SUPPLEMENT TO "AN EFFICIENT ESTIMATOR FOR LOCALLY STATIONARY GAUSSIAN LONG-MEMORY PROCESSES"

By Wilfredo Palma and Ricardo Olea

Department of Statistics
Pontificia Universidad Católica de Chile

This document provides supplementary material related to the paper *An Efficient Estimator for Locally Stationary Gaussian Long-Memory Processes [Palma and Olea (2010)]*. It contains additional examples of calculations of the asymptotic variance of the Whittle estimates, Monte Carlo simulations, a comparison with a kernel maximum likelihood estimation methodology, and two real life applications. Furthermore, it provides proofs of the technical lemmas appearing in that article.

1. Introduction. In what follows, the definitions of models, parameters, covariance matrices, etc., corresponds to those in the manuscript by Palma and Olea (2010). Similarly, theorem numbers refer to those in that paper. This document is organized as follows. Section 2 contains additional illustrations of the calculation of the asymptotic variance of the Whittle estimates. In particular, it provides explicit expressions for the matrix Γ appearing in Theorem 2.2 of Palma and Olea (2010). Section 3 analyzes the selection of the tuning parameters N and S from an empirical computational perspective. Theoretical optimal selection on these values is not available yet, and it is a fundamental topic of future research. Several additional Monte Carlo experiments are discussed in Section 4. In this section we study the finite sample performance of the proposed Whittle estimates and compare its performance to a recently proposed kernel maximum likelihood estimator. Section 5 provides applications of the proposed methodology to two real life time series. Finally, proofs of the lemmas appearing in Palma and Olea (2010) are provided in the Appendix.

2. Asymptotic Variance. Two additional examples illustrating the calculation of the asymptotic variance of the Whittle estimators are provided in this section.

AMS 2000 subject classifications: Primary 62M10; secondary 60G15

Keywords and phrases: Nonstationarity, local stationarity, long-range dependence, Whittle estimation, consistency, asymptotic normality, efficiency

EXAMPLE 2.1. (*Linear trend model*) Consider the LSFN model

$$Y_{t,T} = \sigma\left(\tfrac{t}{T}\right)(1-B)^{-d\left(\frac{t}{T}\right)}\varepsilon_t = \sigma\left(\tfrac{t}{T}\right)\sum_{j=0}^{\infty}\eta_j\left(\tfrac{t}{T}\right)\varepsilon_{t-j}, \tag{1}$$

for $t = 1, \ldots, T$, $\{\varepsilon_t\}$ is a Gaussian white noise sequence with zero mean and unit variance. Assume that the long-memory parameter $d(u)$ of model (1) is specified by a linear trend and $\sigma(u)$ is constant over time,

$$d(u) = \alpha_0 + \alpha_1\, u, \qquad \sigma(u) = \beta_0, \tag{2}$$

for $u \in [0,1]$. In this case the parameter vector is $(\alpha_0, \alpha_1, \beta_0)$ and the matrix Γ can be written as in Eq. (13) of Palma and Olea (2010) with

$$\Gamma_\alpha = \frac{\pi^2}{36}\begin{pmatrix} 6 & 3 \\ 3 & 2 \end{pmatrix}, \qquad \Gamma_\beta = \frac{2}{\beta_0^2}.$$

Thus, the asymptotic variance of $\widehat{\alpha} = (\widehat{\alpha}_0, \widehat{\alpha}_1)$ can be written as

$$\mathrm{Var}(\widehat{\alpha}) \sim \frac{12}{\pi^2 T}\begin{pmatrix} 2 & -3 \\ -3 & 6 \end{pmatrix},$$

and the asymptotic variance of $\widehat{d}(u) = \widehat{\alpha}_0 + \widehat{\alpha}_1 u$ is given by

$$V(u) = \mathrm{Var}(\widehat{\alpha}_0) + 2\,\mathrm{cov}(\widehat{\alpha}_0, \widehat{\alpha}_1)u + \mathrm{Var}(\widehat{\alpha}_1)u^2 \sim \frac{24}{T\pi^2}(1 - 3\,u + 3\,u^2). \tag{3}$$

Figure 1 displays this function for $T = 1$. Note that this variance is minimal at $u = 0.5$ and maximal at $u = 0\,\text{or}\,1$.

EXAMPLE 2.2. (*Cubic trend model*) Consider the LSFN process (1) with long-memory parameter satisfying a cubic trend model and constant variance,

$$d(u) = \tau_0 + \tau_1\, u^3, \qquad \sigma(u) = \beta_0, \tag{4}$$

where $0 < d(u) < \frac{1}{2}$. Now the parameter vector is $(\tau_0, \tau_1, \beta_0)$ and the matrix Γ can be written as in Eq. (13) of Palma and Olea (2010) with

$$\Gamma_\tau = \frac{\pi^2}{168}\begin{pmatrix} 28 & 7 \\ 7 & 4 \end{pmatrix}, \qquad \Gamma_\beta = \frac{2}{\beta_0^2}.$$

Therefore, the asymptotic variance of the estimator $\widehat{\tau} = (\widehat{\tau}_0, \widehat{\tau}_1)$ is given by

$$\mathrm{Var}(\widehat{\tau}) \sim \frac{8}{3T\pi^2}\begin{pmatrix} 4 & -7 \\ -7 & 28 \end{pmatrix}.$$

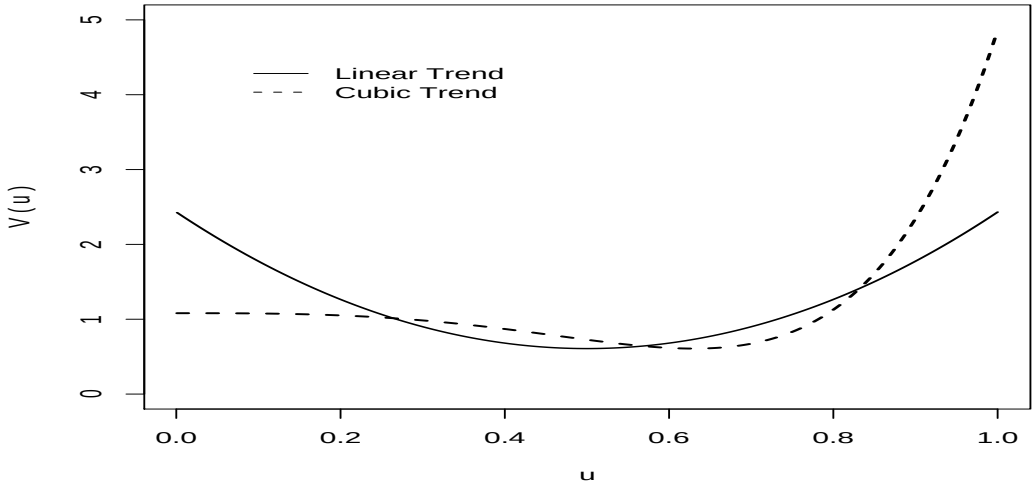

FIG 1. Asymptotic variance of $\widehat{d}(u)$: Linear trend model (heavy line) and Cubic trend model (broken line).

From this formula, the asymptotic variance of $\widehat{d}(u) = \widehat{\tau}_1 + \widehat{\tau}_1 u^3$ is given by

$$V(u) = \mathrm{Var}(\widehat{\tau}_0) + 2\,\mathrm{cov}(\widehat{\tau}_0, \widehat{\tau}_1)u^3 + \mathrm{Var}(\widehat{\tau}_1)u^6 \sim \frac{16}{3T\pi^2}(2 - 7\,u^3 + 14\,u^6). \tag{5}$$

This function is also depicted in Figure 1 for $T = 1$. In this cubic trend case the variance reaches its minimum value at $u = 4^{-1/3} \approx 0.63$ and its maximum at $u = 1$.

3. Bandwidth Selection. Observe that as a consequence of the central limit theorem established in Palma and Olea (2010), the asymptotic distribution of the estimator $\widehat{\theta}$ depends only on the sample size T and it does not depend explicitly on N, M and S, as long as these values satisfy Assumption A3. Nevertheless, in practice it is necessary to provide general rules for selecting these values. There an extensive literature about the bandwidth selection problem in different contexts, see for instance Abramson (1982), Silverman (1986), Stefanski (1990), Brockmann et al. (1993), Robinson (1994), Dahlhaus and Giraitis (1998), Beran and Feng (2002a,b) and Beran (2009), among others. In this section we analyze this problem for the proposed local Whittle estimation methodology. First, we investigate the numerical complexity of the proposed algorithm. Second, we explore the empirical optimal selection of N, M and S for finite samples through several numerical

experiments. However, obtaining theoretical optimality of the tuning parameters N and S requires further research.

3.1. *Computational Issues.* Observe that in order to evaluate the locally stationary Whittle log-likelihood, the FFT must be computed for every one of the M blocks of data, $I_N(u_1,\lambda),\ldots,I_N(u_M,\lambda)$. Given that the size of every block is N, the calculation of all these FFTs is order $\mathcal{O}(MN\log N)$ or equivalently, $\mathcal{O}\left(TNS^{-1}\log N\right)$. Consequently, large block sizes N as well as small shifts S increase the number of calculations.

3.2. *Empirical Optimality.* In order to investigate appropriate selection of N, M and S in a practical setting, we study in detail the following two illustrative examples. Note that other illustrations are discussed in Section 4 of Palma and Olea (2010). Consider fitting the linear and cubic trend models introduced in Examples 2.1 and 2.2, respectively, with sample sizes $T=500$ and $T=1,000$. To assess which values of N and S are optimal for these finite sample situations, we have run a number of simulations for combinations of $N=40,\ldots,400$ with steps $\Delta N=20$ and $S=6,\ldots,300$ with variable steps $\Delta S=2$ to $\Delta S=10$, such that we obtain a higher resolution near the optimal location. The samples of the LSFN processes are generated by using the innovation algorithm, see (Brockwell and Davis, 1991, p.172) with

$$\mathrm{E}[Y_{s,T}Y_{t,T}]=\sigma\left(\tfrac{s}{T}\right)\sigma\left(\tfrac{t}{T}\right)\frac{\Gamma\left[1-d\left(\frac{s}{T}\right)-d\left(\frac{t}{T}\right)\right]\Gamma\left[s-t+d\left(\frac{s}{T}\right)\right]}{\Gamma\left[1-d\left(\frac{s}{T}\right)\right]\Gamma\left[d\left(\frac{s}{T}\right)\right]\Gamma\left[s-t+1-d\left(\frac{t}{T}\right)\right]},$$

for $s,t=1,\ldots,T$, $s\geq t$. The performance of the Whittle estimates can be measured through the integrated mean squared error of $\widehat{d}(u)$ given by

$$\mathrm{IMSE}[\,\widehat{d}\,]=\int_{\Delta}^{1-\Delta}\mathrm{MSE}[\,\widehat{d}(u)\,]\,du\approx\frac{\Delta}{m\,J}\sum_{j=1}^{J}\sum_{k=1}^{m}[\,\widehat{d}_k(u_j)-d(u_j)\,]^2,$$

where $u_j=\Delta+(1-2\Delta)j/T$, $j=1,\ldots,J$, $J=(1-2\Delta)T/\Delta$, and m is the number of repetitions. In this case, we have chosen $\Delta=0.2$ and $m=100$. Figures 2 and 3 show the contour curves of the empirical IMSE of the Whittles estimates of the linear trend model $\widehat{d}(u)=\widehat{\alpha}_0+\widehat{\alpha}_1\,u$, with parameters $\alpha_0=0.15$ and $\alpha_1=0.30$. The darkest zones indicate the minimal IMSE. From Figure 2 for $T=500$, the optimal selection is close to $N\approx 130$ and $S\approx 25$. Besides, from Figure 3 for $T=1,000$ the optimal selection in this case seems to be near $N\approx 160$ and $S\approx 50$. On the other hand, Figures 4 and 5 show the empirical IMSE of the Whittles estimates of the cubic trend model $\widehat{d}(u)=\widehat{\tau}_0+\widehat{\tau}_1\,u^3$ with parameters $\tau_0=0.05$ and $\tau_1=0.4$. Observe that from Figure 4 for $T=500$, the optimal selection appears to be close

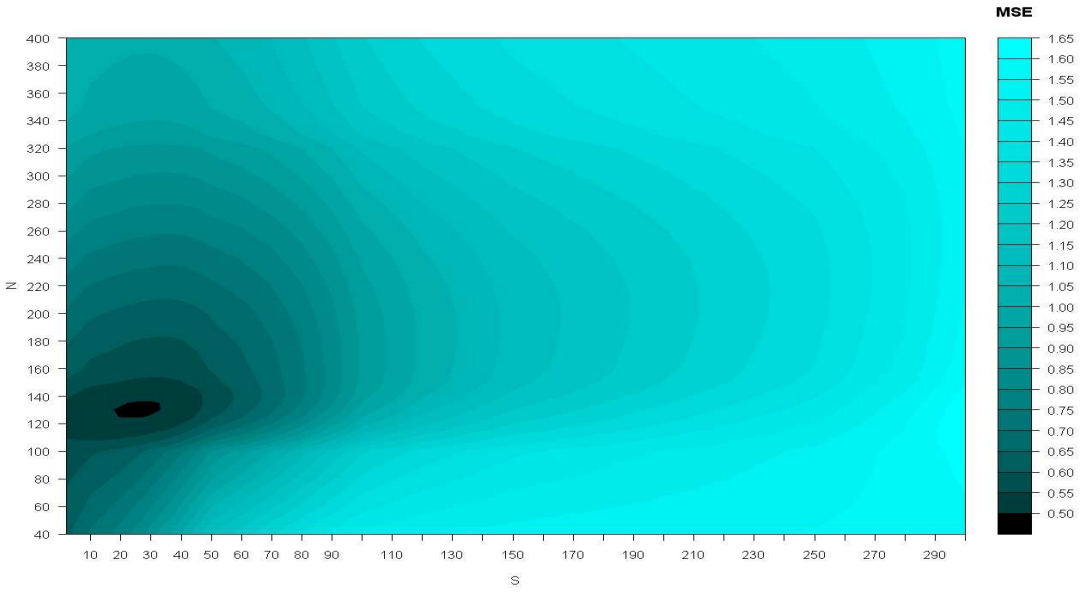

FIG 2. Contour curves of the empirical IMSE of Whittle estimator for the linear trend model with $T = 500$.

