

Locally stationary functional time series*

Anne van Delft

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum

Germany

e-mail: anne.vandelft@rub.de

and

Michael Eichler

Department of Quantitative Economics

Maastricht University

P.O. Box 616, 6200 MD Maastricht

The Netherlands

e-mail: m.eichler@maastrichtuniversity.nl

Abstract: The literature on time series of functional data has focused on processes of which the probabilistic law is either constant over time or constant up to its second-order structure. Especially for long stretches of data it is desirable to be able to weaken this assumption. This paper introduces a framework that will enable meaningful statistical inference of functional data of which the dynamics change over time. We put forward the concept of local stationarity in the functional setting and establish a class of processes that have a functional time-varying spectral representation. Subsequently, we derive conditions that allow for fundamental results from nonstationary multivariate time series to carry over to the function space. In particular, time-varying functional ARMA processes are investigated and shown to be functional locally stationary according to the proposed definition. As a side-result, we establish a Cramér representation for an important class of weakly stationary functional processes. Important in our context is the notion of a time-varying spectral density operator of which the properties are studied and uniqueness is derived. Finally, we provide a consistent nonparametric estimator of this operator and show it is asymptotically Gaussian using a weaker tightness criterion than what is usually deemed necessary.

MSC 2010 subject classifications: Primary 62M10; secondary 62M15.

Keywords and phrases: Functional data analysis, locally stationary processes, spectral analysis, kernel estimator.

Received May 2017.

*This work has been supported in part by Maastricht University, the contract “Projet d’Actions de Recherche Concertées” No. 12/17-045 of the “Communauté française de Belgique” and by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823, Project A1, C1, A7) of the German Research Foundation (DFG). The authors sincerely thank the Editor and three Referees for their constructive comments that helped to produce an improved revision of the original paper.

1. Introduction

In functional data analysis, the variables of interest take the form of smooth functions that vary randomly between repeated observations or measurements. Thus functional data are represented by random smooth functions $X(\tau)$, $\tau \in D$, defined on a continuum D . Examples of functional data are concentration of fine dust as a function of day time, the growth curve of children as functions of age, or the intensity as a function of wavelength in spectroscopy. Because functional data analysis deals with inherently infinite-dimensional data objects, dimension reduction techniques such as functional principal component analysis (FPCA) have been a focal point in the literature. Fundamental for these methods is the existence of a Karhunen-Loève decomposition of the process (Karhunen, 1947; Loève, 1948). Some noteworthy early contributions are Kleffe (1973); Grenander (1981); Dauxois, Pousse and Romain (1982); Besse and Ramsay (1986). For an introductory overview of the main functional data concepts we refer to Ramsay and Silverman (2005) and Ferraty and Vieu (2006).

Most techniques to analyze functional data are developed under the assumption of independent and identically distributed functional observations and focus on capturing the first- and second-order structure of the process. A variety of functional data is however collected sequentially over time. In such cases, the data can be described by a *functional time series* $\{X_t(\tau)\}_{t \in \mathbb{Z}}$. Since such data mostly show serial dependence, the assumption of *i.i.d.* repetitions is violated. Examples of functional time series in finance are bond yield curves, where each function is the yield of the bond as a function of time to maturity (e.g. Bowsher and Meeks, 2008; Hays, Shen and Huang, 2012) or the implied volatility surface of a European call option as a function of moneyness and time to maturity. In demography, mortality and fertility rates are given as a function of age (e.g. Erbas, Hyndman and Gertig, 2007; Hyndman and Ullah, 2007; Hyndman and Booth, 2008), while in geophysical sciences, magnetometers record the strength and direction of the magnetic field every five seconds. Due to the wide range of applications, functional time series and the development of techniques that allow to relax the *i.i.d.* assumption have received an increased interest in recent years.

The literature on functional time series has mainly centered around stationary linear models (Mas, 2000; Bosq, 2002; Dehling and Sharipov, 2005) and prediction methods (Antoniadis, Paparoditis and Sapatinas, 2006; Bosq and Blanke, 2007; Aue, Dubart Norinho and Hörmann, 2015). A general framework to investigate the effect of temporal dependence among functional observations on existing techniques has been provided by Hörmann and Kokoszka (2010), who introduce L_m^p approximability as a moment-based notion of dependence.

Violation of the assumption of identically distributed observations has been examined in the setting of change-point detection (e.g. Berkes et al., 2009; Hörmann and Kokoszka, 2010; Aue et al., 2009; Horváth, Hušková and Kokoszka, 2010; Gabrys, Hörmann and Kokoszka, 2010), in the context of functional regression by Yao, Müller and Wang (2005); Cardot and Sarda (2006) and in the context of common principal component models by Benko, Härdle and Kneip (2009).

Despite the growing literature on functional time series, the existing theory has so far been limited to strongly or weakly stationary processes. With the possibility to record, store and analyze functional time series of an increasing length, the common assumption of (weak) stationarity becomes more and more implausible. For instance, in meteorology the distribution of the daily records of temperature, precipitation and cloud cover for a region, viewed as three related functional surfaces, may change over time due to global climate changes. In the financial industry, implied volatility of an option as a function of moneyness changes over time. Other relevant examples appear in the study of cognitive functions such as high-resolution recordings from local field potentials, EEG and MEG. It is widely known that these type of data have a time-varying spectral structure and their statistical treatment requires to take this into account. While heuristic approaches such as localized estimation are readily implemented and applied, a statistical theory for inference from nonstationary functional time series is yet to be developed.

The objective of the current paper is to develop a framework for inference of nonstationary functional time series that allows the derivation of large sample approximations for estimators and test statistics. For this, we extend the concept of locally stationary processes (Dahlhaus, 1996a) to the functional time series setting. We show that fundamental results for multivariate time series can be carried over to the function space, which is a nontrivial task. Our work, which provides a basis for inference of nonstationary functional time series, focuses on frequency domain-based methods and therefore also builds upon the work by Panaretos and Tavakoli (2013a,b). Functional data carry infinite-dimensional intrinsic variation and in order to exploit this rich source of information, it is important to optimally extract defining characteristics to finite dimension via techniques such as functional PCA (FPCA). In the case of stationary dependent functional data, the shape and smoothness properties of the random curves are completely encoded by the spectral density operator, which has been shown to allow for an optimal lower dimension representation via dynamic FPCA (see e.g., Panaretos and Tavakoli, 2013a; Hörmann, Kidziński and Hallin, 2015). Since the assumption of weak stationarity is often too restrictive, we aim to provide the building blocks for statistical inference of nonstationary functional time series and for the development of techniques such as time-varying dynamic FPCA. In particular, our framework will be essential for the development of optimal dimension reduction techniques via a *local* functional Cramér-Karhunen-Loève representation. Such a representation must not only take into account the between- and within curve dynamics but also that these are time-varying. Moreover, the frequency domain arises quite naturally in certain applications such as brain data imaging and is moreover very useful in nonparametric specifications. For example, Aue and van Delft (2017) use our framework to derive a test for stationarity of functional time series against nonstationary alternatives with slowly changing dynamics.

The paper is structured as follows. In section 2, we first introduce some basic notation and methodology for functional data and relate this in a heuristic manner to the concept of locally stationary time series and introduce the definition

of a locally stationary functional time series. In section 3, we demonstrate that time-varying functional ARMA models have a causal solution and are functionally locally stationary according to the definition in section 2. This hinges on the existence of stochastic integrals for operators that belong to a particular Bochner space. In section 4, the time-varying spectral density operator is defined and its properties are derived. In particular, we will show uniqueness of the time-varying spectral density operator. In section 5, we derive the distributional properties of a local nonparametric estimator of the time-varying spectral density operator and deduce a central limit theorem. The results are illustrated by application to a simulated functional autoregressive process in section 6. Technical details and several auxiliary results that are of independent interest are proved in the Appendix.

2. Locally stationary functional time series

Let $X = \{X_t\}_{t=1, \dots, T}$ be a stochastic process taking values in the Hilbert space $H = L^2([0, 1])$ of all real-valued functions that are square integrable with respect to the Lebesgue measure. While current theory for such processes is limited to the case where $\{X_t\}$ is either strictly or weakly stationary, we consider nonstationary processes with dynamics that vary slowly over time and thus can be considered as approximately stationary at a local level.

As an example, consider the functional autoregressive process X given by

$$X_t(\tau) = B_t(X_{t-1})(\tau) + \varepsilon_t(\tau), \quad \tau \in [0, 1],$$

for $t = 1, \dots, T$, where the errors ε_t are independent and identically distributed random elements in H and B_t for $t = 1, \dots, T$ are bounded operators on H . Assuming that the autoregressive operators B_t change only slowly over time, we can still obtain estimates by treating the process as stationary over short time periods. However, since this stationary approximation deteriorates over longer time periods, standard asymptotics based on an increasing sample size T do not provide suitable distributional approximations for the finite sample estimators. Instead we follow the approach by Dahlhaus (1996a, 1997) and define local stationary processes in a functional setting based on an infill asymptotics. The main idea of this approach is that for increasing T the operator B_t is still ‘observed’ on the same interval but on a finer grid, resulting in more and more observations in the time period over which the process can be considered as approximately stationary. Thus we consider a family of functional processes

$$X_{t,T}(\tau) = B_{t/T}(X_{t-1,T})(\tau) + \varepsilon_t(\tau), \quad \tau \in [0, 1], \quad 1 \leq t \leq T,$$

indexed by $T \in \mathbb{N}$ that all depend on the common operators B_u indexed by rescaled time $u = t/T$. Consequently, we in fact examine a triangular array of random functions that share common dynamics as provided by the continuous operator-valued function B_u , $u \in [0, 1]$. For each T , a different ‘level’ of the sequence is thus considered where the dynamics change more slowly for increasing

values of T . We will establish a class of functional time series with a time-varying functional spectral representation that includes interesting processes such as the above example and higher order time-varying functional ARMA models. The framework as provided in this paper will allow to investigate how nonstationarity affects existing methods, such as (dynamic) FPCA, and how these methods should be adjusted in order to be robust for changing characteristics. Similarly as Dahlhaus and Subba Rao (2006) and Vogt (2012) in the case of ordinary time series, we call a functional time series locally stationary if it can be locally approximated by a stationary functional time series. In the following definition, $\|\cdot\|_2$ denotes the L^2 -norm of H .

Definition 2.1 (Local stationarity). A sequence of stochastic processes $\{X_{t,T}\}_{t \in \mathbb{Z}}$ indexed by $T \in \mathbb{N}$ and taking values in H is called locally stationary if for all rescaled times $u \in [0, 1]$ there exists an H -valued strictly stationary process $\{X_t^{(u)}\}_{t \in \mathbb{Z}}$, such that

$$\|X_{t,T} - X_t^{(u)}\|_2 \leq (|\frac{t}{T} - u| + \frac{1}{T}) P_{t,T}^{(u)} \quad a.s.$$

for all $1 \leq t \leq T$, where $P_{t,T}^{(u)}$ is a positive real-valued process such that for some $\rho > 0$ and $C < \infty$ the process satisfies $\mathbb{E}(|P_{t,T}^{(u)}|^\rho) < C$ for all t and T and uniformly in $u \in [0, 1]$.

For the purpose of illustration, a very simple locally stationary functional time series is depicted in figure 1 (A). Note that visual interpretation of a functional time series can be extremely difficult, especially when it is driven by many interacting components. The process in figure 1 (A) is driven solely by two components and is generated as

$$X_{t,T}(\tau) = \xi_{t,T} \phi_1(\tau) + \chi_{t,T} \phi_2(\tau) \quad \tau \in [0, 1] \quad (1)$$

where ϕ_1, ϕ_2 are basis functions of H and the random coefficients $\xi_{t,T}, \chi_{t,T}$ are independent Gaussian time-varying AR(2). The parameters of the time-varying AR(2) models are chosen in such a way that the magnitude and phase of the roots of the characteristic polynomials of $\xi_{t,T}$ and $\chi_{t,T}$ vary cyclically with rescaled time t/T but in opposite direction as time progresses. The corresponding coefficient curves are plotted in figure 1 (C) and 1 (D), respectively. The dependence structure of the two driving components vary from independence to close to unit root behavior and this varying cyclical behavior is also clearly visible in the resulting functional process $\{X_{t,T}\}$. In order to contrast this with behavior observed under stationarity, figure 1 (B) depicts the closely related weakly stationary functional process of (1) where the random coefficients are generated using two stationary AR(2) with parameters specified as the time average of $\xi_{t,T}$ and $\chi_{t,T}$, respectively. A comparison of the plots for the locally stationary and the stationary case shows a clear difference in the dynamics of the two processes, which is particularly discernible in the projections on the Fourier components. The example thus indicates the effect of falsely misspecifying a locally stationary process as stationary on statistical inference.

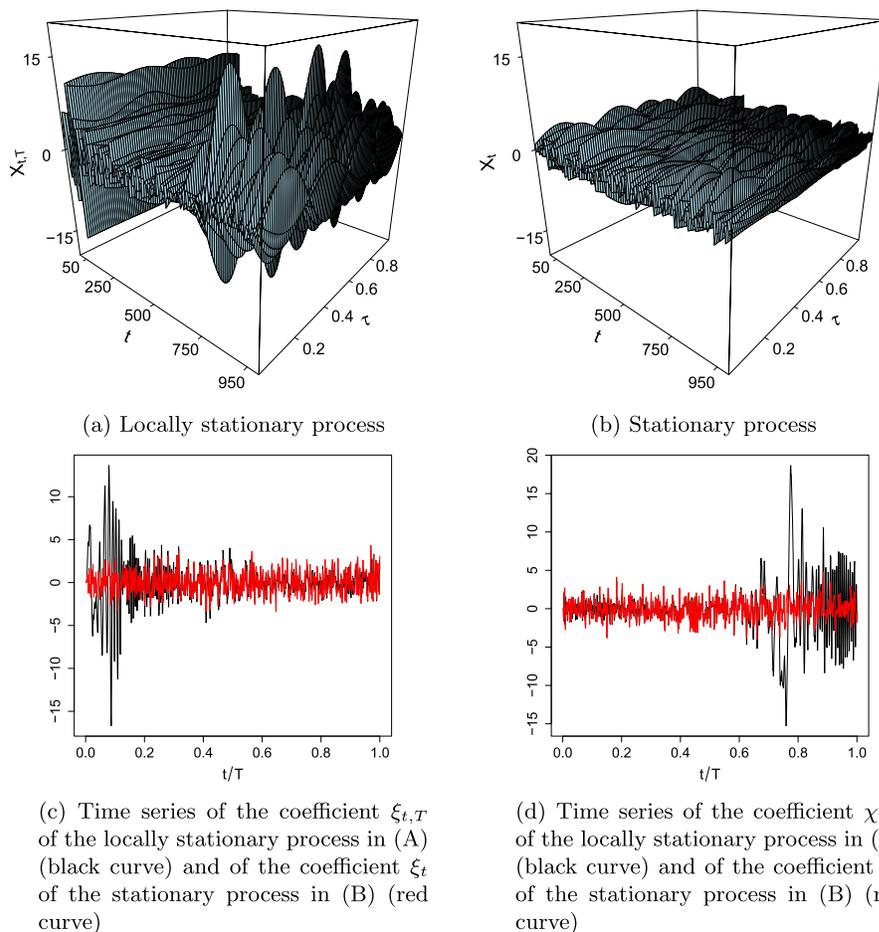


Fig 1: Comparison of stationary and local stationary functional time series

Definition 2.1 is broad and is further investigated in Aue and van Delft (2017). It will allow for the development of statistical inference procedures for nonstationary functional time series and in particular encompasses nonlinear functional models. Nonlinear functional time series is a topic that is relatively unexplored. Possible relevant models that are worth investigating are, for instance, time-varying additive functional regression (Müller and Yao, 2008) and time-varying functional ARCH models (Hörmann, Horváth and Reeder, 2013). However, as the focus of this paper is on frequency domain based methods, it is more appropriate to work with an alternative characterization of local stationarity in terms of spectral representations, which we discuss below. We start by introducing the necessary terminology on operators and spectral representations for stationary functional time series.

2.1. Functional spaces and operators: Notation and terminology

First, we introduce some basic notation and definitions on functional spaces and operators. Let (T, \mathcal{B}) be a measurable space with σ -finite measure μ . Furthermore, let E be a Banach space with norm $\|\cdot\|_E$ and equipped with the Borel σ -algebra. We then define $L_E^p(T, \mu)$ as the Banach space of all strongly measurable functions $f : T \rightarrow E$ with finite norm

$$\|f\|_p = \|f\|_{L_E^p(T, \mu)} = \left(\int \|f(\tau)\|_E^p d\mu(\tau) \right)^{\frac{1}{p}}$$

for $1 \leq p < \infty$ and with finite norm

$$\|f\|_\infty = \|f\|_{L_E^\infty(T, \mu)} = \inf_{\mu(N)=0} \sup_{\tau \in T \setminus N} \|f(\tau)\|_E$$

for $p = \infty$. We note that two functions f and g are equal in L^p , denoted as $f \stackrel{L^p}{=} g$, if $\|f - g\|_p = 0$. If E is a Hilbert space with inner product $\langle \cdot, \cdot \rangle_E$ then $L_E^2(T, \mu)$ is also a Hilbert space with inner product

$$\langle f, g \rangle = \langle f, g \rangle_{L_E^2(T, \mu)} = \int \langle f(\tau), g(\tau) \rangle_E d\mu(\tau).$$

For notational convenience, we use the shorter notation $\|f\|_p$ and $\langle f, g \rangle$ whenever no ambiguity about the space $L_E^p(T, \mu)$ is possible. Similarly, if $T \subset \mathbb{R}^k$ and μ is the Lebesgue measure on T , we omit μ and write $L_E^p(T)$, and if $E = \mathbb{R}$ we write $L^p(T, \mu)$.

Next, an operator A on a Hilbert space H is a function $A : H \rightarrow H$. An operator A is said to be compact if the image of each bounded set under A is relatively compact. If H is separable, there exist orthonormal bases $\{\phi_n\}$ and $\{\psi_n\}$ of H and a monotonically decreasing sequence of non-negative numbers $s_n(A)$, $n \in \mathbb{N}$ converging to zero, such that

$$A f = \sum_{n=1}^{\infty} s_n(A) \langle f, \psi_n \rangle \phi_n \quad (2)$$

for all $f \in H$. The values $s_n(A)$ are called the *singular values* of A and (2) is the *singular value decomposition* of A . For operators on H , we denote the *Schatten p -class* by $S_p(H)$ and its norm by $\|\cdot\|_p$. More specifically, for $p = \infty$, the space $S_\infty(H)$ indicates the space of bounded linear operators equipped with the standard operator norm, while for $1 \leq p < \infty$ the *Schatten p -class* is the subspace of all compact operators A on H such that the sequence $s(A) = (s_n(A))_{n \in \mathbb{N}}$ of singular values of A belongs to ℓ^p ; the corresponding norm is given by $\|A\|_p = \|s(A)\|_p$. For $1 \leq p \leq q \leq \infty$, we have the inclusion $S_p(H) \subseteq S_q(H)$. Two important classes are the trace-class and the Hilbert-Schmidt operators on H , which are given by $S_1(H)$ and $S_2(H)$, respectively. More properties of Schatten-class operators and in particular of Hilbert-Schmidt operators are provided in Appendix B1. Finally, the adjoint of A is denoted by A^\dagger while the

identity and zero operator are given by I_H and O_H , respectively. As usual, the complex conjugate of $z \in \mathbb{C}$ is denoted by \bar{z} and the imaginary number by i .

The main object of this paper are functional time series $X = \{X_t\}$ that take values in the Hilbert space $H = L^2([0, 1])$. More precisely, for some underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\mathbb{H} = L^2_H(\Omega, \mathbb{P})$ be the Hilbert space of all H -valued random variables X with finite second moment $\mathbb{E}\|X\|_2^2 < \infty$. To avoid ambiguities between the norms of H and \mathbb{H} , we write $\|X\|_{\mathbb{H}}$ for the norm in \mathbb{H} and reserve the notation $\|X\|_2$ for the more frequently used norm in H . Throughout the paper, we assume that $X_t \in \mathbb{H}$. For the spectral representation and Fourier analysis of functional time series $\{X_t\}$, we also require the corresponding spaces $H_{\mathbb{C}} = L^2_{\mathbb{C}}([0, 1])$ and $\mathbb{H}_{\mathbb{C}} = L^2_{H_{\mathbb{C}}}(\Omega, \mathbb{P})$. We recall some basic properties of functional time series. First, a functional time series X is called strictly stationary if, for all finite sets of indices $J \subset \mathbb{Z}$, the joint distribution of $\{X_{t+j} \mid j \in J\}$ does not depend on $t \in \mathbb{Z}$. Similarly, X is weakly stationary if its first- and second-order moments exist and are invariant under translation in time. In that case, the mean function m of X is defined as the unique element of H such that

$$\langle m, g \rangle = \mathbb{E}\langle X_t, g \rangle, \quad g \in H.$$

Furthermore, the h -th lag covariance operator \mathcal{C}_h is given by

$$\langle \mathcal{C}_h g_1, g_2 \rangle = \mathbb{E}[\langle g_1, X_0 - m \rangle \langle X_h - m, g_2 \rangle], \quad g_1, g_2 \in H,$$

and belongs to $S_2(H)$. Since $S_2(H)$ is isomorphic to the tensor product, we call \mathcal{C}_h also *autocovariance tensor*. The covariance operator \mathcal{C}_h can alternatively be described by its kernel function c_h satisfying

$$\langle \mathcal{C}_h g_1, g_2 \rangle = \int_0^1 \int_0^1 c_h(\tau, \sigma) g_1(\sigma) g_2(\tau) d\sigma d\tau, \quad g_1, g_2 \in H.$$

In analogy to weakly stationary multivariate time series, where the covariance matrix and spectral density matrix form a Fourier pair, the *spectral density operator* or *tensor* \mathcal{F}_{ω} is given by the Fourier transform of \mathcal{C}_h ,

$$\mathcal{F}_{\omega} = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \mathcal{C}_h e^{-i\omega h}. \quad (3)$$

A sufficient condition for the existence of \mathcal{F}_{ω} in $S_p(H_{\mathbb{C}})$ is $\sum_{h \in \mathbb{Z}} \|\mathcal{C}_h\|_p < \infty$. Since the setting of this paper allows for higher order dependence among the functional observations, we also require the notion of higher order cumulant tensors. The necessary derivations and definitions are given in Appendix B2. Throughout the remainder of this paper, time points in $\{1, \dots, T\}$ will be denoted by t, s or r , while rescaled time points on the interval $[0, 1]$ will be given by u and v . Additionally, angular frequencies are indicated with λ, α, β or ω and functional arguments are denoted by τ, σ .

Finally, we require the notion of stochastic integrals with respect to operator-valued functions. To this end, let \mathcal{B}_{∞} denote the Bochner space $L^2_{S_{\infty}(H_{\mathbb{C}})}([-\pi, \pi])$,

μ) of all strongly measurable functions $U : [-\pi, \pi] \rightarrow S_\infty(H_{\mathbb{C}})$ such that

$$\|U\|_{\mathcal{B}_\infty}^2 = \int_{-\pi}^{\pi} \|U_\omega\|_\infty^2 d\mu(\omega) < \infty,$$

where μ is a measure on the interval $[-\pi, \pi]$ given by $\mu(A) = \int_A \|\mathcal{F}_\omega\|_1 d\omega$ for all Borel sets $A \subseteq [-\pi, \pi]$. The subspace \mathcal{B}_2 is then defined similarly with Hilbert-Schmidt norms replacing the operator norms. We distinguish explicitly between the two spaces as the latter space allows for stronger results to be obtained but excludes interesting processes such as functional autoregressive processes.

2.2. Assumptions

In this section, we collect for better reference the assumptions required in subsequent sections. We start by the main assumptions needed for a frequency domain characterization of local stationarity. In contrast to Panaretos and Tavakoli (2013a), who only consider transfer functions in \mathcal{B}_2 , we also prove the more general case of transfer functions in \mathcal{B}_∞ as it includes the important case of functional autoregressive processes. The necessary results are proved in section B2.3 of the Appendix. Throughout the assumptions and the paper, we refer to $S_p(H_{\mathbb{C}})$ and \mathcal{B}_p with $p = 2$ or $p = \infty$ to make the distinction between the two cases.

- (A1) (i) $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is a weakly stationary white noise process taking values in H with spectral representation $\varepsilon_t = \int_{-\pi}^{\pi} e^{i\omega t} dZ_\omega$, where Z_ω is a 2π -periodic orthogonal increment process taking values in $H_{\mathbb{C}}$;
 (ii) the functional process $X_{t,T}$ with $t = 1, \dots, T$ and $T \in \mathbb{N}$ is given by

$$X_{t,T} = \int_{-\pi}^{\pi} e^{i\omega t} \mathcal{A}_{t,\omega}^{(T)} dZ_\omega \quad \text{a.e. in } \mathbb{H}$$

with transfer operator $\mathcal{A}_{t,\omega}^{(T)} \in \mathcal{B}_p$ and an orthogonal increment process Z_ω .

- (A2) There exists $\mathcal{A} : [0, 1] \times [-\pi, \pi] \rightarrow S_p(H_{\mathbb{C}})$ with $\mathcal{A}_{u,\cdot} \in \mathcal{B}_p$ and $\mathcal{A}_{u,\omega}$ being continuous in u such that for all $T \in \mathbb{N}$

$$\sup_{\omega, t} \|\mathcal{A}_{t,\omega}^{(T)} - \mathcal{A}_{\frac{t}{T},\omega}\|_p = O\left(\frac{1}{T}\right).$$

- (A3) The function $\mathcal{A}_{\cdot,\cdot}$ is Hölder continuous of order $\alpha > 1/2$ in u and ω .
 (A4) The function $\mathcal{A}_{\cdot,\cdot}$ is twice continuously differentiable in u and ω with second derivatives being uniformly bounded in u and ω .

We note that a functional Cramér representation such as in (A1)(ii) can also be obtained when the spectral density operator is not well-defined; we refer the reader to van Delft and Eichler (2018), in which a functional version of Herglotz's theorem is proved and frequency domain representations for stationary time series on the function space are further generalized. For the derivation of asymptotic results for kernel estimators of the spectral density operator, we require additional assumptions on the moments of k -order of the process ε_t in (A1). The following assumption will be imposed for $k \leq 4$ or $k < \infty$.

(A5) The process $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ satisfies $\mathbb{E}\|\varepsilon_0\|_2^k < \infty$ and, with $\mathcal{C}_{t_1, \dots, t_{k-1}}^\varepsilon$ denoting the k -th order cumulant tensor of ε , $\sum_{t_1, \dots, t_{k-1} = -\infty}^{\infty} \|\mathcal{C}_{t_1, \dots, t_{k-1}}^\varepsilon\|_2 < \infty$.

Finally, the following assumptions formulate the conditions imposed on the taper functions, the kernel functions, and the bandwidths used for kernel smoothing.

(A6) The function $h : \mathbb{R} \rightarrow \mathbb{R}^+$ is symmetric with compact support on $[0, 1]$ and is of bounded variation.

(A7) The function $K_f : \mathbb{R} \rightarrow \mathbb{R}^+$ is symmetric, has bounded variation and compact support $[-1, 1]$, and satisfies

$$\begin{aligned} \text{(i)} \quad & \int_{\mathbb{R}} K_f(\omega) d\omega = 1; \\ \text{(ii)} \quad & \int_{\mathbb{R}} \omega K_f(\omega) d\omega = 0. \end{aligned}$$

(A8) The sequences $b_{f,T}$ and $b_{t,T}$ satisfy for $T \rightarrow \infty$ (i) $b_{f,T} \rightarrow 0$ and $b_{t,T} \rightarrow 0$; (ii) $b_{f,T} b_{t,T} T \rightarrow \infty$; (iii) $b_{f,T} \log(b_{t,T} T) \rightarrow 0$; and (iv) $b_{t,T}^2 b_{f,T} \rightarrow 0$.

2.3. Local stationarity in the frequency domain

The original definition of local stationary processes by Dahlhaus (1996a) has been formulated in the frequency domain. The following proposition can be viewed a generalization of Dahlhaus (1996a) to the functional setting.

Proposition 2.2. *Suppose that assumptions (A1) and (A2) hold. Then $\{X_{t,T}\}$ is a locally stationary process in H .*

Proof. For $u \in [0, 1]$, we define the approximating stationary functional process $\{X_t^{(u)}\}_{t \in \mathbb{Z}}$ by

$$X_t^{(u)} = \int_{-\pi}^{\pi} e^{i\omega t} \mathcal{A}_{u,\omega} dZ_\omega.$$

Then we have

$$\|X_{t,T} - X_t^{(u)}\|_2 = \left\| \int_{-\pi}^{\pi} e^{i\omega t} (\mathcal{A}_{t,\omega}^{(T)} - \mathcal{A}_{u,\omega}) dZ_\omega \right\|_2 \leq c P_{t,T}^{(u)}$$

with

$$\begin{aligned} c &= \sup_{\omega} \|\mathcal{A}_{t,\omega}^{(T)} - \mathcal{A}_{u,\omega}\|_{\infty} \\ &\leq \sup_{\omega} \|\mathcal{A}_{t,\omega}^{(T)} - \mathcal{A}_{t/T,\omega}\|_{\infty} + \sup_{\omega} \|\mathcal{A}_{t/T,\omega} - \mathcal{A}_{u,\omega}\|_{\infty} = O\left(\frac{t}{T} + \left|\frac{t}{T} - u\right|\right) \end{aligned}$$

and

$$P_{t,T}^{(u)} = \frac{1}{c} \left\| \int_{-\pi}^{\pi} e^{i\omega t} (\mathcal{A}_{t,\omega}^{(T)} - \mathcal{A}_{u,\omega}) dZ_\omega \right\|_2.$$

Since

$$\mathbb{E}|P_{t,T}^{(u)}|^2 \leq \frac{1}{c^2} \int_{-\pi}^{\pi} \|\mathcal{A}_{t,\omega}^{(T)} - \mathcal{A}_{u,\omega}\|_{\infty}^2 \|\mathcal{F}_\omega\|_1 d\omega \leq \int_{-\pi}^{\pi} \|\mathcal{F}_\omega\|_1 d\omega,$$

the process satisfies the conditions of Definition 2.1 with $\rho = 2$. \square

As in the time series setting, we need the existence of a *transfer operator* $\mathcal{A}_{u,\omega}$ that is continuous in $u \in [0, 1]$ to guarantee locally an approximately stationary behavior without sudden changes. In order to include interesting cases such as autoregressive processes for which a time-varying functional spectral representation with a common continuous transfer operator $\mathcal{A}_{u,\omega}$ does not exist, we require that such a representation only holds approximately by condition (A2).

