JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **46**: 647–673 (2025)
Published online 1 October 2024 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12782

ORIGINAL ARTICLE

LOCAL WHITTLE ESTIMATION IN TIME-VARYING LONG MEMORY SERIES

JOSU ARTECHE[a*] AND LUIS F. MARTINS[b,c]

[a] *UPV/EHU, University of the Basque Country, Bilbao, Spain*
[b] *Business Research Unit (BRU-IUL), Instituto Universitário de Lisboa (ISCTE-IUL), Lisbon, Portugal*
[c] *CIMS, University of Surrey, Surrey, UK*

The memory parameter is usually assumed to be constant in traditional long memory time series. We relax this restriction by considering the memory a time-varying function that depends on a finite number of parameters. A time-varying Local Whittle estimator of these parameters, and hence of the memory function, is proposed. Its consistency and asymptotic normality are shown for locally stationary and locally non-stationary long memory processes, where the spectral behaviour is restricted only at frequencies close to the origin. Its good finite sample performance is shown in a Monte Carlo exercise and in two empirical applications, highlighting its benefits over the fully parametric Whittle estimator proposed by Palma and Olea (2010). Standard inference techniques for the constancy of the memory are also proposed based on this estimator.

Received 11 November 2022; Accepted 4 September 2024

Keywords: Long memory; local Whittle estimation; local stationarity; time-varying parameter.

JEL. C13; C22.

1. INTRODUCTION

The existence of strong persistence in time series has been well documented in various fields of research. See, for example, Palma (2007). It implies that distant observations are significantly correlated, so that the effects of shocks die out very slowly. Detecting such behaviour requires a relatively large number of observations, usually over a long period of time. The models traditionally used for highly dependent series assume that the strength of the persistence, measured by the memory parameter d, is constant over time, which may be difficult to accept for such long time series. For example, the persistence of shocks in economic series may be different in expansions and recessions, implying that the memory parameter changes with the economic situation that exists at each point in time. Indeed, a number of recent papers have found evidence of a persistence that changes over time at some breakpoints (Ray and Tsay, 2002; Song and Bondon, 2012; Martins and Rodrigues, 2014) or in a regime switching long memory context (Haldrup and Nielsen, 2006; Boutahar *et al.*, 2008; Boubaker, 2018). However, these types of models assume an abrupt change in persistence, whereas the behaviour of many time series suggests that a smoothly evolving persistence may be more appropriate than sudden changes. This has led some authors to extend the concept of locally stationary models to accommodate a smoothly varying persistence (Beran, 2009; Palma and Olea, 2010; Roueff and von Sachs, 2011; Bisaglia and Grigoletto, 2021). In these models d is no longer constant, but a smoothly varying function $d(u)$ for $u = t/T$ in a time series from $t = 1$ to $t = T$. These models require estimation of the memory function $d(u)$ to assess the persistence of the series, which can be done in the time domain (Beran, 2009; Boubaker, 2018; Bisaglia and Grigoletto, 2021), in the frequency domain (Palma and Olea, 2010; Wang, 2019; Chan and Palma, 2020) or using log-regression of a series of wavelets (Jensen and Whitcher, 2000; Roueff and von Sachs, 2011). The rescaling $u = t/T$ allows the construction of an asymptotic

*Correspondence to: Josu Arteche, UPV/EHU, University of the Basque Country, Barrio Sarriena s/n, 48940 Leioa, Bizkaia, Bilbao, Spain.
E-mail: josu.arteche@ehu.eus

theory, since increasing the sample size means having more and more observations to identify $d(u)$ in a finer grid, but in the same interval.

Despite the fact that many of these estimators have good asymptotic properties (consistency, asymptotic normality and even efficiency), parametric estimators of locally stationary long memory series suffer from two main drawbacks: first, they are restricted to locally stationary processes, meaning that $d(u) < 1/2$ for every $u \in [0, 1]$, which may be difficult to sustain in many situations. Second, the usual problems of misspecification arise, so that if the parametric model is far from the true data generating process, the estimator may be inconsistent and subject to a large bias. We extend the parametric Whittle estimator proposed by Palma and Olea (2010) in two directions to overcome these two problems: first, we extend the range of possible values of $d(u)$ up to $d(u) < 1$ (for consistency) or $d(u) < 3/4$ (for asymptotic normality). Second, our estimator is based only on a local (in the frequency direction) knowledge of the tine varying spectral density function around the origin, with no need to specify spectral behaviour far from those frequencies, thereby avoiding the effects of misspecification. As usual, this robustness to misspecification comes at the price of lower efficiency when compared with parametric techniques. An additional benefit of our local estimator over the fully parametric Whittle estimator in Palma and Olea (2010) is that Gaussianity is not required, with only linearity being imposed in terms of martingale difference innovations with some finite moments.

Other semi-parametric estimators of $d(u)$ have been proposed by Roueff and von Sachs (2011) and Wang (2019), which similarly avoid the need of a fully and correctly specified data generating process. Their asymptotic properties were derived for a fixed scalar u and no joint asymptotic distribution for a set of u's was offered, which precludes the implementation of valid strategies to test for the constancy of the memory parameter. An additional benefit of our proposal is that the hypothesis of constant d can be easily tested using standard inference techniques.

The rest of the article is structured as follows. Section 2 describes the time-varying long memory (TVLM) processes in a local context such that the spectral behaviour is restricted only around frequencies close to zero. Section 3 introduces the Time-Varying Local Whittle (TVLW) estimator, showing the identifiability of the parameters to be estimated, its consistency and asymptotic normality. Inference is also discussed, focusing on tests of the hypothesis of constant memory. Section 4 analyses its finite sample performance with a Monte Carlo exercise, paying particular attention to testing the constancy of the memory function. Finally Section 5 applies it to some real time series. All the technical details, with proofs of lemmas and theorems, are placed in the Appendix A.

2. TIME-VARYING LONG MEMORY

The triangular array $x_{t,T}$, $t = 1, \dots T$ is a TVLM process if its time-varying spectral density (pseudo spectral density in the locally non-stationary case) function satisfies

$$f_x(u, \lambda) = C(u)\lambda^{-2d_P(u,\xi^0)}(1 + o(1)) \tag{1}$$

as $\lambda \to 0^+$ for $0 < \inf_u C(u) \leq \sup_u C(u) < \infty$ and $d_P(u, \xi^0)$ a function of $u = t/T$ and a finite set of parameters ξ^0 defining the memory at every t. Hereafter the superscript 0 denotes true parameters. We consider TVLM processes satisfying the following assumptions:

A.1 Let $u = t/T$ and denote $v_{t,T} = \sum_{j=0}^{\infty} b_j(u)\varepsilon_{t-j}$ such that $\sup_{u\in[0,1]} \sum_{j=0}^{\infty} b_j^2(u) < \infty$ and $E(\varepsilon_t|F_{t-1}) = 0$, $E(\varepsilon_t^2|F_{t-1}) = 1$ a.s. for $t = 0, \pm 1, \pm 2, \dots$, where F_{t-1} is the σ-field of events generated by ε_s, $s \leq t-1$. If $-1/2 < d_P(u, \xi^0) < 1/2$ then $x_{t,T} = x_0 + v_{t,T}$ and for $1/2 \leq d_P(u, \xi^0) < 1$ then $x_{t,T} = x_0 + \sum_{s=1}^{t} v_{s,T}$ where x_0 is a random variable not depending on t. For $B(u, \lambda) = \sum_{j=0}^{\infty} b_j(u) \exp(-i\lambda j)$ then

$$|B(u, \lambda)| = O(\lambda^{-d_P^v(u,\xi^0)}), \left|\frac{\partial B(u, \lambda)}{\partial u}\right| = O\left(\lambda^{-d_P^v(u,\xi^0)} \log(\lambda) \frac{\partial d_P^v(u, \xi^0)}{\partial u}\right)$$

for all $u \in (0, 1]$.

J. Time Ser. Anal. **46**: 647–673 (2025)
DOI: 10.1111/jtsa.12782

A.2 The time-varying spectral density of $v_{t,T}$ satisfies

$$f_v(u, \lambda) = C(u)\lambda^{-2d_P^v(u,\xi^0)}(1 + o(1))$$

as $\lambda \to 0^+$ for $0 < C(u) < \infty$ for all $u \in [0, 1]$ and $-1/2 < d_P^v(u, \xi^0) = d_P(u, \xi^0) < 1/2$ in the locally stationary case and $-1/2 \leq d_P^v(u, \xi^0) = d_P(u, \xi^0) - 1 < 0$ for a locally non-stationary $x_{t,T}$.

Assumption A.1 avoids the restriction of Gaussianity and only imposes linearity and local stationarity of $v_{t,T}$ with bounded second moments of the innovations. Under this condition the spectral representation of $v_{t,T}$ is

$$v_{t,T} = \int_{-\pi}^{\pi} B\left(\frac{t}{T}, \lambda\right) e^{i\lambda t} \mathrm{d}Z(\lambda), t = 1, 2, \ldots, T,$$

where $\mathrm{d}Z(\lambda)$ is the spectral representation of a centred weak white noise with unit variance. The time-varying spectral density function is then $f_v(u, \lambda) = |B(u, \lambda)|^2/2\pi$. Local non-stationarity is considered as in Velasco (1999a) extending the concept of Type I long memory. The time-varying pseudo spectral density function of $x_{t,T}$ is in this case $f_x(u, \lambda) = |1 - \exp(i\lambda)|^{-2} f_v(u, \lambda)$ such that in both the local stationary and non-stationary cases the spectral or pseudo spectral density function of $x_{t,T}$ satisfies (1) with $d_P(u, \xi^0) \in (-1/2, 1)$ for all $u \in [0, 1]$. Assumption A.2 precludes the possibility of seasonal or cyclical long memory, since no spectral poles are allowed at frequencies far from zero. The analysis could be extended to cover other types of long memory where the spectral density diverges at a positive frequency as in Arteche and Robinson (2000) but it is constrained here to the empirically more popular case of standard long memory at frequency zero. The function $C(u)$ includes the possibly time-varying short memory components and, although it may change with u, it is a finite and positive constant for a given u.

The Locally Stationary AutoRegressive Fractionally Integrated Moving Average model (LSARFIMA) in Palma and Olea (2010) is a natural example satisfying Assumptions A.1 and A.2. It is defined as

$$v_{t,T} - \mu = \sigma(u)(1 - L)^{-d_P^v(u,\xi^0)} \frac{\Phi(u, L)}{\Theta(u, L)} \varepsilon_t, \quad t = 1, \ldots, T, \tag{2}$$

for $u \in [0, 1]$, $\Phi(u, L)$, $\Theta(u, L)$ are autoregressive and moving average polynomials, $\sigma(u)$ is a scale factor, $\sup_u d_P^v(u, \xi^0) < 1/2$ and $\varepsilon_t \sim iid(0, 1)$. The condition $\sup_{u\in[0,1]} d_P^v(u, \xi^0) < 1/2$ entails finite variance and locally stationarity in the sense that they can be locally approximated by a stationary process. See Dahlhaus (1996a, 1996b, 1997) or Dahlhaus and Giraitis (1998) for more details on locally stationary processes. However, condition (1) covers more general cases, including other specifications of the short memory component not belonging to an ARMA context as well as locally non-stationary fractionally integrated ARMA models where $\sup_u d_P(u, \xi^0)$ can be larger than $1/2$ implying non-square summability and infinite variance. Assumption A.1 details how a locally non-stationary specification is achieved by extending the concept of Type I long memory as defined in Velasco (1999a) to allow for a time-varying memory parameter.

Different specifications for $d_P(u, \xi^0)$ are possible, including level shifts (Song and Bondon, 2012; Martins and Rodrigues, 2014) or more complicated functional forms (e.g. Palma and Olea, 2010; Roueff and von Sachs, 2011) capable of capturing a smoothly varying persistence of the series. In particular, we consider that the time-varying memory function $d_P(u, \xi^0)$ is a linear combination of a finite number of Chebyshev polynomials of the form

$$d_P(u, \xi^0) = \sum_{v=0}^{P} \xi_v^0 P_{v,T}(u), \tag{3}$$

where the $(P + 1)$ Chebyshev polynomials $P_{v,T}(t)$ are defined as

$$P_{0,T}(u) = 1, \quad P_{v,T}(u) = \sqrt{2}\cos\left(v\pi\left(u - \frac{0.5}{T}\right)\right), \quad v = 1, 2, 3, \ldots$$

We choose this basis of functions because they allow for smooth transitions and posses convenient mathematical properties. For example, they are orthonormal such that for all integers v, q, $\frac{1}{T}\sum_{t=1}^{T} P_{v,T}(t/T)P_{q,T}(t/T) = \mathbb{1}(v = q)$, where $\mathbb{1}(\cdot)$ is the indicator function. Exploiting this orthonormality property, Bierens and Martins (2010) show that a linear combination of Chebyshev polynomials is a good representation of any square integrable and differentiable real function (see their Lemma 1), so that $d_P(u, \xi^0)$ is a good approximation of a time varying d for large enough P.

3. TVLW ESTIMATION

Given the local (in the frequency dimension) behaviour of the pseudo-spectral density function in (1), the vector of parameters $\xi^0 = (\xi_0^0, \xi_1^0, \ldots, \xi_P^0)'$ can be estimated by minimising a local version of the parametric Whittle function in Dahlhaus (1997) for weakly dependent and Palma and Olea (2010) for long memory locally stationary processes. The TVLW estimator is then defined as

$$(\widehat{C}', \widehat{\xi}_0, \ldots, \widehat{\xi}_P) = \arg\min L_T(C', \xi_0, \ldots, \xi_P),$$

where

$$L_T(C', \xi_0, \ldots, \xi_P) = \sum_{k=1}^{m}\sum_{j=1}^{M}\left\{\log C_j \lambda_k^{-2d_P(u_j,\xi)} + \frac{I_N(u_j, \lambda_k)}{C_j \lambda_k^{-2d_P(u_j,\xi)}}\right\},$$

for $C = (C_1, \ldots, C_M)'$, $C_j = C(u_j)$, $u_j = t_j/T$, $t_j = S(j-1) + N/2$ for $j = 1, \ldots, M$ portions (blocks), where S is the number of positions shifted between two consecutive portions of the full sample and N is the length of the portions in which the full sample is divided such that $T = S(M-1) + N$. Also $d_P(u_j, \xi) = \sum_{v=0}^{P} \xi_v P_{v,T}\left(u_j\right)$ and

$$I_N(u_j, \lambda_k) = |D_N(u_j, \lambda_k)|^2, \tag{4}$$

is the periodogram for each block of N observations at frequency $\lambda_k = 2\pi k/N$ with, for $u \in [0, 1]$

$$D_N(u, \lambda) = \frac{1}{\sqrt{2\pi N}}\sum_{s=0}^{N-1} x_{[uT]-N/2+s+1,T}\, e^{-i\lambda s}.$$

The objective function $L_T(C', \xi_0, \ldots, \xi_P)$ is based on the periodogram computed in M (possibly overlapping) blocks of N observations in a neighbourhood of each $u_j, j = 1, \ldots M$, spanning the interval $\{u_j - N/2, u_j + N/2\}$, with $N/T \to 0$ as $T \to \infty$. It differs from the parametric objective function in Palma and Olea (2010) in that only m frequencies close to the origin are considered. Including the low frequency behaviour of the M periodograms, $I_N(u_j, \lambda_k)$ for $j = 1, \ldots, M$ and $k = 1, \ldots, m$, allows us to detect the possible time-varying character of the memory function. For instance, if we split a time series of $T = 652$ observations into $M = 100$ blocks of length $N = 256$ each, shifting $S = 4$ positions forward each time, we compute the periodogram in the 100 overlapping blocks $(x_{1,652}, x_{2,652}, \ldots, x_{256,652}), (x_{5,652}, x_{7,652}, \ldots, x_{260,652}), \ldots, (x_{397,652}, x_{398,652}, \ldots, x_{652,652})$.

