On the CLT for discrete Fourier transforms of functional time series

Clément Cerovecki[1]* Siegfried Hörmann[1]

[1] Department of Mathematics, Université libre de Bruxelles CP210, Bd. du Triomphe, B-1050 Brussels, Belgium.

Abstract

Abstract. The purpose of this paper is to derive sharp conditions under which the discrete Fourier transform

$$S_n(\theta) := \sum_{t=1}^{n} X_t e^{-\mathrm{i}t\theta}, \quad \theta \in (-\pi, \pi],$$

of a functional time series $(X_t \colon t \geq 1)$ is asymptotically normal. Assuming that the function space is a Hilbert space we prove that a central limit theorem (CLT) holds for almost all frequencies θ if the process (X_t) is stationary, ergodic and purely non-deterministic. Under slightly stronger assumptions we formulate versions which provide a CLT for fixed frequencies as well as for $S_n(\theta_n)$, when $\theta_n \to \theta_0$ is a sequence of fundamental frequencies. In particular we also deduce the regular CLT ($\theta = 0$) under new and very mild assumptions. We show that our results apply to the most commonly studied functional time series.

Keywords: central limit theorem, functional time series, Fourier transform, periodogram, stationarity

*Corresponding author. Email: clement.cerovecki@gmail.com

1 Introduction

Functional time series analysis is a branch of the emerging statistical field of functional data analysis (FDA)—we refer to the monographs [11, 18, 22, 40]. A quickly accessable recent overview is presented in [7]. An excellent literature survey for recent developments in FDA can be found in [12]. This paper serves as the introduction to a special volume of JMVA which contains a number of new contributions to the field.

The need for functional time series (FTS) methodology is easily explained by the fact that many functional data are sequentially sampled and serially correlated by their very nature. A common situation is that a continuous time process is cut into natural segments, such as days. Then there is not just dependence within the individual curves but also across curves and we obtain a time series $(X_t\colon t \in \mathbb{Z})$ with realizations in some function space, i.e., every observation X_t is a random curve $(X_t(\tau)\colon \tau \in \mathcal{U})$ with some continuous domain $\mathcal{U}$. Despite the fact that there exist a variety of recent results related to forecasting functional time series data (e.g., [1, 13, 23, 26, 28, 29, 41]), most of the available FDA methodology is devoted to i.i.d. samples. Functional time series methodology in general has not yet received as much attention as one might expect. One of the first and most seminal contributions is [3]. This monograph is formulating the basic theoretical foundation for FTS. The book focuses on the analysis of functional AR (FAR) processes, which are nowadays among the most popular and best studied FTS models. A core reason for this is that these processes are very convenient for prediction (see e.g. [1, 9, 13, 25, 28]). Further refinements related to FAR modeling can e.g. be found in [8, 30].

The FAR setting also provides a structural dependence framework. For asymptotic inference on time series data imposing a certain dependence assumption is crucial. Typically some form of near-epoch dependence or mixing assumptions, like strong mixing or cumulant mixing, are used for FTS in order establish large sample results (see e.g. [10, 16, 33].) In this paper we will work with so-called *purely non-deterministic processes* (see Assumption 1 below). This framework is very general and includes many commonly employed weak-dependence frameworks.

In our paper, like in most other time series contributions, stationarity is a crucial assumption. To verify this assumption for functional data we refer to [20], who propose corresponding tests.

Lately, there have been some papers devoted to frequency domain analysis for FTS, e.g., [14, 15, 33]. In contrast to the time domain analysis, which is based on analyzing the data sample and the auto-covariance function, the frequency domain analysis is grounded on the *discrete Fourier transform (DFT)*

$$S_n(\theta) = \sum_{t=1}^{n} X_t e^{-\mathrm{i}t\theta}, \quad \theta \in (-\pi, \pi]$$

of some FTS $(X_t \colon t \in \mathbb{Z})$ and its *spectral density operator*,

$$\mathcal{F}_\theta := \sum_{h\in\mathbb{Z}} C_h e^{-\mathrm{i}h\theta}. \tag{1}$$

Here C_h is the lag h covariance operator of the stationary functional time series—the precise definition of C_h is given below in Section 2. The sample $X_1, \ldots, X_n$ and $(C_h \colon h \in \mathbb{Z})$ are equivalent to $(S_n(\theta) \colon \theta \in (-\pi, \pi])$ and $(\mathcal{F}_\theta \colon \theta \in (-\pi, \pi])$, respectively, in the sense that one can be obtained from the other. Depending on the nature of the problem, one or the other approach may be simpler or more effective. For example, [34] and [14] demonstrate that dimension reduction via principal components is more effective (in fact optimal in a certain sense) when done in the frequency domain.

Like for scalar or multivariate time series, the DFT is the main building block for the frequency domain analysis, and hence understanding its asymptotics is a fundamental problem. Moreover, the DFT is of direct interest to statisticians since it is closely related to the *periodogram* which can, for example, be used to detect some underlying periodic behavior of the time series. (See, e.g., [4].) In [17] a variety of test statistics which can be used to reveal a periodic trend in a FTS are proposed and examined. Those tests involve in one or the other way the functional DFT as a building block. It is demonstrated, for example, that a standard functional ANOVA test—which can be used for this problem—is composed in such a way. Unless $(X_t \colon t \in \mathbb{Z})$ is a Gaussian process, the exact distribution of the DFT is infeasible and hence, in order to derive critical values for the tests, a Fourier CLT is needed. Let us also mention that very recently a bootstrap procedure to approximate the distribution of $S_n(\theta)$ has been developed ([35]).

Clearly, when $\theta = 0$ we obtain the regular partial sums process, which is without any doubt of fundamental importance for statistical inference. In [19] asymptotic normality under L^2–m-approximability (see Section 3.3) is derived and used to test for equality in mean of two time series samples. This CLT is a special case of our Theorem 8 below.

In a seminal paper [33] have shown that under regularity assumptions $S_n(\theta)/\sqrt{n}$ converges to a (complex) Gaussian random element with covariance operator $\mathcal{F}_\theta$. (The actual definition of $\mathcal{F}_\theta$ involves scaling by $\frac{1}{2\pi}$, which we omit here.) Hence, $\mathcal{F}_\theta$ can be interpreted as the asymptotic covariance operator of the discrete Fourier transform (DFT). In [33] it is assumed that $\sum_{h\in\mathbb{Z}} \|C_h\|_{\mathcal{T}} < \infty$ (here $\|\cdot\|_{\mathcal{T}}$ denotes the trace norm—see Section 2) in order to assure convergence of the series in (1). It follows that $\mathcal{F}_\theta$ is a nuclear operator, i.e., it has a finite trace. This is an important feature when it comes to verifying tightness of $(S_n(\theta)/\sqrt{n}\colon n \geq 1)$. Regarding the dependence structure, a cumulant type mixing condition for functional data is used. The nice feature of such mixing conditions is that no specific time series model needs to be imposed. Still, this approach requires to compute and bound functional cumulants of all orders, which is generally not an easy task and necessitates moments of all orders. *The main objective of this paper is then to relax these conditions. All our theorems below hold assuming only finite second moments and some very general form of weak dependence.*

For real valued processes asymptotic normality for $S_n(\theta)$ has been obtained under several dependence conditions. Here we only cite the early paper of [42] who considered linear processes, a survey article of [27] and the more recent contributions of [43] and [36]. The latter paper covers a variety of special cases, including strong mixing sequences. It also contains a more detailed literature survey. One of the main results of our article is an extension of the CLT of [36] to functional data. We show the weak convergence of $S_n(\theta)/\sqrt{n}$ for *purely non-deterministic* processes. More precisely, letting $\mathcal{G}_t = \sigma(X_t, X_{t-1}, \ldots)$—the σ-algebra generated by $(X_s\colon s \leq t)$—and $\mathcal{G}_{-\infty} = \bigcap_{t\geq 0} \mathcal{G}_{-t}$ we impose the following assumption.

Assumption 1. *The process $(X_t\colon t \in \mathbb{Z})$ is stationary and ergodic and satisfies $E[X_0|\mathcal{G}_{-\infty}] = 0$ a.s.*

We remark that a conditional expectation for random elements in Hilbert spaces as just stated is well defined if $E\|X_0\| < \infty$ (see, e.g., [3, p.29]). Besides the obligatory existence of second order moments, Assumption 1 will be the only condition needed for the CLT presented below in Theorem 1. Since we will not impose any further condition ensuring summability of the C_h, a tricky part is the construction and definition of the spectral density operator. Our construction will be an indirect one based on a completeness argument in an appropriate Hilbert space.

In Theorem 2 we will give a result which is slightly less general, but is more useful in applications since it will allow for more explicit constructions of $\mathcal{F}_\theta$. In our Theorem 3 we consider the case $\theta = 0$ and derive the CLT for regular partial sums. These main results along with the precise technical setting are presented in Section 2. In Section 3 we consider application of our theorems to so-called Bernoulli shifts. Within this framework we can further refine the asymptotics and consider the weak convergence of $S_n(\theta_n)$ when $(\theta_n \colon n \geq 1)$ is a convergent sequence of *fundamental frequencies*. We also show how the theorems apply in some commonly employed dependence frameworks for functional time series models and compare the required conditions to existing ones in the literature. Proofs are given in Section 4.

2 Main results

We start by introducing further notation and stating the setup precisely.

