T E C H N I C A L

R E P O R T

0417

COMPARISON OF REGRESSION CURVES BASED ON THE ESTIMATION OF THE ERROR DISTRIBUTION

J.C. PARDO-FERNANDEZ, I. VAN KEILEGOM and W. GONZALEZ-MANTEIGA

I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Comparison of regression curves based on the estimation of the error distribution

Juan Carlos Pardo-Fernández [*] Ingrid Van Keilegom [†]
Wenceslao González-Manteiga [‡]

June 9, 2004

Abstract

Assume that (X_1, Y_1) and (X_2, Y_2) are two independent random vectors, that satisfy the non-parametric regression models $Y_j = m_j(X_j) + \sigma_j(X_j)\varepsilon_j$, for $j = 1, 2$, where $m_j(X_j) = E(Y_j|X_j)$ and $\sigma_j^2(X_j) = Var(Y_j|X_j)$ are smooth but unknown regression and variance functions respectively, and the error variable ε_j is independent of X_j. Suppose that ε_1 and ε_2 have the same distribution.

In this paper we introduce a procedure to test the hypothesis $H_0 : m_1(\cdot) = m_2(\cdot)$ of equality of the regression functions. The test is based on the comparison of the empirical distribution of the errors in each model. A Kolmogorov-Smirnov type statistic and a Cramér-von Mises type statistic are considered, and their asymptotic distribution is obtained. The proposed tests can detect local alternatives converging to the null hypothesis at the rate $n^{-1/2}$. We describe a bootstrap procedure in order to approximate the critical values of the test, and present the results of a simulation study, in which the behavior of the tests for small and moderate sample sizes is studied. An application to a real data set is included.

Key Words: Bootstrap; Goodness-of-fit; Heteroscedastic regression; Nonparametric regression.

[*]Departamento de Estatística e I.O. Universidade de Vigo. Financial support from the Spanish Ministry of Science and Technology (with additional European FEDER support) through the project BFM2002-03213 is gratefully acknowledged.

[†]Institut de Statistique. Université catholique de Louvain. Financial support from the IAP research network nr. P5/24 of the Belgian government (Belgian Science Policy) is gratefully acknowledged.

[‡]Departamento de Estatística e I.O. Universidade de Santiago de Compostela. Financial support from the Spanish Ministry of Science and Technology (with additional European FEDER support) through the project BFM2002-03213 and from Xunta de Galicia through project PGIDIT03PXIC20702PN are gratefully acknowledged.

1 Introduction

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_1, Y_1) and (X_2, Y_2) be two independent random vectors, and assume that they satisfy the following non-parametric regression models, for $j = 1, 2$,

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

where the error variable ε_j is independent of X_j, $m_j(X_j) = E(Y_j|X_j)$ is the unknown regression function and $\sigma_j^2(X_j) = Var(Y_j|X_j)$ is the conditional variance function. Suppose that ε_1 and ε_2 have common distribution F_ε, and that X_1 and X_2 have common support R_X. Let (X_{ij}, Y_{ij}), $i = 1, ..., n_j$, be an i.i.d. sample from the distribution of (X_j, Y_j), for $j = 1, 2$. Denote $n = n_1 + n_2$.

We are interested in testing the hypothesis of equality of the two regression functions:

$$H_0 : m_1 = m_2 \qquad \text{versus} \qquad H_a : m_1 \neq m_2. \tag{2}$$

Recent contributions on this problem are Dette and Neumeyer (2001), Speckman et al. (2003), Neumeyer and Dette (2003a,b) among others. A good and recent review on this topic can be found in Neumeyer and Dette (2003b).

The idea of the testing procedure proposed in this paper is to compare the empirical distribution functions of the residuals of the two samples under H_0. More precisely, let $(Y_{ij} - \hat{m}(X_{ij}))/\hat{\sigma}_j(X_{ij})$ estimate the error ε_{ij} under H_0 ($j = 1, 2, i = 1, \ldots, n_j$), where $\hat{m}(\cdot)$ and $\hat{\sigma}_j(\cdot)$ are appropriate kernel estimators of the joint regression function $m(\cdot)$ under H_0 and of the variance function $\sigma_j(\cdot)$. The idea is to construct the empirical distribution functions of these residuals for the two samples and to compare them via Kolmogorov-Smirnov and Cramér-von Mises type statistics. Under H_0, the two estimators both approximate the common error distribution. We establish the asymptotic distribution of these test statistics under H_0 and under a local alternative hypothesis of the form $m_2(\cdot) = m_1(\cdot) + n^{-1/2}r(\cdot)$ for some function $r(\cdot)$. The rate $n^{-1/2}$ at which alternatives are detected is also achieved by the method of Neumeyer and Dette (2003b). We therefore compare our method with theirs in a small simulation study.

The paper 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 real data. In Section 7 we present a test to verify the equality of the distributions of the errors. Finally, Section 8 contains the proofs.

2 Testing procedure

The testing procedure is based on the comparison of two non-parametric estimators of the distribution of the errors F_ε. Let

$$\hat{m}\,(x) = \sum_{j=1}^{2}\sum_{i=1}^{n_j} W_{ij}(x, h_n) Y_{ij} \tag{3}$$

be an estimator of the common regression function $m(x)$ under the null hypothesis H_0, where

$$W_{ij}(x, h_n) = \frac{K((x - X_{ij})h_n^{-1})}{\sum_{j'=1}^{2}\sum_{i'=1}^{n_{j'}} K((x - X_{i'j'})h_n^{-1})}$$

are Nadaraya-Watson type weights, K is a known kernel and h_n is an appropriate bandwidth sequence. Let, for $j = 1, 2$,

$$\hat{\sigma}_j^2(x) = \sum_{i=1}^{n_j} W_{ij}^{(j)}(x, h_n) Y_{ij}^2 - (\hat{m}_j(x))^2 \tag{4}$$

be the estimators of the conditional variances, where

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

and

$$\hat{m}_j(x) = \sum_{i=1}^{n_j} W_{ij}^{(j)}(x, h_n) Y_{ij} \tag{5}$$

(for simplicity we work with the same bandwidth h_n for the estimators $\hat{m}$, $\hat{m}_j$ and $\hat{\sigma}_j$).

For $j = 1, 2$, consider the following estimators of the distribution of the errors

$$\hat{F}_{\varepsilon_j 0}(y) = \frac{1}{n_j}\sum_{i=1}^{n_j} I\left(\frac{Y_{ij} - \hat{m}(X_{ij})}{\hat{\sigma}_j(X_{ij})} \leq y\right). \tag{6}$$

The asymptotic properties of these estimators have been studied in Akritas and Van Keilegom (2001). Under H_0, both $\hat{F}_{\varepsilon_1 0}$ and $\hat{F}_{\varepsilon_2 0}$ are estimators of F_ε. However, under the alternative hypothesis $\hat{F}_{\varepsilon_1 0}$ and $\hat{F}_{\varepsilon_2 0}$ estimate in general different distributions because the errors are calculated with respect to a curve in between the two true regression functions. Therefore, the fact that these two empirical distributions are different gives evidence for the inequality of the regression curves. We propose the following two test statistics: a Kolmogorov-Smirnov type statistic

$$T_{KS} = \sqrt{n} \sup_y |\hat{F}_{\varepsilon_1 0}(y) - \hat{F}_{\varepsilon_2 0}(y)| \tag{7}$$

and a Cramér-von Mises type statistic

$$T_{CM} = n \int (\hat{F}_{\varepsilon_1 0}(y) - \hat{F}_{\varepsilon_2 0}(y))^2 d\hat{F}_{\varepsilon 0}(y), \tag{8}$$

where

$$\hat{F}_{\varepsilon 0}(y) = \frac{1}{n} \sum_{j=1}^{2} \sum_{i=1}^{n_j} I\left(\frac{Y_{ij} - \hat{m}(X_{ij})}{\hat{\sigma}_j(X_{ij})} \leq y \right).$$

Note that the proposed testing procedure is based on the assumption that the errors ε_1 and ε_2 are identically distributed. This assumption is quite common in the context of testing for hypothesis H_0 [see e.g. Dette and Neumeyer (2001), Neumeyer and Dette (2003b) and Speckman *et al.* (2003)], and can easily be verified by using a testing procedure similar to the above one. We will discuss this in Section 7.

Also note that if we are interested in testing the null hypothesis (2) against one-sided alternatives

$$H_a : m_1 \geq m_2 \text{ and } m_1(x) > m_2(x) \text{ for some } x, \tag{9}$$

we can use a Kolmogorov-Smirnov type statistic without absolute values.

3 Main results

Let $F_j(y|x) = P(Y_j \leq y | X_j = x)$ and $F_{X_j}(x) = P(X_j \leq x)$ $(j = 1, 2)$. We need the following regularity assumptions in order to prove our main results:

(A1) For $j = 1, 2$,
(i) X_j is absolutely continuous with compact support R_X and density f_{X_j}.
(ii) f_{X_j}, m_j and σ_j are two times continuously differentiable.
(iii) $\inf_{x \in R_X} f_{X_j}(x) > 0$ and $\inf_{x \in R_X} \sigma_j(x) > 0$.

(A2)
(i) $n_1/n \to p_1 > 0$, $n_2/n \to p_2 > 0$, $p_1 + p_2 = 1$.
(ii) For $j = 1, 2$, $n_j h_n^4 \to 0$ 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 $j = 1,2$, $F_j(y|x)$ is continuous in (x,y), 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.

Let f_ε be the density corresponding to F_ε and let X_{mix} be the random variable with density $f_{mix}(x) = p_1 f_{X_1}(x) + p_2 f_{X_2}(x)$. This variable is a mixture of X_1 and X_2.