The different C_j are finite positive constants, and concentrating their estimators out of the objective function, the TVLW estimator of ξ^0 is defined as

$$\widehat{\xi} = (\widehat{\xi}_0, \ldots, \widehat{\xi}_P) = \arg\min_{\Theta} R_T(\xi),$$

where $R_T(\xi) = M^{-1}\sum_{j=1}^{M} R_{T,j}(\xi)$ for

$$R_{T,j}(\xi) = R_{T,j}(\xi_0, \dots, \xi_P) = \log \widehat{C}_j(\xi) - \frac{2}{m} d_P(u_j, \xi) \sum_{k=1}^{m} \log \lambda_k,$$

$$\widehat{C}_j(\xi) = \widehat{C}_j(\xi_0, \dots, \xi_P) = \frac{1}{m} \sum_{k=1}^{m} \frac{I_N(u_j, \lambda_k)}{\lambda_k^{-2d_P(u_j,\xi)}},$$

with $d_P(u_j, \xi) = \sum_{v=0}^{P} \xi_v P_{v,T}\left(u_j\right)$ and $\Theta = \{\xi : \Delta_1 \le \inf_j d_P(u_j, \xi) \le \sup_j d_P(u_j, \xi) \le \Delta_2; j = 1, 2, \dots, M\}$.

The following assumptions are sufficient to prove the consistency of $\hat{\xi}$ in Theorem 1:

A.3 $\xi^0 \in \Theta$ where $\Theta = \{\xi : \Delta_1 \le \inf_j d_P(u_j, \xi) \le \sup_j d_P(u_j, \xi) \le \Delta_2; j = 1, 2, \dots, M\}$ with $-1/2 < \Delta_1 < \Delta_2 < 1$ and $\Delta_2 - \Delta_1 < 1/2$.

A.4 In a neighbourhood $\lambda \in (0, \epsilon)$ of the origin,

$$\left| \frac{\partial}{\partial \lambda} f_v(u, \lambda) \right| = O\left(\lambda^{-1-2d_P^v(u,\xi^0)} \right),$$

as $\lambda \to 0^+$, for all $u \in (0, 1]$.

A.5 As $T \to \infty$

$$\frac{1}{M} + \frac{N}{SM} \log N \to 0.$$

A.6 As $T \to \infty$

$$\frac{1}{m} + \frac{m}{N} \to 0.$$

Assumption A.3 allows for locally stationary and non-stationary processes but restricts the value of the time-varying memory function to belong to an interval of width 0.5. This restriction is far less severe than other requirements imposed by other authors. For example, Palma and Olea (2010) require $0 < \inf_u d_P(u, \xi^0) < \sup_u d_P(u, \xi^0) < 1/2$. We relax this restriction by allowing for local non-stationarity (but mean reversion) and negative values of the memory parameter. Assumption A.5 imposes the condition that the number of blocks M should go to infinity with T but the shifting period S can diverge or remain fixed. It implies that $N \log N/T \to 0$. This condition permits the identifiability of the vector of parameters ξ as stated in the next lemma. Finally, Assumption A.6 imposes the typical limitation in the rate of divergence of the bandwidth in local Whittle estimation.

Lemma 1 (Identifiability)**.** Consider two $(P+1) \times 1$ vectors ξ^1 and ξ^2. Then, under A.5

- $\xi^1 = \xi^2$ implies $d_P(u_j, \xi^1) = d_P(u_j, \xi^2)$ for $j = 1, .., M$.
- If $d_P(u_j, \xi^1) = d_P(u_j, \xi^2)$ for $j = 1, .., M$, then $\|\xi^1 - \xi^2\|_2 \to 0$ as $T \to \infty$.

Theorem 1. Under Assumptions A.1–A.6, $\hat{\xi} \xrightarrow{p} \xi^0 = (\xi_0^0, \dots, \xi_P^0)$ as $T \to \infty$

The proofs of the lemmas and theorems are in the Appendix A. Theorem 1 is valid for overlapping ($S < N$) and non-overlapping ($S \ge N$) blocks. The next theorem shows the asymptotic normality for the case of non-overlapping blocks. When $S < N$, there are many non-negligible correlation terms between the different blocks in the score of the TVLW estimator, which makes the analytical analysis of their convergence almost untreatable and precludes obtaining the asymptotic distribution. Therefore, the analysis below focuses on the case $S \ge N$. The effects of overlapping blocks on the distribution of the TVLW estimator are analysed in the Monte Carlo of Section 4.

DOI: 10.1111/jtsa.12782
Journal of Time Series Analysis published by John Wiley & Sons Ltd.

The following assumptions are required to get the asymptotic distribution:

B.1 Assumption A.1 holds and $E[\epsilon_t^3|F_{t-1}] = \mu_3$, $E[\epsilon_t^4|F_{t-1}] = \mu_4$ a.s. for $t = 0, \pm 1, \pm 2, \dots$, and finite constants μ_3, μ_4.

B.2 The time-varying spectral density of $v_{t,T}$ satisfies for some $\beta \in [1, 2]$

$$f_v(u, \lambda) = C(u)\lambda^{-2d_P^v(u,\xi^0)}(1 + O(\lambda^\beta)),$$

as $\lambda \to 0$, where $C(u) \in (0, \infty)$ for all $u \in [0, 1]$ where $-1/2 < d_P^v(u, \xi^0) = d_P(u, \xi^0) < 1/2$ in the locally stationary case and $-1/2 \leq d_P^v(u, \xi^0) = d_P(u, \xi^0) - 1 < -1/4$ for a locally non-stationary $x_{t,T}$.

B.3 A.3 holds with $\Delta_2 < 3/4$.

B.4 In a neighbourhood $(0, \epsilon)$ of the origin, $B(u, \lambda)$ is differentiable with respect to λ and

$$\left|\frac{\partial B(u, \lambda)}{\partial \lambda}\right| = \left(\frac{|B(u, \lambda)|}{\lambda}\right),$$

as $\lambda \to 0^+$.

B.5 Assumptions A.5 (for identifiability) and A.6 (for consistency) hold and for any $\gamma > 0$, as $T \to \infty$

$$\frac{m^{1+2\beta}}{N^{2\beta}}(\log m)^2 + \frac{N\sqrt{m}}{SM}\log N + \frac{\log N}{m^\gamma} \to 0.$$

The last term in Assumption B.5 is similarly imposed in assumption 4' in Shimotsu (2007), who deals with Local Whittle estimation in a multi-variate context. Although our focus is univariate, the multiple parameter estimation needs to similarly strengthen the restrictions on m with the last term in Assumption B.5, which is not required in the univariate single-parameter estimation in Robinson (1995b) or Velasco (1999b). This assumption is satisfied for example if $N = S$, $m = O(N^a)$ with $0 < a < \beta/(0.5 + \beta)$ and $M = O(N^b)$ with $b > a/2$, with the $O()$ term here meaning exact proportionality.

Theorem 2. Under Assumptions B.1–B.5,

$$\sqrt{mM}(\hat{\xi} - \xi^0) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4}I_{P+1}\right) \text{ as } T \to \infty,$$

for I_{P+1} the identity matrix of order $P + 1$.

Remark 1. The asymptotic distribution in Theorem 2 is only valid for non-overlapping blocks $S \geq N$. When $S < N$ the correlation between different blocks would at least affect the variance in the asymptotic distribution. In this case some bootstrap approach could be used to approximate the distribution of the TVLW estimator, as for example the local bootstrap in Arteche and Orbe (2016). This procedure exploits the consistency of $\hat{\xi}$ for overlapping and non-overlapping blocks. It is based on resampling the locally Studentized periodogram defined as $I_{jk}\lambda_k^{2d_P(u_j,\hat{\xi})}$ for $I_{jk} = I_N(u_j, \lambda_k)$ in a neighbourhood of every frequency. The bootstrap replicates I_{jk}^*, $k = 1, \dots, m$ for every $j = 1, \dots, M$ are then obtained by resampling locally the studentized periodogram. With those bootstrap replicates the bootstrap objective function can be constructed to obtain the bootstrap estimate ξ^*. Repeating the procedure B times the distribution of $\xi - \xi^0$ can be approximated by the empirical distribution of $\hat{\xi}^* - \hat{\xi}$.

Remark 2. A Hessian based approximation of the variance of $\sqrt{mM}\hat{\xi}$ in the asymptotic distribution is A_{Mm}^{-1} where

$$A_{Mm} = \frac{4}{Mm}\sum_{k=1}^{m} v_k^2 \sum_{j=1}^{M} P_j P_j',$$

for $v_k = \log k - m^{-1}\sum_{l=1}^{m} \log l$ and $P_j = (P_{0,T}(u_j), P_{1,T}(u_j), \ldots, P_{P,T}(u_j))'$. We have found in the simulations that A_{Mm}^{-1} leads to a better approximation of the finite sample variance than the asymptotic expression $4^{-1}I_{P+1}$.

Remark 3. The first restriction in Assumption B.5 implies that $m = O(N^a)$ with a arbitrarily close to $2\beta/(1+2\beta)$. Consider the case $S = N$, which is recommended because it means that all the observations are used in the estimation and leads to higher efficiency. It implies that $T = MN$. Theorem 2 shows that the TVLW estimator $\hat{\xi}$ converges at a rate $O(\sqrt{mM}) = O(N^{a/2}M^{1/2}) = O(T^{a/2}M^{(1-a)/2})$. The LW estimator of a constant d in Robinson (1995b) achieves the rate of convergence $O(T^{a/2})$ with a bandwidth $O(T^a)$. Since $a < 2\beta/(1+2\beta) < 1$, the rate of convergence of the TVLW is faster than that of the LW, tending to equalise as β increases.

Remark 4. The result in Theorem 2 justifies the asymptotic approximation of the distribution of any linear combination of $\hat{\xi}$. For a non-stochastic $P+1$ vector $\iota = (\iota_0, \ldots, \iota_P)'$ the distribution of $\sqrt{mM}\iota'(\hat{\xi}-\xi^0)$ can be asymptotically approximated as $\mathcal{N}\left(0, 4^{-1}\sum_{v=0}^{P}\iota_v^2\right)$. For example, $d_P(u,\hat{\xi}) = \iota'\hat{\xi}$ for $\iota = (P_{0,T}(u), \ldots, P_{P,T}(u))'$ is an estimator of the memory parameter and the distribution of $\sqrt{mM}(d_P(u,\hat{\xi}) - d_P(u,\xi^0))$ can be asymptotically approximated as $\mathcal{N}\left(0, 4^{-1}\sum_{v=0}^{P}P_{v,T}(u)^2\right)$.

Remark 5. The pivotal character of the asymptotic distribution in Theorem 2 permits the implementation of standard inference techniques to test any linear hypothesis of the form $H_0 : R\xi = r$ for some $q\times(P+1)$ matrix R and $q \times 1$ vector r. The next corollary establishes the testing strategy discussing its validity and consistency.

Corollary 1. Consider the test statistic

$$W(R,r) = mM(R\hat{\xi} - r)'\left[RA_{Mm}^{-1}R'\right]^{-1}(R\hat{\xi} - r),$$

for the hypothesis $H_0 : R\xi = r$. If the assumptions in Theorem 2 are satisfied:

(a) Under H_0

$$W(R,r) \overset{d,H_0}{\to} \chi_q^2, \quad \text{as } T \to \infty.$$

(b) Under $H_1 : R\xi = r + \bar{r}$ for $\bar{r}$ a $q \times 1$ vector with constant element $\bar{r}_i \neq 0$ for some or all $i = 1, .., q$,

$$\lim_{T\to\infty} P(W(R,r) > \chi_{q|\alpha}^2) \to 1.$$

for any significance level $\alpha > 0$.

(c) Under $H_1 : R\xi = r + \bar{r}$ for $\bar{r} = (nM)^{-1/2}(\theta_1, \ldots, \theta_q)'$ with constant $\theta_i \neq 0$ for some or all $i = 1, .., q$,

$$W(R,r) \overset{d,H_1}{\to} \chi_q^2\left(4\sum_{i=1}^{q}\theta_i^2\right), \quad \text{as } T \to \infty.$$

Corollary 1(b) shows the consistency of the test based on the asymptotic distribution obtained in (a) and (c) gives the asymptotic distribution against local alternatives. The non-centrality parameter $4\sum_{i=1}^{q}\theta_i^2$ delimits the asymptotic power of the test and its efficiency. A hypothesis of interest is $H_0 : \xi_{i_1} = \ldots = \xi_{i_q} = 0$ against the alternative that one or some of the $q \le P + 1$ equalities are not satisfied, for $i_1, \ldots, i_q \in \{0, 1, \ldots, P\}$. In this case, R is a matrix of zeros except for the elements $[R]_{s,i_s+1} = 1$, $s = 1, \ldots, q$, and r is a vector of zeros. For example, the constancy of the memory parameter implies $q = P$ and $i_1, \ldots, i_q = 1, \ldots, P$, and this hypothesis can be easily tested to assess whether the series under consideration has a constant memory parameter (under H_0) or whether it is a TVLM process instead. A power analysis of this test is included in the Monte Carlo in the next section.

