0536

GOODNESS-OF-FIT TESTS FOR PARAMETRIC MODELS IN CENSORED REGRESSION

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

Goodness-of-fit tests for parametric models in censored regression

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

July 21, 2005

Abstract

In this article we introduce a goodness-of-fit test for parametric regression models when the response variable is right censored. The test is based on the comparison of a parametric estimator and a nonparametric estimator of the distribution of the residuals. Kolmogorov-Smirnov and Cramér-von Mises type statistics are proposed. A bootstrap mechanism is used to approximate the critical values of the test. Some simulations are included and a real data set is analyzed.

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

[*]Departamento de Estatística e IO. Universidade de Vigo. Research 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.

[†]Institut de Statistique. Université catholique de Louvain. Research supported by IAP research network nr. P5/24 of the Belgian government (Belgian Science Policy).

[‡]Departamento de Estatística e IO. Universidade de Santiago de Compostela. Research supported by the Spanish Ministerio de Ciencia y Tecnología (with additional European FEDER support) through the project BFM2002-03213 and from Xunta de Galicia through project PGIDIT03PXIC20702PN.

1 Introduction and statistical model

The aim of regression models consists of describing the relationship between a response and a covariate. In many practical situations parametric regression models are appealing. They describe the relationship between the response and the covariate in a simple way and usually allow for interpretability of the parameters (for instance in linear regression). Nevertheless, if the parametric model fails then the conclusions will be erroneous. Any parametric analysis should be accompanied by a test to check its validity and avoid misspecification and wrong conclusions. This motivates the development of specific goodness-of-fit tests for parametric models in regression.

When the response variable of the regression model is a survival time it is useful to allow for censoring in the statistical model. Fan and Gijbels (1994) motivate the development of analytic tools for censored data: 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.

In the context of censored data the statistical model can be described as follows. Let (X, Y) be a random vector, where Y represents a certain response variable associated to the covariate X. Assume that the response variable Y is subject to random right censoring. This means that there exists a censoring variable C, independent of Y given X, such that we observe $Z = \min\{Y, C\}$ and the indicator of censoring $\Delta = I(Y \le C)$. Consider the following non-parametric regression model:

$$Y = m(X) + \sigma(X)\varepsilon. \tag{1}$$

The error variable ε is independent of X, m is an unknown conditional location function

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

and σ is an unknown conditional scale function representing possible heteroscedasticity

$$\sigma^2(x) = \int_0^1 F^{-1}(s|x)^2 J(s)ds - m^2(x), \tag{3}$$

where $F(\cdot|x)$ is the conditional distribution of Y given $X = x$, $F^{-1}(s|x) = \inf\{y; F(y|x) \ge s\}$ is the corresponding quantile function and $J(s)$ is a score function satisfying $\int_0^1 J(s)ds = 1$ (in general, for any distribution function F we denote $F^{-1}(s) = \inf\{y; F(y) \ge s\}$ for the corresponding quantile function and $\tau_F = \inf\{y; F(y) = 1\}$). Let F_ε be the distribution of the error ε. By construction $\int_0^1 F_\varepsilon^{-1}(s)J(s)ds = 0$ and $\int_0^1 F_\varepsilon^{-1}(s)^2 J(s)ds = 1$. The sample consists of n independent replications (X_i, Z_i, Δ_i), $i = 1, \ldots, n$, from the distribution of (X, Z, Δ).

The choice of the score function J leads to different location and scale functions. In particular if $J(s) = I(0 \le s \le 1)$ then m and σ^2 are the conditional mean and conditional variance respectively. However, it may happen that this choice of J is not appropriate because of the inconsistency of the estimator of the conditional distribution $F(\cdot|x)$ in the right tail due to the censoring. An interesting choice in the context of censored data is $J(s) = (q-p)^{-1} I(p \le s \le q)$, for some $0 \le p \le q \le 1$, which leads to trimmed means and trimmed variances. The conditional median or other conditional quantiles can be seen as limits of trimmed means.

Given a particular parametric class of regression functions $\mathcal{M} = \{m_\theta; \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^p$, for some $p \ge 1$, we are interested in testing the null hypothesis

$$H_0 : m \in \mathcal{M}, \tag{4}$$

versus the general alternative

$$H_1 : m \notin \mathcal{M}.$$

Note that in the case of homoscedasticity, if the parametric model is of the form $m_\theta(\cdot) = \theta_1 + m_{\theta_2}(\cdot)$ (containing a constant as an additive term) and H_0 holds for a particular definition of the location function, that is, for a particular choice of the score function J, then it will necessarily hold for all possible location functions.

The test is based on the comparison of two estimators of the distribution of the errors. Assume that $\hat{\theta}$ is an estimator of the parameter under the null hypothesis and $\hat{m}$ is a nonparametric estimator of the regression function. We compare the distribution of the residuals estimated in a parametric way $(Z_i - m_{\hat{\theta}}(X_i))/\hat{\sigma}(X_i)$ with the distribution of the residuals estimated in a completely nonparametric way $(Z_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$. These residuals are censored because they are calculated with respect to the observed times Z_i and not with respect to the actual (and not observable) times Y_i. We will compare the corresponding Kaplan-Meier estimators of their distributions via Kolmogorov-Smirnov and Cramér-von Mises type statistics. This idea was developed by Van Keilegom, González-Manteiga and Sánchez-Sellero (2005) for non censored data.

Several goodness-of-fit tests for complete data have been proposed in the literature along the last decade. See e.g. Van Keilegom, González-Manteiga and Sánchez-Sellero (2005) for an overview of recent papers on this subject. For censored data Stute, González-Manteiga and Sánchez-Sellero (2000) studied a test based on the marked empirical process of the integrated regression function. Their setup is somewhat different and more restrictive than ours since they assume independence between the response Y and the censoring variable C and focus on the conditional mean.

The rest of this paper is organized as follows. In Section 2 we describe in detail the testing procedure. Section 3 contains some asymptotic results. In Section 4 we propose a bootstrap mechanism to approximate the critical values of the test and in Section 5 we study its practical behavior by means of simulations. A real data set is analyzed in Section 6. Finally, the Appendix contains the proofs of the asymptotic results.

2 Testing procedure

The testing procedure is based on the comparison of two estimators of the distribution of the errors F_ε. This involves nonparametric estimation of the location and scale functions and estimation of the parameter θ under the null hypothesis.

