Accepted Manuscript

Asymptotics, finite-sample comparisons and applications for two-sample tests with functional data

Qing Jiang, Marie Hušková, Simos G. Meintanis, Lixing Zhu

PII: S0047-259X(17)30730-3
DOI: https://doi.org/10.1016/j.jmva.2018.09.002
Reference: YJMVA 4412

To appear in: *Journal of Multivariate Analysis*

Received date : 28 November 2017

Please cite this article as: Q. Jiang, M. Hušková, S.G. Meintanis, L. Zhu, Asymptotics, finite-sample comparisons and applications for two-sample tests with functional data, *Journal of Multivariate Analysis* (2018), https://doi.org/10.1016/j.jmva.2018.09.002

Asymptotics, finite-sample comparisons and applications for two-sample tests with functional data

Qing Jiang[a], Marie Hušková[b], Simos G. Meintanis[c,d,*], Lixing Zhu[a,e]

[a] *School of Statistics, Beijing Normal University, Beijing, P.R. China*
[b] *Department of Statistics, Charles University, Prague, Czech Republic*
[c] *Department of Economics, National and Kapodistrian University of Athens, Athens, Greece*
[d] *Unit for Business Mathematics and Informatics, North-West University, Potchefstroom, South Africa*
(on sabbatical leave from the University of Athens)
[e] *Department of Mathematics, Hong Kong Baptist University, Hong Kong*

Abstract

We consider two-sample tests for functional data with observations which may be uni- or multi-dimensional. The new methods are formulated as L_2-type criteria based on empirical characteristic functions and are convenient from the computational point of view. Asymptotic properties are presented. Simulations and two real data applications are conducted in order to evaluate the performance of the proposed tests vis-à-vis other methods.

Keywords: Empirical characteristic function, Functional data, Two-sample problem

1. Introduction

In recent years, methods for analyzing data expressed in the form of functions have received much attention. As a result there is currently a wide spectrum of models and methods for functional data analysis from mean/covariance function estimation and functional principal component analysis or regression, clustering and classification, including methods for time series modeling of functional objects. For an overview, the reader may refer to [16, 27, 44]. Recent review papers [11, 51] and special issues [19, 20, 35, 50] also provide up-to-date information on research in this area.

This paper addresses the general two-sample problem in the context of functional observations from a nonparametric perspective. Suppose that we observe data $\boldsymbol{X}_{1ij}$ and $\boldsymbol{X}_{2ij}$ arising from two different groups. For each fixed i, $\boldsymbol{X}_{1ij}$ is viewed as a realization of a curve $\boldsymbol{x}_{1i}(t)$ observed at distinct time points $t_{1i1}, \ldots, t_{1im_{1i}}$, and we index the curves by $i \in \{1, \ldots, n_1\}$ for the first group. Likewise suppose that $\boldsymbol{X}_{2ij}$ is a realization of a curve $\boldsymbol{x}_{2i}(t)$, observed at times $t_{2i1}, \ldots, t_{2im_{2i}}$ and indexed by $i \in \{1, \ldots, n_2\}$ for the second group. The observation times t_{1ij}, t_{2ij} are assumed to belong to some closed bounded interval $\mathbb{T}$ that we often take to be $[0, 1]$. Although we work under the assumption of independence between groups, we allow for noise in the observations. Specifically we consider the model

$$\boldsymbol{X}_{1i}(t) = \boldsymbol{x}_{1i}(t) + \boldsymbol{\varepsilon}_{1i}(t), \quad \boldsymbol{X}_{2i}(t) = \boldsymbol{x}_{2i}(t) + \boldsymbol{\varepsilon}_{2i}(t), \tag{1}$$

In this notation, $\boldsymbol{X}_{kij}$ and $\boldsymbol{x}_{kij}$ correspond to $\boldsymbol{X}_{ki}(t)$ and $\boldsymbol{x}_{ki}(t)$ at distinct time points t_{kij} with $k \in \{1, 2\}$. Within Model (1), we wish to test the null hypothesis

$$\mathcal{H}_0 : \forall_{t\in\mathbb{T}}\ \boldsymbol{x}_1(t) =_{\mathcal{L}} \boldsymbol{x}_2(t), \tag{2}$$

where $=_{\mathcal{L}}$ stands for equality in law and $\{\boldsymbol{x}_1(t), t \in \mathbb{T}\}$ denotes a generic instance of the independent and identically distributed (iid) random functions $\{\boldsymbol{x}_{1i}(t), t \in \mathbb{T}\}$ with $i \in \{1, \ldots, n_1\}$, and $\{\boldsymbol{x}_2(t), t \in \mathbb{T}\}$ denotes a generic instance of the iid random functions $\{\boldsymbol{x}_{2i}(t), t \in \mathbb{T}\}$ with $i \in \{1, \ldots, n_2\}$. We also assume that the errors $\{\boldsymbol{\varepsilon}_{1i}(t), t \in \mathbb{T}\}$, $i \in \{1, \ldots, n_1\}$,

*Corresponding author
Email address: simosmei@econ.uoa.gr (Simos G. Meintanis)

and $\{\boldsymbol{\varepsilon}_{2i}(t), t \in \mathbb{T}\}$, $i \in \{1, \ldots, n_2\}$, are iid with zero mean. More specific assumptions are listed in Section 3. Note that the null hypothesis $\mathcal{H}_0$ in (2) refers to equality of marginal distributions, i.e., for each fixed t, which is certainly much narrower that the global version of $\mathcal{H}_0$ that incorporates marginal as well as joint distributions of the two stochastic processes. We nevertheless deal with this more general version of the null hypothesis starting in Section 2.3.

Previous work on the two-sample problem with functional data includes testing for a common location [12, 27–29, 53] and for a common covariance operator [17, 37, 40, 54], both for independent data and functional time series. The more general problem of testing for common principal components is considered in [2]. For a broader treatment of the functional two-sample problem, see [25, 42]. Here however, we deviate from these papers by proposing procedures that rely on the empirical characteristic function (ECF). Apart from other favorable features which will become apparent along the way, note that ECF-based procedures for scalar data are readily extended to multidimensional observations, which is not always true if one uses classical procedures based on the empirical distribution function.

The remainder of the paper is as follows. In Section 2 we introduce the test criteria while in Section 3 we study the large-sample behavior of the new method. Section 4 is devoted to computational and other interesting aspects of the test criteria, and resampling techniques for the implementation of the procedures are provided. In Section 5, we present the results of a Monte Carlo study on the finite-sample properties of the method, along with two illustrations. Finally, we end in Section 6 with conclusions and discussion. Some technical material is deferred to the Appendix. For an earlier version of this paper, the reader is referred to [33].

2. Test statistics

2.1. Univariate case

Our approach for testing $\mathcal{H}_0$ in (2) will be based on the fact that this null hypothesis is tantamount to the identity

$$\forall_{t\in\mathbb{T}}\ \forall_{u\in\mathbb{R}}\ \varphi_{x_1(t)}(u) = \varphi_{x_2(t)}(u) \tag{3}$$

and vice versa. Here and in what follows, $\varphi_{z(t)}(u) = \mathrm{E}(e^{\iota u z(t)})$ with $\iota = \sqrt{-1}$ will denote the characteristic function (CF) of the stochastic quantity $z(t)$. As in [30], we adopt the Fourier formulation (3) of $\mathcal{H}_0$ as our point of departure. Also we assume that the curves $x_{1i}(t)$ and $x_{2i}(t)$ may be recovered following nonparametric techniques (e.g., local linear regression) and write $\widehat{x}_{1i}(t)$ and $\widehat{x}_{2i}(t)$ for the resulting curve estimators. Consider the corresponding ECFs, viz.

$$\widehat{\varphi}_{\widehat{x}_1(t)}(u) = \frac{1}{n_1}\sum_{i=1}^{n_1} e^{\iota u \widehat{x}_{1i}(t)}, \quad \widehat{\varphi}_{\widehat{x}_2(t)}(u) = \frac{1}{n_2}\sum_{i=1}^{n_2} e^{\iota u \widehat{x}_{2i}(t)}, \tag{4}$$

computed from $\widehat{x}_{11}(t), \ldots, \widehat{x}_{1n_1}(t)$ and $\widehat{x}_{21}(t), \ldots, \widehat{x}_{2n_2}(t)$, respectively. Then in view of (3) we suggest the test statistic

$$D_W = \int_{\mathbb{T}}\int_{\mathbb{R}} \widehat{d}_t(u) W(u) du dt, \tag{5}$$

where

$$\widehat{d}_t(u) = \left|\widehat{\varphi}_{\widehat{x}_1(t)}(u) - \widehat{\varphi}_{\widehat{x}_2(t)}(u)\right|^2, \tag{6}$$

and $W > 0$ denotes a weight function such that $\int_{\mathbb{R}} W(u)du < \infty$.

2.2. Multivariate case

The latent curves $\boldsymbol{x}_1(t)$, $\boldsymbol{x}_2(t)$ may also be multidimensional. This is a new area where functional data are observed over time t, but realizations are complex geometrical structures in dimension $p > 1$; see [3, 7, 23, 32] for recent statistical techniques for multivariate functional data. Along the same lines as in the previous section, we will consider testing $\mathcal{H}_0$ with a criterion analogous to (5). However, in order to avoid nonparametric estimation, which is problematic in high dimension, we suggest to modify our assumptions regarding Model (1).

We assume that observations are collected over time $t = t_j$ with $j \in \{1, \ldots, m\}$, for both groups, i.e., we have a common sampling design between the two groups with m being large; see also Section 2.3 for a data-dependent value of m. Moreover we assume that sampling noise is equidistributed between the two groups, viz. $\boldsymbol{\varepsilon}_1(t) =_{\mathcal{L}} \boldsymbol{\varepsilon}_2(t)$, with

a common CF that never vanishes. These assumptions are often realized in practice. Under these assumptions and using the Fourier identities $\varphi_{\boldsymbol{X}_k(t)}(\boldsymbol{u}) = \varphi_{\boldsymbol{x}_k(t)}(\boldsymbol{u})\varphi_{\boldsymbol{\varepsilon}_k(t)}(\boldsymbol{u})$ with $k \in \{1, 2\}$ resulting from Model (1), we conclude that the null hypothesis $\mathcal{H}_0$ in (2) holds if and only if

$$\forall_{t\in\mathbb{T}}\ \forall_{\boldsymbol{u}\in\mathbb{R}^p}\ \varphi_{\boldsymbol{X}_1(t)}(\boldsymbol{u}) = \varphi_{\boldsymbol{X}_2(t)}(\boldsymbol{u}).$$

In view of this fact, we analogously propose the test statistic

$$\Delta_W = \int_{\mathbb{T}}\int_{\mathbb{R}^p} \widehat{\delta}_t(\boldsymbol{u})W(\boldsymbol{u})d\boldsymbol{u}dt, \tag{7}$$

with $W : \mathbb{R}^p \mapsto (0, \infty)$ such that $\int_{\mathbb{R}^p} W(\boldsymbol{u})d\boldsymbol{u} < \infty$, where

$$\widehat{\delta}_t(\boldsymbol{u}) = \left|\widehat{\varphi}_{\boldsymbol{X}_1(t)}(\boldsymbol{u}) - \widehat{\varphi}_{\boldsymbol{X}_2(t)}(\boldsymbol{u})\right|^2, \tag{8}$$

with

$$\widehat{\varphi}_{\boldsymbol{X}_1(t)}(\boldsymbol{u}) = \frac{1}{n_1}\sum_{i=1}^{n_1} e^{\imath\boldsymbol{u}^\top \boldsymbol{X}_{1i}(t)}, \quad \widehat{\varphi}_{\boldsymbol{X}_2(t)}(\boldsymbol{u}) = \frac{1}{n_2}\sum_{i=1}^{n_2} e^{\imath\boldsymbol{u}^\top \boldsymbol{X}_{2i}(t)},$$

which correspond to the ECFs $\varphi_{\boldsymbol{X}_1(t)}(\boldsymbol{u})$ and $\varphi_{\boldsymbol{X}_2(t)}(\boldsymbol{u})$, respectively, considered at fixed time points $t_1, \ldots, t_m$.

Clearly our test statistic is based on a Cramér–von Mises type quadratic distance. While this is by far the dominant approach in the context of ECF-testing, the option of a Kolmogorov–Smirnov ECF statistic based on the supremum norm has also been considered in [9, 43]. However it is well known that a quadratic distance statistic is often more powerful than the corresponding supremum-based distance statistic, and certainly easier to implement particularly in higher dimension; see, e.g., [8]. Therefore, and in line with previous papers in the functional context [1, 25, 42], we opt for the quadratic distance statistic Δ_W in (7) and thereby, among other things, avoid the computational burden of taking a maximum of the ECF process over a high-dimensional $\boldsymbol{u}$-region; see (8).

Note that a multivariate two-sample test may be based on the equality of distributions of all inner products. It is well-known that two p-dimensional random vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ satisfy $\boldsymbol{x}_1 =_{\mathcal{L}} \boldsymbol{x}_2$ if and only if $\boldsymbol{u}^\top\boldsymbol{x}_1 =_{\mathcal{L}} \boldsymbol{u}^\top\boldsymbol{x}_2$, for each direction-vector $\boldsymbol{u} \in \mathbb{R}^p$. In order to circumvent the problem of having to consider all possible directions $\boldsymbol{u}$, some two-sample procedures test the equality over a set of randomly chosen values of $\boldsymbol{u}$, which are often normalized to projections, viz. $\boldsymbol{u}^\top\boldsymbol{u} = 1$. The use of random projection methods has its roots in the celebrated Johnson–Lindenstrauss lemma which states that, for any two vectors in fixed dimension, there is a lower dimension map which is approximately distance-preserving; see [6] for a review of random projection methods in higher dimension.

