FOURIER ANALYSIS OF STATIONARY TIME SERIES IN FUNCTION SPACE[1]

BY VICTOR M. PANARETOS AND SHAHIN TAVAKOLI

Ecole Polytechnique Fédérale de Lausanne

We develop the basic building blocks of a frequency domain framework for drawing statistical inferences on the second-order structure of a stationary sequence of functional data. The key element in such a context is the spectral density operator, which generalises the notion of a spectral density matrix to the functional setting, and characterises the second-order dynamics of the process. Our main tool is the functional Discrete Fourier Transform (fDFT). We derive an asymptotic Gaussian representation of the fDFT, thus allowing the transformation of the original collection of dependent random functions into a collection of approximately independent complex-valued Gaussian random functions. Our results are then employed in order to construct estimators of the spectral density operator based on smoothed versions of the periodogram kernel, the functional generalisation of the periodogram matrix. The consistency and asymptotic law of these estimators are studied in detail. As immediate consequences, we obtain central limit theorems for the mean and the long-run covariance operator of a stationary functional time series. Our results do not depend on structural modelling assumptions, but only functional versions of classical cumulant mixing conditions, and are shown to be stable under discrete observation of the individual curves.

1. Introduction. In the usual context of functional data analysis, one wishes to make inferences pertaining to the law of a continuous time stochastic process $\{X(\tau); \tau \in [0, 1]\}$ on the basis of a collection of T realisations of this stochastic process, $\{X_t(\tau)\}_{t=0}^{T-1}$. These are modelled as random elements of the separable Hilbert space $L^2([0, 1], \mathbb{R})$ of square integrable real functions defined on $[0, 1]$. Statistical analyses typically focus on the first and second-order characteristics of this law [see, e.g., Grenander (1981), Rice and Silverman (1991), Ramsay and Silverman (2005)] and are, for the most part, based on the fundamental Karhunen–Loève decomposition [Karhunen (1947), Lévy (1948), Dauxois, Pousse and Romain (1982), Hall and Hosseini-Nasab (2006)]. Especially the second-order structure of random functions is central to the analysis of functional data, as it is connected with the smoothness properties of the random functions and their optimal finite-dimensional representations [e.g., Adler (1990)]. When functional data are

independent and identically distributed, the entire second-order structure is captured by the covariance operator [Grenander (1981)], or related operators [e.g., Locantore et al. (1999), Kraus and Panaretos (2012)]. The assumption of identical distribution can be relaxed, and this is often done by allowing a varying first-order structure through the inclusion of covariate variables (or functions) in the context of functional regression and analysis of variance models; see Cuevas, Febrero and Fraiman (2002); Cardot and Sarda (2006); Yao, Müller and Wang (2005). Second-order structure has been studied in the "nonidentically distributed" context mostly in terms of the so-called common principal components model [e.g., Benko, Härdle and Kneip (2009)], in a comparison setting, where two functional populations are compared with respect to their covariance structure [e.g., Horváth and Kokoszka (2012), Panaretos, Kraus and Maddocks (2010), Boente, Rodriguez and Sued (2011), Fremdt et al. (2013)], and in the context of detection of sequential changes in the covariance structure of functional observations [Horváth, Hušková and Kokoszka (2010)]; see Horváth and Kokoszka (2012) for an overview.

For sequences of potentially dependent functional data, Gabrys and Kokoszka (2007) and Gabrys, Horváth and Kokoszka (2010) study the detection of correlation. To obtain a complete description of the second-order structure of dependent functional sequences, one needs to consider autocovariance operators relating different lags of the series, as is the case in multivariate time series. This study will usually be carried out under the assumption of stationarity. Research in this context has mostly focused on stationary functional series that are linear. Problems considered include that of the estimation of the second-order structure [e.g., Mas (2000), Bosq (2002), Dehling and Sharipov (2005)] and that of prediction [e.g., Antoniadis and Sapatinas (2003), Ferraty and Vieu (2004), Antoniadis, Paparoditis and Sapatinas (2006)]. It can be said that the linear case is now relatively well understood, and Bosq (2000) and Bosq and Blanke (2007) provide a detailed overview thereof.

Recent work has attempted to move functional time series beyond linear models and construct inferential procedures for time series that are not a priori assumed to be described by a particular model, but are only assumed to satisfy certain weak dependence conditions. Hörmann and Kokoszka (2010) consider the effect that weak dependence can have on the principal component analysis of functional data and propose weak dependence conditions under which they study the stability of procedures that assume independence. They also study the problem of inferring the long-run covariance operator by means of finite-dimensional projections. Horváth, Kokoszka and Reeder (2013) give a central limit theorem for the mean of a stationary weakly dependent functional sequence, and propose a consistent estimator for the long-run covariance operator.

In this paper, rather than focus on isolated characteristics such as the long-run covariance, we consider the problem of inferring the complete second-order structure of stationary functional time series without any structural modelling assumptions, except for cumulant-type mixing conditions. Our approach is to study the problem via Fourier analysis, formulating a frequency domain framework for

weakly dependent functional data. To this aim, we employ suitable generalisations of finite-dimensional notions [e.g., Brillinger (2001), Bloomfield (2000), Priestley (2001)] and provide conditions for these to be well defined.

We encode the complete second-order structure via the spectral density operator, the Fourier transform of the collection of autocovariance operators, seen as operator-valued functions of the lag argument; see Proposition 2.1. We propose strongly consistent and asymptotically Gaussian estimators of the spectral density operator based on smoothing the periodogram operator—the functional analogue of the periodogram matrix; see Theorems 3.6 and 3.7. In this sense, our methods can be seen as functional smoothing, as overviewed in Ferraty and Vieu (2006), but in an operator context; see also, for example, Ferraty et al. (2011a), Ferraty et al. (2011b), Laib and Louani (2010). As a by-product, we also obtain central limit theorems for both the mean and long-run covariance operator of stationary time series paralleling or extending the results of Horváth, Kokoszka and Reeder (2013), but under different weak dependence conditions; see Corollaries 2.4 and 3.8. The key result employed in our analysis is the asymptotic representation of the discrete Fourier transform of a weakly dependent stationary functional process as a collection of independent Gaussian elements of $L^2([0, 1], \mathbb{C})$, the Hilbert space of square integrable complex-valued functions, with mean zero and covariance operator proportional to the spectral density operator at the corresponding frequency (Theorem 2.2). Weak dependence conditions required to yield these results are moment type conditions based on cumulant kernels, which are functional versions of cumulant functions. A noteworthy feature of our results and methodology is that they do not require the projection onto a finite-dimensional subspace, as is often the case with functional time series [Hörmann and Kokoszka (2010), Sen and Klüppelberg (2010)]. Rather, our asymptotic results hold for purely infinite-dimensional functional data.

The paper is organised in seven sections and the supplementary material [Panaretos and Tavakoli (2013)]. The building blocks of the frequency domain framework are developed in Section 2. After some basic definitions and introduction of notation, Section 2.1 provides conditions for the definition of the spectral density operator. The functional version of the discrete Fourier transform is introduced in Section 2.2, where its analytical and asymptotic properties are investigated. Section 2.3 then introduces the periodogram operator and studies its mean and covariance structure. The estimation of the spectral density operator by means of smoothing is considered in Section 3. Section 4 provides a detailed discussion on the weak dependence conditions introduced in earlier sections. The effect of observing only discretely sampled functions is considered in Section 5, where the consistency is seen to persist under conditions on the nature of the discrete sampling scheme. Finite-sample properties are illustrated via simulation in Section 6. Technical background and several lemmas required for the proofs or the main results are provided in a an extensive supplementary material [Panaretos

and Tavakoli (2013)]. One of our technical results, Lemma 7.1, collects some results that may be of independent interest in functional data analysis when seeking to establish tightness in order to extend finite-dimensional convergence results to infinite dimensions; it is given in the main paper, in a short section (Section 7).

2. Spectral characteristics of stationary functional data.

We start out this section with an introduction of some basic definitions and notation. Let $\{X_t\}_{t\in\mathbb{Z}}$ be a functional time series indexed by the integers, interpreted as time. That is, for each t, we understand X_t as being a random element of $L^2([0,1],\mathbb{R})$, with

$$\tau \mapsto X_t(\tau) \in \mathbb{R}, \qquad \tau \in [0,1],$$

denoting its parametrisation. Though all our results will be valid for any separable Hilbert space, we choose to concentrate on $L^2([0,1],\mathbb{R})$, as this is the paradigm for functional data analysis. We denote the inner product in $L^2([0,1],\mathbb{R})$ by $\langle\cdot,\cdot\rangle$, and the induced norm by $\|\cdot\|_2$,

$$\langle f,g\rangle = \left(\int_0^1 f(\tau)g(\tau)\,d\tau\right)^{1/2}, \qquad \|g\|_2 = \langle g,g\rangle^{1/2}, \qquad f,g \in L^2([0,1],\mathbb{R}).$$

Equality of L^2 elements will be understood in the sense of the norm of their difference being zero. The imaginary number will de denoted by i, $i^2 = -1$, and the complex conjugate of $z \in \mathbb{C}$ will be denoted as $\bar{z}$. We also denote $\Delta^{(T)}(\omega) = \sum_{t=0}^{T-1}\exp(-i\omega t)$. The Hermitian adjoint of an operator $\mathscr{A}$ will be denoted as $\mathscr{A}^\dagger$. For a function $g: D \subset \mathbb{R}^n \to \mathbb{C}$, we denote $\|g\|_\infty = \sup_{\mathbf{x}\in D}|g(\mathbf{x})|$.

Throughout, we assume that the series $\{X_t\}_{t\in\mathbb{Z}}$ is strictly stationary: for any finite set of indices $I \subset \mathbb{Z}$ and any $s \in \mathbb{Z}$, the joint law of $\{X_t, t \in I\}$ coincides with that of $\{X_{t+s}, t \in I\}$. If $\mathbb{E}\|X_0\|_2 < \infty$, the mean of X_t is well defined, belongs to $L^2([0,1],\mathbb{R})$, and is independent of t by stationarity, $\mu(\tau) = \mathbb{E}X_t(\tau)$. We also define the autocovariance kernel at lag t by

$$r_t(\tau,\sigma) = \mathbb{E}[(X_{t+s}(\tau) - \mu(\tau))(X_s(\sigma) - \mu(\sigma))], \qquad \tau,\sigma \in [0,1] \text{ and } t,s \in \mathbb{Z}.$$

This kernel is well defined in the L^2 sense if $\mathbb{E}\|X_0\|_2^2 < \infty$; if continuity in mean square of X_t is assumed, then it is also well defined pointwise. Each kernel r_t induces a corresponding operator $\mathscr{R}_t: L^2([0,1],\mathbb{R}) \to L^2([0,1],\mathbb{R})$ by right integration, the *autocovariance operator at lag t*,

$$\mathscr{R}_t h(\tau) = \int_0^1 r_t(\tau,\sigma)h(\sigma)\,d\sigma = \text{cov}[\langle X_0,h\rangle, X_t(\tau)], \qquad h \in L^2([0,1],\mathbb{R}).$$

One of the notions we will employ to quantify the weak dependence among the observations $\{X_t\}$ is that of a *cumulant kernel* of the series; the pointwise definition of a kth *order cumulant kernel* is

$$\text{cum}(X_{t_1}(\tau_1),\ldots,X_{t_k}(\tau_k)) = \sum_{\nu=(\nu_1,\ldots,\nu_p)}(-1)^{p-1}(p-1)!\prod_{l=1}^p \mathbb{E}\left[\prod_{j\in\nu_l}X_{t_j}(\tau_j)\right],$$

where the sum extends over all unordered partitions of $\{1, \ldots, k\}$. Assuming $\mathbb{E}\|X_0\|_2^l < \infty$ for $l \geq 1$ guarantees that the cumulant kernels are well defined in an L^2 sense. A cumulant kernel of order $2k$ gives rise to a corresponding $2k$th order cumulant operator $\mathscr{R}_{t_1, \ldots, t_{2k-1}} : L^2([0, 1]^k, \mathbb{R}) \to L^2([0, 1]^k, \mathbb{R})$, defined by right integration,

$$\mathscr{R}_{t_1, \ldots, t_{2k-1}} h(\tau_1, \ldots, \tau_k)$$
$$= \int_{[0,1]^k} \mathrm{cum}(X_{t_1}(\tau_1), \ldots, X_{t_{2k-1}}(\tau_{2k-1}), X_0(\tau_{2k}))$$
$$\times h(\tau_{k+1}, \ldots, \tau_{2k}) \, d\tau_{k+1} \cdots d\tau_{2k}.$$

2.1. *The spectral density operator.* The autocovariance operators encode all the second-order dynamical properties of the series and are typically the main focus of functional time series analysis. Since we wish to formulate a framework for a frequency domain analysis of the series $\{X_t\}$, we need to consider a suitable notion of Fourier transform of these operators. This we call the *spectral density operator* of $\{X_t\}$, defined rigorously in Proposition 2.1 below. Results of a similar flavour related to Fourier transforms between general Hilbert spaces can be traced back to, for example, Kolmogorov (1978); we give here the precise versions that we will be requiring, for completeness, since those results do not readily apply in our setting.

PROPOSITION 2.1. *Suppose $p = 2$ or $p = \infty$, and consider the following conditions*:

I(p) *the autocovariance kernels satisfy* $\sum_{t \in \mathbb{Z}} \|r_t\|_p < \infty$;

II *the autocovariance operators satisfy* $\sum_{t \in \mathbb{Z}} \||\mathscr{R}_t\||_1 < \infty$, *where* $\||\mathscr{R}_t\||_1$ *is the nuclear norm or Schatten 1-norm; see Paragraph F.1.1 in the supplementary material* [Panaretos and Tavakoli (2013)]. *Then, under* I(p), *for any $\omega \in \mathbb{R}$, the following series converges in* $\| \cdot \|_p$:

$$(2.1) \qquad\qquad f_\omega(\cdot, \cdot) = \frac{1}{2\pi} \sum_{t \in \mathbb{Z}} \exp(-\mathrm{i}\omega t) r_t(\cdot, \cdot).$$

We call the limiting kernel f_ω the spectral density kernel *at frequency ω. It is uniformly bounded and also uniformly continuous in ω with respect to* $\| \cdot \|_p$; *that is, given $\varepsilon > 0$, there exists a $\delta > 0$ such that*

$$|\omega_1 - \omega_2| < \delta \quad\Longrightarrow\quad \|f_{\omega_1} - f_{\omega_2}\|_p < \varepsilon.$$

The spectral density operator $\mathscr{F}_\omega$, the operator induced by the spectral density kernel through right-integration, is self-adjoint and nonnegative definite for all $\omega \in \mathbb{R}$. Furthermore, the following inversion formula holds in $\| \cdot \|_p$:

$$(2.2) \qquad\qquad \int_0^{2\pi} f_\alpha(\tau, \sigma) e^{\mathrm{i}t\alpha} \, d\alpha = r_t(\tau, \sigma) \qquad \forall t, \tau, \sigma.$$

Under only II, *we have*

(2.3)
$$\mathscr{F}_\omega = \frac{1}{2\pi} \sum_{t \in \mathbb{Z}} e^{-i\omega t} \mathscr{R}_t,$$

where the convergence holds in nuclear norm. In particular, the spectral density operators are nuclear, and $\|\mathscr{F}_\omega\|_1 \leq \frac{1}{2\pi} \sum_t \|\mathscr{R}_t\|_1 < \infty.$

PROOF.   See Proposition A.1 in the supplementary material [Panaretos and Tavakoli (2013)].   □

The inversion relationship (2.2), in particular, shows that the autocovariance operators and the spectral density operators comprise a Fourier pair, thus reducing the study of second-order dynamics to that of the study of the spectral density operator.

