JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **40**: 665–692 (2019)
Published online 13 February 2019 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12447

ORIGINAL ARTICLE

INFERENCE FOR THE LAGGED CROSS-COVARIANCE OPERATOR BETWEEN FUNCTIONAL TIME SERIES

GREGORY RICE* AND MARCO SHUM
Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

When considering two or more time series of functional data objects, for instance those derived from densely observed intraday stock price data of several companies, the empirical cross-covariance operator is of fundamental importance due to its role in functional lagged regression and exploratory data analysis. Despite its relevance, statistical procedures for measuring the significance of such estimators are currently undeveloped. We present methodology based on a functional central limit theorem for conducting statistical inference for the cross-covariance operator estimated between two stationary, weakly dependent, functional time series. Specifically, we consider testing the null hypothesis that the two series possess a specified cross-covariance structure at a given lag. Since this test assumes that the series are jointly stationary, we also develop a change-point detection procedure to validate this assumption of independent interest. The most imposing technical hurdle in implementing the proposed tests involves estimating the spectrum of a high dimensional spectral density operator at frequency zero. We propose a simple dimension reduction procedure based on functional principal component analysis to achieve this, which is shown to perform well in a simulation study. We illustrate the proposed methodology with an application to densely observed intraday price data of stocks listed on the New York stock exchange.

Received 14 August 2017; Accepted 21 December 2018

Keywords: Functional data analysis; change point analysis
MOS subject classification: 62, Statistics.

1. INTRODUCTION

Functional time series analysis (FTSA) has grown substantially in the last decade and a half to provide methodology for functional data objects that are obtained sequentially over time. One way such data arise is when dense records of continuous time processes are segmented into collections of curves in some natural way. For example, high frequency records of pollution levels may be segmented to form daily pollution curves, or tick-by-tick asset price data may be used to construct daily intraday price or return curves; see Aue *et al.* (2014) and Kargin and Onatski (2008). Other examples include sequentially observed curves that describe physical phenomena, as arise in functional magnetic resonance imaging or in the observation of dynamic biological systems; see Aston and Kirch (2012), Tavakoli and Panaretos (2016), and Hywood *et al.* (2016). We refer the reader to Ramsay and Silverman (2005), and Ferraty and Vieu (2006) for overviews of the field of functional data analysis, and to Bosq (2000) and Hörmann and Kokoszka (2012) for an introduction to FTSA.

Most of the developments in these two fields focus on analyzing functional data obtained from a single source, for example, intraday price curves derived from a single asset, or in comparing functional data from several independent populations. To give a few examples that are related to this work, Panaretos *et al.* (2010), Paparoditis and Sapatinas (2016), and Pigoli *et al.* (2014) develop methods for performing inference for the covariance operator of functional data. Using a self-normalization approach, a two sample test for the second-order structure with functional time series data that allows for some dependence across the populations is developed in Zhang and Shao (2015).

* Correspondence to: Gregory Rice, Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada. E-mail: grice@uwaterloo.ca

However, in many situations of interest, functional data are obtained simultaneously from two or more sources, for example, intraday price curves derived from several assets. In such cases one often wishes to quantify the second-order dependence relationships between such curves, and a way of achieving this is through the empirical cross-covariance operator. Although the notion of the cross-covariance operator between random elements in a Hilbert space was put forward over forty years ago in Baker (1973), and is discussed in chapter one of Ramsay and Silverman (2005), statistical methodology for estimating and performing further inference for the cross-covariance structure between collections of curves seems to be quite new.

Measuring the cross-covariance between collections of curves has received some attention in the context of multi-variate longitudinal and functional data. Under the assumption that the given data is a simple random sample of multi-variate longitudinal data, Dubin and Müller (2005), Serban *et al.* (2013), and Zhou *et al.* (2008) develop measures of cross-covariance between longitudinal data sources, including measures based on canonical correlation analysis and PCA. PCA of multi-variate functional data is also studied in Chiou *et al.* (2014) and Chiou and Müller (2013), and in Petersen and Müller (2016) an analog of the covariance matrix for multi-variate functional data using Fréchet integration is defined. In each of these cases, the potential effect of temporal dependence among the functional units is not considered.

In the context of bivariate functional time series, the cross-covariance operator and its lagged versions are arguably of greater importance. Methods for lagged functional time series regression, which have recently been put forward in Hörmann *et al.* (2015b) and Pham and Panaretos (2017), are naturally based on the Fourier transform of the lagged cross-covariance operators. Moreover, the initial exploratory analysis of any such series would typically begin by considering the sequence of lagged cross-covariance operators to try to gain insight into the relationship between the series.

Despite the apparent utility of the cross-covariance operator of functional time series, statistical inference for it has not yet been considered, to the best of our knowledge. In Chapter 4 of the seminal monograph on functional linear processes of Bosq (2000), a central limit theorem is given for the covariance and auto-covariance operators of functional time series that may be represented as linear processes. A portmanteau-type test for independence of two functional time series is developed in Horváth and Rice (2015) based on the norms of cross-covariance operators at long lags, but their test assumes under the null hypothesis that the individual series are independent, and is hence not suitable for quantifying the significance of estimates of the cross-covariance between curves.

Additionally, when the data are obtained as bivariate functional time series, it is of interest to know if the cross-covariance structure changes during the observation period. Being able to test for such a feature (i) helps validate the assumption of joint stationarity needed to apply inferential procedures for the cross-covariance operator, (ii) is of use for determining if the regression function changes in the functional lagged regression problem, and (iii) may be of independent interest since the presence and location of such a change point may signify an important event. This problem has also not been addressed, although several authors have considered analogous problems in the context of finite dimensional time series; we refer the reader to Dette *et al.* (2015), Wied *et al.* (2012a), Wied *et al.* (2012b), and Aue *et al.* (2009).

In this article, we consider two types of hypothesis testing problems: tests for a specified cross-covariance structure between two functional time series, for example, that the two series are uncorrelated at a given lag, and change point tests for the covariance structure within a given sample. Two varieties of test statistics are proposed in each of these settings that are based on either the standard L^2 distance or dimension reduction based methods using a suitable principal component basis. These two classes of statistics possess complementary advantages, which we detail by developing the asymptotic properties of each statistic under local alternatives. The asymptotic properties of each test statistic are established assuming a general weak dependence condition similar to the one introduced in Hörmann and Kokoszka (2010), which includes nonlinear time series, as well as the majority of functional time series models studied to date, under mild regularity conditions. We also develop an application of the presented results to visualizing the lagged dependence relationship between two functional time series with an analog of a cross-correlation plot.

This methodology is primarily inspired by a basic observation of Brillinger (1975) that inference for the covariance and/or cross-covariance of time series can be made by performing inference for the mean of a suitably

DOI: 10.1111/jtsa.12447

constructed series. This idea was utilized in Himdi *et al.* (2003) to conduct non-parametric inference for the cross-covariance matrix of finite dimensional time series. Estimation and inference for the mean function of a functional time series was considered in Horváth *et al.* (2012).

The rest of the article is organized as follows. Section 2 contains the main assumptions and notation of the article, as well as asymptotic results for cross-covariance function estimators under these assumptions. The results developed in Section 2 are utilized to develop hypothesis tests in Section 3. Some details about the practical implementation of the methods developed in Section 3 are provided in Section 4, which include methods for overcoming the technical challenge of estimating the eigen-elements of a high-dimensional spectral density operator at frequency zero that arise in the limiting distribution of the test statistics, as well as the development of cross-correlation plot analogs. These testing and estimation procedures were studied by means of Monte Carlo simulation, the results of which we present in Section 5. We illustrate our methodology with an application to cumulative intraday return curves derived from the 1-minute resolution price of Microsoft and Exxon Mobil stock listed on the New York stock exchange from the year 2001 in Section 6. All technical derivations and proofs are provided in the appendices at the end of the article.

2. ASYMPTOTIC PROPERTIES OF CROSS-COVARIANCE FUNCTION ESTIMATES

Before we proceed, we introduce a bit of notation. Let $\langle\cdot,\cdot\rangle_d$ denote the standard inner product on the space $L^2[0,1]^d$ of real valued square integrable functions defined on $[0,1]^d$, and let $\|\cdot\|_d = \langle\cdot,\cdot\rangle_d^{1/2}$. We write f for the function $f(t_1,\ldots,t_d) \in L^2[0,1]^d$ when it does not cause confusion. We use the notation $\int$ to denote $\int_0^1$.

We suppose that $\{(X_i(t), Y_i(s)),\ \ t,s \in [0,1]\}_{i\in\mathbb{Z}}$ is a jointly stationary sequence of real valued stochastic processes whose sample paths are in $L^2[0,1]$ from which we observe a sample of length T, $\{(X_1(t), Y_1(s)), \ldots, (X_T(t), Y_T(s))\}$. For instance, $\{(X_i(t), Y_i(s))\}$ could denote the price of stock X and stock Y on day i at intraday times t and s normalized to the unit interval. We let $\mu_X(t) = EX_0(t)$, and $\mu_Y(s) = EY_0(s)$, and define

$$C_{XY,h}(t,s) = \mathrm{cov}(X_0(t), Y_h(s)) = E\big[(X_0(t)-\mu_X(t))(Y_h(s)-\mu_Y(s))\big],$$

to be the cross-covariance function (or kernel) between $\{X_i\}$ and $\{Y_i\}$ at lag h. $C_{XY,h}$ defines the lag h cross-covariance operator $c_{XY,h} : L^2[0,1] \to L^2[0,1]$ via

$$c_{XY,h}(f)(t) = \int C_{XY,h}(t,s) f(s)\mathrm{d}s.$$

This relationship implies that we may conduct inference for $c_{XY,h}$ by conducting inference for the function $C_{XY,h}$. Based on the sample, we may estimate $C_{XY,h}$ for $h \geq 0$ by

$$\hat{C}_{XY,h}(t,s,x) = \frac{1}{T}\sum_{i=1}^{\lfloor Tx\rfloor}(X_i(t)-\bar{X}(t))(Y_{i+h}(s)-\bar{Y}(s)),\ x \in \left[0, \frac{T-h}{T}\right],$$

which denotes the partial sample estimate of $xC_{XY,h}$ based on the first x proportion of the sample, where

$$\bar{X}(t) = \frac{1}{T}\sum_{i=1}^{T} X_i(t),\ \text{and}\ \bar{Y}(t) = \frac{1}{T}\sum_{i=1}^{T} Y_i(t).$$

The estimator can be defined similarly for $h < 0$. Under mild regularity conditions on the process $\{(X_i(t), Y_i(s))\}$, which are implied by Assumption 2.1, $\hat{C}_{XY,h}(t,s,1)$ is a consistent estimator of $C_{XY,h}(t,s)$ in $L^2[0,1]^2$. The motivation for considering now the partial sample estimates of $C_{XY,h}$ is due to our application to change point testing for the cross-covariance developed below. For now for the sake of simplicity we focus our attention on $C_{XY,0} =: C_{XY}$ and $\hat{C}_{XY,0} =: \hat{C}_{XY}$, that is, we take the lag parameter to be zero, and similar results can be established for lags other than 0.

To derive the asymptotic properties of $\hat{C}_{XY}$, we make use of the following assumption that imposes stationarity, weak dependence, and moment conditions on the functional time series.

Assumption 2.1. (a) There exists a measurable function $g_{XY} : S^\infty \to L^2[0,1] \times L^2[0,1]$, where S is a measurable space, and a sequence of i.i.d. innovations $\{\epsilon_i, : i \in \mathbb{Z}\}$ taking values in S such that $(X_i, Y_i) = g_{XY}(\epsilon_i, \epsilon_{i-1}, \ldots)$.
(b) For all $m \geq 1$, the m-dependent sequence $(X_{i,m}, Y_{i,m}) = g(\epsilon_i, \ldots, \epsilon_{i-m+1}, \epsilon^*_{i-m,m}, \epsilon^*_{i-m-1,m}, \ldots)$ with $\epsilon^*_{i,m}$ being independent copies of $\epsilon_{i,0}$, and $\{\epsilon^*_{i,m} : i \in \mathbb{Z}\}$ independent of $\{\epsilon_i : i \in \mathbb{Z}\}$, satisfies for some $p > 4$,

$$\left(E\|X_i - X_{i,m}\|_1^p\right)^{1/p} = O(m^{-\alpha}), \text{ and } \left(E\|Y_i - Y_{i,m}\|_1^p\right)^{1/p} = O(m^{-\beta})$$

where $\alpha, \beta > 1$.

Assumption 2.1(a) implies that $\{(X_i, Y_i)\}$ is a jointly stationary sequence of Bernoulli shifts in $L^2[0,1] \times L^2[0,1]$ that is driven by an underlying i.i.d. innovation sequence. The space of functional time series models contained within this class is quite large, including the functional ARMA and GARCH processes; see Bosq (2000) and Aue *et al.* (2015). Condition (b) defines a type of L^p-m-approximability condition along the lines of Hörmann and Kokoszka (2010), which can often be easily verified when a time series model for the observations is given. The rate condition on the decay of these coefficients, which is somewhat stronger than the main condition studied in Hörmann and Kokoszka (2010), is used to show that certain kernel lag-window type spectral density operator estimates defined below are consistent. These assumptions could be replaced by mixing conditions and functional versions of cummulant summability conditions as presented in Panaretos and Tavakoli (2013b) and Zhang (2016), which are more comparable to the assumptions in Himdi *et al.* (2003).

