arXiv:1511.04020v3 [stat.ME] 29 Aug 2017

Detecting and dating structural breaks in functional data without dimension reduction*

Alexander Aue† Gregory Rice‡ Ozan Sönmez†

August 31, 2017

Abstract

Methodology is proposed to uncover structural breaks in functional data that is "fully functional" in the sense that it does not rely on dimension reduction techniques. A thorough asymptotic theory is developed for a fully functional break detection procedure as well as for a break date estimator, assuming a fixed break size and a shrinking break size. The latter result is utilized to derive confidence intervals for the unknown break date. The main results highlight that the fully functional procedures perform best under conditions when analogous fPCA based estimators are at their worst, namely when the feature of interest is orthogonal to the leading principal components of the data. The theoretical findings are confirmed by means of a Monte Carlo simulation study in finite samples. An application to annual temperature curves illustrates the practical relevance of the proposed procedures.

Keywords: Change-point analysis; Functional data; Functional principal components; Functional time series; Structural breaks; Temperature data

MSC 2010: Primary: 62G99, 62H99, Secondary: 62M10, 91B84

1 Introduction

This paper considers the problem of detecting and dating structural breaks in functional time series data, and hence lies at the intersection of functional data analysis (FDA) and structural breaks analysis for dependent observations. FDA has witnessed an upsurge in research contributions in the past decade. These are documented, for example, in the comprehensive books by Ramsay and Silverman (2005) and Ferraty and Vieu (2010). Research concerned with structural breaks has a longstanding tradition in both the statistics and econometrics communities. Two recent reviews by Aue and Horváth (2013) and Horváth and Rice (2014) highlight newer developments, the first with a particular focus on time series.

*This research was partially supported by NSF grants DMS 1209226, DMS 1305858 and DMS 1407530. The authors would like to thank the editor, David Dunson, the associate editor and three referees for constructive criticism that has led to a substantial improvement in the quality of this paper.

†Department of Statistics, University of California, Davis, CA 95616, USA, email: [aaue,osonmez]@ucdavis.edu

‡Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada, email: grice@uwaterloo.ca

Early work in functional structural break analysis dealt primarily with random samples of independent curves, the question of interest being whether all curves have a common mean function or whether there are two or more segments of the data that are homogeneous within but heterogeneous without. Berkes et al. (2009) developed statistical methodology to test the null hypothesis of no structural break against the alternative of a (single) break in the mean function assuming that the error terms are independent and identically distributed curves. Aue et al. (2009) quantified the large-sample behavior of a break date estimator under a similar set of assumptions. The work in these two papers was generalized by Aston and Kirch (2012a, b) and Torgovitski (2016) to include functional time series exhibiting weak dependence into the modeling framework. In Zhang et al. (2011), a structural break detection procedure for serially correlated functional time series data is proposed that is based on the self-normalization approach of Shao and Zhang (2010). Structual break detection in the context of functional linear models is considered in Aue et al. (2014) and for spatially distributed functional data in Gromenko et al. (2016). Smooth deviations from stationarity of functional time series in the frequency domain were studied in Aue and van Delft (2017+).

Most of the procedures in FDA, such as those presented in the above cited papers, are based on dimension reduction techniques, primarily using the widely popular functional principal components analysis (fPCA), by which the functional variation in the data is projected onto the directions of a small number of principal curves, and multivariate techniques are then applied to the resulting sequence of score vectors. This is also the case in functional structural break detection, in which after an initial fPCA step multivariate structural break theory is utilized. Despite the fact that functional data are, at least in principle, infinite dimensional, the state of the art in FDA remains to start the analysis with an initial dimension reduction procedure.

Dimension reduction approaches, however, automatically incur a loss of information, namely all information about the functional data that is orthogonal to the basis onto which it is projected. This weakness is easily illustrated in the context of detecting and dating structural breaks in the mean function: if the function representing the mean break is orthogonal to the basis used for dimension reduction, there cannot be a consistent test or estimator for the break date in that basis. This point will be further illustrated by theoretical arguments and in comprehensive numerical studies in Section 4, where other more subtle differences between the competing methods will be highlighted.

The main purpose of this paper is then to develop methodology for detecting and dating structural breaks in functional data without the application of dimension reduction techniques. Here, fully functional test statistics and break date estimators are studied, and their asymptotic theory is developed under the assumption that the model errors satisfy a general weak dependence condition. This theory illuminates a number of potential advantages of the fully functional procedures. For example, it is shown that when the direction of the break is orthogonal to the leading principal components of the data, the estimation of the mean break is asymptotically *improved* when using the fully functional estimator compared to mean breaks of the same size that are contained in the leading principal components. This contrasts with fPCA based techniques in

which such mean breaks are more difficult, if not impossible, to detect, even given arbitrarily large sample sizes. In addition, the assumptions required for the fully functional theory are weaker than the ones used in Aue et al. (2009) and Aston and Kirch (2012a, b), as convergence of the eigenvalues and eigenfunctions of the empirical covariance operator to the eigenvalues of the population covariance operator do not have to be accounted for. These assumptions are typically formulated as finiteness of fourth moment conditions. The relaxation obtained here may be particularly useful for applications to intra-day financial data such as the one-minute log-returns on Microsoft stock discussed in the online supplement Aue et al. (2017+) accompanying this article.

The application presented in Section 5 is concerned with annual temperature curves recorded across different measuring stations in Australia. Structural breaks in these temperature curves are detected with both fPCA and fully functional methods. The sample covariance operator associated with the data has eigenvalues that decay remarkably slowly. A somewhat peculiar feature of fPCA methods in this setting, studied as part of the simulation experiment, is a loss of accuracy in break dating even when the break function loads almost exclusively on the first component. A similar effect is found in the data, where fPCA-based break dates can occur outside of the confidence intervals provided by the fully functional procedure.

Most closely related to the present work are Fremdt et al. (2014), who considered structural break detection using fPCA under an increasing number of projections. Horváth et al. (2014) developed a functional analog of the KPSS test statistic for the purpose of stationarity testing that does not rely on dimension reduction. Sharipov et al. (2016) considered a bootstrap procedure for measuring the significance of the norms of functional CUSUM processes with applications to testing for a structural break in the means of functional observations and in the distribution function of scalar time series observations under a mixing assumption, generalizing the result for the independent, identically distributed case put forward in Tsudaka and Nishiyama (2014). Bucchia and Wendler (2016+) studied general bootstrap procedures for structural break analysis in Hilbert space-valued random fields.

The remainder of the paper is organized as follows. Testing procedures and a break date estimator are introduced in Section 2, along with the main asymptotic results of the paper. The asymptotic properties developed in this section are accompanied by implementation details given in Section 3 and results from a comprehensive simulation study in Section 4. The application to temperature curves is given in Section 5, and Section 6 concludes. Proofs of the main results as well as additional empirical illustrations of the proposed methodology are provided in the online supplement Aue et al. (2017+), henceforth referred to simply as the online supplement. In addition, an R package, `fChange`, has been developed to supplement this article and is available on the Comprehensive R Archive Network. The package contains implementations of all of the testing and estimation procedures introduced below, see Sönmez et al. (2017).

2 Main results

In this paper, a functional data model allowing for a mean function break is considered. It is assumed that the observations $X_1, \ldots, X_n$ are generated from the model

$$X_i = \mu + \delta \mathbb{1}\{i > k^*\} + \varepsilon_i, \qquad i \in \mathbb{Z}, \tag{2.1}$$

where $k^* = \lfloor \theta n \rfloor$, with $\theta \in (0, 1)$, labels the unknown time of the mean break parameterized in terms of the sample size n, μ is the baseline mean function that is distorted by the addition of δ after the break time k^*, $\mathbb{1}_A$ denotes the indicator function of the set A and $\mathbb{Z}$ the set of integers. Each X_i is a real-valued function defined without loss of generality on the unit interval $[0, 1]$. The argument $t \in [0, 1]$ will be used to refer to a particular value $X_i(t)$ of the function X_i. Correspondingly, the quantities μ, δ and ε_i on the right-hand side of (2.1) are functions on $[0, 1]$ as well. Interest is first in testing the structural break hypotheses

$$H_0 \colon \delta = 0 \qquad \text{versus} \qquad H_A \colon \delta \neq 0,$$

and then, in the event that H_A is thought to hold, estimating the break date k^*. Throughout the following assumptions are made, roughly entailing that the innovations $(\varepsilon_i \colon i \in \mathbb{Z})$ are weakly dependent, stationary functional time series. Below, let $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ denote the canonical norm and inner product in $L^2[0, 1]$.

Assumption 2.1. *The innovations $(\varepsilon_i \colon i \in \mathbb{Z})$ satisfy*

(a) there is a measurable function $g \colon S^\infty \to L^2[0, 1]$, where S is a measurable space and independent, identically distributed (iid) innovations $(\epsilon_i \colon i \in \mathbb{Z})$ taking values in S such that $\varepsilon_i = g(\epsilon_i, \epsilon_{i-1}, \ldots)$ for $i \in \mathbb{Z}$;

(b) there are ℓ-dependent sequences $(\varepsilon_{i,\ell} \colon i \in \mathbb{Z})$ such that, for some $p > 2$,

$$\sum_{\ell=0}^{\infty} \big(\mathbb{E}[\|\varepsilon_i - \varepsilon_{i,\ell}\|^p]\big)^{1/p} < \infty,$$

where $\varepsilon_{i,\ell} = g(\epsilon_i, \ldots, \epsilon_{i-\ell+1}, \epsilon^*_{i,\ell,i-\ell}, \epsilon^*_{i,\ell,i-\ell-1}, \ldots)$ with $\epsilon^*_{i,\ell,j}$ being independent copies of $\epsilon_{i,0}$ independent of $(\epsilon_i \colon i \in \mathbb{Z})$.

Processes satisfying Assumption 2.1 were termed L^p-m-approximable by Hörmann and Kokoszka (2010), and cover most stationary functional time series models of interest, including functional AR and ARMA (see Aue et al., 2015; and Bosq, 2000) and functional GARCH processes (see Aue et al., 2017). It is assumed that the underlying error innovations $(\epsilon_i \colon i \in \mathbb{Z})$ are elements of an arbitrary measurable space S. However, in many examples S is itself a function space, and the evaluation of $g(\epsilon_i, \epsilon_{i-1}, ...)$ is a functional of $(\epsilon_j \colon j \leq i)$.

The proposed methodology is based on the (scaled) functional cumulative sum (CUSUM) statistic

$$S^0_{n,k} = \frac{1}{\sqrt{n}} \bigg(\sum_{i=1}^{k} X_i - \frac{k}{n} \sum_{i=1}^{n} X_i \bigg). \tag{2.2}$$

The superscript 0 indicates the tied-down nature of the CUSUM statistic, since $S^0_{n,0} = S^0_{n,n} = 0$ (interpreting an empty sum as zero). Noting that $\|S^0_{n,k}\|$ as a function of k tends to be large at the true break date motivates the use of the max-type structural break detector

$$T_n = \max_{1\le k\le n} \|S^0_{n,k}\|^2$$

to test H_0 versus H_A. Furthermore, the break date estimator for k^* may be taken as

$$\hat{k}^*_n = \min\Big\{k\colon \|S^0_{n,k}\| = \max_{1\le k'\le n} \|S^0_{n,k'}\|\Big\}. \tag{2.3}$$

The main results of this paper concern the large-sample behavior and empirical properties of the test statistic T_n and the estimator k^*_n.

2.1 Asymptotic properties of structural break detector

Under H_0, the limiting behavior of $S^0_{n,k}$ evidently depends on that of the partial sum process of the error terms $(\varepsilon_i\colon i\in\mathbb{Z})$. As this sequence may be weakly serially correlated under Assumption 2.1, the asymptotics of the partial sum process necessarily involve the long-run covariance kernel

$$C_\varepsilon(t,t') = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\varepsilon_0(t), \varepsilon_\ell(t')) \tag{2.4}$$

of the error sequence $(\varepsilon_i\colon i\in\mathbb{Z})$. Note that C_ε constitutes the limiting covariance kernel of $\sqrt{n}$ times the centered sample mean under H_0. It is a well-defined element of $L^2[0,1]^2$ under Assumption 2.1. This kernel was considered initially in Hörmann and Kokoszka (2010). It was also studied in Panaretos and Tavakoli (2012) in the context of spectral analysis of functional time series, and in Horváth et al. (2013) in an application to the functional two sample problem. In addition, C_ε may be used to define a positive definite and symmetric Hilbert–Schmidt integral operator on $L^2[0,1]$, c_ε, given by

$$c_\varepsilon(f)(t) = \int C_\varepsilon(t,s)f(s)ds,$$

which further defines a non-increasing sequence of nonnegative eigenvalues $(\lambda_\ell\colon \ell\in\mathbb{N})$ and a corresponding orthonormal basis of eigenfunctions $(\phi_\ell\colon \ell\in\mathbb{N})$ satisfying

$$c_\varepsilon(\phi_\ell)(t) = \lambda_\ell\phi_\ell(t), \qquad \ell\in\mathbb{N}. \tag{2.5}$$

The eigenvalues of c_ε determine the limiting distribution of T_n as detailed in the following theorem.

Theorem 2.1. *Under Model 2.1, Assumption 2.1 and H_0,*

$$T_n \xrightarrow{\mathcal{D}} \sup_{0\le x\le 1} \sum_{\ell=1}^{\infty} \lambda_\ell B^2_\ell(x) \qquad (n\to\infty), \tag{2.6}$$

where $(B_\ell\colon \ell\in\mathbb{N})$ are independent and identically distributed standard Brownian bridges defined on $[0,1]$.

Theorem 2.1 points to an asymptotically validated test of H_0, namely to reject if the test statistic T_n exceeds the corresponding quantile of the distribution on the right hand side of (2.6). As the limiting distributions depends, in a rather complicated way, on the unknown eigenvalues $(\lambda_\ell \colon \ell \in \mathbb{N})$ and standard Brownian bridges, Monte Carlo simulation can be used to approximate this distribution using estimated eigenvalues. Implementation details are provided in Section 3 below. Theorem 2.1 was also obtained in Sharipov et al. (2016) under a strong mixing condition that is analogous to Assumption 2.1. These authors further developed a block bootstrap methodology to approximate the limiting distribution.

A common assumption made in order for analogous break point detection procedures based on fPCA to be consistent, as studied for example in Berkes et al. (2009) and Aston and Kirch (2012a), is that δ is not orthogonal to the principal component basis used to perform the dimension reduction step. When using the detector T_n no such assumption is needed.

Theorem 2.2. *Under Model 2.1, Assumption 2.1 and H_A, $T_n \xrightarrow{P} \infty$, as $n \to \infty$.*

The proofs of Theorems 2.1 and 2.2 are in the online supplement.

2.2 Asymptotic properties of the break date estimator

Further advantages of the fully functional approach become apparent when studying the asymptotic properties of the break date estimator $\hat{k}_n^*$, which are established below. Two cases are studied: the fixed break situation for which the break size is independent of the sample size, and the shrinking break situation for which the break size converges to zero at a specified rate. In the fixed break case, the following holds.

Theorem 2.3. *If model* (2.1) *holds with $0 \neq \delta \in L^2[0,1]$, and if Assumption 2.1 is satisfied, then*

$$\hat{k}_n^* - k^* \xrightarrow{\mathcal{D}} \min\left\{k \colon P(k) = \sup_{k' \in \mathbb{Z}} P(k')\right\} \qquad (n \to \infty), \tag{2.7}$$

where

$$P(k) = \begin{cases} (1-\theta)\|\delta\|^2 k + \langle \delta, S_{\varepsilon,k} \rangle, & k < 0, \\ -\theta\|\delta\|^2 k + \langle \delta, S_{\varepsilon,k} \rangle, & k \geq 0, \end{cases} \tag{2.8}$$

with

$$S_{\varepsilon,k} = \sum_{i=1}^{k} \varepsilon_i + \sum_{i=-k}^{-1} \varepsilon_i.$$

As one can see in (2.7), the limit distribution of $\hat{k}_n^*$ in the case of a fixed break size depends on the unknown underlying distribution of the error process. This encourages the consideration of a break δ_n that shrinks as a function of the sample size, in which case the limit distribution is the supremum of a two-sided Brownian motion with triangular drift depending on a small set of nuisance parameters, but not otherwise on the distribution of the error sequence $(\varepsilon_i \colon i \in \mathbb{Z})$.

Theorem 2.4. *If model* (2.1) *holds with* $0 \neq \delta = \delta_n \in L^2[0,1]$ *such that* $\|\delta_n\| \to 0$ *but* $n\|\delta_n\|^2 \to \infty$ *and if Assumption 2.1 is satisfied, then*

$$\|\delta_n\|^2\big(\hat{k}_n^* - k^*\big) \xrightarrow{\mathcal{D}} \inf\Big\{x\colon Q(x) = \sup_{x' \in \mathbb{R}} Q(x')\Big\} \qquad (n \to \infty),$$

where $\mathbb{R}$ *denotes the real numbers and*

$$Q(x) = \begin{cases} (1-\theta)x + \sigma W(x), & x < 0, \\ -\theta x + \sigma W(x), & x \geq 0, \end{cases} \tag{2.9}$$

with $(W(x)\colon x \in \mathbb{R})$ *a two-sided Brownian motion, and*

$$\sigma^2 = \lim_{n\to\infty} \iint C_\varepsilon(t,t') \frac{\delta_n(t)\delta_n(t')}{\|\delta_n\|^2} dt dt',$$

where $C_\varepsilon(t,t')$ *is the long-run covariance kernel of* $(\varepsilon_i\colon i \in \mathbb{Z})$ *given in* (2.4)*.*

An interesting consequence of Theorems 2.3 and 2.4 is that mean changes δ that are orthogonal to the primary modes of variation in the data are asymptotically *easier* to detect and estimate. For example, if, under the conditions of Theorem 2.3, δ is orthogonal to the error functions, then the stochastic term in the limit distribution vanishes. Moreover, if the functions δ_n in Theorem 2.4 tend to align with eigenfunctions corresponding to smaller and smaller eigenvalues of the integral operator with kernel C_ε, then σ^2 tends to zero in the definition of $Q(x)$. The proofs of Theorems 2.3 and 2.4 are given in the online supplement.

Theorem 2.4 suggests a confidence interval for k^*.

Corollary 2.1. *Let* $\Xi = \inf\{x\colon Q(x) = \sup_{x' \in \mathbb{R}} Q(x')\}$*. Then, under the conditions of Theorem 2.4 and for* $\alpha \in (0,1)$*, the random interval*

$$\left(\hat{k}_n^* - \frac{\Xi_{1-\alpha/2}}{\|\delta_n\|^2}, \hat{k}_n^* - \frac{\Xi_{\alpha/2}}{\|\delta_n\|^2}\right) \tag{2.10}$$

is an asymptotic $1-\alpha$ *sized confidence interval for* k^**, where* Ξ_q *is the* q*th quantile of* Ξ*.*

The main crux here is that δ_n is unknown and the distribution of Ξ depends on the unknown break fraction θ and the limiting variance parameter σ^2. Consistent estimation techniques for these parameters are discussed in Section 3 below. This confidence interval tends to be conservative in practice due to the fact that it is derived under the assumption of a shrinking break. A thorough empirical study of the break date estimator and the corresponding confidence interval is provided in Section 4.

The last result of this section concerns the large-sample behavior of $\hat{k}_n^*$ if no break is present in the data, that is, if $\delta = 0$ in (2.1).

Theorem 2.5. *If model* (2.1) *holds with* $\delta = 0$*, so that* $X_i = \mu_i + \varepsilon_i$ *for all* $i = 1, \ldots, n$*, and if Assumption 2.1 is satisfied, then*

$$\frac{\hat{k}_n^*}{n} \xrightarrow{\mathcal{D}} \arg\max_{0 \leq x \leq 1} \|\Gamma^0(x,\cdot)\| \qquad (n \to \infty),$$

where Γ^0 *is a bivariate Gaussian process with mean zero and covariance function* $\mathbb{E}[\Gamma^0(x,t)\Gamma^0(x',t')] = (\min\{x,x'\} - xx')C_\varepsilon(t,t')$.

The proof of Theorem 2.5 is provided in the online supplement. Observe that the limiting distribution in Theorem 2.5 is non-pivotal, but it can be approximated via Monte Carlo simulations using an estimator of C_ε. To see this note that, because of the Karhunen–Loéve representation, $\Gamma^0(x,t)$ can be written in the form $\sum_{\ell=1}^\infty \sqrt{\lambda_\ell}\phi_\ell(t)B_\ell(x)$, where $(\lambda_\ell\colon \ell \in \mathbb{N})$ and $(\phi_\ell\colon \ell \in \mathbb{N})$ are the eigenvalues and eigenfunctions of C_ε and $(B_\ell\colon \ell \in \mathbb{N})$ are independent standard Brownian bridges. Computing the norm as required for the limit in Theorem 2.5 yields that

$$\arg\max_{x\in[0,1]} \|\Gamma^0(x,\cdot)\| \stackrel{\mathcal{D}}{=} \arg\max_{x\in[0,1]} \left(\sum_{\ell=1}^\infty \lambda_\ell B_\ell^2(x)\right)^{1/2}.$$

Truncation of the sum under the square-root on the right-hand side gives then approximations to the theoretical limit. For practical purposes population eigenvalues have to be estimated from the data. This can be done following the steps described in Section 3.2.

2.3 Two fPCA based approaches

In the remainder of this section, the fully functional results put forward here are compared to their fPCA counterparts in Berkes et al. (2009), Aue et al. (2009), Aston and Kirch (2012a, b), and Torgovitski (2016). Berkes et al. (2009) and Torgovitski (2016) dealt with detection procedures and Aue et al. (2009) with break dating procedures, while Aston and Kirch (2012a, b) presented both. A short summary of the different approaches follows.

The works of Berkes et al. (2009), Aue et al. (2009) and Aston and Kirch (2012a, b) utilized the eigenvalues, say $\hat\tau_1,\dots,\hat\tau_n$, and eigenfunctions, say $\hat\psi_1,\dots,\hat\psi_n$, of the sample covariance operator $\hat K$ of the observations whose kernel is given by $\hat K(t,t') = n^{-1}\sum_{i=1}^n [X_i(t) - \bar X_n(t)][X_i(t') - \bar X_n(t')]$. In the presence of a mean break as in (2.1), $\hat K(t,t')$ converges as the sample size tends to infinity to the covariance kernel

$$K(t,t') = K_0(t,t') + \theta(1-\theta)\delta(t)\delta(t'),$$

where $K_0(t,t') = \mathbb{E}[\varepsilon_1(t)\varepsilon_1(t')]$ is the covariance kernel of the innovations $(\varepsilon_i\colon i \in \mathbb{Z})$. In particular, the eigenvalues and eigenfunctions of $\hat K(t,t')$ converge to those of $K(t,t')$ under appropriate assumptions that include the finiteness of the fourth moment $\mathbb{E}[\|\varepsilon_1\|^4]$. Choosing a suitable dimension $d \in \{1,\dots,n\}$ allows one to define an fPCA detector based on the maximally selected quadratic form statistic

$$\tilde R_n = \max_{1\le k\le n} \tilde R_{n,k} = \max_{1\le k\le n} \frac{1}{n}\tilde S_{n,k}^T \hat\Sigma_n^{-1}\tilde S_{n,k}, \tag{2.11}$$

and the break point estimator

$$\tilde k_n^* = \min\Big\{k\colon \tilde R_{n,k} = \max_{1\le k'\le n} R_{n,k'}\Big\}, \tag{2.12}$$

where $\tilde{S}_{n,k} = \sum_{i=1}^{k} \hat{\xi}_i - kn^{-1} \sum_{i=1}^{n} \hat{\xi}_i$ and $\hat{\xi}_i = (\hat{\xi}_{i,1}, \ldots, \hat{\xi}_{i,d})^T$ with fPCA scores $\hat{\xi}_{i,\ell} = \langle X_i - \bar{X}_n, \hat{\psi}_\ell \rangle$, and $\hat{\Sigma}_n = \mathrm{diag}(\hat{\tau}_1, \ldots, \hat{\tau}_d)$. For the independent case, the counterparts of Theorems 2.1 and 2.2 were established in Berkes et al. (2009) for a Cramér–von Mises test statistics, and those of Theorems 2.3 and 2.4 in Aue et al. (2009). Aston and Kirch (2012a) considered versions of the test in (2.11) and showed the consistency of $\tilde{k}_n^*$ in the time series case. The performance of R_n and $\tilde{k}_n^*$ depends crucially on the selection of d and the complexity of the break function δ. To briefly illustrate this point, suppose that $K_0(t,t') = \alpha b(t) b(t')$ for some orthonormal function b that is orthogonal to δ. It then follows from elementary calculations that $\hat{\psi}_1$ will be asymptotically orthogonal to δ if and only if $\alpha > \theta(1-\theta)\|\delta\|^2$, and hence under this latter condition one cannot have a consistent fPCA based test or break date estimator if $d = 1$. The use of the fully functional approach to dating break points is therefore especially advantageous in the interesting case of breaks that are sizable but not obvious in the sense that their influence does not show up in the directions of the leading principal components of the data.

This fact was noticed by Torgovitski (2016), who extended the detection procedures in two ways. First, instead of using the spectral decomposition of the covariance operator K, his procedures are based on the long-run covariance operator C_ε and its eigenvalues $\lambda_1, \ldots, \lambda_n$ and eigenfunctions $\phi_1, \ldots, \phi_n$. Second, an alignment is introduced that shifts the detection procedure into the subspace of the potential break, the idea being to significantly improve power, while not majorly compromising the level. The alignment is obtained by modifying the first sample eigenfunction $\hat{\phi}_1$ using

$$\tilde{\phi}_1' = \frac{\hat{\phi}_1}{n^\gamma} + \frac{\hat{s}\tilde{S}_{n,\tilde{k}_n^*}}{\sqrt{n}}, \tag{2.13}$$

where $\gamma \in (0, 1/2)$ is a tuning parameter and $\hat{s} = \mathrm{sign}\langle \hat{\phi}_1, \tilde{S}_{n,\tilde{k}_n^*} \rangle$. Torgovitski (2016) then proposed to replace $\hat{\phi}_1$ with $\hat{\phi}_1' = \tilde{\phi}_1'/\|\tilde{\phi}_1'\|$ in the definition of (2.11), but did not introduce the corresponding break dating procedure.

3 Implementation details

3.1 Estimation of long-run covariance operator

The implementation of the detection procedure and confidence intervals based on the break point estimator requires the estimation of the covariance operator C_ε. Due to its definition as a bi-infinite sum of the lagged autocovariances of the functional time series $(\varepsilon_i \colon i \in \mathbb{Z})$, the following lag-window estimator is used. Let

$$\hat{C}_\varepsilon(t,t') = \sum_{\ell=-\infty}^{\infty} w_\tau\left(\frac{\ell}{h}\right) \hat{\gamma}_\ell(t,t'), \tag{3.1}$$

where the components of this estimator are defined as follows: h is a bandwidth parameter satisfying $h = h(n)$, and $1/h(n) + h(n)/n^{1/2} \to 0$ as $n \to \infty$,

$$\hat{\gamma}_\ell(t,t') = \frac{1}{n} \sum_{i \in \mathcal{I}_\ell} \left[X_i(t) - \bar{X}_i^*(t)\right] \left[X_{i+\ell}(t') - \bar{X}_{i+\ell}^*(t')\right],$$

with $\mathcal{I}_\ell = \{1, \ldots, n-\ell\}$ if $\ell \geq 0$ and $\mathcal{I}_\ell = \{1-\ell, \ldots, n\}$ if $\ell < 0$.

$$\bar{X}_j^*(t) = \begin{cases} \dfrac{1}{\hat{k}_n^*} \displaystyle\sum_{i=1}^{\hat{k}_n^*} X_i(t), & 1 \leq j \leq \hat{k}_n^*, \\ \dfrac{1}{n - \hat{k}_n^*} \displaystyle\sum_{i=\hat{k}_n^*+1}^{n} X_i(t), & \hat{k}_n^* + 1 \leq j \leq n, \end{cases}$$

and w_τ is a symmetric weight function with bounded support of order τ satisfying the standard conditions $w_\tau(0) = 1$, $w_\tau(u) = w_\tau(-u)$, $w_\tau(u) \leq 1$, $w_\tau(u) = 0$ if $|u| > m$ for some $m > 0$, w_τ is continuous, and

$$0 < \mathfrak{q} = \lim_{x \to 0} x^{-\tau}[1 - w_\tau(x)] < \infty. \tag{3.2}$$

Through $\hat{C}_\varepsilon$, eigenvalue estimates $\hat{\lambda}_1, \ldots, \hat{\lambda}_n$ of $\lambda_1, \ldots, \lambda_n$ are defined via the integral operator

$$\hat{\lambda}_\ell \hat{\phi}_\ell(t) = \int \hat{C}_\varepsilon(t, s) \hat{\phi}_\ell(s) ds. \tag{3.3}$$

In order to show consistency of these estimates, a condition supplementary to the weak dependence of the errors $(\varepsilon_i \colon i \in \mathbb{Z})$ given in Assumption 2.1 is needed.

Assumption 3.1. *For some $p > 2$, $\ell(\mathbb{E}[\|\varepsilon_i - \varepsilon_{i,\ell}\|^p])^{1/p} \to 0$ as $\ell \to \infty$.*

Assumption 3.1 is not necessarily stronger than Assumption 2.1, although both are implied by the simple condition that $(\mathbb{E}[\|\varepsilon_i - \varepsilon_{i,\ell}\|^p])^{1/p} = O(\ell^{-\rho})$ for some $\rho > 1$, which is by itself a fairly mild assumption. This condition appears in Horváth et al. (2013). The following result holds.

Proposition 3.1. *Under the conditions of Theorem 2.4 and Assumption 3.1, $\hat{C}_\varepsilon$ in* (3.1) *is a consistent estimator of C_ε in $L^2[0,1]^2$. Moreover, for any fixed $d \in \mathbb{N}$, $\max_{1 \leq \ell \leq d} |\lambda_\ell - \hat{\lambda}_\ell| = o_P(1)$.*

The verification of this result is given in Lemma A.5 of the online supplement. While the proposition guarantees the large-sample accuracy under a reasonably broad set of conditions, producing the estimate $\hat{C}_\varepsilon$ and its eigenvalues satisfying (3.3) in practice requires the choice of a weight function w_τ and bandwidth h. This problem, which is familiar to nonparametric analysis of finite-dimensional time series and spectral density estimation (see, for example, Chapter 7 of Brillinger, 2001), has only recently begun to receive attention in the setting of functional time series.

In the case of long-run covariance function estimation and functional spectral density estimation, Hörmann and Kokoszka (2010) and Panaretos and Tavakoli (2012) utilized Bartlett and Epanechnikov weight functions (see Bartlett 1946; and Wand and Jones, 1995) with bandwidths of the form $h = n^{1/3}$ and $h = n^{1/5}$, respectively. These choices arise from the well-known fact that taking a bandwidth of the form $h = n^{1/(1+2\tau)}$ maximizes the rate at which the mean-squared normed error of the estimator $\hat{C}_\varepsilon$ tends to zero. The performance of the estimator in finite samples can, however, be affected by strong serial correlation in the data, in

which case one should use a larger bandwidth in order to reduce the bias of $\hat{C}_\varepsilon$. An approach that balances these two concerns is to take $h = Mn^{1/(1+2\tau)}$, where the constant M is estimated from the data and increases with the level of serial correlation. It can be shown (see Rice and Shang, 2017) that the optimal constant M in terms of asymptotically minimizing the mean squared normed error of $\hat{C}_\varepsilon$ is of the form

$$M = \left(2\tau \|\mathfrak{q} C_\varepsilon^{(\tau)}\|^2\right)^{1/(1+2\tau)} \left(\left\{ \|C_\varepsilon\|^2 + \left(\int_0^1 C_\varepsilon(u,u)du \right)^2 \right\} \int_{-\infty}^{\infty} w_\tau^2(x)dx \right)^{-1/(1+2\tau)},$$

where $C_\varepsilon^{(\tau)}$ is related to the τth derivative of a spectral density operator evaluated at frequency zero. The unknown quantities in M can be estimated using pilot estimates of C_ε and $C_\varepsilon^{(\tau)}$ to produce an estimated bandwidth $h = \hat{M}n^{1/(1+2\tau)}$. Complete details of this estimation procedure are provided in Section C of the online supplement to the paper.

A comparison of the accuracy in terms of mean-squared normed error of $\hat{C}_\varepsilon$ for a multitude of bandwidth and weight function combinations is provided in Rice and Shang (2017), but a comparative study of how these estimators perform in problems of inference has not been conducted, to the best of our knowledge. With results reported in the online supplement, the proposed break point detection method was compared for all combinations of the Bartlett, Parzen (Parzen, 1957) and a version of the flat-top (Politis and Romano, 1996) weight functions with the four bandwidth choices of $h = n^{1/3}$, $h = n^{1/4}$, $h = n^{1/5}$, and $h = \hat{M}n^{1/(1+2\tau)}$ for the data generating processes considered in the simulation study presented below, as well as some additional processes exhibiting stronger temporal dependence. It was found that when it comes to conducting hypothesis tests and producing confidence intervals as described above with moderately correlated errors, each of these typical choices produced similar results. The difference across weight functions was minuscule, whereas there were some small fluctuations in the empirical sizes of the test of H_0 due to the choice of the bandwidth: no more than a 2% difference when the level was set at 5% over those FAR processes utilized in Section 4, but with expected bigger discrepancies and advantages for the empirical bandwidth when the level of dependence approached non-stationarity. Due to the similarity in performance of each choice for the data generating processes considered below, results are only presented for the bandwidth $h = n^{1/4}$ and the Bartlett weight function below.

3.2 Computation of critical values

To compute the critical values of the limiting distribution of T_n given in Theorem 2.1, say T, the following procedure was employed. Based on the estimator $\hat{C}_\varepsilon$, the first D empirical eigenvalues satisfying (3.3) were computed, where D is taken to be the number of basis elements over which the initial discretely observed functional data are smoothed. By then simulating D independent Brownian bridges on $[0,1]$, $B_\ell(x)$, using the R package `sde`, a realization of T is estimated by

$$\hat{T} = \sup_{0 \le x \le 1} \sum_{\ell=1}^{D} \hat{\lambda}_\ell B_\ell^2(x).$$

This estimation is independently repeated R times, and quantiles of the resulting Monte Carlo distribution are used to produce the appropriate cut-offs. For the results in Sections 4 and 5, R was selected to be $1{,}000$.

3.3 Construction of confidence intervals

This section provides more information on the construction of confidence intervals as defined through Corollary 2.1. To start, let $\hat{\Xi} = \inf\{x\colon \hat{Q}(x) = \sup_{x' \in \mathbb{R}} \hat{Q}(x')\}$ be the sample version of Ξ, where $\hat{Q}$ is an estimated version of Q in (2.9) obtained by plugging in the natural estimators

$$\hat{\theta} = \frac{\hat{k}_n^*}{n} \qquad \text{and} \qquad \hat{\sigma}^2 = \iint \hat{C}_\varepsilon(t,t') \frac{\hat{\delta}_n(t)\hat{\delta}_n(t')}{\|\hat{\delta}_n\|^2} dt dt'$$

in place of their respective population counterparts θ and σ as specified in Theorem 2.4, where

$$\hat{\delta}_n = \frac{1}{n - \hat{k}_n^*} \sum_{i=\hat{k}_n^*+1}^{n} X_i - \frac{1}{\hat{k}_n^*} \sum_{i=1}^{\hat{k}_n^*} X_i,$$

and $\hat{C}_\varepsilon$ an estimator of C_ε as discussed in Section 3.1. All of these estimators are consistent under the conditions of Theorem 2.4; see Lemma A.5 of the online supplement. Let $\hat{\Xi}_q$ denote the qth quantile of the distribution of $\hat{\Xi}$.

Theorem 3.1. *Under the conditions of Theorem 2.4 and Assumption 3.1, for $\alpha \in (0,1)$, the random interval*

$$\left(\hat{k}_n^* - \frac{\hat{\Xi}_{1-\alpha/2}}{\|\hat{\delta}_n\|^2}, \hat{k}_n^* - \frac{\hat{\Xi}_{\alpha/2}}{\|\hat{\delta}_n\|^2} \right)$$

is an asymptotic $1-\alpha$ confidence interval for k^*.

Note that the construction of confidence intervals is aided by the use of the exact form of the maximizers in the limit of Theorem 2.4 as derived in Bhattacharya and Brockwell (1976) and Stryhn (1996), see the supplemental material for more. Since σ^2 and $\hat{\sigma}^2$ are respectively bounded from above by λ_1 and $\hat{\lambda}_1$, the largest eigenvalues of the integral operators with kernels C_ε and $\hat{C}_\varepsilon$, a conservative confidence interval is obtained by replacing $\hat{\sigma}^2$ with $\hat{\lambda}_1$.

4 Simulation Study

4.1 Setting

Following the construction of the data generating processes (DGP's) in Aue et al. (2015), n functional data objects were generated using $D = 21$ Fourier basis functions $v_1, \ldots, v_D$ on the unit interval $[0,1]$. The choice of D corresponds to our study of yearly Australian temperature curves constructed from daily minimum temperature observations that were initially smoothed over this basis. Qualitatively these results remain valid

for larger values of D. Without loss of generality, the initial mean curve μ in 2.1 is assumed to be the zero function. Independent curves were then generated according to

$$\zeta_i = \sum_{\ell=1}^{D} N_{i,\ell} v_\ell,$$

where the $N_{i,\ell}$ are independent normal random variables with standard deviations $\boldsymbol{\sigma} = (\sigma_\ell \colon \ell = 1, \ldots, D)$ used to mimic various decays for the eigenvalues of the covariance and long-run covariance operators. Three distinct situations were considered:

- *Setting 1:* the errors are finite dimensional, using $\sigma_\ell = 1$ for $\ell = 1, 2, 3$ and $\sigma_\ell = 0$ for $\ell = 3, \ldots, D$;
- *Setting 2:* mimics a fast decay of eigenvalues, using $\boldsymbol{\sigma} = (3^{-\ell} \colon \ell = 1, \ldots, D)$;
- *Setting 3:* mimics a slow decay of eigenvalues, using $\boldsymbol{\sigma} = (\ell^{-1} \colon \ell = 1, \ldots, D)$.

Note that the last setting, inspired by the data analysis reported in Section 5, is not to be taken asymptotically in D. Rather it is meant to model a slow decay of the finitely many initial eigenvalues without intending to prescribe the behavior for D tending to infinity.

As innovations, independent curves $\varepsilon_i = \zeta_i$, $i = 1, \ldots, n$, were used. To explore the effect of temporal dependence on the break point estimators, functional autoregressive curves were also considered, which are widely used to model serial correlation of functional data, see Besse et al. (2000) and Antoniadis and Sapatinas (2003). First-order functional autoregressions $\varepsilon_i = \Psi \varepsilon_{i-1} + \zeta_i$, $i = 1, \ldots, n$, were generated (using a burn-in period of 100 initial curves that were discarded). The operator was set up as $\Psi = \kappa \Psi_0$, where the random operator Ψ_0 is represented by a $D \times D$ matrix whose entries consist of independent, centered normal random variables with standard deviations given by $\boldsymbol{\sigma}\boldsymbol{\sigma}'$ as specified by Settings 1–3. A scaling was applied to achieve $\|\Psi_0\| = 1$. The constant κ can then be used to adjust the strength of the temporal dependence. To ensure stationarity of the time series, $|\kappa| = 0.5$ was selected.

To highlight the effect of the distribution of the break function across eigendirections as well as its size relative to the noise level, the following arrangements were made. A class of break functions was studied given by

$$\delta_m = \delta_{m,c} = \sqrt{c}\delta_m^*, \qquad \delta_m^* = \frac{1}{\sqrt{m}} \sum_{\ell=1}^{m} v_\ell, \qquad m = 1, \ldots, D, \tag{4.1}$$

where the normalization ensures that all δ_m^* have unit norm. The role of c is explained below. Note that δ_1 represents the case of a break only in the leading eigendirection of the errors. On the other end of the spectrum is δ_D describing the case of a break that affects all eigendirections uniformly. To relate break size to the natural fluctuations in the innovations, the signal-to-noise ratio

$$\mathrm{SNR} = \frac{\theta(1-\theta)\|\delta_m\|^2}{\mathrm{tr}(C_\varepsilon)} = c\frac{\theta(1-\theta)}{\mathrm{tr}(C_\varepsilon)},$$

was used, where θ denotes the relative location of the break date and C_ε the long-run covariance operator of the ε_i. (Note that since $\|\delta_m^*\| = 1$, in the adopted formulation SNR does not depend on m.) Results are reported choosing c to maintain a prescribed SNR.

Finally, in order to mitigate the effect of the particular shape of the Fourier basis functions and the ordering $v_1, \ldots, v_D$ on the performance of the various procedures, a random permutation π was applied to $1, \ldots, D$ before each simulation run, and the experiment was performed as described above using the permuted ordering $v_{\pi(1)}, \ldots, v_{\pi(D)}$. Combining the previous paragraphs, functional curves $y_i = \delta_m \mathbb{1}\{i > k^*\} + \varepsilon_i$, $i = 1, \ldots, n$, according to (2.1) were generated for $k^* = \lfloor \theta n \rfloor$ with $\theta = 0$ (null hypothesis), and $\theta = 0.25$ and 0.5 (alternative). Both the fully functional procedure and its fPCA counterparts were applied to a variety of settings, with outcomes reported in subsequent sections. All results are based on 1000 runs of the simulation experiments. Additional complementary simulation evidence is presented in the online supplement.

4.2 Level and power of the detection procedures

In this section, the level and power of the proposed detection procedure are compared to the two fPCA-based methods introduced in Section 2.3. In particular, the fPCA-based detector (2.11) was run with three levels of total variation explained (TVE), namely 85%, 90% and 95%. The change-aligned detection procedure (2.13) was set up as in Torgovitski (2016). Critical values for the proposed fully functional detection procedure were obtained through simulation from the limit distribution under the null hypothesis as provided in Theorem 2.1.

Table 4.1 provides the levels for the various detection procedures for the three settings of eigenvalue decays, and iid and FAR(1) data generating processes. For the FAR(1) case, the long-run covariance operator was estimated following the recommendations given in Rice and Shang (2017). The sample sizes under consideration were $n = 50$ and $n = 100$. It can be seen that, even for these rather small to moderate sample sizes, the proposed method kept levels reasonably well across all specifications. This is true to a lesser extent also for the fPCA-based procedures, while the change-aligned version produced the most variable results. The fPCA-based procedures depend, by construction, more explicitly on the behavior of the eigenvalues with levels well adjusted in case of a fast decay. The proposed procedure is fairly robust in all settings.