The previous result leads us to consider time-varying processes of the form

$$X_{t,T} = \sum_{s \in \mathbb{Z}} A_{t,s}^{(T)} \varepsilon_{t-s}, \quad (4)$$

where $\{\varepsilon_s\}_{s \in \mathbb{Z}}$ is a weakly stationary functional white noise process in H and $\{A_{t,s}^{(T)}\}_{s \in \mathbb{Z}}$ are sequences of linear operators for $t = 1, \dots, T$ and $T \in \mathbb{N}$. The following result states the conditions under which such a process satisfies condition (A1).

Proposition 2.3. *Suppose that $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ satisfies assumption (A1) and, for $p = 2$ or $p = \infty$, let $\{A_{t,s}^{(T)}\}_{s \in \mathbb{Z}}$ be a sequence of operators in $S_p(H)$ satisfying $\sum_s \|A_{t,s}^{(T)}\|_p < \infty$ for all $t = 1, \dots, T$ and $T \in \mathbb{N}$. Then the process*

$$X_{t,T} = \sum_{s \in \mathbb{Z}} A_{t,s}^{(T)} \varepsilon_{t-s} \quad (5)$$

satisfies assumption (A1) with $\mathcal{A}_{t,\omega}^{(T)} \in \mathcal{B}_p$.

The proof is relegated to section A1.1 of the Appendix. For $p = 2$, the proposition yields a time-varying version of the corresponding result of Panaretos and Tavakoli (2013a). The more general case $p = \infty$ also includes linear models introduced by Bosq (2000) and Hörmann and Kokoszka (2010) as well as the important class of time-varying functional autoregressive processes, which we discuss in detail in the next section.

Remark 2.4. For processes of the form (4), we can alternatively verify Definition 2.1 in the time domain provided we impose some regularity conditions on the decay of the sequence of filter operators $\{A_{t,s}^{(T)}\}_{s \in \mathbb{Z}}$. For example, sufficient conditions for (A1)–(A2) to be satisfied would be to assume that there exists a positive monotonically decreasing sequence $\{\ell(s)\}_{s \in \mathbb{Z}}$ that satisfies $\sum_{s \in \mathbb{Z}} |s| \ell(s) < \infty$ such that $\sup_{t,T} \|A_{t,s}^{(T)}\|_p < K \ell(s)$ and that there exists a sequence $\{A_{s,u}\}_{s \in \mathbb{N}}$ that satisfies

$$\sup_{t,T} \|A_{t,s}^{(T)} - A_{s,u}\|_p \leq \frac{K \ell(s)}{T},$$

for some constant K independent of T . The local asymptotic theory derived later in this paper relies however on additional smoothness conditions of the approximate transfer operators such as condition (A4). These could then be replaced by $\sup_u \|\frac{\partial^2}{\partial u^2} A_{s,u}\|_p < K \ell(s)$, where $\frac{\partial^2}{\partial u^2} A_{s,u}$ denotes the second-order derivative of the function $u \mapsto A_{s,u}$. Depending on the application, different conditions could be considered. The investigation of necessary restrictions on

the time domain filter operators are beyond the scope of this paper and are left for future work.

3. Locally stationary functional autoregressive processes

Due to its flexibility as well as its simplicity, functional autoregressive processes have been found useful in numerous applications such as economics and medicine, especially for prediction purposes (see e.g., Damon and Guillas, 1982; Besse and Ramsay, 1986; Antoniadis and Sapatinas, 2003, for early work). Despite of being linear in the function space, the filter operators act on a Hilbert space of which the elements can still exhibit arbitrary degrees of nonlinearity and can therefore be seen to be highly nonlinear in terms of scalar records. Most estimation techniques are however still based on the assumption of *i.i.d.* functional errors. This assumption has been relaxed by Bosq (2000), where the assumption of independence of the errors of the causal solution is relaxed to uncorrelatedness in an appropriate sense, and by Hörmann and Kokoszka (2010) for functional AR(1) processes within the framework of L^p - m -approximability.

In this section, we introduce a class of time-varying functional autoregressive processes for which inference and forecasting methods can be developed in a meaningful way. More specifically, we will show that time-varying functional autoregressive processes as well as the more general time-varying functional ARMA processes are locally stationary and that stationary functional ARMA(m, n) processes are a special case. For this we first need to establish that a causal solution exists for time-varying functional AR(m) processes. This is done in the theorem stated below.

Theorem 3.1. *Let $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ be a white noise process in H and let $\{X_{t,T}\}$ be a sequence of time-varying functional AR(m) given by*

$$X_{t,T} = \sum_{j=1}^m B_{\frac{t}{T},j}(X_{t-j,T}) + \varepsilon_t \quad (6)$$

with $B_{u,j} = B_{0,j}$ for $u < 0$ and $B_{u,j} = B_{1,j}$ for $u > 1$. Furthermore, suppose that

- (i) the operators $B_{u,j}$ are continuous in $u \in [0, 1]$ for all $j = 1, \dots, m$;
- (ii) for all $u \in [0, 1]$, the operators satisfy $\sum_{j=1}^m \|B_{u,j}\|_\infty < 1$.

Then (6) has a unique causal solution of the form

$$X_{t,T} = \sum_{l=0}^{\infty} A_{t,l}^{(T)}(\varepsilon_{t-l}) \quad (7)$$

for all $t \in \mathbb{N}$ with $\sup_{t,T} \sum_{l=0}^{\infty} \|A_{t,l}^{(T)}\|_\infty < \infty$.

In order to prove the theorem, note that we can represent the functional

AR(m) process in state space form

$$\underbrace{\begin{pmatrix} X_{t,T} \\ X_{t-1,T} \\ \vdots \\ X_{t-m+1,T} \end{pmatrix}}_{\mathbf{X}_{t,T}^*} = \underbrace{\begin{pmatrix} B_{\frac{t}{T},1} & B_{\frac{t}{T},2} & \cdots & B_{\frac{t}{T},m} \\ I_H & & & O_H \\ & \ddots & & \vdots \\ & & I_H & O_H \end{pmatrix}}_{\mathbf{B}_{\frac{t}{T}}^*} \underbrace{\begin{pmatrix} X_{t-1,T} \\ X_{t-2,T} \\ \vdots \\ X_{t-m,T} \end{pmatrix}}_{\mathbf{X}_{t-1,T}^*} + \underbrace{\begin{pmatrix} \varepsilon_t \\ O_H \\ \vdots \\ O_H \end{pmatrix}}_{\boldsymbol{\varepsilon}_t^*}. \quad (8)$$

Here, $\mathbf{X}_{t,T}^*$ is a m -dimensional random vector taking values in the Hilbert space H^m with inner product $\langle x, y \rangle = \sum_{i=1}^m \langle x_i, y_i \rangle_H$. Furthermore, \mathbf{B}_u^* denotes a matrix of operators and thus is itself an operator on H^m . Consequently, we can write the functional AR(m) process more compactly as

$$\mathbf{X}_{t,T}^* = \mathbf{B}_{\frac{t}{T}}^* (\mathbf{X}_{t-1,T}^*) + \boldsymbol{\varepsilon}_t^*$$

with $\boldsymbol{\varepsilon}_t^* \in L_{H^m}^2(\Omega, \mathbb{P})$.

Proof of Theorem 3.1. In order to show that a causal solution exists in the locally stationary setting, we require the following result which is proved in Appendix A1.2.

Lemma 3.2. *For $u \in [0, 1]$, the assumption $\sum_{j=1}^m \|B_{u,j}\|_\infty < 1$ implies that the operator \mathbf{B}_u^* satisfies $\|\mathbf{B}_u^{*k_o}\|_\infty < 1$ for some $k_o \geq 1, k_o \in \mathbb{Z}$.*

We note that this is a weaker assumption than $\|\mathbf{B}_u^*\|_\infty < 1$. Although $\|\mathbf{B}_u^{*k_o}\|_\infty < 1$ is usually stated as the condition for a causal solution in the stationary case, the condition $\sum_{j=1}^m \|B_{u,j}\|_\infty < 1$ is easier to check in practice. As a consequence of this lemma, the assumption $\sum_{j=1}^m \|B_{u,j}\|_\infty < 1$ for all $u \in [0, 1]$ implies that the spectral radius of \mathbf{B}_u^* satisfies

$$r(\mathbf{B}_u^*) = \sup_{\lambda \in S_u} |\lambda| = \lim_{k \rightarrow \infty} \|\mathbf{B}_u^{*k}\|_\infty^{1/k} < \frac{1}{1 + \delta} \quad (9)$$

for some $\delta > 0$. Observe then that by recursive substitution

$$\mathbf{X}_{t,T}^* = \sum_{l=0}^{\infty} \left(\prod_{s=0}^{l-1} \mathbf{B}_{\frac{t-s}{T}}^* \right) \boldsymbol{\varepsilon}_{t-l}^*.$$

From (8), this implies a solution is given by

$$X_{t,T} = \sum_{l=0}^{\infty} \left[\prod_{s=0}^{l-1} \mathbf{B}_{\frac{t-s}{T}}^* \right]_{1,1} (\varepsilon_{t-l}), \quad (10)$$

where $[\cdot]_{1,1}$ refers to the upper left block element of the corresponding block matrix of operators. In order to prove the theorem we shall proceed in a similar manner as Künsch (1995) and derive that

$$\sup_{t,T} \left\| \left[\prod_{s=0}^{l-1} \mathbf{B}_{\frac{t-s}{T}}^* \right]_{1,1} \right\|_\infty < c\rho^l$$

for some constant c and $\rho < 1$. For this, we require the following lemma.

Lemma 3.3. *Let $B(H)$ be the algebra of bounded linear operators on a Hilbert space. Then for each $A \in B(H)$ and each $\varepsilon > 0$, there exists an invertible element M of $B(H)$ such that $r(A) \leq \|MAM^{-1}\|_\infty \leq r(A) + \varepsilon$.*

Since $B(H)$ forms a unital C^* -algebra, this lemma is a direct consequence of a result in Murphy (1990)[p.74]. Lemma 3.3 together with (9) imply we can specify for fixed u a new operator $M(u) \in B(H)$ such that

$$\|M(u)\mathbf{B}_u^*M^{-1}(u)\|_\infty < \frac{1}{1 + \delta/2}.$$

Because of the continuity of the autoregressive operators in u , we have that for all $u \in [0, 1]$, there exists a neighborhood $\mathcal{V}(u)$ such that

$$\|M(u)\mathbf{B}_v^*M^{-1}(u)\|_\infty < \frac{1}{1 + \delta/3} < 1 \quad \text{for } v \in \mathcal{V}(u).$$

Define now the finite union $\bigcup_{i=1}^r \mathcal{V}(u_i)$ with $\mathcal{V}(u_i) \cap \mathcal{V}(u_l) = \emptyset$ for $i \neq l$. Due to compactness and the fact that $\mathbf{B}_u^* = \mathbf{B}_0^*$ for $u \leq 0$ this union forms a cover of $(-\infty, 1]$. The preceding then implies that there exists a constant c such that

$$\|\mathbf{B}_v^*\|_\infty \leq c\|M(u_i)\mathbf{B}_v^*M^{-1}(u_i)\|_\infty, \quad i = 1, \dots, r.$$

Now, fix t and T and define the set $J_{i,l} = \{s \geq 0 : \frac{t-s}{T} \in \mathcal{V}(u_i)\} \cap \{0, 1, \dots, l-1\}$. Then specify $\rho = \frac{1}{1+\delta/3}$ to obtain

$$\begin{aligned} \left\| \left(\prod_{s=0}^{l-1} \mathbf{B}_{\frac{t-s}{T}}^* \right)_{1,1} \right\|_\infty &\leq \left\| \prod_{s=0}^{l-1} \mathbf{B}_{\frac{t-s}{T}}^* \right\|_\infty \leq \prod_{i=1}^r \left\| \prod_{s \in J_{i,l}} \mathbf{B}_{\frac{t-s}{T}}^* \right\|_\infty \\ &\leq c^m \prod_{i=1}^r \prod_{s \in J_{i,l}} \|M(u_i)\mathbf{B}_{\frac{t-s}{T}}^*M^{-1}(u_i)\|_\infty \\ &\leq c^r \prod_{i=1}^r \rho^{|J_{i,l}|} = c^r \rho^l, \end{aligned}$$

which gives the result. \square

Theorem 3.1 will be used to show that time-varying functional ARMA models for which a functional spectral representation exists satisfy conditions (A1) and (A2) and hence by Proposition 2.2 are locally stationary. Before we can consider general time-varying functional ARMA models we first need the following result, which shows that for time-varying functional autoregressive processes there exists a common continuous transfer operator $\mathcal{A}_{u,\omega}$ that satisfies condition (A2).

Theorem 3.4. *Let $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ be a white noise process in \mathbb{H} and let $\{X_{t,T}\}$ be a sequence of functional autoregressive processes given by*

$$\sum_{j=0}^m B_{\frac{t}{T},j}^*(X_{t-j,T}) = C_{\frac{t}{T}}(\varepsilon_t) \quad (11)$$

with $B_{u,j} = B_{0,j}$, $C_u = C_0$ for $u < 0$ and $B_{u,j} = B_{1,j}$, $C_u = C_1$ for $u > 1$. If the process satisfies, for all $u \in [0, 1]$ and $p = 2$ or $p = \infty$, the conditions

- (i) C_u is an invertible element of $S_\infty(H)$;
 - (ii) $B_{u,j} \in S_p(H)$ for $j = 1, \dots, m$ with $\sum_{j=1}^m \|B_{u,j}\|_l < 1$ and $B_{u,0} = I_H$;
 - (iii) the mappings $u \mapsto B_{u,j}$ for $j = 1, \dots, m$ and $u \mapsto C_u$, are continuous in $u \in [0, 1]$ and differentiable on $u \in (0, 1)$ with bounded derivatives,
- then the process $\{X_{t,T}\}$ satisfies (A2) with

$$\mathcal{A}_{\frac{t}{T}, \omega}^{(T)} = \frac{1}{\sqrt{2\pi}} \left(\sum_{j=0}^m e^{-i\omega j} B_{\frac{t}{T}, j} \right)^{-1} C_{\frac{t}{T}} \quad (12)$$

and thus is locally stationary.

The proof of Theorem 3.4 is relegated to Appendix A1.2. As shown in Theorem 3.1, a sufficient condition for the difference equation (11) to have a causal solution is $\sum_{j=1}^m \|B_{u,j}\|_\infty < 1$ or $\|B_u^{*k_0}\|_\infty < 1$ for some $k_0 \geq 1$. The moving average operators will then satisfy $\sum_{l=0}^\infty \|A_{t,l}^{(T)}\|_\infty < \infty$, and Proposition 2.3 shows that $X_{t,T}$ satisfies assumption (A1) with $\mathcal{A}_{t,\omega}^{(T)} \in \mathcal{B}_\infty$. It follows from (10) that time-varying functional AR(m) processes that have a causal solution with moving average operators satisfying $\sum_{l=0}^\infty \|A_{t,l}^{(T)}\|_2 < \infty$ do not exist. Instead we need at least $A_{t,0}^{(T)}$ to be an invertible element of $S_\infty(H)$ together with $\sum_{j=1}^m \|B_{u,j}\|_2 < 1$. By Proposition B1.6, this case is covered by Proposition 2.2 with $\mathcal{A}_{\frac{t}{T}, \omega} \in S_2(H_{\mathbb{C}})$ in condition (A2). For stationary functional AR(m) processes this is straightforward to verify using back-shift operator notation and by solving for the inverse of the autoregressive lag operator. Under slightly more restrictive assumptions it is possible to obtain uniform convergence results for processes with transfer operators $\mathcal{A}_{t,\omega}^{(T)} \in \mathcal{B}_2$. We will come back to this in sections 4 and 5, in which we consider capturing the changing second-order dependence structure via the time-varying spectral density operator.

Using Theorem 3.4, it is now straightforward to establish that the time-varying functional ARMA processes are locally stationary in the sense of Proposition 2.2. A time-varying functional moving average process of order n has transfer operator

$$\mathcal{A}_{t,\omega}^{(T)} = \frac{1}{\sqrt{2\pi}} \sum_{j=0}^n \Phi_{\frac{t}{T}, j} e^{-i\omega j},$$

where $\Phi_{t/T, j} \in S_p(H)$ are the moving average filter operators. This follows from the spectral representation of the ε_t . Setting $\mathcal{A}_{\frac{t}{T}, \omega} = \mathcal{A}_{t,\omega}^{(T)}$ gives the result. Finally, we can combine this with Theorem 3.4, to obtain that conditions (A1) and (A2) are satisfied for time-varying functional ARMA(m, n) with common continuous transfer operator given by

$$\mathcal{A}_{\frac{t}{T}, \omega} = \frac{1}{\sqrt{2\pi}} C_{\frac{t}{T}} \left(\sum_{j=0}^m e^{-i\omega j} B_{\frac{t}{T}, j} \right)^{-1} \sum_{l=0}^n \Phi_{\frac{t}{T}, l} e^{-i\omega l}. \quad (13)$$

For operators that do not depend on t , this result proves the existence of a well-defined functional Cramér representation for weakly stationary functional ARMA(m, n) processes as discussed in Bosq (2000) or as in Hörmann and Kokoszka (2010). The latter is easily seen by means of an application of the dominated convergence theorem and by defining the m -dependent coupling process by

$$X_{t,T}^{(m)} = g_{t,T}(\varepsilon_t, \dots, \varepsilon_{t-m+1}, \varepsilon_{t-m}^*, \varepsilon_{t-m-1}^{(*)}, \dots)$$

for measurable functions $g_{t,T} : H^\infty \rightarrow H$ with $t = 1, \dots, T$ and $T \in \mathbb{N}$ and where $\{\varepsilon_t^*\}$ is an independent copy of $\{\varepsilon_t\}$.

4. Time-varying spectral density operator

We will now introduce the time-varying spectral density operator and its properties. We will show that the uniqueness property of the time-varying spectral density established by Dahlhaus (1996a) also extends to the infinite dimension. Let $X_{t,T}$ satisfy conditions (A1) and (A2) with $\mathcal{A}_{t,\omega}^{(T)} = \mathcal{A}_{1,\omega}^{(T)}$ for $t < 1$ and $\mathcal{A}_{t,\omega}^{(T)} = \mathcal{A}_{T,\omega}^{(T)}$ for $t > T$. We define the *local autocovariance operator* as the cumulant tensor

$$\mathcal{C}_{u,s}^{(T)} = \text{cov}(X_{[uT-s/2],T}, X_{[uT+s/2],T}), \quad (14)$$

where $[s]$ denotes the largest integer not greater than s . This operator belongs to $S_2(H)$ and hence has a *local autocovariance kernel* $c_{u,s}^{(T)} \in L^2([0, 1]^2)$ given by

$$\langle \mathcal{C}_{u,s}^{(T)} g_1, g_2 \rangle = \int \int c_{u,s}^{(T)}(\tau, \sigma) g_1(\sigma) \overline{g_2(\tau)} d\sigma d\tau, \quad g_1, g_2 \in H. \quad (15)$$

Proposition 4.1. *Suppose (A1) and (A2) are satisfied. Then the local autocovariance operator defined in (14) satisfies $\sum_{s \in \mathbb{Z}} \|\mathcal{C}_{u,s}^{(T)}\|_2 < \infty$.*

The proof can be found in section A1.3 of the Appendix. Proposition 4.1 implies that the Fourier transform of (14) is a well-defined element of $S_2(H_{\mathbb{C}})$ and is given by

$$\mathcal{F}_{u,\omega}^{(T)} = \frac{1}{2\pi} \sum_s \mathcal{C}_{u,s}^{(T)} e^{-i\omega s}. \quad (16)$$

For fixed T , this operator can be seen as a functional generalization of the Wigner-Ville spectral density matrix (Martin and Flandrin, 1985) and we shall therefore refer to it as the *Wigner-Ville spectral density operator* $\mathcal{F}_{u,\omega}^{(T)}$. It is easily shown that the Fourier transform of the autocovariance kernel $c_{u,s}^{(T)}$, for fixed t and T , forms a Fourier pair in L^2 with the kernel of $\mathcal{F}_{u,\omega}^{(T)}$, referred to as the *Wigner-Ville spectral density kernel*

$$f_{u,\omega}^{(T)}(\tau, \sigma) = \frac{1}{2\pi} \sum_{s \in \mathbb{Z}} c_{u,s}^{(T)}(\tau, \sigma) e^{-i\omega s}. \quad (17)$$

More specifically, given $\sum_{s \in \mathbb{Z}} \|c_{u,s}^{(T)}\|_p < \infty$ for $p = 2$ or $p = \infty$, the spectral density kernel is uniformly bounded and uniformly continuous in ω with respect to $\|\cdot\|_p$. Additionally, the inversion formula

$$c_{u,s}^{(T)}(\tau, \sigma) = \int_{-\pi}^{\pi} f_{u,\omega}^{(T)}(\tau, \sigma) e^{is\omega} d\omega \quad (18)$$

holds in $\|\cdot\|_p$ for all s, u, T, τ , and σ . This formula and its extension to higher order cumulant kernels follows directly from an application of the dominated convergence theorem. Under additional assumptions, certain results presented in this paper will hold uniformly rather than in mean square. Sufficient would be to assume that the functional white noise process $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is mean square continuous and that the sequence of operators $\{A_{t,s}^{(T)}\}_{s \in \mathbb{Z}}$ is Hilbert-Schmidt with continuous kernels for all $t = 1, \dots, T$ and $T \in \mathbb{N}$. The process $\{X_{t,T}\}$ is then itself mean square continuous and a slight adjustment in the proof of Proposition 4.1 demonstrates that $\sum_{s \in \mathbb{Z}} \|c_{u,s}^{(T)}\|_{\infty} < \infty$. This is also expected to be the case for yet-to-be-developed concepts such as a time-varying Cramér-Karhunen-Loève representation. The results obtained in the previous section however demonstrate that a representation under these stronger conditions excludes time-varying functional AR(m) models. We will therefore not impose them but merely remark where stronger results could be obtained.

The pointwise interpretation of the L^2 -kernels makes it easy to verify that the Wigner-Ville spectral operator $\mathcal{F}_{u,\omega}^{(T)}$ is 2π -periodic in ω and self-adjoint. Namely, $c_{u,-s}^{(T)}(\sigma, \tau) = c_{u,s}^{(T)}(\tau, \sigma)$ implies $f_{u,\omega}^{\dagger(T)}(\sigma, \tau) = \overline{f_{u,\omega}^{(T)}(\tau, \sigma)}$, where f^{\dagger} the kernel function of the adjoint operator \mathcal{F}^{\dagger} . Moreover, $\mathcal{F}_{\omega}^{\varepsilon}$ is trace-class by Parseval's identity and therefore Proposition B1.3 implies that (16) is actually an element of $S_1(H_{\mathbb{C}})$. We will show in the following that (16) converges in integrated mean square to the *time-varying spectral density operator* defined as

$$\mathcal{F}_{u,\omega} = \mathcal{A}_{u,\omega} \mathcal{F}_{\omega}^{\varepsilon} \mathcal{A}_{u,\omega}^{\dagger}. \quad (19)$$

The time-varying spectral density operator satisfies all of the above properties and is non-negative definite since for every $\psi \in L_{\mathbb{C}}^2([0, 1])$,

$$\langle \mathcal{A}_{u,\omega} \mathcal{F}_{\omega}^{\varepsilon} \mathcal{A}_{u,\omega}^{\dagger} \psi, \psi \rangle = \langle \sqrt{\mathcal{F}_{\omega}^{\varepsilon}} \mathcal{A}_{u,\omega}^{\dagger} \psi, \sqrt{\mathcal{F}_{\omega}^{\varepsilon}} \mathcal{A}_{u,\omega}^{\dagger} \psi \rangle \geq 0,$$

which is a consequence of the non-negative definiteness of $\mathcal{F}_{\omega}^{\varepsilon}$. For any two elements ψ, φ in $L_{\mathbb{C}}^2([0, 1])$, one can interpret the mapping $\omega \mapsto \langle \psi, \mathcal{F}_{u,\omega} \varphi \rangle = \langle \mathcal{F}_{u,\omega} \psi, \varphi \rangle \in \mathbb{C}$ to be the local cross-spectrum of the sequences $\{\langle \psi, X_t^{(u)} \rangle\}_{t \in \mathbb{Z}}$ and $\{\langle \varphi, X_t^{(u)} \rangle\}_{t \in \mathbb{Z}}$. In particular, $\omega \mapsto \langle \psi, \mathcal{F}_{u,\omega} \psi \rangle \geq 0$ can be interpreted as the local power spectrum of $\{\langle \psi, X_t^{(u)} \rangle\}_{t \in \mathbb{Z}}$ for all $u \in [0, 1]$. In analogy to the spectral density matrix in multivariate time series, we will show below that the local spectral density operator completely characterizes the limiting second-order dynamics of the family of functional processes $\{X_{t,T} : t = 1, \dots, T\}_{T \in \mathbb{N}}$.

Theorem 4.2. *Suppose that assumptions (A1) to (A3) hold. Then, for all $u \in (0, 1)$,*

$$\int_{-\pi}^{\pi} \left\| \mathcal{F}_{u,\omega}^{(T)} - \mathcal{F}_{u,\omega} \right\|_2^2 d\omega = o(1) \quad (20)$$

as $T \rightarrow \infty$.

Proof. By definition of the Wigner-Ville operator and Lemma B2.5,

$$\begin{aligned} \mathcal{F}_{u,\omega}^{(T)} &= \frac{1}{2\pi} \sum_s \operatorname{cov} \left(\int_{-\pi}^{\pi} e^{i\lambda[uT-s/2]} \mathcal{A}_{[uT-s/2],\lambda}^{(T)} dZ_\lambda, \right. \\ &\quad \left. \int_{-\pi}^{\pi} e^{i\beta[uT+s/2]} \mathcal{A}_{[uT+s/2],\beta}^{(T)} dZ_\beta \right) e^{-i\omega s} \\ &= \frac{1}{2\pi} \sum_s \int_{-\pi}^{\pi} e^{i\lambda s} \mathcal{A}_{[uT-s/2],\lambda}^{(T)} \mathcal{F}_\lambda^\varepsilon \left(\mathcal{A}_{[uT+s/2],\lambda}^{(T)} \right)^\dagger d\lambda e^{-i\omega s}. \end{aligned}$$

Using identity (B1.3), we have that

$$\mathcal{A}_{[uT-s/2],\lambda}^{(T)} \mathcal{F}_\lambda^\varepsilon \left(\mathcal{A}_{[uT+s/2],\lambda}^{(T)} \right)^\dagger = \left(\mathcal{A}_{[uT-s/2],\lambda}^{(T)} \otimes \mathcal{A}_{[uT+s/2],\lambda}^{(T)} \right) \mathcal{F}_\lambda^\varepsilon.$$

Similarly,

$$\mathcal{F}_{u,\omega} = \frac{1}{2\pi} \sum_s \int_{-\pi}^{\pi} e^{i\lambda s} \left(\mathcal{A}_{u,\lambda} \otimes \mathcal{A}_{u,\lambda}^\dagger \right) \mathcal{F}_\lambda^\varepsilon d\lambda e^{-i\omega s}.$$

We can therefore write the left-hand side of (20) as

$$\int_{-\pi}^{\pi} \left\| \frac{1}{2\pi} \sum_s \int_{-\pi}^{\pi} e^{i\lambda s} \left(\mathcal{A}_{[uT-s/2],\lambda}^{(T)} \otimes \mathcal{A}_{[uT+s/2],\lambda}^{(T)} - \mathcal{A}_{u,\lambda} \otimes \mathcal{A}_{u,\lambda} \right) \mathcal{F}_\lambda^\varepsilon d\lambda e^{-i\omega s} \right\|_2^2 d\omega.$$

Consider the operator

$$G_{s,\lambda}^{(u,T)} = \left(\mathcal{A}_{[uT-s/2],\lambda}^{(T)} \otimes \mathcal{A}_{[uT+s/2],\lambda}^{(T)} - \mathcal{A}_{u,\lambda} \otimes \mathcal{A}_{u,\lambda} \right) \mathcal{F}_\lambda^\varepsilon$$

and its continuous counterpart

$$G_{\frac{s}{2T},\lambda} = \left(\mathcal{A}_{(u-\frac{s}{2T}),\lambda} \otimes \mathcal{A}_{(u+\frac{s}{2T}),\lambda} - \mathcal{A}_{u,\lambda} \otimes \mathcal{A}_{u,\lambda} \right) \mathcal{F}_\lambda^\varepsilon.$$

By Hölder's inequality for operators (Proposition B1.3), both are trace-class and hence Hilbert-Schmidt. Another application of Hölder's inequality together with assumption (A2) yields

$$\begin{aligned} &\left\| \left(\mathcal{A}_{[uT\mp s/2],\lambda}^{(T)} \otimes \left[\mathcal{A}_{[uT\pm s/2],\lambda}^{(T)} - \mathcal{A}_{u\pm\frac{s}{2T},\lambda} \right] \right) \mathcal{F}_\lambda^\varepsilon \right\|_2^2 \\ &\leq \left\| \mathcal{A}_{[uT\mp s/2],\lambda}^{(T)} \right\|_\infty^2 \left\| \mathcal{A}_{[uT\pm s/2],\lambda}^{(T)} - \mathcal{A}_{u\pm\frac{s}{2T},\lambda} \right\|_\infty^2 \left\| \mathcal{F}_\omega^\varepsilon \right\|_2^2 = O\left(\frac{1}{T^2}\right). \end{aligned} \quad (21)$$

By Minkowski's inequality, we obtain

$$\int_{-\pi}^{\pi} \|\mathcal{F}_{u,\omega}^{(T)} - \mathcal{F}_{u,\omega}\|_2^2 d\omega = \int_{-\pi}^{\pi} \|G_{\frac{s}{2T},\omega}\|_2^2 d\omega + o(1).$$

It is therefore sufficient to derive a bound on

$$\int_{-\pi}^{\pi} \|G_{\frac{s}{2T},\omega}\|_2^2 d\omega. \quad (22)$$

A similar argument as in (21) shows that

$$\begin{aligned} & \left\| \left(\mathcal{A}_{(u-\frac{s}{2T}),\omega} \otimes \left(\mathcal{A}_{(u+\frac{s}{2T}),\omega} - \mathcal{A}_{u,\omega} \right) + \left(\mathcal{A}_{(u-\frac{s}{2T}),\omega} - \mathcal{A}_{u,\omega} \right) \otimes \mathcal{A}_{u,\omega} \right) \mathcal{F}_{\omega}^{\varepsilon} \right\|_2 \\ & \leq C \left| \frac{s}{2T} \right|^{\alpha} \end{aligned}$$

for some constant $C > 0$. The operator-valued function $G_{u,\omega}$ is therefore Hölder continuous of order $\alpha > 1/2$ in u . Note that (18) implies the inverse Fourier transform of this operator is a well-defined element of $S_2(H_{\mathbb{C}})$. We can therefore write (22) as

$$\begin{aligned} & \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \sum_{s,s'} e^{-i\omega(s-s')} \left\langle \int_0^{2\pi} G_{\frac{s}{2T},\lambda} e^{is\lambda} d\lambda, \int_0^{2\pi} G_{\frac{s'}{2T},\lambda'} e^{is'\lambda'} d\lambda' \right\rangle \\ & = \frac{1}{2\pi} \sum_{s \in \mathbb{Z}} \|\tilde{G}_{\frac{s}{2T}}\|_2^2, \end{aligned}$$

where \tilde{G}_s can be viewed as the s -th Fourier coefficient operator of $G_{\frac{s}{2T},\lambda}$. Because of Hölder continuity, these operators satisfy $\|\tilde{G}_s\|_2 \leq \|\pi^{\alpha+1} G_{u,\omega}\|_2 |s|^{-\alpha} = O(s^{-\alpha})$. Hence,

$$\sum_{s=n}^{\infty} \|\tilde{G}_{\frac{|s|}{2T}}\|_2^2 = O(n^{1-2\alpha}).$$

Concerning the partial sum $\sum_{s=0}^{n-1} |\hat{g}_s(\tau, \sigma)|^2$, we proceed as in Dahlhaus (1996a) and use summation by parts to obtain

$$\sum_{s=0}^{n-1} \|\tilde{G}_s\|_2^2 = \int_0^{2\pi} \int_0^{2\pi} \sum_{s=0}^{n-1} e^{is(\lambda-\lambda')} \langle G_{\frac{s}{2T},\lambda}, G_{\frac{s'}{2T},\lambda'} \rangle d\lambda d\lambda' = O\left(\frac{n \log(n)}{T^{\alpha}}\right),$$

which follow from the properties of \tilde{G}_s and Lemma B3.1. It is straightforward to see that $\sum_{s=0}^{n-1} \|\tilde{G}_{-s}\|_2^2$ satisfies the same bound. Hence,

$$\begin{aligned} \int_{-\pi}^{\pi} \|\mathcal{F}_{u,\omega}^{(T)} - \mathcal{F}_{u,\omega}\|_2^2 d\omega & = \int_{-\pi}^{\pi} \|G_{\frac{s}{2T},\omega}\|_2^2 d\omega + o(1) \\ & = O\left(n^{1-2\alpha}\right) + O\left(\frac{n \log(n)}{T^{\alpha}}\right). \end{aligned}$$

Choosing an appropriate value $n \ll T$ completes the proof. \square

Intuitively, the value of n such that $n \log(n) T^{-\alpha} \rightarrow 0$ can be seen to determine the length of the data-segment over which the observations are approximately stationary. Only those functional observations $X_{t,T}$ from the triangular array with $t/T \in [u - \frac{n}{T}, u + \frac{n}{T}]$ will effectively contribute to the time-varying spectral density operator at u . As T increases, the width of this interval shrinks and sampling becomes more dense. Because the array shares dynamics through the operator-valued function $\mathcal{A}_{u,\omega}$, which is smooth in u , the observations belonging to this interval will thus become close to stationary as $T \rightarrow \infty$. The theorem therefore implies that, if we have infinitely many observations with the same probabilistic structure around some time point t , the local second-order dynamics of the family are completely characterized by $\mathcal{F}_{u,\omega}$.

