Computational Statistics and Data Analysis 97 (2016) 151–168

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Robust testing for superiority between two regression curves

Graciela Boente [a], Juan Carlos Pardo-Fernández [b,*]

[a] *Departamento de Matemáticas, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and IMAS, CONICET, Ciudad Universitaria, Pabellón 1, 1428, Buenos Aires, Argentina*

[b] *Departamento de Estatística e Investigación Operativa, Universidade de Vigo, Campus Universitario As Lagoas-Marcosende, Vigo, 36310, Spain*

ARTICLE INFO

Article history:
Received 23 June 2015
Received in revised form 30 November 2015
Accepted 1 December 2015
Available online 14 December 2015

Keywords:
Hypothesis testing
Nonparametric regression models
Robust inference
Smoothing techniques

ABSTRACT

The problem of testing the null hypothesis that the regression functions of two populations are equal versus one-sided alternatives under a general nonparametric homoscedastic regression model is considered. To protect against atypical observations, the test statistic is based on the residuals obtained by using a robust estimate for the regression function under the null hypothesis. The asymptotic distribution of the test statistic is studied under the null hypothesis and under root$-n$ local alternatives. A Monte Carlo study is performed to compare the finite sample behaviour of the proposed tests with the classical one obtained using local averages. A sensitivity analysis is carried on a real data set.

1. Introduction

Let us assume that the random vectors $(X_j, Y_j)^{\mathrm{T}} \in \mathbb{R}^2, j = 1, 2$, follow the homoscedastic nonparametric regression models given by

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

where $m_j : \mathbb{R} \to \mathbb{R}$ is a nonparametric smooth function and the error ε_j is independent of the covariate X_j. Throughout this paper, we will not require any moment conditions on the error distributions. As is usual in a robust framework, let us assume that the errors ε_j are such that $\varepsilon_j = \sigma_j\, U_j$, where U_j has a symmetric distribution $G_j(\cdot)$ with scale 1, so that we are able to identify the error's scale, σ_j. When second moments exist, as the case of the classical approach is, these conditions imply that $\mathbb{E}(\varepsilon_j) = 0$ and $\mathrm{VAR}(\varepsilon_j) = \sigma_j^2$, which means that m_j represents the conditional mean, while σ_j^2 equals the residuals variance, i.e., $\sigma_j^2 = \mathrm{VAR}(Y_j - m_j(X_j))$. The nonparametric nature of model (1) offers more flexibility than the standard linear model when modelling a complicated relationship between the response variable and the covariate. In many situations, it is of interest to compare the regression functions m_1 and m_2 to decide if the same functional form appears in both populations. In particular, in this paper we focus on testing the null hypothesis of equality of the regression curves versus a one-sided alternative. Let $\mathcal{R}$ be the common support of the covariates X_1 and X_2 where the comparison will be performed. The null hypothesis to be considered is

$$\mathrm{H}_0 : m_1(x) = m_2(x) \quad \text{for all } x \in \mathcal{R},$$

* Corresponding author. Fax: +34 986 812 401.
E-mail addresses: gboente@dm.uba.ar (G. Boente), juancp@uvigo.es (J.C. Pardo-Fernández).

http://dx.doi.org/10.1016/j.csda.2015.12.002
0167-9473/

while the alternative hypothesis is of the following one-sided type

$$\begin{aligned} &\mathrm{H}_1 : m_1(x) \le m_2(x) \text{ for all } x \in \mathcal{R} \text{ and } m_1(x) < m_2(x) \text{ for } x \in \mathcal{A}, \\ &\quad \text{where } \mathcal{A} \subset \mathcal{R} \text{ is such that } \mathbb{P}(X_j \in \mathcal{A}) > 0, \text{ for } j = 1, 2. \end{aligned} \tag{2}$$

When second moments exist, the problem of testing equality of two regression curves versus one-sided alternatives has been considered by several authors such as Hall et al. (1997), Koul and Schick (1997, 2003) and Neumeyer and Dette (2005), who extended the test proposed in Speckman et al. (2003) to allow for heteroscedasticity. On the other hand, Neumeyer and Pardo-Fernández (2009) introduced a simple root-n test statistic based on the comparison of the sample averages of the estimated residuals, which were computed with respect to a linear convex combination of the kernel regression estimators obtained from each sample.

As is well known, linear kernel regression estimators are sensitive to atypical observations, since they are based on averaging the responses. When estimating the regression function at a value x, the effect of an outlier in the responses will be larger as the distance between the related covariate and the point x is smaller. In this sense, atypical data in the responses in nonparametric regression may lead to a complete distorted estimation which will clearly influence the test statistic and the conclusions of the testing procedure. In this sense, robust estimates are needed in order to provide more reliable estimations and inferences. Beyond the importance of developing robust estimators, the problem of obtaining robust hypothesis testing procedures also deserves attention. In linear regression, recent developments were given, among others, by Salibian-Barrera et al. (2016), where also references to previous robust proposals can be found. However, in the nonparametric setting, robust testing procedures are very scarce. Recently, Dette and Marchlewski (2010) considered a robust test for homoscedasticity in nonparametric regression. On the other hand, under a partly linear regression model, Bianco et al. (2006) proposed a test to study if the nonparametric component equals a fixed given function, while Boente et al. (2013) considered the hypothesis that the nonparametric function is a linear function under a generalized partially linear model. For the problem of testing superiority between two regression curves, Koul and Schick (1997) defined a family of covariate-matched statistics and derived its asymptotic behaviour under the null hypothesis and under root-n local alternatives. This family includes, in particular, a covariate-matched Wilcoxon–Mann–Whitney test based on the sign of all response differences which does not require the existence of second moments. Besides, these authors provide an asymptotic optimality theory allowing to obtain locally asymptotically minimax tests against nonparametric root-n alternatives. To derive these properties, Koul and Schick (1997) assume equal error distributions and equal design densities. In order to avoid these assumptions, Koul and Schick (2003) developed a modified version of one of the covariate-matched statistics based on the response differences of Koul and Schick (1997), but this statistic is not robust when atypical data arise in the responses, as it assumes the existence of second moments. When considering the problem of comparing two or more regression functions, Feng et al. (2015) considered a test for H_0 versus the general alternative $m_1 \neq m_2$ using a generalized likelihood ratio test incorporating a Wilcoxon likelihood function and kernel smoothers, which allows to detect alternatives with rate $\sqrt{nh}$, where h is the bandwidth parameter; however, these authors assume the existence of second moment of the regression errors, so the applicability of their method in a robust context is quite limited.

The aim of this paper is to propose a class of robust tests for H_0 versus H_1 in (2) which allows for possibly different covariate densities and error densities in the two populations. Our proposal combines the ideas of robust smoothing with those given in Neumeyer and Pardo-Fernández (2009) to obtain a procedure detecting root-n alternatives. In Section 2, we recall the definition of the robust estimators. The test statistics is introduced in Section 3, where its asymptotic behaviour under the null hypothesis and root-n local alternatives is also studied. We present the results of a Monte Carlo study in Section 4 and an illustration to a real data set in Section 5. The Appendix A contains some auxiliary results about the robust nonparametric estimator presented in Section 2 and the proof of our main result.

2. Basic definitions and notation

Throughout this paper, we consider independent and identically distributed observations $(X_{ij}, Y_{ij})^{\mathrm{T}}$, $1 \le i \le n_j$, with the same distribution as $(X_j, Y_j)^{\mathrm{T}}, j = 1, 2$. When $\mathbb{E}|Y_j| < \infty$, the regression functions m_j in (1), which in this case equals $\mathbb{E}(Y_j|X_j)$, can be estimated by using the Nadaraya–Watson estimator (see, for example Härdle, 1990). To be more precise, let K be a kernel function (usually a symmetric density) and $h = h_n$ a sequence of strictly positive real numbers. Denote $K_h(u) = h^{-1}K(u/h)$. Then, the classical regression estimators of m_j are defined as

$$\widehat{m}_{j,\mathrm{CL}}(x) = \left\{ \sum_{\ell=1}^{n_j} K_h\left(x - X_{\ell j}\right) \right\}^{-1} \sum_{i=1}^{n_j} K_h\left(x - X_{ij}\right) Y_{ij}. \tag{3}$$

As mentioned in the introduction, the estimators defined in (3) are sensitive to atypical observations, since they are based on averaging the responses. Robust estimates in a non-parametric setting need to be employed to provide estimators insensitive to a single wild spike outlier. Several proposals have been considered and studied in the literature. We can mention, among others, Härdle and Tsybakov (1988) and Boente and Fraiman (1989), who considered robust equivariant estimators under a general heteroscedastic regression model. It is well known that, under a homoscedastic regression model, root-n scale estimators can be obtained. In particular, for fixed designs, scale estimators based on differences are widely used, see, for instance, Rice (1984) and Hall et al. (1990). A robust version of the difference-based estimators was studied in Ghement et al. (2008) and can easily be extended to the situation where X_j are random. Effectively, as in Dette and Munk (1998), let $X_{(1),j} \leq \cdots \leq X_{(n_j),j}$ be the ordered statistics of the explanatory variables of the jth population and denote as $(X_{(1),j}, Y_{D_{1,j},j})^{\mathrm{T}}, \ldots, (X_{(n_j),j}, Y_{D_{n_j,j},j})^{\mathrm{T}}$ the sample of observations ordered according to the values of the explanatory variables, that is, $X_{(\ell),j} = X_{D_{\ell,j},j}$. Then, the estimators defined in Ghement et al. (2008) can be generalized to the present situation by taking the differences $Y_{D_{\ell+1,j},j} - Y_{D_{\ell,j},j}$. Thus, for instance, a robust consistent root-n estimator of σ_j can be obtained as

$$\widehat{\sigma}_j = \frac{1}{\sqrt{2}\Phi^{-1}(3/4)} \operatorname*{median}_{1\leq \ell \leq n_j-1} \left| Y_{D_{\ell+1,j},j} - Y_{D_{\ell,j},j} \right| , \tag{4}$$

where the coefficient $\sqrt{2}\Phi^{-1}(3/4)$ ensures Fisher-consistency for normal errors (Φ^{-1} denotes the quantile function of the standard normal).

Let $\Psi_j : \mathbb{R} \to \mathbb{R}, j = 1, 2$, be bounded and continuous functions and define the function

$$\lambda_j(x, a, \sigma) = \mathbb{E}\left[\Psi_j\left(\frac{Y_j - a}{\sigma} \right) |X_j = x \right]. \tag{5}$$

Note that if (1) holds, Ψ_j is an odd function and the errors have a symmetric distribution, then $\lambda_j(x, m_j, \sigma) = 0$ for any $\sigma > 0$. Hence, to obtain robust estimators of $m_j(x)$, as in Boente and Fraiman (1989), we plug into (5) an estimator of the conditional distribution of $Y_j|X_j = x$ and a robust estimator of the error's scale $\widehat{\sigma}_j$, such as the one defined in (4). The robust nonparametric estimator of $m_j(x)$ is given by

$$\text{the solution } \widehat{m}_j(x) \text{ of } \widehat{\lambda}_j(x, \widehat{m}_j(x), \widehat{\sigma}_j) = 0, \tag{6}$$

where

$$\widehat{\lambda}_j(x, a, \sigma) = \sum_{i=1}^{n_j} K_h\left(x - X_{ij}\right) \Psi_j\left(\frac{Y_{ij} - a}{\sigma} \right). \tag{7}$$

Note that different score functions Ψ_j can be used in the two samples, in this way, we provide a more flexible setting. In the Appendix A, we give general asymptotic results related to the estimator $\widehat{m}_j(x)$ that will be used in the study of the asymptotic behaviour of the test statistic considered below.

3. The test statistic

As mentioned in the introduction, we wish to develop a class of robust tests for H_0 versus H_1 in (2) which allows to detect root-n local alternatives. As in Neumeyer and Pardo-Fernández (2009), let m be any function such that $m_1(x) \leq m(x) \leq m_2(x)$, for all $x \in \mathcal{R}$, and define the random variables, for $j = 1, 2$,

$$\varepsilon_{j0} = Y_j - m(X_j).$$

Let Ψ be an increasing function such that $\mathbb{E}[\Psi(\varepsilon_j/\sigma)]$ exists for any $\sigma > 0$ and $j = 1, 2$, which will be the case if Ψ is a bounded function. Moreover, let $w_j : \mathbb{R} \to \mathbb{R}$ be a non-negative weight function with compact support $\mathcal{S}_j \subset \mathring{\mathcal{R}}$ such that $\mathcal{A} \cap \mathcal{S}_j \neq \emptyset$. Since Ψ is increasing, we obtain

$$\begin{aligned}
\mathbb{E}\left[\Psi\left(\frac{\varepsilon_{10}}{\sigma} \right) w_1(X_1) \right] &= \mathbb{E}\left[\Psi\left(\frac{\varepsilon_1 + m_1(X_1) - m(X_1)}{\sigma} \right) w_1(X_1) \right] \\
&= \mathbb{E}\left[w_1(X_1) \mathbb{E}\left\{ \Psi\left(\frac{\varepsilon_1 + m_1(X_1) - m(X_1)}{\sigma} \right) \mid X_1 \right\} \right] \\
&\leq \mathbb{E}\left[w_1(X_1)\, \mathbb{E}\left\{ \Psi\left(\frac{\varepsilon_1}{\sigma} \right) \mid X_1 \right\} \right] = \mathbb{E}\left[\Psi\left(\frac{\varepsilon_1}{\sigma} \right) \right] \mathbb{E}[w_1(X_1)],
\end{aligned} \tag{8}$$

where the last equality follows since ε_j and X_j are independent. Analogously, we can show that, for any $\sigma > 0$,

$$\mathbb{E}\left[\Psi\left(\frac{\varepsilon_{20}}{\sigma} \right) w_2(X_2) \right] \geq \mathbb{E}\left[\Psi\left(\frac{\varepsilon_2}{\sigma} \right) \right] \mathbb{E}[w_2(X_2)]. \tag{9}$$

Under the null hypothesis H_0, the inequalities in (8) and (9) are actually equalities. However, under the alternative hypothesis, either (8), (9) or both inequalities must be strict when Ψ is strictly increasing and $\mathbb{P}(X_j \in \mathcal{A} \cap \mathcal{S}_j) > 0$. More generally, if $\mathbb{E}[\Psi((\varepsilon_j - a)/\sigma)] < \mathbb{E}[\Psi(\varepsilon_j/\sigma)]$ for any $a > 0$ and $\mathbb{E}[\Psi((\varepsilon_j - a)/\sigma)] > \mathbb{E}[\Psi(\varepsilon_j/\sigma)]$ for any $a < 0$, then we also have strict inequalities under H_1. This holds, for instance, whenever Ψ is a nondecreasing function, strictly increasing in a neighbourhood of 0 and the errors assign positive mass to that neighbourhood. Besides, if $\mathbb{E}[\Psi(U_j)] = \mathbb{E}[\Psi(\varepsilon_j/\sigma_j)] = 0$, for

$j = 1, 2$, which happens, for instance, when Ψ is an odd function and the errors have a symmetric distribution, as is usual when considering score functions in regression models, we get the following chain of equalities and inequalities

$$\mathbb{E}\left[\Psi\left(\frac{\varepsilon_{10}}{\sigma_1}\right) w_1(X_1)\right] \leq \mathbb{E}\left[\Psi\left(\frac{\varepsilon_1}{\sigma_1}\right)\right] \mathbb{E}[w_1(X_1)] = 0 = \mathbb{E}\left[\Psi\left(\frac{\varepsilon_2}{\sigma_2}\right)\right] \mathbb{E}[w_2(X_2)] \leq \mathbb{E}\left[\Psi\left(\frac{\varepsilon_{20}}{\sigma_2}\right) w_2(X_2)\right],$$

where under the null hypothesis all are equalities, but under the alternative one or both inequalities are strict. Therefore, to distinguish H_1 from H_0 it seems reasonable to compare $\mathbb{E}[\Psi(\varepsilon_{10}/\sigma_1)\, w_1(X_1)]$ and $\mathbb{E}[\Psi(\varepsilon_{20}/\sigma_2)\, w_2(X_2)]$.