In this connection, the natural question then concerns the number of directions that are sufficient to completely determine the underlying law. For example in the context of multinormality testing, it is shown in [46] that non-normality is characterized by having the measure of the set of all normal projections equal to zero. Still the issue persists as to how in practice one can implement a test based on a characterization of this type. In this line of research there exist methods of, so-called, random projections relying on the equality of specific, often data-driven, directions $\boldsymbol{u}$; see, e.g., [1, 10, 18, 39]. Alternatively, the approach based on CFs takes into account all possible directions by means of the distributional equality of all trigonometric pairs, viz. $(\sin(\boldsymbol{u}^\top\boldsymbol{x}_1), \cos(\boldsymbol{u}^\top\boldsymbol{x}_1)) =_{\mathcal{L}} (\sin(\boldsymbol{u}^\top\boldsymbol{x}_2), \cos(\boldsymbol{u}^\top\boldsymbol{x}_2))$ for all $\boldsymbol{u} \in \mathbb{R}^p$. Clearly the use of all directions leads to inherently consistent procedures. The trigonometric transformation also has the advantage of being bounded and, as shown below, it leads to computationally convenient procedures.

2.3. Global testing

We extend the test of Section 2.1 to the case of global testing, i.e., testing for the full process including marginals and joint distributions. In this connection we note that, under weak assumptions culminating into the celebrated Kolmogorov consistency theorem, if all finite-dimensional distributions of two processes coincide, then these processes are equal in distribution. In turn it is well-known that each finite-dimensional distribution is completely determined by the corresponding CF. For simplicity of presentation, suppose that the two processes are observed without noise at common times, say $t = t_j = j/m$ with $j \in \{1, \ldots, m\}$. In line with the previous sections, we suggest the test criterion

$$\Upsilon_{W,m} = \int_{\mathbb{R}^m} \left|\widehat{\Phi}_{1m}(\boldsymbol{u}) - \widehat{\Phi}_{2m}(\boldsymbol{u})\right|^2 W(\boldsymbol{u})d\boldsymbol{u}, \tag{9}$$

where

$$\widehat{\Phi}_{km}(\boldsymbol{u}) = \frac{1}{n_k}\sum_{i=1}^{n_k} e^{\mathrm{i}\boldsymbol{u}^\top \mathbb{X}_{ki}},$$

with $\mathbb{X}_{ki} = (X_{ki1}, \ldots, X_{kim})^\top$ for all $i \in \{1, \ldots, n_k\}$ and $k \in \{1, 2\}$, while $X_{kij} = X_{ki}(t_j)$ for all $j \in \{1, \ldots, m\}$.

The difference here is that we allow the number of observation times to grow with the sample sizes n_1, n_2, viz. $m = m_{n_1,n_2} \to \infty$ as $\min(n_1, n_2) \to \infty$. Thus we are in the realm of high dimension. In this connection we also let the weight function depend on t_j, i.e., $W = W_{t_1,\ldots,t_m}$, in a way that will be specified later.

In the context of high dimension, we could also consider groups of local observation times of size L with $m = RL$, and formulate a test statistic as in (9), say Υ_{W,m_r}, on each group of observations corresponding to the specific group of observation times $m_r = \{t_{rL+1}, \ldots, t_{rL+L}\}$ for all $r \in \{0, \ldots, R-1\}$. This is in line with interval-wise testing procedures suggested in [41], which allows to locate portions of the domain that may lead to the rejection of the null hypothesis. Also this approach would somewhat alleviate the effect of high dimension, but this comes at the cost of having to implement a multiple test procedure based on $\Upsilon_{W,m_0}, \ldots, \Upsilon_{W,m_{R-1}}$. In contrast, a test based on $\Upsilon_{W,m_0} + \cdots + \Upsilon_{W,m_{R-1}}$, which could be preferable from the point of view of dimension reduction, would obscure the information drawn from the individual test statistics. Clearly further research is necessary in this direction.

3. Asymptotics

Here we formulate assertions on the limit behavior of Δ_W defined in (7) and of D_W defined in (5). Clearly both criteria implicitly depend on the number of time points at which the curves are observed, and hence since in what follows this number is assumed to be common for both curves, we will write $\Delta_{W,m}$ and $D_{W,m}$, where m denotes the aforementioned number of time points.

We start with formulation of the assumptions for each statistic. They are slightly different. Specifically for $\Delta_{W,m}$ we assume the following.

(A.1) All random functions $\{\boldsymbol{x}_{1i_1}(t),\, t \in \lfloor 0, 1\rfloor\}$, $\{\boldsymbol{\varepsilon}_{1i_1}(t),\, t \in \lfloor 0, 1\rfloor\}$ for $i_1 \in \{1, \ldots, n_1\}$, $\{\boldsymbol{x}_{2i_2}(t),\, t \in \lfloor 0, 1\rfloor\}$, $\{\boldsymbol{\varepsilon}_{2i_2}(t), t \in \lfloor 0, 1\rfloor\}$ for $i_2 \in \{1, \ldots, n_2\}$ are mutually independent.

(A.2) For $i_1 \in \{1, \ldots, n_1\}$ and $i_2 \in \{1, \ldots, n_2\}$, $\{\boldsymbol{\varepsilon}_{1i_1}(t),\, t \in \lfloor 0, 1\rfloor\}$ and $\{\boldsymbol{\varepsilon}_{2i_2}(t),\, t \in \lfloor 0, 1\rfloor\}$ are iid with zero mean and for some $C > 0$ and $\eta > 0$,

$$\mathrm{E}\int_{\mathbb{T}} \|\boldsymbol{\varepsilon}_{1i}(t)\|^{2+\eta} dt \le C, \quad \forall_{t_1,t_2\in\lfloor 0,1\rfloor}\, \mathrm{E}\,\|\boldsymbol{\varepsilon}_{1i}(t_1) - \boldsymbol{\varepsilon}_{1i}(t_2)\|^2 \le C|t_1 - t_2|,$$

where $\|\cdot\|$ is the Euclidean norm.

(A.3) For $i_1 \in \{1, \ldots, n_1\}$, $\{\boldsymbol{x}_{1i}(t),\, t \in \lfloor 0, 1\rfloor\}$ are iid p-dimensional random functions satisfying for some $\eta > 0$ and $C > 0$,

$$\mathrm{E}\int_{\mathbb{T}} \|\boldsymbol{x}_{1i}(t)\|^{2+\eta} dt \le C, \quad \forall_{t_1,t_2\in\lfloor 0,1\rfloor}\, \mathrm{E}\,\|\boldsymbol{x}_{1i}(t_1) - \boldsymbol{x}_{1i}(t_2)\|^2 \le C|t_1 - t_2|.$$

(A.4) For $i_1 \in \{1, \ldots, n_2\}$, $\{\boldsymbol{x}_{2i}(t),\, t \in \lfloor 0, 1\rfloor\}$ are iid p-dimensional random functions satisfying for some $\eta > 0$ and $C > 0$,

$$\mathrm{E}\int_{\mathbb{T}} \|\boldsymbol{x}_{2i}(t)\|^{2+\eta} dt \le C, \quad \forall_{t_1,t_2\in\lfloor 0,1\rfloor}\, \mathrm{E}\,\|\boldsymbol{x}_{2i}(t_1) - \boldsymbol{x}_{2i}(t_2)\|^2 \le C|t_1 - t_2|.$$

(A.5) $\min(n_1, n_2) \to \infty$ in such a way that $n_1/(n_1 + n_2) \to \theta \in (0, 1)$ and also $\min(m_1, m_2) \to \infty$, i.e., m_1 and m_2 tend to infinity together with $\min(n_1, n_2)$.

(A.6) $m_1 = m_2 = m \to \infty$, $t_{1i_1 j} = t_{2i_2 j} = t_j$ for all $i_1 \in \{1, \ldots, n_1\}$, $i_2 \in \{1, \ldots, n_2\}$, and $t_j = j/m$ for all $j \in \{1, \ldots, m\}$.

(A.7) The weight function W is non-negative and such that $W(\pm u_1, \ldots, \pm u_p) = W(u_1, \ldots, u_p)$ for all real $u_1, \ldots, u_p$, and $0 < \int \|\boldsymbol{u}\|^2 W(\boldsymbol{u}) d\boldsymbol{u} < \infty$.

For the test statistic $D_{W,m}$ we consider (A.1), (A.3), (A.4), (A.5), (A.6), (A.7) and additionally the following.

(A.2a) For all $i_1 \in \{1,\dots,n_1\}$, $i_2 \in \{1,\dots,n_2\}$ and $j \in \{1,\dots,m\}$, ε_{1i_1j} and ε_{2i_2j} are independent random variables with zero mean such that $\mathrm{E}|\varepsilon_{ki_kj}|^{2+\lambda} \le C$ for some positive $C > 0$ and some $\lambda > 0$. Moreover, ε_{1i_1j} with $i_1 \in \{1,\dots,n_1\}$ and $j \in \{1,\dots,m\}$ are identically distributed; likewise, ε_{2i_2j} with $i_2 \in \{1,\dots,n_2\}$ and $j \in \{1,\dots,m\}$ are identically distributed.

(A.8) The kernel function K is continuous symmetric positive on $\lfloor -1, +1 \rfloor$ and zero otherwise. Also as $m \to \infty$, the bandwidth $h = h_m$ satisfies $h_m \to 0$ and $mh_m \to \infty$

Most of the assumptions, notably Assumption (A.6), can be weakened at the cost of increased technicality. Notice also that, e.g., a Brownian bridge as well as many Gaussian processes satisfy (A.2) but not (A.2a), and that Assumption (A.8) refers to kernel estimators $\widehat{x_k}$, of the curves x_1, x_2 appearing in (4). In what follows, $\rightsquigarrow$ denotes convergence in distribution.

Theorem 1. *Under Assumptions (A.1)–(A.7) and under the null hypothesis, as* $\min(n_1, n_2) \to \infty$,

$$(n_1 + n_2)\Delta_{W,m} \rightsquigarrow \frac{1}{\theta(1-\theta)} \int_0^1 \int_{\mathbb{R}^p} \{V_\theta(t, \boldsymbol{u})\}^2 W(\boldsymbol{u}) d\boldsymbol{u} dt,$$

where $\{V_\theta(t, \boldsymbol{u}); t \in T, \boldsymbol{u} \in \mathbb{R}^p\}$ *is a Gaussian process with zero mean and covariance structure*

$$\mathrm{cov}\{V_\theta(t_1, \boldsymbol{u}_1), V_\theta(t_2, \boldsymbol{u}_2)\} = \frac{1}{\theta(1-\theta)} \mathrm{cov}[\sin\{\boldsymbol{u}_1^\top \boldsymbol{X}_1(t_1)\} + \cos\{\boldsymbol{u}_1^\top \boldsymbol{X}_1(t_1)\}, \sin\{\boldsymbol{u}_2^\top \boldsymbol{X}_1(t_2)\} + \cos\{\boldsymbol{u}_2^\top \boldsymbol{X}_1(t_2)\}].$$

Theorem 2. *Under Assumptions (A.1)–(A.5) and (A.7) the following holds true, as* $\min(n_1, n_2) \to \infty$*:*

$$\Delta_{W,m} \xrightarrow{\mathcal{P}} \int_0^1 \int_{\mathbb{R}^p} \Big[\mathrm{E}[\sin\{\boldsymbol{u}^\top \boldsymbol{X}_1(t)\} + \cos\{\boldsymbol{u}^\top \boldsymbol{X}_1(t)\} - \sin\{\boldsymbol{u}^\top \boldsymbol{X}_2(t)\} - \cos\{\boldsymbol{u}^\top \boldsymbol{X}_2(t)\}]\Big]^2 W(\boldsymbol{u}) d\boldsymbol{u} dt.$$

Theorem 2 covers alternatives. As soon as the right-hand side of the above limit relation is positive, $(n_1 + n_2)\Delta_{W,m}$ tends to infinity in probability which, together with Theorem 1, implies that the test is consistent.

Next, we formulate analogous assertions for $D_{W,m}$ suitable for the one-dimensional setup ($p = 1$).

Theorem 3. *Under Assumptions (A.1), (A.2a), (A.3)–(A.8) with* $p = 1$ *and under the null hypothesis, as* $\min(n_1, n_2) \to \infty$,

$$(n_1 + n_2) D_{W,m} \rightsquigarrow \frac{1}{\theta(1-\theta)} \int_0^1 \int_{\mathbb{R}} \{\widetilde{V}_\theta(t, u)\}^2 W(u) du dt,$$

where $\{\widetilde{V}_\theta(t, u) : t \in T, u \in \mathbb{R}\}$ *is a Gaussian process with zero mean and covariance structure*

$$\mathrm{cov}\{\widetilde{V}_\theta(t_1, u_1), \widetilde{V}_\theta(t_2, u_2)\} = \frac{1}{\theta(1-\theta)} \mathrm{cov}[\sin\{u_1 x_1(t_1)\} + \cos\{u_1 x_1(t_1)\}, \sin\{u_2 x_1(t_2)\} + \cos\{u_2 x_1(t_2)\}].$$

Theorem 4. *Under Assumptions (A.1), (A.2a), (A.3)–(A.8), the following holds true, as* $\min(n_1, n_2) \to \infty$*:*

$$D_{W,m} \xrightarrow{\mathcal{P}} \int_0^1 \int_{\mathbb{R}} [\mathrm{E}[\sin\{u x_1(t)\} + \cos\{u x_1(t)\} - \sin\{u x_2(t)\} - \cos\{u x_2(t)\}]]^2 W(u) du dt.$$

Theorems 3 and 4 have similar interpretations for $D_{W,m}$ as Theorems 1 and 2 do for $\Delta_{W,m}$.