To examine the power of the detection procedures in finite samples, the break functions δ_m in (4.1) were inserted as described in Section 4.1 with scalings c so that the SNR varied between 0, 0.1, 0.2, 0.3, 0.5, 1 and 1.5. The empirical rejection rates out of 1000 simulations for each test statistic described above are reported as power curves in Figure 4.1 when the errors in (2.1) are iid and distributed according to each of Settings 1, 2, and 3. The sample size in the figure is $n = 50$ and the number of components m in the break functions δ_m are 1, 5, and 20. Note that the plots in Figure 4.1 are not size corrected because it would not qualitatively change the outcomes. Further simulation evidence is provided in the online supplement. The findings of these simulations can be summarized as follows:

- The change aligned test of Torgovitski (2016) was usually outperformed by both the fPCA and fully functional methods for most of the DGP's and sample sizes under consideration.

Setting	DGP	n	Proposed	TVE 85%	TVE 90%	TVE 95%	Aligned
1	iid	50	0.08	0.02	0.03	0.03	0.02
		100	0.06	0.07	0.06	0.06	0.05
	FAR(1)	50	0.07	0.03	0.03	0.02	0.00
		100	0.05	0.07	0.06	0.07	0.02
2	iid	50	0.07	0.06	0.06	0.06	0.09
		100	0.06	0.05	0.06	0.07	0.05
	FAR(1)	50	0.07	0.04	0.04	0.05	0.11
		100	0.06	0.05	0.05	0.05	0.05
3	iid	50	0.04	0.02	0.02	0.00	0.00
		100	0.05	0.07	0.03	0.02	0.01
	FAR(1)	50	0.03	0.01	0.03	0.00	0.00
		100	0.05	0.02	0.02	0.02	0.00

Table 4.1: Empirical sizes for the various detection procedures for two data generation processes. The nominal level was $\alpha = 0.05$.

- The power for the fully functional detection procedure was observed to improve as m increased, as predicted by the theory. Moreover, when the change was largely orthogonal to the errors, as in Setting 1 with $m = 20$, the expected advantage of the fully functional method over the dimension reduction based approaches materialized.
- A particularly interesting example to examine is when $m = 1$ under Setting 3 (with slowly decaying eigenvalues). One notices in this case that, although the change lied fully in the direction of the leading principal component of the errors, the fPCA-based methods were outperformed by the fully functional method, and additionally their performance decayed as TVE increased. Here, the slow decay of eigenvalues adversely affects the fPCA procedure. This contrasts, for example, with the case when $m = 20$ under Setting 2 (with fast decaying eigenvalues), when the fPCA method improved as TVE increased, and ultimately outperformed the fully functional method. This demonstrates that the fPCA method is not guaranteed to beat the proposed detection procedure even when the break is in the leading eigendirection. Moreover, this particular case highlights the fact that increasing TVE may not always lead to improved performance. Note also that this example seems to match well with the situation encountered in an application to Australian temperature curves presented in Section 5.
- In additional simulations reported in the online supplement, the expected improvement in power when n increased was noticed. Additionally, no more power loss than is typical was observed when the model errors are serially correlated rather than independent and identically distributed.

4.3 Performance of the break dating procedures

In order to study the empirical properties of the break date estimator $\hat{k}_n^*$, the break functions δ_m specified in (4.1) of Section 4.1 were utilized again with scaling c chosen to yield SNR values of 0.5 and 1. The break

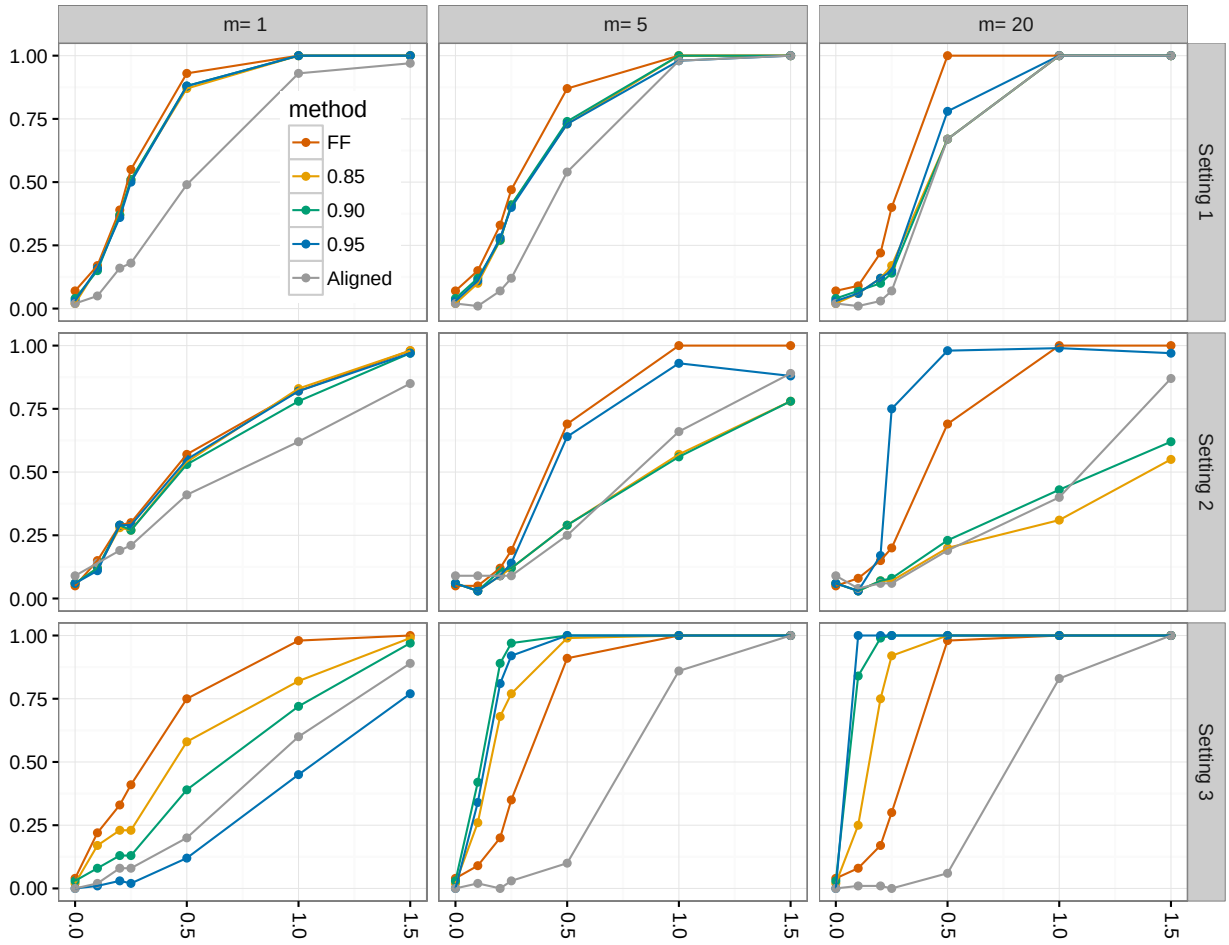

Figure 4.1: Power curves for the various break detection procedures for three different forms of the break functions indexed my m and the three eigenvalues settings for $n = 50$ and independent errors. The x-axis gives different choices of SNR. Observe that "FF" refers to the proposed fully functional method, "0.85", "0.90" and "0.95" correspond to the three levels of TVE in the fPCA procedures, and "Aligned" to the method of Torgovitski (2016).

date was inserted at $\theta = 0.25$, so that the samples before and after the break have a ratio of 1 to 3. As in the previous section, focus is on the small sample size $n = 50$. The results from additional settings are reported in the online supplement. For each setting and choice of m, the estimators $\hat{k}_n^*$ and $\tilde{k}_n^*$ for k^* were computed for the proposed and the fPCA methods, respectively in 1000 independent simulation runs. The results are summarized in the form of box plots in Figure 4.2.

Overall, the proposed method is observed to be competitive, with box plots being narrower or of the same width as those constructed from the fPCA counterparts. It can be seen that the accuracy of the fully functional break date procedure improved for increasing m, spreading the break across a larger number of directions. As expected, the performance of the fPCA procedure was sensitive to the choice of TVE, in a way that often only the best selection of TVE was competitive with the fully functional method. Moreover, in analogy to the same

phenomenon observed in the power study, the fully functional procedure enjoys an advantage when the break loads entirely on the first eigenfunction ($m = 1$) for slowly decaying eigenvalues of the covariance operator (Setting 3).

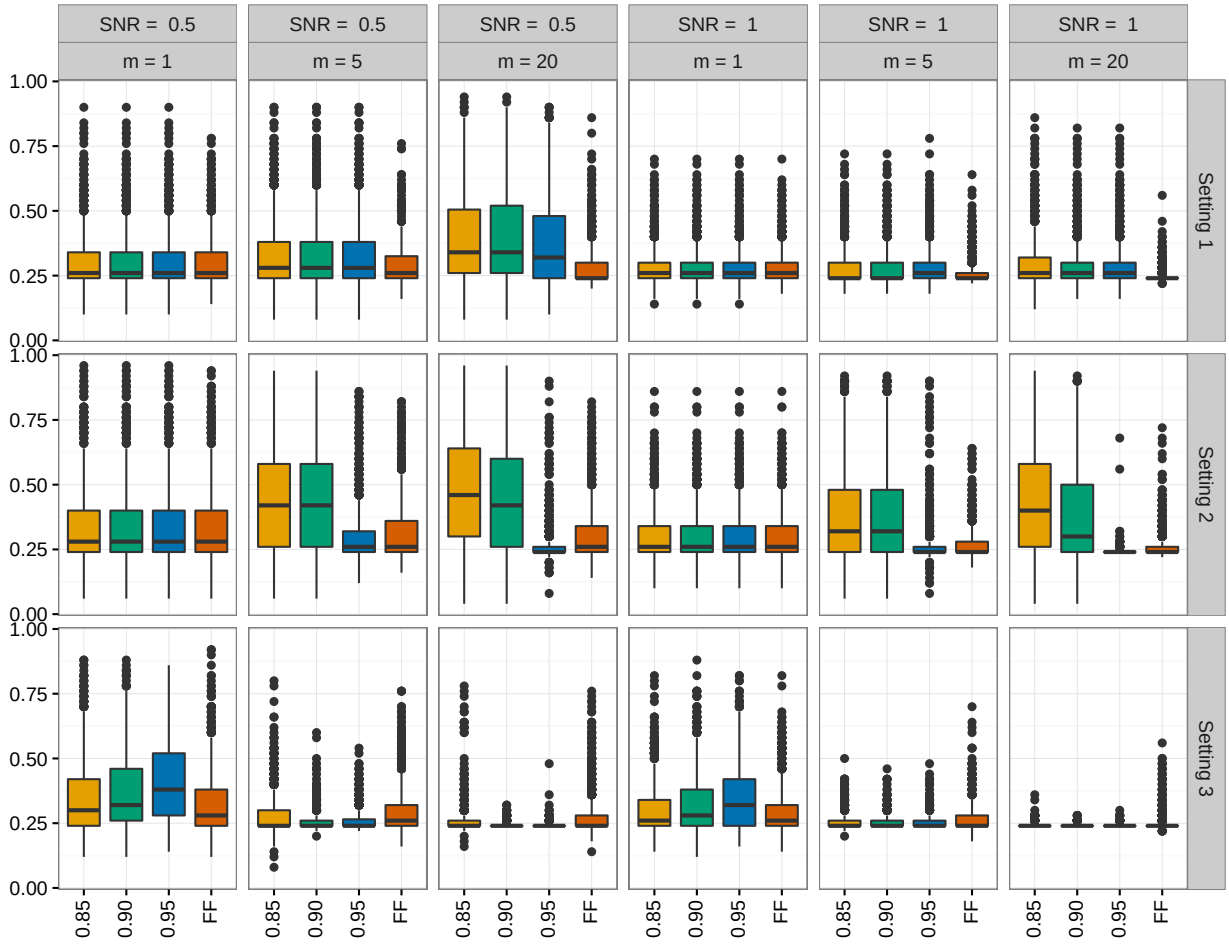

Figure 4.2: Boxplots for the various break dating procedures for three different forms of the break functions indexed my m and two choices of SNR for the three eigenvalues settings, sample size $n = 50$ and independent errors. Labeling of the procedures is as in Figure 4.1.

The confidence intervals computed from Theorem 3.1 are seen to be conservative. As already pointed out after Corollary 2.10, this is due to the fact, that they are based on an asymptotic analysis assuming a shrinking break. For illustration purposes, since this will prove relevant in Section 5, Figure 4.3 gives 95% confidence intervals for the case of Setting 3 with independent errors and sample size $n = 100$. The break function δ_m is inserted in the middle ($\theta = 0.5$), using $m = 1$, 5 and 20 as before. The plots provide further evidence for the theory, as the confidence intervals get significantly narrower when the break function is distributed across a larger number of directions. The case $m = 1$ leads to the widest confidence intervals, which for this case are of little practical relevance. Larger sample sizes and higher SNR lead to the expected improvements, but are not shown here to conserve space. To improve the width of the confidence intervals for small sample sizes

and/or small SNR's, one might entertain some jackknife or bootstrap modifications. This might be pursued in detail elsewhere.

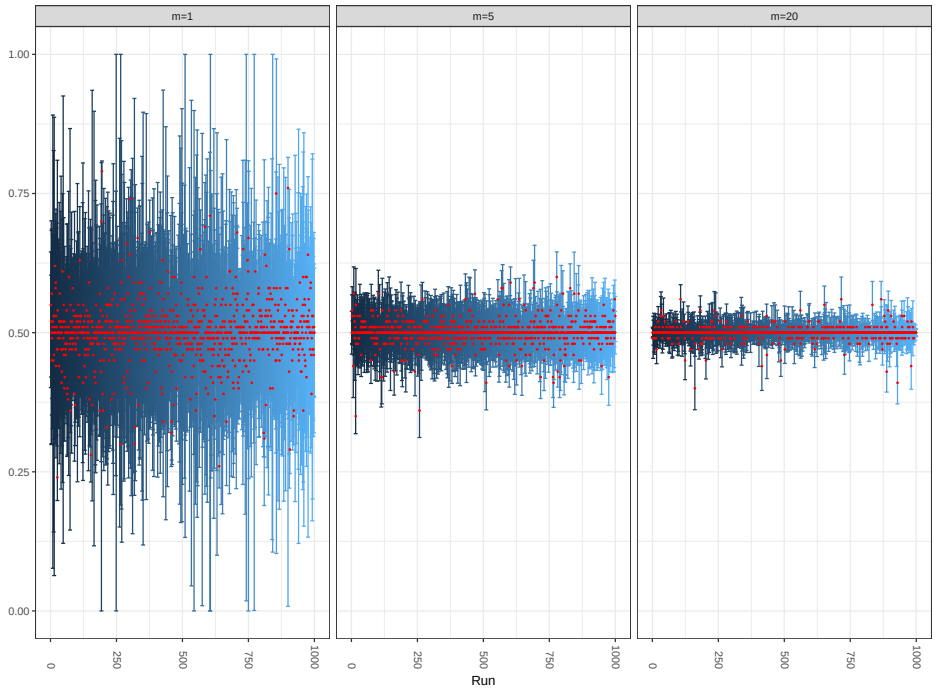

Figure 4.3: Confidence intervals constructed from the fully functional break dating procedure across 1000 simulation runs for Setting 3, sample size $n = 100$, and three types of break functions indexed by m with SNR set to 0.5. For each run, the blue line gives the 95% confidence interval and the red dot the estimated break date.

4.4 Heavy tails

The heavy tail case is only considered for independent curves in Settings 2 with fast decay of eigenvalues of the innovations and break function specified by δ_m in (4.1) with $m = 1$, 5 and 20 as before. Settings 1 and 3 produce results more in favor of the proposed method. Instead of the normal distributions specified in Section 4.1, $\zeta_1, \ldots, \zeta_n$ were chosen to be t-distributed with 2, 3 and 4 degrees of freedom and $\varepsilon_1, \ldots, \varepsilon_{100}$ were defined accordingly. Modifications of the simulation settings presented in this section could potentially be useful for applications to intra-day financial data such as the Microsoft intra-day return data presented as part of the online supplement. Due to the reduced number of finite moments in this setting, the fPCA-based procedure is not theoretically justified, while the fully functional procedure is not justified only for the case of two degrees of freedom.

Results in Figure 4.4 are given for $n = 100$, $k^* = 50$. The summary statistics show the proposed method to be superior in all cases. The proposed method looks in general more favorable in the heavy-tail case than in the time series case of the previous section due to the deteriorated performance in estimating eigenvalues and

eigenfunctions. It can be seen that in all cases the fPCA-based procedure fails to produce reasonable results.

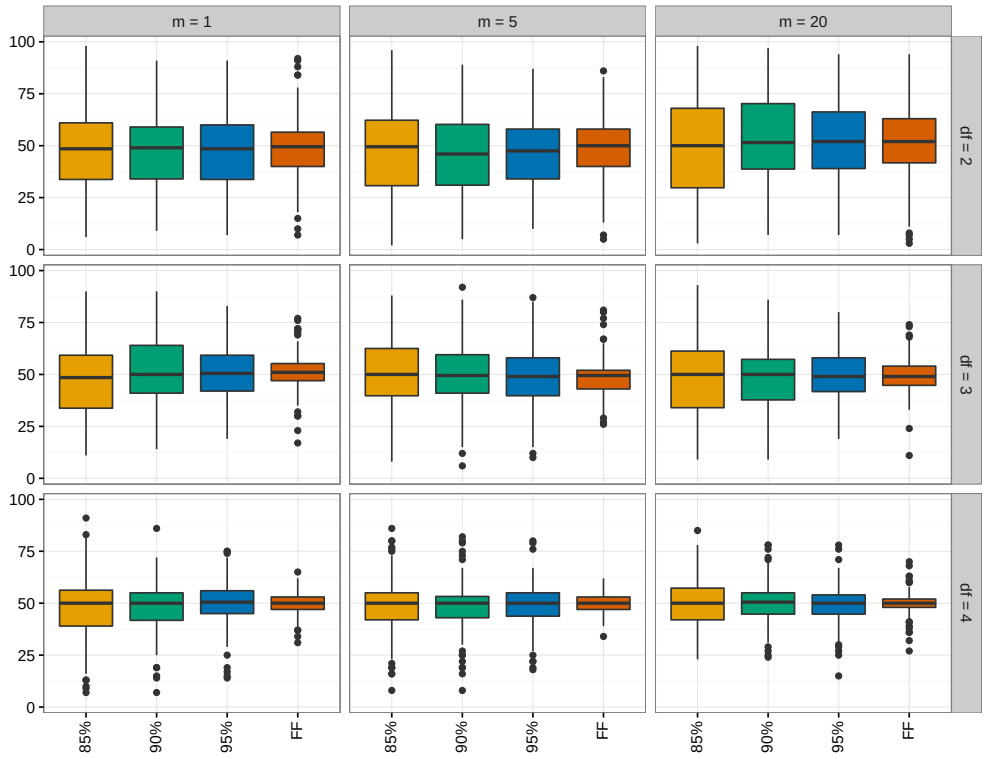

Figure 4.4: Boxplots for the various break dating procedures for three different forms of the break functions indexed my m and t-distributed innovations with 2, 3 and 4 degrees of freedom for Setting 2, sample size $n = 100$, $k^* = 50$ and independent errors. Labeling of the procedures is as in Figure 4.1.

The performance is worst for $\mathrm{df} = 2$ and somewhat comparable for $\mathrm{df} = 3$ and $\mathrm{df} = 4$. The proposed method is seen to work for the latter two cases but its performance deteriorates somewhat for $\mathrm{df} = 2$, a situation that is not theoretically justified.

5 Application to annual temperature curves

In this section, the proposed methodology is applied to annual temperature curves from eight measuring stations in Australia. More precisely, the raw data consists of 365 (366) daily measurements of minimum temperatures that were converted into functional objects using 21 Fourier basis functions. The observations for each of the eight stations are recorded over different time spans, roughly equaling 100 years. The data may be downloaded from The Australian Bureau of Meteorology at the URL `www.bom.gov.au`. For each case, the fully functional break detection procedure rejected the null hypothesis of no structural break in the mean function. Consequently, both functional break dating procedure and fPCA counterpart were applied to locate the time of the mean break. Information on all stations under consideration is provided in Table 5.1. More details may be found in the online supplement.

Station	Range	$\hat{k}_n^*$ (year)	CI (years)	Range of $\tilde{k}_n^*$ (year)
Sydney (Observatory Hill)	1959–2012	1991	(1981, 1994)	1983, **1991**
Melbourne (Regional Office)	1855–2012	1998	(1989, 2000)	1996, **1998**
Boulia Airport	1888–2012	1978	(1954, 1981)	**1978**
Cape Otway Lighthouse	1864–2012	1999	(1949, 2005)	1999, **2000**
Gayndah Post Office	1893–2009	1962	(1952, 1966)	1953, 1962, **1968**
Gunnedah Pool	1876–2011	1985	(1935, 1992)	1979, 1984, **1985**, 1986
Hobart (Ellerslie Road)	1882–2011	1966	(1957, 1969)	**1966**, 1967, 1968, 1969
Robe Comparison	1884–2011	1981	(1954, 1985)	1969, 1974, **1981**

Table 5.1: Summary of results for eight Australian measuring stations. The column labeled $\hat{k}_n^*$ reports the estimated break date using the fully functional method, CI gives the corresponding 95% confidence interval. This is contrasted with the range of break date estimates obtained from using fPCA methods with dimension of the projection space $d = 1, \ldots, 10$. The year in bold is the most frequently chosen break date.

In the following the station Gayndah Post Office is singled out and discussed in more detail. The time series plot of $n = 116$ annual curves recorded in degree Celsius at this station from 1893 to 2009 are given in the upper left panel of Figure 5.1. They exhibit the temperature profile typical for Australia, with higher temperatures in the beginning and end of the year. The corresponding scree plot of sample eigenvalues in the upper right panel of the same figure indicates a slow decay, which Setting 3 in Section 4 sought to mimic. The p-value of the fully functional detection procedure for this station was 0.008. Table 5.1 reports the break date estimate for the fully functional procedure as 1962 and gives a 95% confidence interval spanning the years from 1952 to 1966. In the range considered, the fPCA procedure chose three different years as break dates, namely 1953 (corresponding to $d = 1$ and $\mathrm{TVE} = 0.40$), 1962 (for $d = 3$ and $\mathrm{TVE} = 0.62$), and 1968 (for all other choices of d with TVE reaching 0.92 at $d = 10$). It can therefore be seen that, for any reasonable choice of TVE, the fPCA break date estimate leads to a year that is not included in the 95% confidence interval obtained from the fully functional methodology, even those were shown to be conservative in Section 4. The estimated break function is displayed in the middle panel of Figure 5.1. Almost 90% of the variation in $\|\hat{\delta}\|$ is explained by the first sample eigenfunction, with a rapid decay of contributions from higher sample eigenfunctions. This is displayed in the middle panel of Figure 5.1. The situation is therefore indeed similar to the case displayed in the lower left panels of Figures 4.1 and 4.2, which corresponds to slow decay of eigenfunctions and a break occurring predominantly in the direction of the first mode of variation. That this is a situation beneficial to the proposed procedure is further highlighted in the lower panel of Figure 5.1. Here it can be seen that the estimated SNR of the sample break function decreases significantly with the inclusion of further sample eigenvalues and eigenfunctions into the analysis. In particular, the estimated SNR's are, for $d > 1$, noticeably smaller than the estimated SNR obtained from the fully functional procedure.

The application shows that, while both fully functional and fPCA procedures often work similarly in

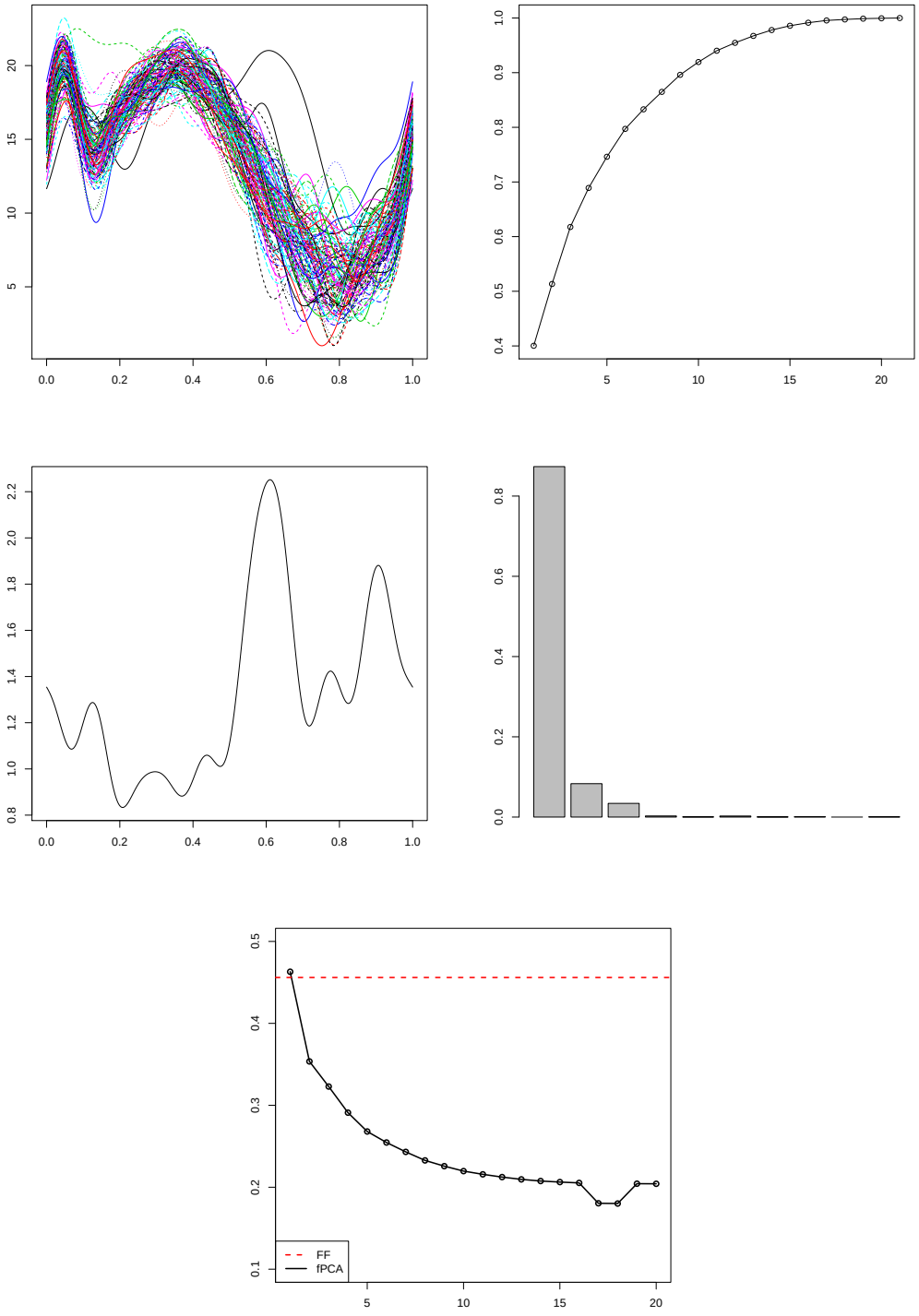

Figure 5.1: Upper panel: Time series plot of annual temperature profiles at Gayndah Post Office (left) and scree plot of eigenvalues from the sample covariance operator of the Gayndah Post Office temperature profiles (right). Middle panel: Estimated break function $\hat{\delta}$ (left) and proportion of variation in $\|\hat{\delta}\|$ explained by the ℓth sample eigenfunction (right). Lower panel: Estimated SNR for the fully functional procedure (straight line) and for the fPCA procedure across varying d.

practice, there are cases when they differ substantially. In the situation discussed in this section, there is evidence to believe that the fully functional method is perhaps more trustworthy. The results of the data application used in combination with the simulation analysis show that one can do worse than the proposed procedure but not obviously better.

6 Conclusions

In this paper, a fully functional methodology was introduced to detect and date mean curve breaks for functional data. The assumptions made allow for time series specifications of the curves and are formulated using the optimal rates for approximations of the data with ℓ-dependent sequences. The assumptions are notably weaker than those usually made in the fPCA context and include heavy-tailed functional observations, making the asymptotic theory developed here widely applicable. In a comprehensive simulation study it is shown that the fully functional method tends to perform better than its fPCA counterpart, with significant performance gains for breaks that do not align well with the directions specified by the largest (few) eigenvalue(s) of the data covariance operator, but also in a number of subtler situations such as breaks concentrated on the first eigendirection with slowly decaying eigenvalues. It is shown in an application to annual temperature curves that the latter situation can be of practical relevance. More generally, this work provides an in-depth study in a specific context of the overarching principle that whenever the signal of interest is not dominant or is "sparse", in the sense that it is not entirely contained in the leading principal components, then alternatives to dimension reduction based methods should be considered and are likely more effective.

References

[1] Antoniadis, A. & T. Sapatinas (2003). Wavelet methods for continuous time prediction using Hilbert-valued autoregressive processes. *Journal of Multivariate Analysis* **87**, 133–158.

[2] Aston, J.A.D. & C. Kirch (2012a). Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis* **109**, 204–220.

[3] Aston, J.A.D. & C. Kirch (2012b). Evaluating stationarity via change-point alternatives with applications to FMRI data. *The Annals of Applied Statistics* **6**, 1906–1948.

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

[5] Aue, A., Gabrys, R., Horváth, L. & P. Kokoszka (2009). Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis* **100**, 2254–2269.

[6] Aue, A., Hörmann, S., Horváth, L. & M. Hušková (2014). Dependent functional linear models with applications to monitoring structural change. *Statistica Sinica* **24**, 1043–1073.

[7] Aue, A. & L. Horváth (2013). Structural breaks in time series. *Journal of Time Series Analysis* **34**, 1–16.

[8] Aue, A., Horváth, L. & D. Pellatt (2017). Functional generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis* **38**, 3–21.

[9] Aue, A., Rice, G. & O. Sönmez (2017+). Online supplement to "Detecting and dating structural breaks in functional data without dimension reduction".

[10] Aue, A. & A. van Delft (2017+). Testing for stationarity of functional time series in the frequency domain. Preprint available at `https://arxiv.org/abs/1701.01741`.

[11] Bartlett, M.S. (1946). The large-sample theory of squential tests. *Proceedings of the Cambridge Philosophical Society* **42**, 239–244.

[12] Berkes, I., Gabrys, R., Horváth, L. & P. Kokoszka (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society, Series B* **71**, 927–946.

[13] Besse, P., Cardot, H. & D. Stephenson (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics* **27**, 673–687.

[14] Bhattacharya, P. & P. Brockwell (1976). The minimum of an additive process with applications to signal estimation and storage theory. *Probability Theory and Related Fields* **37**, 51–75.

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

[16] Bucchia, B. & M. Wendler (2017). Change-point detection and bootstrap for Hilbert space valued random fields. *Journal of Multivariate Analysis* **155**, 344–368.

[17] Ferraty, F. & Vieu, P. (2010). *Nonparametric Functional Data Analysis.* Springer-Verlag, New York.

[18] Fremdt, S. Horváth, L. Kokoszka, P. & J. Steinebach (2014). Functional data analysis with an increasing number of projections. *Journal of Multivariate Analysis* **124**, 313–332.

[19] Gromenko, O. Kokoszka, P. & M. Reimherr (2017). Detection of change in the spatiotemporal mean function. *Journal of the Royal Statistical Society, Series B* **79**, 29–50.

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

[21] Horváth, L. Kokoszka, P. & R. Reeder (2013). Estimation of the mean of of functional time series and a two sample problem. *Journal of the Royal Statistical Society, Series B* **75**, 103–122.

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

[23] Horváth, L. & G. Rice (2014). Extensions of some classical methods in change point analysis. *Test* **23**, 219–255.

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

[25] Panaretos, V. & S. Tavakoli (2012). Fourier analysis of stationary time series in function space. *The Annals of Statistics* **41**, 568–603.

[26] Parzen, E. (1957). On consistent estimates of the spectrum of stationary time series. *The Annals of Mathematical Statistics* **28**, 329–348.

[27] Politis, D.N. & J.P. Romano (1996). On flat-top spectral density estimators for homogeneous random fields. *Journal of Statistical Planning and Inference* **51**, 41–53.

[28] Rice, G. & H.L. Shang (2017). A plug-in bandwidth selection procedure for long run covariance estimation with stationary functional time series. *Journal of Time Series Analysis* **38**, 591–609.

[29] Shao, X. & X. Zhang (2010). Testing for change points in time series. *Journal of the American Statistical Association* **105**, 1228–1240.

[30] Sharipov, O., Tewes, J. & M. Wendler (2016). Sequential block bootstrap in a Hilbert space with application to change point analysis. *Canadian Journal of Statistics* **44**, 300–322.

[31] Sönmez, O., Aue, A. & G. Rice (2017). fChange: Change point analysis in functional data. R package version 0.1.0.

[32] Stryhn, H. (1996). The location of the maximum of asymmetric two-sided Brownian motion with triangular drift. *Statistics & Probability Letters* **29**, 279–284.

[33] Torgovitski, L. (2016). Detecting changes in Hilbert space data based on "repeated" and change-aligned principal components. Preprint available at `http://arxiv.org/abs/1509.07409`.

[34] Tsudaka, K. & Y. Nishiyama (2014). On L^2 space approach to change point analysis. *Journal of Statistical Planning and Inference* **149**, 46-59.

[35] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. Monographs on Statistics and Applied Probability 60. Chapman & Hall, London.

[36] Zhang, X. Shao, X. Hayhoe, K. & D. Wuebbles (2011). Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electronic Journal of Statistics* **5**, 1765–1796.

arXiv:1701.01741v2 [stat.ME] 8 Oct 2017

Testing for stationarity of functional time series in the frequency domain*†

Alexander Aue‡ Anne van Delft§

October 10, 2017

Abstract

Interest in functional time series has spiked in the recent past with papers covering both methodology and applications being published at a much increased pace. This article contributes to the research in this area by proposing stationarity tests for functional time series based on frequency domain methods. Setting up the tests requires a delicate understanding of periodogram- and spectral density operators that are the functional counterparts of periodogram- and spectral density matrices in the multivariate world. Two sets of statistics are proposed. One is based on the eigendecomposition of the spectral density operator, the other on a fixed projection basis. Their properties are derived both under the null hypothesis of stationary functional time series and under the smooth alternative of locally stationary functional time series. The methodology is theoretically justified through asymptotic results. Evidence from simulation studies and an application to annual temperature curves suggests that the tests work well in finite samples.

Keywords: Frequency domain methods, Functional data analysis, Locally stationary processes, Spectral analysis

MSC 2010: Primary: 62G99, 62H99, Secondary: 62M10, 62M15, 91B84

1 Introduction

The aim of this paper is to provide new stationarity tests for functional time series based on frequency domain methods. Particular attention is given to taking into account alternatives allowing for smooth variation as a source of non-stationarity, even though non-smooth alternatives can be covered as well. Functional data analysis has seen an upsurge in research contributions for at least one decade. This is reflected in the growing number of monographs in the area. Readers interested in the current state of statistical inference procedures

*The authors sincerely thank the Associate Editor and two referees for their constructive comments that helped produce a much improved revision of the original paper.

†AA was partially supported by NSF grants DMS 1305858 and DMS 1407530. AvD 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 in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Project A1, C1, A7) of the German Research Foundation (DFG).

‡Department of Statistics, University of California, Davis, CA 95616, USA, email: aaue@ucdavis.edu

§Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany, email: Anne.vanDelft@rub.de

may consult Bosq (2000), Ferraty & Vieu (2010), Horváth & Kokoszka (2012), Hsing & Eubank (2015) and Ramsay & Silverman (2005).

Papers on functional time series have come into the focus more recently and constitute now an active area of research. Hörmann & Kokoszka (2010) introduced a general weak dependence concept for stationary functional time series, while van Delft & Eichler (2016) provided a framework for locally stationary functional time series. Antoniadis & Sapatinas (2003), Aue et al. (2015) and Besse et al. (2000) constructed prediction methodology that may find application across many areas of science, economics and finance. With the exception of van Delft & Eichler (2016), the above contributions are concerned with procedures in the time domain. Complementing methodology in the frequency domain has been developed in parallel. One should mention Panaretos & Tavakoli (2013), who provided results concerning the Fourier analysis of time series in function spaces, and Hörmann et al. (2015), who addressed the problem of dimension reduction for functional time series using dynamic principal components.

The methodology proposed in this paper provides a new frequency domain inference procedure for functional time series. More precisely, tests for second-order stationarity are developed. In the univariate case, such tests have a long history, going back at least to the seminal paper Priestley & Subba Rao (1969), who based their method on the evaluation of evolutionary spectra of a given time series. Other contributions building on this work include von Sachs & Neumann (2000), who used local periodograms and wavelet analysis, and Paparoditis (2009), whose test is based on comparing a local estimate of the spectral density to a global estimate. Dette et al. (2011) and Preuß et al. (2013) developed methods to derive both a measure of and a test for stationarity in locally stationary time series, the latter authors basing their method on empirical process theory. In all papers, interest is in smoothly varying alternatives. The same tests, however, also have power against non-smooth alternatives such as structural breaks or change-points. A recent review discussing methodology for structural breaks in time series is Aue & Horváth (2013), while Aue et al. (2017) is a recent contribution to structural breaks in functional time series.

The proposed test for second-order stationarity of functional time series uses the Discrete Fourier Transform (DFT). Its construction seeks to exploit that the DFTs of a functional time series evaluated at distinct Fourier frequencies are asymptotically uncorrelated if and only if the series is second-order stationary. The proposed method is therefore related to the initial work of Dwivedi & Subba Rao (2011), who put forth similar tests in a univariate framework. Their method has since been generalized to multivariate time series in Jentsch & Subba Rao (2015) as well as to spatial and spatio-temporal data by Bandyopadhyay & Subba Rao (2017) and Bandyopadhyay et al. (2017), respectively. A different version of functional stationarity tests, based on time domain methodology involving cumulative sum statistics (Aue & Horváth, 2013), was given in Horváth et al. (2014).

Building on the research summarized in the previous paragraph, the tests introduced here are the first of their kind in the frequency domain analysis of functional time series. A delicate understanding of the functional DFT is needed in order to derive the asymptotic theory given here. In particular, results on the large-

sample behavior of a quadratic form test statistic are provided both under the null hypothesis of a stationary functional time series and the alternative of a locally stationary functional time series. For this, a weak convergence result is established that might be of interest in its own right as it is verified using a simplified tightness criterion going back to work of Cremers & Kadelka (1986). The main results are derived under the assumption that the curves are observed in their entirety, corresponding to a setting in which functions are sampled on a dense grid rather than a sparse grid. Differences for these two cases have been worked out in Li & Hsing (2010).

The remainder of the paper is organized as follows. Section 2 provides the background, gives the requisite notations and introduces the functional version of the DFT. The exact form of the hypothesis test, model assumptions and the test statistics are introduced in Section 3. The large-sample behavior under the null hypothesis of second-order stationarity and the alternative of local stationarity is established in Sections 4. Empirical aspects are highlighted in Section 5. The proofs are technical and relegated to the Appendix. Several further auxiliary results are proved in the supplementary document Aue & van Delft (2017), henceforth referred to simply as the Online Supplement.

2 Notation and setup

2.1 The function space

A functional time series $(X_t\colon t \in \mathbb{Z})$ will be viewed in this paper as a sequence of random elements on a probability space $(\Omega, \mathcal{A}, P)$ taking values in the separable Hilbert space of real-valued, square integrable functions on the unit interval $[0,1]$. This Hilbert space will be denoted by $H_{\mathbb{R}} = L^2([0,1],\mathbb{R})$. The functional DFT of $(X_t\colon t \in \mathbb{Z})$, to be introduced in Section 2.3, can then be viewed as an element of $H_{\mathbb{C}} = L^2([0,1],\mathbb{C})$, the complex counterpart of $H_{\mathbb{R}}$. While the interval $[0,1]$ provides a convenient parametrization of the functions, the results of this paper continue to hold for any separable Hilbert space.

The complex conjugate of $z \in \mathbb{C}$ is denoted by $\overline{z}$ and the imaginary number by i. The inner product and the induced norm on $H_{\mathbb{C}}$ are given by

$$\langle f, g\rangle = \int_0^1 f(\tau)\overline{g(\tau)}d\tau \qquad \text{and} \qquad \|f\|_2 = \sqrt{\langle f, f\rangle}, \tag{2.1}$$

respectively, for $f, g \in H_{\mathbb{C}}$. Two elements of $H_{\mathbb{C}}$ are tacitly understood to be equal if their difference has vanishing L_2-norm. More generally, for functions $g\colon [0,1]^k \to \mathbb{C}$, the supremum norm is denoted by

$$\|g\|_\infty = \sup_{\tau_1,\ldots,\tau_k\in[0,1]} |g(\tau_1,\ldots,\tau_k)|$$

and the L^p-norm by

$$\|g\|_p = \left(\int_{[0,1]^k} |g(\tau_1,\ldots,\tau_k)|^p \, d\tau_1\cdots d\tau_k\right)^{1/p}.$$

In all of the above, the obvious modifications apply to $H_{\mathbb{R}}$, the canonical Hilbert space in the functional data analysis setting.

Let H stand more generally for $H_{\mathbb{C}}$, unless otherwise stated. An operator A on H is said to be compact if its pre-image is compact. A compact operator admits a *singular value decomposition*

$$A = \sum_{n=1}^{\infty} s_n(A)\, \psi_n \otimes \phi_n, \tag{2.2}$$

where $(s_n(A)\colon n \in \mathbb{N})$, are the *singular values* of A, $(\phi_n\colon n \in \mathbb{N})$ and $(\psi_n\colon n \in \mathbb{N})$ orthonormal bases of H and $\otimes$ denoting the tensor product. The singular values are ordered to form a monotonically decreasing sequence of non-negative numbers. Based on the convergence rate to zero, operators on H can be classified into particular *Schatten p-classes*. That is, for $p \geqslant 1$, the *Schatten p-class* $S_p(H)$ is the subspace of all compact operators A on H such that the sequence $s(A) = (s_n(A)\colon n \in \mathbb{N})$ of singular values of A belongs to the sequence space ℓ^p, that is,

$$A \in S_p(H) \qquad \text{if and only if} \qquad |\!|\!| A |\!|\!|_p = \Bigg(\sum_{n=1}^{\infty} s_n^p(A) \Bigg)^{1/p} < \infty,$$

where $|\!|\!| A |\!|\!|_p$ is referred to as the *Schatten p-norm*. The space $S_p(H)$ together with the norm $|\!|\!| A |\!|\!|_p$ forms a Banach space and a Hilbert space in case $p = 2$. By convention, the space $S_\infty(H)$ indicates the space of bounded linear operators equipped with the standard operator norm. For $1 \leqslant p \leqslant q$, the inclusion $S_p(H) \subseteq S_q(H)$ is valid. 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, in particular of Hilbert–Schmidt operators, are provided in van Delft & Eichler (2016). Finally, the identity operator is denoted by I_H and the zero operator by O_H.