It is clear that to perform the test, consistent estimators of m and σ_j, as those described in Section 2, are needed. Given independent observations $\{(X_{ij}, Y_{ij})^{\mathrm{T}}, i = 1, \ldots, n_j\}, j = 1, 2$, such that $(X_{ij}, Y_{ij})^{\mathrm{T}} \sim (X_j, Y_j)^{\mathrm{T}}$, denote $n = n_1 + n_2$ and let $\widehat{m}_j(x)$ be the robust estimator of $m_j(x)$ given in (6). For a given $x \in \mathcal{R}$, the estimator of the common regression function under the null hypothesis is defined as

$$\widehat{m}(x) = p_1(x)\widehat{m}_1(x) + p_2(x)\widehat{m}_2(x),$$

where $0 \leq p_1(x) \leq 1$ is a given function and $p_2(x) = 1 - p_1(x)$. The test statistic to be considered is

$$T = \left(\frac{n_1 n_2}{n}\right)^{1/2} (\widehat{E}_{20} - \widehat{E}_{10}) = \left(\frac{n_1 n_2}{n}\right)^{1/2} \widehat{E}_0, \tag{10}$$

where

$$\widehat{E}_{j0} = \frac{1}{n_j} \sum_{i=1}^{n_j} \Psi\left(\frac{Y_{ij} - \widehat{m}(X_{ij})}{\widehat{\sigma}_j}\right) w_j(X_{ij}).$$

Note that if $\widehat{\sigma}_j \xrightarrow{p} \sigma_j$ and $\widehat{m}_j$ is uniformly consistent over $\mathcal{S}_\ell$, for $\ell = 1, 2$, then $\widehat{E}_0 \xrightarrow{p} E_0$, where

$$E_0 = \mathbb{E}\left[\Psi\left(\frac{\varepsilon_{20}}{\sigma_2}\right) w_2(X_2)\right] - \mathbb{E}\left[\Psi\left(\frac{\varepsilon_{10}}{\sigma_1}\right) w_1(X_1)\right].$$

Hence, the test will be consistent if $\mathbb{E}[\Psi(U_j)] = 0$ and if, for instance, Ψ is nondecreasing and strictly increasing in a neighbourhood $\mathcal{V}$ of 0 (as is the case of the Huber's score function) and the errors assign positive mass to $\mathcal{V}$. Besides the Huber's score function $\Psi(t) = \min(k, \max(-k, t))$, other possible choices for Ψ are $\Psi(t) = t/\sqrt{1 + t^2/k^2}$ which is a smooth approximation of the Huber function and $\Psi(t) = k \arctan(t/k)$.

The null hypothesis will be rejected for large positive values of the test statistic T. To perform the test for a given significance level, critical values obtained from the (asymptotic) null distribution of T are needed. For that reason, in the sequel, we will analyse the asymptotic distribution of the test statistic. The following assumptions are needed:

A1 $\Psi : \mathbb{R} \to \mathbb{R}$ is a bounded and nondecreasing function. Furthermore, Ψ is twice continuously differentiable with bounded derivatives. Its first and second derivatives, Ψ' and Ψ'', are such that $\nu_j = \mathbb{E}[\Psi'(U_j)] > 0$, for $j = 1, 2$, and $\zeta_1(u) = u\Psi'(u)$ and $\zeta_2(u) = u\Psi''(u)$ are bounded.

A2 For $j = 1, 2$, $\Psi_j : \mathbb{R} \to \mathbb{R}$ are bounded and twice continuously differentiable functions, with bounded derivatives. Besides, the first and second derivatives, Ψ_j' and Ψ_j'', are such that $\nu_{j,j} = \mathbb{E}[\Psi_j'(U_j)] \neq 0$, and $\zeta_{1,j}(u) = u\Psi_j'(u)$ and $\zeta_{2,j}(u) = u\Psi_j''(u)$ are bounded.

A3 For $j = 1, 2$, $w_j : \mathbb{R} \to \mathbb{R}$ are bounded non-negative continuous weight functions with compact support $\mathcal{S}_j \subset \mathring{\mathcal{R}}$ such that $\mathcal{A} \cap \mathcal{S}_j \neq \emptyset$. The function $p_1(x)$ is continuous in a neighbourhood of $\mathcal{S}_j$.

A4 $\mathbb{E}[\Psi_1(aU_1)] = \mathbb{E}[\Psi_2(aU_2)] = \mathbb{E}[\Psi(aU_j)] = 0$ for any $a > 0$ and $j = 1, 2$.

A5 For $j = 1, 2$, the regression function m_j is twice continuously differentiable in a neighbourhood of the support, $\mathcal{R}$, of the density of X_j.

A6 For $j = 1, 2$, the random variable X_j has a density f_j twice continuously differentiable in a neighbourhood of the support $\mathcal{S}_\ell$ of w_ℓ, for $\ell = 1, 2$, and such that $i(f_j) = \inf_{x \in \mathcal{S}_j} f_j(x) > 0$ and $\inf_{t \in \mathcal{S}_j} f_{3-j}(t) > 0$.

A7 The kernel $K : \mathbb{R} \to \mathbb{R}$ is an even, bounded and Lipschitz continuous function with bounded support, say $[-1, 1]$ and such that $\int K(u)du = 1$.

A8 The sample sizes are such that $n_j/n \to \kappa_j$ with $0 < \kappa_j < 1$ as $n = n_1 + n_2 \to \infty$.

A9 The bandwidth sequence is such that $h_n \to 0$, $nh_n/\log n \to \infty$, $\sqrt{nh_n^2}/\log n \to \infty$, $nh^4 \to 0$ as $n \to \infty$.

Assumptions A3 and A5 to A9 are standard conditions in the nonparametric literature, especially when dealing with testing problems. On the other hand, A1 and A2 are usual requirements in a robust setting. In particular, the condition $\nu_j > 0$ in assumption A1 ensures that we get order $n^{1/2}$ for the test statistic. Assumption A4 is a standard assumption to avoid requiring a root-n order of convergence to scale estimators. It holds, for instance, when Ψ_j, $j = 1, 2$, and Ψ are odd functions and the errors U_j have a distribution G_j symmetric around 0. Further comments regarding this assumption are included in the following remark.

Remark 1. In the classical setting, the target is to make inferences on the conditional mean $\mathbb{E}(Y_j|X_j = x)$ and this quantity is obtained by choosing $\Psi_j(t) = t$ in (5). Hence, A4 reduces to the usual requirement that the errors have zero mean. To avoid moment conditions, the practitioner may choose, for instance, $\Psi_j(t) = \text{sgn}(t)$. In this case, inferences are made on the conditional median and A4 means that the error medians are 0. For general score functions Ψ_j, the target is to decide whether the solutions $r_j(x)$ of $\lambda_j(x, a, \sigma_j) = 0$ satisfy H_0 or H_1. When Ψ_j is a strictly increasing function, $r_j(x)$ is the so-called robust conditional location functional as introduced in Boente and Fraiman (1989), who noted that this functional provides a natural extension of the conditional expectation.

Assumption A4 implies that for $j = 1, 2$

$$\mathbb{E}[\Psi_j(U_j)] = 0 \quad \text{and} \quad \mathbb{E}[\Psi(U_j)] = 0. \tag{11}$$

The first equation in (11) means that we have centred the errors with respect to the $M-$location functional related to Ψ_j as defined in Maronna et al. (2006) and ensures that $r_j = m_j$. This property is usually known as Fisher-consistency and guarantees that the target functionals to be compared are the quantities of interest, in our case, the regression functions m_j in model (1). On the other hand, the condition $\mathbb{E}[\Psi(U_j)] = 0$ means that the $M-$location functional related to Ψ also equals 0 and entails that the test based on the statistic T defined in (10) leads to a consistent test for $H_0 : m_1 = m_2$.

To see when (11) holds, we distinguish two situations depending on the symmetry of the error distributions:

- *Symmetric error distributions.* Assume that (Y_j, X_j) satisfies the nonparametric functional regression model (1) and the distribution G_j of U_j is symmetric around 0. As mentioned above, (11) holds for any choice of odd functions Ψ_j and Ψ implying that all robust location conditional estimators are estimating the same quantity, i.e., $r_j = m_j$.
- *Asymmetric error distributions.* When the errors have asymmetric distributions, the situation is somewhat different. As an illustration, assume that W_j has a log-Gamma distribution, that is, $V_j = \exp(W_j) \sim \Gamma(\beta_j, \beta_j)$, where we have used the mean parametrization, i.e., $\mathbb{E}(V_j) = \beta_j$ and $\textsc{Var}(V_j) = \beta_j$. Then, $\mathbb{E}(W_j) = \mu_j = -\log(\beta_j) + \Gamma'(\beta_j)/\Gamma(\beta_j)$, while the median of W_j is $\widetilde{\mu}_j = -\log(\beta_j) + \log(a_j)$, with a_j the median of a $\Gamma(\beta_j, 1)$ distribution. In this asymmetric situation, the classical estimators implicitly consider the model $Y_j = m_j(X_j) + \sigma_j U_j$, where $U_j = W_j - \mu_j$. On the other hand, if the sign function is chosen as score function, the robust location functional r_j is given by $r_j(x) = m_j(x) + c_j$ with $c_j = \sigma_j(\widetilde{\mu}_j - \mu_j)$, so the above model may be written as $Y_j = r_j(X_j) + \sigma_j \widetilde{U}_j$ where now $\widetilde{U}_j = W_j - \widetilde{\mu}_j$ to ensure that the errors $\widetilde{U}_j$ satisfy (11). Note that the same score functions $\Psi_1 = \Psi_2 = \Psi$, the same shape parameters and the same scale $\sigma_1 = \sigma_2$ need to be considered to guarantee that $c_1 = c_2$, so that comparisons between populations are made on the functions of interest. The same arguments apply to other distributions. If no assumption on symmetry is made, an additional assumption of identical errors distribution need to be made to ensure that the difference functions $r_1 - m_1$ and $r_2 - m_2$ are equal and constant, when $\Psi_1 = \Psi_2$. Besides if in addition $\Psi_j = \Psi$, as mentioned above, the model also assumes through (11) that the errors are centred with respect to the $M-$location functional related to Ψ, so that $r_j = m_j$ and the test statistic defined in (10) will still lead to a consistent test for $H_0 : m_1 = m_2$.

For the sake of simplicity, we are assuming that the same bandwidth is used when estimating m_1 and m_2. Similar results can be obtained when different bandwidths are considered as far as both satisfy A9.

The next theorem gives the asymptotic distribution of the test statistic under the null hypothesis and under local alternatives.

Theorem 1. *Assume that* (1) *and* A1 *to* A9 *hold. Let $\widehat{\sigma}_j$ be a consistent estimator of σ_j, $j = 1, 2$. Then,*

(a) *Under $H_0 : m_1 = m_2$, we have that $T \xrightarrow{D} N(0, \sigma_T^2)$ where $\sigma_T^2 = \kappa_1 \tau_2^2 + \kappa_2 \tau_1^2$ with*

$$\tau_j^2 = \mathbb{E}\left[\left(\Psi_j\left(U_j\right)\frac{p_j(X_j)}{f_j(X_j)}\left\{\frac{\sigma_j \nu_{3-j}}{\sigma_{3-j}\nu_{j,j}} w_{3-j}(X_j) f_{3-j}(X_j) - \frac{\nu_j}{\nu_{j,j}} w_j(X_j) f_j(X_j)\right\} + \Psi\left(U_j\right) w_j(X_j)\right)^2\right]. \tag{12}$$

(b) *Under H_1, $T \xrightarrow{p} \infty$.*

(c) *Let $\Delta : \mathbb{R} \to \mathbb{R}$ be such that $\Delta \geq 0$ for all $x \in \mathcal{R}$. Then, under $H_{1n} : m_2(x) = m_1(x) + n^{-1/2}\Delta(x)$, we have that $T \xrightarrow{D} N(c, \sigma_T^2)$ where*

$$c = (\kappa_1\kappa_2)^{1/2}\frac{\nu_2}{\sigma_2}\mathbb{E}\left[\Delta(X_2) w_2(X_2) p_1(X_2)\right] + (\kappa_1\kappa_2)^{1/2}\frac{\nu_1}{\sigma_1}\mathbb{E}\left[\Delta(X_1) w_1(X_1) p_2(X_1)\right].$$

Remark 2. Theorem 1 entails that the asymptotic null distribution of the test statistic is a Normal random variable whose variance depends on unknown quantities. In order to apply the test in practice, a consistent estimator of σ_T^2, say $\widehat{\sigma}_T^2$, is required. Once the estimator is available, a test with asymptotic significance level α can be obtained by rejecting the null hypothesis when the observed value of the test statistic T given in (10) exceeds the critical value $z_{1-\alpha}\widehat{\sigma}_T$, where $z_{1-\alpha}$ is the $(1 - \alpha)$-quantile of a standard Normal.