4. Computations and interpretations

4.1. Univariate case

By assuming smoothness conditions on $x_{1i_1}(t)$ and $x_{2i_2}(t)$, we employ the local linear regression technique to estimate all individual functions $x_{1i_1}(t)$ and $x_{2i_2}(t)$. We compute the value $(\hat{a}_0, \hat{a}_1)$ of the vector (a_0, a_1) that minimizes

$$\sum_{j=1}^{m} \{X_{ki_kj} - a_0 - a_1(t_{i_kj} - t)\}^2 K_{h_{ki_k}}(t - t_{i_kj})$$

for $k \in \{1, 2\}$, where $K_h(\cdot) = h^{-1}K(\cdot/h)$. Here K, the kernel function, is a bounded, symmetric, compactly supported probability density, and h_{ki_k} is the bandwidth for the curve $x_{ki_k}(t)$. Then,

$$\hat{a}_0 = \widehat{x}_{ki_k}(t) = \frac{A_{ki_k2}(t)B_{ki_k0}(t) - A_{ki_k1}(t)B_{ki_k1}(t)}{A_{ki_k0}(t)A_{ki_k2}(t) - A_{ki_k1}(t)^2},$$

where

$$A_{ki_kr} = \frac{1}{mh_{ki_k}} \sum_{j=1}^{m} \left(\frac{t - t_{ikj}}{h_{ki_k}}\right)^r K\left(\frac{t - t_{ikj}}{h_{ki_k}}\right), \quad B_{ki_kr} = \frac{1}{mh_{ki_k}} \sum_{j=1}^{m} X_{ki_kj} \left(\frac{t - t_{ikj}}{h_{ki_k}}\right)^r K\left(\frac{t - t_{ikj}}{h_{ki_k}}\right).$$

Our procedures enjoy the advantage of computational simplicity. To see this we proceed from (6) and by using simple algebra and trigonometric identities, we get

$$\widehat{d}_t(u) = \frac{1}{n_1^2} \sum_{i,\ell=1}^{n_1} \cos[u\{\widehat{x}_{1i}(t) - \widehat{x}_{1\ell}(t)\}] + \frac{1}{n_2^2} \sum_{i,\ell=1}^{n_2} \cos[u\{\widehat{x}_{2i}(t) - \widehat{x}_{2\ell}(t)\}] - \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{\ell=1}^{n_2} \cos[u\{\widehat{x}_{1i}(t) - \widehat{x}_{2\ell}(t)\}]. \tag{10}$$

Then by making use of (10) in (5), we conclude that the test statistic can be written as

$$D_W = \frac{1}{n_1^2} \sum_{i,\ell=1}^{n_1} I_{W,\mathbb{T}}(\widehat{x}_{1i}, \widehat{x}_{1\ell}) + \frac{1}{n_2^2} \sum_{i,\ell=1}^{n_2} I_{W,\mathbb{T}}(\widehat{x}_{2i}, \widehat{x}_{2\ell}) - \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{\ell=1}^{n_2} I_{W,\mathbb{T}}(\widehat{x}_{1i}, \widehat{x}_{2\ell}),$$

where

$$I_{W,\mathbb{T}}(z_1, z_2) = \int_{\mathbb{T}} \int_{\mathbb{R}} \cos[u\{z_1(t) - z_2(t)\}] W(u) du dt. \tag{11}$$

The weight function W in (11) may be chosen in a way that avoids numerical integration in the inner integral $\int \cos\{u(z)\} w(u) du$, but for further details on this we refer to the next section. Then again having computed $\int \cos\{u(z)\} W(u) du = g\{z(t)\}$, say, one also has to compute the outer integral $\int g\{z(t)\} dt$ over $\mathbb{T}$. However, even in the simplest case of local linear smoothers, the closed form obtained for $\widehat{x}_{ki_k}(t)$ is quite complicated and therefore one needs to resort to numerical integration. Nevertheless, integration in closed bounded domains is a well studied numerical problem and there exist several routines available for this purpose. Hence we do not expect any complications to be associated with this part of our procedure. For simplicity we take $\mathbb{T} = [0, 1]$.

4.2. Multivariate case

For simplicity, the integration of time $\mathbb{T}$ is approximated by

$$\int_{\mathbb{T}} \widehat{\delta}_t(\boldsymbol{u}) dt \approx \frac{1}{m} \sum_{j=1}^{m} \widehat{\delta}_{t_j}(\boldsymbol{u}).$$

Then we proceed from (8) by using (10) and obtain

$$\widehat{\delta}_{t_j}(\boldsymbol{u}) = \frac{1}{n_1^2} \sum_{i,\ell=1}^{n_1} \cos\{\boldsymbol{u}^\top(\boldsymbol{X}_{1ij} - \boldsymbol{X}_{1\ell j})\} + \frac{1}{n_2^2} \sum_{i,\ell=1}^{n_2} \cos\{\boldsymbol{u}^\top(\boldsymbol{X}_{2ij} - \boldsymbol{X}_{2\ell j})\} - \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{\ell=1}^{n_2} \cos\{\boldsymbol{u}^\top(\boldsymbol{X}_{1ij} - \boldsymbol{X}_{2\ell j})\}.$$

Consequently the test statistic can be written as

$$\Delta_{W,m} = \frac{1}{m} \sum_{j=1}^{m} \left\{ \frac{1}{n_1^2} \sum_{i,\ell=1}^{n_1} I_W(\boldsymbol{X}_{1ij} - \boldsymbol{X}_{1\ell j}) + \frac{1}{n_2^2} \sum_{i,\ell=1}^{n_2} I_W(\boldsymbol{X}_{2ij} - \boldsymbol{X}_{2\ell j}) - \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{\ell=1}^{n_2} I_W(\boldsymbol{X}_{1ij} - \boldsymbol{X}_{2\ell j}) \right\}, \tag{12}$$

where

$$I_W(\boldsymbol{x}) = \int_{\mathbb{R}^p} \cos(\boldsymbol{u}^\top \boldsymbol{x}) W(\boldsymbol{u}) d\boldsymbol{u}. \tag{13}$$

The criterion (12) is the functional version of a maximum mean discrepancy statistic in the sense of [22] with associated kernel $k(x, y) = I_W(x - y)$. In this connection and as already mentioned, the weight function W in (13) may

be chosen in a way that avoids numerical integration, which is problematic in higher dimension. To see this, recall that the CF of any spherical random variable $\boldsymbol{Z}$ is given by $\varphi_{\boldsymbol{Z}}(u) = \Psi(\|\boldsymbol{u}\|)$, for some family-specific scalar function $\Psi(\cdot)$, where $\|\boldsymbol{u}\|$ denotes the usual Euclidean norm. Hence if $f_{\boldsymbol{Z}}(z)$ denotes the density corresponding to $\varphi_{\boldsymbol{Z}}(\boldsymbol{u})$, we have

$$\int_{\mathbb{R}^p} \cos(\boldsymbol{u}^\top z) f_{\boldsymbol{Z}}(z) dz = \Psi(\|\boldsymbol{u}\|).$$

The latter implies that if $f_{\boldsymbol{Z}}$ is used as weight function W in (13), then the resulting test statistic, say $\Delta_{\Psi,m}$, reduces to

$$\Delta_{\Psi,m} = \frac{1}{m}\sum_{j=1}^{m}\left\{\frac{1}{n_1^2}\sum_{i,\ell=1}^{n_1}\Psi(\|\boldsymbol{X}_{1ij}-\boldsymbol{X}_{1\ell j}\|) + \frac{1}{n_2^2}\sum_{i,\ell=1}^{n_2}\Psi(\|\boldsymbol{X}_{2ij}-\boldsymbol{X}_{2\ell j}\|) - \frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{\ell=1}^{n_2}\Psi(\|\boldsymbol{X}_{1ij}-\boldsymbol{X}_{2\ell j}\|)\right\}. \tag{14}$$

The test criterion in (14) is further advanced by considering specific families of spherically symmetric distributions with a simple CF. Such a family of distributions is the family of spherical stable distributions with $\Psi(u) = e^{-u^\alpha}$, where $\alpha \in (0, 2]$ stands for the characteristic exponent. Interesting special cases of spherical stable distributions are the multivariate Cauchy and normal distributions corresponding to $\alpha = 1$ and $\alpha = 2$, respectively.

We will elaborate here on the case of the spherical stable distribution as weight function in (13). In fact we slightly generalize from the spherical to the elliptical stable law and consider as weight function W the stable density with corresponding CF equal to $e^{-\|\boldsymbol{u}\|^\alpha/\gamma}$. With this choice, the test criterion in (12), say $\Delta_{\alpha,\gamma,m}$, takes the form

$$\Delta_{\alpha,\gamma,m} = \frac{1}{m}\sum_{j=1}^{m}\left\{\frac{1}{n_1^2}\sum_{i,\ell=1}^{n_1}\exp(-\|\boldsymbol{X}_{1ij}-\boldsymbol{X}_{1\ell j}\|^\alpha/\gamma) + \frac{1}{n_2^2}\sum_{i,\ell=1}^{n_2}\exp(-\|\boldsymbol{X}_{2ij}-\boldsymbol{X}_{2\ell j}\|^\alpha/\gamma) - \frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{\ell=1}^{n_2}\exp(-\|\boldsymbol{X}_{1ij}-\boldsymbol{X}_{2\ell j}\|^\alpha/\gamma)\right\}. \tag{15}$$

Interestingly there is a connection between the test criterion in (15) and another two-sample test statistic in the literature. To see this, take a two-term expansion $e^{-\|\boldsymbol{x}\|^\alpha/\gamma} = 1 - \|\boldsymbol{x}\|^\alpha/\gamma + \mathrm{o}(\gamma^{-1})$, as $\gamma \to \infty$, in the test statistic (15), and write after some algebra

$$\lim_{\gamma\to\infty}\gamma\Delta_{\alpha,\gamma,m} = \frac{1}{m}\sum_{j=1}^{m}\left\{\frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{\ell=1}^{n_2}\|\boldsymbol{X}_{1ij}-\boldsymbol{X}_{2\ell j}\|^\alpha - \frac{1}{n_1^2}\sum_{i,\ell=1}^{n_1}\|\boldsymbol{X}_{1ij}-\boldsymbol{X}_{1\ell j}\|^\alpha - \frac{1}{n_2^2}\sum_{i,\ell=1}^{n_2}\|\boldsymbol{X}_{2ij}-\boldsymbol{X}_{2\ell j}\|^\alpha\right\}. \tag{16}$$

The criterion on the right-hand side of (16) is the so-called energy statistic of [47] adapted to the functional context. We mention in this connection that energy statistics have gained considerable popularity lately as they have been employed not just for two-sample testing but also for testing for independence as well as in nonparametric analysis of variance. The reader is referred to the review [48] for more information on energy statistics.

We close this section with a discussion on the weight function. To begin with, it is clear from the arguments presented above that this function does not need to be confined within the family of spherical distributions. In fact the density of any multivariate symmetric around zero distribution may be used in (13), provided that it has a simple CF. By way of example, extra choices available are the uniform distribution over the unit hypercube $[-1, 1]^p$, the Laplace distribution, finite mixtures of symmetric stable or mixtures of Laplace distributions, as well as combinations thereof.

Now the more general issue about a 'good', and potentially optimal, choice for the weight function has several underpinnings. Intuitively speaking this choice should depend on the problem under study, but to motivate the discussion we first note that the ECF is a more accurate estimator of the underlying CF when u is near the origin; see, e.g., [14, 36, 52]. While this observation certainly limits the available choices, it clearly leaves lots of room that includes, e.g., all weight functions discussed in this section, as they all assign maximum weight at $\boldsymbol{u} = \boldsymbol{0}$, and progressively fall-off to zero as $\boldsymbol{u} \to \infty$. In this connection, the ECF literature has hitherto been dominated by the extensive use of $W_\gamma(\boldsymbol{u}) = e^{-\|\boldsymbol{u}\|^2/\gamma}$, which probably has its roots in the central role of the normal distribution in statistics; see, e.g., [4, 26]. Nevertheless, W_γ continues to feature prominently even under nonparametric or semiparametric settings such as testing for independence [13] and testing for the error distribution in regression models [34].

A detailed study of potentially power-efficient weight functions for goodness-of-fit testing in finite dimension was undertaken by Lindsay et al. [38]. Their context is more general in that ECF statistics are a special case, but

nevertheless their suggestions for choosing a proper weight function (kernel in their terminology) are mostly about a simple null hypothesis (without estimated parameters) and quite technical and not at all easy to implement in higher dimension. In fact the authors only provide specific results in the case of testing for normality in finite dimension assisted by Monte Carlo, and still these results are for specific directions (alternatives) away from the null hypothesis.

Similar conclusions hold for the work of Tenreiro [49], who narrows down the search of power optimality by fixing the Gaussian function W_γ as weight function and then performs an analysis of proper values of γ for testing multinormality. It becomes clear from the preceding discussion that the whole issue of weight-function choice and optimality is pretty much open even in finite dimension, and that in the current functional context, it defines a formidable research project in its own right. Therefore in what follows we adapt a more pragmatic approach and provide some practical guidelines on the basis of our Monte Carlo results.

4.3. Global case

Following analogous calculations as in (12) the test statistic in (9) may be written as

$$\Upsilon_{W,m} = \frac{1}{n_1^2}\sum_{i,j=1}^{n_1} I_W(\mathbb{X}_{1i}-\mathbb{X}_{1j}) + \frac{1}{n_2^2}\sum_{i,j=1}^{n_2} I_W(\mathbb{X}_{2i}-\mathbb{X}_{2j}) - \frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} I_W(\mathbb{X}_{1i}-\mathbb{X}_{2j}), \tag{17}$$

with I_W resulting from (13) by setting $p = m$. We suggest to use the density of the zero-mean normal distribution $\mathcal{N}_m(\mathbf{0}, \boldsymbol{\Sigma}^{-1})$ with covariance matrix $\boldsymbol{\Sigma}^{-1}$ as weight function, which leads to

$$I_W(\boldsymbol{x}) = e^{-\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{x}/2}. \tag{18}$$

The suggested form of the covariance matrix is specified by letting the elements depend on the observation times as $\sigma_{j,\ell} = \sigma(t_j, t_\ell)$ for all $j, \ell \in \{1, \ldots, m\}$. In this regard setting $\sigma_{j,\ell} = \min(t_j, t_\ell)$ borrows the covariance structure of a Wiener process, while $\sigma_{j,\ell} = e^{-|t_j - t_\ell|/2}$ of an Ornstein–Uhlenbeck process.

