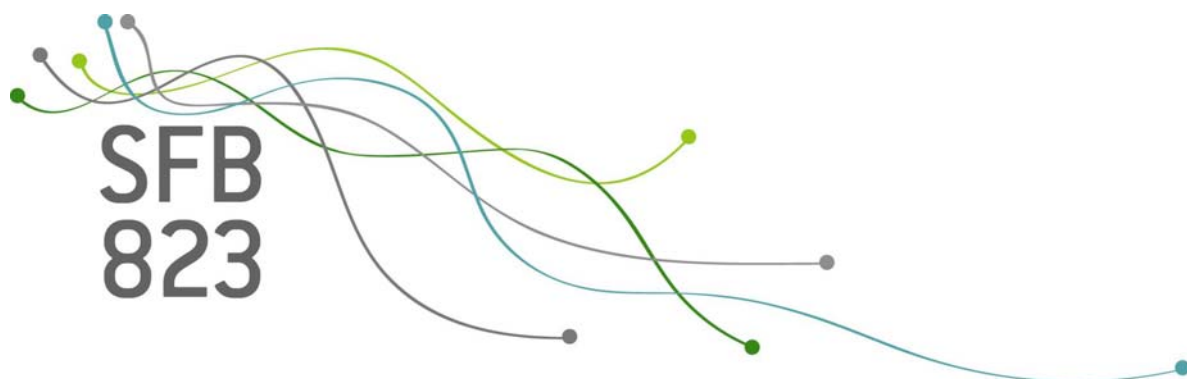# Functional data analysis in the Banach space of continuous functions

Holger Dette, Kevin Kokot, Alexander Aue

Functional data analysis in the Banach space of continuous functions[*]

Holger Dette[†]   Kevin Kokot[†]   Alexander Aue[‡]

October 17, 2017

Abstract

Functional data analysis is typically conducted within the L^2-Hilbert space framework. There is by now a fully developed statistical toolbox allowing for the principled application of the functional data machinery to real-world problems, often based on dimension reduction techniques such as functional principal component analysis. At the same time, there have recently been a number of publications that sidestep dimension reduction steps and focus on a fully functional L^2-methodology. This paper goes one step further and develops data analysis methodology for functional time series in the space of all continuous functions. The work is motivated by the fact that objects with rather different shapes may still have a small L^2-distance and are therefore identified as similar when using an L^2-metric. However, in applications it is often desirable to use metrics reflecting the visualaization of the curves in the statistical analysis. The methodological contributions are focused on developing two-sample and change-point tests as well as confidence bands, as these procedures appear do be conducive to the proposed setting. Particular interest is put on relevant differences; that is, on not trying to test for exact equality, but rather for pre-specified deviations under the null hypothesis.

The procedures are justified through large-sample theory. To ensure practicability, non-standard bootstrap procedures are developed and investigated addressing particular features that arise in the problem of testing relevant hypotheses. The finite sample properties are explored through a simulation study and an application to annual temperature profiles.

Keywords: Banach spaces; Functional data analysis; Time series; Relevant hypotheses; Two-sample tests; Change-point tests; Bootstrap

MSC 2010: 62G10, 62G15, 62M10,

[†]Fakultät für Mathematik, Ruhr-Universität Bochum, Universitätsstraße 150, 44780 Bochum, Germany, emails: [holger.dette, kevin.kokot]@ruhr-uni-bochum.de

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

1 Introduction

This paper proposes new methodology for the analysis of functional data, in particular for the two-sample and change-point settings. The basic set-up considers a sequence of Banach space-valued time series satisfying mixing conditions. The proposed methodology therefore advances functional data analysis beyond the predominant Hilbert space-based methodology. For the latter case, there exists by now a fully fledged theory. The interested reader is referred to the various monographs Ferraty and Vieu [21], Horváth and Kokoszka [26], and Ramsay and Silverman [34] for up-to-date accounts. Most of the available statistical procedures discussed in these monographs are based on dimension reduction techniques such as functional principal component analysis. However, the integral role of smoothness has been discussed at length in Ramsay and Silverman [34] and virtually all functions fit in practice are at least continuous. In such cases dimension reduction techniques can incur a loss of information and fully functional methods can prove advantageous. More recently, Aue et al. [6], Bucchia and Wendler [12] and Horváth et al. [28] discussed fully functional methodology in a Hilbert space framework.

Since all functions utilized for practical purposes are at least continuous, and often smoother than that, it might be more natural to develop methodology for functional data in the space of continuous functions. This is the approach pursued in the present paper. While it might thus be reasonable to build statistical analysis adopting this point of view, there are certain difficulties associated with it. Giving up on the theoretically convenient Hilbert space setting means that substantially more effort has to be put into the derivation of theoretical results, especially if one is interested in the incorporation of dependent functional observations. Section 2 of the main part of this paper gives an introduction to Banach space methodology and states some basic results, in particular an invariance principle for a sequential process in the space of continuous functions.

The theoretical contributions will be utilized for the development of relevant two-sample and change-point tests in Sections 3 and 4, respectively. Here the usefulness of the proposed approach becomes more apparent as differences between two smooth curves are hard to detect in practice. Additionally, small discrepancies might perhaps not even be of importance in many applied situations. Therefore the "relevant" setting is adopted that is not trying to test for exact equality under the null hypothesis, but allows for pre-specified deviations from an assumed null function. For example, if $C(T)$, the space of continuous functions on the compact interval T, is equipped with the sup-norm $\|f\| = \sup_{t \in T} |f(t)|$, and μ_1 and μ_2 are the mean functions corresponding to two samples, interest is in hypotheses of the form

$$H_0 \colon \|\mu_1 - \mu_2\| \leq \Delta \qquad \text{and} \qquad H_1 \colon \|\mu_1 - \mu_2\| > \Delta, \tag{1.1}$$

where $\Delta \geq 0$ denotes a pre-specified constant. The classical case of testing perfect equality, obtained by the choice $\Delta = 0$, is therefore a special case of (1.1). However, in applications it might be reasonable to think about this choice carefully and to define precisely the size of change which one is really interested in. In particular, testing relevant hypotheses avoids the consistency problem as mentioned in Berkson [9], that is: any consistent test will detect any arbitrary small change in the mean functions if the sample size is sufficiently large. One may also view this perspective as a particular form of a bias-variance trade-off. The problem of testing for a relevant difference between two (one-dimensional) means and other (finite-dimensional) parameters has been discussed by numerous authors in biostatistics (see Wellek [41] for a recent review), but to the best of our

knowledge these testing problems have not been considered in the context of functional data. It turns out that from a mathematical point of view the problem of testing relevant (i.e., $\Delta > 0$) hypotheses is substantially more difficult than the classical problem (i.e., $\Delta = 0$). In particular, it is not possible to work with stationarity under the null hypothesis, making the derivation of a limit distribution of a corresponding test statistic or the construction of a bootstrap procedure substantially more difficult.

Section 3 develops corresponding two-sample tests for the Banach space $C(T)$. Section 4 extends these results to the change-point setting (see Aue and Horváth [5] for a recent review of change-point methodology for time series). Here, one has to deal with the additional complexity of locating the unknown time of change. Several new results for change-point analysis of functional data in $C(T)$ are put forward. A specific challenge here is the fact that the asymptotic null distribution of test statistics for hypotheses of the type (1.1) depends on the set of extremal points of the unknown difference $\mu_1 - \mu_2$, and is therefore not distribution free. Most notable for both the two-sample and the change-point problem is the construction of non-standard bootstrap tests for relevant hypotheses to solve this problem. The bootstrap is theoretically validated and then used to determine cut-off values for the proposed procedures.

Another area of application that lends itself naturally to Banach space methodology is that of constructing confidence bands for the mean function of a collection of potentially temporally dependent, continuous functions. There has been recent work by Choi and Reimherr [17] on this topic in a Hilbert space framework for functional parameters of independent functions based on geometric considerations. Here, results for confidence bands for the mean difference in a two-sample framework are added in Section 3.2.1. Natural modifications allow for the inclusion of the one-sample case. One of the main differences between the two approaches is that the proposed bands hold pointwise, while those constructed from Hilbert space theory are valid only in an L^2-sense. This property is appealing for practitioners, because two mean curves can have a rather different shape, yet the L^2-norm of their difference might be very small.

The finite-sample properties of the relevant two-sample and change-point tests and, in particular, the performance of the bootstrap procedures are evaluated with the help of a Monte Carlo simulation study in Section 5. A number of scenarios are investigated, with the outcomes showing that the proposed methodology performs reasonably well. Furthermore, an application to a prototypical data example is given, namely two-sample and cange-point tests for annual temperature profiles recorded at measuring stations in Australia.

The outline of the rest of this paper is as follows. Section 2 introduces the basic notions of the proposed Banach space methodology and gives some preliminary results. Section 3 discusses the two-sample problem and Section 4 is concerned with change-point analysis. Empirical aspects are highlighted in Section 5. Proofs of the main results can be found in an online supplement to this paper.

2 $C(T)$-valued random variables

In this section some basic facts are provided about central limit theorems and invariance principles for $C(T)$-valued random variables, where $C(T)$ is the set of continuous functions from T into the real line $\mathbb{R}$. In what follows, unless otherwise mentioned, $C(T)$ will be equipped with the sup norm $\|\cdot\|$, defined by $\|f\| = \sup_{t \in T} |f(t)|$, thus making $(C(T), \|\cdot\|)$ a Banach space. The natural Borel σ-field $\mathcal{B}(T)$ over $C(T)$ is then generated by the open sets relative to the sup norm $\|\cdot\|$. Measurability

of random variables on $(\Omega, \mathcal{A}, P)$ taking values in $C(T)$ is understood to be with respect to $\mathcal{B}(T)$. The underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is assumed complete. It is further assumed that there is a metric ρ on T such that (T, ρ) is totally bounded. The fact that T is metrizable implies that $C(T)$ is separable and measurability issues are avoided (see Theorem 7.7 in Janson and Kaijser [30]). Moreover, any random variable X in $C(T)$ is tight (see Theorem 1.3 in Billingsley [10]).

Let X be a random variable on $(\Omega, \mathcal{A}, P)$ taking values in $C(T)$. There are different ways to formally introduce expectations and higher-order moments of Banach space-valued random variables (see Janson and Kaijser [30]). The expectation $\mathbb{E}[X]$ of a random variable X in $C(T)$ exists as an element of $C(T)$ whenever $\mathbb{E}[\|X\|] < \infty$. The kth moment exists whenever $\mathbb{E}[\|X\|^k] = \mathbb{E}[\sup_{t \in T} |X(t)|^k] < \infty$. As pointed out in Chapter 11 of Janson and Kaijser [30], kth order moments may be computed through pointwise evaluation as $\mathbb{E}[X(t_1) \cdots X(t_k)]$. The case $k = 2$ is important as it allows for the computation of covariance kernels in a pointwise fashion.

A sequence of random variables $(X_n \colon n \in \mathbb{N})$ converges in distribution or weakly to a random variable X in $C(T)$, whenever it is asymptotically tight and its finite-dimensional distributions converge weakly to the finite-dimensional distributions of X, that is,

$$(X_n(t_1), \ldots, X_n(t_k)) \Rightarrow (X(t_1), \ldots, X(t_k))$$

for any $t_1, \ldots, t_k \in T$ and any $k \in \mathbb{N}$, where the symbol "$\Rightarrow$" indicates convergence in distribution in $\mathbb{R}^k$.

A centered random variable X in $C(T)$ is said to be Gaussian if its finite-dimensional distributions are multivariate normal, that is, for any $t_1, \ldots, t_k$, $(X(t_1), \ldots, X(t_k)) \sim \mathcal{N}_k(0, \Sigma)$, where the (i, j)th entry of the covariance matrix Σ is given by $\mathbb{E}[X(t_i) X(t_j)]$, $i, j = 1, \ldots, k$. The distribution of X is hence completely characterized by its covariance function $k(t, t') = \mathbb{E}[X(t) X(t')]$; see Chapter 2 of Billingsley [10].

In general Banach spaces, deriving conditions under which the central limit theorem (CLT) holds is a difficult task, significantly more complex than the counterpart for real-valued random variables. In Banach spaces, finiteness of second moments of the underlying random variables does not provide a necessary and sufficient condition. Elaborate theory has been developed to resolve the issue, resulting in notions of Banach spaces of type 2 and cotype 2 (see the book Ledoux and Talagrand [31] for an overview). However, the Banach space of continuous functions on a compact interval does not possess the requisite type and cotype properties and further assumptions are needed in order to obtain the CLT, especially to incorporate time series of continuous functions into the framework. To model the dependence of the observations, the notion of φ-mixing triangular arrays $(X_{n,j} \colon n \in \mathbb{N}, \ j = 1, \ldots, n)$ of $C(T)$-valued random variables is introduced; see Bradley [11] and Samur [38]. First, for any two σ-fields $\mathcal{F}$ and $\mathcal{G}$, define

$$\phi(\mathcal{F}, \mathcal{G}) = \sup \left\{ |\mathbb{P}(G|F) - \mathbb{P}(G)| \colon F \in \mathcal{F}, \ G \in \mathcal{G}, \ \mathbb{P}(F) > 0 \right\},$$

where $\mathbb{P}(G|F)$ denotes the conditional probability of G given F. Next, denote by $\mathcal{F}_{k,k'}^n$ the σ-field generated by $(X_{n,j} \colon k \leq j \leq k')$. Then, define the φ-mixing coefficient as

$$\varphi(k) = \sup_{n \in \mathbb{N}, n > k} \max_{k' = 1, \ldots, n-k} \phi(\mathcal{F}_{1,k'}^n, \mathcal{F}_{k'+k,n}^n)$$

4

and call the triangular array $(X_{n,j}\colon n \in \mathbb{N}, j = 1, \ldots, n)$ φ-*mixing* whenever $\lim_{k\to\infty} \varphi(k) = 0$. The φ-mixing property is defined in a similar fashion for a sequence of random variables.

In order to obtain a CLT as well as an invariance principle for triangular arrays of φ-mixing random elements in $C(T)$, the following conditions are imposed.

Assumption 2.1. *Throughout this paper the following conditions are assumed to hold:*

(A1) There is a constant K such that, for all $n \in \mathbb{N}$ and $j = 1, \ldots, n$, $\mathbb{E}[\|X_{n,j}\|^{2+\nu}] \leq K$ for some $\nu > 0$.

(A2) Let $\mathbb{E}[X_{n,j}] = \mu^{(j)}$ for any $n \in \mathbb{N}$ and $j = 1, \ldots, n$. The distributions of the observations in each row only differ in their means, that is, the centered array $(X_{n,j} - \mu^{(j)}\colon n \in \mathbb{N},\ j = 1, \ldots, n)$ is rowwise stationary. Additionally, the covariance structure is the same in each row, that is

$$\mathrm{Cov}(X_{n,j}(t), X_{n,j'}(t')) = \gamma(j - j', t, t')$$

for all $n \in \mathbb{N}$ and $j, j' = 1, \ldots, n$. Note that $\gamma(-j, t, t') = \gamma(j, t', t)$.

(A3) $(X_{n,j}\colon n \in \mathbb{N},\ j = 1, \ldots, n)$ is uniformly Lipschitz, that is, there is a real-valued random variable M with $\mathbb{E}[M^2] < \infty$ such that, for any $n \in \mathbb{N}$ and $j = 1, \ldots, n$, the inequality

$$|X_{n,j}(t) - X_{n,j}(t')| \leq M\rho(t, t')$$

holds almost surely for all $t, t' \in T$.

(A4) $(X_{n,j}\colon n \in \mathbb{N},\ j = 1, \ldots, n)$ is φ-mixing with exponentially decreasing mixing coefficients, that is, there is a constant $a \in [0, 1)$ such that $\varphi(k) \leq a^k$ for any $k \in \mathbb{N}$.

(A5) For any sequence $(r_n)_{n\in\mathbb{N}} \subset \mathbb{N}$ such that $r_n \leq n$, $r_n/n \to 0$ as $n \to \infty$, it follows that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{r_n} (X_{n,j} - \mu^{(j)}) = o_{\mathbb{P}}(1).$$

Note that these assumptions can be formulated for sequences of random variables $(X_n\colon n \in \mathbb{N})$ in $C(T)$ in a similar way. Condition (A5) is satisfied if the distribution of the sums $\sum_{j=1}^{k}(X_{n,j} - \mu^{(j)})$ is symmetric for any $k = 1, \ldots, n$ and $n \in \mathbb{N}$ (see the remark after Proposition 3.1 in Samur [37]). Assumptions (A1)–(A4) imply the following CLT which is proved in Section 6.2 of the online supplement. Throughout this paper the symbol $\rightsquigarrow$ denotes weak convergence in $(C([0,1]))^k$ for some $k \in \mathbb{N}$.

Theorem 2.1. *Let $(X_{n,j}\colon n \in \mathbb{N},\ j = 1, \ldots, n)$ denote a triangular array of random variables in $C(T)$ with expectations $E[X_{n,j}] = \mu^{(j)}$ such that conditions (A1) − (A4) of Assumption 2.1 are*

satisfied. Then,

$$G_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_{n,j} - \mu^{(j)}) \rightsquigarrow Z$$

in $C(T)$, where Z is a centered Gaussian random variable with covariance function

$$C(s,t) = \mathrm{Cov}(Z(s), Z(t)) = \sum_{i=-\infty}^{\infty} \gamma(i,s,t). \tag{2.1}$$

Assumption (A5) will be used to verify a weak invariance principle for the process $(\hat{\mathbb{V}}_n \colon n \in \mathbb{N})$ given by

$$\hat{\mathbb{V}}_n(s) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor sn \rfloor} (X_{n,j} - \mu^{(j)}) + \sqrt{n}\left(s - \frac{\lfloor sn \rfloor}{n}\right)\left(X_{\lfloor sn \rfloor + 1} - \mu^{(\lfloor sn \rfloor + 1)}\right), \tag{2.2}$$

useful for the change-point analysis proposed in Section 4. Note that the process $(\hat{\mathbb{V}}_n(s) \colon s \in [0,1])$ is an element of the Banach space $C([0,1], C(T)) = \{\phi \colon [0,1] \to C(T) \mid \phi \text{ is continuous}\}$, where the norm on this space is given by

$$\sup_{s \in [0,1]} \sup_{t \in T} |\phi(s,t)| = \|\phi\|_{C([0,1] \times T)} \tag{2.3}$$

(note that each element of $C\big([0,1], C(T)\big)$ can equivalently be regarded as an element of $C([0,1] \times T)$). Here and throughout this paper the notation $\|\cdot\|$ is used to denote any of the arising $\sup$-norms as the corresponding space can be identified from the context. The proof of the following result is postponed to Section 6.2 of the online supplement.

Theorem 2.2. *Let $(X_{n,j} \colon n \in \mathbb{N}, j = 1, \ldots, n)$ denote an array of $C([0,1])$-valued random variables such that Assumption 2.1 is satisfied. Then, the weak invariance principle holds, that is,*

$$\hat{\mathbb{V}}_n \rightsquigarrow \mathbb{V} \tag{2.4}$$

in $C([0,1] \times T)$, where $\mathbb{V}$ is a centered Gaussian measure on $C([0,1] \times T)$ characterized by

$$\mathrm{Cov}\big(\mathbb{V}(s,t), \mathbb{V}(s',t')\big) = (s \wedge s')C(t,t'), \tag{2.5}$$

and the long-run covariance function C is given in (2.1).

3 The two-sample problem

From now on, consider the case $T = [0, 1]$, as this is the canonical choice for functional data analysis. Two-sample problems have a long history in statistics and the corresponding tests are among the most applied statistical procedures. For the functional setting, there have been a number of contributions as well. Two are worth mentioning in the present context. Hall and Van Keilegom [24] studied the effect of smoothing when converting discrete observations into functional data. Horváth et al. [27] introduced two-sample tests for L^p-m approximable functional time series based on Hilbert-space theory. In the following, a two-sample test is proposed in the Banach-space framework of Section 2. To this end, consider two independent samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ of $C([0, 1])$-valued random variables. Under (A2) in Assumption 2.1 expectation functions and covariance kernels exist and are denoted by $\mu_1 = \mathbb{E}[X_1]$ and $\mu_2 = \mathbb{E}[Y_1]$, and $k_1(t, t') = \text{Cov}(X_1(t), X_1(t'))$ and $k_2(t, t') = \text{Cov}(Y_1(t), Y_1(t'))$, respectively. Interest is then in the size of the maximal deviation

$$d_\infty = \|\mu_1 - \mu_2\| = \sup_{t \in [0,1]} |\mu_1(t) - \mu_2(t)|$$

between the two mean curves, that is, in testing the hypotheses of a relevant difference

$$H_0 : d_\infty \leq \Delta \qquad \text{versus} \qquad H_1 : d_\infty > \Delta, \tag{3.1}$$

where $\Delta \geq 0$ is a pre-specified constant determined by the user of the test. Note again that the "classical" two-sample problem $H_0 : \mu_1 = \mu_2$ versus $H_0 : \mu_1 \neq \mu_2$ – which, to the best of our knowledge, has not been investigated for $C([0, 1])$-valued data yet – is contained in this setup as the special case $\Delta = 0$. Observe also that tests for relevant differences between two finite-dimensional parameters corresponding to different populations have been considered mainly in the biostatistical literature, for example in Wellek [41]. It is assumed throughout this section that the samples are balanced in the sense that

$$\frac{m}{n + m} \longrightarrow \lambda \in (0, 1) \tag{3.2}$$

as $m, n \to \infty$. Additionally, let $X_1 \ldots, X_m$ and $Y_1, \ldots, Y_n$ be sampled from independent time series $(X_j : j \in \mathbb{N})$ and $(Y_j : j \in \mathbb{N})$ that satisfy conditions (A1)–(A4) of Assumption 2.1. Under these conditions both functional time series satisfy the CLT and it then follows from Theorem 2.1 that

$$\frac{\sqrt{n + m}}{m} \sum_{j=1}^{m} (X_j - \mu_1) \rightsquigarrow \frac{1}{\sqrt{\lambda}} Z_1 \qquad \text{and} \qquad \frac{\sqrt{n + m}}{n} \sum_{j=1}^{n} (Y_j - \mu_2) \rightsquigarrow \frac{1}{\sqrt{1 - \lambda}} Z_2, \tag{3.3}$$

where Z_1 and Z_2 are independent, centered Gaussian processes possessing covariance functions

$$C_1(t, t') = \sum_{j=-\infty}^{\infty} \gamma_1(j, t, t') \qquad \text{and} \qquad C_2(t, t') = \sum_{j=-\infty}^{\infty} \gamma_2(j, t, t'),$$

respectively. Here γ_1 and γ_2, correspond to the respective sequences $(X_j : j \in \mathbb{N})$ and $(Y_j : j \in \mathbb{N})$ and are defined in Assumption 2.1. Now, the weak convergence in (3.3) and the independence of the

samples imply immediately that

$$Z_{m,n} = \sqrt{n+m}\Big(\frac{1}{m}\sum_{j=1}^{m}X_j - \frac{1}{n}\sum_{j=1}^{n}Y_j - (\mu_1 - \mu_2)\Big) \rightsquigarrow Z \tag{3.4}$$

in $C([0,1])$ as $m,n \to \infty$, where $Z = Z_1/\sqrt{\lambda} + Z_2/\sqrt{1-\lambda}$ is a centered Gaussian process with covariance function

$$C(t,t') = \mathrm{Cov}(Z(t), Z(t')) = \frac{1}{\lambda}C_1(t,t') + \frac{1}{1-\lambda}C_2(t,t'). \tag{3.5}$$

Under the convergence in (3.4) the statistic

$$\hat{d}_\infty = \Big\|\frac{1}{m}\sum_{j=1}^{m}X_j - \frac{1}{n}\sum_{j=1}^{n}Y_j\Big\| \tag{3.6}$$

is a reasonable estimator of the maximal deviation $d_\infty = \|\mu_1 - \mu_2\|$, and the null hypothesis in (3.1) is rejected for large values of $\hat{d}_\infty$. In order to develop a test with a pre-specified asymptotic level, the limit distribution of $\hat{d}_\infty$ is determined in the following. For this purpose, let

$$\mathcal{E}^\pm = \big\{t \in [0,1]\colon \mu_1(t) - \mu_2(t) = \pm d_\infty\big\} \tag{3.7}$$

if $d_\infty > 0$, and define $\mathcal{E}^+ = \mathcal{E}^- = [0,1]$ if $d_\infty = 0$. Finally, denote by $\mathcal{E} = \mathcal{E}^+ \cup \mathcal{E}^-$ the set of extremal points of the difference $\mu_1 - \mu_2$ of the two mean functions. The first main result establishes the asymptotic distribution of the statistic $\hat{d}_\infty$.

Theorem 3.1. *If $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are sampled from independent time series $(X_j\colon j \in \mathbb{N})$ and $(Y_j\colon j \in \mathbb{N})$ in $C([0,1])$, each satisfying conditions (A1)–(A4) of Assumption 2.1, then*

$$T_{m,n} = \sqrt{n+m}(\hat{d}_\infty - d_\infty) \xrightarrow{\mathcal{D}} T(\mathcal{E}) = \max\Big\{\sup_{t\in\mathcal{E}^+} Z(t), \sup_{t\in\mathcal{E}^-} -Z(t)\Big\}, \tag{3.8}$$

where the centered Gaussian process Z is given by (3.5) and the sets $\mathcal{E}^+$ and $\mathcal{E}^-$ are defined in (3.7).

It should be emphasized that the limit distribution depends in a complicated way on the set $\mathcal{E}$ of extremal points of the difference $\mu_1 - \mu_2$ and is therefore not distribution free, even in the case of i.i.d. data. In particular, there can be two sets of processes with corresponding mean functions μ_1, μ_2 and $\tilde{\mu}_1, \tilde{\mu}_2$ such that $\|\mu_1 - \mu_2\| = \|\tilde{\mu}_1 - \tilde{\mu}_2\|$. However, the respective limit distributions in Theorem 3.1 will be entirely different if the corresponding sets of extremal points $\mathcal{E}$ and $\tilde{\mathcal{E}}$ do not coincide. The proof of Theorem 3.1 is given in Section 6.3 of the online supplement. In the case $d_\infty = 0$, $\mathcal{E}^+ = \mathcal{E}^- = [0,1]$ and it follows for the random variable $T([0,1])$ in Theorem 3.1 that

$$T = \max_{t\in[0,1]} |Z(t)|. \tag{3.9}$$

8

Here the result is a simple consequence of the weak convergence (3.4) of the process $Z_{m,n}$ (see Theorem 2.1) and the continuous mapping theorem.

