Long-Range Dependent Curve Time Series

Degui Li, Peter M Robinson and Han Lin Shang

University of York, London School of Economics and Australian National University

Seminar, University of Milano-Bicocca, February 14 2019

1 Introduction

Functional time series, or curve processes, arise in many fields, including biology, transportation, environmental science, finance and demography.

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 intra-day price curves of a financial stock, minute-by-minute traffic flow, and electricity price curves.

The following Figure plots intra-day log-return curves for 6 US stocks observed from 2 January 2014 to 31 December 2014.

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, age-specific fertility rates and yield curves.

The following Figure plots annual Australian age-specific fertility rates, where the function support for age lies between 15 and 49.

In either case, the object of interest is a time series of random functions bounded within a finite interval.

The bulk of the functional time series literature assumes stationarity over the temporal dimension, indeed short-range dependence (SRD), even uncorrelatedness.

This work has entailed both parametric and nonparametric modeling, and is relevant to a wide variety of data.

On the other hand, there have been many notable developments, especially over the past thirty years or so, on long-range dependent (LRD), or long memory, time series models.

These describe processes with greater persistence than SRD ones, such that in the stationary case auto-covariances decay very 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, as well as in fields such as economics, finance, agriculture and geophysics have revealed evidence of LRD.

Already some evidence of LRD in functional time series has been found..

We focus initially on temporal sums of (typically regularly-spaced) observations across each curve, establishing asymptotic distribution theory for this simple statistic.

This is then used in justifying some tools of statistical inference.

Our work reflects both the parametric and semiparametric modeling found in the LRD literature.

Suppose $\{X_t : t \in \mathcal{Z}\}$ 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 $\mathcal{Z} = \{0, \pm 1, \ldots\}$.

Much classic literature assumes that $\{X_t\}$ is independent and identically distributed (i.i.d.) across t (e.g. Ramsey and Silverman (2005)).

Generally speaking, two major SRD functional time series structures have been studied.

One extends probabilistic and statistical tools developed for mixing sequences to the functional case (e.g. Ferraty and Vieu (2006)).

The other extends certain linear and nonlinear sequences and martingale or m-dependent approximation techniques to the functional case, 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 (Bosq (2000), Hormann and Kokoszka (2010), Horvath and Kokoszka (2012)) .

For LRD functional time series we decompose the functional observations X_t through projection onto a finite number of sub-spaces, spanned by orthonormal functions, which are defined via eigenanalysis on a covariance function given below.

The dependence degree for the projected curve process varies over different sub-spaces, in particular, the sub-space on which the projection of the curve linear process has the strongest dependence (or the strongest signal), called the dominant sub-space, typically contains much of the information carried by the original curve process and thus plays an important role in empirical studies.

Under mild conditions we establish a central limit theorem for the sum of the LRD curve process and its projection.

We also consider estimation of the long-run covariance function up to multiplication by a rate (cf the spectral density at zero frequency in the time series case), which is crucial in subsequent Functional Principal Component Analysis (FPCA).

Through FPCA we obtain consistent estimates of the orthonormal functions which span the sample dominant sub-space.

We also 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 the dominant sub-space.

In the paper, but not included in the talk, is a functional version of the fractionally integrated ARMA (FARIMA) process which is a natural extension of the time series FARIMA, and a Monte-Carlo study of finite-sample performance which employs it.

Section 2 reviews basic properties and notation for LRD time series.

Section 3 specifies LRD structure for the curve process and gives some limit theorems.

Section 4 constructs estimates of the orthonormal functions which span the dominant sub-space, the dimension of the dominant sub-space and the memory parameter.

Sections 5 and 6 provide two empirical applications.

Section 7 concludes.

2 Time series with LRD

Consider covariance stationary time series X_t, $t = 0, \pm 1, \pm 2, ...$, such that

$$X_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j},, \quad \sum_{j=0}^{\infty} b_j^2 < \infty,$$

for zero-mean, unit variance, uncorrelated ε_t.