We use the term *spectral density operator* by analogy to the multivariate case, in which the Fourier transform of the autocovariance functions is called the spectral density matrix; see, for example, Brillinger (2001). In our case, since the time series takes values in $L^2([0, 1], \mathbb{R})$, the autocovariance functions are in fact *operators* and their Fourier transform is an *operator*, hence the term *spectral density operator*. In light of the inversion formula (2.2), for fixed (τ, σ), we can think of the $\omega \mapsto f_\omega(\tau, \sigma)$ as being a (complex) measure, giving the *distribution of energy between $X_t(\tau)$ and $X_0(\sigma)$ across different frequencies*. That is, $\omega \mapsto f_\omega(\tau, \tau) \geq 0$ gives the power spectrum of the univariate time series $\{X_t(\tau)\}_{t \in \mathbb{Z}}$, while given $\tau \neq \sigma$, $\omega \mapsto f_\omega(\tau, \sigma) \in \mathbb{C}$ gives the cross spectrum of the univariate time series $\{X_t(\tau)\}_{t \in \mathbb{Z}}$ with $\{X_t(\sigma)\}_{t \in \mathbb{Z}}$. When a point-wise interpretation of $\{X_t\}_{t \in \mathbb{Z}}$ is not possible (e.g., because it is only interpretable via L^2 equivalence classes), the spectral density operator admits a weak interpretation as follows: given L^2 elements $\psi \neq \phi$, the mapping $\omega \mapsto \langle \psi, \mathscr{F}_\omega \psi \rangle \geq 0$ is the power spectrum of the univariate time series $\{\langle \psi, X_t \rangle\}_{t \in \mathbb{Z}}$, while $\omega \mapsto \langle \psi, \mathscr{F}_\omega \phi \rangle = \langle \mathscr{F}_\omega \psi, \phi \rangle \in \mathbb{C}$ is the cross spectrum of the univariate time series $\{\langle \psi, X_t \rangle\}_{t \in \mathbb{Z}}$ with the univariate time series $\{\langle \phi, X_t \rangle\}_{t \in \mathbb{Z}}$. In this sense, $\mathscr{F}_\omega$ provides a complete characterisation of the second-order dynamics of the functional process $\{X_t\}$; see also Panaretos and Tavakoli (2013) for the role of the spectral density operator in the spectral representation and the harmonic principal component analysis of functional time series.

2.2. *The functional discrete Fourier transform and its properties.*   In practice, a stretch of length T of the series $\{X_t\}_{t \in \mathbb{Z}}$ will be available, and we will wish to draw inferences on the spectral density operator based on this finite stretch. The main tool that we will employ is the functional version of the discrete Fourier transform (DFT). In particular, define the *functional Discrete Fourier Transform (fDFT)* of $\{X_t\}_{t=0}^{T-1}$ to be

$$\widetilde{X}_\omega^{(T)}(\tau) = (2\pi T)^{-1/2} \sum_{t=0}^{T-1} X_t(\tau) \exp(-i\omega t).$$

It is of interest to note here that the construction of the fDFT does not require the representation of the data in a particular basis. The fDFT transforms the T functional observations to a mapping from $\mathbb{R}$ into $L^2([0, 1], \mathbb{C})$. It straightforwardly inherits some basic analytical properties that its finite-dimensional counterpart satisfies; for example, it is 2π-periodic and Hermitian with respect to ω, and linear with respect to the series $\{X_t\}$.

The extension of the stochastic properties of the multivariate DFT to the fDFT, however, is not as straightforward. It is immediate that $\mathbb{E}\|\widetilde{X}_\omega^{(T)}\|_2^l < \infty$ if $\mathbb{E}\|X_t\|_2^l < \infty$, and hence the fDFT is almost surely in $L^2([0, 1], \mathbb{C})$ if $\mathbb{E}\|X_t\|_2^2 < \infty$. We will see that the asymptotic covariance operator of this object coincides with the spectral density operator. Most importantly, we prove below that the fundamental stochastic property of the multivariate DFT can be adapted and extended to the infinite-dimensional case; that is, under suitable weak dependence conditions, as $T \to \infty$, the fDFT evaluated at distinct frequencies yields independent and Gaussian random elements of $L^2([0, 1], \mathbb{C})$. The important aspect of this limit theorem is that it does not require the assumption of any particular model for the stationary series, and imposes only cumulant mixing conditions. A more detailed discussion of these conditions is provided in Section 4.

THEOREM 2.2 (Asymptotic distribution of the fDFT). *Let $\{X_t\}_{t=0}^{T-1}$ be a strictly stationary sequence of random elements of $L^2([0, 1], \mathbb{R})$, of length T. Assume the following conditions hold*:

(i) $\mathbb{E}\|X_0\|_2^k < \infty$, $\sum_{t_1,\ldots,t_{k-1}=-\infty}^\infty \|\operatorname{cum}(X_{t_1}, \ldots, X_{t_{k-1}}, X_0)\|_2 < \infty, \forall k \geq 2$.
(ii) $\sum_{t \in \mathbb{Z}} \|\!|\mathscr{R}_t|\!\|_1 < \infty$.

Then, for $\omega_{1,T} := \omega_1 = 0$, $\omega_{2,T} := \omega_2 = \pi$, and distinct *integers*

$$s_{3,T}, \ldots, s_{J,T} \in \left\{1, \ldots, \lfloor (T-1)/2 \rfloor\right\}$$

such that

$$\omega_{j,T} := \frac{2\pi s_{j,T}}{T} \overset{T \to \infty}{\longrightarrow} \omega_j, \qquad j = 3, \ldots, J,$$

we have

(2.4) $$\widetilde{X}_{\omega_1}^{(T)} - \sqrt{\frac{T}{2\pi}} \mu \overset{d}{\longrightarrow} \widetilde{X}_{\omega_1} \qquad \text{as } T \to \infty,$$

and $\widetilde{X}_{\omega_{j,T}}^{(T)} \overset{d}{\longrightarrow} \widetilde{X}_{\omega_j}$, as $T \to \infty$, $j = 2, \ldots, J$ where $\{\widetilde{X}_{\omega_j}\}$ are independent mean zero Gaussian elements of $L^2([0, 1], \mathbb{R})$ for $j = 1, 2$, and of $L^2([0, 1], \mathbb{C})$ for $j = 3, \ldots, J$, with covariance operators $\mathscr{F}_{\omega_j}$, respectively.

REMARK 2.3. Though the $\{\omega_{j,T}\}_{j=3}^J$ are distinct for every T, the limiting frequencies $\{\omega_j : j = 3, \ldots, J\}$ need not be distinct.

Note here that condition (i) with $k = 2$ is already required in order to define the spectral density kernel and operator in Proposition 2.1. Condition (i) for $k \geq 3$ is the generalisation of the standard multivariate cumulant condition to the functional case [Brillinger (2001), Condition 2.6.1], and reduces to that exact same condition if the data are finite-dimensional. Condition (ii) is required so that the spectral density operator be a nuclear operator at each ω [which is in turn a necessary condition for the weak limit of the fDFT to be almost surely in $L^2([0,1], \mathbb{C})$]. As we shall see, condition (ii) is, in fact, a sufficient condition for tightness of the fDFT, seen as a functional process indexed by frequency.

PROOF OF THEOREM 2.2. Consider $p_\omega^{(T)}(\tau, \sigma) = \widetilde{X}_\omega^{(T)}(\tau)\widetilde{X}_{-\omega}^{(T)}(\sigma)$, and assume initially that $\mu = 0$. We will treat the case $\mu \neq 0$ at the end of the proof. First we show that for any ω (or sequence ω_T), the sequence of random elements $\widetilde{X}_\omega^{(T)}$, $T = 1, 2, \ldots$, is tight. To do this, we shall use Lemma 7.1. Fix an orthonormal basis $\{\varphi_n\}_{n \geq 1}$ of $L^2([0,1], \mathbb{R})$ and let $H = L^2([0,1], \mathbb{C})$. We notice that $p_\omega^{(T)}$ is a random element of the (complete) tensor product space $H \otimes H$, with scalar product and norm $\langle \cdot, \cdot \rangle_{H \otimes H}$, $\| \cdot \|_{H \otimes H}$, respectively; see Weidmann [(1980), Paragraph 3.4], for instance. Notice that $|\langle \widetilde{X}_\omega, \varphi_n \rangle|^2 = \langle p_\omega^{(T)}, \varphi_n \otimes \varphi_n \rangle$. Since $\mathbb{E}\|p_\omega^{(T)}\|_{H \otimes H} < \infty$ and the projection $\mathcal{P}_n : H \otimes H \to \mathbb{C}$ defined by $\mathcal{P}_n(f) = \langle f, \varphi_n \otimes \varphi_n \rangle_{H \otimes H}$ is continuous and linear, we deduce

$$\mathbb{E}|\langle \widetilde{X}_\omega^{(T)}, \varphi_n \rangle|^2 = \mathbb{E}\mathcal{P}_n p_\omega^{(T)} = \mathcal{P}_n \mathbb{E} p_\omega^{(T)} = \frac{1}{2\pi}\mathcal{P}_n \int_{-\pi}^{\pi} F_T(\omega - \alpha)f_\alpha \, d\alpha$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} F_T(\omega - \alpha)\mathcal{P}_n f_\alpha \, d\alpha \leq \sup_{\alpha \in \mathbb{R}} |\mathcal{P}_n f_\alpha|.$$

The third equality comes from Proposition 2.5 (which is independent of previous results), the fourth equality follows from Tonelli's theorem [Wheeden and Zygmund (1977), page 92], and the last inequality is Young's inequality [Hunter and Nachtergaele (2001), Theorem 12.58]. Notice that $|\mathcal{P}_n f_\alpha| = |\langle \mathscr{F}_\alpha \varphi_n, \varphi_n \rangle| \leq \sum_t |\langle \mathscr{R}_t \varphi_n, \varphi_n \rangle|$ by (2.3). Setting $a_n = \sum_t |\langle \mathscr{R}_t \varphi_n, \varphi_n \rangle|$, which is independent of α and T, we have $\mathbb{E}|\langle \widetilde{X}_\omega^{(T)}, \varphi_n \rangle|^2 \leq a_n$, and $\sum_n a_n \leq \sum_{t \in \mathbb{Z}} \|\mathscr{R}_t\|_1 < \infty$. Therefore, we have proven that $\widetilde{X}_\omega^{(T)}$ is tight. Consequently, the random element $(\widetilde{X}_{\omega_{1,T}}^{(T)}, \ldots, \widetilde{X}_{\omega_{J,T}}^{(T)})$ of $(L^2([0,1], \mathbb{C}))^J$ is also tight. Its asymptotic distribution is therefore determined by the convergence of its finite-dimensional distributions; see, for example, Ledoux and Talagrand [(1991), Paragraph 2.1]. Thus, to complete the proof, it suffices to show that for any $\psi_1, \ldots, \psi_J \in L^2([0,1], \mathbb{C})$,

$$(2.5) \qquad (\langle \widetilde{X}_{\omega_{1,T}}^{(T)}, \psi_1 \rangle, \ldots, \langle \widetilde{X}_{\omega_{J,T}}^{(T)}, \psi_J \rangle) \xrightarrow{d} (\langle \widetilde{X}_{\omega_1}, \psi_1 \rangle, \ldots, \langle \widetilde{X}_{\omega_J}, \psi_J \rangle),$$

where $\widetilde{X}_{\omega_j} \sim \mathcal{N}(0, \mathscr{F}_{\omega_j})$ are independent Gaussian random elements of H, where $H = L^2([0,1], \mathbb{R})$ if $j = 1, 2$ and $H = L^2([0,1], \mathbb{C})$ if $j = 3, \ldots, J$. This is a

consequence of the following claim, which is justified by Brillinger [(2001), Theorem 4.4.1]:

(I) For $j = 1, \ldots, J$, let $\psi_j = \varphi_{2j-1} + i\varphi_{2j}$, where $\varphi_1, \ldots, \varphi_{2J} \in L^2([0, 1], \mathbb{R})$, and $\mathbf{Y}_t = (Y_t(1), \ldots, Y_t(2J)) \in \mathbb{R}^{2J}$ be the vector time series with coordinates $Y_t(l) = \langle X_t, \varphi_l \rangle$. Then $\widetilde{\mathbf{Y}}_{\omega_{j,T}}^{(T)} \xrightarrow{d} \widetilde{\mathbf{Y}}_{\omega_j}$, where $\{\widetilde{\mathbf{Y}}_{\omega_j}\}$ are independent mean zero complex Gaussian random vectors with covariance matrix $\mathbf{F}_{\omega_j}$, $(\mathbf{F}_{\omega_j})_{sl} := F_{\omega_j}(s, l) = \langle \mathscr{F}_{\omega_j} \varphi_l, \varphi_s \rangle$.

For the case $\mu \neq 0$, we only need to consider $j = 1, 2$ since $\widetilde{(X - \mu)}_{\omega_{j,T}}^{(T)} = \widetilde{X}_{\omega_{j,T}}^{(T)}$ for $j = 3, \ldots, J$. We need to show that

$$(2.6) \qquad \widetilde{X}_{\omega_1}^{(T)} - \sqrt{\frac{T}{2\pi}} \mu = (2\pi T)^{-1/2} \sum_{t=0}^{T-1} (X_t - \mu) \xrightarrow{d} \widetilde{X}_0,$$

and also that

$$(2.7) \qquad \widetilde{X}_{\omega_2}^{(T)} = (2\pi T)^{-1/2} \sum_{t=0}^{T-1} (-1)^t X_t \xrightarrow{d} \widetilde{X}_\pi.$$

The weak convergence in (2.6) follows immediately from the case $\mu = 0$. For (2.7), notice that

$$\widetilde{X}_{\omega_2}^{(T)} = (2\pi T)^{-1/2} \sum_{t=0}^{T-1} (-1)^t (X_t - \mu) + \mu (2\pi T)^{-1/2} \sum_{t=0}^{T-1} (-1)^t.$$

The first summand is the discrete Fourier transform of a zero mean random process, and converges to $\widetilde{X}_{\omega_2}$. The second summand is deterministic and bounded by $\|\mu\| (2\pi T)^{-1/2}$, which tends to zero. Finally, the continuous mapping theorem for metric spaces [Pollard (1984)] yields (2.7). □

The theorem has important consequences for the statistical analysis of a functional time series. It essentially allows us to transform a collection of weakly dependent functional data of an unknown distribution, to a collection of approximately independent and Gaussian functional data. In particular, let $\{\omega_{j,T}\}_{j=1}^J$ be J sequences (in T) of frequencies, such that $\omega_{j,T} \xrightarrow{T \to \infty} \omega \neq 0$, for all $1 \leq j \leq J$. Then, provided T is large enough, $\{\widetilde{X}_{\omega_{j,T}}^{(T)}\}_{j=1}^J$ is a collection of J approximately i.i.d. mean zero complex Gaussian random functions with covariance operator $\mathscr{F}_\omega$. The size J of the sample is not allowed to grow with T, however. From a practical point of view, it can be chosen to be large, provided that the $\omega_{j,T}$ are not too far from ω. We will make heavy use of this result in order to construct consistent and asymptotically Gaussian estimators of the spectral density operator by means of the *periodogram kernel*, defined in the next section.

We also remark that the weak convergence relation in equation (2.4) can be re-expressed to trivially yield the corollary:

COROLLARY 2.4 (Central limit theorem for cumulant mixing functional series). *Let $\{X_t\}_{t=0}^T$ be a strictly stationary sequence of random elements of $L^2([0, 1], \mathbb{R})$ of length T satisfying conditions* (i) *and* (ii) *of Theorem* 2.2. *Then*

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=0}^{T-1}X_t(\tau) - \mu(\tau)\right) \xrightarrow{d} \mathcal{N}\left(0, \sum_{t\in\mathbb{Z}}\mathscr{R}_t\right).$$

This provides one of the first instances of central limit theorems for functional series under no structural modelling assumptions beyond weak dependence. To our knowledge, the only other similar result is given in recent work by Horváth, Kokoszka and Reeder (2013), who obtain the same limit under different weak dependence conditions, namely L^p-m-approximability.

The covariance operator of the limiting Gaussian measure is the functional analogue of the long-run covariance matrix from multivariate time series. We will revisit this operator in Section 3, where we will derive a related central limit theorem.

