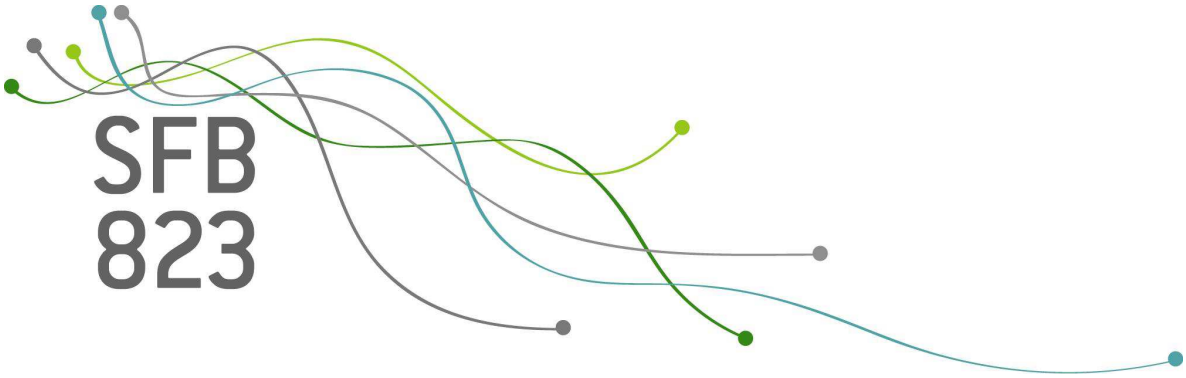
SFB 823

Efficient tests for bio-equivalence in functional data

Holger Dette, Kevin Kokot

Nr. 11/2020

Discussion Paper



Efficient tests for bio-equivalence in functional data

Holger Dette, Kevin Kokot

Ruhr-Universität Bochum

Fakultät für Mathematik

Bochum, Germany

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

April 25, 2020

Abstract

We study the problem of testing the equivalence of functional parameters (such as the mean or variance function) in the two sample functional data problem. In contrast to previous work, which reduces the functional problem to a multiple testing problem for the equivalence of scalar data by comparing the functions at each point, our approach is based on an estimate of a distance measuring the maximum deviation between the two functional parameters. Equivalence is claimed if the estimate for the maximum deviation does not exceed a given threshold. A bootstrap procedure is proposed to obtain quantiles for the distribution of the test statistic and consistency of the corresponding test is proved in the large sample scenario. As the methods proposed here avoid the use of the intersection-union principle they are less conservative and more powerful than the currently available methodology.

Keywords: equivalence tests, functional data, two sample problems, bootstrap, maximum deviation, Banach space valued random variables

1 Introduction

Equivalence tests are nowadays frequently used in drug development to assess similarity of a test and a reference treatment at a controlled type I error. They are very popular in regulatory settings because they reverse the burden of proof compared to a standard test of significance. Therefore they avoid the problem that failing to reject a null hypothesis of no difference is not logically equivalent to deciding for the null hypothesis. Typically equivalence testing is based on a null hypothesis that a scalar parameter of interest, such as the effect difference of two

treatments, is outside an equivalence region defined through an appropriate choice of an interval depending on the metric of equivalence being used. Thus rejecting the null hypothesis means to decide at a controlled type I error that the parameter of interest is in the postulated equivalence region. We refer to the monographs of Wellek (2010) for an overview of the currently available methodology on testing the equivalence of finite dimensional parameters.

On the other hand there are many applications, where the similarity between two populations cannot be appropriately described by a parameter of finite dimension. One obvious situation occurs if treatments involving covariates have to be compared and one is interested in the similarity of the relations between the measured endpoints and the covariates in the two groups. Statistically speaking, this corresponds to the problem establishing the similarity between two regression models and in the last decade considerable efforts have been made to develop methodology to solve this problem. Liu et al. (2009) proposed tests for the hypothesis of equivalence of two linear regression models, while Gsteiger et al. (2011) developed a bootstrap approach using a confidence band for the difference of two non-linear models. These methods are based on the intersection-union principle (see, for example, Berger, 1982) which is used to construct an overall test for equivalence. In a recent paper Dette et al. (2018) showed that equivalence tests based on the intersection-union principle lead to rather conservative decision procedures with low power. As a very powerful alternative they proposed bootstrap tests based on estimates of the maximal deviation between the two curves corresponding to the different treatments. Möllenhoff et al. (2018) demonstrated the superiority of the maximum deviation approach for the comparison of dissolution profiles of two different formulations (see Paixão et al., 2017; Yoshida et al., 2017, for some alternative equivalence tests based on similarity factors). In all these papers, data is finite dimensional and the curves to be compared are defined by parametric regression models with finite dimensional parameters.

Moreover, in the information age, data is often recorded sequentially over time at high resolution and in such instances it is reasonable to model data as functions because the densely sampled observations exhibit certain degrees of dependence and smoothness. As a consequence corresponding parameters such as mean or variance are varying over time and have to be considered as functions as well. The corresponding field in statistics is called functional data analysis and the current state of the art in analyzing functional data is well documented in the monographs by Ramsay and Silverman (2005), Ferraty and Vieu (2010), Horváth and Kokoszka (2012), and Hsing and Eubank (2015). Although numerous statistical concepts such as the comparison of mean functions, covariance operators, principal components, change point analysis have been considered and developed for functional data, the problem of establishing the practical equivalence of two parameters (more precisely parameter functions) for functional data has not found much attention in the literature.

In a recent paper Fogarty and Small (2014) developed methodology for establishing the equivalence between the mean and variance functions from two populations. Their work is motivated by a comparison study of devices for assessing pulmonary function and extends the popular Two One-Sided Testing (TOST) procedure for equivalence testing of scalars (see Schuirmann,

1987; Phillips, 1990, among others) to the functional regime. By the duality between hypotheses testing and confidence intervals their approach is equivalent to the construction of a lower and an upper (pointwise) confidence band for the difference of the two parameters. The test then decides for equivalence if the functions $\kappa_l(\cdot)$ and $\kappa_u(\cdot)$ defining the lower and upper equivalence region for the difference of the two functional parameters are outside of the upper and lower confidence band. Thus their method is similar in spirit to the work of Liu et al. (2009) and Gsteiger et al. (2011) for the comparison of parametric regression models and therefore expected to be rather conservative. A similar comment applies to equivalence tests that can be constructed in the same way using simultaneous confidence bands as developed in Dette et al. (2020) and Liebl and Reimherr (2019).

The purpose of this paper is to develop more efficient procedures to establish equivalence of parameters for the two sample problem in functional data analysis. Our approach is based on an estimate of the maximum deviation between parameter functions (such as the difference of the mean functions or the ratio of the variance functions) and we propose to decide for similarity if the estimated distance is small. In Section 2 we introduce the basic model and review the method of Fogarty and Small (2014). Section 3 is devoted to the construction of a more powerful test for the equivalence of functional parameters, where we concentrate on the mean functions for the sake of brevity. In particular a bootstrap test is developed and its consistency is proved. We also provide a generalization to dependent data and illustrate the superiority of the new test in a small example. In Section 4 we demonstrate the general applicability of our approach and develop methodology for a functional random effect model as considered by Fogarty and Small (2014). We also demonstrate by means of a simulation study that the new tests introduced in this paper are more powerful than the currently available methodology. Finally, all proofs are given in an appendix as they are technically demanding and involve functional data analysis for Banach space valued random variables.

2 Formulation of the problem and state of the art

In this section we state the problem and briefly revisit the approach proposed in Fogarty and Small (2014). To be precise, let $X_{11}(\cdot), \ldots, X_{1m}(\cdot)$ and $X_{21}(\cdot), \ldots, X_{2n}(\cdot)$ denote two independent samples of functional data, which are observed on the interval $[0, 1]$. We denote the mean functions by $\mu_1(\cdot)$ and $\mu_2(\cdot)$ and variance functions by $\sigma_1^2(\cdot)$ and $\sigma_2^2(\cdot)$, respectively (assuming its existence - see Section A for the necessary assumptions). We define $\theta(\cdot) = \mu_1(\cdot) - \mu_2(\cdot)$ and $\lambda(\cdot) = \frac{\sigma_1^2(\cdot)}{\sigma_2^2(\cdot)}$ as measures of similarity between the mean and variance functions, respectively, and consider the hypotheses

$$
\begin{aligned}
&H_0^\theta : \exists\ t \in [0,1] \text{ such that } \theta(t) \notin (\kappa_l(t), \kappa_u(t)) \\
&H_1^\theta : \forall\ t \in [0,1] : \theta(t) \in (\kappa_l(t), \kappa_u(t))
\end{aligned}
\tag{2.1}
$$

and

$$
\begin{aligned}
&H_0^\lambda : \exists\, t \in [0,1] \text{ such that } \lambda(t) \notin (\zeta_l(t), \zeta_u(t)) \\
&H_1^\lambda : \forall\, t \in [0,1] : \lambda(t) \in (\zeta_l(t), \zeta_u(t))\,.
\end{aligned} \tag{2.2}
$$

Here $\kappa_l, \kappa_u, \zeta_l, \zeta_u$ are given functions on the interval $[0,1]$, which define the region of equivalence. These bands have to be developed in cooperation with the experts from the field of application. Usually the band defined by the functions κ_l and κ_u contains the constant function 0 (as one wants to demonstrate the similarity of the functions μ_1 and μ_2) and the band defined by the functions ζ_l and ζ_u contains the constant function 1. Note that the rejection of the null hypothesis in (2.1) means to decide (at a controlled type I error) that the difference of the mean functions is contained in the band defined by the functions κ_l and κ_u and a similar comment applies to the rejection of the null hypothesis in (2.2).

In the following, we concentrate on the mean functions to describe the currently available methodology. Fogarty and Small (2014) combined the intersection-union principle with equivalence testing of scalar parameters to develop tests for the hypotheses (2.1). More precisely, they proposed to test for equivalence in location at each $t \in [0,1]$ and to reject the null hypothesis in (2.1) if all individual tests yield a rejection. For the construction of the individual tests they used a bootstrap version of the Two-One-Sided-Testing (TOST) principle as introduced by Schuirmann (1987). To be precise, if $\hat\theta_{m,n}(t) = \overline{X}_{1\cdot}(t) - \overline{X}_{2\cdot}(t)$ is the common estimate of the mean difference at time $t \in [0,1]$ and

$$
\begin{aligned}
\overline{C}_{1-\alpha,\theta}(t) &= \big[2\hat\theta_{m,n}(t) - q_{1-\alpha}(\hat\theta^*_{m,n}(t)), \infty\big) \\
\underline{C}_{1-\alpha,\theta}(t) &= \big(-\infty, 2\hat\theta_{m,n}(t) - q_{\alpha}(\hat\theta^*_{m,n}(t))\big]
\end{aligned}
$$

are bias corrected percentile-based bootstrap one-sided confidence intervals, then the individual null hypothesis $H^\theta_{0,t} : \theta(t) \notin (\kappa_l(t), \kappa_u(t))$ is rejected in favor of $H^\theta_{1,t} : \theta(t) \in (\kappa_l(t), \kappa_u(t))$ if $\kappa_l(t) \notin \overline{C}_{1-\alpha,\theta}(t)$ **and** $\kappa_u \notin \underline{C}_{1-\alpha,\theta}(t)$, or equivalently

$$
\kappa_l(t) < 2\hat\theta_{m,n}(t) - q_{1-\alpha}(\hat\theta^*_{m,n}(t)) \le 2\hat\theta_{m,n}(t) - q_{\alpha}(\hat\theta^*_{m,n}(t)) < \kappa_u(t)\,. \tag{2.3}
$$

Remark 2.1

(a) It is worthwhile to mention that the concept described here can be used with any type of one-sided confidence intervals. For example, it follows from the proofs of the results in Section 3 (see Section A for more details) that $\sqrt{m+n}(\hat\theta_{m,n}(t) - \theta(t))$ (for fixed t) is asymptotically normal distributed with variance $\sigma^2(t) = \frac{1}{\tau}\sigma_1^2(t) + \frac{1}{1-\tau}\sigma_2^2(t)$, where $\tau = \lim_{m,n\to\infty} m/(m+n)$. Therefore one could use the asymptotic $(1-\alpha)$-confidence intervals

$$
\begin{aligned}
\overline{C}^a_{1-\alpha,\theta}(\theta(t)) &= \Big[\hat\theta_{m,n}(t) - u_{1-\alpha}\frac{\hat\sigma_{m,n}(t)}{\sqrt{m+n}}\,, \infty\Big) \\
\underline{C}^a_{1-\alpha,\theta}(\theta(t)) &= \Big(-\infty, \hat\theta_{m,n}(t) - u_{\alpha}\frac{\hat\sigma_{m,n}(t)}{\sqrt{m+n}}\,\Big]
\end{aligned}
$$

to derive an analogue of the decision rule (2.3), where $\hat{\sigma}^2_{m,n}(t)$ is an appropriate estimate of the asymptotic variance $\sigma^2(t)$ at the point $t \in [0,1]$ and u_α denotes the α-quantile of the standard normal distribution.

(b) Besides the frequentist test described in the previous paragraph Fogarty and Small (2014) also proposed a test within the Bayesian paradigm using Gaussian Processes for modelling the data. Because the focus of this paper is on nonparametric procedures we do not consider this test here.

