Testing relevant hypotheses in functional time series via self-normalization

Holger Dette and Kevin Kokot
Ruhr-Universität Bochum, Germany

and Stanislav Volgushev
University of Toronto, Canada

[Received September 2018. Final revision March 2020]

Summary. We develop methodology for testing relevant hypotheses about functional time series in a tuning-free way. Instead of testing for exact equality, e.g. for the equality of two mean functions from two independent time series, we propose to test the null hypothesis of no relevant deviation. In the two-sample problem this means that an $L^2$-distance between the two mean functions is smaller than a prespecified threshold. For such hypotheses self-normalization, which was introduced in 2010 by Shao, and Shao and Zhang and is commonly used to avoid the estimation of nuisance parameters, is not directly applicable. We develop new self-normalized procedures for testing relevant hypotheses in the one-sample, two-sample and change point problem and investigate their asymptotic properties. Finite sample properties of the tests proposed are illustrated by means of a simulation study and data examples. Our main focus is on functional time series, but extensions to other settings are also briefly discussed.

Keywords: Change point analysis; Cumulative sum; Functional time series; Relevant hypotheses; Self-normalization; Two-sample problems

1. Introduction

Statistics for functional data has found considerable interest in the last 20 years as documented in the various monographs by Ramsay and Silverman (2005), Ferraty and Vieu (2010) and Horváth and Kokoszka (2012) among others. The available methodology includes explorative tools such as shift and feature registration, warping or principal components and methods for statistical inference such as testing of hypotheses and change point analysis. In this context a large portion of the literature attacks the problem of hypotheses testing by considering hypotheses of the form

$$H_0 : d = 0 \quad \text{versus} \quad H_1 : d \neq 0$$

(1.1)

where $d$ is a real-valued parameter such as the norm of the mean function in one sample or the norm of the difference of two mean functions or two covariance operators from two samples. For example Hall and Van Keilegom (2007) studied the effect of smoothing when converting discrete observations into functional data, Horváth et al. (2009) compared linear operators in two functional regression models and Benko et al. (2009) proposed functional principal component analysis for two-sample inference whereas Panaretos et al. (2010) and Fremdt et al. (2013)
considered a test for the equality of covariance operators. More recently Horváth et al. (2013) suggested tests for a comparison of two mean functions from temporally dependent curves under model-free assumptions and Pomann et al. (2016) compared the distributions of two samples by methods which are based on functional principal component analysis. Another important research area in functional data analysis is change point detection and we refer to Berkes et al. (2009), Hörmann and Kokoszka (2010), Aston and Kirch (2012), Zhang et al. (2011), Horváth et al. (2014) and Bucchia and Wendler (2017) among others who have investigated change point problems from various perspectives.

Several researchers have considered methods for independent data. In this case the quantiles for corresponding tests can be easily obtained by asymptotic theory as the unknown quantities in the limit distribution of the test statistics can be reliably estimated (e.g. the asymptotic variance of a standardized mean). However, for functional samples exhibiting temporal dependence, the asymptotic distribution of many commonly used tests involves the long-run variance, which makes the statistical inference substantially more difficult. Several researchers have proposed to estimate the long-run variance (see Kokoszka (2012) or Horváth et al. (2013) among others), but the commonly used estimators depend on regularization parameters. As an alternative, bootstrap methods can be applied to obtain critical values and we refer to Benko et al. (2009), Cuevas et al. (2006), Zhang et al. (2010), Bucchia and Wendler (2017) and Paparoditis and Sapatinas (2016) among many others. A third method to obtain (asymptotically) pivotal test statistics is the concept of self-normalization, which was introduced in the seminal papers of Shao (2010) for the construction of confidence intervals and Shao and Zhang (2010) for change point analysis. More recently it has been developed further for the specific needs of functional data by Zhang et al. (2011) and Zhang and Shao (2015) (see also Shao (2015) for a recent review).

This list of references is by no means complete but a common feature of all these references is that they usually address hypotheses of the form (1.1), which we call ‘classical’ hypotheses in the following discussion. However, in many applications one might not be interested in detecting very small deviations of the parameter $d$ from 0 (often the researcher even knows that $d$ is not exactly equal to 0, before any experiments have been carried out). For example, in change point detection a modification of the statistical analysis for prediction might not be necessary if the difference between the parameters before and after the change point is quite small. This discussion may be viewed as a particular case of the common bias–variance trade-off in statistics. Therefore we argue that one should carefully think about the size of the difference in which one is interested. In particular we propose to replace hypotheses (1.1) by the hypotheses of relevant differences, i.e.

$$H_0 : d \leq \Delta \quad \text{versus} \quad H_1 : d > \Delta,$$

where $\Delta$ is a prespecified constant representing the ‘maximal’ value for the parameter $d$, which can be accepted as not scientifically significant. If the null hypothesis in test (1.2) holds we speak of a null of no relevant difference. This formulation of the testing problem requires the specification of the threshold $\Delta > 0$, which depends on the specific application. Classical hypotheses tests simply use $\Delta = 0$, but we argue that from a practical point of view it might be very reasonable to think about this choice more carefully and to define the size of the change in which one is really interested from a scientific viewpoint.

We also note that the formulation of the testing problem in the form (1.2) avoids the consistency problem that was mentioned in Berkson (1938), i.e. any consistent test will detect any arbitrary small change in the parameters if the sample size is sufficiently large. Moreover, by interchanging the hypotheses, i.e. considering the hypotheses of equivalence,
we can decide on a ‘small parameter’ \( d \) at a controlled type I error (e.g. that the norm \( d \) of the difference between the mean functions of two samples is smaller than a given threshold). Hypotheses of the form (1.2) and (1.3) are called precise hypotheses or relevant hypotheses in the literature (see Berger and Delampady (1987)) and are frequently used in biostatistics. We refer to Chow and Liu (1992) and Wellek (2010) for more details and applications.

In this paper we discuss the problem of testing relevant hypotheses in the context of functional dependent data. We are particularly interested in methods based on self-normalization to avoid estimation of the long-run variance or resampling methods. The construction of efficient long-run variance estimates and resampling techniques is more difficult for testing relevant hypotheses, because—in contrast with classical hypotheses—the null hypothesis usually corresponds to an infinite dimensional set (e.g. the set of mean functions with squared \( L^2 \)-norm less than or equal to \( \Delta \)).

For this purpose we modify the classical approaches to self-normalization-based testing that was proposed by Shao (2010) and Shao and Zhang (2010) to make them applicable to testing relevant hypotheses. Zhang et al. (2011) and Zhang and Shao (2015) also used the concept of self-normalization to develop statistical methodology for functional data analysis. In particular they constructed tests for a change in the mean function and in the lag 1 autocovariance operator and for comparing the covariance operators and associated eigenvalues or eigenvectors from two samples. The main differences between their approach and the methods that are presented here are as follows. First, these references do not consider the problem of testing relevant hypotheses but deal with classical hypotheses of the form (1.1). Thus the present paper addresses a different statistical problem, where currently available methods are not applicable. Second, their approach is based on a dimension reduction projecting the functions on a finite dimensional vector (e.g. principal components), which is then used for the subsequent statistical inference by using common self-normalization techniques. In contrast with their work we can develop a self-normalized test for the problem of testing relevant hypotheses of the form (1.2), which does not require dimension reduction. For this purpose the common concepts of self-normalization must be further extended. This modification is of independent interest besides the field of functional data analysis and is applicable in many other problems.

The remaining part of this paper is organized as follows. Our basic idea is explained in Section 2 for the one- and two-sample case, where it is most transparent. Roughly speaking, we construct an asymptotic confidence interval for the parameter \( d \) to obtain tests for hypotheses of the form (1.2) and (1.3). In Section 3 we address the problem of relevant change point analysis by the new way of self-normalization; here an additional challenge arises from the fact that the change point location is unknown and needs to be estimated. Whereas the methodology in Section 2 and 3 refers to statistical inference for mean functions we illustrate in Section 4 how those ideas can be extended to inference for covariance operators. Some finite sample results are presented in Section 5, where we also illustrate the proposed methodology on two data examples. Here we also provide a brief discussion of self-normalization and estimation of the long-run variance in the context of testing relevant hypotheses. Finally, in an on-line supplement we present additional finite sample results (section A), give the proofs of our results (section B) and discuss extensions beyond functional time series (section C).

The programs that were used to analyse the data can be obtained from

https://rss.onlinelibrary.wiley.com/hub/journal/14679868/series-b-datasets.
2. Relevant hypotheses and self-normalization

Let $T$ be a compact set in $\mathbb{R}^d$ and let $L^2(T)$ denote the Hilbert space of square integrable functions on the set $T$ with the usual inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $\| \cdot \|$. 

2.1. One-sample problems

Let $\{X_n\}_{n \in \mathbb{Z}}$ denote a strictly stationary functional time series where the random variables $X_n$ are elements in $L^2(T)$ (with expectation $\mu := \mathbb{E}[X_1] \in L^2(T)$; see Section 2.1 in B"ucher et al. (2019) for a detailed discussion of expected values in Hilbert spaces). For simplicity we shall assume that $T = [0, 1]$, but all the methods that are proposed in this paper can be generalized to other subsets of $\mathbb{R}^d$. To avoid confusion between the interval $[0, 1]$ corresponding to $\lambda$, which defines the subsample $X_1, \ldots, X_{\lfloor n \lambda \rfloor}$ and the interval $T = [0, 1]$, we write $T$ for the interval $[0, 1]$ belonging to the argument $t$ of $X_n$. On the basis of a sample $X_1, \ldots, X_n$ we are interested in relevant hypotheses regarding the parameter $d = \int_T \mu^2(t) \, dt$, i.e.

$$H_0 : \int_T \mu^2(t) \, dt \leq \Delta \quad \text{versus} \quad H_1 : \int_T \mu^2(t) \, dt > \Delta. \quad (2.1)$$

Define the partial sums

$$S_n(t, \lambda) := \frac{1}{n} \sum_{j=1}^{\lfloor n \lambda \rfloor} X_j(t), \quad \lambda \in [0, 1]; \quad (2.2)$$

then, under suitable assumptions, the statistic $\int_T S_n^2(t, 1) \, dt$ is a consistent estimator of $\int_T \mu^2(t) \, dt$. Consequently a test for the hypotheses (2.1) is obtained by rejecting the null hypothesis of no relevant difference for large values of

$$\hat{\mathbb{V}}_n = \int_T S_n^2(t, 1) \, dt. \quad (2.3)$$

It will be shown in the proof of theorem 1 that under some technical assumptions the asymptotic distribution of an appropriately standardized version of $\hat{\mathbb{V}}_n$ takes the form

$$\sqrt{n} \left\{ \hat{\mathbb{V}}_n - \int_T \mu^2(t) \, dt \right\} \overset{D}{\rightarrow} \mathcal{N}(0, \tau^2)$$

with long-run variance

$$\tau^2 = 4 \int_T \int_T \mu(s) \mu(t) C(s, t) \, ds \, dt, \quad (2.4)$$

where

$$C(s, t) = \text{cov}\{X_0(s), X_0(t)\} + \sum_{l=1}^{\infty} \text{cov}\{X_0(s), X_l(t)\} + \sum_{l=1}^{\infty} \text{cov}\{X_0(s), X_{-l}(t)\} \quad (2.5)$$

is the long-run covariance operator of the process $\{X_n\}_{n \in \mathbb{Z}}$. Here we note that the above weak convergence is also true when $\mu \equiv 0$, in which case the limit is a degenerate normal distribution with a point mass at zero. Unfortunately, the long-run variance $\tau^2$ is difficult to estimate in practice. This motivates us to adopt a self-normalization approach which avoids direct estimation of $\tau^2$. To be more precise let $\nu$ denote a probability measure on the interval $(0, 1)$ and define

$$\hat{\mathbb{V}}_n := \left[ \int_0^1 \left\{ \int_T S_n^2(t, \lambda) \, dt - \lambda^2 \int_T S_n^2(t, 1) \, dt \right\}^2 \nu(\lambda) \right]^{1/2}. \quad (2.6)$$
Table 1. Simulated quantiles (based on 1000 replications) of the distribution of the statistic $W$ defined by expression (2.9), where $\nu$ is the discrete uniform distribution supported on the points 1, $\lambda_i = i/5$ ($i = 1, \ldots, 4$), 2, $\lambda_i = i/20$ ($i = 1, \ldots, 19$), and 3, $\lambda_i = i/100$ ($i = 1, \ldots, 99$)

| Point | 99% quantile | 95% quantile | 90% quantile |
|-------|--------------|--------------|--------------|
| 1     | 18.257       | 10.998       | 7.855        |
| 2     | 16.081       | 10.530       | 7.619        |
| 3     | 16.282       | 10.583       | 7.662        |

