Local Whittle Estimation of Long-Range Dependence for Functional Time Series

Degui Li [*] Peter M. Robinson [†] Han Lin Shang [‡]

Version: December 5, 2020

Abstract

This paper studies stationary functional time series with long-range dependence, and estimates the memory parameter involved. Semiparametric local Whittle estimation is used, where periodogram is constructed from the approximate first score, which is an inner product of the functional observation and estimated leading eigenfunction. The latter is obtained via classical functional principal component analysis. Under the restrictive condition of constancy of the memory parameter over the function support, and other conditions which include rather unprimitive ones on the first score, the estimate is shown to be consistent and asymptotically normal with asymptotic variance free of any unknown parameter, facilitating inference, as in the scalar time series case. Although the primary interest lies in long-range dependence, our methods and theory are relevant to short-range dependent or negative dependent functional time series. A Monte-Carlo study of finite sample performance and an empirical example are included.

Keywords: Long-range dependence, Periodogram, Functional data, Functional principal component analysis, Local Whittle estimation.

[*]Department of Mathematics, University of York, York, YO10 5DD, UK.
[†]Department of Economics, London School of Economics, London WC2A 2AE, UK.
[‡]Department of Actuarial Studies and Business Analytics, Macquarie University, NSW 2109, Australia.

1 Introduction

The past few decades have seen extensive studies and notable developments in modelling long-range dependent (LRD) time series, which appear to exist in many areas such as economics, finance and geophysics. Beran (1994), Robinson (2003) and Giraitis, Koul and Surgailis (2012) provide comprehensive reviews of this topic. The autocovariance/autocorrelation for LRD processes decays to zero more slowly than for short-range dependent (SRD) ones, indeed is not summable, and the spectral density is unbounded at zero frequency. These factors lead to a significant difference in asymptotic theory. One of the most important issues in analysing LRD time series is estimation of the memory (or self-similarity) parameter, which measures dependence strength. The estimated memory parameter plays a crucial role in statistical inference. In general, there are two major approaches to estimation. One is parametric such as Gaussian maximum likelihood (e.g., Fox and Taqqu, 1986; Dahlhaus, 1989) which has the desirable asymptotic properties of root-n consistency (with n denoting the sample size), asymptotic normality and asymptotic efficiency. However, this relies on correct specification of the spectral density over the full frequency band $(-\pi, \pi]$, and becomes inconsistent in the case of misspecification. The other type is semiparametric estimation, which only needs assumptions on the spectral density in a shrinking neighborhood of zero frequency. The local Whittle (LW) or Gaussian semiparametric (Künsch, 1987; Robinson, 1995a) and log periodogram (LP, Robinson, 1995b) are the most commonly-used semiparametric methods. The LP regression has closed form, whereas LW is only implicitly defined but is more efficient. Asymptotics for LP estimation are complicated by the nonlinear functions of the periodogram involved, while LW has been justified under milder conditions. Hence, the main focus of the present paper is the LW method. It has been extensively studied in recent years for stationary and nonstationary time series settings (e.g., Velasco, 1999; Phillips and Shimotsu, 2004; Robinson, 2008).

The aforementioned methodology covers both univariate and finite dimensional multiple time series, but becomes infeasible when the dimension is large. To address this issue, functional time series provide a general framework by using a continuous function to approximate ordered observations. The bulk of the literature studies functional data which are either independent or stationary SRD (e.g., Bosq, 2000; Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Hörmann and Kokoszka, 2010; Horváth and Kokoszka, 2012; Berkes, Horváth and Rice, 2013; Hsing and Eubank, 2015). Li, Robinson and Shang (2020) is among the first to extend the functional framework from SRD to LRD (see also Characiejus and Rauckauskas, 2014; Düker, 2018). They not only establish the central limit theorem for a temporal sum of LRD functional observations, but also develop functional principal component analysis (FPCA) and estimate the memory parameter for the projected process via semiparametric R/S. However, R/S has a very slow convergence rate and performs poorly in finite samples, while LW estimation is known to be more efficient in the traditional time series case. Hence, we extend LW to functional time series.

We restrictively assume the memory parameter is constant across the support of the functional data. Our procedure uses the first score process (with each element defined as an inner product of the functional observation and the leading eigenfunction corresponding to the maximum eigenvalue of the long-run covariance operator to be defined in Section 2), which is univariate and stationary, so it is expected that LW time series asymptotics will still hold in our setting. However, as the first score is latent, we use FPCA to estimate the leading eigenfunction, and approximate the score by an inner product of the functional observation and the estimated eigenfunction. Then we apply LW to the approximate first score process to estimate the memory parameter. Under regularity conditions, we derive consistency and asymptotic normality, analogous to Theorems 1 and 2 of Robinson (1995a), showing that replacement of the latent score process by its approximation has negligible effect. Our model framework also covers SRD and negative dependent (ND) functional processes. The methodology developed in this paper complements the frequency domain analysis for functional time series which has received increasing attention in recent years (e.g., Panaretos and Tavakoli, 2013; Hörmann, Kidzinski and Hallin, 2015; Meyer, Paparoditis and Kreiss, 2020).

