Testing relevant hypotheses in functional time series via self-normalization

Holger Dette, Kevin Kokot, Stanislav Volgushev

Testing relevant hypotheses in functional time series via self-normalization

Holger Dette, Kevin Kokot

Ruhr-Universität Bochum

Fakultät für Mathematik

Bochum, Germany

e-mail: {holger.dette, kevin.kokot}@rub.de

Stanislav Volgushev

University of Toronto

Department of Statisical Sciences

Toronto, Canada

email: stanislav.volgushev@utoronto.ca

September 14, 2018

Abstract

In this paper we develop methodology for testing relevant hypotheses in a tuning-free way. Our main focus is on functional time series, but extensions to other settings are also discussed. Instead of testing for exact equality, for example for the equality of two mean functions from two independent time series, we propose to test a *relevant* deviation under the null hypothesis. In the two sample problem this means that an L^2-distance between the two mean functions is smaller than a pre-specified threshold. For such hypotheses self-normalization, which was introduced by Shao (2010) and Shao and Zhang (2010) and is commonly used to avoid the estimation of nuisance parameters, is not directly applicable. We develop new self-normalized procedures for testing relevant hypotheses in the one sample, two sample and change point problem and investigate their asymptotic properties. Finite sample properties of the proposed tests are illustrated by means of a simulation study and a data example.

Keywords: self normalization, functional time series, two sample problems, change point analysis, CUSUM, relevant hypotheses

AMS Subject Classification:

1 Introduction

Statistics for functional data has found considerable interest in the last twenty years as documented in the various monographs by Ramsay and Silverman (2005), Ferraty and Vieu (2010)

and Horváth and Kokoszka (2012) among others. The available methodology includes explorative tools such as shift and feature registration, warping or principal components, and methods for statistical inference such as testing of hypotheses and change point analysis. In this context a large portion of the literature attacks the problem of hypotheses testing by considering hypotheses of the form

(1.1) $$H_0 : d = 0 \text{ versus } H_1 : d \neq 0$$

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

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

This list of references is by no means complete but a common feature of all of these references is that they usually address hypotheses of the form (1.1), which we call "classical" hypotheses in the following discussion. However, in many applications one might not be interested in detecting very small deviations of the parameter d from 0 (often the researcher even knows that d is not exactly equal to 0, before any experiments have been carried out). For example, in change

point detection a modification of the statistical analysis for prediction might not be necessary if the difference between the parameters before and after the change point is rather small. This discussion may be viewed as a particular case of the common bias variance trade-off in statistics. Therefore we argue that one should carefully think about the size of the difference in which one is interested. In particular we propose to replace the hypotheses (1.1) by the hypotheses of a *relevant difference,* that is

$$(1.2) \qquad\qquad H_0 : d \leq \Delta \quad \text{versus} \quad H_1 : d > \Delta \,,$$

where Δ is a pre-specified constant representing the "maximal" value for the parameter d, which can be accepted as not scientifically significant. This formulation of the testing problem requires the specification of the threshold $\Delta > 0$, which depends on the specific application. "Classical" hypotheses tests simply use $\Delta = 0$, but we argue that from a practical point of view it might be very reasonable to think about this choice more carefully and to define the size of the change in which one is really interested from a scientific viewpoint.

We also note that the formulation of the testing problem in the form (1.2) avoids the consistency problem mentioned in Berkson (1938), that is: any consistent test will detect any arbitrary small change in the parameters if the sample size is sufficiently large. Moreover, by interchanging the hypotheses, that is considering the hypotheses of *equivalence*

$$(1.3) \qquad\qquad H_0 : d > \Delta \quad \text{versus} \quad H_1 : d \leq \Delta \,,$$

one is able to decide for "small parameter" d at a controlled type I error (for example that the norm d of the difference between the mean functions of two samples is smaller than a given threshold). Hypotheses of the form (1.2) and (1.3) are called *precise hypotheses* or *relevant hypotheses* in the literature [see Berger and Delampady (1987)] and are frequently used in biostatistics. We refer to Chow and Liu (1992) and Wellek (2010) for more details and applications.

In this paper we discuss the problem of testing relevant hypotheses in the context of functional dependent data. We are particularly interested in methods based on self-normalization in order to avoid estimation of the long-run variance or resampling methods. For this purpose we modify the classical approaches to self-normalization based testing proposed by Shao (2010) and Shao and Zhang (2010) in order to make them applicable for testing relevant hypotheses. This modification is of independent interest besides the field of functional data analysis and applicable in many other problems. Section 4 contains a discussion of this approach in a general framework. Our basic idea is explained in Section 2 for the one and two sample case, where it is most transparent. Roughly speaking, we construct an asymptotic confidence interval for the parameter d to obtain tests for hypotheses of the form (1.2) and (1.3). In Section 3 we address the problem of relevant change point analysis by the new way of self-normalization; here an additional challenge arises from the fact that the change point location is unknown and needs to be estimated. Some finite sample results are presented in Section 5, where we also illustrate our approach analyzing a small data example. Finally, all technical details are deferred to the Appendix (see Section 6).

2 Relevant hypotheses and self normalization

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

2.1 One sample problems

Let $\{X_n\}_{n \in \mathbb{Z}}$ denote a strictly stationary functional time series where the random variables X_n are elements in $L^2(T)$ (with expectation $\mu := \mathbb{E}[X_1] \in L^2(T)$, see Section 2.1 Bücher et al. (2018) for a detailed discussion of expected values in Hilbert spaces). For the sake of simplicity we will assume that $T = [0, 1]$, but all methods proposed in this paper can be generalized to other subsets of $\mathbb{R}^d$. To avoid confusion between the interval $[0, 1]$ corresponding to λ, which defines the sub-sample $X_1, \ldots, X_{\lfloor n\lambda \rfloor}$ and the interval $T = [0, 1]$ we write T for the interval $[0, 1]$ belonging to the argument t of X_n. Based on a sample $X_1, \ldots, X_n$ we are interested in relevant hypotheses regarding the parameter $d = \int_T \mu^2(t)dt$, that is

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

Define the partial sums

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

then, under suitable assumptions, the statistic $\int_T S_n^2(t, 1)dt$ is a consistent estimator of $\int_T \mu^2(t)dt$. Consequently a test for the hypotheses (2.1) is obtained by rejecting the null hypothesis for large values of

$$(2.2) \qquad \hat{\mathbb{T}}_n = \int_T S_n^2(t, 1)dt \,.$$

It will be shown in the proof of Theorem 2.1 that under some technical assumptions the asymptotic distribution of an appropriately standardized version of $\hat{\mathbb{T}}_n$ takes the form

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

with long-run variance

$$(2.3) \qquad \tau^2 = 4 \int_T \int_T \mu(s)\mu(t)C(s, t)ds\,dt \,,$$

where

$$(2.4) \qquad C(s, t) = \text{Var}(X_0(s)) + \sum_{\ell=1}^{\infty} \text{Cov}(X_0(t), X_\ell(s)) + \sum_{\ell=1}^{\infty} \text{Cov}(X_0(s), X_{-\ell}(t))$$

	99%	95%	90%
1)	18.257	10.998	7.855
2)	16.081	10.530	7.619
3)	16.282	10.583	7.662

Table 1: *Quantiles of the distribution of the statistic $\mathbb{W}$ defined by (2.8), where ν is the discrete uniform distribution supported on the points 1) $\lambda_i = i/5$ ($i = 1, \ldots, 4$), 2) $\lambda_i = i/20$ ($i = 1, \ldots, 19$) and 3) $\lambda_i = i/100$ ($i = 1, \ldots, 99$).*

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

$$(2.5) \qquad \hat{\mathbb{V}}_n := \left(\int_0^1 \left[\int_T S_n^2(t, \lambda) dt - \lambda^2 \int_T S_n^2(t, 1) dt \right]^2 \nu(d\lambda) \right)^{1/2} .$$

As we will show later we have

$$(2.6) \qquad \left(\sqrt{n}(\hat{\mathbb{T}}_n - d), \sqrt{n}\, \hat{\mathbb{V}}_n \right) \xrightarrow{\mathcal{D}} \left(\tau \mathbb{B}(1), \tau \left(\int_0^1 \lambda^2 (\mathbb{B}(\lambda) - \lambda \mathbb{B}(1))^2 \nu(d\lambda) \right)^{1/2} \right),$$

where $\mathbb{B}$ denotes a standard Brownian motion on the interval $[0, 1]$. In particular, this implies that, in the case $\tau \neq 0$, the ratio $(\hat{\mathbb{T}}_n - d)/\hat{\mathbb{V}}_n$ converges to a pivotal distribution. This suggests that a test for (2.1) can be constructed by rejecting the null hypothesis in (2.1), whenever

$$(2.7) \qquad \hat{\mathbb{T}}_n > \Delta + q_{1-\alpha}(\mathbb{W}) \hat{\mathbb{V}}_n ,$$

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

$$(2.8) \qquad \mathbb{W} := \frac{\mathbb{B}(1)}{\left(\int_0^1 \lambda^2 (\mathbb{B}(\lambda) - \lambda \mathbb{B}(1))^2 \nu(d\lambda) \right)^{1/2}} .$$

It is worthwhile to mention that the distribution of $\mathbb{W}$ is not the same as the one in previous work on self-normalization [see for example Shao (2010) or Shao (2015)] and quantiles of this distribution need to be simulated first. In Table 2.1 we display quantiles of this distribution, where ν is the discrete uniform distribution supported on the points $\lambda_i = i/5$ ($i = 1, \ldots, 4$), on the points $\lambda_i = i/20$ ($i = 1, \ldots, 19$) and on the points $\lambda_i = i/100$ ($i = 1, \ldots, 99$), respectively.

Next we prove that the decision rule in (2.7) indeed provides an asymptotic level α test. For this purpose we make the following assumptions [see also Berkes et al. (2013); Horváth et al. (2014)]:

(A1) For all $j \in \mathbb{Z}$ we have $X_j = \mu + \eta_j$, where $(\eta_j)_{j \in \mathbb{Z}}$ is a centered error process which satisfies (A2)–(A4).

(A2) $(\eta_j)_{j \in \mathbb{Z}}$ is a sequence of Bernoulli shifts, that is: there exists a measurable space, say $\mathcal{S}$ and a function $f : \mathcal{S}^\infty \longrightarrow L^2([0, 1])$ such that

$$\eta_j = f(\varepsilon_j, \varepsilon_{j-1}, \ldots) \qquad \text{for all } j \in \mathbb{Z},$$

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

(A3) $\mathbb{E}\|\eta_j\|^{2+\delta} < \infty$ for some $\delta \in (0, 1)$.

(A4) The sequence $(\eta_j)_{j \in \mathbb{Z}}$ can be approximated by ℓ-dependent sequences $(\eta_{j,\ell})_{j \in \mathbb{Z}}$ in the sense that for some $\kappa > 2 + \delta$

$$\sum_{\ell=1}^{\infty} \left(\mathbb{E}\|\eta_0 - \eta_{0,\ell}\|^{2+\delta} \right)^{1/\kappa} < \infty,$$

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

$$\eta_{j,\ell} = f(\varepsilon_j, \varepsilon_{j-1}, \ldots \varepsilon_{j-\ell+1}, \boldsymbol{\varepsilon}^*_{j,\ell})$$
$$\boldsymbol{\varepsilon}^*_{j,\ell} = (\varepsilon^*_{j,\ell,j-\ell}, \varepsilon^*_{j,\ell,j-\ell-1}, \ldots),$$

and the random variables $\varepsilon^*_{j,\ell,k}$ are i.i.d. copies of ε_0, and independent of the sequence $(\varepsilon_j)_{j \in \mathbb{Z}}$.

Theorem 2.1 *Assume that $\Delta > 0$. Under the assumptions (A1)-(A4) the test decision given in (2.7) is an asymptotic level α test for the null hypothesis in (2.1), i.e.*

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

A detailed proof of Theorem 2.1 is given in Section 6.1.1. What follows provides an informal overview of the main steps in the proof. If $\int \mu^2(t)dt \neq 0$ and assumptions (A1) - (A4) hold, it can be shown that

(2.9) $$\left\{ \sqrt{n}\left(\int_T S_n^2(t, \lambda)dt - \lambda^2 \int_T \mu^2(t)dt \right) \right\}_{\lambda \in [0,1]} \rightsquigarrow \left\{ \lambda \tau \mathbb{B}(\lambda) \right\}_{\lambda \in [0,1]},$$

where the symbol $\rightsquigarrow$ means weak convergence in $\ell^\infty([0, 1])$ and τ^2 is defined in (2.3). Now an application of the continuous mapping theorem directly yields the joint weak convergence (2.6). This implies the statement of Theorem 2.1 when $\int \mu^2(t)dt > 0$ after some simple computations.

If $\int \mu^2(t)dt = 0$ it is possible to prove that $\hat{\mathbb{T}}_n = o_{\mathbb{P}}(1), \hat{\mathbb{V}}_n = o_{\mathbb{P}}(1)$. This implies

$$\lim_{n\to\infty} \mathbb{P}\left(\hat{\mathbb{T}}_n > \Delta + q_{1-\alpha}(\mathbb{W})\hat{\mathbb{V}}_n\right) = \lim_{n\to\infty} \mathbb{P}\left(o_{\mathbb{P}}(1) > \Delta\right) = 0\,,$$

where we used that $\Delta > 0$ is fixed.

Remark 2.1 A test for the hypotheses of similarity

$$(2.10) \qquad H_0 : \int_T \mu^2(t)dt > \Delta \quad \text{versus} \quad H_1 : \int_T \mu^2(t)dt \leq \Delta$$

can be obtained by similar arguments. The null hypothesis in (2.10) is rejected, whenever

$$(2.11) \qquad \hat{\mathbb{T}}_n \leq \Delta + q_\alpha(\mathbb{W})\hat{\mathbb{V}}_n\,,$$

where $\hat{\mathbb{T}}_n$ and $\hat{\mathbb{V}}_n$ are defined in (2.2) and (2.5), respectively and $q_\alpha(\mathbb{W})$ is the α-quantile of the distribution of $\mathbb{W}$ defined in (2.8). Similar arguments as given in the proof of Theorem 2.1 show that this test is an asymptotic level α and consistent test for the hypotheses (2.10), that is

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

The details are omitted for the sake of brevity.

2.2 Two sample problems

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

$$D_{m,n}(t, \lambda) := \frac{1}{m}\sum_{j=1}^{\lfloor m\lambda \rfloor} X_j(t) - \frac{1}{n}\sum_{j=1}^{\lfloor n\lambda \rfloor} Y_j(t)\,.$$

From this definition we see that

$$(2.12) \qquad \mathbb{E}[D_{m.n}(t, \lambda)] = \lambda D(t) + O((m \vee n)^{-1})\,.$$

For the sake of brevity we restrict ourselves to the problem of testing the relevant hypotheses

$$(2.13) \qquad H_0 : \int_T D^2(t)dt \leq \Delta \quad \text{versus} \quad H_1 : \int_T D^2(t)dt > \Delta\,,$$

where Δ is a pre-specified threshold. A corresponding test for the hypotheses of similarity can be derived along the lines given in Remark 2.1. Following the discussion in Section 2.1 we propose to reject the null hypothesis in (2.13), whenever

$$(2.14) \qquad \hat{\mathbb{D}}_{m,n} > \Delta + q_{1-\alpha}(\mathbb{W})\hat{\mathbb{V}}_{m,n},$$

where $q_{1-\alpha}(\mathbb{W})$ is the $(1-\alpha)$-quantile of the distribution of the random variable $\mathbb{W}$ in (2.8). The statistics $\hat{\mathbb{D}}_{m,n}$ and $\hat{\mathbb{V}}_{m,n}$ are defined by

$$(2.15) \qquad \hat{\mathbb{D}}_{m,n} = \int_T D^2_{m,n}(t,1)dt,$$

$$(2.16) \qquad \hat{\mathbb{V}}_{m,n} = \Big(\int_0^1 \Big[\int_T D^2_{m,n}(t,\lambda)dt - \lambda^2 \int_T D^2_{m,n}(t,1)dt \Big]^2 \nu(d\lambda) \Big)^{1/2},$$

respectively, where ν is a probability measure on the interval $(0,1)$. The asymptotic properties of this test procedure will be established under the following assumptions.

(B1) The sample sizes satisfy: $m \to \infty$ and $n \to \infty$ and $m/(m+n) \to \rho \in (0,1)$.

(B2) The processes $\{X_n\}_{n\in\mathbb{Z}}$ and $\{Y_n\}_{n\in\mathbb{Z}}$ are independent and satisfy assumptions (A1) - (A4) stated in Section 2.1 with $\mathbb{E}[X_1] = \mu_1, \mathbb{E}[Y_1] = \mu_2$.

Theorem 2.2 *Assume that $\Delta > 0$. Under assumptions (B1)-(B2) the test decision given in (2.14) is an asymptotic level α test for the null hypothesis in (2.13), i.e.*

$$\lim_{n\to\infty} \mathbb{P}\Big(\hat{\mathbb{D}}_{m,n} > \Delta + q_{1-\alpha}(\mathbb{W})\mathbb{V}_{n,m} \Big) = \begin{cases} 0 & \text{if } \int_T D^2(t)dt < \Delta, \\ \alpha & \text{if } \int_T D^2(t)dt = \Delta, \\ 1 & \text{if } \int_T D^2(t)dt > \Delta. \end{cases}$$

Remark 2.2 The statement in Theorem 2.2 continues to hold if the observations X_i, Y_i are generated according to $X_i = \mu_1 + f_1(\varepsilon_i, \varepsilon_{i-1,\dots}), i = 1, \dots, n$ and $Y_i = \mu_2 + f_2(\varepsilon_{n+i}, \varepsilon_{n+i-1,\dots}), i = 1, \dots, m$ where $(\varepsilon_j)_{j\in\mathbb{Z}}$ denotes an i.i.d. sequence of $\mathcal{S}$-valued functions with the property that $\varepsilon_j(t,\omega)$ is jointly measurable as in (A2) and $f_1, f_2 : \mathcal{S}^\infty \to L^2([0,1])$ are functions such that the processes $(f_1(\varepsilon_i, \varepsilon_{i-1,\dots}))_{i\in\mathbb{Z}}$ and $(f_2(\varepsilon_i, \varepsilon_{i-1,\dots}))_{i\in\mathbb{Z}}$ satisfy conditions (A3) and (A4). This essentially corresponds to the setting discussed in Section 3 when the change point location is known.