Theorem 2.1. Under Assumption 2.1, there exists a sequence of Gaussian processes, $\{\Gamma_T(t,s,x),\ \ t,s,x \in [0,1]\}_{T\in\mathbb{N}}$, defined on the same probability space as $\{(X_i, Y_i)\}$, that satisfy

$$\sup_{0\leq x\leq 1} \iint \left(\sqrt{T}\left(\hat{C}_{XY}(t,s,x) - \frac{\lfloor Tx \rfloor}{T} C_{XY}(t,s)\right) - \Gamma_T(t,s,x)\right)^2 \mathrm{d}t\mathrm{d}s = o_P(1), \tag{2.1}$$

where $E\Gamma_T(t,s,x) = 0$, and

$$\mathrm{cov}(\Gamma_T(t,s,x), \Gamma_T(u,v,y)) = \min(x,y) D(t,s,u,v),$$

where $D(t,s,u,v)$ is the long run covariance function of the sequence $\{(X_i(t) - \mu_X(t))(Y_i(s) - \mu_Y(s))\}$, namely

$$D(t,s,u,v) = \sum_{\ell=-\infty}^{\infty} \mathrm{cov}((X_0(t) - \mu_X(t))(Y_0(s) - \mu_Y(s)), (X_\ell(u) - \mu_X(u))(Y_\ell(v) - \mu_Y(v)).$$

Theorem 2.1 provides a Skorokhod–Dudley–Wichura type characterization of an invariance principle for $\hat{C}_{XY}$ that can be utilized to establish the asymptotic properties of continuous functionals of $\hat{C}_{XY}$, and we consider several such statistics in Section 3 to carry out hypothesis testing for C_{XY}. The proof of Theorem 2.1 is given in Appendix A. This result is related to Theorem 2.1 of Horváth *et al.* (2012). The function D describes the asymptotic covariance function of $\sqrt{T}\hat{C}_{XY}(t,s,1)$. D naturally defines a Hilbert-Schmidt integral operator, $\mathfrak{d} : L^2[0,1]^2 \to L^2[0,1]^2$, given by

$$\mathfrak{d}(f)(t,s) = \iint D(t,s,u,v)f(u,v)\mathrm{d}u\mathrm{d}v,$$

which is symmetric and positive definite. $\mathfrak{d}$ defines eigenfunctions φ_i in $L^2[0,1]^2$, and a non-negative sequence of eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ satisfying

$$\mathfrak{d}(\varphi_i)(t,s) = \lambda_i \varphi_i(t,s). \tag{2.2}$$

We define these quantities here as they appear in the limiting distributions and definitions for the test statistics considered below.

3. INFERENCE FOR THE CROSS-COVARIANCE FUNCTION

Theorem 2.1 points to asymptotically validated methods to measure the significance of estimates of C_{XY}. For instance, we may wish to test based on the estimate $\hat{C}_{XY}(\cdot,\cdot,1)$

$$H_{0,1}:\ C_{XY} = C_0 \ \text{ versus } H_{A,1}: C_{XY} \neq C_0,$$

where equality is understood in the $L^2[0,1]^2$ sense, and C_0 is a given function of interest. This null function might be determined from historical data, or taken to be zero to test for zero cross-covariance between X_i and Y_i at a given lag. Since the hypothesis $H_{0,1}$ is well posed only when the sequence $\{(X_i, Y_i)\}$ is, at least weakly, jointly stationary, it is also of interest to determine whether or not this assumption is valid. We frame this as a second hypothesis test of the time homogeneity of the cross-covariance against the 'at most one' change point in the cross-covariance alternative:

$$H_{0,2}:\quad C_{XY}^{(1)} = C_{XY}^{(2)} = \cdots = C_{XY}^{(T)},\ \text{versus}$$

$$H_{A,2}:\ C_1 = C_{XY}^{(1)} = \cdots = C_{XY}^{(k^*)} \neq C_{XY}^{(k^*+1)} = \cdots = C_{XY}^{(T)} = C_2,\ \text{for a } k^* = \lfloor T\theta \rfloor,\quad \theta \in (0,1),$$

where $C_{XY}^{(i)}(t,s) = \operatorname{cov}(X_i(t), Y_i(s))$. We proceed by developing test statistics for each of these hypotheses. To test $H_{0,1}$, we first consider a statistic based on the normalized squared L^2 distance of $\hat{C}_{XY}$ to C_0:

$$F_T = T\|\hat{C}_{XY}(\cdot,\cdot,1) - C_0\|_2^2.$$

Corollary 3.1. Under Assumption 2.1 and $H_{0,1}$,

$$F_T \xrightarrow{\mathcal{D}} \sum_{i=1}^{\infty} \lambda_i \mathcal{N}_i^2,\ \text{as } T \to \infty, \tag{3.1}$$

where $\{\lambda_i,\ i \geq 1\}$ are defined in (2.2), and $\{\mathcal{N}_i, i \geq 1\}$ are independent standard normal random variables.

Corollary 3.1 shows that a test of asymptotic size α of $H_{0,1}$ may be obtained by comparing F_T to the $1-\alpha$ quantile of the limit distribution given in (3.1), which depends on the unknown eigenvalues of the operator $\mathfrak{d}$. To estimate these quantiles, one can estimate a suitably large number of eigenvalues λ_i using an estimate of $\mathfrak{d}$, and then continue by using these estimates to approximate the limiting distribution via Monte Carlo simulation. The details of this implementation are discussed in Section 4, including how to obtain consistent estimates of a finite number of the λ_i's.

DOI: 10.1111/jtsa.12447

The fact that the limiting distribution of F_T is non-pivotal though encourages one to consider alternative test statistics based on projecting $\hat{C}_{XY}$ into finite dimensional subspaces of $L^2[0,1]^2$. A natural choice of the finite dimensional space is the one spanned by the eigenbasis generated by $\mathfrak{d}$. In fact, it is a fairly straightforward calculation to show that for any positive integer p, under Assumption 2.1, the inner products

$$\left\langle \sqrt{T}(\hat{C}_{XY}(\cdot,\cdot,1) - C_{XY}), \varphi_i \right\rangle_2,\ 1 \le i \le p$$

are asymptotically independent, and hence projecting into the directions of φ_i has the effect of partitioning the centered estimator $\hat{C}_{XY}$ into approximately mean zero and independent components. This is similar to the motivation provided for dynamical PCA of Brillinger (1975), which has been studied in the context of functional time series data in Hörmann *et al.* (2015a), and Panaretos and Tavakoli (2013a).

In this direction, let

$$F_{T,p} = \sum_{i=1}^{p} \frac{\left\langle \sqrt{T}(\hat{C}_{XY}(\cdot,\cdot,1) - C_0), \hat{\varphi}_i \right\rangle_2^2}{\hat{\lambda}_i},$$

where p is a user selected fixed positive integer, $\hat{\varphi}_i$, and $\hat{\lambda}_i$, $1 \le i \le p$ are consistent estimates of φ_i and λ_i, $1 \le i \le p$, that is, they satisfy

$$\max_{1\le i\le p} \|\hat{\varphi}_i - \hat{c}_i\varphi_i\|_2 = o_P(1), \text{ and } \max_{1\le i\le p} |\hat{\lambda}_i - \lambda_i| = o_P(1), \tag{3.2}$$

where $\hat{c}_i = \text{sign}(\langle \varphi_i, \hat{\varphi}_i\rangle)$. We discuss in Section 4 how to obtain such estimates of λ_i and $\varphi_i(t,s)$ under the assumption that these eigenvalues are distinct, and give some advice on how to choose p.

Corollary 3.2. Under Assumption 2.1, $\lambda_p > 0$, and $H_{0,1}$,

$$F_{T,p} \xrightarrow{\mathcal{D}} \chi^2(p), \text{ as } T \to \infty,$$

where $\chi^2(p)$ denotes a Chi-squared random variable with p degrees of freedom.

We now turn to the consistency and power properties of F_T and $F_{T,p}$. The following result shows that both statistics diverge at rate T under $H_{A,1}$. This holds generally for F_T, and also for $F_{T,p}$ so long as the difference $C_{XY} - C_0$ is not orthogonal to the first p elements of the principal component basis $\{\varphi_i,\ i \ge 1\}$.

Theorem 3.1. Under Assumption 2.1, $\lambda_p > 0$, and $H_{A,1}$,

$$\frac{F_T}{T} \xrightarrow{P} \|C_{XY} - C_0\|_2^2, \text{ and } \frac{F_{T,p}}{T} \xrightarrow{P} \sum_{i=1}^{p} \frac{\langle C_{XY} - C_0, \varphi_i\rangle_2^2}{\lambda_i}.$$

Moreover, if $C_0 = C_{0,T}$ satisfies $\left\| \sqrt{T}(C_{XY} - C_0) - C_A \right\|_2 \to 0$ as $T \to \infty$ for some element C_A of $L^2[0,1]^2$, then

$$F_T \xrightarrow{\mathcal{D}} \|C_A\|_2^2 + 2\sum_{i=1}^{\infty} \lambda_i^{1/2} \langle C_A, \varphi_i\rangle \mathcal{N}_i + \sum_{i=1}^{\infty} \lambda_i \mathcal{N}_i^2, \tag{3.3}$$

and

$$F_{T,p} \xrightarrow{\mathcal{D}} \sum_{i=1}^{p} \frac{\langle C_A, \varphi_i \rangle_2^2}{\lambda_i} + 2 \sum_{i=1}^{p} \frac{\langle C_A, \varphi_i \rangle_2 \mathcal{N}_i}{\lambda_i^{1/2}} + \sum_{i=1}^{p} \mathcal{N}_i^2.$$

This 'local-alternative' result provides some insight into the complementary strengths and weaknesses of both the norm based and dimension reduction based test statistics. Obviously if $C_{XY} - C_0$ is orthogonal to the first p principal components of $\mathfrak{d}$, then the test statistic $F_{T,p}$ is not expected to have more than trivial power. In contrast the norm based test statistic F_T is expected to have improved power over other alternatives in which $C_{XY} - C_0$ is of the same magnitude, since the second term on the right-hand side (3.3), which has mean zero, will have a smaller variance in this case. Conversely, if $C_{XY} - C_0$ is contained in the subspace spanned by the first p principal components of $\mathfrak{d}$, then $F_{T,p}$ is expected to be more powerful, since then this statistic effectively defines the likelihood ratio test of $H_{0,1}$ assuming the data is Gaussian in the p-dimensional subspace spanned by $\varphi_1, \dots, \varphi_p$.

To test $H_{0,2}$ versus $H_{A,2}$, we define analogously to F_T and $F_{T,p}$,

$$Z_T = T \sup_{0 \le x \le 1} \| \hat{C}_{XY}(\cdot, \cdot, x) - x\hat{C}_{XY}(\cdot, \cdot, 1) \|_2^2.$$

and

$$Z_{T,p} = T \sup_{0 \le x \le 1} \sum_{i=1}^{p} \frac{\langle \hat{C}_{XY}(\cdot, \cdot, x) - x\hat{C}_{XY}(\cdot, \cdot, 1), \hat{\varphi}_i \rangle_2^2}{\hat{\lambda}_i}.$$

Z_T and $Z_{T,p}$ are each maximally selected CUSUM type statistics based on comparing the partial sample estimates of C_{XY} to the estimator from the whole sample. These statistics are similar to those considered in Aue *et al.* (2018), Aston and Kirch (2012), and Sharipov *et al.* (2016). The following corollaries of Theorem 2.1 quantify the large sample behavior of these statistics under $H_{0,2}$.

Corollary 3.3. Under Assumption 2.1 and $H_{0,2}$,

$$Z_T \xrightarrow{\mathcal{D}} \sup_{0 \le x \le 1} \sum_{i=1}^{\infty} \lambda_i B_i^2(x), \text{ as } T \to \infty,$$

where $\{\lambda_i, \ i \ge 1\}$ are defined in (2.2), and $\{B_i(x), i \ge 1, \ x \in [0, 1]\}$ are i.i.d. standard Brownian bridges on [0, 1]. If in addition (3.2) holds, then

$$Z_{T,p} \xrightarrow{\mathcal{D}} \sup_{0 \le x \le 1} \sum_{i=1}^{p} B_i^2(x), \text{ as } T \to \infty,$$

It follows then that a test of $H_{0,2}$ with asymptotic level α is obtained by rejecting if Z_T or $Z_{T,p}$ are larger than the $1 - \alpha$ quantiles of their limiting distributions detailed in Corollary 3.3. These limiting distributions may again be approximated using Monte simulation. The statistics Z_T and $Z_{T,p}$ diverge under $H_{A,2}$ in conjunction with some further regularity conditions. Notably, no orthogonality condition is needed in order for the test statistic $Z_{T,p}$ to diverge, as compared to the consistency result for $F_{T,p}$. This is due to the way the estimates defined below for φ_i behave as the sample size increases under $H_{A,2}$. We address these consistency results in more detail Section A.3 in the appendix.

DOI: 10.1111/jtsa.12447

3.1. Application to Cross-Correlation Plots Between Functional Time Series

When the assumption of joint stationarity is thought to be valid, as can be tested using Z_T or $Z_{T,p}$, Corollary 3.1 can be used to further investigate the lagged dependence relationship between X_i and Y_i through an analog of a cross-correlation plot. The functional cross-correlation coefficient at lag h, $\rho_{XY,h}$, may be defined as

$$\rho_{XY,h} = \frac{\|C_{XY,h}\|_2}{\|[E(X_0(t) - \mu_X(t))^2]^{1/2}\|_1 \|[E(Y_0(t) - \mu_Y(t))^2]^{1/2}\|_1}.$$

A simple application of the Cauchy-Schwarz inequality shows that $0 \leq \rho_{XY,h} \leq 1$. With

$$\hat{\gamma}_{X,\ell}(t,s) = \frac{1}{T}\sum_{i=1}^{T-\ell}(X_i(t) - \bar{X}(t))(X_{i+\ell}(s) - \bar{X}(s)),$$

for $\ell \geq 0$, and

$$\hat{\gamma}_{X,\ell}(t,s) = \frac{1}{T}\sum_{i=1-\ell}^{T}(X_i(t) - \bar{X}(t))(X_{i+\ell}(s) - \bar{X}(s)),$$

for $\ell < 0$, and $\hat{\gamma}_Y(t,s)$ similarly defined, $\rho_{XY,h}$ can be estimated by

$$\hat{\rho}_{XY,h} = \frac{\|\hat{C}_{XY,h}(\cdot,\cdot,1)\|_2}{\left(\int \hat{\gamma}^2_{X,0}(t,t)\mathrm{d}t \int \hat{\gamma}^2_{Y,0}(t,t)\mathrm{d}t\right)^{1/2}}.$$

Assuming that the sequence $\{X_i\}$ is independent of $\{Y_i\}$, one has by Corollary 3.1 that for each fixed h,

$$T\|\hat{C}_{XY,h}(\cdot,\cdot,1)\|_2^2 \xrightarrow{\mathcal{D}} \sum_{j=1}^{\infty} \lambda_j^{ind}\mathcal{N}_j, \tag{3.4}$$

where the λ_j^{ind}'s are the eigenvalues of the covariance operator with kernel

$$D_{ind}(t,s,u,v) = \sum_{\ell=-\infty}^{\infty} \gamma_{X,\ell}(t,u)\gamma_{Y,\ell}(s,v)$$

where $\gamma_{X,\ell}(t,u) = E(X_0(t) - \mu_X(t))(X_\ell(u) - \mu_X(u))$, and γ_Y is similarly defined. D_{ind} may be estimated using an estimator of the form

$$\hat{D}_{ind}(t,s,u,v) = \sum_{\ell=-\infty}^{\infty} W_b\left(\frac{\ell}{B}\right)\hat{\gamma}_{X,\ell}(t,u)\hat{\gamma}_{Y,\ell}(s,v).$$

We consider such estimators in more detail in Section 4, as well as definitions of W_b and B. Furthermore, the λ_i^{ind} can be estimated as in (4.4), and one can estimate the limiting α quantile of $T\|\hat{C}_{XY,h}(\cdot,\cdot,1)\|^2$, $\hat{\Xi}_{ind}(\alpha)$, from (3.4). An approximate size α upper confidence bound for $\rho_{XY,h}$ assuming that the $\{X_i\}$ and $\{Y_i\}$ sequences are independent is then given by

DOI: 10.1111/jtsa.12447

$$\hat{I}_{ind}(\alpha) = \frac{\sqrt{\hat{\Xi}_{ind}(\alpha)}}{\sqrt{T}\left(\int \hat{\gamma}^2_{X,0}(t,t)dt \int \hat{\gamma}^2_{Y,0}(t,t)\mathrm{d}t\right)^{1/2}}.$$

Comparisons of $\hat{\rho}_{XY,h}$ and $\hat{I}_{ind}(\alpha)$ can be used to help identify lags at which $\{X_i\}$ and $\{Y_i\}$ are dependent. We illustrate this in the data application in Section 6 (see Figure 1).