Alternatively one may consider a data-driven weight function by replacing $\boldsymbol{\Sigma}$ with an estimate, say $\widehat{\boldsymbol{\Sigma}}_{n_1,n_2}$, of the population covariance matrix of the pooled sample $\mathbb{X}_{11}, \ldots, \mathbb{X}_{1n_1}, \mathbb{X}_{21}, \ldots, \mathbb{X}_{2n_2}$. The advantage of the latter choice is that then the test statistic becomes affine invariant provided that this estimate satisfies $\widehat{\boldsymbol{\Sigma}}_{n_1,n_2}(\boldsymbol{A}\mathbb{X}_{11}+\boldsymbol{b}, \ldots, \boldsymbol{A}\mathbb{X}_{2n_2}+\boldsymbol{b}) = \boldsymbol{A}\widehat{\boldsymbol{\Sigma}}_{n_1,n_2}(\mathbb{X}_{11}, \ldots, \mathbb{X}_{2n_2})\boldsymbol{A}^\top$, for each nonsingular $m \times m$ matrix $\boldsymbol{A}$ and each $\boldsymbol{b} \in \mathbb{R}^m$.

In this connection we note that since the null hypothesis $\mathcal{H}_0 : \boldsymbol{x}_1(t) =_{\mathcal{L}} \boldsymbol{x}_2(t)$, which is equivalent to $\boldsymbol{X}_1(t) =_{\mathcal{L}} \boldsymbol{X}_2(t)$ due to the assumptions in Section 2.3, is closed with respect to full-rank affine transformations, any genuine test statistic $\mathcal{D} = \mathcal{D}(\mathbb{X}_{11}, \ldots, \mathbb{X}_{2n_2})$ based on $\{\mathbb{X}_{1i_1}\}, \{\mathbb{X}_{2i_2}\}$ with $i_k \in \{1, \ldots, n_k\}$ and $k \in \{1, 2\}$, should also be invariant with respect to such transformations, i.e., we should have

$$\mathcal{D}(\boldsymbol{A}\mathbb{X}_{11} + \boldsymbol{b}, \ldots, \boldsymbol{A}\mathbb{X}_{2n_2} + \boldsymbol{b}) \;=\; \mathcal{D}(\mathbb{X}_{11}, \ldots, \mathbb{X}_{2n_2}).$$

A necessary and sufficient condition for a test statistic $\mathcal{D}$ to be affine invariant is that it is based on the so-called 'Mahanalobis squared radii', given for all $i_k \in \{1, \ldots, n_k\}$, $j \in \{1, \ldots, n_\ell\}$, and $k, \ell \in \{1, 2\}$ with $k \le \ell$, by

$$M_{k,\ell}(i_k, j_\ell) = (\mathbb{X}_{k,i_k} - \mathbb{X}_{\ell,j_\ell})^\top \widehat{\boldsymbol{\Sigma}}^{-1}_{n_1,n_2}(\mathbb{X}_{k,i_k} - \mathbb{X}_{\ell,j_\ell}).$$

We close this section by noting that there exist alternative choices of weighting, such as the one suggested in [25] motivated by the Karhunen–Loève expansion.

4.4. Resampling procedures

As already shown in Section 3, the null distribution of the test statistics considered depends, among other things, on the underlying stochastic properties of the random fields $\boldsymbol{x}_1(t)$ and $\boldsymbol{x}_2(t)$ involved. In order to deal with this issue, we apply appropriate resampling procedures for computing critical points and actually carrying out the tests. To this end, let $\mathcal{D} = \mathcal{D}(\xi_1, \ldots, \xi_n)$ be a generic notation for a test statistic which depends on a sample of size n of observations $\xi_1, \ldots, \xi_n$. Clearly in our case $n = n_1 + n_2$. We will apply the permutation procedure whereby we randomly generate a permutation $b = (b_1, \ldots, b_n)$ of $(1, \ldots, n)$, and compute the test statistic $\mathcal{D}_b = \mathcal{D}(\xi_{b_1}, \ldots, \xi_{b_n})$. The procedure is

repeated for each $b \in \{1, \ldots, B\}$, and the critical point of the test of size α is determined as the corresponding $1 - \alpha$ quantile $\mathcal{D}_{(1-\alpha)B}$ of the values $\mathcal{D}_1, \ldots, \mathcal{D}_B$. The null hypothesis is then rejected if $\mathcal{D} > \mathcal{D}_{(1-\alpha)B}$.

For univariate data, the critical point of the test statistic in (5) is computed by permuting random functions $\{\widehat{x}_{11}(t), \ldots, \widehat{x}_{1n_1}(t), \widehat{x}_{21}(t), \ldots, \widehat{x}_{2n_2}(t)\}$ for $t \in \mathbb{T}$. With multivariate data, defining $\mathbb{X}_{kj} = (X_{kj1}, \ldots, X_{kjm})$, permutations for the criterion in (7) are performed on $\{\mathbb{X}_{11}, \ldots, \mathbb{X}_{1n_1}, \mathbb{X}_{21}, \ldots, \mathbb{X}_{2n_2}\}$. Note that permutation tests for the two-sample problem have already been successfully applied in the context of high-dimensional/functional data; see, e.g., [5, 25].

5. Simulation study

5.1. Study 1: Univariate case

In the univariate case, we generate data $\{(t_{kij}, X_{kij}) : j \in \{1, \ldots, m_{ki}\}\}_{i=1}^{n_k}$ for $k \in \{1, 2\}$ as follows. The sampling design for the curves is assumed balanced, i.e., for all i, $m_{1i} = m_{2i} = m$, and regular. Specifically, suppose that for all $k \in \{1, 2\}$ and $i \in \{1, \ldots, n_k\}$, $t_{ki\cdot}$ are discrete uniform fixed time points on $[0, 1]$. It is assumed that

$$x_{1i}(t) = \sum_{k=1}^{15} e^{-k/2} N_{k1i} \psi_k(t), \quad x_{2i}(t) = \sum_{k=1}^{15} e^{-k/2} N_{k21i} \psi_k(t) + \delta \sum_{k=1}^{15} k^{-2} N_{k22i} \psi_k^*(t),$$

where $N_{k1i}, N_{k21i}, N_{k22i}$ are iid $\mathcal{N}(0, 1)$ variables, and $\delta \geq 0$ controls the deviation from the null hypothesis; we have $\delta = 0$ under $\mathcal{H}_0$. Here $\psi_1(t) \equiv 1$ for all t and $\psi_k(t) = \sqrt{2}\sin\{(k-1)\pi t\}$ are orthonormal basis functions. Also

$$\psi_k^*(t) = \begin{cases} 1 & \text{if } k = 1, \\ \sqrt{2}\sin\{(k-1)\pi(2t-1)\} & \text{if } k \text{ is odd and } k > 1, \\ \sqrt{2}\cos\{(k-1)\pi(2t-1)\} & \text{if } k \text{ is even,} \end{cases}$$

are orthonormal basis functions. The level of significant is set to 0.05 with sample size $n_1 = n_2 \in \{15, 25, 50\}$, and $m \in \{21, 101\}$ time points for each curve. The corresponding inner integral in (11) is computed by averaging the function values over the grid of observation time points. We also tried the versions of the test criterion in (15) that correspond to characteristic exponent $\alpha \in \{0.5, 1.0, 1.5, 2.0\}$, always with $\gamma = 1$. The distribution of each test statistic is approximated by the empirical distribution that is based on $B = 1000$ permutations.

To evaluate the performance of test statistics, we carry out our simulations under two scenarios.

Scenario 1*: The sample noise has the same distribution in the two groups*, i.e., $\varepsilon_1(t) =_{\mathcal{L}} \varepsilon_2(t)$. Due to Fourier identities, there is no need to estimate $\hat{x}_{ki}(t)$, and the testing procedure can be used for X_{kij}. We assumed $\varepsilon_{kij} \sim_{\text{iid}} \mathcal{N}(0, 0.01)$ for all $k \in \{1, 2\}$, $i \in \{1, \ldots, n_k\}$ and $j \in \{1, \ldots, m\}$. The test criterion is denoted by $D_{\alpha,m}$ for simplicity.

Scenario 2*: The sample noise is non-identically distributed between the two groups*. Recall from Section 4.1 that we estimate $x_{1i}(t)$ and $x_{2i}(t)$ by means of local-linear smoothing. The bandwidth is selected by minimizing the cross-validation objective function

$$h = \arg\min_h \left\{ \frac{1}{mn_1} \sum_{i=1}^{n_1} \sum_{j=1}^{m} (X_{1ij} - \widehat{x}_{1ij})^2 + \frac{1}{mn_2} \sum_{i=1}^{n_2} \sum_{j=1}^{m} (X_{2ij} - \widehat{x}_{2ij})^2 \right\},$$

and the test criterion is denoted by $D_{\alpha,m}$. The kernel function is the biquadratic kernel, $K(u) = (15/16)(1 - u^2)^2 \mathbf{1}(|u| \leq 1)$. Without loss of generality, we assumed that $\varepsilon_{1ij} \sim_{\text{iid}} \mathcal{N}(0, \sigma_1^2)$ and $\varepsilon_{2ij} \sim_{\text{iid}} \mathcal{N}(0, \sigma_2^2)$. Then in order to evaluate the influence of the distribution of the sample noise on the test statistics, three pairs of distributions between the two groups are considered: (i) $(\sigma_1, \sigma_2) = (0.01, 0.1)$; (ii) $(\sigma_1, \sigma_2) = (0.1, 0.3)$; (iii) $(\sigma_1, \sigma_2) = (0.1, 0.5)$.

5.1.1. Scenario 1: The sample noise has the same distribution in both groups

Table 1 and Figure 1 display the Type I error and power rates based on simulations when the sample noise distribution is the same for the two groups. The results illustrate that the test controls size well under the null hypothesis with $\delta = 0$ and the power increases for larger values of time points m, sample size $n_1 = n_2$, and the distance parameter δ. Now since our design is identical to that in [25], we were able to compare our results with those therein and found that

Table 1: Rejection rate for the ECF test $D_{\alpha,m}$ at 5% significance based on 1000 permutations when the distribution of the sample noise is the same in both groups.

Sample Size	Time Points	Distance Parameter: δ						
$n_1 = n_2$	m	0	0.2	0.4	0.6	0.8	1	2
				$\alpha = 0.5$				
15	20	0.038	0.061	0.078	0.218	0.634	0.922	1
	100	0.050	0.034	0.116	0.618	0.982	1	1
25	20	0.058	0.052	0.13	0.406	0.908	1	1
	100	0.064	0.044	0.238	0.918	1	1	1
50	20	0.054	0.052	0.192	0.854	1	1	1
	100	0.040	0.062	0.522	1	1	1	1
				$\alpha = 1$				
15	20	0.058	0.048	0.082	0.242	0.652	0.966	1
	100	0.046	0.086	0.136	0.670	0.996	1	1
25	20	0.050	0.062	0.120	0.462	0.958	1	1
	100	0.036	0.076	0.252	0.946	1	1	1
50	20	0.066	0.054	0.206	0.912	1	1	1
	100	0.056	0.046	0.600	1	1	1	1
				$\alpha = 1.5$				
15	20	0.058	0.054	0.080	0.258	0.696	0.976	1
	100	0.046	0.076	0.160	0.728	1	1	1
25	20	0.050	0.070	0.120	0.494	0.976	1	1
	100	0.040	0.076	0.276	0.966	1		1
50	20	0.066	0.052	0.218	0.932	1	1	1
	100	0.060	0.046	0.634	1	1	1	1
				$\alpha = 2$				
15	20	0.046	0.072	0.144	0.388	0.804	0.980	1
	100	0.054	0.094	0.360	0.918	1	1	1
25	20	0.058	0.092	0.226	0.696	0.982	1	1
	100	0.066	0.146	0.542	0.988	1	1	1
50	20	0.074	0.124	0.518	0.980	1	1	1
	100	0.132	0.342	0.976	1	1	1	1

the ECF test statistic is more powerful than the test suggested in that paper. This should not be surprising, since with the ECF method we do not need to estimate the basis functions by smoothing the data when observations are without noise or sampling noise is equidistributed.

Regarding the value of the characteristic exponent α, we observe a nearly monotonic increase of the rejection rate with increasing α. This may temp us to favor test statistics with a larger α. However, a higher rejection rate holds not only under alternatives but also under the null hypothesis, and hence one should weigh better power against the tendency to overshoot the nominal Type I error for larger values of α.

5.1.2. Scenario 2: The sample noise is not identically distributed in the two groups

Tables 2–4 and Figure 2 display the Type I error and power rates based on simulations when the two groups have non-identical sample noise distributions. The results again illustrate that our ECF method performs reasonably well for this scenario too. At the same time, we point out the following findings.

(i) The Type I error grows as the standard deviation of ε_2 increases. Intuitively, the higher the standard deviation, the greater the difference in the observed data, which may lead to higher probability of rejection even when $\mathcal{H}_0$ is true.

(ii) In most cases of Scenario 2, the percentage of rejection drops for larger time points m. This is different from the behavior shown under Scenario 1. The cause of this result may relate to the smoothing procedure. Specifi-

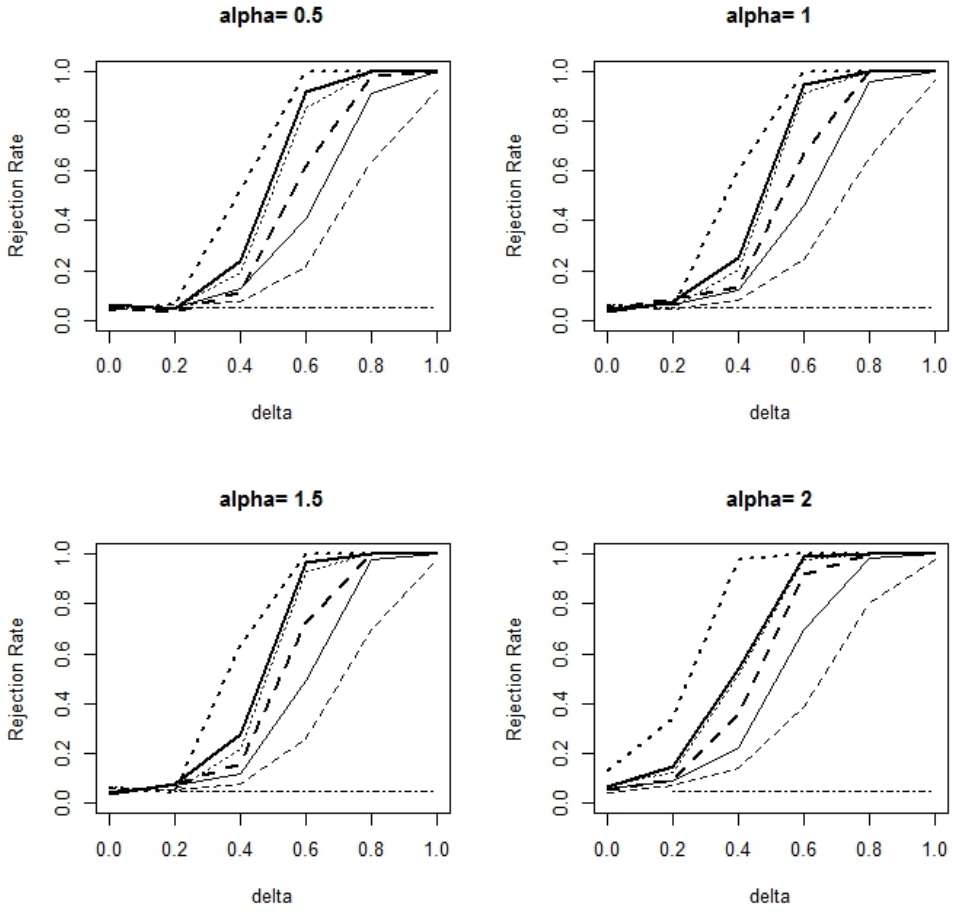

Figure 1: Rejection rate for the ECF test $D_{\alpha,m}$ at 5% significance based on 1000 permutations when the sample noise has the same distribution in the two groups with $n_1 = n_2 = 15$ (dashed line), $n_1 = n_2 = 25$ (solid line) and $n_1 = n_2 = 50$ (dotted line). The thin lines correspond to $m = 20$, and the thick lines to $m = 100$. The null hypothesis holds when $\delta = 0$.

