

# Detecting structural breaks in eigensystems of functional time series

Holger Dette

*Ruhr-Universität Bochum, D-44780 Bochum, e-mail: [Holger.Dette@rub.de](mailto:Holger.Dette@rub.de)*  
and

Tim Kutta

*Ruhr-Universität Bochum, D-44780 Bochum, e-mail: [Tim.Kutta@rub.de](mailto:Tim.Kutta@rub.de)*

**Abstract:** Detecting structural changes in functional data is a prominent topic in statistical literature. However not all trends in the data are important in applications, but only those of large enough influence. In this paper we address the problem of identifying relevant changes in the eigenfunctions and eigenvalues of covariance kernels of  $L^2[0, 1]$ -valued time series. By self-normalization techniques we derive pivotal, asymptotically consistent tests for relevant changes in these characteristics of the second order structure and investigate their finite sample properties in a simulation study. The applicability of our approach is demonstrated analyzing German annual temperature data.

**MSC2020 subject classifications:** 62F05, 62M10.

**Keywords and phrases:** Functional time series, relevant changes, eigenfunctions, eigenvalues, self-normalization.

Received December 2019.

## 1. Introduction

The analysis of functional data has gained increasing attention during the past decades, due to recent advances in computing and data collecting technologies. This surging interest is testified by a rapidly expanding scope of new statistical methods, as reviewed in the monographs of Bosq (2000), Ramsay and Silverman (2005), Ferraty and Vieu (2010), Horváth and Kokoszka (2012) and Hsing and Eubank (2015).

Applications of functional data analysis include such diverse topics as imaging, meteorology, genomics, and economics. The analytical link between these fields lies in modelling observations as random functions, whether they are temperature curves or stock prices. While this approach facilitates the data's interpretation for users, it in exchange poses theoretical challenges, since each observation is now an element of a complex function space.

Consequently dimension reducing procedures play a key role in this field, as they make functional data amenable to the methods of finite dimensional statistics. Among these, *functional principal component analysis* (fPCA) has taken

the most prominent position. As principal component analysis (PCA) for finite dimensions, fPCA is based on projecting the data on linear subspaces that explain most of its variance. These spaces are spanned by the eigenfunctions of the estimated covariance operator. An overview of the mathematical aspects of this procedure can be found in the monographs of Ramsay and Silverman (2005) and Horváth and Kokoszka (2012) and in the survey of Shang (2014). Recently, Aue et al. (2015) used functional principal components for prediction in functional autoregressive models, Shang (2017) constructed forecasts with dynamic updating based on functional principal component regression and Gao et al. (2019) proposed *dynamic* fPCA for forecasting mortality rates. Obviously dimension reducing procedures rest upon the assumption of “stable” eigensystems, i.e. that most of the variance of all data is confined to the same, low-dimensional subspace. This insight has furnished interest in methods to validate this assumption.

One option to investigate the stability of the eigensystem is change point analysis, where one is monitoring a functional time series for structural breaks in the corresponding covariance operators. In particular, Aston and Kirch (2012) and Stoehr et al. (2019) develop a powerful methodology to detect changes in the covariance operator. Similarly changes in the cross-covariance operator have been investigated by Rice and Shum (2019). However, with the specific focus on fPCA it might be reasonable to conduct a refined analysis and to search directly for changes in the eigenfunctions and eigenvalues of the covariance operator. Yet, not much literature can be found in this direction.

The present paper contributes to this problem in two respects. On the one hand we develop a new statistical methodology for change point analysis of the eigenvalues and eigenfunctions of a covariance operator of a functional times series. While a test for a change in the spectrum of the covariance operator has already been proposed by Aue et al. (2018), we are – to the best of our knowledge – not aware of any procedure addressing the problem of change point detection in the eigenfunctions corresponding to a sequence of functional data. On the other hand – in contrast to the cited work, which has its focus on the “classical hypotheses” of strict equality- we propose tests for “precise” hypotheses as introduced in Berger and Delampady (1987). This means that we are NOT interested in asserting arbitrarily small differences in the eigensystem before and after the change versus the hypothesis of exact equality. Rather, we try to detect or reject changes of relevant magnitude.

For example, for the maximal eigenvalue of the covariance operator of the random function  $X_n$ , say  $\tau_{1,n}$ , this means that – in contrast to Aue et al. (2018) – we do not consider the null hypothesis  $\tau_{1,1} = \dots = \tau_{1,N}$ , but develop a statistical methodology to test the hypothesis of no relevant deviation of the eigenvalue before and after the change point, that is

$$H_0 : |\tau_1^{(1)} - \tau_1^{(2)}|^2 \leq \Delta_\tau \quad vs. \quad H_1 : |\tau_1^{(1)} - \tau_1^{(2)}|^2 > \Delta_\tau.$$

Here  $\tau_1^{(1)} := \tau_{1,1} = \dots = \tau_{1, \lfloor N\theta_0 \rfloor} \neq \tau_{1, \lfloor N\theta_0 \rfloor + 1} = \dots = \tau_{1,N} =: \tau_1^{(2)}$  for some  $\theta_0 \in (0, 1)$  and  $\Delta_\tau$  is a given threshold defined by the concrete application (under the null hypothesis the change is considered as not relevant). The hypotheses

regarding the other eigenvalues and eigenfunctions are formulated similarly (see equations (2.4) and (2.5) for more details).

The consideration of relevant hypotheses in the context of change point detection has been introduced in Dette and Wied (2016) and is motivated by the observation that in many applications one is not interested in small changes of a parameter. For example, in forecasting of functional times series, it is not reasonable to use only part of the data if a structural break in an eigenvalue (or eigenfunction) is detected, but the difference before and after the change is rather small. In this case discarding the data before the change could admittedly reduce the prediction bias, but come at the cost of a substantially increased variance due to a smaller sample size used for prediction.

Relevant hypotheses have been considered in statistics to different degrees since the mid 20th century (see Hodges and Lehmann (1954)), and have been investigated intensively in biostatistics, where tests for “bioequivalence” of certain drugs have nowadays become standard (see for example Wellek (2010)). In the context of change point analysis for functional data relevant hypotheses have recently been considered by Dette et al. (2019) for Banach-space valued random variables and by Dette et al. (2018) in Hilbert spaces. The first named paper concentrates on inference regarding the mean functions while Dette et al. (2018) developed tests for a relevant structural break in the mean function and in the covariance operator. The detection of structural breaks in the eigenvalues and eigenfunctions is a substantially more difficult problem due to their implicit definition and statistical tests have mainly been developed for the two sample case (see Zhang and Shao (2015), who consider classical hypotheses and Aue et al. (2019), who discuss relevant hypotheses).

The aim of the present work is to develop statistical methodology for detecting relevant changes in the eigensystem of a functional time series. In Section 2 we introduce the testing problems, define corresponding test statistics and give the main theoretical results. Typically in change point problems of this type estimation of the long run covariance structures is required, which is nearly intractable in the present context, because it involves all eigenvalues and eigenfunctions of the covariance operators before and after the change point (see, for example, Dauxois et al. (1982) or Hall and Hosseini-Nasab (2006) for an explicit representation of the estimated eigenvalues and eigenfunctions in terms of the empirical covariance operator). We propose a self-normalization approach which avoids this problem. In Section 3 we illustrate our approach by virtue of a small simulation study, as well as the investigation of the German weather data. Finally, in Appendices A and B we provide the proofs of our findings and also give some auxiliary results.

## 2. Testing for relevant changes

In this section we provide a precise outline of the testing problems considered in this paper and subsequently present the main theoretical results. Let  $L^2[0, 1]$  denote the Hilbert space of square integrable functions  $f : [0, 1] \rightarrow \mathbb{R}$  equipped

with the common inner product

$$\langle f, g \rangle := \int_0^1 f(t)g(t)dt \quad f, g \in L^2[0, 1].$$

The corresponding norm is denoted by  $\|\cdot\|$ . Notice that according to the induced metric, functions that differ only on a set of Lebesgue mass 0 are identified.

Now suppose we observe a sample of  $N$  random functions  $X_1, \dots, X_N \in L^2[0, 1]$ , where for any  $n \in \{1, \dots, N\}$

$$X_n := \begin{cases} X_n = X_n^{(1)}, & n \leq \theta_0 N \\ X_n = X_n^{(2)}, & n > \theta_0 N. \end{cases} \tag{2.1}$$

Here  $(X_n^{(1)})_{n \in \mathbb{Z}}$  and  $(X_n^{(2)})_{n \in \mathbb{Z}}$  are stationary sequences of centered, random functions in  $L^2[0, 1]$  and  $\theta_0 \in (0, 1)$  is a constant of proportionality. The assumption of vanishing expectations is made for the sake of a simple notation and all results presented in this paper hold in the case  $\mathbb{E}X_1^{(i)} = \mu^{(i)}$  for some  $\mu^{(i)} \in L^2[0, 1]$  ( $i = 1, 2$ ). For a more detailed discussion of this case see Remark 2.7. A general definition of expectations of random functions in  $L^2[0, 1]$  can be found in Bücher et al. (2019). However in the subsequent discussion we will always assume that

$$\mathbb{E}\|X_1^{(i)}\|^2 < \infty, \quad \text{for } i = 1, 2, \tag{2.2}$$

which implies that expectations can be defined point-wise (compare Horváth and Kokoszka (2012), Section 2.2).

Under assumption (2.2) the covariance kernel  $c^{(i)}$  of  $X_1^{(i)}$  ( $i = 1, 2$ ) is almost everywhere defined and given by

$$c^{(i)}(s, t) = \mathbb{E}[X_1^{(i)}(s)X_1^{(i)}(t)].$$

Regarded as a function with two arguments it is an element of  $L^2([0, 1] \times [0, 1])$ , the space of square integrable functions on the unit square, which can be isomorphically identified with the tensor product Hilbert space  $L^2[0, 1] \otimes L^2[0, 1]$  (for details see Weidmann (1980)). We will also denote the induced norm of this space by  $\|\cdot\|$ , since it will always be clear from the context, which space we refer to. By Mercer’s theorem (see König (1986) p. 145) the kernels  $c^{(1)}$  and  $c^{(2)}$  permit the  $L^2$ -expansions

$$c^{(i)}(s, t) = \sum_{j \geq 1} \tau_j^{(i)} v_j^{(i)}(s)v_j^{(i)}(t), \quad (i = 1, 2),$$

where  $v_1^{(i)}, v_2^{(i)}, \dots \in L^2[0, 1]$  are the eigenfunctions of the integral operator

$$g \mapsto \int_0^1 c^{(i)}(\cdot, t)g(t)dt, \quad g \in L^2[0, 1] \tag{2.3}$$

and  $\tau_1^{(i)}, \tau_2^{(i)}, \dots$  are the corresponding eigenvalues. For simplicity of reference we assume for some fixed  $p \in \mathbb{N}$  that the first  $p + 1$  eigenvalues are the largest and

that they are arranged in descending order, i.e.  $\tau_1^{(i)} \geq \dots \geq \tau_{p+1}^{(i)}$ . Furthermore the sets of eigenfunctions are supposed to form orthonormal bases of the space  $L^2[0, 1]$ , which can always be enforced by adding further, orthogonal functions (with corresponding eigenvalues 0).

Based on the sample of observations  $X_1, \dots, X_N$  we want to investigate relevant changes in the eigensystems corresponding to  $c^{(1)}$  and  $c^{(2)}$ . More precisely, for some  $j \in \{1, \dots, p\}$  we test whether the difference of the  $j$ -th eigenvalues  $\tau_j^{(1)}$  and  $\tau_j^{(2)}$  or the  $j$ -th eigenfunctions  $v_j^{(1)}$  and  $v_j^{(2)}$  exceeds a predetermined threshold. To be precise, we consider for a fixed index  $j \in \{1, \dots, p\}$  the hypotheses

$$H_0 : (\tau_j^{(1)} - \tau_j^{(2)})^2 \leq \Delta_\tau \quad \text{vs.} \quad H_1 : (\tau_j^{(1)} - \tau_j^{(2)})^2 > \Delta_\tau \quad (2.4)$$

and

$$H_0 : \|v_j^{(1)} - v_j^{(2)}\|^2 \leq \Delta_v \quad \text{vs.} \quad H_1 : \|v_j^{(1)} - v_j^{(2)}\|^2 > \Delta_v. \quad (2.5)$$

Here  $\Delta_\tau$  and  $\Delta_v$  are prespecified constants, denoting the maximal values for which the distances between the eigenvalues and eigenfunctions are still considered scientifically irrelevant. The particular choice of  $\Delta_\tau$  and  $\Delta_v$  depends on the concrete application. Note also that for  $\Delta_\tau = 0$  or  $\Delta_v = 0$  the hypotheses (2.4) and (2.5) reduce to the classical change point detection problems for eigenvalues and eigenfunctions respectively.

In order to decide whether a relevant change either in the eigenvalues or in the eigenfunctions has occurred we first have to identify the change point  $\theta_0$ .

### 2.1. Change point estimation

The change point estimator is constructed by the CUSUM principle and defined by

$$\hat{\theta} := \frac{1}{N} \operatorname{argmax}_{N\varepsilon \leq k \leq N(1-\varepsilon)} f(k), \quad (2.6)$$

where the function  $f$  is given by

$$f(k) := \frac{k(N-k)}{N^2} \int_0^1 \int_0^1 \left[ \frac{1}{k} \sum_{n=1}^k X_n(s)X_n(t) - \frac{1}{N-k} \sum_{n=k+1}^N X_n(s)X_n(t) \right]^2 ds dt.$$

Note that in definition (2.6) we confine the maximization of  $f$  to a subset of  $\{1, \dots, N\}$  to obtain stable estimators. In practice this restriction is not an issue and even very small values for  $\varepsilon$  can be used in (2.6). We refer to Section 3.3, where we demonstrate the stability of the estimator with respect to the choice of the threshold  $\varepsilon$  by means of a simulation study. In the same section (Remark 3.1) we discuss advantages and disadvantages of our change point estimator and consider alternative approaches. Before we proceed we specify the basic assumptions required for the theoretical statements presented in this paper.

**Assumption 2.1.** Let  $i = 1, 2$ .

1. The sequence of random functions  $(X_n^{(i)})_{n \in \mathbb{Z}}$  consists of Bernoulli shifts, i.e.

$$X_n^{(i)} = g^{(i)}(\epsilon_n, \epsilon_{n-1}, \dots)$$

for some measurable, non-random function  $g^{(i)} : S^\infty \rightarrow L^2[0, 1]$  and i.i.d. innovations  $\{\epsilon_n\}_{n \in \mathbb{Z}}$  with values in a measurable space  $S$ .

2. All random variables  $X_n^{(i)}(t, \omega)$  are jointly measurable in  $(t, \omega)$  for all  $n \in \mathbb{Z}$ .
3.  $\mathbb{E}X_0^{(i)} = 0$  and  $\mathbb{E}\|X_0^{(i)}\|^{4+\delta} < \infty$ , for some  $\delta \in (0, 1)$ .
4. The random functions  $X_n^{(i)}$  are  $m$ -approximable. In other words: for all  $m \in \mathbb{N}$  there exist i.i.d. copies  $\{\epsilon_{n,m,k}^*\}_{k \in \mathbb{Z}}$  of  $\epsilon_0$ , independent of  $\{\epsilon_j\}_{j \in \mathbb{Z}}$ , such that the  $m$ -dependent random functions  $X_{n,m}^{(i)}$

$$X_{n,m}^{(i)} := g^{(i)}(\epsilon_n, \dots, \epsilon_{n-m+1}, \epsilon_{n,m,n-m}^*, \epsilon_{n,m,n-m-1}^*, \dots) \quad (2.7)$$

satisfy

$$\sum_{m \geq 1} (\mathbb{E}\|X_n^{(i)} - X_{n,m}^{(i)}\|^{4+\delta})^{1/\kappa} < \infty, \quad (2.8)$$

for some  $\delta \in (0, 1)$  and  $\kappa > 4 + \delta$ .

Note that these assumptions match those in Berkes et al. (2013), who derived weak invariance principles for  $m$ -approximable sequences. However, the stronger summability condition (2.8) is imposed here, since we are not estimating mean functions, but covariance kernels. We now state a first result concerning the convergence rate of the change point estimator defined in (2.6). The proof follows by similar arguments as given in the proof of Proposition 3.1 in Dette et al. (2018), which are omitted for the sake of brevity.

**Proposition 2.2.** Suppose that Assumption 2.1 holds and that  $c^{(1)} \neq c^{(2)}$ . If  $\varepsilon < \min\{\theta_0, 1 - \theta_0\}$ , then

$$\hat{\theta} = \theta_0 + o_P(1/\sqrt{N}).$$

In the next step we partition the data into two subgroups  $X_1, \dots, X_{N\hat{\theta}}$  and  $X_{N\hat{\theta}+1}, \dots, X_N$ , which we then use to estimate the covariance kernels  $c^{(1)}$  and  $c^{(2)}$  respectively. To be precise, we define for  $\lambda \in [0, 1]$  the estimators

$$\hat{c}^{(1)}(\lambda, s, t) = \frac{1}{\lfloor N\hat{\theta}\lambda \rfloor} \sum_{n=1}^{\lfloor N\hat{\theta}\lambda \rfloor} X_n(s)X_n(t) \quad (2.9)$$

$$\hat{c}^{(2)}(\lambda, s, t) = \frac{1}{\lfloor (N - N\hat{\theta})\lambda \rfloor} \sum_{n=N\hat{\theta}+1}^{N\hat{\theta} + \lfloor (N - N\hat{\theta})\lambda \rfloor} X_n(s)X_n(t) \quad (2.10)$$

and obtain  $\hat{c}^{(1)}(\cdot, \cdot) := \hat{c}^{(1)}(1, \cdot, \cdot)$  and  $\hat{c}^{(2)}(\cdot, \cdot) := \hat{c}^{(2)}(1, \cdot, \cdot)$  as estimators for  $c^{(1)}$  and  $c^{(2)}$  respectively. In (2.9) and (2.10) the quantity  $\lambda \in [0, 1]$  denotes a parameter used for the subsequent self-normalization. If the kernels are degenerate they are interpreted as 0-functions.