4. IMPLEMENTATION

Implementing the testing procedures outlined above requires the estimation of the eigenvalues, and, in case of the dimension reduction based test statistics $F_{T,p}$ and $Z_{T,p}$, the eigenfunctions of $\mathfrak{d}$. We first develop methodology for estimating $\mathfrak{d}$ and its spectrum, and then describe some numerical methods for carrying out this estimation.

4.1. Estimation of D, $\mathfrak{d}$, φ_i, and λ_i

As $\mathfrak{d}$ is simply a scalar multiple of the spectral density operator at frequency zero of the stationary sequence $\{(X_i(t) - \mu_X(t))(Y_i(s) - \mu_Y(s))\}$ in $L^2[0,1]^2$, as defined in Panaretos and Tavakoli (2013a,b), it may be naturally estimated with a kernel lag-window type estimator. Let

$$\hat{D}_T(t,s,u,v) = \sum_{\ell=-\infty}^{\infty} W_b\left(\frac{\ell}{B}\right)\hat{\gamma}_\ell(t,s,u,v), \tag{4.1}$$

where B is a bandwidth parameter satisfying,

$$B = B(T) \to \infty, \ \frac{B}{T^{1/2}} \to 0 \text{ as } T \to \infty, \tag{4.2}$$

and, with $\bar{X}_j(t) = X_j(t) - \bar{X}(t)$ and $\bar{Y}_j(s)$ similarly defined,

$$\hat{\gamma}_\ell(t,s,u,v) = \begin{cases} \frac{1}{T}\sum_{j=1}^{T-\ell}\left(\bar{X}_j(t)\bar{Y}_j(s) - \hat{C}_{XY}(t,s,1)\right)\left(\bar{X}_{j+\ell}(u)\bar{Y}_{j+\ell}(v) - \hat{C}_{XY}(u,v,1)\right), & \ell \geq 0 \\ \frac{1}{T}\sum_{j=1-\ell}^{T}\left(\bar{X}_j(t)\bar{Y}_j(s) - \hat{C}_{XY}(t,s,1)\right)\left(\bar{X}_{j+\ell}(u)\bar{Y}_{j+\ell}(v) - \hat{C}_{XY}(u,v,1)\right), & \ell < 0. \end{cases}$$

We take the function W_b to be a symmetric and continuous weight function with bounded support of order b; see Chapter 7 of Priestley (1981). $\hat{D}_T$ then defines an estimator of $\mathfrak{d}$ by

$$\hat{\mathfrak{d}}_T(f)(t,s) = \iint \hat{D}_T(t,s,u,v)f(u,v)\mathrm{d}u\mathrm{d}v, \tag{4.3}$$

which further defines estimates of the eigenvalues and eigenfunctions of $\mathfrak{d}$ satisfying

$$\hat{\mathfrak{d}}_T(\hat{\varphi}_i)(t,s) = \hat{\lambda}_i\hat{\varphi}_i(t,s). \tag{4.4}$$

In the simulations and application below, we consider the Bartlett weight function, $W_1(x) = (1-|x|)\mathbb{1}(|x|<1)$, which is of order one. We also repeated all simulations using the Parzen weight function, which is of order two, and found that changing the weight function had relatively limited effect relative to the differences in changing the bandwidth; see Andrews (1991) for the definition of the Parzen weight function. To select the bandwidth

B, one could employ any of a number of methods, for example using the criteria discussed in Hörmann and Kokoszka (2010) or by adapting the data driven procedure in Rice and Shang (2017). We study below the choice of $B = T^{1/(2b+1)}$ where $b = 1$ and $b = 2$, which diverge at an optimal rate in terms of minimizing the asymptotic mean squared normed error of the estimator $\hat{D}$ when the weight function is of order b. The choice of $B = T^{1/5}$ is suggested as a heuristic choice for this reason (see Brillinger, 1975), and was also considered in the context of spectral density operator estimation of functional time series in Panaretos and Tavakoli (2013b). We have the following result.

Theorem 4.1. Under Assumption 2.1, $\|D - \hat{D}_T\|_4 = o_P(1)$.

To obtain consistent estimates of the first p elements of the spectrum of $\mathfrak{d}$, we assume for the sake of simplicity that the first p eigenvalues of $\mathfrak{d}$ are distinct, which implies that the corresponding eigenspaces of the first p eigenvalues are one dimensional.

Assumption 4.1. We assume that there exists an integer $p \geq 1$ satisfying that

$$\lambda_1 > \cdots > \lambda_p > \lambda_{p+1} \geq 0$$

where $\{\lambda_i,\ i \geq 1\}$ are defined in (2.2).

Assumption 4.1 could be relaxed by utilizing some of the ideas presented in Reimherr (2015), but we do not pursue those here. Under this assumption, the following result is implied by Theorem 4.1 and the results in Section 6.1 of Gohberg *et al.* (1990).

Corollary 4.1. Under Assumption 2.1 and 4.1, (3.2) holds.

4.2. Numerical Implementation

Although so far we have presented results as if the functional data at hand were observed on their entire domains, in practice the data will consist of only discrete observations of the underlying functions. Let $X_i(t_j)$ and $Y_i(t_j)$, $1 \leq j \leq R$, denote the observed values of the functions $X_i(t)$ and $Y_i(t)$, observed at the common points $\{t_1, \ldots, t_R\}$. We assume here that each functional observation $X_i(t)$ and $Y_i(s)$ are observed at common points in their domains, as this matches our simulations and data example below, although this could be relaxed.

It is straightforward to estimate the test statistics F_T, $F_{T,p}$, Z_T, and $Z_{T,p}$ from the discrete data and simple Riemann sum approximations to the inner products and norms, so long as the eigenvalue and eigenfunction estimates $\hat{\lambda}_i$ and $\hat{\varphi}_i$ are given. However it is less clear how to estimate the eigen-elements satisfying (4.4) from the discrete data. One could in principle estimate the eigenvalues and eigenfunctions $(\lambda_i, \varphi_i(t,s))$, $1 \leq i \leq p$ of $\mathfrak{d}$ by calculating the spectrum of the four-way tensor $\hat{D}_T(t_i, t_j, t_k, t_\ell)$, $1 \leq i, j, k, \ell \leq R$ of dimension R^4, and employing linear interpolation to complete the eigenfunction estimates, but this becomes computationally infeasible for even moderate values of R due to the dimension of the tensor, and the fact that this tensor is typically *dense*.

What we propose instead is a dimension reduction based approach involving the functional principal components (fPC's) of the individual series $\{X_i\}$ and $\{Y_i\}$. Let

$$\hat{c}_X := \left[\frac{1}{T} \sum_{\ell=1}^{T} (X_\ell(t_i) - \bar{X}(t_i))(X_\ell(t_j) - \bar{X}(t_j)) \ :\ 1 \leq i, j \leq R \right]$$

be the sample covariance matrix of the discretized observations of the X sample, and define $\hat{c}_Y$ similarly. Calculating the spectra of $\hat{c}_X$ and $\hat{c}_Y$ yields eigenvalues $\{\nu_{X,1}, \ldots, \nu_{X,R}\}$, $\{\nu_{Y,1}, \ldots, \nu_{Y,R}\}$ and eigenvectors, $\{\theta_{X,1}, \ldots, \theta_{X,R}\}$,

DOI: 10.1111/jtsa.12447

$\{\theta_{Y,1},\dots,\theta_{Y,R}\}$, the latter of which, when multiplied by $\sqrt{R}$, yield discrete approximations to the fPC's of the sequences $\{X_i\}$ and $\{Y_i\}$, $\theta_{X,i}(t_j)$, $\theta_{Y,i}(t_j)$, $1 \le i,j \le R$ respectively. They can be completed to the rest of the unit interval by linear interpolation, which we employ.

We then use the product basis of $L^2[0,1]^4$ generated by these functions to reduce the dimension of $\hat{D}_T$. The projections of $\hat{D}_T$ onto the product basis of the first q_X elements of $\theta_{X,i}(t)$ and q_Y elements $\theta_{Y,i}(t)$ may be stored in a 4-way tensor with $q_X^2 q_Y^2$ elements, $\mathcal{M}$, via

$$\mathcal{M}_{ijkr} = \int \cdots \int \hat{D}_T(t,s,u,v)\theta_{X,i}(t)\theta_{Y,j}(s)\theta_{X,k}(u)\theta_{Y,r}(v)\mathrm{d}t\mathrm{d}s\mathrm{d}u\mathrm{d}v.$$

Each of these elements can be estimated with a simple Riemann sum approximation. The cutoffs q_X and q_Y must be selected by the user. We suggest using the total variance explained approach for this: we take q_X and q_Y so that $\theta_{X,i}(t)$ $1 \le i \le q_X$ and $\theta_{Y,i}(t)$, $1 \le i \le q_Y$, explain at least v% of the variation in each series, where v is close to but strictly smaller than 1; see Chapter 6 of Ramsay and Silverman (2005). This also motivates a natural method to select the integer p in the statistics $F_{T,p}$ and $Z_{T,p}$, for example to take $p = \max\{q_X, q_Y\}$.

The eigenvalues $\{\hat{\lambda}_i, 1 \le i \le q_X q_Y\}$ and eigen-arrays $\{\hat{\mathbf{\Phi}}_i \in \mathbb{R}^{q_X \times q_Y}, 1 \le i \le q_X q_Y\}$ of $\mathcal{M}$ satisfy

$$\mathcal{M}\hat{\mathbf{\Phi}}_\ell = \hat{\lambda}_\ell \hat{\mathbf{\Phi}}_\ell, \quad 1 \le \ell \le q_X q_Y, \text{ where } \mathcal{M}\hat{\mathbf{\Phi}}_\ell[i,j] = \sum_{k=1}^{q_X}\sum_{r=1}^{q_Y} \mathcal{M}_{ijkr}\hat{\mathbf{\Phi}}_\ell[k,r].$$

This eigenvalue problem may be solved numerically by solving for the eigenvalues/vectors of a $q_X q_Y$ by $q_X q_Y$ square matrix that is 'tiled' with the cross-sections of $\mathcal{M}$, as is implemented in `svd.tensor` function in the package `tensorA` in `R`; see van den Boogaart (2010). The eigenfunctions $\varphi_i(t,s)$ may then be estimated with

$$\hat{\varphi}_i(t,s) = \sum_{j=1}^{q_X}\sum_{k=1}^{q_Y} \hat{\mathbf{\Phi}}_i[j,k]\theta_{X,j}(t)\theta_{Y,k}(s).$$

We use these estimates for the $\hat{\lambda}_i'$s and $\hat{\varphi}_i'$s in the simulations and applications below. Beyond that the functions $\theta_{X,i}(t)$ and $\theta_{Y,j}(s)$ may be used to effectively represent the original X and Y series, we also desire that the discrete approximation of the operator $\hat{\mathfrak{d}}$ given by $\mathcal{M}$ is accurate. We may therefore also employ that

$$\frac{\hat{\lambda}_1 + \cdots + \hat{\lambda}_{q_X q_Y}}{\iint \hat{D}_T(t,s,t,s)\mathrm{d}t\mathrm{d}s} \ge v_d,$$

where again v_d is a large percentage. In other words, this condition imposes that the traces of the operators defined by $\mathcal{M}$ and $\hat{D}_T$ agree up to some tolerance defined by v_d. This criterion may be met by increasing q_X and q_Y when necessary.

5. SIMULATION STUDY

5.1. Outline

We now present the results of a small simulation experiment that aimed to study the finite sample properties of the test statistics introduced above. All of the simulations reported below were done using the `R` language (R Development Core Team, 2015). The number of potential experimental settings that could be considered to study Theorem 2.1 and Corollaries 3.1–3.3 is enormous. For the sake of brevity, we mainly focused on demonstrating that for a somewhat rich class of data generating processes (DGP's) exhibiting serial correlation, the tests based on F_T, $F_{T,p}$, Z_T, and $Z_{T,p}$ hold their size well, and that the eigenvalue and eigenfunction estimation procedure explained in Section 4.2 is adequate for such hypothesis testing problems. Towards this goal, we considered the following basic structure for generating synthetic data depending on the parameter $\alpha \in [0, 1]$:

$$X_i(t) = \alpha\epsilon_{c,i}(t) + (1 - \alpha)\epsilon_{x,i}(t), \quad Y_i(t) = \alpha\epsilon_{c,i}(t) + (1 - \alpha)\epsilon_{y,i}(t), \quad 1 \leq i \leq T \tag{5.1}$$

where $\epsilon_{c,i}(t)$, $\epsilon_{x,i}(t)$, and $\epsilon_{y,i}(t)$ are mutually independent sequences that satisfy either of two models given below. In this case, the functional series $X_i(t)$ and $Y_i(t)$ are correlated through their common dependence on $\epsilon_{c,i}(t)$, with the strength of this dependence controlled by α. If $\alpha = 0$ for instance, then the two sequences X_i and Y_i are independent. We took the error sequences to satisfy either:

IID-BM: $\epsilon_{c,i}(t) = W_{c,i}(t)$, $\epsilon_{x,i}(t) = W_{x,i}(t)$, and $\epsilon_{y,i}(t) = W_{y,i}(t)$, where $\{W_{c,i}(t)\}$, $\{W_{x,i}(t)\}$, $\{W_{y,i}(t)\}$ are mutually independent sequences of IID standard Brownian motions.