4. FINITE SAMPLE PERFORMANCE

Here, we study the finite sample performance of the TVLW estimator presented in Section 3 for different values of P and different models. Time-invariant and time-varying long-memory parameters are considered in the model

$$v_{t,T} = (1-L)^{-d_P^v(t/T,\xi^0)} a_t, \qquad t = 1, \dots, T,$$

with a_t being a Gaussian unit variance white noise or a weakly dependent AR(1) process. The true memory is defined as $d_P^v(u,\xi^0) = \sum_{v=0}^{P^0} \xi_v^0 P_{v,T}(u)$ for some P^0 and ξ_v^0, such that $-0.5 \le d_P^v(u,\xi^0) < 0.5$ for all $u \in [0,1]$. The autoregressive coefficient takes the value 0.5, as considered in Henry (2001). The Time-Varying (TV) fractional difference operator is defined as

$$(1-L)^{-d_P^v(t/T,\xi^0)} = \sum_{j=0}^{\infty} c_j(t/T,\xi^0) L^j \ \text{ for } \ c_j(t/T,\xi^0) = \frac{\Gamma[j + d_P^v(t/T,\xi^0)]}{\Gamma(j+1)\Gamma[d_P^v(t/T,\xi^0)]}$$

and the series are generated by truncating the sum in the previous expression as

$$v_{t,T} = \sum_{j=0}^{t+K-1} c_j(t/T,\xi^0) a_{t-j},$$

for $t = 1, \dots, T$ and $T+K$ observations of the weak dependent a_t, say $a_{1-K}, a_{2-K}, \dots, a_T$. The deviation between $v_{t,T}$ and the Type I long memory defined in Assumption A.1 is $\sum_{j=t+K}^{\infty} c_j(t/T,\xi^0) a_{t-j}$ with mean zero and variance $O\left((t+K)^{2d_P^v(t/T,\xi^0)-1}\right)$ because $c_j(t/T,\xi^0)$ is proportional to $j^{d_P^v(t/T,\xi^0)-1}$ as $j \to \infty$ for $d_P^v(t/T,\xi^0) < 0.5$. Then $v_{t,T}$ approaches a locally stationary Type I long memory series for K large enough (see Marinucci and Robinson, 1999; Davidson and Hashimzade, 2009). In this Monte Carlo we use K=100 (similar results, available on request, have been obtained with $K = 1000$). In the locally stationary case $x_{t,T} = v_{t,T}$ such that $d_P(t/T,\xi^0) = d_P^v(t/T,\xi^0) < 0.5$ and for $d_P(t/T,\xi^0) \ge 1/2$ the series $x_{t,T}$ were generated as $\sum_{s=1}^{t} v_{s,T}$ as explained in Assumption A.1 and $d_P(t/T,\xi^0) = d_P^v(t/T,\xi^0) + 1$. See also Johansen and Nielsen (2016) for an analysis of the effects of the initial values in the estimation of non-stationary standard fractionally differenced series.

Two different Data Generating Processes (DGP) are considered: DGP1 is a time-invariant process with $P^0 = 0$ and $d = \xi_0 = 0.4$; and DGP2 is time-varying with $P^0 = 1$, and three different sets of values for $\xi^0 = (\xi_0^0, \xi_1^0)$, namely $\xi^0 = (0.6, 0.1)$ (d in the range 0.458 to 0.741), $\xi^0 = (0.2, 0.2)$ (d ranging from -0.082 to 0.482) and $\xi^0 = (0.3, 0.3)$ (d in the range -0.124 to 0.724). We consider different polynomials $P = 1, 2, 5$ so that the true parameter vector is $\xi^0 = \left(\xi_0^0, 0_P'\right)'$ for DGP1 and $\xi^0 = \left(\xi_0^0, \xi_1^0, 0_{P-1}'\right)'$ for DGP2 where 0_P is a $P \times 1$ vector of zeros. The results do not change significantly for different coefficients ξ^0. All of them are available on request.

We split the time series of T observations into M blocks of length N, shifting S positions forward each time. For the sake of comparison, we use the same (optimal) values proposed by Palma and Olea (2010) for the case of overlapping blocks, that is, the integer parts of $M = 12.629, N = 105, S = 35$ for $T = 512$ and $M = 19.311, N = 200, S = 45$ for $T = 1024$. For non-overlapping blocks, we consider powers of two, namely $M = 8, N = S = 64$ for $T = 512$ and $M = 8, N = S = 128$ for $T = 1024$. With respect to the selection of the bandwidth m (number of frequencies used in the estimation, $\lambda_k = 2\pi k/N$, $k = 1, \dots, m$) we consider three different cases to analyse the sensitivity of the results to the bandwidth: $m = N^\tau$, where $\tau = \{0.6, 0.7, 0.8\}$ for the white noise case, and $\tau = \{0.3, 0.4, 0.5\}$ for the AR(1) process. The lower values of m aim to reduce the biasing effects of the weak dependent noise, as in other local estimators in long memory series.

In this Monte Carlo exercise we compare our semi-parametric estimator with the fully parametric one proposed by Palma and Olea (2010) when a_t follows the weakly dependent AR(1) process under DGP1. Notice that under DGP1, Palma and Olea's model specifies $d(u) = \alpha_0 + \alpha_1 u$ with true values $\alpha_0 = d$ and $\alpha_1 = 0$. For completeness, we consider a correctly specified model (ARFIMA(1,d,0), $p = 1, q = 0$), an over-specified model (ARFIMA(1,d,1),

Table I. Finite sample performance of the TVLW estimator under a time-invariant long-memory parameter (white noise errors)

	T	512			1024		
		EstM	MAD	95%Cov	EstM	MAD	95%Cov
	m	ξ_0 $\xi_1,\dots,\xi_P$	ξ_0 $\xi_1,\dots,\xi_P$	ξ_0 $\xi_1,\dots,\xi_P$	ξ_0 $\xi_1,\dots,\xi_P$	ξ_0 $\xi_1,\dots,\xi_P$	ξ_0 $\xi_1,\dots,\xi_P$
				DGP1 - Overlap			
$P=1$	$N^{0.6}$	0.398 −0.0002	0.063 0.066	0.819 0.785	0.402 −0.0005	0.044 0.047	0.746 0.686
	$N^{0.7}$	0.391 −0.001	0.046 0.046	0.805 0.799	0.398 −0.0008	0.032 0.034	0.732 0.691
	$N^{0.8}$	0.374 0.0003	0.039 0.036	0.755 0.783	0.383 −0.0003	0.028 0.025	0.657 0.709
$P=2$	$N^{0.6}$	0.395 −0.0007	0.063 0.066	0.856 0.815	0.400 −0.001	0.045 0.046	0.784 0.720
	$N^{0.7}$	0.389 −0.0002	0.047 0.048	0.856 0.810	0.396 −0.0007	0.033 0.034	0.759 0.718
	$N^{0.8}$	0.372 −0.0007	0.041 0.036	0.791 0.802	0.382 −0.001	0.028 0.024	0.689 0.732
$P=5$	$N^{0.6}$	0.386 0.0003	0.066 0.063	0.919 0.870	0.396 −0.001	0.045 0.042	0.841 0.814
	$N^{0.7}$	0.384 0.0001	0.050 0.044	0.914 0.881	0.394 −0.0004	0.033 0.030	0.838 0.813
	$N^{0.8}$	0.368 −0.0002	0.043 0.033	0.868 0.887	0.380 −0.001	0.029 0.022	0.760 0.828
				DGP1 – No overlap			
$P=1$	$N^{0.6}$	0.394 −0.002	0.060 0.059	0.933 0.930	0.401 0.0002	0.044 0.046	0.934 0.931
	$N^{0.7}$	0.385 −0.002	0.046 0.043	0.923 0.944	0.396 −0.0004	0.031 0.032	0.939 0.935
	$N^{0.8}$	0.364 −0.001	0.044 0.034	0.841 0.926	0.378 −0.0005	0.029 0.023	0.865 0.938
$P=2$	$N^{0.6}$	0.390 −0.002	0.061 0.061	0.932 0.929	0.398 −0.0007	0.044 0.046	0.934 0.925
	$N^{0.7}$	0.382 −0.001	0.046 0.045	0.924 0.936	0.394 −0.0006	0.032 0.032	0.936 0.937
	$N^{0.8}$	0.362 −0.001	0.046 0.034	0.824 0.920	0.377 −0.001	0.030 0.023	0.853 0.941
$P=5$	$N^{0.6}$	0.379 −0.001	0.063 0.062	0.913 0.923	0.391 −0.0009	0.045 0.046	0.930 0.932
	$N^{0.7}$	0.374 −0.001	0.049 0.045	0.904 0.931	0.389 −0.0006	0.033 0.032	0.935 0.935
	$N^{0.8}$	0.357 −0.001	0.049 0.034	0.786 0.931	0.374 −0.0006	0.032 0.023	0.829 0.936

Notes: The top value in each cell is for $\hat{d} \equiv \hat{\xi}_0$ and the bottom one for the average over $\hat{\xi}_1, \dots, \hat{\xi}_P$. True values: $\xi_0 = 0.4$; $\xi_1 = \cdots = \xi_P = 0$.

$p = 1, q = 1$), and two misspecified models that neglect the existing AR dependence (ARFIMA(0,d,1), $p = 0, q = 1$, and ARFIMA(0,d,0), $p = 0, q = 0$). As in Palma and Olea (2010), we consider two different situations regarding the time-varying nature of the parameters. First, only the memory parameter is considered as being time-varying, while the rest of parameters are constant. In the second specification, all the parameters are allowed to be time-varying as linear functions of u, introducing a higher degree of flexibility. It is important to remember that Palma and Olea's optimal N and S satisfy $N >> S$ and, therefore, their estimator would be expected to do much better in the case of overlapping blocks than under the non-overlapping scheme. We use their 'LS.whittle' function in their R library 'LSTS' to compute the point estimates and confidence intervals.

Tables I–III report the results for DGP1 of our TVLW estimator together with Palma and Olea's results to compare the performance of both estimation methods. Tables IV and V show the results for DGP2 with $(\xi_0, \xi_1) = (0.6, 0.1)$ (under white noise and AR(1) with a constant autorregressive parameter respectively). Table VI presents the empirical size and power of the test of time-constancy of $d(u)$ (cf Remark 5 and Corollary 1) for DGP1 and DGP2 respectively, and both error's models, all based on 1000 replications. Tables I–V present the average (EstM) and the mean absolute deviation (MAD) of the estimates, as well as the 95% coverage rates (95%Cov) of confidence intervals calculated based on the limit distribution in Theorem 2 and Remark 2. The coverage rates are the percentage of cases in which ξ_j^0 falls within the confidence interval $[\hat{\xi}_j \pm 1.96\text{SE}]$, where *se* is the standard error

Table II. Finite sample performance of the TVLW estimator under a time-invariant long-memory parameter (AR(1) errors)

	T	512			1024		
		EstM	MAD	95%Cov	EstM	MAD	95%Cov
	m	ξ_0 $\xi_1,\dots,\xi_P$	ξ_0 $\xi_1,\dots,\xi_P$	ξ_0 $\xi_1,\dots,\xi_P$	ξ_0 $\xi_1,\dots,\xi_P$	ξ_0 $\xi_1,\dots,\xi_P$	ξ_0 $\xi_1,\dots,\xi_P$
				DGP1 – Overlap			
$P = 1$	$N^{0.3}$	0.431 0.004	0.179 0.180	0.839 0.818	0.408 −0.002	0.139 0.145	0.735 0.714
	$N^{0.4}$	0.462 −0.005	0.136 0.132	0.793 0.794	0.424 0.0007	0.095 0.099	0.729 0.720
	$N^{0.5}$	0.508 −0.0006	0.126 0.088	0.613 0.803	0.461 −0.003	0.082 0.068	0.606 0.700
$P = 2$	$N^{0.3}$	0.423 0.0004	0.180 0.185	0.875 0.831	0.399 −0.006	0.142 0.142	0.764 0.731
	$N^{0.4}$	0.455 −0.005	0.135 0.132	0.849 0.829	0.418 −0.002	0.095 0.096	0.771 0.747
	$N^{0.5}$	0.506 0.0005	0.124 0.089	0.692 0.820	0.458 −0.001	0.081 0.067	0.648 0.726
$P = 5$	$N^{0.3}$	0.401 0.002	0.197 0.202	0.929 0.842	0.386 −0.004	0.148 0.140	0.826 0.790
	$N^{0.4}$	0.436 −0.0006	0.142 0.138	0.903 0.854	0.408 −0.003	0.098 0.093	0.843 0.808
	$N^{0.5}$	0.495 −0.0007	0.121 0.088	0.829 0.868	0.452 −0.0002	0.080 0.062	0.760 0.810
				DGP1 – No overlap			
$P = 1$	$N^{0.3}$	0.452 −0.003	0.214 0.206	0.897 0.904	0.420 0.004	0.151 0.150	0.910 0.923
	$N^{0.4}$	0.497 −0.002	0.147 0.119	0.866 0.932	0.450 −0.002	0.099 0.090	0.904 0.924
	$N^{0.5}$	0.566 −0.006	0.172 0.081	0.568 0.929	0.491 −0.003	0.102 0.063	0.749 0.939
$P = 2$	$N^{0.3}$	0.445 −0.009	0.215 0.206	0.896 0.904	0.414 −0.001	0.151 0.152	0.914 0.915
	$N^{0.4}$	0.491 −0.001	0.146 0.123	0.872 0.922	0.445 −0.001	0.099 0.091	0.908 0.921
	$N^{0.5}$	0.562 −0.003	0.169 0.082	0.585 0.927	0.488 −0.0009	0.100 0.063	0.757 0.934
$P = 5$	$N^{0.3}$	0.408 −0.008	0.217 0.219	0.892 0.882	0.388 −0.001	0.156 0.157	0.916 0.897
	$N^{0.4}$	0.469 −0.001	0.143 0.130	0.889 0.902	0.429 −0.0006	0.097 0.093	0.909 0.914
	$N^{0.5}$	0.549 −0.001	0.158 0.085	0.638 0.915	0.478 −0.0006	0.094 0.064	0.787 0.927

Notes: The top value in each cell is for $\hat{d} \equiv \hat{\xi}_0$ and the bottom one for the average over $\hat{\xi}_1, \dots, \hat{\xi}_P$. True values: $\xi_0 = 0.4$; $\xi_1 = \cdots = \xi_P = 0$.

defined in Remark 2. For DGP1 when $P = 2$ and 5 and DGP2 with $P = 5$, the results show the averages over the corresponding individual $\xi_v' s$ when there is more than one $\xi_v = 0$, $v = 1, \dots$. Acknowledging the fact that some positive and negative biases may cancel out, special attention is given to the MAD and the 95%Cov measures.

