ORIGINAL PAPER

Sieve bootstrapping the memory parameter in long-range dependent stationary functional time series

Han Lin Shang[1]

Abstract

We consider a sieve bootstrap procedure to quantify the estimation uncertainty of long-memory parameters in stationary functional time series. We use a semiparametric local Whittle estimator to estimate the long-memory parameter. In the local Whittle estimator, discrete Fourier transform and periodogram are constructed from the first set of principal component scores via a functional principal component analysis. The sieve bootstrap procedure uses a general vector autoregressive representation of the estimated principal component scores. It generates bootstrap replicates that adequately mimic the dependence structure of the underlying stationary process. We first compute the estimated first set of principal component scores for each bootstrap replicate and then apply the semiparametric local Whittle estimator to estimate the memory parameter. By taking quantiles of the estimated memory parameters from these bootstrap replicates, we can nonparametrically construct confidence intervals of the long-memory parameter. As measured by coverage probability differences between the empirical and nominal coverage probabilities at three levels of significance, we demonstrate the advantage of using the sieve bootstrap compared to the asymptotic confidence intervals based on normality.

Keywords Functional principal component analysis · Functional autoregressive fractionally integrated moving average · Vector autoregression · Local Whittle estimator · Long-run covariance function

1 Introduction

The past few decades have seen extensive studies and developments in analyzing long-range dependence (LRD) time series, which appear to exist in many fields, such as agriculture, economics, finance, and geophysics (see, e.g., Beran 1994;

✉ Han Lin Shang
hanlin.shang@mq.edu.au

[1] Department of Actuarial Studies and Business Analytics, Macquarie University, Sydney, Australia

Robinson 2003; Palma 2007; Giraitis et al. 2012; Beran et al. 2013). These books describe stochastic processes with greater persistence than short-range dependent ones. By greater persistence, the autocovariance for LRD processes decays to zero more slowly than for short-range ones. Indeed, the autocovariance is not summable, and the spectral density is unbounded at zero frequency.

One of the most important issues in analyzing LRD time series is estimating the memory parameter, which quantifies the strength of persistence. Li et al. (2020) considered a rescale/range (R/S) estimator to estimate the memory parameter. While the R/S estimator is consistent, it has a slow convergence rate and performs poorly in finite samples, especially small samples. For improving the convergence rate, Li et al. (2021) considered a semiparametric local Whittle estimator, which is known to be more efficient. Further, Li et al. (2021) provided asymptotic results for parametrically constructing confidence intervals of the memory parameter. The asymptotic confidence intervals are the benchmark for our empirical comparison. Through a series of Monte-Carlo studies, Shang (2020) evaluated and compared time-domain and frequency-domain estimators and found that the local Whittle estimator is often one of the most accurate estimators in long-range dependent stationary functional time series. Thus, we consider the local Whittle estimator to estimate the memory parameter, although the sieve bootstrap can also be applied to other memory estimators.

Let $\{\mathcal{X}_t : t \in \mathbb{Z}\}$ be a sequence of functional observations, where each $\mathcal{X}_t$ is a random function of a stochastic process $(\mathcal{X}_t(u) : u \in \mathcal{I})$, $\mathcal{I} \subset \mathbb{R}$ is a compact set, $\mathbb{R}$ is the real line and $\mathbb{Z} = \{0, \pm 1, \dots\}$. Generally speaking, two major stationary short-range dependent functional time series structures have been considered in the literature: one extends small-ball probability for mixing sequences (see, e.g., Ferraty and Vieu 2006; Bathia et al. 2010); the other extends linear and nonlinear sequences and martingale using m-dependent approximation techniques (see, e.g., Bosq 2000; Hörmann and Kokoszka 2010; Horváth and Kokoszka 2012; Rice and Shang 2017). It is assumed that $\mathcal{X}_t = g(\epsilon_t, \epsilon_{t-1}, \dots)$, where $g : \mathcal{S}^\infty \to \mathcal{H}$ and $\{\epsilon_t : t \in \mathbb{Z}\}$ with $\epsilon_t = (\epsilon_t(u) : u \in \mathcal{I})$ is a sequence of independent and identically distributed (i.i.d.) random elements in a measurable space $\mathcal{S}$, and $\mathcal{H}$ is a separable Hilbert space. In Sect. 4, we follow the second structure to generate observations from a stochastic process that follows a functional autoregressive fractionally integrated moving average (ARFIMA) model.

With a time series of functions $(\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n)$, a central issue is to model the temporal dependence accurately. A challenge associated with functional time series is the curse of dimensionality. So, a common practice is to project a time series of functions onto a dominant subspace, such as the first eigenfunction of long-run covariance function (see, e.g., Li et al. 2020, 2021, 2022; Chen and Pun 2021). In so doing, we simplify the problem from a functional to univariate time series analysis.

Long-run covariance and spectral density estimation enjoy a vast literature in functional time series, beginning with the seminal work of Horváth et al. (2012) and Panaretos and Tavakoli (2013). Still, the most commonly used technique is smoothing the periodogram at frequency zero by employing a weight function and bandwidth parameter. While Li et al. (2020) considered a set of linearly decaying weights to estimate the long-run covariance, Rice and Shang (2017) considered a kernel sandwich estimator

and presented a plug-in algorithm to estimate the optimal bandwidth parameter in the kernel sandwich estimator.

Our estimation procedure begins with estimating the long-run covariance function. We then obtain the first functional principal component and its associated scores via eigendecomposition of the estimated long-run covariance function. From a univariate time series of the principal component scores, we apply the semiparametric local Whittle estimator, described in Sect. 2, to estimate the memory parameter. Li et al. (2021) presented theoretical development of the local Whittle estimator for stationary functional time series and provided asymptotic confidence intervals for the memory parameter. We aim to improve the confidence interval calibration through the sieve bootstrap method in Sect. 3. Via a series of simulation studies in Sect. 4, we evaluate and compare the finite-sample performance between the asymptotic and bootstrapped confidence intervals of the memory parameters. The conclusion is given in Sect. 5, along with some idea on how the methodology presented here can be further extended.

2 Estimation of the long-memory parameter

2.1 Estimation of the long-run covariance function

We consider a stationary ergodic functional time series, which exhibits both stationarity and ergodicity. In essence, the stochastic process will not change its statistical properties with time, and its statistical properties can be captured from a single, sufficiently long sample of the process. For such a random process, the long-run covariance function can be defined as

$$
\begin{aligned}
C(u, v) &= \sum_{\ell=-\infty}^{\infty} \gamma_\ell(u, v) \\
&= \sum_{\ell=-\infty}^{\infty} \mathrm{cov}[\mathcal{X}_0(u), \mathcal{X}_\ell(v)] \\
&= \sum_{\ell=-\infty}^{\infty} \mathrm{E}\{[\mathcal{X}_0(u) - \mu(u)][\mathcal{X}_\ell(v) - \mu(v)]\},
\end{aligned}
$$

where $u, v \in \mathcal{I}$ and ℓ denotes a time-series lag variable. A feature of long-memory processes is that the integral operator of $C(u, v)$ is not finite. As a result, the norm of the autocovariance function γ_ℓ of such a process decays much more slowly than that of the usual short-memory functional time series.

While the long-run covariance can be expressed as a bi-infinite summation, its estimation is not trivial. For a finite sample, a natural estimator of $C(u, v)$ is

$$
\widehat{C}_n(u, v) = \frac{1}{n^{1+2d}} \sum_{|\ell|=0}^{|\ell| \leq n} (n - |\ell|) \widehat{\gamma}_\ell(u, v), \tag{1}
$$

where d denotes a memory parameter.