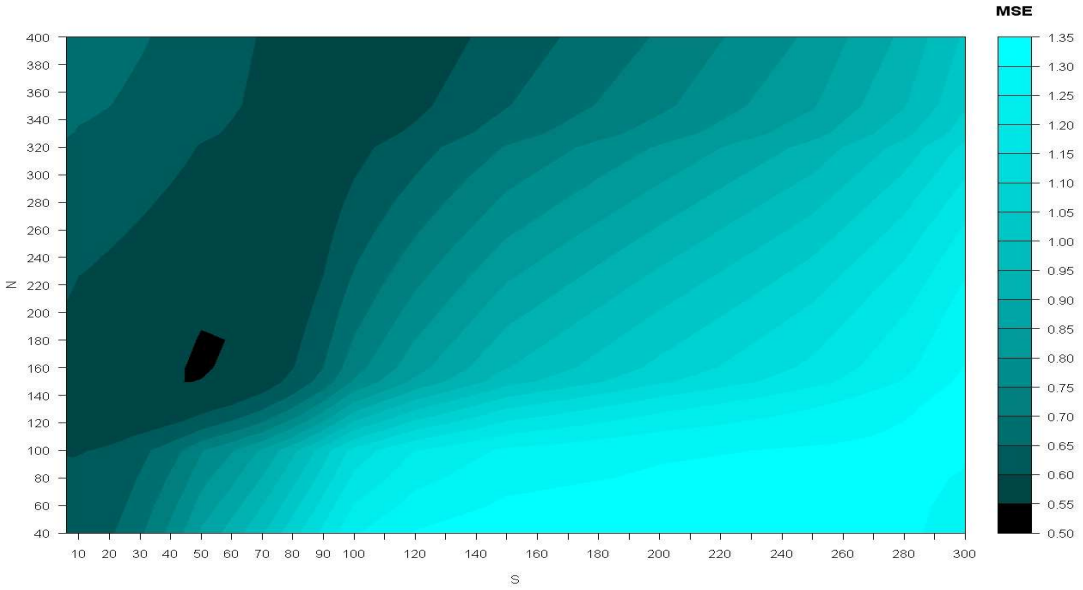

FIG 3. Contour curves of the empirical IMSE of Whittle estimator for the linear trend model with $T = 1,000$.

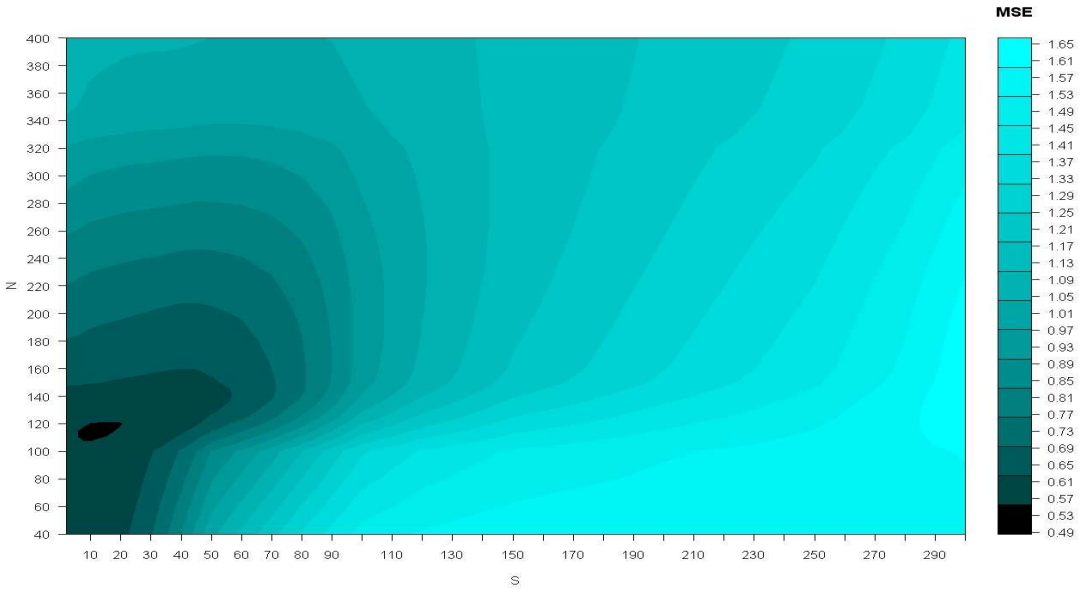

FIG 4. Contour curves of the empirical IMSE of Whittle estimator for the cubic trend model with $T = 500$.

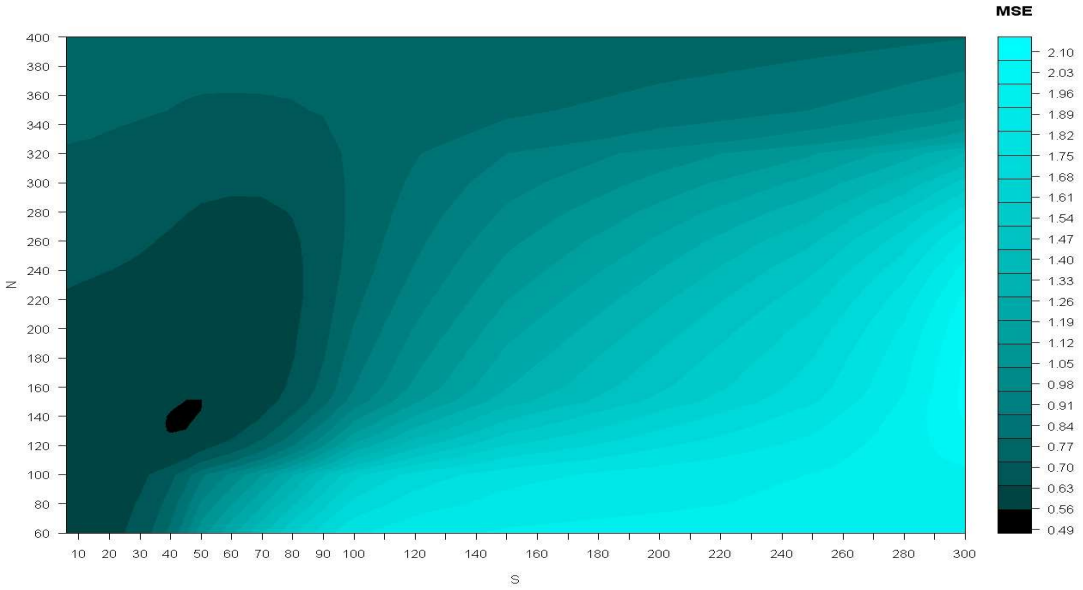

FIG 5. Contour curves of the empirical IMSE of Whittle estimator for the cubic trend model with $T = 1,000$.

to $N \approx 115$ and $S \approx 15$ while from Figure 5 for $T = 1,000$ the optimal selection seems to be $N \approx 140$ and $S \approx 45$. In any case, as suggested by theses plots, there is some flexibility for choosing the values of N and S as long as they belong to the areas with minimal IMSE. Table 1 shows the theoretical and empirical values of the IMSE -multiplied by the sample size T- for both the linear and cubic trend models. The integrated mean squared errors displayed in this table are based on the optimal bandwidth selection of N and S from the contour curves. Note that the empirical IMSE reported in this table are very close to their theoretical optimal counterparts for both sample sizes. Finally, Table 2 provides some general practical guidelines for

TABLE 1
Bandwidth selection: Theoretical and empirical IMSE, multiplied by T.

	Linear Trend Model				Cubic Trend Model			
T	N	S	Optimal	Whittle	N	S	Optimal	Whittle
500	130	25	0.4961	0.5232	115	15	0.4893	0.5262
1,000	160	50	0.4961	0.5484	140	45	0.4893	0.5329

selecting N and S, for different sample sizes. These suggested approximate ranges are based on the several simulations with a variety of models including linear and cubic trends for LSARFIMA models. These values can be used as references for an initial bandwidth selection. Other combinations of N and S should be tried out to verify the stability of the estimated parameters.

TABLE 2
Bandwidth selection: Suggested approximate starting ranges for N and S.

T	N	S
500	100–140	5–40
1000	120–210	30–60
2000	175–280	45–80

4. Simulation Studies. This section investigates the finite sample performance of the Whittle estimators through several Monte Carlo simulations. Besides, we carry out a comparative study of the performances of the Whittle method with a kernel local maximum likelihood estimator.

4.1. *Performance of the Whittle estimates.* In order to gain some insight into the finite sample performance of the Whittle estimator we report next several

Monte Carlo experiments, including linear and cubic trend evolutions of the Hurst parameter. In particular, we investigate the parameter estimation of the LSFN process defined by (1). The first set of simulations was carried out by specifying a linear trend for the long-memory parameter and constant noise variance, as in equation (2) of Example 2.1. The parameters in specification (2) are such that the vector $(\alpha_0, \alpha_1, \beta_0)$ satisfies $0 < d(u) < \frac{1}{2}$ and $\sigma(u) > 0$ for all $u \in [0, 1]$. In the second set of simulations, we consider a cubic trend evolution of the long-memory parameter and a constant noise scale parameter, as in model (4) of Example 2.2. Analogously to the linear trend specification (2), in this cubic trend model the parameter vector $(\tau_0, \tau_1, \beta_0)$ is such that $0 < d(u) < \frac{1}{2}$ and $\sigma(u) > 0$ for all $u \in [0, 1]$. For both sets of simulations, we consider two sample sizes $T = 500$ and $T = 1,000$ with empirical optimal selection of the subdivisons N and S as discussed in Section 3.2. The data taper used in these experiments is the cosine bell function $h(u) = \frac{1}{2}[1 - \cos(2\pi u)]$. The results, based on 1,000 replications, are reported in Tables 3 and 4.

TABLE 3
Whittle estimation for the linear trend model $d(u) = \alpha_0 + \alpha_1 u$ and $\sigma(u) = \beta_0$.