However, Theorem 3.1 provides also the distributional properties of the statistic $\hat{d}_\infty$ in the case $d_\infty > 0$. This is required for testing the hypotheses of a relevant difference between the two mean functions (that is, the hypotheses in (3.1) with $\Delta > 0$) of primary interest here. In this case the weak convergence of an appropriately standardized version of $\hat{d}_\infty$ does *not* follow from the weak convergence (3.4), as the process inside the supremum in (3.6) is not centered. In fact, additional complexity enters in the proofs because even under the null hypothesis observations cannot be easily centered. For details, refer to Section 6.3 of the online supplement.

3.1 Asymptotic inference

3.1.1 Testing the classical hypothesis $H_0\colon \mu_1 \equiv \mu_2$

Theorem 3.1 also provides the asymptotic distributions of the test statistic $\hat{d}_\infty$ in the case of two identical mean functions, that is, if $\mu_1 \equiv \mu_2$. This is the situation investigated in Hall and Van Keilegom [24] and Horváth et al. [27] in Hilbert-space settings. Here it corresponds to the special case $\Delta = 0$ and thus $d_\infty = 0, \mathcal{E}^\pm = [0, 1]$. Consequently,

$$T_{m,n} \xrightarrow{\mathcal{D}} T \qquad (m, n \to \infty),$$

where the random variable T is defined in (3.9). An asymptotic level α test for the classical hypotheses

$$H_0\colon \mu_1 = \mu_2 \qquad \text{versus} \qquad H_1\colon \mu_1 \neq \mu_2 \qquad (3.10)$$

may hence be obtained by rejecting H_0 whenever

$$\hat{d}_\infty > \frac{u_{1-\alpha}}{\sqrt{n+m}}, \qquad (3.11)$$

where $u_{1-\alpha}$ is the $(1-\alpha)$-quantile of the distribution of the random variable T defined in (3.9). Using Theorem 3.1 it is easy to see that the test defined by (3.11) is consistent and has asymptotic level α.

3.1.2 Confidence bands

The methodology developed so far can easily be applied to the construction of simultaneous asymptotic confidence bands for the difference of the mean functions. There is a rich literature on confidence bands for functional data in Hilbert spaces. The available work includes Degras [18], who dealt with confidence bands for nonparametric regression with functional data; Cao et al. [16], who studied simultaneous confidence bands for the mean of dense functional data based on polynomial spline estimators; Cao [15], who developed simultaneous confidence bands for derivatives of functional data when multiple realizations are at hand for each function, exploiting within-curve correlation; and Zheng et al. [42] who treated the sparse case. Most recently Choi and Reimherr [17]

9

extracted geometric features akin to Mahalanobis distances to build confidence bands for functional parameters.

The results presented here are the first of their kind relating to Banach space-valued functional data. The first theorem uses the limit distribution obtained in Theorem 3.1 to construct asymptotic simultaneous confidence bands for the two-sample case. A corresponding bootstrap analog will be developed in the next section. Confidence bands for the one-sample case can be constructed in a similar fashion using standard arguments and the corresponding results are consequently omitted.

Theorem 3.2. *Let the assumptions of Theorem 3.1 be satisfied and, for $\alpha \in (0, 1)$, denote by $u_{1-\alpha}$ the $(1 - \alpha)$-quantile of the random variable T defined in* (3.9) *and define the functions*

$$\mu_{m,n}^{\pm}(t) = \frac{1}{m} \sum_{j=1}^{m} X_j - \frac{1}{n} \sum_{j=1}^{n} Y_j \pm \frac{u_{1-\alpha}}{\sqrt{n+m}}.$$

Then the set $C_{\alpha,m,n} = \left\{ \mu \in C([0,1]) : \mu_{m,n}^{-}(t) \leq \mu(t) \leq \mu_{m,n}^{+}(t) \text{ for all } t \in [0,1] \right\}$ defines a simultaneous asymptotic $(1 - \alpha)$ confidence band for $\mu_1 - \mu_2$, that is,

$$\lim_{m,n\to\infty} \mathbb{P}(\mu_1 - \mu_2 \in C_{\alpha,m,n}) = 1 - \alpha.$$

Note that, unlike their Hilbert-space counterparts, the simultaneous confidence bands given in Theorem 3.2 (and their bootstrap analogs in Section 3.2.1) hold for all $t \in [0,1]$ and not only almost everywhere, making the proposed bands more easily interpretable and perhaps more useful for applications.

3.1.3 Testing for a relevant difference

Recall the definition of the random variable $T(\mathcal{E})$ in Theorem 3.1, then the null hypothesis of no relevant difference in (3.1) is rejected at level α, whenever the inequality

$$\hat{d}_\infty > \Delta + \frac{u_{1-\alpha,\mathcal{E}}}{\sqrt{n+m}} \tag{3.12}$$

holds, where $u_{\alpha,\mathcal{E}}$ denotes the α-quantile of the distribution of $T(\mathcal{E})$ ($\alpha \in (0, 1)$). A conservative test avoiding the use of quantiles depending on the set of extremal points $\mathcal{E}$ can be obtained observing the inequality

$$T(\mathcal{E}) \leq T, \tag{3.13}$$

where the random variable T is defined in (3.9). If u_α denotes the α-quantile of the distribution of T, then (3.13) implies $u_{\alpha,\mathcal{E}} \leq u_\alpha$ and a conservative asymptotic level α test is given by rejecting the null hypothesis in (3.1), whenever the inequality

$$\hat{d}_\infty > \Delta + \frac{u_{1-\alpha}}{\sqrt{n+m}} \tag{3.14}$$

holds. The properties of the tests (3.12) and (3.14) depend on the size of the distance d_∞ and will be explained below. In particular, observe the following properties for the test (3.14):

(a) If $d_\infty < \Delta$, Slutsky's theorem yields that

$$\lim_{n,m\to\infty} \mathbb{P}\Big(\hat{d}_\infty > \Delta + \frac{u_{1-\alpha}}{\sqrt{n+m}}\Big) = \lim_{n,m\to\infty} \mathbb{P}\big(\sqrt{n+m}(\hat{d}_\infty - d_\infty) > \sqrt{n+m}(\Delta - d_\infty) + u_{1-\alpha}\big) = 0.$$

(b) If $d_\infty = \Delta$, we have

$$\limsup_{n,m\to\infty} \mathbb{P}\Big(\hat{d}_\infty > \Delta + \frac{u_{1-\alpha}}{\sqrt{n+m}}\Big) = \limsup_{n,m\to\infty} \mathbb{P}\big(\sqrt{n+m}(\hat{d}_\infty - d_\infty) > \sqrt{n+m}(\Delta - d_\infty) + u_{1-\alpha}\big)$$
$$\leq \lim_{n,m\to\infty} \mathbb{P}\big(\sqrt{n+m}\,(\hat{d}_\infty - d_\infty) > u_{1-\alpha,\mathcal{E}}\big) = \alpha. \qquad (3.15)$$

(c) If $d_\infty > \Delta$, the same calculation as in (a) implies

$$\lim_{n,m\to\infty} \mathbb{P}\Big(\hat{d}_\infty > \Delta + \frac{u_{1-\alpha}}{\sqrt{n+m}}\Big) = 1,$$

proving that the test defined in (3.14) is consistent.

(d) If the mean functions μ_1 and μ_2 define a boundary point of the hypotheses, that is, $d_\infty = \Delta$ and either $\mathcal{E}^+ = [0,1]$ or $\mathcal{E}^- = [0,1]$, then $T(\mathcal{E}) = \max_{t\in[0,1]} Z(t)$ or $T(\mathcal{E}) = \max_{t\in[0,1]} -Z(t)$, and consequently
$$\lim_{n,m\to\infty} \mathbb{P}\Big(\hat{d}_\infty > \Delta + \frac{u_{1-\alpha}}{\sqrt{m+n}}\Big) = \frac{\alpha}{2}.$$

Using similar arguments it can be shown that the test (3.12) satisfies

$$\lim_{n,m\to\infty} \mathbb{P}\Big(\hat{d}_\infty > \Delta + \frac{u_{1-\alpha,\mathcal{E}}}{\sqrt{n+m}}\Big) = \begin{cases} 0 & \text{if } d_\infty < \Delta. \\ \alpha & \text{if } d_\infty = \Delta. \\ 1 & \text{if } d_\infty \geq \Delta. \end{cases}$$

Summarizing, the tests for the hypothesis (3.1) of no relevant difference between the two mean functions defined in (3.12) and (3.14) have asymptotic level α and are consistent. However, the discussion given above also shows that the test (3.14) is conservative, even when $\mathcal{E} = [0,1]$.

3.2 Bootstrap

In order to use the tests (3.11), (3.12) and (3.14) for classical and relevant hypotheses, the quantiles of the distribution of the random variables $T(\mathcal{E})$ and T defined in (3.8) and (3.9) need to be estimated, which depend on certain features of the data generating process. The law $T(\mathcal{E})$ involves the unknown set of extremal points $\mathcal{E}$ of the differences of the mean functions. Moreover, the distributions of $T(\mathcal{E})$ and T depend on the long-run covariance function (3.5). There are methods available in the

literature to consistently estimate the covariance function (see, for example, Horváth et al. [27]). In practice, however, it is difficult to reliably approximate the infinite sums in (3.5) and therefore an easily implementable bootstrap procedure is proposed in the following.

It turns out that a different and non-standard bootstrap procedure will be required for testing relevant hypotheses than for classical hypotheses (and the construction of confidence bands) as in this case the null distribution depends on the set of extremal points $\mathcal{E}$. The corresponding resampling procedure requires a substantially more sophisticated analysis. Therefore the analysis of bootstrap tests for the classical hypothesis and bootstrap confidence intervals is given first and discussion of bootstrap tests for relevant hypotheses is deferred to Section 3.2.2.

3.2.1 Bootstrap confidence intervals and tests for the classical hypothesis $H_0 : \mu_1 = \mu_2$

Following Bücher and Kojadinovic [14] the use of a muliplier block bootstrap is proposed. To be precise, let $(\xi_k^{(1)} \colon k \in \mathbb{N}), \ldots, (\xi_k^{(R)} \colon k \in \mathbb{N})$ and $(\zeta_k^{(1)} \colon k \in \mathbb{N}), \ldots, (\zeta_k^{(R)} \colon k \in \mathbb{N})$ denote independent sequences of independent standard normally distributed random variables and define the $C([0,1])$-valued processes $\hat{B}_{m,n}^{(1)}, \ldots, \hat{B}_{m,n}^{(R)}$ through

$$
\begin{aligned}
\hat{B}_{m,n}^{(r)}(t) = \sqrt{n+m} \Big\{ & \frac{1}{m} \sum_{k=1}^{m-l_1+1} \frac{1}{\sqrt{l_1}} \Big(\sum_{j=k}^{k+l_1-1} X_j(t) - \frac{l_1}{n} \sum_{j=1}^{m} X_j(t) \Big) \xi_k^{(r)} \\
& + \frac{1}{n} \sum_{k=1}^{n-l_2+1} \frac{1}{\sqrt{l_2}} \Big(\sum_{j=k}^{k+l_2-1} Y_j(t) - \frac{l_2}{n} \sum_{j=1}^{n} Y_j(t) \Big) \zeta_k^{(r)} \Big\} \quad (r = 1, \ldots, R)
\end{aligned}
\tag{3.16}
$$

for $t \in [0,1]$, where $l_1, l_2 \in \mathbb{N}$ denote window sizes such that $l_1/m \to 0$ and $l_2/n \to 0$ as $l_1, l_2, m, n \to \infty$. The following result is a fundamental tool for the theoretical investigations of all bootstrap procedures proposed in this paper and is proved in Section 6.3 of the online supplement.

Theorem 3.3. *Suppose that $(X_j \colon j \in \mathbb{N})$ and $(Y_j \colon j \in \mathbb{N})$ satisfy conditions (A1)–(A4) of Assumption 2.1 and let $\hat{B}_{m,n}^{(1)}, \ldots, \hat{B}_{m,n}^{(R)}$ denote the bootstrap processes defined by (3.16) such that $l_1 = m^{\beta_1}$, $l_2 = n^{\beta_2}$ with $0 < \beta_i < \nu_i/(2+\nu_i)$ and ν_i given in Assumption 2.1, $i = 1, 2$. Then,*

$$
(Z_{m,n}, \hat{B}_{m,n}^{(1)}, \ldots, \hat{B}_{m,n}^{(R)}) \rightsquigarrow (Z, Z^{(1)}, \ldots, Z^{(R)})
$$

in $C([0,1])^{R+1}$ as $m, n \to \infty$, where $Z_{m,n}$ is defined in (3.4) and $Z^{(1)}, \ldots, Z^{(R)}$ are independent copies of the centered Gaussian process Z defined by (3.5).

Note that Theorem 3.3 holds under the null hypothesis and alternative. It leads to the following results regarding confidence bands and tests for the classical hypothesis (3.10) based on the the multiplier bootstrap. To this end, note that for the statistics

$$
T_{m,n}^{(r)} = \|\hat{B}_{m,n}^{(r)}\|, \qquad r = 1 \ldots, R,
$$

the continuous mapping theorem yields

$$\left(\sqrt{n+m}\,\hat{d}_\infty,\ T_{m,n}^{(1)},\ldots,T_{m,n}^{(R)}\right) \Rightarrow \left(T,\ T^{(1)},\ldots,T^{(R)}\right), \tag{3.17}$$

where the random variables $T^{(1)},\ldots,T^{(R)}$ are independent copies of the statistic T defined in (3.9). Now, if $T_{m,n}^{\{\lfloor R(1-\alpha)\rfloor\}}$ is the empirical $(1-\alpha)$-quantile of the bootstrap sample $T_{m,n}^{(1)},\ldots,T_{m,n}^{(R)}$, the following results are obtained.

Theorem 3.4. *Let the assumptions of Theorem 3.3 be satisfied and define the functions*

$$\mu_{m,n}^{R,\pm}(t) = \frac{1}{m}\sum_{j=1}^{m}X_j - \frac{1}{n}\sum_{j=1}^{n}Y_j \pm \frac{T_{m,n}^{\{\lfloor R(1-\alpha)\rfloor\}}}{\sqrt{n+m}}.$$

Then, $\hat{C}_{\alpha,m.n}^{R} = \{\mu \in C([0,1])\colon \mu_{m,n}^{R,-}(t) \le \mu(t) \le \mu_{m,n}^{R,+}(t)\ \text{for all}\ t \in [0,1]\}$ defines a simultaneous asymptotic $(1-\alpha)$ confidence band for $\mu_1 - \mu_2$, that is,

$$\lim_{R\to\infty}\liminf_{m,n\to\infty}\mathbb{P}(\mu_1 - \mu_2 \in \hat{C}_{\alpha,m.n}^{R}) \ge 1-\alpha.$$

This section is concluded with a corresponding statement regarding the bootstrap test for the classical hypotheses in (3.10), which rejects the null hypothesis whenever

$$\hat{d}_\infty > \frac{T_{m,n}^{\{\lfloor R(1-\alpha)\rfloor\}}}{\sqrt{n+m}}, \tag{3.18}$$

where the statistic $\hat{d}_\infty$ is defined in (3.6).

Theorem 3.5. *Let the assumptions of Theorem 3.3 be satisfied, then the test* (3.18) *has asymptotic level α and is consistent for the hypotheses* (3.10). *More precisely, under the null hypothesis of no difference in the mean functions,*

$$\lim_{R\to\infty}\limsup_{m,n\to\infty}\mathbb{P}\left(\hat{d}_\infty > \frac{T_{m,n}^{\{\lfloor R(1-\alpha)\rfloor\}}}{\sqrt{n+m}}\right) = \alpha, \tag{3.19}$$

and, under the alternative, for any $R \in \mathbb{N}$,

$$\liminf_{m,n\to\infty}\mathbb{P}\left(\hat{d}_\infty > \frac{T_{m,n}^{\{\lfloor R(1-\alpha)\rfloor\}}}{\sqrt{n+m}}\right) = 1. \tag{3.20}$$

3.2.2 Testing for relevant differences in the mean functions

The problem of constructing an appropriate bootstrap test for the hypotheses of no relevant difference in the mean functions is substantially more complicated. The reason for these difficulties consists

in the fact that in the case of relevant hypotheses the limit distribution of the corresponding test statistic is complicated. In contrast to the problem of testing the classical hypotheses (3.10), where it is sufficient to mimic the distribution of the statistic T in (3.9) (corresponding to the case $\mu_1 \equiv \mu_2$) one requires the distribution of the statistic $T(\mathcal{E})$, which depends in a sophisticated way on the set of extreme points of the (unknown) difference $\mu_1 - \mu_2$. Under the null hypothesis $\|\mu_1 - \mu_2\| \leq \Delta$ these sets can be very different, ranging from a singleton to the full interval $[0, 1]$. As a consequence the construction of a valid bootstrap procedure requires appropriate consistent estimates of the sets $\mathcal{E}^+$ and $\mathcal{E}^-$ introduced in Theorem 3.1.

For this purpose, recall the definition of the Haussdorff distance between two sets $A, B \subset \mathbb{R}$, given by

$$d_H(A, B) = \max \left\{ \sup_{x \in A} \inf_{y \in B} |x - y|, \sup_{y \in B} \inf_{x \in A} |x - y| \right\}$$

and denote by $K([0, 1])$ the set of all compact subsets of the interval $[0, 1]$. First, define estimates of the extremal sets $\mathcal{E}^+$ and $\mathcal{E}^-$ by

$$\hat{\mathcal{E}}_{m,n}^{\pm} = \left\{ t \in [0, 1] \colon \pm (\hat{\mu}_1(t) - \hat{\mu}_2(t)) \geq \hat{d}_\infty - \frac{c_{m,n}}{\sqrt{m+n}} \right\}, \tag{3.21}$$

where $c_{m,n} \sim \log(m+n)$. Our first result shows that the estimated sets $\hat{\mathcal{E}}_{m,n}^+$ and $\hat{\mathcal{E}}_{m,n}^-$ are consistent for $\mathcal{E}^+$ and $\mathcal{E}^-$, respectively.

Theorem 3.6. *Let the assumptions of Theorem 3.3 be satisfied, then*

$$d_H(\hat{\mathcal{E}}_{m,n}^{\pm}, \mathcal{E}^{\pm}) \xrightarrow{\mathbb{P}} 0,$$

where the sets $\hat{\mathcal{E}}_{m,n}^{\pm}$ are defined by (3.21).

The main implication of Theorem 3.6 consists in the fact that the random variable

$$\max_{t \in \hat{\mathcal{E}}_{m,n}^+} \hat{B}_{m,n}(t)$$

converges weakly to the random variable $\max_{t \in \mathcal{E}^+} Z(t)$. Note that $\hat{B}_{m,n} \rightsquigarrow Z$ by Theorem 3.3 and that $d_H(\hat{\mathcal{E}}_{m,n}^+, \hat{\mathcal{E}}^+) \to 0$ in probability by the previous theorem, but the combination of both statements is more delicate and requires a continuity argument which is given in Section 6.3 of the online supplement, where the following result is proved.

Theorem 3.7. *Let the assumptions of Theorem 3.3 be satisfied and define, for $r = 1, \ldots, R$,*

$$K_{m,n}^{(r)} = \max \left\{ \max_{t \in \hat{\mathcal{E}}_{m,n}^+} \hat{B}_{m,n}^{(r)}(t), \max_{t \in \hat{\mathcal{E}}_{m,n}^-} \left(- \hat{B}_{m,n}^{(r)}(t) \right) \right\}. \tag{3.22}$$

Then,

$$\left(\sqrt{n+m}\,(\hat{d}_\infty - d_\infty),\, K_{m,n}^{(1)}, \ldots, K_{m,n}^{(R)}\right) \Rightarrow (T(\mathcal{E}),\, T^{(1)}(\mathcal{E}), \ldots, T^{(R)}(\mathcal{E})), \qquad (3.23)$$

in $\mathbb{R}^{R+1}$, where $d_\infty = \|\mu_1 - \mu_2\|$, the statistic $\hat{d}_\infty$ is defined in (3.6) and the variables $T^{(1)}(\mathcal{E}), \ldots, T^{(R)}(\mathcal{E})$ are independent copies of $T(\mathcal{E})$ defined in Theorem 3.1.

Theorem 3.7 leads to a simple bootstrap test for the hypothesis of no relevant change. To be precise, let $K_{m,n}^{\{\lfloor R(1-\alpha)\rfloor\}}$ denote the empirical $(1-\alpha)$-quantile of the bootstrap sample $K_{m,n}^{(1)}, \ldots, K_{m,n}^{(R)}$, then the null hypothesis of no relevant change is rejected at level α, whenever

$$\hat{d}_\infty > \Delta + \frac{K_{m,n}^{\{\lfloor R(1-\alpha)\rfloor\}}}{\sqrt{n+m}}. \qquad (3.24)$$

The final result of this section shows that the test (3.24) is consistent and has asymptotic level α. The proof is obtained by similar arguments as given in the proof of Theorem 3.5, which are omitted for the sake of brevity.

Theorem 3.8. *Let the assumptions of Theorem 3.3 be satisfied. Then, under the null hypothesis of no relevant difference in the mean functions,*

$$\lim_{R\to\infty} \limsup_{m,n\to\infty} \mathbb{P}\left(\hat{d}_\infty > \Delta + \frac{K_{m,n}^{\{\lfloor R(1-\alpha)\rfloor\}}}{\sqrt{n+m}}\right) = \alpha,$$

and, under the alternative of a relevant difference in the mean functions, for any $R \in \mathbb{N}$,

$$\liminf_{m,n\to\infty} \mathbb{P}\left(\hat{d}_\infty > \Delta + \frac{K_{m,n}^{\{\lfloor R(1-\alpha)\rfloor\}}}{\sqrt{n+m}}\right) = 1.$$

4 Change-point analysis

Change-point problems arise naturally in a number of applications (for example, in quality control, economics and finance; see Aue and Horváth [5] for a recent review). In the functional framework, applications have centered around environmental and climate observations (see Aue et al. [3, 6]) and intra-day finance data (see Horváth et al. [26]). One of the first contributions in the area are Berkes et al. [8] and Aue et al. [4] who developed change-point analysis in a Hilbert space setting for independent data. Generalizations to time series of functional data in Hilbert spaces are due to Aston and Kirch [1, 2]. For Banach-spaces, to the best of our knowledge, the only contributions to change-point analysis available in the literature are due to Račkauskas and Suquet [35, 36], who have provided theoretical work analyzing epidemic alternatives for independent functions based on Hölder norms and dyadic interval decompositions. This section details new results on change-point analysis for $C([0,1])$-valued functional data. The work is the first to systematically exploit a time series structure of the functions as laid out in Section 2.

4.1 Asymptotic inference

More specifically, the problem of testing for a (potentially relevant) change-point is considered for triangular arrays $(X_{n,j}: n \in \mathbb{N}, j = 1, \ldots, n)$ of $C([0,1])$-valued random variables satisfying Assumption 2.1. Denote by $\mu^{(j)} = \mathbb{E}[X_{n,j}] \in C([0,1])$ the expectation of $X_{n,j}$ and assume as in part (A2) of Assumption 2.1 that $\gamma(j - j', t, t') = \mathrm{Cov}(X_{n,j}(t), X_{n,j'}(t'))$ is the covariance kernel common to all random functions in the sample. Parametrize with $s^* \in (\vartheta, 1 - \vartheta)$, where $\vartheta \in (0,1)$ is a constant, the location of the change-point, so that the sequence $(\mu^{(j)})_{j \in \mathbb{N}}$ of mean functions satisfies

$$\mu_1 = \mu^{(1)} = \cdots = \mu^{(\lfloor ns^* \rfloor)} \qquad \text{and} \qquad \mu_2 = \mu^{(\lfloor ns^* \rfloor + 1)} = \cdots = \mu^{(n)}. \tag{4.1}$$

Then, for any $n \in \mathbb{N}$, both $X_{n,1}, \ldots, X_{n,\lfloor ns^* \rfloor}$ and $X_{n,\lfloor ns^* \rfloor + 1}, \ldots, X_{n,n}$ consist of (asymptotically) identically distributed but potentially dependent random functions. Let again $d_\infty = \|\mu_1 - \mu_2\|$ denote the maximal deviation between the mean functions before and after the change-point. Interest is then in testing the hypotheses of a relevant change, that is,

$$H_0: d_\infty \leq \Delta \qquad \text{versus} \qquad H_1: d_\infty > \Delta, \tag{4.2}$$

where $\Delta \geq 0$ is a pre-specified constant. The relevant change-point test setting may be viewed in the context of a bias-variance trade-off. In the time series setting, one is often interested in accurate predictions of future realizations. However, if the stretch of observed functions suffers from a structural break, then only those functions sampled after the change-point should be included in the prediction algorithm because these typically require stationarity. This reduction of observations, however, inevitably leads to an increased variability that may be partially offset with a bias incurred through the relevant approach: if the maximal discrepancy d_∞ in the mean functions remains below a suitably chosen threshold Δ, then the mean-squared prediction error obtained from predicting with the whole sample might be smaller than the one obtained from using only the non-contaminated post-change sample. In applications to financial data, the size of the allowable bias could also be dictated by regulations imposed on, say, investment strategies (Dette and Wied [20] specifically mention Value at Risk as one such example).

