = O\left(\frac{1}{j} \lambda_j^\alpha\right) \quad \text{for any } \alpha \neq 0 \\ \cos \left[d \left(\frac{\lambda_j}{2} + \frac{\pi}{2} \right) \right] &= \cos \left[d \left(\frac{\lambda_{j-1}}{2} + \frac{\pi}{2} \right) \right] + O\left(\frac{1}{n}\right) \end{aligned}$$

and similarly for the sine term.

d) For the bound in d) note that

$$|V_j - V_{j-1}| \leq \left| 2(d - d_0) \log \left(\frac{j}{j-1} \right) \right| + \left| \log \left(\frac{(1 + C_j^0)(1 + C_{j-1})}{(1 + C_{j-1}^0)(1 + C_j)} \right) \right| \quad (19)$$

where the first term on the right hand side is $O(j^{-1}|d-d_0|)$. The second term is the absolute value of

$$\log \left(1 + \frac{C_j^0 - C_j - (C_{j-1}^0 - C_{j-1}) + C_j^0 C_{j-1} - C_{j-1}^0 C_j}{(1 + C_{j-1}^0)(1 + C_j)} \right)$$

Now $C_j^0 - C_j - (C_{j-1}^0 - C_{j-1})$ is equal to

$$(\theta_{10} - \theta_1) \left[\lambda_j^{d_0} A_j(d_0) - \lambda_{j-1}^{d_0} A_{j-1}(d_0) \right] \quad (20)$$

$$+ \theta_1 \left[A_j(d_0)(\lambda_j^{d_0} - \lambda_j^d) - A_{j-1}(d_0)(\lambda_{j-1}^{d_0} - \lambda_{j-1}^d) \right] \quad (21)$$

$$+ \theta_1 \left[\lambda_j^d (A_j(d_0) - A_j(d)) - \lambda_{j-1}^d (A_{j-1}(d_0) - A_{j-1}(d)) \right] \quad (22)$$

$$+ (\theta_{20} - \theta_2)(\lambda_j^{2d_0} - \lambda_{j-1}^{2d_0}) \quad (23)$$

$$+ \theta_2 \left[\lambda_j^{2d_0} - \lambda_{j-1}^{2d_0} - (\lambda_j^{2d} - \lambda_{j-1}^{2d}) \right] \quad (24)$$

$$+ (\theta_{30} - \theta_3) \left[\lambda_j^{1+d_0} B_j(d_0) - \lambda_{j-1}^{1+d_0} B_{j-1}(d_0) \right] \quad (25)$$

$$+ \theta_3 \left[B_j(d_0)(\lambda_j^{1+d_0} - \lambda_j^{1+d}) - B_{j-1}(d_0)(\lambda_{j-1}^{1+d_0} - \lambda_{j-1}^{1+d}) \right] \quad (26)$$

$$+ \theta_3 \left[\lambda_j^{1+d} (B_j(d_0) - B_j(d)) - \lambda_{j-1}^{1+d} (B_{j-1}(d_0) - B_{j-1}(d)) \right] \quad (27)$$

The expression in (20) is $\theta_{10} - \theta_1$ times

$$(\lambda_j^{d_0} - \lambda_{j-1}^{d_0})A_j(d_0) + \lambda_{j-1}^{d_0}(A_j(d_0) - A_{j-1}(d_0)) = O\left(\frac{1}{j}\lambda_j^{d_0}\right) + O\left(\frac{1}{n}\lambda_{j-1}^{d_0}\right) = O\left(\frac{1}{j}\lambda_j^{d_0}\right)$$

Now, apart from the constant θ_1 , (21) is

$$\begin{aligned} & (\lambda_j^{d_0} - \lambda_j^d)[A_j(d_0) - A_{j-1}(d_0)] + A_{j-1}(d_0)[\lambda_j^{d_0} - \lambda_j^d - (\lambda_{j-1}^{d_0} - \lambda_{j-1}^d)] \\ &= O\left(|d-d_0|\lambda_j^{d_0} \log \lambda_j \frac{1}{n}\right) + O\left(|d-d_0|\lambda_j^{d_0} \log \lambda_j \frac{1}{j}\right) \\ &= O\left(|d-d_0|\lambda_j^{d_0} \log \lambda_j \frac{1}{j}\right) \end{aligned}$$