The nonparametric estimators will be constructed using the estimator of the conditional distribution function $F(\cdot|x)$ of Y given the value x of the covariate X when the response is censored introduced by Beran (1981) and studied, among others, by González-Manteiga and Cadarso-Suárez (1994) and Van Keilegom and Veraverbeke (1997):

$$\hat{F}(y|x) = 1 - \prod_{Z_i \leq y, \Delta_i = 1} \left(1 - \frac{W_i(x, h_n)}{\sum_{j=1}^n I(Z_j \geq Z_i) W_j(x, h_n)} \right), \tag{5}$$

where

$$W_i(x, h_n) = \frac{K((x - X_i)/h_n)}{\sum_{j=1}^n K((x - X_j)/h_n)}$$

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

The nonparametric estimator of the location function is

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

and the estimator of the scale function is

$$\hat{\sigma}^2(x) = \int_0^1 \hat{F}^{-1}(s|x)^2 J(s) ds - \hat{m}^2(x). \tag{7}$$

Compute the estimators of the censored residuals

$$\hat{E}_i = \frac{Z_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)} \tag{8}$$

and estimate the distribution of the residuals from the censored sample $(\hat{E}_i, \Delta_i)$, $i = 1, \ldots, n$, through the product-limit estimator introduced by Kaplan and Meier (1958)

$$\hat{F}_\varepsilon(y) = 1 - \prod_{\hat{E}_i \le y, \Delta_i = 1} \left(1 - \frac{1}{\sum_{j=1}^n I(\hat{E}_j \ge \hat{E}_i)}\right). \tag{9}$$

This estimator of the error distribution has been proposed and studied by Van Keilegom and Akritas (1999).

If the null hypothesis is true there exists a value of the parameter θ_0 such that $m = m_{\theta_0}$. For the moment we assume that we have an estimator $\hat{\theta}$ of the true parameter θ_0 under H_0. We will discuss about this issue later. The residuals based on the parametric estimation of the location function $m_{\hat{\theta}}$ are

$$\hat{E}_{i0} = \frac{Z_i - m_{\hat{\theta}}(X_i)}{\hat{\sigma}(X_i)}. \tag{10}$$

Note that we are keeping the nonparametric estimator of the variance function. Estimate the corresponding distribution from the censored sample $(\hat{E}_{i0}, \Delta_i)$, $i = 1, \ldots, n$,

$$\hat{F}_{\varepsilon 0}(y) = 1 - \prod_{\hat{E}_{i0} \le y, \Delta_i = 1} \left(1 - \frac{1}{\sum_{j=1}^n I(\hat{E}_{j0} \ge \hat{E}_{i0})}\right). \tag{11}$$

Under the null hypothesis $\hat{F}_\varepsilon$ and $\hat{F}_{\varepsilon 0}$ are both estimators of F_ε. The fact that there exist differences between these two estimators of the distribution of the errors gives evidence for the alternative hypothesis. This idea is formalized in the following Theorem, in the sense that the equality of the theoretical versions of the distributions considered in (9) and (11), i.e., $F_\varepsilon(y) = P((Y - m(X))/\sigma(X) \le y)$ and $F_{\varepsilon 0}(y) = P((Y - m_{\theta_0}(X))/\sigma(X) \le y)$, for $\theta_0 \in \Theta$, characterizes the null hypothesis. The proof can be found in the Appendix.

Theorem 1 *Let m be a continuous function. Then H_0 holds if and only if there exists $\theta_0 \in \Theta$ such that $F_\varepsilon(y) = F_{\varepsilon 0}(y)$, $-\infty < y < \infty$.*

Let $H_e(y) = P((Z - m(X))/\sigma(X) \le y)$ and let T be any point smaller than τ_{H_e}. The goodness-of-fit test is carried out through the process

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

$-\infty < y \le T$. More precisely, we propose a Kolmogorov-Smirnov type statistic

$$T_{KS} = \sup_{-\infty < y \le T} |\hat{W}(y)| \tag{12}$$

and a Cramér-von Mises type statistic

$$T_{CM} = \int_{-\infty}^{T} \hat{W}^2(y) d\hat{F}_{\varepsilon 0}(y). \tag{13}$$

The null hypothesis (4) is rejected for large values of the test statistics.

Estimation of the parameter θ_0 under the null hypothesis. In the theoretical results we will show in the next section, we assume that the estimator of the parameter under the null hypothesis admits an asymptotic representation of the form

$$\hat{\theta} - \theta_0 = n^{-1} \sum_{i=1}^{n} \vec{\omega}(X_i, Z_i, \Delta_i) + (o_P(n^{-1/2}), \ldots, o_P(n^{-1/2}))^t, \tag{14}$$

where $\vec{\omega}(x, z, \delta) = (\omega_1(x, z, \delta), \ldots, \omega_p(x, z, \delta))^t$ verifies, for $k = 1, \ldots, p$, $E(\omega_k(X, Z, \Delta)) = 0$ and $Var(\omega_k(X, Z, \Delta)) < \infty$. The theory we will develop in Section 3 will be valid for any estimator verifying the representation (14).

Akritas (1996) introduced a parameter estimate for polynomial models which verifies (14). Basically the estimator of the parameter is the least squares estimator based on a preliminary non parametric estimation. Given a polynomial model $m(x) = \theta_1 + \theta_2 x + \cdots + \theta_p x^{p-1}$, the estimate of the parameter $\theta_0 = (\theta_1, \ldots, \theta_p)^t$ is $\hat{\theta} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \hat{\mathbf{m}}$, where $\hat{\mathbf{m}}^t = (\hat{m}(X_1), \ldots, \hat{m}(X_n))$ and $\mathbf{X}$ is the design matrix of dimension $n \times p$ whose (i, k) element is X_i^{k-1}, $i = 1, \ldots, n$, $k = 1, \ldots, p$. We will use this estimator in the simulations of Section 5.

Other references concerning parameter estimation in censored regression can be found in Stute (1999), although most of them are devoted to modelling the conditional mean parametrically.

3 Asymptotic results

In this section we state the main asymptotic results associated to the testing procedure we described in the previous section. In Theorem 2 we give an asymptotic representation for the difference between the two estimators of the distribution of the residuals and in Theorem 3 we state the weak convergence of the corresponding process. In Corollary 4 we obtain the asymptotic distributions of the test statistics under the null hypothesis. These results extend the equivalent ones obtained by Van Keilegom, González-Manteiga and Sánchez-Sellero (2005) for complete data. The proofs can be found in the Appendix.

The notation used in this section is the following: $F(x) = P(X \leq x)$, $F(y|x) = P(Y \leq y|X = x)$, $G(y|x) = P(C \leq y|X = x)$, $H(y|x) = P(Z \leq y|X = x)$, $H_1(y|x) = P(Z \leq y, \Delta = 1|X = x)$, $F_\varepsilon(y) = P(\varepsilon \leq y) = P((Y - m(X))/\sigma(X) \leq y)$ and $F_{\varepsilon 0}(y) = P((Y - m_{\theta_0}(X))/\sigma(X) \leq y)$. We denote $E = (Z - m(X))/\sigma(X)$ and $H_e(y) = P(E \leq y)$, $H_{e1}(y) = P(E \leq y, \Delta = 1)$, $H_e(y|x) = P(E \leq y|X = x)$, $H_{e1}(y|x) = P(E \leq y, \Delta = 1|X = x)$. We also denote $E_0 = (Z - m_{\theta_0}(X))/\sigma(X)$ and $H_{e0}(y) = P(E_0 \leq y)$, $H_{e10}(y) = P(E_0 \leq y, \Delta = 1)$, $H_{e0}(y|x) = P(E_0 \leq y|X = x)$, $H_{e10}(y|x) = P(E_0 \leq y, \Delta = 1|X = x)$. The derivatives of these functions will be denoted with the corresponding lower case letters.