In practice, the population mean $\mu(u)$ can be estimated by its empirical counterpart $\overline{\mathcal{X}}(u) = \frac{1}{n} \sum_{t=1}^{n} \mathcal{X}_t(u)$; and the autocovariance $\gamma_\ell(u, v)$ can also be estimated by

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

From (1), the estimation of the long-run covariance function relies on knowing the value of the memory parameter. However, the first term on the right-hand side of (1) is a constant, and it does not affect the estimation of the orthonormal functions spanning the dominant subspace of functional time series. Following (Li et al. 2020), we perform eigendecomposition on

$$\widehat{\widehat{C}}_n(u, v) = \sum_{|\ell|=0}^{|\ell| \leq n} (n - |\ell|)\widehat{\gamma}_\ell(u, v). \tag{2}$$

From (2), we observe the long-run covariance function is a sum of autocovariance function with decreasing weights. In practice, it is essential to determine the optimal value of ℓ to balance the trade-off between squared bias and variance. In Li et al. (2020), ℓ is chosen as the minimum between sample size n and the number of discretized data points in a function. Alternatively, it is also the approach we take; one could consider a kernel sandwich estimator

$$\widehat{\widehat{C}}_n(u, v) = \sum_{\ell=-\infty}^{\infty} W_q\left(\frac{\ell}{h}\right)\widehat{\gamma}_\ell(u, v),$$

where h is called the bandwidth parameter, and $W_q(\cdot)$ is a symmetric weight function with bounded support of order q. As with any kernel estimator, the choice of kernel function is not as important as the bandwidth parameter. Therefore, Rice and Shang (2017) proposed a plug-in algorithm for determining the optimal bandwidth parameter that minimizes the asymptotic mean-squared normed error between the estimated and actual long-run covariance functions.

2.2 Dynamic functional principal component analysis

Via Mercer's lemma, the estimated long-run covariance function $\widehat{\widehat{C}}_n(u, v)$ can be approximated by

$$\widehat{\widehat{C}}_n(u, v) = \sum_{k=1}^{\infty} \theta_k \phi_k(u)\phi_k(v),$$

where $\theta_1 > \theta_2 > \cdots \geq 0$ are the eigenvalues of $\widehat{\widehat{C}}_n(u, v)$, and $[\phi_1(u), \phi_2(u), \ldots]$ are the orthonormal functional principal components. Because of the inner product in the Hilbert space, we could project a time series of functions onto a set of orthogonal

functional principal components. This results in the Karhunen-Loève expansion of a realization of the stochastic process,

$$\mathcal{X}_t(u) = \overline{\mathcal{X}}(u) + \sum_{k=1}^{\infty} \beta_{t,k}\phi_k(u),$$

where $\beta_{t,k} = \langle \mathcal{X}_t(u) - \overline{\mathcal{X}}(u), \phi_k(u) \rangle$ denotes the kth set of principal component scores for time t.

Since the eigenvalues are ordered in descending order, the first eigenvalue and its associated eigenfunction form the dominant subspace of the functional time series. It is custom to use the first set of principal component scores $\beta_{t,1}$ to determine the memory parameter (see also Li et al. 2020, 2021, 2022; Chen and Pun 2021). Alternatively, one could also consider taking the L_2 norm of several leading principal component scores (see, e.g., Li et al. 2020). The memory parameters estimated from both approaches are similar, so we choose to work with the former.

2.3 Local Whittle estimator

The local Whittle estimator is a Gaussian semiparametric estimation method to estimate the memory parameter based on the periodogram. Introduced by Künsch (1987), Robinson (1995) and Velasco (1999), this frequency-domain estimator does not require the specification of a parametric model for the data. Instead, it depends on the specification of the shape of the spectral density of the time series. The spectral density $f(\delta)$ of a stationary time series is assumed to satisfy

$$f(\delta) \sim G\delta^{-2d}, \qquad \delta \to 0,$$

where $0 < G < \infty$ and $-\frac{1}{2} < d < \frac{1}{2}$. To estimate d, we define an objective function

$$Q(G, d) = \frac{1}{m} \sum_{i=1}^{m} \left\{ \ln(G\delta_i^{-2d}) + \frac{I(\delta_i)}{G\delta_i^{-2d}} \right\},$$

where $\ln(\cdot)$ denotes natural logarithm, $I(\delta_i)$ denotes the sample periodogram, which is the square of discrete Fourier transform of the scores, $\delta_i = (2\pi i)/n$, $i = 1, \ldots, m$, and m is a positive integer and $m = o(n)$ (Robinson 1995). Customarily, $m = \lfloor n^{0.65} + 1 \rfloor$. By minimizing the objective function, we define the estimates

$$(\widehat{G}, \widehat{d}) = \operatorname*{argmin}_{0 < G < \infty,\, d \in \Theta} Q(G, d), \tag{3}$$

where the closed interval of admissible estimates of d, $\Theta = [-\frac{1}{2}, \frac{1}{2}]$ denotes an admissible range for stationary time series. While G and d can be estimated jointly in (3), one could also use an iterative estimation procedure

$$\widehat{d} = \operatorname*{argmin}_{d \in \Theta} R(d)$$

where

$$R(d) = \ln \widehat{G}(d) - \frac{2d}{m} \sum_{i=1}^{m} \ln \delta_i, \qquad \widehat{G}(d) = \frac{1}{m} \sum_{i=1}^{m} \delta_i^{2d} I(\delta_i).$$

Li et al. (2021, Theorem 1) shows that $\sqrt{m}(\widehat{d} - d) \xrightarrow{d} N(0, 1/4)$, as $n \to \infty$. Based on this asymptotic result, we can construct our parametric confidence intervals as:

$$\left[\widehat{d} + \frac{1}{2} \times m^{-\frac{1}{2}} \times Q\left(\frac{\alpha}{2}\right), \quad \widehat{d} + \frac{1}{2} \times m^{-\frac{1}{2}} \times Q\left(1 - \frac{\alpha}{2}\right) \right], \tag{4}$$

where $Q(\cdot)$ represents a quantile function of the standard Gaussian distribution, and α denotes a level of significance, such as $\alpha = 0.2, 0.05$ and 0.01.

3 Sieve bootstrapping

Bootstrapping has been receiving increasing attention in the functional time series literature to quantify estimation or prediction uncertainty. Franke and Nyarige (2019) proposed a residual-based bootstrap for functional autoregressions. Pilavakis et al. (2019) established theoretical results for the moving block and tapered block bootstrap. Shang (2018) considered a maximum entropy bootstrap procedure to study the estimation accuracy of the long-run covariance function. Paparoditis (2018) proposed a sieve bootstrap under the functional autoregressions and derived bootstrap consistency as the sample size and order of autoregression both tend to infinity. From a nonparametric perspective, Zhu and Politis (2017) proposed a kernel estimation of the first-order nonparametric functional autoregression model and its bootstrap approximation.

We revisit the sieve bootstrap of Paparoditis (2018). The basic idea of the sieve bootstrap is to generate a functional time series of pseudo-random elements $(\mathcal{X}_1^*, \mathcal{X}_2^*, \ldots, \mathcal{X}_n^*)$, which appropriately imitate the temporal dependence of the original functional time series. The sieve bootstrapping begins with centering the observed functional time series by computing $\mathcal{Y}_t = \mathcal{X}_t - \overline{\mathcal{X}}_n$, where $\overline{\mathcal{X}}_n = \frac{1}{n} \sum_{t=1}^{n} \mathcal{X}_t$. Via Karhunen-Loève representation, random element $\mathcal{Y}_t$ can be decomposed into two parts:

$$\begin{aligned} \mathcal{Y}_t &= \sum_{k=1}^{\infty} \beta_{k,t} \phi_k \\ &= \sum_{k=1}^{K} \beta_{k,t} \phi_k + \sum_{k=K+1}^{\infty} \beta_{k,t} \phi_k, \\ &= \sum_{k=1}^{K} \beta_{k,t} \phi_k + U_{t,K} \end{aligned} \tag{5}$$

where K denotes the number of retained functional principal components. While the first terms on the right-hand side of (5) are considered as the main driving part of $\mathcal{X}_t$, the second term is treated as white noise.