As we shall show later we have

$$(\sqrt{n}\bar{T}_n - d), \sqrt{n}\hat{\nu}_n \xrightarrow{D} (\tau \mathbb{B}(1), \tau \left[ \int_0^1 \lambda^2 \{ \mathbb{B}(\lambda) - \lambda \mathbb{B}(1) \}^2 \nu(d\lambda) \right]^{1/2}),$$

(2.7)

where $\mathbb{B}$ denotes standard Brownian motion on the interval $[0, 1]$. In particular, this implies that, in the case $\tau \neq 0$, the ratio $(\sqrt{n}\bar{T}_n - d)/\sqrt{n}\hat{\nu}_n$ converges to a pivotal distribution. This suggests that a test for hypotheses (2.1) can be constructed by rejecting the null hypothesis of no relevant difference in hypotheses (2.1), whenever

$$\bar{T}_n > \Delta + q_{1-\alpha}(\mathbb{W}) \hat{\nu}_n,$$

(2.8)

where $q_{1-\alpha}(\mathbb{W})$ denotes the $(1 - \alpha)$-quantile of the distribution of the pivotal random variable

$$\mathbb{W} := \mathbb{B}(1) \left[ \int_0^1 \lambda^2 \{ \mathbb{B}(\lambda) - \lambda \mathbb{B}(1) \}^2 \nu(d\lambda) \right]^{1/2}.$$

(2.9)

It is worthwhile to mention that the distribution of $\mathbb{W}$ is not the same as that in previous work on self-normalization (see for example Shao (2010) or Shao (2015)) and quantiles of this distribution need to be simulated first. In Table 1 we display quantiles of this distribution, where $\nu$ is the discrete uniform distribution supported on the points $\lambda_i = i/5$ ($i = 1, \ldots, 4$), on the points $\lambda_i = i/20$ ($i = 1, \ldots, 19$) and on the points $\lambda_i = i/100$ ($i = 1, \ldots, 99$).

Next we prove that the decision rule (2.8) indeed provides an asymptotic level $\alpha$ test. For this purpose we make the following assumptions (see also Berkes et al. (2013) and Horváth et al. (2014)).

**Assumption 1.** For all $j \in \mathbb{Z}$ we have $X_j = \mu + \eta_j$, where $(\eta_j)_{j \in \mathbb{Z}}$ is a centred error process which satisfies assumptions 2–4.

**Assumption 2.** $(\eta_j)_{j \in \mathbb{Z}}$ is a sequence of Bernoulli shifts, i.e. there is a measurable space, say $\mathcal{S}$, and a function $f : \mathcal{S}^\infty \rightarrow L^2([0, 1])$ such that

$$\eta_j = f(\varepsilon_j, \varepsilon_{j-1}, \ldots)$$

for all $j \in \mathbb{Z}$,

where $(\varepsilon_j)_{j \in \mathbb{Z}}$ is a sequence of independent and identically distributed (IID) $\mathcal{S}$-valued functions, such that $\varepsilon_j(t) = \varepsilon_j(t, \omega)$ is jointly measurable ($j \in \mathbb{Z}$).

**Assumption 3.** $\mathbb{E}\|\eta_j\|^{2+\psi} < \infty$ for some $\psi \in (0, 1)$. 
Assumption 4. The sequence \((\eta_j)_{j \in \mathbb{Z}}\) can be approximated by \(l\)-dependent sequences \((\eta_{j,l})_{j \in \mathbb{Z}}\) in the sense that for some \(\kappa > 2 + \psi\)
\[
\sum_{l=1}^{\infty} \mathbb{E} \|\eta_0 - \eta_{0,l}\|^{2+\psi} \frac{1}{\kappa} < \infty,
\]
where \(\eta_{j,l}\) is defined by
\[
\eta_{j,l} = \mathcal{f}(\varepsilon_{j}, \varepsilon_{j-1}, \ldots, \varepsilon_{j-l+1}, \varepsilon_{j,l}^*),
\]
and the random variables \(\varepsilon_{j,l,k}^*\) are IID copies of \(\varepsilon_0\), and independent of the sequence \((\varepsilon_j)_{j \in \mathbb{Z}}\).

Theorem 1. Assume that \(\Delta > 0\). Under assumptions 1–4 the test decision given in rule (2.8) satisfies
\[
\lim_{n \to \infty} \mathbb{P}\{\hat{T}_n > \Delta + q_{1-\alpha}(\mathbb{W})\hat{\nu}_n\} = \begin{cases} 0 & \text{if } \int_T \mu^2(t)dt < \Delta, \\ \alpha & \text{if } \int_T \mu^2(t)dt = \Delta \text{ and } \tau^2 > 0, \\ 1 & \text{if } \int_T \mu^2(t)dt > \Delta. \end{cases}
\]

A detailed proof of theorem 1 is given in the on-line appendix, section B.1.1. In what follows we provide an informal overview of the main steps in the proof. If \(\int \mu^2(t)dt \neq 0\) and assumptions 1–4 hold, it can be shown that
\[
\left\{ \sqrt{n} \left\{ \int_T S_n^2(t, \lambda)dt - \lambda^2 \int_T \mu^2(t)dt \right\} \right\}_{\lambda \in [0,1]} \to \mathcal{N}\{0, \Sigma(\mathbb{W})\},
\]
where the symbol ‘\(\to\)’ means weak convergence in \(\mathcal{F}(0,1)\) and \(\tau^2\) is defined in expression (2.4). Now an application of the continuous mapping theorem directly yields the joint weak convergence (2.7). This implies the statement of theorem 1 when \(\int \mu^2(t)dt > 0\) after some simple computations. If \(\int \mu^2(t)dt = 0\) it is possible to prove that \(\hat{T}_n = o_P(1)\) and \(\hat{\nu}_n = o_P(1)\). This implies that
\[
\lim_{n \to \infty} \mathbb{P}\{\hat{T}_n > \Delta + q_{1-\alpha}(\mathbb{W})\hat{\nu}_n\} = \lim_{n \to \infty} \mathbb{P}\{o_P(1) > \Delta\} = 0,
\]
where we used that \(\Delta > 0\) is fixed.

Remark 1. In general, the rejection rule (2.8) does not lead to an asymptotic level \(\alpha\) test when \(\Delta = 0\). To see this note that for \(\Delta = 0\) the null hypothesis contains only one point \(\mu \equiv 0\) and we reject the null when \(\hat{T}_n / \hat{\nu}_n > q_{1-\alpha}(\mathbb{W})\). However, a slight extension of the arguments given in the proof of result (2.11) shows that for \(\mu \equiv 0\) we have
\[
\frac{\hat{T}_n}{\hat{\nu}_n} \overset{D}{\to} \mathcal{W} := \int_T \left[ \int_0^1 \left\{ \int_T \Gamma^2(t, 1)dt - \lambda^2 \int_T \Gamma^2(t, 1)dt \right\} \nu(d\lambda) \right]^{1/2},
\]
where \(\Gamma(t, \lambda)\) is a centred Gaussian process with covariance function
\[
\text{cov}\{\Gamma(t, \lambda), \Gamma(s, \lambda')\} = (\lambda \wedge \lambda') C(s, t),
\]
where $C$ is the long-run covariance operator defined in expression (2.5). The distribution of $\tilde{W}$ does not match that of $W$ and is not pivotal. Hence a test based on rejecting $H_0: \mu = 0$ by using decision rule (2.8) will not have asymptotic level $\alpha$.

Remark 2. A test for the hypotheses of equivalence

$$H_0 : \int_T \mu^2(t) dt > \Delta \quad \text{versus} \quad H_1 : \int_T \mu^2(t) dt \leq \Delta$$

(2.12)

can be obtained similarly. The null hypothesis of a relevant difference in test (2.12) is rejected, when

$$\hat{T}_n \leq \Delta + q_{\alpha}(W) \hat{\nu}_n,$$

where $\hat{T}_n$ and $\hat{\nu}_n$ are defined in expressions (2.3) and (2.6) respectively and $q_{\alpha}(W)$ is the $\alpha$-quantile of the distribution of $W$ defined in expression (2.9). Similar arguments to those given in the proof of theorem 1 show that this test is an asymptotic level $\alpha$ and consistent test for hypotheses (2.12), i.e.

$$\lim_{n \to \infty} \mathbb{P}\{ \hat{T}_n \leq \Delta + q_{\alpha}(W) \hat{\nu}_n \} = \begin{cases} 1 & \text{if } \int_T \mu^2(t) dt < \Delta, \\ \alpha & \text{if } \int_T \mu^2(t) dt = \Delta \text{ and } \tau^2 > 0, \\ 0 & \text{if } \int_T \mu^2(t) dt > \Delta. \end{cases}$$

The details have been omitted for brevity.

Remark 3. As pointed out by a referee it is of interest to compare the test (2.8) based on self-normalization with a corresponding test using an estimate of the long-run variance. For this note that such a test rejects the null hypothesis of no relevant difference (2.1), whenever

$$\hat{T}_n > \Delta + u_{1-\alpha} \tau_n / \sqrt{n},$$

(2.13)

where $u_{1-\alpha}$ is the $(1 - \alpha)$-quantile of the standard normal distribution and $\tau_n^2$ is an appropriate estimator of the long-run variance (2.4). In the case of one sample this is still relatively easy. For example, one could use

$$\tau_n^2 = 4 \int_0^1 \int_0^1 S_n(s, 1) S_n(t, 1) \hat{C}_n(s, t) ds dt,$$

(2.14)

where $S_n(t, 1)$ is defined in expression (2.2) and $\hat{C}_n$ is an appropriate estimator of the long-run covariance operator. A numerical illustration of this approach in comparison with self-normalization can be found in Section 5.1.1.