$T = 500, \quad N = 130, \quad S = 25$											
Parameters			Estimates			Theoretical SD			Estimated SD		
α_0	α_1	β_0	$\widehat{\alpha}_0$	$\widehat{\alpha}_1$	$\widehat{\beta}_0$	$\sigma(\widehat{\alpha}_0)$	$\sigma(\widehat{\alpha}_1)$	$\sigma(\widehat{\beta}_0)$	$\widehat{\sigma}(\widehat{\alpha}_0)$	$\widehat{\sigma}(\widehat{\alpha}_1)$	$\widehat{\sigma}(\widehat{\beta}_0)$
0.15	0.30	1.00	0.150	0.281	0.996	0.070	0.121	0.032	0.080	0.136	0.034
0.45	-0.35	1.00	0.428	-0.322	0.999	0.070	0.121	0.032	0.070	0.123	0.034
0.20	0.25	1.00	0.199	0.229	0.994	0.070	0.121	0.032	0.083	0.140	0.035
$T = 1000, \quad N = 160, \quad S = 50$											
Parameters			Estimates			Theoretical SD			Estimated SD		
α_0	α_1	β_0	$\widehat{\alpha}_0$	$\widehat{\alpha}_1$	$\widehat{\beta}_0$	$\sigma(\widehat{\alpha}_0)$	$\sigma(\widehat{\alpha}_1)$	$\sigma(\widehat{\beta}_0)$	$\widehat{\sigma}(\widehat{\alpha}_0)$	$\widehat{\sigma}(\widehat{\alpha}_1)$	$\widehat{\sigma}(\widehat{\beta}_0)$
0.15	0.30	1.00	0.146	0.295	0.997	0.049	0.085	0.022	0.056	0.092	0.024
0.45	-0.35	1.00	0.437	-0.337	0.997	0.049	0.085	0.022	0.052	0.096	0.024
0.20	0.25	1.00	0.200	0.239	0.998	0.049	0.085	0.022	0.058	0.101	0.024

Table 3 exhibits the results from the linear trend model for sample sizes $T = 500$ and $T = 1,000$. We have chosen three combinations of parameters α_0, α_1 and β_0. From this table, note that the means of the estimated parameters are close to their true values. Furthermore, the estimated standard deviations are also close to their theoretical counterparts. Table 4 reports the results for the cubic trend model simulations for sample sizes $T = 500$ and $T = 1,000$. Observe from this table that the estimated parameters are very close to their true values and that the empirical standard deviations are close to their theoretical values.

TABLE 4
Whittle estimation for the cubic trend model $d(u) = \tau_0 + \tau_1\, u^3$ and $\sigma(u) = \beta_0$.

$T = 500, \quad N = 115, \quad S = 15$											
Parameters			Estimates			Theoretical SD			Estimated SD		
τ_0	τ_1	β_0	$\widehat{\tau}_0$	$\widehat{\tau}_1$	$\widehat{\beta}_0$	$\sigma\,(\widehat{\tau}_0)$	$\sigma\,(\widehat{\tau}_1)$	$\sigma(\widehat{\beta}_0)$	$\widehat{\sigma}\,(\widehat{\tau}_0)$	$\widehat{\sigma}\,(\widehat{\tau}_1)$	$\widehat{\sigma}(\widehat{\beta}_0)$
0.05	0.40	1.00	0.054	0.350	0.997	0.046	0.123	0.032	0.042	0.132	0.033
0.45	-0.25	1.00	0.436	-0.235	0.996	0.046	0.123	0.032	0.052	0.161	0.035
0.10	0.30	1.00	0.098	0.259	0.997	0.046	0.123	0.032	0.054	0.158	0.035
$T = 1000, \quad N = 140, \quad S = 45$											
Parameters			Estimates			Theoretical SD			Estimated SD		
τ_0	τ_1	β_0	$\widehat{\tau}_0$	$\widehat{\tau}_1$	$\widehat{\beta}_0$	$\sigma\,(\widehat{\tau}_0)$	$\sigma\,(\widehat{\tau}_1)$	$\sigma(\widehat{\beta}_0)$	$\widehat{\sigma}\,(\widehat{\tau}_0)$	$\widehat{\sigma}\,(\widehat{\tau}_1)$	$\widehat{\sigma}(\widehat{\beta}_0)$
0.05	0.40	1.00	0.049	0.382	0.997	0.033	0.087	0.022	0.033	0.084	0.024
0.45	-0.25	1.00	0.447	-0.254	0.998	0.033	0.087	0.022	0.037	0.109	0.024
0.10	0.30	1.00	0.096	0.291	0.998	0.033	0.087	0.022	0.036	0.104	0.023

In summary, these simulations suggest that the finite sample performance of the proposed Whittle estimators seems to be very good in terms of bias and standard deviations.

4.2. *Comparison of Whittle and kernel estimates.* In this section we compare the performance of the Whittle estimator with a recently proposed kernel local maximum likelihood estimator, see Beran (2009). Suppose that we are interested in estimating the parameter $\theta = d(u)$ at the point $u = u_0$. In the kernel estimation approach, we maximize the local likelihood around u_0,

$$\mathcal{L}(\theta) = \sum_{t=t_0-[T\,b]}^{t_0+[T\,b]} K\left(\frac{t-t_0}{T\,b}\right) e_t^2(\theta),$$

with $t_0 = [T\,u_0]$ where $[\,\cdot\,]$ is the integer part function, $e_t(\theta) = y_t - \sum_{j=1}^{t-1} \pi_j(\theta) y_{t-j}$, $K(\cdot)$ is a nonnegative kernel function such that $K(u) = K(-u)$, $K(u) = 0$ for $|u| > 1$, and b is a bandwidth selection parameter. The optimal choice of this parameter is discussed in detail in Section 4 of Beran (2009).

In order to compare the performance of the Whittle and the kernel estimators, we study the same cubic trend model discussed by Beran (2009), $d(u) = \tau_0 + \tau_1\, u^3$, with kernel $K(u) = \frac{1}{2}\mathbb{1}_{[-1,1]}(u)$, $u_j = 0.2 + \Delta j/T$, $j = 1, \ldots, T/50$, $\Delta = 20$, $\tau_0 = 0.05$ and $\tau_1 = 0.4$. The optimal bandwidth for the kernel estimator was selected by minimizing the empirical IMSE, as suggested by (Beran, 2009, p.7). On the other hand, the empirical optimal bandwidth selection of N and S for the

Whittle estimator was based on the results from Section 3.2 for the cubic trend model.

From a theoretical point of view, one way to compare the performance of the Whittle and kernel estimators is by means of the variance ratio. From Eq. (24) of Beran (2009), the asymptotic variance of the kernel estimator, $\widetilde{d}(u)$, with optimal bandwidth selection is given by, $\mathrm{Var}[\widetilde{d}(u)] \cong \left(3\sqrt{2\tau_1\, u}\; T^{-1}\pi^{-2}\right)^{4/5}$. On the other hand, the asymptotic variance of the Whittle estimator, $\widehat{d}(u)$, is given by expression (5) from Example 2.2. Hence, the variance ratio of these estimators is

$$\frac{\mathrm{Var}[\,\widetilde{d}(u)\,]}{\mathrm{Var}[\,\widehat{d}(u)\,]} \cong \left[\frac{3}{4}\right]^{9/5} \frac{\left(\pi^2\,\tau_1^2\,u^2\,T\right)^{1/5}}{2-7\,u^3+14\,u^6}.$$

Figure 6 displays this ratio for $\tau_1 = 0.4$, $0.2 \le u \le 0.8$, and several values of T.

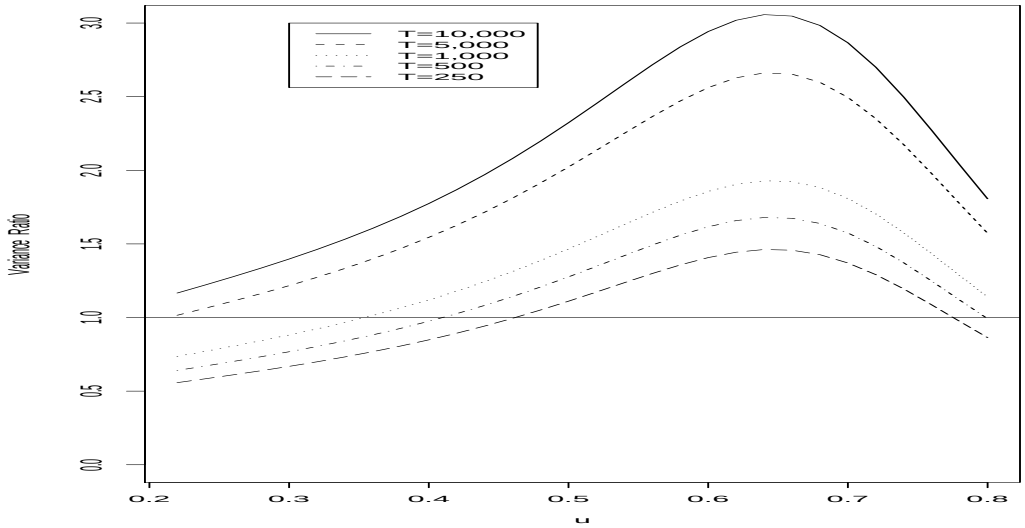

FIG 6. Variance ratio of the kernel and Whittle estimates for the cubic trend model $d(u) = 0.05 + 0.4\,u^3$, for several sample sizes T.

Observe that in these cases, the Whittle estimators exhibit smaller variances than the kernel estimators, for sample sizes larger than $T = 5,000$. In addition, Figure 7 shows the variance ratio of different values of T and τ_1. The heavy line represents the contour level equal to 1. Below that line, the kernel estimator possesses smaller variance than the Whittle estimator while the opposite occurs for the points above that dividing line. As expected from the asymptotic efficiency provided by Theorem 2.3 of Palma and Olea (2010), the Whittle estimator possesses the smallest variance as the sample size T increases.

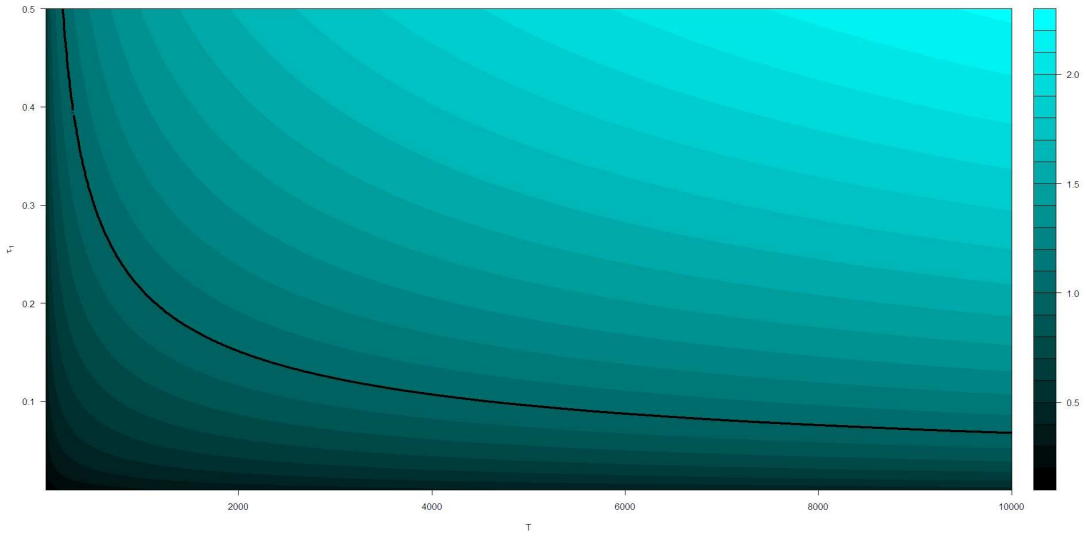

FIG 7. Variance ratio of the kernel and Whittle estimates for the cubic trend model $d(u) = \tau_0 + \tau_1\, u^3$, for different values of τ_1 and T. The heavy line represents a variance ratio equal to 1.

Now, in order to compare the Whittle and the kernel estimator from a empirical point of view, we have carried out two sets of Monte Carlo simulations for the cubic model, with $T = 500$ and $T = 1,000$, respectively. The results, based on 100 repetitions, are summarized in Table 5 and Figures 8 – 9. In Table 5, the IMSE has been multiplied by $T^{4/5}$, analogously to Table 1 of Beran (2009). Observe from Table 5 that the IMSE of the Whittle estimator of $d(u)$ is close to their theoretical optimum while the IMSE of the kernel estimator is almost twice that value. Figure 8 shows boxplots of the estimates from the Whittle (left panel) and kernel (right panel) methods, respectively, for $T = 500$. For the Whittle method, we used $N = 115$ and $S = 15$, the empirical optimal values provided by Figure 4 and Table 1. From the left panel of Figure 8 note that the Whittle estimates follow closely the true model. On the other hand, from the right panel of Figure 8, the kernel method seems to depart from the theoretical model for $u > 0.4$. A similar conclusion is drawn from Figure 9 for the boxplots for sample size $T = 1,000$. Summarizing, from a theoretical point of view, the Whittle method produces efficient estimates with variances decreasing at rate $\mathcal{O}(T^{-1})$. On the other hand, the variance of the kernel estimator decreases at the slower rate $\mathcal{O}(T^{-4/5})$. From a practical perspective, in the examples studied in this section the Whittle estimates seem to outperform the kernel estimates in terms of bias and variance.