The rest of the paper is organised as follows. Section 2 introduces the model assumptions and the infeasible LW using the latent scores. Section 3 constructs an approximation of the first score via FPCA, describes feasible LW estimation and states the main asymptotic theorems. Section 4 assesses finite-sample performance via a Monte-Carlo simulation study, and presents an empirical data analysis using monthly sea surface temperature data. Section 5 concludes the paper. Proofs of the main theoretical results are available in the online supplement. Throughout the paper, we define the Hilbert space $\mathcal{H}$ as a set of measurable functions $z(\cdot)$ such that $\int_{\mathbb{C}} z^2(u)du < \infty$ and the relevant inner product is $\langle z_1, z_2\rangle = \int_{\mathbb{C}} z_1(u)z_2(u)du$, where $\mathbb{C}$ is a compact set. Let $\mathcal{L}_{\mathcal{H}}$ be a space of continuous linear operators from $\mathcal{H}$ to $\mathcal{H}$ equipped with the operator norm defined by $\|\mathbf{L}\| = \sup_{z\in\mathcal{H}}\{\|\mathbf{L}(z)\| : \|z\| \leqslant 1\}$ for $\mathbf{L} \in \mathcal{L}_{\mathcal{H}}$, where $\|z\| = \langle z, z\rangle^{1/2}$ for $z \in \mathcal{H}$. Denote $z_1 \otimes z_2 = \langle z_1, \cdot\,\rangle z_2$ for all $z_1, z_2 \in \mathcal{H}$ and $\mathbf{L}^\star$ as the adjoint of the operator $\mathbf{L}$. Let $a_n \sim b_n$ and $a_n \propto b_n$ denote that $a_n/b_n \to 1$ and $0 < \underline{c} \leqslant |a_n/b_n| \leqslant \bar{c} < \infty$, respectively.

2 Model and assumptions

In this section, we introduce a functional time series model structure covering LRD, SRD and ND functional processes, some technical assumptions as well as an infeasible LW estimation procedure.

2.1 Model framework

Assumption 1. *For an observable functional time series* $X_t = (X_t(u) : u \in \mathbb{C})$, *we impose*

$$X_t = \sum_{j=0}^{\infty} b_j \eta_{t-j}, \quad t = 1, 2, \cdots, \quad \sum_{j=0}^{\infty} b_j^2 < \infty, \tag{2.1}$$

where $(b_j : j \geqslant 0)$ *is a sequence of scalars and* $(\eta_t : t = 0, \pm 1, \pm 2, \cdots)$ *with* $\eta_t = (\eta_t(u) : u \in \mathbb{C})$ *is a sequence of independent and identically distributed (i.i.d.) random functions defined on the compact set* $\mathbb{C}$, *with zero mean and positive definite covariance operator defined by*

$$\mathbf{C}_\eta(x)(u) = \int_{\mathbb{C}} c_\eta(u, v) x(v) dv, \quad x \in \mathcal{H}, \quad c_\eta(u, v) = \mathsf{E}\left[\eta_t(u)\eta_t(v)\right].$$

Remark 1. Since the last quarter of the 20th century, the i.i.d. assumption on innovations has been relaxed in modern and relatively incisive treatments of central limit theory for estimates optimising quadratic forms of linear time series processes, such as the Whittle parametric estimate and the LW estimate: independence has been relaxed to martingale difference conditions, and identity of distribution to a milder homogeneity condition. Analogous relaxations are undoubtedly possible in our functional setting (2.1). Assumption 1 restrictively requires time series dependence structure to be constant across the function support, but allows X_t to be SRD, LRD or ND. Specifically, we may further assume under SRD,

$$\sum_{j=0}^{\infty} |b_j| < \infty, \quad \sum_{j=0}^{\infty} b_j \neq 0; \tag{2.2}$$

while under LRD or ND

$$b_j \sim j^{H_0 - 3/2} \text{ with } 0 < H_0 < 1/2 \text{ or } 1/2 < H_0 < 1 \text{ as } j \to \infty, \tag{2.3}$$

and, in addition, $\sum_{j=0}^{\infty} b_j = 0$ when $0 < H_0 < 1/2$, where H_0 is the memory parameter. Special cases of (2.1) under (2.3) include the parametric model used in Section 4 below and a functional version of the fractionally integrated time series model.