A consistent estimator of σ_T^2 can easily be constructed as

$$\widehat{\sigma}_T^2 = \widehat{\kappa}_1 \widehat{\tau}_1^2 + \widehat{\kappa}_2 \widehat{\tau}_2^2,$$

where, for $j = 1, 2$, $\widehat{\kappa}_j = n_j/n$ and

$$\widehat{\tau}_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} \left[\Psi_j(\widehat{U}_{ij}) \frac{p_j(X_{ij})}{\widehat{f}_j(X_{ij})} \left\{ \frac{\widehat{\sigma}_j \widehat{\nu}_{3-j}}{\widehat{\sigma}_{3-j} \widehat{\nu}_{j,j}} w_{3-j}(X_{ij}) \widehat{f}_{3-j}(X_{ij}) - \frac{\widehat{\nu}_j}{\widehat{\nu}_{j,j}} w_j(X_{ij}) \widehat{f}_j(X_{ij}) \right\} + \Psi(\widehat{U}_{ij}) w_j(X_{ij}) \right]^2,$$

with

$$\widehat{U}_{ij} = \frac{Y_{ij} - \widehat{m}_j(X_{ij})}{\widehat{\sigma}_j}, \qquad \widehat{\nu}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \Psi'(\widehat{U}_{ij}), \qquad \widehat{\nu}_{j,j} = \frac{1}{n_j} \sum_{i=1}^{n_j} \Psi_j'(\widehat{U}_{ij}) \quad \text{and} \quad \widehat{f}_j(x) = \frac{1}{n_j} \sum_{i=1}^{n_j} K_h(x - X_{ij}).$$

Remark 3. Let $\widetilde{f}_j(x) = (\sigma_j \nu_{3-j})(\sigma_{3-j} \nu_j)^{-1} p_j(x) w_{3-j}(x) f_{3-j}(x) + p_{3-j}(x) w_j(x) f_j(x)$. In the particular case $\Psi_1 = \Psi_2 = \Psi$, we have that $\nu_{j,j} = \nu_j$ and therefore the terms $\tau_j^2, j = 1, 2$, that appear in the variance of the asymptotic distribution of the test statistic reduce to $\tau_j^2 = \mathbb{E}\left[\Psi^2(U_j)\right] \mathbb{E}\left[\widetilde{f}_j^2(X_j)/f_j^2(X_j)\right]$.

Remark 4. Statement (c) in Theorem 1 gives the asymptotic distribution of the test statistic under local alternatives and shows that the test detects local alternatives converging to the null hypothesis at the parametric rate $n^{-1/2}$ whenever $\mathbb{E}\left[\Delta(X_2) w_2(X_2) p_1(X_2)\right] > 0$ or $\mathbb{E}\left[\Delta(X_1) w_1(X_1) p_2(X_1)\right] > 0$.

Remark 5. It is worth noticing that the procedure introduced in this paper may be extended to deal with heteroscedasticity, by defining $\widehat{E}_{j0}$ as $n_j^{-1} \sum_{i=1}^{n_j} \Psi\left((Y_{ij} - \widehat{m}(X_{ij}))/\widehat{\sigma}_j(X_{ij})\right) w_j(X_{ij})$, where $\widehat{\sigma}_j(x)$ stands for a robust estimator of the conditional scale function $\sigma_j(x)$. However, to derive the asymptotic behaviour of the corresponding test statistic, which allows to define the critical values, additional assumptions including a uniform Bahadur expansion for $\widehat{\sigma}_j(x)$, as that given for $\widehat{m}_j$ in (A.7), will be needed. We leave this important and challenging problem for future research.

4. Monte Carlo study

In this section, we present the results of a simulation study devoted to illustrate the finite-sample performance of the testing procedure described in Section 3 and to compare its behaviour with that of the test defined in Neumeyer and Pardo-Fernández (2009) and the covariate-matched Wilcoxon–Mann–Whitney statistic given in Koul and Schick (1997). More specifically, for the robust procedure, we use the approximation of the critical values given in Remark 2. In order to make the comparison fair, in the case of the method by Neumeyer and Pardo-Fernández (2009) we restrict ourselves to the homoscedastic case, so we estimate the asymptotic variance of their test statistic under the assumption of constant variance (see their Theorem 1). Tables and figures report the observed frequency of rejections among 1000 simulated data sets with significance level 0.05. We set $m_1(x) = x$ as the regression function in the first population and consider two possibilities for the second population, $m_2(x) = m_1(x) + \Delta_n$ and $m_2(x) = m_1(x) + \Delta_n(\sin(2\pi x) + 1)$. We choose $w_1 = w_2 = \mathbb{I}_{(0,1)}$ and $p_1(x) = 0.5$. From now on, T_{CL} stands for the test proposed in Neumeyer and Pardo-Fernández (2009), while $W^\star$ denotes the covariate-matched Wilcoxon–Mann–Whitney statistic defined in Koul and Schick (1997), where we will also use $W_h^\star$ when indicating the bandwidth h used in its computation.

As mentioned above, the conducted numerical studies aim to compare the performance of the testing procedure described in Section 3 with the testing procedures based on T_{CL} and $W_h^\star$. Several scenarios are considered. In Section 4.1, a common design distribution is considered and the aim is to analyse the performance under different contaminations to evaluate the robustness of the procedure in terms of level and power. In this case, the central model, i.e., the uncontaminated observations correspond to errors having a normal distribution. Section 4.2 summarizes the results of a simulation study conducted to evaluate the performance of the testing procedure described in Section 3, when the errors have an asymmetric distribution. As mentioned in Remark 1, to ensure that all tests are making inferences on the same objects, we will take $\Psi_1 = \Psi_2 = \Psi$ and errors ε_j with the same distribution. No outliers are introduced in this case, since the aim is to evaluate the validity of our proposal when the usual assumption in robustness of symmetric errors is violated. Finally, Section 4.3 considers a situation where different distributions for the design points and the errors are considered between populations. Again, no contamination is considered in this case, since the aim is to study the stability of the procedures under this setting. Recall that the test based on T_{CL} and that described in Section 3 do not make assumptions of a common design density or a common error distribution, while the covariate-matched Wilcoxon–Mann–Whitney statistic $W_h^\star$ assumes that $X_1 \overset{d}{=} X_2$ and $\varepsilon_1 \overset{d}{=} \varepsilon_2$.

4.1. Design points with common density

In this study, the covariates X_j are generated with uniform distribution on $\mathcal{R} = (0, 1)$. We choose $w_1 = w_2 = \mathbb{I}_{(0,1)}$ and $p_1(x) = 0.5$. The following scenarios were considered to simulate the regression errors:

- The first scenario, denoted as C_0, corresponds to the situation where $\varepsilon_j \sim N(0, \sigma_j^2)$, with $\sigma_1 = 0.5$ and $\sigma_2 = 0.75$. In this case no outliers will appear in the data.
- Next, we consider a situation, labelled as $\mathcal{T}_1$, in which $\varepsilon_j \sim \mathcal{C}(0, 25\sigma_j^2)$, where $\mathcal{C}(\mu, \sigma^2)$ stands for the Cauchy distribution with location μ and dispersion σ^2. In this case the errors have no moments.
- We also consider a situation with contaminated gross-errors, labelled as C_{1,π_1,π_2}, in which $\varepsilon_j \sim (1-\pi_j)\,N(0, \sigma_j^2) + \pi_j\,N(0, 25\sigma_j^2)$. We choose $\pi_1 = 0, 0.1$ and $\pi_2 = 0, 0.1$ and select different combinations of the contaminating probabilities, so one or both samples contain outliers.

The robust procedure involves selecting score functions both in the estimation step and when computing the test statistic, as well as choosing smoothing parameters to perform the nonparametric estimation of the regression functions. To analyse the influence of the score functions and the bandwidth choice on the level and power of the test, a preliminary study was carried on. From now on, the results corresponding to our proposal are labelled as $T_{\text{R,H,H}}$, $T_{\text{R,H,T}}$, $T_{\text{R,A,H}}$ and $T_{\text{R,A,T}}$, where the second index indicates the $\Psi-$function used, H being the Huber's function with tuning constant $k_{\text{H}} = 1.345$ and A the function $\psi_k = k\arctan(t/k)$, with $k = 0.9$, and the third index denotes the score function used in the estimation process, that is, H corresponds to $\psi_{k,\text{H}}$ with $k = k_{\text{H}} = 1.345$ while T to the bisquare Tukey's function with constant $k_{\text{T}} = 4.685$. The results corresponding to C_0, to errors with Cauchy distribution and to contaminations C_{1,π_1,π_2} are reported in Tables S.1 to S.3 included in the supplementary material file (see Appendix B). These tables reveal that the results obtained for the test statistics based on the selected bounded score functions are almost equal for all models, independently of the selected score function. Therefore, in the sequel, when considering the robust proposal introduced in this paper, we will restrict our discussion to the results based on $T_{\text{R,H,H}}$.

On the other hand, Tables S.1 to S.3 also show that the choice of the smoothing parameters required to construct the nonparametric estimators does not have a significant impact on the tests either, as the results obtained with different bandwidths are almost equal. Nevertheless, in practice, a data-driven mechanism to choose the required smoothing parameters is desirable. In the numerical studies to be described below, we perform the tests with data-driven bandwidths chosen by least-squares cross-validation for T_{CL} and by robust cross-validation for $T_{\text{R,H,H}}$, as follows. Taking into account that the classical cross-validation criterion (see, for example Härdle, 1990) tries to measure both bias and variance, Bianco and Boente (2007) and Boente and Rodríguez (2010) considered, for partly linear autoregression and partly regression models, a new measure that establishes a trade-off between robust measures of bias and variance. Let $m_j^{(-i)}(x)$ be the smoothers computed with bandwidth h using all the data except (Y_{ij}, X_{ij}) and denote $\widehat{\varepsilon}_{ij}(h) = Y_{ij} - m_j^{(-i)}(X_{ij})$. Let μ_n and σ_n denote robust estimators of location and scale, respectively. For each sample, the robust cross-validation criterion consists of choosing h as the minimizer of

$$\Upsilon_j(h) = \mu_n^2\left(\widehat{\varepsilon}_{ij}(h)\right) + \sigma_n^2\left(\widehat{\varepsilon}_{ij}(h)\right).$$

As location estimator, μ_n, we choose the median, whereas σ_n is taken as a $\tau-$scale estimator.

In order to check the consistency of the test under local alternatives converging to the null hypothesis at a parametric rate, we consider alternatives with $\Delta_n = n^{-1/2}\Delta$, where $\Delta = 0$ (null hypothesis) and $\Delta = 0.5, 2, 4, 6, 8$ (local alternatives) and sample sizes $n_1, n_2 = 50, 100$. Tables S.4 to S.7 available in the supplementary material and Figs. 1 and S.1 of the supplementary file illustrate the behaviour of the tests based on T_{CL} and $T_{\text{R,H,H}}$ in terms of level approximation and power (see Appendix B). All figures depict the results under the central model C_0 in order to have a common reference to study the effect of introducing distributions with no moments or contaminated data. To analyse the level sensitivity of the procedure, we considered an additional contamination model denoted $C_{2,c}$ in which just one observation is modified as follows. We first simulate data as in scenario C_0 and we order the covariates of the first population as $X_{(1),1} \leq \cdots \leq X_{(n_1),1}$. Denote as $(X_{(1),1}, Y_{D_1,1})^{\text{T}}, \ldots, (X_{(n_1),1}, Y_{D_{n_1},1})^{\text{T}}$ the sample of observations ordered according to the values of the explanatory variable. Then, we modify the observation corresponding to the median of the covariates as $X_{(\frac{n_1}{2}),1} = X_{D_{\frac{n_1}{2}},1} = 0.5$ and $Y_{D_{\frac{n_1}{2}},1} = c$.

Under model C_0, both the classical test and the robust test perform almost equally, with a correct approximation of the level and power increasing as the deviation from the null hypothesis gets larger. Since root-n local alternatives are taken, the power is similar for all choices of sample sizes and shows the tests ability to detect these kinds of local alternatives. When the errors have a Cauchy distribution (model $\mathcal{T}_1$), the robust test empirical size is close to the nominal level, while the test T_{CL} provides an underestimated level. Moreover, T_{CL} presents almost no power, while, although some loss is observed with respect to C_0, the power behaviour of $T_{\text{R,H,H}}$ is correct since it is able to detect the considered alternatives. Under the scenarios with contaminated data C_{1,π_1,π_2}, both statistics approximate correctly the level when the sample sizes are equal, but only the robust one gives a correct level approximation when the samples are unbalanced. Moreover, the robust test is more powerful. Finally, scenarios $C_{2,c}$ produce a bad approximation of the level for T_{CL}, yielding to a very liberal test when $c < 0$ and a very conservative test when $c > 0$ as shown in Fig. 2. On the other hand, $T_{\text{R,H,H}}$ also presents some deviations from the nominal level, specially when the samples are unbalanced and $c > 0$, but these deviations are less serious than those of T_{CL}. In this scenario, the power behaviour of $T_{\text{R,H,H}}$ is almost the same as that obtained with normal errors.

With respect to the behaviour of the test based on the covariate-matched Wilcoxon–Mann–Whitney statistic $W^\star$, since there is no automatic way to choose the bandwidth, we report the results obtained when $h = 0.1, 0.15$ and 0.2. Tables S.4 to S.7 show that under C_0 and $\mathcal{T}_1$, the tests based on $T_{\text{R,H,H}}$ and $W^\star$ perform very similarly, as both produce a good level approximation and similar power (in terms of power, $T_{\text{R,H,H}}$ slightly outperforms $W^\star$ under C_0, and the contrary happens