Theorem 1 *Assume (A1)-(A4). Then, under the null hypothesis H_0,*

$$\begin{aligned}
&\hat{F}_{\varepsilon 1 0}(y) - \hat{F}_{\varepsilon 2 0}(y) \\
&= n_1^{-1}\sum_{i=1}^{n_1}\left\{ I\left(\frac{Y_{i1} - m(X_{i1})}{\sigma_1(X_{i1})} \le y\right) + \varphi_1(X_{i1}, Y_{i1}, y)\right\} \\
&- n_2^{-1}\sum_{i=1}^{n_2}\left\{ I\left(\frac{Y_{i2} - m(X_{i2})}{\sigma_2(X_{i2})} \le y\right) + \varphi_2(X_{i2}, Y_{i2}, y)\right\} + o_P(n^{-1/2}),
\end{aligned}$$

uniformly in y, where

$$\begin{aligned}
\varphi_j(u,v,y) &= y f_\varepsilon(y)\frac{(v - m(u))^2 - \sigma_j^2(u)}{2\sigma_j^2(u)} \\
&\quad + (-1)^{j-1} p_j f_\varepsilon(y)\frac{v - m(u)}{f_{mix}(u)}\left(\frac{f_{X_1}(u)}{\sigma_1(u)} - \frac{f_{X_2}(u)}{\sigma_2(u)}\right).
\end{aligned} \tag{10}$$

Note that under the null hypothesis, the function φ_j satisfies $E[\varphi_j(X_j, Y_j, y)] = 0$. Also note that when $f_{X_1} \equiv f_{X_2}$, this representation does not depend on the (common) design density. When in addition the two variance functions are equal, the second term of φ_j vanishes.

Theorem 2 *Assume (A1)-(A4). Then, under the null hypothesis H_0, the process*

$$\hat{W}_0(y) = n^{1/2}(\hat{F}_{\varepsilon 1 0}(y) - \hat{F}_{\varepsilon 2 0}(y))$$

converges weakly to a zero-mean Gaussian process $W_0(y)$ with covariance function

$$\begin{aligned}
&Cov(W_0(y), W_0(y')) \\
&= \sum_{j=1}^{2} p_j^{-1} E\left(\left[I\left(\frac{Y_j - m(X_j)}{\sigma_j(X_j)} \le y\right) + \varphi_j(X_j, Y_j, y)\right]\right. \\
&\qquad \left.\times \left[I\left(\frac{Y_j - m(X_j)}{\sigma_j(X_j)} \le y'\right) + \varphi_j(X_j, Y_j, y')\right]\right) \\
&- \left(p_1^{-1} + p_2^{-1}\right) F_\varepsilon(y) F_\varepsilon(y').
\end{aligned}$$

Corollary 3 *Assume (A1)-(A4). Then, under the null hypothesis H_0,*

$$\begin{aligned} T_{KS} &\xrightarrow{d} \quad \sup_y |W_0(y)| \\ T_{CM} &\xrightarrow{d} \quad \int (W_0(y))^2 dF_\varepsilon(y). \end{aligned}$$

Consider now the limiting behavior of T_{KS} and T_{CM} under the following local alternative:

$$H_{l.a.} : m_{2n} = m_1 + n^{-1/2} r,$$

where r is a function satisfying

$$Var[r(X_j)] < \infty \qquad (j = 1, 2). \tag{11}$$

Theorem 4 *Assume (A1)-(A4) and (11). Then under the alternative hypothesis $H_{l.a.}$*

$$\begin{aligned} &\hat{F}_{\varepsilon_1 0}(y) - \hat{F}_{\varepsilon_2 0}(y) \\ &= n_1^{-1} \sum_{i=1}^{n_1} \left\{ I\left(\frac{Y_{i1} - m_1(X_{i1})}{\sigma_1(X_{i1})} \le y \right) + \tilde{\varphi}_1(X_{i1}, Y_{i1}, y) \right\} - n^{-1/2} d_1 f_\varepsilon(y) \\ &- n_2^{-1} \sum_{i=1}^{n_2} \left\{ I\left(\frac{Y_{i2} - m_{2n}(X_{i2})}{\sigma_2(X_{i2})} \le y \right) + \tilde{\varphi}_2(X_{i2}, Y_{i2}, y) \right\} - n^{-1/2} d_2 f_\varepsilon(y) \\ &+ o_P(n^{-1/2}), \end{aligned}$$

uniformly in y, where $\tilde{\varphi}_1$ and $\tilde{\varphi}_2$ are defined by replacing $m(\cdot)$ by $m_1(\cdot)$ and $m_{2n}(\cdot)$ in the definition of φ_1 and φ_2 respectively,

$$d_j = p_j E\left[\frac{\tilde{r}_j(X_j)}{f_{mix}(X_j)} \left(\frac{f_{X_1}(X_j)}{\sigma_1(X_j)} - \frac{f_{X_2}(X_j)}{\sigma_2(X_j)} \right) \right] + (-1)^{j-1} E\left[\frac{\tilde{r}_j(X_j)}{\sigma_j(X_j)} \right]$$

$(j = 1, 2)$, $\tilde{r}_1(u) = p_2 \frac{f_{X_2}(u)}{f_{mix}(u)} r(u)$ and $\tilde{r}_2(u) = -p_1 \frac{f_{X_1}(u)}{f_{mix}(u)} r(u)$.

Corollary 5 *Assume (A1)-(A4) and (11). Then, under the alternative hypothesis $H_{l.a.}$,*

$$\begin{aligned} T_{KS} &\xrightarrow{d} \quad \sup_y |W_0(y) - d(y)| \\ T_{CM} &\xrightarrow{d} \quad \int (W_0(y) - d(y))^2 dF_\varepsilon(y), \end{aligned}$$

where $d(y) = (d_1 + d_2) f_\varepsilon(y)$.

4 Bootstrap approximation

To apply this testing procedure in practice the asymptotic distribution of the test statistics T_{KS} and T_{CM} can be used to obtain the critical values of the test. These asymptotic distributions, given in Corollary 3, can be estimated by plugging in estimators for $p_j, m, \sigma_j, F_\varepsilon$ and f_ε (and if the design densities are unequal also f_{X_j} and f_{mix}) ($j = 1, 2$). Alternatively, one can use a bootstrap procedure in order to approximate the distributions of the test statistics under the null hypothesis. We will now consider this second option in detail.

First, for $j = 1, 2$ and $i = 1, ..., n_j$, estimate the residuals in a non-parametric way, using the two samples separately, that is

$$\frac{Y_{ij} - \hat{m}_j(X_{ij})}{\hat{\sigma}_j(X_{ij})}. \tag{12}$$

These residuals are then standardized in order to have zero mean and variance one. Let $\tilde{F}_{\varepsilon_j}$ be the empirical distribution of these standardized residuals obtained from each sample.

We propose a smooth bootstrap of residuals [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 can be described with the following steps. For fixed B and for $b = 1, ..., B$,

1. For $j = 1$ and 2, let $\{\varepsilon^*_{ij,b}, i = 1, \ldots, n_j\}$ be an i.i.d. sample from the distribution of $(1 - a_j^2)^{1/2} V_j + a_j Z$, 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 $j = 1$ and 2, define now new responses under the null hypothesis $Y^*_{ij,b}$ by ($i = 1, \ldots, n_j$)
$$Y^*_{ij,b} = \hat{m}(X_{ij}) + \hat{\sigma}_j(X_{ij})\varepsilon^*_{ij,b} \tag{13}$$

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

If we denote $T^*_{KS,(b)}$ and $T^*_{CM,(b)}$ for the order statistics of the statistics obtained in step 3, then $T^*_{KS,([(1-\alpha)B])}$ and $T^*_{CM,([(1-\alpha)B])}$ approximate the $(1 - \alpha)$-quantiles of the distribution of T_{KS} and T_{CM} under the null hypothesis respectively.

5 Simulation study

In this section we study the practical behavior of the proposed bootstrap procedure by means of some simulations. We also compare our method with the procedure developed by Neumeyer and Dette (2003b), which is based on a marked empirical process approach. The comparison is carried out for a selection of the models considered in the simulation section of Neumeyer and Dette's paper (except for model (ii), which is not considered in their paper). More precisely, we consider the following models:

(i)	$m_1(x) = 1;$	$m_2(x) = 1$
(ii)	$m_1(x) = x;$	$m_2(x) = x$
(iii)	$m_1(x) = \sin(2\pi x);$	$m_2(x) = \sin(2\pi x)$
(iv)	$m_1(x) = \exp(x);$	$m_2(x) = \exp(x)$

(v)	$m_1(x) = 1;$	$m_2(x) = 1 + x$
(vi)	$m_1(x) = \exp(x);$	$m_2(x) = \exp(x) + x$
(vii)	$m_1(x) = \sin(2\pi x);$	$m_2(x) = \sin(2\pi x) + x$
(viii)	$m_1(x) = 1;$	$m_2(x) = 1 + \sin(2\pi x)$

Clearly, models $(i)-(iv)$ correspond to the null hypothesis and scenarios $(v)-(viii)$ to the alternative hypothesis. In each case, we consider a homoscedastic and a heteroscedastic situation. In the homoscedastic case the variance functions are (Neumeyer and Dette (2003b))

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

and in the heteroscedastic case the variance functions are

$$\sigma_1^2(x) = \sigma_2^2(x) = \frac{e^x}{\int_0^1 e^x dx}.$$

The common distribution of ε_1 and ε_2 is the standard normal distribution. Similar simulations have been carried out with other distributions for the errors and very similar results were obtained. In all cases the covariates X_1 and X_2 have a uniform distribution on $[0, 1]$.

As we mentioned above, in the theoretical results we work with only one bandwidth for simplicity. Nevertheless, as it was pointed out in Dette and Neumeyer (2001), it is recommended to choose different bandwidths to estimate the variance and regression curves according to the size of the corresponding sample. More precisely, we consider a bandwidth $h = Cn^{-3/10}$ to estimate m under the null hypothesis and two bandwidths $h_j = Cn_j^{-3/10}$ to estimate m_j and σ_j. The bandwidths chosen in this way verify the regularity conditions given in Section 2. In the simulations the cases $C = 1$ and $C = 2$ are presented. In other situations the value of C must be adapted to the support of the regressor variables.

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 1-4 register the proportion of rejections in 1000 trials for sample sizes $(n_1, n_2) = (50, 50), (100, 50)$ and $(100, 100)$ and $B = 200$ bootstrap replications. The significance levels are $\alpha = 0.05$ and $\alpha = 0.10$.