3 Efficient equivalence-testing of functional parameters

In this section we develop an alternative test for the hypotheses (2.1) in the two sample problem, which turns out to be substantially more powerful than the frequentist method proposed by Fogarty and Small (2014). Our approach is based on the estimation of the maximum deviation of the unknown measure of similarity from the equivalence bounds defined in (2.1) and (2.2). To be precise we restrict ourselves again to the difference of the location parameters $\theta = \mu_1 - \mu_2$ and note that the hypotheses in (2.1) can be rewritten as

$$\begin{aligned} H_0^\theta :\ T^\theta &= \max\Big\{ \sup_{t\in[0,1]} \big(-\theta(t) + \kappa_l(t)\big),\ \sup_{t\in[0,1]} \big(\theta(t) - \kappa_u(t)\big)\Big\} \geq 0 \\ H_1^\theta :\ T^\theta &= \max\Big\{ \sup_{t\in[0,1]} \big(-\theta(t) + \kappa_l(t)\big),\ \sup_{t\in[0,1]} \big(\theta(t) - \kappa_u(t)\big)\Big\} < 0\,. \end{aligned} \tag{3.1}$$

The representation of the hypotheses simplifies in the case of symmetric and constant boundaries, that is $\kappa_u(t) = -\kappa_l(t) = \kappa > 0$ for all $t \in [0,1]$, where we obtain for the hypotheses in (3.1)

$$H_0^\theta : \sup_{t\in[0,1]} \big|\theta(t)\big| \geq \kappa\,, \quad H_1^\theta : \sup_{t\in[0,1]} \big|\theta(t)\big| < \kappa\,.$$

For the construction of an efficient test, we define the statistic

$$\hat{T}^\theta_{m,n} = \max\Big\{ \sup_{t\in[0,1]} \big(-\hat{\theta}_{m,n}(t) + \kappa_l(t)\big),\ \sup_{t\in[0,1]} \big(\hat{\theta}_{m,n}(t) - \kappa_u(t)\big)\Big\} \tag{3.2}$$

as an estimator of T^θ, where

$$\hat{\theta}_{m,n} = \overline{X}_{1\cdot} - \overline{X}_{2\cdot}$$

denotes the difference of the sample means, which serves as an estimator of the function $\theta = \mu_1 - \mu_2$. The null hypothesis in (3.1) is then rejected for small values of $\hat{T}^\theta_{m,n}$, where the critical values will be determined by bootstrap (in the independent case by resampling with replacement, in the dependent case by multiplier block bootstrap).

To be precise and motivate our bootstrap assume that $m, n \to \infty$, such that $m/(m+n) \to \tau \in (0,1)$. Then if follows from Theorem A.1 in the online supplement that

$$\sqrt{m+n}\big(\hat{T}^\theta_{m,n} - T^\theta\big) \xrightarrow{\mathcal{D}} Z_{\mathcal{E},\theta} = \max\Big\{ \sup_{t\in\mathcal{E}^l_\theta} \big(-Z(t)\big),\ \sup_{t\in\mathcal{E}^u_\theta} Z(t)\Big\}\,, \tag{3.3}$$

where Z is a Gaussian process with covariance kernel

$$k(s,t) = \frac{1}{\tau}\mathrm{Cov}(X_{11}(s), X_{11}(t)) + \frac{1}{1-\tau}\mathrm{Cov}(X_{21}(s), X_{21}(t)) \tag{3.4}$$

and the sets $\mathcal{E}_\theta^l, \mathcal{E}_\theta^u \subset [0,1]$ contain the points, where the functions $-\theta + \kappa_l$ and $\theta - \kappa_u$ attain the value T^θ, i.e.

$$\mathcal{E}_\theta^l = \left\{t \in [0,1]\colon\ -\theta(t) + \kappa_l(t) = T^\theta\right\}, \quad \mathcal{E}_\theta^u = \left\{t \in [0,1]\colon \theta(t) - \kappa_u(t) = T^\theta\right\}. \tag{3.5}$$

Throughout this paper, these sets are called *extremal sets* and we note that the extremal sets can be empty (but not both at the same time). As a consequence, the limit distribution on the right hand side of (3.3) depends on the covariance kernel k and the extremal sets $\mathcal{E}_\theta^l$ and $\mathcal{E}_\theta^u$ defined by the unknown difference θ between the mean functions μ_1 and μ_2.

For the calculation of quantiles of the distribution of $Z_{\mathcal{E},\theta}$ we propose to use the bootstrap and proceed in two steps:

(1) We estimate the unknown sets of extremal points.

(2) We use the bootstrap to mimic the distribution of the process Z in (3.3).

For the estimation of the extremal sets $\mathcal{E}_\theta^l$ and $\mathcal{E}_\theta^u$, we use the statistics

$$\begin{aligned}
\hat{\mathcal{E}}_\theta^l &= \left\{t \in [0,1]\colon\ -\hat\theta_{m,n}(t) + \kappa_l(t) \geq \hat T_{m,n}^\theta - c\,\frac{\log(m+n)}{\sqrt{m+n}}\right\},\\
\hat{\mathcal{E}}_\theta^u &= \left\{t \in [0,1]\colon \hat\theta_{m,n}(t) - \kappa_u(t) \geq \hat T_{m,n}^\theta - c\,\frac{\log(m+n)}{\sqrt{m+n}}\right\},
\end{aligned} \tag{3.6}$$

where the statistic $\hat T_{m,n}^\theta$ is defined in (3.2) and c is a tuning parameter. For the bootstrap part note that it follows from the arguments given in the proof of Theorem A.1 in the appendix that the statistic on the left hand side of (3.3) is asymptotically equivalent to the statistic

$$\max\left\{\sup_{t\in\mathcal{E}_\theta^l}\big(-\hat Z_{m,n}(t)\big), \sup_{t\in\mathcal{E}_\theta^u}\hat Z_{m,n}(t)\right\}$$

where the process $\hat Z_{m,n}$ is defined by

$$\hat Z_{m,n} = \sqrt{m+n}\left\{\hat\theta_{m,n} - \theta\right\} = \sqrt{m+n}\left\{\overline{X}_{1\cdot} - \mu_1 - (\overline{X}_{2\cdot} - \mu_2)\right\}$$

(by the arguments given in the proof of Theorem A.1 this process converges weakly to the process Z on the right hand side of (3.3)). To mimic the distribution of this process in the independent case we now use resampling with replacement. More precisely, assume for $r = 1, \ldots, R$ that $X_{11}^{*(r)}, \ldots, X_{1m}^{*(r)}$ and $X_{21}^{*(r)}, \ldots, X_{2n}^{*(r)}$ are drawn randomly with replacement from $X_{11}, \ldots, X_{1m}$ and $X_{21}, \ldots, X_{2n}$, respectively, and denote by $\overline{X}_{1\cdot}$ and $\overline{X}_{2\cdot}$ the sample means of both groups. We define

$$\hat Z_{m,n}^{*(r)} = \sqrt{m+n}\left\{\frac{1}{m}\sum_{j=1}^m (X_{1j}^{*(r)} - \overline{X}_{1\cdot}) - \frac{1}{n}\sum_{j=1}^n (X_{2j}^{*(r)} - \overline{X}_{2\cdot})\right\} \tag{3.7}$$

as the r-th bootstrap analogue of the statistic $\hat{Z}_{m,n}$ and a bootstrap version of the random variable on the left hand side of (3.3) by

$$\hat{T}^{\theta,*(r)}_{m,n} = \max\Big\{ \sup_{t\in\hat{\mathcal{E}}^l_\theta} \big(- \hat{Z}^{*(r)}_{m,n}(t)\big),\ \sup_{t\in\hat{\mathcal{E}}^u_\theta} \hat{Z}^{*(r)}_{m,n}(t)\Big\}\,. \tag{3.8}$$

Finally, the null hypothesis in (3.1) is rejected, whenever

$$\sqrt{m+n}\,\hat{T}^\theta_{m,n} < z^{*(R)}_{m,n,\alpha}\,, \tag{3.9}$$

where $z^{*(R)}_{m,n,\alpha}$ is the empirical α-quantile of the bootstrap sample $T^{\theta,*(1)}_{m,n},\dots,T^{\theta,*(R)}_{m,n}$. The following result, which is proved in the appendix, shows that this procedure defines a consistent and asymptotic level α-test for the hypotheses (3.1) (or equivalently for the hypotheses (2.1)).

Theorem 3.1 *Let Assumption A.1 in Section A.2 be satisfied.*

(a) Assume that the null hypothesis H^θ_0 of no equivalence in (2.1) holds, that is $T^\theta \geq 0$. If $T^\theta = 0$, then

$$\lim_{m,n,R\to\infty} \mathbb{P}\big(\sqrt{m+n}\,\hat{T}^\theta_{m,n} < z^{*(R)}_{m,n,\alpha}\big) = \alpha\,.$$

If $T^\theta > 0$, then for any $R\in\mathbb{N}$

$$\lim_{m,n\to\infty} \mathbb{P}\big(\sqrt{m+n}\,\hat{T}^\theta_{m,n} < z^{*(R)}_{m,n,\alpha}\big) = 0\,.$$

(b) If the alternative H^θ_1 of equivalence in (2.1) holds, that is $T^\theta < 0$, we have for any $R\in\mathbb{N}$

$$\liminf_{m,n\to\infty} \mathbb{P}\big(\sqrt{m+n}\,\hat{T}^\theta_{m,n} < z^{*(R)}_{m,n,\alpha}\big) = 1\,.$$

Remark 3.1 The results remain correct in the case of dependent data, where $X_{1,1},\dots X_{1,m}$ and $X_{2,1},\dots X_{2,n}$ are two independent stationary time series. In this case, we propose to use a block multiplier bootstrap to mimic the dependency in the data. To be precise, define a bootstrap process by

$$\begin{aligned}\hat{Z}^{**(r)}_{m,n}(t) =& \sqrt{m+n}\Big\{\frac{1}{m}\sum_{k=1}^{m-l_1+1}\frac{1}{\sqrt{l_1}}\Big(\sum_{j=k}^{k+l_1-1} X_{1j}(t) - \frac{l_1}{m}\sum_{j=1}^{m} X_{1j}(t)\Big)\xi^{(r)}_k \\ &- \frac{1}{n}\sum_{k=1}^{n-l_2+1}\frac{1}{\sqrt{l_2}}\Big(\sum_{j=k}^{k+l_2-1} X_{2j}(t) - \frac{l_2}{n}\sum_{j=1}^{n} X_{2j}(t)\Big)\zeta^{(r)}_k\Big\} \quad (r=1,\dots,R)\end{aligned} \tag{3.10}$$

where $\xi^{(r)}_1,\dots,\xi^{(r)}_m$, $\zeta^{(r)}_1,\dots,\zeta^{(r)}_n$ are independent standard normal distributed random variables and l_1, l_2 are sequences converging to infinity with increasing sample sizes $m,n\to\infty$. The null hypothesis in (2.1) is now rejected, whenever

$$\sqrt{m+n}\,\hat{T}^\theta_{m,n} < z^{**(R)}_{m,n,\alpha}\,, \tag{3.11}$$

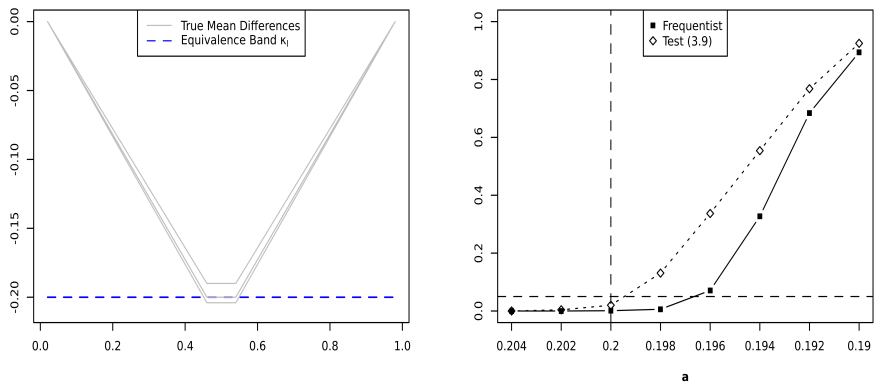


Figure 1: *Left panel: Difference* $\theta = \mu_1 - \mu_2$ *of the mean functions defined by* (3.14) *with fixed* $b_1 = 0.46, b_2 = 0.54$ *and different values for* $a \in \{0.19, 0.204, 0.2\}$. *Right panel: Empirical rejection probabilities of the frequentist test proposed by Fogarty and Small (2014) and the test* (3.9) *for the hypotheses* (3.1) *with* $\kappa_l \equiv -0.2$, $\kappa_u \equiv 0.2$ *and different values of* a.

where $z^{**(R)}_{m,n,\alpha}$ is the empirical α-quantile of the sample $T^{\theta,**(1)}_{m,n}, \ldots, T^{\theta,**(R)}_{m,n}$ and the statistic $\hat{T}^{\theta,**(r)}_{m,n}$ is defined by

$$\hat{T}^{\theta,**(r)}_{m,n} = \max\Big\{ \sup_{t\in\hat{\mathcal{E}}^l_\theta} \big(- \hat{Z}^{**(r)}_{m,n}(t)\big), \sup_{t\in\hat{\mathcal{E}}^u_\theta} \hat{Z}^{**(r)}_{m,n}(t)\Big\} \qquad (r = 1, \ldots, R)\,.$$

In this case - under the assumptions stated in Section A.2.3 - the result in Theorem 3.1 remains valid. Finally we note that this procedure with $l_1 = l_2 = 1$ provides also a valid bootstrap test in the case of independent data.

