

Han Lin Shang*

A Comparison of Hurst Exponent Estimators in Long-range Dependent Curve Time Series

<https://doi.org/10.1515/jtse-2019-0009>

Received February 10, 2019; accepted March 17, 2020

Abstract: The Hurst exponent is the simplest numerical summary of self-similar long-range dependent stochastic processes. We consider the estimation of Hurst exponent in long-range dependent curve time series. Our estimation method begins by constructing an estimate of the long-run covariance function, which we use, via dynamic functional principal component analysis, in estimating the orthonormal functions spanning the dominant sub-space of functional time series. Within the context of functional autoregressive fractionally integrated moving average (ARFIMA) models, we compare finite-sample bias, variance and mean square error among some time- and frequency-domain Hurst exponent estimators and make our recommendations.

Keywords: curve process, dynamic functional principal component analysis, functional ARFIMA, long-run covariance, long-range dependence

1 Introduction

In univariate time series analysis, long memory was brought to prominence by Hurst (1951) and Mandelbrot (1963), and it has subsequently received extensive attention in the literature (see, e. g., Beran 1994; Doukhan, Oppenheim, and Taqqu 2003; Robinson 2003; Embrechts and Maejima 2002; Palma 2007). Of importance in analyzing and modeling long-memory univariate time series is estimating the strength of the long-memory dependence. There are two measures commonly used: The parameter H , known as the Hurst exponent or self-similarity parameter (Mandelbrot and van Ness 1968) and the fractional integration parameter, d , arises from the generalization of autoregressive fractionally integrated moving average (ARFIMA (p, d, q)) models from integer to non-integer values of the integration parameter d . The two parameters are closely related through the simple formula $H = d + \frac{1}{2}$.

***Corresponding author: Han Lin Shang, Ph.D.**, Australian National University, Research School of Finance, Actuarial Studies and Statistics, Level 4, Building 26C, Kingsley St, Acton, Canberra, Australian Capital Territory, 2601, Australia; and Department of Actuarial Studies and Business Analytics, Level 7, 4 Eastern Rd, Macquarie University, NSW, 2109, Australia, E-mail: hanlin.shang@anu.edu.au. <https://orcid.org/0000-0003-1769-6430>

In univariate time series analysis, a number of Hurst exponent estimators have been developed, and theoretical results on the asymptotic properties of various estimators have been obtained. Because the finite-sample properties of these estimators can be quite different from their asymptotic properties, several authors considered an empirical comparison of estimators of H and d . Nine estimators were discussed in some detail by Taqqu, Teverovsky, and Willinger (1995) who performed an empirical investigation of these estimators for a single series length of 10,000 data points, five values of both H and d , and 50 replications. Teverovsky and Taqqu (1997) showed in a simulation study that the differenced variance estimator was unbiased for five values of H (0.5, 0.6, 0.7, 0.8 and 0.9) for series with 10,000 observations whereas the aggregated variance estimator was downwards biased. Jensen (1999) presented a comparison of two estimators based on wavelets, and a Geweke-Porter-Hudak (GPH) estimator for four series lengths (2^7 , 2^8 , 2^9 , 2^{10} observations), five values of d and 1,000 replications. Jeong et al. (2007) performed a comparison of six estimators on simulated fractional Gaussian noise with 2^{15} observations, five values of H and 100 replications.

Long-memory functional time series analysis was recently studied by Li, Robinson, and Shang (2019), who proposed an R/S estimation method for determining long-memory parameter in a functional ARFIMA model, where observations are temporally dependent continuous functions, for example, age-specific fertility rate improvement observed over the years (e. g., Chiou and Müller 2009; Hyndman and Ullah 2007). The functional ARFIMA model can be viewed as a generalization of many parametric models. For example, Bosq (2000) and Bosq and Blanke (2007) provided the functional autoregressive of order 1 (FAR(1)) and derived one-step-ahead forecasts that are based on a regularized form of the Yule-Walker equations. Later, FAR(1) was extended to FAR(p), where the order p can be determined via a sequential hypothesis testing procedure (Kokoszka and Reimherr 2013). Aue, Norinho, and Hörmann (2015) proposed a forecasting method based on vector autoregressive (VAR) forecasts of principal component scores. The method of Aue, Norinho, and Hörmann (2015) can also be viewed as an extension of Hyndman and Shang (2009), where principal component scores are forecast via a univariate time series forecasting method. Klepsch and Klüppelberg (2017) considered the functional moving average (FMA) process and introduced an innovation algorithm to obtain the best linear predictor. Klepsch, Klüppelberg, and Wei (2017) extended the VAR model to vector autoregressive moving average model for modeling and forecasting principal component scores, which can be viewed as a simpler estimation approach of the functional autoregressive moving average. Aue and Klepsch (2017) showed the equivalent relationship between FMA and vector moving average.

A central issue in functional time series analysis is to model the temporal dependence of the functional observations accurately. Following the early work of Li, Robinson, and Shang (2019), we compare the finite-sample estimation accuracy of several Hurst exponent estimators in functional ARFIMA models. Our method constructs an estimate of the long-run covariance function, which we use, via dynamic functional principal component analysis, in estimating the orthonormal functions spanning the dominant sub-space of the curves. Based on the first set of principal component scores, we apply several univariate time series Hurst exponent estimators, and compare their estimation accuracy in terms of bias, variance and mean square error. Our goal is to provide some practical guidance on the method that provides the best estimation accuracy of the Hurst exponent.

The remainder of the paper is outlined as follows. In Section 2, we present two methods for estimating long-run covariance function, from which the dominant set of principal component scores can be obtained. In Section 3, we revisit some long-memory univariate time series estimators for estimating the Hurst exponent. In Section 4, we compare the estimation accuracy of various estimators and make our recommendation.

2 Dynamic Functional Principal Component Analysis

2.1 Estimation of the Long-Run Covariance Function

A time series of functions can be denoted as $\{\mathcal{X}_t, t \in \mathbb{Z}\}$, where $\mathbb{Z} = \{t : t = 0, \pm 1, \dots\}$ and each $\mathcal{X}_t$ is a random function of a stochastic process $\mathcal{X}(\omega)$ where $\omega \in \mathcal{I} \subset \mathbb{R}$ represents a continuum bounded within a finite interval of the real line. Further, let $\{\mathcal{X}_t(\omega)\}_{t \in \mathbb{Z}}$ be a stationary and ergodic functional time series. For a stationary functional time series, the long-run covariance operator is defined as

$$C(\omega, \nu) = \sum_{\ell=-\infty}^{\infty} \gamma_{\ell}(\omega, \nu)$$

$$\gamma_{\ell}(\omega, \nu) = \text{cov}[\mathcal{X}_0(\omega), \mathcal{X}_{\ell}(\nu)],$$

and is a well-defined element of $\mathcal{L}^2(\mathcal{I})^2$ for a compact support interval $\mathcal{I}$, under mild weak dependence and moment conditions. By assuming $\mathcal{X}$ is a continuous and square-integrable function, the function C induces the kernel operator $\mathcal{L}^2(\mathcal{I}) \mapsto \mathcal{L}^2(\mathcal{I})$. Through right integration, C defines a Hilbert-Schmidt integral operator on $\mathcal{L}^2(\mathcal{I})$ given by

$$\mathcal{C}(\mathcal{X})(\omega) = \int_{\mathcal{I}} C(\omega, \nu) \mathcal{X}(\nu) d\nu,$$

whose eigenvalues and eigenfunctions are related to the dynamic functional principal components defined in Hörmann, Kidziński, and Hallin (2015).

In practice, we need estimate $C(\omega, \nu)$ from a finite sample $\mathcal{X}(\omega) = [\mathcal{X}_1(\omega), \dots, \mathcal{X}_n(\omega)]$. Given its definition as a bi-infinite sum, a natural estimator of C is

$$\widehat{C}_n(\omega, \nu) = \frac{1}{n^{3-2\alpha}} \sum_{|\ell|=0}^{|\ell| \leq n} (n - |\ell|) \widehat{y}_\ell(\omega, \nu), \quad (1)$$

where $\alpha = \frac{3}{2} - H$ is the so-called memory parameter, ℓ denotes a lag variable, and

$$\widehat{y}_\ell(\omega, \nu) = \begin{cases} \frac{1}{n} \sum_{j=1}^{n-\ell} [\mathcal{X}_j(\omega) - \overline{\mathcal{X}}(\omega)] [\mathcal{X}_{j+\ell}(\nu) - \overline{\mathcal{X}}(\nu)] & \text{if } \ell \geq 0; \\ \frac{1}{n} \sum_{j=1-\ell}^n [\mathcal{X}_j(\omega) - \overline{\mathcal{X}}(\omega)] [\mathcal{X}_{j+\ell}(\nu) - \overline{\mathcal{X}}(\nu)] & \text{if } \ell < 0. \end{cases}$$

is an estimator of $y_\ell(\omega, \nu)$. In the case of stationary short-memory functional time series, it is known that $\alpha = 1$. From (1), the estimated long-run covariance is obtained by summing all autocovariance functions with linearly decreasing weights. Let W denote the number of grid points in a curve. In Li, Robinson, and Shang (2019), they consider $\ell = \min(n, W)$. For instance, when $W \geq n$, all finite-order lags are utilized. To estimate the value of α , Li, Robinson, and Shang (2019) applied the rescaled range (R/S) estimator of Hurst (1951) to the first set of dynamic principal component scores obtained from eigendecomposition of

$$\widehat{\widehat{C}}_n(\omega, \nu) = \sum_{|\ell|=0}^{|\ell| \leq n} (n - |\ell|) \widehat{y}_\ell(\omega, \nu),$$

since $\frac{1}{n^{3-2\alpha}}$ in (1) is a constant and it does not affect the estimation of the orthonormal functions spanning the dominant sub-space of functional time series.

2.1.1 Kernel Sandwich Estimator

Another long-run covariance estimator is the kernel Sandwich estimator inspired by Andrews (1991) and Andrews and Monahan (1992). It is given by

$$\widehat{C}_{h,q}(\omega, \nu) = \sum_{\ell=-\infty}^{\infty} W_q\left(\frac{\ell}{h}\right) \widehat{y}_\ell(\omega, \nu), \quad (2)$$

where h is called the bandwidth parameter and W_q is a symmetric weight function with bounded support of order q . The kernel sandwich estimator in (2) was introduced in Panaretos and Tavakoli (2012), Horváth, Kokoszka, and Reeder (2013), Rice and Shang (2017), Kokoszka and Reimherr (2017, Chapter 8.5), among

others. As with any kernel estimator, the crucial part is on the estimation of bandwidth parameter h . It can be selected through a data-driven approach, such as the plug-in algorithm of Rice and Shang (2017). The plug-in bandwidth selection method can be summarized as:

1) Compute pilot estimates of $C^{(p)}$, for $p = 0$ and initial order of kernel function q_1 :

$$\widehat{C}_{h_1, q_1}^{(p)}(u, s) = \sum_{\ell=-\infty}^{\infty} W_{q_1}\left(\frac{\ell}{h_1}\right) |\ell|^p \widehat{\gamma}_\ell(u, s),$$

that utilize an initial bandwidth choice $h_1 = h_1(n)$, and weight function W_{q_1} of order q_1 .

2) As established in Berkes, Horváth, and Rice (2016), estimate c_0 by

$$\widehat{c}_0(h_1, q_1, q) = \left(2q \|w \widehat{C}_{h_1, q_1}^{(q)}\|^2\right)^{1/(1+2q)} \left(\left(\|\widehat{C}_{h_1, q_1}^{(0)}\|^2 + \left(\int_0^1 \widehat{C}_{h_1, q_1}^{(0)}(u, u) du \right)^2 \right) \int_{-\infty}^{\infty} W_{q_1}^2(x) dx \right)^{-1/(1+2q)},$$

where q denotes the final order of kernel function, w is a constant depending on the final order of kernel function, and $\int_{-\infty}^{\infty} W_{q_1}^2(x) dx$ is a weight depending on the initial order of kernel function. A list of w and $\int_{-\infty}^{\infty} W_{q_1}^2(x) dx$ values is presented in Table 1.

For the initial kernel function, Rice and Shang (2017) recommend to use flat-top kernel function, i. e., $q_1 = \infty$. For the final kernel function, Rice and Shang (2017) recommend to use Bartlett kernel function, i. e., $q = 1$. Further, there exists w satisfying $0 < w = \lim_{x \rightarrow 0} |x|^{-q} (1 - W_q(x)) < \infty$.

3) Use the bandwidth

$$\widehat{h}_{\text{opt}}(h_1, q_1, q) = \widehat{c}_0(h_1, q_1, q) n^{1/(1+2q)}$$

in the definition of $\widehat{C}_{h, q}$ in (2).

