Detecting and estimating epidemic changes in dependent functional data *

John A D Aston[†] and Claudia Kirch[‡]

February 22, 2011

Abstract

Change point detection in sequences of functional data is examined where the functional observations are dependent. Of particular interest is the case where the change point is an epidemic change (a change occurs and then the observations return to baseline at a later time). The theoretical properties for various tests for at most one change and epidemic changes are derived with a special focus on power analysis. Estimators of the change point location are derived from the test statistics and theoretical properties investigated.

Keywords: change point test, change point estimator, functional data, dimension reduction, power analysis

AMS Subject Classification 2000: 62H15, 62H12, 62M10

1 Introduction

The statistical analysis of functional data has progressed rapidly over the last few years, leading to the possibility of more complex structures being amenable to such techniques. This is particularly true of the complex correlation structure present within and across many functional observed data, requiring methods that can deal both with internal and external dependencies between the observations. Nonparametric techniques for the analysis of functional data are becoming well established (see Ferraty and Vieu [8] or Horváth and Kokoszka [11] for a good overview), and this paper sets out a nonparametric framework for change point analysis within dependent functional data. This extends the work of Berkes et al. [3] and Aue et al. [2] in the i.i.d case as well as of Hörmann and Kokoszka [10] for weakly dependent data all of them for

*This work as well as the position of the second author was financed by the Stifterverband für die Deutsche Wissenschaft by funds of the Claussen-Simon-trust. The first author was also supported by the Engineering and Physical Sciences Research Council (UK) through the CRiSM programme grant and by the project grant EP/H016856/1, and thanks SAMSI for hosting the author during which some of the work was carried out.

†CRiSM, Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK; J.A.D.Aston@warwick.ac.uk

‡Karlsruhe Institute of Technology (KIT), Institute for Stochastics, Kaiserstr. 89, D – 76133 Karlsruhe, Germany; claudia.kirch@kit.edu

at most one change point (AMOC). In the present paper, a wide class of dependency structures is accounted for and two types of change point alternatives are considered, AMOC and epidemic changes, where the observations having changed return to their original state after some unknown time.

Tests and estimators are usually based on dimension-reduction techniques, where it is important that the change is not orthogonal to the projection subspace (for details see Section 2). Most methodology, including those references given above, chooses this subspace based on estimated principle components assuming a general covariance structure within the functional data.

The choice of estimator for the covariance is critical for the power analysis in detecting the change. In particular, a large enough change will switch the estimated principle components in such a way that the change is no longer orthogonal to the projection subspace making it detectable (cf. Theorems 3.2). This switch occurs even for small changes if the underlying covariance structure of the functional data is flat showing that this method yields good results even and especially for underlying covariance structures that are usually seen as being inappropriate for standard prinicipal components analysis. In addition, the theorems characterize detectable changes in terms of the (unobserved) uncontaminated covariance structure, formalising remarks given in Berkes et al. [3].

The paper proceeds as follows. In Section 2, methods for the detection and estimation of change points for dependent functional observations are derived. These methods are presented using an arbitrary orthonormal projection subspace which allows the same general theory to apply regardless of the subspace projection choice. Possible ways of choosing the projection, including issues associated with estimating these projections from the data are detailed in Section 3. The final section gives the details of the proofs.

2 Change-Point Detection Procedures

In this section we investigate change point detection procedures for a mean change in functional observations $X_i(t), t \in \mathcal{Z}$, $i = 1, \ldots, n$, where $\mathcal{Z}$ is some compact set. This setting for independent (functional) observations with at most one change point (AMOC) was investigated by Berkes et al. [3] and for specific weak dependent processes by Hörmann and Kokoszka [10]. We will also allow for dependency (in time) of the functional observations (using meta-assumptions in order to allow for a very general class of dependency) and additionally to the AMOC-Model consider an epidemic change, where after a certain time the mean changes back.

The AMOC-Model is given by

$$X_i(t) = Y_i(t) + \mu_1(t)1_{\{i \leqslant \vartheta n\}} + \mu_2(t)1_{\{\vartheta n < i \leqslant n\}}, \tag{2.1}$$

where the mean functions before and after the change $\boldsymbol{\mu}_j = \mu_j(\cdot)$ as well as the functional time series $\{Y_i(\cdot) : 1 \leqslant i \leqslant n\}$ are elements of $L^2(\mathcal{Z})$, that are ($a.s.$) continuous, $0 < \vartheta \leqslant 1$ describes the position of the change, $\operatorname{E} Y_i(t) = 0$. $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ as well as ϑ are unknown.

The epidemic model is given by

$$X_i(t) = Y_i(t) + \mu_1(t) + (\mu_2(t) - \mu_1(t))1_{\{\vartheta_1 n < i \leqslant \vartheta_2 n\}}, \tag{2.2}$$

where $\boldsymbol{\mu}_j$ and $\{Y_i(\cdot) : 1 \leqslant i \leqslant n\}$ are as above, $0 < \vartheta_1 \leqslant 1$ marks the beginning of the epidemic change, while $\vartheta_1 \leqslant \vartheta_2 \leqslant 1$ marks the end of the epidemic change. $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ as well as ϑ_1, ϑ_2 are unknown.

We begin by considering the corresponding testing problems. Estimators for the point of changes ϑ (resp. ϑ_1, ϑ_2) are related to the test statistics and will be considered in Section 2.2.

2.1 Testing Problem

We are interested in testing the null hypothesis of no change in the mean

$$H_0 : \mathrm{E}\, X_i(\cdot) = \mu_1(\cdot), \quad i = 1, \ldots, n,$$

versus the AMOC alternative

$$\begin{aligned} H_1^{(A)} :& \mathrm{E}\, X_1(\cdot) = \mu_1(\cdot), \quad i = 1, \ldots, \lfloor \vartheta n \rfloor, \quad \text{but} \\ & \mathrm{E}\, X_1(\cdot) = \mu_2(\cdot) \neq \mu_1(\cdot), \quad i = \lfloor n\vartheta \rfloor + 1, \ldots, n, \quad 0 < \vartheta < 1 \end{aligned}$$

respectively versus the epidemic change alternative

$$\begin{aligned} H_1^{(B)} :& \mathrm{E}\, X_1(\cdot) = \mu_1(\cdot), \quad i = 1, \ldots, \lfloor \vartheta_1 n \rfloor, \lfloor \vartheta_2 n \rfloor + 1, \ldots, n, \quad \text{but} \\ & \mathrm{E}\, X_1(\cdot) = \mu_2(\cdot) \neq \mu_1(\cdot), \quad i = \lfloor n\vartheta_1 \rfloor + 1, \ldots, \lfloor \vartheta_2 n \rfloor, \quad 0 < \vartheta_1 < \vartheta_2 < 1. \end{aligned}$$

Note that the null hypothesis corresponds to the cases where $\vartheta = 1$ (AMOC) resp. $\vartheta_1 = \vartheta_2 = 1$ (epidemic change).

It is well known how to test for mean changes in multivariate observations (cf. e.g. Horváth et al. [12]). However, in a functional setting, respectively for high-dimensional data, this is computationally not feasible anymore. Here, the idea is to use a projection into a lower dimensional space and use standard change point statistics for the projected data. In Section 3 we discuss the standard approach using estimated principle components, which was also used by Berkes et al. [3]. In fact even if the first d principle components were known it would still be preferable to use estimated principle components because this leads to a better detectability of changes as Theorem 3.2 b) shows. A discussion of different choices of a subspace as well as different estimation procedures can be found in Aston and Kirch [1].

Let the projection subspace of your choice be spanned by the orthonormal system $\{v_j(\cdot), j = 1, \ldots, d\}$. We assume, that we can estimate v_j by $\widehat{v}_j$ up to a sign. It is possible to weaken this assumption as long as the subspace can consistently be estimated, however the following proofs and Assumption $\mathcal{P}.2$ have to be strengthened to allow for triangular arrays.

Assumption $\mathcal{ON}$. 1. Let $\{v_k(\cdot), k = 1, \ldots, d\}$ and $\{\widehat{v}_k(\cdot), k = 1, \ldots, d\}$ be orthonormal systems. $\widehat{v}_k(\cdot)$, $k = 1, \ldots, d$, are estimators which fulfill under H_0

$$\int (\widehat{v}_k(t) - s_k v_k(t))^2 \, dt = O_P(n^{-1}),$$

where $s_k = \operatorname{sgn}\left(\int v_k(t)\widehat{v}_k(t)\, dt\right)$.

Under alternatives one can not generally expect that the estimator $\widehat{v}_k$ still converges to v_k, however it frequently stabilizes by converging to a different orthonormal system $\{w_k(\cdot) : k = 1, \ldots, d\}$ (cf. Assumption $\mathcal{ON}$.2 and Section 3). Depending on $\{w_k\}$ this may even be a feature not a drawback as it may lead to a better detectability of the change point. Theorem 3.2 shows that this is for example the case if one uses estimated principle components based on the below nonparametric covariance estimator.

We are now ready to explain the main idea of the testing procedure. Denote by $\widehat{\eta}_{i,l}$ the estimated scores, i.e. the projection coefficients of the estimated orthonormal system. To elaborate

$$\widehat{\eta}_{i,l} = \langle X_i, \widehat{v}_l \rangle = \int X_i(t)\widehat{v}_l(t)\,dt, \quad i = 1,\dots,n, \quad l = 1,\dots,d.$$

Then, $\widehat{\boldsymbol{\eta}}_i = (\widehat{\eta}_{i,1},\dots,\widehat{\eta}_{i,d})^T$ is a d-dimensional time series exhibiting the same type of change as the functional sequence $\{X_i(\cdot) : 1 \leqslant i \leqslant n\}$ if the change is not orthogonal to the subspace spanned by $\widehat{v}_1(\cdot),\dots,\widehat{v}_d(\cdot)$. To see this, let

$$\check{\eta}_{i,l} = \langle Y_i, \widehat{v}_l \rangle = \int Y_i(t)\widehat{v}_l(t)\,dt. \tag{2.3}$$

Then it holds

$$\widehat{\eta}_{i,l} = \check{\eta}_{i,l} + 1_{\{i \leqslant \vartheta n\}} \int \mu_1(t)\widehat{v}_l(t)\,dt + 1_{\{i > \vartheta n\}} \int \mu_2(t)\widehat{v}_l(t)\,dt \tag{2.4}$$

in case of AMOC change and an analogous expression for the epidemic change. Consequently, a change is present in the projected data if

$$\int \Delta(t)\widehat{v}_l(t)\,dt \neq 0, \qquad \Delta(t) = \mu_1(t) - \mu_2(t), \qquad \text{for some } l = 1,\dots,d.$$

This representation suggests to use multivariate change point statistics based on $\widehat{\boldsymbol{\eta}}_i$, $i = 1,\dots,n$, which are usually based on

$$\frac{1}{\sqrt{n}}\mathbf{S}_n(k) = \frac{1}{\sqrt{n}}\left(\sum_{1 \leqslant i \leqslant k} \widehat{\boldsymbol{\eta}}_i - \frac{k}{n}\sum_{i=1}^{n} \widehat{\boldsymbol{\eta}}_i\right). \tag{2.5}$$

In fact, if we use an orthonormal basis, i.e. the complete space instead of a subspace, we obtain the Hilbert space analogue of the classic CUSUM change point statistic regardless of the choice of basis, since by $X_i = \sum_{l \geqslant 1} \widehat{\eta}_{i,l}\widehat{v}_l$ and Parsevals identity

$$\max_k \frac{1}{n} \sum_{l \geqslant 1} \left(\sum_{i=1}^{k} \left(\widehat{\eta}_{i,l} - \bar{\widehat{\eta}}_l\right)\right)^2 = \max_k \frac{1}{n} \left\| \sum_{i=1}^{k} (X_i(t) - \bar{X}_i(t)) \right\|^2, \tag{2.6}$$

where $\|\cdot\|$ is the L^2-norm.