X_t has autocovariance function

$$\gamma_j = cov\left(X_t, X_{t+j}\right) = \sum_{k=0}^{\infty} b_k b_{j+k}.$$

X_t can be said to be SRD if

$$\sum_{j=0}^{\infty} \left|\gamma_j\right| < \infty,$$

as is implied by

$$\sum_{j=0}^{\infty} \left|b_j\right| < \infty$$

e.g white noise, stationary ARMA (where γ_j , b_j decay exponentially), and also if

$$b_j \sim j^{-\alpha} \text{ or } \gamma_j \sim j^{1-2\alpha}, \; \alpha > 1.$$

:

X_t can be said to be LRD if

$$\sum_{j=0}^{\infty} \left|\gamma_j\right| = \infty,$$

e.g. if

$$b_j \sim j^{-\alpha} \text{ or } \gamma_j \sim j^{1-2\alpha}, \ 1/2 < \alpha < 1,$$

as is true for FARIMA processes.

We can call α the memory parameter.

Alternative notations:

Self-similarity parameter $H = 3/2 - \alpha$, so $H \in (1/2, 1)$.

Differencing parameter $d = 1 - \alpha$, so $d \in (0, 1/2)$.

3 Modeling structure and large-sample properties

We introduce our model and functional LRD structure, give technical assumptions and state large-sample properties.

We assume a functional linear process

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

where $\{\eta_t : t \in \mathcal{Z}\}$ with $\eta_t = (\eta_t(u) : u \in \mathcal{C})$ is a sequence of i.i.d. (relaxable) random curves, and $\{b_j : j = 0, 1, 2, \cdots\}$ with $b_j = \Big(b_j(u,v) : u, v \in \mathcal{C}\Big)$ is a sequence of kernels with associated integral operators defined by $B_j(x)(u) = \int_{\mathcal{C}} b_j(u,v)x(v)dv$, $x \in \mathcal{H}$.

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 $\langle x_1, x_2 \rangle = \int_{\mathcal{C}} x_1(u)x_2(u)du$.

Define the operator norm of the coefficient operators in the curve process) as $\left\|B_j\right\|_O = \sup\left\{\left\|B_j(x)\right\| : x \in \mathcal{H}, \|x\| = 1\right\}$, where $\| \cdot \|$ is the L_2-norm of square integrable functions on $\mathcal{C}$.

If summability of $\left\|B_j\right\|_O$ over $j = 0, 1, 2, \cdots$ is imposed, the sequence of autocovariance functions between curves X_t and X_{t+k} is absolutely summable.

Then X_t is SRD.

Here we consider the more challenging case when the operator norm of the kernel coefficient operators is not summable.

Let the unnormalised long-run covariance function

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

be positive semi-definite.

Hence, there exist (eigenvalues) $\lambda_{n1} \geq \lambda_{n2} \geq \cdots \geq 0$ and an orthonormal function sequence $\{\psi_{ni} : i = 1, 2, \cdots\}$ with $\psi_{ni} = (\psi_{ni}(u) : u \in \mathcal{C})$ such that

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

For independent curves, eigenanalysis is usually conducted on the covariance function $\mathsf{E}[X_t(u)X_t(v)]$, seeking an optimal linear combination of functional observations with maximum variance function.

However, for curve time series this loses dynamic sample information, and it makes sense instead to use the long-run covariance function which accounts for autocorrelation .

As $c_n(u,v)$ is not normalised, some eigenvalues λ_{ni} diverge to infinity.

In particular, λ_{ni} reflects dependence in the time series

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

We derive

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

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

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

and assume

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

which implies that the curve process converges almost surely.

Write

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

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

It is straightforward to show that for each i

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

where $b_j^i(v) = \int_{\mathcal{C}} b_j(u, v)\psi_{ni}(u)du$, $v \in \mathcal{C}$.

Temporal dependence of x_t^i is determined by the decay rate of the function $b_j^i(v)$ over j.

Assumption 1 *The sequence η_t is composed of i.i.d. (relaxable) random functions in a measurable space with mean zero and positive definite, square integrable covariance function $c_\eta(u;v) = E[\eta_0(u)\eta_0]$.*

Assumption 2 *There exist bounded positive integers κ_0 and $p_1 < p_2 < ... < p_{\kappa_0}$ such that, for $i = p_{k-1} + 1, ..., ; p_k$ with $k = 1, ..., \kappa_0$ and $p_0 = 0$, $b_j^i(u) \sim \rho_i(u) j^{-\alpha_k}$ as $j \to \infty$, $u \in C, 1/2 < \alpha_1 < \alpha_2 < ... < \alpha_{\kappa_0} < 1$ (implying x_t^i is LRD, $i = 1, ..., \kappa_0$) with additional RC implying x_t^i is SRD, $i > \kappa_0$.*

Assumption 2 facilitates use of some well-known limit results developed in the time series LRD literature.