The process $(X_t \colon t \in \mathbb{Z})$ is defined on some probability space $(\Omega, \mathcal{A}, P)$ and takes values in some real *separable Hilbert space* H_0. Although our observations are assumed to be real, the very definition of $S_n(\theta)$ necessitates to adopt a complex setting. So we will henceforth consider the complex Hilbert space $H = H_0 + \mathrm{i}H_0$. Throughout u denotes a generic element in H. Then u is of the form $u = u_0 + \mathrm{i}u_1$ where u_0, u_1 denote generic elements in H_0. We denote $\overline{u} = u_0 - \mathrm{i}u_1$. The space H is equipped with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$ which are induced from the inner product $\langle \cdot, \cdot \rangle_{H_0}$ on H_0, i.e., $\langle u_0 + \mathrm{i}u_1, v_0 + \mathrm{i}v_1 \rangle = \langle u_0, v_0 \rangle_{H_0} + \langle u_1, v_1 \rangle_{H_0} + \mathrm{i}\big(\langle u_1, v_0 \rangle_{H_0} - \langle u_0, v_1 \rangle_{H_0}\big)$. We write $X \in L^p_H(\Omega)$ (short for $X \in L^p_H(\Omega, \mathcal{A}, P)$) to indicate that $E\|X\|^p < \infty$. The space $L^p_H(\Omega)$ is a Banach space and for $p = 2$ again a Hilbert space with inner product $E\langle X, Y \rangle$. The covariance operator $\mathrm{Cov}(X, Y)$ on $L^2_H(\Omega)$ is defined as $\mathrm{Cov}(X, Y)(u) = E\big[(X - EX)\langle u, Y - EY \rangle\big]$ and $\mathrm{Var}(X) := \mathrm{Cov}(X, X)$. We denote by $C_h = \mathrm{Cov}(X_h, X_0)$, $h \in \mathbb{Z}$, the lag h autocovariance operator of the time series. Expectations or other integrals for elements with values in Banach spaces are understood in the sense of Bochner integrals, see, e.g., [32].

In the following $\mathcal{N}_{H_0}(\mu, \Sigma)$ denotes a Gaussian element in H_0 with mean μ and covariance operator Σ. Then $X \sim \mathcal{N}_{H_0}(\mu, \Sigma)$ if and only if the projection $\langle X, u_0 \rangle$ is normally distributed with mean $\langle \mu, u_0 \rangle$ and variance $\langle \Sigma(u_0), u_0 \rangle$. A complex $Z = Z_0 + \mathrm{i}Z_1 \in H$ is said to be Gaussian if (Z_0, Z_1) is a Gaussian element in $H_0 \times H_0$. Define $\mu_i = EZ_i$ $(i = 0, 1)$, $V_{ij} = \mathrm{Cov}(Z_i, Z_j)$ $(i, j \in$

$\{0,1\}$) and set $\mu = \mu_0 + \mathrm{i}\mu_1$. Moreover, set $\Gamma = \mathrm{Var}(Z)$ and let $C(u) = E\big[(Z-\mu)\langle u, \overline{Z-\mu}\rangle\big]$ be the *relation operator* of Z. By simple algebra $\Gamma(u) = V_{00}(u) + V_{11}(u) + \mathrm{i}\big(V_{10}(u) - V_{01}(u)\big)$ and $C(u) = V_{00}(u) - V_{11}(u) + \mathrm{i}\big(V_{01}(u) + V_{10}(u)\big)$. With $\mathrm{Re}(\Gamma) := V_{00} + V_{11}$ and $\mathrm{Im}(\Gamma) := V_{10} - V_{01}$ and analogue definitions for $\mathrm{Re}(C)$ and $\mathrm{Im}(C)$ it follows then that

$$\begin{pmatrix} Z_0 \\ Z_1 \end{pmatrix} \sim \mathcal{N}_{H_0 \times H_0}\left(\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, \frac{1}{2}\begin{bmatrix} \mathrm{Re}(\Gamma + C) & -\mathrm{Im}(\Gamma - C) \\ \mathrm{Im}(\Gamma + C) & \mathrm{Re}(\Gamma - C) \end{bmatrix}\right). \tag{2}$$

Relation (2) implies that the law of Z is determined by μ, Γ and C. We write $Z \sim \mathcal{CN}_H(\mu, \Gamma, C)$. It can be readily shown that $Z \sim \mathcal{CN}_H(0, \Gamma, C)$, if and only if for any $u \in H$ we have

$$\langle Z, u\rangle \sim \mathcal{CN}_{\mathbb{C}}\big(0, \langle \Gamma(u), u\rangle, \langle C(\overline{u}), u\rangle\big).$$

In this paper we are mainly dealing with the *circularly-symmetric* case of Gaussian elements, i.e., when $\mu = 0$ and $C = 0$ (see, e.g., [4, p. 444] for the multivariate case). Then it is common to write $Z \sim \mathcal{CN}_H(0, \Gamma)$. The other important special case is if Z is real (i.e., when $Z_1 = 0$). This is equivalent to say that Γ is real and equals C. Then we can either consider Z as element in H_0 with $Z \sim \mathcal{N}_{H_0}(0, \Gamma)$ or view it as an element in H with $Z \sim \mathcal{CN}_H(0, \Gamma, \Gamma)$. We take the latter point of view.

Below we will consider bounded linear, compact operators $A : H \to H$. Recall that for some orthonormal basis (ONB) $(v_j \colon j \geq 1)$ of H the Hilbert-Schmidt norm of A is $\|A\|_{\mathcal{S}} = \big(\sum_{j\geq 1} \|A(v_j)\|^2\big)^{1/2}$ and the Trace norm of A is $\|A\|_{\mathcal{T}} = \sum_{j\geq 1}\langle (A^*A)^{1/2}(v_j), v_j\rangle$. Both norms are independent of the choice of $(v_j \colon j \geq 1)$. If $\|A\|_{\mathcal{S}} < \infty$ we say that A is Hilbert-Schmidt and if $\|A\|_{\mathcal{T}} < \infty$ we say that A is trace class. We have $\|A\|_{\mathcal{S}} \leq \|A\|_{\mathcal{T}}$. If A is self-adjoint and non-negative definite then $\mathrm{tr}(A) = \|A\|_{\mathcal{T}} = \sum_{j\geq 1}\langle A(v_j), v_j\rangle$. For a zero mean element $X \in L^2_H(\Omega)$ it holds that $\mathrm{tr}\big(\mathrm{Var}(X)\big) = E\|X\|^2$. Finally we recall that a sequence of operators A_n on H is said to converge in the *weak operator topology* to A if $\langle A_n(u), v\rangle \to \langle A(u), v\rangle$ for all $u, v \in H$. Short we write $A_n \xrightarrow{\mathrm{w}} A$.

Theorem 1. *Let $(X_t \colon t \in \mathbb{Z})$ be a sequence in $L^2_{H_0}(\Omega)$ which satisfies Assumption 1. Then for almost every $\theta \in (-\pi, \pi]$ there exists a linear operator $\mathcal{F}_\theta$, which is self-adjoint and non-negative definite such that*

$$\frac{1}{\sqrt{n}} S_n(\theta) \xrightarrow{d} \mathcal{CN}_H\left(0, \mathcal{F}_\theta\right).$$

Moreover we have that

(I) $\frac{1}{n}\mathrm{Var}\big(S_n(\theta)\big) \xrightarrow{\mathrm{w}} \mathcal{F}_\theta$;

(II) $\frac{1}{n}E\|S_n(\theta)\|^2 = \frac{1}{n}\mathrm{tr}\big(\mathrm{Var}\big(S_n(\theta)\big)\big) \to \mathrm{tr}\big(\mathcal{F}_\theta\big) < \infty$;

(III) $C_h = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{F}_\theta\, e^{\mathrm{i}h\theta} d\theta, \qquad \forall h \in \mathbb{Z}$;

(IV) *for almost all* $(\theta, \theta') \in (-\pi, \pi]^2$ *the components of* $(S_n(\theta), S_n(\theta'))/\sqrt{n}$ *are asymptotically jointly Gaussian and independent.*

We call $\mathcal{F}_\theta$ the spectral density operator of $(X_t\colon t \in \mathbb{Z})$ and remark that it is generally not explicitly defined as in (1). In fact, the series in (1) may not be convergent under our mild assumptions. Since

$$\mathcal{F}_{n;\theta} := \frac{1}{n}\mathrm{Var}\big(S_n(\theta)\big) = \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) C_h e^{-\mathrm{i}h\theta},$$

relation **(I)** implies solely that the Cesàro averages of $(C_h e^{-\mathrm{i}h\theta}\colon h \in \mathbb{Z})$ converge (in weak operator topology).

For practical reasons it is useful to know for which frequencies Theorem 1 holds. For example, $\theta = 0$ is an important special case, but the theorem doesn't say if this frequency is part of the exceptional null set or not. We will see that one of the delicate steps in the proof of Theorem 1 is to guarantee existence of the operator $\mathcal{F}_\theta$ and to establish the related convergence in **(I)** and **(II)**. By the extremely mild assumptions we are imposing, we can only assure this for almost every θ. Requiring Assumption 2 below allows us to establish the same result for some fixed frequency θ_0. To formulate this assumption we first introduce the projection operator $\mathcal{P}_k := E[\,\cdot\,|\mathcal{G}_k] - E[\,\cdot\,|\mathcal{G}_{k-1}]$, $k \in \mathbb{Z}$. It is elementary that $\mathcal{P}_k$ are linear operators on $L_H^1(\Omega)$. We note at at this point that a key property which we will need is that $\mathrm{Cov}(\mathcal{P}_k(X), \mathcal{P}_\ell(Y)) = 0$ (the zero operator) for all $X, Y \in L_H^2(\Omega)$ when $k \neq \ell$. That is, projections are strongly orthogonal. (See Lemma 9.)

Assumption 2. *The process* $(X_t\colon t \in \mathbb{Z})$ *is stationary and ergodic and for some* $\theta_0 \in (-\pi, \pi]$ *the following properties hold:*

(A1) $Z_n(\theta_0) := \sum_{t=0}^{n} \mathcal{P}_0(X_t) e^{-\mathrm{i}t\theta_0}$ *is a Cauchy sequence in* $L_H^2(\Omega)$;

(A2) $E\big\|E[S_n(\theta_0)|\mathcal{G}_0]\big\|^2 = o(n)$.