Example 3.1 We have conducted a small simulation study to compare the new bootstrap test (3.9) with the frequentist test proposed by Fogarty and Small (2014). Further numerical results supporting our findings can be found in Section 4.3, where we compare both methods in a functional random effect model.

We have generated functional data as described in Sections 6.3 and 6.4 of Aue et al. (2015), who considered $D = 21$ B-spline basis functions $\nu_1, \ldots, \nu_D$ and defined the random functions $\eta_{11}, \ldots, \eta_{1m}, \eta_{21}, \ldots, \eta_{2n}$ by

$$\eta_{1j} = \sum_{i=1}^{D} N_{1,i,j}\nu_i\,, \quad \eta_{2k} = \sum_{i=1}^{D} N_{2,i,k}\nu_i\,, \qquad j = 1, \ldots, m,\ k = 1, \ldots, n\,, \tag{3.12}$$

where $N_{1,1,1}, \ldots, N_{1,D,m}, N_{2,1,1}, \ldots, N_{2,D,n}$ are independent, normally distributed random variables with expectation zero and variances $\sigma_i^2 = \mathrm{Var}(N_{1,i,j}) = \mathrm{Var}(N_{2,i,k}) = 1/i^2$ $(i = 1, \ldots, D;$ $j = 1, \ldots, m;\ k = 1, \ldots, n)$.

Then two independent samples of independent and identically distributed Gaussian random functions are obtained by

$$X_{1j} = \eta_{1,j} + \mu_1\,, \quad X_{2k} = \eta_{2,k} + \mu_2\,, \tag{3.13}$$

with mean functions

$$\mu_1 \equiv 0\,, \quad \mu_2(t) = \begin{cases} \frac{a}{b_1 - 0.02}(t - 0.02)\,, & t \in [0, b_1) \\ a\,, & t \in [b_1, b_2] \\ \frac{-a}{0.98 - b_2}(t - b_2) + a\,, & t \in (b_2, 1] \end{cases}\,, \tag{3.14}$$

where a, b_1 and b_2 are parameters. The left part of Figure 1 illustrates the difference $\theta = \mu_1 - \mu_2$ of the mean functions for fixed $b_1 = 0.46$, $b_2 = 0.54$ and different values of the parameter a. The equivalence bands, used in the hypotheses (2.1), are defined by $\kappa_l \equiv -0.2$, $\kappa_u \equiv 0.2$. Note that, for any $a > 0.02$, the extremal sets in (3.5) are defined by $\mathcal{E}_\theta^u = \emptyset$, $\mathcal{E}_\theta^l = [b_1, b_2]$ (here $\mathcal{E}_\theta^l = [0.46, 0.54]$) and that the cases $|a| \geq 0.2$ and $|a| < 0.2$ correspond to the null hypothesis of no equivalence and the alternative of equivalent mean functions, respectively. In the right part of Figure 1 we display the empirical rejection probabilities of the frequentist test proposed by Fogarty and Small (2014) and the test defined in (3.9) for different values of $a \in \{0.204, 0.202, \ldots, 0.190\}$ (by symmetry negative values of a yield the same results). Here the extremal sets are estimated by (3.6) with $c = 0.005$.

The sample sizes are $m = n = 100$ and the rejection probabilities are calculated by 1000 simulation runs and 300 bootstrap replications. We observe that the rejection probabilities are strictly smaller than the level 5% for $a > 0.2$ and increase towards 1 for decreasing a beyond 0.2. Both tests slightly underestimate the nominal level at the boundary of the null hypothesis ($a = 0.2$). Moreover, the new test has substantially more power in all considered scenarios under the alternative.

The superiority of the new test is even more visible if the size of the set of extremal points is larger. To illustrate this fact, we consider the mean functions in (3.14) for fixed $a = 0.194$ and different values of b_1 and b_2. The rejection probabilities of the frequentist test proposed by Fogarty and Small (2014) and the test defined in (3.9) are shown in Figure 2 where, for $j = 0, \ldots, 4$, function number j corresponds to the choices $b_1 = 0.5 - 0.08j$, $b_2 = 0.5 + 0.08j$ in the definition of the mean differences in (3.14). The sample sizes are again $m = n = 100$. We observe that only in the case $j = 0$, both tests have comparable power. In all other cases, the new test (3.9) outperforms the test proposed by Fogarty and Small (2014) substantially.

4 A functional random effect model for paired data

In this section we demonstrate that the method introduced in Section 3 for the simple two sample problem of comparing two mean functions is a universally applicable decision rule to decide for the equivalence between two functional parameters from two samples of functional data. For this purpose only the bootstrap procedure has to be adjusted to the situation under consideration.

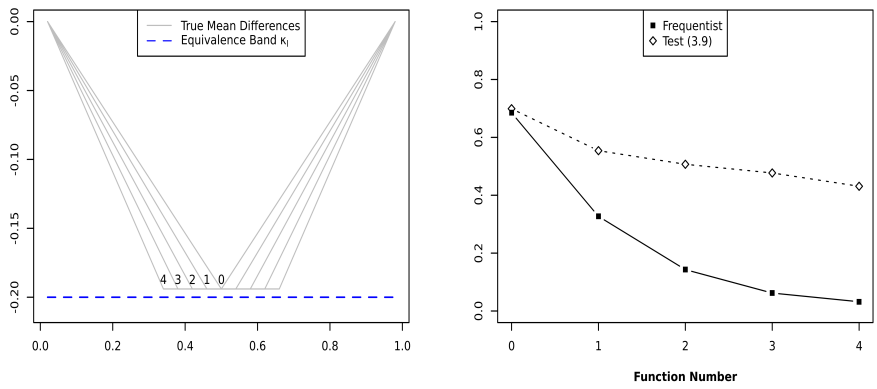


Figure 2: *Left panel: Difference* $\theta = \mu_1 - \mu_2$ *of the mean functions defined by* (3.14) *for fixed* $a = 0.194$ *and different choices of* $b_1 = 0.5 - 0.08j$, $b_2 = 0.5 + 0.08j$, *where* $j = 0, \ldots, 4$. *Right panel: Empirical rejection probabilities of the frequentist test proposed by Fogarty and Small (2014) and the test* (3.9) *for the hypotheses* (3.1) *with* $\kappa_l \equiv -0.2$, $\kappa_u \equiv 0.2$.

As a concrete example (in particular for the sake of comparison with the currently available methodology) we consider a functional analysis of variance model with random effects as proposed by Fogarty and Small (2014) for the analysis of functional data describing the lung volume over time for different patients and different breaths produced by a spirometer (industry standard) and a new device (Structured Light Plethysmography - SLP). While the new SLP holds many advantages, it has to be assured that it produces measurements (practically) equivalent to those produced by the industry standard, before it can be used for diagnoses purposes (see Fogarty and Small, 2014, for more details). There are A patients and for the i-th patient, n_i breaths are recorded simultaneously by both devices leading to paired functional data with cross-covariances between the pairs. The goal is the development of a statistically justified decision rule to decide for or against equivalence of the measurements.

To be precise, we consider pairs of random functions defined by

$$\begin{pmatrix} X_{1,i,j} \\ X_{2,i,j} \end{pmatrix} = \begin{pmatrix} \mu_1 + \varepsilon_{1,i} + \eta_{1,i,j} \\ \mu_2 + \varepsilon_{2,i} + \eta_{2,i,j} \end{pmatrix} \quad j = 1, \ldots, n_i\,,\ i = 1, \ldots, A\,. \tag{4.1}$$

Here μ_1, μ_2 denote the mean functions, the functions $\varepsilon_{1,i}, \varepsilon_{2,i}$ model a random group effect (usually corresponding to different individuals drawn from a larger population) and the functions $\eta_{1,i,j}, \eta_{2,i,j}$ are individual random effects. The random group effects and the individual random effect functions are assumed to be centred and independent and identically distributed, respectively. Furthermore the group effects are independent of the individual ones. Note that the total number of pairs is given by $N = \sum_{i=1}^{A} n_i$.

4.1 Comparing mean functions

For the construction of a test for the hypotheses (3.1), we consider the statistic

$$\hat{T}_N^\theta = \sqrt{A}\,\max\Big\{\sup_{t\in[0,1]}\big(-\hat{\theta}_N(t)+\kappa_l(t)\big),\ \sup_{t\in[0,1]}\big(\hat{\theta}_N(t)-\kappa_u(t)\big)\Big\}\,, \tag{4.2}$$

where $\hat{\theta}_N = \overline{X}_{1\cdot\cdot} - \overline{X}_{2\cdot\cdot}$ and

$$\overline{X}_{\ell\cdot\cdot} = \frac{1}{A}\sum_{i=1}^{A}\frac{1}{n_i}\sum_{j=1}^{n_i} X_{\ell,i,j}\,,\quad \ell=1,2$$

denote the two sample means. The bootstrap analogue of (4.2) is defined as follows. We use the sample means

$$\overline{X}_{\ell,i,\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{\ell,i,j}\,,\quad \ell=1,2\,,\ i=1,\dots,A \tag{4.3}$$

in the different groups to estimate the group effects by

$$\hat{\varepsilon}_{1,i} = \overline{X}_{1,i,\cdot} - \overline{X}_{1,\cdot\cdot}\,,\quad \hat{\varepsilon}_{2,i} = \overline{X}_{2,i,\cdot} - \overline{X}_{2,\cdot\cdot}\,,\quad (i=1,\dots,A).$$

For the bootstrap we draw, for $r=1,\dots,R$, samples $(\hat{\varepsilon}_{1,1}^{\star(r)},\hat{\varepsilon}_{2,1}^{\star(r)}),\dots,(\hat{\varepsilon}_{1,A}^{\star(r)},\hat{\varepsilon}_{2,A}^{\star(r)})$ randomly with replacement from the pairs $(\hat{\varepsilon}_{1,1},\hat{\varepsilon}_{2,1}),\dots,(\hat{\varepsilon}_{1,A},\hat{\varepsilon}_{2,A})$. The bootstrap statistic is then defined by

$$\hat{T}_N^{\theta,\star(r)} = \max\Big\{\sup_{t\in\hat{\mathcal{E}}_\theta^l}\big(-B_N^{\star(r)}(t)\big),\ \sup_{t\in\hat{\mathcal{E}}_\theta^u} B_N^{\star(r)}(t)\Big\}\,, \tag{4.4}$$

where

$$B_N^{\star(r)} = \frac{1}{\sqrt{A}}\sum_{i=1}^{A}\big(\hat{\varepsilon}_{1,i}^{\star(r)} - \hat{\varepsilon}_{2,i}^{\star(r)}\big)\,, \tag{4.5}$$

and the sets $\hat{\mathcal{E}}^l, \hat{\mathcal{E}}^u$ are given by

$$\begin{aligned}\hat{\mathcal{E}}_\theta^l &= \Big\{t\in[0,1]\colon\ -\hat{\theta}_N(t)+\kappa_l(t) \geq \hat{T}_N^\theta - c\,\frac{\log(A)}{\sqrt{A}}\ \Big\}\\ \hat{\mathcal{E}}_\theta^u &= \Big\{t\in[0,1]\colon \hat{\theta}_N(t)-\kappa_u(t) \geq \hat{T}_N^\theta - c\,\frac{\log(A)}{\sqrt{A}}\ \Big\}\,.\end{aligned} \tag{4.6}$$

The consideration of the process $B_N^{\star(r)}$ in (4.5) is motivated by the expansion $\sqrt{A}\,(\hat{\theta}_N-\theta) = \frac{1}{\sqrt{A}}\sum_{i=1}^{A}(\varepsilon_{1,i}-\varepsilon_{2,i}) + o_{\mathbb{P}}(1)$, which is derived in equation (5.5) in the online supplement. The null hypothesis in (3.1) is finally rejected whenever

$$\hat{T}_N^\theta < z_{N,\alpha}^{\star(R)} \tag{4.7}$$

where $z_{N,\alpha}^{\star(R)}$ is the empirical α-quantile of the bootstrap sample $\hat{T}_N^{\theta,\star(1)},\dots,\hat{T}_N^{\theta,\star(R)}$. The following result shows that this decision rule defines a consistent asymptotic level α test for the hypotheses in (2.1).

Theorem 4.1 *Let Assumption A.2 in Section A.3.1 be satisfied and assume that* $A \to \infty$ *and* $\min_i^A n_i \to \infty$ *as* $N \to \infty$.

