Statistica Sinica **17**(2007), 1115-1137

TESTING FOR THE EQUALITY OF k REGRESSION CURVES

Juan Carlos Pardo-Fernández, Ingrid Van Keilegom
and Wenceslao González-Manteiga

*Universidade de Vigo, Université catholique de Louvain
and Universidade de Santiago de Compostela*

Abstract: Assume that (X_j, Y_j) are independent random vectors satisfying the nonparametric regression models $Y_j = m_j(X_j) + \sigma_j(X_j)\varepsilon_j$, for $j = 1, \ldots, k$, where $m_j(X_j) = E(Y_j|X_j)$ and $\sigma_j^2(X_j) = \mathrm{Var}\,(Y_j|X_j)$ are smooth but unknown regression and variance functions respectively, and the error variable ε_j is independent of X_j.

In this article we introduce a procedure to test the hypothesis of equality of the k regression functions. The test is based on the comparison of two estimators of the distribution of the errors in each population. Kolmogorov-Smirnov and Cramér-von Mises type statistics are considered, and their asymptotic distributions are obtained. The proposed tests can detect local alternatives converging to the null hypothesis at the rate $n^{-1/2}$. We describe a bootstrap procedure that approximates the critical values, and present the results of a simulation study in which the behavior of the tests for small and moderate sample sizes is studied. Finally, we include an application to a data set.

Key words and phrases: Bootstrap, Comparison of regression curves, Heteroscedastic regression, nonparametric regression.

1. Introduction. Statistical model

The comparison of two or more groups is an important problem in statistical inference. This comparison can be performed through the regression curves in a non-parametric context. Let (X_j, Y_j) be k independent random vectors, and assume that they satisfy the following non-parametric regression models, for $j = 1, \ldots, k$,

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

where the error variable ε_j, with distribution F_{ε_j}, is independent of X_j, $m_j(X_j) = E(Y_j|X_j)$ is the unknown regression function and $\sigma_j^2(X_j) = \mathrm{Var}\,(Y_j|X_j)$ is the conditional variance function. Suppose that the covariates X_j have common support R_X. Let (X_{ij}, Y_{ij}), $i = 1, \ldots, n_j$, be an i.i.d. sample from the distribution of (X_j, Y_j), for $j = 1, \ldots, k$, and let $n = \sum_{j=1}^k n_j$.

We are interested in testing the null hypothesis of equality of the regression functions

$$H_0 : m_1 = m_2 = \cdots = m_k$$

versus the alternative

$$H_a : m_i \neq m_j \quad \text{for some } i, j \in \{1, 2, \ldots, k\}.$$

The idea of our testing procedure is to compare, in each population, the empirical distribution functions of the residuals with the same distribution function estimated under the null hypothesis. More precisely, fix one population, say j. Let $(Y_{ij} - \hat{m}_j(X_{ij}))/\hat{\sigma}_j(X_{ij})$ estimate the error ε_{ij} and let $(Y_{ij} - \hat{m}(X_{ij}))/\hat{\sigma}_j(X_{ij})$ estimate the same quantity assuming that the null hypothesis holds, where $\hat{m}_j(\cdot)$ is an appropriate kernel estimator of the regression function $m_j(\cdot)$, $\hat{m}(\cdot)$ is an estimator of the common regression function $m(\cdot)$ under H_0, and $\hat{\sigma}_j(\cdot)$ is an estimator of the variance function $\sigma_j(\cdot)$. The idea is to construct the empirical distribution functions of these estimated residuals and to compare them via Kolmogorov-Smirnov and Cramér-von Mises type statistics. Under H_0, both estimators approximate the corresponding error distribution F_{ε_j}. However, if the null hypothesis is not true, they estimate different functions.

The problem of testing for the equality of nonparametric regression curves has been widely treated in the literature. A good, recent review on this topic can be found in Neumeyer and Dette (2003). The contributions of Delgado (1993), Kulasekera (1995), Kulasekera and Wan (1997) and the aforementioned paper by Neumeyer and Dette (2003) are related to the empirical process approach we use here. These papers are mainly devoted to testing for the equality of two regression curves. In practical situations the problem of testing for the equality of more than two regression curves can arise very easily. The extension of the methods for comparing two regression curves to the comparison of more than two curves is not straightforward in many cases. If the comparison is performed pairwise, a correction in the level of the tests must be done, and consequently the power can be affected. This motivates the implementation of general procedures for more than two curves.

We propose several test statistics and establish their asymptotic distribution under H_0 and under a local alternative hypothesis of the form $m_j(\cdot) = m_0(\cdot) + n^{-1/2} r_j(\cdot)$. The rate $n^{-1/2}$ at which alternatives are detected is also achieved by the method of Neumeyer and Dette (2003), based on the comparison of two marked empirical processes of the residuals. We therefore compare our method with theirs in a simulation study.

The article is organized as follows. The testing procedure is described in Section 2 and its main asymptotic results are stated in Section 3. In Section 4 a bootstrap mechanism is introduced in order to approximate the distribution of the test statistics. Sections 5 and 6 present some simulations and an application to data. Finally, Section 7 contains the proofs.

2. Testing Procedure

Let, for $j = 1, \ldots, k$,

$$\hat{m}_j(x) = \sum_{i=1}^{n_j} W_{ij}^{(j)}(x, h_n) Y_{ij},$$
$$\hat{\sigma}_j^2(x) = \sum_{i=1}^{n_j} W_{ij}^{(j)}(x, h_n) Y_{ij}^2 - \hat{m}_j^2(x)$$

be the estimators of the regression curves and conditional variances in each population, where

$$W_{ij}^{(j)}(x, h_n) = \frac{K((x - X_{ij})h_n^{-1})}{\sum_{i'=1}^{n_j} K((x - X_{i'j})h_n^{-1})}$$

are Nadaraya-Watson type weights, K is a known kernel and h_n is an appropriate bandwidth sequence. Let

$$\hat{m}(x) = \sum_{j=1}^{k} \sum_{i=1}^{n_j} W_{ij}(x, h_n) Y_{ij}$$

be an estimator of the common regression function $m(x) = m_1(x) = \cdots = m_k(x)$ under the null hypothesis H_0, where

$$W_{ij}(x, h_n) = \frac{K((x - X_{ij})h_n^{-1})}{\sum_{j'=1}^{k} \sum_{i'=1}^{n_{j'}} K((x - X_{i'j'})h_n^{-1})}.$$

For simplicity we work with the same bandwidth h_n to estimate $\hat{m}$, $\hat{m}_j$ and $\hat{\sigma}_j$. See Section 5 for further discussion about the bandwidth choice. For $j = 1, \cdots, k$, consider the following estimators of the distributions of the errors

$$\hat{F}_{\varepsilon_j}(y) = \frac{1}{n_j} \sum_{i=1}^{n_j} I\left(\frac{Y_{ij} - \hat{m}_j(X_{ij})}{\hat{\sigma}_j(X_{ij})} \le y \right), \tag{2}$$

and the estimators of the distributions of the errors under the null hypothesis

$$\hat{F}_{\varepsilon_j 0}(y) = \frac{1}{n_j} \sum_{i=1}^{n_j} I\left(\frac{Y_{ij} - \hat{m}(X_{ij})}{\hat{\sigma}_j(X_{ij})} \le y \right). \tag{3}$$

The asymptotic properties of these estimators in nonparametric regression models have been studied in Akritas and Van Keilegom (2001). Under H_0, both $\hat{F}_{\varepsilon_j 0}$ and $\hat{F}_{\varepsilon_j}$ are estimators of F_{ε_j}. However, under the alternative hypothesis, $\hat{F}_{\varepsilon_j 0}$ and $\hat{F}_{\varepsilon_j}$ typically estimate different distributions because the residuals are calculated with respect to different curves (the true regression curve in population j and the common regression curve). If these two empirical distributions are different, there is evidence for the inequality of the regression curves.

We perform the comparison between these two estimators of the distribution of the errors in each population using the k-dimensional process $\hat{\mathbf{W}}(y) = (\hat{W}_1(y), \ldots, \hat{W}_k(y))^t$, $-\infty < y < \infty$, where, for $j = 1, \ldots, k$, $\hat{W}_j(y) = n_j^{1/2}(\hat{F}_{\varepsilon_j 0}(y) - \hat{F}_{\varepsilon_j}(y))$. More precisely, we will use the Kolmogorov-Smirnov and Cramér-von Mises type test statistics

$$T_{KS}^1 = \sum_{j=1}^{k} \sup_y |\hat{W}_j(y)| \quad \text{and} \quad T_{CM}^1 = \sum_{j=1}^{k} \int \hat{W}_j^2(y) d\hat{F}_{\varepsilon_j 0}(y).$$

We can also compare the average of the empirical distributions considered in (3),

$$\hat{F}_{\varepsilon 0}(y) = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} I\Big(\frac{Y_{ij} - \hat{m}(X_{ij})}{\hat{\sigma}_j(X_{ij})} \leq y\Big), \tag{4}$$

with the average of the empirical distributions in (2),

$$\hat{F}_{\varepsilon}(y) = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} I\Big(\frac{Y_{ij} - \hat{m}_j(X_{ij})}{\hat{\sigma}_j(X_{ij})} \leq y, \Big), \tag{5}$$

and work with the joint process $\hat{W}(y) = n^{1/2}(\hat{F}_{\varepsilon 0}(y) - \hat{F}_{\varepsilon}(y))$, a linear combination of the components of de multidimensional process $\hat{\mathbf{W}}(y)$. We propose again the Kolmogorov-Smirnov and Cramér-von Mises type test statistics

$$T_{KS}^2 = \sup_y |\hat{W}(y)| \quad \text{and} \quad T_{CM}^2 = \int \hat{W}^2(y) d\hat{F}_{\varepsilon 0}(y).$$