Recall the definition of the sequential empirical process in (2.2), where the argument $s \in [0,1]$ of this process is used to search over all potential change locations. Note that $(\hat{\mathbb{V}}_n(s,t): (s,t) \in [0,1]^2)$ can be regarded as an element of the Banach space $C([0,1]^2)$ (see the discussion before Theorem 2.2). Define the $C([0,1]^2)$-valued process

$$\hat{\mathbb{W}}_n(s,t) = \hat{\mathbb{V}}_n(s,t) - s\hat{\mathbb{V}}_n(1,t), \qquad s, t \in [0,1], \tag{4.3}$$

then, under Assumption 2.1, Theorem 2.2 and the continuous mapping theorem show that

$$\hat{\mathbb{W}}_n \rightsquigarrow \mathbb{W} \tag{4.4}$$

in $C([0,1]^2)$, where $\mathbb{W}(s,t) = \mathbb{V}(s,t) - s\mathbb{V}(1,t)$. In particular, $\mathbb{W}$ is a centered Gaussian measure on $C([0,1]^2)$ defined by

$$\mathrm{Cov}(\mathbb{W}(s,t), \mathbb{W}(s',t')) = (s \wedge s' - ss')C(t,t'). \tag{4.5}$$

In order to define a test for the hypothesis of a relevant change-point defined by (4.2) consider the sequential empirical process $(\hat{\mathbb{U}}_n \colon n \in \mathbb{N})$ on $C([0,1]^2)$ given by

$$\hat{\mathbb{U}}_n(s,t) = \frac{1}{n}\Big(\sum_{j=1}^{\lfloor sn \rfloor} X_{n,j}(t) + n\Big(s - \frac{\lfloor sn \rfloor}{n}\Big)X_{n,\lfloor sn \rfloor+1}(t) - s\sum_{j=1}^{n} X_{n,j}(t)\Big). \qquad (4.6)$$

Evaluating its expected value shows that, in contrast to $\hat{\mathbb{W}}_n$, the process $\hat{\mathbb{U}}_n$ is typically not centered and the equality $\sqrt{n}\,\hat{\mathbb{U}}_n = \hat{\mathbb{W}}_n$ holds only in the case $\mu_1 = \mu_2$. A straightforward calculation shows that

$$\mathbb{E}\big[\hat{\mathbb{U}}_n(s,t)\big] = \big(s \wedge s^* - ss^*\big)\big(\mu_1(t) - \mu_2(t)\big) + o_{\mathbb{P}}(1)$$

uniformly in $(s,t) \in [0,1]^2$. As the function $s \mapsto s \wedge s^* - ss^*$ attains its maximum in the interval $[0,1]$ at the point s^*, the statistic

$$\hat{\mathbb{M}}_n = \sup_{s\in[0,1]} \sup_{t\in[0,1]} |\hat{\mathbb{U}}_n(s,t)| \qquad (4.7)$$

is a reasonable estimate of $s^*(1 - s^*)\,d_\infty = s^*(1 - s^*)\|\mu_1 - \mu_2\|$. It is therefore proposed to reject the null hypothesis in (4.2) for large values of the statistic $\hat{\mathbb{M}}_n$. The following result specifies the asymptotic distribution of $\hat{\mathbb{M}}_n$.

Theorem 4.1. *Assume $d_\infty > 0$, $s^* \in (0,1)$ and let $(X_{n,j}\colon n \in \mathbb{N}, j = 1,\ldots,n)$ be an array of $C([0,1])$-valued random variables satisfying Assumption 2.1. Then*

$$\mathbb{D}_n = \sqrt{n}\big(\hat{\mathbb{M}}_n - s^*(1-s^*)d_\infty\big) \xrightarrow{\ \mathcal{D}\ } D(\mathcal{E}) = \max\Big\{\sup_{t\in\mathcal{E}^+} \mathbb{W}(s^*,t), \sup_{t\in\mathcal{E}^-} -\mathbb{W}(s^*,t)\Big\}, \qquad (4.8)$$

where the statistic $\hat{\mathbb{M}}_n$ is defined in (4.7), $\mathbb{W}$ is the centered Gaussian measure on $C([0,1]^2)$ characterized by (4.5), $\mathcal{E} = \mathcal{E}^+ \cup \mathcal{E}^-$ and the sets $\mathcal{E}^+$ and $\mathcal{E}^-$ are defined in (3.7).

The proof of Theorem 4.1 is given in Section 6.4 of the online supplement. The limit distribution of $\mathbb{D}_n$ is rather complicated and depends on the set $\mathcal{E}$ which might be different for functions $\mu_1 - \mu_2$ with the same sup-norm d_∞ but different corresponding set $\mathcal{E}$. It is also worthwhile to mention that the condition $d_\infty > 0$ is essential in Theorem 4.1. In the remaining case $d_\infty = 0$ the weak convergence of $\hat{\mathbb{M}}_n$ simply follows from $\sqrt{n}\hat{\mathbb{U}}_n = \hat{\mathbb{W}}_n$, (4.4) and the continuous mapping theorem, that is,

$$\sqrt{n}\,\hat{\mathbb{M}}_n \xrightarrow{\ \mathcal{D}\ } \check{T} = \sup_{(s,t)\in[0,1]^2} |\mathbb{W}(s,t)| \qquad (4.9)$$

whenever $d_\infty = 0$.

If $d_\infty > 0$, the true location of the change-point s^* is unknown and therefore has to be estimated from the available data. The next theorem, which is proved in Section 6.4 of the online supplement, proposes one such estimator and specifies its large-sample behavior in form of a rate of convergence.

17

Theorem 4.2. *Assume $d_\infty > 0$, $s^* \in (0,1)$ and let $(X_{n,j} \colon n \in \mathbb{N}, j = 1, \ldots, n)$ be an array of $C([0,1])$-valued random variables satisfying Assumption 2.1, where the random variable M in Assumption (A3) is bounded. Then the estimator*

$$\tilde{s} = \arg \max_{1 \le k < n} \left\| \hat{\mathbb{U}}_n \big(\frac{k}{n}, \cdot \big) \right\| \tag{4.10}$$

satsifies $|\tilde{s} - s^*| = O_\mathbb{P}(n^{-1})$.

Recall that the possible range of change locations is restricted to the open interval $(\vartheta, 1 - \vartheta)$ and define the modified change-point estimator

$$\hat{s} = \max \big\{ \vartheta, \min\{\tilde{s}, 1 - \vartheta\} \big\}, \tag{4.11}$$

where $\tilde{s}$ is given by (4.10). Since $|\hat{s} - s^*| \le |\tilde{s} - s^*|$, it follows that $|\hat{s} - s^*| = O_\mathbb{P}(n^{-1})$ if $d_\infty > 0$, and, if $d_\infty = 0$ suppose that $\hat{s}$ converges weakly to a $[\vartheta, 1 - \vartheta]$-valued random variable $s_{\max}$.

Corollary 4.1. *Let the assumptions of Theorem 4.2 be satisfied and define*

$$\hat{d}_\infty = \frac{\hat{\mathbb{M}}_n}{\hat{s}(1 - \hat{s})} \tag{4.12}$$

as an estimator of d_∞. Then, $\sqrt{n}\big(\hat{d}_\infty - d_\infty\big) \Rightarrow T(\mathcal{E}) = D(\mathcal{E})/[s^*(1 - s^*)]$, where $D(\mathcal{E})$ is defined in (4.8).

Remark 4.1. A consistent level α test for the hypotheses (4.2) is constructed along the lines of the two-sample case discussed in Section 3.

(a) Consider first the case $\Delta > 0$, that is, a relevant hypothesis. If $d_\infty > 0$, implying the existence of a change-point $s^* \in (0,1)$, then the inequality

$$T(\mathcal{E}) \le T = \frac{1}{s^*(1 - s^*)} \sup_{t \in [0,1]} |\mathbb{W}(s^*, t)| \tag{4.13}$$

holds. If $u_{\alpha, \mathcal{E}}$ denotes the quantile of $T(\mathcal{E})$, then $u_{\alpha, \mathcal{E}} \le u_\alpha$ for all $\alpha \in (0,1)$. Consequently, similar arguments as given in Section 3.1.3 show that the test which rejects the null hypothesis of no relevant change if

$$\hat{d}_\infty > \Delta + \frac{u_{1-\alpha}}{\sqrt{n}} \tag{4.14}$$

is consistent and has asymptotic level α. Note that an estimator of the long-run covariance function is needed in order to obtain the α-quantile u_α of the distribution of T. Moreover, the

18

test (4.14) is conservative, even when the set $\mathcal{E}$ of extremal points of the unknown difference $\mu_1 - \mu_2$ is the whole interval $[0, 1]$ (in this case the level is in fact $\alpha/2$ instead of α - see the discussion at the end of Section 3.1.3).

(b) In the case of testing the classical hypotheses $H_0\colon \mu_1 = \mu_2$ versus $H_1\colon \mu_1 \neq \mu_2$, that is $\Delta = 0$, the test described in (4.14) needs to be slightly altered. The asymptotic distribution of $\hat{\mathbb{M}}_n$ under H_0 can be obtained from (4.9) and now it can be seen that rejecting H_0 whenever

$$\hat{d}_\infty > \frac{\check{u}_{1-\alpha}}{\sqrt{n}},$$

where $\check{u}_{1-\alpha}$ denotes the $(1 - \alpha)$-quantile of the distribution of the random variable $\check{T}$ defined by (4.9), yields a consistent asymptotic level α test.

4.2 Bootstrap

In order to avoid the difficulties mentioned in the previous remark, a bootstrap procedure is developed and its consistency is shown. To be precise, denote by

$$\hat{\mu}_1 = \frac{1}{\lfloor \hat{s}n \rfloor} \sum_{j=1}^{\lfloor \hat{s}n \rfloor} X_{n,j} \qquad \text{and} \qquad \hat{\mu}_2 = \frac{1}{\lfloor (1 - \hat{s})n \rfloor} \sum_{j=\lfloor \hat{s}n \rfloor + 1}^{n} X_{n,j}$$

estimators for the expectation before and after the change-point. Let $(\xi_k^{(1)}\colon k \in \mathbb{N}), \ldots, (\xi_k^{(R)}\colon k \in \mathbb{N})$ denote R independent sequences of independent standard normally distributed random variables and consider the $C([0, 1]^2)$-valued processes $\hat{B}_n^{(1)}, \ldots, \hat{B}_n^{(R)}$ defined by

$$
\begin{aligned}
\hat{B}_n^{(r)}(s, t) = & \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor sn \rfloor} \frac{1}{\sqrt{l}} \Big(\sum_{j=k}^{k+l-1} \hat{Y}_{n,j}(t) - \frac{l}{n} \sum_{j=1}^{n} \hat{Y}_{n,j}(t) \Big) \xi_k^{(r)} \\
& + \sqrt{n} \Big(s - \frac{\lfloor sn \rfloor}{n} \Big) \frac{1}{\sqrt{l}} \Big(\sum_{j=\lfloor sn \rfloor + 1}^{\lfloor sn \rfloor + l} \hat{Y}_{n,j}(t) - \frac{l}{n} \sum_{j=1}^{n} \hat{Y}_{n,j}(t) \Big) \xi_{\lfloor sn \rfloor + 1}^{(r)},
\end{aligned}
\tag{4.15}
$$

where $l \in \mathbb{N}$ is a bandwidth parameter satisfying $l/n \to 0$ as $l, n \to \infty$ and

$$\hat{Y}_{n,j} = X_{n,j} - (\hat{\mu}_2 - \hat{\mu}_1) \mathbb{1}\{j > \lfloor \hat{s}n \rfloor\}$$

for $j = 1, \ldots, n$ $(n \in \mathbb{N})$. Note that it is implicitly assumed that $\hat{B}_n^{(r)}((n - l + 1)/n, t) = \hat{B}_n^{(r)}(s, t)$ for any $t \in [0, 1]$ and any $s \in [0, 1]$ such that $\lfloor sn \rfloor > n - l + 1$. Next, define

$$\hat{\mathbb{W}}_n^{(r)}(s, t) = \hat{B}_n^{(r)}(s, t) - s\hat{B}_n^{(r)}(1, t) \ ; \ r = 1, \ldots, R.$$

Theorem 4.3. *Let $\hat{B}_n^{(1)}, \ldots, \hat{B}_n^{(R)}$ denote the bootstrap processes defined by (4.15), where $l = n^\beta$ for some $\beta \in (0, 1/3)$. Further assume that the underlying array $(X_{n,j}\colon j = 1, \ldots, n; n \in \mathbb{N})$ satisfies Assumption 2.1 with the additional requirement that $\nu \geq 2$ in (A1), and suppose that, for any sequence $(r_n\colon n \in \mathbb{N}) \subset \mathbb{N}$ such that $r_n \leq n$, $r_n/n \to 0$ as $n \to \infty$, it follows that*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{r_n} \left(\frac{1}{\sqrt{l}} \sum_{j=i}^{i+l-1} (X_{n,j} - \mu^{(j)}) \right) \xi_i^{(r)} = o_{\mathbb{P}}(1). \tag{4.16}$$

Then, $(\hat{\mathbb{W}}_n, \hat{\mathbb{W}}_n^{(1)}, \ldots, \hat{\mathbb{W}}_n^{(R)}) \rightsquigarrow (\mathbb{W}, \mathbb{W}^{(1)}, \ldots, \mathbb{W}^{(R)})$ in $C([0,1]^2)^{R+1}$, where $\hat{\mathbb{W}}_n$ and $\mathbb{W}$ are defined in (4.3) and (4.5), respectively, and $\mathbb{W}^{(1)}, \ldots, \mathbb{W}^{(R)}$ are independent copies of $\mathbb{W}$.

The proof of Theorem 4.3 is provided in Section 6.4 of the online supplement. Note that condition (4.16) is similar to Assumption (A5) and ensures that the weak invariance principle holds for the bootstrap processes.

We now consider a resampling procedure for the classical hypotheses, that is $\Delta = 0$ in (4.2). For that purpose, define, for $r = 1, \ldots, R$,

$$\check{T}_n^{(r)} = \max \left\{ \left| \hat{\mathbb{W}}_n^{(r)}(s,t) \right| \ : \ s, t \in [0, 1] \right\}. \tag{4.17}$$

Then, by the continuous mapping theorem,

$$(\sqrt{n}\, \hat{\mathbb{M}}_n,\ \check{T}_n^{(1)}, \ldots, \check{T}_n^{(R)}) \Rightarrow (\check{T},\ \check{T}^{(1)}, \ldots, \check{T}^{(R)})$$

in $\mathbb{R}^{R+1}$, where $\check{T}^{(1)}, \ldots, \check{T}^{(R)}$ are independent copies of the random variable $\check{T}$ defined in (4.9). If $\check{T}_n^{\{\lfloor R(1-\alpha) \rfloor\}}$ is the empirical $(1-\alpha)$-quantile of the bootstrap sample $\check{T}_n^{(1)}, \check{T}_n^{(2)}, \ldots, \check{T}_n^{(R)}$, the classical null hypothesis $H_0 : \mu_1 = \mu_2$ of no change point is rejected, whenever

$$\hat{\mathbb{M}}_n > \frac{\check{T}_n^{\{\lfloor R(1-\alpha) \rfloor\}}}{\sqrt{n}} \ . \tag{4.18}$$

It follows by similar arguments as given in Section 6.3 of the online supplement that this test is consistent and has asymptotic level α in the sense of Theorem 3.5, that is

$$\lim_{R \to \infty} \limsup_{n \to \infty} \mathbb{P}_{H_0} \left(\hat{\mathbb{M}}_n > \frac{\check{T}_n^{\{\lfloor R(1-\alpha) \rfloor\}}}{\sqrt{n}} \right) = \alpha \ , \quad \liminf_{n \to \infty} \mathbb{P}_{H_1} \left(\hat{\mathbb{M}}_n > \frac{\check{T}_n^{\{\lfloor R(1-\alpha) \rfloor\}}}{\sqrt{n}} \right) = 1,$$

for any $R \in \mathbb{N}$. The details are omitted for the sake of brevity.

We now continue developing bootstrap methodology for the problem of testing for a relevant change point, that is $\Delta > 0$ in (4.9). It turns out that the theoretical analysis is substantially more complicated as the null hypothesis defines a set in in $C([0,1])$. Similar as in (3.21) the estimates of

the extremal sets $\mathcal{E}^+$ and $\mathcal{E}^-$ are defined by

$$\hat{\mathcal{E}}_n^{\pm} = \left\{ t \in [0,1] \colon \ \pm(\hat{\mu}_1(t) - \hat{\mu}_2(t)) \geq \hat{d}_\infty - \frac{c_n}{\sqrt{n}} \right\}, \tag{4.19}$$

where $c_n \sim \log(n)$ and $\hat{d}_\infty$ is given in (4.12). Consider bootstrap analogs

$$T_n^{(r)} = \frac{1}{\hat{s}(1-\hat{s})} \max\left\{ \max_{t \in \hat{\mathcal{E}}_n^+} \hat{W}_n^{(r)}(\hat{s},t), \ \max_{t \in \hat{\mathcal{E}}_n^-} \left(-\hat{W}_n^{(r)}(\hat{s},t) \right) \right\}, \qquad r = 1, \ldots, R, \tag{4.20}$$

of the statistic $\sqrt{n}(\hat{d}_\infty - d_\infty)$ in Corollary 4.1, where $d_\infty = \|\mu_1 - \mu_2\|$.

Theorem 4.4. *Let the assumptions of Theorem 4.3 be satisfied, then, if $d_\infty > 0$,*

$$(\sqrt{n}(\hat{d}_\infty - d_\infty), \ T_n^{(1)}, \ldots, T_n^{(R)}) \Rightarrow (T(\mathcal{E}), \ T^{(1)}, \ldots, T^{(R)})$$

in $\mathbb{R}^{R+1}$, where $T^{(1)}, \ldots, T^{(R)}$ are independent copies of the random variable $T(\mathcal{E})$ defined in Corollary 4.1.

A test for the hypothesis of a relevant change-point in time series of continuous functions is now obtained by rejecting the null hypothesis in (4.2), whenever

$$\hat{d}_\infty > \Delta + \frac{T_n^{\{\lfloor R(1-\alpha)\rfloor\}}}{\sqrt{n}}, \tag{4.21}$$

where $T_n^{\{\lfloor R(1-\alpha)\rfloor\}}$ is the empirical $(1-\alpha)$-quantile of the bootstrap sample $T_n^{(1)}, T_n^{(2)}, \ldots, T_n^{(R)}$. It follows by similar arguments as given in Section 6.3 of the online supplement that this test is consistent and has asymptotic level α in the sense of Theorem 3.8, that is

$$\lim_{R \to \infty} \limsup_{n \to \infty} \mathbb{P}_{H_0}\left(\hat{d}_\infty > \Delta + \frac{T_n^{\{\lfloor R(1-\alpha)\rfloor\}}}{\sqrt{n}} \right) = \alpha$$

and

$$\liminf_{n \to \infty} \mathbb{P}_{H_1}\left(\hat{d}_\infty > \Delta + \frac{T_n^{\{\lfloor R(1-\alpha)\rfloor\}}}{\sqrt{n}} \right) = 1,$$

for any $R \in \mathbb{N}$. The details are omitted for the sake of brevity.

5 Empirical aspects

In this section, the finite sample properties of the proposed methodology are investigated by means of a small simulation study (see Section 5.1 and 5.2) and its applicability illustrated in a small data example (see Section 5.3). All simulation results presented here are based on $1,000$ runs, the length of the blocks in the bootstrap procedure is $l = 2, l_1 = 2, l_2 = 2$, and the number of bootstrap replications is chosen as $R = 200$ throughout.

5.1 Two sample problems

5.1.1 Classical hypotheses

First, a brief discussion of the bootstrap test (3.18) for the "classical" hypotheses (3.10) is given. For this problem, Horvath et al. [28] proposed a test in a Hilbert-space framework, and therefore a similar scenario as in this paper is considered. Specifically, the sample sizes are chosen as $m = 100$, $n = 200$ and the error processes are given by fAR(1) time series (see Horvath et al. [28]). The left panel of Table 5.1 displays the rejection probabilities of the new test (3.18) for the mean functions

$$\mu_1 \equiv 0 \ , \quad \mu_2(t) = at(1 - t) \tag{5.1}$$

for various values of the parameter a, while the right panel shows results for the functions

$$\mu_1 \equiv 0 \ , \quad \mu_2(t) = 0.1 \frac{(1 - t(1 - t))^k}{\int_0^1 (1 - t(1 - t))^k dt} \tag{5.2}$$

for different values of k. Note that only the model in (5.1) with $a = 0$ corresponds to the null hypothesis. A similar approximation of the nominal level as for the test of Horvath et al. [28] is observed as well as reasonable rejection probabilities under the alternative. For the sake of a comparison the results of the test proposed by Horvath et al. [28] are also displayed, using their statistics $U_{100,200}^{(1)}$ and $U_{100,200}^{(2)}$ on page 109 of their paper (these are the two numbers in brackets). For the models (5.1) the new test is in most cases more powerful than the test proposed by these authors. This superiority is also observed for the models (5.2) if $k = 4, 5$. On the other hand, if $k = 2, 3$, the test of Horvath et al. [28] based on the statistic $U_{100,200}^{(2)}$ yields the best performance, but the new test is always more powerful than the test based on their statistic $U_{100,200}^{(1)}$.

5.1.2 Confidence bands

In order to investigate the finite sample properties of the confidence bands proposed in Section 3.2.1 we investigate a similar scenario as in Sections 6.3 and 6.4 of Aue et al. [3]. To be precise let $D \in \mathbb{N}$, consider B-spline basis functions $\nu_1, \ldots, \nu_D$ (here $D = 21$) and the linear space $\mathbb{H} = \text{span}\{\nu_1, \ldots, \nu_D\}$. Now define independent processes $\varepsilon_1, \ldots, \varepsilon_n \in \mathbb{H} \subset C([0, 1])$ by

$$\varepsilon_j = \sum_{i=1}^{D} N_{i,j} \nu_i, \qquad j = 1, \ldots, n,$$

22

	(5.1)				(5.2)		
a	1%	5%	10%	k	1%	5%	10%
0	2.7 (1.8, 1.9)	7.4 (6.6, 7.2)	13.7 (12.2, 13.5)	2	27.8 (27.6, 57.8)	56.4 (52.6, 77.4)	79 (63.6, 84.9)
0.4	21 (19.4, 12.3)	37.7 (35.9, 26.5)	46.7 (46.7, 36.3)	3	31.3 (27.5, 64.3)	76.4 (49.4, 82.1)	94.9 (61.2, 88.7)
0.6	49.4 (42.1, 29.6)	67.6 (62.2, 51.8)	76.9 (73.1, 62.5)	4	61.2 (28.2, 71.6)	96.8 (52.5, 88.7)	1 (66.7, 93.8)
0.8	74.3 (68.6, 53.8)	87.1 (85.7, 74.6)	91 (91.5, 83.1)	5	90.3 (27.8, 78)	1 (51.6, 91.7)	1 (64.3, 95.5)

Table 5.1: *Simulated rejection probabilities of the bootstrap test* (3.18) *for the hypotheses* (3.10) *(in percent). The mean functions are given by* (5.1) *(left part) and by* (5.2) *(right part), the sample sizes are* $m = 100$ *and* $n = 200$ *and the case* $a = 0$ *corresponds to the null hypotheses. The numbers in brackets represent the results of the two tests proposed by Horvath et al.* [28] *taken from Table 1 in this reference.*

where $N_{1,j}, N_{2,j}, \ldots, N_{D,j}$ are independent, normally distributed random variables with expectation zero and variance $\mathrm{Var}(N_{i,j}) = \sigma_i^2 = 1/i^2$ $(i = 1, \ldots, D;\ j = 1, \ldots, n)$. The fMA(1) process is finally given by $\eta_i = \varepsilon_i + \Theta\varepsilon_{i-1}$, where the operator $\Theta\colon \mathbb{H} \to \mathbb{H}$ (acting on a finite dimensional space) is defined by $\kappa\Psi$ (here $\kappa = 0.5$). The matrix Ψ is chosen randomly, that is, a matrix consisting of normally distributed entries with mean zero and standard deviation $\sigma_i\sigma_j$ is generated and then scaled such that the resulting matrix Ψ has induced norm equal to 1. Finally the two samples are given by

$$X_i = \mu_1 + \eta_i^X \ (i = 1, \ldots, m) \qquad \text{and} \qquad Y_i = \mu_2 + \eta_i^Y \ (i = 1, \ldots, n),$$

where $(\eta_i^X \colon i \in \mathbb{Z})$ and $(\eta_i^Y \colon i \in \mathbb{Z})$ are independent fMA(1) processes distributed as $(\eta_i \colon i \in \mathbb{Z})$. In Table 5.2 we display the simulated coverage percentage and the half width of the confidence band defined in Theorem 3.4, that is

$$\frac{T_{m,n}^{\{\lfloor R(1-\alpha)\rfloor\}}}{\sqrt{m+n}}.$$

The two mean functions are given by

$$\mu_1(t) = 0, \qquad \mu_2(t) = \begin{cases} 0.5t, & t \in [0, \frac{1}{5}] \\ 0.1, & t \in (\frac{1}{5}, \frac{3}{10}] \\ -0.5t + 0.25, & t \in (\frac{3}{10}, \frac{7}{10}] \\ -0.1, & t \in (\frac{7}{10}, \frac{4}{5}] \\ 0.5t - 0.5 & t \in (\frac{4}{5}, 1] \end{cases} \qquad (5.3)$$
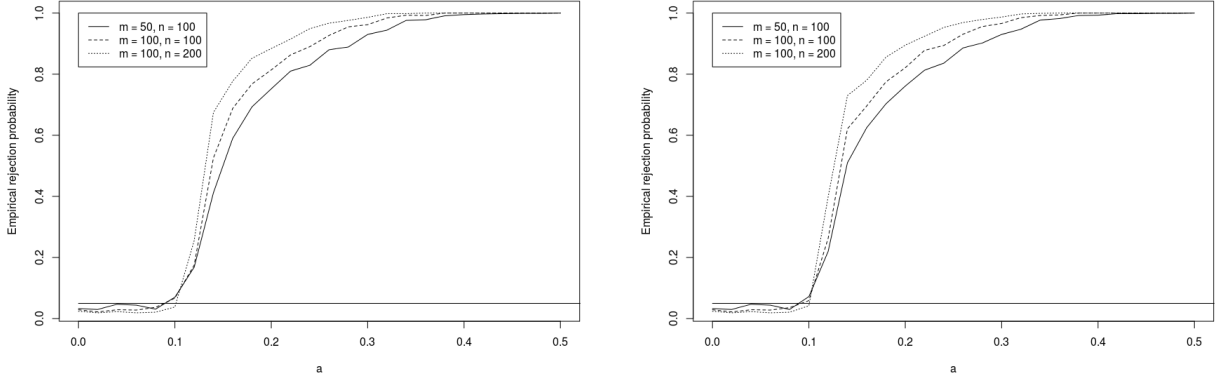
Figure 5.1: *Simulated rejection probabilities of the test* (3.24) *for a relevant difference in the maximal distance between mean functions of size* $\Delta = 0.1$. *The error processes are given by fMA(1) processes. Left panel: model* (5.3); *right panel: model* (5.4).