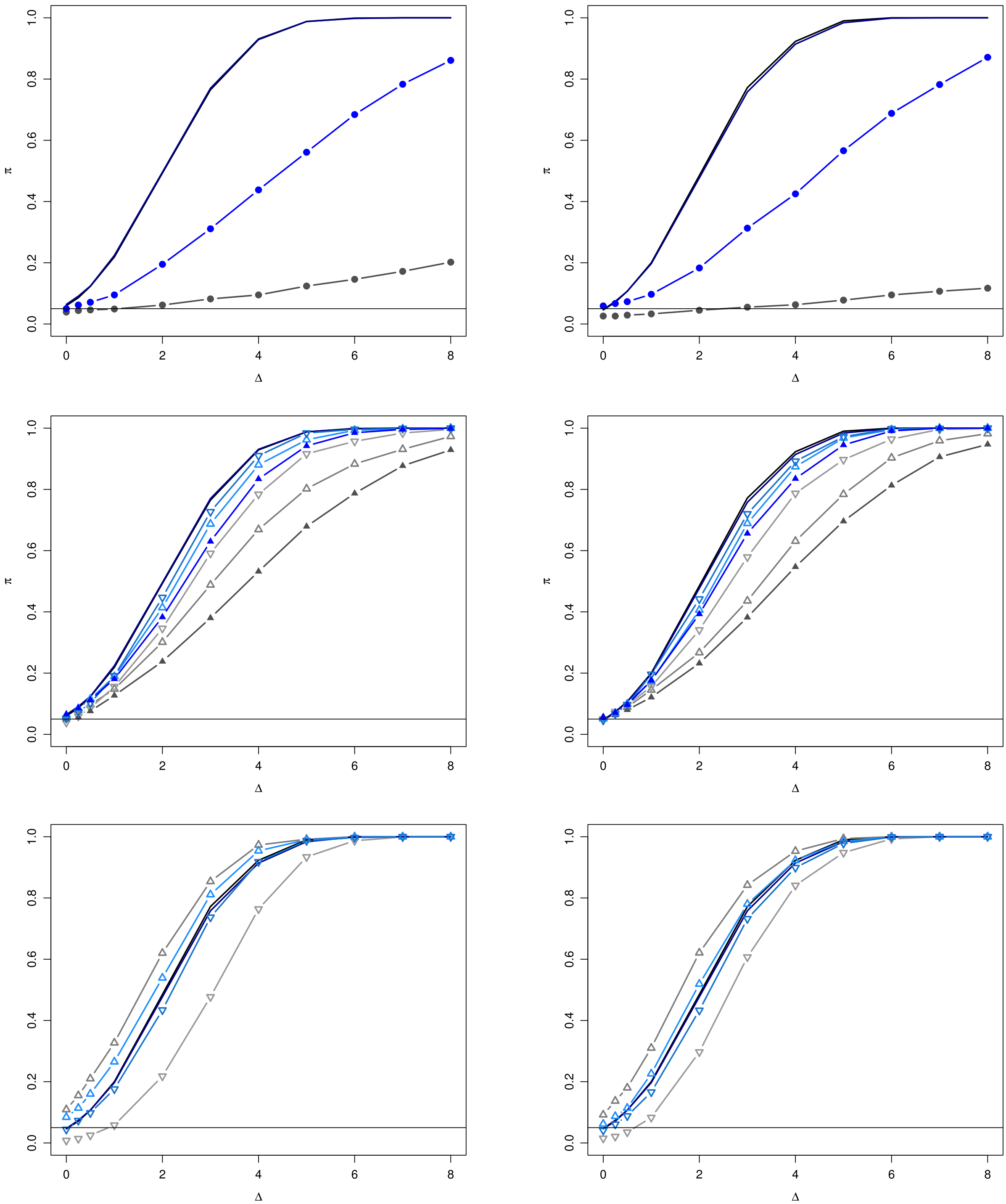

Fig. 1. Frequencies of rejection of T_{CL} (black/grey lines) and $T_{\text{R,H,H}}$ (blue/light blue lines) using the data-driven bandwidths when $n_1 = n_2 = 50$ (left) and $n_1 = n_2 = 100$ (right) and the local alternatives $m_2(x) = m_1(x) + \Delta n^{-1/2}$, with $n = n_1 + n_2$. In all cases, solid black and blue lines represent the power under C_0. Top: Black and blue lines with filled circles give the power under $\mathcal{T}_1$. Centre: Grey and light blue lines give the under C_{1,π_1,π_2}, where the triangles stand for $(\pi_1, \pi_2) = (0, 0.1)$, the inverted triangles for $(\pi_1, \pi_2) = (0.1, 0)$ and the filled ones for $(\pi_1, \pi_2) = (0.1, 0.1)$. Bottom: Grey and light blue lines give the power under $C_{2,c}$, where the triangles stand for $c = -4$, the inverted triangles for $c = 4$. The solid horizontal line indicates the 5%-level. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

under $\mathcal{T}_1$). Under scenarios C_{1,π_1,π_2}, $T_{\text{R,H,H}}$ approximates the level well, whereas $W^\star$ underestimates the level when the samples are unbalanced ($n_1 = 50, n_2 = 100$); the power of $T_{\text{R,H,H}}$ is higher than that of $W^\star$. Similar conclusions can be

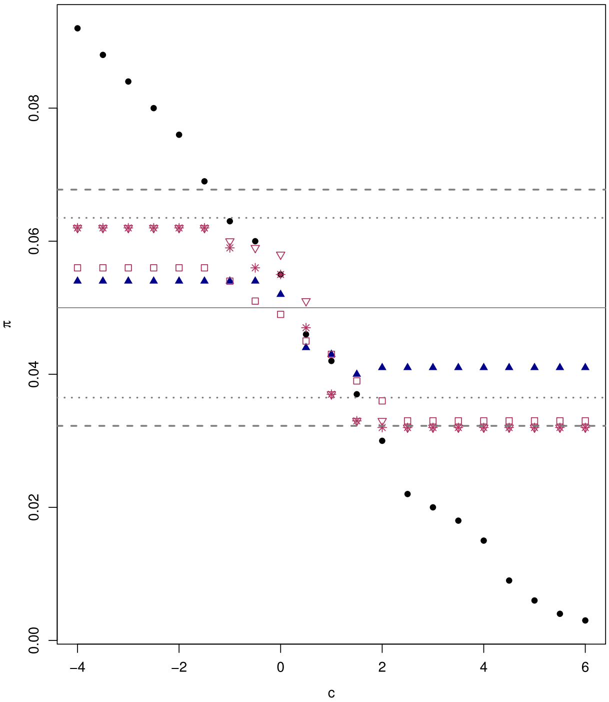

Fig. 2. Empirical size of T_{CL} (filled circles) and $T_{\text{R,H,H}}$ (blue triangles) using the data-driven bandwidths when $n_1 = n_2 = 100$, when the data are generated under $C_{2,c}$. The horizontal solid lines are the nominal level $\alpha = 0.05$ and the dotted and dashed lines represent the acceptance region for testing if the empirical size is significantly different from the nominal level, at level 0.05 and 0.01, respectively. The empirical size of the covariate-matched Wilcoxon–Mann–Whitney statistic $W_h^\star$ is plotted in maroon for different values of h (squares for $h = 0.1$, stars for $h = 0.15$ and inverted triangles for $h = 0.2$).

raised for scenario $C_{2,-4}$. Finally, under $C_{2,4}$, both test statistics tend to underestimate the level, especially with unbalanced samples. To better understand this behaviour, in Fig. 2 we show the proportion of rejections under $C_{2,c}$ for several values of c when $\Delta = 0$ and $n_1 = n_2 = 100$. It seems that the size distortion is less serious for the robust test, as its empirical size remains stable around the nominal level. We hence conclude that the proposed test behaves better than the covariate-matched Wilcoxon–Mann–Whitney test when outliers appear in the sample. Another advantage of our proposal is that it does not require a common density for the design points as does the covariate-matched statistic.

4.2. Asymmetric errors

The goal of this section is to study the performance of the test defined in this paper, when the errors have an asymmetric distribution. In the considered framework, the test statistic T defined in (10) still provides a consistent test to test $\mathrm{H}_0 : m_1 = m_2$, since $\Psi_1 = \Psi_2 = \Psi$ and (11) holds for the centred errors, as explained in Remark 1. Recall that the proof of Theorem 1 requires the stronger assumption A4 which may not hold for the centred log-Gamma errors, so we cannot ensure that the testing procedure described in Remark 2 achieves the nominal level for the Huber's score function. For that reason, this numerical study was conducted to analyse the level and power sensitivity of the test based on $T_{\text{R,H,H}}$ under asymmetric errors.

We generate covariates X_j according to a uniform distribution on $\mathcal{R} = (0, 1)$, while the errors $\varepsilon_j = \sigma_j U_j$ follow a log-Gamma distribution, that is, $V_j = \exp(U_j) \sim \Gamma(\beta_j, \beta_j)$, where for any $\beta > 0$ and $\mu > 0$, we denote by $\Gamma(\beta, \mu)$ the parametrization of the Gamma distribution given by the density $f(v, \beta, \mu) = \beta^\beta \; v^{\beta-1} \; \exp(-\beta\, v/\mu)\{\mu^\beta\, \Gamma(\beta)\}^{-1} I_{v\geq 0}$. Note that, if $V \sim \Gamma(\beta, \mu)$, we have that $\mathbb{E}(V) = \mu$ and $\text{Var}(V) = \mu^2/\beta$, where β is a shape parameter.

We choose $\beta_1 = \beta_2 = \beta = 3$ as well as $\sigma_1 = \sigma_2 = \sigma = 1$. As mentioned in Remark 1, the main reason for taking equal values for β_j and σ_j is to guarantee that we are still testing $m_1 = m_2$ against $m_1 \leq m_2$. The fact that the errors have an asymmetric distribution introduces a shift in the functions solution of (5). For instance, in the classical situation, $\mathbb{E}(Y_j|X_j = x) = m_j(x) + \sigma\, \mathbb{E}(U_j)$, hence the functions to be compared are $r_j(x) = m_j(x) + \sigma \mathbb{E}(U_j)$, where $\mathbb{E}(U_j) = -\log(\tau) + d(\tau)$ with $d(z) = \Gamma'(z)/\Gamma(z)$ being the digamma function. For $\tau = 3$, $d(\tau) \simeq 0.923$ meaning that $r_j(x) \simeq m_j(x) - 0.176\,\sigma$. On the other hand, the $M-$location functional related to the Huber's score function with tuning constant $k_{\text{H}} = 1.345$ is $\mu_{\text{H}} \simeq -0.143$, so the centred errors satisfying (11) are $U_j - \mu_{\text{H}}$ or equivalently the robust conditional location functional solution of (5) is given by $r_j(x) \simeq m_j(x) - 0.143\,\sigma$.

We only report the results when $m_2(x) = m_1(x) + \Delta n^{-1/2}$, since similar ones are obtained using the local alternatives $m_2(x) = m_1(x) + \Delta n^{-1/2}(\sin(2\pi x) + 1)$. The sample sizes considered are $n_1 = n_2 = 50$ and $n_1 = n_2 = 100$. We also compare

Table 1
Frequencies of rejection under the null hypothesis and local alternatives of T_{CL}, $T_{R,H,H}$ and $W^\star$ when the errors have a log-Gamma distribution and $m_2(x) = m_1(x) + \Delta n^{-1/2}$.

h		$n_1 = n_2 = 50$						$n_1 = n_2 = 100$					
	Δ:	0	0.5	2	4	6	8	0	0.5	2	4	6	8
0.1	T_{CL}	0.062	0.123	0.485	0.934	0.999	1.000	0.055	0.101	0.465	0.937	1.000	1.000
	$T_{R,H,H}$	0.080	0.145	0.506	0.933	1.000	1.000	0.058	0.113	0.498	0.944	0.999	1.000
	$W^\star$	0.047	0.093	0.415	0.898	0.996	1.000	0.050	0.090	0.425	0.932	0.999	1.000
0.15	T_{CL}	0.052	0.113	0.495	0.941	0.999	1.000	0.056	0.102	0.467	0.939	1.000	1.000
	$T_{R,H,H}$	0.070	0.132	0.501	0.939	1.000	1.000	0.060	0.108	0.490	0.946	0.999	1.000
	$W^\star$	0.047	0.099	0.446	0.909	0.999	1.000	0.053	0.098	0.448	0.937	0.999	1.000
0.2	T_{CL}	0.055	0.110	0.500	0.943	0.999	1.000	0.057	0.101	0.474	0.939	1.000	1.000
	$T_{R,H,H}$	0.063	0.130	0.501	0.939	1.000	1.000	0.060	0.108	0.488	0.944	0.999	1.000
	$W^\star$	0.046	0.104	0.462	0.923	0.998	1.000	0.054	0.099	0.455	0.940	1.000	1.000
0.25	T_{CL}	0.054	0.106	0.505	0.944	0.999	1.000	0.056	0.101	0.475	0.939	1.000	1.000
	$T_{R,H,H}$	0.067	0.126	0.497	0.937	0.999	1.000	0.061	0.107	0.485	0.944	1.000	1.000
	$W^\star$	0.047	0.105	0.460	0.926	0.998	1.000	0.056	0.102	0.463	0.938	0.999	1.000
0.3	T_{CL}	0.055	0.107	0.507	0.946	0.999	1.000	0.057	0.103	0.477	0.940	1.000	1.000
	$T_{R,H,H}$	0.055	0.125	0.488	0.935	0.999	1.000	0.059	0.106	0.478	0.947	1.000	1.000
	$W^\star$	0.048	0.110	0.469	0.928	0.998	1.000	0.057	0.098	0.466	0.938	0.999	1.000
0.35	T_{CL}	0.054	0.109	0.506	0.946	0.999	1.000	0.055	0.105	0.477	0.942	1.000	1.000
	$T_{R,H,H}$	0.053	0.123	0.486	0.936	0.998	1.000	0.059	0.106	0.475	0.946	1.000	1.000
	$W^\star$	0.046	0.112	0.472	0.927	0.998	1.000	0.056	0.100	0.464	0.938	0.999	1.000
0.4	T_{CL}	0.056	0.112	0.506	0.944	0.999	1.000	0.056	0.104	0.474	0.942	1.000	1.000
	$T_{R,H,H}$	0.051	0.125	0.486	0.936	0.998	1.000	0.059	0.106	0.474	0.944	1.000	1.000
	$W^\star$	0.044	0.115	0.476	0.929	0.998	1.000	0.055	0.098	0.458	0.937	0.999	1.000
0.5	T_{CL}	0.057	0.114	0.498	0.941	0.999	1.000	0.055	0.104	0.467	0.942	1.000	1.000
	$T_{R,H,H}$	0.051	0.131	0.479	0.936	0.997	1.000	0.056	0.108	0.473	0.939	1.000	1.000
	$W^\star$	0.043	0.118	0.479	0.932	0.998	1.000	0.056	0.099	0.463	0.936	0.999	1.000
h_{CV}	T_{CL}	0.052	0.113	0.503	0.934	0.999	1.000	0.053	0.104	0.467	0.940	1.000	1.000
	$T_{R,H,H}$	0.061	0.130	0.498	0.935	0.998	1.000	0.058	0.107	0.478	0.944	1.000	1.000

our procedure with the covariate-matched Wilcoxon–Mann–Whitney statistic $W^\star$ defined in Koul and Schick (1997). We choose different smoothing parameters varying from 0.1 to 0.5 to compute $W_h^\star$. For fair comparisons, we report the observed frequencies of rejection of T_{CL} and $T_{R,H,H}$ using the same bandwidth parameters and also the results obtained using least-squares cross-validation bandwidth for T_{CL} and the robust cross-validation smoothing parameter for $T_{R,H,H}$, denoted by h_{CV}. Table 1 reports the obtained frequencies of rejection. All procedures lead to a similar power behaviour. It is worth noting that in most situations the covariate-matched Wilcoxon–Mann–Whitney statistic leads to an empirical size closer to the nominal one, although the differences obtained with T_{CL} and $T_{R,H,H}$ are well within the Monte Carlo margin of error. In particular, for $W^\star$ and T_{CL} the smallest bandwidth 0.1 leads to the best empirical size, while for the robust procedure $T_{R,H,H}$ a larger bandwidth seems preferable. On the other hand, the cross-validation choice seems to affect more the level performance of $T_{R,H,H}$ than that of T_{CL}, which suggests that in this situation a robust cross-validation procedure based on a robustified deviance may be a better choice.