In order to get limit theorems for $\mathbf{S}_n$ we need to impose certain assumptions on $\{Y_i(\cdot)\}$ as well as the true scores

$$\eta_{i,l} = \int Y_i(t)v_l(t)\,dt \quad i = 1,\dots,n, \quad l = 1,2,\dots. \tag{2.7}$$

Assumption $\mathcal{P}$. 1. The time series $\{Y_i(\cdot) : i \geqslant 1\}$ is centered, stationary and ergodic with

$$\mathrm{E}\,\|Y_1(\cdot)\|^2 = \int \mathrm{E}(Y_1^2(t))\,dt < \infty.$$

Assumption $\mathcal{P}$. 2. Consider $\eta_{i,l}$ as in (2.7), $l = 1,\dots,d$, as well as $\boldsymbol{\eta}_i = (\eta_{i,1},\dots,\eta_{i,d})^T$.

a) The time series $\{\boldsymbol{\eta}_i : i \in \mathbb{Z}\}$ is stationary and short-range dependent i.e.

$$\sum_{i \in \mathbb{Z}} |\operatorname{cov}(\eta_{0,l_1}, \eta_{i,l_2})| < \infty, \quad l_1, l_2 = 1,\dots,d.$$

b) $\{\boldsymbol{\eta}_i\}$ fulfills the following functional limit theorem

$$\left\{ \frac{1}{\sqrt{n}} \sum_{1 \leqslant i \leqslant nx} \boldsymbol{\eta}_i \,:\, 0 \leqslant x \leqslant 1 \right\} \stackrel{D^d[0,1]}{\longrightarrow} \{\boldsymbol{W}_d(x) : 0 \leqslant x \leqslant 1\},$$

where $\boldsymbol{W}_d$ is a d-dimensional Wiener process with positive-definite covariance matrix $\Sigma = \sum_{k \in \mathbb{Z}} \Gamma(k)$, $\Gamma(j) = \mathrm{E}\, \boldsymbol{\eta}_t \boldsymbol{\eta}_{t+h}^T$, $h \geqslant 0$, and $\Gamma(h) = \Gamma(-h)^T$ for $h < 0$.

Assumption $\mathcal{P}.2$ is fulfilled for a large class of functional time series and allows an easy extension of our results to different dependency concepts. Remark 2.1 shows, for instance, the validity of Assumption $\mathcal{P}.2$ in case of strong mixing resp. $L^p - m$-approximable sequences, a weak dependency concept recently introduced to functional data by Hörmann and Kokoszka [10].

Definition 2.1. *A stationary process $\{Y_j : j \in \mathbb{Z}\}$ is called strong mixing with mixing rate r_m if*

$$\sup_{A,B} |P(A \cap B) - P(A)P(B)| = O(r_m), \qquad r_m \to 0,$$

where the supremum is taken over all $A \in \mathcal{A}(Y_0, Y_{-1}, \ldots)$ and $B \in \mathcal{A}(Y_m, Y_{m+1}, \ldots)$.

Definition 2.2. *A stationary (Hilbert-space valued) process $\{Y_j : j \in \mathbb{Z}\} \in L_H^p$ is called $L^p - m$-approximable if $Y_j = f(\epsilon_j, \epsilon_{j-1}, \ldots)$, where ϵ_i are i.i.d., f is measurable and*

$$\sum_{m \geqslant 1} \left(\mathrm{E}\, \|Y_m - Y_m^{(m)}\|^p \right)^{1/p} < \infty,$$

where $Y_j^{(m)} = f(\epsilon_j, \ldots, \epsilon_{j-m+1}, \epsilon'_{j-m}, \epsilon'_{j-m+1}, \ldots)$ and $\{\epsilon'_j\}$ is an independent copy of $\{\epsilon_j\}$.

Remark 2.1. a) If $\{Y_j(\cdot)\}$ is strong mixing then $\{\eta_{i,j} : i = 1, \ldots, n\}$ is strong-mixing with the same rate (cf. Proposition 10.4 in Ferraty and Vieu [8]). Under certain moment conditions in combination with conditions on the mixing rate Davydovs covariance inequality (cf. e.g. Lemma 2.1 in Kuelbs and Philipp [15]) yields $\mathcal{P}.2$ a). $\mathcal{P}.2$ b) can for example be derived from strong invariance principles under similar conditions as given in Kuelbs and Philipp [15].

b) If $\{Y_j(\cdot)\}$ is $L^p - m$-approximable, then $\{\boldsymbol{\eta}_j\}$ is also $L^p - m$-approximable and hence $\mathcal{P}.2$ a) holds if $p = 2$ (cf. Hörmann and Kokoszka [10], Theorem 4.2 and comments below). In this situation $\mathcal{P}.2$ b) follows from Theorem A.2 in Hörmann and Kokoszka [10].

The following lemma gives the null asymptotics in $D[0,1]$ for the process $\mathbf{S}_n(\cdot)$. From this we can easily obtain the null asymptotics of various popular test statistics in our main Theorem 2.1.

Lemma 2.1. *Let Assumptions $\mathcal{ON}.1$ as well as $\mathcal{P}.1$ – $\mathcal{P}.2$ be fulfilled. Then under H_0 it holds*

$$\left\{ \frac{1}{\sqrt{n}} \widetilde{\mathbf{S}}_n(x) : 0 \leqslant x \leqslant 1 \right\} \stackrel{D^d[0,1]}{\longrightarrow} \left\{ \Sigma^{1/2} \mathbf{B}_d(x) : 0 \leqslant x \leqslant 1 \right\},$$

where $\widetilde{\mathbf{S}}_n(x) = (\widetilde{S}_{n,1}(x), \ldots, \widetilde{S}_{n,d}(x))^T$, $s_l = sgn(\int v_k(t) \widehat{v}_k(t)\, dt)$ and

$$\widetilde{S}_{n,l}(x) = s_l S_{n,l}(x) = s_l \sum_{1 \leqslant i \leqslant nx} \left(\widehat{\eta}_{i,l} - \frac{1}{n} \sum_{i=1}^{n} \widehat{\eta}_{i,l} \right), \quad l = 1, \ldots, d. \tag{2.8}$$

Σ is as in Assumption $\mathcal{P}.2$ and $\mathbf{B}_d$ is a standard d-dimensional Brownian bridge.

Remark 2.2. The proof shows that the result remains valid if the rate in Assumption $\mathcal{ON}.1$ is replaced by $o_P(1)$ as well as $\mathcal{P}.1$ by $\sup_{0<x<1}\int\left(\frac{1}{\sqrt{n}}\sum_{1\leqslant i\leqslant nx}Y_i(t)\right)^2 = O_P(1)$. The latter one follows for example from functional central limit theorems for the Banach space valued random variables $\{Y_i(\cdot)\}$.

Since it is not possible to estimate the sign s_l, test statistics should be based on $(S_{n,l}(\cdot))^2$ as $s_l^2=1$. The next theorem gives the null asymptotics for popular statistics for the AMOC-change alternative. Analogous results for weighted versions of the statistics can also be obtained immediately from Lemma 2.1.

Theorem 2.1. *Let the assumptions of Lemma 2.1 hold. Furthermore let $\widehat{\Sigma}$ be a consistent symmetric positive-definite estimator for Σ and $B_l(\cdot)$, $l=1,\ldots,d$, be independent standard Brownian bridges.*

a) The following statistics are suitable to detect AMOC-change alternatives:

$$T_n^{(A1)} = \frac{1}{n^2}\sum_{k=1}^{n}\mathbf{S}_n\left(k/n\right)^T\widehat{\Sigma}^{-1}\mathbf{S}_n\left(k/n\right),$$

$$T_n^{(A2)} = \max_{1\leqslant k\leqslant n}\frac{1}{n}\mathbf{S}_n\left(k/n\right)^T\widehat{\Sigma}^{-1}\mathbf{S}_n\left(k/n\right),$$

$$\text{where } \mathbf{S}_n(x) = \sum_{1\leqslant j\leqslant nx}\left(\widehat{\boldsymbol{\eta}}_j - \frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{\eta}}_i\right).$$

Under H_0 it holds:

$$(i)\quad T_n^{(A1)} \xrightarrow{\mathcal{L}} \sum_{1\leqslant l\leqslant d}\int_0^1 B_l^2(x)\,dx.$$

$$(ii)\quad T_n^{(A2)} \xrightarrow{\mathcal{L}} \sup_{0\leqslant x\leqslant 1}\sum_{1\leqslant l\leqslant d}B_l^2(x).$$

b) The following statistics are suitable to detect epidemic change alternatives.

$$T_n^{(B1)} = \frac{1}{n^3}\sum_{1\leqslant k_1<k_2\leqslant n}\mathbf{S}_n\left(k_1/n,k_2/n\right)^T\widehat{\Sigma}^{-1}\mathbf{S}_n\left(k_1/n,k_2/n\right),$$

$$T_n^{(B2)} = \max_{1\leqslant k_1<k_2\leqslant n}\frac{1}{n}\mathbf{S}_n\left(k_1/n,k_2/n\right)^T\widehat{\Sigma}^{-1}\mathbf{S}_n\left(k_1/n,k_2/n\right),$$

$$\text{where } \mathbf{S}_n(x,y) = \mathbf{S}_n(y)-\mathbf{S}_n(x) = \sum_{nx<j\leqslant ny}\left(\widehat{\boldsymbol{\eta}}_j - \frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{\eta}}_i\right).$$