2.3. *The periodogram kernel and its properties.* The covariance structure of the weak limit of the fDFT given in Theorem 2.2 motivates the consideration of the empirical covariance of the functional DFT as a basis for the estimation of the spectral density operator. Thus, as with the multivariate case, we are led to consider tensor products of the fDFT leading to the notion of a *periodogram kernel*. Define the *periodogram kernel* as

$$p_\omega^{(T)}(\tau, \sigma) = \big[\widetilde{X}_\omega^{(T)}(\tau)\big]\big[\widetilde{X}_\omega^{(T)}(\sigma)\big]^\dagger = \widetilde{X}_\omega^{(T)}(\tau)\widetilde{X}_{-\omega}^{(T)}(\sigma).$$

If we slightly abuse notation and also write $\|\cdot\|_2$ for the norm in $L^2([0, 1]^2, \mathbb{C})$, we have $\|p_\omega^{(T)}\|_2 = \|\widetilde{X}_\omega^{(T)}\|_2^2$, and hence $\mathbb{E}\|p_\omega^{(T)}\|_2^l < \infty$, if $\mathbb{E}\|X_t\|_2^{2l} < \infty$. The expectation of the periodogram kernel is thus well defined, and, letting $a_T = \sum_{t=-T}^T e^{-i\omega t}r_t$, Lemma F.3 yields $\mathbb{E}p_\omega^{(T)} = T^{-1}(a_0 + a_1 + \cdots + a_{T-1})$. That is, the expectation of the periodogram kernel is a Cesàro-sum of the partial sums of the series defining the spectral density kernel. Therefore, in order to probe the properties of the periodogram kernel, we can make use of the Fejér kernel

$$F_T(\omega) = \frac{1}{T}\left(\frac{\sin(T\omega/2)}{\sin(\omega/2)}\right)^2 = \frac{1}{T}|\Delta^{(T)}(\omega)|^2.$$

It will thus be useful to recall some properties of F_T: $\int_{-\pi}^{\pi}F_T = 2\pi$, $F_T(0) = T$, $F_T(\omega) \sim O(T)$ uniformly in ω, and $F_T(2\pi s/T) = 0$ for s an integer with $s \not\equiv 0 \bmod T$. This last property will be used often. We will also be making use of the following cumulant mixing condition, defined for fixed $l \geq 0$ and $k = 2, 3, \ldots$.

CONDITION C(l, k).   For each $j = 1, \ldots, k - 1$,

$$\sum_{t_1, \ldots, t_{k-1}=-\infty}^{\infty} \left(1 + |t_j|^l\right) \left\| \mathrm{cum}(X_{t_1}, \ldots, X_{t_{k-1}}, X_0) \right\|_2 < \infty.$$

With this definition in place, we may determine the exact mean of the periodogram kernel:

PROPOSITION 2.5.   *Assuming that* C($0, 2$) *holds true, we have, for each* $\omega \in \mathbb{R}$,

$$\mathbb{E}\big[p_\omega^{(T)}(\tau, \sigma)\big] = \frac{1}{2\pi} \int_{-\pi}^{\pi} F_T(\omega - \alpha) f_\alpha(\tau, \sigma) \, d\alpha + \frac{1}{2\pi} \mu(\tau)\mu(\sigma) F_T(\omega) \qquad \text{in } L^2.$$

In particular, if $\omega = 2\pi s / T$, *with* s *an integer such that* $s \not\equiv 0 \bmod T$,

$$\mathbb{E}\big[p_\omega^{(T)}(\tau, \sigma)\big] = \frac{1}{2\pi} \int_{-\pi}^{\pi} F_T(\omega - \alpha) f_\alpha(\tau, \sigma) \, d\alpha \qquad \text{in } L^2.$$

PROOF.   See the supplementary material [Panaretos and Tavakoli (2013)], Proposition C.1.   □

In particular, the periodogram kernel is asymptotically unbiased:

PROPOSITION 2.6.   *Let* s *be an integer with* $s \not\equiv 0 \bmod T$. *Then, we have*

$$\mathbb{E}\big[p_{2\pi s/T}^{(T)}(\tau, \sigma)\big] = f_{2\pi s/T}(\tau, \sigma) + \varepsilon_T \qquad \text{in } L^2.$$

The error term ε_T *is* $O(T^{-1})$ *under* C($0, 2$) *and* $o(1)$ *under* C($1, 2$). *In either case, the error term is uniform in integers* $s \not\equiv 0 \bmod T$.

PROOF.   Since $s \not\equiv 0 \bmod T$,

$$\mathbb{E}\big[p_{2\pi s/T}^{(T)}(\tau, \sigma)\big] = \mathrm{cum}\big(\widetilde{X}_{2\pi s/T}^{(T)}(\tau), \widetilde{X}_{-2\pi s/T}^{(T)}(\sigma)\big) = f_{2\pi s/T}(\tau, \sigma) + \varepsilon_t,$$

and the result follows from Theorem B.2 of the supplementary material [Panaretos and Tavakoli (2013)].   □

Having established the mean structure of the periodogram, we turn to the determination of its covariance structure.

THEOREM 2.7.   *Assume* ω_1 *and* ω_2 *are of the form* $2\pi s(T)/T$, *where* $s(T)$ *is an integer,* $s(T) \not\equiv 0 \bmod T$. *We have*

$$\mathrm{cov}\big(p_{\omega_1}^{(T)}(\tau_1, \sigma_1), p_{\omega_2}^{(T)}(\tau_2, \sigma_2)\big) = \eta(\omega_1 - \omega_2) f_{\omega_1}(\tau_1, \tau_2) f_{-\omega_1}(\sigma_1, \sigma_2)$$
$$+ \eta(\omega_1 + \omega_2) f_{\omega_1}(\tau_1, \sigma_2) f_{-\omega_1}(\sigma_1, \tau_2) + \varepsilon_T$$
$$\text{in } L^2,$$

where the function $\eta(x)$ equals one if $x \in 2\pi\mathbb{Z}$, and zero otherwise. The error term ε_T is $o(1)$ under C(0, 2) *and* C(0, 4); $\varepsilon_T \sim O(T^{-1})$ *under* C(1, 2) *and* C(1, 4). *In each case, the error term is uniform in* ω_1, ω_2 [*of the form* $2\pi s(T)/T$ *with* $s(T) \not\equiv 0 \bmod T$].

PROOF. See the supplementary material [Panaretos and Tavakoli (2013)], Theorem C.2. □

3. Estimation of the spectral density operator. The results in the previous section show that the asymptotic covariance of the periodogram is not zero, and hence, as in the multivariate case, the periodogram kernel itself is not a consistent estimator of the spectral density. In this section, we define a consistent estimator, obtained by convolving the periodogram kernel with an appropriate weight function W. To this aim, let $W(x)$ be a real function defined on $\mathbb{R}$ such that:

(1) W is positive, even, and bounded in variation;
(2) $W(x) = 0$ if $|x| \geq 1$;
(3) $\int_{-\infty}^{\infty} W(x)\,dx = 1$;
(4) $\int_{-\infty}^{\infty} W(x)^2\,dx < \infty$.

The assumption of a compact support is not necessary, but will simplify proofs. For a bandwidth $B_T > 0$, write

$$(3.1) \qquad W^{(T)}(x) = \sum_{j \in \mathbb{Z}} \frac{1}{B_T} W\left(\frac{x + 2\pi j}{B_T}\right).$$

Some properties of W can be found in the supplementary material [Panaretos and Tavakoli (2013)]. We define the *spectral density estimator* $f_\omega^{(T)}$ of f_ω at frequency ω as the weighted average of the periodogram evaluated at frequencies of the form $\{2\pi s/T\}_{s=1}^{T-1}$, with weight function $W^{(T)}$,

$$f_\omega^{(T)}(\tau, \sigma) = \frac{2\pi}{T} \sum_{s=1}^{T-1} W^{(T)}\left(\omega - \frac{2\pi s}{T}\right) p_{2\pi s/T}^{(T)}(\tau, \sigma).$$

A consequence of the assumption of compact support worth mentioning is that, in fact, at most $O(T B_T)$ summands of this expression are nonzero. We will show in this section that, under appropriate conditions on the asymptotic behavior of B_T, this estimator retains the property of asymptotic unbiasedness that the periodogram enjoys. We will determine the behaviour of its asymptotic covariance structure and establish consistency in mean square (with respect to the Hilbert–Schmidt norm). Finally, we will determine the asymptotic law of the estimator.

Concerning the mean of the spectral density estimator, we have:

PROPOSITION 3.1.   *Under* $C(1, 2)$, *if* $B_T \to 0$ *and* $B_T T \to \infty$ *as* $T \to \infty$, *then*

$$\mathbb{E} f_\omega^{(T)}(\tau, \sigma) = \int_{\mathbb{R}} W(x) f_{\omega - x B_T}(\tau, \sigma) \, dx + O(B_T^{-1} T^{-1}),$$

where the equality holds in L^2, *and the error terms are uniform in* ω.

PROOF.   See the supplementary material [Panaretos and Tavakoli (2013)], Proposition D.1.   □

Concerning the covariance of the spectral density estimator, we have:

THEOREM 3.2.   *Under* $C(1, 2)$ *and* $C(1, 4)$,

$$\text{cov}\big(f_{\omega_1}^{(T)}(\tau_1, \sigma_1), f_{\omega_2}^{(T)}(\tau_2, \sigma_2)\big)$$

$$= \frac{2\pi}{T} \int_{-\pi}^{\pi} \big\{ W^{(T)}(\omega_1 - \alpha) W^{(T)}(\omega_2 - \alpha) f_\alpha(\tau_1, \tau_2) f_{-\alpha}(\sigma_1, \sigma_2)$$

$$+ W^{(T)}(\omega_1 - \alpha) W^{(T)}(\omega_2 + \alpha) f_\alpha(\tau_1, \sigma_2) f_{-\alpha}(\sigma_1, \tau_2) \big\} \, d\alpha$$

$$+ O(B_T^{-2} T^{-2}) + O(T^{-1}),$$

where the equality holds in L^2, *and the error terms are uniform in* ω.

PROOF.   See the supplementary material [Panaretos and Tavakoli (2013)], Theorem D.2.   □

Noting that $\|W^{(T)}\|_\infty = O(B_T^{-1})$ and $\|f.\|_\infty = O(1)$, a direct consequence of the last result is the following approximation of the asymptotic covariance of the spectral density estimator:

COROLLARY 3.3.   *Under* $C(1, 2)$ *and* $C(1, 4)$,

$$\text{cov}\big(f_{\omega_1}^{(T)}(\tau_1, \sigma_1), f_{\omega_2}^{(T)}(\tau_2, \sigma_2)\big) = O(B_T^{-2} T^{-1}),$$

where the equality holds in L^2, *uniformly in the* ω'*s*.

This bound is not sharp. A better bound is given in the next statement, which, however, is not uniform in ω.

PROPOSITION 3.4.
Assume conditions $C(1, 2)$, $C(1, 4)$, *and that* $B_T \to 0$ *as* $T \to \infty$ *with* $B_T T \to \infty$. *Then*

$$\lim_{T \to \infty} B_T T \, \text{cov}\big(f_{\omega_1}^{(T)}(\tau_1, \sigma_1), f_{\omega_2}^{(T)}(\tau_2, \sigma_2)\big)$$

$$= 2\pi \int_{\mathbb{R}} W(\alpha)^2 \, d\alpha \big\{ \eta(\omega_1 - \omega_2) f_{\omega_1}(\tau_1, \tau_2) f_{-\omega_1}(\sigma_1, \sigma_2)$$

$$+ \eta(\omega_1 + \omega_2) f_{\omega_1}(\tau_1, \sigma_2) f_{-\omega_1}(\sigma_1, \tau_2) \big\}.$$

The function $\eta(x)$ equals one if $x \in 2\pi\mathbb{Z}$, and zero otherwise. The convergence is in L^2 for any fixed ω_1, ω_2. If ω_1, ω_2 depend on T, then the convergence is in L^2, provided $(\omega_1 \pm \omega_2)$ are at a distance of at least $2B_T$ from any multiples of 2π, if not exactly a multiple of 2π.

PROOF. Let $d(x, y)$ denote the distance in $\mathbb{R}/2\pi\mathbb{Z}$. We shall abuse notation and let x, y stand for equivalence classes of real numbers, and also omit the (τ, σ)'s, for the sake of clarity. Theorem 3.2 yields

$$B_T T \operatorname{cov}(f_{\omega_1}^{(T)}, f_{\omega_2}^{(T)})$$

$$(3.2) \qquad = 2\pi B_T \int_{-\pi}^{\pi} W^{(T)}(\omega_1 - \omega_2 - \alpha) W^{(T)}(\alpha) f_{\omega_2 + \alpha} f_{-(\omega_2 + \alpha)} \, d\alpha$$

$$(3.3) \qquad + 2\pi B_T \int_{-\pi}^{\pi} W^{(T)}(\omega_1 + \omega_2 - \alpha) W^{(T)}(\alpha) f_{-(\omega_2 - \alpha)} f_{\omega_2 - \alpha} \, d\alpha$$

$$+ O(B_T^{-1} T^{-1}) + O(B_T).$$

We have employed a change of variables, the fact that $W^{(T)}$ is even, and the fact that both $W^{(T)}$ and f are 2π-periodic. The error terms tend to zero as $B_T \to 0$, $T B_T \to \infty$.

First we show that (3.2) tends to

$$(3.4) \qquad \eta(\omega_1 - \omega_2) f_{\omega_1}(\tau_1, \tau_2) f_{-\omega_1}(\sigma_1, \sigma_2) 2\pi \int_{\mathbb{R}} W(\alpha)^2 \, d\alpha,$$

in L^2, uniformly in all $\omega_1 = \omega_{1,T}$, $\omega_2 = \omega_{2,T}$ such that $\omega_{1,T} \equiv \omega_{2,T}$ or $d(\omega_{1,T} - \omega_{2,T}, 0) \geq 2B_T$ for large T. If $d(\omega_1 - \omega_2, 0) \geq 2B_T$, (3.2) is exactly equal to zero. If $\omega_1 \equiv \omega_2$, we claim that (3.2) tends to

$$(3.5) \qquad f_{\omega}(\tau_1, \tau_2) f_{-\omega}(\sigma_1, \sigma_2) 2\pi \int_{\mathbb{R}} W(\alpha)^2 \, d\alpha.$$

Notice that in this case, (3.2) can be written as $\int_{-\pi}^{\pi} K_T(\alpha) f_{\omega+\alpha} f_{-(\omega+\alpha)} \, d\alpha \times \{\int_{\mathbb{R}} W(\alpha)^2 \, d\alpha\}$, where $K_T(\alpha) = \frac{2\pi}{B_T} [W(\alpha/B_T)]^2 \{\int_{\mathbb{R}} W(\alpha)^2 \, d\alpha\}^{-1}$ is an approximate identity on $[-\pi, \pi]$; see Edwards (1967), Section 3.2. Since the spectral density kernel is uniformly continuous with respect to $\|\cdot\|_2$ (see Proposition 2.1) Lemma F.15 implies that (3.2) tends indeed to (3.5) uniformly in ω with respect to $\|\cdot\|_2$. Hence (3.2) tends to (3.4) in $\|\cdot\|_2$, uniformly in ω's satisfying

$$\omega_{1,T} \equiv \omega_{2,T} \quad \text{or} \quad d(\omega_{1,T} - \omega_{2,T}, 0) \geq 2B_T \qquad \text{for large } T.$$

Similarly, we may show that (3.3) tends to $\eta(\omega_1 + \omega_2) f_{\omega_1}(\tau_1, \sigma_2) f_{-\omega_1}(\sigma_1, \tau_2) \times 2\pi \int_{\mathbb{R}} W(\alpha)^2 \, d\alpha$, uniformly in ω's if $\omega_{1,T} \equiv -\omega_{2,T}$ or $d(\omega_{1,T} + \omega_{2,T}) \geq 2B_T$ for large T. Piecing these results together, we obtain the desired convergence, provided for each T large enough, either $\omega_{1,T} - \omega_{2,T} \equiv 0$, $\omega_{1,T} + \omega_{2,T} \equiv 0$, or

$$d(\omega_{1,T} - \omega_{2,T}, 0) \geq 2B_T \quad \text{and} \quad d(\omega_{1,T} + \omega_{2,T}, 0) \geq 2B_T. \qquad \square$$

REMARK 3.5. In practice, functional data are assumed to be smooth in addition to square-integrable. In such cases, one may hope to obtain stronger results, for example with respect to uniform rather than L^2 norms. Indeed, if the conditions $C(l, k)$ are replaced by the stronger conditions

CONDITION $C'(l, k)$. For each $j = 1, \ldots, k - 1$
$$\sum_{t_1, \ldots, t_{k-1} \in \mathbb{Z}} \left(1 + |t_j|^l\right) \left\|\operatorname{cum}(X_{t_1}, \ldots, X_{t_{k-1}}, X_0)\right\|_\infty < \infty,$$

then the results of Propositions 2.5, 2.6, Theorem 2.7, Proposition 3.1, Theorem 3.2, Corollary 3.3, Proposition 3.4, and Lemma B.1, Theorem B.2 in the supplementary material [Panaretos and Tavakoli (2013)] would hold in the supremum norm with respect to τ, σ.