The value of K is determined as the integer that minimizes ratios of two adjacent empirical eigenvalues given by

$$K = \underset{1 \leq k \leq k_{max}}{\mathrm{argmin}} \left\{ \frac{\widehat{\lambda}_{k+1}}{\widehat{\lambda}_k} \times \mathbb{1}\left(\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1} \geq \tau \right) + \mathbb{1}\left(\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1} < \tau \right) \right\},$$

where $\widehat{\lambda}_k$ represents the kth estimated eigenvalue from the functional principal component analysis, $k_{\max}$ is a pre-specified positive integer, τ is a pre-specified small positive-valued number, and $\mathbb{1}(\cdot)$ is the binary indicator function. Without a prior knowledge about a possible maximum value of k, it is unproblematic to choose a relatively large $k_{\max}$, e.g., $k_{\max} = \#\{k | \widehat{\lambda}_k \geq \frac{1}{n} \sum_{k=1}^{n} \widehat{\lambda}_k\}$ (Ahn and Horenstein 2013). Since small empirical eigenvalues $\widehat{\lambda}_k$ are close to zero, we adopt the threshold constant $\tau = \frac{1}{\ln[\max(\widehat{\lambda}_1, n)]}$ to ensure consistency of k.

We compute sample variance operator $\frac{1}{n} \sum_{t=1}^{n} \mathcal{Y}_t \otimes \mathcal{Y}_t$, and then obtain the first K set of estimated orthonormal eigenfunctions ($\widehat{\phi}_k, k = 1, 2, \ldots, K$) corresponding to the K largest estimated eigenvalues. By projecting a time series of functions onto these orthonormal eigenfunctions, we obtain a time series of estimated, K-dimensional vector of scores; that is,

$$\widehat{\boldsymbol{\beta}}_t = \sum_{\omega=1}^{\mathsf{p}} \widehat{\boldsymbol{A}}_{\omega,\mathsf{p}} \widehat{\boldsymbol{\beta}}_{t-\omega} + \widehat{\boldsymbol{e}}_t,$$

where $t = \mathsf{p} + 1, \mathsf{p} + 2, \ldots, n$, with $\widehat{\boldsymbol{e}}_t$ being the estimated residuals. The order p of the fitted VAR model is chosen using a corrected Akaike information criterion (Hurvich and Tsai 1993), that is, by minimizing

$$\mathrm{AICC}(\mathsf{p}) = n \ln \left| \widehat{\boldsymbol{\Sigma}}_{e,\mathsf{p}} \right| + \frac{n(nK + \mathsf{p}K^2)}{n - K(\mathsf{p} + 1) - 1},$$

over a range of values of p. $\widehat{\boldsymbol{\Sigma}}_{e,\mathsf{p}} = \frac{1}{n} \sum_{t=\mathsf{p}+1}^{n} \widehat{\boldsymbol{e}}_{t,\mathsf{p}} \widehat{\boldsymbol{e}}_{t,\mathsf{p}}^{\top}$ is the variance of the residuals and $\widehat{\boldsymbol{e}}_{t,\mathsf{p}}$ is the residuals after fitting the VAR(p) model to the K-dimensional time series of estimated scores ($\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \ldots, \widehat{\boldsymbol{\beta}}_n$). Computationally, the VARselect function in the vars package (Pfaff 2008) is implemented for selecting the optimal VAR order and estimating parameters.

Given the VAR is an autoregression, it requires a burn-in period to remove the effects of starting values. With a burn-in sample of 100, we generate the vector time series of scores

$$\boldsymbol{\beta}_t^* = \sum_{\omega=1}^{\mathsf{p}} \widehat{\boldsymbol{A}}_{\omega,\mathsf{p}} \boldsymbol{\beta}_{t-\omega}^* + \boldsymbol{e}_t^*, \qquad t = 1, 2, \ldots, n + 100,$$

where we use the starting value $\boldsymbol{\beta}_t^* = \widehat{\boldsymbol{\beta}}_t$ for $t = 1, 2, \ldots, \mathsf{p}$. The bootstrap residuals $\boldsymbol{e}_t^*$ are i.i.d. resampled from the set of centered residuals $\{\widehat{\boldsymbol{e}}_t - \overline{\boldsymbol{e}}_n, t = \mathsf{p} + 1, \mathsf{p} + 2, \ldots, n\}$, and $\overline{\boldsymbol{e}}_n = \frac{1}{n-\mathsf{p}} \sum_{t=\mathsf{p}+1}^n \widehat{\boldsymbol{e}}_t$.

Through the Karhunen-Loève expansion in (5), we obtain bootstrap samples

$$\mathcal{X}_t^* = \overline{\mathcal{X}}_n + \sum_{k=1}^K \boldsymbol{\beta}_t^* \widehat{\phi}_k + U_{t,K}^*,$$

where $U_{t,K}^*$ are i.i.d. resampled from the set $\{\widehat{U}_{t,K} - \overline{U}_n, t = 1, 2, \ldots, n\}$, $\overline{U}_n = \frac{1}{n} \sum_{t=1}^n \widehat{U}_{t,K}$ and $\widehat{U}_{t,K} = \mathcal{Y}_t - \sum_{k=1}^K \widehat{\beta}_{t,k} \widehat{\phi}_k$. We discard the first 100 generated $\mathcal{X}_t^*$ observations and keep $(\mathcal{X}_{101}^*, \mathcal{X}_{102}^*, \ldots, \mathcal{X}_{n+100}^*)$ as the bootstrap generated pseudo functional time series.

4 Numerical studies

4.1 Functional ARFIMA model

We simulate realizations of a stochastic process that follows a functional ARFIMA model. The functional ARFIMA (p, d, q) can be defined as

$$\nabla^d \mathcal{X}_t(u) = \mathcal{Y}_t(u), \quad \nabla = 1 - B, \quad -1/2 < d < 1/2, \tag{6}$$

and

$$\mathcal{Y}_t(u) - \sum_{s=1}^p \int_{\mathcal{I}} \vartheta_s(u, v) \mathcal{Y}_{t-s}(v) \mathrm{d}v = \eta_t(u) + \sum_{s=1}^q \int_{\mathcal{I}} \psi_s(u, v) \eta_{t-s}(v) \mathrm{d}v, \tag{7}$$

where B denotes the backshift operator, η_t denotes the white noise operator, and $\vartheta_s(u, v)$ and $\psi_s(u, v)$ are the kernel functions. Provided the usual conditions on the autoregressive and moving-average operators are satisfied, the models (6) and (7) represent a stationary process if $d < 1/2$ and is invertible if $d > -1/2$.

The functional ARFIMA model can be viewed as a generalization of some widely used parametric time-series models. When $d = 0$, model in (6) and (7) becomes the functional autoregressive moving average of Klepsch et al. (2017). Further, when $q = 0$, it further reduces to the functional autoregressive model of Bosq (2000) and Liu et al. (2016); when $p = 0$, it reduces to the functional moving average model of Chen et al. (2016) and Aue and Klepsch (2017).