(a) Assume that the null hypothesis H_0^θ *of no equivalence in* (2.1) *holds, that is* $T^\theta \geq 0$. *If* $T^\theta = 0$, *then*

$$\lim_{A,\min n_i,R\to\infty} \mathbb{P}\big(\hat{T}_N^\theta < z_{N,\alpha}^{\star(R)}\big) = \alpha\,.$$

If $T^\theta > 0$, *then for any* $R \in \mathbb{N}$

$$\lim_{A,\min n_i\to\infty} \mathbb{P}\big(\hat{T}_N^\theta < z_{N,\alpha}^{\star(R)}\big) = 0\,.$$

(b) If the alternative H_1^θ *of equivalence in* (2.1) *holds, that is* $T^\theta < 0$, *we have for any* $R \in \mathbb{N}$

$$\liminf_{A,\min n_i\to\infty} \mathbb{P}\big(\hat{T}_N^\theta < z_{N,\alpha}^{\star(R)}\big) = 1\,.$$

4.2 Comparing variance functions

Recall the definition of model (4.1) and define (assuming its existence - see Section A.1 for more details)

$$\sigma_1^2(\cdot) = \mathbb{E}\big[\eta_{1,1,1}(\cdot)^2\big]\,, \quad \sigma_2^2(\cdot) = \mathbb{E}\big[\eta_{2,1,1}(\cdot)^2\big] \in C([0,1])$$

as the variance functions of the individual errors $\eta_{1,1,1}, \eta_{2,1,1}$. We are interested in testing the hypotheses (2.2), which can be rewritten as

$$\begin{aligned} &H_0^\lambda:\ T^\lambda = \max\Big\{\sup_{t\in[0,1]}\big(-\log\lambda(t) + \log\zeta_l(t)\big), \sup_{t\in[0,1]}\big(\log\lambda(t) - \log\zeta_u(t)\big)\Big\} \geq 0 \\ &H_1^\lambda:\ T^\lambda < 0\,, \end{aligned} \tag{4.8}$$

where $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$ is the ratio of the two variance functions and $\zeta_l(t), \zeta_u(t)$ are the given equivalence bands. Note that we work with the logarithm of λ to obtain stabilized variances. We define

$$\hat{\sigma}_\ell^2 = \frac{1}{N-A}\sum_{i=1}^{A}\sum_{j=1}^{n_i}\Big(X_{\ell,i,j} - \frac{1}{n_i}\sum_{k=1}^{n_i} X_{\ell,i,k}\Big)^2, \quad \ell = 1,2, \tag{4.9}$$

estimate the variance ratio by $\hat{\lambda} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$ and consider the test statistic

$$\hat{T}_N^\lambda = \sqrt{N}\,\max\Big\{\sup_{t\in[0,1]}\big(-\log\hat{\lambda}(t) + \log\zeta_l(t)\big), \sup_{t\in[0,1]}\big(\log\hat{\lambda}(t) - \log\zeta_u(t)\big)\Big\}\,. \tag{4.10}$$

For the calculation of bootstrap quantiles we adapt resampling with replacement to the random effect model (4.1) and estimate the individual random effects by

$$\hat{\eta}_{1,i,j} = X_{1,i,j} - \overline{X}_{1,i,\cdot}\,, \quad \hat{\eta}_{2,i,j} = X_{2,i,j} - \overline{X}_{2,i,\cdot}$$

for $i = 1, \dots, A$ and $j = 1, \dots, n_i$, where the group means $\overline{X}_{\ell,i,\cdot}$ $(i = 1, \dots, A)$ are defined by (4.3). We now draw with replacement $N = \sum_{i=1}^{A} n_i$ pairs $(\hat{\eta}^{\star(r)}_{1,1,1}, \hat{\eta}^{\star(r)}_{2,1,1}), \dots, (\hat{\eta}^{\star(r)}_{1,A,n_A}, \hat{\eta}^{\star(r)}_{2,A,n_A})$ from $(\hat{\eta}_{1,1,1}, \hat{\eta}_{2,1,1}), \dots, (\hat{\eta}_{1,A,n_A}, \hat{\eta}_{2,A,n_A})$ and define for $r = 1, \dots, R$

$$\hat{T}^{\lambda,\star(r)}_N = \sqrt{N} \max \Big\{ \sup_{t \in \hat{\mathcal{E}}^l_\lambda} \big(- C^{\star(r)}_N(t) \big), \sup_{t \in \hat{\mathcal{E}}^u_\lambda} C^{\star(r)}_N(t) \Big\} \tag{4.11}$$

as the bootstrap analogue of (4.10), where

$$C^{\star(r)}_N = \frac{C^{\star(r)}_{1,N}}{\hat{\sigma}^2_1} - \frac{C^{\star(r)}_{2,N}}{\hat{\sigma}^2_2}, \quad C^{\star(r)}_{\ell,N} = \frac{1}{N-A} \sum_{i=1}^{A} \sum_{j=1}^{n_i} \big\{ (\hat{\eta}^{\star(r)}_{\ell,i,j})^2 - \hat{\sigma}^2_\ell \big\}, \quad \ell = 1, 2 \tag{4.12}$$

and

$$\begin{aligned} \hat{\mathcal{E}}^l_\lambda &= \Big\{ t \in [0,1] \colon \ - \log \hat{\lambda}(t) + \log \zeta_l(t) \geq \hat{T}^\lambda_N - c \, \frac{\log(N)}{\sqrt{N}} \ \Big\} \\ \hat{\mathcal{E}}^u_\lambda &= \Big\{ t \in [0,1] \colon \ \log \hat{\lambda}(t) - \log \zeta_u(t) \geq \hat{T}^\lambda_N - c \, \frac{\log(N)}{\sqrt{N}} \ \Big\}. \end{aligned} \tag{4.13}$$

The consideration of the process $C^{\star(r)}_N$ in (4.12) is motivated by the expansion

$$\sqrt{N}(\log \hat{\lambda} - \log \lambda) = \frac{\sqrt{N}}{N-A} \sum_{i=1}^{A} \sum_{j=1}^{n_i} \Big\{ \frac{(\eta_{1,i,j})^2 - \sigma^2_1}{\sigma^2_1} - \frac{(\eta_{2,i,j})^2 - \sigma^2_2}{\sigma^2_2} \Big\} + o_{\mathbb{P}}(1),$$

which is derived in equation (5.10) in the online supplement.

Finally, the null hypothesis in (2.2) of no equivalence is rejected, whenever

$$\hat{T}^\lambda_N < u^{\star(R)}_{N,\alpha} \tag{4.14}$$

where $u^{\star(R)}_{N,\alpha}$ is the empirical α-quantile of the sample $\hat{T}^{\lambda,\star(1)}_N, \dots, \hat{T}^{\lambda,\star(R)}_N$.

Theorem 4.2 *Let Assumption A.3 in Section A.3.2 be satisfied and assume that $A \to \infty$ and $\min_i^A n_i \to \infty$ as $N \to \infty$.*

(a) Assume that the null hypothesis H^λ_0 of no equivalence in (4.8) holds, that is $T^\lambda \geq 0$. If $T^\lambda = 0$, then

$$\lim_{A, \min n_i, R \to \infty} \mathbb{P}\big(\hat{T}^\lambda_N < u^{\star(R)}_{N,\alpha} \big) = \alpha \,.$$

If $T^\lambda > 0$, then for any $R \in \mathbb{N}$

$$\lim_{A, \min n_i \to \infty} \mathbb{P}\big(\hat{T}^\lambda_N < u^{\star(R)}_{N,\alpha} \big) = 0 \,.$$

(b) If the alternative H^λ_1 of equivalence in (4.8) holds, that is $T^\lambda < 0$, we have for any $R \in \mathbb{N}$

$$\liminf_{A, \min n_i \to \infty} \mathbb{P}\big(\hat{T}^\lambda_N < u^{\star(R)}_{N,\alpha} \big) = 1 \,.$$

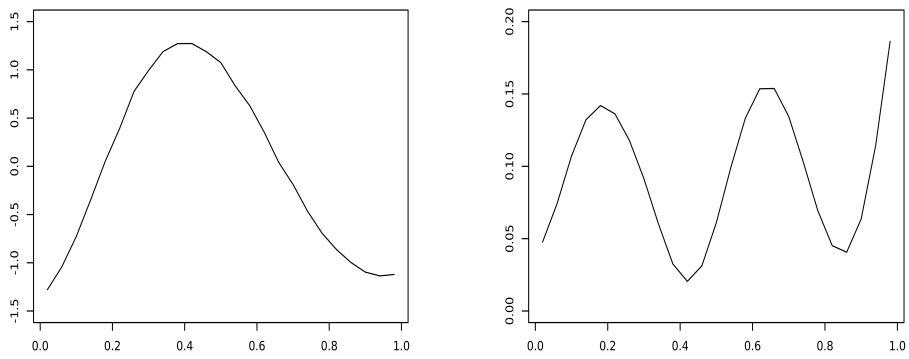


Figure 3: *Expectation function μ_1 (left panel) and variance function σ_1 (right panel) used in the simulation study in Section 4.3.1 and Section 4.3.2.*

4.3 Some numerical results

In this section we illustrate the finite sample properties of the new bootstrap procedure in the functional analysis of variance model (4.1) and also provide a comparison with the method proposed in Fogarty and Small (2014). For this purpose we consider some of the scenarios described in Sections 10.1 and 10.2 of this reference. As a general picture we will demonstrate that the procedure proposed in this paper is more powerful than the frequentist test method developed in Fogarty and Small (2014). Note that Fogarty and Small (2014) also develop a Bayesian test method but it is outperformed by the frequentist test. Therefore, the new bootstrap test is only compared with the frequentist test in the following sections.

For the sake of comparison, we perform the frequentist test of Fogarty and Small (2014) with the same data as the new bootstrap procedure and do not use the exact results displayed in this reference. In each scenario under consideration, we perform 1000 simulation runs and in each run, $R = 300$ bootstrap replicates are generated to calculate the empirical 5% bootstrap quantile. The extremal sets are estimated as in (4.6) and (4.13) with $c = 0.005$, respectively.

4.3.1 Comparison of mean functions

For the mean functions, we consider five different scenarios. The mean function μ_1 is the same in each scenario and can be obtained from the software code provided by Fogarty and Small (2014). It is not defined explicitly and displayed in the left panel of Figure 3. The mean function μ_2 is defined by

$$\mu_2(t) = \mu_1(t) + 0.2 \exp\big(- a_i \, |t - 1/2|\big)$$

(thus the difference has a parametric form), where $a_1 = 0$ and $a_i = 10^{2(i-2)/7}$ for $i = 3, 5, 7, 9$. The differences $\mu_1 - \mu_2$ correspond to the functions 1, 3, 5, 7 and 9 in the left part of Figure 4,

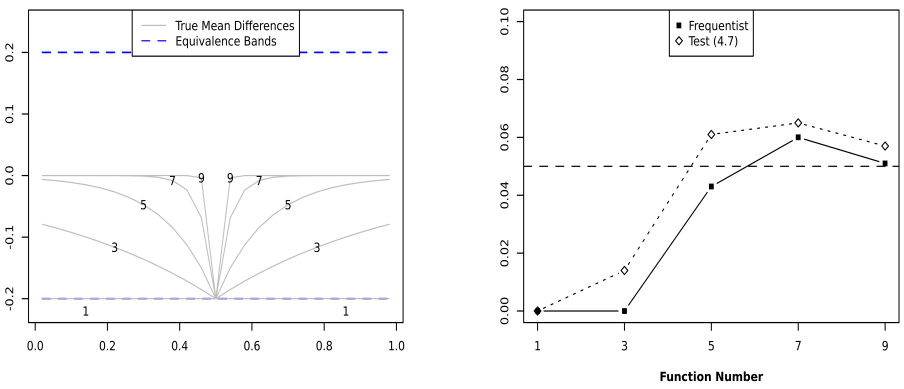


Figure 4: *Approximation of the nominal level by the frequentist test proposed by Fogarty and Small (2014) (called Frequentist) and the test* (4.7) *for the hypotheses* (2.1) *with* $\kappa_l \equiv -0.2$ *and* $\kappa_u \equiv 0.2$. *Left panel: True differences for scenarios* 1, 3, 5, 7 *and* 9. *Right part: simulated nominal level.*

which also shows the equivalence bounds given by $\kappa_l(t) \equiv -0.2$ and $\kappa_u(t) \equiv 0.2$. Note that Fogarty and Small (2014) only investigate the equivalence between the curves on the set $\{t_j = (j-0.5)/25 \colon j = 1, \ldots, 25\}$ in their simulations and for the sake of comparison, we consider the same set here. The variance function σ_1^2 is also the same in each scenario and it is displayed in the right panel of Figure 3. The variance function σ_2^2 is defined by

$$\frac{\sigma_1^2}{\sigma_2^2}(t) = \exp\big(\log(2)\exp\big(-a_i\,|t-1/2|\big)\big) \tag{4.15}$$

where $i = 1, 3, 5, 7, 9$.

The right part of the Figure 4 shows the simulated nominal level of the bootstrap test (4.7) and the frequentist test proposed by Fogarty and Small (2014) for the five cases under consideration. We observe that the frequentist test of Fogarty and Small (2014) approximates the nominal level rather well for the function 9, slightly exceeds the nominal level for function 7 and is conservative in the cases 1, 3 and 5. The test (4.7) shows a similar picture, where it provides a better approximation of the nominal level for the function 3 and slightly exceeds the nominal in the cases 5, 7 and 9.