By Proposition 2.2 the amount of misspecified data is small and therefore we expect the estimated kernels  $\hat{c}^{(1)}(\cdot, \cdot)$  and  $\hat{c}^{(2)}(\cdot, \cdot)$ , and hence their eigensystems to be close to those of  $c^{(1)}$  and  $c^{(2)}$ .

**2.2. Relevant changes in the eigenvalues**

We now proceed to construct a test for the hypothesis (2.4) of a relevant change in the  $j$ -th eigenvalue. For this purpose we define for  $i = 1, 2$  and  $\lambda \in [0, 1]$  the eigenfunctions

$$\hat{v}_{1,\lambda}^{(i)}, \hat{v}_{2,\lambda}^{(i)}, \dots \tag{2.11}$$

and the eigenvalues

$$\hat{\tau}_{1,\lambda}^{(i)} \geq \hat{\tau}_{2,\lambda}^{(i)} \geq \dots \tag{2.12}$$

of the estimates  $\hat{c}^{(i)}(\lambda, \cdot, \cdot)$  (defined in (2.9) and (2.10)). Finally we denote by

$$\hat{\tau}_j^{(i)} := \hat{\tau}_{j,1}^{(i)} ; \quad \hat{v}_j^{(i)} := \hat{v}_{j,1}^{(i)} \quad j = 1, 2, \dots \tag{2.13}$$

the eigensystems of the estimated covariance operators of the full samples  $X_1, \dots, X_{\lfloor N\hat{\theta} \rfloor}$  ( $i = 1$ ) and  $X_{\lfloor N\hat{\theta} \rfloor + 1}, \dots, X_N$  ( $i = 2$ ). Note that the eigenfunctions are only determined up to a sign. Additionally, we define the estimated squared difference of the  $j$ -th eigenvalues by

$$\hat{E}_{j,N}(\lambda) := (\hat{\tau}_{j,\lambda}^{(1)} - \hat{\tau}_{j,\lambda}^{(2)})^2, \quad \lambda \in [0, 1].$$

In view of the testing problem in (2.4) the natural entity of interest is the statistic

$$\hat{E}_{j,N} := \hat{E}_{j,N}(1) = (\hat{\tau}_j^{(1)} - \hat{\tau}_j^{(2)})^2,$$

where  $\hat{\tau}_j^{(i)}$  is defined in (2.13). The null hypothesis (2.4) of no relevant change in the  $j$ -th eigenvalue is now rejected for large values of  $\hat{E}_{j,N}$ . To find critical values for such a test we determine the asymptotic distribution of  $\hat{E}_{j,N}$ , which presupposes the following standard identifiability assumption (see e.g. Horváth and Kokoszka (2012), Hall and Hosseini-Nasab (2006)).

**Assumption 2.3.** *The first  $p+1$  eigenvalues of the covariance kernel  $c^{(i)}$  satisfy  $\tau_1^{(i)} > \dots > \tau_{p+1}^{(i)} > 0$  for  $i = 1, 2$ .*

It will be shown in the Appendix that under the Assumptions 2.1 and 2.3 the statistic  $\hat{E}_{j,N}$  is asymptotically normal in the sense that

$$\sqrt{N}\{\hat{E}_{j,N} - E_j\} \rightarrow_D \mathcal{N}(0, \sigma_E^2), \tag{2.14}$$

where the symbol  $\rightarrow_D$  denotes weak convergence,  $E_j := (\tau_j^{(1)} - \tau_j^{(2)})^2$  is the squared (unknown) difference between the  $j$ -th eigenvalues of the kernels  $c^{(1)}$  and  $c^{(2)}$ , and  $\mathcal{N}(\mu, \sigma^2)$  denotes a normal distribution with mean  $\mu$  and variance  $\sigma^2$ . In particular, if  $\sigma^2 = 0$  this distribution is defined as the point measure with probability mass 1 at the point  $\mu$ . The variance of the normal distribution in (2.14) can be decomposed as

$$\sigma_E^2 = 4|\tau_j^{(1)} - \tau_j^{(2)}|^2 [(\sigma_E^{(1)})^2 \theta_0 + (\sigma_E^{(2)})^2 (1 - \theta_0)], \quad (2.15)$$

where  $\sigma_E^{(1)}$  and  $\sigma_E^{(2)}$  are non-negative parameters depending in a very complicated way on the long run variance of the time series  $\{X_n^{(1)}\}_{n \in \mathbb{Z}}$  and  $\{X_n^{(2)}\}_{n \in \mathbb{Z}}$  and the spectrum and eigenfunctions of the covariance operators  $c^{(1)}$  and  $c^{(2)}$ . A precise definition of these quantities can be found in equations (A.16) and (A.20) in Appendix A. From these representations it can be seen that  $\sigma_E^{(1)}$  and  $\sigma_E^{(2)}$  are notoriously difficult to estimate. We circumvent this problem using self-normalization techniques. This concept has been introduced for change point detection in a seminal paper by Shao and Zhang (2010) and since then been used by many authors. While most of this literature concentrates on classical change point problems, Dette et al. (2018) introduced a novel type of self-normalization for relevant hypotheses and used it to define a self-normalized test for a relevant change in the mean of a time series. In the following we will further develop this concept to detect relevant changes in the spectrum. For this purpose we define a normalizing factor

$$\hat{V}_{j,N} := \left[ \int_0^1 \lambda^4 \{ \hat{E}_{j,N}(\lambda) - \hat{E}_{j,N}(1) \}^2 d\nu(\lambda) \right]^{1/2}, \quad (2.16)$$

where  $\nu$  is a probability measure on the interval  $(0, 1)$ . Even though the specific choice of  $\nu$  in (2.16) is generally not influential, it is numerically convenient to use a discrete measure in applications rather than some mathematically more natural choice like the Lebesgue measure.

The next Proposition is the central building block to prove the feasibility of the normalization approach.

**Proposition 2.4.** *Suppose that Assumptions 2.1 and 2.3 hold, that  $\varepsilon < \min\{\theta_0, 1 - \theta_0\}$ , that  $\sigma_E^{(i)} > 0$  for  $i = 1, 2$  and that  $\tau_j^{(1)} \neq \tau_j^{(2)}$  for some  $j \in \{1, \dots, p\}$ . Then the following weak convergence holds*

$$\sqrt{N}(\hat{E}_{j,N} - E_j, \hat{V}_{j,N}) \rightarrow_D \left( \sigma_E \mathbb{B}(1), \sigma_E \left[ \int_0^1 \lambda^2 (\mathbb{B}(\lambda) - \lambda \mathbb{B}(1))^2 \nu(d\lambda) \right]^{1/2} \right), \quad (2.17)$$

where  $\mathbb{B}$  is a standard Brownian motion.

Combining the weak convergence in (2.17) with the continuous mapping Theorem yields that,

$$\frac{\hat{E}_{j,N} - E_j}{\hat{V}_{j,N}} \rightarrow_D W \quad (2.18)$$



where the random variable

$$W := \frac{\mathbb{B}(1)}{\left\{ \int_0^1 \lambda^2 [\mathbb{B}(\lambda) - \lambda \mathbb{B}(1)]^2 \nu(d\lambda) \right\}^{1/2}} \tag{2.19}$$

is a pivot. Some quantiles of the distribution of  $W$  can be found in Table 1 (where  $\nu$  is a discrete uniform distribution).

TABLE 1  
 $(1 - \alpha)$ -quantiles of the distribution of the statistic  $W$  in (2.19), where  $\nu$  is the uniform distribution on the set  $\{l/K : l = 1, \dots, K - 1\}$ .

|          | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.1$ |
|----------|-----------------|-----------------|----------------|
| $K = 20$ | 16.479          | 9.895           | 7.097          |
| $K = 30$ | 16.248          | 9.925           | 7.149          |

We can now construct an asymptotic level- $\alpha$ -test rejecting the null hypothesis in (2.4), whenever

$$\frac{\hat{E}_{j,N} - \Delta_\tau}{\hat{V}_{j,N}} > q_{1-\alpha}, \tag{2.20}$$

where  $q_{1-\alpha}$  is the asymptotic  $(1 - \alpha)$ -quantile of the distribution of  $W$ . These considerations are summarized in the following theorem.

**Theorem 2.5.** *Suppose that Assumptions 2.1 and 2.3 hold, that  $\varepsilon < \min\{\theta_0, 1 - \theta_0\}$ , and that  $\sigma_E^{(i)} > 0$  for  $i = 1, 2$ . Then for any  $j \in \{1, \dots, p\}$  the rejection rule (2.20) defines an asymptotic level- $\alpha$  and consistent test for the hypothesis (2.4), that is*

$$\mathbb{P}\left(\frac{\hat{E}_{j,N} - \Delta_\tau}{\hat{V}_{j,N}} > q_{1-\alpha}\right) \rightarrow \begin{cases} 0, & E_j < \Delta_\tau \\ \alpha, & E_j = \Delta_\tau \\ 1, & E_j > \Delta_\tau. \end{cases} \tag{2.21}$$

It should be noted that by the same arguments as above a test can be constructed for the hypothesis of a relevant difference in the eigenvalues before and after the change point, that is

$$H_0 : E_j = (\tau_j^{(1)} - \tau_j^{(2)})^2 > \Delta_\tau \quad vs. \quad H_1 : (\tau_j^{(1)} - \tau_j^{(2)})^2 \leq \Delta_\tau. \tag{2.22}$$

The corresponding test rejects if

$$\frac{\hat{E}_{j,N} - \Delta_\tau}{\hat{V}_{j,N}} < q_\alpha \tag{2.23}$$

and the same arguments as in the proof of Theorem 2.5 show that this decision rule defines a consistent and asymptotic level- $\alpha$ -test, that is

$$\mathbb{P}\left(\frac{\hat{E}_{j,N} - \Delta_\tau}{\hat{V}_{j,N}} < q_\alpha\right) \rightarrow \begin{cases} 0, & E_j > \Delta_\tau \\ \alpha, & E_j = \Delta_\tau \\ 1, & E_j < \Delta_\tau. \end{cases}$$

The formulation of the hypothesis in the form (2.23) is useful if one wants to establish the similarity between the eigenvalues at a controlled type-I-error. Hypotheses of the form (2.22) are frequently investigated in biostatistics, in particular in bioequivalence studies (see, for example, Wellek (2010)).

### 2.3. Relevant changes in the eigenfunctions

Similar techniques as in the preceding section can be employed in the analysis of the hypothesis (2.5) of no relevant change in the  $j$ -th eigenfunction. This task is slightly more intricate, as we are now dealing with  $L^2[0, 1]$ -functions instead of real numbers.

Recall the definition of the estimated eigenfunctions in (2.13). As we have already noticed such functions are only determined up to a sign. Thus, to make comparisons meaningful, we always assume that the inner products  $\langle v_j^{(1)}, v_j^{(2)} \rangle$ ,  $\langle v_j^{(1)}, \hat{v}_{j,\lambda}^{(1)} \rangle$ ,  $\langle v_j^{(2)}, \hat{v}_{j,\lambda}^{(2)} \rangle$  and  $\langle \hat{v}_{j,\lambda}^{(1)}, \hat{v}_{j,\lambda}^{(2)} \rangle$  are non-negative for all  $\lambda \in [0, 1]$ . This assumption is solely made for the sake of a clear presentation. It can be dropped if in the testing problem (2.5) and in the subsequently presented test statistic all occurring vector distances  $\|v - v'\|$  are replaced by the  $\min(\|v - v'\|, \|v + v'\|)$ . This is also how the statistic should be understood in applications.

We estimate the squared difference  $D_j := \|v_j^{(1)} - v_j^{(2)}\|^2$  by  $\hat{D}_{j,N} := \hat{D}_{j,N}(1)$ , where the statistic  $\hat{D}_{j,N}(\lambda)$  is defined by

$$\hat{D}_{j,N}(\lambda) = \|\hat{v}_{j,\lambda}^{(1)} - \hat{v}_{j,\lambda}^{(2)}\|^2.$$

Recall that for  $\lambda \in [0, 1]$   $\hat{v}_{j,\lambda}^{(1)}$  and  $\hat{v}_{j,\lambda}^{(2)}$  are defined as the eigenfunctions of the estimated covariance operators from the samples

$$X_1, \dots, X_{\lfloor N\hat{\theta}\lambda \rfloor} \quad \text{and} \quad X_{N\hat{\theta}+1}, \dots, X_{N\hat{\theta}+\lfloor (N-N\hat{\theta})\lambda \rfloor},$$

respectively (see (2.11)). We also introduce the the normalizing factor

$$\hat{U}_{j,N} := \left( \int_0^1 \lambda^4 \left\{ \hat{D}_{j,N}(\lambda) - \hat{D}_{j,N}(1) \right\}^2 \nu(d\lambda) \right)^{1/2}. \tag{2.24}$$

We propose to reject the null hypothesis of no relevant change in the  $j$ -th eigenfunction in (2.5), whenever

$$\frac{\hat{D}_{j,N} - \Delta_v}{\hat{U}_{j,N}} > q_{1-\alpha}, \tag{2.25}$$

where  $q_{1-\alpha}$  is the upper  $\alpha$  quantile of the distribution of  $W$  defined in (2.19). The following result shows that this test has asymptotic level- $\alpha$  and is consistent. The proof can be found in Section A.3 in the appendix.

**Theorem 2.6.** *Suppose that Assumptions 2.1 and 2.3 hold, that  $\varepsilon < \min\{\theta_0, 1 - \theta_0\}$ , that  $j \in \{1, \dots, p\}$  and that the quantities  $\sigma_D^{(1)}, \sigma_D^{(2)}$  defined in (A.33) are*

positive. Then the test defined in (2.25) has asymptotic level  $\alpha$  and is consistent for the hypothesis (2.5), that is

$$\mathbb{P}\left(\frac{\hat{D}_{j,N} - \Delta_v}{\hat{U}_{j,N}} > q_{1-\alpha}\right) \rightarrow \begin{cases} 0, & D_j < \Delta_v \\ \alpha, & D_j = \Delta_v \\ 1, & D_j > \Delta_v. \end{cases} \tag{2.26}$$

We conclude this section with a brief remark, that extends our results to non-centered data. This is of particular importance in applications such as presented in Section 3.4.

**Remark 2.7.** We conclude this section by some application oriented remarks.

1. A careful inspection of the proofs in Appendix A shows that all results in this section remain true if the sequences of random variables  $(X_n^{(i)})_{n \in \mathbb{Z}}$  have non-zero expectation  $\mu^{(i)} \in L^2[0, 1]$  for  $i = 1, 2$ . In this case the estimators of the covariance kernels in (2.9) and (2.10) have to be modified as follows

$$\begin{aligned} \hat{c}^{(1)}(\lambda, s, t) &:= \frac{1}{\lfloor N\hat{\theta}\lambda \rfloor} \sum_{n=1}^{\lfloor N\hat{\theta}\lambda \rfloor} [X_n(s) - \hat{\mu}^{(1)}(s)] [X_n(t) - \hat{\mu}^{(1)}(t)], \\ \hat{c}^{(2)}(\lambda, s, t) &:= \frac{1}{\lfloor (N - N\hat{\theta})\lambda \rfloor} \sum_{n=N\hat{\theta}+1}^{N\hat{\theta}+\lfloor (N - N\hat{\theta})\lambda \rfloor} [X_n(s) - \hat{\mu}^{(2)}(s)] [X_n(t) - \hat{\mu}^{(2)}(t)], \end{aligned}$$

where  $\hat{\mu}^{(1)}$  and  $\hat{\mu}^{(2)}$  denote the empirical mean functions of the samples  $X_1, \dots, X_{\lfloor N\hat{\theta} \rfloor}$  and  $X_{\lfloor N\hat{\theta} \rfloor+1}, \dots, X_N$  respectively.

2. In practice, an important question is the appropriate choice of the threshold, which has to reflect in every particular application the specific scientific context. The classical approach avoids this choice by simply putting  $\Delta = 0$ , but we highly recommend to choose this threshold thoroughly considering the scientific background of the testing problem.

Yet sometimes theoretical arguments can be made to support a proper choice. For instance in the domain of PCA one might be interested, for two samples  $X_1, \dots, X_{\lfloor N\theta_0 \rfloor}$  and  $X_{\lfloor N\theta_0 \rfloor+1}, \dots, X_N$ , in how much variance occurs in the direction  $v_1^{(1)}$  in the second sample. For this purpose we consider the relative variance measure

$$\begin{aligned} R &:= 1 - \frac{\mathbb{E}\langle X_N, v_1^{(2)} \rangle^2 - \mathbb{E}\langle X_N, v_1^{(1)} \rangle^2}{\mathbb{E}\langle X_N, v_1^{(2)} \rangle^2} \\ &= 1 - \frac{\mathbb{E}\langle X, v_1^{(2)} + v_1^{(1)} \rangle \langle X, v_1^{(2)} - v_1^{(1)} \rangle}{\lambda_1^{(2)}}. \end{aligned}$$

If  $v_1^{(1)} = v_1^{(2)}$ , it follows that  $R = 1$  (maximal amount of variance explained) and if  $v_1^{(1)}$  runs orthogonal to the data from the second sample  $R = 0$  (no variance explained). The numerator is upper bounded by  $2\lambda_1^{(2)}\|v_1^{(2)} - v_1^{(1)}\|$ , which yields the lower bound  $R \geq 1 - 2\|v_1^{(2)} - v_1^{(1)}\|$ . One might want to chose a threshold  $\Delta_v$  such that e.g.  $R > 0.7$ , to explain at least 70% of variance along  $v_1^{(2)}$ . In this case we can use the lower bound for  $R$ , to solve for a suitable  $\Delta_v$  ( $\geq \|v_1^{(2)} - v_1^{(1)}\|^2$  under  $H_0$ ).