Similarly, (22) is, apart from θ_1

$$\begin{aligned} & (\lambda_j^d - \lambda_{j-1}^d)[A_j(d_0) - A_j(d)] + \lambda_{j-1}^d[A_j(d_0) - A_{j-1}(d_0) - (A_j(d) - A_{j-1}(d))] \\ &= O\left(\frac{1}{j}\lambda_j^d |d-d_0|\right) + \lambda_{j-1}^d \left[d_0 A'_{j-1}(d_0) - d A'_{j-1}(d) + O\left(\frac{1}{n^2}\right) \right] \\ &= O\left(\frac{1}{j}\lambda_j^d |d-d_0|\right) + \lambda_{j-1}^d \left[O\left(\frac{1}{n^2}\right) + d_0[A'_{j-1}(d_0) - A'_{j-1}(d)] + (d_0-d)A'_{j-1}(d) \right] \\ &= O\left(\frac{1}{j}\lambda_j^d |d-d_0|\right) \end{aligned}$$

for $A'_{j-1}(d) = -\sin[d(\lambda_{j-1}/2 - \pi/2)]\pi/n$. We get similarly the bounds for (25), (26) and (27) with λ_j^{1+d} instead of λ_j^d .

Also (23) is $O(|\theta_{20} - \theta_2|j^{-1}\lambda_j^{2d_0})$ and, apart from the constant θ_2 , (24) is equal to

$$\begin{aligned} & \lambda_j^{2d_0} \left[\left(1 - \frac{1}{j}\right)^{2d} - \left(1 - \frac{1}{j}\right)^{2d_0} \right] + \left[1 - \left(1 - \frac{1}{j}\right)^{2d} \right] (\lambda_j^{2d_0} - \lambda_j^{2d}) \\ &= O\left(\lambda_j^{2d_0} |d - d_0| \frac{1}{j}\right) + O\left(\lambda_j^{2d_0} |d - d_0| \frac{1}{j} \log \lambda_j\right) \\ &= O\left(|d - d_0| \frac{1}{j}\right) \end{aligned}$$

Finally $C_j^0 C_{j-1} - C_{j-1}^0 C_j$ is equal to

$$\begin{aligned} & [C_j^0 - C_j - (C_{j-1}^0 - C_{j-1})]C_{j-1}^0 + (C_j^0 - C_{j-1}^0)(C_{j-1} - C_{j-1}^0) \\ &= O\left(C_{j-1}^0 \frac{1}{j} [|d - d_0| + |\theta_1 - \theta_{10}| \lambda_m^{d_0} + |\theta_2 - \theta_{20}| \lambda_m^{2d_0} + |\theta_3 - \theta_{30}| \lambda_m^{1+d_0}]\right) \end{aligned} \quad (28)$$

from the proof of the bounds in (20)-(27) and because

$$\begin{aligned} C_j^0 - C_{j-1}^0 &= \theta_{10} \lambda_j^{d_0} [A_j(d_0) - A_{j-1}(d_0)] \\ &+ \theta_{10} A_{j-1}(d_0) (\lambda_j^{d_0} - \lambda_{j-1}^{d_0}) + \theta_{20} (\lambda_j^{2d_0} - \lambda_{j-1}^{2d_0}) \\ &+ \theta_{30} \lambda_j^{1+d_0} [B_j(d_0) - B_{j-1}(d_0)] \\ &+ \theta_{30} B_{j-1}(d_0) (\lambda_j^{1+d_0} - \lambda_{j-1}^{1+d_0}) \\ &= O\left(\frac{1}{n} \lambda_j^{d_0}\right) + O\left(\frac{1}{j} \lambda_j^{d_0}\right) + O\left(\frac{1}{j} \lambda_j^{2d_0}\right) = O\left(\frac{1}{j} \lambda_j^{d_0}\right) \end{aligned}$$