We consider generating a time series of functions through a functional ARFIMA(p, d, q) model, where function support $\mathcal{I} = [0, 1]$, $\{\eta_t : t \in \mathbb{Z}\}$ a sequence of i.i.d. standard Brownian motions over $[0, 1]$, in the following two cases:

Case 1: $p = 1, q = 0, d = 0.05, 0.10, \ldots, 0.45, \vartheta_1(u, v) = c_{\mathrm{FAR}} \times \exp\{-(u^2 + v^2)/2\}$,

Case 2: $p = 1, q = 1, \ d = 0.05, 0.10, \dots, 0.45, \ \ \vartheta_1(u, v) = c_{FAR} \times \exp\{-(u^2 + v^2)/2\},$
$\psi_1(u, v) = c_{FMA} \times \min(u, v).$

The constants c_{FAR} and c_{FMA} in $\vartheta_1(u, v)$ and $\psi_1(u, v)$ are selected to ensure that both $\|\vartheta_1\|$ and $\|\psi_1\|$ are smaller than one (see, e.g., Rice and Shang 2017; Kokoszka et al. 2017), so the simulated curve time series are stationary and invertible. We consider the following sets of constants:

Set 1: $\|\vartheta_1\| = \|\psi_1\| = 0.5$, which corresponds to $c_{FAR} = 0.34$ and $c_{FMA} = 1.5$.
Set 2: $\|\vartheta_1\| = \|\psi_1\| = 0.9$, which corresponds to $c_{FAR} = 0.612$ and $c_{FMA} = 4.765$.

The sample sizes are $n = 250$ and 500, each with $R = 200$ replications.

As an illustration, in Fig. 1, we present one replicate of $n = 250$ functional time series generated from a ARFI(1, d) model with $d = 0.2$. Following (Mestre et al. 2021), we plot the functional autocorrelation and partial autocorrelation functions to check the linear temporal dependence in the simulated functional time series. For the ARFI(1, d) model, the functional partial autocorrelation function reveals a significant correlation at lag one. For the ARFIMA(1, d, 1) model, the functional autocorrelation and partial autocorrelation functions do not clearly indicate the autoregressive and moving-average orders.

In Fig. 2, we present one bootstrap sample using the sieve bootstrap method and compute its functional autocorrelation and partial autocorrelation functions. The bootstrap samples obtained from the sieve bootstrap method capture the linear temporal dependence exhibited in the original functional time series.

Via the local Whittle estimator, we estimate the memory parameter $\hat{d} = 0.3101$ and 0.3346 for simulated data generated from the ARFI(1, d) and ARFIMA(1, d, 1)
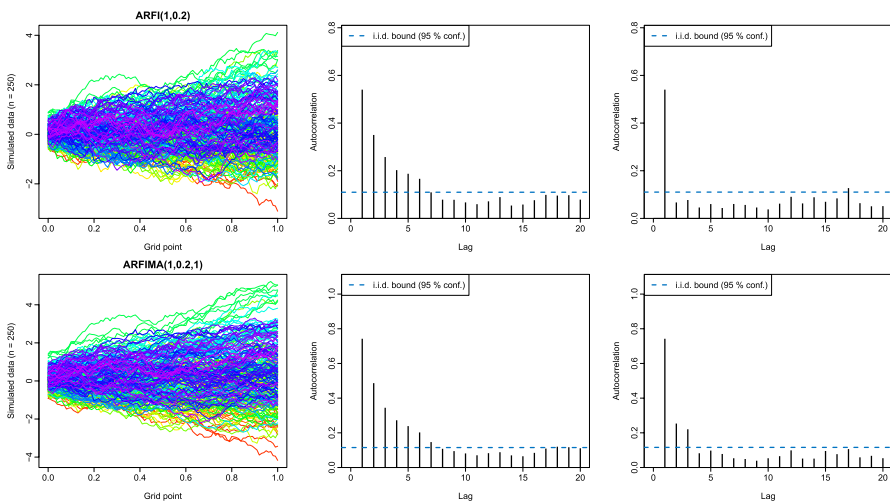

Fig. 1 A simulated functional time series generated from the functional ARFI(1, 0.2) model (top row) and the functional ARFIMA(1, 0.2, 1) model (bottom row) with $n = 250$
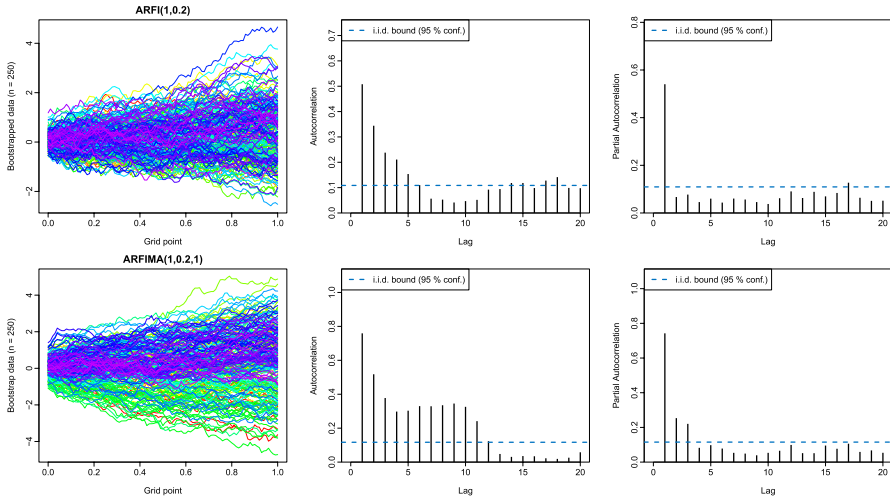
Fig. 2 A bootstrapped functional time series generated from the functional ARFI(1, 0.2) model (top row) and the functional ARFIMA(1, 0.2, 1) model (bottom row)
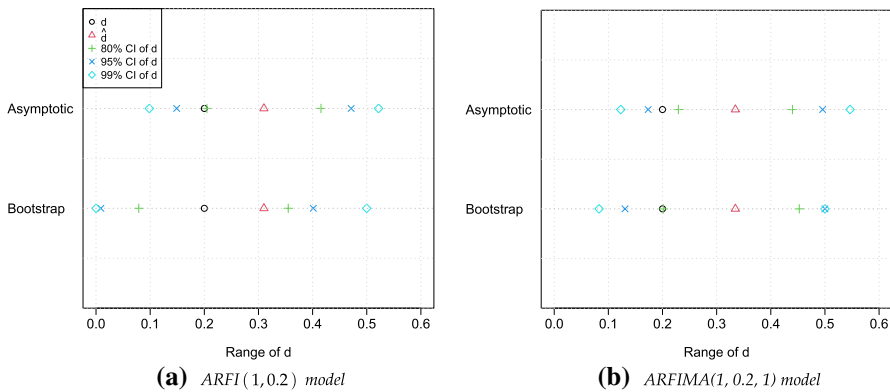
Fig. 3 For one replicate of the simulated functional time series generated from the ARFI(1, 0.2) and ARFIMA(1, 0.2, 1) model, we display the true and estimated memory parameters and the 80%, 95%, and 99% CIs using the asymptotic and sieve bootstrap procedures

models, respectively. Because the sample size $n = 250$ is relatively small, it is expected that there will be a considerable difference between the sample and population parameters. To quantify the estimation uncertainty, we construct 80%, 95%, and 99% CIs using the asymptotic and sieve bootstrap procedures in Fig. 3. The CIs based on the asymptotic normality are constructed from (4), while the CIs based on the sieve bootstrapping are percentiles of 400 bootstrap memory estimates.

Figure 3 displays the CIs constructed by the asymptotic and sieve bootstrap procedures for one simulated functional time series. To assess overall performance, we

repeat the following evaluation criteria for $R = 200$ replicates and report the results in Sect. 4.3.

4.2 Evaluation criteria of the interval forecast accuracy

To measure the interval forecast accuracy, we consider the empirical coverage probability. For each value of d, we compute the empirical coverage probability (ECP) at a level of significance α as

$$\text{ECP}_\alpha = 1 - \frac{1}{R} \sum_{r=1}^{R} \left[\mathbb{1}(d > \widehat{d}_{r,\alpha}^{\text{ub}}) + \mathbb{1}(d < \widehat{d}_{r,\alpha}^{\text{lb}}) \right], \tag{8}$$

where $\mathbb{1}(\cdot)$ represents the binary indicator function, and $(\widehat{d}_{r,\alpha}^{\text{lb}}, \widehat{d}_{r,\alpha}^{\text{ub}})$ represents the lower and upper bounds of asymptotic or bootstrap CIs for the rth replication. From the ECP_α, we compute the coverage probability difference (CPD_α), which is the absolute difference between the empirical and nominal coverage probabilities at various levels of significance.