3. Occasionally specifying a threshold of relevance may be inherently difficult and hence the complementary approach of considering confidence intervals may be more suitable, which is also covered by our theory. More precisely, an asymptotic, one-sided,  $1 - \alpha$  confidence interval for  $D_j$  is given by

$$I_{j,N}^{(v)} := [0, \hat{D}_{j,N} + q_{1-\alpha}\hat{U}_{j,N}].$$

Similarly, a confidence interval for the distance between the  $j$ th eigenvalues of the covariance operators before and after the change point is given by

$$I_{j,N}^{(\tau)} := [0, \hat{E}_{j,N} + q_{1-\alpha}\hat{V}_{j,N}].$$

4. Sometimes it may be of interest to consider a test for relevant changes for two or more thresholds simultaneously. We want to point out that, as long as the number of thresholds is fixed, the convergence in (2.26) guarantees (asymptotic) rejection for all choices of  $\Delta_v < D_j$  simultaneously, with probability 1, as well as asymptotic level- $\alpha$  (since asymptotically the only potential rejection under the null occurs for  $\Delta_v = D_j$ , the probability of which converges to  $\alpha$ ). It is also important to point out that for finite samples the self normalized test statistic is monotonically decreasing in  $\Delta_v$ . In particular if the hypothesis defined in (2.5) is accepted for some  $\Delta_v$ , it is also accepted for all larger  $\Delta_v$ . Correspondingly rejection for a  $\Delta_v$  means rejection for all smaller thresholds. In this sense the results of evaluating the test for multiple choices of  $\Delta_v$  will be logically consistent for the user.

### 3. Finite sample properties

In this section we investigate the performance of the new tests by means of a small simulation study and illustrate potential applications in a data example. All simulations are based on 4000 simulation runs. We are interested in a test of the hypothesis of no relevant differences in the eigenvalues and eigenfunctions as defined in (2.4) and (2.5), respectively. In the subsequent results the measure  $\nu$  in the statistics  $\hat{V}_{j,N}$  and  $\hat{U}_{j,N}$  is the uniform measure on the points  $1/20, 2/20, \dots, 19/20$  (see Table 1,  $K = 20$ , for the critical values of the distribution of  $W$ ). Furthermore we assume that the change point is located at  $\lfloor N/2 \rfloor$ , that is  $\theta_0 = 1/2$ .

**3.1. Relevant changes in the eigenvalues**

We investigate the rejection probabilities of the test (2.25) for the hypothesis of no relevant change in the first and second eigenvalue. To generate data we assume that the observed functions are smoothed over the real Fourier basis of order  $T$ , which is defined for odd  $T$  as

$$\begin{aligned} & \{f_1, \dots, f_T\} \\ & = \left\{1, \sqrt{2} \sin(2\pi x), \dots, \sqrt{2} \sin(\pi(T-1)x), \sqrt{2} \cos(2\pi x), \dots, \sqrt{2} \cos(\pi(T-1)x)\right\}. \end{aligned} \tag{3.27}$$

Following Aue et al. (2009) we set  $T = 21$ , even though higher dimensions are feasible.

We define the covariance kernels in terms of the Fourier basis as follows

$$\begin{aligned} c^{(1)}(s, t) & := \sum_{k=1}^T \tau_k f_k(s) f_k(t) \quad \text{and} \\ c^{(2)}(s, t) & := (1 - \sqrt{E}) \sum_{k=1}^4 \tau_k f_k(s) f_k(t) + \sum_{k=5}^T \tau_k f_k(s) f_k(t), \end{aligned}$$

where  $\tau_k := 1/k^2$  for  $k = 1, \dots, T$  and the parameter  $E$  varies in the interval  $[0, 1]$ . Obviously the squared difference of the  $j$ -th eigenvalues of  $c^{(1)}$  and  $c^{(2)}$  is  $E_j = E/j^4$  for  $j = 1, \dots, 4$  and subsequently 0.

As we have seen in Section 2, the square  $L^2$ -distance between the kernels  $c^{(1)}$  and  $c^{(2)}$  is of importance for the performance of the change point estimator (2.6). In the present case it is given by

$$\int_0^1 \int_0^1 [c^{(1)}(s, t) - c^{(2)}(s, t)]^2 ds dt = E \sum_{k=1}^4 \tau_k^2 = 1.07875 \cdot E.$$

The simulated data is generated by randomly sampling sets of Fourier coefficients according to the above kernels. First we generate  $(N + 1)$  i.i.d. random vectors  $\epsilon_n := (\epsilon_1, \dots, \epsilon_T) \sim \mathcal{N}(0, \text{diag}(\tau_1, \dots, \tau_T))$ ,  $n = 0, \dots, N + 1$ . To introduce potential dependence, we define a matrix  $\Psi \in \mathbb{R}^{T \times T}$  with i.i.d. normally distributed entries  $\Psi_{l,k} \sim \mathcal{N}(0, \psi)$  and consider the coefficients

$$\mathbf{a}_n = \frac{\epsilon_n + \Psi \epsilon_{n-1}}{\sqrt{1 + \psi}}, \quad n = 1, \dots, N.$$

For  $n = \lfloor N\theta_0 \rfloor + 1, \dots, N$  we downscale the first four components of  $\mathbf{a}_n := (a_{n,1}, \dots, a_{n,T})^T$  by a factor  $\sqrt{1 - \sqrt{E}}$ . Finally the process  $\{X_n\}_{n=1, \dots, N}$  is defined by

$$X_n(s) = \sum_{k=1}^T a_{n,k} f_k(s). \tag{3.28}$$

An immediate calculation reveals that for  $n = 1, \dots, \lfloor N\theta_0 \rfloor$  the random variable  $X_n$  has covariance kernel  $c^{(1)}$  and for  $n = \lfloor N\theta_0 \rfloor + 1, \dots, N$  the covariance kernel

of  $X_n$  is given by  $c^{(2)}$ . The dependence of the data is determined by the choice of  $\psi$ . For  $\psi = 0$  we generate i.i.d. data and for  $\psi > 0$  an  $fMA(1)$ -process. In the later case we choose  $\psi$ , such that  $\mathbb{E}\|\Psi\|_{L_1} = 1$ . In each simulation run we use a new realization of  $\Psi$  to generate the complete sample  $X_1, \dots, X_N$ .

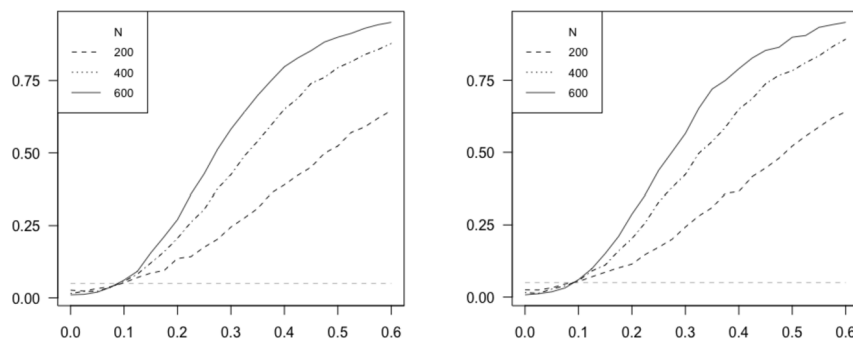


FIG 1. Rejection probabilities of the test (2.20) for the hypothesis (2.4) of no relevant change in the first eigenvalue, where  $\Delta_\tau = 0.1$ . Data comes from an i.i.d. sequence (left) and an  $fMA(1)$ -process (right). The vertical dashed line visualizes the 5% level.

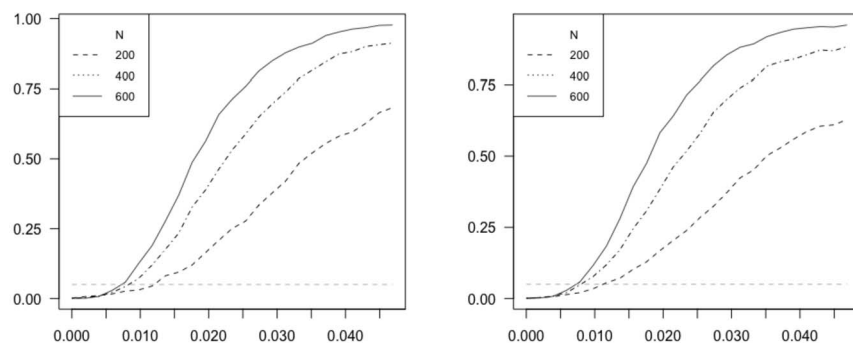


FIG 2. Rejection probabilities of the test (2.20) for the hypothesis (2.4) of no relevant change in the second eigenvalue, where  $\Delta_\tau = 0.005$ . Data comes from an i.i.d. sequence (left) and an  $fMA(1)$ -process (right). The vertical dashed line visualizes the 5% level.

In Figure 1 and 2 we display the rejection probabilities of the test (2.20) for the hypothesis (2.4) of no relevant change in the first and second eigenvalue, with level  $\alpha = 5\%$ . The threshold  $\Delta_\tau$  is given by 0.1 for  $j = 1$  and 0.005 for  $j = 2$  and the tuning parameter in the estimator (2.6) is chosen as  $\varepsilon = 0.05$ .

According to the theoretical discussion in Theorem 2.5 the test should have rejection probabilities smaller, close and larger to  $\alpha$  if  $E_j < \Delta_\tau$  (interior of the null hypothesis),  $E_j = \Delta_\tau$  (boundary of the null hypothesis) and  $E_j > \Delta_\tau$  (alternative), respectively. For the first eigenvalue (Figure 1) we observe a good approximation of the nominal level, at the boundary of the null hypothesis even

if the sample size is  $N = 200$  and a reasonable power. For the second eigenvalue (Figure 2) the test slightly conservative for  $N = 200$  at the boundary of the null hypothesis, but the level is close to  $\alpha$  for  $N = 400$  and  $N = 600$ . A comparison of the left and right part in Figures 1 and 2 shows that dependence in the data has only a small impact on both type I and type II error, even though a subtle increase is visible. Further simulations with different distributions of the Fourier coefficients show that the results are stable in this respect, although heavier tails lead to a loss of power. These results are not reported for the sake of brevity.

### 3.2. Relevant changes in the eigenfunctions

To investigate the finite sample properties of the test (2.25) for the hypothesis of no relevant change in the  $j$ -th eigenfunction in (3.27) we define the covariance kernels

$$c^{(1)}(s, t) := \sum_{k=1}^T \tau_k v_k^{(1)}(s) v_k^{(1)}(t) \quad \text{and} \quad c^{(2)}(s, t) := \sum_{k=1}^T \tau_k v_k^{(2)}(s) v_k^{(2)}(t). \quad (3.29)$$

Here the eigenvalues of  $c^{(1)}$  and  $c^{(2)}$  are the same, that is

$$\tau_k = 1/k^2 \quad \text{for } k = 1, \dots, T,$$

and the respective eigenfunctions  $v_k^{(i)}$  are defined by

$$v_k^{(1)}(s) = f_k(s) \quad (k = 1, \dots, T), \quad v_k^{(2)}(s) = f_k(s) \quad (k = 3, \dots, T),$$

and

$$v_1^{(2)}(s) = \cos(\varphi) f_1(s) + \sin(\varphi) f_2(s), \quad v_2^{(2)}(s) = \cos(\varphi) f_2(s) - \sin(\varphi) f_1(s),$$

where the Fourier basis  $\{f_1, \dots, f_T\}$  is defined in (3.27). In Figure 3 we display representative samples of functions before and after the change has occurred, for a true distance  $\|v_1^{(1)} - v_1^{(2)}\|^2 = 0.3$ . Despite this relatively large difference it is difficult to distinguish the two samples visually.

A simple calculation yields that the  $L^2$ -distance between the kernels is given by

$$\int_0^1 \int_0^1 [c^{(1)}(s, t) - c^{(2)}(s, t)]^2 ds dt = \frac{5[1 - \cos(\varphi)]}{2}$$

and that the distance between the first and second eigenfunctions is

$$\|v_1^{(1)} - v_1^{(2)}\| = \|v_2^{(1)} - v_2^{(2)}\| = \sqrt{2 - 2 \cos(\varphi)}.$$

By construction the two kernels  $c^{(1)}$  and  $c^{(2)}$  in this example are very similar, and therefore the estimation of the change point is a challenging task. Any further difference in the eigensystems would increase the  $L^2$ -distance of the kernels and thus facilitate the problem of change point detection.

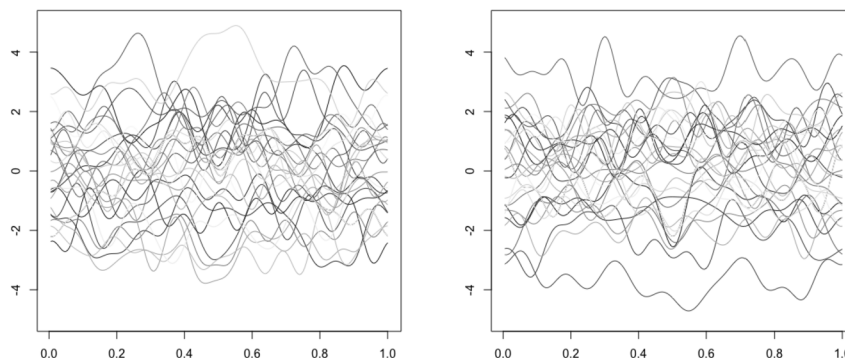


FIG 3. Representative samples of 30 data functions before (left) and after (right) the change. The squared distance between the first eigenvectors of each sample is  $\|v_1^{(1)} - v_1^{(2)}\|^2 = 0.3$ .

The data is generated as described in Section 3.1, where the coefficients  $a_{n,k}$  in the representation (3.28) are given by the vectors

$$\mathbf{a}_n = (a_{n,1}, \dots, a_{n,T}) = \frac{\epsilon_n + \Psi\epsilon_{n-1}}{\sqrt{1 + \psi}}, \quad n = 1, \dots, \lfloor N\theta_0 \rfloor,$$

$$\mathbf{a}_n = (a_{n,1}, \dots, a_{n,T}) = R_{1,2}(\varphi) \frac{\epsilon_n + \Psi\epsilon_{n-1}}{\sqrt{1 + \psi}}, \quad n = \lfloor N\theta_0 \rfloor + 1, \dots, N.$$

Here  $R_{1,2}(\varphi)$  is the rotation matrix of the first two components by an angle  $\varphi$ . Note that for  $n = 1, \dots, \lfloor N\theta_0 \rfloor$  the random variable  $X_n$  in (3.28) has covariance kernel  $c^{(1)}$  and for  $n = \lfloor N\theta_0 \rfloor + 1, \dots, N$  the covariance kernel of  $X_n$  is given by  $c^{(2)}$ .

In Figure 3 and 4 we display the rejection probabilities of the test (2.25) for the hypothesis (2.5) of no relevant change in the first and second eigenfunction (threshold  $\Delta_v = 0.1$ ), with level  $\alpha = 5\%$ . The tuning parameter in the change point estimator (2.6) is given by  $\varepsilon = 0.05$ .

We observe a good approximation of the nominal level at boundary of the null hypothesis ( $D_j = \Delta_v$ ) and the test detects alternatives with decent power. Similar to the investigation of the eigenvalues, additional dependence has little impact on the results. An interesting difference occurs in the consideration of the second eigenfunction. Whereas for the second eigenvalue the self-normalized test (2.20) is slightly more conservative for small sample sizes, we observe that for the second eigenfunction the test (2.25) becomes slightly more liberal in this case.

### 3.3. Details on the change point estimator in (2.6)

In this Section we discuss aspects of the change point estimator defined in (2.6) in greater depth. First we consider what impact the choice of the boundary parameter  $\varepsilon$  has on the performance of the tests. In practice the choice  $\varepsilon = 0.05$  indicates only moderate knowledge, but one might want to use even



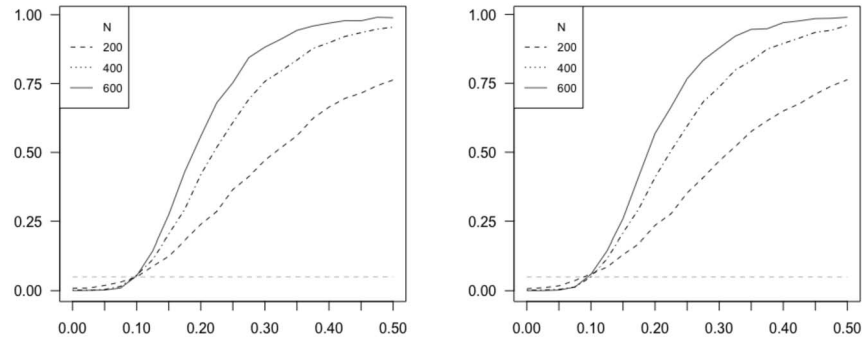


FIG 4. Rejection probabilities of the test (2.25) for the hypothesis (2.5) of no relevant change in the first eigenfunction. Data comes from an i.i.d. sequence (left) and an fMA(1)-process (right). The vertical dashed line visualizes the 5% level.