Table 2: Rejection rate for the ECF test $D_{\alpha,m}$ at 5% significance based on 1000 permutations when $\varepsilon_{1ij} \sim_{\text{iid}} \mathcal{N}(0, 0.01^2)$ and $\varepsilon_{2ij} \sim_{\text{iid}} \mathcal{N}(0, 0.1^2)$.

Sample Size	Time Points	Distance Parameter: δ						
$n_1 = n_2$	m	0	0.5	1	1.5	2	3	4
				$\alpha = 0.5$				
15	20	0.044	0.078	0.452	0.894	0.998	1	1
	100	0.052	0.082	0.454	0.912	0.996	1	1
25	20	0.042	0.100	0.772	1	1	1	1
	100	0.042	0.100	0.722	0.998	1	1	1
50	20	0.048	0.230	0.988	1	1	1	1
	100	0.050	0.196	0.994	1	1	1	1
				$\alpha = 1$				
15	20	0.064	0.068	0.378	0.912	1	1	1
	100	0.052	0.068	0.366	0.866	0.998	1	1
25	20	0.052	0.096	0.710	0.992	1	1	1
	100	0.042	0.098	0.680	0.994	1	1	1
50	20	0.048	0.180	0.99	1	1	1	1
	100	0.052	0.164	0.982	1	1	1	1
				$\alpha = 1.5$				
15	20	0.064	0.056	0.400	0.908	0.998	1	1
	100	0.038	0.050	0.230	0.706	0.970	1	1
25	20	0.050	0.066	0.690	0.996	1	1	1
	100	0.058	0.066	0.422	0.970	1	1	1
50	20	0.040	0.140	0.988	1	1	1	1
	100	0.024	0.128	0.914	1	1	1	1
				$\alpha = 2$				
15	20	0.068	0.056	0.310	0.998	0.998	1	1
	100	0.046	0.050	0.104	0.292	0.740	1	1
25	20	0.044	0.084	0.588	0.996	1	1	1
	100	0.032	0.058	0.128	0.656	0.988	1	1
50	20	0.042	0.094	0.970	1	1	1	1
	100	0.046	0.070	0.396	0.998	1	1	1

cally the variance of estimation is $O\{1/(mh)\}$ in the nonparametric smoothing method, which suggests that the estimation of x_{kij} with a small number of time points m is more likely to fluctuate and hence leads to rejection of $\mathcal{H}_0$ more often than with larger m. The same type of reverse behavior is observed with respect to the value of the characteristic exponent, with larger α now leading to a drop in the percentage of rejection. Thus an overall suggestion of using more data points m and a value of α that is not too small would lead to a compromise test statistic that has reasonably good power but also respects the nominal level of significance.

5.1.3. Comparison with the method of Horváth et al. [28]

We also compare the ECF method and the method HKR in [28] under the same simulation settings, viz.

$$X_{1i}(t) = \mu_1(t) + \varepsilon_{1i}(t), \quad X_{2i}(t) = \mu_2(t) + \varepsilon_{2i}(t),$$

where for all $k \in \{1, 2\}$ and $i \in \{1, \ldots, n_k\}$, $\varepsilon_{ki}(t)$ with $t \in (0, 1)$ are iid Brownian bridges, and $\mu_1(t) = 0$, while $\mu_2(t) = \delta t(1 - t)$. The sample size is set to $n_1 = n_2 = 50$, and the number of time points is $m = 10$ or $m = 30$. The sampling noise then has the same distribution in both groups. Using Fourier identities, we see that the null hypothesis $\mathcal{H}_0$ in (2) holds if and only if

$$\forall_{t \in \mathbb{T}} \; \forall_{\boldsymbol{u} \in \mathbb{R}^p} \; \mathrm{E}[\sin\{\boldsymbol{u}^\top \boldsymbol{Y}(t)\}] = 0,$$

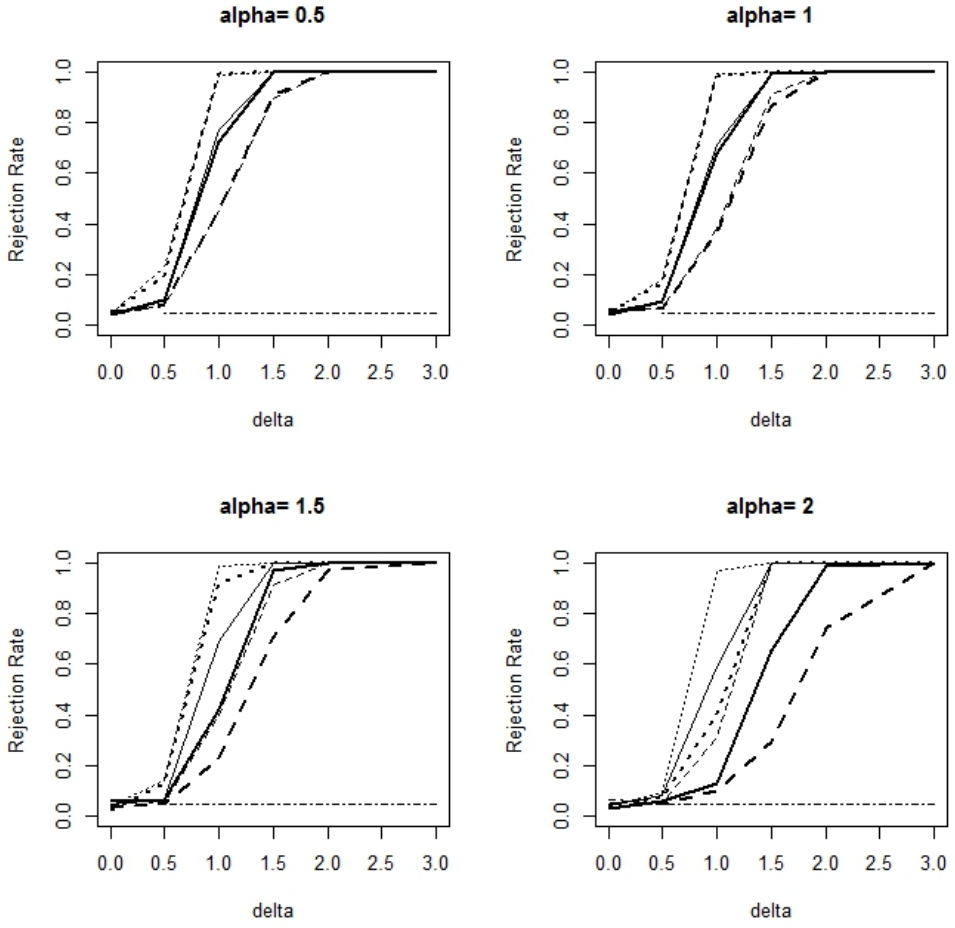

Figure 2: Rejection rate for the ECF test $D_{\alpha,m}$ at 5% significance based on 1000 permutations when the distribution of the sample noise is different between the two groups ($\varepsilon_{1ij} \sim_{\text{iid}} \mathcal{N}(0, 0.01^2)$, $\varepsilon_{2ij} \sim_{\text{iid}} \mathcal{N}(0, 0.1^2)$) with $n_1 = n_2 = 15$ (dashed line), $n_1 = n_2 = 25$ (solid line) and $n_1 = n_2 = 50$ (dotted line). The thin lines correspond to $m = 20$, and the thick lines to $m = 100$. The null hypothesis holds for $\delta = 0$.

Table 3: Rejection rate for the ECF test $D_{\alpha,m}$ at 5% significance based on 1000 permutations when $\varepsilon_{1ij} \sim_{\text{iid}} \mathcal{N}(0, 0.1^2)$ and $\varepsilon_{2ij} \sim_{\text{iid}} \mathcal{N}(0, 0.3^2)$.

Sample Size	Time Points	Distance Parameter: δ						
$n_1 = n_2$	m	0	0.5	1	1.5	2	3	4
				$\alpha = 0.5$				
15	20	0.070	0.090	0.484	0.908	0.996	1	1
	100	0.056	0.080	0.488	0.896	0.996	1	1
25	20	0.054	0.150	0.774	0.992	1	1	1
	100	0.042	0.142	0.744	0.998	1	1	1
50	20	0.050	0.276	0.992	1	1	1	1
	100	0.054	0.286	0.994	1	1	1	1
				$\alpha = 1$				
15	20	0.076	0.088	0.470	0.922	0.998	1	1
	100	0.060	0.066	0.418	0.864	0.994	1	1
25	20	0.058	0.126	0.784	0.992	1	1	1
	100	0.044	0.122	0.682	0.998	1	1	1
50	20	0.042	0.264	0.994	1	1	1	1
	100	0.054	0.238	0.992	1	1	1	1
				$\alpha = 1.5$				
15	20	0.068	0.076	0.424	0.914	0.998	1	1
	100	0.054	0.062	0.244	0.682	0.974	1	1
25	20	0.050	0.104	0.738	0.992	1	1	1
	100	0.038	0.090	0.450	0.982	1	1	1
50	20	0.046	0.228	0.994	1	1	1	1
	100	0.052	0.142	0.940	1	1	1	1
				$\alpha = 2$				
15	20	0.070	0.072	0.338	0.898	0.998	1	1
	100	0.046	0.062	0.090	0.282	0.746	1	1
25	20	0.048	0.096	0.648	0.990	1	1	1
	100	0.032	0.068	0.148	0.698	0.990	1	1
50	20	0.046	0.178	0.978	1	1	1	1
	100	0.040	0.062	0.450	1	1	1	1

where $\boldsymbol{Y}(t) = \boldsymbol{X}_2(t) - \boldsymbol{X}_1(t)$. Then a test statistic analogous to (15) for $\gamma = 1$ is given by

$$\Psi_{\alpha,m} = \frac{1}{m} \sum_{j=1}^{m} \frac{1}{2n^2} \sum_{i,\ell=1}^{n} \{e^{-|Y_i(t_j) - Y_\ell(t_j)|^\alpha} - e^{-|Y_i(t_j) + Y_\ell(t_j)|^\alpha}\}, \tag{19}$$

where $Y_i(t) = X_{2i}(t) - X_{1i}(t)$ for all $i \in \{1, \ldots, n\}$. Table 5 corresponds to $\alpha = 1$ for the test statistic in (19), and to the criterion U_1 of [28].

The results suggest that the performance of the ECF method is comparable to that of the HKR method, while at the same time the ECF is seen to be more robust with respect to the number of points m. However, recall that our test criterion in (19) refers to marginal symmetry while the HKR method is within the full functional context.

5.2. Study 2: Multivariate case

We simulate two random pairs with the same noise distribution with the following setup borrowed from [32]:

$$\boldsymbol{X}_1(t) = \begin{pmatrix} X_{11}(t) \\ X_{12}(t) \end{pmatrix} = \begin{pmatrix} U_3 h_2(t) \\ U_1 h_1(t) + U_3 h_3(t) \end{pmatrix} + \begin{pmatrix} \sqrt{0.1}\, \varepsilon_{11}(t) \\ \sqrt{0.5}\, \varepsilon_{12}(t) \end{pmatrix},$$

$$\boldsymbol{X}_2(t) = \begin{pmatrix} X_{21}(t) \\ X_{22}(t) \end{pmatrix} = \begin{pmatrix} U_3 h_2(t) \\ U_1 h_1(t) + U_3 h_3(t) \end{pmatrix} + \delta \begin{pmatrix} U_2 h_3(t) \\ U_2 h_2(t) \end{pmatrix} + \begin{pmatrix} \sqrt{0.1}\, \varepsilon_{21}(t) \\ \sqrt{0.5}\, \varepsilon_{22}(t) \end{pmatrix},$$

Table 4: Rejection rate for the ECF test $D_{\alpha,m}$ at 5% significance based on 1000 permutations when $\varepsilon_{1ij} \sim_{\text{iid}} \mathcal{N}(0, 0.1^2)$ and $\varepsilon_{2ij} \sim_{\text{iid}} \mathcal{N}(0, 0.5^2)$.