To get some intuition behind **(A1)** and **(A2)** and the general approach we introduce $Z_n^{(k)}(\theta) := \sum_{t=0}^{n} \mathcal{P}_k(X_{t+k})e^{-\mathrm{i}t\theta}$ and write

$$
\begin{aligned}
S_n(\theta) &= \sum_{k=1}^{n} \mathcal{P}_k(S_n(\theta)) + E[S_n(\theta)|\mathcal{G}_0] \\
&= \sum_{k=1}^{n}\sum_{t=k}^{n} \mathcal{P}_k(X_t)e^{-\mathrm{i}t\theta} + E[S_n(\theta)|\mathcal{G}_0] \\
&= \sum_{k=1}^{n} Z_{n-k}^{(k)}(\theta)e^{-\mathrm{i}k\theta} + E[S_n(\theta)|\mathcal{G}_0]. \qquad (3)
\end{aligned}
$$

The variables $Z_n^{(k)}(\theta)$, $k \geq 1$, are strongly orthogonal and by Assumption **(A1)** they have a limit $Z^{(k)}(\theta)$ in $L_H^2(\Omega)$ for $\theta = \theta_0$. Moreover, it is easy to see that $(Z^{(k)}(\theta)\colon k \geq 1)$ is a stationary martingale difference sequence. Together with **(A2)** this guarantees that $S_n(\theta)$ is close to $T_n(\theta) := \sum_{k=1}^{n} Z^{(k)}(\theta)e^{-\mathrm{i}k\theta}$. The partial sum $T_n(\theta)$ is more handy when it comes to study the CLT and to compute the covariance operator. We define

$$
\mathcal{F}_\theta := \mathrm{Var}\big(T_n(\theta)/\sqrt{n}\big) = \mathrm{Var}\big(Z^{(1)}(\theta)\big). \qquad (4)
$$

Theorem 2. *Let $(X_t\colon t \in \mathbb{Z})$ be a sequence in $L_{H_0}^2(\Omega)$ which satisfies Assumption 2 for some $\theta_0 \in (-\pi, \pi]$. Then*

$$
\frac{1}{\sqrt{n}} S_n(\theta_0) \xrightarrow{d} \mathcal{CN}_H\big(0, \mathcal{F}_{\theta_0}, C_{\theta_0}\big),
$$

where $\mathcal{F}_{\theta_0}$ is defined as in (4) *and where $C_{\theta_0} = \mathcal{F}_{\theta_0} I\{\theta_0 \in \{0, \pi\}\}$. Furthermore, the conclusions* **(I)** *and* **(II)** *of Theorem 1 hold for frequency θ_0. If Assumption 2 holds in addition for some $\theta_0' \neq \pm\theta_0$, then conclusion* **(IV)** *of Theorem 1 holds with $(\theta, \theta') = (\theta_0, \theta_0')$.*

As a corollary of this theorem, we obtain the CLT for regular partial sums. It can be phrased as follows.

Theorem 3. *Suppose that $(X_t\colon t \in \mathbb{Z}) \in L_{H_0}^2(\Omega)$. If Assumption 2 holds with $\theta_0 = 0$, then $(X_1 + \cdots + X_n)/\sqrt{n} \xrightarrow{d} \mathcal{N}_{H_0}(0, \mathcal{F}_0)$. We have that $\mathcal{F}_0$ is a non-negative definite, self-adjoint and trace-class operator and that*

$$
\sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) C_h \xrightarrow{w} \mathcal{F}_0.
$$

We conclude this section with a third and again slightly stronger assumption. Here and in the sequel $\nu_p(X) = \big(E\|X\|^p\big)^{1/p}$, $p \geq 1$.

Assumption 3. *The process* $(X_t\colon t \in \mathbb{Z})$ *satisfies Assumption 1 and*

(A3) $\sum_{t=0}^{\infty} \nu_2(\mathcal{P}_0(X_t)) < \infty$.

With this new assumption we can obtain the following useful implications.

Lemma 4. *Assumption 3 implies that*

(i) Assumption 2 holds for all $\theta \in (-\pi, \pi]$*;*

(ii) $\sum_{h\in\mathbb{Z}} \|C_h\|_{\mathcal{S}} < \infty$*;*

(iii) $\mathcal{F}_\theta$ *in* (1) *and* (4) *coincide.*

The proof of Lemma 4 is given in Section 4.

3 Application to Bernoulli shifts

The typical framework we have in mind comprises processes $(X_t\colon t \in \mathbb{Z})$ which can be represented as Bernoulli shifts, i.e.,

$$X_t = f(\varepsilon_t, \varepsilon_{t-1}, \ldots), \tag{5}$$

where $(\varepsilon_t\colon t \in \mathbb{Z})$ is a stationary and ergodic sequence of elements in some normed vector space S and $f : S^{\mathbb{N}} \to H_0$ is measurable. Then $(X_t\colon t \in \mathbb{Z})$ is stationary and ergodic. We remark that in this case we can use in our theorems the filtration $(\mathcal{G}_k\colon k \in \mathbb{Z})$ with $\mathcal{G}_k = \sigma(\varepsilon_k, \varepsilon_{k-1}, \ldots)$. Representation (5) is very common to many time series models. In particular it applies to the two dependence frameworks we are going to discuss below, namely *linear processes* (possibly with dependent noise) and $L^2 - m$*–approximable processes*. These two concepts cover most of the functional time series models studied in the literature.

When $(\varepsilon_t\colon t \in \mathbb{Z})$ are i.i.d., then by Kolmogorov's 0-1 law Assumption 1 applies to all such processes. The following convenient condition thus implies Assumtion 3.

Assumption 4. *The process* $(X_t\colon t \in \mathbb{Z})$ *has representation* (5) *with i.i.d. innovations* $(\varepsilon_t\colon t \in \mathbb{Z})$ *and satisfies* **(A3)***.*

It should be stressed that $(\varepsilon_t \colon t \in \mathbb{Z})$ in (5) need not necessarily be independent in order to yield Assumption 1. For example, if $(\varepsilon_t \colon t \in \mathbb{Z})$ are *strongly mixing* then the tail sigma algebra $\mathcal{G}_{-\infty}$ is again trivial (see [5, p.10]).

For Bernoulli shifts, we can obtain the following refinement of our Theorem 2.

Theorem 5. *Suppose that $(X_t \colon t \in \mathbb{Z})$ are square integrable random elements satisfying Assumption 4. Suppose that $(\theta_n \colon n \geq 1)$ and $(\theta'_n \colon n \geq 1)$ are two sequences of fundamental frequencies (i.e., $\theta_n \in \frac{2\pi}{n}\mathbb{Z}$) with $\theta_n \to \theta$ and $\theta'_n \to \theta'$. Assume further that for all $n \geq 1$ it holds that $\theta_n \neq \pm\theta'_n$ and $\theta_n, \theta'_n \notin \pi\mathbb{Z}$. Then*

$$\frac{1}{\sqrt{n}}(S_n(\theta_n), S_n(\theta'_n)) \xrightarrow{d} \mathcal{CN}_{H \times H}\left(0, \begin{bmatrix} \mathcal{F}_\theta & 0 \\ 0 & \mathcal{F}_{\theta'} \end{bmatrix}\right),$$

where $\mathcal{F}_\theta$ is defined as in (1)*.*

Note that θ_n and θ'_n are allowed to converge to the same limit. The asymptotic Fourier transforms will stay independent. We also stress that if $\theta_n \to \theta \in \{0, \pi\}$, then $S_n(\theta_n)/\sqrt{n}$ and $S_n(\theta)/\sqrt{n}$ have different asymptotics. While for the first we obtain a complex limiting law, the distribution of $S_n(\theta)/\sqrt{n}$ is real.

In the following subsections we explicitly work out three important types of process and verify for each Assumption 4. We can thus deduce, that Theorems 1, 2, 3 and 5 are applicable for these processes.

3.1 Linear processes

Consider a linear process $X_t = \sum_{k \geq 0} \Psi_k(\varepsilon_{t-k})$ where $(\varepsilon_t \colon t \in \mathbb{Z})$ are i.i.d. and zero mean in some Hilbert space H_1 and $\Psi_k \colon H_1 \to H_0$ are bounded linear operators. We denote by $\|\Psi\|_{\mathcal{L}}$ the operator norm.

Theorem 6. *If $X_t \in L^2_{H_0}(\Omega)$ then Theorem 1 holds. If in addition $\kappa := \sum_{k \geq 0} \|\Psi_k\|_{\mathcal{L}} < \infty$ and $\varepsilon_0 \in L^2_{H_1}(\Omega)$ then Assumption 4 holds. Moreover,*

$$\mathcal{F}_\theta = \Psi(\theta) V \Psi(\theta)^*,$$

where $\Psi(\theta) = \sum_{k \geq 0} \Psi_k e^{-\mathrm{i}k\theta}$ and $\Psi(\theta)^*$ is its adjoint operator and $V = \mathrm{Var}(\varepsilon_0)$.

It is easy to see that $\varepsilon_0 \in L^2_{H_1}(\Omega)$ implies that $X_t \in L^2_{H_0}(\Omega)$. Consequently, our Theorem 6 improves the corresponding result in [33], where it is required that $\varepsilon_0 \in L^k_{H_0}(\Omega)$ for all $k \geq 1$.

When $\theta = 0$ we recover the ordinary CLT for the partial sums of $(X_t \colon t \in \mathbb{Z})$ as, e.g., proven in [31]. While for scalar linear processes the CLT only requires square summability of the coefficients, the latter authors prove that in infinite dimensional Hilbert spaces assuming absolute summability is essentially sharp.

Proof. Note that $\mathcal{P}_0(X_t) = \Psi_t(\varepsilon_0)$. Hence, condition **(A3)** follows immediately and thus Assumption 4 holds. The rest follows from the implications of Lemma 4. □

3.2 Linear processes with dependent errors

Consider once again a linear process $X_t = \sum_{k\geq 0} \Psi_k(\delta_{t-k})$, where now $(\delta_t \colon t \in \mathbb{Z})$ has the Bernoulli representation (5) $\delta_t = f(\varepsilon_t, \varepsilon_{t-1}, \dots)$ with i.i.d. innovations $(\varepsilon_t \colon t \in \mathbb{Z})$.

Theorem 7. *Suppose that $(X_t \colon t \in \mathbb{Z})$ is a linear process as stated above, satisfying the summability condition $\sum_{k\geq 0} \|\Psi_k\|_{\mathcal{L}} < \infty$, and further assume that the process $(\delta_t \colon t \in \mathbb{Z})$ itself satisfies condition **(A3)**. Then Assumption 4 holds for $(X_t \colon t \in \mathbb{Z})$.*

For the regular partial sums process, this result compares to [39] who have studied partial sums of linear processes in Banach spaces. They show that the CLT for the innovations transfers to the linear process under summability of $(\|\Psi_k\|_{\mathcal{L}} \colon k \geq 0)$.

