REVSTAT – Statistical Journal
Volume 17, Number 4, October 2019, 523–549

REPEATED MEASURES ANALYSIS FOR FUNCTIONAL DATA USING BOX-TYPE APPROXIMATION — WITH APPLICATIONS

Author: Łukasz Smaga
– Faculty of Mathematics and Computer Science,
Adam Mickiewicz University in Poznań,
Uniwersytetu Poznańskiego 4, 61-614 Poznań, Poland
ls@amu.edu.pl

Received: November 2016 Revised: March 2017 Accepted: August 2017

Abstract:

- The repeated measures analysis for functional data is investigated. In the literature, the test statistic for the two-sample problem when data are from the same subject is considered. Unfortunately, the known permutation and bootstrap approximations for distribution of it may be time-consuming. To avoid this drawback, a Box-type approximation for asymptotic null distribution of that test statistic is proposed. This approximation results in a new testing procedure, which is more efficient from a computational point of view than the known ones. Root-n consistency of the new method is also proved. Via intensive simulation studies, it is found that in terms of size control and power, the new test is comparable with the known tests. An illustrative example of the use of tests in practice is also given.

Key-Words:

- *Box-type approximation; χ^2-type mixture; functional data; repeated measures analysis; two-cumulant approximation; two-sample test.*

AMS Subject Classification:

- 62H15, 62M99.

1. INTRODUCTION

Functional data analysis (FDA) is concerned with data which are viewed as functions defined over some set T. Examples of functional data can be found in several application domains such as meteorology, medicine, economics and many others (for an overview, see Ramsay and Silverman, 2002). Comprehensive surveys about functional data analysis can be found in Ferraty and Vieu (2006), Horváth and Kokoszka (2012), Ramsay *et al.* (2009), Ramsay and Silverman (2002, 2005), Zhang (2013) and in the review papers Cuevas (2014) and Valderrama (2007). Many papers available in the literature are devoted to estimation and classification of functional data, e.g., cluster analysis (Jacques and Preda, 2014; Tokushige *et al.*, 2007; Yamamoto and Terada, 2014), confidence intervals (Lian, 2012), discriminant analysis (Górecki *et al.*, 2014; James and Hastie, 2001; Preda *et al.*, 2007), estimation (Attouch and Belabed, 2014; Chesneau *et al.*, 2013; Cuevas *et al.*, 2006, 2007; Prakasa Rao, 2010), principal component analysis (Berrendero *et al.*, 2011; Boente *et al.*, 2014; Boente and Fraiman, 2000; Jacques and Preda, 2014), variable selection (Gregorutti *et al.*, 2015). Hypothesis testing problems for functional data are also commonly considered, e.g., heteroscedastic ANOVA problem (Cuesta-Albertos and Febrero-Bande, 2010; Zhang, 2013), paired two-sample problem (Martínez-Camblor and Corral, 2011), the one-way ANOVA and MANOVA problem (Abramovich *et al.*, 2004; Cuevas *et al.*, 2004; Horváth and Rice, 2015; Górecki and Smaga, 2015, 2017), testing equality of covariance functions (Zhang, 2013), two-sample Behrens–Fisher problem (Zhang *et al.*, 2010b).

In this paper, the two-sample problem for functional data which are from the same subject (probably submitted to different conditions) is considered. We follow the notation of Martínez-Camblor and Corral (2011). Suppose we have a functional sample consisting of independent trajectories $X_1(t), ..., X_n(t)$ from a stochastic process which may be expressed in the following form

$$X_i(t) = m(t) + \varepsilon_i(t),\ \ t \in [0, 2], \tag{1.1}$$

where $\varepsilon_i(t)$ are random functions with $E(\varepsilon_i(t)) = 0$ and covariance function $\mathbb{C}(s, t)$. Hence, the null hypothesis is of the form

$$H_0\colon m(t) = m(t + 1),\ \ \forall t \in [0, 1]. \tag{1.2}$$

Concerning $t \in [0, 2]$, we ignore (possible) period in which the subject is not monitored.

To illustrate the testing problem described above, we consider the orthosis data. Seven volunteers ($n = 7$) were participated in the experiment. First, they were stepping-in-place without orthosis. Second, they did the same with a spring-loaded orthosis on the right knee. Under each condition, the moment of force at the knee was computed at 256 time points, equally spaced and scaled to the interval $[0, 1]$. So the orthosis data can be represented as curves. We are interested in testing if the mean curves of all volunteers are different under these two conditions (see Figure 1). As the curves obtained without and with orthosis are from the same subjects (volunteers), we have a paired two-sample problem for functional data. The detailed description of the experiment and its analysis are presented in Section 6.

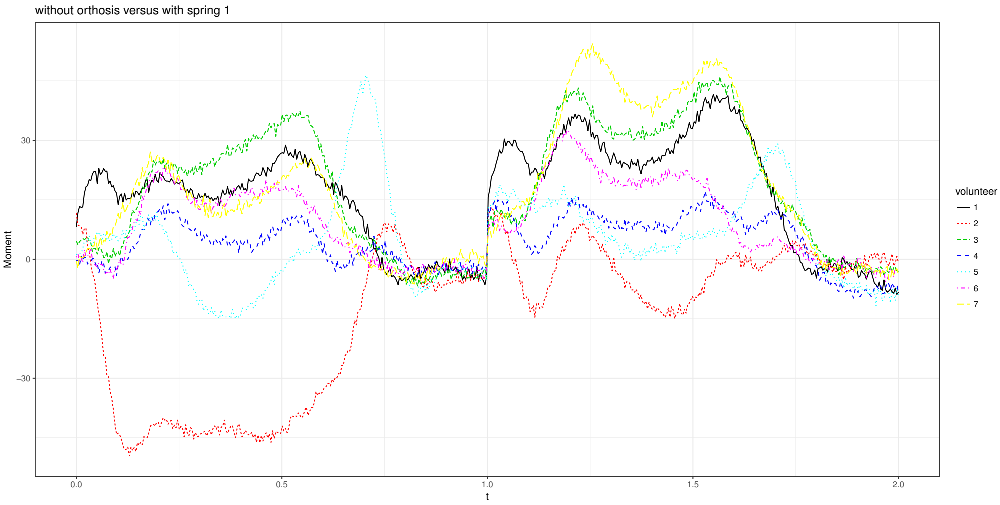

Figure 1: The mean curves of all volunteers of the 10 raw orthosis curves under without orthosis ($t \in [0, 1]$) and with spring 1 ($t \in [1, 2]$) conditions.

For testing (1.2), Martínez-Camblor and Corral (2011) proposed to use the test statistic

$$\mathcal{C}_n = n \int_0^1 \left(\bar{X}(t) - \bar{X}(t+1)\right)^2 \, dt, \tag{1.3}$$

where $\bar{X}(t) = n^{-1} \sum_{i=1}^{n} X_i(t), \; t \in [0, 2]$. This test statistic is based on a simple idea that the null hypothesis should be rejected whenever the "between group variability" measured by the difference between sample means is large enough at a prescribed significance level. As in the standard ANOVA test statistic, appropriate "within group variability" measure may be also contained as denominator in $\mathcal{C}_n$. However, then it seems to be impossible to find the exact sampling distribution of such statistic, even under Gaussianity assumption. Moreover, since Martínez-Camblor and Corral (2011) used an asymptotic test (large sample sizes may be required) and such a denominator tends to some parameter connected with covariance function as $n \to \infty$, the denominator could be replaced by that parameter. Then it could be incorporated to the numerator so that it is only necessary to calculate the asymptotic distribution of the test statistic and replaced by an estimator in that distribution. This reasoning can be used in homoscedastic as well as heteroscedastic case. Bearing in mind this motivation, Martínez-Camblor and Corral (2011) used only the numerator (i.e., $\mathcal{C}_n$), and avoided the homoscedasticity assumption in such a way (see also Cuevas *et al.*, 2004, for similar argumentation).

Martínez-Camblor and Corral (2011) derived a random expression of their test statistic (1.3), and approximated the null distribution by a parametric bootstrap method via re-sampling some Gaussian process involved in the limit random expression of $\mathcal{C}_n$ under the null hypothesis. Moreover, Martínez-Camblor and Corral (2011) considered nonparametric approach and proposed bootstrap and permutation tests. Although these methods work reasonably well in finite samples, they may be time-consuming. In this paper, we present the Box-type approximation (Box, 1954; Brunner *et al.*, 1997; also called two-cumulant approximation, see Zhang, 2013) for the asymptotic distribution of $\mathcal{C}_n$ under the null, and we propose the new test based on this approximation. It is shown to be root-n consistent. The new testing procedure is also much less computationally intensive than the re-sampling and permutation tests of Martínez-Camblor and Corral (2011). Moreover, it is comparable with these tests in terms of size control and power.

This paper is organized as follows. Section 2 presents the Box-type approximation for the asymptotic null distribution of test statistic $\mathcal{C}_n$ and the new test based on this approximation. Its root-n consistency is proved in Section 3. In Section 4, an intensive simulation study providing an idea of the size control and power of the new testing procedure and the tests proposed by Martínez-Camblor and Corral (2011) is given. The comparison of computational time required to perform the considered tests is presented in Section 5. Section 6 contains a real-data example of the use of those tests to the orthosis data. Some concluding remarks are given in Section 7. In the Appendix, proofs of theoretical results, numerical implementation of the new test, R code which performs it and additional simulations are presented.

2. THE TESTING PROCEDURE

In this section, we describe and discuss the new testing procedure for (1.2) which is based on the Box-type approximation for the asymptotic distribution of the test statistic $\mathcal{C}_n$ given by (1.3) under the null.

Let $X_1(t),...,X_n(t)$ be independent trajectories from a stochastic process (with expectation function $m(t)$ and covariance function $\mathbb{C}(s,t)$, $s,t \in [0,2]$) expressed as in (1.1). For theoretical study, we list the following regularity assumptions.

Assumptions:

A1. The mean function $m(t) \in L^2[0,2]$ and $\mathrm{tr}(\mathbb{C}) \stackrel{\text{def}}{=} \int_0^2 \mathbb{C}(t,t)\ dt < \infty$, where $L^2([0,2])$ denotes the set of all square-integrable functions over $[0,2]$.

A2. The subject-effect function $v_1(t) \stackrel{\text{def}}{=} X_1(t) - m(t)$ satisfies

$$E\|v_1\|^4 = E\left(\int_0^2 v_1^2(t)\ dt\right)^2 < \infty.$$

A3. For any $t \in [0,2]$, $\mathbb{C}(t,t) > 0$, and $\max_{t\in[0,2]} \mathbb{C}(t,t) < \infty$.

A4. For any $(s,t) \in [0,2]^2$, $E\left(v_1^2(s)v_1^2(t)\right) < C < \infty$, where C is certain constant independent of any $(s,t) \in [0,2]^2$.

The given assumptions are quite common in functional data analysis literature (see, for instance, Zhang, 2013; Zhang and Liang, 2014). Assumption A1 is regular. It guarantees that as $n \to \infty$, the sample mean function will converge to Gaussian process weakly. Assumptions A2–A4 are additionally imposed to obtain the consistency of estimator of the covariance function. The uniformly boundedness of $E\left(v_1^2(s)v_1^2(t)\right)$ in assumption A4 is satisfied when the subject-effect function $v_1(t)$ is uniformly bounded in probability over $[0,2]$.