3   Relevant change points in functional time series

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

$$(3.1) \qquad X_i = \begin{cases} \mu + f_1(\varepsilon_i, \varepsilon_{i-1,\dots}) & \text{if } i \le N\theta_0, \\ \mu + \delta + f_2(\varepsilon_i, \varepsilon_{i-1,\dots}) & \text{if } i > N\theta_0. \end{cases}$$

Here μ, δ denote deterministic but unknown elements in $L^2(T)$ and $\theta_0 \in (0,1)$ is fixed but unknown. Moreover, $(\varepsilon_j)_{j\in\mathbb{Z}}$ denotes an i.i.d. sequence of $\mathcal{S}$-valued functions with the property that $\varepsilon_j(t,\omega)$ is jointly measurable as in (A2) and $f_1, f_2 : \mathcal{S}^\infty \to L^2(T)$ are functions such that the processes $(f_1(\varepsilon_i, \varepsilon_{i-1,\dots}))_{i\in\mathbb{Z}}$ and $(f_2(\varepsilon_i, \varepsilon_{i-1,\dots}))_{i\in\mathbb{Z}}$ satisfy conditions (A3) and (A4). This setting is general enough to allow for the whole distribution of the observed functional data to change together with their mean.

We aim to construct a test for the relevant hypothesis

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

where Δ is a pre-specified threshold. Note that for known θ_0 a test for H_0 can be constructed in a similar fashion as in Section 2.2. In this section, we will prove that replacing the known change point by an estimator also leads to an asymptotic level α test. To this end we fix a trimming parameter $\varepsilon \in [0, 1/2)$ and define the estimator of the unknown change point θ_0 as

(3.3) $$\hat{\theta} := \frac{1}{N}\text{argmax}_{\lfloor N\varepsilon \rfloor \leq k \leq \lfloor N(1-\varepsilon) \rfloor} \hat{f}(k) \,,$$

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

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

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

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

(3.5) $$\hat{\theta} = \theta_0 + o_{\mathbb{P}}(N^{-1/2}) \,.$$

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

$$D_N^{cp}(t, \lambda, \theta) := \frac{1}{\lfloor N\theta \rfloor}\sum_{j=1}^{\lfloor \lambda\lfloor\theta N\rfloor\rfloor} X_j(t) - \frac{1}{N-\lfloor N\theta \rfloor}\sum_{j=\lfloor\theta N\rfloor+1}^{\lfloor\theta N\rfloor+\lfloor\lambda(N-\lfloor\theta N\rfloor)\rfloor} X_j(t) \,.$$

Following the developments in Section 2.2 let

(3.6) $$\hat{\mathbb{D}}_N^{cp} = \int_T D_N^2(t, 1, \hat{\theta})dt \,,$$

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

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

(3.8) $$\hat{\mathbb{D}}_N^{cp} > \Delta + q_{1-\alpha}(\mathbb{W})\hat{\mathbb{V}}_N^{cp} \,,$$

where $q_{1-\alpha}(\mathbb{W})$ is the $(1-\alpha)$-quantile of the distribution of the random variable $\mathbb{W}$ in (2.8). This test decision is justified in the following theorem.

9

Theorem 3.1 *Assume $\Delta > 0$. If the data is generated according to model (3.1), $\theta_0 \in (\varepsilon, 1 - \varepsilon)$, and the assumptions described right below (3.1) are satisfied, then the test decision in (3.8) leads to a consistent and asymptotic level α test for the relevant hypotheses (3.2), that is*

$$\lim_{n \to \infty} \mathbb{P}\Big(\hat{\mathbb{D}}_N^{cp} > \Delta + q_{1-\alpha}(\mathbb{W})\hat{\mathbb{V}}_N^{cp}\Big) = \begin{cases} 0 & \text{if } \int_T \delta^2(t)dt < \Delta\,, \\ \alpha & \text{if } \int_T \delta^2(t)dt = \Delta\,, \\ 1 & \text{if } \int_T \delta^2(t)dt > \Delta\,. \end{cases}$$

The proof of Proposition 3.1 and Theorem 3.1 is technically difficult and deferred to Section 6.2, but the main idea is as follows. A straightforward calculation shows that the processes $\hat{\mathbb{D}}_N^{cp}$ and $\hat{\mathbb{V}}_N^{cp}$ in (3.6) and (3.7) are continuous functionals of the process

$$\mathbb{Z}_N(\lambda, \hat{\theta}) = \sqrt{N} \int_T \big(D_N^{cp}(t, \lambda, \hat{\theta})^2 - \lambda^2 \delta(t)^2\big)dt\,.$$

Using Proposition 3.1 it can be shown that

$$\sup_{\lambda \in [0,1]} |\mathbb{Z}_N(\lambda, \theta_0) - \mathbb{Z}_N(\lambda, \hat{\theta})| = o_{\mathbb{P}}(1)\,,$$

where θ_0 is the true change point. We can then establish the weak convergence

$$\big\{\mathbb{Z}_N(\lambda, \theta_0)\big\}_{\lambda \in [0,1]} \rightsquigarrow \big\{\lambda \tau_\delta^2 \mathbb{B}(\lambda)\big\}_{\lambda \in [0,1]}\,,$$

where τ_δ^2 is a positive constant depending on the dependence structure of the process. Using the continuous mapping theorem we then find

$$\frac{\hat{\mathbb{D}}_N^{cp}}{\hat{\mathbb{V}}_N^{cp}} \xrightarrow{\mathcal{D}} \mathbb{W}\,,$$

where the random variable $\mathbb{W}$ is defined in (2.8). When $\int_T \delta^2(t)dt > 0$ the assertion of Theorem 3.1 now follows directly. In the remaining case $\int_T \delta^2(t)dt = 0$ one can show $\hat{\mathbb{D}}_N^{cp} = o_{\mathbb{P}}(1), \hat{\mathbb{V}}_N^{cp} = o_{\mathbb{P}}(1)$ and the assertion follows with the same arguments as given in Section 2.1.

4 Extensions beyond functional time series

In this section we briefly discuss how the ideas presented in this paper can be extended beyond the context of functional time series. We begin by introducing a general setup which will be used throughout this section. Let $X_1, ..., X_n$ denote a sample of (potentially dependent) random elements in some measurable space $\mathcal{S}$. Assume that we are interested in inference on a parameter $\mu = \mu_{\mathbb{P}}$ that can be assigned to distributions $\mathbb{P}$ on $\mathcal{S}$. We will assume that μ takes values in $\mathcal{M}$, a subset of a real Hilbert space $\mathcal{H}$ equipped with an inner product $\langle \cdot, \cdot \rangle$ and induced norm by $\|\cdot\|_{\mathcal{H}}$. Further, assume that for each $m \in \mathbb{N}$ there exists a mapping $f_m : \mathcal{S}^m \to \mathcal{M}$, where $f_m(X_1, ..., X_m)$

is interpreted as estimator for μ based on the observations $X_1, ..., X_m$. The situation considered in Sections 1 - 3 corresponds to the choice $\mathcal{S} = L^2([0,1])$, $\mathcal{H} = L^2([0,1])$ and $\mu_{\mathbb{P}}$ is the mean (function) of X_i (see the discussion in Example 4.1 below for more details). Finally, define for $\Lambda \subset [0,1]$ the space of functions

$$\mathcal{B}(\Lambda, \mathcal{H}) := \{f : \Lambda \to \mathcal{H} : \sup_{\lambda \in \Lambda} \|f(\lambda)\|_{\mathcal{H}} < \infty\}$$

equipped with the norm

$$\|g\|_{\mathcal{B}} := \sup_{\lambda \in \Lambda} \|g(\lambda)\|_{\mathcal{H}} \,.$$

This space will be used to characterize the joint behaviour of estimators of μ computed from several sub-samples (with sub-sample proportion corresponding to the index λ). Note that if the set Λ contains only finitely many elements, say $|\Lambda|$, the normed space $(\mathcal{B}(\Lambda, \mathcal{H}), \|\cdot\|_{\mathcal{B}})$ can be identified with the $|\Lambda|$-fold Cartesian product of $\mathcal{H}$ (viewed as a normed space).

4.1 The one sample and two sample case

We begin by considering the one sample case since in this setting the relevant conditions are particularly simple and transparent. The self-normalized statistic is based on a probability measure ν with support $\Lambda_\nu \subset (0,1)$. Using the notation $\hat{\mu}_{1:k} := f_k(X_1, ..., X_k)$ we define random elements in $\mathcal{B}(\Lambda, \mathcal{H})$ through $g_n(\lambda) := \lambda \sqrt{n}(\hat{\mu}_{1:\lfloor \lambda n \rfloor} - \mu)$. Assume that

$$(4.1) \qquad\qquad g_n \rightsquigarrow \mathbb{H} \quad \text{in } \mathcal{B}(\Lambda_\nu \cup 1, \mathcal{H}) \,.$$

Note that we do not require measurability of g_n and weak convergence is defined in the sense of Hoffman-Jorgensen, see Section 1.3 in Van der Vaart and Wellner (1996). Further, assume that the limit $\mathbb{H}$ has the following additional properties:

$$(4.2) \qquad\qquad \left\{ \langle \mathbb{H}(\lambda), \mu \rangle \right\}_{\lambda \in \Lambda_\nu \cup 1} \overset{\mathcal{D}}{=} \sigma^2 \left\{ \lambda \mathbb{B}(\lambda) \right\}_{\lambda \in \Lambda_\nu \cup 1} ,$$

$$(4.3) \qquad\qquad \left(\int_{\Lambda_\nu} \left[\|\mathbb{H}(\lambda)\|_{\mathcal{H}}^2 - \lambda^2 \|\mathbb{H}(1)\|_{\mathcal{H}}^2 \right]^2 \nu(d\lambda) \right)^{-1} = O_P(1) \,,$$

where σ^2 is a (nonnegative) real-valued parameter that can depend on the distribution $\mathbb{P}$ of X_i and $\mathbb{B}$ is a standard Brownian motion on the interval $[0,1]$. Moreover, we assume that $\sigma^2 \neq 0$ if $\mu \neq 0$.

If (4.1) - (4.3) are satisfied, an asymptotic level α and consistent test for the (relevant) hypotheses

$$H_0 : \|\mu\|_{\mathcal{H}}^2 \leq \Delta \quad \text{versus} \quad H_1 : \|\mu\|_{\mathcal{H}}^2 > \Delta$$

is given by rejecting H_0 whenever the inequality (2.7) holds, where the statistic $\hat{\mathbb{T}}_n$ and $\hat{\mathbb{V}}_n$ are now defined by $\hat{\mathbb{T}}_n := \|\hat{\mu}_{1:n}\|_{\mathcal{H}}^2$ and

$$\hat{\mathbb{V}}_n := \left(\int_{\Lambda_\nu} \left[\lambda^2 \|\hat{\mu}_{1:\lfloor n\lambda \rfloor}\|_{\mathcal{H}}^2 - \lambda^4 \|\hat{\mu}_{1:n}\|_{\mathcal{H}}^2 \right]^2 \nu(d\lambda) \right)^{1/2} ,$$

respectively, and $q_{1-\alpha}(\mathbb{W})$ is the $(1-\alpha)$-quantile of the distribution of the random variable $\mathbb{W}$ defined in (2.8).

Example 4.1

(a) In the setting of Section 2.1 the random variables $X_1, ..., X_n$ with distribution $\mathbb{P}$ take values in $\mathcal{S} = L^2([0,1])$, $\mu_\mathbb{P}$ is the mean (function) of X_i which is an element of $\mathcal{H} = L^2([0,1])$, and $f_m(X_1, ..., X_m) = m^{-1}(X_1 + ... + X_m)$. The estimators $\hat{\mu}_{1:\lfloor n\lambda \rfloor}$ then take the form $\hat{\mu}_{1:\lfloor n\lambda \rfloor}(t) = \frac{1}{\lfloor n\lambda \rfloor} \sum_{i=1}^{\lfloor n\lambda \rfloor} X_i(t)$. The limit $\mathbb{H}$ in (4.1) is given by $\mathbb{H}(\lambda) = \Gamma(\cdot, \lambda)$ where the process Γ is defined in equation (6.4).

(b) When $\mathcal{H}$ is $\mathbb{R}$ with $\langle x, y \rangle = xy$, the space $\mathcal{B}(\Lambda, \mathcal{H})$ can be identified with $\ell^\infty(\Lambda)$. Conditions (4.1) and (4.2) follow from the functional CLT

(4.4)
$$\left\{ \lambda \sqrt{n}(\hat{\mu}_{1:\lfloor \lambda n \rfloor} - \mu) \right\}_{\lambda \in \Lambda_\nu \cup 1} \rightsquigarrow \tau^2 \left\{ \mathbb{B}(\lambda) \right\}_{\lambda \in \Lambda_\nu \cup 1}.$$

In this case $\mathbb{H} = \tau^2 \mathbb{B}$ and $\sigma^2 = \mu^2 \tau^2$. Moreover, condition (4.3) follows from elementary properties of the multivariate normal distribution and the assumption that ν is a probability measure on the interval $(0, 1)$.

When μ is the mean of $\mathbb{P}$ and $\hat{\mu}_{1:\lfloor \lambda n \rfloor}$ is the sample mean of the first $\lfloor \lambda n \rfloor$ observations, (4.4) holds under a wide variety of assumptions on the serial dependence of time series. When $\mathcal{S} = \mathbb{R}^d$ and μ is a smooth function of the (multivariate) cdf of X, this condition can be verified using the general framework developed in Volgushev and Shao (2014). This includes quantities such as quantiles or Kendall's τ and other dependence measures.

(c) If $\mathcal{H} = \mathbb{R}^d$ (with the usual inner product), the space $\mathcal{B}(\Lambda, \mathcal{H})$ can be identified with $[\ell^\infty(\Lambda)]^d$. The norm $\|f\|_\mathcal{B} = \sup_{\lambda \in \Lambda} \|f(\lambda)\|_2$ ($f \in \mathcal{B}(\Lambda, \mathbb{R}^d)$) is equivalent to the usual norm $\|(g_1, ..., g_d)\| := \max_{j=1,...,d} \sup_{\lambda \in \Lambda} |g_j(\lambda)|$ where $(g_1, ..., g_d) \in [\ell^\infty(\Lambda)]^d$ and the weak convergence in (4.1) is interpreted as weak convergence in the product space $[\ell^\infty(\Lambda)]^d$ as discussed in Section 1.4 of Van der Vaart and Wellner (1996). Let, $\hat{\mu}_{1:\lfloor \lambda n \rfloor, j}$ and μ_j denote the j'th components of the vectors $\hat{\mu}_{1:\lfloor \lambda n \rfloor}$ and μ, respectively, and let $g_{n,j}(\lambda) := \lambda \sqrt{n}(\hat{\mu}_{1:\lfloor \lambda n \rfloor, j} - \mu_j)$. Conditions (4.1) and (4.2) now follow from a multivariate version of the functional CLT, i.e.

(4.5)
$$(g_{n,1}, ..., g_{n,d})^\top \rightsquigarrow \Sigma(\mathbb{B}_1, ..., \mathbb{B}_d)^\top$$

in $[\ell^\infty(\Lambda)]^d$, where $\mathbb{B}_1, ..., \mathbb{B}_d$ are independent Brownian motions on the interval $[0, 1]$ and Σ denotes a matrix which can depend on $\mathbb{P}$ and is non-singular when $\mu \neq 0$. The functional weak convergence in (4.5) can be verified in a similar fashion as discussed in (b) and details are omitted for the sake of brevity.

Next we briefly discuss the two sample case, where $X_1, ..., X_m$ and $Y_1, ..., Y_n$ are random variables with $X_i \sim \mathbb{P}, Y_i \sim \mathbb{Q}$. For simplicity we shall further assume that the samples $X_1, ..., X_m$ and $Y_1, ..., Y_n$ are independent, while dependence within the samples is explicitly allowed. Let $\mu_\mathbb{P}, \mu_\mathbb{Q}$ denote the parameters of interest corresponding to $\mathbb{P}, \mathbb{Q}$, respectively. Introduce the notation $\hat{\mu}_{1:k}^X := f_k(X_1, ..., X_k), \hat{\mu}_{1:k}^Y := f_k(Y_1, ..., Y_k)$ and let $g_m^X(\lambda) := \lambda \sqrt{m}(\hat{\mu}_{1:\lfloor \lambda m \rfloor}^X - \mu_\mathbb{P})$, $g_n^Y(\lambda) :=$

$\lambda\sqrt{n}(\hat{\mu}_{1:\lfloor\lambda n\rfloor}^{Y} - \mu_{\mathbb{Q}})$. Provided that $m/(m+n) \to \rho \in (0,1)$ and $g_m^X \rightsquigarrow \mathbb{H}^X$ and $g_n^Y \rightsquigarrow \mathbb{H}^Y$ in $\mathcal{B}(\Lambda_\nu \cup 1, \mathcal{H})$ with limiting processes $\mathbb{H}^X, \mathbb{H}^Y$ satisfying (4.2) and (4.3), an asymptotic level α and consistent test for the (relevant) hypotheses

$$H_0 : \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}}^2 \leq \Delta \quad \text{versus} \quad H_1 : \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}}^2 > \Delta$$

is given by rejecting H_0 whenever $\hat{\mathbb{T}}_{m,n} > \Delta + q_{1-\alpha}(\mathbb{W})\hat{\mathbb{V}}_{m,n}$. Here the statistics $\hat{\mathbb{T}}_{m,n}$ and $\hat{\mathbb{V}}_{m,n}$ are defined by $\hat{\mathbb{T}}_{m,n} := \|\hat{\mu}_{1:m}^X - \hat{\mu}_{1:n}^Y\|_{\mathcal{H}}^2$ and

$$\hat{\mathbb{V}}_{m,n} = \left(\int_{\Lambda_\nu} \left\{ \lambda^2 \left\|\hat{\mu}_{1:\lfloor m\lambda\rfloor}^X - \hat{\mu}_{1:\lfloor n\lambda\rfloor}^Y\right\|_{\mathcal{H}}^2 - \lambda^4 \left\|\hat{\mu}_{1:m}^X - \hat{\mu}_{1:n}^Y\right\|_{\mathcal{H}}^2 \right\}^2 \nu(d\lambda) \right)^{1/2},$$

respectively, and $q_{1-\alpha}(\mathbb{W})$ is the $(1-\alpha)$-quantile of the distribution of the random variable $\mathbb{W}$ in (2.8). The discussion in Example 4.1 also applies to the two sample case with obvious modifications.