Combining the results on the asymptotic bias and variance of the spectral density operator, we may now derive the consistency in integrated mean square of the induced estimator for the spectral density operator. Recall that $\mathscr{F}_\omega$ is the integral operator with kernel f_ω, and, similarly let $\mathscr{F}_\omega^{(T)}$ be the operator with kernel $f_\omega^{(T)}$. We have:

THEOREM 3.6. *Provided assumptions* $C(1, 2)$ *and* $C(1, 4)$ *hold,* $B_T \to 0$, $B_T T \to \infty$, *the spectral density operator estimator* $\mathscr{F}_\omega^{(T)}$ *is consistent in integrated mean square, that is,*

$$\operatorname{IMSE}(\mathscr{F}^{(T)}) = \int_{-\pi}^{\pi} \mathbb{E}\left\|\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right\|_2^2 d\omega \to 0, \qquad T \to \infty,$$

where $\|\!|\cdot|\!\|_2$ *is the Hilbert–Schmidt norm* (*the Schatten 2-norm*). *More precisely,* $\operatorname{IMSE}(\mathscr{F}^{(T)}) = O(B_T^2) + O(B_T^{-1} T^{-1})$ *as* $T \to \infty$. *We also have pointwise mean square convergence for a fixed* ω:

$$\mathbb{E}\left\|\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right\|_2^2 = \begin{cases} O(B_T^2) + O(B_T^{-1} T^{-1}), & \text{if } 0 < |\omega| < \pi, \\ O(B_T^2) + O(B_T^{-2} T^{-1}), & \text{if } \omega = 0, \pm\pi \end{cases}$$

as $T \to \infty$.

PROOF. For an integral operator K with a complex-valued kernel $k(\tau, \sigma)$, we will denote by $\overline{K}$ the operator with kernel $\overline{k(\tau, \sigma)}$. Let $\|\!|\cdot|\!\|_2$ be the Hilbert–Schmidt norm. Proposition F.21 yields $\|\!|K|\!\|_2 = \|\!|\overline{K}|\!\|_2$. Further, notice that $f_{-\omega}(\tau, \sigma) = \overline{f_\omega(\tau, \sigma)}$, hence $\mathscr{F}_{-\omega} = \overline{\mathscr{F}_\omega}$. Similarly, $\mathscr{F}_{-\omega}^{(T)} = \overline{\mathscr{F}_\omega^{(T)}}$. Thus, via a change of variables, the IMSE of the spectral density estimator can be written as

$$\int_{-\pi}^{\pi} \mathbb{E}\left\|\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right\|_2^2 d\omega = 2\int_0^{\pi} \mathbb{E}\left\|\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right\|_2^2 d\omega$$

$$= 2\int_0^{\pi} \mathbb{E}\left\|\mathscr{F}_\omega^{(T)} - \mathbb{E}\mathscr{F}_\omega^{(T)}\right\|_2^2 d\omega + 2\int_0^{\pi} \left\|\mathscr{F}_\omega - \mathbb{E}\mathscr{F}_\omega^{(T)}\right\|_2^2 d\omega,$$

which is essentially the usual bias/variance decomposition of the mean square error. Initially, we focus on the variance term. Lemma F.22 yields

$$\int_0^\pi \mathbb{E}\|\|\mathscr{F}_\omega^{(T)} - \mathbb{E}\mathscr{F}_\omega^{(T)}\|\|_2^2 \, d\omega = \int_0^\pi \iint_{[0,1]^2} \mathrm{var}\big(f_\omega^{(T)}(\tau,\sigma)\big) \, d\tau \, d\sigma \, d\omega.$$

Decomposing the outer integral into three terms, $\int_0^\pi = \int_0^{\pi B_T} + \int_{\pi B_T}^{\pi - B_T} + \int_{\pi - B_T}^\pi$, we can use Corollary 3.3 for the first and last summands, and Proposition 3.4 for the second summand to obtain $\int_0^\pi \mathbb{E}\|\|\mathscr{F}_\omega^{(T)} - \mathbb{E}\mathscr{F}_\omega^{(T)}\|\|_2^2 \, d\omega = O(B_T^{-1}T^{-1})$. Turning to the squared bias, Proposition 3.1 yields

$$\int_0^\pi \|\|\mathscr{F}_\omega - \mathbb{E}\mathscr{F}_\omega^{(T)}\|\|_2^2 \, d\omega$$

$$\leq 3 \int_0^\pi \|\|\Big\{\int_{\mathbb{R}} W(x) f_{\omega - x B_T} \, dx - f_\omega \Big\}\|\|_2^2 \, d\omega + O(T^{-2}) + O(B_T^{-2}T^{-2}),$$

where we have used Jensen's inequality and where $\{\int_{\mathbb{R}} W(x) f_{\omega - x B_T} \, dx - f_\omega\}$ denotes the operator with kernel $\int_{\mathbb{R}} W(x) f_{\omega - x B_T}(\tau,\sigma) \, dx - f_\omega(\tau,\sigma)$. Lemma F.4 implies that this difference is of order $O(B_T)$, uniformly in ω. Hence,

$$3 \int_0^\pi \|\|\Big\{\int_{\mathbb{R}} W(x) f_{\omega - x B_T} \, dx - f_\omega \Big\}\|\|_2^2 \, d\omega \leq O(B_T^2).$$

In summary, we have

$$\int_{-\pi}^\pi \mathbb{E}\|\|\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\|\|^2 \, d\omega \leq O(B_T^2) + O(B_T^{-1}T^{-1}).$$

The spectral density estimator $\mathscr{F}_\cdot^{(T)}$ is therefore consistent in integrated mean square if $B_T \to 0$ and $B_T T \to \infty$ as $T \to \infty$.

A careful examination of the proof reveals that the pointwise statement of the theorem follows by a directly analogous argument. □

Finally, if we include some higher-order cumulant mixing conditions, we may obtain the asymptotic distribution of our estimator as being Gaussian.

THEOREM 3.7. *Assume that* $\mathbb{E}\|X_0\|^k < \infty$ *for all* $k \geq 2$ *and*:

(i) $\sum_{t_1,\ldots,t_{k-1}=-\infty}^\infty \|\mathrm{cum}(X_{t_1},\ldots,X_{t_{k-1}},X_0)\|_2 < \infty$, *for all* $k \geq 2$;
(i') $\sum_{t_1,\ldots,t_{k-1}=-\infty}^\infty (1 + |t_j|) \|\mathrm{cum}(X_{t_1},\ldots,X_{t_{k-1}},X_0)\|_2 < \infty$, *for* $k \in \{2,4\}$ *and* $j < k$;
(ii) $\sum_{t \in \mathbb{Z}} (1 + |t|) \|\|\mathscr{R}_t\|\|_1 < \infty$;
(iii) $\sum_{t_1,t_2,t_3 \in \mathbb{Z}} \|\|\mathscr{R}_{t_1,t_2,t_3}\|\|_1 < \infty$.

Then, for any frequencies $\omega_1,\ldots,\omega_J$, *with* $J < \infty$,

$$\sqrt{B_T T}\big(f_{\omega_j}^{(T)} - \mathbb{E}f_{\omega_j}^{(T)}\big) \xrightarrow{d} \hat{f}_{\omega_j}, \qquad j = 1,\ldots,J,$$

where $\hat{f}_{\omega_j}$, $j = 1, \dots, J$, are jointly mean zero complex Gaussian elements in $L^2([0, 1]^2, \mathbb{C})$, with covariance kernel

$$\mathrm{cov}\big(\hat{f}_{\omega_i}(\tau_1, \sigma_1), \hat{f}_{\omega_j}(\tau_2, \sigma_2)\big)$$

$$= 2\pi \int_{\mathbb{R}} W(\alpha)^2 \, d\alpha \{\eta(\omega_i - \omega_j)f_{\omega_i}(\tau_1, \tau_2)f_{-\omega_i}(\sigma_1, \sigma_2)$$

$$+ \eta(\omega_i + \omega_j)f_{\omega_i}(\tau_1, \sigma_2)f_{-\omega_i}(\sigma_1, \tau_2)\}.$$

In particular, we see that $\hat{f}_{\omega_i}$ and $\hat{f}_{\omega_j}$ are independent if $\omega_i \pm \omega_j \not\equiv 0 \bmod 2\pi$, and $\hat{f}_\omega$ is real Gaussian if $\omega \equiv 0 \bmod \pi$.

PROOF.   Let (φ_n) be a basis of $L^2([0, 1], \mathbb{R})$. Then $\{\varphi_{m_1} \otimes \cdots \otimes \varphi_{m_k}\}_{m_1,\dots,m_k \geq 1}$ is a basis of the complex Hilbert space $L^2([0, 1]^k, \mathbb{C})$ [e.g., Kadison and Ringrose (1997)], where $\varphi_{m_1} \otimes \cdots \otimes \varphi_{m_k}(\tau_1, \dots, \tau_k) = \varphi_{m_1}(\tau_1) \times \cdots \times \varphi_{m_k}(\tau_k)$. We denote by $\Phi_\omega(m, n)$ the (m, n)th coordinate of the spectral density matrix and, more generally, we define the kth-order cumulant spectra array $\Phi_{\omega_1,\dots,\omega_{k-1}}$ by

$$\Phi_{\omega_1,\dots,\omega_{k-1}}(m_1, \dots, m_k) = \int_{[0,1]^k} f_{\omega_1,\dots,\omega_{k-1}}(\tau_1, \dots, \tau_k)\varphi_{m_1}(\tau_1) \cdots \varphi_{m_k}(\tau_k) \, d\boldsymbol{\tau}.$$

In other words, the kth-order cumulant spectra array is the scalar product in $L^2([0, 1]^k, \mathbb{C})$ between $f_{\omega_1,\dots,\omega_{k-1}}$ and $\varphi_{m_1} \otimes \cdots \otimes \varphi_{m_k}$.

We also define the periodogram matrix $P_\omega^{(T)}(m, n)$ and the estimator of the spectral density matrix $\Phi_\omega^{(T)}(m, n)$ as the scalar product of $p_\omega^{(T)}$, respectively, $f_\omega^{(T)}$, with $\varphi_m \otimes \varphi_n$. Notice that

$$\Phi_\omega^{(T)}(m, n) = \frac{2\pi}{T} \sum_{s=1}^{T-1} W^{(T)}(\omega - 2\pi s/T) P_{2\pi s/T}^{(T)}(m, n),$$

where $W^{(T)}$ is defined in (3.1). The major steps in the proof are the following two inequalities, along with Lemma 7.1:

(I) For T large enough,

$$T B_T \, \mathrm{var}\big(\Phi_\omega^{(T)}(m, n)\big)$$

(3.6)
$$\leq K\big[\mathrm{sc}_0(m, n, m, n) + \mathrm{sc}_1(m, m)\,\mathrm{sc}_1(n, n)$$

$$+ \mathrm{sc}_0(m, m)\,\mathrm{sc}_0(n, n) + \mathrm{sc}_0(m, n)^2\big],$$

where "sc" stands for "summed cumulant," in particular

$$\mathrm{sc}_0(m_1, \dots, m_k) = \sum_{t_1,\dots,t_{k-1} \in \mathbb{Z}} \big|\mathrm{cum}(\xi_{t_1}(m_1), \dots, \xi_{t_{k-1}}(m_{k-1}), \xi_0(m_k))\big|,$$

$$\mathrm{sc}_1(m_1, \dots, m_k) = \sum_{j=1}^{k-1} \sum_{t_1,\dots,t_{k-1} \in \mathbb{Z}} |t_j|\big|\mathrm{cum}(\xi_{t_1}(m_1), \dots, \xi_{t_{k-1}}(m_{k-1}), \xi_0(m_k))\big|,$$

and $K = 8\|W\|_\infty^2$ is a constant.

(II) We have the following bound:

$$
\sum_{m,n\geq 1} T B_T \operatorname{var}\bigl(\Phi_\omega^{(T)}(m,n)\bigr)
$$

(3.7)

$$
\leq C\Biggl[\Biggl(\sum_{t\in\mathbb{Z}}(1+|t|)\|\mathscr{R}_t\|_1\Biggr)^2 + \sum_{t_1,t_2,t_3\in\mathbb{Z}}\|\mathscr{R}_{t_1,t_2,t_3}\|_1\Biggr]
$$

for some fixed constant $C > 0$. Here, $\mathscr{R}_{t_1,t_2,t_3}$ is the operator on $L^2([0,1]^2,\mathbb{R})$ with kernel $r_{t_1,t_2,t_3}((\tau_1,\tau_2),(\tau_3,\tau_4)) = \operatorname{cum}(X_{t_1},X_{t_2},X_{t_3},X_0)(\tau_1,\tau_2,\tau_3,\tau_4)$. That is, $\mathscr{R}_{t_1,t_2,t_3}f(\tau_1,\tau_2) = \iint_{[0,1]^2} r_{t_1,t_2,t_3}((\tau_1,\tau_2),(\tau_3,\tau_4))f(\tau_3,\tau_4)\,d\tau_3\,d\tau_4$ for $f \in L^2([0,1]^2,\mathbb{R})$.

First we concentrate on establishing (3.6). Recall that

$$
\operatorname{var}\bigl(\Phi_\omega^{(T)}(m,n)\bigr) = (2\pi/T)^2 \sum_{s,l=1}^{T-1} W^{(T)}(\omega - 2\pi s/T)W^{(T)}(\omega - 2\pi l/T)
$$

$$
\times \operatorname{cov}\bigl(P_{2\pi s/T}^{(T)}(m,n), P_{2\pi l/T}^{(T)}(m,n)\bigr).
$$

We need to find an explicit bound on the error terms of Lemma B.1, Theorem B.2 in the supplementary material [Panaretos and Tavakoli (2013)], and Theorem 2.7. An examination of the proof of Lemma B.1 in the supplementary material [Panaretos and Tavakoli (2013)] yields

$$
\Phi_{\omega_1,\dots,\omega_{k-1}}(m_1,\dots,m_k)
$$

$$
= (2\pi)^{-(k-1)} \sum_{t_1,\dots,t_{k-1}=-(T-1)}^{T-1} \exp\Biggl(-\mathrm{i}\sum_{j=1}^{k-1}\omega_j t_j\Biggr)
$$

$$
\times \operatorname{cum}\bigl(\xi_{t_1}(m_1),\dots,\xi_{t_{k-1}}(m_{k-1}),\xi_0(m_k)\bigr)
$$

$$
+ \varepsilon_T^{(\mathrm{B}.1)}(m_1,\dots,m_k),
$$

and $|\varepsilon_T^{(\mathrm{B}.1)}(m_1,\dots,m_k)| \leq (2\pi)^{-(k-1)}(k-1)\operatorname{sc}_0(m_1,\dots,m_k)$. We have used the notation $\varepsilon_T^{(\mathrm{B}.1)}(m_1,\dots,m_k)$ to denote the error term of Lemma B.1, and we shall do likewise for the error term in Theorem B.2 in the supplementary material [Panaretos and Tavakoli (2013)],

$$
T^{k/2}\operatorname{cum}\bigl(\widetilde{\xi}_{\omega_1}^{(T)}(m_1),\dots,\widetilde{\xi}_{\omega_k}^{(T)}(m_k)\bigr)
$$

$$
= (2\pi)^{k/2-1}\Delta^{(T)}\Biggl(\sum_{j=1}^{k}\omega_j\Biggr)\Phi_{\omega_1,\dots,\omega_{k-1}}(m_1,\dots,m_k)
$$

$$
+ \varepsilon_T^{(\mathrm{B}.2)}\Biggl(\sum_{j=1}^{k}\omega_j; m_1,\dots,m_k\Biggr),
$$

where

$$|\varepsilon_T^{(B.2)}(\omega; m_1, \ldots, m_k)|$$

$$\leq 2(2\pi)^{-k/2} \sum_{t_1,\ldots,t_{k-1}=-(T-1)}^{T-1} (|t_1| + \cdots + |t_{k-1}|)$$

$$\times |\operatorname{cum}(\xi_{t_1}(m_1), \ldots, \xi_{t_{k-1}}(m_{k-1}), \xi_0(m_k))|$$

$$+ (2\pi)^{k/2-1}\Delta^{(T)}(\omega)|\varepsilon_T^{(B.1)}(m_1, \ldots, m_k)|$$

$$\leq 2(2\pi)^{-k/2} \operatorname{sc}_1(m_1, \ldots, m_k) + (2\pi)^{-k/2}(k-1)\Delta^{(T)}(\omega) \operatorname{sc}_0(m_1, \ldots, m_k).$$

A less sharp bound (but independent of the frequency) will also be useful,

$$|\varepsilon_T^{(B.2)}(\cdot; m_1, \ldots, m_k)| \leq 3(2\pi)^{-k/2}(k-1)T \operatorname{sc}_0(m_1, \ldots, m_k).$$

We will also need a bound on the spectral density matrix, $|\Phi_{\omega_1,\ldots,\omega_{k-1}}(m_1, \ldots, m_k)| \leq (2\pi)^{-(k-1)} \operatorname{sc}_0(m_1, \ldots, m_k)$.

