Article:

Li, Degui orcid.org/0000-0001-6802-308X, Robinson, Peter M. and Shang, Hanlin (2020) Long-Range Dependent Curve Time Series. Journal of the American Statistical Association. pp. 957-971. ISSN 0162-1459

https://doi.org/10.1080/01621459.2019.1604362

Long-Range Dependent Curve Time Series

Degui Li
Department of Mathematics, University of York

Peter M. Robinson
Department of Economics, London School of Economics

Han Lin Shang
Research School of Finance, Actuarial Studies and Statistics,
Australian National University

Version: March 30, 2019

Abstract

We introduce methods and theory for functional or curve time series with long-range dependence. The temporal sum of the curve process is shown to be asymptotically normally distributed, the conditions for this covering a functional version of fractionally integrated autoregressive moving averages. We also construct an estimate of the long-run covariance function, which we use, via functional principal component analysis, in estimating the orthonormal functions spanning the dominant sub-space of the curves. In a semiparametric context, we propose an estimate of the memory parameter and establish its consistency. A Monte-Carlo study of finite-sample performance is included, along with two empirical applications. The first of these finds a degree of stability and persistence in intra-day stock returns. The second finds similarity in the extent of long memory in incremental age-specific fertility rates across some developed nations.

Keywords: curve process, functional FARIMA, functional principal component analysis, limit theorems, long-range dependence

1 Introduction

Functional or curve time series arise in many fields, including biology, transportation, environmental science, finance and demography (c.f., Chiou & Müller 2009, Kokoszka & Zhang 2012, Chen et al. 2016, Shang 2016). They can be of basically two types. On the one hand, they can arise by separating an almost continuous time record into natural consecutive intervals. Examples include intraday price curves of a financial stock (Kokoszka & Zhang 2012), minute-by-minute traffic flow (Klepsch et al. 2017) and electricity price curves (Chen & Li 2017). Figure 1 plots the intraday log-return curves for 6 US stocks observed from 1 October 2014 to 31 December 2014. This typical example will be analysed in Section 5.2. On the other hand, functional time series can also arise when observations in a time period are considered as finite realisations of a continuous function. Examples include age-specific mortality rates (Chiou & Müller 2009), age-specific fertility rates (Hyndman & Ullah 2007) and yield curves (Hays et al. 2012). Figure 2 plots the annual Australian age-specific fertility rates, where the function support for age lies between 15 and 49. This curve time series data set will be analysed in Section 5.3. In either case, the object of interest is a time series of random functions bounded within a finite interval.

A relatively recent but growing literature has developed methods and theory for functional time series, the bulk of it assuming stationarity over the temporal dimension, indeed short-range dependence (c.f., Bosq 2000, Ferraty & Vieu 2006, Bathia et al. 2010, Hörmann & Kokoszka 2010, Horváth & Kokoszka 2012, Horváth et al. 2014, Laurini 2014, Klepsch & Klüppelberg 2016, Liu et al. 2016, Aue et al. 2017, Aue & Klepsch 2017, Kowal et al. 2017, 2018, Rice & Shang 2017). This work has entailed both parametric and nonparametric modelling and is relevant to a wide variety of data.

On the other hand, there have been many notable developments in the time series literature, especially over the past thirty years or so, on long-range dependent, or long memory, time series models (c.f., Beran 1994, Robinson 2003, Palma 2007, Giraitis et al. 2012, Beran et al. 2013). These describe processes with greater persistence than short-range dependent ones, such that in the stationary case auto-covariances decay very

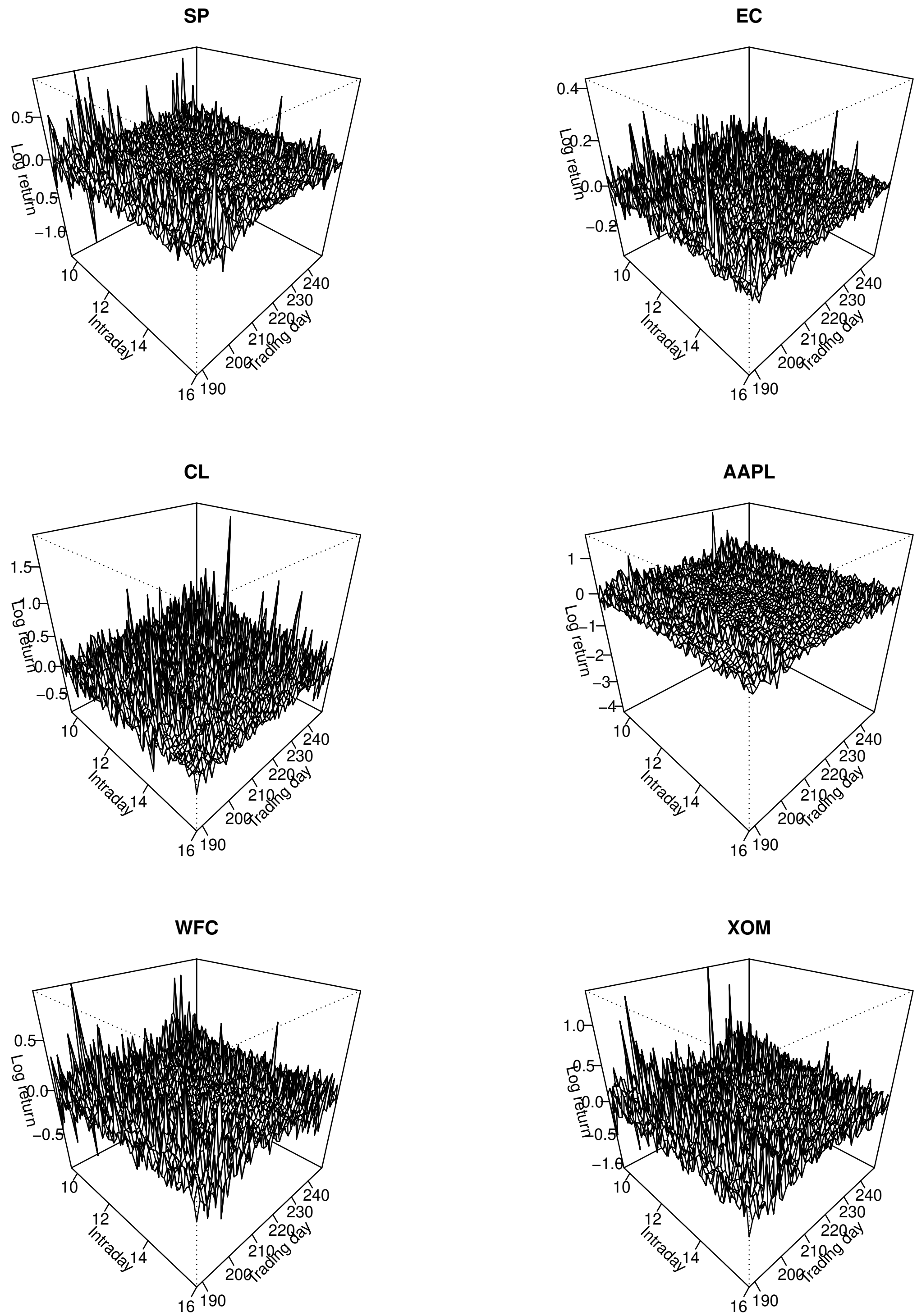


Figure 1: Perspective plots of the intraday log-return curves for 6 US stocks observed between 1 October 2014 and 31 December 2014.

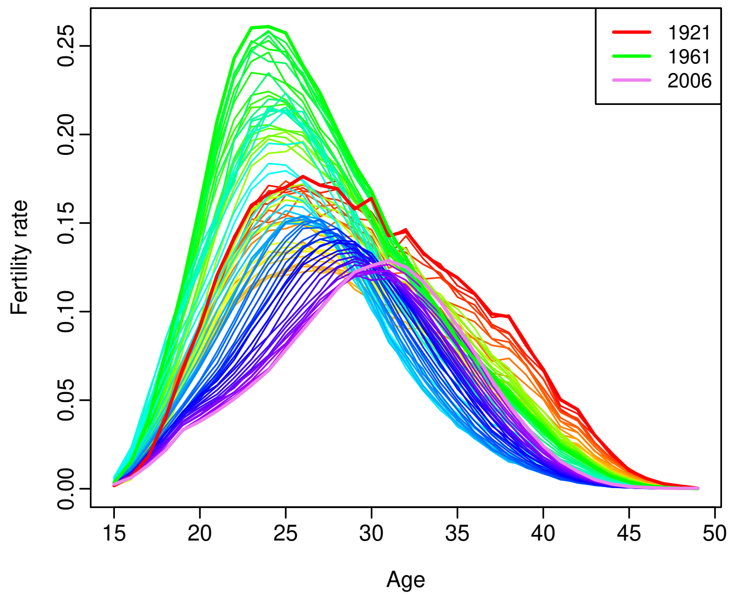


Figure 2: Age-specific fertility rates observed from 1921 to 2006 for ages from 15 through 49 in Australia. Curves are ordered chronologically according to the colours of the rainbow. Curves from the distant past are shown in red, while the most recent ones are in violet.

slowly and the spectral density is unbounded, typically at zero frequency. Time series data arising in a variety of areas of the natural sciences, such as geophysics, as well as in fields such as agriculture, economics and finance, have revealed evidence of long-range dependence.

It is natural then to expect long-range dependence to arise in functional time series. Indeed, Casas & Gao (2008) find evidence of long-range dependence in the daily volatility curve, when treated as a functional time series. However, so far as we know, even basic methodology for analysing functional time series with long-range dependence has not yet been developed. The present paper attempts a start at filling this gap. Given the lack of methods and theory, we focus initially on temporal sums of regularly-spaced observations across each curve, establishing asymptotic distribution theory for this simple statistic. Our result is then used in justifying some tools of statistical inference. Our work reflects both the parametric and semiparametric modelling found in the long-range dependence literature.

We suppose that $\{X_t : t \in \mathcal{Z}\}$ is a sequence of functional observations, where $X_t = (X_t(u) : u \in \mathcal{C})$, $\mathcal{C} \subset \mathcal{R}$ is a compact set, $\mathcal{R}$ is the real line and $\mathcal{Z} = \{0, \pm 1, \dots\}$. Much classic literature such as Ramsay & Silverman (2005) assumes that the functional process $\{X_t\}$ is independent and identically distributed (*i.i.d.*) across t, which seems too restrictive in many practical applications. Generally speaking, two major (stationary) short-range dependent functional time series structures have been studied: one extends probabilistic and statistical tools developed for mixing sequences (Ferraty & Vieu 2006, Bathia et al. 2010); the other extends certain linear and nonlinear sequences and martingale or m-dependent approximation techniques (Bosq 2000, Hörmann & Kokoszka 2010, Horváth & Kokoszka 2012, Rice & Shang 2017), where it is usually assumed that $X_t = g(\varepsilon_t, \varepsilon_{t-1}, \cdots)$, where $g : \mathcal{S}^\infty \to \mathcal{H}$ and $\{\varepsilon_t : t \in \mathcal{Z}\}$ with $\varepsilon_t = (\varepsilon_t(u) : u \in \mathcal{C})$ is a sequence of *i.i.d.* random elements in a measurable space $\mathcal{S}$, and $\mathcal{H}$ is a separable Hilbert space which will be defined in Section 2.1 below.

However, the above weak dependence structures may leave something to be desired. For instance, mixing is not easy to verify. Hörmann & Kokoszka (2010) give examples of nonlinear curve processes with short-range dependence, but this may be unreasonable in others (e.g., the two empirical examples in Sections 5.2 and 5.3). Thus, we study in this paper a curve time series process with long-range dependence. In order to specify the functional dependence structure, we will decompose the functional observations X_t through projection onto a finite number of sub-spaces, spanned by orthonormal basis functions, which are defined via eigenanalysis on a covariance function given in (2.3) below. The dependence degree for the projected curve linear process varies over different sub-spaces, as specified in Assumption 2 and Proposition 1 below. In particular, the sub-space on which the projection has the strongest dependence (or the strongest signal) is called the *dominant sub-space*. This sub-space typically contains most of the information carried by the original curve process and would play an important role in empirical studies. Under mild conditions, we establish a central limit theorem for the sum of the long-range dependent curve process and its projection (i.e., Theorem 1), extending well-known results developed for scalar time series.

