I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0510

COMPARISON OF REGRESSION CURVES WITH CENSORED RESPONSES

PARDO-FERNANDEZ, J. C. and I. VAN KEILEGOM

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

Comparison of regression curves with censored responses

Juan Carlos Pardo-Fernández *

Departamento de Estatística e Investigación Operativa. Universidade de Vigo

Ingrid Van Keilegom †

Institut de Statistique. Université catholique de Louvain

April 2, 2005

Abstract

In this article we introduce a procedure to test the equality of regression functions when the response variables are censored. The test is based on a comparison of Kaplan-Meier estimators of the distribution of the censored residuals. Kolmogorov-Smirnov and Cramér-von Mises type statistics are considered. Some asymptotic results are proved: weak convergence of the process of interest, convergence of the test statistics and behavior of the process under local alternatives. We also describe a bootstrap procedure in order to approximate the critical values of the test. A simulation study and an application to a real data set conclude the paper.

Key words and phrases: Bootstrap; Censored data; Comparison of regression curves; Heteroscedastic regression; Nonparametric regression; Survival analysis.

*Address: E.U. Enxeñería Técnica Industrial. Rúa Torrecedeira, 86. Vigo (36208). Spain. e-mail: juancp@uvigo.es

†Address: Institut de Statistique. Voie du Roman Pays, 20. B-1348 Louvain-la-Neuve. Belgium. e-mail: vankeilegom@stat.ucl.ac.be

1 Introduction. Motivation and statistical model

Regression models are used for describing the relationship between a response and a covariate. In the field of survival analysis it can be useful to allow for censoring in the response variable. For instance, we can consider a model where the survival time (for patients having a certain disease) is the response variable and the age is the covariate. If we can distinguish two or more groups in the population (gender, treated patients and non-treated patients, etc.), we may be interested in testing for the equality of the corresponding regression curves. This kind of test allows to check whether the effect of the covariate over the variable of interest is the same in all the groups.

As it was pointed out in Fan and Gijbels (1994), when the response variable is censored the usual tools of regression (scatter plots, residuals plots, etc.) are not directly applicable to check, at least visually, the shape of the regression curves. This motivates the development of analytic tools in censored regression.

In this context, the statistical model can be described as follows. Let (X_j, Y_j), $j = 1, \ldots, k$, be independent random vectors, where Y_j represents a certain response variable associated to the covariate X_j. Suppose that the covariates have common support R_X. Assume that, for $j = 1, \ldots, k$, the response variable Y_j is subject to random right censoring. This means that there exists a censoring variable C_j, independent of Y_j given X_j, such that we can observe $Z_j = \min\{Y_j, C_j\}$ and the indicator of censoring $\Delta_j = I(Y_j \leq C_j)$. For $j = 1, \ldots, k$, assume that the following non-parametric regression models hold,

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

where the error variable ε_j is independent of X_j, m_j is an unknown conditional location function

$$m_j(x) = \int_0^1 F_j^{-1}(s|x)J(s)ds \tag{2}$$

and σ_j is an unknown conditional scale function representing possible heteroscedasticity

$$\sigma_j^2(x) = \int_0^1 F_j^{-1}(s|x)^2 J(s)ds - m_j^2(x), \tag{3}$$

where $F_j(\cdot|x)$ is the conditional distribution of Y_j given $X_j = x$, $F_j^{-1}(s|x) = \inf\{t; F_j(t|x) \geq s\}$ is the corresponding quantile function and $J(s)$ is a score function satisfying $\int_0^1 J(s)ds = 1$. We denote F_{ε_j} for the distribution of the error ε_j in population j. By construction $\int_0^1 F_{\varepsilon_j}^{-1}(s)J(s)ds = 0$ and $\int_0^1 F_{\varepsilon_j}^{-1}(s)^2 J(s)ds = 1$.

The choice of the function J leads to different location and scale functions. In particular if $J(s) = I(0 \leq s \leq 1)$ then $m_j(x) = E(Y_j|X_j = x)$ is the conditional mean function and $\sigma_j^2(x) = Var(Y_j|X_j = x)$ is the conditional variance function. However, it may happen that this choice of J is not appropriate because of the inconsistency of the estimator of the conditional distribution $F_j(\cdot|x)$ in the right tail due to the censoring . A useful choice is $J(s) = (q-p)^{-1}I(p \leq s \leq q)$, which leads to trimmed means and trimmed variances. The conditional median or other conditional quantiles can be seen as limits of trimmed means.

The samples are $(X_{ij}, Z_{ij}, \Delta_{ij})$, $i = 1, ..., n_j$, from the distribution of (X_j, Z_j, Δ_j), for $j = 1, \ldots, k$. Denote $n = \sum_{j=1}^{k} n_j$.

We are interested in testing the null hypothesis of equality between the location (regression) functions

$$H_0 : m_1 = m_2 = \cdots = m_k, \tag{4}$$

versus the alternative

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

When the distribution of the residuals and the variance functions are the same in all the groups (we do not assume so, but it is an interesting situation), if the null hypothesis holds for a particular definition of the location function, that is for a particular choice of J, then it holds for all possible location functions. However, in a general situation with different variances or different residual distributions, H_0 can be true for a particular choice of the functions m_j and false for another one.

In Pardo-Fernández, Van Keilegom and González-Manteiga (2004) a mechanism of comparison of regression curves for complete data is developed via the estimation of the distribution of the residuals of the models. The idea of the testing procedure proposed in that paper is to compare two estimators of the distribution of the residuals in each population. More precisely, 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 $m_j(\cdot)$ in population j, $\hat{m}(\cdot)$ is an estimator of the joint regression function $m(\cdot)$ under H_0, and $\hat{\sigma}_j^2(\cdot)$ is an estimator of the variance function $\sigma_j^2(\cdot)$. 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, the two estimators approximate the corresponding error distribution F_{ε_j}. However, if the null hypothesis is not true, they

estimate different functions, so a difference between them gives evidence to the inequality of the regression curves. In this paper we will extend that methodology to the situation where the response variable may be censored. Now, because of the censoring in the response variable, we will consider $(Z_{ij} - \hat{m}_j(X_{ij}))/\hat{\sigma}_j(X_{ij})$ and $(Z_{ij} - \hat{m}(X_{ij}))/\hat{\sigma}_j(X_{ij})$ to estimate the *censored* residuals, and we will substitute the empirical distribution by the Kaplan-Meier estimator of the distribution under random censoring (Kaplan and Meier, 1958).

In the case of complete data, the problem of testing for the equality of regression curves has been widely treated in the literature. A good and recent review on this topic can be found in Neumeyer and Dette (2003). To the best of our knowledge, this problem has not been treated in the case of censored responses.

The paper is organized as follows. In Section 2 we will introduce the testing procedure. In Section 3 we will state the main asymptotic results. A bootstrap procedure to approximate the critical points of the test is described in Section 4 and a simulation study is presented in Section 5. Finally, we include an application to real data in Section 6. The proofs of the main results are deferred to the Appendix.

2 Testing procedure

The testing procedure is based on the comparison of two non-parametric estimators of the distribution of the residuals F_{ε_j} in each population. This involves non-parametric estimation of the location and scale functions. All these estimators will be constructed using the estimator of the conditional distribution function $F_j(\cdot|x)$ when the response is censored introduced by Beran (1981):

$$\hat{F}_j(y|x) = 1 - \prod_{Z_{ij} \leq y, \Delta_{ij}=1} \left(1 - \frac{W_{ij}^{(j)}(x, h_n)}{\sum_{l=1}^{n_j} I(Z_{lj} \geq Z_{ij}) W_{lj}^{(j)}(x, h_n)} \right), \tag{5}$$

where

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

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

Now consider the following estimator of the location function for each sample, for $j = 1, \ldots, k$,

$$\hat{m}_j(x) = \int_0^1 \hat{F}_j^{-1}(s|x) J(s) ds, \tag{6}$$

and an estimator of the common location function under the null hypothesis (which we will denote by m) taking into account all the samples

$$\hat{m}(x) = \sum_{j=1}^{k} \frac{n_j}{n} \frac{\hat{f}_j(x)}{\hat{f}_{mix}(x)} \hat{m}_j(x), \tag{7}$$

where,

$$\hat{f}_j(x) = \frac{1}{n_j h_n} \sum_{i=1}^{n_j} K\left(\frac{x - X_{ij}}{h_n}\right)$$

is the kernel estimator of the density f_j of X_j, and

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

Note that $\hat{f}_j(x)$ can be computed in the usual way because the covariates do not suffer from censoring. The estimator of the scale function σ_j from each sample is

$$\hat{\sigma}_j^2(x) = \int_0^1 \hat{F}_j^{-1}(s|x)^2 J(s) ds - \hat{m}_j^2(x). \tag{8}$$

The score function J will be chosen so that $\hat{m}_j(x)$ and $\hat{\sigma}_j^2(x)$ are consistent, even in the case that the tails of the Beran estimator are not consistent.

Compute the estimators of the censored residuals in each sample

$$\hat{E}_{ij} = \frac{Z_{ij} - \hat{m}_j(X_{ij})}{\hat{\sigma}_j(X_{ij})} \tag{9}$$

for $i = 1, \ldots, n_j$, $j = 1, \ldots, k$, and estimate the distribution of the residuals from the censored sample $(\hat{E}_{ij}, \Delta_{ij})$ using the Kaplan-Meier estimator

$$\hat{F}_{\varepsilon_j}(y) = 1 - \prod_{\hat{E}_{ij} \leq y, \Delta_{ij}=1} \left(1 - \frac{1}{\sum_{l=1}^{n_j} I(\hat{E}_{lj} \geq \hat{E}_{ij0})}\right). \tag{10}$$

If the null hypothesis is true, we can estimate the residuals in each sample using the estimator of the common regression function $\hat{m}$, that is

$$\hat{E}_{ij0} = \frac{Z_{ij} - \hat{m}(X_{ij})}{\hat{\sigma}_j(X_{ij})} \tag{11}$$

for $i = 1, \ldots, n_j$, $j = 1, \ldots, k$, and estimate the corresponding distribution from the censored sample $(\hat{E}_{ij0}, \Delta_{ij})$

$$\hat{F}_{\varepsilon_j 0}(y) = 1 - \prod_{\hat{E}_{ij0} \leq y, \Delta_{ij}=1} \left(1 - \frac{1}{\sum_{l=1}^{n_j} I(\hat{E}_{lj0} \geq \hat{E}_{ij0})}\right). \tag{12}$$

Under the null hypothesis, both $\hat{F}_{\varepsilon_j}$ and $\hat{F}_{\varepsilon_j 0}$ are estimators of F_{ε_j}. The fact that there exists some difference between these two estimators of the distribution of the errors gives evidence for the inequality of the location functions. This idea is formalized theoretically in the following Theorem. Note that $\hat{m}(x)$ estimates consistently $m(x) = \sum_{j=1}^{k} p_j \frac{f_j(x)}{f_{mix}(x)} m_j(x)$, where $f_{mix}(x) = \sum_{j=1}^{k} p_j f_j(x)$ is the mixture of the densities of the covariates, provided that $n_j/n \to p_j > 0$. Let $F_{\varepsilon_j}(y) = P((Y_j - m_j(X_j))/\sigma_j(X_j) \leq y)$ and $F_{\varepsilon_j 0}(y) = P((Y_j - m(X_j))/\sigma_j(X_j) \leq y)$ be the theoretical versions (without estimated curves) of the distributions considered in (10) and (12).

Theorem 1 *Assume that m_j is continuous, $j = 1, \ldots, k$ and the moments of order ν of the distributions $F_{\varepsilon_j}(y)$ and $F_{\varepsilon_j 0}(y)$ exist for all $\nu \in \mathbb{N}$. Then $F_{\varepsilon_j}(y) = F_{\varepsilon_j 0}(y)$, $-\infty < y < \infty$, $j = 1, \ldots, k$ if and only if $m_1(x) = \ldots = m_k(x)$ for all $x \in R_X$.*

The equivalence given in the previous result is a theoretical justification of the proposed testing procedure. Its proof can be found in the Appendix.