(left panel) and

$$\mu_1(t) = 0, \qquad \mu_2(t) = \begin{cases} 0.4t, & t \in [0, \frac{1}{4}] \\ 0.1, & t \in (\frac{1}{4}, \frac{3}{4}] \\ -0.4t + 0.4, & t \in (\frac{3}{4}, 1] \end{cases} \qquad (5.4)$$

(right panel). Note that $d_\infty = 0.1$ in both cases. For example, for sample sizes $m = 50$ and $n = 100$ the coverage probability of the 95% uniform confidence band for the difference of the mean functions in model (5.4) is $94.1, \%$ and the width is $2 \cdot 0.34 = 0.68$. We observe a reasonable approximation of the nominal level in all cases under consideration.

	(5.3)			(5.4)		
(m, n)	1%	5%	10%	1%	5%	10%
$(50, 100)$	$(97.5, 0.44)$	$(92.9, 0.34)$	$(88, 0.29)$	$(98.2, 0.44)$	$(94.1, 0.34)$	$(88.1, 0.29)$
$(100, 100)$	$(98.3, 0.36)$	$(94.7, 0.28)$	$(89.3, 0.24)$	$(98.9, 0.36)$	$(95.5, 0.28)$	$(91.2, 0.24)$
$(100, 200)$	$(98.2, 0.31)$	$(94.5, 0.24)$	$(90.4, 0.21)$	$(98.5, 0.31)$	$(94.2, 0.24)$	$(89.7, 0.21)$

Table 5.2: *Simulated coverage probabilities (first number) and half width (second number) of the confidence band for the difference of the two mean functions. The error processes are given by* $fMA(1)$ *processes. Left part: model* (5.3); *right part: model* (5.4).

5.1.3 Testing for a non relevant difference

In this paragraph the finite sample properties of the test (3.24) for the relevant hypotheses of the form (3.1) are investigated, using the same scenario as in Section 5.1.2, that is $fMA(1)$ time series with mean functions defined by (5.3) and (5.4) are used. The constant $c_{m,n}$ in (3.21) is chosen as $0.1 \log(m + n)$.

24

Figure 5.1 shows the rejection probabilities of the test (3.24) for a relevant difference in the maximal distance between mean functions of size $\Delta = 0.1$, where the mean functions are given by (5.3) (left panel) and (5.4) (right panel). The results correspond to the theoretical properties described in Theorem 3.8. More precisely, if $d_\infty = |a| < \Delta$ (interior of the null hypothesis) the probability of rejection is substantially smaller than α and decreases with increasing sample size. At the boundary of the hypotheses, that is $d_\infty = \Delta$, we observe from Table 5.3 that the simulated level is close to the nominal level α. On the other hand, the power of the test is strictly increasing with $d_\infty > \Delta$ (see Figure 5.1).

| | (5.3) | | | (5.4) | | |
(n, m)	1%	5%	10%	1%	5%	10%
$(50, 100)$	2.4	7	13.7	2.5	7.2	13.8
$(100, 100)$	1.7	6.7	11.1	1.7	5.9	11.5
$(100, 200)$	1.2	3.8	9.7	1.2	4.2	10.2

Table 5.3: *Simulated nominal level of the test* (3.24) *for a relevant difference in the mean functions at the boundary of the null hypothesis, that is $d_\infty = \Delta = 0.1$. The error processes are given by fMA(1) processes. Left part: model* (5.3)*; right part: model* (5.4)*.*

5.2 Change-point inference

In this section we investigate the finite sample performance of the change-point tests based on the maximal deviation d_∞ proposed in Section 4. Throughout this section consider a time series of the form

$$X_i = \mu_i + \eta_i^X \ \ (i = 1, \ldots, n) \tag{5.5}$$

where $(\eta_i^X)_{i \in \mathbb{Z}}$ is the fMA(1) process defined in Section 5.1.2 and the sequence of mean functions satisfies (4.1). We investigate the problem of testing for a relevant change of size $\Delta = 0.4$ in the mean functions. The corresponding results are depicted in Table 5.4, where we consider again the model (5.5) with mean functions before and after the change point $s^* = 0.5$ given by (5.3) and (5.4). The parameter c_n in (4.19) is chosen as $0.1 \log(n)$. The case $a = 0.4$ corresponds to the boundary of the hypotheses and here a rather accurate approximation of the nominal level is observed. In the interior of the null hypothesis (that is $a < 0.4$) the rejection probability, for $n = 200, 500$, is strictly smaller than the nominal level and decreasing with an increasing sample size as described at the end of Section 3.1.3 (note that the same arguments also hold for the change point problem). Similarly, under the alternative (i.e. $a > 0.4$) the test shows reasonable rejection probabilities which are increasing with sample size and a.

5.3 Data example

To illustrate the proposed methodology, two applications to annual temperature profiles are reported in this section. Data of this kind were recently used in Aue and van Delft [7] and van Delft et al. [39] in the context of stationarity tests for functional time series and earlier in Fremdt et al. [22] in support of methodology designed to choose the dimension of the projection space obtained with fPCA. For

	a	1%	5%	10%	1%	5%	10%	1%	5%	10%
	n		100			200			500	
	0.37	1.9	4.7	8.4	0.3	0.5	1.1	0	0	0.1
	0.38	2.2	5.2	7.2	0.2	0.6	1.2	0	0	0.1
	0.39	1.8	5	9	0.4	1.1	3.4	0.2	0.5	1.2
(5.3)	0.4	2.8	9.3	17.7	1.3	5.1	10.4	0.9	4.2	8.6
	0.41	6.1	14.0	24.7	5	15	26.2	12.1	29.8	44.4
	0.42	09.3	25.3	40.4	14.8	36.9	58.1	45.9	84.2	95.4
	0.43	19.3	42.2	62.9	42.2	76.4	91.0	91.8	99.4	99.7
	0.37	1.9	4.6	8.2	0.3	0.5	1.1	0	0	0
	0.38	2.1	4.6	7.2	0.1	0.6	1.2	0	0	0.1
	0.39	2.0	5.2	8.7	0.3	1.1	3.1	0.1	0.2	0.8
(5.4)	0.4	2.3	7.8	16.3	1.5	5.4	11.6	0.7	4.2	9.7
	0.41	6.7	17.4	32.6	7.9	21.3	37.3	18.0	43.8	64.9
	0.42	14.6	35.8	54.9	27.7	62.1	81.9	76.1	96.0	99.5
	0.43	32.7	63.9	78.3	68.1	91.8	96.5	98.1	99.7	99.8

Table 5.4: *Simulated nominal level (in percent) of the test (4.21) for the hypotheses (4.2) of a relevant change in the maximal deviation of the mean functions in model (5.5), where $\Delta = 0.4$. The case $a = 0.4$ corresponds to the boundary of the hypotheses, $a < 0.4$ to the null hypothesis and $a > 0.4$ to the alternative. The error processes are given by fMA(1) processes. Upper part: model (5.3); lower part: model (5.4).*

all examples, functions were generated from daily values through representation in a Fourier basis consisting of 49 basis functions, where reasonable deviations from this preset do not qualitatively change the outcome of the analyses to follow.

5.3.1 Two-sample tests

For the two-sample testing problem, annual temperature profiles were obtained from daily temperatures recorded at measuring stations in Cape Otway, a location close to the southernmost point of Australia, and Sydney, a city on the eastern coast of Australia. This led to $m = 147$ respectively $n = 153$ functions for the two samples. Differences in the temperature profiles are expected due to different climate conditions, so the focus of the relevant tests is on working out how big the discrepancy might be. The data considered here is part of the larger data set considered, for example, in Aue and van Delft [7] and van Delft et al. [39].

To set up the test for the hypotheses (3.1), the statistic in (3.6) was computed, resulting in the value

$$\hat{d}_\infty = 5.73.$$

To see whether this is significant, the proposed bootstrap methodology was applied. To estimate the extremal sets in (3.7), the estimators in (3.21) were utilized with $c_{m,n} = 0.1 \log(m+n) = 0.570$. The resulting bootstrap quantiles are reported in the second row of Table 5.5. Also reported in this table are the results of the bootstrap procedure in (3.24) for various levels α and relevance Δ. Note that the maximum difference in mean the functions is achieved at $t = 0.99$, towards the end of December

Δ	99%	97.5%	95%	90%
q	5.138	4.201	3.757	3.009
5.4	TRUE	TRUE	TRUE	TRUE
5.45	FALSE	TRUE	TRUE	TRUE
5.5	FALSE	FALSE	TRUE	TRUE
5.55	FALSE	FALSE	TRUE	TRUE
5.6	FALSE	FALSE	FALSE	FALSE

Table 5.5: *Summary of the bootstrap two sample procedure for relevant hypotheses with varying Δ for the annual temperature curves. The label TRUE refers to a rejection of the null, the label FALSE to a failure to reject the null.*

and consequently during the Australian summer. The results show that there is strong evidence in the data to support the hypothesis that the maximal difference is at least $\Delta = 5.4$, but that there is no evidence that the maximal difference is even larger than $\Delta = 5.6$. Several intermediate values of Δ led to weaker support of the alternative. The left panel of Figure 5.2 displays the difference in mean functions graphically.

5.3.2 Change-point tests

Following Fremdt et al. [22], annual temperature curves were obtained from daily minimum temperatures recorded in Melbourne, Australia. This led to 156 annual temperature profiles ranging from 1856 to 2011 to which the change-point test for the relevant hypotheses in (4.2) was applied based on the rejection decision in (4.14). To compute the test statistic $\hat{d}_\infty$ in (4.12), note that the estimated change-point in (4.11) was $\hat{s} = 0.62$ (corresponding to the year 1962). This gives

$$\hat{d}_\infty = 1.765.$$

To see whether this value leads to a rejection of the null, the multiplier bootstrap procedure was utilized with bandwidth parameter $l = 1$, leading to the rejection rule in (4.21). In order to apply this procedure, first the extremal sets $\hat{\mathcal{E}}^+$ and $\hat{\mathcal{E}}^-$ in (4.19) were selected, choosing $c_n = 0.1 \log n = 0.504$. This yielded the bootstrap quantiles reported in the second row of Table 5.6.

Several values for Δ, determining which deviations are to be considered relevant, were then examined. The results of the bootstrap testing procedure are summarized in Table 5.6. It can be seen that the null hypothesis of no relevant change was rejected at all considered levels for the smaller choice $\Delta = 1.2$. On the other extreme, for $\Delta = 1.4$, the test never rejected. For the intermediate values $\Delta = 1.25, 1.3, 1.35$, the null was rejected at the 2.5%, 5% and 20% level, at the 5% and 10% level, and at the 10% level, respectively. Estimating the mean functions before and after $\hat{s}$ (1962) shows that the maximum difference of the mean functions is approximately 1.765, lending further credibility to the conducted analyses. The right panel of Figure 5.2 displays both mean functions for illustration. It can be seen that the mean difference is maximal during the Australian summer (in February), indicating that the mean functions of minimum temperature profiles have been most drastically changed during the hottest part of the year. The results here are in agreement with the
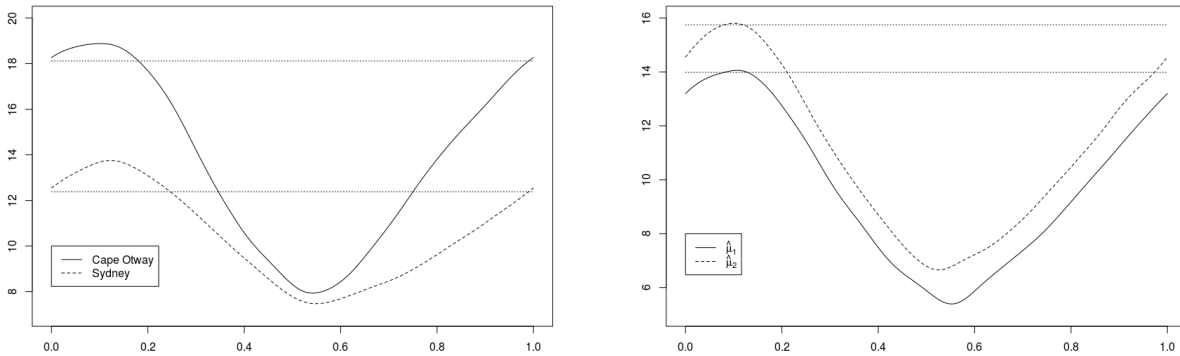
Figure 5.2: *Mean functions for the Australian temperature data. Left panel: Estimated mean functions of the Cape Otway and Sydney series for the two-sample case. Right panel: Estimated mean functions before and after the estimated change-point for the Melbourne temperature series.*

Δ	99%	97.5%	95%	90%
q	6.632	6.278	5.603	4.697
1.2	TRUE	TRUE	TRUE	TRUE
1.25	FALSE	TRUE	TRUE	TRUE
1.3	FALSE	FALSE	TRUE	TRUE
1.35	FALSE	FALSE	FALSE	TRUE
1.4	FALSE	FALSE	FALSE	FALSE

Table 5.6: *Summary of the bootstrap change-point procedure for relevant hypotheses with varying Δ for the annual temperature curves. The label TRUE refers to a rejection of the null, the label FALSE to a failure to reject the null.*

findings put forward in Hughes et al. [29], who reported that average temperatures in Antarctica have risen due to increases in minimum temperatures.

In summary, the results in this section highlight that there is strong evidence in the data for an increase in the mean function of Melbourne annual temperature profiles, with the maximum difference between "before" and "after" mean functions being at least 1.25 degrees centigrade. There is weak evidence that this difference is at least 1.35 degrees centigrade, but there is no support for the relevant hypothesis that it is even larger than that.

References

[1] John A.D. Aston and Claudia Kirch. Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis*, 109:204–220, 2012.

[2] John A.D. Aston and Claudia Kirch. Evaluating stationarity via change-point alternatives with applications to fmri data. *The Annals of Applied Statistics*, 6:1906–1948, 2012.

[3] Alexander Aue, Diogo Dubart Norinho, and Siegfried Hörmann. On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110:378–392, 2015.

[4] Alexander Aue, Robertas Gabrys, Lajos Horváth, and Piotr Kokoszka. Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis*, 100:2254–2269, 2009.

[5] Alexander Aue and Lajos Horváth. Structural breaks in time series. *Journal of Time Series Analysis*, 34:1–16, 2013.

[6] Alexander Aue, Gregory Rice, and Ozan Sönmez. Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society: Series B*, forthcoming.

[7] Alexander Aue and Anne van Delft. Testing for stationarity of functional time series in the frequency domain. *ArXiv e-print 1701.01741v1*, 2017.

[8] Istvan Berkes, Robertas Gabrys, Lajos Horváth, and Piotr Kokoszka. Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B*, 71:927–946, 2009.

[9] Joseph Berkson. Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association*, 33:526–536, 1938.

[10] Patrick Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.

[11] Richard C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2:107–144, 2005.

[12] Beatrice Bucchia and Martin Wendler. Change-point detection and bootstrap for Hilbert space valued random fields. *Journal of Multivariate Analysis*, 155:344–368, 2017.

[13] Axel Bücher, Michael Hoffmann, Mathias Vetter, and Holger Dette. Nonparametric tests for detecting breaks in the jump behaviour of a time-continuous process. *Bernoulli*, 23(2):1335–1364, 2017.

[14] Axel Bücher and Ivan Kojadinovic. A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. *Bernoulli*, 22:927–968, 2016.

[15] Guanqun Cao. Simultaneous confidence bands for derivatives of dependent functional data. *Electronic Journal of Statistics*, 8:2639–2663, 2014.

[16] Guanqun Cao, Lijian Yang, and David Todem. Simultaneous inference for the mean function based on dense functional data. *Journal of Nonparametric Statistics*, 24:359–377, 2012.

[17] Hyunphil Choi and Matthew Reimherr. A geometric approach to confidence regions and bands for functional parameters. *Journal of the Royal Statistical Society: Series B*, forthcoming.

[18] David A. Degras. Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, 21:1735–1765, 2011.

[19] Herold Dehling and Walter Philipp. *Empirical Process Techniques for Dependent Data*, pages 3–113. Birkhäuser Boston, Boston, MA, 2002.

[20] Holger Dette and Dominik Wied. Detecting relevant changes in time series models. *Journal of the Royal Statistical Society: Series B*, 78:371–394, 2016.

[21] Frederic Ferraty and Phillipe Vieu. *Nonparametric Functional Data Analysis*. Springer-Verlag, New York, 2010.

[22] Stefan Fremdt, Lajos Horváth, Piotr Kokoszka, and Josef G. Steinebach. Functional data analysis with increasing number of projections. *Journal of Multivariate Analysis*, 124:313–332, 2014.

[23] Peter Gaenssler, Péter Molnár, and Daniel Rost. On continuity and strict increase of the cdf for the sup-functional of a gaussian process with applications to statistics. *Results in Mathematics*, 51(1):51–60, 2007.

[24] Peter Hall and Ingrid Van Keilegom. Two-sample tests in functional data analysis starting from discrete data. *Statistica Sinica*, 17:1511–1531, 2007.

[25] Samir Ben Hariz, Jonathan J. Wylie, and Qiang Zhang. Optimal rate of convergence for nonparametric change-point estimators for nonstationary sequences. *The Annals of Statistics*, 35:1802–1826, 2007.

[26] Lajos Horváth and Piotr Kokoszka. *Inference for Functional Data with Applications*. Springer-Verlag, New York, 2012.

[27] Lajos Horváth, Piotr Kokoszka, and Ron Reeder. Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B*, 75:103–122, 2013.

[28] Lajos Horváth, Piotr Kokoszka, and Gregory Rice. Testing stationarity of functional time series. *Journal of Econometrics*, 179:66–82, 2014.

[29] Gillian L. Hughes, Suhasini Subba Rao, and Tata Subba Rao. Statistical analysis and time-series models for minimum/maximum temperatures in the Antarctic Peninsula. *Proceedings of the Royal Society A*, 463:241–259, 2007.

[30] Svante Janson and Sten Kaijser. Higher moments of Banach space valued random variables. *Memoirs of the American Mathematical Society*, 238, 2015.

[31] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces*. Springer, 1991.

[32] E. Lehmann. *Testing Statistical Hypotheses, Second Edition*. Chapman and Hall, New York, 1986.

[33] M. Raghavachari. Limiting distributions of Kolmogorov–Smirnov type statistics under the alternative. *The Annals of Statistics*, 1:67–73, 1973.

[34] James O. Ramsay and Bernhard W. Silverman. *Functional Data Analysis*. Springer, New York, second edition, 2005.

[35] Alfredas Račkauskas and Charles Suquet. Hölder norm test statistics for epidemic change. *Journal of Statistical Planning and Inference*, 126:495–520, 2004.

[36] Alfredas Račkauskas and Charles Suquet. Testing epidemic changes of infinite dimensional parameters. *Statistical Inference for Stochastic Processes*, 9:111–134, 2006.

[37] Jorge D. Samur. Convergence of sums of mixing triangular arrays of random vectors with stationary rows. *The Annals of Probability*, 12:390–426, 1984.

[38] Jorge D. Samur. On the invariance principle for stationary ϕ-mixing triangular arrays with infinitely divisible limits. *Probability Theory and Related Fields*, 75:245–259, 1987.

[39] Anne van Delft, Pramita Bagchi, Vaidotas Characiejus, and Holger Dette. A nonparametric test for stationarity in functional time series. *ArXiv e-print 1708.05248*, 2017.

[40] Aad W. Van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications in Statistics*. Springer, New York, 1996.

[41] Stefan Wellek. *Testing Statistical Hypotheses of Equivalence and Noninferiority*. CRC Press, Boca Raton, FL, second edition, 2010.

[42] Shuzhuan Zheng, Lijian Yang, and Wolfgang K. Härdle. A smooth simultaneous confidence corridor for the mean of sparse functional data. *Journal of the American Statistical Association*, 109:661–673, 2014.

6 Online supplement: Proofs

6.1 Convergence of suprema of non-centered processes

This section collects some preliminary results useful in the proofs of the main results. Van der Vaart and Wellner [40] formulate the concept of weak convergence in a slightly more general way, not restricting to sequences of random variables. They state their theory for *nets* of random variables $(X_\alpha \colon \alpha \in A)$, where A is a *directed set*, that is, a non-empty set equipped with a partial order $\leq$ with the additional property that, for any pair $a, b \in A$, there is $c \in A$ such that $a \leq c$ and $b \leq c$. This is a natural extension of the case considered in this paper and it will sometimes be used in the proofs to follow. The weak convergence result presented next is central to establishing limit results for the two-sample and change-point tests.

Theorem 6.1. *Let $(X_\alpha \colon \alpha \in A)$ denote a net of random variables taking values in $C(T)$ and let $\mu \in C(T)$. If $r \colon A \to \mathbb{R}_+$ is defined such that $a_\alpha = log(r_\alpha)/\sqrt{r_\alpha} = o(1)$ and if $Z_\alpha = \sqrt{r_\alpha}(X_\alpha - \mu)$ converges weakly to a Gaussian random variable Z in $C(T)$, then*

$$D_\alpha = \sqrt{r_\alpha}\big(\|X_\alpha\| - \|\mu\|\big) \Rightarrow D(\mathcal{E}) = \max\Big\{\sup_{t\in\mathcal{E}^+} Z(t), \sup_{t\in\mathcal{E}^-} -Z(t)\Big\}$$

in $\mathbb{R}$, where $\Rightarrow$ denotes convergence in distribution, $\mathcal{E} = \mathcal{E}^+ \cup \mathcal{E}^-$ the set of extremal points of μ, divided into $\mathcal{E}^\pm = \{t \in T \colon \mu(t) = \pm\|\mu\|\}$, tacitly adopting the convention $\mathcal{E}^\pm = T$ if $\mu \equiv 0$.

Proof of Theorem 6.1. First notice that, if μ is the zero function, the result is implied by the continuous mapping theorem. Hence, only the case $\|\mu\| > 0$ is considered in the following for which some arguments from Raghavachari [33] are applied. To show that $D_\alpha \xrightarrow{\mathcal{D}} D(\mathcal{E})$, introduce

$$D_\alpha(\mathcal{E}) = \sqrt{r_\alpha}\left(\sup_{t\in\mathcal{E}} |X_\alpha(t)| - \|\mu\|\right)$$

and note that the assertion of Theorem 6.1 directly follows from the following lemmas.

Lemma 6.1. *Under the assumptions of Theorem 6.1, it holds that $D_\alpha(\mathcal{E}) \Rightarrow D(\mathcal{E})$.*

Lemma 6.2. *Under the assumptions of Theorem 6.1, it holds that $R_\alpha = D_\alpha - D_\alpha(\mathcal{E}) = o_\mathbb{P}(1)$.*

Proof of Lemma 6.1. Define the random variable

$$\tilde{D}_\alpha(\mathcal{E}) = \max\Big\{\sup_{t\in\mathcal{E}^+} Z_\alpha(t), \sup_{t\in\mathcal{E}^-} -Z_\alpha(t)\Big\}.$$

From the continuous mapping theorem it follows that $\tilde{D}_\alpha(\mathcal{E}) \Rightarrow D(\mathcal{E})$. Recall that $d_\infty = \|\mu\|$ and rewrite $D_\alpha(\mathcal{E})$ as

$$D_\alpha(\mathcal{E}) = \sqrt{r_\alpha}\left(\sup_{t\in\mathcal{E}} |X_\alpha(t)| - d_\infty\right) = \sqrt{r_\alpha}\max\Big\{\sup_{t\in\mathcal{E}^+} |X_\alpha(t)| - d_\infty, \sup_{t\in\mathcal{E}^-} |X_\alpha(t)| - d_\infty\Big\}$$

$$= \sqrt{r_\alpha} \max \left\{ \sup_{t \in \mathcal{E}^+} X_\alpha(t) - d_\infty, \ \sup_{t \in \mathcal{E}^-} -X_\alpha(t) - d_\infty \right\} + o_{\mathbb{P}}(1)$$

$$= \sqrt{r_\alpha} \max \left\{ \sup_{t \in \mathcal{E}^+} (X_\alpha(t) - \mu(t)), \ \sup_{t \in \mathcal{E}^-} (-X_\alpha(t) + \mu(t)) \right\} + o_{\mathbb{P}}(1)$$

$$= \tilde{D}_\alpha(\mathcal{E}) + o_{\mathbb{P}}(1).$$

The assertion follows.

Proof of Lemma 6.2. First observe that the assumed weak convergence of $Z_\alpha = \sqrt{r_\alpha}(X_\alpha - \mu)$ to Z and the continuous mapping theorem imply that $\|Z_\alpha\| \Rightarrow \|Z\|$. Consequently, Slutsky's lemma yields that

$$\lim_{\alpha \to \infty} \mathbb{P}\left(\|X_\alpha - \mu\| > \frac{a_\alpha}{2} \right) = 0, \tag{6.1}$$

noting that $a_\alpha = \log(r_\alpha)/\sqrt{r_\alpha} = o(1)$ by assumption. The proof of the lemma is now given in two steps.