2.2 Dynamic Functional Principal Component Decomposition

From the long-run covariance $C(w, v)$, we apply functional principal decomposition to extract the functional principal components and their associated scores. With Karhunen-Loève expansion, a stochastic process $\mathcal{X}$ can be expressed as

$$\mathcal{X}_t(\omega) = \mu(\omega) + \sum_{j=1}^{\infty} \beta_{t,j} \phi_j(\omega),$$

Table 1: A list of w and $\int_{-\infty}^{\infty} W_{q_1}^2(x) dx$ values.

Kernel function	W	$\int_{-\infty}^{\infty} W_{q_1}^2(x) dx$
Bartlett	1	2/3
Parzen	6	0.539285
Tukey-Hanning	$\pi^2/4$	3/4
Quadratic Spectral	$18 \times \pi^2/125$	1
Flat-top		4/3

where $\mathcal{X}_t^c(\omega) = \mathcal{X}_t(\omega) - \mu(\omega)$ and $\beta_{t,j}$ is an uncorrelated random variable with zero mean and unit variance. The principal component score $\beta_{t,j}$ is given by the projection of $\mathcal{X}_t^c$ in the direction of the j th eigenfunction ϕ_j , i. e., $\beta_{t,j} = \langle \mathcal{X}_t^c(\omega), \phi_j(\omega) \rangle$. The scores $\beta_j = (\beta_{1,j}, \dots, \beta_{n,j})$ constitute an uncorrelated sequence of random variables with zero mean and variance λ_j which is the j th eigenvalue. They can be interpreted as the weights of the contribution of the functional principal components $\phi_j(\omega)$ to $\mathcal{X}_t^c(\omega)$.

Since the long-run covariance $C(w, v)$ is unknown, the population eigenvalues and eigenfunctions can only be approximated through realizations of $\mathcal{X}(\omega)$. The sample mean and sample covariance are given by

$$\bar{\mathcal{X}}(\omega) = \frac{1}{n} \sum_{t=1}^n \mathcal{X}_t(\omega),$$

$$\hat{C}(\omega, v) = \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{\phi}_j(\omega) \hat{\phi}_j(v),$$

where $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots \geq 0$ are the sample eigenvalues of $\hat{C}(\omega, v)$, and $[\hat{\phi}_1(\omega), \hat{\phi}_2(\omega), \dots]$ are the corresponding orthogonal sample eigenfunctions. The realizations of the stochastic process $\mathcal{X}$ can be written as

$$\mathcal{X}_t(\omega) = \hat{\mu}(\omega) + \sum_{j=1}^{\infty} \hat{\beta}_{t,j} \hat{\phi}_j(\omega), \quad t = 1, 2, \dots, n,$$

where $\hat{\mu}(\omega) = \frac{1}{n} \sum_{t=1}^n \mathcal{X}_t(\omega)$, and $\hat{\beta}_{t,j}$ is the j th estimated principal component score for the t th time period.

Hörmann, Kidziński, and Hallin (2015) showed that kernel Sandwich estimator in (2) is a consistent estimator of the true and unknown long-run covariance, and estimated functional principal components and principal component scores extracted from the estimated long-run covariance are also consistent.

3 Hurst Exponent Estimators

Let the first set of estimated dynamic principal component scores be $\hat{\beta}_1 = \{\hat{\beta}_{1,1}, \hat{\beta}_{2,1}, \dots, \hat{\beta}_{n,1}\}$. Since we consider the first set of scores, we shall replace $\hat{\beta}_1$ by β hereafter. In Li, Robinson, and Shang (2019), they also consider a L_2 norm of multiple sets of scores and find the estimation results remain similar. Due to space constraints, we present our results based on the first set of principal component scores. With the univariate time series of scores β , we evaluate and compare some Hurst exponent estimators from long-memory univariate time-series literature.

The Hurst exponent can be estimated either via time- or frequency-domain based estimators. These estimators can be divided into parametric and semi-parametric ones. The theory of parametric estimators was developed by Dahlhaus (1989) and Fox and Taqqu (1986). Semiparametric estimators of the memory parameter have become popular since they do not require knowing the specific form of the short-memory structure. They are based on the periodograms of the series, and can be categorized into two types: the log-periodogram estimator first proposed by Geweke and Porter-Hudak (1983) and the local-Whittle estimator which is credited to Künsch (1987) and further developed by Robinson (1995a). The log-periodogram estimator is akin to the ordinary least squares and the local-Whittle estimator to the maximum likelihood estimator in the frequency domain.

3.1 Time-domain Based Estimators

In Sections 3.1.1–3.1.5, we present five methods based on a simple linear regression model. In Sections 3.1.6 and 3.1.7, we present two methods based on the R/S estimator.

3.1.1 Aggregated Variance Estimator

The aggregated variance estimator is based on the property of self-similar processes that variances of the aggregated processes decrease at the rate m^{2H-2} as the block size m increases (e. g., Beran 1994; Taqqu, Teverovsky, and Willinger 1995; Teverovsky and Taqqu 1997, Section 4.4). Recall that for a long-range dependent linear process,

$$\text{Var}(\bar{\beta}) \sim Cm^{2H-2},$$

where C is a constant. Consequently,

$$\log_{10} [\text{Var}(\bar{\beta})] \approx \log_{10} C + (2H - 2)\log_{10} m$$

With the predictor variable of $\log_{10} m$ and the response variable of $\log_{10} [\text{Var}(\bar{\beta})]$, we apply a simple linear regression to obtain an estimate of the slope parameter. For instance, one may define the following procedure:

- 1) Divide the time series β into K non-overlapping blocks with block size m and then average within each block, that is considered the aggregated series

$$\bar{\beta}^{(m)}(k) = \frac{1}{m} \sum_{t=(k-1)m+1}^{km} \beta_t, \quad (3)$$

where $k = 1, \dots, K$ denotes a block index and $K = n/m \geq 1$ denotes the number of blocks.

- 2) Compute the overall mean

$$\bar{\beta}^{(m)} = \frac{1}{K} \sum_{k=1}^K \bar{\beta}^{(m)}(k)$$

- 3) For a given m , compute the sample variance of $\bar{\beta}^{(m)}(k)$ as

$$\widehat{\text{Var}}(\bar{\beta}^{(m)}) = \frac{1}{K} \sum_{k=1}^K \left[\bar{\beta}^{(m)}(k) \right]^2 - (\bar{\beta}^{(m)})^2. \quad (4)$$

- 4) Heuristically, when m grows, $\text{Var}(\bar{\beta}^{(m)}) \sim Cm^{2H-2}$. Thus, $\text{Var}(\bar{\beta}^{(m)})$ grows approximately at the rate m^{2H-2} . For different values of $m = 1, \dots, M$, compute (3) and (4) to obtain $\widehat{\text{Var}}(\bar{\beta}) = \{\widehat{\text{Var}}\bar{\beta}^{(1)}, \dots, \widehat{\text{Var}}\bar{\beta}^{(M)}\}$. It is recommended by Taqqu, Teverovsky, and Willinger (1995) and Teverovsky and Taqqu (1997) to choose values of m that are equispaced on a logarithmic scale. Then, regress $\log_{10}[\widehat{\text{Var}}(\bar{\beta})]$ against $\log_{10}(M)$ to obtain regression coefficient $\hat{\theta}_{\text{aggvar}}$. The estimated value of H is given by

$$\hat{H}_{\text{aggvar}} = \frac{\hat{\theta}_{\text{aggvar}} + 2}{2}.$$

3.1.2 Differencing Variance Estimator

To distinguish non-stationarity from long-range dependence, we can difference the variance (see, e. g., Teverovsky and Taqqu 1997). For a given m^* , we compute the difference of the sample variance

$$\text{Var}(\bar{\beta}^{m^*}) = \widehat{\text{Var}}\bar{\beta}^{(m^*+1)} - \widehat{\text{Var}}\bar{\beta}^{(m^*)}, \quad m^* = 1, \dots, M-1. \quad (5)$$

For different values of m^* , compute (3), (4) and (5) to obtain $\widehat{\text{Var}}(\bar{\beta}) = \{\widehat{\text{Var}}\bar{\beta}^{(1)}, \dots, \widehat{\text{Var}}\bar{\beta}^{(M-1)}\}$. Then, regress $\log_{10}[\widehat{\text{Var}}(\bar{\beta})]$ against $\log_{10}(M-1)$ to obtain regression coefficient $\hat{\theta}_{\text{diffvar}}$. The estimated value of H is given by

$$\hat{H}_{\text{diffvar}} = \frac{\hat{\theta}_{\text{diffvar}} + 2}{2}.$$

3.1.3 Absolute Values of the Aggregated Series

Similar to the aggregated variance, the data are split in the same fashion, and the aggregated mean is computed from (3). Instead of computing the sample variance, one finds the sum of the absolute values of the aggregated series, namely

$$\text{abs}(\bar{\beta}^{(m)}) = \frac{1}{K} \sum_{k=1}^K |\bar{\beta}^{(m)}(k)|. \quad (6)$$

For different values of $m = 1, \dots, M$, compute (6) to obtain $\text{abs}(\bar{\beta}) = \{\text{abs}(\bar{\beta}^{(1)}), \dots, \text{abs}(\bar{\beta}^{(M)})\}$. Then, regress $\log_{10}[\text{abs}(\bar{\beta})]$ against $\log_{10}(M)$ to obtain regression coefficient $\hat{\theta}_{\text{absval}}$. The estimated value of H is given by

$$\hat{H}_{\text{absval}} = \hat{\theta}_{\text{absval}} + 1.$$

3.1.4 Higuchi's Method

Similar to the absolute values of the aggregated series, the method of Higuchi (1988) calculates the partial sums $Y(n) = \sum_{t=1}^n \beta_t$ of the time series β , and then finding the normalized length of the curve, namely

$$L(m) = \frac{n-1}{m^3} \sum_{i=1}^m \frac{m}{n-i} \sum_{k=1}^{(n-i)/m} |Y(i+km) - Y[i+(k-1)m]|, \quad (7)$$

where n is the sample size of the time series, m is a block size and $[\cdot]$ denotes the greatest integer function. Since $E[L(m)] \sim Cm^{-D}$ where $D = 2 - H$. For different values of $m = 1, \dots, M$, compute (7) to obtain $\mathbf{L} = \{L(1), \dots, L(M)\}$. Then, regress $\log_{10}(\mathbf{L})$ against $\log_{10}(M)$ to obtain regression coefficient $\hat{\theta}_{\text{Higuchi}}$. The estimated value of H is given by

$$\hat{H}_{\text{Higuchi}} = \hat{\theta}_{\text{Higuchi}} + 2.$$

3.1.5 Detrended Fluctuation Analysis (DFA)

Also known as a variance of residuals or Peng's method, DFA was introduced by Peng et al. (1994) to provide evidence of long memory in deoxyribonucleic acid (DNA) sequences. It consists of the following steps:

- 1) The data series is divided into K nonoverlapping blocks and each block with size m such that $n = mK$.
- 2) Within each of the K blocks, we regress $T_l = \sum_{t=1}^l \beta_t$ against l and estimate the variance of the residuals by

$$S_m^2(k) = \frac{1}{m} \sum_{l=(k-1)m+1}^{km} (T_l - \hat{\zeta}_{0,k} - \hat{\zeta}_{1,k}l)^2, \quad k = 1, \dots, K,$$

where $\hat{\zeta}_{0,k}$ and $\hat{\zeta}_{1,k}$ are least squares regression estimates based on the k th block.

- 3) Compute the average of the variance of the residuals

$$F^2(m) = \frac{1}{K} \sum_{k=1}^K S_m^2(k). \quad (8)$$

- 4) Heuristically, $F^2(m)$ grows at the rate m^{2H} . For different values of $m = 1, \dots, M$, compute (8) to obtain $\mathbf{F}^2 = \{F^2(1), \dots, F^2(M)\}$. Then, regress $\log_{10} \mathbf{F}^2$ against $\log_{10} M$ to obtain regression coefficient $\hat{\theta}_{\text{Peng}}$. The estimated value of H is given by

$$\hat{H}_{\text{Peng}} = \frac{\hat{\theta}_{\text{Peng}}}{2}.$$

The DFA bears a strong resemblance to the variance plot, but instead of assuming stationarity, a fitted linear trend is subtracted from each block (Beran et al. 2013). Therefore, the DFA is less sensitive to the trend exhibited in the data.