Let $H_{e_j}(y) = P((Z_j - m_j(X_j))/\sigma_j(X_j) \leq y)$ and $\tau_{H_{e_j}} = \inf\{y; H_{e_j}(y) = 1\}$. All the asymptotic theory we will develop below is valid up to any point T smaller than $\min_j\{\tau_{H_{e_j}}\}$. The multidimensional process

$$\hat{\mathbf{W}}(y) = (\hat{W}_1(y), \ldots, \hat{W}_k(y))^t,$$

where

$$\hat{W}_j(y) = n_j^{1/2}(\hat{F}_{\varepsilon_j 0}(y) - \hat{F}_{\varepsilon_j}(y)),$$

$-\infty < y \leq T$, will be used to compare the two estimators of the distribution of the residuals in each population. We propose a Kolmogorov-Smirnov type statistic

$$T_{KS} = \sum_{j=1}^{k} \sup_{-\infty < y < T} |\hat{W}_j(y)| \tag{13}$$

and a Cramér-von Mises type statistic

$$T_{CM} = \sum_{j=1}^{k} \int_{-\infty}^{T} \hat{W}_j^2(y) d\hat{F}_{\varepsilon_j 0}(y). \tag{14}$$

The testing procedure consists of rejecting the null hypothesis (4) with significance level α when the value of the statistics T_{KS} or T_{CM} exceeds a certain critical value.

3 Asymptotic results

In this section we state the asymptotic results associated to the testing procedure. In order not to obstruct the description of the results, we defer the regularity assumptions and some auxiliary definitions to the beginning of the Appendix, which also contains the proofs. In the first part we work under the null hypothesis: we give an asymptotic representation for the difference between the two estimators of the distribution of the residuals in each population, we state the weak convergence of the corresponding multidimensional process and the convergence of the test statistics. In the second part we study asymptotic results under local alternatives converging to the null hypothesis at a rate $n^{-1/2}$.

3.1 Asymptotic results under the null hypothesis

The notation in the results below is the following: for $j = 1, \ldots, k$, $F_j(x) = P(X_j \leq x)$, $F_j(y|x) = P(Y_j \leq y|X_j = x)$, $G_j(y|x) = P(C_j \leq y|X_j = x)$, $H_j(y|x) = P(Z_j \leq y|X_j = x)$, $H_{j1}(y|x) = P(Z_j \leq y, \Delta_j = 1|X_j = x)$. We denote $E_j = (Z_j - m_j(X))/\sigma_j(X_j)$ and $H_{e_j}(y) = P(E_j \leq y)$, $H_{e_j1}(y) = P(E_j \leq y, \Delta_j = 1)$, $H_{e_j}(y|x) = P(E_j \leq y|X_j = x)$, $H_{e_j1}(y|x) = P(E_j \leq y, \Delta_j = 1|X_j = x)$. The derivatives of these functions will be denoted with the corresponding lower case letters. Finally, other functions needed in the theoretical results are (for $j = 1, \ldots, k$)

$$\xi_j(z, \delta, y|x) = (1 - F_j(y|x)) \left[- \int_{-\infty}^{y \wedge z} \frac{dH_{j1}(s|x)}{(1 - H_j(s|x))^2} + \frac{I(z \leq y, \delta = 1)}{1 - H_j(z|x)} \right],$$

$$\eta_j(z, \delta|x) = \sigma_j^{-1}(x) \int_{-\infty}^{+\infty} \xi_j(z, \delta, v|x) J(F_j(v|x)) dv,$$

$$\gamma_{j1}(y|x) = \int_{-\infty}^{y} \frac{h_{e_j}(s|x)}{(1 - H_{e_j}(s))^2} dH_{e_j1}(s) + \int_{-\infty}^{y} \frac{dh_{e_j1}(s|x)}{1 - H_{e_j}(s)}.$$

Theorem 2 *Assume (A1)-(A5) and $H_{e_j}(y|x)$, $H_{e_j1}(y|x)$ satisfy (A6). Then, under the null hypothesis H_0, for $j = 1, \ldots, k$,*

$$\hat{F}_{\varepsilon_j 0}(y) - \hat{F}_{\varepsilon_j}(y) = -(1 - F_{\varepsilon_j}(y)) \sum_{l=1}^{k} p_l \left\{ n_l^{-1} \sum_{i=1}^{n_l} \psi_{jl}(X_{il}, Z_{il}, \Delta_{il}, y) \right\} + o_P(n^{-1/2})$$

uniformly in $-\infty < y \leq T$, where

$$\psi_{jl}(x, z, \delta, y) = \left(\frac{f_j(x)}{f_{mix}(x)} \frac{\sigma_l(x)}{\sigma_j(x)} - \frac{I(l = j)}{p_j} \right) \eta_l(z, \delta|x) \gamma_{j1}(y|x).$$

Theorem 3 *Assume (A1)-(A5) and $H_{e_j}(y|x)$, $H_{e_j1}(y|x)$ satisfy (A6). Then under the null hypothesis H_0, the process $\hat{\boldsymbol{W}}(y) = (\hat{W}_1(y), \ldots, \hat{W}_k(y))^t$, $-\infty < y \leq T$, converges weakly to a k-dimensional centered Gaussian process $\boldsymbol{W}(y) = (W_1(y), \ldots, W_k(y))^t$ with covariance structure given by*

$$\begin{aligned} Cov(W_j(y), W_{j'}(y')) =& (p_j p_{j'})^{1/2}(1 - F_{\varepsilon_j}(y))(1 - F_{\varepsilon_{j'}}(y')) \\ & \times \sum_{l=1}^{k} p_l Cov(\psi_{jl}(X_l, Z_l, \Delta_l, y), \psi_{j'l}(X_l, Z_l, \Delta_l, y')). \end{aligned}$$

Corollary 4 *Assume (A1)-(A5) and $H_{e_j}(y|x)$, $H_{e_j1}(y|x)$ satisfy (A6). Then, under the null hypothesis H_0,*

$$\begin{aligned} T_{KS} &\xrightarrow{d} \sum_{j=1}^{k} \sup_{-\infty<y<T} |W_j(y)|, \\ T_{CM} &\xrightarrow{d} \sum_{j=1}^{k} \int_{-\infty}^{T} W_j^2(y) dF_{\varepsilon_j}(y). \end{aligned}$$

3.2 Asymptotic results under local alternatives

Let us study now the limiting behavior of the process $\hat{\mathbf{W}}(y)$ under local alternatives

$$H_{l.a.} : m_j = m^0 + n^{-1/2} r_j,$$

where the functions r_j satisfy

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

In addition, we will use the following condition on the censoring variables. This condition is needed in order to keep the proportion of censoring fixed for any value of n.

(AC) For $j = 1, \ldots, k$, there exist random variables C_j^0 such that $P(C_j \leq y|X_j = x) = P(C_j^0 + n^{-1/2} r_j(x) \leq y|X_j = x)$.

We define $Y_j^0 = m^0(X_j) + \sigma_j(X_j)\varepsilon_j$ and $Z_j^0 = \min\{Y_j^0, C_j^0\}$, and denote $F_j^0(y|x) = P(Y_j^0 \leq y|X_j = x)$, $H_j^0(y|x) = P(Z_j^0 \leq y|X_j = x)$, $H_{j1}^0(y|x) = P(Z_j^0 \leq y, \Delta_j = 1|X_j = x)$,

$$\xi_j^0(z, \delta, y|x) = (1 - F_j^0(y|x)) \left[- \int_{-\infty}^{y \wedge z} \frac{dH_{j1}^0(s|x)}{(1 - H_j^0(s|x))^2} + \frac{I(z \leq y, \delta = 1)}{1 - H_j^0(z|x)} \right],$$

$$\eta_j^0(z,\delta|x) = \sigma_j^{-1}(x)\int_{-\infty}^{+\infty} \xi_j^0(z,\delta,v|x)J(F_j^0(v|x))dv.$$

Theorem 5 *Assume (A1)-(A5), $H_{e_j}(y|x)$ and $H_{e_j1}(y|x)$ satisfy (A6) and (AR), (AC) hold. Then, under the alternative hypothesis $H_{l.a.}$, for $j = 1,\ldots,k$,*

$$\begin{aligned}\hat{F}_{\varepsilon_j 0}(y) - \hat{F}_{\varepsilon_j}(y) = \; & -(1-F_{\varepsilon_j}(y))\sum_{l=1}^{k} p_l \left\{ n_l^{-1}\sum_{i=1}^{n_l}\psi_{jl}^0(X_{il},Z_{il}^0,\Delta_{il},y)\right\} \\ & +n^{-1/2}p_j^{1/2}f_{\varepsilon_j}(y)d_j + o_P(n^{-1/2})\end{aligned}$$

uniformly in $-\infty < y \le T$, where

$$\psi_{jl}^0(x,z,\delta,y) = \left(\frac{f_j(x)}{f_{mix}(x)}\frac{\sigma_l(x)}{\sigma_j(x)} - \frac{I(l=j)}{p_j}\right)\eta_l^0(z,\delta|x)\gamma_{j1}(y|x),$$

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

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

Theorem 6 *Assume (A1)-(A5), $H_{e_j}(y|x)$ and $H_{e_j1}(y|x)$ satisfy (A6) and (AR), (AC) hold. Then, under the alternative hypothesis $H_{l.a.}$, the k-dimensional process $\hat{\boldsymbol{W}}(y), -\infty < y \le T$, converges weakly to $\boldsymbol{W}^0(y) + \boldsymbol{D}(y)$, where $\boldsymbol{D}(y) = (p_1^{1/2}f_{\varepsilon_1}(y)d_1,\ldots,p_k^{1/2}f_{\varepsilon_k}(y)d_k)^t$ and $\boldsymbol{W}^0(y) = (W_1^0(y),\ldots,W_k^0(y))^t$ is the k-dimensional centered Gaussian process with covariance structure given by*

$$\begin{aligned}Cov(W_j^0(y),W_{j'}^0(y')) =& (p_jp_{j'})^{1/2}(1-F_{\varepsilon_j}(y))(1-F_{\varepsilon_{j'}}(y')) \\ & \times \sum_{l=1}^{k} p_l Cov(\psi_{jl}^0(X_l,Z_l^0,\Delta_l,y),\psi_{j'l}^0(X_l,Z_l^0,\Delta_l,y')).\end{aligned}$$

Corollary 7 *Assume (A1)-(A5), $H_{e_j}(y|x)$ and $H_{e_j1}(y|x)$ satisfy (A6) and (AR), (AC) hold. Then, under the alternative hypothesis $H_{l.a.}$,*

$$\begin{aligned}T_{KS} \xrightarrow{d} & \sum_{j=1}^{k}\sup_{-\infty<y<T}|W_j^0(y)+p_j^{1/2}f_{\varepsilon_j}(y)d_j|, \\ T_{CM} \xrightarrow{d} & \sum_{j=1}^{k}\int_{-\infty}^{T}(W_j^0(y)+p_j^{1/2}f_{\varepsilon_j}(y)d_j)^2dF_{\varepsilon_j}(y).\end{aligned}$$

The shift term d_j that appears in the distribution of the test statistics under local alternatives is the same as the one obtained by Pardo-Fernández, Van Keilegom and González-Manteiga (2004) in a complete data situation. The same considerations as in that paper can be made here: d_j is not always different from zero. This means that, although the test is universally consistent in the sense of Theorem 1, the consistency of the test against alternatives converging to the null hypothesis at a rate $n^{-1/2}$ may fail in some particular situations.

4 Bootstrap

In practice, to apply this testing procedure we need the critical values of the test statistics. The asymptotic distributions of the test statistics under the null hypothesis given in Corollary 4 are complicated. Here we consider a bootstrap procedure based on the censored residuals to approximate those values.