4.3. Design points and errors with different distribution

As mentioned above, the aim of this section is to compare the performance of the tests based on T_{CL} and $T_{R,H,H}$ when different distributions for the design points and the errors are considered between populations. In this study, the covariates X_j are generated from Beta distributions on $\mathcal{R} = (0, 1)$, $X_j \sim Be(\beta_{j1}, \beta_{j2})$, $j = 1, 2$ (in particular, we also consider the uniform distribution, which is obtained when $\beta_{11} = \beta_{12} = 1$). The following scenarios were considered to simulate the regression errors:

- The first scenario, denoted $\mathcal{S}_1$, corresponds to the situation where $\varepsilon_j \sim N(0, \sigma_j^2)$, with $\sigma_1 = \sigma_2 = 0.5$.
- In the second scenario, denoted $\mathcal{S}_2$, the errors also have different distributions, that is, we choose $\varepsilon_j = \sigma_j U_j$ where $U_1 \sim N(0, 1)$ and $U_2 \sim \mathcal{DE}$, where $\mathcal{DE}$ stands for the double exponential distribution with density $\exp(-|x|)/2$ and $\sigma_1 = \sigma_2 = 0.5$.

As in Section 4.2, we only report here the results obtained when $m_2(x) = m_1(x) + \Delta n^{-1/2}$, since similar results are obtained with $m_2(x) = m_1(x) + \Delta n^{-1/2}(\sin(2\pi x) + 1)$. Table 2 reports the results for $n_1 = n_2 = 50$ and $n_1 = n_2 = 100$ and different values of β_{ij}. The bandwidths were selected by using cross-validation as in Section 4.1 when estimating the regression function. The test statistics produce very similar results, both in terms of level approximation and in power. The case $X_1 \sim Be(0.5, 0.5)$ and $X_2 \sim Be(2, 2)$ gives a slight overestimation of the size.

Table 2
Frequencies of rejection under the null hypothesis and local alternatives of T_{CL} and $T_{\text{R,H,H}}$ under $\mathcal{S}_1$ and $\mathcal{S}_2$.

	$n_1 = n_2 = 50$						$n_1 = n_2 = 100$					
Δ:	0	0.5	2	4	6	8	0	0.5	2	4	6	8
	scenario $\mathcal{S}_1$ with $X_1 \sim \mathcal{U}(0, 1)$ and $X_2 \sim Be(0.5, 0.5)$											
T_{CL}	0.042	0.108	0.651	0.984	1.000	1.000	0.054	0.112	0.623	0.989	1.000	1.000
$T_{\text{R,H,H}}$	0.045	0.114	0.656	0.984	1.000	1.000	0.050	0.130	0.629	0.986	1.000	1.000
	scenario $\mathcal{S}_1$ with $X_1 \sim \mathcal{U}(0, 1)$ and $X_2 \sim Be(2, 2)$											
T_{CL}	0.044	0.122	0.625	0.984	1.000	1.000	0.048	0.115	0.606	0.992	1.000	1.000
$T_{\text{R,H,H}}$	0.050	0.126	0.616	0.981	1.000	1.000	0.054	0.120	0.609	0.994	1.000	1.000
	scenario $\mathcal{S}_1$ with $X_1 \sim Be(0.5, 0.5)$ and $X_2 \sim Be(2, 2)$											
T_{CL}	0.065	0.137	0.606	0.971	0.999	1.000	0.057	0.130	0.577	0.966	1.000	1.000
$T_{\text{R,H,H}}$	0.066	0.149	0.622	0.971	1.000	1.000	0.068	0.128	0.584	0.969	1.000	1.000
	scenario $\mathcal{S}_2$ with $X_1 \sim \mathcal{U}(0, 1)$ and $X_2 \sim Be(0.5, 0.5)$											
T_{CL}	0.052	0.108	0.452	0.921	0.998	1.000	0.048	0.108	0.484	0.930	0.999	1.000
$T_{\text{R,H,H}}$	0.048	0.116	0.507	0.944	1.000	1.000	0.051	0.125	0.534	0.957	1.000	1.000
	scenario $\mathcal{S}_2$ with $X_1 \sim \mathcal{U}(0, 1)$ and $X_2 \sim Be(2, 2)$											
T_{CL}	0.059	0.111	0.465	0.916	0.998	1.000	0.056	0.117	0.510	0.916	0.997	1.000
$T_{\text{R,H,H}}$	0.062	0.124	0.541	0.955	0.998	1.000	0.063	0.137	0.574	0.969	1.000	1.000
	scenario $\mathcal{S}_2$ with $X_1 \sim Be(0.5, 0.5)$ and $X_2 \sim Be(2, 2)$											
T_{CL}	0.067	0.130	0.489	0.911	0.997	1.000	0.066	0.115	0.461	0.906	0.996	1.000
$T_{\text{R,H,H}}$	0.069	0.143	0.551	0.944	0.999	1.000	0.065	0.115	0.521	0.941	0.999	1.000

5. A real data analysis

Neumeyer and Pardo-Fernández (2009) used a data set from the Data Archive of the *Journal of Applied Econometrics* to illustrate their testing procedure. The data are related to total expenditures of several Dutch households. Particularly, they tested for the equality of the regression curves that explain the relationship between the covariate 'log of the total expenditure' and the response 'log of the expenditure on food' according to the number of household members. The nature of the considered variables justifies the use of a one-sided type test, since it is expected that the food expenditure increases (or, at least, does not decrease) as the size of the household increases. When comparing the households of 3 members (45 observations) and 4 members (73 observations), Neumeyer and Pardo-Fernández (2009) reported a p-value 0.092.

To evaluate if the one-sided test described in Section 3 can be applied for this data set, we first performed the test described in Dette and Marchlewski (2010) to check homoscedasticity in both populations, using the identity function and the Huber's function, which leads to a more resistant procedure, to compute both the regression estimators and the test statistic. In both cases the obtained p-values were larger than 0.5 for the households of 3 and 4 members. We then applied the test procedure described in Section 3 with the Huber's score function, with tuning constant $k_{\text{H}} = 1.345$, both to estimate and to compute the test statistic, $T_{\text{R,H,H}}$, as well as the test statistic described in Neumeyer and Pardo-Fernández (2009) assuming homoscedasticity, T_{CL}. The obtained p-values are 0.125 and 0.102 for T_{CL} and $T_{\text{R,H,H}}$, respectively.

Our purpose here is to illustrate the effect of including an outlier in the data set, in a similar manner as we did under scenario $C_{2,c}$ in our simulation study. We artificially add an observation of the form $(10.74, c)$ to the first sample and then perform the tests with bandwidths chosen by cross-validation as described in Section 4. The value 10.47 corresponds to the sample median of the first population covariate. The obtained p-values of the tests based on T_{CL} and $T_{\text{R,H,H}}$ are reported in Fig. 3 for values of c ranging between 6 and 10. We can observe that the p-values of the classical test present a great variation depending on the value of the contamination c, even leading to a rejection of the null hypothesis when $c \in [6, 7]$. On the other hand, the robust test produces more stable p-values, all of them above 0.05. Figs. S.2 and S.3 in the supplementary material lead to the same conclusions when fixed bandwidths are considered (see Appendix B). Moreover, these figures show that, except for $c = 9.5$, which is a very extreme contamination, the p-values of the robust test are very stable independently of the bandwidth choice, with values always between 0.05 and 0.10.

Finally, we have also considered the situation in which both populations are contaminated in a similar way as described above. Fig. 4 illustrates the p-values behaviour when adding a contaminating observation at the sample median covariate with response taking the value c_1 in the first population and c_2 in the second one. The contaminating values c_j vary on a grid of points between 6 and 11 with a step of 0.25, since the response should be smaller or equal to the covariate, 'log of the total expenditure', whose maximum is 11.5. The bandwidths were selected using cross-validation. Under this contamination, the p-values of the classical test vary in the range $[0.033, 0.659]$ showing its sensitivity and leading to different conclusions depending on the contamination value. Indeed, in a subset of the region $\mathcal{R} = \{(c_1, c_2)$ such that $6 \le c_1 \le 7.75$ and $8.5 \le c_2 \le 11\}$, the p-values of the test based on T_{CL} leads to rejection at 5% level, so the conclusion with respect to the clean data set is reversed. On the other hand, the test based on $T_{\text{R,H,H}}$ is more stable and does not reject the null hypothesis for the considered level.

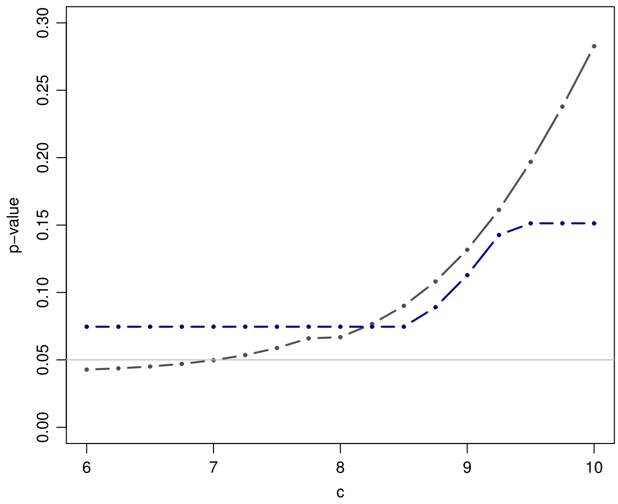

Fig. 3. Illustration on a real data set. p-values of the tests T_{CL} (grey) and $T_{R,H,H}$ (blue) for different values of the contamination c. The solid horizontal line indicates the 5%-level. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

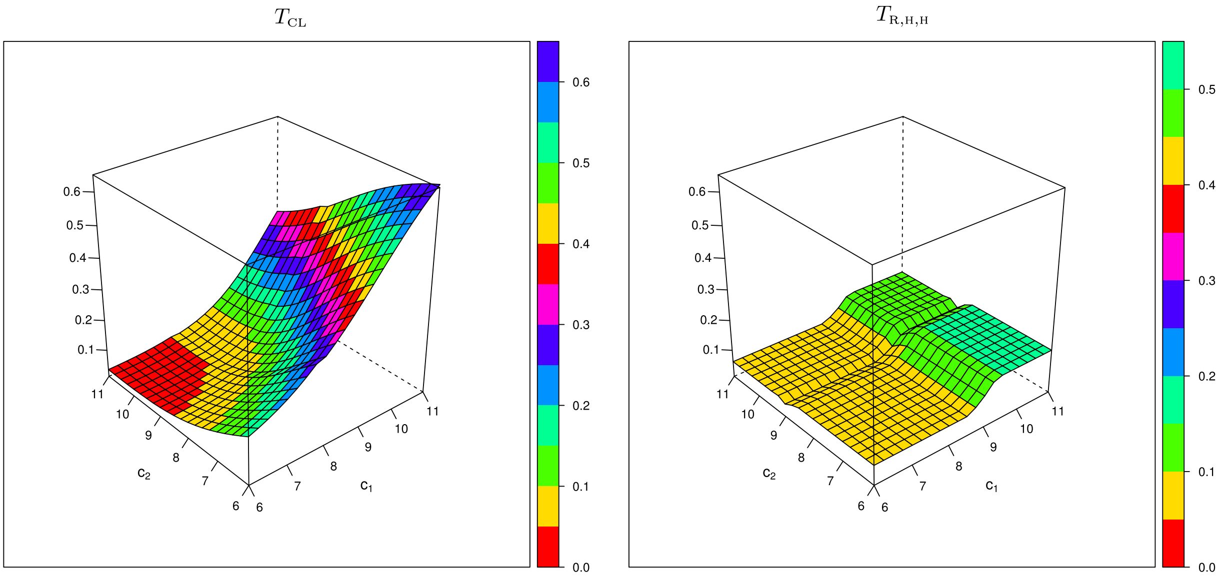

Fig. 4. Illustration on a real data set. Surface plot of the p-values of the tests T_{CL} (left) and $T_{R,H,H}$ (right) as a function of the contamination values c_1 and c_2, when both populations are contaminated and cross-validation bandwidths are considered.

6. Conclusion

In this paper we have studied a new robust method to test for the equality of two regression curves versus a one-sided alternative in a nonparametric setup. The new procedure adapts the ideas in Neumeyer and Pardo-Fernández (2009) to the situation where no moments are assumed for the regression errors. The analysis of the asymptotic distribution of the test statistic reveals that the testing procedure is consistent against local alternatives converging to the null hypothesis at the parametric rate $n^{-1/2}$. Simulations have shown a good practical behaviour of the new test when the critical values are obtained from an approximation of the asymptotic null distribution of the test statistic. If no outliers are present in the sample, the behaviour of the new test is almost equal to that of the classical method, but when outliers appear in the samples, the robust test clearly outperforms the classical procedure. The robust procedure introduced does not assume that the design points have the same density. Besides, when the errors of both populations have a symmetric distribution, it does not require a common distribution for the errors. Finally, the procedure still leads to a consistent test under asymmetric errors if $\Psi_1 = \Psi_2 = \Psi$ and the errors ε_j have the same distribution.

Acknowledgements

The authors wish to thank the Associate Editor and two anonymous referees for valuable comments which led to an improved version of the original paper. This research was partially supported by Grants PIP 112-201101-00339 from CONICET, PICT 2014-0351 from ANPCYT and 20020130100279BA from the Universidad de Buenos Aires at Buenos Aires, Argentina (G. Boente) and Grant MTM2014-55966-P of the Spanish Ministry of Economy and Competitiveness, Spain (J. C. Pardo-Fernández). The research was begun while Juan Carlos Pardo-Fernández was visiting the Universidad de Buenos Aires funded by a scholarship from Santander Universidades (Programa Becas Iberoamérica Jóvenes Profesores e Investigadores España 2014).

Appendix A. Auxiliary results and proof of Theorem 1

A.1. Some results for the robust estimator of the regression function

In this section we give several general results for the robust estimator of the regression function given in (6) that will be used later in the proof of Theorem 1. Strong order of convergence for local M-estimators was studied, among others, by Boente and Fraiman (1991). Recently, Boente and Vahnovan (2015) extended these results to the functional setting, achieving better order the convergence than in the Euclidean setting. For that reason, we will use their results.

From Boente and Vahnovan (2015) we have that, under conditions A2, A4, A5, A7 and A9,

$$\sup_{x\in\mathcal{K}} |\widehat{m}_j(x) - m_j(x)| = O_{a.co.}\left(h^2 + \theta_{n_j}\right), \tag{A.1}$$

where $\theta_{n_j} = \sqrt{\log n_j/(n_j h)}$ for any compact set $\mathcal{K} \subset \mathring{\mathcal{R}}$, where $\mathring{\mathcal{R}}$ stands for the interior of the set $\mathcal{R}$.