		$C = 1$				$C = 2$			
		T_{KS}		T_{CM}		T_{KS}		T_{CM}	
		α		α		α		α	
(n_1, n_2)	Model	0.050	0.100	0.050	0.100	0.050	0.100	0.050	0.100
$(50, 50)$	*(i)*	0.054	0.101	0.047	0.083	0.057	0.105	0.046	0.095
	(ii)	0.054	0.106	0.046	0.088	0.063	0.116	0.046	0.098
	(iii)	0.056	0.102	0.045	0.095	0.060	0.097	0.046	0.101
	(iv)	0.052	0.102	0.045	0.088	0.066	0.109	0.047	0.095
$(100, 50)$	*(i)*	0.058	0.103	0.056	0.104	0.049	0.102	0.055	0.105
	(ii)	0.061	0.112	0.053	0.101	0.053	0.105	0.052	0.107
	(iii)	0.062	0.124	0.059	0.115	0.060	0.113	0.063	0.113
	(iv)	0.059	0.108	0.054	0.100	0.054	0.103	0.057	0.108
$(100, 100)$	*(i)*	0.050	0.106	0.053	0.097	0.056	0.100	0.061	0.099
	(ii)	0.053	0.101	0.054	0.103	0.057	0.109	0.062	0.104
	(iii)	0.070	0.127	0.065	0.108	0.062	0.115	0.076	0.115
	(iv)	0.060	0.111	0.058	0.110	0.062	0.116	0.061	0.114

Table 1. Rejection probabilities under the null hypothesis. The models are homoscedastic with variance functions $\sigma_1^2(x) = 0.25$ *and* $\sigma_2^2(x) = 0.50$.

		$C = 1$				$C = 2$			
		T_{KS}		T_{CM}		T_{KS}		T_{CM}	
		α		α		α		α	
(n_1, n_2)	Model	0.050	0.100	0.050	0.100	0.050	0.100	0.050	0.100
$(50, 50)$	(v)	0.953	0.973	0.970	0.982	0.955	0.975	0.971	0.985
	(vi)	0.944	0.968	0.967	0.983	0.936	0.958	0.953	0.975
	(vii)	0.923	0.963	0.949	0.973	0.847	0.903	0.893	0.936
	$(viii)$	0.070	0.128	0.051	0.097	0.049	0.087	0.048	0.083
$(100, 50)$	(v)	0.965	0.985	0.983	0.993	0.972	0.986	0.985	0.993
	(vi)	0.959	0.981	0.984	0.990	0.943	0.974	0.967	0.983
	(vii)	0.952	0.976	0.977	0.991	0.922	0.954	0.945	0.969
	$(viii)$	0.062	0.115	0.063	0.105	0.059	0.098	0.060	0.106
$(100, 100)$	(v)	0.999	1.000	1.000	1.000	0.999	1.000	1.000	1.000
	(vi)	0.998	1.000	1.000	1.000	0.998	1.000	0.999	1.000
	(vii)	0.996	1.000	1.000	1.000	0.984	0.991	0.996	0.997
	$(viii)$	0.083	0.143	0.073	0.144	0.075	0.138	0.071	0.130

Table 2. Rejection probabilities under the alternative hypothesis. The models are homoscedastic with variance functions $\sigma_1^2(x) = 0.25$ *and* $\sigma_2^2(x) = 0.50$.

		$C = 1$				$C = 2$			
		T_{KS}		T_{CM}		T_{KS}		T_{CM}	
		α		α		α		α	
(n_1, n_2)	Model	0.050	0.100	0.050	0.100	0.050	0.100	0.050	0.100
$(50, 50)$	(i)	0.054	0.116	0.049	0.091	0.059	0.110	0.049	0.095
	(ii)	0.052	0.116	0.048	0.090	0.066	0.117	0.046	0.097
	(iii)	0.054	0.103	0.050	0.096	0.051	0.109	0.048	0.108
	(iv)	0.054	0.111	0.049	0.092	0.067	0.119	0.044	0.095
$(100, 50)$	(i)	0.054	0.105	0.055	0.098	0.045	0.102	0.050	0.097
	(ii)	0.051	0.103	0.050	0.098	0.045	0.099	0.051	0.091
	(iii)	0.056	0.106	0.053	0.094	0.065	0.112	0.054	0.108
	(iv)	0.058	0.111	0.054	0.097	0.055	0.088	0.053	0.094
$(100, 100)$	(i)	0.059	0.103	0.053	0.103	0.064	0.099	0.056	0.101
	(ii)	0.059	0.105	0.054	0.103	0.062	0.106	0.056	0.110
	(iii)	0.068	0.108	0.057	0.106	0.075	0.110	0.068	0.103
	(iv)	0.065	0.107	0.055	0.105	0.056	0.113	0.062	0.109

Table 3. Rejection probabilities under the null hypothesis. The models are heteroscedastic with variance functions $\sigma_1^2(x) = \sigma_2^2(x) = e^x / \int_0^1 e^x dx$

		$C=1$				$C=2$			
		T_{KS}		T_{CM}		T_{KS}		T_{CM}	
		α		α		α		α	
(n_1, n_2)	Model	0.050	0.100	0.050	0.100	0.050	0.100	0.050	0.100
$(50, 50)$	(v)	0.549	0.657	0.590	0.720	0.552	0.668	0.613	0.723
	(vi)	0.552	0.667	0.596	0.711	0.544	0.666	0.587	0.710
	(vii)	0.550	0.665	0.589	0.699	0.505	0.631	0.552	0.672
	$(viii)$	0.069	0.125	0.057	0.120	0.052	0.109	0.056	0.101
$(100, 50)$	(v)	0.662	0.779	0.740	0.829	0.659	0.768	0.745	0.825
	(vi)	0.668	0.773	0.732	0.826	0.646	0.759	0.731	0.817
	(vii)	0.670	0.762	0.725	0.814	0.638	0.745	0.701	0.795
	$(viii)$	0.087	0.141	0.084	0.144	0.073	0.133	0.075	0.123
$(100, 100)$	(v)	0.820	0.898	0.869	0.921	0.822	0.897	0.879	0.920
	(vi)	0.818	0.888	0.874	0.919	0.808	0.884	0.868	0.919
	(vii)	0.810	0.871	0.866	0.914	0.776	0.857	0.850	0.907
	$(viii)$	0.078	0.140	0.097	0.153	0.074	0.149	0.084	0.150

Table 4. Rejection probabilities under the alternative hypothesis. The models are heteroscedastic with variance functions $\sigma_1^2(x) = \sigma_2^2(x) = e^x / \int_0^1 e^x dx$

Tables 1 and 3 show that the level is well approximated in most cases and for both bandwidth choices. The behavior of the power (Tables 2 and 4) is good for models (v), (vi) and (vii). The obtained power is better for the larger sample sizes and in the homoscedastic case. The Cramér-von Mises test gives better power than the Kolmogorov-Smirnov test in most situations. Also note that the choice of the bandwidth has little impact on the rejection probabilities. In Neumeyer and Dette (2003b) only the homoscedastic case is considered, and for a comparison with their simulations we found that our procedure yields better or similar results for the power in most cases.

Model (viii) deserves some special attention. The power under this model is barely higher than the level of the test. This can be explained as follows. Under this model the regression functions are anti-symmetric with respect to the center of the support of the X's and so $\hat{F}_{\varepsilon_1 0}$ and $\hat{F}_{\varepsilon_2 0}$ both estimate the distribution of a zero-mean normal random variable, only the variances are (slightly) different. Hence, for this model it is hard to distinguish the two distributions. Note that this is also clear from Corollary 5.

One solution to this problem consists in carrying out the testing procedure for several subregions of the support of the X's. It suffices to reject H_0 in one of the considered subregions in order to conclude that the regression curves are different. To control the (simultaneous) levels of the tests, a Bonferroni correction of the level of each of the tests can be used.

To apply this technique to model (*viii*), split the interval $[0, 1]$ in $[0, 1/2]$ and $[1/2, 1]$, and set the level equal to 0.025 and 0.05 for each interval in order to control the global level of the test (at most 0.05 and 0.1 respectively). The results are shown in Table 5. All results show a similar behavior to that obtained for models (*v*), (*vi*) and (*vii*). The power is better in the homoscedastic case and for the larger sample sizes. In comparison to Neumeyer and Dette (2003b), our results are now better than theirs for this model.

		$C=1$				$C=2$			
		T_{KS}		T_{CM}		T_{KS}		T_{CM}	
	Hom./	α		α		α		α	
(n_1, n_2)	Het.	0.050	0.100	0.050	0.100	0.050	0.100	0.050	0.100
$(50, 50)$	Hom./	0.955	0.985	0.972	0.996	0.953	0.990	0.970	0.995
	Het.	0.620	0.747	0.695	0.804	0.614	0.760	0.686	0.802
$(100, 50)$	Hom./	0.980	0.991	0.989	0.998	0.978	0.993	0.986	0.997
	Het.	0.784	0.883	0.842	0.924	0.780	0.879	0.846	0.922
$(100, 100)$	Hom./	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Het.	0.923	0.964	0.955	0.979	0.921	0.959	0.955	0.984

Table 5. Rejection probabilities under model (viii), obtained by splitting the interval [0,1] into two parts. For the homoscedastic model, $\sigma_1^2 = 0.25$ *and* $\sigma_2^2 = 0.50$; *for the heteroscedastic model,* $\sigma_1^2(x) = \sigma_2^2(x) = e^x / \int_0^1 e^x\,dx$.

6 Application to real 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 (2003) verified that model (1) holds when X='log of the total monthly expenditure' is considered as covariate and Y='log of the expenditure on food' is the response variable (even a homoscedastic model is verified).

We will compare the regression curves for three groups of households : households consisting of 2 members (159 in total), 3 members (45 in total) and 4 members (74 in total). The bandwidths required for the estimation of the regression and variance functions are taken of the form $h = Cn^{-3/10}$ (as in the simulation study), where C is a constant chosen by a cross-validation procedure. The scatter plots and estimated regression curves are shown in Figures 1, 2 and 3. The values of the test statistics and the corresponding p-values, obtained from 1000 bootstrap replications, are displayed in Table 6.

No. of members	T_{KS}	p-value	T_{CM}	p-value
2 and 3	0.268	0.015	6.765	0.002
3 and 4	0.175	0.370	1.032	0.245
2 and 4	0.362	≤ 0.001	13.176	≤ 0.001

Table 6. Resuls for the econometric data.

Table 6 suggests that the regression functions for households of 2 and 3 members, and also for 2 and 4 members, are different. However it seems that the regression curves are the same when households of 3 and 4 members are considered (note that on the left side of Figure 2 the curves seem quite different, but there are only very few points in that area and so this area has a small contribution to the test statistics).

