Bernoulli **26**(3), 2020, 2383–2399
https://doi.org/10.3150/20-BEJ1199

Frequency domain theory for functional time series: Variance decomposition and an invariance principle

PIOTR KOKOSZKA[1] and NEDA MOHAMMADI JOUZDANI[2]

[1]*Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA. E-mail: Piotr.Kokoszka@colostate.edu*
[2]*Department of Mathematics, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland. E-mail: neda.mohammadijouzdani@epfl.ch*

This paper is concerned with frequency domain theory for functional time series, which are temporally dependent sequences of functions in a Hilbert space. We consider a variance decomposition, which is more suitable for such a data structure than the variance decomposition based on the Karhunen–Loéve expansion. The decomposition we study uses eigenvalues of spectral density operators, which are functional analogs of the spectral density of a stationary scalar time series. We propose estimators of the variance components and derive convergence rates for their mean square error as well as their asymptotic normality. The latter is derived from a frequency domain invariance principle for the estimators of the spectral density operators. This principle is established for a broad class of linear time series models. It is a main contribution of the paper.

Keywords: functional data; invariance principle; spectral analysis; time series; variance decomposition

1. Introduction

Suppose $\{X_t\}$ is a weakly dependent stationary mean zero functional time series. Precise definitions will be given in Section 2. A major tool of Functional Data Analysis (FDA) is the Karhunen-Loéve expansion

$$X_t(u) = \sum_{j=1}^{\infty} \xi_{tj} v_j(u), \qquad E\xi_{tj}^2 = \lambda_j, \tag{1.1}$$

where the functions v_j are the functional principal components (FPCs), and the scalars λ_j are the variance components. Expansion (1.1) and its applications are treated in detail in several monographs, see, for example, Chapter 9 of [22] and Chapter 11 of [27] for introductions, and most Chapters of [7] and [20] for applications. Stationary time series in Hilbert spaces, including aspects of their spectral theory relevant to prediction, are discussed in depth in Chapters 9 and 10 of [36]. The variance decomposition

$$E\|X_t\|^2 = \sum_{j=1}^{\infty} \lambda_j \tag{1.2}$$

1350-7265

has played a fundamental role in FDA. At the most basic level, estimates of the λ_j are used to determine the number of FPCs to be used in dimension reduction through the percentage of explained variance criterion. These estimates also enter into most statistics based on FPCs. While decomposition (1.2) is valid for any stationary functional time series with $E\|X_0\|^2 < \infty$, the functions v_j do not reflect in any way the dependence structure of the time series $\{X_t\}$. If $\{X_t'\}$ is a sequence of i.i.d. functions such that for each t, X_t' has the same covariance operator as X_t, then the series $\{X_t\}$ and $\{X_t'\}$ have the same decompositions (1.1) and (1.2).

To provide a more efficient dimension reduction technique for functional time series, Panaretos and Tavakoli [33] and Hörmann et al. [15] developed theory and methodology for what we call *dynamic* FPCs (DFPCs). Extending the multivariate setting developed in Chapter 9 of [8], they showed that the DFPCs are more suitable to reduce the dimension of functional time series than the usual FPCs. The DFPCs are defined in Section 2, see equations (2.11) and (2.12). They have recently been used by Górecki et al. [13] to develop an effective normality test, and extended to periodically correlated functional time series by Kidzinski et al. [24]. Using approaches based on DFPCs, Hörmann et al. [16] and Pham and Panaretos [35] considered the problem of estimation in a functional regression with dependent error functions.

There is however at present basically no work on an analog of the variance decomposition (1.2) in the context of DFPCs. Such a decomposition will play a similar role for temporally dependent functions X_t as the decomposition (1.2) has played for i.i.d. functional samples. The decomposition derived in Section 2 is implicit in the work of Hörmann et al. [16], but only the consistency of the estimators could be inferred from their work. The original question that motivated this research was to understand second order properties of the estimators of the variance components based on the DFPCs. It turns out that to gain such an understanding, a fairly profound investigation of the foundations of frequency domain estimation of functional time series is needed. We will first explain the motivating questions, and then comment on the chief results presented in this work. Due to a fairly complex structure of the objects we study, we can present only the highlights in the introduction.

In the context of expansion (1.1), if $E\|X_t\|^4 < \infty$ and some simple assumptions hold, then for each $j \geq 1$,

$$\limsup_{N\to\infty} NE(\widehat{\lambda}_j - \lambda_j)^2 < \infty \tag{1.3}$$

and

$$N^{1/2}(\widehat{\lambda}_j - \lambda_j) \xrightarrow{d} N(0, \sigma_j^2), \tag{1.4}$$

where the hat indicates suitable estimators, which are generally the eigenvalues of the sample covariance operator, but other estimators have also been considered, see, for example, Bosq [7]. Relation (1.3) was established by Dauxois et al. [10] under independence, and extended to weakly dependent functional time series by Hörmann and Kokoszka [17]. Relation (1.4) can be derived from the results of Kokoszka and Reimherr [26]. In case of continuous functions satisfying regularity conditions, it follows from the results of Hall and Hosseini-Nasab [14]. Our first objective is to derive analogs of (1.3) and (1.4) in the contexts of dynamic FPCs of Hörmann [15]. We will introduce the variance components Λ_j (see (2.10) and (2.14)) and their estimators $\widehat{\Lambda}_j$ (see (3.1)). The questions are: At what rate does $E(\widehat{\Lambda}_j - \Lambda_j)^2$ tend to zero? Is it true that under a suitable normalization, the difference $\widehat{\Lambda}_j - \Lambda_j$ is asymptotically normal. We will see that the standard rates in (1.3) and (1.4) no longer hold. The second question is particularly delicate. The estimator $\widehat{\Lambda}_j$ is a functional of the spectral density operator viewed as a process on $(-\pi, \pi]$, and to establish the asymptotic normality of $\widehat{\Lambda}_j$ an invariance principle for the process of estimated spectral density operators is needed. We use the term "invariance principle" to emphasize convergence in the metric space of operator-valued functions on $(-\pi, \pi]$ rather than at a set of fixed frequencies. It is not a Donsker's theorem type result for the partial sum process. The invariance principle we prove extends classical results on the normality of spectral density estimators, explained for example, in [1], in two directions. First, we establish an invariance principle, rather than pointwise convergence at a fixed frequency. Second, we do it in the setting of a separable Hilbert space rather than a scalar time series. Combining these two directions leads to the challenge of dealing with tightness in an infinitely dimensional space under dependence in fairly complex inferential problem.