Given functional time series observations $(X_t : t = 1, \cdots, n)$, we define the unnormalised long-run covariance operator:

$$\mathbf{C}_n = \mathsf{E}\left[\sum_{t=1}^{n}\sum_{s=1}^{n} X_t \otimes X_s\right], \text{ or equivalently, } \mathbf{C}_n(z) = \mathsf{E}\left[\sum_{t=1}^{n}\sum_{s=1}^{n} \langle X_t, z\rangle X_s\right] \tag{2.4}$$

for $z \in \mathcal{H}$. Consider an eigenanalysis of $\mathbf{C}_n$ and obtain pairs of eigenvalues and orthonormal eigenfunctions (ρ_{nk}, ψ_{nk}), $k = 1, 2, \cdots$, where $\rho_{n1} \geqslant \rho_{n2} \geqslant \cdots \geqslant 0$ and $\psi_{nk} = (\psi_{nk}(u) : u \in \mathbb{C})$. As $\mathbf{C}_n$ is unnormalised, ρ_{n1} diverges to infinity as n increases and its divergence rate is mainly determined by the decaying rate of b_j defined in Assumption 1. Specifically, when (2.2) holds, by the proof of Theorem 1 in Horváth, Kokoszka and Reeder (2013), we have $\rho_{n1} \propto n$; when (2.3) holds, by Proposition 1 in Li, Robinson and Shang (2020) as well as Lemmas B.3 and B.4 in the online supplement, we have $\rho_{n1} \propto n^{2H_0}$ with H_0 defined in (2.3). As the divergence rate of ρ_{n1} determines the normalisation rate of $\mathbf{C}_n$, we define the normalised long-run covariance operator:

$$\mathbf{C} = \lim_{n\to\infty} \frac{1}{n^{H_*}} \mathbf{C}_n, \quad H_* = \begin{cases} 2H_0, & \text{when (2.3) holds,} \\ 1, & \text{when (2.2) holds,} \end{cases} \tag{2.5}$$

and subsequently obtain (ρ_k, ψ_k), $k = 1, 2, \cdots$, as pairs of eigenvalues and orthonormal functions of $\mathbf{C}$. In particular, we can easily show that $\rho_k = \lim_{n\to\infty} \frac{1}{n^{H_*}} \rho_{nk}$ for $k = 1, 2, \cdots$.

2.2 Infeasible LW estimation

Define $x_t^1 = \int_{\mathbb{C}} X_t(u)\psi_1(u)du$, the inner product of X_t and ψ_1, which is usually referred to as the first score in the functional data analysis. By (2.1) and Assumption 1, we may write

$$x_t^1 = \sum_{j=0}^{\infty} b_j \eta_{t-j}^1, \quad \eta_t^1 = \int_{\mathbb{C}} \eta_t(u)\psi_1(u)du. \tag{2.6}$$

By Assumption 1, $(\eta_t^1, t = 0, \pm 1, \pm 2, \cdots)$ is an i.i.d. sequence with mean zero and positive variance $\sigma_\eta^2 = \mathsf{E}\left[(\eta_t^1)^2\right]$. As in Robinson (1995a), we suppose that the spectral density of x_t^1, denoted by $f_x(\lambda)$, satisfies

$$f_x(\lambda) \sim G_0 \lambda^{1-2H_0} \quad \text{as } \lambda \to 0+, \tag{2.7}$$

where G_0 is an unknown positive constant, $0 < H_0 < 1$, and $\lambda \to 0+$ denotes convergence to zero from above. The decay rates in (2.3) and (2.7) are consistent. Our main interest lies in estimating the memory parameter H_0 via LW.

Define the periodogram of x_t^1 at frequency λ as

$$I_1(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} x_t^1 e^{it\lambda} \right|^2. \tag{2.8}$$

Define $\lambda_j = 2\pi j/n$ and let $m = m_n$ be a bandwidth sequence chosen by the practitioner and satisfying Assumptions 2 and 2* below. Following Robinson (1995a), we deduce from a Gaussian objective function the LW estimate:

$$\widetilde{H} = \arg\min_{H \in \Theta} \widetilde{R}(H), \tag{2.9}$$

where

$$\widetilde{R}(H) = \log \widetilde{G}(H) - \frac{2H-1}{m} \sum_{j=1}^{m} \log \lambda_j, \quad \widetilde{G}(H) = \frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2H-1} I_1(\lambda_j), \tag{2.10}$$

$\Theta = [\Delta_1, \Delta_2]$ with Δ_1 and Δ_2 chosen such that $0 < \Delta_1 < \Delta_2 < 1$. In practical implementation, we may choose Δ_1 and Δ_2 arbitrarily close to 0 and 1, respectively.

Assumption 2. *(i)* $f_x(\lambda)$ *satisfies (2.7), is differentiable in a neighborhood of the origin and*

$$\frac{d}{d\lambda} \log f_x(\lambda) = O(1/\lambda) \ \text{ as } \lambda \to 0+.$$

(ii) $m \to \infty$ *and* $m = o(n)$ *as* $n \to \infty$.