4.2 Testing for relevant change points in a general setting

Next we discuss the problem of testing for relevant change points in the general context introduced in the beginning of Section 4. Assume that $X_i \sim \mathbb{P}$ for $1 \leq i \leq n\theta_0$ and $X_i \sim \mathbb{Q}$ for $n\theta_0 < i \leq n$ where $\theta_0 \in (0,1)$. Technically this is a triangular array model, but we will not stress this in the notation. We explicitly allow $\mathbb{P} = \mathbb{Q}$ which corresponds to the case of no change point in the sequence $X_1, \ldots, X_n$. Let $\mu_{\mathbb{P}}, \mu_{\mathbb{Q}}$ denote the parameters of interest corresponding to $\mathbb{P}, \mathbb{Q}$, respectively. Our aim is to test the hypotheses of a relevant change in the parameter μ of the sequence $X_1, \ldots, X_n$, that is

(4.6) $$H_0 : \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}}^2 \leq \Delta \quad \text{versus} \quad H_1 : \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}}^2 > \Delta .$$

Following the developments in Section 3, we assume that $\hat{\theta}$ is a consistent estimator for θ_0 and introduce the notation $\hat{\mu}_{j:k} := f_{k-j+1}(X_j, \ldots, X_k)$. The null hypothesis in (4.6) is rejected if $\hat{\mathbb{T}}^{cp} > \Delta + q_{1-\alpha}(\mathbb{W})\hat{\mathbb{V}}^{cp}$, where the statstics $\hat{\mathbb{T}}^{cp}$ and $\hat{\mathbb{V}}^{cp}$ are defined by

$$\hat{\mathbb{T}}^{cp} = \|\hat{\mu}_{1:\lfloor n\hat{\theta}\rfloor} - \hat{\mu}_{\lfloor n\hat{\theta}\rfloor+1:n}\|_{\mathcal{H}}^2 ,$$

$$\hat{\mathbb{V}}^{cp} = \left(\int_{\Lambda_\nu} \left\{ \lambda^2 \left\|\hat{\mu}_{1:\lfloor n\lambda\hat{\theta}\rfloor} - \hat{\mu}_{\lfloor n\hat{\theta}\rfloor+1:\lfloor n\hat{\theta}\rfloor+\lfloor n\lambda(1-\hat{\theta})\rfloor}\right\|_{\mathcal{H}}^2 - \lambda^4 \left\|\hat{\mu}_{1:\lfloor n\hat{\theta}\rfloor} - \hat{\mu}_{\lfloor n\hat{\theta}\rfloor+1:n}\right\|_{\mathcal{H}}^2 \right\}^2 \nu(d\lambda) \right)^{1/2},$$

respectively, and $q_{1-\alpha}(\mathbb{W})$ is the $(1-\alpha)$-quantile of the distribution of the random variable $\mathbb{W}$ in (2.8). The asymptotic validity of this test can be verified provided that

(4.7) $$\left\{ \lambda\sqrt{n}(\hat{\mu}_{1:\lfloor \lambda\hat{\theta}n\rfloor} - \hat{\mu}_{1+\lfloor \hat{\theta}n\rfloor+\lfloor \lambda(1-\hat{\theta})n\rfloor:n} + \mu_{\mathbb{Q}} - \mu_{\mathbb{P}}) \right\}_{\lambda\in\Lambda_\nu\cup 1} \rightsquigarrow \mathbb{H} \quad \text{in } \mathcal{B}(\Lambda_\nu \cup 1, \mathcal{H}) ,$$

where the limiting process $\mathbb{H}$ satisfies (4.2) and (4.3). This is a non-trivial high-level assumption since the estimator $\hat{\theta}$ appears on the left-hand side.

In concrete situations, for example in the situation considered in Section 3, it is possible to replace (4.7) by more tractable and elementary assumptions. As a further illustration we will discuss the setting of a real-valued parameter μ that can be represented as smooth functional of the distribution function F of the random variable X with values in $\mathcal{S} = \mathbb{R}^d$. First, we note that there exists a general and well-developed machinery for establishing convergence rates of change point estimators [see, for example Carlstein (1988), Dümbgen (1991) or Hariz et al. (2007), among others], and we assume

$$(4.8) \qquad \hat{\theta} = \theta_0 + o_{\mathbb{P}}(n^{-1/2}),$$

throughout this section, which is satisfied in many cases of practical interest. The following discussion is a little informal and we refer the interested reader to the work of Volgushev and Shao (2014) for some of the technical details that are omitted here for the sake of brevity. Assume that $\mu_{\mathbb{P}} = \Phi(F_{\mathbb{P}}), \mu_{\mathbb{Q}} = \Phi(F_{\mathbb{Q}})$ for some smooth map Φ, where $F_{\mathbb{P}}, F_{\mathbb{Q}}$ denote the cdf of $\mathbb{P}, \mathbb{Q}$, respectively. Let $\hat{\mu}_{i:j} = \Phi(\hat{F}_{i:j})$ where $\hat{F}_{i:j}$ is the empirical CDF of $X_i, ..., X_j$. Under suitable assumptions on the temporal dependence structure of $X_1, ..., X_n$ it is possible to prove the weak convergence

$$\left(\left\{ \lambda\sqrt{n}(\hat{F}_{1:\lfloor n\theta_0\lambda \rfloor}(u) - F_{\mathbb{P}}(u)) \right\}_{\lambda\in[0,1], u\in\mathbb{R}^d}, \left\{ \lambda\sqrt{n}(\hat{F}_{\lfloor n\theta_0 \rfloor+1:\lfloor n\theta_0 \rfloor+\lfloor n\lambda(1-\theta_0) \rfloor}(u) - F_{\mathbb{Q}}(u)) \right\}_{\lambda\in[0,1], u\in\mathbb{R}^d} \right)$$
$$\rightsquigarrow \left(\left\{ \mathbb{G}_1(\lambda, u) \right\}_{\lambda\in[0,1], u\in\mathbb{R}^d}, \left\{ \mathbb{G}_2(\lambda, u) \right\}_{\lambda\in[0,1], u\in\mathbb{R}^d} \right)$$

in $\ell^\infty([0,1] \times \mathbb{R}^d)^2$, where $\mathbb{G}_1, \mathbb{G}_2$ denote two independent centered Gaussian processes with covariance structure of the form $\mathbb{E}[\mathbb{G}_j(s,u)\mathbb{G}_j(t,v)] = (s \wedge t)K_j(u,v)$ $(j = 1,2)$. By elementary calculations we have

$$\sup_{\lambda\geq\varepsilon} \sup_u |\hat{F}_{1:\lfloor n\theta_0\lambda \rfloor}(u) - \hat{F}_{1:\lfloor n\hat{\theta}\lambda \rfloor}(u)| = O_{\mathbb{P}}(|\theta_0 - \hat{\theta}|) = o_{\mathbb{P}}(n^{-1/2}),$$

(note that two empirical cdf which are based on k and $k+l$ observations with an overlap of k observations differ by at most l/k). Similarly, it can be shown that for any $\varepsilon > 0$

$$\sup_{\lambda\geq\varepsilon} \sup_u |\hat{F}_{\lfloor n\theta_0 \rfloor+1:\lfloor n\theta_0 \rfloor+\lfloor n\lambda(1-\theta_0) \rfloor}(u) - \hat{F}_{\lfloor n\hat{\theta} \rfloor+1:\lfloor n\hat{\theta} \rfloor+\lfloor n\lambda(1-\hat{\theta}) \rfloor}(u)| = O_{\mathbb{P}}(|\theta_0 - \hat{\theta}|) = o_{\mathbb{P}}(n^{-1/2}),$$

and the two displays above imply the weak convergence

$$\left(\left\{ \lambda\sqrt{n}(\hat{F}_{1:\lfloor n\hat{\theta}\lambda \rfloor}(u) - F_{\mathbb{P}}(u)) \right\}_{\lambda\in[\varepsilon,1], u\in\mathbb{R}^d}, \left\{ \lambda\sqrt{n}(\hat{F}_{\lfloor n\hat{\theta} \rfloor+1:\lfloor n\hat{\theta} \rfloor+\lfloor n\lambda(1-\hat{\theta}) \rfloor}(u) - F_{\mathbb{Q}}(u)) \right\}_{\lambda\in[\varepsilon,1], u\in\mathbb{R}^d} \right)$$
$$\rightsquigarrow \left(\left\{ \mathbb{G}_1(\lambda, u) \right\}_{\lambda\in[\varepsilon,1], u\in\mathbb{R}^d}, \left\{ \mathbb{G}_2(\lambda, u) \right\}_{\lambda\in[\varepsilon,1], u\in\mathbb{R}^d} \right)$$

in $\ell^\infty([\varepsilon,1] \times \mathbb{R}^d)^2$. Assume that the mapping Φ is Hadamard differentiable at the points $F_{\mathbb{P}}, F_{\mathbb{Q}}$, tangentially to a vector space $V \subset \ell^\infty(\mathbb{R}^d)$, with linear derivatives $d\Phi_{F_{\mathbb{P}}}, d\Phi_{G_{\mathbb{P}}}$ and that

$\mathbb{G}_1(\lambda, \cdot), \mathbb{G}_2(\lambda, \cdot) \in V$ a.s. Then the results in Section 4 of Volgushev and Shao (2014) combined with linearity of $d\Phi_{F_{\mathbb{P}}}, d\Phi_{F_{\mathbb{Q}}}$ and some calculations show that the weak convergence in (4.7) holds for the functional $\mu = \Phi(F)$ with $\hat{\mu}_{i:j} = \Phi(\hat{F}_{i:j})$ provided that $\Lambda_\nu \subset [\varepsilon, 1]$. In particular we have have $\mathbb{H} \stackrel{\mathcal{D}}{=} \sigma^2 \mathbb{B}$ for some $\sigma^2 > 0$.

5 Finite sample properties

In this section we illustrate the finite sample properties of the new procedures by means of a simulation study. Note that one has to specify the measure ν used in the definition of the normalizer (2.5), (2.16) and (3.7) and we use

(5.1)
$$\nu = \frac{1}{19} \sum_{i=1}^{19} \delta_{i/20}$$

throughout this section if not mentioned otherwise; here δ_λ denotes the Dirac measure at the point $\lambda \in [0, 1]$. For example, for this choice the quantity $\hat{\mathbb{V}}_n$ defined in (2.5) is given by

$$\hat{\mathbb{V}}_n = \left[\frac{1}{19} \sum_{i=1}^{19} \left(\int_T S_n^2\big(t, \tfrac{i}{20}\big) dt - \big(\tfrac{i}{20}\big)^2 \int_T S_n^2(t, 1) dt \right)^2 \right]^{1/2}$$

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

5.1 One sample problems

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

(5.2)
$$\mu(t) = \sqrt{2\delta} * \sin(2\pi t)$$

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

$$\eta_j = \sum_{i=1}^{D} N_{i,j} b_i \quad (j = 1, \ldots, n),$$

where $N_{1,1}, N_{2,1}, \ldots, N_{D,n}$ are independent normal distributed random variables with expectation zero and variance $\sigma_i^2 = \text{Var}(N_{i,j}) = 1/i^2$ $(i = 1, \ldots, D; j = 1, \ldots, n)$. Our first example considers independent error processes of the form

(5.3)
$$\varepsilon_j = \eta_j \quad (j \in \mathbb{Z}),$$

15

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

$$(5.4) \qquad\qquad \varepsilon_j = \eta_j + \Theta\eta_{j-1} \quad (j \in \mathbb{Z}).$$

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

In Figure 1 we display the simulated rejection probabilities of the test (2.7) for the hypotheses (2.1), where $\Delta = 0.02$, which corresponds to the value $\delta = 0.02$ in model (5.2). These results show a pattern which is in line with the theoretical findings in Theorem 2.1. For example at the boundary of the null hypotheses, i.e. for $\delta = \Delta = 0.02$, the simulated level is close to the nominal level. In the interior of the null hypothesis ($\delta < \Delta$) the simulated rejection probabilities are strictly smaller than $\alpha = 0.05$, while they are strictly larger than 0.05 in the interior of the alternative, i.e. $\delta > \Delta$.
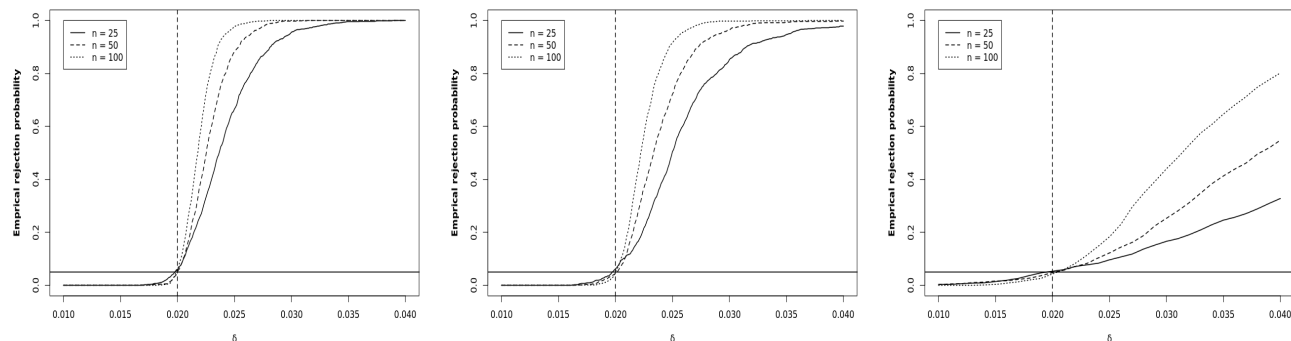

Figure 1: *Simulated rejection probabilities of the test (2.7) for the relevant hypotheses (2.1) with $\Delta = 0.02$. The mean function is given by (5.2) and different error processes are considered. First panel: independent error processes defined by (5.3). Second panel: $fMA(1)$ processes defined by (5.4). Third panel: independent Brownian Bridges*
.

5.2 Two sample problem

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

$$(5.5) \qquad\qquad \mu_1 \equiv 0 , \quad \mu_2(t) = at(1-t)$$

[see Section 4 in Horváth et al. (2013)], such that $\int_0^1 \mu_2^2(t)dt = a^2/30$. We are interested in testing the hypotheses (2.13), that is

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

where $D = \mu_1 - \mu_2$ is the (unknown) difference of the two mean functions and the threshold is given by $\Delta = 0.2^2/30$ (note that this corresponds to the choice $a = 0.2$). In Figure 2 we display the rejection probabilities of the test (2.14) as a function of parameter a for different sample sizes m and n. We consider independent samples, $fMA(1)$ processes (generated as described in Section 5.1) and independent Brownian Bridges as error processes.

We also investigate the important case of dependent samples. In this case we investigate two scenarios. We generate a $fMA(1)$ process $\{\eta_i\}_{i \in \mathbb{Z}}$ as described in Section 5.1 and define

(5.6) $$X_i = \mu_1 + \eta_i , \quad i = 1, \ldots, m ; \quad Y_i = \mu_2 + \eta_{m+i} , \quad i = 1, \ldots, n$$

as the first and

(5.7) $$X_i = \mu_1 + \eta_i , \quad i = 1, \ldots, m ; \quad Y_i = \mu_2 + \sqrt{3}\, \eta_{m+i} , \quad i = 1, \ldots, n .$$

as the second scenario; in both cases $\Delta = 0.3^2/30$. The corresponding rejection probabilities of the test (2.14) are depicted in Figure 3. Overall the test performs well in all settings considered.
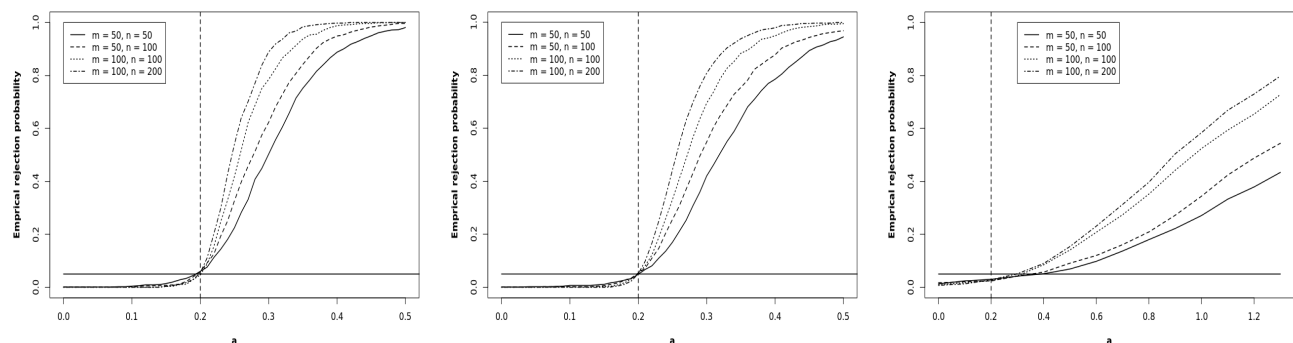

Figure 2: *Simulated rejection probabilities of the test* (2.14) *for the relevant hypotheses* (2.13) *with* $\Delta = 0.2^2/30$. *The mean functions are given by* (5.5) *and different independent error processes are considered. First panel: independent error processes defined by* (5.3). *Second panel: $fMA(1)$ processes defined by* (5.4). *Third panel: Brownian Bridges.*
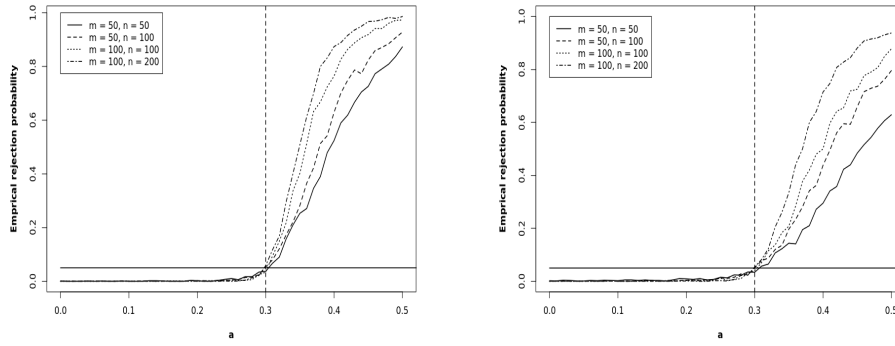
Figure 3: *Simulated rejection probabilities of the test* (2.14) *for the relevant hypotheses* (2.13) *with* $\Delta = 0.3^2/30$ *in the case of dependent samples. The mean functions are given by* (5.5). *Left panel: error processes defined by* (5.6). *Right panel: error processes defined by* (5.7).

