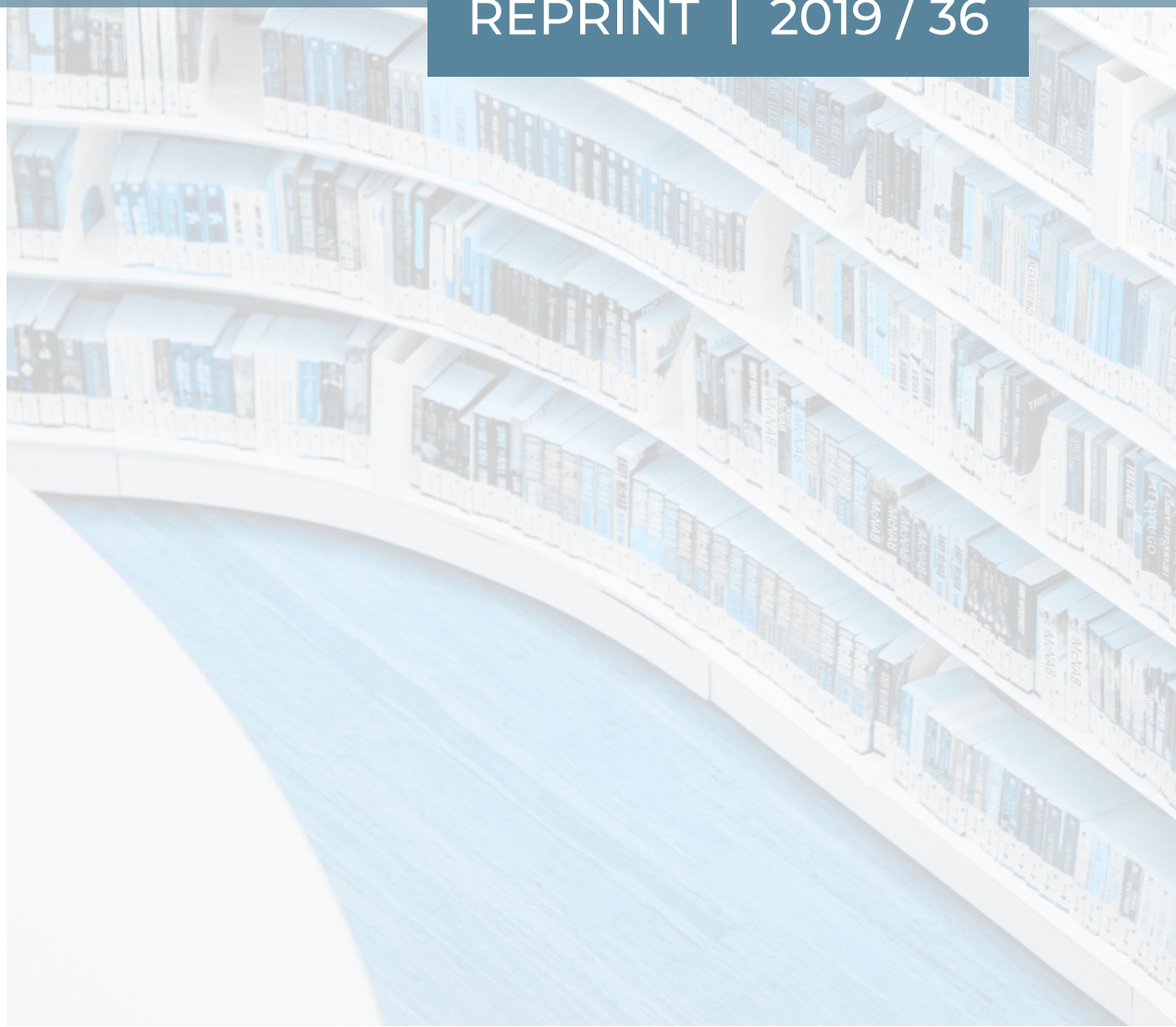
GOODNESS-OF-FIT TESTS FOR CENSORED REGRESSION BASED ON ARTIFICIAL DATA POINTS

González Manteiga, Wenceslao;
Heuchenne, Cédric;
Sánchez Sellero, César and Alessandro Beretta

REPRINT | 2019 / 36



UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

Goodness-of-fit tests for censored regression based on artificial data points

Wenceslao González Manteiga
Departamento de Estatística e I.O.
Universidad de Santiago de Compostela

Cédric Heuchenne
HEC-Université de Liège
ISBA, Université catholique de Louvain

Cesar Sánchez Sellero
Departamento de Estatística e I.O.
Universidad de Santiago de Compostela

Alessandro Beretta
HEC-Université de Liège

Abstract

Suppose we have a location-scale regression model where the location is the conditional mean and the scale is the conditional standard deviation; the response is possibly right-censored, the covariate is fully observed and the error is independent of the covariate. We propose new goodness-of-fit testing procedures for the conditional mean and variance based on an integrated regression function technique which uses artificial data points. We obtain the weak convergence of the resulting processes and study their finite sample behaviour via simulations. Finally, we analyze a data set about unemployment in Galicia.

KEY WORDS: Bootstrap; Goodness-of-fit tests; Kernel method; Least squares estimation; Nonparametric regression; Right censoring; Survival analysis.

1 Introduction

In this paper, we consider the following heteroscedastic regression model

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1.1}$$

where ε is independent of X (one-dimensional), $m(X) = E[Y|X]$ and $\sigma^2(X) = Var[Y|X]$. Suppose also that Y is subject to random right censoring, i.e. instead of observing Y, we only observe (Z, Δ), where $Z = \min(Y, C)$, $\Delta = I(Y \leq C)$ and the random variable C represents the censoring time, which is independent of Y, conditionally on X. Let $(Y_i, C_i, X_i, Z_i, \Delta_i)$ $(i = 1, \ldots, n)$ be n independent copies of (Y, C, X, Z, Δ).