Under assumption A1, by (4.7) in Zhang (2013), we have $E\|X_1\|^2 = \|m\|^2 + \mathrm{tr}(\mathbb{C}) < \infty$. Hence, using the central limit theorem for random elements taking values in a Hilbert space (see, for example, Zhang, 2013, p. 91) and the continuous mapping theorem as in the proof of Theorem 1 in Martínez-Camblor and Corral (2011), under the null hypothesis, we obtain $\mathcal{C}_n \xrightarrow{d} \|\xi\|^2$, as $n \to \infty$, where $\xrightarrow{d}$ denotes convergence in distribution, and $\xi(t), t \in [0,1]$ is a Gaussian process with mean zero and covariance function

$$\mathbb{K}(s,t) = \mathbb{C}(s,t) - \mathbb{C}(s,t+1) - \mathbb{C}(s+1,t) + \mathbb{C}(s+1,t+1),\ s,t \in [0,1] \tag{2.1}$$

(see the proof of Theorem 1 in Martínez-Camblor and Corral, 2011, for more details). Under assumptions A1 and A3, we have $\mathrm{tr}(\mathbb{K})$ is finite, where we use the fact $\mathbb{C}(s,t) \leq (\mathbb{C}(s,s)\mathbb{C}(t,t))^{1/2} \leq \max_{t\in[0,2]} \mathbb{C}(t,t) < \infty$. Thus, Theorem 4.2 in Zhang (2013) implies $\|\xi\|^2$ has the same distribution as $\sum_{k\in\mathbb{N}} \lambda_k A_k$, where A_k, $k = 1,2,...$, is a sequence of independent random variables following a central chi-squared distribution with one degree of freedom, and λ_k, $k = 1,2,...$, is the non-negative sequence, satisfying $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r \geq ... \geq 0$ and $\sum_{k\in\mathbb{N}} \lambda_k^2 < \infty$, of the eigenvalues of $\mathbb{K}(s,t)$ given by (2.1). Since $\mathcal{C}_n \xrightarrow{d} \|\xi\|^2$, as $n \to \infty$, we conclude that

$$\mathcal{C}_n \xrightarrow{d} \mathcal{C}_0^* = \sum_{k\in\mathbb{N}} \lambda_k A_k \tag{2.2}$$

under the null and assumptions A1 and A3. Hence, the test statistic $\mathcal{C}_n$ converges in distribution to a central χ^2-type mixture (see Zhang, 2005), under the null and assumptions A1 and A3. On the basis of (2.2), the asymptotic null distribution of $\mathcal{C}_n$ is known except the unknown eigenvalues λ_k, $k = 1, 2, ...,$ of $\mathbb{K}(s,t)$. These unknown eigenvalues can be estimated by the eigenvalues $\hat{\lambda}_k$, $k = 1, 2, ...,$ of the following estimator of $\mathbb{K}(s,t)$:

$$(2.3)\quad \hat{\mathbb{K}}(s,t) = \hat{\mathbb{C}}(s,t) - \hat{\mathbb{C}}(s,t+1) - \hat{\mathbb{C}}(s+1,t) + \hat{\mathbb{C}}(s+1,t+1),\ \ s,t \in [0,1],$$

where $\hat{\mathbb{C}}(s,t) = (n-1)^{-1}\sum_{i=1}^{n}(X_i(s) - \bar{X}(s))(X_i(t) - \bar{X}(t)),\ \ s,t \in [0,2]$ is the unbiased estimator of $\mathbb{C}(s,t)$ (see Zhang, 2013, p. 108). Moreover, it is often sufficient to use only the positive eigenvalues of $\hat{\mathbb{K}}(s,t)$. With the sample size n growing to infinity, the estimator $\hat{\mathbb{K}}(s,t)$ is consistent in the sense of the following lemma. Let $\xrightarrow{P}$ denote convergence in probability.

Lemma 2.1. *Under the model* (1.1) *and assumptions A1–A4, we have* $\hat{\mathbb{K}}(s,t) \xrightarrow{P} \mathbb{K}(s,t)$ *uniformly over* $[0,1]^2$*, as* $n \to \infty$.

We now apply Box-type approximation (Box, 1954; Brunner *et al.*, 1997) for approximating the asymptotic null distribution of $\mathcal{C}_n$. This method is also known as two-cumulant approximation (see Zhang, 2013). It is an example of the approximation methods using cumulants, which are often considered in functional data analysis (see, for example, Górecki and Smaga, 2015; Zhang, 2013; Zhang and Liang, 2014; Zhang *et al.*, 2010b), so we also may name it "two-cumulant approximation". The key idea of this method is to approximate the distribution of $\mathcal{C}_0^*$ by that of a random variable of the form $\beta\chi_d^2$, where the parameters β and d are determined by matching the first two cumulants or moments of $\mathcal{C}_0^*$ and $\beta\chi_d^2$. By the results of Zhang (2013, Sections 4.3 and 4.5), we have

$$(2.4)\qquad \beta = \frac{\operatorname{tr}(\mathbb{K}^{\otimes 2})}{\operatorname{tr}(\mathbb{K})}, \qquad d = \frac{\operatorname{tr}^2(\mathbb{K})}{\operatorname{tr}(\mathbb{K}^{\otimes 2})},$$

where $\operatorname{tr}(\mathbb{K}) = \int_0^1 \mathbb{K}(t,t)\,dt$ and $\mathbb{K}^{\otimes 2} \overset{\text{def}}{=} \int_0^1 \mathbb{K}(s,u)\mathbb{K}(u,t)\,du$. The approximation of the distribution of $\mathcal{C}_0^*$ by that of $\beta\chi_d^2$ seems to be sensible, since $\mathcal{C}_0^*$ is a χ^2-type mixture which is nonnegative and generally skewed, and so $\beta\chi_d^2$ is. Thus, $\mathcal{C}_0^*$ and $\beta\chi_d^2$ with β and d as in (2.4) have the same range, mean and variance and similar shapes. However, the distributions of these random variables are usually not the same. Moreover, the conditional distributions of the parametric and nonparametric bootstrap and permutation statistics of Martínez-Camblor and Corral (2011) can be different of the distribution of $\beta\chi_d^2$. Fortunately, these distributions are very similar to each other, and the distribution of $\beta\chi_d^2$ can have flexible shapes and be adaptive to different shapes of the underlying null distribution of $\mathcal{C}_n$, which is confirmed by simulation studies of Section 4. From those simulation studies, we can observe that both the previous and new approximations give very similar and satisfactory results for small and moderate sample sizes. The same holds for large samples. For instance, when $n = 2000$, the empirical sizes of the parametric and nonparametric bootstrap, permutation and new testing procedures were equal to 5.2%, 4.8%, 5.2%, 4.8%, respectively, and the empirical power of all tests was equal to 100%. These results suggest that the type I error rate (resp. power) of each test tends to the nominal significance level or to value close to it (resp. one) as $n \to \infty$.

The natural estimators of β and d are obtained by replacing the covariance function $\mathbb{K}(s,t)$ in (2.4) by its estimator $\hat{\mathbb{K}}(s,t)$ given by (2.3), i.e.,

$$\hat{\beta} = \frac{\operatorname{tr}(\hat{\mathbb{K}}^{\otimes 2})}{\operatorname{tr}(\hat{\mathbb{K}})}, \qquad \hat{d} = \frac{\operatorname{tr}^2(\hat{\mathbb{K}})}{\operatorname{tr}(\hat{\mathbb{K}}^{\otimes 2})}. \tag{2.5}$$

Therefore, under the null, $\mathcal{C}_n \sim \hat{\beta}\chi^2_{\hat{d}}$ approximately, and hence the new test (the BT test) for (1.2) is conducted by computing the p-value of the form

$$P(\chi^2_{\hat{d}} > \mathcal{C}_n/\hat{\beta}), \tag{2.6}$$

or for given significance level α, the estimated critical value of $\mathcal{C}_n$ given by

$$\hat{\mathcal{C}}_{n,\alpha} = \hat{\beta}\chi^2_{\hat{d},\alpha}, \tag{2.7}$$

where $\chi^2_{r,\alpha}$ denotes the upper 100α percentile of χ^2_r. The critical region of the new testing procedure is of the form $\{\mathcal{C}_n > \hat{\beta}\chi^2_{\hat{d},\alpha}\}$. In the following theorem, we show that the estimated critical value $\hat{\mathcal{C}}_{n,\alpha}$ tends to theoretical critical value $\mathcal{C}_{0,\alpha} = \beta\chi^2_{d,\alpha}$, as $n \to \infty$. The consistency of the estimators $\hat{\beta}$ and $\hat{d}$ is also proved there.

Theorem 2.1. *Under the assumptions of Lemma 2.1, as $n \to \infty$, we have $\hat{\beta} \xrightarrow{P} \beta$ and $\hat{d} \xrightarrow{P} d$. Moreover, we have $\hat{\mathcal{C}}_{n,\alpha} \xrightarrow{P} \mathcal{C}_{0,\alpha} = \beta\chi^2_{d,\alpha}$, as $n \to \infty$.*

Numerical implementation of the BT test is described in the Appendix. This testing procedure is very easy to implement in the R language (R Core Team, 2015). In the Appendix, we also present and describe the R code which performs the new test.

3. ASYMPTOTIC POWER UNDER LOCAL ALTERNATIVES

In this section, we investigate the asymptotic power of the BT test under two kinds of local alternatives. Power of tests under similar types of alternatives was studied in the literature concerning the functional data analysis (see, for example, Zhang *et al.*, 2010a, Zhang and Liang, 2014). The formulas for the asymptotic powers of the BT test are given in the proofs of Theorems 3.1 and 3.2.

First, we consider the local alternatives of the form $H_{1n}^{(1)}: m(t) - m(t+1) = n^{-\tau/2} d(t), t \in [0,1]$, where $\tau \in [0,1)$ is fixed and $d(t)$ is any fixed real function such that $\|d\| \in (0,\infty)$. So, we study the power behavior when the alternatives tend to the null hypothesis (1.2) with a rate slightly slower than $n^{-1/2}$. In the following result, we establish the asymptotic power of the BT test tends to one, as $n \to \infty$, under $H_{1n}^{(1)}$ and under gaussianity assumption of processes $X_i(t), i = 1, ..., n$ in model (1.1).

Theorem 3.1. *Under model* (1.1), *where* $X_i(t), i = 1, ..., n$ *are Gaussian processes, assumptions A1–A4 and the local alternatives* $H_{1n}^{(1)}$, $\tau \in [0,1)$, *the asymptotic power of the BT test tends to* 1 *as* $n \to \infty$.

We now consider the local alternatives, which tend to the null hypothesis (1.2) with the root-n rate, i.e., $H_{1n}^{(2)}: m(t) - m(t+1) = n^{-1/2} d(t), t \in [0,1]$, where $d(t)$ is any fixed real function such that $\|d\| \in (0,\infty)$. Here, we do not assume gaussianity of the observations, but the asymptotic power of the BT test tending to 1 is obtained when the information provided by $d(t)$ diverges to infinity. This is presented in the following theorem.