Under H_0 it holds:

$$(i)\quad T_n^{(B1)} \xrightarrow{\mathcal{L}} \sum_{1\leqslant l\leqslant d}\int\int_{0\leqslant x<y\leqslant 1}(B_l(x)-B_l(y))^2\,dx\,dy.$$

$$(ii)\quad T_n^{(B2)} \xrightarrow{\mathcal{L}} \sup_{0\leqslant x<y\leqslant 1}\sum_{1\leqslant l\leqslant d}(B_l(x)-B_l(y))^2.$$

Remark 2.3. For the above test statistics estimators of the long-run covariance matrix Σ are needed. Usually, estimators are of the following type:

$$\widehat{\Sigma} = \sum_{|h|\leqslant b_n} w_q(h/b_n)\widehat{\Gamma}(h),$$

for some appropriate weight function w_q and bandwidth b_n where

$$\widehat{\Gamma}(h) = \frac{1}{n}\sum_{j=1}^{n-h}\widehat{\boldsymbol{\eta}}_j\widehat{\boldsymbol{\eta}}_{j+h}^T,\quad h\geqslant 0,\quad \widehat{\Gamma}(h)=\widehat{\Gamma}(-h),\quad h<0.$$

Hörmann and Kokoszka [10] prove consistency of this estimator for weakly dependent data. Politis [16] proposed to use different bandwidths for each entry of the matrix in addition to an automatic bandwidth selection procedure for the class of flat-top weight functions. Generally, in change point analysis it is advisable to adapt the estimators to take a possible change point into account to improve the power of the test. For details in the univariate situation we refer to Hušková and Kirch [13]. More details on problems and solutions for this estimator in the present situation for real data can be found in Aston and Kirch [1].

Now, we turn to the behaviour of the test under the alternative hypothesis. In this case, the estimators $\widehat{v}_k$ cannot in general be expected to converge to $\pm v_k$ anymore, however the following stability assumption is frequently fulfilled (cf. Section 3).

Assumption $\mathcal{ON}$. 2. Let $\{w_k(\cdot), k = 1, \ldots, d\}$ be an orthonormal system, $\{\widehat{v}_k(\cdot), k = 1, \ldots, d\}$ the same estimators as before. Under the alternative it holds

$$\int (\widehat{v}_k(t) - s_k w_k(t))^2 \, dt = o_P(1),$$

where $s_k = \operatorname{sgn}\left(\int w_k(t)\widehat{v}_k(t)\, dt\right)$, i.e. the estimators converge to some contaminated ON-System. Note that w_k usually depends on the type of alternative. For clarity we sometimes write $w_{k,A}$ in case of an AMOC alternative and $w_{k,B}$ in case of an epidemic change alternative.

Lemma 2.2. *Let Assumptions $\mathcal{ON}$.2 and $\mathcal{P}$.1 hold.*

a) Under the AMOC-alternative $H_1^{(A)}$, it holds

$$\sup_{0\leqslant x\leqslant 1}\left|\frac{1}{n}\sum_{1\leqslant i\leqslant nx}\widehat{\eta}_{i,l} - \frac{\lfloor nx\rfloor}{n^2}\sum_{i=1}^{n}\widehat{\eta}_{i,l} - s_l\int \Delta(t) w_{l,A}(t)\, dt \; g_A(x)\right| \xrightarrow{P} 0,$$

where

$$g_A(x) = \begin{cases} x(1-\vartheta), & 0 < x \leqslant \vartheta, \\ \vartheta(1-x), & \vartheta < x < 1, \end{cases}$$

$\Delta(t) = \mu_1(t) - \mu_2(t)$ and $w_{l,A}$ are as in $\mathcal{ON}$.2.

b) Under the epidemic change alternative $H_1^{(B)}$, it holds

$$\sup_{0\leqslant x<y\leqslant 1}\left|\frac{1}{n}\sum_{\lfloor nx\rfloor < i\leqslant ny}\widehat{\eta}_{i,l} - \frac{\lfloor ny\rfloor - \lfloor nx\rfloor}{n^2}\sum_{i=1}^{n}\widehat{\eta}_{i,l} - s_l\int \Delta(t) w_{l,B}(t)\, dt\, g_B(x,y)\right| \xrightarrow{P} 0,$$

where $g_B(x,y) = g_B(y) - g_B(x)$ and

$$g_B(x) = \begin{cases} x(\vartheta_2-\vartheta_1)\int \Delta(t) w_{l,B}(t)\, dt, & x \leqslant \vartheta_1, \\ (\vartheta_1 - x(1-\vartheta_2-\vartheta_1))\int \Delta(t) w_{l,B}(t)\, dt, & \vartheta_1 < x \leqslant \vartheta_2, \\ (x-1)(\vartheta_2-\vartheta_1)\int \Delta(t) w_{l,B}(t)\, dt, & x > \vartheta_2, \end{cases}$$

$\Delta(t) = \mu_1(t) - \mu_2(t)$ and $w_{k,B}(\cdot)$ are as in $\mathcal{ON}$.2.

From the lemma we can conclude that the above tests are consistent in all cases where the change is not orthogonal to the contaminated projection subspace. If one uses the first d estimated principle components a large enough change will switch the projection subspace in such a way that this condition is fulfilled and the change is detectable (cf. Theorem 3.2).

Theorem 2.2. *Let Assumptions $\mathcal{ON}.2$ and $\mathcal{P}.1$ hold, in addition to*

$$\int \Delta(t) w_{k,j}(t)\, dt \neq 0,$$

$\Delta(t) = \mu_1(t) - \mu_2(t)$, for some $k = 1, \ldots, d$ with $j = A$ for a) and $j = B$ for b). For the eigenvalues $\xi_{j,n}$, $j = 1, \ldots, d$, of the estimator $\widehat{\Sigma}$ under the alternative, it holds: $\xi_{j,n} > 0$ as well as $\xi_{j,n} = O_P(1)$.

a) Under the AMOC alternative $H_1^{(A)}$ it holds

$$(i) \quad T_n^{(A1)} \xrightarrow{P} \infty, \qquad (ii) \quad T_n^{(A2)} \xrightarrow{P} \infty.$$

b) Under the epidemic change alternative $H_1^{(B)}$ it holds

$$(i) \quad T_n^{(B1)} \xrightarrow{P} \infty, \qquad (ii) \quad T_n^{(B2)} \xrightarrow{P} \infty.$$

The assumptions on the estimator $\widehat{\Sigma}$ are for example fulfilled if $\widehat{\Sigma} \xrightarrow{P} \Sigma_A$ for some symmetric positive-definite matrix Σ_A.

Remark 2.4. Using Lemma 2.2 one even obtains consistency of the tests for local changes with $\|\Delta_n(\cdot)\| = \|\mu_{1,n}(\cdot) - \mu_{2,n}(\cdot)\| \to 0$ but $\sqrt{n}|\int \Delta_n(t) w_{k,j}(t)\, dt| \to \infty$.

2.2 Estimation of the Change-Point

In this section we consider estimators for the change point ϑ under the AMOC alternative resp. for ϑ_1 and ϑ_2 under the epidemic change alternative and discuss consistency as well as rates of convergence.

First consider the AMOC alternative. Let $\arg\max(a(x) : x) = \min(x : a(x) = \max_y a(y))$ and consider the estimator

$$\widehat{\vartheta} = \arg\max \left(\mathbf{S}_n^T(x) \widehat{\Sigma}^{-1} \mathbf{S}_n(x) : 0 \leqslant x \leqslant 1 \right), \tag{2.9}$$

$$\text{where } \mathbf{S}_n(x) = (S_{n,1}(x), \ldots, S_{n,d}(x))^T, \quad S_{n,l}(x) = \sum_{1 \leqslant i \leqslant nx} \widehat{\eta}_{i,l} - \frac{\lfloor nx \rfloor}{n} \sum_{i=1}^{n} \widehat{\eta}_{i,l}.$$

Consistency of this estimator for i.i.d. observations and a specifically estimated ON-system has been obtained by Berkes et al. [3] and follows immediately from Lemma 2.2. Rates have been obtained by Aue et al. [2] in this situation and their proof can be extended to the dependent situation. To this end we need the following additional assumption.

Assumption $\mathcal{P}$. 3. There exists an increasing sequence $\alpha(n) \to \infty$ such that the contaminated scores $\beta_{i,l} = \int Y_i(t) w_l(t)\, dt$ fulfill a Hájek -Renyi-type inequality:

$$\max_{1 \leqslant k \leqslant n} \frac{\alpha(k)}{k} \left| \sum_{i=1}^{k} \eta_{i,l}^{(j)} \right| = O_P(1), \quad l = 1, \ldots, d.$$

Remark 2.5. Assumption $\mathcal{P}.3$ follows for examples from laws of logarithm, which hold under rather general moment assumptions (cf. Serfling [19]). Under moment and mixing conditions even a law of iterated logarithm can be obtained cf. e.g. Dehling [5].

Theorem 2.3. *Assume that the AMOC model holds. Furthermore, let the assumptions of Theorem 2.2 hold, i.e. the change is detectable, in addition to $\widehat{\Sigma} \xrightarrow{P} \Sigma_A$ for some symmetric positive-definite Σ_A.*

a) Then, the estimator $\widehat{\vartheta}$ is consistent, i.e.

$$\widehat{\vartheta} - \vartheta = o_P(1).$$

b) If additionally Assumption $\mathcal{P}.3$ holds, then

$$n(\widehat{\vartheta} - \vartheta) = O_P(1).$$

Aue et al. [2] additionally obtain the limit distribution of $n(\widehat{\vartheta} - \vartheta)$ in case of i.i.d. data showing that the rate in Theorem 2.3 b) cannot be improved.

In case of an epidemic change alternative we consider the estimator

$$(\widehat{\vartheta}_1, \widehat{\vartheta}_2) = \arg\max\left(\mathbf{S}_n^T(x,y)\widehat{\Sigma}_n^{-1}\mathbf{S}_n(x,y) : 0 \leqslant x < y \leqslant 1\right), \qquad (2.10)$$

$$\text{where } \mathbf{S}_n(x,y) = \mathbf{S}_n(y) - \mathbf{S}_n(x)$$

and $(x_1, y_1) = \arg\max(Z(x,y) : 0 \leqslant x < y \leqslant 1)$ iff $x_1 = \min(0 \leqslant x < 1 : Z(x,y) = \max_{0\leqslant s<t\leqslant 1} Z(s,t)$ for some $y)$ and $y_1 = \max(y > x_1 : Z(x_1,y) = \max_{0\leqslant s<t\leqslant 1} Z(s,t))$.

