Original Article

Journal of Time Series Analysis

First version received March 2009

Published online in Wiley InterScience: 2009

(www.interscience.wiley.com) DOI: 10.1111/j.1467-9892.2009.00638.x

Local Whittle estimation of the memory parameter in presence of deterministic components

Fabrizio Iacone[a,*,†]

We discuss the estimation of the order of integration of a fractional process that may be contaminated by a time-varying deterministic trend or by a break in the mean. We show that in some cases the estimate may still be consistent and asymptotically normally distributed even when the order of magnitude of the spectral density of the fractional process does not dominate the one of the periodogram of the contaminating term. If trimming is introduced, stronger deterministic components may be neglected. The performance of the estimate in small samples is studied in a Monte Carlo experiment.

Keywords: Long memory; persistence; break; deterministic trend; local Whittle estimation.
JEL: C22.

1. INTRODUCTION

We consider a time series x_t, observed at times $t = 1,\ldots,n$, which is composed of two unobservable parts: a deterministic sequence s_t and a zero-mean stochastic process ξ_t such that

$$x_t = \xi_t + s_t. \tag{1}$$

In many economic time series it seems that the deterministic component changes over time: it could include a time trend, for example, or a mean subject to a break. Modelling these terms may sometimes be difficult: in some applications a time trend can be confused with a shift in the mean, or the location of a certain break can be disputed; more often, some features may be neglected altogether.

Incorrectly modelling the deterministic component may result in spurious evidence of strong autocorrelation: Perron (1991) for example considered the same 14 time series for which Nelson and Plosser (1982) did not reject the hypothesis of a unit root using a Dickey–Fuller test, and showed that the conclusion could be reversed in 11 of them if a break was allowed for in 1929. Allowing for fractional integration, unit roots and linear trends may be generalized to

$$\xi_t \in I(\delta) \tag{2}$$

and

$$s_t = \mu t^{\phi - 1/2}, \tag{3}$$

where (2) is the notation commonly used to indicate that ξ_t is integrated of order δ, $\mu \neq 0$ in (3) and δ and $\phi - 1/2$ do not necessarily have to be integers. The consequences of an incorrect specification of the deterministic component in (1) were first analysed by Bhattacharya *et al.* (1983). They showed that the *R/S* statistic computed using x_t, with s_t as in (3) and $\phi \in (0,1/2)$, indicates the presence of the Hurst effect even when $\delta = 0$, the deterministic trend being then mistaken for a stochastic one of order $\delta = \phi$. Trends as in (3) were also considered by Phillips and Shimotsu (2004), who showed that when $\phi > 1$ the local Whittle estimate produces spurious evidence that $\delta = 1$. Undoubtedly, trends as in (3) are of interest because by using fractional powers it is possible to provide a much more refined classification. Also, trends with non-integer powers arise if, for example, d (fractional) differences are taken from a time series with a linear trend, the resulting time series having a time trend with fractional power $1 - d$. For practical purposes, however, it seems that other deterministic components should be considered too. Breaks in the mean are of particular interest: in this case spurious evidence of long memory, when in fact $\delta = 0$, has been discussed, for various estimates, by Teverovsky and Taqqu (1997), Lobato and Savin (1998), Mikosch and Stărică (1999), Diebold and Inoue (2001) and Granger and Hyung (2004).

An incorrect conclusion about the order of integration can lead to the application of inappropriate limit theory, and it can also have important implications for the economic interpretation of the results, for example, because the spurious strong autocorrelation

[a] *University of York*
**Correspondence to: Fabrizio Iacone, Department of Economics and Related studies, University of York, Heslington, York YO10 5DD, UK.*
[†] *E-mail: fi501@york.ac.uk*

could be regarded as a slow response to shocks by the policy-maker or by the agents. It is therefore important to assess in which cases a potential mistake in the specification of the deterministic term may be safely neglected.

In this situation, the estimation of δ may be regarded as analogue to the extraction of a signal from a time series that is contaminated by other terms. Giraitis *et al.* (2001) generalized the class of deterministic components for which the *R/S* and the *V/S* statistics do not report spurious evidence of the Hurst effect; consistent estimation of δ when $\delta > 0$ was discussed by Robinson (1997) for the case in which s_t is a smooth function. Abadir *et al.* (2007) analysed the fully extended local Whittle estimate (which allows for $\delta > 1/2$), and gave general conditions under which neglecting s_t does not affect the limit distribution of the estimate: in terms of (3) these correspond to $\phi \leq \delta$. For the Whittle estimate, however, Heyde and Dai (1996) stated that the condition $\phi < \min(1/4, 1/2 - \delta)$ is sufficient.

Künsch (1986) showed that the Fourier transforms of trends as in (3) (with $\phi < 1/2$) cluster most of their power at the lowest frequencies: the performance of the estimates may therefore be improved by removing those frequencies. Trimming was also discussed by Hurvich *et al.* (2005) for the case in which s_t is a smooth function. Hurvich *et al.* (2005) also showed that tapering may help reduce the required trimming, and possibly improve the small sample performance of the estimate.

In this article, we investigate the consequences of estimating by local Whittle estimation the order of fractional integration in a time series composed by a stationary stochastic process and by a time-varying component. The latter may include a trend as in (3) or a break in the mean.

Our intent is to characterize a class of deterministic components which can be neglected or mis-specified without affecting the limit properties of the estimate. For x_t in (1), ξ_t (2) and s_t in (3), or for s_t being a break in the mean, we derive conditions on δ and ϕ, using $\phi = 1/2$ when the deterministic term s_t is a break in the mean, under which the asymptotic properties of the estimate are not affected by the presence of the deterministic term. These allow for some $\phi > \delta$ even when no trimming is applied, although in this case a minimal number of frequencies is required in the estimation.

We therefore generalize the condition in Abadir *et al.* (2007) at least in the range of values of δ in which both the estimates are defined ($-1/2 < \delta < 1/2$); we also generalize the condition of Heyde and Dai (1996) in the sense that we allow for ϕ to increase, rather than decrease, with δ. When $\delta = 0$ and $\phi = 1/2$, our conditions for consistency and limit normality are the same ones as in Robinson (1997) and in Hurvich *et al.* (2005) respectively; however, we do not require that $\delta \geq 0$ is known in advance, and our conditions are also weaker (and require less trimming, if this is imposed) if δ is larger.

In Section 2, we present the asymptotic theory, and in Section 3 we analyse the small sample properties with a Monte Carlo exercise. We summarize the results in Section 4. The proofs of the theorems are in the Appendix.

2. LOCAL WHITTLE ESTIMATION WITH TIME-VARYING DETERMINISTIC TERMS

We consider a process x_t observed at times $t = 1,\ldots,n$, composed of a deterministic sequence s_t and of a zero-mean stochastic process ξ_t, as in (1). We introduce

$$w_x(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} x_t e^{i\lambda t}$$

as the discrete Fourier transform of x_t, and

$$I_{xx}(\lambda) = w_x(\lambda) w_x(-\lambda)$$

as the periodogram. In the same way we introduce $w_\xi(\lambda)$, $w_s(\lambda)$ for the Fourier transforms of ξ_t and s_t respectively, and $I_{\xi\xi}(\lambda)$, $I_{ss}(\lambda)$ for the periodograms. Finally, we introduce the cross-periodogram between s_t and ξ_t, $I_{s\xi}(\lambda) = w_s(\lambda) w_\xi(-\lambda)$.

For a generic zero-mean stationary process u_t, we also introduce the autocovariance $E(u_t u_{t+s}) = \gamma_{uu}(s)$, $s \in \mathbb{Z}$, and the spectral density $f_{uu}(\lambda)$ so that $\gamma_{uu}(s) = \int_{-\pi}^{\pi} f_{uu}(\lambda)\cos(\lambda s)\mathrm{d}\lambda$. Notice that the process ξ_t has $E(\xi_t) = 0$, but it does not need to be stationary, nor it has to be $I(\delta)$. However, when it is also assumed that ξ_t is stationary, then we indicate its spectral density as $f_{\xi\xi}(\lambda)$.

In the rest of the article we use C for a positive, finite constant, not necessarily always the same one; the operator [.] returns the integer part of a number, whereas ~ indicates that the ratio between the left- and right-hand sides tends to 1.

2.1. The periodograms of some deterministic components

The periodogram of a deterministic fractional trend was first analysed by Künsch (1986), who advocated trimming to remove the potential effects of that term on the estimate; further discussion is in Robinson and Marinucci (2000) and in Phillips and Shimotsu (2004). A reference to the exact order of magnitude of the periodogram of the shift in the mean is in Mikosch and Stărică (1999).

