Universidade de Vigo

A nonparametric ANOVA-type test for regression curves based on characteristic functions

J.C. Pardo-Fernández, M.D. Jiménez-Gamero, A. El Ghouch

Report 12/07

Discussion Papers in Statistics and Operation Research

Departamento de Estatística e Investigación Operativa

Facultade de Ciencias Económicas e Empresariales
Lagoas-Marcosende, s/n · 36310 Vigo
Tfno.: +34 986 812440 - Fax: +34 986 812401
http://webs.uvigo.es/depc05/
E-mail: depc05@uvigo.es

Universidade de Vigo

A nonparametric ANOVA-type test for regression curves based on characteristic functions

J.C. Pardo-Fernández, M.D. Jiménez-Gamero, A. El Ghouch

Report 12/07

Discussion Papers in Statistics and Operation Research

Imprime: GAMESAL

Edita:

Universidade de Vigo
Facultade de CC. Económicas e Empresariales
Departamento de Estatística e Investigación Operativa
As Lagoas Marcosende, s/n 36310 Vigo
Tfno.: +34 986 812440

I.S.S.N: 1888-5756

Depósito Legal: VG 1402-2007

A nonparametric ANOVA-type test for regression curves based on characteristic functions

J.C. Pardo-Fernández[1] *, M.D. Jiménez-Gamero[2] †, A. El Ghouch[3] ‡

[1]Departamento de Estatística e I.O., Universidade de Vigo, Spain

[2]Departamento de Estadística e I.O., Universidad de Sevilla, Spain

[3]ISBA, Université catholique de Louvain, Belgium

Abstract

This article studies a new procedure to test for the equality of k regression curves in a fully nonparametric 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 combination of χ_1^2 random variables. Under certain restrictions on the populations, the asymptotic null distribution of the test statistic is χ_{k-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; regression residuals.

*__Corresponding author.__ 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. Telephone: +34 986 813505. Fax: +34 986 812 401.

†Departamento de Estadística e I.O., Facultad de Matemáticas, Calle Tarfia s.n., 41012 Sevilla, Spain. e-mail: dolores@us.es.

‡Institut de statistique, biostatistique et sciences actuarielles (ISBA), Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. e-mail: anouar.elghouch@uclouvain.be.

1 Introduction

Testing 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, it is also observed 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 nonparametric 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 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 nonparametric.

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. Since the objective is to compare the regression curves, it is reasonable to assume that the covariates have common compact 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 nonparametric settings has been previously treated in the statistical literature. The papers by Delgado (1993), Kulasekera (1995), Neumeyer and Dette (2003), Neumeyer and Pardo-Fernández (2009) and Srihera and Stute (2010), among others, are devoted to the comparison of two curves. Pardo-Fernández et al. (2007) deal with the comparison of $k\,(k \geq 2)$ regression curves. Their approach is based on comparing the distribution functions 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 characteristic function (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 characteristic function (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 and Meintanis (2007, 2010), for testing goodness of fit (gof) for the errors, and Hušková and Meintanis (2009) for testing gof 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 and Meintanis (2009), is that the former usually requires less stringent assumptions for its validity. In addition, from the simulation results for finite sample sizes in these and other related papers, the tests based on the ecf compete very satisfactorily with those based on the empirical cdf (ecdf).

Having in mind the reasons above, 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}, $1 \leq j \leq k$. With this aim, the paper is organized as follows. Section 2 introduces the test statistic, which is of a Cramér-von-Mises type, and provides also an alternative expression, which is useful from a computational point of view. Asymptotic properties are studied in Section 3. It is shown that, under certain weak conditions on the distributions of the errors and of the covariates, the null distribution is proportional to a χ^2_{k-1} distribution. When such conditions do not hold, the null distribution of the test statistic as well as its asymptotic null distributions are both unknown. In order to approximate the null distribution of the test statistic, consistent null distribution estimators are proposed. The behaviour of the test under fixed and local alternatives is also studied. Section 4 contains some numerical results based on simulations to study the practical performance of the test and to compare it with other existing methods. Section 5 concludes. All proofs of the theoretical results are deferred to the Appendix.

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, $1 \leq i \leq k$; for any complex number $z = a + ib$, $Re(z) = a$ is its real part, $\bar{z} = a - ib$ is its conjugate and $|z|$ is its modulus; $N_k(\mu, \Sigma)$ denotes de multivariate normal distribution with mean vector μ and variance-covariance matrix Σ. An unspecified integral denotes integration over the whole real line $\mathbb{R}$.

2 The test statistic

Let (X_{jl}, Y_{jl}), $1 \leq l \leq n_j$, be independent and indentically distributed (iid) 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, $\sigma_j^2(x) = E[\{Y_j - m_j(x)\}^2|X_j = x]$, 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 nonparametric estimators based on kernel smoothing techniques. Let K denote a nonnegative kernel function defined on $\mathbb{R}$ (normally, a symmetric pdf), 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:

$$\hat{m}_j(x) = \sum_{l=1}^{n_j} w_{jl}(x)Y_{jl}, \quad \hat{\sigma}_j^2(x) = \sum_{l=1}^{n_j} w_{jl}(x)Y_{jl}^2 - \hat{m}_j^2(x), \quad \hat{m}_0(x) = \sum_{j=1}^{k} \frac{n_j}{n} \frac{\hat{f}_j(x)}{\hat{f}_{mix}(x)} \hat{m}_j(x),$$

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, for example, Nadaraya-Watson weights

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

which are a particular case of local polynomial weighting (see Fan and Gijbles, 1996). Under the model assumptions that will be stated in the next section, the results in this article are valid for local constant (Nadaraya-Watson) and for local linear estimator.

Based on these estimators, for each population j, $1 \le j \le k$, we construct two samples of residuals:

$$\hat{\varepsilon}_{jl} = \frac{Y_{jl} - \hat{m}_j(X_{jl})}{\hat{\sigma}_j(X_{jl})} \qquad \text{and} \qquad \hat{\varepsilon}_{0jl} = \frac{Y_{jl} - \hat{m}_0(X_{jl})}{\hat{\sigma}_j(X_{jl})}, \tag{3}$$

$1 \le l \le n_j$, whose ecfs are

$$\hat{\varphi}_j(t) = \frac{1}{n_j}\sum_{l=1}^{n_j} \exp(\mathrm{i}t\hat{\varepsilon}_{jl}) \qquad \text{and} \qquad \hat{\varphi}_{0j}(t) = \frac{1}{n_j}\sum_{l=1}^{n_j} \exp(\mathrm{i}t\hat{\varepsilon}_{0jl}), \tag{4}$$

respectively. These ecfs are nothing but (consistent) kernel based nonparametric 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 $\hat{\varphi}_{0j}(t)$, $1 \le j \le k$, using a weighted L_2-distance. More precisely, following the work of Hušková and Meintanis (see Hušková and Meintanis, 2007, 2009, 2010) we define the test statistic

$$T_{1n} \equiv T_{1n}(w) = \sum_{j=1}^{k} \frac{n_j}{n} \int \left|\hat{\varphi}_j(t) - \hat{\varphi}_{0j}(t)\right|^2 w(t)dt, \tag{5}$$

where w is any given non-negative weight function.

The motivation behind the test statistic T_{1n} is the following: T_{1n} converges in probability to (see Theorem 7 below)

$$T_1 \equiv T_1(w) = \sum_{j=1}^{k} p_j \int \left|\varphi_j(t) - \varphi_{0j}(t)\right|^2 w(t)dt. \tag{6}$$

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}. Since this distribution is unknown, as an approximation to it we derive the asymptotic null distribution. This will be done in the next section.