We further consider estimation of the long-run covariance function for the curve process which is crucial in the subsequent *Functional Principal Component Analysis (FPCA)*. FPCA is a natural extension of *Principal Component Analysis (PCA)* in the multivariate setting. The classical PCA conducts eigenanalysis on variance-covariance or correlation matrices, obtaining the eigenvectors (corresponding to the first few largest eigenvalues) and subsequently forming linear combinations of components in the multivariate vector which highlight the variation strongly represented in the original multivariate data. In contrast, FPCA conducts eigenanalysis on variance or covariance functions, obtaining the eigenfunctions (corresponding to the first few largest eigenvalues) and forming linear combinations of functional observations which retain a large proportion of sample variation. A more detailed comparison between PCA and FPCA can be found in Ramsay & Silverman (2005) and Shang (2014). In this paper, through FPCA on the estimate of the long-run covariance function (up to multiplication by a rate), we obtain consistent estimates of the orthonormal functions spanning the sample dominant sub-space. In addition, we introduce two easy-to-implement methods to determine the dimension of the dominant sub-space and consistently estimate the memory parameter for the projected curve process onto this sub-space. We introduce a functional version of the fractionally integrated ARMA (FARIMA) process which is a natural extension of the scalar FARIMA (c.f., Adenstedt 1974, Granger & Joyeux 1980, Hosking 1981) and show that it satisfies our functional dependence structure; it is also employed in a Monte-Carlo study of finite-sample performance.

The rest of the paper is organised as follows. Section 2 specifies the long-range dependence structure for the curve process and gives some relevant limit theorems. Section 3 constructs estimates of the orthonormal functions which span the dominant sub-space, of the dimension of the dominant sub-space, and of the memory parameter. Section 4 studies the functional FARIMA process. Section 5 provides numerical studies including Monte-Carlo simulations and two empirical applications. Section 6 concludes the paper. Detailed proofs of the main theoretical results appear in a supplemental document (Li et al. 2018). Throughout the paper, "$\xrightarrow{\mathrm{d}}$" and "$\xrightarrow{\mathrm{P}}$" denote convergence in distribution and convergence in probability, respectively; "$a_n \sim b_n$" and "$a_n \propto b_n$" denote $a_n/b_n \to 1$ and $0 < c_1 \leqslant a_n/b_n \leqslant c_2 < \infty$, respectively, when n is sufficiently large.

2 Modelling structure and large-sample properties

In this section, we introduce our model and functional long-range dependence structure, give some technical assumptions with discussion, and state relevant large-sample properties.

2.1 Modelling structure

We model our functional time series by

$$X_t(u) = \sum_{j=0}^{\infty} \int_{\mathcal{C}} b_j(u,v)\eta_{t-j}(v)dv, \tag{2.1}$$

where $\{\eta_t : t \in \mathcal{Z}\}$ with $\eta_t = (\eta_t(u) : u \in \mathcal{C})$ is a sequence of *i.i.d.* random curves satisfying Assumption 1 in Section 2.2 below, and $\{b_j : j = 0, 1, 2, \cdots\}$ with $b_j = (b_j(u,v) : u, v \in \mathcal{C})$ is a sequence of kernels with associated integral operators defined by $B_j(x)(u) = \int_{\mathcal{C}} b_j(u,v)x(v)dv$, $x \in \mathcal{H}$. Here the space $\mathcal{H}$ is the set of measurable functions satisfying $\int_{\mathcal{C}} x^2(u)du < \infty$, a separable Hilbert space with inner product $\langle x_1, x_2\rangle = \int_{\mathcal{C}} x_1(u)x_2(u)du$. Let the operator norm of the coefficient operators in (2.1) be defined as

$$\|B_j\|_O = \sup\{\|B_j(x)\| : x \in \mathcal{H}, \|x\| = 1\},$$

where $\|\cdot\|$ is the L_2-norm of square integrable functions on $\mathcal{C}$. If summability of $\|B_j\|_O$ over $j = 0, 1, 2, \cdots$ is imposed, the auto-covariance functions between curve time series X_t and X_{t+k} can be shown to be absolutely summable. As a result, $\{X_t : t \in \mathcal{Z}\}$ defined in (2.1) has classic short-range dependence. This functional process has been systematically studied in the books by Bosq (2000) and Bosq & Blanke (2007), and extended by Hörmann & Kokoszka (2010) and Horváth & Kokoszka (2012).

In this paper, we consider the more challenging case when the operator norm of the kernel coefficient operators in (2.1) is not summable. Let the unnormalised long-run covariance function

$$c_n(u,v) = \mathsf{E}\left[\sum_{t=1}^{n}\sum_{s=1}^{n} X_t(u)X_s(v)\right] \tag{2.2}$$

be positive definite (or positive semi-definite). Hence, there exist $\lambda_{n1} \geqslant \lambda_{n2} \geqslant \cdots \geqslant 0$ and an orthonormal function sequence $\{\psi_{ni} : i = 1, 2, \cdots\}$ with $\psi_{ni} = (\psi_{ni}(u) : u \in \mathcal{C})$ such that

$$\lambda_{ni}\psi_{ni}(u) = \int_{\mathcal{C}} c_n(u,v)\psi_{ni}(v)dv. \tag{2.3}$$

For the case of independent functional data, eigenanalysis is usually conducted on the variance function $\mathsf{E}\left[X_t(u)X_t(v)\right]$, seeking an optimal linear combination of functional observations which "maximise" it (c.f., Ramsay & Silverman 2005). However, for curve time series, this approach loses dynamic sample information and the long-run covariance function is more appropriate (c.f., Horváth & Kokoszka 2012). As $c_n(u,v)$ defined in (2.2) is not normalised, it is easy to show that some eigenvalues λ_{ni} diverge to infinity. In particular, λ_{ni} reflects dependence in the scalar process

$$x_t^i = \int_{\mathcal{C}} X_t(u)\psi_{ni}(u)du, \tag{2.4}$$

which is the inner product of X_t and ψ_{ni}, and is usually referred to as the *score* in functional data analysis. From (2.3) and (2.4), we readily have

$$\lambda_{ni} = \int_{\mathcal{C}}\int_{\mathcal{C}} c_n(u,v)\psi_{ni}(v)\psi_{ni}(u)dvdu = \sum_{t=1}^{n}\sum_{s=1}^{n} \mathsf{E}\left[x_t^i x_s^i\right],$$

indicating that λ_{ni} is the unnormalised long-run variance of $\left\{x_t^i : t \in \mathcal{Z}\right\}$, $i = 1, 2, \cdots$. Note that

$$\|B_j\|_O^2 \leqslant \int_{\mathcal{C}}\int_{\mathcal{C}} b_j^2(u,v)dudv$$

and assume

$$\sum_{j=0}^{\infty} \|B_j\|_O^2 \leqslant \sum_{j=0}^{\infty} \int_{\mathcal{C}}\int_{\mathcal{C}} b_j^2(u,v)dudv < \infty, \tag{2.5}$$

which implies that the curve series defined in (2.1) converge almost surely. Using (2.4) and (2.5), we write

$$X_t(u) = \sum_{i=1}^{\infty} \psi_{ni}(u) \int_{\mathcal{C}} X_t(v)\psi_{ni}(v)dv = \sum_{i=1}^{\infty} x_t^i \cdot \psi_{ni}(u), \tag{2.6}$$

which indicates that we may specify the functional dependence structure for X_t via the scores x_t^i, $i = 1, 2, \cdots$.

Combining (2.1) and (2.4), it is straightforward to show that for each i

$$\begin{aligned} x_t^i &= \sum_{j=0}^{\infty} \int_{\mathcal{C}} \int_{\mathcal{C}} b_j(u,v)\eta_{t-j}(v)\psi_{ni}(u)dudv \\ &= \sum_{j=0}^{\infty} \int_{\mathcal{C}} b_j^i(v)\eta_{t-j}(v)dv, \end{aligned} \tag{2.7}$$

where $b_j^i(v) = \int_{\mathcal{C}} b_j(u,v)\psi_{ni}(u)du$, $v \in \mathcal{C}$. Throughout the paper, we suppress dependence of x_t^i and $b_j^i(\cdot)$ on n unless necessary. From (2.7), the temporal dependence of $\{x_t^i : t \in \mathcal{Z}\}$ is mainly determined by the decay rate of $b_j^i(v)$ over j. This will be specified in Assumption 2 below.

2.2 Assumptions

We next give some technical assumptions which will be used to derive the asymptotic theorems in Section 2.3 below.

Assumption 1. *The sequence* $\{\eta_t : t \in \mathcal{Z}\}$ *in (2.1) is composed of i.i.d. random functions in a measurable space with mean zero and positive definite covariance function defined by* $c_\eta(u,v) = \mathsf{E}[\eta_0(u)\eta_0(v)]$. *Furthermore,*

$$\|C_\eta\|_{\mathcal{S}}^2 := \int_{\mathcal{C}} \int_{\mathcal{C}} c_\eta^2(u,v)dudv < \infty, \tag{2.8}$$

where C_η *is the covariance operator defined by*

$$C_\eta(x)(u) = \int_{\mathcal{C}} c_\eta(u,v)x(v)dv, \quad x \in \mathcal{H},$$

which is Hilbert-Schmidt with $\|C_\eta\|_{\mathcal{S}}$ *being the Hilbert-Schmidt norm.*

Assumption 2. *There exist bounded positive integers κ_0 and $p_1 < p_2 < \cdots < p_{\kappa_0}$ such that, for $i = p_{k-1}+1, \cdots, p_k$ with $k = 1, \cdots, \kappa_0$ and $p_0 = 0$,*

$$b_j^i(u) \sim \rho_i(u) j^{-\alpha_k} \quad \text{as} \quad j \to \infty, \tag{2.9}$$

where $u \in \mathcal{C}$, $1/2 < \alpha_1 < \alpha_2 < \cdots < \alpha_{\kappa_0} < 1$, the functions $\rho_i(\cdot) := \rho_{ni}(\cdot)$ satisfy

$$0 < \bar{\rho}_{p_k} \leqslant \cdots \leqslant \bar{\rho}_{p_{k-1}+1} < \infty \quad \text{for} \quad k = 1, \cdots, \kappa_0$$

with $\bar{\rho}_i$ being the limit of $\left(\mathsf{E}\langle \rho_i, \eta_t\rangle^2\right)^{1/2} = \left(\mathsf{E}\langle \rho_{ni}, \eta_t\rangle^2\right)^{1/2}$ as $n \to \infty$; for $i \geqslant p_{\kappa_0}+1$,

$$\sum_{i=p_{\kappa_0}+1}^{\infty} \sum_{j=0}^{\infty} \left\|b_j^i\right\| < \infty, \quad \max_{i \geqslant p_{\kappa_0}+1} \sup_{1 \leqslant \iota \leqslant 2} \frac{\mathsf{E}\left[\langle b_{\lfloor \iota j \rfloor}^i, \eta_t\rangle^2\right]}{\mathsf{E}\left[\langle b_j^i, \eta_t\rangle^2\right]} = O(1) \quad \text{as} \quad j \to \infty, \tag{2.10}$$

where $\lfloor \cdot \rfloor$ denotes integer part.

Remark 1. Assumption 2 facilitates use of well-known limit results developed for scalar or multivariate long-range dependent processes (c.f., Robinson 1994, 2003, Giraitis et al. 2012). In particular, (2.9) indicates that $\{x_t^i : t \in \mathcal{Z}\}$, $i = p_{k-1}+1, \cdots, p_k$, have the same dependence degree, and moreover, as α_1 is the smallest memory parameter, $\{x_t^i : t \in \mathcal{Z}\}$, $i = 1, \cdots, p_1$, have the strongest degree of long-range dependence (see Proposition 1 and Theorem 1 below). Furthermore, as $\bar{\rho}_1 \geqslant \bar{\rho}_2 \geqslant \cdots \geqslant \bar{\rho}_{p_1}$, the first score process $\{x_t^1 : t \in \mathcal{Z}\}$ has the largest long-run variance, containing the strongest signal from the original curve time series. An approximation of this score process will be used in Section 3.2 to estimate α_1 via the R/S approach. Combining (2.6) and Assumption 2, we have

$$X_t(u) = \sum_{k=1}^{\kappa_0+1} \sum_{i=p_{k-1}+1}^{p_k} x_t^i \psi_{ni}(u) = \sum_{k=1}^{\kappa_0+1} X_{tk}(u), \tag{2.11}$$

where

$$X_{tk}(u) = \sum_{i=p_{k-1}+1}^{p_k} x_t^i \psi_{ni}(u),$$

$p_0 = 0$ and $p_{\kappa_0+1} = \infty$. In (2.11), X_t is decomposed into a summation of $X_{tk} := (X_{tk}(u) : u \in \mathcal{C})$, the projection of X_t onto $\mathcal{S}_k$, $k = 1, \cdots, \kappa_0+1$, where $\mathcal{S}_k = \mathcal{S}_k(\psi_{ni} : p_{k-1}+1 \leqslant i \leqslant p_k)$ is a $(p_k - p_{k-1})$-dimensional sub-space spanned by $\psi_{ni}, i =$

$p_{k-1}+1\cdots,p_k$, defined via (2.3), and $\mathcal{S}_{\kappa_0+1}=\mathcal{S}_{\kappa_0+1}(\psi_{ni}:i\geqslant p_{\kappa_0}+1)$ is the sub-space spanned by the eigenfunctions ψ_{ni}, $i=p_{\kappa_0}+1,p_{\kappa_0}+2,\cdots$. In addition, the positive integer κ_0 is the number of sub-spaces on which the projection of the original curve time series has long-range dependence.

