Journal of Econometrics 212 (2019) 377–392

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A moment-based notion of time dependence for functional time series

Nazarii Salish [a,*], Alexander Gleim [b]

[a] Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe, (Madrid), Spain
[b] Cognite AS, Oksenøyveien 10, 1366 Lysaker, Norway

ARTICLE INFO

Article history:
Received 11 September 2017
Received in revised form 26 October 2018
Accepted 18 March 2019
Available online 8 May 2019

JEL classification:
C18
C55
C14

MSC:
primary 62M10
secondary 62G10
62G20

Keywords:
Functional time series
Time dependencies
Dimension reduction
Principal components
Asymptotics

ABSTRACT

This paper addresses the fundamental topic of time dependence for time series when data points are given as functions. We construct a notion of time dependence through the projections on the basis system extracted from the principal components of normalized sums. This allows us to adapt various scalar time series techniques to the functional data context. In particular, we define dependence based on the autocovariances and cumulants of the projections, covering short and long memory scenarios. This notion naturally applies to linear processes. We illustrate the applicability of this moment based approach through several statistical problems in functional time series: (i) investigating the consistency of the estimator of the functional principal components under short and long memory, (ii) estimating the long-run covariance function and (iii) testing for short memory against the long memory alternative.

1. Introduction

The availability of large amounts of data offers manifold opportunities for researchers to obtain a better understanding of the underlying processes. Examples of such big data sets include high frequency data such as financial transactions, environmental data such as ozone or insolation maps and economic data such as income distributions or yield curves, to name a few. Functional data analysis (FDA) has emerged recently as a field offering suitable statistical tools to describe, model or predict relevant characteristics of such data. (See, e.g., Ramsay and Silverman, 2005, Horváth and Kokoszka, 2012 and Hsing and Eubank, 2015.) In this paper we focus on time series of functional data, which is commonly referred to as functional time series (FTS). In particular, our main aim is to propose a concept of the time dependence for FTS, which enables us to investigate the (limiting) properties of various tools for the inference and analysis of FTS, and present its theoretical foundation and justification.

In the context of classical (i.e., finite dimensional) time series analysis, various notions of dependence have been introduced to account for and analyze relationships between observations across time. For instance, ergodicity or different

* Corresponding author.
E-mail addresses: nsalish@eco.uc3m.es (N. Salish), alexander.gleim@cognite.com (A. Gleim).

https://doi.org/10.1016/j.jeconom.2019.03.007

types of mixing conditions are well established and frequently used (see, e.g., Davidson, 1994). Yet, in the functional context, only few concepts are available to work with time-dependent observations. Existing contributions have largely focused on two frameworks. First, Bosq (2000) in his seminal monograph develops a theory for general linear processes with focus on the first order functional autoregressive model (FAR(1)). This model and its extensions have become the main toolbox for the analysis of FTS. (See, e.g., Besse et al., 2000, Antoniadis et al., 2006, Park and Qian, 2012, Chang et al., 2016, among many others.) The second key reference is Hörmann and Kokoszka (2010) who introduce weak dependence using L^p-m-approximability, which provides a more general framework of dependence compared to FAR(1) since it is not tied to a specific model and applies to both linear and nonlinear dependence.

In this paper we suggest an alternative concept of time dependencies for FTS which is based on the idea of functional principal components (FPCs).[1] Using functional principal component analysis (FPCA) of the normalized sums of covariance functions, each functional observation of a weakly stationary FTS can be expressed via two sequences: a nonstochastic basis system and a stochastic scalar series of projections into this system (as in the Karhunen–Loève representation). It is intuitively appealing to quantify the dependence between functional observations through their respective projections (or scores) since this enables us to adapt various concepts of dependence available for scalar time series to the functional context. We propose a moment notion based on the autocovariances and cumulants of FPCs scores. This notion provides a convenient framework to study the asymptotic properties of tools developed for inference, estimation or forecasting of FTS. We select several problems of recent interest in FTS analysis to illustrate this claim. Proposed notion has been used in scalar time series for investigating various problems (see, e.g., Andrews, 1991, Gonçalves and Kilian, 2007 and Demetrescu et al., 2008), however, it is particularly useful to study scenarios when time series exhibit long memory, which has not yet received attention in the FDA literature. To the best of our knowledge there are only few results available: a limit theorem for partial sums of Banach spaced valued long memory processes has been obtained by Račkauskas and Suquet (2010) and limiting results for polygonal lines processes in Hilbert space are discussed in Račkauskas and Suquet (2011).

In the first part of the paper, we present the framework, the set of assumptions that describe the dependence in FTS, and provide motivation and intuition behind them. We continue our discussion with examples of processes that fulfill these assumptions and additional tools for asymptotic analysis.

In the second part of the paper we illustrate the usefulness of applying the autocovariance and cumulant dependence measure. In particular, three problems are considered. First, we investigate the asymptotic properties of the covariance operator estimator and its eigencomponents. The reason for this is that FPCA has become a cornerstone technique in FDA and many procedures rely on the consistency of FPCs estimates. (See e.g., Reiss and Ogden, 2007, Müller and Yao, 2008, Aue et al., 2015.) We show that convergence of a sample covariance to its population counterpart can be established under long memory of the process, however, it does not have a trivial convergence rate. This property is inherited by the autocovariance operator and the estimated FPCs. In turn it extends the results available for iid observations in Dauxois et al., 1982 and weakly dependent data in Hörmann and Kokoszka, 2010. (More details will be given in Section 3.1.) Second, we establish consistency of the long run covariance estimator under long memory. Results for short memory processes are available, for instance, in Horváth et al. (2013) and Berkes et al. (2016)). Third, we argue that the recently proposed KPSS-type test for stationarity of FTS in Horváth et al. (2014) can also be used against the long memory alternative following (Lee and Schmidt, 1996) and Giraitis et al. (2003).

The remainder of this paper is organized as follows. Section 2 introduces a new notion of dependence for functional time series, offers examples and discusses additional limiting results. In Section 3 we address the theoretical applications mentioned above. Concluding remarks are given in Section 4. All proofs are collected in Appendix.

2. Methodology and assumptions

We shall assume that we observe a series of functional observations $\{X_t(u)\}$ for $u \in [a, b]$ and $t = 1, \ldots, T$, where the interval $[a, b]$ is normalized to $[0, 1]$. For each t an observation X_t belongs to the Hilbert space $H = L^2([0, 1], \|\cdot\|)$ of square integrable functions which is equipped with a norm $\|\cdot\|$ induced by the inner product $\langle x, y\rangle \equiv \int_0^1 x(u)y(u)du$. The object $\{X_t(u)\}_{t=1}^T$ is referred to as functional time series (see e.g., Horváth and Kokoszka, 2012, Chapter 13–16 and Bosq, 2000, for a survey on FTS analysis) and we refer to t as the time index. In what follows the data $\{X_t\}$ are assumed to be given in functional form since the problem of data representation in functional form has been extensively studied in the literature (see, e.g., Ramsay and Silverman, 2005, for a review of the available techniques). We restrict attention to weakly stationary processes $\{X_t(u)\}$. For future reference, let $\mathcal{S}$ denote the space of Hilbert–Schmidt operators from H to H which is equipped with the operator norm $\|\cdot\|_{\mathcal{S}}$ (i.e., for some $\Psi \in \mathcal{S}$, $\|\Psi\|_{\mathcal{S}} = \left(\sum_{i=1}^{\infty}\|\Psi(e_i)\|^2\right)^{1/2}$ for any orthonormal basis $\{e_i\}_{i\geq1}$). The space of bounded linear operators on H is denoted by $\mathcal{L}$ with the norm $\|\Psi\|_{\mathcal{L}} = \sup_{\|x\|\leq1}\{\|\Psi(x)\|,\ x \in H\}$. All random functions are defined on a common probability space $(\Omega, \mathcal{A}, P)$. Let $L_H^p(\Omega, \mathcal{A}, P)$ denote the space of H valued random variables X such that for $p \geq 1$, $\mathbb{E}\|X\|^p < \infty$ or more conveniently, we shall say that X has finite p moments.

Consider a weakly stationary FTS $\{X_t\}_{t=1}^T$ such that $X_t \in L_H^2$ for each $t = 1, \ldots, T$ and an orthonormal basis system $\{\psi_\ell\}_{\ell\geq1}$ on H. Then each random function of a given FTS can be projected on $\{\psi_\ell\}_{\ell\geq1}$ as follows

$$X_t(u) = \mu(u) + \sum_{\ell=1}^{\infty}\theta_{t,\ell}\psi_\ell(u), \tag{1}$$

[1] See, e.g., Karhunen, 1947, Loève, 1946, or Ramsay and Silverman, 2005, for more details of functional principal components in FDA.

where $\mu := \mathbb{E}(X)$ and $\theta_{t,\ell} = \langle X_t - \mu, \psi_\ell \rangle$ denotes the ℓth projection or score of X_t. That is, each random function can be represented via a sequence of scalars $\theta_{t,\ell}$. Instead of defining time dependence between functional elements we consider a notion based on their projections. The existence of (1) and the idea of using projections from Eq. (1) is closely related to the functional principal component analysis and Karhunen–Loève representation. However, in the case of FTS, the selection of a basis $\{\psi_\ell\}_{\ell \geq 1}$ requires a separate analysis which we provide in the following section.

The following assumptions formalize how time dependencies between functional observations $\{X_t\}$ are translated into their score series. Let $\lambda_\ell = \mathbb{E}[\theta_{t,\ell}^2]$ and $\kappa_{\ell_1,\ldots,\ell_p}(\tau_1, \ldots, \tau_{p-1})$ denote the pth order cumulant of $(\theta_{t,\ell_1}, \theta_{t+\tau_1,\ell_2}, \ldots, \theta_{t+\tau_{p-1},\ell_p})$, where $\tau_1, \ldots, \tau_{p-1} \in \mathbb{N}$ are integers (see, e.g., Brillinger, 2001, p.19 for a more detailed description of cumulants). Then we shall assume:

Assumption 1. Define $B_{\ell,m}^{(h)} := \mathbb{E}\left[\theta_{t,\ell}\theta_{t-h,m}\right]$. Then there exists a constant $K > 0$ and some $\beta > 0$ such that

$$\left|B_{\ell,m}^{(h)}\right| \leq K\, h^{-\beta}\sqrt{\lambda_\ell \lambda_m}.$$

Assumption 2. For $p \geq 3$ and some constant $K > 0$, the joint pth order cumulants are absolutely summable

$$\sum_{\tau_1,\ldots,\tau_{p-1}=-\infty}^{\infty} \left|\kappa_{\ell_1,\ldots,\ell_p}(\tau_1, \ldots, \tau_{p-1})\right| \leq K \prod_{j=1}^{p} \lambda_{\ell_j}^{1/2}.$$

In Assumption 1, $B_{\ell,m}^{(h)}$ represents a measure of (auto)covariance between the score series $\left\{\theta_{t,\ell}\right\}$ and the lagged series $\left\{\theta_{t-h,m}\right\}$. It implies that absolute summability of the autocovariances of $\left\{\theta_{t,\ell}\right\}$ is not required across the time dimension. More precisely, under Assumption 1 we have

$$\sum_{h=1}^{T} \left|\mathbb{E}\left[\theta_{t,\ell}\theta_{t-h,m}\right]\right| = \begin{cases} O\left(T^{1-\beta}\sqrt{\lambda_\ell\lambda_m}\right) & \text{for } 0 < \beta < 1, \\ O\left(\log(T)\sqrt{\lambda_\ell\lambda_m}\right) & \text{for } \beta = 1, \\ O\left(\sqrt{\lambda_\ell\lambda_m}\right) & \text{for } \beta > 1, \end{cases}$$

as $T \to \infty$, where parameter β controls the speed of decay of the (auto)covariances of $\left\{\theta_{t,\ell}\right\}$. If $0 < \beta < 1$ then the sum of autocovariances diverges for large T and in what follows we refer to this situation as long memory. Finally, a hyperbolic decay $h^{-\beta}$ in Assumption 1 is used to cover scenarios with short and long memory. However, if one is only interested in short memory processes it is more common in time series literature to assume instead exponential rates of the form ρ^h with $|\rho| < 1$.