We characterize the deterministic component in terms of a bound for the order of magnitude of the periodogram $I_{ss}(\lambda_j)$ at the Fourier frequencies ($\lambda_j = \frac{2\pi j}{n}$, for some integer j, $0 < j < n/2$) used in the local Whittle estimation. Together with the deterministic trend and the shift in the mean, we also consider the single impulse, a deterministic term having $s_t \neq 0$ only for one value of t. Deterministic trends provide a general classification and allow us to compare our results with the rest of the literature, whereas shifts in the mean are important because they are often considered in applied analysis; the single impulse model had less theoretical importance, but we consider it because it emerges when integer differences of a shift in the mean are taken, a procedure that is very common when $\delta > 1/2$. We summarize some results for these three types of deterministic terms in Theorem 1.

Theorem 1. (*i*) *Shift in the mean. If* $s_t = \mu_1$ *for* $t \le [\tau n]$, $s_t = \mu_2$ *for* $t > [\tau n]$, *where* $\tau \in (0,1), |\mu_1| < \infty, |\mu_2| < \infty$ *and* $\mu_1 \neq \mu_2$, *then, for j such that* $0 < j < [n/2]$,

$$I_{ss}(\lambda_j) \le C\lambda_j^{-1} j^{-1}. \tag{4}$$

(*ii*) *Fractional trend. If* $s_0 = 0$, $s_t = O((t+1)^{\phi-1/2})$, $|s_t - s_{t+1}| = O\left(\frac{|s_t|}{t}\right)$, $\phi \in (-1/2,1/2)$, *then, for j such that* $0 < j < [n/2]$,

$$I_{ss}(\lambda_j) \le C\lambda_j^{-2\phi} j^{-1}. \tag{5}$$

(*iii*) *Single impulse. If* $s_t = \mu_3$ for $t = [\tau n]$, $s_t = 0$ for $t \neq [\tau n]$, *where* $\tau \in (0,1)$, $0 < |\mu_3| < \infty$ *then*

$$I_{ss}(\lambda) \le Cn^{-1}.$$

Notice that s_t in case (ii) includes (3), with $\phi \in (-1/2,1/2)$.

These bounds can then be indexed by ϕ, breaks in the mean and single impulses having $\phi = 1/2$ and $\phi = -1/2$ respectively. Exact results may sometimes be provided, as Phillips and Shimotsu (2004) did for trends having $\phi > 1/2$. The bounds as in Theorem 1, however, provide a more general system of classification of the deterministic components: Robinson (1997), for example, showed that the non-parametric mean function he considered has periodogram which can be bounded as in (4).

These bounds have an immediate practical application. Consider the case of ξ_t being a stationary and invertible process, and such that its spectral density can be approximated as $f_{\xi\xi}(\lambda) \sim G\lambda^{-2\delta}$ as $\lambda \to 0^+$, for $G \in (0,\infty)$: roughly speaking, consistent estimation of δ then depends on whether $G\lambda_j^{-2\delta}$ is of order bigger than $I_{ss}(\lambda_j)$ in a certain set of Fourier frequencies, and the orders as stated in Theorem 1 allow a very easy comparison. For example, we can immediately verify that, when $\delta > \phi$, the order of magnitude of $f_{\xi\xi}(\lambda_j)$ dominates that of $I_{ss}(\lambda_j)$ on all the relevant Fourier frequencies. However, it is also possible to obtain consistent estimates if $\delta = \phi$ and indeed if $\delta < \phi$: because of the damping factor j^{-1} in (4) and (5), the effect of the contamination s_t is confined to the lowest frequencies only. Therefore, by including enough frequencies in which $f_{\xi\xi}(\lambda_j)$ dominates, it is possible that the effect of the contamination is rendered negligible. Moreover, by trimming those few frequencies in which $I_{ss}(\lambda_j)$ may dominate, the properties of the estimate of δ may be improved. This treatment is the opposite of the one for a standard signal plus noise model, where the contamination is originated by a weakly dependent or by a fractionally integrated noise of lower order, and the highest frequencies are trimmed instead.

2.2. Robust estimation of the memory parameter

The local Whittle estimate $\hat{\delta}$ is the minimizer, with respect to $d \in [\Delta_1,\Delta_2] \subset (-1/2,1/2)$, of the expression

$$R(d) = \ln\left\{\frac{1}{m-l+1}\sum_{j=l}^{m}\lambda_j^{2d} I_{xx}(\lambda_j)\right\} - 2d\frac{1}{m-l+1}\sum_{j=l}^{m}\ln(\lambda_j), \tag{6}$$

where l and m are integers such that $1 \le l < m \le [n/2]$. This is a slight generalization of the function originally considered by Robinson (1995b), who set $l = 1$: when $l > 1$, one or more of the lowest frequencies are trimmed.

The loss function (6) is based on Giraitis and Robinson (2003), although no time-varying deterministic term was present in x_t in that case.

We discuss consistency of $\hat{\delta}$ in Theorem 2 and limit distribution in Theorem 3 for some cases in which the deterministic component is not necessarily constant.

To prove consistency, we introduce the following assumptions.

Assumption 1. *Let* $\xi_1,\ldots,\xi_n$, $n > 1$, *be a sequence of random variables with* $E(\xi_t) = 0$ *for any t, and let*

$$\zeta_j = \frac{\lambda_j^{2\delta} I_{\xi\xi}(\lambda_j)}{G}. \tag{7}$$

There are $G \in (0,\infty)$ *and* $\Delta_1, \delta, \Delta_2$ *such that* $\delta \in [\Delta_1,\Delta_2] \subset (-1/2,1/2)$, *and*

$$\frac{1}{[\theta m]-l+1}\sum_{j=l}^{[\theta m]}\zeta_j \to_p 1 \tag{8}$$

as $l/m + m/n \to 0$ for any $\theta \in (0,1]$, *and*

$$E(\zeta_j) < C \quad \text{for } l \le j \le m. \tag{9}$$

Assumption 2. *Let* s_t *be such that, for* $0 < j < [n/2]$,

$$I_{ss}(\lambda_j) < C|\lambda_j|^{-2\phi} j^{-1}. \tag{10}$$

Assumption 3. *Let* $l = \max(1,[c_\upsilon n^\upsilon])$, *where* $c_\upsilon \in (0,\infty)$, *and* $m = \max(l+1,[c_\kappa n^\kappa])$, *where* $c_\kappa \in (0,\infty)$, *and* υ *and* κ *are such that*

$$0 \le \upsilon < \kappa < 1 \tag{11}$$

and, for ϕ defined in Assumption 2,

$$\phi < \delta + \frac{1}{2}\frac{\kappa}{1-\upsilon}. \tag{12}$$

Assumption 1 was introduced by Dalla *et al.* (2006; henceforth DGH) and is discussed therein. It is not primitive, but it has the advantage that it may be derived from many different models for ξ_t. For example, under regularity conditions Assumption 1 is met when ξ_t is a stationary and invertible $I(\delta)$ stochastic process. DGH (2006) discussed other models, also allowing for nonstationarity and nonlinearity; they also considered the case in which ξ_t is a signal plus noise process, assuming in that case that the order of integration of the signal is larger than the order of integration of the noise.

We characterized the deterministic component in Assumption 2. It is semi-parametric in the sense that it does not require knowledge of the model, but only of an upper bound. As such, s_t may be constant, or one of the deterministic components characterized in Theorem 1, or a smooth trend as in Robinson (1997) and in Hurvich *et al.* (2005), or a linear combination of several deterministic terms: according to Assumption 2 only the one with periodogram having the largest order is relevant in that case. The price paid for such a general approach is that we cannot introduce tapering as in Hurvich *et al.* (2005) to deal with specific formulations of s_t, because, for example, we do not assume for s_t the smoothness required by these authors.

We characterized l and m in Assumption 3: we restricted m and l to be proportional to n^{κ} and n^{υ} respectively so as to allow a simple computation of the (possibly stochastic) orders of magnitude of the weighted averages of $I_{ss}(\lambda_j)$ and of $I_{s\xi}(\lambda_j)$. When $\delta \geq \phi$, (12) is vacuous, because $\frac{\kappa}{1-\upsilon} > 0$: in this case trimming is never necessary, and we only imposed a very mild restriction on Assumption 4 of Robinson (1995b). When $\phi > \delta$, however condition (12) becomes relevant: for given c_κ, it imposes that a minimal number of frequencies is used, so that enough frequencies in which $G\lambda_j^{-2\delta}$ dominates $I_{ss}(\lambda_j)$ are included, and it only allows to reduce the bandwidth m if trimming is introduced.

The condition (12) indicates that the stronger the autocorrelation (measured by δ), the higher the order of the deterministic trend that can be ignored. Higher trends can also be neglected the larger κ and υ are, because high κ means including more frequencies in which the stochastic rather than the deterministic component dominates the order of magnitude of the periodogram of x_t [owing to the damping factor j^{-1} in (10)]; higher υ has similar implications, because it means that less periodograms in which the deterministic component may be relevant are used in the estimation.