Theorem 2.4. *Assume that the epidemic change model holds. Furthermore, let the assumptions of Theorem 2.2 hold, i.e. the change is detectable, in addition to $\widehat{\Sigma} \xrightarrow{P} \Sigma_A$ for some symmetric positive-definite Σ_A.*

a) Then, the estimator $(\widehat{\vartheta}_1, \widehat{\vartheta}_2)$ is consistent, i.e.

$$(\widehat{\vartheta}_1 - \vartheta_1, \widehat{\vartheta}_2 - \vartheta_2)^T = o_P(1).$$

b) If additionally Assumption $\mathcal{P}.3$ holds, then

$$n(\widehat{\vartheta}_1 - \vartheta_1, \widehat{\vartheta}_2 - \vartheta_2)^T = O_P(1).$$

3 Principal Component Analysis for Subspace Selection

The procedures described in the previous sections depend heavily upon the choice of a subspace spanned by $\{v_k(\cdot), k = 1, \ldots, d\}$ and even more importantly $\{w_k(\cdot), k = 1, \ldots, d\}$. Precisely, we have seen that changes are detectable if they are not orthogonal to the contaminated subspace spanned by $\{w_k(\cdot), k = 1, \ldots, d\}$. A good combination of choice of subspace $\{v_k(\cdot), k = 1, \ldots, d\}$ and estimation procedure can even have the nice property that the contaminated subspace $\{w_k\}$ differs from $\{v_k\}$ in such a way that the change is now detectable using the contaminated subspace. Theorem 3.2 shows that this is the case if the first principle components are chosen according to the below nonparametric estimation procedure.

Classical dimension reduction techniques are often based on the first d principle components, which choose a subspace explaining most of the variance. To this end consider the (spatial) covariance kernel of $Y_i(\cdot)$ given by

$$c(t,s) = \mathrm{E}(Y_i(t)Y_i(s)). \qquad (3.1)$$

The covariance operator $C : \mathcal{L}^2(\mathcal{Z}) \to \mathcal{L}^2(\mathcal{Z})$ is obtained as $Cz = \int_{\mathcal{Z}} c(\cdot,s)z(s)\,ds$. Due to the stationarity of $\{Y_i(\cdot) : 1 \leqslant i \leqslant n\}$ the covariance kernel does not depend on i and is square integrable due to the Cauchy-Schwarz inequality as well as the square integrability of $Y_1(\cdot)$.

Let $\{\lambda_k\}$ be the non-negative decreasing sequence of eigenvalues and $\{v_k(\cdot) : k \geqslant 1\}$ a given set of corresponding orthonormal eigenfunctions of the covariance operator, i.e. they are defined by

$$\int c(t,s)v_l(s)\,ds = \lambda_l v_l(t), \quad l = 1,2,\ldots, \quad t \in \mathcal{Z}. \tag{3.2}$$

Under the above assumptions, the covariance kernel can be written as

$$c(t,s) = \sum_{k=1}^{\infty} \lambda_k v_k(t) v_k(s),$$

and more importantly $Y_i(\cdot)$ can be expressed in terms of the eigenfunctions

$$Y_i(t) = \sum_{l=1}^{\infty} \eta_{i,l} v_l(t), \tag{3.3}$$

where $\{\eta_{i,l} : l = 1,2,\ldots\}$ are uncorrelated with mean 0 and variance λ_l for each i. This property is useful for the above analysis in case of independent data, because then due to the diagonal structure of Σ estimation becomes easier. Also, the test statistics considered above simplify. Unfortunately, for dependent functional data this is no longer true in general as the long-run covariance can be different from zero even if η_{i,l_1} and η_{i,l_2} are uncorrelated for any i.

The infinite sum on the right-hand side converges in $L^2(\mathcal{Z})$ with probability one. Furthermore

$$\eta_{i,l} = \int Y_i(t) v_l(t)\,dt \quad i = 1,\ldots,n, \quad l = 1,2,\ldots. \tag{3.4}$$

More details can for example be found in either Bosq [4] or Horváth and Kokoszka [11].

One may now choose the d eigenfunctions $v_l(\cdot)$, $l = 1,\ldots,d$, belonging to the largest d eigenvalues as a basis for the ON-System needed for the change point procedure. This generates a basis that explains the largest variation of the data of any subspace of size d. ERASE: This variation may or may not include the variation due to any change point present, depending on the size of the change (if the alternative holds) and whether the change is orthogonal to the uncontaminated eigenspace (cf Theorem 3.2).

In practice, the covariance kernel $c(t,s)$ is usually not known but needs to be estimated. The estimators need to be consistent, i.e.

Assumption $\mathcal{C}$. 1. a) Under H_0 the estimated covariance kernel $\widehat{c}_n(t,s)$ is a consistent estimator for the covariance kernel $c(t,s)$ of $\{Y_1(\cdot)\}$ with convergence rate $\sqrt{n}$, i.e.

$$\int\int (\widehat{c}_n(t,s) - c(t,s))^2\,dt\,ds = O_P(n^{-1}).$$

b) The largest $d+1$ eigenvalues corresponding to c satisfy

$$\lambda_1 > \lambda_2 > \ldots > \lambda_d > \lambda_{d+1} > 0.$$

Assumption $\mathcal{C}$.1 b) ensures that the (normalized) eigenfunctions $v_k(\cdot)$ belonging to the first d eigenvectors are identifiable up to their sign.

A natural estimator in a general non-parametric setting is the empirical version of the covariance function

$$\widehat{c}_n(t,s) = \frac{1}{n}\sum_{i=1}^{n}(X_i(t) - \bar{X}_n(t))(X_i(s) - \bar{X}_n(s)), \tag{3.5}$$

where $\bar{X}_n(t) = \frac{1}{n}\sum_{i=1}^{n} X_i(t)$. In case of independent functional observations and for an AMOC change alternative Berkes et al. [3] proved $\mathcal{C}.1$ a). It will be seen in Lemma 3.1 that this also remains true in more general situations.

It is often not possible to find estimators in such a way that they are still consistent under alternatives. However, it is sufficient if they consistently estimate a contaminated covariance kernel $k(t,s)$. In fact, in case of a principle component analysis this is even desirable as it leads to a subspace that will include the change if it is large enough (cf. Theorem 3.2). This is due to the fact that the variability of the empirical covariance function increases where a change is present proportional to the size of the change (cf. (3.6)).

Assumption $\mathcal{C}$. 2. a) Under alternatives H_1 there exists a covariance function $k(t,s)$, such that

$$\int\int (\widehat{c}_n(t,s) - k(t,s))^2 \, dt \, ds \xrightarrow{P} 0.$$

b) The largest $d+1$ contaminated eigenvalues, i.e. the eigenvalues γ_k belonging to $k(t,s)$, fulfill

$$\gamma_1 > \gamma_2 > \ldots > \gamma_d > \gamma_{d+1} > 0.$$

Usually the contaminated covariance function $k(t,s)$ as well as the contaminated eigenvalues γ_k will depend on the type and shape of the change (cf. (3.6)).

As under H_0 Assumption $\mathcal{C}.2$ b) ensures that the (normalised) contaminated eigenfunctions $w_k(\cdot)$ are sign-identifiable. Furthermore it guarantees that the corresponding estimated eigenfunctions correctly estimate the true eigenfunctions (up to the sign). In both situations the assumption makes proofs easier and more understandable, but can be relaxed if the eigenspace can be properly estimated and one works with distributional convergence of triangular arrays instead. It is, however, substantial that $\lambda_d > \lambda_{d+1}$ and $\gamma_d > \gamma_{d+1}$ to have clearly defined eigenspaces.

Theorem 3.1. *Let $\widehat{\lambda}_k$ and $\widehat{v}_k(\cdot)$ be the estimated eigenvalues and eigenfunctions, i.e. the eigenvalues resp. eigenfunctions of $\widehat{c}_n(t,s)$.*

a) Under the null hypothesis and Assumption $\mathcal{C}.1$ it holds for $j = 1, \ldots, d$

$$|\widehat{\lambda}_j - \lambda_j| = O_P(n^{-1/2})$$
$$\int (\widehat{v}_j(t) - s_j v_j(t))^2 \, dt = O_P(n^{-1}),$$

where $v_j(\cdot)$ is an orthonormal set of eigenfunctions up to a sign defined by (3.2) and $s_j = sgn(\int v_j(t)\widehat{v}_j(t)\,dt)$ gives that sign. In particular, Assumption $\mathcal{ON}.1$ holds true.

b) Under alternatives as well as $\mathcal{C}.2$ it holds, $j = 1, \ldots, d$

$$|\widehat{\lambda}_j - \gamma_j| \xrightarrow{P} 0$$
$$\int (\widehat{v}_j(t) - s_j w_j(t))^2 \, dt \xrightarrow{P} 0,$$

where $w_j(\cdot)$, is an orthonormal set of eigenfunctions of $k(t,s)$ up to a sign defined by (3.2) and $s_j = sgn(\int w_j(t)\widehat{v}_j(t)\,dt)$ gives that sign. In particular, Assumption $\mathcal{ON}.2$ holds true.

The following lemma shows that the above remains true even in more general dependence situations.

Lemma 3.1. *a) If $Y_i(\cdot)$ is L^4-m-approximable or strong mixing with mixing rate α_j, $\mathrm{E}\,\|Y_1(\cdot)\|^{4+\delta}<\infty$ and $\sum_{h\geqslant 1}\alpha_h^{\frac{\delta}{4+\delta}}<\infty$, then the rate in $\mathcal{C}.1$ a) holds.*

b) If $\{Y_i(\cdot): i\geqslant 1\}$ fulfills Assumption $\mathcal{P}.1$ and $\widehat{c}_n(t,s)$ is as in (3.5) then Assumption $\mathcal{C}.2$ a) holds with

$$k(t,s)=c(t,s)+\theta(1-\theta)\Delta(t)\Delta(s), \tag{3.6}$$

where

$$\Delta(t)=\mu_1(t)-\mu_2(t),$$
$$\theta=\begin{cases}\vartheta, & AMOC,\\ \vartheta_2-\vartheta_1, & \text{epidemic change.}\end{cases}$$

Furthermore, under H_0, Assumption $\mathcal{C}.1$ holds but without the rate.