TABLE 5
Comparison of Whittle and kernel estimates: IMSE for the cubic trend model, multiplied by $T^{4/5}$.

T	N	S	Optimal	Whittle	Kernel
500	115	15	0.1412	0.1518	0.3225
1000	140	45	0.1229	0.1339	0.3307

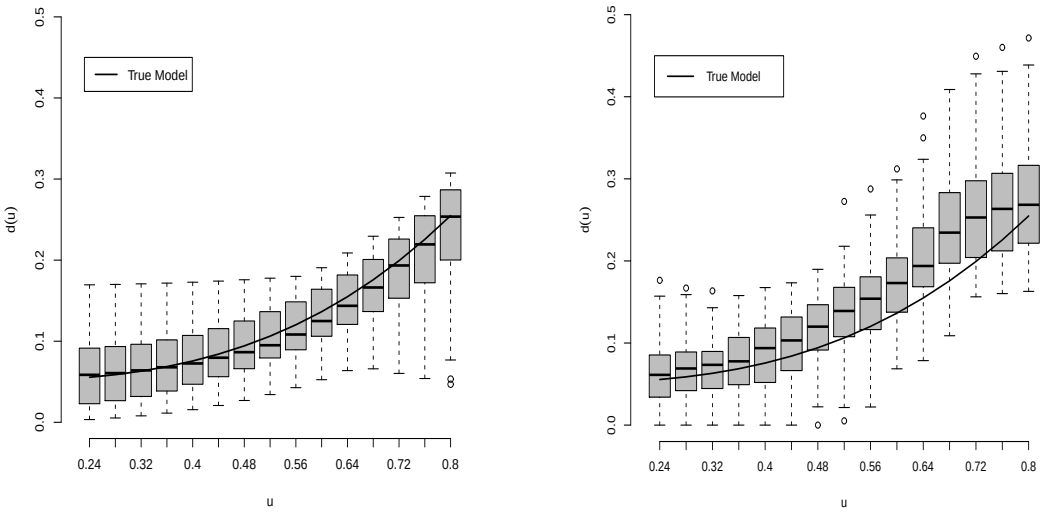

FIG 8. Boxplots of the Whittle (left panel) and kernel (right panel) estimator of $d(u) = \tau_0 + \tau_1\, u^3$, $T = 500$.

However, kernel estimators are valuable estimation tools in situations requiring a flexible methodology for handling complex time-varying parameter structures. Furthermore, in many practical settings Whittle and local smoothing techniques are complementary, as illustrated in the real-life data analysis carried out in Section 5.

5. Data Applications. This section discusses the application of the long-memory locally stationary methodology proposed in Palma and Olea (2010) to the analysis of two real-life time series data. The first example involves a time series consisting of speleothem cave deposit data. Speleothems are mineral deposits formed in caverns. Stalactites and stalagmites are two examples of speleothems. The second illustration is an application to tree ring data, which consist of annual ring width measurements. A major interest in the study of these two time series is the analysis of climatic variables, since the thickness of depositional layers and

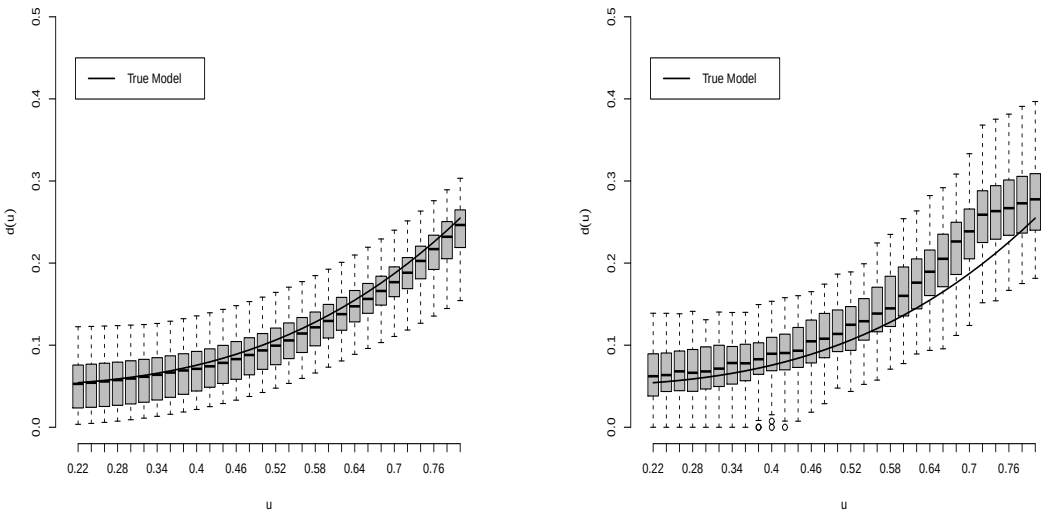

FIG 9. Boxplots of the Whittle (left panel) and kernel (right panel) estimator of $d(u) = \tau_0 + \tau_1\, u^3$, $T = 1,000$.

the tree ring widths are frequently used in many disciplines as climate proxies, see for example Tan et al. (2003) and references therein.

Given the features of the series discussed in this section, we use the simple model setup LSARFIMA with polynomial specifications for the long-memory and the noise scale parameters:

$$d(u) = \alpha_0 + \alpha_1 u + \cdots + \alpha_p u^p, \qquad \sigma(u) = \beta_0 + \beta_1 u + \cdots + \beta_q u^q. \tag{6}$$

In this case, the parameter vector is $\theta = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_0, \beta_1, \ldots, \beta_q)$. Following Dahlhaus (1997), the models are selected by the Akaike's Information Criterion (AIC) and by analyzing the significance of the parameters involved. The AIC is given by

$$\mathrm{AIC}(k) = 2\mathcal{L}_T(\widehat{\theta}) + 2(p + q + 2), \tag{7}$$

where $\mathcal{L}_T$ is the locally stationary Whittle log-likelihood function. Note that the class of LSARFIMA models contains, as a particular case, the family of stationary ARFIMA process. Thus, by looking at the AIC given by (7) and the significance of the parameters, it is possible to gain some insight about whether a stationary model or a locally stationary model fits the data better. The two real-life data examples presented next are meant only as illustrations of the application of the methodology developed in this paper. In these two cases, the locally stationary

models seems to fit better the time series data under study than their stationary counterparts. However, we are not suggesting that those are the only appropriate models. The underlying dynamic of the two data examples could be more complex than the evolutions suggested by the LSARFIMA model fitted. Locally stationary processes offer a flexible and parsimonious approach to the problem. They are flexible because they allow the fitting of a great variety of parameter evolution patterns by adequately choosing the class of basis functions. On the other hand, they constitute a parsimonious approach since the AIC model selection procedure looks for the simplest model that fits the observed data. Furthermore, the LSARFIMA framework allows the formal assessing of statistical hypotheses about the proposed models by means of the properties of the parameter estimates stated in Theorems 2.1–2.3 of Palma and Olea (2010). In particular, this framework allows for the statistical model comparison of stationary and locally stationary ARFIMA models, since ARFIMA processes are are nested in the class of LSARFIMA processes.

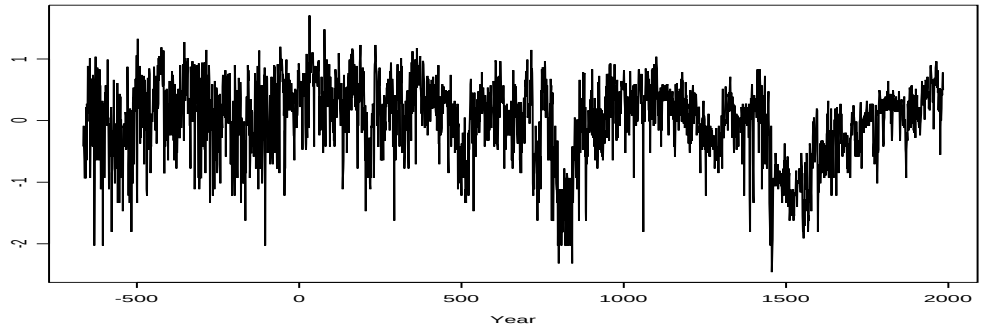

FIG 10. Cave mineral deposit data from 665 BC to 1985 AD.

5.1. *Speleothem Cave Deposit Data.* Figure 10 displays a 2,650-year centered time series of speleothem cave deposit data. This dataset contains stalagmite layer thickness from Shihua Cave, China, from 665 BC to 1985 AD, see Tan et al. (2003). The autocorrelation structure of this series is explored in Figure 11. Panel (a) of this figure shows the empirical autocorrelation function (ACF) while Panel (b) displays the variance plots, see (Beran, 1994, p.92-94) for details. In a variance plot, the broken line represents the expected behavior of the variance of the sample mean of a block of k observations for the short-memory case. On the other hand, the heavy line represents the expected behavior of the variance for a long-memory process. From both panels, this series seems to exhibit long-range dependence. However, a closer look to the empirical ACF of the data reveals that the degree of long-memory seems to change over time. In fact, Panel (a) of Figure 12 shows the sample ACF for

the first 500 observations, Panel (b) reports the sample ACf of 500 observations in the middle part of the series, $t = 1075, \ldots, 1574$, while Panel (c) shows the sample ACF of the last 500 observations, $t = 2151, \ldots, 2650$. Note that the level of strength of the autocorrelations looks different in each 500-year period. Moreover, the strength of the dependence appears to increase over time. On the other hand, from Figure 10, the variance of the series seems to decrease over time. Consistent

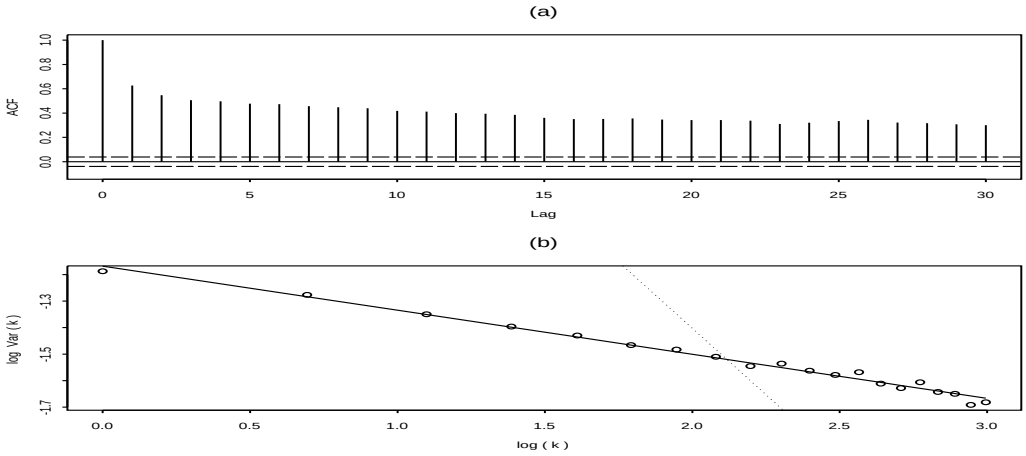

FIG 11. Cave deposit data. (a) Sample ACF, (b) Variance plot.

TABLE 6
Cave Deposit Data: Model Fitting

Parameters	Estimates	SD	z-value
α_0	0.1763	0.0454	3.8796
α_1	0.7730	0.2099	3.6832
α_2	-0.4615	0.2032	-2.2709
β_0	0.5753	0.0137	42.007
β_1	-0.2650	0.0206	-12.853

with this exploratory analysis, we propose the fitting of a locally stationary model to this time series, since this class of processes could capture the nonstationarity of the variance as well as the nonstationarity of the Hurst parameter, all this within a simplified model setup. The model selected by the AIC given by (7) and the significance of the parameters is a LSARFIMA with $p = 2$, $q = 1$ in specification (6) and estimates reported in Table 6. Note that according to the fourth column of this table, all the parameters of this model are statistically significant at the 5% level.