2.2. Two-sample problems

Throughout this section let $\{X_n\}_{n \in \mathbb{Z}}$ and $\{Y_n\}_{n \in \mathbb{Z}}$ denote two strictly stationary functional time series with values in $L^2(T)$. Assume that we observe finite stretches, say $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ from $\{X_n\}_{n \in \mathbb{Z}}$ and $\{Y_n\}_{n \in \mathbb{Z}}$. Denote by $\mu_1 = \mathbb{E}[X_1]$ and $\mu_2 = \mathbb{E}[Y_1]$ the corresponding mean functions and by $D(t) = \mu_1(t) - \mu_2(t)$ their difference and define the partial sum

$$D_{m,n}(t, \lambda) := \frac{1}{m} \sum_{j=1}^{[m\lambda]} X_j(t) - \frac{1}{n} \sum_{j=1}^{[n\lambda]} Y_j(t).$$
From this definition we see that
\[ E[D_{m,n}(t, \lambda)] = \lambda D(t) + O((m \wedge n)^{-1}). \] (2.15)

For brevity we restrict ourselves to the problem of testing the relevant hypotheses
\[ H_0 : \int_T D^2(t)dt \leq \Delta \quad \text{versus} \quad H_1 : \int_T D^2(t)dt > \Delta, \] (2.16)
where \( \Delta \) is a prespecified threshold. A corresponding test for the hypotheses of equivalence can be derived along the lines given in remark 2. Following the discussion in Section 2.1 we propose to reject the null hypothesis of no relevant difference in test (2.16), whenever
\[ \hat{D}_{m,n} > \Delta + q_{1-\alpha}(\mathbb{W}) \hat{V}_{m,n}, \] (2.17)
where \( q_{1-\alpha}(\mathbb{W}) \) is the \((1 - \alpha)\)-quantile of the distribution of the random variable \( \mathbb{W} \) in expression (2.9). The statistics \( \hat{D}_{m,n} \) and \( \hat{V}_{m,n} \) are defined by
\[ \hat{D}_{m,n} = \int_T D^2_{m,n}(t,1)dt, \] (2.18)
\[ \hat{V}_{m,n} = \left[ \int_0^1 \left\{ \int_T D^2_{m,n}(t,\lambda)dt - \lambda^2 \int_T D^2_{m,n}(t,1)dt \right\}^2 \nu(d\lambda) \right]^{1/2} \] (2.19)
respectively, where \( \nu \) is a probability measure on the interval \((0, 1)\). The asymptotic properties of this test procedure will be established under the following assumptions.

**Assumption 5.** The sample sizes satisfy: \( m \to \infty \) and \( n \to \infty \) and \( m/(m+n) \to \rho \in (0, 1) \).

**Assumption 6.** The processes \( \{X_n\}_{n \in \mathbb{Z}} \) and \( \{Y_n\}_{n \in \mathbb{Z}} \) are independent and satisfy assumptions 1–4 stated in Section 2.1 with \( \mathbb{E}[X_1] = \mu_1 \) and \( \mathbb{E}[Y_1] = \mu_2 \).

We also define the quantity
\[ \tau^2_D = 4 \int_T \int_T D(s)D(t) \left\{ \frac{1}{\rho}C_X(s,t) + \frac{1}{1-\rho}C_Y(s,t) \right\} dsdt, \] where \( C_X \) and \( C_Y \) are the long-run covariance operators corresponding to the processes \( \{X_n\}_{n \in \mathbb{Z}} \) and \( \{Y_n\}_{n \in \mathbb{Z}} \) respectively.

**Theorem 2.** Assume that \( \Delta > 0 \). Under assumptions 5 and 6 the test decision given in inequality (2.17) satisfies
\[ \lim_{n \to \infty} \mathbb{P}\{ \hat{D}_{m,n} > \Delta + q_{1-\alpha}(\mathbb{W}) \hat{V}_{m,n} \} = \begin{cases} 0 & \text{if } \int_T D^2(t)dt < \Delta, \\ \alpha & \text{if } \int_T D^2(t)dt = \Delta \text{ and } \tau^2_D > 0, \\ 1 & \text{if } \int_T D^2(t)dt > \Delta. \end{cases} \]

We note that similarly to the one-sample case it can be shown that the rejection rule (2.17) does not lead to an asymptotic level \( \alpha \) test when \( \Delta = 0 \).

**Remark 4.**
(a) The statement in theorem 2 continues to hold if the observations \( X_i \) and \( Y_i \) are generated
In this section we consider data that are generated from the following (triangular array) model:

\[ X_i = \mu_1 + f_1(\varepsilon_i, \varepsilon_{i-1}, \ldots), \quad i = 1, \ldots, n, \quad \text{and} \quad Y_i = \mu_2 + f_2(\varepsilon_{n+i}, \varepsilon_{n+i-1}, \ldots), \quad i = 1, \ldots, m \]

where \((\varepsilon_j)_{j \in \mathbb{Z}}\) denotes an IID sequence of \(S\)-valued functions with the property that \(\varepsilon_j(t, \omega)\) is jointly measurable as in assumption (2) and \(f_1, f_2: S^\infty \to L^2([0, 1])\) are functions such that the processes \((f_1(\varepsilon_i, \varepsilon_{i-1}, \ldots))_{i \in \mathbb{Z}}\) and \((f_2(\varepsilon_i, \varepsilon_{i-1}, \ldots))_{i \in \mathbb{Z}}\) satisfy conditions 3 and 4. This essentially corresponds to the setting that is discussed in Section 3 when the change point location is known.

(b) A test based on estimation of the long-run variance of the statistic \(\hat{\Delta}_{m,n}\) can be constructed along the lines given in remark 3. The details have been omitted for brevity.

Remark 5. As pointed out by the Associate Editor the proposed way of self-normalization is not unique and one could also think about alternative constructions. For instance, one could also use the statistics

\[
\hat{V}_{m,n}^* = \nu^{-\text{ess sup}}_{\lambda \in [0, 1]} \left| \int_T D_{m,n}^2(t, \lambda) dt - \lambda^2 \int_T D_{m,n}^2(t, 1) dt \right|, \\
\hat{V}_{m,n}^{**} = \left. \int_0^1 \left| \int_T D_{m,n}^2(t, \lambda) dt - \lambda^2 \int_T D_{m,n}^2(t, 1) dt \right| \nu(d\lambda) \right| (2.20)
\]

in decision rule (2.8) if the quantile \(q_{1-\alpha}(\mathbb{W})\) is replaced by the \((1 - \alpha)\)-quantile of the random variables

\[
\mathbb{W}^* := \frac{\mathbb{B}(1)}{\nu^{-\text{ess sup}}_{\lambda \in [0, 1]} |\lambda \{\mathbb{B}(\lambda) - \lambda \mathbb{B}^2(1)\}|}, \\
\mathbb{W}^{**} := \frac{\mathbb{B}(1)}{\int_0^1 |\lambda \{\mathbb{B}(\lambda) - \lambda \mathbb{B}^2(1)\}| \nu(d\lambda)} (2.22)
\]

respectively. The self-normalizing factors (2.20) and (2.21) might have some advantages for heavy-tailed data. However, it will be demonstrated in Section 5.2 that the finite sample properties of these two alternative tests are very similar to those of test (2.17).

3. Relevant change points in the mean function

In this section we consider data that are generated from the following (triangular array) model:

\[
X_i = \begin{cases} 
\mu + f_1(\varepsilon_i, \varepsilon_{i-1}, \ldots) & \text{if } i \leq N\theta_0, \\
\mu + \delta + f_2(\varepsilon_i, \varepsilon_{i-1}, \ldots) & \text{if } i > N\theta_0.
\end{cases} 
\]

(3.1)

Here \(\mu\) and \(\delta\) denote deterministic but unknown elements in \(L^2(T)\) and \(\theta_0 \in (0, 1)\) is fixed but unknown. Moreover, \((\varepsilon_j)_{j \in \mathbb{Z}}\) denotes an IID sequence of \(S\)-valued functions with the property that \(\varepsilon_j(t, \omega)\) is jointly measurable as in assumption 2 and \(f_1, f_2: S^\infty \to L^2(T)\) are functions such that the processes \((f_1(\varepsilon_i, \varepsilon_{i-1}, \ldots))_{i \in \mathbb{Z}}\) and \((f_2(\varepsilon_i, \varepsilon_{i-1}, \ldots))_{i \in \mathbb{Z}}\) satisfy conditions 3 and 4. This setting is sufficiently general to allow the whole distribution of the observed functional data to change together with their mean.

We aim to construct a test for the relevant hypothesis

\[
H_0 : \int_T \delta^2(t) dt \leq \Delta \quad \text{versus} \quad H_1 : \int_T \delta^2(t) dt > \Delta
\]

(3.2)

where \(\Delta\) is a prespecified threshold. For known \(\theta_0\) a test for \(H_0\) can be constructed in a similar
fashion to that in Section 2.2. In this section, we shall prove that replacing the known change point by an estimator also leads to an asymptotic level \( \alpha \) test for hypotheses (3.2). For this we fix a trimming parameter \( \varepsilon \in [0, \frac{1}{2}] \) and define the estimator of the unknown change point \( \theta_0 \) as

\[
\hat{\theta} := \frac{1}{N} \arg \max_{[N\varepsilon] + 1 \leq k \leq N - [N\varepsilon]} \hat{f}(k),
\]

where \( \hat{f}(0) = \hat{f}(N) = 0 \) and for \( k = 1, \ldots, N - 1 \)

\[
\hat{f}(k) := \frac{k}{N} \left( 1 - \frac{k}{N} \right) \int_T \left\{ \frac{1}{k} \sum_{j=1}^k X_j(t) - \frac{1}{N-k} \sum_{j=k+1}^N X_j(t) \right\}^2 \, dt.
\]

Our first result shows that the estimator \( \hat{\theta} \) is consistent.

**Proposition 1.** If the data are generated according to model (3.1), \( \int \delta^2(t) \, dt > 0, \theta_0 \in (\varepsilon, 1 - \varepsilon) \), and the assumptions described immediately below model (3.1) are satisfied, then

\[
\hat{\theta} = \theta_0 + o_p(N^{-1/2}).
\]

Next we introduce the test statistic. For arbitrary \( \theta \in [1/N, 1] \) define

\[
D_N^{CP}(t, \lambda, \theta) := \frac{1}{[N\theta]} \sum_{j=1}^{[\lambda(N\theta)]} X_j(t) - \frac{1}{N - [N\theta]} \sum_{j=[\theta N] + 1}^{[\lambda(N-N\theta)]} X_j(t).
\]

Following the developments in Section 2.2 let

\[
\hat{D}_N^{CP} = \int_T D_N^{CP}(t, 1, \hat{\theta})^2 \, dt,
\]

\[
\hat{\nu}_N^{CP} = \left[ \int_0^1 \left\{ \int_T D_N^{CP}(t, \lambda, \hat{\theta})^2 \, dt - \lambda^2 \int_T D_N^{CP}(t, 1, \hat{\theta})^2 \, dt \right\}^2 \, \nu(d\lambda) \right]^{1/2}
\]

respectively, where \( \nu \) is a probability measure on the interval \((0, 1)\). The test for \( H_0 \) takes the form

\[
\hat{D}_N^{CP} > \Delta + q_{1-\alpha}(W) \hat{\nu}_N^{CP},
\]

where \( q_{1-\alpha}(W) \) is the \((1 - \alpha)\)-quantile of the distribution of the random variable \( W \) in expression (2.9). This test decision is justified in the following theorem. For its precise statement we define the quantity

\[
\tau_{\delta, \theta_0}^2 = 4 \int_T \int_T \delta(s) \delta(t) \left\{ \frac{1}{\theta_0} K_1(s, t) + \frac{1}{1 - \theta_0} K_2(s, t) \right\} ds \, dt,
\]

where for \( j = 1, 2 \)

\[
K_j(s, t) = \sum_{h \in \mathbb{Z}} \text{cov}\{\eta_{h}^{(j)}(s), \eta_{h}^{(j)}(t)\}, \quad j = 1, 2,
\]

is the long-run covariance kernel of \( \{\eta_i^{(j)}\}_{i \in \mathbb{Z}} = \{f_j(\varepsilon_i, \varepsilon_{i-1}, \ldots)\}_{i \in \mathbb{Z}} \).