In general, Tables I,II,IV and V show that the finite sample performance of the proposed TVLW estimator seems to be quite good in terms of bias and precision, with estimated coefficients close to the true ones and absolute deviations relatively small (for small m if the noise comes from an AR(1), as it avoids the bias caused by the weak dependence). The results do not seem to be sensitive to the choice of P or to the value of d, performing similarly for stationary and non-stationary values. As expected, the mean absolute deviation (MAD) of the estimates for the models with white noise innovations is smaller than for the AR(1) case. Moreover, comparing overlapping and non-overlapping TVLW estimation, the coverage rate is far from the nominal 95% in the overlapping case when the limiting distribution in Theorem 2 is used. This confirm the suspicion that the requirement of non-overlapping blocks is necessary in Theorem 2. Unlike Palma and Olea's parametric estimator, the use of non-overlapping blocks leads to better results, with coverage rates close to the nominal 95%, even for the smaller T. In fact, our coverage rates compare very favourably with the ones shown by Wang (2019) for his estimators, as they are closer to the 95% level. He considers pointwise GPH and local Whittle estimators for locally stationary long memory processes with a TVLM parameter by fixing $d(u)$ for each $u = u_0 \in [0, 1]$. In this way he obtains an estimator for $d\left(u_0\right)$, discussing its consistency and asymptotic normality for a given u_0. Considering a sequence of values

Table III. Finite sample performance of Palma and Olea's (2010) estimator under a time-invariant long-memory parameter

T	512			1024			512			1024		
	only d linear in u						all linear in u					
	EstM	MAD	95%Cov	EstM	MAD	95%Cov	EstM	MAD	95%Cov	EstM	MAD	95%Cov
	α_0 α_1	α_0 α_1	α_0 α_1	α_0 α_1	α_0 α_1	α_0 α_1	α_0 α_1	α_0 α_1	α_0 α_1	α_0 α_1	α_0 α_1	α_0 α_1
	DGP1 – Overlap											
$p=1,q=0$	0.325 0.002	0.157 0.148	0.801 0.597	0.336 0.002	0.116 0.090	0.811 0.599	0.329 −0.010	0.222 0.384	0.878 0.919	0.324 0.019	0.199 0.334	0.831 0.861
$p=1,q=1$	0.265 0.002	0.219 0.147	0.791 0.598	0.310 0.002	0.143 0.090	0.859 0.595	0.285 −0.037	0.298 0.501	0.851 0.876	0.306 −0.018	0.238 0.406	0.829 0.837
$p=0,q=1$	0.649 0.002	0.251 0.154	0.118 0.590	0.638 0.001	0.238 0.095	0.014 0.611	0.640 0.022	0.245 0.212	0.607 0.959	0.636 0.005	0.236 0.131	0.284 0.958
$p=0,q=0$	0.856 0.003	0.456 0.164	0.001 0.615	0.842 0.001	0.442 0.103	0.000 0.644	0.856 0.003	0.456 0.166	0.017 0.949	0.842 0.001	0.442 0.103	0.000 0.955
	DGP1 – No overlap											
$p=1,q=0$	0.318 0.003	0.177 0.166	0.754 0.503	0.327 0.002	0.144 0.105	0.708 0.530	0.328 −0.003	0.229 0.364	0.829 0.879	0.325 0.008	0.202 0.314	0.770 0.802
$p=1,q=1$	0.229 0.004	0.253 0.164	0.754 0.504	0.282 0.002	0.188 0.105	0.745 0.520	0.248 −0.001	0.328 0.500	0.790 0.787	0.291 −0.021	0.263 0.409	0.757 0.763
$p=0,q=1$	0.662 0.003	0.267 0.172	0.162 0.496	0.647 0.002	0.247 0.111	0.028 0.537	0.658 0.015	0.269 0.243	0.521 0.890	0.645 0.007	0.245 0.154	0.247 0.887
$p=0,q=0$	0.880 0.002	0.480 0.181	0.001 0.516	0.854 0.0006	0.454 0.120	0.000 0.551	0.879 0.001	0.479 0.182	0.016 0.887	0.853 0.0007	0.453 0.121	0.000 0.864

Notes: The top value is for $\hat{d} \equiv \hat{\alpha}_0$ and the bottom one for $\hat{\alpha}_1$. True model is $d(u) = \alpha_0 + \alpha_1 u$, where $\alpha_0 = 0.4, \alpha_1 = 0$. Correct specification: $p = 1, q = 0$.

$u_0 \in [0, 1]$ a discrete series of estimators of $d(u_0)$ is obtained, but their joint asymptotic distribution is not obtained. However, our TVLW estimator provides estimation of the entire continuous path $d(u)$ given the point estimates of $\xi' s$, which allows a standard testing strategy for consistency of the memory parameter. Interestingly, our estimator also compares very favourably with the semi-parametric local Whittle estimator of Robinson (1995b), which is in agreement with the fastest convergence discussed in Remark 3. For $T = 512$, with $m = 44$ (close to $T^{0.6}$ and thus comparable with our TVLW estimator with $m = N^{0.6}$), the EstM of the local Whittle estimator of d equals 0.350, the MAD is 0.082 and the 95%Cov of 0.850 for the ARFIMA(0,d,0) model, which are slightly worse than the results in Table I obtained with the TVLW estimator.

We now turn to Palma and Olea's parametric estimator (Table III). Overall, the results with only d or all the parameters linear in u are very similar and, not surprisingly, the performance of the estimator is better with overlapping blocks. The estimator performs well under a correct specification of the short-memory component, but things get really problematic when the parametric specification is overfitted or wrong: the estimator is imprecise under misspecification (large MAD and poor coverage) and it is biased upwards, with an average estimate of around 0.65 under $p = 0, q = 1$ when the true value is 0.4. Comparing Tables II and III, the main conclusion is that under the correct specification of the DGP our semi-parametric TVLW estimator performs barely as well as the parametric Whittle estimator (the MAD of the TVLW estimator is even lower and has a better percentage coverage with no overlapping blocks), but when the parametric specification of the Whittle estimator is wrong our semi-parametric estimator is definitely the best option.

Table VI shows that the proposed test for time-invariant $d(u)$ is slightly oversized, especially for larger P and AR(1) errors, but the size approaches the nominal 0.05 as the sample size increases. As expected, the power increases with the value of ξ_1 and gets close to one for $T = 1024$ and $\xi_1 = 0.3$ with white noise errors and also with a bandwidth of $m = N^{0.5}$ for AR(1) errors. The power decreases as P gets larger. Increasing P unnecessarily implies that the true DGP is closer to the null hypothesis (remember that $\xi_v = 0$ for $v > 1$) and thus reduces the power of the test. The results are quantitatively and qualitatively similar for a time-varying AR(1) coefficient close to one (results available on request) where the estimates of $\xi_j^0, j > 0$ show a similar small MAD and coverage rates close to 0.95. However, as in the standard LW estimation with constant d, the results for ξ_0^0 deteriorate as the AR coefficient gets close to 1.

Table IV. Finite sample performance of the TVLW estimator under a stationary time-varying long-memory parameter (white noise errors)

	T	512						1024					
		EstM		MAD		95%Cov		EstM		MAD		95%Cov	
	m	ξ_0 ξ_1	$\xi_2,\ldots,\xi_P$	ξ_0 ξ_1	$\xi_2,\ldots,\xi_P$	ξ_0 ξ_1	$\xi_2,\ldots,\xi_P$	ξ_0 ξ_1	$\xi_2,\ldots,\xi_P$	ξ_0 ξ_1	$\xi_2,\ldots,\xi_P$	ξ_0 ξ_1	$\xi_2,\ldots,\xi_P$
						DGP2 – Overlap							
$P=1$	$N^{0.6}$	0.574 0.093	—	0.061 0.063	—	0.820 0.813	—	0.593 0.097	—	0.045 0.047	—	0.726 0.695	—
	$N^{0.7}$	0.564 0.088	—	0.053 0.046	—	0.774 0.807	—	0.586 0.096	—	0.034 0.034	—	0.709 0.691	—
	$N^{0.8}$	0.540 0.084	—	0.062 0.036	—	0.589 0.782	—	0.565 0.092	—	0.038 0.026	—	0.481 0.673	—
$P=2$	$N^{0.6}$	0.573 0.092	0.006	0.063 0.064	0.064	0.865 0.808	0.827	0.592 0.097	0.004	0.046 0.048	0.045	0.768 0.695	0.775
	$N^{0.7}$	0.564 0.087	0.005	0.053 0.047	0.045	0.790 0.803	0.842	0.586 0.096	0.005	0.034 0.034	0.034	0.748 0.699	0.749
	$N^{0.8}$	0.541 0.083	0.007	0.062 0.037	0.036	0.578 0.771	0.831	0.566 0.092	0.006	0.038 0.026	0.025	0.535 0.667	0.744
$P=5$	$N^{0.6}$	0.565 0.094	0.001	0.068 0.075	0.062	0.913 0.822	0.877	0.587 0.098	0.001	0.046 0.050	0.041	0.831 0.760	0.821
	$N^{0.7}$	0.558 0.090	0.001	0.058 0.051	0.043	0.855 0.842	0.885	0.583 0.097	0.002	0.035 0.034	0.030	0.804 0.774	0.822
	$N^{0.8}$	0.537 0.084	0.002	0.065 0.040	0.033	0.687 0.838	0.890	0.564 0.092	0.001	0.039 0.026	0.022	0.594 0.764	0.818
						DGP2 – No overlap							
$P=1$	$N^{0.6}$	0.573 0.094	—	0.065 0.059	—	0.914 0.940	—	0.595 0.098	—	0.045 0.046	—	0.947 0.917	—
	$N^{0.7}$	0.560 0.089	—	0.055 0.044	—	0.868 0.938	—	0.586 0.096	—	0.034 0.033	—	0.919 0.923	—
	$N^{0.8}$	0.528 0.083	—	0.072 0.036	—	0.569 0.931	—	0.560 0.092	—	0.042 0.025	—	0.694 0.928	—
$P=2$	$N^{0.6}$	0.570 0.094	0.003	0.066 0.059	0.062	0.906 0.937	0.932	0.592 0.098	0.004	0.046 0.046	0.046	0.939 0.914	0.924
	$N^{0.7}$	0.558 0.089	0.005	0.057 0.045	0.046	0.854 0.931	0.924	0.584 0.097	0.005	0.035 0.033	0.032	0.913 0.917	0.942
	$N^{0.8}$	0.526 0.083	0.005	0.074 0.036	0.034	0.559 0.928	0.913	0.559 0.092	0.005	0.043 0.025	0.024	0.682 0.928	0.929
$P=5$	$N^{0.6}$	0.559 0.095	0.001	0.071 0.060	0.061	0.871 0.937	0.930	0.585 0.099	0.001	0.047 0.046	0.046	0.930 0.901	0.929
	$N^{0.7}$	0.550 0.089	0.001	0.061 0.045	0.045	0.811 0.936	0.932	0.579 0.097	0.001	0.036 0.033	0.032	0.897 0.917	0.938
	$N^{0.8}$	0.521 0.082	0.002	0.079 0.036	0.033	0.505 0.927	0.932	0.555 0.092	0.001	0.046 0.025	0.024	0.651 0.926	0.930

Notes: In the left, the top value is for $\hat{\xi}_0$ and the bottom one for $\hat{\xi}_1$. In the right, it is the average over $\hat{\xi}_2, \ldots, \hat{\xi}_P$. True values: $\xi_0 = 0.6$, $\xi_1 = 0.1$, and $\xi_2 = \cdots = \xi_P = 0$.

5. EMPIRICAL APPLICATION

Here we discuss the application of the TVLW estimator to two different sets of time series previously considered in the long memory literature. In the first application we analyse the inflation rate in the USA and compare our results with those obtained using the estimator proposed by Palma and Olea (2010), bearing in mind the findings published in the previous study by Martins and Rodrigues (2014). Inflation in developed countries displays very strong persistence, approaching that of a random-walk process (e.g., Fuhrer and Moore, 1995 or Arteche, 2007), and apparently this persistence has been changing over time with a decrease in memory over recent decades. Martins and Rodrigues (2014) only considered a single change in persistence at an unknown date. Using monthly data from January 1951 to December 2009 they found an $I(1)$ process for the annual US inflation rate throughout the entire period, with $\hat{d} = 1.148$ for the period January/1951–April/1982 and $\hat{d} = 0.855$ for the period May/1982–December/2009, but in both regimes the null hypothesis of $d = 1$ was not statistically rejected. We update the inflation rate up to December 2020 and consider the possibility of TVLM in this series.

In the second example, we aim to add some further discussion to the empirical literature on climate variables by looking at the time series of tree ring widths. To compare methods, we use exactly the same tree ring data used by Palma and Olea (2010), who applied their parametric Gaussian long-memory locally stationary methodology

Table V. Finite sample performance of the TVLW estimator under a stationary time-varying long-memory parameter (AR(1) errors)

	T	512						1024					
		EstM		MAD		95%Cov		EstM		MAD		95%Cov	
	m	ξ_0 ξ_1	$\xi_2,\dots,\xi_P$	ξ_0 ξ_1	$\xi_2,\dots,\xi_P$	ξ_0 ξ_1	$\xi_2,\dots,\xi_P$	ξ_0 ξ_1	$\xi_2,\dots,\xi_P$	ξ_0 ξ_1	$\xi_2,\dots,\xi_P$	ξ_0 ξ_1	$\xi_2,\dots,\xi_P$
						DGP2 – Overlap							
$P=1$	$N^{0.3}$	0.596 0.086	—	0.171 0.171	—	0.849 0.840	—	0.600 0.102	—	0.136 0.137	—	0.767 0.731	—
	$N^{0.4}$	0.630 0.084	—	0.127 0.125	—	0.841 0.810	—	0.616 0.098	—	0.093 0.096	—	0.736 0.726	—
	$N^{0.5}$	0.681 0.082	—	0.109 0.086	—	0.702 0.815	—	0.654 0.097	—	0.077 0.064	—	0.622 0.725	—
$P=2$	$N^{0.3}$	0.594 0.084	0.010	0.177 0.175	0.180	0.888 0.833	0.849	0.598 0.102	0.010	0.139 0.138	0.142	0.783 0.725	0.763
	$N^{0.4}$	0.625 0.083	0.001	0.129 0.127	0.124	0.875 0.808	0.868	0.613 0.098	0.005	0.094 0.097	0.097	0.776 0.727	0.749
	$N^{0.5}$	0.678 0.081	0.005	0.110 0.088	0.084	0.762 0.818	0.863	0.651 0.097	0.001	0.077 0.064	0.065	0.683 0.724	0.745
$P=5$	$N^{0.3}$	0.572 0.085	−0.001	0.198 0.227	0.195	0.915 0.806	0.840	0.581 0.106	0.004	0.143 0.164	0.136	0.840 0.750	0.805
	$N^{0.4}$	0.613 0.080	0.0009	0.139 0.155	0.130	0.921 0.810	0.867	0.602 0.100	0.003	0.095 0.106	0.090	0.867 0.764	0.812
	$N^{0.5}$	0.671 0.081	0.003	0.111 0.104	0.087	0.860 0.834	0.866	0.645 0.098	0.002	0.075 0.067	0.059	0.792 0.783	0.816
						DGP2 – No overlap							
$P=1$	$N^{0.3}$	0.628 0.094	—	0.205 0.201	—	0.914 0.920	—	0.618 0.099	—	0.152 0.154	—	0.922 0.928	—
	$N^{0.4}$	0.676 0.085	—	0.133 0.121	—	0.900 0.918	—	0.641 0.098	—	0.096 0.092	—	0.908 0.920	—
	$N^{0.5}$	0.743 0.084	—	0.151 0.082	—	0.651 0.924	—	0.682 0.098	—	0.096 0.063	—	0.790 0.934	—
$P=2$	$N^{0.3}$	0.622 0.094	0.017	0.210 0.204	0.211	0.907 0.915	0.894	0.610 0.100	0.008	0.151 0.156	0.161	0.915 0.924	0.900
	$N^{0.4}$	0.670 0.086	0.011	0.133 0.122	0.122	0.889 0.928	0.912	0.636 0.097	0.004	0.095 0.092	0.092	0.904 0.915	0.918
	$N^{0.5}$	0.739 0.084	0.006	0.149 0.082	0.082	0.654 0.918	0.921	0.679 0.098	−0.0006	0.094 0.064	0.063	0.801 0.936	0.924
$P=5$	$N^{0.3}$	0.588 0.095	−0.0004	0.218 0.213	0.215	0.897 0.899	0.894	0.585 0.100	0.001	0.158 0.164	0.163	0.911 0.902	0.892
	$N^{0.4}$	0.651 0.089	0.003	0.132 0.126	0.130	0.896 0.908	0.898	0.621 0.097	−0.0005	0.094 0.095	0.092	0.907 0.905	0.918
	$N^{0.5}$	0.726 0.084	0.001	0.140 0.087	0.086	0.690 0.914	0.918	0.670 0.095	0.001	0.090 0.064	0.064	0.822 0.931	0.927