Assumption 2 states absolute summability of the joint cumulants of $\left\{\theta_{t,\ell}\right\}$ up to pth order. This allows us to control the temporal dependencies in the pth moments of the score series across the spectral and time dimension. It also implies the finiteness of the pth moment, i.e., $\mathbb{E}\|X_t\|^p < K < \infty$ for all t and some $K > 0$. For more details on how moments are related to cumulants see Appendix A.1. In general this cumulant condition is standard for the time series literature (see, e.g. Andrews, 1991 and Brillinger, 2001) and provides us with a useful measure of the joint statistical dependence of higher order moments and a convenient tool for deriving consistency results and the corresponding rates of convergence. Finally, Assumption 2 can be replaced by a weaker condition that allows for long range dependence in cumulants. For instance, Assumption L in Giraitis et al. (2003) can be adopted for this purpose.

Assumption 2′. For $p \geq 3$, some constant $K > 0$ and $0 < \rho < 1$, the joint pth order cumulants satisfy the condition

$$\sup_{\tau_1} \sum_{\tau_2,\ldots,\tau_{p-1}=-T}^{T} \left|\kappa_{\ell_1,\ldots,\ell_p}(\tau_1, \ldots, \tau_{p-1})\right| \leq K \left(\prod_{j=1}^{p} \lambda_{\ell_j}^{1/2}\right) T^{\rho}.$$

It is common in the long memory literature to tie the parameters β and ρ together such that $\beta = 1 - 2d$ and $\rho = 2d$ with $0 < d < 1/2$ (see, e.g., Giraitis et al., 2003). To keep the framework as general as possible we do not impose any such restrictions on the relationship between β and ρ.

Remark 1 (*Relation to Mixing*)**.** The form of time dependencies assumed in Assumption 1 and 2 is closely related to the concept of mixings on the projections $\{\theta_{t,\ell}\}$. In fact, α-mixing of $\{\theta_{t,\ell}\}$ together with finite p moments imply absolute summability of the joint cumulants up to order p (see, e.g. Andrews, 1991, Lemma 1). Hence, the main difference between the two approaches lies in the way the autocovariances are handled.

Remark 2 (*Relation to $L^p - m$ - Approximability*)**.** As pointed out in Hörmann and Kokoszka (2010) the direct comparison between $L^p - m$ - approximability and classical mixings and therefore our notion of time dependence is not possible, although both methods are moment based. From our point of view the main comparative difference between both approaches is the following. $L^p - m$ - approximability is easy to verify on a broad spectrum of processes and together with the limiting results in Berkes et al. (2013) provide a fairly general theoretical framework for investigating and developing tools for weakly dependent functional data (as illustrated, for example, in Horváth et al., 2013 and Horváth et al., 2014). The notion based on Eq. (1) and the restrictions on the behavior of projections $\{\theta_{t,\ell}\}$ given in Assumptions 1, 2 and 2′ provide an attractive alternative for FTS, since they allow us to adapt conveniently tools already available for the scalar case to FTS settings (as illustrated in Section 3). However, the hyperbolic decay $h^{-\beta}$ of the autocovariances in Assumption 1 and the cumulant conditions in 2 and 2′ are functional analogs of the classical restrictions on long memory processes in scalar time series. Therefore, our framework is tailored to study FTS that have long memory, which to our best knowledge have received little attention in the literature so far and can be seen as complementary to the short memory notion in Hörmann and Kokoszka (2010). To use our notion for short memory processes Assumptions 1, 2 and 2′ have to be appropriately modified. For instance, hyperbolic decay $h^{-\beta}$ should be substituted by exponential decay ρ^{-h}, which is more characteristic for the short memory scenarios.

For some applications an additional assumption might be of use.

Assumption 3. For some $\alpha > 1$ and all $\ell \geq 1$,

$\lambda_\ell - \lambda_{\ell+1} \geq K\ell^{-\alpha-1}$ for some $K > 0$.

Assumption 3 prevents the spacing between adjacent eigenvalues λ_ℓ from being too small and is standard in FDA. It is usually evoked for the estimation of the element of the spectrum of covariance operators and for ill-posed problems such as, for instance, the estimation of the inverse of the covariance operator. Corollary 2 presents an illustration of its use, where the asymptotic properties of the eigenfunction estimators are studied. Additionally, we provide an example in the supplementary Appendix on how Assumptions 1, 2 and 3 can jointly be used to address the ill-posed problem that arises when estimating the FAR model. The form of the restriction in Assumption 3 on the spacing between $\{\lambda_\ell\}$ is taken from seminal work of Hall and Horowitz (2007) on estimating functional linear regressions. A different approach can be found, for example, in Cardot et al. (1999) or Bosq (2000) (cf Theorem 8.7). Recently, Hörmann and Kidziński (2015) showed that for some applications Assumption 3 can be further relaxed to more general settings (in a data driven manner) which we do not pursue in what follows. In this paper we concentrate our attention on Assumptions 1 and 2 which describe the main framework for time dependencies, while Assumption 3 is optional and included for the completeness of the discussion.

It is useful to highlight the following technical implications of Assumption 3: we have that $\lambda_\ell \sim \ell^{-\alpha}$, and Assumption 3 assures jointly with Assumption 1 the absolute summability of the hth autocovariances of $\left\{\theta_{t,\ell}\right\}$ across all components, i.e.,

$$\sum_{\ell=1}^{\infty}\sum_{m=1}^{\infty}\left|\mathbb{E}\left[\theta_{t,\ell}\theta_{t-h,m}\right]\right| \leq \sum_{\ell=1}^{\infty}\sum_{m=1}^{\infty}B_{\ell,m}^{(h)} \leq Kh^{-\beta}.$$

2.1. Selection of the basis

One of the most well known techniques to obtain representation (1) and construct a basis $\{\psi_\ell\}_{\ell\geq 1}$ in FDA is to apply FPCA. The main idea behind FPCA is to use a spectral decomposition of the covariance operator, from which a basis $\{\psi_\ell\}_{\ell\geq 1}$ is extracted as eigenfunctions. Remark 3 provides a more detailed description.

Remark 3. Consider a weakly stationary FTS $\{X_t\}$ with $X_t \in L_H^2$ for every t. That is, X_t possesses a covariance operator $C(x) := \mathbb{E}\left[\langle X_t - \mu, x\rangle (X_t - \mu)\right]$, where $x \in L^2$, $\mu := \mathbb{E}(X_t)$, and C admits the spectral decomposition, i.e.,

$$C(x) = \sum_{\ell=1}^{\infty}\lambda_\ell\langle\psi_\ell, x\rangle\psi_\ell,$$

where $\{\lambda_\ell\}_{\ell\geq 1}$ is the sequence of eigenvalues (in descending order) and $\{\psi_\ell\}_{\ell\geq 1}$ denotes the corresponding sequence of eigenfunctions (i.e., $C(\psi_\ell) = \lambda_\ell\psi_\ell$). If C has full rank then $\{\psi_\ell\}_{\ell\geq 1}$ forms an orthonormal basis system of H. It follows that X_t for every t can be rewritten as in (1) or in other words admits the Karhunen–Loève representation, where a projection $\theta_{t,\ell} = \langle X_t - \mu, \psi_\ell\rangle$ is referred to as the ℓth functional principal component score of X_t. By construction, the elements of the sequence $\left\{\theta_{t,\ell}\right\}_{\ell\geq 1}$ are uncorrelated across the spectral dimension ℓ, have mean zero and variance λ_ℓ. For further readings on FPCA we refer the reader to Ramsay and Silverman (2005) and Kokoszka and Reimherr (2017).

The FPCA presented in Remark 3 is as a key technique in functional data analysis for independent and identically distributed variables. It also has received some attention in FTS applications. For instance, as discussed in Aue et al. (2015) such an approach can be used as a simple and convenient approximation tool for forecasting FTS. However, since FPCA is developed for independent random variables it does not take into account serial dependence and the autocovariance structure of the process. Ignoring serial dependence, naturally embedded in time series, may result in misleading conclusions and inefficient procedures. (See, e.g., Hörmann et al., 2015 for a discussion on dimension reduction inefficiencies.)