Theorem 2. *Under Assumptions* 1, 2 *and* 3,

$$\hat{\delta} \to_p \delta \quad \text{as } n \to \infty.$$

Under a stronger set of conditions, Robinson (1995b) also derived the limit distribution of the estimate $\hat{\delta}$. We repeat these next, updating them so as to take the deterministic component into account as well.

Assumption 1′. *Let ξ_t be a zero-mean, stationary process, with spectral density $f_{\xi\xi}(\lambda)$ such that, for some $\beta \in (0,2]$,*

$$f_{\xi\xi}(\lambda) \sim G\lambda^{-2\delta}(1 + O(\lambda^{\beta})) \quad \text{as } \lambda \to 0^+,$$

where $G \in (0,\infty)$ and $\delta \in [\Delta_1,\Delta_2] \subset (-1/2,1/2)$, $\Delta_1 < \delta < \Delta_2$.

Assumption 2′. *The sequence ξ_t is such that*

$$\xi_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \sum_{j=0}^{\infty} \psi_j^2 < \infty,$$

where

$$E(\varepsilon_t \,|\, \mathcal{F}_{t-1}) = 0, \quad E(\varepsilon_t^2 \,|\, \mathcal{F}_{t-1}) = 1, \ \text{a.s.}, t = 0, \pm 1, \ldots$$
$$E(\varepsilon_t^3 \,|\, \mathcal{F}_{t-1}) = c_1, \quad E(\varepsilon_t^4 \,|\, \mathcal{F}_{t-1}) = c_2, \ \text{a.s.}, t = 0, \pm 1, \ldots$$

in which $\mathcal{F}_t$ is the σ-field generated by ε_s, $s \leq t$, c_1 and c_2 are some finite constants and there exists a random variable ϵ such that $E(\epsilon^2) < \infty$ and for all $\eta > 0$ and some $C > 0$, $P(|\varepsilon_t| > \eta) \leq CP(|\epsilon| > \eta)$.

Assumption 3′. *In a neighbourhood $(0,\alpha)$ of the origin, $\psi(\lambda) = \sum_{s=0}^{\infty} \psi_s e^{i\lambda s}$ is differentiable and*

$$\frac{d}{d\lambda}\psi(\lambda) = O\left(\frac{|\psi(\lambda)|}{\lambda}\right) \quad \text{as } \lambda \to 0^+.$$

Assumption 4′. *Assumptions* 2 *and* 3 *hold and*

$$0 \leq \upsilon < \kappa < 2\beta/(1+2\beta), \tag{13}$$

$$\phi < \delta + \frac{1}{4}\frac{\kappa}{1-\upsilon}. \tag{14}$$

Assumptions 1′ to 3′ require that ξ_t is a stationary process, thus restricting the class of processes considered. These assumptions summarize those proposed for ξ_t in Robinson (1995b), apart from $\Delta_1 < \delta < \Delta_2$. Notice that since the choice of Δ_1 and Δ_2 is free, it is possible to set them so that the condition $\Delta_1 < \delta < \Delta_2$ is always met: DGH (2006), for example, set $\Delta_1 = -1/2$, $\Delta_2 = 1/2$. The largest possible upper bound for κ, $2\beta/(1 + 2\beta)$, is for $\beta = 2$, a class that also includes the case in which ξ_t is an ARFIMA process, see DGH (2006), page 216.

Condition (13) implies Assumption A4′ of Robinson (1995b): when $\delta \geq \phi$, (13) only means that the rate for m in Robinson (1995b) is specified as proportional to powers of n. When $\phi > \delta$, the ranges for κ and υ may be severely restricted by (13) and (14). Condition (14) on ϕ is stronger than that in (12) because if the periodogram $I_{ss}(\lambda_j)$ is proportional to $\lambda_j^{-2\phi} j^{-1}$, when $\phi > \delta$ it may at least induce a positive lower-order bias. Condition (14) then ensures that the contribution from the deterministic component dominates in $I_{xx}(\lambda_j)$ on such a little range of frequencies that this potential effect has order smaller than $1/\sqrt{m}$. Therefore, root-m consistency without trimming is still possible even for some ϕ for which $\delta \geq \phi$ is not met.

Theorem 3. *Under Assumptions* 1′, 2′, 3′ *and* 4′

$$\sqrt{m}(\hat{\delta} - \delta) \rightarrow_d \mathrm{N}\left(0, \frac{1}{4}\right) \quad \textit{as } n \rightarrow \infty. \tag{15}$$

This is a noteworthy result, because it means that it is possible to choose the bandwidth and the trimming so that the limit distribution of the estimate is robust even to a break in the mean. For small values of δ, it may also be necessary to require a certain smoothness in the low frequency approximation of the spectral density in Assumption 1′, but, at least when $\beta = 2$, it is always possible to choose κ and υ so that (13) and (14) are met for any eligible δ.

3. MONTE CARLO EVIDENCE

Conditions (11), (12) and (13), (14) may restrict the range of κ and υ so that the bandwidth m is bound to be close to the trimming parameter l, when the sample is small, thus leaving only a few frequencies available for the estimation. Moreover, when ϕ approaches $\delta + \frac{1}{2}\frac{\kappa}{1-\upsilon}$ or $\delta + \frac{1}{4}\frac{\kappa}{1-\upsilon}$, the performance of the estimate in samples of moderate size may be at odds with the asymptotic properties. We then analysed these situations by means of a small Monte Carlo exercise.

We considered three deterministic structures and two stochastic components; in each situation we compared the estimate with and without trimming.

The case of no deterministic structure, $s_t = 0$, was our benchmark. Following Bhattacharya *et al.* (1983) and other works in the literature, we allowed for a fractional trend and set $s_t = 2t^{-1/4}$, corresponding to $\phi = 1/4$. The last deterministic structure we considered is the shift in the mean, posing it in the middle of the sample, so $s_t = 0$ for $t \leq n/2$ and $s_t = 1$ for $t > n/2$. For the stochastic component ξ_t, we set $\delta = 0$ and $\delta = 0.4$: since it is the difference $\delta - \phi$ that matters asymptotically, we considered in this way quite a wide range of situations. A large δ was also important to analyse a case in which the condition stated by Heyde and Dai (1996) is not met. The data were generated as a sequence of independent standard normals for $\delta = 0$, and using the Davies and Harte (1987) simulator for $\delta = 0.4$.

We set the bandwidth and the trimming parameter as $m = [0.8n^{0.79}]$, $l = [1 + 0.2n^{0.62}]$, and employed $n = 64, 128, 256, 512, 1024$, with 1000 replications. Since the local Whittle estimate does not have a closed-form formulation, we used the log-periodogram regression estimate, trimming the lowest frequencies, as a starting value in the numerical optimization.

In the rest of the section and in the tables, we refer to the three deterministic models for s_t ($s_t = 0$, $s_t = 2t^{-1/4}$, and $s_t = 0$ if $t \leq n/2$, $s_t = 1$ if $t > n/2$) as I, II and III respectively, while we refer to $\delta = 0$ for ξ_t as model a, and to $\delta = 0.4$ as model b; l and 1 distinguish the case in which trimming was applied or not, so for example we use (1,a,III) to indicate the case in which no trimming was applied, $\delta = 0$ and s_t was a break in the mean. Also, we distinguish between $\hat{\delta}^{(1)}$ and $\hat{\delta}^{(l)}$, the former referring to the estimate when $l = 1$, the latter to the estimate when trimming was applied.

Making use of Theorems 2 and 3, all the combinations yield consistent and asymptotically normal estimates under that rather aggressive trimming; without it, root-m convergence may fail for (1,a,II), and consistency for (1,a,III).

In Tables 1 and 2 we reported for $\hat{\delta}^{(1)}$ and $\hat{\delta}^{(l)}$ the average of the deviations of the estimated values from δ (bias), the sample standard deviation (SD) of the estimated values and the one prescribed by the asymptotic theory (ASD) and the root of the sample mean squared error (RMSE). Finally, in the columns $t_{\hat{\delta}^{(1)}}$ and $t_{\hat{\delta}^{(l)}}$ we reported 100 times the percentage in which the standardized statistic $2\sqrt{m}(\hat{\delta} - \delta_0)$ to test H_0:{$\delta = \delta_0$} vs. H_1:{$\delta > \delta_0$} (where δ_0 is 0 or 0.4 according to the situation) exceeded the critical value of a 5% significance test. These statistics were collected in Table 1 for $\delta = 0$, and in Table 2 for $\delta = 0.4$. Notice that two measures are presented there for the ASD: under the column $\hat{\delta}^{(1)}$ we report $1/\sqrt{4m}$, while under the column $\hat{\delta}^{(l)}$ we propose as an alternative reference $1/\sqrt{4S(l,m)}$ where

$$S(l,m) = \sum_{j=l}^{m} v_j^2 \quad \text{where } v_j = \ln j - \frac{1}{m-l+1}\sum_{k=l}^{m} \ln k. \tag{16}$$

The factor $S(l,m)$ is such that $\frac{1}{m}S(l,m) \to 1$ as $1/l + l/m \to 0$, so it does not appear in Theorem 3. It was originally suggested by Geweke and Porter-Hudak (1983) for the formula of the variance of the log-periodogram regression estimate, and its use to explicitly take into account the effect of trimming in small samples was discussed by Deo and Hurvich (2001); for a local Whittle-type estimate it was used by Hurvich *et al.* (2005). This measure is also used in the last column ($t^*_{\hat{\delta}^{(l)}}$) where the empirical size of the test was computed standardizing $(\hat{\delta} - \delta_0)$ by $1/\sqrt{4S(l,m)}$ instead.