Remark 2. Combining (2.9) and (2.10) in Assumption 2 leads to the summability condition (2.5), indicating that the integral operators $B_j(x)(u)=\int_{\mathcal{C}}b_j(u,v)x(v)dv$, $x\in\mathcal{H}$, are Hilbert-Schmidt for $j=0,1,2,\cdots$. Furthermore, the conditions in Assumption 2 determine the decay rate of the operator norms $\|B_j\|_O$ when j goes to infinity. In fact, by the Cauchy-Schwarz inequality and the definition of the operator norm, we have for any $i=p_{k-1}+1,\cdots,p_k$ with $k=1,\cdots,\kappa_0$, as $j\to\infty$,

$$\begin{aligned}
\bar{\rho}_i j^{-\alpha_k} &\sim \left\{\mathsf{E}\left[\int_{\mathcal{C}} b_j^i(v)\eta_t(v)dv\right]^2\right\}^{1/2}\\
&= \left\{\mathsf{E}\left[\int_{\mathcal{C}}\int_{\mathcal{C}} b_j(u,v)\eta_t(v)\psi_{ni}(u)dudv\right]^2\right\}^{1/2}\\
&= \left(\mathsf{E}\left\{\|\eta_t\|^2\left[\int_{\mathcal{C}}\int_{\mathcal{C}} b_j(u,v)\frac{\eta_t(v)}{\|\eta_t\|}\psi_{ni}(u)dudv\right]^2\right\}\right)^{1/2}\\
&\leqslant \left(\mathsf{E}\left\{\|\eta_t\|^2\int_{\mathcal{C}}\left[\int_{\mathcal{C}} b_j(u,v)\frac{\eta_t(v)}{\|\eta_t\|}dv\right]^2du\right\}\right)^{1/2}\left[\int_{\mathcal{C}}\psi_{ni}^2(u)du\right]^{1/2}\\
&\leqslant \left[\mathsf{E}\|\eta_t\|^2\right]^{1/2}\cdot\|B_j\|_O,
\end{aligned}$$

which together with Assumptions 1 and 2, implies that $\|B_j\|_O$ is not summable over $j=0,1,2,\cdots$.

Example 1 below connects (2.9) and (2.10) in Assumption 2 to the kernels b_j in the curve process (2.1), where the eigenfunctions of $c_n(\cdot,\cdot)$ coincide with those of $c_\eta(\cdot,\cdot)$.

Example 1. Let the kernels b_j in (2.1) be defined as $b_0=I$ (the identity operator), and

$$b_j(u,v)=j^{-\alpha_\star}\left[1+O(j^{-1})\right]B(u,v)+b_j^\star B^\star(u,v),\qquad j\geqslant 1,$$

where $1/2<\alpha_\star<1$, $\{b_j^\star:j=1,2,\cdots\}$ is a sequence of real numbers satisfying $\sum_{j=0}^\infty|b_j^\star|<$

∞ and $\sup_{1\leqslant\iota\leqslant 2}\left|b^{\star}_{\lfloor\iota j\rfloor}/b^{\star}_j\right| = O(1)$ as $j\to\infty$,

$$B(u,v)=\sum_{k=1}^{p}\gamma_k^{\star}\psi_k^{\star}(u)\psi_k^{\star}(v),\qquad B^{\star}(u,v)=\sum_{k=p+1}^{\infty}\gamma_k^{\star}\psi_k^{\star}(u)\psi_k^{\star}(v),$$

$\gamma_k^{\star}$, $k=1,2,\cdots$, are positive constants (in non-increasing order) and $\psi_k^{\star}$ are orthogonal eigenfunctions of $c_\eta(\cdot,\cdot)$ defined in Assumption 1 (with corresponding eigenvalues in non-increasing order). For $i=1,\cdots,p$, choosing $\psi_{ni}=\psi_i^{\star}$, we readily have

$$\int_{\mathcal{C}} b_j(u,v)\psi_{ni}(u)du \sim j^{-\alpha_\star}\sum_{k=1}^{p}\gamma_k^{\star}\psi_k^{\star}(v)\int_{\mathcal{C}}\psi_k^{\star}(u)\psi_i^{\star}(u)du = j^{-\alpha_\star}\gamma_i^{\star}\psi_i^{\star}(v)\quad\text{as}\quad j\to\infty,$$

and thus verify (2.9) with $\kappa_0=1$, $p_1=p$, $\alpha_1=\alpha_\star$ and $\rho_i(u)=\gamma_i^{\star}\psi_i^{\star}(u)$. If we further assume that $\sum_{k=1}^{\infty}|\gamma_k^{\star}|<\infty$, with the conditions on $b_j^{\star}$, we can verify (2.10) in Assumption 2.

2.3 Large-sample properties

We next present some large-sample properties, starting with a proposition which gives the orders of the λ_{ni}.

Proposition 1. *Suppose that Assumptions 1 and 2 are satisfied. Then*

$$\lambda_{ni}\sim\theta_i^2 n^{3-2\alpha_k},\quad \theta_i^2=\frac{\bar{\rho}_i^2 c_{\alpha_k}}{(1-\alpha_k)(3-2\alpha_k)}\tag{2.12}$$

for $i=p_{k-1}+1,\cdots,p_k$ *with* $k=1,\cdots,\kappa_0$*, and*

$$\sum_{i=p_{\kappa_0}+1}^{\infty}\lambda_{ni}=O(n),\tag{2.13}$$

where $c_{\alpha_k}=\int_0^\infty x^{-\alpha_k}(1+x)^{-\alpha_k}dx$*,* $\bar{\rho}_i$ *and* α_k *are defined as in Assumption 2.*

Remark 3. It follows from the decomposition (2.11) and Assumption 2 that $X_t(u)$ defined in (2.1) can be decomposed into a summation of κ_0 functional long-range dependent processes $\{X_{tk}:t\in\mathcal{Z}\}$, $k=1,\cdots,\kappa_0$, and a short-range dependent process $\{X_{t,\kappa_0+1}:t\in\mathcal{Z}\}$. From Proposition 1, we may also derive the explicit form of the asymptotic variance in the central limit theorem, see Theorem 1 and Remark 4 below. Furthermore, by Assumption 2 and Proposition 1, we find that α_k reflects the strength of the signal for the projection of X_t on $\mathcal{S}_k$. In particular, the projection of $X_t(u)$ onto $\mathcal{S}_1$ typically contains a large proportion of the information carried by the original process and thus results in the strongest signal, see (2.17) below.

We next investigate the limit distribution of temporal sums, defining

$$T_n(u) = \sum_{t=1}^{n} X_t(u) = \sum_{k=1}^{\kappa_0+1} \sum_{t=1}^{n} X_{tk}(u) = \sum_{k=1}^{\kappa_0+1} T_{nk}(u),$$

where $T_{nk}(u) = \sum_{t=1}^{n} X_{tk}(u)$. For notational simplicity, write $T_n = (T_n(u) : u \in \mathcal{C})$ and $T_{nk} = (T_{nk}(u) : u \in \mathcal{C})$. Proposition 1 suggests that in limit theorems normalisation rates for T_{nk} will differ over $k = 1, \cdots, \kappa_0 + 1$. Our main focus is on the cases $k = 1, \cdots, \kappa_0$ for which T_{nk} is constructed using the functional long-range dependent process X_{tk}, as limit theorems for T_{n,κ_0+1} have been extensively studied (c.f., Bosq 2000, Berkes et al. 2013). The following theorem establishes a central limit theorem for T_{nk}.

Theorem 1. *Suppose that Assumptions 1 and 2 are satisfied, and* $\max_{1\leqslant k\leqslant \kappa_0} (p_k - p_{k-1})$ *is bounded. Then for each* $k = 1, \cdots, \kappa_0$,

$$n^{-(3/2-\alpha_k)} \cdot T_{nk} \xrightarrow{d} Z_k, \tag{2.14}$$

where Z_k *is a Gaussian random element with zero mean and covariance function defined by*

$$\sigma_k(u, v) = \lim_{n\to\infty} \frac{1}{n^{3-2\alpha_k}} \mathsf{E}\left[T_{n,k}(u) T_{n,k}(v)\right] = \lim_{n\to\infty} \frac{1}{n^{3-2\alpha_k}} \sum_{i=p_{k-1}+1}^{p_k} \lambda_{ni} \psi_{ni}(u) \psi_{ni}(v). \tag{2.15}$$

Furthermore, Z_k, $k = 1, \cdots, \kappa_0$, *are mutually independent.*

Remark 4. With the conditions in Assumption 2 and Proposition 1, (2.14) can also be expressed as

$$n^{-(3/2-\alpha_k)} \cdot T_{nk} \xrightarrow{d} \sum_{i=p_{k-1}+1}^{p_k} \theta_i N_i \psi_i, \tag{2.16}$$

where θ_i is defined in Proposition 1, N_i, $i = 1, 2, \cdots, p_{\kappa_0}$, are independent standard normal random variables, and ψ_i is the limit of ψ_{ni} as specified in Section 3 below. In addition, Theorem 1 and Proposition 1 show that T_n is asymptotically dominated by T_{n1} and in particular,

$$n^{-(3-2\alpha_1)} \left\| C_n^{\mathsf{T}} - C_{n1}^{\mathsf{T}} \right\|_{\mathcal{S}} = o(1), \tag{2.17}$$

where C_n^{T} and C_{n1}^{T} are the covariance operators defined by

$$C_n^{\mathsf{T}}(x) = \mathsf{E}\left[\langle T_n, x\rangle T_n\right], \;\; C_{n1}^{\mathsf{T}}(x) = \mathsf{E}\left[\langle T_{n1}, x\rangle T_{n1}\right], \;\; x \in \mathcal{H}.$$

Therefore, we call $\mathcal{S}_1$ the asymptotically *dominant sub-space* and obtain the following result as a direct corollary of Theorem 1.

Corollary 1. *Suppose that the conditions in Theorem 1 are satisfied. Then*

$$n^{-(3/2-\alpha_1)} \cdot T_n \stackrel{d}{\longrightarrow} \mathsf{Z}_1, \tag{2.18}$$

where Z_1 is defined as in Theorem 1.

Remark 5. Theorem 1 and Corollary 1 above not only generalise some classic limit theorems for long memory time series (c.f., Davydov 1970, Robinson 2003, Giraitis et al. 2012) to the functional case, but also extend some existing theorems for short-range dependent curve processes (c.f., Bosq 2000, Horváth & Kokoszka 2012). Theorem 1 and Corollary 1 still hold with slight modification under the more general condition: $b_j^i(u) \sim \rho_i(u)\,[j^{-\alpha_k} l_i(j)]$ as $j \to \infty$, where $l_i(\cdot)$ is a positive slowly varying function (c.f., Bingham et al. 1987) which depends on i. Our limit distribution theory is comparable to theorems in Characiejus & Rauckauskas (2014) and Düker (2018) which consider Hilbert space-valued long-range dependent linear processes with derivation heavily relying on the theory of multiplication operator. However, it seems challenging to directly apply the methodology in Section 3 below to their model framework, making it difficult to achieve dimension reduction via FPCA.

Remark 6. In the above limit results, we assume the dimension p_1 of the dominant sub-space $\mathcal{S}_1$ (and of the other sub-spaces on which there is long-range dependence) is fixed, but we could allow p_1 to diverge slowly with n. The limit distribution results in Theorem 1 and Corollary 1 are still valid after modifying conditions and the proofs in Li et al. (2018). For example, in Corollary 1, we need the implicit restriction that

$$\lim_{n\to\infty} \frac{1}{n^{3-2\alpha_1}} \sum_{i=1}^{p_1} \lambda_{ni} < \infty,$$

which together with Proposition 1, indicates that $\sum_{i=1}^{p_1} \bar{\rho}_i^2$ is bounded. Hence, in case of divergent p_1, $\bar{\rho}_i$ would converge to zero as the index i approaches p_1 (from the left).

3 Estimation of the dominant sub-space

From (2.17) and Corollary 1, the projection of the curve process onto $\mathcal{S}_1$ typically contains the bulk of the information carried by $X_t(u)$ and in particular,

$$\sum_{i=1}^{p_1} \lambda_{ni} \Big/ \sum_{j=1}^{\infty} \lambda_{nj} = 1 + o(1).$$

Therefore, it is important to estimate $\mathcal{S}_1$. As the orthonormal functions $\psi_{n1}, \cdots, \psi_{np_1}$ depend on n, we define their limits as

$$\psi_i = \lim_{n\to\infty} \psi_{ni}, \;\; i = 1, \cdots, p_1. \tag{3.1}$$

In this section, we show how to estimate the long-run covariance function (up to multiplication by a rate) for the curve process and use it to obtain estimates of $\psi_1, \cdots, \psi_{p_1}$ via FPCA. Then we discuss how to consistently estimate p_1 and α_1.