2.2 Dependence structure on the function space

A functional time series $X = (X_t\colon t \in \mathbb{Z})$ is called strictly stationary if, for all finite sets of indices $J \subset \mathbb{Z}$, the joint distribution of $(X_{t+j}\colon j \in J)$ 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. The L^2-setup provides a point-wise interpretation to these moments; that is, the mean function m of X can be defined using the parametrization $m(\tau) = \mathbb{E}[X_t(\tau)]$, $\tau \in [0,1]$, and the autocovariance kernel c_h at lag $h \in \mathbb{Z}$ by

$$c_h(\tau, \tau') = \mathrm{Cov}(X_{t+h}(\tau), X_t(\tau')), \qquad \tau, \tau' \in [0,1]. \tag{2.3}$$

Both m and c_h are well defined in the L^2-sense if $\mathbb{E}[\|X_0\|_2^2] < \infty$. Each kernel c_h induces a corresponding *autocovariance operator* $\mathcal{C}_h$ on $H_{\mathbb{R}}$ by

$$\mathcal{C}_h\, g(\tau) = \int_0^1 c_h(\tau, \tau')\, g(\tau')\, d\tau' = \mathbb{E}\big[\langle g, X_0 \rangle\, X_h(\tau)\big], \tag{2.4}$$

for all $g \in H_{\mathbb{R}}$. In analogy to weakly stationary multivariate time series, where the covariance matrix and spectral density matrix form a Fourier pair, the *spectral density operator* $\mathcal{F}_\omega$ is given by the Fourier transform of $\mathcal{C}_h$,

$$\mathcal{F}_\omega = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \mathcal{C}_h\, e^{-\mathrm{i}\omega h}. \tag{2.5}$$

A sufficient condition for the existence of $\mathcal{F}_\omega$ in $S_p(H_\mathbb{C})$ is $\sum_{h\in\mathbb{Z}} |||\mathcal{C}_h|||_p < \infty$.

As in Panaretos & Tavakoli (2013), higher-order dependence among the functional observations is defined through cumulant mixing conditions (Brillinger, 1981; Brillinger & Rosenblatt, 1967). For this, the notion of higher-order cumulant tensors is required; see Appendix S1 for their definition and a discussion on their properties for nonstationary functional time series. Point-wise, after setting $t_k = 0$ because of stationarity, the k-th order cumulant function of the process X can be defined by

$$c_{t_1,\ldots,t_{k-1}}(\tau_1,\ldots,\tau_k) = \sum_{\nu=(\nu_1,\ldots,\nu_p)} (-1)^{p-1}\,(p-1)! \prod_{l=1}^{p} \mathbb{E}\Big[\prod_{j\in\nu_l} X_{t_j}(\tau_j)\Big], \tag{2.6}$$

where the summation is over all unordered partitions of $\{1,\ldots,k\}$. The quantity in (2.6) will be referred to as the *k-th order cumulant kernel* if it is properly defined in the L^2-sense. A sufficient condition for this to be satisfied is $\mathbb{E}[\|X_0\|_2^k] < \infty$. The cumulant kernel $c_{t_1,\ldots,t_{2k-1}}(\tau_1,\ldots,\tau_{2k})$ induces a *2k-th order cumulant operator* $\mathcal{C}_{t_1,\ldots,t_{2k-1}}$ through right integration:

$$\mathcal{C}_{t_1,\ldots,t_{2k-1}} g(\tau_1,..,\tau_k) = \int_{[0,1]^k} c_{t_1,\ldots,t_{2k-1}}(\tau_1,\ldots,\tau_{2k}) g(\tau_{k+1},\ldots,\tau_{2k})\, d\tau_{k+1}\cdots d\tau_{2k},$$

which maps from $L^2([0,1]^k,\mathbb{R})$ to $L^2([0,1]^k,\mathbb{R})$. Similar to the case $k=2$, this operator will form a Fourier pair with a $2k$-th order cumulant spectral operator given summability with respect to $|||\cdot|||_p$ is satisfied. The $2k$-th order cumulant spectral operator is specified as

$$\mathcal{F}_{\omega_1,\ldots,\omega_{2k-1}} = (2\pi)^{1-2k} \sum_{t_1,\ldots,t_{2k-1}\in\mathbb{Z}} \mathcal{C}_{t_1,\ldots,t_{2k-1}} \exp\Big(-\mathrm{i}\sum_{j=1}^{2k-1}\omega_j\, t_j\Big), \tag{2.7}$$

where the convergence is in $|||\cdot|||_p$. Under suitable regularity conditions, the corresponding kernels also form a Fourier pair. Note that a similar reasoning can be applied to $2(k+1)$-th order cumulant operators and their associated cumulant spectral operators, using more complex notation.

2.3 The functional discrete Fourier transform

The starting point of this paper is the following proposition that characterizes second-order stationary behavior of a functional time series in terms of a spectral representation.

Proposition 2.1. *A zero-mean, $H_\mathbb{R}$-valued stochastic process $(X_t\colon t\in\mathbb{Z})$ admits the representation*

$$X_t = \int_{-\pi}^{\pi} e^{\mathrm{i}t\omega} dZ_\omega \qquad \textit{a.s. a.e.,} \tag{2.8}$$

where $(Z_\omega\colon \omega\in(-\pi,\pi])$ is a right-continuous functional orthogonal-increment process, if and only if it is weakly stationary.

If the process is not weakly stationary, then a representation in the frequency domain is not necessarily well-defined and certainly not with respect to complex exponential basis functions. It can be shown that a time-dependent functional Cramér representation holds true if the time-dependent characteristics of the process are

captured by a Bochner measurable mapping that is an evolutionary operator-valued mapping in time direction. More details can be found in van Delft & Eichler (2016). In the following the above proposition is utilized to motivate the proposed test procedure.

In practice, the stretch $X_0, \dots, X_{T-1}$ is observed. If the process is weakly stationary, the *functional Discrete Fourier Transform* (fDFT) $D_\omega^{(T)}$ can be seen as an estimate of the increment process Z_ω. This is in accordance with the spatial and spatio-temporal arguments put forward in Jentsch & Subba Rao (2015) and Bandyopadhyay et al. (2017), respectively. At frequency ω, the fDFT is given by

$$D_\omega^{(T)} = \frac{1}{\sqrt{2\pi T}} \sum_{t=0}^{T-1} X_t e^{-\mathrm{i}\omega t}, \tag{2.9}$$

while the functional time series itself can be represented through the inverse fDFT as

$$X_t = \sqrt{\frac{2\pi}{T}} \sum_{j=0}^{T-1} D_{\omega_j}^{(T)} e^{\mathrm{i}\omega_j t}. \tag{2.10}$$

Under weak dependence conditions expressed through higher-order cumulants, the fDFTs evaluated at distinct frequencies yield asymptotically independent Gaussian random elements in $H_{\mathbb{C}}$ (Panaretos & Tavakoli, 2013). The fDFT sequence of a Hilbert space stationary process is in particular asymptotically uncorrelated at the canonical frequencies $\omega_j = 2\pi j/T$. For weakly stationary processes, it will turn out specifically that, for $j \neq j'$ or $j \neq T - j'$, the covariance of the fDFT satisfies $\langle \mathrm{Cov}(D_{\omega_j}^{(T)}, D_{\omega_{j'}}^{(T)}) g_1, g_2 \rangle = O(1/T)$ for all $g_1, g_2 \in H$. This fundamental result will be extended to locally stationary functional time series in this paper by providing a representation of the cumulant kernel of the fDFT sequence in terms of the (time-varying) spectral density kernel. Similar to the above, the reverse argument (uncorrelatedness of the functional DFT sequence implies weak stationarity) can be shown by means of the inverse fDFT. Using expression (2.9), the covariance operator of X_t and $X_{t'}$ in terms of the fDFT sequence can be written as

$$\begin{aligned} \mathcal{C}_{t,t'} &= \mathbb{E}[X_t \otimes X_{t'}] \\ &= \frac{2\pi}{T} \sum_{j,j'=0}^{T-1} \mathbb{E}[D_{\omega_j}^{(T)} \otimes D_{\omega_{j'}}^{(T)}] e^{\mathrm{i}t\omega_j - \mathrm{i}t'\omega_{j'}} \\ &= \frac{2\pi}{T} \sum_{j=0}^{T-1} \mathbb{E}[I_{\omega_j}^{(T)}] e^{\mathrm{i}\omega_j (t-t')} \qquad (2.11) \\ &= \mathcal{C}_{t-t'}, \end{aligned}$$

where the equality in (2.11) holds in an L^2-sense when $\langle \mathbb{E}[D_{\omega_j}^{(T)} \otimes D_{\omega_{j'}}^{(T)}] g_1, g_2 \rangle = 0$ for all $g_1, g_2 \in H$ with $j \neq j'$ or $j \neq T - j'$ and where $I_{\omega_j}^{(T)} = D_{\omega_{j'}}^{(T)} \otimes D_{\omega_{j'}}^{(T)}$ is the periodogram tensor. This demonstrates that the autocovariance kernel of a second-order stationary functional time series is obtained and, hence, that an uncorrelated DFT sequence implies second-order stationarity up to lag T. Below, the behavior of the fDFT under the smooth alternative of locally stationary functional time series is derived. These properties will then be exploited to set up a testing framework for functional stationarity. Although this work is therefore related

to Dwivedi & Subba Rao (2011) — who similarly exploited analogous properties of the DFT of a stationary time series — it should be emphasized that the functional setting is a nontrivial extension.

3 The functional stationarity testing framework

This section gives precise formulations of the hypotheses of interest, states the main assumptions of the paper and introduces the test statistics. Throughout, interest is in testing the null hypothesis

$$H_0\colon (X_t\colon t\in\mathbb{Z}) \text{ is a stationary functional time series}$$

versus the alternative

$$H_A\colon (X_t\colon t\in\mathbb{Z}) \text{ is a locally stationary functional time series,}$$

where locally stationary functional time series are defined as follows.

Definition 3.1. *A stochastic process $(X_t\colon t\in\mathbb{Z})$ taking values in $H_{\mathbb{R}}$ is said to be locally stationary if*

(1) $X_t = X_t^{(T)}$ for $t = 1,\dots,T$ and $T\in\mathbb{N}$; and

(2) for any rescaled time $u\in[0,1]$, there is a strictly stationary process $(X_t^{(u)}\colon t\in\mathbb{Z})$ such that

$$\big\|X_t^{(T)} - X_t^{(u)}\big\|_2 \leqslant \Big(\Big|\tfrac{t}{T} - u\Big| + \tfrac{1}{T}\Big) P_{t,T}^{(u)} \qquad a.s.,$$

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

Note that, under H_A, the process constitutes a triangular array of functions. Inference methods are then based on in-fill asymptotics as popularized in Dahlhaus (1997) for univariate time series. The process is therefore observed on a finer grid as T increases and more observations are available at a local level. A rigorous statistical framework for locally stationary functional time series was recently provided in van Delft & Eichler (2016). These authors established in particular that linear functional time series can be defined by means of a functional time-varying Cramér representation and provided sufficient conditions in the frequency domain for the above definition to be satisfied.

Based on the observations in Section 2.3, a test for weak stationarity can be set up exploiting the uncorrelatedness of the elements in the sequence $(D_{\omega_j}^{(T)})$. Standardizing these quantities is a delicate issue as the spectral density operators $\mathcal{F}_{\omega_j}^{(T)}$ are not only unknown but generally not globally invertible. Here, a statistic based on projections is considered. Let $(\psi_l\colon l\in\mathbb{N})$ be an orthonormal basis of $H_{\mathbb{C}}$. Then, $(\psi_l\otimes\psi_{l'}\colon l,l'\in\mathbb{N})$ is an orthonormal basis of $L^2([0,1]^2,\mathbb{C})$ and, by definition of the Hilbert–Schmidt inner product on the algebraic tensor product space $H\otimes H$,

$$\langle\mathbb{E}\big[D_{\omega_{j_1}}^{(T)}\otimes D_{\omega_{j_2}}^{(T)}\big], \psi_l\otimes\psi_{l'}\rangle_{H\otimes H} = \mathbb{E}[\langle D_{\omega_{j_1}}^{(T)},\psi_l\rangle\overline{\langle D_{\omega_{j_2}}^{(T)},\psi_{l'}\rangle}] = \mathrm{Cov}\big(\langle D_{\omega_{j_1}}^{(T)},\psi_l\rangle, \overline{\langle D_{\omega_{j_2}}^{(T)},\psi_{l'}\rangle}\big).$$

This motivates to set up test statistics based on the quantities

$$\gamma_h^{(T)}(l,l') = \frac{1}{T}\sum_{j=1}^{T} \frac{\langle D_{\omega_j}^{(T)}, \psi_l\rangle\overline{\langle D_{\omega_{j+h}}^{(T)}, \psi_{l'}\rangle}}{\sqrt{\langle \mathcal{F}_{\omega_j}(\psi_l), \psi_l\rangle\langle \mathcal{F}_{\omega_{j+h}}^{\dagger}(\psi_{l'}), \psi_{l'}\rangle}}, \qquad h = 1,\dots,T-1. \tag{3.1}$$

A particularly interesting (random) projection choice is provided by the eigenfunctions $\phi_l^{\omega_j}$ and $\phi_l^{\omega_{j+h}}$ corresponding to the l-th largest eigenvalues of the spectral density operators $\mathcal{F}_{\omega_j}$ and $\mathcal{F}_{\omega_{j+h}}$, respectively. In this case, based on orthogonality relations the statistic can be simplified to

$$\gamma_h^{(T)}(l) = \frac{1}{T}\sum_{j=1}^{T} \frac{\langle D_{\omega_j}^{(T)}, \phi_l^{\omega_j}\rangle\overline{\langle D_{\omega_{j+h}}^{(T)}, \phi_l^{\omega_{j+h}}\rangle}}{\sqrt{\lambda_l^{-\omega_j}\lambda_l^{\omega_{j+h}}}}, \qquad h = 1,\dots,T-1, \tag{3.2}$$

and thus depends on l only. In the following the notation $\gamma_h^{(T)}$ is used to refer both to a fixed projection basis used for dimension reduction at both frequencies ω_j and ω_{j+h} as in (3.1) as well as for the random projection based method in (3.2) using two separate Karhunen–Loève decompositions at the two frequencies of $\mathcal{F}_{\omega_j}$ and $\mathcal{F}_{\omega_{j+h}}$, respectively. The write-up will generally present theorems for both tests if the formulation allows, but will focus on the latter case if certain aspects have to be treated differently. The counterparts for the former case are then collected in the appendix for completeness.

In practice, the unknown spectral density operators $\mathcal{F}_{\omega_j}$ and $\mathcal{F}_{\omega_{j+h}}$ are to be replaced with consistent estimators $\hat{\mathcal{F}}_{\omega_j}^{(T)}$ and $\hat{\mathcal{F}}_{\omega_{j+h}}^{(T)}$. Similarly the population eigenvalues and eigenfunctions need to be replaced by the respective sample eigenvalues and eigenfunctions. The estimated quantity corresponding to (3.1) and (3.2) will be denoted by $\hat{\gamma}_h^{(T)}$. Note that the large-sample distributional properties of tests based on (3.1) and (3.2) are similar. Some additional complexity enters for the statistic in (3.2) because the unknown eigenfunctions can only be identified up to rotation on the unit circle and it has to be ensured that the proposed test statistic is invariant to this rotation; see Section E.2 for the details. As an estimator of $\mathcal{F}_{\omega}^{(T)}$, take

$$\hat{\mathcal{F}}_{\omega}^{(T)} = \frac{2\pi}{T}\sum_{j=1}^{T} K_b(\omega-\omega_j)\big([D_{\omega_j}^{(T)}]\otimes[D_{\omega_j}^{(T)}]^{\dagger}\big), \tag{3.3}$$

where $K_b(\cdot) = \frac{1}{b}K(\frac{\cdot}{b})$ is a window function satisfying the following conditions.

Assumption 3.1. Let $K\colon [-\frac{1}{2},\frac{1}{2}] \to \mathbb{R}$ be a positive, symmetric window function with $\int K(x)dx = 1$ and $\int K(x)^2dx < \infty$ that is periodically extended, i.e., $K_b(x) = \frac{1}{b}K(\frac{x\pm 2\pi}{b})$.

The periodic extension is to include estimates for frequencies around $\pm\pi$. Further conditions on the bandwidth b are imposed below (Section 4) to determine the large-sample behavior of $\hat{\gamma}_h^{(T)}$ under both the null and the alternative. The replacement of the unknown operators with consistent estimators requires to derive the order of the difference

$$\sqrt{T}\big|\gamma_h^{(T)} - \hat{\gamma}_h^{(T)}\big|. \tag{3.4}$$

It will be shown in the next section that, for appropriate choices of the bandwidth b, this term is negligible under both null and alternative hypothesis.

To set up the test statistic, it now appears reasonable to extract information across a range of directions $l, l' = 1, \dots, L$ and a selection of lags $h = 1, \dots, \bar{h}$, where $\bar{h}$ denotes an upper limit. Build therefore first the $L \times L$ matrix $\hat{\Gamma}_h^{(T)} = (\hat{\gamma}_h^{(T)}(l, l') \colon l, l' = 1, \dots, L)$ and construct the vector $\hat{\boldsymbol{\gamma}}_h^{(T)} = \mathrm{vec}(\hat{\Gamma}_h^{(T)})$ by vectorizing $\hat{\Gamma}_h^{(T)}$ via stacking of its columns. Define

$$\hat{\boldsymbol{\beta}}_h^{(T)} = \boldsymbol{e}^\top \hat{\boldsymbol{\gamma}}_h^{(T)}, \tag{3.5}$$

where $\boldsymbol{e}$ is a vector of dimension L^2 whose elements are all equal to one and $^\top$ denotes transposition. Note that for (3.2), a somewhat simplified expression is obtained, since the off-diagonal terms $l \neq l'$ do not have to be accounted for. Dropping these components reduces the number of terms in the sum (3.5) from L^2 to L, even though the general form of the equation continues to hold. Choose next a collection $h_1, \dots, h_M$ of lags each of which is upper bounded by $\bar{h}$ to pool information across a number of autocovariances and build the (scaled) vectors

$$\sqrt{T}\hat{\boldsymbol{b}}_M^{(T)} = \sqrt{T}\big(\Re\hat{\boldsymbol{\beta}}_{h_1}^{(T)}, \dots, \Im\hat{\boldsymbol{\beta}}_{h_M}^{(T)}, \Im\hat{\boldsymbol{\beta}}_{h_1}^{(T)}, \dots, \Im\hat{\boldsymbol{\beta}}_{h_M}^{(T)}\big)^\top,$$

where $\Re$ and $\Im$ denote real and imaginary part, respectively. Finally, set up the quadratic form

$$\hat{Q}_M^{(T)} = T(\hat{\boldsymbol{b}}_M^{(T)})^\top \hat{\Sigma}_M^{-1} \hat{\boldsymbol{b}}_M^{(T)}, \tag{3.6}$$

where $\hat{\Sigma}_M$ is an estimator of the asymptotic covariance matrix of the vectors $\boldsymbol{b}_M^{(T)}$ which are defined by replacing $\hat{\gamma}_h^{(T)}$ with $\gamma_h^{(T)}$ in the definition of $\hat{\boldsymbol{b}}_M^{(T)}$. The statistic $\hat{Q}_M^{(T)}$ will be used to test the null of stationarity against the alternative of local stationarity. Note that this quadratic form depends on the tuning parameters L and M. These effects will be briefly discussed in Section 5.

4 Large-sample results

4.1 Properties under the null of stationarity

The following gives the main requirements under stationarity of the functional time series that are needed to establish the asymptotic behavior of the test statistics under the null hypothesis.

Assumption 4.1 (Stationary functional time series). *Let $(X_t \colon t \in \mathbb{Z})$ be a stationary functional time series with values in $H_{\mathbb{R}}$ such that*

(i) $\mathbb{E}[\|X_0\|_2^k] < \infty$,

(ii) $\sum_{t_1,\dots,t_{k-1}=-\infty}^{\infty}(1 + |t_j|^\ell)\|c_{t_1,\dots,t_{k-1}}\|_2 < \infty$ *for all* $1 \leqslant j \leqslant k-1$,

for some fixed values of $k, \ell \in \mathbb{N}$.

The conditions of Assumption 4.1 ensure that the k-th order cumulant spectral density kernel

$$f_{\omega_1,\dots,\omega_{k-1}}(\tau_1, \dots, \tau_k) = \frac{1}{(2\pi)^{k-1}} \sum_{t_1,\dots,t_{k-1}=-\infty}^{\infty} c_{t_1,\dots,t_{k-1}}(\tau_1, \dots, \tau_k) e^{-\mathrm{i}(\sum_{j=1}^k \omega_j t_j)} \tag{4.1}$$

is well-defined in L^2 and is uniformly continuous in ω with respect to $\|\cdot\|_2$. Additionally, the parameter ℓ controls the smoothness of f_ω in the sense that, for all $i \leqslant \ell$,

$$\sup_{\omega} \left\| \frac{\partial^i}{\partial \omega^i} f_\omega \right\|_2 < \infty. \tag{4.2}$$

A proof of these facts can be found in Panaretos & Tavakoli (2013). Through right-integration, the function (4.1) induces a k-th order cumulant spectral density operator $\mathcal{F}_{\omega_1,\ldots,\omega_{k-1}}$ which is Hilbert–Schmidt. The following theorem establishes that the scaled difference between $\gamma_h^{(T)}$ and $\hat{\gamma}_h^{(T)}$ is negligible in large samples.

Theorem 4.1. *Let Assumptions 3.1 and 4.1 be satisfied with $k = 8$ and $\ell = 2$ and assume further that $\inf_\omega \langle \mathcal{F}(\psi), \psi \rangle > 0$ for $\psi \in H$. Then, for any fixed h,*

$$\sqrt{T} |\gamma_h^{(T)} - \hat{\gamma}_h^{(T)}| = O_p\left(\frac{1}{\sqrt{bT}} + b^2 \right) \qquad (T \to \infty).$$

The proof is given in Section S2 of the Online Supplement. For the eigen-based statistic (3.2), the condition $\inf_\omega \langle \mathcal{F}(\psi), \psi \rangle > 0$ reduces to $\inf_\omega \lambda_L^\omega > 0$. In practice this means that only those eigenfunctions should be included that belong to the l-largest eigenvalues in the quadratic form (3.6). This is discussed in more detail in Section 5, where a criterion is proposed to choose L. Theorem 4.1 shows that the distributional properties of $\hat{\gamma}_h^{(T)}$ are asymptotically the same as those of $\gamma_h^{(T)}$, or both versions (3.1) and (3.2) of the test, provided that the following extra condition on the bandwidth holds.

Assumption 4.2. *The bandwidth b satisfies $b \to 0$ such that $bT \to \infty$ as $T \to \infty$.*

Note that these rates are in fact necessary for the estimator in (3.3) to be consistent, see Panaretos & Tavakoli (2013), and therefore do not impose an additional constraint under H_0. The next theorem derives the second-order structure of $\gamma_h^{(T)}$ for the eigenbased statistics (3.2). It shows that the asymptotic variance is uncorrelated for all lags h and that there is no correlation between the real and imaginary parts.

Theorem 4.2. *Let Assumption 4.1 be satisfied with $k = \{2, 4\}$. Then,*

$$\begin{aligned} T \operatorname{Cov}\Big(\Re\gamma_{h_1}^{(T)}(l_1), \Re\gamma_{h_2}^{(T)}(l_2) \Big) &= T \operatorname{Cov}\Big(\Im\gamma_{h_1}^{(T)}(l_1), \Im\gamma_{h_2}^{(T)}(l_2) \Big) \\ &= \begin{cases} \dfrac{\delta_{l_1,l_2}}{2} + \dfrac{1}{4\pi} \displaystyle\int\int \frac{\langle \mathcal{F}_{\omega,-\omega-\omega_h,-\omega'}(\phi_{l_2}^{\omega'} \otimes \phi_{l_2}^{\omega'+\omega_h'}), \phi_{l_1}^{\omega} \otimes \phi_{l_1}^{\omega+\omega_h} \rangle}{\sqrt{\lambda_{l_1}^{\omega} \lambda_{l_2}^{-\omega'} \lambda_{l_1}^{-\omega-\omega_h} \lambda_{l_2}^{\omega'+\omega_h'}}} d\omega d\omega', & \text{if } h_1 = h_2 = h, \\ O(\frac{1}{T}), & \text{if } h_1 \neq h_2, \end{cases} \end{aligned}$$

where $\delta_{i,j} = 1$ if $i = j$ and 0 otherwise. Furthermore,

$$T \operatorname{Cov}(\Re\gamma_{h_1}^{(T)}(l_1), \Im\gamma_{h_2}^{(T)}(l_2)) = O\left(\frac{1}{T} \right)$$

uniformly in $h_1, h_2 \in \mathbb{Z}$.

The proof of Theorem 4.2 is given in Appendix D.1, its counterpart for (3.1) is stated as Theorem C.1 in Appendix C. Note that the results in the theorem use at various instances the fact that the k-th order spectral density operator at frequency $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_k)^T \in \mathbb{R}^k$ is equal to the k-th order spectral density operator at frequency $-\boldsymbol{\omega}$ in the manifold $\sum_{j=1}^k \omega_j \mod 2\pi$.

With the previous results in place, the large-sample behavior of the quadratic form statistics $\hat{Q}_M^{(T)}$ defined in (3.6) can be derived. This is done in the following theorem.

Theorem 4.3. *Let Assumptions 3.1 and 4.2 be satisfied. Let Assumption 4.1 be satisfied with $k \geqslant 1$ and $\ell = 2$ and assume that $\inf_\omega \lambda_L^\omega > 0$. Then,*

(a) For any collection $h_1, \ldots, h_M$ bounded by $\bar{h}$,

$$\sqrt{T}\boldsymbol{b}_M^{(T)} \xrightarrow{\mathcal{D}} \mathcal{N}_{2M}(\mathbf{0}, \Sigma_0) \qquad (T \to \infty),$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution and $\mathcal{N}_{2M}(\mathbf{0}, \Sigma_0)$ a $2M$-dimensional normal distribution with mean $\mathbf{0}$ and diagonal covariance matrix $\Sigma_0 = \operatorname{diag}(\sigma_{0,m}^2 \colon m = 1, \ldots, 2M)$ whose elements are

$$\sigma_{0,m}^2 = \lim_{T\to\infty} \sum_{l_1,l_2=1}^{L} T\operatorname{Cov}\big(\Re\gamma_{h_m}^{(T)}(l_1), \Re\gamma_{h_m}^{(T)}(l_2)\big), \qquad m = 1, \ldots, M,$$

and $\sigma_{0,M+m}^2 = \sigma_{0,m}^2$. The explicit form of the limit is given by Theorem 4.2.

(b) Using the result in (a), it follows that

$$\hat{Q}_M^{(T)} \xrightarrow{\mathcal{D}} \chi_{2M}^2 \qquad (T \to \infty),$$

where χ_{2M}^2 is a χ^2-distributed random variable with $2M$ degrees of freedom.

The proof of Theorem 4.3 is provided in Appendix D.1. Part (b) of the theorem can now be used to construct tests with asymptotic level α. Theorem C.2 contains the result for the test based on (3.1). To better understand the power of the test, the next section investigates the behavior under the alternative of local stationarity.

4.2 Properties under the alternative

This section contains the counterparts of the results in Section 4.1 for locally stationary functional time series. The following conditions are essential for the large-sample results to be established here.

Assumption 4.3. Assume $(X_t^{(T)} \colon t \leqslant T, T \in \mathbb{N})$ and $(X_t^{(u)} \colon t \in \mathbb{Z})$ are as in Definition 3.1 and let $\kappa_{k;t_1,\ldots,t_{k-1}}$ be a positive sequence in $L^2([0,1]^k, \mathbb{R})$ independent of T such that, for all $j = 1, \ldots, k-1$ and some $\ell \in \mathbb{N}$,

$$\sum_{t_1,\ldots,t_{k-1}\in\mathbb{Z}} (1 + |t_j|^\ell)\|\kappa_{k;t_1,\ldots,t_{k-1}}\|_2 < \infty. \tag{4.3}$$

Suppose furthermore that there exist representations

$$X_t^{(T)} - X_t^{(t/T)} = Y_t^{(T)} \qquad \text{and} \qquad X_t^{(u)} - X_t^{(v)} = (u-v)Y_t^{(u,v)}, \tag{4.4}$$

for some processes $(Y_t^{(T)} \colon t \leqslant T, T \in \mathbb{N})$ and $(Y_t^{(u,v)} \colon t \in \mathbb{Z})$ taking values in $H_{\mathbb{R}}$ whose k-th order joint cumulants satisfy

(i) $\|\mathrm{cum}(X_{t_1}^{(T)}, \dots, X_{t_{k-1}}^{(T)}, Y_{t_k}^{(T)})\|_2 \leqslant \frac{1}{T}\|\kappa_{k;t_1-t_k,\dots,t_{k-1}-t_k}\|_2$,

(ii) $\|\mathrm{cum}(X_{t_1}^{(u_1)}, \dots, X_{t_{k-1}}^{(u_{k-1})}, Y_{t_k}^{(u_k,v)})\|_2 \leqslant \|\kappa_{k;t_1-t_k,\dots,t_{k-1}-t_k}\|_2$,

(iii) $\sup_u \|\mathrm{cum}(X_{t_1}^{(u)}, \dots, X_{t_{k-1}}^{(u)}, X_{t_k}^{(u)})\|_2 \leqslant \|\kappa_{k;t_1-t_k,\dots,t_{k-1}-t_k}\|_2$,

(iv) $\sup_u \|\frac{\partial^\ell}{\partial u^\ell}\mathrm{cum}(X_{t_1}^{(u)}, \dots, X_{t_{k-1}}^{(u)}, X_{t_k}^{(u)})\|_2 \leqslant \|\kappa_{k;t_1-t_k,\dots,t_{k-1}-t_k}\|_2$.

Note that these assumptions are generalizations of the ones in Lee & Subba Rao (2016), who investigated the properties of quadratic forms of stochastic processes in a finite-dimensional setting. For fixed u_0, the process $(X_t^{(u_0)} \colon t \in \mathbb{Z})$ is stationary and thus the results of van Delft & Eichler (2016) imply that the *local k-th order cumulant spectral kernel*

$$f_{u_0;\omega_1,\dots,\omega_{k-1}}(\tau_1, \dots, \tau_k) = \frac{1}{(2\pi)^{k-1}} \sum_{t_1,\dots,t_{k-1} \in \mathbb{Z}} c_{u_0;t_1,\dots,t_{k-1}}(\tau_1, \dots, \tau_k) e^{-\mathrm{i}\sum_{l=1}^{k-1}\omega_l t_l} \tag{4.5}$$

exists, where $\omega_1, \dots, \omega_{k-1} \in [-\pi, \pi]$ and

$$c_{u_0;t_1,\dots,t_{k-1}}(\tau_1, \dots, \tau_k) = \mathrm{cum}\big(X_{t_1}^{(u_0)}(\tau_1), \dots, X_{t_{k-1}}^{(u_0)}(\tau_{k-1}), X_{t_0}^{(u_0)}(\tau_k)\big) \tag{4.6}$$

is the corresponding local cumulant kernel of order k at time u_0. The quantity $f_{u,\omega}$ will be referred to as the *time-varying spectral density kernel* of the stochastic process $(X_t^{(T)} \colon t \leqslant T, T \in \mathbb{N})$. Under the given assumptions, this expression is formally justified by Lemma S1.2.

Because of the standardization necessary in $\hat{\gamma}^{(T)}$, it is of importance to consider the properties of the estimator (3.3) in case the process is locally stationary. The next theorem shows that it is a consistent estimator of the integrated time-varying spectral density operator

$$G_\omega = \int_0^1 \mathcal{F}_{u,\omega} du,$$

where the convergence is uniform in $\omega \in [-\pi, \pi]$ with respect to $|||\cdot|||_2$. This therefore becomes an operator-valued function in ω that acts on H and is independent of rescaled time u.

Theorem 4.4 (**Consistency and uniform convergence**). *Suppose $(X_t^{(T)} \colon t \leqslant T, T \in \mathbb{N})$ satisfies Assumption 4.3 for $\ell = 2$ and consider the estimator $\hat{\mathcal{F}}_\omega^{(T)}$ in* (3.3) *with bandwidth fulfilling Assumption 3.1. Then,*

(i) $\mathbb{E}[|||\hat{\mathcal{F}}_\omega^{(T)} - G_\omega|||_2^2] = O((bT)^{-1} + b^4)$;

(ii) $\sup_{\omega \in [-\pi,\pi]} |||\hat{\mathcal{F}}_\omega^{(T)} - G_\omega|||_2 \xrightarrow{p} 0$,

uniformly in $\omega \in [-\pi, \pi]$.

The proof of Theorem 4.4 is given in Section S4 of the Appendix. Under the conditions of this theorem, the sample eigenelements $(\hat{\lambda}_l^\omega, \hat{\phi}_l^\omega : l \in \mathbb{N})$ of $\hat{\mathcal{F}}_\omega$ are consistent for the eigenelements $(\tilde{\lambda}_l^\omega, \tilde{\phi}_l^\omega : l \in \mathbb{N})$ of G_ω; see Mas & Menneteau (2003).

To obtain the distributional properties of $\hat{\gamma}^{(T)}$, it is necessary to replace the denominator with its deterministic limit. In analogy with Theorem 4.1, the theorem below gives conditions on the bandwidth for which this is justified under the alternative.

Theorem 4.5. *Let Assumption 4.3 be satisfied with $k = 8$ and $\ell = 2$ and assume that $\inf_\omega \langle G_\omega(\psi), \psi\rangle > 0$ for all $\psi \in H_{\mathbb{C}}$. Then,*

$$\sqrt{T}|\gamma_h^{(T)} - \hat{\gamma}_h^{(T)}| = O_p\bigg(\frac{1}{\sqrt{bT}} + b^2 + \frac{1}{b\sqrt{T}}\bigg) \qquad (T \to \infty).$$

The proof of Theorem 4.5 is given in Section S2 of the Online Supplement. The theorem shows that for $\hat{\gamma}_h^{(T)}$ to have the same asymptotic sampling properties as $\gamma_h^{(T)}$, additional requirements on the bandwidth b are needed. These are stated next.

Assumption 4.4. *The bandwidth b satisfies $b \to 0$ such that $b\sqrt{T} \to \infty$ as $T \to \infty$.*

The conditions imply that the bandwidth should tend to zero at a slower rate than in the stationary case. The conditions on the bandwidth imposed in this paper are in particular weaker than the ones required by Dwivedi & Subba Rao (2011) in the finite-dimensional context.

The dependence structure of $\gamma_h^{(T)}$ under the alternative is more involved than under the null of stationarity because the mean is nonzero for $h \neq 0 \mod T$. Additionally, the real and imaginary components of the covariance structure are correlated. The following theorem is for the test based on (3.2).

Theorem 4.6. *Let Assumption 4.3 be satisfied with $k = \{2, 4\}$. Then, for $h = 1, \ldots, T-1$,*

$$\mathbb{E}\big[\gamma_h^{(T)}(l)\big] = \frac{1}{2\pi}\int_0^{2\pi}\int_0^1 \frac{\langle \mathcal{F}_{u,\omega}\tilde{\phi}_l^{\omega+\omega_h}, \tilde{\phi}_l^\omega\rangle e^{-\mathrm{i}2\pi uh}}{\sqrt{\tilde{\lambda}_l^\omega \tilde{\lambda}_l^{\omega+\omega_h}}} du d\omega + O\bigg(\frac{1}{T}\bigg) = O\bigg(\frac{1}{h^2}\bigg) + O\bigg(\frac{1}{T}\bigg). \tag{4.7}$$

The covariance structure satisfies

1. $T\,\mathrm{Cov}(\Re\gamma_{h_1}^{(T)}(l_1), \Re\gamma_{h_2}^{(T)}(l_2)) =$
$$\frac{1}{4}\big[\Sigma_{h_1,h_2}^{(T)}(l_1,l_2) + \acute{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2) + \grave{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2) + \bar{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2)\big] + R_T,$$
2. $T\,\mathrm{Cov}(\Re\gamma_{h_1}^{(T)}(l_1,l_2), \Im\gamma_{h_2}^{(T)}(l_3,l_4)) =$
$$\frac{1}{4\mathrm{i}}\big[\Sigma_{h_1,h_2}^{(T)}(l_1,l_2) - \acute{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2) + \grave{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2) - \bar{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2)\big] + R_T,$$
3. $T\,\mathrm{Cov}(\Im\gamma_{h_1}^{(T)}(l_1,l_2), \Im\gamma_{h_2}^{(T)}(l_3,l_4)) =$
$$\frac{1}{4}\big[\Sigma_{h_1,h_2}^{(T)}(l_1,l_2) - \acute{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2) - \grave{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2) + \bar{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2)\big] + R_T,$$

where $\|R_T\|_2 = O(T^{-1})$ *and where* $\Sigma^{(T)}_{h_1,h_2}(l_1,l_2), \acute{\Sigma}^{(T)}_{h_1,h_2}(l_1,l_2)\grave{\Sigma}^{(T)}_{h_1,h_2}(l_1,l_2), \bar{\Sigma}^{(T)}_{h_1,h_2}(l_1,l_2)$ *are derived in the Appendix and general expressions for* (3.1) *are defined in equations (S.5.1)–(S.5.4) of the Online Supplement.*

The proof of Theorem 4.6 is given in Section S4 of the Appendix. Its companion theorem is stated as Theorem C.3 in Section C. The last result in this section concerns the asymptotic of $\hat{Q}^{(T)}_M$ in (3.6) in the locally stationary setting. Before stating this result, observe that the previous theorem shows that a noncentrality parameter will have to enter, since the mean of $\gamma^{(T)}_h(l)$ is nonzero. Henceforth, the limit of (4.7) shall be denoted by

$$\mu_h(l) = \frac{1}{2\pi}\int_0^{2\pi}\int_0^1 \frac{\langle \mathcal{F}_{u,\omega}\tilde{\phi}_l^{\omega+\omega_h}, \tilde{\phi}_l^{\omega}\rangle e^{-\mathrm{i}2\pi uh}}{\sqrt{\tilde{\lambda}_l^{\omega}\tilde{\lambda}_l^{\omega+\omega_h}}} du d\omega$$

and its vectorization by $\boldsymbol{\mu}_h$. Following Paparoditis (2009) and Dwivedi & Subba Rao (2011), there is an intuitive interpretation of the degree of nonstationarity that can be detected in these functions. For fixed l and small h, they can be seen to approximate the Fourier coefficients of the function $\langle \mathcal{F}_{u,\omega}\tilde{\phi}_l^{\omega+\omega_h}, \tilde{\phi}_l^{\omega}\rangle/(\tilde{\lambda}_l^{\omega}\tilde{\lambda}_l^{\omega+\omega_h})^{1/2}$. More specifically, for small h and $T \to \infty$, they approximate

$$\vartheta_{h,j}(l) = \frac{1}{2\pi}\int_0^{2\pi}\int_0^1 \frac{\langle \mathcal{F}_{u,\omega}\tilde{\phi}_l^{\omega+\omega_h}, \tilde{\phi}_l^{\omega}\rangle}{\sqrt{\tilde{\lambda}_l^{\omega}\tilde{\lambda}_l^{\omega+\omega_h}}} e^{\mathrm{i}2\pi uh - \mathrm{i}j\omega} du d\omega.$$

Thus, $\mu_h(l) \approx \vartheta_{h,0}(l)$. If the process is weakly stationary, then $\mathcal{F}_{u,\omega} \equiv \mathcal{F}_\omega$ and the eigenelements reduce to $\lambda_l^{\omega}, \lambda_l^{\omega+h}$ and $\phi_l^{\omega}, \phi_l^{\omega+h}$, respectively, and hence the integrand of the coefficients does not depend on u. All Fourier coefficients are zero except $\vartheta_{0,j}(l)$. In particular, $\vartheta_{0,0}(l) = 1$. Observe that, for the statistic (3.1), the above becomes $\vartheta_{h,j}(l,l')$ and, for $h = 0$, the off-diagonal elements yield coherence measures. The mean functions can now be seen to reveal long-term non-stationary behavior. Unlike testing methods based on segments in the time domain, the proposed method is consequently able to detect smoothly changing behavior in the temporal dependence structure. A precise formulation of the asymptotic properties of the test statistic in (3.2) under H_A is given in the next theorem.

Theorem 4.7. *Let Assumptions 3.1 and 4.4 be satisfied. Let Assumption 4.3 be satisfied* $k \geqslant 1$ *and* $\ell = 2$ *and assume that* $\inf_\omega \lambda_L^\omega > 0$*. Then,*

(a) For any collection $h_1, \ldots, h_M$ *bounded by* $\bar{h}$*,*

$$\sqrt{T}\boldsymbol{b}_M^{(T)} \xrightarrow{\mathcal{D}} \mathcal{N}_{2M}(\boldsymbol{\mu}, \Sigma_A) \qquad (T \to \infty),$$

where $\mathcal{N}_{2M}(\boldsymbol{\mu}, \Sigma_A)$ *denotes a* $2M$*-dimensional normal distribution with mean vector* $\boldsymbol{\mu}$*, whose first* M *components are* $\Re \boldsymbol{e}^T\boldsymbol{\mu}_{h_m}$ *and last* M *components are* $\Im \boldsymbol{e}^T\boldsymbol{\mu}_{h_m}$*, and block covariance matrix*

$$\Sigma_A = \begin{pmatrix} \Sigma_A^{(11)} & \Sigma_A^{(12)} \\ \Sigma_A^{(21)} & \Sigma_A^{(22)}. \end{pmatrix}$$

whose $M \times M$ *blocks are, for* $m, m' = 1, \ldots, M$*, given by*

$$\Sigma_A^{(11)}(m,m') = \lim_{T\to\infty}\sum_{l_1,l_2=1}^{L}\frac{1}{4}\big[\Sigma^{(T)}_{h_1,h_2}(l_1,l_2) + \acute{\Sigma}^{(T)}_{h_1,h_2}(l_1,l_2) + \grave{\Sigma}^{(T)}_{h_1,h_2}(l_1,l_2) + \bar{\Sigma}^{(T)}_{h_1,h_2}(l_1,l_2)\big],$$

$$\Sigma_A^{(12)}(m,m') = \lim_{T\to\infty} \sum_{l_1,l_2=1}^{L} \frac{1}{4\mathrm{i}}\big[\Sigma_{h_1,h_2}^{(T)}(l_1,l_2) - \acute{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2) + \grave{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2) - \bar{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2)\big],$$

$$\Sigma_A^{(22)}(m,m') = \lim_{T\to\infty} \sum_{l_1,l_2=1}^{L} \frac{1}{4}\big[\Sigma_{h_1,h_2}^{(T)}(l_1,l_2) - \acute{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2) - \grave{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2) + \bar{\Sigma}_{h_1,h_2}^{(T)}(l_1,l_2)\big],$$

where $\Sigma_{h_m,h_{m'}}^{(T)}(l_1,l_2), \acute{\Sigma}_{h_m,h_{m'}}^{(T)}(l_1,l_2), \grave{\Sigma}_{h_m,h_{m'}}^{(T)}(l_1,l_2), \bar{\Sigma}_{h_m,h_{m'}}^{(T)}(l_1,l_2)$ *are as in Theorem 4.6.*

(b) Using the result in (a), it follows that

$$\hat{Q}_M^{(T)} \xrightarrow{\mathcal{D}} \chi^2_{\mu,2M}, \qquad (T\to\infty),$$

where $\chi^2_{\mu,2M}$ *denotes a generalized noncentral* χ^2 *random variable with noncentrality parameter* $\mu = \|\boldsymbol{\mu}\|_2^2$ *and* $2M$ *degrees of freedom.*

The proof of Theorem 4.7 can be found in Appendix S4. Theorem C.4 contains the result for (3.1).