Remark 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*

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

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 $\{(X_{jl}, Y_{jl}),\ 1 \leq l \leq n_j\}$ are iid observations from (X_j, Y_j) and the sets $\{(X_{1l}, Y_{1l}),\ 1 \leq l \leq n_1\}$, ..., $\{(X_{jk}, Y_{jk}),\ 1 \leq l \leq n_k\}$ are independent.

Assumption (A):

(A.1) For $1 \leq j \leq k$: (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 $1 \leq j \leq k$: the samples 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 $w(t) \geq 0$, 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 $\hat{m}_0$. Observe that no restriction on the distribution of the errors have been done, like the existence of a pdf. So the results in this paper could be used to compare two or more regression functions when the distribution of the errors is arbitrary: continuous, discrete or mixed. This is not the case for many methods that can be found in the literature, which usually assume continuity of the errors.

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

$$\begin{aligned}
\sigma_{jj} =& \quad 1 - 2p_j E\left\{\frac{f_j(X_j)}{f_{mix}(X_j)}\right\} + p_j \sum_{r=1}^{k} p_r E\left\{\frac{\sigma_r^2(X_r)}{\sigma_j^2(X_r)}\frac{f_j^2(X_r)}{f_{mix}^2(X_r)}\right\}, \\
\sigma_{jv} =& \quad \sqrt{p_j p_v}\sum_{r=1}^{k} p_r E\left\{\frac{\sigma_r^2(X_r)}{\sigma_j(X_r)\sigma_v(X_r)}\frac{f_j(X_r)f_v(X_r)}{f_{mix}^2(X_r)}\right\} \\
& -\sqrt{p_j p_v} E\left\{\frac{\sigma_v(X_v)}{\sigma_j(X_v)}\frac{f_j(X_v)}{f_{mix}(X_v)} + \frac{\sigma_j(X_j)}{\sigma_v(X_j)}\frac{f_v(X_j)}{f_{mix}(X_j)}\right\}, \quad j \neq v.
\end{aligned} \tag{7}$$

Theorem 2 *Under Assumption (A), if H_0 is true, then*

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

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 $a_j = \int t^2 |\varphi_j(t)|^2 w(t)dt$, $1 \leq j \leq k$. The results the Theorems and Corollaries below will hold whenever trace$(\mathcal{A}\Sigma) > 0$. Before stating the results, we briefly discuss this condition. Observe that

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

The quantities σ_{jj} in (7) 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 the expression above and Assumption A, it is clear that $\sigma_{jj} > 0$, for all j. Therefore, to ensure trace$(\mathcal{A}\Sigma) > 0$ we only need to ensure that $a_j > 0$ for some j. An easy way to get $a_j > 0$ is by taking $w(t) > 0$, for t in a neighborhood of the origin.

The following assumption will appear in the statement of some of the results below:

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

Corollary 3 *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 2.*

In other words, the limiting distribution of nT_{1n} under H_0 is a linear combination of independent chi-square variables, $\sum_{j=1}^{k} \beta_j \chi^2_{1,j}$, where $\chi^2_{1,1}, \ldots, \chi^2_{1,k}$ are independent chi-square 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_j's, and on the distribution of the covariates through σ_{jv}'s. They also depend on the unknown design densities, f_j, and the conditional variance functions, σ_j^2. So to use Theorem 2 in practice, one first need to find a consistent estimator, say $\hat{\beta}_j$, for every β_j, $1 \leq j \leq k$. This can be easily done via plug-in method using the kernel estimators (as defined above) 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 done via Monte Carlo method or some numerical method (see for example Kotz et al, 1967, Castaño-Martínez et al, 2005). With such a distribution, we can finally get the critical value and/or the p-value for the test based on T_{1n}. The next result states the validity of this procedure.

Let $W_{1n} = \sum_{j=1}^{k} \hat{\beta}_j \chi^2_{1j}$, where $\chi^2_{11}, \ldots, \chi^2_{1k}$ are independent chi-square variables with one degree of freedom and $\hat{\beta}_1, \ldots, \hat{\beta}_k$ are the eigenvalues of $\hat{\mathcal{A}}\hat{\Sigma}$, with $\hat{\mathcal{A}} = diag(\hat{a}_1, \ldots, \hat{a}_k)$, $\hat{\Sigma} = (\hat{\sigma}_{jv})_{1\le j,v\le k}$,

$$\hat{a}_j = \frac{-1}{\begin{pmatrix} n_j \\ 2 \end{pmatrix}} \sum_{1\le r<s\le n_j} D_2 I_w(\hat{\varepsilon}_{jr} - \hat{\varepsilon}_{js}), \quad 1 \le j \le k, \tag{8}$$

$D_2 I_w(t) = \frac{\partial^2}{\partial t^2} I_w(t)$, I_w as defined in Remark 1, $\hat{\sigma}_{jj} = 1 - 2\hat{p}_j\hat{\mu}_j + \hat{p}_j \sum_{r=1}^{k} \hat{p}_r \hat{\mu}_{jjr}$, $\hat{\sigma}_{jv} = \sqrt{\hat{p}_j\hat{p}_v} \sum_{r=1}^{k} \hat{p}_r \hat{\mu}_{jvr} - \sqrt{\hat{p}_j\hat{p}_v}(\hat{\mu}_{jv} + \hat{\mu}_{vj})$, $\quad j \ne v$, with

$$\hat{p}_j = \frac{n_j}{n}, \quad \hat{\mu}_j = \frac{1}{n_j}\sum_{l=1}^{n_j} \frac{\hat{f}_j(X_{jl})}{\hat{f}_{mix}(X_{jl})}, \quad \hat{\mu}_{jv} = \frac{1}{n_v}\sum_{l=1}^{n_v} \frac{\hat{\sigma}_v(X_{vl})}{\hat{\sigma}_j(X_{vl})} \frac{\hat{f}_j(X_{vl})}{\hat{f}_{mix}(X_{vl})},$$

$$\hat{\mu}_{jvr} = \frac{1}{n_r}\sum_{l=1}^{n_r} \frac{\hat{\sigma}^2_r(X_{rl})}{\hat{\sigma}_j(X_{rl})\hat{\sigma}_v(X_{rl})} \frac{\hat{f}_j(X_{rl})\hat{f}_v(X_{rl})}{\hat{f}^2_{mix}(X_{rl})},$$

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

Theorem 4 *Under Assumptions (A) and (B),*

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

Remark 5 *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*

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

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

Corollary 6 *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*

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

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 (8).

The result in Corollary 3 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 nT_{1n} or any consistent estimator of it. The following Theorem shows that, with probability tending to 1, T_{1n} behaves (asymptotically) like T_1, see (6). This will allow us to derive the consistency of our test.

Theorem 7 *Suppose that Assumption (A) hold. Then, $T_{1n} = T_1 + o_p(1)$, where T_1 is as defined in (6).*

As an immediate consequence of this theorem, we conclude that, for adequate choices of the weight function, the test $\Psi_{s,\alpha}$ is consistent against any fixed alternative, that is to say, it will asymptotically reject H_0 with probability one if it is not true. This property is formally stated in the following corollary.

Corollary 8 *Suppose that Assumption (A) hold. If w is such that $T_1 > 0$ whenever $m_r \neq m_s$, for some $1 \leq r, s \leq k$, $r \neq s$, then $\lim_{n\to\infty} P(\Psi_{1,\alpha} = 1) = 1$, for any $0 < \alpha < 1$.*

Since two distinct characteristic functions can be equal in a finite interval (see, for example, Feller, 1971; p. 479), a general way to ensure that $T_1 > 0$ whenever $m_r \neq m_s$, for some $1 \leq r, s \leq k$, $r \neq s$, is to take $w(t) > 0$, for all $t \in \mathbb{R}$. For instance, one can take w as the pdf of a normal law.