We now turn to Theorem 2.7: for $s, l = 1, \ldots, T - 1$,

$$\operatorname{cov}\big(P_{2\pi s/T}^{(T)}(m, n), P_{2\pi l/T}^{(T)}(m, n)\big)$$

$$= (2\pi/T)\Phi_{2\pi s/T, -2\pi s/T, 2\pi l/T}(m, n, m, n) + T^{-2}\varepsilon_T^{(B.2)}(\cdot; m, n, m, n)$$

$$+ \delta_{s,l}\big[\Phi_{2\pi s/T}(m, m)\Phi_{-2\pi s/T}(n, n) + \Phi_{2\pi s/T}(m, m)T^{-1}\varepsilon_T^{(B.2)}(\cdot; n, n)$$

$$+ \Phi_{-2\pi s/T}(n, n)T^{-1}\varepsilon_T^{(B.2)}(\cdot; m, m)\big]$$

$$+ \delta_{s+l,T}\big[\Phi_{2\pi s/T}(m, n)\Phi_{-2\pi s/T}(n, m)$$

$$+ \Phi_{2\pi s/T}(m, n)T^{-1}\varepsilon_T^{(B.2)}(\cdot; n, m)$$

$$+ \Phi_{-2\pi s/T}(n, m)T^{-1}\varepsilon_T^{(B.2)}(\cdot; m, n)\big]$$

$$+ T^{-2}\bigg[\varepsilon_T^{(B.2)}\bigg(\frac{2\pi(s-l)}{T}; m, m\bigg)\varepsilon_T^{(B.2)}\bigg(-\frac{2\pi(s-l)}{T}; n, n\bigg)$$

$$+ \varepsilon_T^{(B.2)}\bigg(\frac{2\pi(s+l)}{T}; m, n\bigg)\varepsilon_T^{(B.2)}\bigg(-\frac{2\pi(s+l)}{T}; n, m\bigg)\bigg],$$

where $\delta_{s,l} = 1$ if $s = l$, and zero otherwise. Using the previous bounds, and the fact that $\operatorname{sc}_0(m, n) = \operatorname{sc}_0(n, m)$, we obtain

$$|\operatorname{cov}\big(P_{2\pi s/T}^{(T)}(m, n), P_{2\pi l/T}^{(T)}(m, n)\big)|$$

$$\leq \frac{1}{4\pi^2}\big[4T^{-2} \operatorname{sc}_1(m, m) \operatorname{sc}_1(n, n) + 10T^{-1} \operatorname{sc}_0(m, n, m, n)$$

$$+ 8\delta_{s,l} \operatorname{sc}_0(m, m) \operatorname{sc}_0(n, n) + 8\delta_{s+l,T} \operatorname{sc}_0(m, n)^2\big],$$

and hence

$$
T B_T \left| \mathrm{var}\big(\Phi_\omega^{(T)}(m,n)\big) \right|
$$

$$
\leq B_T \left[T^{-1} \sum_{s=1}^{T-1} W^{(T)}(\omega - 2\pi s/T) \right]^2
$$

$$
\times \left[4T^{-1} \mathrm{sc}_1(m,m)\, \mathrm{sc}_1(n,n) + 10\, \mathrm{sc}_0(m,n,m,n) \right]
$$

$$
+ 8\, \mathrm{sc}_0(m,m)\, \mathrm{sc}_0(n,n) B_T T^{-1}
$$

$$
\times \sum_{s=1}^{T-1} \big(W^{(T)}(\omega - 2\pi s/T)\big)^2 + 8\, \mathrm{sc}_0(m,n)^2 B_T T^{-1}
$$

$$
\times \sum_{s=1}^{T-1} W^{(T)}(\omega - 2\pi s/T)\, W^{(T)}(\omega + 2\pi s/T).
$$

Since at most $\frac{T B_T}{\pi} + 1$ of the summands are nonzero and $\|W^{(T)}\|_\infty \leq B_T^{-1}\|W\|_\infty$ by Lemma F.11, we obtain $[T^{-1}\sum_{s=1}^{T-1} W^{(T)}(\omega - 2\pi s/T)]^2 \leq \pi^{-2}\|W\|_\infty^2$, and $B_T T^{-1}\sum_{s=1}^{T-1}(W^{(T)}(\omega - 2\pi s/T))^2 \leq \pi^{-1}\|W\|_\infty^2$, for large T. Similarly $|B_T T^{-1}\sum_{s=1}^{T-1} W^{(T)}(\omega - 2\pi s/T) W^{(T)}(\omega + 2\pi s/T)| \leq \pi^{-1}\|W\|_\infty^2$ for large T. Since $B_T \to 0$, for T large enough, we have

$$
T B_T \left| \mathrm{var}\big(\Phi_\omega^{(T)}(m,n)\big) \right| \leq \|W\|_\infty^2 \cdot \big[\mathrm{sc}_0(m,n,m,n) + \mathrm{sc}_1(m,m)\, \mathrm{sc}_1(n,n)
$$

$$
+ 8\, \mathrm{sc}_0(m,m)\, \mathrm{sc}_0(n,n) + 8\, \mathrm{sc}_0(m,n)^2 \big].
$$

Now (3.6) follows immediately by setting $K = 8\|W\|_\infty^2$.

To prove (3.7), notice that, for large T, inequality (3.6) gives us

$$
\sum_{m,n \geq 1} T B_T \, \mathrm{var}\big(\Phi_\omega^{(T)}(m,n)\big)
$$

$$
\leq K \left[\sum_{m,n \geq 1} \mathrm{sc}_0(m,n,m,n) + \left(\sum_{m \geq 1} \mathrm{sc}_1(m,m) \right)^2 + \left(\sum_{m \geq 1} \mathrm{sc}_0(m,m) \right)^2 \right.
$$

$$
\left. + \sum_{m,n \geq 1} \mathrm{sc}_0(m,n)^2 \right].
$$

Notice that $\mathrm{cum}(\xi_{t_1}(m), \xi_{t_2}(n), \xi_{t_3}(m), \xi_0(n)) = \langle \mathcal{R}_{t_1,t_2,t_3}\varphi_m \otimes \varphi_n, \varphi_m \otimes \varphi_n \rangle$, hence $\sum_{m,n \geq 1} \mathrm{sc}_0(m,n,m,n) \leq \sum_{t_1,t_2,t_3 \in \mathbb{Z}} \||\mathcal{R}_{t_1,t_2,t_3}\||_1$. We also have $\mathrm{cum}(\xi_t(m), \xi_0(n)) = \langle \mathcal{R}_t\varphi_n, \varphi_m \rangle$, hence $\sum_{m \geq 1} \mathrm{sc}_0(m,m) \leq \sum_{t \in \mathbb{Z}} \||\mathcal{R}_t\||_1$. Using the Cauchy–Schwarz inequality and Parseval's identity, we also obtain $\sum_{m,n \geq 1} \mathrm{sc}_0(m,n)^2 \leq (\sum_{t \in \mathbb{Z}} \||\mathcal{R}_t\||_1)^2$. Similarly, $\sum_{m,n \geq 1} \mathrm{sc}_1(m,m)\, \mathrm{sc}_1(n,n) \leq (\sum_{t \in \mathbb{Z}} |t| \||\mathcal{R}_t\||_1)^2$. Inequality (3.7) is then established by noticing that both $\sum_t \||\mathcal{R}_t\||_1$ and $\sum_t |t| \||\mathcal{R}_t\||_1$ are bounded by $\sum_{t \in \mathbb{Z}} (1 + |t|) \||\mathcal{R}_t\||_1$, and setting $C = 3K$.

We can now put (3.6) and (3.7) to use in order to establish the main result. We first show that $\sqrt{T B_T}(f_{\omega_j}^{(T)} - \mathbb{E}f_{\omega_j}^{(T)})$ is tight. Choose an orthonormal basis φ_n of $L^2([0,1], \mathbb{R})$. Notice that

$$\mathbb{E}[\langle \sqrt{T B_T}(f_{\omega_j}^{(T)} - \mathbb{E}f_{\omega_j}^{(T)}), \varphi_m \otimes \varphi_n \rangle^2] = T B_T \operatorname{var}(\Phi_\omega^{(T)}(m,n)).$$

Since $(\varphi_m \otimes \varphi_n)_{n,m \geq 1}$ is an orthonormal basis of $L^2([0,1]^2, \mathbb{C})$, the tightness of $\sqrt{T B_T}(f_{\omega_j}^{(T)} - \mathbb{E}f_{\omega_j}^{(T)})$ follows from (3.6), (3.7) and Lemma 7.1. Therefore the vector $\sqrt{T B_T}(f_{\omega_1}^{(T)} - \mathbb{E}f_{\omega_1}^{(T)}, \ldots, f_{\omega_J}^{(T)} - \mathbb{E}f_{\omega_J}^{(T)})$ is also tight in $(L^2([0,1]^2, \mathbb{C}))^J$. Applying Brillinger (2001), Theorem 7.4.4, to the finite-dimensional distributions of this vector completes the proof. $\square$

Note here that condition (i) for $k = 2$ is $\sum_{t \in \mathbb{Z}} \|r_t\|_2 < \infty$, which guarantees that the spectral density operator is continuous in ω with respect to the Hilbert–Schmidt norm. If in addition we want it to be continuous in τ, σ we need to assume the stronger conditions $\sum_{t \in \mathbb{Z}} \|r_t\|_\infty < \infty$, and that each r_t is continuous.

When $\omega = 0$, the operator $2\pi \mathscr{F}_\omega$ reduces to the long-run covariance operator $\sum_{t \in \mathbb{Z}} \mathscr{R}_t$, the limiting covariance operator of the empirical mean. Correspondingly, $2\pi \mathscr{F}_0^{(T)}$ is an estimator of the long-run covariance operator that is consistent in mean square for the long-run covariance, under no structural modelling assumptions. A similar estimator was also considered in Horváth, Kokoszka and Reeder (2013), who derived weak consistency under L^p-m-approximability weak dependence conditions. Hörmann and Kokoszka (2010) studied this problem by projecting onto a finite-dimensional subspace. However, neither of these papers considers functional central limit theorems for the estimator of the long-run covariance operator; taking $\omega = 0$, in Theorem 3.7, we obtain such a result:

COROLLARY 3.8. *Under the conditions of Theorem 3.7, we have*

$$\sqrt{B_T T}(2\pi \mathscr{F}_0^{(T)} - 2\pi \mathbb{E}\mathscr{F}_0^{(T)}) \xrightarrow{d} \mathcal{N}(0, (2\pi)^{3/2}\|W\|_2^2 \mathfrak{C}),$$

where $\mathfrak{C}$ is the integral operator on $L^2([0,1]^2, \mathbb{R})$ with kernel

$$c(\tau_1, \sigma_1, \tau_2, \sigma_2) = \{f_0(\tau_1, \tau_2)f_0(\sigma_1, \sigma_2) + f_0(\tau_1, \sigma_2)f_0(\sigma_1, \tau_2)\}.$$

We remark that the limiting Gaussian random operator is purely real.

4. Weak dependence, tightness and projections. Our results on the asymptotic Gaussian representations of the discrete Fourier transform and the spectral density estimator (Theorems 2.2 and 3.7) effectively rest upon two sets of weak dependence conditions: (1) the summability of the nuclear norms of the autocovariance operators (at various rates), and (2) the summability of the cumulant kernels of all orders (at various rates). The roles of these two sets of weak dependence

conditions are distinct. The first is required in order to establish tightness of the sequence of discrete Fourier transforms and spectral density estimators of the underlying process. Tightness allows one to then apply the Cramér–Wold device, and to determine the asymptotic distribution by considering finite-dimensional projections; see, for example, Ledoux and Talagrand (1991). The role of the second set of weak dependence conditions, then, is precisely to allow the determination of the asymptotic law of the projections, thus identifying the stipulated limiting distribution via tightness.

Therefore, in principle, one can replace the second set of weak dependence conditions with a set of conditions that allow for the discrete Fourier transforms and spectral density estimators of the *vector* time series of the projections to be asymptotically Gaussian, jointly in any finite number of frequencies. Our approach was to generalise the cumulant multivariate conditions of Brillinger (2001), which do not require structural assumptions further to stationarity. Alternatively, one may pursue generalizations of multivariate conditions involving α-mixing and summable cumulants of order 2, 4, and 8 as in Hannan (1970), Chapter IV, Paragraph 4 and Rosenblatt (1984, 1985), though α-mixing can also be a strong condition. Adding more structure, for example, in the context of linear processes, one can focus on extending weaker conditions requiring finite fourth moments and summable coefficients [Anderson (1994), Hannan (1970)].

For the case of nonlinear moving-average representations of the form $\xi_t = G(\varepsilon_t, \varepsilon_{t-1}, \ldots)$, where G is a measurable function, and $\{\varepsilon_j\}$ are i.i.d. random variables, several results exist; however, none of them are (yet) established for vector time series. For instance Shao and Wu (2007) show that if the second moment of ξ_t is finite and

$$\sum_{k=0}^{\infty} \sqrt{\mathbb{E}\big|\mathbb{E}[\xi_k - \xi_{k+1}|\mathcal{F}_0]\big|^2} < \infty,$$

where $\mathcal{F}_0$ is the sigma-algebra generated by $\{\varepsilon_0, \varepsilon_{-1}, \ldots\}$, then the discrete Fourier transforms of ξ_t are asymptotically Gaussian, jointly for a finite number of frequencies. Furthermore, Shao and Wu (2007) establish the asymptotic normality of the spectral density estimator at distinct frequencies under the moment condition $\mathbb{E}|\xi_t|^{4+\delta} < \infty$, and the following coupling condition: there exist $\alpha > 0$, $C > 0$ and $\rho \in (0, 1)$ such that

$$(4.1) \qquad \mathbb{E}\big|\xi_t - \xi_t'\big|^\alpha < C\rho^t \qquad \forall t = 0, 1, \ldots,$$

where $\xi_t' = G(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0', \varepsilon_{-1}', \ldots)$ and $(\varepsilon_k')_{k \in \mathbb{Z}}$ is an i.i.d. copy of $(\varepsilon_k)_{k \in \mathbb{Z}}$. Notice that (4.1) is related to (in fact stronger than) the L^p-m-approximability condition of Hörmann and Kokoszka (2010). Under the weaker conditions $\mathbb{E}|\xi_t|^4 < \infty$, and

$$\sum_{t=0}^{\infty} \big(\mathbb{E}|\xi_t - \check{\xi}_t|^4\big)^{1/4} < \infty,$$

where $\check{\xi}_t = G(\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_t)$ and ε_0' is an i.i.d. copy of ε_0, Liu and Wu (2010) establish that the spectral density estimator at a *fixed* frequency is asymptotically Gaussian. The idea behind these coupling conditions is to approximate the series ξ_t by m-dependent series, for which derivation of asymptotic results is easier. We also mention that, under milder conditions, Peligrad and Wu (2010) establish that for almost all $\omega \in (0, 2\pi)$, the discrete Fourier transform at ω is marginally asymptotically normal.

The weak dependence conditions pursued in this paper have the advantage of not requiring additional structure, at the price of being relatively strong if additional structure could be assumed. For example, if a process is linear, the cumulant conditions will be satisfied provided all moments exist and the coefficient operators are summable in an appropriate sense, as shown in the proposition below. As mentioned above, we conjecture that four moments and summability of the coefficients would suffice in the linear case; however, a more thorough study of weak dependence conditions for the linear case is outside the scope of the present paper.