IID-EXP: $\epsilon_{c,i}(t) = E_{c,i}(t)$, $\epsilon_{x,i}(t) = E_{x,i}(t)$, and $\epsilon_{y,i}(t) = E_{y,i}(t)$, where

$$E_{i,c}(t) = \sum_{j=1}^{3} e_{i,j}\phi_j(t),$$

$e_{i,j}$ are i.i.d. centered exponential random variable with parameter one, and $\{\phi_i(t)\}$ is the standard Fourier basis. $E_{i,x}$ and $E_{i,y}$ are similarly defined.

FAR(1): $\epsilon_{c,i}(t) = \int \Phi(t,s)\epsilon_{c,i}(s)ds + W_{c,i}(t)$, $\epsilon_{x,i}(t) = \int \Phi(t,s)\epsilon_{x,i}(s)ds + W_{x,i}(t)$, and $\epsilon_{y,i}(t) = \int \Phi(t,s)\epsilon_{y,i}(s)ds + W_{y,i}(t)$, where $\{W_{c,i}(t)\}$, $\{W_{x,i}(t)\}$, and $\{W_{y,i}(t)\}$ are defined above, and $\Phi(t,s) = \min(t,s)$.

FMA(1): $\epsilon_{c,i}(t) = W_{c,i}(t) + W_{c,i-1}(t)$, $\epsilon_{x,i}(t) = W_{x,i}(t) + W_{x,i-1}(t)$, and $\epsilon_{y,i}(t) = W_{y,i}(t) + W_{y,i-1}(t)$, where $\{W_{c,i}(t)\}$, $\{W_{x,i}(t)\}$, and $\{W_{y,i}(t)\}$ are defined above.

The sequence $\{(X_i, Y_i)\}$ satisfying (5.1) with error sequences satisfying either of the above models satisfy Assumption 2.1. The choice of the Brownian motions for the innovation sequences in the FAR and FMA processes is partially motivated by our application to intraday returns data below, see Figure 2. The process IID-EXP is meant to both study an example in which the underlying process is not Gaussian, and to also consider a process of finite dimension to evaluate whether the projection based test statistics perform better than their norm based counterparts in this case. For each setting of T and α, the data was generated on an equally spaced grid on $[0, 1]$ with $R = 100$ points, and the simulation was repeated 1000 times for each setting to calculate the empirical size and power curves presented below. We chose $q_X = q_Y = 3$ in the estimation of procedure for the eigenvalues and eigenfunctions in Section 4, which in the vernacular of fPCA corresponds to a total variance explained of around 93% on average for each sequence X_i and Y_i in case of the IID-BM, FAR(1) and FMA(1) processes. We considered larger values of q_X and q_Y, and found that it did not have much of an effect on the results, although the computational time increases substantially as these parameters increase. We similarly took $p = 3$ in the definition of $F_{T,p}$ and $Z_{T,p}$.

To study the size of the test of $H_{0,1}$ based on F_T and $F_{T,3}$, we generated data according to (5.1) for each setting of the error sequence with $\alpha = 0$, and performed tests with nominal levels of 5%, and 1% of $H_{0,1} : C_{XY} = 0$. The rejection rates from 1000 simulations from each statistic are reported in Table I. One thing that was clear based on these simulations was that the tests based on either statistic tended to be somewhat oversized. This issue improves as T increases, as predicted by Corollaries 3.1 and 3.2, and all tests achieved quite good size once $T \geq 300$. The presence of serial correlation increases the size inflation of the tests, however this is fairly well controlled by the simple kernel lag-window estimator of the operator $\mathfrak{d}$. This could be improved further by increasing the bandwidth at the cost of increased computational time, according to unreported simulations. The fact that the empirical levels are quite close to nominal and improve with increasing T is indicative that the eigenvalue/eigenfunction estimation described in Section 4.2 is performing adequately for this application.

DOI: 10.1111/jtsa.12447

Table I. Empirical sizes with nominal levels of 5%, and 1% for a test of $H_{0,1} : C_{XY} = 0$ where the data was generated according to (5.1) with $\alpha = 0$

Statistic: F_T									
		IID-BM		IID-EXP		FAR(1)		FMA(1)	
B	T	5%	1%	5%	1%	5%	1%	5%	1%
$T^{1/5}$	50	0.061	0.019	0.044	0.003	0.091	0.019	0.101	0.032
	100	0.052	0.010	0.039	0.004	0.069	0.017	0.094	0.033
	300	0.062	0.010	0.057	0.005	0.061	0.018	0.074	0.020
$T^{1/3}$	50	0.086	0.024	0.048	0.011	0.099	0.023	0.104	0.027
	100	0.072	0.014	0.040	0.005	0.095	0.032	0.083	0.026
	300	0.062	0.016	0.056	0.016	0.065	0.015	0.067	0.021
Statistic: $F_{T,3}$									
		IID-BM		IID-EXP		FAR(1)		FMA(1)	
B	T	5%	1%	5%	1%	5%	1%	5%	1%
$T^{1/5}$	50	0.061	0.019	0.042	0.006	0.077	0.023	0.070	0.012
	100	0.052	0.010	0.035	0.006	0.065	0.013	0.071	0.015
	300	0.062	0.010	0.043	0.011	0.057	0.017	0.057	0.016
$T^{1/3}$	50	0.091	0.022	0.041	0.004	0.098	0.030	0.089	0.025
	100	0.069	0.020	0.035	0.007	0.101	0.027	0.070	0.015
	300	0.060	0.016	0.048	0.013	0.061	0.013	0.059	0.018

We also studied the empirical power of F_T and $F_{T,3}$ by testing $H_{0,1} : C_{XY} = 0$ with data following (5.1), both with FAR(1) and IID-EXP errors, and α increasing from 0 to 1. The results of this simulation are reported as power curves in Figure 3. We observed that both tests were powerful for large enough values of α. Interestingly, the simple norm based test F_T possessed better power than $F_{T,3}$ in all the examples that we considered, including the IID-EXP case, in which the dimension of the $\{X_i\}$ and $\{Y_i\}$ series is equal to q_X and q_Y respectively. The only way that we can think to explain this is that the difference stems mostly from the way estimation error of the eigenvalues λ_i of $\mathfrak{d}$ affects each test statistic. In the case of the norm based test, errors in estimating small eigenvalues have a negligible effect, since they manifest as small changes in the estimated limiting distribution, while the projection based tests can be greatly affected by small estimation errors of the eigenvalues. A similar observation is made in the context of change point analysis in Aue *et al.* (2018).

For testing $H_{0,2}$ based on Z_T and $Z_{T,3}$, we briefly present the results of another empirical size and power study. In this case we applied each statistic to data generated according to (5.1) with α ranging from zero to one, and for $T = 100, 300$, and 1000. As expected in this case the value of α made little difference, and the empirical size from 1000 independent simulations for $\alpha = 0$ each two values of the bandwidth are presented Table II. We saw that in general the change point tests based on Z_T and $Z_{T,3}$ tended to be undersized, with this being more pronounced for the statistic $Z_{T,3}$, but this improved with increasing sample size.

To study the power of the change point tests under $H_{A,2}$, we considered two DGP exhibiting a change point in the cross-covariance structure at lag zero. We generated observations (X_i, Y_i) for $1 \le i \le \lfloor T/2 \rfloor$ from model (5.1), and then generated (X_i, Y_i), $\lfloor T/2 \rfloor + 1 \le i \le T$ from either

- Scenario 1, Change in direction of relationship:

$$X_i(t) = -\alpha\varepsilon_{c,i}(t) + (1-\alpha)\varepsilon_{x,i}(t), \quad Y_i(t) = \alpha\varepsilon_{c,i}(t) + (1-\alpha)\varepsilon_{y,i}(t), \quad \lfloor T/2 \rfloor + 1 \le i \le T$$

- Scenario 2, Change in strength of relationship:

$$X_i(t) = (1-\alpha)\varepsilon_{c,i}(t) + \alpha\varepsilon_{x,i}(t), \quad Y_i(t) = (1-\alpha)\varepsilon_{c,i}(t) + \alpha\varepsilon_{y,i}(t), \quad \lfloor T/2 \rfloor + 1 \le i \le T$$

DOI: 10.1111/jtsa.12447

Table II. Empirical sizes with nominal levels of 5%, and 1% for a test of $H_{0,2}\ :\ C_{XY}^{(1)} = \ldots = C_{XY}^{(T)}$ where the data was generated according to (5.1) with $\alpha = 0$

Statistic: Z_T									
		IID-BM		IID-EXP		FAR(1)		FMA(1)	
B	T	5%	1%	5%	1%	5%	1%	5%	1%
$T^{1/5}$	50	0.026	0.001	0.015	0.000	0.023	0.001	0.034	0.003
	100	0.040	0.006	0.016	0.003	0.046	0.004	0.045	0.003
	300	0.050	0.007	0.033	0.006	0.052	0.011	0.067	0.018
	1000	0.064	0.017	0.049	0.007	0.072	0.012	0.080	0.012
$T^{1/3}$	50	0.024	0.000	0.004	0.000	0.034	0.003	0.027	0.000
	100	0.035	0.002	0.010	0.000	0.048	0.005	0.041	0.003
	300	0.059	0.013	0.037	0.003	0.078	0.010	0.059	0.009
	1000	0.078	0.018	0.047	0.007	0.067	0.022	0.058	0.011

Statistic: $Z_{T,3}$									
		IID-BM		IID-EXP		FAR(1)		FMA(1)	
B	T	5%	1%	5%	1%	5%	1%	5%	1%
$T^{1/5}$	50	0.003	0.000	0.000	0.000	0.006	0.000	0.000	0.000
	100	0.017	0.002	0.018	0.001	0.028	0.003	0.002	0.000
	300	0.043	0.002	0.043	0.001	0.046	0.005	0.024	0.002
	1000	0.073	0.009	0.038	0.008	0.068	0.011	0.070	0.017
$T^{1/3}$	50	0.000	0.000	0.002	0.000	0.003	0.000	0.001	0.000
	100	0.008	0.000	0.012	0.000	0.011	0.000	0.009	0.000
	300	0.028	0.002	0.028	0.003	0.032	0.003	0.038	0.006
	1000	0.059	0.007	0.057	0.011	0.055	0.006	0.042	0.008

Table III. Empirical power for change point tests with nominal levels of 5%, and 1% where the data was generated according to Scenarios 1 and 2 with errors following FAR(1) and IID-EXP. In each case $B = T^{1/3}$

DGP: FAR(1)					
		Z_T		$Z_{T,3}$	
Scenario	T	5%	1%	5%	1%
1	50	0.040	0.005	0.009	0.000
	100	0.086	0.015	0.043	0.007
	300	0.155	0.040	0.089	0.011
	1000	0.349	0.164	0.218	0.086
2	50	0.058	0.005	0.014	0.001
	100	0.100	0.020	0.041	0.003
	300	0.212	0.074	0.112	0.029
	1000	0.494	0.284	0.345	0.148

DGP: IID-EXP					
		Z_T		$Z_{T,3}$	
Scenario	T	5%	1%	5%	1%
1	50	0.008	0.000	0.002	0.000
	100	0.022	0.000	0.014	0.007
	300	0.125	0.032	0.084	0.011
	1000	0.447	0.207	0.346	0.086
2	50	0.007	0.000	0.000	0.000
	100	0.025	0.003	0.016	0.000
	300	0.102	0.016	0.075	0.008
	1000	0.344	0.133	0.330	0.121

DOI: 10.1111/jtsa.12447

In Scenario 1 we set $\alpha = 0.2$ and Scenario 2 we set $\alpha = 0.46$ so that in both cases $\|C_{XY}^{(1)} - C_{XY}^{(T)}\|$ are approximately equal for the purpose of comparison. The empirical power for several values of T and errors following the FAR(1) and IID-EXP DGP are given in Table III. We did not notice much of a difference in power between Scenarios 1 and 2. Again we noticed that the norm based methods out-performed the projection based methods when the errors followed an FAR(1) model, but the methods were more comparable for IID-EXP errors.

6. APPLICATION TO CUMULATIVE INTRADAY RETURNS

A natural example of functional time series data are those derived from densely recorded asset price data, such as intraday stock price data. Recently there has been an upsurge in quantitative research focused on analyzing the information contained within curves constructed from such data; we refer the reader to Barndorff-Nielsen and Shephard (2004), Wang and Zou (2010), Gabrys *et al.* (2010), Müller *et al.* (2011), and Kokoszka and Reimherr (2013). Price curves associated with popular companies are commonly displayed at websites and applications like yahoo.com/finance.