κ_0 is the number of sub-spaces on which the projection of $X_t(u)$ has LRD.

The $\left\{x_t^i : t \in \mathcal{Z}\right\}$, $i = p_{k-1} + 1, \cdots, p_k$, have the same dependence degree.

The series x_t^1 will be used to estimate α_1.

We have

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

where

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

and $p_{\kappa_0+1} = \infty$.

Here X_t is decomposed into a summation of $X_{tk} = (X_{tk}(u) : u \in \mathcal{C})$, the projection of the functional observations X_t onto the sub-spaces $\mathcal{S}_k$, $k = 1, \cdots, \kappa_0 + 1$, where $\mathcal{S}_k = \mathcal{S}_k(\psi_{ni} : p_{k-1} + 1 \leq i \leq p_k)$ is a $(p_k - p_{k-1})$-dimensional sub-space spanned by ψ_{ni}, $i = p_{k-1} + 1 \cdots, p_k$, and $\mathcal{S}_{\kappa_0+1} = \mathcal{S}_{\kappa_0+1}(\psi_{ni} : i \geq p_{\kappa_0} + 1)$ is the sub-space spanned by the eigenfunctions ψ_{ni}, $i = p_{\kappa_0} + 1, p_{\kappa_0} + 2, \cdots$.

Proposition 1. *Under Assumptions 1 and 2,* $\lambda_{ni} \sim n^{3-\alpha_k}$, $i = p_{k-1} + 1, ..., p_k$ *with* $k = 1, ..., \kappa_0$ *and* $\sum_{i=p_{k_0}+1}^{\infty} \lambda_{ni} = O(n)$.

The functional process X_t can correspondingly be decomposed into a summation of κ_0 functional LRD processes X_{tk}, $k = 1, \cdots, \kappa_0$, and a functional SRD process X_{t,κ_0+1}.

From Proposition 1 we may derive the explicit form of the asymptotic variance

Define

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

where $T_{nk}(u) = \sum_{t=1}^{n} X_{tk}(u)$.

Put $T_n =(T_n(u) : u \in \mathcal{C})$ 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 the cases $k = 1, \cdots, \kappa_0$ for which T_{nk} depends on the functional LRD process X_{tk}, as limit theorems for T_{n,κ_0+1} have been studied.

We have the following CLT.

Theorem 1 *If Assumptions 1 and 2 hold and* $\max_{1\le k\le\kappa_0}(p_k - p_{k-1})$ *is bounded, for each* $k = 1, \cdots, \kappa_0$,

$$n^{-(3/2-\alpha_k)} T_{nk} \xrightarrow{d} \mathsf{Z}_k,$$

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

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

Furthermore, the Gaussian random elements Z_k, $k = 1, \cdots, \kappa_0$, *are mutually independent.*

The asymptotic normal distribution can be also expressed as

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

where N_i, $i = 1, 2, \cdots, p_{\kappa_0}$, are independent standard normal random variables, and ψ_i is the limit of ψ_{ni}.

Theorem 1 and Proposition 1 show that T_n is asymptotically dominated by T_{n1} and in particular,

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

where $\| \cdot \|_{\mathcal{S}}$ is the Hilbert-Schmidt norm of the Hilbert-Schmidt operator and C_n^T and C_{n1}^T are the covariance operators defined by

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

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

Corollary 1 *If the conditions in Theorem1 are satisfied,*

$$n^{-(3/2-\alpha_1)}T_n \xrightarrow{d} \mathsf{Z}_1.$$

Theorem 1 and Corollary 1 not only extend classic limit theorems for LRD time series to the functional case, but also extend some theorems for SRD functional processes to LRD.

They hold with slight modification when we generalise to $b_j(i) \sim j^{-\alpha_k} l_i(j)$ as $j \to \infty$, where $l_i(\cdot)$ is a slowly varying function.

We assume that the dimension p_1 of the dominant sub-space $\mathcal{S}_1$ is fixed.

But we could allow $p_1 \rightarrow \infty$ as $n \rightarrow \infty$.

Our limit distribution results still hold under modified conditions.

4 Estimation

The projected functional process onto $\mathcal{S}_1$ contains much of the information carried by the original functional process and in particular,

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

As the orthonormal functions $\psi_{n1}, \cdots, \psi_{np_1}$ depend on n, we define their limits as

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

We show how to estimate the long-run covariance function (up to multiplication by a rate) for the functional process and use it to obtain estimates of the limit orthonormal functions $\psi_1, \cdots, \psi_{p_1}$ via FPCA.