The functions $H_{e0}(y)$, $H_{e10}(y)$, $H_e(y)$ and $H_{e1}(y)$ are estimated by the empirical distribution functions based on the corresponding estimated censored residuals

$$\hat{H}_{e0}(y) = n^{-1} \sum_{i=1}^{n} (\hat{E}_{i0} \leq y),$$

$$\hat{H}_{e10}(y) = n^{-1} \sum_{i=1}^{n} (\hat{E}_{i0} \leq y, \Delta_i = 1),$$

$$\hat{H}_e(y) = n^{-1} \sum_{i=1}^{n} (\hat{E}_i \leq y)$$

and

$$\hat{H}_{e1}(y) = n^{-1} \sum_{i=1}^{n} (\hat{E}_i \leq y, \Delta_i = 1)$$

The following functions also appear in the results below:

$$\xi(z, \delta, y|x) = (1 - F(y|x)) \left[- \int_{-\infty}^{y \wedge z} \frac{dH_1(s|x)}{(1 - H(s|x))^2} + \frac{I(z \leq y, \delta = 1)}{1 - H(z|x)} \right],$$

$$\eta(z, \delta|x) = \sigma^{-1}(x) \int_{-\infty}^{+\infty} \xi(z, \delta, v|x) J(F(v|x)) dv,$$

$$\zeta(z, \delta|x) = \sigma^{-1}(x) \int_{-\infty}^{+\infty} \xi(z, \delta, v|x) J(F(v|x)) \frac{v - m(x)}{\sigma(x)} dv,$$

$$\gamma_0(y|x) = \int_{-\infty}^{y} \frac{h_{e0}(s|x) h_{e10}(s)}{(1 - H_{e0}(s))^2} ds + \int_{-\infty}^{y} \frac{dh_{e10}(s|x)}{1 - H_{e0}(s)},$$

The regularity assumptions needed for the theoretical results are

(A1) (i) X has convex and compact support R_X.
(ii) f is two times continuously differentiable and $\inf_{x \in R_X} f(x) > 0$.
(iii) m and σ are two times continuously differentiable and $\inf_{x \in R_X} \sigma(x) > 0$.

(A2) (i) $nh_n^4 \to 0$ and $nh_n^{3+2\delta}(\log h_n^{-1})^{-1} \to \infty$ for some $\delta > 0$.
(ii) K is a symmetric density function with compact support and K is twice continuously differentiable.
(iii) J is twice continuously differentiable on the interior of its support, $\int_0^1 J(s)ds = 1$ and $J(s) \geq 0$ for all $0 \leq s \leq 1$.
(iv) Let $\tilde{T}_x$ be any value less than the upper bound of the support of $H(\cdot|x)$ such that $\inf_{x\in R_X}(1 - H(\tilde{T}_x|x)) > 0$. Then there exist $0 \leq s_0 \leq s_1 \leq 1$ such that $s_1 \leq \inf_x F(\tilde{T}_x|x)$, $s_0 \leq \inf\{s \in [0,1], J(s) \neq 0\}$, $s_1 \geq \sup\{s \in [0,1], J(s) \neq 0\}$ and $\inf_{x\in R_X} \inf_{s_0\leq s\leq s_1} f(F^{-1}(s|x)|x) > 0$.

(A3) The functions η and ζ are twice continuously differentiable with respect to x and their first and second derivatives are bounded, uniformly in $x \in R_X$, $z < \tilde{T}_x$ and δ.

(A4) $H(y|x)$, $H_1(y|x)$, $H_{e0}(y|x)$, $H_{e10}(y|x)$, $H_e(y|x)$ and $H_{e1}(y|x)$ are continuously differentiable with respect to x and y up to order three.

(A5) (i) Θ is a compact subset of $\mathbb{R}^p$.
(ii) $m_\theta(x)$ is two times continuously differentiable with respect to θ for all $x \in R_X$.

Assuming (A5) and the representation given in (14), a Taylor expansion of $m_\theta(u)$ as function of θ around θ_0 leads to

$$m_{\hat{\theta}}(u) - m_{\theta_0}(u) = n^{-1}\sum_{i=1}^n \varphi_{\theta_0}(u, X_i, Z_i, \Delta_i) + o_P(n^{-1/2}),$$

where

$$\varphi_{\theta_0}(u, x, z, \delta) = \left(\frac{\partial m_\theta(u)}{\partial\theta_1}\bigg|_{\theta=\theta_0}, \ldots, \frac{\partial m_\theta(u)}{\partial\theta_p}\bigg|_{\theta=\theta_0} \right) \vec{\omega}(x, z, \delta). \tag{15}$$

Theorem 2 *Assume (A1)-(A5). Then, under the null hypothesis H_0,*

$$\hat{F}_{\varepsilon 0}(y) - \hat{F}_\varepsilon(y) = (1 - F_\varepsilon(y))n^{-1}\sum_{i=1}^n \psi_{\theta_0}(X_i, Z_i, \Delta_i, y) + o_P(n^{-1/2})$$

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

$$\psi_{\theta_0}(x, z, \delta, y) = \int \sigma^{-1}(u)\varphi_{\theta_0}(u, x, z, \delta)f(u)\gamma_0(y|u)du + \gamma_0(y|x)\eta(z, \delta|x).$$

Theorem 3 *Assume (A1)-(A5). Then, under the null hypothesis H_0, the process $\hat{W}(y) = n^{1/2}(\hat{F}_{\varepsilon 0}(y) - \hat{F}_\varepsilon(y))$, $-\infty < y \leq T$ converges weakly to a centered Gaussian process $W(y)$ with covariance function*

$$Cov(W(y), W(y')) = (1 - F_\varepsilon(y))(1 - F_\varepsilon(y'))E(\psi_{\theta_0}(X, Z, \Delta, y), \psi_{\theta_0}(X, Z, \Delta, y')).$$

Corollary 4 *Assume (A1)-(A5). Then, under the null hypothesis* H_0*,*

$$T_{KS} \xrightarrow{d} \sup_{-\infty<y<T} |W(y)|$$
$$T_{CM} \xrightarrow{d} \int_{-\infty}^{T} W^2(y) dF_\varepsilon(y).$$

4 Bootstrap approximation

The asymptotic distributions of the test statistics under the null hypothesis given in Corollary 4 are complicated. We propose a bootstrap procedure in order to approximate the critical values of the test in practical situations. The resampling procedure is based on a smoothed version of the 'naive bootstrap' described in Efron (1981) and studied in Akritas (1986).