The specific data that we consider was obtained from www.nasdaq.com, and consists of 1 minute resolution closing prices of a single share of Microsoft (ticker `MSFT`), and Exxon Mobil (ticker `XOM`) stocks from 2 January to 31 December 2001, which comprises data from 248 trading days ($T = 248$) with $R = 389$ observations per day. We applied the methods introduced in Sections 3 to study the cross-covariance structure between these two companies stock prices on the intraday scale. Let $P_{M,i}(t_j)$, and $P_{X,i}(t_j)$ $i = 1, \ldots, T, j = 1, \ldots, R$, denote the prices of Microsoft and Exxon Mobil stock on day i at intraday time t_j respectively. The first three price curves of each series constructed from the raw price data and linear interpolation are displayed in the left hand panel of Figure 2. The functional time series of price curves are evidently non-stationary due to frequent level shifts and volatility, and hence we considered the following transformation of these curves akin to taking the log returns for scalar price data:

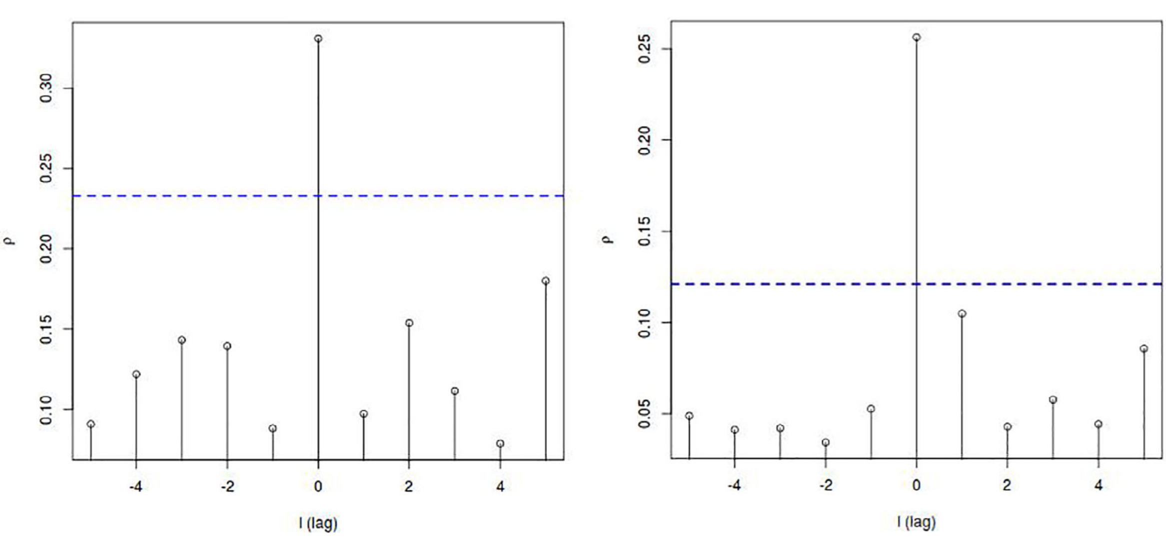

Figure 1. The left and right hand panels show plots of $\hat{\rho}_{XY,h}$ for $-5 \leq h \leq 5$ versus $\hat{I}_{ind}(0.95)$ (blue line) before and after the date 22 March 2001 respectively. The plots indicate the cross-covariance at lags other than zero are not significantly different from zero [Color figure can be viewed at wileyonlinelibrary.com]

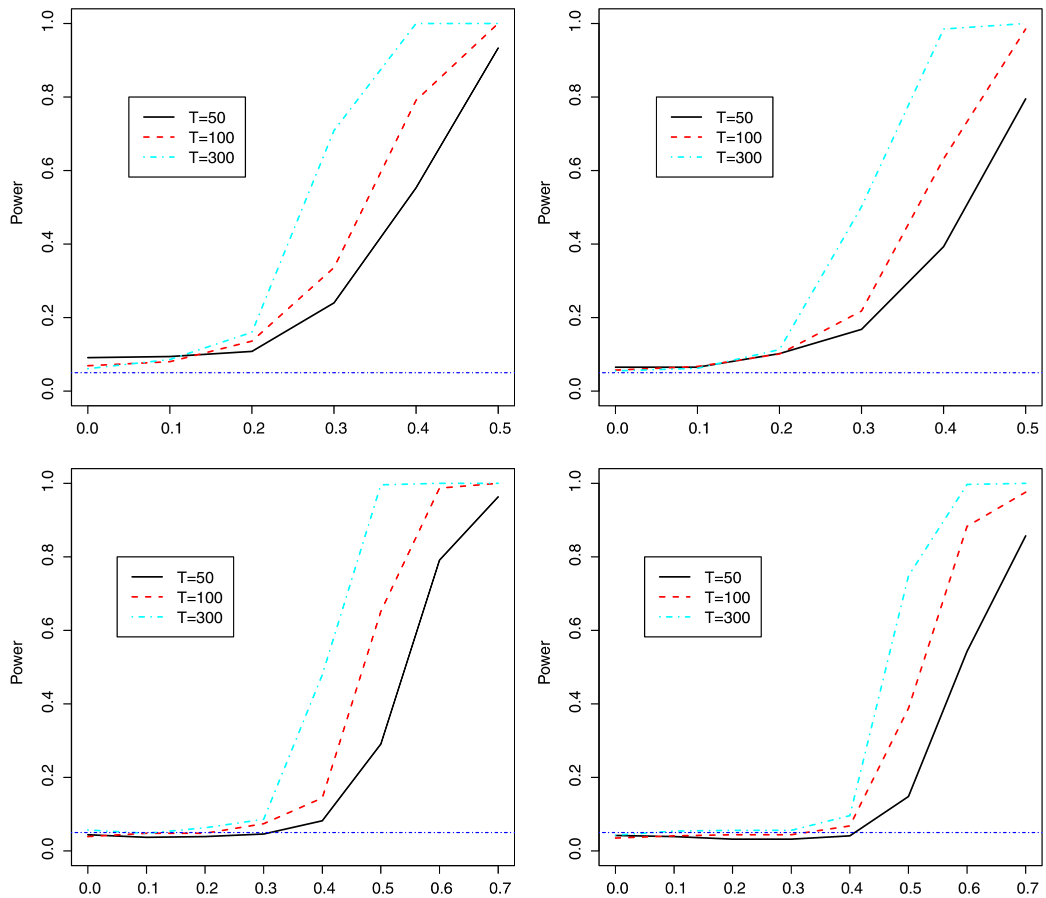

Figure 2. The left hand panel displays three intraday price curves derived from the one-minute resolution closing prices of a single share of XOM and MSFT. The right hand panel displays the corresponding CIDR curves [Color figure can be viewed at wileyonlinelibrary.com]

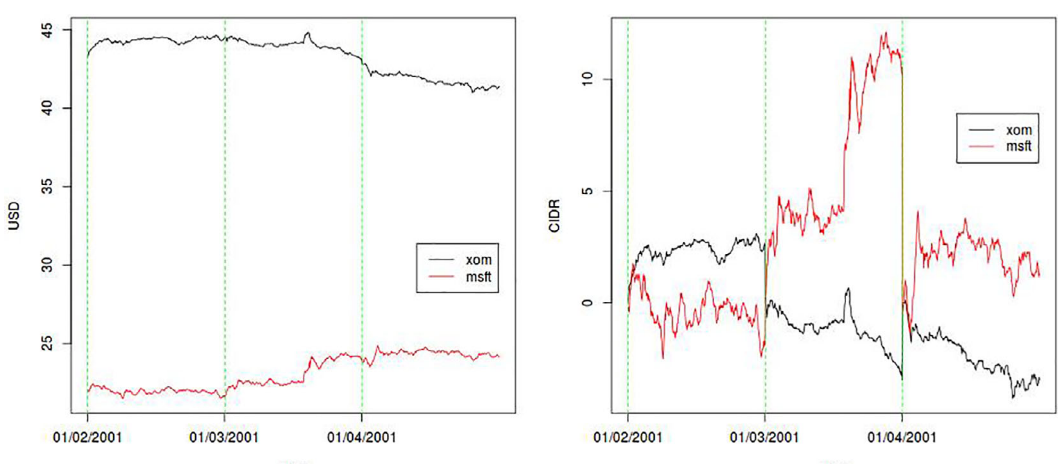

Figure 3. Power curves as a function of α ranging from 0 to 0.8 at increments of 0.1, for a test $H_{0,1}\,:\,C_{XY}=0$ with level of 5% with $B=T^{1/5}$ applied to (X_i, Y_i) following (5.1) with FAR(1) errors (top) and IID-EXP errors (bottom) using F_T (left), and $F_{T,3}$ (right) [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1111/jtsa.12447

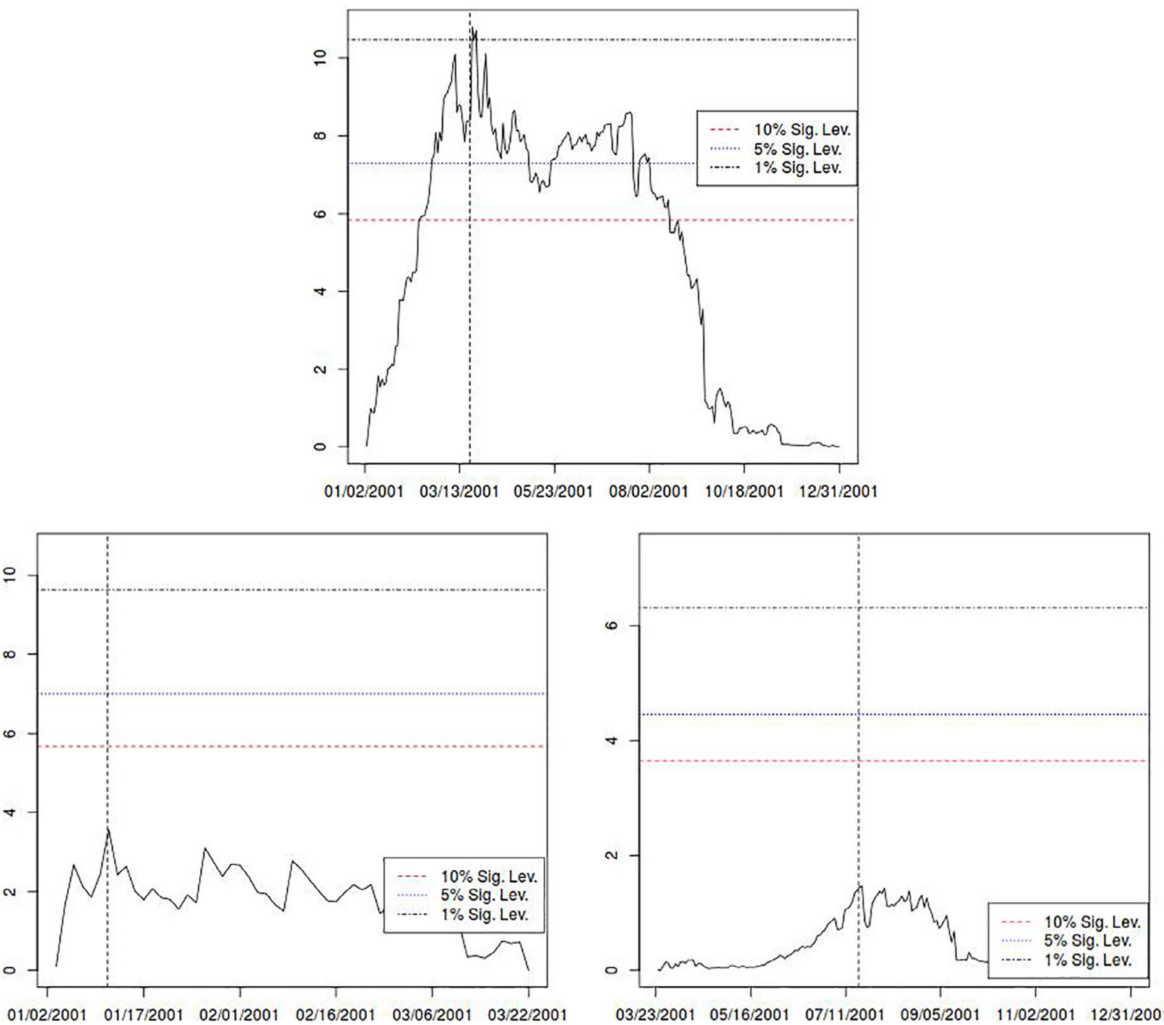

Figure 4. The top panel displays a 'CUSUM chart' of $\|\hat{C}_{XY}(\cdot,\cdot,t/T)-(t/T)\hat{C}_{XY}(\cdot,\cdot,1)\|_2^2$ versus t calculated from the `XOM` and `MSFT` price data from 2001. The maximum value Z_T is significant to the 0.01 level, and is achieved at the value t corresponding to 22 March. The lower left and lower right panels show similar CUSUM charts for the sub samples of data before and after this point respectively [Color figure can be viewed at wileyonlinelibrary.com]

Definition 1. Suppose $P_i(t_j), i = 1, \dots, T, j = 1, \dots, R$, is the price of a financial asset at time t_j on day i. The functions

$$r_i(t_j) = 100[\ln P_i(t_j) - \ln P_i(t_1)], \quad j = 1, 2, \dots, R, \quad i = 1, \dots, T,$$

are called the *cumulative intraday returns* (CIDR's).

Since the logarithm is increasing, the CIDR curves have nearly the same shape as the original daily price curves, but the assumption of stationarity is much more plausible for these curves. According to their definition, the CIDR's always start from zero, so level stationarity is enforced, and taking the logarithm helps reduce potential scale inflation. The stationarity of CIDR curves derived from intraday stock price data was argued empirically in Horváth *et al.* (2014).

DOI: 10.1111/jtsa.12447

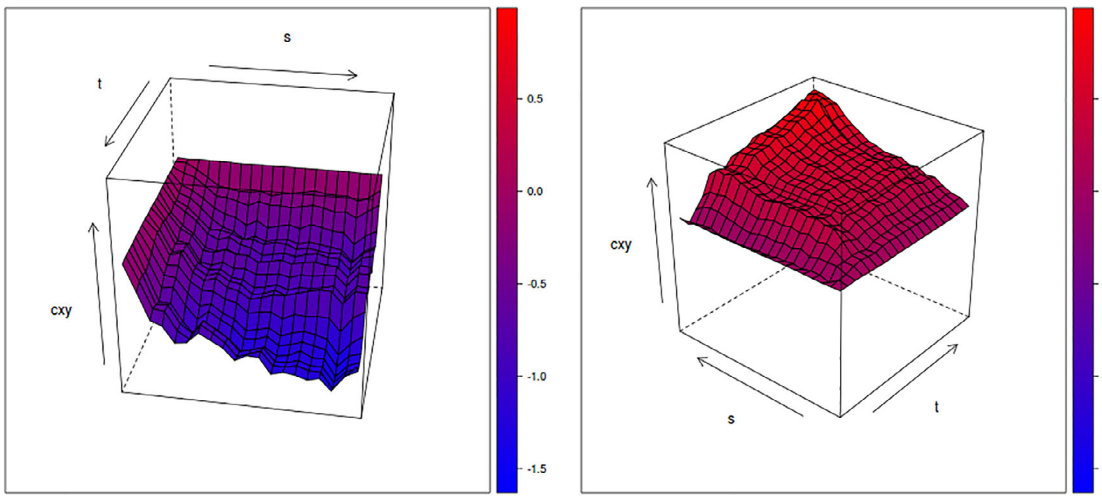

Figure 5. The left and right hand panels display the estimator $\hat{C}_{XY}$ before and after the date 22 March 2001 respectively. Both of these surface estimates were measured to be significantly different from zero, as the test of $H_{0,1}\ :\ C_{XY} = 0$ based on F_T yielded p-values less than 0.000 in both cases [Color figure can be viewed at wileyonlinelibrary.com]

Let $X_i(t) = r_{M,i}(t)$ and $Y_i(t) = r_{X,i}(t)$, $1 \le i \le T$, denote the CIDR curves derived from the Microsoft and Exxon Mobil stock price data and linear interpolation respectively. The first three CIDR curves of each series are plotted in the right hand panel of Figure 2.