**Theorem 3.** Assume that \( \Delta > 0 \). If the data are generated according to model (3.1), \( \theta_0 \in (\varepsilon, 1 - \varepsilon) \), and the assumptions described immediately below model (3.1) hold, then the test decision in expression (3.8) satisfies
Testing Relevant Hypotheses in Functional Time Series

\[
\lim_{n \to \infty} P\{ \hat{D}_{n}^{\text{CP}} > \Delta + q_{1-\alpha}(\mathbb{W}) \hat{V}_{n}^{\text{CP}} \} = \begin{cases} 
0 & \text{if } \int_{T}^{\delta^{2}(t)dt < \Delta}, \\
\alpha & \text{if } \int_{T}^{\delta^{2}(t)dt = \Delta \text{ and } \tau_{\delta,\theta_{0}}^{2} > 0}, \\
1 & \text{if } \int_{T}^{\delta^{2}(t)dt > \Delta}.
\end{cases}
\]

Similarly to the one-sample and two-sample case, rejection rule (3.8) does not lead to an asymptotic level \(\alpha\) test when \(\Delta = 0\).

The proof of proposition 1 and theorem 3 is technically difficult and has been deferred to on-line appendix section B.2, but the main idea is as follows. A straightforward calculation shows that the processes \(\hat{D}_{n}^{\text{CP}}\) and \(\hat{V}_{n}^{\text{CP}}\) in expressions (3.6) and (3.7) are continuous functionals of the process

\[
Z_{n}(\lambda, \hat{\theta}) = \sqrt{N} \int_{T}^{\{ D_{n}^{\text{CP}}(t, \lambda, \hat{\theta})^{2} - \lambda^{2} \delta(t)^{2} \} dt.
\]

Using proposition 1 it can be shown that

\[
\sup_{\lambda \in [0,1]} |Z_{n}(\lambda, \theta_{0}) - Z_{n}(\lambda, \hat{\theta})| = o_{P}(1),
\]

where \(\theta_{0}\) is the true change point. We can then establish the weak convergence

\[
\left\{ Z_{n}(\lambda, \theta_{0}) \right\}_{\lambda \in [0,1]} \rightsquigarrow \left\{ \lambda \tau_{\delta,\theta_{0}}^{2} \mathcal{B}(\lambda) \right\}_{\lambda \in [0,1]},
\]

where \(\tau_{\delta,\theta}^{2}\) is defined in expression (3.9). Using the continuous mapping theorem we then find that

\[
\frac{\hat{D}_{n}^{\text{CP}}}{\hat{V}_{n}^{\text{CP}}} \overset{D}{\to} \mathbb{W},
\]

where the random variable \(\mathbb{W}\) is defined in expression (2.9). When \(\int_{T}^{\delta^{2}(t)dt > 0}\), the assertion of theorem 3 now follows directly. In the remaining case \(\int_{T}^{\delta^{2}(t)dt = 0}\) one can show that \(\hat{D}_{n}^{\text{CP}} = o_{P}(1)\) and \(\hat{V}_{n}^{\text{CP}} = o_{P}(1)\) and the assertion follows with the same arguments as given in Section 2.1.

Remark 6. In this remark we briefly explain how—in principle—a test can be constructed by using an appropriate estimate of the long-run variance of the statistic \(\hat{D}_{n}^{\text{CP}}\) in expression (3.6). A careful inspection of the proof of theorem 3 (see in particular the discussion at the end of the proof of expression (B.31) in the on-line supplement) shows that

\[
\sqrt{N} \left\{ \hat{D}_{n}^{\text{CP}} - \int_{T}^{\delta^{2}(t)dt} \right\} \overset{D}{\to} \mathcal{N}(0, \tau_{\delta,\theta_{0}}^{2}),
\]

where the asymptotic variance is defined in expression (3.9). If \(\hat{\tau}_{n}^{2}\) denotes an estimator of \(\tau_{\delta,\theta_{0}}^{2}\), an asymptotic level \(\alpha\) test is obtained by rejecting the null hypothesis in test (3.2), whenever

\[
\hat{D}_{n}^{\text{CP}} > \Delta + \frac{\hat{\tau}_{n}^{2}}{\sqrt{N}} u_{1-\alpha}.
\]

One possibility to estimate \(\tau_{\delta,\theta_{0}}^{2}\) is to replace \(\theta_{0}\) by the estimator \(\hat{\theta}\) that is defined in expression (3.3) and to use appropriate plug-in estimators for the covariance kernels \(K_{j}\) and unknown
difference $\delta$ based on the subsamples $\{X_i : i \leq \lfloor N\hat{\theta}\rfloor\}$ and $\{X_i : i > \lfloor N\hat{\theta}\rfloor\}$. Details have been omitted for brevity.

**Remark 7.** The extension of the methodology to the analysis of multiple-change-point problems is of great practical interest and is briefly indicated here. Let $0 < \theta_1 < \theta_2 < \ldots < \theta_K < 1$ denote the unknown change points and assume that

$$X_i = \mu + \delta_{j-1} + f(\epsilon_i, \epsilon_{i-1}, \ldots)$$

if $\lfloor N\theta_{j-1}\rfloor + 1 \leq i \leq \lfloor N\theta_j\rfloor$, $1 \leq j \leq K + 1$, (3.11)

where $\delta_0 = 0; \theta_0 = 0$ and $\theta_{K+1} = 1$. For simplicity we consider the same function $f$ as filter for the error process on the $j$th segment (in contrast with model (3.1)) and further consider $K$ as known. Defining the vector of integrated squared differences

$$\Psi := \left(\int T \delta_1^2(t) dt, \int T \{\delta_2(t) - \delta_1(t)\}^2 dt, \ldots, \int T \{\delta_K(t) - \delta_{K-1}(t)\}^2 dt\right)^T,$$

there are several possibilities to formulate relevant hypotheses in this setting. Here we shall focus on

$$H_0^{L^2} : \sum_{j=1}^K \Psi_j \leq \Delta,$$

which corresponds to the null that the sum of all integrated squared changes does not exceed a threshold $\Delta$ and

$$H_0^{\infty} : \max_{j=1}^K \Psi_j \leq \Delta,$$

meaning that no single integrated squared change exceeds $\Delta$. Note that $H_0^{\infty}$ can be equivalently formulated as the intersection of the following hypotheses:

$$H_0^{(j)} : \Psi_j \leq \Delta \quad j = 1, \ldots, K.$$

(3.14)

For testing either of the hypotheses, we propose first to estimate the multiple change points by adapting one of the commonly used methods such as binary segmentation or wild binary segmentation (see for example Vostrikova (1981), Harchaoui and Lévy-Leduc (2010), Fryzlewicz (2014) and Zhang and Lavitas (2018) among many others) to dependent functional data. The resulting estimator is denoted by $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_K)$ and we put $\hat{\theta}_{K+1} = 1$ and $\hat{\theta}_0 = 0$. Next we consider generalizations of the statistics $\hat{D}_N^{cp}$ and $\hat{\nu}_N^{cp}$ in expressions (3.6) and (3.7) on each segment:

$$S_j := \{X_i \lfloor N\hat{\theta}_{j-1}\rfloor + 1 \leq i \leq \lfloor N\hat{\theta}_{j+1}\rfloor\}, \quad j = 1, \ldots, K.$$

More precisely, define $\hat{N}_j := \lfloor N\hat{\theta}_j\rfloor - \lfloor N\hat{\theta}_{j-1}\rfloor$, $j = 1, \ldots, K + 1$, as the sample size between the $(j - 1)$st and $j$th estimated change point and let

$$\hat{D}_{N,j}^{cp}(t, \lambda, \hat{\theta}) = \frac{1}{\hat{N}_j} \sum_{i=1}^{\lfloor N\hat{\theta}_j\rfloor - 1} X_{\lfloor N\hat{\theta}_j\rfloor - 1 + i}(t) - \frac{1}{\hat{N}_{j+1}} \sum_{i=1}^{\lfloor N\hat{\theta}_{j+1}\rfloor} X_{\lfloor N\hat{\theta}_{j+1}\rfloor + i}(t), \quad j = 1, \ldots, K.$$

Further, define

$$\hat{\nu}_N^{cp,j}(\lambda, \hat{\theta}) := \int_{T} \hat{D}_{N,j}^{cp}(t, \lambda, \hat{\theta})^2 dt, \quad j = 1, \ldots, K.$$

(3.15)

With this preparation, we first discuss a test for the hypothesis $H_0^{L^2}$ in expression (3.12). Define
\[
\hat{D}_N^{L^2}(\lambda, \hat{\theta}) := \sum_{j=1}^{K} \hat{D}_N^{cp}(\lambda, \hat{\theta}),
\]

\[
\hat{V}_N^{L^2}(\hat{\theta}) := \left[ \int_0^1 \left\{ \hat{D}_N^{L^2}(\lambda, \hat{\theta}) - \lambda^2 \hat{D}_N^{L^2}(1, \hat{\theta}) \right\}^2 \nu(d\lambda) \right]^{1/2}.
\]

We propose to reject \( H_0^{L^2} \) whenever

\[
\hat{D}_N^{L^2}(1, \hat{\theta}) > \Delta + q_{1-\alpha}(W) \hat{V}_N^{L^2}(\hat{\theta}),
\]

where \( q_{1-\alpha}(W) \) denotes the \((1 - \alpha)\)-quantile of the random variable \( W \) defined in expression (2.9). In section B.3 of the on-line supplement, we show that this provides a consistent and asymptotic level \( \alpha \) test for \( H_0^{L^2} \) under the following conditions.

**Condition 1.** The data are generated from model (3.11) with \( f, \{\varepsilon_i\}_{i \in \mathbb{Z}} \) satisfying assumptions 2–4.

**Condition 2.** The true number of change points is \( K \) (i.e. all entries of \( \Psi \) are non-zero and there are no other change points) and \( \hat{\theta}_j = \theta_j + o_P(N^{-1/2}) \), \( j = 1, \ldots, K \).

These conditions are made for a simpler presentation. It is possible to generalize condition 1 to the case where the filter \( f \) changes in each segment as in model (3.1). Similarly, condition 2 can be weakened to for example \( \Psi_j = 0 \) for some values \( j \). In this case the requirement \( \hat{\theta}_j = \theta_j + o_P(N^{-1/2}) \) for \( j \) with \( \Phi_j = 0 \) is not realistic and must be replaced by a different condition. Details have been omitted for brevity.