The spectral density operator was introduced by Panaretos and Tavakoli [34], and has been used in several applications, to the papers listed above, we can add [2,19,29] and [38]. It has been known since the work of Kuelbs [28] that a central limit theorem in a Hilbert space implies an invariance principle, but we do not use this result; the structure of the process we consider is different. Influential papers on time domains central limit theorem in a Hilbert space are [32] and [31]. A time domain invariance principle for approximable, a type of weak dependence condition we use as well, functional time series was established by Berkes et al. [5].

The paper is organized as follows. The mathematical framework and general assumptions are introduced in Section 2, where the frequency domain variance decomposition is also specified. Section 3 is dedicated to the study of the asymptotics of the MSE the estimators $\widehat{\Lambda}_j$. Their asymptotic normality is established in Section 4 as a consequence of the invariance principle discussed above. All proofs, except simple, illustrative arguments, are presented in the Supplementary Material [25]. Proofs of the results stated in Sections 2 and 3 are presented in Section A. Additional results related to Theorem 3.1 are presented in Section B. Proofs of the results stated in Section 4 are presented in Section C. The main part of the paper is self-contained, but occasional references to results and formulas given in the supplement are given to help the reader navigate it. These references begin with letters A, B or C.

2. Preliminaries

Suppose $\{X_t, t \in \mathbb{Z}\}$ is a stationary, weakly dependent, mean zero time series of functions in $\mathcal{H} := L^2([0, 1])$. The interval $[0, 1]$ with the Lebesgue measure is used for conceptual guidance, and can be replaced in all formulas by a measure space with total measure equal to 1. To emphasize it, we write $\int$ in place of $\int_0^1$. The Hilbert space $\mathcal{H}$ of square integrable functions is equipped with the usual inner product

$$\langle f, g \rangle_{\mathcal{H}} = \int f(u) g(u)\, du, \qquad f, g \in L^2[0, 1],$$

and the corresponding norm $\|\cdot\|_{\mathcal{H}}$. The notation $\mathcal{L}$ is used to denote the space of bounded linear operators on $\mathcal{H}$. The subspaces of Hilbert-Schmidt operators and nuclear operators are denoted by $\mathcal{S}$ and $\mathcal{N}$, respectively, see, for example, [7]. We use these as subscripts to distinguish the corresponding norms or inner products. An integral operator $K \in \mathcal{S}$ is an operator which admits the representation

$$K(f)(u) = \int k(u, v) f(v)\, dv, \quad f \in L^2[0, 1],$$

with $k(u, v) \in L^2([0, 1] \times [0, 1])$, that is,

$$\iint \left| k(u, v) \right|^2 du\, dv = \|k\|_2^2 = \|K\|_{\mathcal{S}}^2 < \infty.$$

For any $f, g \in \mathcal{H}$, $f \otimes g : \mathcal{H} \to \mathcal{H}$ denotes a bounded linear operator defined by $f \otimes g : h \mapsto \langle h, g \rangle_{\mathcal{H}} f$. We see that $f \otimes g$ is an integral operator with the kernel $f(\cdot)g(\cdot)$. Notation $\otimes_{\mathcal{S}}$ indicates a similar concept for operators on the space $\mathcal{S}$. Notation $L^p_{\mathcal{H}}(\Omega, \mathcal{A}, \mathbb{P})$ is used to denote the space of $\mathcal{H}$-valued random functions defined on the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with finite pth moment, that is, the random function X belongs to the space $L^p_{\mathcal{H}}(\Omega, \mathcal{A}, \mathbb{P})$ if and only if $\mu_p(X) = (E\|X\|^p)^{\frac{1}{p}} < \infty$.

We assume that $\{X_t, t \in \mathbb{Z}\}$ satisfies $\mu_2(X_t) < \infty$, $EX_t = 0$, and is second–order stationary, that is, the lag–h autocovariance kernel $c_h(u, v) := E[X_{t+h}(u)X_t(v)]$ does not depend on t. The condition

$$\sum_{h \in \mathbb{Z}} \|c_h\|_2 < \infty \tag{2.1}$$

guarantees the convergence, in $\|\cdot\|_2$, of the series

$$f_\theta(u, v) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} c_h(u, v) e^{-ih\theta}, \quad u, v \in [0, 1].$$

The kernels $f_\theta(\cdot, \cdot), \theta \in (-\pi, \pi]$, called the spectral density kernels, define integral Hilbert-Schmidt operators $\mathcal{F}_\theta$, which are called the spectral density operators. *In the remainder of the paper, we assume that* (2.1) *holds, and so the spectral density operators exist.*

We now define estimators $\hat{f}_\theta(u, v)$, which are kernels of the estimators $\widehat{\mathcal{F}}_\theta$. For fixed $u, v \in [0, 1]$ and series length N, we estimate $c_h(u, v)$ by

$$\hat{c}_h(u, v) = \frac{1}{N} \sum_{k=1}^{N-h} X_{h+k}(u) X_k(v), \quad h \geq 0; \qquad \hat{c}_h(u, v) = \frac{1}{N} \sum_{k=1}^{N-|h|} X_k(u) X_{k+|h|}(v), \quad h < 0,$$

which reduces to

$$\hat{c}_h(u, v) = \frac{1}{N} \sum_{k=1}^{N} X_{h+k}(u) X_k(v), \tag{2.2}$$

by setting the terms with impossible subscripts to zero. We use the following estimator of the spectral density kernel $f_\theta(u, v)$:

$$\hat{f}_\theta(u, v) = \frac{1}{2\pi} \sum_{|h| \le q} w\left(\frac{h}{q}\right) \hat{c}_h(u, v) e^{-ih\theta}, \tag{2.3}$$

where $q = q(N)$ is a bandwidth function. To lighten the notation, the argument N is often suppressed. We impose the following assumption on the bandwidth function q and the weight function w.

Assumption 2.1. The following conditions hold as $N \to \infty$:

$$q(N) = o(N), \qquad q(N) \to \infty$$

and for each fixed h,

$$w\left(\frac{h}{q}\right) \to 1.$$

The weight function $w(\cdot)$ is even, i.e. $w(s) = w(-s)$, continuous on $[-1, 1]$ and hence bounded, that is,

$$\forall s \in [-1, 1], \qquad \left|w(s)\right| \le b \quad \text{for some } b. \tag{2.4}$$

Continuity of $w(\cdot)$ is needed to establish Lemma 4.3, other lemmas require only (2.4).