Theorem 3.2. *Under model* (1.1), *assumptions A1–A4 and the local alternatives* $H_{1n}^{(2)}$, *as* $n \to \infty$, *the asymptotic power of the BT test tends to* 1 *as* $\|d\| \to \infty$.

Theorems 3.1 and 3.2 indicate that the BT test can detect the local alternatives $H_{1n}^{(1)}$ and $H_{1n}^{(2)}$ with probability tending to one under the assumptions given above. By the definition of Zhang and Liang (2014), we obtain that the BT test is root-n consistent.

4. SIMULATIONS

Simulations are conducted to compare the empirical sizes (type I error rates) and powers of the BT test with those of Martínez-Camblor and Corral (2011). As we mentioned, Martínez-Camblor and Corral (2011) proposed three approximation methods for the null distribution of $\mathcal{C}_n$ based on the asymptotic distribution (the A test), on bootstrap (the B test), and on permutation (the P test). Additional simulations considering different dependency structure than that in this section are given in the Appendix. All simulations were conducted with the help of the R computing environment (R Core Team, 2015).

4.1. Description of the simulation experiments

To be consistent with the results of Martínez-Camblor and Corral (2011) for the A, B and P tests, we present similar simulation experiments to those in that paper. We generated $X_i(t) = m_1(t) + \varepsilon_{i1}(t)$ and $X_i(t+1) = m_2(t) + \varepsilon_{i2}(t)$ for $t \in [0,1]$, $i = 1, ..., n$, where $m_j(t)$ and $\varepsilon_{ij}(t)$ are described below. Sample sizes $n = 25, 35, 50$ are considered. Let

$$
\begin{aligned}
&m_{0,1}(t) = \sqrt{6t/\pi}\exp(-6t)I_{[0,1]}(t), \quad m_{1,1}(t) = \sqrt{13t/(2\pi)}\exp(-13t/2)I_{[0,1]}(t),\\
&m_{2,1}(t) = \sqrt{11t/(2\pi)}\exp(-11t/2)I_{[0,1]}(t), \quad m_{3,1}(t) = \sqrt{5}t^{2/3}\exp(-7t)I_{[0,1]}(t),\\
&m_{0,2}(t) = (\sin(2\pi t^2))^5 I_{[0,1]}(t), \quad m_{1,2}(t) = (\sin(2\pi t^2))^3 I_{[0,1]}(t),\\
&m_{2,2}(t) = (\sin(2\pi t^2))^7 I_{[0,1]}(t), \quad m_{3,2}(t) = (\sin(2\pi t^{(9/5)}))^3 I_{[0,1]}(t).
\end{aligned}
$$

Figure 2 depicts the shapes of $m_{i,j}(t)$. Because of the choice of $m_i(t), i = 1, 2$, we considered eight models. In models M0–M3, $m_1 = m_{0,1}$ and $m_2 = m_{j,1}$, $j = 0, ..., 3$ respectively, and in models M4–M7, $m_1 = m_{0,2}$ and $m_2 = m_{j,2}$, $j = 0, ..., 3$ respectively.

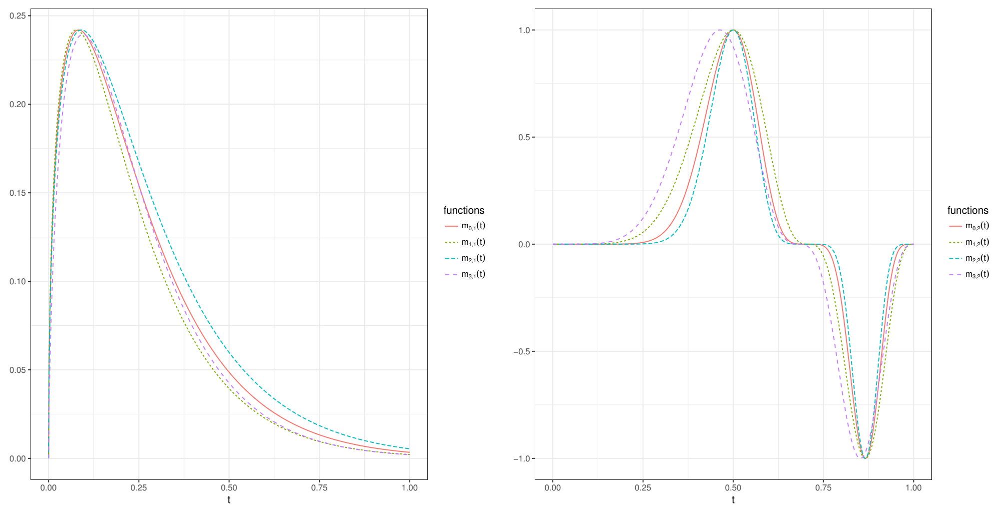

Figure 2: The shapes of functions $m_{i,j}(t), t \in [0,1]$ used in simulations of Section 4.

Three different types of errors were considered. In the *normal* case, $\varepsilon_{i1}(t) = \xi B_{i1}(t)$ and $\varepsilon_{i2}(t) = \rho\varepsilon_{i1}(t) + \xi\sqrt{1-\rho^2}B_{i2}(t)$, where $\rho = 0, 0.25, 0.5$, B_{i1} and B_{i2} are two independent standard Brownian Bridges, and $\xi = 0.05$ for models M0–M3 and $\xi = 0.5$ for the remaining. In the *lognormal* (resp. *mixed*) case, the error functions are $\exp(\varepsilon_{ij}(t))$, $j = 1, 2$ (resp. $\varepsilon_{i1}(t)$ and $\exp(\varepsilon_{i2}(t))$), where $\varepsilon_{ij}(t)$ are as above. The errors functions $\exp(\varepsilon_{ij}(t))$ are adequately centered.

In practice, the functional data are not usually continuously observed. The points, at which the functional data are observed, are called the design time points. So, the processes $X_i(t), X_i(t+1), t \in [0,1]$ were generated in discretized versions $X_i(t_r)$, $X_i(t_r+1)$, for $r = 1, ..., I$ and for $I = 26, 101, 251$, where the values t_r were chosen equispaced in the interval $[0,1]$.

Under various parameter configurations, the empirical sizes and powers (as percentages) of the tests were calculated at the nominal significance level $\alpha = 5\%$ and based on 1000 replications. In Tables 1–7, the results for models M0–M6 are displayed. For model M7, the empirical powers were always 100%. The empirical power in the omitted rows in these tables is always 100%. Similarly as in Martínez-Camblor and Corral (2011), the p-values of the A, B and P tests were estimated from 1000 replications.

Table 1: Empirical sizes (as percentages) of all tests obtained in model M0. The column "R" refers to different residual types (N – normal, L – lognormal, M – mixed).

R	n	I	26				101				251			
		ρ	A	B	P	BT	A	B	P	BT	A	B	P	BT
N	25	0.00	$\underline{\mathbf{6.9}}$	**6.8**	6.0	**6.5**	6.1	6.2	4.9	6.1	5.3	5.4	5.0	5.3
		0.25	**6.8**	$\underline{\mathbf{7.1}}$	6.4	**6.8**	**6.8**	$\underline{\mathbf{6.9}}$	5.9	$\underline{\mathbf{7.1}}$	5.2	5.6	5.1	5.3
		0.50	$\underline{\mathbf{7.2}}$	$\underline{\mathbf{7.6}}$	6.4	$\underline{\mathbf{7.1}}$	$\underline{\mathbf{7.2}}$	$\underline{\mathbf{7.0}}$	6.2	**6.7**	5.1	5.7	4.6	5.2
	35	0.00	5.5	6.1	5.7	6.1	4.8	4.5	4.0	4.7	4.3	4.5	4.1	4.3
		0.25	5.4	6.1	5.4	5.5	5.4	5.6	5.1	5.1	4.2	4.4	3.6	4.2
		0.50	5.9	6.0	5.1	5.9	5.4	5.6	5.2	5.7	4.4	4.3	3.7	4.2
	50	0.00	5.4	5.2	4.8	5.2	6.0	6.1	5.7	5.8	**6.5**	5.8	5.7	6.1
		0.25	5.2	5.7	5.1	5.0	6.3	**6.6**	**6.5**	6.4	6.3	5.7	5.9	5.7
		0.50	5.3	5.4	5.0	5.0	$\underline{\mathbf{7.2}}$	$\underline{\mathbf{7.1}}$	6.3	$\underline{\mathbf{7.0}}$	5.4	5.1	4.9	5.2
L	25	0.00	**6.8**	**6.7**	**6.7**	6.4	5.4	5.7	5.7	5.5	5.1	4.8	5.5	5.0
		0.25	$\underline{\mathbf{7.2}}$	$\underline{\mathbf{7.4}}$	$\underline{\mathbf{7.2}}$	$\underline{\mathbf{7.1}}$	5.1	5.3	5.1	4.9	5.7	5.8	5.7	5.6
		0.50	$\underline{\mathbf{7.2}}$	$\underline{\mathbf{8.0}}$	$\underline{\mathbf{7.9}}$	$\underline{\mathbf{7.6}}$	5.2	5.3	5.3	5.4	5.6	5.9	6.0	5.9
	35	0.00	5.4	5.3	4.8	5.3	5.5	5.0	5.2	5.2	4.8	5.0	4.8	4.8
		0.25	4.7	4.3	4.6	4.5	5.2	5.2	5.0	5.2	5.0	4.8	4.8	4.7
		0.50	4.7	4.9	4.7	4.6	5.9	5.9	5.7	5.7	4.7	4.6	4.5	4.6
	50	0.00	5.0	5.3	4.9	5.1	5.3	5.3	4.9	4.8	4.9	4.9	5.2	5.3
		0.25	5.8	5.7	5.7	5.6	5.0	5.1	4.5	4.7	4.7	4.9	4.6	4.8
		0.50	5.5	5.5	5.7	5.8	5.0	5.3	5.1	5.1	4.4	4.6	4.4	4.6
M	25	0.00	5.1	5.6	4.8	5.2	5.6	6.0	5.6	5.6	5.8	5.6	5.8	5.6
		0.25	5.7	5.3	5.2	5.0	5.2	5.6	5.4	5.6	6.2	6.0	5.8	5.7
		0.50	5.8	6.1	5.6	5.9	5.9	5.9	5.6	6.1	6.1	6.1	5.3	6.0
	35	0.00	5.3	5.5	5.2	5.3	4.9	4.7	4.5	4.7	4.5	4.8	4.7	4.8
		0.25	4.8	5.0	4.7	4.9	4.9	5.3	4.8	5.0	4.8	5.0	4.9	4.6
		0.50	4.9	5.3	4.7	4.8	5.1	5.2	4.5	5.3	4.6	5.7	4.4	4.6
	50	0.00	5.1	5.1	5.0	5.1	5.5	5.3	5.4	5.2	5.8	5.7	5.6	5.8
		0.25	4.9	4.9	4.9	4.7	5.3	5.5	5.0	5.0	6.2	6.1	5.5	6.1
		0.50	4.5	4.8	4.5	4.6	5.7	5.8	5.5	5.5	5.9	6.4	5.6	6.0

Table 2: Empirical powers (as percentages) of all tests obtained in model M1. The column "R" refers to different residual types (N – normal, L – lognormal, M – mixed). The empirical power in the omitted rows is always 100%.