The procedures to test for the equality of regression curves are consistent in the sense that the equality of the regression curves is equivalent to the equality of the distribution functions we are comparing. We state this in more detail in the following theorem. Assume that $n_j/n \to p_j > 0$. Note that $\hat{m}$ consistently estimates the function $m(x) = \sum_{j=1}^{k} p_j [(f_j(x))/(f_{mix}(x))] m_j(x)$, where f_j is the density of X_j and $f_{mix}(x) = \sum_{j=1}^{k} p_j f_j(x)$ is the density of the mixture of the covariates. Consider the theoretical versions (without estimated curves) of the empirical distributions we have defined in (3), (2), (4) and (5): $F_{\varepsilon_j 0}(y) = P((Y_j - m(X_j))\sigma_j^{-1}(X_j) \leq y)$, $F_{\varepsilon_j}(y) = P((Y_j - m_j(X_j))\sigma_j^{-1}(X_j) \leq y)$, $F_{\varepsilon 0}(y) = \sum_{j=1}^k p_j P((Y_j - m(X_j))\sigma_j^{-1}(X_j) \leq y)$ and $F_{\varepsilon}(y) = \sum_{j=1}^k p_j P((Y_j - m_j(X_j))\sigma_j^{-1}(X_j) \leq y)$.

Theorem 1. *Assume that m_j is continuous, for $j = 1, \ldots, k$.*

1. $F_{\varepsilon_j 0}(y) = F_{\varepsilon_j}(y)$, $-\infty < y < \infty$, *for all* $j = 1, \ldots, k$ *if and only if* $m_1(x) = m_2(x) = \cdots = m_k(x)$, *for all* $x \in R_X$.
2. $F_{\varepsilon 0}(y) = F_{\varepsilon}(y)$, $-\infty < y < \infty$, *if and only if* $m_1(x) = m_2(x) = \cdots = m_k(x)$, *for all* $x \in R_X$.

The equivalences given in Theorem 1 are just theoretical justifications for the testing procedures we have proposed. This result involves unknown functions that are estimated in the actual testing procedures.

3. Main results

Let $F_j(y|x) = P(Y_j \leq y | X_j = x)$ and $F_j(x) = P(X_j \leq x)$, for $j = 1, \ldots, k$. We need the following regularity assumptions in order to prove our main results.

(A1) For $j = 1, \ldots, k$,
(i) X_j is absolutely continuous with compact support R_X and density f_j;
(ii) f_j, m_j and σ_j are two times continuously differentiable;
(iii) $\inf_{x \in R_X} f_j(x) > 0$ and $\inf_{x \in R_X} \sigma_j(x) > 0$.

(A2) For $j = 1, \ldots, k$,
(i) $n_j/n \to p_j > 0$;
(ii) $n_j h_n^4 \to 0$ and $n_j h_n^{3+2\delta}(\log h_n^{-1})^{-1} \to \infty$ for some $\delta > 0$.

(A3) K is a symmetric density function with compact support and K is twice continuously differentiable.

(A4) For $j = 1, \ldots, k$, $F_j(y|x)$ is continuous in (x, y) and differentiable with respect to y, $F_j'(y|x)$ is continuous in (x, y) and $\sup_{x,y} |y^2 F_j'(y|x)| < \infty$. The same holds for all other partial derivatives of $F_j(y|x)$ with respect to x and y up to order two.

Finally, let f_{ε_j} be the density corresponding to F_{ε_j}.

Theorem 2. *Assume* (A1)$-$(A4). *Then, under the null hypothesis H_0, for* $j = 1, \ldots, k$

$$\hat{F}_{\varepsilon_j 0}(y) - \hat{F}_{\varepsilon_j}(y)$$
$$= f_{\varepsilon_j}(y) \sum_{l=1}^k p_l \left\{ \frac{1}{n_l} \sum_{i=1}^{n_l} \frac{Y_{il} - m(X_{il})}{\sigma_j(X_{il})} \Big(\frac{f_j(X_{il})}{f_{mix}(X_{il})} - \frac{I(l = j)}{p_j} \Big) \right\} + o_P(n^{-\frac{1}{2}}),$$

uniformly in y.

Theorem 3. *Assume* (A1)$-$(A4). *Then, under the null hypothesis* H_0*, the* k*-dimensional process* $\hat{\boldsymbol{W}}(y) = (\hat{W}_1(y), \ldots, \hat{W}_k(y))^t$ *converges weakly to* $\boldsymbol{W}(y) = (f_{\varepsilon_1}(y)W_1, \ldots, f_{\varepsilon_k}(y)W_k)^t$*, where* $W_1, \ldots, W_k$ *are normal random variables with mean zero and covariance structure*

$$\begin{aligned}&\mathrm{Cov}\,(W_j, W_{j\prime})\\ &\quad = p_j^{\frac{1}{2}} p_{j\prime}^{\frac{1}{2}} \sum_{l=1}^{k} p_l E\Big[\frac{\sigma_l^2(X_l)}{\sigma_j(X_l)\sigma_{j\prime}(X_l)}\Big(\frac{f_j(X_l)}{f_{mix}(X_l)} - \frac{I(l=j)}{p_j}\Big)\Big(\frac{f_{j\prime}(X_l)}{f_{mix}(X_l)} - \frac{I(l=j\prime)}{p_{j\prime}}\Big)\Big].\end{aligned}$$

Corollary 4. *Assume* (A1)$-$(A4). *Then, under the null hypothesis* H_0*,*

$$\begin{aligned}T_{KS}^1 &\xrightarrow{d} \sum_{j=1}^{k} |W_j| \sup_y |f_{\varepsilon_j}(y)|, & T_{CM}^1 &\xrightarrow{d} \sum_{j=1}^{k} W_j^2 \int f_{\varepsilon_j}^2(y)dF_{\varepsilon_j}(y),\\ T_{KS}^2 &\xrightarrow{d} \sup_y |W(y)|, & T_{CM}^2 &\xrightarrow{d} \int W^2(y)dF_{\varepsilon}(y),\end{aligned}$$

where $W(y) = \sum_{j=1}^{k} p_j^{1/2} f_{\varepsilon_j}(y)W_j$ *and* $F_\varepsilon(y) = \sum_{j=1}^{k} p_j F_{\varepsilon_j}(y)$.

Consider now the limiting behavior of the test statistics under the local alternatives $H_{l.a.} : m_j = m_0 + n^{-1/2} r_j$, where the functions r_j satisfy

(A5) (i) r_j is two times continuously differentiable, for $j = 1, \ldots, k$,
(ii) $\mathrm{Var}\,[r_j(X_l)] < \infty$, for $j = 1, \ldots, k$ and $l = 1, \ldots, k$.

Theorem 5. *Assume* (A1)$-$(A5). *Then, under the alternative hypothesis* $H_{l.a.}$*, the* k*-dimensional process* $\hat{\boldsymbol{W}}(y) = (\hat{W}_1(y), \ldots, \hat{W}_k(y))^t$ *converges weakly to* $\boldsymbol{W}(y) + \boldsymbol{D}(y)$*, where* $\boldsymbol{W}(y)$ *is defined in Theorem* 3 *and* $\boldsymbol{D}(y) = (p_1^{1/2} f_{\varepsilon_1}(y)d_1, \ldots, p_k^{1/2} f_{\varepsilon_k}(y)d_k)^t$*, with*

$$d_j = E\Big[\frac{R(X_j) - r_j(X_j)}{\sigma_j(X_j)}\Big],$$

and $R(u) = \sum_{j=1}^{k} p_j [f_j(u)/f_{mix}(u)] r_j(u)$.

Corollary 6. *Assume* (A1)$-$(A5). *Then, under the alternative hypothesis* $H_{l.a.}$*,*

$$\begin{aligned}T_{KS}^1 &\xrightarrow{d} \sum_{j=1}^{k} |W_j + p_j^{\frac{1}{2}} d_j| \sup_y |f_{\varepsilon_j}(y)|, & T_{CM}^1 &\xrightarrow{d} \sum_{j=1}^{k} (W_j + p_j^{\frac{1}{2}} d_j)^2 \int f_{\varepsilon_j}^2(y)dF_{\varepsilon_j}(y),\\ T_{KS}^2 &\xrightarrow{d} \sup_y |W(y) + d(y)|, & T_{CM}^2 &\xrightarrow{d} \int (W(y) + d(y))^2 dF_{\varepsilon}(y),\end{aligned}$$

where $d(y) = \sum_{j=1}^{k} p_j f_{\varepsilon_j}(y) d_j$*, the random variables* W_j *are defined in the statement of Theorem* 3*, and* $W(y)$ *and* $F_\varepsilon(y)$ *are defined in the statement of Corollary* 4.

We can analyze in more detail the effect of the local alternatives if we consider the simpler situation of two regression curves where one of the curves is fixed and the other one varies with n. The null hypothesis is $H_0 : m_1 = m_2$ and the alternative $H_{l.a.} : m_2 = m_1 + n^{-1/2}r$. In this situation $d_1 = p_2\ E\ [(f_2(X_1)r(X_1))\ / (f_{mix}(X_1)\sigma_1(X_1))]$ and $d_2 = -p_1\ E[(f_1(X_2)r(X_2))\ /(f_{mix}(X_2)\sigma_2(X_2))]$, and these values may be zero in some cases. Nevertheless, there are important situations with consistency against alternatives converging to the null hypothesis at a rate $n^{-1/2}$, such as the one-sided alternatives (when r is a positive function). And, of course, the testing procedure is universally consistent in the sense of Theorem 1.