For $i = 1, \ldots, n$, estimate the censored residuals in a nonparametric way

$$\hat{E}_i = \frac{Z_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)}. \tag{16}$$

From the censored sample of estimated residuals $(\hat{E}_i, \Delta_i)$, $i = 1, \ldots, n$, compute the Kaplan-Meier estimator $\hat{F}_\varepsilon$ and standardize these residuals in order to verify the initial assumption of having location function 0 and scale function 1 (if $\lambda_1 = \int \hat{F}_\varepsilon^{-1}(s)J(s)ds$ and $\lambda_2 = (\int \hat{F}_\varepsilon^{-1}(s)^2 J(s)ds - \lambda_1^2)^{1/2}$ then the standardized residuals are $\tilde{E}_i = (\hat{E}_i - \lambda_1)/\lambda_2$).

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

1. For $i = 1, \ldots, n$:

   - Let $Y_{i,b}^* = m_{\hat{\theta}}(X_i) + \hat{\sigma}(X_i)\varepsilon_{i,b}^*$, where $\varepsilon_{i,b}^* = V_{i,b} + aS_{i,b}$, $V_{i,b}$ is drawn from $\hat{F}_\varepsilon$ (standardized), $S_{i,b}$ is a random variable with mean zero and variance one to introduce a small perturbation in the residuals (the perturbation is controlled by the constant a). Note that the bootstrap responses follow the null hypothesis by construction.
   - Select at random a $C_{i,b}^*$ from a smoothed version of $\hat{G}(\cdot|X_i)$, which is the Beran estimator of the conditional distribution $G(\cdot|X_i)$ of the censoring variable obtained by replacing Δ_i by $1 - \Delta_i$ in the expression of $\hat{F}(\cdot|X_i)$.

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

2. The bootstrap sample is $\{(X_i, Z^*_{i,b}, \Delta^*_{i,b}), i = 1, \ldots, n\}$.

3. Let $T^*_{KS,b}$ and $T^*_{CM,b}$ be the test statistics calculated with the bootstrap sample.

Let $T^*_{KS,(b)}$ be the b-th order statistic of $T^*_{KS,1}, \ldots, T^*_{KS,B}$, and analogously for $T^*_{CM,(b)}$. 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 Simulations

We present some simulation results in order to study the finite-sample behavior of the goodness-of-fit test and the bootstrap approximation of the critical values.

The regression and variance functions are those corresponding to the choice $J(s) = 0.75^{-1} I(0 \leq s \leq 0.75)$. We consider the following regression functions:

$$
\begin{array}{ll}
(i) & m(x) = x \\
(ii) & m(x) = x + 0.6(x - 0.5) \\
(iii) & m(x) = x + 2x^2 \\
(iv) & m(x) = x + 0.5\sin(4\pi x)
\end{array}
$$

The variance function is $\sigma^2(x) = 0.5$. The covariate is uniformly distributed in the interval $[0, 1]$ and the error is exponentially distributed, transformed such that $\int_0^1 F_\varepsilon^{-1}(s)J(s)ds = 0$ and $\int_0^1 F_\varepsilon^{-1}(s)^2 J(s)ds = 1$. The censoring variable is $C = m(X) + \sigma(X)\rho$, where ρ is independent of ε and has survival function $1 - F_\rho(y) = (1 - F_\varepsilon(y))^\beta$, with $\beta = 1/3$ (25% of censoring) and $\beta = 1$ (50% of censoring).

We will test for two different null hypotheses: a complete linear model

$$H_0 : m(x) = \theta_1 + \theta_2 x \tag{17}$$

(in this case models (i) and (ii) correspond to the null hypothesis and (iii) and (iv) correspond to the alternative hypothesis) and a linear model through the origin

$$H_0 : m(x) = \theta x \tag{18}$$

(in this case model (i) corresponds to the null hypothesis and models (ii)-(iv) correspond to the alternative hypothesis). The parameter is estimated by using the method proposed by Akritas (1996) for polynomial models described in Section 2.

Table 1 shows the rejection probabilities in 1000 trials for sample sizes $n = 100$ and $n = 200$ and significance levels $\alpha = 0.05$ and $\alpha = 0.10$. We choose the kernel of Epanechnikov $K(u) = 0.75(1-u^2)I(|u| < 1)$ to calculate the weights that appear in the Beran estimator. The bandwidth was chosen of the form $h = cn^{-3/10}$ and the cases $c = 0.75$ and $c = 1$ are displayed. In the bootstrap we use $B = 200$ replications, $a = n^{-3/10}$ and $S_{i,b}$ is a standard normal.

The threshold T in the definition of the test statistics was chosen to be the largest observed value of the sample of censored residuals estimated under the null hypothesis. The level is well approximated in most cases, although this approximation gets worse when the data are heavily censored (50% for censoring). The behavior of the power is as expected: it increases with the sample size and it decreases with the amount of censored data. Model (*iii*) is very difficult to distinguish from a linear model, especially when the amount of censored data is 50%.

We believe that the choice of the threshold T may have an impact on the power. When the data are heavily censored, the Kaplan-Meier estimators have large jumps in the right tail of the distribution. This may produce large values of the test statistics even under the null hypothesis. In Table 2 we have repeated the same simulations by using as a threshold the quantile of order $\hat{F}_{\varepsilon 0}^{-1}(\hat{F}_{\varepsilon 0}(+\infty) - 0.10)$ to avoid this problem. Now the level behaves reasonably well and the power is better than in the previous table, especially when the amount of censoring is 50% and the sample size is 200.

6 Data analysis

We illustrate the proposed goodness-of-fit test on a data set concerning male patients suffering from larynx cancer, diagnosed and treated during the period 1970-1978 in the Netherlands. More details about this data set can be found in Kardaun (1983). The variable of interest is the time between first treatment and death. At the end of the study 40 patients were alive (their survival times are censored). Heuchenne and Van Keilegom (2005) suggested a linear model to explain the relationship between the log of the age of the patient at diagnosis as a covariate and the log of the survival time as response. These authors work with the conditional mean.

Figure 1 shows the data and regression curves estimated nonparametrically with score functions $J(s) = 0.50^{-1}I(0.25 \leq s \leq 0.75)$ and $J(s) = 0.75^{-1}I(0 \leq s \leq 0.75)$. We believe that the choice of the score function is not crucial here since the data seem to be

Table 1: Rejection probabilities (models i-iv) of the tests based on T_{KS} and T_{CM}. The threshold T is the largest observed value of the sample of censored residuals estimated under the null hypothesis.