3.1 Estimation of the orthonormal functions ψ_i

We start by estimating the long-run covariance function as it plays a key role in estimating the ψ_i, $i = 1, \cdots, p_1$. In order to carry out statistical inference on T_n defined in Section 2.3, for example to set confidence regions, we also need to consistently estimate the limiting covariance function implied by Corollary 1, and given by

$$c(u, v) = \lim_{n\to\infty} \frac{1}{n^{3-2\alpha_1}} \mathsf{E}\left[T_n(u) T_n(v)\right] = \lim_{n\to\infty} \frac{1}{n^{3-2\alpha_1}} c_n(u, v). \tag{3.2}$$

Consider the case of "I(0)" scalar time series z_t, given for example by a linear process:

$$z_t = \sum_{j=0}^{\infty} b_j^{\diamond} \epsilon_{t-j}, \quad 0 < \left| \sum_{j=0}^{\infty} b_j^{\diamond} s^j \right| < \infty \tag{3.3}$$

for s on the unit circle of the complex plane, where $\{\epsilon_t : t \in \mathbb{Z}\}$ is a sequence of *i.i.d.* scalar random variables with zero mean and finite and positive variance σ^2. The asymptotic variance of $n^{-1/2} \sum_{t=1}^{n} z_t$ is $\sigma^2 \big(\sum_{j=0}^{\infty} b_j^{\diamond} \big)^2$, equivalently 2π times the spectral density of z_t at zero frequency. The latter can be consistently estimated using nonparametric spectral density estimation. A great deal has been made of this topic in the econometric literature, in the context of more general mean-like statistics, and with the introduction of such terminology as "Heteroskedastic and Autocorrelation Consistent (HAC)" and "long-run variance" estimation, but not only are these estimates based heavily on ideas from the classical nonparametric spectral estimation literature, but the idea of studentising a sample mean by such an estimate goes back to Jowett (1955) and Hannan (1957), so relevant statistical literature long precedes the econometric work. If

$$b_0^{\diamond} = 1, \;\; b_j^{\diamond} \sim j^{-\alpha} \;\text{ with }\; 1/2 < \alpha < 1 \;\text{ as }\; j \to \infty$$

in the scalar linear process (3.3), z_t is not $I(0)$ but $I(d)$, $0 < d := 1 - \alpha < 1/2$, and its spectral density diverges at zero frequency, so such methods cannot be used for inference on $n^{-H} \sum_{t=1}^{n} z_t$ with $H = 3/2 - \alpha$. However, in the latter case, Robinson (1994, 2005) and Abadir et al. (2009) develop and justify suitable studentisations, depending in part on consistent estimates of H (indeed these apply also to "antipersistent" series $I(d)$ with $-1/2 < d < 0$, i.e., $0 < H < 1/2$, where the spectral density vanishes at frequency zero).

In our functional time series setting, let $\overline{X}_n(u) = \frac{1}{n} \sum_{t=1}^{n} X_t(u)$ and

$$\bar{r}_k(u, v) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-k} \left[X_t(u) - \overline{X}_n(u) \right] \left[X_{t+k}(v) - \overline{X}_n(v) \right], & k \geqslant 0, \\ \frac{1}{n} \sum_{t=1}^{n-|k|} \left[X_{t+|k|}(u) - \overline{X}_n(u) \right] \left[X_t(v) - \overline{X}_n(v) \right], & k < 0, \end{cases}$$

where $|k| \leqslant m$, $m = m_n$ is a user-chosen bandwidth sequence satisfying $m \to \infty$ and

$m = o(n)$. If α_1 is known a priori, define

$$\bar{c}_m(u,v) = \frac{1}{m^{3-2\alpha_1}} \sum_{|k| \leqslant m} (m - |k|)\, \bar{r}_k(u,v) \tag{3.4}$$

as an estimate of $c(u,v)$. Let $\overline{C}_m$ and C be the operators defined by

$$\overline{C}_m(x)(u) = \int_{\mathcal{C}} \bar{c}_m(u,v)x(v)dv, \quad C(x)(u) = \int_{\mathcal{C}} c(u,v)x(v)dv, \quad x \in \mathcal{H}.$$

The following proposition shows the consistency of $\overline{C}_m$.

Proposition 2. *Suppose that the conditions in Theorem 1 are satisfied,* $\mathsf{E}\left[\|\eta_t\|^4\right] < \infty$ *and* $m \propto n^\gamma$ *with* $0 < \gamma < \min\{1/(4\alpha_1 - 2), 1\}$*. Then*

$$\left\|\overline{C}_m - C\right\|_{\mathcal{S}} = o_P(1). \tag{3.5}$$

Remark 7. If α_1 is unknown, (3.4) is infeasible. For an estimate $\widetilde{\alpha}_1$ such that $\widetilde{\alpha}_1 - \alpha_1 = o_P(1/\log n)$, consider

$$\widetilde{c}_m(u,v) = \frac{1}{m^{3-2\widetilde{\alpha}_1}} \sum_{|k| \leqslant m} (m - |k|)\, \bar{r}_k(u,v).$$

Note that

$$\begin{aligned} \widetilde{c}_m(u,v) - c(u,v) &= \widetilde{c}_m(u,v) - \bar{c}_m(u,v) + \bar{c}_m(u,v) - c(u,v) \\ &= \left(\frac{1}{m^{3-2\widetilde{\alpha}_1}} - \frac{1}{m^{3-2\alpha_1}}\right) \sum_{|k| \leqslant m} (m - |k|)\, \bar{r}_k(u,v) + \\ &\quad [\bar{c}_m(u,v) - c(u,v)]\,, \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} m^{2\widetilde{\alpha}_1 - 3} - m^{2\alpha_1 - 3} &= O_P\left(|\widetilde{\alpha}_1 - \alpha_1| \log m\right) \cdot m^{2\alpha_1 - 3} \\ &= o_P\left(\log m / \log n\right) \cdot m^{2\alpha_1 - 3} \\ &= o_P(1) \cdot m^{2\alpha_1 - 3} \end{aligned} \tag{3.7}$$

by standard calculation and using $\widetilde{\alpha}_1 - \alpha_1 = o_P(1/\log n)$. By (3.5)–(3.7),

$$\left\|\widetilde{C}_m - C\right\|_{\mathcal{S}} = o_P(1),$$

where the operator $\widetilde{C}_m$ is defined similarly to $\overline{C}_m$ but with $\bar{c}_m(u,v)$ replaced by $\widetilde{c}_m(u,v)$.

We next estimate the orthonormal functions ψ_i, $i = 1, \cdots, p_1$, defined in (3.1). From Proposition 1, the p_1 largest eigenvalues $\lambda_1, \cdots, \lambda_{p_1}$ of $c(u, v)$ defined in (3.1) are positive, bounded away from zero and infinity and satisfy $\lambda_i = \lim_{n\to\infty} \lambda_{ni}/n^{3-2\alpha_1}$, and the sum of the remaining eigenvalues tends to zero. Furthermore, by (2.3), (3.1) and (3.2), we may show that $\psi_1, \cdots, \psi_{p_1}$ are the eigenfunctions of $c(u, v)$ corresponding to the p_1 largest eigenvalues. Hence, we can implement FPCA on $\widetilde{c}_m(u, v)$, a consistent estimate of $c(u, v)$, and estimate $\psi_1, \cdots, \psi_{p_1}$. But since $\widetilde{c}_m(u, v)$ is proportional to

$$\widehat{c}_m(u, v) = \sum_{|k| \leqslant m} (m - |k|)\, \bar{r}_k(u, v), \tag{3.8}$$

it is obvious that the eigenfunctions via FPCA of $\widetilde{c}_m(u, v)$ are the same as those of $\widehat{c}_m(u, v)$, and using (3.8) does not require estimating α_1. Hence, for practical applications we consider eigenanalysis on $\widehat{c}_m(u, v)$ and let $\widehat{\psi}_1, \cdots, \widehat{\psi}_{p_1}$ be the eigenfunctions of $\widehat{c}_m(u, v)$ corresponding to the p_1 largest eigenvalues. The following theorem shows that $\widehat{\psi}_i$ consistently estimates ψ_i and ψ_{ni} (up to sign change), $i = 1, \cdots, p_1$.

Theorem 2. *Suppose that the conditions in Proposition 2 are satisfied, p_1 is known, and $0 < \lambda_{p_1} < \cdots < \lambda_1 < \infty$. Then,*

$$\max_{1\leqslant i\leqslant p_1} \left\| \widehat{\psi}_i - \tau_i \psi_i \right\| = o_P(1) \tag{3.9}$$

and

$$\max_{1\leqslant i\leqslant p_1} \left\| \widehat{\psi}_i - \tau_{ni} \psi_{ni} \right\| = o_P(1) \tag{3.10}$$

if $\tau_i = \tau_{ni}$, where $\tau_i = \mathsf{sign}(\langle \widehat{\psi}_i, \psi_i \rangle)$ and $\tau_{ni} = \mathsf{sign}(\langle \widehat{\psi}_i, \psi_{ni} \rangle)$.

3.2 Estimation of α_1

From Proposition 1 and Theorem 1, the first score process $\{x_t^1 : t \in \mathbb{Z}\}$ defined in (2.4) is a long-range dependent linear process with memory parameter α_1. We estimate α_1 by the so-called R/S method introduced by Hurst (1951, 1956) to study the behaviour of the Nile and various reservoirs. A number of other memory parameter estimates have better statistical properties than the R/S estimate (which, for example, is clearly inefficient in the case of Gaussian innovations), but the latter has the advantage of making relatively quick and easy use of our asymptotic results. Define

$$\mathsf{R}_n = \max_{1\leqslant k\leqslant n} \sum_{t=1}^{k} \left(x_t^1 - \overline{x}^1\right) - \min_{1\leqslant k\leqslant n} \sum_{t=1}^{k} \left(x_t^1 - \overline{x}^1\right)$$

and

$$\mathsf{S}_n^* = \left[\frac{1}{n}\sum_{t=1}^{n} \left(x_t^1 - \overline{x}^1\right)^2\right]^{1/2}, \qquad \overline{x}^1 = \frac{1}{n}\sum_{t=1}^{n} x_t^1. \tag{3.11}$$

The R/S statistic may be defined by $\mathsf{R}_n/\mathsf{S}_n^*$. From the argument in Section 3.1, the eigenfunction ψ_{n1} can be consistently estimated by $\widehat{\psi}_1$, which suggests that we may approximate x_t^1 by

$$\widehat{x}_t^1 = \int_{\mathcal{C}} X_t(u)\widehat{\psi}_1(u)du.$$

Define $\widehat{\mathsf{R}}_n$ and $\widehat{\mathsf{S}}_n^*$ like R_n and S_n^* but with x_t^1 replaced by $\widehat{x}_t^1$. Then a feasible R/S statistic is $\widehat{\mathsf{R}}_n/\widehat{\mathsf{S}}_n^*$ with the following asymptotic distribution.

Proposition 3. *Suppose that the conditions in Theorems 1 and 2 are satisfied. Then we have*

$$\frac{1}{n^{H_1}}\frac{\widehat{\mathsf{R}}_n}{\widehat{\mathsf{S}}_n^*} \xrightarrow{P} \mathsf{V}, \tag{3.12}$$

where $H_1 = 3/2 - \alpha_1$*,*

$$\mathsf{V} = \left\{\mathsf{E}\left[(x_t^1)^2\right]\right\}^{-1/2}\theta_1\left\{\sup_{0\leqslant r\leqslant 1}\left[\mathsf{B}_{H_1}(r) - r\mathsf{B}_{H_1}(1)\right] - \inf_{0\leqslant r\leqslant 1}\left[\mathsf{B}_{H_1}(r) - r\mathsf{B}_{H_1}(1)\right]\right\},$$

θ_1 *is defined in Proposition 1, and* $\mathsf{B}_H(\cdot)$ *is a fractional Brownian motion with index* H.