4. Bootstrap approximation

To apply this testing procedure in practice the asymptotic distribution of the test statistics can be used to obtain the critical values of the test. These asymptotic distributions, given in Corollary 4, can be estimated by plugging in estimators for $p_j, m, \sigma_j, F_{\varepsilon_j},\ f_{\varepsilon_j},\ f_j$ and f_{mix}. Alternatively, one can use a bootstrap procedure to approximate the distributions of the test statistics under the null hypothesis. We now consider this second option in detail.

First, for $j = 1, \ldots, k$ and $i = 1, \ldots, n_j$, estimate the residuals in a nonparametric way, using each sample separately, that is $(Y_{ij} - \hat{m}_j(X_{ij}))/\hat{\sigma}_j(X_{ij})$. These residuals are then standardized to have mean zero and variance one. Let $\tilde{F}_{\varepsilon_j}$ be the empirical distribution of the standardized residuals obtained from the jth sample.

We propose a smooth bootstrap of the residuals. Note that the asymptotic representation given in Theorem 2 involves the density of the residuals f_{ε_j}. This suggests that a smoothed version of the bootstrap of the residuals must be used. In the bootstrap of the residuals the samples are drawn from the empirical distribution, while in the smooth bootstrap the resamples are drawn from an estimate of the corresponding density. See Freedman (1981) for the bootstrap of the residuals and, e.g., Davison and Hinkley (1997) or Silverman and Young (1987) for the smoothing in the bootstrap.

The bootstrap procedure is described as follows. Let b index the bootstrap run, $b = 1, \ldots, B$.

1. For $j = 1, \ldots, k$, let $\{\varepsilon^*_{ij,b},\ i = 1, \ldots, n_j\}$ be an i.i.d. sample from the distribution of $(1 - a_j^2)^{1/2}V_j + a_jZ$, where V_j has distribution $\tilde{F}_{\varepsilon_j}$ and Z is, e.g., a standard normal random variable. The constants a_j, which determine the amount of smoothing in the bootstrap, are related to the sample size in each sample.
2. For $j = 1, \ldots, k$ and $i = 1, \ldots, n_j$, define new responses under the null hypothesis $Y^*_{ij,b} = \hat{m}(X_{ij}) + \hat{\sigma}_j(X_{ij})\varepsilon^*_{ij,b}$.

3. Let $T^{1*}_{KS,b}$, $T^{1*}_{CM,b}$, $T^{2*}_{KS,b}$ and $T^{2*}_{CM,b}$ be the test statistics obtained from the bootstrap samples $\{(X_{ij}, Y^*_{ij,b}),\ i = 1,\ldots,n_j\}$, $j = 1,\ldots,k$.

Since in Step 2 the bootstrap resamples are constructed under the null hypothesis of equal regression functions, we approximate the distribution of the test statistics under the null hypothesis. If we let $T^{1*}_{KS,(b)}$ be the order statistics of the values $T^{1*}_{KS,1},\ldots,T^{1*}_{KS,B}$ obtained in Step 3, and analogously for $T^{1*}_{CM,(b)}$, $T^{2*}_{KS,(b)}$ and $T^{2*}_{CM,(b)}$, then $T^{1*}_{KS,([(1-\alpha)B])}$, $T^{1*}_{CM,([(1-\alpha)B])}$, $T^{2*}_{KS,([(1-\alpha)B])}$ and $T^{2*}_{CM,([(1-\alpha)B])}$ approximate the $(1-\alpha)$-quantiles of the distribution of T^1_{KS}, T^1_{CM}, T^2_{KS} and T^2_{CM} under the null hypothesis, respectively.

5. Simulation study

Here we study the behavior of the bootstrap procedure by means of some simulations. In order to be able to compare with other methods in the literature, in the first part we restrict our study to the comparison of two regression curves. In particular, we compare our method with the procedure developed by Neumeyer and Dette (2003) that is based on a marked empirical process approach. The comparison is carried out by selecting the following models, considered in the simulation section of the above mentioned paper (except for model (ii), not considered there):

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

In each case, we consider a homoscedastic and a heteroscedastic situation. In the homoscedastic case, the variance functions are

$$\sigma_1^2(x) = 0.25 \quad \text{and} \quad \sigma_2^2(x) = 0.50, \tag{6}$$

and in the heteroscedastic case, the variance functions are

$$\sigma_1^2(x) = \sigma_2^2(x) = \frac{e^x}{\int_0^1 e^t dt}. \tag{7}$$

The distribution of ε_1 and ε_2 is the standard normal distribution. Other simulations have been carried out with other distributions for the errors, and similar results were obtained. In all cases the covariates X_1 and X_2 are uniformly distributed on the interval $[0, 1]$.

For the nonparametric estimation of the regression and variance curves we use the kernel of Epanechnikov: $K(u) = 0.75(1 - u^2)I(|u| < 1)$. Concerning the amount of smoothing we apply in the bootstrap, we recommend different constants a_j depending on the sample sizes n_j. We work in all cases with $a_j = 2n_j^{-3/10}$. Considered as a bandwidth, a_j chosen in this way is a small bandwidth to estimate the density of the errors (in our simulations, a standard normal). Tables register the proportion of rejections in 1,000 trials for sample sizes $(n_1, n_2) = (50, 50), (100, 50)$ and $(100, 100)$ based on $B = 200$ bootstrap replications. The significance levels are $\alpha = 0.05$ and $\alpha = 0.10$.

In the theoretical results we work with only one bandwidth h_n, and in simulations we have found better approximation of the level when the same bandwidth is used to estimate the common regression curve and the regression curves in each population, especially for 'oscillating' functions, as in model (iii). This can be explained as follows: when using different bandwidths the estimation of the regression curve in one population can be oversmoothed with respect to the estimation of the common regression function, and then the test can detect differences when the curves are really the same.

From a theoretical point of view, the optimal bandwidth order, $n^{-1/5}$, for estimating regression functions is excluded by assumption (A2-ii). To obtain the order of the optimal bandwidth, the second order terms of the representation given in Theorem 2 are needed. This is however beyond the scope of this paper. Some interesting comments about the choice of the smoothing parameter in testing problems can be found in Zhang (2003). In practice, we recommend performing the test for a reasonable range of bandwidths and studying the significance trace (see Hart (1997, p.160)), as we do in Section 6.

We have also observed that the choice of the bandwidth does not represent a big impact on the rejection probabilities. This is illustrated in Table 1, which shows results under the null hypothesis -models (ii) and (iii)- and under the alternative hypothesis -models (v) and (vii)- of the tests based on T^1_{KS} and T^1_{CM} for bandwidths of the form $h = Cn^{-3/10}$ and different values of the constant C. Note that this bandwidth satisfies the regularity conditions given in Section 2. Similar results were obtained for the other models, and for the tests based on T^2_{KS} and T^2_{CM}. In the rest of the tables we only use $C = 1$. In other situations, the value of C must be adapted to the support of the regressor variables.

Table 2 shows that the level is well-approximated in most cases. The approximation is better for the tests based on the statistics T^1_{KS} and T^1_{CM}. The tests based on T^2_{KS} and T^2_{CM} seem to be somewhat conservative. The behavior of the power (see Table 3) of the tests based on T^1_{KS} and T^1_{CM} is good for models (v), (vi) and (vii). On the other hand, the tests based on T^2_{KS} and T^2_{CM} give good power for model (viii). In both cases the power obtained is better for larger sample sizes. The Cramér-von Mises test gives better power than the Kolmogorov-Smirnov test in most situations.

Table 1. Rejection probabilities under models (ii), (iii), (v) and (vii) of the tests based on T^1_{KS} and T^1_{CM}. The models are homoscedastic, with variances given in (6). The significance level is $\alpha = 0.05$.

		T^1_{KS}			T^1_{CM}		
(n_1, n_2)	C :	0.5	1	1.5	0.5	1	1.5
$(50, 50)$	(ii)	0.053	0.051	0.056	0.053	0.054	0.054
	(iii)	0.066	0.064	0.058	0.058	0.071	0.055
	(v)	0.939	0.948	0.952	0.960	0.972	0.972
	(vii)	0.950	0.945	0.898	0.966	0.963	0.943
$(100, 50)$	(ii)	0.055	0.048	0.058	0.061	0.050	0.058
	(iii)	0.060	0.067	0.060	0.065	0.076	0.071
	(v)	0.983	0.983	0.982	0.992	0.992	0.992
	(vii)	0.986	0.977	0.964	0.990	0.984	0.979
$(100, 100)$	(ii)	0.056	0.054	0.058	0.051	0.053	0.056
	(iii)	0.049	0.058	0.058	0.050	0.061	0.057
	(v)	1.000	1.000	1.000	1.000	1.000	1.000
	(vii)	1.000	0.999	0.999	1.000	1.000	0.999

Table 2. Rejection probabilities under the null hypothesis -models (i) to (iv)- of the tests based on T^1_{KS}, T^1_{CM}, T^2_{KS} and T^2_{CM}. The models are homoscedastic, with variances given in (6), and heteroscedastic, with variances given in (7).