PROPOSITION 4.1.   *Let $X_t = \sum_{s \in \mathbb{Z}} A_s \varepsilon_{t-s}$ be a linear process with $\mathbb{E}\|\varepsilon_0\|_2^p < \infty$ for all $p \geq 1$, and $\sum_{s \in \mathbb{Z}}(1 + |s|^l)\|a_s\|_2 < \infty$ for some positive integer l, where a_s is the kernel of A_s. Then for all fixed $k = 1, 2, \ldots$, X_t satisfies* C(l, k),

$$\sum_{t_1, \ldots, t_{k-1} \in \mathbb{Z}} (1 + |t_j|^l)\|\mathrm{cum}(X_{t_1}, \ldots, X_{t_{k-1}}, X_0)\|_2 < \infty, \qquad \forall j = 1, \ldots, k-1.$$

Furthermore,

$$\sum_{t \in \mathbb{Z}}(1 + |t|^l)|\|\mathscr{R}_t\||_1 < \infty,$$

$$\sum_{t_1, t_2, t_3} \||\mathrm{cum}(X_{t_1}, X_{t_2}, X_{t_3}, X_0)\||_1 < \infty,$$

where we view $\mathrm{cum}(X_{t_1}, X_{t_2}, X_{t_3}, X_0)$ *as an operator on* $L^2([0,1]^2, \mathbb{R})$; *see Section 2.*

PROOF.   See Proposition E.1 in the supplementary material [Panaretos and Tavakoli (2013)].   □

5. The effect of discrete observation.   In practice, functional data are often observed on a discrete grid, subject to measurement error, and smoothing is employed to make the transition into the realm of smooth functions. This section considers the stability of the consistency of our estimator of the spectral density operator with respect to discrete observation of the underlying stationary functional process. Since our earlier results do not a priori require any smoothness of the functional data, except perhaps smoothness that is imposed by our weak dependence conditions, we consider a "minimal" scenario where the curves are

only assumed to be continuous in mean square. Under this weak assumption, we formalise the asymptotic discrete observation framework via observation on an increasingly dense grid subject to measurement error of variance decreasing at a certain rate [e.g., Hall and Vial (2006)]. In principle, one may drop the assumption that the noise variance decreases at a certain rate at the expense of smoothness assumptions on the curves that would suffice for smoothers constructed via the noisy sampled curves to converge to the true curves, at a corresponding mean squared error rate.

Let Γ be the grid $0 = \tau_1 < \tau_2 < \cdots < \tau_M < \tau_{M+1} = 1$ on $[0, 1]$, with $M = M(T)$ being a function of T such that $M(T) \to \infty$ as $T \to \infty$, and

$$|\Gamma| = \sup_{j=1,\ldots,M+1} \tau_j - \tau_{j-1} \to 0, \qquad M \to \infty.$$

Assume we observe the curves X_t on this grid (except possibly at τ_{M+1}), additively corrupted by measurement error, represented by independent and identically distributed random variables $\{\varepsilon_{tj}\}$ (and independent of the X_t's),

$$y_{tj} = X_t(\tau_j) + \varepsilon_{tj},$$

with $\mathbb{E}\varepsilon_{tj} = 0$ and $\sqrt{\text{var}(\varepsilon_{tj})} = \sigma(M)$. Our goal is to show that our estimator of $\mathscr{F}_\omega^{(T)}$, when constructed on the basis of the y_{tj}'s, retains its consistency for the true spectral density operator. To construct our estimator on the basis of discrete observations, we use the following (naive) proxy of the true X_t,

$$_{\varepsilon,s}X_t(\tau) = y_{tj} \qquad \text{if } \tau_j \leq \tau < \tau_{j+1},$$

and define the step-wise version of X_t,

$$_sX_t(\tau) = X_t(\tau_j) \qquad \text{if } \tau_j \leq \tau < \tau_{j+1}.$$

Just as the spectral density kernel estimator $f_\omega^{(T)}$ is a functional of the X_t's, we can define $_{\varepsilon,s}f_\omega^{(T)}$ and $_sf_\omega^{(T)}$, as the corresponding functionals of the $_{\varepsilon,s}X_t$'s, $_sX_t$, respectively. The same can also be done for f_ω, $\mathscr{F}_\omega^{(T)}$, $p_\omega^{(T)}$, $\widetilde{X}_\omega^{(T)}$. We then have the following stability result.

THEOREM 5.1. *Under* C(1, 2), *if* $\mathbb{E}\varepsilon_{tj}^4 < \infty$, $\sigma^2(M) = o(B_T)$, $B_T = o(1)$ *such that* $T B_T \to \infty$, *and if*

(5.1) $$\text{each } r_t \text{ is continuous, and } \sum_t \|r_t\|_\infty < \infty$$

holds, then

$$\int_{-\pi}^{\pi} \mathbb{E}\||_{\varepsilon,s}\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega^{(T)}\||_2^2 \, d\omega \to 0, \qquad T \to \infty.$$

Moreover, we also have pointwise mean square convergence for a fixed ω,

$$\mathbb{E}\||_{\varepsilon,s}\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega^{(T)}\||_2^2 \to 0, \qquad T \to \infty$$

under the same conditions if $0 < |\omega| < \pi$, and under the stronger condition $TB_T^2 \to \infty$ if $\omega = 0, \pm\pi$.

PROOF OF THEOREM 5.1. First, we use the triangle inequality,

$$\int_{-\pi}^{\pi} \mathbb{E}\||_{\varepsilon,s}\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega^{(T)}\||_2^2 \, d\omega = \int_{-\pi}^{\pi} \iint \mathbb{E}|_{\varepsilon,s}f_\omega^{(T)} - f_\omega^{(T)}|^2 \, d\omega$$

$$(5.2) \qquad\qquad \leq 2\int_{-\pi}^{\pi} \iint \mathbb{E}|_{\varepsilon,s}f_\omega^{(T)} - {}_sf_\omega^{(T)}|^2 \, d\omega$$

$$(5.3) \qquad\qquad + 2\int_{-\pi}^{\pi} \iint \mathbb{E}|_sf_\omega^{(T)} - f_\omega^{(T)}|^2 \, d\omega.$$

The inner integrals are on $[0, 1]^2$ with respect to $d\tau \, d\sigma$. First, we deal with the first summand,

$$|_{\varepsilon,s}f_\omega^{(T)} - {}_sf_\omega^{(T)}|^2 = 2\pi T^{-2}\left|\sum_{l=0}^{T-1} W^{(T)}(\omega - 2\pi l/T)({}_{\varepsilon,s}p_{2\pi l/T}^{(T)} - {}_sp_{2\pi l/T}^{(T)})\right|^2$$

$$\leq O(T^{-1})\sum_{l=0}^{T-1}[W^{(T)}(\omega - 2\pi l/T)]^2|_{\varepsilon,s}p_{2\pi l/T}^{(T)} - {}_sp_{2\pi l/T}^{(T)}|^2,$$

where we have used Jensen's inequality. We claim that, if $\tau_j \leq \tau < \tau_{j+1}$ and $\tau_k \leq \sigma < \tau_{k+1}$,

$$|_{\varepsilon,s}p_\omega^{(T)}(\tau,\sigma) - {}_sp_\omega^{(T)}(\tau,\sigma)|^2 \leq 3|_s\widetilde{X}_\omega^{(T)}(\tau)|^2|\widetilde{\varepsilon}_{-\omega}^{(T)}(k)|^2$$

$$+ 3|\widetilde{\varepsilon}_\omega^{(T)}(j)\widetilde{\varepsilon}_{-\omega}^{(T)}(k)|^2 + 3|\widetilde{\varepsilon}_\omega^{(T)}(j)|^2|_s\widetilde{X}_{-\omega}^{(T)}(\sigma)|^2,$$

where $\widetilde{\varepsilon}_\omega^{(T)}(j) = (2\pi T)^{-1/2}\sum_{l=0}^{T-1} e^{-i\omega t}\varepsilon_{tj}$. To see this, we note that

$$_{\varepsilon,s}p_\omega^{(T)}(\tau,\sigma) - {}_sp_\omega^{(T)}(\tau,\sigma) = {}_{\varepsilon,s}\widetilde{X}_\omega^{(T)}(\tau) \cdot {}_{\varepsilon,s}\widetilde{X}_{-\omega}^{(T)}(\sigma) - {}_s\widetilde{X}_\omega^{(T)}(\tau) \cdot {}_s\widetilde{X}_{-\omega}^{(T)}(\sigma)$$

$$= ({}_{\varepsilon,s}\widetilde{X}_\omega^{(T)}(\tau) - {}_s\widetilde{X}_\omega^{(T)}(\tau)) \cdot {}_{\varepsilon,s}\widetilde{X}_{-\omega}^{(T)}(\sigma)$$

$$+ {}_s\widetilde{X}_\omega^{(T)}(\tau) \cdot ({}_{\varepsilon,s}\widetilde{X}_{-\omega}^{(T)}(\sigma) - {}_s\widetilde{X}_{-\omega}^{(T)}(\sigma))$$

$$= {}_s\widetilde{X}_\omega^{(T)}(\tau)\widetilde{\varepsilon}_{-\omega}^{(T)}(k) + \widetilde{\varepsilon}_\omega^{(T)}(j)\widetilde{\varepsilon}_{-\omega}^{(T)}(k)$$

$$+ \widetilde{\varepsilon}_\omega^{(T)}(j)_s\widetilde{X}_{-\omega}^{(T)}(\sigma),$$

since $_{\varepsilon,s}\widetilde{X}_\omega^{(T)}(\tau) = {}_s\widetilde{X}_\omega^{(T)}(\tau) + \widetilde{\varepsilon}_\omega^{(T)}(j)$, and similarly if we replace σ by τ and j by k. Our claim thus follows from Jensen's inequality.

In order to bound the expectation of $|_{\varepsilon,s}p_\omega^{(T)}(\tau,\sigma) - {}_s p_\omega^{(T)}(\tau,\sigma)|^2$, we will first compute the expectation, conditional on the σ-algebra generated by the X_t's, which we will denote by $\mathbb{E}_X$, and then use the tower property. As an intermediate step, we claim that $\mathbb{E}_X|\widetilde{\varepsilon}_\omega^{(T)}(j)|^2 = O(\sigma^2(M))$,

$$\mathbb{E}_X|\widetilde{\varepsilon}_\omega^{(T)}(j)\widetilde{\varepsilon}_{-\omega}^{(T)}(k)|^2 = \begin{cases} O(\sigma^4(M)), & \text{if } j \neq k, \\ O(\sigma^4(M)) + O(T^{-1}), & \text{if } j = k, \end{cases}$$

uniformly in j, k (notice that all $\mathbb{E}_X$ can be replaced by $\mathbb{E}$ since the ε_{tj}'s are independent of the X_t's). To establish this, notice that $|\widetilde{\varepsilon}_\omega^{(T)}(j)|^2 = \widetilde{\varepsilon}_\omega^{(T)}(j)\widetilde{\varepsilon}_{-\omega}^{(T)}(j)$, hence $\mathbb{E}_X|\widetilde{\varepsilon}_\omega^{(T)}(j)|^2 = (2\pi T)^{-1}\sum_{t,s=0}^{T-1} e^{-i\omega(t-s)}\mathbb{E}[\varepsilon_{tj}\varepsilon_{sj}]$. The summand is equal to $\sigma^2(M)$ if $t = s$, and zero otherwise (by independence of the ε's), hence the first statement follows directly. The case $j \neq k$ follows from the first statement, once the independence of $\widetilde{\varepsilon}_\omega^{(T)}(j)$ and $\widetilde{\varepsilon}_{-\omega}^{(T)}(k)$ has been noticed. We can now turn to the case $j = k$. First notice that

$$\mathbb{E}_X|\widetilde{\varepsilon}_\omega^{(T)}(j)\widetilde{\varepsilon}_{-\omega}^{(T)}(j)|^2 = (2\pi T)^{-2}\sum_{t_1,t_2,t_3,t_4=0}^{T-1} e^{-i\omega[(t_1-t_2)+(t_3-t_4)]}\mathbb{E}_X(\varepsilon_{t_1}\varepsilon_{t_2}\varepsilon_{t_3}\varepsilon_{t_4}),$$

where we have written ε_t instead of ε_{tj} for tidiness. The expectation of the product of the ε's is equal to zero if at least one of the t_l's is different from the all the other ones (by independence). So we may assume that each ε_{t_l} appears at least twice. There can be therefore $2 - r$ distinct terms in $\varepsilon_{t_1}\varepsilon_{t_2}\varepsilon_{t_3}\varepsilon_{t_4}$, where $r = 0$ or 1. If $r = 0$, $\mathbb{E}_X(\varepsilon_{t_1}\varepsilon_{t_2}\varepsilon_{t_3}\varepsilon_{t_4}) = \sigma^4(M)$, and if $r = 1$, $\mathbb{E}_X(\varepsilon_{t_1}\varepsilon_{t_2}\varepsilon_{t_3}\varepsilon_{t_4}) = \mathbb{E}_X\varepsilon^4 = \mathbb{E}\varepsilon^4$. Thus $\mathbb{E}_X|\widetilde{\varepsilon}_\omega^{(T)}(j)\widetilde{\varepsilon}_{-\omega}^{(T)}(j)|^2 = (2\pi T)^{-2}[N_0\sigma^4(M) + N_1\mathbb{E}\varepsilon^4]$, where N_r is the number of ways we can assign integers $t_1,\ldots,t_4$ in $\{0,\ldots,T-1\}$ such that each t_l appears at least twice and exactly $2 - r$ distinct integers appear. Simple combinatorics yield $N_0 = \binom{4}{2}T(T-1) = 6T(T-1)$, and $N_1 = T$, and so the case case $j \neq k$ follows directly since $\mathbb{E}\varepsilon^4 < \infty$.

We can now bound $\mathbb{E}_X|_{\varepsilon,s}p_{2\pi l/T}^{(T)} - {}_s p_{2\pi l/T}^{(T)}|^2$,

$$\begin{aligned}
\mathbb{E}_X|_{\varepsilon,s}p_{2\pi l/T}^{(T)} - {}_s p_{2\pi l/T}^{(T)}|^2 &\leq 3|_s\widetilde{X}_{2\pi l/T}^{(T)}(\tau)|^2\mathbb{E}_X|\widetilde{\varepsilon}_{-2\pi l/T}^{(T)}(k)|^2 \\
&\quad + 3\mathbb{E}_X|\widetilde{\varepsilon}_{2\pi l/T}^{(T)}(j)\widetilde{\varepsilon}_{-2\pi l/T}^{(T)}(k)|^2 \\
&\quad + 3|_s\widetilde{X}_{-2\pi l/T}^{(T)}(\sigma)|^2\mathbb{E}_X|\widetilde{\varepsilon}_{2\pi l/T}^{(T)}(j)|^2 \\
&\leq O(\sigma^2(M))[|_s\widetilde{X}_{2\pi l/T}^{(T)}(\tau)|^2 + |_s\widetilde{X}_{-2\pi l/T}^{(T)}(\sigma)|^2] \\
&\quad + O(\sigma^4(M)) + O(T^{-1}).
\end{aligned}$$

Since $|\widetilde{X}_\omega^{(T)}(\tau)|^2 = |p_\omega^{(T)}(\tau,\tau)|$, Proposition 2.6, Remark 3.5 and (5.1) yield that $\int \mathbb{E}|_s \widetilde{X}_{2\pi l/T}^{(T)}(\tau)|^2 \, d\tau = O(1)$. Using the tower property, we obtain

$$\iint \mathbb{E}|_{\varepsilon,s} p_{2\pi l/T}^{(T)} - {}_s p_{2\pi l/T}^{(T)}|^2 \leq O(\sigma^2(M)) + O(T^{-1}),$$

uniformly in $l = 1, \ldots, T-1$ under the assumptions of this theorem. Thus

$$\iint \mathbb{E}|_{\varepsilon,s} f_\omega^{(T)} - {}_s f_\omega^{(T)}|^2$$

$$\leq O(T^{-1}) \sum_{l=0}^{T-1} [W^{(T)}(\omega - 2\pi l/T)]^2 \cdot \iint \mathbb{E}|_{\varepsilon,s} p_{2\pi l/T}^{(T)} - {}_s p_{2\pi l/T}^{(T)}|^2$$

$$= O(B_T^{-1}\sigma^2(M)) + O(B_T T)^{-1},$$

uniformly in ω. Hence we obtain the bound on the expectation of first summand (5.2),

$$\int_{-\pi}^{\pi} \iint \mathbb{E}|_{\varepsilon,s} p_{2\pi l/T}^{(T)} - {}_s p_{2\pi l/T}^{(T)}|^2 \, d\omega = O(B_T^{-1}\sigma^2(M)) + O(B_T T)^{-1},$$

under the assumptions of the theorem.