The aim of this paper is to test the hypothesis

$$H_0 : \Psi \in \mathcal{M} \text{ versus } H_1 : \Psi \notin \mathcal{M}, \tag{1.2}$$

where $\mathcal{M} = \{\Psi_\vartheta : \vartheta \in \Theta\}$ is a class of parametric functions, $\Psi(\cdot)$ is either $m(\cdot)$ or $\sigma^2(\cdot)$ and $\Theta \subset I\!R^D$.

The approach used in this paper was introduced by Stute (1997) and is based on an estimator of the integrated function $\Psi(\cdot)$,

$$I(x) = \int_{-\infty}^{x} \Psi(z) dF_X(z),$$

where $F_X(x) = P(X \leq x)$. Following the lines of Stute (1997), the corresponding integrated process is given by

$$IP(x) = n^{-1/2} \sum_{i=1}^{n} \left(\psi(X_i, Y_i) - \Psi\left(X_i\right)\right) I\left(X_i \leq x\right), \tag{1.3}$$

using the fact that $I(x) = E[1_{\{X \leq x\}} \psi(X, Y)]$, where $E[\psi(X, Y)|X] = \Psi(X)$. Therefore, $\psi(X, Y) = Y$ or $(Y - m(X))^2$ and may depend on a vector of parameters according to the required test. When censored data are present, extensions of methods proposed by Heuchenne and Laurent (2017) as well as Heuchenne and Van Keilegom (2007, 2008a) are used to estimate the parameters of $\Psi(\cdot)$ (possibly $\psi(\cdot, \cdot)$) and replace censored $\psi(\cdot, \cdot)$ by artificial versions which can be considered as uncensored.

Although a number of goodness-of-fit tests exists for the regression function with censored data, few results are obtained for the conditional variance and especially for a function to test which is nonlinear instead of polynomial. Dette and Heuchenne (2012) proposed a goodness-of-fit test for any scale function but not adapted to the estimation of the conditional variance (a parametric scale function is estimated and tested but its validation does not imply the validation of the corresponding standard deviation). Stute, González Manteiga and Sánchez Sellero (2000) developped a goodness-of-fit test for censored nonlinear regression but it suffers from restrictive assumptions. This is due to the

use of the bivariate Kaplan-Meier estimator of Stute (1993). It assumes that (1) Y and C are independent (unconditionally on X) and that (2) $P(Y \leq C|X,Y) = P(Y \leq C|Y)$, which is satisfied when e.g. C is independent of X. Both assumptions are often violated in practice.

The paper is organized as follows. In the next section, the testing procedure is described in detail. Section 3 summarizes the main asymptotic results, including the weak convergence of the proposed process (the extension of $IP(x)$ to censored data) to a Gaussian process. In Section 4, we present the results of a simulation study, in which the new procedure is compared with the method of Stute, González Manteiga and Sánchez Sellero (2000). Section 5 applies the proposed techniques to a study of unemployment in Galicia whereas the online Supplementary Material of the paper contains the assumptions, functions and proofs needed to obtain the main results of Section 3.

2 Notations and description of the method

The idea of the proposed method consists of first estimating the unknown functions $\psi(\cdot,\cdot)$ due to censored observations, and second comparing those so-obtained artificial functions with a parametric estimation of $\Psi(\cdot)$ via the classical process (1.3). Define

$$\psi^{k*}(X,Z,\Delta) = \psi^k(X,Y)\Delta + E[\psi^k(X,Y)|Y > C, X](1-\Delta),\ k = 0,1,2,$$

and note that $E(\psi^k(X,Y)|X) = E(\psi^{k*}(X,Z,\Delta)|X) = \Psi_{\theta_k}(X)$, $k = 0,1,2$, under the null hypothesis ($\Psi_{\theta_k}(X) = \Psi(X)$ if H_0 is true). The index k indicates to which test corresponds the new data point $\psi^{k*}(X,Z,\Delta)$. Indeed,

1. for $k = 0$, $\psi^0(X,Y) = Y$ corresponding to a goodness-of-fit test for the conditional mean m,

2. for $k = 1$, $\psi^1(X,Y) = (Y - m_{\theta_0}(X))^2$ corresponding to a goodness-of-fit test for the conditional variance σ^2, assuming that the conditional mean has a known parametric form (and the true vector of parameters is defined by θ_0),

3. for $k = 2$, $\psi^2(X,Y) = (Y - m(X))^2$ corresponding to a goodness-of-fit test for the conditional variance σ^2, not assuming any parametric form for the conditional mean m.

Hence, we can work in the sequel with the variable $\psi^{k*}(X,Z,\Delta)$ instead of $\psi^k(X,Y)$. In order to estimate $\psi^{k*}(X,Z,\Delta)$ for a censored observation, we first need to introduce a number of notations.

Let $m^0(\cdot)$ be any location function and $\sigma^0(\cdot)$ be any scale function, meaning that $m^0(x) = Q(F(\cdot|x))$ and $\sigma^0(x) = S(F(\cdot|x))$ for some functionals Q and S that satisfy $Q(F_{aY+b}(\cdot|x)) = aQ(F_Y(\cdot|x)) + b$ and $S(F_{aY+b}(\cdot|x)) = aS(F_Y(\cdot|x))$, for all $a \geq 0$ and $b \in I\!R$ (here $F_{aY+b}(\cdot|x)$ denotes the conditional distribution of $aY + b$ given $X = x$). Let $\varepsilon^0 = (Y - m^0(X))/\sigma^0(X)$. Then, it can be easily seen that if model (1.1) holds (i.e. ε is independent of X), then ε^0 is also independent of X.