The convergence result (3.12) motivates estimating α_1 by

$$\widehat{\alpha}_1 = 3/2 - \widehat{H}_1, \qquad \widehat{H}_1 = \log\left(\widehat{\mathsf{R}}_n/\widehat{\mathsf{S}}_n^*\right)/\log n.$$

Using Proposition 3, we may show that

$$\widehat{\alpha}_1 - \alpha_1 = O_P\left(\log^{-1} n\right). \tag{3.13}$$

Although $\widehat{\alpha}_1$ is consistent, the convergence rate in (3.13) is so slow that we cannot even achieve the consistency of $\widetilde{C}_m$ described in Remark 7 if $\widehat{\alpha}_1$ is used to construct $\widetilde{c}_m(u,v)$, but, as discussed in Section 3.1, we do not require an estimate of α_1 when conducting eigenanalysis of $\widehat{c}_m(u, v)$. We conjecture that a faster rate of convergence can be obtained if the local Whittle method is used to estimate α_1 (c.f., Robinson 1995), which will be left for future research. Whereas the classical R/S method focuses on estimating the Hurst coefficient by $\log(\widehat{R}_n/\widehat{S}_n^*)/\log n$, Lo (1991) considers a hypothesis testing procedure to detect long memory by a modified statistic, where S_n^* in (3.11) is replaced by the square root of a consistent and positive estimate of the long-run variance. Our R/S estimate uses only the first score process, and when the estimated dimension $\widehat{p}_1$ of $\mathcal{S}_1$ (see Section 3.3 below) exceeds 1, it seems natural to normalise components in the multivariate score process $\left(\widehat{x}_t^i,\ i = 1, \cdots, \widehat{p}_1\right)$, where $\widehat{x}_t^i$, $i = 1, \cdots, \widehat{p}_1$, are the approximate score processes defined analogously to $\widehat{x}_t^1$. Specifically, we construct

$$\widehat{x}_t^N = \sqrt{\sum_{i=1}^{\widehat{p}_1} \left(\widehat{x}_t^i\right)^2}, \qquad t = 1, \cdots, n,$$

and then apply the R/S method to the normalised score process $\{\widehat{x}_t^N : t \in \mathcal{Z}\}$. More details are given Appendix B.1 in the supplemental document.

3.3 Estimation of p_1

Another critical issue in practical implementation is to determine p_1, the dimension of the dominant sub-space $\mathcal{S}_1$. One commonly-used method is to select the first few eigenfunctions (corresponding to the first few largest eigenvalues) of $\widehat{c}_m(\cdot, \cdot)$ so that a pre-determined amount, say 85%, of the total variation is accounted for. Horváth & Kokoszka (2012) call this the CPV (cumulative percentage of total variance) method, but it can be difficult to establish its consistency. Instead, we use a simple ratio method introduced by Lam & Yao (2012) to determine p_1. Specifically, letting $\widehat{\lambda}_{m,i}$ be the i^{th} largest eigenvalue of $\widehat{c}_m(\cdot, \cdot)$, we estimate p_1 by

$$\widehat{p}_1 = \arg \min_{1 \leqslant i \leqslant \bar{P}} \frac{\widehat{\lambda}_{m,i+1}}{\widehat{\lambda}_{m,i}}, \tag{3.14}$$

where $\bar{P}$ is a pre-specified positive integer and $0/0 = 1$. To reduce estimation error in practical implementation, we set $\widehat{\lambda}_{m,i}/\widehat{\lambda}_{m,1}$ as 0 if its absolute value is smaller than $\epsilon_\star$, a pre-specified small positive number, so

$$\frac{\widehat{\lambda}_{m,i+1}}{\widehat{\lambda}_{m,i}} = \frac{\widehat{\lambda}_{m,i+1}/\widehat{\lambda}_{m,1}}{\widehat{\lambda}_{m,i}/\widehat{\lambda}_{m,1}} = 0/0 = 1, \tag{3.15}$$

if both $\left|\widehat{\lambda}_{m,i}/\widehat{\lambda}_{m,1}\right|$ and $\left|\widehat{\lambda}_{m,i+1}/\widehat{\lambda}_{m,1}\right|$ are smaller than $\epsilon_\star$. The following proposition shows the consistency of $\widehat{p}_1$ defined in (3.14).

Proposition 4. *Suppose that the conditions in Proposition 2 are satisfied. Then we have* $\widehat{p}_1 \xrightarrow{P} p_1$.

The proof of the above proposition is given in the supplemental document. The ratio method will be used in our numerical studies to estimate the dimension of $\mathcal{S}_1$. Other dimension selection methods proposed in the literature such as cross-validation (Hall & Hosseini-Nasab 2006), bootstrap (Hall & Vial 2006), minimum description length (Poskitt & Sengarapillai 2013), and Akaike information criterion (Li et al. 2013) may also be applicable to estimate p_1.

4 The functional FARIMA model

In this section, we study the functional FARIMA(p, d, q) process defined by

$$\triangledown^d X_t(u) = Y_t(u), \quad \triangledown = 1 - B, \quad -1/2 < d < 1/2, \tag{4.1}$$

and

$$Y_t(u) - \sum_{i=1}^{p} \int_{\mathcal{C}} \phi_i\,(u,v)\, Y_{t-i}(v) dv = \eta_t(u) + \sum_{i=1}^{q} \int_{\mathcal{C}} \varphi_i\,(u,v)\, \eta_{t-i}(v) dv, \tag{4.2}$$

where B denotes the backshift operator, $\{\eta_t : t \in \mathcal{Z}\}$ satisfies Assumption 1 in Section 2.2, and $\phi_i(u,v)$ and $\varphi_i(u,v)$ are the kernels with associated integral operators defined by $\int_{\mathcal{C}} \phi_i(u,v)x(v)dv$ and $\int_{\mathcal{C}} \varphi_i(u,v)x(v)dv$, respectively, $x \in \mathcal{H}$, and such that $Y_t(u)$ is stationary with respect to t. When $d = 0$, (4.1) becomes the functional ARMA(p, q) of Klepsch et al. (2017), while when $q = 0$, it further reduces to the functional AR(p) model of Bosq (2000) and Liu et al. (2016), and when $p = 0$, it reduces to the functional MA(q) model of Chen et al. (2016) and Aue & Klepsch (2017). We stress that d is fixed over the set $\mathcal{C}$ and since the main interest of the present paper is long-range dependence, we only consider $d \in (0, 1/2)$. We will show that, under mild conditions, Assumption 2 and Proposition 1 in Section 2 hold for the functional FARIMA, see Proposition 5 and Remark 8 below. This model will be employed in Section 5 to generate curve time series in simulation, where finite-sample performance of the methodology proposed in Section 3 will be examined. However, we do not give full consideration to the FARIMA model to conserve on space.

For notational simplicity, let $\phi_i = \phi_i(\cdot, \cdot)$ and $\varphi_i = \varphi_i(\cdot, \cdot)$. As in Bosq (2000) and Klepsch et al. (2017), for the functional ARMA(p, q) process $\{Y_t\}$ with $Y_t = (Y_t(u) : u \in \mathcal{C})$, we can write

$$\overline{\mathbf{Y}}_t(u) = \int_{\mathcal{C}} \overline{\boldsymbol{\phi}}(u, v)\overline{\mathbf{Y}}_{t-1}(v)dv + \sum_{i=0}^{q} \int_{\mathcal{C}} \overline{\boldsymbol{\varphi}}_i(u, v)\overline{\boldsymbol{\eta}}_{t-i}(v)dv, \tag{4.3}$$

where

$$\begin{aligned}
\overline{\mathbf{Y}}_t(u) &= [Y_t(u), \cdots, Y_{t-p+1}(u)]^\tau, \qquad \overline{\boldsymbol{\eta}}_t(u) = [\eta_t(u), 0, \cdots, 0]^\tau, \\
\overline{\boldsymbol{\phi}}(\cdot, \cdot) &= \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ I & O & \cdots & O \\ \vdots & \vdots & \vdots & \vdots \\ O & \cdots & I & O \end{bmatrix}, \quad \overline{\boldsymbol{\varphi}}_i(\cdot, \cdot) = \begin{bmatrix} \varphi_i & O & \cdots & O \\ O & O & \cdots & O \\ \vdots & \vdots & \vdots & \vdots \\ O & O & \cdots & O \end{bmatrix},
\end{aligned}$$

$\varphi_0 = I$, I and O denote the identity and zero operators, respectively, and "τ" denotes transposition. Let $\mathcal{H}^p$ be the cartesian product of p copies of $\mathcal{H}$ (c.f., Chapter 5 of Bosq 2000). The following assumption ensures existence of a unique stationary and causal solution to (4.3).

Assumption 3. *Let $\varphi_1, \cdots, \varphi_q$ be the Hilbert-Schmidt kernels in the sense that*

$$\int_{\mathcal{C}} \int_{\mathcal{C}} \varphi_i^2(u, v)dudv < \infty, \;\; i = 1, \cdots, q,$$

and let $\phi_1, \cdots, \phi_q$ satisfy

$$\sum_{i=1}^{p} \left(\int_{\mathcal{C}} \int_{\mathcal{C}} \phi_i^2(u, v)dudv \right)^{1/2} < 1.$$

Following Theorem 3.8 of Klepsch et al. (2017), by (4.3) and Assumption 3 above, we readily have

$$\overline{\mathbf{Y}}_t = \sum_{i=0}^{\infty} \mathbf{A}_i(\overline{\boldsymbol{\eta}}_{t-i}), \tag{4.4}$$

where $\overline{\mathbf{Y}}_t = \left(\overline{\mathbf{Y}}_t(u) : u \in \mathcal{C}\right)$, $\overline{\boldsymbol{\eta}}_t = (\overline{\boldsymbol{\eta}}_t(u) : u \in \mathcal{C})$, $\mathbf{A}_i$ is the integral operator on $\mathcal{H}^p$ with kernel

$$\boldsymbol{\alpha}_i(u,v) = \begin{cases} \sum_{k=0}^{i} \int_{\mathcal{C}} \overline{\boldsymbol{\phi}}^{i-k}(u,u_1)\overline{\boldsymbol{\varphi}}_k(u_1,v)du_1, & 0 \leqslant i \leqslant q-1, \\ \sum_{k=0}^{q} \int_{\mathcal{C}} \overline{\boldsymbol{\phi}}^{i-k}(u,u_1)\overline{\boldsymbol{\varphi}}_k(u_1,v)du_1, & i \geqslant q, \end{cases}$$

and $\overline{\boldsymbol{\phi}}^0 = \mathrm{I}$. Letting $\pi(u_1, \cdots, u_p) = u_1$, we obtain the following MA(∞) representation:

$$Y_t = \sum_{i=0}^{\infty} \pi\left[\mathbf{A}_i(\overline{\boldsymbol{\eta}}_{t-i})\right]. \tag{4.5}$$

Note that

$$X_t = \nabla^{-d} Y_t = (1-B)^{-d}\, Y_t = \sum_{i=0}^{\infty} \beta_{i,-d} B^i Y_t = \sum_{i=0}^{\infty} \beta_{i,-d} Y_{t-i}, \tag{4.6}$$

where, by Stirling's formula,

$$\beta_{i,-d} = \beta_i^* + \beta_i^\diamond, \quad \beta_i^* = \frac{1}{\Gamma(d)} i^{-1+d}, \quad \beta_i^\diamond = O\left(i^{-2+d}\right), \tag{4.7}$$

and $\Gamma(\cdot)$ is the gamma function. A combination of (4.5) and (4.6) leads to

$$X_t = \sum_{i=0}^{\infty} \beta_{i,-d} \sum_{j=0}^{\infty} \pi\left[\mathbf{A}_j(\overline{\boldsymbol{\eta}}_{t-i-j})\right]. \tag{4.8}$$

Define the operator $\mathbf{A}_\infty$ by

$$\mathbf{A}_\infty(x)(u) = \sum_{j=0}^{\infty} \mathbf{A}_j(x)(u) = \sum_{j=0}^{\infty} \int_{\mathcal{C}} \boldsymbol{\alpha}_j(u,v)x(v)dv, \quad x \in \mathcal{H},\ u \in \mathcal{C}. \tag{4.9}$$

Assumption 4. *Let*

$$c^\diamond(u,v) = \mathsf{E}\left\{\pi\left[\mathbf{A}_\infty(\overline{\boldsymbol{\eta}}_0)(u)\right]\pi\left[\mathbf{A}_\infty(\overline{\boldsymbol{\eta}}_0)(v)\right]\right\}, \quad u,v \in \mathcal{C}.$$

There exist $0 < \lambda_{p_\diamond}^\diamond < \cdots < \lambda_1^\diamond < \infty$ *and orthonormal functions* $\psi_1^\diamond(\cdot), \cdots, \psi_{p_\diamond}^\diamond(\cdot)$ *such that*

$$\lambda_k^\diamond \psi_k^\diamond(u) = \int_{\mathcal{C}} c^\diamond(u,v)\psi_k^\diamond(v)dv, \quad k = 1, \cdots, p_\diamond, \tag{4.10}$$

and $\lambda_k^\diamond = 0$ *when* $k \geqslant p_\diamond + 1$. *The positive integer* $p_\diamond$ *is fixed.*

Let

$$\overline{X}_t = \sum_{i=0}^{\infty} \beta_{i,-d} \sum_{j=0}^{\infty} \pi\left[\mathbf{A}_j(\overline{\boldsymbol{\eta}}_{t-i})\right] = \sum_{i=0}^{\infty} \beta_{i,-d} \cdot \pi\left[\mathbf{A}_\infty(\overline{\boldsymbol{\eta}}_{t-i})\right], \tag{4.11}$$

and $\widetilde{X}_t = X_t - \overline{X}_t$. The following proposition describes the functional dependence structure of $\overline{X}_t$ and $\widetilde{X}_t$, respectively.