Next, we define L^p-approximable sequences, which form a large subclass of strictly stationary weakly dependent functional time series. (A term that has traditionally been used is "L^p–m-approximable", but since the m is not part of the definition, we use a simpler term.)

Definition 2.1. A functional time series $\{X_t\}$ is said to be L^p-approximable if it takes values in $L^p_{\mathcal{H}}(\Omega, \mathcal{A}, \mathbb{P})$ and admits the representation

$$X_t = f(\varepsilon_t, \varepsilon_{t-1}, \dots),$$

where $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a sequence of i.i.d. random elements taking values in some measurable space S, and f is a measurable function from S^∞ to $\mathcal{H}$. Moreover, if the sequence $\{\varepsilon'_t, t \in \mathbb{Z}\}$ is an independent copy of $\{\varepsilon_t, t \in \mathbb{Z}\}$, then

$$\sum_{m=1}^{\infty} \mu_p\big(X_m - X_m^{(m)}\big) < \infty, \tag{2.5}$$

where $X_t^{(m)}$ is defined as

$$X_t^{(m)} = f\big(\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-m+1}, \varepsilon'_{t-m}, \varepsilon'_{t-m-1}, \dots\big).$$

In the proofs, we choose a different independent copy for each truncation level m. The essence of Definition 2.1 is that the impact of innovations ε_t far back in the past becomes negligible; they can be replaced by independent copies, and the effect of this replacement is quantified by (2.5). Conditions in Definition 2.1 have been shown to hold for all known stationary models for temporally dependent functions, assuming the parameters of these models satisfy nonrestrictive conditions, see [17,18] or Chapter 16 of [20]. For scalar time series, conditions similar in spirit were used by Shao and Wu [37], Berkes et al. [4], Zhou [40], and for functional time series by Horváthet al. [21], Zhang [39], Bardsley et al. [3], to name just a few references.

The general assumption of approximability specified in Definition 2.1 is however not sufficient to establish asymptotic properties of frequency domain functional principal component analysis. Even in the case of scalar time series, additional assumptions are needed. They are relatively simple for linear processes (absolute summability of impulse response coefficients), for example, Chapter 7 of [9], but become more complicated for more general models (conditions on fourth order cumulants), for example, Chapter 8 of [1]. In our setting, which involves both general weak dependence and an infinite directional variance decomposition, we impose the following assumption.

Assumption 2.2. We assume that the functional time series $\{X_t\}$ is L^4-approximable in the sense of Definition 2.1 and

$$\sup_{h>0} \sum_{r=1}^{\infty} \iint_{[0,1]^2} \left|\operatorname{Cov}\left(X_0(v)\left(X_h(u) - X_h^{(h)}(u)\right), X_r^{(r)}(v) X_{r+h}^{(r+h)}(u)\right)\right| du\, dv < \infty. \tag{2.6}$$

The usual sufficient condition for the consistency of the kernel estimator of the long–run variance of a scalar time series is $\sum_h \sum_r \sum_s |\kappa(h,r,s)| < \infty$, where $\kappa(h,r,s)$ are fourth order cumulants. Condition (2.6) is, in a sense, weaker because $\sum_h$ is replaced by $\sup_h$, and there is only a single sum with respect to r. A similar modification, but using cumulants, was introduced by Giraitis et al. [12]. The general structure of condition (2.6) reflects the spirit of Definition 2.1 and the fact that autocovariances are now functions of $[0,1] \times [0,1]$. It is a very weak assumption. For linear processes, the left-hand side of (2.6) vanishes, as stated in Proposition 2.1 below, whose proof illustrates the meaning of this condition.

Proposition 2.1. *Consider the linear process*

$$X_t = \sum_{j=0}^{\infty} \Psi_j(\varepsilon_{t-j}), \tag{2.7}$$

with i.i.d. innovations $\varepsilon_i \in L^p_{\mathcal{H}}(\Omega, \mathcal{A}, \mathbb{P})$ *and* $\Psi_j \in \mathcal{L}$. *If the sequence* (2.7) *is* L^p*-approximable for some* $p \geq 2$, *then the left-hand side of* (2.6) *vanishes.*

A simple condition for the L^p-approximability of the process (2.7) , for $p \geq 2$, is

$$\sum_{m=1}^{\infty} \sum_{j=m}^{\infty} \|\Psi_j\|_{\mathcal{L}} < \infty, \tag{2.8}$$

see Proposition 16.1 of [20].

We conclude this section with results stating the decomposition of variance based on frequency domain FPCA. The jth eigenvalue and eigenfunction of the spectral density operator $\mathcal{F}_\theta$ are denoted, respectively, by $\lambda_j(\theta)$ and $\varphi_j(\theta)$. Both Theorem 2.1 and Proposition 2.2 are proven in Section A.

Theorem 2.1. *Let $\{X_t\}$ be a second order stationary process satisfying (2.1) and $\lambda_j(\theta)$ be the jth eigenvalue of the spectral density operator $\mathcal{F}_\theta$. Then*

$$E\|X_t\|^2 = \sum_{j=1}^{\infty} \Lambda_j, \tag{2.9}$$

where

$$\Lambda_j = \int_{-\pi}^{\pi} \lambda_j(\theta)\, d\theta. \tag{2.10}$$

Theorem 2.1 provides a formula for the Λ_j, which allows us to define estimators $\widehat{\Lambda}_j$ in Section 3. There is however another representation of the Λ_j. Recall that the $\varphi_j(\theta)$ are the eigenfunctions of the spectral density operator $\mathcal{F}_\theta$, and set

$$\phi_{jk} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\theta} \varphi_j(\theta)\, d\theta. \tag{2.11}$$

The functions ϕ_{jk} are the DFPCs. An analog of (1.1) is

$$X_t(u) = \sum_{j=1}^{\infty} \sum_{l\in\mathbb{Z}} Y_{j,t+l} \phi_{jl}(u). \tag{2.12}$$

The Y_{jt} are the dynamic FPC scores defined by $Y_{jt} = \sum_{l\in\mathbb{Z}} \langle X_{t-l}, \phi_{jl}\rangle$. Since $E[Y_{j,t+l} Y_{i,t+k}] = 0$ if $i \neq j$, we see that

$$E\|X_t\|^2 = \sum_{j=1}^{\infty} E\left\| \sum_{l\in\mathbb{Z}} Y_{j,t+l} \phi_{jl} \right\|^2. \tag{2.13}$$

One may thus expect that the following result holds.