We have for $F_\varepsilon^0(y) = P(\varepsilon^0 \leq y)$ and $E_i^0 = (Z_i - m^0(X_i))/\sigma^0(X_i)$, $i = 1, \ldots, n$,

$$\psi^{k*}(X_i, Z_i, \Delta_i) = \psi^k(X_i, Y_i)\Delta_i + \frac{\int_{E_i^0}^{\infty} \psi^k(X_i, m^0(X_i) + \sigma^0(X_i)y)\, dF_\varepsilon^0(y)}{1 - F_\varepsilon^0(E_i^0)}(1 - \Delta_i),$$

$k = 0, 1, 2$, for the following choices of m^0 and σ^0 :

$$m^0(x) = \int_0^1 F^{-1}(s|x)J(s)\, ds, \quad \sigma^{02}(x) = \int_0^1 F^{-1}(s|x)^2 J(s)\, ds - m^{02}(x), \tag{2.1}$$

where $F(y|x) = P(Y \leq y|x)$, $F^{-1}(s|x) = \inf\{y; F(y|x) \geq s\}$ is the quantile function of Y given x and $J(s)$ is a given score function satisfying $\int_0^1 J(s)\, ds = 1$. When $J(s)$ is chosen appropriately (namely put to zero in the right tail, there where the quantile function cannot be estimated in a consistent way due to the right censoring), $m^0(x)$ and $\sigma^0(x)$ can be estimated consistently. The distribution $F(y|x)$ in (2.1) is replaced by the Beran (1981) estimator, defined by (in the case of no ties) :

$$\hat{F}(y|x) = 1 - \prod_{Z_i \leq y, \Delta_i = 1} \Big\{1 - \frac{W_i(x, a_n)}{\sum_{j=1}^n I(Z_j \geq Z_i) W_j(x, a_n)}\Big\}, \tag{2.2}$$

where

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

K is a kernel function and $\{a_n\}$ a bandwidth sequence. Therefore,

$$\hat{m}^0(x) = \int_0^1 \hat{F}^{-1}(s|x)J(s)\, ds \text{ and } \hat{\sigma}^{02}(x) = \int_0^1 \hat{F}^{-1}(s|x)^2 J(s)\, ds - \hat{m}^{02}(x) \tag{2.3}$$

estimate $m^0(x)$ and $\sigma^{02}(x)$. Next,

$$\hat{F}_\varepsilon^0(y) = 1 - \prod_{\hat{E}_{(i)}^0 \leq y, \Delta_{(i)} = 1} \left(1 - \frac{1}{n - i + 1}\right), \tag{2.4}$$

denotes the Kaplan-Meier (1958)-type estimator of F_ε^0 (in the case of no ties), where $\hat{E}_i^0 = (Z_i - \hat{m}^0(X_i))/\hat{\sigma}^0(X_i)$, $\hat{E}_{(i)}^0$ is the i-th order statistic of $\hat{E}_1^0, \ldots, \hat{E}_n^0$ and $\Delta_{(i)}$ is the corresponding censoring indicator. This estimator has been studied in detail by Van

Keilegom and Akritas (1999). This leads to the following estimators for $\psi^k(X_i, Y_i)(k = 0, 1)$:

$$\tilde{\psi}_T^{0*}(X_i, Z_i, \Delta_i) = \hat{Y}_{Ti}^* = Y_i\Delta_i + \Big\{\hat{m}^0(X_i) + \frac{\hat{\sigma}^0(X_i)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^0 \wedge T)} \int_{\hat{E}_i^0 \wedge T}^T y\, d\hat{F}_\varepsilon^0(y)\Big\}(1 - \Delta_i), \quad (2.5)$$

$$\tilde{\psi}_T^{1*}(X_i, Z_i, \Delta_i) = (Y_i - \hat{m_{\theta_0}}(X_i))_T^{2*} = (Y_i - m_{\theta_0}(X_i))^2\Delta_i + \Big\{(\hat{m}^0(X_i) - m_{\theta_0}(X_i))^2 \quad (2.6)$$
$$+ \frac{\int_{\hat{E}_i^0 \wedge T}^T \hat{\sigma}^0(X_i) y[\hat{\sigma}^0(X_i)y + 2(\hat{m}^0(X_i) - m_{\theta_0}(X_i))]\, d\hat{F}_\varepsilon^0(y)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^0 \wedge T)}\Big\}(1 - \Delta_i),$$

where $m_{\theta_0}(\cdot)$ in (2.6) is replaced by $m(\cdot)$ to obtain the expression of $\hat{\psi}_T^{2*}(X_i, Z_i, \Delta_i)$, $T < \tau_{H_\varepsilon^0}$, $H_\varepsilon^0(y) = P(E^0 \leq y)$ for $E^0 = (Z - m^0(X))/\sigma^0(X)$, and $\tau_F = \inf\{y : F(y) = 1\}$ for any distribution F. Truncations by the constant T in the above integrals and denominators are due to right censoring (however, when $\tau_{F_\varepsilon^0} \leq \tau_{G_\varepsilon^0}$, for $G_\varepsilon^0(y) = P(C^0 \leq y)$ and $C^0 = (C - m^0(X))/\sigma^0(X)$, T can be chosen arbitrarily close to $\tau_{F_\varepsilon^0}$). In $\tilde{\psi}_T^{1*}(X_i, Z_i, \Delta_i)$, θ_0 can be replaced by its estimator obtained by the method of Heuchenne and Van Keilegom (2008a) to obtain $\hat{\psi}_T^{1*}(X_i, Z_i, \Delta_i)$, while similarly $\tilde{\psi}_T^{2*}(X_i, Z_i, \Delta_i)$ can be defined with $m(\cdot)$ which can be replaced by a nonparametric estimator, say $\tilde{m}_T(X_i)$, developed, by example, in Heuchenne and Van Keilegom (2008b, 2010). For the sake of simplicity, we therefore use $\hat{\psi}_T^{k*}(X_i, Z_i, \Delta_i)$ to denote fully estimated quantities (with estimated θ_0 or $m(\cdot)$) for $k = 0, 1, 2$.