Before estimating and measuring the significance of the cross-covariance between these functional time series of CIDR's, we first tested for its homogeneity within the sample, a test of $H_{0,2}$, using the test statistic Z_T. For this analysis, we took $q_X = q_M = 3$ for the method outline in Section 4.2, which corresponded to approximately 94% variance explained in each series. A 'CUSUM chart' of $c(t) = T\|\hat{C}_{XY}(\cdot,\cdot,t/T) - (t/T)\hat{C}_{XY}(\cdot,\cdot,1)\|_2^2$ versus t is given in Figure 4 with the corresponding 10%, 5% and 1% significance levels of the estimated limiting distribution in Corollary 3.1. The statistic Z_T far exceeded the 1% level, which indicates that there is strong evidence that the covariance relationship is heterogenous within the sample. Also apparent in the plot, the largest difference between the partial sample cross-covariance estimate occurs on 22 March 2001. We segmented the data into two subsamples before and after this point of lengths $T_1 = 56$ and $T_2 = 192$ respectively, and again tested for the homogeneity of the cross-covariance within each sub-sample. In both sub-samples the homogeneity could not be rejected with any significance. Interestingly, the sample including the intraday `XOM` and `MSFT` returns data before and after the terrorist attacks on 11 September 2001 did not seem to exhibit a change point in the cross-covariance relationship. The date 22 March 2001 does however seem to be in the proximity of some fairly important events relating to the world oil market. On 17 March, OPEC had announced that they were cutting oil production by 4%, and just three days later the largest off shore oil rig in the world, Petrobras 36, sank off of the coast of Brazil.

To measure the significance of the estimates of the cross-covariance before and after the change point, we applied a test of $H_{0,1}\ :\ C_{XY} = 0$ to each sub-sample. The null hypothesis was strongly rejected in both cases with p-values equal to zero out to three decimal places. The cross-covariance surfaces are displayed in Figure 5, from which we can see that the shape of the cross-covariance is quite different before and after the initial change point: the surface goes from being negative to predominantly positive. Plots of $\hat{\rho}_{XY,h}$ versus $\hat{I}_{ind}(0.95)$ in Figure 1 indicate that the cross-covariance between these curves at other lags is not significant at any reasonable level.

ACKNOWLEDGEMENTS

We thank the associate editor and two anonymous referees for their insightful and constructive comments that led to substantial improvements to this article. The first author is partially supported by the Natural Science and Engineering Research Council of Canada's Discovery and Accelerator grants.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

DATA AVAILABILITY STATEMENT

The data used to produce this example and supporting code are available in an online supporting information to this article.

REFERENCES

Andrews D. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**: 817–858.

Aston J, Kirch C. 2012. Detecting and estimating epidemic changes in dependent functional data. *Journal of Multivariate Analysis* **109**: 204–220.

Aue A, Hörmann S, Horváth L, Reimherr M. 2009. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics* **37**: 4046–4087.

Aue A, Horváth L, Pellatt DF. 2015. *Functional Generalized Autoregressive Conditional Heteroskedasticity*. Davis, CA: University of California, Davis. Technical report.

Aue A, Norinho DD, Hörmann S. 2014. On the prediction of stationary functional time series. *Journal of the American Statistical Association* **110**: 378–392.

Aue A, Rice G, Sönmez O. 2018. Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society, Ser. B.* **80**: 509–529.

Baker C. 1973. Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society* **186**: 273–289.

Barndorff-Nielsen OE, Shephard N. 2004. Econometric analysis of realized covariance: high frequency based covariance, regression and correlation in financial economics. *Econometrica* **72**: 885–925.

Berkes I, Horváth L, Rice G. 2013. Weak invariance principles for sums of dependent random functions. *Stochastic Processes and Their Applications* **123**: 385–403.

Bosq D. 2000. *Linear Processes in Function Spaces*. New York: Springer.

Brillinger DR. 1975. *Time Series: Data Analysis and Theory*. New York: Holt.

Chiou J-M, Chen Y-T, Yang Y-F. 2014. Multivariate functional principal component analysis: a normalization approach. *Statistica Sinica* **24**: 1571–1596.

Chiou J-M, Müller H-G. 2013. Linear manifold modeling of multivariate functional data. *Journal of the Royal Statistical Society Series B* **76**: 605–626.

Dette H, Wu W, Zhou Z. 2015. Change point analysis of second order characteristics in non-stationary time series. Technical Report.

Dubin J, Müller H-G. 2005. Dynamical correlation for multivariate longitudinal data. *Journal of the American Statistical Association* **100**: 872–881.

Ferraty F, Vieu P. 2006. *Nonparametric Functional Data Analysis: Theory and Practice*. New York: Springer.

Gabrys R, Horváth L, Kokoszka P. 2010. Tests for error correlation in the functional linear model. *Journal of the American Statistical Association* **105**: 1113–1125.

Gohberg I, Golberg S, Kaashoek MA. 1990. Classes of linear operators. In *Operator Theory: Advances and Applications*, Vol. 49. Basel: Birkhaüser.

Himdi KE, Roy R, Duchesne P. 2003. Tests for non-correlation of two multivariate time series: a nonparametric approach. *IMS Lecture Notes Monogr. Ser., Mathematical Statistics and Applications: Festschrift for Constance van Eeden* **42**: 397–416.

Hörmann S, Kidziski L, Hallin M. 2015a. Dynamic functional principal components. *Journal of the Royal Statistical Society, Series B* **77**: 319–348.

Hörmann S, Kidziski L, Kokoszka P. 2015b. Estimation in functional lagged regression. *Journal of Time Series Analysis* **36**: 541–561.

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

Hörmann S, Kokoszka P. 2012. Functional time series. In *Time Series. Handbook of Statistics*, Vol. 30, Rao CR, Rao TS (eds.). Amsterdam: Elsevier.

Horváth L, Hušková M, Rice G. 2013. Testing independence for functional data. *Journal of Multivariate Analysis* **117**: 100–119.

Horváth L, Kokoszka P, Reeder R. 2012. Estimation of the mean of functional time series and a two sample problem. *Journal of the Royal Statistical Society (B)*: 74.

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

Horváth L, Rice G. 2015. Testing for independence between two functional time series. *Journal of Econometrics* **189**: 371–382.

Hywood J, Read MN, Rice G. 2016. Statistical analysis of spatially homogeneous dynamic agent-based processes using functional time series analysis. *Spatial Statistics* **17**: 199–219.

Jirak M. 2013. On weak invariance principles for sums of dependent random functionals. *Statistics and Probability Letters* **83**: 2291–2296.

Kargin V, Onatski A. 2008. Curve forecasting by functional autoregression. *Journal of Multivariate Analysis* **99**: 2508–2526.

Kokoszka P, Reimherr M. 2013. Asymptotic normality of the principal components of functional time series. *Stochastic Processes and their Applications* **123**: 1546–1562.

Müller H-G, Sen R, Stadtmüller U. 2011. Functional data analysis for volatility. *Journal of Econometrics* **165**: 233–245.

Panaretos VM, Kraus D, Maddocks JH. 2010. Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *Journal of the American Statistical Association* **105**: 670–682.

Panaretos V. M., Tavakoli S. 2013a. Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series. *Stochastic Processes and their Applications* **123**(7): 2779–2807.

Panaretos VM, Tavakoli S. 2013b. Fourier analysis of stationary time series in function space. *Annals of Statistics* **41**(2): 568–603.

Paparoditis E, Sapatinas T. 2016. Bootstrap-based *k* sample testing for functional data. *Biometrika* **103**: 727–733.

Petersen A, Müller H-G. 2016. Fréchet integration and adaptive metric selection for interpretable covariance of multivariate functional data. *Biometrika* **103**: 103–120.

Pham T, Panaretos V. 2017. Methodology and convergence rates for functional time series regression. Technical Report.

Pigoli D, Aston J, Dryden I, Secchi P. 2014. Distances and inference for covariance operators. *Biometrika* **101**: 409–422.

Priestley MB. 1981. *Spectral Analysis and Time Series*. New York: Academic Press.

R Development Core Team. 2015. *R: A Language and Environment for Statistical Computing. R Package Version 2.2.0*. Vienna: R Foundation for Statistical Computing.

Ramsay JO, Silverman BW. 2005. *Functional Data Analysis*. New York: Springer.

Reimherr M. 2015. Functional regression with repeated eigenvalues. *Statistics and Probability Letters* **107**: 62–70.

Rice G, Shang HL. 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.

Serban N, Staicu A-M, Carroll R-J. 2013. Multilevel cross-dependent binary longitudinal data. *Biometrics* **69**: 903–913.

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

Tavakoli S, Panaretos VM. 2016. Detecting and localizing differences in functional time series dynamics: a case study in molecular biophysics. *Journal of the American Statistical Association* **111**(515): 1020–1035.

Torgovitski L. 2016. *Darling–Erdös-type CUSUM-procedure for Functional Data II*: University of Cologne. Working paper.

van den Boogaart K. G. 2010. *tensorA: Advanced tensors arithmetic with named indices. R package version 0.36.* https://CRAN.R-project.org/package=tensorA.

Wang Y, Zou J. 2010. Vast volatility matrix estimation for high-frequency financial data. *The Annals of Statistics* **38**: 953–978.

Wied D, Arnold M, Bissantz N, Ziggel D. 2012a. A new fluctuation test for constant variances with applications to finance. *Metrika* **75**: 1111–1127.

Wied D, Krämer W, Dehling H. 2012b. Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory* **28**: 570–589.

Zhang X. 2016. White noise testing and model diagnostic checking for functional time series. *Journal of Econometrics* **194**: 76–95.

Zhang X, Shao X. 2015. Two sample inference for the second-order property of temporally dependent functional data. *Bernoulli* **21**: 909–929.

Zhou L, Huang J, Carroll R-J. 2008. Joint modelling of paired sparse functional data using principal components. *Biometrika* **95**: 601–619.

APPENDIX A. PROOFS OF TECHNICAL RESULTS

A.1. Proof of Theorem 2.1

To prove Theorem 2.1, we may assume without loss of generality that $\mu_X = \mu_Y = 0$. Let

$$\tilde{C}_{XY}(t,s,x) = \frac{1}{T}\sum_{i=1}^{\lfloor Tx \rfloor} X_i(t)Y_i(s).$$

DOI: 10.1111/jtsa.12447

Lemma A.1. Under Assumption 2.1,

$$\sup_{0\le x\le 1} \|\hat{C}_{XY}(\cdot,\cdot,x) - \tilde{C}_{XY}(\cdot,\cdot,x)\|_2 = O_P\left(\frac{1}{T}\right).$$

Proof. According to the definitions of $\hat{C}_{XY}$ and $\tilde{C}_{XY}$ and the triangle inequality,

$$\begin{aligned}\|\hat{C}_{XY}(\cdot,\cdot,x) - \tilde{C}_{XY}(\cdot,\cdot,x)\|_2 &= \frac{1}{T}\left\|\sum_{i=1}^{\lfloor Tx\rfloor}(X_i-\bar{X})(Y_i-\bar{Y}) - X_iY_i\right\|_2 \\ &= \frac{1}{T}\left\|\sum_{i=1}^{\lfloor Tx\rfloor} -\bar{X}Y_i - X_i\bar{Y} + \bar{X}\bar{Y}\right\|_2 \\ &\le \frac{1}{T}\left\|\sum_{i=1}^{\lfloor Tx\rfloor}\bar{X}Y_i\right\|_2 + \frac{1}{T}\left\|\sum_{i=1}^{\lfloor Tx\rfloor}X_i\bar{Y}\right\|_2 + \frac{1}{T}\left\|\sum_{i=1}^{\lfloor Tx\rfloor}\bar{X}\bar{Y}\right\|_2 \\ &= G_1(x) + G_2(x) + G_3(x).\end{aligned} \tag{A1}$$

For the term $G_1(x)$, we have that

$$G_1(x) = \frac{1}{T}\left\|\sum_{i=1}^{\lfloor Tx\rfloor}\bar{X}Y_i\right\|_2 = \frac{1}{T}\|\bar{X}\|_1\left\|\sum_{i=1}^{\lfloor Tx\rfloor}Y_i\right\|_1. \tag{A2}$$

Assumption 2.1 implies that both the series X_i and Y_i satisfy the conditions of Theorem 3.3 of Berkes *et al.* (2013), from which it follows that

$$\|\bar{X}\|_1 = O_P\left(\frac{1}{\sqrt{T}}\right), \text{ and } \sup_{0\le x\le 1}\left\|\sum_{i=1}^{\lfloor Tx\rfloor}Y_i\right\|_1 = O_P\left(\sqrt{T}\right).$$

This combined with (A2) implies that $\sup_{0\le x\le 1} G_1(x) = O_P(1/T)$. Parallel arguments show that $\sup_{0\le x\le 1} G_2(x) = O_P(1/T)$ and $\sup_{0\le x\le 1} G_3(x) = O_P(1/T)$, from which the result follows in light of (A1). □

Lemma A.2. Under Assumption 2.1 and if $q = p/2$ with p defined in Assumption 2.1, then

$$(E\|X_iY_i - X_{i,m}Y_{i,m}\|_2^q)^{1/q} = O(m^{-\gamma}),$$

where $\gamma = \min(\alpha,\beta)$, and $X_{i,m}$ and $Y_{i,m}$ are defined in Assumption 2.1.