R	n	I	26				101				251			
		ρ	A	B	P	BT	A	B	P	BT	A	B	P	BT
N	25	0.00	40.5	40.4	38.2	39.5	39.8	39.9	36.6	39.5	38.5	38.6	36.7	38.8
		0.25	49.2	49.6	46.8	48.9	51.4	51.5	48.3	51.0	50.2	49.8	47.4	49.5
		0.50	68.5	68.0	66.0	68.0	70.2	71.0	68.7	69.8	69.3	69.6	66.6	69.1
	35	0.00	53.1	53.8	51.6	52.7	52.1	52.0	50.2	52.3	54.4	54.5	53.1	54.5
		0.25	65.7	66.5	65.2	65.9	64.7	65.2	63.2	65.3	66.9	66.8	65.7	66.3
		0.50	85.4	85.7	84.6	84.9	81.6	81.3	80.1	81.1	85.0	85.0	84.2	84.8
	50	0.00	69.2	68.9	68.2	68.8	68.7	69.3	68.3	68.6	69.8	68.6	68.4	69.1
		0.25	81.8	81.1	81.2	81.1	81.0	81.4	80.9	81.1	82.2	82.9	81.7	82.4
		0.50	93.9	93.9	93.7	93.9	94.5	94.5	94.1	94.5	94.3	94.0	94.0	94.0
L	25	0.00	98.8	98.1	99.0	98.6	99.0	99.2	99.3	98.9	99.2	99.3	99.6	99.4
M	25	0.00	62.6	62.8	61.2	61.8	61.6	61.9	61.3	61.7	61.3	62.0	60.6	61.6
		0.25	67.2	67.6	65.7	67.1	67.3	68.6	66.4	67.5	67.6	67.6	66.8	67.3
		0.50	73.2	74.0	72.1	73.6	74.1	74.6	73.2	74.7	75.0	75.5	73.3	74.3
	35	0.00	76.8	77.5	76.5	76.9	77.2	76.5	76.9	76.7	78.6	79.8	78.4	78.8
		0.25	82.7	82.7	82.0	82.8	82.6	83.0	81.6	82.1	84.7	84.5	83.5	84.8
		0.50	87.5	88.1	87.4	87.1	87.3	87.2	86.5	87.0	88.5	88.1	88.2	88.3
	50	0.00	89.9	90.7	90.0	90.0	90.9	91.0	91.2	91.4	91.0	91.1	90.8	91.3
		0.25	93.7	93.8	92.9	93.6	94.3	94.0	93.9	93.9	94.4	94.3	94.5	94.5
		0.50	96.1	96.0	95.8	96.1	97.0	97.1	96.9	97.1	97.3	97.5	97.2	97.4

Table 3: Empirical powers (as percentages) of all tests obtained in model M2. The column "R" refers to different residual types (N – normal, L – lognormal, M – mixed). The empirical power in the omitted rows is always 100%.

R	n	I	26				101				251			
		ρ	A	B	P	BT	A	B	P	BT	A	B	P	BT
N	25	0.00	49.3	49.0	46.9	48.5	50.0	50.1	48.0	49.6	49.1	49.8	46.5	48.7
		0.25	60.7	62.1	58.8	60.6	61.9	62.5	59.5	61.7	63.0	62.8	59.2	62.2
		0.50	78.6	79.0	77.0	78.6	79.5	79.4	78.5	79.3	79.4	79.4	77.7	79.1
	35	0.00	62.4	61.6	60.3	61.2	65.4	65.2	64.1	64.8	63.1	62.2	60.7	62.1
		0.25	76.8	77.3	75.9	77.2	78.1	78.1	77.3	78.0	75.4	75.4	73.5	74.5
		0.50	90.9	91.4	90.3	91.0	91.5	91.6	90.7	91.5	89.8	89.9	89.6	90.1
	50	0.00	79.9	80.2	79.2	79.9	79.9	80.1	78.7	79.8	79.7	80.5	79.2	79.6
		0.25	90.7	90.4	89.6	90.3	90.1	90.0	89.5	90.0	90.2	89.9	89.5	89.7
		0.50	97.9	97.8	97.6	97.8	98.5	98.2	98.3	98.3	98.1	98.3	97.8	98.2
L	25	0.00	100	100	100	99.9	99.9	99.9	100	100	100	99.9	100	100
M	25	0.00	74.1	74.2	72.8	73.5	75.5	75.4	74.4	74.8	75.4	76.0	74.3	75.3
		0.25	79.9	79.8	79.2	79.8	81.7	81.5	80.5	81.1	80.9	81.6	81.2	81.1
		0.50	85.7	85.6	84.0	85.5	86.8	87.5	85.9	87.2	85.7	85.8	85.1	85.7
	35	0.00	87.5	87.8	87.5	87.4	89.3	89.7	89.2	89.3	87.9	88.0	87.2	88.0
		0.25	92.7	92.5	92.2	92.4	93.0	92.4	92.3	92.9	92.0	91.8	91.8	91.7
		0.50	95.9	96.0	95.7	95.9	95.9	95.8	95.8	95.7	95.2	94.8	94.6	95.0
	50	0.00	97.8	97.6	97.8	97.7	97.0	96.9	97.0	96.7	97.6	97.7	97.7	97.5
		0.25	98.9	98.8	98.8	98.9	98.5	98.5	98.6	98.5	98.7	98.6	98.7	98.6
		0.50	99.4	99.2	99.2	99.5	99.3	99.4	99.4	99.3	99.4	99.4	99.4	99.5

Table 4: Empirical powers (as percentages) of all tests obtained in model M3. The column "R" refers to different residual types (N – normal, L – lognormal, M – mixed). The empirical power in the omitted rows is always 100%.

R	n	I	26				101				251			
		ρ	A	B	P	BT	A	B	P	BT	A	B	P	BT
N	25	0.00	19.3	19.6	18.0	18.4	31.4	31.6	30.9	30.9	31.4	32.7	31.7	31.7
		0.25	25.1	25.2	23.6	24.3	46.9	46.2	47.3	45.3	46.3	46.3	47.0	44.5
		0.50	39.6	39.6	38.9	39.2	76.2	75.9	78.2	75.4	80.0	81.7	81.9	80.5
	35	0.00	26.0	25.7	26.0	26.3	46.6	45.9	47.3	46.1	50.5	50.7	50.1	49.6
		0.25	37.1	37.7	36.8	36.9	67.2	68.3	69.7	67.8	73.2	73.7	74.6	73.3
		0.50	60.3	60.4	60.4	59.7	96.2	95.5	96.2	96.0	97.7	97.5	98.2	97.7
	50	0.00	40.3	41.2	40.1	39.8	77.5	77.8	78.5	77.2	82.9	82.9	83.7	82.5
		0.25	54.1	53.9	54.5	53.3	95.0	94.9	96.1	95.1	97.6	97.7	97.4	97.7
		0.50	84.6	84.2	84.6	83.4	100	100	100	100	100	100	100	100
L	25	0.00	70.4	71.7	75.6	71.5	97.9	97.9	98.8	98.1	99.2	99.1	99.4	99.3
		0.25	91.5	91.9	94.4	91.1	99.8	99.8	100	99.7	100	100	100	100
		0.50	99.9	99.9	99.9	99.8	100	100	100	100	100	100	100	100
	35	0.00	93.0	92.7	94.7	92.6	100	100	100	100	100	100	100	100
		0.25	99.9	99.9	100	99.7	100	100	100	100	100	100	100	100
	50	0.00	100	100	99.9	100	100	100	100	100	100	100	100	100
M	25	0.00	33.5	34.1	34.6	32.9	62.8	62.6	64.8	62.2	66.6	67.3	69.9	67.0
		0.25	39.6	40.4	40.6	39.4	71.1	71.5	75.1	70.6	77.1	77.6	79.4	76.6
		0.50	46.8	46.9	47.4	45.6	79.1	80.0	82.0	79.3	84.7	84.4	85.5	84.2
	35	0.00	48.6	48.7	50.0	48.0	87.2	87.0	89.0	87.6	90.7	90.6	92.8	90.6
		0.25	57.2	56.8	58.3	56.9	92.8	93.0	93.5	92.6	95.3	95.6	96.2	95.3
		0.50	65.3	65.5	66.1	65.3	96.0	96.1	96.6	96.1	97.9	97.8	98.1	98.0
	50	0.00	75.0	75.1	75.8	74.8	99.4	99.6	99.6	99.6	99.8	99.5	99.8	99.9
		0.25	83.0	82.8	83.1	81.8	99.8	99.8	100	99.8	100	100	99.9	100
		0.50	89.4	89.1	89.3	88.8	99.9	99.9	100	100	100	100	100	100

4.2. Results

In this subsection, we describe the simulation results for the new method and the tests of Martínez-Camblor and Corral (2011).

Tables 1 and 5 display the empirical sizes of the tests obtained in models M0 and M4. Based on the binomial proportion confidence interval, for the nominal level $\alpha = 5\%$, the empirical size over the 1000 independent replications should belong to the interval $[3.6\%, 6.4\%]$ (resp. $[3.2\%, 6.8\%]$) with probability 95% (resp. 99%). Therefore in Tables 1 and 5, when the rejection proportions are outside the 95% significance limits, they are displayed in bold, and when they are outside the 99% significance limits they are underlined. The results for the BT test and the tests proposed in Martínez-Camblor and Corral (2011) are generally quite satisfactory, and the nominal level is well respected in most cases by the tests. Their empirical sizes are rarely larger than the upper endpoint of the 95% confidence interval, and they are not less than lower endpoint of that interval.

Under normal and mixed cases, the B test is the most liberal of all the tests, and it is slightly more liberal than the A and BT tests, which are more liberal than the P test. Nevertheless, the P test is not conservative. Under lognormal case, the empirical sizes do not express such a tendency in general. In model M0 and normal case, the empirical sizes of all tests decrease when I increases for $n = 25, 35$, and they increase when $n = 50$. In the other cases of model M0 and in model M4, this observation is not true generally, and the behavior of the empirical sizes is more complicated when I increases. Summarising, the new test respects the nominal level a bit better than the A and B tests and may be more liberal than the P test.

Table 5: Empirical sizes (as percentages) of all tests obtained in model M4. The column "R" refers to different residual types (N – normal, L – lognormal, M – mixed).