Proposition 2.2. *Under the assumptions of Theorem 2.1,*

$$\Lambda_j = E\left\| \sum_{l\in\mathbb{Z}} Y_{j,t+l} \phi_{jl} \right\|^2. \tag{2.14}$$

3. Bounds on mean squared error

The λ_j in (1.2) are the eigenvalues of the covariance operator. They are typically estimated by the eigenvalues of the sample covariance operator, and are denoted $\hat{\lambda}_j$. These estimated components of variance form the basis of a large number of FDA procedures. The theoretical justification of these procedures relies on the bound $E|\hat{\lambda}_j - \lambda_j|^2 = O(N^{-1})$. Our objective in this section is to derive an analogous bound for the Λ_j defined in Theorem 2.1 and their estimators $\widehat{\Lambda}_j$ defined in the following.

Recall that the jth eigenvalue and eigenfunction of the spectral density operator $\mathcal{F}_\theta$ are denoted, respectively, by $\lambda_j(\theta)$ and $\varphi_j(\theta)$. Define $\hat{\lambda}_j(\theta)$ and $\widehat{\varphi}_j(\theta)$ to be, respectively, the jth eigenvalue and eigenfunction of the operator $\widehat{\mathcal{F}}_\theta$ with the kernel $\hat{f}_\theta(\cdot,\cdot)$ defined in (2.3). Formula (2.10) suggests the estimator

$$\widehat{\Lambda}_j := \int_{-\pi}^{\pi} \hat{\lambda}_j(\theta)\, d\theta. \tag{3.1}$$

Observe that

$$\begin{aligned} E|\Lambda_j - \widehat{\Lambda}_j|^2 &= E\left|\int_{-\pi}^{\pi} \lambda_j(\theta)\, d\theta - \int_{-\pi}^{\pi} \hat{\lambda}_j(\theta)\, d\theta\right|^2 \\ &\le E\int_{-\pi}^{\pi} \left|\lambda_j(\theta) - \hat{\lambda}_j(\theta)\right|^2 d\theta \\ &\le E\int_{-\pi}^{\pi} \|\mathcal{F}_\theta - \widehat{\mathcal{F}}_\theta\|_{\mathcal{S}}^2\, d\theta, \end{aligned} \tag{3.2}$$

where the bound (3.2) follows from the inequality $\|\cdot\|_{\mathcal{L}} < \|\cdot\|_{\mathcal{S}}$ and Lemma 2.2 in [20]. We thus see that the key to obtain an upper bound on $E|\Lambda_j - \widehat{\Lambda}_j|^2$ is to establish an upper bound on (3.2). The following proposition establishes such a bound.

Proposition 3.1. *Consider the estimator $\widehat{\mathcal{F}}_\theta$ with kernel* (2.3)*. If Assumptions* 2.1 *and* 2.2 *hold, then*

$$E\int_{-\pi}^{\pi} \|\mathcal{F}_\theta - \widehat{\mathcal{F}}_\theta\|_{\mathcal{S}}^2\, d\theta = O\big(P_1(N) + P_2(N) + q/N\big), \tag{3.3}$$

where

$$P_1(N) = \sum_{|h|>q} \|C_h\|_{\mathcal{S}}^2, \qquad P_2(N) = \sum_{|h|\le q}\left[w\left(\frac{h}{q}\right)\left(1 - \frac{|h|}{N}\right) - 1\right]^2 \|C_h\|_{\mathcal{S}}^2.$$

It follows from the proof of Proposition 3.1, that the terms P_1 and P_2 are exact in the sense that their multiples appear in the decomposition of the left-hand side of (3.3). The remainder term is $O(q/N)$ under the assumptions of Proposition 3.1. We see, that the best rate our approach allows us to obtain is given by

$$\limsup_{N\to\infty} \sup_{j\ge 1} \frac{N}{q} E|\Lambda_j - \widehat{\Lambda}_j|^2 < \infty. \tag{3.4}$$

We will see in Section 4 that this rate cannot be improved because $\sqrt{N/q}(\Lambda_j - \widehat{\Lambda}_j)$ converges in distribution to a nondegenerate limit. These relations imply that the distance between $\widehat{\Lambda}_j$ and

Λ_j is of the order $\sqrt{q/N}$, and is asymptotically larger than the corresponding distance between the static population variances λ_j and their estimators $\hat{\lambda}_j$.

For relation (3.4) to hold, we need $P_1(N) = O(q/N)$ and $P_2(N) = O(q/N)$. Assumption (2.1) allows us only to conclude $P_1(N) \to 0$ ($\|C_h\|_{\mathcal{S}} = \|c_h\|_2$), so $P_1(N) = O(q/N)$ is an effective restriction. Similarly, $P_2(N) \to 0$, by dominated convergence, and $P_2(N) = O(q/N)$ is an effective restriction, which reflects an interplay between the rates of decay of $w(\cdot) - 1$ and of the squared norms $\|C_h\|_{\mathcal{S}}^2$. Many abstract assumptions could be formulated to ensure that these conditions hold. Instead, we consider the most common case in which q is a power function of N. Before doing so, we state the following corollary, which follows from the above discussion.

Corollary 3.1. *Under the assumptions of Proposition* 3.1, $E \int_{-\pi}^{\pi} \|\mathcal{F}_\theta - \widehat{\mathcal{F}}_\theta\|_{\mathcal{S}}^2 \, d\theta \to 0$.

We now turn to the assumptions that guarantee a specific rate of decay to zero.

Assumption 3.1. The bandwidth function q satisfies $q(N) = cN^p$, for some $p \in (0, 1)$ and some $c > 0$.

Assumption 3.2. The tail series $\sum_{|h|>q} \|C_h\|_{\mathcal{S}}^2$ tends to zero with rate N^{p-1} or faster, where p is as in Assumption 3.1.