7 Testing the equality of the error distributions

As we mentioned in Section 2 the testing procedure we have introduced for the null hypothesis of equality of two regression functions is based on the assumption that the errors ε_1 and ε_2 are identically distributed. This assumption can, however, easily be verified by using a testing procedure similar to the above one, and is related to the assumption of the equality of the error variances in Analysis of Variance, see e.g. Bartlett's test, Cochran's test and Hartley's test (Seber, 1977).

Denote by F_{ε_j} the distribution of ε_j and by f_{ε_j} its corresponding density. Now we are interested in testing the null hypothesis

$$H_{0\varepsilon} : F_{\varepsilon_1} = F_{\varepsilon_2}. \tag{14}$$

Let F_ε (respectively f_ε) be the common distribution (respectively density) function of ε_1 and ε_2 under $H_{0\varepsilon}$. Let, for $j = 1, 2$,

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

where $\hat{m}_j$ and $\hat{\sigma}_j$ are given in (5) and (4). To test the hypothesis $H_{0\varepsilon}$, we consider the following test statistics:

$$S_{KS} = \sqrt{n} \sup_y |\hat{F}_{\varepsilon_1}(y) - \hat{F}_{\varepsilon_2}(y)|$$

$$S_{CM} = n \int [\hat{F}_{\varepsilon_1}(y) - \hat{F}_{\varepsilon_2}(y)]^2 d\hat{F}_\varepsilon(y),$$

where

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

Theorem 6 *Assume (A1)-(A4). Then, under the null hypothesis $H_{0\varepsilon}$,*

$$\begin{aligned}
&\hat{F}_{\varepsilon_1}(y) - \hat{F}_{\varepsilon_2}(y) \\
&= n_1^{-1} \sum_{i=1}^{n_1} \left\{ I\left(\frac{Y_{i1} - m_1(X_{i1})}{\sigma_1(X_{i1})} \le y \right) + \psi_1(X_{i1}, Y_{i1}, y) \right\} \\
&- n_2^{-1} \sum_{i=1}^{n_2} \left\{ I\left(\frac{Y_{i2} - m_2(X_{i2})}{\sigma_2(X_{i2})} \le y \right) + \psi_2(X_{i2}, Y_{i2}, y) \right\} + o_P(n^{-1/2})
\end{aligned}$$

uniformly in y, where

$$\psi_j(u, v, y) = y f_\varepsilon(y) \frac{(v - m_j(u))^2 - \sigma_j^2(u)}{2\sigma_j^2(u)} + f_\varepsilon(y) \frac{v - m_j(u)}{\sigma_j(u)}. \tag{15}$$

The proof of the following results can be obtained in a very similar way as that of Theorem 2 and Corollary 3.

Theorem 7 *Assume (A1)-(A4). Then, under the null hypothesis $H_{0\varepsilon}$, the process*

$$\hat{W}(y) = n^{1/2}(\hat{F}_{\varepsilon_1}(y) - \hat{F}_{\varepsilon_2}(y))$$

converges weakly to a zero-mean Gaussian process $W(y)$ with covariance function

$$\begin{aligned}
&Cov(W(y), W(y')) \\
&= \sum_{j=1}^{2} p_j^{-1} E\left(\left[I\left(\frac{Y_j - m_j(X_j)}{\sigma_j(X_j)} \le y \right) + \psi_j(X_j, Y_j, y) \right] \right. \\
&\qquad \left. \times \left[I\left(\frac{Y_j - m_j(X_j)}{\sigma_j(X_j)} \le y' \right) + \psi_j(X_j, Y_j, y') \right] \right) \\
&- \left(p_1^{-1} + p_2^{-1} \right) F_\varepsilon(y) F_\varepsilon(y').
\end{aligned}$$

Corollary 8 *Assume (A1)-(A4). Then, under the null hypothesis $H_{0\varepsilon}$,*

$$S_{KS} \xrightarrow{d} \sup_y |W(y)|$$
$$S_{CM} \xrightarrow{d} \int (W(y))^2 \, dF_\varepsilon(y).$$

8 Proofs

Before proving the main results, we first consider the following three lemmas.

Lemma 9 *Assume (A1)-(A3). Then, under the null hypothesis H_0 (for $j = 1$ or 2),*

$$\int \frac{\hat{m}(x) - m(x)}{\sigma_j(x)} f_{X_j}(x)dx = n^{-1} \sum_{k=1}^{2} \sum_{i=1}^{n_k} \frac{Y_{ik} - m(X_{ik})}{f_{mix}(X_{ik})} \frac{f_{X_j}(X_{ik})}{\sigma_j(X_{ik})} + o_P(n^{-1/2}).$$

Proof. Consider the kernel estimator of the density of the mixture $f_{mix}(x)$

$$\hat{f}_{mix}(x) = \frac{1}{nh_n} \sum_{k=1}^{2} \sum_{i=1}^{n_k} K\left(\frac{x - X_{ik}}{h_n}\right). \tag{16}$$

It is well known that (see for example Wand and Jones, 1995) $\hat{f}_{mix}(x) = f_{mix}(x) + O(h_n^2) + O_P((nh_n)^{-1/2})$, uniformly in x. By assumption (A2-ii) $O(h_n^2) = O((nh_n)^{-1/2})$, hence $\hat{f}_{mix}(x) - f_{mix}(x) = O_P((nh_n)^{-1/2})$ 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

$$\begin{aligned} \hat{m}(x) - m(x) &= \frac{1}{nh_n \hat{f}_{mix}(x)} \sum_{k=1}^{2} \sum_{i=1}^{n_k} K\left(\frac{x - X_{ik}}{h_n}\right) (Y_{ik} - m(x)) \\ &= \frac{1}{nh_n f_{mix}(x)} \sum_{k=1}^{2} \sum_{i=1}^{n_k} K\left(\frac{x - X_{ik}}{h_n}\right) (Y_{ik} - m(x)) + O_P((nh_n)^{-1})), \end{aligned}$$

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

$$\begin{aligned} &\int \frac{\hat{m}(x) - m(x)}{\sigma_j(x)} f_{X_j}(x)dx \\ &= \frac{1}{nh_n} \sum_{k=1}^{2} \sum_{i=1}^{n_k} \int \frac{K((x - X_{ik})h_n^{-1})(Y_{ik} - m(x))}{\sigma_j(x)} \frac{f_{X_j}(x)}{f_{mix}(x)} dx + o_P(n^{-1/2}). \end{aligned}$$

Denote $L(x) = (Y_{ik} - m(x)) f_{X_j}(x) (f_{mix}(x)\sigma_j(x))^{-1}$. Using the change of variable $u = (x - X_{ik})h_n^{-1}$, a Taylor expansion of second order of L around X_{ik} and assumption (A3), we have that

$$\begin{aligned}
&\int \frac{K((x-X_{ik})h_n^{-1})(Y_{ik}-m(x))}{\sigma_j(x)} \frac{f_{X_j}(x)}{f_{mix}(x)} dx \\
&= h_n \int K(u) L(X_{ik}+h_n u)du \\
&= h_n L(X_{ik}) \int K(u)du + h_n^2 L'(X_{ik}) \int uK(u)du + O(h_n^3) \\
&= h_n L(X_{ik}) + O(h_n^3).
\end{aligned}$$

Then, using assumption (A2),

$$\int \frac{\hat{m}(x)-m(x)}{\sigma_j(x)} f_{X_j}(x)dx = \frac{1}{n} \sum_{k=1}^{2} \sum_{i=1}^{n_k} \frac{Y_{ik}-m(X_{ik})}{\sigma_j(X_{ik})} \frac{f_{X_j}(X_{ik})}{f_{mix}(X_{ik})} + o_P(n^{-1/2}).$$

Lemma 10 *Assume (A1)-(A3). Then (for $j=1$ or 2),*

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

Proof. Similar to the proof of the previous lemma.

Lemma 11 *Assume (A1)-(A3). Then (for $j=1$ or 2),*

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

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 $O_P((n_j h_n)^{-1})$, and hence we can write

$$\hat{\sigma}_j(x)-\sigma_j(x) = \frac{\hat{\sigma}_j^2(x)-\sigma_j^2(x)}{2\sigma_j(x)} + O_P((n_j h_n)^{-1})$$

uniformly in x, and if we denote $\tilde{\sigma}_j^2(x) = \sum_{i=1}^{n_j} W_{ij}^{(j)}(x,h_n)(Y_{ij}-m_j(x))^2$,

$$\begin{aligned}
\hat{\sigma}_j^2(x) &= \sum_{i=1}^{n_j} W_{ij}^{(j)}(x,h_n)Y_{ij}^2 - \hat{m}_j^2(x) = \sum_{i=1}^{n_j} W_{ij}^{(j)}(x,h_n)(Y_{ij}^2-\hat{m}_j^2(x)) \\
&= \sum_{i=1}^{n_j} W_{ij}^{(j)}(x,h_n)(Y_{ij}-m_j(x))^2 - (m_j(x)-\hat{m}_j(x))^2 \\
&= \tilde{\sigma}_j^2(x) + O_P((n_j h_n)^{-1}),
\end{aligned}$$

because $\hat{m}_j(x) - m_j(x) = O_P((n_j h_n)^{-1/2})$. Since $\sigma_j^2(x)$ can be considered as the regression function of the variable $(Y_{ij} - m_j(x))^2$, we have that $\tilde{\sigma}_j^2(x) - \sigma_j^2(x) = O_P((n_j h_n)^{-1/2})$. Now we can write

$$
\begin{aligned}
\tilde{\sigma}_j^2(x) - \sigma_j^2(x) &= \frac{1}{n_j h_n \hat{f}_{X_j}(x)} \sum_{i=1}^{n_j} K((x - X_{ij})h_n^{-1}) \left\{ (Y_{ij} - m_j(x))^2 - \sigma_j^2(x) \right\} \\
&= \frac{1}{n_j h_n f_{X_j}(x)} \sum_{i=1}^{n_j} K((x - X_{ij})h_n^{-1}) \left\{ (Y_{ij} - m_j(x))^2 - \sigma_j^2(x) \right\} \\
&+ (1 - \hat{f}_{X_j}(x) f_{X_j}(x)^{-1})(\tilde{\sigma}_j^2(x) - \sigma_j^2(x)) \\
&= \frac{1}{n_j h_n f_{X_j}(x)} \sum_{i=1}^{n_j} K\left((x - X_{ij}) h_n^{-1}\right) \left\{ (Y_{ij} - m_j(x))^2 - \sigma_j^2(x) \right\} \\
&+ O_P((n_j h_n)^{-1}).
\end{aligned}
$$