4.3 Simulation results

We consider the asymptotic CIs based on (4). Implemented by Li et al. (2021), this parametric approach of constructing CIs provides symmetric lower and upper CIs. In contrast, the sieve bootstrapping produces a set of bootstrap functional time series. For each of the $B = 400$ bootstrap samples, we can implement the local Whittle estimator in Sect. 2 to compute an estimate of the memory parameter. By taking quantiles of all bootstrap memory parameter estimates, we obtain a nonparametric approach to constructing CIs. This nonparametric approach constructs lower and upper bounds that can be asymmetric.

In Table 1, we present the ECP_α in (8) and CPD_α for three levels of significance. Under a ARFI(1, d) model, these results are averaged over $R = 200$ replicates for $n = 250$ and 500. The CIs constructed by the sieve bootstrapping generally have a better calibration than those constructed by the asymptotic normality results. By better calibration, the empirical coverage probabilities are closer to the nominal coverage probabilities. This finding is not surprising since the CIs of the sieve bootstrap can be asymmetric around the sample estimate of the memory parameter. For the sieve bootstrap method, the empirical coverage probabilities gradually deviate from the nominal coverage probabilities as the memory parameter d increases from 0.05 to 0.45. This phenomenon holds for the functional time series with a moderate temporal dependence, and it becomes less so for the stronger dependence. With various significance levels, we observe that the mean CPDs obtained from the sieve bootstrap are generally smaller than those obtained from the asymptotic normality. As the sample size increases from $n = 250$ to 500, the CPD becomes smaller for both approaches. When the temporal dependence is

Table 1 With $n = 250$ and 500, 200 replicates of the functional time series are generated from the ARFI(1, d) model with $\|\vartheta_1\| = \|\psi_1\| = 0.5$ or 0.9

$\|\cdot\|$	n	Criterion	Asymptotic CIs				Bootstrap CIs		
			d	$\alpha = 0.20$	0.05	0.01	$\alpha = 0.20$	0.05	0.01
0.5	250	ECP	0.05	0.515	0.705	0.855	0.730	0.940	0.985
			0.10	0.525	0.715	0.845	0.710	0.945	0.990
			0.15	0.495	0.710	0.845	0.710	0.945	0.990
			0.20	0.475	0.695	0.845	0.715	0.935	0.985
			0.25	0.480	0.695	0.820	0.695	0.905	0.980
			0.30	0.505	0.690	0.995	0.680	0.890	0.985
			0.35	0.485	0.985	0.995	0.650	0.865	0.975
			0.40	0.980	0.985	0.990	0.600	0.865	0.970
			0.45	0.980	0.985	0.995	0.590	0.880	0.970
		CPD	0.05	0.285	0.245	0.135	0.070	0.010	0.005
			0.10	0.275	0.235	0.145	0.090	0.005	0.000
			0.15	0.305	0.240	0.145	0.090	0.005	0.000
			0.20	0.325	0.255	0.145	0.085	0.015	0.005
			0.25	0.320	0.255	0.170	0.105	0.045	0.010
			0.30	0.295	0.260	0.005	0.120	0.060	0.005
			0.35	0.315	0.035	0.005	0.150	0.085	0.015
			0.40	0.180	0.035	0.000	0.200	0.085	0.020
			0.45	0.180	0.035	0.005	0.210	0.070	0.020
			Mean	0.276	0.177	0.084	**0.124**	**0.042**	**0.009**
	500	ECP	0.05	0.665	0.865	0.940	0.845	0.955	0.985
			0.10	0.630	0.870	0.935	0.840	0.960	0.985
			0.15	0.630	0.870	0.935	0.830	0.955	0.985
			0.20	0.640	0.880	0.935	0.830	0.945	0.990
			0.25	0.635	0.875	0.930	0.810	0.945	0.985
			0.30	0.635	0.865	0.935	0.750	0.925	0.980
			0.35	0.620	0.870	1	0.715	0.905	0.970
			0.40	0.590	1	1	0.675	0.880	0.955
			0.45	0.975	1	1	0.655	0.870	0.960
		CPD	0.05	0.135	0.085	0.050	0.045	0.005	0.005
			0.10	0.170	0.080	0.055	0.040	0.010	0.005
			0.15	0.170	0.080	0.055	0.030	0.005	0.005
			0.20	0.160	0.070	0.055	0.030	0.005	0
			0.25	0.165	0.075	0.060	0.010	0.005	0.005
			0.30	0.165	0.085	0.055	0.050	0.025	0.010
			0.35	0.180	0.080	0.010	0.085	0.045	0.020
			0.40	0.210	0.050	0.010	0.125	0.070	0.035
			0.45	0.175	0.050	0.010	0.145	0.080	0.030
			Mean	0.170	0.073	0.040	**0.062**	**0.028**	**0.013**

Table 1 (continued)

‖ · ‖	n	Criterion	Asymptotic CIs				Bootstrap CIs		
			d	$\alpha = 0.20$	0.05	0.01	$\alpha = 0.20$	0.05	0.01
0.9	250	ECP	0.05	0.745	0.895	0.960	0.820	0.945	0.985
			0.10	0.735	0.890	0.960	0.815	0.945	0.975
			0.15	0.730	0.890	0.955	0.800	0.915	0.970
			0.20	0.740	0.885	0.950	0.810	0.920	0.970
			0.25	0.730	0.890	0.945	0.805	0.910	0.965
			0.30	0.720	0.890	0.960	0.815	0.900	0.965
			0.35	0.720	0.925	0.960	0.800	0.905	0.960
			0.40	0.875	0.925	0.960	0.795	0.915	0.955
			0.45	0.865	0.935	0.960	0.810	0.920	0.960
		CPD	0.05	0.055	0.055	0.030	0.020	0.005	0.005
			0.10	0.065	0.060	0.030	0.015	0.005	0.015
			0.15	0.070	0.060	0.035	0.000	0.035	0.020
			0.20	0.060	0.065	0.040	0.010	0.030	0.020
			0.25	0.070	0.060	0.045	0.005	0.040	0.025
			0.30	0.080	0.060	0.030	0.015	0.050	0.025
			0.35	0.080	0.025	0.030	0.000	0.045	0.030
			0.40	0.075	0.025	0.030	0.005	0.035	0.035
			0.45	0.065	0.015	0.030	0.010	0.030	0.030
			Mean	0.069	0.047	0.033	**0.009**	**0.031**	**0.023**
	500	ECP	0.05	0.740	0.895	0.965	0.785	0.930	0.990
			0.10	0.740	0.900	0.970	0.760	0.925	0.975
			0.15	0.740	0.905	0.975	0.705	0.925	0.975
			0.20	0.745	0.910	0.975	0.700	0.940	0.975
			0.25	0.730	0.915	0.970	0.700	0.935	0.980
			0.30	0.735	0.915	0.965	0.705	0.940	0.985
			0.35	0.750	0.915	0.980	0.775	0.940	0.985
			0.40	0.745	0.945	0.980	0.795	0.930	0.985
			0.45	0.865	0.945	0.980	0.820	0.945	0.980
		CPD	0.05	0.060	0.055	0.025	0.015	0.020	0.000
			0.10	0.060	0.050	0.020	0.040	0.025	0.015
			0.15	0.060	0.045	0.015	0.095	0.025	0.015
			0.20	0.055	0.040	0.015	0.100	0.010	0.015
			0.25	0.070	0.035	0.020	0.100	0.015	0.010

Table 1 (continued)

$\| \cdot \|$	n	Criterion	Asymptotic CIs				Bootstrap CIs		
			d	$\alpha = 0.20$	0.05	0.01	$\alpha = 0.20$	0.05	0.01
			0.30	0.065	0.035	0.025	0.095	0.010	0.005
			0.35	0.050	0.035	0.010	0.025	0.010	0.005
			0.40	0.055	0.005	0.010	0.005	0.020	0.005
			0.45	0.065	0.005	0.010	0.020	0.005	0.010
			Mean	0.060	0.034	0.017	**0.055**	**0.016**	**0.009**

For each replicate, we compute empirical coverage probabilities and coverage probability differences for the CIs constructed by the asymptotic and sieve bootstrap procedures at three levels of significance. Empirical coverage probability is abbreviated as ECP, while coverage probability difference is abbreviated as CPD. The smallest averaged CPD values are highlighted in bold for each level of significance

stronger, we observe a further improvement in estimation accuracy from the sieve bootstrapping.