We specify the range of p for two specific, commonly used kernels, the Bartlett kernel

$$w(s) = \begin{cases} 1 - |s|, & 0 \le |s| \le 1, \\ 0, & |s| > 1, \end{cases} \tag{3.5}$$

and the Parzen kernel

$$w(s) = \begin{cases} 1 - 6|s|^2 + 6|s|^3, & |s| \le \dfrac{1}{2}, \\ 2\big(1 - |s|^3\big), & \dfrac{1}{2} \le |s| \le 1, \\ 0, & |s| > 1. \end{cases} \tag{3.6}$$

Theorem 3.1. *Suppose Assumption* 2.2 *holds. Then, relation* (3.4) *holds in any of the following two cases*:

(i) *the Bartlett kernel is used*, $\sum_{h \in \mathbb{Z}} |h|^2 \|C_h\|_{\mathcal{S}}^2 < \infty$, *and Assumptions* 3.1 *and* 3.2 *hold for* $p \in [\frac{1}{3}, 1)$;

(ii) *the Parzen kernel is used*, $\sum_{h \in \mathbb{Z}} |h|^4 \|C_h\|_{\mathcal{S}}^2 < \infty$, *and Assumptions* 3.1 *and* 3.2 *hold for* $p \in [\frac{1}{7}, 1)$.

Remark 3.1. The optimal rate of convergence of $\sup_{j \ge 1} E|\Lambda_j - \widehat{\Lambda}_j|^2$ to zero, in terms of N only, depends on the factors: (1) the kernel used; (2) the rate at which the norms $\|C_h\|_{\mathcal{S}}$ decay to zero. Under the assumptions of Theorem 3.1, we obtain the rate $N^{-2/3}$ for the Bartlett Kernel, and $N^{-6/7}$ for the Parzen kernel. It might thus appear that the Parzen kernel should be recommended, but this is true only if $\sum_{h \in \mathbb{Z}} |h|^4 \|C_h\|_{\mathcal{S}}^2 < \infty$. If it is assumed, for example, that the

$\|C_h\|_{\mathcal{S}}$ decay exponentially fast, then the Parzen kernel is superior. Under weaker assumptions on the rate of decay of the $\|C_h\|_{\mathcal{S}}$, only slower can be claimed. These issues are discussed in greater detail in Section B of the Supplementary Material [25], which also studies other kernels.

To summarize, the results of this section show that if Λ_j is the variance component given by (2.10), equivalently by (2.14), and $\widehat{\Lambda}_j$ its estimator given by (3.1), then $E|\widehat{\Lambda}_j - \Lambda_j|^2 \sim q/N$, where q is the bandwidth used in the estimator (2.3). The result is proven for the Bartlett and Parzen kernels. In Section 4, we focus on the asymptotic distribution. The results of the present section provide guidance for the results of Section 4, but do not follow from them as convergence of moments cannot be directly inferred from weak convergence.

We conclude this section by noting that a different estimator of $\mathcal{F}_\theta$, based on locally averaging a suitably defined functional periodogram, was used in Section 3 of [34]. Abusing notation by denoting their estimator also by $\widehat{\mathcal{F}}_\theta$, their Theorem 3.6 shows that the left-hand side of (3.3) tends to zero with N. As explained above, such a result also holds for the estimator defined by (2.3), requiring only Assumption 2.1, but it cannot be used to obtain the rate in (3.4); additional information about the kernel and the rates of decay of autocovariance operators and of the bandwidth is needed. Theorem 3.6 of [34] also shows that for a *fixed* θ, different from 0 or $\pm\pi$, $\|\mathcal{F}_\theta - \widehat{\mathcal{F}}_\theta\|_{\mathcal{S}}^2 = O(q^{-2}) + O(q/N)$, providing a different bound.

4. An invariance principle for estimated spectral density operators

In this section, and in Section C, we will work with the space $L^2_{\mathcal{S}}((-\pi,\pi])$ of square integrable $\mathcal{S}$-valued functions on the interval $(-\pi,\pi]$. More precisely,

$$L^2_{\mathcal{S}}((-\pi,\pi]) = \left\{\Psi : (-\pi,\pi] \longrightarrow \mathcal{S}, \|\Psi\|^2_{L^2_{\mathcal{S}}((-\pi,\pi])} = \int_{-\pi}^{\pi} \|\Psi(\theta)\|^2_{\mathcal{S}}\, d\theta < \infty\right\}.$$

To adhere to notation used in related papers, the argument θ will be indicated as a subscript in the context of spectral density operators. The space $L^2_{\mathcal{S}}((-\pi,\pi])$ is a separable Hilbert space with the inner product $\langle\Psi(\cdot),\Phi(\cdot)\rangle = \int_{-\pi}^{\pi}\langle\Psi(\theta),\Phi(\theta)\rangle_{\mathcal{S}}\, d\theta$.

We will show that asymptotic distribution of the frequency domain variance components studied in this paper can be established under the following assumption, which uses the concept of a Gaussian distribution in a Hilbert space. There are several equivalent definition of a Gaussian random element in such a space, see, for example, Section 5.2 in [30]. For example, X is Gaussian with mean zero, if and only if each projection $\langle X, h\rangle$ is a normal random variable with mean zero.

Assumption 4.1. The following weak convergence holds in $L^2_{\mathcal{S}}((-\pi,\pi])$:

$$\sqrt{N/q}\left\{\widehat{\mathcal{F}}_\theta - \mathcal{F}_\theta, \theta \in (-\pi,\pi]\right\} \xrightarrow{d} G(0,\Sigma), \tag{4.1}$$

where $G(0,\Sigma)$ denotes a Gaussian distribution on $L^2_{\mathcal{S}}((-\pi,\pi])$ with mean zero and a covariance operator Σ.

The invariance principle (4.1) is of separate and more general interest, and our specific results can be viewed as some of its applications. Since (4.1) is a weak convergence in the metric space $L^2_{\mathcal{S}}((-\pi,\pi])$, it can be combined with the continuous mapping theorem to obtain distributions of statistics based on the estimated spectral density estimator. We will show later in this section that (4.1) holds for a broad class of linear processes (2.7). No invariance principle of this type is currently available. Returning to the estimator of [34], their Theorem 3.7 essentially states that for any *fixed* frequency θ, $\sqrt{N/q}(\hat{f}_\theta - E\hat{f}_\theta)$ has a Gaussian limit in $L^2([0,1]\times[0,1])$. This will follow from (4.1) for the estimator defined by (2.3), if one can suitably bound the difference $E\hat{f}_\theta - f_\theta$. For this, additional assumptions are needed, which are generally satisfied. For statistical applications centering with $\mathcal{F}_\theta$, rather than $E\widehat{\mathcal{F}}_\theta$, is more convenient as it directly leads to statements on the convergence of an estimator to its target. Before verifying (4.1) for a specific class of models, we state a theorem it implies. In analogy to the commonly assumed condition that the eigenvalues λ_j of the covariance operator satisfy $\lambda_1 > \lambda_2 > \cdots$, we need the following assumption.