		T^1_{KS}		T^1_{CM}		T^2_{KS}		T^2_{CM}	
(n_1, n_2)	α :	0.050	0.100	0.050	0.100	0.050	0.100	0.050	0.100
				Homoscedastic models					
$(50, 50)$	(i)	0.051	0.093	0.052	0.105	0.056	0.104	0.049	0.101
	(ii)	0.051	0.101	0.054	0.102	0.049	0.095	0.055	0.100
	(iii)	0.064	0.110	0.071	0.122	0.047	0.097	0.066	0.121
	(iv)	0.057	0.102	0.055	0.108	0.048	0.092	0.055	0.096
$(100, 50)$	(i)	0.055	0.099	0.055	0.107	0.048	0.088	0.050	0.105
	(ii)	0.048	0.100	0.050	0.106	0.054	0.093	0.055	0.101
	(iii)	0.067	0.117	0.076	0.127	0.067	0.130	0.073	0.130
	(iv)	0.048	0.097	0.050	0.111	0.059	0.099	0.055	0.110
$(100, 100)$	(i)	0.052	0.088	0.050	0.106	0.039	0.090	0.044	0.087
	(ii)	0.054	0.088	0.053	0.107	0.044	0.077	0.044	0.092
	(iii)	0.058	0.110	0.061	0.110	0.045	0.091	0.060	0.104
	(iv)	0.052	0.095	0.059	0.104	0.044	0.086	0.049	0.095
				Heteroscedastic models					
$(50, 50)$	(i)	0.052	0.097	0.050	0.095	0.044	0.095	0.057	0.094
	(ii)	0.047	0.095	0.053	0.103	0.043	0.095	0.051	0.088
	(iii)	0.049	0.096	0.052	0.103	0.044	0.088	0.047	0.087
	(iv)	0.042	0.090	0.054	0.094	0.038	0.086	0.046	0.089
$(100, 50)$	(i)	0.054	0.099	0.052	0.114	0.042	0.085	0.049	0.087
	(ii)	0.048	0.096	0.053	0.102	0.032	0.083	0.038	0.085
	(iii)	0.060	0.116	0.063	0.121	0.055	0.101	0.041	0.090
	(iv)	0.050	0.099	0.053	0.102	0.039	0.079	0.046	0.082
$(100, 100)$	(i)	0.050	0.096	0.056	0.099	0.039	0.067	0.043	0.076
	(ii)	0.052	0.097	0.056	0.100	0.034	0.079	0.039	0.073
	(iii)	0.054	0.098	0.056	0.107	0.038	0.093	0.051	0.101
	(iv)	0.052	0.093	0.056	0.104	0.038	0.083	0.044	0.074

Table 3. Rejection probabilities under the alternative hypothesis -models (v) to (viii)- of the tests based on T^1_{KS}, T^1_{CM}, T^2_{KS} and T^2_{CM}. The models are homoscedastic, with variances given in (6), and heteroscedastic, with variances given in (7).

		T^1_{KS}		T^1_{CM}		T^2_{KS}		T^2_{CM}	
(n_1, n_2)	α :	0.050	0.100	0.050	0.100	0.050	0.100	0.050	0.100
				Homoscedastic models					
(50, 50)	(v)	0.948	0.978	0.972	0.986	0.698	0.816	0.902	0.951
	(vi)	0.950	0.973	0.969	0.986	0.702	0.802	0.903	0.950
	(vii)	0.945	0.972	0.963	0.977	0.639	0.779	0.874	0.941
	(viii)	0.213	0.370	0.158	0.312	0.688	0.795	0.844	0.922
(100, 50)	(v)	0.983	0.994	0.992	0.998	0.858	0.917	0.966	0.983
	(vi)	0.983	0.990	0.991	0.995	0.839	0.901	0.968	0.985
	(vii)	0.977	0.986	0.984	0.994	0.809	0.890	0.953	0.973
	(viii)	0.286	0.436	0.180	0.342	0.777	0.869	0.901	0.960
(100, 100)	(v)	1.000	1.000	1.000	1.000	0.969	0.989	0.997	0.999
	(vi)	1.000	1.000	1.000	1.000	0.964	0.983	0.997	0.998
	(vii)	0.999	1.000	1.000	1.000	0.952	0.980	0.993	0.998
	(viii)	0.430	0.647	0.324	0.562	0.961	0.985	0.998	0.998
				Heteroscedastic models					
(50, 50)	(v)	0.596	0.717	0.638	0.762	0.233	0.369	0.359	0.479
	(vi)	0.583	0.712	0.643	0.753	0.233	0.359	0.363	0.489
	(vii)	0.586	0.701	0.626	0.748	0.223	0.355	0.347	0.469
	(viii)	0.122	0.199	0.086	0.167	0.376	0.535	0.562	0.692
(100, 50)	(v)	0.760	0.854	0.816	0.886	0.341	0.486	0.492	0.639
	(vi)	0.740	0.847	0.815	0.881	0.343	0.481	0.489	0.632
	(vii)	0.749	0.844	0.810	0.870	0.321	0.463	0.465	0.598
	(viii)	0.178	0.302	0.136	0.240	0.513	0.645	0.709	0.808
(100, 100)	(v)	0.899	0.952	0.928	0.962	0.445	0.583	0.649	0.767
	(vi)	0.912	0.953	0.929	0.962	0.440	0.578	0.645	0.763
	(vii)	0.890	0.951	0.920	0.958	0.428	0.566	0.639	0.753
	(viii)	0.213	0.341	0.136	0.261	0.673	0.782	0.882	0.932

The method proposed by Neumeyer and Dette (2003) is based on the marked empirical processes (with $l = 1, 2$) $\hat{R}^l(x) = n^{-1}\sum_{i=1}^{n_1} e^l_{i1} I(X_{i1} \le x) - n^{-1}\sum_{i=1}^{n_2} e^l_{i2}$ $I(X_{i2} \le x)$, where $e^1_{ij} = (n_{3-j}/n)(Y_{ij} - \hat{m}(X_{ij}))\hat{f}_{mix}(X_{ij})\hat{f}_{3-j}(X_{ij})$ and $e^2_{ij} = (n/n_j)(Y_{ij} - \hat{m}(X_{ij}))\hat{f}_j^{-1}(X_{ij})$, for $j = 1, 2$. The test statistics are $K^l = \sup_x$ $|R^l(x)|$ $(l = 1, 2)$ and their distributions are approximated by a wild bootstrap procedure. For the sake of comparison we show, in Table 4, the results for the tests based on K^1 and K^2 using the same samples as in the previous tables. Since the level is well-approximated, we only show the results corresponding to the power behavior. For the estimation of the needed functions, bandwidth selection, and bootstrap, we have kept exactly the setting described in Neumeyer and Dette (2003). We found that our procedure based on T^1_{KS} and T^1_{CM} yields better or similar results for the power in most cases for models (v), (vi) and (vii), whereas for model (viii) we obtained better results with the tests based on T^2_{KS} and T^2_{CM}.

Table 4. Rejection probabilities under the alternative hypothesis -models (v) to (viii)- of the tests based on K^1 and K^2 (Neumeyer and Dette (2003)). The models are homoscedastic, with variances given in (6), and heteroscedastic, with variances given in (7).

		Homoscedastic models				*Heteroscedastic models*			
		K^1		K^2		K^1		K^2	
(n_1, n_2)	α :	0.050	0.100	0.050	0.100	0.050	0.100	0.050	0.100
$(50, 50)$	(v)	0.935	0.960	0.888	0.920	0.616	0.720	0.625	0.723
	(vi)	0.933	0.961	0.823	0.871	0.616	0.721	0.593	0.682
	(vii)	0.921	0.956	0.928	0.959	0.602	0.690	0.638	0.738
	(viii)	0.697	0.819	0.479	0.635	0.335	0.494	0.206	0.347
$(100, 50)$	(v)	0.971	0.983	0.943	0.963	0.753	0.823	0.745	0.826
	(vi)	0.966	0.983	0.895	0.924	0.754	0.825	0.697	0.792
	(vii)	0.966	0.980	0.973	0.984	0.732	0.814	0.781	0.850
	(viii)	0.738	0.866	0.572	0.718	0.419	0.573	0.268	0.440
$(100, 100)$	(v)	0.999	1.000	0.992	0.995	0.902	0.945	0.889	0.933
	(vi)	0.998	1.000	0.961	0.970	0.903	0.944	0.863	0.910
	(vii)	0.999	1.000	1.000	1.000	0.885	0.933	0.908	0.953
	(viii)	0.967	0.996	0.873	0.937	0.657	0.800	0.480	0.648

Our method is valid for more than two curves. We explore now the behavior of the testing procedure in a three regression curves setup. We consider the models

$$(ix)\ m_1(x) = m_2(x) = m_3(x) = x;$$
$$(x)\ m_1(x) = x,\ m_2(x) = x + 0.25,\ m_3(x) = x + 0.5;$$
$$(xi)\ m_1(x) = x,\ m_2(x) = 0.5,\ m_3(x) = 1 - x.$$

The variance functions are

$$\sigma_1^2(x) = \sigma_2^2(x) = \sigma_3^2(x) = 0.5. \tag{8}$$

As in the previous simulated models, the covariates are uniformly distributed in [0, 1] and the errors are distributed as a standard normal. The choice of the kernel, the bandwidth and the amount of smoothing in the bootstrap is the same as in the previous simulations: K is the Epanechnikov kernel, $h = n^{-3/10}$ and $a_j = 2n_j^{-3/10}$. The results obtained with samples sizes $(50, 50, 50)$, $(100, 50, 50)$, $(100, 100, 50)$ and $(100, 100, 100)$ and significance levels $\alpha = 0.05$ and $\alpha = 0.10$ are shown in Table 5. The tests based on T^1_{KS} and T^1_{CM} approximate the level well, while the tests based on T^2_{KS} and T^2_{CM} seem a bit conservative, as happened in the two-curve examples. In all cases the power increases with sample size. The first version of the test statistics produces better power in model (x) and the second version gives better power in model (xi), as happened with model (viii). In these two models -(viii) and (xi)- all the distributions we are comparing have expectation zero, and the differences appear in other features of the distributions.

