Scandinavian Journal of Statistics Theory and Applications

Scandinavian Journal of Statistics, Vol. 42: 197–213, 2015
doi: 10.1111/sjos.12102

A Non-parametric ANOVA-type Test for Regression Curves Based on Characteristic Functions

JUAN CARLOS PARDO-FERNÁNDEZ
Departamento de Estatística e I.O., Universidade de Vigo

MARÍA DOLORES JIMÉNEZ-GAMERO
Departamento de Estadística e I.O., Universidad de Sevilla

ANOUAR EL GHOUCH
ISBA, Université Catholique de Louvain

ABSTRACT. This article studies a new procedure to test for the equality of k regression curves in a fully non-parametric context. The test is based on the comparison of empirical estimators of the characteristic functions of the regression residuals in each population. The asymptotic behaviour of the test statistic is studied in detail. It is shown that under the null hypothesis, the distribution of the test statistic converges to a finite combination of independent chi-squared random variables with one degree of freedom. The coefficients in this linear combination can be consistently estimated. The proposed test is able to detect contiguous alternatives converging to the null at the rate $n^{-1/2}$. The practical performance of the test based on the asymptotic null distribution is investigated by means of simulations.

Key words: comparison of regression curves, empirical characteristic function, non-parametric regression, regression residuals

1. Introduction

Testing for the equality of the means of k populations ($k \geq 2$) is a classical problem in Statistics. When the populations are assumed to follow a normal distribution with equal variance, then the ANOVA F-test is the classical way to perform the test.

In this paper, we consider a more general setting. We assume that in each population along with the response variable, Y, we observe another variable, X, the covariate, so that the mean and the variance of the response variable depend on the values of the covariate. More specifically, let $(X_j, Y_j), 1 \leq j \leq k$, be k-independent random vectors satisfying general non-parametric regression models

$$Y_j = m_j(X_j) + \sigma_j(X_j)\varepsilon_j, \tag{1}$$

where $m_j(x) = E(Y_j \mid X_j = x)$ is the regression function, $\sigma_j^2(x) = Var(Y_j \mid X_j = x)$ is the conditional variance function and ε_j is the regression error, which is assumed to be independent of X_j. Note that by construction, $E(\varepsilon_j) = 0$ and $Var(\varepsilon_j)$=1. The covariate X_j is continuous with density function f_j. The regression functions, the variance functions, the distribution of the errors and the distribution of the covariates are unknown, and no parametric models are assumed for them. Under this framework, our approach is fully non-parametric.

In this conditional setting, the hypothesis of equality of means is stated in terms of the conditional means or regression functions

$$H_0 : m_1 = m_2 = \ldots = m_k,$$

or in other words, the mean effect of the covariates over the responses is equal in the k populations. Because the objective is to compare the regression curves, it is reasonable to assume that the covariates have common support. The alternative hypothesis is

$$H_1 : H_0 \text{ is not true.}$$

Note that this testing problem contains the simpler case described in the first paragraph as a particular case by only eliminating the covariates in the models.

The problem of testing for the equality of regression curves in non-parametric settings has been previously treated in the statistical literature. The majority of the available papers either are devoted to the comparison of only two curves or impose some restrictive assumptions such as fixed design, equal sample sizes, identical design points and homoscedasticity of the residuals. Examples of such works include Delgado (1993), Kulasekera (1995), Munk & Dette (1998), Neumeyer & Pardo-Fernández (2009) and Srihera & Stute (2010), among many others. To the best of our knowledge, the most related works to ours are the papers of Neumeyer & Dette (2003) and of Pardo-Fernández *et al.* (2007). These papers proposed and studied a procedure to test the hypothesis of equality of $k(k \geq 2)$ regression functions based on the comparison of marked empirical processes of the residuals for the former and the comparison of distribution functions of the residuals for the latter.

In this work, we investigate a new test procedure based on a weighted process generated by the characteristic functions (CFs) of the residuals. Compared with the competitors, the main advantages of the proposed method are the following. First, the test is universally consistent for any fixed alternative without any restrictions on the weight function involved in the definition of the test statistic (except for the fact that it should be positive) and without any restrictions on the distribution of the regression errors. Thus, the method can be applied when the distributions of the errors are arbitrary: continuous, discrete or mixed. This is illustrated in the simulation study where an example with errors having a mixed distribution is given. Second, the test can be used to check the equality of any number of regression curves and can detect local alternatives converging to the null at the rate $n^{-1/2}$. Third, the critical values or the p-values can be obtained from the asymptotic null distribution of the proposed test statistic. This is not the case for most of the existing methods, which typically rely on bootstrap. In our case, although we study a bootstrap version of our test statistic in order to perform a comparison with other methods, the bootstrap is not needed because we are able to obtain the desired level and a good power even for moderate sample sizes by using the asymptotic null distribution. This is clearly shown for many scenarios considered in the simulation study.

Our approach is based on comparing the CFs of the regression errors. More specifically, let $\varepsilon_j = \{Y_j - m_j(X_j)\}/\sigma_j(X_j)$ be the regression error in population j. Let m_0 be the common regression curve under the null hypothesis and define

$$\varepsilon_{0j} = \{Y_j - m_0(X_j)\}/\sigma_j(X_j) = \varepsilon_j + \{m_j(X_j) - m_0(X_j)\}/\sigma_j(X_j), \tag{2}$$

$1 \leq j \leq k$. It turns out that the null hypothesis H_0 is true if and only if, for all $1 \leq j \leq k$, the random variables ε_j and ε_{0j} have the same distribution (see theorem 1 in Pardo-Fernández *et al.*, 2007). This assessment can be interpreted in terms of the cumulative distribution function (CDF) or in terms of any other function characterizing the probability law of the errors. Pardo-Fernández *et al.* (2007) restricted their attention to the CDF.

The probability law of any random variable X is also characterized by its CF, $\varphi(t) = E\{\exp(\mathrm{i}tX)\}$. Recent years have witnessed an increasing number of proposals for hypothesis testing whose test statistics measure deviations between the empirical CF (ECF) of the available data and an estimator of the CF under the null hypothesis. In the line of the setting considered in this paper, that is, by assuming that the data are generated by regression models, are the papers by Jiménez-Gamero *et al.* (2005) and Hušková & Meintanis (2007, 2010), for testing goodness of fit for the errors, and Hušková & Meintanis (2009) for testing goodness of fit of the regression function to a parametric function. An advantage of the CF approach over the one based on the CDF, as observed in Hušková & Meintanis (2009), is that the former usually requires less stringent assumptions for its validity. In addition, from the simulation results that can be found in the literature (see, for example, Hušková & Meintanis, 2010), the tests based on the ECF compete very satisfactorily with those based on the empirical CDF (ECDF).

Having in mind the reasons earlier, the purpose of the present paper is to test H_0 by comparing consistent estimators of the CFs of the random variables ε_j and ε_{0j}, say $\hat{\varphi}_j(t)$ and $\hat{\varphi}_{0j}(t)$, respectively, $1 \le j \le k$. To measure deviations between these estimators, we consider a Cramér–von Mises type test statistic. In order to derive the asymptotic null distribution of the test statistic, we first give a stochastic expansion for the differences $\hat{\varphi}_j(t) - \hat{\varphi}_{0j}(t)$. As a consequence of this expansion, it is shown that the test statistic converges in law to a finite linear combination of independent chi-squared variables. Moreover, under certain weak conditions on the distributions of the errors and the covariates, the asymptotic null distribution is proportional to a χ^2_{k-1} distribution. We provide consistent estimators of the coefficients appearing in this linear combination, which allow us to propose a consistent estimator of the asymptotic null distribution. The behaviour of the test under fixed and local alternatives is also studied. Specifically, it is shown that the proposed test is able to detect any fixed alternative and contiguous alternatives converging to the null at the rate $n^{-1/2}$, where n denotes the total sample size. In contrast to the procedure in Pardo-Fernández *et al.* (2007), to derive these properties, we do not require further assumptions on the distribution on the errors, such as that they have a probability density. For finite sample, the quality of the proposed approximation of the null distribution of the test statistic and the power are investigated numerically. From this numerical study, we conclude that the proposed approximation of the null distribution works adequately for moderate sample sizes, and in terms of power, the proposed test competes very satisfactorily with those based on the ECDF.