Proof. We only have to prove condition **(A3)**. It holds that

$$\begin{aligned}\sum_{t\geq 0} \nu_2\big(\mathcal{P}_0(X_t)\big) &\leq \sum_{t\geq 0}\sum_{k\geq 0} \nu_2\big(\mathcal{P}_0(\Psi_k(\delta_{t-k})\big) \\ &\leq \sum_{k\geq 0} \|\Psi_k\|_{\mathcal{L}} \sum_{t\geq 0} \nu_2\big(\mathcal{P}_0(\delta_{t-k})\big) < \infty.\end{aligned}$$

□

3.3 $L^2 - m$–approximable processes

[16] have used the concept of *$L^p - m$–approximability* for analyzing dependent functional data. Then a process $(X_t \colon t \in \mathbb{Z})$ is said to be $L^p - m$–approximable if X_t has representation (5) with i.i.d. innovations and

$$\sum_{m=1}^{\infty} \nu_p(X_0 - X_0^{(m)}) < \infty,$$

where $X_0^{(m)} = f(\varepsilon_0, \ldots, \varepsilon_{0-m+1}, \tilde{\varepsilon}_{-m}, \tilde{\varepsilon}_{-m-1}, \ldots)$ for some independent copy $(\tilde{\varepsilon}_t \colon t \in \mathbb{Z})$ of $(\varepsilon_t \colon t \in \mathbb{Z})$. In [16] it is shown that this concept applies to many stationary and non-stationary functional time series models, including, for example, functional ARCH. The concept is somewhat related to *near epoch dependence (NED)* often employed in the econometrics literature. See, e.g., [37]. If this condition holds with $p > 2$, then by a recent result of [2] a weak invariance principle for the partial sums process holds. This result has been sharpened by [24] who proved the same invariance principle under $p = 2$ and also under a milder coupling condition.

Theorem 8. *Suppose that $(X_t \colon t \in \mathbb{Z})$ is $L^2 - m$–approximable. Then Assumption 4 holds.*

Proof. We first note that under $L^2 - m$-approximability $X_s^{(s)}$ is independent of $\mathcal{G}_0$. First note that the construction yields $\nu^2\big(\mathcal{P}_0(X_s)\big) = E\|\mathcal{P}_0(X_s - X_s^{(s)})\|^2$. The right hand side can be bounded by

$$\begin{aligned} &2E\left(\|E[X_s - X_s^{(s)}|\mathcal{G}_0]\|^2 + \|E[X_s - X_s^{(s)}|\mathcal{G}_{-1}]\|^2\right) \\ &\quad \leq 4E\|X_0 - X_0^{(s)}\|^2 = 4\nu_2^2\big(X_0 - X_0^{(s)}\big). \end{aligned}$$

□

4 Proofs

Some crucial ideas in our proofs come from [36]. To extend these ideas to the functional setup, several non-trivial steps need to be added. In particular, showing existence of the spectral density operator is delicate. Here it arises as a limiting point in some appropriate Hilbert space—see Lemma 10 and the discussion thereafter. Next, verifying tightness requires extra efforts

(Lemmas 11, 14, 15, 16). One step in our proofs is to apply [36] to the projected Fourier transforms. For the proof of Theorem 1 we need to deal with the fact that each projection will come with its own exceptional set of frequencies where the CLT might fail and one difficulty is to make sure that the exceptional set for the functional DFT is still a null-set (Proposition 13).

4.1 Preliminary lemmas

We start with a lemma that discusses basic properties of the projection operators $\mathcal{P}_k$.

Lemma 9. *Let $X, Y \in L^2_H(\Omega)$.*

(i) For integers $k \neq \ell$ we have the strong orthogonality relation

$$\operatorname{Cov}\big(\mathcal{P}_k(X), \mathcal{P}_\ell(Y)\big) = 0;$$

(ii) If X is $\mathcal{G}_0$-measurable, then $X = \sum_{t\geq 0} \mathcal{P}_{-t}(X) + E[X|\mathcal{G}_{-\infty}]$ almost surely and in $L^2_{H_0}(\Omega)$;

(iii) If X is $\mathcal{G}_0$-measurable then under Assumption 1

$$\sum_{t\geq 0} E\|\mathcal{P}_{-t}(X)\|^2 = E\|X\|^2.$$

We notice that *(iii)* can be viewed as Parseval-type identity.

Proof. *(i)* We have to show that $E\Big[\langle \mathcal{P}_k(X), v\rangle \overline{\langle \mathcal{P}_\ell(Y), u\rangle}\Big] = 0 \quad \forall u, v \in H$. It holds that $\langle \mathcal{P}_k(X), v\rangle = \mathcal{P}_k(\langle X, v\rangle)$ and hence it is enough to restrict to the scalar case. The result follows then straight forwardly from elementary properties for conditional expectations of real valued random variables.

(ii) Note that $\sum_{t=0}^n \mathcal{P}_{-t}(X) = X - E[X|\mathcal{G}_{-n-1}]$. We consider the decaying sequence of σ-algebras $(\mathcal{G}_{-k}\colon k \geq 0)$. For any integrable random variable $X \in H$ the process $(E[X|\mathcal{G}_{-k}]\colon k \geq 0)$ is a reverse martingale with values in H (see, e.g., [6]). It converges a.s. to $E[X|\mathcal{G}_{-\infty}]$. If X is square integrable then convergence also holds in $L^2_{H_0}(\Omega)$.

(iii) By Assumption 1 and *(i)* of this lemma it follows that

$$\sum_{t\geq 0} E\|\mathcal{P}_{-t}(X)\|^2 = \lim_{n\to\infty} E\|X - E[X|\mathcal{G}_{-n}]\|^2 = E\|X\|^2.$$

□

Let $G = L^2_H(\Omega)$ and consider the Hilbert space $L^2_G((-\pi,\pi],\mathcal{B},\lambda)$, with $\mathcal{B}$ and λ being the Borel σ-field and the Lebesgue measure on $(-\pi,\pi]$, respectively. For simplicity we write $L^2_G((-\pi,\pi])$. This space is equipped with inner product $(V,W) = \int_{-\pi}^{\pi} E\langle V(\theta), W(\theta)\rangle d\theta$ and norm $|||V||| = \sqrt{(V,V)}$.

Lemma 10. *Define* $Z_n = Z_n(\theta) := \sum_{t=0}^n \mathcal{P}_0(X_t)e^{-\mathrm{i}t\theta}$. *Then* (Z_n) *is a Cauchy sequence in* $L^2_G((-\pi,\pi])$, *if and only if* $\sum_{t\geq 0} E\|\mathcal{P}_0(X_t)\|^2 < \infty$. *Moreover, under Assumption 1 the latter summability condition holds.*

We remark that this lemma provides a slightly weaker version of Assumption 2, part 1.

Proof. Using stationarity and the orthogonality of the functions $\theta \mapsto e^{-\mathrm{i}t\theta}$, $\theta \in (-\pi,\pi]$, $(t \in \mathbb{Z})$ we obtain for $m < n$

$$|||Z_n - Z_m|||^2 = \int_{-\pi}^{\pi} E\left\|\sum_{t=m+1}^{n} \mathcal{P}_0(X_t)e^{-\mathrm{i}t\theta}\right\|^2 d\theta = 2\pi \sum_{t=m+1}^{n} E\|\mathcal{P}_0(X_t)\|^2$$
$$= 2\pi \sum_{t=m+1}^{n} E\|\mathcal{P}_{-t}(X_0)\|^2.$$

The result follows by point *(iii)* of Lemma 9. □

It follows under Assumption 1 that there exists an element $Z \in L^2_G((-\pi,\pi])$ with $|||Z_n - Z||| \to 0$. This in turn has some important implications.

(P1) Since

$$|||Z|||^2 = \int_{-\pi}^{\pi} E\|Z(\theta)\|^2 d\theta < \infty,$$

we conclude that $E\|Z(\theta)\|^2 < \infty$ for all $\theta \in M_0 = (-\pi,\pi]\backslash N_0$ where $\lambda(N_0) = 0$. Hence, for all $\theta \in M_0$ the covariance operator $\mathcal{F}_\theta := \mathrm{Var}\big(Z(\theta)\big)$ is well defined, self-adjoint and non-negative definite. The denotation $\mathcal{F}_\theta$ is intentional. As we will see later *it is defining the spectral density operator* (compare to (4)). Since $\mathrm{tr}\big(\mathcal{F}_\theta\big) = E\|Z(\theta)\|^2$, this operator is trace class. For $\theta \in N_0$ we set $\mathcal{F}_\theta = 0$.

(P2) There exists a sequence (n_k) such that $E\|Z_{n_k}(\theta) - Z(\theta)\|^2 \to 0$ for all $\theta \in M_1 := (-\pi,\pi]\backslash N_1$, where $\lambda(N_1) = 0$.

(P3) By construction the mapping $\theta \mapsto Z(\theta) \in G$ is measurable, and the mapping from $G \to \mathcal{S}$ (the set of Hilbert-Schmidt operators on H) with

$Z(\theta) \mapsto \mathrm{Var}\big(Z(\theta)\big)$, is continuous. Hence, $\theta \to \mathcal{F}_\theta$ is measurable as a mapping from $(-\pi, \pi]$ to the space $\mathcal{S}$, which is known to be a separable Hilbert space. Consequently the integral in **(III)** of Theorem 1 is well defined.

The next lemma will be used in the proof of tightness and implies part **(II)** of Theorem 1.