First, for $j = 1, \ldots, k$ and $i = 1, \ldots, n_j$, estimate the censored residuals in a non parametric way, using each sample separately

$$\hat{E}_{ij} = \frac{Z_{ij} - \hat{m}_j(X_{ij})}{\hat{\sigma}_j(X_{ij})}. \tag{15}$$

From the censored sample of estimated residuals $\{(\hat{E}_{ij}, \Delta_{ij}), i = 1, \ldots, n_j\}$ compute the Kaplan-Meier estimator $\hat{F}_{\varepsilon_j}$ and 'standardize' these residuals in order to verify the initial assumption of having 'location function' 0 and 'scale function' 1. The standardized residuals are $\tilde{E}_{ij} = (\hat{E}_{ij} - \lambda_j)/\rho_j$, where $\lambda_j = \int \hat{F}_{\varepsilon_j}^{-1}(s)J(s)ds$ and $\rho_j = (\int \hat{F}_{\varepsilon_j}^{-1}(s)^2 J(s)ds - \lambda_j^2)^{1/2}$.

For resampling the censored residuals we use the 'naive bootstrap' described in Efron (1981) and studied in Akritas (1986). Different approaches of smooth bootstrap for censored data were considered in González-Manteiga, Cao and Marron (1996).

The bootstrap procedure we propose consists of the following steps. For fixed B and for $b = 1, \ldots, B$,

1. For each $j = 1, \ldots, k$ and $i = 1, \ldots, n_j$:

   - Let $Y^*_{ij,b} = \hat{m}(X_{ij}) + \hat{\sigma}_j(X_{ij})\varepsilon^*_{ij,b}$, where $\varepsilon^*_{ij,b} = V_{ij,b} + a_j Z_{ij,b}$, $V_{ij,b}$ is drawn from $\hat{F}_{\varepsilon_j}$ (standardized), and $Z_{ij,b}$ is a random variable with mean zero and variance one.
   - Select at random a $C^\star_{ij,b}$ from a smoothed version of $\hat{G}_j(\cdot|X_{ij})$, which is the Beran estimator of $G_j(\cdot|X_{ij})$ obtained by replacing Δ_{ij} by $1-\Delta_{ij}$ in the expression for $\hat{F}_j(\cdot|X_{ij})$.

- Let $Z^\star_{ij,b} = \min(Y^\star_{ij,b}, C^\star_{ij,b})$ and $\Delta^\star_{ij,b} = I(Y^\star_{ij,b} \leq C^\star_{ij,b})$.

2. The bootstrap samples are, for $j = 1, \ldots, k$, $\{(X_{ij}, Z^*_{ij,b}, \Delta^*_{ij,b}), i = 1, \ldots, n_j\}$.

3. Let $T^*_{KS,b}$ and $T^*_{CM,b}$ be the test statistics obtained from the bootstrap samples.

If we denote $T^*_{KS,(b)}$ and $T^*_{CM,(b)}$ for the order 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 present some simulations in order to study the practical behavior of the proposed bootstrap procedure. We restrict ourselves to two-sample situations ($k = 2$). More precisely, we consider the following models:

$$\begin{array}{lll} (i) & m_1(x) = x; & m_2(x) = x \\ (ii) & m_1(x) = \exp(x); & m_2(x) = \exp(x) \\ (iii) & m_1(x) = x; & m_2(x) = x + 1 \\ (iv) & m_1(x) = \exp(x); & m_2(x) = \exp(x) + x \end{array}$$

Clearly, models (i) and (ii) correspond to the null hypothesis and models (iii) and (iv) to the alternative hypothesis. In each case we consider a homoscedastic and a heteroscedastic situation. In the homoscedastic case the variances are

$$\sigma_1^2(x) = 0.25 \quad \text{and} \quad \sigma_2^2(x) = 0.50, \tag{16}$$

while in the heteroscedastic case the variance functions are

$$\sigma_1^2(x) = \frac{e^x}{\int_0^1 e^t dt} \quad \text{and} \quad \sigma_2^2(x) = \frac{e^{2x}}{\int_0^1 e^{2t} dt}. \tag{17}$$

Note that in the heteroscedastic case the variances are larger than in the homoscedastic case.

The censoring variables are

$$C_j = m_j(X_j) + \sigma_j(X_j)\rho_j,$$

where ρ_j has survival function $1 - F_\rho(y) = (1 - F_\varepsilon(y))^\beta$. This mechanism of censoring can be seen as a 'conditional proportional hazards model' (see Koziol and Green (1976) for

the 'proportional hazards model') and it allows us to have the same amount of censoring over all the support of the covariates. The proportion of censored data is $(1+\beta)^{-1}$. In the tables we consider $\beta = 1/3$ (25% of censoring) and $\beta = 1$ (50% of censoring).

In the theoretical results we have used only one bandwidth. As in Pardo-Fernández, Van Keilegom and González-Manteiga (2004), we have found that the bandwidth has not a big impact on the results of the tests, but it is recommendable to use the same bandwidth to estimate m and m_j. The variance functions could be estimated with different bandwidths. In these simulations we use a bandwidth of the type $h = Cn^{-3/10}$ to estimate m, m_j and σ_j $(j = 1, 2)$. The bandwidths chosen in this way verify the regularity conditions assumed in the theoretical results. In the tables the cases $C = 1$ and $C = 1.5$ are shown. This will allow us to check the test sensitivity to the change of the bandwidth.

In Tables 1 and 2 the distribution of the errors is *Exponential*, and in this case, the regression and variance functions are those corresponding to expressions (2) and (3) with the choice $J(s) = 0.75^{-1}I(0 \leq s \leq 0.75)$ for the score function (trimmed mean and variance). For the test statistics in (13) and (14) we take as the threshold T the value corresponding to the quantile 75% of the combined sample of the estimated residuals under the null hypothesis. Note that all these choices are reasonable for the models and censoring mechanisms we have considered. We work with $a_j = n_j^{-3/10}$ in the smooth bootstrap.

Table 1 shows the proportion of rejections in 1000 trials for sample sizes $(n_1, n_2) = (50, 50)$, $(100, 50)$ and $(100, 100)$, and when the expected amount of censored data is 25%. Table 2 shows the proportion of rejections in 1000 trials for sample sizes $(n_1, n_2) = (100, 100)$, $(200, 100)$ and $(200, 200)$ when the expected amount of censored data is 50%. In all cases we worked with $B = 200$ bootstrap replications and significance levels $\alpha = 0.05$ and $\alpha = 0.10$. Larger samples sizes for models with 50% of censored data are justified by the difficulty of those models.

The approximation of the level -models (i) and (ii)- is good in most cases. The results for models (iii) and (iv) show that the tests gain power as the sample sizes increase. In most cases the test based on T_{CM} gives better results than the test based on T_{KS}, and we also observe that the choice of the bandwidth has little impact on the rejection probabilities.

		$C=1$				$C=1.5$			
		T_{KS}		T_{CM}		T_{KS}		T_{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)	0.049	0.084	0.048	0.093	0.044	0.095	0.051	0.091
	(ii)	0.042	0.089	0.048	0.094	0.044	0.090	0.051	0.094
	(iii)	0.984	0.993	0.989	0.996	0.984	0.994	0.989	0.994
	(iv)	0.505	0.629	0.550	0.674	0.489	0.643	0.560	0.667
$(100, 50)$	(i)	0.059	0.112	0.057	0.104	0.061	0.098	0.067	0.098
	(ii)	0.067	0.107	0.058	0.102	0.055	0.096	0.058	0.100
	(iii)	0.999	0.999	0.998	0.999	1.000	1.000	1.000	1.000
	(iv)	0.557	0.724	0.541	0.690	0.580	0.720	0.568	0.699
$(100, 100)$	(i)	0.055	0.109	0.059	0.103	0.056	0.109	0.058	0.102
	(ii)	0.060	0.117	0.062	0.103	0.057	0.105	0.058	0.110
	(iii)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(iv)	0.863	0.924	0.888	0.935	0.862	0.923	0.879	0.932
					Heteroscedastic models				
$(50, 50)$	(i)	0.045	0.090	0.044	0.094	0.044	0.085	0.046	0.090
	(ii)	0.042	0.085	0.043	0.089	0.045	0.090	0.048	0.089
	(iii)	0.771	0.849	0.792	0.864	0.763	0.853	0.777	0.866
	(iv)	0.243	0.360	0.256	0.357	0.245	0.346	0.257	0.356
$(100, 50)$	(i)	0.048	0.087	0.044	0.086	0.051	0.102	0.048	0.092
	(ii)	0.050	0.086	0.046	0.088	0.048	0.089	0.049	0.091
	(iii)	0.909	0.955	0.880	0.939	0.913	0.956	0.892	0.947
	(iv)	0.236	0.343	0.191	0.283	0.246	0.363	0.221	0.313
$(100, 100)$	(i)	0.066	0.112	0.057	0.116	0.059	0.117	0.059	0.111
	(ii)	0.061	0.111	0.059	0.112	0.061	0.109	0.056	0.110
	(iii)	0.976	0.987	0.973	0.989	0.971	0.985	0.974	0.988
	(iv)	0.472	0.582	0.476	0.583	0.465	0.584	0.474	0.581

Table 1: *Rejection probabilities (models i-iv) of the tests based on T_{KS} and T_{CM} when the expected amount of censored data is 25%. The models are homoscedastic, with variances given in (16), and heteroscedastic, with variances given in (17).*

		$C=1$				$C=1.5$			
		T_{KS}		T_{CM}		T_{KS}		T_{CM}	
(n_1, n_2)	α :	0.050	0.100	0.050	0.100	0.050	0.100	0.050	0.100
					Homoscedastic models				
$(100, 100)$	(i)	0.052	0.093	0.055	0.093	0.042	0.081	0.049	0.083
	(ii)	0.051	0.083	0.050	0.099	0.037	0.075	0.046	0.087
	(iii)	0.995	0.997	0.995	0.997	0.995	0.997	0.994	0.996
	(iv)	0.751	0.855	0.813	0.903	0.723	0.838	0.792	0.888
$(200, 100)$	(i)	0.072	0.126	0.078	0.137	0.055	0.108	0.057	0.116
	(ii)	0.071	0.119	0.074	0.139	0.064	0.111	0.064	0.114
	(iii)	1.000	1.000	0.999	1.000	0.999	1.000	1.000	1.000
	(iv)	0.808	0.903	0.841	0.910	0.851	0.922	0.866	0.933
$(200, 200)$	(i)	0.053	0.101	0.059	0.105	0.052	0.090	0.057	0.098
	(ii)	0.055	0.094	0.059	0.103	0.055	0.092	0.058	0.098
	(iii)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(iv)	0.980	0.992	0.988	0.993	0.982	0.991	0.986	0.993
					Heteroscedastic models				
$(100, 100)$	(i)	0.032	0.067	0.042	0.089	0.038	0.077	0.043	0.089
	(ii)	0.035	0.070	0.041	0.081	0.039	0.071	0.045	0.086
	(iii)	0.932	0.964	0.942	0.972	0.905	0.953	0.926	0.966
	(iv)	0.387	0.520	0.441	0.578	0.370	0.495	0.430	0.565
$(200, 100)$	(i)	0.050	0.103	0.065	0.125	0.045	0.098	0.055	0.104
	(ii)	0.053	0.091	0.062	0.122	0.046	0.092	0.052	0.110
	(iii)	0.987	0.995	0.991	0.997	0.990	0.997	0.990	0.998
	(iv)	0.381	0.500	0.366	0.478	0.381	0.513	0.381	0.512
$(200, 200)$	(i)	0.044	0.081	0.049	0.090	0.041	0.082	0.048	0.082
	(ii)	0.046	0.086	0.049	0.089	0.046	0.086	0.046	0.089
	(iii)	0.998	0.999	0.998	1.000	0.998	1.000	0.998	1.000
	(iv)	0.688	0.787	0.722	0.812	0.669	0.786	0.722	0.800

Table 2: *Rejection probabilities (models i-iv) of the tests based on T_{KS} and T_{CM} when the expected amount of censored data is 50%. The models are homoscedastic, with variances given in (16), and heteroscedastic, with variances given in (17).*

6 Application to real data

We illustrate our testing procedure with an application to the Small Cell Lung Cancer Data. The data set is available in Ying, Jung and Wei (1995) and consists of lifetimes of patients suffering from small cell lung cancer. The patients are divided into two groups to follow two different treatments (Group A and Group B). The first group consists of 62 patients (15 censored) and the second group consists of 59 patients (8 censored). We consider the base 10 log of the survival time (in days) as response variable and the age as covariate. The support of the covariate was transformed into the interval [0,1]. We work with different values for the bandwidth needed in the estimation, ranging from 0.15 to 0.40.