Then we discuss how to determine p_1, and how to estimate α_1.

In order to carry out statistical inference on the function T_n defined above, for example to set confidence regions, we need to consistently estimate the limiting covariance function

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

For "$I(0)$" time series z_t, given for example by a linear process:

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

for s on the unit circle of the complex plane, where $\{\epsilon_t : t \in \mathcal{Z}\}$ 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 $\sigma^2 \left(\sum_{j=0}^{\infty} b_j \right)^2$, 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 1950’s statistical literature, so relevant statistical literature long precedes the econometric work.

If

$$b_0 = 1, \quad b_j \sim j^{-\alpha} \text{ as } j \to \infty$$

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 $\sum_{t=1}^{n} z_t$.

But in the latter case, suitable studentisations have been developed, depending in part on consistent estimates of d (indeed these apply also to "antipersistent" series $I(d)$ with $-1/2 < d = 1 - \alpha < 0$, i.e., $0 < H = 3/2 - \alpha < 1/2$, where the spectral density vanishes at frequency zero).

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

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

where $|k| \leq m$, $m = m_n$ satisfies $m \to \infty$ and $m = o(n)$.

If the memory parameter α_1 is known a priori, define

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

as an estimate of $c(u,v)$.

Let $\overline{C}_m$ and C be the operators defined by

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

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

Proposition 2 *If the conditions of Theorem 1 are satisfied,* $E\|\eta_t\|^4 < \infty$ *and* $m \sim n^\gamma$, $0 < \gamma < 1/(4\alpha_1 - 2)$,

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

The estimate $\overline{c}_m(u,v)$ is infeasible if α_1 is unknown.

For an estimate $\widetilde{\alpha}_1$ such that $\widetilde{\alpha}_1 - \alpha_1 = o_P(1/\log n)$,

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

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

We next estimate the orthonormal functions ψ_i, $i = 1, \cdots, p_1$.

From Proposition 1, the p_1 largest eigenvalues $\lambda_1, \cdots, \lambda_{p_1}$ of $c(u, v)$ 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, we may show that $\psi_1, \cdots, \psi_{p_1}$ are the eigenfunctions of $c(u, v)$ corresponding to the first p_1 largest eigenvalues.

Hence, we can implement FPCA on $\widetilde{c}_m(u, v)$, a consistent estimate of $c(u, v)$, and estimate the orthonormal functions $\psi_1, \cdots, \psi_{p_1}$.

But since $\widetilde{c}_m(u,v)$ is proportional to

$$\widehat{c}_m(u,v) = \sum_{|k|\le m} (m-|k|)\,\overline{r}_k(u,v),$$

the eigenfunctions via FPCA of $\widetilde{c}_m(u,v)$ are the same as those of $\widehat{c}_m(u,v)$, and don't require estimating α_1.

Hence we can do 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 first 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 *If the conditions of Propostion 2 are satisfied, p_1 is known, and $0 < \lambda_{p_1} < \cdots < \lambda_1 < \infty$, then*

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

and

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

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

We estimate α_1 by the R/S method introduced to study the behavior 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

$$R_n = \max_{1 \le k \le n} \sum_{t=1}^{k} \left(x_t^1 - \overline{x}^1 \right) - \min_{1 \le k \le n} \sum_{t=1}^{k} \left(x_t^1 - \overline{x}^1 \right)$$

and

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

The eigenfunction ψ_1 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 $\widehat{R}_n$ and $\widehat{S}_n^*$ like R_n and S_n^* but with x_t^1 replaced by $\widehat{x}_t^1$ and a feasible R/S statistic $\widehat{R}_n/\widehat{S}_n^*$

Proposition 3 *Under the conditions of Theorems 1 and 2,* $n^{-H_1}\widehat{R}_n/\widehat{S}_n^* \to_d V$, *where* $H_1 = 3/2 - \alpha_1$ *and*

$$V = \left(E\left(x_t^1\right)^2\right)^{-1/2}\left(\sup_{0\leq r\leq 1}\left|B_{H_1}(r) - rB_{H_1}(1)\right| - \inf_{0\leq r\leq 1}\left|B_{H_1}(r) - rB_{H_1}(1)\right|\right)$$

$B_H(r)$ *being fractional Brownian motion with index* H*.*

This motivates estimating α_1 by

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

Using Proposition 3, we may show that

$$\widehat{\alpha}_1 - \alpha_1 = O_P\,(1/\log n)$$

Although $\widehat{\alpha}_1$ is consistent, its convergence rate is thus very slow, so we even cannot achieve the consistency of $\widetilde{C}_m$ if $\widehat{\alpha}_1$ is used to construct $\widetilde{c}_m(u,v)$.