In Section 2 we have seen that the tests have asymptotic power one and the estimators are consistent if the change $\Delta(\cdot)$ is not orthogonal to the contaminated projection subspace, which depends directly on both the change point and the change itself. The following theorem allows a characterisation of detectable changes in terms of the non-contaminated projection subspace and even more importantly shows that the change has a tendency to influence the contaminated projection subspace in such a way that it becomes detectable.

Theorem 3.2. *a) Let $\{v_l, l=1,\ldots,d\}$ be the eigenfunctions belonging to the largest d eigenvalues of c as well as $\{w_l, l=1,\ldots,d\}$ the eigenfunctions belonging to the largest d eigenvalues of k as in (3.6).*
For $\Delta(t)=\mu_1(t)-\mu_2(t)$ it holds

$$\int\Delta(t)v_l(t)\,dt\neq 0 \text{ for some } l=1,\ldots,d$$
$$\Longrightarrow\int\Delta(t)w_l(t)\,dt\neq 0 \text{ for some } l=1,\ldots,d.$$

This shows, that any change that is not orthogonal to the non-contaminated subspace is detectable.

b) Let $\Delta_D(t)=D\,\Delta(t)$, $\int\Delta^2(t)\,dt\neq 0$. Then, there exists $D_0>0$ such that

$$\int\Delta_D(t)w_{1,D}(t)\,dt\neq 0$$

for all $|D|\geqslant D_0$, where $w_{1,D}$ is the eigenfunction belonging to the largest eigenvalue of the contaminated covariance kernel $k_D(t,s)=c(t,s)+\theta(1-\theta)\Delta_D(t)\Delta_D(s)$. This shows, that any large enough change is detectable.

Remark 3.1. Theorem 3.2 a) shows that we are able to detect at least all changes that are not orthogonal to the non-contaminated subspace spanned by the first d principle components. Part b) shows that frequently changes can be detected even if they are orthogonal to the non-contaminated ON-System. The reason is that large mean changes lead to a larger variability of the empirical covariance function and thus the contaminated covariance function $k(t,s)=c(t,s)+\theta(1-\theta)\Delta(t)\Delta(s)$ in the components that are not orthogonal to the change, while not changing the variability in the components that are orthogonal. In the following example such a change in the subspace takes place: Let $\{b_j: j\geqslant 1\}$ be an orthonormal basis of the continuous functions on $\mathcal{Z}$. Furthermore X,Y are i.i.d. $N(0,1)$, and $Y(t)=2Xb_1(t)+Yb_2(t)$.

Obviously $c(t,s)$ has the eigenvalues 4 with eigenfunction b_1 as well as the eigenvalue 1 with eigenfunction b_2 in addition to the eigenvalue 0. As in Lemma 3.1 let $\theta = 1/2$ and consider $\Delta(t) = 4b_2(t)$ which for $d = 1$ is obviously orthogonal to b_1, but it is easy to see that the eigenvalues of $k(t,s)$, are now 5 corresponding to b_2 and 4 corresponding to b_1 in addition to the eigenvalue 0. This shows that the mean change is no longer orthogonal to the space spanned by the eigenfunction corresponding to the largest eigenvalue, which is the one spanned by b_2.

An immediate corollary to Theorem 3.2 also gives rise to a surprising fact for multivariate data. PCA is well known to work poorly as a representation of the data when the covariance matrix of the multivariate observations is close to a multiple of the identity matrix. In fact, the scree plot will be linear in nature in the case when the covariance is an exact multiple of the identity implying that there is no effective sparse representation of the data. As a contrast, by the proof of the theorem above, it is optimal for detecting a change point if the uncontaminated covariance matrix is a multiple of the identity matrix. In this case, choosing only a single principal component from the contaminated covariance will guarantee that the power of detection is asymptotically one. Thus PCA based change point detection (for either epidemic or AMOC) works best when PCA itself works worst for the uncontaminated system regardless of the direction of the change.

This fact also translates over to functional data, but because the eigenvalues are square summable, the degenerate case will not occur. However, in situations where the eigenvalues decay very rapidly in the uncontaminated case, changes orthogonal to the eigenfunctions corresponding to the largest uncontaminated eigenvalues will be required to be bigger if they are supposed to be detectable than in situations with more slowly decreasing eigenvalues.

4 Proofs

4.1 Proofs of Section 2

Most of the proofs in this section follow the ideas of proofs given in either Berkes et al. [3] (for the proofs of Subsection 2.1) or Aue et al. [2] (for the proofs of Subsection 2.2) for AMOC situations in the simpler situation of i.i.d. functional data using a subspace obtained from principle components analysis, which allows to consider only the simpler situation, where $\widehat{\Sigma}$ is a diagonal matrix.

Proof of Lemma 2.1. First note that under H_0

$$\widehat{\eta}_{i,l} - \bar{\widehat{\eta}}_l = \check{\eta}_{i,l} - \bar{\check{\eta}}_l,$$

where (e.g.) $\bar{\widehat{\eta}}_l = \frac{1}{n}\sum_{i=1}^n \widehat{\eta}_{i,l}$. Furthermore

$$\begin{aligned}
&\sup_{0<x<1}\left|\frac{1}{\sqrt{n}}\sum_{1\leqslant i\leqslant nx} s_l\check{\eta}_{i,l} - \frac{1}{\sqrt{n}}\sum_{1\leqslant i\leqslant nx}\eta_{i,l}\right| \\
&= \sup_{0<x<1}\left|\int\left(\frac{1}{n}\sum_{1\leqslant i\leqslant nx} Y_i(t)\right)\sqrt{n}\,(s_l\widehat{v}_l(t) - v_l(t))\,dt\right| \\
&\leqslant \sup_{0<x<1}\left(\int\left(\frac{1}{n}\sum_{1\leqslant i\leqslant nx} Y_i(t)\right)^2\right)^{1/2}\left(n\int (s_l\widehat{v}_l(t) - v_l(t))^2\right)^{1/2} = o_P(1).
\end{aligned}$$

The last line follows since by ergodicity and stationarity the following law of large numbers holds (cf. e.g. Ranga Rao [18])

$$\left\| \frac{1}{n} \sum_{i=1}^{n} Y_i(\cdot) \right\|_{L^2(\mathcal{Z})} \to 0 \qquad a.s., \tag{4.1}$$

hence by standard arguments

$$\sup_{0<x<1} \int \left(\frac{1}{n} \sum_{1 \leqslant i \leqslant nx} Y_i(t) \right)^2 dt = o_P(1). \tag{4.2}$$

The second factor is $O_P(1)$ by $\mathcal{ON}.1$ and the fact that $s_l^2 = 1, s_l^{-1} = s_l$.

The assertion now follows from Assumption $\mathcal{P}.2$. ■

Proof of Theorem 2.1. The assertions of the theorem follow immediately from Lemma 2.1 and the fact that $s_l^2 = 1$. ■

Proof of Lemma 2.2. Concerning a), by (2.4) and (4.2) it holds uniformly in $x \leqslant \vartheta$

$$\begin{aligned}
&\frac{1}{n} \sum_{1 \leqslant i \leqslant nx} \widehat{\eta}_{i,l} - \frac{\lfloor nx \rfloor}{n^2} \sum_{i=1}^{n} \widehat{\eta}_{i,l} \\
&= x(1-\vartheta) \left(\int \mu_1(t) \widehat{v}_l(t)\, dt - \int \mu_2(t) \widehat{v}_l(t)\, dt \right) + o_P(1) \\
&= x(1-\vartheta) s_l \int (\mu_1(t) - \mu_2(t)) w_{l,A}(t)\, dt + o_P(1),
\end{aligned}$$

where the last line follows from Assumption $\mathcal{ON}.2$. Analogously one obtains uniformly in $x > \vartheta$

$$\begin{aligned}
&\frac{1}{n} \sum_{1 \leqslant i \leqslant nx} \widehat{\eta}_{i,l} - \frac{\lfloor nx \rfloor}{n^2} \sum_{i=1}^{n} \widehat{\eta}_{i,l} \\
&= \vartheta(1-x) s_l \int (\mu_1(t) - \mu_2(t)) w_{l,A}(t)\, dt + o_P(1),
\end{aligned}$$

which finishes the proof of a).

Concerning b) note that for the epidemic change alternative one gets analogously uniformly in x

$$\begin{aligned}
&\frac{1}{n} \sum_{1 \leqslant i \leqslant nx} \widehat{\eta}_{i,l} - \frac{\lfloor nx \rfloor}{n^2} \sum_{i=1}^{n} \widehat{\eta}_{i,l} \\
&= o_P(1) + \begin{cases} x(\vartheta_2 - \vartheta_1) \int \Delta(t) w_{l,B}(t)\, dt, & x \leqslant \vartheta_1, \\ (\vartheta_1 - x(1 - \vartheta_2 - \vartheta_1)) \int \Delta(t) w_{l,B}(t)\, dt, & \vartheta_1 < x \leqslant \vartheta_2, \\ (x-1)(\vartheta_2 - \vartheta_1) \int \Delta(t) w_{l,B}(t)\, dt, & x > \vartheta_2, \end{cases}
\end{aligned}$$

yielding the assertion. ■

Proof of Theorem 2.2. Consider $\mathbf{d} = (\int \Delta(t) w_{1,A}(t)\, dt, \ldots, \int \Delta(t) w_{d,A}(t)\, dt)^T$ with $\mathbf{d}^T \mathbf{d} > 0$ by assumption. Due to the assumptions on $\widehat{\Sigma}$ there exists a unitary matrix U such that

$$\widehat{\Sigma}^{-1} = U^T \Xi^{-1} U, \qquad \Xi^{-1} = (\xi_{j,n}^{-1} 1_{\{j=l\}})_{j,l=1,\ldots,d}.$$

Moreover, by assumption there exists $c > 0$ such that $P(\min_{j=1,\ldots,d} 1/\xi_{j,n} \geqslant c) \to 1$ and on this set $\mathbf{d}^T\widehat{\Sigma}^{-1}\mathbf{d} \geqslant c\,\mathbf{d}^T\mathbf{d}$, hence with $D = c\mathbf{d}^T\mathbf{d} > 0$

$$P\left(\mathbf{d}^T\widehat{\Sigma}^{-1}\mathbf{d} \geqslant D\right) = 1 + o(1).$$

By Lemma 2.2 we obtain

$$T_n^{(A1)} = n\left(\mathbf{d}^T\widehat{\Sigma}^{-1}\mathbf{d}\left(\int_0^1 g_A^2(x)\,dx + o(1)\right) + o_P(1)\right) \xrightarrow{P} \infty$$

and analogously $T_n^{(B1)} \xrightarrow{P} \infty$. Furthermore

$$T_n^{(A2)} \geqslant n\left(\mathbf{d}^T\widehat{\Sigma}^{-1}\mathbf{d}\,g_A^2(\vartheta) + o_P(1)\right) \xrightarrow{P} \infty$$

and analogously $T_n^{(B2)} \xrightarrow{P} \infty$. ■

Proof of Theorem 2.3. Lemma 2.2 implies

$$\sup_{0\leqslant x\leqslant 1}\left|\frac{1}{n^2}\mathbf{S}_n^T(x)\widehat{\Sigma}^{-1}\mathbf{S}_n(x) - g_A^2(x)\mathbf{d}^T\Sigma_A^{-1}\mathbf{d}\right| = o_P(1),$$

where $\mathbf{d} = (\int\Delta(t)w_{1,A}(t)\,dt,\ldots,\int\Delta(t)w_{d,A}(t)\,dt)^T$. Since $\mathbf{d}^T\Sigma_A^{-1}\mathbf{d} > 0$ and $g_A^2(\cdot)$ has a unique maximum at $x = \vartheta$ and is continuous, assertion a) follows by standard arguments, noting that $\widehat{\vartheta} = \arg\max \mathbf{S}_n^T(x)\widehat{\Sigma}^{-1}\mathbf{S}_n(x)/n^2$.