Despite the smallness of the samples, the results were broadly in line with the theory, at least if we only consider the main features. We found that the bias was always quite small but for the case (1,a,III). Not surprisingly, (1,a,II) was the only other one exhibiting a certain systematic deviation from the true value. Given that the periodogram of the deterministic component may still dominate in the frequencies closer to 0, a small residual bias, which should vanish at a rate faster than root-m, can still appear in small samples even when the conditions for Theorem 3 are met: this residual bias was largest in the case (1,b,III) and it seems fair to conjecture that it depends on the gap $\delta - \phi$ and on the trimming.

The bias generated by the deterministic component did not affect the dispersion at least if the hypotheses for Theorem 2 were met. Trimming, however, had a strong effect on the dispersion. The approximation $(2\sqrt{m})^{-1}$ derived in Theorem 3 did not seem very close to it even in the largest sample, whereas $(2\sqrt{S(l,m)})^{-1}$ yielded satisfactory results.

Table 1. Monte Carlo bias, standard deviation (SD) and root mean square error (RMSE) of the estimate and size of 5% tests H_0:$\{\delta = \delta_0\}$, $\delta = 0$

		Bias		SD		ASD		RMSE		Size of 5% tests		
s	n	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(l)}$	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(l)}$	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(l)}$	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(l)}$	$t_{\hat{\delta}^{(1)}}$	$t_{\hat{\delta}^{(l)}}$	$t^*_{\hat{\delta}^{(l)}}$
I	64	−0.020	−0.020	0.144	0.225	0.109	0.207	0.145	0.226	0.071	0.182	0.046
	128	−0.015	−0.020	0.101	0.171	0.083	0.162	0.102	0.172	0.056	0.171	0.044
	256	−0.011	−0.014	0.073	0.118	0.063	0.114	0.074	0.119	0.049	0.165	0.041
	512	−0.007	−0.012	0.053	0.083	0.048	0.087	0.053	0.084	0.047	0.143	0.041
	1024	−0.005	−0.006	0.038	0.060	0.036	0.059	0.038	0.061	0.043	0.141	0.046
II	64	0.056[a]	0.015	0.141	0.225	0.109	0.207	0.152[a]	0.226	0.182[a]	0.226	0.062
	128	0.062[a]	0.008	0.095	0.173	0.083	0.162	0.114[a]	0.173	0.209[a]	0.219	0.062
	256	0.065[a]	0.008	0.066	0.118	0.063	0.114	0.092[a]	0.118	0.294[a]	0.210	0.063
	512	0.065[a]	0.005	0.047	0.083	0.048	0.087	0.080[a]	0.083	0.390[a]	0.184	0.069
	1024	0.062[a]	0.008	0.035	0.060	0.036	0.059	0.072[a]	0.060	0.532[a]	0.193	0.068
III	64	0.234[b]	0.063	0.106	0.216	0.109	0.207	0.257[b]	0.225	0.689[b]	0.315	0.100
	128	0.246[b]	0.031	0.067	0.164	0.083	0.162	0.255[b]	0.167	0.943[b]	0.256	0.075
	256	0.247[b]	0.027	0.043	0.118	0.063	0.114	0.251[b]	0.121	1.000[b]	0.259	0.082
	512	0.249[b]	0.014	0.028	0.082	0.048	0.087	0.251[b]	0.083	1.000[b]	0.233	0.063
	1024	0.252[b]	0.017	0.019	0.060	0.036	0.059	0.252[b]	0.062	1.000[b]	0.240	0.089

[a]The limit normal distribution as does not follow from Theorem 3.
[b]Consistency as does not follow from Theorem 2.
ASD (asymptotic SD) is $(2\sqrt{m})^{-1}$ in column $\hat{\delta}^{(1)}$, and $(2\sqrt{S(l,m)})^{-1}$ in column $\hat{\delta}^{(l)}$.
$t^*_{\hat{\delta}^{(l)}}$ is the empirical size of the 5% tests standardized by $(2\sqrt{S(l,m)})^{-1}$.

Table 2. Monte Carlo bias, standard deviation (SD) and root mean square error (RMSE) of the estimates and size of 5% tests H_0:$\{\delta = \delta_0\}$, $\delta = 0.4$

		Bias		SD		ASD		RMSE		Size of 5% tests		
s	n	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(l)}$	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(l)}$	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(l)}$	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(l)}$	$t_{\hat{\delta}^{(1)}}$	$t_{\hat{\delta}^{(l)}}$	$t^*_{\hat{\delta}^{(l)}}$
I	64	−0.029	−0.042	0.148	0.143	0.109	0.207	0.151	0.223	0.062	0.130	0.029
	128	−0.019	−0.034	0.108	0.172	0.083	0.162	0.109	0.176	0.069	0.162	0.031
	256	−0.012	−0.022	0.074	0.140	0.063	0.114	0.075	0.121	0.056	0.129	0.034
	512	−0.007	−0.015	0.053	0.118	0.048	0.087	0.054	0.084	0.052	0.115	0.036
	1024	−0.004	−0.011	0.040	0.109	0.036	0.059	0.040	0.058	0.046	0.103	0.031
II	64	−0.016	−0.029	0.143	0.160	0.109	0.207	0.144	0.212	0.079	0.150	0.031
	128	−0.011	−0.028	0.107	0.184	0.083	0.162	0.108	0.176	0.076	0.172	0.035
	256	−0.006	−0.016	0.075	0.156	0.063	0.114	0.075	0.120	0.068	0.142	0.037
	512	−0.002	−0.012	0.054	0.136	0.048	0.087	0.054	0.084	0.076	0.130	0.044
	1024	0.000	−0.009	0.041	0.116	0.036	0.059	0.041	0.058	0.058	0.108	0.035
III	64	0.039	−0.014	0.147	0.191	0.109	0.207	0.153	0.221	0.158	0.181	0.041
	128	0.036	−0.016	0.107	0.202	0.083	0.162	0.113	0.173	0.177	0.190	0.049
	256	0.029	−0.014	0.075	0.163	0.063	0.114	0.080	0.122	0.149	0.148	0.040
	512	0.024	−0.010	0.054	0.126	0.048	0.087	0.059	0.082	0.146	0.120	0.040
	1024	0.022	−0.006	0.039	0.120	0.036	0.059	0.045	0.057	0.160	0.113	0.039

ASD (asymptotic SD) is $(2\sqrt{m})^{-1}$ in column $\hat{\delta}^{1}$, and $(2\sqrt{S(l,m)})^{-1}$ in column $\hat{\delta}^{(l)}$.
$t^*_{\hat{\delta}^{(l)}}$ is the empirical size of the 5% tests standardized by $(2\sqrt{S(l,m)})^{-1}$.

Since trimming increased the dispersion, $\hat{\delta}^{(1)}$ was always superior in RMSE sense in the cases in which the conditions for Theorem 3 were met, despite the potential residual bias; even when only the conditions for consistency were met, $\hat{\delta}^{(1)}$ was still roughly equivalent to $\hat{\delta}^{(l)}$, the latter only being clearly superior in a RMSE sense when the gap $\phi - \delta$ was so large that the conditions for consistency of $\hat{\delta}^{(1)}$ failed altogether.

We also analysed the limit normality as in (15) by looking at how often the test statistic to test H_0:{$\delta = \delta_0$} vs. H_1:{$\delta > \delta_0$} exceeded the 5% critical value.

The residual bias had a certain impact even when the conditions for Theorem 3 still held; when these were not met, as in (1,a,II) and (1,a,III) the empirical sizes were well above 5%. Trimming also shifted the size above 5%, as we can see by comparing the sizes $t_{\hat{\delta}^{(1)}}$ and $t_{\hat{\delta}^{(l)}}$ in the benchmark situation of no trends, but when the dispersion inflation factor was taken into account, as in $t^*_{\hat{\delta}^{(l)}}$, the empirical size was satisfactory.

If trimming is not necessary, this Monte Carlo exercise confirmed that it is advisable to set $l = 1$, especially when the sample is small. However, when there is no additional information on ϕ and δ, then it may be a safer strategy. A correction for the variance should, anyway, be taken into account.