The paper is organized as follows. Section 2 introduces the test statistic and also provides an alternative expression, which is useful from a computational point of view. The asymptotic null distribution of the test statistic and the behaviour of the test under fixed and local alternatives are studied in Section 3. Section 4 reports a summary of a numerical study conducted to study the practical performance of the test and to compare it with other existing methods. Section 5 concludes the paper. All proofs of the theoretical results are deferred to the Supporting information.

The following notation will be used along the paper: P_0 denotes probability assuming that H_0 is true; E_0 denotes expectation assuming that H_0 is true; P_* denotes the conditional probability law, given the data; all limits in this paper are taken when $n \to \infty$; $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution; $\xrightarrow{P}$ denotes convergence in probability; $\xrightarrow{a.s.}$ denotes the almost sure convergence; if $x \in \mathbb{R}^k$, with $x' = (x_1, \ldots, x_k)$, then $diag(x)$ is the $k \times k$ diagonal matrix whose (i,i) entry is x_i, $1 \le i \le k$; for any complex number $z = a + \mathrm{i}b$, $Re(z) = a$ is its real part, $\bar{z} = a - \mathrm{i}b$ is its conjugate and $|z|$ is its modulus; $N_k(\mu, \Sigma)$ denotes the multivariate normal distribution with mean vector μ and variance-covariance matrix Σ; an unspecified integral denotes integration over the whole real line $\mathbb{R}$; for a given non-negative real-valued function w and for any complex-valued measurable function g, we denote $\|g\|_w = \left(\int |g(t)|^2 w(t) dt\right)^{1/2}$ to the norm in the Hilbert space $L^2(\mathbb{R}, w)$.

2. The test statistic

Let $(X_{jl}, Y_{jl}), 1 \leq l \leq n_j$ be independent and identically distributed observations from $(X_j, Y_j), 1 \leq j \leq k$. Let $f_j(x)$ be the probability density function (PDF) of $X_j, n = \sum_{j=1}^k n_j$, and let $f_{mix}(x) = \sum_{j=1}^k p_j f_j(x)$ be the PDF of the mixture of covariates according to the weights $p_1, \ldots, p_k$, where $p_j = \lim n_j/n$. In order to estimate the errors, we first need to estimate the regression functions, $m_j(x) = E(Y_j | X_j = x)$, the variance functions, $\sigma_j^2(x) = E\left[\{Y_j - m_j(x)\}^2 | X_j = x\right]$, and the common regression function under $H_0, m_0(x) = \sum_{j=1}^k p_j\{f_j(x)/f_{mix}(x)\}m_j(x)$. With this aim, we use non-parametric estimators on the basis of kernel smoothing techniques. Let K denote a non-negative kernel function defined on $\mathbb{R}$, and let $0 < h_n \equiv h \to 0$ be the bandwidth or smoothing parameter and $K_h(x) = h^{-1}K(x/h)$. We use the following estimators for the functions m_j, σ_j^2 and m_0:

$$\hat{m}_j(x) = \sum_{l=1}^{n_j} w_{jl}(x) Y_{jl}, \quad \hat{\sigma}_j^2(x) = \sum_{l=1}^{n_j} w_{jl}(x) Y_{jl}^2 - \hat{m}_j^2(x),$$

$$\hat{m}_0(x) = \sum_{j=1}^k \frac{n_j}{n} \frac{\hat{f}_j(x)}{\hat{f}_{mix}(x)} \hat{m}_j(x),$$

where

$$\hat{f}_j(x) = n_j^{-1} \sum_{l=1}^{n_j} K_h(x - X_{jl}), \quad \hat{f}_{mix}(x) = \sum_{j=1}^k \frac{n_j}{n} \hat{f}_j(x),$$

$1 \leq j \leq k$. The quantities w_{jl} are either the local-linear weights given by

$$w_{jl}(x) = \frac{K_h(X_{jl} - x)\left\{S_{2,n_j}(x) - (X_{jl} - x)S_{1,n_j}\right\}}{S_{0,n_j}(x)S_{2,n_j}(x) - S_{1,n_j}^2(x)},$$

with $S_{k,n_j}(x) = \sum_{l=1}^{n_j}(X_{jl} - x)^k K_h(X_{jl} - x), k = 0, 1, 2$, or the Nadaraya–Watson weights

$$w_{jl}(x) = \frac{K_h(X_{jl} - x)}{\sum_{v=1}^{n_j} K_h(X_{jv} - x)}.$$

Both are particular cases of local-polynomial weighting (Fan & Gijbels, 1996). Under the model assumptions that will be stated in the next section, the results in this article are valid for local-linear and for Nadaraya–Watson (local-constant) estimators.

On the basis of these estimators, for each population $j, 1 \leq j \leq k$, we construct two samples of residuals:

$$\hat{\varepsilon}_{jl} = \frac{Y_{jl} - \hat{m}_j(X_{jl})}{\hat{\sigma}_j(X_{jl})} \quad \text{and} \quad \hat{\varepsilon}_{0jl} = \frac{Y_{jl} - \hat{m}_0(X_{jl})}{\hat{\sigma}_j(X_{jl})},$$

$1 \leq l \leq n_j$, whose ECFs are

$$\hat{\varphi}_j(t) = \frac{1}{n_j} \sum_{l=1}^{n_j} \exp(\mathrm{i}t\hat{\varepsilon}_{jl}) \quad \text{and} \quad \hat{\varphi}_{0j}(t) = \frac{1}{n_j} \sum_{l=1}^{n_j} \exp\left(\mathrm{i}t\hat{\varepsilon}_{0jl}\right),$$

respectively. These ECFs are nothing but (consistent) kernel-based non-parametric estimators of the population CFs $\varphi_j(t) = E\{\exp(\mathrm{i}t\varepsilon_j)\}$ and $\varphi_{0j}(t) = E\{\exp(\mathrm{i}t\varepsilon_{0j})\}$, respectively, where ε_{0j} is as defined in (2). The testing procedure consists of comparing $\hat{\varphi}_j(t)$ and $\hat{\varphi}_{0j}(t), 1 \leq j \leq$

k, using a weighted L_2-distance. More precisely, following the work of Hušková & Meintanis (2007, 2009, 2010), we define the test statistic

$$T_{1n} \equiv T_{1n}(w) = \sum_{j=1}^{k} \frac{n_j}{n} \|\hat{\varphi}_j(t) - \hat{\varphi}_{0j}(t)\|_w^2 ,$$

where w is any given non-negative weight function with finite integral, $\int w(t)dt < \infty$. The presence of the weight function w in the integrals appearing in the expression of T_{1n} is necessary in order to ensure their finiteness, because $\|\hat{\varphi}_j(t) - \hat{\varphi}_{0j}(t)\|_w^2 \le 4\int w(t)dt$, for all j.

The motivation behind the test statistic T_{1n} is the following: T_{1n} converges in probability to (theorem 3)

$$T_1 \equiv T_1(w) = \sum_{j=1}^{k} p_j \|\varphi_j(t) - \varphi_{0j}(t)\|_w^2 . \tag{3}$$

Under $H_0, \varphi_j(t) = \varphi_{0j}(t)$ for all t and for $1 \le j \le k$, and thus T_1 vanishes. As a consequence, under H_0, T_{1n} should be 'very small'. We then conclude that any value of T_{1n}, which is 'significantly large', should lead to the rejection of H_0. In practice, given a significance level, a threshold value above which H_0 is rejected needs to be established. To this end, we need to study the null distribution of T_{1n}. Because this distribution is unknown, as an approximation to it, we derive the asymptotic null distribution. This will be carried out in the next section.