Notes: In the left, the top value is for $\hat{\xi}_0$ and the bottom one for $\hat{\xi}_1$. In the right, it is the average over $\hat{\xi}_2, \dots, \hat{\xi}_P$. True values: $\xi_0 = 0.6$, $\xi_1 = 0.1$, and $\xi_2 = \cdots = \xi_P = 0$.

in the supplementary material. These annual data extend from 0 AD to 1989 AD. Based on the sample ACF, they concluded that the strength of the persistence decreases with time. Palma and Olea (2010) estimated a Gaussian locally stationary ARFIMA(1,d,1) model with a linear time-varying d and found that $\hat{d}$ dropped from 0.328 in 0 AD to 0.139 in 1989 AD. We reassess these findings below using our alternative TVLW semi-parametric estimator.

For both illustrations, we select P in the specification of $d_P(u,\xi)$ by testing the statistical significance of the coefficients using non-overlapping blocks and the results in Theorem 2 and Remark 2. We consider the range of $P = 0, 1, 2, \dots, 10$ polynomials, doing backwards elimination starting from the largest model $P_{\max}$. Using the test statistic in Corollary 1, we also test for the constancy of the memory parameter $d_P(u,\xi)$. Moreover, we calculate the confidence intervals of $d_P(u,\xi)$ as usual from the asymptotic normality in Theorem 2 and $\mathrm{Var}(\hat{d}_P(u,\hat{\xi})) = \sum_{v=0}^{P} \mathrm{var}(\hat{\xi}_v) P^2_{v,T}(u)$ using the asymptotic independence of the different $\hat{\xi}_v$. Finally, the goodness-of-fit of the estimated memory function is assessed by studying the long-memory properties of the filtered time series obtained by plugging the estimated $d_P(u,\hat{\xi})$ into the expansion of $(1-L)^{d_P(u,\hat{\xi})}$. The statistical significance of all the ξ's in the filtered series is then tested to assess that no memory remains in the filtered series, implying that our estimator correctly captures all the persistence of the series.

DOI: 10.1111/jtsa.12782 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Table VI. Empirical size and power for the test of time-constancy of $d(u)$

	T / m		512			1024	
			Size				
$P=1$	$(N^{0.6}, N^{0.3})$		(0.070, 0.096)			(0.069, 0.077)	
	$(N^{0.7}, N^{0.4})$		(0.056, 0.068)			(0.065, 0.076)	
	$(N^{0.8}, N^{0.5})$		(0.074, 0.071)			(0.062, 0.061)	
$P=2$	$(N^{0.6}, N^{0.3})$		(0.079, 0.120)			(0.080, 0.099)	
	$(N^{0.7}, N^{0.4})$		(0.079, 0.093)			(0.061, 0.095)	
	$(N^{0.8}, N^{0.5})$		(0.077, 0.081)			(0.069, 0.078)	
$P=5$	$(N^{0.6}, N^{0.3})$		(0.128, 0.202)			(0.093, 0.166)	
	$(N^{0.7}, N^{0.4})$		(0.078, 0.168)			(0.081, 0.133)	
	$(N^{0.8}, N^{0.5})$		(0.097, 0.124)			(0.083, 0.097)	
			Power				
		$\xi_0 = 0.6$, $\xi_1 = 0.1$	$\xi_0 = 0.2$, $\xi_1 = 0.2$	$\xi_0 = 0.3$, $\xi_1 = 0.3$	$\xi_0 = 0.6$, $\xi_1 = 0.1$	$\xi_0 = 0.2$, $\xi_1 = 0.2$	$\xi_0 = 0.3$, $\xi_1 = 0.3$
$P=1$	$N^{0.6}$	0.288	0.713	0.971	0.473	0.936	1.000
	$N^{0.3}$	0.110	0.159	0.254	0.137	0.259	0.437
	$N^{0.7}$	0.393	0.896	0.997	0.711	0.994	1.000
	$N^{0.4}$	0.116	0.264	0.481	0.170	0.485	0.760
	$N^{0.8}$	0.541	0.979	0.999	0.883	1.000	1.000
	$N^{0.5}$	0.164	0.457	0.769	0.276	0.716	0.959
$P=2$	$N^{0.6}$	0.233	0.616	0.939	0.371	0.892	0.999
	$N^{0.3}$	0.131	0.170	0.243	0.139	0.249	0.397
	$N^{0.7}$	0.332	0.831	0.990	0.625	0.990	1.000
	$N^{0.4}$	0.121	0.234	0.404	0.152	0.423	0.696
	$N^{0.8}$	0.460	0.961	0.999	0.827	1.000	1.000
	$N^{0.5}$	0.145	0.388	0.712	0.231	0.655	0.937
$P=5$	$N^{0.6}$	0.196	0.516	0.862	0.308	0.808	0.990
	$N^{0.3}$	0.192	0.209	0.248	0.221	0.286	0.391
	$N^{0.7}$	0.269	0.711	0.977	0.485	0.974	1.000
	$N^{0.4}$	0.182	0.275	0.395	0.175	0.361	0.583
	$N^{0.8}$	0.357	0.883	0.999	0.704	1.000	1.000
	$N^{0.5}$	0.185	0.356	0.616	0.204	0.527	0.871

Notes: $m = N^{0.6}, N^{0.7}, N^{0.8}$ are for white noise errors (See DGP1 in Table I for size and DGP2 in Table IV for power); $m = N^{0.3}, N^{0.4}, N^{0.5}$ are for AR(1) errors (See DGP1 in Table II for size and DGP2 in Table V for power).

5.1. Inflation Rates

The US inflation rate series, defined as the annual growth rate of the consumer price index, spans from January 1951 to December 2020 and can be obtained from the OECD website https://data.oecd.org/price/inflation-cpi.htm. Figure 1 shows the plot of this time series. As we observe, the US inflation rate was so volatile that it even led to a period of deflation (from March to October 2009).

To keep the range of estimated memory parameters within the region where the asymptotic theory in Section 2 remains valid, the estimator is applied to the first-differences of the inflation to get estimates of $d_P^{\mathrm{FD}}(u,\xi)$. The memory of the inflation series is then obtained as $d_P(u,\xi) = d_P^{\mathrm{FD}}(u,\xi) + 1$.

Based on the Monte Carlo simulations, we consider $M = 12$ non-overlapping blocks shifting $S = N = 2^6$ observations at each block. Thus, we have $T = 768$ observations (from January 1957 to December 2020). The number of frequencies considered are $m = \lfloor N^{0.6} \rfloor$ and $m = \lfloor N^{0.7} \rfloor$. The results for this non-overlapping scheme are in Table VII, where we report the selected P, the estimated parameters $\hat{\xi}_0, \hat{\xi}_1, \dots, \hat{\xi}_P$, the standard errors, the 95% confidence interval based on Theorem 2 and Remark 2, and the test statistic for memory constancy. With both bandwidths the constancy of the memory is rejected, resulting in $P = 3$ Chebyshev polynomials needed if $m = \lfloor N^{0.7} \rfloor$ is used and $P = 2$ with $m = \lfloor N^{0.6} \rfloor$.

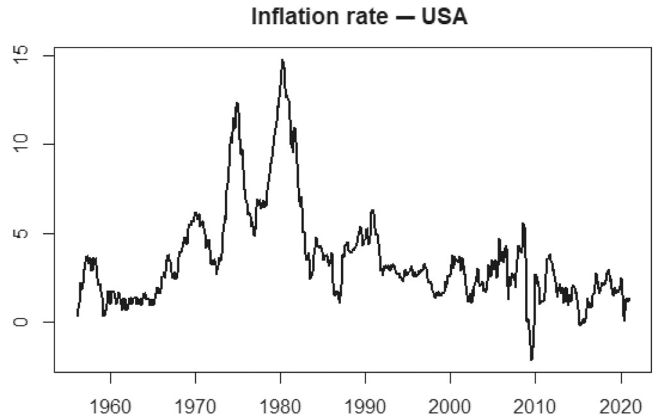

Figure 1. US inflation rate series

Table VII. TVLW estimation: US Inflation

P	2	3
m	$N^{0.6}$	$N^{0.7}$
ξ_0	$0.134^{**}_{0.057}$	$0.209^{***}_{0.043}$
	[0.021,0.247]	[0.124,0.295]
ξ_1	$0.099^{*}_{0.057}$	$0.117^{***}_{0.043}$
	[−0.013,0.212]	[0.032,0.203]
ξ_2	$-0.105^{*}_{0.057}$	$-0.043_{0.043}$
	[−0.218,0.007]	[−0.129,0.041]
ξ_3		$-0.109^{**}_{0.043}$
		[−0.194,-0.023]
StatConst	6.347^{**}	14.512^{***}

Notes: *,**,*** denote 10%, 5%, 1% significant levels; point estimates on top and the SEs below; 95%Cov in square brackets; $M = 12$; $S = N = 2^6$; $T = 768$; StatConst is the test statistic for constant $d(u)$.

Figure 2 shows the time-varying estimated long-memory function $d_3(u, \hat{\xi})$ with $m = N^{0.7}$. The form of the estimated memory with $m = N^{0.6}$ is very similar and is thus not included. The adequacy of $d_3(u, \hat{\xi})$ to estimate the memory of the series is assessed by testing the significance of the different elements in ξ in the fractionally differenced series. The test statistic in Corollary 1 with $R = I_{P+1}$ gives a p-value of 0.200 for $P = 10$ and it ranges from 0.111 for $P = 1$ to 0.601 for $P = 6$, which evidences that $d_3(u, \hat{\xi})$ captures all the memory of the series.

The results in Figure 2 confirm the main findings of Martins and Rodrigues (2014), but are at the same time more specific in terms of the degree of nonlinear persistence of the inflation rate throughout the whole period. Indeed, for the US economy, the assumption of an $I(1)$ process for the inflation rate[1] (cf. Martins and Rodrigues, 2014, and others) seems to be incorrect for one specific period. At the end of the 70s/early 80s, $d(u)$ is certainly greater than one, with a confidence interval ranging from 1.2 to 1.6. In the remaining periods (before the mid 70s and after the early 90s), we have that $d(u) = 1$ is inside the 95% confidence set so that unit root is a possibility for a given u. The evolution of the memory function in Figure 2 shows that the persistence tends to be greater in periods of high inflation than in periods of low inflation, with the peaks in the 70s related to the recession caused by the oil crises.

[1] Following Palma and Olea (2010), we estimated the stationary ARFIMA model with time-invariant coefficients using the Haslett–Raftery method in R. The point estimate for d is 1.00005 for the fractional noise model and 1.238 for the $(1, d, 0)$ model.

DOI: 10.1111/jtsa.12782 Journal of Time Series Analysis published by John Wiley & Sons Ltd.

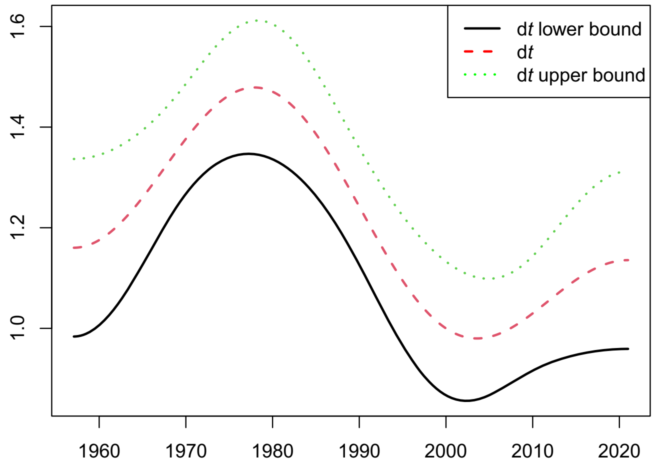

Figure 2. TVLW estimation of $d_P(u, \xi)$: US inflation

The Great Recession of 2008/2009 saw a further increase in persistence. This suggests that negative shocks are more persistent than positive ones.

To complete the analysis, we also apply Palma and Olea's parametric estimator to the US inflation rate. The same strategy prior-differencing-adding-one strategy is here used with the purpose of bounding the memory function between the values where the theory applies, i.e. $0 < \inf_u d_P(u, \xi) < \sup d_P(u, \xi) < 1/2$. We consider their optimal pair $N = 105$ and $S = 35$ and, following their approach, the LSARFIMA models are selected from the Akaike Information Criterion and by analysing the significance of the parameters involved. The model selected includes one AR and one MA coefficient ($p = q = 1$), the function $d(u)$ is of power 4, the SD and the AR are linear functions of u, and the MA coefficient is constant for all u. All parameters are statistically significant, with the exception of the constant and linear components of $d(u)$. Noticeably, the estimated AR function is $\phi(u) = -0.984 + 1.833u$ which means that for values of u close to the boundaries (u near 0 or 1)the AR coefficient is near one, in absolute value. The estimated values for the parameter $d(u)$ is represented in Figure 3 ($p = q = 1$), where $\hat{d}_P(u) = 0.162 - 0.193u + 9.451u^2 - 25.499u^3 + 16.221u^4$. Note that the validity of Palma and Olea's estimator cannot be guaranteed because the condition that $d_P(u, \xi) < 0.5$ for all u is not satisfied. However our TVLW estimator remains valid for a larger range of values and, contrary to Palma and Oleas's estimator, the asymptotic theory can be safely used for the estimates obtained.