Finally, 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} verifies analogous conditions to those stated in assumption (A) for the functions m_j, and the functions r_j satisfy

$$E\{r_j^2(X_l)\} < \infty, \quad 1 \leq j, l \leq k. \tag{9}$$

Theorem 9 *Under Assumption (A) and the alternative hypothesis $H_{1,n}$, if (9) holds, then*

$$\sqrt{n_j}\,\{\hat{\varphi}_j(t) - \hat{\varphi}_{0j}(t)\} = it\varphi_j(t)(Z_j + \sqrt{p_j}\mu_j) + R_j(t), \quad \textit{with} \quad \int R_j(t)^2 w(t)dt = o_p(1),$$

where $Z = (Z_1, \ldots, Z_k)'$ is as in Theorem (2) and $\mu' = (\sqrt{p_1}\mu_1, \ldots, \sqrt{p_k}\mu_k)$, with

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

Corollary 10 *Under Assumption (A) and the alternative hypothesis $H_{1,n}$, if (9) holds, then $nT_{1n} \xrightarrow{\mathcal{L}} (Z+\mu)'\mathcal{A}(Z+\mu)$, where Z is as defined in Theorem 2 and μ is as in Theorem 9.*

Although the test based on the rule $\Psi_{1,\alpha}$ is fully nonparametric, it is able to detect local alternatives converging to the null hypothesis at the rate $n^{-1/2}$ whenever $\mu'\mathcal{A} \neq 0$.

The paper by Pardo-Fernández et al. (2007) studies two Kolmogorov-Smirnov and two Cramér-von Mises type statistics for testing H_0. 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

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

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 (see Remark 1)

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

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

Corollary 11 *Let $\mathcal{B} = diag(p)\mathcal{C}diag(p)$, where $p = (\sqrt{p_1}, \ldots, \sqrt{p_k})'$ and $\mathcal{C} = (c_{jv})_{1 \le j,v \le k}$ is the matrix with elements*

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

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 2.

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 $\Sigma = I_k - pp'$, with $p' = (\sqrt{p_1}, \ldots, \sqrt{p_k})$; 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

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

In the first case ($\text{trace}(\mathcal{B}\Sigma) > 0$), the asymptotic null distribution of T_{2n} is analogous to the distribution of T_{1n}, that is, a combination of chi-square random variables multiplied by the eigenvalues of $\mathcal{B}\Sigma$, which can be estimated as in Theorem 4. In the second case ($\text{trace}(\mathcal{B}\Sigma) = 0$), a deeper analysis of the asymptotic distribution is required. However, from a practical point of view, this analysis is somehow useless since 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}.

Remark 12 *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} and nT_{2n}. Nevertheless, from a computational point of view, the estimators in Theorem 4 and Corollary 6 are less time consuming. The finite sample performance of both approximations are numerically investigated in Section 4.*

4 Numerical results

In this section we report the results of an experiment carried out to study of the practical behaviour of the proposed testing procedure by means of simulations. We investigate the approximation given in Theorem 4 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 scenario:

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. The regression errors ε_1 and ε_2 are $N(0, 1)$.

Nonparametric estimation of the regression functions is performed by the local-linear estimator described in Section 2. For the estimation of the variance functions, we prefer the local-constant estimator (Nadaraya-Watson), since the local-linear can 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.

The bandwidth selection is often a complicated issue in situations involving in nonparametric smoothing. For a bandwidth of the form $h = Cn^a$, assumption (A.5) implies that $-0.5 < a < -0.25$, so the optimal bandwidth in estimation (which corresponds to $a = -0.2$ for the Nadaraya-Watson estimator and for the local-linear estimator) is excluded. In these simulations we present results for fixed bandwidths of the form $h = Cn^{-0.375}$, with $C = 1, 1.5, 2$. This choice provides reasonable values for the considered setups.

As weighting function $w(t)$ we take the pdf of a normal random variable with mean zero and standard deviation σ_w. In Table 1 we briefly investigate the effect of changing the parameter σ_w in models (iv) (level approximation) and (vi) (power) when the critical

Table 1: 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.

			α:	0.100	0.100	0.100	0.050	0.050	0.050	0.010	0.010	0.010
model	σ_w	(n_1, n_2)	C:	1.0	1.5	2.0	1.0	1.5	2.0	1.0	1.5	2.0
(iv)	0.50	(50,50)		0.119	0.103	0.103	0.066	0.054	0.047	0.012	0.010	0.005
	0.50	(100,50)		0.116	0.089	0.091	0.057	0.042	0.037	0.012	0.009	0.009
	0.50	(100,100)		0.100	0.082	0.078	0.046	0.041	0.039	0.012	0.009	0.007
	0.75	(50,50)		0.131	0.111	0.106	0.071	0.060	0.048	0.012	0.010	0.007
	0.75	(100,50)		0.123	0.100	0.098	0.060	0.049	0.041	0.011	0.010	0.011
	0.75	(100,100)		0.111	0.090	0.083	0.049	0.043	0.039	0.012	0.009	0.008
	1.00	(50,50)		0.140	0.123	0.111	0.074	0.064	0.049	0.013	0.012	0.008
	1.00	(100,50)		0.137	0.102	0.102	0.067	0.051	0.042	0.013	0.009	0.011
	1.00	(100,100)		0.122	0.097	0.084	0.052	0.046	0.041	0.012	0.009	0.008
(vi)	0.50	(50,50)		0.972	0.972	0.960	0.945	0.940	0.919	0.829	0.814	0.775
	0.50	(100,50)		0.993	0.996	0.996	0.984	0.988	0.987	0.956	0.951	0.931
	0.50	(100,100)		1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.997	0.996
	0.75	(50,50)		0.973	0.974	0.967	0.947	0.945	0.929	0.838	0.824	0.796
	0.75	(100,50)		0.993	0.997	0.996	0.985	0.992	0.989	0.961	0.952	0.938
	0.75	(100,100)		1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.997	0.997
	1.00	(50,50)		0.974	0.976	0.970	0.953	0.949	0.936	0.839	0.831	0.803
	1.00	(100,50)		0.994	0.997	0.997	0.986	0.994	0.990	0.962	0.957	0.939
	1.00	(100,100)		1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.997	0.997

values are approximated from the estimated asymptotic distribution of T_{n1} as explained in Theorem 4. The results are quite homogeneous, so this parameter does not seem to have an important impact on the results. In the rest of the simulations we only show results for $\sigma_w = 1$.

Tables 2 and 3 also show results for the test based on the estimated asymptotic distribution of T_{n1}. The level (models i-iv) is slightly overestimated 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 and in the heteroscedastic cases.

As mentioned before, the tests proposed in Pardo-Fernández et al. (2007) are based on a bootstrap approximation of the null distribution of the test statistics. In order to establish a fair comparison, we have also applied the bootstrap to our test statistic. Besides, we have also incorporated here the test statistic T_{n2}, for which the asymptotic null distribution is difficult to approximate. Table 4 shows the results of the tests based on T_{n1} and T_{n2} 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, we see that the approximation of the level is good for all test statistics, except for T^2_{KS}. Compared to the asymptotic approximation, the bootstrap approximation improves the behaviour of the test statistic T_{1n} for small sample sizes. In terms of power, for models (v)-(vii), the test based on T_{1n} is very similar to the test based on T^1_{CM}, and both outperform the other ones. For model (viii), the best power is achieved by T_{2n}. Also note in that model T_{1n} reaches a reasonable power and it is much better than T^1_{CM}.