Lemma 11. *Under Assumption 1 we have for all* $\theta \in M_2 = (-\pi, \pi] \backslash N_2$ *with* $\lambda(N_2) = 0$ *that*

$$\mathrm{tr}\big(\mathcal{F}_{n;\theta}\big) = \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) E\langle X_h, X_0\rangle e^{-\mathrm{i}h\theta} \to \mathrm{tr}\big(\mathcal{F}_\theta\big) < \infty.$$

Proof. Set $c_h := (2\pi)^{-1} \int_{-\pi}^{\pi} E\|Z(\theta)\|^2 e^{\mathrm{i}h\theta} d\theta$. Using (i) we infer from the Fejér-Lebesgue theorem that

$$\sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) c_h e^{-\mathrm{i}h\theta} \to E\|Z(\theta)\|^2 = \mathrm{tr}\big(\mathcal{F}_\theta\big) < \infty \quad \text{for almost all } \theta.$$

We define M_2 as the set of convergence points. We show now that $c_h = E\langle X_h, X_0\rangle$. Using Lemma 10 and continuity of $|||\cdot|||$, it can be readily shown that

$$c_h = \lim_{n\to\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} E\|Z_n(\theta)\|^2 e^{\mathrm{i}h\theta} d\theta.$$

Without loss of generality assume $h \geq 0$. Using stationarity we deduce

$$\begin{aligned}\frac{1}{2\pi} \int_{-\pi}^{\pi} E\|Z_n(\theta)\|^2 e^{\mathrm{i}h\theta} d\theta &= \sum_{t=0}^{n} \sum_{s=0}^{n} E\langle \mathcal{P}_0(X_t), \mathcal{P}_0(X_s)\rangle \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-\mathrm{i}(t-s-h)\theta} d\theta \\ &= \sum_{t=h}^{n} E\langle \mathcal{P}_0(X_t), \mathcal{P}_0(X_{t-h})\rangle = \sum_{t=h}^{n} E\langle \mathcal{P}_{-t}(X_0), \mathcal{P}_{-t}(X_{-h})\rangle.\end{aligned}$$

Since by Lemma 9 *(i)* the terms $\mathcal{P}_{-t}(X_0)$ and $\mathcal{P}_{-s}(X_{-h})$ are orthogonal in $L^2_{H_0}(\Omega)$ for $s \neq t$, it follows that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} E\|Z_n(\theta)\|^2 e^{\mathrm{i}h\theta} d\theta = E\left\langle \sum_{t=h}^{n} \mathcal{P}_{-t}(X_0), \sum_{s=h}^{n} \mathcal{P}_{-s}(X_{-h}) \right\rangle.$$

By Assumption 1 and Lemma 9 *(ii)*

$$\sum_{t=h}^{n} \mathcal{P}_{-t}(X_0) = \sum_{t=h}^{n} \mathcal{P}_{-t}(E[X_0|\mathcal{G}_{-h}]) \stackrel{L^2_{H_0}(\Omega)}{\longrightarrow} E[X_0|\mathcal{G}_{-h}]. \tag{6}$$

Similarly, $\sum_{s=h}^{n} \mathcal{P}_{-s}(X_{-h}) \stackrel{L^2_{H_0}(\Omega)}{\longrightarrow} E[X_{-h}|\mathcal{G}_{-h}] = X_{-h}$. And hence, by continuity of the inner product, $c_h = E\big\langle E[X_0|\mathcal{G}_{-h}], X_{-h}\big\rangle = E\langle X_h, X_0\rangle$. □

The next lemma yields property **(III)** of Theorem 1.

Lemma 12. *The operators* $\mathcal{F}_\theta$ *define the* spectral density operators *of* $(X_t\colon t \in \mathbb{Z})$ *at frequency* θ*. This is*

$$C_h = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{F}_\theta \, e^{\mathrm{i}h\theta} d\theta, \qquad \forall h \in \mathbb{Z}.$$

Proof. We have seen in (P3) that the mapping $\theta \mapsto \mathcal{F}_\theta$ is measurable. The integrand is valued in the separable Hilbert space $\mathcal{S}$. Since $\int_{-\pi}^{\pi} \|\mathcal{F}_\theta\|_{\mathcal{S}} d\theta \leq \int_{-\pi}^{\pi} \mathrm{tr}\big(\mathcal{F}_\theta\big) d\theta < \infty$ we know that $\mathcal{F}_\theta$ is strongly integrable and hence we can define (in the sense of a Bochner integral) $I = \int_{-\pi}^{\pi} \mathcal{F}_\theta \, e^{\mathrm{i}h\theta} d\theta$. Let $u, v \in H$. Since Bochner integrals are interchangeable with bounded linear operators we obtain

$$\begin{aligned}\langle I(v), u\rangle &= \int_{-\pi}^{\pi} \langle \mathcal{F}_\theta(v), u\rangle e^{\mathrm{i}h\theta} d\theta = \int_{-\pi}^{\pi} E\Big[\langle Z(\theta), u\rangle \overline{\langle Z(\theta), v\rangle}\Big] e^{\mathrm{i}h\theta} d\theta \\ &= \lim_{n\to\infty} \int_{-\pi}^{\pi} E\Big[\langle Z_n(\theta), u\rangle \, \overline{\langle Z_n(\theta), v\rangle}\Big] e^{\mathrm{i}h\theta} d\theta.\end{aligned}$$

The last equality can be deduced from $|||Z_n - Z||| \to 0$. Assume now without loss of generality that $h \geq 0$. Similar arguments as in Lemma 11 lead to

$$\frac{1}{2\pi}\langle I(v), u\rangle = \lim_{n\to\infty} E\bigg[\bigg\langle \sum_{t=0}^{n-h} \mathcal{P}_{-t}(X_h), u\bigg\rangle \bigg\langle v, \sum_{s=0}^{n-h} \mathcal{P}_{-s}(X_0)\bigg\rangle\bigg].$$

From (6) it follows that

$$\frac{1}{2\pi}\langle I(v), u\rangle = E\big[\langle E[X_h|\mathcal{G}_0], u\rangle \, \langle v, X_0\rangle\big] = \langle C_h(v), u\rangle.$$

Since u and v are arbitrary in H, we can infer that $I = 2\pi C_h$. □

We show next that the projections $\langle S_n(\theta), u\rangle/\sqrt{n}$, $u \in H$, converge weakly to $\langle S_0(\theta), u\rangle$, where $S_0(\theta) \sim \mathcal{CN}_H(0, \mathcal{F}_\theta)$ is the limiting complex Gaussian element and where $\mathcal{F}_\theta$ is defined as in (P1). The first step towards this result is given by the following proposition.

Proposition 13. *Under Assumption 1 there exists for all $u \in H$ a set $\widetilde{N} \subset (-\pi, \pi]$ with $\lambda(\widetilde{N}) = 0$, such that on $\widetilde{M} = (-\pi, \pi]\backslash\widetilde{N}$ the following holds:*

(a) $\lim_{n\to\infty} \mathrm{Var}(\langle S_n(\theta), u\rangle)/n = \lim_{n\to\infty}\langle \mathcal{F}_{n;\theta}(u), u\rangle = \langle \mathcal{F}_\theta(u), u\rangle$;

(b) $\langle S_n(\theta), u\rangle/\sqrt{n} \xrightarrow{d} \mathcal{CN}_{\mathbb{C}}(0, \langle \mathcal{F}_\theta(u), u\rangle)$;

Proof. We first show that there exists for all $u \in H$ a set $N_u \subset (-\pi, \pi]$ with $\lambda(N_u) = 0$, such that on $M_u = (-\pi, \pi]\backslash N_u$ **(a)** and **(b)** hold.

Let $u \in$ and let $\mathcal{G}_k(u)$ be the filtration of the process $(\langle X_t, u\rangle \colon t \in \mathbb{Z})$. From the results in [36] we obtain that $\mathrm{Var}(\langle S_n(\theta), u\rangle)/n \to f^u(\theta)$ for some function $f^u(\theta)$ which is finite on M_u. More precisely, slightly adapting the proofs of Lemmas 4.1. and 4.2. in their article we obtain that the $L^2(\Omega)$ limit

$$D^u(\theta) := \lim_{n\to\infty} \sum_{t=0}^{n} \mathcal{P}_0(\langle X_t, u\rangle)e^{-\mathrm{i}t\theta} = \lim_{n\to\infty}\langle Z_n(\theta), u\rangle \tag{7}$$

exists on M_u and that $f^u(\theta) = \mathrm{Var}\big(D^u(\theta)\big)$. (Directly using their arguments would require to use the projection operator $\mathcal{P}_0^u(\cdot) = E[\,\cdot\,|\mathcal{G}_0(u)] - E[\,\cdot\,|\mathcal{G}_{-1}(u)]$.) We assume without loss of generality that $M = M_0 \cap M_1$ is a subset of M_u, otherwise replace M_u by $M_u \cap M$. We now determine $f^u(\theta)$. By result (P2) of Section 4.1 it follows that $E(\langle Z_{n_k}(\theta), u\rangle - \langle Z(\theta), u\rangle)^2 \to 0$ for every $u \in H$ and all $\theta \in M_1$. Hence, by result (P1) in the same section, we get

$$\mathrm{Var}\big(\langle Z_{n_k}(\theta), u\rangle\big) \to \langle \mathcal{F}_\theta(u), u\rangle < \infty,$$

for all $\theta \in M$ and all $u \in H$, which implies on M_u the relation $f^u(\theta) = \langle \mathcal{F}_\theta(u), u\rangle < \infty$. This shows part **(a)** on M_u.

We have $E[\langle X_t, u\rangle|\mathcal{G}_{-\infty}] = \langle E[X_t|\mathcal{G}_{-\infty}], u\rangle$ and by Assumption 1 this is equal to zero. The tower property of conditional expectations implies $E[\langle X_t, u\rangle|\mathcal{G}_{-\infty}(u)] = 0$ and hence on M_u **(b)** directly follows from [36]. Let us note that their CLT result is stated for real valued time series, but this requirement is not needed. Hence we can apply it for the time series $(\langle X_t, u\rangle \colon t \geq 1)$ which takes values in $\mathbb{C}$ when $u \in H$.

It remains to prove that for all u in H we can find a common exceptional set of Lebesgue measure 0. To this end let H' be a dense and countable subset of H. We set $\widetilde{M} = \cap_{u' \in H'} M_u$. Then $(-\pi, \pi] \backslash \widetilde{M}$ has Lebesgue measure 0. Furthermore, for all $u' \in H'$ and $\theta \in \widetilde{M}$ **(a)** and **(b)** hold. The objective is now to extend this result to all $u \in H$. For **(a)** we observe that

$$\begin{aligned}
&\big|\langle \mathcal{F}_{n;\theta}(u), u\rangle - \langle \mathcal{F}_\theta(u), u\rangle\big| \\
&\quad \leq \big|\langle \mathcal{F}_{n;\theta}(u), u\rangle - \langle \mathcal{F}_{n;\theta}(u'), u'\rangle\big| + \big|\langle \mathcal{F}_\theta(u'), u'\rangle - \langle \mathcal{F}_\theta(u), u\rangle\big| \\
&\qquad + \big|\langle \mathcal{F}_{n;\theta}(u'), u'\rangle - \langle \mathcal{F}_\theta(u'), u'\rangle\big| \\
&\quad \leq \Big[\mathrm{tr}\big(\mathcal{F}_{n;\theta}\big) + \mathrm{tr}\big(\mathcal{F}_\theta\big)\Big] \times \Big[\big(\|u\| + \|u'\|\big) \times \|u - u'\|\Big] \\
&\qquad + \big|\langle \mathcal{F}_{n;\theta}(u'), u'\rangle - \langle \mathcal{F}_\theta(u'), u'\rangle\big|.
\end{aligned}$$

Since we can assume without loss of generality that $M_2 \subset \widetilde{M}$, it follows that for all $\theta \in \widetilde{M}$

$$\limsup_{n\to\infty} \big|\langle \mathcal{F}_{n;\theta}(u), u\rangle - \langle \mathcal{F}_\theta(u), u\rangle\big| \leq 4\varepsilon(\|u\| + 1)\,\mathrm{tr}\big(\mathcal{F}_\theta\big),$$

if $\|u - u'\| \leq \varepsilon \leq 1$. Since ε can be chosen arbitrarily small result **(a)** follows.