Proposition 5. *Suppose that Assumptions 1, 3 and 4 are satisfied.*

(i) The curve process $\{\overline{X}_t\}$ is long-range dependent, and Proposition 1 in Section 2.3 holds with $p_1 = p_\diamond$, $\kappa_0 = 1$ and

$$\lambda_{nj}^\diamond \sim \frac{c_{1-d}\lambda_j^\diamond}{d(1+2d)\Gamma^2(d)} n^{2d+1}, \quad j = 1, \cdots, p_\diamond, \tag{4.12}$$

where $\lambda_j^\diamond$ and $p_\diamond$ are defined in Assumption 4, c_{1-d} is defined as in Proposition 1, and $\lambda_{nj}^\diamond$ is the j^{th} largest eigenvalue of

$$c_n^\diamond(u,v) = \mathsf{E}\left[\sum_{t=1}^{n}\sum_{s=1}^{n} \overline{X}_t(u)\overline{X}_s(v)\right], \quad u,v \in \mathcal{C}.$$

(ii) The curve process $\{\widetilde{X}_t\}$ is stationary and short-range dependent.

Remark 8. Let $\mathcal{S}_\diamond$ be a $p_\diamond$-dimensional sub-space spanned by the orthonormal eigenfunctions $\psi_1^\diamond, \cdots, \psi_{p_\diamond}^\diamond$ defined in (4.10). The proof of Proposition 5(i) shows that the projection of the functional FARIMA(p, d, q) process $\{X_t\}$ onto the space $\mathcal{S}_\diamond$ is long-range dependent. In addition, from Proposition 5(ii), we may show that $\mathcal{S}_\diamond$ is the dominant sub-space. Furthermore, Corollary 1 holds for $\{X_t\}$. The estimation methodology developed in Section 3 can also be used to estimate $\psi_1^\diamond, \cdots, \psi_{p_\diamond}^\diamond$ and the parameter $d = 1 - \alpha_1$, and determine the dimension $p_\diamond$, see the simulation study in Section 5.1 below.

Remark 9. In this section, we limit the discussion to the case that d is fixed over $\mathcal{C}$, ensuring that the theory in Section 2 and the methodology in Section 3 are applicable. This assumption is restrictive, but, motivated by Characiejus & Rauckauskas (2014) and Düker (2018), using the multiplication operator, it may be possible to extend (4.1) to

$$\nabla^{d(u)} X_t(u) = Y_t(u), \quad 0 < d(u) < 1/2,\ u \in \mathcal{C}. \tag{4.13}$$

A different technique would be needed to derive the relevant limit theory for (4.13), while the methodology proposed in Section 3 would need to be substantially generalised. We will leave this to future study.

5 Numerical studies

This section provides Monte-Carlo studies of finite-sample performance and two empirical applications. In fact the data employed here, in common with many other data to be found in practice, do not consist of continuous records over a finite interval, but rather ones over discrete grids. We thus adopt the linear interpolation algorithm of Hyndman et al. (2018) in the R software (R Core Team 2018) to convert discrete data points into a continuous function before implementing the developed methodology.

5.1 Simulation study

We generate X_t by the functional FARIMA(p, d, q) model in (4.1) and (4.2), with $\mathcal{C} = [0, 1]$ and $\{\eta_t : t \in \mathcal{Z}\}$ a sequence of *i.i.d.* standard Brownian motions over $[0, 1]$, in the following two cases:

Case 1: $p = 1, d = 0.2, q = 0, \phi_1(u, v) = 0.34 \times \exp\{-(u^2 + v^2)/2\}$,

Case 2: $p = 1, d = 0.2, q = 1, \phi_1(u, v) = 0.34 \times \exp\{-(u^2 + v^2)/2\}, \varphi_1(u, v) = \frac{3}{2}\min(u, v)$.

The choice of constants in ϕ_1 and φ_1 ensures that both $\|\phi_1\|$ and $\|\varphi_1\|$ are smaller than one (c.f., Rice & Shang 2017, Kokoszka et al. 2017), so the simulated curve time series are stationary and invertible. The sample sizes are $n = 500, 1000$ and 2000, with 1000 replications.

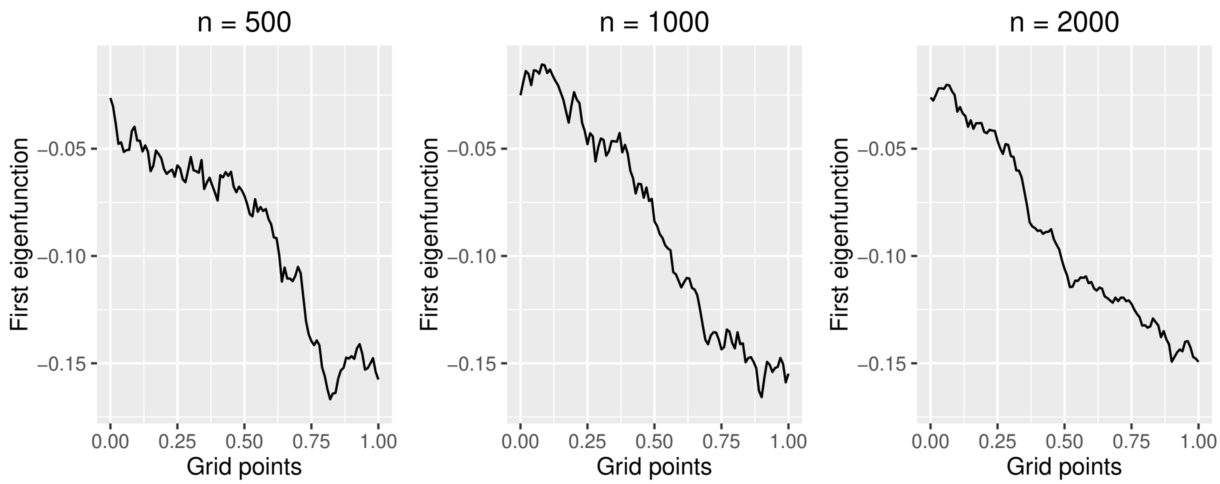
The number p_1 of orthonormal functions in the dominant sub-space $\mathcal{S}_1$ is estimated by either the CPV method to explain at least 85% of total variation or the ratio method defined in (3.14) with $\bar{P} = \lfloor U/2 \rfloor$, where $U = 101$ is the number of equi-spaced discretised points (in simulating the curve samples). As shown in Table 1, for the FARIMA$(1, 0.2, 0)$, over 92% of replications estimate p_1 by $\widehat{p}_1 = 1$ using the ratio method; and for the FARIMA(1, 0.2, 1), over 97% of them estimate it as 1. The CPV method overestimates p_1 more often than the ratio method, especially for the FARIMA$(1, 0.2, 0)$.

Table 1: The numbers of replications producing different estimates $\widehat{p}_1$ of $p_1 = 1$ based on the ratio method and the CPV method.

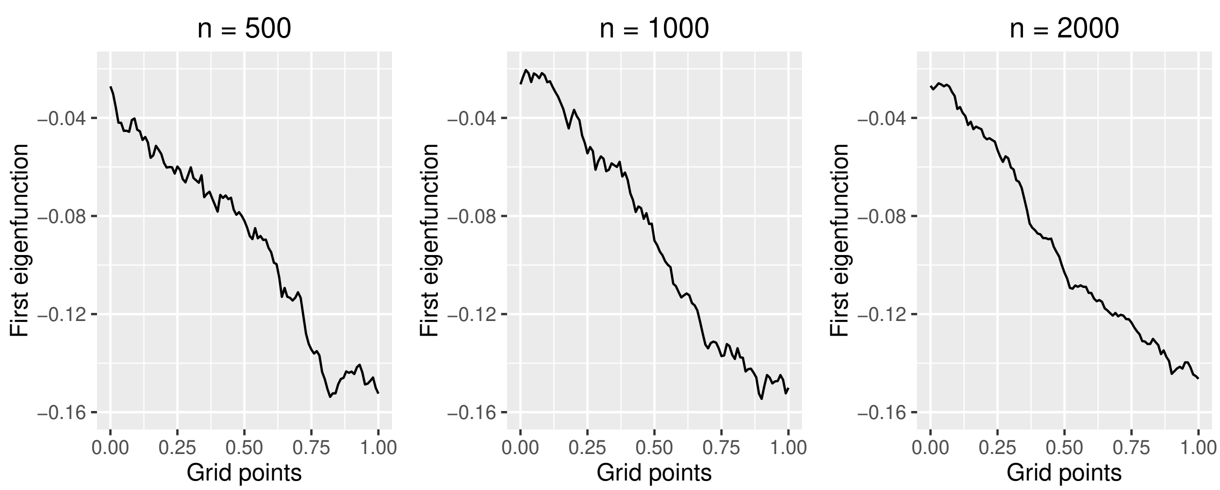
	Functional FARIMA$(1, 0.2, 0)$						Functional FARIMA$(1, 0.2, 1)$					
	Ratio method			CPV method			Ratio method			CPV method		
$\widehat{p}_1$	500	1000	2000	500	1000	2000	500	1000	2000	500	1000	2000
1	920	973	995	784	851	893	978	997	1000	975	995	1000
2	65	27	5	208	148	107	20	3		25	5	
3	7			8	1		2					
4	5											
6	2											
7	1											

Figures 3a and 3b plot the estimated eigenfunction corresponding to the largest eigenvalue after implementing FPCA on $\widehat{c}_m(u, v)$ defined in (3.5) for the FARIMA(1, 0.2, 0) and FARIMA(1, 0.2, 1), respectively, where, for simplicity, we select $m = 101$ that is the same as U, the number of discretised points in simulating curves, and smaller than the sample size (while we have no reason for arguing that U should influence the value of m). The estimated eigenfunction exhibits similar shape among all three sample sizes, and the estimated eigenfunction for the FARIMA(1, 0.2, 1) differs slightly in shape from that for the FARIMA(1, 0.2, 0) due to the moving average component.

For each of the 1000 replications, we estimate $d = 1 - \alpha_1$ using the R/S method and the results are presented using boxplots in Figure 4, from which we find that R/S performs reasonably well, with accuracy improving as n increases. As the estimated number of orthonormal functions spanning $\mathcal{S}_1$ can be more than one (in particular when n is either 500 or 1000), in the online supplement, we also consider normalising a set of principal



(a) FARIMA$(1, 0.2, 0)$ process



(b) FARIMA$(1, 0.2, 1)$ process

Figure 3: The first empirical eigenfunction for $n = 500, 1000$ and 2000.

component scores (when $\widehat{p}_1 > 1$) and then apply the R/S method to the normalised score process.

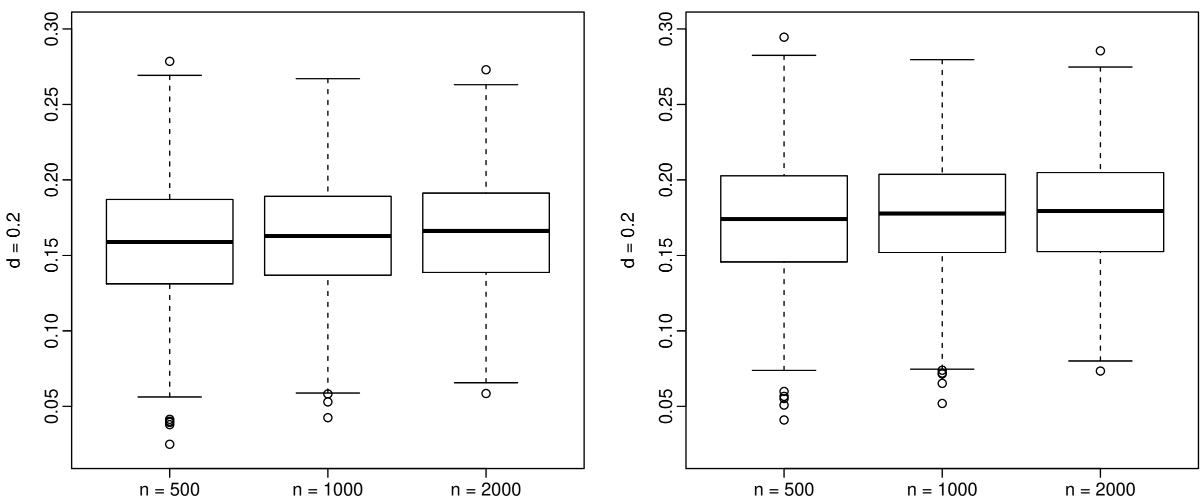


Figure 4: Boxplots for R/S estimates of d with the left one for FARIMA(1, 0.2, 0) and the right one for FARIMA(1, 0.2, 1).