The above theorem provides a promising result. It is well-known from the time series setting that a Cramér representation as given in Proposition 2.2 is in general not unique (e.g. Priestley, 1981). However, Theorem 4.2 shows that the uniqueness property as proved by Dahlhaus (1996a) generalizes to the functional setting. That is, if the family of H -valued processes $\{X_{t,T} : t = 1, \dots, T\}_{T \in \mathbb{N}}$ has a representation with common transfer operator $\mathcal{A}_{u,\omega}$ that operates on $H_{\mathbb{C}}$ and that is continuous in u , then the time-varying spectral density operator will be uniquely determined from the triangular array. This uniqueness of the time-varying spectral density operator is expected to be extremely valuable in the development of inference methods. For example, it would be of interest to determine whether this result will allow to develop Quasi Likelihood methods to fit parametric models in the functional setting. Such an extension is not direct and has to take into account the compactness of the operator and the properties of Toeplitz operators in the infinite dimension. This is however beyond the scope of this paper and the authors will consider this in future work.

Remark 4.3. If assumptions (A1) and (A2) hold with $p = 2$, we have by continuity of the inner product that the kernel $a_{u,\omega}$ of $\mathcal{A}_{u,\omega}$ is well-defined in $L^2_{\mathbb{C}}([0, 1]^2)$ and hence is uniformly Hölder continuous of order $\alpha > 1/2$ in both u and ω . If we thus additionally assume that the $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ are mean square continuous and the operator $\mathcal{A}_{u,\omega}$ is an element of \mathcal{B}_2 of which the Hilbert-Schmidt component has a kernel that is continuous in its functional arguments, the error holds in uniform norm.

5. Estimation

The time-varying spectral density operator as defined in section 4.2 allows to capture the complete second-order structure of a functional time series with possibly changing dynamics. In order to consider inferential techniques such as dynamic FPCA for nonstationary functional time series, functional Whittle likelihood methods or other general testing procedures, we require a consistent estimator for the time-varying spectral density operator. In this section, we present a nonparametric estimator of the time-varying spectral density operator. It should be noted that this requires a careful consideration of certain concepts on the function space, details of which are relegated to sections B2–B3 of the

Appendix, and that there are some discrepancies compared to existing results available in the Euclidean setting.

The section is structured as follows. First, we define a functional version of the segmented periodogram and derive its mean and covariance structure. We then consider a smoothed version of this operator and show its consistency. Finally, we provide a central limit theorem for the proposed estimator of the time-varying spectral density operator. Proofs of this section can be found in section A1.4 of the Appendix.

5.1. The functional segmented periodogram tensor

The general idea underlying inference methods in the setting of locally stationary processes is that the process $X_{t,T}$ can be considered to be close to some stationary process, say $X_t^{(u_0)}$, on a reasonably small data-segment around u_0 . If this segment is described by $\{t : |\frac{t}{T} - u_0| \leq b_t/2\}$ for some bandwidth b_t , classical estimation methods from the stationary framework can be applied on this stretch. The estimated value is subsequently assigned to be the value of the parameter curve at the midpoint u_0 of the segment. The entire parameter curve of interest in time-direction can then be obtained by shifting the segment. We will also apply this technique in the functional setting.

First, let the length of the stretch considered for estimation be denoted by N_T , where N_T is even and $N_T \ll T$. In the following, we will drop the explicit dependence of N on T and simply write $N = N_T$. Then the local version of the *functional Discrete Fourier Transform (fDFT)* is defined as

$$D_{u,\omega}^{(T)} = \sum_{s=0}^{N-1} h_{s,N} X_{[uT]-N/2+s+1,T} e^{-i\omega s}, \quad (23)$$

where $h_{s,N}$ is a data taper of length N . It is clear that $D_{u,\omega}^{(T)}$ is a 2π -periodic function in ω that takes values in H_C . The data-taper is used to improve the finite-sample properties of the estimator (Dahlhaus, 1988): firstly, it mitigates spectral leakage, which is the transfer of frequency content from large peaks to surrounding areas and is also a problem in the stationary setting. Secondly, it reduces the bias that stems from the degree of nonstationarity of the process on the given data-segment, that is, the fact that we use the observations $X_{t,T}$ for estimation rather than the unknown stationary process $X_t^{(u_0)}$. We define the data-taper by a function $h : [0, 1] \rightarrow \mathbb{R}$ and setting $h_{s,N} = h(\frac{s}{N})$; the taper function h should decay smoothly to zero at the endpoints of the interval while being essentially equal to 1 in the central part of the interval. Thus the taper gives more weight to data-points closer to the midpoint.

As a basis for estimation of the time-varying spectral density operator, we consider the normalized tensor product of the local functional Discrete Fourier Transform. This leads to the concept of a segmented or *localized periodogram tensor*

$$I_{u,\omega}^{(T)} = (2\pi H_{2,N}(0))^{-1} D_{u,\omega}^{(T)} \otimes D_{u,\omega}^{(T)}, \quad (24)$$

where

$$H_{k,N}(\omega) = \sum_{s=0}^{N-1} h_{s,N}^k e^{-i\omega s} \quad (25)$$

is the finite Fourier transform of the k -th power of the data-taper. Given the moments are well-defined in $L^2_{\mathbb{C}}([0,1]^2)$, the corresponding *localized periodogram kernel* is given by

$$I_{u,\omega}^{(T)}(\tau, \sigma) = (2\pi H_{2,N}(0))^{-1} D_{u,\omega}^{(T)}(\tau) D_{u,-\omega}^{(T)}(\sigma). \quad (26)$$

Similar to the stationary case, sufficient conditions for the existence of the higher order moments of the localized periodogram tensor are obtained from

$$\|I_{u,\omega}^{(T)}\|_2^\rho = (2\pi H_{2,N}(0))^{-\rho} \|D_{u,\omega}^{(T)}\|_2^{2\rho}, \quad (27)$$

which implies that $\mathbb{E}\|I_{u,\omega}^{(T)}\|_2^\rho < \infty$ if $\mathbb{E}\|D_{u,\omega}^{(T)}\|_2^{2\rho} < \infty$ or, in terms of moments of X , $\mathbb{E}\|X_{t,T}\|_2^{2\rho} < \infty$.

To ease notation, we denote $t_{u,r} = \lfloor uT \rfloor - N/2 + r + 1$ to be the r -th element of the data-segment with midpoint u . For $u_j = j/T$ we also write $t_{j,r} = t_{u_j,r}$ and abbreviate $u_{j,r} = t_{j,r}/T$. The following result is used throughout the rest of the paper.

Proposition 5.1. *Suppose that assumption (A1) holds with $\mathcal{A}_{t,\omega}^{(T)} \in \mathcal{B}_\infty$, and additionally $\sup_{\omega_1, \dots, \omega_{k-1}} \|\mathcal{F}_{\omega_1, \dots, \omega_{k-1}}^\varepsilon\|_2 < \infty$. Then*

$$\begin{aligned} \text{cum}(X_{t_{r_1}, T}, \dots, X_{t_{r_k}, T}) &= \int_{\Pi^k} e^{i(\lambda_1 r_1 + \dots + \lambda_k r_k)} \left(\mathcal{A}_{t_{r_1}, \lambda_1}^{(T)} \otimes \dots \otimes \mathcal{A}_{t_{r_k}, \lambda_k}^{(T)} \right) \\ &\quad \times \eta(\lambda_1 + \dots + \lambda_k) \mathcal{F}_{\lambda_1, \dots, \lambda_{k-1}}^\varepsilon d\lambda_1 \dots d\lambda_k, \end{aligned} \quad (28)$$

where the equality holds in the tensor product space $H_{\mathbb{C}} \otimes \dots \otimes H_{\mathbb{C}}$. Moreover, for fixed $t \in \{1, \dots, T\}$ and $T \in \mathbb{N}$, the k -th order cumulant spectral tensor of the linear functional process $\{X_{t,T}\}$,

$$\mathcal{F}_{\lambda_1, \dots, \lambda_{k-1}}^{(t,T)} = \left(\mathcal{A}_{t_{r_1}, \lambda_1}^{(T)} \otimes \dots \otimes \mathcal{A}_{t_{r_{k-1}}, \lambda_{k-1}}^{(T)} \otimes \mathcal{A}_{t_{r_k}, -\lambda_+}^{(T)} \right) \mathcal{F}_{\lambda_1, \dots, \lambda_{k-1}}^\varepsilon,$$

where $\lambda_+ = \lambda_1 + \dots + \lambda_{k-1}$, is well-defined in the tensor product space $\bigotimes_{i=1}^k H_{\mathbb{C}}$ with kernel $f_{\lambda_1, \dots, \lambda_{k-1}}^{(t,T)}(\tau_1, \dots, \tau_k)$.

Note that under the stronger condition $\mathcal{A}_{t,\omega}^{(T)} \in \mathcal{B}_2$, the tensor $\mathcal{F}_{\lambda_1, \dots, \lambda_{k-1}}^{(t,T)}$ will be trace-class for all $k \geq 2$. The above proposition implies that the higher order cumulant tensor of the local fDFT can be written as

$$\begin{aligned} \text{cum}(D_{u,\omega_1}^{(T)}, \dots, D_{u,\omega_k}^{(T)}) &= \int_{\Pi^k} \left(H_N(\mathcal{A}_{t_{u,\bullet}, \lambda_1}^{(T)}, \omega_1 - \lambda_1) \otimes \dots \otimes H_N(\mathcal{A}_{t_{u,\bullet}, \lambda_k}^{(T)}, \omega_k - \lambda_k) \right) \\ &\quad \times \eta(\lambda_1 + \dots + \lambda_k) \mathcal{F}_{\lambda_1, \dots, \lambda_{k-1}}^\varepsilon d\lambda_1 \dots d\lambda_k. \end{aligned} \quad (29)$$

Here, the function $H_N(G_\bullet, \omega)$ and similarly $H_{k,N}(G_\bullet, \omega)$ generalize the definitions of H_N and $H_{k,N}$ to

$$H_{k,N}(G_\bullet, \omega) = \sum_{s=0}^{N-1} h_{s,N}^k G_s e^{-i\omega s} \quad (30)$$

with $H_N(G_\bullet, \omega) = H_{1,N}(G_\bullet, \omega)$, where in our setting $G_s \in \mathcal{B}_\infty$ for all $s \in \mathbb{N}_0$. For $G_\bullet = I_{H_\mathbb{C}}$, we get back the original definitions of H_N and $H_{k,N}$. The convolution property of H_N straightforwardly generalizes to

$$\int_{\Pi} H_{k,N}(A_\bullet, \alpha + \gamma) \otimes H_{l,N}(B_\bullet, \beta - \gamma) d\gamma = 2\pi H_{k+l,N}(A_\bullet \otimes B_\bullet, \alpha + \beta), \quad (31)$$

where $(A_r)_{r=0,\dots,N-1}$ and $(B_r)_{r=0,\dots,N-1}$ are vectors of tensors or operators.

From the taper function h , we derive the smoothing kernel K_t in rescaled time u by

$$K_t(x) = \frac{1}{H_2} h\left(x + \frac{1}{2}\right)^2 \quad (32)$$

for $x \in [-\frac{1}{2}, \frac{1}{2}]$ and zero elsewhere; furthermore, we define the bandwidth $b_{t,T} = N/T$ that corresponds to segments of length N , and set $K_{t,T}(x) = \frac{1}{b_{t,T}} K_t\left(\frac{x}{b_{t,T}}\right)$. Finally, we define the kernel-specific constants

$$\kappa_t = \int_{\mathbb{R}} x^2 K_t(x) dx \quad \text{and} \quad \|K_t\|_2^2 = \int_{\mathbb{R}} K_t(x)^2 dx.$$

The first order and second order properties of the segmented functional periodogram can now be determined.

Theorem 5.2. *Suppose that assumptions (A1), (A2), (A4), and (A5) with $k \leq 4$ hold. Then the mean and covariance structure of the local functional periodogram are given by*

$$\begin{aligned} \mathbb{E}\langle I_{u,\omega}^{(T)} g_1, g_2 \rangle &= \langle \mathcal{F}_{u,\omega} g_1, g_2 \rangle + \frac{1}{2} b_{t,T}^2 \kappa_t \frac{\partial^2}{\partial u^2} \langle \mathcal{F}_{u,\omega} g_1, g_2 \rangle \\ &\quad + o(b_{t,T}^2) + O\left(\frac{\log(b_{t,T} T)}{b_{t,T} T}\right), \end{aligned}$$

and

$$\begin{aligned} \text{cov}\left(\langle I_{u,\omega_1}^{(T)}, g_1 \otimes g_2 \rangle, \langle I_{u,\omega_2}^{(T)}, g_3 \otimes g_4 \rangle\right) &= H_{2,N}\left(\langle \mathcal{F}_{\frac{t_u}{T}, \omega_1} g_3, g_1 \rangle, \omega_1 - \omega_2\right) H_{2,N}\left(\langle \mathcal{F}_{\frac{t_u}{T}, -\omega_1} g_4, g_2 \rangle, \omega_2 - \omega_1\right) \\ &\quad + H_{2,N}\left(\langle \mathcal{F}_{\frac{t_u}{T}, \omega_1} g_4, g_1 \rangle, \omega_1 + \omega_2\right) H_{2,N}\left(\langle \mathcal{F}_{\frac{t_u}{T}, -\omega_1} g_3, g_2 \rangle, -\omega_1 - \omega_2\right) \\ &\quad + O\left(\frac{\log(N)}{N}\right) + O\left(\frac{1}{N}\right), \end{aligned}$$

for all $g_1, g_2, g_3, g_4 \in H_\mathbb{C}$.

The proof exploits assumption (A2) and is based on the theory of L -functions (Dahlhaus, 1983), which allows to provide upper bound conditions on the data-taper function. Details of the extension of the latter to the functional setting can be found in section B3 of the Appendix.

5.2. Consistent estimation

Theorem 5.2 shows that the segmented periodogram tensor is not a consistent estimator. In order to obtain a consistent estimator we proceed by smoothing the estimator over different frequencies. That is, we consider convolving the segmented periodogram tensor with a window function in frequency direction

$$\hat{\mathcal{F}}_{u,\omega}^{(T)} = \frac{1}{b_{f,T}} \int_{\Pi} K_f\left(\frac{\omega - \lambda}{b_{f,T}}\right) I_{u,\lambda}^{(T)} d\lambda, \quad (33)$$

where $b_{f,T}$ denotes the bandwidth in frequency direction. To ease notation, we also write $K_{f,T}(\omega) = \frac{1}{b_{f,T}} K_f\left(\frac{\omega}{b_{f,T}}\right)$. Additionally we use subsequently

$$\kappa_f = \int_{\mathbb{R}} \omega^2 K_f(\omega) d\omega \quad \text{and} \quad \|K_f\|_2^2 = \int_{\mathbb{R}} K_f^2(\omega) d\omega$$

as an abbreviation for kernel-specific constants.

Theorem 5.3 (Properties of the estimator $\hat{\mathcal{F}}_{u,\omega}^{(T)}$). *Suppose that assumptions (A1), (A2), and (A4) to (A7) with $p = \infty$ and $k \leq 4$. Then the estimator*

$$\hat{\mathcal{F}}_{u,\omega}^{(T)} = \int_{\Pi} K_{f,T}(\omega - \lambda) I_{u,\lambda}^{(T)} d\lambda \quad (34)$$

has mean

$$\begin{aligned} \mathbb{E}\langle \hat{\mathcal{F}}_{u,\omega}^{(T)} g_1, g_2 \rangle &= \langle \mathcal{F}_{u,\omega} g_1, g_2 \rangle + \frac{\kappa_t b_{t,T}^2}{2} \frac{\partial^2}{\partial u^2} \langle \mathcal{F}_{u,\omega} g_1, g_2 \rangle + \frac{\kappa_f b_{f,T}^2}{2} \frac{\partial^2}{\partial \omega^2} \langle \mathcal{F}_{u,\omega} g_1, g_2 \rangle \\ &\quad + o(b_{t,T}^2) + o(b_{f,T}^2) + O\left(\frac{\log(b_{t,T} T)}{b_{t,T} T}\right), \end{aligned} \quad (35)$$

and covariance structure

$$\begin{aligned} &\text{cov}\left(\langle \hat{\mathcal{F}}_{u,\omega_1}^{(T)}, g_1 \otimes g_2 \rangle, \langle \hat{\mathcal{F}}_{u,\omega_2}^{(T)}, g_3 \otimes g_4 \rangle\right) \\ &= \frac{2\pi \|K_t\|_2^2}{b_{t,T} T} \int_{\Pi} K_{f,T}(\omega_1 - \lambda_1) K_{f,T}(\omega_2 - \lambda_1) \langle \mathcal{F}_{u,\lambda_1} g_3, g_1 \rangle \langle \mathcal{F}_{u,-\lambda_1} g_4, g_2 \rangle d\lambda_1 \\ &\quad + \frac{2\pi \|K_t\|_2^2}{b_{f,T} T} \int_{\Pi} K_{f,T}(\omega_1 - \lambda_1) K_{f,T}(\omega_2 + \lambda_1) \langle \mathcal{F}_{u,\lambda_1} g_4, g_1 \rangle \langle \mathcal{F}_{u,-\lambda_1} g_3, g_2 \rangle d\lambda_1 \\ &\quad + O\left(\frac{\log(b_{t,T} T)}{b_{t,T} T}\right) + O\left(\frac{b_{t,T}}{T}\right) + O((b_{t,T} b_{f,T} T)^{-2}) \end{aligned} \quad (36)$$

for all $g_1, g_2, g_3, g_4 \in H_{\mathbb{C}}$.

The proof follows from a multivariate Taylor expansion and an application of Lemma P4.1 of Brillinger (1981). We note that the covariance has greatest magnitude for $\omega_1 \pm \omega_2 \equiv 0 \pmod{2\pi}$, where the weight is concentrated in a band of width $O(b_{f,T})$ around ω_1 and ω_2 respectively. The above result shows that the bandwidths need to decay to zero with an appropriate rate in order to obtain consistency.

Proposition 5.4. *Under assumptions (A1), (A2), and (A4) to (A8) with $p = \infty$ and $k \leq 4$, we have*

$$\begin{aligned} \lim_{T \rightarrow \infty} b_{t,T} b_{f,T} T \operatorname{cov}(\langle \hat{\mathcal{F}}_{u,\omega_1}^{(T)}, g_1 \otimes g_2 \rangle, \langle \hat{\mathcal{F}}_{u,\omega_2}^{(T)}, g_3 \otimes g_4 \rangle) \\ = 2\pi \|K_t\|_2^2 \|K_f\|_2^2 \eta(\omega_1 - \omega_2) \langle \mathcal{F}_{u,\omega_1} g_3, g_1 \rangle \langle \mathcal{F}_{u,-\omega_1} g_4, g_2 \rangle \\ + 2\pi \|K_t\|_2^2 \|K_f\|_2^2 \eta(\omega_1 + \omega_2) \langle \mathcal{F}_{u,\omega_1} g_4, g_1 \rangle \langle \mathcal{F}_{u,-\omega_1} g_3, g_2 \rangle \end{aligned} \quad (37)$$

for all $g_1, g_2, g_3, g_4 \in H_{\mathbb{C}}$ and for fixed ω_1, ω_2 . If ω_1, ω_2 depend on T then the convergence holds provided that $\liminf_{T \rightarrow \infty} |(\omega_{1,T} \pm \omega_{2,T}) \bmod 2\pi| > \varepsilon$ for some $\varepsilon > 0$.

The proof follows straightforwardly from a change of variables and a functional generalization of approximate identities (e.g., Edwards, 1967).

Corollary 5.5. *Under the same assumptions*

$$\left\| \operatorname{cov}(\hat{\mathcal{F}}_{u,\omega_1}^{(T)}, \hat{\mathcal{F}}_{u,\omega_2}^{(T)}) \right\|_2 = O\left(\frac{1}{b_{f,T} b_{t,T} T}\right)$$

uniformly in $\omega_1, \omega_2 \in [-\pi, \pi]$ and $u \in [0, 1]$.

Proof. Note that $\|K_{f,T}\|_{\infty} = O\left(\frac{1}{b_{f,T}}\right)$ and $\|K_{f,T}\|_1 = 1$, from which it is easy to see that

$$\sup_{\omega} \left\| \int_{\Pi} K_{f,T}(\omega + \lambda) K_{f,T}(\lambda) d\lambda \right\|_2 = O\left(\frac{1}{b_{f,T}}\right).$$

Since $\|\mathcal{F}_{u,\omega}\|_2$ is moreover uniformly bounded in u and ω , the statement follows directly from equation (A1.15). \square

Remark 5.6. Theorem 5.2, Theorem 5.3, Proposition 5.4, and Corollary 5.5 can be shown to hold in a stronger sense under additional assumptions. Namely, if the respective set of assumptions hold with $p = 2$ then the kernel function of the transfer operator $\mathcal{A}_{t,\omega}^{(T)}$ is well-defined. If this function is continuous and the white noise process $\{\varepsilon_t\}$ is moreover mean square continuous with $\sup_{\omega_1, \dots, \omega_{k-1}} \|f_{\omega_1, \dots, \omega_{k-1}}^{\varepsilon}\|_{\infty} < \infty$ for $k \leq 4$, the aforementioned statements hold in uniform norm.

Theorem 5.7 (Convergence in integrated mean square). *Suppose that assumptions (A1), (A2), and (A4) to (A8) with $k = 2, 4$ hold. Then the spectral density operator is consistent in integrated mean square. More precisely, we have*

$$\begin{aligned} \operatorname{IMSE}(\hat{\mathcal{F}}_{u,\omega}^{(T)}) &= \int_{\Pi} \mathbb{E} \|\hat{\mathcal{F}}_{u,\omega}^{(T)} - \mathcal{F}_{u,\omega}\|_2^2 d\omega \\ &= O((b_{t,T} b_{f,T} T)^{-1}) + o(b_{t,T}^2 + b_{f,T}^2 + (b_{t,T} T)^{-1} \log(b_{t,T} T)). \end{aligned}$$

Since it is uniform in $\omega \in \Pi$, we have pointwise mean square convergence where the error also satisfies $\mathbb{E} \|\hat{\mathcal{F}}_{u,\omega}^{(T)} - \mathcal{F}_{u,\omega}\|_2^2 = O\left(\frac{1}{b_t b_f T}\right) + o\left(b_t^2 + b_f^2 + \frac{\log b_t T}{b_t T}\right)$.

The proof follows almost straightforwardly from decomposing the above in terms of its variance and its squared bias.

Remark 5.8 (Discrete observations). In practice, functional data commonly are observed not continuously but only on a discrete grid. In this case, the above consistency results continue to hold only under additional regularity assumptions. To illustrate the effect of discrete observations, suppose that, for given T , we observe the random functions $X_t(\tau)$, $t = 1, \dots, T$, on the grid points $0 \leq \tau_1 < \dots < \tau_M \leq 1$. The corresponding discrete data $x_{i,t} = X_t(\tau_i)$ are interpolated to yield functions $X_t^M(\tau)$ in some M -dimensional subspace H_M of H . Furthermore, let $P_M X_t$ be the orthogonal projection of X_t onto H_M . Then the mean square error of the estimator $\hat{\mathcal{F}}_{u,\omega}(X^M)$ based on the discrete observation can be bounded by application of Minkowski's inequality by

$$\begin{aligned} & \mathbb{E} \|\hat{\mathcal{F}}_{u,\omega}(X^M) - \mathcal{F}_{u,\omega}\|_2^2 \\ & \leq 2 \mathbb{E} \|\hat{\mathcal{F}}_{u,\omega}(X^M) - \hat{\mathcal{F}}_{u,\omega}(P_M X)\|_2^2 + 2 \mathbb{E} \|\hat{\mathcal{F}}_{u,\omega}(P_M X) - \mathcal{F}_{u,\omega}\|_2^2. \end{aligned} \quad (38)$$

Here, the first term can be interpreted as the error due to discretization. In the second term, the mean square error of $\hat{\mathcal{F}}_{u,\omega}(P_M X)$ can be rewritten as

$$\begin{aligned} & \mathbb{E} \|\hat{\mathcal{F}}_{u,\omega}(P_M X) - \mathcal{F}_{u,\omega}\|_2^2 \\ & = \mathbb{E} \|\hat{\mathcal{F}}_{u,\omega}(P_M X) - \mathcal{F}_{u,\omega}(P_M X)\|_2^2 + \|\mathcal{F}_{u,\omega}(P_M X) - \mathcal{F}_{u,\omega}\|_2^2, \end{aligned}$$

where $\mathcal{F}_{u,\omega}(P_M X)$ is the time-varying spectral density operator of the process $P_M X_t$. The first term describes the estimation error and, by Parseval's equality, is bounded by

$$\mathbb{E} \|\hat{\mathcal{F}}_{u,\omega}(P_M X) - \mathcal{F}_{u,\omega}(P_M X)\|_2^2 \leq \mathbb{E} \|\hat{\mathcal{F}}_{u,\omega}(X) - \mathcal{F}_{u,\omega}\|_2^2$$

and hence converges to zero as $T \rightarrow \infty$ in a local stationary framework. Finally, the second term is the approximation error due to replacing the functions X_t by their projections $P_M X_t$. Under suitable regularity conditions on the functions X_t and for an increasingly dense grid, the approximation error can be made arbitrarily small. Furthermore, the discretization error $\|X_t^M - P_M X_t\|_2$ converges to zero such that for an appropriate rate of $M \rightarrow \infty$, the error due to discretization in 38 also tends down to zero. The detailed derivations are similar to those in section 5 of Panaretos and Tavakoli (2013b) and are therefore omitted.

5.3. Weak convergence of the empirical process

The results of the previous section give rise to investigating the limiting distribution of $\hat{\mathcal{F}}_{u,\omega}^{(T)}$, the local estimator of the spectral density operator. We will proceed by showing that joint convergence of its kernel $\hat{f}_{u,\omega}^{(T)}$ to complex Gaussian elements in $L_{\mathbb{C}}^2([0, 1]^2)$ can be established.