The proof of part **(b)** follows along similar lines of arguments. Just compare the characteristic functions of the real and complex part of $\langle S_n(\theta), u\rangle$ to the corresponding normal ones. □

4.2 Tightness

The following technical lemma will be crucial for showing tightness.

Lemma 14. *Consider sequences $(p_j^{(n)} : j \geq 1)$, $n \geq 0$, with the following properties: (a) $p_j^{(n)} \geq 0$ for all j, n; (b) $\lim_n p_j^{(n)} = p_j^{(0)}$; (c) $\sum_{j\geq 1} p_j^{(0)} = p < \infty$; (d) $\lim_n \sum_{j\geq 1} p_j^{(n)} = p$; (e) $\sum_{j\geq 1} p_j^{(n)} < \infty$ for all $n \geq 1$. Then*

$$\lim_{m\to\infty} \sup_n \sum_{j>m} p_j^{(n)} = 0.$$

Proof. Fix an $\varepsilon > 0$. We have to show that for $m \geq m(\varepsilon)$ we have $\sum_{j\geq m} p_j^{(n)} < \varepsilon$ for all $n \geq 1$.

By (c) we can choose $m_1 = m_1(\varepsilon)$ such that $\sum_{j\geq m} p_j^{(0)} < \varepsilon/3$ for all $m \geq m_1$. Furthermore, by (b) we can choose a large enough $n_1 = n_1(\varepsilon)$ such that for all $n \geq n_1$ we have $|\sum_{j=1}^{m_1} p_j^{(0)} - \sum_{j=1}^{m_1} p_j^{(n)}| < \varepsilon/3$. Next, by possibly further enlarging n_1 we deduce from (c) and (d) that $|\sum_{j\geq 1} p_j^{(n)} - \sum_{j\geq 1} p_j^{(0)}| < \varepsilon/3$. Consequently, for $n \geq n_1$, we have

$$\begin{aligned}\sum_{j>m_1} p_j^{(n)} &= \sum_{j\geq 1} p_j^{(n)} - \sum_{j=1}^{m_1} p_j^{(n)} \\ &\leq |\sum_{j\geq 1} p_j^{(n)} - \sum_{j\geq 1} p_j^{(0)}| + |\sum_{j\geq 1} p_j^{(0)} - \sum_{j=1}^{m_1} p_j^{(0)}| + |\sum_{j=1}^{m_1} p_j^{(0)} - \sum_{j=1}^{m_1} p_j^{(n)}| \\ &< \varepsilon.\end{aligned}$$

Because of (a) this bound is still valid for all $m \geq m_1$. For the n_1 just chosen, we can find an $m_2 = m_2(\varepsilon)$, such that $\sum_{j>m_2} p_j^{(n)} < \varepsilon$ for all $n \leq n_1$. This is because of (d) and (e) we know that $\sup_{n\geq 1} \sum_{j\geq 1} p_j^{(n)} < \infty$. And again, because of (a) we know also that $\sum_{j>m} p_j^{(n)} < \varepsilon$ for all $m \geq m_2$ and $n \leq n_1$. Hence, set $m(\varepsilon) = \max\{m_1, m_2\}$. $\square$

Lemma 15. *Take some ONB $(v_j \colon j \geq 1)$ of H. Lemma 14 applies with $p_j^{(n)} = \langle \mathcal{F}_{n;\theta}(v_j), v_j \rangle$, $p_j^{(0)} = \langle \mathcal{F}_\theta(v_j), v_j \rangle$ for all $\theta \in \widetilde{M}$.*

Proof. We can assume that $(v_j \colon j \geq 1)$ belongs to the dense subset H' which was defined in Proposition 13. Relation (a) is trivial. Relation (b) follows from part **(a)** of Proposition 13. Relation (c) holds because $\mathcal{F}_\theta$ is nuclear on $\widetilde{M}$. And similarly relation (e) holds because $\mathcal{F}_{n;\theta}$ is nuclear for any n. Finally note that (d) can be reformulated as $\operatorname{tr}(\mathcal{F}_{n;\theta}) \to \operatorname{tr}(\mathcal{F}_\theta)$. By Lemma 11 this holds for almost all $\theta \in M_2 \subset \widetilde{M}$. $\square$

Lemma 16. *Under Assumptions 1 the sequence $(S_n(\theta)/\sqrt{n} \colon n \geq 1)$ is tight for all $\theta \in \widetilde{M}$.*

Proof. Let $\varepsilon > 0$. We consider the sequences $0 < \ell_k \nearrow \infty$ and $0 < N_k \nearrow \infty$ and define

$$K = \bigcap_{k=1}^{\infty} \left\{ x \in H : \sum_{j>N_k} |\langle v_j, x \rangle|^2 \leq \frac{1}{\ell_k} \right\}.$$

Just as in [3] (p. 52) we can see that it is a compact subset of H. We now have that

$$\Pr\left(S_n(\theta)/\sqrt{n} \in K\right) \geq 1 - \sum_{k=1}^{\infty} \ell_k \sum_{j>N_k} E\left|\langle S_n(\theta)/\sqrt{n}, v_j\rangle\right|^2$$

$$= 1 - \sum_{k=1}^{\infty} \ell_k \sum_{j>N_k} \langle \mathcal{F}_{n;\theta}(v_j), v_j\rangle,$$

where we used the σ-subadditivity and the Markov inequality. By Lemma 15 we know that

$$\sup_n \sum_{j\geq m} \langle \mathcal{F}_{n;\theta}(v_j), v_j\rangle \to 0 \quad (m \to \infty).$$

Therefore, for any $\varepsilon > 0$, we can choose increasing sequences (ℓ_k) and (N_k) such that

$$\ell_k \sum_{j>N_k} \langle \mathcal{F}_{n;\theta}(v_j), v_j\rangle \leq \varepsilon 2^{-k}.$$

□

4.3 Proofs of Lemma 4 and Theorems 1, 2 and 5

Proof of Lemma 4. It is easy to see that **(A3)** implies **(A1)** for all $\theta \in (-\pi, \pi]$. Now we prove **(A2)**. By part *(iii)* of Lemma 9 we have

$$E\|E[S_n(\theta)|\mathcal{G}_0]\|^2 = \sum_{j\geq 0} E\|\mathcal{P}_{-j}(E[S_n(\theta)|\mathcal{G}_0])\|^2 = \sum_{j\geq 0} E\|\mathcal{P}_{-j}(S_n(\theta))\|^2.$$

Therefore we may conclude that

$$E\|E[S_n(\theta)|\mathcal{G}_0]\|^2 \leq \sum_{j=0}^{\infty} E \sum_{s,t=1}^{n} |\langle \mathcal{P}_{-j}(X_t), \mathcal{P}_{-j}(X_s)\rangle|$$

$$= \sum_{j=0}^{\infty} \sum_{s,t=1}^{n} E|\langle \mathcal{P}_{-j}(X_t), \mathcal{P}_{-j}(X_s)\rangle|$$

$$\leq \sum_{s=1}^{n} \sum_{j=0}^{\infty} \left(\sum_{t=0}^{\infty} \nu_2(\mathcal{P}_0(X_{t+j}))\right) \nu_2(\mathcal{P}_0(X_{s+j})) = o(n).$$

With the help of Lemma 9 we obtain

$$\|C_h\|_{\mathcal{S}} = \|EX_h \otimes X_0\|_{\mathcal{S}} \leq \sum_{k=0}^{\infty} \|E\mathcal{P}_{-k}(X_h) \otimes \mathcal{P}_{-k}(X_0)\|_{\mathcal{S}}$$
$$\leq \sum_{k=0}^{\infty} \nu_2\left(\mathcal{P}_0(X_{h+k})\right) \nu_2\left(\mathcal{P}_0(X_k)\right).$$

Consequently

$$\sum_{h=-\infty}^{\infty} \|C_h\|_{\mathcal{S}} \leq \left(\sum_{h=0}^{\infty} \nu_2\left(\mathcal{P}_0(X_h)\right)\right)^2 < \infty.$$

This proves *(ii)* and implies that expression (1) is well defined and continuous.

Since now we know that Assumption 2 holds, we infer from Theorem 2 that the Cèsaro means $\mathcal{F}_{n;\theta} \stackrel{\mathrm{w}}{\to} \mathcal{F}_\theta$, with $\mathcal{F}_\theta$ defined as in (4). But *(ii)* implies that also the regular partial sums $\sum_{|h|<n} C_h e^{-\mathrm{i}h\theta}$ converge to the same limit. □

Proof of Theorem 1. Parts **(II)** and **(III)** of Theorem 1 follow directly from Lemmas 11 and 12. Part **(I)** can be deduced from the polarization identity for self-adjoint operators Γ

$$\langle \Gamma(x), y\rangle = \frac{1}{4}\big[\langle \Gamma(x+y), x+y\rangle - \langle \Gamma(x-y), x-y\rangle$$
$$+ \mathrm{i}\langle \Gamma(x+\mathrm{i}y), x+\mathrm{i}y\rangle - \mathrm{i}\langle \Gamma(x-\mathrm{i}y), x-\mathrm{i}y\rangle\big],$$

and from part **(a)** of Proposition 13. Next, the asymptotic normality of $S_n(\theta)/\sqrt{n}$ for all $\theta \in \widetilde{M}$ follows from the corresponding convergence of the projections (Proposition 13, part **(b)**) and the tightness shown in Lemma 16.