3.1.6 Rescaled Range (R/S) Estimator

The R/S estimator was introduced by Hurst (1951) for estimating the minimum capacity of a dam. The R/S estimator is one of the first methods for estimating Hurst exponent. Although many Hurst exponent estimators have better statistical properties than the R/S estimator (which, for example, is inefficient in the case of Gaussian innovations), it is a simple method that computes fast (see, e. g., Li, Robinson, and Shang 2019). Given a time series of scores β , calculation of the R/S statistic has the following steps:

- 1) Calculate the range

$$R_n = \max_{1 \leq T \leq n} \sum_{t=1}^T (\beta_t - \bar{\beta}) - \min_{1 \leq T \leq n} \sum_{t=1}^T (\beta_t - \bar{\beta}), \quad \bar{\beta} = \frac{1}{n} \sum_{t=1}^n \beta_t$$

- 2) Calculate the scale

$$S_n = \sqrt{\frac{1}{n-1} \sum_{t=1}^n (\beta_t - \bar{\beta})^2}.$$

If β_t is second-order stationary, then S_n^2 converges in probability to $\text{Var}(\beta_t)$ (Beran et al. 2013, p.410).

3) The R/S estimator may be defined by

$$\hat{H}_{RS} = \frac{\log_{10}(R_n/S_n)}{\log_{10}n}.$$

The plot of $\log_{10}(R_n/S_n)$ against $\log_{10}n$ is also known as “pox plots”.

3.1.7 Rescaled Adjusted Range Estimator

While the R/S estimator is applied to the original time series, the rescaled adjusted range estimator is implemented to the partial sum of the original time series (see, e. g., Mandelbrot 1975; Mandelbrot and Taqqu 1979; Mandelbrot and Wallis 1969). For a univariate time series of principal component scores β with the partial sum

$$Y(n) = \sum_{t=1}^n \beta_t$$

and sample variance

$$S^2(n) = \frac{1}{n} \sum_{t=1}^n \beta_t^2 - \left[\frac{1}{n} Y(n) \right]^2,$$

the rescaled adjusted range estimator is given by

$$\frac{R_n}{S_n} = \frac{1}{S(n)} \left\{ \max_{1 \leq T \leq n} \left[Y(T) - \frac{T}{n} Y(n) \right] - \min_{1 \leq T \leq n} \left[Y(T) - \frac{T}{n} Y(n) \right] \right\}.$$

Choosing logarithmically equidistant values of n , regress $\log_{10}(R_n/S_n)$ against $\log_{10}(n)$ to obtain regression coefficient $\hat{\theta}_{RAR}$. The estimated value of H is given by

$$\hat{H}_{RAR} = \hat{\theta}_{RAR}.$$

3.2 Frequency-Domain Based Estimators

3.2.1 (Smoothed) Periodogram Estimator

With a univariate time series of scores β , the periodogram can be defined as

$$w(\lambda_j) = (2\pi n)^{-1/2} \sum_{t=1}^n \beta_t \exp^{it\lambda_j}, \quad I(\lambda_j) = |w(\lambda_j)|^2, \quad (9)$$

where $\lambda_j = 2\pi j/n$ denotes the set of harmonic frequencies, $j = 1, \dots, J$ where J is a positive integer, and $i^2 = -1$. Since the periodogram is a measure of autocovariance, it can also be expressed as

$$\begin{aligned}
 I(\lambda_j) &= \frac{1}{2\pi} \sum_{s=1-n}^{n-1} R(s) \cos(s\lambda_j) \\
 &= \frac{1}{2\pi} \left\{ R(0) + 2 \sum_{s=1}^{n-1} R(s) \cos(s\lambda_j) \right\}, \quad \lambda_j \in [-\pi, \pi]
 \end{aligned}$$

where $R(s)$ denotes the sample autocovariance function, i. e.,

$$R(s) = \frac{1}{n} \sum_{t=1}^{n-s} (\beta_t - \bar{\beta})(\beta_{t+s} - \bar{\beta}), \quad s = 0, \pm 1, \dots, \pm(n-1),$$

where $\bar{\beta}$ is the sample mean of the time series of scores.

Because $I(\lambda)$ is an estimator of the spectral density, a time series with long-range dependence should have a periodogram which is proportional to $|\lambda|^{1-2H} = |\lambda|^{-2d}$ close to the origin (Taqqu, Teverovsky, and Willinger 1995). Thus, regress the logarithm of the periodogram for different values of λ against $\log_{10}(\lambda)$ to obtain regression coefficient $\hat{\theta}_{\text{per}}$. The estimated value of H is given by

$$\hat{H}_{\text{per}} = \frac{1 - \hat{\theta}_{\text{per}}}{2}.$$

As advocated by Taqqu, Teverovsky, and Willinger (1995), we use only the lowest 10% of the frequencies for the regression, since the proportionality above 10% only holds for λ close to the origin.

The frequency axis is divided into logarithmically equidistant boxes, and the periodogram values corresponding to the frequencies inside the box are averaged, to obtain smoothed periodogram. The periodogram values at very low frequencies are remained, while the rest are divided into 60 boxes (see, e. g., Taqqu, Teverovsky, and Willinger 1995). By regressing the logarithm of the smoothed periodogram against frequencies, we obtain regression coefficient $\hat{\theta}_{\text{boxper}}$. To achieve the robustness in the least square fitting, we use a robust linear model. The estimated value of H is given by

$$\hat{H}_{\text{boxper}} = \frac{1 - \hat{\theta}_{\text{boxper}}}{2}.$$

3.2.2 (Smoothed) Geweke-Porter-Hudak Estimator

In the univariate ARFIMA (p, d, q) models, Geweke and Porter-Hudak (1983) proposed a semiparametric estimator of d based on the first J periodogram ordinates given in (9). Let β be a stationary time series with spectral density

$$\begin{aligned}
f(\lambda) &= |1 - \exp^{-i\lambda}|^{-2d} f_*(\lambda) \\
&\sim |\lambda|^{-2d} f_*(\lambda) \\
&\sim C|\lambda|^{-2d}
\end{aligned} \tag{10}$$

as $\lambda \rightarrow 0$, where $-\frac{1}{2} < d < \frac{1}{2}$. Recall that the empirical estimate to the spectral density is the periodogram given in (10),

$$\log_{10} f(\lambda) \sim \log_{10} C + db(\lambda),$$

where $b(\lambda) = -2\log_{10}(\lambda)$. In practice, we replace $f(\lambda)$ by its empirical analogy $I(\lambda)$, thus

$$\log_{10} I(\lambda) \sim \log_{10} C + db(\lambda).$$

By a simple linear regression, Geweke and Porter-Hudak (1983) suggested the least-square estimator

$$\begin{aligned}
\hat{d}_{\text{GPH}} &= \frac{\sum_{j=1}^J (b_j - \bar{b}) \log_{10} I(\lambda_j)}{\sum_{j=1}^J (b_j - \bar{b})^2}, \\
&= \frac{-\frac{1}{2} \sum_{j=1}^J [\log_{10}(\lambda_j) - \overline{\log_{10}(\lambda_j)}] \log_{10} I(\lambda_j)}{\sum_{j=1}^J [\log_{10}(\lambda_j) - \overline{\log_{10}(\lambda_j)}]^2}
\end{aligned}$$

where $b_j = -2\log_{10}(\lambda_j)$, $\bar{b} = \frac{1}{J} \sum_{j=1}^J b_j$ and $\overline{\log_{10}(\lambda_j)} = \frac{1}{J} \sum_{j=1}^J \log_{10}(\lambda_j)$. Note that $\lambda_j = 2\pi j/n$ for $j = 1, \dots, J$ are the J smallest Fourier frequencies. The number J acts as a bandwidth parameter. Following Geweke and Porter-Hudak (1983), we choose $J = \sqrt{n}$.

Further, Robinson (1995b) showed that this estimator is consistent and has a central limit theorem of the form

$$\sqrt{J}(\hat{d}_{\text{GPH}} - d) \xrightarrow{d} N\left(0, \frac{\pi^2}{24}\right).$$

Reisen (1994) considered a smoothed periodogram using the Parzen lag window, for estimating the parameter d . Let $I_s(\lambda)$ denote a smoothed periodogram of the form

$$I_s(\lambda) = \frac{1}{2\pi} \sum_{s=-h}^h K\left(\frac{s}{h}\right) R(s) \cos(s \times \lambda), \quad \lambda \in [-\pi, \pi],$$

where $K(u)$ is called the lag window generator, a fixed continuous even function in the range $-1 < u < 1$, with $K(0) = 1$ and $K(-u) = K(u)$. The bandwidth parameter h

is a function of n , and it is customarily chosen as $n^{\frac{9}{10}}$. The Parzen lag window generator has the following form:

$$K(u) = \begin{cases} 1 - 6u^2 + 6|u|^3 & |u| \leq \frac{1}{2}; \\ 2(1 - |u|)^3 & |u| \leq 1; \\ 0 & |u| > 1. \end{cases}$$

The smoothed periodogram estimator can be written as

$$\begin{aligned} \hat{d}_{\text{SGPH}} &= \frac{\sum_{j=1}^J (b_j - \bar{b}) \log_{10} I_s(\lambda_j)}{\sum_{j=1}^J (b_j - \bar{b})^2}, \\ &= \frac{-\frac{1}{2} \sum_{j=1}^J [\log_{10}(\lambda_j) - \overline{\log_{10}(\lambda_j)}] \log_{10} I_s(\lambda_j)}{\sum_{j=1}^J [\log_{10}(\lambda_j) - \overline{\log_{10}(\lambda_j)}]^2}. \end{aligned}$$

3.2.3 Wavelet Estimator

This estimator computes the discrete wavelet transform, and obtains the wavelet coefficient $w_{j,k}$ associated with a mean zero $I(d)$ process with $-\frac{1}{2} < d < \frac{1}{2}$. The wavelet coefficient $w_{j,k}$ as $j \rightarrow 0$ are distributed $N(0, \sigma^2 2^{-2jd})$, where σ^2 is a finite constant (Jensen 1999). The variance $\sigma^2 2^{-2jd}$ depends on the scaling parameter j but is independent of the translation parameter k . We define

$$R(j) = \sigma^2 2^{-2jd}$$

be the wavelet coefficient's variance at scale j . Taking the logarithm transformation of $R(j)$, we obtain

$$\log_{10} R(j) = \log_{10} \sigma^2 - d \log_{10} 2^j,$$

where d can be estimated via ordinary least squares. Since $R(j)$ is a population quantity, we estimate it by the sample variance of the wavelet coefficients as

$$\hat{R}(j) = \frac{1}{2^j} \sum_{k=0}^{2^j-1} w_{j,k}^2.$$

3.2.4 Local Whittle Estimator

The local Whittle estimator is a Gaussian semiparametric estimation method to estimate the Hurst exponent based on the periodogram. It is first introduced by Künsch (1987) and later developed by Robinson (1995a), Velasco (1999) and subsequent authors. The local Whittle method does not require the specification of a

parametric model for the data. It only relies on the specification of the shape of the spectral density of the time series β .

Note that the spectral density $f(\lambda)$ of a stationary time series is usually assumed to satisfy that

$$f(\lambda) \sim G\lambda^{1-2H} = G\lambda^{-2d}, \quad \text{as } \lambda \rightarrow 0+,$$

where $0 < G < \infty$, $0 < H < 1$ and $-\frac{1}{2} < d < \frac{1}{2}$.

Define $Q(G, d)$ as the objective function

$$Q(G, d) = \frac{1}{m_\diamond} \sum_{j=1}^{m_\diamond} \left\{ \ln(G\lambda_j^{-2d}) + \frac{I(\lambda_j)}{G\lambda_j^{-2d}} \right\}, \quad (11)$$

where $\lambda_j = (2\pi j)/n$, $j = 1, \dots, m_\diamond$, and $m_\diamond$ is a positive integer satisfying $m_\diamond < n/2$, $m_\diamond \rightarrow \infty$ and $m_\diamond = o(n)$ (see, e. g., Robinson 1995a). As in Robinson (1995a), we define the estimates

$$(\hat{G}, \hat{d}) = \underset{0 < G < \infty, d \in \Theta}{\operatorname{argmin}} Q(G, d),$$

where the closed interval of admissible estimates of true value of the self-similarity measure d_0 , $\Theta = [\nabla_1, \nabla_2]$, ∇_1 and ∇_2 are numbers picked such that $-\frac{1}{2} < \nabla_1 < \nabla_2 < \frac{1}{2}$ as defined in Robinson (1995a). Alternatively, we may obtain

$$\hat{d} = \underset{d \in \Theta}{\operatorname{argmin}} R(d)$$

where

$$R(d) = \ln \hat{G}(d) - \frac{2d}{m_\diamond} \sum_{j=1}^{m_\diamond} \ln \lambda_j, \quad \hat{G}(d) = \frac{1}{m_\diamond} \sum_{j=1}^{m_\diamond} \lambda_j^{2d} I(\lambda_j).$$

Further, Robinson (1995a) showed that $\hat{d}$ is a consistent estimator of d_0 , and $\sqrt{m_\diamond}(\hat{d} - d_0) \rightarrow N\left(0, \frac{1}{4}\right)$ as $n \rightarrow \infty$.