$$\begin{aligned} C_{j-1}^0 - C_{j-1} &= (\theta_{10} - \theta_1) \lambda_{j-1}^{d_0} A_{j-1}(d_0) + \theta_1 A_{j-1}(d_0) (\lambda_{j-1}^{d_0} - \lambda_{j-1}^d) \\ &+ \theta_1 \lambda_{j-1}^d [A_{j-1}(d_0) - A_{j-1}(d)] + (\theta_{20} - \theta_2) \lambda_{j-1}^{2d_0} + \theta_2 (\lambda_{j-1}^{2d_0} - \lambda_{j-1}^d) \\ &+ (\theta_{30} - \theta_3) \lambda_{j-1}^{1+d_0} B_{j-1}(d_0) + \theta_3 B_{j-1}(d_0) (\lambda_{j-1}^{1+d_0} - \lambda_{j-1}^{1+d}) \\ &+ \theta_3 \lambda_{j-1}^{1+d} [B_{j-1}(d_0) - B_{j-1}(d)] \\ &= O(|\theta_{10} - \theta_1| \lambda_m^{d_0} + \lambda_m^{d_0} |d - d_0| \log \lambda_m + |\theta_{20} - \theta_2| \lambda_m^{2d_0} + |\theta_{30} - \theta_3| \lambda_m^{1+d_0}) \end{aligned}$$

b) The bound for V_j^\dagger in b) uses the following relations

$$\begin{aligned} \log \lambda_j - \frac{1}{m} \sum \log \lambda_k &= \log j - \frac{1}{m} \sum \log k = O(\log m) \\ \log \left(\frac{1 + C_j^0}{1 + C_j} \right) &= \log \left(1 + \frac{C_j^0 - C_j}{1 + C_j} \right) = O(|C_j^0 - C_j|) = O(\lambda_m^d + \lambda_m^{d_0}) \end{aligned}$$

e) For $d \in \Delta'_n$ we get e) using also the following bounds

$$\begin{aligned} \lambda_j^d &= \lambda_j^{d_0} (1 + o(1)) \quad , \quad \lambda_j^{2d} = \lambda_j^{2d_0} (1 + o(1)) \\ A_j(d) &= A_j(d_0) + O(|d - d_0|) \quad , \quad B_j(d) = B_j(d_0) + O(|d - d_0|) \end{aligned}$$

The bounds of c) and f) are equally obtained noting that $\log \lambda_m - m^{-1} \sum \log \lambda_k = \log m - m^{-1} \sum \log k = 1 + o(1)$. \square

Corollary 2 By Lemmas 3 and 4 $\sup_{\Delta'_n \times \Theta} \left| m^{-1} \sum_{j=1}^m V_j^\dagger U_{zj} \right|$ is bounded in probability by

$$O_p \left[\left(\frac{1}{\sqrt{m}} + \frac{\log^2 m}{m^{2(1-d_0)}} \right) \sup_{\Delta'_n \times \Theta} [|d - d_0| + |\theta_{10} - \theta_1| \lambda_m^{d_0} + |\theta_{20} - \theta_2| \lambda_m^{2d_0} + |\theta_{30} - \theta_3| \lambda_m^{1+d_0}] \right]$$