In time series analysis one of the most essential elements used for constructing tools for inference or estimation are normalized sums of the process. This observation provides a natural motivation for reconstructing FPCA with normalized sums instead of single observations as in the conventional FPCA. Moreover, this approach allows us to take into account the structure of autocovariances of FTS and incorporate a long memory property that is consistent with the scalar time series literature. To this end we first need to introduce several notations. Consider FTS $\{X_t\}$ and its demeaned counterpart $\{Z_t := X_t - \mu\}$. Then long run covariance operator of $\{X_t\}$ is defined as

$$\Gamma(x) = \int_0^1 \gamma(u, v)x(v)\mathrm{d}v,$$

where $\gamma(u, v)$ is the long run covariance kernel

$$\begin{aligned}\gamma(u, v) &= \mathbb{E}\left[Z_0(u), Z_0(v)\right] + \sum_{h\geq 1}\mathbb{E}\left[Z_0(u)Z_h(v)\right] + \sum_{h\geq 1}\mathbb{E}\left[Z_0(v)Z_h(u)\right] \\ &= \sum_{h=-\infty}^{\infty} c_h(u, v), \end{aligned} \tag{2}$$

where $c_h(u, v)$ denotes the kernel of the autocovariance operator

$$C_h(x) = \mathbb{E}\left(\langle X_{t-h} - \mu, x\rangle (X_t - \mu)\right) \text{ with } x \in H.$$

It is shown in Horváth et al. (2013) that the series in (2) is convergent in $L^2([0, 1] \times [0, 1])$. Further, to accommodate the case when the FTS has a long memory we have to adjust the definition of $\gamma(u, v)$ appropriately. Long memory of a weakly stationary scalar process $\{X_t\}$ is typically described by hyperbolic decay $h^{-\beta}$ (for $0 < \beta < 1$) of the covariance function. (See, e.g., Beran, 1994 definition 2.1 and for a general overview of the long memory literature.) We generalize this concept to FTS as

$$c_h(u, v) \sim k(u, v)\, h^{-\beta}, \text{ with } 0 < \beta < 1, \tag{3}$$

where $k(u, v)$ is a deterministic positive definite function in $L^2([0, 1] \times [0, 1])$, that is, $\int\int k^2(u, v)\mathrm{d}u\mathrm{d}v < \infty$ and $\int\int k(u, v)x(u)x(v)\mathrm{d}u\mathrm{d}v > 0$ for $x \in L^2([0, 1])$. The asymptotic similarity "$\sim$" in (3) is defined in L^2 sense on the unit square $[0, 1] \times [0, 1]$, that is, $\int_0^1\int_0^1 \left(c_h(u, v) - K\, k(u, v)h^{-\beta}\right)^2 \mathrm{d}u\mathrm{d}v \to 0$ as $h \to \infty$, where K is a positive constant. Under (3) the definition of the long run covariance kernel $\gamma(u, v)$ has to be normalized by $T^{1-\beta}$ to assure convergence in $L^2([0, 1] \times [0, 1])$, i.e.,

$$\gamma_\beta(u, v) := \lim_{T\to\infty} \sum_{h=-T}^{T} \frac{c_h(u, v)}{T^{1-\beta}}.$$

Fix $L \geq 1$ and consider any orthonormal system $e_1, \ldots, e_L$ in H. Further, let $S_T^{(\beta)} = \frac{1}{T^{\max\{1-\beta/2, 1/2\}}}\sum_{i=1}^T X_t$, where the normalization $\frac{1}{T^{\max\{1-\beta/2, 1/2\}}}$ is chosen to account for both short memory with $\beta > 1$ and long memory with $0 < \beta < 1$ of the FTS. The system $e_1, \ldots, e_L$ is asymptotically optimal (in the mean square sense) if it minimizes the distance between $S_T^{(\beta)}$ and its projection $\widetilde{S}_T^{(\beta)} = \sum_{\ell=1}^L \langle S_T^{(\beta)}, e_\ell\rangle e_\ell$ on the subspace spanned by it, i.e.,

$$\begin{aligned}\lim_{T\to\infty} \mathbb{E}\left\|S_T^{(\beta)} - \widetilde{S}_T^{(\beta)}\right\|^2 &= \lim_{T\to\infty} \int_0^1\int_0^1 \frac{1}{T^{\max\{2-\beta, 1\}}}\sum_{s,t=1}^T c_{s-t}(u, v)\mathrm{d}u\mathrm{d}v \\ &\quad - \lim_{T\to\infty}\sum_{\ell=1}^L \int_0^1\int_0^1 \frac{1}{T^{\max\{2-\beta, 1\}}}\sum_{s,t=1}^T c_{s-t}(u, v)e_\ell(u)e_\ell(v)\mathrm{d}u\mathrm{d}v. \end{aligned} \tag{4}$$

Notice that,

$$\begin{aligned}\lim_{T\to\infty} \frac{1}{T^{\max\{2-\beta, 1\}}}\sum_{s,t=1}^T c_{s-t}(u, v) &= \lim_{T\to\infty}\sum_{h=-T}^T \left(1 - \frac{|h|}{T}\right)\frac{c_h(u, v)}{T^{\max\{1-\beta, 0\}}} \\ &= \begin{cases}\gamma^\beta(u, v) & \text{for } 0 < \beta < 1, \\ \gamma(u, v) & \text{for } 1 < \beta. \end{cases}\end{aligned}$$

Hence, (4) can be rewritten as $\int_0^1\int_0^1 \gamma^*(u, v)\mathrm{d}u\mathrm{d}v - \sum_{\ell=1}^L\int_0^1\int_0^1 \gamma^*(u, v)e_\ell(u)e_\ell(v)\mathrm{d}u\mathrm{d}v$, where $\gamma^*(u, v) = \gamma^\beta(u, v)$ if $0 < \beta < 1$ and $\gamma^*(u, v) = \gamma(u, v)$ if $\beta > 1$. Finally, by using arguments identical to standard FPCA we have that (4) is minimized when $e_1, \ldots, e_L$ is selected as eigenfunctions of the long run covariance Γ^* with kernel $\gamma^*(u, v)$. This establishes the last step in constructing Assumptions 1, 2 and 2′, where the basis $\{\psi_\ell\}$ is selected as eigenfunctions of the long run covariance function $\gamma^*(u, v)$.

Remark 4. One of the important implications from selecting a basis $\{\psi_\ell\}$ through normalized sums of X_t is that it allows us to define long memory in FTS in line with the long memory concept for scalar time series. In fact, notice that the speed of decay of the projections' autocovariances in Assumption 1 is implied by (3). That is, for fixed $\ell \geq 1$ and $m \geq 1$ we have

$$\mathbb{E}\left[\theta_{t,\ell}\theta_{t-h,m}\right] = \mathbb{E}[\langle X_t - \mu, \psi_l\rangle\langle X_{t-h} - \mu, \psi_m\rangle] = \int_0^1\int_0^1 c_h(u,v)\psi_\ell(u)\psi_m(v)\mathrm{d}u\mathrm{d}v$$
$$\sim h^{-\beta}\int_0^1\int_0^1 k(u,v)\psi_\ell(u)\psi_m(v)\mathrm{d}u\mathrm{d}v.$$

Remark 5 (*Dynamic FPCA*)**.** Alternatively, in the case of weak dependence, Assumptions 1 and 2 can be defined using score series from the dynamic FPCA as described in Hörmann et al. (2015) and Panaretos and Tavakoli (2013a). Dynamic score series are by construction uncorrelated in all lags and levels, which automatically simplifies Assumption 1. That is, $B^{(h)}_{\ell,m} = 0$ if $\ell \neq m$. However, to tackle the higher order dependencies, some form of Assumption 2 is still required across components.

2.2. Examples

To illustrate the range of applicability of the moment-based notion given in Assumptions 1, 2 and 2′ we give below several examples:

Gaussian noise. Consider a zero mean Gaussian white noise $\{X_t\}$ with $X_t \in L^2_H$. (See, e.g., Bosq, 2000, Section 3 for a definition of white noise in the Hilbert Space.) It is well known that the corresponding score series $\{\theta_{t,\ell}\}$ obtained from the Karhunen–Loève expansion is independent Gaussian. In this case, the fourth order cumulants are zero, hence the cumulant condition in Assumption 2 (with $p = 4$) is verified trivially as well as Assumption 1 with $\beta > 1$.

Linear process. Consider a general linear process $\{X_t\} \in L^4_H$, i.e.,

$$X_t = \sum_{i=0}^{\infty} \Phi_i\left(\varepsilon_{t-i}\right), \tag{5}$$

where the errors $\varepsilon_t \in L^4_H$ are iid with zero mean. If a sequence of bounded linear operators $\{\Phi_i\}_{i=1}^{\infty}$ satisfies

$$\sum_{i=0}^{\infty}\|\Phi_i\|_{\mathcal{L}} < \infty,$$

then $\{X_t\}$ satisfies Assumption 1 with $\beta > 1$ and Assumption 2 (with $p = 4$). See Appendix A.2 for a proof.

Linear process with long memory. To illustrate the existence of a functional process with the long memory property consider example (5) with condition

$$\|\Phi_i\|_{\mathcal{L}} \sim i^{d-1},$$

where $0 < d < 1/2$. Then the process $\{X_t\}$ satisfies Assumptions 1 and 2′ with $\beta = 1 - 2d < 1$ and $\rho = 2d$. The proof follows immediately from the arguments presented in Giraitis et al. (2003, p. 276) for scalar time series.

2.3. Additional results for linear processes

A combination of Assumptions 1 and 2′ with a generalized linear process representation as in (5) provide a fairly general and simple to handle framework for FTS analysis. For instance, it can be used directly to generalize a CLT for scalar linear processes with long memory (see, e.g., Hosking, 1996, Theorem 1) to FTS settings. The CLT for H-valued linear processes with short memory can be found in Merlevède et al. (1997).

Theorem 1. *Assume that a weakly stationary process $\{X_t\}$ satisfies a general linear representation*

$$X_t = \sum_{i=1}^{\infty}\int \phi_i(u,v)\varepsilon_{t-i}(v)\,\mathrm{d}v,$$

where $\{\phi_i(u,v)\}$ denotes kernels of a sequence of bounded linear operators $\{\Phi_i\}$ such that

$$\|\Phi_i\|_{\mathcal{L}} \sim i^{d-1}, \text{ with } 0 < d < 1/2. \tag{6}$$

Further, $\{\varepsilon_t\}$ denotes an iid H-valued sequence with zero mean and a covariance operator C_ε. Let $S_T = 1/T\sum_{t=1}^{T} X_t$, then we have

$$T^{1/2-d}S_T \xrightarrow{d} \boldsymbol{\Gamma}\left(0, \Omega_X\right) \text{ as } T \to \infty,$$

where $\boldsymbol{\Gamma}(0, \Omega_X)$ *is a centered Gaussian H-valued random variable with the covariance operator*

$$\Omega_X = \left(\sum_{i=1}^{\infty} \Phi_i\right) C_\varepsilon \left(\sum_{i=1}^{\infty} \Phi_i^*\right),$$

and Φ_i^* *denotes an adjoint operator of* Φ_i.

Remark 6. This CLT is not a trivial extension of the CLT for scalar linear processes with long memory presented for instance in Hosking (1996) Theorem 1. The CLT for scalar linear processes is based on the seminal Theorem 18.6.5 of Ibragimov and Linnik (1971). As shown in Merlevède et al. (1997) there is no analogue of Ibragimov and Linnik (1971) result in Hilbert spaces. In particular, if $\sum_{i=1}^{\infty} \|\Phi_i\|_{\mathcal{L}} < \infty$ is violated without any additional assumptions on the behavior of Φ_i the tightness of $T^{1/2-d}S_T$ may fail. Therefore, Eq. (6) is key to establish the limiting result in Theorem 1 together with Corollary 1.1 in Merlevède (1996).

3. Applications

To illustrate the applicability of the moment based notion in this section we provide an asymptotic analysis of the following problems in the FTS literature: the estimation of the functional principal components, the estimation of the long-run covariance function and the KPSS-type test for long memory.