We now turn to the second summand (5.3). First notice that

$$\int_{-\pi}^{\pi} \iint \mathbb{E}|_s f_\omega^{(T)} - f_\omega^{(T)}|^2 \, d\omega = 2 \int_0^{\pi} \iint \mathbb{E}|_s f_\omega^{(T)} - f_\omega^{(T)}|^2 \, d\omega,$$

since ${}_s f_{-\omega}^{(T)} = \overline{{}_s f_\omega^{(T)}}$ and $f_{-\omega}^{(T)} = \overline{f_\omega^{(T)}}$. Using the decomposition

$$\mathbb{E}|_s f_\omega^{(T)} - f_\omega^{(T)}|^2 = \mathrm{cov}[{}_s f_\omega^{(T)} - f_\omega^{(T)}, {}_s f_\omega^{(T)} - f_\omega^{(T)}] + |\mathbb{E}[{}_s f_\omega^{(T)} - f_\omega^{(T)}]|^2,$$

the covariance term can be written as sums and differences of four terms of the form $\mathrm{cov}(f_\omega^{(T)}(\sigma_1, \sigma_2), f_\omega^{(T)}(\sigma_3, \sigma_4))$, for some σ_l's. The important thing here is that each of these terms can be bounded in L^2—independently of the σ_l's—using Corollary 3.3 and Proposition 3.4,

$$\mathrm{cov}[{}_s f_\omega^{(T)} - f_\omega^{(T)}, {}_s f_\omega^{(T)} - f_\omega^{(T)}]$$

$$= \begin{cases} O(B_T^{-2}T^{-1}) + O(T^{-1}), & \text{if } \omega \in [0, B_T] \cup [\pi - B_T, \pi], \\ O(B_T^{-1}T^{-1}), & \text{if } \omega \in [B_T, \pi - B_T] \end{cases}$$

in L^2. Hence decomposing $\int_0^{\pi} = \int_0^{B_T} + \int_{B_T}^{\pi-B_T} + \int_{\pi-B_T}^{\pi}$, we obtain

$$\int_0^{\pi} \iint \mathrm{cov}[{}_s f_\omega^{(T)} - f_\omega^{(T)}, {}_s f_\omega^{(T)} - f_\omega^{(T)}] \, d\omega = O(B_T^{-1}T^{-1}),$$

if $B_T \to 0$.

In order to bound $|\mathbb{E}[_sf_\omega^{(T)} - f_\omega^{(T)}]|^2$, we use Proposition 3.1 and Lemma F.4 (with $p = 1$),

$$\iint |\mathbb{E}[_sf_\omega^{(T)} - f_\omega^{(T)}]|^2 \le 4 \iint |_sf_\omega - f_\omega|^2 + O(B_T^2) + O(T^{-2}) + O(B_T T)^{-2},$$

uniformly in ω. Thus

$$\int_0^\pi \iint |\mathbb{E}[_sf_\omega^{(T)} - f_\omega^{(T)}]|^2 \, d\omega$$

$$\le 4 \int_{-\pi}^\pi \iint |_sf_\omega - f_\omega|^2 \, d\omega + O(B_T^2) + O(T^{-2}) + O(B_T T)^{-2}.$$

The quantity $\iint |_sf_\omega - f_\omega|^2$ is in fact the the squared distance between $_sf_\omega$ and f_ω in the space $L^2([0,1]^2, \mathbb{C})$. Under (5.1), $f_\omega(\tau, \sigma)$ is uniformly continuous in ω, τ, σ; since $_sf_\omega$ is a step-wise approximation of f_ω, we obtain

$$\sup_{\omega \in [-\pi, \pi]} \iint |_sf_\omega - f_\omega|^2 \to 0, \qquad M \to \infty.$$

Piecing these results together, we obtain

$$\int_0^\pi \iint \mathbb{E}|_sf_\omega^{(T)} - f_\omega^{(T)}|^2 \, d\omega = o(1) + O(B_T^{-1} T^{-1}) + O(B_T^2),$$

and therefore

$$\int_{-\pi}^\pi \iint \mathbb{E}|_{\varepsilon,s}f_\omega^{(T)} - f_\omega^{(T)}|^2 \, d\omega = O(\sigma^2(M) B_T^{-1}) + O(B_T^{-1} T^{-1}) + O(B_T^2) + o(1),$$

where the $o(1)$ term comes from the L^2 distance between $_sf_\omega$ and f_ω. Under our assumptions, the right-hand side tends to zero as $T \to \infty$.

A careful examination of the proof reveals that the pointwise statement of the theorem follows with a directly analogous argument. $\square$

REMARK 5.2. The use of Proposition 3.4 was valid in this context, but requires some attention. Indeed, it relies on Lemma F.15 in the supplementary material [Panaretos and Tavakoli (2013)], applied to $g_{(\tau,\sigma)}(\alpha) = {}_sf_\alpha^{(T)}(\tau, \sigma)$. Remark F.16 in the supplementary material [Panaretos and Tavakoli (2013)] tells us that the convergence of the convolution integral depends on the uniform continuity parameter $\delta(\varepsilon)$, which here will depend on the size of the sampling grid $M = M(T)$; in other words, $\delta(\varepsilon) = \delta(\varepsilon, M)$. But notice that since (5.1) holds,

$$\|_sf_{\omega_1} - {}_sf_{\omega_2}\|_2 \le \sup_{0 \le \tau, \sigma \le 1} |_sf_{\omega_1}(\tau, \sigma) - {}_sf_{\omega_2}(\tau, \sigma)|$$

$$= \sup_{\tau, \sigma = \tau_1, \ldots, \tau_M} |f_{\omega_1}(\tau, \sigma) - f_{\omega_2}(\tau, \sigma)|$$

$$\le \sup_{0 \le \tau, \sigma \le 1} |f_{\omega_1}(\tau, \sigma) - f_{\omega_2}(\tau, \sigma)|,$$

hence we can choose a $\delta(\varepsilon)$ that is independent of M, and the application of Proposition 3.4 is valid.

6. Numerical simulations. In order to probe the finite sample performance of our estimators (in terms of IMSE), we have performed numerical simulations on stationary functional time series admitting a linear representation

$$X_t = \sum_{s=0}^{10} A_s \varepsilon_{t-s}.$$

We have taken the collection of innovation functions $\{\varepsilon_t\}$ to be independent Wiener processes on $[0, 1]$, which we have represented using a truncated Karhunen–Loève expansion,

$$\varepsilon_t(\tau) = \sum_{k=1}^{1000} \xi_{k,t} \sqrt{\lambda_k} e_k(\tau).$$

Here $\lambda_k = 1/[(k - 1/2)^2 \pi^2]$, $\xi_{k,t}$ are independent standard Gaussian variables and $e_k(\tau) = \sqrt{2} \sin[(k - 1/2)\pi\tau]$ is orthonormal system in $L^2([0, 1], \mathbb{R})$ [Adler (1990)]. We have constructed the operators A_s so that their image be contained within a 50-dimensional subspace of $L^2([0, 1], \mathbb{R})$, spanned by an orthonormal basis $\psi_1, \ldots, \psi_{50}$. Representing ε_t in the e_k basis, and A_s in the $\psi_m \otimes e_k$ basis, we obtain a matrix representation of the process X_t as $\mathbf{X}_t = \sum_{s=0}^{10} \mathbf{A}_s \boldsymbol{\varepsilon}_{t-s}$, where $\mathbf{X}_t$ is a 50×1 matrix, each $\mathbf{A}_s$ is a 50×1000 matrix, and each $\boldsymbol{\varepsilon}_t$ is a 1000×1 matrix.

We simulated a stretch of X_t, $t = 0, \ldots, T - 1$ for $T = 2^n$, with $n = 7, 8, \ldots, 15$. Typical functional data sets would range between $T = 2^6$ and $T = 2^8$ data points. We constructed the matrices $\mathbf{A}_s$, as random Gaussian matrices with independent entries, such that elements in row j where $N(0, j^{-2\alpha})$ distributed. When $\alpha = 0$, the projection of each ε_t onto the subspace spanned by each $\psi_m, m = 1, \ldots, 50$ has (roughly) a comparable magnitude. A positive value of α, for example, $\alpha = 1$ means that the projection of ε_t onto the subspace spanned by ψ_j will have smaller magnitude for larger j's.

For comparison purposes, we also carried out analogous simulations, but with $\lambda_k = 1$, that is, the variance of the innovations ε_t being equal to one in each direction $e_n, n = 1, \ldots, 1000$. In the sequel, we will refer to these as the simulations with "white noise innovations," and to the previous ones as "Wiener innovations." The white noise process is, of course, not a true white noise process, but a projection of a white noise process. However, it does represent a case of a "rough" innovation process, which we present here as an extreme scenario.

For each T, we generated 200 simulation runs which we used to compute the IMSE by approximating the integral

$$2 \int_0^\pi \mathbb{E} \big\| \mathscr{F}_\omega - \mathscr{F}_\omega^{(T)} \big\|_2^2 \, d\omega$$

by a weighted sum over the finite grid $\Gamma = \{\pi j/10; j = 0, \ldots, 9\}$. We chose $B_T = T^{-1/5}$ [e.g., Grenander and Rosenblatt (1957), Paragraph 4.7, Brillinger (2001),

Paragraph 7.4] and $W(x)$ to be the *Epanechnikov kernel* [e.g., Wand and Jones (1995)], $W(x) = \frac{3}{4}(1 - x^2)$ if $|x| < 1$, and zero otherwise. The results are shown in a log-log scale in Figure 1, for $\alpha = 2$. The slopes of the least square lines passing through the medians of the simulation results show that $\text{IMSE}(\mathscr{F}^{(T)}) \propto T^\beta$, with $\beta \approx -0.797$ for the white noise innovations, and $\beta \approx -0.796$ for the Wiener innovations. According to Theorem 3.6, the decay of the $\text{IMSE}(\mathscr{F}^{(T)})$ is bounded by

$$C_1 T^{-2/5} + C_2 T^{-4/5} \approx C_1 T^{-0.4} \qquad \text{(if } T \text{ is large)}$$

for some constants C_1, C_2.

In order to gain a visual appreciation of the accuracy of the estimators, we construct plots to compare the true and estimated spectral density kernels in Figures 2 and 3, for the Wiener and white noise cases, respectively. For practical purposes, we set $\alpha = 2$, as for the simulation of the IMSEs. We simulated $X_t = A_0 \varepsilon_t + A_1 \varepsilon_{t-1}$, where $\varepsilon_t(\tau)$ lies on the subspace of $L^2([0, 1], \mathbb{R})$ spanned by the basis $e_1, \ldots, e_{100}$, and the operators A_0, A_1 lie in the subspace spanned by $(\psi_m \otimes e_k)_{m=1,\ldots,51; k=1,\ldots,100}$. Since the target parameter is a complex-valued function defined over a two-dimensional rectangle, some information loss must be incurred when representing it graphically. We chose to suppress the phase component of the spectral density kernel, plotting only its amplitude, $|f_\omega(\tau, \sigma)|$, for all $(\tau, \sigma) \in [0, 1]^2$ and for selected frequencies ω (the spectral density kernel is seen to be smooth in ω, so this does not entail a significant loss of information). For various choices of sample size T, we have replicated the realisation of the process, and the corresponding kernel density estimator for the particular frequency. Each time, we plotted the contours in superposition, in order to be able to visually appreciate the variability in the estimators: tangled contour lines where no clear systematic pattern emerges signify a region of high variability, whereas aligned contour lines that adhere to a recognisable shape represent regions of low variability. As is expected, the "smoother" the innovation process, the less variable the results appear to be, and the variability decreases for larger values of T.

7. Background results and technical statements. Statements and proofs of intermediate results in functional analysis and probability in function space that are required in our earlier formal derivations, can be found in the supplementary material [Panaretos and Tavakoli (2013)]. This supplement also collects some known results and facts for the reader's ease. We include here a useful lemma that provides an easily verifiable L^2 moment condition that is sufficient for tightness to hold true. It collects arguments appearing in the proof of Bosq (2000), Theorem 2.7, and its proof can also be found in the supplementary material [Panaretos and Tavakoli (2013)].
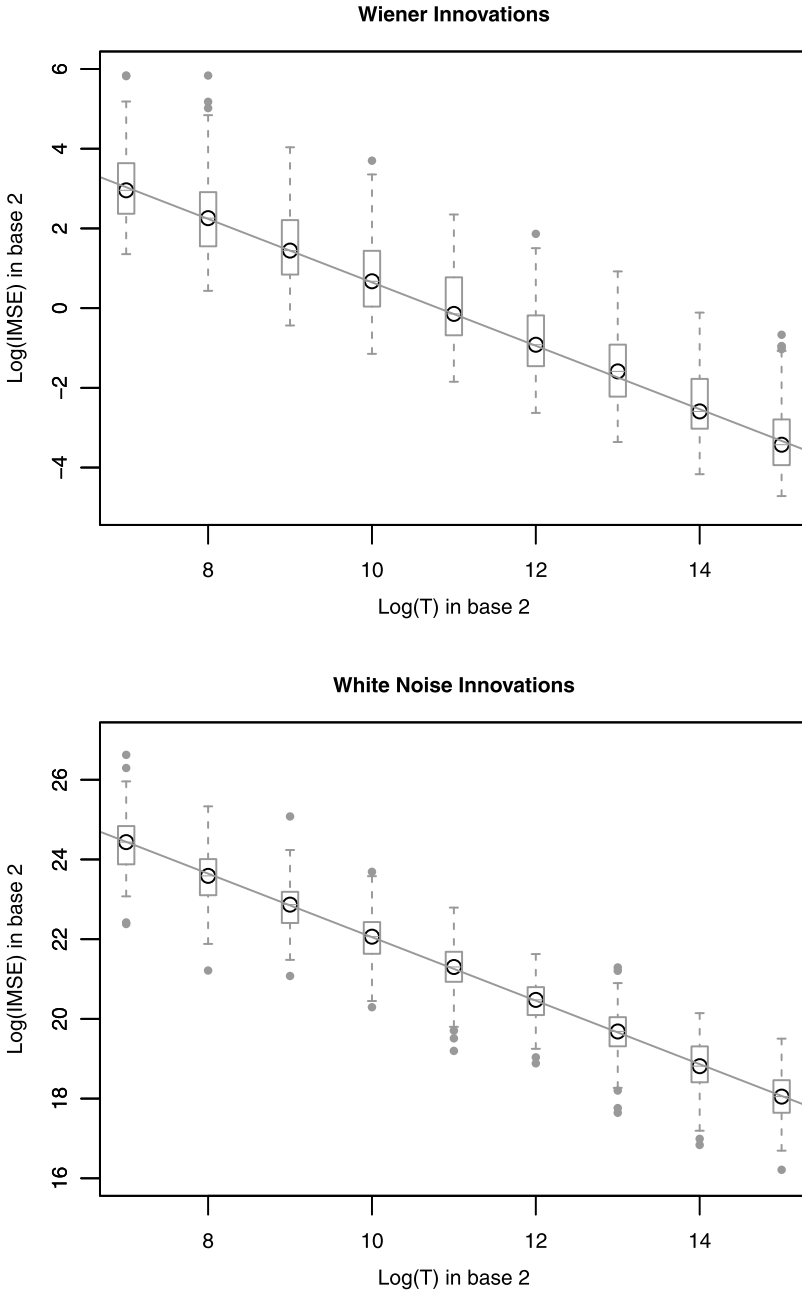
FIG. 1. *The results of the simulated ISE in a log–log scale, with* $\alpha = 2$*. The upper and lower plot correspond to the Wiener Innovations and the White Noise Innovation setups, respectively. The dots correspond to the median of the results of the simulations, and the lines are the least square lines of the medians. The boxplots summarise the distribution of the ISE for the* 200 *simulation runs. Though the ranges of the y-axes are different, the scales are the same, and the two least square lines are indeed almost parallel.*
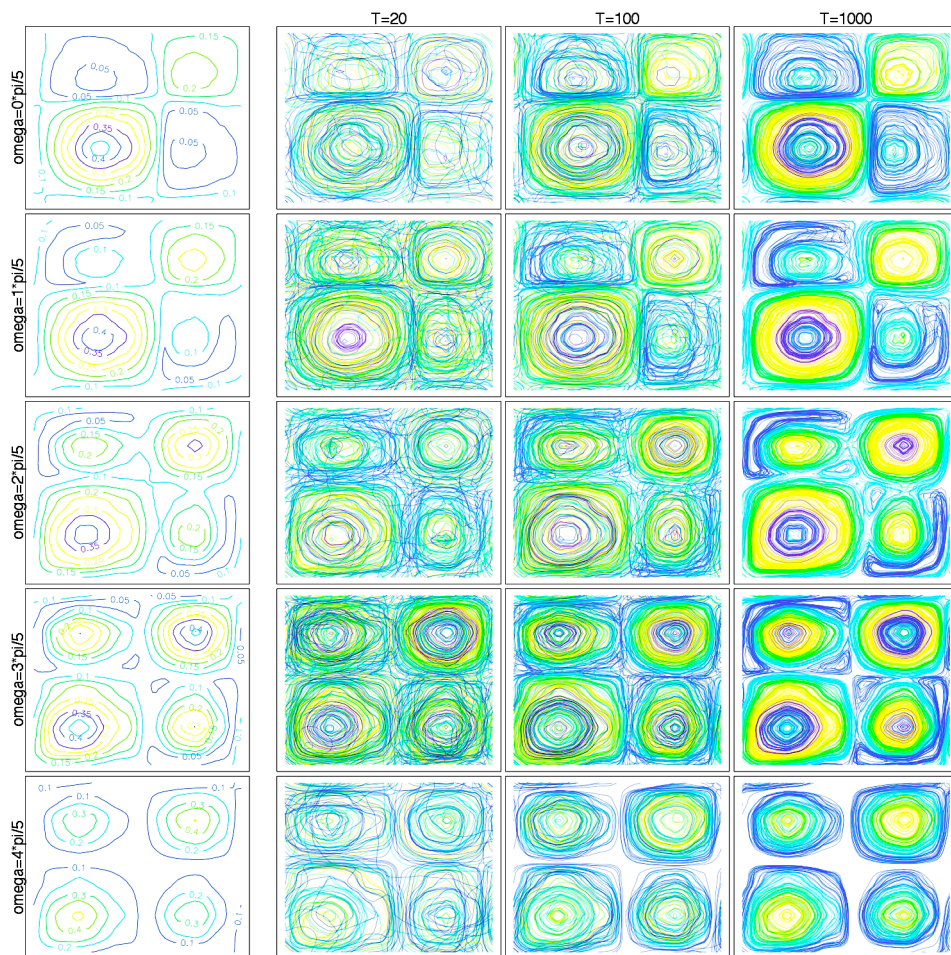
FIG. 2. *Contour plots for the amplitude of the true and estimated spectral density kernel when the innovation process consists of Wiener processes. Each row corresponds to a different frequency ($\omega = k\pi/5$, $k = 0, 1, \ldots, 4$, going from top to bottom). The first column contains the contour plots of the true amplitudes of the kernel at each corresponding frequency. The rest of the columns correspond to the estimated contours for different sample sizes ($T = 20, 100, 1000$ from left to right). Twenty estimates, corresponding to twenty replications of the process, have been superposed in order to provide a visual illustration of the variability. The contours plotted always correspond to the same level curves and use the same colour-coding in each row.*