Consider the sequence of random elements $(\hat{E}_{u,\omega}^{(T)})_{T \in \mathbb{N}}$ in $L_{\mathbb{C}}^2([0, 1]^2)$, where

$$\hat{E}_{u,\omega}^{(T)} = \sqrt{b_{t,T} b_{f,T} T} (\hat{f}_{u,\omega}^{(T)} - \mathbb{E}[\hat{f}_{u,\omega}^{(T)}])$$

for fixed $\omega \in [-\pi, \pi]$ and $u \in [0, 1]$. In order to establish convergence in $L_{\mathbb{C}}^2([0, 1]^2)$, it is more appropriate to consider the representation of $\hat{E}_{u,\omega}^{(T)}$ with

respect to some orthonormal basis. For this, let $\{\psi_m\}_{m \in \mathbb{N}}$ be an orthonormal basis of $H_{\mathbb{C}}$. Then $\{\psi_{mn}\}_{m,n \in \mathbb{N}}$ with $\psi_{mn} = \psi_m \otimes \psi_n$ forms an orthonormal basis of $L^2_{\mathbb{C}}([0, 1]^2)$, and $\hat{E}_{u,\omega}^{(T)}$ equals

$$\hat{E}_{u,\omega}^{(T)} = \sum_{m,n \in \mathbb{N}} \langle \hat{E}_{u,\omega}^{(T)}, \psi_{mn} \rangle \psi_{mn}.$$

Hence, the distribution of $\hat{E}_{u,\omega}^{(T)}$ is fully characterized by the finite-dimensional distribution of the coefficients of its basis representation. Furthermore, weak convergence of $\hat{E}_{u,\omega}^{(T)}$ will follow from the weak convergence of $(\langle \hat{E}_{u,\omega}^{(T)}, \psi_{mn} \rangle)_{m,n \in \mathbb{N}}$ in the sequence space $\ell^2_{\mathbb{C}}$. Subsequently, we identify $\hat{E}_{u,\omega}^{(T)}$ with its dual $(\hat{E}_{u,\omega}^{(T)})^* \in L^2_{\mathbb{C}}([0, 1]^2)^*$ and write

$$\hat{E}_{u,\omega}^{(T)}(\phi) = \langle \hat{E}_{u,\omega}^{(T)}, \phi \rangle$$

for all $\phi \in L^2_{\mathbb{C}}([0, 1]^2)$.

To show convergence to a Gaussian functional process, we make use of the following result by Cremers and Kadelka (1986), which weakens the tightness condition usually employed to prove weak convergence and generalizes earlier results by Grinblat (1976).

Lemma 5.9. *Let (T, \mathcal{B}, μ) be a measure space, let $(E, |\cdot|)$ be a Banach space, and let $(X_n)_{n \in \mathbb{N}}$ be a sequence of random elements in $L^p_E(T, \mu)$ such that*

- (i) *the finite-dimensional distributions of X_n converge weakly to those of a random element X_0 in $L^p_E(T, \mu)$ and*
- (ii) $\limsup_{n \rightarrow \infty} \mathbb{E} \|X_n\|_p^p \leq \mathbb{E} \|X_0\|_p^p$.

Then X_n converges weakly to X_0 in $L^p_E(T, \mu)$.

In our setting, the weak convergence of the process $\hat{E}_{u,\omega}^{(T)}$ in $L^2_{\mathbb{C}}([0, 1]^2)$ will follow from the joint convergence of $\hat{E}_{u,\omega}^{(T)}(\psi_{m_1, n_1}), \dots, \hat{E}_{u,\omega}^{(T)}(\psi_{m_k, n_k})$ for all $k \in \mathbb{N}$ and the condition

$$\mathbb{E} \|\hat{E}_{u,\omega}^{(T)}\|_2^2 = \sum_{m,n \in \mathbb{N}} \mathbb{E} |\hat{E}_{u,\omega}^{(T)}(\psi_{mn})|^2 \rightarrow \sum_{m,n \in \mathbb{N}} \mathbb{E} |E_{u,\omega}(\psi_{mn})|^2 = \mathbb{E} \|E_{u,\omega}\|_2^2 \quad (39)$$

as $T \rightarrow \infty$. In contrast, Panaretos and Tavakoli (2013b) employ the slightly stronger condition

$$|\hat{E}_{u,\omega}^{(T)}(\psi_{mn})|^2 \leq \phi_{mn}$$

for all $T \in \mathbb{N}$ and $m, n \in \mathbb{N}$ and some sequence $(\phi_{mn}) \in \ell^1$. In fact, the condition corresponds in our setting to the one given in Grinblat (1976). Finally, we note that condition (39) is sufficient for our purposes, but recently it has been shown (Bogachev and Miftakhov, 2015) that it can be further weakened to

$$\sup_{T \in \mathbb{N}} \mathbb{E} \|\hat{E}_{u,\omega}^{(T)}\|_2^2 < \infty.$$

For the convergence of the finite-dimensional distributions, we show convergence of the cumulants of all orders to that of the limiting process. For the first

and second order cumulants of $\hat{E}_{u,\omega}^{(T)}(\psi_{mn})$, this follows from Theorem 5.3. It therefore remains to show that all cumulants of higher order vanish asymptotically.

Proposition 5.10. *Suppose that assumptions (A1), (A2), and (A4) to (A8) for some $k \geq 3$ hold. Then, for all $u \in [0, 1]$ and for all $\omega_i \in [-\pi, \pi]$ and $m_i, n_i \in \mathbb{N}$ for $i = 1, \dots, k$, we have*

$$\text{cum}(\hat{E}_{u,\omega_1}^{(T)}(\psi_{m_1 n_1}), \dots, \hat{E}_{u,\omega_k}^{(T)}(\psi_{m_k n_k})) = o(1) \quad (40)$$

as $T \rightarrow \infty$.

The distributional properties of the functional process can now be summarized in the following theorem.

Theorem 5.11 (Weak convergence). *Suppose that assumptions (A1), (A2), and (A4) to (A8) for all $k \in \mathbb{N}$ hold. Then*

$$(\hat{E}_{u,\omega_j}^{(T)})_{j=1,\dots,J} \xrightarrow{\mathcal{D}} (E_{u,\omega_j})_{j=1,\dots,J}, \quad (41)$$

where E_{u,ω_j} , $j = 1, \dots, J$, are jointly Gaussian elements in $L^2_{\mathbb{C}}([0, 1]^2)$ with means $\mathbb{E}(E_{u,\omega_i}(\psi_{mn})) = 0$ and covariances

$$\begin{aligned} & \text{cov}(E_{u,\omega_i}(\psi_{mn}), E_{u,\omega_j}(\psi_{m'n'})) \\ &= 2\pi \|K_t\|_2^2 \|K_f\|_2^2 \left[\eta(\omega_i - \omega_j) \langle \mathcal{F}_{u,\omega_i} \psi_{m'}, \psi_m \rangle \langle \mathcal{F}_{u,-\omega_i} \psi_{n'}, \psi_n \rangle \right. \\ & \quad \left. + \eta(\omega_i + \omega_j) \langle \mathcal{F}_{u,\omega_i} \psi_{n'}, \psi_m \rangle \langle \mathcal{F}_{u,-\omega_i} \psi_{m'}, \psi_n \rangle \right] \end{aligned} \quad (42)$$

for all $i, j \in 1, \dots, J$ and $m, m', n, n' \in \mathbb{N}$.

Proof of Theorem 5.11. For condition 39, we note that

$$\mathbb{E} \|\hat{E}_{u,\omega}^{(T)}\|_2^2 = \int_{[0,1]^2} \text{var}(\hat{E}_{u,\omega}^{(T)}(\tau, \sigma)) d\tau d\sigma = b_{t,T} b_{f,T} T \|\text{var}(\mathcal{F}_{u,\omega}^{(T)})\|_2^2$$

and it therefore is satisfied by Theorem 5.3. Together with the convergence of the finite-dimensional distributions this proves the asserted weak convergence. \square

6. Numerical simulations

To illustrate the performance of the estimator in finite samples, we consider a time-varying functional time series with representation

$$X_{t,T} = B_{\frac{t}{T},1}(X_{t-1,T}) + \varepsilon_t, \quad (43)$$

where $B_{u,1} \in \mathcal{B}_{\infty}$ is continuous in $u \in [0, 1]$ and where $\{\varepsilon_t\}$ is a collection of independent innovation functions. In order to generate the process, let $\{\psi_i\}_{i \in \mathbb{N}}$ be an orthonormal basis of H and denote the vector of the first k Fourier coefficients of $X_{t,T}$ by $\mathbf{X}_t^{(T)} = (\langle X_{t,T}, \psi_1 \rangle, \dots, \langle X_{t,T}, \psi_k \rangle)'$. Similar to Hörmann,

Kidziński and Hallin (2015), we exploit that the linearity of the autoregressive operator implies the first k Fourier coefficients, for k large, approximately satisfy a VAR(1) equation. That is,

$$\mathbf{X}_t^{(T)} \approx \mathfrak{B}_{\frac{t}{T},1} \mathbf{X}_{t-1}^{(T)} + \boldsymbol{\varepsilon}_t \quad \forall t, T, \quad (44)$$

where $\boldsymbol{\varepsilon}_t = (\langle \varepsilon_t, \psi_1 \rangle, \dots, \langle \varepsilon_t, \psi_k \rangle)'$ and $\mathfrak{B}_{\frac{t}{T},1} = (\langle B_{\frac{t}{T},1}(\psi_i), \psi_j \rangle, 1 \leq i, j \leq k)$. Correspondingly, the local spectral density kernel will satisfy

$$f_{u,\omega}^{(T)}(\tau, \sigma) \approx \lim_{i,j \rightarrow \infty} \sum_{i,j=1}^k \mathfrak{f}_{u,\omega,i,j}^{(T)} \psi_i(\tau) \psi_j(\sigma),$$

where $\mathfrak{f}_{u,\omega}^{(T)}$ is the spectral density matrix of the Fourier coefficients in (44). Implementation was done in R together with the `fda` package. For the simulations, we choose the Fourier basis functions on $[0, 1]$. The construction of the estimator in (33) requires specification of smoothing kernels and corresponding bandwidths in time- as well as frequency direction. Although the choice of the smoothing kernels usually does not affect the performance significantly, bandwidth selection is a well-known problem in nonparametric statistics. As seen from Theorem 5.3, both bandwidths influence the bias-variance relation. Depending on the persistence of the autoregressive process a smaller bandwidth in frequency direction is desirable around the peak (at $\lambda = 0$ for the above process), while slow changes in time direction allow for tapering (i.e., smoothing in time direction) over more functional observations. It would therefore be of interest to develop an adaptive procedure as proposed in van Delft and Eichler (2015) to select the bandwidth parameters. Investigation of this is however beyond the scope of the present paper. In the examples below, the bandwidths were set fixed to $b_{t,T} = T^{-1/6}$ and $b_{f,T} = 2T^{-1/5} - b_{t,T}$. We chose as smoothing kernels

$$K_t(x) = K_f(x) = 6\left(\frac{1}{4} - x^2\right) \quad x \in \left[-\frac{1}{2}, \frac{1}{2}\right],$$

which have been shown to be optimal in the time series setting (Dahlhaus, 1996b).

In order to construct the matrix $\mathfrak{B}_{\frac{t}{T},1}$, we first generate a matrix A_u with entries that are mutually independent Gaussian where the (i, j) -th entry has variance

$$ui^{-2c} + (1 - u)e^{-i-j}.$$

The entries will tend to zero as $i, j \rightarrow \infty$, because the operator $B_{\frac{t}{T},1}$ is required to be bounded. The matrix $\mathfrak{B}_{u,1}$ is consequently obtained as $\mathfrak{B}_{u,1} = \eta A_u / \|A_u\|_\infty$. The value of η thus determines the persistence of the process. Additionally, the collection of innovation functions $\{\varepsilon_t\}$ is specified as a linear combination of the Fourier basis functions with independent zero-mean Gaussian coefficients such that the l -th coefficient $\langle \varepsilon_t, \psi_l \rangle$ has variance $1/[(l - 1.5)\pi]^2$. The parameters were set to $c = 3$ and $\eta = 0.4$. To visualize the variability of the estimator, figure 2 depicts the amplitude of the true spectral density kernel

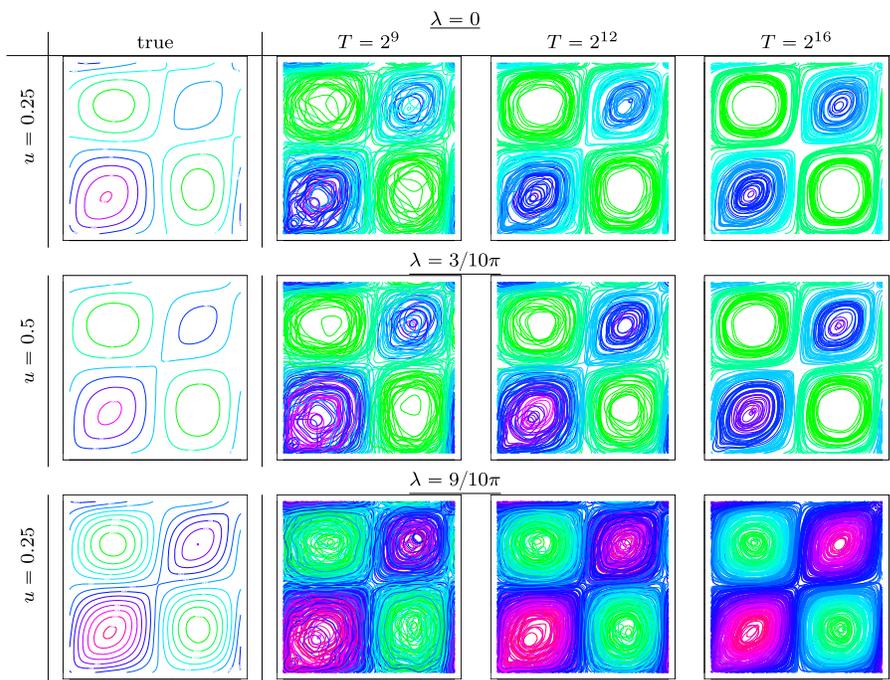


Fig 2: Contour plots of the true and estimated spectral density of the FAR(1) at different time points at frequencies $\lambda = 0$, $\lambda = \frac{3}{10}\pi$ and $\lambda = \frac{9}{10}\pi$.

of the process for various values of u and λ with 20 replications of the corresponding estimator superposed for different sample sizes T . For each row, the same level curves were used where each level curve has the same color-coding within that row. The first two rows of figure 2 give the different levels for the estimator around the peak in frequency direction, while the last row provides contour plots further away from the peak. Increasing the sample size leads to less variability, as can be seen from the better aligned contour lines. It can also be observed that the estimates become more stable as we move further away from the peak. Nevertheless, the peaks and valleys are generally reasonably well captured even for the contour plots in the area around the peak.

As a second example, we consider a FAR(2) with the location of the peak varying with time. More specifically, the Fourier coefficients are now obtained by means of a VAR(2)

$$\mathbf{X}_t^{(T)} = \mathfrak{B}_{\frac{t}{T},1} \mathbf{X}_{t-1}^{(T)} + \mathfrak{B}_{\frac{t}{T},2} \mathbf{X}_{t-2}^{(T)} + \varepsilon_t,$$

where $\mathfrak{B}_{u,1} = \eta_{u,1} A_{u,1} / \|A_{u,1}\|_\infty$ and $\mathfrak{B}_{u,2} = \eta_{u,2} A_{u,2} / \|A_{u,2}\|_\infty$. The entries of the matrices $A_{u,1}$ and $A_{u,2}$ are mutually independent and are generated such that $[A_{u,1}]_{i,j} = \mathcal{N}(0, e^{-(i-3)-(j-3)})$ and $[A_{u,2}]_{i,j} = \mathcal{N}(0, (i^{8/2} + j^{2/2})^{-1})$,

respectively. The norms are specified as

$$\eta_{u,1} = 0.4 \cos(1.5 - \cos(\pi u)) \text{ and } \eta_{u,2} = -0.5.$$

This will result in the peak to be located at $\lambda = \arccos(0.3 \cos[1.5 - \cos(\pi u)])$. The collection of innovation functions $\{\varepsilon_t\}$ is chosen such that the l -th coefficient $\langle \varepsilon_t, \psi_l \rangle$ has variance $1/[(l - 2.65)\pi]^2$. Figure 3 provides the contour plots for different local time values where the frequency was set to $\lambda = 1.5 - \cos(\pi u)$, i.e., the direction in which most change in time-direction is visible in terms of amplitude. We observe good results in terms of identifying the peaks and valleys overall where again the variability clearly reduces for $T > 512$. For the value $u = 0.5$, one is really close to the location of a peak and observe wrongful detection of a small peak in the middle of the contour plot. This is an indication some over-smoothing occurs which, to some extent, is difficult to prevent for autoregressive models, even in the stationary time series case.

7. Concluding remarks

This paper forms a basis for the development of statistical techniques and methods for the analysis of nonstationary functional time series. We have provided a theoretical framework for meaningful statistical inference of functional time series with dynamics that change slowly over time. For this, the notion of local stationarity was introduced for time series on the function space. We focused on a class of functional locally stationary processes for which a time-varying functional Cramér representation exists. The second-order characteristics of processes belonging to this class are completely captured by the time-varying spectral density operator. We moreover introduced time-varying functional ARMA processes and showed that these belong to the class of locally stationary functional processes. In the last section, we considered the nonparametric estimation of the time-varying spectral density operator. To derive the asymptotic distribution, a weaker tightness criterion is used than what is common in the existing literature. The results derived in this paper give rise to consider Quasi-likelihood methods on the function space as well as the development of prediction and appropriate dimension reduction techniques for nonstationary time series on the function space. This is left for future work.

Appendix A1: Proofs

In this appendix, we prove the main theoretical results of the paper.

A1.1. Proofs of Section 2

Proof of Proposition 2.3. For fixed $t \in \{1, \dots, T\}$ and $T \in \mathbb{N}$, let $U_{s,\omega} = e^{i\omega(t-s)} A_{t,s}^{(T)}$. We have

$$\mathcal{T}(U_{s,\cdot}) = \int_{-\pi}^{\pi} e^{i\omega(t-s)} A_{t,s}^{(T)} dZ_{\omega} = A_{t,s}^{(T)} \varepsilon_{t-s},$$

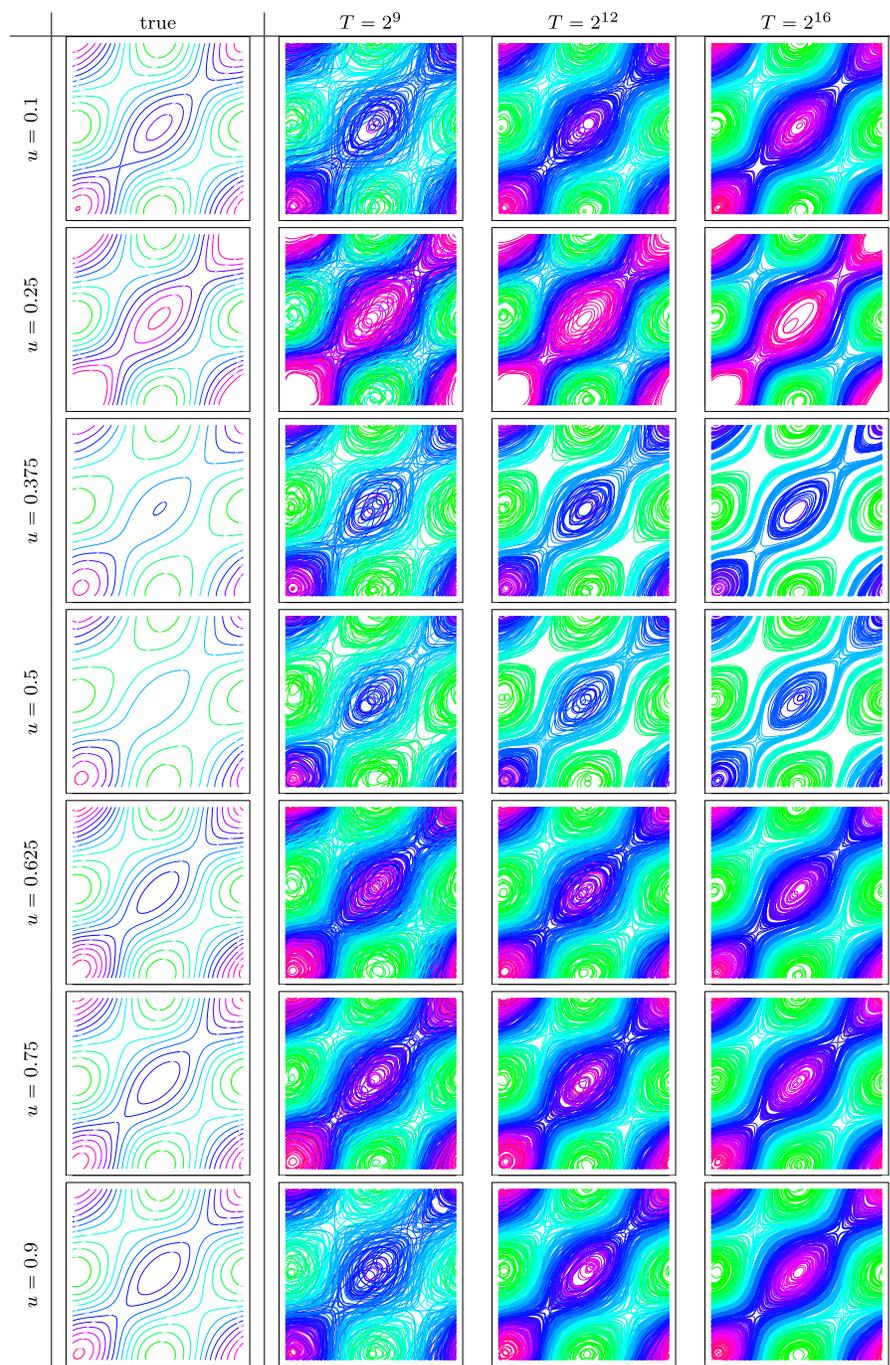


Fig 3: Contour plots of the true and estimated spectral density of the FAR(2) at different time points for $\lambda = 1.5 - \cos(\pi u)$.

where \mathcal{T} is the mapping defined by linear extension of $\mathcal{T}(U \mathbf{1}_{[\alpha, \beta]}) = U(Z_\beta - Z_\alpha)$ (see section B2.3). By definition of the operator $U_{s, \cdot}$, we have $\|U_{s, \cdot}\|_{\mathcal{B}_\infty}^2 \leq \|A_{t, s}^{(T)}\|_\infty^2 \int_{-\pi}^\pi \|\mathcal{F}_\omega^\varepsilon\|_1 d\omega < \infty$ and thus $U_{s, \cdot} \in \mathcal{B}_\infty$. Similarly, $\sum_s \mathcal{T}(U_{s, \cdot}) \in \mathcal{B}_\infty$ from which it follows that

$$\lim_{N \rightarrow \infty} \sum_{|s| \leq N} e^{i\omega(t-s)} A_{t, s}^{(T)} = e^{i\omega t} \sum_{s \in \mathbb{Z}} e^{-i\omega s} A_{t, s}^{(T)} = e^{i\omega t} \mathcal{A}_{t, \omega}^{(T)} \in \mathcal{B}_\infty.$$

The continuity of the mapping \mathcal{T} then implies

$$X_{t, T} = \sum_s \mathcal{T}(U_{s, \cdot}) = \mathcal{T}\left(\sum_s U_{s, \cdot}\right) = \int_{-\pi}^\pi e^{i\omega t} \mathcal{A}_{t, \omega}^{(T)} dZ_\omega \quad \text{a.e. in } \mathbb{H}. \quad \square$$

A1.2. Proofs of Section 3

Proof of Lemma 3.2. We follow the lines of Bosq (2000, Theorem 5.2, Corollary 5.1). To ease notation, we shall write I and O for the identity and zero operator on H , respectively while we denote the identity operator on H^m by I_{H^m} . Consider the bounded linear operator $\tilde{P}_u(\lambda)$ on H

$$\tilde{P}_u(\lambda) = \lambda^m I - \lambda^{m-1} B_{u,1} - \dots - \lambda B_{u,m-1} - B_{u,m}, \quad \lambda \in \mathbb{C}.$$

It is straightforward to derive that, under the assumption $\sum_{j=1}^m \|B_{u,j}\|_\infty < 1$, non-invertibility of $\tilde{P}_u(\lambda)$ implies that λ has modulus strictly less than 1. Define the invertible matrices $U_u(\lambda)$ and $M_u(\lambda)$ on the complex extension H^m by

$$U_{u,ij}(\lambda) = \begin{cases} \lambda^{j-i} I_H & \text{if } j \geq i, \\ O_H & \text{otherwise} \end{cases}$$

for $i, j = 1, \dots, m$ and

$$M_u(\lambda) = \begin{pmatrix} O_{H^{(m-1) \times 1}} & -I_{H^{m-1}} \\ P_{u,0}(\lambda) & P_{u,1}(\lambda), \dots, P_{u,m-1}(\lambda) \end{pmatrix},$$

where $P_{u,0}(\lambda) = I$ and $P_{u,j}(\lambda) = \lambda_u P_{u,j-1}(\lambda) - B_{u,j}$ for $j = 1, \dots, m$. Then

$$M_u(\lambda) \begin{pmatrix} \lambda I_{H^m} - \mathbf{B}_u^* \\ O_{H^1 \times (m-1)} \end{pmatrix} U_u(\lambda) = \begin{pmatrix} I_{H^{(m-1)}} & O_{H^{(m-1) \times 1}} \\ O_{H^1 \times (m-1)} & \tilde{P}_u(\lambda) \end{pmatrix},$$

from which it follows that $(\lambda I_{H^m} - \mathbf{B}_u^*)$ is not invertible when $\tilde{P}_u(\lambda)$ is not invertible. In other words, the spectrum S_u of \mathbf{B}_u^* over the complex extension of H^m , which is a closed set, satisfies

$$\begin{aligned} S_u &= \{\lambda \in \mathbb{C} : \lambda I_{H^m} - \mathbf{B}_u^* \text{ not invertible}\} \subset \{\lambda \in \mathbb{C} : \tilde{P}_u(\lambda) \text{ not invertible}\} \\ &= \{\lambda \in \mathbb{C} : |\lambda| < 1\}. \end{aligned}$$

Hence, the assumption that $\sum_{j=1}^m \|B_{u,j}\|_\infty < 1$ for all u , implies that the spectral radius of \mathbf{B}_u^* satisfies

$$r(\mathbf{B}_u^*) = \sup_{\lambda \in S_u} |\lambda| = \lim_{k \rightarrow \infty} \|\mathbf{B}_u^{*k}\|_\infty^{1/k} < \frac{1}{1 + \delta} \quad (\text{A1.1})$$

for some $\delta > 0$. The equality is a well-known result for the spectral radius of bounded linear operators¹ and can for example be found in Dunford and Schwartz (1958). From (A1.1) it is now clear that there exists a $k_0 \in \mathbb{Z}$, $\alpha \in (0, 1)$ and a constant c_1 such that for all $k \geq k_0$

$$\|\mathbf{B}_u^{*k}\|_\infty < c_1 \alpha^k. \quad (\text{A1.2})$$

Finally, it has been shown in Bosq (2000, p.74) that this is equivalent to the condition $\|\mathbf{B}_u^{*k_0}\|_\infty < 1$ for some integer $k_0 \geq 1$. \square

Proof of Theorem 3.4. The moving average representation (7) and the difference equation (11) together imply that the process can be represented as

$$X_{t,T} = \sum_{l=0}^{\infty} A_{t,l}^{(T)} C_{\frac{t-l}{T}}^{-1} \sum_{j=0}^m B_{\frac{t-l}{T},j} (X_{t-l-j,T}).$$

Using the linearity of the operators and applying a change of variables $l' = l + j$, this can be written as

$$X_{t,T} = \sum_{l'=0}^{\infty} \sum_{j=0}^m A_{t,l'-j}^{(T)} C_{\frac{t-l'+j}{T}}^{-1} B_{\frac{t-l'+j}{T},j} (X_{t-l',T}),$$

where $A_{t,l'-j}^{(T)} = O_H$ for $l' < j$. For a purely nondeterministic solution we require

$$\sum_{j=0}^m A_{t,l'-j}^{(T)} C_{\frac{t-l'+j}{T}}^{-1} B_{\frac{t-l'+j}{T},j} = \begin{cases} I_H & \text{if } l' = 0, \\ O_H & \text{if } l' \neq 0. \end{cases} \quad (\text{A1.3})$$

Because ε_t is white noise in \mathbb{H} , it has spectral representation

$$\varepsilon_t = (2\pi)^{-1/2} \int_{-\pi}^{\pi} e^{i\omega t} dZ_\omega, \quad t \in \mathbb{Z}. \quad (\text{A1.4})$$

Since a solution of the form (7) exists, we also have

$$X_{t,T} = \int_{-\pi}^{\pi} e^{i\omega t} \mathcal{A}_{t,\omega}^{(T)} dZ_\omega,$$

where $\mathcal{A}_{t,\omega}^{(T)} = \frac{1}{\sqrt{2\pi}} \sum_{l=0}^{\infty} A_{t,l}^{(T)} e^{-i\omega l}$. Substituting the spectral representations of $X_{t,T}$ and ε_t into (11), we get together with the linearity of the operators $B_{u,j}$ and $\mathcal{A}_{t,\omega}^{(T)}$

$$\int_{-\pi}^{\pi} \sum_{j=0}^m e^{i\omega(t-j)} B_{\frac{t}{T},j} \mathcal{A}_{t-j,\omega}^{(T)} dZ_\omega = (2\pi)^{-1/2} \int_{-\pi}^{\pi} e^{i\omega t} C_{\frac{t}{T}} dZ_\omega,$$

¹Gelfand's formula.