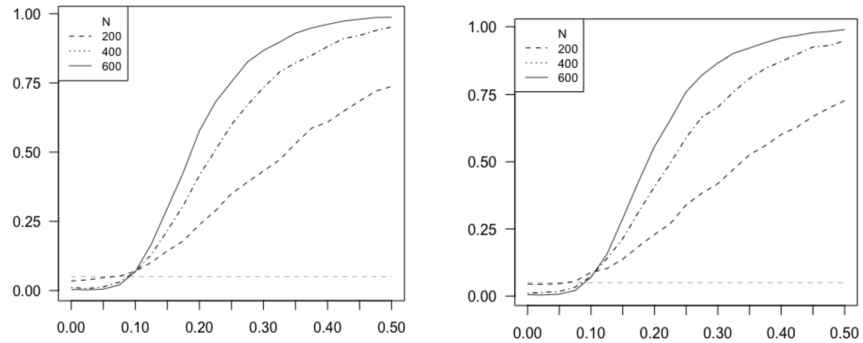
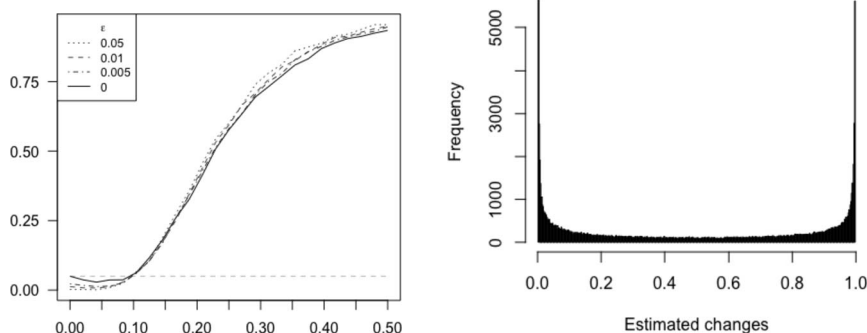


FIG 5. Rejection probabilities of the test (2.25) for the hypothesis (2.5) of no relevant change in the second eigenfunction. Data comes from an i.i.d. sequence (left) and an fMA(1)-process (right). The vertical dashed line visualizes the 5% level.

smaller values, or more rigorously put  $\varepsilon = 0$ . We therefore consider the model of the preceding section, for the first eigenfunction with an i.i.d. sample, of size  $N = 400$  and investigate the impact of different choices  $\varepsilon = 0, 0.005, 0.01, 0.05$  on the performance of the test (2.25) for the hypothesis (2.5) of no relevant difference in the first eigenfunction. The results are displayed in Figure 6a and indicate a high stability with respect to the choice of  $\varepsilon$ . Whereas the power of the test is hardly influenced by the choice of  $\varepsilon$ , we observe that the choice  $\varepsilon = 0$  produces comparatively high rejection probabilities under the null hypothesis, particularly for small values of  $D_j$ . This effect can be explained as follows. Under the alternative the two samples  $X_1, \dots, X_{\lfloor \theta_0 N \rfloor}$  and  $X_{\lfloor \theta_0 N \rfloor + 1}, \dots, X_N$  have quite distinct covariance structures and so the change point estimator will perform well, regardless of the choice of  $\varepsilon$ . However, if  $D_j$  is close to 0, the samples will be nearly indistinguishable, such that the change point  $\theta_0$  is estimated less accurately. In such cases the test has larger rejection probabilities.



(a) Simulated rejection probabilities of the test (2.25) for the hypothesis (2.5) of no relevant change in the first eigenfunction ( $\Delta_v = 0.1$ ). The sample size is  $N = 400$  and different choices of the tuning parameter  $\varepsilon$  in the estimator (2.6) are considered.

(b) Histogram of 100.000 realizations of the change point estimator (2.6) with tuning parameter  $\varepsilon = 0$ . The sample size is  $N = 400$ .

FIG 6.

This interpretation is visualized by Figure 6b where we display a histogram of the estimated change points using the estimator (2.6) with  $\varepsilon = 0$  and true distance of the eigenfunctions  $D_j = 0$ . We observe that in this case the estimator frequently localizes the change point in the first or last bin. For the problem of testing for relevant differences in the eigenvalues as considered in Section 3.1, similar effects can be observed, which are not reported to avoid redundancy.

We conclude this Section by pointing out some advantages and weaknesses of the change point estimator for finite samples, as well as potential alternatives.

**Remark 3.1.** The change point estimator  $\hat{\theta}$  defined in (2.6) and considered throughout this paper detects the change  $\theta_0$ , by searching for structural breaks in the covariance operators of the samples. The benefit of this approach is, that the estimate is based not only on information from one, but from all principal components simultaneously. However a potential drawback of this approach might be that subtle changes in later elements of the eigensystems could be difficult to detect, because they are weighted less. Under these circumstances one might consider solutions tailored to the specific testing problem, such as

$$\hat{\theta}_j^{(\tau)} := \operatorname{argmax}_{N\varepsilon \leq k \leq N(1-\varepsilon)} \frac{k(N-k)}{N^2} (\hat{\tau}_{k,j}^{(1)} - \hat{\tau}_{k,j}^{(2)})^2 \quad (3.30)$$

$$\hat{\theta}_j^{(v)} := \operatorname{argmax}_{N\varepsilon \leq k \leq N(1-\varepsilon)} \frac{k(N-k)}{N^2} \|\hat{v}_{k,j}^{(1)} - \hat{v}_{k,j}^{(2)}\|^2, \quad (3.31)$$

where  $\hat{\tau}_{k,j}^{(1)}, \hat{\tau}_{k,j}^{(2)}$  and  $\hat{v}_{k,j}^{(1)}, \hat{v}_{k,j}^{(2)}$  are the  $j$ th eigenvalues and eigenfunctions of the estimated kernels

$$\hat{c}^{(1)}(s, t, k) := \frac{1}{k} \sum_{n=1}^k X_n(s)X_n(t), \quad \hat{c}^{(2)}(s, t, k) := \frac{1}{N-k} \sum_{n=k+1}^N X_n(s)X_n(t).$$

respectively. These solutions promise to focus more specifically on changes in the component under investigation.

While estimates of the form (3.30) and (3.31) might have their place in practice, it is important to note that in general these change point estimators are not consistently estimating the location  $\theta_0$  of the change point, even if a change in the  $j$ th component occurs. More precisely, consider an example of two dimensional data, from two samples of i.i.d. normally distributed random variables

$$\begin{aligned} X_1, \dots, X_{\lfloor N\theta_0 \rfloor} &\sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right) \\ X_{\lfloor N\theta_0 \rfloor + 1}, \dots, X_N &\sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right). \end{aligned}$$

The first eigenvector (discrete eigenfunction) of the first sample is  $(1, 0)^\top$  and the first one of the second sample  $(0, 1)^\top$ . These eigenvectors have maximal distance, and more specifically the two data sets live on perpendicular subspaces of  $\mathbb{R}^2$ . A change could not be easier to detect. However it can be shown that for large  $N$

$$\|\hat{v}_{k,1}^{(1)} - \hat{v}_{k,1}^{(2)}\|^2 \approx \begin{cases} 1, & 2\lfloor N\theta_0 \rfloor > k > 2\lfloor N\theta_0 \rfloor - N \\ 0, & \text{else,} \end{cases}$$

where  $\hat{v}_{k,1}^{(1)}$  and  $\hat{v}_{k,1}^{(2)}$  are the estimated, first eigenvectors from the samples  $X_1, \dots, X_k$  and  $X_{k+1}, \dots, X_n$ . In particular the estimator  $\hat{\theta}^{(v)}$  defined in (3.31) converges to  $1/2$  for any true  $\theta_0 \in (1/4, 3/4)$ , due to the weighting factor  $k(N-k)/N^2$ . Similar examples of complete failure exist for the eigenvalues. It hence seems advisable to employ component specific change point estimators with some caution.

### 3.4. A data example

In this subsection we apply the new methodology to identify relevant changes in temperature measurements in northern Germany. The data consists of daily temperature averages, published by the national meteorological agency ‘‘Der Deutsche Wetterdienst’’ ([https://www.dwd.de/DE/Home/home\\_node.html](https://www.dwd.de/DE/Home/home_node.html)) in the state of Bremen (Lat=53.1°, Long= 008.7°) over the years 1890 to 2018. Due to incomplete data in the first years, as well as immediately after WWII, the years 1890–1893 as well as 1945–1946 have been removed from the data. This leaves us with 123 years of daily measurements.

We smooth this data over the Fourier basis defined in (3.27), for  $T = 41$ , a choice between Fremdt et al. (2014) ( $T=49$ ) and Aue et al. (2009) ( $T=21$ ), which allows a reasonable approximation of the daily temperatures and still reflects general trends. However it is worth noticing that our subsequent findings are quite robust with respect to different choices of  $T$ .

For the tuning in the change point estimator (2.6) we choose  $\varepsilon = 0.01$ , which yields  $N\hat{\theta} = 92$  as an estimated change point. This corresponds to the year 1988

and we can calculate the respective eigenfunctions and eigenvalues, before and after the change. Exemplarily we show in Figure 7 the first three eigenfunctions before and after the estimated change point. We observe that the first eigenfunctions are quite similar, but larger differences can be found between the second and third eigenfunctions (see lower panel in Figure 7).

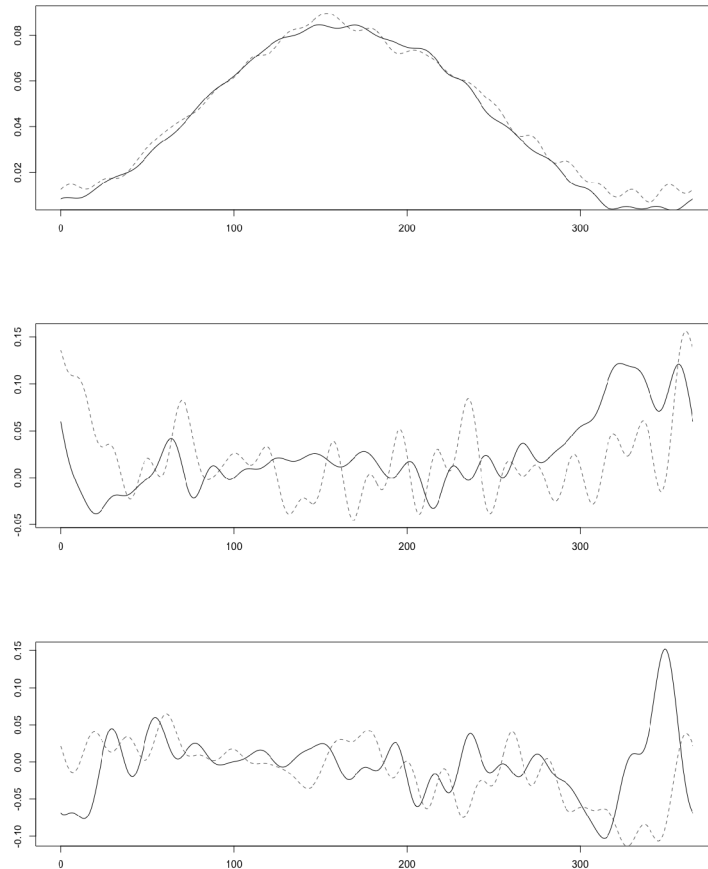


FIG 7. Eigenfunctions of the estimated covariance operators from 1894–1988 (solid) and 1989–2018 (dotted). Upper panel: First eigenfunctions  $\hat{v}_1^{(1)}$  and  $\hat{v}_1^{(2)}$ . The curves differ only slightly. Is this relevant? Middle panel: Second eigenfunctions  $\hat{v}_2^{(1)}$  and  $\hat{v}_2^{(2)}$ . The functions are more dissimilar, revealing very limited common trends. Lower panel: Third eigenfunctions  $\hat{v}_3^{(1)}$  and  $\hat{v}_3^{(2)}$ . These functions are rather different, close to orthogonality.

For the first five eigenfunctions we have applied the test (2.25) for the hypothesis (2.5) of no relevant change to see whether we can find relevant differences for varying sizes of  $\Delta_v$ . To make our results more interpretable we translate the measure of similarity  $\Delta_v$  into an angle  $\varphi$ , i.e., if the squared distance of two eigenfunctions differs by at least  $\Delta_v \approx 0.58$  the “geometric angle” between them is at least  $\varphi = \pi/4$ . This is again due to the fact that the angle  $\varphi$  between

two eigenfunctions  $v, w$  determines their distance

$$\|v - w\| = \sqrt{2 - 2 \cos(\varphi)}.$$

In Table 2 we display the decisions of the test (2.25) for the hypothesis (2.5) of no relevant changes in the eigenfunctions where different thresholds are considered. We observe that the test does not detect relevant changes between the first and the second eigenfunctions. In contrast, the eigenfunctions of larger order display significant differences, which confirms the visualization in Figure 7. The test detects relevant changes in the third, fourth and fifth eigenfunctions for nearly all thresholds (and the same holds true for eigenfunctions of larger order).

TABLE 2  
*Relevance of the differences of the first five eigenfunctions for different, relevant angles  $\varphi$ . Acceptance of the null hypothesis in (2.5) with  $\Delta = \sqrt{2 - 2 \cos(\varphi)}$  is denoted by “TRUE” ( $p$ -value > 10%), rejection by “FALSE”. For rejections, superindices indicate the probability of a less extreme event under the null.*

| Eigenfunctions     | $i = 1$ | $i = 2$ | $i = 3$                  | $i = 4$                  | $i = 5$                  |
|--------------------|---------|---------|--------------------------|--------------------------|--------------------------|
| $\varphi = \pi/16$ | TRUE    | TRUE    | FALSE <sup>&gt;99%</sup> | FALSE <sup>&gt;99%</sup> | FALSE <sup>&gt;99%</sup> |
| $\varphi = \pi/8$  | TRUE    | TRUE    | FALSE <sup>&gt;99%</sup> | FALSE <sup>&gt;95%</sup> | FALSE <sup>&gt;99%</sup> |
| $\varphi = \pi/4$  | TRUE    | TRUE    | FALSE <sup>&gt;99%</sup> | FALSE <sup>&gt;95%</sup> | FALSE <sup>&gt;99%</sup> |
| $\varphi = 2\pi/5$ | TRUE    | TRUE    | FALSE <sup>&gt;95%</sup> | TRUE                     | FALSE <sup>&gt;90%</sup> |

To fully appreciate these results we have to take the estimated eigenvalues into account. The first five estimates before and after the change point are given by

$$\begin{aligned} \hat{\tau}_1^{(1)} &= 38.38, & \hat{\tau}_2^{(1)} &= 0.56, & \hat{\tau}_3^{(1)} &= 0.26, & \hat{\tau}_4^{(1)} &= 0.24, & \hat{\tau}_5^{(1)} &= 0.21 \\ \hat{\tau}_1^{(2)} &= 41.15, & \hat{\tau}_2^{(2)} &= 0.45, & \hat{\tau}_3^{(2)} &= 0.40, & \hat{\tau}_4^{(2)} &= 0.34, & \hat{\tau}_5^{(2)} &= 0.29. \end{aligned}$$

The rapid decay of the eigenvalues indicates, that most of the data’s variance – in fact about 90% for each sample – is explained by the first principal components. Due to the similarity of the first eigenfunctions a low-dimensional representation of all data may be given, by projecting on a common, low-dimensional function space, a process much facilitating subsequent analysis. In contrast, due to great dissimilarities of the higher order eigenfunctions, finding a common space that captures, say 95% of all variance will require much higher dimensions.

Beyond such fPCA-related considerations, the eigenvalues encode further, valuable information about the data. They indicate how strongly each eigenfunction contributes to the measured functions. Displaying the eigenvalues 2–12 (those still of larger order than 0.01) in the below graphic reveals a striking trend: The estimated eigenvalues of the time period from 1894–1988 are decaying faster than those of 1989–2018.

It should be noted that this trend persists if we base our estimates on a change point earlier in time, even if we split the data in equally sized halves. Of

course contamination of the second data set then leads to smaller differences, but the underlying trend of slower decay of the eigenvalues from the earlier time period remains visible. This indicates that the observed effects are not simply due to a suggestive change point selection.

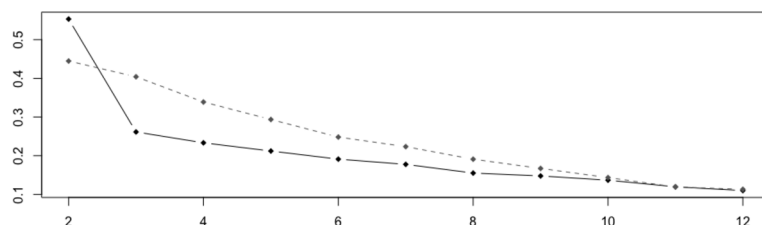


FIG 8. Eigenvalues of order 2–12 of the estimated covariance operators from 1884–1988 (solid) and 1989–2018 (dotted).

To establish the relevance of these differences, we consider each eigenvalue with an individual threshold of relevance, suited to its magnitude (proportional to the size of  $\hat{\tau}_j^{(1)}$ ) and apply the test (2.20).