			$h = 0.75n^{-3/10}$				$h = n^{-3/10}$			
			T_{KS}		T_{CM}		T_{KS}		T_{CM}	
% Cens.	n	Model	0.050	0.100	0.050	0.100	0.050	0.100	0.050	0.100
					$H_0 : m(x) = \theta_1 + \theta_2 x$					
25%	100	(i)	0.054	0.098	0.049	0.097	0.043	0.086	0.042	0.096
		(ii)	0.044	0.097	0.048	0.105	0.039	0.080	0.041	0.094
		(iii)	0.124	0.202	0.130	0.220	0.099	0.165	0.122	0.210
		(iv)	0.290	0.423	0.307	0.472	0.201	0.331	0.214	0.389
	200	(i)	0.052	0.108	0.047	0.102	0.050	0.114	0.053	0.110
		(ii)	0.044	0.110	0.047	0.106	0.038	0.096	0.051	0.106
		(iii)	0.260	0.380	0.263	0.375	0.236	0.365	0.267	0.387
		(iv)	0.727	0.838	0.760	0.877	0.604	0.761	0.723	0.848
50%	100	(i)	0.039	0.118	0.065	0.135	0.020	0.080	0.036	0.115
		(ii)	0.038	0.103	0.055	0.136	0.030	0.086	0.042	0.116
		(iii)	0.038	0.106	0.049	0.117	0.036	0.084	0.040	0.097
		(iv)	0.066	0.174	0.070	0.194	0.050	0.124	0.046	0.126
	200	(i)	0.027	0.110	0.055	0.125	0.021	0.070	0.045	0.118
		(ii)	0.029	0.098	0.050	0.118	0.020	0.081	0.045	0.112
		(iii)	0.040	0.108	0.063	0.137	0.023	0.087	0.045	0.114
		(iv)	0.155	0.310	0.247	0.435	0.080	0.207	0.156	0.327
					$H_0 : m(x) = \theta x$					
25%	100	(i)	0.056	0.112	0.056	0.107	0.057	0.108	0.049	0.098
		(ii)	0.311	0.427	0.356	0.496	0.283	0.382	0.336	0.466
		(iii)	0.439	0.571	0.471	0.601	0.343	0.456	0.438	0.552
		(iv)	0.256	0.398	0.148	0.293	0.225	0.385	0.132	0.279
	200	(i)	0.065	0.118	0.071	0.121	0.061	0.114	0.060	0.117
		(ii)	0.556	0.667	0.596	0.726	0.531	0.636	0.593	0.709
		(iii)	0.749	0.827	0.743	0.824	0.688	0.765	0.732	0.807
		(iv)	0.703	0.850	0.661	0.865	0.707	0.845	0.693	0.864
50%	100	(i)	0.041	0.103	0.060	0.152	0.030	0.078	0.056	0.117
		(ii)	0.098	0.195	0.176	0.321	0.061	0.148	0.134	0.277
		(iii)	0.130	0.242	0.215	0.391	0.061	0.154	0.176	0.331
		(iv)	0.127	0.259	0.079	0.169	0.098	0.213	0.065	0.137
	200	(i)	0.044	0.112	0.059	0.141	0.027	0.088	0.053	0.118
		(ii)	0.166	0.295	0.335	0.499	0.135	0.248	0.307	0.471
		(iii)	0.280	0.426	0.426	0.590	0.190	0.324	0.394	0.561
		(iv)	0.327	0.527	0.356	0.564	0.274	0.457	0.359	0.551

Table 2: Rejection probabilities (models i-iv) of the tests based on T_{KS} and T_{CM}. The threshold T is the quantile of order $\hat{F}_{\varepsilon 0}^{-1}(\hat{F}_{\varepsilon 0}(+\infty) - 0.10)$ of the sample of censored residuals estimated under the null hypothesis.

			$h = 0.75n^{-3/10}$				$h = n^{-3/10}$			
			T_{KS}		T_{CM}		T_{KS}		T_{CM}	
% Cens.	n	Model	0.050	0.100	0.050	0.100	0.050	0.100	0.050	0.100
					$H_0 : m(x) = \theta_1 + \theta_2 x$					
25%	100	(i)	0.063	0.110	0.051	0.107	0.053	0.108	0.054	0.108
		(ii)	0.056	0.112	0.052	0.116	0.052	0.092	0.046	0.100
		(iii)	0.142	0.226	0.152	0.241	0.127	0.206	0.152	0.241
		(iv)	0.340	0.491	0.378	0.542	0.263	0.411	0.290	0.466
	200	(i)	0.062	0.128	0.055	0.114	0.061	0.131	0.061	0.115
		(ii)	0.057	0.130	0.051	0.115	0.061	0.114	0.058	0.109
		(iii)	0.302	0.410	0.277	0.385	0.301	0.431	0.309	0.438
		(iv)	0.768	0.868	0.789	0.898	0.704	0.823	0.782	0.886
50%	100	(i)	0.032	0.060	0.050	0.111	0.019	0.044	0.037	0.112
		(ii)	0.034	0.064	0.052	0.114	0.029	0.055	0.043	0.090
		(iii)	0.064	0.116	0.089	0.146	0.051	0.090	0.059	0.126
		(iv)	0.192	0.333	0.196	0.361	0.112	0.230	0.106	0.258
	200	(i)	0.034	0.071	0.055	0.109	0.036	0.072	0.050	0.101
		(ii)	0.034	0.075	0.054	0.112	0.032	0.059	0.042	0.097
		(iii)	0.114	0.204	0.125	0.226	0.090	0.172	0.120	0.196
		(iv)	0.500	0.669	0.611	0.757	0.392	0.561	0.526	0.670
					$H_0 : m(x) = \theta x$					
25%	100	(i)	0.062	0.119	0.059	0.108	0.063	0.119	0.049	0.104
		(ii)	0.321	0.447	0.370	0.513	0.299	0.406	0.352	0.484
		(iii)	0.462	0.592	0.496	0.617	0.365	0.479	0.455	0.576
		(iv)	0.266	0.421	0.166	0.317	0.250	0.416	0.158	0.319
	200	(i)	0.069	0.123	0.070	0.122	0.066	0.118	0.062	0.118
		(ii)	0.559	0.680	0.606	0.734	0.547	0.644	0.606	0.722
		(iii)	0.766	0.840	0.753	0.833	0.709	0.787	0.740	0.817
		(iv)	0.713	0.869	0.677	0.877	0.735	0.858	0.714	0.887
50%	100	(i)	0.049	0.093	0.072	0.140	0.037	0.072	0.054	0.111
		(ii)	0.185	0.274	0.298	0.418	0.154	0.230	0.262	0.392
		(iii)	0.254	0.369	0.365	0.501	0.157	0.256	0.326	0.466
		(iv)	0.228	0.373	0.142	0.254	0.185	0.326	0.111	0.196
	200	(i)	0.060	0.098	0.067	0.128	0.043	0.084	0.051	0.112
		(ii)	0.393	0.519	0.514	0.654	0.336	0.461	0.496	0.638
		(iii)	0.550	0.670	0.628	0.749	0.451	0.580	0.603	0.719
		(iv)	0.612	0.782	0.594	0.777	0.598	0.747	0.614	0.779