Given the operator $\mathcal{A}_{\frac{t}{T},\omega}$ satisfies equation (12), the previous implies we can write

$$\begin{aligned} \frac{1}{\sqrt{2\pi}}C_{\frac{t}{T}} &= \sum_{j=0}^m e^{-i\omega j} B_{\frac{t}{T},j} \mathcal{A}_{\frac{t}{T},\omega} \\ &= \sum_{j=0}^m e^{-i\omega j} B_{\frac{t}{T},j} \mathcal{A}_{\frac{t-j}{T},\omega} + \sum_{j=0}^m e^{-i\omega j} B_{\frac{t}{T},j} (\mathcal{A}_{\frac{t}{T},\omega} - \mathcal{A}_{\frac{t-j}{T},\omega}). \end{aligned}$$

From the last equation, it follows that

$$\begin{aligned} \sum_{j=0}^m e^{i\omega(t-j)} B_{\frac{t}{T},j} (\mathcal{A}_{\frac{t-j}{T},\omega}^{(T)} - \mathcal{A}_{\frac{t-j}{T},\omega}) &= \sum_{j=0}^m e^{i\omega(t-j)} B_{\frac{t}{T},j} (\mathcal{A}_{\frac{t}{T},\omega} - \mathcal{A}_{\frac{t-j}{T},\omega}) \\ &= C_{\frac{t}{T}} \Omega_{t,\omega}^{(T)}, \end{aligned} \quad (\text{A1.5})$$

where $\Omega_{t,\omega}^{(T)} = O_H$, $t \leq 0$. We will show that this operator is of order $O(\frac{1}{T})$ in $S_p(H_{\mathbb{C}})$. By Proposition B1.6, the smooth transfer operator satisfies $\mathcal{A}_{u,\omega} \in S_p(H_{\mathbb{C}})$. Under the conditions of Theorem 3.4, we have that for any element $\psi \in H_{\mathbb{C}}$ and fixed $\omega \in \Pi$, the mapping $u \mapsto \mathcal{A}_{u,\omega}(\psi)(\tau)$ is continuous and, from the properties of the $B_{u,j}$, is differentiable and has bounded derivatives with respect to u . By the Mean Value Theorem, we obtain

$$\sup_{t,\omega} \|\mathcal{A}_{\frac{t}{T},\omega} - \mathcal{A}_{\frac{t-j}{T},\omega}\|_p \leq \sup_{t,\omega} \sup_{\frac{t-j}{T} < u_0 < \frac{t}{T}} \left| \frac{j}{T} \right| \left\| \frac{\partial}{\partial u} \mathcal{A}_{u,\omega} \Big|_{u=u_0} \right\|_p = O\left(\frac{1}{T}\right),$$

for all $\omega \in \Pi$, uniformly in u . It then easily follows from equation (A1.5) and Proposition B1.3 that $\|C_{\frac{t}{T}} \Omega_{t,\omega}^{(T)}\|_p = O(\frac{1}{T})$ uniformly in t, ω . From (A1.3), we additionally have

$$\begin{aligned} \sum_{l=0}^t A_{t,l}^{(T)} \Omega_{t-l,\omega}^{(T)} &= \sum_{l=0}^t \sum_{j=0}^m A_{t,l-j}^{(T)} C_{\frac{t-l+j}{T}}^{-1} B_{\frac{t-l+j}{T},j} e^{i\omega(t-l)} [\mathcal{A}_{\frac{t-l}{T},\omega}^{(T)} - \mathcal{A}_{\frac{t-l}{T},\omega}] \\ &= e^{i\omega t} [\mathcal{A}_{t,\omega}^{(T)} - \mathcal{A}_{\frac{t}{T},\omega}]. \end{aligned}$$

Since the moving average operators are either in $S_2(H)$ or in $S_{\infty}(H)$, the above together with another application of Hölder's inequality for operators yields

$$\sup_{t,\omega} \|\mathcal{A}_{t,\omega}^{(T)} - \mathcal{A}_{\frac{t}{T},\omega}\|_p \leq \sup_{t,\omega} \left(\|A_{t,0}^{(T)}\|_{\infty} \|\Omega_{t,\omega}^{(T)}\|_p + \sum_{l=1}^t \|A_{t,l}^{(T)}\|_p \|\Omega_{t-l,\omega}^{(T)}\|_p \right) \leq \frac{K}{T},$$

for some constant K independent of T . We remark that the state space representation of the previous section allow similarly to derive that (A1) and (A2) still hold for $p = \infty$ under the weaker assumption of Lemma 3.2. \square

A1.3. Proofs of Section 4

Proof of Proposition 4.1. For fixed t and T , we have by Minkowski's inequality

$$\sum_s \|\text{cum}(X_{[uT-s/2],T}, X_{[uT+s/2],T})\|_2$$

$$\begin{aligned}
&= \sum_s \left\| \frac{1}{2\pi} \int_{\Pi} (\mathcal{A}_{[uT-s/2], \lambda_1}^{(T)} \otimes \mathcal{A}_{[uT+s/2], -\lambda_1}^{(T)}) \mathcal{F}_{\lambda_1}^{\varepsilon} e^{i\lambda_1 s} d\lambda_1 \right\|_2 \\
&= \sum_{s: \{(1 \leq [uT-s/2] \leq T) \cup (1 \leq [uT+s/2] \leq T)\}} \|\mathcal{C}_{u,s}^{(T)}\|_2 + \sum_{s: \{(1 \leq [uT-s/2] \leq T) \cup (1 \leq [uT+s/2] \leq T)\}^c} \|\mathcal{C}_{u,s}^{(T)}\|_2
\end{aligned}$$

where $\{\cdot\}^c$ denotes the complement event. Now since $\mathcal{A}_{t,\omega}^{(T)} = \mathcal{A}_{0,\omega}$ for $t < 1$ and $\mathcal{A}_{t,\omega}^{(T)} = \mathcal{A}_{0,\omega}$ for $t > T$, we can write

$$\begin{aligned}
&= \sum_{s:B} \left\| \frac{1}{2\pi} \int_{\Pi} (\mathcal{A}_{[uT-s/2], \lambda_1}^{(T)} \otimes \mathcal{A}_{[uT+s/2], -\lambda_1}^{(T)}) \mathcal{F}_{\lambda_1}^{\varepsilon} e^{i\lambda_1 s} d\lambda_1 \right\|_2 \\
&+ \sum_{s:B^c} \left\| \frac{1}{2\pi} \int_{\Pi} (\mathcal{A}_{0,\lambda_1} \otimes \mathcal{A}_{1,-\lambda_1}) \mathcal{F}_{\lambda_1}^{\varepsilon} e^{i\lambda_1 s} d\lambda_1 \right\|_2, \tag{A1.6}
\end{aligned}$$

where $B = \{(1 \leq [uT-s/2] \leq T) \cup (1 \leq [uT+s/2] \leq T)\}$. Because the first sum is finite, an application of proposition B1.3 implies it can be bounded by

$$K \sup_{t,T,\omega} \|\mathcal{A}_{t,\omega}^{(T)}\|_{\infty}^2 \|\mathcal{F}^{\varepsilon}\|_2 < \infty,$$

for some constant K . For the second term, we note that

$$\frac{1}{2\pi} \int_{\Pi} (\mathcal{A}_{0,\lambda_1} \otimes \mathcal{A}_{1,-\lambda_1}) \mathcal{F}_{\lambda_1}^{\varepsilon} e^{i\lambda_1 s} d\lambda_1 = \text{cum}(X_{t+s}^{(0)}, X_t^{(1)}).$$

It thus corresponds to the cross-covariance operator of the two stationary processes $X_t^{(0)}$ and $X_t^{(1)}$ at lag s . By Proposition B2.4, we can alternatively express this as

$$\text{cum}(X_s^{(0)}, X_0^{(1)}) = \sum_{l,k} (A_{0,l} \otimes A_{1,k}) \text{cum}(\varepsilon_{t+s-l}, \varepsilon_{t-k}).$$

Using then that ε_t is functional white noise, we find for the second term in (A1.6)

$$\begin{aligned}
\sum_{s:B^c} \|\mathcal{C}_{u,s}^{(T)}\|_2 &\leq \left\| \sum_{l,k \in \mathbb{Z}} (A_{0,l} \otimes A_{1,k}) \text{cum}(\varepsilon_0, \varepsilon_0) \right\|_2 \\
&\leq \sum_{l \in \mathbb{Z}} \|A_{0,l}\|_{\infty} \sum_{k \in \mathbb{Z}} \|A_{1,k}\|_{\infty} \|\text{cum}(\varepsilon_0, \varepsilon_0)\|_2 < \infty.
\end{aligned}$$

The result now follows. \square

A1.4. Proofs of Section 5

Proof of Proposition 5.1. We have by Theorem B2.2 and by Proposition B2.1,

$$\begin{aligned}
&\text{cum} \left(\int_{\Pi} e^{i\lambda_1 r_1} A_{t_{r_1}, \lambda_1}^{(T)} dZ_{\lambda_1}, \dots, \int_{\Pi} e^{i\lambda_k r_k} A_{t_{r_k}, \lambda_k}^{(T)} dZ_{\lambda_k} \right) \\
&= \int_{\Pi} \dots \int_{\Pi} \text{cum} (e^{i\lambda_1 r_1} A_{t_{r_1}, \lambda_1}^{(T)} dZ_{\lambda_1}, \dots, e^{i\lambda_k r_k} A_{t_{r_k}, \lambda_k}^{(T)} dZ_{\lambda_k})
\end{aligned}$$

$$\begin{aligned}
 &= \int_{\Pi} \cdots \int_{\Pi} (e^{i\lambda_1 r_1} \mathcal{A}_{t_{r_1}, \lambda_1}^{(T)} \otimes \cdots \otimes e^{i\lambda_k r_k} \mathcal{A}_{t_{r_k}, \lambda_k}^{(T)}) \text{cum}(dZ_{\lambda_1}, \dots, dZ_{\lambda_k}) \\
 &= \int_{\Pi^k} e^{i(\lambda_1 r_1 + \cdots + \lambda_k r_k)} (\mathcal{A}_{t_{r_1}, \lambda_1}^{(T)} \otimes \cdots \otimes \mathcal{A}_{t_{r_k}, \lambda_k}^{(T)}) \\
 &\quad \times \eta(\lambda_1 + \cdots + \lambda_k) \mathcal{F}_{\lambda_1, \dots, \lambda_{k-1}}^{\varepsilon} d\lambda_1 \cdots d\lambda_k,
 \end{aligned}$$

where the equality holds in the tensor product space $H_{\mathbb{C}} \otimes \cdots \otimes H_{\mathbb{C}}$. Note that the last line corresponds to the inversion formula of the cumulant tensor of order k . For fixed $t \in \{1, \dots, T\}$ and $T \in \mathbb{N}$, the k -th order cumulant spectral tensor of the linear functional process $\{X_{t,T}\}$ can thus be given by

$$\mathcal{F}_{\lambda_1, \dots, \lambda_{k-1}}^{(t,T)} = \left(\mathcal{A}_{t_{r_1}, \lambda_1}^{(T)} \otimes \cdots \otimes \mathcal{A}_{t_{r_{k-1}}, \alpha_{k-1}}^{(T)} \otimes \mathcal{A}_{t_{r_k}, -\lambda_+}^{(T)} \right) \mathcal{F}_{\lambda_1, \dots, \lambda_{k-1}}^{\varepsilon}, \quad (\text{A1.7})$$

and is well-defined in the tensor product space $\bigotimes_{i=1}^k H_{\mathbb{C}}$. In particular, Proposition B1.3 implies the corresponding operator is Hilbert-Schmidt for $k \geq 2$

$$\begin{aligned}
 \|\mathcal{F}_{\lambda_1, \dots, \lambda_{2k-1}}\|_2 &\leq \|\mathcal{A}_{t_{r_1}, \lambda_1}^{(T)} \otimes \cdots \otimes \mathcal{A}_{t_{r_{k-1}}, \alpha_{k-1}}^{(T)} \otimes \mathcal{A}_{t_{r_k}, -\lambda_+}^{(T)}\|_{\infty} \|\mathcal{F}_{\lambda_1, \dots, \lambda_{2k-1}}^{\varepsilon}\|_2 \\
 &= \left(\sup_{t_{r_j}, \lambda_j} \|\mathcal{A}_{t_{r_j}, \lambda_j}^{(T)}\|_{\infty} \right)^{2k} \|\mathcal{F}_{\lambda_1, \dots, \lambda_{2k-1}}^{\varepsilon}\|_2 < \infty.
 \end{aligned}$$

We therefore have that the kernel function $f_{\lambda_1, \dots, \lambda_{k-1}}^{(t,T)}(\tau_1, \dots, \tau_k)$ is a properly defined element in $L_{\mathbb{C}}^2([0, 1]^2)$. In case $k = 2$, we moreover have that $\mathcal{F}_{\lambda_1} \in S_1(H_{\mathbb{C}})$. This follows by the fact that the ε_t are white noise and thus $\|\mathcal{F}_{\lambda_1}^{\varepsilon}\|_1 \leq \sum_t \|\text{cum}(\varepsilon_t, \varepsilon_0)\|_1 = \|\mathcal{C}_0^{\varepsilon}\|_1 = \mathbb{E}\|\varepsilon_0\|_2^2 < \infty$. \square

Proof of Theorem 5.2. Under assumption (A1)–(A2) we have for all $t = 1, \dots, T$ and $T \in \mathbb{N}$ that $X_{t,T}$ are locally stationary random elements in H . Therefore, by Proposition 5.1 and (29),

$$\begin{aligned}
 \mathbb{E}(I_{u_j, \omega}^{(T)}) &= \frac{1}{2\pi H_{2,N}(0)} \text{cum}(D_{u_j, \omega}^{(T)}, D_{u_j, -\omega}^{(T)}) \\
 &= \frac{1}{2\pi H_{2,N}(0)} \int_{\Pi} \left(H_N(\mathcal{A}_{t_j, \bullet, \lambda}^{(T)}, \omega - \lambda) \otimes H_N(\mathcal{A}_{t_j, \bullet, -\lambda}^{(T)}, \lambda - \omega) \right) \mathcal{F}_{\lambda}^{\varepsilon} d\lambda.
 \end{aligned}$$

In order to replace the transfer function kernels with their continuous approximations, we write

$$\begin{aligned}
 &\mathcal{A}_{t_j, r, \lambda}^{(T)} \otimes \mathcal{A}_{t_j, s, -\lambda}^{(T)} - \mathcal{A}_{u_j, r, \omega} \otimes \mathcal{A}_{u_j, s, -\omega} \\
 &= (\mathcal{A}_{t_j, r, \lambda}^{(T)} - \mathcal{A}_{u_j, r, \omega}) \otimes \mathcal{A}_{t_j, s, -\lambda}^{(T)} + \mathcal{A}_{u_j, r, \omega} \otimes (\mathcal{A}_{t_j, s, -\lambda}^{(T)} - \mathcal{A}_{u_j, s, -\omega}).
 \end{aligned} \quad (\text{A1.8})$$

We focus on finding a bound on the first term as the second term can be bounded similarly. Since $H_N(\cdot, \cdot)$ is linear in its first argument, we have by the triangle inequality

$$\begin{aligned} & \left\| H_N(\mathcal{A}_{t_j, r, \lambda}^{(T)} - \mathcal{A}_{u_j, r, \omega}, \omega - \lambda) \right\|_\infty \\ & \leq \left\| H_N(\mathcal{A}_{t_j, \bullet, \lambda}^{(T)} - \mathcal{A}_{u_j, \bullet, \lambda}, \omega - \lambda) \right\|_\infty + \left\| H_N(\mathcal{A}_{u_j, \bullet, \lambda} - \mathcal{A}_{u_j, \bullet, \omega}, \omega - \lambda) \right\|_\infty. \end{aligned}$$

For the first term of this expression, assumption (A2) and Lemma B3.3 imply

$$\left\| \sum_{r=0}^{N-1} h_{r, N}(\mathcal{A}_{t_j, r, \lambda}^{(T)} - \mathcal{A}_{u_j, r, \lambda}) e^{-ir(\omega - \lambda)} \right\|_\infty \leq C \frac{N}{T} \quad (\text{A1.9})$$

for some generic constant C independent of T . Next, we consider the second term. Similarly as in the proof of Lemma B3.3, we have

$$\begin{aligned} & H_N(\mathcal{A}_{u_j, \bullet, \lambda} - \mathcal{A}_{u_j, \bullet, \omega}, \omega - \lambda) \\ & = H_N(\omega - \lambda) (\mathcal{A}_{u_j, \lambda} - \mathcal{A}_{u_j, \omega}) + H_N(\omega - \lambda) (\mathcal{A}_{u_j, N-1, \lambda} - \mathcal{A}_{u_j, N-1, \omega}) \\ & \quad - \sum_{r=0}^{N-1} \left[(\mathcal{A}_{u_j, r, \lambda} - \mathcal{A}_{u_j, r-1, \lambda}) - (\mathcal{A}_{u_j, r, \omega} - \mathcal{A}_{u_j, r-1, \omega}) \right] H_s(\omega - \lambda). \end{aligned}$$

Since the transfer function operator is twice continuously differentiable in u and ω , we find by two applications of the mean value theorem

$$\begin{aligned} & \left\| (\mathcal{A}_{u_j, r, \lambda} - \mathcal{A}_{u_j, r-1, \lambda}) - (\mathcal{A}_{u_j, r, \omega} - \mathcal{A}_{u_j, r-1, \omega}) \right\|_\infty \\ & \leq \sup_{u \in [0, 1], \omega \in \Pi} \left\| \frac{\partial^2 \mathcal{A}_{u, \omega}}{\partial u \partial \omega} \right\| \frac{|\lambda - \omega|}{T}. \end{aligned}$$

Hence we obtain the upper bound

$$\left\| H_N(\mathcal{A}_{u_j, \bullet, \lambda} - \mathcal{A}_{u_j, \bullet, \omega}, \omega - \lambda) \right\|_\infty \leq C L_N(\omega - \lambda) |\omega - \lambda| + C \frac{N}{T} L_N(\omega - \lambda) |\omega - \lambda|.$$

Moreover, Lemma B3.3 implies

$$\left\| H_N(\mathcal{A}_{t_j, \bullet, -\lambda}^{(T)}, \omega - \lambda) \right\|_\infty \leq C L_N(\omega - \lambda).$$

With these bounds and Proposition B1.3 and Lemma B3.1, we now obtain

$$\begin{aligned} & \int_\Pi \left\| \left(H_N(\mathcal{A}_{u_j, \bullet, \lambda} - \mathcal{A}_{u_j, \bullet, \omega}, \omega - \lambda) \otimes H_N(\mathcal{A}_{t_j, \bullet, -\lambda}^{(T)}, \omega - \lambda) \right) \mathcal{F}_\lambda^\varepsilon \right\|_2 d\lambda \\ & \leq \int_\Pi \left\| H_N(\mathcal{A}_{u_j, \bullet, \lambda} - \mathcal{A}_{u_j, \bullet, \omega}, \omega - \lambda) \right\|_\infty \left\| H_N(\mathcal{A}_{t_j, \bullet, -\lambda}^{(T)}, \omega - \lambda) \right\|_\infty \left\| \mathcal{F}_\lambda^\varepsilon \right\|_2 d\lambda \\ & \leq C \int_\Pi L_N(\omega - \lambda)^2 d\lambda \leq C \log(N). \end{aligned}$$

The second term of (A1.8) is similar and thus the error from replacing $\mathcal{A}_{t_j, r, \lambda}^{(T)}$ and $\mathcal{A}_{t_j, s, -\lambda}^{(T)}$ by $\mathcal{A}_{u_j, r, \omega}$ and $\mathcal{A}_{u_j, s, -\omega}$, respectively, is of order $O\left(\frac{\log(N)}{N}\right)$ in L^2 .

The expectation of the periodogram tensor can therefore be written as

$$\begin{aligned}
 \mathbb{E}(I_{u_j, \omega}^{(T)}) &= \frac{1}{2\pi H_{2,N}(0)} \int_{\Pi} \left(H_N(\mathcal{A}_{t_j, \bullet, \lambda}^{(T)}, \omega - \lambda) \otimes H_N(\mathcal{A}_{t_j, \bullet, -\lambda}^{(T)}, \lambda - \omega) \right) \mathcal{F}_{\lambda}^{\varepsilon} d\lambda \\
 &= \frac{1}{2\pi H_{2,N}(0)} \int_{\Pi} \left(H_N(\mathcal{A}_{u_j, \bullet, \omega}^{(T)}, \omega - \lambda) \otimes H_N(\mathcal{A}_{u_j, \bullet, -\omega}^{(T)}, \lambda - \omega) \right) \mathcal{F}_{\lambda}^{\varepsilon} d\lambda + R_T \\
 &= \frac{1}{H_{2,N}(0)} H_{2,N}(\mathcal{A}_{u_j, \bullet, \omega} \otimes \mathcal{A}_{u_j, \bullet, -\omega}, 0) \mathcal{F}_{\omega}^{\varepsilon} + R_T \\
 &= \frac{1}{H_{2,N}(0)} H_{2,N}(\mathcal{F}_{u_j, \bullet, \omega}, 0) + R_T
 \end{aligned}$$

where the remainder term R_T is of order $O\left(\frac{\log(N)}{N}\right)$. Correspondingly, the local periodogram kernel is given by

$$\mathbb{E}(I_{u_j, \omega}^{(T)}(\tau, \sigma)) = \frac{1}{H_{2,N}(0)} \sum_{r=1}^N h_{r,N}^2 f_{u_j, r, \omega}(\tau, \sigma) + O\left(\frac{\log(N)}{N}\right).$$

Since by the conditions of the theorem, the operator-valued function $\mathcal{A}_{u, \omega}$ is twice continuously differentiable with respect to u , Theorem B1.8 implies that the spectral density operator $\mathcal{F}_{u, \omega}$ is also twice continuously differentiable in $u \in (0, 1)$. Hence, by a Taylor approximation of $\mathcal{F}_{u_j, r, \omega}$ about u_j , we find for the mean of the periodogram tensor

$$\mathbb{E}(I_{u_j, \omega}^{(T)}) = \mathcal{F}_{u_j, \omega} + \frac{1}{2} b_{t, T}^2 \kappa_t \frac{\partial^2 \mathcal{F}_{u, \omega}}{\partial u^2} \Big|_{u=u_j} + O\left(\frac{\log(N)}{N}\right),$$

where we have used the definition in (32) of the smoothing kernel K_t in time direction. As by the assumption on the taper function this kernel is symmetric about zero, the first order term in the Taylor approximation is zero.

This proves the first part of from Theorem 5.2. For the covariance, we note that the product theorem for cumulants (see section B2) and the fact that the means are zero imply

$$\begin{aligned}
 \text{cov}(I_{u_j, \omega_1}^{(T)}, I_{u_j, \omega_2}^{(T)}) &= \frac{1}{4\pi^2 H_{2,N}(0)^2} \left[\text{cum}(D_{u_j, \omega_1}^{(T)}, D_{u_j, -\omega_1}^{(T)}, D_{u_j, -\omega_2}^{(T)}, D_{u_j, \omega_2}^{(T)}) \right. \\
 &\quad + S_{1423} \left(\text{cum}(D_{u_j, \omega_1}^{(T)}, D_{u_j, \omega_2}^{(T)}) \otimes \text{cum}(D_{u_j, -\omega_1}^{(T)}, D_{u_j, -\omega_2}^{(T)}) \right) \\
 &\quad \left. + S_{1324} \left(\text{cum}(D_{u_j, \omega_1}^{(T)}, D_{u_j, -\omega_2}^{(T)}) \otimes \text{cum}(D_{u_j, -\omega_1}^{(T)}, D_{u_j, \omega_2}^{(T)}) \right) \right], \tag{A1.10}
 \end{aligned}$$

where S_{ijkl} denotes the permutation operator on $\otimes_{i=1}^4 L_{\mathbb{C}}^2([0, 1])$ that permutes the components of a tensor according to the permutation $(1, 2, 3, 4) \mapsto (i, j, k, l)$, that is, $S_{ijkl}(x_1 \otimes \cdots \otimes x_4) = x_i \otimes \cdots \otimes x_l$.

We first show that the first term of this expression is of lower order than the other two. By (29), the cumulant is equal to

$$\int_{\Pi^4} H_N(\mathcal{A}_{t_j, \cdot, \lambda_1}^{(T)}, \omega_1 - \lambda_1) \otimes H_N(\mathcal{A}_{t_j, \cdot, \lambda_2}^{(T)}, -\omega_1 - \lambda_2) \otimes H_N(\mathcal{A}_{t_j, \cdot, \lambda_3}^{(T)}, -\omega_2 - \lambda_3) \\ \otimes H_N(\mathcal{A}_{t_j, \cdot, \lambda_4}^{(T)}, \omega_2 - \lambda_4) \eta(\lambda_1 + \dots + \lambda_4) \mathcal{F}_{\lambda_1, \lambda_2, \lambda_3}^\varepsilon d\lambda_1 \cdots d\lambda_4$$

and hence, by Lemma B3.3, is bounded in L^2 -norm by

$$C \int_{\Pi^3} L_N(\omega_1 - \lambda_1) L_N(-\omega_1 - \lambda_2) L_N(-\omega_2 - \lambda_3) L_N(\lambda_1 + \lambda_2 + \lambda_3 + \omega_2) d\lambda_1 d\lambda_2 d\lambda_3 \\ \leq C \log(N)^2 \int_{\Pi^3} L_N(\omega_2 + \lambda_3)^2 d\lambda_3 \leq C N \log(N)^2.$$

Next we consider the second term of (A1.10). A similar derivation as for the expectation of the periodogram tensor shows that the term equals

$$\int_{\Pi^2} H_N(\mathcal{A}_{t_j, \bullet, \lambda_1}^{(T)}, \omega_1 - \lambda_1) \otimes H_N(\mathcal{A}_{t_j, \bullet, -\lambda_1}^{(T)}, \omega_2 + \lambda_1) \otimes H_N(\mathcal{A}_{t_j, \bullet, \lambda_2}^{(T)}, -\omega_1 - \lambda_2) \\ \otimes H_N(\mathcal{A}_{t_j, \bullet, -\lambda_2}^{(T)}, \lambda_2 - \omega_2) \mathcal{F}_{\lambda_1}^\varepsilon \otimes \mathcal{F}_{\lambda_2}^\varepsilon d\lambda_1 d\lambda_2 \\ = \int_{\Pi^2} H_N(\mathcal{A}_{u_j, \bullet, \lambda_1}, \omega_1 - \lambda_1) \otimes H_N(\mathcal{A}_{u_j, \bullet, -\lambda_1}, \omega_2 + \lambda_1) \otimes H_N(\mathcal{A}_{u_j, \bullet, \lambda_2}, -\omega_1 - \lambda_2) \\ \otimes H_N(\mathcal{A}_{u_j, \bullet, -\lambda_2}, \lambda_2 - \omega_2) \mathcal{F}_{\lambda_1}^\varepsilon \otimes \mathcal{F}_{\lambda_2}^\varepsilon d\lambda_1 d\lambda_2 + R_T \\ = H_{2,N}(\mathcal{A}_{u_j, \bullet, \omega_1} \otimes \mathcal{A}_{u_j, \bullet, -\omega_1}, \omega_1 - \omega_2) \otimes H_{2,N}(\mathcal{A}_{u_j, \bullet, \omega_1} \otimes \mathcal{A}_{u_j, \bullet, -\omega_1}, \omega_2 - \omega_1) \\ \times \mathcal{F}_{\omega_1}^\varepsilon \otimes \mathcal{F}_{-\omega_1}^\varepsilon \\ = H_{2,N}(\mathcal{F}_{u_j, \bullet, \omega_1}, \omega_1 - \omega_2) \otimes H_{2,N}(\mathcal{F}_{u_j, \bullet, -\omega_1}, \omega_2 - \omega_1).$$

Proceeding in an analogous matter for the third term of (A1.10), we obtain the stated result. \square

Proof of Theorem 5.3. By Theorem 5.2, the expectation of the periodogram tensor can be written as

$$\mathbb{E}(I_{u_j, \omega}^{(T)}) = H_{2,N}(\mathcal{F}_{u_j, \bullet, \omega}, 0) + R_T = \frac{1}{H_{2,N}(0)} \sum_{r=1}^N h_{r,N}^2 \mathcal{F}_{u_j, r, \omega} + O\left(\frac{\log(N)}{N}\right),$$

where the remainder term R_T is of order $O\left(\frac{\log(N)}{N}\right)$. Because the operator-valued function $\mathcal{A}_{u, \omega}$ is twice differentiable with respect to both u and ω , it follows from Theorem B1.8 that the tensor $\mathcal{F}_{u, \omega}$ is twice continuously differentiable in both u , and ω . We can therefore apply a Taylor expansion of $\mathcal{F}_{u_j, r, \omega}$ about to the point $x = (u_j, \omega_o)$ to obtain

$$\mathcal{F}_{u_j, r, \omega} = \mathcal{F}_{u_j, \omega_o} + \left(\frac{r - N/2}{T}\right) \frac{\partial}{\partial u} \mathcal{F}_{u, \omega} \Big|_{(u, \omega)=x} + (\omega - \lambda) \frac{\partial}{\partial \omega} \mathcal{F}_{u, \omega} \Big|_{(u, \omega)=x} \\ = \frac{1}{2} \left(\frac{r - N/2}{T}\right)^2 \frac{\partial^2}{\partial u^2} \mathcal{F}_{u, \omega} \Big|_{(u, \omega)=x} + \frac{1}{2} (\omega - \lambda)^2 \frac{\partial^2}{\partial \omega^2} \mathcal{F}_{u, \omega} \Big|_{(u, \omega)=x} \\ + \left(\frac{r - N/2}{T}\right) (\omega - \lambda) \left(\frac{\partial^2}{\partial u \partial \omega} \mathcal{F}_{u, \omega} \Big|_{(u, \omega)=x} + \frac{\partial^2}{\partial \omega \partial u} \mathcal{F}_{u, \omega} \Big|_{(u, \omega)=x} \right) + R_{T,p}, \quad (\text{A1.11})$$

where the remainder can generally be bounded by

$$\begin{aligned} R_{T,p} &= \sum_{i_1, i_2 \in \mathbb{N}: i_1 + i_2 > p} \frac{b_{t,T}^{i_1} |\omega - \alpha|^{i_2}}{(i_1)! (i_2)!} \sup_{u, \omega} \left\| \frac{\partial^{i_1 + i_2}}{\partial u^{i_1} \partial \omega^{i_2}} f_{u, \omega} \right\|_2 \\ &= \sum_{i_1, i_2 \in \mathbb{N}: i_1 + i_2 = p} o(b_{t,T}^{i_1} |\omega - \alpha|^{i_2}) \quad p \geq 2. \end{aligned} \quad (\text{A1.12})$$

In order to derive the mean of the estimator, we set $vb_{t,T} = \frac{r-N/2}{T}$ and recall that the taper function relates to a smoothing kernel K_t in time direction by

$$K_t(v) = \frac{1}{H_2} h^2 \left(\frac{v+1}{2} \right) \quad (\text{A1.13})$$

for $v \in [-\frac{1}{2}, \frac{1}{2}]$ with bandwidth $b_{t,T} = N/T$. It then follows from (A1.11) that a Taylor expansion about to the point $x = (u_j, \omega_o)$ yields

$$\begin{aligned} \mathbb{E}(\hat{\mathcal{F}}_{u_j, \omega_o}^{(T)}) &= \mathcal{F}_{u_j, \omega_o} + \sum_{i=1}^2 \frac{1}{i!} b_{t,T}^i \int v^i K_t(v) dv \int_{\Pi} K_f(\alpha) d\alpha \frac{\partial^i \mathcal{F}_{u, \omega}}{\partial u^i} \Big|_{(u, \omega) = x} \\ &\quad + \sum_{i=1}^2 \frac{1}{i!} b_{t,T}^i \int_{\Pi} \alpha^i K_f(\alpha) d\alpha \int K_t(v) dv \frac{\partial^i \mathcal{F}_{u, \omega}}{\partial \omega^i} \Big|_{(u, \omega) = x} \\ &\quad + \frac{1}{2} b_{t,T} b_{f,T} \int v K_t(v) dv \int_{\Pi} \alpha K_f(\alpha) d\alpha \left[\frac{\partial^2 \mathcal{F}_{u, \omega}}{\partial u \partial \omega} + \frac{\partial^2 \mathcal{F}_{u, \omega}}{\partial \omega \partial u} \right] \Big|_{(u, \omega) = x} + R_{T,p}. \end{aligned}$$

Because the smoothing kernels are symmetric around 0, we obtain

$$\begin{aligned} \mathbb{E}(\hat{\mathcal{F}}_{u_j, \omega_o}^{(T)}) &= \mathcal{F}_{u_j, \omega_o} + \frac{1}{2} b_{t,T}^2 \kappa_t \frac{\partial^2}{\partial u^2} \mathcal{F}_{u, \omega} \Big|_{(u, \omega) = x} + \frac{1}{2} b_{f,T}^2 \kappa_f 2 \cdot 1 \frac{\partial^2}{\partial \omega^2} \mathcal{F}_{u, \omega} \Big|_{(u, \omega) = x} \\ &\quad + o(b_{t,T}^2) + o(b_{f,T}^2) + O\left(\frac{\log(b_{t,T} T)}{b_{t,T} T}\right), \end{aligned} \quad (\text{A1.14})$$

where the error terms follow from (A1.12) and Theorem 5.2, respectively. This establishes Result *i*) of Theorem 5.3.