Assumption 2*. *(i)*

$$f_x(\lambda) \sim G_0 \lambda^{1-2H_0} \left(1 + O(\lambda^{\delta_1})\right) \ \text{ as } \lambda \to 0+, \tag{2.11}$$

where $0 < \delta_1 \leqslant 3$, *and* $\beta(\lambda) := \sum_{j=0}^{\infty} b_j e^{ij\lambda}$ *is differentiable in a neighborhood of the origin, with*

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

(ii) $m \to \infty$ *and* $m^{1+2\delta_1}(\log m)^2 = o\left(n^{2\delta_1}\right)$ *as* $n \to \infty$ *with* δ_1 *defined as in (2.11).*

Remark 2. Assumptions 2 and 2* above are similar to the conditions used in Robinson (1995a). Assumption 2 is imposed to establish consistency of $\widetilde{H}$. The stronger Assumption 2* is imposed to derive asymptotic normality.

By Assumption 1, x_t^1 defined in (2.6) is a univariate stationary linear process, where $\sum_{j=0}^{\infty} b_j^2 < \infty$ and (η_t^1) is an i.i.d. sequence with mean zero and variance $\sigma_\eta^2 > 0$. This implies that Assumption A3 in Robinson (1995a) is satisfied (in fact, this only requires innovations and centred squared innovations in the linear process to be martingale differences). From Theorems 1 and 2 in Robinson (1995a), we readily have the following proposition.

Proposition 1. *Suppose that Assumptions 1 and 2 are satisfied and* $\Delta_1 < H_0 < \Delta_2$. *Then, (i)* $\widetilde{H}$ *is weakly consistent; (ii) under Assumption 2* and assuming* $\mathsf{E}\left[\|\eta_t\|^4\right] < \infty$,

$$m^{1/2}\left(\widetilde{H} - H_0\right) \xrightarrow{d} \mathsf{N}(0,\ 1/4), \ \text{ as } n \to \infty. \tag{2.12}$$

As in Robinson (1995a) the lack of bias entailed in the centring at H_0 is due to Assumption 2*. Unfortunately, $\widetilde{H}$ is practically infeasible as the eigenfunction ψ_1 and the first score x_t^1 are unobservable. In Section 3 below, we use FPCA to consistently estimate ψ_1 (up to possible sign change), obtain an approximation to x_t^1, and subsequently construct feasible LW estimation of H_0, for which Proposition 1 continues to hold.

3 Main methodology and theory

In this section, we first introduce classical FPCA to estimate ψ_1 and a feasible LW estimation method using the approximate first score process, and then state the main asymptotic results including consistency and asymptotic normality for the feasible LW.

3.1 FPCA and feasible LW estimation

In order to construct feasible LW estimation of H_0, we need to approximate the latent score x_t^1. This can be done by consistently estimating the eigenfunction ψ_1. The latter can be achieved via FPCA of a sample version of the long-run covariance operator $\mathbf{C}_n$ defined in (2.4). Let $\overline{X}_n = \frac{1}{n}\sum_{t=1}^n X_t$, and

$$\mathbf{R}_{n,k} = \begin{cases} \frac{1}{n}\sum_{t=1}^{n-k}\left(X_t - \overline{X}_n\right) \otimes \left(X_{t+k} - \overline{X}_n\right), & k \geqslant 0, \\ \frac{1}{n}\sum_{t=1}^{n-|k|}\left(X_{t+|k|} - \overline{X}_n\right) \otimes \left(X_t - \overline{X}_n\right), & k < 0, \end{cases}$$

where $|k| \leqslant q$, and q is a tuning parameter satisfying some mild restrictions. Define

$$\overline{\mathbf{C}}_n = \sum_{t=1}^{q}\sum_{s=1}^{q}\mathbf{R}_{n,t-s} = \sum_{|k|\leqslant q}(q-|k|)\mathbf{R}_{n,k}, \quad \widetilde{\mathbf{C}}_n = \frac{1}{q^{H_*}}\overline{\mathbf{C}}_n, \tag{3.1}$$

where H_* is defined as in (2.5). Note that $\widetilde{\mathbf{C}}_n$ is a natural extension of the classic heteroskedasticity and autocorrelation consistent estimator due to Hannan (1957) and the nonparametric spectral density estimation literature, and subsequently heavily developed in the econometric literature, as well as in inference on the memory parameter by Robinson (2005) and Abadir, Distaso and Giraitis (2009). However, as H_* in the normalisation factor depends on the unknown parameter H_0, we must replace H_0 by an estimate with sufficient convergence rate. As $\overline{\mathbf{C}}_n$ is proportional to $\widetilde{\mathbf{C}}_n$, the sample eigenfunctions obtained via FPCA of $\overline{\mathbf{C}}_n$ are the same as via FPCA of $\widetilde{\mathbf{C}}_n$. A significant advantage of $\overline{\mathbf{C}}_n$ is that we do not require any prior information or a preliminary estimate of H_0 and thus it is applicable no matter whether the underlying functional time series are LRD, SRD or ND. Let $\overline{\psi}_1$ be the eigenfunction of $\overline{\mathbf{C}}_n$ corresponding to the maximum eigenvalue.