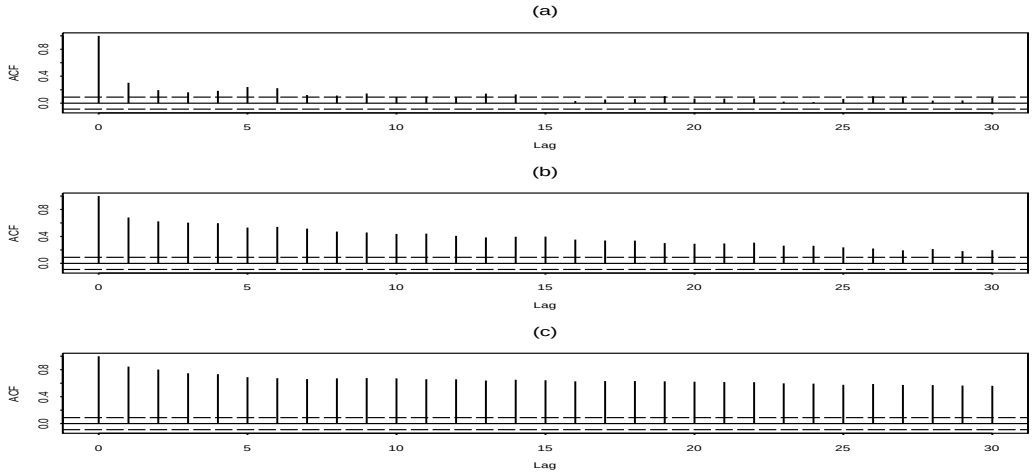

FIG 12. Cave deposit data. Sample ACF: (a) Observations 1 to 500, (b) observations 1075 to 1575, and (c) observations 2150 to 2650.

The evolutions of the long-memory coefficient, $d(u)$, and the variance scale, $\sigma^2(u)$, are depicted in Panels (a)–(b) of Figure 13, respectively. For comparison purposes,

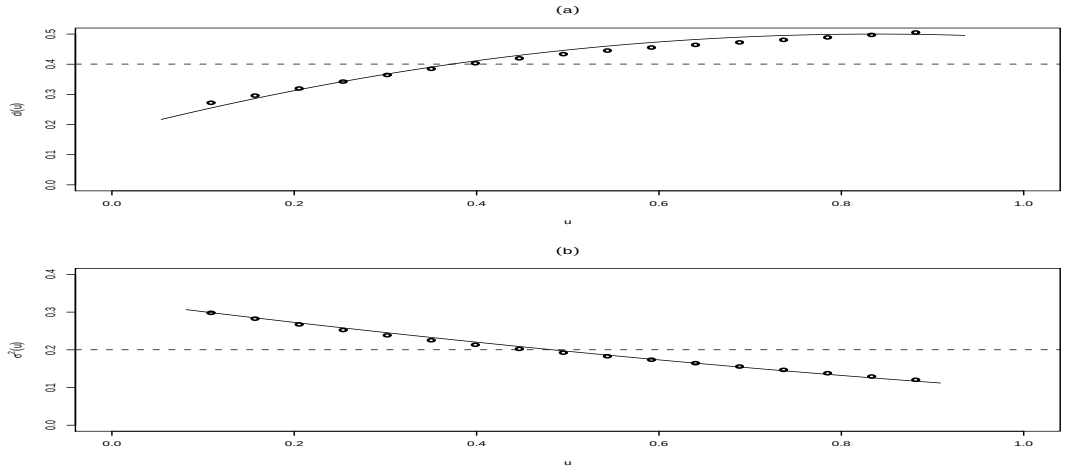

FIG 13. Cave deposit data. (a) Estimates of the long-memory parameter. (b) Estimates of the noise variance. In both panels the heavy line represents the locally stationary ARFIMA model, the horizontal broken line indicates the stationary ARFIMA model and the dots represent the cubic spline smoothing.

these plots show three estimation approaches: a locally stationary LSARFIMA process, a stationary ARFIMA process and a cubic spline smoothing procedure. The LSARFIMA model was fitted by means of an R computational program which is available from the authors upon request. The stationary ARFIMA model was fitted using the Haslett-Raftery method in R. On the other hand, the cubic spline smoothing is an heuristic approach that allows a simple and direct insight about the structure of the dynamic of the parameters, see for example Görg (2007). This approach consists in estimating the values of the parameters d and σ^2 for blocks of N observations, shifting S values each time, and then smoothing the parameter values by a cubic spline procedure. In this case, we have used the R implementation of the cubic spline smoothing method, see for example Hastie and Tibshirani (1990) and Chambers and Hastie (1992).

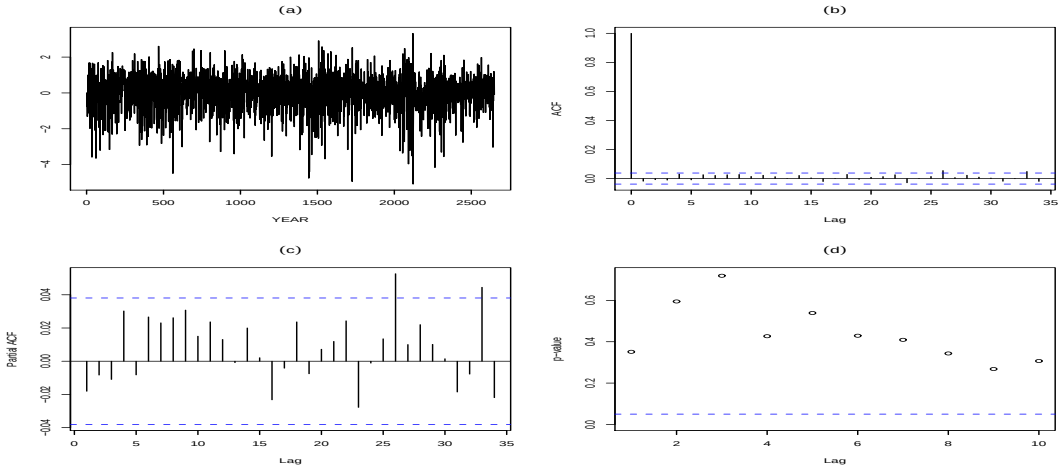

FIG 14. Cave deposit data: Residual analysis. (a) Residuals from the fitted model, (b) Sample ACF, (c) Sample partial ACF, (d) Ljung-Box tests.

In Figure 13, the heavy line represents the estimated values for the parameters $d(u)$ and $\sigma^2(u)$ for the fitted LSARFIMA process. The dotted lines represent these values for a stationary ARFIMA$(0, d, 0)$ process, fitted to the same data assuming d and σ constant. In this case the estimates of these parameters are $\widehat{d} = 0.4005$ and $\widehat{\sigma} = 0.4474$. Furthermore, the dots represent the cubic spline smoothing estimates. As expected from the exploratory analysis, the long-range dependence parameter increases over time while the variance decreases over time.

A residual analysis for the fitted LSARFIMA model is presented in Figure 14. Panel (a) of this figure shows the residuals from the model. Panels (b) and (c) displays the sample ACF and PACF, respectively, while Panel (d) exhibits the

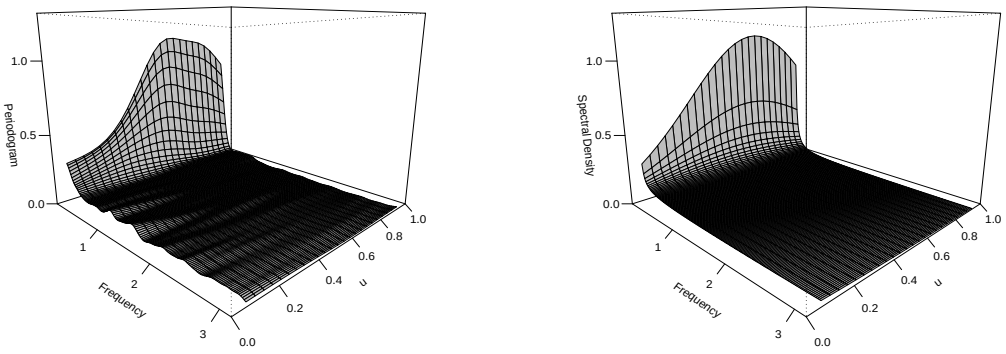

FIG 15. Cave deposit data. Left: Time-varying smoothed periodogram. Right: Theoretical spectral density of the fitted model.

Ljung-Box whiteness tests. No significant autocorrelations or partial autocorrelations are observed in Panels (b)–(c), respectively. Additionally, from Panel (d), the Ljung-Box portmanteau statistics for testing the whiteness of these residuals indicate that the white noise hypothesis is not rejected for all the lags considered, at the 5% significance level. Finally, Figure 15 illustrates the performance of the fitted LSARFIMA model from a spectral analysis perspective. The panel on the left shows the smoothed periodogram of the cave deposit data while the panel on the right displays the theoretical time-varying spectral density of the fitted model. Note that the time-frequency structure of these two panels are very similar, indicating that the proposed LSARFIMA model seems to fit appropriately the spectral dynamic of this time series.

5.2. *Tree ring data.* Figure 16 displays a series centered annual pinus longaeva tree ring width measurements at Mammoth Creek, Utah, from 0 AD to 1989 AD. These data, available at the National Climatic Data Center, are reported by Graybill (1990). The sample ACF of these data, displayed in Panel (a) of Figure 17, shows significant autocorrelations at large lags. In addition, the corresponding variance plot, see Panel (b) of Figure 17, provides further evidence of long-memory behavior. Nevertheless, a closer look at the autocorrelation structure by computing the sample ACF of data stretches, see Figure 18, suggests that the strength of the dependence seems to decay over time. In fact, the values of the sample ACF of the first 500 observations, see Panel (a), are higher than the corresponding sample ACF values for the other two 500-year periods considered, see Panels (b)–(c).

To account for this loss of memory over time, a LSARFIMA class of processes

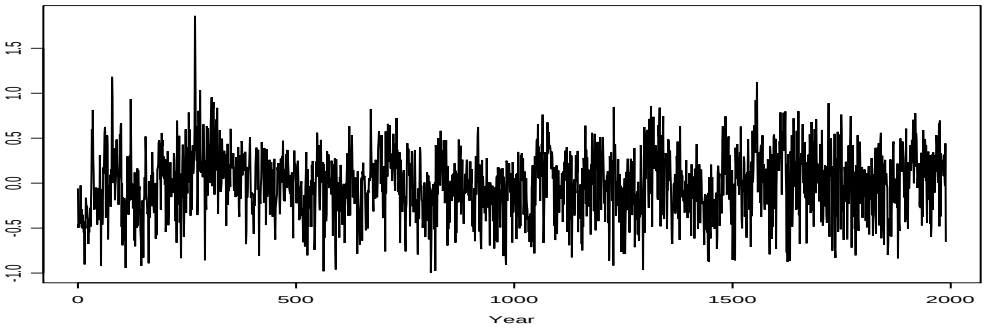

FIG 16. Tree ring data, from 0 to 1989 AD.

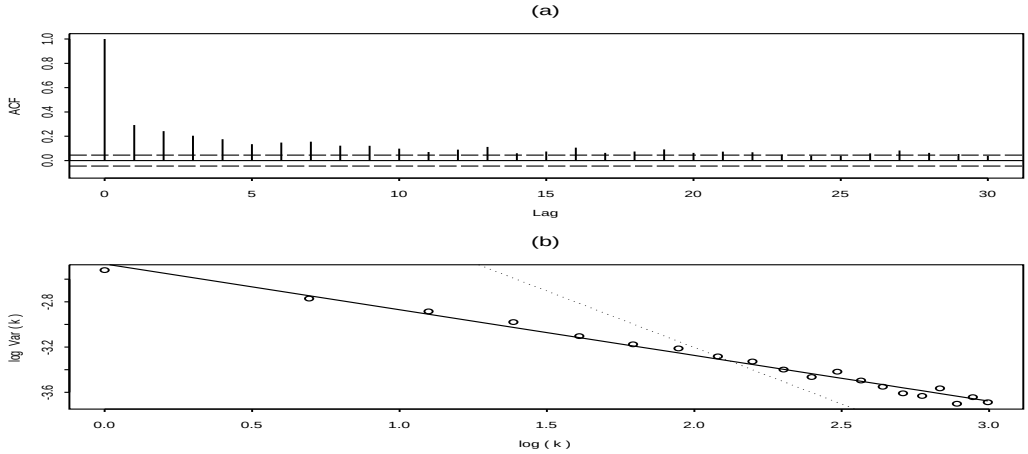

FIG 17. Tree ring data: (a) Sample ACF, (b) Variance plot.

is proposed to fit these observations. The model selected by the AIC given by (7) and the significance of the parameters is a LSARFIMA with $p = 1$, $q = 1$ in specification (6) and estimates reported in Table 7. Note that according to the fourth column of this table, all the parameters of this model are statistically significant at the 5% level.