For the proof of the covariance structure, we note that

$$\text{cov}(\hat{\mathcal{F}}_{u, \omega_1}^{(T)}, \hat{\mathcal{F}}_{u, \omega_2}^{(T)}) = \int_{\Pi^2} K_{f,T}(\omega_1 - \lambda_1) K_{f,T}(\omega_2 - \lambda_2) \text{cov}(I_{u, \lambda_1}^{(T)}, I_{u, \lambda_2}^{(T)}) d\lambda_1 d\lambda_2,$$

where by Theorem 5.2

$$\begin{aligned} &\text{cov}(I_{u, \lambda_1}^{(T)}, I_{u, \lambda_2}^{(T)}) \\ &= \frac{1}{4\pi^2 H_{4,N}(0)^2} \left[S_{1423} \left(H_{2,N}(\mathcal{F}_{u_\bullet, \lambda_1}, \lambda_1 - \lambda_2) \otimes H_{2,N}(\mathcal{F}_{u_\bullet, -\lambda_1}, \lambda_2 - \lambda_1) \right) \right. \\ &\quad \left. + S_{1324} \left(H_{2,N}(\mathcal{F}_{u_\bullet, \lambda_1}, \lambda_1 + \lambda_2) \otimes H_{2,N}(\mathcal{F}_{u_\bullet, -\lambda_1}, -\lambda_1 - \lambda_2) \right) \right] \\ &\quad + O\left(\frac{\log(N)}{N}\right). \end{aligned}$$

We treat the two terms of the covariance tensor separately. Starting with the first term, we have

$$\begin{aligned}
& \left\| \int_{\Pi^2} K_{f,T}(\omega_1 - \lambda_1) \left[K_{f,T}(\omega_2 - \lambda_2) \right. \right. \\
& \quad \times \left. \left. \left[H_{2,N}(F_{u_\bullet, \lambda_1}, \lambda_1 - \lambda_2) \otimes H_{2,N}(\mathcal{F}_{u_\bullet, -\lambda_1}, \lambda_2 - \lambda_1) \right] \right. \right. \\
& \quad \left. \left. - K_{f,T}(\omega_2 - \lambda_1) |H_{2,N}(\lambda_1 - \lambda_2)|^2 (\mathcal{F}_{u, \lambda_1} \otimes \mathcal{F}_{u, -\lambda_1}) \right] d\lambda_1 d\lambda_2 \right\|_2 \\
& \leq \left\| \int_{\Pi^2} K_{f,T}(\omega_1 - \lambda_1) K_{f,T}(\omega_2 - \lambda_2) \right. \\
& \quad \times \left[H_{2,N}(F_{u_\bullet, \lambda_1}, \lambda_1 - \lambda_2) \otimes H_{2,N}(\mathcal{F}_{u_\bullet, -\lambda_1}, \lambda_2 - \lambda_1) \right. \\
& \quad \left. - |H_{2,N}(\lambda_1 - \lambda_2)|^2 (\mathcal{F}_{u, \lambda_1} \otimes \mathcal{F}_{u, -\lambda_1}) \right] d\lambda_1 d\lambda_2 \Big\|_2 \\
& \quad + \left\| \int_{\Pi^2} K_{f,T}(\omega_1 - \lambda_1) \left[K_{f,T}(\omega_2 - \lambda_2) - K_{f,T}(\omega_2 - \lambda_1) \right] \right. \\
& \quad \left. \times |H_{2,N}(\lambda_1 - \lambda_2)|^2 (\mathcal{F}_{u, \lambda_1} \otimes \mathcal{F}_{u, -\lambda_1}) d\lambda_1 d\lambda_2 \right\|_2.
\end{aligned}$$

Since $\mathcal{F}_{u, \lambda}$ is uniformly Lipschitz continuous in u , we have $\|\mathcal{F}_{u_r, \lambda} - \mathcal{F}_{u, \lambda}\|_2 \leq C \frac{N}{T}$ and hence the first term on the right hand side is bounded by

$$C \int_{\Pi^2} b_{f,T}^2 L_{\frac{1}{b_{f,T}}}(\omega_1 - \lambda_1)^2 L_{\frac{1}{b_{f,T}}}(\omega_2 - \lambda_2)^2 L_N(\lambda_1 - \lambda_2)^2 \frac{N}{T} d\lambda_1 d\lambda_2 \leq C \frac{N^2}{b_{f,T} T}.$$

For the second term, we exploit uniform Lipschitz continuity of the kernel function K_f to get the upper bound

$$C \int_{\Pi^2} K_{f,T}(\omega_1 - \lambda_1)^2 b_{f,T}^{-2} |\lambda_1 - \lambda_2| L_N(\lambda_1 - \lambda_2)^2 d\lambda_1 d\lambda_2 \leq C \frac{\log(N)}{b_{f,T}^2}.$$

In total we obtain

$$\|\text{cov}(\hat{\mathcal{F}}_{u, \omega_1}^{(T)}, \hat{\mathcal{F}}_{u, \omega_2}^{(T)})\|_2 = O\left(\frac{\log(N)}{b_{f,T}^2 N^2}\right) + O\left(\frac{1}{b_{f,T} T}\right) + O\left(\frac{\log(N)}{N}\right)$$

uniformly in $\omega_1, \omega_2 \in [-\pi, \pi]$ and $u \in [0, 1]$. \square

Proof of Proposition 5.4. A change of variables shows that (36) can be written as

$$\begin{aligned}
& b_{t,T} b_{f,T} T \text{cov}(\langle \hat{\mathcal{F}}_{u, \omega_1}^{(T)}, g_1 \otimes g_2 \rangle_{H_{\mathbb{C}} \otimes H_{\mathbb{C}}}, \langle \hat{\mathcal{F}}_{u, \omega_2}^{(T)}, g_3 \otimes g_4 \rangle_{H_{\mathbb{C}} \otimes H_{\mathbb{C}}}) \\
& = 2\pi b_{f,T} \|K_t\|_2^2 \int_{\Pi} K_{f,T}(\omega_1 - \omega_2 - \lambda) K_{f,T}(\lambda) \\
& \quad \times \langle \mathcal{F}_{u, \omega_2 - \lambda} g_3, g_1 \rangle \langle \mathcal{F}_{u, -\omega_2 - \lambda} g_4, g_2 \rangle d\lambda \\
& \quad + 2\pi b_{f,T} \|K_t\|_2^2 \int_{\Pi} K_{f,T}(\omega_1 + \omega_2 - \lambda) K_{f,T}(\lambda)
\end{aligned}$$

$$\begin{aligned} & \times \langle \mathcal{F}_{u, -\omega_2 + \lambda} g_4, g_1 \rangle \langle \mathcal{F}_{u, \omega_2 - \lambda} g_3, g_2 \rangle d\lambda \\ & + O(b_{f,T} \log(b_{t,T} T)) + O(b_{t,T}^2 b_{f,T}) + O((b_{t,T} b_{f,T} T)^{-1}). \end{aligned} \quad (\text{A1.15})$$

The error terms will tend to zero under assumption (A8). Since the product of the two kernels in the first integral is exactly zero whenever $|\lambda - (\omega_1 - \omega_2)| > b_{f,T}$ or $\lambda > b_{f,T}$, the first integral vanishes for large enough T unless $\omega_1 = \omega_2$. For $\omega_1 = \omega_2$, the integral in the first term becomes

$$\int_{\Pi} K_{f,T}(-\lambda) K_{f,T}(\lambda) \langle \mathcal{F}_{u, \omega_1 + \lambda} g_3, g_1 \rangle \langle \mathcal{F}_{u, -\omega_1 - \lambda} g_4, g_2 \rangle d\lambda$$

and further by symmetry of the kernel

$$= \int_{\Pi} K_{f,T}(\lambda)^2 \langle \mathcal{F}_{u, \omega_1 + \lambda} g_3, g_1 \rangle \langle \mathcal{F}_{u, -\omega_1 - \lambda} g_4, g_2 \rangle d\lambda.$$

We note that $\|K_f\|_2^{-2} K_{f,T}(\lambda)^2$ satisfies the properties of an approximate identity (e.g., Edwards, 1967). Hence application of Lemma F.15 of Panaretos and Tavakoli (2013b), which covers approximate identities in a functional setting, yields that the integral converges to

$$\|K_f\|_2^2 \langle \mathcal{F}_{u, \omega_1} g_3, g_1 \rangle \langle \mathcal{F}_{u, -\omega_1} g_4, g_2 \rangle,$$

with respect to $\|\cdot\|_2$. Since the integral in the second term in (A1.15) vanishes unless $\omega_1 = -\omega_2$, we can apply a similar argument, which proves the proposition. \square

Proof of Theorem 5.7. We decompose the difference in terms of its variance and its squared bias. That is,

$$\begin{aligned} & \int_{\Pi} \mathbb{E} \|\hat{\mathcal{F}}_{u,\omega}^{(T)} - \mathbb{E} \hat{\mathcal{F}}_{u,\omega}^{(T)} + \mathbb{E} \hat{\mathcal{F}}_{u,\omega}^{(T)} - \mathcal{F}_{u,\omega}\|_2^2 d\omega \\ & = \int_{\Pi} \mathbb{E} \|\hat{\mathcal{F}}_{u,\omega}^{(T)} - \mathbb{E} \hat{\mathcal{F}}_{u,\omega}^{(T)}\|_2^2 d\omega + \int_{\Pi} \mathbb{E} \|\mathbb{E} \hat{\mathcal{F}}_{u,\omega}^{(T)} - \mathcal{F}_{u,\omega}\|_2^2 d\omega. \end{aligned} \quad (\text{A1.16})$$

The cross term cancels which is easily seen by noting that $\mathbb{E}(\hat{\mathcal{F}}_{u,\omega}^{(T)} - \mathbb{E}(\hat{\mathcal{F}}_{u,\omega}^{(T)})) = O_{H_{\mathbb{C}}}$ and hence

$$\mathbb{E} \left(\langle \hat{\mathcal{F}}_{u,\omega}^{(T)} - \mathbb{E}(\hat{\mathcal{F}}_{u,\omega}^{(T)}), \mathbb{E}(\hat{\mathcal{F}}_{u,\omega}^{(T)} - \mathcal{F}_{u,\omega}) \rangle_{H_{\mathbb{C}} \otimes H_{\mathbb{C}}} \right) = 0$$

for all $u \in [0, 1]$ and $\omega \in [-\pi, \pi]$. Consider the first term of (A1.16). Self-adjointness of $\hat{\mathcal{F}}_{u,\omega}^{(T)}$ and $\mathbb{E} \|X_{t,T}\|_2^4 < \infty$ imply that

$$\begin{aligned} & \text{tr}(\text{cov}(\hat{\mathcal{F}}_{u,\omega}^{(T)}, \hat{\mathcal{F}}_{u,\omega}^{(T)})) \\ & = \sum_{n,m=1}^{\infty} \langle \mathbb{E}[(\hat{\mathcal{F}}_{u,\omega}^{(T)} - \mathbb{E}(\hat{\mathcal{F}}_{u,\omega}^{(T)})) \otimes (\hat{\mathcal{F}}_{u,\omega}^{(T)} - \mathbb{E}(\hat{\mathcal{F}}_{u,\omega}^{(T)})) \psi_{nm}, \psi_{nm} \rangle \end{aligned}$$

$$= \mathbb{E} \sum_{n,m=1}^{\infty} |\langle (\hat{\mathcal{F}}_{u,\omega}^{(T)} - \mathbb{E}(\hat{\mathcal{F}}_{u,\omega}^{(T)})), \psi_{nm} \rangle_2|^2 = \mathbb{E} \|\hat{\mathcal{F}}_{u,\omega}^{(T)} - \mathbb{E}\hat{\mathcal{F}}_{u,\omega}^{(T)}\|_2^2 < \infty$$

for some orthonormal basis $\{\psi_{nm}\}$ of $L_{\mathbb{C}}^2([0, 1]^2)$. By Fubini's theorem and Corollary 5.5, we thus find

$$\int_{\Pi} \mathbb{E} \|\hat{\mathcal{F}}_{u,\omega}^{(T)} - \mathbb{E}\hat{\mathcal{F}}_{u,\omega}^{(T)}\|_2^2 d\omega = \int_{\Pi} \int_{[0,1]^2} \text{var}(\hat{f}_{u,\omega}^{(T)}(\tau, \sigma)) d\tau d\sigma d\omega = O\left(\frac{1}{b_t b_f T}\right).$$

Theorem 5.3 then yields for the second term of (A1.16)

$$\begin{aligned} \int_{\Pi} \|\mathcal{F}_{u,\omega} - \mathbb{E}\hat{\mathcal{F}}_{u,\omega}^{(T)}\|_2^2 d\omega &= \int_{\Pi} \int_{[0,1]^2} |f_{u,\omega}(\tau, \sigma) - \mathbb{E}\hat{f}_{u,\omega}^{(T)}(\tau, \sigma)|^2 d\tau d\sigma d\omega \\ &= O\left(b_t^2 + b_f^2 + \frac{\log b_t T}{b_t T}\right)^2. \end{aligned}$$

□

Proof of Proposition 5.10. We have

$$\begin{aligned} \text{cum}(\hat{E}_{u,\omega_1}^{(T)}(\psi_{m_1 n_1}), \dots, \hat{E}_{u,\omega_k}^{(T)}(\psi_{m_k n_k})) &= \frac{(b_{t,T} b_{f,T} T)^{k/2}}{H_{2,N}(0)^k} \int_{\Pi^k} \prod_{j=1}^k K_{f,T}(\omega_j - \lambda_j) \\ &\times \text{cum}\left(D_{u,\omega_1}^{(T)}(\psi_{m_1}) D_{u,-\omega_1}^{(T)}(\psi_{n_1}), \dots, D_{u,\omega_k}^{(T)}(\psi_{m_k}) D_{u,-\omega_k}^{(T)}(\psi_{n_k})\right) d\lambda_1 \cdots d\lambda_k, \end{aligned} \quad (\text{A1.17})$$

where $D_{u,\omega}^{(T)}(\phi) = \langle D_{u,\omega}^{(T)}, \phi \rangle$ for $\phi \in L_{\mathbb{C}}^2([0, 1])$. Application of the product theorem for cumulants (e.g. Brillinger, 1981, Theorem 2.3.2) yields for the cumulant

$$\begin{aligned} \text{cum}(D_{u,\omega_1}^{(T)}(\psi_{m_1}) D_{u,-\omega_1}^{(T)}(\psi_{n_1}), \dots, D_{u,\omega_k}^{(T)}(\psi_{m_k}) D_{u,-\omega_k}^{(T)}(\psi_{n_k})) \\ = \sum_{i.p.} \prod_{l=1}^M \text{cum}(D_{u,\gamma_p}^{(T)}(\psi_{\tau_p}), p \in P_l), \end{aligned} \quad (\text{A1.18})$$

where the summation extends over all indecomposable partitions $P = \{P_1, \dots, P_M\}$ of the table

$$\begin{array}{cc} (1, 0) & (1, 1) \\ \vdots & \vdots \\ (k, 0) & (k, 1) \end{array},$$

and, for $p = (i, j)$, $\gamma_p = \gamma_{ij} = (-1)^j \lambda_i$ as well as $r_p = r_{ij} = m_i^{1-j} n_i^j$ for $i = 1, \dots, k$ and $j \in \{0, 1\}$. For the next steps, we further denote the elements of P_l with $|P_l| = d_l$ by p_{l1}, \dots, p_{ld_l} . Then, by (29), we obtain further for the above cumulant

$$\begin{aligned} \sum_{i.p.} \prod_{l=1}^M \int_{\Pi^{d_l-1}} \int_{[0,1]^{d_l}} \left[\bigotimes_{s=1}^{d_l} H_N(\mathcal{A}_{i_{u,\bullet}, \alpha_s}^{(T)}, \gamma_{p_{ls}} - \alpha_s) \mathcal{F}_{\alpha_1, \dots, \alpha_{d_l-1}}^\varepsilon \right] (\tau_1, \dots, \tau_{d_l}) \\ \times \prod_{s=1}^{d_l} \overline{\psi_{r_{p_{ls}}}(\tau_s)} d\tau_1 \cdots d\tau_{d_l} \eta(\alpha_1 + \dots + \alpha_{d_l}) d\alpha_1 \cdots d\alpha_{d_l}. \end{aligned} \quad (\text{A1.19})$$

Noting that the inner integral is a inner product in the tensor product space, we get

$$\begin{aligned}
 & \left| \left\langle \bigotimes_{s=1}^{d_l} H_N(\mathcal{A}_{t_{u,\bullet},\alpha_s}^{(T)}, \gamma_{p_{l_s}} - \alpha_s) \mathcal{F}_{\alpha_1, \dots, \alpha_{d_l-1}}^\varepsilon, \bigotimes_{s=1}^{d_l} \psi_{r_{p_{l_s}}} \right\rangle \right| \\
 & \leq \left\| \bigotimes_{s=1}^{d_l} H_N(\mathcal{A}_{t_{u,\bullet},\alpha_s}^{(T)}, \gamma_{p_{l_s}} - \alpha_s) \mathcal{F}_{\alpha_1, \dots, \alpha_{d_l-1}}^\varepsilon \right\|_2 \left\| \bigotimes_{s=1}^{d_l} \psi_{r_{p_{l_s}}} \right\|_2 \\
 & \leq \left\| \bigotimes_{s=1}^{d_l} H_N(\mathcal{A}_{t_{u,\bullet},\alpha_s}^{(T)}, \gamma_{p_{l_s}} - \alpha_s) \right\|_\infty \left\| \mathcal{F}_{\alpha_1, \dots, \alpha_{d_l-1}}^\varepsilon \right\|_2 \prod_{s=1}^{d_l} \left\| \psi_{r_{p_{l_s}}} \right\|_2.
 \end{aligned}$$

Noting that by Lemma B3.3

$$\left\| H_N(\mathcal{A}_{t_{u,\bullet},\alpha_s}^{(T)}, \gamma_{p_{l_s}} - \alpha_s) \right\|_\infty \leq K L_N(\gamma_{p_{l_s}} - \alpha_s)$$

for some constant K , we get together with $\left\| \mathcal{F}_{\alpha_1, \dots, \alpha_{d_l-1}}^\varepsilon \right\|_2 \leq K'$ as an upper bound for (A1.19)

$$K \sum_{i.p.} \prod_{l=1}^M \int_{\Pi^{d_l}} \prod_{s=1}^{d_l} L_N(\gamma_{p_{l_s}} - \alpha_s) \eta(\alpha_1 + \dots + \alpha_{d_l}) d\alpha_1 \cdots d\alpha_{d_l}$$

and further by repeated use of Lemma B3.1(v)

$$\leq K \sum_{i.p.} \prod_{l=1}^M L_N(\bar{\gamma}_l) \log(N)^{d_l-1} \leq K \log(N)^{2k-M} \sum_{i.p.} \prod_{l=1}^M L_N(\bar{\gamma}_l).$$

Substituting the upper bound for the cumulant in (A1.17) and noting that $\frac{1}{N} H_{2,N}(0) \rightarrow \|h\|_2^2$ as $N \rightarrow \infty$, we find

$$\begin{aligned}
 & \left| \text{cum}(\hat{E}_{u,\omega_1}^{(T)}(\psi_{m_1 n_1}), \dots, \hat{E}_{u,\omega_k}^{(T)}(\psi_{m_k n_k})) \right| \\
 & \leq \frac{C b_{f,T}^{k/2} \log(N)^{2k-M}}{N^{k/2}} \sum_{i.p.} \int_{\Pi^k} \prod_{j=1}^k K_{f,T}(\omega_j - \lambda_j) \prod_{l=1}^M L_N(\bar{\gamma}_l) d\lambda_1 \cdots d\lambda_k.
 \end{aligned} \tag{A1.20}$$

It is sufficient to show that for each indecomposable partition $\{P_1, \dots, P_M\}$ the corresponding term in the above sum tends to zero. First, suppose that $M = k$. Bounding the factors $K_{f,T}(\omega_i - \lambda_i)$ by $\|K_f\|_\infty / b_{f,T}$ for $i = 2, \dots, k$ and integrating over $\lambda_3, \dots, \lambda_k$, we obtain by Lemma B3.2(i) as an upper bound

$$\begin{aligned}
 & \frac{C \log(N)^{2k-M}}{b_{f,T}^{k/2-1} N^{k/2}} \int_{\Pi^2} K_{f,T}(\omega_1 - \lambda_1) L_N(\lambda_1 \pm \lambda_2)^2 d\lambda_1 d\lambda_2 \\
 & \leq \frac{C \log(N)^{2k-2}}{b_{f,T}^{k/2-2} N^{k/2}} \int_{\Pi^2} L_{b_{f,T}^{-1}}(\omega_1 - \lambda_1)^2 L_N(\lambda_1 \pm \lambda_2)^2 d\lambda_2 d\lambda_1 \\
 & \leq \frac{C \log(N)^{2k-2}}{b_{f,T}^{k/2-2} N^{k/2}} \int_{\Pi} N L_{b_{f,T}^{-1}}(\omega_1 - \lambda_1)^2 d\lambda_1 \leq \frac{C \log(N)^{2k-2}}{(b_{f,T} N)^{k/2-1}},
 \end{aligned}$$

where we have $K_{f,T}(\omega) \leq b_{f,T} L_{b_{f,T}^{-1}}(\omega)$ and repeatedly Lemma B3.1(iv). Next, if $M < k$ we select variables $\lambda_{i_1}, \dots, \lambda_{i_{k-2}}$ according to Lemma B3.2(ii) and bound all corresponding factors $K_{f,T}(\omega_{i_j} - \lambda_{i_j})$ for $j = 1, \dots, k-2$ by $\|K_{f,T}\|_{\infty}/b_{f,T}$. Then integration over the $k-2$ selected variables yields the upper bound

$$\begin{aligned} & \frac{C \log(N)^{3k-M-2}}{b_{f,T}^{k/2-2} N^{k/2-1}} \int_{\Pi^2} K_{f,T}(\omega_{i_{k-1}} - \lambda_{i_{k-1}}) K_{f,T}(\omega_{i_k} - \lambda_{i_k}) d\lambda_{i_{k-1}} d\lambda_{i_k} \\ & \leq \frac{C b_{f,T} \log(N)^{3k-M-2}}{b_{f,T}^{k/2-1} N^{k/2-1}}, \end{aligned}$$

since $\|K_{f,T}\|_1 = 1$. Since $b_{f,T} N = b_{f,T} b_{t,T} T \rightarrow \infty$ and $k/2 - 1 > 0$, the upper bounds tend to zero as $T \rightarrow \infty$, which completes the proof. \square

Appendix B

In Appendix B, we provide additional technical material necessary to complete the proofs of the main part of this paper. Section B1 contains background material on operator theory. Section B2 provide background on the higher order structure of random functions as well as an important result on the existence of a stochastic integral, necessary to define functional Cramér representations. Section B3 introduces the necessary background on tapering on function spaces.

Appendix B1: Some operator theory

We start with a general characterization of a tensor product of a finite sequence of vector spaces, which in particular holds for sequences of Hilbert spaces.

Definition B1.1 (Algebraic tensor product of Banach spaces). Given a finite sequence of vector spaces V_1, \dots, V_k over an arbitrary field \mathbb{F} , we define the *algebraic tensor product* $V_1 \otimes \dots \otimes V_k$ as a vector space with a multi-linear map $V_1 \times \dots \times V_k \rightarrow W$ given by $(f_1, \dots, f_k) \rightarrow (f_1 \otimes \dots \otimes f_k)$ such that, for every linear map $\mathcal{T} : V_1 \times \dots \times V_k \rightarrow W$, there is unique k-linear map $\tilde{\mathcal{T}} : V_1 \times \dots \times V_k \rightarrow W$ that satisfies

$$\mathcal{T}(f_1, \dots, f_k) = \tilde{\mathcal{T}}(f_1 \otimes \dots \otimes f_k).$$

Here, uniqueness is meant up to isomorphisms. The tensor product can be viewed as a linearized version of the product space $V_1 \times \dots \times V_k$ satisfying equivalence relations of the form $a(v_1, v_2) \sim (av_1, v_2) \sim (v_1, av_2)$ where $a \in \mathbb{K}$ and $v_1 \in V_1, v_2 \in V_2$, which induce a quotient space. These relationships uniquely identify the points in the product space $V_1 \times \dots \times V_k$ that yield multi-linear relationships. In a way, the tensor product $\bigotimes_{j=1}^k V_j$ can thus be viewed as the ‘freest’ way to put the respective different vector spaces V_1, \dots, V_k together. We mention in particular that the algebraic tensor product satisfies the associative law, i.e., $(V_1 \otimes V_2) \otimes V_3 = V_1 \otimes (V_2 \otimes V_3)$, and hence it will often be sufficient to restrict attention to $k = 2$.

The algebraic tensor product of two Hilbert spaces H_1 and H_2 is itself not a Hilbert space. We can however construct a Hilbert space by considering the inner product acting on $H_1 \otimes H_2$ given by

$$\langle x \otimes y, x' \otimes y' \rangle_{H_1 \otimes H_2} = \langle x, x' \rangle \langle y, y' \rangle, \quad x, x' \in H_1, y, y' \in H_2$$

and then taking the completion with respect to the induced norm $\|\cdot\|_{H_1 \otimes H_2}$. The completed space, denoted by $H_1 \widehat{\otimes} H_2$, is identifiable with the Hilbert-Schmidt operators and is referred to as the *Hilbert Schmidt tensor product*. Throughout this work, when reference is made to the tensor product space of Hilbert spaces we mean the latter space. When no confusion can arise, we shall moreover abuse notation slightly and denote $H_1 \widehat{\otimes} H_2$ simply by $H_1 \otimes H_2$.

Definition B1.2. The tensor product $(A \otimes B) \in S_P(H) \otimes S_p(H) \cong S_p(S_p(H))$ between two operators $A, B \in S_p(H)$ is defined as

$$(A \otimes B)(x \otimes y) = Ax \otimes By \quad (\text{B1.1})$$

for $x, y \in H$. It follows straightforwardly from the property

$$(x \otimes y)z = \langle z, y \rangle x, \quad z \in H, \quad (\text{B1.2})$$

that for any $C \in S_p(H)$, we have the identity

$$(A \otimes B)C = ACB^\dagger, \quad (\text{B1.3})$$

where B^\dagger denotes the adjoint operator of B .

Proposition B1.3 (Hölder's Inequality for operators). *Let H be a separable Hilbert space and $A, B \in S_\infty(H)$. Then the composite operator AB also defines a bounded linear operator over H , i.e., $AB \in S_\infty(H)$. This operation satisfies the associative law. Moreover, let $1 \leq p, q, r \leq \infty$, such that $\frac{1}{r} = \frac{1}{q} + \frac{1}{p}$. If $A \in S_q(H)$ and $B \in S_p(H)$ then $AB \in S_r(H)$ and*

$$\|AB\|_r \leq \|A\|_q \|B\|_p.$$

Proposition B1.4. *Let $H = L^2_{\mathbb{C}}(T, \mu)$ be a separable Hilbert space, where (T, μ) is a measure space. The functions $a, b, c \in L^2_{\mathbb{C}}(T \times T, \mu \otimes \mu)$ induce operators A, B, C on H such that for all $x \in H$*

$$Ax(\tau) = \int_{\mathcal{D}} a(\tau, \sigma)x(\sigma)d\mu(\sigma), \quad (\text{B1.4})$$

and the composition operator AB has kernel

$$[AB](\tau, \sigma) = \int_{\mathcal{D}} a(\tau, \mu_1)b(\mu_1, \sigma)d\mu_1 \quad (\text{B1.5})$$

for all $\tau, \sigma \in T$ μ -almost everywhere. The tensor product operator $(A \otimes B) \in S_2(S_2(H))$ in composition with C has kernel

$$[(A \otimes B)C](\tau, \sigma) = \int_{\mathcal{D}} \int_{\mathcal{D}} a(\tau, \mu_1)\overline{b(\sigma, \mu_2)}c(\mu_1, \mu_2)d\mu_1 d\mu_2. \quad (\text{B1.6})$$

Because $(A \otimes B)C$ has a well defined kernel in $L^2_{\mathbb{C}}(T \times T, \mu \otimes \mu)$, it can moreover be viewed as an operator on H . Using identity (B1.3), this is the operator ACB^\dagger , where B^\dagger has kernel $b^\dagger(\mu_2, \sigma) = \overline{b(\sigma, \mu_2)}$.