We have performed the test of equality of the regression curves of the two curves using as score function $J(s) = 0.75^{-1}I(0 \leq s \leq 0.75)$ and $J(s) = 0.50^{-1}I(0.25 \leq s \leq 0.75)$. The second choice of the function J produces curves closer to the conditional median. The obtained results are very similar. The p-values are obtained from 1000 bootstrap replications. Figure 1 shows the estimated curves, using $h = 0.30$ as a bandwidth.

When testing for the equality of the curves, the null hypothesis is clearly rejected in all cases, with p-values smaller than 0.02 for the statistic T_{KM} and smaller than 0.005 for T_{CM}. However, it seems reasonable to suppose that the regression curves differ only by a shift (see Figure 1). A test to check that can be obtained by transforming the response variables in $Z'_{ij} = Z_{ij} - t_j$, for $j = 1, 2$ and $i = 1, \ldots, n_j$, where $t_j = n^{-1} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \hat{m}_j(X_{ij})$. In this case the p-values are larger than 0.55 for the statistic T_{KM} and larger than 0.67 for T_{CM}. All these results are summarized in Figures 2 and 3, which show graphs of the p-values versus the bandwidth.

Appendix: Notation, auxiliary results and proofs

Proof of Theorem 1. Assume that $F_{\varepsilon_j}(y) = F_{\varepsilon_j 0}(y)$, for $-\infty < y < \infty$. We write

$$P\left(\frac{Y_j - m(X_j)}{\sigma_j(X_j)} \leq y\right) = P\left(\frac{Y_j - m_j(X_j)}{\sigma_j(X_j)} + \frac{m_j(X_j) - m(X_j)}{\sigma_j(X_j)} \leq y\right),$$

for all y, or equivalently

$$\begin{aligned} &P\left(\exp\left\{\frac{Y_j - m(X_j)}{\sigma_j(X_j)}\right\} \leq y\right) \\ &= P\left(\exp\left\{\frac{Y_j - m_j(X_j)}{\sigma_j(X_j)}\right\}\exp\left\{\frac{m_j(X_j) - m(X_j)}{\sigma_j(X_j)}\right\} \leq y\right), \end{aligned}$$

for all y. Since $(Y_j - m_j(X_j))/\sigma_j(X_j)$ and X_j are independent, it follows that

$$\begin{aligned} &E\left[\left(\exp\left\{\frac{Y_j - m(X_j)}{\sigma_j(X_j)}\right\}\right)^{2\nu}\right] \\ &= E\left[\left(\exp\left\{\frac{Y_j - m_j(X_j)}{\sigma_j(X_j)}\right\}\right)^{2\nu}\right] E\left[\left(\exp\left\{\frac{m_j(X_j) - m(X_j)}{\sigma_j(X_j)}\right\}\right)^{2\nu}\right], \end{aligned}$$

for all ν. Then

$$E\left[\left(\exp\left\{\frac{m_j(X_j) - m(X_j)}{\sigma_j(X_j)}\right\}\right)^{2\nu}\right] = 1,$$

for all ν. Carleman condition (see e.g. Feller, 1966) ensures that

$$P\left(\exp\left\{\frac{m_j(X_j) - m(X_j)}{\sigma_j(X_j)}\right\} = 1\right) = 1$$

or

$$P\left(\frac{m_j(X_j) - m(X_j)}{\sigma_j(X_j)} = 0\right) = 1,$$

and this clearly implies that $m_j(x) = m(x)$ for all $j = 1, \ldots, k$ and for all $x \in R_X$, except for a set of points of probability zero. The continuity allows extending the equality of the regression curves to the whole support of the covariates. The converse implication is trivial.

Before starting the proofs of the results in Section 3, we state some additional notation ($j = 1, \ldots, k$)

$$\xi_{e_j}(z, \delta, y) = (1 - F_{\varepsilon_j}(y))\left[-\int_{-\infty}^{y \wedge z} \frac{dH_{e_j 1}(s)}{(1 - H_{e_j}(s))^2} + \frac{I(z \leq y, \delta = 1)}{1 - H_{e_j}(z)}\right],$$

$$\zeta_j(z, \delta | x) = \sigma_j^{-1}(x) \int_{-\infty}^{+\infty} \xi_j(z, \delta, v | x) J(F_j(v|x)) \frac{v - m_j(x)}{\sigma_j(x)} dv,$$

$$\gamma_{j2}(y|x) = \int_{-\infty}^{y} \frac{sh_{e_j}(s|x)}{(1-H_{e_j}(s))^2} dH_{e_j1}(s) + \int_{-\infty}^{y} \frac{d(sh_{e_j1}(s|x))}{1-H_{e_j}(s)}.$$

The functions $H_{e_j}(y)$ and $H_{e_j1}(y)$ are estimated in two different ways. First, we estimate them using the empirical distribution function of the censored residuals from each sample

$$\hat{H}_{e_j}(y) = n_j^{-1} \sum_{i=1}^{n_j} I(\hat{E}_{ij} \leq y) \text{ and } \hat{H}_{e_j1}(y) = n_j^{-1} \sum_{i=1}^{n_j} I(\hat{E}_{ij} \leq y, \Delta_{ij} = 1).$$

On the other hand, when working under the null hypothesis, we use the censored residuals based on the estimator of the common regression function

$$\hat{H}_{e_j0}(y) = n_j^{-1} \sum_{i=1}^{n_j} I(\hat{E}_{ij0} \leq y) \text{ and } \hat{H}_{e_j10}(y) = n_j^{-1} \sum_{i=1}^{n_j} I(\hat{E}_{ij0} \leq y, \Delta_{ij} = 1).$$

We list below the regularity assumptions we need for the proof of the main asymptotic results.

(A1) For $j = 1, \ldots, k$,
(i) X_j is absolutely continuous with compact support R_X.
(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) $n_j/n \to p_j > 0$, and $\sum_{j=1}^{k} p_j = 1$.
(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)
(i) K is a symmetric density function with compact support and K is twice continuously differentiable.
(ii) J is twice continuously differentiable in the interior of its support, $\int_0^1 J(s)ds = 1$ and $J(s) \geq 0$ for all $0 \leq s \leq 1$.
(iii) For $j = 1, \ldots, k$, let $\tilde{T}_{xj}$ be any value less than the upper bound of the support of $H_j(\cdot|x)$ such that $\inf_{x \in R_X}(1 - H_j(\tilde{T}_{xj}|x)) > 0$. Then there exist $0 \leq s_{0j} \leq s_{1j} \leq 1$ such that $s_{1j} \leq \inf_x F_j(\tilde{T}_{xj}|x)$, $s_{0j} \leq \inf\{s \in [0,1], J(s) \neq 0\}$, $s_{1j} \geq \sup\{s \in [0,1], J(s) \neq 0\}$ and $\inf_{x \in R_X} \inf_{s_{0j} \leq s \leq s_{1j}} f_j(F_j^{-1}(s|x)|x) > 0$.

(A4) For $j = 1, \ldots, k$, the functions η_j and ζ_j are twice continuously differentiable with respect to x and their first and second derivatives are bounded, uniformly in $x \in R_X$, $z < \tilde{T}_{xj}$ and δ.

In conditions (A5) and (A6) we use the generic notation $L(y|x)$ for a conditional distribution or subdistribution function, and denote $l(y|x) = L'(y|x)$ for their derivative with respect to y, $\dot{L}(y|x)$ their derivative with respect to x, and similar notation for higher order derivatives.

(A5) Let L be $H_j(y|x)$ or $H_{j1}(y|x)$, for $j = 1, \ldots, k$.
(i) $L(y|x)$ is continuous.
(ii) $l(y|x) = L'(y|x)$ exists, is continuous in (x, y), and $\sup_{x,y} |yL'(y|x)| < \infty$.
(iii) $L''(y|x)$ exists, is continuous in (x, y), and $\sup_{x,y} |y^2 L''(y|x)| < \infty$.
(iv) $\dot{L}(y|x)$ exists, is continuous in (x, y), and $\sup_{x,y} |y\dot{L}(y|x)| < \infty$.
(v) $\ddot{L}(y|x)$ exists, is continuous in (x, y), and $\sup_{x,y} |y^2\ddot{L}(y|x)| < \infty$.
(vi) $\dot{L}'(y|x)$ exists, is continuous in (x, y), and $\sup_{x,y} |y\dot{L}'(y|x)| < \infty$.

(A6)
(i) $l(y|x) = L'(y|x)$ exists, is continuous in (x, y), and $\sup_{x,y} |yL'(y|x)| < \infty$.
(ii) $L''(y|x)$ exists, is continuous in (x, y), and $\sup_{x,y} |y^2 L''(y|x)| < \infty$.
(iii) $\ddot{L}(y|x)$ exists, is continuous in (x, y), and $\sup_{x,y} |y\ddot{L}(y|x)| < \infty$.
(iv) $\ddot{L}'(y|x)$ exists, is continuous in (x, y), and $\sup_{x,y} |y\ddot{L}'(y|x)| < \infty$.

First we set four auxiliary lemmas, and then we prove the main results.

Lemma 8 *Assume (A1)-(A5) and $H_{e_j}(y|x)$ satisfy (A6). Then under the null hypothesis H_0, for $j = 1, \ldots, k$,*

$$\begin{aligned}
&\hat{H}_{e_j0}(y) - H_{e_j}(y) \\
&= \frac{1}{n_j}\sum_{i=1}^{n_j} I(E_{ij} \le y) - H_{e_j}(y) - \frac{1}{n_j}\sum_{i=1}^{n_j} y h_{e_j}(y|X_{ij})\zeta_j(Z_{ij}, \Delta_{ij}|X_{ij}) \\
&-\frac{1}{n}\sum_{l=1}^{k}\sum_{i=1}^{n_l} h_{e_j}(y|X_{il})\frac{f_j(X_{il})}{f_{mix}(X_{il})}\frac{\sigma_l(X_{il})}{\sigma_j(X_{il})}\eta_l(Z_{il}, \Delta_{il}|X_{il}) + o_P(n^{-1/2}),
\end{aligned}$$

uniformly in $-\infty < y \le T$.

Proof. From the proof of Proposition A.2 in Van Keilegom and Akritas (1999), we have that

$$\begin{aligned}
&\hat{H}_{e_j0}(y) - H_{e_j}(y) = \frac{1}{n_j}\sum_{i=1}^{n_j} I(E_{ij} \le y) - H_{e_j}(y) \qquad (18) \\
&+ \int h_{e_j}(y|x)\frac{\hat{m}(x) - m(x)}{\sigma_j(x)} f_j(x)dx + \int y h_{e_j}(y|x)\frac{\hat{\sigma}_j(x) - \sigma_j(x)}{\sigma_j(x)} f_j(x)dx \\
&+ o_P(n_j^{-1/2}),
\end{aligned}$$

uniformly in $-\infty < y \leq T$. The last term is $o_P(n_j^{-1/2})$ because of the uniform consistency of $\hat{m}$ and $\hat{\sigma}_j$. The consistency of $\hat{\sigma}_j$ is given in Proposition 4.5 in Van Keilegom and Akritas (1999). The consistency of $\hat{m}$ can be obtained using the consistency of $\hat{m}_l$ (also given in Proposition 4.5 in Van Keilegom and Akritas, 1999), the consistency of $\hat{f}_l$ and $\hat{f}_{mix}$ and taking into account the relation

$$\begin{aligned}\hat{m}(x) - m(x) &= \sum_{l=1}^{k} \frac{n_l}{n} \frac{\hat{f}_l(x)}{\hat{f}_{mix}(x)} (\hat{m}_l(x) - m(x)) \\ &= \sum_{l=1}^{k} \frac{n_l}{n} \frac{f_l(x)}{f_{mix}(x)} (\hat{m}_l(x) - m(x)) + o_P(n^{-1/2}),\end{aligned}$$

uniformly in x.