Using (A2-ii) write

$$
\begin{aligned}
&\int \frac{\hat{\sigma}_j(x) - \sigma_j(x)}{\sigma_j(x)} f_{X_j}(x) dx \\
&= \int \frac{\tilde{\sigma}_j^2(x) - \sigma_j^2(x)}{2\sigma_j^2(x)} f_{X_j}(x) dx + O_P((n_j h_n)^{-1}) \\
&= \frac{1}{n_j h_n} \sum_{i=1}^{n_j} \int \frac{K((x - X_{ij})h_n^{-1}) \left\{ (Y_{ij} - m_j(x))^2 - \sigma_j^2(x) \right\}}{2\sigma_j^2(x)} dx + o_P(n_j^{-1/2}).
\end{aligned}
$$

Using a Taylor expansion of second order we obtain

$$
\begin{aligned}
&\int \frac{\hat{\sigma}_j(x) - \sigma_j(x)}{\sigma_j(x)} f_{X_j}(x) dx \\
&= \frac{1}{n_j} \sum_{i=1}^{n_j} \frac{(Y_{ij} - m_j(X_{ij}))^2 - \sigma_j^2(X_{ij})}{2\sigma_j^2(X_{ij})} + O(h_n^2) + o_P(n_j^{-1/2}) \\
&= \frac{1}{n_j} \sum_{i=1}^{n_j} \frac{(Y_{ij} - m_j(X_{ij}))^2 - \sigma_j^2(X_{ij})}{2\sigma_j^2(X_{ij})} + o_P(n_j^{-1/2}),
\end{aligned}
$$

and this concludes the proof of the lemma.

Proof of Theorem 1. First we will study the asymptotic behavior of $\hat{F}_{\varepsilon_j 0}(y) - F_\varepsilon(y)$, ($j = 1$ or 2). We will use some results and proofs from Akritas and Van Keilegom (2001). They assume that the functions m, m_j, and σ_j are L-functionals depending on a certain score function J. In our case these 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 extended to $J \equiv 1$. From the proof of Theorem 1 in Akritas and Van Keilegom (2001) we then have that

$$\hat{F}_{\varepsilon_j 0}(y) - F_\varepsilon(y) = n_j^{-1} \sum_{i=1}^{n_j} I\left(\frac{Y_{ij} - m(X_{ij})}{\sigma_j(X_{ij})} \leq y\right) - F_\varepsilon(y) \tag{17}$$
$$+ f_\varepsilon(y) \int \frac{y(\hat{\sigma}_j(x) - \sigma_j(x)) + \hat{m}(x) - m(x)}{\sigma_j(x)} f_{X_j}(x)dx + R_{n_j}(y),$$

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 (17) remains valid when the errors are estimated with $\hat{m}$, because their Lemma 1 holds in that case.

The integral on the right side above can be decomposed as follows:

$$\int \frac{y\left(\hat{\sigma}_j(x) - \sigma_j(x)\right) + \hat{m}(x) - m(x)}{\sigma_j(x)} f_{X_j}(x)\, dx$$
$$= y \int \frac{\hat{\sigma}_j(x) - \sigma_j(x)}{\sigma_j(x)} f_{X_j}(x)\, dx + \int \frac{\hat{m}(x) - m(x)}{\sigma_j(x)} f_{X_j}(x)\, dx.$$

Now, using Lemma 9, Lemma 11 and the fact that under the null hypothesis $m_1(x) = m_2(x) = m(x)$,

$$\hat{F}_{\varepsilon_j 0}(y) - F_\varepsilon(y) = n_j^{-1} \sum_{i=1}^{n_j} I\left(\frac{Y_{ij} - m(X_{ij})}{\sigma_j(X_{ij})} \leq y\right) - F_\varepsilon(y)$$
$$+ y f_\varepsilon(y)\, n_j^{-1} \sum_{i=1}^{n_j} \frac{(Y_{ij} - m(X_{ij}))^2 - \sigma_j^2(X_{ij})}{2\sigma_j^2(X_{ij})}$$
$$+ f_\varepsilon(y)\, n^{-1} \sum_{k=1}^{2} \sum_{i=1}^{n_k} \frac{Y_{ik} - m(X_{ik})}{f_{mix}(X_{ik})} \frac{f_{X_j}(X_{ik})}{\sigma_j(X_{ik})} + o_P(n^{-1/2})$$

uniformly in y. Note that $n_j = O(n)$ because of condition (A2-i).

We can write the difference between the two estimators for the distribution of the error as follows:

$$\hat{F}_{\varepsilon_1 0}(y) - \hat{F}_{\varepsilon_2 0}(y)$$
$$= n_1^{-1} \sum_{i=1}^{n_1} I\left(\frac{Y_{i1} - m(X_{i1})}{\sigma_1(X_{i1})} \leq y\right) + y f_\varepsilon(y)\, n_1^{-1} \sum_{i=1}^{n_1} \frac{(Y_{i1} - m(X_{i1}))^2 - \sigma_1^2(X_{i1})}{2\sigma_1^2(X_{i1})}$$
$$+ f_\varepsilon(y) \frac{n_1}{n} n_1^{-1} \sum_{i=1}^{n_1} \frac{Y_{i1} - m(X_{i1})}{f_{mix}(X_{i1})} \left(\frac{f_{X_1}(X_{i1})}{\sigma_1(X_{i1})} - \frac{f_{X_2}(X_{i1})}{\sigma_2(X_{i1})}\right)$$

$$
-n_2^{-1}\sum_{i=1}^{n_2} I\left(\frac{Y_{i2}-m\left(X_{i2}\right)}{\sigma_2\left(X_{i2}\right)}\leq y\right)-yf_\varepsilon\left(y\right)n_2^{-1}\sum_{i=1}^{n_2}\frac{\left(Y_{i2}-m\left(X_{i2}\right)\right)^2-\sigma_2^2\left(X_{i2}\right)}{2\sigma_2^2\left(X_{i2}\right)}
$$

$$
+f_\varepsilon\left(y\right)\frac{n_2}{n}n_2^{-1}\sum_{i=1}^{n_2}\frac{Y_{i2}-m\left(X_{i2}\right)}{f_{mix}\left(X_{i2}\right)}\left(\frac{f_{X_1}\left(X_{i2}\right)}{\sigma_1\left(X_{i2}\right)}-\frac{f_{X_2}\left(X_{i2}\right)}{\sigma_2\left(X_{i2}\right)}\right)+o_P(n^{-1/2}).
$$

Now note that (for $j=1$ or 2)

$$
\begin{aligned}
&f_\varepsilon\left(y\right)\frac{n_j}{n}n_j^{-1}\sum_{i=1}^{n_j}\frac{Y_{ij}-m\left(X_{ij}\right)}{f_{mix}\left(X_{ij}\right)}\left(\frac{f_{X_1}\left(X_{ij}\right)}{\sigma_1\left(X_{ij}\right)}-\frac{f_{X_2}\left(X_{ij}\right)}{\sigma_2\left(X_{ij}\right)}\right)\\
&=f_\varepsilon\left(y\right)p_jn_j^{-1}\sum_{i=1}^{n_j}\frac{Y_{ij}-m\left(X_{ij}\right)}{f_{mix}\left(X_{ij}\right)}\left(\frac{f_{X_1}\left(X_{ij}\right)}{\sigma_1\left(X_{ij}\right)}-\frac{f_{X_2}\left(X_{ij}\right)}{\sigma_2\left(X_{ij}\right)}\right)+o_P(n_j^{-1/2}),
\end{aligned}
$$

since $n_j^{-1}\sum_{i=1}^{n_j}\frac{Y_{ij}-m(X_{ij})}{f_{mix}(X_{ij})}\left(\frac{f_{X_1}(X_{ij})}{\sigma_1(X_{ij})}-\frac{f_{X_2}(X_{ij})}{\sigma_2(X_{ij})}\right)=O_P(n^{-1/2})$ and $n_j/n\to p_j$. Hence,

$$
\begin{aligned}
&\hat F_{\varepsilon_1 0}\left(y\right)-\hat F_{\varepsilon_2 0}\left(y\right)\\
&=n_1^{-1}\sum_{i=1}^{n_1}\left\{I\left(\frac{Y_{i1}-m\left(X_{i1}\right)}{\sigma_1\left(X_{i1}\right)}\leq y\right)+\varphi_1\left(X_{i1},Y_{i1},y\right)\right\}\\
&-n_2^{-1}\sum_{i=1}^{n_2}\left\{I\left(\frac{Y_{i2}-m\left(X_{i2}\right)}{\sigma_2\left(X_{i2}\right)}\leq y\right)+\varphi_2\left(X_{i2},Y_{i2},y\right)\right\}+o_P(n^{-1/2}),
\end{aligned}
$$

uniformly in y, where φ_j, $j=1,2$, are the functions defined in (10).

Proof of Theorem 2. In order to prove the weak convergence of the process we will use the techniques developed in van der Vaart and Wellner (1996) involving the theory of bracketing numbers.

First we will prove the weak convergence of the following process of i.i.d random variables (for $j=1$ or 2)

$$
\hat W_{j0}\left(y\right)=n_j^{-1/2}\sum_{i=1}^{n_j}\left\{I\left(\frac{Y_{ij}-m\left(X_{ij}\right)}{\sigma_j\left(X_{ij}\right)}\leq y\right)+\varphi_j\left(X_{ij},Y_{ij},y\right)-F_\varepsilon\left(y\right)\right\}.
$$

Consider the class of functions

$$
\mathcal{F}_j=\left\{I\left(\frac{Y_j-m\left(X_j\right)}{\sigma_j\left(X_j\right)}\leq y\right)+\varphi_j\left(X_j,Y_j,y\right)-F_\varepsilon\left(y\right);\quad -\infty<y<\infty\right\}.
$$

$\hat W_{j0}\left(y\right)$ is the $\mathcal{F}_j$-indexed empirical process (see p. 80 in van der Vaart and Wellner, 1996).