Remark 2.1. From lemma 1 in Alba-Fernández *et al.* (2008), an alternative expression for T_{1n}, which is useful from a computational point of view, is given by

$$nT_{1n} = \sum_{j=1}^{k} \frac{1}{n_j} \left\{ \sum_{l,s=1}^{n_j} I_w\left(\hat{\varepsilon}_{jl} - \hat{\varepsilon}_{js}\right) + \sum_{l,s=1}^{n_j} I_w\left(\hat{\varepsilon}_{0jl} - \hat{\varepsilon}_{0js}\right) - 2\sum_{l,s=1}^{n_j} I_w\left(\hat{\varepsilon}_{jl} - \hat{\varepsilon}_{0js}\right) \right\},$$

where $I_w(t) = \int \cos(tx)w(x)dx$. If w is a PDF with CF φ_w then $I_w(t) = Re\{\varphi_w(t)\}$, which clearly coincides with φ_w when w is a symmetric PDF.

3. Asymptotics

In order to study the limit behaviour of the test statistic T_{1n}, we first need to introduce some assumptions on the models (1) and on the available data. Recall that we are assuming that $\{(X_{jl}, Y_{jl}),\ 1 \le l \le n_j\}$ are independent and identically distributed observations from (X_j, Y_j) and the sets $\{(X_{1l}, Y_{1l}),\ 1 \le l \le n_1\}, \ldots, \{(X_{jk}, Y_{jk}),\ 1 \le l \le n_k\}$ are independent.

Assumption (A):

- **(A.1)** For $1 \le j \le k$: (i) X_j has a compact support R. (ii) f_j, m_j and σ_j are two times continuously differentiable on R. (iii) $\inf_{x \in R} f_j(x) > 0$ and $\inf_{x \in R} \sigma_j(x) > 0$.
- **(A.2)** For $1 \le j \le k$: the sample sizes satisfy $\lim n_j/n = p_j$, where $0 < p_j < 1$.
- **(A.3)** K is a twice continuously differentiable symmetric PDF with compact support.
- **(A.4)** The weight function satisfies $w(t) \ge 0$, for all $t \in \mathbb{R}$, and $\int t^4 w(t)dt < \infty$.
- **(A.5)** $nh_n^4 \to 0$ and $nh_n^2/\ln n \to \infty$.

These assumptions are mainly needed to guarantee the uniform consistency of the kernel estimators $\hat{f}_j, \hat{\sigma}_j, \hat{m}_j$ and $\hat{m}_0$. Unlike the methods based on the ECDF, observe that we do not impose any restriction on the distribution of the errors, like the existence of a PDF. So the results in this paper could be used to compare two or more regression functions when the distributions of the errors are arbitrary: continuous, discrete or mixed. An example with errors having a mixed distribution is given in Section 4.

3.1. Asymptotic null distribution

The following theorem gives an asymptotic approximation for $\sqrt{n_j}\{\hat{\varphi}_j(t) - \hat{\varphi}_{0j}(t)\}, 1 \le j \le n_j$, which will allow us to derive the asymptotic null distribution of the test statistic T_{1n}, given in the subsequent corollary. Let $\Sigma = (\sigma_{jv})_{1 \le j,v \le k}$ be the matrix whose elements are

$$\begin{aligned} \sigma_{jj} &= 1 - 2p_j E\left\{\frac{f_j(X_j)}{f_{mix}(X_j)}\right\} + p_j \sum_{r=1}^{k} p_r E\left\{\frac{\sigma_r^2(X_r)}{\sigma_j^2(X_r)} \frac{f_j^2(X_r)}{f_{mix}^2(X_r)}\right\}, \\ \sigma_{jv} &= \sqrt{p_j p_v} \sum_{r=1}^{k} p_r E\left\{\frac{\sigma_r^2(X_r)}{\sigma_j(X_r)\sigma_v(X_r)} \frac{f_j(X_r) f_v(X_r)}{f_{mix}^2(X_r)}\right\} \\ &\quad - \sqrt{p_j p_v} E\left\{\frac{\sigma_v(X_v)}{\sigma_j(X_v)} \frac{f_j(X_v)}{f_{mix}(X_v)} + \frac{\sigma_j(X_j)}{\sigma_v(X_j)} \frac{f_v(X_j)}{f_{mix}(X_j)}\right\}, \quad j \ne v. \end{aligned} \tag{4}$$

Theorem 1. *Under assumptions (A.1)–(A.3) and (A.5), if H_0 is true, then*

$$\sqrt{n_j}\{\hat{\varphi}_j(t) - \hat{\varphi}_{0j}(t)\} = \mathrm{i}t\varphi_j(t) Z_j + tR_{1j}(t) + t^2 R_{2j}(t),$$

where $\sup_t |R_{sj}(t)| = o_P(1), s = 1,2$ *and* $Z := (Z_1, \ldots, Z_k)' \sim N_k(0, \Sigma)$.

Define the diagonal matrix $\mathcal{A} = diag(a_1, \ldots, a_k)$, where $a_j = \|t\varphi_j(t)\|_w^2, 1 \le j \le k$. The results in the theorems and the corollaries in the succeeding text will hold whenever trace$(\mathcal{A}\Sigma) > 0$. Before stating the results, we briefly discuss this condition. Observe that

$$\text{trace}(\mathcal{A}\Sigma) = \sum_{j=1}^{k} a_j \sigma_{jj} > 0 \text{ if and only if } a_j > 0 \text{ and } \sigma_{jj} > 0 \text{ for some } j, 1 \le j \le k.$$

The quantities σ_{jj} in (4) can be also expressed as

$$\sigma_{jj} = p_j \sum_{l=1}^{k} p_l E\left[\frac{\sigma_l^2(X_l)}{\sigma_j^2(X_l)} \left\{\frac{f_j(X_l)}{f_{mix}(X_l)} - \frac{I(l=j)}{p_l}\right\}^2\right],$$

where $I(A)$ denotes the indicator function of the set A. From assumptions (A.1)(iii) and (A.2), it follows that $\sigma_{jj} > 0$ for all j. Thus, to ensure trace$(\mathcal{A}\Sigma) > 0$, we only need to ensure that $a_j > 0$ for some j. An easy way to obtain $a_j > 0$ is by taking $w(t) > 0$, for t in a neighbourhood of the origin.

The following assumption will appear in the statement of some of the results in the succeeding text.

Assumption (B): $a_j > 0$ for some $1 \le j \le k$.

Corollary 1. *Under assumptions (A) and (B), if H_0 is true, then $nT_{1n} \xrightarrow{\mathcal{L}} W_1 = Z'\mathcal{A}Z$, where Z is as in theorem 1.*

In other words, the limiting distribution of nT_{1n} under H_0 is a finite linear combination of independent chi-squared variables, $\sum_{j=1}^{k} \beta_j \chi_{1,j}^2$, where $\chi_{1,1}^2, \ldots, \chi_{1,k}^2$ are independent chi-squared random variates with one degree of freedom and $\beta_1, \ldots, \beta_k$ are the eigenvalues of

$\mathcal{A}\Sigma$. Unfortunately, the quantities β_j in this linear combination are unknown. They depend on the distribution of the errors through the $a_1, \ldots, a_k$ and on the distribution of the covariates through $\Sigma = (\sigma_{jv})_{1 \le j,v \le k}$. They also depend on the unknown design densities, f_j, and the conditional variance functions, σ_j^2. So to use theorem 1 in practice, one first needs to find a consistent estimator, say $\hat{\beta}_j$, for every β_j, $1 \le j \le k$. This can be easily carried out via plug-in method using the estimators defined earlier instead of the unknown functions φ_j, f_j, f_{mix} and σ_j^2. In order to perform the test, we also need to approximate the distribution of $\sum_{j=1}^k \hat{\beta}_j \chi^2_{1,j}$, which can be carried out via Monte Carlo method or some numerical method (see, for example, Kotz *et al.*, 1967, Castaño-Martínez & López-Blázquez, 2005). With such a distribution, we can finally obtain the critical value and/or the p-value for the test based on T_{1n}. The next result states the validity of this procedure.

