CrossMark

A two-sample test for the error distribution in nonparametric regression based on the characteristic function

G. I. Rivas-Martínez[1] · M. D. Jiménez-Gamero[2] ·
J. L. Moreno-Rebollo[2]

Abstract A test for the equality of error distributions in two nonparametric regression models is proposed. The test statistic is based on comparing the empirical characteristic functions of the residuals calculated from independent samples of the models. The asymptotic null distribution of the test statistic cannot be used to estimate its null distribution because it is unknown, since it depends on the unknown common error distribution. To approximate the null distribution, a weighted bootstrap estimator is studied, providing a consistent estimator. The finite sample performance of this approximation as well as the power of the resulting test are evaluated by means of a simulation study. The procedure can be extended to testing for the equality of $d > 2$ error distributions.

✉ M. D. Jiménez-Gamero
  dolores@us.es

  G. I. Rivas-Martínez
  gusyri@hotmail.com

  J. L. Moreno-Rebollo
  jlmoreno@us.es

[1] Laboratorio de Sistemas de Potencia y Control, Universidad Nacional de Asunción, San Lorenzo, Paraguay

[2] Departamento de Estadística e Investigación Operativa, Universidad de Sevilla, Seville, Spain

1 Introduction

Let (X_1, Y_1) and (X_2, Y_2) be two independent random vectors. Assume that they satisfy the general nonparametric regression models,

$$Y_k = m_k(X_k) + \sigma_k(X_k)\varepsilon_k, \tag{1}$$

where $m_k(x) = E(Y_k \mid X_k = x)$ is the regression function, $\sigma_k^2(x) = Var(Y_k \mid X_k = x)$ is the conditional variance function and ε_k is the regression error, which is assumed to be independent of X_k, $k = 1, 2$. By construction $E(\varepsilon_k) = 0$ and $Var(\varepsilon_k)=1$. The regression functions, the variance functions, the distributions of the error and that of the covariates are unknown and no parametric models are assumed for them. We are interested in testing for the equality of the error distributions, that is, in tests of the null hypothesis

$$H_0 : F_1 = F_2,$$

versus the alternative

$$H_1 : F_1 \neq F_2,$$

where F_1, F_2 stand for the cumulative distribution function (CDF) of ε_1 and ε_2, respectively. In view of the uniqueness of the characteristic function (CF), the null hypothesis can equivalently be stated as

$$H_0 : C_1 = C_2,$$

versus the alternative

$$H_1 : C_1 \neq C_2,$$

where C_k denotes the CF corresponding to F_k, that is, $C_k(t) = \int \exp(\mathrm{i}tx)\, dF_k(x) = R_k(t) + \mathrm{i}I_k(t), k = 1, 2$.

The equality of the error distributions is a usual assumption in several statistical problems such as that of testing for the equality of regression curves (see for example, Young and Bowman 1995; Hall and Hart 1990; Kulasekera and Wang 2001). The equality of the error distributions may considerably simplify some procedures. For instance, under H_0, the asymptotic null distribution of the test statistic for the equality of regression functions in Pardo-Fernández et al. (2015a) coincides with the one of the classical ANOVA test for comparing means (it also requires the equality of the densities of the covariates). Another example is given by the asymptotic null distribution of the test statistic for the equality of variance functions in Pardo-Fernández et al. (2015b) based on CFs. The authors prove that, when the covariates are identically distributed and the error distributions are equal, the asymptotic distribution coincides with that of the classical Levene test for comparing variances. Thus, the problem of testing for H_0 is of considerable practical interest.

The problem of testing whether two samples come from the same population has generated a considerable amount of research. Many different approaches have been proposed to deal with this problem when the data are observable (see, for example, the references in Meintanis 2005; Alba-Fernández et al. 2008; Hobza et al. 2014; Baringhaus and Kolbe 2015; Modarres 2016). In our setting the errors are not observable

and the inference must be based on the residuals, which are not independent even if the original data are. The number of proposals to deal with this case is not so big. Mora (2005) proposed tests for testing H_0 when the regression models are linear; for the more general model given in (1), Pardo-Fernández (2007)—PF07 from now on- also proposed tests for H_0. These two papers study Kolgomorov–Smirnov (KS) and Cramér–von Misses (CvM) type test statistics based on the empirical CDFs of the residuals. Since the null distribution of these test statistics is unknown, these papers use a smooth bootstrap to approximate the critical values. Two main problems with these procedures are: they assume strong conditions on the distributions of the errors which, among other things, are supposed to have a smooth density; in addition, although quite easy to implement, the bootstrap can become computationally expensive as the sample sizes of the data increase.

This paper proposes a test for H_0 that is based on comparing the empirical CF (ECF) of the residuals in samples from the models. It can be seen as a residual version of the test in Alba-Fernández et al. (2008), designed for the two-sample problem when observable independent and identically distributed (IID) data are available from each population. A weighted bootstrap (WB) estimator, in the sense of Burke (2000), is proposed to approximate the critical values. This method has been previously used in Kojadinovic and Yan (2012) and Ghoudi and Rémillard (2014), to approximate the null distribution of goodness-of-fit tests based on the empirical CDF, in Jiménez-Gamero and Kim (2015), to approximate the null distribution of goodness-of-fit tests based on the ECF, in Quessy and Éthier (2012), for the two-sample problem for dependent data, and in Jiménez-Gamero et al. (2016), for the two-sample problem for observable independent data, among others. In view of the good properties of the WB in these and other papers, it is also expected to work well for approximating the null distribution of the test statistic considered in this paper.

Compared to the tests in PF07, the procedure suggested in this paper has two main advantages. First, it assumes less stringent assumptions on the distribution of the regression errors. Specifically, we do not assume that the error distribution has a probability density function. Thus, the method can be applied when such distribution is arbitrary: continuous, discrete or mixed. Secondly, the WB approximation is computationally more efficient than that based on the smooth bootstrap.

The paper is organized as follows. Section 2 describes the test statistic. The problem of approximating the null distribution of the proposed test statistic is dealt with in Sect. 3, where the use of a WB estimator is studied. The consistency of the resulting null distribution estimator is proved. It is also shown that the resulting test is consistent, in the sense of being able to detect any alternative. Some practical issues are addressed in Sect. 4. Section 5 reports the results of some simulation experiments designed to study the finite sample performance of the proposed approximation, to compare it with other methods as well as a real data application. From this numerical study it is concluded that the WB approximation works, in the sense of providing levels close to the nominal values, and that the power of the test is comparable or even greater than the power of the test based on the empirical CDF. Section 6 shows how the proposed test can be extended to the general case of comparison of $d > 2$ error distributions. All proofs and technical details are deferred to the last section.

The following notation will be used along the paper: all vectors are column vectors; the superscript T denotes transpose; $1_n \in \mathbb{R}^n$ has all its components equal to 1; if $x \in \mathbb{R}^k$, with $x^T = (x_1, \ldots, x_k)$, then $diag(x)$ is the $k \times k$ diagonal matrix whose (i, i) entry is x_i, $1 \leq i \leq k$; P_0, E_0 and Cov_0 denote probability, expectation and covariance, respectively, by assuming that the null hypothesis is true; P_*, E_* and Cov_* denote the conditional probability law, expectation and covariance, given the data, respectively; $\overset{\mathcal{L}}{\to}$ denotes convergence in distribution; $\overset{P}{\to}$ denotes convergence in probability; $\overset{a.s.}{\to}$ denotes the almost sure convergence; for any complex number $z = a + ib$, $|z|$ is its modulus; an unspecified integral denotes integration over the whole real line $\mathbb{R}$; for a given non-negative real-valued function ω we denote $\| \cdot \|_\omega$ to the norm and $\langle \cdot, \cdot \rangle_\omega$ to the scalar product in the Hilbert space $L^2(\omega) = \{g : \mathbb{R} \to \mathbb{C}, \int |g(t)|^2 \omega(t)dt < \infty\}$.

2 The test statistic

Let (X_{kj}, Y_{kj}), $j = 1, 2, \ldots, n_k$, be an IID sample from (X_k, Y_k) satisfying (1), and let $\varepsilon_{k1}, \ldots, \varepsilon_{kn_k}$ denote the associated errors, $k = 1, 2$. Since the hypothesis H_0 establishes the equality of the distributions of the errors ε_{kj} and they are not observable, the inference must be based on the residuals,

$$\hat{\varepsilon}_{kj} = \frac{Y_{kj} - \hat{m}_k(X_{kj})}{\hat{\sigma}_k(X_{kj})}, \quad j = 1, 2, \ldots, n_k, \tag{2}$$

where $\hat{m}_k$ and $\hat{\sigma}_k$ are estimators of m_k and σ_k, respectively, $k = 1, 2$. Several choices are possible for $\hat{m}_k$ and $\hat{\sigma}_k$. Here, as in PF07, we use the following kernel estimators for the regression function m_k and the variance function σ_k^2,

$$\hat{m}_k(x) = \sum_{j=1}^{n_k} W_{kj}(x; h_k)Y_{kj}, \quad x \in S,$$

$$\hat{\sigma}_k^2(x) = \sum_{j=1}^{n_k} W_{kj}(x; h_k) \left\{ Y_{kj} - \hat{m}_k(x) \right\}^2, \quad x \in S,$$

where

$$W_{kj}(x; h_k) = \frac{K_{h_k}(X_{kj} - x)}{\sum_{s=1}^{n_k} K_{h_k}(X_{ks} - x)}, \quad x \in S,$$

$K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h})$, $K(\cdot)$ is a kernel and h_k is the bandwidth, satisfying certain conditions that will be specified later. The proposed test statistic takes the form

$$T_{n_1,n_2} = \|\hat{C}_1 - \hat{C}_2\|_\omega^2,$$

where

$$\hat{C}_k(t) = \frac{1}{n_k} \sum_{j=1}^{n_k} \exp(it\hat{\varepsilon}_{kj}) = \hat{R}_k(t) + i\hat{I}_k(t), \tag{3}$$

$$\hat{R}_k(t) = \frac{1}{n_k}\sum_{j=1}^{n_k}\cos(t\hat{\varepsilon}_{kj}), \quad \hat{I}_k(t) = \frac{1}{n_k}\sum_{j=1}^{n_k}\sin(t\hat{\varepsilon}_{kj}),$$

$k = 1, 2$, and $\omega(t)$ is a non-negative function.

In order to study properties of T_{n_1,n_2} some assumptions will be required. Next we list them.

(A.1) All limits in this paper are taken when $n_1, n_2 \to \infty$ in such a way that

$$\lim \frac{n_1}{N} = \tau, \quad \text{for some} \quad \tau \in (0, 1),$$

where $N = n_1 + n_2$.

(A.2) The weight function $\omega(t)$ is a non-negative symmetric function,

$$\omega(t) = \omega(-t), \quad \forall t,$$

and $\int t^4\omega(t)dt < \infty$.