R	n	I	26				101				251			
		ρ	A	B	P	BT	A	B	P	BT	A	B	P	BT
		0.00	5.3	5.2	4.7	5.3	5.7	6.1	5.4	5.5	5.1	5.0	4.2	4.6
	25	0.25	5.4	5.4	5.0	5.3	5.2	5.4	5.1	5.4	5.4	5.4	4.3	5.4
		0.50	5.2	5.1	4.6	5.2	5.0	5.1	4.5	4.9	5.7	5.8	4.7	5.6
		0.00	4.9	5.0	4.3	4.7	**6.8**	**6.7**	6.4	**6.7**	4.7	5.0	4.5	4.8
N	35	0.25	5.5	5.7	4.9	5.7	**6.8**	**7.0**	**6.5**	**6.8**	4.9	4.8	4.4	4.9
		0.50	5.9	6.3	5.6	6.3	**6.7**	**6.6**	6.3	6.4	4.9	5.0	5.1	5.2
		0.00	6.3	6.1	5.7	6.0	5.8	5.7	5.5	5.9	4.9	5.1	4.9	4.9
	50	0.25	6.3	6.4	5.9	6.4	5.3	5.5	4.8	5.2	5.3	5.5	5.0	5.4
		0.50	5.9	6.4	5.4	6.2	5.3	5.4	5.0	5.2	5.5	5.8	5.2	5.6
		0.00	4.0	4.4	4.4	4.4	5.6	5.6	5.5	5.6	5.4	5.3	5.2	5.2
	25	0.25	3.9	4.0	4.0	4.0	6.0	6.0	6.2	5.9	5.6	5.5	5.4	5.4
		0.50	3.7	4.1	4.1	3.8	5.9	6.3	6.4	6.1	4.8	5.0	5.1	4.8
		0.00	4.4	4.9	4.9	4.9	4.6	4.7	4.9	4.8	5.4	5.7	5.2	5.3
L	35	0.25	4.2	4.1	4.5	4.2	4.2	4.4	4.8	4.5	5.3	5.3	5.5	5.1
		0.50	4.4	4.6	4.8	4.5	4.6	4.9	5.6	5.2	5.4	5.2	5.1	5.2
		0.00	5.0	5.2	5.2	5.0	4.9	5.4	5.1	5.2	4.8	5.5	5.1	5.2
	50	0.25	5.4	5.2	5.7	5.6	5.3	5.2	5.0	5.0	4.9	5.9	5.5	5.4
		0.50	5.8	5.7	5.7	5.8	6.2	5.8	5.6	5.7	5.3	5.7	5.6	5.4
		0.00	5.6	5.1	5.3	5.4	5.7	6.0	5.9	5.6	6.3	6.2	6.1	6.2
	25	0.25	5.5	5.3	5.1	5.4	5.7	6.4	5.4	5.9	6.2	6.2	6.1	6.3
		0.50	5.6	5.5	5.3	5.4	**6.6**	**6.8**	6.2	**6.6**	**6.6**	6.2	6.1	5.9
		0.00	5.4	5.8	5.5	5.5	5.8	5.9	5.7	5.8	4.7	4.8	4.6	4.6
M	35	0.25	5.2	5.4	5.1	5.1	5.9	5.8	5.9	5.9	5.1	5.4	4.6	5.1
		0.50	5.8	5.6	4.9	5.1	5.9	6.1	5.6	5.8	5.4	5.3	5.2	5.1
		0.00	**7.8**	**8.0**	**7.5**	**7.5**	6.4	**7.0**	**6.6**	**6.7**	5.7	6.1	**6.5**	5.9
	50	0.25	**8.6**	**8.5**	**7.9**	**8.6**	6.4	**6.6**	6.4	6.4	6.0	**6.5**	**6.5**	6.2
		0.50	**8.5**	**8.5**	**7.9**	**8.5**	**6.6**	**6.7**	6.1	**6.6**	6.2	**6.7**	**6.7**	**6.5**

Table 6: Empirical powers (as percentages) of all tests obtained in model M5. The column "R" refers to different residual types (N – normal, L – lognormal, M – mixed). The empirical power in the omitted rows is always 100%.

R	n	I	26				101				251			
		ρ	A	B	P	BT	A	B	P	BT	A	B	P	BT
N	25	0.00	71.9	71.9	72.0	71.5	74.3	74.2	73.7	73.3	72.0	73.2	73.9	72.2
		0.25	89.4	89.2	89.7	89.0	90.5	89.5	90.5	89.6	87.3	88.4	89.0	87.5
		0.50	99.3	99.2	99.6	99.5	99.2	99.0	99.4	99.2	99.3	99.2	99.4	99.1
	35	0.00	91.4	91.8	91.5	91.4	92.2	93.6	93.5	92.8	92.4	92.4	92.5	91.8
		0.25	99.1	98.9	99.2	99.1	99.1	99.1	99.4	99.1	99.1	99.1	99.5	99.3
	50	0.00	99.4	99.6	99.8	99.7	99.7	99.6	99.7	99.7	99.6	99.8	99.8	99.8
L	25	0.00	99.4	99.2	99.3	98.9	99.6	99.6	99.7	99.5	99.3	99.3	99.6	99.4
		0.25	99.9	99.9	99.9	99.9	100	100	100	100	100	100	100	100
	35	0.00	99.9	99.9	99.9	99.9	100	100	100	100	100	100	100	100
M	25	0.00	95.8	96.6	97.3	96.4	97.4	97.7	98.1	97.9	96.5	96.2	96.6	96.3
		0.25	98.8	98.6	98.7	98.7	99.1	99.2	99.3	99.2	98.5	98.8	99.0	98.7
		0.50	99.8	99.8	99.8	99.9	99.8	99.9	99.8	99.8	99.8	99.8	99.7	99.8
	35	0.00	99.9	99.9	99.9	99.9	99.9	99.8	100	100	99.9	99.9	99.9	99.9

Table 7: Empirical powers (as percentages) of all tests obtained in model M6. The column "R" refers to different residual types (N – normal, L – lognormal, M – mixed). The empirical power in the omitted rows is always 100%.

R	n	I	26				101				251			
		ρ	A	B	P	BT	A	B	P	BT	A	B	P	BT
N	25	0.00	23.1	23.4	22.6	22.5	23.1	23.0	22.8	23.2	23.3	22.8	22.2	23.0
		0.25	31.7	32.1	30.8	30.8	32.9	32.3	32.0	31.4	30.6	30.6	30.5	30.3
		0.50	51.9	51.9	51.4	50.3	53.8	54.0	55.1	52.8	51.3	52.5	52.1	50.7
	35	0.00	32.3	33.0	32.2	32.1	35.5	35.3	35.0	34.5	33.3	33.0	32.0	33.2
		0.25	47.8	47.8	48.6	47.5	49.2	49.8	49.9	49.6	47.4	48.9	48.9	48.1
		0.50	76.3	76.6	76.7	75.4	77.3	77.3	78.9	76.8	75.2	75.2	77.3	75.6
	50	0.00	52.8	50.9	52.9	51.4	54.1	54.7	55.1	53.8	56.9	57.4	56.2	56.7
		0.25	74.5	73.3	75.6	74.2	75.2	75.4	76.2	75.5	76.3	76.4	76.4	75.7
		0.50	95.3	95.4	96.3	95.6	95.4	95.8	96.6	95.9	96.6	96.7	96.9	96.3
L	25	0.00	61.4	61.1	63.8	62.2	63.7	64.3	65.7	63.4	61.6	63.2	64.8	61.8
		0.25	76.8	77.3	79.1	77.6	79.2	79.7	82.1	79.7	78.6	80.2	81.7	79.0
		0.50	93.8	94.5	95.7	94.6	95.3	95.7	96.2	95.4	95.8	96.0	96.9	96.0
	35	0.00	82.1	82.8	83.9	82.7	82.8	82.7	84.6	83.4	82.5	82.5	84.4	82.7
		0.25	93.3	93.4	94.0	93.1	95.3	95.1	95.8	95.2	94.3	94.5	95.4	94.5
		0.50	99.5	99.9	99.6	99.6	99.7	99.7	99.8	99.7	99.5	99.5	99.6	99.5
	50	0.00	95.9	95.7	96.5	95.9	96.6	96.9	97.2	97.1	96.8	96.8	97.6	97.3
		0.25	98.9	99.1	99.0	98.8	99.5	99.5	99.6	99.6	99.6	99.6	99.8	99.6
M	25	0.00	33.9	34.5	35.4	33.6	34.4	35.7	36.0	34.1	34.7	35.9	36.4	34.3
		0.25	41.0	40.7	41.4	39.8	41.9	42.2	43.7	42.0	42.1	42.4	43.7	41.8
		0.50	47.5	48.0	49.3	47.3	49.7	50.9	51.3	49.1	50.4	49.9	51.5	49.6
	35	0.00	50.8	50.7	52.3	50.4	51.0	51.6	53.6	51.7	51.4	52.0	53.9	51.8
		0.25	60.5	61.7	62.8	60.4	63.4	61.9	64.6	62.3	60.7	60.9	62.2	60.2
		0.50	71.8	71.9	73.4	71.6	75.8	76.1	76.9	75.7	72.9	74.0	74.4	73.1
	50	0.00	74.8	74.3	75.8	74.2	73.6	73.3	75.2	73.4	75.4	74.2	76.4	74.6
		0.25	84.6	85.3	85.9	84.6	84.6	84.0	85.3	83.9	86.9	86.1	87.1	86.2
		0.50	92.7	93.6	93.0	92.5	92.8	93.2	93.5	92.8	94.4	94.6	94.4	94.1

The empirical powers of the testing procedures obtained in models M1–M3 and M5–M6 are given in Tables 2–4 and 6–7. Similarly to the empirical sizes, the empirical powers are also quite satisfactory. The observed differences among the empirical powers of all tests are very small. In models M1–M2, the B test is usually a bit better than the other tests, while in models M3 and M5–M6, the P test has such property. In models M1–M2 and M5–M6, the empirical powers of each test are similar among different I's, while in model M3, they increase when I increases. They also increase with n or ρ. Since in models M3 and M6 the functions m_1 and m_2 are very close to each other, the observed empirical powers are usually moderate. In the other models, they are generally quite high even for small n and ρ in all considered situations. Thus, the empirical powers of the BT test are comparable with those of the tests proposed by Martínez-Camblor and Corral (2011), and their behavior is quite satisfactory.

5. SPEED COMPARISON

In this section, we study how the computational time required to perform the A, B, P and BT tests depends on the number of observations n and the number of design time points.

In the experiments, the functional data were generated under models considered in Section 4. We changed $n = 100, 200, ..., 1000$ and $I = 500, 1000$. As an example, Figure 3 shows the execution times against n for obtaining the final p-values of the A, B, P and BT tests when the data were generated as in model M4 under normal case and $\rho = 0.5$, $I = 500$ or $I = 1000$. The results obtained in the other models are similar.

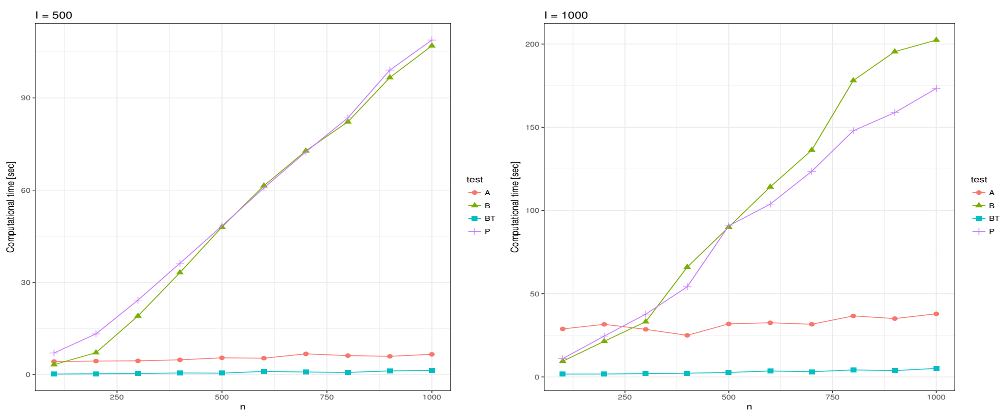

Figure 3: The execution times versus n for obtaining the final p-values of the A, B, P and BT tests when the number I of design time points in $[0, 1]$ as well as in $[1, 2]$ is equal to 500 or 1000. The data were generated as in model M4 under normal case and $\rho = 0.5$.