5 Empirical results

This section reports the results of an illustrative simulation study designed to verify that the large-sample theory is useful for applications to finite samples. The test is subsequently applied to annual temperature curves data. The findings provide guidelines for a further fine-tuning of the test procedures to be investigated further in future research.

5.1 Simulation setting

To generate functional time series, the general strategy applied, for example in the papers by Aue et al. (2015) and Hörmann et al. (2015), is utilized. For this simulation study, all processes are build on a Fourier basis representation on the unit interval $[0,1]$ with basis functions $\psi_1,\ldots,\psi_{15}$. Note that the lth Fourier coefficient of a pth-order functional autoregressive, FAR(p), process $(X_t\colon t\in\mathbb{Z})$ satisfies

$$\begin{aligned}\langle X_t,\psi_l\rangle &= \sum_{l'=1}^{\infty}\sum_{t'=1}^{p}\langle X_{t-t'},\psi_l\rangle\langle A_{t'}(\psi_l),\psi_{l'}\rangle + \langle\varepsilon_t,\psi_l\rangle \\ &\approx \sum_{l'=1}^{L_{\max}}\sum_{t'=1}^{p}\langle X_{t-t'},\psi_l\rangle\langle A_{t'}(\psi_l),\psi_{l'}\rangle + \langle\varepsilon_t,\psi_l\rangle, \end{aligned} \tag{5.1}$$

the quality of the approximation depending on the choice of $L_{\max}$. The vector of the first $L_{\max}$ Fourier coefficients $\mathbf{X}_t = (\langle X_t,\psi_1\rangle,\ldots,\langle X_t,\psi_{L_{\max}}\rangle)^\top$ can thus be generated using the pth-order vector autoregressive, VAR(p), equations

$$\mathbf{X}_t = \sum_{t'=1}^{p}\mathbf{A}_{t'}\mathbf{X}_{t-t'} + \boldsymbol{\varepsilon}_t,$$

where the (l,l') element of $\mathbf{A}_{t'}$ is given by $\langle A_{t'}(\psi_l),\psi_{l'}\rangle$ and $\boldsymbol{\varepsilon}_t = (\langle\varepsilon_t,\psi_1\rangle,\ldots,\langle\varepsilon_t,\psi_{L_{\max}}\rangle)^\top$. The entries of the matrices $\mathbf{A}_{t'}$ are generated as $\mathcal{N}(0,\nu_{l,l'}^{(t')})$ random variables, with the specifications of $\nu_{l,l'}$ given below.

To ensure stationarity or the existence of a causal solution (see Bosq, 2000; van Delft & Eichler, 2016, for the stationary and locally stationary case, respectively), the norms $\kappa_{t'}$ of $\boldsymbol{A}_{t'}$ are required to satisfy certain conditions, for example, $\sum_{t'=1}^{p} |||\boldsymbol{A}_{t'}|||_\infty < 1$. The functional white noise, FWN, process is included in (5.1) setting $p = 0$.

All simulation experiments were implemented by means of the `fda` package in R and any result reported in the remainder of this section is based on 1000 simulation runs.

5.2 Specification of tuning parameters

The test statistics in (3.6) depends on the tuning parameters L, determining the dimension of the projection spaces, and M, the number of lags to be included in the procedure. In the following a criterion will be set up to choose L, while for M only two values were entertained because the selection is less critical for the performance as long as it is not chosen too large. Note that an optimal choice for L is more cumbersome to obtain in the frequency domain than in the time domain because a compromise over a number of frequencies is to be made, each of which may individually lead to a different optimal L than the one globally selected. To begin with, consider the eigenbased version of the test. The theory provided in the previous two sections indicates that including too small eigenvalues into the procedure would cause instability. On the other hand, the number of eigenvalues used should explain a reasonable proportion of the total variation in the data. These two requirements can be conflicting and a compromise is sought implementing a thresholding approach through the criterion

$$L = L_M(\zeta_1, \zeta_2, \xi) = \max\left\{\ell\colon \zeta_1 < \frac{1}{T}\sum_{j=1}^{T} \frac{\sum_{l=1}^{\ell} \lambda_l^{\omega_j}}{\sum_{l'=1}^{L_{\max}} \lambda_{l'}^{\omega_j}} < \zeta_2 \text{ and } \frac{\inf_j \lambda_\ell^{\omega_j}}{\inf_j \lambda_1^{\omega_j}} > \xi\right\} - \lfloor \log(M) \rfloor, \tag{5.2}$$

where ω_j denotes the j-th Fourier frequency and $\lfloor\cdot\rfloor$ integer part.[1] Extensive simulation studies have shown that the choices $\zeta_1 = 0.70$, $\zeta_2 = 0.90$ and $\xi = 0.15$ work well when the eigenvalues display a moderate decay. Then, the performance of the procedure appears robust against moderate deviations from these presets.

For processes with a very steep decay of the eigenvalues, however, this rule might pick L on the conservative side. To take this into account, the criterion is prefaced with a preliminary check for fast decay that leads to a relaxation of the parameter values ζ_2 and ξ if satisfied. The condition to be checked is

$$C(\xi_1, \xi_2, \xi_3) = \left\{\frac{\inf_j \lambda_2^{\omega_j}}{\inf_j \lambda_1^{\omega_j}} < \xi_1 \text{ and } \frac{\inf_j \lambda_3^{\omega_j}}{\inf_j \lambda_1^{\omega_j}} < \xi_2 \text{ and } \frac{\inf_j \lambda_4^{\omega_j}}{\inf_j \lambda_1^{\omega_j}} < \xi_3\right\}, \tag{5.3}$$

with $\xi_1 = 0.5$, $\xi_2 = 0.25$ and $\xi_3 = 0.125$. Observe that (5.3) provides a quantification of what is termed a 'fast decay' in this paper. If it is satisfied, L is chosen using the criterion (5.2) but now with $\zeta_2 = 0.995$ and $\xi = 0.01$. Condition (5.3) is in particular useful for processes for which the relative decay of the eigenvalues tends to be quadratic, a behavior more commonly observed for nonstationary functional time series. A further automation of the criterion is interesting and may be considered in future research as it is beyond the scope

[1] The rule is $\max\{1, L\}$ in the exceptional cases that (5.2) returns a nonpositive value.

of the present paper. The test statistic $\hat{Q}_M^{(T)}$ in (3.6) is then set up with the above choice of L and with $h_m = m$ and $M = 1$ and 5. A rejection is reported if the simulated test statistic value exceeds the critical level prescribed in part (b) of Theorem 4.3.

Correspondingly, for the fixed projection basis, the test statistic $\hat{Q}_M^{(T)}$ in (3.6) is set up to select those directions $\ell \in \{1, \ldots, L_{\max}\}$ for which $\langle \mathcal{F}_\omega \psi_\ell, \psi_\ell \rangle$ are largest and which explain at least 90 percent of total variation (averaged over frequencies). Furthermore, $h_m = m$ for $m = 1, \ldots, M$ with $M = 1$ and 5. A rejection is reported as above using the critical level prescribed in part (b) of Theorem C.2.

The performance of both tests is evaluated below in a variety of settings. To distinguish between the two approaches, refer to the eigenbased statistic as $\hat{Q}_{M,e}^{(T)}$ and to the fixed projection statistic as $\hat{Q}_{M,f}^{(T)}$. Estimation of the spectral density operator and its eigenelements, needed to compute the two statistics, was achieved using (3.3) with the concave smoothing kernel $K(x) = 6(0.25 - x^2)$ with compact support on $x \in [-1/2, 1/2]$ and bandwidth $b = T^{-1/5}$.

5.3 Estimating the fourth-order spectrum

The estimation of the matrix Σ_M is a necessary ingredient in the application of the proposed stationarity test. Generally, the estimation of the sample (co)variance can influence the power of tests as has been observed in a number of previous works set in similar but nonfunctional contexts. Among these contributions are Paparoditis (2009) and Jin et al. (2015), who used the spectral density of the squares, Nason (2013), who worked with locally stationary wavelets, Dwivedi & Subba Rao (2011), who focused on Gaussianity of the observations, and Jentsch & Subba Rao (2015), who employed a stationary bootstrap procedure. A different idea was put forward by Bandyopadhyay & Subba Rao (2017) and Bandyopadhyay et al. (2017). These authors utilized the notion of orthogonal samples to estimate the variance, falling back on a general estimation strategy developed in Subba Rao (2016).

To estimate the tri-spectrum for the eigenbased statistics, the estimator proposed in Brillinger & Rosenblatt (1967) was adopted to the functional context in the following way. Note first that in order to utilize the results of Theorem 4.3, the limiting covariance structure given through discretized population terms of the form

$$\sum_{j_1, j_2 = 1}^{T} \frac{(2\pi)}{T^2} \langle \mathcal{F}_{\omega_{j_1}, -\omega_{j_1+h_1}, -\omega_{j_2}} (\phi_{l_2}^{\omega_{j_2}} \otimes \phi_{l_2}^{\omega_{j_2+h_1}}), \phi_{l_1}^{\omega_{j_1}} \otimes \phi_{l_1}^{\omega_{j_1+h_1}} \rangle + O\left(\frac{1}{T^2}\right)$$

has to be estimated. To do this, consider the raw estimator

$$\begin{aligned} &I_{\alpha_1,\alpha_2,\alpha_3,\alpha_4}^{(T)}(\hat{\phi}_{l_1}^{\omega_{j_1}}, \hat{\phi}_{l_1}^{\omega_{j_2}}, \hat{\phi}_{l_2}^{\omega_{j_3}}, \hat{\phi}_{l_2}^{\omega_{j_4}}) \\ &\qquad = \frac{1}{(2\pi)^3 T} \sum_{t_1,t_2,t_3,t_4=1}^{T} \langle \mathcal{C}_{t_1,t_2,t_3,t_4} (\hat{\phi}_{l_2}^{\omega_{j_3}} \otimes \hat{\phi}_{l_2}^{\omega_{j_4}}), \hat{\phi}_{l_1}^{\omega_{j_1}} \otimes \hat{\phi}_{l_1}^{\omega_{j_2}} \rangle e^{-\mathrm{i} \sum_{j=1}^{4} \alpha_j}, \end{aligned}$$

and observe that this is the fourth-order periodogram estimator of the random variables $\langle X_t, \hat{\phi}_l^{\omega_j} \rangle$. This estimator is to be evaluated for a combination of frequencies $\alpha_1, \ldots, \alpha_4$ that lie on the principal manifold but not in any proper submanifold (see below) and is an unbiased but inconsistent estimator under the null

hypothesis. The verification of this claim is given in the Online Supplement. To construct a consistent version from the raw estimate, smooth over frequencies. The elements of $\hat{\Sigma}_m$ can then be obtained from the more standard second-order estimators above and from fourth-order estimators of the form

$$\frac{(2\pi)^3}{(b_4T)^3}\sum_{k_1,k_2,k_3,k_4} K_4\Big(\frac{\omega_{j_1}-\alpha_{k_1}}{b_4},\ldots,\frac{\omega_{j_4}-\alpha_{k_4}}{b_4}\Big)\Phi(\alpha_{k_1},\ldots,\alpha_{k_4})I^{(T)}_{\alpha_{k_1},\ldots,\alpha_{k_4}}(l_1,l_2), \tag{5.4}$$

where $K_4(x_1,\ldots,x_4)$ is a smoothing kernel with compact support on $\mathbb{R}^4$ and where $\Phi(\alpha_1,\alpha_2,\alpha_3,\alpha_4)=1$ if $\sum_{j=1}^4\alpha_j\equiv 0 \mod 2\pi$ such that $\sum_{j\in J}\alpha_j\not\equiv 0 \mod 2\pi$ where J is any non-empty subset of $\{1,2,3,4\}$ and $\Phi(\alpha_1,\alpha_2,\alpha_3,\alpha_4)=0$ otherwise. The function Φ therefore controls the selection of frequencies in the estimation, ensuring that no combinations that lie on a proper submanifold are chosen. This is important because, for $k>2$, the expectation of k-th order periodograms at such submanifolds possibly diverges. This was pointed out in Brillinger & Rosenblatt (1967, Page 163). The estimator in (5.4) is consistent under the null if the bandwidth b_4 satisfies $b_4\to 0$ but $b_4^{-3}T\to\infty$ as $T\to\infty$. As this estimator is required for various combinations of l_1 and l_2, it is worthwhile to mention that computational complexity increases rapidly. The implementation was therefore partially done with the compiler language `C++` and the `Rcpp`-package in `R`.

The simulations and application to follow below were conducted with $K_4(x_1,\ldots,x_4)=\prod_{j=1}^4 K(x_j)$, where K is as in Section 5.2 and bandwidth $b_4=T^{-1/6}$. It should be noted that the outcomes were not sensitive with respect to the bandwidth choices.

5.4 Finite sample performance under the null

Under the null hypothesis of stationarity the following data generating processes, DGPs, were studied:

(a) The Gaussian FWN variables $\varepsilon_1,\ldots,\varepsilon_T$ with coefficient variances $\mathrm{Var}(\langle\varepsilon_t,\psi_l\rangle)=\exp((l-1)/10)$;

(b) The FAR(2) variables $X_1,\ldots,X_T$ with operators specified through the respective variances $\nu^{(1)}_{l,l'}=\exp(-l-l')$ and $\nu^{(2)}_{l,l'}=1/(l+l'^{3/2})$ and Frobenius norms $\kappa_1=0.75$ and $\kappa_2=-0.4$, and innovations $\varepsilon_1,\ldots,\varepsilon_T$ as in (a);

(c) The FAR(2) variables $X_1,\ldots,X_T$ as in (b) but with Frobenius norms $\kappa_1=0.4$ and $\kappa_2=0.45$.

The sample sizes under consideration are $T=2^n$ for $n=6,\ldots,10$, so that the smallest sample size consists of 64 functions and the largest of 1024. The processes in (a)–(c) comprise a range of stationary scenarios. DGP (a) is the simplest model, specifying an independent FWN process. DGPs (b) and (c) exhibit second-order autoregressive dynamics of different persistence, with the process in (c) possessing the stronger temporal dependence.

The empirical rejection levels for the processes (a)–(c) can be found in Table 5.1. It can be seen that the empirical levels for both statistics with $M=1$ are generally well adjusted with slight deviations in a few cases. The performance of the statistics with $M=5$ is similar, although the rejection levels for the eigenbased statistics tend to be a little conservative for models B and C.

			% level			% level			% level			% level	
	T	$\hat{Q}^{(T)}_{1,e}$	5	1	$\hat{Q}^{(T)}_{5,e}$	5	1	$\hat{Q}^{(T)}_{1,f}$	5	1	$\hat{Q}^{(T)}_{5,f}$	5	1
(a)	64	1.32	4.00	0.90	8.17	3.40	0.30	1.31	3.60	0.60	8.65	3.90	1.20
	128	1.40	5.20	0.60	8.98	5.00	0.90	1.38	5.00	0.80	9.07	5.10	1.40
	256	1.53	5.40	1.30	8.89	4.50	0.80	1.31	5.60	0.80	9.23	5.30	1.00
	512	1.49	5.50	1.30	9.35	5.50	1.00	1.38	5.20	1.00	9.34	4.90	0.90
	1024	1.44	5.70	1.10	9.17	4.00	0.70	1.36	5.00	1.20	9.51	4.60	0.70
(b)	64	1.23	2.70	0.10	7.94	2.80	0.80	1.41	2.90	0.60	8.59	4.80	1.60
	128	1.40	5.20	0.80	8.80	3.60	0.70	1.41	5.40	0.90	8.91	5.10	1.20
	256	1.41	5.70	1.70	8.89	2.90	0.40	1.34	4.60	0.90	9.21	5.20	1.00
	512	1.29	6.10	1.00	9.11	5.30	0.90	1.34	5.40	1.10	9.36	4.30	0.50
	1024	1.42	4.70	2.00	8.91	3.70	0.70	1.44	4.40	0.90	9.46	4.20	0.70
(c)	64	1.21	3.90	0.80	8.25	2.00	0.30	1.36	3.50	0.60	8.62	3.90	1.20
	128	1.39	5.80	1.40	8.86	5.00	1.30	1.39	5.90	0.90	8.81	4.00	1.20
	256	1.42	4.50	1.50	8.90	3.00	0.50	1.38	4.50	0.90	9.37	4.30	1.20
	512	1.49	5.40	1.10	9.26	3.70	0.60	1.39	6.50	1.10	9.36	5.00	0.70
	1024	1.39	4.50	0.90	9.29	3.90	0.50	1.34	5.90	1.00	9.66	5.70	0.80

Table 5.1: Median of test statistic values and rejection rates of $\hat{Q}^{(T)}_{M,e}$ and $\hat{Q}^{(T)}_{M,f}$ at the 1% and 5% asymptotic level for the processes (a)–(c) for various choices of M and T. All table entries are generated from 1000 repetitions.

T	(a)		(b)		(c)		(d)		(e)		(f)	
64	8.20	14.00	8.17	14.01	8.20	14.01	7.99	14.01	5.09	12.45	7.84	14.00
	0.88	0.93	0.88	0.92	0.88	0.92	0.88	0.92	0.93	0.93	0.88	0.93
128	9.89	14.00	9.77	14.01	9.87	14.00	8.95	14.01	5.94	12.35	9.19	14.00
	0.90	0.93	0.89	0.92	0.89	0.92	0.89	0.92	0.94	0.93	0.88	0.93
256	10.52	14.00	10.06	14.01	10.32	14.00	8.88	14.02	6.46	12.48	10.04	14.00
	0.89	0.93	0.88	0.92	0.88	0.92	0.89	0.92	0.93	0.93	0.88	0.93
512	11.00	14.00	11.00	14.02	11.00	14.00	10.00	14.01	6.65	12.37	11.00	14.00
	0.89	0.94	0.89	0.92	0.89	0.93	0.89	0.92	0.94	0.93	0.89	0.93
1024	11.00	14.00	11.00	14.03	11.00	14.00	11.00	14.02	6.44	12.33	11.00	14.00
	0.88	0.94	0.89	0.92	0.88	0.93	0.89	0.92	0.94	0.94	0.88	0.94

Table 5.2: Average choices of L (top entries in each row specified by T) and aTVE (bottom entries) for $M = 1$, various sample sizes and processes (a)–(f). For each process, the left column is for the eigenbased statistics, the right column for the fixed projection statistics.

For the eigenbased statistics set up with $M = 1$, Table 5.2 displays the average choices of L according to (5.2) and the corresponding average total variation explained (aTVE) $T^{-1}\sum_{j=1}^{T}\big(\sum_{l=1}^{L}\lambda_l^{\omega_j}/\sum_{l'=1}^{L_{\max}}\lambda_{l'}^{\omega_j}\big)$. It can be seen that larger sample sizes lead to the selection of larger L, as more degrees of freedom become available. Moreover, aTVE tends to be close to 0.90 for the models under consideration. Some evidence on closeness between empirical and limit densities for the statistics $\hat{Q}^{(T)}_{5,e}$ and $\hat{Q}^{(T)}_{5,f}$ are provided in Figure 5.1. Naturally, the numbers for the fixed projection based statistics are more homogeneous.

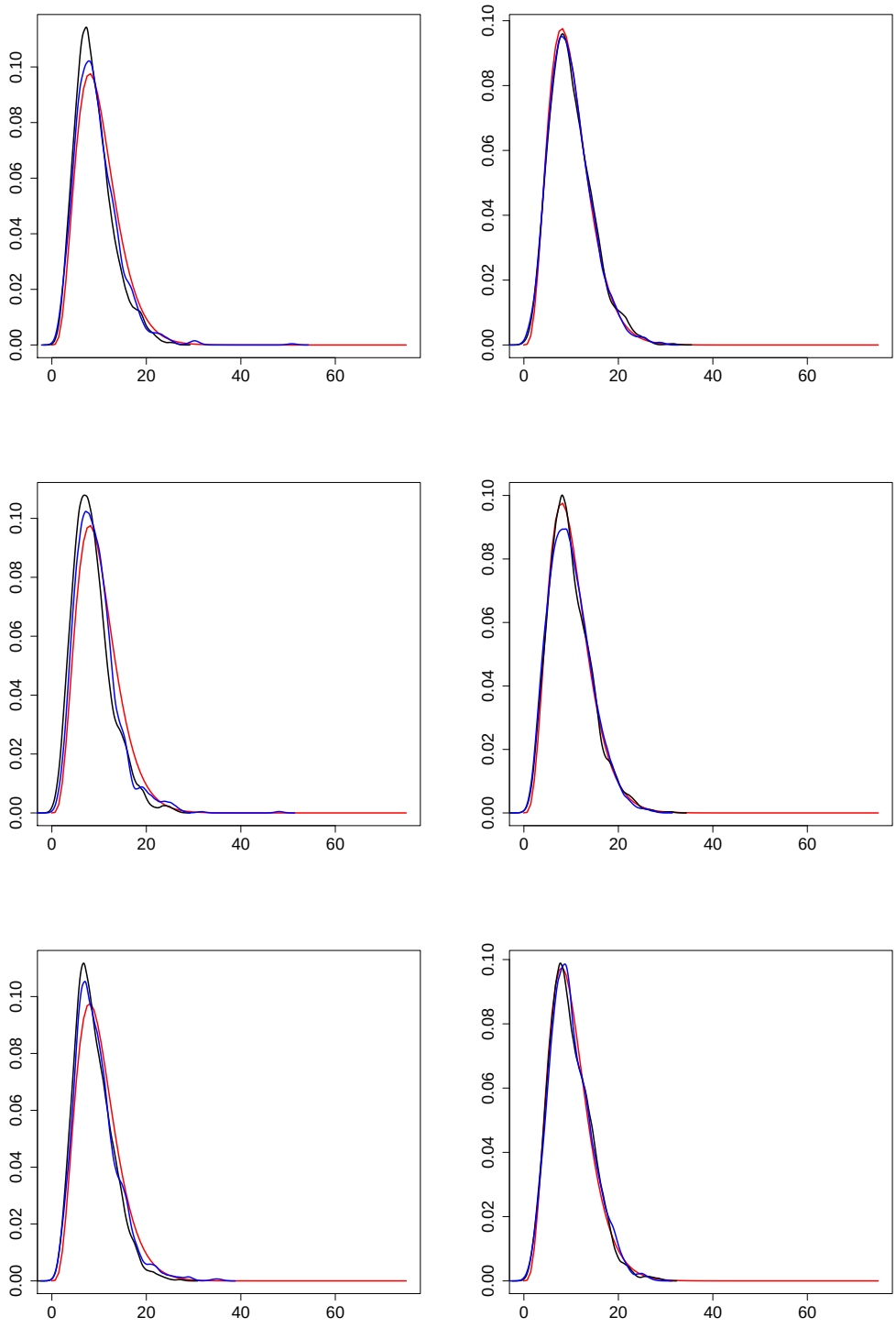

Figure 5.1: black: Empirical density of $\hat{Q}_{5,e}^{(T)}$ (black) and $\hat{Q}_{5,f}^{(T)}$ (blue) for $T = 64$ (left panel) and $T = 512$ (right panel) for DGPs (a)–(c) (top to bottom). Red: The corresponding chi-squared densities predicted under the null.

5.5 Finite sample performance under the alternative

Under the alternative, the following data generating processes are considered:

(d) The tvFAR(1) variables $X_1, \ldots, X_T$ with operator specified through the variances $\nu^{(1)}_{l,l'} = \exp(-l - l')$ and Frobenius norm $\kappa_1 = 0.8$, and innovations given by (a) with added multiplicative time-varying variance

$$\sigma^2(t) = \frac{1}{2} + \cos\left(\frac{2\pi t}{1024}\right) + 0.3\sin\left(\frac{2\pi t}{1024}\right);$$

(e) The tvFAR(2) variables $X_1, \ldots, X_T$ with both operators as in (d) but with time-varying Frobenius norm

$$\kappa_{1,t} = 1.8\cos\left(1.5 - \cos\left(\frac{4\pi t}{T}\right)\right),$$

constant Frobenius norm $\kappa_2 = -0.81$, and innovations as in (a);

(f) The structural break FAR(2) variables $X_1, \ldots, X_T$ given in the following way.

- For $t \leqslant 3T/8$, the operators are as in (b) but with Frobenius norms $\kappa_1 = 0.7$ and $\kappa_2 = 0.2$, and innovations as in (a);
- For $t > 3T/8$, the operators are as in (b) but with Frobenius norms $\kappa_1 = 0$ and $\kappa_2 = -0.2$, and innovations as in (a) but with variances $\mathrm{Var}(\langle \varepsilon_t, \psi_l \rangle) = 2\exp((l-1)/10)$.

All other aspects of the simulations are as in Section 5.4. The processes studied under the alternative provide intuition for the behavior of the proposed tests under different deviations from the null hypothesis. DGP (d) is time-varying only through the innovation structure, in the form of a slowly varying variance component. The first-order autoregressive structure is independent of time. DGP (e) is a time-varying second-order FAR process for which the first autoregressive operator varies with time. The final DGP in (f) models a structural break, a different type of alternative. Here, the process is not locally stationary as prescribed under the alternative in this paper, but piecewise stationary with the two pieces being specified as two distinct FAR(2) processes.

The empirical power of the various test statistics for the processes in (d)–(f) are in Table 5.3. Power results are roughly similar across the selected values of M for both statistics. For DGP (d), power is low for the small sample sizes $T = 64$ for the eigenbased statistics and even lower for the fixed projection statistics. It is at 100% for all T larger or equal to 256. The low power is explained by the form of the time-varying variance which takes 1024 observations to complete a full cycle of the sine and cosine components. In the situation of the smallest sample size, this slowly varying variance appears more stationary, explaining why rejections of the null are less common. Generally, the eigenbased statistics performs better than its fixed projection counterpart for this process.

DGP (e) shows lower power throughout. The reason for this is that the form of local stationarity under consideration here is more difficult to separate from stationary behavior expected under the null hypothesis.

			% level			% level			% level			% level	
	T	$\hat{Q}^{(T)}_{1,e}$	5	1	$\hat{Q}^{(T)}_{5,e}$	5	1	$\hat{Q}^{(T)}_{1,f}$	5	1	$\hat{Q}^{(T)}_{5,f}$	5	1
(d)	64	6.39	54.30	30.00	14.33	29.40	12.30	2.68	17.20	4.00	11.08	13.20	4.80
	128	66.97	99.90	99.80	86.94	99.80	99.40	18.94	98.80	93.30	35.79	98.10	90.20
	256	611.13	100.00	100.00	747.09	100.00	100.00	130.23	100.00	100.00	206.05	100.00	100.00
	512	1462.97	100.00	100.00	1506.20	100.00	100.00	269.31	100.00	100.00	347.91	100.00	100.00
	1024	1550.04	100.00	100.00	1627.24	100.00	100.00	195.85	100.00	100.00	251.64	100.00	100.00
(e)	64	5.29	48.90	46.30	15.78	47.20	45.40	3.93	44.20	39.40	14.74	44.40	40.50
	128	6.05	50.20	46.30	20.74	52.00	48.40	4.20	45.30	42.10	15.47	46.50	43.20
	256	5.67	49.50	45.80	19.71	52.80	47.90	4.04	43.70	40.40	15.67	46.10	43.10
	512	6.35	50.65	47.55	25.80	56.96	51.45	3.91	44.40	41.80	17.37	48.50	44.30
	1024	9.80	52.06	50.35	42.40	65.83	59.50	4.64	46.80	43.80	22.10	54.80	49.00
(f)	64	23.84	97.00	93.10	30.79	89.70	79.60	6.19	52.50	23.50	15.81	33.60	14.60
	128	30.37	97.00	93.20	40.66	90.70	83.10	11.97	90.30	69.90	23.45	80.70	51.60
	256	80.27	100.00	100.00	87.25	100.00	100.00	23.10	99.90	98.60	38.49	99.80	98.10
	512	288.79	100.00	100.00	298.20	100.00	100.00	46.99	100.00	100.00	68.60	100.00	100.00
	1024	781.13	100.00	100.00	824.65	100.00	100.00	94.85	100.00	100.00	127.87	100.00	100.00

Table 5.3: Median of test statistic values and rejection rates of $\hat{Q}^{(T)}_{M,e}$ and $\hat{Q}^{(T)}_{M,f}$ at the 1% and 5% asymptotic level for the processes (d)–(f) for various choices of M and T. All table entries are generated from 1000 repetitions.

This is corroborated in Figure 5.2, where the center of the empirical version of the non-central generalized chi-squared limit under the alternative is more closely aligned with that of the standard chi-squared limit expected under the null hypothesis than for the cases (d) and (f). Both test statistics behave similarly across the board.

The results for DGP (f) indicate that the proposed statistics have power against structural break alternatives. This statement is supported by Figure 5.2. Here the eigenbased version picks up power quicker than the fixed projection counterpart as the sample size T increases. All statistics work well once T reaches 256.

Going back to Table 5.2, it further seen that L is under H_A generally chosen in a similar way as under H_0, with one notable exception being the process under (e), where L tends to be significantly smaller. Under the alternative, aTVE tends to be marginally larger for models (d)–(f) than for the models (a)–(c) considered under the null hypothesis.

5.6 Finite sample performance under non-Gaussian observations

In this section, the behavior of the eigenbased test under non-Gaussianity is further investigated through the following processes:

(g) The FAR(2) variables $X_1, \ldots, X_T$ as in (b) but with both independent t_{19}-distributed FWN and independent $\beta(6,6)$-distributed FWN;

(h) The tvFAR(1) variables $X_1, \ldots, X_T$ as in (d) but with independent t_{10}-distributed FWN and independent $\beta(6,6)$-distributed FWN.

For direct comparison, both t_{19}- and $\beta(6,6)$-distributions were standardized to conform to zero mean and unit variance as the standard normal. The sample sizes considered were $T = 64$ and $T = 128$, since these are most

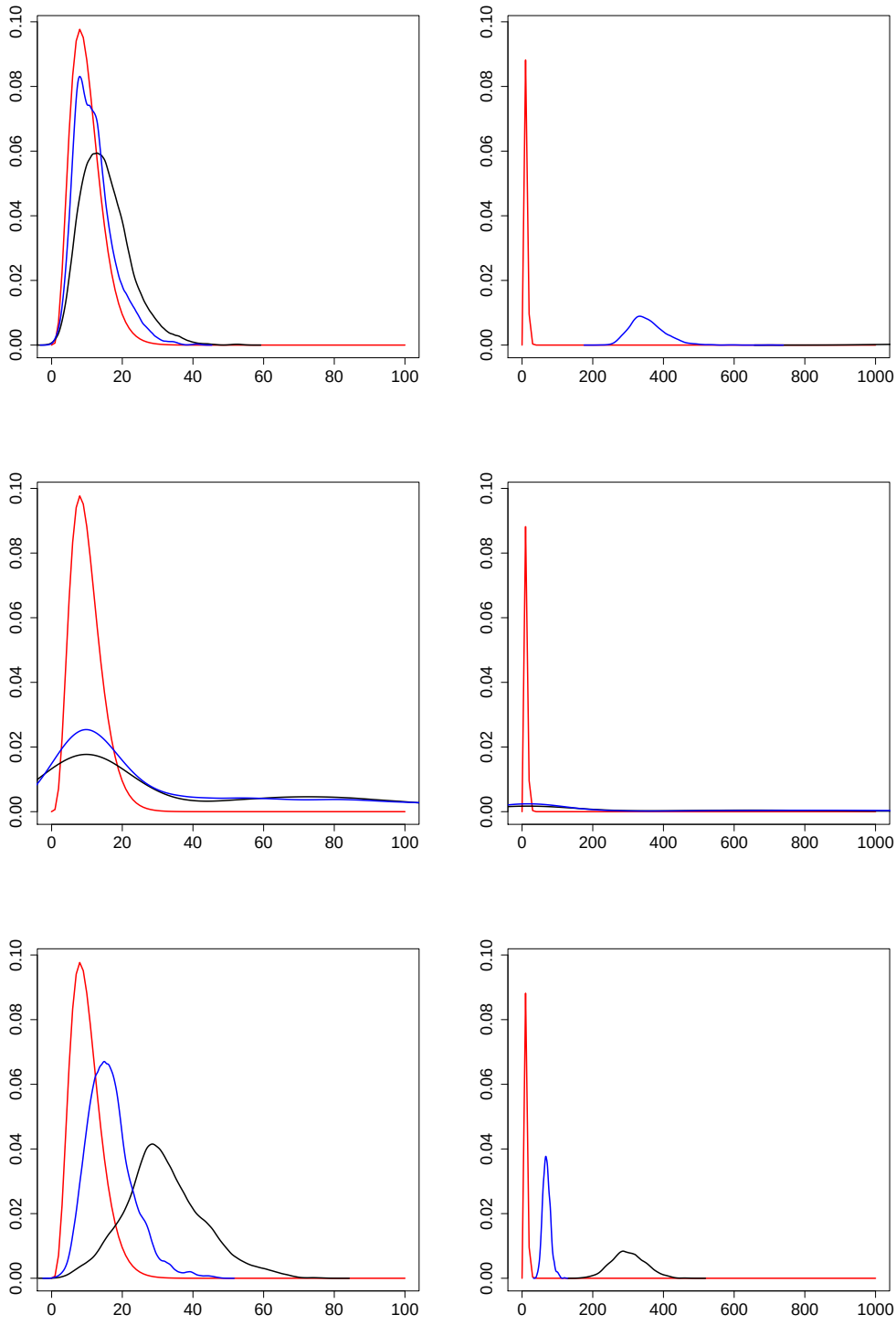

Figure 5.2: Empirical density of $\hat{Q}_{5,e}^{(T)}$ (black) and $\hat{Q}_{5,f}^{(T)}$ (blue) for $T = 64$ (left panel) and $T = 512$ (right panel) for DGPs (d)–(f) (top to bottom). Note that 2.5% outliers have been removed for Model (e). Red: The corresponding chi-squared densities predicted under the null.

similar to the ones observed for the temperature data discussed in Section 5.7, and $M = 1, 3$ and 5. All other aspects are as detailed in Section 5.4. These additional simulations were designed to shed further light on the effect of estimating the fourth-order spectrum in situations deviating from the standard Gaussian setting. Note in particular that the t_{19}-distribution serves as an example for leptokurtosis (the excess kurtosis is 0.4) and the $\beta(6, 6)$ distribution for platykurtosis (the excess kurtosis is -0.4). Process (g) showcases the behavior under the null, while process (h) highlights the performance under the alternative. The corresponding results are given in Table 5.4 and can be readily compared with corresponding outcomes for the Gaussian processes (b) and (d) in Tables 5.1–.5.3.

						% level			% level	
		T	L	aTVE	$\hat{Q}^{(T)}_{1,e}$	5	1	$\hat{Q}^{(T)}_{5,e}$	5	1
(g)	t	64	8.13	0.88	1.33	3.30	0.30	8.24	2.40	0.70
		128	9.73	0.89	1.22	4.90	1.40	8.80	3.90	1.00
	β	64	8.18	0.88	1.16	2.90	0.60	7.61	2.40	0.40
		128	9.80	0.89	1.34	4.60	0.90	8.97	4.00	0.60
(h)	t	64	4.32	0.98	7.44	59.70	40.10	16.55	42.80	26.80
		128	4.57	0.98	85.85	100.00	100.00	98.84	99.80	99.50
	β	64	4.42	0.98	7.07	58.60	37.80	16.14	39.30	22.50
		128	4.65	0.98	88.78	100.00	100.00	98.57	99.90	99.80

Table 5.4: Median of test statistic values and rejection rates of $\hat{Q}^{(T)}_{M,e}$ at the 1% and 5% asymptotic level for the processes (g) and (h), where t and β indicate t_{19}- and $\beta(6, 6)$-distributed innovations, respectively. The values for L and aTVE correspond to $M = 1$. All table entries are generated from 1000 repetitions.

It can be seen from the results in Table 5.4 that the proposed procedures perform roughly as expected. First, under the null hypothesis for process (g) and for $T = 64$, levels tend to be conservative for all M. For $T = 128$, levels are well adjusted for $M = 1$ and still decent for $M = 5$. Both t_{19} and $\beta(6, 6)$ variables tend to be produce levels that differ only little from their Gaussian counterparts in Table 5.1. Second, under the alternative for process (h), powers align roughly as for the Gaussian case in Table 5.3. Comparing to Table 5.2, it can be seen that for process (h), the chosen L are lower than in the Gaussian case. Overall, the simulation results reveal that the estimation of the fourth-order spectrum does not lead to a marked decay in performance.

5.7 Application to annual temperature curves

To give an instructive data example, the proposed method was applied to annual temperature curves recorded at several measuring stations across Australia over the last century and a half. The exact locations and lengths of the functional time series are reported in Table 5.5, and the annual temperature profiles recorded at the Gayndah station are displayed for illustration in the left panel of Figure 5.3. To test whether these annual temperature profiles constitute stationary functional time series or not, the proposed testing method was utilized, using specifications similar to those in the simulation study. Focusing here on the eigenbased version, the test

statistic $\hat{Q}^{(T)}_{M,e}$ in (3.6) was applied with L chosen according to (5.2) and $h_m = m$ for $m = 1, \ldots, M$, where $M = 1$ and 5.

The testing results are summarized in Table 5.5. It can be seen that stationarity is rejected in favor of the alternative at the 1% significance level at all measuring stations for $\hat{Q}^{(T)}_{5,e}$ and for all but two measuring stations for $\hat{Q}^{(T)}_{1,e}$, the exceptions being Melbourne and Sydney, where the test is rejected at the 5% level with a p-value of 0.014, and Sydney, where the p-value is approximately 0.072. The clearest rejection of the null hypothesis was found for the measuring station at Gunnedah Pool for both choices of M. The values of L chosen with (5.2) range from 4 to 8 and are similar to the values observed in the simulation study. The right-hand side of Figure 5.3 shows the decay of eigenvalues associated with the different measuring stations. Moreover, for $M = 1$ ($M = 5$) aTVE ranges from 0.762 (0.705) in Hobart to 0.882 (0.847) in Melbourne.

Station	T	L	aTVE	$\hat{Q}^{(T)}_{1,e}$	L	aTVE	$\hat{Q}^{(T)}_{5,e}$
Boulia	120	7	0.864	11.623	6	0.823	65.463
Robe	129	8	0.876	13.262	7	0.837	39.339
Cape Otway	149	5	0.844	35.382	4	0.798	75.876
Gayndah	117	6	0.840	14.638	5	0.792	55.967
Gunnedah	133	4	0.825	41.494	3	0.760	76.805
Hobart	121	5	0.762	20.405	4	0.705	38.288
Melbourne	158	8	0.882	5.274	7	0.847	54.879
Sydney	154	8	0.873	8.591	7	0.837	28.359

Table 5.5: Summary of results for eight Australian measuring stations. The column labeled T reports the sample size, L gives the value chosen by (5.2), aTVE is the average total variation explained as used in this criterion.

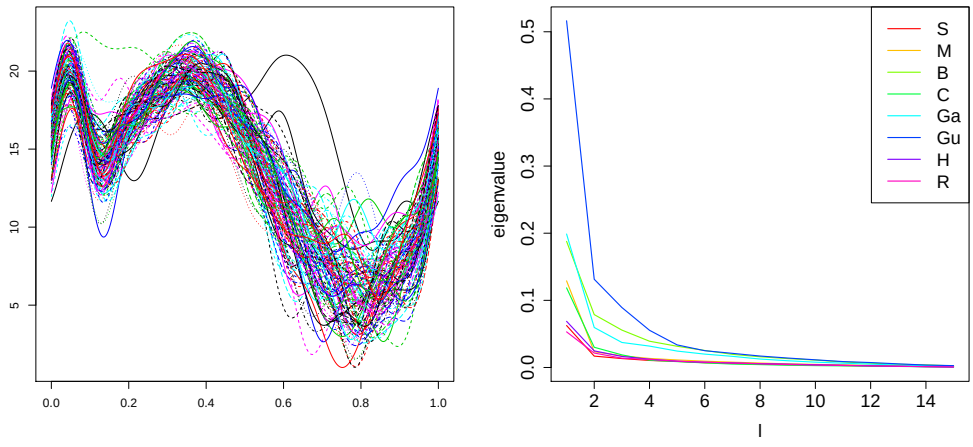

Figure 5.3: Plot of annual temperature curves at Gayndah station (left) and of eigenvalue decay across different measuring stations (right).

6 Conclusions and future work

In this paper methodology for testing the stationarity of a functional time series is put forward. The tests are based on frequency domain analysis and exploit that fDFTs at different canonical frequencies are uncorrelated if and only if the underlying functional time series are stationary. The limit distribution of the quadratic form-type test statistics has been determined under the null hypothesis as well as under the alternative of local stationarity. Finite sample properties were highlighted in simulation experiments with various data generating processes and an application to annual temperature profiles, where deviations from stationarity were detected.

The empirical results show promise for further applications to real data, but future research has to be devoted to a further fine-tuning of the proposed method; for example, an automated selection of frequencies h_m outside of the standard choice $h_m = m$ for all $m = 1, \ldots, M$. This can be approached through a more refined analysis of the size of the various $\hat{\boldsymbol{\beta}}_{h_m}^{(T)}$ in (3.5) whose real and imaginary part make up the vector $\hat{\mathbf{b}}_M^{(T)}$ in the test statistics $\hat{Q}_M^{(T)}$.