In general, for any classes of functions $\mathcal{G}_1$ and $\mathcal{G}_2$, define $\mathcal{G}_1+\mathcal{G}_2=\{g_1+g_2;\ g_1\in\mathcal{G}_1,g_2\in\mathcal{G}_2\}$ and $\mathcal{G}_1\cdot\mathcal{G}_2=\{g_1\cdot g_2;\ g_1\in\mathcal{G}_1,g_2\in\mathcal{G}_2\}$. In this way the class $\mathcal{F}_j$ can be written as

$$
\mathcal{F}_j=\mathcal{F}_{j1}+\mathcal{F}_{j2}+\mathcal{F}_{j3}\mathcal{F}_{j4}+\mathcal{F}_{j5}\mathcal{F}_{j6},
$$

where

$$\mathcal{F}_{j1} = \left\{ I\left(\frac{Y_j - m(X_j)}{\sigma_j(X_j)} \leq y\right); \quad -\infty < y < \infty \right\},$$
$$\mathcal{F}_{j2} = \{-F_\varepsilon(y); \quad -\infty < y < \infty\},$$
$$\mathcal{F}_{j3} = \{y f_\varepsilon(y); \quad -\infty < y < \infty\},$$
$$\mathcal{F}_{j4} = \left\{ \frac{(Y_j - m(X_j))^2 - \sigma_j^2(X_j)}{2\sigma_j^2(X_j)}; \quad -\infty < y < \infty \right\},$$
$$\mathcal{F}_{j5} = \left\{(-1)^{j-1} f_\varepsilon(y); \quad -\infty < y < \infty\right\},$$
$$\mathcal{F}_{j6} = \left\{ p_j \frac{Y_j - m(X_j)}{f_{mix}(X_j)} \left(\frac{f_{X_1}(X_j)}{\sigma_1(X_j)} - \frac{f_{X_2}(X_j)}{\sigma_2(X_j)} \right); \quad -\infty < y < \infty \right\}.$$

The classes $\mathcal{F}_{j4}$ and $\mathcal{F}_{j6}$ consist of only one function, hence $N_{[]}(\varepsilon, \mathcal{F}_{j4}, L_2(P)) = 1$ and $N_{[]}(\varepsilon, \mathcal{F}_{j6}, L_2(P)) = 1$, where $N_{[]}$ is the bracketing number, P is the measure of probability corresponding to the joint distribution of (X_j, Y_j) and $L_2(P)$ is the L_2-norm.

The classes $\mathcal{F}_{j2}$, $\mathcal{F}_{j3}$ and $\mathcal{F}_{j5}$ do not depend on (X_j, Y_j). They are classes of constant functions, considered as functions of (X_j, Y_j). All these functions are bounded because of condition (A4). Let $M \geq 1$ be such that $\sup_y |f_\varepsilon(y)| \leq M$ and $\sup_y |y f_\varepsilon(y)| \leq M$. If $\varepsilon < 2M$ then $N_{[]}(\varepsilon, \mathcal{F}_{jl}, L_2(P)) \leq 2M\varepsilon^{-1}$, $l = 2, 3, 5$. If $\varepsilon > 2M$, then $N_{[]}(\varepsilon, \mathcal{F}_{jl}, L_2(P)) = 1$, $l = 2, 3, 5$.

The functions in the class $\mathcal{F}_{j1}$ are decreasing in $(Y_j - m(X_j))/\sigma_j(X_j)$. Then, by Theorem 2.7.5 in van der Vaart and Wellner (1996), $N_{[]}(\varepsilon, \mathcal{F}_{j1}, L_2(P)) = O(\exp(K\varepsilon^{-1}))$ for some $K > 0$. Obviously, since the functions in $\mathcal{F}_{j1}$ are bounded with values in $[0, 1]$, if $\varepsilon > 1$ then $N_{[]}(\varepsilon, \mathcal{F}_{j1}, L_2(P)) = 1$.

By Theorem 2.10.6 in van der Vaart and Wellner (1996)

$$N_{[]}(\varepsilon, \mathcal{F}_j, L_2(P)) \leq \prod_{l=1}^{6} N_{[]}(\varepsilon, \mathcal{F}_{jl}, L_2(P)),$$

and taking logarithms $\log N_{[]}(\varepsilon, \mathcal{F}_j, L_2(P)) \leq \sum_{l=1}^{6} \log N_{[]}(\varepsilon, \mathcal{F}_{jl}, L_2(P))$. Now,

$$\begin{aligned}
&\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}_j, L_2(P))} d\varepsilon \\
&\leq \int_0^1 \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}_{j1}, L_2(P))} d\varepsilon + \int_0^{2M} \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}_{j2}, L_2(P))} d\varepsilon \\
&+ \int_0^{2M} \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}_{j3}, L_2(P))} d\varepsilon + \int_0^{2M} \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}_{j5}, L_2(P))} d\varepsilon,
\end{aligned}$$

and we can conclude that the integral $\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}_j, L_2(P))} d\varepsilon$ is finite and then, by Theorem 2.5.6. in van der Vaart and Wellner (1996) the class of functions $\mathcal{F}_j$ is Donsker.

The weak convergence of the process $\hat{W}_{j0}(y)$ now follows from pages 81 and 82 of the aforementioned book. The limit process, $W_{j0}(y)$, is a zero-mean Gaussian process with covariance function

$$\begin{aligned}
&\operatorname{Cov}\left(W_{j0}(y), W_{j0}(y')\right) \\
&= E\left(\left[I\left(\frac{Y_j - m(X_j)}{\sigma_j(X_j)} \le y\right) + \varphi_j(X_j, Y_j, y)\right]\left[I\left(\frac{Y_j - m(X_j)}{\sigma_j(X_j)} \le y'\right) + \varphi_j(X_j, Y_j, y')\right]\right) \\
&- F_\varepsilon(y) F_\varepsilon(y').
\end{aligned}$$

Now, our process of interest can be written as

$$\begin{aligned}
\hat{W}_0(y) &= n^{1/2}\left(\hat{F}_{\varepsilon 1 0}(y) - \hat{F}_{\varepsilon 2 0}(y)\right) \\
&= n^{1/2} n_1^{-1} \sum_{i=1}^{n_1} \left\{ I\left(\frac{Y_{i1} - m(X_{i1})}{\sigma_1(X_{i1})} \le y\right) + \varphi_1(X_{i1}, Y_{i1}, y) \right\} \\
&- n^{1/2} n_2^{-1} \sum_{i=1}^{n_2} \left\{ I\left(\frac{Y_{i2} - m(X_{i2})}{\sigma_2(X_{i2})} \le y\right) + \varphi_2(X_{i2}, Y_{i2}, y) \right\} + o_P(1) \\
&= \left(n_1^{-1} n\right)^{1/2} \hat{W}_{10}(y) - \left(n_2^{-1} n\right)^{1/2} \hat{W}_{20}(y) + o_P(1).
\end{aligned}$$

The processes $\hat{W}_{10}(y)$ and $\hat{W}_{20}(y)$ are independent and by assumption (A2-i) we have that $nn_1^{-1} \to p_1^{-1}$ and $nn_2^{-1} \to p_2^{-1}$. So, using the first part of this proof, we can conclude that the process $\hat{W}_0(y)$ converges weakly to a zero-mean Gaussian process $W_0(y)$ with covariance function

$$\begin{aligned}
&\operatorname{Cov}\left(W_0(y), W_0(y')\right) \\
&= p_1^{-1} \operatorname{Cov}\left(W_{10}(y), W_{10}(y')\right) + p_2^{-1} \operatorname{Cov}\left(W_{20}(y), W_{20}(y')\right).
\end{aligned}$$

Proof of Corollary 3. The convergence of the Kolmogorov-Smirnov type statistic follows directly from the continuous mapping theorem.

For the Cramér-von Mises type statistic it suffices to show that $d\hat{F}_\varepsilon(y)$ can be replaced by $dF_{\varepsilon 0}(y)$. Using the weak convergence of the processes $n^{1/2}(\hat{F}_{\varepsilon_1 0}(y) - \hat{F}_{\varepsilon_2 0}(y))$ and $n^{1/2}(\hat{F}_{\varepsilon_j 0}(y) - F_\varepsilon(y))$ $(j = 1, 2)$, and the Skorohod construction (see Serfling, 1980) we can write

$$\sup_y |n^{1/2}(\hat{F}_{\varepsilon_1 0}(y) - \hat{F}_{\varepsilon_2 0}(y)) - W_0(y)| \to_{a.s.} 0 \tag{18}$$

$$\sup_y |\hat{F}_{\varepsilon_j 0}(y) - F_\varepsilon(y)| \to_{a.s.} 0 \tag{19}$$

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

$$\begin{aligned}
&\left| n \int (\hat{F}_{\varepsilon_1 0}(y) - \hat{F}_{\varepsilon_2 0}(y))^2 d\hat{F}_{\varepsilon 0}(y) - \int W_0^2(y) dF_\varepsilon(y) \right| \\
&\leq \left| \int (n(\hat{F}_{\varepsilon_1 0}(y) - \hat{F}_{\varepsilon_2 0}(y))^2 - W_0^2(y)) d\hat{F}_{\varepsilon 0}(y) \right| \\
&+ \left| \int W_0^2(y) d(\hat{F}_{\varepsilon 0}(y) - F_\varepsilon(y)) \right|
\end{aligned}$$

The first term of the right hand side of the above inequality is $o(1)$ a.s. due to (18). For the second term note that

$$\hat{F}_{\varepsilon 0}(y) = \frac{n_1}{n} \hat{F}_{\varepsilon_1 0}(y) + \frac{n_2}{n} \hat{F}_{\varepsilon_2 0}(y)$$

and

$$\hat{F}_{\varepsilon 0}(y) - F_\varepsilon(y) = \frac{n_1}{n} (\hat{F}_{\varepsilon_1 0}(y) - F_\varepsilon(y)) + \frac{n_2}{n} (\hat{F}_{\varepsilon_2 0}(y) - F_\varepsilon(y)).$$

Hence it follows from (19) that $\sup_y |\hat{F}_{\varepsilon 0}(y) - F_\varepsilon(y)| \to_{a.s.} 0$. 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

$$\left| \int W_0^2(y) d(\hat{F}_{\varepsilon 0}(y) - F_\varepsilon(y)) \right| \to_{a.s.} 0.$$

This concludes the proof of Corollary 3.