Proposition 2 in Section 3.2 below shows that $\overline{\psi}_1$ is consistent for ψ_1, so it is sensible to approximate the first score x_t^1 by $\overline{x}_t^1 = \langle X_t, \overline{\psi}_1 \rangle$. Following the LW estimation procedure with x_t^1 replaced by $\overline{x}_t^1$, we can obtain a feasible estimator of H_0. Specifically, as in (2.8), we define the periodogram of $\overline{x}_t^1$ at frequency λ as

$$\overline{I}_1(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} \overline{x}_t^1 e^{it\lambda} \right|^2, \tag{3.2}$$

and analogously to Section 2.2, we consider the estimate

$$\overline{H} = \arg\min_{H \in \Theta} \overline{R}(H) = \arg\min_{H \in \Theta} \left\{ \log \overline{G}(H) - \frac{2H-1}{m} \sum_{j=1}^{m} \log \lambda_j \right\} \tag{3.3}$$

with $\overline{G}(H) = \frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2H-1} \overline{I}_1(\lambda_j)$. Section 3.2 below will show that the asymptotic results stated in Proposition 1 still hold for the feasible LW estimate $\overline{H}$.

3.2 Main asymptotic theory

We first give some restrictions on the coefficients b_j which are needed to prove consistency for $\widetilde{\mathbf{C}}_n$ and $\overline{\psi}_1$ in Proposition 2 below.

Assumption 3. *(i) For SRD, (2.2) holds and* $\sum_{j=0}^{\infty} \sum_{k=j}^{\infty} |b_k| < \infty$.

(ii) For LRD, (2.3) holds for $1/2 < H_0 < 1$.

(iii) For ND, $b_j = j^{H_0 - 3/2}(1 + O(j^{\delta_2}))$ *with* $\delta_2 < H_0 - 1/2$ *for* $0 < H_0 < 1/2$, *and* $\sum_{j=0}^{\infty} b_j = 0$.

With Assumptions 1 and 3 above, we can derive the following proposition, which will play a crucial role in deriving the main asymptotic result (Theorem 1).

Proposition 2. *Let Assumptions 1 and 3 be satisfied,* $q = o(n^{1/2})$, *and* $\mathsf{E}\left[\|\eta_t\|^4\right] < \infty$. *Then (i)* $\left\| \widetilde{\mathbf{C}}_n - \mathbf{C} \right\| = o_P(1)$; *and (ii) if also that* $0 \leqslant \rho_2 < \rho_1 < \infty$, $\left\| \overline{\psi}_1 - \tau_1 \psi_1 \right\| = o_P(1)$ *with* $\tau_1 = \mathrm{sign}\left(\langle \overline{\psi}_1, \psi_1 \rangle \right)$.

Remark 3. As the functional linear process X_t is allowed to be LRD, SRD or ND, Proposition 2 has wider applicability than existing results developed for LRD (Proposition 2 in Li, Robinson and Shang, 2020) or SRD (Theorem 4.1 in Hörmann and Kokoszka, 2010). Proposition 2(ii) is critical to ensure that the feasible LW has the same asymptotic distribution as the infeasible LW. Furthermore, by assuming that $0 \leqslant \rho_{p+1} < \rho_p < \cdots < \rho_1 < \infty$ for a positive integer p, we have $\left\| \overline{\psi}_k - \tau_k \psi_k \right\| = o_P(1)$ with $\tau_k = \mathrm{sign}\left(\langle \overline{\psi}_k, \psi_k \rangle \right)$ for any $1 \leqslant k \leqslant p$.

Theorem 1. *Suppose that Assumptions 1, 2* and 3 are satisfied,* $q = o(n^{1/2})$*, and* $\mathsf{E}\left[\|\eta_t\|^4\right] < \infty$*. Then (i)* $\overline{H}$ *is weakly consistent; and (ii)*

$$m^{1/2}\left(\overline{H} - H_0\right) \xrightarrow{d} N(0,\ 1/4), \quad \text{as } n \to \infty. \tag{3.4}$$

Remark 4. Replacing ψ_1 by its estimate $\overline{\psi}_1$ thus has negligible impact. This is mainly due to application of Proposition 2(ii) and Assumption 2*. As in Proposition 1(ii), the asymptotic variance in (3.4) is free of any nuisance parameter, facilitating statistical inference of the parameter H_0.

4 Numerical studies

We now present both simulation and empirical studies to examine numerical performance of the proposed feasible LW estimation method in finite samples.