Sample Size	Time Points	Distance Parameter: δ						
$n_1 = n_2$	m	0	0.5	1	1.5	2	3	4
				$\alpha = 0.5$				
15	20	0.084	0.152	0.536	0.948	0.988	1	1
	100	0.062	0.118	0.536	0.914	0.994	1	1
25	20	0.120	0.242	0.838	0.998	1	1	1
	100	0.090	0.224	0.872	1	1	1	1
50	20	0.172	0.484	0.998	1	1	1	1
	100	0.148	0.534	1	1	1	1	1
				$\alpha = 1$				
15	20	0.088	0.136	0.524	0.950	0.994	1	1
	100	0.064	0.088	0.454	0.896	0.992	1	1
25	20	0.104	0.204	0.838	1	1	1	1
	100	0.090	0.190	0.796	1	1	1	1
50	20	0.128	0.500	0.996	1	1	1	1
	100	0.136	0.478	1	1	1	1	1
				$\alpha = 1.5$				
15	20	0.092	0.102	0.448	0.950	0.996	1	1
	100	0.048	0.068	0.286	0.728	0.988	1	1
25	20	0.076	0.176	0.818	0.996	1	1	1
	100	0.070	0.106	0.592	0.992	1	1	1
50	20	0.120	0.448	1	1	1	1	1
	100	0.084	0.252	0.994	1	1	1	1
				$\alpha = 2$				
15	20	0.078	0.086	0.372	0.934	0.994	1	1
	100	0.052	0.060	0.118	0.312	0.794	1	1
25	20	0.082	0.154	0.770	1	1	1	1
	100	0.062	0.076	0.182	0.770	0.992	1	1
50	20	0.086	0.370	0.996	1	1	1	1
	100	0.068	0.100	0.576	1	1	1	1

where the three basic functions are respectively

$$h_1(t) = (1 - |10t - 5|)_+ , \quad h_2(t) = (1 - |10t - 3|)_+ , \quad h_3(t) = (1 - |10t - 7|)_+ ,$$

and $U_1 \sim \mathcal{N}(0.5, 1/12)$, $U_2 \sim \mathcal{N}(1, 1/12)$, $U_3 \sim \mathcal{N}(0, 2/3)$. For the noise distribution, we have $\varepsilon_{11}, \varepsilon_{12}, \varepsilon_{21}, \varepsilon_{22} \sim \mathcal{N}(0, 1)$, with $\delta \geq 0$ again controlling deviation from the null hypothesis; we have $\delta = 0$ under $\mathcal{H}_0$.

Table 6 shows the Type I error and power rates based on simulations when samples are bivariate and the distribution of the sample noise is the same in two groups. The results are much the same as those in the one-dimensional case displayed in Table 1, only now the test criterion is less sensitive to the choice of the characteristic exponent α.

5.3. Study 3: Global test

Given the observation time points $t_j = j/m$ with $j \in \{0, \ldots, m\}$, we simulated $X_1(t)$ as a Wiener process and set $X_2(t) = (1 + \delta)Z(t)$, where $Z(t)$ is another Wiener process, independent of $X_1(t)$. Thus $\delta = 0$ corresponds to the null hypothesis. Observe that here observations are assumed to be without noise. The level of significant is 0.05, with sample size $n_1 = n_2 = n$, where $n \in \{15, 25, 50\}$, and $m \in \{15, 25\}$, for each curve. We consider the test statistic in (17)–(18), with the matrix $\mathbf{\Sigma}$ corresponding to a Wiener process. The distribution of this test statistic was approximated by the empirical distribution based on $B = 1000$ permutations and the results are shown in Table 7.

Another round of results are shown in Table 8 for the global test statistic when the two samples are as above but with X_1 and Z generated from two independent Ornstein–Uhlenbeck processes and using an Ornstein–Uhlenbeck

Table 5: Rejection rate for the ECF test $\Psi_{\alpha,m}$, $\alpha = 1$, and the HKR method at 5% significance with m time points.

δ	$m = 10$		$m = 30$	
	ECF	HKR	ECF	HKR
0.0	0.052	0	0.048	0.061
0.1	0.058	0	0.078	0.071
0.2	0.098	0	0.096	0.108
0.3	0.122	0	0.132	0.174
0.4	0.206	0	0.188	0.246
0.5	0.268	0	0.284	0.331
0.6	0.358	0	0.382	0.448
0.7	0.470	0	0.498	0.530
0.8	0.578	0	0.610	0.665
0.9	0.634	0	0.712	0.751
1.0	0.770	0	0.800	0.811
1.1	0.812	0	0.854	0.889
1.2	0.910	0	0.896	0.928
1.3	0.926	0	0.952	0.960

covariance matrix $\mathbf{\Sigma}$ in the test statistic (17)–(18). Both sets of results are consistent and qualitatively analogous to our previous findings, showing a good control of the level and reasonable behavior with respect to the sample size n, as well as with respect to the distance parameter δ and the number m of observational times. As already mentioned in Section 4.3, a test statistic with the covariance matrix $\mathbf{\Sigma}$ in (18) being data-dependent is also possible, but too time consuming as in this case a high-dimensional matrix $\mathbf{\Sigma}$ and its inverse must be computed for each Monte Carlo sample.

Table 6: Rejection rate for ECF test $\Delta_{\alpha,m}$ at 5% significance based on 1000 permutations when samples are two-dimensional data and sample noise is equidistributed between the two groups.

Sample Size	Time Points	Distance Parameter: δ					
$n_1 = n_2$	m	0	0.2	0.4	0.6	0.8	1
				$\alpha = 0.5$			
15	20	0.038	0.076	0.178	0.458	0.804	0.966
	100	0.048	0.122	0.448	0.970	1	1
50	20	0.054	0.144	0.792	0.998	1	1
	100	0.046	0.428	1	1	1	1
				$\alpha = 1$			
15	20	0.052	0.080	0.186	0.468	0.818	0.97
	100	0.046	0.118	0.458	0.972	1	1
50	20	0.054	0.152	0.794	0.998	1	1
	100	0.046	0.429	1	1	1	1
				$\alpha = 1.5$			
15	20	0.042	0.070	0.198	0.486	0.818	0.968
	100	0.054	0.108	0.480	0.940	1	1
50	20	0.050	0.162	0.794	0.998	1	1
	100	0.044	0.444	1	1	1	1
				$\alpha = 2$			
15	20	0.042	0.068	0.186	0.490	0.798	0.954
	100	0.054	0.108	0.486	0.978	1	1
50	20	0.042	0.158	0.784	0.998	1	1
	100	0.042	0.400	1	1	1	1

Table 7: Rejection rate for ECF test $\Upsilon_{W,m}$ at 5% significance based on 1000 permutations when samples come from Wiener processes.

Sample Size	Time Points	Distance Parameter: δ					
$n_1 = n_2$	m	0	0.1	0.2	0.3	0.4	0.5
15	15	0.060	0.060	0.190	0.270	0.410	0.570
	25	0.050	0.090	0.150	0.420	0.520	0.640
25	15	0.040	0.140	0.180	0.410	0.650	0.700
	25	0.050	0.080	0.330	0.530	0.630	0.760
50	15	0.060	0.090	0.460	0.670	0.930	0.950
	25	0.060	0.160	0.480	0.670	0.830	0.920

Table 8: Rejection rate for ECF test $\Upsilon_{W,m}$ at 5% significance based on 1000 permutations when samples come from Ornstein–Uhlenbeck processes.

Sample Size	Time Points	Distance Parameter: δ					
$n_1 = n_2$	m	0	0.1	0.2	0.3	0.4	0.5
15	15	0.045	0.065	0.145	0.250	0.370	0.550
	25	0.055	0.095	0.210	0.355	0.570	0.590
25	15	0.045	0.095	0.175	0.355	0.625	0.720
	25	0.060	0.115	0.285	0.490	0.815	0.840
50	15	0.060	0.140	0.325	0.670	0.890	0.930
	25	0.050	0.140	0.405	0.700	0.810	0.890

5.4. Application 1: MCO data

In order to simulate reduced blood flow for a period of one hour, Ruiz-Meana et al. [45] measured the mithochondiral calcium overload (MCO) in two groups (control and treatment) every 10 seconds during an hour in isolated mouse cardiac cells. For the treatment group, a dose of cariporide was administered, which is believed to inhibit cell death due to oxidative stress. In fact, for technical reasons, the original experiment was performed twice, using both the "intact", original cells and "permeabilized" cells, a condition related to the mitochondrial membrane. The data were made available by Febrero-Bande and Oviedo de la Fuente [15] through the R package `fda.usc`.

Figure 3 illustrates that the two groups of data shown, one with intact cells and the other group with permeabilized cells, may or may not behave in the same way, but a definite conclusion is not easily reached by simple visual inspection. We applied our ECF test $D_{\alpha,m}$, $\alpha = 1$, $m = 342$, together with the HKR statistic in [28]. The resulting pairs of p-values shown in Table 9 lead to rejection of the null hypothesis for both the intact cells and the permeabilized cells, a conclusion consistent with the conclusion drawn in [21].

Table 9: Rejection rate for the ECF test $D_{\alpha,m}$, $\alpha = 1$, and the HKR method to test the equality between the treatment and control calcium concentrations curves in the intact and permeabilized experiments.

	ECF	HKR
Intact	$9e^{-5}$	0
Permeabilized	0.004	0

5.5. Application 2: Australian temperature data

We apply the proposed method to temperature data collected by 224 weather stations across Australia; see [25]. The data set consists of monthly mean maximum temperatures in degrees Celsius. Figure 4 shows the temperature curves of the 224 weather stations during 1933, 1953, 1973 and 1993. Our interest is whether weather patterns in Australia have changed during this time period. In order to test this hypothesis, the time period is split into four periods, namely from 1914 to 1933 (period 1), from 1934 to 1953 (period 2), from 1954 to 1973 (period 3) and from

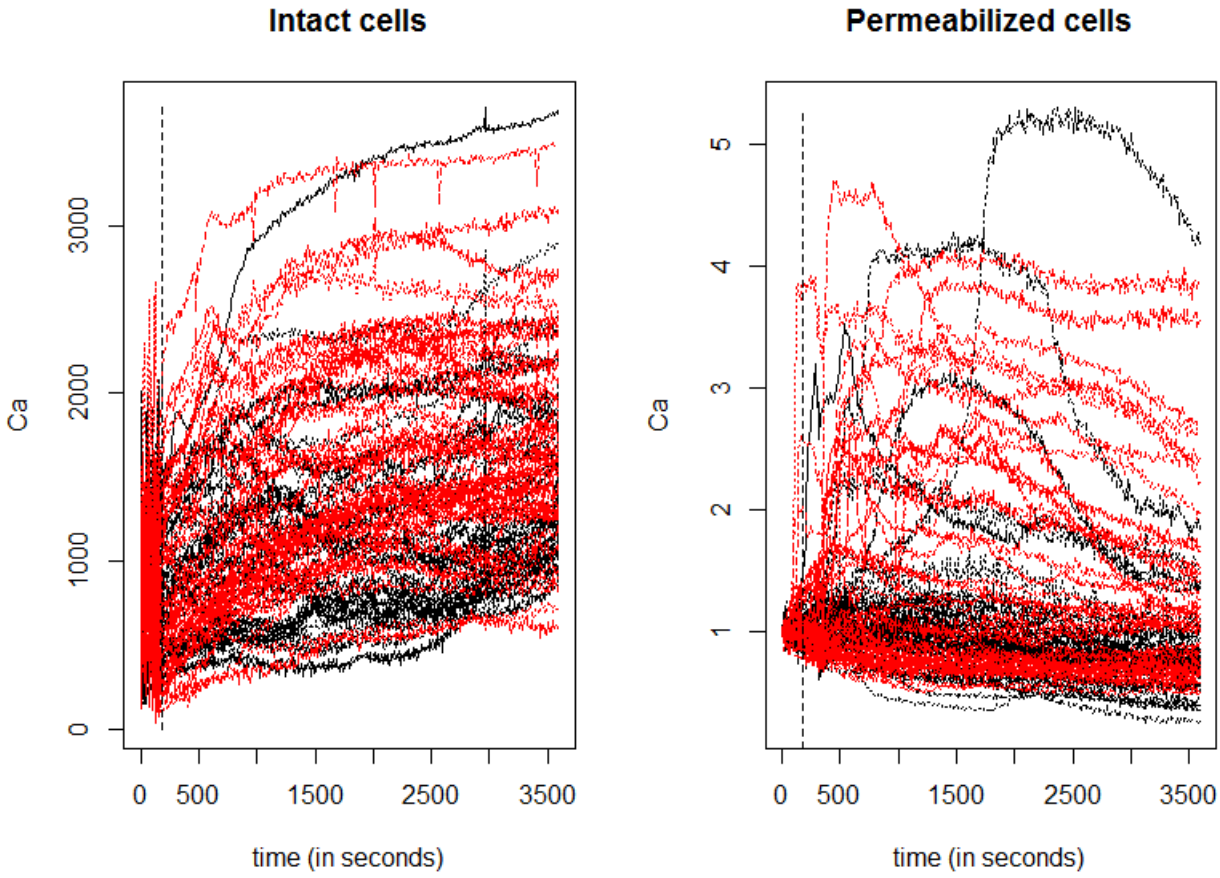

Figure 3: The left panel shows data in control and treatment groups of intact cells. The right panel shows data in control and treatment groups of permeabilized cells. The black line represents the control group and the red line represents the treatment group.

1974 to 1993 (period 4). Then we carry out the proposed pairwise tests on each of the six possible pairs with our criterion $D_{\alpha,m}$, $\alpha = 1$, $m = 12 \times 20$.

The corresponding p-values are reported in Table 10. At the 5% level, two comparisons (period 1 vs. 4, and period 2 vs. 4) lead to significant p-values. Hence our test suggests that the mean maximum temperature in Australia has changed over time, and specifically more significantly so in the last period; see [25] for analogous findings.

Table 10: Rejection rate for the ECF test $D_{\alpha,m}$, $\alpha = 1$, of the pairwise tests based on the Australian weather data.