Lemma 5 For $(d, \theta) \in \Delta'_n \times \Theta$

$$\begin{aligned} \frac{1}{m} \sum_{j=1}^m (V_j^\dagger)^2 &= 4(d - d_0)^2(1 + o(1)) \\ &+ (\theta_{10} - \theta_1)^2 \cos^2 \left(\frac{d_0 \pi}{2} \right) \lambda_m^{2d_0} \frac{d_0^2}{(2d_0 + 1)(1 + d_0)^2} (1 + o(1)) \\ &+ (\theta_{20} - \theta_2)^2 \lambda_m^{4d_0} \frac{4d_0^2}{(4d_0 + 1)(1 + 2d_0)^2} (1 + o(1)) \\ &+ (\theta_{30} - \theta_3)^2 \sin^2 \left(\frac{d_0 \pi}{2} \right) \lambda_m^{2(d_0+1)} \frac{(1 + d_0)^2}{(2d_0 + 3)(2 + d_0)^2} (1 + o(1)) \\ &+ 2(\theta_{10} - \theta_1)(\theta_{20} - \theta_2) \cos \left(\frac{d_0 \pi}{2} \right) \lambda_m^{3d_0} \frac{2d_0^2}{(3d_0 + 1)(d_0 + 1)(1 + 2d_0)} (1 + o(1)) \\ &- 2(\theta_{10} - \theta_1)(\theta_{30} - \theta_3) \cos \left(\frac{d_0 \pi}{2} \right) \sin \left(\frac{d_0 \pi}{2} \right) \lambda_m^{1+2d_0} \frac{d_0}{2(d_0 + 1)(d_0 + 2)} (1 + o(1)) \\ &- 2(\theta_{20} - \theta_2)(\theta_{30} - \theta_3) \sin \left(\frac{d_0 \pi}{2} \right) \lambda_m^{1+3d_0} \frac{2d_0(1 + d_0)}{(3d_0 + 2)(2d_0 + 1)(d_0 + 2)} (1 + o(1)) \\ &+ 4(d - d_0)(\theta_{10} - \theta_1) \cos \left(\frac{d_0 \pi}{2} \right) \lambda_m^{d_0} \frac{d_0}{(d_0 + 1)^2} (1 + o(1)) \\ &+ 4(d - d_0)(\theta_{20} - \theta_2) \lambda_m^{2d_0} \frac{2d_0}{(2d_0 + 1)^2} (1 + o(1)) \\ &+ 4(d - d_0)(\theta_{30} - \theta_3) \lambda_m^{1+d_0} \frac{1 + d_0}{(d_0 + 2)^2} (1 + o(1)) \end{aligned}$$

where the $o(1)$ terms are uniform over $(d, \theta) \in \Delta'_n \times \Theta$.

Proof:

$$\frac{1}{m} \sum_{j=1}^m (V_j^\dagger)^2 = 4(d - d_0)^2 \frac{1}{m} \sum (\log^\dagger \lambda_j)^2 \quad (29)$$

$$+ \frac{1}{m} \sum \left[\log^\dagger \left(\frac{1 + C_j^0}{1 + C_j} \right) \right]^2 \quad (30)$$

$$+ \frac{4(d - d_0)}{m} \sum \log^\dagger \lambda_j \log \left(\frac{1 + C_j^0}{1 + C_j} \right) \quad (31)$$

The right hand side of (29) is $4(d - d_0)^2(1 + O(m^{-1} \log^2 m))$ where hereafter the $o()$ and

$O()$ terms hold uniformly over $(d, \theta) \in \Delta'_n \times \Theta$. Now

$$\log \left(\frac{1 + C_j^0}{1 + C_j} \right) = \log \left(1 + \frac{C_j^0 - C_j}{1 + C_j} \right) = C_j^0 - C_j - \frac{C_j(C_j^0 - C_j)}{1 + C_j} + O[(C_j^0 - C_j)^2]$$

and $C_j^0 - C_j$ is equal to

$$(\theta_{10} - \theta_1)A_j(d_0)\lambda_j^{d_0} + (\theta_{20} - \theta_2)\lambda_j^{2d_0} + (\theta_{30} - \theta_3)B_j(d_0)\lambda_j^{1+d_0} + O(|d - d_0|\lambda_j^{d_0} \log \lambda_j)$$

since $\lambda_j^d = \lambda_j^{d_0} + O(|d - d_0|\lambda_j^{d_0} \log \lambda_j)$ such that

$$\begin{aligned} \log \left(\frac{1 + C_j^0}{1 + C_j} \right) &= (\theta_{10} - \theta_1)A_j(d_0)\lambda_j^{d_0} + (\theta_{20} - \theta_2)\lambda_j^{2d_0} + (\theta_{30} - \theta_3)B_j(d_0)\lambda_j^{1+d_0} \\ &+ O(|d - d_0|\lambda_j^{d_0} \log \lambda_j + |\theta_{10} - \theta_1|\lambda_j^{2d_0} + |\theta_{20} - \theta_2|\lambda_j^{3d_0} + |\theta_{30} - \theta_3|\lambda_j^{1+2d_0}) \end{aligned}$$