First using Proposition 4.8 in Van Keilegom and Akritas (1999)

$$\begin{aligned}&\hat{m}(x) - m(x) \\ = \; & -\frac{1}{nh_n} \frac{1}{f_{mix}(x)} \sum_{l=1}^{k} \sum_{i=1}^{n_l} \sigma_l(x) K\left(\frac{x - X_{il}}{h_n}\right) \eta_l(Z_{il}, \Delta_{il}|x) + o_P(n^{-1/2}),\end{aligned}$$

uniformly in x.

The two integrals in (18) will be analyzed separately. The first integral becomes

$$\begin{aligned}&\int h_{e_j}(y|x) \frac{\hat{m}(x) - m(x)}{\sigma_j(x)} f_j(x) dx \\ = \; & -\frac{1}{nh_n} \sum_{l=1}^{k} \sum_{i=1}^{n_l} \int h_{e_j}(y|x) \frac{f_j(x)}{f_{mix}(x)} \frac{\sigma_l(x)}{\sigma_j(x)} \eta_l(Z_{il}, \Delta_{il}|x) K\left(\frac{x - X_{il}}{h_n}\right) dx + o_P(n^{-1/2}).\end{aligned}$$

Using the change of variable $u = (x - X_{il})h_n^{-1}$, a Taylor expansion of second order around X_{il} and assumptions (A2-ii),(A3-i) and (A4) we obtain

$$\begin{aligned}&\int h_{e_j}(y|x) \frac{\hat{m}(x) - m(x)}{\sigma_j(x)} f_j(x) dx \\ &= -\frac{1}{n} \sum_{l=1}^{k} \sum_{i=1}^{n_l} h_{e_j}(y|X_{il}) \frac{f_j(X_{il})}{f_{mix}(X_{il})} \frac{\sigma_l(X_{il})}{\sigma_j(X_{il})} \eta_l(Z_{il}, \Delta_{il}|X_{il}) + o_P(n^{-1/2}).\end{aligned}$$

From Proposition 4.9 of Van Keilegom and Akritas (1999) and a Taylor expansion as we did above, we obtain a similar result for the second integral in (18)

$$\int y h_{e_j}(y|x) \frac{\hat{\sigma}_j(x) - \sigma_j(x)}{\sigma_j(x)} f_j(x) dx = -\frac{1}{n_j} \sum_{i=1}^{n_j} y h_{e_j}(y|X_{ij}) \zeta_j(Z_{ij}, \Delta_{ij}|X_{ij}) + o_P(n^{-1/2}).$$

The result stated in the Lemma now follows immediately.

Lemma 9 *Assume (A1)-(A5) and $H_{e_j1}(y|x)$ satisfy (A6). Then under the null hypothesis H_0, for $j = 1, \ldots, k$,*

$$\begin{aligned}
&\hat{H}_{e_j10}(y) - H_{e_j1}(y) \\
= \;& \frac{1}{n_j}\sum_{i=1}^{n_j} I(E_{ij} \le y, \Delta_{ij} = 1) - H_{e_j1}(y) - \frac{1}{n_j}\sum_{i=1}^{n_j} y h_{e_j1}(y|X_{ij})\zeta_j(Z_{ij}, \Delta_{ij}|X_{ij}) \\
& -\frac{1}{n}\sum_{l=1}^{k}\sum_{i=1}^{n_l} h_{e_j1}(y|X_{il})\frac{f_j(X_{il})}{f_{mix}(X_{il})}\frac{\sigma_l(X_{il})}{\sigma_j(X_{il})}\eta_l(Z_{il}, \Delta_{il}|X_{il}) + o_P(n_j^{-1/2}),
\end{aligned}$$

uniformly in $-\infty < y \le T$.

Proof. Similar to the proof of Lemma 8.

Lemma 10 *Assume (A1)-(A5) and $H_{e_j}(y|x)$ satisfy (A6). Then, for $j = 1, \ldots, k$,*

$$\begin{aligned}
&\hat{H}_{e_j}(y) - H_{e_j}(y) \\
&= \frac{1}{n_j}\sum_{i=1}^{n_j} I(E_{ij} \le y) - H_{e_j}(y) - \frac{1}{n_j}\sum_{i=1}^{n_j} y h_{e_j}(y|X_{ij})\zeta_j(Z_{ij}, \Delta_{ij}|X_{ij}) \\
&-\frac{1}{n_j}\sum_{i=1}^{n_j} h_{e_j}(y|X_{ij})\eta_j(Z_{ij}, \Delta_{ij}|X_{ij}) + o_P(n_j^{-1/2}),
\end{aligned}$$

uniformly in $-\infty < y \le T$.

Proof. This is Proposition A.2 in Van Keilegom and Akritas (1999).

Lemma 11 *Assume (A1)-(A5) and $H_{e_j1}(y|x)$ satisfy (A6). Then, for $j = 1, \ldots, k$,*

$$\begin{aligned}
&\hat{H}_{e_j1}(y) - H_{e_j1}(y) \\
&= \frac{1}{n_j}\sum_{i=1}^{n_j} I(E_{ij} \le y, \Delta_{ij} = 1) - H_{e_j1}(y) - \frac{1}{n_j}\sum_{i=1}^{n_j} y h_{e_j1}(y|X_{ij})\zeta_j(Z_{ij}, \Delta_{ij}|X_{ij}) \\
&-\frac{1}{n_j}\sum_{i=1}^{n_j} h_{e_j1}(y|X_{ij})\eta_j(Z_{ij}, \Delta_{ij}|X_{ij}) + o_P(n_j^{-1/2}),
\end{aligned}$$

uniformly in $-\infty < y \le T$.

Proof. Similar to the previous one.

Proof of Theorem 2. From the proof of Theorem 3.1 in Van Keilegom and Akritas (1999), we have that

$$\begin{aligned}
&\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j}(y) \\
&= (1 - F_{\varepsilon_j}(y)) \left[\int_{-\infty}^{y} \frac{\hat{H}_{e_j}(s) - H_{e_j}(s)}{(1 - H_{e_j}(s))^2} dH_{e_j 1}(s) + \int_{-\infty}^{y} \frac{d(\hat{H}_{e_j 1}(s) - H_{e_j 1}(s))}{1 - H_{e_j}(s)} \right] \\
&+ o_P(n^{-1/2}).
\end{aligned}$$

As in the proof of Lemma 8, the last terms of the above expressions are $o_P(n^{-1/2})$ because of the consistency of $\hat{m}$ and $\hat{\sigma}_j$. Applying Lemmas 8 and 9

$$\begin{aligned}
&\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j}(y) \\
&= \frac{1}{n_j} \sum_{i=1}^{n_j} \xi_{e_j}(E_{ij}, \Delta_{ij}, y) - \frac{1}{n_j} \sum_{i=1}^{n_j} (1 - F_{\varepsilon_j}(y)) \zeta_j(Z_{ij}, \Delta_{ij} | X_{ij}) \gamma_{j2}(y | X_{ij}) \\
&- \frac{1}{n} \sum_{l=1}^{k} \sum_{i=1}^{n_l} (1 - F_{\varepsilon_j}(y)) \frac{f_j(X_{il})}{f_{mix}(X_{il})} \frac{\sigma_l(x)}{\sigma_j(X_{il})} \eta_l(Z_{il}, \Delta_{il} | X_{il}) \gamma_{j1}(y | X_{il}) + o_P(n^{-1/2}).
\end{aligned}$$

Analogously,

$$\begin{aligned}
&\hat{F}_{\varepsilon_j}(y) - F_{\varepsilon_j}(y) \\
&= (1 - F_{\varepsilon_j}(y)) \left[\int_{-\infty}^{y} \frac{\hat{H}_{e_j}(s) - H_{e_j}(s)}{(1 - H_{e_j}(s))^2} dH_{e_j 1}(s) + \int_{-\infty}^{y} \frac{d(\hat{H}_{e_j 1}(s) - H_{e_j 1}(s))}{1 - H_{e_j}(s)} \right] \\
&+ o_P(n_j^{-1/2})
\end{aligned}$$

and applying Lemmas 10 and 11

$$\begin{aligned}
&\hat{F}_{\varepsilon_j}(y) - F_{\varepsilon_j}(y) \\
&= \frac{1}{n_j} \sum_{i=1}^{n_j} \xi_{e_j}(E_{ij}, \Delta_{ij}, y) - \frac{1}{n_j} \sum_{i=1}^{n_j} (1 - F_{\varepsilon_j}(y)) \zeta_j(Z_{ij}, \Delta_{ij} | X_{ij}) \gamma_{j2}(y | X_{ij}) \\
&- \frac{1}{n_j} \sum_{i=1}^{n_j} (1 - F_{\varepsilon_j}(y)) \eta_j(Z_{ij}, \Delta_{ij} | X_{ij}) \gamma_{j1}(y | X_{ij}) + o_P(n_j^{-1/2}).
\end{aligned}$$

By writing $\hat{F}_{\varepsilon_j 0}(y) - \hat{F}_{\varepsilon_j}(y) = (\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j}(y)) - (\hat{F}_{\varepsilon_j}(y) - F_{\varepsilon_j}(y))$, the representation given in the statement of the Theorem follows immediately.

Proof of Theorem 3. We will use the Cramér-Wold device (see e.g. Serfling, 1980) to prove the weak convergence of the multidimensional process $\hat{\mathbf{W}}(y)$ by showing the weak

convergence of any linear combination of its components. Let $\hat{V}(y) = \sum_{j=1}^{k} a_j \hat{W}_j(y)$ be one of these linear combinations.