3.2.5 Local Whittle Estimator With Tapering

Velasco (1999) showed that it is possible to estimate consistently the Hurst exponent of non-stationary processes using the local Whittle estimator by tapering the observations. Let the tapered periodogram of β be $I_p(\lambda_j)$, and define $Q_p(G, d)$ as the objective function

$$Q_p(G, d) = \frac{p}{m_\diamond} \sum_j^{m_\diamond} \left\{ \ln(G\lambda_j^{-2d}) + \frac{I_p(\lambda_j)}{G\lambda_j^{-2d}} \right\}, \quad (12)$$

where all the summations run for $j = p, 2p, \dots, m_\diamond$, assuming $m_\diamond/p$ is integer. Define

the closed interval of admissible estimate of d_0 , $\Theta = [\nabla_1, \nabla_2]$, ∇_1 and ∇_2 are numbers picked such that $0 < \nabla_1 < \nabla_2 < d^*$ and $p \geq d^* + \frac{1}{2}$ where d^* is the maximum value of d we can estimate with tapers of order p , and d^* may lie in a region where β is non-stationary. When $p = 1$, (12) reduces to (11).

As in Velasco (1999), we define the estimates

$$(\hat{G}_p, \hat{d}_p) = \underset{0 < G < \infty, d \in \Theta}{\operatorname{argmin}} Q_p(G, d).$$

Alternatively, we may obtain

$$\hat{d}_p = \underset{d \in \Theta}{\operatorname{argmin}} R_p(d).$$

where

$$R_p(d) = \ln \hat{G}_p(d) - 2d \frac{p}{m_\circ} \sum_j^{m_\circ} \ln \lambda_j, \quad \hat{G}_p(d) = \frac{p}{m_\circ} \sum_j^{m_\circ} \lambda_j^{2d} I_p(\lambda_j).$$

The tapered periodogram includes only frequencies λ_j for $j = p, 2p, \dots, m_\circ$. The periodogram for non-stationary processes is equivalent to the periodogram for stationary processes evaluated at these frequencies (Velasco 1999).

3.2.6 Modified Local Whittle Estimator

Hou and Perron (2014) proposed a modified local Whittle estimator that has good properties under local contamination. These contaminations include processes whose spectral density functions dominate at low frequencies, such as random level shifts, deterministic level shifts and deterministic trends (Hou and Perron 2014). The data generating process is given

$$z_t = c + \beta_t + u_t,$$

where β_t is a process with memory parameter $d \in [0, \frac{1}{2}]$ and c is a constant. When $d = 0$, β_t is a short-memory process. The process u_t is the low frequency contamination. For a given sample size n , we define the periodogram of process z_t to be $I_z(\lambda_j)$ and $f_z(\lambda_j) = E[I_z(\lambda_j)]$. Since the periodogram of u_t is of order $O_p(\lambda_j^{-2}/n)$, we add a term $(G_u \lambda_j^{-2}/n)$ to the spectral density function of β to govern the low frequency contamination. The modified spectral density function is $f_j = G_0 \lambda_j^{-2d} + G_u \lambda_j^{-2}/n$. Let $\theta = G_u/G_0$ be the noise-to-signal ratio, the modified spectral density function is

$$f_j = G_0(\lambda_j^{-2d} + \theta \lambda_j^{-2}/n) = G_0 \cdot g_j,$$

where

$$g_j = \lambda_j^{-2d} + \theta \lambda_j^{-2}/n.$$

The modified local Whittle estimator is

$$(\hat{d}_\diamond, \hat{\theta}_\diamond) = \underset{d, \theta}{\operatorname{argmin}} J_{m_\diamond}(d, \theta),$$

where

$$J_{m_\diamond} = \ln \left\{ \frac{1}{m_\diamond} \sum_{j=1}^{m_\diamond} \left[\ln(g_j) + \frac{I_z(\lambda_j)}{g_j} \right] \right\}.$$

3.2.7 Exact Local Whittle Estimator

The local Whittle estimator is based on an approximation of $I_\beta(\lambda_j) \sim \lambda_j^{-2d} I_u(\lambda_j)$, where β denotes the original time series and u denotes the noise process. Shimotsu and Phillips (2005) proposed an exact local Whittle estimator that uses a corrected discrete Fourier transform of β to approximate periodogram $I_u(\lambda_j)$. They consider the fractional process β generated by the model

$$\begin{aligned} (1-L)^{d_0} \beta_t &= u_t I\{t \geq 1\}, \quad t = 0, \pm 1, \dots, \\ \beta_t &= (1-L)^{-d_0} u_t I\{t \geq 1\}. \end{aligned}$$

The discrete Fourier transform of a time series β_t evaluated at frequency λ as

$$\begin{aligned} w_\beta(\lambda) &= (2\pi n)^{-\frac{1}{2}} \sum_{t=1}^n \beta_t \exp^{it\lambda} \\ I_\beta(\lambda) &= |w_\beta(\lambda)|^2 \\ I_u(\lambda) &= I_{\Delta^{d_0}\beta}(\lambda) \end{aligned}$$

Define $Q(G, d)$ as the objective function

$$Q_{m_\diamond}(G, d) = \frac{1}{m_\diamond} \sum_{j=1}^{m_\diamond} \left\{ \ln(G\lambda_j^{-2d}) + \frac{I_{\Delta^d\beta}(\lambda_j)}{G\lambda_j^{-2d}} \right\},$$

where $I_{\Delta^d\beta}(\lambda_j)$ is the periodogram of

$$\Delta^d \beta_t = (1-L)^d \beta_t = \sum_{k=0}^t \frac{(-d)_k}{k!} \beta_{t-k}.$$

As in Robinson (1995a), we define the estimates

$$(\hat{G}, \hat{d}) = \underset{0 < G < \infty, d \in \Theta}{\operatorname{argmin}} Q(G, d),$$

where $\Theta = [\Delta_1, \Delta_2]$, Δ_1 and Δ_2 are numbers picked such that $-\infty < \Delta_1 < \Delta_2 < \infty$. Alternatively, we may obtain

$$\hat{d} = \underset{d \in \Theta}{\operatorname{argmin}} R(d)$$

where

$$R(d) = \ln \hat{G}(d) - \frac{2d}{m_\circ} \sum_{j=1}^{m_\circ} \ln \lambda_j, \quad \hat{G}(H) = \frac{1}{m_\circ} \sum_{j=1}^{m_\circ} I_{\Delta^d \beta}(\lambda_j). \quad (13)$$

3.2.8 Two-Step Local Whittle Estimator

The two-step local Whittle estimator of Shimotsu (2010) consists of two steps: First, let $\hat{d}$ denotes the first-stage estimator obtained from (13) and then we define the two-step ELW estimator, coined as ELW2S, as

$$\hat{d}_{\text{ELW2S}} = \hat{d} - \frac{R'(\hat{d})}{R''(\hat{d})}, \quad (14)$$

where $R(d)$ is the objective function defined in (13), ' and '' denote the first- and second-order derivatives, respectively. To improve its finite-sample performance, one can iterate the two-step procedure in (14) and obtain the updated estimator $\hat{d}_{\text{ELW2S}}$ within a certain tolerance level.

Shimotsu (2010) considered the following data generating process

$$\beta_t = \mu_0 + \beta_t^0, \quad \beta_t^0 = (1 - L)^{-d_0} u_t I\{t \geq 1\},$$

where $\mu_0 = E(\beta_t)$ is a fixed unknown quantity. An estimator for μ_0 is the sample average $\bar{\beta} = \frac{1}{n} \sum_{t=1}^n \beta_t$. The resulting memory parameter estimator is given as

$$\hat{d} = \underset{d \in \Theta}{\operatorname{argmin}} R^\circ(d),$$

where

$$R^\circ(d) = \ln \hat{G}^\circ(d) - \frac{2d}{m_\circ} \sum_{j=1}^{m_\circ} \ln \lambda_j, \quad \hat{G}^\circ(d) = \frac{1}{m_\circ} \sum_{j=1}^{m_\circ} I_{\Delta^d(\beta - \hat{\mu})}(\lambda_j),$$

where $I_{\Delta^d(\beta - \hat{\mu})}(\lambda_j)$ is the periodogram of $\Delta^d(\beta - \hat{\mu})$.

4 Numerical Studies

Numerical studies of finite-sample performance are provided via simulation and empirical applications.

4.1 The Functional ARFIMA Model

We study the functional ARFIMA (p, d, q) process defined by

$$\nabla^d \mathcal{X}_t(u) = Y_t(u), \quad \nabla = 1 - B, \quad -1/2 < d < 1/2, \quad (15)$$

and

$$Y_t(u) - \sum_{i=1}^p \int_{\mathcal{I}} \phi_i(u, v) Y_{t-i}(v) dv = \eta_t(u) + \sum_{i=1}^q \int_{\mathcal{I}} \psi_i(u, v) \eta_{t-i}(v) dv, \quad (16)$$

where B denotes the backshift operator, $\{\eta_t\}$ denotes the noise operator, and $\phi_i(u, v)$ and $\psi_i(u, v)$ are the kernels with the associated integral operators defined by $\int_{\mathcal{I}} \phi_i(u, v) x(v) dv$ and $\int_{\mathcal{I}} \psi_i(u, v) x(v) dv$, respectively, $x \in \mathcal{H}$, and such that $Y_t(u)$ is stationary with respect to t . Note that

$$\begin{aligned} \mathcal{X}_t &= \nabla^{-d} Y_t = (1 - B)^{-d} Y_t \\ &= \sum_{i=0}^{\infty} \beta_{i,-d} B^i Y_t \\ &= \sum_{i=0}^{\infty} \beta_{i,-d} Y_{t-i}, \end{aligned} \quad (17)$$

where, by Stirling's formula,

$$\begin{aligned} \beta_{i,-d} &= \beta_i^* + \beta_i^\diamond \\ \beta_i^* &= \frac{1}{\Gamma(d)} i^{-1+d} \\ \beta_i^\diamond &= O(i^{-2+d}), \end{aligned}$$

where $\Gamma(\cdot)$ is the gamma function. For a stationary series Y_t , it can be simulated from the following moving average (MA) (∞) representation of the functional autoregressive moving average (ARMA) (p, q) process:

$$Y_t = \sum_{i=0}^{\infty} \pi[A_i(\bar{\eta}_{t-i})], \quad (18)$$

where $\pi(u_1, \dots, u_p) = u_1$, and A_i denotes the integral operator in Hilbert space (see, e. g., Klepsch, Klüppelberg, and Wei 2017; Li, Robinson, and Shang 2019). By combining (16) with (15), we obtain

$$\mathcal{X}_t = \sum_{i=0}^{\infty} \beta_{i,-d} \sum_{j=0}^{\infty} \pi[A_j(\bar{\eta}_{t-i-j})].$$

In our simulation studies, we implement a truncation, i. e., we use the first n samples of $\bar{\eta}$ where n denotes the sample size as the burn-in, and keep the remaining $n + 100$ samples of $\bar{\eta}$ as our simulated realizations.

When $d = 0$, model (13) and (14) becomes the functional ARMA (p, q) of Klepsch, Klüppelberg, and Wei (2017), while when $q = 0$, it further reduces to the functional autoregressive (AR) (p) model of Bosq (2000) and Liu, Xiao, and Chen (2016), and when $p = 0$, it reduces to the functional MA (q) model of Chen, Lei, and Tu (2016) and Aue and Klepsch (2017).

For notational simplicity, we let $\phi_i = \phi_i(\cdot, \cdot)$ and $\psi_i = \psi_i(\cdot, \cdot)$. As in Bosq (2000) and Klepsch, Klüppelberg, and Wei (2017), for the functional ARMA (p, q) process $\{Y_t\}$, we can write

$$\bar{Y}_t(u) = \int_{\mathcal{I}} \bar{\phi}(u, v) \bar{Y}_{t-1}(v) dv + \sum_{i=0}^q \int_{\mathcal{I}} \bar{\psi}_i(u, v) \bar{\eta}_{t-i}(v) dv,$$

where

$$\begin{aligned} \bar{Y}_t(u) &= [Y_t(u), \dots, Y_{t-p+1}(u)]^\top, & \bar{\eta}_t(u) &= [\eta_t(u), 0, \dots, 0]^\top, \\ \bar{\phi}(\cdot, \cdot) &= \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_p \\ I & O & \dots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & \dots & I & O \end{bmatrix}, & \bar{\psi}(\cdot, \cdot) &= \begin{bmatrix} \psi_1 & O & \dots & O \\ O & O & \dots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & O \end{bmatrix}, \end{aligned}$$

$\psi_0 = I$, I and O denote the identity and zero operators, respectively.