First of all, the BT test is the fastest among all considered ones, as was expected. It may be extremely faster than the testing procedures of Martínez-Camblor and Corral (2011), and works at most a few seconds. The execution time for the A test almost does not depend on the number of observations. This follows from that in the implementation of the A test, the data are used only to calculate the value of test statistic and the estimator of covariance function (This is done only once.). However, the execution time for this testing procedure increases significantly with an increase of the number of design time points, since the generation of artificial trajectories of the Gaussian process ξ described in Section 2 strongly depends on it. In most cases, the nonparametric bootstrap and permutation methods are the slowest ones. Their execution times are quite similar and increase much with an increase of n or I.

Summarizing, the BT test works very fast even for big data sets, in contrast to the other testing procedures under consideration.

6. APPLICATIONS TO THE ORTHOSIS DATA

In this section, we apply the new test and the testing procedures proposed by Martínez-Camblor and Corral (2011) to real-data example, using orthosis data, which are available on the website of Professor Jin-Ting Zhang (`http://www.stat.nus.edu.sg/ zhangjt/books/Chapman/FANOVA.htm`). These data were used for illustrative purposes in many problems for functional data (see, for instance, Abramovich *et al.*, 2004; Górecki and Smaga, 2015; Zhang and Liang, 2014).

Abramovich *et al.* (2004) reported the orthosis data were acquired and computed in an experiment by Dr. Amarantini David and Dr. Martin Luc (Laboratoire Sport et Performance Motrice, EA 597, UFRAPS, Grenoble University, France). The aim of their research was to investigate how muscle copes with an external perturbation. Seven young male volunteers participated in the experiment. They wore a spring-loaded orthosis of adjustable stiffness under the following four experimental conditions: a control condition (without orthosis); an orthosis condition (with orthosis); and two spring conditions (with spring 1 or with spring 2) in which stepping-in-place was perturbed by fitting a spring-loaded orthosis onto the right knee joint. All volunteers tried all four conditions 10 times for 20 seconds each. In order to avoid possible perturbations in the initial and final parts of the experiment, only the central 10 seconds were used in the study. The resultant moment of force at the knee was derived by means of body segment kinematics recorded with a sampling frequency of 200 Hz. For each stepping-in-place replication, the resultant moment was computed at 256 time points, equally spaced and scaled to the interval $[0, 1]$ so that a time interval corresponded to an individual gait cycle.

For illustrative purposes, we use the orthosis data under the first (without orthosis) and third (with spring 1) experimental conditions. For each volunteer, we calculate the mean curve of the 10 raw orthosis curves under these conditions. Figure 1 depicts the resulting curves. Of interest is to test if the mean curves of all volunteers are different under these two conditions ($t \in [0, 1]$ — without orthosis; $t \in [1, 2]$ — with spring 1). This is a paired two-sample problem for functional data. We applied the A, B, P and BT tests to this problem and the p-values of these tests are equal to 0.001, 0, 0, 0.0008123766 respectively. Hence all testing procedures suggest that the mean curves of all volunteers under without orthosis and with spring 1 conditions are unlikely the same. From Figure 1, however, we observe that the mean curves may be the same at the last stage of the experiment, i.e., for $t \in [0.8, 1] \cup [1.8, 2]$. In this case, the p-values of the A, B, P and BT tests are equal to 0.201, 0.204, 0.241, 0.2368321 respectively, and hence we fail to reject the equality of mean curves of all volunteers under without orthosis and with spring 1 conditions over $[0.8, 1] \cup [1.8, 2]$. Zhang and Liang (2014) also observed similar behavior of orthosis curves at the last stage of the experiment and confirmed its evidence by using appropriate tests. However, they considered the orthosis data under all four experimental conditions in the context of the functional analysis of variance.

7. CONCLUSIONS

In this paper, we studied the paired two-sample problem for functional data. We proposed the test for this problem based on the test statistic considered by Martínez-Camblor and Corral (2011) and the Box-type approximation for its asymptotic null distribution. This testing procedure is root-n consistent, easy to implement and much less computationally intensive than the re-sampling and permutation tests of Martínez-Camblor and Corral (2011). Moreover, it is comparable with those tests in terms of size control and power, and its finite sample behavior is very satisfactory. The illustrative real-data example indicates that the decisions suggested by the new test and the testing procedures of Martínez-Camblor and Corral (2011) seem to be similar in practice.

APPENDIX

A. Proofs

In the proofs, we use similar techniques as in Zhang *et al.* (2010a) and Zhang and Liang (2014).

Proof of Lemma 2.1: Under assumptions A1–A4, from the proof of Theorem 4.17 in Zhang (2013), it follows that $\hat{\mathbb{C}}(s,t) \xrightarrow{P} \mathbb{C}(s,t)$ uniformly over $[0,2]^2$, as $n \to \infty$. Hence, by (2.3) and the continuous mapping theorem, we obtain $\hat{\mathbb{K}}(s,t) \xrightarrow{P} \mathbb{K}(s,t)$. $\square$

Proof of Theorem 2.1: By Lemma 2.1, we obtain $\hat{\mathbb{K}}(s,t) \xrightarrow{P} \mathbb{K}(s,t)$ uniformly over $[0,1]^2$. Hence

$$\lim_{n\to\infty} \operatorname{tr}(\hat{\mathbb{K}}) = \int_0^1 \lim_{n\to\infty} \hat{\mathbb{K}}(t,t)\; dt = \int_0^1 \mathbb{K}(t,t)\; dt = \operatorname{tr}(\mathbb{K}),$$
$$\lim_{n\to\infty} \operatorname{tr}(\hat{\mathbb{K}}^{\otimes 2}) = \int_0^1\int_0^1 \lim_{n\to\infty} \hat{\mathbb{K}}^2(s,t)\; dsdt = \int_0^1\int_0^1 \mathbb{K}^2(s,t)\; dsdt = \operatorname{tr}(\mathbb{K}^{\otimes 2}).$$

Therefore, by (2.5) and (2.7) and the continuous mapping theorem, we conclude that

$$\hat{\beta} = \frac{\operatorname{tr}(\hat{\mathbb{K}}^{\otimes 2})}{\operatorname{tr}(\hat{\mathbb{K}})} \xrightarrow{P} \frac{\operatorname{tr}(\mathbb{K}^{\otimes 2})}{\operatorname{tr}(\mathbb{K})} = \beta, \quad \hat{d} = \frac{\operatorname{tr}^2(\hat{\mathbb{K}})}{\operatorname{tr}(\hat{\mathbb{K}}^{\otimes 2})} \xrightarrow{P} \frac{\operatorname{tr}^2(\mathbb{K})}{\operatorname{tr}(\mathbb{K}^{\otimes 2})} = d$$

and $\hat{\mathcal{C}}_{n,\alpha} = \hat{\beta}\chi^2_{\hat{d},\alpha} \xrightarrow{P} \beta\chi^2_{d,\alpha}$, as $n \to \infty$. The theorem is proved. $\square$

Proof of Theorem 3.1: Under the local alternatives $H_{1n}^{(1)}$, we have

(1.1)
$$\begin{aligned}\mathcal{C}_n &= n\int_0^1 \left((\bar{X}(t)-m(t)) - (\bar{X}(t+1)-m(t+1)) + (m(t)-m(t+1))\right)^2 dt\\ &= \int_0^1 \left(n^{1/2}(\bar{X}(t)-m(t)) - n^{1/2}(\bar{X}(t+1)-m(t+1)) + n^{(1-\tau)/2}d(t)\right)^2 dt.\end{aligned}$$

Under gaussianity assumption, Theorem 4.14 of Zhang (2013, p. 109) implies $n^{1/2}(\bar{X}(t)-m(t)), t\in[0,2]$ is a Gaussian process with mean zero and covariance function $\mathbb{C}(s,t)$. Hence, the processes $n^{1/2}(\bar{X}(t)-m(t))$ and $n^{1/2}(\bar{X}(t+1)-m(t+1))$ for $t\in[0,1]$ are also Gaussian processes with such parameters.

Thus, $n^{1/2}(\bar{X}(t)-m(t))-n^{1/2}(\bar{X}(t+1)-m(t+1))+n^{(1-\tau)/2}d(t)$ is a Gaussian process with mean $n^{(1-\tau)/2}d(t)$ and covariance function $\mathbb{K}(s,t)$ given by (2.1). By the assumption of $d\in L^2([0,1])$ and since $\mathrm{tr}(\mathbb{K})$ is finite as noted in Section 2, from Theorem 4.2 in Zhang (2013, p. 86), it follows that $\mathcal{C}_n$ has the same distribution as $\sum_{r=1}^{l}\lambda_r A_r+n^{1-\tau}\sum_{r=l+1}^{\infty}\Delta_r^2$, where $A_r\sim\chi_1^2(n^{1-\tau}\lambda_r^{-1}\Delta_r^2)$ are independent, λ_r are the decreasing-ordered eigenvalues of $\mathbb{K}(s,t)$, $\Delta_r\overset{\mathrm{def}}{=}\int_0^1 d(t)\phi_r(t)dt$, $\phi_r(t)$ are the associated eigenfunctions of $\mathbb{K}(s,t)$, $r=1,2,\dots$, and l is the number of all positive eigenvalues. The possibility of $l=\infty$ is permitted. Using above observation and since $\sum_{r=1}^{\infty}\Delta_r^2=\|d\|^2$, we calculate the expected value and variance of the test statistic as follows

$$
\begin{aligned}
E(\mathcal{C}_n)&=\sum_{r=1}^{l}\lambda_r E(A_r)+n^{1-\tau}\sum_{r=l+1}^{\infty}\Delta_r^2\\
&=\sum_{r=1}^{l}\lambda_r(1+n^{1-\tau}\lambda_r^{-1}\Delta_r^2)+n^{1-\tau}\sum_{r=l+1}^{\infty}\Delta_r^2\\
&=\sum_{r=1}^{l}\lambda_r+n^{1-\tau}\|d\|^2=\mathrm{tr}(\mathbb{K})+n^{1-\tau}\|d\|^2,\\
Var(\mathcal{C}_n)&=\sum_{r=1}^{l}\lambda_r^2 Var(A_r)=2\sum_{r=1}^{l}\lambda_r^2(1+2n^{1-\tau}\lambda_r^{-1}\Delta_r^2)=2\sum_{r=1}^{l}\lambda_r^2+4n^{1-\tau}\Delta_\lambda^2\\
&=2\mathrm{tr}(\mathbb{K}^{\otimes 2})+4n^{1-\tau}\Delta_\lambda^2,
\end{aligned}
$$

where $\Delta_\lambda^2\overset{\mathrm{def}}{=}\sum_{r=1}^{l}\lambda_r\Delta_r^2$. The rest of the proof is divided into two cases.