We have also briefly 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:

(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, 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 where obtained for these values. The results are shown in Table 5. As in Tables 2 and 3, the level is well approximated for large sample sizes and the behaviour in terms of power is correct.

Table 2: Observed rejection proportions of the test based on the asymptotic distribution of T_{1n} under models (i)-(viii). The models are homoscedastic.

		α:	0.100	0.100	0.100	0.050	0.050	0.050	0.010	0.010	0.010
model	(n_1, n_2)	C:	1.0	1.5	2.0	1.0	1.5	2.0	1.0	1.5	2.0
(i)	(50,50)		0.148	0.135	0.133	0.081	0.070	0.072	0.015	0.013	0.011
	(100,50)		0.143	0.118	0.114	0.069	0.058	0.058	0.014	0.010	0.012
	(100,100)		0.126	0.107	0.105	0.056	0.048	0.045	0.012	0.011	0.013
(ii)	(50,50)		0.146	0.132	0.126	0.078	0.069	0.061	0.012	0.012	0.009
	(100,50)		0.139	0.112	0.111	0.071	0.054	0.053	0.013	0.010	0.012
	(100,100)		0.124	0.104	0.095	0.053	0.047	0.044	0.012	0.011	0.010
(iii)	(50,50)		0.109	0.097	0.118	0.059	0.046	0.061	0.007	0.007	0.012
	(100,50)		0.111	0.078	0.089	0.054	0.034	0.040	0.007	0.004	0.007
	(100,100)		0.086	0.067	0.085	0.039	0.032	0.036	0.009	0.006	0.004
(iv)	(50,50)		0.140	0.123	0.111	0.074	0.064	0.049	0.013	0.012	0.008
	(100,50)		0.137	0.102	0.102	0.067	0.051	0.042	0.013	0.009	0.011
	(100,100)		0.122	0.097	0.084	0.052	0.046	0.041	0.012	0.009	0.008
(v)	(50,50)		0.999	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000
	(100,50)		1.000	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000
	(100,100)		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(vi)	(50,50)		0.974	0.976	0.970	0.953	0.949	0.936	0.839	0.831	0.803
	(100,50)		0.994	0.997	0.997	0.986	0.994	0.990	0.962	0.957	0.939
	(100,100)		1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.997	0.997
(vii)	(50,50)		0.976	0.979	0.976	0.947	0.955	0.953	0.817	0.835	0.851
	(100,50)		0.993	0.998	0.998	0.987	0.995	0.992	0.957	0.953	0.955
	(100,100)		1.000	1.000	1.000	1.000	1.000	1.000	0.994	0.996	0.996
(viii)	(50,50)		0.838	0.688	0.567	0.659	0.484	0.367	0.318	0.179	0.117
	(100,50)		0.922	0.826	0.712	0.777	0.625	0.502	0.450	0.270	0.165
	(100,100)		0.982	0.953	0.907	0.937	0.859	0.767	0.705	0.519	0.368

Table 3: Observed rejection proportions of the test based on the asymptotic distribution of T_{1n} under models (i)-(viii). The models are heteroscedastic.

		α:	0.100	0.100	0.100	0.050	0.050	0.050	0.010	0.010	0.010
model	(n_1, n_2)	C:	1.0	1.5	2.0	1.0	1.5	2.0	1.0	1.5	2.0
(i)	(50,50)		0.147	0.127	0.122	0.075	0.072	0.065	0.014	0.010	0.009
	(100,50)		0.146	0.116	0.106	0.072	0.054	0.052	0.017	0.011	0.010
	(100,100)		0.129	0.104	0.097	0.060	0.048	0.046	0.012	0.010	0.009
(ii)	(50,50)		0.145	0.120	0.111	0.071	0.065	0.055	0.014	0.009	0.006
	(100,50)		0.142	0.110	0.102	0.070	0.052	0.048	0.017	0.009	0.010
	(100,100)		0.130	0.101	0.092	0.059	0.049	0.047	0.012	0.009	0.008
(iii)	(50,50)		0.099	0.089	0.122	0.055	0.045	0.056	0.006	0.005	0.013
	(100,50)		0.109	0.071	0.084	0.052	0.034	0.044	0.010	0.005	0.005
	(100,100)		0.091	0.068	0.081	0.038	0.029	0.034	0.009	0.006	0.005
(iv)	(50,50)		0.141	0.111	0.100	0.072	0.060	0.050	0.011	0.009	0.005
	(100,50)		0.140	0.104	0.090	0.067	0.049	0.040	0.015	0.008	0.008
	(100,100)		0.127	0.094	0.082	0.059	0.045	0.043	0.012	0.008	0.006
(v)	(50,50)		0.999	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000
	(100,50)		0.999	1.000	1.000	0.998	1.000	1.000	0.998	1.000	1.000
	(100,100)		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(vi)	(50,50)		0.961	0.964	0.944	0.932	0.927	0.904	0.801	0.787	0.746
	(100,50)		0.992	0.997	0.994	0.984	0.985	0.980	0.945	0.926	0.904
	(100,100)		0.999	1.000	1.000	0.998	1.000	1.000	0.992	0.994	0.991
(vii)	(50,50)		0.965	0.966	0.964	0.931	0.928	0.938	0.771	0.788	0.807
	(100,50)		0.994	0.998	0.996	0.983	0.988	0.987	0.932	0.927	0.920
	(100,100)		1.000	1.000	1.000	0.999	1.000	1.000	0.992	0.992	0.991
(viii)	(50,50)		0.917	0.837	0.800	0.807	0.673	0.597	0.453	0.316	0.249
	(100,50)		0.939	0.902	0.828	0.884	0.758	0.665	0.583	0.405	0.291
	(100,100)		0.991	0.990	0.975	0.979	0.955	0.921	0.864	0.763	0.661

Table 4: Observed rejection proportions of the tests based on the bootstrap approximation of T_{1n} and T_{2n}, T_{KS}^1, T_{KS}^2, T_{CM}^1 and T_{CM}^2 under models (i)-(viii). The significance level is $\alpha = 0.05$ and the bandwidth is constructed with $C = 1.0$.