Finally, the functions $\hat{\psi}_T^{k*}(X_i, Z_i, \Delta_i)$ (resp $\Psi_{\vartheta_{nk}^T}(X_i)$), $k = 0, 1, 2$, replace $\psi(X_i, Y_i)$ (resp $\Psi(X_i)$), $i = 1, \ldots, n$, in (1.3) for which we define

$$\vartheta_{nk}^T := \text{argmin}_{\vartheta_k \in \Theta_k} \sum_{i=1}^n [\hat{\psi}_T^{k*}(X_i, Z_i, \Delta_i) - \Psi_{\vartheta_k}(X_i)]^2, \quad (2.7)$$

as estimators for the parameters describing $\mathcal{M}_k = \{\Psi_{\vartheta_k} : \vartheta_k \in \Theta_k\}$ (Θ_k is a compact subset of $\mathbb{R}^{D_k}$, D_k is a positive integer and $\Psi_\vartheta(\cdot)$ is either $m_\vartheta(\cdot)$ or $\sigma_\vartheta^2(\cdot)$, the tested parametric variance), the class of parametric functions corresponding to the goodness-of-fit test k, $k = 0, 1, 2$. Note that the estimation procedure proposed in (2.7) when $k = 2$ and $m(\cdot)$ is estimated by the method proposed in Heuchenne and Van Keilegom (2010) has been studied in Heuchenne and Laurent (2017). In order to focus on the primary issues, we assume the existence of a well-defined minimizer for (2.7). Solutions for those problems can be obtained using an (iterative) procedure for nonlinear minimization problems, like e.g. a Newton-Raphson procedure. Since $\hat{\psi}_T^{1*}(X_i, Z_i, \Delta_i) = \hat{\psi}_T^{1*}(X_i, Z_i, \Delta_i, \vartheta_{n0}^T) = (Y_i - \hat{m_{\vartheta_{n0}^T}}(X_i))_T^{2*}$, $i = 1, \ldots, n$, we will use in the sequel $\hat{\psi}_T^{k*}(X_i, Z_i, \Delta_i, \vartheta_{n0}^T) = \hat{\psi}_{Ti}^{k*}(\vartheta_{n0}^T)$, $k = 0, 1, 2$ (especially to develop the proofs). Therefore, we consider the following expression

$$ICP_k(x) = n^{-1/2} \sum_{i=1}^n (\hat{\psi}_{Ti}^{k*}(\vartheta_{n0}^T) - \Psi_{\vartheta_{nk}^T}(X_i)) I(X_i \leq x),\ k = 0, 1, 2. \quad (2.8)$$

More precisely, we propose a Kolmogorov-Smirnov type statistic

$$T_{KSI,k} = \sup_{x \in R_X} |ICP_k(x)|$$

and a Cramer-von Mises type statistic

$$T_{CMI,k} = \int_{R_X} (ICP_k(x))^2 d\hat{F}_X(x),$$

where $\hat{F}_X(\cdot)$ is the empirical distribution of the X-values. The null hypothesis (1.2) is rejected for large values of the test statistics.

Remark 2.1 (Test with parametric variance) In the case $k = 0$, we test a parametric form for the conditional mean without assuming any parametric form for the conditional variance. We could consider such a parametric form introducing it at the denominator of each term of (2.8) for $k = 0$. This would be equivalent to define $\psi(X, Y) = Y/\sigma_\theta(X)$ for some θ. An estimator for the vector of parameters θ could be obtained by example using (2.7) for $k = 2$ and the analytic form of the corresponding test statistics would be straightforward.

3 Asymptotic results

We start by developing an asymptotic representation for the expression (2.8) under the null hypothesis and where the remaining term is $o_P(n^{-1/2})$ uniformly in x. This will allow us to obtain the weak convergence of the process $ICP_k(x)$, $k = 0, 1, 2$. Finally, the asymptotic distributions of the proposed test statistics are obtained. The assumptions, proofs and involved functions in the results below are given in the Supplementary Material of the paper. The index k in the results below may be confusing since it is an additional notation in already complicated expressions; the reader may omit it in a first step and focus on the general issue: the idea of Theorem 3.1 hereunder is valid for a range of functions among which the ones proposed at the beginning of Section 2. We could apply it to other cases: for example, using another definition for $m(\cdot)$ and/or $\sigma(\cdot)$, like, e.g., a truncated mean or a robust measure of scale, can lead to similar artificial functions (with some specificities).