Using the representation given in Theorem 2

$$\begin{aligned}
&\sum_{j=1}^{k} a_j \hat{W}_j(y) = \sum_{j=1}^{k} a_j n_j^{1/2} (\hat{F}_{\varepsilon_j 0}(y) - \hat{F}_{\varepsilon_j}(y)) \\
&= -\sum_{j=1}^{k} a_j n_j^{1/2} (1 - F_{\varepsilon_j}(y)) \times \\
&\times \left\{ \sum_{l=1}^{k} p_l n_l^{-1} \sum_{i=1}^{n_l} \frac{f_j(X_{il})}{f_{mix}(X_{il})} \frac{\sigma_l(X_{il})}{\sigma_j(X_{il)}} \eta_l(Z_{il}, \Delta_{il}|X_{il}) \gamma_{j1}(y|X_{il}) \right. \\
&\left. - \frac{1}{n_j} \sum_{i=1}^{n_j} \eta_j(Z_{ij}, \Delta_{ij}|X_{ij}) \gamma_{j1}(y|X_{ij}) \right\} + o_P(1) \\
&= \sum_{l=1}^{k} \frac{1}{n_l^{1/2}} \sum_{i=1}^{n_l} \varphi_l(X_{il}, Z_{il}, \Delta_{il}, y) + o_P(1),
\end{aligned}$$

where

$$\begin{aligned}
&\varphi_l(x, z, \delta, y) \\
&= -\eta_l(z, \delta|x) \left\{ \sum_{j=1}^{k} a_j (p_j p_l)^{1/2} (1 - F_{\varepsilon_j}(y)) \frac{f_j(x)}{f_{mix}(x)} \frac{\sigma_l(x)}{\sigma_j(x)} \gamma_{j1}(y|x) - a_l (1 - F_{\varepsilon_l}(y)) \gamma_{l1}(y|x) \right\}.
\end{aligned}$$

Denote, for $l = 1, \ldots, k$,

$$\hat{V}_l(y) = n_l^{-1/2} \sum_{i=1}^{n_l} \varphi_l(X_{il}, Z_{il}, \Delta_{il}, y).$$

With the notation of van der Vaart and Wellner (1996), if we consider the class of functions $\mathcal{F}_l = \{(x, z, \delta) \longrightarrow \varphi_l(x, z, \delta, y), -\infty < y < T\}$, then the process $\hat{V}_l(y)$ is the $\mathcal{F}_l$-indexed process. 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 $\mathcal{G}_1 \mathcal{G}_2 = \{g_1 g_2; g_1 \in \mathcal{G}_1, g_2 \in \mathcal{G}_2\}$. The class $\mathcal{F}_l$ can be written as $\mathcal{F}_l = \sum_{j=1}^{k+1} \mathcal{F}_{lj}^1 \mathcal{F}_{lj}^2$, where, for $j =, 1, \ldots, k$,

$$\mathcal{F}_{lj}^1 = \left\{ (x, z, \delta) \longrightarrow -\eta_l(z, \delta|x) a_j (p_j p_l)^{1/2} \frac{f_j(x)}{f_{mix}(x)} \frac{\sigma_l(x)}{\sigma_j(x)}, -\infty < y \leq T \right\},$$

$$\mathcal{F}_{lj}^2 = \left\{ (x, z, \delta) \longrightarrow (1 - F_{\varepsilon_j}(y)) \gamma_{j1}(y|x), -\infty < y \leq T \right\},$$

$$\mathcal{F}_{l,k+1}^1 = \left\{ (x, z, \delta) \longrightarrow \eta_l(z, \delta|x) a_l, -\infty < y \leq T \right\}.$$

and

$$\mathcal{F}_{l,k+1}^2 = \left\{ (x, z, \delta) \longrightarrow (1 - F_{\varepsilon_l}(y)) \gamma_{l1}(y|x), -\infty < y \leq T \right\}.$$

The functions in classes $\mathcal{F}_{lj}^2$ are bounded uniformly in y, as well as their first derivatives. Let M be a bound for the absolute value of all these functions. If $\varepsilon < 2M$ then their bracketing numbers are $N_{[]}(\varepsilon, \mathcal{F}_{lj}^2, L_2(P)) = O(\exp(K\varepsilon^{-1}))$, where $N_{[]}$ is the bracketing number, P is the measure of probability corresponding to the joint distribution of (X_l, Z_l, Δ_l) and $L_2(P)$ is the L_2-norm. If $\varepsilon \geq 2M$ then $N_{[]}(\varepsilon, \mathcal{F}_{lj}^2, L_2(P)) = 1$. Since the classes $\mathcal{F}_{lj}^1$ consist of only one function, hence the bracketing numbers of the product classes $\mathcal{F}_{lj}^1\mathcal{F}_{lj}^2$ verify the same conditions as those of the classes $\mathcal{F}_{lj}^2$.

By Theorem 2.10.6 in van der Vaart and Wellner (1996), which relates the bracketing number of a sum of classes of functions to the bracketing numbers of each class, we obtain

$$N_{[]}(\varepsilon, \mathcal{F}_l, L_2(P)) \leq \prod_{j=1}^{k+1} N_{[]}(\varepsilon, \mathcal{F}_{lj}^2, L_2(P)).$$

Now, we have

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}_l, L_2(P))}d\varepsilon \leq \sum_{j=1}^{k+1} \int_0^{2M} \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}_{lj}^2\mathcal{F}_{j3}, L_2(P))}d\varepsilon$$

and then the integral $\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}_l, L_2(P))}d\varepsilon$ is finite. This implies that the class of functions $\mathcal{F}_l$ is Donsker by Theorem 2.5.6 in van der Vaart and Wellner (1996). 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 $\mathrm{Cov}(V_l(y), V_l(y')) = \mathrm{Cov}(\varphi_l(X_l, Z_l, \Delta_l, y), \varphi_l(X_l, Z_l, \Delta_l, y'))$.

Write $\hat{V}(y) = \sum_{l=1}^k \hat{V}_l(y)$. The processes $\hat{V}_l(y)$ are independent. Using the first part of this proof, we conclude that the process $\hat{V}(y)$ converges weakly to a zero-mean Gaussian process, $V(y)$, with covariance function

$$\mathrm{Cov}(V(y), V(y')) = \sum_{l=1}^k \mathrm{Cov}(\varphi_l(X_l, Z_l, \Delta_l, y), \varphi_l(X_l, Z_l, \Delta_l, y')).$$

Finally, since we have verified the weak convergence of $\hat{V}(y)$, and using the Cramér-Wold device, we can conclude that the k-dimensional process $\hat{\mathbf{W}}(y)$ converges weakly to a centered k-dimensional Gaussian process with covariance structure given in the statement of the Theorem.

Proof of Corollary 4. The weak convergence of the k-dimensional process $\hat{\mathbf{W}}(y)$ and the continuous mapping theorem ensure the convergence of T_{KS}.

For the statistic T_{CM}, we will prove that

$$\int_{-\infty}^T \hat{W}_j^2(y)d\hat{F}_{\varepsilon_j 0}(y) \to_d \int_{-\infty}^T W_j^2(y)dF_{\varepsilon_j}(y). \tag{19}$$

The weak convergence of the processes $\hat{W}_j(y)$ and $n^{1/2}(\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j}(y))$, and the Skorohod construction (see Serfling, 1980) yield

$$\sup_{-\infty<y\leq T} |\hat{W}_j(y) - W_j(y)| \to_{a.s.} 0, \tag{20}$$

$$\sup_{-\infty<y\leq T} |\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j}(y)| \to_{a.s.} 0. \tag{21}$$

Now write

$$\left|\int_{-\infty}^{T} \hat{W}_j^2(y) d\hat{F}_{\varepsilon_j 0}(y) - \int_{-\infty}^{T} W_j^2(y) dF_{\varepsilon_j}(y)\right|$$
$$\leq \left|\int_{-\infty}^{T} (\hat{W}_j^2(y) - W_j^2(y)) d\hat{F}_{\varepsilon_j 0}(y)\right| + \left|\int_{-\infty}^{T} W_j^2(y) d(\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j}(y))\right|.$$

The first term of the right hand side of the above inequality is $o(1)$ a.s. due to (20). The trajectories of the limit process $W_j(y)$ are bounded and continuous almost surely. Then, by applying Helly-Bray Theorem (see p. 97 in Rao, 1965) to each of these trajectories and taking into account (21), we obtain

$$\left|\int_{-\infty}^{T} W_j^2(y) d(\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j}(y))\right| \to_{a.s.} 0.$$

This concludes the proof of the Corollary.

Before proving the asymptotic results concerning the behavior of the process under the alternative hypothesis, we introduce some notation and some general considerations. Under $H_{l.a.}$, the estimator of the common regression curve $\hat{m}$ estimates $m_n(x) = m^0(x) + n^{-1/2}R(x)$, where $R(x) = \sum_{l=1}^{k} p_l \frac{f_l(x)}{f_{mix}(x)} r_l(x)$, and $\hat{m}_j(x)$ estimates $m_{jn}(x) = m^0(x) + n^{-1/2} r_j(x)$. The censored residuals with respect to m_n are $E_{j0} = (Z_j - m_n(X_j))/\sigma_j(X_j)$, and with respect to m_{jn} the residuals are $E_j = (Z_j - m_{jn}(X_j))/\sigma_j(X_j)$ (for simplicity we keep the same notation of subsection 3.1 for the last ones, although they depend on n). We have the relation

$$E_{j0} = \frac{Z_j - m_n(X_j)}{\sigma_j(X_j)} = E_j + \frac{m_{jn}(X_j) - m_n(X_j)}{\sigma_j(X_j)} = E_j - n^{-1/2}\frac{R(X_j) - r_j(X_j)}{\sigma_j(X_j)}.$$

Note also that now $\hat{F}_{\varepsilon_j 0}(y)$ estimates $F_{\varepsilon_j 0}(y) = P((Y_j - m_n(X_j))/\sigma_j(X_j) \leq y)$. We denote $H_{e_j 0}(y) = P(E_{j0} \leq y)$, $H_{e_{j0}1}(y) = P(E_{j0} \leq y, \Delta_j = 1)$, $H_{e_j 0}(y|x) = P(E_{j0} \leq y | X_j = x)$, $H_{e_{j0}1}(y|x) = P(E_{j0} \leq y, \Delta_j = 1 | X_j = x)$.

Lemma 12 *Assume (A1)-(A5) and $H_{e_j}(y|x)$ satisfy (A6) and (AR) holds. Then, under the alternative hypothesis $H_{l.a.}$, for any $j = 1, \ldots, k$,*

$$\begin{aligned}\hat{H}_{e_j0}(y) - H_{e_j0}(y) &= \frac{1}{n_j}\sum_{i=1}^{n_j} I(E_{ij} \le y) - H_{e_j}(y) - \frac{1}{n_j}\sum_{i=1}^{n_j} y h_{e_j}(y|X_{ij})\zeta_j(Z_{ij}, \Delta_{ij}|X_{ij}) \\ &\quad -\frac{1}{n}\sum_{l=1}^{k}\sum_{i=1}^{n_l} h_{e_j}(y|X_{il})\frac{f_j(X_{il})}{f_{mix}(X_{il})}\frac{\sigma_l(X_{il})}{\sigma_j(X_{il})}\eta_l(Z_{il}, \Delta_{il}|X_{il}) + o_P(n_j^{-1/2}),\end{aligned}$$

uniformly in $-\infty < y \le T$.

Proof. We follow the proof of Proposition A.2 in Van Keilegom and Akritas (1999) and write

$$\begin{aligned}&\hat{H}_{e_j0}(y) - H_{e_j0}(y) = n_j^{-1}\sum_{i=1}^{n_j} I(E_{ij0} \le y) - H_{e_j0}(y) \qquad (22)\\ &+ \int H_{e_j0}\left(\frac{y\hat{\sigma}_j(x) + \hat{m}(x) - m_n(x)}{\sigma_j(x)}\middle| x\right) f_j(x)dx - \int H_{e_j0}(y|x)f_j(x)dx + o_P(n^{-1/2}),\end{aligned}$$

uniformly in $-\infty < y \le T$. Note that the remainder term in (22) is $o_P(n^{-1/2})$ provided that $\hat{m} - m_n$ satisfies Propositions 4.5, 4.6 and 4.7 in Van Keilegom and Akritas (1999). These propositions can be shown to hold true under standard lines of proof.