To estimate a_j, we replace $\varphi_j(t)$ by $\hat{\varphi}_j(t)$ in its expression obtaining

$$\begin{aligned}\tilde{a}_j &= \int t^2 |\hat{\varphi}_j(t)|^2 w(t)dt = \frac{1}{n_j^2} \sum_{r,s=1}^{n_j} \int t^2 \cos\left\{t\left(\hat{\varepsilon}_{jr} - \hat{\varepsilon}_{js}\right)\right\} w(t)dt \\ &= \frac{n_j - 1}{n_j}\hat{a}_j + \frac{1}{n_j}\int t^2 w(t)dt,\end{aligned}$$

where

$$\hat{a}_j = \frac{-1}{\binom{n_j}{2}} \sum_{1 \le r < s \le n_j} D_2 I_w\left(\hat{\varepsilon}_{jr} - \hat{\varepsilon}_{js}\right), \quad 1 \le j \le k, \tag{5}$$

$D_2 I_w(t) = \frac{\partial^2}{\partial t^2} I_w(t) = -\int t^2 \cos(tu) w(u)du$ and I_w is as defined in Remark 2.1. Because $\int t^2 w(t)dt$ is a constant term, we estimate a_j by $\hat{a}_j$, which resembles a U-statistic. Let $\hat{\mathcal{A}} = diag\,(\hat{a}_1, \ldots, \hat{a}_k)$ and $\hat{\Sigma} = \left(\hat{\sigma}_{jv}\right)_{1 \le j,v \le k}$, with $\hat{\sigma}_{jj} = 1 - 2\hat{p}_j\hat{\mu}_j + \hat{p}_j \sum_{r=1}^k \hat{p}_r \hat{\mu}_{jjr}$, $\hat{\sigma}_{jv} = \sqrt{\hat{p}_j \hat{p}_v} \sum_{r=1}^k \hat{p}_r \hat{\mu}_{jvr} - \sqrt{\hat{p}_j \hat{p}_v}\left(\hat{\mu}_{jv} + \hat{\mu}_{vj}\right)$, $j \ne v$,

$$\begin{aligned}\hat{p}_j &= \frac{n_j}{n}, \quad \hat{\mu}_j = \frac{1}{n_j}\sum_{l=1}^{n_j} \frac{\hat{f}_j(X_{jl})}{\hat{f}_{mix}(X_{jl})}, \quad \hat{\mu}_{jv} = \frac{1}{n_v}\sum_{l=1}^{n_v} \frac{\hat{\sigma}_v(X_{vl})}{\hat{\sigma}_j(X_{vl})} \frac{\hat{f}_j(X_{vl})}{\hat{f}_{mix}(X_{vl})}, \\ \hat{\mu}_{jvr} &= \frac{1}{n_r}\sum_{l=1}^{n_r} \frac{\hat{\sigma}_r^2(X_{rl})}{\hat{\sigma}_j(X_{rl})\hat{\sigma}_v(X_{rl})} \frac{\hat{f}_j(X_{rl})\hat{f}_v(X_{rl})}{\hat{f}^2_{mix}(X_{rl})},\end{aligned}$$

$1 \le j, v, r \le k$.

Let $W_{1n} = \sum_{j=1}^k \hat{\beta}_j \chi^2_{1j}$, where $\chi^2_{11}, \ldots, \chi^2_{1k}$ are independent chi-squared variables with one degree of freedom and $\hat{\beta}_1, \ldots, \hat{\beta}_k$ are the eigenvalues of $\hat{\mathcal{A}}\hat{\Sigma}$.

Theorem 2. *Under assumptions (A) and (B),*

$$\sup_x |P_0\{nT_{1n} \le x\} - P_*(W_{1n} \le x)| \xrightarrow{P} 0.$$

Remark 3.1. If all the covariates have the same distribution, $f_1 = \ldots = f_k$, and all variance functions are equal, $\sigma_1 = \ldots = \sigma_k$, then

$$\Sigma = I_k - pp', \quad p' = \left(\sqrt{p_1}, \ldots, \sqrt{p_k}\right). \tag{6}$$

In this case, it is easy to see that Σ has two different eigenvalues: 0, with multiplicity 1, and 1, with multiplicity $k-1$. Therefore, if it is also assumed that the laws of the errors are such that

$a = a_1 = \ldots = a_k$ (for instance, if they also have the same distribution), then $a^{-1} n T_{1n}(w) \xrightarrow{\mathcal{L}} \sum_{j=1}^{k-1} \chi^2_{1j} = \chi^2_{k-1}$, which coincides with the null distribution of the classical ANOVA test for comparing means. To obtain a consistent null distribution estimator of $nT_{1n}(w)$ in this case, it suffices to have a consistent estimator of a.

Corollary 2. *Suppose that assumptions (A) and (B) hold. If all covariates have the same distribution, all variance functions are equal and the laws of the errors are such that $a = a_1 = \ldots = a_k$, then*

$$\sup_x |P_0\{nT_{1n} \leq x\} - P_*(W_{01n} \leq x)| \xrightarrow{P} 0,$$

where $W_{01n} = \hat{a}\chi^2_{k-1}$, with $\hat{a} = \sum_{j=1}^{k} \hat{p}_j \hat{a}_j$, and $\hat{a}_j$ is as defined by (5).

The result in corollary 1 tells us that $nT_{1n} = O_P(1)$. As a decision rule for testing H_0 against H_1, we propose to use $\Psi_{1,\alpha} = I(nT_{1n} > t_{1,\alpha})$, where $t_{1,\alpha}$ is the $1 - \alpha$ percentile of the null distribution of nT_{1n} or any consistent estimator of it.

3.2. Consistency

In this section, we study the test statistic T_{1n} when the alternative hypothesis is fixed. The following theorem shows that with probability tending to 1, T_{1n} behaves (asymptotically) like T_1, see (3). This will allow us to derive the consistency of the test $\Psi_{1,\alpha}$.

Theorem 3. *Suppose that assumption (A) holds. Then, $T_{1n} = T_1 + o_p(1)$, where T_1 is as defined in (3).*

As an immediate consequence of the above theorem and corollary 1, we conclude that the test $\Psi_{1,\alpha}$ is consistent against all fixed alternatives. This property is formally stated in the following corollary.

Corollary 3. *Suppose that assumption (A) and the alternative hypothesis H_1 hold. If $w(t) > 0$, for all $t \in \mathbb{R}$, then $\lim_{n\to\infty} P(\Psi_{1,\alpha} = 1) = 1$, for any $0 < \alpha < 1$.*

It is known that two distinct CFs can be equal in a finite interval (see, for example, Feller, 1971, p. 479). In order to ensure that $T_1 > 0$ whenever $m_r \neq m_s$, for some $1 \leq r, s \leq k, r \neq s$, we made the assumption that $w > 0$. It is important to note that this assumption does not involve any characteristic of the underlying data-generating procedure. For instance, taking w to be the PDF of, for example, a normal law guarantees the universal consistency of our test. In opposition, some existing works made rather restrictive assumptions that exclude certain type of alternatives. For example, the test of Srihera & Stute (2010) may not be able to detect a difference between two crossing curves.