Step 1: First, define the sets

$$\mathcal{E}_\alpha^\pm = \{ t \in T : |\pm d_\infty - \mu(t)| \le a_\alpha \} = \{ t \in T : \pm \mu(t) \ge d_\infty - a_\alpha \}$$

on which the function $\pm \mu$ is within a_α from its extremal value d_∞, and let $\mathcal{E}_\alpha = \mathcal{E}_\alpha^+ \cup \mathcal{E}_\alpha^-$. It will be shown in the following that D_α can be replaced with $D_\alpha(\mathcal{E}_\alpha)$ in the definition of R_α without changing its asymptotic behavior. Here, $D_\alpha(\mathcal{E}_\alpha)$ is defined as $D_\alpha(\mathcal{E})$, using $\mathcal{E}_\alpha$ in place of $\mathcal{E}$. A corresponding definition is used for $D_\alpha(T \backslash \mathcal{E}_\alpha)$. Then,

$$0 \le R_\alpha = D_\alpha - D_\alpha(\mathcal{E}) = \max \left\{ D_\alpha(\mathcal{E}_\alpha) - D_\alpha(\mathcal{E}), D_\alpha(T \backslash \mathcal{E}_\alpha) - D_\alpha(\mathcal{E}) \right\}. \tag{6.2}$$

The second term in the maximum on the right-hand side of (6.2) is negligible as the following considerations show. Note that

$$D_\alpha(T \backslash \mathcal{E}_\alpha) - D_\alpha(\mathcal{E}) = \sqrt{r_\alpha} \left(\sup_{t \in T \backslash \mathcal{E}_\alpha} |X_\alpha(t) - \mu(t) + \mu(t)| - \sup_{t \in \mathcal{E}} |X_\alpha(t)| \right)$$

$$\le \sqrt{r_\alpha} \left(\sup_{t \in T \backslash \mathcal{E}_\alpha} (|Y_\alpha(t)| + |\mu(t)|) - \sup_{t \in \mathcal{E}} |X_\alpha(t)| \right),$$

where $Y_\alpha = X_\alpha - \mu$ is the centered version of X_α. The definition of $\mathcal{E}_\alpha$ yields that

$$\sqrt{r_\alpha} \left(\sup_{t \in T \backslash \mathcal{E}_\alpha} (|Y_\alpha(t)| + |\mu(t)|) - \sup_{t \in \mathcal{E}} |X_\alpha(t)| \right) < \sqrt{r_\alpha} \left(\sup_{t \in T \backslash \mathcal{E}_\alpha} |Y_\alpha(t)| + d_\infty - a_\alpha - \sup_{t \in \mathcal{E}} |X_\alpha(t)| \right)$$

$$\le \sqrt{r_\alpha} \sup_{t \in T} |Y_\alpha(t)| - \log(r_\alpha) - \sqrt{r_\alpha} \left(\sup_{t \in \mathcal{E}} |X_\alpha(t)| - d_\infty \right).$$

33

Observe next that

$$\lim_{\alpha \to \infty} \mathbb{P}\left(\sqrt{r_\alpha} \sup_{t \in T} |Y_\alpha(t)| - \log(r_\alpha) - \sqrt{r_\alpha}\left(\sup_{t \in \mathcal{E}} |X_\alpha(t)| - d_\infty\right) > 0\right)$$

$$\leq \lim_{\alpha \to \infty} \mathbb{P}\left(\left|\sqrt{r_\alpha} \sup_{t \in T} |Y_\alpha(t)| - \sqrt{r_\alpha}\left(\sup_{t \in \mathcal{E}} |X_\alpha(t)| - d_\infty\right)\right| > \log(r_\alpha)\right)$$

$$\leq \lim_{\alpha \to \infty} \left\{\mathbb{P}\left(\sqrt{r_\alpha} \sup_{t \in T} |Y_\alpha(t)| > \frac{\log(r_\alpha)}{2}\right) + \mathbb{P}\left(\sqrt{r_\alpha}\left|\sup_{t \in \mathcal{E}} |X_\alpha(t)| - d_\infty\right| > \frac{\log(r_\alpha)}{2}\right)\right\} = 0.$$

The first term on the right-hand side converges to 0 because of (6.1) and, recalling the weak convergence shown in Lemma 6.1, it follows by similar arguments that the second term converges to 0. On the other hand, the first term in (6.2) is always nonnegative as $\mathcal{E}$ is a subset of $\mathcal{E}_\alpha$. Therefore,

$$0 \leq R_\alpha = D_\alpha(\mathcal{E}_\alpha) - D_\alpha(\mathcal{E}) + o_{\mathbb{P}}(1) = \sqrt{r_\alpha}\left(\sup_{t \in \mathcal{E}_\alpha} |X_\alpha(t)| - \sup_{t \in \mathcal{E}} |X_\alpha(t)|\right) + o_{\mathbb{P}}(1)$$

holds and it suffices to evaluate $D_\alpha(\mathcal{E}_\alpha) - D_\alpha(\mathcal{E})$.

Step 2: Define $R_\alpha^\pm = D_\alpha(\mathcal{E}_\alpha^\pm) - D_\alpha(\mathcal{E}^\pm)$ and observe that then $0 \leq R_\alpha \leq \max\{R_\alpha^+, R_\alpha^-\} + o_{\mathbb{P}}(1)$. For the completion of the proof it is necessary to show that $R_\alpha^\pm = o_{\mathbb{P}}(1)$. To this end, write

$$0 \leq R_\alpha^+ = \sqrt{r_\alpha}\left(\sup_{t \in \mathcal{E}_\alpha^+} |X_\alpha(t)| - \sup_{t \in \mathcal{E}^+} |X_\alpha(t)|\right) = \sqrt{r_\alpha}\left(\sup_{t \in \mathcal{E}_\alpha^+} X_\alpha(t) - \sup_{t \in \mathcal{E}^+} X_\alpha(t)\right) + o_{\mathbb{P}}(1)$$

$$= \sqrt{r_\alpha}\left(\sup_{t \in \mathcal{E}_\alpha^+} \left(X_\alpha(t) - d_\infty\right) - \sup_{t \in \mathcal{E}^+} \left(X_\alpha(t) - d_\infty\right)\right) + o_{\mathbb{P}}(1)$$

$$\leq \sqrt{r_\alpha}\left(\sup_{t \in \mathcal{E}_\alpha^+} \left(X_\alpha(t) - \mu(t)\right) - \sup_{t \in \mathcal{E}^+} \left(X_\alpha(t) - \mu(t)\right)\right) + o_{\mathbb{P}}(1)$$

$$= \sqrt{r_\alpha}\left(\sup_{t \in \mathcal{E}_\alpha^+} Y_\alpha(t) - \sup_{t \in \mathcal{E}^+} Y_\alpha(t)\right) + o_{\mathbb{P}}(1).$$

Define $\mathcal{E}^+(\gamma) = \{s \in T \colon \exists\, t \in \mathcal{E}^+ \text{ with } \rho(t, s) < \gamma\}$ and $\delta_\alpha = 2\inf\{\gamma > 0 \colon \mathcal{E}_\alpha^+ \subset \mathcal{E}^+(\gamma)\}$. Since $\mathcal{E}_\alpha^+ \subset \mathcal{E}^+(\delta_\alpha)$ the above expression can be bounded by

$$\sqrt{r_\alpha}\left(\sup_{t \in \mathcal{E}_\alpha^+} Y_\alpha(t) - \sup_{t \in \mathcal{E}^+} Y_\alpha(t)\right) + o_{\mathbb{P}}(1) \leq \sqrt{r_\alpha}\left(\sup_{t \in \mathcal{E}^+(\delta_\alpha)} Y_\alpha(t) - \sup_{t \in \mathcal{E}^+} Y_\alpha(t)\right) + o_{\mathbb{P}}(1)$$

$$\leq \sqrt{r_\alpha} \sup_{\rho(s,t) < \delta_\alpha} |Y_\alpha(s) - Y_\alpha(t)| + o_{\mathbb{P}}(1).$$

Because of equicontinuity, it remains to show that $\lim_{\alpha \to \infty} \delta_\alpha = 0$. Now, the sequence $(\delta_\alpha \colon \alpha \in A)$ decreases in α and $\delta_\alpha > 0$ so that $\lim_{\alpha \to \infty} \delta_\alpha$ exists. By construction, $\mathcal{E}_\alpha^+ \subset \mathcal{E}^+(\delta_\alpha)$ but $\mathcal{E}_\alpha^+ \not\subset \mathcal{E}^+(\delta_\alpha/4)$. There is hence a subsequence $(s_\alpha \colon \alpha \in A) \subset \mathcal{E}_\alpha^+$ such that $\rho(s_\alpha, t) \geq \delta_\alpha/4$ for all $t \in \mathcal{E}^+$ and all $\alpha \in A$. The compactness of T implies that $(s_\alpha \colon \alpha \in A)$ contains a convergent subsequence $(s_{\alpha_\beta} \colon \beta \in A)$. It follows thus that $\lim_{\beta \to \infty} s_{\alpha_\beta} = s \in T$ and $d_\infty = \lim_{\beta \to \infty} \mu(s_{\alpha_\beta}) = \mu(s)$ because μ is continuous. Consequently $s \in \mathcal{E}^+$ but on the other hand $\rho(s_{\alpha_\beta}, s) \geq \delta_\alpha/4$ so that $\lim_{\alpha \to \infty} \delta_\alpha$ has to be 0. The lemma is now proven because similar arguments imply that $R_\alpha^- = o_{\mathbb{P}}(1)$.

6.2 Proofs of the results in Section 2

Proof of Theorem 2.1. Note that the results of Section 1.5 in Van der Vaart and Wellner [40] also hold, if the space ℓ^∞ is replaced by $C(T)$. Consequently, the assertion follows from the convergence of the finite dimensional distributions of $(G_n\colon n \in \mathbb{N})$ and the existence of a metric ρ such that $(G_n\colon n \in \mathbb{N})$ is uniformly ρ-equicontinuous in probability.

In a first step, convergence of finite-dimensional distributions is verified. Let $q \in \mathbb{N}$ and $t_1, \ldots, t_q \in T$. By the Cramér-Wold device, the convergence of the finite-dimensional distributions follows from

$$Z_n = \sum_{j=1}^q c_j G_n(t_j) = \sum_{j=1}^q c_j \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(X_{n,i}(t_j) - \mu^{(i)}(t_j) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,n} \Rightarrow \bar{Z} = \sum_{j=1}^q c_j Z(t_j),$$

(6.3)

where $Z_{i,n} = \sum_{j=1}^q c_j (X_{n,i}(t_j) - \mu^{(i)}(t_j))$ and $c_1, \ldots, c_q \in \mathbb{R}$ denote arbitrary constants.

In order to show the weak convergence of Z_n to $\bar{Z}$, a blocking technique similar to Lemma A.1 in Bücher and Kojadinovic [14] is utilized. Each block consists of a big subblock followed by a small subblock. For $1/(2 + 2\nu) < \eta_b < \eta_s < 1/2$, where $\nu > 0$ is the same as in Assumption (A1), define the length of the small and the big subblocks as $s_n = \lfloor n^{1/2 - \eta_s} \rfloor$ and $b_n = \lfloor n^{1/2 - \eta_b} \rfloor$, respectively. In total there are therefore $k_n = \lfloor n/(b_n + s_n) \rfloor$ blocks. Now consider the random variables

$$B_{j,n} = \sum_{i=(j-1)(b_n+s_n)+1}^{(j-1)(b_n+s_n)+b_n} Z_{i,n} \qquad \text{and} \qquad S_{j,n} = \sum_{i=(j-1)(b_n+s_n)+b_n+1}^{j(b_n+s_n)} Z_{i,n}, \qquad j = 1, \ldots, k_n,$$

corresponding to the sums of the j-th big and small subblock, respectively. It directly follows that

$$Z_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{k_n} B_{j,n} + \frac{1}{\sqrt{n}} \sum_{j=1}^{k_n} S_{j,n} + \frac{1}{\sqrt{n}} R_n,$$

(6.4)

where $R_n = \sum_{i=k_n(b_n+s_n)+1}^n Z_{i,n}$ denotes the sum of the remaining terms $Z_{i,n}$ after the last small subblock. Now the variance of Z_n can be written as

$$\mathrm{Var}(Z_n) = \mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{k_n} B_{j,n} \right) + \frac{2}{n} \sum_{j,j'=1}^{k_n} \mathbb{E}[B_{j,n} S_{j',n}] + \frac{2}{n} \sum_{j=1}^{k_n} \mathbb{E}[B_{j,n} R_n]$$

$$+ \frac{1}{n} \sum_{j,j'=1}^{k_n} \mathbb{E}[S_{j,n} S_{j',n}] + \frac{2}{n} \sum_{j=1}^{k_n} \mathbb{E}[S_{j,n} R_n] + \frac{1}{n} \mathbb{E}[R_n^2].$$

(6.5)

It will be shown that each term except the first on the right-hand side of (6.5) converges to zero. As a consequence, it follows that

$$|Z_n - 1/\sqrt{n} \sum_{j=1}^{k_n} B_{j,n}| = |1/\sqrt{n} \sum_{j=1}^{k_n} S_{j,n} + 1/\sqrt{n} R_n| = o_{\mathbb{P}}(1), \tag{6.6}$$

$$\mathrm{Var}(Z_n) = \mathrm{Var}\left(\sum_{j=1}^{k_n} B_{j,n} \right) + o(1). \tag{6.7}$$

First show that the second term on the right side of (6.5) converges to zero. To this end,

$$\mathbb{E}[B_{j,n} S_{j',n}] = \sum_{i=(j-1)(b_n+s_n)+1}^{(j-1)(b_n+s_n)+b_n} \sum_{i'=(j'-1)(b_n+s_n)+b_n+1}^{j'(b_n+s_n)} \mathbb{E}[Z_{i,n} Z_{i',n}].$$

Note that $\sigma(X(t)) \subset \sigma(X)$ for any random variable X in $C([0,1])$ and $t \in [0,1]$, where $\sigma(X)$ denotes the σ-field generated by X (see Problem 1.7.1 in Van der Vaart and Wellner [40]). Hence, $\phi(\sigma(X(t)), \sigma(Y(s))) \le \phi(\sigma(X), \sigma(Y))$ for any $C([0,1])$-valued random variables X, Y and any $s, t \in [0,1]$. Using this fact and formula (3.17) in Dehling and Philipp [19] leads to the bound

$$|\mathbb{E}[Z_{i,n} Z_{i',n}]| \le \sum_{l,l'=1}^{q} |c_l c_{l'}||\mathrm{Cov}(X_i(t_l), X_{i'}(t_{l'}))| \tag{6.8}$$

$$\le 2\varphi(|i-i'|)^{1/2} \sum_{l,l'=1}^{q} |c_l c_{l'}| \, \mathbb{E}[X_i(t_l)^2]^{1/2} \mathbb{E}[X_{i'}(t_{l'})^2]^{1/2} \lesssim \varphi(|i-i'|)^{1/2},$$

where the last inequality follows from Assumption (A1) and the symbol $\lesssim$ means less or equal up to a constant independent of n. This gives

$$|\mathbb{E}[B_{j,n} S_{j,n}]| \lesssim \sum_{i=(j-1)(b_n+s_n)+1}^{(j-1)(b_n+s_n)+b_n} \sum_{i'=(j-1)(b_n+s_n)+b_n+1}^{j(b_n+s_n)} \varphi(|i-i'|)^{1/2} \lesssim \sum_{i=1}^{b_n+s_n-1} i\varphi(i)^{1/2} < \infty. \tag{6.9}$$

The last inequality holds, since Assumption (A4) yields $\sum_{i=1}^{\infty} i\varphi(i)^{1/2} \le \sum_{i=1}^{\infty} ia^{i/2} < \infty$. Similarly it can be shown that $|\mathbb{E}[B_{j,n} S_{j-1,n}]| < \infty$. For $j' > j+1$ and $j > j'+1$ there is at least one big subblock between the observations and since $\varphi(\cdot)$ is monotonically decreasing, it follows that

$$|\mathbb{E}[B_{j,n} S_{j',n}]| \lesssim \sum_{i=(j-1)(b_n+s_n)+1}^{(j-1)(b_n+s_n)+b_n} \sum_{i'=(j'-1)(b_n+s_n)+b_n+1}^{j'(b_n+s_n)} \varphi(|i-i'|)^{1/2} = O(b_n s_n \varphi(b_n)^{1/2}).$$

Overall, it follows that

$$\frac{2}{n} \sum_{j,j'=1}^{k_n} \mathbb{E}[B_{j,n} S_{j',n}] = O(n^{-1} k_n) + O(n^{-1} k_n^2 b_n s_n \varphi(b_n)^{1/2}) = O(b_n^{-1}) + O(n b_n^{-1} s_n a^{b_n/2}) = o(1),$$

using that $n^{-1} k_n = (b_n + s_n)^{-1} = O(b_n^{-1})$ and that $s_n/b_n \to 0$ and $na^{b_n/2} \to 0$.

For the third term in (6.5), proceed in a similar way as in (6.9) to get

$$\mathbb{E}[B_{k_n,n} R_n] \lesssim \sum_{i=(k_n-1)(b_n+s_n)+1}^{(k_n-1)(b_n+s_n)+b_n} \sum_{i'=k_n(b_n+s_n)+1}^{n} \varphi(|i - i'|)^{1/2} \lesssim \sum_{i=1}^{b_n+(n-k_n(s_n+b_n))-1} i\varphi(i)^{1/2} < \infty.$$

In the case of $j < k_n$, there is again at least one big subblock between the observations in $\mathbb{E}[B_{j,n} R_n]$ and therefore it follows that

$$|\mathbb{E}[B_{j,n} R_n]| \lesssim \sum_{i=(j-1)(b_n+s_n)+1}^{(j-1)(b_n+s_n)+b_n} \sum_{i'=k_n(b_n+s_n)+1}^{n} \varphi(|i - i'|)^{1/2} = O(b_n^2 \varphi(b_n)^{1/2}),$$

since $n - k_n(s_n + b_n) \le s_n + b_n = O(b_n)$. Altogether, the calculations above yield

$$\frac{2}{n} \sum_{j=1}^{k_n} \mathbb{E}[B_{j,n} R_n] = O(n^{-1} k_n b_n^2 \varphi(b_n)^{1/2}) + O(n^{-1}) = O(b_n a^{b_n/2}) + O(n^{-1}),$$

which converges to zero.

Now consider the fourth term in (6.5) and use (6.8) to conclude

$$\mathbb{E}[S_{j,n} S_{j',n}] \lesssim \sum_{i=(j-1)(b_n+s_n)+b_n+1}^{j(b_n+s_n)} \sum_{i'=(j'-1)(b_n+s_n)+b_n+1}^{j'(b_n+s_n)} \varphi(|i - i'|)^{1/2}. \tag{6.10}$$

For $j = j'$, it holds that

$$\mathbb{E}[S_{j,n}^2] \lesssim \sum_{i,i'=1}^{s_n} \varphi(|i - i'|)^{1/2} \lesssim \sum_{i=0}^{s_n-1} (s_n - i)\varphi(i)^{1/2} \lesssim s_n \sum_{i=0}^{\infty} a^{i/2} \lesssim s_n \tag{6.11}$$

and therefore $\mathbb{E}[S_{j,n}^2] = O(s_n)$. Since there is always at least one big subblock between two small subblocks, it follows that $\mathbb{E}[S_{j,n} S_{j',n}] = O(s_n^2 \varphi(b_n)^{1/2})$ for $j \neq j'$. Hence,

$$\frac{1}{n} \sum_{j,j'=1}^{k_n} \mathbb{E}[S_{j,n} S_{j',n}] = O(n^{-1} k_n s_n) + O(n^{-1} k_n^2 s_n^2 \varphi(b_n)^{1/2}) = O(b_n^{-1} s_n) + O(n b_n^{-2} s_n^2 a^{b_n/2}),$$

37

which converges to zero, since $b_n^{-1} s_n \to 0$ and $n a^{b_n/2} \to 0$ as $n \to \infty$.

For the fifth term in (6.5), use similar arguments as for the third term and for the last term, use the same arguments as in (6.11) to get

$$\frac{1}{n} \mathbb{E}[R_n^2] = O(n^{-1}(n - k_n(b_n + s_n))) = O(n^{-1} b_n) \to 0.$$

From (6.6) it follows that it suffices to show the convergence $n^{-1/2} \sum_{j=1}^{k_n} B_{j,n} \Rightarrow \bar{Z}$ in order to establish (6.3). For that purpose let $\psi_{j,n}(t) = \exp(itn^{-1/2} B_{j,n})$ and define $\mathbb{E}\big[\prod_{j=1}^{k_n} \psi_{j,n}(t) \big]$ as the characteristic function of $n^{-1/2} \sum_{j=1}^{k_n} B_{j,n}$. Let $B'_{1,n}, \ldots, B'_{k_n,n}$ denote independent random variables such that $B_{j,n}$ and $B'_{j,n}$ are equally distributed $(j = 1, \ldots, k_n)$ and define $\prod_{j=1}^{k_n} \mathbb{E}[\psi_{j,n}(t)]$ as the characteristic function of $n^{-1/2} \sum_{j=1}^{k_n} B'_{j,n}$. Then

$$\left| \mathbb{E}\Big[\prod_{j=1}^{k_n} \psi_{j,n}(t) \Big] - \prod_{j=1}^{k_n} \mathbb{E}\Big[\psi_{j,n}(t) \Big] \right| \le \sum_{i=1}^{k_n} \left| \prod_{j=1}^{i-1} \mathbb{E}\Big[\psi_{j,n}(t) \Big] \right| \left| \mathbb{E}\Big[\prod_{j=i}^{k_n} \psi_{j,n}(t) \Big] - \mathbb{E}\Big[\psi_{i,n}(t) \Big] \mathbb{E}\Big[\prod_{j=i+1}^{k_n} \psi_{j,n}(t) \Big] \right|$$

$$\lesssim k_n \max_{1 \le i \le k_n - 1} \phi\Big(\sigma\left(\psi_{i,n}(t)\right), \sigma\Big(\prod_{j=i+1}^{k_n} \psi_{j,n}(t) \Big) \Big) \lesssim k_n \varphi(s_n) = o(1),$$

where Lemma 3.9 of Dehling and Philipp [19] was used for the the second inequality, while the third inequality follows, since there are always s_n observations between two big subblocks. Hence, it suffices to show that $n^{-1/2} \sum_{j=1}^{k_n} B'_{j,n}$ converges in distribution to $\bar{Z}$ in order to establish (6.3). For that purpose, utilize the Lindeberg–Feller central limit theorem for triangular arrays. It is first shown that $\mathrm{Var}(n^{-1/2} \sum_{j=1}^{k_n} B'_{j,n})$ converges to the variance of $\bar{Z}$.

Recall that the random variables $B'_{1,n}, \ldots, B'_{k_n,n}$ are independent and have the same distributions as $B_{1,n}, \ldots, B_{k_n,n}$, respectively. Thus,

$$\mathrm{Var}\Big(\frac{1}{\sqrt{n}} \sum_{j=1}^{k_n} B'_{j,n} \Big) = \frac{1}{n} \sum_{j=1}^{k_n} \mathrm{Var}(B'_{j,n}) = \mathrm{Var}\Big(\frac{1}{\sqrt{n}} \sum_{j=1}^{k_n} B_{j,n} \Big) - \frac{1}{n} \sum_{\substack{j,j'=1 \\ j \ne j'}}^{k_n} \mathbb{E}[B_{j,n} B_{j',n}]. \quad (6.12)$$

It is already known that $\mathrm{Var}(n^{-1/2} \sum_{j=1}^{k_n} B_{j,n}) = \mathrm{Var}(Z_n) + o(1)$. Recall the calculations in (6.10) and the subsequent discussion to note that it can be shown in a similar way that $\mathbb{E}[B_{j,n} B_{j',n}] = O(b_n^2 \varphi(s_n)^{1/2})$, for $j \ne j'$. Consequently,

$$\frac{1}{n} \sum_{\substack{j,j'=1 \\ j \ne j'}}^{k_n} \mathbb{E}[B_{j,n} B_{j',n}] = O(n^{-1} k_n^2 b_n^2 \varphi(s_n)^{1/2}).$$

Since $n^{-1}k_n = O((b_n + s_n)^{-1}) = O(b_n^{-1})$, $O(n^{-1}k_n^2 b_n^2 \varphi(s_n)^{1/2}) = O(na^{s_n/2}) \to 0$, this means that $\mathrm{Var}(n^{-1/2}\sum_{j=1}^{k_n} B'_{j,n}) = \mathrm{Var}(Z_n) + o(1)$.

Now consider the variance of $\bar{Z}$. Recall the definition of $\bar{Z}$ in (6.3) and note that its variance satisfies

$$\mathrm{Var}(\bar{Z}) = \mathrm{Var}\Big(\sum_{l=1}^{q} c_l Z(t_l)\Big) = \sum_{l,l'=1}^{q} c_l c_{l'} \mathrm{Cov}(Z(t_l), Z(t_{l'})) = \sum_{l,l'=1}^{q} c_l c_{l'} C(t_l, t_{l'}). \qquad (6.13)$$

The variance of Z_n can be expressed as

$$\mathrm{Var}(Z_n) = \mathrm{Var}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{i,n}\Big) = \frac{1}{n}\sum_{i,i'=1}^{n} \mathrm{Cov}(Z_{i,n}, Z_{i',n}) = \frac{1}{n}\sum_{l,l'=1}^{q} c_l c_{l'} \sum_{i,i'=1}^{n} \mathrm{Cov}(X_{n,i}(t_l), X_{n,i'}(t_{l'}))$$

$$= \frac{1}{n}\sum_{l,l'=1}^{q} c_l c_{l'} \sum_{i,i'=1}^{n} \gamma(i - i', t_l, t_{l'}) = \frac{1}{n}\sum_{l,l'=1}^{q} c_l c_{l'} \sum_{i=-n}^{n} (n - |i|)\gamma(i, t_l, t_{l'}) = \mathrm{Var}(\bar{Z}) + o(1)$$

by the dominated convergence theorem. Consequently, $\mathrm{Var}(n^{-1/2}\sum_{j=1}^{k_n} B'_{j,n}) = \mathrm{Var}(\bar{Z}) + o_{\mathbb{P}}(1)$ follows from the previous discussion.