Theorem 3.1 *(i) Assume (A1)-(A10) (in the Supplementary Material of the paper). Then, under the null hypothesis H_0,*

$$n^{-1} \sum_{i=1}^{n} (\hat{\psi}_{Ti}^{k*}(\vartheta_{n0}^T) - \Psi_{\vartheta_{nk}^T}(X_i)) I(X_i \leq x) = n^{-1} \sum_{i=1}^{n} \chi_k^x(X_i, Z_i, \Delta_i, \theta_0^T, \theta_k^T) + R_n(x),$$

$k = 0, 1, 2$, *where* $\sup\{|R_n(x)|; x \in R_X\} = o_P(n^{-1/2})$, R_X *is the support of* X, $\chi_k^x(X_i, Z_i, \Delta_i, \theta_0^T, \theta_k^T)$ *is defined in the Supplementary Material of the paper and* θ_k^T *in Remark 3.1 hereunder.*

(ii) Next, under the same assumptions, if $\psi_T^{k*}(\theta_0^T)$, $k = 0, 1, 2$, *follows a model such that* $E[\psi_T^{k*}(\theta_0^T)|X] = \Psi_{\theta_k^T}(X)$, *then, under the null hypothesis* H_0, *the process* $ICP_k(x) = n^{-1/2}\sum_{i=1}^n(\hat{\psi}_{Ti}^{k*}(\vartheta_{n0}^T) - \Psi_{\vartheta_{nk}^T}(X_i))I(X_i \leq x)$, $x \in R_X$, *converges weakly to a centered gaussian process* $W_k(x)$ *with covariance function*

$$Cov(W_k(x), W_k(x')) = E[\chi_k^x(X, Z, \Delta, \theta_0^T, \theta_k^T)\chi_k^{x'}(X, Z, \Delta, \theta_0^T, \theta_k^T)].$$

(iii) Finally, under the same assumptions and the null hypothesis H_0,

$$T_{KSI,k} \xrightarrow{d} \sup_{x \in R_X} |W_k(x)|,$$

$$T_{CMI,k} \xrightarrow{d} \int_{R_X} W_k^2(x) dF_X(x),$$

k=0,1,2.

Remark 3.1 (Non zero mean asymptotic representation) As it is clear from the definitions of $\hat{\psi}_{Ti}^{k*}(\vartheta_{n0}^T)$, $\Psi_{\vartheta_{nk}^T}(X_i)$ and ϑ_{nk}^T for $k = 0, 1, 2$, expression (2.8) is actually estimating

$$n^{-1/2}\sum_{i=1}^n(\psi_{Ti}^{k*}(\theta_0^T) - \Psi_{\theta_k^T}(X_i))I(X_i \leq x), \ k = 0, 1, 2, \tag{3.1}$$

where

$$\psi_{Ti}^{0*} = Y_{Ti}^* = Y_i\Delta_i + \Big\{m^0(X_i) + \frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^0 \wedge T)}\int_{E_i^0 \wedge T}^T y\, dF_\varepsilon^0(y)\Big\}(1 - \Delta_i), \tag{3.2}$$

$$\begin{aligned}\psi_{Ti}^{1*}(\theta_0^T) = (Y_i - m_{\theta_0^T}(X_i))_T^{2*} = (Y_i - m_{\theta_0^T}(X_i))^2\Delta_i + \Big\{&(m^0(X_i) - m_{\theta_0^T}(X_i))^2 \\ &+ \frac{\int_{E_i^0 \wedge T}^T \sigma^0(X_i)y[\sigma^0(X_i)y + 2(m^0(X_i) - m_{\theta_0^T}(X_i))]\, dF_\varepsilon^0(y)}{1 - F_\varepsilon^0(E_i^0 \wedge T)}\Big\}(1 - \Delta_i),\end{aligned} \tag{3.3}$$

where ψ_{Ti}^{2*} is obtained by replacing $m_{\theta_0^T}(X_i)$ in $\psi_{Ti}^{1*}(\theta_0^T)$ by $m_T(X_i)$, $i = 1, \ldots, n$ ($m_T(\cdot)$ is the stochastic limit of $\tilde{m}_T(\cdot)$ when $n \to \infty$), $\Psi_{\theta_0^T}(\cdot) = m_{\theta_0^T}(\cdot)$, $\Psi_{\theta_p^T}(\cdot) = \sigma_{\theta_p^T}^2(\cdot)$, $p = 1, 2$, and $\theta_k^T = (\theta_{k1}^T, \ldots, \theta_{kD_k}^T)$, $k = 0, 1, 2$, are the unique parameters which minimize

$$E[\{E(\psi_T^{k*}(\theta_0^T)|X) - \Psi_{\vartheta_k}(X)\}^2]$$

(see hypothesis (A10) in the Supplementary Material of the paper). A consequence of those limits is that the expression $n^{-1}\sum_{i=1}^{n}(\psi_{Ti}^{k*}(\theta_0^T)-\Psi_{\theta_k^T}(X_i))I(X_i\leq x)$ in the asymptotic representation of Theorem 3.1 has a mean different from zero. This is due to the use of the estimator (2.4) which is inconsistent in the right tails. That can lead to errors when testing parametric hypothesis. However, as also studied in Heuchenne and Van Keilegom (2010), pages 443-444, the inconsistent region of (2.4) is smaller than inconsistent regions of other distribution estimators for censored data (like, e.g., the Beran estimator). In addition, $\psi_T^{k*}(\theta_0^T)$, $\Psi_{\theta_k^T}(X)$, θ_k^T can be made arbitrarily close to $\psi^{k*}(X,Z,\Delta,\theta_0)$, $\Psi_{\theta_k}(X)$, θ_k, $k=0,1,2$, provided $\tau_{F_\varepsilon^0}\leq\tau_{G_\varepsilon^0}$. Under that assumption, $\psi_T^{k*}(\theta_0^T)-\Psi_{\theta_k^T}(X)$, $k=0,1,2$, can then be made arbitrarily close to a random variable with zero conditional mean (under H_0).