4.2 Simulation Study

We consider generating the curve time series $\mathcal{X}_t$ through a functional ARFIMA (p, d, q) model defined in (13) and (14), where $\mathcal{I} = [0, 1]$, $\{\eta_t, t \in \mathbb{Z}\}$ is a sequence of independent and identically distributed standard Brownian motions over $[0, 1]$, and the following two cases are covered:

Case 1: $p = 1, q = 0, \phi_1(u, v) = 0.34 \times \exp \{-(u^2 + v^2)/2\}$,

Case 2: $p = 1, q = 1, \phi_1(u, v) = 0.34 \times \exp \{-(u^2 + v^2)/2\}, \psi_1(u, v) = \frac{3}{2} \min(u, v)$,

where $d = 0.05, 0.10, \dots, 0.40$ in both cases. The choice of the constants in the definitions of ϕ_1 and ψ_1 ensures that both $\|\phi_1\|$ and $\|\psi_1\|$ are smaller than one (c.f., Rice and Shang 2017), so the simulated curve time series are stationary and invertible. The sample sizes employed are $n = 250, 500, 1,000$ with 1,000 replications.

For a given estimation method, we obtain $B = 1,000$ estimated values of self-similar parameter d , namely $\hat{d}_b$ for $b = 1, \dots, 1,000$. We compute their bias, variance (σ^2) and mean squared error (MSE), given below

$$\begin{aligned}\text{Bias} &= \frac{1}{B} \sum_{b=1}^B (d - \hat{d}_b), \\ \sigma^2 &= \frac{1}{B-1} \left[\sum_{b=1}^B \left(\hat{d}_b - \frac{1}{B} \sum_{b=1}^B \hat{d}_b \right)^2 \right], \\ \text{MSE} &= \frac{1}{B} \sum_{b=1}^B (\hat{d}_b - d)^2.\end{aligned}$$

The bias measures the tendency to over or under-estimate the long-memory parameter. The variance measures how far a set of estimated long-memory parameters are spread out from their mean. The MSE provides some information on the estimation accuracy of the long-memory parameter.

Under the functional ARFIMA $(1, d, 0)$ model, in Table 2, we evaluate and compare the finite-sample bias, variance, and MSE with the long-run covariance estimated from Li, Robinson, and Shang (2019). The R/S estimator produces the smallest variance. The local Whittle and Hou-Perron estimator with tapering produces the smallest bias, while the Peng's estimator produces the smallest MSE.

Under the functional ARFIMA $(1, d, 1)$ model, in Table 3, we evaluate and compare the finite-sample bias, variance, and MSE with the long-run covariance estimated from Li, Robinson, and Shang (2019). The R/S estimator produces the smallest variance. The local Whittle and Hou-Perron estimators produce the smallest bias for various sample sizes, while the local Whittle estimator produces the smallest MSE.

In the supplement, we also consider the same functional models with the long-run covariance estimated from Rice and Shang (2017). Despite two ways of estimating the long-run covariance function, the summary statistics of the estimation bias, variance and MSE differ marginally and our recommendation for the best estimator, in terms of bias, variance, and MSE, remains the same.

The ranking of estimators might depend on the particular short-memory dependence of the time series, so we have also explored some variants of Case 1 and Case 2 for a selected subset of estimates in Table 4.

By altering the coefficient in the kernel function, it changes the temporal dependence structure from weak to strong dependence in the ARFIMA $(1, d, 0)$ case. In the ARFIMA $(1, d, 0)$ model, the coefficient 0.34 produces a L_2 norm of 0.5, implying a moderate temporal dependence. To have the L_2 norm of ϕ_1 to be 0.1, we change that coefficient from 0.34 to 0.068. To have the L_2 norm of ϕ_1 to be 0.9, we change that coefficient from 0.34 to 0.612. From the MSE of the functional ARFIMA $(1, d, 0)$ model in Table 2, the Peng estimator produces the smallest estimation error, followed by the local Whittle estimator. From Table 4, we observe that the

Table 2: With sample size $n = 250, 500, 1,000$, we evaluate and compare the finite-sample bias, variance, and MSE of the estimation error Δd under the functional ARFIMA $(1, d, 0)$ model, where the long-run covariance was estimated from Li, Robinson, and Shang (2019).

n	Statistic	Estimator	d							Overall	
			0.05	0.10	0.15	0.20	0.25	0.30	0.35		0.40
250	Bias	$\hat{\epsilon}_{\text{aggrvar}}$	-0.087	-0.098	-0.109	-0.121	-0.134	-0.148	-0.164	-0.182	-0.130
		$\hat{\epsilon}_{\text{diffvar}}$	0.228	0.224	0.223	0.227	0.228	0.232	0.239	0.226	0.228
		$\hat{\epsilon}_{\text{absval}}$	0.028	0.017	0.006	-0.007	-0.020	-0.034	-0.050	-0.069	-0.016
		$\hat{\epsilon}_{\text{Higuchi}}$	-0.044	-0.053	-0.061	-0.071	-0.082	-0.095	-0.111	-0.132	-0.081
		$\hat{\epsilon}_{\text{Peng}}$	0.049	0.039	0.030	0.021	0.014	0.007	0.001	-0.004	0.020
		$\hat{\epsilon}_{\text{RS}}$	0.046	0.016	-0.014	-0.044	-0.074	-0.105	-0.137	-0.170	-0.060
		$\hat{\epsilon}_{\text{RAR}}$	0.120	0.094	0.066	0.039	0.011	-0.018	-0.047	-0.077	0.023
		$\hat{\epsilon}_{\text{per}}$	0.037	0.049	0.061	0.073	0.085	0.099	0.114	0.128	0.081
		$\hat{\epsilon}_{\text{boxper}}$	0.132	0.132	0.132	0.132	0.132	0.131	0.132	0.134	0.132
		$\hat{\epsilon}_{\text{GPH}}$	0.027	0.030	0.033	0.037	0.041	0.045	0.051	0.055	0.040
		$\hat{\epsilon}_{\text{SGPH}}$	-0.032	-0.033	-0.034	-0.034	-0.031	-0.028	-0.023	-0.018	-0.029
		$\hat{\epsilon}_{\text{Wavelet}}$	0.052	0.048	0.044	0.040	0.036	0.032	0.028	0.025	0.038
		$\hat{\epsilon}_{\text{Local}_W}$	-0.018	-0.018	-0.017	-0.016	-0.014	-0.011	-0.009	-0.010	-0.014
		$\hat{\epsilon}_{\text{Local}_W_T}$	0.022	0.020	0.017	0.013	0.007	-0.002	-0.013	-0.029	0.004
		$\hat{\epsilon}_{\text{Hou_Perron}}$	-0.015	-0.017	-0.019	-0.022	-0.025	-0.027	-0.031	-0.036	-0.024
		Variance	$\hat{\epsilon}_{\text{ELW}}$	0.060	0.061	0.062	0.063	0.064	0.066	0.068	0.070
$\hat{\epsilon}_{\text{ELW2S}}$	0.040		0.040	0.039	0.046	0.049	0.049	0.046	0.045	0.044	
$\hat{\epsilon}_{\text{aggrvar}}$	0.030		0.029	0.029	0.029	0.028	0.027	0.026	0.025	0.028	
$\hat{\epsilon}_{\text{diffvar}}$	0.042		0.041	0.043	0.046	0.049	0.048	0.052	0.055	0.047	
$\hat{\epsilon}_{\text{absval}}$	0.030		0.030	0.029	0.029	0.028	0.027	0.026	0.024	0.028	
$\hat{\epsilon}_{\text{Higuchi}}$	0.015		0.016	0.017	0.018	0.019	0.020	0.020	0.019	0.018	
$\hat{\epsilon}_{\text{Peng}}$	0.005		0.006	0.006	0.007	0.007	0.007	0.008	0.008	0.007	
$\hat{\epsilon}_{\text{RS}}$	0.002		0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	
$\hat{\epsilon}_{\text{RAR}}$	0.014		0.014	0.014	0.015	0.015	0.014	0.014	0.014	0.014	
$\hat{\epsilon}_{\text{per}}$	0.036		0.035	0.034	0.033	0.033	0.032	0.032	0.032	0.033	

Table 2: (continued)

<i>n</i>	Statistic	Estimator	<i>d</i>								Overall
			0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	
MSE		$\hat{\epsilon}_{\text{boxper}}$	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009
		$\hat{\epsilon}_{\text{GPH}}$	0.045	0.044	0.043	0.042	0.042	0.041	0.041	0.041	0.042
		$\hat{\epsilon}_{\text{SGPH}}$	0.024	0.025	0.025	0.025	0.025	0.025	0.026	0.026	0.025
		$\hat{\epsilon}_{\text{Wavelet}}$	0.076	0.076	0.076	0.076	0.076	0.077	0.078	0.079	0.077
		$\hat{\epsilon}_{\text{Local}_W}$	0.009	0.009	0.009	0.009	0.009	0.009	0.008	0.007	0.009
		$\hat{\epsilon}_{\text{Local}_W_T}$	0.045	0.044	0.042	0.039	0.036	0.031	0.027	0.022	0.036
		$\hat{\epsilon}_{\text{Hou-Perron}}$	0.015	0.016	0.016	0.018	0.019	0.021	0.023	0.027	0.019
		$\hat{\epsilon}_{\text{ELW}}$	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
		$\hat{\epsilon}_{\text{ELW2S}}$	0.020	0.020	0.019	0.020	0.019	0.018	0.018	0.020	0.019
		$\hat{\epsilon}_{\text{aggvar}}$	0.037	0.039	0.041	0.043	0.046	0.049	0.053	0.058	0.046
		$\hat{\epsilon}_{\text{diffvar}}$	0.094	0.091	0.093	0.097	0.101	0.102	0.109	0.106	0.099
		$\hat{\epsilon}_{\text{absval}}$	0.031	0.030	0.029	0.029	0.029	0.028	0.028	0.029	0.029
		$\hat{\epsilon}_{\text{Higuchi}}$	0.017	0.019	0.021	0.023	0.026	0.029	0.032	0.037	0.025
		$\hat{\epsilon}_{\text{Peng}}$	0.008	0.007	0.007	0.007	0.007	0.007	0.008	0.008	0.007
		$\hat{\epsilon}_{\text{RS}}$	0.004	0.002	0.002	0.004	0.007	0.013	0.021	0.031	0.011
		$\hat{\epsilon}_{\text{RAR}}$	0.028	0.023	0.019	0.016	0.015	0.015	0.016	0.020	0.019
		$\hat{\epsilon}_{\text{per}}$	0.037	0.037	0.037	0.038	0.040	0.042	0.045	0.049	0.041
		$\hat{\epsilon}_{\text{boxper}}$	0.027	0.027	0.027	0.026	0.027	0.027	0.027	0.027	0.027
		$\hat{\epsilon}_{\text{GPH}}$	0.046	0.045	0.044	0.043	0.044	0.043	0.044	0.044	0.044
		$\hat{\epsilon}_{\text{SGPH}}$	0.025	0.026	0.026	0.026	0.026	0.026	0.026	0.027	0.026
		$\hat{\epsilon}_{\text{Wavelet}}$	0.078	0.078	0.078	0.077	0.078	0.078	0.079	0.079	0.078
		$\hat{\epsilon}_{\text{Local}_W}$	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.007	0.009
		$\hat{\epsilon}_{\text{Local}_W_T}$	0.049	0.048	0.045	0.042	0.037	0.032	0.027	0.022	0.038
		$\hat{\epsilon}_{\text{Hou-Perron}}$	0.015	0.016	0.017	0.018	0.019	0.021	0.024	0.028	0.020
		$\hat{\epsilon}_{\text{ELW}}$	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.012	0.011
		$\hat{\epsilon}_{\text{ELW2S}}$	0.021	0.021	0.021	0.022	0.021	0.021	0.021	0.022	0.021