Proof of Theorem 4. First note that $\hat{m}(x)$ can be written as

$$\hat{m}(x) = \frac{n_1}{n} \frac{\hat{f}_{X_1}(x)}{\hat{f}_{mix}(x)} \hat{m}_1(x) + \frac{n_2}{n} \frac{\hat{f}_{X_2}(x)}{\hat{f}_{mix}(x)} \hat{m}_2(x)$$

(where $\hat{f}_{X_j}(x) = (n_j h_n)^{-1} \sum_{i=1}^{n_j} K((x - X_{ij})/h_n)$ and $\hat{f}_{mix}(x)$ is defined in (16)), which estimates consistently the function

$$m_n(x) = p_1 \frac{f_{X_1}(x)}{f_{mix}(x)} m_1(x) + p_2 \frac{f_{X_2}(x)}{f_{mix}(x)} m_{2n}(x).$$

Under $H_{l.a.}$, m_n can be expressed in the following way

$$m_n(x) = m_1(x) + n^{-1/2} \tilde{r}_1(x) \qquad \text{and} \qquad m_n(x) = m_{2n}(x) + n^{-1/2} \tilde{r}_2(x), \tag{20}$$

where the functions $\tilde{r}_j$ are defined in the statement of the theorem $(j = 1, 2)$. Following the same steps as in the proof of Theorem 1, we can write

$$\begin{aligned}
&\hat{F}_{\varepsilon 1 0}(y) - \hat{F}_{\varepsilon 2 0}(y) \\
&= n_1^{-1} \sum_{i=1}^{n_1} I\left(\frac{Y_{i1} - m_n(X_{i1})}{\sigma_1(X_{i1})} \leq y\right) + y f_\varepsilon(y) n_1^{-1} \sum_{i=1}^{n_1} \frac{(Y_{i1} - m_1(X_{i1}))^2 - \sigma_1^2(X_{i1})}{2\sigma_1^2(X_{i1})} \\
&+ f_\varepsilon(y) p_1 n_1^{-1} \sum_{i=1}^{n_1} \frac{Y_{i1} - m_n(X_{i1})}{f_{mix}(X_{i1})} \left(\frac{f_{X_1}(X_{i1})}{\sigma_1(X_{i1})} - \frac{f_{X_2}(X_{i1})}{\sigma_2(X_{i1})}\right) \\
&- n_2^{-1} \sum_{i=1}^{n_2} I\left(\frac{Y_{i2} - m_n\left(X_{i2}\right)}{\sigma_2\left(X_{i2}\right)} \leq y\right) - y f_\varepsilon(y) n_2^{-1} \sum_{i=1}^{n_2} \frac{(Y_{i2} - m_{2n}(X_{i2}))^2 - \sigma_2^2(X_{i2})}{2\sigma_2^2(X_{i2})} \\
&+ f_\varepsilon(y) p_2 n_2^{-1} \sum_{i=1}^{n_2} \frac{Y_{i2} - m_n(X_{i2})}{f_{mix}(X_{i2})} \left(\frac{f_{X_1}(X_{i2})}{\sigma_1(X_{i2})} - \frac{f_{X_2}(X_{i2})}{\sigma_2(X_{i2})}\right) + o_P(n^{-1/2}),
\end{aligned}$$

and taking into account (20),

$$\begin{aligned}
&\hat{F}_{\varepsilon 1 0}\left(y\right) - \hat{F}_{\varepsilon 2 0}\left(y\right) \\
&= n_1^{-1} \sum_{i=1}^{n_1} \left\{ I\left(\frac{Y_{i1} - m_1\left(X_{i1}\right)}{\sigma_1\left(X_{i1}\right)} - n^{-1/2} \frac{\tilde{r}_1(X_{i1})}{\sigma_1\left(X_{i1}\right)} \leq y\right) + \tilde{\varphi}_1\left(X_{i1}, Y_{i1}, y\right) \right\} \\
&- f_\varepsilon(y) p_1 n^{-1/2} n_1^{-1} \sum_{i=1}^{n_1} \frac{\tilde{r}_1(X_{i1})}{f_{mix}(X_{i1})} \left(\frac{f_{X_1}(X_{i1})}{\sigma_1(X_{i1})} - \frac{f_{X_2}(X_{i1})}{\sigma_2(X_{i1})}\right) \\
&- n_2^{-1} \sum_{i=1}^{n_2} \left\{ I\left(\frac{Y_{i2} - m_{2n}\left(X_{i2}\right)}{\sigma_2\left(X_{i2}\right)} - n^{-1/2} \frac{\tilde{r}_2(X_{i2})}{\sigma_2\left(X_{i2}\right)} \leq y\right) + \tilde{\varphi}_2\left(X_{i2}, Y_{i2}, y\right) \right\} \\
&- f_\varepsilon(y) p_2 n^{-1/2} n_2^{-1} \sum_{i=1}^{n_2} \frac{\tilde{r}_2(X_{i2})}{f_{mix}(X_{i2})} \left(\frac{f_{X_1}(X_{i2})}{\sigma_1(X_{i2})} - \frac{f_{X_2}(X_{i2})}{\sigma_2(X_{i2})}\right) + o_P(n^{-1/2}), \qquad (21)
\end{aligned}$$

where the functions $\tilde{\varphi}_j$ are defined in the statement of the theorem. Let us now study each term of this representation. First,

$$\begin{aligned}
&f_\varepsilon(y) p_j n^{-1/2} n_j^{-1} \sum_{i=1}^{n_j} \frac{\tilde{r}_j(X_{ij})}{f_{mix}(X_{ij})} \left(\frac{f_{X_1}(X_{ij})}{\sigma_1(X_{ij})} - \frac{f_{X_2}(X_{ij})}{\sigma_2(X_{ij})}\right) \\
&= n^{-1/2} f_\varepsilon(y) b_j + o_P(n^{-1/2}), \qquad (22)
\end{aligned}$$

where b_j is the first term of d_j. An application of the proof of Lemma 1 of Akritas and Van Keilegom (2001) yields that

$$\begin{aligned}
&\sup_y |n_j^{-1} \sum_{i=1}^{n_j} \left\{ I(\varepsilon_{ij} - n^{-1/2} \tilde{r}_j(X_{ij}) \sigma_j^{-1}(X_{ij}) \leq y) - I(\varepsilon_{ij} \leq y) \right. \\
&\left. - P(\varepsilon_j - n^{-1/2} \tilde{r}_j(X_j) \sigma_j^{-1}(X_j) \leq y) + P(\varepsilon_j \leq y) \right\} | = o_P(n^{-1/2})
\end{aligned}$$

(where $\varepsilon_{i1} = (Y_{i1} - m_1(X_{i1}))/\sigma_1(X_{i1})$ and $\varepsilon_{i2} = (Y_{i2} - m_{2n}(X_{i2}))/\sigma_2(X_{i2})$). Now, considering the following probability as a function of y and applying a Taylor expansion, we have that

$$\begin{aligned}
&P(\varepsilon_j - n^{-1/2}\tilde{r}_j(X_j)\sigma_j^{-1}(X_j) \leq y) \\
&= \int P(\varepsilon_j - n^{-1/2}\tilde{r}_j(x)\sigma_j^{-1}(x) \leq y | X_j = x) f_{X_j}(x)dx \\
&= P(\varepsilon_j \leq y) - n^{-1/2}E[\tilde{r}_j(X_j)\sigma_j^{-1}(X_j)]f_\varepsilon(y) + o(n^{-1/2}).
\end{aligned}$$

Hence we can write

$$\begin{aligned}
&n_j^{-1}\sum_{i=1}^{n_j} I\left(\varepsilon_{ij} - n^{-1/2}\tilde{r}_j(X_{ij})\sigma_j^{-1}(X_{ij}) \leq y\right) \\
&= n_j^{-1}\sum_{i=1}^{n_j} I\left(\varepsilon_{ij} \leq y\right) - n^{-1/2}E[\tilde{r}_j(X_j)\sigma_j^{-1}(X_j)]f_\varepsilon(y) + o_P(n^{-1/2}). \qquad (23)
\end{aligned}$$

The result now follows from (21), (22) and (23).

Proof of Collorary 5. First note that, under $H_{l.a.}$, $(Y_{i1} - m_1(X_{i1}))/\sigma_1(X_{i1}) = \varepsilon_{i1}$ and $(Y_{i2} - m_{2n}(X_{i2}))/\sigma_2(X_{i2}) = \varepsilon_{i2}$. On the other hand, under H_0, $(Y_{i1} - m(X_{i1}))/\sigma_1(X_{i1}) = \varepsilon_{i1}$ and $(Y_{i2} - m(X_{i2}))/\sigma_2(X_{i2}) = \varepsilon_{i2}$. Therefore it can be easily shown that the distribution of

$$\begin{aligned}
&n_1^{-1}\sum_{i=1}^{n_1}\left\{I\left(\frac{Y_{i1} - m_1(X_{i1})}{\sigma_1(X_{i1})} \leq y\right) + \tilde{\varphi}_1(X_{i1}, Y_{i1}, y)\right\} \\
&- n_2^{-1}\sum_{i=1}^{n_2}\left\{I\left(\frac{Y_{i2} - m_{2n}(X_{i2})}{\sigma_2(X_{i2})} \leq y\right) + \tilde{\varphi}_2(X_{i2}, Y_{i2}, y)\right\}
\end{aligned}$$

under $H_{l.a}$, equals the distribution of

$$\begin{aligned}
&n_1^{-1}\sum_{i=1}^{n_1}\left\{I\left(\frac{Y_{i1} - m(X_{i1})}{\sigma_1(X_{i1})} \leq y\right) + \varphi_1(X_{i1}, Y_{i1}, y)\right\} \\
&- n_2^{-1}\sum_{i=1}^{n_2}\left\{I\left(\frac{Y_{i2} - m(X_{i2})}{\sigma_2(X_{i2})} \leq y\right) + \varphi_2(X_{i2}, Y_{i2}, y)\right\}
\end{aligned}$$

under H_0. Hence, the result follows from Theorem 2 and 4 and the proof of Corollary 3.