Panels (a) and (b) of Figure 19 show the evolutions of the long-memory parameter, $d(u)$, and the variance scale, $\sigma^2(u)$, respectively. Similarly to the data application discussed in Section 5.1, this figure includes three estimation approaches: a LSARFIMA process, a stationary ARFIMA process and cubic spline smoothing. In Figure 19, the heavy line represents the estimated values for the parameters

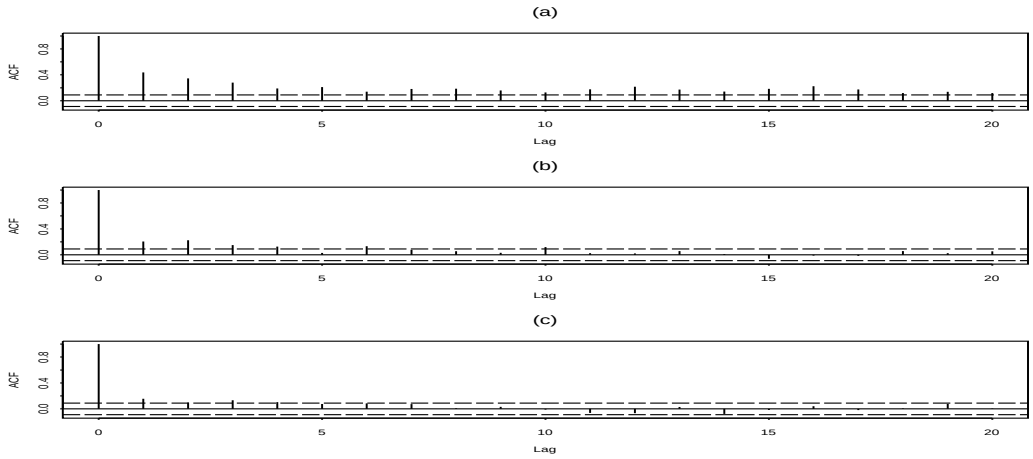

FIG 18. Tree ring data. Sample ACF: (a) Observations 1 to 500, (b) observations 750 to 1250, and (c) observations 1490 to 1990.

TABLE 7
Tree Ring Data: Model Fitting

Parameters	Estimates	SD	z-value
α_0	0.3285	0.0349	9.3998
α_1	-0.1890	0.0605	-3.1223
β_0	0.3609	0.0162	22.2197
β_1	-0.3087	0.0753	-4.0983
β_2	0.3884	0.0749	5.1877

$d(u)$ and $\sigma^2(u)$ for the LSARFIMA process. The dotted lines represent these values for a stationary ARFIMA$(0, d, 0)$ process, fitted to the same data assuming d and σ constant. In this case the estimates for these parameters are $\widehat{d} = 0.2415$ and $\widehat{\sigma} = 0.3287$. Furthermore, the dots represent the cubic spline smoothing estimates. As suggested by the exploratory analysis, the fitted model indicates that the long-range dependence parameter decreases over time. On the other hand, the variance seems to possess a more complex evolution. Figure 20 exhibits four panels exploring the structure of the residuals. Panel (a) of this figure displays the residuals from the fitted LSARFIMA model. Panels (b) and (c) show the sample ACF and PACF, respectively, while Panel (d) exhibits the Ljung-Box whiteness tests. No significant autocorrelations or partial autocorrelations are observed in Panels (b)–(c). This conclusion is formally supported by the Ljung-Box tests of Panel (d) which indicate that the white noise null hypothesis is not rejected for all

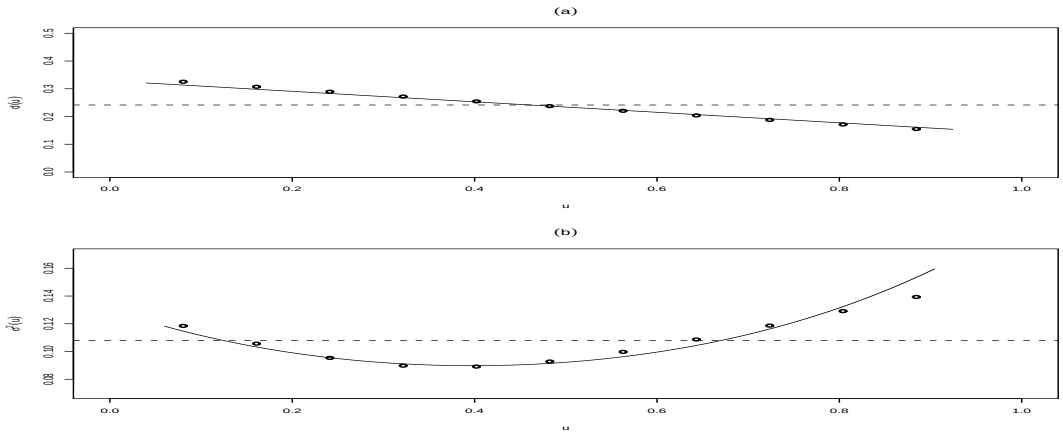

FIG 19. Tree ring data. a) Estimates of the long-memory parameter. (b) Estimates of the noise variance. In both panels the heavy line represents the locally stationary ARFIMA model, the horizontal broken line indicates the stationary ARFIMA model and the dots represent the cubic spline smoothing.

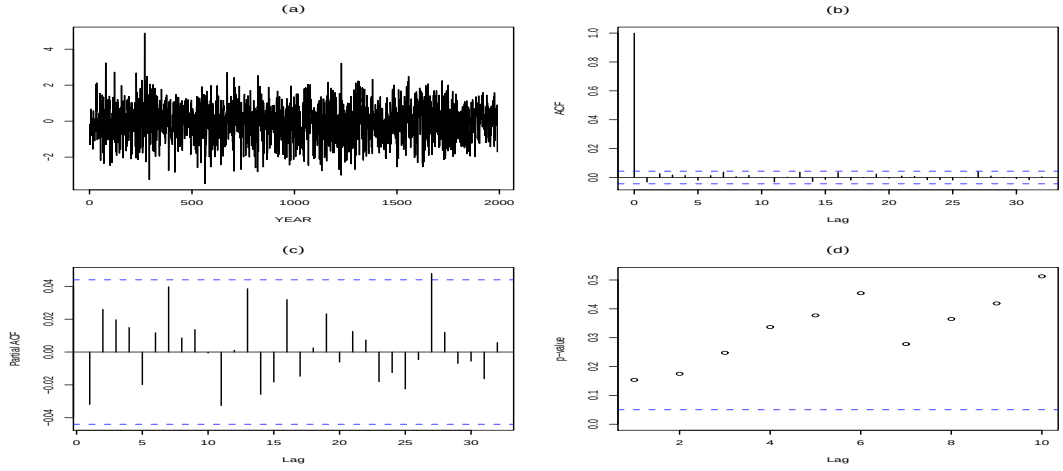

FIG 20. Tree ring data: Residual analysis. (a) Residuals from the fitted model, (b) Sample ACF, (c) Sample partial ACF, (d) Ljung-Box tests.

the lags considered, at the 5% level of significance.

Finally, Figure 21 compares the empirical and theoretical spectrum for the tree ring data. The panel on the left shows the smoothed periodogram of the tree ring series while the panel on the right displays the theoretical time-varying spectral

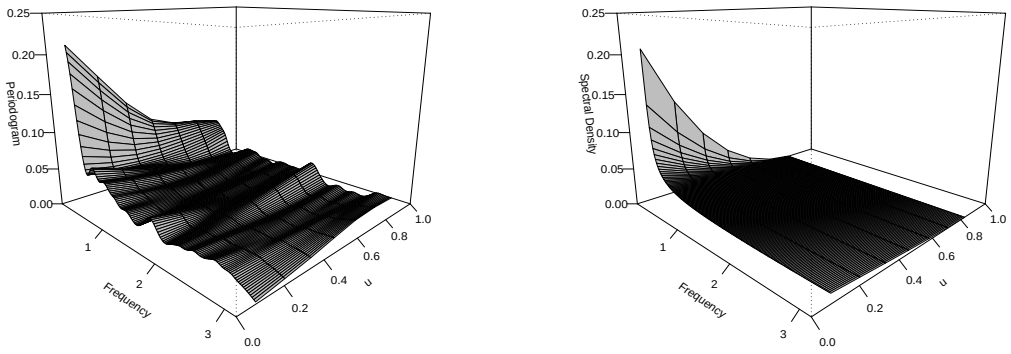

FIG 21. Tree ring data. Left: Time-varying smoothed periodogram. Right: Theoretical spectral density of the fitted model.

density of the fitted model. From this figure, the time-frequency structure of these two panels are very similar, indicating that the proposed LSARFIMA model seems to fit appropriately the spectral dynamic of this time series data. Observe that these tree ring data were also analyzed by Beran (2009). The results from our analysis agree with those from Beran's work since in both cases the fitted long memory parameter decreases monotonically, as indicated by our Figure 19(a) and Figure 8(c) of Beran (2009).

In the two real-life data applications examined above, both time series exhibit a time-varying long-range dependence structure and heteroscedasticity which are not adequately accounted for by fitting a stationary ARFIMA process. Moreover, that dependence structures and noise scale evolutions seem to be better fitted by a locally stationary ARFIMA model. Thus, these examples reveal that the long-memory locally stationary processes can be very helpful for modeling real-life time series data with complex time-varying parameter structures.

6. Appendix. This appendix contains the proofs of nine auxiliary lemmas used to prove the theorems stated in Palma and Olea (2010).

PROOF. (LEMMA 1) By Fubini's theorem and finite sum we have that

$$\begin{aligned}\sum_{k=-n}^{n}\widehat{g}(u,k) &= \int_{-\pi}^{\pi} g(u,\lambda)\sum_{k=-n}^{n} e^{-ik\lambda}\,d\lambda \\ &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\phi(u,\lambda+\omega)f(u,\omega)\sum_{k=-n}^{n} e^{-ik\lambda}\,d\lambda\,d\omega, \\ &= \int_{-\pi}^{\pi} f(u,\omega)\sum_{k=-n}^{n}\left\{\int_{-\pi}^{\pi}\phi(u,\lambda+\omega)e^{-ik(\lambda+\omega)}\,d\lambda\right\}e^{ik\omega}d\omega \\ &= 2\pi\int_{-\pi}^{\pi} f(u,\omega)\sum_{k=-n}^{n}\widehat{\phi}(u,\omega)e^{ik\omega}d\omega.\end{aligned}$$

Let $S_n(\omega)=\sum_{k=-n}^{n}\widehat{\phi}(u,k)\,e^{ik\omega}$, be the partial sum of the Fourier series of $\phi(u,\cdot)$, where

$$\widehat{\phi}(u,k)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\phi(u,\lambda)\,e^{ik\lambda}\,d\lambda.$$

Since $\phi(u,\cdot)$ has continuous derivative, $\phi(u,\cdot)$ is of bounded variation. Then, by the Dirichlet-Jordan criterion (Zygmund, 1959, Theorem 8.1(ii)), the Fourier series $S_n(\omega)$ converges uniformly to $\phi(u,\omega)$. Therefore, since $f(u,\cdot)$ is integrable we have

$$\sum_{k=-n}^{n}\widehat{g}(u,k)=2\pi\int_{-\pi}^{\pi}f(u,\omega)S_n(\omega)\,d\omega\to 2\pi\int_{-\pi}^{\pi}f(u,\omega)\phi(u,\omega)\,d\omega=2\pi g(u,0),$$

as $n\to\infty$.