Corollary B1.5. Let $A_i, i = 1, \dots, k$ for k finite belong to $S_p(H)$ and let

$$\psi = (\psi_1 \otimes \dots \otimes \psi_k)$$

be an element of $\bigotimes_{i=1}^k H$. Then we have that the linear mapping

$$\mathcal{A} = (A_1 \otimes \dots \otimes A_k)$$

satisfies i) $\|\mathcal{A}\psi\|_2 < \infty$ and ii) $\|\mathcal{A}\|_p < \infty$.

Proof of Corollary B1.5. For i), we have by proposition B1.3,

$$\begin{aligned} \|\mathcal{A}\psi\|_2 &= \|(A_1 \otimes \dots \otimes A_k)\psi\|_2 = \|(A_1 \otimes \dots \otimes A_k)\|_\infty \|\psi\|_2 \\ &\leq \|(A_2 \otimes \dots \otimes A_k)\|_\infty \|A_1\|_\infty \|\psi\|_2 \\ &\leq \prod_{i=1}^k \|A_i\|_\infty \|\psi\|_2 \leq \prod_{i=1}^k \|A_i\|_p \|\psi\|_2 < \infty. \end{aligned}$$

In case $p = 2$, the latter equals $\prod_{i=1}^k \|a_i\|_2 \|\psi\|_2$ by proposition B1.7. Property ii) holds since for any $A_1, A_2 \in S_p(H)$, we have $\|A_1 \otimes A_2\|_p = \|A_1\|_p \|A_2\|_p$. To illustrate the second property, observe that if $p = 2$ we obtain

$$\begin{aligned} \|\mathcal{A}\|_2^2 &= \|A_1 \otimes \dots \otimes A_k\|_2^2 = \int_{[0,1]^{2k}} |a_1(\tau_1, \mu_1) \dots a_k(\tau_k, \mu_k)|^2 d\tau_1 \dots d\tau_k d\mu_1 \dots d\mu_k \\ &= \int_{[0,1]^2} a_1(\tau_1, \mu_1) \overline{a_1(\tau_1, \mu_1)} d\tau_1, d\mu_1 \dots \int_{[0,1]^2} a_k(\tau_k, \mu_k) \overline{a_k(\tau_k, \mu_k)} d\tau_k d\mu_k \\ &= \|a_1\|_2^2 \dots \|a_k\|_2^2 < \infty. \end{aligned} \quad \square$$

Proposition B1.6 (Neumann series). Let A be a bounded linear operator on H and I_H be the identity operator. If $\|A\|_\infty < 1$, the operator $I_H - A$ has a unique bounded inverse on H given by

$$(I_H - A)^{-1} = \sum_{k=0}^{\infty} A^k. \quad (\text{B1.7})$$

If $A \in S_2(H)$ with $\|A\|_2 < 1$, then this equality holds in Hilbert-Schmidt norm.

Proof. We only show the case $A \in S_2(H)$. Note that the space $S_2(H)$ is a Hilbert space. Then for $m < n$,

$$\left\| \sum_{k=0}^m A^k - \sum_{k=0}^n A^k \right\|_2 \leq \sum_{k=m+1}^n \|A\|_2^k \leq \frac{\|A\|_2^{m+1}}{1 - \|A\|_2},$$

which shows that the partial sum forms a Cauchy sequence and hence has a limit A^* in $S_2(H)$. Furthermore, we have

$$(I_H - A)A^* = \lim_{n \rightarrow \infty} (I_H - A) \sum_{k=0}^n A^k = \lim_{n \rightarrow \infty} (I_H - A^{n+1}) = I_H$$

in $S_2(H)$, which shows that A^* is the inverse of $I_H - A$. □

Proposition B1.7 (Hilbert-Schmidt operators as kernel operator). *Let $H = L^2_{\mathbb{C}}(T, \mu)$ be a separable Hilbert space, where (T, μ) is a measure space, and let A be an operator on H . Then $A \in S_2(H)$ if and only if it is an integral operator, that is, there exists a function $a \in L^2_{\mathbb{C}}(T \times T, \mu \otimes \mu)$ such that*

$$Ax(\tau) = \int a(\tau, \sigma) x(\sigma) d\mu(\sigma)$$

for all $\tau \in T$ μ -almost everywhere. Moreover, we have $\|A\|_2 = \|a\|_2$.

Proof. First, suppose A is an integral operator on H with kernel $a \in L^2_{\mathbb{C}}(T \times T, \mu \otimes \mu)$. Because H is separable, it has a countable orthonormal basis $\{\psi_n\}_{n \in \mathbb{N}}$. For fixed $\tau \in M$, the function $a_{\tau}(\sigma) = a(\tau, \sigma)$ defines a measurable function on $L^2_{\mathbb{C}}(T, \mu)$. We can therefore write

$$A\psi_n(\tau) = \int a(\tau, \sigma) \psi_n(\sigma) d\mu(\sigma) = \langle a_{\tau}, \overline{\psi_n} \rangle.$$

Observe that $\{\overline{\psi_n}\}_{n \geq 1}$ also forms a orthonormal basis of H . An application of the Cauchy-Schwarz Inequality gives $|\langle a_{\tau}, \overline{\psi_n} \rangle|^2 \leq \|a_{\tau}\|^2 \|\overline{\psi_n}\|^2 < \infty$ and therefore

$$\sum_{n=1}^m |\langle a_{\tau}, \overline{\psi_n} \rangle|^2 \leq \sum_{n=1}^{\infty} |\langle a_{\tau}, \overline{\psi_n} \rangle|^2 = \|a_{\tau}\|_2^2 < \infty,$$

by Parseval's Identity. Hence, as a corollary of the Monotone and Dominated Convergence Theorem we find

$$\begin{aligned} \|A\|_2^2 &= \sum_{n=1}^{\infty} \|A\psi_n\|^2 = \lim_{m \rightarrow \infty} \sum_{n=1}^m |\langle a_{\tau}, \overline{\psi_n} \rangle|^2 d\tau = \int \lim_{m \rightarrow \infty} \sum_{n=1}^m |\langle a_{\tau}, \overline{\psi_n} \rangle|^2 d\tau \\ &= \int \|a_{\tau}\|^2 d\tau = \iint |a(\tau, \sigma)|^2 d\sigma d\tau = \|a\|_2^2 < \infty, \end{aligned}$$

showing A is Hilbert Schmidt and $\|A\|_2 = \|a\|_2$. Now suppose A is Hilbert Schmidt. In this case, we have by definition $\sum_{n=1}^{\infty} \|A\psi_n\|^2 < \infty$ and consequently the series $\sum_{n=1}^{\infty} A\psi_n$ converges in $L^2_{\mathbb{C}}(T, \mu)$. Therefore the function

$$a(\tau, \sigma) = \sum_{n=1}^{\infty} A\psi_n(\tau) \overline{\psi_n(\sigma)}$$

will be well-defined on $L^2_{\mathbb{C}}(T \times T, \mu \otimes \mu)$. Hence, for any element $x \in L^2_{\mathbb{C}}(T, \mu)$, the Dominated Convergence Theorem yields

$$\begin{aligned} Ax(\tau) &= A\left(\lim_{m \rightarrow \infty} \sum_{n=1}^m \langle x, \psi_n \rangle \psi_n\right)(\tau) = \lim_{m \rightarrow \infty} \sum_{n=1}^m \langle x, \psi_n \rangle A\psi_n(\tau) \\ &= \lim_{m \rightarrow \infty} \sum_{n=1}^m \left(\int x(\sigma) \overline{\psi_n(\sigma)} d\sigma \right) A\psi_n(\tau) \\ &= \lim_{m \rightarrow \infty} \left(\int x(\sigma) \sum_{n=1}^m \overline{\psi_n(\sigma)} A\psi_n(\tau) d\sigma \right) \\ &= \int x(\sigma) \sum_{n \geq 1} \overline{\psi_n(\sigma)} A\psi_n(\tau) d\sigma = \int x(\sigma) a(\tau, \sigma) d\sigma. \quad \square \end{aligned}$$

Theorem B1.8 (Product Rule on Banach spaces). *Let E, F_1, F_2, G be Banach spaces and let $U \subset E$ be open. Suppose that $f : U \rightarrow F_1$, and $G : U \in F_2$ are Fréchet differentiable of order k . Let $Z(\cdot, \cdot) : F_1 \times F_2 \rightarrow G$ be a continuous bilinear map. Then, $Z(f, g) : U \rightarrow G$ is Fréchet differentiable of order k and*

$$\frac{\partial Z}{\partial u}(f(u), g(u)) = Z\left(\frac{\partial f(u)}{\partial u}, g(u)\right) + Z\left(f(u), \frac{\partial g(u)}{\partial u}\right). \quad (\text{B1.8})$$

For the proof, see for example Nelson (1969).

Appendix B2: Moment and cumulant tensors

Let X be a random element on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ that takes values in a separable Hilbert space H . More precisely, we endow H with the topology induced by the norm on H and assume that $X : \Omega \rightarrow H$ is Borel-measurable. Then the mean $\mathbb{E}(X)$ of X in H exists and is given by

$$\mathbb{E}(X) = \sum_{i \in \mathbb{N}} \mathbb{E}(\langle X, \psi_i \rangle) \psi_i,$$

where $(\psi_i)_{i \in \mathbb{N}}$ is an orthonormal basis of H , provided that $\mathbb{E}(\|X\|_2^2) < \infty$.

For higher moments, it is appropriate to consider these as tensors in a tensor product space $H \otimes \cdots \otimes H$ of appropriate dimension. More precisely, let X_1, \dots, X_k be random elements in H . Then we define the moment tensor $\mathbb{E}(X_1 \otimes \cdots \otimes X_k)$ by

$$\mathbb{E}(X_1 \otimes \cdots \otimes X_k) = \sum_{i_1, \dots, i_k \in \mathbb{N}} \mathbb{E}\left(\prod_{j=1}^k \langle X_j, \psi_{i_j} \rangle\right) \psi_{i_1} \otimes \cdots \otimes \psi_{i_k}.$$

Similarly, we define the cumulant tensor $\text{cum}(X_1, \dots, X_k)$ by

$$\text{cum}(X_1, \dots, X_k) = \sum_{i_1, \dots, i_k \in \mathbb{N}} \text{cum}(\langle X_1, \psi_{i_1} \rangle, \dots, \langle X_k, \psi_{i_k} \rangle) \psi_{i_1} \otimes \cdots \otimes \psi_{i_k}. \quad (\text{B2.9})$$

The cumulants on the right hand side are as usual given by

$$\text{cum}(\langle X_1, \psi_{i_1} \rangle, \dots, \langle X_k, \psi_{i_k} \rangle) = \sum_{\nu=(\nu_1, \dots, \nu_p)} (-1)^{p-1} (p-1)! \prod_{r=1}^p \mathbb{E}\left(\prod_{j \in \nu_r} \langle X_j, \psi_{i_j} \rangle\right),$$

where the summation extends over all unordered partitions ν of $\{1, \dots, k\}$.

More generally, we also require the case where the X_i are themselves tensors, that is, $X_i = \otimes_{j=1}^{l_i} X_{ij}$, $i = 1, \dots, k$, for random elements X_{ij} in H with $j = 1, \dots, l_i$ and $i = 1, \dots, k$. In this case, the joint cumulant tensor $\text{cum}(X_1, \dots, X_k)$ is given by an appropriate generalization of the product theorem for cumulants (Brillinger, 1981, Theorem 2.3.2) to the tensor case,

$$\begin{aligned} & \text{cum}(X_1, \dots, X_k) \\ &= \sum_{r_{11}, \dots, r_{kl_k} \in \mathbb{N}} \sum_{\nu=(\nu_1, \dots, \nu_p)} \prod_{n=1}^p \text{cum}(\langle X_{ij}, \psi_{r_{ij}} \rangle | (i, j) \in \nu_n) \psi_{r_{11}} \otimes \cdots \otimes \psi_{r_{kl_k}}, \end{aligned}$$

where the summation extends over all indecomposable partitions $\nu = (\nu_1, \dots, \nu_p)$ of the table

$$\begin{array}{ccc} (1, 1) & \cdots & (1, l_1) \\ \vdots & \ddots & \vdots \\ (k, 1) & \cdots & (k, l_k). \end{array}$$

Formally, we also abbreviate this by

$$\text{cum}(X_1, \dots, X_k) = \sum_{\nu=(\nu_1, \dots, \nu_p)} S_\nu \left(\bigotimes_{n=1}^p \text{cum}(X_{ij} | (i, j) \in \nu_n) \right), \quad (\text{B2.10})$$

where S_ν is the permutation that maps the components of the tensor back into the original order, that is, $S_\nu \left(\bigotimes_{r=1}^p \bigotimes_{(i,j) \in \nu_r} X_{ij} \right) = X_{11} \otimes \cdots \otimes X_{kl_k}$.

Next, let A_1, \dots, A_k linear bounded operators on H . As in Appendix B1, let $A_1 \otimes \cdots \otimes A_k$ be the operator on $H \otimes \cdots \otimes H$ given by

$$(A_1 \otimes \cdots \otimes A_k)(x_1 \otimes \cdots \otimes x_k) = (A_1 x_1) \otimes \cdots \otimes (A_k x_k)$$

for all $x_1, \dots, x_k \in H$. The next proposition states that moment tensors—and hence also cumulant tensors by the above definitions—transform linearly.

Proposition B2.1. *Let A_1, \dots, A_k be bounded linear operators on H and let X_1, \dots, X_k be random elements in H . Then*

$$(A_1 \otimes \cdots \otimes A_k) \mathbb{E}(X_1 \otimes \cdots \otimes X_k) = \mathbb{E}((A_1 X_1) \otimes \cdots \otimes (A_k X_k)). \quad (\text{B2.11})$$

Proof. Let $\{\psi_i\}_{i \in \mathbb{N}}$ be an orthonormal basis of H . Using the definition of a moment tensor, we get

$$\begin{aligned} & (A_1 \otimes \cdots \otimes A_k) \mathbb{E}(X_1 \otimes \cdots \otimes X_k) \\ &= \sum_{i_1, \dots, i_k \in \mathbb{N}} \mathbb{E} \left(\prod_{j=1}^k \langle X_j, \psi_{i_j} \rangle \right) (A_1 \psi_{i_1}) \otimes \cdots \otimes (A_k \psi_{i_k}) \end{aligned}$$

and further, by representing $A_j \psi_{i_j}$ with respect to the chosen orthonormal basis,

$$\begin{aligned} &= \sum_{i_1, \dots, i_k \in \mathbb{N}} \sum_{n_1, \dots, n_k \in \mathbb{N}} \mathbb{E} \left(\prod_{j=1}^k \langle X_j, \psi_{i_j} \rangle \right) \prod_{j=1}^k \langle A_j \psi_{i_j}, \psi_{n_j} \rangle (\psi_{n_1} \otimes \cdots \otimes \psi_{n_k}) \\ &= \sum_{n_1, \dots, n_k \in \mathbb{N}} \mathbb{E} \left[\prod_{j=1}^k \left\langle A_j \left(\sum_{i_j \in \mathbb{N}} \langle X_j, \psi_{i_j} \rangle \right), \psi_{n_j} \right\rangle \right] (\psi_{n_1} \otimes \cdots \otimes \psi_{n_k}) \\ &= \sum_{n_1, \dots, n_k \in \mathbb{N}} \mathbb{E} \left[\prod_{j=1}^k \langle A_j X_j, \psi_{n_j} \rangle \right] (\psi_{n_1} \otimes \cdots \otimes \psi_{n_k}) \\ &= \mathbb{E}((A_1 X_1) \otimes \cdots \otimes (A_k X_k)), \end{aligned}$$

where we have used linearity of the operators, of the inner product, and of the ordinary mean. \square

As a direct consequence of the above proposition, we also have linearity of cumulant tensors. More precisely, for $i = 1, \dots, k$, let X_i be a random tensor in $\otimes_{j=1}^k H$ and let A_i be a linear bounded operator on the same tensor product space. Then

$$(A_1 \otimes \cdots \otimes A_k) \text{cum}(X_1, \dots, X_k) = \text{cum}(A_1 X_1, \dots, A_k X_k). \quad (\text{B2.12})$$

B2.1. Higher order dependence under functional stationarity

For stationary processes, we follow convention and write the cumulant tensor (B2.9) as a function of $k-1$ elements, i.e., we denote it by $\mathcal{C}_{t_1, \dots, t_{k-1}}$. Under regularity conditions this operator is in $S_2(H)$ and we can define the corresponding k -th order cumulant kernel $c_{t_1, \dots, t_{k-1}}$ of the process X by

$$\begin{aligned} & \mathcal{C}_{t_1, \dots, t_{k-1}} \\ &= \sum_{i_1, \dots, i_k \in \mathbb{N}} \int_{[0,1]^k} c_{t_1, \dots, t_{k-1}}(\tau_1, \dots, \tau_k) \prod_{j=1}^k \psi_{i_j}(\tau_j) d\tau_1 \cdots d\tau_k \psi_{i_1} \otimes \cdots \otimes \psi_{i_k}. \end{aligned} \quad (\text{B2.13})$$

A sufficient condition that is often imposed for this to hold is $\mathbb{E}\|X_0\|_2^k < \infty$. Similar to the second-order case, the tensor (B2.9) will form a Fourier pair with a k -th order cumulant spectral operator given summability with respect to $\|\cdot\|_p$ is satisfied. The k -th order cumulant spectral tensor is specified as

$$\mathcal{F}_{\omega_1, \dots, \omega_{k-1}} = (2\pi)^{1-k} \sum_{t_1, \dots, t_{k-1} \in \mathbb{Z}} \mathcal{C}_{t_1, \dots, t_{k-1}} \exp\left(-i \sum_{j=1}^{k-1} \omega_j t_j\right), \quad (\text{B2.14})$$

where the convergence is in $\|\cdot\|_p$. Properties on the kernels that are relevant in the time-dependent framework are discussed in section 4 of the main paper.

Theorem B2.2. *Let $\{X_t\}_{t \in \mathbb{Z}}$ be a stationary stochastic process in $L^2([0, 1])$ such that $\mathbb{E}\|X_0\|_2^k < \infty$ for all $k \in \mathbb{N}$ and $\sum_{t_1, \dots, t_{k-1} = -\infty}^{\infty} \|\mathcal{C}_{t_1, \dots, t_{k-1}}\|_2 < \infty$. Furthermore let*

$$Z_\omega^{(N)} = \frac{1}{2\pi} \sum_{t=-N}^N X_t \int_{-\pi}^{\omega} e^{-i\lambda t} d\lambda.$$

Then there exists a 2π -periodic stochastic process $\{Z_\omega\}_{\omega \in \mathbb{R}}$ taking values in $L^2_{\mathbb{C}}([0, 1])$ with $\bar{Z}_\omega = Z_{-\omega}$ such that $\lim_{N \rightarrow \infty} \mathbb{E}\|Z_\omega^{(N)} - Z_\omega\|_2^2 = 0$. Furthermore, $\{Z_\omega\}$ equals almost everywhere the functional orthogonal increment process of the Cramer representation of $\{X_t\}$, that is,

$$X_t = \int_{-\pi}^{\pi} e^{i\omega t} dZ_\omega \quad \text{a.e. in } \mathbb{H}.$$

Finally, we have for $k \geq 2$

$$\text{cum}(Z_{\omega_1}, \dots, Z_{\omega_k}) = \int_{-\pi}^{\omega_1} \cdots \int_{-\pi}^{\omega_k} \eta\left(\sum_{j=1}^k \lambda_j\right) \mathcal{F}_{\alpha_1, \dots, \alpha_{k-1}} d\alpha_1 \cdots d\alpha_k, \quad (\text{B2.15})$$

which holds almost everywhere and in L^2 .

The final statement of the above theorem suggests the use of the differential notation

$$\begin{aligned} & \text{cum}(dZ_{\omega_1}(\tau_1), \dots, dZ_{\omega_k}(\tau_k)) \\ &= \eta(\omega_1 + \dots + \omega_k) f_{\omega_1, \dots, \omega_{k-1}}(\tau_1, \dots, \tau_k) d\omega_1 \cdots d\omega_k. \end{aligned}$$

Proof of Theorem B2.2. The theorem generalizes Theorem 4.6.1 of Brillinger (1981). Let μ be the measure on the interval $[-\pi, \pi]$ given by

$$\mu(A) = \int_A \|\mathcal{F}_\omega\|_1 d\omega,$$

for all Borel sets $A \subseteq [-\pi, \pi]$. Similar to the time series setting, it has been shown (Panaretos and Tavakoli, 2013a) that there is a unique isomorphism \mathcal{T} of $\overline{\text{sp}}\{X_t\}_{t \in \mathbb{Z}}$ onto $L^2_{\mathbb{C}}([-\pi, \pi], \mu)$ such that

$$\mathcal{T} X_t = e^{it}.$$

for all $t \in \mathbb{Z}$. The process defined by $Z_\omega = \mathcal{T}^{-1}(1_{(-\pi, \omega]}(\cdot))$ is then a functional orthogonal increment process of which the second order properties are completely determined by the spectral density operator \mathcal{F} . We have

$$\mathcal{T}(Z_\omega - Z_\nu) = 1_{(\nu, \omega]}(\cdot), \quad -\pi < \nu < \omega < \pi,$$

and for $b_j \in \mathbb{C}$, $j = 1, \dots, N$

$$\mathcal{T} \left(\sum_{j=1}^N b_j X_{t_j} \right) = \sum_{j=1}^N b_j e^{it_j(\cdot)}.$$

For the first part of the proof, we shall use that the function $1_{(-\pi, \omega]}(\cdot)$ can be approximated by the N -th order Fourier series approximation

$$b_N(\lambda) = \sum_{|t| \leq N} \tilde{b}_{\omega, t} e^{it\lambda},$$

where the Fourier coefficients are given by

$$\tilde{b}_{\omega, t} = \frac{1}{2\pi} \int_{-\pi}^{\pi} 1_{(-\pi, \omega]}(\lambda) e^{-it\lambda} d\lambda. \quad (\text{B2.16})$$

The approximation satisfies the properties listed in the following proposition (Brockwell and Davis, 1991, Proposition 4.11.2).

Proposition B2.3. *Let $\{b_N\}_{N \geq 1}$ be the sequence of functions defined in (B2.16). Then for $-\pi < \nu < \omega < \pi$,*

- (i) $\sup_{\lambda \in [-\pi, \pi] \setminus \mathcal{E}} |b_N(\lambda) - 1_{(\nu, \omega]}(\lambda)| \rightarrow 0$ as $N \rightarrow \infty$, where \mathcal{E} is an open subset of $[-\pi, \pi]$ containing both ν and ω ;
- (ii) $\sup_{\lambda \in [-\pi, \pi]} |b_N(\lambda)| \leq C < \infty$ for all $N \geq 1$.

Note then that we can write

$$Z_\omega^{(N)} = \frac{1}{2\pi} \sum_{|t| \leq N} X_t \int_{-\pi}^{\pi} 1_{(-\pi, \omega]}(\lambda) e^{it\lambda} d\lambda = \sum_{|t| \leq N} \tilde{b}_{\omega, t} X_t,$$

where $\{\tilde{b}_{\omega, t}\}_{t \in \mathbb{N}}$ are the Fourier coefficients of the indicator function $1_{(-\pi, \omega]}$. Therefore,

$$\begin{aligned} & \text{cum}(Z_{\omega_1}^{(N)}, \dots, Z_{\omega_k}^{(N)}) \\ &= \sum_{|t_1|, \dots, |t_k| \leq N} \tilde{b}_{\omega_1, t_1} \cdots \tilde{b}_{\omega_k, t_k} \text{cum}(X_{t_1}, \dots, X_{t_k}) \end{aligned}$$

and by stationarity of the process X_t

$$\begin{aligned} &= \sum_{|t_1|, \dots, |t_k| \leq N} \tilde{b}_{\omega_1, t_1} \cdots \tilde{b}_{\omega_k, t_k} \int_{\Pi^k} e^{i(\alpha_1 t_1 + \dots + \alpha_k t_k)} \eta\left(\sum_{j=1}^k \alpha_j\right) \\ & \quad \times \mathcal{F}_{\alpha_1 \dots \alpha_{k-1}} d\alpha_1 \cdots d\alpha_k \\ &= \int_{\Pi^k} \eta\left(\sum_{j=1}^k \alpha_j\right) \mathcal{F}_{\alpha_1 \dots \alpha_{k-1}} \prod_{i=1}^k \sum_{|t_i| < N} \left(\int_{-\pi}^{\omega_i} e^{-it_i \lambda_i} d\lambda_i \right) e^{i\alpha_i t_i} d\alpha_1 \cdots d\alpha_k \\ &= \int_{\Pi^k} \eta\left(\sum_{j=1}^k \alpha_j\right) \mathcal{F}_{\alpha_1 \dots \alpha_{k-1}} b_{\omega_1, N}(\alpha_1) \cdots b_{\omega_k, N}(\alpha_k) d\alpha_1 \cdots d\alpha_k. \end{aligned}$$

To show convergence, recall that the kernel function $\mathcal{F}_{\alpha_1 \dots \alpha_{k-1}}$ is bounded and uniformly continuous in the manifold $\sum_{j=1}^k \alpha_j \equiv 0 \pmod{2\pi}$ with respect to $\|\cdot\|_2$. An application of Hölder's inequality yields

$$\begin{aligned} & \left\| \int_{\Pi^k} \eta\left(\sum_{j=1}^k \alpha_j\right) \mathcal{F}_{\alpha_1, \dots, \alpha_{k-1}} \left[\prod_{i=1}^k b_{\omega_i, N}(\alpha_i) - \prod_{i=1}^k 1_{(-\pi, \omega_i]}(\alpha_i) \right] d\alpha_1 \cdots d\alpha_k \right\|_2 \\ & \leq \sup_{\alpha_1, \dots, \alpha_{k-1}} \|\mathcal{F}_{\alpha_1 \dots \alpha_{k-1}}\|_2 \int_{\Pi^k} \left| \prod_{i=1}^k b_{\omega_i, N}(\alpha_i) - \prod_{i=1}^k 1_{(-\pi, \omega_i]}(\alpha_i) \right| d\alpha_1 \cdots d\alpha_k. \end{aligned}$$

A standard telescoping argument together with Proposition B2.3 gives

$$\begin{aligned} & \leq K \int_{\Pi^k} \sum_{j=1}^k \prod_{l=1}^{j-1} |b_{\omega_l, N}(\alpha_l)| \prod_{l=j+1}^k |1_{(-\pi, \omega_1]}(\alpha_l)| \\ & \quad \times |b_{\omega_j, N}(\alpha_j) - 1_{(-\pi, \omega_j]}(\alpha_j)| d\alpha_1 \cdots d\alpha_k \\ & \leq K k \left(\sup_{1 \leq j \leq k} \sup_{\alpha} |b_{\omega_j, N}(\alpha)| \right)^{k-1} \sup_{\omega} \int_{\Pi} |b_{\omega, N}(\alpha) - 1_{(-\pi, \omega]}(\alpha)| d\alpha \rightarrow 0 \end{aligned}$$

as $N \rightarrow \infty$. Hence, the dominated convergence theorem implies

$$\begin{aligned}
 & \lim_{N \rightarrow \infty} \text{cum}(Z_{\omega_1}^{(N)}, \dots, Z_{\omega_k}^{(N)}) \\
 &= \frac{1}{(2\pi)^k} \int_{\Pi^k} 1_{(-\pi, \omega_1]}(\alpha_1) \cdots 1_{(-\pi, \omega_k]}(\alpha_k) \mathcal{F}_{\alpha_1 \dots \alpha_{k-1}} \eta \left(\sum_{j=1}^k \alpha_j \right) d\alpha_1 \cdots d\alpha_k \\
 &= \frac{1}{(2\pi)^k} \int_{-\pi}^{\omega_1} \cdots \int_{-\pi}^{\omega_k} \eta \left(\sum_{j=1}^k \lambda_j \right) \mathcal{F}_{\alpha_1 \dots \alpha_{k-1}} d\lambda_1 \cdots d\lambda_k \\
 &= \text{cum}(Z_{\omega_1}, \dots, Z_{\omega_k}), \tag{B2.17}
 \end{aligned}$$

which establishes the L^2 convergence in (B2.15). The almost everywhere convergence is proved similarly by replacing \mathcal{F} by $f(\tau_1, \dots, \tau_k)$. In order to show that $X_t = \int_{-\pi}^{\pi} e^{i\omega t} dZ_\omega$ with probability 1, it remains to show that

$$\mathbb{E} \left\| X_t - \int_{-\pi}^{\pi} e^{i\omega t} dZ_\omega \right\|_2^2 = 0. \tag{B2.18}$$

We refer to Panaretos and Tavakoli (2013a) for a proof. \square

B2.2. Higher order dependence for time-dependent linear models

Proposition B2.4. *Let $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ be a functional i.i.d. process in H with $\mathbb{E} \|\varepsilon_0\|_2^k < \infty$, $k \in \mathbb{N}$ and let $\{A_{t,s}^{(T)}\}_{s \in \mathbb{Z}}$ be a sequence of operators in $S_\infty(H)$ satisfying $\sum_s \|A_{t,s}^{(T)}\|_\infty < \infty$ for all $t = 1, \dots, T$ and $T \in \mathbb{N}$. Then the process $X_{t,T}^{(N)} = \sum_{|s| \leq N} A_{t,s}^{(T)} \varepsilon_{t-s}$ has the following properties:*

- (i) $X_{t,T}^{(N)}$ converges to a process $X_{t,T}$ in $L_H^k(\Omega, \mathbb{P})$;
- (ii) $\text{cum}(X_{t_1,T}, \dots, X_{t_k,T}) = \left(\sum_{s_1 \in \mathbb{Z}} A_{t_1,s_1}^{(T)} \otimes \cdots \otimes \sum_{s_k \in \mathbb{Z}} A_{t_k,s_k}^{(T)} \right) \text{cum}(\varepsilon_{t_1-s_1}, \dots, \varepsilon_{t_k-s_k})$, where the convergence is with respect to $\|\cdot\|_2$.