3.3. Local alternatives

In this section, we study the limiting behaviour of the test statistic under local alternatives converging to the null hypothesis at the rate $n^{-1/2}$. Specifically, let us consider the following local alternative hypothesis

$$H_{1,n} : m_j = m_{00} + n^{-1/2} r_j, \quad 1 \leq j \leq k,$$

where m_{00} is assumed to be two times continuously differentiable and the functions r_j satisfy

$$E\left\{r_j^2(X_l)\right\} < \infty, \quad 1 \le j, l \le k. \tag{7}$$

Theorem 4. *Under assumption (A) and the alternative hypothesis $H_{1,n}$, if (7) holds, then*

$$\sqrt{n_j}\,\{\hat{\varphi}_j(t) - \hat{\varphi}_{0j}(t)\} = \mathrm{i}t\varphi_j(t)\left(Z_j + \sqrt{p_j}\mu_j\right) + R_j(t),$$

where $\|R_j\|_w = o_p(1)$, $Z = (Z_1, \ldots, Z_k)'$ is as in theorem 1 and

$$\mu_j = \sum_{v=1}^{k} p_v E\left\{\frac{f_v(X_j) r_v(X_j)}{f_{mix}(X_j)\sigma_j(X_j)}\right\} - E\left\{\frac{r_j(X_j)}{\sigma_j(X_j)}\right\}, \quad 1 \le j \le k.$$

Corollary 4. *Under assumption (A) and the alternative hypothesis $H_{1,n}$, if (7) holds, then $nT_{1n} \xrightarrow{\mathcal{L}} (Z+\mu)'\mathcal{A}(Z+\mu)$, where Z is as defined in theorem 1 and $\mu' = \left(\sqrt{p_1}\mu_1, \ldots, \sqrt{p_k}\mu_k\right)$.*

We conclude that although the test based on the rule $\Psi_{1,\alpha}$ is fully non-parametric, it is able to detect local alternatives converging to the null hypothesis at the rate $n^{-1/2}$ whenever $\mu'\mathcal{A} \ne 0$.

3.4. *A second test statistic*

The paper by Pardo-Fernández *et al.* (2007) studies two Kolmogorov–Smirnov and two Cramér–von Mises type statistics for testing H_0 based on the ECDF of the residuals. Our test statistic T_{1n} can be seen as the CF analogue of their first Cramér–von Mises type statistic. An ECF version of their second Cramér–von Mises type statistic is

$$T_{2n} = \|\hat{\varphi}(t) - \hat{\varphi}_0(t)\|_w^2,$$

where $\hat{\varphi}(t) = \sum_{j=1}^k \frac{n_j}{n}\hat{\varphi}_j(t)$ and $\hat{\varphi}_0(t) = \sum_{j=1}^k \frac{n_j}{n}\hat{\varphi}_{0j}(t)$, which are consistent estimators of $\varphi(t) = \sum_{j=1}^k p_j\varphi_j(t)$ and $\varphi_0(t) = \sum_{j=1}^k p_j\varphi_{0j}(t)$, respectively. The motivation of this statistic is that the equality of $\varphi(t)$ and $\varphi_0(t)$ also characterizes the null hypothesis.

The same steps followed in the analysis of T_{1n} can be used to study T_{2n}. In particular, T_{2n} can be computed as (Remark 2.1)

$$n^2 T_{2n} = \sum_{j,v=1}^{k}\sum_{l=1}^{n_j}\sum_{s=1}^{n_v}\left\{I_w\left(\hat{\varepsilon}_{jl} - \hat{\varepsilon}_{vs}\right) + I_w\left(\hat{\varepsilon}_{0jl} - \hat{\varepsilon}_{0vs}\right) - 2I_w\left(\hat{\varepsilon}_{jl} - \hat{\varepsilon}_{0vs}\right)\right\}.$$

The asymptotic null distribution of T_{2n} is given in the following result, which is analogous to corollary 1.

Corollary 5. *Let $\mathcal{B} = diag(p)\mathcal{C}diag(p)$, where p is as defined in (6) and $\mathcal{C} = (c_{jv})_{1\le j,v\le k}$ is the matrix with elements*

$$c_{jv} = \int t^2 Re\left\{\varphi_j(t)\overline{\varphi_v(t)}\right\} w(t)dt, \qquad 1 \le j, v, \le k.$$

Under assumption (A), if H_0 is true and trace$(\mathcal{B}\Sigma) > 0$, then $nT_{2n} \xrightarrow{\mathcal{L}} W_2 = Z'\mathcal{B}Z$, where Z is as in theorem 1.

In contrast to the case of T_{1n}, there is no easy way of ensuring that trace($\mathcal{B}\Sigma$) > 0. To see this fact, consider, for example, the case with $f_1 = \ldots = f_k$ and $\sigma_1 = \ldots = \sigma_k$. In this situation, we saw that Σ has the expression (6); if in addition the errors are such that $c = c_{jv}, 1 \le j, v \le k$, then trace($\mathcal{B}\Sigma$) $= 0$, and thus, the distribution of nT_{2n} is degenerate for any choice of the weight function w.

The asymptotic distribution of T_{2n} under H_0 depends on certain properties of the populations, which are typically unknown, and it can be summarized as

$$nT_{2n} = \begin{cases} O_p(1) & \text{if} \quad \mathrm{trace}(\mathcal{B}\Sigma) > 0, \\ o_p(1) & \text{if} \quad \mathrm{trace}(\mathcal{B}\Sigma) = 0. \end{cases}$$

In the first case (trace($\mathcal{B}\Sigma$) > 0), the asymptotic null distribution of T_{2n} is analogous to the distribution of T_{1n}, that is, a combination of chi-squared random variables multiplied by the eigenvalues of $\mathcal{B}\Sigma$, which can be estimated as in theorem 2. In the second case (trace($\mathcal{B}\Sigma$) $= 0$), a deeper analysis of the asymptotic distribution is required. However, from a practical point of view, this analysis is somehow useless because the practitioner would not know which one of the two situations apply for a given data set. Because of these reasons, we have focused on the test statistic T_{1n}.

4. Numerical results

In this section, we report the results of an experiment carried out to study the practical behaviour of the proposed testing procedure by means of simulations. We investigate the approximation given in theorem 2 and also the bootstrap approximation used in Pardo-Fernández *et al.* (2007) in order to compare their tests with ours. In all cases, the tables display the observed proportion of rejections in 1000 simulated data sets.

Firstly, in a two-population ($k = 2$) framework, the following regression models are considered:

(i) $m_1(x) = m_2(x) = 1$;
(ii) $m_1(x) = m_2(x) = x$;
(iii) $m_1(x) = m_2(x) = \sin(2\pi x)$;
(iv) $m_1(x) = m_2(x) = \exp(x)$;
(v) $m_1(x) = x, m_2(x) = 1 + x$;
(vi) $m_1(x) = \exp(x), m_2(x) = \exp(x) + x$;
(vii) $m_1(x) = \sin(2\pi x), m_2(x) = \sin(2\pi x) + x$;
(viii) $m_1(x) = 1, m_2(x) = 1 + \sin(2\pi x)$.

Models (i)–(iv) are under the null hypothesis, and models (v)–(viii) are under the alternative. For the scale functions, in each case, we study a homoscedastic and a heteroscedastic scenarios:

Homoscedastic models (S1): $\sigma_1(x) = 0.50; \sigma_2(x) = \sqrt{0.50}$.
Heteroscedastic models (S2): $\sigma_1(x) = \frac{7}{6}0.50x + \frac{1}{2}0.50; \sigma_2(x) = \frac{7}{8}\sqrt{0.50}x + \frac{1}{2}\sqrt{0.50}$.

The covariates X_1 and X_2 have distributions $Beta(1.5, 2)$ and $Beta(2, 1.5)$, respectively. This choice of the distributions of the covariates motivates the models of the scale functions in the heteroscedastic case, as they verify that $E[\sigma_1(X_1)] = 0.50$ and $E[\sigma_2(X_2)] = \sqrt{0.50}$, so the homoscedastic case and the heteroscedastic case are somehow comparable. If not mentioned otherwise, the regression errors ε_1 and ε_2 are $N(0, 1)$, although other distributions will also be considered in Section 4.3.

Non-parametric estimation of the regression functions is performed by the local-linear estimator. For the estimation of the variance functions, we prefer the local-constant estimator (Nadaraya–Watson), because the local-linear may produce negative values. In both cases, the kernel function is the kernel of Epanechnikov $K(u) = 0.75(1 - u^2)I(-1 < u < 1)$, which have some optimal properties. In the next section, we will discuss the choice of the smoothing parameter.