Another promising route of research is as follows. Figure 6.1 provides contour plots of squared modulus of $\hat{\gamma}_1^{(T)}$ for model (b) under the null and for models (d)–(f) under the alternative. The contours are obtained from averaging (over the simulation runs) the aggregated contributions $\hat{\gamma}_1^{(T)}$ and projecting these onto a Fourier basis of dimension 15. It can be seen that the magnitude of the contours provides another indicator for how easy or hard it may be to reject the null hypothesis. The top row in the figure is for the stationary DGP (b). For any of the sample sizes considered, the magnitude across $[0, 1]^2$ remains small, as expected under the null. The behavior under the alternative is markedly different, but the specifics depend on the type of alternative. For the time-varying noise process (d), the contribution of non-stationarity is at the diagonal, with the magnitude along this ridge depending on the sample size. For DGP (e), the form of non-stationarity creates very different contours. The structural break process (f) induces non-stationarity in the contours in a similar way as DGP (d), with most concentration occurring at the diagonal for all sample sizes. Any future refinement of the tests will have to take these features into account.

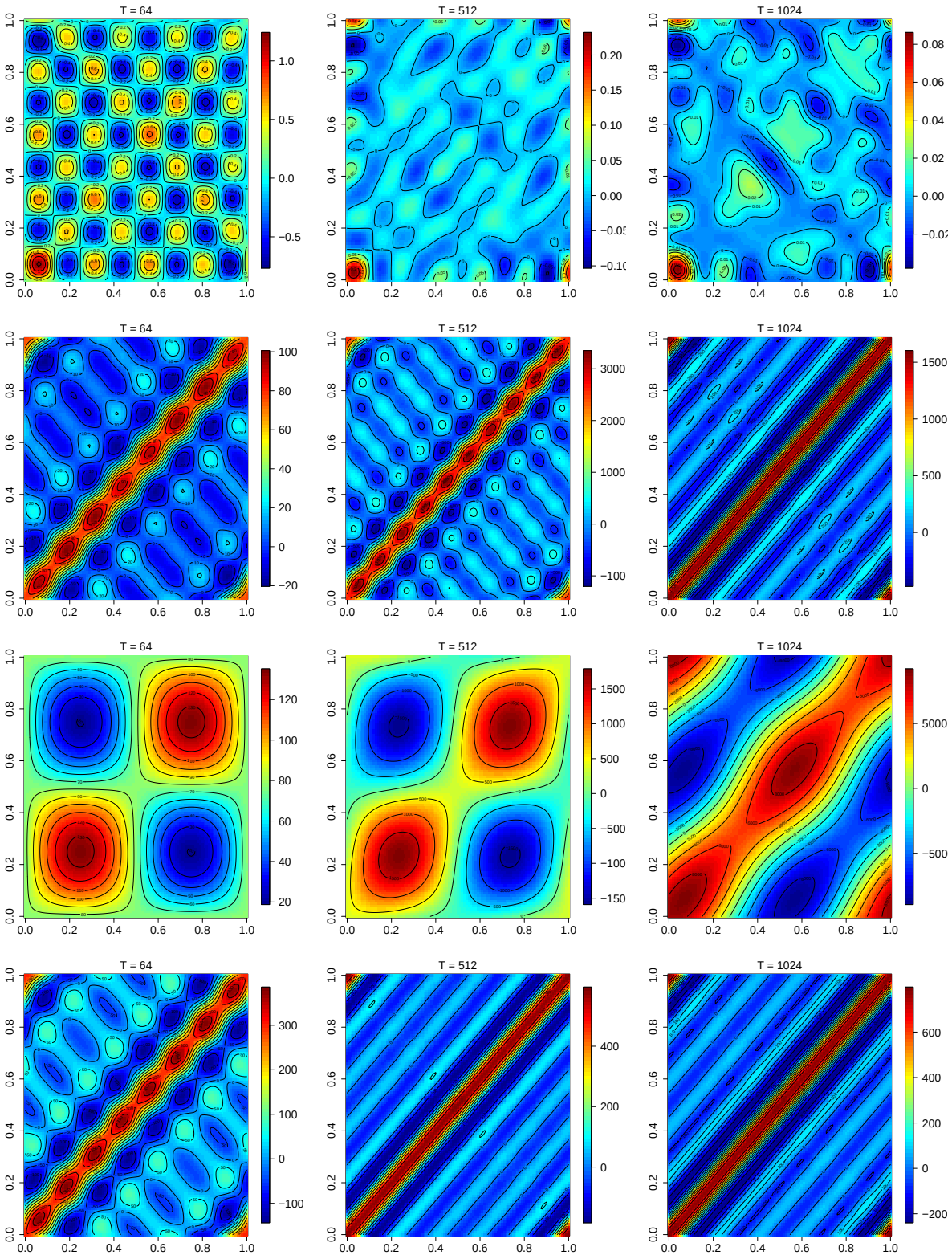

Figure 6.1: Contour plots of $\hat{\gamma}_1^{(T)}$ for $L^\star$ for various sample sizes T and DGPs (b) and (d)–(f) (top to bottom).

A A functional Cramér representation

Proof of Proposition 2.1. Let $(X_t\colon t \in \mathbb{Z})$ be a zero-mean, $H_{\mathbb{R}}$-valued stochastic process that is weakly stationary. It has been shown (Panaretos & Tavakoli, 2013) that for processes with $\sum_{h\in\mathbb{Z}} |||\mathcal{C}_h|||_1 < \infty$, there exists an isomorphic mapping between the subspaces $\overline{\text{sp}}(X_t\colon t \in \mathbb{Z})$ of $\mathbb{H}$ and $\overline{\text{sp}}(e^{\mathrm{i}t\cdot}\colon t \in \mathbb{Z})$ of $L^2([-\pi,\pi], \mathscr{B}, |||\mathcal{F}_\omega|||_1 d\omega)$. As a consequence, X admits the representation

$$X_t = \int_{-\pi}^{\pi} e^{\mathrm{i}t\lambda} dZ_\lambda \qquad \text{a.s. a.e.,} \tag{A.1}$$

where $(Z_\lambda\colon \lambda \in (-\pi,\pi])$ is a right-continuous, functional orthogonal-increment process. Conversely, for any process with the above representation, orthogonality of the increments and Theorem S2.2 of van Delft & Eichler (2016) imply

$$\begin{aligned}\text{Cov}(X_t, X_s) &= \mathbb{E}\Big[\int_{-\pi}^{\pi} e^{\mathrm{i}t\lambda_1} dZ_{\lambda_1} \otimes \int_{-\pi}^{\pi} e^{\mathrm{i}s\lambda_2} dZ_{\lambda_2}\Big] \\ &= \int_{-\pi}^{\pi} e^{\mathrm{i}(t-s)\lambda} \mathcal{F}_\lambda d\lambda = \mathcal{C}_{t-s},\end{aligned}$$

showing that a process that admits representation (A.1) must be weakly stationary. □

B Properties of functional cumulants

For random elements $X_1, \ldots, X_k$ in a Hilbert space H, the *moment tensor of order k* can be defined as

$$\mathbb{E}\big[X_1 \otimes \cdots \otimes X_k\big] = \sum_{l_1,\ldots l_k \in \mathbb{N}} \mathbb{E}\Big[\prod_{t=1}^{k} \langle X_t, \psi_{l_t}\rangle\Big] (\psi_{l_1} \otimes \cdots \otimes \psi_{l_k}),$$

where the elementary tensors $(\psi_{l_1} \otimes \cdots \otimes \psi_{l_k}\colon l_1, \ldots, l_k \in \mathbb{N})$ form an orthonormal basis in the tensor product space $\bigotimes_{j=1}^{k} H$. The latter follows since $(\psi_l\colon l \in \mathbb{N})$ is an orthornormal basis of the separable Hilbert space H. Similarly, define the *k-th order cumulant tensor* by

$$\text{cum}\big(X_1, \ldots, X_k\big) = \sum_{l_1,\ldots l_k \in \mathbb{N}} \text{cum}\Big(\prod_{t=1}^{k} \langle X_t, \psi_{l_t}\rangle\Big) (\psi_{l_1} \otimes \cdots \otimes \psi_{l_k}), \tag{B.1}$$

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

$$\text{cum}\big(\langle X_1, \psi_{l_1}\rangle, \ldots, \langle X_k, \psi_{l_k}\rangle\big) = \sum_{\nu=(\nu_1,\ldots,\nu_p)} (-1)^{p-1}\,(p-1)! \prod_{r=1}^{p} \mathbb{E}\Big[\prod_{t\in\nu_r} \langle X_t, \psi_{l_t}\rangle\Big],$$

the summation extending over all unordered partitions ν of $\{1, \ldots, k\}$. The following is a generalization of the product theorem for cumulants (Brillinger, 1981, Theorem 2.3.2).

Theorem B.1. *Consider the tensor $X_t = \bigotimes_{j=1}^{J_t} X_{tj}$ for random elements X_{tj} in H with $j = 1, \ldots, J_t$ and $t = 1, \ldots, k$. Let $\nu = \{\nu_1, \ldots \nu_p\}$ be a partition of $\{1, \ldots, k\}$. The joint cumulant tensor* $\text{cum}(X_1, \ldots, X_k)$ *is given by*

$$\text{cum}(X_1, \ldots, X_k) = \sum_{r_{11},\ldots,r_{kJ_t}} \sum_{\nu=(\nu_1,\ldots,\nu_p)} \prod_{n=1}^{p} \text{cum}\big(\langle X_{tj}, \psi_{r_{tj}}\rangle | (t,j) \in \nu_n\big)\, \psi_{r_{11}} \otimes \cdots \otimes \psi_{r_{kJ_t}},$$

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

$$\begin{matrix} (1,1) & \cdots & (1,J_1) \\ \vdots & \ddots & \vdots \\ (k,1) & \cdots & (k,J_t). \end{matrix}$$

Formally, abbreviate this by

$$\operatorname{cum}(X_1, \dots, X_k) = \sum_{\nu=(\nu_1,\dots,\nu_p)} S_\nu \Big(\otimes_{n=1}^p \operatorname{cum}\big(X_{tj} | (t,j) \in \nu_n\big)\Big),$$

where S_ν is the permutation that maps the components of the tensor back into the original order, that is, $S_\nu(\otimes_{r=1}^p \otimes_{(t,j)\in\nu_r} X_{tj}) = X_{11} \otimes \cdots \otimes X_{kJ_t}$.

Next, results for cumulants of the fDFT are stated under both stationarity and local stationarity regimes.

Lemma B.1 (**Cumulants of the fDFT under stationarity**). *Let $(X_t \colon t \in \mathbb{Z})$ be a k-th order stationary sequence taking values in $H_\mathbb{R}$ that satisfies Assumption 4.1. The cumulant operator of the fDFT then satisfies*

$$\operatorname{cum}\big(D^{(T)}_{\omega_{j_1}}, \dots, D^{(T)}_{\omega_{j_k}}\big) = \frac{(2\pi)^{k/2-1}}{T^{k/2}} \Delta_T^{(\sum_{l=1}^k \omega_{j_l})} \mathcal{F}_{\omega_{j_1},\dots,\omega_{j_{k-1}}} + R_{T,k}, \tag{B.2}$$

where the function $\Delta_T^{(\omega)} = T$ for $\omega \equiv 0 \mod 2\pi$, $\Delta_T^{(\omega_k)} = 0$ for $k \not\equiv 0 \mod T$ and the remainder satisfies $\|R_{T,k}\|_2 = O(T^{-k/2})$.

The proof of this lemma can be found in Panaretos & Tavakoli (2013). To give the analog of Lemma B.1 in the locally stationary case, a number of auxiliary statements must be derived first. Because of space constraints, these statements and their proofs are relegated to the Online supplement. Lemma S1.1 and Lemma S1.2 allow to derive the that the cumulant tensors of the local fDFT can be expressed in terms of the time-varying spectral operator. At the Fourier frequencies, the time-varying spectral operator can in turn be shown to possess a well-defined Fourier transform. The properties of the resulting Fourier coefficients make apparent that the dependence structure of the local fDFT behaves in a very specific manner that is based on the distance of the frequencies. The coefficients additionally provide an upper bound on the norm of the cumulant operator. This is summarized in the next lemma, which is the locally stationary version of Lemma B.1.

Lemma B.2 (**Cumulants of the fDFT under local stationarity**). *Let $(X_{t,T} \colon t \leqslant T, T \in \mathbb{N})$ be a k-th order locally stationary process in H satisfying Assumption 4.3. The cumulant operator of the local fDFT satisfies*

$$\begin{aligned} \operatorname{cum}\big(D^{(T)}_{\omega_{j_1}}, \dots, D^{(T)}_{\omega_{j_k}}\big) &= \frac{(2\pi)^{k/2-1}}{T^{k/2}} \sum_{t=0}^{T-1} \mathcal{F}_{t/T;\omega_{j_1},\dots,\omega_{j_{k-1}}} e^{-\mathrm{i}\sum_{l=1}^k t\omega_{j_l}} + R_{k,T} \\ &= \frac{(2\pi)^{k/2-1}}{T^{k/2-1}} \tilde{\mathcal{F}}_{j_1+\dots+j_k;\omega_{j_1},\dots,\omega_{j_{k-1}}} + R_{k,T}, \end{aligned} \tag{B.3}$$

where $\|R_{k,T}\|_2 = O(T^{-k/2})$ and the operator

$$\tilde{\mathcal{F}}_{s;\omega_{j_1},\dots,\omega_{j_{k-1}}} = \int_0^1 \mathcal{F}_{u;\omega_{j_1},\dots,\omega_{j_{k-1}}} e^{-\mathrm{i}2\pi s u} du \tag{B.4}$$

denotes the s-th Fourier coefficient of $\mathcal{F}_{u;\omega_{j_1},\dots,\omega_{j_{k-1}}}$ and is Hilbert–Schmidt. for some constant $C > 0$.

In case the process does not depend on u, we have $\tilde{\mathcal{F}}_{s;\omega_{j_1},\ldots,\omega_{j_{k-1}}} = O_H$ for $s \neq 0$. That is, the operator $\tilde{\mathcal{F}}_{s;\omega_{j_1},\ldots,\omega_{j_{2k-1}}}$ maps any $\psi \in L^2([0,1]^k, \mathbb{C})$ to the origin for $s \neq 0$. The corollary below is a direct consequence of Lemma B.2.

Corollary B.1. *If Assumption 4.3 holds with $\ell = 2$, then*

(i) $\|\!|\mathrm{cum}(D^{(T)}_{\omega_{j_1}}, \ldots, D^{(T)}_{\omega_{j_k}})|\!\|_2 \leqslant \dfrac{C}{T^{k/2-1}|j_1 + \cdots + j_k|^2} + O\Big(\dfrac{1}{T^{k/2}}\Big)$;

(ii) $\sup\limits_{\omega} \sum\limits_{s\in\mathbb{Z}} |\!|\!|\tilde{\mathcal{F}}_{s;\omega}|\!|\!|_2 \leqslant \infty$.

C Companion results for the test defined through (3.1)

Theorem C.1. *Let Assumption 4.1 be satisfied with $k = \{2, 4\}$. Then,*

$$
\begin{aligned}
&T\,\mathrm{Cov}\Big(\Re\gamma^{(T)}_{h_1}(l_1, l_1'), \Re\gamma^{(T)}_{h_2}(l_2, l_2')\Big) \\
&= T\,\mathrm{Cov}\Big(\Im\gamma^{(T)}_{h_1}(l_1, l_1'), \Im\gamma^{(T)}_{h_2}(l_2, l_2')\Big) \\
&= \begin{cases}
\dfrac{1}{4\pi}\Bigg(\displaystyle\int \frac{\mathcal{F}^{(l_1,l_2)}_{\omega}\mathcal{F}^{(l_1',l_2')}_{-\omega-\omega_h}}{\sqrt{\mathcal{F}^{(l_1,l_1)}_{\omega}\mathcal{F}^{(l_2,l_2)}_{-\omega}\mathcal{F}^{(l_1',l_1')}_{-\omega-\omega_h}\mathcal{F}^{(l_2',l_2')}_{\omega+\omega_h}}} d\omega + \int \frac{\mathcal{F}^{(l_1,l_2')}_{\omega}\mathcal{F}^{(l_1',l_2)}_{-\omega-\omega_h}}{\sqrt{\mathcal{F}^{(l_1,l_1)}_{\omega}\mathcal{F}^{(l_2',l_2')}_{-\omega}\mathcal{F}^{(l_1',l_1')}_{-\omega-\omega_h}\mathcal{F}^{(l_2,l_2)}_{\omega+\omega_h}}} d\omega \\
\qquad + \displaystyle\int\int \frac{\mathcal{F}^{(l_1\,l_1',l_2\,l_2')}_{\omega,-\omega-\omega_h,-\omega'}}{\sqrt{\mathcal{F}^{(l_1,l_1)}_{\omega}\mathcal{F}^{(l_1',l_1')}_{-\omega-\omega_h}\mathcal{F}^{(l_2,l_2)}_{-\omega'}\mathcal{F}^{(l_2',l_2')}_{\omega'+\omega_h}}} d\omega d\omega'\Bigg), & \text{if } h_1 = h_2 = h. \\
O(\frac{1}{T}), & \text{if } h_1 \neq h_2.
\end{cases}
\end{aligned}
$$

where $\mathcal{F}^{(l,l')}_{\omega} = \langle \mathcal{F}_{\omega}(\psi_{l'}), \psi_l\rangle$, $\mathcal{F}^{(l_1l_1',l_2l_2')}_{\omega_1,\omega_2,\omega_3} = \langle \mathcal{F}_{\omega_1,\omega_2,\omega_3}(\psi_{l_2l_2'}), \psi_{l_1l_1'}\rangle$ and $\psi_{ll'} = \psi_l \otimes \psi_{l'}$. Furthermore,

$$
T\,\mathrm{Cov}(\Re\gamma^{(T)}_{h_1}(l_1, l_1'), \Im\gamma^{(T)}_{h_2}(l_2, l_2')) = O\Big(\frac{1}{T}\Big)
$$

uniformly in $h_1, h_2 \in \mathbb{Z}$.

Theorem C.2. *If the condition $\inf_\omega \lambda^{\omega}_L > 0$ is replaced with $\inf_\omega \langle \mathcal{F}(\psi), \psi\rangle > 0$ for all $\psi \in H_{\mathbb{C}}$, then parts (a) and (b) of Theorem 4.3 are retained using*

$$
\sigma^2_{0,m} = \lim_{T\to\infty} \sum_{l_1,l_1',l_2,l_2'=1}^{L} T\mathrm{Cov}\big(\Re\gamma^{(T)}_{h_m}(l_1, l_1'), \Re\gamma^{(T)}_{h_m}(l_2, l_2')\big), \qquad m = 1, \ldots, M,
$$

where the explicit form of the limit is given by Theorem C.1.

Theorem C.3. *Let Assumption 4.3 be satisfied with $k = \{2, 4\}$. Then, for $h = 1, \ldots, T-1$,*

$$
\mathbb{E}\big[\gamma^{(T)}_h(l, l')\big] = \frac{1}{2\pi}\int_0^{2\pi}\int_0^1 \frac{\mathcal{F}^{(l,l')}_{u,\omega} e^{-\mathrm{i}2\pi uh}}{(G^{(l,l)}_{\omega} G^{(l',l')}_{\omega+\omega_h})^{1/2}} du d\omega + O\Big(\frac{1}{T}\Big) = O\Big(\frac{1}{h^2}\Big) + O\Big(\frac{1}{T}\Big). \tag{C.1}
$$

The covariance structure satisfies

1. $T\operatorname{Cov}(\Re\gamma_{h_1}^{(T)}(l_1,l_2),\Re\gamma_{h_2}^{(T)}(l_3,l_4)) =$
$$\frac{1}{4}\big[\Sigma_{h_1,h_2}^{(T)}(\boldsymbol{l}_4)+\acute{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4)+\grave{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4)+\bar{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4)\big]+R_T,$$

2. $T\operatorname{Cov}(\Re\gamma_{h_1}^{(T)}(l_1,l_2),\Im\gamma_{h_2}^{(T)}(l_3,l_4)) =$
$$\frac{1}{4\mathrm{i}}\big[\Sigma_{h_1,h_2}^{(T)}(\boldsymbol{l}_4)-\acute{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4)+\grave{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4)-\bar{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4)\big]+R_T,$$

3. $T\operatorname{Cov}(\Im\gamma_{h_1}^{(T)}(l_1,l_2),\Im\gamma_{h_2}^{(T)}(l_3,l_4)) =$
$$\frac{1}{4}\big[\Sigma_{h_1,h_2}^{(T)}(\boldsymbol{l}_4)-\acute{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4)-\grave{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4)+\bar{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4)\big]+R_T,$$

where $\Sigma_{h_1,h_2}^{(T)}(\boldsymbol{l}_4),\acute{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4),\grave{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4),\bar{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4)$ *are defined in equations (S.5.1)–(S.5.4) of the Online Supplement. and* $\|R_T\|_2=O(T^{-1})$.

Theorem C.4. *If the condition* $\inf_\omega\lambda_L^\omega>0$ *is replaced with* $\inf_\omega\langle\mathcal{F}(\psi),\psi\rangle>0$ *for all* $\psi\in H_{\mathbb{C}}$*, then parts (a) and (b) of Theorem C.4 are retained using* Σ_A *defined through*

$$\Sigma_A^{(11)}(m,m')=\lim_{T\to\infty}\sum_{l_1,l_2,l_3,l_4=1}^{L}\frac{1}{4}\big[\Sigma_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4)+\acute{\Sigma}_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4)+\grave{\Sigma}_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4)+\bar{\Sigma}_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4)\big],$$
$$\Sigma_A^{(12)}(m,m')=\lim_{T\to\infty}\sum_{l_1,l_2,l_3,l_4=1}^{L}\frac{1}{4\mathrm{i}}\big[\Sigma_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4)-\acute{\Sigma}_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4)+\grave{\Sigma}_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4)-\bar{\Sigma}_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4)\big],$$
$$\Sigma_A^{(22)}(m,m')=\lim_{T\to\infty}\sum_{l_1,l_2,l_3,l_4=1}^{L}\frac{1}{4}\big[\Sigma_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4)-\acute{\Sigma}_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4)-\grave{\Sigma}_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4)+\bar{\Sigma}_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4)\big],$$

where $\Sigma_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4),\acute{\Sigma}_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4),\grave{\Sigma}_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4),\bar{\Sigma}_{h_m,h_{m'}}^{(T)}(\boldsymbol{l}_4)$ *are as in Theorem C.3.*

D First and second order dependence structure

D.1 Under the null hypothesis of stationarity

Proof of Theorem 4.2. Using Lemma B.1, for $h=1,\dots,T-1$, and

$$\langle x,y\rangle\overline{\langle z,w\rangle}=\langle(x\otimes z)w,y\rangle=\langle x\otimes z,y\otimes w\rangle_{HS}\qquad x,y,z,w\in H,$$

where $\langle\cdot,\cdot\rangle_{HS}$ denotes the Hilbert–Schmidt inner product, the expectation of (3.2) is given by

$$\mathbb{E}\big[\sqrt{T}\gamma_h^{(T)}(l)\big]=\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\left[\frac{1}{T}\Delta_T^{(h)}\frac{\langle\mathcal{F}_{\omega_j},\phi_l^{\omega_j}\otimes\phi_l^{\omega_{j+h}}\rangle_{HS}}{\sqrt{\lambda_l^{\omega_j}\lambda_l^{\omega_{j+h}}}}+O\Big(\frac{1}{T}\Big)\right]=O\Big(\frac{1}{\sqrt{T}}\Big).\tag{D.1}$$

Additionally, Theorem B.1 implies that the covariance structure of the fDFT's is given by

$$\begin{aligned}\operatorname{Cov}\big(D_{\omega_{j_1}}^{(T)}\otimes D_{\omega_{j_1+h_1}}^{(T)},D_{\omega_{j_2}}^{(T)}\otimes D_{\omega_{j_2+h_2}}^{(T)}\big)&=\operatorname{cum}(D_{\omega_{j_1}}^{(T)},D_{-\omega_{j_1+h_1}}^{(T)},D_{-\omega_{j_2}}^{(T)},D_{\omega_{j_2+h_2}}^{(T)})\\&\quad+S_{1324}\big(\operatorname{cum}(D_{\omega_{j_1}}^{(T)},D_{-\omega_{j_2}}^{(T)})\otimes\operatorname{cum}(D_{-\omega_{j_1+h_1}}^{(T)},D_{\omega_{j_2+h_2}}^{(T)})\big)\end{aligned}$$

$$+ S_{1423}\big(\mathrm{cum}(D^{(T)}_{\omega_{j_1}}, D^{(T)}_{\omega_{j_2+h_2}}) \otimes \mathrm{cum}(D^{(T)}_{-\omega_{j_1+h_1}}, D^{(T)}_{-\omega_{j_2}})\big), \quad \text{(D.2)}$$

where S_{ijkl} denotes the permutation operator on $\otimes_{i=1}^4 L^2_{\mathbb{C}}([0,1])$ that permutes the components of a tensor according to the permutation $(1,2,3,4) \mapsto (i,j,k,l)$, that is, $S_{ijkl}(x_1 \otimes \cdots \otimes x_4) = x_i \otimes \cdots \otimes x_l$. Denote then the elementary tensor $\psi_{ll'} = \psi_l \otimes \psi_{l'}$. Write $D^{(l)}_{\omega} = \langle D^{(T)}_{\omega}, \psi_l \rangle$, $\mathcal{F}^{(l,l')}_{\omega} = \langle \mathcal{F}_{\omega}(\psi_{l'}), \psi_l \rangle$ and $\mathcal{F}^{(lm,l'm')}_{\omega_{j_1},\omega_{j_2},\omega_{j_3}} = \langle \mathcal{F}_{\omega_{j_1},\omega_{j_2},\omega_{j_3}}(\psi_{l'm'}), \psi_{lm} \rangle$. Under the conditions of Theorem 4.2, Lemma B.2 implies for the denominator in (3.1) that

$$\begin{aligned}
&\frac{1}{T}\sum_{j_1,j_2=1}^{T}\Bigg(\frac{(2\pi)}{T^2}\mathcal{F}^{(l_1\,l_2,l_3\,l_4)}_{\omega_{j_1},-\omega_{j_1+h_1},-\omega_{j_2}}\Delta_T^{(\omega_{h_2}-\omega_{h_1})} + O\Big(\frac{1}{T^2}\Big) \\
&+ \Big[\mathcal{F}^{(l_1,l_3)}_{\omega_{j_1}}\frac{1}{T}\Delta_T^{(\omega_{j_1}-\omega_{j_2})} + O\Big(\frac{1}{T}\Big)\Big]\Big[\mathcal{F}^{(l_2,l_4)}_{-\omega_{j_1+h_1}}\frac{1}{T}\Delta_T^{(\omega_{j_1+h_1}-\omega_{j_2+h_2})} + O\Big(\frac{1}{T}\Big)\Big] \\
&+ \Big[\mathcal{F}^{(l_1,l_4)}_{\omega_{j_1}}\frac{1}{T}\Delta_T^{(\omega_{j_1}+\omega_{j_2+h_2})} + O\Big(\frac{1}{T}\Big)\Big]\Big[\mathcal{F}^{(l_2,l_3)}_{-\omega_{j_1+h_1}}\frac{1}{T}\Delta_T^{(-\omega_{j_1+h_1}-\omega_{j_2})} + O\Big(\frac{1}{T}\Big)\Big]\Bigg).
\end{aligned}$$

For the statistic (3.2) this structure becomes

$$\begin{aligned}
&\frac{1}{T}\sum_{j_1,j_2=1}^{T}\Bigg(\frac{(2\pi)}{T^2}\langle\mathcal{F}_{\omega_{j_1},-\omega_{j_1+h_1},-\omega_{j_2}}(\phi_{l_2}^{\omega_{j_2}}\otimes\phi_{l_2}^{\omega_{j_2+h_2}}), \phi_{l_1}^{\omega_{j_1}}\otimes\phi_{l_1}^{\omega_{j_1+h_1}}\rangle\Delta_T^{(\omega_{h_2}-\omega_{h_1})} + O\big(\frac{1}{T^2}\big) \\
&+ \Big[\langle\mathcal{F}_{\omega_{j_1}}\phi_{l_2}^{\omega_{j_2}}, \phi_{l_1}^{\omega_{j_1}}\rangle\frac{1}{T}\Delta_T^{(\omega_{j_1}-\omega_{j_2})} + O\big(\frac{1}{T}\big)\Big]\Big[\langle\mathcal{F}_{-\omega_{j_1+h_1}}\phi_{l_2}^{-\omega_{j_2+h_2}}, \phi_{l_1}^{-\omega_{j_1+h_1}}\rangle\frac{1}{T}\Delta_T^{(\omega_{j_1+h_1}-\omega_{j_2+h_2})} + O\big(\frac{1}{T}\big)\Big] \\
&+ \big[\langle\mathcal{F}_{\omega_{j_1}}\phi_{l_2}^{-\omega_{j_2+h_2}}, \phi_{l_1}^{\omega_{j_1}}\rangle\frac{1}{T}\Delta_T^{(\omega_{j_1}+\omega_{j_2+h_2})} + O\big(\frac{1}{T}\big)\Big]\Big[\langle(\mathcal{F}_{-\omega_{j_1+h_1}}\phi_{l_2}^{\omega_{j_2}}, \phi_{l_1}^{-\omega_{j_1+h_1}}\rangle\frac{1}{T}\Delta_T^{(-\omega_{j_1+h_1}-\omega_{j_2})} + O\big(\frac{1}{T}\big)\Big]\Bigg).
\end{aligned}$$

Using that the spectral density operator is self-adjoint and the projection basis are its eigenfunctions, it follows

$$\langle\mathcal{F}_{\omega_{j_1}}(\phi_{l_2}^{\omega_{j_2}}), \phi_{l_1}^{\omega_{j_1}}\rangle = \langle\phi_{l_2}^{\omega_{j_2}}, \mathcal{F}_{\omega_{j_1}}(\phi_{l_1}^{\omega_{j_1}})\rangle = \lambda_{l_1}^{\omega_{j_1}}\langle\phi_{l_2}^{\omega_{j_2}}, \phi_{l_1}^{\omega_{j_1}}\rangle.$$

The covariance strucure of the numerator of (3.2) then reduces to

$$\begin{aligned}
&= \frac{1}{T}\sum_{j_1,j_2=1}^{T}\Bigg(\frac{(2\pi)}{T^2}\langle\mathcal{F}_{\omega_{j_1},-\omega_{j_1+h_1},-\omega_{j_2}}(\phi_{l_2}^{\omega_{j_2}}\otimes\phi_{l_2}^{\omega_{j_2+h_2}}), \phi_{l_1}^{\omega_{j_1}}\otimes\phi_{l_1}^{\omega_{j_1+h_1}}\rangle\Delta_T^{(\omega_{h_2}-\omega_{h_1})} + O\big(\frac{1}{T^2}\big) \\
&+ \Big[\lambda_{l_1}^{\omega_{j_1}}\langle\phi_{l_2}^{\omega_{j_2}}, \phi_{l_1}^{\omega_{j_1}}\rangle\frac{1}{T}\Delta_T^{(\omega_{j_1}-\omega_{j_2})} + O\big(\frac{1}{T}\big)\Big]\Big[\lambda_{l_1}^{-\omega_{j_1+h_1}}\langle\phi_{l_2}^{-\omega_{j_2+h_2}}, \phi_{l_1}^{-\omega_{j_1+h_1}}\rangle\frac{1}{T}\Delta_T^{(\omega_{j_1+h_1}-\omega_{j_2+h_2})} + O\big(\frac{1}{T}\big)\Big] \\
&+ \big[\lambda_{l_1}^{\omega_{j_1}}\langle\phi_{l_2}^{-\omega_{j_2+h_2}}, \phi_{l_1}^{\omega_{j_1}}\rangle\frac{1}{T}\Delta_T^{(\omega_{j_1}+\omega_{j_2+h_2})} + O\big(\frac{1}{T}\big)\Big]\Big[\lambda_{l_1}^{-\omega_{j_1+h_1}}\langle\phi_{l_2}^{\omega_{j_2}}, \phi_{l_1}^{-\omega_{j_1+h_1}}\rangle\frac{1}{T}\Delta_T^{(-\omega_{j_1+h_1}-\omega_{j_2})} + O\big(\frac{1}{T}\big)\Big].
\end{aligned}$$

In case $h_1 \neq h_2$, the first line is of order $O(T^{-1})$. In the second and third line, it can be seen that the cross terms will be of order $O(T^{-1})$ uniformly in ω_j. The first product term in the second line will only be of order $O(1)$ if $j_1 = j_2$ and $j_1 + h_1 = j_2 + h_2$, while in the third line this requires $j_1 = -j_2 - h_2$ and $j_2 = -j_1 - h_1$. It can then be derived that $T\mathrm{Cov}(\gamma^{(T)}_{h_1}(l_1), \overline{\gamma^{(T)}_{h_2}(l_2)}) = O(T^{-1})$ for $h_2 \neq T - h_1$. Since

$$\Re\gamma^{(T)}_{h_1}(l_1) = \frac{1}{2}(\gamma^{(T)}_{h_1}(l_1) + \overline{\gamma^{(T)}_{h_1}(l_1)})$$

and

$$\Im\gamma_{h_1}^{(T)}(l_1) = \frac{1}{2\mathrm{i}}\big(\gamma_{h_1}^{(T)}(l_1) - \overline{\gamma_{h_1}^{(T)}(l_1)}\big),$$

it follows therefore

$$T\mathrm{Cov}\big(\Re\gamma_{h_1}^{(T)}(l_1, \Im\gamma_{h_2}^{(T)}(l_2)\big) = O(T^{-1})$$

uniformly in h_1, h_2. All together, the above derivation yields

$$\begin{aligned} T\,\mathrm{Cov}\big(\Re\gamma_{h_1}^{(T)}(l_1), \Re\gamma_{h_2}^{(T)}(l_2)\big) &= T\,\mathrm{Cov}\big(\Im\gamma_{h_1}^{(T)}(l_1), \Im\gamma_{h_2}^{(T)}(l_2)\big) \\ &= \frac{T}{2}\mathrm{Cov}\big(\gamma_{h_1}^{(T)}(l_1), \gamma_{h_2}^{(T)}(l_2)\big), \end{aligned}$$

and thus Lipschitz-continuity and orthogonality of the eigenfunctions imply that the limiting covariance structure of (3.2) is given by

$$\begin{aligned} &T\,\mathrm{Cov}\Big(\Re\gamma_{h_1}^{(T)}(l_1), \Re\gamma_{h_2}^{(T)}(l_2)\Big) = T\,\mathrm{Cov}\Big(\Im\gamma_{h_1}^{(T)}(l_1), \Im\gamma_{h_2}^{(T)}(l_2)\Big) \\ &= \frac{\delta_{l_1,l_2}}{2} + \frac{1}{4\pi}\int\int \frac{\langle \mathcal{F}_{\omega,-\omega-\omega_h,-\omega'}(\phi_{l_2}^{\omega'} \otimes \phi_{l_2}^{\omega'+\omega_h'}), \phi_{l_1}^{\omega} \otimes \phi_{l_1}^{\omega+\omega_h}\rangle}{\sqrt{\lambda_{l_1}^{\omega}\lambda_{l_2}^{-\omega'}\lambda_{l_1}^{-\omega-\omega_h}\lambda_{l_2}^{\omega'+\omega_h'}}} d\omega d\omega'. \end{aligned}$$

This completes the proof. □

D.2 Under the alternative hypothesis of local stationarity

Proof of Theorem 4.4. (i) In order to prove the first assertion of the theorem, introduce the bias-variance decomposition

$$\begin{aligned} &\mathbb{E}\Big[\big\|\hat{\mathcal{F}}_{\omega}^{(T)} - \mathbb{E}\big[\hat{\mathcal{F}}_{\omega}^{(T)}\big] + \mathbb{E}\big[\hat{\mathcal{F}}_{\omega}^{(T)}\big] - G_{\omega}\big\|_2^2\Big] \\ &\quad = \mathbb{E}\Big[\big\|\hat{\mathcal{F}}_{\omega}^{(T)} - \mathbb{E}\big[\hat{G}_{\omega}\big]\big\|_2^2\Big] + \mathbb{E}\Big[\big\|\mathbb{E}\big[\hat{\mathcal{F}}_{\omega}^{(T)}\big] - G_{\omega}\big\|_2^2\Big]. \end{aligned} \tag{D.3}$$

The cross terms cancel because $\mathbb{E}[\langle \hat{\mathcal{F}}_{\omega}^{(T)} - \mathbb{E}[\hat{\mathcal{F}}_{\omega}^{(T)}], \mathbb{E}[\hat{\mathcal{F}}_{\omega}^{(T)}] - G_{\omega}\rangle_{H\otimes H}]$ and $\mathbb{E}[\hat{\mathcal{F}}_{\omega}^{(T)} - \mathbb{E}[\hat{\mathcal{F}}_{\omega}^{(T)}]] = O_H$. Now, by Lemma B.1,

$$\mathrm{cum}\big(D_{\omega}^{(T)}, D_{-\omega}^{(T)}\big) = \frac{1}{T}\sum_{t=0}^{T-1} \mathcal{F}_{t/T,\omega} + R_{T,2} = G_{\omega}^{(T)} + R_{T,2},$$

where $\|R_{T,2}\|_2 = O(T^{-1})$. Convolution of the cumulant tensor with the smoothing kernel and a subsequent Taylor expansion give

$$\mathbb{E}\big[\hat{\mathcal{F}}_{\omega}^{(T)}\big] = \frac{2\pi}{bT}\sum_{-=1}^{T} K_b(\omega - \omega_j)\mathrm{cum}\big(D_{\omega}^{(T)}, D_{-\omega}^{(T)}\big) = G_{\omega} + \epsilon_{b,T},$$

where $\|\epsilon_{b,T}\|_2 = O(b^2 + (bT))^{-1}$. The interchange of summations is justified by Fubini's Theorem, since $\sup_{\omega,u}\|f_{u,\omega}\|_2 < \infty$ and $\sup_{\omega,u}\|\frac{\partial^2}{\partial\omega_2}f_{u,\omega}\|_2 < \infty$. Here, the error term $(bT)^{-1}$ follows from discretization of the window function (see, for example, Lemma P5.1 of Brillinger, 1981). Note that the integral approximation

in time direction does not change the error term because of Lipschitz continuity in u. Thus, the second term in (D.3) satisfies

$$\mathbb{E}\big[\|\mathbb{E}\hat{\mathcal{F}}_{\omega}^{(T)} - G_{\omega}\|_2^2\big] = O\bigg(b^2 + \frac{1}{bT}\bigg)^2.$$

To bound the first term of the right-hand side in (D.3), observe that, for $\omega \neq \omega'$,

$$\begin{aligned}
\operatorname{cum}(D_{\omega}^{(T)}, D_{\omega'}^{(T)}) &= \frac{1}{2\pi T}\sum_{t,t'=1}^{T} \operatorname{cum}(X_t^{(T)}, X_{t'}^{(T)}) e^{-\mathrm{i}(t\omega + t'\omega')} \\
&= \frac{1}{2\pi T}\sum_{t,t'=1}^{T} \mathcal{C}_{t/T;t'-t}(\tau, \tau') e^{-\mathrm{i}(t'-t)\omega' - \mathrm{i}t(\omega+\omega')} + \epsilon_{t-t',T} \\
&= \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{|h|\leqslant T-t} \operatorname{cum}(X_t^{(T)}, X_{t+h}^{(T)}) e^{-\mathrm{i}(h\omega') - \mathrm{i}t(\omega+\omega')} + R_{T,2} \\
&= \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{|h|\leqslant T-t} \mathcal{F}_{t/T,h} e^{-\mathrm{i}(h\omega') - \mathrm{i}t(\omega+\omega')} + R_{T,2}, \qquad \text{(D.4)}
\end{aligned}$$

where Lemma S1.1 was applied to obtain the second equality sign in combination with

$$\|\epsilon_T\|_2 = \frac{1}{2\pi T}\sum_{t_1=1}^{T} \frac{1+|t_1-t_2|}{T} \|\kappa_{2;t_1-t_2}\|_2 = O\bigg(\frac{1}{T}\bigg)$$

by (4.3). Decompose the corresponding local autocovariance operator of (D.4) as

$$\frac{1}{2\pi T}\bigg(\sum_{h=0}^{T-1}\sum_{t=1}^{T-h} \mathcal{C}_{t/T;h} e^{-\mathrm{i}(h\omega') - \mathrm{i}t(\omega+\omega')} + \sum_{h=-T+1}^{-1}\sum_{t=1}^{T-|h|} \mathcal{C}_{t/T;h} e^{-\mathrm{i}(h\omega') - \mathrm{i}t(\omega+\omega')}\bigg) + R_{T,2}. \qquad \text{(D.5)}$$

Under Assumption 4.3,

$$\begin{aligned}
\bigg\|\frac{1}{2\pi T}\sum_{h=0}^{T-1}\sum_{t=1}^{T-h} \mathcal{C}_{t/T,h} e^{-\mathrm{i}(h\omega') - \mathrm{i}t(\omega+\omega')}\bigg\|_2 &\leqslant \frac{1}{2\pi T}\sum_{h=0}^{T-1}\bigg|\sum_{t=1}^{T-h} e^{-\mathrm{i}t(\omega+\omega')}\bigg| \big\|\mathcal{C}_{t/T,h}\big\|_2 \\
&\leqslant \frac{1}{2\pi T}\sum_{h=0}^{T-1} \big|\Delta_{T-h}^{(\omega+\omega')}\big| \big\|\mathcal{C}_{t/T,h}\big\|_2 \leqslant \frac{C}{T}\sum_{h\in\mathbb{Z}} |h| \|\kappa_{2;h}\|_2 = O\bigg(\frac{1}{T}\bigg)
\end{aligned}$$

for some constant C. A similar derivation shows the same bound holds for the second term of (D.5). It can therefore be concluded that $\big\|\operatorname{cum}(D_{\omega}^{(T)}, D_{\lambda}^{(T)})\big\|_2 = O(T^{-1})$ uniformly in $\omega \neq \omega', 0 \leqslant \omega, \omega' < \pi$. Furthermore, Lemma B.1 and Minkowski's Inequality yield

$$\begin{aligned}
\Big\|\operatorname{cum}(D_{\omega}^{(T)}, D_{-\omega}^{(T)}, D_{\omega'}^{(T)}, D_{-\omega'}^{(T)})\Big\|_2 &\leqslant \frac{1}{T}\bigg\|\frac{1}{T}\sum_{t=0}^{T-1} \mathcal{F}_{\frac{t}{T},\omega,-\omega,\omega'}\bigg\|_2 + O\bigg(\frac{1}{T^2}\bigg) \\
&= \frac{1}{T}\Big\|G_{\omega,-\omega,\omega'}^{(T)}\Big\|_2 + O\bigg(\frac{1}{T^2}\bigg) = O\bigg(\frac{1}{T}\bigg).
\end{aligned}$$

The last equality follows since $\sup_{u,\omega} \|\mathcal{F}_{t/T,\omega,-\omega,\omega'}\|_2 \leqslant \sum_{h_1,h_2,h_3\in\mathbb{Z}} \|\kappa_{3;h_1,h_2,h_3}\|_2 = O(1)$ by Assumption 4.3. Therefore the product theorem for cumulant tensors (Theorem B.1) implies that

$$\operatorname{Cov}\big(I_{\omega}^{(T)}, I_{\omega'}^{(T)}\big) = \operatorname{cum}\big(D_{\omega}^{(T)}, D_{-\omega}^{(T)}, D_{-\omega'}^{(T)}, D_{\omega'}^{(T)}\big)$$

$$+ S_{1324}\Big(\operatorname{cum}\big(D_{\omega}^{(T)}, D_{-\omega'}^{(T)}\big) \otimes \operatorname{cum}\big(D_{-\omega}^{(T)}, D_{\omega'}^{(T)}\big)\Big)$$
$$+ S_{1423}\Big(\operatorname{cum}\big(D_{\omega}^{(T)}, D_{\omega'}^{(T)}\big) \otimes \operatorname{cum}\big(D_{-\omega}^{(T)}, D_{-\omega'}^{(T)}\big)\Big), \tag{D.6}$$

where S_{ijkl} denotes the permutation operator on $\otimes_{i=1}^{4} L^2_{\mathbb{C}}([0,1])$ that permutes the components of a tensor according to the permutation $(1,2,3,4) \mapsto (i,j,k,l)$, that is, $S_{ijkl}(x_1 \otimes \cdots \otimes x_4) = x_i \otimes \cdots \otimes x_l$. It is clear from (D.6) that $\|\!|\operatorname{Cov}(I_{\omega}^{(T)}, I_{\omega'}^{(T)})|\!\|_2 = O(T^{-1})$ for $\omega' \neq \omega, 0 \leqslant \omega, \omega' < \pi$, while for $\omega' = \omega$ it follows that

$$\begin{aligned}\|\!|\operatorname{Cov}\big(I_{\omega}^{(T)}, I_{\omega}^{(T)}\big)|\!\|_2 \leqslant & \Big\| S_{1324}\Big(\operatorname{cum}\big(D_{\omega}^{(T)}, D_{-\omega}^{(T)}\big) \otimes \operatorname{cum}\big(D_{-\omega}^{(T)}, D_{\omega}^{(T)}\big)\Big)\Big\|_2 \\ & + \Big\| S_{1423}\Big(\operatorname{cum}\big(D_{\omega}^{(T)}, D_{\omega}^{(T)}\big) \otimes \operatorname{cum}\big(D_{-\omega}^{(T)}, D_{-\omega}^{(T)}\big)\Big)\Big\|_2 + R_{T,2} = O(1). \end{aligned} \tag{D.7}$$

Furthermore,

$$\begin{aligned}\operatorname{Cov}\big(\hat{\mathcal{F}}_{\omega}^{(T)}, \hat{\mathcal{F}}_{\omega}^{(T)}\big) = & \left(\frac{2\pi}{T}\right)^2 \sum_{j,j'=1}^{T} K_b(\omega - \omega_j) K_b(\omega - \omega_{j'}) \\ & \times \frac{1}{T^2}\Bigg(\sum_{t,t'=0}^{T-1} \mathcal{F}_{t/T,\omega_j} \otimes \mathcal{F}_{t/T,-\omega_j} e^{-\mathrm{i}(t-t')(\omega_j - \omega_{j'})} \\ & + \sum_{t,t'=0}^{T-1} \mathcal{F}_{t/T,\omega_j} \otimes \mathcal{F}_{t/T,-\omega_j} e^{-\mathrm{i}(t-t')(\omega_j + \omega_{j'})} \Bigg) + R_{T,2}. \end{aligned} \tag{D.8}$$

Hence,

$$\begin{aligned}\Big\|\operatorname{Cov}(\hat{\mathcal{F}}_{\omega}^{(T)}, \hat{\mathcal{F}}_{\omega}^{(T)})\Big\|_2 \leqslant & \quad \sup_{u,\omega} \|\!|\mathcal{F}_{u,\omega}|\!\|_2^2 \left(\frac{2\pi}{T}\right)^2 \sum_{j,j'=1}^{T} K_b(\omega - \omega_j) K_b(\omega - \omega_{j'}) \frac{|\Delta_T^{(\omega_j - \omega_{j'})}|^2}{T^2} \\ & + \sup_{u,\omega} \|\!|\mathcal{F}_{u,\omega}|\!\|_2^2 \left(\frac{2\pi}{T}\right)^2 \sum_{j,j'=1}^{T} K_b(\omega - \omega_j) K_b(\omega - \omega_{j'}) \frac{|\Delta_T^{(\omega_j + \omega_{j'})}|^2}{T^2} + O\left(\frac{1}{T}\right) \\ & = O\left(\frac{1}{bT}\right). \end{aligned} \tag{D.9}$$

Together with the equivalence of the Hilbert–Schmidt norm of the operator and the L_2–norm of its kernel, the above implies then that the second term of (D.3) satisfies

$$\mathbb{E}\big[\|\!|\hat{\mathcal{F}}_{\omega}^{(T)} - \mathbb{E}\hat{\mathcal{F}}_{\omega}^{(T)}|\!\|_2^2\big] = \int_{[0,1]^2} \operatorname{Var}(\hat{f}_{\omega}^{(T)}(\tau,\tau'))\, d\tau d\tau' = O\left(\frac{1}{bT}\right)$$

uniformly in $\omega \in [-\pi, \pi]$. This establishes (i).