Remark 3.2 (Local alternative hypothesis) The behaviour of the process can also be studied under the alternative hypothesis. By example, in the conditional mean case, a local (Pitman) alternative of the type $H_{1n}: m(x)=m_{\theta_0}(x)+n^{-1/2}r(x)$ is considered in the sequel. In order to keep the proportion of censoring fixed for any value of n, we use in this context the following assumption on the censoring variable. There exists a random variable C_0, not depending on n, such that $P(C\leq y|X)=P(C_0+n^{-1/2}r(X)\leq y|X)$. Next, we define $Y_0=m_{\theta_0}(X)+\sigma(X)\varepsilon$, $Z_0=Y_0\wedge C_0$ and assume that $E[r^2(X)]<\infty$. We also replace $E[\{E(Y_T^*|X)-m_{\vartheta_0}(X)\}^2]$ of the assumption (A10) in the Supplementary Material of the paper by $E[\{E(Y_{0T}^*|X)-m_{\vartheta_0}(X)\}^2]$, where $Y_{0T}^*=Y_T^*-n^{-1/2}r(X)$. C_0, Y_0, Z_0 and Y_{0T}^* do not depend on n by construction. Theoretically, their use makes the main parts of the asymptotic representations under H_{1n} independent of n and equal to the asymptotic representations obtained under the null hypothesis.

More precisely,

$$\begin{aligned}n^{-1/2}\sum_{i=1}^{n}(\hat{Y}_{Ti}^*-m_{\vartheta_{n0}^T}(X_i))I(X_i\leq x) \;=\; & n^{-1/2}\sum_{i=1}^{n}(\hat{Y}_{Ti}^*-Y_{Ti}^*)I(X_i\leq x)\\ & +n^{-1/2}\sum_{i=1}^{n}(Y_{0Ti}^*+n^{-1/2}r(X_i)-m_{\theta_0^T}(X_i))I(X_i\leq x)\\ & +n^{-1/2}\sum_{i=1}^{n}(m_{\theta_0^T}(X_i)-m_{\vartheta_{n0}^T}(X_i))I(X_i\leq x). \qquad (3.4)\end{aligned}$$

The treatment of the first term on the right hand side of the above expression exactly follows the lines the proof of Theorem 3.1 (see Lemma 1.1 and the term Ω_{1n}^x in the Supplementary Material). The resulting asymptotic representation is equal to the one that would have been constructed with replications of Y_0, C_0 and Z_0 defined above. For

the third term on the right hand side of (3.4) and especially

$$n^{-1/2}\sum_{i=1}^{n}\left[\frac{\partial m_{\theta_0^T}(X_i)}{\partial \vartheta_0}\right]^t(\theta_0^T-\vartheta_{n0}^T)I(X_i\leq x),$$

where $[\cdot]^t$ denotes the transpose of a vector (or a matrix), we follow the lines of Theorems 3.1 and 3.2 in Heuchenne and Van Keilegom (2008a) to get

$$\begin{aligned}\theta_0^T-\vartheta_{n0}^T &= -\Omega^{-1}n^{-1}\sum_{i=1}^{n}(\hat{Y}_{Ti}^*-Y_{Ti}^*)\frac{\partial m_{\theta_0^T}(X_i)}{\partial \vartheta_0} \qquad (3.5)\\ &\quad -\Omega^{-1}n^{-1}\sum_{i=1}^{n}(Y_{0Ti}^*+n^{-1/2}r(X_i)-m_{\theta_0^T}(X_i))\frac{\partial m_{\theta_0^T}(X_i)}{\partial \vartheta_0}+o_P(n^{-1/2}),\end{aligned}$$

where the matrix $\Omega=(\Omega_{jk})$ $(j,k=1,\ldots,D_0)$ is defined by

$$\Omega_{jk}=E\left[\frac{\partial m_{\theta_0^T}(X)}{\partial \vartheta_{0j}}\frac{\partial m_{\theta_0^T}(X)}{\partial \vartheta_{0k}}-\{Y_{0T}^*-m_{\theta_0^T}(X)\}\frac{\partial^2 m_{\theta_0^T}(X)}{\partial \vartheta_{0j}\partial \vartheta_{0k}}\right]$$

and ϑ_{0j} is the j^{th} component of ϑ_0, $j=1,\ldots,D_0$. We treat the first term on the right hand side of (3.5) like the first term on the right hand side of (3.4), i.e., in the same way as in Theorem 3.2 of Heuchenne and Van Keilegom (2008a). That also leads to an asymptotic representation equal to the one that would have been obtained with replications of Y_0, C_0 and Z_0. Next, if we do not consider the parts involving $n^{-1/2}r(X_i)$ in the second terms on the right hand sides of (3.4) and (3.5), we finally obtain the asymptotic representation of Theorem 3.1 for replications of Y_0, C_0 and Z_0. The difference with Theorem 3.1(i) comes therefore from the terms involving $n^{-1/2}r(X_i)$ which, gathered, lead to the additional term

$$n^{-1/2}b(x)=n^{-1/2}\{E[r(X)I(X\leq x)]-\sum_{d=1}^{D_0}\Omega_d^{-1}E[r(X)\frac{\partial m_{\theta_0^T}(X)}{\partial \vartheta_0}]E[\frac{\partial m_{\theta_0^T}(X)}{\partial \vartheta_{0d}}I(X\leq x)]\},$$

where Ω_d^{-1} represents the d^{th} row of the matrix Ω^{-1}. As a consequence, we will have for the resulting statistics under H_{1n},

$$T_{KSI,0}\xrightarrow{d}\sup_{x\in R_X}|W_0(x)+b(x)|$$

and

$$T_{CMI,0}\xrightarrow{d}\int_{R_X}(W_0(x)+b(x))^2dF_X(x).$$

To better interpret $b(x)$ and the alternative hypotheses of the type $H_{1n}^c: m(x)=m_{\theta_0}(x)+r_n(x)$, where $r_n(x)$ tends to 0 when n tends to infinity, we end this remark with the two following situations.