The main conclusion is that the parametric and semi-parametric estimators show a similar evolution in the persistence of the series: large values of the memory function around the 80s and, after the 80s, they show a downward trend until 2000, when they begin to rise again.

5.2. Tree Ring

The tree ring dataset used is the same as in Palma and Olea (2010) and represents annual tree ring width measurements (tree ring standardised growth index) at Mammoth Creek, Utah, as reported by Graybill (1990). Figure 4 shows the observed time series.

We consider the non-overlapping case with $M = 15$ blocks of observations, shifting $S = N = 2^7$ observations at each block, with a total of $T = 1920$ datapoints (from 70 AD to 1989 AD). Using the same strategy as before based on Theorem 2 we find that $P = 0$ is selected for bandwidths $m = \lfloor N^{0.5} \rfloor$ and $m = \lfloor N^{0.6} \rfloor$, implying a constant d. However, for $m = \lfloor N^{0.4} \rfloor$ there is statistical evidence against a constant $d(u)$ with a decreasing tendency being observed after $t = 1250$ (around year 1320 AD). See Table VIII for the results of the TVLW estimation and the

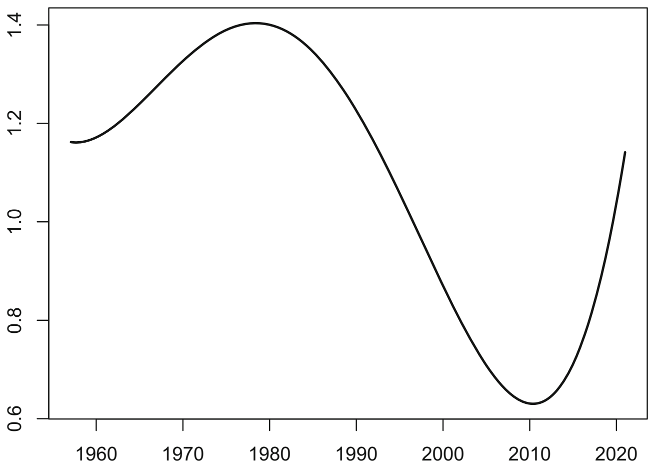

Figure 3. US inflation: P&O estimation ($p = q = 1$)

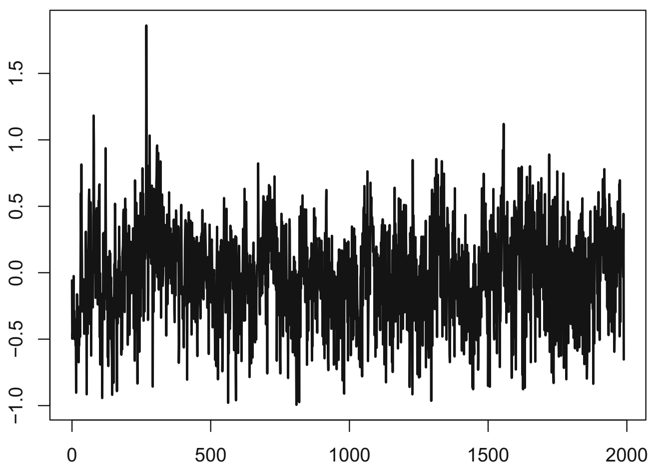

Figure 4. Tree ring width

test of constant memory parameter and Figure 5 for the time-varying estimated memory function $d_P(u, \hat{\xi})$ and its 95% confidence bands when $m = \lfloor N^{0.4} \rfloor$.

As in Palma and Olea (2010) we obtain evidence for a time-varying memory for the tree ring data when $m = \lfloor N^{0.4} \rfloor$. However, we do not observe a decline in $d(u)$ over the whole sample but only in the second half. Noticeably, during the first half of the sample $d_P(u, \hat{\xi})$ is consistently around or even greater than 0.5, in contrast to the smaller values obtained by Palma and Olea (2010). Closer to the end of the sample (after $t = 1575$) the value $d(u) = 0$ belongs to the confidence set, implying that the tree rings series may have lost its long memory since 1645. These differences with Palma and Olea (2010) could perhaps indicate a model misspecification caused by the presence of a short memory component that is not considered in their estimation procedure. To compare the adequacy of both estimated memory functions, we evaluate the persistence remaining in the filtered series using both estimated

DOI: 10.1111/jtsa.12782 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Table VIII. TVLW estimates: Tree ring width

P	3
m	$N^{0.4}$
ξ_0	$0.354^{***}_{0.076}$
	[0.203,0.505]
ξ_1	$0.109_{0.076}$
	[−0.041,0.260]
ξ_2	$-0.149^{*}_{0.076}$
	[−0.300,0.001]
ξ_3	$0.150^{*}_{0.076}$
	[−0.0002,0.301]
StatConst	9.618^{**}

Notes: *,**,*** denote 10%, 5%, 1% significant levels; Point estimates on top and the SEs below; 95%IC in square brackets; $M = 15$; $S = N = 2^7$; $T = 1920$; StatConst is the test statistic for constant $d(u)$.

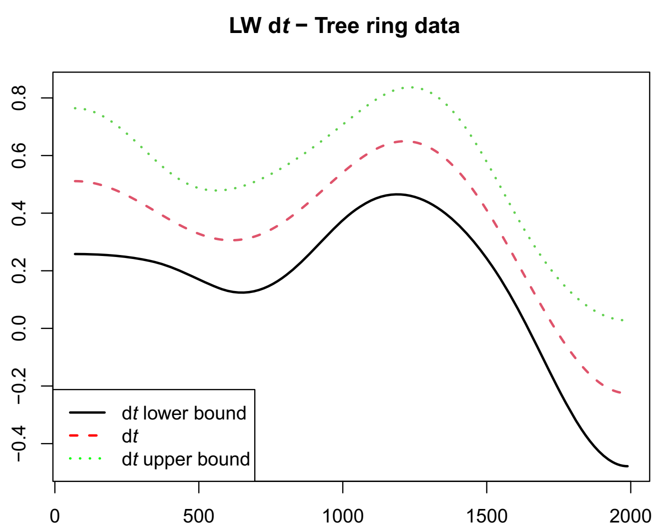

Figure 5. TVLW estimates: Tree ring width

memories: $d_3(u, \hat{\xi}) = 0.354 + 0.109P_{1,T}(u) - 0.149P_{2,T}(u) + 0.150P_{3,T}(u)$ and Palma and Olea's estimate $\hat{d}(u) = 0.3285 - 0.189u$. The test for no memory ($\xi_v = 0$ for $v = 0, 1, \ldots, P$) in the series filtered with $d(u, \hat{\xi})$ gives a p-value of 0.276 for $P = 10$, with p-values of the tests for joint significance ranging from 0.110 (for $P = 6$) to 0.995 (for $P = 1$). However, when the series is filtered with Palma and Olea's estimated $\hat{d}(u)$, the test for no memory gives a p-value of 0.020 for $P = 10$, 0.442 for $P = 1$ and between 0.003 (for $P = 6$) to 0.091 (for $P = 2$) for the rest of values of P. These results suggest that the TVLW estimates capture the persistence of the tree ring series better than the parametric estimator in Palma and Olea (2010).

6. CONCLUSION

This article proposes a semi-parametric TVLW estimator of a time-varying memory function. Its main advantages over the parametric Whittle estimator of Palma and Olea (2010) are the robustness to misspecification and to non-Gaussian distributions, and its validity under local non-stationarity and non-invertibility. The price to pay is the usual loss of efficiency if the model is correctly specified. Other semi-parametric estimators, such as those proposed by Roueff and von Sachs (2011) and Wang (2019) also share some of these characteristics, but their results cannot be directly used to test for a constant memory.

The TVLW estimator is based on a specification of $d(u)$ as a linear combination of Chebyshev's polynomials. We conjecture that some other orthonormal functions can also be used, but a comparison with other specifications is beyond the scope of this article and left to future research. An analysis of the robustness of the TVLW estimator to a wrong specification of $d(u)$ is also interesting but it requires a thorough theoretical and Monte Carlo analysis to assess the impact of such misspecification, and is also left to future research.

ACKNOWLEDGEMENTS

Josu Arteche research was funded by the Spanish Ministry of Science and Innovation and ERDF grant PID2019-105183GB-I00, Ministry of Science, Innovation and Universities grant PID2023-150480NB-I00 and UPV/EHU Econometrics Research Group (Basque Government grant IT1359-19). We thank Ricardo Olea for some help with the library LSTS. Luis F. Martins research was funded by the FCT – Fundacao para a Ciencia e a Tecnologia (National Science and Technology Foundation) through grant UIDB/00315/2020.

CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this article can be freely obtained from the OECD website https://data.oecd.org/price/inflation-cpi.htm.

REFERENCES

Arteche J. 2007. The analysis of seasonal long memory: the case of Spanish inflation. *Oxford Bulletin of Economics and Statistics* **69**:749–772.

Arteche J, Orbe J. 2016. A bootstrap approximation for the distribution of the Local Whittle estimator. *Computational Statistics and Data Analysis* **100**:645–660.

Arteche J, Robinson PM. 2000. Semiparametric inference in seasonal and cyclical long memory processes. *Journal of Time Series Analysis* **21**:1–27.

Beran J. 2009. On parameter estimation for locally stationary long-memory processes. *Journal of Statistical Planning and Inference* **139**:900–915.

Bierens HJ, Martins LF. 2010. Time varying cointegration. *Econometric Theory* **26**:1453–1490.

Bisaglia L, Grigoletto M. 2021. A new time-varying model for forecasting long-memory series. *Statistical Methods & Applications* **30**:139–155.

Boubaker H. 2018. A Generalized ARFIMA model with smooth transition fractional integration parameter. *Journal of Time Series Econometrics* **10**:1–20.

Boutahar M, Dufrenot G, Peguin-Feissolle A. 2008. A simple fractionally integrated model with a time varying long memory parameter. *Computational Economics* **31**:225–241.

Chan NH, Palma W. 2020. On the estimation of locally stationary long-memory processes. *Statistica Sinica* **30**:111–134.

Dahlhaus R. 1996a. On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications* **62**:139–168.

Dahlhaus R. 1996b. *Asymptotic statistical inference for nonstationary processes with evolutionary spectra*. In *Athens Conference on Applied Probability and Time Series, Vol. II. Lecture Notes in Statistics*, Robinson PM, Rosenblatt M (eds.). Springer-Verlag, New York; 145–159.

Dahlhaus R. 1997. Fitting time series models to non-stationary processes. *The Annals of Statistics* **25**:1–37.

Dahlhaus R, Giraitis L. 1998. On the optimal segment length for parameter estimates for locally stationary time series. *Journal of Time Series Analysis* **19**:629–655.

Davidson J, Hashimzade N. 2009. Type I and type II fractional Brownian motions: a reconsideration. *Computational Statistics and Data Analysis* **53**:2089–2106.

Fuhrer J, Moore G. 1995. Inflation persistence. *Quarterly Journal of Economics* **110**:127–159.

Gradshteyn IS, Ryzhik IM. 1994. *Table of Integrals, Series and Products*, 5th ed. Academic Press, San Diego.
Haldrup N, Nielsen MO. 2006. A regime switching long memory model for electricity prices. *Journal of Econometrics* **135**:349–376.
Henry M. 2001. Robust automatic bandwidth for long memory. *Journal of Time Series Analysis* **22**:293–316.
Hurvich CM, Moulines E, Soulier P. 2005. Estimating long memory in volatility. *Econometrica* **73**:1283–1328.
Jensen MJ, Whitcher B. 2000. Time-varying long memory in volatility: detection and estimation with wavelets. Technical Report, EURANDOM.
Johansen S, Nielsen MO. 2016. The role of initial values in conditional sum-of-squares estimation of nonstationary fractional time series models. *Econometric Theory* **32**:1095–1139.
Marinucci D, Robinson PM. 1999. Alternative forms of fractional Brownian motion. *Journal of Statistical Inference and Planning* **80**:111–122.
Martins LF, Rodrigues PMM. 2014. Testing for persistence change in fractionally integrated models: an application to world inflation rates. *Computational Statistics and Data Analysis* **76**:502–522.
Palma W. 2007. *Long-Memory Time Series: Theory and Methods*. Wiley Series in Probability and Statistics Wiley, United Kingdom.
Palma W, Olea R. 2010. An efficient estimator for locally stationary Gaussian long-memory processes. *The Annals of Statistics* **38**:2958–2997.
Ray BK, Tsay RS. 2002. Bayesian methods for change-point detection in long-range dependent processes. *Journal of Time Series Analysis* **23**:687–705.
Robinson PM. 1995a. Log-periodogram regression of time series with long-range dependence. *Annals of Statistics* **23**:1048–1072.
Robinson PM. 1995b. Gaussian semiparametric estimation of long-range dependence. *Annals of Statistics* **23**:1630–1661.
Roueff F, von Sachs R. 2011. Locally stationary long memory estimation. *Stochastic Processes and their Applications* **121**:813–844.
Shimotsu K. 2007. Gaussian semiparametric estimation of multivariate fractionally integrated processes. *Journal of Econometrics* **137**:277–310.
Song L, Bondon P. 2012. Piecewise FARIMA models for long-memory time series. *Journal of Statistical Computation and Simulation* **82**:1367–1382.
Velasco C. 1999a. Non-stationary log-periodogram regression. *Journal of Econometrics* **91**:325–371.
Velasco C. 1999b. Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* **20**:87–127.
Wang L. 2019. Time varying long memory parameter estimation for locally stationary long memory processes. *Communications in Statistics – Theory and Methods* **48**:2529–2547.

APPENDIX A: PROOFS OF THE LEMMAS AND THEOREMS

In all the proofs κ denotes a positive constant, $0 < \kappa < \infty$, possibly different in every situation.

Proof of Lemma 1. Denote $P_j = (P_{0,T}(u_j), P_{1,T}(u_j), \dots, P_{P,T}(u_j))'$, such that $d_P(u_j, \xi) = \xi' P_j$. The first bullet is then obvious. For the second one note that $d_P(u_j, \xi^1) = d_P(u_j, \xi^2)$ implies $(\xi^1 - \xi^2)' P_j = 0$ and taking sums for $j = 1, \dots, M$, $(\xi^1 - \xi^2)' M^{-1} \sum_{j=1}^{M} P_j P_j' (\xi^1 - \xi^2) = 0$ such that the statement is proved if $M^{-1} \sum_{j=1}^{M} P_{v,T}(u_j) P_{k,T}(u_j) \to \mathbb{1}(v = k)$ as $M \to \infty$.