Then

$$\begin{aligned} \frac{1}{m} \sum \log^2 \left(\frac{1 + C_j^0}{1 + C_j} \right) &= (\theta_{10} - \theta_1)^2 \cos^2 \left(\frac{d_0\pi}{2} \right) \lambda_m^{2d_0} \frac{1}{2d_0 + 1} (1 + o(1)) \\ &+ (\theta_{20} - \theta_2)^2 \lambda_m^{4d_0} \frac{1}{4d_0 + 1} (1 + o(1)) \\ &+ (\theta_{30} - \theta_3)^2 \sin^2 \left(\frac{d_0\pi}{2} \right) \lambda_m^{2(1+d_0)} \frac{1}{2d_0 + 3} (1 + o(1)) \\ &+ 2(\theta_{10} - \theta_1)(\theta_{20} - \theta_2) \cos \left(\frac{d_0\pi}{2} \right) \lambda_m^{3d_0} \frac{1}{3d_0 + 1} (1 + o(1)) \\ &- 2(\theta_{10} - \theta_1)(\theta_{30} - \theta_3) \cos \left(\frac{d_0\pi}{2} \right) \sin \left(\frac{d_0\pi}{2} \right) \lambda_m^{1+2d_0} \frac{1}{2d_0 + 2} (1 + o(1)) \\ &- 2(\theta_{20} - \theta_2)(\theta_{30} - \theta_3) \sin \left(\frac{d_0\pi}{2} \right) \lambda_m^{1+3d_0} \frac{1}{3d_0 + 2} (1 + o(1)) \\ &+ O(\lambda_m^{2d_0} (d - d_0)^2 \log^2 \lambda_m + |d - d_0| |\theta_{10} - \theta_1| \lambda_m^{2d_0} \log \lambda_m \\ &+ |d - d_0| |\theta_{20} - \theta_2| \lambda_m^{3d_0} \log \lambda_m + |d - d_0| |\theta_{30} - \theta_3| \lambda_m^{1+2d_0} \log \lambda_m) \end{aligned}$$

$$\begin{aligned}
\left[\frac{1}{m} \sum \log \left(\frac{1 + C_j^0}{1 + C_j} \right) \right]^2 &= (\theta_{10} - \theta_1)^2 \cos^2 \left(\frac{d_0 \pi}{2} \right) \lambda_m^{2d_0} \frac{1}{(d_0 + 1)^2} (1 + o(1)) \\
&+ (\theta_{20} - \theta_2)^2 \lambda_m^{4d_0} \frac{1}{(2d_0 + 1)^2} (1 + o(1)) \\
&+ (\theta_{30} - \theta_3)^2 \sin^2 \left(\frac{d_0 \pi}{2} \right) \lambda_m^{2+2d_0} \frac{1}{(d_0 + 2)^2} (1 + o(1)) \\
&+ 2(\theta_{10} - \theta_1)(\theta_{20} - \theta_2) \cos \left(\frac{d_0 \pi}{2} \right) \lambda_m^{3d_0} \frac{1}{(1 + d_0)(2d_0 + 1)} (1 + o(1)) \\
&- 2(\theta_{10} - \theta_1)(\theta_{30} - \theta_3) \cos \left(\frac{d_0 \pi}{2} \right) \sin \left(\frac{d_0 \pi}{2} \right) \lambda_m^{1+2d_0} \frac{(1 + o(1))}{(1 + d_0)(d_0 + 2)} \\
&- 2(\theta_{20} - \theta_2)(\theta_{30} - \theta_3) \sin \left(\frac{d_0 \pi}{2} \right) \lambda_m^{1+3d_0} \frac{1}{(1 + 2d_0)(d_0 + 2)} (1 + o(1)) \\
&+ O(\lambda_m^{2d_0} (d - d_0)^2 \log^2 \lambda_m + |d - d_0| |\theta_{10} - \theta_1| \lambda_m^{2d_0} \log \lambda_m \\
&+ |d - d_0| |\theta_{20} - \theta_2| \lambda_m^{3d_0} \log \lambda_m + |d - d_0| |\theta_{30} - \theta_3| \lambda_m^{1+2d_0} \log \lambda_m)
\end{aligned}$$