Proof. We have according to the triangle inequalities in $L^2[0,1]^2$ and for $(E(\cdot)^q)^{1/q}$ that

$$\begin{aligned}\left(E\|X_iY_i - X_{i,m}Y_{i,m}\|_2^q\right)^{1/q} &= \left(E\|X_iY_i - X_iY_{i,m} + X_{i,m}Y_i - X_{i,m}Y_{i,m}\|_2^q\right)^{1/q} \\ &\le \left(E\left(\|X_iY_i - X_iY_{i,m}\|_2 + \|X_{i,m}Y_i - X_{i,m}Y_{i,m}\|_2\right)^q\right)^{1/q} \\ &\le \left(E\|X_iY_i - X_iY_{i,m}\|_2^q\right)^{1/q} + \left(E\|X_{i,m}Y_i - X_{i,m}Y_{i,m}\|_2^q\right)^{1/q} \\ &= \left(E\|X_i\|_1^q\|Y_i - Y_{i,m}\|_1^q\right)^{1/q} + \left(E\|Y_i\|_1^q\|X_i - X_{i,m}\|_1^q\right)^{1/q}.\end{aligned} \tag{A3}$$

According to the Cauchy–Schwarz inequality and stationarity, we have that

$$E\|X_i\|_1^q\|Y_i - Y_{i,m}\|_1^q \le \left(E\|X_i\|_1^{2q}\right)^{1/2}E\left(\|Y_i - Y_{i,m}\|_1^{2q}\right)^{1/2} = \left(E\|X_0\|_1^{2q}\right)^{1/2}E\left(\|Y_0 - Y_{0,m}\|_1^{2q}\right)^{1/2}. \tag{A4}$$

One obtains a similar bound with the roles of X and Y swapped, which combined with the last line of (A3) we get that (with $p = 2q$)

$$\left(E\|X_iY_i - X_{i,m}Y_{i,m}\|_2^q\right)^{1/q} \leq \left(E\|X_0\|_1^p\right)^{1/p}\left(E\|Y_0 - Y_{0,m}\|_1^p\right)^{1/p} + \left(E\|Y_0\|_1^p\right)^{1/p}\left(E\|X_0 - X_{0,m}\|_1^p\right)^{1/p}.$$

Since both $(E\|X_0\|_1^p)^{1/p}$ and $(E\|Y_0\|_1^p)^{1/p}$ are finite, and according to Assumption 2.1 $\left(E\|X_i - X_{i,m}\|^p\right)^{1/p} = O(m^{-\alpha})$, and $\left(E\|Y_i - Y_{i,m}\|^p\right)^{1/p} = O(m^{-\beta})$, we obtain that

$$\left(E\|X_iY_i - X_{i,m}Y_{i,m}\|_2^q\right)^{1/q} = O(m^{-\gamma}),$$

where $\gamma = \min(\alpha, \beta)$ as needed. □

To complete the proof of Theorem 2.1, we employ Theorem 1.1 of Berkes *et al.* (2013) (cf. Theorem 1.2 in Jirak, 2013) applied to stochastic process with sample paths in $L^2[0,1]^2$. The original Theorem 1.1 in Berkes *et al.* (2013) is given for random functions with sample paths in $L^2[0,1]$ and the following extension is simple to obtain from this. We present this result here for ease of reference:

Theorem 1.1 of Berkes et al. (2013) in $L^2[0,1]^2$ Consider a mean zero strictly stationary sequence of stochastic processes $\{\xi_i(t,s),\ t,s \in [0,1],\ i \in \mathbb{Z}\}$. Suppose there exists a function $f : S^\infty \to L^2[0,1]^2$, and i.i.d. innovations $\{\epsilon_i,\ : i \in \mathbb{Z}\}$ taking values in S such that $\xi_i = f(\epsilon_i, \epsilon_{i-1}, \ldots)$. Suppose $\xi_{i,m}$ is defined as in Assumption 2.1(b), and let $v_m^{(p,q)} = (E\|\xi_0 - \xi_{0,m}\|_2^q)^{1/p}$. If there exists $q > p > 2$ so that

$$\sum_{m=1}^{\infty} v_m^{(p,q)} < \infty,$$

then with

$$S_T(t,s,x) = \frac{1}{\sqrt{T}} \sum_{i=1}^{\lfloor Tx \rfloor} \xi_i(t,s), \quad t,s,x \in [0,1],$$

there exists a sequence of Gaussian processes $\{\Gamma_T(t,s,x),\ t,s,x \in [0,1],\ T \in \mathbb{N}\}$, (defined on the same but perhaps enlarged probability space) such that

$$\sup_{x\in[0,1]} \iint [S_T(t,s,x) - \Gamma_T(t,s,x)]^2 \mathrm{d}t\mathrm{d}s = o_P(1),$$

where for all T, $\Gamma_T(t,s,x)$ has mean zero, and covariance function

$$E[\Gamma_T(t,s,x)\Gamma_T(u,v,y)] = \min\{x,y\} \sum_{\ell=-\infty}^{\infty} E[\xi_0(t,s)\xi_\ell(u,v)].$$

Proof of Theorem 2.1. According to Lemma A.1, it is enough to prove Theorem 2.1 for the process $\tilde{C}_{XY}(t,s,x)$. $\tilde{C}_{XY}(t,s,x) - (\lfloor Tx \rfloor/T)C_{XY}(t,s)$ is the partial sum of the random functions $\xi_i(t,s) = X_i(t)Y_i(s) - EX_0(t)Y_0(s)$ which form a mean zero and stationary sequence in $L^2[0,1]^2$. Let $\xi_{i,m}(t,s) = X_{i,m}(t)Y_{i,m}(s) - EX_0(t)Y_0(s)$, denote an m-dependent approximation to ξ_i. It follows directly from A.2 that

$$\left(E\|\xi_i - \xi_{i,m}\|_2^q\right)^{1/q} = O(m^{-\gamma}),$$

where $\gamma > 1$, hence for some $p < q$,

$$\left(E\|\xi_i - \xi_{i,m}\|_2^q\right)^{1/p} = O(m^{-\zeta}),$$

DOI: 10.1111/jtsa.12447

where $\zeta > 1$. This gives that

$$\sum_{m=1}^{\infty} \left(E\|\xi_i - \xi_{i,m}\|_2^q\right)^{1/p} < \infty.$$

Therefore the sequence $\{\xi_i\}$ satisfies the conditions of Theorem 1.1 of Berkes *et al.* (2013), and hence with

$$\Xi_T(t,s,x) := \sqrt{T}\left(\tilde{C}_{XY}(t,s,x) - \frac{\lfloor Tx \rfloor}{T} C_{XY}(t,s)\right) = \frac{1}{\sqrt{T}} \sum_{i=1}^{\lfloor Tx \rfloor} \xi_i(t,s),$$

then

$$\sup_{0\le x\le 1} \iint \left(\Xi_T(t,s,x) - \Gamma_T(t,s,x)\right)^2 \mathrm{d}t\mathrm{d}s = o_P(1), \tag{A5}$$

where Γ_T has the properties attributed in Theorem 2.1, which establishes the result.

□

To see how Corollaries 3.1–3.3 follow from this result, we note that for all T,

$$\{\Gamma_T(t,s,x) : \ t,s,x \in [0,1]\} \stackrel{D}{=} \{\Gamma_0(t,s,x) : \ t,s,x \in [0,1]\},$$

where

$$\Gamma_0(t,s,x) = \sum_{\ell=1}^{\infty} \lambda_\ell^{1/2} W_\ell(x) \varphi_\ell(t,s),$$

with $\{W_\ell(t),\ t \in [0,1],\ \ell \in \mathbb{N}\}$ being i.i.d. standard Brownian motions, and $(\lambda_\ell\ \varphi_\ell)$ being the eigenvalues and eigenfunctions of the operator $\mathfrak{d}$. This follows from simply calculating the mean and covariance function of Γ_0 and using Mercer's theorem.

Proof of Theorem 3.1. The first part of the theorem follows directly from the ergodic theorem in Hilbert spaces; see for example Appendix A of Horváth *et al.* (2013). To get the second result, we have using the assumption that $\sqrt{T}(C_{XY} - C_0) \stackrel{L^2[0,1]^2}{\to} C_A$ and Theorem 2.1 that, in case of F_T,

$$\begin{aligned}
F_T &= T\|\hat{C}_{XY}(\cdot,\cdot,1) - C_{XY} + C_{XY} - C_0\|_2^2 \\
&= T\langle(\hat{C}_{XY}(\cdot,\cdot,1) - C_{XY}) + (C_{XY} - C_0), (\hat{C}_{XY}(\cdot,\cdot,1) - C_{XY}) + (C_{XY} - C_0)\rangle_2 \\
&= T\|\hat{C}_{XY}(\cdot,\cdot,1) - C_{XY}\|_2^2 + 2\langle\sqrt{T}(\hat{C}_{XY}(\cdot,\cdot,1) - C_{XY}), \sqrt{T}(C_{XY} - C_0)\rangle_2 + T\|C_{XY} - C_0\|_2^2 \\
&\stackrel{\mathcal{D}}{\to} \sum_{i=1}^{\infty} \lambda_i \mathcal{N}_i^2 + 2\sum_{i=1}^{\infty} \lambda_i^{1/2} \langle C_A, \varphi_i\rangle \mathcal{N}_i + \|C_A\|_2^2, \quad (T \to \infty),
\end{aligned}$$

as needed. The result for $F_{T,p}$ follows similarly. □

A.2. Proof of Theorem 4.1

Let

$$\tilde{D}_T(t,s,u,v) = \sum_{\ell=-\infty}^{\infty} W_b\left(\frac{\ell}{B}\right) \tilde{\gamma}_\ell(t,s,u,v), \tag{A6}$$

with B defined in (4.2), W_b being a bounded, symmetric, and continuous weight function of order b, and

$$\tilde{\gamma}_\ell(t,s,u,v) = \begin{cases} \frac{1}{T}\sum_{j=1}^{T-\ell}\left(X_j(t)Y_j(s) - EX_0(t)Y_0(s)\right)\left(X_{j+\ell}(u)Y_{j+\ell}(v) - EX_0(u)Y_0(v)\right), & \ell \geq 0 \\ \frac{1}{T}\sum_{j=1-\ell}^{T}\left(X_j(t)Y_j(s) - EX_0(u)Y_0(v)\right)\left(X_{j+\ell}(t)Y_{j+\ell}(s) - EX_0(u)Y_0(v)\right), & \ell < 0. \end{cases}$$

Lemma A.3. Under Assumption 2.1,

$$\|\hat{D} - \tilde{D}\|_4 = o_P(1).$$

Proof. According to the definition of $\hat{\gamma}_\ell$, when $\ell \geq 0$, we have that

$$\begin{aligned} \hat{\gamma}_\ell(t,s,u,v) &= \frac{1}{T}\sum_{j=1}^{T-\ell}(\bar{X}_j(t)\bar{Y}_j(s) - \hat{C}_{XY}(t,s,1))(\bar{X}_{j+\ell}(u)\bar{Y}_{j+\ell}(v) - \hat{C}_{XY}(u,v,1)) \\ &= \frac{1}{T}\sum_{j=1}^{T-\ell}\Bigg((X_j(t)Y_j(s) - EX_0(t)Y_0(s)) - \bar{X}(t)Y_j(s) + \bar{X}(t)\bar{Y}(s) - X_j(t)\bar{Y}(s) \\ &\quad + (EX_0(t)X_0(s) - \hat{C}(t,s,1))\Bigg) \times \Bigg(X_{j+\ell}(u)Y_{j+\ell}(v) - EX_0(u)Y_0(v) - \bar{X}(u)Y_{j+\ell}(v) \\ &\quad + \bar{X}(u)\bar{Y}(v) - X_{j+\ell}(u)\bar{Y}(v) + (EX_0(u)X_0(v) - \hat{C}(u,v,1))\Bigg) \\ &= \tilde{\gamma}(t,s,u,v) + \sum_{p=1}^{24} R_{p,\ell}(t,s,u,v), \end{aligned} \tag{A7}$$

where the terms $R_{p,\ell}$ represent the remaining 24 terms in the definition of $\hat{\gamma}_\ell$ obtained by completing the multiplication that results in 25 terms, the first of which corresponds to $\tilde{\gamma}_\ell$. We now aim to show that for all p and ℓ, $E\|R_{p,\ell}\|_4 = O(T^{-1/2})$. For $R_{1,\ell}$, we have by the triangle inequality, a couple of applications of the Cauchy-Schwarz inequality, and the assumed stationarity that

$$\begin{aligned} E\|R_{1,\ell}\|_4 &= \frac{1}{T}E\left\|\sum_{j=1}^{T-\ell}(X_j(t)Y_j(s) - EX_0(t)Y_0(s))\bar{X}(u)Y_{j+\ell}(v)\right\|_4 \\ &\leq \frac{1}{T}\sum_{j=1}^{T-\ell}E\|(X_j(t)Y_j(s) - EX_0(t)Y_0(s))\bar{X}(u)Y_{j+\ell}(v)\|_4 \\ &= \frac{1}{T}\sum_{j=1}^{T-\ell}E\|(X_j(t)Y_j(s) - EX_0(t)Y_0(s))\|_2\|\bar{X}\|_1\|Y_{j+\ell}\|_1 \\ &\leq \frac{1}{T}\sum_{j=1}^{T-\ell}\left(E\|(X_j(t)Y_j(s) - EX_0(t)Y_0(s))\|_2^2\right)^{1/2}\left(E(\|\bar{X}\|_1\|Y_{j+\ell}\|_1)^2\right)^{1/2} \\ &\leq \frac{1}{T}\sum_{j=1}^{T-\ell}\left(E\|(X_j(t)Y_j(s) - EX_0(t)Y_0(s))\|_2^2\right)^{1/2}\left(E\|\bar{X}\|_1^4\right)^{1/4}\left(E\|Y_{j+\ell}\|_1^4\right)^{1/4} \\ &= \frac{T-\ell}{T}\left(E\|(X_0(t)Y_0(s) - EX_0(t)Y_0(s))\|_2^2\right)^{1/2}\left(E\|\bar{X}\|_1^4\right)^{1/4}\left(E\|Y_0\|_1^4\right)^{1/4}. \end{aligned} \tag{A8}$$