4.1. The choice of the smoothing parameter

The choice of the smoothing parameter or bandwidth is certainly a delicate issue in any non-parametric procedure. For estimation purposes, it is well known that the bandwidth controls the trade-off between bias and variance of the estimator. In the context of testing, this problem has not been studied in detail yet. González-Manteiga & Crujeiras (2013) gave a very recent review about goodness-of-fit problems in non-parametric regression, including the comparison of regression curves. In the discussion of the paper, the authors say that the bandwidth selection for tests based on smoothing is a 'really tough problem' and 'it is far from being solved'. This conclusion was also raised by several discussants of the paper (see, for example, the discussions of Sperlich (2013) and de Uña-Álvarez (2013) to the aforementioned article). We also share that opinion.

Although a detailed study on this topic is still missing, in the context of comparing regression curves several practical proposals have been made, sometimes explicitly, sometimes implicitly. We will review here three of relatively recent and relevant papers on the topic. (i) In Neumeyer & Dette (2003), the proposed methodology allows for the use of the optimal bandwidth for estimation (of order $n^{-1/5}$) and bandwidths based on the classical rule of the thumb are employed. (ii) In Pardo-Fernández *et al.* (2007), the theory does not allow for the use of the optimal bandwidth in estimation (which is also the case in the current piece of research), but some practical recommendations are suggested: first, when estimating the regression curves to be compared, it is recommended to use a common bandwidth; second, in practical applications, the test can be performed for a reasonable range of bandwidths, and the obtained p-values can be analysed. (iii) In Srihera & Stute (2010), nearest-neighbour estimators are used, and the optimal smoothing parameters for estimation are also excluded by the theory; in simulations, fixed values are used.

In a recent paper about testing for the distribution of the regression error, Heuchenne & Van Keilegom (2010) suggested several possibilities for the choice of the smoothing parameter based on cross-validation techniques. We have checked the practical performance of one of their proposal in our context (the one referred as method f, which is the one recommended by the authors). A summary of the obtained results is given in Table 1, which shows the approximation of the level of the test based on T_{1n} for models (i)–(iv) with sample sizes $n_1 = n_2 = 100$. The weight function, w, is the PDF of a standard normal (see the discussion about the role of the weight function in the next section). The critical values are obtained from the approximation of the asymptotic distribution given in theorem 2. The level approximation is good for model (iv) but not correct for models (i) and (iii). Besides, we have observed that the proposed cross-validation procedure tends to pick very large bandwidths (often, the largest value in the allowed interval). Although the use of cross-validation bandwidths might be a reasonable choice in some cases, we are not sure that they offer a global solution to the problem because their practical performance is not always satisfactory.

In the simulations contained in the rest of this section, we essay to study the general behaviour of the proposed test, and therefore, we prefer to consider non-data dependent bandwidths. We take a bandwidth depending on the sample size of the form $h_n = Cn^a$. According

Table 1. *Empirical level of the test based on the asymptotic distribution of T_{1n} for homoscedastic and heteroscedastic models and smoothing parameters chosen by cross-validation with $n_1 = n_2 = 100$*

		Homoscedastic models			Heteroscedastic models		
Model	α :	0.100	0.050	0.010	0.100	0.050	0.010
(i)		0.154	0.064	0.018	0.148	0.062	0.012
(ii)		0.136	0.052	0.012	0.124	0.052	0.010
(iii)		0.152	0.088	0.016	0.152	0.086	0.012
(iv)		0.106	0.048	0.010	0.106	0.046	0.010

to assumption (A.5), the allowed values for a are $-0.5 < a < -0.25$. We choose to take the exponent in the middle of the interval of allowed values, $a = -0.375$, and then several values of the constant C are taken into account.

4.2. The weight function

In the present section, we discuss the role of the weight function w. For our test to be consistent, it is only needed a positive w satisfying $\int t^4 w(t)dt < \infty$. However, from the results in Section 3, it is clear that the choice of the weight function affects the power. In fact, from corollary 4, the asymptotic power of the test $\Psi_{1,\alpha}$ is given by

$$P\left((Z+\mu)'\mathcal{A}(Z+\mu) > t_{1,\alpha}\right). \tag{8}$$

This probability depends on w through $\mathcal{A} = diag(a_1, \ldots, a_k)$, where $a_j = \int t^2 |\varphi_j(t)|^2 w(t)dt, 1 \le j \le k$. Clearly, the optimal weight (the one that maximizes the power) depends on the CFs $\varphi_1, \ldots, \varphi_k$ and on $\mu_1, \ldots, \mu_k$. Because all these quantities are unknown and involve the unobservable residuals, formula (8) is of little help in practice.

An alternative approach is to use different weight functions $w_1, \ldots, w_k$ for the different populations. This leads to the following test statistic:

$$\sum_{j=1}^{k} \frac{n_j}{n} \|\hat{\varphi}_j(t) - \hat{\varphi}_{0j}(t)\|^2_{w_j}.$$

By theorem 1, this quantity has the same asymptotic properties as T_{1n}. In fact, one only needs to replace $a_j = \|t\varphi_j(t)\|^2_w$ by $\bar{a}_j = \|t\varphi_j(t)\|^2_{w_j}$ for the results given in Section 3 to continue to hold. Following the guidelines in Epps & Pulley (1983) (see also Epps, 2005, Jiménez-Gamero *et al.*, 2009, and Hušková & Meintanis, 2010), a reasonable choice for $w_j(t)$ is $|\varphi_j(t)|^2$. This choice attempts to give high weight where the statistic $\hat{\varphi}_j(t)$ is a relatively precise estimator of $\varphi_j(t)$. The problem is that $\varphi_j(t)$ is unknown and needs to be estimated. Taking $w_j(t) \propto |\hat{\varphi}_j(t)|^2$ is not possible because $\int |\hat{\varphi}_j(t)|^2\, dt = \infty$, since $\hat{\varphi}_j(t)$ is a periodic function. To overcome this difficulty, we could consider a kernel smoothing estimator, but its application requires to assume rather strong assumptions on the distribution of the errors.

A more practical method avoiding the aforementioned difficulties is to use a parametric density function as a weight. The density function should put most of the weight near the origin, because the ECF estimates more accurately the population CF around $t = 0$. This approach is connected with the comparison of kernel density estimators as follows. Let S be a PDF symmetric around the origin, and let w be such that $w^{1/2}(x) = (2\pi)^{-1/2} \int \exp(itx)S(t)dt$. Lemma 2.1 in Fan (1998) (see also Anderson *et al.*, 1994, Henze *et al.*, 2005, Hušková & Meintanis, 2012, and Meintanis, 2013) shows that

$$\int |\hat{\varphi}_j(t) - \hat{\varphi}_{0j}(t)|^2 w_j(t)dt = \int \left\{\hat{f}_j(t) - \hat{f}_{0j}(t)\right\}^2 dt,$$

where

$$\hat{f}_j(t) = \frac{1}{n}\sum_{l=1}^{n_j} S\left(t - \hat{\varepsilon}_{jl}\right), \quad \hat{f}_{0j}(t) = \frac{1}{n}\sum_{l=1}^{n_j} S\left(t - \hat{\varepsilon}_{0jl}\right),$$

that is to say, for adequate choices of the weight function w_j, the quantity $\|\hat{\varphi}_j - \hat{\varphi}_{0j}\|_w^2$ coincides with the integral of the squared of the difference between kernel estimators of the PDF of the errors, both estimated with bandwidth $\eta = 1$. This bandwidth can be taken arbitrarily by considering $S_\eta(x) = \eta^{-1}S(x/\eta)$ instead of $S = S_1$.