There is no restriction in assuming that the weight function $\omega(t)$ is symmetric because otherwise by defining $\omega_1(t) = 0.5\{\omega(t)+\omega(-t)\}$, which is clearly symmetric, we have that

$$\|C_1 - C_2\|_\omega = \|C_1 - C_2\|_{\omega_1},$$

for any two CFs C_1 and C_2. Note that the symmetry of ω implies that

$$T_{n_1,n_2} = \|\hat{R}_1 - \hat{R}_2 + \hat{I}_1 - \hat{I}_2\|_\omega^2.$$

The following assumption will be required to ensure that $\hat{m}_k$ and $\hat{\sigma}_k$ provide consistent estimators of m_k and σ_k, respectively.

(A.3) For $k = 1, 2$,
   (i) $E(\varepsilon_k^4) < \infty$.
   (ii) X_k is absolutely continuous with compact support S and density function f_k.
   (iii) f_k, m_k and σ_k are twice continuously differentiable.
   (iv) $\inf_{x \in S} f_k > 0$ and $\inf_{x \in S} \sigma_k > 0$.
   (v) $n_k h_k^4 \to 0$ and $n_k h_k^2 / \ln n_k \to \infty$.
   (vi) K is a symmetric density function with compact support and twice continuously differentiable.

For simplicity we assume that the same kernel function, K, is used in both populations. Nevertheless, the results to be stated remain true if different kernels are used, whenever they satisfy Assumption (A.3)(vi).

In order to give a sound justification of T_{n_1,n_2} as a test statistic for testing H_0 we next derive its limit.

Theorem 1 *Suppose that Assumptions (A.1)–(A.3) hold, then $T_{n_1,n_2} \xrightarrow{P} \kappa = \|C_1 - C_2\|_\omega^2$.*

Note that $\kappa \geq 0$. If H_0 is true then $\kappa = 0$. Since two distinct characteristic functions can be equal in a finite interval (Feller 1971, p. 506), a general way to ensure that $\kappa = 0$ iff H_0 is true is to take ω positive for almost all (with respect to the Lebesgue measure) points in $\mathbb{R}$. Thus, a reasonable test for testing H_0 should reject the null hypothesis for large values of T_{n_1,n_2}. Now, to determine what are large values we must calculate its null distribution, or at least an approximation to it. This is the topic of the next section.

3 Approximating the null distribution

The null distribution of T_{n_1,n_2} is clearly unknown, so it must be approximated. We first try to estimate it by means of its asymptotic null distribution.
   Let

$$R_k'(t) = -\int x \sin(tx) d F_k(x), \quad I_k'(t) = \int x \cos(xt) d F_k(t),$$

$k = 1, 2$. Note that under the null hypothesis $R_1(t) = R_2(t) = R(t)$, $I_1(t) = I_2(t) = I(t)$, $R_1'(t) = R_2'(t) = R'(t)$ and $I_1'(t) = I_2'(t) = I'(t)$.

Theorem 2 *Suppose that Assumptions (A.1)–(A.3) hold. If H_0 is true, then*

$$\frac{n_1 n_2}{N} T_{n_1,n_2} \xrightarrow{\mathcal{L}} \|Z\|_\omega^2,$$

where $\{Z(t), \ t \in \mathbb{R}\}$ is a centered Gaussian process on $L_2(\omega)$ with covariance kernel $\varrho_0(t, s) = Cov_0\{Z_0(\varepsilon; t), Z_0(\varepsilon; s)\}$ and

$$Z_0(\varepsilon; t) = \cos(t\varepsilon) + t\varepsilon I(t) - t\frac{\varepsilon^2 - 1}{2} R'(t) - R(t)$$
$$+ \sin(t\varepsilon) - t\varepsilon R(t) - t\frac{\varepsilon^2 - 1}{2} I'(t) - I(t).$$

Remark 1 The asymptotic null distribution of $\frac{n_1 n_2}{N} T_{n_1,n_2}$ can be expressed as

$$\|Z\|_\omega^2 \overset{d}{=} \sum_{j \geq 1} \lambda_j Z_j^2, \tag{4}$$

where $\overset{d}{=}$ stands from the equality in distribution, $Z_1, Z_2, \ldots$ are independent standard normal variables and the set $\{\lambda_j\}_{j \geq 1}$ are the non-zero eigenvalues of the integral equation

$$\int \varrho_0(t, s) g_j(t) w(t) dt = \lambda_j g_j(s)$$

with corresponding eigenfunctions $\{g_j(\cdot)\}_{j \geq 1}$.

From Remark 1 it becomes evident that the asymptotic null distribution of T_{n_1,n_2} does not provide a useful approximation to its null distribution since it depends on the unknown common distribution. So, we next study another way of approximating it by means of a WB estimator.

Let

$$C_\tau(t) = \tau C_1(t) + (1 - \tau)C_2(t) = R_\tau(t) + iI_\tau(t).$$

Let $\xi_{11}, \ldots, \xi_{1n_1}, \xi_{21}, \ldots, \xi_{2n_2}$ be IID random variates with mean 0 and variance 1, which are independent of $(X_{11}, Y_{11}), \ldots, (X_{1n_1}, Y_{1n_1}), (X_{21}, Y_{21}), \ldots, (X_{2n_2}, Y_{2n_2})$. We define the following WB version of T_{n_1,n_2},

$$T^*_{1,n_1,n_2} = \|C_1^* - C_2^*\|^2_\omega,$$

where

$$C_k^*(t) = \frac{1}{n_k} \sum_{j=1}^{n_k} \xi_{kj} Z_{k,\tau}(\varepsilon_{kj}; t),$$

$$
\begin{aligned}
Z_{k,\tau}(\varepsilon; t) = {}& \cos(t\varepsilon) + t\varepsilon I_k(t) - t\frac{\varepsilon^2 - 1}{2}R_k'(t) - R_\tau(t) \\
& + \sin(t\varepsilon) - t\varepsilon R_k(t) - t\frac{\varepsilon^2 - 1}{2}I_k'(t) - I_\tau(t),
\end{aligned}
\tag{5}
$$

$k = 1, 2$. The next result gives the weak limit of the conditional distribution of T^*_{1,n_1,n_2}, given the data $(X_{11}, Y_{11}), \ldots, (X_{1n_1}, Y_{1n_1}), (X_{21}, Y_{21}), \ldots, (X_{2n_2}, Y_{2n_2})$.

Theorem 3 *Suppose that Assumptions (A.1)–(A.3) hold, then*

$$\sup_x \left| P_* \left\{ \frac{n_1 n_2}{N} T^*_{1,n_1,n_2} \le x \right\} - P\{T_\tau \le x\} \right| \xrightarrow{P} 0,$$

where $T_\tau = \|Z_\tau\|^2_\omega$, $\{Z_\tau(t), t \in \mathbb{R}\}$ *is a centered Gaussian process on* $L_2(\omega)$ *with covariance kernel* $\varrho_\tau(t, s) = (1 - \tau)\varrho_{1,\tau}(t, s) + \tau\varrho_{2,\tau}(t, s)$ *and* $\varrho_{k,\tau}(t, s) = E\{Z_{k,\tau}(\varepsilon_k; t)Z_{k,\tau}(\varepsilon_k; s)\}$, $k = 1, 2$.

The result in Theorem 3 is valid whether the null hypothesis is true or not. If H_0 holds, then the kernels $\varrho_0(t, s)$ and $\varrho_\tau(t, s)$ coincide. Therefore, a direct consequence of Theorems 2 and 3 is that the conditional distribution of T^*_{1,n_1,n_2}, given the data, provides a consistent estimator of the distribution of T_{n_1,n_2} when H_0 is true. However, from a practical point of view, this result is useless because the function $Z_{k,\tau}(\varepsilon_{kj}; t)$ depends on the non-observable error ε_{kj} and on the unknown values of the functions $R_k(t), I_k(t), R_k'(t)$ and $I_k'(t), j = 1, \ldots, n_j, k = 1, 2$. To overcome these difficulties we replace ε_{kj} by $\hat\varepsilon_{kj}$, $R_k(t)$ by $\hat{R}_k(t)$, $I_k(t)$ by $\hat{I}_k(t)$, $R_k'(t)$ by $\hat{R}_k'(t)$ and $I_k'(t)$ by $\hat{I}_k'(t)$, with

$$\hat{R}_k'(t) = -\frac{1}{n_k} \sum_{j=1}^{n_k} \hat\varepsilon_{kj} \sin(t\hat\varepsilon_{kj}), \quad \hat{I}_k'(t) = \frac{1}{n_k} \sum_{j=1}^{n_k} \hat\varepsilon_{kj} \cos(t\hat\varepsilon_{kj}).$$

So, instead of T^*_{1,n_1,n_2}, now we consider

$$T^*_{2,n_1,n_2} = \|\hat{U}^*_1 - \hat{U}^*_2\|^2_\omega,$$

where

$$
\begin{aligned}
\hat{U}^*_k(t) = \tfrac{1}{n_k} \sum_{j=1}^{n_k} \Bigg\{ &\cos(t\hat{\varepsilon}_{kj}) + t\hat{\varepsilon}_{kj}\hat{I}_k(t) - t\tfrac{\hat{\varepsilon}^2_{kj}-1}{2}\hat{R}'_k(t) - \hat{R}_\tau(t) \\
&+ \sin(t\hat{\varepsilon}_{kj}) - t\hat{\varepsilon}_{kj}\hat{R}_k(t) - t\tfrac{\hat{\varepsilon}^2_{kj}-1}{2}\hat{I}'_k(t) - \hat{I}_\tau(t) \Bigg\} \xi_{kj},
\end{aligned}
\tag{6}
$$

$k = 1, 2$, and

$$\hat{R}_\tau(t) = \frac{n_1}{N}\hat{R}_1(t) + \frac{n_2}{N}\hat{R}_2(t), \quad \hat{I}_\tau(t) = \frac{n_1}{N}\hat{I}_1(t) + \frac{n_2}{N}\hat{I}_2(t).$$

The next theorem states that replacing ε_{kj} by $\hat{\varepsilon}_{kj}, \ldots, I'_k(t)$ by $\hat{I}'_k(t)$ in the expression of T^*_{1,n_1,n_2} has no asymptotic effect, in the sense that both T^*_{1,n_1,n_2} and T^*_{2,n_1,n_2} have the same conditional asymptotic distribution, given the data. Observe that all quantities involved in the definition of T^*_{2,n_1,n_2} are known, thus, in principle, one could be able to know, or at least to accurately approximate its conditional distribution, given the data. This practical issue will be handled in Sect. 4.

Theorem 4 *Suppose that Assumptions (A.1)–(A.3) hold, then*

$$\sup_x \left| P_* \left\{ \frac{n_1 n_2}{N} T^*_{2,n_1,n_2} \le x \right\} - P\{T_\tau \le x\} \right| \overset{P}{\longrightarrow} 0,$$

where T_τ is as defined in Theorem 3.

The result in Theorem 4 is valid whether the null hypothesis H_0 is true or not. As observed before for T^*_{1,n_1,n_2}, an immediate consequence of this fact and Theorem 2 is the following.