Table 5. Rejection probabilities under models (ix) to (xi) of the tests based on T^1_{KS}, T^1_{CM}, T^2_{KS} and T^2_{CM}. The models are homoscedastic, with variances given in (8).

		T^1_{KS}		T^1_{CM}		T^2_{KS}		T^2_{CM}	
(n_1, n_2, n_3)	α :	0.050	0.100	0.050	0.100	0.050	0.100	0.050	0.100
(50, 50, 50)	(ix)	0.049	0.088	0.051	0.096	0.041	0.068	0.036	0.072
	(x)	0.818	0.886	0.890	0.929	0.274	0.391	0.430	0.555
	(xi)	0.089	0.190	0.077	0.146	0.290	0.434	0.475	0.601
(100, 50, 50)	(ix)	0.054	0.090	0.058	0.097	0.043	0.086	0.031	0.083
	(x)	0.937	0.975	0.975	0.991	0.420	0.541	0.596	0.710
	(xi)	0.140	0.232	0.106	0.193	0.462	0.596	0.662	0.763
(100, 100, 50)	(ix)	0.050	0.098	0.057	0.107	0.039	0.073	0.047	0.094
	(x)	0.929	0.958	0.940	0.962	0.365	0.516	0.841	0.912
	(xi)	0.171	0.283	0.090	0.176	0.420	0.569	0.495	0.616
(100, 100, 100)	(ix)	0.056	0.109	0.061	0.119	0.055	0.088	0.059	0.100
	(x)	0.989	0.996	0.993	0.998	0.523	0.677	0.760	0.850
	(xi)	0.201	0.311	0.133	0.236	0.576	0.708	0.835	0.907

Also note that all the simulation results are based on fixed alternatives. When working with local alternatives converging to the null hypothesis at a rate $n^{-1/2}$, models (viii) and (xi) produce no power (see Theorem 5).

6. Application to data

We illustrate our testing procedure by means of data from the Data Archive of the Journal of Applied Econometrics, consisting of monthly expenditures of several Dutch households. The data are registered in Dutch guilders and correspond to the period from October 1986 to September 1987. Einmahl and Van Keilegom (2007) verified that model (1) holds when X=‘log of the total monthly expenditure’ is considered as a covariate and Y=‘log of the expenditure on food’ is the response variable (even a homoscedastic model is verified).

We compare the regression curves for three groups of households: households consisting of two members (159 in total), three members (45 in total) and four members (73 in total). Figure 1 shows the scatter plots and estimated regression curves based on the cross-validation bandwidths. We have transformed the support of the covariates to the interval $[0, 1]$ and performed the test for a wide range of bandwidths, going from 0.15 to 0.35. The p-values are based on 1,000 bootstrap replications. Since the results obtained with the test statistics T^1_{KS} and T^1_{CM} are very similar to those obtained with T^2_{KS} and T^2_{CM}, we only discuss the first ones.

When testing for the equality of the three regression curves we obtained p-values smaller than 0.003 for T^1_{KS} and smaller than 0.001 for T^1_{CM} for any bandwidth considered in the range. There is a strong evidence for the inequality of the three regression curves.

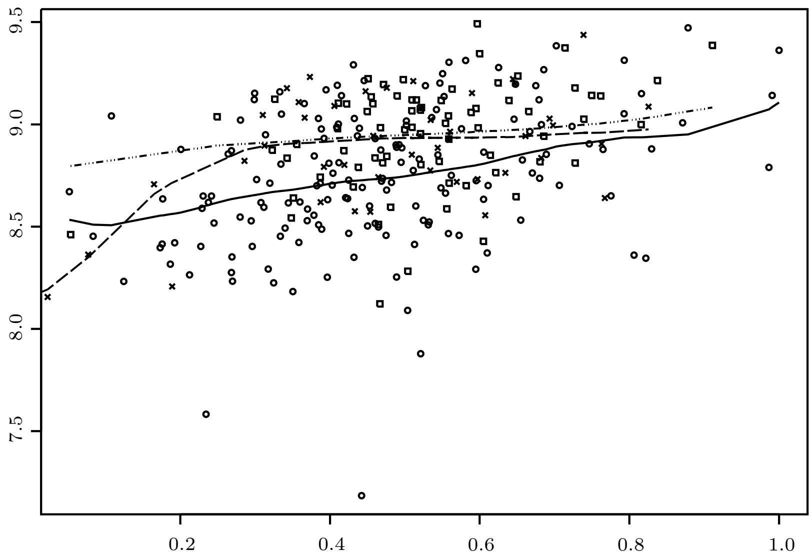

Figure 1. Scatter plot of 'log(food expenditure)' versus 'log(total expenditure)' and estimated regression curves of households consisting of two members (circles and solid line), three members (crosses and dashed line) and four members (squares and dashed dotted line).

It makes sense then to test for the equality of each set of two curves. When we test for the equality of the regression curves corresponding to households consisting of two and three members the p-values are smaller than 0.02 for T^1_{KS}, and smaller than 0.002 for T^1_{CM}. More extreme p-values (all of them smaller than 0.001) were observed when testing for households of two and four members. However the p-values for households consisting of three and four members were between 0.32 and 0.61 for T^1_{KS}, and between 0.53 and 0.68 for T^1_{CM}. It seems that the regression curves are the same when households of three and four members are considered (note that on the left side of Figure 1 the curves seem quite different, but there are very few points in that area and they make a small contribution to the test statistics). All the results are summarized in Figure 2.

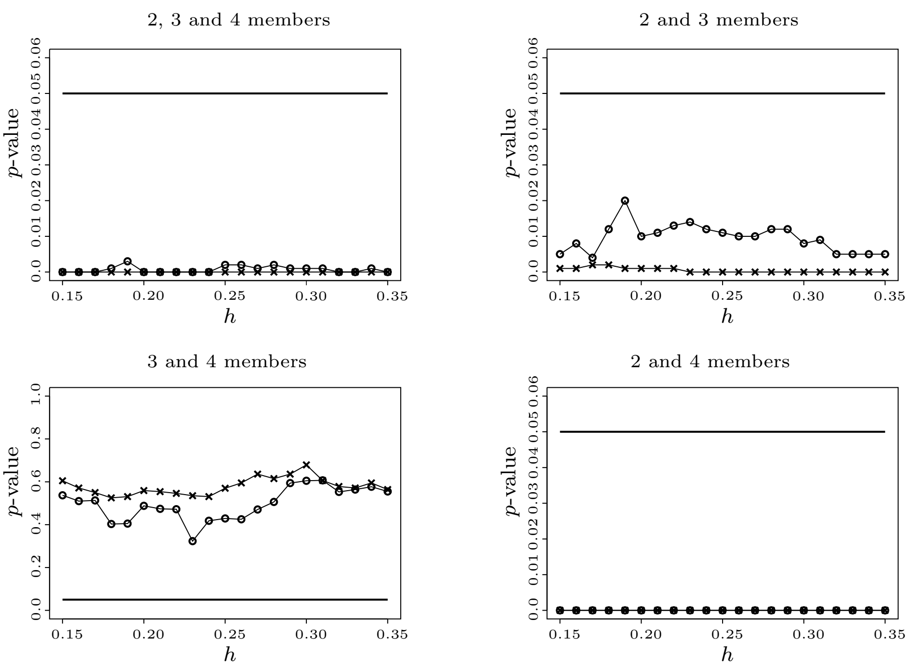

Figure 2. Graphics of the p-values as function of the bandwidth h obtained from 1,000 bootstrap replications with the test statistics T^1_{KS} (line with circles) and T^1_{CM} (line with crosses). The solid horizontal line corresponds to a p-value of 0.05.

7. Proofs

Proof of Theorem 1. 1st part. Assume $F_{\varepsilon_j 0}(y) = F_{\varepsilon_j}(y)$. This implies that the first and the second moment of these distributions are the same. From the first moment we have that $E[(Y_j - m(X_j))/\sigma_j(X_j)] = E[(Y_j - m_j(X_j))/\sigma_j(X_j)] = 0$. The second moment of $F_{\varepsilon_j 0}(y)$ can be written as $\mathrm{Var}\,[(Y_j - m(X_j))/\sigma_j(X_j)] = \mathrm{Var}\,[(Y_j - m_j(X_j))/\sigma_j(X_j)] + E[(m_j(X_j) - m(X_j))^2/\sigma_j^2(X_j)]$. We are assuming that $\mathrm{Var}\,[(Y_j - m(X_j))\sigma_j^{-1}(X_j)] = \mathrm{Var}\,[(Y_j - m_j(X_j))\sigma_j^{-1}(X_j)]$, hence $E((m_j(X_j) - m(X_j))^2\sigma_j^{-2}(X_j)) = 0$, and this implies $m_j(x) = m(x)$, for all $j = 1, \ldots, k$, and for all $x \in R_X$, except for a set of points of probability zero. The continuity of m_j allows one to extend the equality to all $x \in R_X$.

2nd part. The result is obtained by equating the first and second moments of $F_{\varepsilon 0}(y)$ and $F_\varepsilon(y)$, as in the 1st part.

The converse implications are trivial.

Before proving the main results in Section 3, we state three lemmas.