The aforementioned observation does not narrow down the spectrum of possibilities. So further criteria must be taken into account. From a practical point of view, the ease of computation of T_{1n} (also T_{2n}) is closely related to the choice of w. This is specially appealing if instead of using the approximation given in theorem 2, one wishes to employ a bootstrap approximation, which requires to evaluate the test statistic in a high number of artificial samples. In this sense, a good choice is $w(t) \propto \exp\left\{-t^2/2\sigma_w^2\right\}$, which is tantamount to estimate the PDF of the errors by using the normal kernel and bandwidth σ_w. Nevertheless, at this point, we must say that the choice of σ_w cannot be guided by standard results on kernel density estimation, because in such a case the asymptotic results in this paper are no longer true. A theoretical study of the optimal choice of σ_w in terms of Bahadur slopes can be found in Tenreiro (2009) for the problem of testing goodness-of-fit for the normal distribution. The results in the cited paper show that the optimal choice of σ_w depends on the alternative, which is unknown in practice. Because of this reason, we did not pursue this line. Neither too large, nor too small, values of σ_w are appropriate because, proceeding as in Henze *et al.* (2005), we obtain

$$\lim_{\sigma_w \to 0} \sigma_w^{-2} T_{1n} = \sum_{j=1}^{k} \frac{n_j}{n}\left(\bar{\hat{\varepsilon}}_j - \bar{\hat{\varepsilon}}_{0j}\right)^2, \tag{9}$$

with $\bar{\hat{\varepsilon}}_j = \frac{1}{n_j}\sum_{j=1}^{n_j}\hat{\varepsilon}_{jl}, \bar{\hat{\varepsilon}}_{0j} = \frac{1}{n_j}\sum_{j=1}^{n_j}\hat{\varepsilon}_{0jl}, 1 \le j \le k$. The right-side of (9) estimates $\theta = \sum_{j=1}^{k} p_j\{E(\varepsilon_j) - E(\varepsilon_{0j})\}^2$. In certain situations, θ might be zero, even when H_0 is not true, and therefore, the resulting test based on small σ_w would not be consistent against all fixed alternatives. A similar phenomenon occurs when $\sigma_w \to \infty$.

In the context of testing goodness of fit for the distribution of the errors in non-parametric regression models, the simulations in Hušková & Meintanis (2010) reveal that taking $\sigma_w = 1$ gives good results. We have also investigated numerically the effect of changing the parameter σ_w in models (iv) (level approximation) and (vi) (power) when the critical values are approximated from the asymptotic distribution of T_{1n} as explained in theorem 2. Table 2 (which is just a part of a larger simulation study) summarizes the obtained results. It can be seen that the results are quite homogeneous with slightly better level with $\sigma_w = 1$, which is the value that we will consider in the rest of the simulations.

4.3. Results of the test based on the asymptotic null distribution

In this section, we present the results of the test based on the approximation of the asymptotic null distribution of T_{1n} as given in theorem 2. Table S1 (homoscedastic models; see Supporting information) and Table S2 (heteroscedastic models; see Supporting information) display the results for bandwidths of the form $h = Cn^{-0.375}$, with $C = 1, 1.5, 2$, which provide reasonable values for the considered setups. The level—models (i)–(iv)—is slightly overestimated

Table 2. *Observed rejection proportions of the test based on the asymptotic distribution of T_{1n} for homoscedastic models (iv) and (vi) and for different choices of the parameter σ_w with $n_1 = n_2 = 100$*

Model	σ_w	α: / C:	0.100 / 1.0	0.100 / 1.5	0.100 / 2.0	0.050 / 1.0	0.050 / 1.5	0.050 / 2.0	0.010 / 1.0	0.010 / 1.5	0.010 / 2.0
(iv)	0.50		0.100	0.082	0.078	0.046	0.041	0.039	0.012	0.009	0.007
	0.75		0.111	0.090	0.083	0.049	0.043	0.039	0.012	0.009	0.008
	1.00		0.122	0.097	0.084	0.052	0.046	0.041	0.012	0.009	0.008
	1.25		0.133	0.100	0.086	0.056	0.047	0.042	0.012	0.010	0.008
	1.50		0.138	0.109	0.093	0.065	0.051	0.044	0.013	0.010	0.008
(vi)	0.50		1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.997	0.996
	0.75		1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.997	0.997
	1.00		1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.997	0.997
	1.25		1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.997	0.997
	1.50		1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.997	0.997

Table 3. *Observed rejection proportions of the test based on the asymptotic distribution of T_{1n} for different error distribution with $n_1 = n_2 = 100$*

Error distribution	Model	α: / C:	0.100 / 1.0	0.100 / 1.5	0.100 / 2.0	0.050 / 1.0	0.050 / 1.5	0.050 / 2.0	0.010 / 1.0	0.010 / 1.5	0.010 / 2.0
Mixed	(ii)		0.104	0.094	0.093	0.051	0.050	0.047	0.010	0.007	0.008
	(iv)		0.105	0.090	0.083	0.049	0.046	0.044	0.011	0.008	0.007
	(vi)		0.996	1.000	0.999	0.992	0.998	0.998	0.985	0.993	0.990
	(viii)		0.989	0.997	0.994	0.987	0.992	0.985	0.950	0.917	0.836
Exponential	(ii)		0.149	0.125	0.115	0.076	0.063	0.055	0.015	0.014	0.010
	(iv)		0.146	0.115	0.105	0.077	0.061	0.046	0.015	0.012	0.008
	(vi)		0.997	0.999	0.999	0.997	0.999	0.998	0.992	0.990	0.988
	(viii)		0.998	1.000	0.995	0.995	0.990	0.981	0.948	0.904	0.812

The models are for heteroscedastic.

for small sample sizes, but the approximation improves as the sample sizes increase, reaching a good approximation for $(n_1, n_2) = (100, 100)$. The test also reaches good power, both in the homoscedastic case and in the heteroscedastic case.

We have also run simulations with error distributions different from the normal distribution. In particular, we have considered two cases: (a) errors with mixed distribution of the form

$$\varepsilon_j = \begin{cases} 0 & \text{with probability } 0.5, \\ N(0, \sqrt{2}) & \text{with probability } 0.5, \end{cases}$$

$j = 1, 2$, which in practice could model a case where the observations come from two devices, one of them with no measurement error; and (b) errors with a recentred exponential distribution, that is, $\varepsilon_j + 1 \sim Exponential(1)$, for $j = 1, 2$. Table 3 displays a brief summary of the obtained results for models (ii), (iv), (vi) and (viii) under heteroscedasticity and sample sizes $n_1 = n_2 = 100$. In the case of the mixed distribution, the approximation of the level is very good. In the case on the exponential distribution, we observe an overestimation of the level, probably caused by a bad approximation of the asymptotic null distribution due to the asymmetry of the error distribution (although not shown in the table, better approximations are achieved for larger sample sizes). Both cases show good power. The same kind of conclusions can be established for the rest of the models, which are not shown here.

4.4. Results based on a bootstrap approximation

The aim of this subsection is to compare the power of the test proposed in this paper with those in Pardo-Fernández *et al.* (2007). To approximate the null distribution of their test statistics, Pardo-Fernández *et al.* (2007) employed a bootstrap procedure based on smoothed residuals (see also Neumeyer, 2009, for a theoretical justification). Of course, the same bootstrap procedure could be used to approximate the null distribution of nT_{1n}. Nevertheless, from a computational point of view, the estimators in theorem 2 and corollary 2 are less time consuming. In order to establish a fair comparison, we have also estimated the null distribution of the test proposed in this paper by using the bootstrap algorithm defined in the aforementioned paper. Besides, we have also incorporated here the test statistic T_{2n}, for which the asymptotic null distribution is difficult to approximate. Table S3 (Supporting information) shows the results of the tests based on T_{1n} and T_{2n}, and the four tests proposed in Pardo-Fernández *et al.* (2007), which are denoted by $T^1_{KS}, T^2_{KS}, T^1_{CM}$ and T^2_{CM}. For the sake of brevity of the presentation of the table, we restrict ourselves to the significance level $\alpha = 0.05$ and bandwidth with $C = 1$ (similar results have been obtained for other significance levels and other specifications of the bandwidth). In terms of level approximation, we can see that it is good for all test statistics, except for T^2_{KS}. Compared with the asymptotic approximation, the bootstrap approximation improves the behaviour of the test statistic T_{1n} for small sample sizes. For models (v)–(vii), the highest power is achieved by the test based on the ECF, T_{1n}. For model (viii), the highest power is achieved by T_{2n}, which is also based on the ECF. Note that in this model, T_{1n} reaches reasonable power and it is much better than its ECDF-based analogue T^1_{CM}. Summarizing, for the models considered, the test based on T_{1n} presents, as a whole, the best behaviour.