Assumption 4.2. The index j is such that for each $\theta \in (-\pi,\pi]$, $\lambda_j(\theta) - \lambda_{j+1}(\theta) > 0$ and $\lambda_{j-1}(\theta) - \lambda_j(\theta) > 0$, with only the first condition needed if $j = 1$.

By Proposition 7(a) of [15], the functions $\theta \mapsto \lambda_j(\theta)$ are continuous, so Assumption 4.2 implies that

$$\inf_{\theta\in(-\pi,\pi]}\big[\lambda_j(\theta) - \lambda_{j+1}(\theta)\big] > 0 \quad \text{and} \quad \inf_{\theta\in(-\pi,\pi]}\big[\lambda_{j-1}(\theta) - \lambda_j(\theta)\big] > 0. \tag{4.2}$$

(The interval $(-\pi,\pi]$ is viewed as a compact unit circle in the complex plane.) [15] allow the functions $\lambda_j(\theta) - \lambda_{j+1}(\theta)$ to have finitely many zeros, but for our stronger results we need a positive separation at each θ.

Theorem 4.1. *If Assumptions 2.1, 2.2, 4.1 and 4.2 hold, then*

$$\sqrt{N/q}(\widehat{\Lambda}_j - \Lambda_j) = \sqrt{\frac{N}{q}}\int_{-\pi}^{\pi}\big(\hat{\lambda}_j(\theta) - \lambda_j(\theta)\big)\,d\theta \xrightarrow{d} N\big(0,\sigma_j^2\big). \tag{4.3}$$

Denoting the limit in (4.1) by $\{Z(\theta), \theta \in (-\pi,\pi]\}$, the asymptotic variance is given by

$$\sigma_j^2 = E\bigg\{\int_{-\pi}^{\pi}\big\langle Z(\theta), \varphi_j(\theta)\otimes\varphi_j(\theta)\big\rangle_{\mathcal{S}}\,d\theta\bigg\}^2.$$

Our next objective is to show that the invariance principle (4.1) holds for functional linear processes specified in the following assumption.

Assumption 4.3. The functional time series $\{X_t\}$ has representation (2.7) with i.i.d. mean zero innovation $\{\varepsilon_t\}$ satisfying $E\|\varepsilon_t\|^8 < \infty$ and the coefficients Ψ_j satisfying (2.8).

As the first step, we establish the invariance principle with centering by the expectation, a result of independent value, which holds under weaker assumptions.

Theorem 4.2. *If Assumptions 2.1 and 4.3 hold, then the limiting distribution, as $N \to \infty$, of the $L^2_{\mathcal{S}}((-\pi,\pi])$-valued random element*

$$\sqrt{N/q}\{\widehat{\mathcal{F}}_\theta - E\widehat{\mathcal{F}}_\theta; \theta \in (-\pi,\pi]\}$$

is mean zero Gaussian.

The proof of Theorem 4.2 follows from Lemmas 4.1, 4.2, 4.3 and 4.4. All these lemmas are proven in Section C of the Supplementary Material [25] under the heading PROOF OF THEOREM 4.2.

The first lemma is a direct consequence of Theorem 4.2 in [6].

Lemma 4.1. *Let $\{Z_N\}_{N\geq 1}$ be a sequence of random elements with values in some general Hilbert space $\mathbb{H}$. Suppose the sequence $\{Z_N\}_{N\geq 1}$ admits the following decomposition*

$$Z_N = Y_{N,d} + X_{N,d}, \quad d \geq 1, N \geq 1,$$

in which

(i) *the sequence $\{X_{N,d}\}$ tends to zero in mean square, as d tends to infinity, uniformly in N, that is,* $\lim_{d\to\infty} \sup_{N\geq 1} E\|X_{N,d}\|^2 = 0$,

(ii) *for each fixed d, the sequence $\{Y_{N,d}\}$ tends in law to some random element Y_d, as N tends to infinity,*

(iii) *the limits Y_d tends in law to some random element Y, as d tends to infinity.*

Then, we conclude that the Z_N tend in law to Y.

We will work with the truncation

$$X_{t,d} = \sum_{j=0}^{d} \Psi_j(\varepsilon_{t-j}),$$

which forms a $d-$dependent process. The corresponding autocovariance operator and its estimator are denoted by $C_{h,d}$ and $\hat{c}_{h,d}$, respectively. Similarly, we use $\mathcal{F}_{\theta,d}$ and $\widehat{\mathcal{F}}_{\theta,d}$ to denote the spectral density operator and its estimator.

The next lemma states that condition (i) of Lemma 4.1 holds the space $\mathbb{H} = L^2_{\mathcal{S}}((-\pi,\pi])$, with $X_{N,d}$ defined by (4.4).

Lemma 4.2. *Under Assumptions 2.1 and 4.3, the $L^2_{\mathcal{S}}((-\pi,\pi])$-valued random element*

$$\{\sqrt{N/q}(\widehat{\mathcal{F}}_\theta - \widehat{\mathcal{F}}_{\theta,d} - E(\widehat{\mathcal{F}}_\theta - \widehat{\mathcal{F}}_{\theta,d})), \theta \in (-\pi,\pi]\} \tag{4.4}$$

tends to zero in mean square sense, as d tends to infinity, uniformly on N, that is,

$$\lim_{d\to\infty} \sup_{N\geq 1} E\|\sqrt{N/q}(\widehat{\mathcal{F}}_{\cdot} - \widehat{\mathcal{F}}_{\cdot,d} - E(\widehat{\mathcal{F}}_{\cdot} - \widehat{\mathcal{F}}_{\cdot,d}))\|^2_{L^2_{\mathcal{S}}((-\pi,\pi])} = 0. \tag{4.5}$$

Next, we verify condition (ii) of Lemma 4.1 with $Y_{N,d}$ defined by the left-hand side of (4.6).

Lemma 4.3. *Under Assumptions 2.1 and 4.3, for each fixed d, as $N \to \infty$,*

$$\sqrt{N/q}\big(\widehat{\mathcal{F}}_{\theta,d} - E\widehat{\mathcal{F}}_{\theta,d}; \theta \in (-\pi,\pi]\big) \to G_d \quad \text{in } L^2_{\mathcal{S}}\big((-\pi,\pi]\big), \tag{4.6}$$

where G_d is mean zero Gaussian.