In Table 2, we present the ECP_α in (8) and CPD_α for three levels of significance. Under an ARFIMA(1, d, 1) model, these results are averaged over $R = 200$ replicates for $n = 250$ and 500. The CIs constructed by the sieve bootstrapping also have a better calibration than those constructed by the asymptotic normality. As the sample size increases from $n = 250$ to 500, the CPD generally becomes smaller for both approaches. When the temporal dependence is stronger, we observe a further improvement in estimation accuracy from the sieve bootstrapping.

4.4 Simulation of functional time series via discrete Fourier transform

Following an early work of Li et al. (2021), we also consider another data generating process by using an algorithm of Davies and Harte (1987) to simulate functional time series. Let $\mathcal{X}_t$ be a 'fractional noise' process with autocovariance

$$\gamma_j = \frac{1}{2}\left(|j + 1|^{1+2d} - 2|j|^{1+2d} + |j - 1|^{1+2d}\right),$$

where $d = -0.3, -0.15, 0, 0.15, 0.3$. These parameter values are chosen to reflect negative dependent, short-range dependent and long-range dependent properties. For each n, let $g_k := g_{n,k}$, $k = 0, 1, \ldots, 2n - 1$, be the discrete Fourier transform of the real sequence $\{\gamma_0, \gamma_1, \ldots, \gamma_{n-1}, \gamma_n, \gamma_{n-1}, \ldots, \gamma_1\}$, i.e.,

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

and $g_k = g_{2n-k}$ for $k = n, \ldots, 2n - 1$. Let η_t be an i.i.d. standard Brownian motion sequence over a domain [0, 1] and then we construct

Table 2 With $n = 250$ and 500, $R = 200$ replicates of the functional time series are generated from the ARFIMA(1, d, 1) model with $\|\vartheta_1\| = \|\psi_1\| = 0.5$ or 0.9

$\|\cdot\|$	n	Criterion	d	Asymptotic CIs			Bootstrap CIs		
				$\alpha = 0.20$	0.05	0.01	$\alpha = 0.20$	0.05	0.01
0.5	250	ECP	0.05	0.420	0.605	0.790	0.575	0.860	0.965
			0.10	0.415	0.610	0.790	0.590	0.860	0.955
			0.15	0.380	0.610	0.795	0.600	0.860	0.960
			0.20	0.380	0.595	0.795	0.595	0.855	0.970
			0.25	0.365	0.595	0.795	0.595	0.830	0.970
			0.30	0.375	0.575	1.000	0.600	0.840	0.965
			0.35	0.380	0.985	1.000	0.580	0.835	0.960
			0.40	0.980	0.985	1.000	0.575	0.825	0.950
			0.45	0.985	0.985	1.000	0.555	0.815	0.940
		CPD	0.05	0.380	0.345	0.200	0.225	0.090	0.025
			0.10	0.385	0.340	0.200	0.210	0.090	0.035
			0.15	0.420	0.340	0.195	0.200	0.090	0.030
			0.20	0.420	0.355	0.195	0.205	0.095	0.020
			0.25	0.435	0.355	0.195	0.205	0.120	0.020
			0.30	0.425	0.375	0.010	0.200	0.110	0.025
			0.35	0.420	0.035	0.010	0.220	0.115	0.030
			0.40	0.180	0.035	0.010	0.225	0.125	0.040
			0.45	0.185	0.035	0.010	0.245	0.135	0.050
			Mean	0.361	0.246	0.114	**0.215**	**0.108**	**0.031**
	500	ECP	0.05	0.585	0.810	0.925	0.770	0.935	0.985
			0.10	0.575	0.820	0.925	0.780	0.940	0.980
			0.15	0.580	0.810	0.925	0.785	0.935	0.975
			0.20	0.585	0.815	0.925	0.770	0.925	0.970
			0.25	0.565	0.805	0.930	0.725	0.915	0.970
			0.30	0.545	0.815	0.925	0.665	0.905	0.965
			0.35	0.535	0.800	1	0.640	0.880	0.965
			0.40	0.515	1	1	0.630	0.865	0.960
			0.45	0.980	1	1	0.630	0.860	0.960
		CPD	0.05	0.215	0.140	0.065	0.030	0.015	0.005
			0.10	0.225	0.130	0.065	0.020	0.010	0.010
			0.15	0.220	0.140	0.065	0.015	0.015	0.015
			0.20	0.215	0.135	0.065	0.030	0.025	0.020
			0.25	0.235	0.145	0.060	0.075	0.035	0.020
			0.30	0.255	0.135	0.065	0.135	0.045	0.025
			0.35	0.265	0.150	0.010	0.160	0.070	0.025
			0.40	0.285	0.050	0.010	0.170	0.085	0.030
			0.45	0.180	0.050	0.010	0.170	0.090	0.030
			Mean	0.233	0.119	0.046	**0.089**	**0.043**	**0.020**

Table 2 (continued)

‖ · ‖	n	Criterion	d	Asymptotic CIs			Bootstrap CIs		
				$\alpha = 0.20$	0.05	0.01	$\alpha = 0.20$	0.05	0.01
0.9	250	ECP	0.05	0.710	0.870	0.970	0.850	0.975	1
			0.10	0.710	0.875	0.960	0.825	0.955	0.995
			0.15	0.675	0.870	0.955	0.820	0.950	0.995
			0.20	0.685	0.875	0.960	0.815	0.950	0.990
			0.25	0.690	0.880	0.955	0.815	0.950	0.990
			0.30	0.705	0.880	0.975	0.800	0.950	0.985
			0.35	0.695	0.945	0.975	0.820	0.945	0.985
			0.40	0.910	0.945	0.980	0.820	0.935	0.980
			0.45	0.900	0.950	0.980	0.815	0.935	0.980
		CPD	0.05	0.090	0.080	0.020	0.050	0.025	0.010
			0.10	0.090	0.075	0.030	0.025	0.005	0.005
			0.15	0.125	0.080	0.035	0.020	0.000	0.005
			0.20	0.115	0.075	0.030	0.015	0.000	0.000
			0.25	0.110	0.070	0.035	0.015	0.000	0.000
			0.30	0.095	0.070	0.015	0.000	0.000	0.005
			0.35	0.105	0.005	0.015	0.020	0.005	0.005
			0.40	0.110	0.005	0.010	0.020	0.015	0.010
			0.45	0.100	0.000	0.010	0.015	0.015	0.010
			Mean	0.104	0.051	0.022	**0.020**	**0.007**	**0.006**
	500	ECP	0.05	0.780	0.920	0.970	0.830	0.960	0.995
			0.10	0.790	0.925	0.970	0.815	0.965	0.985
			0.15	0.785	0.925	0.975	0.785	0.955	0.975
			0.20	0.785	0.930	0.975	0.775	0.960	0.985
			0.25	0.790	0.935	0.975	0.780	0.950	0.985
			0.30	0.780	0.930	0.975	0.790	0.955	0.985
			0.35	0.800	0.930	0.985	0.815	0.950	0.985
			0.40	0.795	0.950	0.980	0.825	0.950	0.985
			0.45	0.900	0.955	0.990	0.845	0.945	0.980
		CPD	0.05	0.020	0.030	0.020	0.030	0.010	0.005
			0.10	0.010	0.025	0.020	0.015	0.015	0.005
			0.15	0.015	0.025	0.015	0.015	0.005	0.015
			0.20	0.015	0.020	0.015	0.025	0.010	0.005
			0.25	0.010	0.015	0.015	0.020	0.000	0.005
			0.30	0.020	0.020	0.015	0.010	0.005	0.005