4.5. The case of three populations

We have also investigated the test based on the estimated asymptotic null distribution of T_{1n} in the case of three populations ($k = 3$). Now the regression models are as follows:

(ix) $m_1(x) = m_2(x) = m_3(x) = 1$.
(x) $m_1(x) = m_2(x) = m_3(x) = x$.
(xi) $m_1(x) = x, m_2(x) = x + 0.2, m_3(x) = x + 0.4$.
(xii) $m_1(x) = x, m_2(x) = x, m_3(x) = x + 0.25$.
(xiii) $m_1(x) = 0.5, m_2(x) = x, m_3(x) = 1 - x$.
(xiv) $m_1(x) = 0, m_2(x) = \sin(2\pi x), m_3(x) = -\sin(2\pi x)$.

Models (ix)–(x) are under the null hypothesis, and models (xi)–(xiv) are under the alternative. We only consider homoscedastic models with scale functions $\sigma_1(x) = \sqrt{0.25}, \sigma_2(x) = \sqrt{0.25}$ and $\sigma_3(x) = \sqrt{0.50}$. The covariates X_1, X_2 and X_3 are $Beta(1.5, 2), Beta(2, 1.5)$ and $Beta(2, 2)$, respectively, and all regression errors are $N(0, 1)$. As in the previous cases, a bandwidth of the form $h = Cn^{-0.375}$ is chosen, but now the $C = 2, 2.5, 3$ are displayed. Other choices for C were also tried, but better results were obtained for these values. The results are shown in Table S4 (Supporting information). As in Tables S1 and S2, the level is well approximated for large sample sizes, and the behaviour in terms of power is correct.

5. Conclusions

A test for the comparison of k regression functions has been proposed and studied under a totally non-parametric setting. The test statistic compares the ECF of the residuals in each population with the ECF of the residuals under the null hypothesis. For adequate choices of the weight function involved in the definition of the test statistic, the resulting test is consistent against any fixed alternative and is able to detect contiguous alternatives converging to the null at a rate $n^{-1/2}$. To derive these properties, we have assumed certain assumptions, which are weaker than those required by those based on the ECDF. Specifically, no requirement is imposed on the distributions of the errors. An estimation of the asymptotic null distribution has been proposed as an estimator of the null distribution of the test statistic. In the cases tried in our numerical experiments, it is observed that this approximation works, in the sense of providing type I errors close to the nominal values, specially when the sample sizes are at least 100. For smaller sample sizes, it is recommended to approximate the null distribution through a bootstrap mechanism.

Acknowledgements

The authors thank the anonymous referees for their constructive comments and suggestions that helped to improve the presentation. J. C. Pardo-Fernández has been financially supported by grant MTM2011-23204 (FEDER support included) of the Spanish Ministry of Science and Innovation. M. D. Jiménez-Gamero has been financially supported by grant UJA2013/08/01 (University of Jaén and Caja Rural Provincial of Jaén). A. El Ghouch acknowledges financial support from IAP research network P6/03 of the Belgian Government (Belgian Science Policy) and from the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'.

References

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

Anderson, N. H., Hall, P. & Titterington, D. M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *J. Multivariate Anal.* **50**, 41–54.

Castaño-Martínez, A. & López-Blázquez, F. (2005). Distribution of a sum of weighted central chi-square variables. *Comm. Statist. Theory Methods* **34**, 515–524.

de Uña-Álvarez, J. (2013). Comments on: an updated review of goodness-of-fit tests for regression models. *TEST* **22**, 414–418.

Delgado, M. A. (1993). Testing the equality of nonparametric regression curves. *Statist. Probab. Lett.* **17**, 199–204.

Epps, T. W. (2005). Tests for location-scale families based on the empirical characteristic function. *Metrika* **62**, 99–114.

Epps, T. W. & Pulley, L. B. (1983). A test for normality based on the empirical characteristic function. *Biometrika* **70**, 723–726.

Fan, Y. (1998). Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. *Econom. Theory* **14**, 604–621.

Fan, J. & Gijbels. I. (1996). *Local polynomial modelling and its applications*, Chapman & Hall, London.

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

González-Manteiga, W. & Crujeiras, R. (2013). An updated review of goodness-of-fit tests for regression models. *TEST* **22**, 361–411.

Henze, N., Klar, B. & Zhu, L. X. (2005). Checking the adequacy of the multivariate semiparametric location shift model. *J. Multivariate Anal.* **93**, 238–265.

Heuchenne, C. & Van Keilegom, I. (2010). Goodness-of-fit tests for the error distribution in nonparametric regression. *Comput. Statist. Data Anal.* **54**, 1942–1951.

Hušková, M. & Meintanis, S. G. (2007). Omnibus tests for the error distribution in the linear regression model. *Statistics* **41**, 363–376.

Hušková, M. & Meintanis, S. G. (2009). Goodness-of-fit tests for parametric regression models based on empirical characteristic functions. *Kybernetika* **45**, 960–971.

Hušková, M. & Meintanis, S. G. (2010). Tests for the error distribution in nonparametric possibly heteroscedastic regression models. *TEST* **19**, 92–112.
Hušková, M. & Meintanis, S. G. (2012). Tests for symmetric error distribution in linear and nonparametric regression models. *Comm. Statist. Simulation Comput.* **41**, 833–851.
Jiménez-Gamero, M. D., Muñoz-García, J. & Pino-Mejías, R. (2005). Testing goodness of fit for the distribution of errors in multivariate linear models. *J. Multivariate Anal.* **95**, 301–322.
Jiménez-Gamero, M. D., Alba-Fernández, V., Muñoz-García, J. & Chalco-Cano, Y. (2009). Goodness-of-fit tests based on empirical characteristic functions. *Comput. Statist. Data Anal.* **53**, 3957–3971.
Kotz, S., Johnson, N. L. & Boyd, D. W. (1967). Series representations of quadratic forms in normal variables. I. Central case. *Ann. Math. Statist.* **38**, 823–837.
Kulasekera, K. B. (1995). Comparison of regression curves using quasi-residuals. *J. Amer. Statist. Assoc.* **90**, 1085–1093.
Meintanis, S. G. (2013). Comments on: an updated review of goodness-of-fit tests for regression models. *TEST* **22**, 432–436.
Munk, A. & Dette, H. (1998). Nonparametric comparison of several regression functions: exact and asymptotic theory. *Ann. Statist.* **26**, 2339–2368.
Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of non-parametric regression residuals. *Scand. J. Statist.* **36**, 204–228.
Neumeyer, N. & Dette, H. (2003). Nonparametric comparison of regression curves: an empirical process approach. *Ann. Statist.* **31**, 880–920.
Neumeyer, N. & Pardo-Fernández, J. C. (2009). A simple test for comparing regression curves versus one-sided alternatives. *J. Statist. Plann. Inference* **139**, 4006–4016.
Pardo-Fernández, J. C., Van Keilegom, I. & González-Manteiga, W. (2007). Testing for the equality of k regression curves. *Statist. Sinica* **17**, 1115–1137.
Sperlich, S. (2013). Comments on: an updated review of goodness-of-fit tests for regression models. *TEST* **22**, 419–427.
Srihera, R. & Stute, W. (2010). Nonparametric comparison of regression functions. *J. Multivariate Anal.* **101**, 2039–2059.
Tenreiro, C. (2009). On the choice of the smoothing parameter for the BHEP goodness-of-fit. *Comput. Statist. Data Anal.* **53**, 1038–1053.

Received April 2013, in final form April 2014

Juan Carlos Pardo-Fernández, Departamento de Estatística e I.O., Facultade de Ciencias Económicas e Empresariais, Campus Universitario As Lagoas-Marcosende, 36310 Vigo, Spain.
E-mail: juancp@uvigo.es

Supporting information

Additional Supporting information may be found in the online version of this article at the publisher's web site.