		Homoscedastic models						Heteroscedastic models					
model	(n_1, n_2)	T_{1n}	T_{2n}	T_{KS}^1	T_{KS}^2	T_{CM}^1	T_{CM}^2	T_{1n}	T_{2n}	T_{KS}^1	T_{KS}^2	T_{CM}^1	T_{CM}^2
(i)	(50,50)	0.051	0.043	0.046	0.024	0.052	0.041	0.047	0.047	0.043	0.021	0.048	0.041
	(100,50)	0.047	0.048	0.052	0.036	0.047	0.041	0.040	0.048	0.052	0.037	0.045	0.045
	(100,100)	0.042	0.043	0.041	0.033	0.038	0.037	0.042	0.042	0.041	0.027	0.041	0.040
(ii)	(50,50)	0.052	0.049	0.048	0.029	0.055	0.042	0.049	0.050	0.045	0.028	0.047	0.045
	(100,50)	0.046	0.045	0.052	0.037	0.048	0.042	0.042	0.048	0.048	0.030	0.046	0.046
	(100,100)	0.042	0.042	0.041	0.031	0.043	0.043	0.043	0.047	0.038	0.027	0.041	0.041
(iii)	(50,50)	0.057	0.051	0.052	0.042	0.059	0.049	0.057	0.059	0.045	0.040	0.055	0.054
	(100,50)	0.047	0.050	0.047	0.044	0.048	0.049	0.045	0.055	0.048	0.036	0.050	0.054
	(100,100)	0.042	0.049	0.033	0.031	0.043	0.046	0.042	0.047	0.034	0.033	0.035	0.044
(iv)	(50,50)	0.055	0.053	0.052	0.033	0.055	0.042	0.054	0.053	0.045	0.029	0.048	0.041
	(100,50)	0.047	0.046	0.052	0.037	0.045	0.050	0.043	0.049	0.049	0.032	0.048	0.046
	(100,100)	0.042	0.044	0.039	0.029	0.042	0.041	0.043	0.044	0.040	0.027	0.042	0.042
(v)	(50,50)	1.000	1.000	1.000	0.979	1.000	0.998	1.000	1.000	1.000	0.979	1.000	0.996
	(100,50)	1.000	1.000	1.000	0.995	1.000	1.000	1.000	1.000	1.000	0.995	1.000	1.000
	(100,100)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(vi)	(50,50)	0.937	0.784	0.895	0.438	0.924	0.678	0.912	0.639	0.868	0.348	0.911	0.548
	(100,50)	0.988	0.913	0.980	0.652	0.989	0.871	0.981	0.791	0.968	0.479	0.981	0.734
	(100,100)	1.000	0.986	0.999	0.843	1.000	0.969	0.999	0.935	0.998	0.685	0.999	0.892
(vii)	(50,50)	0.950	0.776	0.912	0.431	0.942	0.649	0.926	0.619	0.891	0.326	0.930	0.532
	(100,50)	0.991	0.916	0.986	0.640	0.992	0.872	0.985	0.813	0.975	0.500	0.987	0.739
	(100,100)	1.000	0.987	1.000	0.838	1.000	0.970	1.000	0.937	1.000	0.680	1.000	0.889
(viii)	(50,50)	0.584	0.936	0.188	0.586	0.164	0.822	0.726	0.963	0.324	0.692	0.299	0.891
	(100,50)	0.712	0.971	0.307	0.688	0.250	0.891	0.804	0.984	0.380	0.766	0.341	0.932
	(100,100)	0.898	1.000	0.418	0.935	0.386	0.992	0.957	1.000	0.688	0.962	0.648	0.998

Table 5: Observed rejection proportions of the tests based on the asymptotic distribution of T_{1n} under models (ix)-(xiv).

		α:	0.100	0.100	0.100	0.050	0.050	0.050	0.010	0.010	0.010
model	(n_1, n_2, n_3)	C:	2.00	2.50	3.00	2.00	2.50	3.00	2.00	2.50	3.00
(ix)	(50,50,50)		0.132	0.128	0.137	0.068	0.071	0.074	0.009	0.011	0.011
	(100,50,50)		0.128	0.126	0.124	0.078	0.075	0.071	0.012	0.014	0.013
	(100,100,50)		0.138	0.131	0.130	0.080	0.081	0.085	0.021	0.020	0.019
	(100,100,100)		0.122	0.124	0.121	0.057	0.057	0.061	0.013	0.011	0.010
(x)	(50,50,50)		0.125	0.114	0.112	0.062	0.059	0.062	0.008	0.008	0.007
	(100,50,50)		0.116	0.116	0.113	0.065	0.069	0.068	0.008	0.012	0.011
	(100,100,50)		0.127	0.123	0.121	0.076	0.072	0.070	0.018	0.017	0.016
	(100,100,100)		0.109	0.105	0.099	0.055	0.054	0.054	0.012	0.010	0.009
(xi)	(50,50,50)		0.909	0.899	0.899	0.839	0.830	0.815	0.620	0.608	0.588
	(100,50,50)		0.971	0.966	0.960	0.925	0.923	0.918	0.782	0.764	0.757
	(100,100,50)		0.975	0.970	0.968	0.952	0.949	0.946	0.837	0.832	0.818
	(100,100,100)		0.996	0.996	0.996	0.995	0.991	0.990	0.956	0.953	0.953
(xii)	(50,50,50)		0.657	0.635	0.622	0.528	0.507	0.496	0.269	0.249	0.244
	(100,50,50)		0.647	0.638	0.624	0.535	0.518	0.517	0.306	0.294	0.289
	(100,100,50)		0.694	0.676	0.669	0.590	0.568	0.558	0.350	0.329	0.322
	(100,100,100)		0.889	0.883	0.885	0.823	0.813	0.804	0.643	0.630	0.619
(xiii)	(50,50,50)		0.393	0.378	0.364	0.248	0.241	0.220	0.078	0.070	0.063
	(100,50,50)		0.416	0.390	0.374	0.264	0.252	0.238	0.087	0.081	0.069
	(100,100,50)		0.391	0.366	0.350	0.234	0.221	0.215	0.071	0.067	0.062
	(100,100,100)		0.552	0.534	0.511	0.374	0.353	0.355	0.153	0.140	0.130
(xiv)	(50,50,50)		1.000	1.000	0.999	1.000	0.999	0.997	0.996	0.986	0.972
	(100,50,50)		1.000	1.000	1.000	1.000	1.000	0.999	0.995	0.992	0.981
	(100,100,50)		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(100,100,100)		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

5 Conclusions

A test for the comparison of k regression functions has been proposed and studied. The test can be seen as a cf version of a ecdf based test previously proposed in Pardo-Fernández et al. (2007). Both tests are developed under a totally nonparametric setting, and they share some desirable asymptotic properties such as consistency under both fixed and contiguous alternatives. Nevertheless, the conditions required to develop our theory are less restrictive, which is a quite desirable property. In addition, the simulations carried out reveal that the behavior of both tests are very close in terms of power. 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.

6 Appendix

We now sketch the proofs of the results stated in Section 3. With this aim, we first observe that under the considered assumptions, for $1 \le j \le k$,

$$\sup_t |\hat{m}_j(t) - m_j(t)| = o_p(n_j^{-1/4}),$$

$$\sup_t |\hat{\sigma}_j(t) - \sigma_j(t)| = o_p(n_j^{-1/4}),$$

$$\sup_t |\hat{f}_j(t) - f_j(t)| = o_p(n_j^{-1/4}).$$

Proof of Theorem 2 By Taylor's Theorem,

$$\begin{aligned}\sqrt{n_j}\,\{\hat{\varphi}_j(t) - \hat{\varphi}_{0j}(t)\} =& \sqrt{n_j}\left\{\frac{1}{n_j}\sum_l \exp(\mathrm{i}t\hat{\varepsilon}_{jl}) - \frac{1}{n_j}\sum_l \exp(\mathrm{i}t\hat{\varepsilon}_{0jl})\right\}\\ &= \mathrm{i}t\frac{1}{\sqrt{n_j}}\sum_l (\hat{\varepsilon}_{jl} - \hat{\varepsilon}_{0jl})\exp(\mathrm{i}t\hat{\varepsilon}_{0jl}) + t^2 R_{1n_j}(t),\end{aligned}$$

where

$$\sup_t |R_{1n_j}(t)| \le \sqrt{n_j}O_p(1)\{\sup_t |\hat{m}_j(t) - \hat{m}_0(t)|\}^2 = o_p(1).$$