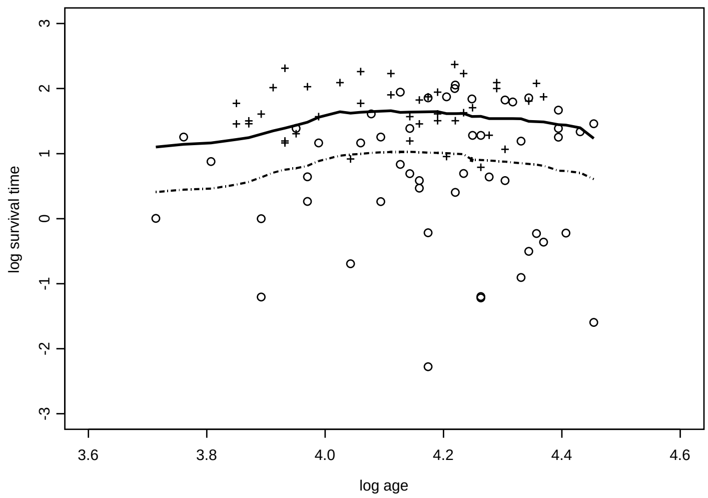

Figure 1: *Scatter plot of 'log(survival time)' versus 'log(age)' (crosses for uncensored data and circles for censored data) and estimated regression curves with* $J(s) = 0.50^{-1}I(0.25 \le s \le 0.75)$ *(solid line) and* $J(s) = 0.75^{-1}I(0 \le s \le 0.75)$ *(dashed line).*

homoscedastic. The two estimated curves are almost parallel.

We have applied our test to verify the claimed linear model with both choices of the function J. The obtained results were very similar, so we will only discuss the results corresponding to $J(s) = 0.50^{-1}I(0.25 \le s \le 0.75)$. We have performed the test over a wide range of bandwidths (from 0.15 to 0.35) and we have calculated the p-values based on 1000 bootstrap replications. The threshold T was taken to be the quantile of order $\hat{F}_{\varepsilon 0}^{-1}(\hat{F}_{\varepsilon 0}(+\infty) - 0.10)$. The Kolmogorov-Smirnov type statistic T_{KS} produced p-values between 0.29 and 0.99. On the other hand, the Cramér-von Mises type statistic T_{CM} gave p-values between 0.27 and 0.95. The hypothesis of linearity can then be clearly accepted.

Heuchenne and Van Keilegom (2005) also gave bootstrap confidence intervals for the parameters of the linear regression. The interval corresponding to the slope of the regression line contains zero. Hence it is reasonable to test for a constant model instead of the complete linear model. We have applied the goodness-of-fit test to check the constant model and the obtained p-values were between 0.12 and 0.73 for T_{KS} and between 0.08 and 0.80 for T_{CM}. It seems that the constant model can also be accepted in this example.

All these results are summarized in Figure 2.

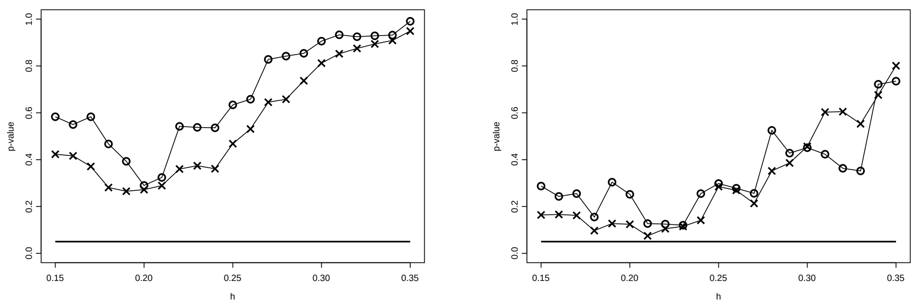

Figure 2: *Graphs of the p-values as function of the bandwidth h when testing for a linear model (left) and for a constant model (right) with the test statistics T_{KS} (line with circles) and T_{CM} (line with crosses). The score function is $J(s) = 0.50^{-1} I(0.25 \leq s \leq 0.75)$. The solid horizontal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

Appendix: Proofs

Proof of Theorem 1. The direct implication is trivial. On the other hand, assume that there exists a θ_0 such that $F_{\varepsilon 0}(y) = F_\varepsilon(y)$. We can write

$$P\left(\frac{Y - m_{\theta_0}(X)}{\sigma(X)} \leq y\right) = P\left(\frac{Y - m(X)}{\sigma(X)} + \frac{m(X) - m_{\theta_0}(X)}{\sigma(X)} \leq y\right),$$

or equivalently

$$P\left(\exp\left\{\frac{Y - m_{\theta_0}(X)}{\sigma(X)}\right\} \leq y\right) = P\left(\exp\left\{\frac{Y - m(X)}{\sigma(X)}\right\} \exp\left\{\frac{m(X) - m_{\theta_0}(X)}{\sigma(X)}\right\} \leq y\right),$$

for all y. The residuals $(Y - m(X))/\sigma(X)$ and the covariate X are independent, hence the moments of the distributions above verify the relation

$$E\left[\left(\exp\left\{\frac{Y - m_{\theta_0}(X)}{\sigma(X)}\right\}\right)^{2\nu}\right]$$
$$= E\left[\left(\exp\left\{\frac{Y - m(X)}{\sigma(X)}\right\}\right)^{2\nu}\right] E\left[\left(\exp\left\{\frac{m(X) - m_{\theta_0}(X)}{\sigma(X)}\right\}\right)^{2\nu}\right],$$

and hence

$$E\left[\left(\exp\left\{\frac{m(X) - m_{\theta_0}(X)}{\sigma(X)}\right\}\right)^{2\nu}\right] = 1,$$

for all $\nu \in \mathbb{N}$. Carleman's condition (see e.g. Feller, 1966) ensures that

$$P\left(\exp\left\{\frac{m(X) - m_{\theta_0}(X)}{\sigma(X)}\right\} = 1\right) = 1$$

or

$$P\left(\frac{m(X) - m_{\theta_0}(X)}{\sigma(X)} = 0\right) = 1.$$

This and the continuity of m implies the equality of $m(x)$ and $m_{\theta_0}(x)$ for all $x \in R_X$.