Finally, we mention that it is fair to expect that the performance of the estimate in small samples may also depend on the scale of the deterministic component: for example, on μ^2 when the deterministic component is a trend as in (3), and on $(\mu_1 - \mu_2)^2$ when it is a shift in the mean. In the latter case, the location of the break, τ, may also be of interest. Further simulations (not included here) showed that Theorems 2 and 3 provided a poorer guide to the small sample performances when $(\mu_1 - \mu_2)^2$ was larger, and the same may be conjectured for μ^2. We also found that the small sample bias induced by the break in the mean decreased slightly when the break was not set in the middle of the sample, but was closer to either the beginning or the end of the sample.

4. CONCLUSION

We have studied the local Whittle estimate of the memory parameter in the presence of time-varying deterministic components. We have found that it is possible to define the class of deterministic components that do not affect the asymptotic properties of the estimates by means of a simple bound on $\phi - \delta$ (using $\phi = 1/2$ for the break in the mean), and that this class may be further extended when the lowest frequencies are trimmed.

The Monte Carlo experiment has shown that the asymptotic properties can be observed already in samples of moderate size, but also that trimming may inflate the variance. We have discussed a correction factor that, at least in our experiment, has improved the precision of the approximation of the dispersion.

APPENDIX

We present the proofs of the theorems in the first subsection; some technical lemmas which we used in the arguments are discussed in the second subsection.

A.1. Proofs of the theorems

Proof of Theorem 1. Part (i), shift in the mean. First, we show that $|\sum_{t=1}^n e^{i\lambda t}s_t| \le C|\lambda|^{-1}$ when $0 < |\lambda| < \pi$:

$$\left|\sum_{t=1}^{n} s_t e^{i\lambda t}\right| = \left|\mu_1 \sum_{t=1}^{[\tau n]} e^{i\lambda t} + \mu_2 \sum_{t=[\tau n]+1}^{n} e^{i\lambda t}\right| = \left|(\mu_1 - \mu_2)\sum_{t=1}^{[\tau n]} e^{i\lambda t} + \mu_2 \sum_{t=1}^{n} e^{i\lambda t}\right| \le C|\lambda|^{-1},$$

where we used $|\sum_{t=r}^s e^{i\lambda t}| \le \frac{C(s-r)}{1+(s-r)|\lambda|}$, as discussed in Lem 3.2 of Robinson and Marinucci (2001). The periodogram on $|\lambda| \in (0,\pi)$ is then bounded as $I_{ss}(\lambda) \le \frac{C}{n}|\lambda|^{-2}$, and (4) follows after replacing λ by $2\pi j/n$ ($0 < j < [n/2]$).

Part (ii), fractional trend. By Lem 5 of Robinson and Marinucci (2003), $|\sum_{t=1}^n s_t e^{i\lambda t}| \le C|\lambda|^{-\phi-1/2}$, so $I_{ss}(\lambda) \le C/n|\lambda|^{-2\phi-1}$. The bound (5) then follows on replacing $2\pi j/n$ (with $0 < j < [n/2]$) to λ.

Part (iii), single impulse. In this case, $I_{ss}(\lambda) = \frac{1}{2\pi n}|\mu_3 e^{i\lambda[\tau n]}|^2 = \frac{(\mu_3)^2}{2\pi n}$. □

Proof of Theorem 2. To simplify the notation, let

$$m^* = m - l + 1. \tag{A.1}$$

Notice that, by Assumption 3, this means that

$$m/m^* < C \text{ for any } n \quad \text{and} \quad m/m^* \to 1 \text{ as } n \to \infty. \tag{A.2}$$

We follow DGH (2006). Let

$$\eta_j^* = \frac{I_{xx}(\lambda_j)}{G\lambda_j^{-2\delta}},$$

following the proofs of Thm 1 and Lem 2 of DGH, it can be seen that $\hat{\delta}$ is a consistent estimate of δ if

$$E(\eta_j^*) \le C \quad \text{for all } l \le j \le m \tag{A.3}$$

and

$$\frac{1}{[\theta m] - l + 1}\sum_{j=l}^{[\theta m]} \eta_j^* \to_p 1 \quad \text{as } n \to \infty, \tag{A.4}$$

where l and m satisfy (11). Letting

$$v_j = G^{-1}\lambda_j^{2\delta} I_{\xi s}(\lambda_j),\ \ v_j^* = G^{-1}\lambda_j^{2\delta} I_{s\xi}(\lambda_j),\ \ z_j = G^{-1}\lambda_j^{2\delta} I_{ss}(\lambda_j) \tag{A.5}$$

and recalling (7),

$$\eta_j^* = \zeta_j + v_j + v_j^* + z_j. \tag{A.6}$$

Since $\frac{1}{m^*}\sum_{j=l}^{m}\zeta_j \to_p 1$ as $n \to \infty$ as assumed in (8), condition (A.4) requires that

$$\frac{1}{[\theta m] - l + 1}\sum_{j=l}^{[\theta m]} (v_j + v_j^* + z_j) \to_p 0 \quad \text{as } n \to \infty. \tag{A.7}$$

We only discuss $\theta = 1$, the cases $\theta \in (0,1)$ following in the same way. From Assumption 2,

$$\frac{1}{m^*}\sum_{j=l}^{m} z_j = \begin{cases} O\left(\frac{1}{m}\left(\frac{m}{n}\right)^{2(\delta-\phi)}\right) & \text{if } \delta > \phi \\ O\left(\frac{\ln m}{m}\right) & \text{if } \delta = \phi \\ O\left(\frac{1}{m}\left(\frac{l}{n}\right)^{2(\delta-\phi)}\right) & \text{if } \delta < \phi. \end{cases} \tag{A.8}$$

When $\delta > \phi$ or $\delta = \phi$, it can be verified that these bounds are $o(1)$ using (11); when $\delta < \phi$, it is possible to verify that the bound is $o(1)$ using (12).

Next, notice that $|I_{\xi s}(\lambda_j)|^2 = I_{\xi\xi}(\lambda_j) I_{ss}(\lambda_j)$, so $|v_j| \le \zeta_j^{1/2} z_j^{1/2}$. Hence,

$$\left|\frac{1}{m^*}\sum_{j=l}^{m} v_j\right| \le \frac{1}{m^*}\sum_{j=l}^{m}|v_j| \le \frac{1}{m^*}\sum_{j=l}^{m}\zeta_j^{1/2} z_j^{1/2} \le \left(\frac{1}{m^*}\sum_{j=l}^{m}\zeta_j\right)^{1/2}\left(\frac{1}{m^*}\sum_{j=l}^{m} z_j\right)^{1/2}$$

by the Schwarz inequality. By (8), $\left(\frac{1}{m^*}\sum_{j=l}^{m}\zeta_j\right)^{1/2} = O_p(1)$, whereas $\left(\frac{1}{m^*}\sum_{j=l}^{m} z_j\right) = o(1)$ using the bounds in (A.8) and applying the conditions (11) and, when $\delta < \phi$, (12), so $\left|\frac{1}{m^*}\sum_{j=l}^{m} v_j\right| = o_p(1)$. Therefore, (A.7) holds.

To prove (A.3), let

$$r = c_\varphi n^\varphi \quad \text{for some } c_\varphi \in (0, \infty) \quad \text{and} \quad \varphi = \frac{2(\phi - \delta)}{2(\phi - \delta) + 1}. \tag{A.9}$$

Since

$$E(\eta_j^*) = E(\zeta_j) + 2E\left|\mathrm{Re}(v_j)\right| + z_j \le 2E(\zeta_j) + 2z_j, \tag{A.10}$$

noticing that $E(\zeta_j) \le C$ for all $l \le j \le m$ by (9), (A.3) follows from (A.10) if $z_j \le C$ for all $l \le j \le m$. Assumption 2 implies that

$$z_j \le C\left(\frac{j}{n}\right)^{2(\delta-\phi)} j^{-1} \quad \text{for all } l \le j \le m,$$

so when $\delta \ge \phi$, (A.3) follows immediately. When $\phi > \delta$, for any $j \ge r$

$$C\left(\frac{j}{n}\right)^{2(\delta-\phi)} j^{-1} \le C n^{\varphi(2(\delta-\phi)-1)-2(\delta-\phi)} \le C,$$

because, for n large enough, $l > r$ when $\upsilon \ge \varphi$. Therefore, if either $\delta \ge \phi$ or $\upsilon \ge \varphi$, $\hat{\delta} \to_p \delta$ follows as in DGH.

It only remains to discuss the case in which $\phi > \delta$ and $\varphi > \upsilon$.