Assume that Ψ_j is twice continuously differentiable, with first and second derivatives Ψ_j' and Ψ_j'', respectively. Then, from (7) and denoting $w_{ij}(x) = K_h\left(x - X_{ij}\right)$, we have the following expansion

$$\begin{aligned}0 &= \frac{1}{n_j}\sum_{i=1}^{n_j} w_{ij}(x)\Psi_j\left(\frac{Y_{ij} - \widehat{m}_j(x)}{\widehat{\sigma}_j}\right)\\ &= \frac{1}{n_j}\sum_{i=1}^{n_j} w_{ij}(x)\Psi_j\left(\frac{Y_{ij} - m_j(x)}{\widehat{\sigma}_j}\right) + \frac{m_j(x) - \widehat{m}_j(x)}{\widehat{\sigma}_j}\widehat{A}_j(x,\widehat{\sigma}_j),\end{aligned}$$

where

$$\widehat{A}_j(x,\widehat{\sigma}_j) = \frac{1}{n_j}\sum_{i=1}^{n_j} w_{ij}(x)\Psi_j'\left(\frac{Y_{ij} - m_j(x)}{\widehat{\sigma}_j}\right) - \frac{1}{2}\,\frac{\widehat{m}_j(x) - m_j(x)}{\widehat{\sigma}_j}\,\frac{1}{n_j}\sum_{i=1}^{n_j} w_{ij}(x)\Psi_j''\left(\frac{Y_{ij} - m_j(x) + \xi_{ij}(x)}{\widehat{\sigma}_j}\right),$$

with $\xi_{ij}(x)$ an intermediate point between 0 and $m_j(x) - \widehat{m}_j(x)$. Hence, we obtain the following representation

$$\widehat{m}_j(x) - m_j(x) = \widehat{A}_j(x,\widehat{\sigma}_j)^{-1}\,\frac{\widehat{\sigma}_j}{n_j}\sum_{i=1}^{n_j} w_{ij}(x)\Psi_j\left(\frac{Y_{ij} - m_j(x)}{\widehat{\sigma}_j}\right). \tag{A.2}$$

The expansion (A.2) will be helpful when deriving the asymptotic behaviour of the test statistic. Note that since the density, f_j, of X_j, is twice continuously differentiable and Ψ_j'' is bounded from (A.1) we get that

$$\sup_{x\in\mathcal{K}}\left|\widehat{A}_j(x,\widehat{\sigma}_j) - \frac{1}{n_j}\sum_{i=1}^{n_j} w_{ij}(x)\Psi_j'\left(\frac{Y_{ij} - m_j(x)}{\widehat{\sigma}_j}\right)\right| = O_{a.co.}\left(h^2 + \theta_{n_j}\right).$$

Hence, standard arguments and the consistency of $\widehat{\sigma}_j$ allow to show that

$$\sup_{x\in\mathcal{K}} |\widehat{A}_j(x,\widehat{\sigma}_j) - f_j(x)\nu_{j,j}| = O_{a.co.}\left(h^2 + \theta_{n_j}\right),$$

where $\nu_{j,j} = \mathbb{E}[\Psi_j'\left(U_j\right)]$, so

$$\sup_{x\in\mathcal{K}} |\widehat{B}_j(x,\widehat{\sigma}_j)| = O_{a.co.}\left(h^2 + \theta_{n_j}\right), \tag{A.3}$$

with $\widehat{B}_j(x,\widehat{\sigma}_j) = \widehat{A}_j^{-1}(x,\widehat{\sigma}_j) - \left(f_j(x)\nu_{j,j}\right)^{-1}$. Thus, if we denote

$$\widehat{L}_j(x,\widehat{\sigma}_j) = \frac{1}{n_j}\sum_{i=1}^{n_j} w_{ij}(x)\Psi_j\left(\frac{Y_{ij} - m_j(x)}{\widehat{\sigma}_j}\right),$$

$$\Lambda_j(x,u,\sigma) = \mathbb{E}\left\{\Psi_j\left(\frac{Y_j - m_j(x)}{\sigma}\right) \mid X_j = u\right\} = \mathbb{E}\left\{\Psi_j\left(\frac{\sigma_j U_j + m_j(u) - m_j(x)}{\sigma}\right)\right\},$$

we have that

$$\begin{aligned}\widehat{M}_j(x,\widehat{\sigma}_j) &= \widehat{m}_j(x) - m_j(x) - \frac{\widehat{\sigma}_j}{f_j(x)\nu_{j,j}}\widehat{L}_j(x,\widehat{\sigma}_j) = \widehat{B}_j(x,\widehat{\sigma}_j)\widehat{\sigma}_j\widehat{L}_j(x,\widehat{\sigma}_j)\\ &= \widehat{B}_j(x,\widehat{\sigma}_j)\widehat{\sigma}_j\frac{1}{n_j}\sum_{i=1}^{n_j}K_h(x-X_{ij})\left\{\Psi_j\left(\frac{Y_{ij}-m_j(x)}{\widehat{\sigma}_j}\right) - \Lambda_j(x,X_{ij},\widehat{\sigma}_j)\right\}\\ &\quad + \widehat{B}_j(x,\widehat{\sigma}_j)\widehat{\sigma}_j\frac{1}{n_j}\sum_{i=1}^{n_j}K_h(x-X_{ij})\Lambda_j(x,X_{ij},\widehat{\sigma}_j)\\ &= \widehat{B}_j(x,\widehat{\sigma}_j)\widehat{\sigma}_j\widehat{M}_{j,1}(x,\widehat{\sigma}_j) + \widehat{B}_j(x,\widehat{\sigma}_j)\widehat{\sigma}_j\widehat{M}_{j,2}(x,\widehat{\sigma}_j). \end{aligned} \tag{A.4}$$

As in Ferraty et al. (2010), we easily obtain that

$$\begin{aligned}\sup_{\sigma\in\left[\frac{\sigma_j}{2},2\sigma_j\right]}\sup_{x\in\mathcal{K}}|\widehat{M}_{j,1}(x,\sigma)| &= \sup_{\sigma\in[\frac{\sigma_j}{2},2\sigma_j]}\sup_{x\in\mathcal{K}}\left|\frac{1}{n_j}\sum_{i=1}^{n_j}K_h(x-X_{ij})\left\{\Psi_j\left(\frac{Y_{ij}-m_j(x)}{\sigma}\right) - \Lambda_j(x,X_{ij},\sigma)\right\}\right|\\ &= O_{a.co.}\left(\theta_{n_j}\right), \end{aligned} \tag{A.5}$$

where $\theta_{n_j} = \sqrt{\log n_j/(n_j h)}$. On the other hand, using that $\mathbb{E}[\Psi_j\left(\sigma_j U_j/\sigma\right)] = 0$ for any $\sigma > 0$, a Taylor's expansion of order two leads to

$$\Lambda_j(x,u,\sigma) = \frac{m_j(u)-m_j(x)}{\sigma}\mathbb{E}\left[\Psi_j'\left(\frac{\sigma_j U_j}{\sigma}\right)\right] + \frac{1}{2}\frac{(m_j(u)-m_j(x))^2}{\sigma^2}\mathbb{E}\left[\Psi_j''\left(\frac{\sigma_j U_j + \xi_{ij}(u,x)(m_j(u)-m_j(x))}{\sigma}\right)\right],$$

which, together with the fact that Ψ_j' and Ψ_j'' are bounded, implies that

$$\begin{aligned}\sup_{\sigma\in[\frac{\sigma_j}{2},2\sigma_j]}\sup_{x\in\mathcal{K}}|\widehat{M}_{j,2}(x,\sigma)| &= \sup_{\sigma\in[\frac{\sigma_j}{2},2\sigma_j]}\sup_{x\in\mathcal{K}}\left|\frac{1}{n_j}\sum_{i=1}^{n_j}K_h(x-X_{ij})\Lambda_j(x,X_{ij},\sigma)\right|\\ &= O_{a.co.}\left(h^2+\theta_{n_j}\right) \end{aligned} \tag{A.6}$$

Therefore, (A.3), (A.5), (A.6) and the consistency of $\widehat{\sigma}_j$ yield to

$$\sup_{x\in\mathcal{K}}|\widehat{M}_j(x,\widehat{\sigma}_j)| = \sup_{x\in\mathcal{K}}\left|\widehat{m}_j(x) - m_j(x) - \frac{\widehat{\sigma}_j}{f_j(x)\nu_{j,j}}\widehat{L}_j(x,\widehat{\sigma}_j)\right| = O_{a.co.}\left(h^2+\theta_{n_j}^2\right). \tag{A.7}$$

It is worth noting that, analogous arguments to those considered in Theorem 4.4 in Boente and Vahnovan (2015) together with the previous computations allow to show that (A.1) and (A.7) still hold when $m_2(x) = m_1(x) + n^{-1/2}\,\Delta(x)$, i.e., under the considered root-n local alternatives in which $(X_{i2}, Y_{i2})^{\mathrm{T}}$, $1 \le i \le n_2$, correspond to a triangular array with $Y_{i2} = Y_{i2}^{(n)} = m_1(X_{i2}) + n^{-1/2}\,\Delta(X_{i2}) + \sigma_2 U_{i2}$.

A.2. Proof of Theorem 1

We first state some technical results collected in a Lemma whose proof can be found in the supplementary material available online (see Appendix B).

Lemma A.1. *Assume that (1) and A1 to A9 hold. Let $\widehat{\sigma}_j$ be a consistent estimator of σ_j. For any fixed $j = 1, 2$, denote $\widehat{R}_1(\sigma) = (1/n_j^2)\sum_{1\le i\ne\ell\le n_j} Z_{i\ell}(\sigma)$ and $\widehat{R}_2(\sigma,\widetilde{\sigma}) = 1/(n_j\, n_{3-j})\sum_{i=1}^{n_j}\sum_{\ell=1}^{n_{3-j}} W_{i\ell}(\sigma,\widetilde{\sigma})$ where*

$$\begin{aligned} Z_{i\ell}(\sigma) &= \Psi'\left(\frac{\sigma_j\, U_{ij}}{\sigma}\right)\Psi_j'\left(\frac{\sigma_j U_{\ell j}}{\sigma}\right)\frac{p_j(X_{ij})w_j(X_{ij})}{f_j(X_{ij})}K_h(X_{ij}-X_{\ell j})\left(m_j(X_{\ell j}) - m_j(X_{ij})\right)\\ W_{i\ell}(\sigma,\widetilde{\sigma}) &= \Psi'\left(\frac{\sigma_j\, U_{ij}}{\sigma}\right)\Psi_{3-j}'\left(\frac{\sigma_{3-j} U_{\ell,3-j}}{\widetilde{\sigma}}\right)\frac{p_{3-j}(X_{ij})w_j(X_{ij})}{f_{3-j}(X_{ij})}K_h(X_{ij}-X_{\ell,3-j})\left(m_{3-j}(X_{\ell,3-j}) - m_{3-j}(X_{ij})\right). \end{aligned}$$

Then,

(a) $\sup_{\sigma\in\mathcal{I}_j}\left|\mathbb{E}\left[\widehat{R}_1(\sigma)\right]\right| = o_{\mathbb{P}}(n^{-1/2})$ *and* $\sup_{\sigma\in\mathcal{I}_j,\widetilde{\sigma}\in\mathcal{I}_{3-j}}\left|\mathbb{E}\left[\widehat{R}_2(\sigma,\widetilde{\sigma})\right]\right| = o_{\mathbb{P}}(n^{-1/2})$, *where $\mathcal{I}_s = [\sigma_s/2, 2\sigma_s]$, for $s = j$, $3-j$.*

(b) *There exists a constant* $C > 0$ *not depending on n such that for all* $n \geq n_0$

$$\sup_{\sigma \in \mathcal{I}_j} \mathbb{P}\left(\sqrt{n}|\widehat{R}_1(\sigma) - \mathbb{E}\widehat{R}_1(\sigma)| > \epsilon\right) \leq C\left(\frac{h^2}{n_j} + h^2\right)$$

$$\sup_{\sigma \in \mathcal{I}_j, \widetilde{\sigma} \in \mathcal{I}_{3-j}} \mathbb{P}(\sqrt{n}|\widehat{R}_2(\sigma, \widetilde{\sigma}) - \mathbb{E}\widehat{R}_2(\sigma, \widetilde{\sigma})| > \epsilon) \leq C\left(\frac{h^2}{n_j} + h^4\right)$$

(c) $\sup_{\sigma \in \mathcal{I}_j} \left|\widehat{R}_1(\sigma) - \mathbb{E}\left[\widehat{R}_1(\sigma)\right]\right| = o_{\mathbb{P}}(n^{-1/2})$ *and* $\sup_{\sigma \in \mathcal{I}_j, \widetilde{\sigma} \in \mathcal{I}_{3-j}} \left|\widehat{R}_2(\sigma, \widetilde{\sigma}) - \mathbb{E}\left[\widehat{R}_2(\sigma, \widetilde{\sigma})\right]\right| = o_{\mathbb{P}}(n^{-1/2})$.