Corollary 1 *If H_0 is true and the assumptions in Theorem 4 hold, then*

$$\sup_x \left| P_* \left\{ \frac{n_1 n_2}{N} T^*_{2,n_1,n_2} \le x \right\} - P\left\{ \frac{n_1 n_2}{N} T_{n_1,n_2} \le x \right\} \right| \overset{P}{\longrightarrow} 0.$$

Let $\alpha \in (0, 1)$. For testing H_0 we consider

$$
\Psi_* = \begin{cases} 1, & \text{if } T_{n_1,n_2} \ge t^*_{2,n_1,n_2,\alpha}, \\ 0, & \text{otherwise}, \end{cases}
$$

where $t^*_{2,n_1,n_2,\alpha}$ is the $1-\alpha$ percentile of the conditional distribution of T^*_{2,n_1,n_2}, or equivalently, $\Psi_* = 1$ if $p^* \le \alpha$, where $p^* = P_* \left\{ T^*_{2,n_1,n_2} \ge T_{n_1,n_2,obs} \right\}$ and $T_{n_1,n_2,obs}$ is the observed value of the test statistic. The result in Corollary 1 states that Ψ_* is

asymptotically correct, in the sense that its type I error probability is asymptotically equal to the nominal value α.

From Theorems 1, 2 and 4, it readily follows the next result.

Corollary 2 *Suppose that H_0 is not true, the assumptions in Theorem 4 hold and ω is such that*

$$\kappa = \|C_1 - C_2\|_\omega^2 > 0, \tag{7}$$

then $P(\Psi_* = 1) \to 1$.

Corollary 2 shows that, if ω is such that (7) holds, then the test Ψ_* is consistent in the sense of being able to asymptotically detect any (fixed) alternative. As discussed before, a general way to ensure (7) is to take ω positive for almost all (with respect to the Lebesgue measure) points in $\mathbb{R}$.

Remark 2 The results so far stated keep on being true if instead of using the raw multipliers, $\xi_{11}, \ldots, \xi_{1n_1}, \xi_{21}, \ldots, \xi_{2n_2}$, we use the centered multipliers, $\xi_{11} - \bar{\xi}_1, \ldots, \xi_{1n_1} - \bar{\xi}_1, \xi_{21} - \bar{\xi}_2, \ldots, \xi_{2n_2} - \bar{\xi}_2$, where $\bar{\xi}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} \xi_{jk}$, $k = 1, 2$, as suggested in Burke (2000) and Kojadinovic and Yan (2012).

Remark 3 In Remark 1 we saw that the null distribution of T_{n_1,n_2} is a linear combination of independent χ^2 variables, the weights in that linear combination being the eigenvalues $\{\lambda_j\}_{j\geq 1}$ of certain functional. Routine algebra shows that the conditional distribution of T_{2,n_1,n_2}^*, given the data, can be also expressed as a linear combination of a certain finite set of variables, $\sum_{j=1}^N \hat{\lambda}_j W_j^2$, where $\hat{\lambda}_1, \ldots, \hat{\lambda}_N$ are the eigenvalues of the symmetric $N \times N$-matrix M_2, that will be defined in next section (see Eqs. (9) and (10)), and $(W_1, \ldots, W_N) = (\xi_{11}, \ldots, \xi_{1n_1}, \xi_{21}, \ldots, \xi_{2n_2})H$, H being the matrix containing the eigenvectors associated to the eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_N$, that is, $M_2 = H \, diag(\hat{\lambda}_1, \ldots, \hat{\lambda}_N)H^T$. What really happens is that the set $\{\hat{\lambda}_j\}_{j=1}^N$ converges to $\{\lambda_j\}_{j\geq 1}$ (see Delhing and Mikosch 1994).

Remark 4 From Remark 3 it becomes evident that the conditional distribution of T_{2,n_1,n_2}^*, given the data, depends on the distribution of $(W_1, \ldots, W_N)$. The distribution of this random vector is, in general, unknown. For the special case where the multipliers come from a standard normal distribution, the vector $(W_1, \ldots, W_N)$ has independent components distributed according to a standard normal distribution, and thus the conditional distribution of T_{2,n_1,n_2}^*, given the data, is a finite linear combination of independent χ^2 variables, where the weights in the linear combination are $\hat{\lambda}_1, \ldots, \hat{\lambda}_N$. Note that in this special case the WB distribution of the test statistic is of the same type as its asymptotic null distribution.

4 On the practical calculation

This section describes some computational issues related to the calculation of the test statistic T_{n_1,n_2} and the WB approximation to its null distribution.

4.1 Calculation of the test statistic

Let

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \quad A = \begin{pmatrix} \frac{1}{n_1^2} 1_{n_1} 1_{n_1}^T & \frac{-1}{n_1 n_2} 1_{n_1} 1_{n_2}^T \\ \frac{-1}{n_1 n_2} 1_{n_2} 1_{n_1}^T & \frac{1}{n_2^2} 1_{n_2} 1_{n_2}^T \end{pmatrix},$$

with $M_{rs} = (\varphi_\omega(\hat{\varepsilon}_{rj} - \hat{\varepsilon}_{sv}))_{1 \le j \le n_r, \, 1 \le v \le n_v}$, $r, s = 1, 2$, and

$$\varphi_\omega(x) = \int \cos(t'x)\omega(t)dt. \tag{8}$$

Let v be the vector of $\mathbb{R}^N$ with the first n_1 components equal to $1/n_1$ and the rest equal to $-1/n_2$. In practice, the test statistic T_{n_1,n_2} can be computed by using the following expression (see Lemma 1 in Alba-Fernández et al. 2008)

$$\begin{aligned} T_{n_1,n_2} &= \frac{1}{n_1^2} \sum_{j,r=1}^{n_1} \varphi_\omega(\hat{\varepsilon}_{1j} - \hat{\varepsilon}_{1r}) + \frac{1}{n_2^2} \sum_{l,v=1}^{n_2} \varphi_\omega(\hat{\varepsilon}_{2l} - \hat{\varepsilon}_{2v}) \\ &\quad - \frac{2}{n_1 n_2} \sum_{j=1}^{n_1} \sum_{l=1}^{n_2} \varphi_\omega(\hat{\varepsilon}_{1j} - \hat{\varepsilon}_{2l}) \\ &= v^T M v \\ &= 1_N^T M_1 1_N, \end{aligned}$$

with $M_1 = M \odot A$, $\odot$ denoting the Hadamard product.

The WB version of T_{n_1,n_2}, T^*_{2,n_1,n_2}, can be expressed as

$$T^*_{2,n_1,n_2} = \xi^T M_2 \xi,$$

with $\xi^T = (\xi_{11}, \ldots, \xi_{1n_1}, \xi_{21}, \ldots, \xi_{2n_2})$ and

$$M_2 = M_3 \odot A, \quad M_3 = \begin{pmatrix} M_{3,11} & M_{3,12} \\ M_{3,21} & M_{3,22} \end{pmatrix}, \tag{9}$$

$$M_{3,rs} = \left(\int Z_{r,\tau}(\hat{\varepsilon}_{rj}; t) Z_{s,\tau}(\hat{\varepsilon}_{sv}; t)\omega(t)dt \right)_{1 \le j \le n_r, \, 1 \le v \le n_v}, \tag{10}$$

$r, s = 1, 2$. An explicit expression for M_3 is given in the Appendix.

4.2 Calculation of the WB distribution of the test statistic

Normal multipliers. As observed in Remark 4, if the multipliers have a normal distribution then, conditional on the data, T^*_{2,n_1,n_2} is distributed as $W = \sum_{j=1}^N \hat{\lambda}_j \chi^2_{1,j}$,

where $\hat{\lambda}_1, \ldots, \hat{\lambda}_N$ are the eigenvalues of M_2 and $\chi^2_{1,1}, \ldots, \chi^2_{1,N}$ are independent variables having a chi-squared distribution with 1 degree of freedom. The law of W can be numerically approximated by using, for example, Imhof's method (Imhof 1961). In this case, the WB estimator of the p-value can be calculated as follows:

Algorithm 1 1. Calculate the residuals $\hat{\varepsilon}_{11}, \ldots, \hat{\varepsilon}_{1n_1}, \hat{\varepsilon}_{21}, \ldots, \hat{\varepsilon}_{2n_2}$.
2. Calculate the observed value of the test statistic T_{n_1,n_2}, $T_{n_1,n_2,obs}$.
3. Calculate the eigenvalues of M_2, $\hat{\lambda}_1, \ldots, \hat{\lambda}_N$.
4. Approximate the p-value by $\hat{p}^* = P_* \left(\sum_{j=1}^N \hat{\lambda}_j \chi^2_{1,j} > T_{n_1,n_2,obs} \right)$.

Arbitrary multipliers. As also observed in Remark 4, the WB distribution of T_{n_1,n_2} is unknown for arbitrary multipliers. Nevertheless, the WB p-value estimator can be easily approximated by simulation as follows. Let $\Delta(u) = 1$ if $u > 0$ and $\Delta(u) = 0$ if $u \leq 0$.

Algorithm 2 1. Calculate the residuals $\hat{\varepsilon}_{11}, \ldots, \hat{\varepsilon}_{1n_1}, \hat{\varepsilon}_{21}, \ldots, \hat{\varepsilon}_{2n_2}$.
2. Calculate the observed value of the test statistic T_{n_1,n_2}, $T_{n_1,n_2,obs}$.
3. Calculate M_2.
4. For some large integer B, repeat the following steps for every $b \in \{1, \ldots, B\}$:
    (a) Generate $\xi_{11}, \ldots, \xi_{1n_1}, \xi_{21}, \ldots, \xi_{2n_2}$ IID variables with mean 0 and variance 1.
    (b) Calculate $T^{*b}_{2,n_1,n_2} = \xi^T M_2 \xi$.
5. Approximate the p-value by $\hat{p}^* = \frac{1}{B} \sum_{b=1}^B \Delta(T^{*b}_{2,n_1,n_2} - T_{n_1,n_2,obs})$.

5 Numerical results

5.1 Finite sample performance

The properties so far studied are asymptotic. In order to empirically investigate the performance of the proposed test for finite sample sizes, we carried out a simulation experiment. The objective of this experiment is fourfold: first, to study the goodness of the WB approximation to the null distribution of the test statistic; second, to analyze the WB approximation in terms of power, comparing it to the power that results when the bootstrap employed in PF07 is used to approximate the null distribution of the proposed test statistic (denoted as Boot in the tables); third, to compare the power of the proposed test to the CvM type test in PF07 (denoted as CM in the tables) (the KS test is not considered in our simulation study because, in the simulations carried out in PF07, it was less powerful than the CvM test); and finally, to compare the WB and the bootstrap approximations in terms of the CPU time required. This section reports and summarizes the obtained results. All computations in this paper have been performed by using programs written in the R language (R Core Team 2015). Specifically, to numerically approximate the WB p-value by Imhof's method the R package `CompQuadForm` (Duchesne and Lafaye de Micheaux 2010) was used.