Again by Taylor's Theorem,

$$\mathrm{i}t\frac{1}{\sqrt{n_j}}\sum_l(\hat{\varepsilon}_{jl}-\hat{\varepsilon}_{0jl})\exp(\mathrm{i}t\hat{\varepsilon}_{0jl})=\mathrm{i}t\frac{1}{\sqrt{n_j}}\sum_l(\hat{\varepsilon}_{jl}-\hat{\varepsilon}_{0jl})\exp(\mathrm{i}t\varepsilon_{jl})+t^2R_{2n_j}(t),$$

with

$$\sup_t|R_{2n_j}(t)|\leq O_p(1)\sup_t|\hat{m}_j(t)-\hat{m}_0(t)|\left\{\sqrt{n_j}\sup_t|m_0(t)-\hat{m}_0(t)|+\right.$$

$$\left.+\sup_t|\hat{\sigma}_j(t)-\sigma_j(t)|\frac{1}{\sqrt{n_j}}\sum_l|\varepsilon_{jl}|\right\}=o_p(1).$$

From the definition of the residuals,

$$\mathrm{i}t\frac{1}{\sqrt{n_j}}\sum_l(\hat{\varepsilon}_{jl}-\hat{\varepsilon}_{0jl})\exp(\mathrm{i}t\varepsilon_{jl})=\hat{A}_j(t)+tR_{3n_j}(t),$$

where

$$\hat{A}_j(t)=\mathrm{i}t\frac{1}{\sqrt{n_j}}\sum_l\frac{\hat{m}_0(X_{jl})-\hat{m}_j(X_{jl})}{\sigma_j(X_{jl})}\exp(\mathrm{i}t\varepsilon_{jl}),\tag{10}$$

and

$$\sup_t|R_{3n_j}(t)|\leq O_p(1)\sqrt{n_j}\sup_t|\hat{m}_j(t)-\hat{m}_0(t)|\sup_t|\hat{\sigma}_j(t)-\sigma_j(t)|=o_p(1).$$

Under H_0,

$$\hat{m}_0(x)-m_0(x)=\sum_{v=1}^{k}w_v(x)\left\{\hat{m}_v(x)-m_v(x)\right\}+o_p(n^{-1/2}),$$

uniformly in x, where $w_v(x)=\frac{n_v}{n}\frac{f_v(x)}{f_{mix}(x)}$. Therefore, $\hat{A}_j(t)=\hat{A}_{2j}(t)-\hat{A}_{1j}(t)+tR_{4n_j}(t)$, with $\sup_t|R_{4n_j}(t)|=o_P(1)$,

$$\hat{A}_{2j}(t)=\sum_{v=1}^{k}\frac{\mathrm{i}t}{\sqrt{n_j}}\sum_{l=1}^{n_j}\frac{w_v(X_{jl})}{\sigma_j(X_{jl})}\left\{\hat{m}_v(X_{jl})-m_v(X_{jl})\right\}\exp(it\varepsilon_{jl}),\tag{11}$$

and

$$\hat{A}_{1j}(t)=\frac{\mathrm{i}t}{\sqrt{n_j}}\sum_{l=1}^{n_j}\frac{1}{\sigma_j(X_{jl})}\left\{\hat{m}_j(X_{jl})-m_j(X_{jl})\right\}\exp(it\varepsilon_{jl}).\tag{12}$$

Standard arguments (see e.g. Masry, 1996) show that, under the assumed conditions,

$$\hat{m}_v(x)-m_v(x)=\frac{1}{n_vf_v(x)}\sum_{s=1}^{n_v}K_h(x-X_{vs})\sigma_v(X_{vs})\varepsilon_{vs}+o_p(n_v^{-1/2}),$$

uniformly in x. Thus,

$$\hat{A}_{2j}(t)=\mathrm{i}t\frac{1}{n\sqrt{n_j}}\sum_{v=1}^{k}n_jn_vU_{v,j}(t)+tR_{5n_j}(t),$$

where $\sup_t |R_{5n_j}(t)| = o_P(1)$ and

$$U_{v,j}(t) = \frac{1}{n_j n_v} \sum_{l=1}^{n_j} \sum_{s=1}^{n_v} S_n(X_{jl}, \varepsilon_{jl}; X_{vs}, \varepsilon_{vs}; t),$$

with

$$S_n(X_{jl}, \varepsilon_{jl}; X_{vs}, \varepsilon_{vs}; t) = \frac{\sigma_v(X_{vs})}{f_{mix}(X_{jl})\sigma_j(X_{jl})} K_h(X_{jl} - X_{vs})\varepsilon_{vs} \exp(\mathrm{i}t\varepsilon_{jl}).$$

If $j \neq v$, then, for every t, $U_{v,j}(t)$ is a two sample U-statistic of degree $(1,1)$ with kernel $S_n(X_{jl}, \varepsilon_{jl}; X_{vs}, \varepsilon_{vs}; t)$. Its Hájek projection is given by

$$\hat{U}_{v,j}(t) = \varphi_j(t) \frac{1}{n_v} \sum_{s=1}^{n_v} \varepsilon_{vs} \sigma_v(X_{vs}) \tilde{h}_n(X_{vs}),$$

where $\tilde{h}_n(x) = E\left(\frac{K_h(X_j - x)}{f_{mix}(X_j)\sigma_j(X_j)}\right)$. We have (see e.g. Theorem 11.1 and the proof of Theorem 12.6 in van der Vaart, 1998)

$$E[\{U_{v,j}(t) - \hat{U}_{v,j}(t)\}^2] = Var\{U_{v,j}(t)\} - Var\{\hat{U}_{v,j}(t)\} = \frac{1}{n_v n_j} E\{S_n^2(X_j, \varepsilon_j; X_v, \varepsilon_v; t)\}.$$

Because $E\{S_n^2(X_j, \varepsilon_j; X_v, \varepsilon_v; t)\} = O(1/h) = o(n)$, uniformly in t, we obtain $E[\{U_{v,j}(t) - \hat{U}_{v,j}(t)\}^2] = o(n^{-1})$, uniformly in t. This together with the fact that $\tilde{h}_n(x) = \frac{f_j(x)}{f_{mix}(x)\sigma_j(x)} + O(h^2)$, uniformly in x, implies

$$U_{v,j}(t) = \varphi_j(t) \frac{1}{n_v} \sum_{s=1}^{n_v} \frac{f_j(X_{vs})\sigma_v(X_{vs})}{f_{mix}(X_{vs})\sigma_j(X_{vs})} \varepsilon_{vs} + R_{v,j}(t), \quad \text{with} \quad \sup_t |R_{v,j}(t)| = o_p(n^{-1/2}). \tag{13}$$

If $j = v$, then $U_{j,j}(t)$ can be written as

$$U_{j,j}(t) = \frac{K(0)}{n_j^2 h} \sum_{s=1}^{n_j} \frac{\varepsilon_{js} \exp(it\varepsilon_{js})}{f_{mix}(X_{js})} + \frac{n_j - 1}{2n_j} U_j(t),$$

where, for every t, $U_j(t)$ is a one sample U-statistic of degree 2 with kernel $S_n(X_{jl}, \varepsilon_{jl}; X_{js}, \varepsilon_{js}; t) + S_n(X_{js}, \varepsilon_{js}; X_{jl}, \varepsilon_{jl}; t)$. Arguments very similar to those employed for the case $j \neq v$ can be used to show that $U_j(t) = 2\hat{U}_{j,j}(t) + R_j(t)$, with $\sup_t |R_j(t)| = o_p(n^{-1/2})$. Since

$$\sqrt{n} \frac{K(0)}{n_j^2 h} \left| \sum_{s=1}^{n_j} \frac{\varepsilon_{js} \exp(it\varepsilon_{js})}{f_{mix}(X_{js})} \right| \leq \frac{M}{\sqrt{nh^2}} \frac{1}{n_j} \sum_{l=1}^{n_j} |\varepsilon_{jl}| = o_P(1), \quad \forall t,$$

for some constant $M > 0$, we get that $U_{j,j}(t)$ also satisfies (13) with $j = v$, and thus

$$\hat{A}_{2j}(t) = \mathrm{i}t \frac{p_j}{\sqrt{n_j}} \varphi_j(t) \sum_{v=1}^{k} \sum_{s=1}^{n_v} \frac{f_j(X_{vs})}{f_{mix}(X_{vs})} \frac{\sigma_v(X_{vs})}{\sigma_j(X_{vs})} \varepsilon_{vs} + tR_{2j}(t), \quad \text{with} \quad \sup_t |R_{2j}(t)| = o_p(1). \tag{14}$$