Period	Period	p-value
1	2	0.0553
1	3	0.0519
1	4	0.0199
2	3	0.0521
2	4	0.0391
3	4	0.1296

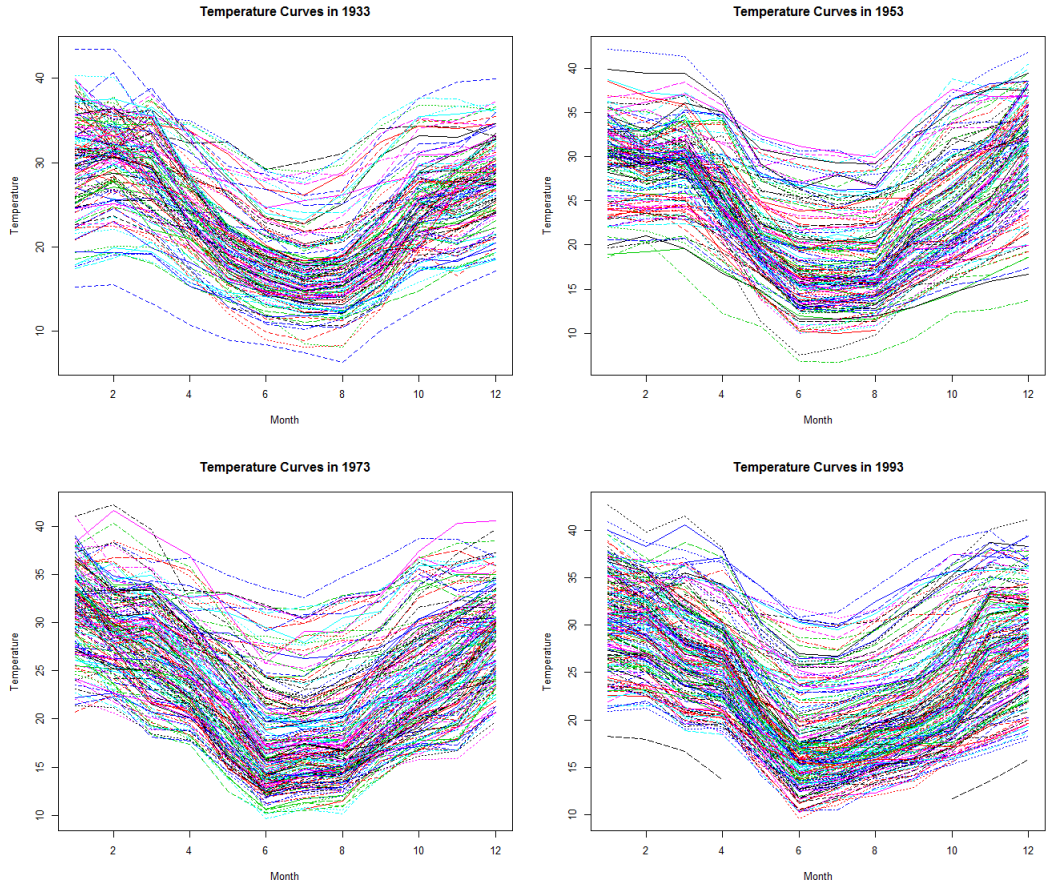

Figure 4: Temperature curves for the 224 weather stations in 1933 (top left), 1953 (top right), 1973 (bottom left), 1993 (bottom right).

6. Conclusion

In this paper, we suggested a new procedure for testing the two-sample null hypothesis with functional data. The procedure measures the distance between (marginal or joint) characteristic functions rather than the classical distance between distribution functions. We studied the limit behavior of the new test criteria both under the null and under alternatives. Moreover our simulations show that specific versions of the proposed method work well in different sampling scenarios, in that they control the nominal size well under the null and at the same time show good power performance that is enhanced with larger sample size and distance from the null hypothesis. The method is an adaptation to the functional data set up of earlier Fourier methods for the same problem with perfectly observed iid data and consequently inherits certain desirable properties from that context such as computational convenience, and the possibility of immediate extension to multidimensional observations, which is not always true if one employs classical procedures, and good performance vis-à-vis the latter procedures whenever these are applicable. A couple of real data examples illustrate the applicability of the new procedures.

Appendix: Proofs

Here we present auxiliary results for the proofs of Theorems 1–4. In what follows D stands for a generic constant. Denoting $g(z) = \cos(z) + \sin$ for all $z \in \mathbb{R}$, and using elementary properties of sin and cos, we find

$$\Delta_{W,m} = \frac{1}{m}\sum_{j=1}^{m}\int_{\mathbb{R}^p}\left\{\frac{1}{n_1}\sum_{i=1}^{n_1} g(\boldsymbol{u}^\top \boldsymbol{X}_{1ij}) - \frac{1}{n_2}\sum_{i=1}^{n_2} g(\boldsymbol{u}^\top \boldsymbol{X}_{2ij})\right\}^2 W(\boldsymbol{u})d\boldsymbol{u}$$

and

$$D_{W,m} = \frac{1}{m}\sum_{j=1}^{m}\int_{\mathbb{R}}\left\{\frac{1}{n_1}\sum_{i=1}^{n_1} g(u\widehat{x}_{1ij}) - \frac{1}{n_2}\sum_{i=1}^{n_2} g(u\widehat{x}_{2ij})\right\}^2 W(u)du.$$

At every $u \in \mathbb{R}^p$ and $t \in \lfloor 0, 1\rfloor$, we define the processes

$$V_{n_1}(\boldsymbol{u}, t) = \frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}[g\{\boldsymbol{u}^\top \boldsymbol{X}_{1i}(t)\} - \mathrm{E}g\{\boldsymbol{u}^\top \boldsymbol{X}_{1i}(t)\}], \quad V_{n_2}(\boldsymbol{u}, t) = \frac{1}{\sqrt{n_2}}\sum_{i=1}^{n_2}[g\{\boldsymbol{u}^\top \boldsymbol{X}_{2i}(t)\} - \mathrm{E}g\{\boldsymbol{u}^\top \boldsymbol{X}_{2i}(t)\}],$$

and $V_{n_1,n_2}(\boldsymbol{u}, t) = \sqrt{(n_1+n_2)/n_1}V_{n_1}(\boldsymbol{u}, t) - \sqrt{(n_1+n_2)/n_2}V_{n_2}(\boldsymbol{u}, t)$.

Observe that under Assumptions (A.1)–(A.7), the random variables $V_{n_1}(\boldsymbol{u}, t)$ and $V_{n_2}(\boldsymbol{u}, t)$ are, for fixed $\boldsymbol{u}$ and t, sums of iid random variables with zero mean, finite variances $\mathrm{var}\{g(\boldsymbol{u}^\top \boldsymbol{X}_{1i}(t))\}$ and $\mathrm{var}\{g(\boldsymbol{u}^\top \boldsymbol{X}_{2i}(t))\}$, respectively, and if additionally the null hypothesis holds true, then $\mathrm{E}[g\{\boldsymbol{u}^\top \boldsymbol{X}_{1i_1}(t)\}] = \mathrm{E}[g\{\boldsymbol{u}^\top \boldsymbol{X}_{2i_2}(t)\}]$ for all $t \in \lfloor 0, 1\rfloor$, $i_1 \in \{1, \ldots, n_1\}$, $i_2 \in \{1, \ldots, n_2\}$, and $\boldsymbol{u} \in \mathbb{R}^p$. Here is a useful basic lemma.

Lemma 5. *Under Assumptions (A.1)–(A.5) and (A.7), the following holds true:*

(a) For any compact subset F of $\mathbb{R}^p$ and any $t \in [0, 1]$, one has

$$\sup_{n_1,n_2\in\mathbb{N}} \mathrm{E}\int_0^1\int_F \{V_{n_1,n_2}(\boldsymbol{u}, t)\}^2 W(\boldsymbol{u})d\boldsymbol{u}dt < \infty.$$

(b) There exist $a > 0$, $C > 0$, such that for any $t_1, t_2 \in [0, 1]$ and $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{R}^p$, one has

$$\sup_{n_1,n_2\in\mathbb{N}} \mathrm{E}|V^2_{n_1,n_2}(\boldsymbol{u}_1, t_1) - V^2_{n_1,n_2}(\boldsymbol{u}_2, t_2)| \le C(\|\boldsymbol{u}_1 - \boldsymbol{u}_2\|^a + |t_1 - t_2|^a).$$

(c) The marginal distribution of $\{V_{n_1,n_2}(\boldsymbol{u}, t)\}$ converges to the marginal distribution of a Gaussian process $\{V_\theta(\boldsymbol{u}, t)\}$ with covariance structure defined, for all $t_1, t_2 \in [0, 1]$ and $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{R}^p$, by

$$\mathrm{cov}\{V_\theta(\boldsymbol{u}_1, t_1), V_\theta(\boldsymbol{u}_2, t_2)\} = \frac{1}{\theta(1-\theta)}\mathrm{cov}\Big[\sqrt{1-\theta}\, g\{\boldsymbol{u}_1^\top \boldsymbol{X}_1(t_1)\} + \sqrt{\theta}\, g\{\boldsymbol{u}_1^\top \boldsymbol{X}_2(t_1)\},$$
$$\sqrt{1-\theta}\, g\{\boldsymbol{u}_2^\top \boldsymbol{X}_1(t_2)\} + \sqrt{\theta}\, g\{\boldsymbol{u}_2^\top \boldsymbol{X}_2(t_2)\}\Big].$$

Proof. For any $t \in [0, 1]$ and any $\boldsymbol{u}$, the random vectors $\boldsymbol{X}_{k1}(t), \ldots, \boldsymbol{X}_{kn_k}(t)$ are iid random variables with $\mathrm{E}\{V_{kn_k}(\boldsymbol{u}, t)\} = 0$ and $\mathrm{E}\{V^2_{kn_k}(\boldsymbol{u}, t)\} = \mathrm{var}[g\{\boldsymbol{u}^\top \boldsymbol{X}_{ki_k}(t)\}]^2 \le D$ for all $k \in \{1, 2\}$ and some $D > 0$ as g is bounded. From this, assertion (a) follows. From the assumptions, the boundedness of sine and cosine and the mean value theorem, for some $D > 0$,

$$\mathrm{E}[g\{\boldsymbol{u}_1^\top \boldsymbol{X}_{ki_k}(t_1)\} - g\{\boldsymbol{u}_2^\top \boldsymbol{X}_{ki_k}(t_2)\}]^2 \le D\,\mathrm{E}\{\boldsymbol{u}_1^\top \boldsymbol{X}_{ki_k}(t_1) - \boldsymbol{u}_2^\top \boldsymbol{X}_{ki_k}(t_2)\}^2 \le D\|\boldsymbol{u}_1 - \boldsymbol{u}_2\|^2 + |t_1 - t_2|^2.$$

Now (b) holds true because by Hölder's inequality and Minkowski's inequality, for $k \in \{1, 2\}$, and for some $D > 0$,

$$\begin{aligned}
\mathrm{E}|V^2_{n_1,n_2}(\boldsymbol{u}_1, t_1) - V^2_{n_1,n_2}(\boldsymbol{u}_2, t_2)| &\le [\mathrm{E}|V_{n_1,n_2}(\boldsymbol{u}_1, t_1) - V_{n_1,n_2}(\boldsymbol{u}_2, t_2)|^2 \times \mathrm{E}|V_{n_1,n_2}(\boldsymbol{u}_1, t_1) + V_{n_1,n_2}(\boldsymbol{u}_2, t_2)|^2]^{1/2} \\
&\le D\{\mathrm{E}|V_{n_1,n_2}(\boldsymbol{u}_1, t_1) - V_{n_1,n_2}(\boldsymbol{u}_2, t_2)|^2\}^{1/2} \\
&\le D[\mathrm{E}[g\{\boldsymbol{u}_1^\top \boldsymbol{X}_{ki_k}(t_1)\} - g\{\boldsymbol{u}_2^\top \boldsymbol{X}_{ki_k}(t_2)\}]^2]^{1/2} \\
&\le D(\|\boldsymbol{u}_1 - \boldsymbol{u}_2\| + |t_1 - t_2|).
\end{aligned}$$

For any $t \in [0,1]$ and $u \in \mathbb{R}^p$, by the assumptions $V_{n_1,n_2}(\boldsymbol{u},t)$ is the linear combination of independent random variables with zero mean and finite moment of the order $2+\eta$. Thus the assumptions of the Central Limit Theorem are satisfied and we find $V_{n_1,n_2}(\boldsymbol{u},t) \rightsquigarrow V_\theta(\boldsymbol{u},t)$. Additionally, the covariance structure of $V_\theta(\boldsymbol{u},t)$ can be computed easily; we find

$$\operatorname{cov}\{V_\theta(\boldsymbol{u}_1,t_1), V_\theta(\boldsymbol{u}_2,t_2)\} = \frac{1}{\theta}\operatorname{cov}[g\{\boldsymbol{u}_1^\top \boldsymbol{X}_1(t_1)\}, g\{\boldsymbol{u}_2^\top \boldsymbol{X}_1(t_2)\}] + \frac{1}{(1-\theta)}\operatorname{cov}[g\{\boldsymbol{u}_1^\top \boldsymbol{X}_2(t_1)\}, g\{\boldsymbol{u}_2^\top \boldsymbol{X}_2(t_2)\}].$$

Hence (c) holds true, and the proof of the lemma is complete. □

Theorem 6. *Under Assumptions (A.1)–(A.5) and (A.7) the following holds true, as* $\min(n_1,n_2)\to\infty$*:*

$$\int_0^1\int_{\mathbb{R}^p}\{V_{n_1,n_2}(\boldsymbol{u},t)\}^2 W(\boldsymbol{u})d\boldsymbol{u}dt \rightsquigarrow \frac{1}{\theta(1-\theta)}\int_0^1\int_{\mathbb{R}^p}\{V_\theta(\boldsymbol{u},t)\}^2 W(\boldsymbol{u})d\boldsymbol{u}dt,$$

where $\{V_\theta(\boldsymbol{u},t) : t\in T, \boldsymbol{u}\in\mathbb{R}^p\}$ *is a Gaussian process with zero mean and covariance structure*

$$\operatorname{cov}\{V_\theta(\boldsymbol{u}_1,t_1), V_\theta(\boldsymbol{u}_2,t_2)\} = \frac{1}{\theta}\operatorname{cov}[g\{\boldsymbol{u}_1^\top \boldsymbol{X}_1(t_1)\}, g\{\boldsymbol{u}_2^\top \boldsymbol{X}_1(t_2)\}] + \frac{1}{1-\theta}\operatorname{cov}[g\{\boldsymbol{u}_1^\top \boldsymbol{X}_2(t_1)\}, g\{\boldsymbol{u}_2^\top \boldsymbol{X}_2(t_2)\}].$$