Note that $\widehat{\vartheta}$ is obtained as the arg max of $Q_n(x) := \mathbf{S}_n(x)^T\widehat{\Sigma}^{-1}\mathbf{S}_n(x)$. This is equivalent to $\widehat{\vartheta} = \widehat{k}/n$ and $\widehat{k} = \arg\max(Q_n(k/n) - Q_n(\lfloor\vartheta n\rfloor/n) : k = 1,\ldots,n)$. The key to the proof is now the following decomposition for $k \leqslant k^\circ := \lfloor n\vartheta\rfloor$ which generalizes equation (4.1) in Aue et al. [2] for situations where $\widehat{\Sigma}$ has no diagonal shape. Since for a symmetric matrix C it holds $(a-b)^TC(a+b) = a^TCa - b^TCb$ we get by (2.4)

$$Q_n(k/n) - Q_n(k^\circ/n) = (\mathbf{A}_k^{(1)} + \widehat{\mathbf{d}}B_k^{(1)})^T\widehat{\Sigma}^{-1}(\mathbf{A}_k^{(2)} + \widehat{\mathbf{d}}B_k^{(2)}), \tag{4.3}$$

for $\check{\eta}$ as in (2.3), $\mathbf{A}_k^{(j)} = (A_{k,1}^{(j)},\ldots,A_{k,d}^{(j)})^T$, $j = 1,2$, where

$$A_{k,l}^{(1)} = -\sum_{i=k+1}^{k^\circ}\check{\eta}_{i,l} - \frac{k-k^\circ}{n}\sum_{i=1}^{n}\check{\eta}_{i,l}, \quad A_{k,l}^{(2)} = \sum_{i=1}^{k}\check{\eta}_{i,l} + \sum_{i=1}^{k^\circ}\check{\eta}_{i,l} - \frac{k+k^\circ}{n}\sum_{i=1}^{n}\check{\eta}_{i,l},$$

$\widehat{\mathbf{d}} = (\widehat{d}_1,\ldots,\widehat{d}_d)^T$ with $\widehat{d}_l = \int(\mu_1(t)-\mu_2(t))\widehat{v}_l(t)\,dt$ and

$$B_k^{(1)} = (k-k^\circ)\frac{n-k^\circ}{n}, \quad B_k^{(2)} = (k+k^\circ)\frac{n-k^\circ}{n}.$$

Now, we prove that $B_k^{(1)}B_k^{(2)}\,\mathbf{d}^T\widehat{\Sigma}\mathbf{d}$ is the dominating term. Let

$$\begin{aligned} L_{n,k} &= -(k^\circ-k)(k+k^\circ)\left(\frac{n-k^\circ}{n}\right)^2, \\ \text{i.e. } |L_{n,k}| &\geqslant (k^\circ-k)n\left(\vartheta(1-\vartheta)^2 + o(1)\right). \end{aligned} \tag{4.4}$$

Then,

$$B_k^{(1)}B_k^{(2)}\,\widehat{\mathbf{d}}^T\widehat{\Sigma}\widehat{\mathbf{d}} = L_{n,k}\,\widehat{\mathbf{d}}^T\widehat{\Sigma}\widehat{\mathbf{d}} = L_{n,k}(\mathbf{d}^T\Sigma_A\mathbf{d} + o_P(1)), \tag{4.5}$$

since by assumption $\widehat{\Sigma} \xrightarrow{P} \Sigma_A$ and by Theorem 3.1 and the Cauchy-Schwarz inequality it holds for $d_l = \int(\mu_1(t)-\mu_2(t))w_{l,A}(t)\,dt$,

$$\left|\widehat{d}_l - s_ld_l\right| \leqslant \|\mu_1(\cdot)-\mu_2(\cdot)\|\|\widehat{v}_l - s_lw_{l,A}\| = o_P(1).$$

Similarly

$$\max_{1\leqslant k\leqslant k^\circ-N}\frac{|\widehat{d}_l B_k^{(1)}|}{k^\circ-k}=O_P(1),\qquad \max_{1\leqslant k\leqslant k^\circ-N}\frac{|\widehat{d}_l B_k^{(2)}|}{n}=O_P(1). \tag{4.6}$$

Concerning $A_{k,l}^{(1)}$ it first holds by stationarity of $\check{\eta}_{\cdot,l}$

$$\max_{1\leqslant k\leqslant k^\circ-N}\frac{1}{k^\circ-k}\left|\sum_{i=k+1}^{k^\circ}\check{\eta}_{i,l}\right|\overset{\mathcal{L}}{=}\max_{N\leqslant k<k^\circ}\frac{1}{k}\left|\sum_{i=1}^{k}\check{\eta}_{i,l}\right|.$$

By (4.1) and Theorem 3.1 we get

$$\begin{aligned}&\left|\max_{1\leqslant k\leqslant n}\frac{1}{k}\left|\sum_{i=1}^{k}\check{\eta}_{i,l}\right|-\max_{1\leqslant k\leqslant n}\frac{1}{k}\left|\sum_{i=1}^{k}\eta_{i,l}^{(A)}\right|\right|\\&\leqslant\|\widehat{v}_l(t)-s_k w_{l,A}(t)\|\max_{1\leqslant k\leqslant n}\left\|\frac{1}{k}\sum_{i=1}^{k}Y_i(\cdot)\right\|=o_P(1)\end{aligned}$$

as well as by Assumption $\mathcal{P}.3$

$$\max_{N\leqslant k\leqslant n}\frac{1}{k}\left|\sum_{i=1}^{k}\eta_{i,l}^{(A)}\right|\leqslant\alpha(N)^{-1}O_P(1),$$

which becomes arbitrarily small as $N\to\infty$. This shows

$$\max_{1\leqslant k\leqslant k^\circ-N}\frac{|A_{k,l}^{(1)}|}{k^\circ-k}\leqslant\alpha(N)^{-1}O_P(1)+o_P(1). \tag{4.7}$$

Since by the Cauchy-Schwarz inequality and (4.2)

$$\max_{1\leqslant k\leqslant n}\frac{1}{n}\left|\sum_{i=1}^{k}\eta_{i,l}^{(A)}\right|\leqslant\|w_{l,A}\|\max_{1\leqslant k\leqslant n}\left\|\frac{1}{n}\sum_{i=1}^{k}Y_i(\cdot)\right\|=o_P(1),$$

we get

$$\max_{1\leqslant k\leqslant k^\circ-N}\frac{|A_{k,l}^{(2)}|}{n}=o_P(1). \tag{4.8}$$

From $\widehat{\Sigma}\xrightarrow{P}\Sigma_A$, (4.4), (4.7) and (4.8) we can conclude

$$\max_{1\leqslant k\leqslant k^\circ-N}\frac{\left|\mathbf{A}_k^{(1)T}\widehat{\Sigma}\mathbf{A}_k^{(2)}\right|}{L_{n,k}}=O_P(1)\max_{l=2,\ldots,d}\max_{1\leqslant k\leqslant k^\circ-N}\frac{|A_{k,l}^{(1)}|}{k^\circ-k}\max_{1\leqslant k\leqslant k^\circ-N}\frac{|A_{k,l}^{(2)}|}{n}=o_P(1)$$

and similarly using additionally (4.6) it holds

$$\begin{aligned}&\max_{1\leqslant k\leqslant k^\circ-N}\frac{\left|\mathbf{A}_k^{(1)T}\widehat{\Sigma}\mathbf{d}B_k^{(2)}\right|}{L_{n,k}}\leqslant\alpha(N)^{-1}O_P(1)+o_P(1),\\&\max_{1\leqslant k\leqslant k^\circ-N}\frac{\left|B_k^{(1)}\mathbf{d}^T\widehat{\Sigma}\mathbf{A}_k^{(2)}\right|}{L_{n,k}}=o_P(1).\end{aligned}$$

This in addition to (4.5) and

$$\max_{k\leqslant k^\circ-N}L_{n,k}\leqslant-Nk^\circ\left(1-\frac{k^\circ}{n}\right)^2<0 \tag{4.9}$$

we obtain

$$\begin{aligned}
&P(n\widehat{\vartheta} \leqslant n\vartheta - N) = P(\widehat{k} \leqslant k^\circ - N) \\
&= P\left(\max_{k\leqslant k^\circ - N} Q_n\left(\frac{k}{n}\right) - Q_n\left(\frac{k^\circ}{n}\right) \geqslant \max_{k>k^\circ-N} Q_n\left(\frac{k}{n}\right) - Q_n\left(\frac{k^\circ}{n}\right)\right) \\
&\leqslant P\left(\max_{k\leqslant k^\circ - N} Q_n\left(\frac{k}{n}\right) - Q_n\left(\frac{k^\circ}{n}\right) \geqslant 0\right) \\
&\leqslant P\left(\left(\mathbf{d}^T\Sigma_A\mathbf{d} + o_P(1) + \alpha(N)^{-1}O_P(1)\right)\max_{k\leqslant k^\circ-N} L_{n,k} \geqslant 0\right) \\
&\leqslant P\left(\alpha(N)^{-1}O_P(1) \geqslant \mathbf{d}^T\Sigma_A\mathbf{d} + o_P(1)\right) \leqslant P(O_P(1) \geqslant \alpha(N)\,\mathbf{d}^T\Sigma_A\mathbf{d}) + o_P(1),
\end{aligned}$$

which becomes arbitrarily small if $N \to \infty$, since by assumption $\mathbf{d}^T\Sigma_A\mathbf{d} > 0$.