since $A_j(d) = \cos(d\pi/2)(1 + O(\lambda_j))$ and similarly for $B_j(d_0)$. The required result for (30) comes by the difference between both. Finally the last three terms of the right hand side of the Lemma come similarly from (31). \square

The following lemma adapts Lemma 4 in Sun and Phillips (2003) to the non stationary case, allowing also for correlation between signal and noise.

Lemma 6 *Let $0 < d_0 < 3/4$ and $c_{kn} = c_k$ be a triangular array for which*

$$\max_k |c_k| = o(m), \quad \sum_{k=[1+m^\alpha]}^m c_k^2 \sim \rho m, \quad \sum_{k=[1+m^\alpha]}^m |c_k|^p = O(m),$$

for all $p \geq 1$ and $1 > \alpha > \max\{1/2, 1/4(1 - d_0)\}$. Then

$$\frac{1}{\sqrt{m}} \sum_{k=[1+m^\alpha]}^m c_k U_{zk} \xrightarrow{d} N \left(0, \frac{\pi^2}{6} \rho \right)$$

Proof: Note first that such an α always exists because $d_0 < 3/4$. The main difference with respect to Lemma 4 in Sun and Phillips (2003) comes from the possibility of $d_0 \geq 1/2$. In view of Lemma 1, in order to have the error terms in the covariance matrix of the normalized discrete Fourier transforms to be $o(m^{-1/2})$ we need to consider only Fourier frequencies λ_k for $m^\alpha < k \leq m$ (compare with the trimming in Velasco 1999). The result follows then as in Robinson (1995), Velasco (1999) or Sun and Phillips (2003).

Acknowledgements

Research supported by Spanish *Ministerio de Ciencia y Tecnología* and *FEDER* grant SEJ2007-61362/ECON and Basque Government grant IT-334-07 (UPV/EHU Econometrics Research Group). I thank Carlos Velasco, Javier Hualde, P.M. Robinson and Fabrizio Iacone for helpful comments.

References

- [1] Andersen, T.G., Bollerslev, T. (1998). Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *Int. Econ. Rev.* 39, 885-905.
- [2] Andersen, T.G., Bollerslev, T., Diebold, F.X, Labys, P. (2003). Modeling and Forecasting Realized Volatility. *Econometrica* 71, 579-625.
- [3] Arteche, J. (2004). Gaussian Semiparametric Estimation in Long Memory in Stochastic Volatility and Signal Plus Noise Models, *J. Econometrics* 119, 131-154.
- [4] Arteche, J. (2006). Semiparametric estimation in perturbed long memory series, *Comput. Stat. Data An.* 52, 2118-2141.
- [5] Arteche, J., Orbe, J. (2009). Bootstrap based bandwidth choice for log-periodogram regression, *J. Time Ser. Anal.* 30, 591-617.
- [6] Bollerslev, T., Jubinski, D. (1999). Equity Trading Volume and Volatility: Latent Information Arrivals and Common Long-Run Dependencies, *J. Bus. Econ. Stat.* 17, 9-21.
- [7] Bound, J., Brown, C., Duncan, G.J., Rodgers, W.L. (1994). Evidence on the validity of cross-sectional and longitudinal labor market data, *J. Labor Econ.* 12, 345-268.
- [8] Breidt, F.J., Crato, N., de Lima, P. (1998). The Detection and Estimation of Long Memory in Stochastic Volatility. *J. Econometrics* 83, 325-348.
- [9] Davidson, R., MacKinnon, J.G. (2004). *Econometric Theory and Methods*. Oxford University Press, New York.
- [10] De Jong, F., Mahieu, R., Schotman, P. (1998). Price discovery in the foreign exchange market: an empirical analysis of the yen/dmark rate, *J. Int. Money Finan.* 17, 5-27.