(d) $\widehat{R}_1(\widehat{\sigma}_j) \stackrel{p}{\to} 0$ *and* $\widehat{R}_2(\widehat{\sigma}_j, \widehat{\sigma}_{3-j}) \stackrel{p}{\to} 0$

Proof of Theorem 1. We begin by obtaining an expansion for $\widehat{E}_{j0}$ that will allow us to derive the asymptotic distribution of T. Using that Ψ is twice continuously differentiable, a Taylor's expansion of order two leads to

$$\begin{aligned}
\widehat{E}_{j0} &= \frac{1}{n_j}\sum_{i=1}^{n_j} \Psi\left(\frac{Y_{ij} - \widehat{m}(X_{ij})}{\widehat{\sigma}_j}\right) w_j(X_{ij}) \\
&= \frac{1}{n_j}\sum_{i=1}^{n_j} \Psi\left(\frac{\sigma_j U_{ij} + m_j(X_{ij}) - \widehat{m}(X_{ij})}{\widehat{\sigma}_j}\right) w_j(X_{ij}) \\
&= \frac{1}{n_j}\sum_{i=1}^{n_j} \Psi\left(\frac{\sigma_j U_{ij}}{\widehat{\sigma}_j}\right) w_j(X_{ij}) + \frac{1}{\widehat{\sigma}_j}\frac{1}{n_j}\sum_{i=1}^{n_j} \Psi'\left(\frac{\sigma_j U_{ij}}{\widehat{\sigma}_j}\right)(m_j(X_{ij}) - \widehat{m}(X_{ij}))w_j(X_{ij}) \\
&\quad + \frac{1}{2\widehat{\sigma}_j^2}\frac{1}{n_j}\sum_{i=1}^{n_j} \Psi''\left(\frac{\xi_{ij}}{\widehat{\sigma}_j}\right)(m_j(X_{ij}) - \widehat{m}(X_{ij}))^2 w_j(X_{ij}) \\
&= T_{j1}(\widehat{\sigma}_j) + \frac{1}{\widehat{\sigma}_j} T_{j2} + T_{j3},
\end{aligned}$$

where $\xi_{ij} = \sigma_j U_{ij} + \theta_{ij}(m_j(X_{ij}) - \widehat{m}(X_{ij}))$ and θ_{ij} is an intermediate point in [0, 1]. Using that $\widehat{\sigma}_j \stackrel{p}{\to} \sigma_j$ and that ζ is bounded, standard empirical process arguments allow to show that $T_{j1}(\widehat{\sigma}_j)$ has the same asymptotic behaviour as $T_{j1}(\sigma_j)$, i.e., $\sqrt{n}\{T_{j1}(\widehat{\sigma}_j) - T_{j1}(\sigma_j)\} = o_{\mathbb{P}}(1)$. On the other hand, using that Ψ'' is bounded, from **A9** and (A.1), we get $T_{j3} = o_{\mathbb{P}}(n^{-1/2})$. Hence we have that

$$\widehat{E}_{j0} = \frac{1}{n_j}\sum_{i=1}^{n_j} \Psi\left(U_{ij}\right) w_j(X_{ij}) + \frac{1}{\widehat{\sigma}_j} T_{j2} + o_{\mathbb{P}}(n^{-1/2}). \tag{A.8}$$

The term T_{j2} needs to be further analysed. Note that $m_j(x) = p_1(x)m_j(x) + p_2(x)m_j(x)$, so $m_j(x) - \widehat{m}(x) = p_1(x)\{m_j(x) - \widehat{m}_1(x)\} + p_2(x)\{m_j(x) - \widehat{m}_2(x)\} = \sum_{s=1}^{2} p_s(x)\{m_j(x) - \widehat{m}_s(x)\}$ which leads to

$$m_j(x) - \widehat{m}(x) = \sum_{s=1}^{2} p_s(x)\{m_j(x) - m_s(x)\} + \sum_{s=1}^{2} p_s(x)\{m_s(x) - \widehat{m}_s(x)\}.$$

Hence,

$$T_{j2} = \frac{1}{n_j}\sum_{i=1}^{n_j} \Psi'\left(\frac{\sigma_j U_{ij}}{\widehat{\sigma}_j}\right)(m_j(X_{ij}) - \widehat{m}(X_{ij}))w_j(X_{ij}) = \widehat{T}_{j2,1}(\widehat{\sigma}_j) - \widehat{T}_{j2,2}(\widehat{\sigma}_j),$$

where

$$\begin{aligned}
\widehat{T}_{j2,1}(\sigma) &= \sum_{s=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j} \Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right) p_s(X_{ij})(m_j(X_{ij}) - m_s(X_{ij}))w_j(X_{ij}), \\
\widehat{T}_{j2,2}(\sigma) &= \sum_{s=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j} \Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right) p_s(X_{ij})(\widehat{m}_s(X_{ij}) - m_s(X_{ij}))w_j(X_{ij}).
\end{aligned}$$

We have the following expression for $\widehat{T}_{j2,1}(\sigma)$

$$\widehat{T}_{j2,1}(\sigma) = (-1)^j \frac{1}{n_j}\sum_{i=1}^{n_j} \Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right)(m_2(X_{ij}) - m_1(X_{ij}))w_j(X_{ij})p_{3-j}(X_{ij}), \tag{A.9}$$

Note that under the null hypothesis $\widehat{T}_{j2,1}(\widehat{\sigma}_j) = 0$. So, using (A.8), we obtain that

$$\widehat{E}_{j0} = \frac{1}{n_j}\sum_{i=1}^{n_j} \Psi\left(U_{ij}\right) w_j(X_{ij}) + \frac{1}{\widehat{\sigma}_j}\widehat{T}_{j2,1}(\widehat{\sigma}_j) - \frac{1}{\widehat{\sigma}_j}\widehat{T}_{j2,2}(\widehat{\sigma}_j) + o_{\mathbb{P}}(n^{-1/2}). \tag{A.10}$$

To study the term $\widehat{T}_{j2,2}(\widehat{\sigma}_j)$, we will use the representation for $\widehat{m}_s(x) - m_s(x)$ given in (A.7) with $\mathcal{K} = \mathcal{S}_j$, which also holds under $n^{1/2}$ local alternatives, so

$$\begin{aligned}
\widehat{T}_{j2,2}(\sigma) &= \sum_{s=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j}\Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right) p_s(X_{ij})(\widehat{m}_s(X_{ij}) - m_s(X_{ij}))w_j(X_{ij}) \\
&= \sum_{s=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j}\Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right) p_s(X_{ij})w_j(X_{ij})\frac{\widehat{\sigma}_s}{f_s(X_{ij})\nu_{s,s}}\widehat{L}_s(X_{ij},\widehat{\sigma}_s) \\
&\quad + \sum_{s=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j}\Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right) p_s(X_{ij})\widehat{M}_s(X_{ij},\widehat{\sigma}_s)w_j(X_{ij}) \\
&= \widehat{R}_{j,1}(\sigma) + \widehat{R}_{j,2}(\sigma),
\end{aligned} \tag{A.11}$$

where $\widehat{M}_j(x,\sigma)$ is given in (A.4). Hence, (A.7) and the fact that $\sqrt{nh_n^2}/\log n \to \infty$ and $nh^4 \to 0$ imply that, for $s = 1, 2$, $\max_i |\widehat{M}_s(X_{ij},\widehat{\sigma}_s)| = o_{\mathbb{P}}(n^{-1/2})$ so using that $0 \le p_s \le 1$ and that Ψ' and w_j are bounded, we get that $\widehat{R}_{j,2}(\widehat{\sigma}_j) = o_{\mathbb{P}}(n^{-1/2})$. Therefore, we only have to study the behaviour of $\widehat{R}_{j,1}(\widehat{\sigma}_j)$. Note that

$$\begin{aligned}
\widehat{R}_{j,1}(\sigma) &= \sum_{s=1}^{2}\frac{\widehat{\sigma}_s}{\nu_{s,s}}\frac{1}{n_j}\sum_{i=1}^{n_j}\Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right)\frac{p_s(X_{ij})w_j(X_{ij})}{f_s(X_{ij})}\widehat{L}_s(X_{ij},\widehat{\sigma}_s) \\
&= \sum_{s=1}^{2}\frac{\widehat{\sigma}_s}{\nu_{s,s}}\frac{1}{n_j}\sum_{i=1}^{n_j}\Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right)\frac{p_s(X_{ij})w_j(X_{ij})}{f_s(X_{ij})}\frac{1}{n_s}\sum_{\ell=1}^{n_s}K_h(X_{ij} - X_{\ell s})\Psi_s\left(\frac{Y_{\ell s} - m_s(X_{ij})}{\widehat{\sigma}_s}\right).
\end{aligned}$$

Using that $Y_{\ell s} = \sigma_s U_{\ell s} + m_s(X_{\ell s})$ and applying a second order Taylor's expansion, we obtain that $\widehat{R}_{j,1}(\sigma) = \widehat{R}_{j,1,1}(\sigma) + \widehat{R}_{j,1,2}(\sigma) + \widehat{R}_{j,1,3}(\sigma)$ where

$$\begin{aligned}
\widehat{R}_{j,1,1}(\sigma) &= \sum_{s=1}^{2}\frac{\widehat{\sigma}_s}{\nu_{s,s}}\frac{1}{n_j}\sum_{i=1}^{n_j}\Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right)\frac{p_s(X_{ij})w_j(X_{ij})}{f_s(X_{ij})}\frac{1}{n_s}\sum_{\ell=1}^{n_s}K_h(X_{ij} - X_{\ell s})\Psi_s\left(\frac{\sigma_s U_{\ell s}}{\widehat{\sigma}_s}\right), \\
\widehat{R}_{j,1,2}(\sigma) &= \sum_{s=1}^{2}\frac{1}{\nu_{s,s}}\frac{1}{n_j}\sum_{i=1}^{n_j}\Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right)\frac{p_s(X_{ij})w_j(X_{ij})}{f_s(X_{ij})}\frac{1}{n_s}\sum_{\ell=1}^{n_s}K_h(X_{ij} - X_{\ell s})\Psi_s'\left(\frac{\sigma_s U_{\ell s}}{\widehat{\sigma}_s}\right)\left(m_s(X_{\ell s}) - m_s(X_{ij})\right), \\
\widehat{R}_{j,1,3}(\sigma) &= \sum_{s=1}^{2}\frac{1}{\widehat{\sigma}_s\nu_{s,s}}\frac{1}{n_j}\sum_{i=1}^{n_j}\Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right)\frac{p_s(X_{ij})w_j(X_{ij})}{f_s(X_{ij})} \\
&\quad \times \frac{1}{n_s}\sum_{\ell=1}^{n_s}K_h(X_{ij} - X_{\ell s})\Psi_s''\left(\frac{\sigma_s U_{\ell s}}{\widehat{\sigma}_s} + \theta_{\ell s}\left(m_s(X_{\ell s}) - m_s(X_{ij})\right)\right)\left(m_s(X_{\ell s}) - m_s(X_{ij})\right)^2,
\end{aligned}$$

where $0 < \theta_{\ell s} < 1$. Using that K has bounded support and m_j is Lipschitz, we get that $|m_s(X_{\ell s}) - m_s(X_{ij})| \le Ch$ where $K_h(X_{ij} - X_{\ell s}) \ne 0$. Thus, the boundness of Ψ', Ψ_s'', w_j and the fact that w_j has support on $\mathcal{S}_j$ and $\inf_{x\in\mathcal{S}_j} f_j(x) > 0$ together with the consistency of $\widehat{\sigma}_s$ and the assumption $nh^4 \to 0$ entail that $\widehat{R}_{j,1,3}(\widehat{\sigma}_j) = o_{\mathbb{P}}(n^{-1/2})$. Note that $\widehat{R}_{j,1,2}(\widehat{\sigma}_j) = \nu_{j,j}^{-1}\widehat{R}_{j,1,2}^{(1)}(\widehat{\sigma}_j) + \nu_{3-j,3-j}^{-1}\widehat{R}_{j,1,2}^{(2)}(\widehat{\sigma}_j,\widehat{\sigma}_{3-j})$, where $\widehat{R}_{j,1,2}^{(1)}(\sigma) = n_j^{-2}\sum_{1\le i\ne\ell\le n_j} Z_{i\ell}^{(j)}$ and $\widehat{R}_{j,1,2}^{(2)}(\sigma,\widetilde{\sigma}) = n_j^{-1}n_{3-j}^{-1}\sum_{i=1}^{n_j}\sum_{\ell=1}^{n_{3-j}} W_{i\ell}^{(j)}$ with

$$\begin{aligned}
Z_{i\ell}^{(j)} &= \Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right)\Psi_j'\left(\frac{\sigma_j U_{\ell j}}{\sigma}\right)\frac{p_j(X_{ij})w_j(X_{ij})}{f_j(X_{ij})}K_h(X_{ij} - X_{\ell j})\left(m_j(X_{\ell j}) - m_j(X_{ij})\right), \\
W_{i\ell}^{(j)} &= \Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right)\Psi_{3-j}'\left(\frac{\sigma_{3-j} U_{\ell,3-j}}{\widetilde{\sigma}}\right)\frac{p_{3-j}(X_{ij})w_j(X_{ij})}{f_{3-j}(X_{ij})}K_h(X_{ij} - X_{\ell,3-j})\left(m_{3-j}(X_{\ell,3-j}) - m_{3-j}(X_{ij})\right).
\end{aligned}$$

Lemma A.1 and the fact that $\widehat{\sigma}_j$ is consistent lead us to $\widehat{R}_{j,1,2}(\widehat{\sigma}_j) = o_{\mathbb{P}}(n^{-1/2})$.

To deal with $\widehat{R}_{j,1,1}(\sigma)$, we rearrange the sum to obtain that $\widehat{R}_{j,1,1}(\sigma) = \sum_{s=1}^{2} \nu_{s,s}^{-1}\widehat{R}^{(s)}_{j,1,1}(\sigma, \widehat{\sigma}_s)$, where

$$\begin{aligned}\widehat{R}^{(s)}_{j,1,1}(\sigma, \widetilde{\sigma}) &= \widetilde{\sigma}\frac{1}{n_j}\sum_{i=1}^{n_j}\Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right)\frac{p_s(X_{ij})w_j(X_{ij})}{f_s(X_{ij})}\frac{1}{n_s}\sum_{\ell=1}^{n_s}K_h(X_{ij}-X_{\ell s})\Psi_s\left(\frac{\sigma_s U_{\ell s}}{\widetilde{\sigma}}\right)\\ &= \frac{1}{n_s}\sum_{\ell=1}^{n_s}\widetilde{\sigma}\Psi_s\left(\frac{\sigma_s U_{\ell s}}{\widetilde{\sigma}}\right)\frac{1}{n_j}\sum_{i=1}^{n_j}\Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right)\frac{p_s(X_{ij})w_j(X_{ij})}{f_s(X_{ij})}K_h(X_{ij}-X_{\ell s}).\end{aligned}$$

Using that $\mathbb{E}[\Psi_s(\sigma_s U_{\ell s}/\sigma)] = 0$ for any $\sigma > 0$, that $\widehat{\sigma}_j$ is a consistent estimator of σ_j and that $\zeta_{1,j}(u) = u\Psi_j'(u)$ is bounded, we easily get that

$$\widehat{R}^{(s)}_{j,1,1}(\widehat{\sigma}_j, \widehat{\sigma}_s) = \sigma_s\nu_j\frac{1}{n_s}\sum_{\ell=1}^{n_s}\Psi_s\left(U_{\ell s}\right)\frac{p_s(X_{\ell s})w_j(X_{\ell s})f_j(X_{\ell s})}{f_s(X_{\ell s})} + o_{\mathbb{P}}(n^{-1/2}). \tag{A.12}$$

From (A.10), (A.11), the fact that $\widehat{T}_{j2,2}(\sigma) = \sum_{s=1}^{2}(1/\nu_{s,s})\widehat{R}^{(s)}_{j,1,1}(\sigma, \widehat{\sigma}_s) + o_{\mathbb{P}}(n^{-1/2})$ and (A.12) we obtain that

$$\widehat{E}_{j0} = \frac{1}{n_j}\sum_{i=1}^{n_j}\Psi\left(U_{ij}\right)w_j(X_{ij}) + \frac{1}{\widehat{\sigma}_j}\widehat{T}_{j2,1}(\widehat{\sigma}_j) - \frac{\nu_j}{\sigma_j}\sum_{s=1}^{2}\frac{\sigma_s}{\nu_{s,s}}\frac{1}{n_s}\sum_{\ell=1}^{n_s}\Psi_s\left(U_{\ell s}\right)\frac{p_s(X_{\ell s})w_j(X_{\ell s})f_j(X_{\ell s})}{f_s(X_{\ell s})} + o_{\mathbb{P}}(n^{-1/2}). \tag{A.13}$$