Finally, verify the Lindeberg condition for the random variables $B'_{j,n}$. Using Hölder's inequality with $p = 1 + \nu/2$ and $q = (2 + \nu)/\nu$ (here ν is the same as in Assumption (A1)) and Markov's inequality yields

$$\mathbb{E}\big[(B'_{j,n})^2 \mathbb{1}\{|B'_{j,n}| > \sqrt{n}\delta\}\big] = \mathbb{E}\big[B_{j,n}^2 \mathbb{1}\{|B_{j,n}| > \sqrt{n}\delta\}\big] \leq \mathbb{E}\big[|B_{j,n}|^{2+\nu}\big]^{2/(2+\nu)} \mathbb{P}\big(|B_{j,n}| > \sqrt{n}\delta\big)^{\nu/(2+\nu)}$$

$$\leq \mathbb{E}\big[|B_{j,n}|^{2+\nu}\big]^{2/(2+\nu)} \mathbb{E}\big[|B_{j,n}|^{2+\nu}\big]^{\nu/(2+\nu)} (n^{1/2}\delta)^{-\nu} = \mathbb{E}\big[|B_{j,n}|^{2+\nu}\big](n^{1/2}\delta)^{-\nu}.$$

Minkowski's inequality gives

$$\mathbb{E}\big[|B_{j,n}|^{2+\nu}\big]^{1/(2+\nu)} \leq \sum_{i=(j-1)(b_n+s_n)+1}^{(j-1)(b_n+s_n)+b_n} \mathbb{E}\big[|Z_{i,n}|^{2+\nu}\big]^{1/(2+\nu)} = O(b_n),$$

where the last estimate follows from

$$\max_{1\leq i\leq n} \mathbb{E}\big[|Z_{i,n}|^{2+\nu}\big]^{1/(2+\nu)} \leq \max_{1\leq i\leq n} \sum_{l=1}^{q} |c_l| \mathbb{E}\big[|X_{n,i}(t_l) - \mu^{(i)}(t_l)|^{2+\nu}\big]^{1/(2+\nu)} < \infty$$

by Assumption (A1). Combining these estimates gives the Lindeberg condition, that is,

$$\frac{1}{n}\sum_{j=1}^{k_n}\mathbb{E}\big[(B'_{j,n})^2\mathbb{1}\{|B'_{j,n}| > \sqrt{n}\delta\}\big] = O(n^{-1}k_n b_n^{2+\nu}n^{-\nu/2}) = O(b_n^{1+\nu}n^{-\nu/2}) = O(n^{1/2-\eta_b(1+\nu)}) = o(1),$$

since by assumption $\eta_b > 1/(2+2\nu)$.

The proof is completed showing that the process $(G_n)_{n\in\mathbb{N}} \subset C([0,1])$ is uniformly ρ-equicontinuous in probability with respect to the metric $\rho(s,t) = |s-t|$. With the notation $Y_{n,j} := X_{n,j} - \mu^{(j)}$ we obtain

$$
\begin{aligned}
\|G_n(s) - G_n(t)\|_2^2 &= \frac{1}{n}\sum_{j,j'=1}^{n}\mathbb{E}\bigg[\big(Y_{n,j}(s) - Y_{n,j}(t)\big)\big(Y_{n,j'}(s) - Y_{n,j'}(t)\big)\bigg] \\
&\leq \frac{2}{n}\sum_{j=0}^{n-1}(n-j)\mathbb{E}\big[\big(Y_{n,1}(s) - Y_{n,1}(t)\big)\big(Y_{n,1+j}(s) - Y_{n,1+j}(t)\big)\big] \qquad (6.14)
\end{aligned}
$$

(note that each row of the array $\{Y_{n,j} : n \in \mathbb{N}, j = 1,\ldots,n\}$ is stationary). Using Assumption (A3), straightforward calculations yield, for any $i = 1,\ldots,n$,

$$\mathbb{E}\big[|Y_{n,i}(s) - Y_{n,i}(t)|^2\big]^{1/2} \lesssim |s-t|.$$

The inequality above and (3.17) in Dehling and Philipp [19] together with Assumption (A4) imply, for any $j = 1,\ldots,n$,

$$\mathbb{E}\big[(Y_{n,1}(s) - Y_{n,1}(t))(Y_{n,1+j}(s) - Y_{n,1+j}(t))\big] \lesssim |s-t|^2\,\varphi(j)^{1/2} \lesssim |s-t|^2\,a^{j/2}.$$

Therefore it follows that

$$\|G_n(s) - G_n(t)\|_2^2 \lesssim |s-t|^2\sum_{j=0}^{\infty}a^{j/2} \lesssim |s-t|^2,$$

and we obtain from Theorem 2.2.4 in Van der Vaart and Wellner [40]

$$
\begin{aligned}
\Big\|\sup_{\rho(s,t)\leq\delta}|G_n(s) - G_n(t)|\Big\|_2 &\lesssim \int_0^{\eta}\sqrt{D(\nu,\rho)}\,d\nu + \delta D(\eta,\rho) \\
&\lesssim \int_0^{\eta}\frac{1}{\sqrt{\nu}}\,d\nu + \frac{\delta}{\eta} = 2\sqrt{\eta} + \frac{\delta}{\eta}.
\end{aligned}
$$

where $D(\eta,\rho) = c\lceil\frac{1}{\eta}\rceil$ is the packing number with respect to the metric $\rho(s,t) = |s-t|$. Markov's

inequality now yields, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{\rho(s,t)\leq\delta} |G_n(s) - G_n(t)| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\left\|\sup_{\rho(s,t)\leq\delta}|G_n(s) - G_n(t)|\right\|_2^2 \lesssim \frac{1}{\varepsilon^2}\left[2\sqrt{\eta} + \frac{\delta}{\eta}\right]$$

and, since $\eta > 0$ is arbitrary, we have

$$\lim_{\delta\searrow 0} \limsup_{n\to\infty} \mathbb{P}\left(\sup_{\rho(s,t)\leq\delta} |G_n(s) - G_n(t)| > \varepsilon\right) = 0.$$

This means that $(G_n)_{n\in\mathbb{N}}$ is asymptotically uniformly ρ-equicontinuous in probability and the assertion in Theorem 2.1 follows. $\qquad\square$

Proof of Theorem 2.2. Theorem 2.1 implies that $n^{-1/2}\sum_{j=1}^n (X_{n,j} - \mu^{(j)}) \rightsquigarrow Z$ in $C([0,1])$, where Z is centered Gaussian with covariance function $\mathrm{Cov}(Z(s), Z(t)) = C(s,t)$ for $s, t \in [0,1]$. Moreover, Assumption (A4) together with Remark 3.6.4 in Samur [38] yield, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} n\mathbb{P}\left(\frac{1}{\sqrt{n}}\|X_{n,1} - \mu^{(1)}\| > \varepsilon\right) = 0. \tag{6.15}$$

The assertion now follows from Corollary 3.5 in Samur [38]. $\qquad\square$

6.3 Proofs of the results of Section 3

Proof of Theorem 3.1. Note that $\mathbb{N} \times \mathbb{N}$ is a directed set and $[0,1]$ is compact. Now by the weak convergence in (3.4), the claim follows directly from Theorem 6.1. $\qquad\square$

Proof of Theorem 3.2. Note that

$$\{\mu_1 - \mu_2 \in C_{\alpha,m,n}\} = \left\{\sup_{t\in[0,1]} \sqrt{n+m}\,\left|\frac{1}{m}\sum_{j=1}^m \widetilde{X}_j(t) - \frac{1}{n}\sum_{j=1}^n \widetilde{Y}_j(t)\right| \leq u_{1-\alpha}\right\},$$

where $\widetilde{X}_j = X_j - \mu_1$ and $\widetilde{Y}_j = Y_j - \mu_2$. Therefore, it follows from the discussion in Section 3.1.1 (applied to the random variables $\widetilde{X}_j$ and $\widetilde{Y}_j$) that

$$\lim_{m,n\to\infty} \mathbb{P}(\mu_1 - \mu_2 \in C_{\alpha,m,n}) = \lim_{m,n\to\infty} \mathbb{P}(T_{m,n} \leq u_{1-\alpha}) = 1 - \alpha.$$

This is the assertion. $\qquad\square$

Proof of Theorem 3.3. Using the notations

$$V_m = \frac{\sqrt{n+m}}{m} \sum_{j=1}^{m} (X_j - \mu_1) \qquad \text{and} \qquad W_n = \frac{\sqrt{n+m}}{n} \sum_{j=1}^{n} (Y_j - \mu_2)$$

write $Z_{m,n} = V_m + W_n$. Next define

$$U_m^{(r)} = \frac{\sqrt{n+m}}{m} \sum_{i=1}^{m-l_1+1} \frac{1}{\sqrt{l_1}} \Big(\sum_{j=i}^{i+l_1-1} X_j(t) - \frac{l_1}{m} \sum_{j=1}^{m} X_j(t) \Big) \xi_i^{(r)},$$

$$G_n^{(r)} = \frac{\sqrt{n+m}}{n} \sum_{i=1}^{n-l_2+1} \frac{1}{\sqrt{l_2}} \Big(\sum_{j=i}^{i+l_2-1} Y_j(t) - \frac{l_2}{n} \sum_{j=1}^{n} Y_j(t) \Big) \varsigma_i^{(r)},$$

then $(Z_{m,n}, \hat{B}_{m,n}^{(1)}, \ldots, \hat{B}_{m,n}^{(R)}) = (V_m, U_m^{(1)}, \ldots, U_m^{(R)}) + (W_n, G_n^{(1)}, \ldots, G_n^{(R)})$, and with similar but easier arguments as in the proof of Theorem 4.3, it can be shown that

$$(V_m, U_m^{(1)}, \ldots, U_m^{(R)}) \rightsquigarrow \tfrac{1}{\sqrt{\lambda}} (Z_1, Z_1^{(1)}, \ldots, Z_1^{(R)}),$$

$$(W_n, G_n^{(1)}, \ldots, G_n^{(R)}) \rightsquigarrow \tfrac{1}{\sqrt{1-\lambda}} (Z_2, Z_2^{(1)}, \ldots, Z_2^{(R)})$$

in $C([0,1])^{R+1}$. Since the two vectors are independent, it directly follows that $(Z_{m,n}, \hat{B}_{m,n}^{(1)}, \ldots, \hat{B}_{m,n}^{(R)}) \rightsquigarrow (Z, Z^{(1)}, \ldots, Z^{(R)})$ in $C([0,1])^{R+1}$. $\qquad\square$

Proof of Theorems 3.4 and 3.5. The statement (3.19) of Theorem 3.5 is a direct consequence of Proposition F.1 in the online supplement of Bücher and Kojadinovic [14] (note that the continuity of the random variable T is implied by the results in Gaenssler et al. [23]).

Now Theorem 3.4 follows from the well-known relation between confidence sets and tests for simple hypotheses (see for example Lehmann [32], p. 214), observing that

$$C_{\alpha,m,n} = \left\{ f \in C([0,1]) \colon \left\| \hat{\mu}_1 - \hat{\mu}_2 - f \right\| \leq \frac{T_{m,n}^{\{\lfloor R(1-\alpha) \rfloor\}}}{\sqrt{n+m}} \right\}$$

is defined by the acceptance region of an asymptotic level α test for the hypotheses $H_{0,f} \colon \mu_1 - \mu_2 \equiv f$ versus $H_{0,f} \colon \mu_1 - \mu_2 \not\equiv f$, which rejects the null hypothesis, whenever

$$\left\| \hat{\mu}_1 - \hat{\mu}_2 - f \right\| > \frac{T_{m,n}^{\{\lfloor R(1-\alpha) \rfloor\}}}{\sqrt{n+m}}.$$

The remaining statement regarding consistency follows from similar arguments as given in Bücher

et al. [13]. Under the alternative, Theorem 3.1 yields, for any $K \in \mathbb{N}$,

$$\lim_{m,n\to\infty} \mathbb{P}(\sqrt{n+m}\,\hat{d}_\infty \geq K) = \lim_{m,n\to\infty} \mathbb{P}(\sqrt{n+m}(\hat{d}_\infty - d_\infty) + \sqrt{n+m}\,d_\infty \geq K) = 1. \quad (6.16)$$

Moreover, under the alternative, Theorem 3.3 together with the continuous mapping theorem yield, for any $R \in \mathbb{N}$,

$$\lim_{K\to\infty} \lim_{m,n\to\infty} \mathbb{P}(T_{m,n}^{\{\lfloor R(1-\alpha)\rfloor\}} > K) \leq \lim_{K\to\infty} \lim_{m,n\to\infty} \mathbb{P}(\max_{r=1,\dots,R} T_{m,n}^{(r)} > K) = 0. \quad (6.17)$$

The assertion in (3.20) now follows from (6.16) and (6.17). $\qquad\square$

Proof of Theorem 3.6. Straightforward calculations show that

$$\begin{aligned}
\mathbf{D}_{m,n} &= \left|\hat{d}_\infty - d_\infty\right| + \sup_{t\in[0,1]} \left|\hat{\mu}_1(t) - \hat{\mu}_2(t) - (\mu_1(t) - \mu_2(t))\right| \\
&\geq \left|\hat{d}_\infty - d_\infty\right| + \sup_{t\in\mathcal{E}^+} \left|\hat{\mu}_1(t) - \hat{\mu}_2(t) - d_\infty\right| \\
&= \sup_{t\in\mathcal{E}^+} \left(\left|\hat{d}_\infty - d_\infty\right| + \left|d_\infty - (\hat{\mu}_1(t) - \hat{\mu}_2(t))\right|\right) \\
&\geq \sup_{t\in\mathcal{E}^+} \left|\hat{d}_\infty - (\hat{\mu}_1(t) - \hat{\mu}_2(t))\right| \\
&\geq \hat{d}_\infty - \inf_{t\in\mathcal{E}^+} \left(\hat{\mu}_1(t) - \hat{\mu}_2(t)\right),
\end{aligned}$$

and (3.4) and Theorem 3.1 yield (observing the order of the sequence $c_{m,n}$) that $\mathbb{P}\left(\frac{c_{m,n}}{\sqrt{m+n}} \geq \mathbf{D}_{m,n}\right) \to 1$ as $m, n \to \infty$. Therefore,

$$\mathbb{P}(\mathcal{E}^+ \subset \hat{\mathcal{E}}_{m,n}^+) = \mathbb{P}\left(\inf_{t\in\mathcal{E}^+}(\hat{\mu}_1(t) - \hat{\mu}_2(t)) \geq \hat{d}_\infty - \frac{c_{m,n}}{\sqrt{m+n}}\right) \geq \mathbb{P}\left(\frac{c_{m,n}}{\sqrt{m+n}} \geq \mathbf{D}_{m,n}\right) \to 1$$

as $m, n \to \infty$. This means that, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{x\in\mathcal{E}^+} \inf_{y\in\hat{\mathcal{E}}_{m,n}^+} |x - y| > \varepsilon\right) \to 0 \quad (6.18)$$

as $m, n \to \infty$. On the other hand, the inequality

$$\begin{aligned}
\sup_{x_{m,n}\in\hat{\mathcal{E}}_{m,n}^+} |\mu_1(x_{m,n}) - \mu_2(x_{m,n}) - d_\infty| &\leq \sup_{x_{m,n}\in\hat{\mathcal{E}}_{m,n}^+} |\mu_1(x_{m,n}) - \mu_2(x_{m,n}) - \hat{d}_\infty| + |\hat{d}_\infty - d_\infty| \\
&\leq \sup_{x_{m,n}\in\hat{\mathcal{E}}_{m,n}^+} \left(|\mu_1(x_{m,n}) - \mu_2(x_{m,n}) - (\hat{\mu}_1(x_{m,n}) - \hat{\mu}_2(x_{m,n}))| + |\hat{d}_\infty - d_\infty| + \frac{c_{m,n}}{\sqrt{m+n}}\right)
\end{aligned}$$

43

shows that

$$\mathbb{P}\Big(\sup_{x_{m,n}\in\hat{\mathcal{E}}^+_{m,n}} |\mu_1(x_{m,n}) - \mu_2(x_{m,n}) - d_\infty| \le 2\frac{c_{m,n}}{\sqrt{m+n}}\Big)$$
$$\ge \mathbb{P}\Big(\sup_{t\in[0,1]} |\mu_1(t) - \mu_2(t) - (\hat{\mu}_1(t) - \hat{\mu}_2(t))| + |\hat{d}_\infty - d_\infty| \le \frac{c_{m,n}}{\sqrt{m+n}}\Big) \to 1$$

as $m, n \to \infty$. Therefore,

$$\sup_{x_{m,n}\in\hat{\mathcal{E}}^+_{m,n}} |\mu_1(x_{m,n}) - \mu_2(x_{m,n}) - d_\infty| \to 0 \tag{6.19}$$

(outer) almost surely, which implies

$$\sup_{x_{m,n}\in\hat{\mathcal{E}}^+_{m,n}} \inf_{y\in\mathcal{E}^+} |x_{m,n} - y| \to 0 \tag{6.20}$$

(outer) almost surely as $m, n \to \infty$. To see this, assume the contrary. Then there would exist a sequence $(x_{m,n})$ in $\hat{\mathcal{E}}^+_{m,n}$ with $\inf_{y\in\mathcal{E}^+} |x_{m,n} - y| \ge \varepsilon$. Now consider the closed set $\mathcal{U}_\varepsilon = \{t \in \hat{\mathcal{E}}^+_{m,n} : |t - s| \ge \varepsilon, \forall s \in \mathcal{E}^+\}$. Then $x_{m,n} \in \mathcal{U}_\varepsilon$ and

$$\max\{\mu_1(t) - \mu_2(t) : t \in \mathcal{U}_\varepsilon\} < d_\infty,$$

which contradicts (6.19) and proves (6.20). Combining (6.18) and (6.20), for any $\varepsilon > 0$,

$$\mathbb{P}\big(d_H(\hat{\mathcal{E}}^+_{m,n}, \mathcal{E}^+) > \varepsilon\big) = \mathbb{P}\Big(\max\Big\{ \sup_{x\in\hat{\mathcal{E}}^+_{m,n}} \inf_{y\in\mathcal{E}^+} |x - y|, \sup_{x\in\mathcal{E}^+} \inf_{y\in\hat{\mathcal{E}}^+_{m,n}} |x - y|\Big\} > \varepsilon\Big) \to 0$$

as $m, n \to \infty$. In the same way it can be shown that $\mathbb{P}\big(d_H(\hat{\mathcal{E}}^-_{m,n}, \mathcal{E}^-) > \varepsilon\big) \to 0$ as $m, n \to \infty$, and the assertion of of Lemma 3.6 follows. $\qquad\square$

Proof of Theorem 3.7. The proof is a direct consequence of the following lemma, which might be of own interest and will be proved at the end of this section. $\qquad\square$

Lemma 6.3. *Let $((X_\alpha, X_\alpha^{(1)}, \dots X_\alpha^{(R)}) : \alpha \in A)$ be a net of random variables in $C([0,1])^{R+1}$ and $((M_\alpha^{(1)}, M_\alpha^{(2)}) : \alpha \in A)$ a net of random elements in the set $K([0,1])$ of all compact subsets of the interval $[0,1]$. Furthermore, let $((X_\alpha, X_\alpha^{(1)}, \dots X_\alpha^{(R)}) : \alpha \in A)$ converge weakly to $(X, X^{(1)}, \dots, X^{(R+1)})$ in $C([0,1])^{R+1}$, where $X^{(1)}, \dots, X^{(R+1)}$ are independent copies of X, and $((M_\alpha^{(1)}, M_\alpha^{(2)}) : \alpha \in A)$ converge in probability to the non random sets $(M^{(1)}, M^{(2)})$ in $K([0,1])^2$,*

that is

$$\mathbb{P}\big(\max \big\{d_H(M_\alpha^{(1)}, M^{(1)}), d_H(M_\alpha^{(2)}, M^{(2)})\big\} > \varepsilon \big) \to 0$$

for all $\varepsilon > 0$. Then the random variables

$$Y_\alpha = \max \Big\{ \max_{t \in M^{(1)}} X_\alpha(t), \max_{t \in M^{(2)}} (-X_\alpha(t)) \Big\} \quad \textit{and} \quad Y_\alpha^{(r)} = \max \Big\{ \max_{t \in M_\alpha^{(1)}} X_\alpha^{(r)}(t), \max_{t \in M_\alpha^{(2)}} (-X_\alpha^{(r)}(t)) \Big\},$$

($r = 1, \ldots, R$) satisfy

$$(Y_\alpha, Y_\alpha^{(1)}, \ldots, Y_\alpha^{(R+1)}) \Rightarrow (Y, Y^{(1)}, \ldots, Y^{(R)})$$

in $\mathbb{R}^{R+1}$ where $Y = \max \big\{ \max_{t \in M^{(1)}} X(t), \max_{t \in M^{(2)}} (-X(t)) \big\}$ and $Y^{(1)}, \ldots, Y^{(R)}$ are independent copies of Y.

Theorem 3.3, Lemma 3.6 and Lemma 6.3 show that

$$\big(\tilde{D}_{m,n}(\mathcal{E}), \ K_{m,n}^{(1)}, \ldots, K_{m,n}^{(R)} \big) \Rightarrow (T(\mathcal{E}), \ T^{(1)}(\mathcal{E}), \ldots, T^{(R)}(\mathcal{E})),$$

in $\mathbb{R}^{R+1}$, where

$$\tilde{D}_{m,n}(\mathcal{E}) = \sqrt{n+m} \ \max \Big\{ \sup_{t \in \mathcal{E}^+} Z_{m,n}, \sup_{t \in \mathcal{E}^-} (-Z_{m,n}) \Big\}$$

and the random variable $Z_{m,n}$ is defined by (3.4). The proofs of Lemmas 6.1 and 6.2 yield that

$$\sqrt{n+m}(\hat{d}_\infty - d_\infty) = \tilde{D}_{m,n}(\mathcal{E}) + o_\mathbb{P}(1),$$

thus completing the proof of Theorem 3.7.

Proof of Lemma 6.3. Only the convergence $Y_\alpha^{(1)} \Rightarrow Y^{(1)}$ is shown, since the general case follows from very similar albeit notationally more complex arguments.

Let $\mathcal{S}$ denote the space $C([0,1]) \times K([0,1])^2$ equipped with the metric d defined by

$$d((f, A, B), (g, A', B')) = \max \big\{ \|f - g\|_\infty, d_H(A, A'), d_H(B, B') \big\}$$

for any $(f, A, A'), (g, B, B') \in S$. Slutsky's theorem then yields $(X_\alpha^{(1)}, M_\alpha^{(1)}, M_\alpha^{(2)}) \rightsquigarrow (X, M^{(1)}, M^{(2)})$ in $\mathcal{S}$. Now consider the function

$$\mathcal{H} \colon \mathcal{S} \to \mathbb{R}, \ (f, A, A') \mapsto \max\{\max_{t \in A} f(t), \max_{t \in A'}(-f(t))\}$$

and show that it is continuous. For this purpose it is sufficient to prove that the maps

$$h \colon \mathcal{S} \to \mathbb{R}, \ (f, A, A') \mapsto \max_{t \in A} f(t) \qquad \text{and} \qquad h^- \colon \mathcal{S} \to \mathbb{R} \ (f, A, A') \mapsto \max_{t \in A'} (-f(t))$$

are continuous (since the maximum of two continuous functions is continuous). Exemplarily, focus on the map h (the corresponding statement for h^- follows from closely analogous arguments). Let $\varepsilon > 0$ be arbitrary and $(f, A, A'), (g, B, B') \in \mathcal{S}$ such that

$$d((f, A, A'), (g, B, B')) < \min\{\delta, \varepsilon/2\},$$

where $\delta > 0$ is chosen such that, for any $t, s \in T$ with $|t - s| < \delta$, the inequality $|g(t) - g(s)| < \varepsilon/2$ holds (note that g is uniformly continuous on $[0, 1]$). Then,

$$|h(f, A, A') - h(g, B, B')| = |\max_{t \in A} f(t) - \max_{t \in B} g(t)| \leq |\max_{t \in A} f(t) - \max_{t \in A} g(t)| + |\max_{t \in A} g(t) - \max_{t \in B} g(t)|,$$

where the first term can be bounded by $\max_{t \in A} |f(t) - g(t)| \leq \|f - g\|_\infty < \varepsilon/2$. For the second term, first consider the case where $\max_{t \in A} g(t) \geq \max_{t \in B} g(t)$ and obtain

$$|\max_{t \in A} g(t) - \max_{t \in B} g(t)| = \max_{t \in A} g(t) - \max_{t \in B} g(t) = g(t_1) - \max_{t \in B} g(t),$$

where $t_1 \in \arg\max_{t \in A} g(t)$. Since $d_H(A, B) < \delta$, there is a $t_2 \in B$ such that $|t_1 - t_2| < \delta$. Thus,

$$g(t_1) - \max_{t \in B} g(t) \leq g(t_1) - g(t_2) < \varepsilon/2 \ .$$

For the second case, use the same arguments to show that $|\max_{t \in A} g(t) - \max_{t \in B} g(t)| < \varepsilon/2$. This yields

$$|h(f, A, A') - h(g, B, B')| < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

verifying that h is continuous.