5.3 Change point problem

We begin by considering the model (3.1) with $\theta_0 = 0.5, \mu = 0, \delta(t) = at(1-t)$, the errors are i.i.d. from (5.3). The trimming parameter ε for estimating the change point location is set to 0.05. Data are generated with $a = 0, 0.02, \ldots, 0.5$ and then empirical rejection probabilities are calculated using $\Delta = 0.1^2/30, 0.2^2/30, 0.3^2/30, 0.4^2/30$, respectively. These probabilities are shown in Figure 4. From Theorem 3.1, we expect that the probability of rejection should be close to α at the boundary of the hypotheses $(\int_0^1 D^2(t)dt = \Delta)$, strictly smaller than α in the interior of the null hypothesis $(\int_0^1 D^2(t)dt < \Delta)$ and larger than α in the interior of the alternative $(\int_0^1 D^2(t)dt > \Delta)$. This pattern is clearly observed for relevant hypotheses with theshold $\Delta \geq 0.2^2/30$. On the other hand the proposed test is oversized if relevant hypotheses with $\Delta = 0.1^2/30$ are tested (see the left upper panel in Figure 4). The reason for this behaviour consists in the fact that change point tests for relevant hypotheses require a precise estimate of the change point (see the definition of the statistics $\hat{\mathbb{D}}_N^{cp}$ and $\hat{\mathbb{V}}_N^{cp}$ in (3.6) and (3.7)). For small values of a it is extremely difficult to estimate the true change point location, and an imprecise estimation of the change point results in a less accurate approximation of the nominal level. The difficulty of estimating the true change point location for small values of a is further illustrated in Figure 5 where we show the histogram of the corresponding estimator of the change point for $a = 0.1, 0.2, 0.3, 0.4$ with sample size is $N = 200$.
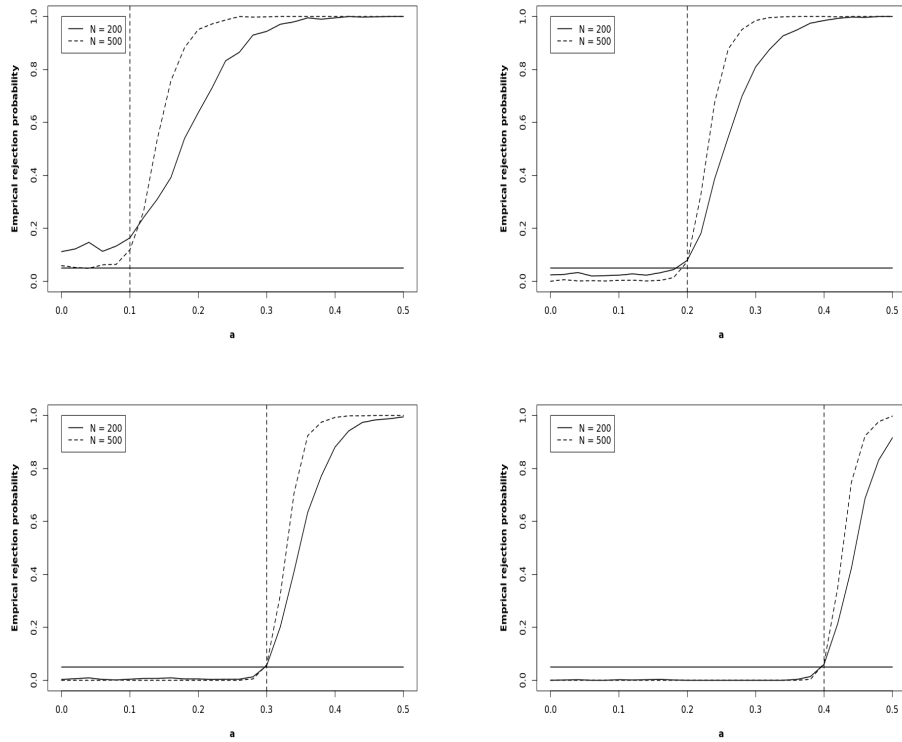
Figure 4: *Simulated rejection probabilities of the test* (3.8) *for the relevant hypotheses* (3.2) *with* $\Delta = 0.1^2/30, 0.2^2/30, 0.3^2/30, 0.4^2/30$. *Data is generated according to model* (3.1) *with* $\theta_0 = 0.5, \mu = 0, \delta(t) = at(1-t)$, *for* $a = 0, 0.02, \ldots, 0.5$, *and the errors are i.i.d. defined by* (5.3). *The tuning parameter is set to* $\varepsilon = 0.05$.
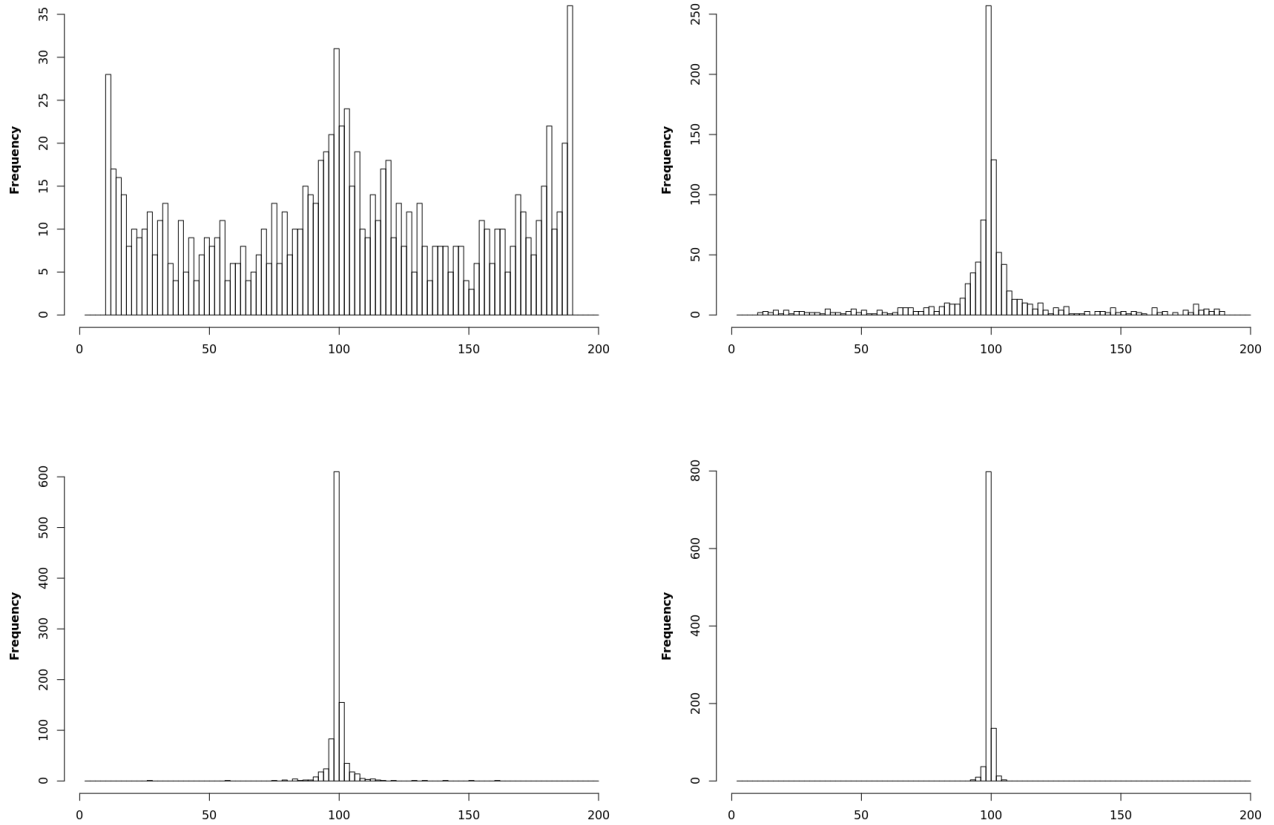
Figure 5: *Histogram of the change point estimator $\hat{\theta}$ defined in (3.3). The sample size is $N = 200$ and data is generated according to model (3.1) with $\theta_0 = 0.5, \mu = 0, \delta(t) = at(1-t)$, for $a = 0, 0.02, \ldots, 0.5$, and the errors are i.i.d. defined by (5.3). The tuning parameter is set to $\varepsilon = 0.05$.*

Next, we investigate two scenarios similar to those described at the end of Section 5.2, i.e we generate a $fMA(1)$ process $\{\eta_i\}_{i \in \mathbb{Z}}$ as described in Section 5.1 and define

$$(5.8) \qquad X_i = \mu + \eta_i , \quad i = 1, \ldots, \lfloor \theta_0 N \rfloor ; \quad X_i = \mu + \delta + \eta_i , \quad i = \lfloor \theta_0 N \rfloor + 1, \ldots, N .$$

as the first and

$$(5.9) \quad X_i = \mu + \eta_i , \quad i = 1, \ldots, \lfloor \theta_0 N \rfloor ; \quad X_i = \mu + \delta + \sqrt{3}\, \eta_i , \quad i = \lfloor \theta_0 N \rfloor + 1, \ldots, N .$$

as the second scenario. The functions μ, δ are as described in the beginning of this section. The corresponding rejection probabilities of the test (3.8) are depicted in Figure 6 where we restrict our attention to the case $\Delta = 0.3^2/30$ for the sake of brevity. We find that for both error settings the test performs reasonably well.
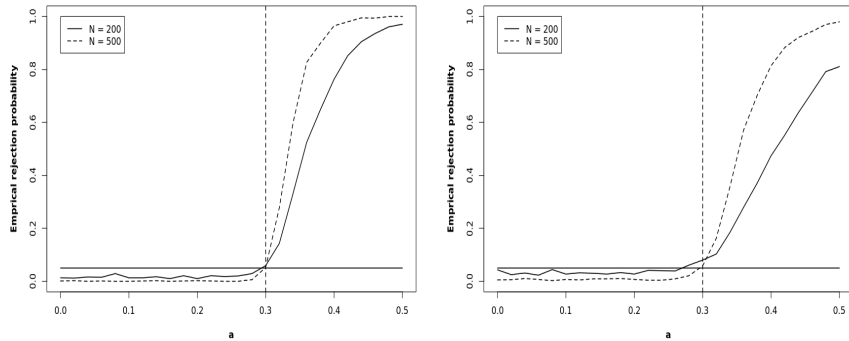
Figure 6: *Simulated rejection probabilities of the test* (3.8) *for the relevant hypotheses* (3.2) *with* $\Delta = 0.3^2/30$ *in the case of* $fMA(1)$ *samples. The mean function after the change point is given by* (5.5) *and the mean function before the change point is the zero function. Left panel: error processes defined by* (5.6). *Right panel: error processes defined by* (5.7).

5.4   Data example

In this section we consider an application of the methodology developed in Section 2.2 to Australian temperature data. The data consists of daily minimum temperatures collected at different meteorological stations in Australia. Following Fremdt et al. (2014) we project the daily values of each year on a Fourier basis consisting of 49 basis functions resulting in annual temperature curves for each location under consideration. These authors investigate the temperature data to illustrate methodology designed to choose the dimension of the projection space obtained with fPCA and in Aue and van Delft (2017) the data is considered in the context of stationarity tests for functional time series. In what follows, we use ν defined by (5.1), i.e. we use the quantiles of the distribution of $\mathbb{W}$ in (2.8) which are displayed in the second line of Table 2.1.

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

In order to calculate the test decision in (2.14) for the hypotheses defined in (2.13), we computed the statistic in (2.15) and the normalizer in (2.16). We obtained $\hat{\mathbb{D}}_{m,n} = 4.115$, $\hat{\mathbb{V}}_{m,n} = 1.201$ and in Table 2, the test decisions are displayed for several choices of the level α and the threshold parameter Δ. In Figure 7 we display the two estimated mean functions graphically.

The results in Table 2 provide no evidence for an integrated squared mean difference larger than $\Delta = 12.1$ but on the other extreme there is strong evidence that it exceeds $\Delta = 9.6$. Choosing Δ between 9.7 and 11.2 led to rejecting the null at level $\alpha \geq 5\%$ and for $\Delta \in [11.3, 12]$, the test rejected the null only at level $\alpha \geq 10\%$, which means weaker support of the alternative.

21

Δ	99%	95%	90%
9.6	TRUE	TRUE	TRUE
9.7	FALSE	TRUE	TRUE
11.2	FALSE	TRUE	TRUE
11.3	FALSE	FALSE	TRUE
12	FALSE	FALSE	TRUE
12.1	FALSE	FALSE	FALSE

Table 2: *Summary of the two sample test 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.*
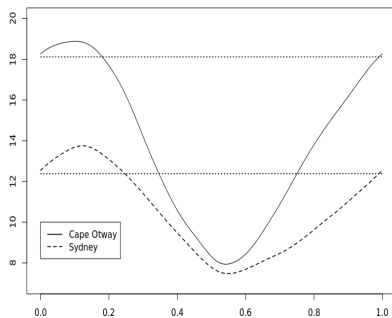

Figure 7: *Mean functions of the Cape Otway and Sydney series for the two sample case.*

References

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

Aue, A., Dubart Norinho, D., and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110:378–392.

Aue, A. and van Delft, A. (2017). Testing for stationarity of functional time series in the frequency domain. *ArXiv e-print 1701.01741v1.*

Benko, M., Haerdle, W., and Kneip, A. (2009). Common functional principal components. *Ann. Statist.*, 37(1):1–34.

Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, 2(3):317–335.

Berkes, I., Gabrys, R., Horváth, L., and Kokoszka, P. (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):927–946.

Berkes, I., Horváth, L., and Rice, G. (2013). Weak invariance principles for sums of dependent random functions. *Stochastic Processes and their Applications*, 123(2):385 – 403.

Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association*, 33(203):526–536.

Bucchia, B. and Wendler, M. (2017). Change-point detection and bootstrap for hilbert space valued random fields. *J. Multivariate Analysis*, 155:344–368.

Bücher, A., Dette, H., and Heinrichs, F. (2018). Detecting deviations from second-order stationarity in locally stationary functional time series. *arXiv:1808.04092*.

Carlstein, E. (1988). Nonparametric change-point estimation. *Ann. Statist.*, 16(1):188–197.

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

Cuevas, A., Febrero, M., and Fraiman, R. (2006). On the use of the bootstrap for estimating functions with functional data. *Computational Statistics & Data Analysis*, 51(2):1063 – 1074.

DasGupta, A. (2008). Asymptotic theory of statistics and probability.

Dümbgen, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *The Annals of Statistics*, 19(3):1471–1495.

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

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

Fremdt, S., Steinebach, J., Horvath, L., and Kokoszka, P. (2013). Testing the equality of covariance operators in functional samples. *Scandinavian Journal of Statistics*, 40(1):138–152.

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

Hariz, S. B., Wylie, J. J., and Zhang, Q. (2007). Optimal rate of convergence for nonparametric change-point estimators for nonstationary sequences. *Ann. Statist.*, 35(4):1802–1826.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Ann. Statist.*, 38(3):1845–1884.

Horváth, L., Kokoska, P., and Reimherr, M. (2009). Two sample inference in functional linear models. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 37(4):571–591.

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

Horváth, L., Kokoszka, P., and Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B*, 75:103–122.

Horváth, L., Kokoszka, P., and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, 179(1):66 – 82.

Kokoszka, P. (2012). Dependent functional data. *ISRN Probability and Statistics*, 2012.

Panaretos, V. M., Kraus, D., and Maddocks, J. H. (2010). Second-order comparison of gaussian random functions and the geometry of dna minicircles. *Journal of the American Statistical Association*, 105(490):670–682.

Paparoditis, E. and Sapatinas, T. (2016). Bootstrap-based testing of equality of mean functions or equality of covariance operators for functional data. *Biometrika*, 103(3):727–733.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer Series in Statistics. Springer New York.

Pomann, G.-M., Staicu, A.-M., and Ghosh, S. (2016). A two-sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(3):395–414.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York, second edition.

Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):343–366.

Shao, X. (2015). Self-normalization for time series: A review of recent developments. *Journal of the American Statistical Association*, 110(512):1797–1817.

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, 105(491):1228–1240.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications in Statistics*. Springer, New York.

Volgushev, S. and Shao, X. (2014). A general approach to the joint asymptotic analysis of statistics from sub-samples. *Electronic Journal of Statistics*, 8(1):390–431.

Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority*. CRC Press, Boca Raton, FL, second edition.

Zhang, C., Peng, H., and Zhang, J.-T. (2010). Two samples tests for functional data. *Communications in Statistics - Theory and Methods*, 39(4):559–578.

Zhang, X. and Shao, X. (2015). Two sample inference for the second-order property of temporally dependent functional data. *Bernoulli*, 21(2):909–929.

Zhang, X., Shao, X., Hayhoe, K., and Wuebbles, D. J. (2011). Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electron. J. Statist.*, 5:1765–1796.

6 Appendix: Proofs

We begin with some technical preliminaries, the notation introduced in this section will be used throughout all the proofs. Define the set of functions

$$\mathcal{G} := \left\{ f : T \times [0,1] \to \mathbb{R} : \sup_{\lambda \in [0,1]} \int_T f^2(t,\lambda) dt < \infty \right\}.$$

Equip this set with the norm

$$\|f - g\|_{\mathcal{G}} := \sup_{\lambda \in [0,1]} \left\{ \int_T [f(t,\lambda) - g(t,\lambda)]^2 dt \right\}^{1/2}$$

to obtain a normed vector space. We will frequently work with random elements with values in this space. Here, random elements need not be measurable and we will make use of the general theory of outer probabilities (see Chapter 1 in Van der Vaart and Wellner (1996)) where appropriate without explicitly mentioning this.