3.1. Estimation of functional principal components

Various tools in FDA use FPCs directly (see, e.g., Reiss and Ogden, 2007, Müller and Yao, 2008, and Aue et al., 2015). However, in practice scores and other FPCs (C and its eigenvalues and eigenfunctions) are not known and have to be estimated. Therefore, it is necessary to verify the convergence of the estimated FPC to their population counterparts. Consistency results are available for independent observations (see, e.g., Dauxois et al., 1982) and for L^4-m-dependent functional data (see e.g., Hörmann and Kokoszka, 2010). In what follows we present the asymptotic properties of the estimated FPCs under the concept of time dependence stated in Assumptions 1, 2 and 2′. In particular, we show how long memory influence the convergence speed of the FPCs estimators.

Suppose we observe a sequence $X_1, \ldots, X_T$. The standard estimators for the mean function, μ, and the covariance operator, $C(x)$, are given by the following sample averages

$$\hat{\mu}(u) = \frac{1}{T}\sum_{t=1}^{T} X_t(u), \tag{7}$$

$$\widehat{C}_T(x) = \frac{1}{T}\sum_{t=1}^{T} \langle X_t - \hat{\mu}, x\rangle \left(X_t(u) - \hat{\mu}(u)\right), \quad x \in L^2. \tag{8}$$

Further, eigenvalues and eigenfunction of $C(x)$ are denoted as η_ℓ and ω_ℓ, respectively. Using $\widehat{C}_T(x)$, we can estimate η_ℓ and ω_ℓ from the eigenequation

$$\widehat{C}_T(\widehat{\omega}_\ell) = \widehat{\eta}_\ell\, \widehat{\omega}_\ell,$$

where $\hat{\eta}_\ell$ and $\widehat{\omega}_\ell$ denote the estimators of the corresponding variables. Typically estimates of eigenelements ($\hat{\eta}_\ell$ and $\widehat{\omega}_\ell$) can be obtained for $1 \le \ell \le L$ with arbitrary fixed but small L (i.e., $L < T$). Ramsay and Silverman (2005, Section 6.4) discuss practical and computational methods for solving eigenequations.

Lemma 1. *For a weakly stationary FTS* $\{X_t\}_{t=1}^{T}$ *that fulfills Assumption 1 such that* $X_t \in L_H^2$ *for all t we have*

$$\limsup_{T\to\infty} \frac{1}{\max\left\{T^{-\beta}, T^{-1}\right\}} \mathbb{E}\left\|\hat{\mu}_T - \mu\right\|^2 < \infty.$$

Furthermore, if Assumption 1 is replaced by $B_{\ell,m}^{(h)} \sim K\, h^{-\beta}\sqrt{\lambda_\ell\lambda_m}$ *then we also have*

$$\liminf_{T\to\infty} \frac{1}{\max\left\{T^{-\beta}, T^{-1}\right\}} \mathbb{E}\left\|\hat{\mu}_T - \mu\right\|^2 > 0.$$

Lemma 1 presents the consistency result for the estimator of μ. It follows that the upper bound for the convergence rate is given as $O\left(\max\left\{T^{-\beta}, T^{-1}\right\}\right)$. Moreover, this is also a lower bound if Assumption 1 is strengthened to a more conventional form as $\left|B_{\ell,m}^{(h)}\right| \sim K\, h^{-\beta}\sqrt{\lambda_\ell\lambda_m}$. For the applications considered in the rest of this paper the obtained results remain unchanged if the data are centered prior to the analysis, given the rate of convergence of $\hat{\mu}_T$. Hence, in what follows we shall assume without loss of generality that X_t has mean equal to zero for all $t = 1, \ldots, T$. The following result establishes the consistency of estimator (8).

Theorem 2. *Suppose that $\{X_t\}$ is a weakly stationary FTS such that $X_t \in L_H^4$ and fulfills Assumptions 1 and 2 (with $p = 4$) then*

$$\limsup_{T\to\infty} T^{\beta^*}\,\mathbb{E}\left\|\widehat{C}_T - C\right\|_{\mathcal{S}}^2 < \infty,$$

where $\beta^* := \min\{2\beta, 1\}$. If we have Assumption 2′ instead of Assumption 2 then

$$\limsup_{T\to\infty} T^{d^*}\mathbb{E}\left\|\widehat{C}_T - C\right\|_{\mathcal{S}}^2 < \infty,$$

where $d^* := \min\{2\beta, (1-\rho)\}$. Furthermore, if the conditions in Assumptions 1 and 2′ (or 2) are replaced by $B_{\ell,m}^{(h)} \sim K\,h^{-\beta}\sqrt{\lambda_\ell\lambda_m}$ and $\sup_{\tau_1}\sum_{\tau_2,\ldots,\tau_{p-1}=-T}^{T}\kappa_{\ell_1,\ldots,\ell_p}(\tau_1,\ldots,\tau_{p-1}) \sim K\left(\prod_{j=1}^{p}\lambda_{\ell_j}^{1/2}\right)T^{\rho}$ then we also have

$$\liminf_{T\to\infty} T^{d^*}\,\mathbb{E}\left\|\widehat{C}_T - C\right\|_{\mathcal{S}}^2 > 0.$$

Theorem 2 implies that the upper bound for the rate of convergence is given as $O\left(T^{-\beta^*}\right)$ or $O\left(T^{-d^*}\right)$ depending on the chosen assumption for the cumulants. This bound will also be a lower bound if Assumptions 1 and 2 or 2′ are strengthened such that they impose orders of magnitude instead of upper bounds on the behavior of the autocovariances and cumulants. According to Theorem 2 the fastest convergence rate T^{-1} can be achieved for the estimator $\widehat{C}_T$ when $\beta \geq 1/2$ and there is no long memory in higher order moments. This extends previously obtained results in Bosq (2000) and Hörmann and Kokoszka (2010) showing that the rate T^{-1} can also be obtained for processes that potentially have long range dependencies. In other words, Theorem 2 extends the available result for scalar time series (see, e.g., Hosking, 1996) to functional settings: the absolute summability of the autocovariances is not necessary to obtain convergence rate T^{-1} of the covariance estimator to the population counterpart. Finally, if one is only interested in establishing the consistency of the covariance operator estimator, Assumption 1 can be relaxed to $B_{\ell,m}^{(h)} \leq Kb_h\sqrt{\lambda_\ell, \lambda_m}$ with $\sum_{h=1}^{\infty} h^{-1}b_h < \infty$. This condition allows for a slow decay of the time dependencies represented by component b_h that can even be of logarithmic order $b_h = O\left(\ln(h)^{-1-\beta}\right)$ for $\beta > 0$.

The autocovariance operator defined as

$$C_h(x) = \mathbb{E}\left[\langle X_t, x\rangle X_{t-h}\right],$$

for $t = 1, \ldots, T$ and some h, can be estimated similarly by the sample analogue

$$\widehat{C}_{h,T}(x) = \frac{1}{T}\sum_{i=1}^{T-h}\langle X_t, x\rangle\left(X_{t+h}(t)\right), \tag{9}$$

and the following holds for any autocovariance operator of order h.

Corollary 1. *Suppose that $\{X_t\}$ is a weakly stationary FTS such that $X_t \in L_H^4$ and fulfills Assumptions 1 and 2′ (with $p = 4$) then*

$$\limsup_{T\to\infty} T^{d^*}\,\mathbb{E}\left\|\widehat{C}_{h,T} - C_h\right\|_{\mathcal{S}}^2 < \infty.$$

Furthermore, if we have $B_{\ell,m}^{(h)} \sim K\,h^{-\beta}\sqrt{\lambda_\ell\lambda_m}$ and $\sup_{\tau_1}\sum_{\tau_2,\ldots,\tau_{p-1}=-T}^{T}\kappa_{\ell_1,\ldots,\ell_p}(\tau_1,\ldots,\tau_{p-1}) \sim K\left(\prod_{j=1}^{p}\lambda_{\ell_j}^{1/2}\right)T^{\rho}$ then

$$\liminf_{T\to\infty} T^{d^*}\,\mathbb{E}\left\|\widehat{C}_{h,T} - C_h\right\|_{\mathcal{S}}^2 > 0.$$

Using Theorem 2 and Lemma 3.2 of Hörmann and Kokoszka (2010) explicit bounds for the mean squared error of the eigenelement estimators can be established.

Corollary 2. *Suppose that $\{X_t\}$ is a weakly stationary FTS such that $X_t \in L_H^4$ and fulfills Assumptions 1 and 2′ (with $p = 4$). Then, for all $1 \leq \ell \leq L$*

$$\mathbb{E}\left(\left|\hat{\eta}_\ell - \eta_\ell\right|^2\right) = O\left(T^{-d^*}\right) \quad \text{and} \quad \mathbb{E}\left(\|a_\ell\widehat{\omega}_\ell - \omega_\ell\|^2\right) = O\left(\delta_\ell^2 T^{-d^*}\right)$$

as $T\to\infty$, where $a_\ell := sign(\langle\widehat{\omega}_\ell, \omega_\ell\rangle)$, $\delta_\ell := \max\{(\eta_{\ell-1}-\eta_\ell)^{-1}, (\eta_\ell - \eta_{\ell+1})^{-1}\}$. Further, if Assumption 3 also holds we have

$$\sup_{1\leq\ell\leq L}\mathbb{E}\left(\|a_\ell\widehat{\omega}_\ell - \omega_\ell\|^2\right) = O\left(L^{2(1+\alpha)}T^{-d^*}\right)$$

as $L\to\infty$ and $T\to\infty$.

The results in Corollary 2 indicate that, as ℓ increases, it becomes more difficult to estimate the eigenfunctions ω_ℓ associated with η_ℓ since the expected L^2 error is proportional to δ_ℓ^2. As a consequence, the spacing between adjacent eigenvalues $\{\eta_\ell\}$ cannot decrease too fast. Further, by controlling the speed at which $|\eta_\ell - \eta_{\ell+1}|$ decreases we can allow L to increase with sample size T. This is done by evoking Assumption 3 which implies that $L \to \infty$ such that $L = o\left(T^{d^*/2(1+\alpha)}\right)$ as $T \to \infty$. Finally, the estimator $\widehat{\omega}_\ell$ of ω_ℓ is only identified up to a change in sign. As it is standard in the literature, we shall tacitly assume that the sign of $\widehat{\omega}_\ell$ is chosen such that $\int \widehat{\omega}_\ell \omega_\ell \geq 0$.