4.1 Monte-Carlo simulations

We use an algorithm of Davies and Harte (1987) to simulate functional time series observations. Let X_t be a "fractional noise" process with autocovariance $\gamma_j = \frac{1}{2}\left(|j+1|^{2H_0} - 2|j|^{2H_0} + |j-1|^{2H_0}\right)$, where $H_0 = 0.2, 0.35, 0.5, 0.65$ and 0.8. These parameter values are chosen to reflect ND, SRD and LRD properties. For each n, let $g_k := g_{n,k}$, $k = 0, 1, \cdots, 2n-1$, be the discrete Fourier transform of the real sequence $\{\gamma_0, \gamma_1, \cdots, \gamma_{n-1}, \gamma_n, \gamma_{n-1}, \cdots, \gamma_1\}$, i.e.,

$$g_k = \gamma_0 + 2\sum_{j=1}^{n-1} \gamma_j \cos\left(\frac{\pi k j}{n}\right) + \gamma_n \cos(k\pi), \quad k = 0, 1, \cdots, n-1,$$

and $g_k = g_{2n-k}$ for $k = n, \cdots, 2n-1$. Let (η_t) be an i.i.d. standard Brownian motion sequence over $[0,1]$ and define

$$\widetilde{\eta}_t = \begin{cases} \eta_t, & 1 \leqslant t \leqslant n-1, \\ \eta_{2n-t}, & n+1 \leqslant t \leqslant 2n-1, \\ \sqrt{2}\eta_t, & t = 0, n. \end{cases}$$

Then we construct

$$X_t = \frac{1}{2n^{1/2}}\left[\sqrt{2}\eta_0 g_0^{1/2} + \sqrt{2}\eta_n g_n^{1/2} + 2\sum_{k=1}^{n-1} \eta_k g_k^{1/2} \cos\left(\frac{\pi k t}{n}\right)\right], \quad 0 \leqslant t \leqslant n. \tag{4.1}$$

We take $n = 250, 500, 1000$ with 2000 replications.

We choose $m = n^{4/5}$ that lies between the lower and upper bounds recommended by Lobato and Robinson (1998), and obtain the LW estimates $\overline{H}_b$, $b = 1, \cdots, 2000$. We compute the Monte-Carlo bias and mean squared error (MSE). The results are reported in Table 1. Biases are always positive for LRD cases ($H_0 = 0.65, 0.80$) and negative otherwise, and are slowly decreasing (in absolute value) in n, except for the SRD case ($H_0 = 0.50$) where they slightly increase. MSE decreases slowly in n. Overall, bias and MSE are somewhat worst in ND cases ($H_0 = 0.20, 0.35$).

Table 1: Monte-Carlo bias and MSE of the feasible LW based on 2000 replications.

H_0	$n = 250$		$n = 500$		$n = 1000$	
	Bias	MSE	Bias	MSE	Bias	MSE
0.20	-0.0858	0.0107	-0.0776	0.0080	-0.0692	0.0059
0.35	-0.0333	0.0046	-0.0294	0.0028	-0.0256	0.0017
0.50	-0.0020	0.0033	-0.0036	0.0019	-0.0041	0.0011
0.65	0.0197	0.0037	0.0131	0.0021	0.0087	0.0011
0.80	0.0379	0.0048	0.0262	0.0026	0.0183	0.0014

We next use the normal approximation in Theorem 1(ii) to conduct statistical inference of the memory parameter. Specifically, for each replication, we construct a nominal $100(1-\alpha)\%$ confidence interval for H_0 as follows:

$$\left(\overline{H}_b - m^{-1/2} z_{\alpha/2}/2,\ \ \overline{H}_b + m^{-1/2} z_{\alpha/2}/2\right),$$

where z_α denotes the upper α-quantile of the standard normal distribution and $b = 1, \cdots, 2000$. For $H_0 = 0.35, 0.5, 0.65$, we report in Table 2 the empirical coverage probabilities for $\alpha = 0.05, 0.01, 0.001$. The coverage probabilities tend to be too small, especially in ND cases, where they actually get markedly worse with increasing n.

4.2 Application to monthly sea surface temperatures

We next consider a time series of average monthly sea surface temperature from January, 1950 to December, 2019, available online at `https://www.cpc.ncep.noaa.gov/data/indices/ersst5.nino.mth.81-10.ascii`. These temperatures are measured by moored buoys in the "Niño region". The function support $\mathbb{C}$ is the time interval between January and December in each calendar year, and a linear interpolation algorithm is used to produce time series of continuous