Proof of Theorem 2. Since we are working under the null hypothesis, there exists θ_0 such that $m = m_{\theta_0}$. From the proof of Proposition A.2 in Van Keilegom and Akritas (1999), we have that

$$\begin{aligned}\hat{H}_{e0}(y) - H_{e0}(y) =& \frac{1}{n}\sum_{i=1}^{n} I(E_{i0} \le y) - H_{e0}(y) \\ &+ \int h_{e0}(y|x)\frac{m_{\hat{\theta}}(x) - m_{\theta_0}(x)}{\sigma(x)} f(x)dx \\ &+ \int y h_{e0}(y|x)\frac{\hat{\sigma}(x) - \sigma(x)}{\sigma(x)} f(x)dx + o_P(n^{-1/2}),\end{aligned} \tag{19}$$

uniformly in $-\infty < y \le T$. The last term is $o_P(n^{-1/2})$ because of the uniform consistency of $m_{\hat{\theta}}$ and $\hat{\sigma}$. The consistency of $\hat{\sigma}$ is given by Proposition 4.5 in Van Keilegom and Akritas (1999), and the consistency of $m_{\hat{\theta}}$ can be obtained in a similar way.

Define the class of functions $M_\Theta(R_X) = \{x \to (m_\theta(x) - m(x))/\sigma(x), \theta \in \Theta\}$. Firstly, this class verifies $P((m_{\hat{\theta}}(x) - m(x))/\sigma(x) \in M_\Theta(R_X)) \to 1$ as $n \to \infty$ because of the consistency of the parameter estimate. Secondly, the bracketing number $N_{[]}(\lambda^2, M_\Theta(R_X), L_2(P)) = O(\lambda^{-2p})$ for any $\lambda > 0$ because of the compactness of the parametric space Θ. This bracketing number is smaller than for the class $C_1^{1+\delta}(R_X)$ defined in Lemma A.1 in Van Keilegom and Akritas (1999). Then we can replace the class $C_1^{1+\delta}(R_X)$ by the class $M_\Theta(R_X)$ in that Lemma and this justifies expression (19).

Using the expression (15), the first integral in (19) can be written as

$$\begin{aligned}&\int h_{e0}(y|x)\frac{m_{\hat{\theta}}(x) - m_{\theta_0}(x)}{\sigma(x)} f(x)dx \\ &= n^{-1}\sum_{i=1}^{n}\int h_e(y|u)\sigma^{-1}(u)\varphi_{\theta_0}(u, X_i, Z_i, \Delta_i) f(u)du + o_P(n^{-1/2}).\end{aligned}$$

From Proposition 4.9 in Van Keilegom and Akritas (1999) and a Taylor expansion, the second integral in (19) becomes

$$\int y h_{e0}(y|x)\frac{\hat{\sigma}(x) - \sigma(x)}{\sigma(x)} f(x)dx = -n^{-1}\sum_{i=1}^{n} y h_{e0}(y|X_i)\zeta(Z_i, \Delta_i|X_i) + o_P(n^{-1/2}).$$

Hence

$$\hat{H}_{e0}(y) - H_{e0}(y) = n^{-1}\sum_{i=1}^{n} I(E_{i0} \le y) - H_{e0}(y) + n^{-1}\sum_{i=1}^{n}\int h_{e0}(y|u)\sigma^{-1}(u)\varphi_{\theta_0}(u, X_i, Z_i, \Delta_i)f(u)du - n^{-1}\sum_{i=1}^{n} y h_{e0}(y|X_i)\zeta(Z_i, \Delta_i|X_i) + o_P(n^{-1/2}), \tag{20}$$

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

Similarly it can be proved that

$$\hat{H}_{e10}(y) - H_{e10}(y) = n^{-1}\sum_{i=1}^{n} I(E_{i0} \le y, \Delta_i = 1) - H_{e10}(y) + n^{-1}\sum_{i=1}^{n}\int h_{e10}(y|u)\sigma^{-1}(u)\varphi_{\theta_0}(u, X_i, Z_i, \Delta_i)f(u)du - n^{-1}\sum_{i=1}^{n} y h_{e10}(y|X_i)\zeta(Z_i, \Delta_i|X_i) + o_P(n^{-1/2}), \tag{21}$$

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

Proposition A.2 in Van Keilegom and Akritas (1999) ensures that

$$\hat{H}_{e}(y) - H_{e}(y) = n^{-1}\sum_{i=1}^{n} I(E_{i} \le y) - H_{e}(y) - n^{-1}\sum_{i=1}^{n} h_e(y|X_i)\eta(Z_i, \Delta_i|X_i) - n^{-1}\sum_{i=1}^{n} y h_{e}(y|X_i)\zeta(Z_i, \Delta_i|X_i) + o_P(n^{-1/2}), \tag{22}$$

and

$$\hat{H}_{e1}(y) - H_{e1}(y) = n^{-1}\sum_{i=1}^{n} I(E_{i} \le y, \Delta_i = 1) - H_{e1}(y) - n^{-1}\sum_{i=1}^{n} h_{e1}(y|X_i)\eta(Z_i, \Delta_i|X_i) - n^{-1}\sum_{i=1}^{n} y h_{e1}(y|X_i)\zeta(Z_i, \Delta_i|X_i) + o_P(n^{-1/2}), \tag{23}$$

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

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

$$\begin{aligned}
&\hat{F}_{\varepsilon 0}(y) - F_{\varepsilon 0}(y) \qquad (24)\\
&= (1 - F_{\varepsilon 0}(y)) \left[\int_{-\infty}^{y} \frac{\hat{H}_{e0}(s) - H_{e0}(s)}{(1 - H_{e0}(s))^2} dH_{e10}(s) + \int_{-\infty}^{y} \frac{d(\hat{H}_{e10}(s) - H_{e10}(s))}{1 - H_{e0}(s)} \right]\\
&+ o_P(n^{-1/2})
\end{aligned}$$

and

$$\begin{aligned}
&\hat{F}_{\varepsilon}(y) - F_{\varepsilon}(y) \qquad (25)\\
&= (1 - F_{\varepsilon}(y)) \left[\int_{-\infty}^{y} \frac{\hat{H}_{e}(s) - H_{e}(s)}{(1 - H_{e}(s))^2} dH_{e1}(s) + \int_{-\infty}^{y} \frac{d(\hat{H}_{e1}(s) - H_{e1}(s))}{1 - H_{e}(s)} \right]\\
&+ o_P(n^{-1/2}).
\end{aligned}$$

Clearly under H_0, it holds that $E_{i0} = E_i$, $F_{\varepsilon 0} = F_\varepsilon$, $H_{e0} = H_e$, $H_{e10} = H_{e1}$, $h_{e0} = h_e$ and $h_{e10} = h_{e1}$. By writing $\hat{F}_{\varepsilon 0}(y) - \hat{F}_\varepsilon(y) = (\hat{F}_{\varepsilon 0}(y) - F_\varepsilon(y)) - (\hat{F}_\varepsilon(y) - F_\varepsilon(y))$ and substituting (20), (21) in (24) and (22), (23) in (25) the representation given in the statement of the Theorem follows immediately.