Given the inequality (12), there is ρ such that $0 < \rho < \kappa$ and

$$\phi < \delta + \frac{1}{2}\frac{\kappa - \rho}{1 - \upsilon}. \tag{A.11}$$

Let

$$\Theta = \{d : \delta - \rho/2 \le d \le \Delta_2\}.$$

If $\Delta_1 \geq \delta - \rho/2$, following DGH, consistency follows if for any ε there exists $\alpha > 0$ such that

$$P\left(\inf_{d\in[\Delta_1,\Delta_2],|d-\delta|\geq\varepsilon} U_n(d) - U_n(\delta) \geq \alpha\right) \to 1 \quad \text{as } n \to \infty,$$

where

$$U_n(d) = \ln\left(\frac{1}{m^*}\sum_{j=l}^{m}\left(\frac{j}{m}\right)^{2d} I_{xx}(\lambda_j)\right) - \frac{2d}{m^*}\sum_{j=l}^{m}\ln\left(\frac{j}{m}\right). \tag{A.12}$$

DGH showed that

$$U_n(d) - U_n(\delta) = \ln L_n(d) - \ln L_n(\delta) + 2d - 2\delta + o(1).$$

where

$$L_n(d) = \frac{1}{m^*}\sum_{j=l}^{m}\left(\frac{j}{m}\right)^{2(d-\delta)}\eta_j^*,$$

When $l = 1$, and this also holds for any l meeting our assumptions. From Lemma A.1, as $n \to \infty$, uniformly in $d \in \Theta$,

$$U_n(d) - U_n(\delta) = \ln L(d) - \ln L(\delta) + 2d - 2\delta + o_p(1),$$

where

$$L(d) = \int_0^1 x^{2(d-\delta)}\mathrm{d}x = (1 + 2(d-\delta))^{-1},$$

and, since $-\ln(1 + x) + x > 0$ for $x > -1$ and $x \neq 0$, for $|d - \delta| \geq \varepsilon$,

$$-\ln(1 + 2(d-\delta)) + 2d - 2\delta \geq \alpha_\varepsilon > 0.$$

We therefore proved that, if $\Delta_1 \geq \delta - \rho/2$, $\hat{\delta} \to_p \delta$.

Otherwise, consider

$$\hat{\delta}_1 = \arg\min_{d\in[\Delta_1,\delta-\rho/2]} R(d), \widehat{\delta}_0 = \arg\min_{d\in[\delta-\rho/2,\Delta_2]} R(d).$$

Because by Lemma A.2 $U_n(d) - U_n(\delta)$ is a convex function, then

$$\hat{\delta} = \hat{\delta}_1 1(\hat{\delta}_0 = \delta - \rho/2) + \hat{\delta}_0 1(\hat{\delta}_0 > \delta - \rho/2),$$

where 1(.) is the indicator function. Notice now that $\hat{\delta}_0 \to_p \delta$, because we have already shown that $\hat{\delta} \to_p \delta$, if $\Delta_1 \geq \delta - \rho/2$. Therefore, $1(\hat{\delta}_0 = \delta - \rho/2) = o_p(1)$, and then $\hat{\delta}_1 1(\hat{\delta}_0 = \delta - \rho/2) = o_p(1)$ because $\hat{\delta}_1 \in [\Delta_1, \delta - \rho/2]$, so $\hat{\delta}_1 = O_p(1)$. On the other hand because $\hat{\delta}_0 \to_p \delta > \delta - \rho/2$; then, $\hat{\delta}_0 1(\hat{\delta}_0 > \delta - \rho/2) \to_p \delta$. Therefore, $\hat{\delta} \to_p \delta$ by Slutsky's theorem. □

Proof of Theorem 3. Proceeding as in DGH, we let

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

where

$$T_n(d) = 2\frac{1}{m^*}\sum_{j=l}^{m}\left(\frac{j}{m}\right)^{2(d-\delta)} v_j\eta_j^*, V_n(d) = \frac{1}{m^*}\sum_{j=l}^{m}\left(\frac{j}{m}\right)^{2(d-\delta)}\eta_j^*$$

for v_j defined in (16).

For $\rho > 0$ and such that (A.11) holds, take $\varepsilon < \min(\rho/2, \Delta_2 - \delta, \delta - \Delta_1)$, and rewrite

$$\begin{aligned}\hat{\delta} - \delta &= (\hat{\delta} - \delta)1(|\hat{\delta} - \delta| \leq \varepsilon) + (\hat{\delta} - \delta)1(|\hat{\delta} - \delta| > \varepsilon) \\ &= (\hat{\delta} - \delta)1(|\hat{\delta} - \delta| \leq \varepsilon) + (\hat{\delta} - \delta)o_p(1)\end{aligned} \tag{A.13}$$

where (A.13) follows by Theorem 2.

When $|\hat{\delta} - \delta| \leq \varepsilon$, then $V_n(\hat{\delta}) \geq \frac{1}{m^*}\sum_{j=l}^{m}\left(\frac{j}{m}\right)^{2\varepsilon}\eta_j^*$. For $\phi \leq \delta$ or $\varphi \leq \upsilon$, where φ is defined in (A.9),

$$\frac{1}{m^*}\sum_{j=l}^{m}\left(\frac{j}{m}\right)^{2\varepsilon}\eta_j^* \to_p \int_0^1 x^{2\varepsilon}\mathrm{d}x > 0 \tag{A.14}$$

by Lem 2 of DGH as in Eqn (81) of DGH. We now prove the convergence in (A.14) also for $\phi > \delta$ and $\varphi > \upsilon$. This follows by Lemma A.1, because the uniform convergence in probability for $d \in \Theta$ also implies pointwise convergence in probability for any d, as long as

$d \in \Theta$. Take then $d = \delta + \varepsilon$: the convergence in (A.14) follows provided that the hypotheses of Lemma A.1 are met. Given $\phi > \delta$ and $\varphi > \upsilon$, (A.11), $\varepsilon < \min(\rho/2, \Delta_2 - \delta, \delta - \Delta_1)$ and $\Delta_1 < \delta < \Delta_2$, then $\delta + \varepsilon \in \Theta$ and the hypotheses of Lemma A.1 are met, so the convergence in (A.14) follows.

Given the assumptions on ε and that $\Delta_1 < \delta < \Delta_2$, then $\hat{\delta}$ is interior to $[\Delta_1, \Delta_2]$, so $\frac{\partial U_n(d)}{\partial d}|_{d=\hat{\delta}} = 0$ and $T_n(\hat{\delta}) = 0$, and by the mean value theorem,

$$T_n(\hat{\delta}) = T_n(\delta) + \frac{\partial T_n(d)}{\partial d}|_{\delta=\delta^*}(\hat{\delta} - \delta), \quad |\delta^* - \delta| \le |\hat{\delta} - \delta|.$$

DGH already showed that for constant s_t, $l = 1$,

$$\frac{\partial T_n(d)}{\partial d}|_{\delta=\delta^*} \to_p 4, \tag{A.15}$$

and this can be generalized starting the summation in any l instead, provided that (11) holds. We now show (A.15) when s_t meets Assumption 2 instead. For this purpose, we replace (A.6) in the formula of $\frac{\partial T_n(d)}{\partial d}|_{\delta=\delta^*}$ and we show that the additional terms owing to $I_{ss}(\lambda_j)$ and $I_{\xi s}(\lambda_j)$ are negligible in probability. To prove this, we first notice that, bounding

$$|v_j| \le C \ln m, \tag{A.16}$$

$$\left| 4 \frac{1}{m^*} \sum_{j=l}^{m} v_j \left(\frac{j}{m}\right)^{2(\delta^*-\delta)} \ln\left(\frac{j}{m}\right) z_j \right| \le C \ln^2 m \frac{1}{m} \sum_{j=l}^{m} \left(\frac{j}{m}\right)^{2(\delta^*-\delta)} z_j$$
$$\le C \ln^2 m \frac{1}{m} \sum_{j=l}^{m} \left(\frac{j}{m}\right)^{-2\varepsilon} z_j \le C \ln^2 m \frac{m^{2\varepsilon}}{m} n^{2(\phi-\delta)} l^{-2\varepsilon+2(\delta-\phi)} = o(1) \tag{A.17}$$

recalling (A.11) and $\rho > 2\varepsilon$. Next, bounding again $|v_j| \le \zeta_j^{1/2} z_j^{1/2}$,

$$\left| \frac{1}{m^*} \sum_{j=l}^{m} v_j \left(\frac{j}{m}\right)^{2(\delta^*-\delta)} \ln\left(\frac{j}{m}\right) v_j \right| \le C \frac{\ln^2 m}{m} \sum_{j=l}^{m} \left(\frac{j}{m}\right)^{-2\varepsilon} |v_j|$$
$$\le C \left(\frac{\ln^2 m}{m} \sum_{j=l}^{m} \left(\frac{j}{m}\right)^{-2\varepsilon} \zeta_j \right)^{1/2} \left(\frac{\ln^2 m}{m} \sum_{j=l}^{m} \left(\frac{j}{m}\right)^{-2\varepsilon} z_j \right)^{1/2}. \tag{A.18}$$

Using the bound in Thm 2 of Robinson (1995a), $E(\zeta_j) \le C$, so $\frac{\ln^2 m}{m} \sum_{j=l}^{m} \left(\frac{j}{m}\right)^{-2\varepsilon} \zeta_j = O_p(\ln^2 m)$, whereas, because of (A.17), $\frac{\ln^2 m}{m} \sum_{j=l}^{m} \left(\frac{j}{m}\right)^{-2\varepsilon} z_j = O(\ln^2 m \frac{m^{2\varepsilon}}{m} n^{2(\phi-\delta)} l^{-2\varepsilon+2(\delta-\phi)})$. Therefore, by (A.11) and $\rho > 2\varepsilon$, the bound in (A.18) is $o_p(1)$.