Lemma 7. *Assume* (A1)$-$(A3). *Then, under the null hypothesis* H_0*, for any*

$j \in \{1, \ldots, k\}$,

$$\int \frac{\hat{m}(x) - m(x)}{\sigma_j(x)} f_j(x) dx = \frac{1}{n} \sum_{l=1}^{k} \sum_{i=1}^{n_l} \frac{Y_{il} - m(X_{il})}{\sigma_j(X_{il})} \frac{f_j(X_{il})}{f_{mix}(X_{il})} + o_P(n^{-\frac{1}{2}}).$$

Proof. Let $\hat{f}_{mix}(x) = (nh_n)^{-1} \sum_{l=1}^{k} \sum_{i=1}^{n_l} K((x - X_{il})h_n^{-1})$ be the kernel estimator of the density of the mixture $f_{mix}(x)$. By properties of the kernel estimator (see for example Wand and Jones (1995)) and assumption (A2-ii), we have that $\hat{f}_{mix}(x) - f_{mix}(x) = O_P((nh_n)^{-1/2})$ and $\hat{f}_{mix}(x) f_{mix}^{-1}(x) - 1 = O_P((nh_n)^{-1/2})$. Similarly $\hat{m}(x) - m(x) = O_P((nh_n)^{-1/2})$. Then we obtain

$$\hat{m}(x) - m(x) = \frac{1}{nh_n f_{mix}(x)} \sum_{l=1}^{k} \sum_{i=1}^{n_l} K\Big(\frac{x - X_{il}}{h_n}\Big)(Y_{il} - m(x)) + O_P((nh_n)^{-1})),$$

uniformly in x. Taking this into account with condition (A2-ii), the integral becomes

$$\int \frac{\hat{m}(x) - m(x)}{\sigma_j(x)} f_j(x) dx$$
$$= \frac{1}{nh_n} \sum_{l=1}^{k} \sum_{i=1}^{n_l} \int \frac{K((x - X_{il})h_n^{-1})(Y_{il} - m(x))}{\sigma_j(x)} \frac{f_j(x)}{f_{mix}(x)} dx + o_P(n^{-\frac{1}{2}}).$$

Let $L(x) = (Y_{il} - m(x)) f_j(x) (f_{mix}(x)\sigma_j(x))^{-1}$. Using the change of variable $u = (x - X_{il})h_n^{-1}$, a Taylor expansion of second order of L around X_{il}, and assumption (A3), we immediately obtain the result.

Lemma 8. *Assume* (A1)–(A3). *Then, for any* $j \in \{1, \ldots, k\}$,

$$\int \frac{\hat{m}_j(x) - m_j(x)}{\sigma_j(x)} f_j(x) dx = \frac{1}{n_j} \sum_{i=1}^{n_j} \frac{Y_{ij} - m_j(X_{ij})}{\sigma_j(X_{ij})} + o_P(n_j^{-\frac{1}{2}}).$$

Proof. The proof is similar to that of the previous lemma.

Lemma 9. *Assume* (A1)–(A3). *Then, for any* $j \in \{1, \ldots, k\}$

$$\int \frac{\hat{\sigma}_j(x) - \sigma_j(x)}{\sigma_j(x)} f_j(x) dx = \frac{1}{n_j} \sum_{i=1}^{n_j} \frac{(Y_{ij} - m_j(X_{ij}))^2 - \sigma_j^2(X_{ij})}{2\sigma_j^2(X_{ij})} + o_P(n_j^{-\frac{1}{2}}).$$

Proof. Write

$$\hat{\sigma}_j(x) - \sigma_j(x) = \frac{\hat{\sigma}_j^2(x) - \sigma_j^2(x)}{2\sigma_j(x)} - \frac{(\hat{\sigma}_j(x) - \sigma_j(x))^2}{2\sigma_j(x)}.$$

From Proposition 3 in Akritas and Van Keilegom (2001), the second term is $O_P((n_jh_n)^{-1})$, and hence we can write $\hat\sigma_j(x)-\sigma_j(x) = (\hat\sigma_j^2(x)-\sigma_j^2(x))/(2\sigma_j(x))+O_P((n_jh_n)^{-1})$, uniformly in x. Since $\hat m_j(x)-m_j(x) = O_P((n_jh_n)^{-1/2})$, it is easy to see that $\hat\sigma_j^2(x) = \tilde\sigma_j^2(x)+O_P((n_jh_n)^{-1})$, where $\tilde\sigma_j^2(x) = \sum_{i=1}^{n_j} W_{ij}^{(j)}(x,h_n)(Y_{ij}-m_j(x))^2$. If we consider $\sigma_j^2(x)$ as the regression function of the variable $(Y_{ij}-m_j(x))^2$, we have that $\tilde\sigma_j^2(x)-\sigma_j^2(x) = O_P((n_jh_n)^{-1/2})$ and

$$\tilde\sigma_j^2(x)-\sigma_j^2(x) = \frac{1}{n_jh_nf_j(x)}\sum_{i=1}^{n_j}K\Big(\frac{x-X_{ij}}{h_n}\Big)[(Y_{ij}-m_j(x))^2-\sigma_j^2(x)]+O_P((n_jh_n)^{-1}).$$

Using the previous expression and (A2-ii), we obtain

$$\int\frac{\hat\sigma_j(x)-\sigma_j(x)}{\sigma_j(x)}f_j(x)dx = \frac{1}{n_jh_n}\sum_{i=1}^{n_j}\int\frac{K((x-X_{ij})h_n^{-1})((Y_{ij}-m_j(x))^2-\sigma_j^2(x))}{2\sigma_j^2(x)}dx+o_P(n_j^{-\frac12}).$$

Using a Taylor expansion of second order, we obtain the representation given in the statement of the lemma.

Proof of Theorem 2. Write

$$\hat F_{\varepsilon_j0}(y)-\hat F_{\varepsilon_j}(y) = (\hat F_{\varepsilon_j0}(y)-F_{\varepsilon_j}(y))-(\hat F_{\varepsilon_j}(y)-F_{\varepsilon_j}(y)). \tag{9}$$

First we study the asymptotic behavior of $\hat F_{\varepsilon_j0}(y)-F_{\varepsilon_j}(y)$. We use some results and proofs from Akritas and Van Keilegom (2001). These authors assume that the functions m, m_j, and σ_j are L-functionals depending on a certain score function J. In our case the functionals are the conditional mean and variance that correspond to $J\equiv 1$. This choice of J is not covered by the results in Akritas and Van Keilegom (2001). However, it is easy to check that the results in that paper can be suitably extended. From the proof of Theorem 1 in Akritas and Van Keilegom (2001), we have that

$$\begin{aligned}\hat F_{\varepsilon_j0}(y)-F_{\varepsilon_j}(y) &= \frac{1}{n_j}\sum_{i=1}^{n_j}I\Big(\frac{Y_{ij}-m(X_{ij})}{\sigma_j(X_{ij})}\le y\Big)-F_{\varepsilon_j}(y)\\ &\quad+f_{\varepsilon_j}(y)\int\frac{y(\hat\sigma_j(x)-\sigma_j(x))+\hat m(x)-m(x)}{\sigma_j(x)}f_j(x)dx+R_{n_j}(y),\end{aligned} \tag{10}$$

where $\sup_y|R_{n_j}(y)| = o_P(n_j^{-1/2})$. Note that in Akritas and Van Keilegom (2001) the estimation of the distribution of the residuals is considered from one sample. This means that, with our notation, the error ε_{ij} is estimated by $(Y_{ij}-$

$\hat{m}_j(X_{ij}))/\hat{\sigma}_j(X_{ij})$. However, the decomposition given in (10) remains valid when the errors are estimated with $\hat{m}$, because their Lemma 1 holds in that case.

Using Lemma 7, Lemma 9, and the fact that under the null hypothesis $m = m_1 = \cdots = m_k$,

$$\begin{aligned}\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j}(y) = &\frac{1}{n_j}\sum_{i=1}^{n_j} I\Big(\frac{Y_{ij} - m(X_{ij})}{\sigma_j(X_{ij})} \le y\Big) - F_{\varepsilon_j}(y) \\ &+ y f_{\varepsilon_j}(y)\frac{1}{n_j}\sum_{i=1}^{n_j}\frac{(Y_{ij} - m(X_{ij}))^2 - \sigma_j^2(X_{ij})}{2\sigma_j^2(X_{ij})} \\ &+ f_{\varepsilon_j}(y)\frac{1}{n}\sum_{l=1}^{k}\sum_{i=1}^{n_l}\frac{Y_{il} - m(X_{il})}{\sigma_j(X_{il})}\frac{f_j(X_{il})}{f_{mix}(X_{il})} + o_P(n^{-\frac{1}{2}}) \qquad (11)\end{aligned}$$

uniformly in y. Analogously, from the proof of Theorem 1 in Akritas and Van Keilegom (2001) and Lemmas 8 and 9,

$$\begin{aligned}\hat{F}_{\varepsilon_j}(y) - F_{\varepsilon_j}(y) = &\frac{1}{n_j}\sum_{i=1}^{n_j} I\Big(\frac{Y_{ij} - m(X_{ij})}{\sigma_j(X_{ij})} \le y\Big) - F_{\varepsilon_j}(y) \\ &+ y f_{\varepsilon_j}(y)\frac{1}{n_j}\sum_{i=1}^{n_j}\frac{(Y_{ij} - m(X_{ij}))^2 - \sigma_j^2(X_{ij})}{2\sigma_j^2(X_{ij})} \\ &+ f_{\varepsilon_j}(y)\frac{1}{n_j}\sum_{i=1}^{n_j}\frac{Y_{ij} - m(X_{ij})}{\sigma_j(X_{ij})} + o_P(n_j^{-\frac{1}{2}}) \qquad (12)\end{aligned}$$

uniformly in y. The representation given in the statement of the theorem follows from (9), (11) and (12).