Proof of Theorem 3. Following the notation of van der Vaart and Wellner (1996), if we define the class of functions

$$\mathcal{F} = \left\{(x, z, \delta) \longrightarrow (1 - F_\varepsilon(y))\psi_{\theta_0}(x, z, \delta, y), -\infty < y \le T\right\},$$

it is clear that the asymptotic behavior of our process of interest $\hat{W}(y)$ is determined by the asymptotic behavior of the $\mathcal{F}$-indexed process.

We will use the decomposition $\mathcal{F} = \sum_{k=1}^{p+1} \mathcal{F}_k^1 \mathcal{F}_k^2$ (for any classes of functions $\mathcal{G}_1$ and $\mathcal{G}_2$, we denote, in general, $\mathcal{G}_1 + \mathcal{G}_2 = \{g_1 + g_2; g_1 \in \mathcal{G}_1, g_2 \in \mathcal{G}_2\}$ and $\mathcal{G}_1\mathcal{G}_2 = \{g_1 g_2; g_1 \in \mathcal{G}_1, g_2 \in \mathcal{G}_2\}$), where, for $k = 1, \ldots, p$,

$$\mathcal{F}_k^1 = \left\{ (x, z, \delta) \longrightarrow (1 - F_\varepsilon(y)) \int \sigma^{-1}(u) \left(\left. \frac{\partial m_\theta(u)}{\partial \theta_k} \right|_{\theta = \theta_0} \right) f(u)\gamma_0(y|u)du, -\infty < y \le T \right\},$$

$$\mathcal{F}_k^2 = \{(x, z, \delta) \longrightarrow \omega_k(x, z, \delta), -\infty < y \le T\}$$

and

$$\mathcal{F}_{p+1}^1 = \{(u, z, \delta) \longrightarrow (1 - F_\varepsilon(y))\gamma_0(y|x), -\infty < y \le T\},$$

$$\mathcal{F}_{p+1}^2 = \{(u, z, \delta) \longrightarrow \eta(z, \delta|x), -\infty < y \le T\}.$$

For $k = 1, \ldots, p+1$, the class $\mathcal{F}_k^1$ consists of uniformly bounded functions of y. If M is a bound for the absolute value of those functions then their bracketing number is $N_{[]}(\lambda, \mathcal{F}_k^1, L_2(P)) = O(\exp(K\lambda^{-1}))$ for $\lambda < 2M$ and some $K > 0$, and $N_{[]}(\lambda, \mathcal{F}_k^1, L_2(P)) = 1$ for $\lambda > 2M$, where P is the measure of probability corresponding to the joint distribution of (X, Z, Δ) and $L_2(P)$ is the L_2-norm. Since the class $\mathcal{F}_k^2$ consists of only one function, the bracketing number of the product class $\mathcal{F}_k^1\mathcal{F}_k^2$ is the same as the bracketing number of the class $\mathcal{F}_k^1$.

Theorem 2.10.6 in van der Vaart and Wellner (1996) can be applied here to obtain

$$N_{[]}(\lambda, \mathcal{F}, L_2(P)) \leq \prod_{k=1}^{p+1} N_{[]}(\lambda, \mathcal{F}_k^1, L_2(P)),$$

and hence

$$\int_0^\infty \sqrt{\log N_{[]}(\lambda, \mathcal{F}, L_2(P))}d\lambda \leq \sum_{k=1}^{p+1} \int_0^{2M} \sqrt{\log N_{[]}(\lambda, \mathcal{F}_k^1, L_2(P))}d\lambda < \infty.$$

This proves that $\mathcal{F}$ is Donsker by Theorem 2.5.6 in van der Vaart and Wellner (1996). The weak convergence of the process $\hat{W}(y)$ now follows from pages 81 and 82 of the aforementioned book.

Proof of Corollary 4. The convergence of T_{KS} follows directly from the weak convergence of the process $\hat{W}(y)$ and the continuous mapping theorem.

The convergence of T_{CM} requires some more detail. If we apply the Skorohod construction (see Serfling, 1980) to the processes $\hat{W}(y)$ and $n^{1/2}(\hat{F}_{\varepsilon 0}(y) - F_\varepsilon(y))$ we obtain

$$\sup_{-\infty < y \leq T} |\hat{W}(y) - W(y)| \to_{a.s.} 0, \tag{26}$$

and

$$\sup_{-\infty < y \leq T} |\hat{F}_{\varepsilon 0}(y) - F_\varepsilon(y)| \to_{a.s.} 0. \tag{27}$$

Write

$$\begin{aligned} &\left| \int_{-\infty}^T \hat{W}^2(y) d\hat{F}_{\varepsilon 0}(y) - \int_{-\infty}^T W^2(y) dF_\varepsilon(y) \right| \\ &\leq \left| \int_{-\infty}^T (\hat{W}^2(y) - W^2(y)) d\hat{F}_{\varepsilon 0}(y) \right| + \left| \int_{-\infty}^T W^2(y) d(\hat{F}_{\varepsilon 0}(y) - F_\varepsilon(y)) \right|. \end{aligned}$$

We will show that the expression above is negligible, and this suffices to obtain the convergence of the Cramér-von Mises type statistic.

The first term of the right hand side of the above inequality is $o(1)$ a.s. due to (26). The limit process $W(y)$ has bounded and continuous trajectories almost surely. By taking into account (27) and applying the Helly-Bray Theorem (see p. 97 in Rao, 1965) to each of the trajectories of $W(y)$, we obtain

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

References

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

Akritas, M. G. (1996). On the use of nonparametric regression techniques for fitting parametric regression models. *Biometrics* **52**, 1342-1362.

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. and Cadarso-Suárez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *J. Nonparametr. Statist.* **4**, 65-78.

Heuchenne, C. and Van Keilegom, I. (2005). Polynomial regression with censored data based on preliminary nonparametric estimation (conditionally accepted by *Ann. Inst. Statist. Math.*).

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

Kardaun, O. (1983). Statistical survival analysis of male larynx-cancer patients - a case study. *Statist. Neerlandica* **37**, 103-125.

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.

Stute, W. (1999). Nonlinear censored regression. *Statist. Sinica* **9**, 1089-1102.

Stute, W.; González-Manteiga, W. and Sánchez-Sellero, C. (2000). Nonparametric model checks in censored regression. *Comm. Statist. Theory Methods* **29**, 1611-1629.

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

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

Van Keilegom, I., González-Manteiga, W. and Sánchez-Sellero, C. (2005). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution (submitted).

Van Keilegom, I. and Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Statist. Math.* **49**, 467-491.