Proof of Theorem 6. From the proof of Theorem 1 in Akritas and Van Keilegom (2001) we have that

$$\begin{aligned}
&\hat{F}_{\varepsilon_j}(y) - F_{\varepsilon_j}(y) \\
&= n_j^{-1} \sum_{i=1}^{n_j} I\left(\frac{Y_{ij} - m_j(X_{ij})}{\sigma_j(X_{ij})} \le y \right) - F_{\varepsilon_j}(y) \\
&+ f_{\varepsilon_j}(y) \int \frac{y(\hat{\sigma}_j(x) - \sigma_j(x)) + \hat{m}_j(x) - m_j(x)}{\sigma_j(x)} f_{X_j}(x)dx + R_{n_j}(y),
\end{aligned}$$

where $\sup_y |R_{n_j}(y)| = o_P(n_j^{-1/2})$. The integral in the expression above can be decomposed as follows

$$\begin{aligned}
&\int \frac{y(\hat{\sigma}_j(x) - \sigma_j(x)) + \hat{m}_j(x) - m_j(x)}{\sigma_j(x)} f_{X_j}(x)dx \\
&= y \int \frac{\hat{\sigma}_j(x) - \sigma_j(x)}{\sigma_j(x)} f_{X_j}(x)dx + \int \frac{\hat{m}_j(x) - m_j(x)}{\sigma_j(x)} f_{X_j}(x)dx.
\end{aligned}$$

Now, using Lemma 10 and Lemma 11,

$$\begin{aligned}
\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j}(y) &= n_j^{-1} \sum_{i=1}^{n_j} I\left(\frac{Y_{ij} - m_j(X_{ij})}{\sigma_j(X_{ij})} \le y \right) - F_{\varepsilon}(y) \\
&+ y f_{\varepsilon_j}(y) n_j^{-1} \sum_{i=1}^{n_j} \frac{(Y_{ij} - m_j(X_{ij}))^2 - \sigma_j^2(X_{ij})}{2\sigma_j^2(X_{ij})} \\
&+ f_{\varepsilon_j}(y) n_j^{-1} \sum_{i=1}^{n_j} \frac{Y_{ij} - m_j(X_{ij})}{\sigma_j(X_{ij})} + o_P(n_j^{-1/2}) \\
&= n_j^{-1} \sum_{i=1}^{n_j} \left\{ I\left(\frac{Y_{ij} - m_j(X_{ij})}{\sigma_j(X_{ij})} \le y \right) + \psi_j(X_{ij}, Y_{ij}, y) \right\} - F_{\varepsilon}(y) + o_P(n_j^{-1/2}),
\end{aligned}$$

uniformly in y, where the function ψ_j is defined in (15). To conclude the proof it suffices to follow the same steps as in the proof of Theorem 1.

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.

Dette, H. and Neumeyer, N. (2001) Nonparametric analysis of covariance. *Ann. Statist.* **29**, 1361-1400.

Einmahl, J. H. J. and Van Keilegom, I. (2003). Goodness-of-fit tests in nonparametric regression (submitted).

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

Neumeyer, N. and Dette, H. (2003a) A note on one-sided non parametric analysis of covariance by ranking residuals (submitted).

Neumeyer, N. and Dette, H. (2003b) 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.

Seber, G. A. F. (1977) *Linear Regression Analysis.* 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.

Speckman, P. L., Chin, J. E., Hewett, J. E. and Bertelson, S. E. (2003) A one-sided test adjusting for covariates by ranking residuals following smoothing (submitted).

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.

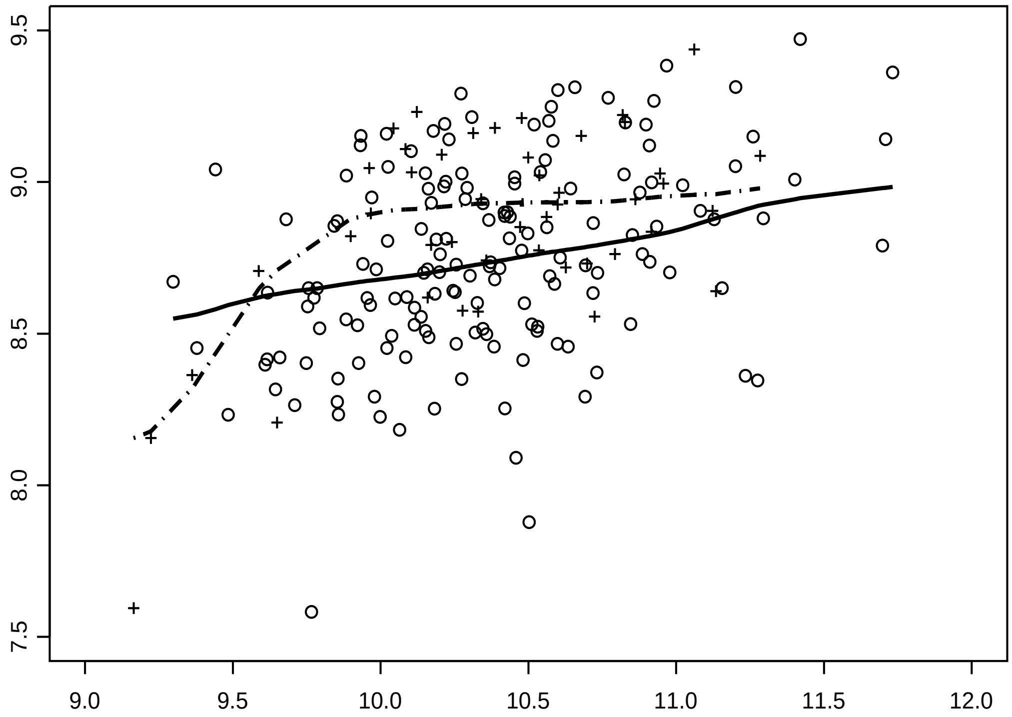

Figure 1: *Scatter plot of 'log(food espenditure)' versus 'log(total expenditure)' and estimated regression curves of households consisting of 2 members (circles and solid line) and 3 members (crosses and dashed line).*

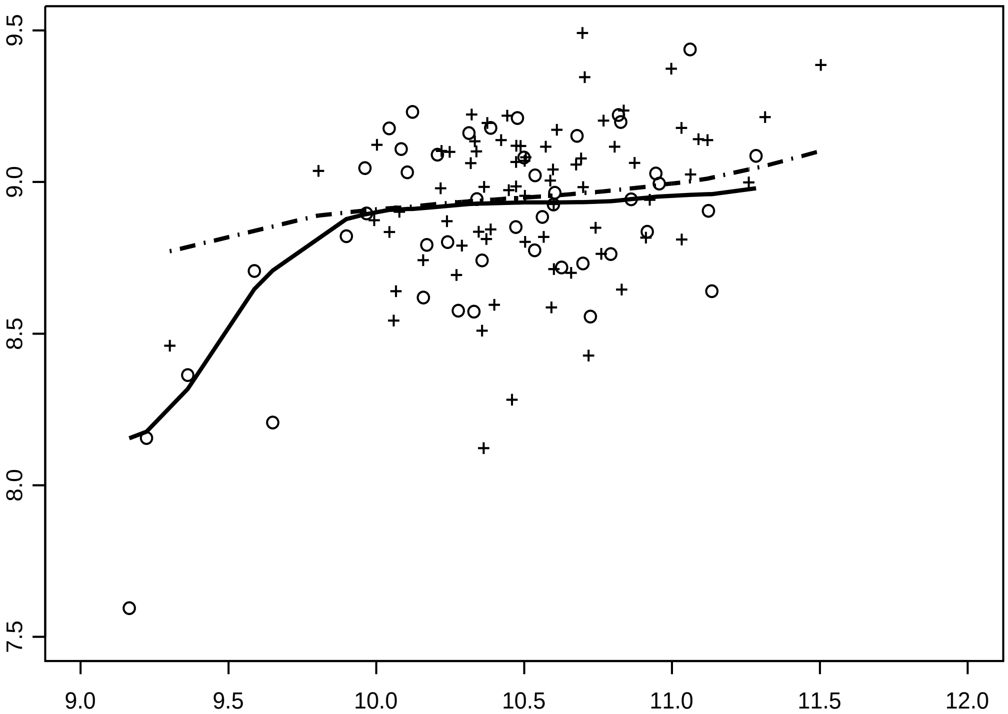

Figure 2: *Scatter plot of 'log(food espenditure)' versus 'log(total expenditure)' and estimated regression curves of households consisting of 3 members (circles and solid line) and 4 members (crosses and dashed line).*

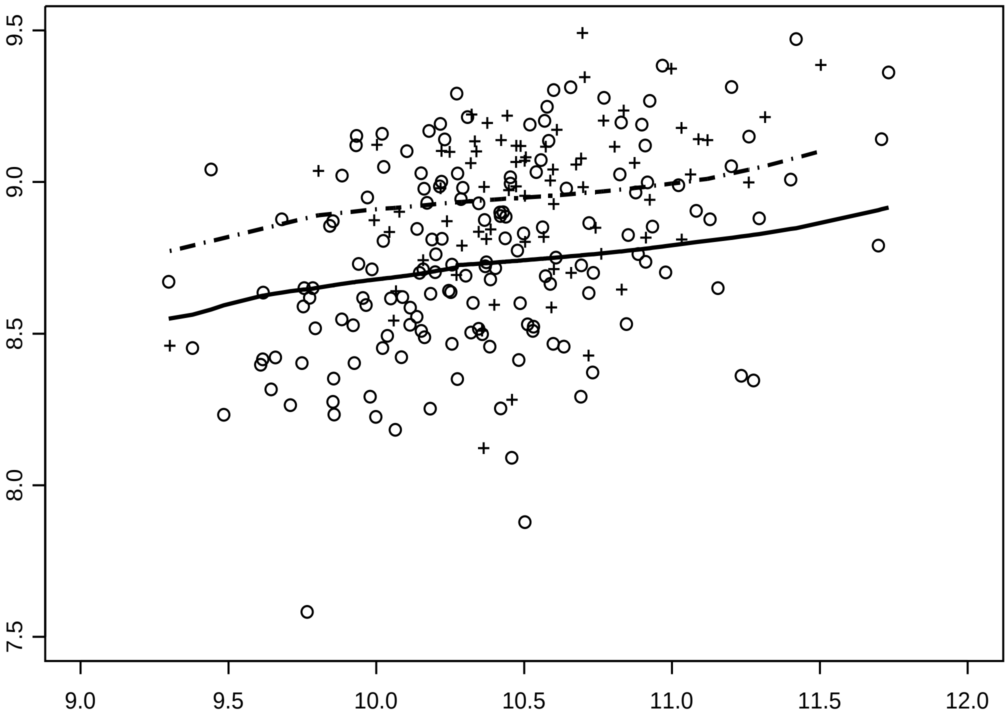

Figure 3: *Scatter plot of 'log(food espenditure)' versus 'log(total expenditure)' and estimated regression curves of households consisting of 2 members (circles and solid line) and 4 members (crosses and dashed line).*