Next we study the power of the two tests for the hypotheses (2.1). The mean function μ_1 is given in the left panel of Figure 3 and μ_2 is defined by

$$\mu_2(t) = \mu_1(t) - b_i \cos(2\pi t) - c_i$$

$b_i = 0.05 - 0.1 \cdot (i-1)/14$ and $c_i = 0.15 - 0.3 \cdot (i-1)/14$ for $i = 1, \ldots, 8$. The variance function σ_1^2 is given in the right panel of Figure 3 and σ_2^2 is defined by

$$\frac{\sigma_1^2}{\sigma_2^2}(t) = (0.1\cos(2\pi t) + 1.8)^{d_i}\,, \tag{4.16}$$

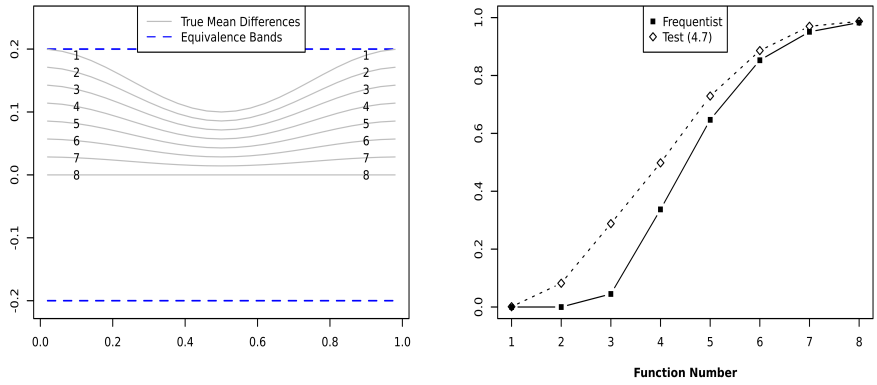


Figure 5: *Power comparison of the frequentist test proposed by Fogarty and Small (2014) (called Frequentist) and the test* (4.7) *for the hypotheses* (2.1) *with* $\kappa_l \equiv -0.2$ *and* $\kappa_u \equiv 0.2$. *Left panel: true mean difference for each scenario (1-8). Right panel: simulated rejection probabilities.*

where $d_i = -1 + 2 \cdot (i-1)/14$ for $i = 1, \ldots, 8$. The mean differences are depicted in the left part of Figure 5. We observe that the frequentist test of Fogarty and Small (2014) is outperformed by the new test (4.7) proposed in this paper. While the differences between the test (4.7) and the frequentist test of Fogarty and Small (2014) are small in scenarios $6-8$ (because the power of both tests is close to 1), we observe substantial advantages of the new test (4.7) for the functions $2-5$.

4.3.2 Variance functions

In this section, we consider the same scenarios as in the previous section and investigate the finite sample properties of the tests for the equivalence of the variance functions of the two samples. For the different scenarios, the decision rule in (4.14) is applied in order to decide for the null or the alternative hypothesis which are defined by (4.8) or equivalently by (2.2). The results are then compared with those of the frequentist test developed in Fogarty and Small (2014). In the left part of Figure 6, we display the true ratio of the variance functions for each considered scenario in (4.15) as well as the equivalence bands defined by $\zeta_l \equiv 0.5, \zeta_u \equiv 2$. The right part of this figure displays the simulated nominal level of the bootstrap test (4.14) and the frequentist test proposed by Fogarty and Small (2014) on the boundary of the null hypothesis. Similar to the results for testing the equivalence of the means, the frequentist approximates the test level slightly better than the bootstrap test in the scenarios $5, 7$ and 9. In scenario 1, both tests are conservative. The same is true for scenario 3 but in this case, the empirical rejection probability of the new test is closer to the nominal level.

The true ratio of the variance functions for the considered scenarios under the alternative hypothesis and the used equivalence bands $\zeta_l \equiv 1/1.9, \zeta_u \equiv 1.9$ are displayed in the left part of Figure 7. Only the functions 1 - 5 in (4.16) are considered since both tests always reject

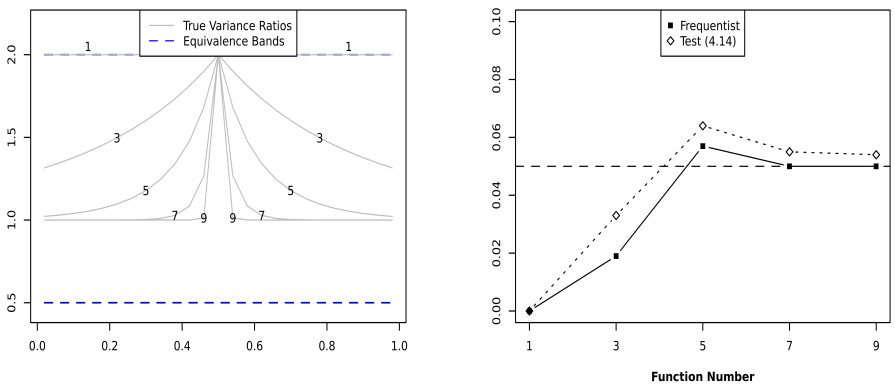


Figure 6: *Approximation of the nominal level by the frequentist test proposed by Fogarty and Small (2014) (called Frequentist) and the test* (4.14) *for the hypotheses* (2.2) *with* $\zeta_l \equiv 0.5$ *and* $\zeta_u \equiv 2$. *Left part: True ratio of the variance functions in the scenarios* $1, 3, 5, 7$ *and* 9 *in* (4.15). *Right part: Simulated rejection probabilities.*

the null hypothesis in the cases $6 - 8$. The rejection rates of the two tests corresponding to the five considered scenarios are displayed in the right panel of Figure 7. We observe a superior performance of the new bootstrap test (4.14) in all the considered scenarios, where in the scenarios 1, 4 and 5 the differences are very small.

Acknowledgements This research was partially supported by the Collaborative Research Center 'Statistical modeling of nonlinear dynamic processes' (*Sonderforschungsbereich 823, Teilprojekt A1, C1*) and the Research Training Group 'High-dimensional phenomena in probability - fluctuations and discontinuity' (*RTG 2131*). The authors are grateful to Martina Stein, who typed parts of this manuscript with considerable technical expertise and to Dr. Colin Fogarty for sending us the code of the procedures developed by Fogarty and Small (2014).

References

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.

Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics*, 24:295–300.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2:107–144.

Bücher, A. and Kojadinovic, I. (2019). A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. *Journal of Theoretical Probability*, 32:1145–1165.

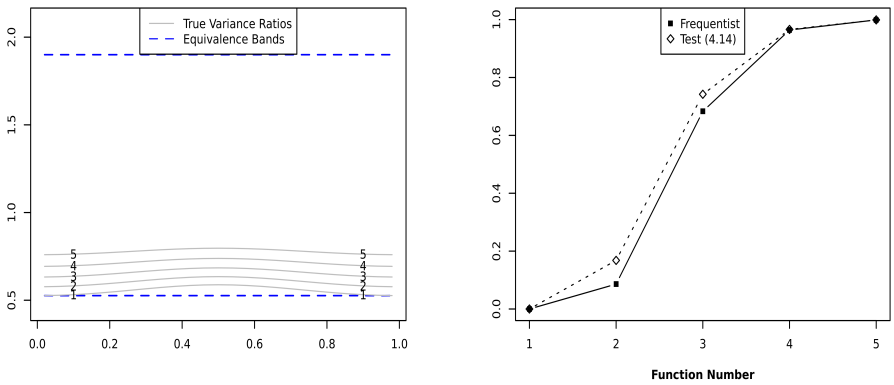


Figure 7: *Power comparison of the frequentist test proposed by Fogarty and Small (2014) (called Frequentist) and the test* (4.14) *for the hypotheses* (2.2) *with* $\zeta_l \equiv 1/1.9$ *and* $\zeta_u \equiv 1.9$. *Left part: True ratio of the variance functions in the scenarios* $1-5$ *in* (4.16). *Right part: Empirical rejection probabilities.*

Cárcamo, J., Rodríguez, L.-A., and Cuevas, A. (2020). Directional differentiability for supremum-type functionals: statistical applications. *Bernoulli, to appear; ArXiv e-print 1902.01136*.

Dette, H., Kokot, K., and Aue, A. (2020). Functional data analysis in the banach space of continuous functions. *Annals of Statistics, to appear; ArXiv e-print 1710.07781v2*.

Dette, H., Möllenhoff, K., Volgushev, S., and Bretz, F. (2018). Equivalence of regression curves. *Journal of the American Statistical Association*, 113:711–729.

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

Fogarty, C. B. and Small, D. S. (2014). Equivalence testing for functional data with an application to comparing pulmonary function devices. *Ann. Appl. Stat.*, 8(4):2002–2026.

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

Gsteiger, S., Bretz, F., and Liu, W. (2011). Simultaneous confidence bands for nonlinear regression models with application to population pharmacokinetic analyses. *Journal of Biopharmaceutical Statistics*, 21(4):708–725.

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

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to linear Operators*. Wiley, New York.

Janson, S. and Kaijser, S. (2015). Higher moments of Banach space valued random variables. *Memoirs of the American Mathematical Society*, 238.

Liebl, D. and Reimherr, M. (2019). Fast and fair simultaneous confidence bands for fundtional parameters. *arXiv:1910.00131*.

Liu, W., Bretz, F., Hayter, A. J., and Wynn, H. P. (2009). Assessing non-superiority, non-inferiority of equivalence when comparing two regression models over a restricted covariate region. *Biometrics*,

65(4):1279–1287.

Möllenhoff, K., Dette, H., Kotzagiorgis, E., Volgushev, S., and Collignon, O. (2018). Regulatory assessment of drug dissolution profiles comparability via maximum deviation. *Statistics in Medicine*, 37(20):2968–2981.

Paixão, P., Gouveia, L. F., Silva, N., and Morais, J. A. (2017). Evaluation of dissolution profile similarity - Comparison between the f_2, the multivariate statistical distance and the f_2 bootstrapping methods. *European Journal of Pharmaceutics and Biopharmaceutics*, 79:29–50.

Phillips, K. F. (1990). Power of the two one-sided tests procedure in bioequivalence. *Journal of pharmacokinetics and biopharmaceutics*, 18(2):137–144.

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

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of pharmacokinetics and biopharmaceutics*, 15(6):657–680.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

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

Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority*. CRC Press.

Yoshida, H., Shibata, H., Izutsu, K. I., and Goda, Y. (2017). Comparison of dissolution similarity assessment methods for products with large variations: f_2 statistics and model-independent multivariate confidence region procedure for dissolution profiles of multiple oral products. *Biological and Pharmaceutical Bulletin*, 40(5):722–725.

A Appendix: theoretical justification

In this section we provide proofs and the necessary assumptions for our main theoretical results. We begin with some basic facts about Banach space valued random variables.

A.1 $C([0,1])$-valued random variables

Throughout this paper we assume that all random variables are elements of the space $C([0,1])$ of all continuous functions from the compact set $[0,1]$ into $\mathbb{R}$. The space $C([0,1])$ is equipped with the sup-norm defined by $\|f\|_\infty = \sup_{t\in[0,1]} |f(t)|$. It is assumed that the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is complete and speaking of measurability is always meant with respect to the natural Borel σ-field $\mathcal{B}([0,1])$ (generated by the open sets relative to the sup-norm $\|\cdot\|_\infty$). Theorem 11.7 in Janson and Kaijser (2015) implies that $C([0,1])$ is separable and measurability issues are avoided. Completeness and separability of $C([0,1])$ directly imply that any random variable X in $C([0,1])$ is tight (see Theorem 1.3 in Billingsley, 1968)).

Expectations and higher-order moments of $C([0,1])$-valued random variables can be defined formally in different ways, for example through injective tensor products (see Janson and Kaijser, 2015). Denote by $\mathbb{E}[X]$ the expectation of a random variable X in $C([0,1])$ and note that it exists as an element of $C([0,1])$ whenever $\mathbb{E}[\|X\|_\infty] < \infty$. Generally, the kth moment of X exists as an element of $C([0,1]^k)$ whenever $\mathbb{E}[\|X\|_\infty^k] = \mathbb{E}[\sup_{t\in[0,1]} |X(t)|^k] < \infty$ and it can be computed through pointwise evaluation as $\mathbb{E}[X(t_1)\cdots X(t_k)]$ (see Chapter 11 of Janson and Kaijser, 2015)). In particular, covariance kernels of random variables in $C([0,1])$ can be computed in a pointwise fashion and the variance function of X can be defined by $\sigma^2(t) = \mathbb{E}[(X(t) - \mathbb{E}[X(t)])^2]$. A random variable $X \in C([0,1])$ is said to be Gaussian if all finite dimensional vectors $(X(t_1), \ldots, X(t_k))$ are multivariate normal distributed (for any $t_1, \ldots, t_k \in [0,1]$ and $k \in \mathbb{N}$). The distribution of Gaussian random variables in $C([0,1])$ is completely characterized by its expectation and its covariance function (see Chapter 2 of Billingsley, 1968). Throughout this paper, weak convergence in $C([0,1])$ is denoted by the symbol "$\rightsquigarrow$" and the symbol $\xrightarrow{\mathcal{D}}$ denotes weak convergence of a finite dimensional random variable.