3.2. Estimation of long-run covariance

The long-run covariance operator is fundamental in time series analysis since it often plays a part in the asymptotic approximation of tools used for both functional and scalar time series. For instance, in the scalar case the variance of the sample mean converges to the long run variance (see, e.g., Anderson, 1971). In the FTS context (Horváth et al., 2013) study two sample mean testing problem and Horváth et al. (2014) suggest a test for stationarity of FTS. In both cases the limiting distributions depend on the long run covariance function. In this section we show that the asymptotic properties of the long run covariance operator can be easily established under the moment based notion suggested in Assumptions 1 and 2 and 2′.

Let $w_h(q)$ denote the Bartlett weights, i.e.,

$$w_h(q) = 1 - \frac{|h|}{q+1},$$

where $1 \leq h \leq q$. Then the estimator of the long run kernel $\gamma(u, v)$ is

$$\widehat{\gamma}_T(u, v) = \widehat{c}_0(u, v) + \sum_{h=1}^{q} w_h(q)\left(\widehat{c}_h(u, v) + \widehat{c}_h(v, u)\right), \tag{10}$$

where $\widehat{c}_h(u, v)$ is the estimator of the kernel of the autocovariance operator C_h, i.e.,

$$\widehat{c}_h(u, v) = \frac{1}{T}\sum_{t=h+1}^{T} X_t(u)X_{t-h}(v) \quad \text{with} \quad 0 \leq h \leq T-1.$$

Since the results for the short memory case are already established in the literature (see, e.g., Horváth et al., 2013, Panaretos and Tavakoli, 2013b and Berkes et al., 2016) in the following theorem we establish consistency of the estimator $\widehat{\gamma}_T(u, v)$ in the long memory case.

Theorem 3. *Let $\{X_t\}$ be a stationary process such that $X_t \in L^4_H$ for all t and let $q \to \infty$, $q/T \to 0$ as $T \to \infty$. Then under Assumption 1 with $0 < \beta < 1$ and Assumption 2′*

$$\limsup_{T\to\infty, q\to\infty} \max\left\{\left(\frac{T}{q}\right)^{2\beta}, \frac{T^{1-\rho}}{q^{2\beta-1}}\right\} \mathbb{E}\int_0^1\int_0^1 \left(\frac{\widehat{\gamma}_T(u, v)}{q^{1-\beta}} - \gamma_\beta(u, v)\right)^2 \mathrm{d}u\mathrm{d}v < \infty.$$

Further, if in addition $B^{(h)}_{\ell,m} \sim K\, h^{-\beta}\sqrt{\lambda_\ell\lambda_m}$ and $\sup_{\tau_1}\sum_{\tau_2,\ldots,\tau_{p-1}=-T}^{T} \kappa_{\ell_1,\ldots,\ell_p}(\tau_1, \ldots, \tau_{p-1}) \sim K\left(\prod_{j=1}^{p}\lambda_{\ell_j}^{1/2}\right)T^{\rho}$ then

$$\liminf_{T\to\infty, q\to\infty} \max\left\{\left(\frac{T}{q}\right)^{2\beta}, \frac{T^{1-\rho}}{q^{2\beta-1}}\right\} \mathbb{E}\int_0^1\int_0^1 \left(\frac{\widehat{\gamma}_T(u, v)}{q^{1-\beta}} - \gamma_\beta(u, v)\right)^2 \mathrm{d}u\mathrm{d}v > 0.$$

Theorem 3 shows that the limiting behavior of $\widehat{\gamma}_T$ is not trivial under long memory and hinges on the decay speed of the autocovariances (expressed through β) and cumulants (expressed through ρ). Also notice that our results are in line with those obtained for scalar time series. To see this, tie together parameters β and ρ such that $\beta = 1-2d$ and $\rho = 2d$, where $0 < d < 1/2$. From Theorem 3 we then have

$$\mathbb{E}\int_0^1\int_0^1 \left(\frac{1}{q^{2d}}\widehat{\gamma}_T(u, v) - \gamma_{1-2d}(u, v)\right)^2 \mathrm{d}u\mathrm{d}v = O\left(\max\left\{\left(\frac{q}{T}\right)^{2-4d}, \left(\frac{q}{T}\right)^{1-2d} q^{-2d}\right\}\right) = o(1),$$

as $T \to \infty$, $q \to \infty$ and $q/T \to 0$. Hence, the estimator of the long-run covariance operator in (10) converges to the population counterpart with the same rate as the estimator of the long run variance for the scalar time series under long memory (see, e.g., Theorem 3.1 in Giraitis et al., 2003)

3.3. Testing for long memory

This section presents an extension of the results of the recently proposed KPSS-type test in Horváth et al. (2014). For scalar time series it is established that the KPSS test of Kwiatkowski et al. (1992) can be used to test for long memory (see, e.g., Lee and Schmidt, 1996 and Giraitis et al., 2003). We show that the statistics considered in Horváth et al. (2014) to test for stationarity of FTS can be extended to account for the long memory alternative using the time dependence notion defined in Assumptions 1, 2 and 2′ and the results developed for scalar time series.

We start by aligning our assumptions of short and long memory in FTS with those for scalar times series given in Giraitis et al. (2003). In the following, let $W(r)$ and $W^{(d)}(r)$ with $0 \le r \le 1$ denote the standard and fractional Wiener process (Brownian motion), respectively. Then $\mathfrak{B}(r)$ and $\mathfrak{B}^{(d)}(r)$ are the standard and fractional Brownian Bridge, respectively. Let $\mathbf{s}_T(r,u)$ be the demeaned partial sums of X_t: $\mathbf{s}_T(r,u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} (X_t(u) - \mathbb{E}(X_t(u)))$. The null hypothesis of **short memory** is expressed as:

(i) Assumption 1 with $\beta > 1$ holds;
(ii) Assumption 2 with $p = 4$ holds.
(iii) It holds that for every T we can define a sequence of Gaussian processes $Z_T(r,u)$ such that $\{Z_T(r,u), 0 \le r,t \le 1\} \overset{\mathcal{D}}{=} \{Z(r,u), 0 \le r,t \le 1\}$, $Z(r,u) = \sum_{\ell=1}^{\infty} \sqrt{\lambda_\ell} W_\ell(r) \psi_\ell(u)$ and

$$\sup_{0 \le r \le 1} \int_0^1 (\mathbf{s}_T(r,u) - Z_T(r,u))^2 \, \mathrm{d}u = o_p(1), \text{ as } T \to \infty, \tag{11}$$

where $\{\lambda_\ell\}_\ell$ and $\{\psi_\ell\}_\ell$ denote eigenvalues and eigenfunctions of the long-run covariance of $\{X_t\}_{t=1}^T$, and $\{W_\ell\}_{\ell=1}^{\infty}$ are independent and identically distributed Wiener processes.

The **long memory** alternative is:

(i) Assumption 1 holds with $\beta = 1 - 2d$, where $0 < d < 1/2$;
(ii) Assumption 2′ holds with $p = 4$ and $\rho = 2d$.
(iii) It holds that for every T we can define a sequence of Gaussian processes $Z_T^{(d+1/2)}(r,u)$ such that $\{Z_T^{(d+1/2)}(r,u), 0 \le r,t \le 1\} \overset{\mathcal{D}}{=} \{Z^{(d+1/2)}(r,u), 0 \le r,t \le 1\}$, $Z^{(d+1/2)}(r,u) = \sum_{\ell=1}^{\infty} \sqrt{\lambda_\ell^{(d)}} W_{\ell,T}^{(d+1/2)}(r) \psi_\ell^{(d)}(u)$ and

$$\sup_{0 \le r \le 1} \int_0^1 \left(\frac{1}{T^d} \mathbf{s}_T(r,u) - Z_T^{(d+1/2)}(r,u) \right)^2 \mathrm{d}u = o_p(1), \text{ as } T \to \infty, \tag{12}$$

where $\{\lambda_\ell^{(d)}\}$ and $\{\psi_\ell^{(d)}\}$ are the eigenvalues and eigenfunctions of $\gamma^{(d)}(u,v)$, and $\{W_\ell^{(d)}\}$ are independently and identically distributed fractional Wiener processes with parameter d.

By stating (11) and (12) we intend to keep the theory below applicable for all processes that satisfy these invariance principles. Conditions under which (11) and (12) hold are well understood and studied in the case of scalar time series. (See, for instance, Billingsley (1968), Herrndorf (1984), Phillips (1987) and Lee and Schmidt, 1996 just to name a few.) In the case of dependent functional data several results are available. Merlevède et al. (1997) show that the linear process in (5) with $\sum_{i=1}^{\infty} \|\Phi_i\|_{\mathcal{L}} < \infty$ (i.e., under Assumption 1 and $\beta > 1$, and Assumption 2) satisfies the (functional) central limit theorem. Račkauskas and Suquet (2011) discuss limiting theorems for Hilbert-valued linear processes under long memory. Furthermore, Chen and White (1998) establish the central limit and the functional central limit theorem for Hilbert-valued stochastic processes under near epoch dependency on mixing processes. Also the invariance principle is available for FTS under L_p-m approximability in Berkes et al. (2013).

The test statistic takes the form:

$$R_T = \int_0^1 \int_0^1 Z_T^2(r,u) \mathrm{d}r \mathrm{d}u,$$

where $Z_T(r,u) = S_T(r,u) - rS_T(1,u)$ with $S_T(r,u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} X_t(u)$ for $0 \le r,u \le 1$. Under the null hypothesis of short memory we have

$$R_T \xrightarrow{D} \sum_{\ell=1}^{\infty} \lambda_\ell \int_0^1 \mathfrak{B}_\ell^2(r) \mathrm{d}r, \tag{13}$$

where $\{\mathfrak{B}_\ell\}$ is a sequence of independent Brownian bridges. It follows from (13) that the limiting distribution of R_T depends on the eigenvalues of the long-run covariance operator. To construct the statistics with limiting distribution free of parameters Horváth et al. (2014) considered a second statistic based on the projections on $\widehat{\psi}_\ell(u)$ given as

$$R_T(L) = \sum_{\ell=1}^{L} \frac{1}{\widehat{\lambda}_\ell} \int_0^1 \langle Z_T(r,\cdot), \widehat{\psi}_\ell \rangle^2 \mathrm{d}r,$$

Under the null hypothesis and given that estimators $\widehat{\lambda}_\ell$ and $\widehat{\psi}_\ell$ are consistent the limiting distribution is expressed as

$$R_T(L) \xrightarrow{D} \sum_{\ell=1}^{L} \int_0^1 \mathfrak{B}_\ell^2(r) \mathrm{d}r.$$

The asymptotic behavior of R_T and $R_T(L)$ tests under long memory hinges on two facts: (i) the asymptotic properties of the long run kernel estimator under Assumptions 1 and 2′ (given in Theorem 3) and (ii) the invariance principle in (12). To see this notice first that

$$\frac{1}{T^d}Z_T(r,u) = V_T^{(d)}(r,u) + \mu(u)\left(\frac{\lfloor Tr\rfloor - Tr}{T^{1/2+d}}\right),$$

where $V_T^{(d)}(r,u) = \frac{1}{T^d}\mathbf{s}_T(r,u) - \frac{1}{T^d} r\,\mathbf{s}_T(r,u)$. This implies that $\frac{1}{T^{2d}}R_T = \int_0^1\int_0^1 V_T^{(d)}(r,u)\mathrm{d}u\mathrm{d}r + o_p(1)$ and from (12) we have

$$\int_0^1\int_0^1 V_T^{(d)}(r,u)\mathrm{d}u\mathrm{d}r \xrightarrow{D} \sum_{\ell=1}^{\infty}\lambda_\ell \int_0^1 \left(\mathfrak{B}_\ell^{(1/2+d)}(r)\right)^2 \mathrm{d}r. \tag{14}$$

Putting together (14), Theorem 3 and Lemma 3.2 of Hörmann and Kokoszka (2010) yields

$$\left(\frac{q}{T}\right)^{2d} R_T(L) \xrightarrow{D} \sum_{\ell=1}^{L}\int_0^1 \left(\mathfrak{B}_\ell^{(1/2+d)}(r)\right)^2 \mathrm{d}r,$$

where $q \to \infty$, $q/T \to 0$ as $T \to \infty$. With the last result we can conclude that the $R_T(L)$ test in Horváth et al. (2014) in addition to change point, integrated and deterministic trend alternatives have also power against the long memory alternative.