But as discussed above we do not require an estimate of α_1 when conducting eigenanalysis.

We conjecture that a faster rate of convergence and a normal limit distribution can be obtained if, say, local Whittle is used to estimate α_1.

Another critical issue in practical implementation is to determine p_1, the dimension of the dominant sub-space $\mathcal{S}_1$.

Letting $\widehat{\lambda}_{m,i}$ be the i^{th} largest eigenvalue of $\widehat{c}_m(\cdot,\cdot)$, we might estimate p_1 by

$$\widehat{p}_1 = \arg\min_{1\le i\le P} \frac{\widehat{\lambda}_{m,i+1}}{\widehat{\lambda}_{m,i}},$$

where P is a pre-specified positive integer and $0/0 = 1$. .

Proposition 4 *Under the conditions of Proposition 2,* $\widehat{p}_1 \rightarrow p$*.*

5 Application to US stocks

We apply our methodology to intraday log-returns of six stocks over a time period of one year from 2 January 2014 to 31 December 2014, containing 249 trading days after removing the half trading day 24 December 2014.

Let $p_t(u)$ denote the price of a stock on day t at time u and intraday log-returns defined by

$$X_t(u) = \ln p_t(u) - \ln p_t(u-h),$$

where h is a time window typically chosen as 1, 5 or 15 minutes.

Only $h = 5$ and therefore 5-minute log-returns are considered, to avoid microstructure noise.

Since all stocks trade from 9:30am to 4:00pm, there are 78 measurements per day and $n = 249$ curves denoted by $\{X_1, X_2, \ldots, X_{249}\}$ with $X_t = (X_t(u) : u \in \mathcal{C})$, where $\mathcal{C}$ is the time interval between 9:30am and 4:00pm.

The intraday log-returns for the six stocks over 249 trading days are plotted in the following Figure.

The high-dimensionality of the intraday observations make the application of existing univariate/multivariate LRD estimation procedures impractical.

The estimated eigenfunction corresponding to the maximum eigenvalue and their corresponding principal component scores are plotted in the following Figures.

The persistence in the pattern of the functional time series is quite strong, and we estimate α_1 by R/S.

For these six data sets, the estimates of α_1 are presented in the following Table, from which we may conclude a consistent, persistent pattern.

6 Application to age-specific fertility rates

Annual Australian fertility rates from 1921 to 2006 for each age from 15 to 49 were obtained from the Australian Bureau of Statistics (Cat.No.3105.0.65.001, Table 38).

There are $n = 86$ curves from 1921 to 2006 denoted by $\{X_1, X_2, \dots, X_{86}\}$, where $\mathcal{C}$ for age lies between 15 and 49.

These give 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.

By analysing changes in fertility rates as a function of both age and calendar year, fertility rates have changed slowly over time, reflecting changing social conditions affecting fertility, as shown in the following Figure.

For example, there is an increase in fertility in all age groups around the end of World War II (1945), a rapid increase in fertility during the 1960s which corresponds to the baby boom period, and an increase in fertility at higher ages in more recent years caused by a delay in child-bearing.

The estimated eigenfunction corresponding to the largest eigenvalue, along with its corresponding principal component scores, are given in the following Figure.

Using a stationarity test, based on a p-value of 0.12, we cannot reject at 5% level the null hypothesis of stationarity.

We estimate α_1 by R/S to be 0.7051.

.

We also consider three other developed countries, all with relatively long series commencing before 1950.

The estimates of α_1 , all of which are less than 0.71, along with their respective time periods and the p-values of the stationarity test, are shown in the following Table.

7 Conclusion

We have introduced a curve linear process with LRD, derived some relevant methodology and asymptotic theorems, and apllied them in two empirical data sets.

Various extensions might be pursued.

As one example, a hypothesis testing procedure based on a modified R/S statisticmight be developed to detect the presence of LRD in functional time series.

As another stationary LRD functional time series might be extended to nonstationarity, leading to study of a functional FARIMA(p, d, q) with $d = d(u) > 1/2$, and a functional version of fractional cointegration.