Thus, we have that

$$\begin{aligned}\widehat{E}_{20} &= \frac{1}{\widehat{\sigma}_2}\widehat{T}_{22,1}(\widehat{\sigma}_2) + \frac{1}{n_2}\sum_{\ell=1}^{n_2}\Psi\left(U_{\ell 2}\right)w_2(X_{\ell 2}) - \frac{\nu_2}{\nu_{2,2}}\frac{1}{n_2}\sum_{\ell=1}^{n_2}\Psi_2\left(U_{\ell 2}\right)\frac{p_2(X_{\ell 2})w_2(X_{\ell 2})f_2(X_{\ell 2})}{f_2(X_{\ell 2})}\\ &\quad - \frac{\nu_2\sigma_1}{\sigma_2\nu_{1,1}}\frac{1}{n_1}\sum_{\ell=1}^{n_1}\Psi_1\left(U_{\ell 1}\right)\frac{p_1(X_{\ell 1})w_2(X_{\ell 1})f_2(X_{\ell 1})}{f_1(X_{\ell 1})} + o_{\mathbb{P}}(n^{-1/2}),\\ \widehat{E}_{10} &= \frac{1}{\widehat{\sigma}_1}\widehat{T}_{12,1}(\widehat{\sigma}_1) + \frac{1}{n_1}\sum_{\ell=1}^{n_1}\Psi\left(U_{\ell 1}\right)w_1(X_{\ell 1}) - \frac{\nu_1}{\nu_{1,1}}\frac{1}{n_1}\sum_{\ell=1}^{n_1}\Psi_1\left(U_{\ell 1}\right)\frac{p_1(X_{\ell 1})w_1(X_{\ell 1})f_1(X_{\ell 1})}{f_1(X_{\ell 1})}\\ &\quad - \frac{\nu_1\sigma_2}{\sigma_1\nu_{2,2}}\frac{1}{n_2}\sum_{\ell=1}^{n_2}\Psi_2\left(U_{\ell 2}\right)\frac{p_2(X_{\ell 2})w_1(X_{\ell 2})f_1(X_{\ell 2})}{f_2(X_{\ell 2})} + o_{\mathbb{P}}(n^{-1/2}),\end{aligned}$$

so that

$$\begin{aligned}\widehat{E}_{20} - \widehat{E}_{10} &= \frac{1}{n_1}\sum_{\ell=1}^{n_1}\Psi_1\left(U_{\ell 1}\right)\frac{p_1(X_{\ell 1})}{f_1(X_{\ell 1})}\left\{\frac{\nu_1}{\nu_{1,1}}w_1(X_{\ell 1})f_1(X_{\ell 1}) - \frac{\nu_2\sigma_1}{\sigma_2\nu_{1,1}}w_2(X_{\ell 1})f_2(X_{\ell 1})\right\} - \Psi\left(U_{\ell 1}\right)w_1(X_{\ell 1})\\ &\quad - \frac{1}{n_2}\sum_{\ell=1}^{n_2}\Psi_2\left(U_{\ell 2}\right)\frac{p_2(X_{\ell 2})}{f_2(X_{\ell 2})}\left\{\frac{\nu_2}{\nu_{2,2}}w_2(X_{\ell 2})f_2(X_{\ell 2}) - \frac{\nu_1\sigma_2}{\sigma_1\nu_{2,2}}w_1(X_{\ell 2})f_1(X_{\ell 2})\right\} - \Psi\left(U_{\ell 2}\right)w_2(X_{\ell 2})\\ &\quad + \frac{1}{\widehat{\sigma}_2}\widehat{T}_{22,1}(\widehat{\sigma}_2) - \frac{1}{\widehat{\sigma}_1}\widehat{T}_{12,1}(\widehat{\sigma}_1) + o_{\mathbb{P}}(n^{-1/2})\\ &= S_{2,n_2} - S_{1,n_1} + \frac{1}{\widehat{\sigma}_2}\widehat{T}_{22,1}(\widehat{\sigma}_2) - \frac{1}{\widehat{\sigma}_1}\widehat{T}_{12,1}(\widehat{\sigma}_1) + o_{\mathbb{P}}(n^{-1/2}),\end{aligned}$$

where

$$\begin{aligned}S_{2,n_2} &= \frac{1}{n_2}\sum_{\ell=1}^{n_2}\Psi_2\left(U_{\ell 2}\right)\frac{p_2(X_{\ell 2})}{f_2(X_{\ell 2})}\left\{\frac{\nu_1\sigma_2}{\sigma_1\nu_{2,2}}w_1(X_{\ell 2})f_1(X_{\ell 2}) - \frac{\nu_2}{\nu_{2,2}}w_2(X_{\ell 2})f_2(X_{\ell 2})\right\} + \Psi\left(U_{\ell 2}\right)w_2(X_{\ell 2})\\ S_{1,n_1} &= \frac{1}{n_1}\sum_{\ell=1}^{n_1}\Psi_1\left(U_{\ell 1}\right)\frac{p_1(X_{\ell 1})}{f_1(X_{\ell 1})}\left\{\frac{\nu_2\sigma_1}{\sigma_2\nu_{1,1}}w_2(X_{\ell 1})f_2(X_{\ell 1}) - \frac{\nu_1}{\nu_{1,1}}w_1(X_{\ell 1})f_1(X_{\ell 1})\right\} + \Psi\left(U_{\ell 1}\right)w_1(X_{\ell 1}).\end{aligned}$$

Therefore, the test statistic can be written as

$$T = \left(\frac{n_1}{n}\right)^{1/2} n_2^{1/2}S_{2,n_2} - \left(\frac{n_2}{n}\right)^{1/2} n_1^{1/2}S_{1,n_1} + \widehat{\Delta}_{n_1,n_2} + o_{\mathbb{P}}(n^{-1/2}),$$

where

$$\widehat{\Delta}_{n_1,n_2} = \left(\frac{n_1}{n}\frac{n_2}{n}\right)^{1/2} n^{1/2}\left(\frac{1}{\widehat{\sigma}_2}\widehat{T}_{22,1}(\widehat{\sigma}_2) - \frac{1}{\widehat{\sigma}_1}\widehat{T}_{12,1}(\widehat{\sigma}_1)\right).$$

For $j = 1, 2$, the term $n_j^{1/2} S_{j,n_j}$ is asymptotically normally distributed with mean 0 and variance τ_j^2 given in (12).

(a) Under H_0, $\widehat{T}_{22,1}(\widehat{\sigma}_2) = 0$ and $\widehat{T}_{12,1}(\widehat{\sigma}_1) = 0$, thus $\widehat{\Delta}_{n_1,n_2} = 0$. Therefore, $T \xrightarrow{D} N(0, \kappa_1\tau_2^2 + \kappa_2\tau_1^2)$ under H_0, concluding the proof of (a).

(b) To analyse the asymptotic behaviour of the test statistic under H_1 recall the representation given in (A.9). Let $\mathcal{P}_\ell = \{x \in \mathcal{R},\ p_{3-\ell}(x) > 0\}$. Denote $\mathcal{J} = \{j \in \{1, 2\}$ such that $\mathcal{P}_j \cap \mathcal{S}_j \cap \mathcal{A} \neq \emptyset\}$. Taking into account that $p_2(x) = 1 - p_1(x)$, we have that $\text{card}(\mathcal{J}) \geq 1$. Let

$$T_{j2,1}(\sigma) = \mathbb{E}[\widehat{T}_{j2,1}(\sigma)] = (-1)^j \mathbb{E}\left[\Psi'\left(\frac{\sigma_j U_j}{\sigma}\right)\right] \mathbb{E}\left[(m_2(X_j) - m_1(X_j))w_j(X_j)p_{3-j}(X_j)\right].$$

Since $\widehat{\sigma}_j$ is a consistent estimator of σ_j, under H_1 we have that $\sqrt{n}\{\widehat{T}_{j2,1}(\widehat{\sigma}_j) - T_{j2,1}(\widehat{\sigma}_j)\}$ has the same asymptotic distribution as $\sqrt{n}\{\widehat{T}_{j2,1}(\sigma_j) - T_{j2,1}(\sigma_j)\}$, which is asymptotically normally distributed with asymptotic variance $\kappa_j^{-1}\, \mathbb{E}[(\Psi'(U_j))^2]\mathbb{E}[(m_2(X_j) - m_1(X_j))^2 w_j^2(X_j)p_{3-j}^2(X_{ij})] \neq 0$ for $j \in \mathcal{J}$. Hence, if $\mathcal{J} = \{1, 2\}$, using that $\nu_j = \mathbb{E}[\Psi'(U_{ij})] > 0$, we have that, under H_1, $\sqrt{n}(-1)^j T_{j2,1}(\sigma_j) \to +\infty$ for $j \in \mathcal{J}$. This implies that $\sqrt{n}(-1)^j \widehat{T}_{j2,1}(\widehat{\sigma}_j) \to +\infty$, for $j \in \mathcal{J}$, so that $\widehat{\Delta}_{n_1,n_2} \to +\infty$. Otherwise, $\text{card}(\mathcal{J}) = 1$ and denote j_0 its unique element. Then, $p_{j_0}(x) = 0$ in $\mathcal{S}_{3-j_0} \cap \mathcal{A}$ so $\sqrt{n} T_{j_0 2,1}(\sigma_j) = 0$ while $\sqrt{n}(-1)^{3-j_0} T_{3-j_0 2,1}(\sigma_j) \to +\infty$, which concludes the proof of (b).

(c) Finally, we now consider the behaviour under $H_{1,n}$. In this case, $m_2(x) = m_1(x) + n^{-1/2}\Delta(x)$, so

$$n^{1/2}\widehat{T}_{j2,1}(\sigma) = (-1)^j \frac{1}{n_j}\sum_{i=1}^{n_j} \Psi'\left(\frac{\sigma_j U_{ij}}{\sigma}\right) \Delta(X_{ij}) w_j(X_{ij}) p_{3-j}(X_{ij}).$$

Using standard empirical process arguments, it is easy to see that the consistency of $\widehat{\sigma}_j$, implies that $n^{1/2}\widehat{T}_{j2,1}(\widehat{\sigma}_j) \xrightarrow{p} (-1)^j \nu_j \mathbb{E}\left[\Delta(X_j)w_j(X_j)p_{3-j}(X_j)\right]$. Thus, we have that

$$n^{1/2}\left(\frac{1}{\widehat{\sigma}_2}\widehat{T}_{22,1}(\widehat{\sigma}_2) - \frac{1}{\widehat{\sigma}_1}\widehat{T}_{12,1}(\widehat{\sigma}_1)\right) \xrightarrow{p} \frac{\nu_2}{\sigma_2}\mathbb{E}\left[\Delta(X_2)w_2(X_2)p_1(X_2)\right] + \frac{\nu_1}{\sigma_1}\mathbb{E}\left[\Delta(X_1)w_1(X_1)p_2(X_1)\right] = d.$$

Taking into account that $n_j/n \to \kappa_j$, we get that $\widehat{\Delta}_{n_1,n_2} \xrightarrow{p} (\kappa_1\kappa_2)^{1/2} d$ under $H_{1,n}$. Therefore, under H_{1n}, the test statistic T converges in distribution to $N\left((\kappa_1\kappa_2)^{1/2} d, \kappa_1\tau_2^2 + \kappa_2\tau_1^2\right)$, concluding the proof. □

Appendix B. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.csda.2015.12.002.

References

Bianco, A., Boente, G., 2007. Robust estimators under a semiparametric partly linear autoregression model: Asymptotic behaviour and bandwidth selection. J. Time Series Anal. 28, 274–306.
Bianco, A., Boente, G., Martínez, E., 2006. Robust tests in semiparametric partly linear models. Scand. J. Stat. 2, 435–450.
Boente, G., Cao, R., González-Manteiga, W., Rodríguez, D., 2013. Testing in generalized partially linear models: A robust approach. Statist. Probab. Lett. 83, 203–212.
Boente, G., Fraiman, R., 1989. Robust nonparametric regression estimation. J. Multivariate Anal. 29, 180–198.
Boente, G., Fraiman, R., 1991. Strong uniform convergence rates for some robust equivariant nonparametric regression estimates for mixing processes. Internat. Statist. Rev. 59, 355–372.
Boente, G., Rodríguez, D., 2010. Robust inference in generalized partially linear models. Comput. Statist. Data Anal. 54, 2942–2966.
Boente, G., Vahnovan, A., 2015. Strong convergence of robust equivariant nonparametric functional regression estimators. Statist. Probab. Lett. 100, 1–11.
Dette, H., Marchlewski, M., 2010. A robust test for homoscedasticity in nonparametric regression. J. Nonparametr. Stat. 22, 723–736.
Dette, H., Munk, A., 1998. Testing heteroscedasticity in nonparametric regression. J. R. Stat. Soc. Ser. B 60, 693–708.
Feng, L., Zou, C., Wang, Z., Zhu, L., 2015. Robust comparison of regression curves. TEST 24, 185–204.
Ferraty, F., Laksaci, A., Vieu, Ph., 2010. Rate of uniform consistency for nonparametric estimates with functional variables. J. Statist. Plann. Inference 140, 335–352.
Ghement, I., Ruiz, M., Zamar, R., 2008. Robust estimation of error scale in nonparametric regression models. J. Statist. Plann. Inference 138, 3200–3216.
Hall, P., Huber, C., Speckman, P.L., 1997. Covariate-matched one-sided tests for the difference between functional means. J. Amer. Statist. Assoc. 92, 1074–1083.
Hall, P., Kay, J., Titterington, D., 1990. Asymptotically optimal difference-based estimation of variance in nonparametric regression. Biometrika 77, 521–528.
Härdle, W., 1990. Applied Nonparametric Regression. Cambridge University Press.
Härdle, W., Tsybakov, A.B., 1988. Robust nonparametric regression with simultaneous scale curve estimation. Ann. Statist. 16, 120–135.
Koul, H.L., Schick, A., 1997. Testing for the equality of two nonparametric regression curves. J. Statist. Plann. Inference 65, 293–314.
Koul, H.L., Schick, A., 2003. Testing for superiority among two regression curves. J. Statist. Plann. Inference 117, 15–33.
Maronna, R., Martin, R., Yohai, V., 2006. Robust Statistics, Theory and Methods. John Wiley & Sons, Ltd..
Neumeyer, N., Dette, H., 2005. A note on one–sided nonparametric analysis of covariance by ranking residuals. Math. Methods Statist. 14, 80–104.
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.
Rice, J., 1984. Bandwidth choice for nonparametric regression. Ann. Statist. 12, 1215–1230.
Salibian-Barrera, M., Van Aelst, S., Yohai, V.J., 2016. Robust tests for linear regression models based on τ-estimates. Comput. Statist. Data Anal. 93, 436–455.
Speckman, P.L., Chin, J.E., Hewett, J.E., Bertelson, S.E., 2003. A one-sided test adjusting for covariates by ranking residuals following smoothing. Unpublished manuscript available at http://www.stat.missouri.edu/~speckman/report/srrt.ps.gz.