(ii) The second part of the proof proceeds along similar lines as Paparoditis (2009). An application of Minkowski's inequality yields

$$\begin{aligned}\|\!|\hat{\mathcal{F}}_{\omega}^{(T)} - G_{\omega}|\!\|_2 \leqslant & \left\| \frac{2\pi}{T} \sum_{j=1}^{T} K_b(\omega_{j'} - \omega_j) \Big[\operatorname{cum}(D_{\omega_j}^{(T}, D_{-\omega_j}^{(T)}) - \frac{1}{T} \sum_{t=1} \mathcal{F}_{t/T,\omega_j} \Big] \right\|_2 \\ & + \left\| \frac{2\pi}{T} \sum_{j=1}^{T} K_b(\omega_{j'} - \omega_j) \frac{1}{T} \sum_{t=1} \big[f_{t/T,\omega_j} - \mathcal{F}_{t/T,\omega} \big] \right\|_2 \end{aligned}$$

$$+ \left\| \left(\frac{2\pi}{T} \sum_{j=1}^{T} K_b(\omega_{j'} - \omega_j) - 1 \right) \frac{1}{T} \sum_{t=1} \mathcal{F}_{t/T,\omega} \right\|_2 .$$

Markov's inequality together with (i), which is not affected by the discretization of the integral, imply the first term tends to zero. Since the spectral operator is Lipschitz continuous in ω, the second term is bounded by

$$\left| 1 - \int_{-\pi}^{\pi} K_b(\omega_k - \omega') d\omega' \right| |b| = O\left(1 + \frac{1}{bT}\right) O(b) = O(b).$$

Finally, the third term is seen to be of order $O((bT)^{-1})$. □

Proof of Theorem 4.6. By Lemma B.2, the expectation of $\gamma_h^{(T)}(l, l')$ satisfies

$$\mathbb{E}\big[\gamma_h^{(T)}(l,l')\big] = \frac{1}{T} \sum_{j=1}^{T} \left[\frac{1}{T} \sum_{t=0}^{T-1} \frac{\mathcal{F}^{(l,l')}_{t/T,\omega_j}}{(G^{(l,l)}_{\omega_j} G^{(l',l')}_{\omega_{j+h}})^{1/2}} e^{-\mathrm{i}t\omega_h} + O\left(\frac{1}{T}\right) \right] = O\left(\frac{1}{h^2} + \frac{1}{T}\right),$$

for all $h = 1, \ldots, T-1$. Using then Lipschitz continuity of the spectral operators taking as basis the eigenfunctions of spectral density operator, expression (C.1) becomes

$$\mathbb{E}\big[\gamma_h^{(T)}(l)\big] = \frac{1}{2\pi} \int_0^{2\pi} \int_0^1 \frac{\langle \mathcal{F}_{u,\omega} \tilde{\phi}_l^{\omega+\omega_h}, \tilde{\phi}_l^{\omega} \rangle e^{-\mathrm{i}2\pi uh}}{(\tilde{\lambda}_l^{\omega} \tilde{\lambda}_l^{\omega+\omega_h})^{1/2}} du d\omega + O\left(\frac{1}{T}\right) = O\left(\frac{1}{h^2}\right) + O\left(\frac{1}{T}\right).$$

Note once more that

$$\Re\gamma_h^{(T)}(l) = \frac{1}{2}\Big(\gamma_h^{(T)}(l) + \overline{\gamma_h^{(T)}(l)}\Big) \quad \text{and} \quad \Im\gamma_h^{(T)}(l) = \frac{1}{2\mathrm{i}}\Big(\gamma_h^{(T)}(l) - \overline{\gamma_h^{(T)}(l)}\Big).$$

Under the alternative, these are in fact correlated and four separate cases will have to be considered:

(i) $$\begin{aligned}\mathrm{Cov}\big(\Re\gamma_{h_1}^{(T)}(l_1), \Re\gamma_{h_2}^{(T)}(l_2)\big) = \frac{1}{4}\Big[&\mathrm{Cov}\big(\gamma_{h_1}^{(T)}(l_1), \gamma_{h_2}^{(T)}(l_2)\big) + \mathrm{Cov}\big(\gamma_{h_1}^{(T)}(l_1), \overline{\gamma_{h_2}^{(T)}(l_2)}\big) \\ &+ \mathrm{Cov}\big(\overline{\gamma_{h_1}^{(T)}(l_1)}, \gamma_{h_2}^{(T)}(l_2)\big) + \mathrm{Cov}\big(\overline{\gamma_{h_1}^{(T)}(l_1)}, \overline{\gamma_{h_2}^{(T)}(l_2)}\big)\Big],\end{aligned}$$

(ii) $$\begin{aligned}\mathrm{Cov}\big(\Re\gamma_{h_1}^{(T)}(l_1), \Im\gamma_{h_2}^{(T)}(l_2)\big) = \frac{1}{4\mathrm{i}}\Big[&\mathrm{Cov}\big(\gamma_{h_1}^{(T)}(l_1), \gamma_{h_2}^{(T)}(l_2)\big) - \mathrm{Cov}\big(\gamma_{h_1}^{(T)}(l_1), \overline{\gamma_{h_2}^{(T)}(l_2)}\big) \\ &+ \mathrm{Cov}\big(\overline{\gamma_{h_1}^{(T)}(l_1)}, \gamma_{h_2}^{(T)}(l_2)\big) - \mathrm{Cov}\big(\overline{\gamma_{h_1}^{(T)}(l_1)}, \overline{\gamma_{h_2}^{(T)}(l_2)}\big)\Big],\end{aligned}$$

(iii) $$\begin{aligned}\mathrm{Cov}\big(\Im\gamma_{h_1}^{(T)}(l_1), \Re\gamma_{h_2}^{(T)}(l_2)\big) = \frac{1}{4\mathrm{i}}\Big[&\mathrm{Cov}\big(\gamma_{h_1}^{(T)}(l_1), \gamma_{h_2}^{(T)}(l_2)\big) + \mathrm{Cov}\big(\gamma_{h_1}^{(T)}(l_1), \overline{\gamma_{h_2}^{(T)}(l_2)}\big) \\ &- \mathrm{Cov}\big(\overline{\gamma_{h_1}^{(T)}(l_1)}, \gamma_{h_2}^{(T)}(l_2)\big) - \mathrm{Cov}\big(\overline{\gamma_{h_1}^{(T)}(l_1)}, \overline{\gamma_{h_2}^{(T)}(l_2)}\big)\Big],\end{aligned}$$

(iv) $$\begin{aligned}\mathrm{Cov}\big(\Im\gamma_{h_1}^{(T)}(l_1), \Im\gamma_{h_2}^{(T)}(l_2)\big) = \frac{1}{4}\Big[&\mathrm{Cov}\big(\gamma_{h_1}^{(T)}(l_1), \gamma_{h_2}^{(T)}(l_2)\big) - \mathrm{Cov}\big(\gamma_{h_1}^{(T)}(l_1), \overline{\gamma_{h_2}^{(T)}(l_2)}\big) \\ &- \mathrm{Cov}\big(\overline{\gamma_{h_1}^{(T)}(l_1)}, \gamma_{h_2}^{(T)}(l_2)\big) + \mathrm{Cov}\big(\overline{\gamma_{h_1}^{(T)}(l_1)}, \overline{\gamma_{h_2}^{(T)}(l_2)}\big)\Big].\end{aligned}$$

The expressions for the covariance structure of $\sqrt{T}\gamma_h^{(T)}$ and its conjugate can now be derived using Lemma B.2. For example,

$$\mathrm{Cov}\big(\langle D_{\omega_{j_1}}, \phi_{l_1}^{\omega_{j_1}} \rangle \overline{\langle D_{\omega_{j_1+h_1}}, \phi_{l_1}^{\omega_{j_1+h_1}} \rangle}, \langle D_{\omega_{j_2}}, \phi_{l_2}^{\omega_{j_2}} \rangle \overline{\langle D_{\omega_{j_2+h_2}}, \phi_{l_2}^{\omega_{j_2+h_2}} \rangle}\big)$$

$$
\begin{aligned}
=&\Big[\langle\tilde{\mathcal{F}}_{j_1-j_2;\omega_{j_1}}\phi_{l_2}^{\omega_{j_2}},\phi_{l_1}^{\omega_{j_1}}\rangle+O\Big(\frac{1}{T}\Big)\Big]\Big[\langle\tilde{\mathcal{F}}_{-j_1-h_1+j_2+h_2;-\omega_{j_1+h_1}}\phi_{l_2}^{-\omega_{j_2+h_2}},\phi_{l_1}^{-\omega_{j_1+h_1}}\rangle+O\Big(\frac{1}{T}\Big)\Big]\\
&+\Big[\langle\tilde{\mathcal{F}}_{j_1+j_2+h_2;\omega_{j_1}}\phi_{l_2}^{-\omega_{j_2+h_2}},\phi_{l_1}^{\omega_{j_1}}\rangle+O\Big(\frac{1}{T}\Big)\Big]\Big[\langle\tilde{\mathcal{F}}_{-j_1-h_1-j_2;-\omega_{j_1+h_1}}\phi_{l_2}^{\omega_{j_2}},\phi_{l_1}^{-\omega_{j_1+h_1}}\rangle+O\Big(\frac{1}{T}\Big)\Big]\\
&+\frac{2\pi}{T}\langle\tilde{\mathcal{F}}_{-h_1+h_2;\omega_{j_1},-\omega_{j_1+h_1},-\omega_{j_2}}(\phi_{l_2}^{\omega_{j_2}}\otimes\phi_{l_2}^{\omega_{j_2+h_2}}),\phi_{l_1}^{\omega_{j_1}}\otimes\phi_{l_1}^{\omega_{j_1+h_1}}\rangle+O\Big(\frac{1}{T^2}\Big)
\end{aligned}
$$

and therefore

$$
\begin{aligned}
&\mathrm{Cov}\big(\sqrt{T}\gamma_{h_1}^{(T)}(l_1),\sqrt{T}\gamma_{h_2}^{(T)}(l_2)\big)\\
&\quad=\frac{1}{T}\sum_{j_1,j_2=1}^{T}\Bigg\{\frac{2\pi}{T}\frac{\langle\tilde{\mathcal{F}}_{-h_1+h_2;\omega_{j_1},-\omega_{j_1+h_1},-\omega_{j_2}}(\phi_{l_2}^{\omega_{j_2}}\otimes\phi_{l_2}^{\omega_{j_2+h_2}}),\phi_{l_1}^{\omega_{j_1}}\otimes\phi_{l_1}^{\omega_{j_1+h_1}}\rangle}{(\lambda_{l_1}^{\omega_{j_1}}\lambda_{l_1}^{-\omega_{j_1+h_1}}\lambda_{l_2}^{-\omega_{j_2}}\lambda_{l_2}^{\omega_{j_2+h_2}})^{1/2}}\\
&\qquad+\frac{\langle\tilde{\mathcal{F}}_{j_1-j_2;\omega_{j_1}}\phi_{l_2}^{\omega_{j_2}},\phi_{l_1}^{\omega_{j_1}}\rangle\langle\tilde{\mathcal{F}}_{-j_1-h_1+j_2+h_2;-\omega_{j_1+h_1}}\phi_{l_2}^{-\omega_{j_2+h_2}},\phi_{l_1}^{-\omega_{j_1+h_1}}\rangle}{(\lambda_{l_1}^{\omega_{j_1}}\lambda_{l_1}^{-\omega_{j_1+h_1}}\lambda_{l_2}^{-\omega_{j_2}}\lambda_{l_2}^{\omega_{j_2+h_2}})^{1/2}}\\
&\qquad+\frac{\langle\tilde{\mathcal{F}}_{j_1+j_2+h_2;\omega_{j_1}}\phi_{l_2}^{-\omega_{j_2+h_2}},\phi_{l_1}^{\omega_{j_1}}\rangle\langle\tilde{\mathcal{F}}_{-j_1-h_1-j_2;-\omega_{j_1+h_1}}\phi_{l_2}^{\omega_{j_2}},\phi_{l_1}^{-\omega_{j_1+h_1}}\rangle}{(\lambda_{l_1}^{\omega_{j_1}}\lambda_{l_1}^{-\omega_{j_1+h_1}}\lambda_{l_2}^{-\omega_{j_2}}\lambda_{l_2}^{\omega_{j_2+h_2}})^{1/2}}\\
&\qquad+O\Bigg(\frac{1}{T}\Bigg[\frac{\langle\tilde{\mathcal{F}}_{j_1-j_2;\omega_{j_1}}\phi_{l_2}^{\omega_{j_2}},\phi_{l_1}^{\omega_{j_1}}\rangle}{(\lambda_{l_1}^{\omega_{j_1}}\lambda_{l_2}^{-\omega_{j_2}})^{1/2}}+\frac{\langle\tilde{\mathcal{F}}_{-j_1-h_1+j_2+h_2;-\omega_{j_1+h_1}}\phi_{l_2}^{-\omega_{j_2+h_2}},\phi_{l_1}^{-\omega_{j_1+h_1}}\rangle}{(\lambda_{l_1}^{-\omega_{j_1+h_1}}\lambda_{l_2}^{\omega_{j_2+h_2}})^{1/2}}\\
&\qquad\quad+\frac{\langle\tilde{\mathcal{F}}_{j_1+j_2+h_2;\omega_{j_1}}\phi_{l_2}^{-\omega_{j_2+h_2}},\phi_{l_1}^{\omega_{j_1}}\rangle}{(\lambda_{l_1}^{\omega_{j_1}}\lambda_{l_2}^{\omega_{j_2+h_2}})^{1/2}}+\frac{\langle\tilde{\mathcal{F}}_{-j_1-h_1-j_2;-\omega_{j_1+h_1}}\phi_{l_2}^{\omega_{j_2}},\phi_{l_1}^{-\omega_{j_1+h_1}}\rangle}{(\lambda_{l_1}^{-\omega_{j_1+h_1}}\lambda_{l_2}^{-\omega_{j_2}})^{1/2}}\Bigg]+\frac{1}{T^2}\Bigg)\Bigg\}.
\end{aligned}
$$

The more general expressions are provided in the Online Supplement. □

E Weak convergence

The proof of the distributional properties of $\hat{\gamma}^{(T)}$ as stated in Theorem 4.3 and 4.7 are established in this section. The proof consists of a few steps. First, we derive the properties of $\gamma^{(T)}$, i.e., in which the spectral density operators and the corresponding eigenelements are known. For this, we investigate the distributional properties of the operator

$$
w_h^{(T)}=\frac{1}{T}\sum_{j=1}^{T}D_{\omega_j}^{(T)}\otimes D_{\omega_{j+h}}^{(T)}\quad h=1,\dots T-1. \tag{E.1}
$$

Given appropriate rates of the bandwidths, replacing the denominator of (3.1) with, respectively, consistent estimates of the spectral density operators and its eigenvalues will follow from theorems 4.1 and 4.5. The proofs of these theorems can be found in the Online Supplement. For the empirical counterpart of (3.2), it then remains to show that we can replace the projection basis with the sample versions of the eigenfunctions. In particular, we shall below that

$$
\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\big\langle D_{\omega_j}^{(T)}\otimes D_{\omega_{j+h}}^{(T)},\hat{\phi}_{l,T}^{\omega_j}\otimes\hat{\phi}_{l,T}^{\omega_{j+h}}\big\rangle_{HS}\xrightarrow{D}\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\big\langle D_{\omega_j}^{(T)}\otimes D_{\omega_{j+h}}^{(T)},\phi_{l}^{\omega_j}\otimes\phi_{l}^{\omega_{j+h}}\big\rangle_{HS}.
$$

E.1 Weak convergence on the function space

To demonstrate weak convergence of (E.1), we shall make use of a result (Cremers & Kadelka, 1986), which considerably simplifies the verification of the usual tightness condition often invoked in weak convergence proofs. In particular, the following lemma indicates that weak convergence of the functional process will almost directly follow from the weak convergence of the finite dimensional distributions once it is weakly tight in the following sense.

Lemma E.1. *Let $(\mathscr{T}, \mathcal{A}, \mu)$ be a measure space, let $(B, |\cdot|)$ be a Banach space, and let $X = (X_n \colon n \in \mathbb{N})$ be a sequence of random elements in $L^p_B(\mathscr{T}, \mu)$ such that*

(i) the finite-dimensional distributions of X converge weakly to those of a random element X_0 in $L^p_B(\mathscr{T}, \mu)$;

(ii) $\limsup_{n\to\infty} \mathbb{E}[\|X_n\|_p^p] \leqslant \mathbb{E}[\|X_0\|_p^p]$.

Then, X converges weakly to X_0 in $L^p_B(\mathscr{T}, \mu)$.

To apply it in the present context, consider the sequence $(\hat{E}_h^{(T)} \colon T \in \mathbb{N})$ of random elements in $L^2([0,1]^2, \mathbb{C})$, for $h = 1, \dots, T-1$ defined through

$$\hat{E}_h^{(T)} = \sqrt{T}\Big(w_h^{(T)} - \mathbb{E}\big[w_h^{(T)}\big]\Big).$$

For $\psi_{ll'} = \psi_l \otimes \psi_l'$ an orthonormal basis of $L^2([0,1]^2, \mathbb{C})$, we can represent the process as

$$\hat{E}_h^{(T)} = \sum_{l,l'=1}^{\infty} \langle \hat{E}_h^{(T)}, \psi_{ll'} \rangle \psi_{ll'}$$

from which it is easily seen the finite-dimensional distributions of the basis coefficients provide a complete characterization of the distributional properties of $\hat{E}_h^{(T)}$. Thus, weak convergence of $(\langle \hat{E}_h^{(T)}, \psi_{ll'} \rangle \colon l, l' \in \mathbb{N})$ in the sequence space $\ell^2_{\mathbb{C}}$ will imply weak convergence of the process $(\hat{E}_h^{(T)} \colon T \in \mathbb{N})$. To formalize this, we put the functional $\hat{E}_h^{(T)}$ in duality with $(\hat{E}_h^{(T)})^* \in L^2([0,1]^2, \mathbb{C})^*$ through the pairing

$$\hat{E}_h^{(T)}(\phi) = \langle \hat{E}_h^{(T)}, \phi \rangle$$

for all $\phi \in L^2([0,1]^2, \mathbb{C})^*$. The conditions of Lemma E.1 can now be verified. For the first, the following theorem establishes that the finite-dimensional distributions converge weakly to a Gaussian process both under the null and the alternative.

Theorem E.1. *Under Assumption 4.1 or Assumption 4.3, for all $l_i, l_i' \in \mathbb{N}$, $h_i = 1, \dots, T-1$, $i = 1, \dots, k$ and $k \geqslant 3$,*

$$\operatorname{cum}\Big(\hat{E}_{h_1}^{(T)}(\psi_{l_1 l_1'}), \dots, \hat{E}_{h_k}^{(T)}(\psi_{l_k l_k'})\Big) = o(1) \qquad (T \to \infty).$$

Its proof can be found in Section S3 of the Online Supplement. The second condition of Lemma E.1 will be satisfied if

$$\mathbb{E}\big[\|\hat{E}_h^{(T)}\|_2^2\big] = \sum_{l,l'=1}^{\infty} \mathbb{E}\big[\big|\hat{E}_h^{(T)}(\psi_{ll'})\big|^2\big] \to \sum_{l,l'=1}^{\infty} \mathbb{E}\big[\big|E_h(\psi_{ll'})\big|^2\big] = \mathbb{E}\big[\|E_h\|_2^2\big] \qquad (T \to \infty), \tag{E.2}$$

with E_h denoting the limiting process. Using Theorem (E.1) and (E.2) weak convergence of the functional process can now be determined, distinguishing between the real and imaginary parts.

Theorem E.2 (Weak convergence under the null). *Let $(X_t \colon t \in \mathbb{Z})$ be a stochastic process taking values in $H_{\mathbb{R}}$ satisfying Assumption 4.1 with $\ell = 2$. Then,*

$$\big(\Re\hat{E}_{h_i}^{(T)}, \Im\hat{E}_{h_i}^{(T)} \colon i = 1, \dots, k\big) \xrightarrow{D} (\mathcal{R}_{h_i}, \mathcal{I}_{h_i} \colon i = 1, \dots, k), \tag{E.3}$$

where $\mathcal{R}_{h_1}, \mathcal{I}_{h_2}$, $h_1, h_2 \in \{1, \dots, T-1\}$, are jointly Gaussian elements in $L^2([0,1]^2, \mathbb{C})$ with means $\mathbb{E}[\mathcal{R}_{h_1}(\psi_{ll'})] = \mathbb{E}[\mathcal{I}_{h_2}(\psi_{ll'})] = 0$ and covariances

$$\begin{aligned}
&\operatorname{Cov}\big(\mathcal{R}_{h_1}(\psi_{l_1l_1'}), \mathcal{R}_{h_2}(\psi_{l_2l_2'})\big) && (E.4)\\
&\quad = \operatorname{Cov}\big(\mathcal{I}_{h_1}(\psi_{l_1l_1'}), \mathcal{I}_{h_2}(\psi_{l_2l_2'})\big) \\
&\quad = \frac{1}{4\pi}\int_0^{2\pi} \big\langle \mathcal{F}_{\omega}(\psi_{l_2}), \psi_{l_1}\big\rangle \big\langle \mathcal{F}_{-\omega-\omega_h}(\psi_{l_2'}), \psi_{l_1}\big\rangle d\omega && (E.5)\\
&\qquad + \frac{1}{4\pi}\int_0^{2\pi} \big\langle \mathcal{F}_{\omega}(\psi_{l_2'}), \psi_{l_1}\big\rangle \big\langle \mathcal{F}_{-\omega-\omega_h}(\psi_{l_2}), \psi_{l_1'}\big\rangle d\omega \\
&\qquad + \frac{1}{4\pi}\int_0^{2\pi}\int_0^{2\pi} \big\langle \mathcal{F}_{\omega,-\omega-\omega_h,-\omega'}(\psi_{l_2l_2'}), \psi_{l_1l_1'}\big\rangle d\omega d\omega'
\end{aligned}$$

for all $h_1 = h_2$ and l_1, l_1', l_2, l_2', and 0 otherwise. In addition,

$$\operatorname{Cov}(\mathcal{R}_{h_1}(\psi_{l_1l_1'}), \mathcal{I}_{h_2}(\psi_{l_2l_2'})) = 0$$

uniformly in h_1, h_2 and l_1, l_1', l_2, l_2'.

Proof. The covariance structure is the more general version of Theorem 4.2 as derived in Appendix D.1 while the convergence of the finite–dimensional distributions from Theorem E.1. It then remains to verify that the condition (ii) of Lemma E.1 is satisfied. This follows from the covariance structure since

$$\mathbb{E}\big[\|\hat{E}_h^{(T)}\|_2^2\big] = \int_{[0,1]^2} \operatorname{Var}\big(\hat{E}_h^{(T)}(\tau,\tau')\big) d\tau d\tau' \leqslant T\|\operatorname{Var}(w_h^{(T)})\|_2^2 = 2\|\operatorname{Var}(\mathcal{R}_h)\|_2^2.$$

This completes the proof. □

Under the alternative, a similar result is obtained.

Theorem E.3 (Weak convergence under the alternative). *Let $(X_t \colon t \in \mathbb{Z})$ be a stochastic process taking values in $H_{\mathbb{R}}$ satisfying Assumption 4.3 with $\ell = 2$. Then,*

$$\big(\Re\hat{E}_{h_i}^{(T)}, \Im\hat{E}_{h_i}^{(T)} \colon i = 1, \dots, k\big) \xrightarrow{d} (\mathcal{R}_{h_i}, \mathcal{I}_{h_i} \colon i = 1, \dots, k), \tag{E.6}$$

where $\mathcal{R}_{h_1}, \mathcal{I}_{h_2}$, $h_1, h_2 \in \{1, \dots, T-1\}$, *are jointly Gaussian elements in* $L^2([0,1]^2, \mathbb{C})$ *with means* $\mathbb{E}[\mathcal{R}_{h_1}(\psi_{ll'})] = \mathbb{E}[\mathcal{I}_{h_2}(\psi_{ll'})] = 0$ *and covariance structure*

1. $\mathrm{Cov}(\mathcal{R}_{h_1}(\psi_{l_1 l_1'}), \mathcal{R}_{h_2}(\psi_{l_2 l_2'})) =$
$$\frac{1}{4}\big[\Upsilon_{h_1,h_2}(\psi_{l_1 l_1'\, l_2 l_2'}) + \acute{\Upsilon}_{h_1,h_2}(\psi_{l_1 l_1'\, l_2 l_2'}) + \grave{\Upsilon}_{h_1,h_2}(\psi_{l_1 l_1'\, l_2 l_2'}) + \bar{\Upsilon}_{h_1,h_2}(\psi_{l_1 l_1'\, l_2 l_2'})\big]$$
2. $\mathrm{Cov}(\mathcal{R}_{h_1}(\psi_{l_1 l_1'}), \mathcal{I}_{h_2}(\psi_{l_2 l_2'})) =$
$$\frac{1}{4\mathrm{i}}\big[\Upsilon_{h_1,h_2}(\psi_{l_1 l_1'\, l_2 l_2'}) - \acute{\Upsilon}_{h_1,h_2}(\psi_{l_1 l_1'\, l_2 l_2'}) + \grave{\Upsilon}_{h_1,h_2}(\psi_{l_1 l_1'\, l_2 l_2'}) - \bar{\Upsilon}_{h_1,h_2}(\psi_{l_1 l_1'\, l_2 l_2'})\big]$$
3. $\mathrm{Cov}(\mathcal{I}_{h_1}(\psi_{l_1 l_1'}), \mathcal{I}_{h_2}(\psi_{l_2 l_2'})) =$
$$\frac{1}{4}\big[\Upsilon_{h_1,h_2}(\psi_{l_1 l_1'\, l_2 l_2'}) - \acute{\Upsilon}_{h_1,h_2}(\psi_{l_1 l_1'\, l_2 l_2'}) - \grave{\Upsilon}_{h_1,h_2}(\psi_{l_1 l_1'\, l_2 l_2'}) + \bar{\Upsilon}_{h_1,h_2}(\psi_{l_1 l_1'\, l_2 l_2'})\big]$$

for all h_1, h_2 *and* l_1, l_1', l_2, l_2', *and where* $\Upsilon_{h_1,h_2}, \acute{\Upsilon}_{h_1,h_2}, \grave{\Upsilon}_{h_1,h_2}$ *and* $\bar{\Upsilon}_{h_1,h_2}$ *are given in* (S4.5)–(S4.7).

Proof. The covariance structure is fully derived in the Online Supplement while the convergence of the finite–dimensional distributions follows again from Theorem E.1. Condition (ii) of Lemma E.1 is satisfied since

$$\mathbb{E}\big[\|\hat{E}_h^{(T)}\|_2^2\big] = \int_{[0,1]^2} \mathrm{Var}\big(\hat{E}_h^{(T)}(\tau,\tau')\big) d\tau d\tau' \leqslant T\|\mathrm{Var}(w_h^{(T)})\|_2^2 = \|\mathrm{Var}(\mathcal{R}_h)\|_2^2 + \|\mathrm{Var}(\mathcal{I}_h)\|_2^2,$$

which completes the proof. □

E.2 Replace projection basis with estimates

We now focus on replacing the projection basis with estimates of the eigenfunctions of the spectral density operators. It can be shown (Mas & Menneteau, 2003) that for appropriate rates of the bandwidth b for which the estimated spectral density operator is a consistent estimator of the true spectral density operator, the corresponding estimated eigenprojectors $\hat{\Pi}_l^{\omega} = \hat{\phi}_l^{\omega} \otimes \hat{\phi}_l^{\omega}$ are consistent for the eigenprojectors $\Pi_{l,T}^{\omega}$. However, the estimated eigenfunctions are not unique and are only identified up to rotation on the unit circle. In order to show that replacing the eigenfunctions with estimates does not affect the limiting distribution, we therefore first have to consider the issue of rotation. More specifically, when we estimate $\phi_l^{\omega_j}$, a version $\hat{z}_l \hat{\phi}_l^{\omega_j}$ where $\hat{z}_l \in \mathbb{C}$ with modulus $|\hat{z}_l| = 1$ is obtained. We can therefore not guarantee that an estimated version $\hat{z}_l \hat{\phi}_{l,T}^{\omega_j}$ is close to the true eigenfunction $\hat{\phi}_l^{\omega_j}$. It is therefore essential that our test statistic is invariant to this rotation. To show this, write

$$\Psi_h(l,j) = \langle D_{\omega_j}^{(T)}, \hat{\phi}_l^{\omega_j} \rangle \overline{\langle D_{\omega_{j+h}}^{(T)}, \hat{\phi}_l^{\omega_{j+h}} \rangle}$$

and let $\boldsymbol{\Psi}(h, L) = \mathrm{vec}(\Psi_h(l,j))$ for the stacked vector of dimension $L(T-h) \times 1$, then (3.5) can be written as

$$\hat{\boldsymbol{\beta}}_h^{(T)} = e^{\top} \boldsymbol{\Psi}(h, L)$$

Let us then construct the matrix

$$Z_L^j = \begin{pmatrix} \hat{z}_1^j & \cdots & & \\ \vdots & \hat{z}_2^j & & \\ & & \ddots & \\ & & & \hat{z}_L^j \end{pmatrix}$$

and the block diagonal matrix $Z_L^{1:T} = diag(Z_L^j : j = 1, \ldots, T)$ and additionally the kronecker product

$$\boldsymbol{Z}(h, L) = Z_L^{1:T-h} \otimes Z_L^{h:T}.$$

This object is then of dimension $L(T-h) \times L(T-h)$ and is diagonal with elements $\hat{z}_l^j \overline{\hat{z}}_l^{j+h}, l = 1, \ldots, L; j = 1, \ldots, T-h$. If we rotate the eigenfunctions on the unit circle, we obtain a version

$$\hat{\boldsymbol{\beta}}_h^{(T)} = e^\top \boldsymbol{Z}(h, L)\boldsymbol{\Psi}(h, L),$$

in which case we can write

$$\sqrt{T}\mathcal{Z}_{L,M}\hat{\boldsymbol{b}}_M^{(T)} = \sqrt{T}\big(\Re\hat{\boldsymbol{\beta}}_{h_1}^{(T)}, \ldots, \Im\hat{\boldsymbol{\beta}}_{h_M}^{(T)}, \Im\hat{\boldsymbol{\beta}}_{h_1}^{(T)}, \ldots, \Im\hat{\boldsymbol{\beta}}_{h_M}^{(T)}\big)^\top,$$

where the block diagonal matrix is given by

$$\mathcal{Z}_L = diag(\mathfrak{R}\boldsymbol{Z}(h_1, L), \ldots, \mathfrak{R}\boldsymbol{Z}(h_M, L), \mathfrak{S}\boldsymbol{Z}(h_1, L), \ldots \mathfrak{S}\boldsymbol{Z}(h_M, L))^\top.$$

However this rotation then also implies that $\hat{\Sigma}_M$ becomes $\mathcal{Z}_{L,M}\hat{\Sigma}_M\mathcal{Z}_{L,M}^\top$ and therefore

$$T(\hat{\boldsymbol{b}}_M^{(T)})^\top(\mathcal{Z}_{L,M})^\top[\mathcal{Z}_{L,M}\hat{\Sigma}_M\mathcal{Z}_{L,M}{}^\top]^{-1}\mathcal{Z}_{L,M}\hat{\boldsymbol{b}}_M^{(T)} = \hat{Q}_M^{(T)}$$

showing our statistic does not depend on rotation of the estimated eigenfunctions. In the rest of the proof, we can therefore focus on estimates $\hat{\phi}_{l,T}^{\omega_{j+h}}$ and $\hat{\phi}_{l,T}^{\omega_{j+h}}$ and ignore their respective unknown rotations $\hat{z}_l^j$ and $\overline{\hat{z}}_l^{j+h}$.

Given the estimates of the spectral density operators are consistent, note that the consistency of the estimated eigenprojectors together with continuity of the tensor product imply that

$$\sup_\omega \|\|\widehat{\Pi}_{l,T,h}^\omega - \Pi_{l,h}^\omega\|\|_2^2 \xrightarrow{p} 0 \qquad T \to \infty, \tag{E.7}$$

where the projectors are given by $\Pi_{l,h}^\omega := \phi_l^{\omega_j} \otimes \phi_l^{\omega_{j+h}}$. In the previous section, we moreover showed

$$\frac{1}{\sqrt{T}}\sum_{j=1}^T D_{\omega_j}^{(T)} \otimes D_{\omega_{j+h}}^{(T)} \xrightarrow{\mathcal{D}} w_h^{(T)}$$

Using Portmanteau's Theorem, this means there exists a large M_δ s.t.

$$\limsup_{T\to\infty} \mathbb{P}\Big(\Big\|\frac{1}{\sqrt{T}}\sum_{j=1}^T D_{\omega_j}^{(T)} \otimes D_{\omega_{j+h}}^{(T)}\Big\|_2 \geqslant M_\delta\Big) \leqslant \mathbb{P}\Big(\big\|w_h^{(T)}\big\|_2 \geqslant M_\delta\Big) \leqslant \delta, \delta > 0.$$

The Cauchy-Schwarz inequality and the definition of the Hilbert-Schmidt inner product yield

$$\limsup_{T\to\infty}\mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\big\langle D_{\omega_j}^{(T)}\otimes D_{\omega_{j+h}}^{(T)},\hat{\phi}_{l,T}^{\omega_j}\otimes\hat{\phi}_{l,T}^{\omega_{j+h}}-\phi_l^{\omega_j}\otimes\phi_l^{\omega_{j+h}}\big\rangle_{HS}\Big|>\epsilon\Big)$$

$$\leqslant\limsup_{T\to\infty}\mathbb{P}\Big(\int\int\Big|\frac{1}{\sqrt{T}}\sum_{j=1}^{T}D_{\omega_j}^{(T)}(\tau)D_{-\omega_{j+h}}^{(T)}(\tau')\big[\hat{\phi}_{l,T}^{-\omega_j}(\tau)\hat{\phi}_{l,T}^{\omega_{j+h}}(\tau')-\phi_l^{-\omega_j}(\tau)\phi_l^{\omega_{j+h}}(\tau')\big]\Big|d\tau d\tau'>\epsilon\Big)$$

and using (E.7)

$$\leqslant\limsup_{T\to\infty}\mathbb{P}\Big(\sup_{\omega}|||\hat{\phi}_{l,T}^{\omega}\otimes\hat{\phi}_{l,T}^{\omega_h}-\phi_l^{\omega}\otimes\phi_l^{\omega_h}|||_2|||\frac{1}{\sqrt{T}}\sum_{j=1}^{T}D_{\omega_j}^{(T)}\otimes D_{\omega_{j+h}}^{(T)}|||_2>\epsilon\Big)$$

$$\leqslant\limsup_{T\to\infty}\mathbb{P}(\sup_{\omega}|||\hat{\phi}_{l,T}^{\omega}\otimes\hat{\phi}_{l,T}^{\omega_h}-\phi_l^{\omega}\otimes\phi^{\omega_h}|||_2>\epsilon/M_\delta\Big)+\limsup_{T\to\infty}\mathbb{P}\Big(|||\frac{1}{\sqrt{T}}\sum_{j=1}^{T}D_{\omega_j}^{(T)}\otimes D_{\omega_{j+h}}^{(T)}|||_2\geqslant M_\delta\Big)$$

$$\leqslant 0+\delta$$

for every $\delta>0$. Given the bandwidth condition (Assumption 4.2) is satisfied, we therefore finally obtain

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\big\langle D_{\omega_j}^{(T)}\otimes D_{\omega_{j+h}}^{(T)},\hat{\phi}_l^{\omega_{j+h}}\otimes\hat{\phi}_l^{\omega_j}\big\rangle_{HS}\xrightarrow{\mathcal{D}}\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\big\langle D_{\omega_j}^{(T)}\otimes D_{\omega_{j+h}}^{(T)},\phi_l^{\omega_{j+h}}\otimes\phi_l^{\omega_j}\big\rangle_{HS}\tag{E.8}$$

as $T\to\infty$.

References

Antoniadis, A. & T. Sapatinas (2003). Wavelet methods for continuous time prediction using Hilbert-valued autoregressive processes. *Journal of Multivariate Analysis* **87**, 133–158.

Aue, A., Dubart Nourinho, D. & S. Hörmann (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association* **110**, 378–392.

Aue, A. & L. Horváth (2013). Structural breaks in time series. *Journal of Time Series Analysis* **34**, 1–16.

Aue, A., Rice, G. & O. Sönmez (2017). Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society, Series B*, forthcoming.

Aue, A. & A. van Delft (2017). Online supplement to "Testing for stationarity of functional time series in the frequency domain".

Bandyopadhyay, S. & S. Subba Rao (2017). A test for stationarity for irregularly spaced spatial data. *Journal of the Royal Statistical Society, Series B* **79**, 95–123.

Bandyopadhyay, S., Jentsch, C. & S. Subba Rao (2017). A spectral domain test for stationarity of spatio-temporal data. *Journal of Time Series Analysis* **38**, 326–351.

Besse, P., Cardot, H. & D. Stephenson (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics* **27**, 673–687.

Bosq, D. (2000). *Linear Processes in Function Spaces.* Springer-Verlag, New York.

Brillinger, D. (1981). *Time Series: Data Analysis and Theory.* McGraw Hill, New York.