Three specifications for the functions m_k and σ_k were considered:

S1: $Y_{kj} = X_{kj} + X^2_{kj} + (X_{kj} + 0.5)\varepsilon_{kj}, 1 \leq j \leq n_k, k = 1, 2,$

S2: $Y_{kj} = X_{kj} + 0.5\varepsilon_{kj}$, $1 \le j \le n_k$, $k = 1, 2$,

S3: $Y_{1j} = X_{1j} + X_{1j}^2 + (X_{1j} + 0.5)\varepsilon_{1j}$, $1 \le j \le n_1$, and $Y_{2j} = X_{2j} + 0.5\varepsilon_{2j}$, $1 \le j \le n_2$,

with $X_{kj} \sim U(0, 1)$, $1 \le j \le n_k$, $k = 1, 2$. For each of these specifications for m_k and σ_k, the following three cases were considered for the error distribution:

(i) $\varepsilon_1, \varepsilon_2 \sim N(0, 1)$,

(ii) $\varepsilon_1 \sim N(0, 1)$, $\varepsilon_2 \sim E(1) - 1$,

(iii) $\varepsilon_1 \sim N(0, 1)$, $\varepsilon_2 \sim U(-\sqrt{3}, \sqrt{3})$,

where $E(1)$ stands for a negative exponential law with mean 1. Case (i) corresponds to the null hypothesis, while cases (ii) and (iii) are alternatives.

To estimate the regression function and the conditional variance the Epanechnikov kernel $K(u) = 0.75 \times (1 - u^2)$ was employed.

As weight function for T_{n_1, n_2} we took $\omega(t) = \exp(-\beta t^2)$. This weight function has been considered in many other test statistics involving ECFs (see, for example, the tests in Alba-Fernández et al. 2008; Meintanis 2005; Meintanis et al. 2015; Pardo-Fernández et al. 2015a, b, among others).

Another issue is the choice of h. Since the choice of the bandwidth for tests based on smoothing remains an open issue (see, for example, de Uña-Álvarez 2013; González-Manteiga and Crujeiras 2013; Sperlich 2013), we proceeded as in the simulation studies in PF07 and Pardo-Fernández et al. (2015a): we took $h_k = c \times n_k^{-a}$, where c and a are real constants. To decide the values for a and c, we performed an extensive simulation experiment with the aim of selecting those values giving type I error probabilities closest to the nominal values. We also tried several values for β, specifically $\beta = \{0.05, 0.10, 0.15, 0.20, 0.25\}$. In general, better results -in the sense of agreement between the observed type I error probabilities and the target values- were obtained for $\beta = 0.15$. Because of this reason we fixed $\beta = 0.15$ in all simulations.

1000 samples with sizes $n_1, n_2 \in \{50, 100\}$ were generated for each case and each specification for m_k and σ_k. For each sample, to approximate the WB p-value of the observed value of the test statistic, we applied Algorithm 2, with raw and centered multipliers generated from a standard normal distribution and $B = 1000$, and Algorithm 1. In simulations we observed that, as expected, these approximations provided quite similar values. Nevertheless, the WB with centered multipliers gives slightly better results, in the sense of yielding type I error probabilities which are a bit closer to the nominal values than the other two. Because of this reason, we recommend its use. All results displayed in the tables were obtained by using Algorithm 2 with centered multipliers. To approximate the bootstrap p-value we proceeded as in PF07, generating 200 bootstrap samples. The tables report the fractions of p-values less than or equal to 0.05 and 0.10.

Tables 1, 2, 3 display the results for the level. Looking at them it can be concluded that for n_1, $n_2 = 100$ all choices for a and c in these tables give values very similar to the true value of α, for all specifications and for all tests. In general, $a = 0.3$ and $c = 1.0$ give quite reasonable results, so we set these values for a and c to study the power.

Table 4 displays the results for the power. In case (ii) all tests have similar power for all considered specifications; in case (iii) the test proposed in this paper exhibits

Table 1 Percentage of rejections at the significance levels 5% (upper entry) and 10% (lower entry) for case (i) and specification S1

a	n_1, n_2	$c = 1.00$			$c = 1.25$			$c = 1.50$		
		CM	Boot	WB	CM	Boot	WB	CM	Boot	WB
0.30	50, 50	4.60	4.50	6.40	6.00	6.00	6.10	6.00	6.10	5.90
		8.60	14.50	10.90	11.00	12.50	11.30	10.50	11.00	11.80
	50, 100	5.40	6.50	5.10	4.50	4.00	5.10	4.00	4.40	5.10
		10.40	9.00	9.60	9.00	8.90	10.00	9.50	10.20	10.30
	100, 100	4.40	4.20	5.60	6.00	5.50	5.30	4.00	5.40	4.80
		9.40	9.30	10.20	10.50	11.50	10.00	9.00	10.50	9.60
0.35	50, 50	6.50	6.50	6.30	6.00	4.50	6.40	6.00	6.00	6.20
		10.50	13.50	11.10	10.50	12.50	10.80	11.50	12.50	11.20
	50, 100	4.00	4.00	5.10	4.50	5.50	5.00	5.50	5.00	4.90
		11.00	9.50	9.50	10.50	9.50	9.80	9.00	8.00	10.00
	100, 100	5.50	4.00	5.20	5.00	5.20	5.60	4.50	5.50	5.50
		11.50	9.20	9.60	11.50	9.00	10.00	11.00	12.00	9.60
0.40	50, 50	6.00	5.00	6.10	4.00	6.00	6.40	6.00	4.50	6.40
		9.00	13.00	10.10	11.00	14.00	10.20	11.50	14.50	10.80
	50, 100	5.00	6.50	5.20	4.30	4.00	5.10	6.50	4.50	4.90
		12.00	12.50	9.30	10.50	9.00	9.60	9.00	9.00	9.50
	100, 100	6.00	4.10	5.70	5.70	5.50	5.20	4.70	5.50	5.60
		9.50	9.00	9.40	10.90	11.50	9.70	11.00	8.50	9.90
0.45	50, 50	6.00	3.50	5.70	6.00	6.00	5.90	4.50	6.50	6.40
		8.50	13.00	11.80	9.50	13.50	11.40	11.50	13.50	11.20
	50, 100	6.50	5.50	5.00	5.00	6.00	5.40	7.00	6.50	5.10
		10.00	12.50	10.30	12.50	13.00	9.00	11.00	10.00	9.30
	100, 100	6.00	6.00	5.80	5.80	5.50	5.70	4.60	4.50	5.30
		10.90	11.00	10.00	9.50	9.00	9.40	11.50	9.00	9.70

larger power than the one based on the empirical CDF. As for the WB and bootstrap approximations to the null distribution of T_{n_1,n_2}, the bootstrap test is slightly more powerful than the one based on the WB approximation. Nevertheless, as the sample size increases, the power of both tests become closer. This fact was also observed in Kojadinovic and Yan (2012) and Ghoudi and Rémillard (2014) for goodness-of-fit tests based on the empirical CDF. As will be seen a bit later, the practical importance of this fact resides in that for larger sample sizes the bootstrap becomes extremely time consuming.

We also compared the bootstrap and the WB approximations in terms of the required computing time. To calculate the WB approximation we used Algorithm 1 (denoted as WB1 in Table 5) and Algorithm 2 (denoted as WB2 in Table 5). For the comparisons to be fair, we took $B = 1000$ for the bootstrap and the WB2 estimators. As for the raw and the centered multipliers, the difference in the required computing time is negligible. Table 5 shows the CPU time consumed in seconds to get a p-value for testing the

Table 2 Percentage of rejections at the significance levels 5% (upper entry) and 10% (lower entry) for case (i) and specification S2

a	n_1, n_2	$c = 1.00$			$c = 1.25$			$c = 1.50$		
		CM	Boot	WB	CM	Boot	WB	CM	Boot	WB
0.30	50, 50	5.90	6.00	5.90	7.00	6.50	6.50	6.50	6.00	6.20
		11.30	11.00	11.20	12.00	11.00	11.40	12.00	9.50	11.60
	50, 100	5.30	4.50	4.50	4.50	4.00	5.10	5.50	4.00	4.40
		10.60	9.50	10.50	11.50	9.00	10.10	9.00	9.00	9.30
	100, 100	5.90	4.00	5.50	4.50	4.50	5.60	5.50	5.50	5.30
		10.90	11.50	10.30	9.00	9.10	10.20	9.00	11.00	10.10
0.35	50, 50	5.00	5.00	6.20	7.00	5.50	5.90	6.50	6.50	6.20
		9.00	13.00	11.60	10.00	11.00	11.10	11.50	11.00	11.60
	50, 100	5.50	5.50	4.90	5.00	4.00	4.60	4.00	6.50	4.90
		11.50	10.50	10.60	11.50	10.50	10.30	12.50	10.00	10.00
	100, 100	4.50	4.50	5.90	4.50	5.00	5.40	4.00	5.50	5.50
		9.50	9.00	9.60	9.50	9.00	10.10	9.00	9.50	9.80
0.40	50, 50	5.00	4.50	6.30	5.50	5.00	6.30	6.00	6.00	5.90
		8.00	11.50	11.70	10.00	13.00	11.60	10.00	11.00	11.10
	50, 100	5.00	4.50	4.90	5.50	5.50	5.00	4.50	4.00	4.60
		12.00	11.00	9.60	10.00	10.50	10.20	11.00	9.50	10.10
	100, 100	4.50	5.50	5.70	4.50	5.50	5.90	5.00	6.00	5.50
		9.00	9.50	9.30	9.50	9.50	9.50	10.00	10.00	10.20
0.45	50, 50	5.00	4.50	6.00	5.00	4.50	6.40	5.50	5.00	6.30
		8.50	10.00	12.00	11.00	12.00	11.60	9.00	13.00	11.40
	50, 100	5.50	4.00	4.80	6.50	4.50	4.80	5.50	5.00	4.80
		12.50	11.00	10.00	12.00	12.00	9.20	11.50	10.50	10.10
	100, 100	4.50	5.50	6.60	4.00	4.00	5.80	5.50	6.00	6.00
		9.00	9.50	10.00	9.00	9.00	9.20	9.00	11.00	9.50

equality of the error distribution for several sample sizes. Looking at this table it becomes evident that WB2 is more efficient than the bootstrap approximation, in terms of the required computing time, specially for larger sample sizes. The difference between WB1 and WB2 is really small. The gain in computational efficiency of the WB over the bootstrap stems from the fact that one does not have to re-estimate m and σ at each iteration, which slows down the process considerably; by contrast, for the WB approximation, once the matrix M_3 is calculated, the WB replicates $T_{2,n_1,n_2}^{*1}, \ldots, T_{2,n_1,n_2}^{*B}$ are calculated very rapidly.