TABLE 3

Differences of the eigenvalues from the samples 1894–1988 and 1989–2018 for varying sizes of  $\Delta_\tau$ . Relevant differences with probability  $\geq 95\%$  are marked in gray and  $\geq 99\%$  in black. White cells suggest no rejection ( $p$ -value  $> 10\%$ ).

| $\Delta_\tau \backslash j$ | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    |
|----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $\hat{\tau}_j^{(1)}/50$    | Black | White | Gray  | Black | Black | White | White | White | White | White | White | White |
| $\hat{\tau}_j^{(1)}/100$   | Black | White | Gray  | Black | Black | Black | Black | Black | White | White | White | White |
| $\hat{\tau}_j^{(1)}/200$   | Black | White | Black | Black | Black | Black | Black | Black | Black | White | White | White |

The visual inspection of the eigenvalues is consistent with the testing results. The eigenvalues of the covariance operators differ up to order 9, with decreasing relevance.

One practical interpretation of these differences may be as follows: The faster decay of the eigenvalues of  $\hat{c}^{(1)}$  compared to those of  $\hat{c}^{(2)}$ , indicates that the observations from 1894–1988 are less influenced by higher order eigenfunctions, than those from 1989–2018. Given that the eigenfunctions become more irregular with increasing order (compare Figure 7), this might imply rougher data, i.e. more short term variability temperatures recently, than in the first part of the 20th century.

### Acknowledgements

This work has been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823, Teilprojekt A1, C1) of the German Research Foundation (DFG). The authors would like to

thank the referees for their constructive comments on an earlier version of this manuscript.

**Appendix A: Proofs**

For clarity of presentation we confine ourselves to the case  $j = 1$ , i.e. differences in the first eigenfunction and eigenvalue. The general case follows by exactly the same arguments.

An important feature of the proofs is the replacement of the estimated change point  $\hat{\theta}$ , by the deterministic, true change point  $\theta_0$ . If we knew the true change point, we could construct the ideal, estimated covariance kernels by  $\tilde{c}^{(1)}(1, s, t)$  and  $\tilde{c}^{(2)}(1, s, t)$  where for  $\lambda \in [0, 1]$

$$\begin{aligned} \tilde{c}^{(1)}(\lambda, s, t) &= \frac{1}{\lfloor \lfloor N\theta_0 \rfloor \lambda \rfloor} \sum_{n=1}^{\lfloor \lfloor N\theta_0 \rfloor \lambda \rfloor} X_n(s)X_n(t), & (A.1) \\ \tilde{c}^{(2)}(\lambda, s, t) &= \frac{1}{\lfloor (N - \lfloor N\theta_0 \rfloor)\lambda \rfloor} \sum_{n=N\theta_0+1}^{\lfloor N\theta_0 \rfloor + \lfloor (N - \lfloor N\theta_0 \rfloor)\lambda \rfloor} X_n(s)X_n(t). \end{aligned}$$

These kernels, as well as the corresponding eigensystems

$$\tilde{v}_{1,\lambda}^{(i)}, \tilde{v}_{2,\lambda}^{(i)}, \dots \quad \text{and} \quad \tilde{\tau}_{1,\lambda}^{(i)} \geq \tilde{\tau}_{2,\lambda}^{(i)} \geq \dots$$

for  $i = 1, 2$ , will be frequently referred to in the following section.

**A.1. Proof of Proposition 2.4**

Recall the definition of the eigenvalues  $\hat{\tau}_{1,\lambda}^{(i)}$  for  $i = 1, 2$  of the estimated kernels  $\hat{c}^{(i)}(\lambda, \cdot, \cdot)$  in (2.9) and (2.10). In order to show the proposition, we establish the weak convergence

$$\begin{aligned} \{H_N(\lambda)\}_{\lambda \in [0,1]} &:= \left\{ \sqrt{N}\lambda^2 [(\hat{\tau}_{1,\lambda}^{(1)} - \hat{\tau}_{1,\lambda}^{(2)})^2 - (\tau_1^{(1)} - \tau_1^{(2)})^2] \right\}_{\lambda \in [0,1]} & (A.2) \\ &\rightarrow_D \{ \lambda \sigma_E \mathbb{B}(\lambda) \}_{\lambda \in [0,1]}, \end{aligned}$$

where  $\mathbb{B}$  is a standard Brownian motion. The statement then follows by an application of the continuous mapping Theorem, as

$$\sqrt{N}(\hat{E}_{1,N} - E_1, \hat{V}_{1,N}) = \left( H_N(1), \left[ \int_0^1 \{H_N(\lambda) - \lambda^2 H_N(1)\}^2 d\nu(\lambda) \right]^{1/2} \right).$$

The proof of (A.2) consists of two steps.

**Step 1:** First we demonstrate that using the change point estimate  $\hat{\theta}$  is asymptotically as good as knowing the true location  $\theta_0$  of the change point, or more precisely

$$\sqrt{N}\lambda^2 [(\hat{\tau}_{1,\lambda}^{(1)} - \hat{\tau}_{1,\lambda}^{(2)})^2 - (\hat{\tau}_{1,\lambda}^{(1)} - \hat{\tau}_{1,\lambda}^{(2)})^2] = o_P(1) \tag{A.3}$$

uniformly with respect to  $\lambda \in [0, 1]$ . To establish this equality we show

$$\lambda|\tilde{\tau}_{1,\lambda}^{(1)} - \hat{\tau}_{1,\lambda}^{(1)}| = o_P(1/\sqrt{N}) \tag{A.4}$$

uniformly in  $\lambda \in [0, 1]$ . Deducing (A.3) from (A.4) is then a simple calculation. To obtain an upper bound on the left side of (A.4), we employ Lemma 2.2 from Horváth and Kokoszka (2012), which yields

$$\lambda|\tilde{\tau}_{1,\lambda}^{(1)} - \hat{\tau}_{1,\lambda}^{(1)}| \leq \lambda\|\tilde{c}^{(1)}(\lambda, \cdot, \cdot) - \hat{c}^{(1)}(\lambda, \cdot, \cdot)\|. \tag{A.5}$$

For the difference of the kernels we obtain

$$\begin{aligned} & \lambda[\tilde{c}^{(1)}(\lambda, s, t) - \hat{c}^{(1)}(\lambda, s, t)] \tag{A.6} \\ &= \frac{1}{\lfloor N\theta_0 \rfloor} \sum_{n=1}^{\lfloor \lfloor N\theta_0 \rfloor \lambda \rfloor} X_n(t)X_n(s) - \frac{1}{N\hat{\theta}} \sum_{n=1}^{\lfloor N\hat{\theta}\lambda \rfloor} X_n(t)X_n(s) + \mathcal{O}_P(N^{-1}) \\ &= \frac{\text{sign}(\theta_0 - \hat{\theta})}{\lfloor N(\theta_0 \vee \hat{\theta}) \rfloor} \sum_{n=\lfloor \lfloor \hat{\theta} \wedge \theta_0 \rfloor N \rfloor \lambda}^{\lfloor \lfloor \hat{\theta} \vee \theta_0 \rfloor N \rfloor \lambda} X_n(t)X_n(s) \\ &+ \left( \frac{1}{\lfloor N\theta_0 \rfloor} - \frac{1}{N\hat{\theta}} \right) \sum_{n=1}^{\lfloor \lfloor \hat{\theta} \wedge \theta_0 \rfloor N \rfloor \lambda} X_n(t)X_n(s) + \mathcal{O}_P(N^{-1}), \end{aligned}$$

where the second equality follows by a straightforward rearrangement of the terms. Notice that  $|1/\hat{\theta} - N/\lfloor N\theta_0 \rfloor| = o_P(1/\sqrt{N})$ , which follows immediately by Proposition 2.2. An application of the triangle inequality shows that the  $L^2([0, 1] \times [0, 1])$ -norm of the second term on right side of (A.6) can be bounded by

$$\left| \frac{N}{\lfloor N\theta_0 \rfloor} - \frac{1}{\hat{\theta}} \right| \frac{1}{N} \sum_{n=1}^N \|X_n \otimes X_n\| = o_P(1/\sqrt{N}),$$

where we have applied Birkhoff’s Theorem. The  $L^2([0, 1] \times [0, 1])$ -norm of the first term on the right of (A.6) is also of order  $o_P(1/\sqrt{N})$ . Counting the summands we see that centering only yields a further term of order  $o_P(1/\sqrt{N})$ . We now use Theorem B.1 from Appendix B to get

$$\begin{aligned} & \sup_{\lambda} \left\{ \int_0^1 \int_0^1 \left[ \frac{\text{sign}(\theta_0 - \hat{\theta})}{\lfloor N(\theta_0 \wedge \hat{\theta}) \rfloor} \sum_{n=\lfloor \lfloor \hat{\theta} \wedge \theta_0 \rfloor N \rfloor \lambda}^{\lfloor \lfloor \hat{\theta} \vee \theta_0 \rfloor N \rfloor \lambda} [X_n(t)X_n(s) - \mathbb{E}[X_n(t)X_n(s)]] \right]^2 dsdt \right\}^{1/2} \tag{A.7} \\ &= \mathcal{O}_P \left( \frac{1}{\sqrt{N}} \sup_{\lambda} \left\{ \int_0^1 \int_0^1 \left[ \Gamma_N(\lambda[\hat{\theta} \vee \theta_0]N)/N, s, t) - \Gamma_N(\lambda[\hat{\theta} \wedge \theta_0]N)/N, s, t) \right]^2 dsdt \right\}^{1/2} \right), \end{aligned}$$



where  $\Gamma_N$  is a Gaussian process defined by

$$\Gamma_N(x, s, t) := \sum_{l \geq 1} \Lambda_l \mathbb{B}_{l,N}(x) \Psi_l(s, t),$$

$\{\mathbb{B}_{l,N} : l \in \mathbb{N}\}$  are independent Brownian motions,  $\{\Lambda_l : l \in \mathbb{N}\}$  are non-negative, square summable values and  $\{\Psi_l : l \in \mathbb{N}\}$  is an orthonormal basis of  $L^2([0, 1] \times [0, 1])$ . Enlarging the set over which we maximize, we see that the supremum in (A.7) is bounded by

$$\sup_{|\lambda - \mu| \leq |\hat{\theta} - \theta_0|} \left( \int_0^1 \int_0^1 [\Gamma_N(\lambda, s, t) - \Gamma_N(\mu, s, t)]^2 ds dt \right)^{1/2},$$

and the definition of  $\Gamma_N$  yields

$$\begin{aligned} & \sup_{|\lambda - \mu| \leq |\hat{\theta} - \theta_0|} \left( \int_0^1 \int_0^1 [\Gamma_N(\lambda, s, t) - \Gamma_N(\mu, s, t)]^2 ds dt \right)^{1/2} \\ &= \sup_{|\lambda - \mu| \leq |\hat{\theta} - \theta_0|} \left( \sum_{l \geq 1} \Lambda_l^2 [\mathbb{B}_{l,N}(\lambda) - \mathbb{B}_{l,N}(\mu)]^2 \right)^{1/2}. \end{aligned}$$

Its expectation is bounded by

$$\left( \mathbb{E} \left[ \sup_{|\lambda - \mu| \leq |\hat{\theta} - \theta_0|} (\mathbb{B}(\lambda) - \mathbb{B}(\mu))^2 \right] \sum_{l \geq 1} \Lambda_l^2 \right)^{1/2},$$

where  $\mathbb{B}$  is a standard Brownian motion. The expectation converges to 0 by application of the dominated convergence Theorem together with the almost sure continuity of the paths of the Brownian motion (see Billingsley, Section 37). These considerations yield (A.4) and hence (A.3).

**Step 2:** We now prove the weak convergence (A.2). Step 1 and straightforward calculations yield

$$\begin{aligned} & \sqrt{N} \lambda^2 [(\tilde{\tau}_{1,\lambda}^{(1)} - \tilde{\tau}_{1,\lambda}^{(2)})^2 - (\tau_1^{(1)} + \tau_1^{(2)})^2] \\ &= \sqrt{N} \lambda^2 [\tilde{\tau}_{1,\lambda}^{(1)} - \tilde{\tau}_{1,\lambda}^{(2)} - \tau_1^{(1)} - \tau_1^{(2)}]^2 \\ & \quad + 2\sqrt{N} \lambda^2 (\tau_1^{(1)} - \tau_1^{(2)}) (\tilde{\tau}_{1,\lambda}^{(1)} - \tilde{\tau}_{1,\lambda}^{(2)} - \tau_1^{(1)} + \tau_1^{(2)}) \\ &= 2\sqrt{N} \lambda^2 (\tau_1^{(1)} - \tau_1^{(2)}) (\tilde{\tau}_{1,\lambda}^{(1)} - \tilde{\tau}_{1,\lambda}^{(2)} - \tau_1^{(1)} + \tau_1^{(2)}) + o_P(1). \end{aligned}$$

For further analysis of the quantity  $G_N(\lambda) = \sqrt{N} \lambda^2 (\tilde{\tau}_{1,\lambda}^{(1)} - \tilde{\tau}_{1,\lambda}^{(2)} - \tau_1^{(1)} + \tau_1^{(2)})$  we use Proposition B.2, in Appendix B, which gives

$$G_N(\lambda) = \sqrt{N} \lambda^2 (\tilde{\tau}_{1,\lambda}^{(1)} - \tilde{\tau}_{1,\lambda}^{(2)} - \tau_1^{(1)} + \tau_1^{(2)}) = R_N^{(1)}(\lambda) + R_N^{(2)}(\lambda) + o_P(1),$$

where

$$R_N^{(1)}(\lambda) = \frac{\sqrt{N} \lambda}{\lfloor N \theta_0 \rfloor} \sum_{n=1}^{\lfloor \lfloor N \theta_0 \rfloor \lambda \rfloor} \int_0^1 \int_0^1 (X_n^{(1)}(s) X_n^{(1)}(t) - c^{(1)}(s, t)) v_1^{(1)}(s) v_1^{(1)}(t) ds dt$$

$$R_N^{(2)}(\lambda) = \frac{\sqrt{N}\lambda}{[N(1-\theta_0)]} \sum_{n=[N\theta_0]+1}^{[[N(1-\theta_0)]\lambda]+[N\theta_0]} \int_0^1 \int_0^1 (X_n^{(2)}(s)X_n^{(2)}(t) - c^{(2)}(s,t))v_1^{(2)}(s)v_1^{(2)}(t)dsdt.$$

Weak convergence of the process  $\{G_N(\lambda)\}_{\lambda \in [0,1]}$  now follows by an application of the continuous mapping Theorem and the weak convergence of the vector valued process

$$(R_N^{(1)}(\lambda), R_N^{(2)}(\lambda))_{\lambda \in [0,1]}. \tag{A.8}$$

For a proof of this statement we show asymptotic tightness and convergence of the finite dimensional distributions. We therefore introduce the random variables

$$Y_n^{(i)} := \int_0^1 \int_0^1 [X_n^{(i)}(s)X_n^{(i)}(t) - c^{(i)}(s,t)]v_1^{(i)}(s)v_1^{(i)}(t)dsdt \quad i = 1, 2 \tag{A.9}$$

and

$$Y_{n,m}^{(i)} := \int_0^1 \int_0^1 [X_{n,m}^{(i)}(s)X_{n,m}^{(i)}(t) - c^{(i)}(s,t)]v_1^{(i)}(s)v_1^{(i)}(t)dsdt \quad i = 1, 2, \tag{A.10}$$

where the random functions  $X_{n,m}^{(i)}$  are defined in (2.7). Asymptotic tightness can be shown coordinate-wise, such that we verify it exemplarily for the first component. This can be rewritten as the process

$$(R_N^{(1)}(\lambda))_{\lambda \in [0,1]} = \left( \frac{\sqrt{N}\lambda}{[N\theta_0]} \sum_{n=1}^{[[N\theta_0]\lambda]} Y_n^{(1)} \right)_{\lambda \in [0,1]}. \tag{A.11}$$

As tightness of a stochastic process  $(G(\lambda))_{\lambda \in [0,1]}$  in  $\ell^\infty[0,1]$  implies tightness of  $(\lambda G(\lambda))_{\lambda \in [0,1]}$ , it will suffice to show tightness of  $(R_N^{(1)}(\lambda)/\lambda)_{\lambda \in [0,1]}$ . To prove this assertion we note

$$\begin{aligned} & \lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} \mathbb{P} \left( \sup_{|r-q| < \delta} \left| \frac{\sqrt{N}}{[N\theta_0]} \left( \sum_{n=1}^{[[N\theta_0]q]} Y_n^{(1)} - \sum_{n=1}^{[[N\theta_0]r]} Y_n^{(1)} \right) \right| > \varepsilon \right) \tag{A.12} \\ & \leq \lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} \mathbb{P} \left( \sup_{|r-q| < \delta} \left| \frac{\sqrt{N}}{[N\theta_0]} \left( \sum_{n=1}^{[[N\theta_0]q]} Y_{n,m}^{(1)} - \sum_{n=1}^{[[N\theta_0]r]} Y_{n,m}^{(1)} \right) \right| > \varepsilon/2 \right) \\ & + 2 \lim_{N \rightarrow \infty} \mathbb{P} \left( \sup_{q \in [0,1]} \left| \frac{\sqrt{N}}{[N\theta_0]} \sum_{n=1}^{[[N\theta_0]q]} Y_n^{(1)} - Y_{n,m}^{(1)} \right| > \varepsilon/4 \right). \end{aligned}$$

Adapting the proof of Lemma 2.1 in Berkes et al. (2013) shows that under the assumptions stated in Section 2

$$\lim_{m \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{P} \left( \max_{k=1, \dots, N} \left\| \frac{1}{\sqrt{N}} \sum_{n=1}^k (X_n^{(i)} \otimes X_n^{(i)} - X_{n,m}^{(i)} \otimes X_{n,m}^{(i)}) \right\| > x \right) = 0 \tag{A.13}$$

for any  $x > 0$ . By this result and the Cauchy-Schwarz inequality, the second term on the right side of (A.12) converges to 0 as  $m \rightarrow \infty$ .