6.1 Proofs of the results in Section 2

First, we state some preliminary results that will be useful throughout the proofs. Define

$$(6.1) \qquad \tilde{S}_n(t,\lambda) = \frac{1}{n} \sum_{j=1}^{\lfloor n\lambda \rfloor} (X_j(t) - \mu(t)), \quad \lambda \in [0,1].$$

Then it follows from Theorem 1.1 in Berkes et al. (2013), that there exists a sequence of measurable random elements in $\mathcal{G}$, say $\{\Gamma_n(t,\lambda)\}_{\lambda,t\in[0,1]}$, such that

$$(6.2) \qquad \sup_{\lambda \in [0,1]} \int_T \left(\sqrt{n}\, \tilde{S}_n(t,\lambda) - \Gamma_n(t,\lambda) \right)^2 dt = o_{\mathbb{P}}(1)$$

$$(6.3) \qquad \{\Gamma_n(t,\lambda)\}_{\lambda,t\in[0,1]} \stackrel{\mathcal{D}}{=} \{\Gamma(t,\lambda)\}_{\lambda,t\in[0,1]},$$

where Γ is defined by

$$(6.4) \qquad \Gamma(t,\lambda) = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t)W_i(\lambda).$$

$\{W_i\}_{i\in\mathbb{N}}$ is a sequence of independent Brownian motions and λ_i, ϕ_i are the eigenvalues and (orthonormal) eigenfunctions of the integral operator corresponding to the covariance kernel

$$(6.5) \qquad C(s,t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s)\phi_i(t)$$

defined in (2.4), that is

$$(6.6) \qquad \lambda_i \phi_i(s) = \int_T C(t,s)\phi_i(t)dt \qquad (i \in \mathbb{N}).$$

Note that Berkes et al. (2013) also prove that $\sum_k \lambda_k < \infty$ (see their Lemma 2.2) and that

$$(6.7) \qquad \sup_{0 \leq \lambda \leq 1} \int_T \Gamma^2(t, \lambda) dt < \infty \quad a.s.$$

The latter implies that for any square integrable function $\zeta : [0, 1] \to \mathbb{R}$ the process

$$\left\{ \int_T \zeta(t) \Gamma(t, \lambda) dt \right\}_{\lambda \in [0,1]}$$

can be viewed as an element of $\ell^\infty([0, 1])$ and that the same is true for the process $\{\int_T \Gamma^2(t, \lambda) dt\}_{\lambda \in [0,1]}$. Moreover, summability of the sequence $(\lambda_k)_{k \in \mathbb{N}}$ together with properties of the modulus of continuity of Brownian motions implies that for any positive sequence $(\kappa_k)_{k \in \mathbb{N}}$ such that $\kappa_n \to 0$, it follows

$$\sup_{\substack{\nu, \lambda \in [0,1]: \\ |\nu - \lambda| \leq \kappa_n}} \int_T \{\Gamma(t, \lambda) - \Gamma(t, \nu)\}^2 dt$$

$$= \sup_{\substack{\nu, \lambda \in [0,1]: \\ |\nu - \lambda| \leq \kappa_n}} \int_T \sum_{i=1}^\infty \sum_{j=1}^\infty \sqrt{\lambda_i \lambda_j} \{W_i(\lambda) - W_i(\nu)\} \{W_j(\lambda) - W_j(\nu)\} \phi_i(t) \phi_j(y) dt$$

$$= \sup_{\substack{\nu, \lambda \in [0,1]: \\ |\nu - \lambda| \leq \kappa_n}} \sum_{i=1}^\infty \sum_{j=1}^\infty \sqrt{\lambda_i \lambda_j} \{W_i(\lambda) - W_i(\nu)\} \{W_j(\lambda) - W_j(\nu)\} \int_T \phi_i(t) \phi_j(y) dt$$

$$= \sup_{\substack{\nu, \lambda \in [0,1]: \\ |\nu - \lambda| \leq \kappa_n}} \sum_{i=1}^\infty \lambda_i \{W_i(\lambda) - W_i(\nu)\}^2$$

$$\leq \sum_{i=1}^\infty \lambda_i \sup_{\substack{\nu, \lambda \in [0,1]: \\ |\nu - \lambda| \leq \kappa_n}} \{W_i(\lambda) - W_i(\nu)\}^2 = o_{\mathbb{P}}(1)$$

where the last line follows since by Fubini's Theorem

$$\mathbb{E}\left[\sum_{i=1}^\infty \lambda_i \sup_{\substack{\nu, \lambda \in [0,1]: \\ |\nu - \lambda| \leq \kappa_n}} \{W_i(\lambda) - W_i(\nu)\}^2 \right] = \mathbb{E}\left[\sup_{\substack{\nu, \lambda \in [0,1]: \\ |\nu - \lambda| \leq \kappa_n}} \{W_1(\lambda) - W_1(\nu)\}^2 \right] \sum_{i=1}^\infty \lambda_i = o(1).$$

This implies

$$(6.8) \qquad \sup_{\substack{\nu, \lambda \in [0,1]: \\ |\nu - \lambda| \leq \kappa_n}} \int_T \{\Gamma^2(t, \lambda) - \Gamma^2(t, \nu)\}^2 dt = o_{\mathbb{P}}(1) \quad \kappa_n \to 0.$$

6.1.1 Proof of Theorem 2.1

The main ingredients of the proof are the convergence result stated in (2.9) when $\int_T \mu^2(t) dt > 0$ and the bounds $\hat{\mathbb{T}}_n = o_{\mathbb{P}}(1), \hat{\mathbb{V}}_n = o_{\mathbb{P}}(1)$ when $\int_T \mu^2(t) dt = 0$. We begin by considering

26

the case $\int_T \mu^2(t)dt = 0$. In that case we have $\tilde{S}_n(t,\lambda) \equiv S_n(t,\lambda)$ for and hence by (6.2) $\sup_{\lambda\in[0,1]} \int_T S_n^2(t,\lambda)dt = o_\mathbb{P}(1)$ which implies $\hat{\mathbb{T}}_n = o_\mathbb{P}(1)$, $\hat{\mathbb{V}}_n = o_\mathbb{P}(1)$.

For the case $\int_T \mu^2(t)dt > 0$ note that a straightforward calculation shows

$$
\begin{aligned}
\sqrt{n}\int_T \big(S_n^2(t,\lambda) - \lambda^2\mu^2(t)\big)dt &= \sqrt{n}\int_T \big(S_n(t,\lambda) - \lambda\mu(t)\big)^2 dt + 2\sqrt{n}\int_T \lambda\mu(t)\big(S_n(t,\lambda) - \lambda\mu(t)\big)dt \\
&= \sqrt{n}\int_T \tilde{S}_n^2(t,\lambda)dt + 2\sqrt{n}\int_T \lambda\mu(t)\tilde{S}_n(t,\lambda)dt + o_\mathbb{P}(1) \\
&= \frac{1}{\sqrt{n}}\int_T \big(\sqrt{n}\tilde{S}_n(t,\lambda) - \Gamma_n(t,\lambda)\big)^2 dt + \frac{1}{\sqrt{n}}\int_T \Gamma_n^2(t,\lambda)dt \\
&\quad - \frac{2}{\sqrt{n}}\int_T \big(\sqrt{n}\tilde{S}_n(t,\lambda) - \Gamma_n(t,\lambda)\big)\Gamma_n(t,\lambda)dt \\
&\quad + 2\int_T \lambda\mu(t)\big(\sqrt{n}\tilde{S}_n(t,\lambda) - \Gamma_n(t,\lambda)\big)dt + 2\int_T \lambda\mu(t)\Gamma_n(t,\lambda)dt + o_\mathbb{P}(1) \\
&= 2\int_T \lambda\mu(t)\Gamma_n(t,\lambda)dt + o_\mathbb{P}(1)
\end{aligned}
$$

uniformly with respect to $\lambda \in [0,1]$, where we repeatedly used (6.2) and the Cauchy-Schwarz inequality. Therefore we obtain from (6.3) and Slutsky's Lemma that

$$
(6.9) \qquad \Big\{\sqrt{n}\Big(\int_T S_n^2(t,\lambda)dt - \lambda^2\int_T \mu^2(t)dt\Big)\Big\}_{\lambda\in[0,1]} \rightsquigarrow \Big\{2\lambda\int_T \mu(t)\Gamma(t,\lambda)dt\Big\}_{\lambda\in[0,1]}
$$

in $\ell^\infty([0,1])$ (recall that by the discussion at the beginning of this section the process on the right hand side is an element of $\ell^\infty([0,1])$), and observing (6.4), (6.5) and (6.6) it follows by a straightforward calculation that

$$
\mathrm{Cov}\Big(\int_T \mu(t)\Gamma(t,\lambda)dt, \int_T \mu(t)\Gamma(t,\lambda')dt\Big) = (\lambda\wedge\lambda')\int_T\int_T \mu(t)\mu(s)C(s,t)dsdt.
$$

Therefore

$$
\Big\{2\lambda\int_T \mu(t)\Gamma(t,\lambda)dt\Big\}_{\lambda\in[0,1]} \overset{\mathcal{D}}{=} \big\{\lambda\tau\mathbb{B}(\lambda)\big\}_{\lambda\in[0,1]},
$$

where $\mathbb{B}$ denotes a standard Brownian motion on the interval $[0,1]$ and τ^2 is defined in (2.3). Consequently the statement (2.9) in Section 2 follows from (6.9) and an application of the continuous mapping theorem observing that the mapping

$$
\mathbb{Z} \mapsto \frac{\mathbb{Z}(1)}{\big(\int_0^1 (\mathbb{Z}(\lambda) - \lambda^2\mathbb{Z}(1))^2\nu(d\lambda)\big)^{1/2}}
$$

from the measurable functions in $\ell^\infty([0,1])$ onto $\mathbb{R}$ is continuous at points $f \in \ell^\infty([0,1])$ with $\int_0^1 (f(\lambda) - \lambda^2 f(1))^2\nu(d\lambda) \neq 0$. This yields

$$
(6.10) \qquad \frac{\hat{\mathbb{T}}_n - d}{\hat{\mathbb{V}}_n} \overset{\mathcal{D}}{\to} \frac{\mathbb{B}(1)}{\big(\int_0^1 \lambda^2(\mathbb{B}(\lambda) - \lambda\mathbb{B}(1))^2\nu(d\lambda)\big)^{1/2}} = \mathbb{W}.
$$

27

In the discussion following Theorem 2.1 we already argued that

$$\lim_{n\to\infty} \mathbb{P}\Big(\hat{\mathbb{T}}_n > \Delta + q_{1-\alpha}(\mathbb{W})\hat{\mathbb{V}}_n\Big) = 0\,,$$

whenever $\int_T \mu^2(t)dt = 0$. When $0 < \int_T \mu^2(t)dt < \Delta$, we have

$$\lim_{n\to\infty} \mathbb{P}\Big(\hat{\mathbb{T}}_n > \Delta + q_{1-\alpha}(\mathbb{W})\hat{\mathbb{V}}_n\Big) = \lim_{n\to\infty} \mathbb{P}\Big(\frac{\hat{\mathbb{T}}_n - d}{\hat{\mathbb{V}}_n} > \frac{\sqrt{n}(\Delta - d)}{\sqrt{n}\,\hat{\mathbb{V}}_n} + q_{1-\alpha}(\mathbb{W})\Big) = 0$$

since $\sqrt{n}\,\hat{\mathbb{V}}_n = O_{\mathbb{P}}(1), \frac{\hat{\mathbb{T}}_n - d}{\hat{\mathbb{V}}_n} = O_{\mathbb{P}}(1)$, $\hat{\mathbb{V}}_n \geq 0$ a.s. and $\sqrt{n}(\Delta - d) \to +\infty$. In the case $d = \int_T \mu^2(t)dt = \Delta$ we conclude

$$\lim_{n\to\infty} \mathbb{P}\Big(\hat{\mathbb{T}}_n > \Delta + q_{1-\alpha}(\mathbb{W})\hat{\mathbb{V}}_n\Big) = \lim_{n\to\infty} \mathbb{P}\Big(\frac{\hat{\mathbb{T}}_n - d}{\hat{\mathbb{V}}_n} > q_{1-\alpha}(\mathbb{W})\Big) = \alpha$$

and, if $d > \Delta$, we have

$$\lim_{n\to\infty} \mathbb{P}\Big(\hat{\mathbb{T}}_n > \Delta + q_{1-\alpha}(\mathbb{W})\hat{\mathbb{V}}_n\Big) = \lim_{n\to\infty} \mathbb{P}\Big(\frac{\hat{\mathbb{T}}_n - d}{\hat{\mathbb{V}}_n} > \frac{\sqrt{n}(\Delta - d)}{\sqrt{n}\,\hat{\mathbb{V}}_n} + q_{1-\alpha}(\mathbb{W})\Big) = 1$$

since $\sqrt{n}\,\hat{\mathbb{V}}_n = O_{\mathbb{P}}(1)$, $\hat{\mathbb{V}}_n \geq 0$ a.s. and $\sqrt{n}(\Delta - d) \to -\infty$. $\qquad\square$

6.1.2 Proof of Theorem 2.2

The processes $\{X_m\}_{m\in\mathbb{Z}}$ and $\{Y_n\}_{n\in\mathbb{Z}}$ satisfy assumptions (B2) and thus admit the representation $X_j = \mu_1 + \eta_j^X, Y_j = \mu_2 + \eta_j^Y$ where $(\eta_j^X)_{j\in\mathbb{Z}}$ and $(\eta_j^Y)_{j\in\mathbb{Z}}$ denote centered error processes that both satisfy (A2)-(A4). Define

$$\tilde{D}_{m,n}(t, \lambda) := S_m^X(t, \lambda) - S_n^Y(t, \lambda)\,,$$

where the processes S_m^X and S_n^Y are given by

$$S_m^X(t, \lambda) = \frac{1}{m}\sum_{j=1}^{\lfloor m\lambda\rfloor}\big(X_j(t) - \mu_1(t)\big) = \frac{1}{m}\sum_{j=1}^{\lfloor m\lambda\rfloor}\eta_j^X(t)\,,$$

$$S_n^Y(t, \lambda) = \frac{1}{n}\sum_{j=1}^{\lfloor n\lambda\rfloor}\big(Y_j(t) - \mu_2(t)\big) = \frac{1}{n}\sum_{j=1}^{\lfloor n\lambda\rfloor}\eta_j^Y(t)\,,$$

respectively. A similar calculation as given in Section 6.1.1 shows that

$$\begin{aligned}
\mathbb{Z}_{m,n}(\lambda) &:= \sqrt{n+m}\int_T\big(D_{m,n}^2(t, \lambda) - \lambda^2 D^2(t)\big)dt \\
&= 2\sqrt{n+m}\int_T \lambda D(t)\big(D_{m,n}(t, \lambda) - \lambda D(t)\big)dt + o_{\mathbb{P}}(1) \\
&= 2\sqrt{n+m}\int_T \lambda D(t)\tilde{D}_{m,n}(t, \lambda)dt + o_{\mathbb{P}}(1) = \mathbb{Z}_m^X(\lambda) - \mathbb{Z}_n^Y(\lambda) + o_{\mathbb{P}}(1)\,,
\end{aligned}$$

28

where we use the fact that $\lambda\mu_1 = \frac{1}{m}\sum_{i=1}^{\lfloor m\lambda \rfloor} \mu_1 + o(1)$ uniformly in λ in the third equality and the processes $\{\mathbb{Z}_m^X(\lambda)\}_{\lambda \in [0,1]}$ and $\{\mathbb{Z}_n^Y(\lambda)\}_{\lambda \in [0,1]}$ are given by

$$(6.11) \qquad \mathbb{Z}_m^X(\lambda) := 2\sqrt{n+m} \int_T \lambda D(t) S_m^X(t,\lambda)dt\,,$$

$$(6.12) \qquad \mathbb{Z}_n^Y(\lambda) := 2\sqrt{n+m} \int_T \lambda D(t) S_n^Y(t,\lambda)dt\,,$$

respectively. As the times series $\{X_n\}_{n\in\mathbb{Z}}$ and $\{Y_n\}_{n\in\mathbb{Z}}$ satisfy assumptions (A1) - (A4) it follows from the proof of Theorem 2.1 that the processes $\{\mathbb{Z}_m^X(\lambda)\}_{\lambda \in [0,1]}$ and $\{\mathbb{Z}_n^Y(\lambda)\}_{\lambda \in [0,1]}$ converge weakly in $\ell^\infty([0,1])$ with both limits corresponding to scaled Brownian motions. Furthermore, both processes are independent and therefore

$$\{\mathbb{Z}_{m,n}(\lambda)\}_{\lambda\in[0,1]} \rightsquigarrow \{\lambda\tau_D\mathbb{B}(\lambda)\}_{\lambda\in[0,1]}$$

in $\ell^\infty([0,1])$, where $\{\mathbb{B}(\lambda)\}_{\lambda\in[0,1]}$ is a Brownian motion and τ_D is a real number depending on the auto-covariance structures of $\{X_m\}_{m\in\mathbb{Z}}$ and $\{Y_n\}_{n\in\mathbb{Z}}$.

The assertion now follows exactly in the same way as in the proof of Theorem 2.1 and the details are omitted for the sake of brevity. $\qquad\square$

6.2 Proofs of the results in Section 3

6.2.1 A technical result

In this section we prove a useful intermediate result. Now consider the situation which is described in model (3.1) (see also Remark 2.2). Set

$$\eta_i^{(1)} = f_1(\varepsilon_i, \varepsilon_{i-1,\dots}), \quad i = 1,\dots,N$$
$$\eta_i^{(2)} = f_2(\varepsilon_i, \varepsilon_{i-1,\dots}), \quad i = 1,\dots,N$$

where $f_1, f_2, (\varepsilon_j)_{j\in\mathbb{Z}}$ satisfy the conditions in (A2), $\mathbb{E}[\eta_1^{(1)}] = \mathbb{E}[\eta_1^{(2)}] = 0$, and $\eta_i^{(1)}, \eta_i^{(2)}$ satisfy assumptions (A3), (A4).