Finally, we ran simulations when the error distributions come from a mixed distribution. Specifically, we considered the following cases:

(iv)

$$\varepsilon_1, \varepsilon_2 \sim \begin{cases} 0, & \text{with probability } 0.20, \\ N(0, \sqrt{5/4}) & \text{with probability } 0.80, \end{cases}$$

Table 3 Percentage of rejections for at the significance levels 5% (upper entry) and 10% (lower entry) for case (i) and specification S3

a	n_1, n_2	$c = 1.00$			$c = 1.25$			$c = 1.50$		
		CM	Boot	WB	CM	Boot	WB	CM	Boot	WB
0.30	50, 50	4.00	7.00	6.40	7.50	6.50	6.10	6.00	7.00	5.90
		10.00	12.50	10.90	11.50	13.50	11.30	12.50	10.50	11.80
	50, 100	5.20	5.00	5.10	6.50	5.50	5.10	6.00	4.00	5.10
		11.60	11.00	9.60	10.50	9.00	10.00	11.50	9.00	10.30
	100, 100	5.80	4.00	5.60	4.50	4.50	5.30	4.50	5.50	4.80
		9.80	9.50	10.20	10.50	11.50	10.00	11.50	11.00	9.60
0.35	50, 50	4.00	4.00	6.20	5.50	6.50	6.60	7.00	6.50	6.80
		11.50	13.50	11.40	13.50	12.00	11.10	11.50	13.00	12.40
	50, 100	6.50	6.50	5.00	7.00	7.50	5.20	4.50	5.50	4.90
		11.00	11.50	9.50	10.00	9.50	9.80	9.00	11.50	10.20
	100, 100	4.00	4.00	5.50	4.00	6.00	5.40	6.50	6.50	5.50
		9.00	9.10	10.10	10.00	12.50	10.30	9.50	11.50	10.40
0.40	50, 50	5.50	5.50	6.10	7.50	5.00	6.30	5.50	5.50	6.40
		9.50	12.00	11.70	11.50	13.50	11.40	13.00	12.50	11.20
	50, 100	5.00	5.50	5.10	6.50	5.00	4.90	6.00	4.00	5.00
		12.50	12.50	9.50	11.00	12.00	9.60	11.50	10.00	9.70
	100, 100	4.50	5.50	5.90	4.00	4.00	5.50	4.80	4.00	5.30
		9.50	11.50	9.50	9.00	9.50	10.10	11.00	12.00	10.30
0.45	50, 50	5.50	4.00	6.00	6.00	6.00	6.50	4.50	4.50	6.30
		11.50	14.50	12.00	9.50	12.00	11.80	11.50	13.50	11.30
	50, 100	7.00	5.50	5.40	5.00	5.00	5.10	4.50	5.00	4.70
		13.00	11.50	10.40	12.00	13.50	9.50	10.50	11.50	9.70
	100, 100	5.50	5.00	6.10	5.50	5.50	5.90	4.30	6.00	5.60
		10.50	11.50	10.00	9.00	9.00	9.60	11.50	9.00	10.20

(v)

$$\varepsilon_1 \sim \begin{cases} 0, & \text{with probability } 0.20, \\ N(0, \sqrt{5/4}) & \text{with probability } 0.80, \end{cases}$$

$$\varepsilon_2 \sim \begin{cases} 0, & \text{with probability } 0.50, \\ N(0, \sqrt{2}) & \text{with probability } 0.50. \end{cases}$$

Case (iv) corresponds to the null hypothesis and case (v) is an alternative. In practice, these cases could model a situation where the observations come from two devices, one of them with no measurement error. The test in PF07 cannot be applied in this setting because it requires the error distribution to have a smooth density. Table 6 displays the obtained results for the test proposed in this paper for $n_1, n_2 = 100$. Again the empirical levels are close to the target values and the test has power against the alternative.

Table 4 Percentage of rejections at the significance levels 5% and 10% for cases (ii) and (iii) and specifications S1–S3

	Case	n_1, n_2	$\alpha = 0.05$			$\alpha = 0.10$		
			CM	Boot	WB	CM	Boot	WB
S1	(ii)	50, 50	76.00	75.60	75.10	85.60	85.30	84.00
		50, 100	88.20	87.60	88.30	92.60	92.70	90.80
		100, 100	95.60	97.70	96.80	98.00	99.40	98.70
	(iii)	50, 50	14.40	23.20	20.60	27.60	35.20	29.10
		50, 100	27.20	29.70	26.20	40.60	44.20	42.00
		100, 100	33.40	49.00	46.60	47.60	64.20	61.10
S2	(ii)	50, 50	76.20	75.90	71.00	81.00	83.90	82.70
		50, 100	83.60	87.50	86.90	89.20	92.80	91.80
		100, 100	96.70	97.70	97.90	97.90	99.10	98.70
	(iii)	50, 50	16.50	24.40	18.70	23.00	35.80	29.30
		50, 100	20.80	29.00	24.60	31.90	43.60	38.00
		100, 100	40.50	50.00	47.00	55.00	64.90	61.80
S3	(ii)	50, 50	72.20	73.00	70.10	82.80	84.00	78.90
		50, 100	85.60	84.50	83.50	90.00	91.80	90.20
		100, 100	98.00	97.00	96.00	99.20	98.80	97.70
	(iii)	50, 50	18.40	27.30	20.30	31.60	38.90	32.80
		50, 100	28.60	31.90	27.30	43.20	48.00	43.60
		100, 100	37.60	53.60	51.20	54.40	68.00	65.40

Table 5 CPU time in seconds for the calculation of a p-value

n_1, n_2	Boot/WB2	WB1	WB2
50, 50	13.12	1.10	1.20
50, 100	19.43	1.20	1.35
100, 100	25.05	1.32	1.45
100, 150	36.13	1.41	1.70
150, 150	35.63	1.53	1.85

5.2 Real data analysis

Finally, we applied the proposed test to a real data set. To estimate the p-value we applied Algorithm 2 with $B = 1000$. Several values for a and c were tried. Next we briefly describe it.

Young and Bowman (1995) proposed a method for testing the equality and parallelism of two or more smooth curves. Their method assumes that the errors are equally distributed in each population. They applied their method to a data set consisting of the yield (g/plant) and density (plants/m^2) of White Spanish Onions from two South Australian localities, namely Purnong Landing (first group, 42 observations) and Virginia (second group, 42 observations). This data set is available in the R package sm

Table 6 Percentage of rejections at the significance levels 5% (upper entry) and 10% (lower entry) for cases (iv) and (v) and specifications S1–S3

	a	$c = 1.00$		$c = 1.25$		$c = 1.50$	
		(iv)	(v)	(iv)	(v)	(iv)	(v)
S1	0.30	5.00	52.00	5.50	51.00	5.00	47.50
		10.50	70.00	10.00	68.50	10.00	66.50
	0.35	5.50	52.00	5.50	51.50	5.00	50.50
		10.00	69.50	10.50	70.00	9.50	69.50
	0.40	5.40	50.50	5.50	51.50	5.50	52.00
		10.00	69.50	10.00	69.50	10.00	70.50
S2	0.30	5.50	47.50	5.00	44.00	5.00	42.50
		10.00	65.50	9.00	60.00	9.00	56.50
	0.35	5.00	50.50	5.00	48.00	5.50	45.00
		9.00	68.00	10.50	65.50	9.50	60.50
	0.40	5.00	51.50	5.00	50.00	5.00	49.00
		10.00	71.50	10.00	68.00	10.00	65.00
S3	0.30	5.00	51.50	5.00	48.00	5.00	46.50
		10.00	69.50	10.50	66.00	11.50	62.50
	0.35	5.00	51.50	5.00	51.00	5.00	49.00
		10.50	69.50	10.00	69.50	10.50	67.50
	0.40	5.50	55.00	5.00	52.00	5.50	51.50
		10.50	70.00	10.50	69.50	9.50	69.00

Table 7 p-values for the data set

a	$\beta = 0.05$			$\beta = 0.15$			$\beta = 0.25$		
	$c = 1.00$	$c = 1.25$	$c = 1.50$	$c = 1.00$	$c = 1.25$	$c = 1.50$	$c = 1.00$	$c = 1.25$	$c = 1.50$
0.30	0.961	0.743	0.496	0.944	0.570	0.393	0.920	0.521	0.363
0.35	0.959	0.943	0.760	0.989	0.850	0.579	0.976	0.840	0.529
0.40	0.863	0.970	0.946	0.861	0.996	0.883	0.790	0.989	0.848
0.45	0.570	0.903	0.971	0.422	0.915	0.996	0.345	0.846	0.988

(Bowman and Azzalini 2014). Table 7 displays the estimated p-values when using the test proposed in this paper for testing the equality of the error distributions. As in Young and Bowman (1995), the test was applied on the logarithm of the data. Looking at this table we see that the equality of the error distribution cannot be rejected.

6 Testing for the equality of $d > 2$ error distributions

The proposed test can be extended to testing for the equality of $d > 2$ error distributions as follows. Let (X_k, Y_k), $1 \leq k \leq d$, be d independent random vectors satisfying the general nonparametric regression model (1), $1 \leq k \leq d$. Let F_k and $C_k = R_k + iI_k$ denote de CDF and the CF of ε_k, respectively, $1 \leq k \leq d$. Suppose that independent

samples are available from each population: $(X_{k1}, Y_{k1}), \ldots, (X_{kn_k}, Y_{kn_k})$, $1 \leq k \leq d$. Let $N = n_1 + \cdots + n_d$. For testing

$$H_{0d} : F_1 = \cdots = F_d \quad \Leftrightarrow \quad C_1 = \cdots = C_d,$$

against the general alternative

$$H_{1d} : H_{0d} \text{ is not true,}$$

for observable data, Hušková and Meintanis (2008) have proposed to compare the ECF associated to the sample from each population to the ECF of all available data which, under H_{0d}, estimates the common CF, say $C = C_1 = \cdots = C_d$. A residual version of such test can be used for testing H_{0d} in our setting. Specifically, let $\hat{\varepsilon}_{kj}$, $1 \leq j \leq n_k$, $1 \leq k \leq d$, be defined as in (2) and let

$$T_N = \sum_{k=1}^{d} n_k \|\hat{C}_k - \hat{C}\|_\omega^2,$$

where $\hat{C}_k$ is as defined in (3) and

$$\hat{C} = \frac{1}{N} \sum_{k=1}^{d} n_k \hat{C}_k.$$

Analogue results to those given in Theorems 1, 2 and 4 can be given for T_N. Next we state them without proofs because they closely follow those provided for $d = 2$.

Theorem 5 *Suppose that $n_k/N \to \tau_k > 0$, $1 \leq k \leq d$, Assumptions (A.2) and (A.3) hold for all $1 \leq k \leq d$, then $\frac{1}{N} T_N \xrightarrow{P} \sum_{k=1}^{d} \tau_k \|C_k - C_0\|_\omega^2$, with $C_0 = \sum_{k=1}^{d} \tau_k C_k$.*

Theorem 6 *Suppose that assumptions in Theorem 5 hold. If H_0 is true, then*

$$T_N \xrightarrow{\mathcal{L}} \sum_{k=1}^{d} \|Z_k - \sqrt{\tau_k} Z_0\|_\omega^2,$$

where $\{Z_k(t), t \in \mathbb{R}\}$, $k = 1, \ldots, d$, are d IID centered Gaussian processes on $L_2(\omega)$ with covariance kernel $\varrho_0(t, s)$ as defined in Theorem 2 and $Z_0 = \sum_{k=1}^{d} \sqrt{\tau_k} Z_k$.