LEMMA 7.1 (Criterion for tightness in Hilbert space). *Let H be a (real or complex) separable Hilbert space, and $X_T : \Omega \to H$, $T = 1, 2, \ldots$, be a sequence of random variables. If for some complete orthonormal basis $\{e_n\}_{n \geq 1}$ of H, we have $\mathbb{E}|\langle X_T, e_n \rangle|^2 \leq a_n$, $n = 1, 2, \ldots$, for all large T, and $\sum_{n \geq 1} a_n < \infty$, then $\{X_T\}_{T \geq 1}$ is tight.*
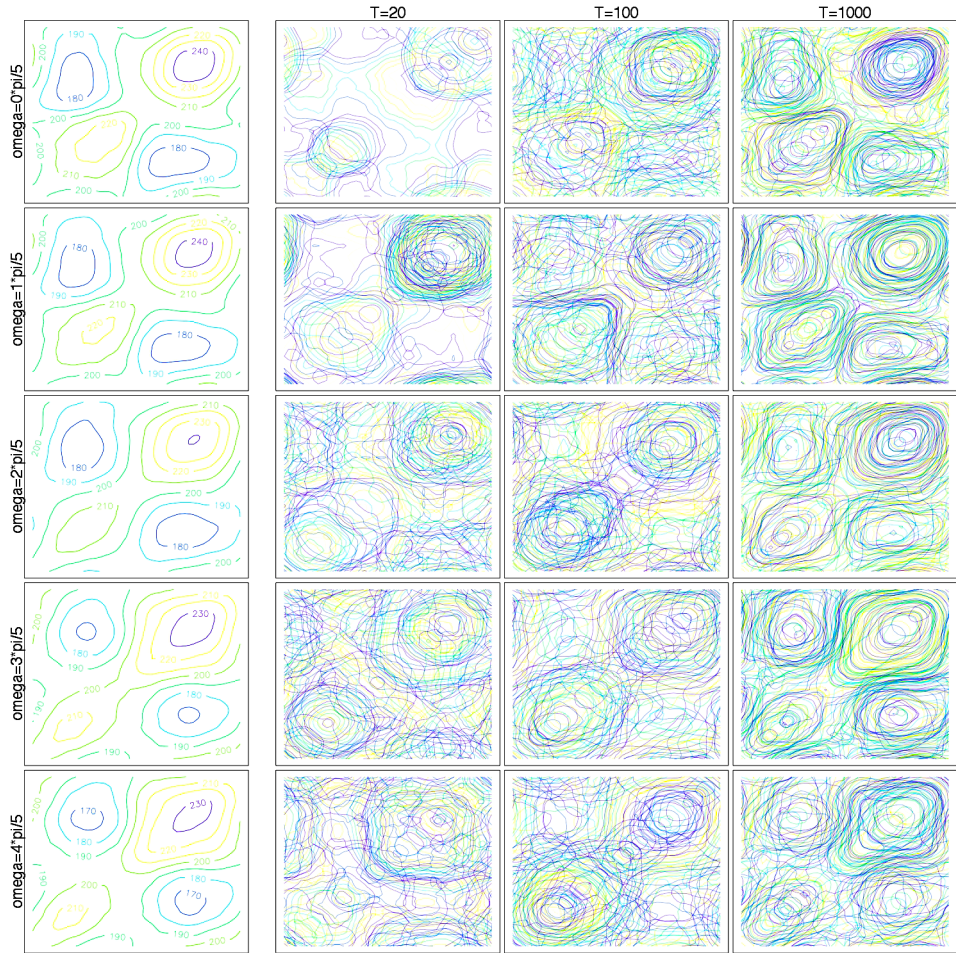
FIG. 3.   *Contour plots for the amplitude of the true and estimated spectral density kernel when the innovation process consists of white noise processes. Each row corresponds to a different frequency ($\omega = k\pi/5$, $k = 0, 1, \ldots, 4$, going from top to bottom). The first column contains the contour plots of the true amplitudes of the kernel at each corresponding frequency. The rest of the columns correspond to the estimated contours for different sample sizes ($T = 20, 100, 1000$ from left to right). Twenty estimates, corresponding to twenty replications of the process, have been superposed in order to provide a visual illustration of the variability. The contours plotted always correspond to the same level curves and use the same colour-coding in each row.*

Acknowledgements.   Our thanks go the Editor, Associate Editor and three referees for their careful reading and thoughtful comments.

SUPPLEMENTARY MATERIAL

Online Supplement: "Fourier Analysis of Stationary Time Series in Function Space" (DOI: [10.1214/13-AOS1086SUPP](); .pdf). The online supplement con-

tains the proofs that were omitted, and several additional technical results used in this paper.

REFERENCES

ADLER, R. J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **12**. IMS, Hayward, CA. MR1088478

ANDERSON, T. W. (1994). *The Statistical Analysis of Time Series*. Wiley, New York.

ANTONIADIS, A., PAPARODITIS, E. and SAPATINAS, T. (2006). A functional wavelet-kernel approach for time series prediction. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 837–857. MR2301297

ANTONIADIS, A. and SAPATINAS, T. (2003). Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *J. Multivariate Anal.* **87** 133–158. MR2007265

BENKO, M., HÄRDLE, W. and KNEIP, A. (2009). Common functional principal components. *Ann. Statist.* **37** 1–34. MR2488343

BLOOMFIELD, P. (2000). *Fourier Analysis of Time Series*: *An Introduction*, 2nd ed. Wiley, New York. MR1884963

BOENTE, G., RODRIGUEZ, D. and SUED, M. (2011). Testing the equality of covariance operators. In *Recent Advances in Functional Data Analysis and Related Topics* 49–53. Physica-Verlag/Springer, Heidelberg. MR2815560

BOSQ, D. (2000). *Linear Processes in Function Spaces*: *Theory and Applications. Lecture Notes in Statistics* **149**. Springer, New York. MR1783138

BOSQ, D. (2002). Estimation of mean and covariance operator of autoregressive processes in Banach spaces. *Stat. Inference Stoch. Process.* **5** 287–306. MR1943835

BOSQ, D. and BLANKE, D. (2007). *Inference and Prediction in Large Dimensions*. Wiley, Chichester. MR2364006

BRILLINGER, D. R. (2001). *Time Series*: *Data Analysis and Theory. Classics in Applied Mathematics* **36**. SIAM, Philadelphia, PA. MR1853554

CARDOT, H. and SARDA, P. (2006). Linear regression models for functional data. In *The Art of Semiparametrics* 49–66. Physica-Verlag/Springer, Heidelberg. MR2234875

CUEVAS, A., FEBRERO, M. and FRAIMAN, R. (2002). Linear functional regression: The case of fixed design and functional response. *Canad. J. Statist.* **30** 285–300. MR1926066

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

DEHLING, H. and SHARIPOV, O. S. (2005). Estimation of mean and covariance operator for Banach space valued autoregressive processes with dependent innovations. *Stat. Inference Stoch. Process.* **8** 137–149. MR2121674

EDWARDS, R. (1967). *Fourier Series*: *A Modern Introduction*. Holt, Rinehart & Winston, New York.

FERRATY, F. and VIEU, P. (2004). Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. *J. Nonparametr. Stat.* **16** 111–125. MR2053065

FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis*: *Theory and Practice*. Springer, New York. MR2229687

FERRATY, F., GOIA, A., SALINELLI, E. and VIEU, P. (2011a). Recent advances on functional additive regression. In *Recent Advances in Functional Data Analysis and Related Topics* 97–102. Physica-Verlag/Springer, Heidelberg. MR2815567

FERRATY, F., LAKSACI, A., TADJ, A. and VIEU, P. (2011b). Kernel regression with functional response. *Electron. J. Stat.* **5** 159–171. MR2786486

FREMDT, S., STEINEBACH, J., HORVÁTH, L. and KOKOSZKA, P. (2013). Testing the equality of covariance operators in functional samples. *Scand. J. Stat.* **40** 138–152.

GABRYS, R., HORVÁTH, L. and KOKOSZKA, P. (2010). Tests for error correlation in the functional linear model. *J. Amer. Statist. Assoc.* **105** 1113–1125. MR2752607

GABRYS, R. and KOKOSZKA, P. (2007). Portmanteau test of independence for functional observations. *J. Amer. Statist. Assoc.* **102** 1338–1348. MR2412554

GRENANDER, U. (1981). *Abstract Inference.* Wiley, New York. MR0599175

GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series.* Wiley, New York. MR0084975

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

HALL, P. and VIAL, C. (2006). Assessing the finite dimensionality of functional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 689–705. MR2301015

HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York. MR0279952

HÖRMANN, S. and KOKOSZKA, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38** 1845–1884. MR2662361

HORVÁTH, L., HUŠKOVÁ, M. and KOKOSZKA, P. (2010). Testing the stability of the functional autoregressive process. *J. Multivariate Anal.* **101** 352–367. MR2564345

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications.* Springer, New York. MR2920735

HORVÁTH, L., KOKOSZKA, P. and REEDER, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **75** 103–122.

HUNTER, J. K. and NACHTERGAELE, B. (2001). *Applied Analysis.* World Scientific, River Edge, NJ. MR1829589

KADISON, R. V. and RINGROSE, J. R. (1997). *Fundamentals of the Theory of Operator Algebras. Graduate Studies in Mathematics* **15**. Amer. Math. Soc., Providence, RI.

KARHUNEN, K. (1947). Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae. Ser. A I Math.-Phys.* **1947** 79. MR0023013

KOLMOGOROV, A. (1978). *Stationary Sequences in Hilbert Space.* National Translations Center [John Crerar Library], Chicago.

KRAUS, D. and PANARETOS, V. M. (2012). Disperson operators and resistant second-order functional data analysis. *Biometrika* **99** 813–832.

LAIB, N. and LOUANI, D. (2010). Nonparametric kernel regression estimation for functional stationary ergodic data: Asymptotic properties. *J. Multivariate Anal.* **101** 2266–2281. MR2719861

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*: Isoperimetry and Processes. *Ergebnisse der Mathematik und Ihrer Grenzgebiete* (3) [*Results in Mathematics and Related Areas* (3)] **23**. Springer, Berlin. MR1102015

LÉVY, P. (1948). *Processus stochastiques et mouvement Brownien. Suivi d'une note de M. Loève.* Gauthier-Villars, Paris. MR0029120

LIU, W. and WU, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory* **26** 1218–1245. MR2660298

LOCANTORE, N., MARRON, J. S., SIMPSON, D. G., TRIPOLI, N., ZHANG, J. T. and COHEN, K. L. (1999). Robust principal component analysis for functional data. *TEST* **8** 1–73. MR1707596

MAS, A. (2000). Estimation d'opérateurs de corrélation de processus linéaires fonctionnels: lois limites, déviations modérées. Ph.D. thesis, Université Paris VI.

PANARETOS, V. M., KRAUS, D. and MADDOCKS, J. H. (2010). Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *J. Amer. Statist. Assoc.* **105** 670–682. MR2724851

PANARETOS, V. M. and TAVAKOLI, S. (2013). Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series. *Stochastic Process. Appl.* To appear. DOI:10.1016/j.spa.2013.03.015, available at http://www.sciencedirect.com/science/article/pii/S0304414913000793.

PANARETOS, V. M. and TAVAKOLI, S. (2013). Supplement to "Fourier analysis of stationary time series in function space." DOI:10.1214/13-AOS1086SUPP.

PELIGRAD, M. and WU, W. B. (2010). Central limit theorem for Fourier transforms of stationary processes. *Ann. Probab.* **38** 2009–2022. MR2722793

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York. MR0762984

PRIESTLEY, M. B. (2001). *Spectral Analysis and Time Series*, *Vol. I and II*. Academic Press, San Diego.

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. MR2168993

RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53** 233–243. MR1094283

ROSENBLATT, M. (1984). Asymptotic normality, strong mixing and spectral density estimates. *Ann. Probab.* **12** 1167–1180. MR0757774

ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser, Boston, MA. MR0885090

SEN, R. and KLÜPPELBERG, C. (2010). Time series of functional data. Unpublished manuscript. Available at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.185.2739.

SHAO, X. and WU, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35** 1773–1801. MR2351105

WAND, M. P. and JONES, M. C. (1995). *Kernel Smoothing. Monographs on Statistics and Applied Probability* **60**. Chapman & Hall, London. MR1319818

WEIDMANN, J. (1980). *Linear Operators in Hilbert Spaces. Graduate Texts in Mathematics* **68**. Springer, New York. MR0566954

WHEEDEN, R. L. and ZYGMUND, A. (1977). *Measure and Integral*: *An Introduction to Real Analysis. Pure and Applied Mathematics* **43**. Dekker, New York. MR0492146

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903. MR2253106

SECTION DE MATHÉMATIQUES
ECOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
1015 LAUSANNE
SWITZERLAND
E-MAIL: victor.panaretos@epfl.ch
       shahin.tavakoli@epfl.ch