For the first term we define for  $h = 1, \dots, m$

$$Q(h, T) := \{n \leq T : n \bmod m = h\}$$

and split the sample  $Y_{1,m}^{(1)}, \dots, Y_{\lfloor N\theta_0 \rfloor, m}^{(1)}$  in  $m$  groups of independent identically distributed random variables, which gives

$$\begin{aligned} & \mathbb{P}\left(\sup_{|r-q|<\delta} \frac{\sqrt{N}}{\lfloor N\theta_0 \rfloor} \left| \sum_{n=1}^{\lfloor \lfloor N\theta_0 \rfloor q \rfloor} Y_{n,m}^{(1)} - \sum_{n=1}^{\lfloor \lfloor N\theta_0 \rfloor r \rfloor} Y_{n,m}^{(1)} \right| > \varepsilon/2\right) \\ &= \mathbb{P}\left(\sup_{|r-q|<\delta} \frac{\sqrt{N}}{\lfloor N\theta_0 \rfloor} \left| \sum_{h=1}^m \sum_{n \in Q(h, \lfloor \lfloor N\theta_0 \rfloor q \rfloor)} Y_{n,m}^{(1)} - \sum_{n \in Q(h, \lfloor \lfloor N\theta_0 \rfloor r \rfloor)} Y_{n,m}^{(1)} \right| > \varepsilon/2\right) \\ &\leq \mathbb{P}\left(\sup_{|r-q|<\delta} m \max_{1 \leq h \leq m} \frac{\sqrt{N}}{\lfloor N\theta_0 \rfloor} \left| \sum_{n \in Q(h, \lfloor \lfloor N\theta_0 \rfloor q \rfloor)} Y_{n,m}^{(1)} - \sum_{n \in Q(h, \lfloor \lfloor N\theta_0 \rfloor r \rfloor)} Y_{n,m}^{(1)} \right| > \varepsilon/2\right) \\ &\leq m \mathbb{P}\left(\sup_{|r-q|<\delta} \frac{\sqrt{N}}{\lfloor N\theta_0 \rfloor} \left| \sum_{n \in Q(1, \lfloor \lfloor N\theta_0 \rfloor q \rfloor)} Y_{n,m}^{(1)} - \sum_{n \in Q(1, \lfloor \lfloor N\theta_0 \rfloor r \rfloor)} Y_{n,m}^{(1)} \right| > \varepsilon/(2m)\right). \end{aligned}$$

In the last step we have used that all sums are identically distributed and we have assumed without loss of generality that  $\lfloor \lfloor N\theta_0 \rfloor q \rfloor$  and  $\lfloor \lfloor N\theta_0 \rfloor r \rfloor$  are divisible by  $m$  (otherwise the remaining term, is asymptotically negligible). Taking the limit with respect to  $N$  we observe that the right side converges to 0, which follows due to asymptotic tightness of partial sums of independent random variables, as presented in van der Vaart and Wellner (1996).

The remaining part of this proof consists in verifying the marginal convergence of (A.8). More precisely, we prove by an application of the Cramer-Wold device for a finite number of parameters  $0 \leq \lambda_1 \leq \dots \leq \lambda_K \leq 1$  weak convergence of the random vector

$$[R_N^{(1)}(\lambda_1), R_N^{(2)}(\lambda_1), \dots, R_N^{(1)}(\lambda_K), R_N^{(2)}(\lambda_K)].$$

Again (A.13) and the Cauchy-Schwarz inequality show that we can replace this vector by its  $m$ -dependent version  $[R_{N,m}^{(1)}(\lambda_1), R_{N,m}^{(2)}(\lambda_1), \dots, R_{N,m}^{(1)}(\lambda_K), R_{N,m}^{(2)}(\lambda_K)]$ , where

$$\begin{aligned} & (R_{N,m}^{(1)}(\lambda), R_{N,m}^{(2)}(\lambda)) \\ &= \left( \frac{\sqrt{N}\lambda}{\lfloor N\theta_0 \rfloor} \sum_{n=1}^{\lfloor \lfloor N\theta_0 \rfloor \lambda \rfloor} Y_{n,m}^{(1)}, \frac{\sqrt{N}\lambda}{\lfloor N(1-\theta_0) \rfloor} \sum_{n=\lfloor N\theta_0 \rfloor + 1}^{\lfloor \lfloor N(1-\theta_0) \rfloor \lambda \rfloor + \lfloor N\theta_0 \rfloor} Y_{n,m}^{(2)} \right)_{\lambda \in [0,1]}. \end{aligned}$$

Now let  $a_1, \dots, a_K$  and  $b_1, \dots, b_K$  be arbitrary real numbers and consider the sum

$$\sum_{j=1}^K a_j R_{N,m}^{(1)}(\lambda_j) + b_j R_{N,m}^{(2)}(\lambda_j) \tag{A.14}$$

$$\begin{aligned}
 &= \sqrt{N} \sum_{j=1}^K \left\{ \sum_{n=\lfloor \lambda_{j-1} N \theta_0 \rfloor}^{\lfloor \lambda_j N \theta_0 \rfloor} Y_{n,m}^{(1)} \frac{\lambda_j}{\lfloor N \theta_0 \rfloor} \sum_{l=1}^j a_l \right\} \\
 &\quad + \sqrt{N} \sum_{j=1}^K \left\{ \sum_{n=\lfloor \lambda_{j-1} N \rfloor + \lfloor N \theta_0 \rfloor}^{\lfloor \lambda_j N \rfloor + \lfloor N \theta_0 \rfloor} Y_{n,m}^{(2)} \frac{\lambda_j}{\lfloor N(1-\theta_0) \rfloor} \sum_{l=1}^j b_l \right\}
 \end{aligned}$$

(here we put  $\lambda_0 = 0$ ). To establish weak convergence we use the central limit theorem from Berk (1973) for  $m$ -dependent triangular arrays of random variables. The only non-trivial assumption of this theorem in our case is the convergence of the variance of (A.14), which will be established next. As we can see the covariance of the two groups for  $i = 1, 2$  converges to 0, since

$$\begin{aligned}
 &\frac{1}{N} \text{Cov} \left( \sum_{j=1}^K \sum_{n=\lambda_{j-1} \lfloor N \theta_0 \rfloor}^{\lfloor \lambda_j \lfloor N \theta_0 \rfloor \rfloor} Y_{n,m}^{(1)} \frac{N \lambda_j}{\lfloor N \theta_0 \rfloor} \sum_{l=1}^j a_l, \right. \\
 &\quad \left. \sum_{j=1}^K \sum_{n=\lfloor \lambda_{j-1} N \rfloor + \lfloor N \theta_0 \rfloor}^{\lfloor \lambda_j N \rfloor + \lfloor N \theta_0 \rfloor} Y_{n,m}^{(2)} \frac{N \lambda_j}{\lfloor N(1-\theta_0) \rfloor} \sum_{l=1}^j b_l \right)
 \end{aligned}$$

is of order  $o(1)$  due to  $m$ -dependence. Consequently, we can investigate the variance of each of the two terms in (A.14) separately. Iterating the argument yields that we may confine ourselves to the variance of the terms

$$\begin{aligned}
 &\sqrt{N} \sum_{n=\lambda_{j-1} \lfloor N \theta_0 \rfloor}^{\lfloor \lambda_j \lfloor N \theta_0 \rfloor \rfloor} Y_{n,m}^{(1)} \frac{\lambda_j}{\lfloor N \theta_0 \rfloor} \sum_{l=1}^j a_l \quad \text{and} \\
 &\sqrt{N} \sum_{n=\lfloor \lambda_{j-1} N \rfloor + \lfloor N \theta_0 \rfloor}^{\lfloor \lambda_j N \rfloor + \lfloor N \theta_0 \rfloor} Y_{n,m}^{(2)} \frac{\lambda_{j'}}{\lfloor N(1-\theta_0) \rfloor} \sum_{l=1}^{j'} b_l,
 \end{aligned}$$

for  $j, j' = 1, \dots, K$  separately. The subsequent convergence arguments are now the same for all remaining terms and we only exemplarily consider the first one and obtain

$$\begin{aligned}
 &\text{Var} \left\{ \sqrt{N} \sum_{n=1}^{\lfloor \lambda_1 \lfloor N \theta_0 \rfloor \rfloor} Y_{n,m}^{(1)} \frac{\lambda_1}{\lfloor N \theta_0 \rfloor} a_1 \right\} = \frac{\lambda_1^2 N a_1^2}{\lfloor N \theta_0 \rfloor^2} \text{Var} \left\{ \sum_{n=1}^{\lfloor \lambda_1 \lfloor N \theta_0 \rfloor \rfloor} Y_{n,m}^{(1)} \right\} \quad (\text{A.15}) \\
 &= \frac{\lambda_1^2 \lfloor \lambda_1 \lfloor N \theta_0 \rfloor \rfloor N a_1^2}{\lfloor N \theta_0 \rfloor^2} \sum_{|k| \leq m} (1 - |k| / \lfloor \lambda_1 \lfloor N \theta_0 \rfloor \rfloor) \text{Cov}(Y_{0,m}^{(1)}, Y_{k,m}^{(1)}) \\
 &\rightarrow \lambda_1^3 \frac{a_1^2}{\theta_0} \sum_{|k| \leq m} \text{Cov}(Y_{0,m}^{(1)}, Y_{k,m}^{(1)}).
 \end{aligned}$$

Finally we have to show that for  $m \rightarrow \infty$  the sum on the right-hand side converges to

$$(\sigma_E^{(1)})^2 = \text{Var}(Y_0^{(1)}) + 2 \sum_{l \geq 1} \text{Cov}(Y_0^{(1)}, Y_l^{(1)}) > 0, \quad (\text{A.16})$$

which follows from

$$C_m := \sum_{|k| \leq m} [\mathbb{C}ov(Y_0^{(1)}, Y_k^{(1)}) - \mathbb{C}ov(Y_{0,m}^{(1)}, Y_{k,m}^{(1)})] \rightarrow 0. \tag{A.17}$$

Note that  $\sigma_E^{(1)}$  is positive by assumption. To prove (A.17) consider the estimate

$$C_m \leq \sum_{|k| \leq m} |\mathbb{C}ov(Y_0^{(1)}, Y_k^{(1)} - Y_{k,m}^{(1)})| + \sum_{|k| \leq m} |\mathbb{C}ov(Y_{0,m}^{(1)} - Y_0^{(1)}, Y_{k,m}^{(1)})|.$$

Each term can be bounded according to Cauchy-Schwarz by

$$\sum_{|k| \leq m} \mathbb{E}[(Y_{0,m}^{(1)} - Y_0^{(1)})^2]^{1/2} \mathbb{E}[(Y_0^{(1)})^2]^{1/2} = m \mathbb{E}[(Y_0^{(1)})^2]^{1/2} \mathbb{E}[(Y_{0,m}^{(1)} - Y_0^{(1)})^2]^{1/2}. \tag{A.18}$$

The expectation on the right can be further analyzed plugging in the definitions of  $Y_0$  and  $Y_{0,m}$  (see (A.9), (A.10)).

$$\begin{aligned} & \mathbb{E}[(Y_{0,m}^{(1)} - Y_0^{(1)})^2]^{1/2} \\ &= \mathbb{E}\left[\left\{ \int_0^1 \int_0^1 [X_{0,m}^{(1)}(s)X_{0,m}^{(1)}(t) - X_0^{(1)}(s)X_0^{(1)}(t)]v_1^{(1)}(s)v_1^{(1)}(t)dsdt \right\}^2\right]^{1/2} \\ &\leq \mathbb{E}[\|X_0 \otimes X_0 - X_{0,m} \otimes X_{0,m}\|^2]^{1/2} \\ &\leq 2\mathbb{E}[\|X_0 - X_{0,m}\|^2\|X_0\|^2 + \|X_0 - X_{0,m}\|^2\|X_{0,m}\|^2]^{1/2} \\ &\leq 4\mathbb{E}[\|X_0 - X_{0,m}\|^4]^{1/4}\mathbb{E}[\|X_0\|^4]^{1/4}. \end{aligned}$$

Using this estimate in (A.18) yields the upper bound

$$C_m \leq L(2m + 1)\mathbb{E}[\|X_0 - X_{0,m}\|^4]^{1/4}, \tag{A.19}$$

for some fixed constant  $L > 0$ . According to Assumption 2.1 there exists a sufficiently small number  $\eta > 0$ , such that the sequence  $\mathbb{E}[\|X_0 - X_{0,m}\|^4]^{1/4+\eta}$  is summable, and we obtain

$$\lim_{m \rightarrow \infty} (2m + 1)\mathbb{E}[\|X_0 - X_{0,m}\|^4]^{1/4} = 0.$$

Analogously to  $\sigma_E^{(1)}$  we define the long run variance for the random variables  $Y_n^{(2)}$  as

$$(\sigma_E^{(2)})^2 = \mathbb{V}ar(Y_0^{(2)}) + 2 \sum_{l \geq 1} \mathbb{C}ov(Y_0^{(2)}, Y_l^{(2)}) > 0 \tag{A.20}$$

(recall that  $\sigma_E^{(2)}$  is positive by assumption). We have now verified all conditions of the Theorem in Berk (1973) and this implies convergence of the finite

dimensional distributions. Consequently it follows that

$$\begin{aligned} & \left( \frac{\sqrt{N}\lambda}{[N\theta_0]} \sum_{n=1}^{\lfloor [N\theta_0]\lambda \rfloor} Y_n^{(1)}, \frac{\sqrt{N}\lambda}{[N(1-\theta_0)]} \sum_{n=\lfloor [N\theta_0]\lambda \rfloor+1}^{\lfloor [N\theta_0] + \lfloor [N(1-\theta_0)]\lambda \rfloor \rfloor} Y_n^{(2)} \right)_{\lambda \in [0,1]} \\ & \rightarrow_D \left( \sigma_E^{(1)} \lambda \mathbb{B}_1(\lambda) / \sqrt{\theta_0}, \sigma_E^{(2)} \lambda \mathbb{B}_2(\lambda) / \sqrt{1-\theta_0} \right)_{\lambda \in [0,1]}, \end{aligned}$$

where  $\mathbb{B}_1$  and  $\mathbb{B}_2$  are independent, standard Brownian motions. Combining these considerations with the continuous mapping theorem shows that

$$\begin{aligned} & \left( \sqrt{N}\lambda^2 [(\tilde{\tau}_{1,\lambda}^{(1)} - \tilde{\tau}_{1,\lambda}^{(2)})^2 - (\tau_1^{(1)} - \tau_1^{(2)})^2] \right)_{\lambda \in [0,1]} \\ & \rightarrow_D \left( 2|\tau_1^{(1)} - \tau_1^{(2)}| \lambda [\sigma_E^{(1)} / \sqrt{\theta_0} \mathbb{B}_1(\lambda) + \sigma_E^{(2)} / \sqrt{1-\theta_0} \mathbb{B}_2(\lambda)] \right)_{\lambda \in [0,1]}, \end{aligned}$$

which completes the proof. □

### A.2. Proof of Theorem 2.5

Recall that the probability of rejecting the null hypothesis in (2.5) is given by

$$\mathbb{P}\left(\frac{\hat{E}_{1,N} - \Delta_\tau}{\hat{V}_{1,N}} > q_{1-\alpha}\right) = \mathbb{P}\left(\frac{\hat{E}_{1,N} - E_1}{\hat{V}_{1,N}} > q_{1-\alpha} + \frac{\Delta_\tau - E_1}{\hat{V}_{1,N}}\right).$$

Suppose that  $E_1 > 0$  i.e. that the first eigenvalues of  $c^{(1)}$  and  $c^{(2)}$  are different. Then by Proposition 2.2,

$$\frac{\hat{E}_{1,N} - E_1}{\hat{V}_{1,N}} \rightarrow_D W, \tag{A.21}$$

where the random variable  $W$  is defined in (2.19). The same result shows that  $\hat{V}_{1,N}$  is of order  $o_P(1)$  which implies that  $(\Delta_\tau - E_1)/\hat{V}_{1,N}$  converges to  $+\infty$  if  $\Delta_\tau > E_1$  and to  $-\infty$  if  $\Delta_\tau < E_1$ , both in probability. This implies consistency and level  $\alpha$  as stated in (2.21), in the case  $E_1 > 0$ .

Now suppose that  $E_1 = 0$ . In this case we show that  $\hat{E}_{1,N}$  and  $\hat{V}_{1,N}$  are of order  $o_P(1)$ . Then, with probability converging to 1,  $\hat{E}_{1,N}/\hat{V}_{1,N}$  is asymptotically negligible compared to  $\Delta_\tau/\hat{V}_{1,N}$  which converges to infinity. First, suppose that the kernels  $c^{(1)}$  and  $c^{(2)}$  are not equal. Proposition 2.2 implies that we may replace the change point estimator in numerator  $\hat{E}_{1,N}$  and denominator  $\hat{V}_{1,N}$  by the actual change point, only incurring asymptotically vanishing terms. More precisely, the denominator in (A.21) equals

$$\begin{aligned} \hat{V}_{1,N} &= \left( \int_0^1 \lambda^4 [(\tilde{\tau}_{1,\lambda}^{(1)} - \tilde{\tau}_{1,\lambda}^{(2)})^2 - (\tilde{\tau}_{1,1}^{(1)} - \tilde{\tau}_{1,1}^{(2)})^2]^2 \nu(d\lambda) \right)^{1/2} + o_P(1) \\ &= \left( \int_0^1 \lambda^4 [(\tilde{\tau}_{1,\lambda}^{(1)} - \tau_1^{(1)} + \tau_1^{(2)} - \tilde{\tau}_{1,\lambda}^{(2)})^2 - (\tilde{\tau}_{1,1}^{(1)} - \tau_1^{(1)} + \tau_1^{(2)} - \tilde{\tau}_{1,1}^{(2)})^2]^2 \nu(d\lambda) \right)^{1/2} \\ &\quad + o_P(1), \end{aligned}$$

where we have used the equality of the eigenvalues. Now Proposition B.2 in Appendix B shows that  $\hat{V}_{1,N} = o_P(1)$ . Finally, similar, but simpler arguments show that  $\hat{E}_{1,N} = o_P(1)$ .