□

PROOF. (LEMMA 2) Summation by parts yields

$$\begin{aligned}H_N\left[\phi\left(\tfrac{\cdot}{T},\gamma\right)h\left(\tfrac{\cdot}{N}\right),\lambda\right]-\phi\left(\tfrac{\cdot}{T},\gamma\right)H_N(\lambda) &= \left[\phi\left(\tfrac{N-1}{T},\gamma\right)-\phi\left(\tfrac{t}{T},\gamma\right)\right]H_N\left[h\left(\tfrac{\cdot}{N}\right),\lambda\right] \\ &\quad-\sum_{s=1}^{N-1}\left[\phi\left(\tfrac{s}{T},\gamma\right)-\phi\left(\tfrac{s-1}{T},\gamma\right)\right]H_N\left[h\left(\tfrac{\cdot}{N}\right),\lambda\right].\end{aligned}$$

But, by the mean value theorem

$$\left|\phi\left(\tfrac{s}{T},\gamma\right)-\phi\left(\tfrac{s-1}{T},\gamma\right)\right|\le\sup_{0\le u\le\frac{N}{T}}\left|\tfrac{\partial}{\partial u}\phi(u,\gamma)\right|\,\tfrac{1}{T}.$$

But, by assumption we have

$$\sup_{u\le\frac{N}{T}}\left|\tfrac{\partial}{\partial u}\phi(u,\gamma)\right|\le K\sup_{0\le u\le 1}|\gamma|^{-2d(u)}\le K\,|\gamma|^{-2d}.$$

Thus,

$$\left|\phi\left(\tfrac{s}{T},\gamma\right)-\phi\left(\tfrac{s-1}{T},\gamma\right)\right|\leq K\,\tfrac{|\gamma|^{-2d}}{T}.$$

Furthermore, by (Dahlhaus, 1988, Lemma 5.4) $|H_s(\lambda)|\leq K\,L_N(\lambda)$. Consequently,

$$\left|H_N\left[\phi\left(\tfrac{\cdot}{T},\gamma\right)\,h\left(\tfrac{\cdot}{N}\right),\lambda\right]-\phi\left(\tfrac{t}{N},\gamma\right)\,H_N(\lambda)\right|\leq K\;\;|\gamma|^{-2d}\,L_N(\lambda),$$

as required. □

Proof. (Lemma 3) Assume first that $\ell<0$ and let $n=-\ell>0$. By defining the positive function $g(x)=(x-1)^{-2d_1}-x^{-2d_1}$ for $x>1$, we can write the integral $I(\ell)$ as

$$I(\ell)=\int_1^n g(x)\,(n-x)^{d_1+d_2-1}\,dx+\int_n^\infty g(x)\,(x-n)^{d_1+d_2-1}\,dx,$$

where $n\geq 1$. Therefore, we can write

$$\begin{aligned}
&\int_1^n g(x)\,(n-x)^{d_1+d_2-1}\,dx\\
&=\int_1^n (x-1)^{-2d_1}\,(n-x)^{d_1+d_2-1}\,dx-\int_1^n x^{-2d_1}\,(n-x)^{d_1+d_2-1}\,dx\\
&=(n-1)^{d_2-d_1}\,B(1-2d_1,d_1+d_2)-\int_1^n x^{-2d_1}\,(n-x)^{d_1+d_2-1}\,dx\\
&=(n-1)^{d_2-d_1}\,B(1-2d_1,d_1+d_2)-\int_0^n x^{-2d_1}\,(n-x)^{d_1+d_2-1}\,dx\\
&\qquad+\int_0^1 x^{-2d_1}\,(n-x)^{d_1+d_2-1}\,dx
\end{aligned}$$

$$\begin{aligned}
&=(n-1)^{d_2-d_1}\,B(1-2d_1,d_1+d_2)-n^{d_2-d_1}\,B(1-2d_1,d_1+d_2)\\
&\qquad+\int_0^1 x^{-2d_1}\,(n-x)^{d_1+d_2-1}\,dx\\
&=\left[(n-1)^{d_2-d_1}-n^{d_2-d_1}\right]\,B(1-2d_1,d_1+d_2)+\int_0^1 x^{-2d_1}\,(n-x)^{d_1+d_2-1}\,dx\\
&\leq K\,n^{d_2-d_1-1}+K\,n^{d_1+d_2-1}\leq\mathcal{O}(n^{d_1+d_2-1}),
\end{aligned}$$

where $B(\cdot,\cdot)$ is the Beta function. On the other hand,

$$\int_n^\infty g(x)\,(x-n)^{d_1+d_2-1}\,dx\leq K\int_n^\infty x^{-2d_1-1}\,(x-n)^{d_1+d_2-1}\,dx\leq K\,n^{d_2-d_1-1}.$$

Therefore, $I(\ell)=\mathcal{O}((-\ell)^{d_1+d_2-1})$, $\ell<0$. Now, for $\ell\geq 0$ we have

$$I(\ell)\leq\ell^{\,d_1+d_2-1}\int_1^\infty g(x)\,dx,$$

but $0 \leq g(x) \leq K\, x^{-2d_1-1}$. Therefore,

$$\int_1^\infty g(x)\, dx \leq K \int_1^\infty x^{-2d_1-1}\, dx \leq K.$$

Hence, the result holds. □

PROOF. (LEMMA 4) Since ϕ is positive and symmetric in λ, there exist $Q(u)^{-1/2}$ such that $Q(u)^{-1} = Q(u)^{-1/2}Q(u)^{-1/2}$. Let $Y = Q(u)^{-1/2}X$. Then

$$X'X = Y'Q(u)Y = \int_{-\pi}^{\pi} \phi(u,\lambda) \left| \sum_{j=1}^{N} e^{i\lambda j} Y_j \right|^2 d\lambda \geq C \int_{-\pi}^{\pi} |\lambda|^{2d(u)} \left| \sum_{j=1}^{N} e^{i\lambda j} Y_j \right|^2 d\lambda.$$

Thus,

$$\frac{X'X}{Y'Y} \geq 2\pi C\, \frac{\int_{-\pi}^{\pi} |\lambda|^{2d(u)} |\sum e^{i\lambda j} Y_j|^2 d\lambda}{\int_{-\pi}^{\pi} |\sum e^{i\lambda j} Y_j|^2 d\lambda}.$$

Define $h^*(\lambda) = \frac{|\sum_{j=1}^{N} e^{i\lambda j} Y_j|^2}{\int_{-\pi}^{\pi} |\sum_{j=1}^{N} e^{i\lambda j} Y_j|^2 d\lambda}$. Then, this is a probability function over $[-\pi, \pi]$ satisfying $\int_{-\pi}^{\pi} h^*(\lambda) d\lambda = 1$ and $h^*(\lambda) \leq \frac{N}{2\pi}$. Consequently,

$$\frac{X'X}{Y'Y} \geq C \int_{-\pi}^{\pi} |\lambda|^{2d(u)} h^*(\lambda) d\lambda \geq C \inf_{h \in \mathcal{P}} \int_{-\pi}^{\pi} |\lambda|^{2d(u)} h(\lambda) d\lambda,$$

where $\mathcal{P} = \{h : \text{probability density in } [-\pi, \pi], h \leq N/2\pi\}$. Consequently,

$$\frac{X'X}{Y'Y} \geq C \frac{N}{\pi} \int_0^{\frac{\pi}{N}} \lambda^{2d(u)} d\lambda = \frac{\pi^{2d(u)} C}{2d(u)+1} N^{-2d(u)}.$$

Hence $Y'Y = X'Q(u)^{-1}(\phi)X \leq KX'XN^{2d(u)}$ where K is the positive constant $K = \pi^{-2d(u)}[2d(u)+1]C^{-1}$. □

PROOF. (LEMMA 5) By decomposing $X \in \mathbb{R}^{NM}$ as $X = (X(u_1)', \ldots, X(u_M)')'$, with $X(u) = (X_1(u), \ldots, X_N(u))$ and u_j are the indices defined in Section 2 of Palma and Olea (2010), we can write

$$X'[Q(\phi)^{-1} - Q(\varphi)]X = \sum_{j=1}^{M} X(u_j)'[Q(u_j)^{-1} - Q(\varphi_j)]X(u_j)$$

where $\varphi_j = \phi(u_j, \cdot)^{-1}/4\pi^2$. Note that $Q(u_j)^{-1}$ exists since $\phi(u,\lambda) > 0$. Furthermore, since ϕ is symmetric in $\lambda, Q(U_j)^{-1}$ can be decomposed as

$$Q(u_j)^{-1} = Q(u_j)^{-1/2} Q(u_j)^{-1/2}.$$

Therefore,

$$\begin{aligned} &X'[Q(\phi)^{-1} - Q(\varphi)]X \\ &= \sum_{j=1}^{M} [Q(u_j)^{-1/2} X(u_j)]^1 [I - Q(u_j)^{1/2} Q(\varphi_j) Q(u_j)^{1/2}][Q(u_j)^{-1/2} X(u_j)]. \end{aligned}$$

Hence,

$$|X'[Q(\phi)^1 - Q(\psi)]X| \leq \sum_{j=1}^{M} [X(u_j)' Q(u_j)^{-1} X(u_j)] |I - Q(u_j)^{1/2} Q(\varphi_j) Q(u_j)^{1/2}|.$$

By Lemma 6, $X(u_j)'Q(u_j)^{-1}X(u_j) \leq KX(u_j)'X(u_j)N^{2d(u_j)}$. On the other hand by Theorem 1.a of Fox and Taqqu (1987) we have that

$$|I - Q(u_j)^{1/2} Q(\varphi_j) Q(u_j)^{1/2}|^2 = \mathrm{tr}[I - 2Q(\varphi_j)Q(u_j) + Q(\varphi_j)^2 Q(u_j)^2] = \mathcal{O}(N).$$

Therefore, $|X'[Q(\phi)^{-1} - Q(\varphi)]X| \leq KN^{2d+\frac{1}{2}} \sum_{j=1}^{M} X'(u_j)X(u_j)$ as required. □

Proof. (Lemma 6) Assume that the vector $X \in \mathbb{R}^{MN}$ is decomposed as $X' = (X_1', \ldots, X_M')$ where $X_i' = (X_i(1), \ldots, X_i(N)) \in \mathbb{R}^N$. Thus, $X'RX = \sum_{ij=1}^{M} X_i' R_{ij} X_j$. In turn, by (**??**)

$$\begin{aligned} R_{ij}(r,s) &= \mathrm{E}(Y_{t_i - N/2+1+r,T}\, Y_{t_j - N/2+1+s,T}) \\ &= h\left(\tfrac{r}{N}\right) h\left(\tfrac{s}{N}\right) \int_{-\pi}^{\pi} e^{2\lambda(t_i - t_j + s - r} A^{\circ}_{t_i - N/2+r+1}(\lambda) \overline{A^{\circ}_{t_i - N/2+s+1}(-\lambda)} d\lambda, \\ &= h\left(\tfrac{r}{N}\right) h\left(\tfrac{s}{N}\right) \int_{-\pi}^{\pi} e^{2\lambda(t_i - t_j + s - r} A(u_i, \lambda) \overline{A(u_j, -\lambda)} d\lambda + \mathcal{O}\left(\tfrac{1}{T}\right) \\ &\equiv h\left(\tfrac{r}{N}\right) h\left(\tfrac{s}{N}\right) \widehat{f}(u_i, u_j, t_i - t_j + r - s) + \mathcal{O}\left(\tfrac{1}{T}\right). \end{aligned}$$

Therefore,

$$X_i' R_{ij} X_j = \sum_{r,s=1}^{N} h\left(\tfrac{r}{N}\right) h\left(\tfrac{s}{N}\right) \widehat{f}(u_i, u_j, t_i - t_j + r - s) X_i(r) X_j(s) + \mathcal{O}\left(\tfrac{N^2}{T}\right),$$

and

$$X'RX = \sum_{i,j=1}^{M} \sum_{r,s=1}^{N} h\left(\tfrac{r}{N}\right) h\left(\tfrac{s}{N}\right) \widehat{f}(u_i, u_j, t_i - t_j + r - s) X_i(r) X_j(s) + \mathcal{O}\left(\tfrac{M^2 N^2}{T}\right).$$

Thus,

$$
\begin{aligned}
|X'RX| &\leq \sum_{i,j_1}^{M}\sum_{r,s=1}^{N} h\left(\tfrac{r}{N}\right) h\left(\tfrac{s}{N}\right) |t_i - t_j|^{d(u_i)+d(u_j)-1} \left|\omega\left(\tfrac{i}{M},\tfrac{r}{N}\right)\omega\left(\tfrac{i}{M},\tfrac{r}{N}\right)\right| \\
&+ \mathcal{O}\left(\frac{M^2N^2}{T}\right) \\
&\leq KM^2N^2T^{-1}\int_0^1\int_0^1\int_0^1\int_0^1 h(x)h(y)|u-v|^{d(u)+d(v)-1}T^{d(u)+d(v)} \\
&\qquad \times |\omega(u,x)||\omega(v,y)|dx\,dy\,du\,dv + \mathcal{O}\left(\frac{M^2N^2}{T}\right). \qquad (8)
\end{aligned}
$$

where $\omega : [0,1]\times[0,1] \to \mathbb{R}$ is an integrable function such that $X_i(r) = \omega(\frac{i}{M}, \frac{r}{N})$ for $i = 1,\ldots,M$ and $r = 1,\ldots,N$. On the other hand, by Lemma 6 with $\delta = 2d+1$ we have,

$$X'A^{-1}X \;=\; X'A(\varphi)X + \mathcal{O}(X'XN^{\delta}),$$

and,

$$X'A(\varphi)X \;=\; \sum_{j=1}^{M} X_j'A(\varphi)X_j = \sum_{j=1}^{M}\sum_{r,s=1}^{N} X_j(r)X_j(s)\widehat{\varphi}(u,j,r-s),$$

where $\varphi(\lambda) = [4\pi^2\phi(\lambda)]^{-1}$. Thus,

$$
\begin{aligned}
X'A(\varphi)X &= K\sum_{j=1}^{M}\sum_{r,s=1}^{N} \omega\left(\tfrac{j}{M},\tfrac{r}{N}\right)\omega\left(\tfrac{j}{M},\tfrac{s}{N}\right)|\tfrac{r}{N}-\tfrac{s}{N}|^{2d_j-1}\,\frac{1}{N^2M}(MN^{2d_j+1}) \\
&\geq KMN\int_0^1\int_0^1\int_0^1 |\omega(u,x)||\omega(u,y)||x-y|^{2d(u)-1}N^{2d(u)}dx\,dy\,du. \qquad (9)
\end{aligned}
$$