Lemma 6.1 *In the setting above consider a fixed (but arbitrary) function ζ in $L^2(T)$. For $\lambda \in [0,1]$ define the processes*

$$\tilde{Z}_N^{(k)}(\lambda) := \frac{1}{\sqrt{N}} \sum_{i=1}^{\lfloor N\lambda \rfloor} \int_T \eta_i^{(k)}(t)\zeta(t)dt, \quad k = 1,2\,,$$

then

$$(\tilde{Z}_N^{(1)}, \tilde{Z}_N^{(2)})^\top \rightsquigarrow \Sigma^{1/2}(\mathbb{B}_1, \mathbb{B}_2)^\top \quad in\ \ell^\infty([0,1])^2\,,$$

where $\mathbb{B}_1, \mathbb{B}_2$ are two independent standard Brownian motions on the interval $[0,1]$ and Σ is a symmetric 2×2 matrix with finite entries given by

$$\Sigma_{ij} = \sum_{h\in\mathbb{Z}} \int_T \int_T Cov(\eta_0^{(i)}(s), \eta_h^{(j)}(t))\zeta(s)\zeta(t)dsdt\,.$$

Proof It suffices to prove tightness of the processes $\tilde{Z}_N^{(1)}, \tilde{Z}_N^{(2)}$ individually and joint finite-dimensional convergence. Tightness of $\tilde{Z}_N^{(k)}$ follows from weak convergence of $\tilde{Z}_N^{(k)}$, which can be established by an application of Theorem 2.1 in Berkes et al. (2013) and similar arguments as given in Section 6.1.1. Thus it remains to prove that for arbitrary $s_1, ..., s_L \in [0, 1]$ we have

$$(\tilde{Z}_N^{(1)}(s_1), \tilde{Z}_N^{(2)}(s_1), ..., \tilde{Z}_N^{(1)}(s_L), \tilde{Z}_N^{(2)}(s_L)) \rightsquigarrow \mathcal{N}(0, \Sigma(s_1, ..., s_L))$$

where $\Sigma(s_1, ..., s_L)$ denotes the covariance matrix of the vector $(\mathbb{G}_1(s_1), \mathbb{G}_2(s_1), ..., \mathbb{G}_1(s_L), \mathbb{G}_2(s_L))$ and $(\mathbb{G}_1, \mathbb{G}_2)^\top := \Sigma(\mathbb{B}_1, \mathbb{B}_2)^\top$. Following Berkes et al. (2013) we define the random variables

$$\eta_{j,m}^{(k)} := f_k(\varepsilon_j, \varepsilon_{j-1}, \dots \varepsilon_{j-m+1}, \boldsymbol{\varepsilon}_{j,m}^*), \quad k = 1, 2, m \in \mathbb{N},$$

where $\boldsymbol{\varepsilon}_{j,m}^* = (\varepsilon_{j,m,j-m}^*, \varepsilon_{j,m,j-m-1}^*, \dots)$ is given in Assumption (A4). Let Σ_m denote matrices with entries (below we shall prove that all entries are finite for any $m \geq 1$)

$$(\Sigma_m)_{ij} := \sum_{|h| \leq m} \int_T \int_T \mathrm{Cov}(\eta_{0,m}^{(i)}(s), \eta_{h,m}^{(j)}(t))\zeta(s)\zeta(t)dsdt,$$

define $(\mathbb{G}_{1,m}, \mathbb{G}_{2,m})^\top := \Sigma_m^{1/2}(\mathbb{B}_1, \mathbb{B}_2)^\top$ and

$$\tilde{Z}_{N,m}^{(k)}(\lambda) := \frac{1}{\sqrt{N}} \sum_{i=1}^{\lfloor N\lambda \rfloor} \int_T \eta_{i,m}^{(k)}(t)\zeta(t)dt, \quad k = 1, 2.$$

By an application of Example 11 in Chapter IV of Pollard (1984) it suffices to prove that

(i) $\Sigma_m \to \Sigma$ as $m \to \infty$.

(ii) For any $\delta > 0, k \in \{1, 2\}, \ell \in \{1, ..., L\}$ we have

$$\lim_{m \to \infty} \limsup_{N \to \infty} \mathbb{P}\left(|\tilde{Z}_{N,m}^{(k)}(s_\ell) - \tilde{Z}_N^{(k)}(s_\ell)| > \delta \right) = 0.$$

(iii) For any fixed m we have

$$(\tilde{Z}_{N,m}^{(1)}(s_1), \tilde{Z}_{N,m}^{(2)}(s_1), ..., \tilde{Z}_{N,m}^{(1)}(s_L), \tilde{Z}_{N,m}^{(2)}(s_L)) \rightsquigarrow \mathcal{N}(0, \Sigma_m(s_1, ..., s_L)).$$

In order to show the claim in (i), we prove $(\Sigma_m)_{ij} \to \Sigma_{ij}$ as $m \to \infty$ for $i, j = 1, 2$. For $i = j$ this assertion directly follows from Lemma 2.2 in Berkes et al. (2013). For $i \neq j$, one can use similar arguments as in the proof of the latter Lemma. More precisely, assume without loss of generality that $i = 1, j = 2$ and show that

(6.13) $|\Sigma_{12}| < \infty,$

(6.14) $|(\Sigma_m)_{12}| < \infty, \quad m \geq 1,$

(6.15) $(\Sigma_m)_{12} \to \Sigma_{12}, \quad \text{as } m \to \infty.$

30

From the estimate

(6.16)
$$|\Sigma_{12}| \leq \left| \int_T \int_T \mathbb{E}[\eta_0^{(1)}(s)\eta_0^{(2)}(t)]\zeta(s)\zeta(t)dsdt \right| + 2\left| \sum_{h=1}^{\infty} \int_T \int_T \mathbb{E}[\eta_0^{(1)}(s)\eta_h^{(2)}(t)]\zeta(s)\zeta(t)dsdt \right|,$$

(6.13) follows if each of the terms above is finite. For the first term, we apply the Cauchy-Schwarz inequality to obtain by (A3)

$$\left| \int_T \int_T \mathbb{E}[\eta_0^{(1)}(s)\eta_0^{(2)}(t)]\zeta(s)\zeta(t)dsdt \right| \leq \|\zeta\|^2 \left(\int_T \int_T \mathbb{E}[\eta_0^{(1)}(s)\eta_0^{(2)}(t)]^2 dsdt \right)^{1/2}$$

$$\leq \|\zeta\|^2 \left(\int_T \int_T \mathbb{E}[\eta_0^{(1)}(s)^2]\mathbb{E}[\eta_0^{(2)}(t)^2] dsdt \right)^{1/2}$$

$$= \|\zeta\|^2 \left(\mathbb{E}\|\eta_0^{(1)}\|^2 \mathbb{E}\|\eta_0^{(2)}\|^2 \right)^{1/2} < \infty.$$

We proceed with the second term in (6.16) and drop the constant 2. For any $i \geq 1$, $\mathbb{E}[\eta_0^{(1)}\eta_{i,i}^{(2)}] = \mathbb{E}[\eta_0^{(1)}]\mathbb{E}[\eta_{i,i}^{(2)}] = 0$ since $\eta_0^{(1)}$ and $\eta_{i,i}^{(2)}$ are independent. Applying the triangle inequality and the Cauchy-Schwarz inequality yields

(6.17)
$$\left| \sum_{h=1}^{\infty} \int_T \int_T \mathbb{E}[\eta_0^{(1)}(s)\eta_h^{(2)}(t)]\zeta(s)\zeta(t)dsdt \right|$$

$$\leq \sum_{h=1}^{\infty} \|\zeta\|^2 \left(\int_T \int_T \mathbb{E}[\eta_0^{(1)}(s)\eta_h^{(2)}(t)]^2 dsdt \right)^{1/2}$$

$$= \sum_{h=1}^{\infty} \|\zeta\|^2 \left(\int_T \int_T \mathbb{E}[\eta_0^{(1)}(s)(\eta_h^{(2)}(t) - \eta_{h,h}^{(2)}(t))]^2 dsdt \right)^{1/2}$$

$$\leq \sum_{h=1}^{\infty} \|\zeta\|^2 \left(\int_T \int_T \mathbb{E}[\eta_0^{(1)}(s)^2]\mathbb{E}[(\eta_h^{(2)}(t) - \eta_{h,h}^{(2)}(t))^2] dsdt \right)^{1/2}$$

$$= \|\zeta\|^2 \mathbb{E}\left[\|\eta_0^{(1)}\|^2\right]^{1/2} \sum_{h=1}^{\infty} \mathbb{E}\left[\|\eta_0^{(2)} - \eta_{0,h}^{(2)}\|^2\right]^{1/2}.$$

Due to condition (A3) we have $\mathbb{E}\left[\|\eta_0^{(1)}(s)\|^2\right]^{1/2} < \infty$ and by an application of the Hölder inequality we get

$$\sum_{h=1}^{\infty} \mathbb{E}\left[\|\eta_h^{(2)} - \eta_{h,h}^{(2)}\|^2\right]^{1/2} \leq \sum_{h=1}^{\infty} \mathbb{E}\left[\|\eta_0^{(2)} - \eta_{0,h}^{(2)}\|^{2+\delta}\right]^{1/(2+\delta)}$$

which is finite by (A4). This completes the proof of (6.13).

For $(\Sigma_m)_{12}$, we proceed similarly. We have $\mathbb{E}[\eta_0^{(1)}(s)\eta_0^{(2)}(t)] = \mathbb{E}[\eta_{0,m}^{(1)}(s)\eta_{0,m}^{(2)}(t)]$ and therefore

$$\left| \int_T \int_T \mathbb{E}[\eta_{0,m}^{(1)}(s)\eta_{0,m}^{(2)}(t)]\zeta(s)\zeta(t)dsdt \right| < \infty.$$

Note that the vectors $(\eta_{0,m}^{(1)}, \eta_{h,m}^{(2)})$ and $(\eta_0^{(1)}, \eta_{h,m}^{(2)})$ have the same distribution for all $h = 1, \ldots, m$. Furthermore, the vectors $(\eta_{h,m}^{(1)}, \eta_{h,h}^{(2)})$ and $(\eta_0^{(1)}, \eta_{0,h}^{(2)})$ have the same distribution for all $h = 1, \ldots, m$ (this follows from the definition of all quantities involved and the i.i.d. structure of the $\varepsilon_{i,j,m}^*$). Thus, using similar arguments as in (6.17), we obtain

$$\left| \sum_{h=1}^m \int_T \int_T \mathbb{E}[\eta_{0,m}^{(1)}(s)\eta_{h,m}^{(2)}(t)]\zeta(s)\zeta(t)dsdt \right| \leq \|\zeta\|^2 \sum_{h=1}^m \left(\int_T \int_T \mathbb{E}[\eta_{0,m}^{(1)}(s)\eta_{h,m}^{(2)}(t)]^2 dsdt \right)^{1/2}$$

$$= \|\zeta\|^2 \sum_{h=1}^m \left(\int_T \int_T \mathbb{E}[\eta_0^{(1)}(s)\eta_{h,m}^{(2)}(t)]^2 dsdt \right)^{1/2} \leq \|\zeta\|^2 \mathbb{E}[\|\eta_0^{(1)}\|^2]^{1/2} \sum_{h=1}^m \mathbb{E}[\|\eta_{h,m}^{(2)} - \eta_{h,h}^{(2)}\|^2]^{1/2}$$

$$\leq \|\zeta\|^2 \mathbb{E}[\|\eta_0^{(1)}\|^2]^{1/2} \sum_{h=1}^\infty \mathbb{E}[\|\eta_0^{(2)} - \eta_{0,h}^{(2)}\|^2]^{1/2} < \infty$$

which proves (6.14).

In order to establish (6.15), we begin by observing that

$$\left| \sum_{h=0}^\infty \int_T \int_T \mathbb{E}[\eta_0^{(1)}(s)\eta_h^{(2)}(t)]\zeta(s)\zeta(t)dsdt - \sum_{h=0}^m \int_T \int_T \mathbb{E}[\eta_{0,m}^{(1)}(s)\eta_{h,m}^{(2)}(t)]\zeta(s)\zeta(t)dsdt \right|$$

$$\leq \left| \sum_{h=1}^m \int_T \int_T \{\mathbb{E}[\eta_0^{(1)}(s)\eta_h^{(2)}(t)] - \mathbb{E}[\eta_{0,m}^{(1)}(s)\eta_{h,m}^{(2)}(t)]\}\zeta(s)\zeta(t)dsdt \right|$$

$$+ \sum_{h=m+1}^\infty \left| \int_T \int_T \mathbb{E}[\eta_0^{(1)}(s)\eta_h^{(2)}(t)]\zeta(s)\zeta(t)dsdt \right|$$

$$\leq \sum_{h=1}^m \left| \int_T \int_T \mathbb{E}\left[\eta_0^{(1)}(s)\{\eta_h^{(2)}(t) - \eta_{h,m}^{(2)}(t)\}\right]\zeta(s)\zeta(t)dsdt \right|$$

$$+ \|\zeta\|^2 \mathbb{E}[\|\eta_0^{(1)}\|^2]^{1/2} \sum_{h=m+1}^\infty \mathbb{E}[\|\eta_0^{(2)} - \eta_{0,h}^{(2)}\|^2]^{1/2},$$

where the last inequality follows by similar arguments as (6.17). Now the second term converges to zero as $m \to \infty$ and for the first term we obtain

$$\sum_{h=1}^m \left| \int_T \int_T \mathbb{E}\left[\eta_0^{(1)}(s)\{\eta_h^{(2)}(t) - \eta_{h,m}^{(2)}(t)\}\right]\zeta(s)\zeta(t)dsdt \right|$$

$$\leq \|\zeta\|^2 \sum_{h=1}^m \left(\int_T \int_T \mathbb{E}\left[\eta_0^{(1)}(s)\{\eta_h^{(2)}(t) - \eta_{h,m}^{(2)}(t)\}\right]^2 dsdt \right)^{1/2}.$$

Now by (A4) we have for any fixed h

$$\left(\int_T \int_T \mathbb{E}\left[\eta_0^{(1)}(s)\{\eta_h^{(2)}(t) - \eta_{h,m}^{(2)}(t)\}\right]^2 dsdt \right)^{1/2} \leq \mathbb{E}[\|\eta_0^{(1)}\|^2]^{1/2} \mathbb{E}[\|\eta_h^{(2)} - \eta_{h,m}^{(2)}\|^2]^{1/2}$$

$$= \mathbb{E}[\|\eta_0^{(1)}\|^2]^{1/2} \mathbb{E}[\|\eta_0^{(2)} - \eta_{0,m}^{(2)}\|^2]^{1/2} \to 0.$$

Since also by similar arguments as in (6.17) and in the proof of (6.14)

$$\left(\int_T \int_T \mathbb{E}\left[\eta_0^{(1)}(s)\{\eta_h^{(2)}(t) - \eta_{h,m}^{(2)}(t)\} \right]^2 dsdt \right)^{1/2}$$

$$\leq 2 \left\{ \left(\int_T \int_T \mathbb{E}\left[\eta_0^{(1)}(s)\eta_h^{(2)}(t) \right]^2 dsdt \right)^{1/2} + \left(\int_T \int_T \mathbb{E}\left[\eta_0^{(1)}(s)\eta_{h,m}^{(2)}(t) \right]^2 dsdt \right)^{1/2} \right\}$$

$$\leq 4 \, \mathbb{E}\left[\|\eta_0^{(1)}\|^2 \right]^{1/2} \mathbb{E}\left[\|\eta_0^{(2)} - \eta_{0,h}^{(2)}\|^2 \right]^{1/2}$$

and since the right-hand side is summable over $h \geq 1$ it follows that

$$\sum_{h=1}^{m} \left| \int_T \int_T \mathbb{E}\left[\eta_0^{(1)}(s)\{\eta_h^{(2)}(t) - \eta_{h,m}^{(2)}(t)\} \right] \zeta(s)\zeta(t)dsdt \right| \to 0, \quad m \to \infty$$

by the dominated convergence theorem for series.

The assertion in (6.15) follows and this also completes the proof of (i).

The claim in (ii) follows by a direct application of Lemma 2.1 in Berkes et al. (2013).