Now let $\xi_{1,1}, \ldots, \xi_{1,n_1}, \ldots, \xi_{d,1}, \ldots, \xi_{d,n_d}$ be IID random variates with mean 0 and variance 1, which are independent of the data, $(X_{k1}, Y_{k1}), \ldots, (X_{kn_k}, Y_{kn_k})$, $1 \leq k \leq d$. Let

$$T_N^* = \sum_{k=1}^{d} n_k \|\hat{U}_k^* - \hat{U}_0^*\|_\omega^2,$$

where $\hat{U}_k^*$ is as defined in (6) with $\hat{R}_\tau$ and $\hat{I}_\tau$ replaced by $\hat{R}_{\tau_1,...,\tau_d}$ and $\hat{I}_{\tau_1,...,\tau_d}$, respectively,

$$\hat{R}_{\tau_1,...,\tau_d}(t) = \frac{1}{N}\sum_{k=1}^{d} n_k \hat{R}_k(t), \quad \hat{I}_{\tau_1,...,\tau_d}(t) = \frac{1}{N}\sum_{k=1}^{d} n_k \hat{I}_k(t),$$

and

$$\hat{U}_0^* = \frac{1}{N}\sum_{k=1}^{d} n_k \hat{U}_k^*.$$

Theorem 7 *Suppose that assumptions in Theorem 5 hold, then*

$$\sup_x \left| P_* \left\{ T_N^* \le x \right\} - P \left\{ T_{\tau_1,...,\tau_d} \le x \right\} \right| \xrightarrow{P} 0,$$

where $T_{\tau_1,...,\tau_d} = \sum_{k=1}^{d} \| Z_{k,\tau_1,...,\tau_k} - \sqrt{\tau_k} Z_{0,\tau_1,...,\tau_k} \|_\omega^2$, $\{ Z_{k,\tau_1,...,\tau_k}(t),\ t \in \mathbb{R} \}$, $1 \le k \le d$, *are independent centered Gaussian processes on* $L_2(\omega)$ *with covariance kernel*

$$\varrho_{k,\tau_1,...,\tau_k}(t,s) = E\{ Z_{k,\tau_1,...,\tau_k}(\varepsilon_k; t) Z_{k,\tau_1,...,\tau_k}(\varepsilon_k; s) \},$$

$Z_{k,\tau_1,...,\tau_k}(\varepsilon_k; t)$ *is defined as in* (5) *with* R_τ *and* I_τ *replaced by* $R_{\tau_1,...,\tau_d}$ *and* $I_{\tau_1,...,\tau_d}$, *respectively,*

$$R_{\tau_1,...,\tau_d}(t) = \sum_{k=1}^{d} \tau_k R_r(t), \quad I_{\tau_1,...,\tau_d}(t) = \sum_{k=1}^{d} \tau_k I_r(t),$$

and $Z_{0,\tau_1,...,\tau_k} = \sum_{k=1}^{d} \sqrt{\tau_k} Z_{k,\tau_1,...,\tau_k}$.

Similar results to those stated in Corollaries 1 and 2 for T_{n_1,n_2}^* can be given for T_N^*. To save space we omit them.

Appendix

An expression for M_3

Let

$$\varphi'_\omega(t) = \frac{\partial}{\partial t}\varphi_\omega(t), \quad \varphi''_\omega(t) = \frac{\partial^2}{\partial t^2}\varphi_\omega(t),$$

where φ_ω is as defined in (8). Let D_1, D_2 the matrices defined similarly to M with φ_ω replaced by φ_ω' and φ_ω'', respectively. Let

$$e_1 = (\hat{\varepsilon}_{11}, \ldots, \hat{\varepsilon}_{1n_1}, \hat{\varepsilon}_{21}, \ldots, \hat{\varepsilon}_{2n_2})^T,$$

$$e_2 = \left(\frac{\hat{\varepsilon}_{11}^2 - 1}{2}, \ldots, \frac{\hat{\varepsilon}_{1n_1}^2 - 1}{2}, \frac{\hat{\varepsilon}_{21}^2 - 1}{2}, \ldots, \frac{\hat{\varepsilon}_{2n_2}^2 - 1}{2}\right)^T,$$

$$d_1^T = 1_N^T D_1/N,$$

$$d_2^T = 1_N^T (e_1 1_N^T \odot D_1)/N,$$

$$c^T = 1_N^T M/N,$$

$$a_1 = 1_N^T D_2 1_N/N^2,$$

$$a_2 = 1_N^T (e_1 1_N^t \odot D_2) 1_N/N^2,$$

$$a_3 = -1_N^T (e_1 1_N^t \odot D_1) 1_N/N^2,$$

$$a_4 = e_1^T D_2 e_1/N^2,$$

$$a_5 = 1_N^T M 1_N/N^2.$$

With this notation,

$$M_3 = M - e_1 d_1^T - d_1 e_1^T - e_2 d_2^T - d_2 e_2^T - 1_N c^T - c 1_N^T - a_1 e_1 e_1^T$$
$$- a_2 \left(e_1 e_2^T + e_2 e_1^T\right) - a_3 \left(e_2 1_N^T + 1_N e_2^T\right) - a_4 e_2 e_2^T + a_5 1_N 1_N^T.$$

Proofs

We now sketch the proofs of the results stated in the previous sections, as well as some preliminary results. Observe that under Assumption (A.3) (see, for example, Masry 1996)

$$\sup_{x \in S} |\hat{m}_k(x) - m_k(x)| = o_P\left(n_k^{-1/4}\right),$$
$$\sup_{x \in S} |\hat{\sigma}_k(x) - \sigma_k(x)| = o_P\left(n_k^{-1/4}\right),$$

(11)

$k = 1, 2$. Let

$$\tilde{C}_k(t) = \frac{1}{n_k} \sum_{j=1}^{n_k} \exp(it\varepsilon_{kj}) = \tilde{R}_k(t) + i\tilde{I}_k(t),$$

$$\tilde{R}_k(t) = \frac{1}{n_k} \sum_{j=1}^{n_k} \cos(t\varepsilon_{kj}), \quad \tilde{I}_k(t) = \frac{1}{n_k} \sum_{j=1}^{n_k} \sin(t\varepsilon_{kj}),$$

$k = 1, 2$.

Proof of Theorem 1 From Lemma 10(i) in Pardo-Fernández et al. (2015a) and (11),

$$T_{n_1,n_2} = \|\tilde{C}_1 - \tilde{C}_2\|_\omega^2 + o_P(1). \tag{12}$$

Theorem 2 in Alba-Fernández et al. (2008) asserts that

$$\|\tilde{C}_1 - \tilde{C}_2\|_\omega^2 \xrightarrow{a.s.} \kappa. \tag{13}$$

The result follows from (12) and (13). $\qquad\square$

Lemma 1 *If Assumptions (A.2) and (A.3) hold, then*

$$\frac{n_1 n_2}{N} T_{n_1,n_2} = \|Z_{n_1,n_2}\|_\omega^2,$$

with

$$Z_{n_1,n_2}(t) = \sqrt{\frac{n_2}{N}} U_1(t) - \sqrt{\frac{n_1}{N}} U_2(t),$$

$$U_k(t) = U_{0k}(t) + t\rho_{k,1}(t) + t^2 r_{k,2}(t),$$

$$U_{0k}(t) = \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} \left\{ \cos(t\varepsilon_{kj}) + t\varepsilon_{kj} I_k(t) - t\frac{\varepsilon_{kj}^2 - 1}{2} R_k'(t) - R_\tau + \sin(t\varepsilon_{kj}) \right.$$
$$\left. - t\varepsilon_{kj} R_k(t) - t\frac{\varepsilon_{kj}^2 - 1}{2} I_k'(t) - I_\tau \right\},$$

$\sup_t |r_{k,s}(t)| = o_P(N^{-1/2})$, $k, s = 1, 2$.

Proof We have that

$$\frac{n_1 n_2}{N} T_{n_1,n_2} = \|Z_{n_1,n_2}^0\|_\omega^2,$$

where $Z_{n_1,n_2}^0(t) = \sqrt{\frac{n_2}{N}} U_1^0(t) - \sqrt{\frac{n_1}{N}} U_2^0(t)$, $U_k^0(t) = \sqrt{n_k}\{\hat{C}_k(t) - C_\tau(t)\}$, $k = 1, 2$. From Lemma 10 in Pardo-Fernández et al. (2015b),

$$U_k^0(t) = \sqrt{n_k}\tilde{C}_k(t) + A_{k,1}(t) + A_{k,2}(t) - \sqrt{n_k}C_\tau(t) + t\rho_{k,1}(t) + t^2\rho_{k,2}(t), \quad k = 1, 2,$$

where

$$A_{k,1}(t) = it \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} \exp(it\varepsilon_{kj}) \left(\frac{\sigma_k(X_{kj}) - \hat{\sigma}_k(X_{kj})}{\sigma_k(X_{kj})} \right) \varepsilon_{kj},$$

$$A_{k,2}(t) = it \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} \exp(it\varepsilon_{kj}) \left(\frac{m_k(X_{kj}) - \hat{m}_k(X_{kj})}{\sigma_k(X_{kj})} \right)$$

and $\sup_t |\rho_{k,s}(t)| = o_P(N^{-1/2})$, $k, s = 1, 2$. From Lemma 11 in Pardo-Fernández et al. (2015b),

$$A_{k,1}(t) = -\frac{t}{2} C_k'(t) \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} (\varepsilon_{kj}^2 - 1) + t\rho_{k,3}(t), \quad \sup_t |\rho_{k,3}(t)| = o_P(1),$$

with $C_k'(t) = R_k'(t) + iI_k'(t)$. From the proof of Theorem 1 in Pardo-Fernández et al. (2015a),

$$A_{k,2}(t) = -it C_k(t) \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} \varepsilon_{kj} + t\rho_{k,4}(t), \quad \sup_t |\rho_{k,4}(t)| = o_P(1).$$

All above facts and Assumption (A.2) imply that $\|Z_{n_1,n_2}^0\|_\omega^2 = \|Z_{n_1,n_2}\|_\omega^2$. This completes the proof. □

Proof of Theorem 2 Let us continue with the notation in the statement of Lemma 1. By the central limit theorem for IID random elements in Hilbert spaces, $\{U_{01}(t), t \in \mathbb{R}\}$ converges to a centered Gaussian process $U^{(1)}$ on $L^2(w)$ with covariance structure $\varrho_0(t, s)$. By the independence of the two samples, $\{U_{02}(t), t \in \mathbb{R}\}$ converges in distribution to an independent copy $U^{(2)}$ of $U^{(1)}$. As, for constants a and b satisfying $a^2 + b^2 = 1$, the centered process $Z(t) = aU^{(1)}(t) + bU^{(2)}(t)$ has covariance structure $\varrho_0(t, s)$, and since $(\sqrt{1 - n_1/N})^2 + (\sqrt{n_1/N})^2 = 1$ and n_1/N converges to τ, it follows that $\{Z_{n_1,n_2}(t), t \in \mathbb{R}\}$ converges in law to $\{Z(t), t \in \mathbb{R}\}$, under H_0. Finally, the result follows from the continuous mapping theorem. □

Proof of Theorem 3 Note that

$$\frac{n_1 n_2}{N} T_{1,n_1,n_2}^* = \|Z_{n_1,n_2}^*\|_\omega^2,$$

where

$$Z_{n_1,n_2}^* = \sqrt{\frac{n_2}{N}} U_1^*(t) - \sqrt{\frac{n_1}{N}} U_2^*(t),$$

where $U_k^*(t) = \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} \xi_{kj} Z_{k,\tau}(\varepsilon_{kj}; t)$, $k = 1, 2$.