We will analyze in detail each term of the expression above. First

$$\begin{aligned}H_{e_j0}(y) &= \int H_{e_j0}(y|x)f_j(x)dx \qquad (23)\\ &= \int H_{e_j}(y|x)f_j(x)dx + \int h_{e_j}(y|x)\frac{m_n(x) - m_{jn}(x)}{\sigma_j(x)}f_j(x)dx + o(n^{-1/2})\\ &= H_{e_j}(y) + n^{-1/2}E\left[h_{e_j}(y|X_j)\frac{R(X_j) - r_j(X_j)}{\sigma_j(X_j)}\right] + o(n^{-1/2}),\end{aligned}$$

and

$$\begin{aligned}&\int H_{e_j0}\left(\frac{y\hat{\sigma}_j(x) + \hat{m}(x) - m_n(x)}{\sigma_j(x)}\middle| x\right) f_j(x)dx \qquad (24)\\ &= \int H_{e_j}(y|x)f_j(x)dx + \int h_{e_j}(y|x)\frac{y\hat{\sigma}_j(x) + \hat{m}(x) - m_{jn}(x) - y\sigma_j(x)}{\sigma_j(x)}f_j(x)dx + o(n^{-1/2})\\ &= H_{e_j}(y) + \int h_{e_j}(y|x)\frac{y(\hat{\sigma}_j(x) - \sigma_j(x)) + \hat{m}(x) - m_n(x)}{\sigma_j(x)}f_j(x)dx\\ &+ n^{-1/2}E\left[h_{e_j}(y|X_j)\frac{R(X_j) - r_j(X_j)}{\sigma_j(X_j)}\right] + o(n^{-1/2}).\end{aligned}$$

An application of the proof of Lemma A.1 in Van Keilegom and Akritas (1999) yields

$$\sup_y |n_j^{-1}\sum_{i=1}^{n_j}\big\{I(E_{ij}-n^{-1/2}(R(X_{ij})-r_j(X_{ij}))\sigma_j^{-1}(X_{ij})\le y)$$
$$-P(E_j-n^{-1/2}(R(X_j)-r_j(X_j))\sigma_j^{-1}(X_j)\le y)-I(E_{ij}\le y)+P(E_j\le y)\big\}|$$
$$=o_P(n_j^{-1/2})$$

Considering the following probability as a function of y and developing a Taylor expansion we obtain

$$P(E_j-n^{-1/2}(R(X_j)-r_j(X_j))\sigma_j^{-1}(X_j)\le y)$$
$$=\int P(E_j-n^{-1/2}(R(X_j)-r_j(X_j))\sigma_j^{-1}(X_j)\le y|X_j=x)f_j(x)dx$$
$$=\int P(E_j\le y|X_j=x)f_j(x)dx+n^{-1/2}E[h_{e_j}(y|X_j)(R(X_j)-r_j(X_j))\sigma_j^{-1}(X_j)]+o(n^{-1/2})$$
$$=P(E_j\le y)+n^{-1/2}E[h_{e_j}(y|X_j)(R(X_j)-r_j(X_j))\sigma_j^{-1}(X_j)]+o(n^{-1/2}),$$

and hence

$$n_j^{-1}\sum_{i=1}^{n_j}I(E_{ij0}\le y)=n_j^{-1}\sum_{i=1}^{n_j}I(E_{ij}-n^{-1/2}(R(X_{ij})-r_j(X_{ij}))\sigma_j^{-1}(X_{ij})\le y) \tag{25}$$
$$=n_j^{-1}\sum_{i=1}^{n_j}I(E_{ij}\le y)+n^{-1/2}E[h_{e_j}(y|X_j)(R(X_j)-r_j(X_j))\sigma_j^{-1}(X_j)]+o_P(n^{-1/2}).$$

Substituting (23), (24) and (25) in (22), we obtain

$$\hat{H}_{e_j0}(y)-H_{e_j0}(y)=n_j^{-1}\sum_{i=1}^{n_j}I(E_{ij}\le y)-H_{e_j}(y)$$
$$+\int h_{e_j}(y|x)\frac{y(\hat{\sigma}_j(x)-\sigma_j(x))+\hat{m}(x)-m_n(x)}{\sigma_j(x)}f_j(x)dx+o_P(n^{-1/2}).$$

Since $\hat{m}(x)-m_n(x)=\sum_{l=1}^k\frac{n_l}{n}\frac{f_l(x)}{f_{mix}(x)}(\hat{m}_l(x)-m_{ln}(x))+o_P(n^{-1/2})$, the integral on the right-hand side of the expression above can be handled in a similar way as in the proof of Lemma 8.

Lemma 13 *Assume (A1)-(A5) and $H_{e_j1}(y|x)$ satisfy (A6) and (AR) holds. Then, under*

the alternative hypothesis $H_{l.a.}$, for any $j = 1, \ldots, k$,

$$\begin{aligned}
&\hat{H}_{e_j 10}(y) - H_{e_j 10}(y) \\
= \; & \frac{1}{n_j}\sum_{i=1}^{n_j} I(E_{ij} \le y, \Delta_{ij} = 1) - H_{e_j 1}(y) - \frac{1}{n_j}\sum_{i=1}^{n_j} y h_{e_j 1}(y|X_{ij})\zeta_j(Z_{ij}, \Delta_{ij}|X_{ij}) \\
& -\frac{1}{n}\sum_{l=1}^{k}\sum_{i=1}^{n_l} h_{e_j 1}(y|X_{il})\frac{f_j(X_{il})}{f_{mix}(X_{il})}\frac{\sigma_l(X_{il})}{\sigma_j(X_{il})}\eta_l(Z_{il}, \Delta_{il}|X_{il}) + o_P(n_j^{-1/2}),
\end{aligned}$$

uniformly in $-\infty < y \le T$.

Proof. Similar to the proof of Lemma 12.

Proof of Theorem 5. First we write

$$\begin{aligned}
\int_{-\infty}^{y} \frac{d(\hat{H}_{e_j 10}(s) - H_{e_j 10}(s))}{1 - H_{e_j 0}(s)} &= \int_{-\infty}^{y} \frac{d(\hat{H}_{e_j 10}(s) - H_{e_j 10}(s))}{1 - H_{e_j}(s)} \\
&+ \int_{-\infty}^{y} \left(\frac{1}{1 - H_{e_j 0}(s)} - \frac{1}{1 - H_{e_j}(s)}\right) d(\hat{H}_{e_j 10}(s) - H_{e_j 10}(s)).
\end{aligned} \tag{26}$$

The proof of Corollary A.5 in Van Keilegom and Akritas (1999) can be adapted here to show that

$$\sup_{-\infty < y \le T} \left| \int_{-\infty}^{y} \left(\frac{1}{1 - H_{e_j 0}(s)} - \frac{1}{1 - H_{e_j}(s)}\right) d(\hat{H}_{e_j 10}(s) - H_{e_j 10}(s)) \right| = o_P(n^{-1/2}). \tag{27}$$

Indeed, equation (23) in the proof of Lemma 12 says that $H_{e_j 0}(y) - H_{e_j}(y) = O(n^{-1/2})$. Note that this order is not stochastic and better than the equivalent one needed in the above-mentioned proof. It suffices to follow the same steps to obtain (27). Hence the last term of the expression (26) is $o_P(n^{-1/2})$, and we obtain

$$\int_{-\infty}^{y} \frac{d(\hat{H}_{e_j 10}(s) - H_{e_j 10}(s))}{1 - H_{e_j 0}(s)} = \int_{-\infty}^{y} \frac{d(\hat{H}_{e_j 10}(s) - H_{e_j 10}(s))}{1 - H_{e_j}(s)} + o_P(n^{-1/2}). \tag{28}$$

Similarly to equation (23), it holds that

$$H_{e_j 10}(y) = \int H_{e_j 1}\left(\frac{y\sigma_j(x) + m_n(x) - m_{jn}(x)}{\sigma_j(x)}\middle| x\right) f_j(x)dx,$$

and taking derivatives and a Taylor expansion of $h_{e_j 1}$ around y

$$h_{e_j 10}(y) = \int h_{e_j 1}\left(\frac{y\sigma_j(x) + m_n(x) - m_{jn}(x)}{\sigma_j(x)}\middle| x\right) f_j(x)dx = h_{e_j 1}(y) + O(n^{-1/2}).$$

It follows that

$$\frac{h_{e_j10}(s)}{(1-H_{e_j0}(s))^2} = \frac{h_{e_j1}(s)}{(1-H_{e_j}(s))^2}$$
$$+ h_{e_j10}(s)\left(\frac{1}{(1-H_{e_j0}(s))^2} - \frac{1}{(1-H_{e_j}(s))^2}\right) + \frac{1}{(1-H_{e_j}(s))^2}(h_{e_j10}(s) - h_{e_j1}(s))$$
$$= \frac{h_{e_j1}(s)}{(1-H_{e_j}(s))^2} + O(n^{-1/2}),$$

and since $\hat{H}_{e_j0}(y) - H_{e_j0}(y) = O_P(n^{-1/2})$, we obtain

$$\int_{-\infty}^{y} \frac{\hat{H}_{e_j0}(s) - H_{e_j0}(s)}{(1-H_{e_j0}(s))^2} dH_{e_j10}(s) = \int_{-\infty}^{y} \frac{\hat{H}_{e_j0}(s) - H_{e_j0}(s)}{(1-H_{e_j}(s))^2} dH_{e_j1}(s) + o_P(n^{-1/2}). \quad (29)$$

Using (28) and (29), as in the proof of Theorem 2, we have that

$$\hat{F}_{\varepsilon_j0}(y) - F_{\varepsilon_j0}(y)$$
$$= (1-F_{\varepsilon_j0}(y))\left[\int_{-\infty}^{y} \frac{\hat{H}_{e_j0}(s) - H_{e_j0}(s)}{(1-H_{e_j}(s))^2} dH_{e_j1}(s) + \int_{-\infty}^{y} \frac{d(\hat{H}_{e_j10}(s) - H_{e_j10}(s))}{1-H_{e_j}(s)}\right]$$
$$+ o_P(n^{-1/2}).$$

From the proof of Theorem 2, we also have

$$\hat{F}_{\varepsilon_j}(y) - F_{\varepsilon_j}(y)$$
$$= (1-F_{\varepsilon_j}(y))\left[\int_{-\infty}^{y} \frac{\hat{H}_{e_j}(s) - H_{e_j}(s)}{(1-H_{e_j}(s))^2} dH_{e_j1}(s) + \int_{-\infty}^{y} \frac{d(\hat{H}_{e_j1}(s) - H_{e_j1}(s))}{1-H_{e_j}(s)}\right]$$
$$+ o_P(n_j^{-1/2}).$$

Now write

$$F_{\varepsilon_j0}(y) = P\left(\frac{Y_j - m_n(X_j)}{\sigma_j(X_j)} \le y\right) = P\left(\frac{Y_j - m_{jn}(X_j)}{\sigma_j(X_j)} - n^{-1/2}\frac{R(X_j) - r_j(X_j)}{\sigma_j(X_j)} \le y\right)$$
$$= \int P\left(\frac{Y_j - m_{jn}(X_j)}{\sigma_j(X_j)} - n^{-1/2}\frac{R(X_j) - r_j(X_j)}{\sigma_j(X_j)} \le y \,\middle|\, X_j = x\right) f_j(x)dx.$$

If we consider the probability inside the integral as a function of y and apply a Taylor expansion, we obtain

$$F_{\varepsilon_j0}(y) = F_{\varepsilon_j}(y) + n^{-1/2} f_{\varepsilon_j}(y) E\left[\frac{R(X_j) - r_j(X_j)}{\sigma_j(X_j)}\right] + o(n^{-1/2}). \quad (30)$$

Straightforward calculations lead to $\eta_j(Z_j, \Delta_j|X_j) = \eta_j^0(Z_j^0, \Delta_j|X_j)$. Following the same steps as in the proof of Theorem 2, using Lemmas 12 and 13, and taking into

account that $F_{\varepsilon_j 0}(y) = F_{\varepsilon_j}(y) + O(n^{-1/2})$, we obtain the expressions

$$\begin{aligned}
&\hat{F}_{\varepsilon_j 0}(y) - F_{\varepsilon_j 0}(y) \qquad (31)\\
&= \frac{1}{n_j}\sum_{i=1}^{n_j} \xi_{e_j}(E_{ij}, \Delta_{ij}, y) - \frac{1}{n_j}\sum_{i=1}^{n_j}(1 - F_{\varepsilon_j}(y))\zeta_j(Z_{ij}, \Delta_{ij}|X_{ij})\gamma_{j2}(y|X_{ij})\\
&-\frac{1}{n}\sum_{l=1}^{k}\sum_{i=1}^{n_l}(1 - F_{\varepsilon_j}(y))\frac{f_j(X_{il})}{f_{mix}(X_{il})}\frac{\sigma_l(x)}{\sigma_j(X_{il})}\eta_l^0(Z_{il}^0, \Delta_{il}|X_{il})\gamma_{j1}(y|X_{il}) + o_P(n^{-1/2}),
\end{aligned}$$

and

$$\begin{aligned}
&\hat{F}_{\varepsilon_j}(y) - F_{\varepsilon_j}(y) \qquad (32)\\
&= \frac{1}{n_j}\sum_{i=1}^{n_j} \xi_{e_j}(E_{ij}, \Delta_{ij}, y) - \frac{1}{n_j}\sum_{i=1}^{n_j}(1 - F_{\varepsilon_j}(y))\zeta_j(Z_{ij}, \Delta_{ij}|X_{ij})\gamma_{j2}(y|X_{ij})\\
&-\frac{1}{n_j}\sum_{i=1}^{n_j}(1 - F_{\varepsilon_j}(y))\eta_j^0(Z_{ij}^0, \Delta_{ij}|X_{ij})\gamma_{j1}(y|X_{ij}) + o_P(n_j^{-1/2}).
\end{aligned}$$

Finally, by combining expressions (30), (31) and (32) we obtain the representation given in the statement of the Theorem. The leading term of the obtained representation does not depend on n, because the functions η_j^0 are defined in terms of distributions of random variables which do not depend on n and the functions γ_{j1} are defined in terms of distributions of residuals.