Case 1. Let $\Delta_r=0$ for all $r=1,\dots,l$. Then, $\mathcal{C}_n$ has the same distribution as

$$
\sum_{r=1}^{l}\lambda_r A_r+n^{1-\tau}\sum_{r=1}^{\infty}\Delta_r^2=\sum_{r=1}^{l}\lambda_r A_r+n^{1-\tau}\|d\|^2,
$$

where $A_r\sim\chi_1^2$. Hence, the distributions of $\mathcal{C}_n$ and $\mathcal{C}_0^*+n^{1-\tau}\|d\|^2$ are the same, where $\mathcal{C}_0^*$ is given in (2.2). Theorem 2.1 implies the asymptotic power of the BT test is of the form $P(\mathcal{C}_n>\hat{\mathcal{C}}_{n,\alpha})=P(\mathcal{C}_0^*>\mathcal{C}_{0,\alpha}-n^{1-\tau}\|d\|^2)+o(1)$, and it is easy to see that this power tends to 1, as $n\to\infty$.

Case 2. Let $\Delta_r\neq 0$ for some $r\in\{1,\dots,l\}$. Since $A_r\sim\chi_1^2(n^{1-\tau}\lambda_r^{-1}\Delta_r^2)$, it has the same distribution as $(Y_r+n^{(1-\tau)/2}\lambda_r^{-1/2}\Delta_r)^2$, where $Y_r\sim N(0,1)$. Thus, the distribution of $\mathcal{C}_n$ is as that of $\sum_{r=1}^{l}\lambda_r Y_r^2+2n^{(1-\tau)/2}\Delta_\lambda Y+n^{1-\tau}\sum_{r=1}^{\infty}\Delta_r^2$, where $Y\overset{\mathrm{def}}{=}\sum_{r=1}^{l}\lambda_r^{1/2}\Delta_r Y_r/\Delta_\lambda\sim N(0,1)$. Therefore, $(\mathcal{C}_n-E(\mathcal{C}_n))/Var(\mathcal{C}_n)$ has the same distribution as

$$
\frac{\sum_{r=1}^{l}\lambda_r(Y_r^2-1)}{\sqrt{2\mathrm{tr}(\mathbb{K}^{\otimes 2})+4n^{1-\tau}\Delta_\lambda^2}}+\frac{2n^{(1-\tau)/2}\Delta_\lambda Y}{\sqrt{2\mathrm{tr}(\mathbb{K}^{\otimes 2})+4n^{1-\tau}\Delta_\lambda^2}}.
$$

Since $\tau \in [0, 1)$, $\mathrm{tr}(\mathbb{K}^{\otimes 2})$ is finite (by the Cauchy–Schwarz inequality) and $0 < \Delta_\lambda^2 < \lambda_1 \sum_{r=1}^{l} \Delta_r^2 \le \lambda_1 \|d\|^2 < \infty$, we have

$$\frac{\sum_{r=1}^{l} \lambda_r (Y_r^2 - 1)}{\sqrt{2\mathrm{tr}(\mathbb{K}^{\otimes 2}) + 4n^{1-\tau}\Delta_\lambda^2}} \xrightarrow{p} 0,$$

and

$$\frac{2n^{(1-\tau)/2}\Delta_\lambda Y}{\sqrt{2\mathrm{tr}(\mathbb{K}^{\otimes 2}) + 4n^{1-\tau}\Delta_\lambda^2}} = \frac{2\Delta_\lambda Y}{\sqrt{2\mathrm{tr}(\mathbb{K}^{\otimes 2})/n^{1-\tau} + 4\Delta_\lambda^2}} \xrightarrow{d} Y \sim N(0,1),$$

as $n \to \infty$. By Theorem 2.1, we obtain

$$P(\mathcal{C}_n > \hat{\mathcal{C}}_{n,\alpha}) = 1 - \Phi\left(\frac{\mathcal{C}_{0,\alpha} - \mathrm{tr}(\mathbb{K}) - n^{1-\tau}\|d\|^2}{\sqrt{2\mathrm{tr}(\mathbb{K}^{\otimes 2}) + 4n^{1-\tau}\Delta_\lambda^2}} \right) + o(1),$$

where Φ is the cumulative distribution function $N(0, 1)$. Hence, $P(\mathcal{C}_n > \hat{\mathcal{C}}_{n,\alpha}) \to 1$, as $n \to \infty$, because $\tau \in [0, 1)$ and $\mathcal{C}_{0,\alpha}$, $\mathrm{tr}(\mathbb{K})$, $\mathrm{tr}(\mathbb{K}^{\otimes 2})$ and $\Delta_\lambda^2 > 0$ are finite. □

Proof of Theorem 3.2: Under the local alternatives $H_{1n}^{(2)}$, by (1.1), we have

$$\mathcal{C}_n = \int_0^1 \left(n^{1/2}(\bar{X}(t) - m(t)) - n^{1/2}(\bar{X}(t+1) - m(t+1)) + d(t) \right)^2 dt.$$

Similarly as in the proof of Theorem 1 in Martínez-Camblor and Corral (2011), we obtain $n^{1/2}(\bar{X}(t) - m(t)) - n^{1/2}(\bar{X}(t+1) - m(t+1)) + d(t) \xrightarrow{d} \xi_d(t)$, as $n \to \infty$, where $\xi_d(t), t \in [0, 1]$ is a Gaussian process with mean $d(t)$ and covariance function $\mathbb{K}(s, t)$ given by (2.1). Hence, by the continuous mapping theorem, we have $\mathcal{C}_n \xrightarrow{d} \|\xi_d\|^2$, as $n \to \infty$. Since $d \in L^2([0, 1])$ and $\mathrm{tr}(\mathbb{K}) < \infty$ (see Section 2), Theorem 4.2 in Zhang (2013, p. 86) shows that $\|\xi_d\|^2$ has the same distribution as $\sum_{r=1}^{l} \lambda_r A_r + \sum_{r=l+1}^{\infty} \delta_r^2$, where $A_r \sim \chi_1^2(\lambda_r^{-1}\delta_r^2)$ are independent, λ_r are the decreasing-ordered eigenvalues of $\mathbb{K}(s, t)$, $\delta_r \stackrel{\mathrm{def}}{=} \int_0^1 d(t)\phi_r(t)dt$, $\phi_r(t)$ are the associated eigenfunctions of $\mathbb{K}(s, t)$, $r = 1, 2, ...$, and l is the number of all positive eigenvalues ($l = \infty$ is possible). Since A_r has the same distribution as $(Y_r + \lambda_r^{-1/2}\delta_r)^2$, $Y_r \sim N(0, 1)$, the distribution of $\|\xi_d\|^2$ is the same as that of $\sum_{r=1}^{l} \lambda_r Y_r^2 + 2\delta_\lambda Y + \sum_{r=1}^{\infty} \delta_r^2$, where $\delta_\lambda^2 \stackrel{\mathrm{def}}{=} \sum_{r=1}^{l} \lambda_r \delta_r^2$ and $Y \stackrel{\mathrm{def}}{=} \sum_{r=1}^{l} \lambda_r^{1/2}\delta_r Y_r / \delta_\lambda \sim N(0, 1)$. Observing that $\sum_{r=1}^{\infty} \delta_r^2 = \|d\|^2$ and by Theorem 2.1, the asymptotic power of the BT test, as $n \to \infty$, is given by $P(\mathcal{C}_n > \hat{\mathcal{C}}_{n,\alpha}) = P(\mathcal{C}_0^* + 2\delta_\lambda Y + \|d\|^2 > \mathcal{C}_{0,\alpha}) + o(1)$, where $\mathcal{C}_0^*$ and $\mathcal{C}_{0,\alpha}$ are given in (2.2) and Theorem 2.1. The rest of the proof runs as in the proof of Proposition 4 in Zhang and Liang (2014) taking $\delta^2 = \|d\|^2$. □

B. Numerical implementation

As we mentioned in Subsection 4.1, in practice, the n functional observations are not continuously observed. Each function is usually observed on a grid of design time points. In this paper, all individual functions $X_i(t)$ for $t \in [0,1]$ and $t \in [1,2]$ in the simulations and the example (also in the function `BT.test` given in the next section) are assumed to be observe on a common grid of design time points that are equally spaced in $[0,1]$ and in $[1,2]$. To implement the new test when the design time points are different for different individual functions, one first has to reconstruct the functional sample from the observed discrete functional sample using some smoothing technique, then discretize each individual function of the reconstructed functional sample on a common grid of time points, and finally apply the test accordingly (see Zhang, 2013, or Zhang and Liang, 2014, for more details).

Assume that $0 = t_1 \leq t_2 \leq \cdots \leq t_p = 1$ and $1 = t_1 + 1 \leq t_2 + 1 \leq \cdots \leq t_p + 1 = 2$ denote a grid of design time points that are equally spaced in $[0,1]$ and in $[1,2]$, at which the data are observed. Then, we have

$$\mathcal{C}_n = n\int_0^1 (\bar{X}(t) - \bar{X}(t+1))^2\,dt \approx \frac{n}{p}\sum_{i=1}^{p}(\bar{X}(t_i) - \bar{X}(t_i+1))^2 = \frac{1}{p}\mathcal{C}_n^0,$$

$$\mathrm{tr}(\hat{\mathbb{K}}) = \int_0^1 \hat{\mathbb{K}}(t,t)\,dt \approx \frac{1}{p}\sum_{i=1}^{p}\hat{\mathbb{K}}(t_i,t_i) = \frac{1}{p}\mathrm{trace}(\hat{\mathbf{K}}),$$

$$\mathrm{tr}(\hat{\mathbb{K}}^{\otimes 2}) = \int_0^1\int_0^1 \hat{\mathbb{K}}^2(s,t)\,dsdt \approx \frac{1}{p^2}\sum_{i=1}^{p}\sum_{j=1}^{p}\hat{\mathbb{K}}^2(t_i,t_j) = \frac{1}{p^2}\mathrm{trace}(\hat{\mathbf{K}}^2),$$

where $\hat{\mathbf{K}} = (\hat{\mathbb{K}}(t_i,t_j))_{i,j=1}^{p}$. For example, similar approximations have previously been used by Zhang (2013, p. 117), and Zhang and Liang (2014). If the number p is very small, then we can first reconstruct the data as described in the last paragraph and then discretize the reconstructed functions on a greater number of design time points. The estimated parameters $\hat{\beta}$ and $\hat{d}$ in (2.5) are approximately expressed as

$$\hat{\beta} \approx \frac{\mathrm{trace}(\hat{\mathbf{K}}^2)}{p\cdot\mathrm{trace}(\hat{\mathbf{K}})} = \frac{1}{p}\hat{\beta}^0, \quad \hat{d} \approx \frac{\mathrm{trace}^2(\hat{\mathbf{K}})}{\mathrm{trace}(\hat{\mathbf{K}}^2)},$$

and hence the approximation of the p-value given in (2.6) is of the form

$$P(\chi^2_{\hat{d}} > \mathcal{C}_n/\hat{\beta}) \approx P(\chi^2_{\hat{d}} > \mathcal{C}_n^0/\hat{\beta}^0).$$