First, it will be shown that conditional on $(X_{11}, Y_{11}), \ldots, (X_{1n_1}, Y_{1n_1})$, $\{U_1^*(t), t \in \mathbb{R}\}$ converges in law to $\{U_{1\tau}(t), t \in \mathbb{R}\}$ on $L_2(\omega)$, where $\{U_{1\tau}(t), t \in \mathbb{R}\}$ is a centered Gaussian process on $L_2(\omega)$ with covariance kernel $\varrho_{1,\tau}(t, s)$. To achieve this result we will apply Theorem 1.1 in Kundu et al. (2000). Next we will show that conditions (i)–(iii) in that theorem hold.

Note that $E_*\{\xi_{1j} Z_{1,\tau}(\varepsilon_{1j}; t)\} = 0$, $1 \leq j \leq n_1$. Denote

$$c_{n_1}(t, s) = Cov_*\{U_1^*(t), U_1^*(s)\} = \frac{1}{n_1} \sum_{j=1}^{n_1} Z_{1,\tau}(\varepsilon_{1j}; t) Z_{1,\tau}(\varepsilon_{1j}; s).$$

From the strong law of large numbers,

$$c_{n_1}(t, s) \xrightarrow{a.s.} \varrho_{1,\tau}(t, s), \quad \forall s, t \in \mathbb{R}. \tag{14}$$

Note also that

$$|c_{n_1}(t, s)| \le g(t, s), \quad \forall s, t \in \mathbb{R}, \tag{15}$$

with

$$g(t, s) = \varpi_1 + \varpi_2(|t| + |s|) + \varpi_3 |t||s|, \quad a.s.$$

for certain positive constants $\varpi_1, \varpi_2, \varpi_3$.

Let $\{e_k, k \ge 0\}$ be an orthonormal basis of $L_2(\omega)$. Let V_1 denote the covariance operator of U_1^* and let V_τ denote the covariance operator of $U_{1\tau}(t)$. From (14) and (15), by the dominated convergence theorem,

$$
\begin{aligned}
\lim \langle V_1 e_k, e_l \rangle_\omega &= \lim \int c_{n_1}(s, t) e_k(t) e_l(s) \omega(t) \omega(s) dt ds \\
&= \int \varrho_{1,\tau}(t, s) e_k(t) e_l(s) \omega(t) \omega(s) dt ds = \langle V_\tau e_k, e_l \rangle_\omega \quad a.s.
\end{aligned}
$$

Thus taking $a_{kl} = \langle V_\tau e_k, e_l \rangle_\omega$, the condition (i) in the aforementioned Theorem 1.1 holds. To check the condition (ii), by monotone convergence theorem, Parseval's relation and dominated convergence theorem, it follows

$$
\begin{aligned}
\lim \sum_{k=0}^\infty \langle V_1 e_k, e_l \rangle_\omega &= \lim \sum_{k=0}^\infty E_* \{ \langle U_1^*(t), e_k \rangle_\omega^2 \} = \lim E_* \{ \| U_1^*(t) \|_\omega^2 \} \\
&= \int \lim c_{n_1}(t, t) \omega(t) dt = \int \varrho_{1,\tau}(t, t) \omega(t) dt \\
&= E \| U_{1\tau} \|_\omega^2 < \infty \quad a.s.
\end{aligned}
$$

Before verifying condition (iii), we first notice that

$$
\left| \langle \frac{1}{\sqrt{n_1}} \xi_{1j} Z_{1,\tau}(\varepsilon_{1j}; t), e_k \rangle_\omega \right| \le \frac{|\xi_{1j}|}{\sqrt{n_1}} \left(\int Z_{1,\tau}^2(\varepsilon_{1j}; t) \omega(t) dt \right)^{1/2} \le g(t, t) \frac{|\xi_{1j}|}{\sqrt{n_1}}.
$$

From the above inequality,

$$
\begin{aligned}
\sum_{j=1}^{n_1} E_* \left[\langle \frac{1}{\sqrt{n_1}} \xi_{1j} Z_{1,\tau}(\varepsilon_{1j}; t), e_k \rangle_\omega^2 I \left\{ \left| \langle \frac{1}{\sqrt{n_1}} \xi_{1j} Z_{1,\tau}(\varepsilon_{1j}; t), e_k \rangle_\omega \right| > \epsilon \right\} \right] \\
\le \frac{g^2(t,t)}{n_1} \sum_{j=1}^{n_1} E_* \left[\xi_{1j}^2 I\{|\xi_{1j}| > g^{-1}(t, t)\epsilon \sqrt{n_1}\} \right] \\
= g^2(t, t) E_* \left[\xi_{1j}^2 I\{|\xi_{1j}| > g^{-1}(t, t)\epsilon \sqrt{n_1}\} \right] \to 0,
\end{aligned}
$$

$\forall \epsilon > 0, \forall k \ge 0$.

Analogously, conditional on $(X_{21}, Y_{21}), \ldots, (X_{2n_2}, Y_{2n_2})$, $\{U_2^*(t), t \in \mathbb{R}\}$ converges in law to $\{U_{2\tau(t)}, t \in \mathbb{R}\}$ on $L_2(\omega)$, where $\{U_{2\tau}(t), t \in \mathbb{R}\}$ is a centered Gaussian process on $L_2(\omega)$ with covariance kernel $\varrho_{2,\tau}(t, s)$. A similar argument to that in the proof of Theorem 2 shows that, conditional on $(X_{11}, Y_{11}), \ldots, (X_{1n_1}, Y_{1n_1})$, $(X_{21}, Y_{21}), \ldots, (X_{2n_2}, Y_{2n_2})$, $\{Z_{1,n_1,n_2}^*(t), t \in \mathbb{R}\}$ converges in law to $\{Z_\tau(t), t \in \mathbb{R}\}$ on $L_2(\omega)$. □

Lemma 2 *Suppose that Assumption (A.3) holds, then*

(a) $\frac{1}{n_k} \sum_{j=1}^{n_k} (\varepsilon_{kj} - \hat{\varepsilon}_{kj})^2 = o_P(1), \ k = 1, 2.$

(b) $\frac{1}{n_k} \sum_{j=1}^{n_k} (\hat{\varepsilon}_{kj}^2 - \varepsilon_{kj}^2)^2 = o_P(1), \ k = 1, 2.$

(c) $\frac{1}{n_k} \sum_{j=1}^{n_k} (\hat{\varepsilon}_{kj}^2 - 1)^2 = O_P(1), \ k = 1, 2.$

(d) $\frac{1}{n_k} \sum_{j=1}^{n_k} \hat{\varepsilon}_{kj}^2 = O_P(1), \ k = 1, 2.$

Proof The difference between the residuals and the errors can be written as follows

$$\hat{\varepsilon}_{kj} - \varepsilon_{kj} = \varepsilon_{kj} \left(\frac{\sigma_k(X_{kj}) - \hat{\sigma}_k(X_{kj})}{\hat{\sigma}_k(X_{kj})} \right) + \left(\frac{m_k(X_{kj}) - \hat{m}_k(X_{kj})}{\hat{\sigma}_k(X_{kj})} \right), \quad (16)$$

$k = 1, 2.$ The results in (a)–(d) follow from (11) and (16). □

Lemma 3 *Suppose that Assumptions (A.1)–(A.3) hold, then*

(a) $\|t(\hat{R}_k - R_k)\|_\omega^2 = o_P(1), \ \|t(\hat{I}_k - I_k)\|_\omega^2 = o_P(1), \quad k = 1, 2.$

(b) $\|R_\tau - \hat{R}_\tau\|_\omega^2 = o_P(1), \quad \|I_\tau - \hat{I}_\tau\|_\omega^2 = o_P(1),$

(c) $\|t(R_k' - \hat{R}_k')\|_\omega^2 = o_P(1), \ \|t(I_k' - \hat{I}_k')\|_\omega^2 = o_P(1). \quad k = 1, 2.$

Proof (a) By the mean value theorem,

$$t\{\hat{R}_k(t) - \tilde{R}_k(t)\} = -t^2 r_k(t).$$

From Lemma 2(a) and the Cauchy–Schwarz inequality,

$$\sup_t |r_k(t)| \leq \frac{1}{n_k} \sum_{j=1}^{n_k} |\hat{\varepsilon}_{kj} - \varepsilon_{kj}| \leq \left(\frac{1}{n_k} \sum_{j=1}^{n_k} (\hat{\varepsilon}_{kj} - \varepsilon_{kj})^2 \right)^{1/2} = o_P(1).$$

Therefore,

$$\|t(\hat{R}_k - \tilde{R}_k)\|_\omega^2 \leq \sup_t |r_k(t)|^2 \int t^4 \omega(t) dt = o_P(1). \quad (17)$$

We also have that

$$\|t(R_k - \tilde{R}_k)\|_\omega^2 = o_P(1). \quad (18)$$

Finally, (17) and (18) both imply that $\|t(\hat{R}_k - R_k)\|_\omega^2 = o_P(1)$. The proof for $\|t(\hat{I}_k - I_k)\|_\omega^2$ is parallel. The proof of parts (b) and (c) follow similar steps. □