Finally, the asymptotic independence relation **(IV)** can be obtained by verifying that the projections $(\langle S_n(\theta), u\rangle, \langle S_n(\theta'), u'\rangle)/\sqrt{n}$ converge for any u and u' in H to a bivariate complex Gaussian vector with independent components. Similarly as remarked by [36] this amounts to combining the scalar proof with a Wold argument. We sketch the main steps:

Define $\sigma_1^2 = \sigma_1^2(\theta, u) = \langle \mathcal{F}_\theta(u), u\rangle$ and $\sigma_2^2 = \sigma_2^2(\theta', u') = \langle \mathcal{F}_{\theta'}(u'), u'\rangle$ and assume below that $u, u' \in H$ and $\theta, \theta' \in \widetilde{M}$ (as defined in Proposition 1) with the additional constraints $\theta \neq \theta'$ and $\theta \neq -\theta'$ (which only exclude an additional null-set of $(-\pi,\pi]^2$). We will prove that for any such choice of u,u',θ,θ'

$$(\langle S_n(\theta),u\rangle + \langle S_n(\theta'),u'\rangle)\sqrt{n} \xrightarrow{d} \mathcal{CN}_{\mathbb{C}}(0,\sigma_1^2+\sigma_2^2,c), \tag{8}$$

where $c := \big(\langle \mathcal{F}_\theta(\bar{u}),u\rangle I\{\theta\in\{0,\pi\}\} + \langle \mathcal{F}_\theta(\bar{u}),u\rangle I\{\theta'\in\{0,\pi\}\}\big)$. From this the asymptotic independence and joint Gaussianity can be deduced.

From Proposition 1 and its proof we know that $(\langle Z_n(\theta),u\rangle\colon n\geq 1)$ given in (7) converges in $L^2(\Omega)$ to some variable $D^u(\theta)$ with variance $\sigma_1^2(\theta,u)<\infty$. This provides the ingredient for the construction of the martingale approximation as given in [36, p. 11f.]. Following their construction we obtain stationary and ergodic martingale difference sequences $(D_k^u(\theta)\colon k\geq 1)$ such that $D_k^u(\theta)\stackrel{d}{=}D^u(\theta)$ and such that

$$\langle S_n(\theta),u\rangle = \sum_{k=1}^n D_k^u(\theta)e^{-ik\theta} + R_n \quad \text{and} \quad \langle S_n(\theta'),u'\rangle = \sum_{k=1}^n D_k^{u'}(\theta')e^{-ik\theta'} + R_n',$$

where $E|R_n|^2 + E|R_n'|^2 = o(n)$. (See also the discussion after our Theorem 2.) For simplicity we write henceforth $D_k = D_k^{u'}(\theta')$ and $D_k' = D_k^{u'}(\theta')$ and set $Y_k := D_k e^{-ik\theta} + D_k' e^{-ik\theta'}$. Then $\sum_{k=1}^n Y_k$ is itself a martingale to which we will apply a martingale CLT. Due to

$$S_n := \langle S_n(\theta),u\rangle + \langle S_n(\theta'),v\rangle = \sum_{k=1}^n Y_k + R_n + R_n'$$

this CLT carries over to S_n. The Lindeberg condition for this martingale is easily checked. As for the asymptotic *variance* we have

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^n E[|Y_k|^2\,|\mathcal{G}_{k-1}] &= \frac{1}{n}\sum_{k=1}^n E[|D_k|^2\,|\mathcal{G}_{k-1}] + \frac{1}{n}\sum_{k=1}^n E[D_k\overline{D_k'}\,|\mathcal{G}_{k-1}]e^{-ik(\theta-\theta')} \\ &\quad + \frac{1}{n}\sum_{k=1}^n E[D_k'\overline{D_k}\,|\mathcal{G}_{k-1}]e^{-ik(\theta'-\theta)} + \frac{1}{n}\sum_{k=1}^n E[|D_k'|^2\,|\mathcal{G}_{k-1}] \\ &\xrightarrow{a.s.} \quad E|D_1|^2 + E|D_1'|^2 = \sigma_1^2+\sigma_2^2,\end{aligned}$$

where we applied the ergodic theorem and Lemma 5 in [43] for the middle terms. This Lemma applies if $\theta\neq\theta'$. If $\theta+\theta'\neq 0$ we obtain by the same lemma for the asymptotic *relation*

$$\frac{1}{n}\sum_{k=1}^{n} E[Y_k^2\,|\mathcal{G}_{k-1}] = \frac{1}{n}\sum_{k=1}^{n} E[D_k^2\,|\mathcal{G}_{k-1}]e^{-2ik\theta}$$
$$+\frac{1}{n}\sum_{k=1}^{n} E[2D_kD_k' e^{-ik(\theta+\theta')}\,|\mathcal{G}_{k-1}] + \frac{1}{n}\sum_{k=1}^{n} E[(D_k')^2\,|\mathcal{G}_{k-1}]e^{-2ik\theta'}$$
$$\xrightarrow{a.s.} \quad ED_1^2 I\{\theta\in\{0,\pi\}\} + E(D_1')^2 I\{\theta\in\{0,\pi\}\} = c.$$

□

Proof of Theorem 2. Let $\theta_0 \in (-\pi,\pi]$ be such that Assumption 2 is satisfied. For notational convenience we use for the proof $\theta=\theta_0$. Due to relation (3) and stationarity, we have that

$$\begin{aligned}\frac{E\|S_n(\theta)\|^2}{n} &= \frac{1}{n}\sum_{k=1}^{n} E\|Z_{n-k}^{(k)}(\theta)\|^2 + \frac{1}{n}E\big\|E[S_n(\theta)|\mathcal{G}_0]\big\|^2\\ &= E\|Z^{(1)}(\theta)\|^2 \;+\; o(1),\end{aligned}$$

hence **(II)** holds. Moreover, we have

$$\frac{E|\langle S_n(\theta),u\rangle|^2}{n} = E|\langle Z^{(1)}(\theta),u\rangle|^2 \;+\; o(1)$$

and with the polarization identity this shows **(I)**. The CLT for the projections $\langle S_n(\theta),u\rangle/\sqrt{n}$, as well as its bivariate version when considering two frequencies, can be shown by the same martingale approximation as in the proof of Theorem 1. Also the proof of tightness of $S_n(\theta)/\sqrt{n}$ doesn't require any new ideas and can be shown along the same lines as in the proof of Theorem 1. □

Proof of Theorem 5. The first step is again to show that $S_n(\theta_n)$ can be approximated by a martingale just as in the case when the frequency is fixed. From Lemma 4 we know that both conditions of Assumption 2 are satisfied for any frequency. We may thus define $Z^{(k)}(\theta) := \sum_{t=0}^{\infty}\mathcal{P}_k(X_{k+t})e^{-\mathrm{i}t\theta}$ (as a limit in $L_H^2(\Omega)$) and the variables $\widetilde{S}_n(\theta) := \sum_{k=1}^{n} Z^{(k)}(\theta)e^{-\mathrm{i}k\theta_n}$. Then

$$S_n(\theta_n) - \widetilde{S}_n(\theta) = \sum_{k=1}^{n}\Big(\mathcal{P}_k(S_n(\theta_n)) - Z^{(k)}(\theta)e^{-\mathrm{i}k\theta_n}\Big) + E[S_n(\theta_n)|\mathcal{G}_0].$$

From the proof of Lemma 4 we infer that $\|E[S_n(\theta_n)|\mathcal{G}_0]\| = o(n)$. Furthermore, using orthogonality, we get

$$\begin{aligned}
&\nu_2^2\left(\sum_{k=1}^{n}\left(\mathcal{P}_k(S_n(\theta_n)) - Z^{(k)}(\theta)e^{-\mathrm{i}k\theta_n}\right)\right)\\
&\quad = \sum_{k=1}^{n}\nu_2^2\left(\sum_{t=k}^{n}\mathcal{P}_k(X_t)e^{-\mathrm{i}t\theta_n} - Z^{(k)}(\theta)e^{-\mathrm{i}k\theta_n}\right)\\
&\quad = \sum_{k=1}^{n}\nu_2^2\left(Z_{n-k}^{(k)}(\theta_n) - Z^{(k)}(\theta)\right).
\end{aligned}$$

It is easy to show that **(A3)** and $\theta_n \to \theta$ imply that the last term is $o(n)$.

Fix some arbitrary $u, u' \in H$, define σ_1^2 and σ_2^2 as in the proof of Theorem 1. Moreover, we set $D_k := \lim_{n\to\infty}\langle Z_n^{(k)}(\theta_n), u\rangle$ and $D_k' = \lim_{n\to\infty}\langle Z_n^{(k)}(\theta_n'), u'\rangle$ and $Y_k^n := D_k e^{-\mathrm{i}k\theta_n} + D_k' e^{-\mathrm{i}k\theta_n'}$. We show that

$$(\langle\widetilde{S}_n(\theta), u\rangle + \langle\widetilde{S}_n(\theta'), u'\rangle)/\sqrt{n} = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} Y_k^n \xrightarrow{d} \mathcal{CN}_{\mathbb{C}}(0, \sigma_1^2 + \sigma_2^2).$$

To this end we apply a martingale CLT to the array of martingale differences $(Y_k^n\colon 1 \le k \le n)$, $n \ge 1$. The Lindeberg condition is easily checked and it remains again to determine the asymptotic *variance* and *relation*. In analogy to the proof of Theorem 1 we get

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^{n} E[|Y_k^n|^2\,|\mathcal{G}_{k-1}] &= \frac{1}{n}\sum_{k=1}^{n} E[|D_k|^2\,|\mathcal{G}_{k-1}] + \frac{1}{n}\sum_{k=1}^{n} E[D_k\overline{D_k'}\,|\mathcal{G}_{k-1}]e^{-\mathrm{i}k(\theta_n-\theta_n')}\\
&\quad + \frac{1}{n}\sum_{k=1}^{n} E[D_k'\overline{D_k}\,|\mathcal{G}_{k-1}]e^{-\mathrm{i}k(\theta_n'-\theta_n)} + \frac{1}{n}\sum_{k=1}^{n} E[|D_k'|^2\,|\mathcal{G}_{k-1}],
\end{aligned}$$

and by the exact same arguments the first and the last term converge to σ_1^2 and σ_2^2, respectively. In order to deal with the middle term we set $U_k := E[D_k\overline{D_k'}\,|\mathcal{G}_{k-1}]$ and our target is to prove that

$$\frac{1}{n}\sum_{k=1}^{n} U_k e^{-\mathrm{i}k(\theta_n-\theta_n')} \xrightarrow[n\to\infty]{\Pr} 0.$$