Proof of Theorem 3. The Cramér-Wold device (see e.g., Serfling (1980)) ensures that the weak convergence of a multidimensional process is equivalent to the weak convergence of any linear combination of its components. Consider then a linear combination of the components of the process $\hat{\mathbf{W}}(y)$, say $\hat{V}(y) = \sum_{j=1}^{k} a_j n_j^{1/2}(\hat{F}_{\varepsilon_j 0}(y) - \hat{F}_{\varepsilon_j}(y))$. Using the representation given in Theorem 2 it is not difficult to obtain that

$$\sum_{j=1}^{k} a_j n_j^{\frac{1}{2}}(\hat{F}_{\varepsilon_j 0}(y) - \hat{F}_{\varepsilon_j}(y)) = \sum_{l=1}^{k}\frac{1}{n_l^{\frac{1}{2}}}\sum_{i=1}^{n_l}\psi_l(X_{il}, Y_{il}, y) + o_P(1),$$

where

$$\psi_l(u, v, y) = \frac{v - m(u)}{\sigma_l(u)}\Big(\sum_{j=1}^{k} a_j p_l^{\frac{1}{2}} p_j^{\frac{1}{2}} f_{\varepsilon_j}(y)\frac{\sigma_l(u)}{\sigma_j(u)}\frac{f_j(u)}{f_{mix}(u)} - a_l f_{\varepsilon_l}(y)\Big).$$

Write $\hat{V}_l(y) = n_l^{-1/2}\sum_{i=1}^{n_l}\psi_l(X_{il}, Y_{il}, y)$, for $l = 1, \ldots, k$. Consider the class of functions $\mathcal{F}_l = \{(u, v) \longrightarrow \psi_l(u, v, y), -\infty < y < \infty\}$. The process $\hat{V}_l(y)$ is the $\mathcal{F}_l$-indexed process (see page 80 in van der Vaart and Wellner (1996)). In general, for any classes of functions $\mathcal{G}_1$ and $\mathcal{G}_2$, define $\mathcal{G}_1 + \mathcal{G}_2 = \{g_1 + g_2; g_1 \in \mathcal{G}_1, g_2 \in \mathcal{G}_2\}$. With this notation the class $\mathcal{F}_l$ can be written as $\mathcal{F}_1 = \sum_{j=1}^{k+1}\mathcal{F}_{lj}$, where, for $j = 1, \ldots, k$,

$$\mathcal{F}_{lj} = \Big\{(u, v) \longrightarrow a_j p_l^{\frac{1}{2}} p_j^{\frac{1}{2}} f_{\varepsilon_j}(y)\frac{\sigma_l(u)}{\sigma_j(u)}\frac{f_j(u)}{f_{mix}(u)}\frac{v - m(u)}{\sigma_l(u)}, -\infty < y < \infty\Big\},$$

$$\mathcal{F}_{l,k+1} = \Big\{(u, v) \longrightarrow -a_l f_{\varepsilon_l}(y)\frac{v - m(u)}{\sigma_l(u)}, -\infty < y < \infty\Big\}.$$

All these classes follow the same pattern, they factorize in a part not depending on y and a bounded function of y. Let M be such that $\sup_{y,j=1,\ldots,k}|f_{\varepsilon_j}(y)| < M$. Then $N_{[]}(\delta, \mathcal{F}_{lj}, L_2(P)) \le 2M\delta^{-1}$ if $\delta < 2M$ and $N_{[]}(\delta, \mathcal{F}_{lj}, L_2(P)) = 1$ if $\delta > 2M$, where $N_{[]}$ is the bracketing number, P is the probability measure corresponding to the joint distribution of (X_l, Y_l) and $L_2(P)$ is the L_2-norm.

Theorem 2.10.6 in van der Vaart and Wellner (1996) ensures that $\log N_{[]}(\delta, \mathcal{F}_l, L_2(P)) \le \sum_{j=1}^{k+1}\log N_{[]}(\delta, \mathcal{F}_{lj}, L_2(P))$, and consequently the integral $\int_0^\infty \sqrt{\log N_{[]}(\delta, \mathcal{F}_l, L_2(P))}d\delta$ is finite. Then, by Theorem 2.5.6. in van der Vaart and Wellner (1996), the class of functions $\mathcal{F}_l$ is Donsker. The weak convergence of the process $\hat{V}_l(y)$ now follows from pages 81 and 82 of the aforementioned book. The limit process, $V_l(y)$, is a zero-mean Gaussian process with covariance function $\mathrm{Cov}\,(V_l(y), V_l(y')) = E[\psi_l(X_l, Y_l, y)\psi_l(X_l, Y_l, y')]$.

Our process of interest can be written as $\hat{V}(y) = \sum_{l=1}^{k}\hat{V}_l(y)$, and the processes $\hat{V}_l(y)$ are independent. So, using the first part of this proof, we can conclude that $\hat{V}(y)$ converges weakly to a zero-mean Gaussian process $V(y)$ with covariance function $\mathrm{Cov}\,(V(y), V(y')) = \sum_{l=1}^{k}E[\psi_l(X_l, Y_l, y)\psi_l(X_l, Y_l, y')]$.

Applying the Cramér-Wold device, we obtain the weak convergence of the k-dimensional process $\hat{\mathbf{W}}(y)$. Note as well that the representation given in Theorem 2 for $\hat{W}_j(y)$ factorizes in a deterministic component $f_{\varepsilon_j}(y)$ and a sum of independent random variables with mean zero not depending on y. Therefore $\hat{\mathbf{W}}(y)$ converges to $(f_{\varepsilon_1}(y)W_1, \ldots, f_{\varepsilon_k}(y)W_k)^t$, where $W_1, \ldots, W_k$ are normal random variables with mean zero and covariance structure given in the statement of the theorem.

Proof of Corollary 4. The weak convergence of $\hat{\mathbf{W}}(y)$ given in Theorem 3 ensures the weak convergence of each of its components to $f_{\varepsilon_j}(y)W_j$. The convergence of T_{KS}^1 follows directly from the Continuous Mapping Theorem. For T_{CM}^1 we write

$$\int f_{\varepsilon_j}(y)^2 d\hat{F}_{\varepsilon_j 0}(y) = \int f_{\varepsilon_j}(y)^2 dF_{\varepsilon_j 0}(y) + \int f_{\varepsilon_j}(y)^2 d(\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j 0}(y)).$$

Using integration by parts we obtain

$$\Big|\int f_{\varepsilon_j}(y)^2 d(\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j 0}(y))\Big| \le \sup_y |\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j}(y)| \sup_y |f'_{\varepsilon_j}(y)| = o_P(1),$$

since $\sup_y |\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j}(y)| = o_P(1)$ due to Theorem 2 in Akritas and Van Keilegom (2001), and $\sup_y |f'_{\varepsilon_j}(y)| < \infty$ due to assumption (A4). Hence $\int f^2_{\varepsilon_j}(y) d\hat{F}_{\varepsilon_j 0}(y) = \int f^2_{\varepsilon_j}(y) dF_{\varepsilon_j}(y) + o_P(1)$. This concludes the proof of the convergence of T^1_{CM}.

The process $\hat{W}(y)$ can be expressed as $\hat{W}(y) = \hat{V}(y) + o_P(1)$, where $\hat{V}(y)$ is a particular linear combination of the components of $\hat{\mathbf{W}}(y)$ as considered at the start of the proof of Theorem 3, putting $a_j = p_j^{1/2}$. The corresponding limit process is the mean-zero Gaussian process $W(y)$ defined in the statement of the corollary. As in the first part of this proof, the Continuous Mapping Theorem ensures the convergence of T^2_{KS}.

For T^2_{CM}, it suffices to show that $d\hat{F}_\varepsilon(y)$ can be replaced by $dF_\varepsilon(y)$. Using the weak convergence of the processes $\hat{W}(y)$ and $n^{1/2}(\hat{F}_{\varepsilon 0}(y) - F_\varepsilon(y))$, and the Skorohod construction (see Serfling (1980)) we can write

$$\sup_y |\hat{W}(y) - W(y)| \to_{a.s.} 0 \quad \text{and} \quad \sup_y |\hat{F}_{\varepsilon 0}(y) - F_{\varepsilon 0}(y)| \to_{a.s.} 0 \tag{13}$$

(we use for simplicity the same notation as for the original processes). Now

$$\begin{aligned}
&\Big|\int \hat{W}^2(y) d\hat{F}_{\varepsilon 0}(y) - \int W_0^2(y) dF_{\varepsilon 0}(y)\Big| \\
&\quad \le \Big|\int (\hat{W}^2(y) - W^2(y)) d\hat{F}_{\varepsilon 0}(y)\Big| + \Big|\int W^2(y) d(\hat{F}_{\varepsilon 0}(y) - F_{\varepsilon 0}(y))\Big|.
\end{aligned}$$

The first term of the right hand side of the above inequality is $o(1)$ a.s. due to the first expression in (13). For the second term, taking into account the second expression in (13), and since the trajectories of the limit process $W_0(y)$ are bounded and continuous almost surely, we can apply the Helly-Bray Theorem (see e.g., Rao (1965, p.97)), to each of these trajectories and conclude that $|\int W^2(y) d(\hat{F}_{\varepsilon 0}(y) - F_{\varepsilon 0}(y))| \to_{a.s.} 0$. This concludes the proof of Corollary 4.