In the case of equality  $c^{(1)} = c^{(2)}$  the estimator of the change point assumes some uninformative value inside the interval  $[\varepsilon, 1 - \varepsilon]$  and we obtain

$$\begin{aligned} \hat{V}_{1,N}^2 &\leq 4 \sup_{\lambda} \left( \lambda^2 (\hat{\tau}_{1,\lambda}^{(1)} - \tau_1^{(1)} + \tau_1^{(2)} - \hat{\tau}_{1,\lambda}^{(2)})^2 \right)^2 \\ &\leq 64 \sup_{\lambda} \left( \lambda^2 (\hat{\tau}_{1,\lambda}^{(1)} - \tau_1^{(1)})^2 \right)^2 + 64 \sup_{\lambda} \left( \lambda^2 (\tau_1^{(2)} - \hat{\tau}_{1,\lambda}^{(2)})^2 \right)^2. \end{aligned}$$

Consider now the inequality

$$\left( \lambda^2 (\hat{\tau}_{1,\lambda}^{(1)} - \tau_1^{(1)})^2 \right)^2 \leq \sup_{\theta \in [\varepsilon, 1 - \varepsilon]} \left\| \frac{1}{[N\theta]} \sum_{n=1}^{\lfloor [N\theta]\lambda \rfloor} (X_n \otimes X_n - c^{(1)}) \right\|^4 + o_P(1)$$

which follows from Lemma 2.2 in Horváth and Kokoszka (2012). The above expression is of order  $o_P(1)$  uniformly in  $\lambda$  by Lemma B.1 in the supplementary material of Aue et al. (2018). Applying the same argument to the second term yields that  $\hat{V}_{1,N} = o_P(1)$ . The corresponding arguments for the estimate  $\hat{E}_{1,N} = o_P(1)$  are similar and therefore again omitted.  $\square$

**A.3. Proof of Theorem 2.6**

Recall the definition of the eigenfunctions  $\hat{v}_{1,\lambda}^{(i)}$  of the estimated kernels  $\hat{c}^{(i)}(\lambda, \cdot, \cdot)$  defined in (2.9) and (2.10). Similarly as for the proof of Theorem 2.5 we prove the weak convergence

$$\{H_N(\lambda)\}_{\lambda \in [0,1]} \rightarrow_D \{\lambda \sigma_D \mathbb{B}(\lambda)\}_{\lambda \in [0,1]}, \tag{A.22}$$

where the process  $\{H_N(\lambda)\}_{\lambda \in [0,1]}$  is defined by

$$\begin{aligned} H_N(\lambda) &:= \sqrt{N} \left[ \int_0^1 \left( \hat{v}_{1,\lambda}^{(1)}(t) - \hat{v}_{1,\lambda}^{(2)}(t) \right)^2 \lambda^2 dt \right. \\ &\quad \left. - \int_0^1 \left( v_1^{(1)}(t) - v_1^{(2)}(t) \right)^2 \lambda^2 dt \right] \end{aligned}$$

The result then follows by similar arguments as given in the proof of Theorem 2.5 and the continuous mapping theorem, which implies the weak convergence of the tuple

$$\left( \sqrt{N}(\hat{D}_{1,N} - D_1), \sqrt{N}\hat{U}_{1,N} \right) = \left( H_N(1), \left[ \int_0^1 \{H_N(\lambda) - \lambda^2 H_N(1)\}^2 d\nu(\lambda) \right]^{1/2} \right).$$

First we replace the estimated by the true change point showing

$$\sqrt{N}\lambda^2 \left[ \int_0^1 \left( \hat{v}_{1,\lambda}^{(1)}(t) - \tilde{v}_{1,\lambda}^{(2)}(t) \right)^2 dt \right] \tag{A.23}$$

$$- \int_0^1 \left( \hat{v}_{1,\lambda}^{(1)}(t) - \hat{v}_{1,\lambda}^{(2)}(t) \right)^2 dt \Big] = o_P(1)$$

uniformly in  $\lambda \in [0, 1]$ . To establish this equality we prove

$$\lambda^2 \|\tilde{v}_{1,\lambda}^{(1)} - \hat{v}_{1,\lambda}^{(1)}\| = o_P(1/\sqrt{N}) \tag{A.24}$$

uniformly in  $\lambda \in [0, 1]$ . (A.24) then follows from (A.23) by a simple application of the Cauchy-Schwarz inequality. Note that we may confine ourselves to considering  $\lambda \in (1/\sqrt{N}, 1)$ , since for  $\lambda \in (0, 1/\sqrt{N})$  the left side of (A.24) is upper bounded by  $2/N$ . To derive (A.24) for  $\lambda > 1/\sqrt{N}$  we use Lemma 2.3 from Horváth and Kokoszka (2012) and obtain

$$\lambda^2 \|\tilde{v}_{1,\lambda}^{(1)} - \hat{v}_{1,\lambda}^{(1)}\| \leq \lambda^2 \frac{2\sqrt{2} \|\tilde{c}^{(1)}(\lambda, \cdot, \cdot) - \hat{c}^{(1)}(\lambda, \cdot, \cdot)\|}{\tilde{\tau}_{1,\lambda}^{(1)} - \tilde{\tau}_{2,\lambda}^{(1)}}, \tag{A.25}$$

where  $\tilde{\tau}_{1,\lambda}^{(1)}$  and  $\hat{\tau}_{2,\lambda}^{(1)}$  are the eigenvalues of covariance kernels  $\tilde{c}^{(1)}(\lambda, \cdot, \cdot)$  defined in (A.1). We now consider numerator and denominator separately.

Beginning with the denominator we first notice that by consistency of the estimated eigenvalues  $\tilde{\tau}_{1,\lambda}^{(1)}$  and  $\tilde{\tau}_{2,\lambda}^{(1)}$  we have  $\tilde{\tau}_{1,\lambda}^{(1)} - \tilde{\tau}_{2,\lambda}^{(1)} = \tau_1^{(1)} - \tau_2^{(1)} + o_P(1)$ . In particular, since  $\tau_1^{(1)} > \tau_2^{(1)}$  by Assumption 2.3 it is bounded away from 0 with probability converging to 1. To see the consistency of the eigenvalues, we use the following upper bound

$$|\tilde{\tau}_{1,\lambda}^{(1)} - \tau_1^{(1)}| \leq \|\tilde{c}^{(1)}(\lambda, \cdot, \cdot) - c^{(1)}(\cdot, \cdot)\| = \mathcal{O}_P(N^{-1/4} \log(N)),$$

where the second equality follows from Lemma B.1 in the supplementary material of Aue et al. (2018) and holds uniformly in  $\lambda \in (1/\sqrt{N}, 1)$ . Applying the same argument to the second eigenvalue yields also consistency of  $\tilde{\tau}_{2,\lambda}^{(1)}$ .

In the proof of Proposition 2.2 (step 1), we have already shown that the numerator on the right side of (A.25) is of order  $o_P(1/\sqrt{N})$  and hence (A.23) follows.

We now turn to an investigation of the process

$$K_N(\lambda) := \sqrt{N} \left[ \int_0^1 \left( \tilde{v}_{1,\lambda}^{(1)}(t) - \tilde{v}_{1,\lambda}^{(2)}(t) \right)^2 \lambda^2 dt - \int_0^1 \left( v_1^{(1)}(t) - v_1^{(2)}(t) \right)^2 \lambda^2 dt \right],$$

for which a simple calculation shows

$$\begin{aligned} K_N(\lambda) &= \sqrt{N} \left[ \int_0^1 \left( \tilde{v}_{1,\lambda}^{(1)}(t) - \tilde{v}_{1,\lambda}^{(2)}(t) - v_1^{(1)}(t) + v_1^{(2)}(t) \right)^2 \lambda^2 dt \right] \tag{A.26} \\ &\quad + 2\sqrt{N} \left[ \int_0^1 \left( v_1^{(1)}(t) - v_1^{(2)}(t) \right) \left( \tilde{v}_{1,\lambda}^{(1)}(t) - \tilde{v}_{1,\lambda}^{(2)}(t) - v_1^{(1)}(t) + v_1^{(2)}(t) \right) \lambda^2 dt \right] \\ &= 2\sqrt{N} \left[ \int_0^1 \left( v_1^{(1)}(t) - v_1^{(2)}(t) \right) \left( \tilde{v}_{1,\lambda}^{(1)}(t) - \tilde{v}_{1,\lambda}^{(2)}(t) - v_1^{(1)}(t) + v_1^{(2)}(t) \right) \lambda^2 dt \right] \\ &\quad + o_P(1). \end{aligned}$$



For the second equality we have used Proposition 2.1 from Aue et al. (2019). In order to determine the limiting behavior of this expression, we make several technically helpful transformations beginning with a linearization.

Similar calculations as in the proof of Proposition 2.3 in Aue et al. (2019) yield the representation

$$K_N(\lambda) = 2\lambda\sqrt{N} \left( \frac{1}{\lfloor N\theta_0 \rfloor} \sum_{n=1}^{\lfloor \lfloor N\theta_0 \rfloor \lambda \rfloor} \bar{X}_n^{(1)} \right) \tag{A.27}$$

$$+ \frac{1}{\lfloor N(1-\theta_0) \rfloor} \sum_{n=\lfloor N\theta_0 \rfloor+1}^{\lfloor N\theta_0 \rfloor + \lfloor \lfloor N(1-\theta_0) \rfloor \lambda \rfloor} \bar{X}_n^{(2)} \Big) + o_P(1), \tag{A.28}$$

where the random variables  $\bar{X}_n^{(1)}$  and  $\bar{X}_n^{(2)}$  are defined by

$$\bar{X}_n^{(i)} := \int_0^1 \tilde{X}_n^{(i)}(s, t) f^{(i)}(s, t) ds dt \quad i = 1, 2, \tag{A.29}$$

$$\tilde{X}_n^{(i)}(s, t) = X_n^{(i)}(s)X_n^{(i)}(t) - \mathbb{E}[X_n^{(i)}(s)X_n^{(i)}(t)] \quad i = 1, 2 \tag{A.30}$$

and

$$f^{(i)}(s, t) = -v_1^{(i)}(s) \sum_{k \geq 2} \frac{v_k^{(i)}(t)}{\tau_1^{(i)} - \tau_k^{(i)}} \int_0^1 v_k^{(i)}(u) v_1^{(3-i)}(u) du.$$

Notice that  $f^{(i)} \in L^2[0, 1]$ . Weak convergence of the process  $\{K_N(\lambda)\}_{\lambda \in [0, 1]}$  defined in (A.27) follows from weak convergence of the two dimensional process

$$\left\{ \frac{\sqrt{N}\lambda}{\lfloor N\theta_0 \rfloor} \sum_{n=1}^{\lfloor \lfloor N\theta_0 \rfloor \lambda \rfloor} \bar{X}_n^{(1)}, \frac{\sqrt{N}\lambda}{\lfloor N(1-\theta_0) \rfloor} \sum_{n=\lfloor N\theta_0 \rfloor+1}^{\lfloor N\theta_0 \rfloor + \lfloor \lfloor N(1-\theta_0) \rfloor \lambda \rfloor} \bar{X}_n^{(2)} \right\}_{\lambda \in [0, 1]} \tag{A.31}$$

and the continuous mapping theorem.

Similar arguments as given in Aue et al. (2019) show that the components of (A.31)

$$\left( \lambda \frac{\sqrt{N}}{\lfloor N\theta_0 \rfloor} \sum_{n=1}^{\lfloor \lfloor N\theta_0 \rfloor \lambda \rfloor} \bar{X}_n^{(1)} \right)_{\lambda \in [0, 1]} \quad \text{and}$$

$$\left( \lambda \frac{\sqrt{N}}{\lfloor N(1-\theta_0) \rfloor} \sum_{n=\lfloor N\theta_0 \rfloor+1}^{\lfloor N\theta_0 \rfloor + \lfloor \lfloor N(1-\theta_0) \rfloor \lambda \rfloor} \bar{X}_n^{(2)} \right)_{\lambda \in [0, 1]}$$

converge weakly to stochastic processes of the form  $\sigma_D^{(1)} \sqrt{\theta_0} \lambda \mathbb{B}_1(\lambda)$  and  $\sigma_D^{(2)} \sqrt{1-\theta_0} \lambda \mathbb{B}_2(\lambda)$  for some suitable constants  $\sigma_D^{(1)}$  and  $\sigma_D^{(2)}$  (see equation (A.33) below), where  $\mathbb{B}_1$  and  $\mathbb{B}_2$  are independent, standard Brownian motions. In particular both processes are asymptotically tight and consequently the vector in (A.31) is also asymptotically tight. To complete the proof of weak convergence

of (A.31), it therefore remains to prove the convergence of the finite dimensional distributions.

For this purpose we replace the random variables  $\tilde{X}_n^{(i)}$  in (A.30) by their  $m$ -dependent analogues

$$\bar{X}_{n,m}^{(i)} := \int_0^1 \tilde{X}_{n,m}^{(i)}(s,t) f^{(i)}(s,t) ds dt \quad i = 1, 2 \tag{A.32}$$

where

$$\tilde{X}_{n,m}^{(i)}(s,t) := X_{n,m}^{(i)}(s)X_{n,m}^{(i)}(t) - \mathbb{E}[X_{n,m}^{(i)}(s)X_{n,m}^{(i)}(t)] \quad (i = 1, 2)$$

and the variables  $X_{n,m}^{(i)}$  are defined in (2.7). The norm of

$$\begin{aligned} & \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor \lambda \lfloor N\theta_0 \rfloor \rfloor} (\bar{X}_n^{(1)} - \bar{X}_{n,m}^{(1)}) \\ &= \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor \lambda \lfloor N\theta_0 \rfloor \rfloor} \int_0^1 \int_0^1 (\tilde{X}_n^{(1)}(s,t) - \tilde{X}_{n,m}^{(1)}(s,t)) f^{(1)}(s,t) ds dt \end{aligned}$$

is bounded by

$$\begin{aligned} & \max_{k=1, \dots, \lfloor N\theta_0 \rfloor} \left\| \frac{1}{\sqrt{N}} \sum_{n=1}^k (X_n^{(1)} \otimes X_n^{(1)} - \mathbb{E}X_n^{(1)} \otimes X_n^{(1)} \right. \\ & \quad \left. - X_{n,m}^{(1)} \otimes X_{n,m}^{(1)} + \mathbb{E}X_{n,m}^{(1)} \otimes X_{n,m}^{(1)}) \right\| \|f^{(1)}\| \\ &= \max_{k=1, \dots, \lfloor N\theta_0 \rfloor} \left\| \frac{1}{\sqrt{N}} \sum_{n=1}^k (X_n^{(1)} \otimes X_n^{(1)} - X_{n,m}^{(1)} \otimes X_{n,m}^{(1)}) \right\| \|f^{(1)}\| \end{aligned}$$

and consequently converges to 0 in probability according to (A.13) if  $m \rightarrow \infty$ . The case  $i = 2$  can be treated analogously. Therefore it is sufficient to prove the convergence of the finite dimensional distributions of the vector

$$\left( \frac{\sqrt{N}\lambda}{\lfloor N\theta_0 \rfloor} \sum_{n=1}^{\lfloor \lfloor N\theta_0 \rfloor \lambda \rfloor} \bar{X}_{n,m}^{(1)}, \frac{\sqrt{N}\lambda}{\lfloor N(1-\theta_0) \rfloor} \sum_{n=\lfloor N\theta_0 \rfloor+1}^{\lfloor N\theta_0 \rfloor + \lfloor \lfloor N(1-\theta_0) \rfloor \lambda \rfloor} \bar{X}_{n,m}^{(2)} \right)_{\lambda \in [0,1]},$$

which can be shown in the same way as in the proof of Step 2 in Proposition 2.4. Finally, we define

$$(\sigma_D^{(i)})^2 = \text{Var}(\bar{X}_0^{(i)}) + 2 \sum_{l \geq 1} \text{Cov}(\bar{X}_0^{(i)}, \bar{X}_l^{(i)}) > 0 \quad (i = 1, 2) \tag{A.33}$$

(these quantities are positive by assumption) and obtain

$$\left( \frac{\sqrt{N}\lambda}{\lfloor N\theta_0 \rfloor} \sum_{n=1}^{\lfloor \lfloor N\theta_0 \rfloor \lambda \rfloor} \bar{X}_n^{(1)}, \frac{\sqrt{N}\lambda}{\lfloor N(1-\theta_0) \rfloor} \sum_{n=\lfloor N\theta_0 \rfloor+1}^{\lfloor N\theta_0 \rfloor + \lfloor \lfloor N(1-\theta_0) \rfloor \lambda \rfloor} \bar{X}_n^{(2)} \right)_{\lambda \in [0,1]}$$

$$\rightarrow_D \left( \sigma_D^{(1)} \lambda \mathbb{B}_1(\lambda) / \sqrt{\theta_0}, \sigma_D^{(2)} \lambda \mathbb{B}_2(\lambda) / \sqrt{1 - \theta_0} \right)_{\lambda \in [0,1]}.$$