Proof of Theorem 6. The leading term of the representation given in Theorem 5 when working under $H_{l.a.}$

$$n^{1/2}\sum_{l=1}^{k} p_l \left\{ n_l^{-1}\sum_{i=1}^{n_l} \psi_{jl}^0(X_{il}, Z_{il}^0, \Delta_{il}, y)\right\}$$

equals the leading term of the representation given under H_0 in Theorem 2

$$n^{1/2}\sum_{l=1}^{k} p_l \left\{ n_l^{-1}\sum_{i=1}^{n_l} \psi_{jl}(X_{il}, Z_{il}, \Delta_{il}, y)\right\},$$

where m^0 in the first expression above plays the role of m in the second one. Hence the asymptotic behavior is the same and the weak convergence follows immediately.

Proof of Corollary 7. The convergence of the test statistics under the alternative hypothesis $H_{l.a.}$ can be obtained in the same way as the proof of Corollary 4, by only taking into account the weak convergence of the process established in Theorem 6.

Acknowledgements

The research of Juan Carlos Pardo-Fernández is supported by the Spanish Ministerio de Ciencia y Tecnología (with additional European FEDER support), through the project BFM2002-03213, and by the Vicerreitorado de Investigación of the Universidade de Vigo. The research of Ingrid Van Keilegom is supported by IAP research network nr. P5/24 of the Belgian government (Belgian Science Policy).

An important part of this work was done when Juan Carlos Pardo-Fernández visited the Institut de Statistique (Université catholique de Louvain, Belgium), with a travel grant from the Dirección Xeral de Investigación e Desenvolvemento of the Xunta de Galicia.

References

Akritas, M. (1986) Bootstrapping the Kaplan-Meier estimator. *J. Amer. Statist. Assoc.* **81**, 1032-1038.

Beran, R. (1981) Nonparametric regression with randomly censored survival data. Technical report, University of California, Berkeley.

Efron, B (1981) Censored data and the bootstrap. *J. Amer. Statist. Assoc.* **76**, 312-319.

Fan, J. and Gijbels, I. (1994) Censored regression: local linear approximations and their applications. *J. Amer. Statist. Assoc.* **89**, 560-570.

Feller, W. (1966) *An Introduction to Probability Theory and Its Applications*, Vol.II. Wiley, New York.

González-Manteiga, W., Cao, R. and Marron, J. S. (1996) Bootstrap selection of the smoothing parameter in nonparametric hazard estimation. *J. Amer. Statist. Assoc.* **91**, 1130-1140.

Kaplan, E. L. and Meier, P. (1958) Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

Koziol, J. A. and Green,S. B. (1976) A Cramér-von Mises statistic for randomly censored data. *Biometrika* **63**, 465-474.

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

Pardo-Fernández, J. C., Van Keilegom, I. and González-Manteiga, W. (2004) Testing for the equality of k regression curves. *Submitted.*

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.

van der Vaart, A. W. and Wellner, J. A. (1996) *Weak convergence and empirical processes.* Springer, New York.

Van Keilegom, I. and Akritas, M. (1999) Transfer of tail information in censored regression models. *Ann. Statist.* **27**, 1745-1784.

Ying, Z., Jung, S.H. and Wei, J. (1995) Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90**, 178-184.

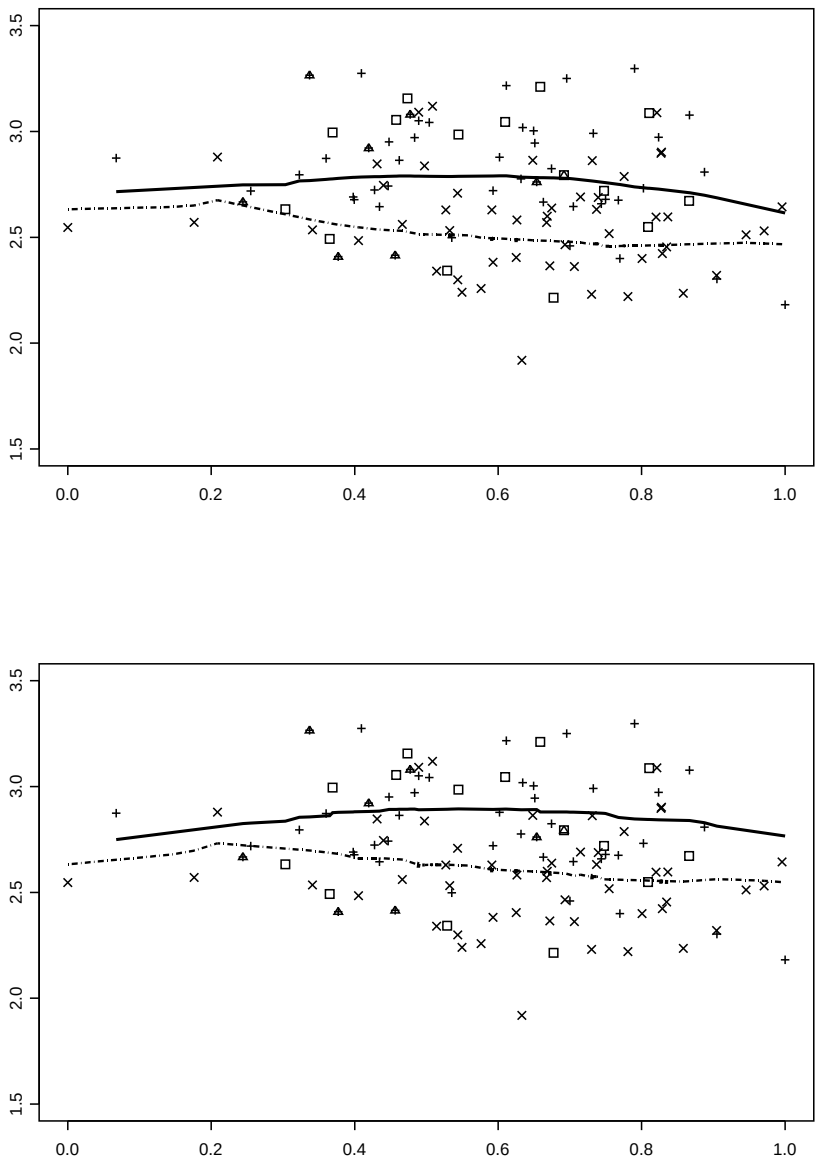

Figure 1: *Scatter plot of 'log$_{10}$(survival time)' versus 'age' (rescaled to* $[0,1]$*) and estimated regression curves of Group A (solid line, + for uncensored data,* $\square$ *for censored data) and Group B (dashed line,* $\times$ *for uncensored data,* $\triangle$ *for censored data), with* $J(s) = 0.75^{-1}I(0 \leq s \leq 0.75)$ *(top) and* $J(s) = 0.50^{-1}I(0.25 \leq s \leq 0.75)$ *(bottom).*

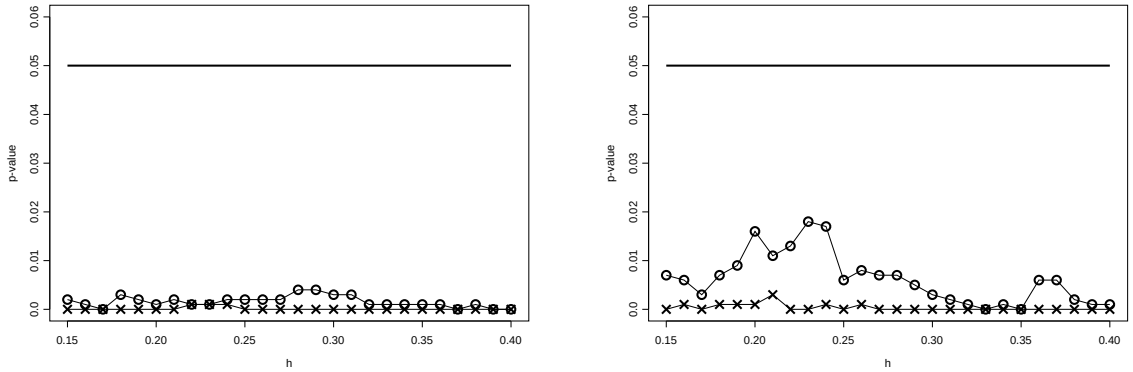

Figure 2: *Graphs of the p-values as function of the bandwidth h when testing for the equality of the regression curves with the test statistics T_{KS} (line with circles) and T_{CM} (line with crosses). The curves were estimated using $J(s) = 0.75^{-1}I(0 \leq s \leq 0.75)$ (left) and $J(s) = 0.50^{-1}I(0.25 \leq s \leq 0.75)$ (right). The solid horizontal line corresponds to a p-value of 0.05.*

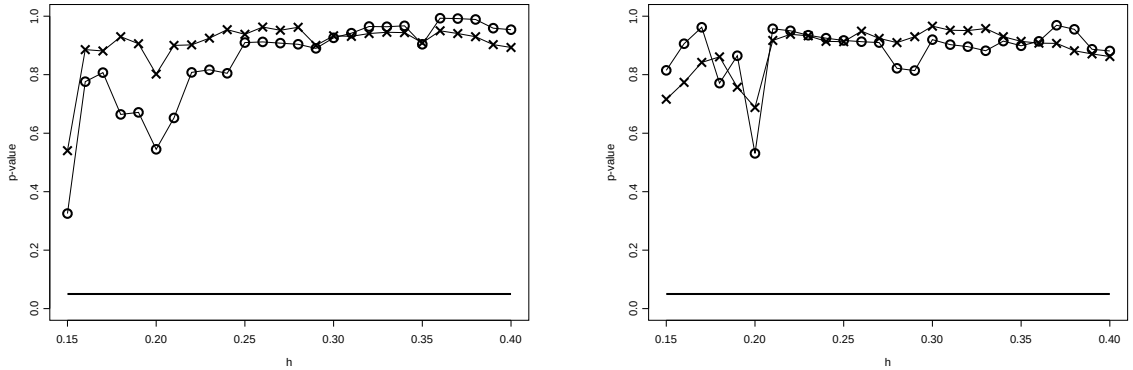

Figure 3: *Graphs of the p-values as function of the bandwidth h when testing for constant difference between the regression curves with the test statistics T_{KS} (line with circles) and T_{CM} (line with crosses). The curves were estimated using $J(s) = 0.75^{-1}I(0 \leq s \leq 0.75)$ (left) and $J(s) = 0.50^{-1}I(0.25 \leq s \leq 0.75)$ (right). The solid horizontal line corresponds to a p-value of 0.05.*