To prove this orthonormality property of the Chebyshev polynomials note that

$$P_{v,T}(u_j) P_{k,T}(u_j) = 2\cos\left(v\pi\left[\frac{t_j - 0.5}{T}\right]\right)\cos\left(k\pi\left[\frac{t_j - 0.5}{T}\right]\right)$$
$$= \cos\left((v-k)\pi\left[\frac{t_j - 0.5}{T}\right]\right) + \cos\left((v+k)\pi\left[\frac{t_j - 0.5}{T}\right]\right).$$

For $a = 0$, $\cos(a\pi(t_j - 0.5)/T)$ is obviously one, and for $a \neq 0$,

$$\frac{1}{M}\sum_{j=1}^{M}\cos\left(a\pi\left[\frac{t_j - 0.5}{T}\right]\right) = \cos\left(\frac{a\pi}{2T}(N - S - 1)\right)\frac{1}{M}\sum_{j=1}^{M}\cos\left(\frac{a\pi S}{2T}(2j-1)\right), \tag{A1}$$

$$- \sin\left(\frac{a\pi}{2T}(N - S - 1)\right)\frac{1}{M}\sum_{j=1}^{M}\sin\left(\frac{a\pi S}{2T}(2j-1)\right), \tag{A2}$$

noting that $t_j = S(j-1) + N/2$. Now, under Assumption A.4, as $T \to \infty$,

$$\cos\left(\frac{a\pi}{2T}(N-S-1)\right) \to 1,$$

$$\sin\left(\frac{a\pi}{2T}(N-S-1)\right) \sim \frac{N}{MS} \to 0,$$

$$\frac{1}{M}\sum_{j=1}^{M}\cos\left(\frac{a\pi S}{2T}(2j-1)\right) = \frac{\sin\left(\frac{Ma\pi S}{T}\right)}{2M\sin\left(\frac{a\pi S}{2T}\right)}$$

$$\frac{1}{M}\sum_{j=1}^{M}\sin\left(\frac{a\pi S}{2T}(2j-1)\right) = \frac{\sin^2\left(\frac{Ma\pi S}{2T}\right)}{M\sin\left(\frac{a\pi S}{2T}\right)},$$

where $a \sim b$ means that $a/b \to \kappa$ and the last two equalities come from formulae 1.342.3 and 1.342.4 in Gradshteyn and Ryzhik (1994). Also note that under Assumption A.4, and because $T = S(M-1) + N$

$$\sin\left(\frac{a\pi S}{2T}\right) \sim \frac{1}{M}$$

$$\sin\left(\frac{Ma\pi S}{T}\right) = \sin(a\pi) + O\left(\frac{1}{M} + \frac{N}{MS}\right) = O\left(\frac{1}{M} + \frac{N}{MS}\right)$$

$$\sin\left(\frac{Ma\pi S}{2T}\right) \to \sin\left(\frac{a\pi}{2}\right) = (-1)^{(a-1)/2} I(a \text{ odd}),$$

as $T \to \infty$. Thus $M^{-1}\sum_{j=1}^{M} P_{v,T}(u_j) P_{k,T}(u_j) \to \mathbb{1}(v = k)$ as $T \to \infty$ (note that Assumption A.5 implies $M \to \infty$). ■

For simplicity of notation denote hereafter $d_j(\xi) = d_P(u_j, \xi)$ and $d_j(\xi^0) = d_P(u_j, \xi^0)$ and $\overline{v}$ is the complex conjugate of v.

Lemma 2. Let $v_{jk} = D_N(u_j, \lambda_k)/C_j^{1/2}\lambda_k^{-d_j(\xi^0)}$. Under Assumptions A.1, A.2 (B.2), A.4 and A.5, for any sequence of positive integers $k = k(N)$ and $l = l(N)$, $k > l$, as $T, N \to \infty$,

(a) $Ev_{jk}\overline{v}_{jk} = 1 + O\left(\frac{\log k}{k^{2\alpha_j}} + \frac{k^\beta}{N^\beta} + \frac{N\log N}{T}\right)$, with the second term arising under B.2,

(b) $Ev_{jk}v_{jk} = O\left(\frac{\log k}{k^{2\alpha_j}} + \frac{N\log N}{T}\right)$,

(c) $Ev_{jk}\overline{v}_{jl} = O\left(\frac{\log k}{l^{2\alpha_j}} + \frac{N\log N}{T}\right)$,

(d) $Ev_{jk}v_{jk} = O\left(\frac{\log k}{l^{2\alpha_j}} + \frac{N\log N}{T}\right)$,

where $\alpha_j = \min(1/2, 1 - d_j(\xi^0))$.

Proof. The proof uses the results in theorem 2 in Robinson (1995a) and theorem 1 in Velasco (1999a). The only differences come from the TV character of our models. Thus, we only focus here on these differences, appealing to Robinson (1995a) and Velasco (1999a) for the rest of the proof. Consider for example the proof of (a) for the stationary case. Note first that under Assumption B.2, $f_v(u_j, \lambda_l) - C_j\lambda_l^{-2d_j(\xi^0)} = O(\lambda_l^{\beta-2d_j(\xi^0)})$ and under A.2 is $o(\lambda_l^{-2d_j(\xi^0)})$. Also, using Assumption A.1 $E\left(D_N(u_j, \lambda_k)\overline{D_N(u_j, \lambda_k)}\right)$ is equal to

$$\frac{1}{2\pi N}\sum_{t=1}^{N}\sum_{r=1}^{N}E(x_{[u_jT]-N/2+t,T}x_{[u_jT]-N/2+r,T})e^{-i\lambda_k(t-r)}$$
$$=\sum_{t=1}^{N}\sum_{r=1}^{N}\int_{-\pi}^{\pi}B\Big(u_j-\frac{N}{2T}+\frac{t}{T},\lambda\Big)B\Big(u_j-\frac{N}{2T}+\frac{r}{T},-\lambda\Big)\frac{1}{2\pi N}e^{i(\lambda-\lambda_k)(t-r)}\mathrm{d}\lambda$$
$$=\int_{-\pi}^{\pi}\left[B\big(u_j,\lambda\big)B\big(u_j,-\lambda\big)+O\left(\frac{N\log|\lambda|}{T}|\lambda|^{-2d_P(\bar{u},\xi^0)}\right)\right]K(\lambda-\lambda_k)\mathrm{d}\lambda$$
$$=\int_{-\pi}^{\pi}\left[f_v(u_j,\lambda)+O\left(\frac{N\log|\lambda|}{T}|\lambda|^{-2d_P(\bar{u},\xi^0)}\right)\right]K(\lambda-\lambda_k)\mathrm{d}\lambda,$$

for $\bar{u}$ in the interval $\left\{u_j-\frac{N}{2T},u_j+\frac{N}{2T}\right\}$ where $K(\lambda)=(2\pi N)^{-1}\sum_{t=1}^{N}\sum_{r=1}^{N}e^{-i(\lambda)(t-r)}$ is Fejer's kernel. Then

$$E\Big(D_N(u_j,\lambda_k)\overline{D_N(u_j,\lambda_k)}\Big)-f_v(u_j,\lambda_k)$$
$$=\int_{-\pi}^{\pi}\big(f_v(u_j,\lambda)-f_v(u_j,\lambda_k)\big)K(\lambda-\lambda_k)\mathrm{d}\lambda$$
$$+O\left(\int_{-\pi}^{\pi}\frac{N\log|\lambda|}{T}\left[|\lambda|^{-2d_P(\bar{u},\xi^0)}-|\lambda_k|^{-2d_P(\bar{u},\xi^0)}+|\lambda_k|^{-2d_P(\bar{u},\xi^0)}\right]\right)$$
$$=O\left(\lambda_k^{-2d_j(\xi^0)}\left[\frac{\log k}{k}+\frac{N\log N}{T}\right]\right),$$

where the final bounds are obtained by splitting the integral as in Robinson (1995a, proof of theorem 2a) with a different treatment in frequencies close to and far from zero, and using the properties that the integral of $K(\lambda)$ over $(-\pi,\pi)$ is 1, $K(\lambda)=O(N^{-1}|\lambda|^{-2})$ and $|\lambda_k|^{-d_P(\bar{u},\xi^0)}=|\lambda_k|^{-d_j(\xi^0)}(1+O(\frac{N}{T}\log N))$ under Assumption A.5 for $\bar{u}$ in the interval $\{u_j-\frac{N}{2T},u_j+\frac{N}{2T}\}$. This proves the result in a) for the stationary case. The non-stationary case can be proved similarly using the results in the proof of theorem 1 in Velasco (1999a). The steps needed for the proof of (b), (c), and (d) are similar and are thus omitted. ■

Proof of Theorem 1. We need to prove that $\|\hat{\xi}-\xi^0\|=o_p(1)$, where $\|\cdot\|$ denotes the supremum norm. Note first that $R_{T,j}(\xi)=R_j(\xi)+\log(\hat{C}_j(\xi)/C_j(\xi))$ where

$$R_j(\xi)=\log(C_j(\xi))-\frac{2}{m}d_j(\xi)\sum_{k=1}^{m}\log\lambda_k$$
$$C_j(\xi)=\frac{C_j}{m}\sum_{k=1}^{m}\lambda_k^{2\left(d_j(\xi)-d_j(\xi^0)\right)}.$$

By concavity of the log function for $\xi\in\Theta$,

$$R_j(\xi)\geq\log C_j+\frac{1}{m}\sum_{k=1}^{m}\log\left(\lambda_k^{2(d_j(\xi)-d_j(\xi^0))}\right)-\frac{2}{m}d_j(\xi)\sum_{k=1}^{m}\log\lambda_k=R_j(\xi^0),$$

and thus every $R_j(\xi)$ achieves a minimum at ξ^0. Define now $R(\xi) = M^{-1}\sum_{j=1}^{M} R_j(\xi)$ such that

$$
\begin{aligned}
0 \le R(\hat{\xi}) - R(\xi^0) &= R(\hat{\xi}) - R_T(\hat{\xi}) + R_T(\hat{\xi}) - R_T(\xi^0) + R_T(\xi^0) - R(\xi^0) \\
&\le \frac{1}{M}\sum_{j=1}^{M}\left\{ \log\left(\frac{\hat{C}_j(\xi^0)}{C_j(\xi^0)}\right) - \log\left(\frac{\hat{C}_j(\hat{\xi})}{C_j(\hat{\xi})}\right)\right\},
\end{aligned}
$$

because $R_T(\hat{\xi}) - R_T(\xi^0) \le 0$. Define now the $R^{P+1} \to R$ function $K_j(s)$ as

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

Denote $K_j^i(s) = \partial K_j(s)/\partial s_i$, $K_j^{il}(s) = \partial K_j(s)/\partial s_i \partial s_l$, $K_j'(s)$ is the vector of first derivatives $[K_j'(s)]_i = K_j^i(s)$ and $K_j''(s)$ is the matrix of second derivatives $[K_j''(s)]_{il} = K_j^{il}(s)$. Since $K_j(0) = K_j^i(0) = 0$ for $i = 0, 1, \ldots, P$ and considering the continuity and twice differentiability of $K_j(s)$, the mean value theorem gives for $s \in \Theta$,

$$
K_j(s) = \frac{1}{2}s'K_j''(c)s,
$$

for some $\|c\| \le \|s\|$. The $i \times l$ element of the matrix $K_j''(c)$ is of the form $K_j^{il}(c) = 4P_i(u_j)P_l(u_j)c_j(c)$ for

$$
c_j(c) = \frac{\frac{1}{m}\sum_k k^{2c'P_j}\log^2 k \frac{1}{m}\sum_k k^{2c'P_j} - \left(\frac{1}{m}\sum_k k^{2c'P_j}\log k\right)^2}{\left(\frac{1}{m}\sum_k k^{2c'P_j}\right)^2}
$$

such that $c_j(c) > 0$ for $j = 1, \ldots, M$ by Cauchy–Schwarz inequality. Then for $j = 1, \ldots, M$

$$
\begin{aligned}
K(s) &\ge \frac{2}{M}\sum_{j=1}^{M} s'P_jP_j's\,c_j(c) \\
&\ge \min_{(c,j)}\{c_j(c)\}\, s'\frac{2}{M}\sum_{j=1}^{M}P_jP_j's \\
&\ge \kappa s's,
\end{aligned}
$$

for a positive generic constant κ (which in what follows can be different in each situation) and large enough T, since under Assumption A.5 $\frac{1}{M}\sum_{j=1}^{M}P_jP_j' \to I$ (see Lemma 1). Replacing s by $(\hat{\xi} - \xi^0)$, then,

$$
\begin{aligned}
\sum_{v=0}^{P}(\hat{\xi}_i - \xi_i^0)^2 &\le \kappa K(\hat{\xi} - \xi^0) = \kappa(R(\hat{\xi}) - R(\xi^0)) \\
&\le \kappa\frac{1}{M}\sum_{j=1}^{M}\left\{ \log\left(\frac{\hat{C}_j(\xi^0)}{C_j(\xi^0)}\right) - \log\left(\frac{\hat{C}_j(\hat{\xi})}{C_j(\hat{\xi})}\right)\right\} \\
&\le \kappa\frac{1}{M}\sum_{j=1}^{M}\sup_{\xi\in\Theta}\left|\log\left(\frac{\hat{C}_j(\xi)}{C_j(\xi)}\right)\right|,
\end{aligned}
$$

DOI: 10.1111/jtsa.12782 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

and the consistency is thus proved if $\sup_\Theta |(\hat{C}_j(\xi) - C_j(\xi))/C_j(\xi)| = o_p(1)$ for every $j = 1, \ldots, M$. Now

$$\frac{\hat{C}_j(\xi) - C_j(\xi)}{C_j(\xi)} = \sum_{k=1}^{m} c_{jk}\left(\frac{I_{jk}}{g_{jk}} - 1\right),$$

where $I_{j,k} = I_N(u_j, \lambda_k)$, $g_{jk} = C_j \lambda_k^{-2d_j(\xi^0)}$ and

$$c_{jk} = \frac{1}{m}\frac{\lambda_k^{2(d_j(\xi)-d_j(\xi^0))}}{\frac{1}{m}\sum_{k=1}\lambda_k^{2(d_j(\xi)-d_j(\xi^0))}} = \frac{1}{m}\frac{k^{2(d_j(\xi)-d_j(\xi^0))}}{\frac{1}{m}\sum_{k=1}k^{2(d_j(\xi)-d_j(\xi^0))}}.$$