The third term can be treated analogously. For some $m \geq 1$ set $U_k^m = E[U_k|\varepsilon_{k-1}, \ldots, \varepsilon_{k-m-1}]$. Then $(U_k^m \colon k \geq 1)$ is a strictly stationary and m-dependent sequence which satisfies $U_k^m \xrightarrow{L^1} E[U_k|\varepsilon_{k-1}, \ldots] = U_k$ $(m \to \infty)$. Thus, for any $\epsilon > 0$ we can find a large enough $m \geq 1$ such that $E|U_1 - U_1^m| \leq \epsilon^2$. Now we set $\alpha_n := \theta_n - \theta_n'$ and note that by our assumptions $\sum_{k=1}^n e^{-\mathrm{i}k\alpha_n} = 0$. It follows that

$$
\begin{aligned}
&\Pr\left(\Big|\frac{1}{n}\sum_{k=1}^n U_k e^{-\mathrm{i}k\alpha_n}\Big| > \epsilon\right) \\
&\quad \leq \Pr\left(\Big|\frac{1}{n}\sum_{k=1}^n (U_k - U_k^m) e^{-\mathrm{i}k\alpha_n}\Big| > \epsilon/2\right) + \Pr\left(\Big|\frac{1}{n}\sum_{k=1}^n U_k^m e^{-\mathrm{i}k\alpha_n}\Big| > \epsilon/2\right) \\
&\quad \leq \frac{\mathrm{E}\,|U_1 - U_1^m|}{\epsilon/2} + \Pr\left(\Big|\frac{1}{n}\sum_{j=1}^m \sum_{\substack{1\leq k\leq n\\ m|(k-j)}} (U_k^m - EU_k^m) e^{-\mathrm{i}k\alpha_n}\Big| > \epsilon/2\right) \\
&\quad \leq \epsilon/2 + \sum_{j=1}^m \Pr\left(\Big|\frac{1}{n/m} \sum_{\substack{1\leq k\leq n\\ m|(k-j)}} (U_k^m - EU_k^m) e^{-\mathrm{i}k\alpha_n}\Big| > \epsilon/2\right).
\end{aligned} \tag{9}
$$

Here $m|\ell$ signifies that m is a divisor of ℓ. For $k = j, j+m, j+2m, \ldots$ the terms U_k^m are i.i.d. Thus, we can apply the weighted law of large numbers in [38] to obtain that for large enough $n \geq n_0(\epsilon, m)$ each of the probabilities in (9) is $\leq \epsilon/(2m)$.

By the same kind of arguments it can be seen that the asymptotic relation is zero.

Finally, in order to show the tightness of $S_n(\theta_n)/\sqrt{n}$, we need to apply Lemma 14 with $p_j^{(n)} = E|\langle S_n(\theta_n), v_j\rangle|^2/n$, and $p_j^{(0)} = E|\langle \mathcal{F}_\theta(v_j), v_j\rangle| = E|\langle Z^{(1)}(\theta), v_j\rangle|^2$, where $(v_j)_{j\geq 1}$ is an arbitrary ONB. Using the previous approximation $S_n(\theta_n) = \widetilde{S}_n(\theta) + o_{L^2}(\sqrt{n})$, it is easy to check that conditions (a)-(e) are satisfied. With this we can conclude along the lines of Section 4.2. □

Acknowledgement

We would like to thank two anonymous referees and the associate editor for their constructive comments, which helped to improve the presentation of the paper. This research was supported by the Communauté française de Belgique—Actions de Recherche Concertées (2010–2015) and from the Interuniversity Attraction Poles Programme (IAP-network P7/06), Belgian Science Policy Office. Clément Cerovecki acknowledges support from the F.R.S.-FNRS Fond de la Recherche Scientifique, Rue d'Egmont 5, B-1000 Bruxelles.

References

[1] Aue, A., Dubart Norinho, D., and Hörmann, S. (2015). *On the prediction of stationary functional time series.* Journal of the American Statistical Association, 110, 509, 378–392.

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

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

[4] Brockwell, P.J. and Davis, R.A. (1991). Time Series: Theory and Methods. Springer, New York.

[5] Bradley, R.C. (2005). *Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions.* Probability Surveys Vol. 2, 107–144.

[6] Chatterji, S.D. (1964). *A note on the convergence of Banach-space valued martingales.* Mathematische Annalen, Volume 153, Issue 2, pp 142–149.

[7] Cuevas, A. (2014). *A partial overview of the theory of statistics with functional data.* Journal of Statistical Planning and Inference 147, 1-23.

[8] Damon, J., Guillas, S. (2005). *Estimation and simulation of autoregressive Hilbertian processes with exogenous variables.* Statistical Inference for Stochastic Processes, 8, 185–204.

[9] Didericksen, D., Kokoszka, P., and Zhang, X. (2012). *Empirical properties of forecasts with the functional autoregressive model.* Computational Statistics, 27, 285–298.

[10] Ferraty, F., Goia, A., and Vieu, P. (2002). *Functional nonparametric model for time series: a fractal approach for dimension reduction.* Test 11, 317–344.

[11] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice.* Springer Science & Business Media.

[12] Goia, A. and Vieu, P. (2016). *An introduction to recent advances in high/infinite dimensional statistics.* Journal of Multivariate Analysis 146 (2016): 1–6.

[13] Guillas, S., Bakare, A., Morley, J., Simons, R. (2011). *Functional autoregressive forecasting of long-term seabed evolution.* Journal of Coastal Conservation, 15, 337–351.

[14] Hörmann, S., Kidziński, L. and Hallin, M. (2015). *Dynamic Functional Principal Component.* Journal of the Royal Statistical Society: Series B, 77, 319–348.

[15] Hörmann, S., Kidziński, L. and Kokoszka, P. (2014). *Estimation in functional lagged regression.* Journal of Time Series Analysis, 36, 541–561.

[16] Hörmann, S., Kokoszka, P. (2010). *Weakly dependent functional data.* The Annals of Statistics, 38, 1845–1884.

[17] Hörmann, S., Kokoszka, P. and Nisol, G. (2016). *Detection of periodicity in functional time series.* arXiv:1607.02017, working paper.

[18] Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications.* (Vol. 200). Springer Science & Business Media.

[19] Horváth, L. and Kokoszka, P. and Reeder, R. (2013). *Estimation of the mean of functional time series and a two-sample problem.* JRSSB, 75.1, 103–122.

[20] Horváth, L., Kokoszka, P., and Rice, G. (2014). *Testing stationarity of functional time series.* Journal of Econometrics, 179, 66–82.

[21] Horváth, L., Rice, G., and Whipple, S. (2016). *Adaptive bandwidth selection in the estimation of the long run covariance of functional time series.* Computational Statistics and Data Analysis, 100, 676–693.

[22] Hsing, T., and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators.* John Wiley & Sons.

[23] Hyndman, R.J. and Shang, H.L. (2009). *Forecasting functional time series.* Journal of the Korean Statistical Society, 38, No. 3, 199–211.

[24] Jirak, M. (2013). *On weak invariance principles for sums of dependent random functionals.* Statistics and Probability Letters, 83, 2291–2296.

[25] Kargin, V. and Onatski, A. (2008). *Curve forecasting by functional autoregression.* Journal of Multivariate Analysis, 99, 2508–2526.

[26] Klepsch, J and Klüppelberg, C. (2016). *An Innovations Algorithm for the prediction of functional linear processes.* arXiv:1607.05874, working paper.

[27] Kokoszka, P. and Mikosch, T. (2000). *The periodogram at the Fourier frequencies.* Stochastic Processes and their Applications, 86, 49–79.

[28] Kowal, D.R., Matteson, D.S., and Ruppert, D. (2016). *Functional autoregression for sparsely sampled data.* arXiv:1603.02982, working paper.

[29] Liebl, D. (2013). *Modeling and forecasting electricity prices: A functional data perspective.* The Annals of Applied Statistics, 7, 1562–1592.

[30] Mas, A. (2007). *Weak convergence in the functional autoregressive model.* Journal of Multivariate Analysis, 98, 1231–1261.

[31] Merlevède, F., Peligrad, M. and Utev, S. (1997). *Sharp Conditions for the CLT of Linear Processes in a Hilbert Space.* Journal of Theoretical Probability, 10, No. 3, 681–693.

[32] Mikusiński, J. (1978). *The Bochner Integral.* Birkhauser, Basel.

[33] Panaretos, V.M. and Tavakoli, S. (2013). *Fourier analysis of stationonary time series in function spaces.* The Annals of Statistics, 41, No. 2, 568–603.

[34] Panaretos, V.M. and Tavakoli, S. (2013). *Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series.* Stochastic Processes and their Applications, 123, 2779–2807.

[35] Paparoditis, E. (2016) *Sieve Bootstrap for Functional Time Series.* arXiv:1609.06029, working paper.

[36] Peligrad, M. and Wu, W.B. (2010). *Central limit theorem for Fourier transforms of stationary processes.* The Annals of Probability, 38, No. 5, 2009–2022.

[37] Pötscher, B.M. and Prucha, I.R. (1997). *Dynamic Nonlinear Econometric Models.* Springer.

[38] Pruitt, W. E.(1966), *Summability of independent random variables.* Journal of Mathematical and Mechanics 15: 769-776.

[39] Račkauskas, A. and Suquet, Ch. (2010). *On limit theorems for Banach-space-valued linear processes.* Lithuanian Mathematical Journal, 50, No. 1, 71–87.

[40] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis.* Springer, New York, second edition.

[41] Ruiz-Medina, M. D., Romano, E., and Fernàndez-Pascual, R. (2016). *Plug-in prediction intervals for a special class of standard ARH(1) processes.* Journal of Multivariate Analysis, 146, 138–150.

[42] Walker, A.M. (1965). *Some asymptotic results for the periodogram of a stationary time series.* Journal of the Australian Mathematical Society, 5, No. 1, 107–128.

[43] Wu, W.B. (2004). *Fourier transforms of stationary processes.* Proceeding of the American society, 133, 285–293.