1. if $r_n(x) = n^{-a}r(x)$ with $a > 0.5$, the asymptotic normality for each value of x in Theorem 3.1 is still valid and the power of the test tends theoretically to the level the test.

2. if $r_n(x) = n^{-a}r(x)$ with $a < 0.5$, the asymptotic normality for each value of x in Theorem 3.1 is not valid anymore; the asymptotic law for each value of x is still normal but there is a *location shift* defined by $n^{1/2-a}b(x)$. If $b(x) \neq 0$, i.e., $r(x)$ is chosen in such a way that the resulting $b(x) \neq 0$ on some subinterval of the support of X, the power of the test tends to 1 when n tends to infinity.

4 Practical implementation and simulations

In this section, we study the finite sample behavior of the different test statistics. We are interested in the behavior of the percentage of simulated samples for which the null hypothesis is rejected. The simulations are carried out for samples of size $n = 100$ and the results are obtained by using 10000 simulations. We develop simulations for the three proposed goodness-of-fit tests and the two corresponding statistics ($T_{KSI,k}$ and $T_{CMI,k}$, $k = 0, 1, 2$).

The problem of testing the goodness-of-fit of a parametric model for the conditional mean, when the response variable is subject to random right censoring, was also considered by Stute et al. (2000). They proposed the following process

$$WP(x)^{-1/2} \sum_{i=1}^{n} W_{in} \left(Z_i - m_{\vartheta_{n0}^S}(X_i) \right) I(X_i \leq x),$$

where W_{in} are the Kaplan-Meier weights attached to the censored sample (Z_i, Δ_i) $(i = 1, \ldots, n)$, $S \leq \tau_H$ (for $H(y)$, the distribution of the observable Z) and ϑ_{n0}^S is obtained by

$$\vartheta_{n0}^S = \min_{\vartheta_0 \in \Theta_0} \sum_{i=1}^{n} W_{in} \left(Z_i - m_{\vartheta_0}(X_i) \right)^2. \tag{4.1}$$

Again, Kolmogorov-Smirnov and Cramer-von Mises statistics can be obtained from the process WP,

$$T_{KSW} = \sup_{x \in R_X} |WP(x)|, \qquad T_{CMW} = \int (WP(x))^2 \, d\hat{F}_X(x).$$

Therefore, in the regression case, we compare those methods with the ones proposed in this paper.

First, we describe chosen characteristics of the proposed methods. For the score function J, we recommend the choice $J(s) = b^{-1}I(0 \leq s \leq b)$ $(0 \leq s \leq 1)$, where $b = \min_{1 \leq i \leq n} \hat{F}(+\infty|X_i)$. In this way, the region where the Beran estimators $\hat{F}(\cdot|X_1), \ldots,$

$\hat{F}(\cdot|X_n)$ are inconsistent is not used, and on the other hand, we exploit to a maximum the 'consistent' region. For $K(x)$, we work with the Epanechnikov kernel function $K(x) = (3/4)(1-x^2)I(|x| \leq 1)$. In order to improve the behavior near the boundaries of the covariate space, we use the reflection method to compute all kernel estimations. The point T can be chosen larger (or equal) than the last order statistic $\hat{E}^0_{(n)}$ of the estimated residuals $\hat{E}^0_i$, $i = 1, \ldots, n$. In this way, all the (unconditional) Kaplan-Meier jumps in (2.5) or (2.6) are considered. Next, for each method, equations (4.1) and (2.7) have an explicit solution in the considered models. Finally, the last order statistic on which each global Kaplan-Meier estimator is constructed may be censored. In this case, it is redefined as uncensored.

In order to develop simulations for the testing procedures, two tables (for Kolmogorov-Smirnov and Cramer-von Mises statistics) of critical values were constructed for each method and each parametric model supposed under H_0. So, for each value of the couple (parameter, bandwidth), test statistics were computed on 1000 samples of size 100 providing in this way estimations of their distributions under the null. Next, other samples were simulated. For each of them, the values of the statistics were compared with the corresponding critical values for the chosen bandwidth and for the estimated parameter. Each of the following tables contains the percentages of rejections obtained for different deviations from the null and different values of the bandwidth. The sample size is also 100 and each percentage of rejection is obtained from 10000 replicates.

For the goodness-of-fit test 1, the first simulated model is constructed in this way:

$$\begin{aligned} Y &= 5X + a(X) + \varepsilon, \\ C &= 5X + a(X) + 1 + \varepsilon^*, \end{aligned} \tag{4.2}$$

where X is uniform on the interval $[0, 1]$, ε and ε^* are standard normal, independent of X, ε is independent of ε^*, and $a(x)$ is a function that indicates the deviation from the null hypothesis which consists in the parametric model

$$H_0 : m(x) = \vartheta_0 x, \tag{4.3}$$

where $\vartheta_0 \in I\!R$ is an unknown parameter. It is easy to see that, under this model,

$$P(\Delta = 0|X) = \Phi(-1/\sqrt{2}),$$

which is independent of X, where Φ is the standard normal distribution function.

Table 1 gives the rejection percentages of null hypothesis (4.3) when the model (4.2) has different shapes of the deviation $a(x)$ and when different values of the bandwidth are used. Under the null hypothesis, $a(x) = 0$, the level of the test, 5%, is respected

by the four tests. Under two of the alternative models the four tests show a similar behaviour, with the tests based on $WP(x)$ being sligthly more powerful, while under the third alternative model, the tests based on $ICP_0(x)$ are much more powerful.