The continuous mapping theorem gives

$$\begin{aligned} & \left( \sqrt{N} \left[ \int_0^1 \left( \hat{v}_{1,\lambda}^{(1)}(t) - \hat{v}_{1,\lambda}^{(2)}(t) \right)^2 \lambda^2 dt - \int_0^1 \left( v_1^{(1)}(t) - v_1^{(2)}(t) \right)^2 \lambda^2 dt \right] \right)_{\lambda \in [0,1]} \\ \rightarrow_D & \left( 2[\lambda \sigma_D^{(1)} / \sqrt{\theta_0} \mathbb{B}_1(\lambda) + \lambda \sigma_D^{(2)} / \sqrt{1 - \theta_0} \mathbb{B}_2(\lambda)] \right)_{\lambda \in [0,1]}. \end{aligned}$$

Now the same steps as in the proof of Theorem 2.5 yield the desired result.  $\square$

## Appendix B: Technical details and supplementary results

### B.1. Weak convergence of the covariance kernel

In this section we provide an adaption of Theorem 1.1 in Berkes et al. (2013) to the estimation of covariance kernels. Let  $(X_n)_{n \in \mathbb{Z}}$  be a sequence of random functions satisfying Assumption 2.1 and consider the sequential process

$$S_N(x, s, t) := \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \{X_n(s)X_n(t) - \mathbb{E}X_n(s)X_n(t)\}. \tag{B.1}$$

Thus we are interested in a sum of random elements  $X_n \otimes X_n \in L^2([0, 1] \times [0, 1])$ . These products can be approximated by products of the  $m$ -dependent random functions  $X_{n,m} \otimes X_{n,m}$ , where the random variables  $X_{n,m}$  are defined in Assumption 2.1 (note that  $X_n$  and  $X_{n,m}$  have the same distribution). Using Assumption 2.1 and the notation  $\delta' = \delta/2$  and  $\kappa' = \kappa/2$  we obtain for a suitable constant  $K > 0$

$$\begin{aligned} & \sum_{m \geq 1} \left( \mathbb{E} \|X_n \otimes X_n - \mathbb{E}X_n \otimes X_n - X_{n,m} \otimes X_{n,m} + \mathbb{E}X_{n,m} \otimes X_{n,m}\|^{2+\delta'} \right)^{1/\kappa'} \\ = & \sum_{m \geq 1} \left( \mathbb{E} \|X_n \otimes X_n - X_{n,m} \otimes X_{n,m}\|^{2+\delta'} \right)^{1/\kappa'} \\ \leq & K \sum_{m \geq 1} \left( \mathbb{E} \|X_n\|^{4+2\delta'} \mathbb{E} \|X_n - X_{n,m}\|^{4+2\delta'} \right)^{1/(2\kappa')} \\ & + K \sum_{m \geq 1} \left( \mathbb{E} \|X_{n,m}\|^{4+2\delta'} \mathbb{E} \|X_n - X_{n,m}\|^{4+2\delta'} \right)^{1/(2\kappa')} \\ = & 2K \left( \mathbb{E} \|X_0\|^{4+2\delta'} \right)^{1/(2\kappa')} \sum_{m \geq 1} \left( \mathbb{E} \|X_n - X_{n,m}\|^{4+2\delta'} \right)^{1/(2\kappa')} < \infty. \end{aligned}$$

This consideration demonstrates that we have the same approximation properties as required in Berkes et al. (2013) for the random functions  $X_n \otimes X_n$ .

The corresponding “long run covariance kernel” for the sequence  $(X_n \otimes X_n)_{n \in \mathbb{Z}}$  can be defined as

$$C(q, r, s, t) := \mathbb{E}[X_0(q)X_0(r)X_0(s)X_0(t)] + \sum_{n \geq 1} \mathbb{E}[X_0(q)X_0(r)X_n(s)X_n(t)] + \sum_{n \geq 1} \mathbb{E}[X_n(q)X_n(r)X_0(s)X_0(t)].$$

By analogous arguments as presented in Lemma 2.2 of Berkes et al. (2013), it can be observed that  $C$  is square integrable and thus defines a Hilbert-Schmidt operator (see e.g. Bump (1996) p. 168). It thus follows that there exists a spectral decomposition of the integral operator with kernel  $C$ . Let us call its eigenfunctions  $\Psi_1, \Psi_2, \dots$  and its corresponding eigenvalues  $\Lambda_1, \Lambda_2, \dots$ . With this eigensystem we can define the Gaussian process

$$\Gamma(x, s, t) := \sum_{l \geq 1} \Lambda_l \mathbb{B}_l(x) \Psi_l(s, t),$$

where  $\mathbb{B}_l$  are independent Brownian motions for all  $l \geq 1$  and  $x \in [0, 1]$ . We now state an analogue of Berkes’ Theorem 1.1. The proof runs along the same lines as in Berkes et al. (2013) and is therefore omitted.

**Theorem B.1.** *Suppose Assumptions 2.1 hold. Then there exists for every  $N$  a Gaussian process  $\Gamma_N =_D \Gamma$  such that*

$$\sup_x \int_0^1 \int_0^1 (S_N(x, t, s) - \Gamma_N(x, s, t))^2 ds dt = o_P(1).$$

**B.2. Eigenvalue-expansion**

In this section we investigate a stochastic linearization of the estimated eigenvalues of the empirical covariance operator. For this purpose let  $(X_n)_{n \in \mathbb{Z}}$  be a stationary, functional time series, with vanishing mean function, that complies to the Assumptions 2.1 and 2.3. We call the corresponding covariance kernel  $c$ , its eigenvalues  $\tau_1 \geq \tau_2 \geq \dots$  and its eigenfunctions  $v_1, v_2, \dots$ . For the data sample  $X_1, \dots, X_N$  we define the sequential estimator of the covariance kernel

$$\hat{c}(\lambda, s, t) := \frac{1}{[N\lambda]} \sum_{n=1}^{[N\lambda]} X_n(s)X_n(t). \tag{B.2}$$

Its eigenvalues are denoted by  $\hat{\tau}_{1,\lambda} \geq \hat{\tau}_{2,\lambda} \geq \dots$  and its eigenfunctions by  $\hat{v}_{1,\lambda}, \hat{v}_{2,\lambda}, \dots$

**Proposition B.2.** *Under the above assumptions we have for any  $j \in \{1, \dots, p\}$*

$$\sup_{\lambda \in [0,1]} \left| \lambda(\hat{\tau}_{j,\lambda} - \tau_j) - \frac{1}{\sqrt{N}} \int_0^1 \hat{Z}_N(s, t, \lambda) v_j(s) v_j(t) ds dt \right| = o_P(1/\sqrt{N}), \tag{B.3}$$

where

$$\hat{Z}_N(s, t, \lambda) := \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor N\lambda \rfloor} (X_n(s)X_n(t) - c(s, t)).$$

Furthermore, with  $\kappa$  defined in Assumption 2.1, 4. we have

$$\sup_{\lambda \in [0,1]} |\sqrt{\lambda}(\hat{\tau}_{j,\lambda} - \tau_j)| = \mathcal{O}_P(\log(N)^{1/\kappa}/\sqrt{N}). \tag{B.4}$$

*Proof.* We begin by noticing that

$$\sup_{\lambda \in [0,1]} \left| \frac{\lfloor N\lambda \rfloor}{N} \int_0^1 \int_0^1 c(s, t)v_j(s)v_j(t)dsdt - \lambda\tau_j \right| \tag{B.5}$$

$$= \left| \frac{\lfloor N\lambda \rfloor}{N} \tau_j - \lambda\tau_j \right| = \mathcal{O}(1/N). \tag{B.6}$$

Recalling the definition of  $\hat{c}(\lambda, s, t)$  as defined in (B.2), we use (B.5) to rewrite (B.3) as

$$\begin{aligned} & \sup_{\lambda \in [0,1]} \lambda \left| \hat{\tau}_{j,\lambda} - \int_0^1 \int_0^1 \hat{c}(\lambda, s, t)v_j(s)v_j(t)dsdt \right| + \mathcal{O}_P(1/N) \\ &= \sup_{\lambda \in [0,1]} \lambda \left| \int_0^1 \int_0^1 \hat{c}(\lambda, s, t)[\hat{v}_{j,\lambda}(s)\hat{v}_{j,\lambda}(t) - v_j(s)v_j(t)]dsdt \right| + \mathcal{O}_P(1/N) \\ &\leq \sup_{\lambda \in [0,1]} \lambda \left| \int_0^1 \int_0^1 \hat{c}(\lambda, s, t)[\hat{v}_{j,\lambda}(s)\hat{v}_{j,\lambda}(t) - \hat{v}_{j,\lambda}(s)v_j(t)]dsdt \right| \\ &+ \sup_{\lambda \in [0,1]} \lambda \left| \int_0^1 \int_0^1 \hat{c}(\lambda, s, t)[\hat{v}_{j,\lambda}(s)v_j(t) - v_j(s)v_j(t)]dsdt \right| + \mathcal{O}_P(1/N) \\ &= A + B + \mathcal{O}_P(1/N), \end{aligned}$$

where the last equality defines the terms  $A$  and  $B$  in an obvious way. We now investigate the terms  $A$  and  $B$  separately. For the term  $A$  we observe that  $\hat{v}_j$  is the eigenfunction of the integral operator associated with  $\hat{c}$ , which gives

$$\begin{aligned} A &= \sup_{\lambda \in [0,1]} \lambda \left| \hat{\tau}_{j,\lambda} \int_0^1 \hat{v}_{j,\lambda}(t)[\hat{v}_{j,\lambda}(t) - v_j(t)]dt \right| \\ &= \sup_{\lambda \in [0,1]} \hat{\tau}_{j,\lambda} \|\hat{v}_{j,\lambda} - v_j\|^2 \lambda/2 = o_P(1/\sqrt{N}). \end{aligned}$$

Here the second equality follows by the parallelogram law and in the last step we used the estimate

$$\sup_{\lambda \in [0,1]} \|\hat{v}_{j,\lambda} - v_j\|^2 \lambda = o_P(1/\sqrt{N}) \tag{B.7}$$

(see Proposition 2.1 in Aue et al. (2019)) and

$$\sup_{\lambda \in [0,1]} \hat{\tau}_{j,\lambda} \leq \sup_{\lambda \in [0,1]} (|\hat{\tau}_{j,\lambda} - \tau_j| + \tau_j) = \mathcal{O}_P(1),$$

according to Lemma B.1 in the supplementary material for (Aue et al., 2018). Turning to term  $B$  we see that

$$B \leq \sup_{\lambda \in [0,1]} \lambda \left| \int_0^1 \int_0^1 [\hat{c}(\lambda, s, t) - c(s, t)] v_j(t) [\hat{v}_{j,\lambda}(s) - v_j(s)] ds dt \right| \\ + \sup_{\lambda \in [0,1]} \lambda \left| \int_0^1 \int_0^1 c(s, t) v_j(t) [\hat{v}_{j,\lambda}(s) - v_j(s)] ds dt \right| =: R_1 + R_2,$$

where the last equality defines the random variables  $R_1$  and  $R_2$  in an obvious manner. For the term  $R_1$  we obtain by the Cauchy-Schwarz inequality

$$R_1 \leq \sup_{\lambda \in [0,1]} \sqrt{\lambda} \|\hat{c}(\cdot, \cdot, \lambda) - c\| \|v_j\| (\sqrt{\lambda} \|\hat{v}_{j,\lambda} - v_j\|).$$

Again by Lemma B.1 from the supplement of (Aue et al., 2018) we observe that

$$\sup_{\lambda \in [0,1]} \sqrt{\lambda} \|\hat{c}(\cdot, \cdot, \lambda) - c\| = \mathcal{O}(\log^{1/\kappa} / \sqrt{N}).$$

(B.7) shows that  $R_1 = o_P(1/\sqrt{N})$ . We use similar arguments and obtain

$$R_2 = \sup_{\lambda \in [0,1]} \lambda \tau_j \left| \int_0^1 v_j(t) [\hat{v}_{j,\lambda}(t) - v_j(t)] dt \right| \\ = \sup_{\lambda \in [0,1]} (\lambda \|\hat{v}_{j,\lambda} - v_j\|^2) \tau_j / 2 = o_P(1/\sqrt{N}).$$

Combining these considerations proves the first assertion (B.3). For a proof of (B.4) we note that

$$\sup_{\lambda \in [0,1]} |\sqrt{\lambda}(\hat{\tau}_{j,\lambda} - \tau_j)| \leq \sup_{\lambda \in [0,1]} \left\| \frac{\sqrt{\lambda}}{[N\lambda]} \sum_{n=1}^{[N\lambda]} X_n \otimes X_n - c \right\| + \mathcal{O}_P(N^{-1}) \\ = \mathcal{O}_P(\log(N)^{1/\kappa} N^{-1/2}).$$

The first inequality follows from bounding the eigenvalue distance by the operator distance and this again by the  $L^2$ -distance of the kernels. The second one follows by a Lemma B.1 in the supplementary material for (Aue et al., 2018).  $\square$

**References**

Aston, J. A. and Kirch, C. (2012). Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis*, 109:204–220. [MR2922864](#)  
 Aue, A., Dette, H., and Rice, G. (2019). Two-sample tests for relevant differences in the eigenfunctions of covariance operators. [arXiv:1909.06098v1](#).

- Aue, A., Dubart Nourinho, D., and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110:378–392. [MR3338510](#)
- Aue, A., Gabrys, R., Horváth, L., and Kokoszka, P. (2009). Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis*, 100(10):2254–2269. [MR2560367](#)
- Aue, A., Rice, G., and Sönmez, O. (2018). Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society, Ser. B*, 80(3):509–529. [MR3798876](#)
- Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, 2:317–335. [MR0920141](#)
- Berk, K. N. (1973). A central limit theorem for  $m$ -dependent random variables with unbounded  $m$ . *Annals of Probability*, 1(2):352–354. [MR0350815](#)
- Berkes, I., Horváth, L., and Rice, G. (2013). Weak invariance principles for sums of dependent random functions. *Stochastic Processes and their Applications*, 123(2):385–403. [MR3003356](#)
- Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer-Verlag, New York. [MR1783138](#)
- Bücher, A., Dette, H., and Heinrichs, F. (2019). Detecting deviations from second-order stationarity in locally stationary functional time series. *Annals of the Institute of Statistical Mathematics*, to appear, [arXiv:1808.04092](#). [MR4113273](#)
- Bump, D. (1996). *Automorphic Forms and Representations*. Cambridge University Press, Cambridge. [MR1431508](#)
- Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *Journal of multivariate analysis*, 12(1):136–154. [MR0650934](#)
- Dette, H., Kokot, K., and Aue, A. (2019). Functional data analysis in the Banach space of continuous functions. *Annals of Statistics*, to appear, [arXiv:1710.07781](#). [MR4102692](#)
- Dette, H., Kokot, K., and Volgushev, S. (2018). Testing relevant hypotheses in functional time series via self-normalization. [arXiv:1809.06092](#). [MR4112779](#)
- Dette, H. and Wied, D. (2016). Detecting relevant changes in time series models. *Journal of the Royal Statistical Society: Series B*, 78:371–394. [MR3454201](#)
- Ferraty, F. and Vieu, P. (2010). *Nonparametric Functional Data Analysis*. Springer-Verlag, New York. [MR2229687](#)
- Fremdt, S., Horváth, L., Kokoszka, P., and Steinebach, J. (2014). Functional data analysis with an increasing number of projections. *Journal of Multivariate Analysis*, 124:313–332. [MR3147328](#)
- Gao, Y., Shang, H. L., and Yang, Y. (2019). High-dimensional functional time series forecasting: An application to age-specific mortality rates. *Journal of Multivariate Analysis*, 170:232–243. Special Issue on Functional Data Analysis and Related Topics. [MR3913038](#)

- Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components. *Journal of the Royal Statistical Society, Ser. B*, 68:109–126. [MR2212577](#)
- Hodges, J. and Lehmann, E. (1954). Testing the approximate validity of statistical hypotheses. *Journal of the Royal Statistical Society, Ser. B*, 16(2):261–268. [MR0069461](#)
- Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer Series in Statistics, New York. [MR2920735](#)
- Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to linear Operators*. Wiley, New York. [MR3379106](#)
- König, H. (1986). *Eigenvalue Distribution of Compact Operators*. Birkhäuser, Basel. [MR0889455](#)
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York. [MR2168993](#)
- Rice, G. and Shum, M. (2019). Inference for the lagged cross-covariance operator between functional time series. *Journal of Time Series Analysis*, 40(5):665–692. [MR3995663](#)
- Shang, H. L. (2014). A survey of functional principal component analysis. *ASTA Advances in Statistical Analysis*, 98:121–142. [MR3254025](#)
- Shang, H. L. (2017). Functional time series forecasting with dynamic updating: An application to intraday particulate matter concentration. *Econometrics and Statistics*, 1:184–200. [MR3669996](#)
- Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, 105(491):1228–1240. [MR2752617](#)
- Stoehr, C., Aston, J., and Kirch, C. (2019). Detecting changes in the covariance structure of functional time series with application to fMRI data. [arXiv:1903.00288](#).
- van der Vaart, A. and Wellner, J. (1996). *Weak convergence and Empirical Processes. With Applications to Statistics*. Springer Series in Statistics, New York. [MR1385671](#)
- Weidmann, J. (1980). *Linear Operators in Hilbert spaces*. Springer, New York. [MR0566954](#)
- Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority*. CRC Press, Boca Raton, FL, second edition. [MR2676002](#)
- Zhang, X. and Shao, X. (2015). Two sample inference for the second-order property of temporally dependent functional data. *Bernoulli*, 21(2):909–929. [MR3338651](#)