Finally, we verify condition (iii) of Lemma 4.1.

Lemma 4.4. *Under Assumptions 2.1 and 4.3, the limit G_d in (4.6) converges in $L^2_{\mathcal{S}}((-\pi,\pi])$ to a mean zero Gaussian distribution.*

To replace the centering with the expectation by centering with $\mathcal{F}_{\cdot}$, additional assumptions are needed, namely

$$\frac{N}{q}\sum_{|h|>q}\|C_h\|^2_{\mathcal{S}} \to 0; \tag{4.7}$$

$$\frac{N}{q}\sum_{|h|\le q}\left|1 - w\left(\frac{h}{q}\right)\left(1-\frac{|h|}{N}\right)\right|^2\|C_h\|^2_{\mathcal{S}} \to 0. \tag{4.8}$$

As the proof of Theorem 4.3 shows, the left-hand sides of (4.7) and (4.8) are (up to multiplicative constants) exact difference terms, so these conditions cannot be improved.

Theorem 4.3. *If, in addition to Assumptions 2.1 and 4.3, conditions (4.7) and (4.8) hold, then (4.1) holds.*

Theorem 4.3 can be used to establish invariance principle (4.1) for specific functional time series models whose dependence can be quantified by the rate of decay of the norms $\|C_h\|_{\mathcal{S}}$, and for specific kernels w and bandwidth rates $q = q(N)$. The coefficient of $\|C_h\|^2_{\mathcal{S}}$ in (4.8) can be directly computed for any specific kernel w, like the Bartlett and Parzen kernels (3.5) and (3.6). For example, for the Bartlett kernel (3.5),

$$1 - w\left(\frac{h}{q}\right)\left(1-\frac{|h|}{N}\right) = \frac{N|h| + q|h| - |h|^2}{qN}.$$

The sum in (4.8) can then be easily bounded from above using a specific bound on the rate of decay of the $\|C_h\|^2_{\mathcal{S}}$. If these norms decay exponentially fast, conditions (4.7) and (4.8) will generally hold. We state a corollary which connects to the results of Section 3, and which can be proven by the method outlined above.

Corollary 4.1. *Suppose the functional time series $\{X_t\}$ satisfies Assumption 4.3. If the conditions of Theorem 3.1 hold, but replacing intervals $[\frac{1}{3},1)$ and $[\frac{1}{7},1)$ by open intervals $(\frac{1}{3},1)$ and $(\frac{1}{7},1)$, respectively, and the series $\sum_{|h|>q}\|C_h\|_{\mathcal{S}}$ tends to zero with rate $N^{(p-1-\alpha)}$, for some $\alpha > 0$, then (4.1) holds.*

Theorem 4.2 is difficult to prove; its proof takes up most of Section C of the Supplementary Material [25]. We conclude this section with comments which provide some context. The proof of Theorem 4.2 requires, after numerous other steps, an application of a central limit theorem for triangular arrays in the space $L^2_{\mathcal{S}}((-\pi,\pi])$. We could not find a suitable theorem in an abstract separable Hilbert space, which would be applicable in our context. We established a custom designed result as Lemma C.1 in the Supplementary Material [25], which is proven using key assumptions of Theorem 9 of [11] in our specialized setting. We state a simplified version of Lemma C.1 (Theorem 4.5), which is not directly applicable in our setting, but which may be applicable, possibly with some adjustments, in other settings. The proof is provided in Section C to enhance possible modifications and customizations. Our Theorem 4.5 is an extension of the following result used in Chapter 8 of [1] to establish the asymptotic normality of the spectral density estimator at a fixed frequency.

Theorem 4.4 (Anderson [1]). *Suppose $\zeta_{j,N}$ are mean zero rowwise i.i.d. and satisfy the following conditions*: (a) $E\zeta_{1,N}^2 = r_N\sigma_Z^2$, $r_N \to 1$, (b) $\sup_{N\ge 1} E\zeta_{1,N}^4 < \infty$. *Then*, $M_N^{-1/2}\sum_{j=1}^{M_N}\zeta_{j,N} \xrightarrow{d} N(0,\sigma_Z^2)$, *for any sequence $M_N \to \infty$.*

Theorem 4.4 essentially follows from the Lindeberg–Feller theorem, see, for example, Theorem 5.12 in [23].

Theorem 4.5. *Suppose for each j and N, $Z_{j,N}$ is a mean zero random element of a separable Hilbert space $\mathbb{H}$. For a fixed N, the $Z_{j,N}$ are i.i.d. with*

$$\kappa_2 := \sum_{N\ge 1} E\|Z_{j,N}\|^2 < \infty. \tag{4.9}$$

Moreover, for each $e \in \mathbb{H}$,

$$\kappa_e := \sup_{N\ge 1} E\langle e, Z_{j,N}\rangle^4 < \infty \tag{4.10}$$

and

$$E\langle e, Z_{1,N}\rangle^2 = r_N\sigma_e^2, \quad r_N \to 1. \tag{4.11}$$

Then, as $N \to \infty$, for any sequence $M_N \to \infty$,

$$M_N^{-1/2}\sum_{j=1}^{M_N} Z_{j,N} \xrightarrow{d} G,$$

where G is a mean zero Gaussian element of $\mathbb{H}$.

Supplementary Material

Supplement (DOI: 10.3150/20-BEJ1199SUPP; .pdf). All proofs, except simple, illustrative arguments, are presented in the Supplementary Material [25]. Proofs of the results stated in Sections 2 and 3 are presented in Section A. Additional results related to Theorem 3.1 are presented in Section B. Proofs of the results stated in Section 4 are presented in Section C.