C. R code

The new test is performed by the R function `BT.test` given below. The notations in the program are consistent with or similar to those used in the paper. The argument `x` is a data frame or matrix of data, whose each row is a discretized version of a function $X_i(t)$, $t \in [0, 2]$, $i = 1, ..., n$. It means that the columns of `x` represent the values of the sample functions at the design time points. The number of columns is even, and the first half of them is connected with the design time points in $[0, 1]$, and the second half with those in $[1, 2]$. As outputs, we obtain value of test statistic and p-value of the test.

```
BT.test = function(x){
 n = nrow(x); p = ncol(x); CC = var(x)
 Cn = n*sum((colMeans(x[, 1:(p/2)]) - colMeans(x[, (p/2+1):p]))^2)
 KK = CC[1:(p/2), 1:(p/2)] - CC[1:(p/2), (p/2+1):p] -
      CC[(p/2+1):p, 1:(p/2)] + CC[(p/2+1):p, (p/2+1):p]
 A = sum(diag(KK)); B = sum(diag(KK%*%KK)); beta = B/A; d = (A^2)/B
 p.value = 1 - pchisq(Cn/beta, d)
 return(c(Cn/(p/2), p.value))
}
```

D. Additional simulations

In this section, we present some additional simulations suggested by one of the reviewers. The simulation models are similar to those in Subsection 4.1, but we consider the functional autoregressive process of order one ($\mathrm{FAR}_f(1)$) instead of compound symmetric dependency structure. The $\mathrm{FAR}_f(1)$ process was considered, for example, by Didericksen *et al.* (2012) or Horváth and Rice (2015). The error functions are generated in the following way:

$$\varepsilon_{ij}(t) = \eta \int_0^1 f(t,u)\varepsilon_{i,j-1}(u)\, du + \xi B_{ij}(t),\ t \in [0,1], i = 1, ..., n, j = 1, 2, \tag{4.1}$$

where f is a kernel, $\eta = 0.005, 0.125$ and $\xi = 0.05, 0.5$ for models M0–M3 and M4–M7, respectively, and B_{ij} are independent standard Brownian Bridges. If $\|f\| < 1$, then (4.1) has a unique stationary and ergodic solution (see Bosq, 2000, $\eta = \xi = 1$). We consider $f(t, u) = c \exp((t^2 + u^2)/2)$, where $c = 0.3416$ so that $\|f\| \approx 0.5$ (see Horváth and Rice, 2015). To obtain ε_{i0}, we use $\varepsilon_{i,-2} = \xi B_i$, where B_i is a standard Brownian Bridge, and then $\varepsilon_{i,-1}$ and ε_{i0} are generated according to (4.1). The errors functions ε_{ij} are adequately centered. The results are given in Table 8. They are very satisfactory. The conclusions are similar to those obtained in Subsection 4.2.

Table 8: Empirical sizes and powers (as percentages) of the A, B, P and BT tests obtained in $\mathrm{FAR}_f(1)$ case. The column "M" refers to different models.

M	I	26				101				251			
	n	A	B	P	BT	A	B	P	BT	A	B	P	BT
M0	25	4.7	4.6	4.6	4.6	5.5	5.3	5.3	5.4	4.4	4.4	4.5	4.4
	35	4.4	4.9	4.5	4.6	5.4	5.6	5.4	5.1	5.3	5.3	5.3	4.9
	50	4.9	4.5	5.3	4.8	4.8	4.8	4.3	4.9	6.0	5.2	5.5	5.6
M1	25	99.3	99.2	99.7	99.3	99.5	99.5	99.7	99.6	99.6	99.3	99.7	99.1
	35	100	100	100	100	100	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100	100	100	100
M2	25	100	100	100	100	100	100	100	100	100	100	100	100
	35	100	100	100	100	100	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100	100	100	100
M3	25	69.2	68.5	74.5	69.7	97.9	98.2	99.0	98.0	99.2	99.5	99.4	99.4
	35	94.9	94.7	96.5	95.3	100	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100	100	100	100
M4	25	5.2	5.5	4.9	5.1	4.3	4.4	4.1	4.5	6.1	6.4	5.7	5.9
	35	**6.6**	**6.7**	6.2	**6.6**	5.2	5.6	5.3	5.3	4.8	4.8	4.2	4.8
	50	5.5	5.4	4.9	5.5	5.9	5.5	5.9	5.7	5.4	5.3	4.8	5.1
M5	25	79.4	79.8	80.3	79.0	84.6	84.1	84.8	84.0	83.9	84.4	84.2	83.5
	35	96.1	95.9	96.6	95.8	97.0	97.0	97.5	97.3	96.9	96.8	97.5	97.0
	50	100	100	100	100	100	99.9	100	100	100	99.9	100	100
M6	25	30.1	29.7	29.4	29.2	29.7	30.1	29.9	29.4	31.8	32.0	31.6	30.9
	35	41.5	41.2	42.2	41.3	43.0	43.6	42.9	42.2	44.7	43.7	44.6	43.4
	50	58.7	57.7	58.6	57.4	64.8	65.2	65.9	63.9	65.9	65.8	66.7	65.5
M7	25	100	100	100	100	100	100	100	100	100	100	100	100
	35	100	100	100	100	100	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100	100	100	100

ACKNOWLEDGMENTS

The author is thankful to the referees for their constructive suggestions which improved the presentation of this article.

REFERENCES

[1] Abramovich, F.; Antoniadis, A.; Sapatinas, T. and Vidakovic, B. (2004). Optimal testing in a fixed-effects functional analysis of variance model, *International Journal of Wavelets, Multiresolution, Information Processing*, **2**, 323–349.

[2] Attouch, M.K. and Belabed, F.Z. (2014). The k nearest neighbors estimation of the conditional hazard function for functional data, *REVSTAT — Statistical Journal*, **12**, 273–297.

[3] Berrendero, J.R.; Justel, A. and Svarc, M. (2011). Principal components for multivariate functional data, *Computational Statistics & Data Analysis*, **55**, 2619–2634.

[4] Boente, G.; Barrera, M.S. and Tyler, D.E. (2014). A characterization of elliptical distributions, some optimality properties of principal components for functional data, *Journal of Multivariate Analysis*, **131**, 254–264.

[5] Boente, G. and Fraiman, R. (2000). Kernel-based functional principal components, *The Annals of Statistics*, **20**, 655–674.

[6] Bosq, D. (2000). *Linear Processes in Function Spaces*, Springer, New York.

[7] Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification, *The Annals of Mathematical Statistics*, **25**, 290–302.

[8] Brunner, E.; Dette, H. and Munk, A. (1997). Box-type approximations in nonparametric factorial designs, *Journal of the American Statistical Association*, **92**, 1494–1502.

[9] Chesneau, C.; Kachour, M. and Maillot, B. (2013). Nonparametric estimation for functional data by wavelet thresholding, *REVSTAT — Statistical Journal*, **11**, 211–230.

[10] Cuesta-Albertos, J.A. and Febrero-Bande, M. (2010). A simple multiway ANOVA for functional data, *Test*, **19**, 537–557.

[11] Cuevas, A. (2014). A partial overview of the theory of statistics with functional data, *Journal of Statistical Planning and Inference*, **147**, 1–23.

[12] Cuevas, A.; Febrero, M. and Fraiman, R. (2004). An ANOVA test for functional data, *Computational Statistics & Data Analysis*, **47**, 111–122.

[13] 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**, 1063–1074.

[14] Cuevas, A.; Febrero, M. and Fraiman, R. (2007). Robust estimation, classification for functional data via projection-based depth notions, *Computational Statistics*, **22**, 481–496.

[15] Didericksen, D.; Kokoszka, P. and Zhang, X. (2012). Empirical properties of forecasts with the functional autoregressive model, *Computational Statistics*, **27**, 285–298.

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

[17] Górecki, T.; Krzyśko, M. and Waszak, Ł. (2014). Functional discriminant coordinates, *Communications in Statistics — Theory and Methods*, **43**, 1013–1025.

[18] Górecki, T. and Smaga, Ł. (2015). A comparison of tests for the one-way ANOVA problem for functional data, *Computational Statistics*, **30**, 987–1010.

[19] Górecki, T. and Smaga, Ł. (2017). Multivariate analysis of variance for functional data, *Journal of Applied Statistics*, **44**, 2172–2189.

[20] Gregorutti, B.; Michel, B. and Saint-Pierre, P. (2015). Grouped variable importance with random forests, application to multiple functional data analysis, *Computational Statistics & Data Analysis*, **90**, 15–35.

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

[22] Horváth, L. and Rice, G. (2015). Testing equality of means when the observations are from functional time series, *Journal of Time Series Analysis*, **36**, 84–108.

[23] Jacques, J. and Preda, C. (2014). Functional data clustering: a survey, *Advances in Data Analysis and Classification*, **8**, 231–255.

[24] James, G.M. and Hastie, T.J. (2001). Functional linear discriminant analysis for irregularly sampled curves, *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, **63**, 533–550.

[25] Lian, H. (2012). Empirical likelihood confidence intervals for nonparametric functional data analysis, *Journal of Statistical Planning and Inference*, **142**, 1669–1677.

[26] Martínez-Camblor, P. and Corral, N. (2011). Repeated measures analysis for functional data, *Computational Statistics & Data Analysis*, **55**, 3244–3256.

[27] Prakasa Rao, B.L.S. (2010). Nonparametric density estimation for functional data by delta sequences, *Brazilian Journal of Probability and Statistics*, **24**, 468–478.

[28] Preda, C.; Saporta, G. and Lévéder, C. (2007). PLS classification of functional data, *Computational Statistics*, **22**, 223–235.

[29] Ramsay, J.O.; Hooker, G. and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*, Springer.

[30] Ramsay, J.O. and Silverman, B.W. (2002). *Applied Functional Data Analysis. Methods and Case Studies*, Springer, New York.

[31] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis, 2nd edition*, Springer, New York.

[32] R Core Team (2015). R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org/`

[33] Tokushige, S.; Yadohisa, H. and Inada, K. (2007). Crisp and fuzzy k-means clustering algorithms for multivariate functional data, *Computational Statistics*, **22**, 1–16.

[34] Valderrama, M.J. (2007). An overview to modelling functional data, *Computational Statistics*, **22**, 331–334.

[35] Yamamoto, M. and Terada, Y. (2014). Functional factorial K-means analysis, *Computational Statistics & Data Analysis*, **79**, 133–148.

[36] Zhang, J.T. (2005). Approximate and asymptotic distributions of chi-squared-type mixtures with applications, *Journal of the American Statistical Association*, **100**, 273–285.

[37] Zhang, J.T. (2013). *Analysis of Variance for Functional Data*, Chapman & Hall, London.

[38] Zhang, J.T. and Liang, X. (2014). One-way ANOVA for functional data via globalizing the pointwise F-test, *Scandinavian Journal of Statistics*, **41**, 51–71.

[39] Zhang, J.T.; Liang, X. and Xiao, S. (2010a). On the two-sample Behrens–Fisher problem for functional data, *Journal of Statistical Theory and Practice*, **4**, 571–587.

[40] Zhang, C.Q.; Peng, H. and Zhang, J.T. (2010b). Two-sample tests for functional data, *Communications in Statistics — Theory and Methods*, **39**, 559–578.