A.2 Proofs of the results in Section 3

In this section we provide rigorous arguments for the statements made in Section 3. In the following discussion $BL_1(C([0,1]))$ denotes the space of bounded (by 1) Lipschitz functions from $C([0,1])$ into $\mathbb{R}$. That is the set of all functions $h : C([0,1]) \to \mathbb{R}$ with $\|h\|_\infty \leq 1$ and $|h(x) - h(y)| \leq \|x - y\|_\infty (= \sup_{t\in[0,1]} |x(t) - y(t)|)$ for any $x, y \in C([0,1])$.

A.2.1 Basic assumptions and a limit theorem for the maximum deviation estimate

For the proofs of the results in Section 3 we make the following assumption, which guarantees the existence of the central limit theorem for independent random variables in $C([0,1])$ (see the

discussion at the beginning of the proof of Theorem A.1).

Assumption A.1 *Let* $(X_{1j}\colon j \in \mathbb{N})$ *and* $(X_{2j}\colon j \in \mathbb{N})$ *denote two independent sequences of* $C([0,1])$*-valued random variables such that each sequence has independent identically distributed elements and assume that the following conditions are satsified:*

(A1) There exist constants $\nu_1, \nu_2 > 0$*,* K *such that, for all* $j \in \mathbb{N}$ *and* $i = 1, 2$*,*

$$\mathbb{E}\big[\|X_{ij}\|_\infty^{2+\nu_i}\big] \le K\,, \quad \mathbb{E}\big[\|X_{ij}\|_\infty^4\big] < \infty\,.$$

(A2) There exists a constant $\vartheta > 1/4$ *and a real-valued non-negative random variable* M_i *with* $\mathbb{E}\big[M_i^4\big] < \infty$ *such that, for* $i = 1, 2$ *and any* $j \in \mathbb{N}$*, the inequality*

$$|X_{ij}(s) - X_{ij}(t)| \le M_i\, \rho(s,t) = M_i\, |s-t|^\vartheta$$

holds almost surely for all $s, t \in [0,1]$*.*

Theorem A.1 *If Assumption A.1 holds, we have*

$$\sqrt{n+m}\big(\hat T^\theta_{m,n} - T^\theta\big) \xrightarrow{\mathcal{D}} Z_{\mathcal{E},\theta} = \max\Big\{ \sup_{t\in\mathcal{E}^l_\theta} \big(-Z(t)\big),\, \sup_{t\in\mathcal{E}^u_\theta} Z(t)\Big\}\,, \tag{5.1}$$

where T^θ *and* $\hat T^\theta_{m,n}$ *are defined by* (3.1) *and* (3.2)*, respectively and the extremal sets* $\mathcal{E}^l_\theta$ *and* $\mathcal{E}^u_\theta$ *are defined by* (3.5)*.*

Proof. Note that condition (A2) and the fact that $\vartheta > 1/4$ imply $\int_0^{\tilde\tau} D(\omega,\rho)^{1/4}\, d\omega < \infty$ for some $\tilde\tau > 0$ where $D(\omega,\rho)$ denotes the packing number with respect to the metric $\rho(s,t) = |s-t|^\vartheta$ that is the maximal number of ω-seperated points in $[0,1]$ (see Van der Vaart and Wellner, 1996). Therefore it follows from Theorem 2.1 in Dette et al. (2020) that

$$\hat Z_{m,n} = \sqrt{m+n}\,(\hat\theta_{m,n} - \theta) \rightsquigarrow Z \tag{5.2}$$

in $C([0,1])$ as $m, n \to \infty$ where Z is a (tight) centred Gaussian random function in $C([0,1])$ with covariance kernel as defined in (3.4) (see also Remark 2.1 (b) in Dette et al., 2020). Observing the estimate

$$\begin{aligned}
&\sqrt{m+n}\big(\hat T^\theta_{m,n} - T^\theta\big)\\
&= \sqrt{m+n}\Big(\max\Big\{ \sup_{t\in[0,1]} \big(-\hat\theta_{m,n}(t) + \kappa_l(t)\big),\, \sup_{t\in[0,1]} \big(\hat\theta_{m,n}(t) - \kappa_u(t)\big)\Big\} - T^\theta\Big)\\
&= \sqrt{m+n}\Big(\max\Big\{ \sup_{t\in\mathcal{E}^l_\theta} \big(-\hat\theta_{m,n}(t) + \kappa_l(t)\big),\, \sup_{t\in\mathcal{E}^u_\theta} \big(\hat\theta_{m,n}(t) - \kappa_u(t)\big)\Big\} - T^\theta\Big) + o_{\mathbb{P}}(1)\\
&= \sqrt{m+n}\max\Big\{ \sup_{t\in\mathcal{E}^l_\theta} \big(-\hat\theta_{m,n}(t) + \kappa_l(t) - T^\theta\big),\, \sup_{t\in\mathcal{E}^u_\theta} \big(\hat\theta_{m,n}(t) - \kappa_u(t) - T^\theta\big)\Big\} + o_{\mathbb{P}}(1)\\
&= \sqrt{m+n}\max\Big\{ \sup_{t\in\mathcal{E}^l_\theta} \big(-\hat\theta_{m,n}(t) + \theta(t)\big),\, \sup_{t\in\mathcal{E}^u_\theta} \big(\hat\theta_{m,n}(t) - \theta(t)\big)\Big\} + o_{\mathbb{P}}(1)\,,
\end{aligned}$$

the assertion of Theorem A.1 follows from (5.2) and the continuous mapping theorem. □

A.2.2 Proof of Theorem 3.1

We begin showing that the bootstrap process $\hat{Z}^{*(r)}_{m,n}$ converges conditionally given the data $(X_{lj}\colon j \in \mathbb{N};\ l = 1, 2)$, to the same limit as $\hat{Z}_{m,n}$. More precisely, this means

$$\sup_{h \in BL_1(C([0,1]))} |\mathbb{E}_M h(\hat{Z}^{*(r)}_{m,n}) - \mathbb{E}h(Z)| = o_{\mathbb{P}}(1) \tag{5.3}$$

as $m, n \to \infty$ where $\mathbb{E}_M$ denotes the conditional expectation given the data $(X_{lj}\colon j \in \mathbb{N};\ l = 1, 2)$ and the random variable Z is defined by (3.4) (see for example Section 23.2.1 in Van der Vaart, 1998).

Note that the convergence in (5.3) holds under the null and under the alternative hypothesis. By the continuous mapping theorem and similar arguments as given in Lemma B.3 of Dette et al. (2020) it follows that the bootstrap statistic $\hat{T}^{\theta,*(r)}_{m,n}$ defined by (3.8) converges conditionally given the data $(X_{lj}\colon j \in \mathbb{N};\ l = 1, 2)$ to the same limit as $\sqrt{m+n}\,(\hat{T}^{\theta}_{m,n} - T^{\theta})$ that is $Z_{\mathcal{E},\theta}$ (see (3.3)). If $T^{\theta} = 0$, Lemma 4.2 in Bücher and Kojadinovic (2019) directly implies the first assertion of Theorem 3.1 that is

$$\lim_{m,n,R\to\infty} \mathbb{P}\big(\sqrt{m+n}\,\hat{T}^{\theta}_{m,n} < z^{*(R)}_{m,n,\alpha}\big) = \alpha$$

(note that the continuity of the random variable $Z_{\mathcal{E},\theta}$ is implied by the results in Gaenssler et al. (2007)). If $T^{\theta} \neq 0$, write

$$\mathbb{P}\big(\sqrt{m+n}\,\hat{T}^{\theta}_{m,n} < z^{*(R)}_{m,n,\alpha}\big) = \mathbb{P}\big(\sqrt{m+n}\,(\hat{T}^{\theta}_{m,n} - T^{\theta}) + \sqrt{m+n}\,T^{\theta} < z^{*(R)}_{m,n,\alpha}\big)\,.$$

Then it follows from (5.1), (5.3) and simple arguments that, for any $R \in \mathbb{N}$,

$$\lim_{m,n\to\infty} \mathbb{P}\big(\sqrt{m+n}\,\hat{T}^{\theta}_{m,n} < z^{*(R)}_{m,n,\alpha}\big) = 0 \quad \text{and} \quad \liminf_{m,n\to\infty} \mathbb{P}\big(\sqrt{m+n}\,\hat{T}^{\theta}_{m,n} < z^{*(R)}_{m,n,\alpha}\big) = 1$$

if $T^{\theta} > 0$ and $T^{\theta} \leq 0$, respectively. This proves the remaining assertions of Theorem 3.1.

In order to prove the convergence in (5.3), we will utilize the link between weak convergence in the Banach space of continuous functions $C([0, 1])$ and weak convergence of empirical processes in the space of bounded functions $l^{\infty}(\mathcal{F})$ from an appropriately defined function space $\mathcal{F}$ into $\mathbb{R}$. In fact we use the CLT derived in the proof of Theorem A.1 (see equation (5.2)) and conclude that Theorem 23.7 in Van der Vaart (1998) can be applied to show weak convergence of the empirical bootstrap process in $l^{\infty}(\mathcal{F})$ conditionally given the data $(X_{lj}\colon j \in \mathbb{N};\ l = 1, 2)$. Afterwards it will be argued that this again implies the convergence in (5.3).

For any $t \in [0, 1]$ and $x \in C([0, 1])$, consider the canonical projection $\pi_t\colon C([0, 1]) \to \mathbb{R}$ with $\pi_t(x) = x(t)$, define the function class

$$\mathcal{F} = \{\pi_t\colon t \in [0, 1]\}$$

and note that this class is a subset of $C([0, 1])^*$, the dual space of $C([0, 1])$. Defining the map $x^{**} : \mathcal{F} \to \mathbb{R}$ by $x^{**}(\pi_t) = \pi_t(x) = x(t)$, for any $x \in C([0, 1])$, leads to an isometric identification

of $C([0,1])$ with a subset of $l^\infty(\mathcal{F})$ and in the following, this subset is denoted by $C([0,1])^{**}$. Both, $C([0,1])$ and $l^\infty(\mathcal{F})$, are equipped with the respective sup-norm $\|\cdot\|_\infty$ and it is clear that $\|x\|_\infty = \|x^{**}\|_\infty$. For any $C([0,1])$-valued random variable X the corresponding random variable $X^{**} \in C([0,1])^{**} \subset l^\infty(\mathcal{F})$ is defined by

$$X^{**}(\pi_t) = \delta_X \pi_t = \int_{C([0,1])} \pi_t(x)\delta_X(dx) = X(t)$$

where δ_X denotes the dirac measure. Next, show that the weak convergence of a sequence of random variables in $C([0,1])$, that is $X_n \rightsquigarrow X$, is equivalent to the weak convergence $X_n^{**} \rightsquigarrow X^{**}$ in $l^\infty(\mathcal{F})$ (see also Section 2.1.4 in Van der Vaart and Wellner, 1996). Following Section 1.12 in Van der Vaart and Wellner (1996) (note that each random variable $X \in C([0,1])$ is separable), weak convergence of a sequence $X_n \subset C([0,1])$ to a separable random variable $X \in C([0,1])$, denoted by $X_n \rightsquigarrow X$, is equivalent to

$$\sup_{h \in BL_1(C([0,1]))} |\mathbb{E}^* h(X_n) - \mathbb{E}h(X)| = o(1)\,.$$

Each function $h \in BL_1(C([0,1]))$ can be identified by the function $h^{**} : C([0,1])^{**} \to \mathbb{R}$ defined through $h^{**}(x^{**}) = h(x)$ for any $x^{**} \in C([0,1])^{**}$. Note that $\|h^{**}\|_\infty = \|h\|_\infty \leq 1$ and

$$|h^{**}(x^{**}) - h^{**}(y^{**})| = |h(x) - h(y)| \leq \|x - y\|_\infty = \|x^{**} - y^{**}\|_\infty.$$

Thus $h^{**} \in BL_1(C([0,1])^{**})$ and

$$\begin{aligned} o(1) &= \sup_{h \in BL_1(C([0,1]))} |\mathbb{E}^* h(X_n) - \mathbb{E}h(X)| \\ &= \sup_{h^{**} \in BL_1(C([0,1])^{**})} |\mathbb{E}^* h^{**}(X_n^{**}) - \mathbb{E}h^{**}(X^{**})| \\ &= \sup_{g \in BL_1(l^\infty(\mathcal{F}))} |\mathbb{E}^* g|_{C([0,1])^{**}}(X_n^{**}) - \mathbb{E}g|_{C([0,1])^{**}}(X^{**})| \\ &= \sup_{g \in BL_1(l^\infty(\mathcal{F}))} |\mathbb{E}^* g(X_n^{**}) - \mathbb{E}g(X^{**})| \end{aligned} \tag{5.4}$$

where $g|_{C([0,1])^{**}}$ denotes the restriction of the function g to $C([0,1])^{**}$. Consequently $X_n^{**} \rightsquigarrow X^{**}$ in $l^\infty(\mathcal{F})$ (by construction, X and X^{**} can be defined on the same original probability space and we have that X^{**} is separable if and only if X is separable).