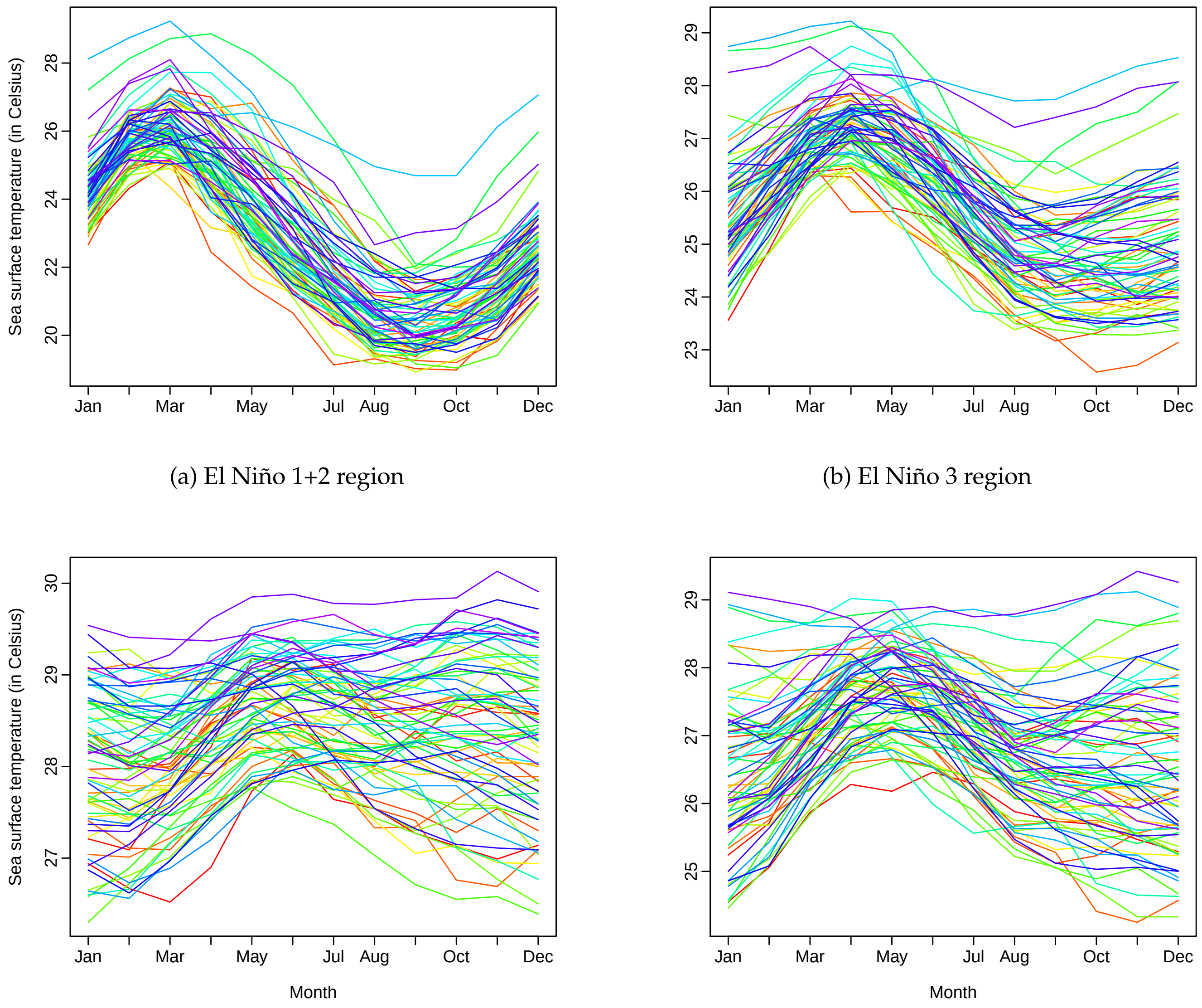
Table 2: Empirical coverage probabilities for $H_0 = 0.35, 0.50, 0.65$ and $\alpha = 0.05, 0.01$ and 0.001

		H_0		
$1-\alpha$	n	0.35	0.50	0.65
0.95	250	0.8900	0.9370	0.9245
	500	0.8785	0.9330	0.9300
	1000	0.8625	0.9365	0.9360
0.99	250	0.9590	0.9855	0.9795
	500	0.9535	0.9820	0.9835
	1000	0.9500	0.9800	0.9830
0.999	250	0.9910	0.9975	0.9965
	500	0.9860	0.9960	0.9980
	1000	0.9870	0.9955	0.9970

functions. Thus $n = 70$, a relatively small sample size for estimates with nonparametric rate. We chose $m = n^{4/5} \approx 30$ in the feasible LW, but with such a small n one cannot be confident that any m would achieve modest bias or imprecision. In Figure 1, we present rainbow plots of the monthly sea surface temperatures for four El Niño regions. The functional stationarity test of Horváth, Kokoszka and Rice (2014) never rejects the null of stationarity with large p-values in Table 3. Applying the feasible LW estimate to the first set of the estimated functional principal component scores, we obtain the estimated memory parameters together with nominal 95% confidence intervals for each El Niño region, reported in Table 3. The rather large confidence intervals, reflecting the smallness of m, make it difficult to draw conclusions about whether LRD, SRD or ND assumption applies for El Niño 1+2, 3 and 3+4 regions, whereas the result for region 4 is consistent with LRD. Note that all the upper bounds of the confidence intervals are smaller than the boundary value of nonstationarity, confirming the result of the functional stationarity test.