Analogous arguments for $k \geqslant k^\circ + N$ show that $P(n\widehat{\vartheta} \geqslant n\vartheta + N)$ becomes arbitrarily small as $N \to \infty$, which finishes the proof. ■

Proof of Theorem 2.4. Note that $g_B(x)$ is continuous and has a unique maximum at $x = \vartheta_1$ and a unique minimum at $x = \vartheta_2$, hence $g_B(x,y) = g_B(y) - g_B(x)$ is continuous and has a unique (for $x < y$) maximum at $(\vartheta_1, \vartheta_2)$. Then, the proof of a) is completely analogous to the proof of a) of Theorem 2.3.

The proof of b) is close to the proof of Theorem 2.3 b), we therefore only sketch it here. Let $Q_n(k_1,k_2) := \mathbf{S}_n(k_1/n, k_2/n)^T\widehat{\Sigma}^{-1}\mathbf{S}_n(k_1/n, k_2/n))$, then $(\widehat{k}_1, \widehat{k}_2) = \arg\max(Q_n(k_1,k_2) - Q_n(k_1^\circ, k_2^\circ))$ where $\widehat{\vartheta}_j = \widehat{k}_j/n$.

Note that by an analogous expression to (2.4) it holds ($k_j^\circ := \lfloor n\vartheta_j \rfloor$, $a_+ = \max(a,0)$)

$$S_{n,l}\left(\frac{k}{n}\right) = \sum_{j=1}^{k}\check{\eta}_{j,l} - \frac{k}{n}\sum_{j=1}^{n}\check{\eta}_{j,l} + \widehat{d}_l\left(\frac{k(k_2^\circ - k_1^\circ)}{n} - (\min(k,k_2^\circ) - k_1^\circ)_+\right),$$

where $\widehat{d}_l = \int(\mu_1(t) - \mu_2(t))\widehat{v}_l\,dt$ and $\widehat{\mathbf{d}} = (\widehat{d}_1, \ldots, \widehat{d}_n)^T$. Analogously to (4.3) we get

$$\begin{aligned}
&Q_n(k_1,k_2) - Q_n(k_1^\circ, k_2^\circ) \\
&= \left(\mathbf{S}_n\left(\frac{k_2}{n}\right) - \mathbf{S}_n\left(\frac{k_2^\circ}{n}\right) - \mathbf{S}_n\left(\frac{k_1}{n}\right) + \mathbf{S}_n\left(\frac{k_1^\circ}{n}\right)\right)^T \\
&\qquad \widehat{\Sigma}\left(\mathbf{S}_n\left(\frac{k_2}{n}\right) - \mathbf{S}_n\left(\frac{k_1}{n}\right) + \mathbf{S}_n\left(\frac{k_2^\circ}{n}\right) - \mathbf{S}_n\left(\frac{k_1^\circ}{n}\right)\right) \\
&= (\mathbf{A}^{(1)}_{k_1,k_2} - \widehat{\mathbf{d}}\,B^{(1)}_{k_1,k_2})^T\,\widehat{\Sigma}\,(\mathbf{A}^{(2)}_{k_1,k_2} - \widehat{\mathbf{d}}\,B^{(2)}_{k_1,k_2}),
\end{aligned}$$

where $\mathbf{A}^{(j)}_{k_1,k_2} = (A^{(j)}_{k_1,k_2,1}, \ldots, A^{(j)}_{k_1,k_2,d})^T$, $j = 1, 2$, and

$$\begin{aligned}
A^{(1)}_{k_1,k_2,l} &= z_2\sum_{j=m_2+1}^{M_2}\check{\eta}_{j,l} - z_1\sum_{j=m_1+1}^{M_1}\check{\eta}_{j,l} - \frac{k_2 - k_2^\circ}{n}\sum_{j=1}^{n}\check{\eta}_{j,l} + \frac{k_1 - k_1^\circ}{n}\sum_{j=1}^{n}\check{\eta}_{j,l}, \\
A^{(2)}_{k_1,k_2,l} &= \sum_{j=1}^{k_2}\check{\eta}_{j,l} - \sum_{j=1}^{k_1}\check{\eta}_{j,l} + \sum_{j=1}^{k_2^\circ}\check{\eta}_{j,l} - \sum_{j=1}^{k_1^\circ}\check{\eta}_{j,l} - \frac{k_2 - k_1 + k_2^\circ - k_1^\circ}{n}\sum_{j=1}^{n}\check{\eta}_{j,l}, \\
B^{(1)}_{k_1,k_2} &= (m_2 - k_1^\circ)_+ - (k_2^\circ - k_1^\circ) - (\min(k_1,k_2^\circ) - k_1^\circ)_+ - \frac{k_2^\circ - k_1^\circ}{n}(k_2 - k_2^\circ - k_1 + k_1^\circ), \\
B^{(2)}_{k_1,k_2} &= (m_2 - k_1^\circ)_+ - (\min(k_1,k_2^\circ) - k_1^\circ)_+ + (k_2^\circ - k_1^\circ) - \frac{k_2^\circ - k_1^\circ}{n}(k_2^\circ - k_1^\circ + k_2 - k_1)
\end{aligned}$$

with $z_j = 1$ if $k_j > k_j^\circ$ and $z_j = -1$ else, $m_j = \min(k_j, k_j^\circ)$, $M_j = \max(k_j, k_j^\circ)$. We will show that the deterministic part is dominating as long as $\max(|k_1 - k_1^\circ|, |k_2 -$

$k_2^\circ|) > N$. Here, the problem is that the maximum needs to be divided into six parts (instead of just two as in the proof of Theorem 2.3). Let

$$L_{n,k_1,k_2} := B^{(1)}_{k_1,k_2} B^{(2)}_{k_1,k_2}$$

In all six cases one can then show that analogously to (4.5)

$$\frac{|L_{n,k_1,k_2}|}{n \max(|k_1 - k_1^\circ|, |k_2 - k_2^\circ|)} \geqslant c + o(1) > 0 \tag{4.10}$$

as well as analogously to (4.9)

$$\max L_{n,k_1,k_2} < 0. \tag{4.11}$$

Due to limitations of space we only give the proof exemplary for the case where $0 \leqslant k_1 \leqslant k_1^\circ < k_2 \leqslant k_2^\circ$ and $\max(|k_1 - k_1^\circ|, |k_2 - k_2^\circ|) > N$. The other cases are not completely analogous but similar arguments can be used. In the above case we obtain

$$-B^{(1)}_{k_1,k_2} = (k_2^\circ - k_2)\left(1 - \frac{k_2^\circ - k_1^\circ}{n}\right) + (k_1^\circ - k_1)\frac{k_2^\circ - k_1^\circ}{n}, \tag{4.12}$$

and hence there exists $c_1 > 0$ such that

$$\frac{-B^{(1)}_{k_1,k_2}}{\max(|k_1 - k_1^\circ|, |k_2 - k_2^\circ|)} \geqslant c_1 + o(1).$$

Similarly

$$\begin{aligned}
B^{(2)}_{k_1,k_2} &= k_2 - k_1^\circ + k_2^\circ - k_1^\circ - \frac{k_2^\circ - k_1^\circ}{n}\left(k_2^\circ - k_1^\circ + k_2 - k_1\right) \\
&= k_2\left(1 - \frac{k_2^\circ - k_1^\circ}{n}\right) + k_1\left(\frac{k_2^\circ - k_1^\circ}{n}\right) - k_1^\circ + (k_2^\circ - k_1^\circ)\left(1 - \frac{k_2^\circ - k_1^\circ}{n}\right) \\
&\geqslant -k_1^\circ \frac{k_2^\circ - k_1^\circ}{n} + (k_2^\circ - k_1^\circ)\left(1 - \frac{k_2^\circ - k_1^\circ}{n}\right) = (k_2^\circ - k_1^\circ)\left(1 - \frac{k_2^\circ}{n}\right),
\end{aligned}$$

hence there exists $c_2 > 0$ such that

$$\frac{B^{(2)}_{k_1,k_2}}{n} \geqslant c_2 + o(1),$$

proving (4.10) and (4.11).

It is easy to see that analogously to (4.6)

$$\max_{\substack{k_1 < k_2 \\ \max(|k_1 - k_1^\circ|, |k_2 - k_2^\circ|) > N}} \frac{|B^{(2)}_{k_1,k_2}|}{n} = O_P(1),$$

as well as by a case-by-case study as above, for the exemplary case cf. (4.12), analogously to (4.6)

$$\max_{\substack{k_1 < k_2 \\ \max(|k_1 - k_1^\circ|, |k_2 - k_2^\circ|) > N}} \frac{|B^{(1)}_{k_1,k_2}|}{\max(|k_1 - k_1^\circ|, |k_2 - k_2^\circ|)} = O_P(1).$$

As in the proof of Theorem 2.3 it holds $|\widehat{d_l} - s_l d_l| = o_P(1)$. Analogously to (4.7) we get

$$\max_{\substack{k_1 < k_2 \\ \max(|k_1 - k_1^\circ|, |k_2 - k_2^\circ|) > N}} \frac{|A^{(1)}_{k_1,k_2,l}|}{\max(|k_1 - k_1^\circ|, |k_2 - k_2^\circ|)} = \alpha(N)^{-1} O_P(1) + o_P(1),$$

as well as analogously to (4.8)

$$\max_{\substack{k_1<k_2\\ \max(|k_1-k_1^\circ|,|k_2-k_2^\circ|)>N}} \frac{|A^{(2)}_{k_1,k_2,l}|}{n} = o_P(1).$$

The proof can now be completed as the proof of Theorem 2.3. ∎

4.2 Proofs of Section 3

Proof of Theorem 3.1. The assertion follows immediately from the assumptions and Lemmas 4.2 and 4.3 of Bosq [4]. ∎