Brillinger, D. & M. Rosenblatt (1967). Asymptotic theory of estimates of k-th order spectra. In *Spectral Analysis of Time Series (Ed. B. Harris)*, Wiley, New York, pages 153–188.

Cremers, H. & D. Kadelka (1986). On weak convergence of integral functions of stochastic processes with applications to processes taking paths in L_p^E. *Stochastic Processes and their Applications* **21**, 305–317.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* **25**, 1–37.

Dette, H., Preuß, P. & M. Vetter (2011). A measure of stationarity in locally stationary processes with applications to testing. *Journal of the American Statistical Association* **106**, 1113–1124.

Dwivedi, Y. & Subba Rao, S. (2011). A test for second-order stationarity of a time series based on the discrete Fourier transform. *Journal of Time Series Analysis* **32**, 68–91.

Ferraty, F. & Vieu, P. (2010). *Nonparametric Functional Data Analysis.* Springer-Verlag, New York.

Hörmann, S., Kidziński, Ł. & M. Hallin (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society, Series B* **77**, 319–348.

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

Horváth, L. & P. Kokoszka (2012). *Inference for Functional Data with Applications.* Springer-Verlag, New York.

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

Hsing, T. & R. Eubank (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators.* Wiley, New York.

Jentsch, C. & S. Subba Rao (2015). A test for second order stationarity of a multivariate time series. *Journal of Econometrics* **185**, 124–161.

Jin, L., Wang, S. & H. Wang (2015). A new non-parametric stationarity test of time series in the time domain. *Journal of the Royal Statistical Society, Series B* **77**, 893–922.

Lee, J. & S. Subba Rao (2016). A note on general quadratic forms of nonstationary stochastic processes. Technical Report, Texas A&M University.

Li, Y. & T. Hsing (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics* **38**, 3321–3351.

Mas, A. & L. Menneteau (2003). Perturbation approach applied to the asymptotic study of random operators. In: *Hoffmann-Jørgensen et al. (eds.). High dimensional probability III*. Birkhäuser, Boston, pages 127–134.

Nason, G. (2013). A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *Journal of the Royal Statistical Society, Series B* **75**, 879–904.

Panaretos, V. & S. Tavakoli (2013). Fourier analysis of stationary time series in function space. *The Annals of Statistics* **41**, 568–603.

Paparoditis, E. (2009). Testing temporal constancy of the spectral structure of a time series. *Bernoulli* **15**, 1190–1221.

Preuß, P., Vetter, M. & H. Dette (2013). A test for stationarity based on empirical processes. *Bernoulli* **19**, 2715–2749.

Priestley, M.B. & T. Subba Rao (1969). A test for non-stationarity of time-series. *Journal of the Royal Statistical Society, Series B* **31**, 140–149.

Ramsay, J.O. & B.W. Silverman (2005). *Functional Data Analysis (2nd ed.).* Springer-Verlag, New York.

Subba Rao, S. (2016). Orthogonal samples for estimators in time series. Preprint available at `https://arxiv.org/abs/1611.00398`.

Van Delft, A. & M. Eichler (2016). Locally stationary functional time series. Preprint, available at `https://arxiv.org/pdf/1602.05125v2.pdf`.

Von Sachs, R. & M.H. Neumann (1999). A wavelet-based test for stationarity. *Journal of Time Series Analysis* **21**, 597–613.

Online Supplement to "Testing for stationarity of functional time series in the frequency domain"[1]

Alexander Aue[2] Anne van Delft[3]

October 10, 2017

Abstract

This supplement contains additional technical material necessary to complete the proofs of theorems of the main paper Aue & van Delft (2016). Section S1 contains the proofs of several auxiliary lemmas stated in Appendix A of the main paper. Section S2 contains results on finding bounds for the denominator of the test statistic. Sections S3 deals with convergence of the finite-dimensional distributions. Section S4 establishes the asymptotic covariance structure of the test under the alternative of local stationarity.

Keywords: Frequency domain methods, Functional data analysis, Locally stationary processes, Spectral analysis

MSC 2010: Primary: 62G99, 62H99, Secondary: 62M10, 62M15, 91B84

S1 Properties of functional cumulants under local stationarity

Lemma S1.1. *Let Assumption 4.3 be satisfied and* $c_{u;t_1,\ldots,t_{k-1}}$ *as given in* (4.6)*. Then,*

$$\Big\|\mathrm{cum}\big(X_{t_1}^{(T)},\ldots,X_{t_{k-1}}^{(T)},X_{t_k}^{(T)}\big)-c_{t_1/T;t_1-t_k,\ldots,t_{k-1}-t_k}\Big\|_2 \leqslant \left(\frac{k}{T}+\sum_{j=1}^{k-1}\Big|\frac{t_j-t_k}{T}\Big|\right)\|\kappa_{k;t_1-t_k,\ldots,t_{k-1}-t_k}\|_2.$$

Proof of Lemma A.2. By linearity of the cumulant operation, consecutively taking differences leads, by equation (4.4) of the main paper and the triangle inequality, to

$$\Big\|\mathrm{cum}\big(X_{t_1}^{(T)},\ldots,X_{t_k}^{(T)}\big)-\mathrm{cum}\big(X_{t_1}^{(t_1/T)},\ldots,X_{t_k}^{(t_k/T)}\big)\Big\|_2 \leqslant K\frac{k}{T}\|\kappa_{k;t_1-t_k,\ldots,t_{k-1}-t_k}\|_2,$$

using part (i) of Assumption 4.3. By (4.4),

$$X_{t_j}^{(t_j/T)}-X_{t_j}^{(t_k/T)}=\frac{(t_j-t_k)}{T}Y_{t_j}^{(t_j/T,t_k/T)}. \tag{S1.1}$$

[1]AA was partially supported by NSF grants DMS 1305858 and DMS 1407530. AvD was partially supported 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).

[2]Department of Statistics, University of California, Davis, CA 95616, USA, email: aaue@ucdavis.edu

[3]Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany, email: Anne.vanDelft@rub.de

Similarly,

$$\Big\|\mathrm{cum}\big(X_{t_1}^{(t_1/T)},\ldots,X_{t_k}^{(t_k/T)}\big)-c_{t_1/T;t_1-t_k,\ldots,t_{k-1}-t_k}\Big\|_2 \leqslant \sum_{j=1}^{k-1}\frac{|t_j-t_k|}{T}\|\kappa_{k;t_1-t_k,\ldots,t_{k-1}-t_k}\|_2,$$

which follows from part (iii) of Assumption 4.3 Minkowski's inequality then implies the lemma. □

Lemma S1.2. *Consider a sequence of functional processes* $(X_t^{(T)}\colon t\leqslant T, T\in\mathbb{N})$ *satisfying Assumption 4.3 with* $k=2$ *and* $\ell=2$*. Then, this triangular array uniquely characterizes the time-varying local spectral density operator*

$$\mathcal{F}_{u,\omega}=\frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\mathfrak{C}_{u,h}e^{-\mathrm{i}\omega h}, \tag{S1.2}$$

which belongs to $S_2(H)$*. Its kernel satisfies*

(i) $\sup_{u,\omega}\|\frac{\partial^i}{\partial u^i}f_{u,\omega}\|_2<\infty$ *for* $i=1,2$*,*

(ii) $\sup_{u,\omega}\|\frac{\partial^i}{\partial^i\omega}f_{u,\omega}\|_2<\infty$ *for* $i=1,2$*,*

(iii) $\sup_{u,\omega}\|\frac{\partial^2}{\partial\omega\partial u}f_{u,\omega}\|_2<\infty$.

Proof of Lemma A.3. Using Lemma A.2, it is straightforward to show that $(X_t^T\colon t\leqslant T, T\in\mathbb{N})$ uniquely determines the time-varying spectral density operator, that is,

$$\int_{-\pi}^{\pi}\|\mathcal{F}_{u,\omega}^{(T)}-\mathcal{F}_{u,\omega}\|_2^2\,d\omega=o(1)\qquad(T\to\infty). \tag{S1.3}$$

Existence of the derivatives follow from the dominated convergence theorem and the product rule for differentiation in Banach spaces (Nelson, 1969). □

Proof of Lemma A.4. The first line of (A.4) follows on replacing the cumulants $\mathrm{cum}(X_{t_1}^{(T)},\ldots,X_{t_{k-1}}^{(T)},X_{t_k}^{(T)})$ with $c_{t_k/T;t_1-t_k,\ldots,t_{k-1}-t_k}$ and Lemma A.2. The second line because the discretization of the integral is an operation of order $O(T^{-2})$.

By Assumption 4.3 (iv), the kernel of $u\mapsto\frac{\partial}{\partial u}\mathcal{F}_{u;\omega_1,\ldots,\omega_{k-1}}$ satisfies

$$\Big\|\sup_u\frac{\partial}{\partial u}f_{u;\omega_1,\ldots,\omega_{k-1}}\Big\|_2\leqslant\frac{1}{(2\pi)^{k-1}}\sum_{t_1,\ldots,t_k}\|\kappa_{k;t_1-t_k,\ldots,t_{k-1}-t_k}\|_2<\infty.$$

The dominated convergence theorem therefore yields

$$\sup_{u,\omega_1,\ldots,\omega_{k-1}}\Big\|\frac{\partial}{\partial u}f_{u,\omega_1,\ldots,\omega_{k-1}}\Big\|_2<\infty. \tag{S1.4}$$

Finally, integration by parts for a periodic function in $L^2([0,1]^k)$ with existing n-th directional derivative in u, yields

$$\begin{aligned}
&\|\tilde{f}_{s;\omega_{j_1},\ldots,\omega_{j_{k-1}}}\|_2^2 \\
&= \int_{[0,1]^k} \left| \left[\frac{\frac{\partial^{n-1}}{\partial u^{n-1}} f_{u;\omega_{j_1},\ldots,\omega_{j_{k-1}}}(\boldsymbol{\tau})}{(-\mathrm{i}2\pi s)^{n-1}} e^{-\mathrm{i}s2\pi u} \right]_0^1 - \int_0^1 \frac{e^{-\mathrm{i}s2\pi u}}{(-\mathrm{i}2\pi s)^n} \frac{\partial^n}{\partial u^n} f_{u;\omega_{j_1},\ldots,\omega_{j_{k-1}}}(\boldsymbol{\tau}) du \right|^2 d\boldsymbol{\tau} \\
&= \int_{[0,1]^{k+2}} \frac{1}{(2\pi s)^{2n}} e^{\mathrm{i}2\pi s(u-v)} \frac{\partial^2}{\partial u^2} f_{u;\omega_{j_1},\ldots,\omega_{j_{k-1}}}(\boldsymbol{\tau}) \frac{\partial^2}{\partial v^2} f_{v;\omega_{j_1},\ldots,\omega_{j_{k-1}}}(\boldsymbol{\tau}) d\boldsymbol{\tau} du dv \\
&\leqslant \frac{1}{(2\pi s)^{2n}} \int_{[0,1]^{k+2}} \left| \frac{\partial^2}{\partial u^2} f_{u;\omega_{j_1},\ldots,\omega_{j_{k-1}}}(\boldsymbol{\tau}) \frac{\partial^2}{\partial v^2} f_{v;\omega_{j_1},\ldots,\omega_{j_{k-1}}}(\boldsymbol{\tau}) \right| d\boldsymbol{\tau} du dv \\
&\leqslant \frac{1}{(2\pi s)^{2n}} \int_{[0,1]^2} \left\| \frac{\partial^2}{\partial u^2} f_{u;\omega_{j_1},\ldots,\omega_{j_{k-1}}} \right\|_2 \left\| \frac{\partial^2}{\partial v^2} f_{v;\omega_{j_1},\ldots,\omega_{j_{k-1}}} \right\|_2 du dv \\
&\leqslant \frac{1}{(2\pi s)^{2n}} \left(\sup_u \left\| \frac{\partial^2}{\partial u^2} f_{u;\omega_{j_1},\ldots,\omega_{j_{k-1}}} \right\|_2 \right)^2 < \infty,
\end{aligned}$$

where the Cauchy–Schwarz inequality was applied in the second-to-last equality. The interchange of integrals is justified by Fubini's theorem. Thus,

$$\sup_{\omega_1,..,\omega_{k-1}} \|\tilde{f}_{s;\omega_{j_1},\ldots,\omega_{j_{k-1}}}\|_2 \leqslant \frac{1}{(2\pi)^{2n}} \sup_{u,\omega_1,\ldots,\omega_n} \left\| \frac{\partial^n}{\partial u^n} f_{u;\omega_{j_1},\ldots,\omega_{j_{k-1}}} \right\|_2 |s|^{-n}. \tag{S1.5}$$

and (A.6) follows from Assumption 4.3 (iv). □

Proof of Corollary A.1. For completeness, we elaborate on part (ii) of the corollary. Note that

$$\|\!|\tilde{\mathcal{F}}_{0:\omega}|\!\|_2 \leqslant \sup_{\omega,u} \|\!|\mathcal{F}_{u,\omega}|\!\|_2 < \sum_h \|\kappa_{2,h}\|_2 < \infty.$$

The p-harmonic series for $p = 2$ then yields

$$\sup_{\omega} \sum_{s\in\mathbb{Z}} \|\!|\tilde{\mathcal{F}}_{s;\omega}|\!\|_2 \leqslant \sum_h \|\kappa_{2,h}\|_2 \left(1 + \frac{1}{(2\pi)^4} \frac{\pi^2}{3}\right) < \infty, \tag{S1.6}$$

where the constant $(2\pi)^{-4}$ follows from (S1.5). □

S2 Error bound for the denominator of the test statistic

A bound needs to be obtained on the error resulting from the replacement of the unknown spectral operators with consistent estimators. It will be sufficient to consider a bound on

$$\sqrt{T}|\gamma_h^{(T)}(l_1,l_2) - \hat{\gamma}_h^{(T)}(l_1,l_2)|$$

for all $l_1, l_2 \in \mathbb{N}$, where $\gamma^{(T)}$ is defined in equation (3.1) of the main paper. Consider the function $g(x) = x^{-1/2}$, $x > 0$, and notice that

$$\gamma_h^{(T)}(l_1,l_2) - \hat{\gamma}_h^{(T)}(l_1,l_2) = \frac{1}{T} \sum_{j=1}^{T} D_{\omega_j}^{(l_1)} D_{-\omega_{j+h}}^{(l_2)} \Big[g(\mathcal{F}_{\omega_j}^{(l_1,l_1)} \mathcal{F}_{\omega_{j+h}}^{(l_2,l_2)}) - g(\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)} \hat{\mathcal{F}}_{\omega_{j+h}}^{l_2,l_2)}) \Big].$$

Given the assumption $\inf_\omega \langle \mathcal{F}_\omega(\psi), \psi \rangle > 0$ for all $\psi \in H$ is satisfied, continuity of the inner product implies that for fixed l_1, l_2 the mean value theorem may be applied to find

$$\begin{aligned}
&\gamma_h^{(T)}(l_1, l_2) - \hat{\gamma}_h^{(T)}(l_1, l_2) \\
&\qquad = \frac{1}{T} \sum_{j=1}^{T} D_{\omega_j}^{(l_1)} D_{-\omega_{j+h}}^{(l_2)} \left[\frac{\partial g(x)}{\partial x} \Big|_{x = \breve{\mathcal{F}}_{\omega_j}^{(l_1,l_1)} \breve{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)}} \big(\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)} \hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)} - \mathcal{F}_{\omega_j}^{(l_1,l_1)} \mathcal{F}_{\omega_{j+h}}^{(l_2,l_2)} \big) \right],
\end{aligned}$$

where $\breve{\mathcal{F}}_{\omega_j}^{(l_1,l_1)} \breve{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)}$ lies in between $\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)} \hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)}$ and $\mathcal{F}_{\omega_j}^{(l_1,l_1)} \mathcal{F}_{\omega_{j+h}}^{(l_2,l_2)}$. Because of uniform convergence of $\hat{\mathcal{F}}_{\omega_j} \in S_2(H)$ with respect to ω, it follows that

$$\sqrt{T} |\gamma_h^{(T)}(l_1, l_2) - \hat{\gamma}_h^{(T)}(l_1, l_2)| = O_p(1) \left| \frac{1}{T} \sum_{j=1}^{T} \frac{D_{\omega_j}^{(l_1)} D_{-\omega_{j+h}}^{(l_2)}}{(\mathcal{F}_{\omega_j}^{(l_1,l_1)} \mathcal{F}_{\omega_{j+h}}^{(l_2,l_2)})^{3/2}} \big(\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)} \hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)} - \mathcal{F}_{\omega_j}^{(l_1,l_1)} \mathcal{F}_{\omega_{j+h}}^{(l_2,l_2)} \big) \right|. \tag{S2.1}$$

With these preliminary results at hand, the first main theorems can be verified.

Proof of Theorems 4.1 and 4.5. In order to prove both theorems, partition (S2.1) as follows

$$J_1(l, l_2) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} \frac{D_{\omega_j}^{(l_1)} D_{-\omega_{j+h}}^{(l_2)} - \mathbb{E}[D_{\omega_j}^{(l_1)} D_{-\omega_{j+h}}^{(l_2)}]}{(\mathcal{F}_{\omega_j}^{(l_1,l_1)} \mathcal{F}_{\omega_{j+h}}^{(l_2,l_2)})^{3/2}} \hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)} \big(\hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)} - \mathcal{F}_{\omega_{j+h}}^{(l_2,l_2)} \big), \tag{S2.2}$$

$$J_2(l, l_2) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} \frac{D_{\omega_j}^{(l_1)} D_{-\omega_{j+h}}^{(l_2)} - \mathbb{E}[D_{\omega_j}^{(l_1)} D_{-\omega_{j+h}}^{(l_2)}]}{(\mathcal{F}_{\omega_j}^{(l_1,l_1)} \mathcal{F}_{\omega_{j+h}}^{(l_2,l_2)})^{3/2}} \mathcal{F}_{\omega_{j+h}}^{(l_2,l_2)} \big(\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)} - \mathcal{F}_{\omega_j}^{(l_1,l_1)} \big), \tag{S2.3}$$

$$J_3(l, l_2) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} \frac{\mathbb{E}[D_{\omega_j}^{(l_1)} D_{-\omega_{j+h}}^{(l_2)}]}{(\mathcal{F}_{\omega_j}^{(l_1,l_1)} \mathcal{F}_{\omega_{j+h}}^{(l_2,l_2)})^{3/2}} \big(\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)} \hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)} - \mathbb{E}[\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)} \hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)}] \big), \tag{S2.4}$$

$$J_4(l, l_2) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} \frac{\mathbb{E}[D_{\omega_j}^{(l_1)} D_{-\omega_{j+h}}^{(l_2)}]}{(\mathcal{F}_{\omega_j}^{(l_1,l_1)} \mathcal{F}_{\omega_{j+h}}^{(l_2,l_2)})^{3/2}} \big(\mathbb{E}[\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)} \hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)}] - \mathcal{F}_{\omega_j}^{(l_1,l_1)} \mathcal{F}_{\omega_{j+h}}^{(l_2,l_2)} \big). \tag{S2.5}$$

The proof of both theorems are based on Corollary S2.1 and Lemmas S2.1-S2.4 below. These results together with an application of the Cauchy–Schwarz inequality yield

(i) $|J_1| = O_p\left(\frac{1}{\sqrt{bT}} + b^2 \right)$,

(ii) $|J_2| = \begin{cases} O_p\left(\frac{1}{\sqrt{bT}} + \frac{b^2}{\sqrt{T}} \right) & \text{under Assumption 4.1,} \\ O_p\left(\frac{1}{\sqrt{bT}} + b^2 \right) & \text{under Assumption 4.3,} \end{cases}$

(iii) $|J_3| = \begin{cases} O_p\left(\frac{1}{\sqrt{bT}} \right) & \text{under Assumption 4.1,} \\ O_p\left(\frac{1}{\sqrt{bT}} \right) & \text{under Assumption 4.3,} \end{cases}$

(iv) $|J_4| = \begin{cases} O\Big(b^2 + \dfrac{1}{bT}\Big) & \text{under Assumption 4.1.} \\ O\Big(\sqrt{T}b^2 + \dfrac{1}{b\sqrt{T}}\Big) & \text{under Assumption 4.3.} \end{cases}$

Minkowski's Inequality then gives the result. □

Corollary S2.1. *Under Assumption 4.1*

$$\mathbb{E}\big[\|\hat{\mathcal{F}}_\omega^{(T)} - \mathcal{F}_\omega\|_2^2\big] = O\Big(\frac{1}{bT} + b^4\Big) \qquad (T \to \infty), \tag{S2.6}$$

while, under Assumption 4.3,

$$\mathbb{E}\big[\|\hat{\mathcal{F}}_\omega^{(T)} - G_\omega\|_2^2\big] = O\Big(\frac{1}{bT} + b^4\Big) \qquad (T \to \infty). \tag{S2.7}$$

Proof. See Panaretos & Tavakoli (2013, Theorem 3.6) and Theorem 4.4, respectively. □

Lemma S2.1. *Let* $(g_j^{(l_1,l_2)} : j \in \mathbb{Z})$ *be a bounded sequence in* $\mathbb{C}$ *for all* $l_1, l_2 \in \mathbb{N}$ *such that* $\inf_j g_j^{(l_1,l_2)} > 0$. *Under Assumption 4.1 and under Assumption 4.3,*

$$\mathbb{E}\Big[\Big|\frac{1}{\sqrt{T}}\sum_{j=1}^{T} g_j^{(l_1,l_2)}\big(D_{\omega_j}^{(l_1)}D_{-\omega_{j+h}}^{(l_2)} - \mathbb{E}\big[D_{\omega_j}^{(l_1)}D_{-\omega_{j+h}}^{(l_2)}\big]\big)\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)}\Big|^2\Big] = O(1).$$

Proof. Observe that

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\frac{1}{\sqrt{T}}\sum_{j=1}^{T} g_j^{(l_1,l_2)}\big(D_{\omega_j}^{(l_1)}D_{-\omega_{j+h}}^{(l_2)} - \mathbb{E}\big[D_{\omega_j}^{(l_1)}D_{-\omega_{j+h}}^{(l_2)}\big]\big)\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)}\Big|^2\Big] \\
&= \mathbb{E}\Big[\Big|\frac{1}{\sqrt{T}}\sum_{j=1}^{T} g_j^{(l_1,l_2)}\big(D_{\omega_j}^{(l_1)}D_{-\omega_{j+h}}^{(l_2)} - E\big[D_{\omega_j}^{(l_1)}D_{-\omega_{j+h}}^{(l_2)}\big]\big)\frac{2\pi}{bT}\sum_{j'=1}^{T} K\Big(\frac{\omega_{j'}}{b}\Big)D_{\omega_{j-j'}}^{(l_1)}D_{\omega_{j'-j}}^{(l_1)}\Big|^2\Big] \\
&= \frac{1}{T}\Big(\frac{2\pi}{bT}\Big)^2 \sum_{j_1,j_2=1}^{T} g_{j_1,j_2}^{(l_1,l_2)} \sum_{j_1',j_2'=1}^{T} K\Big(\frac{\omega_{j_1'}}{b}\Big)K\Big(\frac{\omega_{j_2'}}{b}\Big)\mathbb{E}\Big[\big\{\big(D_{\omega_{j_1}}^{(l_1)}D_{-\omega_{j_1+h}}^{(l_2)} - \mathbb{E}\big[D_{\omega_{j_1}}^{(l_1)}D_{-\omega_{j_1+h}}^{(l_2)}\big]\big) \\
&\qquad \times D_{\omega_{j_1-j_1'}}^{(l_1)}D_{\omega_{j_1'-j_1}}^{(l_1)}\big\}\big\{\big(D_{-\omega_{j_2}}^{(l_1)}D_{\omega_{j_2+h}}^{(l_2)} - \mathbb{E}\big[D_{-\omega_{j_2}}^{(l_1)}D_{\omega_{j_2+h}}^{(l_2)}\big]\big)D_{\omega_{j_2'-j_2}}^{(l_1)}D_{\omega_{j_2-j_2'}}^{(l_1)}\big\}\Big].
\end{aligned}$$

Expanding the expectation in terms of cumulants, one obtains the structure

$$\begin{aligned}
&\mathbb{E}[(X - \mathbb{E}X)Y][(W - EW)Z] \\
&\quad = E[XYWZ] - E[W]E[XYZ] - E[YWZ[E[X] + E[W]E[X]E[YZ] \\
&\quad = \mathrm{cum}(X,Y,W,Z) + \mathrm{cum}(X,Y,W)\mathrm{cum}(Z) + \mathrm{cum}(X,W,Z)\mathrm{cum}(Y) \\
&\qquad + \mathrm{cum}(X,Y)\mathrm{cum}(W,Z) + \mathrm{cum}(X,W)\mathrm{cum}(Y,Z) + \mathrm{cum}(X,Z)\mathrm{cum}(Y,W),
\end{aligned}$$

where X, Y, W, Z are products of random elements of H. Hence, by the product theorem for cumulants, only those products of cumulants have to be considered that lead to indecomposable partitions of the matrix below

or of any sub-matrix (with the same column structure) with the exception that (Y) or (Z) is allowed to be decomposable but not within the same partition.

$$\begin{array}{ll} (X) & D^{(l_1)}_{\omega_{j_1}} \quad D^{(l_2)}_{-\omega_{j_1+h}} \\ (Y) & D^{(l_1)}_{\omega_{j_1-j_1'}} \quad D^{(l_1)}_{\omega_{j_1'-j_1}} \\ (W) & D^{(l_1)}_{-\omega_{j_2}} \quad D^{(l_2)}_{\omega_{j_2+h}} \\ (Z) & D^{(l_1)}_{\omega_{j_2'-j_2}} \quad D^{(l_1)}_{\omega_{j_2-j_2'}} \end{array} \tag{S2.8}$$

That is, the order of the error belonging to those partitions has to be investigated for which the cumulant terms that contain an element of the sets $\{D^{(l_1)}_{\omega_{j_1}}, D^{(l_2)}_{-\omega_{j_1+h}}\}$ and of $\{D^{(l_1)}_{-\omega_{j_2}} D^{(l_2)}_{\omega_{j_2+h}}\}$ also contain at least one element from another set. Because the process has zero-mean, it suffices consider partitions for which $m_i \geqslant 2$. By Lemma A.1 of the main paper, a cumulant of order k upscaled by order T will be of order $O(T^{-k/2+1})$ under H_0 and $O(T^{-k/2+2})$ under the alternative. This directly implies that only terms of the following form have to be investigated:

$$\frac{1}{T}\left(\frac{2\pi}{bT}\right)^2 \sum_{j_1,j_2=1}^{T} \sum_{j_1',j_2'=1}^{T} K\left(\frac{\omega_{j_1'}}{b}\right) K\left(\frac{\omega_{j_2'}}{b}\right) \mathrm{cum}_4 \mathrm{cum}_2 \mathrm{cum}_2, \tag{S2.9}$$

$$\frac{1}{T}\left(\frac{2\pi}{bT}\right)^2 \sum_{j_1,j_2=1}^{T} \sum_{j_1',j_2'=1}^{T} K\left(\frac{\omega_{j_1'}}{b}\right) K\left(\frac{\omega_{j_2'}}{b}\right) \mathrm{cum}_3 \mathrm{cum}_3 \mathrm{cum}_2, \tag{S2.10}$$

$$\frac{1}{T}\left(\frac{2\pi}{bT}\right)^2 \sum_{j_1,j_2=1}^{T} \sum_{j_1',j_2'=1}^{T} K\left(\frac{\omega_{j_1'}}{b}\right) K\left(\frac{\omega_{j_2'}}{b}\right) \mathrm{cum}_2 \mathrm{cum}_2 \mathrm{cum}_2 \mathrm{cum}_2. \tag{S2.11}$$

However, for a fixed partition $P = \{P_1, \ldots, P_M\}$,

$$T \prod_{j=1}^{M} O\left(\frac{1}{T^{m_j/2-1}}\right),$$

from which it is also clear that (S2.9) and (S2.10) are at most going to be of order $O(1)$ under the alternative and of lower order under H_0. The term (S2.11) could possibly of order $O(T)$. Further analysis can therefore be restricted to partitions of this form. As mentioned above, partitions in which a term contains an element of the sets $\{D^{(l_1)}_{\omega_{j_1}}, D^{(l_2)}_{-\omega_{j_1+h}}\}$ and of $\{D^{(l_1)}_{-\omega_{j_2}} D^{(l_2)}_{\omega_{j_2+h}}\}$ must contain at least one element from another set. It follows that partitions in which $m_i = 2$ for all $i \in \{1, \ldots, M\}$, without any restrictions on the summations, are decomposable. We find the term of highest order is thus of the type

$$\begin{aligned} &\frac{1}{T} \sum_{j_1,j_2=1}^{T} g^{(l_1,l_2)}_{j_1,j_2} \left(\frac{2\pi}{bT}\right)^2 \sum_{j_1',j_2'=1}^{T} K\left(\frac{\omega_{j_1'}}{b}\right) K\left(\frac{\omega_{j_2'}}{b}\right) \mathrm{cum}\big(D^{(l_2)}_{-\omega_{j_1+h}}, D^{(l_2)}_{\omega_{j_2+h}}\big) \mathrm{cum}\big(D^{(l_1)}_{-\omega_{j_2}}, D^{(l_1)}_{\omega_{j_1}}\big) \\ &\quad \times \mathrm{cum}\big(D^{(l_1)}_{\omega_{j_1-j_1'}} D^{(l_1)}_{\omega_{j_1-j_1'}}\big) \mathrm{cum}\big(D^{(l_1)}_{\omega_{j_2'-j_2}} D^{(l_1)}_{\omega_{j_2-j_2'}}\big) = \sup_j g_j^{2(l_1,l_2)} O(TT^{-1}) = O(1), \end{aligned}$$

under the null of stationarity and under the alternative, where the error is uniform in ω. The bound in case of the alternative follows from Corollary A.1 of the main paper. The result follows since, by the positive definiteness of the spectral density operators, $\sup_j |g_j^{(l_1,l_2)}| < \infty$ for all $l_1, l_2 \in \mathbb{N}$. □

Lemma S2.2. *Let* $(g_j^{(l_1,l_2)} : j \in \mathbb{Z})$ *be a bounded sequence in* $\mathbb{C}$ *for all* $l_1, l_2 \in \mathbb{N}$ *such that* $\inf_{l_1,l_2} g_j^{(l_1,l_2)} > 0$. *Then,*

$$E\bigg[\bigg|\frac{1}{\sqrt{T}}\sum_{j=1}^{T} g_j^{(l_1,l_2)}\big(D_{\omega_j}^{(l_1)}D_{-\omega_{j+h}}^{(l_2)} - \mathbb{E}\big[D_{\omega_j}^{(l_1)}D_{-\omega_{j+h}}^{(l_2)}\big]\big)\bigg|^2\bigg] = \begin{cases} O\Big(\frac{1}{T}\Big) & \text{under } H_0. \\ O(1) & \text{under } H_1. \end{cases}$$

Proof. Notice that

$$\begin{aligned}
&E\bigg[\bigg|\frac{1}{\sqrt{T}}\sum_{j=1}^{T} g_j^{(l_1,l_2)}\big(D_{\omega_j}^{(l_1)}D_{-\omega_{j+h}}^{(l_2)} - \mathbb{E}\big[D_{\omega_j}^{(l_1)}D_{-\omega_{j+h}}^{(l_2)}\big]\big)\bigg|^2\bigg] \\
&= \frac{1}{T}\sum_{j_1,j_2=1}^{T}\Big(g_{j_1,j_2}^{(l_1,l_2)}\mathbb{E}\big[D_{\omega_{j_1}}^{(l_1)}D_{-\omega_{j_1+h}}^{(l_2)}D_{-\omega_{j_2}}^{(l_1)}D_{\omega_{j_2+h}}^{(l_2)}\big] - \mathbb{E}\big[D_{\omega_{j_1}}^{(l_1)}D_{-\omega_{j_1+h}}^{(l_2)}\big]\mathbb{E}\big[D_{-\omega_{j_2}}^{(l_1)}D_{\omega_{j_2+h}}^{(l_2)}\big]\Big) \\
&= \frac{1}{T}\sum_{j_1,j_2=1}^{T}\Big(g_{j_1,j_2}^{(l_1,l_2)}\mathrm{cum}\big(D_{\omega_{j_1}}^{(l_1)}, D_{-\omega_{j_1+h}}^{(l_2)}, D_{-\omega_{j_2}}^{(l_1)}D_{\omega_{j_2+h}}^{(l_2)}\big) + \\
&\quad \mathrm{cum}\big(D_{\omega_{j_1}}^{(l_1)}, D_{-\omega_{j_2}}^{(l_1)}\big)\mathrm{cum}\big(D_{-\omega_{j_1+h}}^{(l_2)}, D_{\omega_{j_2+h}}^{(l_2)}\big) + \mathrm{cum}\big(D_{\omega_{j_1}}^{(l_1)}, D_{\omega_{j_2+h}}^{(l_2)}\big)\mathrm{cum}\big(D_{-\omega_{j_1+h}}^{(l_2)}, D_{-\omega_{j_2}}^{(l_1)}\big)\Big).
\end{aligned}$$

Under the null, this is therefore of the order $O(T/T^2 + 1/T) = O(1/T)$, where is uniform over ω. Under the alternative, by Corrolary A.1 of the main paper, the last term can be estimated by

$$\begin{aligned}
&\frac{1}{T}\sum_{j_1,j_2=1}^{T} g_{j_1,j_2}^{(l_1,l_2)}\bigg(\frac{1}{T}G_{\omega_{j_1},-\omega_{j_1+h},-\omega_{j_2}}^{(l_1\,l_2,l_1\,l_2)} + \tilde{\mathcal{F}}_{j_1-j_2;\omega_{j_1}}^{(l_1,l_1)}\tilde{\mathcal{F}}_{-j_1+j_2;-\omega_{j_1+h}}^{(l_2,l_2)} \\
&\qquad + \tilde{\mathcal{F}}_{j_1+j_2+h;\omega_{j_1}}^{(l_1,l_2)}\tilde{\mathcal{F}}_{-j_1-j_2-h;-\omega_{j_1+h}}^{(l_2,l_1)} + O\Big(\frac{1}{T}\Big)\bigg) \\
&\leqslant \sup_j |g_j^{(l_1,l_2)}|^2\frac{1}{T}\sum_{j_1,j_2=1}^{T}\bigg(\frac{1}{T}G_{\omega_{j_1},-\omega_{j_1+h},-\omega_{j_2}}^{(l_1\,l_2,l_1\,l_2)} + \tilde{\mathcal{F}}_{j_1-j_2;\omega_{j_1}}^{(l_1,l_1)}\tilde{\mathcal{F}}_{-j_1+j_2;-\omega_{j_1+h}}^{(l_2,l_2)} \\
&\qquad + \tilde{\mathcal{F}}_{j_1+j_2+h;\omega_{j_1}}^{(l_1,l_2)}\tilde{\mathcal{F}}_{-j_1-j_2-h;-\omega_{j_1+h}}^{(l_2,l_1)} + O\Big(\frac{1}{T}\Big)\bigg) \\
&\leqslant \sup_j |g_j^{(l_1,l_2)}|^2\bigg(\sum_{t_1,t_2,t_3}\|\kappa_{4;t_1,t_2,t_3}\|_2\|\psi_{l_1}\|_2^2|\psi_{l_2}\|_2^2 + C\sum_t\|\kappa_{2;t}\|_2\bigg) = O(1),
\end{aligned}$$

for some constant C. □

Lemma S2.3. *Let* $(g_j^{(l_1,l_2)} : j \in \mathbb{Z})$ *be a bounded sequence in* $\mathbb{C}$ *for all* $l_1, l_2 \in \mathbb{N}$ *such that* $\inf_{l_1,l_2} g_j^{(l_1,l_2)} > 0$. *Then,*

$$\begin{aligned}
&\mathbb{E}\bigg[\bigg|\frac{1}{\sqrt{T}}\sum_{j=1}^{T} g_j^{(l_1,l_2)}\mathbb{E}\big[D_{\omega_j}^{(l_1)}D_{-\omega_{j+h}}^{(l_2)}\big]\Big(\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)}\hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)} - \mathbb{E}\big[\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)}\hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)}\big]\Big)\bigg|\bigg] \\
&\qquad = \begin{cases} O\Big(\frac{1}{\sqrt{b}T}\Big) & \textit{under Assumption 4.1.} \\ O\Big(\frac{1}{\sqrt{bT}}\Big) & \textit{under Assumption 4.3.} \end{cases}
\end{aligned}$$

Proof. Observe first that by the Cauchy–Schwarz inequality,

$$\mathbb{E}[|J_3(l_1,l_2)|] \leqslant \sup_j |g_j^{(l_1,l_2)} \mathbb{E}\big[D_{\omega_j}^{(l_1)} D_{-\omega_{j+h}}^{(l_2)}\big]| \\ \times \left(\mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\big(\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)}\hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)} - \mathbb{E}\big[\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)}\hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)}\big]\big)\right|^2\right]\right)^{1/2},$$

which follows because the term over which the supremum is taken is deterministic. In particular, it is of order $O(T^{-1})$ under the null and $O(h^{-2})$ under the alternative. To find a bound on

$$\mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\big(\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)}\hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)} - \mathbb{E}\big[\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)}\hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)}\big]\big)\right|^2\right],$$

we proceed similarly as in the proof of Lemma S2.1. Observe that,

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\big(\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)}\hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)} - \mathbb{E}\hat{\mathcal{F}}_{\omega_j}^{(l_1,l_1)}\hat{\mathcal{F}}_{\omega_{j+h}}^{(l_2,l_2)}\big)\right|^2\right] \\
=&\frac{1}{T}\left(\frac{2\pi}{bT}\right)^2 \sum_{j_1,j_2=1}^{T}\sum_{j_1',j_2',j_3',j_4'=1}^{T}\prod_{i=1}^{4}K\left(\frac{\omega_{j_i'}}{b}\right) \\
&\times \mathbb{E}\Big[\big(D_{\omega_{j_1-j_1'}}^{(l_1)} D_{\omega_{j_1'-j_1}}^{(l_1)} D_{\omega_{j_1+h-j_2'}}^{(l_2)} D_{\omega_{j_2'-j_1-h}}^{(l_2)} - E[D_{\omega_{j_1-j_1'}}^{(l_1)} D_{\omega_{j_1'-j_1}}^{(l_1)} D_{\omega_{j_1+h-j_2'}}^{(l_2)} D_{\omega_{j_2'-j_1-h}}^{(l_2)}]\big) \\
&\times \big(D_{\omega_{j_3'-j_2}}^{(l_1)} D_{\omega_{j_2-j_3'}}^{(l_1)} D_{\omega_{j_4'-j_2-h}}^{(l_2)} D_{\omega_{j_2+h-j_4'}}^{(l_2)} - \mathbb{E}[D_{\omega_{j_3'-j_2}}^{(l_1)} D_{\omega_{j_2-j_3'}}^{(l_1)} D_{\omega_{j_4'-j_2-h}}^{(l_2)} D_{\omega_{j_2+h-j_4'}}^{(l_2)}]\big)\Big].
\end{aligned}$$

Write

$$\mathbb{E}[(X-\mathbb{E}X)][(Y-EY)] = \mathrm{cum}(X,Y) - \mathrm{cum}(X)\mathrm{cum}(Y)$$

for products X, W of random elements of H. When expanding this in terms of cumulants, we only have to consider those products of cumulants that lead to indecomposable partitions of the rows of the matrix below

$$\begin{array}{llllll}
(X) & D_{\omega_{j_1-j_1'}}^{(l_1)} & D_{\omega_{j_1'-j_1}}^{(l_1)} & D_{\omega_{j_1+h-j_2'}}^{(l_2)} & D_{\omega_{j_2'-j_1-h}}^{(l_2)} \\
(Y) & D_{\omega_{j_3'-j_2}}^{(l_1)} & D_{\omega_{j_2-j_3'}}^{(l_1)} & D_{\omega_{j_4'-j_2-h}}^{(l_2)} & D_{\omega_{j_2+h-j_4'}}^{(l_2)}
\end{array} \tag{S2.12}$$

In order to satisfy this, in every partition there must be at least one term that contains both an element of X and of Y. A similar reasoning as in the proof of S2.1 indicates we will only have to consider partitions where $2 \leqslant m_i \leqslant 4$ for $i = 1, \ldots, M$. In case of stationarity we only have to consider those with $m_i = 2$ for all $i = 1, \ldots, M$. In both cases at least one restriction in terms of the summation must occur in order for the partition to be decomposable. In particular, it can be verified that the partition of highest order is of the form

$$\begin{aligned}
&\frac{1}{T}\sum_{j_1,j_2=1}^{T}\left(\frac{2\pi}{bT}\right)^4 \sum_{j_1',j_2',j_3',j_4'=1}^{T}\prod_{i=1}^{4}K\left(\frac{\omega_{j_i'}}{b}\right)\mathrm{cum}(D_{\omega_{j_4'-j_2-h}}^{(l_2)} D_{\omega_{j_2+h-j_4'}}^{(l_2)})\mathrm{cum}(D_{\omega_{j_1+h-j_2'}}^{(l_2)} D_{\omega_{j_2'-j_1-h}}^{(l_2)}) \\
&\qquad \times \mathrm{cum}(D_{\omega_{j_1-j_1'}}^{(l_1)}, D_{\omega_{j_3'-j_2}}^{(l_1)})\mathrm{cum}(D_{\omega_{j_1'-j_1}}^{(l_1)}, D_{\omega_{j_2-j_3'}}^{(l_1)})
\end{aligned}$$

$$= \frac{1}{T}\sum_{j_1,j_2=1}^{T}\left(\frac{2\pi}{bT}\right)^4 \sum_{j_1',j_2',j_3',j_4'=1}^{T}\prod_{i=1}^{4}K\left(\frac{\omega_{j_i'}}{b}\right)\left[\tilde{\mathcal{F}}^{(l_2,l_2)}_{0;\omega_{-j_2+h-j_4'}}+O\left(\frac{1}{T}\right)\right]\left[\tilde{\mathcal{F}}^{(l_2,l_2)}_{0;\omega_{j_1-j_2'+h}}O\left(\frac{1}{T}\right)\right]$$
$$\times\left[\tilde{\mathcal{F}}^{(l_1,l_1)}_{j_1-j_2-j_1'+j_3';\omega_{j_1-j_1'}}+O\left(\frac{1}{T}\right)\right]\left[\tilde{\mathcal{F}}^{(l_1,l_1)}_{j_2-j_1-j_3'+j_1';\omega_{j_1'-j_1}}+O\left(\frac{1}{T}\right)\right]$$
$$\leqslant C\frac{1}{bT}\sup_{\omega}|G^{2(l_2,l_2)}_{\omega}|\frac{1}{T}\sum_{j_1,j_2=1}^{T}\left[\tilde{\mathcal{F}}^{(l_1,l_1)}_{j_1-j_2;\omega_{j_1-j_1'}}+O\left(\frac{1}{T}\right)\right]\left[\tilde{\mathcal{F}}^{(l_1,l_1)}_{j_2-j_1;\omega_{j_1'-j_1}}+O\left(\frac{1}{T}\right)\right]=O\left(\frac{1}{bT}\right),$$

since $\|\tilde{\mathcal{F}}_{j_1-j_2-j_1'+j_3';\omega_{j_1-j_1'}}\|_2 \leqslant C|j_1-j_2-j_1'+j_3'|^{-2}$ and $\|K(x/b)\|_\infty = O(1)$, where the bandwidth leads to only bT nonzero terms in the summation over j_1. The same bound can be shown to hold under stationarity. As before, the error is uniform with respect to ω which follows again from Corollary A.1 of the main paper. □