Proof of Theorem 4 $\frac{n_1 n_2}{N} T_{2,n_1,n_2}^*$ can be expressed as $\frac{n_1 n_2}{N} T_{2,n_1,n_2}^* = D_1 + D_2 + 2D_3$, where $D_3^2 \leq D_1 D_2$, $D_1 = \frac{n_1 n_2}{N} T_{1,n_1,n_2}^*$, $D_2 = \frac{n_1 n_2}{N} \|(\hat{U}_1^* - \hat{U}_2^*) - (C_1^* - C_2^*)\|_\omega^2$. From Theorem 3,

$$\sup_x |P_*(D_1 \leq x) - P(T_\tau \leq x)| \xrightarrow{P} 0,$$

Thus, to show the result it suffices to see that $D_2 = o_{p_*}(1)$ in probability. With this aim, observe that D_2 can be expressed as

$$D_2 = \sum_{k=1,2} \sum_{j=1}^{8} S_{k,j} + \sum_{k,s=1,2} \sum_{1 \leq l \neq j \leq 8} S_{k,s,j,l},$$

with $S_{k,s,j,l}^2 \le S_{k,j} S_{s,l}$, $1 \le j, l \le 8$, $k, s = 1, 2$,

$$S_{k,1} = \frac{N-n_k}{N} \| \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} \{\cos(t\hat{\varepsilon}_{kj}) - \cos(t\varepsilon_{kj})\} \xi_{k,j} \|_\omega^2,$$

$$S_{k,2} = \frac{N-n_k}{N} \| \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} \{\sin(t\hat{\varepsilon}_{kj}) - \sin(t\varepsilon_{kj})\} \xi_{k,j} \|_\omega^2,$$

$$S_{k,3} = \frac{N-n_k}{N} \| \frac{t}{\sqrt{n_k}} \sum_{j=1}^{n_k} (\hat{\varepsilon}_{kj} \hat{R}_k - \varepsilon_{kj} R_k) \xi_{k,j} \|_\omega^2,$$

$$S_{k,4} = \frac{N-n_k}{N} \| \frac{t}{\sqrt{n_k}} \sum_{j=1}^{n_k} (\hat{\varepsilon}_{kj} \hat{I}_k - \varepsilon_{kj} I_k) \xi_{k,j} \|_\omega^2,$$

$$S_{k,5} = \frac{N-n_k}{N} \| \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} (R_\tau - \hat{R}_\tau) \xi_{1,k} \|_\omega^2,$$

$$S_{k,6} = \frac{N-n_k}{N} \| \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} (I_\tau - \hat{I}_\tau) \xi_{k,j} \|_\omega^2,$$

$$S_{k,7} = \frac{N-n_k}{N} \| \frac{t}{2\sqrt{n_k}} \sum_{j=1}^{n_k} \{(\hat{\varepsilon}_{kj}^2 - 1) \hat{I}_k' - (\varepsilon_{kj}^2 - 1) I_k'\} \xi_{k,j} \|_\omega^2,$$

$$S_{k,8} = \frac{N-n_k}{N} \| \frac{t}{2\sqrt{n_k}} \sum_{j=1}^{n_k} \{(\hat{\varepsilon}_{kj}^2 - 1) R_k' - (\varepsilon_{kj}^2 - 1) \hat{R}_k'\} \xi_{k,j} \|_\omega^2,$$

$k = 1, 2$. We will show that $S_{k,j} = o_{p_*}(1)$ in probability, $1 \le j \le 8$, $k = 1, 2$.

By the mean value theorem,

$$S_{1,1} = \frac{n_2}{N} \frac{1}{n_1} \sum_{j,m=1}^{n_1} \xi_{1j} \xi_{1m} (\hat{\varepsilon}_{1j} - \varepsilon_{1j})(\hat{\varepsilon}_{1m} - \varepsilon_{1m}) \int t^2 \sin(t\tilde{\varepsilon}_{1j}) \sin(\tilde{\varepsilon}_{1m}) \omega(t) dt,$$

where $\tilde{\varepsilon}_{1j} = \alpha_{1j} \varepsilon_{1j} + (1 - \alpha_{1j}) \hat{\varepsilon}_{1j}$, for some $\alpha_{1j} \in (0, 1)$. Then, from Lemma 2(a),

$$E_*(S_{1,1}) \le \frac{n_2}{N} \frac{1}{n_1} \sum_{j=1}^{n_1} (\varepsilon_{1j} - \hat{\varepsilon}_{1j})^2 \int t^2 \omega(t) dt = o_p(1),$$

which implies $S_{1,1} = o_{p_*}(1)$ in probability. Analogously, $S_{2,1} = o_{p_*}(1)$, $S_{1,2} = o_{p_*}(1)$, $S_{2,2} = o_{p_*}(1)$ in probability.

Observe that $S_{1,3} = S_{13} + S_{23} + 2S_{33}$, with $S_{33}^2 \le S_{13} S_{23}$,

$$S_{13} = \frac{n_2}{N} \frac{1}{n_1} \sum_{j,m=1}^{n_1} (\hat{\varepsilon}_{1j} - \varepsilon_{1j})(\hat{\varepsilon}_{1m} - \varepsilon_{1m}) \xi_{1j} \xi_{1m} \| t R_1 \|_\omega^2,$$

$$S_{23} = \frac{n_2}{N} \frac{1}{n_1} \sum_{j,m=1}^{n_1} \hat{\varepsilon}_{1j} \hat{\varepsilon}_{1m} \xi_{1j} \xi_{1m} \| t (\hat{R}_1 - R_1) \|_\omega^2.$$

From Lemma 2(a) it follows that $E_*(S_{13}) = o_P(1)$ and thus $S_{13} = o_{p_*}(1)$ in probability. From Lemma 2(d) and Lemma 3(a) it follows that $E_*(S_{23}) = o_P(1)$ and thus $S_{23} = o_{p_*}(1)$ in probability. Therefore, $S_{1,3} = o_{p_*}(1)$ in probability. Analogously, $S_{2,3} = o_{p_*}(1)$, $S_{1,4} = o_{p_*}(1)$ and $S_{2,4} = o_{p_*}(1)$ in probability.

Since $S_{1,5} = \frac{n_2}{N} \left(\frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} \xi_{kj} \right)^2 \| R_\tau - \hat{R}_\tau \|_\omega^2$, the central limit theorem and Lemma 3(b) imply that $E_*(S_{1,5}) = o_P(1)$ and thus $S_{1,5} = o_{p_*}(1)$ in probability. Analogously, $S_{2,5} = o_{p_*}(1)$, $S_{1,6} = o_{p_*}(1)$ and $S_{1,6} = o_{p_*}(1)$ in probability.

Observe that $S_{1,7} = S_{17} + S_{27} + 2S_{37}$, with $S_{37}^2 \le S_{17}S_{27}$,

$$S_{17} = \frac{n_2}{N}\frac{1}{4}\frac{1}{n_1}\sum_{j,m=1}^{n_1}(\hat{\varepsilon}_{1j}^2 - 1)(\hat{\varepsilon}_{1m}^2 - 1)\xi_{1j}\xi_{1m}\|t(\hat{I}_1' - I_1')\|_\omega^2,$$

$$S_{27} = \frac{n_2}{N}\frac{1}{4}\frac{1}{n_1}\sum_{j,m=1}^{n_1}(\hat{\varepsilon}_{1j}^2 - \varepsilon_{1j}^2)(\hat{\varepsilon}_{1m}^2 - \varepsilon_{1m}^2)\xi_{1j}\xi_{1m}\|tI_k'\|_\omega^2.$$

From Lemma 2(c) and Lemma 3(c), it follows that $E_*(S_{17}) = o_P(1)$ and thus $S_{17} = o_{p_*}(1)$ in probability. From Lemma 2(b) it follows that $E_*(S_{27}) = o_P(1)$ and thus $S_{27} = o_{p_*}(1)$, in probability. Therefore, $S_{1,7} = o_{p_*}(1)$, in probability. Analogously, $S_{2,7} = o_{p_*}(1)$, $S_{1,8} = o_{p_*}(1)$, and $S_{2,8} = o_{p_*}(1)$ in probability. This completes the proof.

References

Alba-Fernández V, Jiménez-Gamero MD, Muñoz-García J (2008) A test for the two-sample problem based on empirical characteristic functions. Comput Stat Data Anal 52:3730–3748

Baringhaus L, Kolbe D (2015) Two-sample tests based on empirical Hankel transforms. Stat Pap 56:597–617

Bowman AW, Azzalini, A (2014) R package 'sm': nonparametric smoothing methods (version 2.2-5.4). http://www.stats.gla.ac.uk/~adrian/sm, http://azzalini.stat.unipd.it/Book_sm

Burke MD (2000) Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap. Stat Probab Lett 46:13–20

de Uña-Álvarez J (2013) Comments on: an updated review of Goodness-of-Fit tests for regression models. Test 22:414–418

Delhing H, Mikosch T (1994) Random quadratic forms and the bootstrap for U-statistics. J Multivar Anal 51:392–413

Duchesne P, Lafaye de Micheaux P (2010) Computing the distribution of quadratic forms: further comparisons between the Liu-Tang-Zhang approximation and exact methods. Comput Stat Data Anal 54:858–862

Feller W (1971) An introduction to probability theory and its applications, vol 2. Wiley, New York

Ghoudi K, Rémillard B (2014) Comparison of specification tests for GARCH models. Comput Stat Data Anal 76:291–300

González-Manteiga W, Crujeiras R (2013) An updated review of Goodness-of-Fit tests for regression models. Test 22:361–411

Hall P, Hart JD (1990) Bootstrap test for difference between means in nonparametric regression. J Am Stat Assoc 85:1039–1049

Hobza T, Morales D, Pardo L (2014) Divergence-based test of homogeneity for spatial data. Stat Pap 55:1059–1077

Hušková M, Meintanis SG (2008) Tests for the multivariate k-sample problem based on the empirical characteristic function. J Nonparametr Stat 20:263–277

Imhof JP (1961) Computing the distribution of quadratic forms in normal variables. Biometrika 48:419–426

Jiménez-Gamero MD, Alba-Fernández MV, Jodrá P, Barranco-Chamorro I (2016) Fast tests for the two-sample problem based on the empirical characteristic function. Math Comput Simul. doi:10.1016/j.matcom.2016.09.007

Jiménez-Gamero MD, Kim H-M (2015) Fast goodness-of-fit test based on the characteristic function. Comput Stat Data Anal 89:172–191

Kojadinovic I, Yan J (2012) Goodness-of-fit testing based on a weighted bootstrap: a fast sample alternative to the parametric bootstrap. Can J Stat 40:480–500

Kulasekera KB, Wang J (2001) A test of the equality of regression curves using Gâteaux scores. Austral N Z J Stat 43:89–99

Kundu S, Majumdar S, Mukherjee K (2000) Central limit theorems revisited. Stat Probab Lett 47:265–275

Masry E (1996) Multivariate local polynomial regression for time series: uniform strong consistency and rates. J Time Ser Anal 17:571–600

Meintanis SG (2005) Permutation tests for homogeneity based on the empirical characteristic function. J Nonparametr Stat 17:583–592

Meintanis SG, Ngatchou-Wandji J, Taufer E (2015) Goodness-of-fit tests for multivariate stable distributions based on the empirical characteristic function. J Multivar Anal 140:171–192

Modarres R (2016) Multinomial interpoint distances. Stat Pap. doi:10.1007/s00362-016-0766-7

Mora J (2005) Comparing distribution functions of errors in linear models: a nonparametric approach. Stat Probab Lett 73:425–432

Pardo-Fernández JC (2007) Comparison of error distributions in nonparametric regression. Stat Probab Lett 77:350–356

Pardo-Fernández JC, Jiménez-Gamero MD, El Ghouch A (2015a) A nonparametric ANOVA-type test for regression curves based on characteristic functions. Scand J Stat 42:197–213

Pardo-Fernández JC, Jiménez-Gamero MD, El Ghouch A (2015b) Tests for the equality of conditional variance functions in nonparametric regression. Electron J Stat 9:1826–1851

Quessy JF, Éthier F (2012) Cramér-von Mises and characteristic function tests for the two and k-sample problems with dependent data. Comput Stat Data Anal 56:2097–2111

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

Sperlich S (2013) Comments on: an updated review of Goodness-of-Fit tests for regression models. Test 22:419–427

Young SG, Bowman AW (1995) Non-parametric analysis of covariance. Biometrics 51:920–931