Table 2: (continued)

n	Statistic	Estimator	d								Overall			
			0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40				
MSE		$\hat{\epsilon}_{\text{GPH}}$	0.028	0.028	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027
		$\hat{\epsilon}_{\text{SGPH}}$	0.016	0.016	0.016	0.016	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017
		$\hat{\epsilon}_{\text{Wavelet}}$	0.027	0.027	0.027	0.027	0.027	0.026	0.026	0.026	0.026	0.026	0.026	0.027
		$\hat{\epsilon}_{\text{Local}_W}$	0.006	0.006	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.005	0.005
		$\hat{\epsilon}_{\text{Local}_W_T}$	0.023	0.023	0.023	0.022	0.021	0.020	0.018	0.015	0.015	0.015	0.020	0.020
		$\hat{\epsilon}_{\text{Hou_Perron}}$	0.008	0.008	0.009	0.009	0.011	0.011	0.013	0.014	0.010	0.010	0.004	0.004
		$\hat{\epsilon}_{\text{ELW}}$	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		$\hat{\epsilon}_{\text{ELW2S}}$	0.006	0.006	0.008	0.009	0.009	0.007	0.006	0.007	0.007	0.007	0.007	0.007
		$\hat{\epsilon}_{\text{aggrvar}}$	0.021	0.023	0.025	0.028	0.031	0.034	0.038	0.042	0.030	0.030	0.068	0.068
		$\hat{\epsilon}_{\text{diffvar}}$	0.069	0.068	0.068	0.067	0.067	0.067	0.066	0.070	0.019	0.018	0.019	0.018
		$\hat{\epsilon}_{\text{absval}}$	0.017	0.017	0.018	0.018	0.019	0.020	0.021	0.023	0.027	0.027	0.018	0.018
		$\hat{\epsilon}_{\text{Higuchi}}$	0.011	0.013	0.015	0.017	0.019	0.021	0.023	0.027	0.009	0.009	0.012	0.012
		$\hat{\epsilon}_{\text{Peng}}$	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.005	0.004	0.004	0.004	0.004
		$\hat{\epsilon}_{\text{RS}}$	0.003	0.002	0.002	0.003	0.007	0.012	0.018	0.027	0.009	0.009	0.012	0.012
		$\hat{\epsilon}_{\text{RAR}}$	0.018	0.014	0.012	0.010	0.009	0.010	0.011	0.015	0.012	0.012	0.016	0.016
		$\hat{\epsilon}_{\text{per}}$	0.015	0.015	0.015	0.016	0.016	0.017	0.018	0.020	0.016	0.016	0.020	0.020
		$\hat{\epsilon}_{\text{boxper}}$	0.009	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
		$\hat{\epsilon}_{\text{GPH}}$	0.028	0.028	0.027	0.027	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
		$\hat{\epsilon}_{\text{SGPH}}$	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017
		$\hat{\epsilon}_{\text{Wavelet}}$	0.028	0.027	0.027	0.027	0.027	0.026	0.026	0.026	0.027	0.027	0.027	0.027
		$\hat{\epsilon}_{\text{Local}_W}$	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.005	0.006	0.006	0.006	0.006
		$\hat{\epsilon}_{\text{Local}_W_T}$	0.023	0.023	0.023	0.022	0.021	0.020	0.018	0.015	0.020	0.018	0.015	0.020
		$\hat{\epsilon}_{\text{Hou_Perron}}$	0.009	0.009	0.009	0.010	0.011	0.012	0.014	0.015	0.011	0.011	0.015	0.011
		$\hat{\epsilon}_{\text{ELW}}$	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
		$\hat{\epsilon}_{\text{ELW2S}}$	0.007	0.007	0.008	0.010	0.010	0.008	0.007	0.007	0.008	0.007	0.007	0.008

Table 2: (continued)

<i>n</i>	Statistic	Estimator	<i>d</i>								Overall
			0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	
MSE		$\hat{\epsilon}_{\text{GPH}}$	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.020	0.019
		$\hat{\epsilon}_{\text{SGPH}}$	0.011	0.011	0.011	0.011	0.012	0.012	0.012	0.012	0.012
		$\hat{\epsilon}_{\text{Wavelet}}$	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012
		$\hat{\epsilon}_{\text{Local}_W}$	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
		$\hat{\epsilon}_{\text{Local}_W_T}$	0.014	0.014	0.014	0.014	0.013	0.013	0.012	0.010	0.013
		$\hat{\epsilon}_{\text{Hou-Perron}}$	0.004	0.004	0.005	0.005	0.005	0.005	0.006	0.006	0.005
		$\hat{\epsilon}_{\text{ELW}}$	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
		$\hat{\epsilon}_{\text{ELW2S}}$	0.003	0.003	0.004	0.006	0.006	0.005	0.004	0.004	0.004
		$\hat{\epsilon}_{\text{aggvar}}$	0.014	0.015	0.016	0.018	0.020	0.022	0.026	0.030	0.020
		$\hat{\epsilon}_{\text{diffvar}}$	0.048	0.046	0.047	0.047	0.047	0.047	0.048	0.049	0.047
		$\hat{\epsilon}_{\text{absval}}$	0.012	0.012	0.012	0.012	0.013	0.014	0.015	0.017	0.013
		$\hat{\epsilon}_{\text{Higuchi}}$	0.005	0.006	0.007	0.008	0.009	0.010	0.011	0.014	0.009
		$\hat{\epsilon}_{\text{peng}}$	0.003	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003
		$\hat{\epsilon}_{\text{RS}}$	0.003	0.001	0.001	0.003	0.006	0.010	0.016	0.024	0.008
		$\hat{\epsilon}_{\text{RAR}}$	0.015	0.012	0.009	0.008	0.007	0.008	0.009	0.012	0.010
		$\hat{\epsilon}_{\text{per}}$	0.006	0.006	0.007	0.007	0.007	0.008	0.008	0.009	0.007
		$\hat{\epsilon}_{\text{boxper}}$	0.003	0.003	0.003	0.003	0.003	0.003	0.004	0.004	0.003
		$\hat{\epsilon}_{\text{GPH}}$	0.019	0.019	0.019	0.019	0.019	0.019	0.020	0.020	0.019
		$\hat{\epsilon}_{\text{SGPH}}$	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012
		$\hat{\epsilon}_{\text{Wavelet}}$	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012
		$\hat{\epsilon}_{\text{Local}_W}$	0.004	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
		$\hat{\epsilon}_{\text{Local}_W_T}$	0.014	0.014	0.014	0.014	0.013	0.013	0.012	0.010	0.013
		$\hat{\epsilon}_{\text{Hou-Perron}}$	0.004	0.005	0.005	0.005	0.005	0.005	0.006	0.007	0.005
		$\hat{\epsilon}_{\text{ELW}}$	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
		$\hat{\epsilon}_{\text{ELW2S}}$	0.003	0.003	0.004	0.007	0.007	0.006	0.004	0.004	0.005

The smallest error is highlighted in bold for each memory parameter *d*.

Table 3: (continued)

n	Statistic	Estimator	d								Overall
			0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	
MSE		$\hat{\epsilon}_{per}$	0.037	0.036	0.036	0.036	0.035	0.035	0.035	0.035	0.036
		$\hat{\epsilon}_{boxper}$	0.009	0.009	0.009	0.009	0.009	0.009	0.010	0.009	0.009
		$\hat{\epsilon}_{GPH}$	0.047	0.047	0.046	0.046	0.046	0.045	0.045	0.046	0.046
		$\hat{\epsilon}_{SGPH}$	0.025	0.026	0.026	0.027	0.027	0.028	0.028	0.029	0.027
		$\hat{\epsilon}_{Wavelet}$	0.077	0.076	0.075	0.075	0.075	0.076	0.076	0.076	0.076
		$\hat{\epsilon}_{Local_W}$	0.009	0.009	0.009	0.009	0.009	0.009	0.008	0.007	0.009
		$\hat{\epsilon}_{Local_W_T}$	0.045	0.044	0.042	0.039	0.036	0.031	0.027	0.022	0.036
		$\hat{\epsilon}_{Hou_Perron}$	0.014	0.014	0.015	0.016	0.017	0.018	0.020	0.022	0.017
		$\hat{\epsilon}_{ELW}$	0.008	0.008	0.008	0.008	0.008	0.008	0.007	0.007	0.008
		$\hat{\epsilon}_{ELW2S}$	0.019	0.019	0.020	0.019	0.018	0.018	0.019	0.020	0.019
		$\hat{\epsilon}_{aggrvar}$	0.040	0.042	0.045	0.049	0.053	0.057	0.062	0.067	0.052
		$\hat{\epsilon}_{diffvar}$	0.101	0.100	0.101	0.104	0.102	0.103	0.101	0.101	0.102
		$\hat{\epsilon}_{absval}$	0.032	0.032	0.032	0.032	0.032	0.033	0.034	0.035	0.033
		$\hat{\epsilon}_{Higuchi}$	0.017	0.019	0.021	0.023	0.026	0.029	0.033	0.038	0.026
		$\hat{\epsilon}_{Peng}$	0.017	0.015	0.013	0.012	0.011	0.010	0.010	0.010	0.012
		$\hat{\epsilon}_{RS}$	0.007	0.003	0.002	0.003	0.006	0.011	0.019	0.029	0.010
		$\hat{\epsilon}_{RAR}$	0.034	0.026	0.020	0.016	0.013	0.012	0.013	0.016	0.019
		$\hat{\epsilon}_{per}$	0.039	0.040	0.041	0.042	0.044	0.046	0.049	0.053	0.044
		$\hat{\epsilon}_{boxper}$	0.173	0.172	0.171	0.170	0.169	0.168	0.166	0.164	0.169
		$\hat{\epsilon}_{GPH}$	0.048	0.047	0.047	0.046	0.046	0.046	0.047	0.048	0.047
		$\hat{\epsilon}_{SGPH}$	0.026	0.027	0.028	0.028	0.029	0.029	0.029	0.030	0.028
		$\hat{\epsilon}_{Wavelet}$	0.104	0.102	0.099	0.097	0.096	0.095	0.093	0.092	0.097
		$\hat{\epsilon}_{Local_W}$	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.007	0.009
		$\hat{\epsilon}_{Local_W_T}$	0.049	0.048	0.045	0.042	0.037	0.032	0.027	0.022	0.038
		$\hat{\epsilon}_{Hou_Perron}$	0.014	0.015	0.015	0.016	0.017	0.018	0.020	0.022	0.017
		$\hat{\epsilon}_{ELW}$	0.020	0.020	0.020	0.020	0.020	0.020	0.021	0.021	0.020

Table 3: (continued)

n	Statistic	Estimator	d								Overall
			0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	
500	Bias	$\hat{\epsilon}_{ELW2S}$	0.028	0.028	0.029	0.029	0.028	0.028	0.029	0.031	0.029
		$\hat{\epsilon}_{aggr}$	-0.068	-0.080	-0.092	-0.106	-0.120	-0.136	-0.153	-0.170	-0.115
		$\hat{\epsilon}_{diff}$	0.228	0.222	0.222	0.223	0.214	0.214	0.213	0.215	0.219
		$\hat{\epsilon}_{absval}$	0.016	0.004	-0.009	-0.022	-0.037	-0.052	-0.070	-0.087	-0.032
		$\hat{\epsilon}_{Higuchi}$	-0.028	-0.037	-0.046	-0.055	-0.066	-0.079	-0.095	-0.114	-0.065
		$\hat{\epsilon}_{Peng}$	0.083	0.073	0.063	0.055	0.047	0.039	0.033	0.027	0.052
		$\hat{\epsilon}_{RS}$	0.064	0.034	0.004	-0.026	-0.057	-0.089	-0.121	-0.154	-0.043
	Variance	$\hat{\epsilon}_{RAR}$	0.122	0.096	0.069	0.042	0.014	-0.014	-0.044	-0.074	0.026
		$\hat{\epsilon}_{per}$	0.033	0.039	0.046	0.054	0.061	0.069	0.078	0.089	0.059
		$\hat{\epsilon}_{boxper}$	0.286	0.285	0.285	0.285	0.285	0.285	0.285	0.285	0.285
		$\hat{\epsilon}_{GPH}$	0.010	0.011	0.013	0.014	0.017	0.020	0.024	0.029	0.017
		$\hat{\epsilon}_{SGPH}$	-0.029	-0.031	-0.032	-0.032	-0.031	-0.029	-0.026	-0.021	-0.029
		$\hat{\epsilon}_{Wavelet}$	0.108	0.102	0.097	0.092	0.088	0.084	0.079	0.075	0.091
		$\hat{\epsilon}_{Local_W}$	-0.009	-0.009	-0.009	-0.008	-0.007	-0.006	-0.005	-0.005	-0.007
500	Bias	$\hat{\epsilon}_{Local_W_T}$	0.022	0.022	0.021	0.020	0.018	0.014	0.007	-0.005	0.015
		$\hat{\epsilon}_{Hou_Perron}$	-0.000	-0.002	-0.002	-0.003	-0.005	-0.007	-0.010	-0.011	-0.005
		$\hat{\epsilon}_{ELW}$	0.062	0.062	0.062	0.062	0.063	0.064	0.065	0.066	0.063
		$\hat{\epsilon}_{ELW2S}$	0.052	0.050	0.053	0.059	0.065	0.063	0.056	0.057	0.057
		$\hat{\epsilon}_{aggr}$	0.017	0.018	0.018	0.018	0.019	0.018	0.018	0.017	0.018
		$\hat{\epsilon}_{diff}$	0.022	0.022	0.022	0.022	0.022	0.023	0.024	0.024	0.023
		$\hat{\epsilon}_{absval}$	0.017	0.018	0.018	0.019	0.019	0.019	0.019	0.017	0.018
	Variance	$\hat{\epsilon}_{Higuchi}$	0.010	0.011	0.012	0.013	0.014	0.014	0.014	0.014	0.013
		$\hat{\epsilon}_{Peng}$	0.003	0.003	0.003	0.004	0.004	0.004	0.004	0.005	0.004
		$\hat{\epsilon}_{RS}$	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
		$\hat{\epsilon}_{RAR}$	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		$\hat{\epsilon}_{per}$	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
		$\hat{\epsilon}_{boxper}$	0.286	0.285	0.285	0.285	0.285	0.285	0.285	0.285	0.285
		$\hat{\epsilon}_{SGPH}$	0.010	0.011	0.013	0.014	0.017	0.020	0.024	0.029	0.017