Finally, we briefly comment on testing hypothesis (3.13) by using self-normalization, which is a substantially more challenging problem. Although it is possible to construct self-normalized test statistics for each of the hypotheses \( H_0^{(j)} \) in expression (3.14) separately, note that neighbouring segments \( S_j \) and \( S_{j+1} \) overlap so in the limit those self-normalized statistics become dependent. It can further be shown that, although the marginal distributions are pivotal, the resulting joint distribution is not pivotal any more. Therefore one option to construct a test for hypothesis (3.13) is to apply a multiple-testing correction after testing each \( H_0^{(j)} \) separately on the basis of self-normalization. Problems of this type have been discussed more intensively in the context of testing simultaneously several hypotheses of equivalence of the form (1.3); see Munk and Pflüger (1999) and Wang et al. (1999). These references indicate the general difficulty in constructing tests for multiple precise hypotheses by using the joint distribution of a vector of test statistics, and we leave a detailed investigation of this problem for future research.

### 4. Inference for covariance operators

In this section we briefly discuss extensions of the methodology in Sections 2 and 3 to test similar hypotheses regarding the covariance operators of functional time series. For brevity we omit the one-sample case and focus on the two-sample case and the change point setting. The problem of testing the classical hypothesis has found considerable attention in the literature and we refer to Panaretos et al. (2010), Fremdt et al. (2013), Guo et al. (2016) and Paparoditis and Sapatinas (2016), who developed methodology for comparing covariance operators for independent functional data. Pilavakis et al. (2019) proposed a test for the equality of the lag 0 auto-covariance operators of \( K \) functional time series, and Sharipov and Wendler (2019) considered bootstrap-based statistical inference for covariance operators of functional time series. Change
point analysis for covariance operators was developed by Aston and Kirch (2012) and Stoehr et al. (2019), among others, whereas Zhang and Shao (2015) and Aue et al. (2020) discussed statistical inference tools for the eigensystem of covariance operators.

We shall repeatedly make use of the following strengthening of assumptions 3 and 4 on a sequence of $L^2(T)$-valued random elements $\{\eta_j\}_{j \in \mathbb{Z}}$ that satisfy assumption 2.

**Assumption 3’.** $\mathbb{E}\|\eta_0\|^{4+\psi} < \infty$ for some $\psi \in (0, 1)$.

**Assumption 4’.** The sequence $(\eta_j)_{j \in \mathbb{Z}}$ can be approximated by $l$-dependent sequences $(\eta_{j,l})_{j \in \mathbb{Z}}$ in the sense that for some $\kappa > 4 + \psi$

$$\sum_{l=1}^{\infty} (\mathbb{E}\|\eta_0 - \eta_{0,l}\|^{4+\psi})^{1/\kappa} < \infty,$$

where $\eta_{j,l}$ is defined by

$$\eta_{j,l} = f(\varepsilon_j, \varepsilon_{j-1}, \ldots, \varepsilon_{j-l+1}, \varepsilon_{j,l}),$$

and $\varepsilon_{j,l} = (\varepsilon_{j,l,j-l,1}, \varepsilon_{j,j-1,l,1}, \ldots),$ and the random variables $\varepsilon_{j,l,k}$ are IID copies of $\varepsilon_0$, and independent of the sequence $(\varepsilon_j)_{j \in \mathbb{Z}}$.

### 4.1. Two-sample problem

Given two samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ from independent, strictly stationary functional time series $\{X_t\}_{t \in \mathbb{Z}}$ and $\{Y_t\}_{t \in \mathbb{Z}}$, we are interested in testing the null hypothesis of no relevant difference in the covariance operators $C^X$ and $C^Y$ where

$$C^X(s, t) := \mathbb{E}\{X_0(s) - \mathbb{E}[X_0(s)]\} \{X_0(t) - \mathbb{E}[X_0(t)]\}$$

and $C^Y$ is defined similarly. Thus we investigate the hypotheses

$$H_0^C : d_C = \int_T \int_T D_C^2(s, t)dsdt \leq \Delta \quad \text{versus} \quad H_1^C : d_C > \Delta,$$  

(4.1)

where $D_C(s, t) = C^X(s, t) - C^Y(s, t)$ denotes the difference of the covariance operators at the points $s, t \in T$. We reject the null hypothesis in test (4.1), whenever

$$\hat{D}^C_{m,n} > \Delta + q_{1-\alpha}(\mathbb{W})\hat{\mathbb{V}}_{m,n},$$

(4.2)

where $q_{1-\alpha}(\mathbb{W})$ denotes the $(1 - \alpha)$-quantile of the random variable $\mathbb{W}$ defined in expression (2.9),

$$\hat{D}^C_{m,n} = \int_T \int_T D_{m,n}^2(s, t, 1)dsdt,$$

$$\hat{\mathbb{V}}^C_{m,n} = \left[\int_0^1 \left\{ \int_T \int_T D_{m,n}^2(s, t, \lambda)dsdt - \lambda^2 \int_T \int_T D_{m,n}^2(s, t, 1)dsdt \right\}^{1/2} \nu(d\lambda) \right]^{1/2},$$

and $D_{m,n}$ is the partial sum defined by

$$D_{m,n}(s, t, \lambda) = \frac{1}{m-1} \sum_{j=1}^{[m\lambda]} \left\{ X_j(s) - \frac{1}{\lfloor m\lambda \rfloor} \sum_{i=1}^{[m\lambda]} X_i(s) \right\} \left\{ X_j(t) - \frac{1}{\lfloor m\lambda \rfloor} \sum_{i=1}^{[m\lambda]} X_i(t) \right\}$$

$$- \frac{1}{n-1} \sum_{j=1}^{[n\lambda]} \left\{ Y_j(s) - \frac{1}{\lfloor n\lambda \rfloor} \sum_{i=1}^{[n\lambda]} Y_i(s) \right\} \left\{ Y_j(t) - \frac{1}{\lfloor n\lambda \rfloor} \sum_{i=1}^{[n\lambda]} Y_i(t) \right\}.$$
The following result shows that decision rule (4.2) defines a consistent asymptotic level \( \alpha \) test for the hypotheses \( H^C_0 \) versus \( H^C_1 \) provided that \( \Delta > 0 \). For a precise statement, define

\[
\tau^2_{DC} = 4 \int_T \int_T \int_T \int_T D_C(s, t) D_C(s', t') \left[ \frac{1}{\rho} C_X \{(s, t), (s', t')\} + \frac{1}{1 - \rho} C_Y \{(s, t), (s', t')\} \right] ds'ds'dt'
\]

where \( \rho \) comes from assumption 5 and \( C_X \) and \( C_Y \) denote the long-run covariance kernels corresponding to the time series \( \{X_t \otimes X_{t'}\}_{t \in \mathbb{Z}} \) and \( \{Y_t \otimes Y_{t'}\}_{t \in \mathbb{Z}} \) respectively.

**Theorem 4.** Assume that \( \Delta > 0 \). Let assumption 5 from Section 2.2 hold and assume that the functional time series \( \{X_t\}_{t \in \mathbb{Z}} \) and \( \{Y_t\}_{t \in \mathbb{Z}} \) satisfy assumption 1 with means \( \mu_X \) and \( \mu_Y \) and errors \( \eta^X_i \) and \( \eta^Y_i \), and assumption 2, and that \( \eta^X_i \) and \( \eta^Y_i \) satisfy assumptions 3 and 3'.

Finally, assume that \( \nu \) puts no mass in a neighbourhood of zero. Then

\[
\lim_{n \to \infty} \mathbb{P}\{\hat{\Delta}^C_{m,n} > \Delta + q_{1-\alpha}(\mathbb{W}) \cap \nu^C_{n,m}\} = \begin{cases} 0 & \text{if } d_C < \Delta, \\ \alpha & \text{if } d_C = \Delta \text{ and } \tau^2_{DC} > 0, \\ 1 & \text{if } d_C > \Delta. \end{cases}
\]

Similarly to the setting of testing for a relevant difference in the means it can be shown that rejection rule (4.2) does not lead to an asymptotic level \( \alpha \) test when \( \Delta = 0 \).

### 4.2. Change point problem

For an extension of the methodology to testing for relevant changes in the covariance structure of a time series, we shall assume that data are generated from the model

\[
X_i = \begin{cases} \mu + f_1(e_i, e_{i-1}, \ldots) & \text{if } i \leq N\theta_0, \\ \mu + f_2(e_i, e_{i-1}, \ldots) & \text{if } i > N\theta_0. \end{cases}
\]

(4.3)

Define \( \eta^{(k)}_i := f_k(e_i, e_{i-1}, \ldots), k = 1, 2, \) and denote by \( C_1 \) and \( C_2 \) the covariance operators of the process \( X_i \) before and after the structural break. We are now interested in testing the hypotheses

\[
H^C_0 : d^C_{CP} = \int_T \int_T \{C_1(s, t) - C_2(s, t)\}^2 ds dt \leq \Delta \quad \text{versus} \quad H^C_1 : d^C_{CP} > \Delta.
\]

(4.4)

Similarly to Section 3 we first construct an estimator for the unknown change point location \( \theta_0 \). For this we define

\[
\bar{X}_{l,k}(t) := \frac{1}{k-l+1} \sum_{i=l}^{k} X_i(t)
\]

and consider the covariance estimators

\[
\hat{C}_{1:k}(s, t) := \frac{1}{k-1} \sum_{j=1}^{k} \{X_j(s) - \bar{X}_{1:k}(s)\} \{X_j(t) - \bar{X}_{1:k}(t)\},
\]

\[
\hat{C}_{k+1:N}(s, t) := \frac{1}{N-k-1} \sum_{j=k+1}^{N} \{X_j(s) - \bar{X}_{k+1:N}(s)\} \{X_j(t) - \bar{X}_{k+1:N}(t)\}
\]

for \( k = 2, \ldots, N - 2 \). Next fix a trimming parameter \( \varepsilon \in [0, \frac{1}{2}] \) and define the estimator

\[
\hat{\theta}^{cov} := \frac{1}{N} \arg \max_{|N_k|+1 \leq k \leq N-|N_k|} \hat{f}^{cov}(k),
\]

where \( \hat{f}^{cov}(0) = \hat{f}^{cov}(N) = 0 \) and for \( k = 1, \ldots, N - 1 \)
Following the approach in Section 3, for arbitrary $\theta \in [2/N, 1 - 1/N]$ define
\[
D_{N}^{\text{cp,cov}}(s, t, \lambda, \theta) = \frac{1}{[N\theta] - 1} \sum_{j=1}^{[[N\theta]\lambda]} \{ X_j(s) - \bar{X}_{1:|[N\theta]\lambda]}(s) \{ X_j(t) - \bar{X}_{1:|[N\theta]\lambda]}(t) \} - \frac{1}{N - [N\theta] - 1} \sum_{j=[N\theta] + 1}^{[[N\theta]\lambda]} \{ X_j(s) - \bar{X}_{[N\theta] + 1:|[N\theta]\lambda]}(s) \{ X_j(t) - \bar{X}_{[N\theta] + 1:|[N\theta]\lambda]}(t) \},
\]
and consider the statistics
\[
\hat{D}_{N}^{\text{cp,cov}} = \int_{T} \int_{T} D_{N}^{\text{cp,cov}}(s, t, 1, \hat{\theta})^2 ds \, dt,
\]
\[
\check{V}_{N}^{\text{cp,cov}} = \left[ \int_{0}^{1} \left\{ \int_{T} \int_{T} D_{N}^{\text{cp,cov}}(s, t, \lambda, \hat{\theta})^2 dt - \lambda^2 \int_{T} \int_{T} D_{N}^{\text{cp,cov}}(s, t, 1, \hat{\theta})^2 dt \right\}^2 \nu(d\lambda) \right]^{1/2},
\]
where $\nu$ is a probability measure on the interval $(0, 1)$. The test for hypotheses (4.4) rejects $H_0^C$, whenever
\[
\hat{D}_{N}^{\text{cp,cov}} > \Delta + q_{1-\alpha}(\mathbb{W}) \check{V}_{N}^{\text{cp,cov}},
\]
where $q_{1-\alpha}(\mathbb{W})$ is the $(1 - \alpha)$-quantile of the distribution of the random variable $\mathbb{W}$ in expression (2.9). For a precise statement of the next theorem, we define
\[
\tau_{C, \theta_0}^2 = 4 \int_{T} \int_{T} \int_{T} \int_{T} \delta_C(s, t) \delta_C(s', t') \left[ \frac{1}{\theta_0} K_1\{ (s, t), (s', t') \} + \frac{1}{1 - \theta_0} K_2\{ (s, t), (s', t') \} \right] ds \, dr \, ds' \, dt',
\]
where $\delta_C(s, t) = C_2(s, t) - C_1(s, t)$ and $K_j$ is the long-run covariance kernel of $\{ \eta^{(j)} \otimes \eta^{(j)} \}_{i \in \mathbb{Z}}$.