4. Conclusion

In this paper a time dependence concept for functional observations is proposed. It is based on the FPCA of the normalized sums. In particular, time dependence in FTS is quantified through the autocovariances and cumulants of projections into the basis extracted from the long run covariance operator. This construction allows us to describe the long memory property in FTS. Furthermore, we present three theoretical illustrations using this new moment based notion of dependence. In particular, estimation of the FPCs, estimation of the long-run covariance function and KPSS-type tests are investigated.

Acknowledgments

We are thankful to the co-editor Jianqing Fan, the associate editor, two anonymous referees, Joerg Breitung, Miguel Delgado, Matei Demetrescu, Lajos Horváth, Alois Kneip, Piotr Kokoszka, Dominik Liebl, Mirjam Salish and Carlos Velasco for their constructive comments and suggestions that have substantially improved the original manuscript. The bulk of the work was conducted while Alexander Gleim was a Ph.D. student at the Bonn Graduate School of Economics and later employee at Statkraft AS.

The first author gratefully acknowledges the financial support from the German Research Foundation (Deutsche Forschungsgemeinschaft) through grant DE 1617/2-1, the Ministerio Economía y Competitividad (Spain), grant MDM 2014-0431, and Comunidad de Madrid, Spain, MadEco-CM (S2015/HUM-3444)

Appendix

A.1. Auxiliary results

The following combinatorial representation of pth order moments in terms of joint cumulants is often used for proofs and is stated here for future reference. For a set of random variables $x_1, \ldots, x_p$ one has

$$\mathbb{E}\left[x_1 \cdot \ldots \cdot x_p\right] = \sum_{\pi}\prod_{B\in\pi}\kappa_{(x_i : i\in B)}, \tag{A.1}$$

were π cycles through all possible partitions of the set $\{1, 2, \ldots, p\}$ and B cycles through all blocks of partition π. For instance, zero mean random variables satisfy the following expressions: $\kappa_{(x_1,x_2)} = \mathbb{E}\left[x_1x_2\right]$ for $p = 2$, $\kappa_{(x_1x_2x_3)} = \mathbb{E}\left[x_1x_2x_3\right]$ for $p = 3$ and

$$\begin{aligned}\kappa_{(x_1,x_2,x_3,x_4)} &= \mathbb{E}\left[x_1x_2x_3x_4\right] - \mathbb{E}\left[x_1x_2\right]\mathbb{E}[x_3x_4] \\ &\quad - \mathbb{E}\left[x_1x_3\right]\mathbb{E}[x_2x_4] - \mathbb{E}\left[x_1x_4\right]\mathbb{E}[x_2x_3]\,.\end{aligned} \tag{A.2}$$

A.2. Appendix: Proofs

Linear Process: Proof

First note that both $\{X_t\}$ and $\{\varepsilon_t\}$ admit Karhunen–Loève representation, i.e.,

$$X_t = \sum_{\ell=1}^{\infty} \theta_{t,\ell}\psi_\ell \quad \text{and} \quad \varepsilon_t = \sum_{m=1}^{\infty} \xi_{t,m}\phi_m,$$

where $\{\psi_\ell\}$ and $\{\phi_m\}$ are the corresponding sequences of orthogonal eigenfunctions and $\{\theta_{t,\ell}\}$ and $\{\xi_{t,m}\}$ are the corresponding sequences of FPC scores of long-run covariance functions (notice, since ε_t are iid then the long-run covariance will coincide with the covariance function). Then from (5) we have

$$\theta_{t,\ell} = \sum_{i=0}^{\infty}\sum_{m=1}^{\infty} \xi_{t-i,m} \langle \Phi_i(\phi_m), \psi_\ell \rangle = \sum_{i=0}^{\infty}\sum_{m=1}^{\infty} \xi_{t-i,m} a_{i,ml}, \tag{A.3}$$

where $a_{i,ml} = \langle \Phi_i(\phi_m), \psi_\ell \rangle$.

From relation (A.2) and following arguments of Anderson (1971, p. 466–467) we have

$$\kappa_{\ell_1,\ldots,\ell_p}(\tau_1, \tau_2, \tau_3) = \sum_{m=1}^{\infty} c_m^{(4)} \sum_{i=0}^{\infty} a_{i,ml} a_{i+\tau_1,ml} a_{i+\tau_2,ml} a_{i+\tau_3,ml},$$

where $c_m^{(4)}$ denotes a 4-order cumulant of $\xi_{t,m}$. Since $\varepsilon_t \in L_H^4$ we have $\sum_{m=1}^{\infty} |c_m^{(4)}| < \infty$. Further, $\sum_{i=0}^{\infty} |a_{i,ml}| \le \sum_{i=0}^{\infty} \|\Phi_i\|_{\mathcal{L}} \|\phi_m\| \|\psi_\ell\| = \sum_{i=0}^{\infty} \|\Phi_i\|_{\mathcal{L}} < \infty$. Then we have

$$\sum_{\tau_1,\tau_2,\tau_3=-\infty}^{\infty} \left|\kappa_{\ell_1,\ldots,\ell_p}(\tau_1, \ldots, \tau_{p-1})\right| \le \sum_{m=1}^{\infty} |c_m^{(4)}| \prod_{j=1}^{4} \left(\sum_{i=0}^{\infty} |a_{i,ml_j}| \right) < \infty. \quad \square$$

Proof of Lemma 1. We have

$$\begin{aligned}
\mathbb{E}\left\|\hat{\mu} - \mu\right\|^2 &= \frac{1}{T^2}\sum_{t,s=1}^{T} \mathbb{E}\langle X_t - \mu, X_s - \mu\rangle = \frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{\ell=1}^{\infty} \mathbb{E}\left[\theta_{t,\ell}\theta_{s,\ell}\right] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{\ell=1}^{\infty} \mathbb{E}\left[\theta_{t,\ell}^2\right] + \frac{1}{T^2}\sum_{t\neq s=1}^{T}\sum_{\ell=1}^{\infty} \mathbb{E}\left[\theta_{t,\ell}\theta_{s,\ell}\right] \\
&= O\left(T^{-1}\right) + \frac{1}{T^2}\sum_{t\neq s=1}^{T}\sum_{\ell=1}^{\infty} \mathbb{E}\left[\theta_{t,\ell}\theta_{s,\ell}\right],
\end{aligned}$$

where the last inequality comes from the fact that $X_t \in L_H^2$ which in turn implies $\sum_{\ell=1}^{\infty} \lambda_\ell < \infty$. Further, invoking Assumption 1 gives

$$\begin{aligned}
\left|\frac{1}{T^2}\sum_{t\neq s=1}^{T}\sum_{\ell=1}^{\infty} \mathbb{E}\left[\theta_{t,\ell}\theta_{s,\ell}\right]\right| &= \left|\frac{2}{T^2}\sum_{h=1}^{T-1}\sum_{t=h+1}^{T}\sum_{\ell=1}^{\infty} \mathbb{E}\left[\theta_{t,\ell}, \theta_{t-h,\ell}\right]\right| \\
&\le \frac{2}{T^2}\sum_{h=1}^{T-1}\sum_{t=h+1}^{T}\sum_{\ell=1}^{\infty} \left|B_{\ell,\ell}^{(h)}\right| \\
&\le \frac{C}{T^2}\sum_{h=1}^{T-1}(T-h)h^{-\beta}\sum_{\ell=1}^{\infty} \lambda_\ell = O\left(\max\left\{T^{-\beta}, T^{-1}\right\}\right).
\end{aligned}$$

This concludes the first part of the lemma. The proof for the lim inf follows identical lines as above where inequalities are replaced with "$\sim$". $\square$

Proof of Theorem 1. First note that since $\|\Phi_i\|_{\mathcal{L}} \sim i^{-(1-d)}$ then $\|\Phi_i\|_{\mathcal{L}}^2$ are summable across i, i.e.,

$$\sum_{i=0}^{\infty} \|\Phi_i\|_{\mathcal{L}}^2 < \infty. \tag{A.4}$$

Second, $\sigma_T^2 := \mathbb{E}\,\|TS_T\|^2$ diverge as $T \to \infty$. To see this fact consider $\mathbb{E}\,\|S_T\|^2$

$$\mathbb{E}\,\|S_T\|^2 = \frac{1}{T}\sum_{s,t=1}^{T}\int_0^1\int_0^1 \mathbb{E}\left[X_t(u)X_s(v)\right]\mathrm{d}u\mathrm{d}v = \frac{1}{T^2}\sum_{s,t=1}^{T}\int_0^1\int_0^1 c_{t-s}(u,v)\mathrm{d}u\mathrm{d}v$$
$$= T^{-(1-2d)}\int_0^1\int_0^1\sum_{h=-T}^{T}\left(1-\frac{|h|}{T}\right)\frac{c_{t-s}(u,v)}{T^{2d}}\mathrm{d}u\mathrm{d}v.$$

It is straightforward to see that the summands in the last equality converge to the kernel of the long run covariance operator $\gamma_{1-2d}(u,v)$ as $T \to \infty$ (see Section 2.1 for the definition of $\gamma_{1-2d}(u,v)$). That is, $\mathbb{E}\,\|S_T\|^2 \sim T^{-(1-2d)}\int_0^1\int_0^1 \gamma_{1-2d}(u,v)\mathrm{d}u\mathrm{d}v$, which gives us

$$\mathbb{E}\,\|TS_T\|^2 \sim T^{1+2d}\int_0^1\int_0^1 \gamma_{1-2d}(u,v)\mathrm{d}u\mathrm{d}v. \tag{A.5}$$