We now show that

$$\sqrt{m} T_n(\delta) \to_d \mathrm{N}(0, 4). \tag{A.19}$$

This was already showed by Robinson (1995b) for constant s_t, $l = 1$, and, given that $\frac{\partial T_n(d)}{\partial d}|_{\delta=\delta^*} = 4 + o_p(1)$, the result (A.19) also holds when $l > 1$ in view of results in Theorem 2 of Robinson (1995b); also see part 3 of Theorem A.1 of Hurvich *et al.* (2005). To verify that a time-varying deterministic component does not affect (A.19) provided that Assumption 4′ is met, we verify that

$$\sqrt{m} \frac{1}{m^*} \sum_{j=l}^{m} (v_j (v_j + v_j^*)) \to_p 0, \tag{A.20}$$

$$\sqrt{m} \frac{1}{m^*} \sum_{j=l}^{m} (v_j z_j) \to 0. \tag{A.21}$$

By the triangle inequality, (A.2) and (A.16),

$$\left| \sqrt{m} \frac{1}{m^*} \sum_{j=l}^{m} (v_j z_j) \right| \le C \frac{\sqrt{m}}{m} \sum_{j=l}^{m} (|v_j| z_j) \le C \frac{\sqrt{m}}{m} \ln m \sum_{j=l}^{m} z_j. \tag{A.22}$$

By (A.8), when $\delta > \phi$ (A.22) is $O(\frac{\sqrt{m}}{m} (\frac{m}{n})^{2(\delta-\phi)} \ln m)$, when $\delta = \phi$ it is $O(\frac{\sqrt{m}}{m} \ln^2 m)$ and when $\delta < \phi$ it is $O(\frac{\sqrt{m}}{m} (\frac{l}{n})^{2(\delta-\phi)} \ln m)$. The bounds when $\delta \ge \phi$ are $o(1)$ in view of (13), and the bound when $\delta < \phi$ is $o(1)$ also using condition (14), so (A.21) holds.

Next, notice that

$$E\left(\frac{\sqrt{m}}{m^*} \sum_{j=l}^{m} (v_j v_j) \right)^2 \tag{A.23}$$

$$= \frac{m}{(m^*)^2} \sum_{j=l}^{m} v_j^2 G^{-1} \lambda_j^{2\delta} w_s(\lambda_j) E(G^{-1} \lambda_j^{2\delta} w_\xi(-\lambda_j) w_\xi(\lambda_j)) w_s(-\lambda_j) \tag{A.24}$$

$$+ \frac{m}{(m^*)^2} \sum_{j=l}^{m} \sum_{k=l,,k\neq j}^{m} v_j v_k G^{-1} \lambda_j^{\delta} w_s(\lambda_j) E(G^{-1} \lambda_j^{\delta} w_\xi(-\lambda_j) \lambda_k^{\delta} w_\xi(\lambda_k)) \lambda_k^{\delta} w_s(-\lambda_k). \tag{A.25}$$

In view of Assumption 2, (A.16) and Thm 2 of Robinson (1995a), the term in (A.24) is

$$O\left(\frac{\ln^2 m}{m} \sum_{j=l}^{m} \lambda_j^{2\delta} w_s(\lambda_j) w_s(-\lambda_j)\right),$$

which is $o_p(1)$ by the same arguments used in the discussion of the order of magnitude of (A.22).

Next, since $w_s(\lambda_j) = O(\sqrt{w_s(\lambda_j) w_s(-\lambda_j)}) = O((j/n)^{-\phi} j^{-1/2})$, using Thm 2 of Robinson (1995a), (A.25) is

$$\begin{aligned}
&O\left(\frac{1}{m} \ln^2 m \sum_{j=l}^{m} \sum_{k=l}^{j-1} (j/n)^{\delta-\phi} j^{-1/2} (k^{-1} \ln j)(k/n)^{\delta-\phi} k^{-1/2}\right) \\
&= O\left(\frac{1}{m} \left(\frac{1}{n}\right)^{2(\delta-\phi)} \ln^3 m \sum_{j=l}^{m} \sum_{k=l}^{j-1} j^{\delta-\phi} j^{-1/2} k^{-1} k^{\delta-\phi} k^{-1/2}\right) \\
&= \begin{cases}
O(\frac{1}{m} (\frac{l}{n})^{2(\delta-\phi)} \ln^3 m) & \text{if } \delta < \phi - 1/2 \\
O(\frac{n}{ml} \ln^4 m) & \text{if } \delta = \phi - 1/2 \\
O(\frac{1}{m} (\frac{1}{n})^{2(\delta-\phi)} \ln^3 m \; l^{\delta-\phi-1/2} \; m^{\delta-\phi+1/2}) & \text{if } \phi - 1/2 < \delta < \phi + 1/2 \\
O((\frac{1}{n}) \ln^4 m) & \text{if } \delta = \phi + 1/2 \\
O(\frac{1}{m} (\frac{m}{n})^{2(\delta-\phi)} \ln^3 m) & \text{if } \delta > \phi + 1/2.
\end{cases}
\end{aligned}$$

Notice that (13) and (14) imply that

$$\frac{1}{\sqrt{m}} \left(\frac{l}{n}\right)^{2(\delta-\phi)} = o(1).$$

We can therefore rewrite the bound for the case $\delta < \phi - 1/2$ as $O(\frac{1}{\sqrt{m}} (\frac{l}{n})^{2(\delta-\phi)} \frac{\ln^3 m}{\sqrt{m}}) = o(1)$, and, when $\delta = \phi - 1/2$, noticing that $\frac{1}{\sqrt{m}} (\frac{l}{n})^{2(\delta-\phi)} = \frac{n}{\sqrt{m}\, l}$, as $O(\frac{n}{\sqrt{m}\, l} \frac{\ln^4 m}{\sqrt{m}}) = o(1)$. For $\phi - 1/2 < \delta \leq \phi$, we rewrite the bound as $O(\frac{1}{\sqrt{m}} (\frac{l}{n})^{2(\delta-\phi)} \; (\frac{m}{l})^{\delta-\phi} \; \frac{\ln^3 m}{\sqrt{l}}) = o(1)$, while for $\phi \leq \delta < \phi + 1/2$ it is $O((\frac{m}{n})^{(\delta-\phi)} \; (\frac{l}{n})^{(\delta-\phi)} \; \frac{\ln^3 m}{\sqrt{l}\sqrt{m}}) = o(1)$. Finally, it is immediate to verify that the bound is $o(1)$ when $\delta \geq \phi + 1/2$. Therefore, (A.23) is $o(1)$, which implies (A.20). □

Combining these results with (A.13), (15) follows.

A.2. Technical lemmas

Lemma A.1. *Under the assumptions of Theorem 2 and $\phi > \delta$, $\varphi > \upsilon$, where φ is defined in* (A.9), *m^* defined in* (A.1),

$$\Theta = \{d : \delta - \rho/2 \leq d \leq \Delta_2\},$$

where $\rho > 0$ is such that (A.11) *holds, then*

$$\sup_{d \in \Theta} \left| \frac{1}{m^*} \sum_{j=l}^{m} \left(\frac{j}{m}\right)^{2(d-\delta)} \frac{I_{xx}(\lambda_j)}{G \lambda_j^{-2\delta}} - \int_0^1 x^{2(d-\delta)} \mathrm{d}x \right| \to_p 0. \tag{A.26}$$

Proof. We first verify that $r/m \to 0$ as $n \to \infty$, where r is defined in (A.9). From (12),

$$\kappa \geq (1 - \upsilon) 2(\phi - \delta),$$

so a sufficient condition for $\kappa > \varphi$ to hold is

$$(1 - \upsilon) 2(\phi - \delta) > \frac{2(\phi - \delta)}{2(\phi - \delta) + 1}$$

and, solving for υ,

$$\upsilon < \frac{2(\phi - \delta)}{2(\phi - \delta) + 1}.$$

Because $E(\eta_j^*) \le C$ for $r \le j \le m$ (see the discussion following eqn A.10) and (A.4), from the proof of Lem 2 of DGH,

$$\sup_{d\in\Theta}\left|\frac{1}{m^*}\sum_{j=r}^{m}\left(\frac{j}{m}\right)^{2(d-\delta)}\eta_j^* - \int_0^1 x^{2(d-\delta)}\mathrm{d}x\right| \to_p 0.$$