**Theorem 5.** Assume that the data are generated according to model (4.3), $\theta_0 \in (\varepsilon, 1 - \varepsilon)$, and that $(\varepsilon_j)_{j \in \mathbb{Z}}$ together with $f_1$ and $f_2$ satisfy assumptions 2, 3’ and 4’. Further assume that $\nu$ puts no mass in a neighbourhood of zero. Then decision rule (4.6) satisfies
\[
\lim_{n \to \infty} \mathbb{P}\{ \hat{D}_{N}^{\text{cp,cov}} > \Delta + q_{1-\alpha}(\mathbb{W}) \check{V}_{N}^{\text{cp,cov}} \} = \begin{cases} 0 & \text{if } d_{C}^{\text{cp}} < \Delta, \\ \alpha & \text{if } d_{C}^{\text{cp}} = \Delta \text{ and } \tau_{C, \theta_0}^2 > 0, \\ 1 & \text{if } d_{C}^{\text{cp}} > \Delta. \end{cases}
\]

Similarly to the setting of testing for a relevant change point in the mean it can be shown that rejection rule (4.6) does not lead to an asymptotic level $\alpha$ test when $\Delta = 0$.

## 5. Finite sample properties

In this section we illustrate the finite sample properties of the new procedures by means of a simulation study. Note that we must specify the measure $\nu$ that is used in the definition of the normalizer (2.6), (2.19) and (3.7) and we use $\nu = (1/19) \sum_{i=1}^{19} \delta_{i/20}$ throughout this section if not mentioned otherwise; here $\delta_{\lambda}$ denotes the Dirac measure at the point $\lambda \in [0, 1]$. For example, for this choice the quantity $\check{V}_n$ that is defined in expression (2.6) is given by
\[ \hat{V}_n = \left( \frac{1}{19} \sum_{i=1}^{19} \left\{ \int_T S_n^2(t, i/20) \, dt - \left( \frac{i}{20} \right)^2 \int_T S_n^2(t, 1) \, dt \right\} \right)^{1/2} \]

and the other expressions are obtained similarly. In the following sections we discuss the one-sample case, the two-sample case and change point detection separately. All results are based on 1000 simulation runs.

5.1. One-sample problems

We consider a process \( \{X_n\}_{n \in \mathbb{N}} \) with expectation function

\[ \mu(t) = \sqrt{2\delta} \sin(2\pi t) \]  

and different error processes, where we investigate a similar scenario to that in Aue et al. (2015) (see sections 6.3 and 6.4 in Aue et al. (2015)). More precisely, let \( X_n = \mu + \varepsilon_n \) and consider B-spline basis functions \( b_1, \ldots, b_D \) \((D \in \mathbb{N})\). Define the linear space \( \mathbb{H} = \text{span}\{b_1, \ldots, b_D\} \subset L^2([0, 1]) \) and independent processes \( \eta_1, \ldots, \eta_n \in \mathbb{H} \) by

\[ \eta_j = \sum_{i=1}^{D} N_{i,j} b_i, \quad j = 1, \ldots, n, \]  

where \( N_{1,1}, N_{2,1}, \ldots, N_{D,n} \) are independent \( N(0, \sigma_i^2) \) \((i = 1, \ldots, D; j = 1, \ldots, n)\) distributed random variables. Our first example considers independent error processes of the form

\[ \varepsilon_j = \eta_j, \quad j \in \mathbb{Z}, \]  

whereas the second example investigates a functional moving average fMA(1) process given by

\[ \varepsilon_j = \eta_j + \Theta \eta_{j-1}, \quad j \in \mathbb{Z}. \]  

Here the operator \( \Theta : \mathbb{H} \to \mathbb{H} \) (acting on finite dimensional spaces) is defined by the matrix \( \Theta = (\Theta_{ij})_{i,j=1}^{D} \in \mathbb{R}^{D \times D} \), where the entries \( \Theta_{ij} \) are normally distributed with mean 0 and standard deviation \( \kappa \sigma_i \sigma_j \) and \( \kappa \) is a scaling factor such that the resulting matrix \( \Theta \) has (induced) spectral norm equal to 0.7. The operator \( \Theta \) is newly generated in every simulation run (see sections 6.3 and 6.4 in Aue et al. (2015) for a similar approach) and we use \( D = 21 \). The third error structure under consideration is independent Brownian bridges.

In Fig. 1 we display the simulated rejection probabilities of test (2.8) for hypotheses (2.1), where \( \Delta = 0.02 \), which corresponds to the value \( \delta = 0.02 \) in model (5.1). These results show a pattern which is in line with the theoretical findings in theorem 1. For example at the boundary of the null hypotheses, i.e. for \( \delta = \Delta = 0.02 \), the simulated level is close to the nominal level. In the interior of the null hypothesis \( \delta < \Delta \) the simulated rejection probabilities are strictly smaller than \( \alpha = 0.05 \), whereas they are strictly larger than 0.05 in the interior of the alternative, i.e. \( \delta > \Delta \).

5.1.1. Estimating the long-run variance

It is of interest to compare the procedure based on self-normalization with the test (2.13) that is defined in remark 3, which uses an estimate of the long-run variance. For this comparison, we also implement the (practically infeasible) test which rejects the null hypothesis of no relevant difference whenever

\[ \hat{T}_n > \Delta + u_{1-\alpha} \frac{\tau}{\sqrt{n}}, \]  

where \( \Delta = 0.02 \) and \( u_{1-\alpha} \) is the critical value of the standard normal distribution.
i.e. we use the true asymptotic standard deviation $\tau$ instead of its estimate $\hat{\tau}_n$. Throughout this section we consider fAR(1) error processes defined by

$$\varepsilon_j = \eta_j + \kappa \varepsilon_{j-1}, \quad j \in \mathbb{Z},$$

(5.6)

for some $\kappa \in (0, 1)$ and expectation function $\mu$ as in equation (5.1). The random functions $\eta_j$, for $j = 1, \ldots, n$, are defined as in expression (5.2) (again with $D = 21$) but, in this section, we use the Fourier functions defined by $b_1 \equiv 1$ and

$$b_j(t) = \begin{cases} \sqrt{2} \sin(j\pi t), & j \text{ is even}, \\ \sqrt{2} \cos((j-1)\pi t), & j > 1 \text{ is odd} \end{cases}$$

(5.7)

as (orthonormal) basis functions such that an explicit calculation of the long-run variance becomes easier. Thus we have $\text{cov}\{\varepsilon_0(s), \varepsilon_i(t)\} = \text{cov}\{\eta_0(s), \eta_0(t)\} \kappa^i/(1 - \kappa)^2$ which yields
Fig. 2. Approximation of the test level for various values of \( \kappa \) (errors are fAR(1) processes defined by expression (5.6)) (■, true long-run variance; ●, estimated long-run variance; ▲, self-normalization): (a) \( \Delta = 0.5, n = 100 \); (b) \( \Delta = 1.5, n = 100 \); (c) \( \Delta = 0.5, n = 200 \); (d) \( \Delta = 1.5, n = 200 \)

\[
\tau^2 = 4 \int_T \int_T \mu(s)\mu(t)C(s,t)ds\ dt = \frac{4}{(1-\kappa)^2} \sum_{i=1}^D \frac{1}{T^2} \left( \int_0^1 \mu(t)b_i(t)dt \right)^2.
\]

To obtain an estimate of the long-run covariance function \( C \), we use the opt_bandwidth function from the R package fChange with the Bartlett kernel (both as kern_type and as kern_type_ini).

In Fig. 2, we compare the approximation of the nominal level of the three tests for various values of \( n, \kappa \) and \( \Delta \) at the boundary of the null hypothesis, i.e. for \( \int_T \mu^2(t)dt = \Delta \). We observe that the self-normalized test performs well across all settings considered with only a slight inflation of level for the most difficult case \( \kappa = 0.8, n = 100 \). In contrast, even for a large sample size \( n = 200 \), the tests based on the estimated and true (asymptotic) long-run variance exceed their nominal level for all values of \( \kappa \) considered with especially large over-rejections for \( \kappa > 0.5 \). Interestingly, the test based on the estimated long-run variance performs slightly better...
compared with the test with the true asymptotic long-run variance when \( n = 100 \). A similar pattern can be observed for data that are more heavy tailed. For brevity, additional details are deferred to section A.5 in the on-line supplement.

5.2. Two-sample problem

We begin considering the case of two independent (stationary) samples \( X_1, \ldots, X_m \) and \( Y_1, \ldots, Y_n \), with \( \mathbb{E}[X_j] = \mu_1 \) and \( \mathbb{E}[Y_j] = \mu_2 \), where the mean functions are given by

\[
\begin{align*}
\mu_1 &\equiv 0, \\
\mu_2(t) &= at(1-t)
\end{align*}
\]

(see section 4 in Horváth et al. (2013) for a similar approach), such that \( \int_0^1 \mu_2^2(t) dt = a^2/30 \). We are interested in testing hypotheses (2.16), i.e.

\[
H_0: \int_T D^2(t) dt \leq \Delta \quad \text{versus} \quad H_1: \int_T D^2(t) dt > \Delta,
\]

where \( D = \mu_1 - \mu_2 \) is the (unknown) difference of the two mean functions and the threshold is given by \( \Delta = 0.2^2/30 \) (note that this corresponds to the choice \( a = 0.2 \)). We consider independent samples, fMA(1) processes (generated as described in Section 5.1) and independent Brownian bridges as error processes.

In Fig. 3 we display the rejection probabilities of test (2.17) as a function of the parameter \( a \) for various sample sizes \( m \) and \( n \). We observe that the test yields a good approximation of the nominal level at the boundary \( a = 0.2^2/30 \) and detects the alternatives with reasonable power. Further results for dependent samples are presented in section A.2 of the on-line supplement and show a similar picture.

We conclude this section by investigating the effect of more heavy-tailed data and compare test (2.17) with the two tests that were obtained by the alternative self-normalizations in expressions (2.20) and (2.21). To be precise these tests reject the null hypothesis of no relevant difference, whenever

\[
\hat{D}_{m,n} > \Delta + q_{1-\alpha}(\mathbb{W}^*) \hat{\psi}_{m,n}^*,
\]

\[
\hat{D}_{m,n} > \Delta + q_{1-\alpha}(\mathbb{W}^{**}) \psi_{m,n}^*,
\]

where \( q_{1-\alpha}(\mathbb{W}^*) \) and \( q_{1-\alpha}(\mathbb{W}^{**}) \) are the \((1-\alpha)\)-quantiles of \( \mathbb{W}^* \) and \( \mathbb{W}^{**} \) in expression (2.22).