Table 3: (continued)

n	Statistic	Estimator	d								Overall	
			0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40		
MSE		$\hat{\epsilon}_{\text{boxper}}$	0.003	0.003	0.003	0.003	0.003	0.003	0.004	0.004	0.003	0.003
		$\hat{\epsilon}_{\text{GPH}}$	0.029	0.029	0.028	0.028	0.028	0.028	0.029	0.028	0.028	0.028
		$\hat{\epsilon}_{\text{SGPH}}$	0.016	0.017	0.017	0.017	0.017	0.018	0.018	0.018	0.017	0.017
		$\hat{\epsilon}_{\text{Wavelet}}$	0.027	0.027	0.027	0.027	0.027	0.026	0.026	0.026	0.027	0.027
		$\hat{\epsilon}_{\text{Local}_W}$	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.005	0.006	0.006
		$\hat{\epsilon}_{\text{Local}_W_T}$	0.023	0.023	0.022	0.022	0.020	0.019	0.017	0.013	0.020	0.020
		$\hat{\epsilon}_{\text{Hou_Perron}}$	0.008	0.008	0.008	0.008	0.009	0.011	0.012	0.013	0.010	0.010
		$\hat{\epsilon}_{\text{ELW}}$	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		$\hat{\epsilon}_{\text{ELW2S}}$	0.006	0.007	0.009	0.009	0.008	0.006	0.006	0.007	0.007	0.007
		$\hat{\epsilon}_{\text{aggrvar}}$	0.021	0.024	0.027	0.030	0.033	0.037	0.041	0.046	0.032	0.032
		$\hat{\epsilon}_{\text{diffvar}}$	0.074	0.071	0.072	0.072	0.068	0.069	0.069	0.070	0.071	0.071
		$\hat{\epsilon}_{\text{absval}}$	0.017	0.018	0.019	0.019	0.020	0.021	0.023	0.025	0.020	0.020
		$\hat{\epsilon}_{\text{Higuchi}}$	0.011	0.013	0.014	0.016	0.018	0.021	0.023	0.027	0.018	0.018
		$\hat{\epsilon}_{\text{Peng}}$	0.010	0.008	0.007	0.007	0.006	0.006	0.005	0.005	0.007	0.007
		$\hat{\epsilon}_{\text{RS}}$	0.006	0.003	0.002	0.002	0.005	0.010	0.016	0.026	0.009	0.009
		$\hat{\epsilon}_{\text{RAR}}$	0.023	0.017	0.013	0.010	0.008	0.008	0.010	0.013	0.013	0.013
		$\hat{\epsilon}_{\text{per}}$	0.015	0.016	0.016	0.017	0.018	0.019	0.020	0.022	0.018	0.018
		$\hat{\epsilon}_{\text{boxper}}$	0.085	0.085	0.085	0.085	0.085	0.085	0.085	0.085	0.085	0.085
		$\hat{\epsilon}_{\text{GPH}}$	0.029	0.029	0.028	0.029	0.029	0.029	0.029	0.029	0.029	0.029
		$\hat{\epsilon}_{\text{SGPH}}$	0.017	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018
		$\hat{\epsilon}_{\text{Wavelet}}$	0.038	0.037	0.036	0.035	0.034	0.033	0.032	0.032	0.035	0.035
		$\hat{\epsilon}_{\text{Local}_W}$	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.005	0.006	0.006
		$\hat{\epsilon}_{\text{Local}_W_T}$	0.023	0.023	0.023	0.022	0.021	0.019	0.017	0.013	0.020	0.020
		$\hat{\epsilon}_{\text{Hou_Perron}}$	0.008	0.008	0.008	0.008	0.009	0.011	0.012	0.013	0.010	0.010
		$\hat{\epsilon}_{\text{ELW}}$	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		$\hat{\epsilon}_{\text{ELW2S}}$	0.009	0.010	0.011	0.012	0.012	0.010	0.009	0.010	0.010	0.010

Table 3: (continued)

n	Statistic	Estimator	d								Overall	
			0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40		
MSE		$\hat{\epsilon}_{\text{GPH}}$	0.019	0.019	0.020	0.019	0.019	0.020	0.020	0.020	0.020	0.020
		$\hat{\epsilon}_{\text{SGPH}}$	0.011	0.011	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012
		$\hat{\epsilon}_{\text{Wavelet}}$	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012
		$\hat{\epsilon}_{\text{Local}_W}$	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
		$\hat{\epsilon}_{\text{Local}_W_T}$	0.014	0.014	0.014	0.014	0.013	0.013	0.011	0.009	0.013	0.013
		$\hat{\epsilon}_{\text{Hou_Perron}}$	0.004	0.004	0.004	0.005	0.005	0.005	0.005	0.006	0.005	0.005
		$\hat{\epsilon}_{\text{ELW}}$	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
		$\hat{\epsilon}_{\text{ELW2S}}$	0.003	0.004	0.005	0.006	0.005	0.004	0.003	0.004	0.004	0.004
		$\hat{\epsilon}_{\text{aggar}}$	0.013	0.015	0.016	0.018	0.020	0.023	0.027	0.031	0.020	0.020
		$\hat{\epsilon}_{\text{diffvar}}$	0.052	0.050	0.051	0.051	0.050	0.050	0.050	0.051	0.051	0.051
		$\hat{\epsilon}_{\text{absval}}$	0.012	0.012	0.012	0.012	0.013	0.014	0.016	0.018	0.014	0.014
		$\hat{\epsilon}_{\text{Higuchi}}$	0.005	0.006	0.006	0.007	0.008	0.010	0.011	0.014	0.008	0.008
		$\hat{\epsilon}_{\text{Peng}}$	0.006	0.005	0.005	0.004	0.004	0.003	0.003	0.003	0.004	0.004
		$\hat{\epsilon}_{\text{RS}}$	0.005	0.002	0.001	0.002	0.004	0.008	0.014	0.022	0.007	0.007
		$\hat{\epsilon}_{\text{RAR}}$	0.019	0.014	0.011	0.008	0.007	0.007	0.008	0.011	0.010	0.010
		$\hat{\epsilon}_{\text{per}}$	0.007	0.007	0.007	0.008	0.008	0.009	0.009	0.010	0.008	0.008
		$\hat{\epsilon}_{\text{boxper}}$	0.038	0.038	0.038	0.039	0.039	0.039	0.039	0.040	0.039	0.039
		$\hat{\epsilon}_{\text{GPH}}$	0.019	0.019	0.020	0.019	0.019	0.020	0.020	0.020	0.020	0.020
		$\hat{\epsilon}_{\text{SGPH}}$	0.012	0.012	0.012	0.012	0.012	0.012	0.013	0.013	0.012	0.012
		$\hat{\epsilon}_{\text{Wavelet}}$	0.017	0.017	0.017	0.016	0.016	0.015	0.015	0.015	0.016	0.016
		$\hat{\epsilon}_{\text{Local}_W}$	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
		$\hat{\epsilon}_{\text{Local}_W_T}$	0.014	0.014	0.014	0.014	0.013	0.013	0.011	0.009	0.013	0.013
		$\hat{\epsilon}_{\text{Hou_Perron}}$	0.004	0.004	0.004	0.005	0.005	0.005	0.005	0.006	0.005	0.005
		$\hat{\epsilon}_{\text{ELW}}$	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		$\hat{\epsilon}_{\text{ELW2S}}$	0.004	0.004	0.006	0.008	0.008	0.006	0.005	0.005	0.006	0.006

The smallest error is highlighted in bold for each memory parameter *d*.

Table 4: With sample size $n = 250, 500, 1,000$, we evaluate and compare the finite-sample MSE ($\times 100$) of the estimation error Δd under the functional ARFIMA $(1, d, 0)$ and ARFIMA $(1, d, 1)$ models, where the long-run covariance was estimated from Li, Robinson, and Shang (2019).

ρ	Estimator	n	d								Overall
			0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	
DGP = ARFIMA (1, d , 0)											
0.9	Peng	250	0.96	0.87	0.81	0.78	0.76	0.77	0.78	0.81	0.82
		500	0.53	0.47	0.43	0.41	0.41	0.42	0.43	0.45	0.44
		1,000	0.33	0.29	0.27	0.26	0.26	0.26	0.27	0.29	0.28
	Local_W	250	0.95	0.94	0.94	0.93	0.92	0.90	0.86	0.76	0.90
		500	0.61	0.61	0.61	0.61	0.60	0.60	0.58	0.54	0.60
		1,000	0.35	0.35	0.35	0.35	0.35	0.34	0.34	0.33	0.35
0.1	Peng	250	0.50	0.56	0.63	0.70	0.76	0.83	0.89	0.94	0.73
		500	0.27	0.31	0.35	0.39	0.44	0.48	0.52	0.55	0.41
		1,000	0.17	0.19	0.22	0.25	0.28	0.30	0.33	0.35	0.26
	Local_W	250	0.92	0.91	0.89	0.86	0.83	0.80	0.76	0.68	0.83
		500	0.63	0.62	0.62	0.61	0.60	0.59	0.56	0.51	0.59
		1,000	0.36	0.36	0.35	0.35	0.35	0.34	0.34	0.32	0.35
DGP = ARFIMA (1, d , 1)											
0.9	Local_W	250	0.96	0.96	0.96	0.96	0.96	0.95	0.89	0.73	0.92
		500	0.59	0.58	0.58	0.58	0.59	0.58	0.57	0.51	0.57
		1,000	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.32	0.34
	Peng	250	1.80	1.58	1.40	1.26	1.15	1.08	1.02	0.99	1.29
		500	1.05	0.90	0.79	0.70	0.64	0.59	0.56	0.55	0.72
		1,000	0.65	0.55	0.48	0.42	0.38	0.36	0.34	0.33	0.44

Table 4: (continued)

ρ	Estimator	n	d							Overall	
			0.05	0.10	0.15	0.20	0.25	0.30	0.35		0.40
0.1	Local_W	250	0.92	0.90	0.89	0.86	0.85	0.80	0.76	0.68	0.83
		500	0.62	0.62	0.61	0.61	0.60	0.59	0.56	0.51	0.59
	Peng	1,000	0.36	0.35	0.35	0.35	0.34	0.34	0.34	0.32	0.34
		250	0.50	0.55	0.61	0.67	0.73	0.80	0.85	0.90	0.70
		500	0.27	0.30	0.33	0.38	0.42	0.46	0.50	0.53	0.40
		1,000	0.17	0.19	0.21	0.24	0.26	0.29	0.31	0.33	0.25

Peng estimator still performs better than the local Whittle estimator under various degrees of dependence.

Similarly, by altering the coefficients in the kernel functions, it alters the temporal dependence structure from weak to strong dependence in the ARFIMA $(1, d, 1)$ case. In the functional ARFIMA $(1, d, 1)$ model, the coefficients $(0.34, 1.5)$ produces a L_2 norm of 0.5 for both the AR and MA components implying a moderate temporal dependence. To have the L_2 norm of ϕ_1 and ψ_1 to be 0.1, we change the coefficients from $(0.34, 1.5)$ to $(0.068, 0.059)$. To have the L_2 norm of ϕ_1 and ψ_1 to be 0.9, we change the coefficients from $(0.34, 1.5)$ to $(0.612, 4.765)$. From the MSE of the functional ARFIMA $(1, d, 0)$ model, in Table 3, the local Whittle estimator produces the smallest estimation error, followed by the Peng estimator. From Table 4, we observe that the local Whittle estimator still performs better than the Peng estimator under the strong dependence, but not so under the weak dependence. From this example, various degrees of short-memory temporal dependence can affect the estimation accuracy of the long-memory parameter.