We now show that the contribution of the summation over the frequencies $l \le j < r$ is asymptotically negligible uniformly in $d \in \Theta$. For this purpose, letting

$$J_m(d) = \frac{1}{m^*}\sum_{j=l}^{r-1}\left(\frac{j}{m}\right)^{2(d-\delta)}\eta_j^*,$$

we obtain

$$\begin{aligned}E\sup_{d\in\Theta}|J_m(d)| &= E\left|\frac{1}{m^*}\sum_{j=l}^{r-1}\left(\frac{j}{m}\right)^{2((\delta-\rho/2)-\delta)}\eta_j^*\right| \\ &\le E\left(\frac{2}{m^*}\sum_{j=l}^{r-1}\left(\frac{j}{m}\right)^{-\rho}\zeta_j\right) + \frac{2}{m^*}\sum_{j=l}^{r-1}\left(\frac{j}{m}\right)^{-\rho} z_j\end{aligned}$$

for ζ_j defined in (7) and υ_j and z_j defined in (A.5), and the inequality follows from (A.10). Notice that $E(\frac{1}{m^*}\sum_{j=l}^{r-1}(\frac{j}{m})^{-\rho}\zeta_j) = O_p((\frac{r}{m})^{1-\rho}) = o_p(1)$. The second term can be bounded by

$$\begin{aligned}&C\frac{1}{m}\left(\frac{1}{m}\right)^{-\rho}\left(\frac{1}{n}\right)^{2(\delta-\phi)}\sum_{j=l}^{r} j^{-\rho+2(\delta-\phi)-1} \\ &\le Cm^{-1}m^{\rho}n^{2(\phi-\delta)}l^{-\rho+2(\delta-\phi)} \le Cn^{-\kappa+\rho\kappa}n^{2(\phi-\delta)}n^{2\upsilon(-\rho/2+(\delta-\phi))},\end{aligned}$$

which is negligible as $n \to \infty$ provided that

$$-\kappa + \rho\kappa + 2(\phi-\delta) + 2\upsilon(-\rho/2 + (\delta-\phi)) < 0,$$

and rearranging terms,

$$\kappa \ge 2(1-\upsilon)(\phi-\delta) + \rho(\kappa-\upsilon). \tag{A.27}$$

Since, from (A.11), $\kappa > 2(1-\upsilon)(\phi-\delta)+\rho$, (A.27) holds if $\rho(\kappa-\upsilon) \le \rho$. Since $\rho > 0$, this inequality is simplified to $(\kappa-\upsilon) \le 1$, which is met given (11). Combining these results, we obtain

$$E\sup_{d\in\Theta}|J_m(d)| \to 0,$$

so (A.26) follows. □

Lemma A.2. *Under the assumptions of Theorem* 2, *for* $U_n(d)$ *defined in* (A.12), $U_n(d)-U_n(\delta)$ *is a convex function.*

Proof. Let

$$a_j(d) = \left(\frac{j}{m}\right)^{2(d-\delta)}\eta_j^* \quad \text{and} \quad A_n(d) = \ln\left(\frac{1}{m^*}\sum_{j=l}^{m}a_j(d)\right)$$

so that $U_n(d) - U_n(\delta) = A_n(d) - \frac{2d}{m^*}\sum_{j=l}^{m}\ln(\frac{j}{m}) - A_n(\delta) + \frac{2\delta}{m^*}\sum_{j=l}^{m}\ln(\frac{j}{m})$. The sign of $\frac{\partial^2 A_n(d)}{\partial d^2}$ then is equal to the sign of

$$\left(\sum_{j=l}^{m}\ln^2\left(\frac{j}{m}\right)a_j(d)\right)\left(\sum_{k=l}^{m}a_k(d)\right) - \left(\sum_{j=l}^{m}\ln\left(\frac{j}{m}\right)a_j(d)\right)^2. \tag{A.28}$$

Letting $x_k = \sqrt{a_k(d)}$, $y_j = \ln(\frac{j}{m})\sqrt{a_j(d)}$, (A.28) is $(\sum_{j=l}^{m}y_j^2)(\sum_{k=l}^{m}x_k^2) - (\sum_{j=l}^{m}x_jy_j)^2$, which is non-negative by the Schwarz inequality. Moreover, because x_k and y_j are not linearly dependent, the expression in (A.28) is negative. The function $A_n(d)$ is then convex. Therefore, $U_n(d)$ and $U_n(d)-U_n(\delta)$ are also convex functions. □

Acknowledgements

This article is derived from the author's PhD dissertation at London School of Economics. Financial support from the Economic and Social Research Council through grant R000239936 and from the Dennis Sargan Memorial Fund is gratefully acknowledged. The aurthor is very grateful to Peter M. Robinson, to Peter C.B. Phillips and to Stepana Lazarova for many helpful discussions and suggestions. He also thanks Giovanni Urga, Myunghwan Seo, Andy R. Tremayne, Valentina Corradi, Roderick J. McCrorie, Patrick Marsh for many comments that improved the presentation.

REFERENCES

Abadir, K. M., Distaso, W. and Giraitis, L. (2007) Nonstationarity-extended local Whittle estimation. *Journal of Econometrics* **141**, 1353–84.

Bhattacharya, R. N., Gupta, V. K. and Waymire, E. (1983) The Hurst effect under trends. *Journal of Applied Probability* **20**, 649–62.

Dalla, V., Giraitis, L. and Hidalgo, J. (2006) Consistent estimation of the memory parameter for nonlinear time series. *Journal of Time Series Analysis* **27**, 211–51.

Davies, R. B. and Harte, D. S. (1987) Tests for Hurst effect. *Biometrika* **74**, 95–101.

Deo, R. S. and 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.

Diebold, F. X. and Inoue, A. (2001) Long memory and structural change. *Journal of Econometrics* **105**, 131–59.

Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis* **4**, 221–38.

Giraitis, L. and Robinson, P. M. (2003) Edgeworth expansions for semiparametric Whittle estimation of long memory. *Annals of Statistics* **31**, 1325–75.

Giraitis, L., Kokoszka, P. and Leipus, R. (2001) Testing for long memory in the presence of a general trend. *Journal of Applied Probability* **38**, 1033–54.

Granger, C. W. J. and Hyung, N. (2004) Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance* **11**, 399–421.

Heyde, C. C. and Dai, W. (1996) On the robustness to small trends of estimation based on the smoothed periodogram. *Journal of Time Series Analysis* **17**, 141–50.

Hurvich, C., Lang, G. and Soulier, P. (2005) Estimation of long memory in the presence of a smooth nonparametric trend. *Journal of the American Statistical Association* **100**, 853–71.

Künsch, H. R. (1986). Discrimination between monotonic trends and long range dependence. *Journal of Applied Probability* **23**, 1025–30.

Lobato, I. N. and Savin, N. E. (1998). Real and spurious long memory properties of the stock market data. *Journal of Business and Economic Statistics* **16**, 261–83.

Mikosch, T. and Stărică, C. (1999) Change of structure in financial time series, long range dependence and the GARCH model. Working Paper, Department of Statistics, University of Pennsylvania.

Nelson, C. M. and Plosser, C. I. (1982) Trends and random walk in macroeconomic time series. *Journal of Monetary Economics* **10**, 139–62.

Perron, P. (1991). The great crash, the oil price shock and the unit root hypothesis. *Econometrica*, **57**, 1361–401.

Phillips, P. C. B. and Shimotsu, K. (2004) Local Whittle estimation in nonstationary and unit root cases. *Annals of Statistics* **32**, 656–92.

Robinson, P. M. (1995a). Log-periodogram regression of time series with long range dependence. *Annals of Statistics* **23**, 1048–72.

Robinson, P. M. (1995b). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* **23**, 1630–61.

Robinson, P. M. (1997). Large sample inference for nonparametric regression with dependent errors. *Annals of Statistics* **25**, 2054–83.

Robinson, P. M. and Marinucci, D. (2000) The averaged periodogram for nonstationary vector stochastic processes. *Statistical Inference for Stochastic Processes* **3**, 149–60.

Robinson, P. M. and Marinucci, D. (2001) Narrow band analysis of nonstationary processes. *Annals of Statistics* **29**, 947–86.

Robinson, P. M. and Marinucci, D. (2003) Semiparametric frequency domain analysis of fractional cointegration. In *Time series with Long Memory* (ed. P. M. Robinson). Oxford: Oxford University press, pp. 334–73.

Teverovsky, V. and Taqqu, M. (1997) Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimate. *Journal of Time Series Analysis* **18**, 279–304.