In Fig. 4(a), we display the rejection probabilities of tests (2.17), (5.9) and (5.10) in the situation that was considered in Fig. 3. More precisely, the sample sizes are \( m = 50 \) and \( n = 100 \), the mean functions are given by expression (5.8) and the error process is an fMA(1) process defined by expression (5.4). We observe very similar behaviour for all three tests under consideration.

Next we investigate a similar situation for more heavy-tailed data and consider similar error processes to those used in Kraus and Panaretos (2012), i.e.

\[
\eta_i(t) = \frac{1}{\sqrt{10}} \sum_{k=1}^{10} \{k^{-3/2}/\sqrt{2} \sin(2\pi kt) V_{i,k} + 3^{-k/2}/\sqrt{2} \cos(2\pi kt) W_{i,k}\},
\]

\[
\tilde{\eta}_j(t) = \frac{1}{\sqrt{10}} \sum_{k=1}^{10} \{k^{-3/2}/\sqrt{2} \sin(2\pi kt) \tilde{V}_{j,k} + 3^{-k/2}/\sqrt{2} \cos(2\pi kt) \tilde{W}_{j,k}\}
\]

\((i = 1, \ldots, m, j = 1, \ldots, n)\). Here the random variables \( V_{i,k}, \tilde{V}_{j,k}, \tilde{W}_{j,k} \) and \( W_{i,k} \) are independent \( t_5 \)-distributed random variables scaled to have unit variance. Fig. 4(b) shows the empirical rejec-
Fig. 3. Simulated rejection probabilities of test (2.17) for the relevant hypotheses (2.16) with $\Delta = 0.2^2/30$ (the mean functions are given by expression (5.8) and different independent error processes are considered) (———, $m = 50$, $n = 50$; ——, $m = 50$, $n = 100$; ———, $m = 100$, $n = 100$; ————, $m = 100$, $n = 200$): (a) independent error processes defined by expression (5.3); (b) fMA(1) processes defined by expression (5.4); (c) Brownian bridges

5.3. Change point problem

We begin considering model (3.1) with $\theta_0 = 0.5$, $\mu = 0$ and $\delta(t) = at(1-t)$, and the errors are IID from expression (5.3). The trimming parameter $\epsilon$ for estimating the change point location is set to 0.05. Data are generated with $a = 0, 0.02, \ldots, 0.5$ and then empirical rejection probabilities are calculated by using $\Delta = 0.1^2/30, 0.2^2/30, 0.3^2/30$. These probabilities are shown in Fig. 5.

From theorem 3, we expect that the probability of rejection should be close to $\alpha$ at the boundary of the hypotheses ($\int_0^1 D^2(t)dt = \Delta$), strictly smaller than $\alpha$ in the interior of the null hypothesis ($\int_0^1 D^2(t)dt < \Delta$) and larger than $\alpha$ in the interior of the alternative ($\int_0^1 D^2(t)dt > \Delta$).
Fig. 4. Simulated rejection probabilities of tests (2.17), (5.9) and (5.10) by using different self-normalizing factors (the mean functions are given by expression (5.8) and different independent error processes are considered with sample sizes $m = 50$ and $n = 100$): the threshold is defined as $\Delta = 0.2^2/30$ and the errors are fMA(1) processes given by (a) expression (5.4) and (b) expression (5.11).

This pattern is clearly observed for relevant hypotheses with threshold $\Delta \geq 0.2^2/30$. However, the test proposed is oversized if relevant hypotheses with $\Delta = 0.1^2/30$ are tested (see Fig. 5(a)). This is because change point tests for relevant hypotheses require a precise estimate of the change point. For small values of $a$ it is extremely difficult to estimate the true change point location, and an imprecise estimation of the change point results in a less accurate approximation of the nominal level. The difficulty of estimating the true change point location for small values of $a$
Fig. 5. Simulated rejection probabilities of test (3.8) for the relevant hypotheses (3.2) with (a) $\Delta = 0.1^2/30$, (b) $\Delta = 0.2^2/30$ and (c) $\Delta = 0.3^2/30$ (data are generated according to model (3.1) with $\theta_0 = 0.5$, $\mu = 0$ and $\delta(t) = a(1 - t)$, for $a = 0, 0.02, \ldots, 0.5$, and the errors are iid defined by expression (5.3); the tuning parameter is set to $\varepsilon = 0.05$): ————, $N = 200$; ————, $N = 500$

is further illustrated in Fig. 6 where we show the histogram of the corresponding estimator of the change point for $a = 0.1, 0.2, 0.3$ with sample size $N = 200$.

Next, we investigate the properties of our test with dependent error processes, i.e. we generate an fMA(1) process $\{\eta_i\}_{i \in \mathbb{Z}}$ as described in Section 5.1 and define

$$X_i = \mu + \eta_i, \quad i = 1, \ldots, \lfloor \theta_0 N \rfloor, \quad (5.12)$$

$$X_i = \mu + \delta \eta_i, \quad i = \lfloor \theta_0 N \rfloor + 1, \ldots, N,$$

$$X_i = \mu + \eta_i, \quad i = 1, \ldots, \lfloor \theta_0 N \rfloor, \quad (5.13)$$

$$X_i = \mu + \delta \sqrt{3} \eta_i, \quad i = \lfloor \theta_0 N \rfloor + 1, \ldots, N,$$

as the first and second scenario. The functions $\mu$ and $\delta$ are as described at the beginning of this
Fig. 6. Histogram of the change point estimator $\hat{\theta}$ defined in expression (3.3); size $N = 200$ data are generated according to model (3.1) with $\theta_0 = 0.5, \mu = 0$ and $\delta(t) = at(1-t)$, for $a = 0.1, 0.2, 0.3$, and the errors are IID defined by expression (5.3); the tuning parameter is set to $\varepsilon = 0.05$

section. The corresponding rejection probabilities of test (3.8) are depicted in Fig. 7 where we restrict our attention to the case $\Delta = 0.3^2/30$ for brevity. We find that for both error settings the test performs reasonably well.

5.4. Results for covariance operators
In this section we investigate the finite sample properties of the test for precise hypotheses regarding the covariance operators as introduced in Section 4.

5.4.1. Two-sample problem
For brevity we display only results for fMA(1) processes which are defined by

$$X_j = \eta_j + \kappa \eta_{j-1}, \quad Y_i = \tilde{\eta}_i + \kappa \tilde{\eta}_{i-1} \quad j = 1, \ldots, m, \quad i = 1, \ldots, n,$$

with $\kappa = 0.7$. The error processes are given by

$$\eta_j = \sum_{l=1}^{D} N_{l,j} b_l,$$
Fig. 7. Simulated rejection probabilities of test (3.8) for the relevant hypotheses (3.2) with $\Delta = 0.3^2/30$ in the case of fMA(1) samples (the mean function after the change point is given by expression (5.8) and the mean function before the change point is the zero function) $(\ldots, N = 200; \ldots, N = 500)$: (a) error processes defined by expression (5.12); (b) error processes defined by expression (5.13).

\[ \tilde{\eta}_i = a \sum_{l=1}^{D} N'_{i,j'} b_l, \]

for $j = 1, \ldots, m$ and $i = 1, \ldots, n$, where the coefficients $N_{i,j}$ and $N'_{i,j'}$ are independent $N(0, \sigma^2)$ ($i = 1, \ldots, D = 21, j = 1, \ldots, m, j' = 1, \ldots, n$) distributed random variables and the (orthonormal) basis functions $b_1, \ldots, b_D$ are defined in expression (5.7) ($b_1 \equiv 1$). Similarly to Section 6.3 in Aue et al. (2015), we consider two scenarios for the variance structure of the random coefficients, namely, for any $j = 1, \ldots, m$ and $j' = 1, \ldots, n$, scenarios A and B respectively:
Fig. 8. Simulated rejection probabilities of test (4.2) for hypotheses (4.1) of a relevant difference between the covariance operators of two FMA(1) processes ($\Delta = (1 - 1.5^2 \gamma_i \sigma_i^4 (1 + 0.7^2) \gamma_i) \quad (\text{---}, \ m = 100, \ n = 100; \quad ---, \ m = 200, \ n = 200; \quad ----, \ m = 500, \ n = 500)$: (a) variance scenario A; (b) variance scenario B

$$\begin{align*}
\sigma_i^2 &= \text{var}(N_{i,j}) = \text{var}(N_{i,j}^\prime) = 1/i^2, \quad i = 1, \ldots, D, \\
\sigma_i^2 &= 1.2^{-2i}, \quad i = 1, \ldots, D. 
\end{align*}$$

(5.15)

In this case $X_i$ is a multiple of $Y_j$ in distribution and the distance between the covariance operators is given by

$$\int_0^1 \int_0^1 D^2(s,t)\text{d}s\text{d}t = (1 - a^2)^2 \sum_{i=1}^{D} \sigma_i^4 (1 + \kappa^2)^2$$
Fig. 9. Simulated rejection probabilities of test (4.6) for the hypotheses (4.4) of a relevant change point in the covariance operator of an fMA(1) process \( \Delta = (1 - 1.5^2)^2 \sum_{i=1}^D \sigma_i^2 (1 + 0.7^2)^2 \) \( \sigma_0 = 0.72 \): (a) variance scenario A; (b) variance scenario B (see Paparoditis and Sapatinas (2016) for a similar approach). The empirical rejection probabilities of test (4.2) for various values of \( a \) are displayed in Fig. 8, where the case \( a = 1.5 \) corresponds to the boundary of the hypotheses. We observe a similar pattern to that for the comparison of the mean functions, where test (4.2) is slightly more conservative in variance scenario A. Additional results with independent data and heavy-tailed errors show a similar picture and can be found in section A.3 of the on-line supplement.

5.4.2. Change point problem

In this section we investigate the test for a relevant change in the covariance operator, which
Table 2. Summary of the two-sample test for relevant hypotheses with varying $\Delta$ for the annual temperature curves†

| $\Delta$ | 99% quantile | 95% quantile | 90% quantile |
|----------|--------------|--------------|--------------|
| 9.0      | True         | True         | True         |
| 9.1      | False        | True         | True         |
| 10.7     | False        | True         | True         |
| 10.8     | False        | False        | True         |
| 11.7     | False        | False        | True         |
| 11.8     | False        | False        | False        |

†The label ‘True’ refers to a rejection of the null, and the label ‘False’ to a failure to reject the null.

was developed in Section 4.2. For this we consider an fMA(1) process $X'_1, \ldots, X'_N$ defined by expression (5.14) (with $\kappa = 0.7$), where the basis functions and variances $\sigma^2_1, \ldots, \sigma^2_D$ are given by expressions (5.7) and (5.15) respectively. The data $X_1, \ldots, X_N$ are defined by

$$X_j = \begin{cases} X'_j, & j \leq \lfloor N\theta_0 \rfloor, \\ aX'_j, & j > \lfloor N\theta_0 \rfloor \end{cases}$$

for $j = 1, \ldots, N$, where the change point is given by $\theta_0 = 0.5$. In Fig. 9 we show the rejection probabilities of test (4.6) for hypotheses (4.4), where the threshold is given by

$$\Delta = (1 - 1.5^2)^2 \sum_{i=1}^{D} \sigma_i^4(1 + 0.7^2)^2.$$ 

Overall, we observe similar behaviour to that for the test for a relevant change point in the mean functions. Further simulations with independent and heavy-tailed data show similar patterns; see section A.4 of the on-line supplement.