The discussion at the beginning of this proof shows that

$$Y_\alpha^{(1)} = \max \big\{ \max_{t \in M_\alpha^{(1)}} X_\alpha^{(1)}(t), \ \max_{t \in M_\alpha^{(2)}} (-X_\alpha^{(1)}(t)) \big\} \Rightarrow Y^{(1)}.$$

The proof is complete. $\qquad\square$

6.4 Proofs of the results in Section 4

Proof of Theorem 4.1. Recall the notations (2.2), (4.3) and (4.6) to obtain

$$\sqrt{n}\left(\hat{\mathbb{U}}_n(s,t) - \left(s \wedge s^* - ss^*\right)\left(\mu_1(t) - \mu_2(t)\right)\right)$$
$$= \hat{\mathbb{W}}_n(s,t) + \sqrt{n}\left(E[\hat{\mathbb{U}}_n(s,t)] - \left(s \wedge s^* - ss^*\right)\left(\mu_1(t) - \mu_2(t)\right)\right)$$
$$= \hat{\mathbb{W}}_n(s,t) + o_{\mathbb{P}}(1)$$

uniformly with respect to $(s,t) \in [0,1]^2$. An application of (4.4) shows that the weak convergence

$$\sqrt{n}\left(\hat{\mathbb{U}}_n(s,t) - \left(s \wedge s^* - ss^*\right)\left(\mu_1(t) - \mu_2(t)\right) \colon (s,t) \in [0,1]^2\right) \rightsquigarrow \left(\mathbb{W}(s,t) \colon (s,t) \in [0,1]^2\right)$$

follows. Now the representation

$$\mathbb{D}_n = \sqrt{n}\left(\sup_{(s,t)\in[0,1]^2}\left|\hat{\mathbb{U}}_n(s,t)\right| - \sup_{(s,t)\in[0,1]^2}\left|\left(s \wedge s^* - ss^*\right)\left(\mu_1(t) - \mu_2(t)\right)\right|\right)$$

and Theorem 6.1 yield the assertion of Theorem 4.1. $\qquad\square$

Proof of Theorem 4.2. In order to prove the assertion, use Corollary 2 in Hariz et al. [25]. Let $\mathcal{M}$ be the space of all signed finite measures on $C([0,1])$, define π_t as the canonical projection $C([0,1]) \ni x \to x(t)$ and consider the class $\mathcal{F} = \{\pi_t \colon t \in [0,1]\}$. Note that

$$N(\nu) = \sup_{t\in[0,1]}\left|\int_{C([0,1])}\pi_t(x)\,\nu(dx)\right|$$

defines a semi-norm on $\mathcal{M}$. In particular, if $P = \mathbb{P}^{X_1}$ and $Q = \mathbb{P}^{X_n}$ are the distributions on $C([0,1])$ before and after the change-point, it holds that

$$\int_{C([0,1])}\pi_t(x)(P-Q)(dx) = \int_{C([0,1])}\pi_t(x)(\mathbb{P}^{X_1} - \mathbb{P}^{X_n})(dx) = \int_{\Omega}\pi_t(X_1)d\mathbb{P} - \int_{\Omega}\pi_t(X_n)d\mathbb{P}$$
$$= \int_{\Omega}X_1(t)d\mathbb{P} - \int_{\Omega}X_n(t)d\mathbb{P} = \mathbb{E}[X_1(t)] - \mathbb{E}[X_n(t)] = \mu_1(t) - \mu_2(t),$$

and therefore $N(P-Q) = \|\mu_1 - \mu_2\| > 0$. The estimator of the change-point can now be rewritten as $\tilde{s} = n^{-1}\min\left(\arg\max_{1\leq k<n}\{N(D_k)\}\right)$, where

$$D_k = \frac{k}{n}\left(1 - \frac{k}{n}\right)\left(\frac{1}{k}\sum_{i=1}^{k}\delta_{X_{n,i}} - \frac{1}{n-k}\sum_{i=k+1}^{n}\delta_{X_{n,i}}\right)$$

47

and δ_x denotes the Dirac measure at the point $x \in C([0,1])$. The assertion of Theorem 4.2 follows from Corollary 2 and Remark 2 in Hariz et al. [25] if the following conditions can be verified.

There exist constants $C > 0$ and $\xi > 0$ such that, for any p,

$$\sup_{t \in [0,1]} \sup_{k+p \leq n} \mathbb{E}\left[\left(\sum_{i=k}^{k+p} \left(\pi_t(X_{n,i}) - \mathbb{E}[\pi_t(X_{n,i})]\right)\right)^2\right] \leq Cp^{2-\xi}. \tag{6.21}$$

For any $\varepsilon > 0$,

$$N_{[\,]}(\varepsilon, \mathcal{F}, \|\cdot\|_{\mathcal{G}}) < \infty , \tag{6.22}$$

where $\|\cdot\|_{\mathcal{G}}$ is a norm on a space $\mathcal{G}$ (containing $\mathcal{F}$), which satisfies $(|P(|\pi_t|)| + |Q(|\pi_t|)|) \leq \|\pi_t\|_{\mathcal{G}}$ for any $t \in [0,1]$ and $N_{[\,]}(\varepsilon, \mathcal{F}, \|\cdot\|_{\mathcal{G}})$ denotes the *bracketing number*. Moreover, for any $\nu \in \mathcal{M}$ and $f : C([0,1]) \to \mathbb{R}$, define

$$\nu(f) = \int_{C([0,1])} f(x)\, \nu(dx).$$

Since Assumption (A3) is satisfied with a bounded random variable M, only consider the subspace of $C([0,1])$ that consists of all functions that are Lipschitz continuous with a uniform constant c. Therefore,

$$|\pi_s(x) - \pi_t(x)| = |x(s) - x(t)| \leq c|s - t|$$

for any $s, t \in [0,1]$. It follows from Theorem 2.7.11 in Van der Vaart and Wellner [40] that

$$N_{[\,]}(2\tilde{\varepsilon}\|c\|_{\mathcal{G}}, \mathcal{F}, \|\cdot\|_{\mathcal{G}}) \leq N(\tilde{\varepsilon}, [0,1], |\cdot|).$$

In the equation above $N(\tilde{\varepsilon}, [0,1], |\cdot|)$ denotes the covering number, that is the minimal number of balls of radius $\tilde{\varepsilon}$ needed to cover the unit interval $[0,1]$. Note that

$$N(\tilde{\varepsilon}, [0,1], |\cdot|) = \left\lceil \frac{1}{2\tilde{\varepsilon}} \right\rceil < \infty$$

and therefore (6.22) is satisfied.

Using Assumptions (A1), (A4) and (3.17) in Dehling and Philipp [19], leads to

$$
\sup_{t\in[0,1]} \sup_{k+p\leq n} \mathbb{E}\left[\left(\sum_{i=k}^{k+p}\left(\pi_t(X_{n,i}) - \mathbb{E}[\pi_t(X_{n,i})]\right)\right)^2\right] = \sup_{t\in[0,1]} \sup_{k+p\leq n} \mathbb{E}\left[\left(\sum_{i=k}^{k+p}\left(X_{n,i}(t) - \mathbb{E}[X_{n,i}(t)]\right)\right)^2\right]
$$

$$
= \sup_{t\in[0,1]} \sup_{k+p\leq n} \sum_{i,j=k}^{k+p} \mathrm{Cov}(X_{n,i}(t), X_{n,j}(t)) \lesssim \sup_{k+p\leq n} \sum_{i,j=k}^{k+p} \varphi(|i-j|)^{1/2}
$$

$$
\lesssim \sum_{i=0}^{p}(p+1-i)\varphi(i)^{1/2} \lesssim (p+1)\sum_{i=0}^{\infty} a^{i/2} \lesssim p
$$

which means that (6.21) is satisfied for $\xi = 1$. $\qquad\square$

Proof of Theorem 4.3. The weak convergence

$$
(\hat{\mathbb{V}}_n, \hat{B}_n^{(1)}, \ldots, \hat{B}_n^{(R)}) \rightsquigarrow (\mathbb{V}, \mathbb{V}^{(1)}, \ldots, \mathbb{V}^{(R)}) \tag{6.23}
$$

in $C([0,1]^2)^{R+1}$ will be verified, where $\hat{\mathbb{V}}_n$ is defined in (2.2) and $\mathbb{V}^{(1)}, \ldots, \mathbb{V}^{(R)}$ are independent copies of $\mathbb{V}$ defined in (2.5). The assertion then follows from the continuous mapping theorem. We only consider the case $d_\infty > 0$ since, for $d_\infty = 0$, the assertion follows by similar arguments. The proof of (6.23) is complicated and consists of a series of steps, which are described first.

(1) Replace the estimates $\hat{\mu}_1$ and $\hat{\mu}_2$ in (4.15) by the true functions μ_1 and μ_2 and consider the process

$$
\tilde{B}_n^{(k)}(s,t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} \frac{1}{\sqrt{l}} \left(\sum_{j=i}^{i+l-1}\left(Y_{n,j}(t) - \mu_1(t)\right)\right)\xi_i^{(k)}
$$

$$
+ \sqrt{n}\left(s - \frac{\lfloor sn \rfloor}{n}\right)\frac{1}{\sqrt{l}}\left(\sum_{j=\lfloor sn \rfloor+1}^{\lfloor sn \rfloor+l}\left(Y_{n,j}(t) - \mu_1(t)\right)\right)\xi_{\lfloor sn \rfloor+1}^{(k)},
$$

where $Y_{n,j} = X_{n,j} - (\mu_2 - \mu_1)\mathbb{1}\{j > \lfloor s^*n \rfloor\}$. For this process, show the weak convergence

$$
(\hat{\mathbb{V}}_n, \tilde{B}_n^{(1)}, \ldots, \tilde{B}_n^{(R)}) \rightsquigarrow (\mathbb{V}, \mathbb{V}^{(1)}, \ldots, \mathbb{V}^{(R)}). \tag{6.24}
$$

(2) Next show that

$$
\sup_{(s,t)\in[0,1]^2} |\tilde{B}_n^{(k)}(s,t) - \bar{B}_n^{(k)}(s,t)| = o_\mathbb{P}(1), \tag{6.25}
$$

49

where the process $\bar{B}_n^{(k)}$ is defined by

$$\bar{B}_n^{(k)}(s,t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} \frac{1}{\sqrt{l}} \Big(\sum_{j=i}^{i+l-1} Y_{n,j}(t) - \frac{l}{n} \sum_{j=1}^{n} Y_{n,j}(t) \Big) \xi_i^{(k)}$$

$$+ \sqrt{n} \Big(s - \frac{\lfloor sn \rfloor}{n} \Big) \frac{1}{\sqrt{l}} \Big(\sum_{j=\lfloor sn \rfloor+1}^{\lfloor sn \rfloor+l} Y_{n,j}(t) - \frac{l}{n} \sum_{j=1}^{n} Y_{n,j}(t) \Big) \xi_{\lfloor sn \rfloor+1}^{(k)}. \tag{6.26}$$

(3) Finally establish the assertion (6.23), proving for $k = 1, \ldots, R$, that

$$\sup_{(s,t)\in[0,1]^2} |\hat{B}_n^{(k)}(s,t) - \bar{B}_n^{(k)}(s,t)| = o_{\mathbb{P}}(1) . \tag{6.27}$$

Combining (1)–(3) completes the proof. $\qquad\square$

Proof of (6.24). The weak convergence of the process will be shown through proving the weak convergence of its finite-dimensional distributions and asymptotic tightness.

(A) Convergence of the finite-dimensional distributions. This part uses similar arguments as given in the proof of Theorem 2.1 and detailed arguments are only given when substantial differences occur. For the sake of brevity and simpler notations, consider the case $R = 1$ and prove

$$\hat{Z}_n = \sum_{j=1}^{q} c_j \hat{\mathbb{V}}_n(s_j, u_j) + \sum_{j=1}^{q} d_j \tilde{B}_n^{(1)}(t_j, v_j) \Rightarrow Z = \sum_{j=1}^{q} c_j \mathbb{V}(s_j, u_j) + \sum_{j=1}^{q} d_j \mathbb{V}^{(1)}(t_j, v_j) \tag{6.28}$$

for any $q \in \mathbb{N}$, $(u_1, s_1, v_1, t_1), \ldots, (u_q, s_q, v_q, t_q) \in [0,1]^4$ and arbitrary constants $c_1, d_1, \ldots, c_q, d_q \in \mathbb{R}$. The convergence of the finite-dimensional distributions of the process $(\hat{\mathbb{V}}_n, \tilde{B}_n^{(1)})$ then follows from an application of the Cramér–Wold device. For this purpose define

$$Z_n = \sum_{j=1}^{q} c_j \check{\mathbb{V}}_n(s_j, u_j) + \sum_{j=1}^{q} d_j \check{B}_n^{(1)}(t_j, v_j),$$

where

$$\check{\mathbb{V}}_n(s,t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor sn \rfloor} (X_{n,j} - \mu^{(j)}) \ , \ \ \check{B}_n^{(1)}(s,t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} \frac{1}{\sqrt{l}} \Big(\sum_{j=i}^{i+l-1} \big(X_{n,j}(t) - \mu^{(j)}(t) \big) \Big) \xi_i^{(1)}.$$

Using Assumption (A2) and the fact that $(\xi_i^{(1)})_{i\in\mathbb{N}}$ is independent of $(X_{n,j} : n \in \mathbb{N}, j = 1, \ldots, n)$, it

can be seen that $\|\hat{Z}_n - Z_n\|_2 \lesssim \frac{1}{\sqrt{n}} + \frac{\sqrt{l}}{\sqrt{n}} \to 0$. Therefore (6.28) follows from

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(Z_{i,n} + Z_{i,n}^{(1)} \right) \Rightarrow Z, \tag{6.29}$$

where (note that $X_{n,j} - \mu^{(j)} = Y_{n,j} - \mu_1$)

$$Z_{i,n} = \sum_{j=1}^{q} c_j (Y_{n,i}(u_j) - \mu_1(u_j)) \mathbb{1}\{i \le \lfloor s_j n \rfloor\}, \tag{6.30}$$

$$Z_{i,n}^{(1)} = \xi_i^{(1)} \sum_{j=1}^{q} d_j \frac{1}{\sqrt{l}} \left(\sum_{k=i}^{i+l-1} (Y_{n,k}(v_j) - \mu_1(v_j)) \right) \mathbb{1}\{i \le \lfloor t_j n \rfloor\}. \tag{6.31}$$

A blocking technique will again be utilized to investigate the weak convergence of this sum (see the proof of Theorem 2.1). To this end, let η_b and η_s denote two constants such that

$$\frac{\beta(2+\nu)+1}{2+2\nu} < \eta_b < \eta_s < \frac{1}{2}, \tag{6.32}$$

and define the length of the small and the big subblocks as $s_n = \lfloor n^{1/2-\eta_s} \rfloor$ and $b_n = \lfloor n^{1/2-\eta_b} \rfloor$, respectively. Note that $(\beta(2+\nu)+1)/(2+2\nu) < 1/2$ (since we assumed that $\beta < 1/3$) and that we have $k_n = \lfloor n/(b_n + s_n) \rfloor$ blocks in total. Now introducing the sums

$$B_{j,n} = \sum_{i=(j-1)(b_n+s_n)+1}^{(j-1)(b_n+s_n)+b_n} (Z_{i,n} + Z_{i,n}^{(1)}) \quad \text{and} \quad S_{j,n} = \sum_{i=(j-1)(b_n+s_n)+b_n+1}^{j(b_n+s_n)} (Z_{i,n} + Z_{i,n}^{(1)}), \tag{6.33}$$

of the terms in the j-th big and small subblock, respectively ($j = 1, \ldots, k_n$), the representations (6.4) and (6.5) in the proof of Theorem 2.1 are obtained, adopting appropriate redefinitions of the quantities $B_{j,n}$, $S_{j,n}$ and R_n involved in this representation. It can be shown that

$$|Z_n - 1/\sqrt{n} \sum_{j=1}^{k_n} B_{j,n}| = |1/\sqrt{n} \sum_{j=1}^{k_n} S_{j,n} + 1/\sqrt{n} R_n| = o_{\mathbb{P}}(1) \tag{6.34}$$

$$\operatorname{Var}(Z_n) = \operatorname{Var}(\sum_{j=1}^{k_n} B_{j,n}) + o(1), \tag{6.35}$$

proving that each term on the right-hand side of (6.5), with the exception of the first, converges to zero in probability. Exemplarily, consider the fourth term to indicate the differences to the proof of

Theorem 2.1. Note that

$$\mathbb{E}[S_{j,n}S_{j',n}] = \sum_{i=(j-1)(b_n+s_n)+b_n+1}^{j(b_n+s_n)} \sum_{i'=(j'-1)(b_n+s_n)+b_n+1}^{j'(b_n+s_n)} \mathbb{E}[Z_{i,n}Z_{i',n}] + \mathbb{E}[Z_{i,n}^{(1)}Z_{i',n}^{(1)}]$$

and it follows from (3.17) in Dehling and Philipp [19] together with Assumption (A1) that

$$|\mathbb{E}[Z_{i,n}Z_{i',n}]| \leq \sum_{j,j'=1}^{q} |c_j c_{j'}||\mathrm{Cov}(Y_i(u_j), Y_{i'}(u_{j'}))| \lesssim \varphi(|i-i'|)^{1/2}, \qquad (6.36)$$

as the σ-field generated by a $C([0,1])$-valued random variable X always contains the σ-field generated by $X(t)$ for a fixed $t \in [0,1]$. Moreover, $\mathbb{E}[(Z_{i,n}^{(1)})^2] \leq \mathrm{const} < \infty$, which follows from the representation

$$\mathbb{E}[(Z_{i,n}^{(1)})^2] = \frac{1}{l}\sum_{j,j'=1}^{q} d_j d_{j'} \sum_{i',i''=i}^{i+l-1} \mathrm{Cov}(Y_{n,i'}(v_j), Y_{n,i''}(v_{j'}))\mathbb{1}\{i \leq \lfloor t_j n \rfloor\}\mathbb{1}\{i \leq \lfloor t_{j'} n \rfloor\}$$

and the fact that the covariance in the last expression can be estimated by (here C denotes a constant)

$$\frac{1}{l}\sum_{i',i''=i}^{i+l-1} |\mathrm{Cov}(Y_{n,i'}(v_j), Y_{n,i''}(v_{j'}))| \leq C + \frac{2}{l}\sum_{i'=i}^{i+l-2}\sum_{i''=i'+1}^{i+l-1} |\mathrm{Cov}(Y_{n,i'}(v_j), Y_{n,i''}(v_{j'}))|$$

$$\lesssim C + \frac{2}{l}\sum_{i'=i}^{i+l-2}\sum_{i''=i'+1}^{i+l-1} \varphi(|i'-i''|)^{1/2} \leq C + \frac{2}{l}\sum_{i'=1}^{l-1}(l-i')a^{i'/2} < \infty,$$

using Assumptions (A1), (A4) and (3.17) in Dehling and Philipp [19]). Combining this result with (6.36) leads to

$$\mathbb{E}[S_{j,n}^2] \lesssim \sum_{i,i'=(j-1)(b_n+s_n)+b_n+1}^{j(b_n+s_n)} \varphi(|i-i'|)^{1/2} + s_n \leq 2\sum_{i=0}^{s_n-1}(s_n-i)\varphi(i)^{1/2} + s_n = O(s_n) . \quad (6.37)$$

With similar arguments it follows that $\mathbb{E}[S_{j,n}S_{j',n}] = O(s_n^2\varphi(b_n)^{1/2})$ for $j \neq j'$ since there is at least one big subblock between the observations. Hence,

$$\frac{1}{n}\sum_{j,j'=1}^{k_n} \mathbb{E}[S_{j,n}S_{j',n}] = O(n^{-1}k_n s_n) + O(n^{-1}k_n^2 s_n^2\varphi(b_n)^{1/2}) = O(b_n^{-1}s_n) + O(nb_n^{-2}s_n^2 a^{b_n/2}) = o(1)$$

as $b_n^{-1}s_n \to 0$ and $na^{b_n/2} \to 0$. It can be shown by similar arguments that the second, third, fifth and

52

sixth term in (6.5) are of order $o(1)$ and (6.34) and (6.35) therefore follow.

Equation (6.34) implies that it suffices to show that $n^{-1/2} \sum_{j=1}^{k_n} B_{j,n}$ converges in distribution to Z to prove (6.29). Using similar arguments as in the proof of Theorem 2.1, it can be shown that

$$
\left| \mathbb{E}\Big[\prod_{j=1}^{k_n} \psi_{j,n}(t) \Big] - \prod_{j=1}^{k_n} \mathbb{E}\left[\psi_{j,n}(t) \right] \right| \lesssim k_n \varphi(s_n) = O(k_n a^{s_n}) = o(1) \,,
$$

where $\mathbb{E}[\ \prod_{j=1}^{k_n} \psi_{j,n}(t)]$ is the characteristic function of $n^{-1/2} \sum_{j=1}^{k_n} B_{j,n}$, $\prod_{j=1}^{k_n} \mathbb{E}\left[\psi_{j,n}(t) \right]$ is the characteristic function of $n^{-1/2} \sum_{j=1}^{k_n} B'_{j,n}$ and $B'_{1,n}, \ldots, B'_{k_n,n}$ are independent random variables such that $B_{j,n}$ and $B'_{j,n}$ have the same distribution for any $j = 1, \ldots k_n$. Therefore (6.29) follows from

$$
n^{-1/2} \sum_{j=1}^{k_n} B'_{j,n} \Rightarrow Z, \tag{6.38}
$$

which can be established by the Lindeberg–Feller central limit theorem for triangular arrays. Similar arguments as given in the discussion following (6.12) give

$$
\mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{k_n} B'_{j,n} \right) = \mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{k_n} B_{j,n} \right) + o(1) = \mathrm{Var}(Z_n) + o(1). \tag{6.39}
$$

We now show that $\mathrm{Var}(Z_n)$ converges to the variance of the random variable Z defined in (6.28), which is given by

$$
\begin{aligned}
\mathrm{Var}(Z) &= \mathrm{Var}\Big(\sum_{j=1}^{q} c_j \mathbb{V}(s_j, u_j) \Big) + \mathrm{Var}\Big(\sum_{j=1}^{q} d_j \mathbb{V}^{(1)}(t_j, v_j) \Big) \\
&= \sum_{j,j'=1}^{q} c_j c_{j'} (s_j \wedge s_{j'}) C(u_j, u_{j'}) + \sum_{j,j'=1}^{q} d_j d_{j'} (t_j \wedge t_{j'}) C(v_j, v_{j'}) \,.
\end{aligned} \tag{6.40}
$$

On the other hand (observing that $\mathrm{Cov}(Z_{i,n}, Z^{(1)}_{i',n}) = 0$),

$$
\mathrm{Var}(Z_n) = \frac{1}{n} \sum_{i,i'=1}^{n} \Big(\mathbb{E}[Z_{i,n} Z_{i',n}] + \mathbb{E}[Z^{(1)}_{i,n} Z^{(1)}_{i',n}] \Big) = \frac{1}{n} \sum_{i,i'=1}^{n} \mathbb{E}[Z_{i,n} Z_{i',n}] + \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[(Z^{(1)}_{i,n})^2] \,.
$$

$$
\tag{6.41}
$$

The second term in this expression satisfies

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(Z_{i,n}^{(1)})^2] &=\frac{1}{n}\sum_{i=1}^{n}\sum_{j,j'=1}^{q}d_jd_{j'}\frac{1}{l}\sum_{i_1',i_2'=i}^{i+l-1}\mathrm{Cov}(Y_{i_1'}(v_j),Y_{i_2'}(v_{j'}))\mathbb{1}\big\{i\le\lfloor(t_j\wedge t_{j'})n\rfloor\big\}\\
&=\sum_{j,j'=1}^{q}d_jd_{j'}\frac{\lfloor(t_j\wedge t_{j'})n\rfloor}{n}\sum_{i'=-(l-1)}^{l-1}\frac{l-|i'|}{l}\,\gamma(i',v_j,v_{j'})\\
&\to\sum_{j,j'=1}^{q}d_jd_{j'}(t_j\wedge t_{j'})\sum_{i=-\infty}^{\infty}\gamma(i,v_j,v_{j'})=\sum_{j,j'=1}^{q}d_jd_{j'}(t_j\wedge t_{j'})C(v_j,v_{j'}),
\end{aligned}
$$

where the dominated convergence theorem was used in the last step. For the first term in (6.41), assume without loss of generality that $s_j\le s_{j'}$ and note that

$$
\begin{aligned}
\frac{1}{n}\sum_{i,i'=1}^{n}\mathbb{E}[Z_{i,n}Z_{i',n}] &=\frac{1}{n}\sum_{j,j'=1}^{q}c_jc_{j'}\sum_{i=1}^{\lfloor s_jn\rfloor}\sum_{i'=1}^{\lfloor s_{j'}n\rfloor}\mathrm{Cov}(Y_i(u_j),Y_{i'}(u_{j'}))\\
&=\sum_{j,j'=1}^{q}c_jc_{j'}\left(\frac{1}{n}\sum_{i=1}^{\lfloor s_jn\rfloor}\sum_{i'=1}^{\lfloor s_jn\rfloor}\gamma(i-i',u_j,u_{j'})+\frac{1}{n}\sum_{i=1}^{\lfloor s_jn\rfloor}\sum_{i'=\lfloor s_jn\rfloor+1}^{\lfloor s_{j'}n\rfloor}\mathrm{Cov}(Y_i(u_j),Y_{i'}(u_{j'}))\right)\\
&=\sum_{j,j'=1}^{q}c_jc_{j'}s_j\sum_{i=-\infty}^{\infty}\gamma(i,u_j,u_{j'})+o(1)=\sum_{j,j'=1}^{q}c_jc_{j'}s_j\,C(u_j,u_{j'})+o(1),
\end{aligned}
$$

where the dominated convergence theorem was used again for the first term and the bound

$$
\frac{1}{n}\sum_{i=1}^{\lfloor s_jn\rfloor}\sum_{i'=\lfloor s_jn\rfloor+1}^{\lfloor s_{j'}n\rfloor}|\mathrm{Cov}(Y_i(u_j),Y_{i'}(u_{j'}))|\lesssim\frac{1}{n}\sum_{i=1}^{\lfloor s_{j'}n\rfloor-1}i\varphi(i)^{1/2}\le\frac{1}{n}\sum_{i=1}^{\lfloor s_{j'}n\rfloor-1}ia^{i/2}=o(1)
$$

for the second term. Observing (6.39)–(6.41), it follows that $\mathrm{Var}(n^{-1/2}\sum_{j=1}^{k_n}B'_{j,n})=\mathrm{Var}(Z)+o(1)$ as postulated.