5.2 Application to US stocks

We now apply our methodology to intraday log-returns of the six US stocks listed in Table 2. We select a time period of one year from 2 January 2014 to 31 December 2014, containing 249 trading days after removing a half trading day on 24 December 2014. Let $SP_t(u)$ denote the price of a stock on day t at time u. Following Aue et al. (2017) and Kokoszka & Reimherr (2017), we introduce the intraday log-returns:

$$X_t(u) = \ln SP_t(u) - \ln SP_t(u-h), \tag{5.1}$$

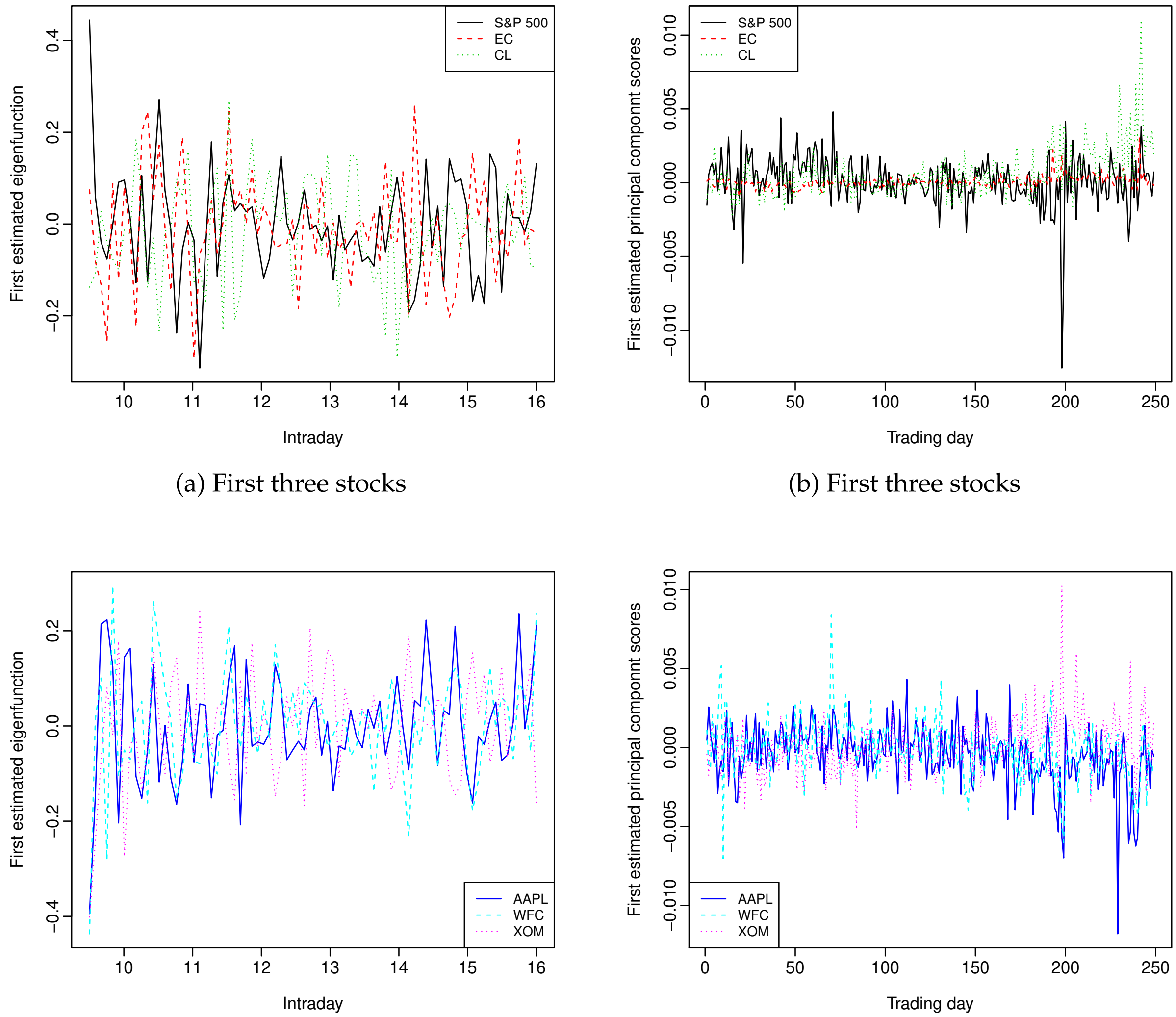
where h is a time window typically chosen as 1, 5 or 15 minutes. We take $h = 5$, to avoid microstructure noise in 1-minute log returns; see also Aue et al. (2017) who propose a functional GARCH model to fit such data. In Section 1, Figure 1 plots these intraday log-return curves observed in the last quarter of 2014. Since all stocks trade from

9:30am to 4:00pm, there are 78 measurements per day and $n = 249$ curves denoted by $\{X_1, X_2, \cdots, X_{249}\}$ with $X_t = (X_t(u) : u \in \mathcal{C})$, where $\mathcal{C}$ is the time interval between 9:30am and 4:00pm. The high-dimensionality of the intraday observations makes application of univariate/multivariate long-range dependence procedures impractical. With the linear interpolation algorithm of Hyndman et al. (2018), we not only convert discrete data points into a continuous function but also fill in missing values for the WFC and XOM stocks, ensuring each daily curve is of the same length.

Table 2: Sectors and stocks used in this study.

Sector	Symbol	Full Name
Index	S&P 500	Standard & Poor 500 Index
Currency Exchange	EC	Euro to Dollar
Futures	CL	Crude Oil (WTI Sweet Light) Futures
Technology	AAPL	Apple Inc
Financials	WFC	Wells Fargo & Company
Energy	XOM	Exxon Mobile Corporation

Using FPCA, the estimated eigenfunction corresponding to the maximum eigenvalue is displayed in Figures 5a and 5c, with their corresponding principal component scores in Figures 5b and 5d. When estimating the long-run covariance function, we choose the value of m as 78, the same as the number of discrete points in a trading day. We estimate α_1 using R/S and the estimated first principal component scores, see Table 3, from which we conclude that these six curve time series show a persistent pattern and do not change rapidly over time. Furthermore, we estimate α_1 for the first and second halves of the sample (with the first half containing 124 curves and the second half 125) to examine whether a structural break might cause the long memory; the results, also presented in Table 3, suggest it is not. In addition, to quantify estimation uncertainty, in the supplement we implement a bootstrap method introduced by Shang (2018) to construct confidence intervals for the R/S estimates. Table 4 reports $\widehat{p}_1$, based on the CPV and ratio methods. As above, the ratio method tends to select smaller $\widehat{p}_1$ than CPV.



(a) First three stocks (b) First three stocks

(c) Last three stocks (d) Last three stocks

Figure 5: The first estimated eigenfunctions and principal component scores for the intraday log-return curves of the six stocks.

5.3 Application to age-specific fertility rates

Annual Australian fertility rates from 1921 to 2006 for each age from 15 to 49 are obtained from the Australian Bureau of Statistics (Cat.No.3105.0.65.001, Table 38). There are $n = 86$

Table 3: Estimates of α_1 for the intraday log-return curves of the six stocks.

Sample period	S&P 500	EC	CL	AAPL	WFC	XOM
All 249 curves	0.8761	0.7872	0.7525	0.7985	0.8005	0.8117
First 124 curves	0.8390	0.7870	0.7909	0.8197	0.7935	0.7915
Last 125 curves	0.8179	0.7791	0.7535	0.8033	0.8450	0.8272

Table 4: Estimation of the dimension of $\mathcal{S}_1$ for the intraday log-return curves.

Method	S&P 500	EC	CL	AAPL	WFC	XOM
CPV	4	5	4	4	4	4
Ratio	2	1	2	3	4	4

curves from 1921 to 2006 denoted by $\{X_1, X_2, \cdots, X_{86}\}$, where the function support $\mathcal{C}$ for age lies between 15 and 49. Fertility rates are defined as the number of live births during the calendar year, according to the age of the mother, per 1000 of the female resident population of the same age on 30 June, and have changed slowly over time, as shown in Figure 2, reflecting the changing social conditions affecting fertility. For example, there is an increase in fertility in all age groups around the end of World War II (1945), a rapid increase during the 1960s corresponding to the baby boom period (McDonald 2000), and an increase at higher ages in more recent years caused by delay in child-bearing. Figure 2 suggests that the functional time series are non-stationary, so to generate series which we hope are close to stationarity, we implement the transformation (Haberman & Renshaw 2012)

$$\overline{X}_t(u_i) = 2 \cdot \frac{1 - X_t(u_i)/X_{t-1}(u_i)}{1 + X_t(u_i)/X_{t-1}(u_i)} = 2 \cdot \frac{X_{t-1}(u_i) - X_t(u_i)}{X_{t-1}(u_i) + X_t(u_i)}, \qquad i = 1, \cdots, I, \tag{5.2}$$

where $X_t(u_i)$ denotes fertility rate for age u_i in year t, and I denotes the number of discrete ages. The denominator in (5.2) attempts to avoid the small phase difference between the numerator and denominator. The $\overline{X}_t(u_i)$ are called incremental fertility rates. In Figure 6, we present these rates in Australia from 1922 to 2006.

Using FPCA for the transformed curve time series $\{\overline{X}_1, \cdots, \overline{X}_{86}\}$, the estimated eigenfunction, corresponding to the largest eigenvalue, along with its corresponding principal component scores are given in Figure 7. The value of m in the long-run covariance function estimation is chosen as 29. This estimated eigenfunction seems to capture the so-called fertility postponement happening in many developed countries in recent years.

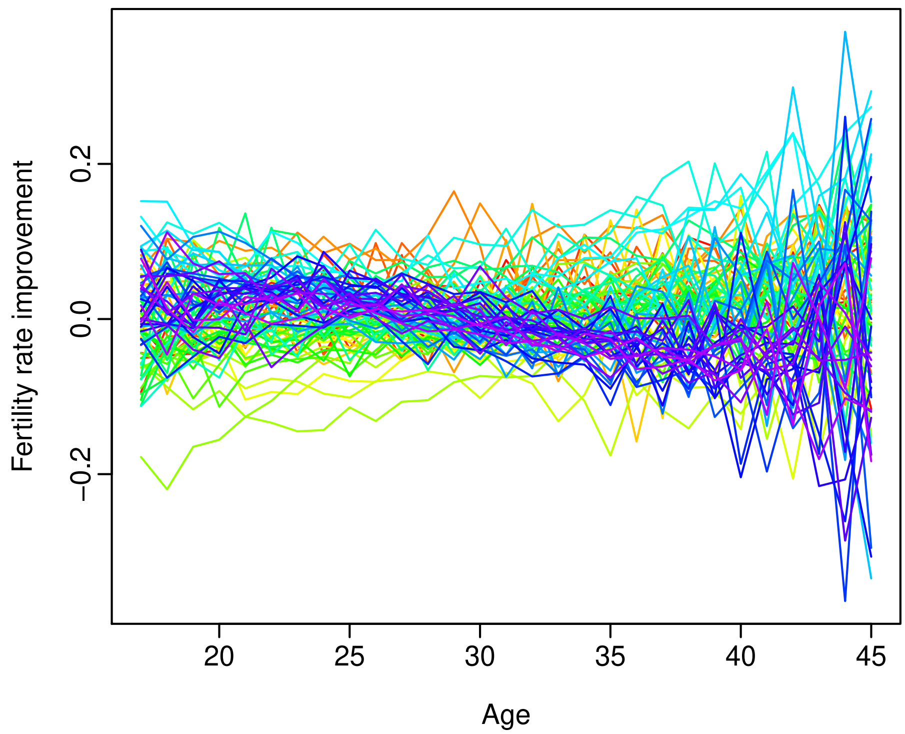


Figure 6: Australian incremental fertility rates from 1922 to 2006.

In addition to Australia, we further consider 13 other developed countries for which data are available in Human Fertility Database (2016). These are all developed countries with relatively long data series, see Table 5. The transformation defined in (5.2) is applied to these age-specific fertility rates, and a stationarity test proposed by Horváth et al. (2014) is used to examine whether or not the transformed functional time series is stationary. Under the null hypothesis of stationarity, we present the p-values in Table 5, from which we cannot reject the null hypothesis at the 10% level of significance.