If additionally (A.6) is fulfilled then also, as $\min(n_1,n_2,m)\to\infty$*,*

$$\frac{1}{m}\sum_{j=1}^m\int_{\mathbb{R}^p}\{V_{n_1,n_2}(\boldsymbol{u},j/m)\}^2 W(\boldsymbol{u})d\boldsymbol{u} \rightsquigarrow \frac{1}{\theta(1-\theta)}\int_0^1\int_{\mathbb{R}^p}\{V_\theta(\boldsymbol{u},t)\}^2 W(\boldsymbol{u})d\boldsymbol{u}dt.$$

Proof. Under our assumptions, the assertion of Lemma 5 holds true, hence the assumptions of Theorem 22 in [31] are satisfied and by this theorem we can conclude the first assertion. In view of the assumptions, the second assertion also holds true, and the proof of the theorem is complete. □

Proof of Theorem 1. Under the null hypothesis is true and on account of Assumption (A.2) we note that for all $\boldsymbol{u}$, t, $i_1\in\{1,\dots,n_1\}$ and $i_2\in\{1,\dots,n_2\}$, $\mathrm{E}[g\{\boldsymbol{u}^\top \boldsymbol{X}_{1i_1}(t)\}] = \mathrm{E}[g\{\boldsymbol{u}^\top \boldsymbol{X}_{2i_2}(t)\}]$, and therefore the assertions of Theorem 6 remain true if in the definition of $\{V_{n_k}(\boldsymbol{u},t)$, $\boldsymbol{u}\in\mathbb{R}^p, t\in\lfloor 0,1\rfloor\}$ the terms $\mathrm{E}g\{\boldsymbol{u}^\top \boldsymbol{X}_{1i_1}(t)\} = \mathrm{E}g\{\boldsymbol{u}^\top \boldsymbol{X}_{2i_2}(t)\}$ are omitted. Theorem 1 follows. □

Proof of Theorem 2. It follows immediately from Theorem 6. □

Next, consider the assumptions (A.1), (A.2a), (A.3)–(A.8) and study the behavior of

$$\int_0^1\int_{\mathbb{R}^p}\{\widehat{V}_{n_1,n_2}(\boldsymbol{u},t)\}^2 W(\boldsymbol{u})d\boldsymbol{u}dt,$$

where $\widehat{V}_{n_1,n_2}(\boldsymbol{u},t)$ is defined as $V_{n_1,n_2}(\boldsymbol{u},t)$ with $\boldsymbol{X}_{ki_k}(t)$ replaced by $\widehat{\boldsymbol{x}}_{ki_k}(t)$ that is defined in Section 4.1 in the univariate case. Then we can write

$$\widehat{\boldsymbol{x}}_{ki_k}(t) = \frac{1}{mh}\sum_{j=1}^m \boldsymbol{X}_{ki_kj}\widetilde{K}\Big(\frac{t-j/m}{h}\Big),$$

where the kernel $\widetilde{K}$ is determined from the definition in Section 4.1 that has similar properties as K. We use the decomposition

$$\widehat{\boldsymbol{x}}_{ki_k}(t) = \frac{1}{mh}\sum_{j=1}^m \boldsymbol{x}_{ki_kj}\widetilde{K}\Big(\frac{t-j/m}{h}\Big) + \frac{1}{mh}\sum_{j=1}^m \boldsymbol{\varepsilon}_{ki_kj}\widetilde{K}\Big(\frac{t-j/m}{h}\Big) \equiv \widetilde{\boldsymbol{x}}_{ki_k}(t) + \widetilde{\boldsymbol{\varepsilon}}_{ki_k}(t)$$

and the Taylor expansion

$$g\{\boldsymbol{u}\widehat{\boldsymbol{x}}_{ki_k}(t)\} = g\{\boldsymbol{u}\widetilde{\boldsymbol{x}}_{ki_k}(t)\} + \boldsymbol{u}\widetilde{\boldsymbol{\varepsilon}}_{ki_k}(t)\frac{\partial g(z)}{\partial z}\Big|_{z=\boldsymbol{u}\boldsymbol{x}_{ki_k}(t)} + \{\boldsymbol{u}\boldsymbol{\varepsilon}_{ki_k}(t)\}^2 R_{ki_k}(t),$$

where $R_{ki_k}(t)$ is a remainder. Under these assumptions for any fixed t, the random vectors $\widehat{\boldsymbol{x}}_{ki_k}(t)$ are independent. Going through the proofs of Lemma 5 and Theorem 6, we find that under Assumptions (A.1), (A.2a), (A.3)–(A.8), the assertions of Theorem 6 remain true if $V_{n_1,n_2}(\boldsymbol{u},t)$ is replaced by $\widehat{V}_{n_1,n_2}(\boldsymbol{u},t)$. □

Proofs of Theorem 3 and 4. The proofs of both theorems follow from the above considerations and Theorem 6. □

Acknowledgments

The authors are grateful to the Editor-in-Chief, Guest Associate Editor, and referees for comments and suggestions which led to an improved manuscript. M. Hušková was partially supported by grant GAČR 15-09663S. S.G. Meintanis was partially supported by grant 11699 of the Special Account for Research Grants (ΕΛΚΕ) of the National and Kapodistrian University of Athens.

References

[1] R. Bárcenas, J. Ortega, A.J. Quiroz, Quadratic forms of the empirical processes for the two-sample problem for functional data, Test 26 (2017) 503–526.
[2] M. Benko, W.K. Härdle, A. Kneip, Common functional principal components, Ann. Stat. 37 (2009) 1–34.
[3] J.R. Berrendero, A. Justel, M. Svarc, Principal components for multivariate functional data, Comput. Statist. Data Anal. 55 (2011) 2619–2634.
[4] M. Bilodeau, P. Lafaye de Micheaux, A multivariate empirical characteristic function test of independence with normal marginals, J. Multivariate Anal. 95 (2005) 345–369.
[5] A. Cabaña, A.M. Estrada, J. Peña, A.J. Quiroz, Permutation tests in the two-sample problem for functional data. In: G. Aneiros, E.G. Bongiorno, R. Cao, P. Vieu (Eds.), Functional Statistics and Related Fields: Contributions to Statistics, Springer, New York, pp. 77–85, 2017.
[6] T.I. Cannings, R.J. Samworth, Random-projection ensemble classification, J. R. Stat. Soc. B (Stat. Methodol.) 79 (2017) 959–1035.
[7] J.M. Chiou, H.G. Müller, Linear manifold modeling of multivariate functional data, J. R. Stat. Soc. B (Stat. Methodol.) 76 (2014) 605–626.
[8] S.N. Chiu, K.I. Liu, Generalized Cramér–von Mises goodness-of-fit tests for multivariate distributions, Comput. Statist. Data Anal. 53 (2009) 3817–3834.
[9] S. Csörgő, Testing for normality in arbitrary dimension, Ann. Statist. 14 (1986) 708–732.
[10] J.A. Cuesta-Albertos, E. del Barrio, R. Fraiman, C. Matrán, The random projection method in goodness of fit for functional data, Comput. Statist. Data Anal. 51 (2007) 4814–4831.
[11] A. Cuevas, A partial overview of the theory of statistics with functional data, J. Stat. Plann. Infer. 147 (2014) 1–23.
[12] A. Cuevas, M. Febrero, R. Fraiman, An ANOVA test for functional data, Comput. Statist. Data Anal. 47 (2004) 111–122.
[13] Z. Du, Testing for serial independence of panel errors, Comput. Statist. Data Anal. 76 (2014) 248–261.
[14] T.W. Epps, Characteristic functions and their empirical counterparts: Geometric interpretations and applications to statistical inference, Amer. Statist. 47 (1993) 33–38.
[15] M. Febrero-Bande, M. Oviedo de la Fuente, Statistical Computing in Functional Data Analysis: The R Package `fda.usc`, J. Stat. Softw. 51 (2012) 1–28.
[16] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis, Springer, New York, 2006.
[17] S. Fremdt, J. Steinebach, L. Horváth, P. Kokoszka, Testing the equality of covariance operators in functional samples, Scand. J. Stat. 40 (2013) 138–152.
[18] A.K. Ghosh, M. Biswas, Distribution-free high-dimensional two-sample tests based on discriminating hyperplanes, Test 25 (2016) 525–547.
[19] A. Goia, P. Vieu, Editorial: An introduction to recent advances in high/infinite dimensional statistics, J. Multivariate Anal. 146 (2016) 1–6.
[20] W. González-Manteiga, P. Vieu, Editorial: Statistics for Functional Data, Comput. Statist. Data Anal. 51 (2007) 4788–4792.
[21] K.B. Gregory, R.J. Carroll, V. Baladandayuthapani, S.N. Lahiri, A two-sample test for equality of means in high dimension, J. Amer. Stat. Assoc. 110 (2015) 681–696.
[22] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf, A. Smola, A kernel two-sample test, J. Mach. Learn. Res. 13 (2012) 723–773.
[23] O. Gromenko, P. Kokoszka, M. Reimherr, Detection of change in spatiotemporal mean function, J. R. Stat. Soc. B (Stat. Methodol.) 79 (2017) 29–50.
[25] P. Hall, N. Tajvidi, Permutation tests for equality of distributions in high-dimensional settings, Biometrika 89 (2002) 359–374.
[25] P. Hall, I. Van Keilegom, Two-sample tests in functional data analysis starting from discrete data, Stat. Sinica 17 (2007) 1511–1531.
[26] N. Henze, T. Wagner, A new approach to the BHEP tests for multivariate normality, J. Multivariate Anal. 62 (1997) 1–23.
[27] L. Horváth, P. Kokoszka, Inference for Functional Data with Applications, Springer, New York, 2012.
[28] L. Horváth, P. Kokoszka, R. Reeder, Estimation of the mean of functional time series and a two sample problem, J. R. Stat. Soc. B (Stat. Methodol.) 75 (2013) 103–122.
[29] L. Horváth, G. Rice, Testing equality of means when observations are from functional time series, J. Time Ser. Anal. 36 (2015) 84–108.
[30] M. Hušková, S.G. Meintanis, Tests for the multivariate *k*-sample problem based on the empirical characteristic function, J. Nonparametr. Statist. 20 (2008) 263–277.
[31] I. Ibragimov, R. Has'minskiĭ, Statistical Estimation: Asymptotic Theory, Springer, New York, 1981.
[32] J. Jacques, C. Preda, Model-based clustering for multivariate functional data, Comput. Statist. Data Anal. 71 (2014) 92–106.
[33] Q. Jiang, S.G. Meintanis, L.X. Zhu, Two-sample tests for multivariate functional data. In: G. Aneiros, E.G. Bongiorno, R. Cao, P. Vieu (Eds.), Functional Statistics and Related Fields: Contributions to Statistics, Springer, New York, pp. 145–154, 2017.
[34] M.D. Jiménez-Gamero, J. Muñoz-García, R. Pino-Mejías, Testing goodness-of-fit for the distribution of errors in multivariate linear models, J. Multivariate Anal. 95 (2005) 301–322.
[35] P. Kokoszka, H. Oja, B. Park, L. Sangalli, Editorial: Special issue on functional data analysis, ECOSTA 1 (2017) 99–100.
[36] I.A. Koutrouvelis, Regression-type estimation of the parameters of stable laws, J. Amer. Statist. Assoc. 75 (1980) 918–928.
[37] D. Kraus, V.M. Panaretos, Dispersion operators and resistant second-order functional data analysis, Biometrika 99 (2012) 813–832.

[38] B. Lindsay, M. Markatou, S. Ray, Kernels, degrees of freedom, and power properties of quadratic distance goodness-of-fit tests, J. Amer. Statist. Assoc. 109 (2014) 395–410.
[39] A. Nieto-Reyes, J.A. Cuesta-Albertos, F. Gamboa, A random projection based test of Gaussianity for stationary processes, Comput. Statist. Data Anal. 75 (2014) 124–141.
[40] V.M. Panaretos, D. Kraus, J.H. Maddocks, Second-order comparison of Gaussian random functions and the geometry of DNA minicircles, J. Amer. Statist. Assoc. 105 (2010) 670–682.
[41] A. Pini, S. Vantini, Interval-wise testing for functional data, J. Nonparametr. Statist. 29 (2017) 407–424.
[42] G.M. Pomann, A.M. Staicu, S. Ghosh, two-sample hypothesis testing for functional data with application to a diffusion tensor imaging study of multiple sclerosis, J. R. Stat. Soc. C (Appl. Stat.) 65 (2016) 395–414.
[43] J. Pudelko, On a new affine invariant and consistent test for multivariate normality, Probab. Math. Statist. 25 (2005) 43–54.
[44] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, Springer, New York, 2005.
[45] M. Ruiz-Meana, D. Garcia-Dorado, P. Pina, J. Inserte, L. Agull, J. Soler, Cariporide preserves mitochondrial proton gradient and delays ATP depletion in cardiomyocytes during ischemic conditions, Amer. J. Physiol. Heart Circul. Physiol. 285 (2003) H999–H1006.
[46] Y. Shao, M. Zhou, A characterization of multivariate normality through univariate projections, J. Multivariate Anal. 101 (2010) 2637–2640.
[47] G. Székely, M. Rizzo, Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method, J. Classif. 22 (2005) 151–183.
[48] G. Székely, M. Rizzo, Energy statistics: A class of statistics based on distances, J. Stat. Plan. Infer. 143 (2013) 249–272.
[49] C. Tenreiro, On the choice of the smoothing parameter for the BHEP goodness-of-fit test, Comput. Statist. Data Anal. 53 (2009) 1038–1053.
[50] M.J. Valderama, An overview to modelling functional data, Comput. Statist. 22 (2007) 331–334.
[51] J.L. Wang, J.M. Chiou, H.G. Müller, Annual Review of Statistics and its Application 3 (2016) 257–295.
[52] A.H. Welsh, Implementing empirical characteristic function procedures, Statist. Probab. Lett. 4 (1986) 65–67.
[53] C. Zhang, H. Peng, J.T. Zhang, Two-sample tests for functional data, Commun. Stat. Theory Meth. 39 (2010) 559–578.
[54] X. Zhang, X. Shao, Two sample inference for the second-order property of temporally dependent functional data, Bernoulli 21 (2015) 909–929.