For a proof of the Lindeberg condition, use Hölder's inequality (with $p=1+\nu/2$ and $q=(2+\nu)/\nu$, where ν is the same as in Assumption (A1)) and Markov's inequality to obtain

$$
\begin{aligned}
\Delta_n &=\frac{1}{n}\sum_{j=1}^{k_n}\mathbb{E}\big[(B'_{j,n})^2\mathbb{1}\{|B'_{j,n}|>\sqrt{n}\delta\}\big]=\frac{1}{n}\sum_{j=1}^{k_n}\mathbb{E}\big[B_{j,n}^2\mathbb{1}\{|B_{j,n}|>\sqrt{n}\delta\}\big]\\
&\le\frac{1}{n}\sum_{j=1}^{k_n}\mathbb{E}\big[|B_{j,n}|^{2+\nu}\big]^{2/(2+\nu)}\mathbb{P}\big(|B_{j,n}|>\sqrt{n}\delta\big)^{\nu/(2+\nu)}\le\frac{1}{n}\sum_{j=1}^{k_n}\mathbb{E}\big[|B_{j,n}|^{2+\nu}\big](n^{1/2}\delta)^{-\nu}.\quad(6.42)
\end{aligned}
$$

Now, observing the definition of $Z_{i,n}$, $Z_{i,n}^{(1)}$ in (6.30), (6.31), respectively, and Assumption (A1), it

can be seen that $\max_{1 \leq i \leq n} \mathbb{E}\big[|Z_{i,n}|^{2+\nu}\big]^{1/(2+\nu)} < \infty$ and $\max_{1 \leq i \leq n} \mathbb{E}\big[|Z_{i,n}^{(1)}|^{2+\nu}\big]^{1/(2+\nu)} \lesssim \sqrt{l}$, and the Lindeberg condition follows from (6.42) observing that the representation of $B_{j,n}$ in (6.33) and Minkowski's inequality give $\mathbb{E}\big[|B_{j,n}|^{2+\nu}\big]^{1/(2+\nu)} = O(b_n l^{1/2})$, that is,

$$\Delta_n = O\Big(\frac{k_n b_n^{2+\nu} l^{(2+\nu)/2}}{n^{1+\nu/2}}\Big) = O\Big(\frac{b_n^{1+\nu} l^{(2+\nu)/2}}{n^{\nu/2}}\Big) = O(n^{1/2-\eta_b(1+\nu)} l^{(2+\nu)/2}) = O(n^{1/2-\eta_b(1+\nu)} n^{\beta(2+\nu)/2}),$$

which converges to zero by the assumption (6.32).

(B) Asymptotic tightness: Since $(\hat{\mathbb{V}}_n)_{n \in \mathbb{N}}$ converges weakly to $\mathbb{V}$, the process $(\hat{\mathbb{V}}_n)_{n \in \mathbb{N}}$ is asymptotically tight and it remains to show that $(\tilde{B}_n^{(k)})_{n \in \mathbb{N}}$ is asymptotically tight (for any $k = 1, \ldots, R$, note that marginal asymptotic tightness implies joint asymptotic tightness).

Let $s, t \in [0, 1]$ be arbitrary and define $\varepsilon_{n,j} = Y_{n,j} - \mu_1$ for any $n \in \mathbb{N}$ and $j = 1, \ldots, n$, then, since $\xi_1^{(k)}, \ldots, \xi_n^{(k)}$ are independent of $\varepsilon_{n,1}, \ldots, \varepsilon_{n,n}$ with $\mathbb{E}[\xi_j^{(k)}] = 0$ and $\mathrm{Var}(\xi_j^{(k)}) = 1$,

$$
\begin{aligned}
\|\tilde{B}_n^{(k)}(1,s) - \tilde{B}_n^{(k)}(1,t)\|_2^2 =& \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\Big[\frac{1}{l} \sum_{j=i}^{i+l-1} \sum_{j'=i}^{i+l-1} \big(\varepsilon_{n,j}(s) - \varepsilon_{n,j}(t)\big)\big(\varepsilon_{n,j'}(s) - \varepsilon_{n,j'}(t)\big)\Big] \\
\leq& \frac{2}{l} \sum_{j=0}^{l-1} (l-j) \mathbb{E}\big[\big(\varepsilon_{n,1}(s) - \varepsilon_{n,1}(t)\big)\big(\varepsilon_{n,1+j}(s) - \varepsilon_{n,1+j}(t)\big)\big], \quad (6.43)
\end{aligned}
$$

utilizing the fact that each row of the array $(\varepsilon_{n,j}\colon n \in \mathbb{N}, j = 1, \ldots, n)$ is stationary. Assumption (A3) implies $\mathbb{E}\big[|\varepsilon_{n,i}(s) - \varepsilon_{n,i}(t)|^2\big]^{1/2} \lesssim |s - t|$, and (3.17) in Dehling and Philipp [19] yields

$$\|\tilde{B}_n^{(k)}(1,s) - \tilde{B}_n^{(k)}(1,t)\|_2^2 \lesssim \frac{2}{l} \sum_{j=0}^{l-1} (l-j)|s-t|^2 \, a^{j/2} \lesssim |s-t|^2 \sum_{j=0}^{\infty} a^{j/2} \lesssim |s-t|^2 \, .$$

Now consider the metric $\rho(s,t) = |s - t|$ on the interval $[0,1]$ and define $D(\eta, \rho) = \big\lceil \frac{1}{\eta} \big\rceil$ as the corresponding packing number, then Theorem 2.2.4 in Van der Vaart and Wellner [40] shows that

$$\Big\| \sup_{\rho(s,t) \leq \delta} |\tilde{B}_n^{(k)}(1,s) - \tilde{B}_n^{(k)}(1,t)| \Big\|_2 \leq K'\Big[\int_0^\eta \sqrt{D(\nu, \rho)} \, d\nu + \delta D(\eta, \rho) \Big] \lesssim 2\sqrt{\eta} + \frac{\delta}{\eta},$$

and Markov's inequality now yields, for any $\varepsilon > 0$,

$$\mathbb{P}\Big(\sup_{\rho(s,t) \leq \delta} |\tilde{B}_n^{(k)}(1,s) - \tilde{B}_n^{(k)}(1,t)| > \varepsilon \Big) \lesssim \frac{1}{\varepsilon^2} \Big[2\sqrt{\eta} + \frac{\delta}{\eta} \Big].$$

Since $\eta > 0$ is arbitrary, it follows that the process $(\tilde{B}_n^{(k)}(1,\cdot))_{n \in \mathbb{N}}$ is asymptotically uniformly ρ-equicontinuous in probability. Moreover, the finite-dimensional distributions of $(\tilde{B}_n^{(k)}(1,\cdot))_{n \in \mathbb{N}}$

converge weakly to the finite-dimensional distribution of $\mathbb{V}^{(k)}(1,\cdot)$ and therefore it follows that $\tilde{B}_n^{(k)}(1,\cdot) \rightsquigarrow \mathbb{V}^{(k)}(1,\cdot)$ in $C([0,1])$ (see the discussion at the beginning of the proof of Theorem 2.1).

The asymptotic tightness of the process $\tilde{B}_n^{(k)}$ in $C([0,1]^2)$ is now a consequence of Corollary 3.5 in Samur [38]. To be precise, note that $\tilde{B}_n^{(k)}(1,t) = n^{-1/2}\sum_{i=1}^n D_{n,i}^{(k)}(t)$, where

$$D_{n,i}^{(k)}(t) = \frac{1}{\sqrt{l}}\Big(\sum_{j=i}^{i+l-1}\big(Y_{n,j}(t) - \mu_1(t)\big)\Big)\xi_i^{(k)} = \frac{1}{\sqrt{l}}\Big(\sum_{j=i}^{i+l-1}\big(X_{n,j}(t) - \mu^{(j)}(t)\big)\Big)\xi_i^{(k)}.$$

The array $(D_{n,i}^{(k)} : n \in \mathbb{N}, i = 1, \dots, n)$ is φ-mixing and (by Assumption (A1) and Markov's inequality)

$$n\mathbb{P}\Big(\frac{1}{\sqrt{n}}\|D_{n,1}^{(k)}\| > \varepsilon\Big) \lesssim \frac{1}{\varepsilon^4}\frac{1}{n}\frac{1}{l^2}\mathbb{E}\Big[\Big(\sum_{j=1}^l \|X_{n,j} - \mu^{(j)}\|\Big)^4\Big] \lesssim \frac{1}{\varepsilon^4}\frac{l^2}{n} = \frac{1}{\varepsilon^4}n^{2\beta-1}.$$

Therefore, since $\beta < 1/3$ by assumption, $\lim_{n\to\infty} n\mathbb{P}(n^{-1/2}\|D_{n,1}^{(k)}\| > \varepsilon) = 0$. By the previous discussion and (4.16), use Corollary 3.5 in Samur [38] to obtain $(\tilde{B}_n^{(k)})_{n\in\mathbb{N}} \rightsquigarrow \mathbb{V}$ in $C([0,1]^2)$, which finally implies that $(\tilde{B}_n^{(k)})_{n\in\mathbb{N}}$ is asymptotically tight. $\qquad\square$

Proof of (6.25). Note that

$$\|\bar{B}_n^{(k)} - \tilde{B}_n^{(k)}\| \leq \Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor sn\rfloor}\frac{1}{\sqrt{l}}\Big(l\mu_1(t) - \frac{l}{n}\sum_{j=1}^n Y_{n,j}(t)\Big)\xi_i^{(k)}\Big\|$$

$$+ \sup_{s,t\in[0,1]}\Big|\sqrt{n}\Big(s - \frac{\lfloor sn\rfloor}{n}\Big)\frac{1}{\sqrt{l}}\Big(l\mu_1(t) - \frac{l}{n}\sum_{j=1}^n Y_{n,j}(t)\Big)\xi_{\lfloor sn\rfloor+1}^{(k)}\Big|$$

$$\leq \sup_{s\in[0,1]}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor sn\rfloor}\xi_i^{(k)}\Big| \times \frac{\sqrt{l}}{\sqrt{n}}\sup_{t\in[0,1]}\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^n(X_{n,j}(t) - \mu^{(j)}(t))\Big|$$

$$+ \frac{1}{\sqrt{n}}\max_{i=1}^n|\xi_i^{(k)}| \times \frac{\sqrt{l}}{\sqrt{n}}\sup_{t\in[0,1]}\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^n\big(X_{n,j}(t) - \mu^{(j)}(t)\big)\Big|.$$

Therein, $\sup_{s\in[0,1]}|n^{-1/2}\sum_{i=1}^{\lfloor sn\rfloor}\xi_i^{(k)}| = O_{\mathbb{P}}(1)$, $\max_{i=1}^n|\xi_i^{(k)}| = O_{\mathbb{P}}(\sqrt{\log n})$, while all other terms are of order $O_{\mathbb{P}}(\sqrt{l/n}) = o_{\mathbb{P}}(1)$, by Theorem 2.1. $\qquad\square$

Proof of (6.27). Recall the definition of $\bar{B}_n^{(k)}(s,t)$ in (6.26) and define

$$\check{B}_n^{(k)}(s,t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} \frac{1}{\sqrt{l}} \left(\sum_{j=i}^{i+l-1} \check{Y}_{n,j}(t) - \frac{l}{n} \sum_{j=1}^{n} \check{Y}_{n,j}(t) \right) \xi_i^{(k)}$$

$$+ \sqrt{n}\left(s - \frac{\lfloor sn \rfloor}{n}\right) \frac{1}{\sqrt{l}} \left(\sum_{j=\lfloor sn \rfloor+1}^{\lfloor sn \rfloor+l} \check{Y}_{n,j}(t) - \frac{l}{n} \sum_{j=1}^{n} \check{Y}_{n,j}(t) \right) \xi_{\lfloor sn \rfloor+1}^{(k)},$$

where $\check{Y}_{n,j} = X_{n,j} - (\hat{\mu}_2 - \hat{\mu}_1)\mathbb{1}\{j > \lfloor s^*n \rfloor\}$. Then,

$$\sup_{(s,t)\in[0,1]^2} |\check{B}_n^{(k)}(s,t) - \bar{B}_n^{(k)}(s,t)| \leq U_n^{(1,1)} + U_n^{(1,2)} + U_n^{(2)} = o_{\mathbb{P}}(1), \tag{6.44}$$

where

$$U_n^{(1,1)} = \sup_{(s,t)\in[0,1]^2} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} \frac{1}{\sqrt{l}} \left(\sum_{j=i}^{i+l-1} \varepsilon_n(t)\mathbb{1}\{j > \lfloor s^*n \rfloor\} \right) \xi_i^{(k)} \right|, \tag{6.45}$$

$$U_n^{(1,2)} = \sup_{(s,t)\in[0,1]^2} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} \frac{1}{\sqrt{l}} \left(\frac{l}{n} \sum_{j=1}^{n} \varepsilon_n(t)\mathbb{1}\{j > \lfloor s^*n \rfloor\} \right) \xi_i^{(k)} \right|, \tag{6.46}$$

$$U_n^{(2)} = \sup_{(s,t)\in[0,1]^2} \left| \frac{\sqrt{n}}{\sqrt{l}}\left(s - \frac{\lfloor sn \rfloor}{n}\right) \left(\sum_{j=\lfloor sn \rfloor+1}^{\lfloor sn \rfloor+l} \varepsilon_n(t)\mathbb{1}\{j > \lfloor s^*n \rfloor\} - \frac{l}{n} \sum_{j=1}^{n} \varepsilon_n(t)\mathbb{1}\{j > \lfloor s^*n \rfloor\} \right) \xi_{\lfloor sn \rfloor+1}^{(k)} \right|,$$

and $\varepsilon_n = \hat{\mu}_2 - \mu_2 - (\hat{\mu}_1 - \mu_1)$. To prove (6.44), it will be shown that all terms on the right-hand side of (6.44) converge to zero in probability, concentrating on $U_n^{(1,1)}$ and $U_n^{(1,2)}$ for the sake of brevity (the term $U_n^{(2)}$ can be treated similarly). At the end of this proof it will be verified that

$$\sup_{t\in[0,1]} |\varepsilon_n(t)| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right). \tag{6.47}$$

Direct calculations observing (6.47) yield

$$U_n^{(1,2)} \leq \sup_{s\in[0,1]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} \xi_i^{(k)} \right| \times \sqrt{l} \sup_{t\in[0,1]} |\varepsilon_n(t)| = O_{\mathbb{P}}\left(\sqrt{\frac{l}{n}}\right) = o_{\mathbb{P}}(1). \tag{6.48}$$

For the first term on the right side of (6.44) similar arguments yield

$$U_n^{(1,1)} = \sup_{t\in[0,1]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor s^*n \rfloor} \frac{1}{\sqrt{l}} \left(\sum_{j=i}^{i+l-1} \varepsilon_n(t)\mathbb{1}\{j > \lfloor s^*n \rfloor\} \right) \xi_i^{(k)} \right| + o_{\mathbb{P}}(1)$$

57

$$\leq \sum_{i=\lfloor s^*n\rfloor-l+1}^{\lfloor s^*n\rfloor} |\xi_i^{(k)}| \frac{\sqrt{l}}{\sqrt{n}} \sup_{t\in[0,1]} |\varepsilon_n(t)| + o_\mathbb{P}(1) = O_\mathbb{P}\Big(\frac{l^{3/2}}{n}\Big) + o_\mathbb{P}(1) = O_\mathbb{P}(n^{3/2\beta-1}) + o_\mathbb{P}(1) = o_\mathbb{P}(1),$$

which follows using (6.47), $\beta < 1/3$ (by assumption) and $\sum_{i=\lfloor s^*n\rfloor-l+1}^{\lfloor s^*n\rfloor} |\xi_i^{(k)}| = O_\mathbb{P}(l)$, the latter relation implied by Markov's inequality.

Therefore (6.44) holds and observing the definition of $\hat{B}_n^{(k)}$ in (4.15), it is next shown that

$$\sup_{(s,t)\in[0,1]^2} |\hat{B}_n^{(k)}(s,t) - \check{B}_n^{(k)}(s,t)| \leq Z_n^{(1)} + Z_n^{(2)} + Z_n^{(3)} + Z_n^{(4)} = o_\mathbb{P}(1), \qquad (6.49)$$

where

$$Z_n^{(1)} = \sup_{(s,t)\in[0,1]^2} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn\rfloor} \frac{1}{\sqrt{l}} \sum_{j=i}^{i+l-1} (\hat{\mu}_1(t) - \hat{\mu}_2(t))(\mathbb{1}\{j > \lfloor s^*n\rfloor\} - \mathbb{1}\{j > \lfloor \hat{s}n\rfloor\})\xi_i^{(k)} \Big|$$

$$Z_n^{(2)} = \sup_{(s,t)\in[0,1]^2} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn\rfloor} \frac{1}{\sqrt{l}} \frac{l}{n} \sum_{j=1}^{n} (\hat{\mu}_1(t) - \hat{\mu}_2(t))(\mathbb{1}\{j > \lfloor s^*n\rfloor\} - \mathbb{1}\{j > \lfloor \hat{s}n\rfloor\})\xi_i^{(k)} \Big|,$$

$$Z_n^{(3)} = \sup_{(s,t)\in[0,1]^2} \Big| \Big(s - \frac{\lfloor sn\rfloor}{n}\Big)\frac{\sqrt{n}}{\sqrt{l}} \sum_{j=\lfloor sn\rfloor+1}^{\lfloor sn\rfloor+l} (\hat{\mu}_1(t) - \hat{\mu}_2(t))(\mathbb{1}\{j > \lfloor s^*n\rfloor\} - \mathbb{1}\{j > \lfloor \hat{s}n\rfloor\})\xi_{\lfloor sn\rfloor+1}^{(k)} \Big|,$$

$$Z_n^{(4)} = \sup_{(s,t)\in[0,1]^2} \Big| \Big(s - \frac{\lfloor sn\rfloor}{n}\Big)\frac{\sqrt{n}}{\sqrt{l}} \frac{l}{n} \sum_{j=1}^{n} (\hat{\mu}_1(t) - \hat{\mu}_2(t))(\mathbb{1}\{j > \lfloor s^*n\rfloor\} - \mathbb{1}\{j > \lfloor \hat{s}n\rfloor\})\Big)\xi_{\lfloor sn\rfloor+1}^{(k)} \Big|.$$

As $\big|\mathbb{1}\{j > \lfloor s^*n\rfloor\} - \mathbb{1}\{j > \lfloor \hat{s}n\rfloor\}\big| = \mathbb{1}\{\lfloor(\hat{s}\wedge s^*)n\rfloor < j \leq \lfloor(\hat{s}\vee s^*)n\rfloor\}$, it follows that

$$Z_n^{(2)} \leq \sup_{(s,t)\in[0,1]^2} \Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn\rfloor} \frac{|\xi_i^{(k)}|}{\sqrt{l}} \frac{l}{n} \sum_{j=1}^{n} |\hat{\mu}_1(t) - \hat{\mu}_2(t)| \mathbb{1}\{\lfloor(\hat{s}\wedge s^*)n\rfloor < j \leq \lfloor(\hat{s}\vee s^*)n\rfloor\}\Big) \quad (6.50)$$

$$\leq \|\hat{\mu}_1 - \hat{\mu}_2\|_\infty \Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} |\xi_i^{(k)}|\Big) \frac{\sqrt{l}}{n} \big(n|\hat{s} - s^*| + 1\big) = O_\mathbb{P}\Big(\sqrt{\frac{l}{n}}\Big) = o_\mathbb{P}(1),$$

using that $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} |\xi_i^{(k)}| = O_\mathbb{P}(\sqrt{n})$ (by Markov's inequality) and $\|\hat{\mu}_1 - \hat{\mu}_2\| = O_\mathbb{P}(1)$ (implied by (6.47)).

For the first term on the right-side of (6.49), use similar arguments as in (6.50) to obtain

$$Z_n^{(1)} \leq \|\hat{\mu}_1 - \hat{\mu}_2\|_\infty \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{\sqrt{l}}\Big(\sum_{j=i}^{i+l-1} \mathbb{1}\{\lfloor(\hat{s}\wedge s^*)n\rfloor < j \leq \lfloor(\hat{s}\vee s^*)n\rfloor\}\Big)|\xi_i^{(k)}|$$

$$= \left\| \hat{\mu}_1 - \hat{\mu}_2 \right\|_\infty \frac{1}{\sqrt{n}} \sum_{i=\lfloor(\hat{s}\wedge s^*)n\rfloor-l+2}^{\lfloor(\hat{s}\vee s^*)n\rfloor} \frac{1}{\sqrt{l}} \left(\sum_{j=i}^{i+l-1} \mathbb{1}\{\lfloor(\hat{s}\wedge s^*)n\rfloor < j \leq \lfloor(\hat{s}\vee s^*)n\rfloor\} \right) |\xi_i^{(k)}|$$

$$\leq \left\| \hat{\mu}_1 - \hat{\mu}_2 \right\|_\infty \frac{\sqrt{l}}{\sqrt{n}} \sum_{i=\lfloor(\hat{s}\wedge s^*)n\rfloor-l+2}^{\lfloor(\hat{s}\vee s^*)n\rfloor} |\xi_i^{(k)}| = O_\mathbb{P}(l^{3/2}n^{-1/2}) = O_\mathbb{P}(n^{3\beta/2-1/2}) = o_\mathbb{P}(1),$$

since $\sum_{i=\lfloor(\hat{s}\wedge s^*)n\rfloor-l+2}^{\lfloor(\hat{s}\vee s^*)n\rfloor} |\xi_i^{(k)}| = O_\mathbb{P}(l)$, which follows from $|\hat{s} - s^*| = O_\mathbb{P}(n^{-1})$. Similarly, one can show $Z_n^{(3)} = o_\mathbb{P}(1)$, $Z_n^{(4)} = o_\mathbb{P}(1)$ and therefore (6.49) holds, which implies (observing (6.44)) the assertion (6.27).

In order to prove the remaining statement (6.47), note that

$$\sup_{t\in[0,1]} \left| \hat{\mu}_1(t) - \mu_1(t) \right| \leq \frac{1}{\sqrt{n}} \frac{n}{\lfloor \hat{s}n \rfloor} \left(Q_n^{(1)} + Q_n^{(2)} \right), \tag{6.51}$$

where

$$Q_n^{(1)} = \sup_{t\in[0,1]} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor(\hat{s}\wedge s^*)n\rfloor} \left(X_{n,j}(t) - \mu_1(t) \right) \right|,$$

$$Q_n^{(2)} = \sup_{t\in[0,1]} \left| \frac{1}{\sqrt{n}} \sum_{j=\lfloor(\hat{s}\wedge s^*)n\rfloor+1}^{\lfloor(\hat{s}\vee s^*)n\rfloor} \left(X_{n,j}(t) - \mu_1(t) \right) \right|.$$

Recall the definition of $\hat{\mathbb{V}}_n$ in (2.2), then Theorem 2.2 and the extended continuous mapping theorem (see Theorem 1.11.1 in Van der Vaart and Wellner [40]) yield

$$Q_n^{(1)} = \sup_{t\in[0,1]} \left| \hat{\mathbb{V}}_n \left(\lfloor(\hat{s}\wedge s^*)n\rfloor/n, t \right) \right| = O_\mathbb{P}(1) . \tag{6.52}$$

On the other hand,

$$\lim_{p\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(Q_n^{(2)} > p \right) \leq \lim_{p\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(Q_n^{(2)} > p , \, \left| \lfloor \hat{s}n \rfloor - \lfloor s^*n \rfloor \right| \leq p \right)$$

$$+ \lim_{p\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\left| \lfloor \hat{s}n \rfloor - \lfloor s^*n \rfloor \right| > p \right).$$

The second term in this inequality is of order $o_\mathbb{P}(1)$, since $|\hat{s} - s^*| = O_\mathbb{P}(n^{-1})$. For the first term use Markov's inequality and Assumption (A1) to obtain

$$\lim_{p\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(Q_n^{(2)} > p , \, \left| \lfloor \hat{s}n \rfloor - \lfloor s^*n \rfloor \right| \leq p \right) \leq \lim_{p\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{j=\lfloor s^*n\rfloor-p}^{\lfloor s^*n\rfloor+p} \left(\|X_{n,j}\| + \|\mu_1\| \right) > p \right)$$

$$\lesssim \lim_{p \to \infty} \limsup_{n \to \infty} \frac{1}{\sqrt{n}} = 0 \,.$$

This means that $Q_n^{(2)} = O_{\mathbb{P}}(1)$. Observing the calculations in (6.51) and (6.52) leads to

$$\sup_{t \in [0,1]} \left| \hat{\mu}_1(t) - \mu_1(t) \right| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}} \right) \quad \text{and} \quad \sup_{t \in [0,1]} \left| \hat{\mu}_2(t) - \mu_2(t) \right| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}} \right), \qquad (6.53)$$

the second estimate following similarly. This yields (6.47) and completes the proof of (6.27).   □

Proof of Theorem 4.4. In order to prove the assertion define

$$\mathbf{D}_n = |\hat{d}_\infty - d_\infty| + \sup_{t \in [0,1]} |\hat{\mu}_1(t) - \hat{\mu}_2(t) - (\mu_1(t) - \mu_2(t))|.$$

Observing (6.53) and Corollary 4.1, yields $\mathbb{P}(c_n/\sqrt{n} \geq \mathbf{D}_n) \to 1$ as $n \to \infty$, and the same arguments as given in the proof of Lemma 3.6 show $d_h(\hat{\mathcal{E}}_n^\pm, \mathcal{E}^\pm) \xrightarrow{\mathbb{P}^*} 0$, where the sets $\hat{\mathcal{E}}_n^+$ and $\hat{\mathcal{E}}_n^-$ are now defined in (4.19). Theorem 4.3 and Lemma 6.3 lemma imply

$$(\tilde{D}_n(\mathcal{E}), \, T_n^{(1)}, \ldots, T_n^{(R)}) \Rightarrow (T(\mathcal{E}), \, T^{(1)}, \ldots, T^{(R)})$$

in $\mathbb{R}^{R+1}$, where

$$\tilde{D}_n(\mathcal{E}) = \frac{1}{\hat{s}(1 - \hat{s})} \max \Big\{ \max_{t \in \mathcal{E}^+} \hat{W}_n(\hat{s}, t), \max_{t \in \mathcal{E}^-}(-\hat{W}_n(\hat{s}, t)) \Big\}$$

and the random variable $\hat{W}_n$ is defined by (4.3). From the discussion following Theorem 4.2 and the proofs of Lemma 6.1 and Lemma 6.2, $\tilde{D}_n(\mathcal{E}) = \mathbb{D}_n/[\hat{s}(1 - \hat{s})] + o_{\mathbb{P}}(1) = \sqrt{n}(\hat{d}_\infty - d_\infty) + o_{\mathbb{P}}(1)$, where $\mathbb{D}_n$ is defined by (4.8). This implies the assertion.   □