In the supplement, we also consider the same functional models with the long-run covariance estimated from Rice and Shang (2017). Despite two ways of estimating the long-run covariance function, the summary statistics of the MSE differ marginally, and our recommendation for the best estimator, in terms of MSE, remains the same.

5 Conclusion

Through a series of simulation studies, we identify the most accurate estimator for estimating the long-memory parameter in a functional ARFIMA model. For a functional ARFIMA $(1, d, 0)$ with various d values, the local Whittle estimator with tapering produces the smallest bias; the rescaled range estimator produces the smallest variance, and the Peng estimator produces the smallest mean square error. For a functional ARFIMA $(1, d, 1)$ with various d values, the local Whittle estimator produces the smallest bias with $n = 250$ and the Hou-Perron estimator produces the smallest bias with $n = 500$ and $n = 1,000$; the rescaled range estimator produces the smallest variance; and the local Whittle estimator produces the smallest mean square error.

For the functional ARFIMA $(1, d, 0)$ model, the Peng estimator consistently produces the smallest estimation error under weak, moderate, and strong dependence. For the functional ARFIMA $(1, d, 1)$ model, the local Whittle estimator produces the smallest estimation errors for moderate and strong dependence, while the Peng estimator produces the smallest estimation error for weak dependence. In summary, various degrees of short-memory temporal dependence can affect the estimation accuracy of the long-memory parameter.

There are several ways in which the present study can be further extended, and we briefly outline three:

- 1) Estimation of a time-varying long-memory parameter.
- 2) Consider a wavelet-based or Fourier-based multivariate Whittle estimation for a multivariate time series of principal component scores. The multivariate estimation method should be more efficient than a univariate estimation method, subject to each set of principal component scores shows a similar degree of persistence. Should this condition fails to satisfy, it may lead to biased estimates.
- 3) Propose tests for detecting the presence of long-memory.

References

- Andrews, D. 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59 (3): 817–58, [https://doi.org/10.1016/0304-4076\(95\)01788-7](https://doi.org/10.1016/0304-4076(95)01788-7).
- Andrews, D. W. K., and J. C. Monahan. 1992. "An improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator." *Econometrica* 60 (4): 953–66, <https://doi.org/10.2307/2951574>.
- Aue, A., and J. Klepsch. 2017. *Estimating Functional Time Series by Moving Average Model Fitting*, Technical Report. Davies: University of California. <https://arxiv.org/abs/1701.00770>.
- Aue, A., D. D. Norinho, and S. Hörmann. 2015. "On the Prediction of Stationary Functional Time Series." *Journal of the American Statistical Association* 110 (509): 378–92, <https://doi.org/10.1080/01621459.2014.909317>.
- Beran, J. 1994. *Statistics for Long Memory Processes*. New York: Chapman & Hall/CRC Press.
- Beran, J., Y. Feng, S. Ghosh, and R. Kulik. 2013. *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Berlin, Heidelberg: Springer.
- Berkes, I., L. Horváth, and G. Rice. 2016. "On the Asymptotic Normality of Kernel Estimators of the Long Run Covariance of Functional Time Series." *Journal of Multivariate Analysis* 144: 150–75, <https://doi.org/10.1016/j.jmva.2015.11.005>.
- Bosq, D. 2000. *Linear Processes in Function Spaces*. New York: Lecture notes in Statistics.
- Bosq, D., and D. Blanke. 2007. *Inference and Prediction in Large Dimensions*. Chichester: John Wiley & Sons.
- Chen, S. X., L. Lei, and Y. Tu. 2016. "Functional Coefficient Moving Average Model with Applications to Forecasting Chinese CPI." *Statistica Sinica* 26 (4): 1649–72, <https://doi.org/10.5705/ss.2014.196t>.
- Chiou, J.-M., and H.-G. Müller. 2009. "Modeling Hazard Rates as Functional Data for the Analysis of Cohort Lifetables and Mortality Forecasting." *Journal of the American Statistical Association* 104 (486): 572–85, <https://doi.org/10.1198/jasa.2009.0023>.
- Dahlhaus, R. 1989. "Efficient Parameter Estimation for Self-Similar Processes." *The Annals of Statistics* 17 (4): 1749–66, <https://doi.org/10.1214/aos/1176347393>.
- Doukhan, P., G. Oppenheim, and M. Taqqu, eds (2003). *Theory and Applications of Long-Range Dependence*. Boston, MA: Birkhäuser.

- Embrechts, P., and M. Maejima. 2002. *Selfsimilar Processes*. Princeton: Princeton University Press.
- Fox, R., and M. S. Taqqu. 1986. "Large-sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series." *The Annals of Statistics* 14 (2): 517–32.
- Geweke, J., and S. Porter-Hudak. 1983. "The Estimation and Application of Long Memory Time Series Models." *Journal of Time Series Analysis* 4 (4): 221–37, <https://doi.org/10.1111/j.1467-9892.1983.tb00371.x>.
- Higuchi, T. 1988. "Approach to an Irregular Time Series on the Basis of the Fractal Theory." *Physica D* 31 (2): 277–83, [https://doi.org/10.1016/0167-2789\(88\)90081-4](https://doi.org/10.1016/0167-2789(88)90081-4).
- Hörmann, S., L. Kidziński, and M. Hallin. 2015. "Dynamic Functional Principal Components." *Journal of the Royal Statistical Society: Series B* 77 (2): 319–48.
- Horváth, L., P. Kokoszka, and R. Reeder. 2013. "Estimation of the Mean of Functional Time Series and A Two Sample Problem." *Journal of the Royal Statistical Society: Series B* 75 (1): 103–22, <https://doi.org/10.1111/j.1467-9868.2012.01032.x>.
- Hou, J., and P. Perron. 2014. "Modified Local Whittle Estimator for Long Memory Processes in the Presence of Low Frequency (and other) Contaminations." *Journal of Econometrics* 182: 309–28, <https://doi.org/10.1016/j.jeconom.2014.05.004>.
- Hurst, H. E. 1951. "Long-term Storage Capacity of Reservoirs." *Transactions of the American Society of Civil Engineers* 116 (1): 770–99.
- Hyndman, R. J., and H. L. Shang. 2009. "Forecasting Functional Time Series (with Discussion)." *Journal of the Korean Statistical Society* 38 (3): 199–221, <https://doi.org/10.1016/j.jkss.2009.06.002>.
- Hyndman, R. J., and M. Ullah. 2007. "Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach." *Computational Statistics and Data Analysis* 51 (10): 4942–56, <https://doi.org/10.1016/j.csda.2006.07.028>.
- Jensen, M. J. 1999. "Using Wavelets to Obtain a Consistent Ordinary Least Squares Estimator of the Long-Memory Parameter." *Journal of Forecasting* 18 (1): 17–32.
- Jeong, H.-D. J., J.-S. R. Lee, D. McNickle, and K. Pawlikowski. 2007. "Comparison of Various Estimators in Simulated FGN." *Simulation and Modelling Practice and Theory* 15 (9): 1173–91, <https://doi.org/10.1016/j.simpat.2007.08.004>.
- Klepsch, J., and C. Klüppelberg. 2017. "An Innovations Algorithm for the Prediction of Functional Linear Processes." *Journal of Multivariate Analysis* 155: 252–71, <https://doi.org/10.1016/j.jmva.2017.01.005>.
- Klepsch, J., C. Klüppelberg, and T. Wei. 2017. "Prediction of Functional ARMA Processes with an Application to Traffic Data." *Econometrics and Statistics* 1: 128–49, <https://doi.org/10.1016/j.ecosta.2016.10.009>.
- Kokoszka, P., and M. Reimherr. 2013. "Determining the Order of the Functional Autoregressive Model." *Journal of Time Series Analysis* 34 (1): 116–29, <https://doi.org/10.1111/j.1467-9892.2012.00816.x>.
- Kokoszka, P., and M. Reimherr. 2017. *Introduction to Functional Data Analysis*. Boca Raton: Chapman & Hall/CRC Press.
- Künsch, H. R. 1987. "Statistical Aspects of Self-Similar Processes." In *Proceedings of the World Congress of the Bernoulli Society*, Vol. 1, Tashkent, 67–74.
- Li, D., P. M. Robinson, and H. L. Shang. 2019. "Long-range Dependent Curve Time Series." *Journal of the American Statistical Association: Theory and Methods*. <https://doi.org/10.1080/01621459.2019.1604362>.

- Liu, X., H. Xiao, and R. Chen. 2016. "Convolutional Autoregressive Models for Functional Time Series." *Journal of Econometrics* 194 (2): 263–82, <https://doi.org/10.1016/j.jeconom.2016.05.006>.
- Mandelbrot, B. B. 1963. "The Variation of Certain Speculative Prices." *The Journal of Business* 36 (4): 394–419.
- Mandelbrot, B. B. 1975. "Limit Theorems on the Self-Normalized Range for Weakly and Strongly Dependent Processes." *Probability Theory and Related Fields* 31 (4): 271–85, <https://doi.org/10.1007/BF00534968>.
- Mandelbrot, B. B., and M. S. Taqqu. 1979. "Robust R/S Analysis of Long Run Serial Correlation." In *Proceedings of the 42nd Session of the International Statistical Institute, Manila*, 69–104.
- Mandelbrot, B. B., and J. W. van Ness. 1968. "Fractional Brownian Motions, Fractional Noises and applications." *SIAM Review* 10 (4): 422–37.
- Mandelbrot, B. B., and J. R. Wallis. 1969. "Robustness of the Rescaled Range R/S in the Measurement of Noncyclic Long Run Statistical Dependence." *Water Resources Research* 5 (5): 967–88, <https://doi.org/10.1029/WR005i005p00967>.
- Palma, W. 2007. *Long-Memory Time Series: Theory and Methods*. Hoboken, New Jersey: Wiley.
- Panaretos, V. M., and S. Tavakoli. 2012. "Fourier Analysis of Stationary Time Series in Function Space." *Annals of Statistics* 41 (2): 568–603, <https://doi.org/10.1214/13-AOS1086>.
- Peng, C. K., S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger. 1994. "Mosaic Organization of DNA Nucleotides." *Physical Review E* 49 (2): 1685–9, <https://doi.org/10.1103/physreve.49.1685>.
- Reisen, V. A. 1994. "Estimation of the Fractional Difference Parameter in the ARIMA (p, d, q) Model Using the Smoothed Periodogram." *Journal of Time Series Analysis* 15 (3): 335–0, <https://doi.org/10.1111/j.1467-9892.1994.tb00198.x>.
- Rice, G., and H. L. Shang. 2017. "A Plug-in Bandwidth Selection Procedure for Long-Run Covariance Estimation with Stationary Functional Time Series." *Journal of Time Series Analysis* 38 (4): 591–609, <https://doi.org/10.1111/jtsa.12229>.
- Robinson, P. M. 1995a. "Gaussian Semiparametric Estimation of Long Range Dependence." *The Annals of Statistics* 23 (5): 1630–61.
- Robinson, P. M. 1995b. "Log-periodogram Regression of Time Series with Long Range Dependence." *The Annals of Statistics* 23 (3): 1048–72.
- Robinson, P. M., ed. (2003). *Time Series with Long Memory*. Oxford: Oxford University Press.
- Shimotsu, K. 2010. "Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend." *Econometric Theory* 26: 501–40.
- Shimotsu, K., and P. C. B. Phillips. 2005. "Exact Local Whittle Estimation of Fractional Integration." *The Annals of Statistics* 33 (4): 1890–933.
- Taqqu, M., V. Teverovsky, and W. Willinger. 1995. "Estimators for Long-Range Dependence: An Empirical Study." *Fractals* 3 (4): 785–98.
- Teverovsky, V., and M. Taqqu. 1997. "Testing for Long-Range Dependence in the Presence of Shifting Means or A Slowly Declining Trend, Using A Variance-Type Estimator." *Journal of Time Series Analysis* 18 (3): 279–304, <https://doi.org/10.1111/1467-9892.00050>.
- Velasco, C. 1999. "Gaussian Semiparametric Estimation of Non-Stationary Time Series." *Journal of Time Series Analysis* 20 (1): 87–127, <https://doi.org/10.1111/1467-9892.00127>.