Table 3: For monthly sea surface temperature data from four El Niño regions with various coordinates: test p-values and estimates of the memory parameter with confidence intervals (CI)

region	coordinate	p-value	$\overline{H}$	95%-CI
Niño 1+2 region	$0-10°$ South, $90-80°$ West	0.647	0.5810	(0.4021, 0.7599)
Niño 3 region	5° North - 5° South, $150-90°$ West	0.609	0.5252	(0.3463, 0.7041)
Niño 4 region	5° North - 5° South, 160° East - 150° West	0.505	0.7299	(0.5510, 0.9088)
Niño 3+4 region	5° North - 5° South, $170-120°$ West	0.731	0.5032	(0.3243, 0.6821)



(a) El Niño 1+2 region

(b) El Niño 3 region

(c) El Niño 4 region

(d) El Niño 3+4 region

Figure 1: Rainbow plots for displaying monthly sea surface temperature at four El Niño regions from January, 1950 to December, 2019.

5 Conclusions

In this paper we have introduced feasible LW estimates of the memory parameter for stationary functional time series which are LRD, SRD or ND. Under regularity conditions which restrictively require constancy of the memory parameter across function support, we derive asymptotic theory including weak consistency and asymptotic normality, comparable to the classic time series LW asymptotic theory (Robinson, 1995a). As a crucial preliminary step in the estimation procedure, we use FPCA to estimate the leading eigenfunction corresponding to the maximum eigenvalue of the estimated long-run covariance operator, and subsequently obtain an approximation to the (latent) first score process. Monte-Carlo simulations find the LW estimate performs well in finite samples, and an empirical example is included.

Acknowledgements

The authors would like to thank two reviewers for the comments, which helped to improve the paper. The first author is partially supported by the National Natural Science Foundation of China (No. 72033002).

Supplemental materials

The online supplemental materials contain the detailed proofs of the main asymptotic theorems together with some technical lemmas.

Data availability statement

The link for the empirical data set is provided in Section 4.2.

References

Abadir, K., Distaso, W. and Giraitis, L. (2009). Two estimators of the long-run variance: beyond short memory. *Journal of Econometrics* **150**, 56–70.

Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman & Hall, New York.

Berkes, I., Horváth, L. and Rice, G. (2013). Weak invariance principles for sums of dependent random functions. *Stochastic Processes and Their Applications* **123**, 385–403.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer, New York.

Characiejus, V. and Račkauskas, A. (2014). Operator self-similar processes and functional central limit theorems. *Stochastic Processes and Their Applications* **124**, 2605–2627.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar process. *Annals of Statistics* **17**, 1749–1766.

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

Düker, M. (2018). Limit theorems for Hilbert space-valued linear processes under long range dependence. *Stochastic Processes and Their Applications* **128**, 1439–1465.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York.

Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics* **14**, 517–532.

Giraitis, L., Koul, H. and Surgailis, D. (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press, London.

Hannan, E. J. (1957). The variance of the mean of a stationary process. *Journal of the Royal Statistical Society: Series B* **19**, 282–285.

Hörmann, S., Kidzinski, L. and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society: Series B* **77**, 319–348.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Annals of Statistics* **38**, 1845–1884.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer, New York.

Horváth, L., Kokoszka, P. and Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B* **75**, 103–122.

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics* **179**, 66–82.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley Series in Probability and Statistics.

Künsch, H. R. (1987). Statistical aspects of self-similar processes. In *Proceedings of the First World Congress of the Bernoulli Society* (Y. A. Prohorov and V. V. Sazonov, eds.), VNU Science Press, Utrecht, pp. 67–74.

Li, D., Robinson, P. M. and Shang, H. L. (2020). Long-range dependent curve time series. *Journal of the American Statistical Association* **115**, 957–971.

Lobato, I. N. and Robinson, P. M. (1998). A nonparametric test for I(0). *Review of Economic Studies* **65**(3), 475–495.

Meyer, M., Paparoditis, E. and Kreiss, J. P. (2020). Extending the validity of frequency domain bootstrap methods to general stationary processes. *Annals of Statistics* **48**, 2404–2427.

Panaretos, V. M. and Tavakoli, S. (2013). Fourier analysis of stationary time series in function space. *Annals of Statistics* **41**, 568–603.

Phillips, P. C. B. and Shimotsu, K. (2004). Local Whittle estimation in nonstationary and unit root cases. *Annals of Statistics* **32**, 656–692.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis* (2nd edition). Springer, New York.

Robinson, P. M. (1995a). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* **23**, 1630–1661.

Robinson, P. M. (1995b). Log-periodogram regression of time series with long range dependence. *Annals of Statistics* **23**, 1048–1072.

Robinson, P. M. (2003). *Time Series with Long Memory*. Oxford University Press, Oxford.

Robinson, P. M. (2005). Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction. *Econometric Theory* **21**, 171–180.

Robinson, P. M. (2008). Multiple local Whittle estimation in stationary systems. *Annals of Statistics* **36**, 2508–2530.

Sheather S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society Series B* **53**, 683–690.

Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* **20**, 87–127.