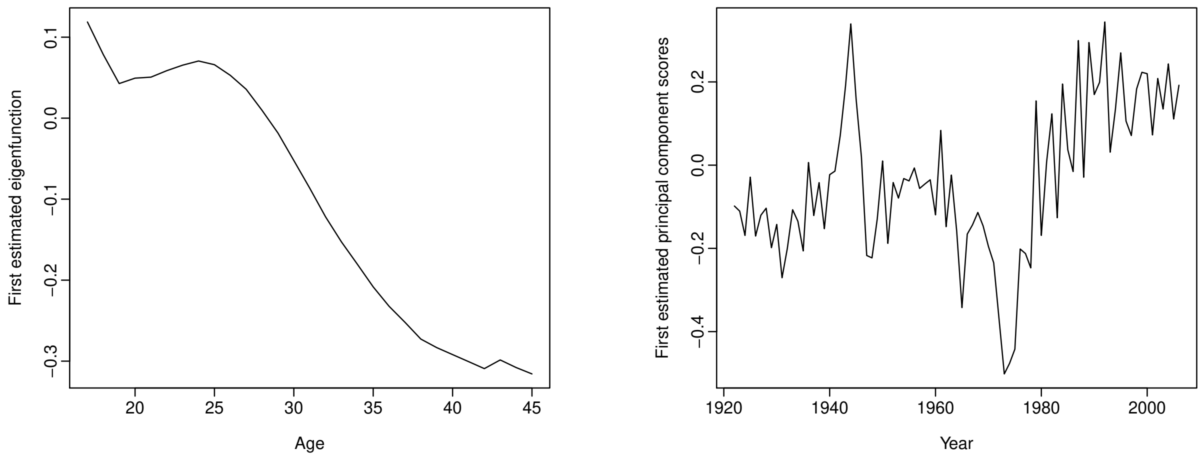


Figure 7: The first estimated eigenfunctions and principal component scores for the Australian incremental fertility rates.

Table 5: The p-values of the stationarity test for various developed countries.

Country	Data period	p-value	Country	Data period	p-value
Australia	1921:2006	0.241	Canada	1921:2011	0.365
Denmark	1916:2016	0.112	Finland	1939:2012	0.218
France	1946:2015	0.119	Germany	1956:2013	0.108
Iceland	1960:2012	0.108	Italy	1954:2014	0.140
Netherland	1950:2016	0.151	Spain	1922:2014	0.521
Sweden	1891:2016	0.113	Switzerland	1932:2014	0.137
UK	1974:2016	0.165	USA	1933:2014	0.238

Using the estimated first principal component scores, through implementing FPCA, we obtain R/S estimates of α_1 (see Table 6), which together with the results in Table 5, indicate that the age-specific incremental fertility rates for the selected developed countries exhibit stationary and long-range dependence structure, with all but one of the estimated values less than 0.85. These results not only confirm the features shown in the functional time series plots but also allow us to quantify the strength of long-range dependence. In the supplemental document, we report the confidence intervals of the R/S estimates and the estimated dimension of $\mathcal{S}_1$ using the CPV and ratio methods.

Table 6: Estimates of α_1 for the age-specific incremental fertility rates from the 14 developed countries available in the Human Fertility Database (2016).

Country	R/S estimates	Country	R/S estimates
Australia	0.7676	Canada	0.7667
Denmark	0.7305	Finland	0.7973
France	0.7335	Germany	0.7297
Iceland	0.9081	Italy	0.7459
Netherland	0.7292	Spain	0.8432
Sweden	0.7192	Switzerland	0.7367
UK	0.7927	USA	0.7636

6 Conclusion

This paper has introduced a functional or curve linear process with long-range dependence and derived some relevant asymptotic theorems. The functional dependence structure is specified via the projections of the curve process onto different sub-spaces spanned by additive orthonormal functions. Under regularity conditions, we have established a central limit theorem. In particular, we have shown that the projection of the curve linear process onto the (asymptotically) dominant sub-space contains most of the sample information carried by the original curve process. The orthonormal functions that span the dominant sub-space are estimated via classical FPCA on the estimated long-run covariance function (up to multiplication by a rate). We also provide easy-to-implement and asymptotically justified methods to estimate the memory parameter and dimension of the dominant sub-space, respectively. These methodologies are illustrated via simulations and empirical applications to US stock prices and age-specific fertility rates, both of which appear to be long-range dependent over time.

The paper might be extended in several directions. First, hypothesis testing based on the modified R/S statistic in Lo (1991) can be developed to detect the presence of long memory in curve time series. Second, we might attempt to extend stationarity to nonstationarity (e.g., the functional FARIMA(p, d, q) with $d > 1/2$). Other possible extensions would allow the memory parameter to vary over $\mathcal{C}$ and a functional version of fractional cointegration. Some of these are currently pursued in a separate project.

SUPPLEMENTARY MATERIAL

The supplemental document contains the detailed proofs of the theoretical results and gives additional numerical results.

Acknowledgements

The authors would like to thank an Associate Editor and a reviewer for their insightful comments and suggestions, which substantially improved the paper. They also thank Professor Hong Miao for providing the financial data set.

References

Abadir, K. M., Distaso, W. & Giraitis, L. (2009), 'Two estimators of the long-run variance: Beyond short memory', *Journal of Econometrics* **150**(1), 56–70.

Adenstedt, R. (1974), 'On large-sample estimation of the mean of a stationary random sequence', *The Annals of Statistics* **2**(6), 1095–1107.

Aue, A., Horváth, L. & Pellatt, D. F. (2017), 'Functional generalized autoregressive conditional heteroskedasticity', *Journal of Time Series Analysis* **38**(1), 3–21.

Aue, A. & Klepsch, J. (2017), Estimating functional time series by moving average model fitting, Technical report, University of California, Davies.
URL: https://arxiv.org/abs/1701.00770

Bathia, N., Yao, Q. & Ziegelmann, F. (2010), 'Identifying the finite dimensionality of curve time series', *The Annals of Statistics* **38**(6), 3352–3386.

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

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

Berkes, I., Horváth, L. & Rice, G. (2013), 'Weak invariance principles for sums of dependent random functions', *Stochastic Processes and Their Applications* **123**(2), 385–403.

Bingham, N., Goldie, C. & Teugels, J. (1987), *Regular Variation*, Cambridge University Press, Cambridge.

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

Bosq, D. & Blanke, D. (2007), *Inference and Prediction in Large Dimensions*, Dunod and John Wiley & Sons, Chichester.

Casas, I. & Gao, J. (2008), 'Econometric estimation in long-range dependent volatility models', *Journal of Econometrics* **147**(1), 72–83.

Characiejus, V. & Rauckauskas, A. (2014), 'Operator self-similar processes and functional central limit theorems', *Stochastic Processes and Their Applications* **124**(8), 2605–2627.

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

Chen, Y. & Li, B. (2017), 'An adaptive functional autoregressive forecast model to predict electricity prince curves', *Journal of Business and Economic Statistics* **35**(3), 371–388.

Chiou, J.-M. & Müller, H.-G. (2009), 'Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting', *Journal of the American Statistical Association* **104**(486), 572–585.

Davydov, Y. A. (1970), 'The invariance principle for stationary processes', *Theory of Probability and Its Applications* **15**(3), 487–498.

Düker, M. (2018), 'Limit theorems for Hilbert space-valued linear processes under long range dependence', *Stochastic Processes and Their Applications* **128**(5), 1439–1465.

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

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

Granger, C. W. J. & Joyeux, R. (1980), 'An introduction to long-memory time series models and fractional differencing', *Journal of Time Series Analysis* **1**(1), 15–29.

Haberman, S. & Renshaw, A. (2012), 'Parametric mortality improvement rate modelling and projecting', *Insurance: Mathematics and Economics* **50**(3), 309–333.

Hall, P. & Hosseini-Nasab, M. (2006), 'On properties of functional principal components analysis', *Journal of the Royal Statistical Society Series B* **68**(1), 109–126.

Hall, P. & Vial, C. (2006), 'Assessing the finite dimensionality of functional data', *Journal of the Royal Statistical Society Series B* **68**, 689–705.

Hannan, E. J. (1957), 'The variance of the mean of a stationary process', *Journal of the Royal Statistical Society Series B* **19**(2), 282–285.

Hays, S., Shen, H. & Huang, J. Z. (2012), 'Functional dynamic factor models with application to yield curve forecasting', *Annals of Applied Statistics* **6**(3), 870–894.

Hörmann, S. & Kokoszka, P. (2010), 'Weakly dependent functional data', *The Annals of Statistics* **38**(3), 1845–1884.

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

Horváth, L., Kokoszka, P. & Rice, G. (2014), 'Testing stationarity of functional time series', *Journal of Econometrics* **179**(1), 66–82.

Hosking, J. R. M. (1981), 'Fractional differencing', *Biometrika* **68**(1), 165–176.

Human Fertility Database (2016), *Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria)*. Available at www.humanfertility.org (data downloaded on 10/August/2016).

Hurst, H. E. (1951), 'Long-term storage capacity of reservoirs', *Transactions of the American Society of Civil Engineers* **116**(1), 770–799.

Hurst, H. E. (1956), 'Methods of using long-term storage in reservoirs', *Proceedings of the Institution of Civil Engineers* **5**(5), 519–543.

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E. & Yasmeen, F. (2018), *forecast: Forecasting functions for time series and linear models*. R package version 8.4.
URL: https://CRAN.R-project.org/package=forecast

Hyndman, R. & Ullah, M. S. (2007), 'Robust forecasting of mortality and fertility rates: A functional data approach', *Computational Statistics and Data Analysis* **51**(10), 4942–4956.

Jowett, G. H. (1955), 'The comparison of means of sets of observations from sections of independent stochastic series', *Journal of the Royal Statistical Society Series B* **17**(2), 208–227.

Klepsch, J. & Klüppelberg, C. (2016), An innovations algorithm for the prediction of functional linear processes, Working paper, Technische Universität München.
URL: https://arxiv.org/abs/1607.05874

Klepsch, J., Klüppelberg, C. & Wei, T. (2017), 'Prediction of functional ARMA processes with an application to traffic data', *Econometrics and Statistics* **1**, 128–149.

Kokoszka, P. & Reimherr, M. (2017), *Introduction to Functional Data Analysis*, CRC Press, Boca Raton.

Kokoszka, P., Rice, G. & Shang, H. L. (2017), 'Inference for the autocovariance of a functional time series under conditional heteroscedasticity', *Journal of Multivariate Analysis* **162**, 32–50.

Kokoszka, P. & Zhang, X. (2012), 'Functional prediction of intraday cumulative returns', *Statistical Modelling* **12**(4), 377–398.

Kowal, D. R., Matteson, D. S. & Ruppert, D. (2017), 'A Bayesian multivariate functional dynamic linear model', *Journal of the American Statistical Association* **112**(518), 733–744.

Kowal, D. R., Matteson, D. S. & Ruppert, D. (2018), 'Functional autoregression for sparsely sampled data', *Journal of Business and Economic Statistics* **forthcoming**.

Lam, C. & Yao, Q. (2012), 'Factor modelling for high-dimensional time series: Inference for the number of ffactor', *The Annals of Statistics* **40**(2), 694–726.

Laurini, M. P. (2014), 'Dynamic functional data analysis with non-parametric state space models', *Journal of Applied Statistics* **41**(1), 142–163.

Li, D., Robinson, P. M. & Shang, H. L. (2018), Supplement to "long-range dependent curve time series".

Li, Y., Wang, N. & Carroll, R. (2013), 'Selecting the number of principal components in functional data', *Journal of the American Statistical Association* **108**(504), 1284–1294.

Liu, X., Xiao, H. & Chen, R. (2016), 'Convolutional autoregressive models for functional time series', *Journal of Econometrics* **194**(2), 263–282.

Lo, A. W. (1991), 'Long-term memory in stock market prices', *Econometrica* **59**(5), 1279–1313.

McDonald, P. (2000), 'Low fertility in Australia: Evidence, causes and policy responses', *People and Place* **8**(2), 6–21.

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

Poskitt, D. S. & Sengarapillai, A. (2013), 'Description length and dimensionality reduction in functional data analysis', *Computational Statistics and Data Analysis* **58**, 98–113.

R Core Team (2018), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: https://www.R-project.org/

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

Rice, G. & Shang, H. L. (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.

Robinson, P. M. (1994), 'Semiparametric analysis of long memory time series', *The Annals of Statistics* **22**(1), 515–539.

Robinson, P. M. (1995), 'Gaussian semiparametric estimation of long range dependence', *The Annals of Statistics* **23**(5), 1630–1661.

Robinson, P. M. (2005), 'Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction', *Econometric Theory* **21**(1), 171–180.

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

Shang, H. L. (2014), 'A survey of functional principal component analysis', *AStA Advance in Statistical Analysis* **98**(2), 121–142.

Shang, H. L. (2016), 'Mortality and life expectancy forecasting for a group of populations in developed countries: A multilevel functional data method', *The Annals of Applied Statistics* **10**(3), 1639–1672.

Shang, H. L. (2018), 'Bootstrap methods for stationary functional time series', *Statistics and Computing* **28**(1), 1–10.