References

[1] Anderson, T.W. (1971). *The Statistical Analysis of Time Series*. New York: Wiley. MR0283939
[2] Aue, A. and van Delft, A. (2020). Testing for stationarity of functional time series in the frequency domain. *Ann. Statist.* To appear.
[3] Bardsley, P., Horváth, L., Kokoszka, P. and Young, G. (2017). Change point tests in functional factor models with application to yield curves. *Econom. J.* **20** 86–117. MR3636962 https://doi.org/10.1111/ectj.12075
[4] Berkes, I., Hörmann, S. and Schauer, J. (2011). Split invariance principles for stationary processes. *Ann. Probab.* **39** 2441–2473. MR2932673 https://doi.org/10.1214/10-AOP603
[5] Berkes, I., Horváth, L. and Rice, G. (2013). Weak invariance principles for sums of dependent random functions. *Stochastic Process. Appl.* **123** 385–403. MR3003356 https://doi.org/10.1016/j.spa.2012.10.003
[6] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396
[7] Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications. Lecture Notes in Statistics* **149**. New York: Springer. MR1783138 https://doi.org/10.1007/978-1-4612-1154-9
[8] Brillinger, D.R. (1975). *Time Series: Data Analysis and Theory. International Series in Decision Processes*. New York: Holt, Rinehart and Winston, Inc. MR0443257
[9] Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods. Springer Series in Statistics*. New York: Springer. MR2839251
[10] Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* **12** 136–154. MR0650934 https://doi.org/10.1016/0047-259X(82)90088-4
[11] Garling, D.J.H. (1976). Functional central limit theorems in Banach spaces. *Ann. Probab.* **4** 600–611. MR0423449 https://doi.org/10.1214/aop/1176996030
[12] Giraitis, L., Kokoszka, P., Leipus, R. and Teyssière, G. (2003). Rescaled variance and related tests for long memory in volatility and levels. *J. Econometrics* **112** 265–294. MR1951145 https://doi.org/10.1016/S0304-4076(02)00197-5
[13] Górecki, T., Hörmann, S., Horváth, L. and Kokoszka, P. (2018). Testing normality of functional time series. *J. Time Series Anal.* **39** 471–487. MR3819053 https://doi.org/10.1111/jtsa.12281
[14] Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 109–126. MR2212577 https://doi.org/10.1111/j.1467-9868.2005.00535.x
[15] Hörmann, S., Kidziński, Ł. and Hallin, M. (2015). Dynamic functional principal components. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 319–348. MR3310529 https://doi.org/10.1111/rssb.12076
[16] Hörmann, S., Kidziński, Ł. and Kokoszka, P. (2015). Estimation in functional lagged regression. *J. Time Series Anal.* **36** 541–561. MR3356268 https://doi.org/10.1111/jtsa.12114
[17] Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38** 1845–1884. MR2662361 https://doi.org/10.1214/09-AOS768

[18] Hörmann, S. and Kokoszka, P. (2012). Functional time series. In *Time Series* (C.R. Rao and T.S. Rao, eds.). *Handbook of Statistics* **30**. Amsterdam: Elsevier.

[19] Hörmann, S., Kokoszka, P. and Nisol, G. (2018). Testing for periodicity in functional time series. *Ann. Statist.* **46** 2960–2984. MR3851761 https://doi.org/10.1214/17-AOS1645

[20] Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. *Springer Series in Statistics*. New York: Springer. MR2920735 https://doi.org/10.1007/978-1-4614-3655-3

[21] Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series. *J. Econometrics* **179** 66–82. MR3153649 https://doi.org/10.1016/j.jeconom.2013.11.002

[22] Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis*, *with an Introduction to Linear Operators*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR3379106 https://doi.org/10.1002/9781118762547

[23] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications* (*New York*). New York: Springer. MR1876169 https://doi.org/10.1007/978-1-4757-4015-8

[24] Kidziński, Ł., Kokoszka, P. and Mohammadi Jouzdani, N. (2018). Principal components analysis of periodically correlated functional time series. *J. Time Series Anal.* **39** 502–522. MR3819055 https://doi.org/10.1111/jtsa.12283

[25] Kokoszka, P. and Mohammadi Jouzdani, N. (2020). Supplement to "Frequency domain theory for functional time series: Variance decomposition and an invariance principle." https://doi.org/10.3150/20-BEJ1199SUPP

[26] Kokoszka, P. and Reimherr, M. (2013). Asymptotic normality of the principal components of functional time series. *Stochastic Process. Appl.* **123** 1546–1562. MR3027890 https://doi.org/10.1016/j.spa.2012.12.011

[27] Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. *Texts in Statistical Science Series*. Boca Raton, FL: CRC Press. MR3793167

[28] Kuelbs, J. (1973). The invariance principle for Banach space valued random variables. *J. Multivariate Anal.* **3** 161–172. MR0328986 https://doi.org/10.1016/0047-259X(73)90020-1

[29] Leucht, A., Paparoditis, E. and Sapatinas, T. (2018). Testing equality of spectral density operators for functional linear processes. Preprint, Technische Universität Braunschweig. Available at arXiv:1804.03366.

[30] Linde, W. (1986). *Probability in Banach Spaces—Stable and Infinitely Divisible Distributions*, 2nd ed. Chichester: Wiley. MR0874529

[31] Merlevède, F. (2003). On the central limit theorem and its weak invariance principle for strongly mixing sequences with values in a Hilbert space via martingale approximation. *J. Theoret. Probab.* **16** 625–653. MR2009196 https://doi.org/10.1023/A:1025668415566

[32] Merlevède, F., Peligrad, M. and Utev, S. (1997). Sharp conditions for the CLT of linear processes in a Hilbert space. *J. Theoret. Probab.* **10** 681–693. MR1468399 https://doi.org/10.1023/A:1022653728014

[33] Panaretos, V.M. and Tavakoli, S. (2013). Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series. *Stochastic Process. Appl.* **123** 2779–2807. MR3054545 https://doi.org/10.1016/j.spa.2013.03.015

[34] Panaretos, V.M. and Tavakoli, S. (2013). Fourier analysis of stationary time series in function space. *Ann. Statist.* **41** 568–603. MR3099114 https://doi.org/10.1214/13-AOS1086

[35] Pham, T. and Panaretos, V.M. (2018). Methodology and convergence rates for functional time series regression. *Statist. Sinica* **28** 2521–2539. MR3839872

[36] Pourahmadi, M. (2001). *Foundations of Time Series Analysis and Prediction Theory*. *Wiley Series in Probability and Statistics*: *Applied Probability and Statistics*. New York: Wiley. MR1849562

[37] Shao, X. and Wu, W.B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35** 1773–1801. MR2351105 https://doi.org/10.1214/009053606000001479

[38] van Delft, A. and Eichler, M. (2020). A note on Herglotz's theorem for time series on function spaces. *Stochastic Process*. *Appl*. To appear.
[39] Zhang, X. (2016). White noise testing and model diagnostic checking for functional time series. *J*. *Econometrics* **194** 76–95. MR3523520 https://doi.org/10.1016/j.jeconom.2016.04.004
[40] Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *J*. *Amer*. *Statist*. *Assoc*. **108** 726–740. MR3174655 https://doi.org/10.1080/01621459.2013.787184

Received May 2019 and revised November 2019