Proof of Theorem 5. Under $H_{l.a.}$, $\hat{m}(x)$ estimates $m_n(x) = m_0(x) + n^{-1/2}R(x)$, where $R(x) = \sum_{j=1}^k p_j [f_j(x)/f_{mix}(x)] r_j(x)$, $\hat{m}_j$ estimates $m_{jn}(x) = m_0(x) + n^{-1/2} r_j(x)$ and $\hat{F}_{\varepsilon_j 0}(y)$ estimates $F_{\varepsilon_j 0}(y) = P((Y_j - m_n(X_j))\sigma_j^{-1}(X_j) \le y)$. Considering the following probability as a function of y, and applying a Taylor expansion, we obtain

$$F_{\varepsilon_j 0}(y) = P\Big(\frac{Y_j - m_{jn}(X_j)}{\sigma_j(X_j)} - n^{-\frac{1}{2}} \frac{R(X_j) - r_j(X_j)}{\sigma_j(X_j)} \le y\Big)$$

$$= \int P\Big(\frac{Y_j - m_{jn}(X_j)}{\sigma_j(X_j)} - n^{-\frac{1}{2}} \frac{R(X_j) - r_j(X_j)}{\sigma_j(X_j)} \le y \Big| X_j = x\Big) f_j(x)dx$$
$$= F_{\varepsilon_j}(y) + n^{-\frac{1}{2}} f_{\varepsilon_j}(y) E\Big[\frac{R(X_j) - r_j(X_j)}{\sigma_j(X_j)}\Big] + o(n^{-\frac{1}{2}}). \tag{14}$$

Following the same steps as in the proof of Theorem 2,

$$\begin{aligned}
&\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j 0}(y) \\
&\quad = \frac{1}{n_j} \sum_{i=1}^{n_j} I\Big(\frac{Y_{ij} - m_n(X_{ij})}{\sigma_j(X_{ij})} \le y\Big) - F_{\varepsilon_j 0}(y) \\
&\qquad + f_{\varepsilon_j}(y) \int \frac{y(\hat{\sigma}_j(x) - \sigma_j(x))}{\sigma_j(x)} f_j(x)dx \\
&\qquad + f_{\varepsilon_j}(y) \int \frac{\hat{m}(x) - m_n(x)}{\sigma_j(x)} f_j(x)dx + o_P(n^{-\frac{1}{2}}).
\end{aligned} \tag{15}$$

An application of the proof of Lemma 1 of Akritas and Van Keilegom (2001) shows that

$$\begin{aligned}
&\sup_y \Big| \frac{1}{n_j} \sum_{i=1}^{n_j} \Big\{ I\Big(\frac{Y_{ij} - m_{jn}(X_{ij})}{\sigma_j(X_{ij})} - n^{-\frac{1}{2}} \frac{R(X_{ij}) - r_j(X_{ij})}{\sigma_j(X_{ij})} \le y\Big) \\
&\qquad - I\Big(\frac{Y_{ij} - m_{jn}(X_{ij})}{\sigma_j(X_{ij})} \le y\Big) \\
&\qquad - P\Big(\frac{Y_j - m_{jn}(X_j)}{\sigma_j(X_j)} - n^{-\frac{1}{2}} \frac{R(X_j) - r_j(X_j)}{\sigma_j(X_j)} \le y\Big) \\
&\qquad + P\Big(\frac{Y_j - m_{jn}(X_j)}{\sigma_j(X_j)} \le y\Big) \Big\} \Big| \\
&\quad = o_P(n^{-\frac{1}{2}}).
\end{aligned}$$

Taking (14) into account, we can write

$$\begin{aligned}
&\frac{1}{n_j} \sum_{i=1}^{n_j} I\Big(\frac{Y_{ij} - m_n(X_{ij})}{\sigma_j(X_{ij})} \le y\Big) \\
&\quad = \frac{1}{n_j} \sum_{i=1}^{n_j} I\Big(\frac{Y_{ij} - m_{jn}(X_{ij})}{\sigma_j(X_{ij})} \le y\Big) \\
&\qquad + n^{-\frac{1}{2}} f_{\varepsilon_j}(y) E\Big[\frac{R(X_j) - r_j(X_j)}{\sigma_j(X_j)}\Big] + o_P(n^{-\frac{1}{2}}).
\end{aligned} \tag{16}$$

As in the proof of Theorem 3 we have

$$
\begin{aligned}
&\hat{F}_{\varepsilon_j}(y) - F_{\varepsilon_j}(y) \\
&\quad = \frac{1}{n_j}\sum_{i=1}^{n_j} I\Big(\frac{Y_{ij} - m_{jn}(X_{ij})}{\sigma_j(X_{ij})} \le y\Big) - F_{\varepsilon_j}(y) \\
&\qquad + f_{\varepsilon_j}(y)\int \frac{y(\hat{\sigma}_j(x) - \sigma_j(x))}{\sigma_j(x)} f_j(x)dx \\
&\qquad + f_{\varepsilon_j}(y)\int \frac{\hat{m}_j(x) - m_{jn}(x)}{\sigma_j(x)} f_j(x)dx + o_P(n^{-\frac{1}{2}}). \qquad (17)
\end{aligned}
$$

By combining (15), (16) and (17) we obtain

$$
\begin{aligned}
&\hat{F}_{\varepsilon_j 0}(y) - \hat{F}_{\varepsilon_j}(y) \\
&\quad = f_{\varepsilon_j}(y)\int \frac{\hat{m}(x) - m_n(x)}{\sigma_j(x)} f_j(x)dx \\
&\qquad - f_{\varepsilon_j}(y)\int \frac{\hat{m}_j(x) - m_{jn}(x)}{\sigma_j(x)} f_j(x)dx \\
&\qquad + n^{-\frac{1}{2}} f_{\varepsilon_j}(y) E\Big[\frac{R(X_j) - r_j(X_j)}{\sigma_j(X_j)}\Big] + o_P(n^{-\frac{1}{2}}),
\end{aligned}
$$

and following a similar development as in Lemmas 8 and 9, we can write

$$
\begin{aligned}
\hat{F}_{\varepsilon_j 0}(y) - \hat{F}_{\varepsilon_j}(y) &= f_{\varepsilon_j}(y)\sum_{l=1}^{k} p_l \Big\{\frac{1}{n_l}\sum_{i=1}^{n_l} \frac{Y_{il} - m_{ln}(X_{il})}{\sigma_j(X_{il})}\Big(\frac{f_j(X_{il})}{f_{mix}(X_{il})} - \frac{I(l=j)}{p_j}\Big)\Big\} \\
&\quad + n^{-\frac{1}{2}} f_{\varepsilon_j}(y) E\Big[\frac{R(X_j) - r_j(X_j)}{\sigma_j(X_j)}\Big] + o_P(n^{-\frac{1}{2}}). \qquad (18)
\end{aligned}
$$

Note that $(Y_l - m_{ln}(X_l))/\sigma_j(X_{il}) = \varepsilon_l$, which does not depend on n. The leading term of the representation given in (18) under $H_{l.a.}$ is the same as the leading term of the representation given in Theorem 3 under the null hypothesis. Therefore their limit distributions are the same. This concludes the proof.

Proof of Corollary 6. The proof is similar to that of Corollary 4, taking into account the weak convergence of $\hat{\mathbf{W}}(y)$ under $H_{l.a.}$ given in Theorem 5.

Acknowledgements

The research of Juan Carlos Pardo-Fernández is supported by Ministerio de Educación y Ciencia (project MTM2005-00820, with additional European FEDER support), Vicerreitorado de Investigación of the Universidade de Vigo and Dirección Xeral de Investigación e Desenvolvemento of the Xunta de Galicia. The research of Ingrid Van Keilegom is supported by IAP research network nr. P5/24 of the Belgian government (Belgian Science Policy). The research of Wenceslao González-Manteiga is supported Ministerio de Educación y Ciencia (project MTM2005-00820, with additional European FEDER support) and Xunta de Galicia (project PGIDIT03PXIC20702PN). This paper benefitted from the comments of two referees, an associate editor and the Co-Editor.

References

Akritas, M. G. and Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scand. J. Statist.* **28**, 549-568.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Applications.* Cambridge University Press.

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

Einmahl, J. and Van Keilegom, I. (2007). Specification tests in nonparametric regression. *J. Econometrics.* To appear.

Freedman, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9**, 1218-1228.

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests.* Springer, New York.

Kulasekera, K. B. (1995). Comparison of regression curves using quasi-residuals. *J. Amer. Statist. Assoc.* **90**, 1085-1093.

Kulasekera, K. B. and Wang, J. (1997). Smoothing parameter selection for power optimality in testing of regression curves. *J. Amer. Statist. Assoc.* **92**, 500-511.

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

Rao, C. R. (1965). *Linear Statistical Inference and its Applications.* Wiley, New York.

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

Silverman, B. W. and Young, G. A. (1987). The bootstrap: To smooth or not to smooth? *Biometrika* **74**, 469-479.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing.* Chapman and Hall, London.

Zhang, C. (2003). Calibrating the degrees of freedom for automatic data smoothing and effective curve checking. *J. Amer. Statist. Assoc.* **98**, 609-628.

Departamento de Estatística e IO, Universidade de Vigo. EU Enxeñería Técnica Industrial, Rúa Torrecedeira 86, Vigo (36208), Spain.

E-mail: juancp@uvigo.es

Institut de Statistique, Université catholique de Louvain. Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium.

E-mail: vankeilegom@stat.ucl.ac.be

Departamento de Estatística e IO, Universidade de Santiago de Compostela. Facultade de Matemáticas, Campus Universitario Sur, Santiago de Compostela (15782), Spain.

E-mail: wenceslao@usc.es

(Received January 2005; accepted February 2006)