For a proof of claim (iii) note that for each fixed m the sequence

$$\left(\int_T \eta_{i,m}^{(1)}(t)\zeta(t)dt , \int_T \eta_{i,m}^{(2)}(t)\zeta(t)dt \right)_{i \in \mathbb{Z}}$$

form a collection of stationary, m-dependent random vectors with finite variance. Now (iii) follows by a straightforward application of the Cramer-Wold device and the CLT for m-dependent random variables, see for instance Theorem 9.1 in DasGupta (2008). $\qquad \square$

6.2.2 Proof of Proposition 3.1

Step 1: Recall the definition of $\hat{f}$ in (3.4). We begin by proving the following preliminary result

(6.18) $$\{ \mathbb{G}_N(\theta) \}_{\theta \in [0,1]} := \{ \sqrt{N}(\hat{f}(\lfloor N\theta \rfloor) - d(\theta)) \}_{\theta \in [0,1]} \rightsquigarrow \{ \mathbb{G}(\theta) \}_{\theta \in [0,1]}$$

in $\ell^\infty([0,1])$ as $N \to \infty$, where

(6.19) $$d(\theta) := \tilde{d}(\theta) \int_T \delta(t)^2 dt, \quad \tilde{d}(\theta) = \theta(1-\theta) \begin{cases} (\theta_0/\theta)^2, & 1 > \theta > \theta_0 \\ ((1-\theta_0)/(1-\theta))^2, & 0 < \theta \leq \theta_0 \end{cases}$$

$\tilde{d}(0) = \tilde{d}(1) = 0$, and the process $\mathbb{G}$ is a random element in $\ell^\infty([0,1])$ with a.s. continuous sample paths. To this end define for $k = 1, ..., N-1$

$$A_N(t,k) := \frac{1}{k}\sum_{j=1}^{k}(X_j(t) - \mathbb{E}[X_j(t)]) - \frac{1}{N-k}\sum_{j=k+1}^{N}(X_j(t) - \mathbb{E}[X_j(t)])$$

$$B_N(t,k) := \frac{1}{k}\sum_{j=1}^{k}\mathbb{E}[X_j(t)] - \frac{1}{N-k}\sum_{j=k+1}^{N}\mathbb{E}[X_j(t)]$$

and let $A_N(t, N) = A_N(t, 0) = B_N(t, N) = B_N(t, 0) \equiv 0$. With those definitions we can write

(6.20)
$$\hat{f}(k) = \int_T (A_N(t, k) + B_N(t, k))^2 dt \, \frac{k}{N}\left(1 - \frac{k}{N}\right)$$
$$= \left\{ \int_T A_N(t, k)^2 dt + 2\int_T A_N(t, k)B_N(t, k)dt + \int_T B_N(t, k)^2 dt \right\} \frac{k}{N}\left(1 - \frac{k}{N}\right).$$

From Theorem 1.1 in Berkes et al. (2013) it follows that

(6.21)
$$\frac{k}{N}\left(1 - \frac{k}{N}\right) \int_T A_N(t, k)^2 dt = o_{\mathbb{P}}(N^{-1/2}),$$

uniformly with respect to k. For $1 \le k \le k_0 := \lfloor N\theta_0 \rfloor$, straightforward calculations yield

$$B_N(t, k) = \frac{1}{k}\sum_{j=1}^{k}\mu(t) - \frac{1}{N-k}\sum_{j=k+1}^{k_0}\mu(t) - \frac{1}{N-k}\sum_{j=k_0+1}^{N}(\mu(t) + \delta(t))$$
$$= \mu(t)\left(1 - \frac{k_0 - k}{N - k}\right) - (\mu(t) + \delta(t))\frac{N - k_0}{N - k}$$
$$= \mu(t)\left(1 - \frac{\theta_0 - k/N}{1 - k/N}\right) - (\mu(t) + \delta(t))\frac{1 - \theta_0}{1 - k/N} + O(N^{-1})$$
$$= -\frac{1 - \theta_0}{1 - k/N}\delta(t) + O(N^{-1})$$

and in the case $N > k > k_0$ we have (again uniformly in k)

$$B_N(t, k) = \frac{1}{k}\sum_{j=1}^{k_0}\mu(t) + \frac{1}{k}\sum_{j=k_0+1}^{k}(\mu(t) + \delta(t)) - \frac{1}{N-k}\sum_{j=k+1}^{N}(\mu(t) + \delta(t))$$
$$= \frac{\theta_0}{k/N}\mu(t) + \frac{k/N - \theta_0}{k/N}(\mu(t) + \delta(t)) - \frac{1 - k/N}{1 - k/N}(\mu(t) + \delta(t)) + O(N^{-1})$$
$$= -\frac{\theta_0}{k/N}\delta(t) + O(N^{-1}).$$

Hence we obtain

(6.22)
$$\frac{k}{N}\left(1 - \frac{k}{N}\right)\int_T B_N(t, k)^2 dt = d(k/N) + O(N^{-1}),$$

uniformly with respect to k and

$$\frac{k}{N}\left(1 - \frac{k}{N}\right)\int_T A_N(t, k)B_N(t, k)dt = \int_T A_N(t, k)\delta(t)dt \, \tilde{d}(k/N) + o_{\mathbb{P}}(N^{-1/2}).$$

Therefore we obtain from (6.20), (6.21), (6.22), Lipschitz continuity of $\theta \mapsto d(\theta), \theta \mapsto \tilde{d}(\theta)$ and the line above

(6.23)
$$\mathbb{G}_N(\theta) = 2\sqrt{N}\left\{\int_T A_N(t, \lfloor N\theta \rfloor)\delta(t)dt \, \tilde{d}(\theta)\right\} + o_{\mathbb{P}}(1)$$

uniformly with respect to $\theta \in [0,1]$. In order to investigate the leading term on the right hand side observe for any $\theta \in [1/N, 1)$ the representation

$$
A_N(t, \lfloor \theta N \rfloor) = \frac{1}{\lfloor \theta N \rfloor} \left(\sum_{j=1}^{\lfloor (\theta \wedge \theta_0) N \rfloor} \eta_j^{(1)}(t) + \mathbb{1}\{\theta_0 < \theta\} \sum_{j=\lfloor \theta_0 N \rfloor + 1}^{\lfloor \theta N \rfloor} \eta_j^{(2)}(t) \right)
$$
$$
- \frac{1}{N - \lfloor \theta N \rfloor} \left(\mathbb{1}\{\theta_0 \geq \theta\} \sum_{j=\lfloor \theta N \rfloor + 1}^{\lfloor \theta_0 N \rfloor} \eta_j^{(1)}(t) + \sum_{j=\lfloor (\theta \vee \theta_0) N \rfloor + 1}^{N} \eta_j^{(2)}(t) \right)
$$
$$
= \frac{1}{\lfloor \theta N \rfloor} \sum_{j=1}^{\lfloor (\theta \wedge \theta_0) N \rfloor} \eta_j^{(1)}(t) - \frac{\mathbb{1}\{\theta_0 \geq \theta\}}{N - \lfloor \theta N \rfloor} \sum_{j=\lfloor \theta N \rfloor + 1}^{\lfloor \theta_0 N \rfloor} \eta_j^{(1)}(t)
$$
$$
+ \frac{\mathbb{1}\{\theta_0 < \theta\}}{\lfloor \theta N \rfloor} \sum_{j=\lfloor \theta_0 N \rfloor + 1}^{\lfloor \theta N \rfloor} \eta_j^{(2)}(t) - \frac{1}{N - \lfloor \theta N \rfloor} \sum_{j=\lfloor (\theta \vee \theta_0) N \rfloor + 1}^{N} \eta_j^{(2)}(t),
$$

which yields

$$
2\sqrt{N} \int_T A_N(t, \lfloor N\theta \rfloor) \delta(t) dt
$$
$$
= \frac{N}{\lfloor \theta N \rfloor} \tilde{\mathbb{Z}}_N^{(1)}(\theta \wedge \theta_0) - \mathbb{1}\{\theta_0 \geq \theta\} \frac{N}{N - \lfloor \theta N \rfloor} (\tilde{\mathbb{Z}}_N^{(1)}(\theta_0) - \tilde{\mathbb{Z}}_N^{(1)}(\theta))
$$
$$
+ \mathbb{1}\{\theta_0 < \theta\} \frac{N}{\lfloor \theta N \rfloor} (\tilde{\mathbb{Z}}_N^{(2)}(\theta) - \tilde{\mathbb{Z}}_N^{(2)}(\theta_0)) - \frac{N}{N - \lfloor \theta N \rfloor} (\tilde{\mathbb{Z}}_N^{(2)}(1) - \tilde{\mathbb{Z}}_N^{(2)}(\theta_0 \vee \theta)) \, ,
$$

where

(6.24)
$$
\tilde{\mathbb{Z}}_N^{(i)}(\lambda) = \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor \lambda N \rfloor} \int_T \eta_j^{(i)}(t) \delta(t) dt \, .
$$

Finally, note that we have

$$
\sup_{\theta \in [1/N, 1)} \left| \tilde{d}(\theta) \frac{N}{\lfloor \theta N \rfloor} - \frac{\tilde{d}(\theta)}{\theta} \right| = o(1), \quad \sup_{\theta \in [1/N, 1)} \left| \tilde{d}(\theta) \frac{N}{N - \lfloor \theta N \rfloor} - \frac{\tilde{d}(\theta)}{1 - \theta} \right| = o(1) \, .
$$

Hence Lemma 6.1, Slutskys Lemma and the continuous mapping theorem yield

(6.25)
$$
2\sqrt{N} \left\{ \int_T A_N(t, \lfloor N\theta \rfloor) \delta(t) dt \, \tilde{d}(\theta) \right\}_{\theta \in [0,1]} \rightsquigarrow \{\mathbb{G}(\theta)\}_{\theta \in [0,1]} \, .
$$

Combing (6.25) with (6.23) gives us the weak convergence in (6.18).

Step 2: Given the weak convergence in (6.18) we are ready to prove (3.5). The proof will proceed in three steps. First, we show that $\hat{\theta} = \theta_0 + o_{\mathbb{P}}(1)$. In the second step we show that

(6.26)
$$
\hat{\theta} = \theta_0 + o_{\mathbb{P}}(N^{-1/4}).
$$

35

In the final step we derive (3.5).

Observe that the function $\theta \mapsto d(\theta)$, defined in (6.19), is strictly increasing in $[0, \theta_0]$ and strictly decreasing in $(\theta_0, 1]$. Therefore, for any $\delta > 0$, there is an $\varepsilon > 0$ such that $|\theta - \theta_0| > \delta$ implies $d(\theta_0) - d(\theta) > \varepsilon$. Now let $\delta > 0$ be arbitrary and assume $|\hat{\theta} - \theta| > \delta$. Using that $\hat{k} := N\hat{\theta}$ is the maximizer of the function $k \mapsto f(k)$, the result from Step 1 and the previously mentioned monotonicity property, we obtain

(6.27)
$$0 \geq \hat{f}(k_0) - \hat{f}(\hat{k}) = \hat{f}(k_0) - d(\theta_0) - (\hat{f}(\hat{k}) - d(\hat{\theta})) + d(\theta_0) - d(\hat{\theta})$$
$$= O_{\mathbb{P}}(N^{-1/2}) + d(\theta_0) - d(\hat{\theta}) > O_{\mathbb{P}}(N^{-1/2}) + \varepsilon$$

for some $\varepsilon > 0$, where $k_0 := \lfloor N\theta_0 \rfloor$. This means that

$$\mathbb{P}(|\hat{\theta} - \theta_0| > \delta) \leq \mathbb{P}(O_{\mathbb{P}}(N^{-1/2}) < -\varepsilon) \to 0$$

as $N \to \infty$ and therefore, $\hat{\theta}$ converges to the true change point θ_0 in probability.

Next we show that $|\hat{\theta} - \theta_0| = O_{\mathbb{P}}(N^{-1/2})$. Making a Taylor expansion of d at the point θ_0, we obtain, as $\theta \to \theta_0$,

$$d(\theta) = d(\theta_0) + c\big(-(\theta - \theta_0)\mathbb{1}\{\theta > \theta_0\} + (\theta - \theta_0)\mathbb{1}\{\theta \leq \theta_0\}\big) + O\big((\theta - \theta_0)^2\big)$$

for some constant $c > 0$. Therefore, as $\theta \to \theta_0$, we can find a constant $\delta > 0$ such that

(6.28)
$$d(\theta_0) - d(\theta) \geq \delta |\theta - \theta_0| + O((\theta - \theta_0)^2).$$

Since $\hat{\theta}$ is a consistent estimator of θ_0 (by the discussion in the previous paragraph), we can use this property and similar arguments as in (6.27) to obtain

$$0 \leq \hat{f}(\hat{k}) - \hat{f}(k_0) = O_{\mathbb{P}}(N^{-1/2}) + d(\hat{\theta}) - d(\theta_0) \leq O_{\mathbb{P}}(N^{-1/2}) - \delta |\hat{\theta} - \theta_0|$$

which means that $|\hat{\theta} - \theta_0| = O_{\mathbb{P}}(N^{-1/2})$.

Thus, with probability converging to 1, we have $\hat{\theta} \in \mathrm{argmax}_{\theta : |\theta - \theta_0| \leq N^{-1/4}} \hat{f}(\lfloor N\theta \rfloor)$. Since the process $\mathbb{G}_N$ in (6.18) is stochastically equicontinuous, we get

$$|\hat{f}(\hat{k}) - \hat{f}(k_0) - (d(\hat{\theta}) - d(\theta_0))| \leq \sup_{\theta : |\theta - \theta_0| \leq N^{-1/4}} |\hat{f}(\lfloor N\theta \rfloor) - \hat{f}(k_0) - (d(\theta) - d(\theta_0))|$$
$$\leq \sup_{\theta, \theta' : |\theta - \theta'| \leq N^{-1/4}} N^{-1/2} |\mathbb{G}_N(\theta) - \mathbb{G}_N(\theta')|$$
$$= o_{\mathbb{P}}(N^{-1/2}).$$

Using this rate and the bound in (6.28) yields

$$0 \leq \hat{f}(\hat{k}) - \hat{f}(k_0) \leq d(\hat{\theta}) - d(\theta_0) + |\hat{f}(\hat{k}) - \hat{f}(k_0) - (d(\hat{\theta}) - d(\theta_0))|$$
$$\leq d(\hat{\theta}) - d(\theta_0) + o_{\mathbb{P}}(N^{-1/2})$$
$$\leq -\delta |\theta_0 - \hat{\theta}| + o_{\mathbb{P}}(N^{-1/2})$$

which finally implies $|\hat{\theta} - \theta_0| = o_{\mathbb{P}}(N^{-1/2})$. $\qquad\square$

6.2.3 Proof of Theorem 3.1

We begin by stating some useful technical results and notations. Define

$$
(6.29) \qquad \tilde{S}_N^{(k)}(t, \lambda) = \frac{1}{N} \sum_{j=1}^{\lfloor \lambda N \rfloor} \eta_j^{(k)}(t), \quad k = 1, 2
$$

where $\eta_j^{(1)} := f_1(\varepsilon_j, \varepsilon_{j-1,\ldots}), \eta_j^{(2)} := f_2(\varepsilon_j, \varepsilon_{j-1,\ldots})$ for $j \in \mathbb{Z}$. Since f_1, f_2 satisfy assumptions (A3), (A4), it follows from Theorem 1.1 in Berkes et al. (2013) that there exist random elements in $\mathcal{G}$ (recall the beginning of Section 6), say $\Gamma_N^{(i)}$, with

$$
(6.30) \qquad \sup_{\lambda \in [0,1]} \int_T (\sqrt{N} \tilde{S}_N^{(i)}(t, \lambda) - \Gamma_N^{(i)}(t, \lambda))^2 dt = o_P(1), \quad i = 1, 2,
$$

where each $\Gamma_N^{(i)}$ satisfies the analogue of (6.3)-(6.8) with covariance kernels corresponding to $\eta_i^{(1)}$ and $\eta_i^{(2)}$, respectively.

First consider the case $\int \delta^2(t) dt \neq 0$. Recalling that $\hat{\theta} = \theta_0 + o_{\mathbb{P}}(N^{-1/2})$ by Proposition 3.1 we proceed in several steps. First, we show that for the process

$$
\mathbb{Z}_N(\lambda, \theta) = \sqrt{N} \int_T \left(D_N^{cp}(t, \lambda, \theta)^2 - \lambda^2 \delta(t)^2 \right) dt,
$$

we have

$$
(6.31) \qquad \left\{ \mathbb{Z}_N(\lambda, \theta_0) \right\}_{\lambda \in [0,1]} \rightsquigarrow \left\{ \lambda \tau_\delta^2 \mathbb{B}(\lambda) \right\}_{\lambda \in [0,1]}
$$

in $\ell^\infty([0,1])$, where $\{\mathbb{B}(\lambda)\}_{\lambda \in [0,1]}$ is a Brownian motion and τ_δ is a parameter depending on the covariance structure of $\{(\eta_j^{(1)}, \eta_j^{(2)})\}_{j \in \mathbb{Z}}$. Second we prove

$$
(6.32) \qquad \sup_{\lambda \in [0,1]} |\mathbb{Z}_N(\lambda, \theta_0) - \mathbb{Z}_N(\lambda, \hat{\theta})| = o_{\mathbb{P}}(1),
$$

where $\hat{\theta}$ is the estimator of θ_0 defined in (3.3). Finally we can again use the same arguments as in the proof of Theorem 2.1 to obtain the assertion.

Next, consider the case $\int_T \delta^2(t) dt = 0$. It suffices to show that $\hat{\mathbb{D}}_N^{cp} = o_{\mathbb{P}}(1), \hat{\mathbb{V}}_N^{cp} = o_{\mathbb{P}}(1)$. To this end define the partial sum process

$$
W_N(t, \lambda) := \frac{1}{N} \left\{ \sum_{i=1}^{\lfloor N(\lambda \wedge \theta_0) \rfloor} \eta_i^{(1)}(t) + \sum_{i=\lfloor N\theta_0 \rfloor + 1}^{\lfloor N(\lambda \vee \theta_0) \rfloor} \eta_i^{(2)}(t) \right\}
$$

and observe that by (6.30) and some elementary computations we have

$$
\sup_{\lambda \in [0,1]} \int_T W_N^2(t, \lambda) dt = o_{\mathbb{P}}(1).
$$

Next, observing that

$$D_N^{cp}(t, \lambda, \theta)$$
$$= \frac{N}{\lfloor N\theta \rfloor} W_N\left(t, \frac{\lfloor \lambda \lfloor N\theta \rfloor \rfloor}{N}\right) - \frac{N}{N - \lfloor N\theta \rfloor} \left\{ W_N\left(t, \frac{\lfloor N\theta \rfloor + \lfloor \lambda(N - \lfloor N\theta \rfloor) \rfloor}{N}\right) - W_N\left(t, \frac{\lfloor N\theta \rfloor}{N}\right) \right\},$$

some elementary calculations taking into account that by definition $\hat{\theta} \in [\varepsilon, 1 - \varepsilon]$ show that

$$\hat{\mathbb{V}}_N^{cp} \leq 4 \sup_{\lambda \in [0,1], \theta \in [\varepsilon, 1-\varepsilon]} \int_T \{D_N^{cp}(t, \lambda, \theta)\}^2 dt \lesssim \frac{1}{\varepsilon^2} \sup_{\lambda \in [0,1]} \int_T W_N^2(t, \lambda) dt = o_{\mathbb{P}}(1).$$

Similar but simpler arguments show that $\hat{\mathbb{D}}_N^{cp} = o_{\mathbb{P}}(1)$ and this completes the proof in the case $\int_T \delta^2(t) dt = 0$.