Table 2 (continued)

$\|\cdot\|$	n	Criterion	d	Asymptotic CIs			Bootstrap CIs		
				$\alpha = 0.20$	0.05	0.01	$\alpha = 0.20$	0.05	0.01
			0.35	0.000	0.020	0.005	0.015	0.000	0.005
			0.40	0.005	0.000	0.010	0.025	0.000	0.005
			0.45	0.100	0.005	0.000	0.045	0.005	0.010
			Mean	0.022	0.018	0.013	0.022	**0.006**	**0.007**

For each replicate, we compute empirical coverage probabilities and coverage probability differences for the CIs constructed by the asymptotic and sieve bootstrap procedures at three levels of significance. Empirical coverage probability is abbreviated as ECP, while coverage probability difference is abbreviated as CPD. The smallest averaged CPD values are highlighted in bold for each level of significance

$$\mathcal{X}_t = \frac{1}{2n^{1/2}} \left[\sqrt{2}\eta_0 g_0^{1/2} + \sqrt{2}\eta_n g_n^{1/2} + 2 \sum_{k=1}^{n-1} \eta_k g_k^{1/2} \cos\left(\frac{\pi kt}{n}\right) \right], \qquad 0 \le t \le n. \tag{9}$$

We take $n = 250$ and 500 with $R = 200$ replications. For each replication, we generate $B = 400$ bootstrap pseudo functional time series.

As an illustration, in Fig. 4, we present one replicate of $n = 250$ functional time series generated from (9) with $d = -0.3, 0$ and 0.3 to reflect negative dependence, short-range dependence and long-range dependence. We plot the functional autocorrelation and partial autocorrelation functions to examine the behavior of the linear temporal dependence.

In Table 3, we present the ECP_α and CPD_α for three levels of significance. These results are averaged over $R = 200$ replicates for $n = 250$ and 500. As measured by the averaged CPD, the CIs constructed by the sieve bootstrapping generally have a better calibration than those constructed by the asymptotic normality results. By better calibration, the ECPs are closer to the nominal coverage probabilities. The CIs obtained from the sieve bootstrapping have a better calibration when $d \le 0$, whereas the CIs obtained from the asymptotic normality have a better calibration when $d > 0$. Our finding further confirms the validity of the asymptotic CIs for long-memory functional time series. As the sample size increases from $n = 250$ to 500, the CPD becomes smaller for the parametric approach based on the asymptotic normality.

5 Conclusion

Not all long-memory estimators have an asymptotic distribution. For those other than the local Whittle estimator, it may not be possible to construct parametric confidence intervals of the memory parameter. In contrast, the sieve bootstrapping generates pseudo stationary functional time series, where the temporal dependence in the original functional time series is adequately preserved. For each bootstrap replicate,
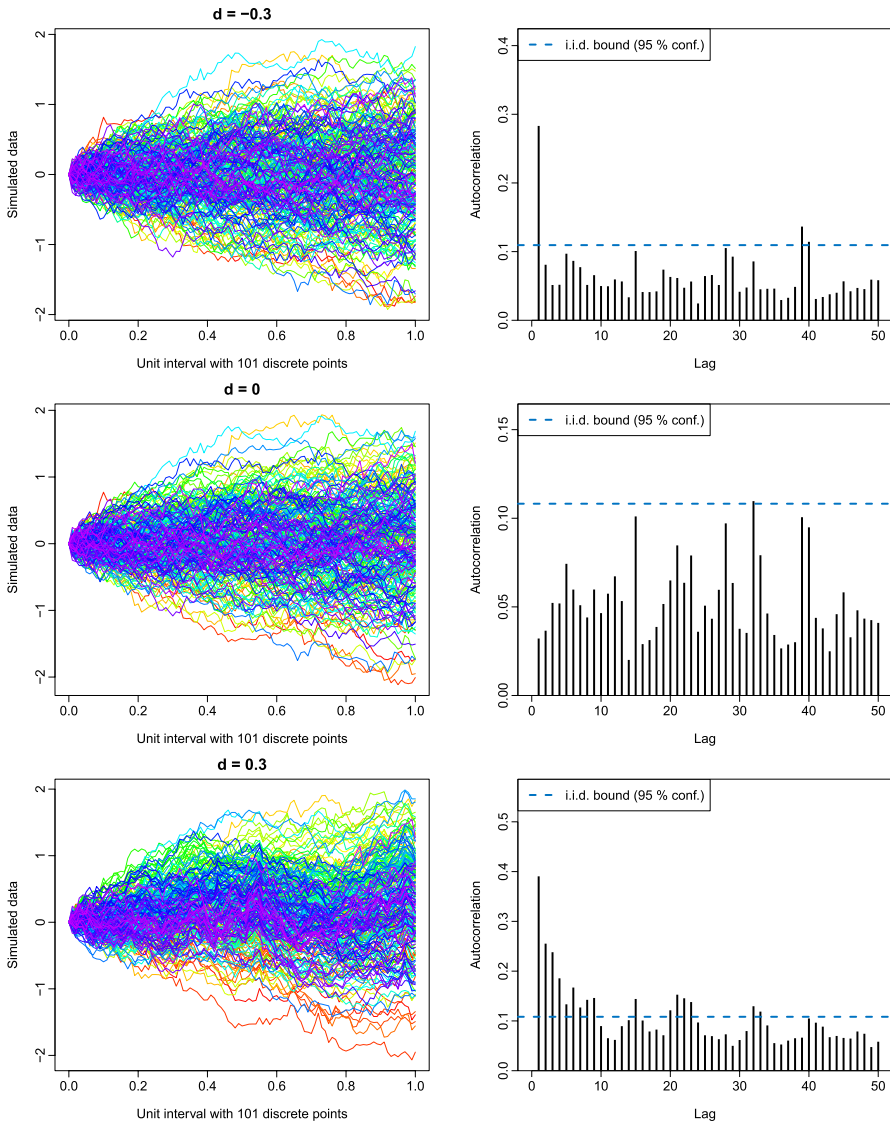
Fig. 4 With the same random seed, we simulate functional time series with $d = -0.3, 0$, and 0.3. These d values are chosen to reflect negative dependence, short-range dependence, and long-range dependence in the series. For each simulated series, we also display its corresponding functional autocorrelation function

any long-memory estimator can be applied. Therefore, the sieve bootstrap method presents a general framework for constructing confidence intervals of the memory parameter. Using the ARFI(1, d) and ARFIMA(1, d, 1) models, via a series of simulation studies, we evaluate and compare the empirical and nominal coverage probabilities between the asymptotic and bootstrap confidence intervals using the local