Since the envelope function of $\mathcal{F}$, $F(x) = \|x\|_\infty$, is finite for any $x \in C([0,1])$ and the convergence in (5.2) together with the previous discussion means that $\mathcal{F}$ is a Donsker class, Theorem 23.7 in Van der Vaart (1998) can be applied. For this purpose, note that

$$\hat{Z}_{m,n}^{*(r)} = \frac{\sqrt{m+n}}{\sqrt{m}} \hat{Z}_{1,m}^{*(r)} + \frac{\sqrt{m+n}}{\sqrt{n}} \hat{Z}_{2,n}^{*(r)},$$

where

$$\hat{Z}_{1,m}^{*(r)} = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} (X_{1j}^{*(r)} - \overline{X}_{1\cdot}) = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} (M_{1j}^{(r)} - 1) X_{1j}$$

and $\hat{Z}_{2,n}^{*(r)}$ is defined analogously. The random vector $(M_{11}^{(r)}, \ldots, M_{1m}^{(r)})$ follows a multinomial distribution with parameters m, $p_j = 1/m$, $j = 1, \ldots, m$. The corresponding empirical bootstrap process can be written as

$$(\hat{Z}_{1,m}^{*(r)})^{**} = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} (M_{1j}^{(r)} - 1)\delta_{X_{1j}} .$$

Now, Theorem 23.6 in Van der Vaart (1998) implies that

$$\sup_{g \in BL_1(l^\infty(\mathcal{F}))} |\mathbb{E}_M g((\hat{Z}_{1,m}^{*(r)})^{**}) - \mathbb{E}g(Z_1^{**})| = o_{\mathbb{P}}(1)$$

where $\mathbb{E}_M$ denotes the conditional expectation given the data $(X_{lj} \colon j \in \mathbb{N};\ l = 1, 2)$ and Z_1^{**} is the (unconditional) limit of $\sqrt{m}(\overline{X}_{1\cdot} - \mu_1)^{**}$. Similar arguments as in (5.4) and the subsequent discussion yield that this equation is equivalent to

$$\sup_{h \in BL_1(C([0,1]))} |\mathbb{E}_M h(\hat{Z}_{1,m}^{*(r)}) - \mathbb{E}h(Z_1)| = o_{\mathbb{P}}(1)$$

which means that the sequence $\hat{Z}_{1,m}^{*(r)}$ converges conditionally given the data $(X_{lj} \colon j \in \mathbb{N};\ l = 1, 2)$ to Z_1 in $C([0, 1])$. The corresponding statement for $\hat{Z}_{2,n}^{*(r)}$ can be derived similarly. Since $\hat{Z}_{1,m}^{*(r)}$ and $\hat{Z}_{2,n}^{*(r)}$ are independent (5.3) now follows, which completes the proof of Theorem 3.1.

A.2.3 Proof of Remark 3.1

In this section we consider the case of dependent data and give some arguments why the decision rule in (3.11) based on the multiplier bootstrap process defined by (3.10) yields a consistent and asymptotic level α-test. For that consider the dependency concept of φ-mixing (see for example Bradley (2005)). Denote by $\mathbb{P}(G|F)$ the conditional probability of G given F and, for any two σ-fields $\mathcal{F}$ and $\mathcal{G}$, define

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

For a given stationary sequence $(\eta_j \colon j \in \mathbb{N})$ of random variables in $C([0, 1])$, denote by $\mathcal{F}_k^{k'}$ the σ-field generated by $(\eta_j \colon k \leq j \leq k')$. Then, the kth *φ-mixing coefficient* of $(\eta_j \colon j \in \mathbb{N})$ is defined by

$$\varphi(k) = \sup_{k' \in \mathbb{N}} \phi(\mathcal{F}_1^{k'}, \mathcal{F}_{k'+k}^{\infty})$$

and the stationary time series $(\eta_j \colon j \in \mathbb{N})$ is called *φ-mixing* whenever the sequence of mixing coefficients converges to zero as $k \to \infty$.

The statement in Remark 3.1 is correct, if the following assumptions are satisfied:

(B1) $(X_{1,j} \colon j \in \mathbb{N})$ and $(X_{2,j} \colon j \in \mathbb{N})$ are independent stationary time series satisfying conditions (A1) and (A2) in Assumption A.1.

(B2) Both sequences are φ-mixing and the mixing coefficients satisfy, for $i = 1, 2$,

$$\sum_{k=1}^{\infty} k^{1/(1/2-\bar{\tau}_i)} \varphi_i(k)^{1/2} < \infty\,, \quad \sum_{k=1}^{\infty} (k+1)\varphi_i(k)^{1/4} < \infty\,,$$

for some $\bar{\tau}_i \in (1/(2+2\nu_i), 1/2)$ where the constant ν_i is the same as in (A1).

(B3) The window parameters l_1, l_2 in the definition of the bootstrap processes in (3.10) are defined by $l_1 = m^{\beta_1}$, $l_2 = n^{\beta_2}$ such that

$$0 < \beta_i < \nu_i/(2+\nu_i)\,, \quad \bar{\tau}_i > (\beta_i(2+\nu_i)+1)/(2+2\nu_i)\,,$$

and the constants ν_i and $\bar{\tau}_i$ are given in (A1) and (B2), respectively ($i = 1, 2$).

Under these assumptions it follows from Theorem 2.1 in Dette et al. (2020) that the CLT in (5.2) also holds in the dependent case, where the limiting process Z is a Gaussian process with covariance kernel

$$C(s,t) = \frac{1}{\tau} C_1(s,t) + \frac{1}{1-\tau} C_2(s,t)$$

and $C_i(s,t) = \sum_{j=-\infty}^{\infty} \mathrm{Cov}(X_{i0}(s), X_{ij}(t))$ ($i = 1, 2$). Similarly, by Theorem 3.3 in the same reference the bootstrap process (3.10) satisfies

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

in $C([0,1])^{R+1}$ as $m, n \to \infty$ where $Z^{(1)}, \ldots Z^{(R)}$ are independent copies of Z. Note that this is equivalent to the corresponding statement in (5.3) (by Lemma 2.2 in Bücher and Kojadinovic, 2019) and therefore, the statement of Remark 3.1 now follows by similar arguments as given in the proof of Theorem 3.1.

A.3 Proofs of the results in Section 4

A.3.1 Proofs of the results in Section 4.1

Recall the model defined in (4.1) and assume the following:

Assumption A.2 *The differences of the error terms are sampled from the independent sequences* $((\varepsilon_{1,i} - \varepsilon_{2,i})\colon i \in \mathbb{N})$ *and* $((\eta_{1,i,j} - \eta_{2,i,j})\colon i, j \in \mathbb{N})$ *where each sequence has independent and identically distributed elements and satisfies Assumption A.1.*

Proof of Theorem 4.1. Theorem 2.1 in Dette et al. (2020) implies

$$\frac{1}{\sqrt{A}} \sum_{i=1}^{A} (\varepsilon_{1,i} - \varepsilon_{2,i}) \rightsquigarrow Z_\varepsilon \quad \text{and} \quad \frac{1}{\sqrt{N}} \sum_{i=1}^{A} \sum_{j=1}^{n_i} (\eta_{1,i,j} - \eta_{2,i,j}) \rightsquigarrow Z_\eta$$

in $C([0,1])$ as $A \to \infty$, $\min_{i=1}^{A} n_i \to \infty$ where Z_ε and Z_η are centred Gaussian processes with covariance kernels

$$k_\varepsilon(s,t) = \mathrm{Cov}(\varepsilon_{1,1}(s) - \varepsilon_{2,1}(s), \varepsilon_{1,1}(t) - \varepsilon_{2,1}(t))\,,$$
$$k_\eta(s,t) = \mathrm{Cov}(\eta_{1,1,1}(s) - \eta_{2,1,1}(s), \eta_{1,1,1}(t) - \eta_{2,1,1}(t))\,,$$

respectively. Then we have

$$\begin{aligned}\sqrt{A}\,(\hat{\theta}_N - \theta) &= \frac{1}{\sqrt{A}} \sum_{i=1}^{A} \Big\{ \varepsilon_{1,i} - \varepsilon_{2,i} + \frac{1}{n_i} \sum_{j=1}^{n_i} (\eta_{1,i,j} - \eta_{2,i,j}) \Big\} \\ &= \frac{1}{\sqrt{A}} \sum_{i=1}^{A} (\varepsilon_{1,i} - \varepsilon_{2,i}) + o_{\mathbb{P}}(1) \rightsquigarrow Z_\varepsilon\end{aligned} \tag{5.5}$$

in $C([0,1])$ as $A \to \infty$, $\min_{i=1}^{A} n_i \to \infty$ and similar arguments as in the proof of Theorem A.1 yield

$$\hat{T}_N^\theta \xrightarrow{\mathcal{D}} Z_{\mathcal{E},\theta} = \max\Big\{ \sup_{t \in \mathcal{E}_\kappa^l} \big(- Z_\varepsilon(t) \big), \sup_{t \in \mathcal{E}_\kappa^u} Z_\varepsilon(t) \Big\}\,, \tag{5.6}$$

where the statistic $\hat{T}_N^\theta$ is defined by (4.2).

In order to establish the second equality in (5.5), we define

$$\tilde{\eta}_N = \frac{1}{\sqrt{A}} \sum_{i=1}^{A} \frac{1}{n_i} \sum_{j=1}^{n_i} (\eta_{1,i,j} - \eta_{2,i,j}) \tag{5.7}$$

and show that this process converges to zero in probability in $C([0,1])$ as $A \to \infty$, $\min_{i=1}^{A} n_i \to \infty$. For this purpose we show that the finite dimensional distributions converge to zero and prove that $\tilde{\eta}_N$ is asymptotically ρ-equicontinuous in probability, where $\rho(s,t) = |s-t|^\theta$. This proves $\tilde{\eta}_N \rightsquigarrow 0$ in $C([0,1])$ under the stated assumptions (see Theorem 7.5 in Billingsley, 1968). By the Cramér-Wold device, convergence of the finite dimensional distributions to zero is equivalent to

$$\sum_{k=1}^{q} c_k \tilde{\eta}_N(t_k) = \frac{1}{\sqrt{A}} \sum_{i=1}^{A} \frac{1}{n_i} \sum_{j=1}^{n_i} \sum_{k=1}^{q} c_k (\eta_{1,i,j}(t_k) - \eta_{2,i,j}(t_k)) \xrightarrow{\mathcal{D}} 0 \tag{5.8}$$

for any $t_1, \ldots, t_q \in [0,1]$ and $q \in \mathbb{N}$. Using that the differences $\eta_{1,i,j} - \eta_{2,i,j}$, $i = 1, \ldots, A$, $j = 1, \ldots, n_i$, are independent and the fact that $\mathbb{E}[\eta_{1,i,j}(t_k) - \eta_{2,i,j}(t_k)] = 0$ yields

$$\begin{aligned}\mathbb{E}\Big[\Big(\sum_{k=1}^{q} c_j \tilde{\eta}_N(t_k) \Big)^2 \Big] &= \frac{1}{A} \sum_{i=1}^{A} \frac{1}{n_i^2} \sum_{j=1}^{n_i} \mathbb{E}\Big[\Big(\sum_{k=1}^{q} c_k (\eta_{1,i,j}(t_k) - \eta_{2,i,j}(t_k)) \Big)^2 \Big] \\ &\lesssim \frac{1}{A} \sum_{i=1}^{A} \frac{1}{n_i} \leq \frac{1}{\min_{i=1}^{A} n_i} \to 0\end{aligned}$$

where we also used assumption (A1) and the symbol "$\lesssim$" means less or equal up to a constant factor independent of $A, n_1, \dots n_A$. This proves (5.8) and thus the convergence of the finite dimensional distributions to 0.