Similarly, it can be proved that

$$\hat{A}_{1j}(t) = \mathrm{i}t\frac{1}{\sqrt{n_j}}\varphi_j(t)\sum_{s=1}^{n_j}\varepsilon_{js} + tR_{1j}(t), \quad \text{with} \quad \sup_t |R_{1j}(t)| = o_p(1). \tag{15}$$

We conclude that, under H_0,

$$\sqrt{n_j}\left\{\hat{\varphi}_j(t) - \hat{\varphi}_{0j}(t)\right\} = \mathrm{i}t\varphi_j(t)Z_{n,j} + tR_1(t) + t^2R_2(t), \quad \text{with} \quad \sup_t |R_s(t)| = o_p(1), \; s = 1, 2,$$

with

$$Z_{n,j} = \frac{\sqrt{p}_j}{\sqrt{n}}\sum_{v=1}^{k}\sum_{s=1}^{n_v}\frac{f_j(X_{vs})}{f_{mix}(X_{vs})}\frac{\sigma_v(X_{vs})}{\sigma_j(X_{vs})}\varepsilon_{vs} - \frac{1}{\sqrt{n}_j}\sum_{s=1}^{n_j}\varepsilon_{js}. \tag{16}$$

The result follows by applying the CLT to $Z_n = (Z_{n,1}, \ldots, Z_{n,k})'$. □

Proof of Corollary 3 From the definition of T_{1n} and the result in Theorem 2 we get,$nT_{1n} = \sum_{j=1}^{k} Z_j^2 \int t^2|\varphi_j(t)|^2 w(t)dt + o_p(1) = Z^t\mathcal{A}Z + o_p(1) \xrightarrow{\mathcal{L}} W_1$. □

Proof of Theorem 4 Since asymptotically, under H_0, $T_{1n}(w)$ is a continuous function of the quantities a_j, p_j and σ_{rj}^2, $1 \leq r, j \leq k$, to prove the result it suffices to prove the consistency of the estimators of these quantities. From assumption (A.2), $\hat{p}_j \to p_j$, $1 \leq j \leq k$. Next, to prove the consistency of $\hat{a}_j$ we will consider the following equivalent expression

$$\hat{a}_j = \frac{n_j}{n_j - 1}\left\{\int t^2 \left|\hat{\varphi}_j(t)\right|^2 w(t)dt\right\} - \frac{1}{n_j - 1}\int t^2 w(t)dt.$$

Observe that for $1 \leq j \leq k$ and $l = 1, \ldots, n_j$,

$$\hat{\varepsilon}_{jl} - \varepsilon_{jl} = \left(\frac{\sigma_j(X_{jl})}{\hat{\sigma}_j(X_{jl})} - 1\right)\varepsilon_{jl} - \frac{\hat{m}_j(X_{jl}) - m_j(X_{jl})}{\hat{\sigma}_j(X_{jl})}.$$

By Taylor's Theorem,

$$n_j^{-1}\sum_l \exp(\mathrm{i}t\hat{\varepsilon}_{jl}) = n_j^{-1}\sum_l \exp(\mathrm{i}t\varepsilon_{jl}) + tR_{1n_j}(t), \quad \text{with} \quad \sup_t |R_{1n_j}(t)| = o_P(1), \tag{17}$$

and thus,

$$\hat{a}_j = U_j + o_P(1), \quad 1 \leq j \leq k, \tag{18}$$

with

$$U_j = \frac{-1}{\binom{n_j}{2}}\sum_{1 \leq r < s \leq n_j} D_2 I_w(\varepsilon_{jr} - \varepsilon_{js}), \quad 1 \leq j \leq k,$$

which is a degree-2 U-statistic and therefore (see Serfling, 1980)

$$U_j = a_j + o_P(1), \quad 1 \leq j \leq k. \tag{19}$$

Now, from (18) and (19), we get $\hat{a}_j = a_j + o_P(1)$, $1 \le j \le k$.

From the assumed conditions and the SLLN, we get

$$\hat{\mu}_j = \frac{1}{n_j}\sum_{l=1}^{n_j}\frac{\hat{f}_j(X_{jl})}{\hat{f}_{mix}(X_{jl})} = \frac{1}{n_j}\sum_{l=1}^{n_j}\frac{f_j(X_{jl})}{f_{mix}(X_{jl})} + o_p(n^{-1/4}) = E\left\{\frac{f_j(X_j)}{f_{mix}(X_j)}\right\} + o_P(1),$$

$1 \le j \le k$. The consistency of $\hat{\mu}_{jv}$ and $\hat{\mu}_{jvu}$ can be analogously dealt with, $1 \le j, v, u \le k$. $\square$

Proof of Theorem 6 The proof follows from the proof of Theorem 4, so we omit it. $\square$

Proof of Theorem 7 Under the assumed conditions (17) holds. Similarly, $n_j^{-1}\sum_l \exp(\mathrm{i}t\hat{\varepsilon}_{0jl}) = n_j^{-1}\sum_l \exp(\mathrm{i}t\varepsilon_{0jl}) + tR_{2n_j}(t)$, with $\sup_t |R_{2n_j}(t)| = o_P(1)$, where $\varepsilon_{0jl} = \{Y_{jl} - m_0(X_{jl}\}/\sigma_j(X_{jl})$ Let $\tilde{\varphi}_j(t) = n_j^{-1}\sum_l \exp(\mathrm{i}t\varepsilon_{jl})$ and $\tilde{\varphi}_{0j}(t) = n_j^{-1}\sum_l \exp(\mathrm{i}t\varepsilon_{0jl})$. We conclude that

$$T_{1n} = \sum_{j=1}^{k}\frac{n_j}{n}\int |\tilde{\varphi}_j(t) - \tilde{\varphi}_{0j}(t)|^2\, w(t)dt + o_p(1) = \sum_{j=1}^{k}\frac{n_j}{n}V_j + o_p(1),$$

where $V_j = \dfrac{1}{n_j^2}\displaystyle\sum_{r,s=1}^{n_j}\{I_w(\varepsilon_{jr} - \varepsilon_{js}) + I_w(\varepsilon_{0jr} - \varepsilon_{0js}) - 2I_w(\varepsilon_{jr} - \varepsilon_{0js})\}$, with I_w as defined in Remark 1. For $1 \le j \le k$, V_j is a V-statistic of degree 2 with a bounded kernel and thus (see Serfling, 1980) it converges to it expected value which is

$$\int |\varphi_j(t) - \varphi_{0j}(t)|^2 w(t)dt.$$

This concludes the proof. $\square$

Proof of Theorem 9 We first note that under $H_{1,n}$,

$$m_0 = m_{00} + \frac{1}{\sqrt{n}}r_0, \quad \text{with} \quad r_0(t) = \sum_{v=1}^{k} p_v \frac{f_v(t)}{f_{mix}(t)} r_v(t).$$