The proof now follows as in Robinson (1995b) and Velasco (1999b) adapting some steps to the locally stationary and non-stationary character of the processes we deal with. Note that

$$\sum_{k=1}^{m} c_{jk}\left(\frac{I_{jk}}{g_{jk}} - 1\right) = \sum_{k=1}^{m} c_{jk}(r_{jk} + s_{jk}), \tag{A3}$$

$$+ \sum_{k=1}^{m} c_{jk}\frac{1}{N}\sum_{s=0}^{N-1}(\varepsilon^2_{t_j-N/2+s+1} - 1), \tag{A4}$$

for $r_{jk} = I_{jk}/g_{jk} - 2\pi I^{\varepsilon}_{jk}$, $I^{\varepsilon}_{jk} = I^{\varepsilon}_N(u_j, \lambda_k)$ the periodogram of the innovations in Assumption A.1 and $s_{jk} = N^{-1}\sum_s \sum_{t\neq s}^{N-1} \varepsilon_{t_j-N/2+s+1}\varepsilon_{t_j-N/2+t+1}e^{i(t-s)\lambda_k}$. By summation by parts (A3) is equal to

$$\sum_{k=1}^{m-1}(c_{jk} - c_{jk+1})\sum_{l=1}^{k}(r_{jl} + s_{jl}) + c_{jm}\sum_{k=1}^{m}(r_{jk} + s_{jk}) = Z_1(\xi) + Z_2(\xi).$$

To get a bound for $E|r_{jl}|$ note that

$$r_{jl} = \left(1 - \frac{g_{jl}}{f_{jl}}\right)\frac{I_{jl}}{g_{jl}} + \left(\frac{I_{jl}}{f_{jl}} - 2\pi I^{\varepsilon}_{jl}\right),$$

for $f_{jl} = f_x(u_j, \lambda_l)$. Now we need the following results

$$E\left|\frac{I_{jl}}{g_{jl}}\right| \le \kappa, l = 1, 2, \ldots, m, \tag{A5}$$

$$\left|1 - \frac{g_{jl}}{f_{jl}}\right| = o(1), l = 1, 2, \ldots, m, \tag{A6}$$

$$E\left|\frac{I_{jl}}{f_{jl}} - 2\pi I^{\varepsilon}_{jl}\right| = O\left(\frac{\log^{1/2} l}{l^{\alpha_j}} + \frac{N^{1/2}\log^{1/2} N}{T^{1/2}}\right), \tag{A7}$$

for a generic positive constant $0 < \kappa < \infty$ and $\alpha_j = \min(1/2, 1 - d_j(\xi^0))$. (A6) is easily obtained from Assumption A.2. (A5) can be deduced from Lemma 2. Finally, the result in (A7) can be similarly proved using formulae (3.17) in Robinson (1995b) and (A2) in Velasco (1999b) together with Lemma 2.

These bounds lead to $\sum_{l=1}^{k} E|r_{jl}| = o(k) + O(k^{1-\alpha_j}\log^{1/2}k)$. Note also that $c_{jk} - c_{jk+1} = c_{jk}O(k^{-1})$ and thus

$$
\begin{aligned}
&\sum_{k=1}^{m-1} |c_{jk} - c_{jk+1}| \sum_{l=1}^{k} E|r_{jl}| \\
&= \sum_{k=1}^{m-1} c_{jk} o(1) + \sum_{k=1}^{m-1} c_{jk} O(k^{-\alpha_j}\log^{1/2}k) \\
&= o(1) + O\left(\frac{\frac{\log^{1/2}m}{m} \sum_k k^{2(d_j(\xi)-d_j(\xi^0))-\alpha_j}}{\frac{1}{m}\sum_k k^{2(d_j(\xi)-d_j(\xi^0))}} \right) \\
&= o(1) + O\left(\frac{\log^{1/2}m}{m} \frac{\left[m^{2(d_j(\xi)-d_j(\xi^0))-\alpha_j+1} + \log m \right]}{m^{2(d_j(\xi)-d_j(\xi^0))}} \right) \\
&= o(1),
\end{aligned}
$$

because $2(d_j - d_j^0) + 1 > 0$ under Assumption A.3 and $\alpha_j > 0$. Now since $E[(\sum_{l=1}^{k} s_{jl})^2] = O(k)$ (see Robinson, 1995b, formula (3.20), or Hurvich *et al.*, 2005, p. 1325), proceeding similarly as before

$$
\begin{aligned}
&\sup_{\Theta} \sum_{k=1}^{m-1} (c_{jk} - c_{jk+1}) \sum_{l=1}^{k} s_{jl} = O_p\left(\sup_{\Theta} \sum_{k=1}^{m-1} |c_{jk}| k^{-1/2} \right) \\
&= O_p\left(\sup_{\Theta} \frac{\sum_k k^{2(d_j(\xi)-d_j(\xi^0))-1/2}}{\sum_k k^{2(d_j(\xi)-d_j(\xi^0))}} \right) = O_p\left(\sup_{\Theta} \frac{m^{2(d_j(\xi)-d_j(\xi^0))+1/2} + 1}{m^{2(d_j(\xi)-d_j(\xi^0))}} \right) = o_p(1),
\end{aligned}
$$

and thus $\sup_{\Theta} Z_1(\xi) = o_p(1)$. Proceeding similarly we also get $\sup_{\Theta} Z_2(\xi) = o_p(1)$. Finally, (A4) is clearly $o_p(1)$ under Assumption A.6 (see formula (3.19) in Robinson, 1995b). ■

Proof of Theorem 2. Note that

$$
(\hat{\xi} - \xi^0) = \left(\left. \frac{\mathrm{d}^2 R_T(\xi)}{\mathrm{d}\xi \mathrm{d}\xi'} \right|_{\xi=\bar{\xi}} \right)^{-1} \left. \frac{\mathrm{d}R_T(\xi)}{\mathrm{d}\xi} \right|_{\xi=\xi^0},
$$

for $\bar{\xi}$ such that $\|\bar{\xi} - \xi^0\| \leq \|\hat{\xi} - \xi^0\|$. The theorem is thus proven if

(A) Hessian convergence:

$$
\left. \frac{\mathrm{d}^2 R_T(\xi)}{\mathrm{d}\xi \mathrm{d}\xi'} \right|_{\xi=\bar{\xi}} \xrightarrow{p} 4I_{P+1}.
$$

(B) Score convergence:

$$
\sqrt{mM} \left. \frac{\mathrm{d}R_T(\xi)}{\mathrm{d}\xi} \right|_{\xi=\xi^0} \xrightarrow{d} \mathcal{N}(0, 4I_{P+1}).
$$

B) Hessian convergence: Fix $\delta > 0$ and let $M = \{\xi : \log^4 N \|\xi - \xi^0\| \leq \delta\}$. We first show that $P(\bar{\xi} \notin M) \to 0$. Proceeding as in the proof of consistency, $P(\log^4 N \sum_{v=0}^{P} (\hat{\xi} - \xi^0)^2 > \delta) \to 0$ if $\sup_{\Theta} |(\hat{C}_j(\xi) - C_j(\xi))/$

DOI: 10.1111/jtsa.12782
Journal of Time Series Analysis published by John Wiley & Sons Ltd.

$C_j(\xi)| = o_p(\log^{-4}N)$. Noting now that $|1 - g_{jl}^{-1}f_{jl}| = O(\lambda_j^{\beta})$ then $\sum_{l=1}^{k} E|r_{jl}| = O(k^{\beta+1}N^{\beta} + k^{1-\alpha_j}\log^{1/2}k + kN^{1/2}\log^{1/2}NT^{-1/2})$ and proceeding as in the proof of consistency, $\sup_{\Theta} Z_1(\xi)$ is

$$O_p\left(\left[\frac{m}{N}\right]^{\beta} + \frac{\log^{3/2}m}{m^{\delta}} + \frac{N^{1/2}\log^{3/2}m}{T^{1/2}}\right) = o_p(\log^{-4}N),$$

for some $\delta > 0$ under Assumptions B.3–B.5. Similarly, $\sup_{\Theta} Z_2(\xi)$ is

$$O_p\left(\left[\frac{m}{N}\right]^{\beta} + \frac{\log^{1/2}m}{m^{\alpha^*}} + \frac{N^{1/2}\log^{3/2}m}{T^{1/2}}\right) = o_p(\log^{-4}N),$$

under B.5, for $\alpha^* = \min_j \alpha_j$. Finally, (A4) has mean zero and variance $O(N^{-1})$ under Assumption B.5 and $S \geq N$. Therefore, we can restrict to values $\overline{d} \in M$.

Now the matrix of second derivatives has (i, h)th element

$$\left[\frac{\mathrm{d}^2 R_T(\xi)}{\mathrm{d}\xi \mathrm{d}\xi'}\right]_{ih} = \frac{4}{M}\sum_{j=1}^{M} \frac{C_{j2}^{i-1,h-1}(\xi)\hat{C}_j(\xi) - C_{j1}^{i-1,0}(\xi)C_{j1}^{0,h-1}(\xi)}{\hat{C}_j(\xi)^2},$$

where for $a = 1, 2$,

$$C_{ja}^{ih}(\xi) = \frac{1}{m}\sum_{k=1}^{m} \frac{I_{j,k}}{\lambda_k^{-2d_j(\xi)}} \log^a \lambda_k P_{i,T}(u_j)P_{h,T}(u_j).$$

Since $|\lambda_k^{2(d_j(\xi)-d_j(\xi^0))} - 1| \leq 2|d_j(\xi) - d_j(\xi^0)| \log \lambda_k \max \lambda_k^{2(d_j(\xi)-d_j(\xi^0))} = O_p(|d_j(\xi) - d_j(\xi^0)|N^{1/logN} \log N) = O_p(\log^{-3}N)$ in M, then $|C_{ja}^{ih}(\overline{\xi}) - C_{ja}^{ih}(\overline{\xi^0})| = O_p(\log^{a-3}N)$. Denote

$$J_{ja}^{ih} = \frac{C_j}{m}\sum_{k=1}^{m} \log^a \lambda_k P_{i,T}(u_j)P_{h,T}(u_j).$$

Then

$$\begin{aligned}
|C_{ja}^{ih}(\xi^0) - J_{ja}^{ih}| &\leq \frac{C_j}{m}\sum_{k=1}^{m} \log^a \lambda_k P_{i,T}(u_j)P_{h,T}(u_j)\left|\frac{I_{j,k}}{g_{jk}} - 1\right| \\
&\leq \frac{C_j}{m}\sum_{k=1}^{m} \log^a \lambda_k P_{i,T}(u_j)P_{h,T}(u_j)\left|r_{jk} + s_{jk} + \frac{1}{N}\sum_{s=0}^{N-1}(\varepsilon_{t_j-N/2+s+1}^2 - 1)\right| \\
&= O_p\left(\left[\frac{m}{N}\right]^{\beta}\log^a N + \frac{\log^a N \log^{1/2} m}{m^{\alpha^*}} + \frac{\log^a N}{N^{1/2}} + \frac{N^{1/2}\log^{1/2}}{T^{1/2}}\right) \\
&= O_p(\log^{a-3}N),
\end{aligned}$$

under Assumption B.5. Then

$$\left[\left.\frac{\mathrm{d}^2 R_T(\xi)}{\mathrm{d}\xi \mathrm{d}\xi'}\right|_{\xi=\overline{\xi}}\right]_{ih} = \frac{4}{M}\sum_{j=1}^{M} \frac{C_{j2}^{i-1,h-1}(\xi^0)\hat{C}_j(\xi^0) - C_{j1}^{i-1,0}(\xi^0)C_{j1}^{h-1,0}(\xi^0) + O_p(\log^{-1}N)}{\hat{C}_j(\xi^0)^2 + O_p(\log^{-3}N)}$$

$$= \frac{4}{M}\sum_{j=1}^{M}\frac{J_{j2}^{i-1,h-1}C_j - J_{j1}^{i-1,0}J_{j1}^{h-1,0} + O_p(\log^{-1}N)}{C_j^2 + O_p(\log^{-3}N)}$$
$$= \frac{4}{M}\sum_{j=1}^{M}P_{i-1,T}(u_j)P_{h-1,T}(u_j) + o(1) + o_p(1)$$
$$= 4\mathbb{1}(i = h) + o(1) + o_p(1),$$

because $J_{j2}^{ih}C_j - J_{j1}^{i0}J_{j1}^{h0}$ is

$$\frac{C_j^2}{m}P_{i,T}(u_j)P_{h,T}(u_j)\sum_{k=1}^{m}v_k^2 = C_j^2P_{i,T}(u_j)P_{h,T} + o(1).$$

(B) Score convergence: The result follows if for any vector $\eta = (\eta_0, \dots, \eta_P)'$

$$\sqrt{mM}\eta' \left.\frac{\mathrm{d}R_T(\xi)}{\mathrm{d}\xi}\right|_{\xi=\xi^0} \xrightarrow{d} \mathcal{N}(0, 4\eta'\eta). \tag{A8}$$

Noting that $\hat{C}_j(\xi^0) \xrightarrow{p} C_j$, the left-hand side of (A8) is asymptotically equivalent to

$$\frac{2}{\sqrt{M}}\sum_{j=1}^{M}\sum_{v=0}^{P}\eta_v P_{v,T}(u_j)Z_j \text{ for } Z_j = \frac{1}{\sqrt{m}}\sum_{k=1}^{m}v_k\frac{I_{jk}}{g_{jk}}.$$

Proceeding as in Robinson (1995, theorem 2) and Velasco (1999, theorem 3) we get that for every $j = 1, .., M$, $Z_j \xrightarrow{d} \mathcal{N}(0, 1)$. Then, taking into account the asymptotic independence of Z_j and Z_l, $l \neq j$ (guaranteed by the fact that $S \geq N$) and that $M^{-1}\sum_{j=1}^{M}P_jP_j' \to I_{P+1}$ we get the desired result. ■

Proof of Corollary 1. The proof of a) comes directly from Theorem 2 taking into account that $A_{Mm} \to 4I_{P+1}$ as $T \to \infty$. To prove (b) and (c) note that under the assumptions in Theorem 2 and H_1, $\sqrt{4mM}(R\hat{\xi} - r - \overline{r}) \xrightarrow{d,H_1} \mathcal{N}_q(0, R'R)$ or

$$\sqrt{4mM}(R\hat{\xi} - r) \xrightarrow{d,H_1} \mathcal{N}_q(\overline{r}\sqrt{4mM}, R'R),$$

such that $W(R, r) \xrightarrow{p,H_1} \infty$ under the constant specification of $\overline{r}$ in b) and $W(R, r) \xrightarrow{d,H_1} \chi_q^2(4\sum_{i=1}^{q}\theta_i^2)$ under the local specification in (c). ■

DOI: 10.1111/jtsa.12782
Journal of Time Series Analysis published by John Wiley & Sons Ltd.