Proof of Lemma 3.1. Assertion a) for L^p-m-approximable sequences has been proven in Hörmann and Kokoszka [10], Theorem 3.1. The proof for mixing sequences is very similar, where we use the version of Davydovs covariance inequality for Hilbert space valued random variables due to Dehling and Philipp [6] (Lemma 2.2) ($t^{-1}+r^{-1}+s^{-1}=1$):

$$|\mathrm{E}\langle Y_1(\cdot),Y_{1+h}(\cdot)\rangle - \langle \mathrm{E}\,Y_1(\cdot),\mathrm{E}\,Y_{1+h}(\cdot)\rangle| \leqslant 15\alpha_h^{1/t}\left(\mathrm{E}\,\|Y_1(\cdot)\|^r\right)^{1/r}\left(\mathrm{E}\,\|Y_{1+h}(\cdot)\|^s\right)^{1/s}, \tag{4.13}$$

where $\langle\cdot,\cdot\rangle=\langle\cdot,\cdot\rangle_{\mathcal{Z}}$ is the scalar product on $L^2(\mathcal{Z})$. It holds

$$\begin{aligned}&\int\int(\widehat{c}_n(t,s)-c(t,s))^2\,dt\,ds\\ &\leqslant 2\int\int\left(\frac{1}{n}\sum_{i=1}^n Y_i(t)Y_i(s)-\mathrm{E}\,Y_1(t)Y_1(s)\right)^2 dt\,ds + 2\left\|\frac{1}{n}\sum_{i=1}^n Y_i(\cdot)\right\|^4.\end{aligned}$$

$Z_i(t,s)=Y_i(t)Y_i(s)\in L^2(\mathcal{Z}\times\mathcal{Z})$ is strong mixing with mixing rate α_h. Some calculations and (4.13) yield

$$\begin{aligned}&n\,\mathrm{E}\int\int\left(\frac{1}{n}\sum_{i=1}^n Y_i(t)Y_i(s)-\mathrm{E}\,Y_1(t)Y_1(s)\right)^2 dt\,ds\\ &=\sum_{|h|<n}\left(1-\frac{|h|}{n}\right)\left(\mathrm{E}\langle Z_1,Z_{1+h}\rangle_{\mathcal{Z}\times\mathcal{Z}}-\langle \mathrm{E}\,Z_1,\mathrm{E}\,Z_{1+h}\rangle_{\mathcal{Z}\times\mathcal{Z}}\right)\\ &\leqslant c\left(\mathrm{E}\,\|Z_1\|^{2+\delta/2}_{\mathcal{Z}\times\mathcal{Z}}\right)^{\frac{2}{2+\delta/2}}\sum_{h\geqslant 1}\alpha_h^{\frac{\delta}{4+\delta}}\\ &\leqslant c'\left(\mathrm{E}\,\|Y_1(\cdot)\|^{4+\delta}_{\mathcal{Z}}\right)^{\frac{4}{4+\delta}}<\infty,\end{aligned}$$

for some constants $c,c'>0$. Hence $\int\int\left(\frac{1}{n}\sum_{i=1}^n Y_i(t)Y_i(s)-\mathrm{E}\,Y_1(t)Y_1(s)\right)^2 dt\,ds = O_P(n^{-1})$. Analogously one obtains

$$\left\|\frac{1}{n}\sum_{i=1}^n Y_i(\cdot)\right\|^2 = O_P(n^{-1}).$$

Assertion b) follows analogously to Berkes et al. [3], proof of Lemma 1, on using (4.1). ∎

Proof of Theorem 3.2. For the proof of a) let γ_l be the eigenvalue of $k(\cdot,\cdot)$ belonging to the w_l and λ_l the eigenvalue of $c(\cdot,\cdot)$ belonging to v_l. We prove the contrapositive. To this end assume

$$\int_{\mathcal{Z}}\Delta(t)w_l(t)\,dt=0\quad l=1,\ldots,d.$$

This implies

$$\int c(t,s)w_l(s)\,ds = \int k(t,s)w_l(s)\,ds - \theta(1-\theta)\Delta(t)\int \Delta(s)w_l(s)\,ds$$
$$= \gamma_l w_l(t), \quad l = 1,\ldots,d. \tag{4.14}$$

This shows that $\gamma_l, l = 1,\ldots,d$, are eigenvalues of $c(t,s)$ with eigenfunctions w_l. Hence, there exist $r_1, r_2, \ldots, r_d$, $r_s \neq r_t$ for $s \neq t$, such that $\gamma_l = \lambda_{r_l}$ and $w_l = \pm v_{r_l}$.

Recall the min-max principle for the l-largest eigenvalue β_l of a compact non-negative operator Γ in a Hilbert space with inner product $\langle\cdot,\cdot\rangle$ (cf. e.g. Gohberg et al. [9], Theorem 4.9.1)

$$\beta_l = \min_{\substack{\mathcal{S}\subset\mathcal{Z}\\ \dim(\mathcal{S})=l-1}} \max_{\substack{x\perp\mathcal{S}\\ \|x\|=1}} \langle\Gamma x, x\rangle\,. \tag{4.15}$$

For the covariance operator it holds

$$\langle Cx, x\rangle = \int_{\mathcal{Z}}\int_{\mathcal{Z}} x(t)c(t,s)x(s)dtds$$
$$\leqslant \int_{\mathcal{Z}}\int_{\mathcal{Z}} x(t)c(t,s)x(s)dtds + \theta(1-\theta)\left(\int_{\mathcal{Z}} x(t)\Delta(t)\,dt\right)^2 = \langle Kx, x\rangle,$$

where $\langle\cdot,\cdot\rangle$ is the scalar product in the Hilbert space $\mathcal{L}^2(\mathcal{Z})$ and $Cx = \int_{\mathcal{Z}} c(\cdot,s)x(s)\,ds$ and an analogous expression for K. Hence we can conclude from the min-max principle

$$\lambda_l = \min_{\substack{\mathcal{S}\subset\mathcal{Z}\\ \dim(\mathcal{S})=l-1}} \max_{\substack{x\perp\mathcal{S}\\ \|x\|=1}} \langle Cx, x\rangle \leqslant \min_{\substack{\mathcal{S}\subset\mathcal{Z}\\ \dim(\mathcal{S})=l-1}} \max_{\substack{x\perp\mathcal{S}\\ \|x\|=1}} \langle Kx, x\rangle = \gamma_l. \tag{4.16}$$

In particular

$$\lambda_1 \leqslant \gamma_1 = \lambda_{r_1} \leqslant \lambda_1,$$

hence $\lambda_1 = \gamma_1$. Analogously one can deduct inductively that $\lambda_l = \gamma_l$, $l = 2,\ldots,d$. This implies $w_l = \pm v_l$ and hence $\int \Delta(t)v_l(t)\,dt = \pm\int \Delta(t)w_l(t)\,dt = 0$, $l = 1,\ldots,d$.

Concerning b) let $\|x_0\| = 1$ with $\int \Delta(t)x_0(t)\,dt \neq 0$. By (4.15)

$$\gamma_{1,D} = \max_{\|x\|=1}\left(\int\int x(t)c(t,s)x(s)\,dt\,ds + D^2\theta(1-\theta)\left(\int \Delta(t)x(t)\,dt\right)^2\right)$$
$$\geqslant D^2\theta(1-\theta)\left(\int \Delta(t)x_0(t)\,dt\right)^2 \to \infty$$

as $D^2 \to \infty$, where $\gamma_{1,D}$ is the largest eigenvalue belonging to the contaminated covariance kernel k_D associated with the change Δ_D.

In particular, there exists $D_0 > 0$ such that for all $|D| \geqslant D_0$ it holds $\gamma_{1,D} > \lambda_1$, where λ_1 is the largest eigenvalue of the uncontaminated covariance kernel c.

Suppose that for those D it holds $\int \Delta_D(t)w_{1,D}(t) = 0$, then by (4.15)

$$\gamma_{1,D} = \int\int w_{1,D}(t)k_D(t,s)w_{1,D}(s)\,dt\,ds$$
$$= \int\int w_{1,D}(t)c(t,s)w_{1,D}(s)\,dt\,ds + \theta(1-\theta)\left(\int \Delta_D(t)w_{1,D}(t)\,dt\right)^2$$
$$\leqslant \max_{\|x\|=1}\int\int x(t)c(t,s)x(s)\,dt\,ds = \lambda_1,$$

which is a contradiction. Hence $\int \Delta_D(t)w_{1,D}(t) \neq 0$ for all $|D| \geqslant D_0$. ■

References

[1] Aston, J. A. D. and Kirch, C. Estimation of the distribution of change-points with application to fMRI data *Technical Report, 2011.*

[2] Aue, A., Gabrys, R., Horváth, L., and Kokoszka, P. Estimation of a change-point in the mean function of functional data. *J. Multivariate Anal.*, 100:2254–2269, 2009.

[3] Berkes, I., Gabrys, R., Horváth, L., and Kokoszka, P. Detecting changes in the mean of functional observations. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 71:927–946, 2009.

[4] Bosq, D. *Linear Processes in Function Spaces.* Springer, 2000.

[5] Dehling, H. Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Wahrsch. verw. Geb.*, 63:393–432, 1983.

[6] Dehling, H. and Philipp, W. Almost sure invariance principles for weakly dependent vector-valued random variables. *Ann. Probab.*, 10:689–701, 1982.

[7] Dudley, R. M. and Philipp, W. Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. verw. Geb.*, 62:509–552, 1983.

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

[9] Gohberg, I., Goldberg, S., and Kaashoek, M. A. *Basic classes of linear operators.* Birkhäuser, Boston, 2003.

[10] Hörmann, S. and Kokoszka, P. Weakly dependent functional data. *Ann. Statist.*, 38:1845–1884, 2010.

[11] Horváth, L. and Kokoszka, P. *Inference for Functional Data with Applications.* Book in preparation. 2011.

[12] Horváth, L., Kokoszka, P., and Steinebach, J. Testing for changes in multivariate dependent observations with an application to temperature changes. *J. Multivariate Anal.*, 68:96–119, 1999.

[13] Hušková, M. and Kirch, C. A note on studentized confidence intervals in change-point analysis. *Comput. Statist.*, 25:269–289, 2010.

[14] Kokoszka, P., and Leipus, R. Change-point in the mean of dependent observations. *Statist. Probab. Lett.*, 40:385–393, 1998.

[15] Kuelbs, J., and Philipp, W. Almost sure invariance principles for partial sums of mixing b-valued random variables. *Ann. Probab.*, 8:1003–1036, 1980.

[16] Politis, D.N. Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. To appear in *Econometric Theory*. Preprint: Department of Economics, UCSD, Paper 2005-03R, http://repositories.cdlib.org/ucsdecon/2005-03R.

[17] Ramsay, J. O. and Silverman, B. W. *Functional Data Analysis.* Springer, Berlin, 2nd edition, 2005.

[18] Ranga Rao, R. Relation between weak and uniform convergence of measures with applications. *Ann. Math. Statist.*, 33:659–680, 1962.

[19] Serfling, R.J. Convergence properties of S_n under moment restrictions. *Ann. Math. Statist.*, 41:1235–1248, 1970.