5.5. Data illustrations

5.5.1. Two-sample test

In this section we consider an application of the methodology that was developed in Section 2.2 to Australian temperature data. The data consist of daily minimum temperatures collected at various meteorological stations in Australia. Following Fremdt et al. (2014) we project the daily values of each year on a Fourier basis consisting of 49 basis functions, resulting in annual temperature curves for each location under consideration. Fremdt et al. (2014) investigated the temperature data to illustrate methodology that was designed to choose the dimension of the projection space obtained with functional principal component analysis and in Aue and van Delft (2019) the data were considered in the context of stationarity tests for functional time series.

We investigate annual data curves obtained from the meteorological stations in Cape Otway (1865–2011) and Sydney (1859–2011). Cape Otway is in the south of Australia and Sydney is a city on the eastern coast of Australia. There is a distance of approximately 1000 km between them such that differences in the temperature profiles are expected and the task of the relevant two-sample test is now to specify how big the difference might be. The samples consist of $m = 147$ and $n = 153$ temperature curves respectively.

To calculate the test decision (2.17) for the hypotheses that are defined in expression (2.16), we
Fig. 10. (a) Mean functions of the Cape Otway (—) and Sydney (---) series for the two-sample case and (b) mean curves of the river flow for the periods 1910–1964 (——) and 1965–2014 (- - - -)

The results in Table 2 provide no evidence for an integrated squared mean difference that is larger than $\Delta = 11.8$ but at the other extreme there is strong evidence that it exceeds $\Delta = 9$. Choosing $\Delta$ between 9.1 and 10.7 led to rejecting the null at level $\alpha \geq 5\%$ and for $\Delta \in [10.8, 11.7]$, the test rejected the null only at level $\alpha \geq 10\%$, which means weaker support of the alternative.
Table 3. Summary of the change point test for relevant hypotheses with varying \( \Delta \) for the annual river flow curves†

| \( \Delta \) | 99\% quantile | 95\% quantile | 90\% quantile |
|-------------|---------------|---------------|---------------|
| 0.72        | False         | False         | True          |
| 0.73        | False         | False         | False         |

†The labels ‘True’ and ‘False’ refer to a rejection of the null and failure to reject the null respectively.

5.5.2. Change point test

In this section we consider daily flows (in cubic metres per second) of the river Chemnitz at Göritzhain (in the east of Germany), where data were recorded for the years 1909–2014. One hydrological year (different definitions are possible but we consider the same as Sharipov et al. (2016)) starts at November 1st and ends at October 31st, which means that we consider the hydrological years 1910–2014. Note that Sharipov et al. (2016) considered the years 1910–2012. We regard data from each year as one flow curve resulting in a sample of size \( N = 105 \).

In the definition of the change point estimator, we use the trimming parameter \( \varepsilon = 0.1 \) and obtain the year 1964 as a possible change point. This is the same year as was identified by Sharipov et al. (2016). In Fig. 10(b), we display the mean of the curves before and after the estimated change point. Applying the test defined by expression (3.8) for two values of \( \Delta \) leads to the test decisions in Table 3. For \( \Delta = 0.72 \), we reject the null hypothesis of no relevant change at level \( \alpha = 0.1 \) and we do not reject the null at level \( \alpha = 0.05 \). For \( \Delta \geq 0.73 \), we do not reject the null at the test levels under consideration.

Acknowledgements

This work has been supported in part by the Collaborative Research Center ‘Statistical modeling of nonlinear dynamic processes’ (Sonderforschungsbereich 823, Teilprojekt A1, C1 and T1) of the German Research Foundation and a discovery grant (to Stanislav Volgushev) from the Natural Sciences and Engineering Research Council of Canada. Parts of this paper were written while H. Dette was visiting the Department of Statistics, University of Toronto, and this author thanks the institute for its hospitality. The authors thank Tim Kutta for helpful discussions regarding some of the material on covariance operators. The authors are also grateful to the two referees and the Associate Editor whose constructive comments led to a substantial improvement of an earlier version of this paper.

References

Aston, J. A. and Kirch, C. (2012) Detecting and estimating changes in dependent functional data. *J. Multiv. Anal.*, 109, 204–220.
Aue, A. and van Delft, A. (2019) Testing for stationarity of functional time series in the frequency domain. *Ann. Statist.*, to be published.
Aue, A., Dubart Norinho, D. and Hörmann, S. (2015) On the prediction of stationary functional time series. *J. Am. Statist. Ass.*, 110, 378–392.
Aue, A., Rice, G. and Sönmez, O. (2020) Structural break analysis for spectrum and trace of covariance operators. *Environmetrics*, 31, article env.2617.
Benko, M., Haerdle, W. and Kneip, A. (2009) Common functional principal components. *Ann. Statist.*, 37, 1–34.
Berger, J. O. and Delampady, M. (1987) Testing precise hypotheses. *Statist. Sci.*, 2, 317–335.
Berkes, I., Gabrys, R., Horváth, L. and Kokoszka, P. (2009) Detecting changes in the mean of functional observations. *J. R. Statist. Soc. B*, 71, 927–946.

Berkes, I., Horváth, L. and Rice, G. (2013) Weak invariance principles for sums of dependent random functions. *Stoch. Processes Appl.*, 123, 385–403.

Berkson, J. (1938) Some difficulties of interpretation encountered in the application of the chi-square test. *J. Am. Statist. Ass.*, 33, 526–536.

Bucchia, B. and Wendler, M. (2017) Change-point detection and bootstrap for Hilbert space valued random fields. *J. Multiv. Anal.*, 155, 344–368.

Bücher, A., Dette, H. and Heinrichs, F. (2019) Detecting deviations from second-order stationarity in locally stationary functional time series. *Ann. Inst. Statist. Math.*, to be published.

Chow, S.-C. and Liu, P.-J. (1992) *Design and Analysis of Bioavailability and Bioequivalence Studies*. New York: Dekker.

Cuevas, A., Febrero, M. and Fraiman, R. (2006) On the use of the bootstrap for estimating functions with functional data. *Computat Statist. Data Anal.*, 51, 1063–1074.

Ferraty, F. and Vieu, P. (2010) *Nonparametric Functional Data Analysis*. New York: Springer.

Fremdt, S., Horváth, L., Kokoszka, P. and Steinebach, J. G. (2014) Functional data analysis with increasing number of projections. *J. Multiv. Anal.*, 124, 313–332.

Fremdt, S., Steinebach, J. Horváth, L. and Kokoszka, P. (2013) Testing the equality of covariance operators in functional samples. *Scand. J. Statist.*, 40, 138–152.

Fryzlewicz, P. (2014) Wild binary segmentation for multiple change-point detection. *Ann. Statist.*, 42, 2243–2281.

Guo, J., Zhou, B. and Zhang, J.-T. (2016) A supremum-norm based test for the equality of several covariance functions. *Computat Statist. Data Anal.*, 124, 15–26.

Hall, P. and Van Keilegom, I. (2007) Two-sample tests in functional data analysis starting from discrete data. *Statist. Sin.*, 17, 1511–1531.

Harchaoui, Z. and Lévy-Leduc, C. (2010) Multiple change-point estimation with a total variation penalty. *J. Am. Statist. Ass.*, 105, 1480–1493.

Hörmann, S. and Kokoszka, P. (2010) Weakly dependent functional data. *Ann. Statist.*, 38, 1845–1884.

Horváth, L., Kokoszka, P. and Reimherr, M. (2009) Two sample inference in functional linear models. *Can. J. Statist.*, 37, 571–591.

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

Horváth, L., Kokoszka, P. and Reeder, R. (2013) Estimation of the mean of functional time series and a two-sample problem. *J. R. Statist. Soc. B*, 75, 103–122.

Horváth, L., Kokoszka, P. and Rice, G. (2014) Testing stationarity of functional time series. *J. Econometr.*, 179, 66–82.

Kokoszka, P. (2012) Dependent functional data. *ISRN Probab. Statist.*, article 958254.

Kraus, D. and Panaretos, V. M. (2012) Dispersion operators and resistant second-order functional data analysis. *Biometrika*, 99, 813–832.

Munk, A. and Pflüger, R. (1999) $(1-\alpha)$-equivariant confidence rules for convex alternatives are $\alpha/2$ level tests—with applications to the multivariate assessment of bioequivalence. *J. Am. Statist. Ass.*, 94, 1311–1319.

Panaretos, V. M., Kraus, D. and Maddocks, J. H. (2010) Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *J. Am. Statist. Ass.*, 105, 670–682.

Paparoditis, E. and Sapatinas, T. (2016) Bootstrap-based testing of equality of mean functions or equality of covariance operators for functional data. *Biometrika*, 103, 727–733.

Pilavakis, D., Paparoditis, E. and Sapatinas, T. (2019) Testing equality of autocovariance operators for functional time series. *Preprint arXiv:1901.08535*.

Pomann, G.-M., Staicu, A.-M. and Ghosh, S. (2016) A two-sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis. *Appl. Statist.*, 65, 395–414.

Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*, 2nd edn. New York: Springer.

Shao, X. (2010) A self-normalized approach to confidence interval construction in time series. *J. R. Statist. Soc. B*, 72, 343–366.

Shao, X. (2015) Self-normalization for time series: a review of recent developments. *J. Am. Statist. Ass.*, 110, 1797–1817.

Shao, X. and Zhang, X. (2010) Testing for change points in time series. *J. Am. Statist. Ass.*, 105, 1228–1240.

Sharipov, O., Tews, J. and Wendler, M. (2016) Sequential block bootstrap in a Hilbert space with application to change point analysis. *Can. J. Statist.*, 44, 300–322.

Sharipov, O. S. and Wendler, M. (2019) Bootstrapping covariance operators of functional time series. *Preprint arXiv:1904.06721*.

Stoehr, C., Aston, J. and Kirch, C. (2019) Detecting changes in the covariance structure of functional time series with application to fMRI data. *Preprint arXiv:1903.00288*.

Vostríkova, L. Y. (1981) Detecting “disorder” in multidimensional random processes. *Sov. Math. Dokl.*, 24, 55–59.

Wang, W., Gene Hwang, J. and Dasgupta, A. (1999) Statistical tests for multivariate bioequivalence. *Biometrika*, 86, 395–402.

Wellek, S. (2010) *Testing Statistical Hypotheses of Equivalence and Noninferiority*, 2nd edn. Boca Raton: CRC Press.
Zhang, C., Peng, H. and Zhang, J.-T. (2010) Two samples tests for functional data. *Communs Statist. Theory Meth.*, 39, 559–578.
Zhang, T. and Lavitas, L. (2018) Unsupervised self-normalized change-point testing for time series. *J. Am. Statist. Ass.*, 113, 637–648.
Zhang, X. and Shao, X. (2015) Two sample inference for the second-order property of temporally dependent functional data. *Bernoulli*, 21, 909–929.
Zhang, X., Shao, X., Hayhoe, K. and Wuebbles, D. J. (2011) Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electron. J. Statist.*, 5, 1765–1796.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

‘Supplement to “Testing relevant hypotheses in functional time series via self-normalization”’. 