Proof of (6.31). Define the processes

$$S_N^{(1)}(t, \lambda, \theta) = \frac{1}{\lfloor \theta N \rfloor} \sum_{j=1}^{\lfloor \lambda \lfloor \theta N \rfloor \rfloor} (X_j(t) - \mu(t))$$

$$S_N^{(2)}(t, \lambda, \theta) = \frac{1}{N - \lfloor \theta N \rfloor} \sum_{j=\lfloor \theta N \rfloor + 1}^{\lfloor \theta N \rfloor + \lfloor \lambda(N - \lfloor \theta N \rfloor) \rfloor} (X_j(t) - \mu(t) - \delta(t))$$

and similar to the calculations in Section 6.1.1 we can write

$$\mathbb{Z}_N(\lambda, \theta) = \sqrt{N} \int_T \left(D_N^{cp}(t, \lambda, \theta) - \lambda\delta(t)\right)^2 dt + 2\sqrt{N} \int_T \lambda\delta(t)\left(D_N^{cp}(t, \lambda, \theta) - \lambda\delta(t)\right) dt$$

(6.33)
$$= \sqrt{N} \int_T \left(S_N^{(1)}(t, \lambda, \theta) - S_N^{(2)}(t, \lambda, \theta)\right)^2 dt$$

$$+ 2\sqrt{N} \int_T \lambda\delta(t)\left(S_N^{(1)}(t, \lambda, \theta) - S_N^{(2)}(t, \lambda, \theta)\right) dt + o_{\mathbb{P}}(1)$$

uniformly in $\lambda \in [0, 1], \theta \in (\varepsilon, 1 - \varepsilon)$ for any $\varepsilon > 0$. For $\theta = \theta_0$, the first term at the end of the calculation above converges to zero (as in the two sample case) and the second term can be rewritten such that Lemma 6.1 can be applied

(6.34)
$$\mathbb{Z}_N(\lambda, \theta_0) = 2\sqrt{N} \int_T \lambda\delta(t)\left(S_N^{(1)}(t, \lambda, \theta_0) - S_N^{(2)}(t, \lambda, \theta_0)\right) dt + o_{\mathbb{P}}(1)$$
$$= \mathbb{Z}_N^{(1)}(\lambda, \theta_0) - \mathbb{Z}_N^{(2)}(\lambda, \theta_0) + o_{\mathbb{P}}(1),$$

where

(6.35)
$$\mathbb{Z}_N^{(i)}(\lambda, \theta) = 2\sqrt{N} \int_T \lambda\delta(t) S_N^{(i)}(t, \lambda, \theta) dt, \ i = 1, 2.$$

38

We can rewrite the processes in (6.34) as

$$\mathbb{Z}_N^{(1)}(\lambda, \theta_0) - \mathbb{Z}_N^{(2)}(\lambda, \theta_0) = 2\frac{N}{\lfloor\theta_0 N\rfloor}\lambda\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor\lambda\lfloor\theta_0 N\rfloor\rfloor}\int_T \eta_j^{(1)}(t)\delta(t)dt$$

$$- 2\frac{N}{N - \lfloor\theta_0 N\rfloor}\lambda\frac{1}{\sqrt{N}}\sum_{j=\lfloor\theta_0 N\rfloor+1}^{\lfloor\theta_0 N\rfloor+\lfloor\lambda(N-\lfloor\theta_0 N\rfloor)\rfloor}\int_T \eta_j^{(2)}(t)\delta(t)dt$$

$$= \frac{2\lambda}{\theta_0}\tilde{\mathbb{Z}}_N^{(1)}(\lfloor\lambda\lfloor\theta_0 N\rfloor\rfloor/N)$$

$$- \frac{2\lambda}{1-\theta_0}\left\{\tilde{\mathbb{Z}}_N^{(2)}((\lfloor\theta_0 N\rfloor + \lfloor\lambda(N - \lfloor\theta_0 N\rfloor)\rfloor)/N) - \tilde{\mathbb{Z}}_N^{(2)}(\theta_0)\right\} + o_{\mathbb{P}}(1),$$

where the remainder is uniform in $\lambda \in [0, 1]$ and the processes $\tilde{\mathbb{Z}}_N^{(1)}$ and $\tilde{\mathbb{Z}}_N^{(2)}$ are defined in (6.24). An application of Lemma 6.1 with $\zeta = \delta$, asymptotic equicontinuity of the sample paths of $\tilde{\mathbb{Z}}_N^{(i)}$, and the continuous mapping theorem yield the assertion in (6.31). To see that the limit has the right structure, observe that

$$(6.36)\qquad \mathbb{Z}_N^{(1)}(\lambda, \theta_0) - \mathbb{Z}_N^{(2)}(\lambda, \theta_0) \rightsquigarrow \frac{2\lambda}{\theta_0}\tilde{\mathbb{Z}}^{(1)}(\lambda\theta_0) - \frac{2\lambda}{1-\theta_0}\left\{\tilde{\mathbb{Z}}^{(2)}(\theta_0 + \lambda(1-\theta_0)) - \tilde{\mathbb{Z}}^{(2)}(\theta_0)\right\}$$

where

$$\tilde{\mathbb{Z}}^{(1)} = \tilde{\Sigma}_{11}\mathbb{B}_1 + \tilde{\Sigma}_{12}\mathbb{B}_2;\qquad \tilde{\mathbb{Z}}^{(2)} = \tilde{\Sigma}_{21}\mathbb{B}_1 + \tilde{\Sigma}_{22}\mathbb{B}_2,$$

and $\tilde{\Sigma}_{ij}$ denotes the ij-th entry of $\Sigma^{1/2}$. By straightforward calculations one obtains

$$\text{Cov}\big(\tilde{\mathbb{Z}}^{(1)}(\lambda_1\theta_0), \tilde{\mathbb{Z}}^{(2)}(\theta_0 + \lambda_2(1-\theta_0)) - \tilde{\mathbb{Z}}^{(2)}(\theta_0)\big) = 0,$$
$$\text{Cov}(\tilde{\mathbb{Z}}^{(1)}(\lambda_1\theta_0), \tilde{\mathbb{Z}}^{(1)}(\lambda_2\theta_0)) = (\lambda_1 \wedge \lambda_2)\theta_0(\tilde{\Sigma}_{11}^2 + \tilde{\Sigma}_{12}^2).$$

Furthermore, $\tilde{\mathbb{Z}}^{(2)}(\theta_0 + \lambda(1-\theta_0)) - \tilde{\mathbb{Z}}^{(2)}(\theta_0)$ has the same distribution as $\tilde{\mathbb{Z}}^{(2)}(\lambda(1-\theta_0))$ and

$$\text{Cov}(\tilde{\mathbb{Z}}^{(2)}(\lambda_1(1-\theta_0)), \tilde{\mathbb{Z}}^{(2)}(\lambda_2(1-\theta_0))) = (\lambda_1 \wedge \lambda_2)(1-\theta_0)(\tilde{\Sigma}_{21}^2 + \tilde{\Sigma}_{22}^2).$$

Combining the calculations above, we can conclude that the limit process in (6.36) is of the form as claimed in (6.31).

Proof of (6.32). For $\theta = \hat{\theta}$, we show that the first term at the end of the calculation in (6.33) vanishes by proving

$$(6.37)\qquad\qquad N^{1/4}\sup_{\lambda\in[0,1]}\|S_N^{(i)}(\,\cdot\,,\lambda,\hat{\theta})\| = o_{\mathbb{P}}(1)$$

for $i = 1, 2$. Since both cases are similar we only consider the case $i = 1$. Write

$$(6.38)\qquad\qquad N^{1/4}S_N^{(1)}(t, \lambda, \hat{\theta}) = \frac{1}{N^{1/4}}\frac{N}{\lfloor\hat{\theta}N\rfloor}\big(Q_N^{(1)}(t, \lambda, \hat{\theta}) + Q_N^{(2)}(t, \lambda, \hat{\theta})\big)$$

39

where

$$Q_N^{(1)}(t, \lambda, \hat{\theta}) = \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor \lambda \lfloor \hat{\theta} N \rfloor \rfloor \wedge \lfloor \theta_0 N \rfloor} (X_j(t) - \mu(t)) = \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor \lambda \lfloor \hat{\theta} N \rfloor \rfloor \wedge \lfloor \theta_0 N \rfloor} \eta_j^{(1)}(t)$$

(6.39)

$$Q_N^{(2)}(t, \lambda, \hat{\theta}) = \frac{1}{\sqrt{N}} \sum_{j=\lfloor \lambda \lfloor \hat{\theta} N \rfloor \rfloor \wedge \lfloor \theta_0 N \rfloor + 1}^{\lfloor \lambda \lfloor \hat{\theta} N \rfloor \rfloor} (X_j(t) - \mu(t)) .$$

We have by (6.30) and the properties of $\Gamma_N^{(i)}$

$$\lim_{p \to \infty} \limsup_{N \to \infty} \mathbb{P}\left(\sup_{\lambda \in [0,1]} \|Q_N^{(1)}(\;\cdot\;, \lambda, \hat{\theta})\| > p \right)$$

$$\leq \lim_{p \to \infty} \limsup_{N \to \infty} \mathbb{P}\left(\sup_{\lambda \in [0,1]} \sqrt{N} \|\tilde{S}_N^{(1)}(\cdot, \lambda)\| > p \right)$$

$$= \lim_{p \to \infty} \limsup_{N \to \infty} \mathbb{P}\left(\sup_{\lambda \in [0,1]} \int_T \{\Gamma_N^{(1)}(t, \lambda)\}^2 dt > p^2 \right) + o(1) = 0$$

where $\tilde{S}_N^{(1)}$ is defined in (6.29). Therefore $\sup_{\lambda \in [0,1]} \|Q_N^{(1)}(\;\cdot\;, \lambda, \hat{\theta})\| = O_{\mathbb{P}}(1)$. The second term in (6.39) is zero if $\lfloor \lambda \lfloor \hat{\theta} N \rfloor \rfloor \leq \lfloor \theta_0 N \rfloor$ or, if $\lfloor \lambda \lfloor \hat{\theta} N \rfloor \rfloor > \lfloor \theta_0 N \rfloor$, we can write

$$\sup_{\lambda \in [0,1]} \|Q_N^{(2)}(\;\cdot\;, \lambda, \hat{\theta})\| = \sup_{\lambda \in [0,1]} \left\| \frac{1}{\sqrt{N}} \sum_{j=\lfloor \theta_0 N \rfloor + 1}^{\lfloor \lambda \lfloor \hat{\theta} N \rfloor \rfloor} (\eta_j^{(2)} + \delta) \right\| .$$

The number of terms in the sum above is bounded by the distance between θ_0 and $\hat{\theta}$ in the sense that

(6.40)

$$\sup_{\lambda \in [0,1]} \left\| \frac{1}{\sqrt{N}} \sum_{j=\lfloor \theta_0 N \rfloor + 1}^{\lfloor \lambda \lfloor \hat{\theta} N \rfloor \rfloor} (\eta_j^{(2)} + \delta) \right\|$$

$$\lesssim \sup_{\substack{\nu, \lambda \in [0,1]: \\ |\nu - \lambda| \leq |\theta_0 - \hat{\theta}|}} \sqrt{N} \|\tilde{S}_N^{(2)}(\;\cdot\;, \nu) - \tilde{S}_N^{(2)}(\;\cdot\;, \lambda)\| + \frac{\lfloor \hat{\theta} N \rfloor - \lfloor \theta_0 N \rfloor}{\sqrt{N}} \|\delta\|$$

$$= \sup_{\substack{\nu, \lambda \in [0,1]: \\ |\nu - \lambda| \leq |\theta_0 - \hat{\theta}|}} \sqrt{\int_T \{\Gamma_N^{(2)}(t, \nu) - \Gamma_N^{(2)}(t, \lambda)\}^2 dt} + \frac{\lfloor \hat{\theta} N \rfloor - \lfloor \theta_0 N \rfloor}{\sqrt{N}} \|\delta\| + o_{\mathbb{P}}(1) ,$$

where the last equality follows by (6.30) and the definition of $\| \cdot \|$. The first term in (6.40) converges to zero in probability since (6.8) holds with $\Gamma_N^{(2)}$ instead of Γ and since by (3.5) $|\theta_0 - \hat{\theta}| = o_{\mathbb{P}}(1/\sqrt{N})$. The latter also implies that the second term converges to zero in probability. Therefore we have $\sup_{\lambda \in [0,1]} \|Q_N^{(2)}(\;\cdot\;, \lambda, \hat{\theta})\| = o_{\mathbb{P}}(1)$. Recalling (6.38), we conclude that (6.37) holds in the case $i = 1$ and similar arguments yield the statement for $i = 2$. This means that we

40

can continue the calculations in (6.33) for $\theta = \hat{\theta}$ and obtain

$$\sqrt{N}\int_T \big(S_N^{(1)}(t,\lambda,\hat{\theta}) - S_N^{(2)}(t,\lambda,\hat{\theta})\big)^2 dt + 2\sqrt{N}\int_T \lambda\delta(t)\big(S_N^{(1)}(t,\lambda,\hat{\theta}) - S_N^{(2)}(t,\lambda,\hat{\theta})\big)dt$$

$$= 2\sqrt{N}\int_T \lambda D^{cp}(t)\big(S_N^{(1)}(t,\lambda,\hat{\theta}) - S_N^{(2)}(t,\lambda,\hat{\theta})\big)dt + o_{\mathbb{P}}(1)$$

$$= \mathbb{Z}_N^{(1)}(\lambda,\hat{\theta}) - \mathbb{Z}_N^{(2)}(\lambda,\hat{\theta}) + o_{\mathbb{P}}(1),$$

where $\mathbb{Z}_N^{(i)}$, for $i = 1, 2$, is defined by (6.35). In order to prove (6.32) it consequently remains to show

$$\sup_{\lambda\in[0,1]} |\mathbb{Z}_N^{(i)}(\lambda,\hat{\theta}) - \mathbb{Z}_N^{(i)}(\lambda,\theta_0)| = o_{\mathbb{P}}(1)$$

for $i = 1, 2$. For that purpose we write

$$\sup_{\lambda\in[0,1]} |\mathbb{Z}_N^{(1)}(\lambda,\hat{\theta}) - \mathbb{Z}_N^{(1)}(\lambda,\theta_0)|$$

$$= \sup_{\lambda\in[0,1]} \left| 2\sqrt{N}\int_T \lambda\delta(t)\bigg\{ \bigg(\frac{1}{\lfloor\hat{\theta}N\rfloor} - \frac{1}{\lfloor\theta_0 N\rfloor}\bigg) \sum_{j=1}^{\lfloor\lambda\lfloor(\hat{\theta}\wedge\theta_0)N\rfloor\rfloor} \big(X_j(t) - \mu(t)\big) \right.$$

$$\left. + \big(\mathbb{1}\{\hat{\theta}\geq\theta_0\} - \mathbb{1}\{\hat{\theta}<\theta_0\}\big)\frac{1}{\lfloor(\hat{\theta}\vee\theta_0)N\rfloor} \sum_{j=\lfloor\lambda\lfloor(\hat{\theta}\wedge\theta_0)N\rfloor\rfloor+1}^{\lfloor\lambda\lfloor(\hat{\theta}\vee\theta_0)N\rfloor\rfloor} \big(X_j(t) - \mu(t)\big) \bigg\}dt \right|$$

$$\leq 2\|\delta\|\lambda\bigg\{ \frac{N}{\lfloor\theta_0 N\rfloor}\frac{|\lfloor\theta_0 N\rfloor - \lfloor\hat{\theta}N\rfloor|}{\lfloor\hat{\theta}N\rfloor} \sup_{\lambda\in[0,1]} \bigg\|\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor\lambda\lfloor(\hat{\theta}\wedge\theta_0)N\rfloor\rfloor} (X_j - \mu)\bigg\|$$

$$+ \frac{N}{\lfloor(\hat{\theta}\vee\theta_0)N\rfloor} \sup_{\lambda\in[0,1]} \bigg\|\frac{1}{\sqrt{N}}\sum_{j=\lfloor\lambda\lfloor(\hat{\theta}\wedge\theta_0)N\rfloor\rfloor+1}^{\lfloor\lambda\lfloor(\hat{\theta}\vee\theta_0)N\rfloor\rfloor} (X_j - \mathbb{E}[X_j])\bigg\|$$

$$+ \frac{N}{\lfloor(\hat{\theta}\vee\theta_0)N\rfloor} \sup_{\lambda\in[0,1]} \frac{1}{\sqrt{N}} \sum_{j=\lfloor\lambda\lfloor(\hat{\theta}\wedge\theta_0)N\rfloor\rfloor+1}^{\lfloor\lambda\lfloor(\hat{\theta}\vee\theta_0)N\rfloor\rfloor} \|\delta\| \bigg\}$$

$$\lesssim \frac{|\theta_0 - \hat{\theta}|}{\theta_0\hat{\theta}} \sup_{\lambda\in[0,1]} \sqrt{N}\|\tilde{S}_N^{(1)}(\,\cdot\,,\lambda)\|$$

$$+ \frac{1}{\hat{\theta}\vee\theta_0}\bigg(\sum_{i=1}^{2} \sup_{\substack{\nu,\lambda\in[0,1]:\\|\nu-\lambda|\leq|\theta_0-\hat{\theta}|}} \sqrt{N}\big\|\tilde{S}_N^{(i)}(\,\cdot\,,\nu) - \tilde{S}_N^{(i)}(\,\cdot\,,\lambda)\big\| + \frac{|\lfloor\hat{\theta}N\rfloor - \lfloor\theta_0 N\rfloor|}{\sqrt{N}}\|\delta\| \bigg) + o(1)$$

$$= o_{\mathbb{P}}(1).$$

The last equality holds since $|\theta_0 - \hat{\theta}| = o_{\mathbb{P}}(1/\sqrt{N})$, $\sup_{\lambda\in[0,1]} \sqrt{N}\|\tilde{S}_N^{(1)}(\,\cdot\,,\lambda)\| = O_{\mathbb{P}}(1)$ and since

$$\sup_{\substack{\nu,\lambda\in[0,1]:\\|\nu-\lambda|\leq|\theta_0-\hat{\theta}|}} \sqrt{N}\big\|\tilde{S}_N^{(i)}(\,\cdot\,,\nu) - \tilde{S}_N^{(i)}(\,\cdot\,,\lambda)\big\| = \sup_{\substack{\nu,\lambda\in[0,1]:\\|\nu-\lambda|\leq|\theta_0-\hat{\theta}|}} \sqrt{\int_T \{\Gamma_N^{(i)}(t,\nu) - \Gamma_N^{(i)}(t,\lambda)\}^2 dt} + o_{\mathbb{P}}(1)$$

which is $o_{\mathbb{P}}(1)$ by similar arguments as given right after (6.40). Similar arguments prove

$$\sup_{\lambda \in [0,1]} |\mathbb{Z}_N^{(2)}(\lambda, \hat{\theta}) - \mathbb{Z}_N^{(2)}(\lambda, \theta_0)| = o_{\mathbb{P}}(1)$$

which finally implies (6.32). $\square$