Thus from (9),

$$\frac{1}{\sqrt{n_j}}\sum_l \{m_j(X_{jl}) - m_0(X_{jl})\}^2 = \frac{1}{n\sqrt{n_j}}\sum_l \{r_j(X_{jl}) - r_0(X_{jl})\}^2 = O_P(n^{-1/2}), \quad 1 \le j \le k. \tag{20}$$

Now, taking into account (20) and following the same steps as those given in the proof of Theorem 2, we get

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

with $\hat{A}_j(t)$ as defined in (10) and $\sup_t |R_{sn_j}(t)| = o_p(1),\ s = 1, 2$. Under $H_{1,n}$,

$$\hat{m}_0(x) - m_0(x) = \sum_{v=1}^{k} \frac{n_v}{n} \frac{f_v(x)}{f_{mix}(x)} \{\hat{m}_v(x) - m_v(x)\} + \frac{1}{\sqrt{n}} \sum_{v=1}^{k} \left\{ \frac{n_v}{n} \frac{\hat{f}_v(x)}{\hat{f}_{mix}(x)} - p_v \frac{f_v(x)}{f_{mix}(x)} \right\} r_v(x) + o_p(n^{-1/2}),$$

uniformly in x. Therefore, $\hat{A}_j(t) = \tilde{A}_{2j}(t) - \hat{A}_{1j}(t) + \hat{A}_{3j}(t) + tR_{3n_j}(t)$, with $\sup_t |R_{3n_j}(t)| = o_P(1)$, $\hat{A}_{1j}(t)$ as in (12) $\tilde{A}_{2j}(t) = \hat{A}_{2j}(t) + \mathrm{i}t\hat{A}_{22j}(t)$, $\hat{A}_{2j}(t)$ as in (11),

$$\hat{A}_{22j}(t) = \frac{1}{\sqrt{nn_j}} \sum_{v=1}^{k} \sum_{l} \left\{ \frac{n_v}{n} \frac{\hat{f}_v(X_{jl})}{\hat{f}_{mix}(X_{jl})} - p_v \frac{f_v(X_{jl})}{f_{mix}(X_{jl})} \right\} r_v(X_{jl}) \exp(\mathrm{i}t\varepsilon_{jl}),$$

$$\hat{A}_{3j}(t) = it \frac{1}{\sqrt{n_j}} \sum_{l} \frac{m_j(X_{jl}) - m_0(X_{jl})}{\sigma_j(X_{jl})} \exp(\mathrm{i}t\varepsilon_{jl}).$$

From (9), $\sup_t |\hat{A}_{22j}(t)| = o_P(1)$. Now let us consider the Hilbert space $\mathcal{H} = L^2(\mathbb{R}, w)$ of (equivalence classes of) measurable functions $g : \mathbb{R} \to \mathbb{R}$, with $\|g\|_w^2 = \int g(t)^2 w(t)dt < \infty$. From the SLLN in Hilbert spaces, $\hat{A}_{3j}(t) = it\varphi_j(t)\sqrt{p_j}\mu_j + R_{3j}(t)$, with $\|R_{3j}(t)\|_w = o_P(1)$. Finally, taking into account (14) and (15), we conclude that under $H_{1,n}$

$$\sqrt{n_j}\,\{\hat{\varphi}_j(t) - \hat{\varphi}_{0j}(t)\} = \mathrm{i}t\varphi_j(t)(Z_{n,j} + \sqrt{p_j}\mu_j) + R_j(t), \quad \text{with} \quad \|R_j(t)\|_w = o_p(1),$$

with $Z_{n,j}$ as defined in (16). The result follows from the CLT. □

Proof of Corollary 11 Analogously to the proof of Corollary 3, from the definition of T_{2n} and the result in Theorem 2, we get

$$nT_{2n} = \int \left| \sum_{j=1}^{k} \sqrt{p_j} t\varphi_j(t) Z_j \right|^2 w(t)dt + o_p(1) =$$

$$\int \left\{ \sum_{j=1}^{k} \sqrt{p_j} t\varphi_j(t) Z_j \right\} \left\{ \sum_{v=1}^{k} \sqrt{p_v} t\overline{\varphi_v(t)} Z_v \right\} w(t)dt + o_p(1) =$$

$$\sum_{j,v=1}^{k} \sqrt{p_j p_v} Z_j Z_v \int t^2 \varphi_j(t)\overline{\varphi_v(t)} w(t)dt + o_p(1).$$

Note that the coefficient of $Z_j Z_v$ in the non-negligible term in the above expression is

$$\sqrt{p_j p_v} \left\{ \int t^2 \varphi_j(t)\overline{\varphi_v(t)} w(t)dt + \int t^2 \varphi_v(t)\overline{\varphi_j(t)} w(t)dt \right\} =$$

$$2\sqrt{p_j p_v} Re \left\{ \varphi_j(t)\overline{\varphi_v(t)} \right\} = 2\sqrt{p_j p_v} Re \left\{ \varphi_v(t)\overline{\varphi_j(t)} \right\}.$$

Thus,

$$nT_{2n} = Z^t \mathcal{B} Z + o_p(1) \xrightarrow{\mathcal{L}} W_2. \quad \square$$

Acknowledgements

J.C. Pardo-Fernández has been financially supported by Grants MTM2008-03129 and MTM2011-23204 (FEDER support included) of the Spanish Ministerio de Ciencia e Innovación and 10PXIB300068PR of the Xunta de Galicia. M.D. Jiménez-Gamero has been partially supported by grant MTM2008-00018 of the Spanish Ministerio de Ciencia e Innovación. 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

Akritas, M.G., Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, 28, 549–567.

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

Castaño-Martínez, A., López-Blázquez, F. (2005). Distribution od a sum of weighted central chi-square variables. *Communications in Statistics-Theory and Methods*, 34, 515–524.

Delgado, M.A. (1993). Testing the equality of nonparametric regression curves. *Statistics and Probability Letters*, 17, 199–204.

Fan, J., Gijbels. I. (1996). *Local Polynomial Modelling and Its Applications. Chapman & Hall.*

Feller, W. (1971). *An Introduction to Probability Theory and its Applications, Vol 2.* Wiley.

Hušková, M., Meintanis, S. (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.

Jiménez-Gamero, M.D., Muñoz-García, J., Pino-Mejías, R. (2005). Testing goodness of fit for the distribution of errors in multivariate linear models. *Journal of Multivariate Analalysis*, 95, 301–322.

Kotz, S., Johnson, N.L., Boyd, D.W. (1967). Series representations of quadratic forms in normal variables. I. Central case. *The Annals of Mathematical Statistics*, 38 , 823–837.

Kulasekera, K.B. (1995). Comparison of regression curves using quasi-residuals. *Journal of the American Statistical Association*, 90, 1085–1093.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, 17, 571–599.

Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of non-parametric regression residuals. *Scandinavian Journal of Statistics*, 36, 204–228.

Neumeyer, N., Dette, H. (2003). Nonparametric comparison of regression curves: an empirical process approach. *The Annals of Statistics*, 31, 880-920.

Neumeyer, N., Pardo-Fernández, J.C. (2009). A simple test for comparing regression curves versus one-sided alternatives. *Journal of Statistical Planning and Inference*, 139, 4006-4016.

Pardo-Fernández, J.C., Van Keilegom, I., González-Manteiga, W. (2007). Testing for the equality of k regression curves. *Statistica Sinica*, 17, 1115-1137.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley.

Srihera, R., Stute, W. (2010). Nonparametric comparison of regression functions. *Journal of Multivariate Analysis*, 101, 2039–2059.

van der Vaart, A. W. (1998). Asymptotic Statistics. Cambridge University Press.