Table 3 With $n = 250$ and 500, 200 replicates of the functional time series are generated from the ARFI(1, d) model

n	Criterion	d	Asymptotic CIs			Bootstrap CIs		
			$\alpha = 0.20$	0.05	0.01	$\alpha = 0.20$	0.05	0.01
250	ECP	− 0.30	0.475	0.690	1	0.855	0.99	1
		− 0.15	0.595	0.765	0.885	0.860	0.960	0.995
		0	0.650	0.830	0.905	0.805	0.960	0.980
		0.15	0.675	0.845	0.920	0.580	0.875	0.965
		0.30	0.690	0.850	0.930	0.660	0.805	0.910
	CPD	− 0.30	0.325	0.260	0.010	0.055	0.040	0.010
		− 0.15	0.205	0.185	0.105	0.060	0.010	0.005
		0	0.150	0.120	0.085	0.005	0.010	0.010
		0.15	0.125	0.105	0.070	0.220	0.075	0.025
		0.30	0.110	0.100	0.060	0.140	0.145	0.080
		Mean	0.183	0.154	0.066	**0.096**	**0.056**	**0.026**
500	ECP	− 0.30	0.525	0.730	0.885	0.725	0.875	0.975
		− 0.15	0.590	0.820	0.920	0.845	0.990	0.990
		0	0.625	0.850	0.940	0.850	0.940	0.970
		0.15	0.695	0.870	0.940	0.555	0.780	0.945
		0.30	0.705	0.875	0.950	0.675	0.825	0.905
	CPD	− 0.30	0.275	0.220	0.105	0.075	0.075	0.015
		− 0.15	0.210	0.130	0.070	0.045	0.040	0.000
		0	0.175	0.100	0.050	0.050	0.010	0.020
		0.15	0.105	0.080	0.050	0.245	0.170	0.045
		0.30	0.095	0.075	0.040	0.125	0.125	0.085
		Mean	0.172	0.121	0.063	**0.108**	**0.084**	**0.033**

For each replicate, we compute empirical coverage probabilities and coverage probability differences for the CIs constructed by the asymptotic and sieve bootstrap procedures at three levels of significance. Empirical coverage probability is abbreviated as ECP, while coverage probability difference is abbreviated as CPD

The smaller averaged CPD values are highlighted in bold

Whittle estimator. Averaged over nine d values from 0.05 to 0.45, the sieve bootstrap confidence intervals produce a better calibration than the asymptotic confidence intervals, especially when the functional time series exhibits a stronger dependence. In addition, we also consider a simulation study where functional time series can exhibit negative dependent, short-range, and long-range dependent. Averaged over five d values from − 0.30 to 0.30, the sieve bootstrap confidence intervals produce a better calibration than the asymptotic confidence intervals, especially when the memory parameter $d \leq 0$.

There are several ways in which the present paper could be further extended. We briefly mention four below: (1) In the local Whittle estimator, the bandwidth parameter $m = \lfloor n^{0.65} + 1 \rfloor$ is used in the simulation studies. It is possible to explore other values centered around 0.65. (2) In the sieve bootstrap, we generate $B = 400$

bootstrap functional time series. It is possible to explore other values of bootstrap replication to achieve possible improvement in calibration. (3) Shang (2020) presented a comparison of point forecast accuracy, in terms of bias, variance, and mean square error, among several time-domain and frequency-domain memory estimators. The sieve bootstrap can evaluate interval forecast accuracy, in terms of empirical coverage probability, for those memory estimators. 4) With the sieve bootstrapping, it is possible to develop hypothesis tests for testing the presence of long memory.

References

Ahn, S.C., Horenstein, A.R.: Eigenvalue ratio test for the number of factors. Econometrica **81**(3), 1203–1227 (2013)

Aue, A., Klepsch, J.: Estimating functional time series by moving average model fitting. Technical report, University of California, Davis. (2017). arXiv:1701.00770

Bathia, N., Yao, Q., Ziegelmann, F.: Identifying the finite dimensionality of curve time series. Ann. Stat. **38**(6), 3352–3386 (2010)

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

Beran, J., Feng, Y., Ghosh, S., Kulik, R.: Long-Memory Processes: Probabilistic Properties and Statistical Methods. Springer, Berlin (2013)

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

Chen, Y., Pun, C.S.: Functional unit root test. Technical report, Nanyang Technological University (2021). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3761262

Chen, S.X., Lei, L., Tu, Y.: Functional coefficient moving average model with applications to forecasting Chinese CPI. Stat. Sin. **26**(4), 1649–1672 (2016)

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

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

Franke, J., Nyarige, E. G.: A residual-based bootstrap for functional autoregressions. Technical report, Technische Universität Kaiserslautern. (2019). arxiv:1905.07635

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

Hörmann, S., Kokoszka, P.: Weakly dependent functional data. Ann. Stat. **38**(3), 1845–1884 (2010)

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

Horváth, L., Kokoszka, P., Reeder, R.: Estimation of the mean of functional time series and a two sample problem. J. R. Stat. Soc. B **75**, 103–122 (2012)

Hurvich, C.M., Tsai, C.-L.: A corrected Akaike information criterion for vector autoregressive model selection. J. Time Ser. Anal. **14**(3), 271–279 (1993)

Klepsch, J., Klüppelberg, C., Wei, T.: Prediction of functional ARMA processes with an application to traffic data. Econom. Stat. **1**, 128–149 (2017)

Kokoszka, P., Rice, G., Shang, H.L.: Inference for the autocovariance of a functional time series under conditional heteroscedasticity. J. Multivar. Anal. **162**, 32–50 (2017)

Künsch, H.R.: Statistical aspects of self-similar processes. In: Proceedings of the World Congress of the Bernoulli Society, vol. 1, Tashkent, pp. 67–74 (1987)

Li, D., Robinson, P.M., Shang, H.L.: Long-range dependent curve time series. J. Am. Stat. Assoc. Theory Methods **115**(530), 957–971 (2020)

Li, D., Robinson, P.M., Shang, H.L.: Local Whittle estimation of long-range dependence for functional time series. J. Time Ser. Anal. **42**(5–6), 685–695 (2021)

Li, D., Robinson, P.M., Shang, H.L.: Nonstationary fractionally integrated functional time series. Bernoulli (2022) **(in press)**

Liu, X., Xiao, H., Chen, R.: Convolutional autoregressive models for functional time series. J. Econom. **194**(2), 263–282 (2016)

Mestre, G., Portela, J., Rice, G., San Roque, A.M., Alonso, E.: Functional time series model identification and diagnosis by means of auto- and partial autocorrelation analysis. Comput. Stat. Data Anal. **155**, 107108 (2021)

Palma, W.: Long-Memory Time Series. Wiley, Hoboken (2007)

Panaretos, V.M., Tavakoli, S.: Fourier analysis of stationary time series in function space. Ann. Stat. **41**(2), 568–603 (2013)

Paparoditis, E.: Sieve bootstrap for functional time series. Ann. Stat. **46**(6B), 3510–3538 (2018)

Pfaff, B.: Analysis of Integrated and Cointegrated Time Series with R, 2nd edn. Springer, New York (2008)

Pilavakis, D., Paparoditis, E., Sapatinas, T.: Moving block and tapered block bootstrap for functional time series with an application to the K-sample mean problem. Bernoulli **25**(4B), 3496–3526 (2019)

Rice, G., Shang, H.L.: A plug-in bandwidth selection procedure for long run covariance estimation with stationary functional time series. J. Time Ser. Anal. **38**(4), 591–609 (2017)

Robinson, P.M.: Gaussian semiparametric estimation of long range dependence. Ann. Stat. **23**(5), 1630–1661 (1995)

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

Shang, H.L.: Bootstrap methods for stationary functional time series. Stat. Comput. **28**(1), 1–10 (2018)

Shang, H.L.: A comparison of Hurst exponent estimators in long-range dependent curve time series. J. Time Ser. Econom. **12**(1), 1–39 (2020)

Velasco, C.: Gaussian semiparametric estimation of non-stationary time series. J. Time Ser. Anal. **20**(1), 87–127 (1999)

Zhu, T., Politis, D.N.: Kernel estimates of nonparametric functional autoregression models and their bootstrap approximation. Electron. J. Stat. **11**(2), 2876–2906 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.