Now, by combining (8) and (9) we get

$$\left|\frac{X'RX}{X'A^{-1}X}\right| \leq K\frac{MN}{T}B(N,T),$$

where

$$
\begin{aligned}
&B(N,T) \\
&= \int_0^1\int_0^1\int_0^1\int_0^1 h(x)h(y)|u-v|^{d(u)+d(v)-1}T^{d(u)+d(v)}|\omega(u,x)||\omega(v,y)|dx\,dy\,du\,dv \\
&\qquad \times\left[\int_0^1\int_0^1\int_0^1 |\omega(u,x)||\omega(u,y)||x-y|^{2d(u)-1}N^{2d(u)}dx\,dy\,du\right]^{-1}.
\end{aligned}
$$

Given that $0 \leq d(u), d(v) \leq d$, for all $u, v \in [0,1]$ we conclude

$$B(N,T) \leq K \left(\frac{T}{N}\right)^{2d},$$

proving the result. □

PROOF. (LEMMA 7) Consider a real number $\alpha \in (0,1]$ satisfying $1-\alpha = \mathcal{O}(1/N)$. Then, we can write

$$|C(N)| \leq K \log N \sum_{t=0}^{N-1} \sum_{k=N-t}^{N-1} \frac{1}{k^2} = K \log N \left\{ \sum_{t=1}^{\alpha N} \sum_{k=N-t}^{N-1} \frac{1}{k^2} + \sum_{t=\alpha N+1}^{N-1} \sum_{k=N-t}^{N-1} \frac{1}{k^2} \right\}.$$

For $\frac{t}{N} < \alpha < 1$ we have

$$\sum_{k=N-t}^{N-1} \frac{1}{k^2} = \frac{1}{N} \sum_{k=N-t}^{N-1} \left(\frac{N}{k}\right)^2 \frac{1}{N} \leq K \frac{1}{N} \int_{1-\frac{t}{N}}^{1} \frac{dx}{x^2} \leq K \frac{1}{N} \frac{t/N}{1-t/N}.$$

Thus,

$$\sum_{t=1}^{\alpha N} \sum_{k=N-t}^{N-1} \frac{1}{k^2} \leq \frac{K}{N} \sum_{t=1}^{\alpha N} \frac{t/N}{1-t/N} \leq K \int_0^{\alpha} \frac{x}{1-x} dx \leq -K \log(1-\alpha). \tag{10}$$

Since $1-\alpha = \mathcal{O}(\frac{1}{N})$, we have $-\log(1-\alpha) \sim \log N$. On the other hand, for $\alpha N + 1 \leq t \leq N-1$ we can write

$$\sum_{k=N-t}^{N-1} \frac{1}{k^2} \leq \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6},$$

Therefore,

$$\sum_{t=\alpha N+1}^{N-1} \sum_{k=N-t}^{N-1} \frac{1}{k^2} \leq \sum_{t=\alpha N+1}^{N-1} \frac{\pi^2}{6} \leq N(1-\alpha)\frac{\pi^2}{6} = \mathcal{O}(1). \tag{11}$$

Finally, by combining inequalities (10) and (11) we obtain the result. □

PROOF. (LEMMA 8) Note that for $k \geq 1$ we have $k^2 - d^2 \geq (k-1)^2$. Hence, for

any $\alpha \in (0,1]$ we can write

$$\begin{aligned}|D(N,T)| &\leq C\frac{\log}{N}N\sum_{t=0}^{N-1}\sum_{k=N-t}^{N-1}\frac{1}{(k-1)^2}\left(\frac{t}{T}\right)\\ &= C\frac{\log N}{T}\sum_{t=0}^{N-1}\sum_{k=N-t}^{N=2}\frac{1}{k^2}\left(\frac{t}{N}\right)\\ &= C\frac{\log N}{T}\left\{\sum_{t=2}^{\alpha N}\sum_{k=N-t}^{N-2}\frac{1}{k^2}\left(\frac{t}{N}\right)+\sum_{t=\alpha N+1}^{N-1}\sum_{k=N-t}^{N-2}\frac{1}{k^2}\left(\frac{t}{N}\right)\right\}.\end{aligned}$$

For $2 \leq t \leq \alpha N$, we have

$$\sum_{k=N-t}^{N-2}\frac{1}{k^2} = \frac{1}{N}\sum_{k=N-t}^{N-2}\left(\frac{N}{k}\right)^2\frac{1}{N} \leq K\frac{1}{N}\int_{1-\frac{t}{N}}^{1}x^{-2}dx \leq K\frac{1}{N}\left[\frac{(\frac{t}{N})}{1-(\frac{t}{N})}\right],$$

where $\frac{t}{N} \leq \alpha < 1$. Hence,

$$\sum_{t=2}^{\alpha N}\sum_{k=N-t}^{N-2}\frac{1}{k^2}\,(\tfrac{t}{N}) \leq K\frac{1}{N}\sum_{t=2}^{\alpha N}\frac{(\frac{t}{N})^2}{1-(\frac{t}{N})^2} \leq -K\log(1-\alpha).$$

Now, by choosing $\alpha = 1-\mathcal{O}(\frac{1}{N})$ we have that $-\log(1-\alpha) = \mathcal{O}(\log N)$. The second term can be bounded as follows. Since

$$\sum_{k=N-t}^{N-2}\frac{1}{k^2} \leq \sum_{k=1}^{N}\frac{1}{k^2} \leq \frac{\pi^2}{6},$$

we have

$$\sum_{t=\alpha N+1}^{N-1}\sum_{k=N-t}^{N-2}\frac{1}{k^2}\left(\frac{t}{N}\right) \leq \frac{\pi^2}{6}\sum_{t=\alpha N+1}^{N-1}\left(\frac{t}{N}\right) \leq KN\int_{\alpha}^{1}x dx = \frac{K}{2}N(1-\alpha^2).$$

But, $1-\alpha^2 = 1-\left[1-\mathcal{O}\left(\frac{1}{N}\right)\right]^2 = \mathcal{O}\left(\frac{1}{N}\right)$, so that

$$\left|\sum_{t=\alpha N+1}^{N-1}\sum_{k=N-t}^{N-2}\frac{1}{k^2}\frac{t}{N}\right| \leq K.$$

Consequently, $|D(N,T)| \leq K\frac{\log^2 N}{T}$, as required. □

PROOF. (LEMMA 9) Consider first the case $z \in [0,1]$. Note that the above integral can be written as $I(z) = I_1(z) + I_2(z)$ where

$$I_1(z) = \int_0^z (z-x)^{\alpha-1} \int_1^\infty (y-x)^{-\beta}(y-z)^{\alpha-1}\, dy\, dx,$$

and

$$I_2(z) = \int_z^1 (x-z)^{\alpha-1} \int_1^\infty (y-x)^{-\beta}(y-z)^{\alpha-1}\, dy\, dx.$$

Now, $I_1(z)$ can be written as

$$I_1(z) = \int_0^z x^{\alpha-\beta-1} \int_{1-z}^\infty (y/x+1)^{-\beta} y^{\alpha-1}\, dy\, dx.$$

An application of Eq. (3.194-2) of Gradshteyn and Ryzhik (2000) yields

$$I_1(z) = \frac{(1-z)^{2\alpha-\beta}}{\beta-\alpha} \int_0^z \left(\frac{x}{1-z}\right)^{\alpha-1} F\left(\beta, \beta-\alpha, \beta-\alpha+1, -\frac{x}{1-z}\right) \frac{dx}{1-z},$$

where $F(\cdot)$ is the hypergeometric function. By changing variables, this integral can be written as

$$I_1(z) = \frac{(1-z)^{2\alpha-\beta}}{\beta-\alpha} \int_0^{\frac{z}{1-z}} x^{\alpha-1} F(\beta, \beta-\alpha, \beta-\alpha+1, -x)\, dx.$$

Consequently,

$$I_1(z) \leq \frac{(1-z)^{2\alpha-\beta}}{\beta-\alpha} \int_0^\infty x^{\alpha-1} F(\beta, \beta-\alpha, \beta-\alpha+1, -x)\, dx.$$

But, by Eq. (7.511) of Gradshteyn and Ryzhik (2000) we conclude $I_1(z) \leq K(1-z)^{2\alpha-\beta}$. By analogous calculations, the second integral is given by

$$\begin{aligned} I_2(z) &= \frac{(1-z)^{\alpha-\beta}}{\beta-\alpha} \int_z^1 (x-z)^{\alpha-1} F\left(\beta, \beta-\alpha, \beta-\alpha+1, \frac{x-z}{1-z}\right) dx \\ &= \frac{(1-z)^{\alpha-\beta}}{\beta-\alpha} \int_0^1 x^{\alpha-1} F(\beta, \beta-\alpha, \beta-\alpha+1, x)\, dx. \end{aligned}$$

Hence, by Eq. (7.512) of Gradshteyn and Ryzhik (2000) we conclude $I_2(z) = K(1-z)^{2\alpha-\beta}$. On the other hand, by similar procedures we obtain $I(z) = K(z-1)^{2\alpha-\beta}$ for the case $z \in (1, 1+\delta]$, proving the result. □

Acknowledgement. This research was partially supported from Fondecyt Grant 1085239.

References.

I. S. Abramson. On bandwidth variation in kernel estimates—a square root law. *Ann. Statist.*, 10(4):1217–1223, 1982.

J. Beran. On parameter estimation for locally stationary long-memory processes. *J.Statist. Plann. Inference*, 139:900–915, 2009.

J. Beran. *Statistics for Long-Memory Processes*, volume 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1994.

J. Beran and Y. Feng. Iterative plug-in algorithms for SEMIFAR models—definition, convergence, and asymptotic properties. *J. Comput. Graph. Statist.*, 11(3):690–713, 2002a.

J. Beran and Y. Feng. SEMIFAR models—a semiparametric approach to modelling trends, long-range dependence and nonstationarity. *Comput. Statist. Data Anal.*, 40(2):393–419, 2002b.

M. Brockmann, T. Gasser, and E. Herrmann. Locally adaptive bandwidth choice for kernel regression estimators. *J. Amer. Statist. Assoc.*, 88(424):1302–1309, 1993.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, New York, Second edition, 1991.

J. M. Chambers and T. J. Hastie. *Statistical Models in S*. Wadsworth & Brooks/Cole, Pacific Grove, Ca, 1992.

R. Dahlhaus. Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Ann. Statist.*, 16(2):808–841, 1988.

R. Dahlhaus. Fitting time series models to nonstationary processes. *Ann. Statist.*, 25(1):1–37, 1997.

R. Dahlhaus and L. Giraitis. On the optimal segment length for parameter estimates for locally stationary time series. *J. Time Ser. Anal.*, 19(6):629–655, 1998.

R. Fox and M. S. Taqqu. Central limit theorems for quadratic forms in random variables having long-range dependence. *Probab. Theory Related Fields*, 74(2):213–240, 1987.

M. G. Görg. Long memory versus strucutral breaks: A time-varying memory approach. Master Thesis, 2007.

I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academic, San Diego, CA, sixth edition, 2000.

D. A. Graybill. Pinus longaeva tree ring data. Mammoth Creek, Utah, 1990.

T. J. Hastie and R. J. Tibshirani. *Generalized additive models*, volume 43 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London, 1990.

W. Palma and R. Olea. An efficient estimator for locally stationary Gaussian long-memory processes. *Ann. Statist.*, 2010.

P. M. Robinson. Rates of convergence and optimal spectral bandwidth for long range dependence. *Probab. Theory Related Fields*, 99(3):443–473, 1994.

B. W. Silverman. *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1986.

L. A. Stefanski. Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.*, 9(3):229–235, 1990.

M. Tan, T.S. Liu, J. Hou, X. Qin, H. Zhang, and T. Li. Cyclic rapid warming on centennial-scale revealed by a 2650-year stalagmite record of warm season temperature. *Geophysical Research Letters*, 30:191–194, 2003.

A. Zygmund. *Trigonometric Series. Vol. I*. Cambridge University Press, Cambridge, 1959.

Department of Statistics
Pontificia Universidad Católica de Chile