Third,

$$\sup_T \frac{\sum_{i=-\infty}^{\infty}\left\|\sum_{t=1}^{T}\Phi_{t-i}\right\|_{\mathcal{L}}^2}{\sigma_T^2} < \infty. \tag{A.6}$$

This fact follows from $\|\Phi_i\|_{\mathcal{L}} \sim i^{-(1-d)}$ and (A.5), i.e.,

$$\sup_T \frac{\sum_{i=-\infty}^{\infty}\left\|\sum_{t=1}^{T}\Phi_{t-i}\right\|_{\mathcal{L}}^2}{\sigma_T^2} \le \sup_T \frac{\sum_{i=-\infty}^{\infty}\sum_{t=1}^{T}\|\Phi_{t-i}\|_{\mathcal{L}}^2}{T^{1+2d}\int_0^1\int_0^1\gamma_{1-2d}(u,v)\mathrm{d}u\mathrm{d}v}$$
$$= \frac{T\sum_{i=-\infty}^{\infty}\|\Phi_i\|_{\mathcal{L}}^2}{T^{1+2d}\int_0^1\int_0^1\gamma_{1-2d}(u,v)\mathrm{d}u\mathrm{d}v} < \infty,$$

since $0 < \int_0^1\int_0^1 \gamma_{1-2d}(u,v)\mathrm{d}u\mathrm{d}v < \infty$, $\sum_{i=-\infty}^{\infty}\|\Phi_i\|_{\mathcal{L}}^2 < \infty$ by (A.4) and $T^{-2d} \to 0$ as $T \to \infty$. Finally, let $\{e_l\}_{l=1}^{\infty}$ be an orthonormal basis of H and

$$\sigma_{l,m} = \lim_{T\to\infty}\frac{\mathbb{E}\left(\langle TS_T, e_l\rangle\langle TS_T, e_m\rangle\right)}{\sigma_T^2} \quad \text{for } l, m \ge 1. \tag{A.7}$$

If the limit in (A.7) exist than the following holds trivially

$$\sum_{l=1}^{\infty}\sigma_{l,l} = 1, \tag{A.8}$$

since $\mathbb{E}\,\|TS_T\|^2 = \sum_{l=1}^{\infty}\mathbb{E}\left(\langle TS_T, e_l\rangle\langle TS_T, e_l\rangle\right)$ for any full orthonormal basis of H. To show the existence of $\sigma_{l,m}$ we use the same arguments as in for (A.5)

$$\frac{\mathbb{E}\left(\langle TS_T, e_l\rangle\langle TS_T, e_m\rangle\right)}{\sigma_T^2} = \frac{T^{1+2d}\int_0^1\int_0^1\sum_{h=-T}^{T}\left(1-\frac{|h|}{T}\right)\frac{c_{t-s}(u,v)}{T^{2d}}e_l(u)e_m(u)\mathrm{d}u\mathrm{d}v}{T^{1+2d}\int_0^1\int_0^1\sum_{h=-T}^{T}\left(1-\frac{|h|}{T}\right)\frac{c_{t-s}(u,v)}{T^{2d}}\mathrm{d}u\mathrm{d}v}.$$

Since $\sum_{h=-T}^{T}\left(1-\frac{|h|}{T}\right)\frac{c_{t-s}(u,v)}{T^{2d}}$ converge to $\gamma_{1-2d}(u,v)$ and $e_l(u)$, $e_m(v)$ are orthogonal we have

$$\sigma_{l,m} \longrightarrow \frac{\int_0^1\int_0^1\gamma_{1-2d}(u,v)e_l(u)e_m(u)\mathrm{d}u\mathrm{d}v}{\int_0^1\int_0^1\gamma_{1-2d}(u,v)\mathrm{d}u\mathrm{d}v} \quad \text{as } T \to \infty$$

and $|\sigma_{l,m}| < \infty$.

Then putting together (A.4)–(A.8) allows us to make a use of Corollary 1.1 in Merlevède (1996), which concludes this proof. □

Proof of Theorem 2. First note, since $X_t \in L_H^2$ then $\sum_{\ell=1}^{\infty}\lambda_\ell < \infty$ and $\sum_{\ell=1}^{\infty}\lambda_\ell^2 < \infty$. Then,

$$\mathbb{E}\left\|\widehat{C}_T - C\right\|_{\mathcal{S}}^2 = \sum_{\ell=1}^{\infty}\mathbb{E}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\langle X_t, \psi_\ell\rangle X_t - \mathbb{E}\left[\langle X_t, \psi_\ell\rangle X_t\right]\right)\right\|^2$$
$$= \frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{\ell=1}^{\infty}\left(\sum_{m=1}^{\infty}\mathbb{E}\left[\theta_{t,\ell}\theta_{s,\ell}\theta_{t,m}\theta_{s,m}\right] - \lambda_\ell^2\right) \tag{A.9}$$

$$= \frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{\ell=1}^{\infty}\left(\mathbb{E}\left[\theta_{t,\ell}^2\theta_{s,\ell}^2\right]-\lambda_\ell^2\right)$$
$$+ \frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{\ell\neq m=1}^{\infty}\mathbb{E}\left[\theta_{t,\ell}\theta_{s,\ell}\theta_{t,m}\theta_{s,m}\right] := a_T + b_T. \quad (A.10)$$

It follows from relation (A.1) that

$$a_T = \frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{\ell=1}^{\infty}\left(\kappa_{\ell,\ell,\ell,\ell}(0,|t-s|,|t-s|) + 2\mathbb{E}\left[\theta_{t,\ell}\theta_{s,\ell}\right]^2\right),$$

where $\frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{\ell=1}^{\infty}\kappa_{\ell,\ell,\ell,\ell}(0,|t-s|,|t-s|) = O\left(T^{\rho-1}\right)$ by Assumption 2 and

$$\begin{aligned}
\frac{2}{T^2}\sum_{t,s=1}^{T}\sum_{\ell=1}^{\infty}\mathbb{E}\left[\theta_{t,\ell}\theta_{s,\ell}\right]^2 &= \frac{2}{T^2}\sum_{t\neq s=1}^{T}\sum_{\ell=1}^{\infty}\mathbb{E}\left[\theta_{t,\ell}\theta_{s,\ell}\right]^2 + \frac{2}{T}\sum_{\ell=1}^{\infty}\lambda_\ell^2\\
&\leq \frac{4}{T^2}\sum_{h=1}^{T-1}\sum_{t=h+1}^{T}\sum_{\ell=1}^{\infty}\left(B_{\ell,\ell}^{(h)}\right)^2 + \frac{2}{T}\sum_{\ell=1}^{\infty}\lambda_\ell^2\\
&\leq \frac{B}{T}\sum_{h=1}^{T-1}h^{-2\beta}\sum_{\ell=1}^{\infty}\lambda_\ell^2 + \frac{2}{T}\sum_{\ell=1}^{\infty}\lambda_\ell^2\\
&= O\left(\max\left\{T^{-2\beta}, T^{-1}\right\}\right).
\end{aligned}$$

Similar arguments apply to term b_T, i.e.,

$$\frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{\ell\neq m=1}^{\infty}\mathbb{E}\left[\theta_{t,\ell}\theta_{s,\ell}\theta_{t,m}\theta_{s,m}\right] = \frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{\ell\neq m=1}^{\infty}\left(\kappa_{\ell,\ell,m,m}(0,|t-s|,|t-s|)+\right.$$
$$\left.+\,\mathbb{E}\left[\theta_{t,\ell}\theta_{s,\ell}\right]\mathbb{E}\left[\theta_{t,m}\theta_{s,m}\right] + \mathbb{E}\left[\theta_{t,\ell}\theta_{s,m}\right]\mathbb{E}\left[\theta_{t,m}\theta_{s,\ell}\right]\right) \quad (A.11)$$

by relation (A.1). The first terms on the r.h.s of (A.11) is $O\left(T^{\rho-1}\right)$ by Assumption 2. The second and the third terms on the r.h.s of (A.11) are $O\left(\max\{T^{-2\beta}, T^{-1}\}\right)$ by the same arguments as above. In particular, for the third term we have

$$\begin{aligned}
\frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{\ell\neq m=1}^{\infty}\mathbb{E}\left[\theta_{t,\ell}\theta_{s,m}\right]\mathbb{E}\left[\theta_{t,m}\theta_{s,\ell}\right] &\leq \frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{\ell\neq m=1}^{\infty}\left(B_{\ell,m}^{(t-s)}\right)^2 = \frac{2}{T^2}\sum_{h=1}^{T-1}\sum_{t=h+1}^{T}\sum_{\ell\neq m=1}^{\infty}\left(B_{\ell,m}^{(h)}\right)^2\\
&\leq \frac{B}{T}\sum_{h=1}^{T-1}h^{-2\beta}\sum_{\ell\neq m=1}^{\infty}\lambda_\ell\lambda_m = O\left(\max\left\{T^{-2\beta}, T^{-1}\right\}\right).
\end{aligned}$$

Putting together rates for a_T and b_T yield the first part of statement of the theorem. The proof for the lim inf follows identical lines as above, where inequalities are replaced with "$\sim$" □

Proof of Theorem 3. From (1) and arguments in Giraitis et al. (2003, proof of Theorem 3.1) it follows directly that

$$\int_0^1\int_0^1 \frac{1}{q^{1-\beta}}\mathbb{E}[\widehat{\gamma}_T(u,v)] - \gamma_\beta(u,v)\mathrm{d}u\mathrm{d}v \to 0.$$

Hence, to proof the statement of the theorem it suffices to study the limit

$$\lim_{T\to\infty,q\to\infty}\max\left\{\frac{T^{2\beta}}{q^2},\frac{T^{1-\rho}}{q}\right\}\mathbb{E}\int_0^1\int_0^1\left(\widehat{\gamma}_T(u,v)-\mathbb{E}[\widehat{\gamma}_T(u,v)]\right)^2\mathrm{d}u\mathrm{d}v.$$