- [11] Deo, R.S., Hurvich, C.M. (2001). On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models, *Econometric Theory* 17, 686-710.
- [12] Deo, R., Hurvich, C., Lu, Y. (2006). Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment. *J. Econometrics* 131, 29-58.
- [13] Giraitis, L., Robinson, P.M., Samarov, A. (2000). Adaptive semiparametric estimation of the memory parameter, *J. Multivariate Anal.* 72, 183-207.
- [14] Granger, C.W.J. (1966). The typical spectral shape of an economic variable. *Econometrica* 34, 150-161.
- [15] Harvey, A.C. (1998). Long memory in stochastic volatility. In: Knight, J., Satchell, S. (Eds.), *Forecasting Volatility in Financial Markets*, Oxford: Butterworth-Haineman, 307-320.
- [16] Hasbrouck, J. (1993). Assessing the quality of a security market: a new approach to transaction-cost measurement. *Rev. Finan. Stud.* 6, 191-212.
- [17] Henry, M., Robinson, P.M. (1996). Bandwidth choice in Gaussian semiparametric estimation of long range dependence, in: Robinson, P.M., Rosenblatt, M., (Eds.), *Athens Conference on Applied Probability and Time Series, Vol. II. Lecture Notes in Statistics* 115, New York: Springer-Verlag, 220-232.
- [18] Hurvich, C.M., Soulier, P. (2002). Testing for long memory in volatility. *Econometric Theory* 18, 1291-1308.
- [19] Hurvich, C.M., Deo, R., Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter in a long-memory time series. *J. Time Ser. Anal.* 19, 19-46.
- [20] Hurvich, C.M., Moulines, E., Soulier, P. (2005). Estimating long memory in volatility, *Econometrica* 73, 1283-1328.
- [21] Hurvich, C.M., Moulines, E., Soulier, P. (2008). Corrigendum to "Estimating Long Memory in Volatility", *Econometrica* 76, 661-662.

- [22] Hurvich, C.M., Ray, B.K. (2003). The Local Whittle Estimator of Long-Memory Stochastic Volatility, *J. Finan. Econometrics* 1, 445-470.
- [23] Hurvich, C.M., Soulier, P. (2002). Testing for long memory in volatility, *Econometric Theory* 18, 1291-1308.
- [24] Lieberman, O., Phillips, P.C.B. (2008). Refined Inference on Long Memory in Realized Volatility. *Econometric Rev.* 27, 254-267.
- [25] Martens, M., van Dijk, D., de Pooter, M. (2009). Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements, *Int. J. Forecasting* 25, 282-303.
- [26] Robinson, P.M. (1995). Log-periodogram regression of time series with long-range dependence. *Ann. Statist.* 23, 1048-1072.
- [27] Robinson, P.M. (2008). Multiple local whittle estimation in stationary systems. *Ann. Statist.* 36, 2508-2530.
- [28] Sun, Y., Phillips, P.C.B. (2003). Nonlinear log-periodogram regression for perturbed fractional processes, *J. Econometrics* 115, 355-389.
- [29] Sun, Y., Phillips, P.C.B. (2004). Understanding the Fisher Equation, *J. Appl. Econometrics* 19, 869-886.
- [30] Veiga, H. (2006). A two factor Long Memory Stochastic Volatility model, *Working Paper 06-13(03) Statistics and Econometrics Series*, Universidad Carlos III de Madrid.
- [31] Velasco, C. (1999). Non-stationary log-periodogram regression, *J. Econometrics* 91, 325-371.
- [32] Velasco, C. (2000). Non Gaussian log-periodogram regression. *Econometric Theory* 16, 44-79.