Lemma S2.4. *Let $(g_j^{(l_1,l_2)}: j\in\mathbb{Z})$ be a bounded sequence in $\mathbb{C}$ for all $l_1,l_2\in\mathbb{N}$ such that $\inf_{l_1,l_2} g_j^{(l_1,l_2)}>0$. Then,*

$$\left|\frac{1}{\sqrt{T}}\sum_{j=1}^{T}g_j^{(l_1,l_2)}\mathbb{E}\big[D^{(l_1)}_{\omega_j}D^{(l_2)}_{-\omega_{j+h}}\big]\big(\mathbb{E}\big[\hat{\mathcal{F}}^{(l_1,l_1)}_{\omega_j}\hat{\mathcal{F}}^{(l_2,l_2)}_{\omega_{j+h}}\big]-\mathcal{F}^{(l_1,l_1)}_{\omega_j}\mathcal{F}^{(l_2,l_2)}_{\omega_{j+h}}\big)\right|$$
$$=\begin{cases}O\left(b^2+\dfrac{1}{bT}\right) & \textit{under Assumption 4.1.}\\ O\left(\sqrt{T}b^2+\dfrac{1}{b\sqrt{T}}\right) & \textit{under Assumption 4.3.}\end{cases}$$

Proof. First note that

$$\left|\frac{1}{\sqrt{T}}\sum_{j=1}^{T}g_j^{(l_1,l_2)}\mathbb{E}\big[D^{(l_1)}_{\omega_j}D^{(l_2)}_{-\omega_{j+h}}\big]\right|=\begin{cases}O\left(\dfrac{1}{\sqrt{T}}\right) & \text{under Assumption 4.1.}\\ O(\sqrt{T}) & \text{under Assumption 4.3.}\end{cases}$$

Observe next that

$$\mathbb{E}\big[\hat{\mathcal{F}}^{(l_1,l_1)}_{\omega_j}\hat{\mathcal{F}}^{(l_2,l_2)}_{\omega_{j+h}}\big]-\mathcal{F}^{(l_1,l_1)}_{\omega_j}\mathcal{F}^{(l_2,l_2)}_{\omega_{j+h}}$$
$$=\left(\frac{2\pi}{bT}\right)^2\sum_{j_1,j_2=1}^{T}K\left(\frac{\omega_{j_1}}{b}\right)K\left(\frac{\omega_{j_2}}{b}\right)\mathbb{E}\big[D^{(l_1)}_{\omega_{j-j_1}}D^{(l_1)}_{\omega_{j_1-j}}D^{(l_2)}_{\omega_{j+h-j_2}}D^{(l_2)}_{\omega_{j_2-j-h}}\big]-\mathcal{F}^{(l_1,l_1)}_{\omega_j}\mathcal{F}^{(l_2,l_2)}_{\omega_{j+h}}$$
$$=\left(\frac{2\pi}{bT}\right)^2\sum_{j_1,j_2=1}^{T}K\left(\frac{\omega_{j_1}}{b}\right)K\left(\frac{\omega_{j_2}}{b}\right)\Big(\mathrm{cum}(D^{(l_1)}_{\omega_{j-j_1}},D^{(l_2)}_{\omega_{j+h-j_2}})\mathrm{cum}(D^{(l_1)}_{\omega_{j_1-j}},D^{(l_2)}_{\omega_{j_2-j-h}})$$
$$+\mathrm{cum}(D^{(l_1)}_{\omega_{j-j_1}},D^{(l_2)}_{\omega_{j_2-j-h}})\mathrm{cum}(D^{(l_1)}_{\omega_{j_1-j}},D^{(l_2)}_{\omega_{j+h-j_2}})\Big)+O\left(b^2+\frac{1}{bT}\right).$$

Here, it was used that $\mathbb{E}[|\hat{\mathcal{F}}^{(l_1,l_1)}_{\omega_j}-\mathcal{F}^{(l_1,l_1)}_{\omega_j}|]\leqslant\mathbb{E}[\|\hat{\mathcal{F}}_{\omega_j}-\mathcal{F}_{\omega_j}\|_2]\|\psi_{l_1}\|_2\|\psi_{l_1}\|_2=O(b^2+1/bT)$ under H_0. The same bound holds under the alternative, where $\mathcal{F}_\omega$ is replaced with the integrated spectrum G_ω. Under the alternative, write

$$\left(\frac{2\pi}{bT}\right)^2\sum_{j_1,j_2=1}^{T}K\left(\frac{\omega_{j_1}}{b}\right)K\left(\frac{\omega_{j_2}}{b}\right)\big(\tilde{\mathcal{F}}^{(l_1,l_2)}_{j_2-j_1-h;\omega_{j-j_1}}\tilde{\mathcal{F}}^{(l_1,l_2)}_{j_1+h-j_2;\omega_{j_1-j}}+\tilde{\mathcal{F}}^{(l_1,l_2)}_{h-j_1-j_2;\omega_{j-j_1}}\tilde{\mathcal{F}}^{(l_1,l_2)}_{j_1+j_2-h;\omega_{j_1-j}}$$

$$+ O\Big(\frac{1}{T}\Big) + O\Big(b^2 + \frac{1}{bT}\Big) = O\Big(b^2 + \frac{1}{bT}\Big)$$

where Corollary A.1 of the main paper was applied and where we used that the bandwidth leads to only bT nonzero terms in the summation. Under H_0, a similar argument shows that the term is of order $O(b^2 + 1/bT)$. The result now follows. □

S3 Convergence of finite-dimensional distributions

Theorem S3.1. *Let Lemma A.1 be satisfied for some finite $k \geqslant 3$. Then, for all $l_i, l_i' \in \mathbb{N}$ and $h_i \in \mathbb{Z}$ with $i = 1, \ldots, k$,*

$$\frac{1}{T^{k/2}} \mathrm{cum}\Big(w_{h_1}^{(T)}(\psi_{l_1 l_1'}), \ldots, w_{h_k}^{(T)}(\psi_{l_k l_k'})\Big) = o(1) \qquad (T \to \infty), \tag{S3.1}$$

where $w_h^{(T)}(\psi_{ll'}) = \langle w_h^{(T)}, \psi_{ll'} \rangle$ and $(\psi_{ll'} : l, l' \in \mathbb{N})$ an orthonormal basis of $L^2([0,1]^2, \mathbb{C})$

Proof. The proof is given in three parts, the first of which provides the outset, the second gives the arguments for the stationary case, while the third deals with the locally stationary situation.

(1) Preliminaries. Fix $\tau_1, \tau_2 \in [0,1]$ and $h = 1, \ldots, T-1$. It will be shown that the finite-dimensional distributions of $(w_h^{(T)}(\tau_1, \tau_2) : T \in \mathbb{N})$ converge to a Gaussian distribution by proving that the higher order cumulants of the terms $\sqrt{T} w_h^{(T)}(\psi_{ll'}) = \sqrt{T} \langle w_h^{(T)}, \psi_{ll'} \rangle$ vanish asymptotically. To formulate this, consider an array of the form

$$\begin{matrix} (1,1) & (1,2) \\ \vdots & \vdots \\ (k,1) & (k,2) \end{matrix} \tag{S3.2}$$

and let the value $s = ii'$ correspond to entry (i, i'). For a partition $P = \{P_1, \ldots, P_Q\}$, the elements of a set P_q will be denoted by $s_{q1}, \ldots, s_{qm_q}$ where $|P_q| = m_q$ is the corresponding number of elements in P_q. Associate with entry s the frequency index $j_s = j_{ii'} = (-1)^{i'-1}(j_i + h_i^{i'-1})$, Fourier frequency $\lambda_{j_s} = \frac{2\pi j_s}{T}$ and the basis function index $v_s = v_{ii'} = l_i^{2-i'} l_i'^{i'-1}$ for $i = 1, \ldots, k$ and $i' = 1, 2$. An application of the product theorem for cumulants yields

$$\begin{aligned} &\mathrm{cum}\Big(\sum_{j_1=1}^{T} D_{\omega_{j_1}}^{(l_1)} D_{-\omega_{j_1+h_1}}^{(l_1')}, \ldots, \sum_{j_k=1}^{T} D_{\omega_{j_k}}^{(l_k)} D_{-\omega_{j_k+h_k}}^{((l_k')} \Big) \\ &\quad = \sum_{j_1, \ldots, j_k} \sum_{i.p.} \mathrm{cum}(D_{\lambda_{j_s}}^{(v_s)} : s \in P_1) \cdots \mathrm{cum}(D_{\lambda_{j_s}}^{(v_s)} : s \in P_Q), \end{aligned}$$

where the summation extends over all indecomposable partitions $P = \{P_1, \ldots, P_Q\}$ of (S3.2). Because X_t has zero-mean, the number of elements within each set must satisfy $m_q \geqslant 2$. To ease notation, write $D_{\omega_{j_k}}^{(l)} = \langle D_{\omega_{j_k}}^{(T)}, \psi_l \rangle$ and

$$(\mathcal{F}_{t/T;\lambda_{j_s}}^{(v_s)} : s \in P_q) = \langle f_{t/T;\lambda_{j_{q1}}, \ldots, \lambda_{j_{qm_q-1}}}, \otimes_{i'=1}^{m_q} \psi_{v_{s_{qi'}}} \rangle,$$

noting that the latter quantity is well-defined. An application of Lemma A.1 of the main paper yields

$$\sum_{i.p.}\prod_{q=1}^{Q}\operatorname{cum}\big(D^{(v_s)}_{\lambda_{j_s}}\colon s\in P_q\big) \tag{S3.3}$$
$$=\sum_{i.p.}\prod_{q=1}^{Q}\bigg[\frac{(2\pi)^{m_q/2-1}}{T^{m_q/2}}\bigg(\sum_{t=0}^{T-1}\mathcal{F}^{(v_s)}_{t/T;\lambda_{j_s}}e^{-\mathrm{i}\sum_s t\lambda_{j_s}}\colon s\in P_q\bigg)+O\bigg(\frac{1}{T^{m_q/2}}\bigg)\bigg],$$

where under the null $\mathcal{F}^{(v_s)}_{t/T;\lambda_{j_s}}=\mathcal{F}^{(v_s)}_{\lambda_{j_s}}$. In the following, the proof is separated into the cases where the true process is stationary and where it is locally stationary.

(2) Proof under stationarity. Recall that $\sup_{\boldsymbol{\omega}}|||\mathcal{F}_{\omega_{j_1},\ldots,\omega_{j_{k'-1}}}|||2<\infty$ for all $k'\leqslant k$, and thus, by the Cauchy–Schwarz inequality, $\sup_{\boldsymbol{\omega}}|\mathcal{F}^{(v_s)}_{\lambda_{j_s}}|<\infty$ for $s\in P_q$ and $q=1,\ldots,Q$. Therefore,

$$\sum_{i.p.}\prod_{q=1}^{Q}\operatorname{cum}(D^{(v_s)}_{\lambda_{j_s}}\colon s\in P_q)$$
$$\leqslant\sum_{i.p.}\prod_{q=1}^{Q}\bigg[\frac{(2\pi)^{m_q/2-1}K_q}{T^{m_q/2}}\Delta^{(T)}\bigg(\sum_{s\in P_q}\lambda_{j_s})+O\bigg(\frac{1}{T^{m_q/2}}\bigg)\bigg]$$

for some constants $K_1,\ldots,K_Q$ independent of T. Due to the functions $\Delta^{(T)}$, there are Q constraints if $Q<k$ or if $Q=k$ and there exists h_{i_1} and h_{i_2} such that $h_{i_1}\neq h_{i_2}$ for $i_1,i_2\in\{1,\ldots,k\}$. On the other hand, if the size of the partition is equal to k and $h_{i_1}=h_{i_2}$ for all $i_1,i_2\in\{1,\ldots,k\}$, there are $Q-1$ constraints. This implies that

$$\frac{1}{T^{n/2}}\operatorname{cum}\bigg(\sum_{j_1=1}^{T}D^{(l_1)}_{\omega_{j_1}}D^{(l_1')}_{-\omega_{j_1+h_1}},\ldots,\sum_{j_n=1}^{T}D^{(l_k)}_{\omega_{j_k}}D^{(l_k')}_{-\omega_{j_k+h_k}}\bigg)$$
$$=O(T^{-n/2}T^{n-(Q-1)}T^{-2n/2}T^{Q})$$
$$=O(T^{-n/2+1}).$$

The cumulants of order $k\geqslant 3$ will therefore tend to 0 as $T\to\infty$.

(3) Proof under local stationarity. Write (S3.3) in terms of the Fourier coefficients as

$$\frac{1}{T^{n/2}}\sum_{j_1,\ldots,j_k=1}^{T}\sum_{i.p.}\prod_{q=1}^{Q}\operatorname{cum}\big(D^{(v_s)}_{\lambda_{k_s}}\colon s\in P_q\big)$$
$$=\frac{1}{T^{n/2}}\sum_{j_1,\ldots,j_k=1}^{T}\sum_{i.p.}\prod_{q=1}^{Q}\bigg[\frac{(2\pi)^{m_q/2-1}}{T^{m_q/2-1}}\big(\tilde{\mathcal{F}}^{(v_s)}_{\sum_s j_s;\lambda_{j_s}}\colon s\in P_q\big)+O\bigg(\frac{1}{T^{m_q/2}}\bigg)\bigg].$$

Note that, by Corollary A.1 and the Cauchy–Schwarz inequality,

$$\sum_{j=1}^{T}|\tilde{\mathcal{F}}^{(v_s)}_{\sum_s j_s;\lambda_{j_s}}|\leqslant\sup_{\omega}\sum_{j\in\mathbb{Z}}\|\tilde{\mathcal{F}}_{j;\omega}\|_2\prod_{i=1}^{m_q}\|\psi_{v_{qi}}\|_2<\infty,\qquad s\in P_q,$$

for all $q = 1, \dots, Q$. If $Q < k$ or if $Q = k$ and there are h_{i_1} and h_{i_2} such that $h_{i_1} \neq h_{i_2}$ for $i_1, i_2 \in \{1, \dots, k\}$ within the same set, then there is dependence on Q of the k sums $j_1, \dots, j_n$. On the other hand, if the size of the partition is equal to k and $h_{i_1} = h_{i_2}$ for all $i_1, i_2 = 1, \dots, k$, then there are $Q-1$ constraints on $j_1, \dots, j_n$. Thus, similar to the stationary case, it follows that the order is

$$O(T^{-k/2} T^{k-Q+1} T^{-2k/2+Q}) = O(T^{-k/2+1}),$$

hence giving the result. □

S4 Proofs under the alternative hypothesis of local stationarity

Completion of the proof of Theorem 4.6. To find the expressions for the covariance structure of $\sqrt{T}\gamma_h^{(T)}$ and its complex conjugate, use Theorem A.1 and Lemma A.2 of the main paper to write

$$\begin{aligned}
&\operatorname{Cov}\big(D^{(l_1)}_{\omega_{j_1}} D^{(l_2)}_{-\omega_{j_1+h_1}}, D^{(l_3)}_{\omega_{j_2}} D^{(l_4)}_{-\omega_{j_2+h_2}}\big) \\
&\quad = \frac{2\pi}{T} \tilde{\mathcal{F}}^{(l_1\, l_2, l_3\, l_4)}_{(-h_1+h_2;\omega_{j_1},-\omega_{j_1+h_1},-\omega_{j_2})} + O\bigg(\frac{1}{T^2}\bigg) \\
&\qquad + \bigg[\tilde{\mathcal{F}}^{(l_1,l_3)}_{(j_1-j_2;\omega_{j_1})} + O\bigg(\frac{1}{T}\bigg)\bigg]\bigg[\tilde{\mathcal{F}}^{(l_2,l_4)}_{(-j_1-h_1+j_2+h_2;-\omega_{j_1+h_1})} + O\bigg(\frac{1}{T}\bigg)\bigg] \\
&\qquad + \bigg[\tilde{\mathcal{F}}^{(l_1,l_4)}_{(j_1+j_2+h_2;\omega_{j_1})} + O\bigg(\frac{1}{T}\bigg)\bigg]\bigg[\tilde{\mathcal{F}}^{(l_2,l_3)}_{(-j_1-h_1-j_2,-\omega_{j_1+h_1})} + O\bigg(\frac{1}{T}\bigg)\bigg].
\end{aligned}$$

Thus,

$$\begin{aligned}
&\operatorname{Cov}\big(\sqrt{T}\gamma_{h_1}^{(T)}(l_1,l_2), \sqrt{T}\gamma_{h_2}^{(T)}(l_3,l_4)\big) \\
&= \frac{1}{T}\sum_{j_1,j_2=1}^{T} \bigg\{ \frac{2\pi}{T} \frac{\tilde{\mathcal{F}}^{(l_1\, l_2, l_3\, l_4)}_{(-h_1+h_2;\omega_{j_1},-\omega_{j_1+h_1},-\omega_{j_2})}}{(G^{(l_1,l_1)}_{\omega_{j_1}} G^{(l_2,l_2)}_{-\omega_{j_1+h_1}} G^{(l_3,l_3)}_{-\omega_{j_2}} G^{(l_4,l_4)}_{\omega_{j_2+h_2}})^{1/2}} + \frac{\tilde{\mathcal{F}}^{(l_1,l_3)}_{(j_1-j_2;\omega_{j_1})} \tilde{\mathcal{F}}^{(l_2,l_4)}_{(-j_1-h_1+j_2+h_2;-\omega_{j_1+h_1})}}{(G^{(l_1,l_1)}_{\omega_{j_1}} G^{(l_3,l_3)}_{-\omega_{j_2}} G^{(l_2,l_2)}_{-\omega_{j_1+h_1}} G^{(l_4,l_4)}_{\omega_{j_2+h_2}})^{1/2}} \\
&\qquad + \frac{\tilde{\mathcal{F}}^{(l_1,l_4)}_{(j_1+j_2+h_2;\omega_{j_1})} \tilde{\mathcal{F}}^{(l_2,l_3)}_{(-j_1-h_1-j_2,-\omega_{j_1+h_1})}}{(G^{(l_1,l_1)}_{\omega_{j_1}} G^{(l_4,l_4)}_{\omega_{j_2+h_2}} G^{(l_2,l_2)}_{-\omega_{j_1+h_1}} G^{(l_3,l_3)}_{-\omega_{j_2}})^{1/2}} \\
&\qquad + O\bigg(\frac{1}{T}\bigg[\frac{\tilde{\mathcal{F}}^{(l_1,l_3)}_{(j_1-j_2;\omega_{j_1})}}{(G^{(l_1,l_1)}_{\omega_{j_1}} G^{(l_3,l_3)}_{-\omega_{j_2}})^{1/2}} + \frac{\tilde{\mathcal{F}}^{(l_2,l_4)}_{(-j_1-h_1+j_2+h_2;-\omega_{j_1+h_1})}}{(G^{(l_2,l_2)}_{-\omega_{j_1+h_1}} G^{(l_4,l_4)}_{\omega_{j_2+h_2}})^{1/2}} \\
&\qquad\quad + \frac{\tilde{\mathcal{F}}^{(l_1,l_4)}_{(j_1+j_2+h_2;\omega_{j_1})}}{(G^{(l_1,l_1)}_{\omega_{j_1}} G^{(l_4,l_4)}_{\omega_{j_2+h_2}})^{1/2}} + \frac{\tilde{\mathcal{F}}^{(l_2,l_3)}_{(-j_1-h_1-j_2,-\omega_{j_1+h_1})}}{(G^{(l_2,l_2)}_{-\omega_{j_1+h_1}} G^{(l_3,l_3)}_{-\omega_{j_2}})^{1/2}}\bigg] + \frac{1}{T^2}\bigg)\bigg\}.
\end{aligned}$$

By Corollary A.1 (ii), this equals

$$\begin{aligned}
\Sigma^{(T)}_{h_1,h_2}(\boldsymbol{l}_4) &= T \operatorname{Cov}\big(\gamma_{h_1}^{(T)}(l_1,l_2), \gamma_{h_2}^{(T)}(l_3,l_4)\big) \qquad \text{(S4.1)} \\
&= \frac{1}{T}\sum_{j_1,j_2=1}^{T} \mathcal{G}^{(l_1,l_2,l_3,l_4)}_{j_1,j_2}\Big(\tilde{\mathcal{F}}^{(l_1,l_3)}_{(j_1-j_2;\omega_{j_1})} \tilde{\mathcal{F}}^{(l_2,l_4)}_{(-j_1-h_1+j_2+h_2;-\omega_{j_1+h_1})} \\
&\qquad + \tilde{\mathcal{F}}^{(l_1,l_4)}_{(j_1+j_2+h_2;\omega_{j_1})} \tilde{\mathcal{F}}^{(l_2,l_3)}_{(-j_1-h_1-j_2,-\omega_{j_1+h_1})} + \frac{2\pi}{T} \tilde{\mathcal{F}}^{(l_1\, l_2, l_3\, l_4)}_{(-h_1+h_2;\omega_{j_1},-\omega_{j_1+h_1},-\omega_{j_2})}\Big)
\end{aligned}$$

$$+ O\Big(\frac{1}{T}\Big),$$

where $\mathcal{G}_{j_1,j_2}^{(l_1,l_2,l_3,l_4)} = (G_{\omega_{j_1}}^{(l_1,l_1)} G_{\omega_{j_1+h_1}}^{(l_2,l_2)} G_{\omega_{j_2}}^{(l_3,l_3)} G_{\omega_{j_2+h_2}}^{(l_4,l_4)})^{-1/2}$. Similarly,

$$\begin{aligned}
\acute{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4) &= T\,\mathrm{Cov}\big(\gamma_{h_1}^{(T)}(l_1,l_2), \overline{\gamma_{h_2}^{(T)}(l_3,l_4)}\big) \qquad \text{(S4.2)}\\
&= \frac{1}{T}\sum_{j_1,j_2=1}^{T} \mathcal{G}_{j_1,j_2}^{(l_1,l_2,l_3,l_4)} \Big(\tilde{\mathcal{F}}_{(j_1+j_2;\omega_{j_1})}^{(l_1,l_3)} \tilde{\mathcal{F}}_{(-j_1-h_1-j_2-h_2;-\omega_{j_1+h_1})}^{(l_2,l_4)} \\
&\quad + \tilde{\mathcal{F}}_{(j_1-j_2-h_2;\omega_{j_1})}^{(l_1,l_4)} \tilde{\mathcal{F}}_{(-j_1-h_1+j_2;-\omega_{j_1+h_1})}^{(l_2,l_3)} + \frac{2\pi}{T} \tilde{\mathcal{F}}_{(-h_1-h_2;\omega_{j_1},-\omega_{j_1+h_1},\omega_{j_2})}^{(l_1\,l_2,l_3\,l_4)} \Big) \\
&\quad + O\Big(\frac{1}{T}\Big),\\
\bar{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4) &= T\,\mathrm{Cov}\big(\overline{\gamma_{h_1}^{(T)}(l_1,l_2)}, \overline{\gamma_{h_2}^{(T)}(l_3,l_4)}\big) \qquad \text{(S4.3)}\\
&= \frac{1}{T}\sum_{j_1,j_2=1}^{T} \mathcal{G}_{j_1,j_2}^{(l_1,l_2,l_3,l_4)} \Big(\tilde{\mathcal{F}}_{(-j_1+j_2;-\omega_{j_1})}^{(l_1,l_3)} \tilde{\mathcal{F}}_{(j_1+h_1-j_2-h_2;\omega_{j_1+h_1})}^{(l_2,l_4)} \\
&\quad + \tilde{\mathcal{F}}_{(-j_1-j_2-h_2;-\omega_{j_1})}^{(l_1,l_4)} \tilde{\mathcal{F}}_{(j_1+h_1+j_2;\omega_{j_1+h_1})}^{(l_2,l_3)} + \frac{2\pi}{T} \tilde{\mathcal{F}}_{(h_1-h_2;-\omega_{j_1},\omega_{j_1+h_1},\omega_{j_2})}^{(l_1\,l_2,l_3\,l_4)} \Big) \\
&\quad + O\Big(\frac{1}{T}\Big),
\end{aligned}$$

and

$$\begin{aligned}
\grave{\Sigma}_{h_1,h_2}^{(T)}(\boldsymbol{l}_4) &= T\,\mathrm{Cov}\big(\overline{\gamma_{h_1}^{(T)}(l_1,l_2)}, \gamma_{h_2}^{(T)}(l_3,l_4)\big) \qquad \text{(S4.4)}\\
&= \frac{1}{T}\sum_{j_1,j_2=1}^{T} \Big(\mathcal{G}_{j_1,j_2}^{(l_1,l_2,l_3,l_4)} \tilde{\mathcal{F}}_{(-j_1-j_2;-\omega_{j_1})}^{(l_1,l_3)} \tilde{\mathcal{F}}_{(j_1+h_1+j_2+h_2;\omega_{j_1+h_1})}^{(l_2,l_4)} \\
&\quad + \tilde{\mathcal{F}}_{(-j_1+j_2+h_2;-\omega_{j_1})}^{(l_1,l_4)} \tilde{\mathcal{F}}_{(j_1+h_1-j_2;\omega_{j_1+h_1})}^{(l_2,l_3)} + \frac{2\pi}{T} \tilde{\mathcal{F}}_{(h_1+h_2;-\omega_{j_1},\omega_{j_1+h_1},-\omega_{j_2})}^{(l_1\,l_2,l_3\,l_4)} \Big) \\
&\quad + O\Big(\frac{1}{T}\Big).
\end{aligned}$$

This completes the proof. □

Similarly, we find for the covariance structure of Theorem E.3:

$$\begin{aligned}
\Upsilon_{h_1,h_2}(\psi_{l_1 l_1'\, l_2 l_2'}) = \lim_{T\to\infty} \frac{1}{T} \sum_{j_1,j_2=1}^{T} \Big(&\big\langle \tilde{\mathcal{F}}_{j_1-j_2;\omega_{j_1}}(\psi_{l_2}), \psi_{l_1} \big\rangle \big\langle \tilde{\mathcal{F}}_{-j_1-h_1+j_2+h_2;-\omega_{j_1+h_1}}(\psi_{l_2'}), \psi_{l_1'} \big\rangle \\
&+ \big\langle \tilde{\mathcal{F}}_{j_1+j_2+h_2;\omega_{j_1}}(\psi_{l_2'}), \psi_{l_1} \big\rangle \big\langle \tilde{\mathcal{F}}_{-j_1-h_1-j_2,-\omega_{j_1+h_1}}(\psi_{l_2}), \psi_{l_1'} \big\rangle \\
&+ \frac{2\pi}{T} \big\langle \tilde{\mathcal{F}}_{(-h_1+h_2;\omega_{j_1},-\omega_{j_1+h_1},-\omega_{j_2})}(\psi_{l_2\,l_2'}), \psi_{l_1,l_1'} \big\rangle \Big), \qquad \text{(S4.5)}
\end{aligned}$$

$$\acute{\Upsilon}_{h_1,h_2}(\psi_{l_1 l_1' \, l_2 l_2'}) = \lim_{T\to\infty} \frac{1}{T} \sum_{j_1,j_2=1}^{T} \Big(\big\langle \tilde{\mathcal{F}}_{j_1+j_2;\omega_{j_1}}(\psi_{l_2}), \psi_{l_1} \big\rangle \big\langle \tilde{\mathcal{F}}_{-j_1-h_1-j_2-h_2;-\omega_{j_1+h_1}}(\psi_{l_2'}), \psi_{l_1'} \big\rangle$$
$$+ \big\langle \tilde{\mathcal{F}}_{j_1-j_2-h_2;\omega_{j_1}}(\psi_{l_2'}), \psi_{l_1} \big\rangle \big\langle \tilde{\mathcal{F}}_{-j_1-h_1+j_2;-\omega_{j_1+h_1}}(\psi_{l_2}), \psi_{l_1'} \big\rangle$$
$$+ \frac{2\pi}{T} \big\langle \tilde{\mathcal{F}}_{(-h_1-h_2;\omega_{j_1},-\omega_{j_1+h_1},\omega_{j_2})}(\psi_{l_2 \, l_2'}), \psi_{l_1 \, l_1'} \big\rangle \Big), \tag{S4.6}$$

$$\bar{\Upsilon}_{h_1,h_2}(\psi_{l_1 l_1' \, l_2 l_2'}) = \lim_{T\to\infty} \frac{1}{T} \sum_{j_1,j_2=1}^{T} \Big(\big\langle \tilde{\mathcal{F}}_{-j_1+j_2;-\omega_{j_1}}(\psi_{l_2}), \psi_{l_1} \big\rangle \big\langle \tilde{\mathcal{F}}_{j_1+h_1-j_2-h_2;\omega_{j_1+h_1}}(\psi_{l_2'}), \psi_{l_1'} \big\rangle$$
$$+ \big\langle \tilde{\mathcal{F}}_{-j_1-j_2-h_2;-\omega_{j_1}}(\psi_{l_2'}), \psi_{l_1} \big\rangle \big\langle \tilde{\mathcal{F}}_{j_1+h_1+j_2;\omega_{j_1+h_1}}(\psi_{l_2}), \psi_{l_1'} \big\rangle$$
$$+ \frac{2\pi}{T} \big\langle \tilde{\mathcal{F}}_{(h_1-h_2;-\omega_{j_1},\omega_{j_1+h_1},\omega_{j_2})}(\psi_{l_2 \, l_2'}), \psi_{l_1 \, l_1'} \big\rangle \Big) \tag{S4.7}$$

and

$$\hat{\Upsilon}_{h_1,h_2}(\psi_{l_1 l_1' \, l_2 l_2'}) = \lim_{T\to\infty} \frac{1}{T} \sum_{j_1,j_2=1}^{T} \Big(\big\langle \tilde{\mathcal{F}}_{-j_1-j_2;-\omega_{j_1}}(\psi_{l_2}), \psi_{l_1} \big\rangle \big\langle \tilde{\mathcal{F}}_{j_1+h_1+j_2+h_2;\omega_{j_1+h_1}}(\psi_{l_2'}), \psi_{l_1'} \big\rangle$$
$$+ \big\langle \tilde{\mathcal{F}}_{-j_1+j_2+h_2;-\omega_{j_1}}(\psi_{l_2'}), \psi_{l_1} \big\rangle \big\langle \tilde{\mathcal{F}}_{j_1+h_1-j_2;\omega_{j_1+h_1}}(\psi_{l_2}), \psi_{l_1'} \big\rangle$$
$$+ \frac{2\pi}{T} \big\langle \tilde{\mathcal{F}}_{(h_1+h_2;-\omega_{j_1},\omega_{j_1+h_1},-\omega_{j_2})}(\psi_{l_2 \, l_2'}), \psi_{l_1 \, l_1'} \big\rangle \Big) \tag{S4.8}$$

S5 Results for the fourth-order spectrum

Using a basis expansion, the expectation of the fourth-order periodogram operator can be expressed in terms the cumulants of the upscaled fDFTs. We have

$$\begin{aligned}
\mathbb{E}\big[I^{(T)}_{\omega_{j_1},\omega_{j_2},\omega_{j_3},\omega_{j_4}}\big] &= \frac{1}{(2\pi)^3 T}(2\pi T)^2 D^{(T)}_{\omega_{j_1}} \otimes D^{(T)}_{\omega_{j_2}} \otimes D^{(T)}_{\omega_{j_3}} \otimes D^{(T)}_{\omega_{j_4}} \\
&= \frac{T}{(2\pi)} \Big(\operatorname{cum}\big(D^{(T)}_{\omega_{j_1}}, \ldots, D^{(T)}_{\omega_{j_4}}\big) + \operatorname{cum}\big(D^{(T)}_{\omega_{j_1}}, D^{(T)}_{\omega_{j_2}}\big) \otimes \operatorname{cum}\big(D^{(T)}_{\omega_{j_3}}, D^{(T)}_{\omega_{j_4}}\big) \\
&\quad + S_{1324}(\operatorname{cum}\big(D^{(T)}_{\omega_{j_1}}, D^{(T)}_{\omega_{j_3}}\big) \otimes \operatorname{cum}\big(D^{(T)}_{\omega_{j_2}}, D^{(T)}_{\omega_{j_4}}\big)) \\
&\quad + S_{1423}(\operatorname{cum}\big(D^{(T)}_{\omega_{j_1}}, D^{(T)}_{\omega_{j_4}}\big))) \otimes \operatorname{cum}\big(D^{(T)}_{\omega_{j_2}}, D^{(T)}_{\omega_{j_3}}\big) \Big),
\end{aligned} \tag{S5.1}$$

where Theorem B.1 has been applied to reach the second equality. Note that, due to the inclusion of the function Φ, only those terms are to be considered for which the frequencies satisfy $j_4 = -j_1 - j_2 - j_3$ in such a way that $j_1 \neq j_2$ $j_1 \neq j_3$ and $j_2 \neq j_3$ and $j_1 \neq j_4$. For such values not contained in a proper submanifold,

the products of second-order cumulant tensors are at most of order $O(T^{-2})$ in an L^2 sense under the null hypothesis. Using Lemma B.1, it follows that, under H_0,

$$\Big\| \mathbb{E}\big[I^{(T)}_{\omega_{j_1},\omega_{j_2},\omega_{j_3},\omega_{j_4}}\big] - \frac{T}{(2\pi)}\frac{2\pi}{T^2}\mathcal{F}_{\omega_{j_1},\omega_{j_2},\omega_{j_3}} \Big\|_2 = O\Big(\frac{T}{T^2}\Big) = O\Big(\frac{1}{T}\Big)$$

and hence asymptotically unbiased. Using a standard smoothing kernel argument and a subsequent Taylor expansion, the estimator (5.4) is readily shown to satisfy

$$\big\| \mathbb{E}\big[\hat{\mathcal{F}}^{(T)}_{\omega_{j_1},\omega_{j_2},\omega_{j_3},\omega_{j_4}}\big] - \mathcal{F}_{\omega_{j_1},\omega_{j_2},\omega_{j_3}} \big\|_2 = O\Big(\frac{1}{b_4 T} + b_4^2\Big)$$

and hence

$$\big\| \mathbb{E}\int\int \hat{\mathcal{F}}^{(T)}_{\omega,-\omega+\omega_h,-\omega',\omega'+\omega'_h} d\omega d\omega' - \int\int \mathcal{F}_{\omega,-\omega+\omega_h,-\omega',\omega'+\omega'_h} d\omega d\omega' \big\|_2 = O\Big(\frac{1}{b_4 T} + b_4^2\Big).$$

Under the alternative, (S5.1) continues to hold. Again only those combinations of $j_1,\dots,j_4$ have to be considered that are on the principal manifold but not on a proper submanifold. By Lemma B.2, the first term of the decomposition in (S5.1) now yields the time-integrated fourth order spectral density operator plus an error term

$$\frac{T}{(2\pi)}\Big(\frac{(2\pi)}{T^2}\sum_{t=0}^{T-1}\mathcal{F}_{t/T;\omega_{j_1},\dots,\omega_{j_4-1}} e^{-\mathrm{i}\sum_{l=1}^{k} t\omega_{j_l}} + R_{4,T}\Big),$$

where $R_{4,T} = O(T^{-2})$ and hence multiplied by T leads to an error of $O(T^{-1})$. Additionally, there are three terms of the form

$$\begin{aligned}\frac{T}{(2\pi)}\Big(&\frac{1}{T^2}\sum_{t,t'=0}^{T-1}\mathcal{F}_{t/T,\omega_{j_1}}\otimes\mathcal{F}_{t/T,\omega_{j_3}} e^{-\mathrm{i}(t(\omega_{j_1+j_2})+t'(\omega_{j_3+j_4}))}\\ &+ R_{2,T}\frac{1}{T}\sum_{t=0}^{T-1}\mathcal{F}_{t/T,\omega_{j_1}} e^{-\mathrm{i}t(\omega_{j_1+j_2})} + R_{2,T}\frac{1}{T}\sum_{t=0}^{T-1}\mathcal{F}_{t/T,\omega_{j_1}} e^{-\mathrm{i}t(\omega_{j_3+j_4})} + R_{2,T}R_{2,T}\Big).\end{aligned}$$

Note that Φ only selects those terms for which the frequencies satisfy $j_4 = -j_1 - j_2 - j_3$ such that $j_1 \neq j_2$ $j_1 \neq j_3$ and $j_2 \neq j_3$. Let $\mathscr{R}$ be a remainder element of $S_2(H)$. The second-order terms can be written as

$$\frac{T}{(2\pi)}\Big(\frac{1}{T^2}\sum_{t,t'=0}^{T-1}\mathcal{F}_{t/T,\omega_{j_1}}\otimes\mathcal{F}_{t/T,\omega_{j_3}} e^{-\mathrm{i}(t-t')(\omega_{j_1+j_2})} + \frac{\mathscr{R}}{T}\frac{2}{|j_1+j_2|^2}\Big).$$

Therefore, consider bounding their smoothed versions

$$\begin{aligned}&\frac{(2\pi)^3}{(b_4T)^3}\sum_{k_1,k_2,k_3} K_4\Big(\frac{\omega_{j_1}-\alpha_{k_1}}{b_4},\dots,\frac{\omega_{j_4}+\sum_{i=1}^{3}\alpha_{k_i}}{b_4}\Big)\Phi(\alpha_{k_1},\dots,\alpha_{k_4})\\ &\quad\Big(\frac{1}{(2\pi)T}\sum_{t,t'=0}^{T-1}\mathcal{F}_{t/T,\alpha_{k_1}}\otimes\mathcal{F}_{t'/T,\alpha_{k_3}} e^{-\mathrm{i}(t-t')(\alpha_{k_1+k_2})} + \frac{1}{(2\pi)}\frac{2\mathscr{R}}{|k_1+k_2|^2}\Big).\end{aligned}$$

Consider the leading term first. The double sum and nonnegativity of the smoothing kernels allows to derive the upper bound

$$\frac{(2\pi)^3}{(b_4T)^3}\sum_{k_1,k_2,k_3}\Big\|K_4\Big(\frac{\omega_{j_1}-\alpha_{k_1}}{b_4},\dots,\frac{\omega_{j_4}+\sum_{i=1}^3\alpha_{k_i}}{b_4}\Big)\Phi(\boldsymbol{\alpha_k})\frac{1}{(2\pi)T}\sum_{t,t'=0}^{T-1}\mathcal{F}_{t/T,\alpha_{k_1}}\otimes\mathcal{F}_{t/T,\alpha_{k_3}}e^{-\mathrm{i}(t-t')(\alpha_{k_1+k_2})}\Big\|_2$$
$$\leqslant\frac{(2\pi)^2}{b_4^3(T)^4}\sum_{k_1,k_2,k_3}K_4\Big(\frac{\omega_{j_1}-\alpha_{k_1}}{b_4},\dots,\frac{\omega_{j_4}+\sum_{i=1}^3\alpha_{k_i}}{b_4}\Big)\Phi(\boldsymbol{\alpha_k})\Big|\sum_{t,t'=0}^{T-1}e^{-\mathrm{i}(t-t')(\alpha_{k_1+k_2})}\Big|\|\mathcal{F}_{t/T,\alpha_{k_1}}\otimes\mathcal{F}_{t/T,\alpha_{k_3}}\|_2,$$

but since submanifolds, i.e., $k_1=k_2$, are not allowed via Φ, the above can be bounded by

$$\sup_{u,\omega}\|\mathcal{F}_{u,\omega}\|_2^2\frac{(2\pi)^2}{b_4^3(T)^4}\sum_{k_1,k_2,k_3}K_4\Big(\frac{\omega_{j_1}-\alpha_{k_1}}{b_4},\dots,\frac{\omega_{j_4}+\sum_{i=1}^3\alpha_{k_i}}{b_4}\Big)\Phi(\boldsymbol{\alpha_k})\Big|\Delta_T^{(\alpha_{k_1+k_2})}\Big|^2$$
$$\leqslant\Big(\sup_{u,\omega}\|\mathcal{F}_{u,\omega}\|_2^2O(\frac{1}{b_4T})\Big)=O\Big(\frac{1}{b_4T^2}\Big).$$

Now consider the two error terms. A change of variables gives

$$\frac{(2\pi)^3}{(b_4T)^3}\sum_{k_1,k_2,k_3}K_4\Big(\frac{\omega_{k_1}}{b_4},\dots,\frac{\omega_{k_4}}{b_4}\Big)\Phi(\boldsymbol{\omega_j-\alpha_k})\frac{\mathscr{R}}{|j_1-k_1+j_2-k_2|^2}$$

and by the p-harmonic series this is of order

$$\Big\|\frac{(2\pi)^3}{(b_4T)^3}\sum_{k_2,k_3,l}K_4\Big(\frac{\omega_l}{b_4},\dots,\frac{\omega_{k_4}}{b_4}\Big)\Phi(\boldsymbol{\omega_j-\alpha_k})\frac{\mathscr{R}}{|l|^2}\Big\|_2=O\Big(\frac{1}{b_4T}\Big).$$

Therefore, a similar argument as under the null yields

$$\Big\|\mathbb{E}\Big[\int\int\hat{\mathcal{F}}^{(T)}_{\omega,-\omega+\omega_h,-\omega',\omega'+\omega_h'}d\omega d\omega'\Big]-\int\int G_{\omega,-\omega+\omega_h,-\omega',\omega'+\omega_h'}d\omega d\omega'\Big\|_2=O\Big(\frac{1}{b_4T}+b_4^2\Big),$$

where $G_{\omega,-\omega+\omega_h,-\omega',\omega'+\omega_h'}$ is the time-integrated fourth-order spectral density operator acting on $S_2(H_{\mathbb{C}})$.

References

Aue, A. & A. van Delft (2017). Testing for stationarity of functional time series in the frequency domain. Preprint.

Cremers, H. & D. Kadelka (1986). On weak convergence of integral functions of stochastic processes with applications to processes taking paths in L_p^E. *Stochastic Processes and their Applications* **21**, 305–317.

Nelson, E. (1969). *Topics in Dynamics, Volume I: Flows.* Princeton University Press.

Panaretos, V. & S. Tavakoli (2013). Fourier analysis of stationary time series in function space. *The Annals of Statistics* **41**, 568–603.