Proof of Proposition B2.4. For the first equality, we need to show that

$$\lim_{N \rightarrow \infty} \mathbb{E} \|X_{t,T}^{(N)} - X_{t,T}\|_2^k = 0.$$

We will do this by demonstrating that the the tail series $X_{t,T}^{-(N)} = \sum_{s=N+1}^M A_{t,s}^{(T)} \varepsilon_{t-s}$ converges. Since $\|A_{t,s}^{(T)} \varepsilon_t\|_2 \leq \|A_{t,s}^{(T)}\|_\infty \|\varepsilon_t\|_2$, an application of the generalized Hölder's Inequality yields

$$\begin{aligned}
 \mathbb{E} \|X_{t_1,T}^{-(N)}\|_2^k &\leq \sum_{s_1, \dots, s_k = N+1}^M \|A_{t_1,s_1}^{(T)}\|_\infty \cdots \|A_{t_k,s_k}^{(T)}\|_\infty \mathbb{E} [\|\varepsilon_{t_1-s_1}\|_2 \cdots \|\varepsilon_{t_k-s_k}\|_2] \\
 &\leq \sum_{|s_1|, \dots, |s_k| > N} \|A_{t_1,s_1}^{(T)}\|_\infty \cdots \|A_{t_k,s_k}^{(T)}\|_\infty \mathbb{E} [\|\varepsilon_{t_1-s_1}\|_2^k \cdots \|\varepsilon_{t_k-s_k}\|_2^k]^{1/k} \\
 &\leq \left(\sum_{|s| > N} \|A_{t,s}^{(T)}\|_\infty \right)^k \mathbb{E} \|\varepsilon_0\|_2^k < \infty,
 \end{aligned}$$

uniformly in M . Hence, $\lim_{N \rightarrow \infty} (\mathbb{E} \|X_{t,T}^{-\langle N \rangle}\|_2^k)^{1/k} = 0$.

We now prove (2). By Proposition B2.1 and (1), we have

$$\begin{aligned} & \text{cum}(A_{t_1, s_1}^{(T)} \varepsilon_{t_1 - s_1}, \dots, A_{t_k, s_k}^{(T)} \varepsilon_{t_k - s_k}) \\ &= \left(A_{t_1, s_1}^{(T)} \otimes \dots \otimes A_{t_k, s_k}^{(T)} \right) \text{cum}(\varepsilon_{t_1 - s_1}, \dots, \varepsilon_{t_k - s_k}). \end{aligned}$$

It is therefore sufficient to show that

$$\begin{aligned} & \text{cum}\left(\sum_{s_1 \in \mathbb{Z}} A_{t_1, s_1}^{(T)} \varepsilon_{t_1 - s_1}, \dots, \sum_{s_k \in \mathbb{Z}} A_{t_k, s_k}^{(T)} \varepsilon_{t_k - s_k} \right) \\ &= \sum_{s_1, \dots, s_k \in \mathbb{Z}} \text{cum}(A_{t_1, s_1}^{(T)} \varepsilon_{t_1 - s_1}, \dots, A_{s_k, t_k, T} \varepsilon_{t_k - s_k}). \end{aligned}$$

Let $\{\psi_l\}_{l \in \mathbb{N}}$ be an orthonormal basis of H . Then $\{\psi_{l_1} \otimes \dots \otimes \psi_{l_k}\}_{l_1, \dots, l_k \geq 1}$ forms an orthonormal basis $\bigotimes_{j=1}^k H$. For the partial sums

$$\sum_{s_j=1}^N A_{t_j, s_j}^{(T)} \varepsilon_{t_j - s_j}, \quad j = 1, \dots, k,$$

we obtain by virtue of the triangle inequality, the Cauchy-Schwarz Inequality and generalized Hölder Inequality

$$\begin{aligned} \mathbb{E} \left\| \prod_{j=1}^k \sum_{s_j=1}^N A_{t_j, s_j}^{(T)} \varepsilon_{t_j - s_j} \psi_l \right\|_1 &\leq \prod_{j=1}^k \mathbb{E} \left\| \sum_{s_j=1}^N A_{t_j, s_j}^{(T)} \varepsilon_{t_j - s_j} \psi_l \right\|_1 \\ &\leq \left(\sup_{t, T} \sum_{s \in \mathbb{Z}} \|A_{t, s}^{(T)}\|_\infty \right)^k \mathbb{E} \|\varepsilon_0\|_2^k < \infty. \end{aligned}$$

The result now follows by the dominated convergence theorem. \square

B2.3. Existence of the stochastic integral

In order to provide sufficient conditions for local stationarity of functional processes in terms of spectral representations, we turn to investigating the conditions under which stochastic integrals $\int_{-\pi}^{\pi} U_\omega dZ_\omega$ for $S_\infty(H_{\mathbb{C}})$ -valued functions U_ω are well-defined. For this, let μ be a measure on the interval $[-\pi, \pi]$ given by

$$\mu(A) = \int_A \|\mathcal{F}_\omega\|_1 d\omega \tag{B2.19}$$

for all Borel sets $A \subseteq [-\pi, \pi]$, and let $\mathcal{B}_\infty = L_{S_\infty(H_{\mathbb{C}})}^2([-\pi, \pi], \mu)$ be the corresponding Bochner space of all strongly measurable functions $U : [-\pi, \pi] \rightarrow S_\infty(H_{\mathbb{C}})$ such that

$$\|U\|_{\mathcal{B}_\infty}^2 = \int_{\Pi} \|U_\omega\|_\infty^2 d\mu(\omega) < \infty. \tag{B2.20}$$

Panaretos and Tavakoli (2013a) showed that the stochastic integral is well defined in $\mathbb{H}_{\mathbb{C}}$ for operators that belong to the Bochner space $\mathcal{B}_2 = L_{S_2(H_{\mathbb{C}})}^2([-\pi, \pi],$

μ), which is a subspace of \mathcal{B}_∞ . In particular, it contains all functions $U : [-\pi, \pi] \rightarrow S_2(H_{\mathbb{C}})$ of the form

$$U_\omega = g(\omega) I + A_\omega,$$

where g and A are, respectively, \mathbb{C} and $S_2(H_{\mathbb{C}})$ -valued functions that are both càdlàg with a finite number of jumps and A additionally satisfies $\int_{\Pi} \|A_\omega\|_2^2 d\mu(\omega) < \infty$. Here, continuity in $S_2(H_{\mathbb{C}})$ is meant with respect to the operator norm $\|\cdot\|_\infty$. Because the space \mathcal{B}_2 is too restrictive to include interesting processes such as general functional autoregressive processes, we first show that the integral is properly defined in $\mathbb{H}_{\mathbb{C}}$ for all elements of \mathcal{B}_∞ . To do so, consider the subspace $\mathcal{Q}_0 \subset \mathcal{B}_\infty$ of step functions spanned by elements $U \mathbf{1}_{[\alpha, \beta]}$ for $U \in S_\infty(H_{\mathbb{C}})$ and $\alpha < \beta \in [-\pi, \pi]$. Additionally, denote its closure by $\mathcal{Q} = \overline{\mathcal{Q}_0}$. Define then the mapping $\mathcal{T} : \mathcal{Q}_0 \mapsto \mathbb{H}_{\mathbb{C}}$ by linear extension of

$$\mathcal{T}(U \mathbf{1}_{[\alpha, \beta]}) = U(Z_\beta - Z_\alpha). \quad (\text{B2.21})$$

The following lemma shows that the image of \mathcal{T} is in $\mathbb{H}_{\mathbb{C}}$.

Lemma B2.5. *Let X_t be a functional process with spectral representation $X_t = \int_{-\pi}^{\pi} e^{i\omega t} dZ_\omega$ for some functional orthogonal increment process Z_ω that satisfies $\mathbb{E}\|Z_\omega\|_2^2 = \int_{-\pi}^{\omega} \|\mathcal{F}_\lambda\|_1 d\lambda$. Then for $U_1, U_2 \in S_\infty(H_{\mathbb{C}})$ and $\alpha, \beta \in [-\pi, \pi]$*

$$(i) \quad \langle U_1 Z_\alpha, U_2 Z_\beta \rangle_{\mathbb{H}_{\mathbb{C}}} = \text{tr} \left(U_1 \left[\int_{-\pi}^{\alpha \wedge \beta} \mathcal{F}_\omega d\omega \right] U_2^\dagger \right)$$

and

$$(ii) \quad \|U_1 Z_\alpha\|_{\mathbb{H}_{\mathbb{C}}}^2 \leq \|U_1\|_\infty^2 \int_{-\pi}^{\alpha} \|\mathcal{F}_\lambda\|_1 d\lambda.$$

Proof of Lemma B2.5. Firstly, we note that by Cauchy-Schwarz inequality

$$\begin{aligned} \mathbb{E} \int_0^1 |U_1 Z_\alpha(\tau) U_2 Z_\beta(\tau)| d\tau &\leq \mathbb{E} \|U_1 Z_\alpha\|_2 \|U_2 Z_\beta\|_2 \leq \|U_1\|_\infty \|U_2\|_\infty \mathbb{E} \|Z_\alpha\|_2 \|Z_\beta\|_2 \\ &\leq \|U_1\|_\infty \|U_2\|_\infty \int_{-\pi}^{\alpha \wedge \beta} \|\mathcal{F}_\lambda\|_1 d\lambda < \infty. \end{aligned} \quad (\text{B2.22})$$

Secondly, $U_1 Z_\alpha$ and $U_2 Z_\beta$ are elements in $H_{\mathbb{C}}$ and therefore the (complete) tensor product $U_1 Z_\alpha \otimes U_2 \overline{Z_\beta}$ belongs to $S_2(H_{\mathbb{C}})$. By Proposition B1.7, it is thus a kernel operator with kernel $[U_1 Z_\alpha \otimes U_2 \overline{Z_\beta}](\tau, \sigma) = U_1 Z_\alpha(\tau) \overline{U_2 Z_\beta(\sigma)}$. An application of Fubini's Theorem yields

$$\begin{aligned} \mathbb{E} \int_0^1 U_1 Z_\alpha(\tau) \overline{U_2 Z_\beta(\tau)} d\tau &= \int_0^1 E(U_1 Z_\alpha \otimes U_2 \overline{Z_\beta})(\tau, \tau) d\tau \\ &= \int_0^1 (U_1 \otimes U_2) E(Z_\alpha \otimes \overline{Z_\beta})(\tau, \tau) d\tau = \int_0^1 (U_1 \otimes U_2) \int_{-\pi}^{\alpha \wedge \beta} \mathcal{F}_\omega d\omega(\tau, \tau) d\tau \\ &= \int_0^1 U_1 \int_{-\pi}^{\alpha \wedge \beta} \mathcal{F}_\omega d\omega(\tau, \tau) U_2^\dagger d\tau, \end{aligned}$$

where the second equality follows because the expectation commutes with bounded operators for integrable random functions (Proposition B2.1) and the last equality follows from the identity (B1.3) of definition B1.2. This shows the first result of Lemma B2.5. The second result follows straightforwardly from (B2.22). \square

It is easily seen from the previous lemma that for $\lambda_1 > \lambda_2 \geq \lambda_3 > \lambda_4$

$$\langle U_1(Z_{\lambda_1} - Z_{\lambda_2}), U_2(Z_{\lambda_3} - Z_{\lambda_4}) \rangle_{\mathbb{H}_{\mathbb{C}}} = 0,$$

demonstrating orthogonality of the increments is preserved. Since every element $U_n \in \mathcal{Q}_0$ can be written as $\sum_{j=1}^n U_j \mathbf{1}_{[\lambda_j, \lambda_{j+1})}$ the lemma moreover implies

$$\begin{aligned} \|\mathcal{T}(U)\|_{\mathbb{H}_{\mathbb{C}}}^2 &= \sum_{j,k=1}^n \langle U_j(Z_{\lambda_{j+1}} - Z_{\lambda_j}), U_k(Z_{\lambda_{k+1}} - Z_{\lambda_k}) \rangle_{\mathbb{H}_{\mathbb{C}}} \\ &= \sum_{j=1}^n \|U_j(Z_{\lambda_{j+1}} - Z_{\lambda_j})\|_{\mathbb{H}_{\mathbb{C}}}^2 \\ &\leq \sum_{j=1}^n \|U_j\|_{\infty}^2 \int_{\lambda_j}^{\lambda_{j+1}} \|\mathcal{F}_{\alpha}\|_1 d\alpha = \|U\|_{\mathcal{B}_{\infty}}^2. \end{aligned}$$

The mapping $\mathcal{T} : \mathcal{Q}_0 \mapsto \mathbb{H}_{\mathbb{C}}$ is therefore continuous. Together with the completeness of the space $\mathbb{H}_{\mathbb{C}}$ this establishes that, for every sequence $\{U_n\}_{n \geq 1} \subset \mathcal{Q}_0$ converging to some element $U \in \overline{\mathcal{Q}}$, the sequence $\{\mathcal{T}(U_n)\}_{n \geq 1}$ forms a Cauchy sequence in $\mathbb{H}_{\mathbb{C}}$ with limit $\mathcal{T}(U) = \lim_{n \rightarrow \infty} \mathcal{T}(U_n)$. By linearity and continuity of the mapping \mathcal{T} , the limit is independent of the choice of the sequence. Furthermore, since \mathcal{Q}_0 is the subspace spanned by step functions that are square integrable on $[-\pi, \pi]$ with respect to the finite measure μ and hence is dense in $L^2_{S_{\infty}(\mathbb{H}_{\mathbb{C}})}([-\pi, \pi], \mu)$, we have $\mathcal{B}_{\infty} \subseteq \overline{\mathcal{Q}}$. Since $\|\mathcal{T}(U)\|_{\mathbb{H}_{\mathbb{C}}} \leq \|U\|_{\mathcal{B}_{\infty}}$, the above extension is well-defined for all $U \in \mathcal{B}_{\infty}$.

Appendix B3: Data taper

In order to show convergence of the higher order cumulants of the estimator in (33), we will make use of two lemmas from Dahlhaus (1997) (Lemma A.4 and A.5 resp.). Both rely on the function $L_T : \mathbb{R} \rightarrow \mathbb{R}, T \in \mathbb{R}^+$, which is the 2π -periodic extension of

$$L_T(\lambda) = \begin{cases} T & \text{if } |\lambda| \leq 1/T, \\ 1/|\lambda| & \text{if } 1/T \leq |\lambda| \leq \pi. \end{cases} \quad (\text{B3.23})$$

The function L_T satisfies some nice properties. The following lemma lists those required in the current paper:

Lemma B3.1. *Let $k, l, T \in \mathbb{N}, \lambda, \alpha, \omega, \mu, \gamma \in \mathbb{R}$ and $\Pi : (-\pi, \pi]$. The following inequalities then hold with a constant C independent of T .*

- (i) $L_T(\lambda)$ is monotone increasing in T and decreasing in $\lambda \in [0, \pi]$;
- (ii) $|\lambda|L_T(\lambda) \leq C$ for all $|\lambda| \leq \pi$;
- (iii) $\int_{\Pi} L_T(\lambda)d\lambda \leq C \log T$;
- (iv) $\int_{\Pi} L_T(\lambda)^k d\lambda \leq C T^{k-1}$ for $k > 1$;
- (v) $\int_{\Pi} L_T(\alpha - \lambda) L_T(\lambda + \gamma) d\lambda \leq C L_T(\alpha + \gamma) \log T$.

In addition, we also make use of Lemma 2 from Eichler (2007).

Lemma B3.2. *Let $\{P_1, \dots, P_m\}$ be an indecomposable partition of the table*

$$\begin{array}{cc} \alpha_1 & -\alpha_1 \\ \vdots & \vdots \\ \alpha_n & -\alpha_n \end{array}$$

with $n \geq 3$. For $P_j = \{\gamma_{j1}, \dots, \gamma_{jd_j}\}$, let $\bar{\gamma}_j = \gamma_{j1} + \dots + \gamma_{jd_j}$.

- (i) *If $m = n$ then for any $n - 2$ variables $\alpha_{i_1}, \dots, \alpha_{i_{n-2}}$ we have*

$$\int_{\Pi^{k-2}} \prod_{j=1}^n L_T(\bar{\gamma}_j) d\alpha_{i_1} \cdots d\alpha_{i_{n-2}} \leq C L_N(\alpha_{i_{n-1}} \pm \alpha_{i_n})^2 \log(T)^{n-2}.$$

- (ii) *If $m < n$ then there exists $n - 2$ variables $\alpha_{i_1}, \dots, \alpha_{i_{n-2}}$ such that*

$$\int_{\Pi^{k-2}} \prod_{j=1}^n L_T(\bar{\gamma}_j) d\alpha_{i_1} \cdots d\alpha_{i_{n-2}} \leq C T \log(T)^{n-2}.$$

The usefulness of the L_T function stems from the fact that it gives an upperbound for the function $H_{k,N}$ which was defined in section 5. Namely, we have

$$|H_{k,N}^{(\lambda)}| \leq L_N(\lambda), \forall k \in \mathbb{N}. \tag{B3.24}$$

We also require an adjusted version of Lemma A.5 of Dahlhaus (1997):

Lemma B3.3. *Let $N, T \in \mathbb{N}$. Suppose h is a data-taper of bounded variation and let the operator-valued function $G_u : [0, 1] \rightarrow S_p(H)$ be continuously differentiable in u such that $\|\frac{\partial G_u}{\partial u}\|_p < \infty$ uniformly in u . Then we have for $0 \leq t \leq N$,*

$$\begin{aligned} H_N(G_{\frac{\bullet}{T}}, \omega) &= H_N(\omega)G_{\frac{t}{T}} + O\left(\sup_u \left\| \frac{\partial}{\partial u} G_u \right\|_p \frac{N}{T} L_N(\omega)\right) \\ &= O\left(\sup_{u \leq N/T} \|G_u\|_p \frac{N}{T} L_N(\omega) + \sup_u \left\| \frac{\partial}{\partial u} G_u \right\|_p \frac{N}{T} L_N(\omega)\right), \end{aligned} \tag{B3.25}$$

where $H_N(G_{\bullet}, \omega)$ is as in (30) The same holds if $G_{\frac{\bullet}{T}}$ on the left hand side is replaced by operators $G_{\bullet}^{(T)}$ for which $\sup_s \|G_{\bullet}^{(T)} - G_{\frac{\bullet}{T}}\|_p = O(\frac{1}{T})$.

Proof. Summation by parts gives

$$\begin{aligned} H_N(G_{\frac{\bullet}{T}}, \omega) - H_N(\omega)G_{\frac{t}{T}} &= \sum_{s=0}^{N-1} [G_{\frac{s}{T}} - G_{\frac{t}{T}}] h_{s,N} e^{-i\omega s} \\ &= - \sum_{s=0}^{N-1} [G_{\frac{s}{T}} - G_{\frac{s-1}{T}}] H_s(h_{\bullet,N}, \omega) \\ &\quad + [G_{\frac{N-1}{T}} - G_{\frac{t}{T}}] H_N(\omega). \end{aligned}$$

It has been shown in Dahlhaus (1988) that $|H_s(h_{\bullet,N}, \omega)| \leq KL_s(\omega) \leq KL_N(\omega)$. The result in (B3.25) then follows since

$$\|G_b - G_a\|_p \leq \sup_{a < \xi < b} \left\| \frac{\partial}{\partial u} G_u \Big|_{u=\xi} \right\|_p |b - a|, \quad a, b \in \mathbb{R},$$

by the Mean Value Theorem. The lemma holds additionally for operators $G_{\bullet}^{(T)}$ that satisfy $\sup_s \|G_{\bullet}^{(T)} - G_{\frac{\bullet}{T}}\|_p = O(\frac{1}{T})$. This is a consequence of Minkowski's inequality since

$$\begin{aligned} &\|H_N(G_{\bullet}^{(T)} - G_{\frac{\bullet}{T}}, \omega) + H_N(G_{\frac{\bullet}{T}}, \omega)\|_p \\ &= \|H_N(G_{\bullet}^{(T)} - G_{\frac{\bullet}{T}}, \omega)\|_p + \|H_N(G_{\frac{\bullet}{T}}, \omega)\|_p \\ &= O\left(\frac{N}{T} + L_N(\lambda)\right) = O(L_N(\lambda)). \end{aligned} \tag{B3.26}$$

Hence, the replacement error is negligible compared to the error of B3.25. \square

If $p = 2$, the above implies that the kernel function $g_u \in H_{\mathbb{C}}^2$ of G_u satisfies

$$\begin{aligned} \|H_N(g_{\frac{\bullet}{T}}, \omega) - H_N(\omega)g_{\frac{t}{T}}\|_2 &= R_{1,N}, \\ \|H_N(g_{\frac{\bullet}{T}}, \omega)\| &= R_{2,N} + R_{1,N}, \end{aligned}$$

where

$$\begin{aligned} \|R_{1,N}\|_2 &= O\left(\sup_u \left\| \frac{\partial}{\partial u} g_u \right\|_p \frac{N}{T} L_N(\omega)\right), \\ \|R_{2,N}\|_2 &= O\left(\sup_{u \leq N/T} \|g_u\|_p \frac{N}{T} L_N(\omega)\right). \end{aligned} \tag{B3.27}$$

Similarly if $g_{\frac{\bullet}{T}}$ on the left hand side is replaced by the kernel function $g_{\bullet}^{(T)} \in H_{\mathbb{C}}^2$ of $G_{\bullet}^{(T)}$. If the kernels are bounded uniformly in their functional arguments, Lemma A.5 of Dahlhaus (1997) is pointwise applicable.

References

ANTONIADIS, A., PAPANODITIS, E. and SAPATINAS, T. (2006). A functional wavelet-kernel approach for time series prediction. *Journal of the Royal Statistical Society Ser. B* **86** 837–857. [MR2301297](#)

- ANTONIADIS, A. and SAPATINAS, T. (2003). Wavelet methods for continuous time predictions using Hilbert-valued autoregressive processes. *Journal of Multivariate Analysis* **87** 133–158. [MR2007265](#)
- AUE, A., DUBART NORINHO, D. and HÖRMANN, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association* **110** 378–392. [MR3338510](#)
- AUE, A. and VAN DELFT, A. (2017). Testing for stationarity of functional time series in the frequency domain. Preprint, <https://arxiv.org/pdf/1701.01741.pdf>.
- AUE, A., GABRYS, R., HORVÁTH, L. and KOKOSZKA, P. (2009). Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis* **100** 2254–2269. [MR2560367](#)
- BENKO, M., HÄRDLE, W. and KNEIP, A. (2009). Common functional component analysis. *Annals of Statistics* **37** 1–34. [MR2488343](#)
- BERKES, I., GABRYS, R., HORVÁTH, L. and KOKOSZKA, P. (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society Ser. B* **71** 927–946. [MR2750251](#)
- BESSE, P. and RAMSAY, J. O. (1986). Principal components analysis of samples functions. *Psychometrika* **51** 285–311. [MR0848110](#)
- BOGACHEV, V. I. and MIFTAKHOV, A. F. (2015). On weak convergence of finite-dimensional and infinite-dimensional distributions of random processes. Preprint 15102, Collaborative Research Centre 701, University of Bielefeld. [MR3571407](#)
- BOSQ, D. (2000). *Linear Processes in Function Spaces*. Springer, New York. [MR1783138](#)
- BOSQ, D. (2002). Estimation of mean and covariance operator of autoregressive processes in Banach spaces. *Statistical inference for Stochastic Processes* **5** 287–306.
- BOSQ, D. and BLANKE, D. (2007). *Inference and Prediction in Large Dimensions*. Wiley, Chichester.
- BOWSER, C. G. and MEEKS, R. (2008). The Dynamics of Economic Functions: Modeling and Forecasting the Yield Curve. *Journal of the American Statistical Association* **103** 1419–1437.
- BRILLINGER, D. R. (1981). *Time Series: Data Analysis and Theory*. McGraw Hill, New York.
- BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*. Springer, New York.
- CARDOT, H. and SARDA, P. (2006). Linear regression models for functional data. In *The Art of Semiparametrics* (S. Sperlich and G. Aydinli, eds.) 49–66. Springer, Heidelberg.
- CREMERS, H. and KADELKA, D. (1986). On weak convergence of integral functionals of stochastic processes with applications to processes taking paths in L_p^E . *Stochastic Processes and Their Applications* 305–317.
- DAHLHAUS, R. (1983). Spectral analysis with tapered data. *Journal of Time Series Analysis* **4** 163–175.

- DAHLHAUS, R. (1988). Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Annals of Statistics* **16** 808–841.
- DAHLHAUS, R. (1996a). On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Process and their Applications* **62** 139–168. [MR1388767](#)
- DAHLHAUS, R. (1996b). *Asymptotic statistical inference for nonstationary processes with evolutionary spectra*. In: Robinson, P. M., Rosenblatt, M. (Eds.), Athens Conference on Applied Probability and Time Series Analysis, vol. II. Springer-Verlag, New York. [MR1466743](#)
- DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics* **25** 1–37. [MR1429916](#)
- DAHLHAUS, R. and SUBBA RAO, S. (2006). Statistical inference for time-varying ARCH processes. *Annals of Statistics* **34** 1077–1114. [MR2278352](#)
- DAMON, J. and GUILLAS, S. (1982). The inclusion of exogeneous variables in functional autoregressive ozone forecasting. *Environmetrics* **13** 759–774.
- 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. *Journal of Multivariate Analysis* **12** 136–154. [MR0650934](#)
- DEHLING, H. and SHARIPOV, O. S. (2005). Estimation of mean and covariance operator for Banach space valued autoregressive processes with independent innovations. *Statistical inference for stochastic processes* **8** 137–149. [MR2121674](#)
- DUNFORD, N. and SCHWARTZ, J. T. (1958). *Linear Operators, Part I: General Theory*. Wiley, New York. [MR0117523](#)
- EDWARDS, R. (1967). *Fourier series: a modern introduction*. Holt, Rinehart and Winston, New York. [MR0222538](#)
- EICHLER, M. (2007). A frequency-domain based test for independence between stationary time series. *Metrika* **65** 133–157. [MR2288054](#)
- ERBAS, B., HYNDMAN, R. J. and GERTIG, D. M. (2007). Forecasting age-specific breast cancer mortality using functional data models. *Statistics in Medicine* **26** 458–470. [MR2312725](#)
- FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis*. Springer, New York. [MR2229687](#)
- GABRYS, R., HÖRMANN, S. and KOKOSZKA, P. (2010). Monitoring the intraday volatility pattern. Technical Report, Utah State University.
- GRENANDER, U. (1981). *Abstract Inference*. Wiley Series in Probability and Mathematical Statistics **IX**. John Wiley & Sons, New York. [MR0599175](#)
- GRINBLAT, L. Š. (1976). A limit theorem for measurable random processes and its applications. *Proceedings of the American Mathematical Society* **61** 371–376. [MR0423450](#)
- HAYS, S., SHEN, H. and HUANG, J. Z. (2012). Functional dynamic factor models with application to yield curve forecasting. *Annals of Applied Statistics* **6** 870–894. [MR3012513](#)
- HÖRMANN, S., HORVÁTH, L. and REEDER, R. (2013). A functional version of the ARCH model. *Econometric Theory* **29** 267–288. [MR3042756](#)

- HÖRMANN, S., KIDZIŃSKI, L. and HALLIN, M. (2015). Dynamic functional principal components. *The Royal Statistical Society: Series B* **77** 319–348. [MR3310529](#)
- HÖRMANN, S. and KOKOSZKA, P. (2010). Weakly dependent functional data. *The Annals of Statistics* **38** 1845–1884. [MR2662361](#)
- HORVÁTH, L., HUŠKOVÁ, M. and KOKOSZKA, P. (2010). Testing the stability of the functional autoregressive process. *Journal of Multivariate Analysis* **21** 352–367. [MR2564345](#)
- HYNDMAN, R. J. and BOOTH, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting* **24** 323–342.
- HYNDMAN, R. J. and ULLAH, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis* **51** 4942–4956. [MR2364551](#)
- KARHUNEN, K. (1947). Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Annales Academiae Scientiarum Fennicae, Ser. A.I. Math.-Phys.* **37** 1–79. [MR0023013](#)
- KLEFFE, J. (1973). Principal components of random variables with values in a separable Hilbert space. *Mathematische Operationsforschung Statistik* **5** 391–406. [MR0391402](#)
- KÜNSCH, H. R. (1995). A note on causal solution for locally stationary AR processes. Preprint, ETH Zurich.
- LOÈVE, M. (1948). *Fonctions aléatoires du second ordre. Supplement to P. Lévy, Processus stochastiques et Mouvement Brownian*. Gauthier-Villars, Paris. [MR0190953](#)
- MARTIN, W. and FLANDRIN, P. (1985). Wigner-Ville spectral analysis of non-stationary processes. *IEEE Trans. Signal Process* **33** 1461–1470.
- MAS, A. (2000). Estimation d’opérateurs de corrélation de processus linéaires fonctionnels: lois limites, déviations modérées, PhD thesis, Université Paris VI.
- MÜLLER, H. G. and YAO, F. (2008). Functional additive models. *Journal of the American Statistical Association* **103** 1534–1544. [MR2504202](#)
- MURPHY, G. J. (1990). *C*-algebras and Operator Theory*. Academic Press. [MR1074574](#)
- NELSON, E. (1969). *Topics in Dynamics, Volume I: Flows*. University Press, Princeton.
- PANARETOS, V. M. and TAVAKOLI, S. (2013a). Cramér-Karhunen-Loève representation and harmonic analysis of functional time series. *Stochastic Processes and their Applications* **123** 2779–2807.
- PANARETOS, V. M. and TAVAKOLI, S. (2013b). Fourier analysis of stationary time series in function space. *The Annals of Statistics* **41** 568–603.
- PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series, Vol. 2*. Academic Press, London.
- RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York.

- VAN DELFT, A. and EICHLER, M. (2015). Data-adaptive estimation of time-varying spectral densities. Preprint Maastricht University, <https://arxiv.org/pdf/1512.00825.pdf>.
- VAN DELFT, A. and EICHLER, M. (2018). A note on Herglotz's theorem for time series on function spaces. ArXiv preprint.
- VOGT, M. (2012). Nonparametric regression for locally stationary time series. *Annals of Statistics* **40** 2601–2633.
- YAO, F., MÜLLER, H. G. and WANG, L. J (2005). Functional linear regression for longitudinal data. *Annals of Statistics* **33** 2873–2903.