In order to verify the equicontinuity condition, we utilize Theorem 2.2.4 in Van der Vaart and Wellner (1996). For any $i = 1, \dots, A$, $j = 1, \dots, n_i$, define $\eta_{i,j} = \eta_{1,i,j} - \eta_{2,i,j}$. Then, for any $s, t \in [0,1]$, we have

$$\begin{aligned}
\mathbb{E}\big[|\tilde\eta_N(s) - \tilde\eta_N(t)|^4\big]^{1/4} &= \mathbb{E}\Big[\Big|\frac{1}{\sqrt{A}}\sum_{i=1}^{A}\frac{1}{n_i}\sum_{j=1}^{n_i}(\eta_{i,j}(s) - \eta_{i,j}(t))\Big|^4\Big]^{1/4} \\
&= \frac{1}{\sqrt{A}}\Big(\sum_{i=1}^{A}\frac{1}{n_i^4}\mathbb{E}\Big[\Big(\sum_{j=1}^{n_i}(\eta_{i,j}(s) - \eta_{i,j}(t))\Big)^4\Big] \\
&\quad + \sum_{i\neq i'}^{A}\frac{1}{n_i^2}\frac{1}{n_{i'}^2}\mathbb{E}\Big[\Big(\sum_{j=1}^{n_i}(\eta_{i,j}(s) - \eta_{i,j}(t))\Big)^2\Big]\mathbb{E}\Big[\Big(\sum_{j=1}^{n_{i'}}(\eta_{i',j}(s) - \eta_{i',j}(t))\Big)^2\Big]\Big)^{1/4}.
\end{aligned}$$

Using assumptions (A1) and (A2) it follows that

$$\begin{aligned}
\mathbb{E}\Big[\Big(\sum_{j=1}^{n_i}(\eta_{i,j}(s) - \eta_{i,j}(t))\Big)^4\Big] &= \sum_{j=1}^{n_i}\mathbb{E}\big[(\eta_{i,j}(s) - \eta_{i,j}(t))^4\big] \\
&\quad + \sum_{j\neq j'}^{n_i}\mathbb{E}\big[(\eta_{i,j}(s) - \eta_{i,j}(t))^2\big]\mathbb{E}\big[(\eta_{i,j'}(s) - \eta_{i,j'}(t))^2\big] \\
&\lesssim (n_i + n_i^2)\rho(s,t)^4\,, \\
\mathbb{E}\Big[\Big(\sum_{j=1}^{n_i}(\eta_{i,j}(s) - \eta_{i,j}(t))\Big)^2\Big] &= \sum_{j=1}^{n_i}\mathbb{E}\big[(\eta_{i,j}(s) - \eta_{i,j}(t))^2\big] \lesssim n_i\rho(s,t)^2
\end{aligned}$$

which yields

$$\mathbb{E}\big[|\tilde\eta_N(s) - \tilde\eta_N(t)|^4\big]^{1/4} \lesssim \frac{1}{\sqrt{A}}\Big(\sum_{i=1}^{A}\frac{1}{n_i^2} + \sum_{i\neq i'}^{A}\frac{1}{n_i n_{i'}}\Big)^{1/4}\rho(s,t) \lesssim \rho(s,t)\,.$$

Now we obtain from Theorem 2.2.4 in Van der Vaart and Wellner (1996) and Markov's inequality that

$$\begin{aligned}
\mathbb{P}\Big(\sup_{\rho(s,t)\leq\delta}|\tilde\eta_N(s) - \tilde\eta_N(t)| > \varepsilon\Big) &\leq \frac{1}{\varepsilon^4}\mathbb{E}\big[|\tilde\eta_N(s) - \tilde\eta_N(t)|^4\big] \\
&\lesssim \Big(\int_0^\tau D(\omega,\rho)^{1/4}d\omega + \delta D(\tau,\rho)^{1/2}\Big)^4
\end{aligned}$$

for any $\varepsilon, \delta, \tau > 0$. The discussion at the beginning of the proof of Theorem A.1 and the fact that $\tau > 0$ is arbitrary finally imply

$$\lim_{\delta\searrow 0}\limsup_{A,\min n_i\to\infty}\mathbb{P}\Big(\sup_{\rho(s,t)\leq\delta}|\tilde\eta_N(s) - \tilde\eta_N(t)| > \varepsilon\Big) = 0$$

which means that $\tilde{\eta}_N = o_{\mathbb{P}}(1)$ since we already proved the convergence of the finite dimensional distributions to zero.

The bootstrap process in (4.5) can be written

$$
\begin{aligned}
B_N^{\star(r)} &= \frac{1}{\sqrt{A}} \sum_{i=1}^{A} (M_i^{(r)} - 1)(\overline{X}_{1,i,\cdot} - \overline{X}_{2,i,\cdot}) \\
&= \frac{1}{\sqrt{A}} \sum_{i=1}^{A} (M_i^{(r)} - 1)\big\{\varepsilon_{1,i} - \varepsilon_{2,i} + \frac{1}{n_i} \sum_{j=1}^{n_i} (\eta_{1,i,j} - \eta_{2,i,j})\big\} \qquad (5.9) \\
&= \frac{1}{\sqrt{A}} \sum_{i=1}^{A} (M_i^{(r)} - 1)(\varepsilon_{1,i} - \varepsilon_{2,i}) + o_{\mathbb{P}}(1)
\end{aligned}
$$

where $(M_1^{(r)}, \ldots, M_A^{(r)})$ follows a multinomial distribution with parameters A, $p_j = 1/A$, $j = 1, \ldots, A$ and the last estimate follows by the same arguments as given in the derivation of the second equality in (5.5).

Observe that we have a central limit theorem by the argument given in equation (5.5) and that the bootstrap process has a stochastic expansion given in (5.9). Therefore it follows by similar arguments as given in the proof of Theorem 3.1 that $B_N^{\star(r)}$ converges conditionally given $((\varepsilon_{1,i} - \varepsilon_{2,i})\colon i \in \mathbb{N})$ to Z_ε. The continuous mapping theorem and similar arguments as given in the proof of Lemma B.3 of Dette et al. (2020) yield that the bootstrap statistic defined by (4.4) converges conditionally given the data $((\varepsilon_{1,i} - \varepsilon_{2,i})\colon i \in \mathbb{N})$ to the limit $Z_{\mathcal{E},\theta}$ in (5.6). Furthermore, the assertions in Theorem 4.1 follow by the same arguments given in the proof of Theorem 3.1.

A.3.2 Proofs of the results in Section 4.2

Assumption A.3 *Let $(\eta_{1,i,j}\colon i,j \in \mathbb{N})$ and $(\eta_{2,i,j}\colon i,j \in \mathbb{N})$ denote two (possibly dependent) sequences of $C([0,1])$-valued random variables such that each sequence has independent identically distributed elements and assume that the following conditions are satsified:*

(D1) There exist constants $\nu_1, \nu_2 > 0$, K such that, for all $i,j \in \mathbb{N}$ and $l = 1,2$,

$$\mathbb{E}\big[\|\eta_{l,i,j}\|_\infty^{4+\nu_l}\big] \leq K\,, \quad \mathbb{E}\big[\|\eta_{l,i,j}\|_\infty^{8}\big] < \infty\,.$$

(D2) There exist constants $\vartheta > 1/4$, $\tilde{K}$ and real-valued non-negative random variables M_1 and M_2 with $\mathbb{E}\big[\|\eta_{l,i,j}\|_\infty^4 M_l^4\big] \leq \tilde{K} < \infty$ such that, for $l = 1,2$ and any $i,j \in \mathbb{N}$, the inequality

$$|\eta_{l,i,j}(s) - \eta_{l,i,j}(t)| \leq M_l\, \rho(s,t) = M_l\, |s-t|^\vartheta$$

holds almost surely for all $s,t \in [0,1]$.

Proof of Theorem 4.2. It is easy to see that Assumption A.3 implies that the sequences of squared individual errors $(\eta_{1,i,j}^2\colon i,j \in \mathbb{N})$ and $(\eta_{2,i,j}^2\colon i,j \in \mathbb{N})$ satisfy Assumption A.1. We have

for $l = 1, 2$

$$\begin{aligned}\hat{\sigma}_l^2 &= \frac{1}{N-A}\sum_{i=1}^{A}\sum_{j=1}^{n_i}\Big(X_{l,i,j} - \frac{1}{n_i}\sum_{k=1}^{n_i} X_{l,i,k}\Big)^2 \\ &= \frac{1}{N-A}\sum_{i=1}^{A}\sum_{j=1}^{n_i}\Big(\eta_{l,i,j} - \frac{1}{n_i}\sum_{k=1}^{n_i} \eta_{l,i,k}\Big)^2 \\ &= \frac{1}{N-A}\sum_{i=1}^{A}\sum_{j=1}^{n_i}(\eta_{l,i,j})^2 + o_{\mathbb{P}}\Big(\frac{1}{\sqrt{N}}\Big)\ ,\end{aligned}$$

where the last inequality follows from similar arguments as given in the discussion after equation (5.7). Now we have

$$\begin{aligned}&\sqrt{N}\big(\hat{\sigma}_1^2 - \sigma_1^2, \hat{\sigma}_2^2 - \sigma_2^2\big) \\ &\quad = \sqrt{N}\Big(\frac{1}{N-A}\sum_{i=1}^{A}\sum_{j=1}^{n_i}\big\{(\eta_{1,i,j})^2 - \sigma_1^2\big\}, \frac{1}{N-A}\sum_{i=1}^{A}\sum_{j=1}^{n_i}\big\{(\eta_{2,i,j})^2 - \sigma_2^2\big\}\Big) + o_{\mathbb{P}}(1) \rightsquigarrow (Z_1, Z_2)\end{aligned}$$

in $C([0,1])^2$ as $A \to \infty$, $\min_{i=1}^{A} n_i \to \infty$. This follows from the fact that by Theorem 2.1 in Dette et al. (2020) each component converges individually to its corresponding limiting process in $C([0,1])$. Therefore, both elements are asymptotically tight and marginal asymptotic tightness implies joint asymptotic tightness. The convergence of the finite dimensional distributions follows from the ordinary multidimensional central limit theorem. By the delta-method as stated in Proposition 2.1 in Cárcamo et al. (2020) we obtain the convergence

$$\begin{aligned}&\sqrt{N}\bigg(\log\Big(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}\Big) - \log\Big(\frac{\sigma_1^2}{\sigma_2^2}\Big)\bigg) \\ &\quad = \frac{\sqrt{N}}{N-A}\sum_{i=1}^{A}\sum_{j=1}^{n_i}\Big\{\frac{(\eta_{1,i,j})^2 - \sigma_1^2}{\sigma_1^2} - \frac{(\eta_{2,i,j})^2 - \sigma_2^2}{\sigma_2^2}\Big\} + o_{\mathbb{P}}(1) \rightsquigarrow Z = \frac{Z_1}{\sigma_1^2} - \frac{Z_2}{\sigma_2^2}\end{aligned} \tag{5.10}$$

in $C([0,1])$, and similar arguments as in the proof of Theorem A.1 yield

$$\hat{T}_N^{\lambda} \xrightarrow{\mathcal{D}} Z_{\mathcal{E},\theta} = \max\Big\{\sup_{t\in\mathcal{E}_\zeta^l}\big(-Z(t)\big), \sup_{t\in\mathcal{E}_\zeta^u} Z(t)\Big\}, \tag{5.11}$$

wh ere the statistic $\hat{T}_N^{\lambda}$ is defined by (4.10). Turning to the bootstrap process we have for $l = 1, 2$,

$r = 1, \ldots, R$

$$
\begin{aligned}
C_{l,N}^{\star(r)} &= \frac{1}{N-A} \sum_{i=1}^{A} \sum_{j=1}^{n_i} \big\{ (\hat{\eta}_{l,i,j}^{\star(r)})^2 - \hat{\sigma}_l^2 \big\} \\
&= \frac{1}{N-A} \sum_{i=1}^{A} \sum_{j=1}^{n_i} \big\{ (\hat{\eta}_{l,i,j}^{\star(r)})^2 - (\hat{\eta}_{l,i,j})^2 \big\} + o_{\mathbb{P}}(1) \\
&= \frac{1}{N-A} \sum_{i=1}^{A} \sum_{j=1}^{n_i} (M_{i,j}^{(r)} - 1) \Big(\eta_{l,i,j} - \frac{1}{n_i} \sum_{k=1}^{n_i} \eta_{l,i,k} \Big)^2 + o_{\mathbb{P}}(1) \\
&= \frac{1}{N-A} \sum_{i=1}^{A} \sum_{j=1}^{n_i} (M_{i,j}^{(r)} - 1) (\eta_{l,i,j})^2 + o_{\mathbb{P}}\Big(\frac{1}{\sqrt{N}} \Big) ,
\end{aligned}
$$

where the last estimate follows from similar arguments as those used to derive the second equality in (5.5) and the vector $(M_{1,1}^{(r)}, \ldots, M_{1,n_1}^{(r)}, \ldots, M_{A,1}^{(r)}, \ldots, M_{A,n_A}^{(r)})$ follows a multinomial distribution with parameters N and equal probabilities $1/N$. Then

$$
\begin{aligned}
\sqrt{N}\, \tilde{C}_N^{\star(r)} &= \sqrt{N} \Big(\frac{C_{1,N}^{\star(r)}}{\sigma_1^2} - \frac{C_{2,N}^{\star(r)}}{\sigma_2^2} \Big) \\
&= \frac{\sqrt{N}}{N-A} \sum_{i=1}^{A} \sum_{j=1}^{n_i} (M_{i,j}^{(r)} - 1) \Big\{ \frac{(\eta_{1,i,j})^2}{\sigma_1^2} - \frac{(\eta_{2,i,j})^2}{\sigma_2^2} \Big\} + o_{\mathbb{P}}(1)
\end{aligned}
$$

and this process converges conditionally given $(\eta_{1,i,j} \colon i, j \in \mathbb{N})$ and $(\eta_{2,i,j} \colon i, j \in \mathbb{N})$ to the same limit as the limit in (5.10). Finally it can be shown that

$$
\sqrt{N} \| C_N^{\star(r)} - \tilde{C}_N^{\star(r)} \|_\infty = \sqrt{N} \| C_{1,N}^{\star(r)} (1/\sigma_1^2 - 1/\hat{\sigma}_1^2) - C_{2,N}^{\star(r)} (1/\sigma_2^2 - 1/\hat{\sigma}_2^2) \|_\infty = o_{\mathbb{P}}(1)
$$

and by the continuous mapping theorem and similar arguments as given in the proof of Lemma B.3 of Dette et al. (2020), the bootstrap statistic defined by (4.11) converges conditionally given the data $(X_{lj} \colon j \in \mathbb{N};\ l = 1, 2)$ to the limit in (5.11).

The assertion of Theorem 4.2 now follows by the same arguments given in the proof of Theorem 3.1.