Denote $\widetilde{B}_{\ell,k}^{(h)} = \frac{1}{T}\sum_{t=h+1}^{T}\theta_{t,\ell}\theta_{t-h,k}$. Then

$$\mathbb{E}\int_0^1\int_0^1\left(\widehat{\gamma}_T(u,v)-\mathbb{E}[\widehat{\gamma}_T(u,v)]\right)^2\mathrm{d}u\mathrm{d}v = \sum_{\ell,k}\mathbb{E}\left[\sum_{|h|\leq q}w_h(q)\left(\widetilde{B}_{\ell,k}^{(h)} - B_{\ell,k}^{(h)}\right)\right]^2$$
$$= \sum_{\ell,k}\sum_{|h|,|i|\leq q}\left(1-\frac{|h|}{q+1}\right)\left(1-\frac{|i|}{q+1}\right)Cov\left(\widetilde{B}_{\ell,k}^{(h)},\widetilde{B}_{\ell,k}^{(i)}\right) \quad (A.12)$$

where the equality follows from Eq. (1). Using Eq. (A.1) $Cov\left(\widetilde{B}^{(h)}_{\ell,k}, \widetilde{B}^{(i)}_{\ell,k}\right)$ can be rewritten as sum of two terms

$$\begin{aligned} Cov\left(\widetilde{B}^{(h)}_{\ell,k}, \widetilde{B}^{(i)}_{\ell,k}\right) &= \frac{1}{T^2}\sum_{t=1}^{T-|h|}\sum_{s=1}^{T-|i|} Cov\left(\theta_{t,\ell}\theta_{t+|h|,k}, \theta_{s,\ell}\theta_{s+|i|,k}\right) \\ &= \frac{1}{T^2}\sum_{t=1}^{T-|h|}\sum_{s=1}^{T-|i|}\left(B^{t-s}_{\ell,\ell}B^{t-s+|h|-|i|}_{k,k} + B^{t-s-|i|}_{\ell,k}B^{t-s+|h|}_{k,\ell}\right) \\ &+ \frac{1}{T^2}\sum_{t=1}^{T-|h|}\sum_{s=1}^{T-|i|}\kappa_{\ell,k,\ell,k}(|h|, s-t, s-t+|i|) := a_T + b_T. \end{aligned}$$

Then (A.12) can also be split into two parts as

$$\begin{aligned} \mathbb{E}\int_0^1\int_0^1 (\widehat{\gamma}_T(u,v) - \mathbb{E}[\widehat{\gamma}_T(u,v)])^2\,\mathrm{d}u\mathrm{d}v &= \sum_{\ell,k}\sum_{|h|,|i|\le q}\left(1-\frac{|h|}{q+1}\right)\left(1-\frac{|i|}{q+1}\right)a_T \\ &+ \sum_{\ell,k}\sum_{|h|,|i|\le q}\left(1-\frac{|h|}{q+1}\right)\left(1-\frac{|i|}{q+1}\right)b_T \\ &:= A_T + B_T. \end{aligned}$$

Using Assumption 1 we have that a_T converges to the following integral

$$\begin{aligned} \frac{T^{2\beta}}{\lambda_\ell\lambda_k}a_T &\sim K\frac{1}{T^2}\sum_{t=1}^{T-|h|}\sum_{s=1}^{T-|i|}\left(\left|\frac{t-s}{T}\right|^{-\beta}\left|\frac{t-s-|i|+|h|}{T}\right|^{-\beta} + \left|\frac{t-s-|i|}{T}\right|^{-\beta}\left|\frac{t-s+|h|}{T}\right|^{-\beta}\right) \\ &\underset{T\to\infty}{\to} 2K\int_0^1\int_0^1 |t-s|^{-2\beta}\,\mathrm{d}t\mathrm{d}s = \frac{4K}{(1-2\beta)(2-2\beta)} < \infty, \end{aligned}$$

where the last limit does not depend on h and i since $|i|, |h| < q$ and $q/T \to 0$. Hence, for A_T we have

$$\begin{aligned} \lim_{T\to\infty, q\to\infty}\frac{T^{2\beta}}{q^2}A_T &= \lim_{q\to\infty}\frac{1}{q^2}\sum_{|h|,|i|\le q}\left(1-\frac{|h|}{q+1}\right)\left(1-\frac{|i|}{q+1}\right)\frac{4K\sum_{\ell,k}\lambda_\ell\lambda_k}{(1-\beta)(2-\beta)} \\ &\underset{T\to\infty}{\to} \frac{4K\sum_{\ell,k}\lambda_\ell\lambda_k}{(1-\beta)(2-\beta)}\left(\int_0^1 (1-t)\,\mathrm{d}t\right)^2 < \infty. \end{aligned}$$

Using similar arguments the rate of convergence is derived for B_T using Assumption 2′, i.e.,

$$\begin{aligned} &\frac{1}{T^2}\sum_{\ell,k}\sum_{|h|\le q}\left(\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{|i|\le q}\kappa_{\ell,k,\ell,k}(|h|, s-t, s-t+|i|)\right) \\ &\sim \frac{1}{T}\sum_{\ell,k}\sum_{|h|\le q}\left(\sum_{|t|,|i|\le 2T}\kappa_{\ell,k,\ell,k}(|h|, t, i)\right) \\ &\sim K\frac{q}{T^{1-\rho}}\sum_{\ell,k}\lambda_\ell\lambda_k, \end{aligned}$$

where $\sum_{\ell,k}\lambda_\ell\lambda_k < \infty$. This concludes the proof of the theorem. □

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2019.03.007.

References

Anderson, T., 1971. The Statistical Analysis of Time Series. John Wiley and Sons, New York.
Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.
Antoniadis, A., Paparoditis, E., Sapatinas, T., 2006. A functional wavelet–kernel approach for time series prediction. J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (5), 837–857.
Aue, A., Norinho, D.D., Hörmann, S., 2015. On the prediction of stationary functional time series. J. Amer. Statist. Assoc. 110 (509), 378–392.
Beran, J., 1994. Statistics for Long Meory Processes. Chapman and Hall.
Berkes, I., Horváth, L., Rice, G., 2013. Weak invariance principles for sums of dependent random functions. Stochastic Process. Appl. 123 (2), 385–403.

Berkes, I., Horváth, L., Rice, G., 2016. On the asymptotic normality of kernel estimators of the long run covariance of functional time series. J. Multivariate Anal. 144, 150–175.
Besse, P.C., Cardot, H., Stephenson, D.B., 2000. Autoregressive forecasting of some functional climatic variations. Scand. J. Stat. 27 (4), 673–687.
Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York.
Bosq, D., 2000. Linear Processes in Function Spaces. Springer, New York.
Brillinger, D.R., 2001. Time Series: Data Analysis and Theory. Society for Industrial and Applied Mathematics, Philadelphia.
Cardot, H., Ferraty, F., Sarda, P., 1999. Functional linear model. Statist. Probab. Lett. 45, 11–22.
Chang, Y., Kim, C.S., Park, J.Y., 2016. Nonstationarity in time series of state densities. J. Econometrics 192 (1), 152–167.
Chen, X., White, H., 1998. Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications. Econometric Theory 14 (2), 260–284.
Dauxois, J., Pousse, A., Romain, Y., 1982. Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. J. Multivariate Anal. 12 (1), 136–154.
Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press, Oxford.
Demetrescu, M., Kuzin, V., Hassler, U., 2008. Long memory testing in the time domain. Econometric Theory 24 (1), 176–215.
Giraitis, L., Kokoszka, P., Leipus, R., Teyssière, G., 2003. Rescaled variance and related tests for long memory in volatility and levels. J. Econometrics 112 (2), 265–294.
Gonçalves, S., Kilian, L., 2007. Asymptotic and bootstrap inference for AR(∞) processes with conditional heteroskedasticity. Econometric Rev. 26 (6), 609–641.
Hall, P., Horowitz, J.L., 2007. Methodology and convergence rates for functional linear regression. Ann. Statist. 35 (1), 70–91.
Herrndorf, N., 1984. A functional central limit theorem for weakly dependent sequences of random variables. Ann. Probab. 12 (1), 141–153.
Hörmann, S., Kidziński, Ł., 2015. A note on estimation in Hilbertian linear models. Scand. J. Stat. 42 (1), 43–62.
Hörmann, S., Kidziński, Ł., Hallin, M., 2015. Dynamic functional principal components. J. R. Stat. Soc. Ser. B Stat. Methodol. 77 (2), 319–348.
Hörmann, S., Kokoszka, P., 2010. Weakly dependend functional data. Ann. Statist. 38, 1845–1884.
Horváth, L., Kokoszka, P., 2012. Inference for Functional Data with Applications. Springer, New York.
Horváth, L., Kokoszka, P., Reeder, R., 2013. Estimation of the mean of functional time series and a two-sample problem. J. R. Stat. Soc. Ser. B Stat. Methodol. 75 (1), 103–122.
Horváth, L., Kokoszka, P., Rice, G., 2014. Testing stationarity of functional time series. J. Econometrics 179 (1), 66–82.
Hosking, J.R., 1996. Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. J. Econometrics 73 (1), 261–284.
Hsing, T., Eubank, R., 2015. Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. In: Wiley Series in Probability and Statistics, Wiley.
Ibragimov, I., Linnik, Y., 1971. Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, Groningen.
Karhunen, K., 1947. Über Lineare Methoden in der Wahrscheinlichkeitsrechnung. Ann. Acad. Sci. Fenn. Math. 37, 1–79.
Kokoszka, P., Reimherr, M., 2017. Introduction to Functional Data Analysis. Chapman and Hall/CRC.
Kwiatkowski, D., Phillips, P.C., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. J. Econometrics 54 (1), 159–178.
Lee, D., Schmidt, P., 1996. On the power of the KPSS test of stationarity against fractionally-integrated alternatives. J. Econometrics 73, 285–302.
Loève, M., 1946. Fonctions aléatoires du second ordre. Rev. Sci. 84, 195–206.
Merlevède, F., 1996. Central limit theorem for linear processes with values in a Hilbert space. Stochastic Process. Appl. 65 (1), 103–114.
Merlevède, F., Peligrad, M., Utev, S., 1997. Sharp conditions for the CLT of linear processes in a Hilbert space. J. Theoret. Probab. 10 (3), 681–693.
Müller, H.-G., Yao, F., 2008. Functional additive models. J. Amer. Statist. Assoc. 103, 1534–1544.
Panaretos, V.M., Tavakoli, S., 2013a. Cramer-Karhunen-Loeve representation and harmonic principal component analysis of functional time series. Stoch. Process. Appl. 123 (7), 2779–2807.
Panaretos, V.M., Tavakoli, S., 2013b. Fourier analysis of stationary time series in function space. Ann. Statist. 41 (2), 568–603.
Park, J.Y., Qian, J., 2012. Functional regression of continuous state distributions. J. Econometrics 167 (2), 397–412.
Phillips, P.C.B., 1987. Time series regression with a unit root. Econometrica 55 (2), 277–301.
Račkauskas, A., Suquet, C., 2010. On limit theorems for banach-space-valued linear processes. Lith. Math. J. 50 (1), 71–87.
Račkauskas, A., Suquet, C., 2011. Operator fractional Brownian motion as limit of polygonal lines processes in Hilbert space. Stoch. Dyn. 11 (1), 49–70.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis, second ed. Springer, New York.
Reiss, P.T., Ogden, R.T., 2007. Functional principal component regression and functional partial least squares. J. Amer. Statist. Assoc. 102 (479), 984–996.