  J. Time Ser. Anal. **40**: 665–692 (2019)
DOI: 10.1111/jtsa.12447

It follows from Proposition 3.1 of Torgovitski (2016) that under Assumption 2.1, $(E\|\bar{X}\|_1^4)^{1/4} = O(T^{-1/2})$, and also according to Assumption 2.1, $(E\|Y_0\|_1^4)^{1/4} = O(1)$ and $(E\|(X_0(t)Y_0(s) - EX_0(t)Y_0(s))\|_2^2)^{1/2} = O(1)$. This implies with (A8) that $E\|R_{1,\ell}\|_4 = O(T^{-1/2})$. Similar arguments using the fact that $(E\|\bar{Y}\|_1^4)^{1/4} = O(T^{-1/2})$ and $\|EX_0(t)Y_0(s) - \hat{C}_{XY}(t,s,1)\|_2 = O(T^{-1/2})$ can be applied to get that $E\|R_{p,\ell}\| = O(T^{-1/2})$, $2 \le p \le 24$. This implies via the triangle inequality and (A7) that

$$E\|\hat{\gamma}_\ell - \tilde{\gamma}_\ell\|_4 = O(T^{-1/2}),$$

from which it follows that

$$E\left\|\sum_{\ell=1}^{\infty} W_b\left(\frac{\ell}{B}\right)(\hat{\gamma}_\ell - \tilde{\gamma}_\ell)\right\|_4 \le \sum_{\ell=1}^{\infty} W_b\left(\frac{\ell}{B}\right)E\|(\hat{\gamma}_\ell - \tilde{\gamma}_\ell)\|_4 = O(hT^{-1/2}) = o(1), \tag{A9}$$

by (4.2) and the fact that W_b has bounded support. A parallel argument may be used to show that for $\ell < 0$, that $E\|\hat{\gamma}_\ell - \tilde{\gamma}_\ell\|_4 = O(T^{-1/2})$, from which we obtain that

$$E\left\|\sum_{\ell=-\infty}^{-1} W_b\left(\frac{\ell}{B}\right)(\hat{\gamma}_\ell - \tilde{\gamma}_\ell)\right\|_4 = O(hT^{-1/2}) = o(1). \tag{A10}$$

(A9) and (A10) together with the triangle inequality imply that

$$E\|\hat{D} - \tilde{D}\|_4 \le E\left\|\sum_{\ell=1}^{\infty} W_b\left(\frac{\ell}{B}\right)(\hat{\gamma}_\ell - \tilde{\gamma}_\ell)\right\|_4 + E\left\|\sum_{\ell=-\infty}^{-1} W_b\left(\frac{\ell}{B}\right)(\hat{\gamma}_\ell - \tilde{\gamma}_\ell)\right\|_4 = o(1),$$

and this implies the result with Markov's inequality. □

Lemma A.4. Under Assumption 2.1,

$$\|\tilde{D} - D\|_4 = o_P(1).$$

Proof. To show this, we use Theorem 2.2 of Horváth *et al.* (2012). Noting that $\tilde{\gamma}_\ell$ is the autocovariance estimator based on a sample of size T at lag ℓ of the mean zero random functions $\{\xi_i(t,s),\ i \in \mathbb{Z}\}$, where, recalling from above, $\xi_i(t,s) = X_i(t)Y_i(s) - EX_0(t)Y_0(s)$. With $\xi_{i,m}(t,s) = X_{i,m}(t)Y_{i,m}(s) - EX_0(t)Y_0(s)$, Theorem 2.2 of Horváth *et al.* (2012) states that Lemma A.4 holds if both,

1. $\lim_{m\to\infty} m(E\|\xi_i - \xi_{i,m}\|_2^2)^{1/2} = 0$, and
2. $$\sum_{m=1}^{\infty}\left(E\|\xi_i - \xi_{i,m}\|_2^2\right)^{1/2} < \infty.$$

These conditions both follow from Lemma A.2 and Lyapounovs inequality.

□

Proof of Theorem 4.1. The Theorem follows directly from Lemmas A.3 and A.4. □

A.3. Consistency of Z_T and $Z_{T,p}$

To show that the tests based on Z_T and $Z_{T,p}$ are consistent, we require that the series before and after the change are stationary and are of the form outlined in Assumption (2.1).

Assumption A.1. Under $H_{A,2}$, we assume that there exist measurable functions $g^{(1)}_{XY} : S^\infty \to L^2[0,1] \times L^2[0,1]$ and $g^{(2)}_{XY} : S^\infty \to L^2[0,1] \times L^2[0,1]$, where S is a measurable space, and a sequence of i.i.d. innovations $\{\epsilon_i, : i \in \mathbb{Z}\}$ taking values in S such that $(X_i, Y_i) = g^{(1)}_{XY}(\epsilon_i, \epsilon_{i-1}, \ldots)$ for $i \le k^*$, and $(X_i, Y_i) = g^{(2)}_{XY}(\epsilon_i, \epsilon_{i-1}, \ldots)$ for $i > k^*$, such that, for each i, (X_i, Y_i) satisfies Assumption 2.1 (b), and $H_{A,2}$ holds.

Assumption A.1 holds under a number of conceivable models for observations (X_i, Y_i) following $H_{A,2}$. For instance, if

$$X_i(t) = \alpha_i \varepsilon_{c,i}(t) + (1-\alpha_i)\varepsilon_{x,i}(t), \quad Y_i(t) = \alpha_i \varepsilon_{c,i}(t) + (1-\alpha_i)\varepsilon_{y,i}(t),$$

where the $\varepsilon'_{\cdot,i}s$ are i.i.d. functional innovations (or themselves approximable Bernoulli shifts), and $\alpha_1 = \cdots = \alpha_{k^*} \ne \alpha_{k^*+1} = \cdots = \alpha_{T^*+1}$, then Assumption A.1 holds. Interestingly, as mentioned in Section 3, no orthogonality condition is needed in order for the test statistic $Z_{T,p}$ to diverge since, under $H_{A,2}$, the estimates for φ_i in (4.4) converge into the direction of $C_1 - C_2$ due to a diverging in magnitude rank one perturbation of the operator $\hat{\mathfrak{d}}$ under $H_{A,2}$.

Theorem A.1. Under $H_{A,2}$ and Assumption A.1 holds, then

$$Z_T \xrightarrow{P} \infty.$$

If in addition the estimates $\hat{\lambda}_i$ and $\hat{\varphi}_i$ are defined according to (4.4), then

$$Z_{T,p} \xrightarrow{P} \infty.$$

We provide here a shortened version of the proof, since the notation is quite tedious to develop in full detail, and some similar calculations are given in the proof of A.3.

Outline of proof. It follows from Assumption A.1 and the conditions on the bandwidth B in (4.2) that

$$\left\| \frac{1}{k^*} \sum_{i=1}^{k^*} (X_i(t) - \mu_X(t))(Y_i(s) - \mu_Y(s)) - C_1(t,s) \right\|_2 = o_P(h/\sqrt{T}),$$

and

$$\left\| \frac{1}{T-k^*} \sum_{i=k^*+1}^{T} (X_i(t) - \mu_X(t))(Y_i(s) - \mu_Y(s)) - C_2(t,s) \right\|_2 = o_P(h/\sqrt{T}).$$

From this it is straightforward to establish the first part of Theorem A.1, and so we instead focus on the second part, which is more tedious. First we must show that, as $T \to \infty$, $\hat{\varphi}_1$ asymptotically tends to go into the direction of $C_1 - C_2$. Recalling that under $H_{A,2}$, $k^* = \lfloor T\theta \rfloor$, we may obtain from the above two bounds along the lines of the proof of Lemma A.3 that if

$$\gamma^*_\ell(t,s,u,v) = \begin{cases} \frac{1}{T}\sum_{j=1}^{T-\ell}\Big(X_j(t)Y_j(s) - (\theta C_1(t,s) + (1-\theta)C_2(t,s))\Big) \\ \qquad \times\Big(X_{j+\ell}(u)Y_{j+\ell}(v) - (\theta C_1(u,v) + (1-\theta)C_2(u,v))\Big), & \ell \ge 0 \\ \frac{1}{T}\sum_{j=1-\ell}^{T}\Big(X_j(t)Y_j(s) - (\theta C_1(t,s) + (1-\theta)C_2(t,s)\Big) \\ \qquad \times\Big(X_{j+\ell}(u)Y_{j+\ell}(v) - (\theta C_1(u,v) + (1-\theta)C_2(u,v))\Big), & \ell < 0. \end{cases}$$

DOI: 10.1111/jtsa.12447

and

$$D_T^*(t,s,u,v) = \sum_{\ell=-\infty}^{\infty} W_b\left(\frac{\ell}{B}\right)\gamma_\ell^*(t,s,u,v), \tag{A11}$$

then $\|\hat{D} - D^*\| = o_P(1)$. Suppose $\ell \geq 0$. We may then write

$$\gamma_\ell^*(t,s,u,v) = \sum_{i=1}^{3} U_{\ell,T}^{(i)}(t,s,u,v),$$

where

$$U_{\ell,T}^{(1)}(t,s,u,v) = \frac{1}{T}\sum_{j=1}^{k^*-\ell}\Big(X_j(t)Y_j(s) - (\theta C_1(t,s) + (1-\theta)C_2(t,s))\Big) \times \Big(X_{j+\ell}(u)Y_{j+\ell}(v) - (\theta C_1(u,v) + (1-\theta)C_2(u,v))\Big),$$

$$U_{\ell,T}^{(2)}(t,s,u,v) = \frac{1}{T}\sum_{j=k^*-\ell+1}^{k^*}\Big(X_j(t)Y_j(s) - (\theta C_1(t,s) + (1-\theta)C_2(t,s))\Big) \times \Big(X_{j+\ell}(u)Y_{j+\ell}(v) - (\theta C_1(u,v) + (1-\theta)C_2(u,v))\Big),$$

$$U_{\ell,T}^{(3)}(t,s,u,v) = \frac{1}{T}\sum_{j=k^*}^{T-\ell}\Big(X_j(t)Y_j(s) - (\theta C_1(t,s) + (1-\theta)C_2(t,s))\Big) \times \Big(X_{j+\ell}(u)Y_{j+\ell}(v) - (\theta C_1(u,v) + (1-\theta)C_2(u,v))\Big).$$

In each term in the summand defining $U_{\ell,T}^{(1)}$, we may add and subtract $C_1(t,s)$ and $C_1(u,v)$ respectively, and then by completing the multiplication in the summand of $U_{\ell,T}^{(1)}$, we get that $U_{\ell,T}^{(1)}(t,s,u,v) = \sum_{i=1}^{4} V_{\ell,T}^{(1,i)}(t,s,u,v)$, where

$$V_{\ell,T}^{(1,1)}(t,s,u,v) = \frac{(1-\theta)^2(k^*-\ell)}{T}(C_1(t,s) - C_2(t,s))(C_1(u,v) - C_2(u,v)),$$

$$V_{\ell,T}^{(1,2)}(t,s,u,v) = (1-\theta)(C_1(t,s) - C_2(t,s))\frac{1}{T}\sum_{j=1}^{k^*-\ell}(X_{j+\ell}(u)Y_{j+\ell}(v) - C_1(u,v)),$$

$$V_{\ell,T}^{(1,3)}(t,s,u,v) = (1-\theta)(C_1(u,v) - C_2(u,v))\frac{1}{T}\sum_{j=1}^{k^*-\ell}(X_j(t)Y_j(s) - C_1(t,s)),$$

and

$$V_{\ell,T}^{(1,4)}(t,s,u,v) = \frac{1}{T}\sum_{j=1}^{k^*-\ell}(X_j(t)Y_j(s) - C_1(t,s))(X_{j+\ell}(u)Y_{j+\ell}(v) - C_1(u,v)).$$

DOI: 10.1111/jtsa.12447

Combining these decompositions, we have that

$$\sum_{\ell=-\infty}^{\infty} W_b\left(\frac{\ell}{B}\right) U^{(1)}_{\ell,T}(t,s,u,v) = \sum_{i=1}^{4}\sum_{\ell=-\infty}^{\infty} W_b\left(\frac{\ell}{B}\right) V^{(1,i)}_{\ell,T}(t,s,u,v), \tag{A12}$$

and for the first term,

$$\sum_{\ell=-\infty}^{\infty} W_b\left(\frac{\ell}{B}\right) V^{(1,1)}_{\ell,T}(t,s,u,v) = (1-\theta)^2(C_1(t,s)-C_2(t,s))(C_1(u,v)-C_2(u,v))\sum_{\ell=-\infty}^{\infty}\frac{(k^*-\ell)}{T}W_b\left(\frac{\ell}{B}\right).$$

According to the definitions of k^*, W_b and B, $\sum_{\ell=-\infty}^{\infty}((k^*-\ell)/T)W_b(\ell/B) = O(h)$, and so

$$\left\|(1-\theta)^2(C_1(t,s)-C_2(t,s))(C_1(u,v)-C_2(u,v))\sum_{\ell=-\infty}^{\infty}\frac{(k^*-\ell)}{T}W_b\left(\frac{\ell}{B}\right)\right\|_4 = O(h).$$

One can show using Assumption A.1 that the remaining terms in (A12) are of lower asymptotic order in norm. Similarly with $U^{(3)}_{\ell,T}$ in place of $U^{(1)}_{\ell,T}$ in (A12), and following a similar decomposition, we get that the leading order term is of the form $(C_1(t,s)-C_2(t,s))(C_1(u,v)-C_2(u,v))$ multiplied by a term on the order of B. The corresponding term with $U^{(2)}_{\ell,T}$ is asymptotically negligible. Since $\hat{\varphi}_1$ must maximize the quadratic form as a function of v:

$$\iiiint \hat{D}_T(t,s,u,v)v(t,s)v(u,v)\mathrm{d}t\mathrm{d}s\mathrm{d}u\mathrm{d}v,$$

it follows that

$$\|\hat{\varphi}_1 - s_1(C_1-C_2)/\|C_1-C_2\|_2\|_2 = o_P(1),$$

where s_1 is 1 or -1, and $\hat{\lambda}_1 = O(h)$. Therefore, again using Assumption A.1

$$\begin{aligned} Z_{T,p} &= T\sup_{0\le x\le 1}\sum_{i=1}^{p}\frac{\langle \hat{C}_{XY}(\cdot,\cdot,x) - x\hat{C}_{XY}(\cdot,\cdot,1), \hat{\varphi}_i\rangle_2^2}{\hat{\lambda}_i} \\ &\ge T\sum_{i=1}^{p}\frac{\langle \hat{C}_{XY}(\cdot,\cdot,\theta) - \theta\hat{C}_{XY}(\cdot,\cdot,1), \hat{\varphi}_i\rangle_2^2}{\hat{\lambda}_i} \\ &\ge T\frac{\langle \hat{C}_{XY}(\cdot,\cdot,\theta) - \theta\hat{C}_{XY}(\cdot,\cdot,1), \hat{\varphi}_1\rangle_2^2}{\hat{\lambda}_1} \\ &= T\frac{\langle \theta(1-\theta)(C_1-C_2),(C_1-C_2)\rangle_2^2}{\hat{\lambda}_1} + o\left(\frac{T}{B}\right)\xrightarrow{P}\infty, \end{aligned}$$

giving the result. □

  J. Time Ser. Anal. **40**: 665–692 (2019)
DOI: 10.1111/jtsa.12447