$a(x)$	a_n	T_{KSW}	T_{CMW}	T_{KSI}	T_{CMI}
0	0.20	5.00	5.12	4.76	4.78
0	0.25	5.14	5.23	4.70	4.67
0	0.30	5.07	5.13	4.62	4.44
x^2	0.20	24.23	24.82	20.03	19.07
x^2	0.25	24.90	25.35	20.02	19.13
x^2	0.30	24.34	24.72	19.21	18.14
$x * \exp(x)$	0.20	59.56	57.95	46.23	42.94
$x * \exp(x)$	0.25	57.85	56.74	46.05	42.22
$x * \exp(x)$	0.30	58.50	57.80	44.74	40.47
$\sin(2\pi x)$	0.20	51.33	47.46	90.12	85.35
$\sin(2\pi x)$	0.25	51.56	47.55	91.67	87.73
$\sin(2\pi x)$	0.30	52.64	48.14	93.05	89.86

Table 1: *Percentage of rejections of (4.3) for model (4.2) under the null hypothesis, $a(x) = 0$, and under three alternatives (nominal level 5%).*

Table 1 provides a number of tools for intuitive understanding of the new method. The most important one is the fitting of the the curve under the null on the samples generated under an alternative model. More precisely, we consider a theoretical distance between models (hereafter abbreviated by TDM): a curve measuring the distance between the true alternative curve and the curve under the null using the asymptotic values of the least squares estimators obtained under the true alternative model. For example, an integral (over a part or the entire support of X, R_X) of the absolute value of the difference between both curves could be used. According to the alternative model, TDM is relatively small (for the second and third models in Table 1, the line leaves its position under the null to fit approximately well the alternative models) or larger (for the fourth model in Table 1, the line is perturbed by the alternative samples but does not fit well the bumps of a sinus function). So, when TDM is large, alternatives should be more easily detected. According

to Heuchenne and Van Keilegom (2008a), Stute's method suffers from restrictive conditions leading to increases of biases and variances of resulting estimators. On the other side, constructing new data points (2.5) using model (1.1) enables to add information improving the fit of a curve to a data set. For each simulated sample, the obtained least squares estimators determine the critical value we use and therefore the corresponding distribution of the statistic. If TDM is small, we obtain slightly weaker rejected proportions for the new method with respect Stute's method because statistics constructed with the new method often correspond to acceptance regions of distributions (small TDM combined with well fitting method) while statistics obtained by Stute's method more often reach tails of distributions (variable least squares estimators determining statistics distributions for which a smaller proportion of generated samples corresponds). If TDM is larger, the above characteristics of Stute's method still appear while use of the model in the new method allows to obtain less variable, well fitted least squares estimators and therefore easier detection of alternatives. For the fourth model of Table 1, this effect is still more pronounced if the value of the bandwidth parameter increases (since that decreases the variances of least squares estimators). Note that this is not true if the increasing of a_n leads to samples too easily fittable by the curve under the null (see second and third models of Table 1).

For the goodness-of-fit tests 2 and 3, the simulated model is constructed in this way:

$$\begin{aligned} Y &= 1 + 2X + \sigma(X)\,\varepsilon, \\ C &= 1 + 2X + 1 + \sigma(X)\,\varepsilon^*, \end{aligned} \tag{4.4}$$

where X is uniform on the interval $[0,1]$, ε and ε^* are standard normal, independent of X, ε is independent of ε^*, and $\sigma^2(X) = Var[Y|X]$. Different functions $\sigma(x)$ are considered, one of them under the null and two under the alternative, where the null hypothesis consists of a constant conditional variance,

$$H_0 : \sigma(x) = \vartheta_0, \tag{4.5}$$

where $\vartheta_0 \in I\!R$ is an unknown parameter.

Table 2 gives the rejection percentages of null hypothesis (4.5) when the model (4.4) has different shapes of the conditional standard deviation $\sigma(x)$ and when different values of the bandwidth are used. The columns headed by the caption, $k = 1$, contain the rejection percentages corresponding to the goodness-of-fit test for the constant conditional variance, assuming that the conditional mean is linear, whereas the columns headed by the caption, $k = 2$, are obtained from the test that is constructed without assuming any parametric form for the conditional mean (in this case, the conditional mean is estimated using the

method of Heuchenne and Van Keilegom, 2010). Under the null hypothesis, $\sigma(x) = 1$ and the level of the test, 5%, is respected by the four tests.

Under the alternative models, the test constructed assuming the parametric model, $k = 1$, is more powerful than the test constructed without this assumption, which could be expected. But it is interesting to see that the difference in power is very small.

		$k=1$		$k=2$	
$\sigma(x)$	a_n	T_{KSI}	T_{CMI}	T_{KSI}	T_{CMI}
1	0.20	4.14	4.74	6.02	6.26
1	0.25	4.52	4.81	5.77	5.54
1	0.30	4.16	4.56	5.80	4.55
$exp(x)$	0.20	65.80	74.85	61.74	72.16
$exp(x)$	0.25	71.70	78.29	68.59	75.22
$exp(x)$	0.30	73.72	79.66	70.89	76.22
$(1+x)^2$	0.20	83.61	89.96	81.29	88.77
$(1+x)^2$	0.25	89.28	93.03	88.19	92.12
$(1+x)^2$	0.30	91.72	94.03	90.39	93.08

Table 2: *Percentage of rejections of (4.5) for model (4.4) under the null hypothesis and under two alternatives (nominal level 5%).*

5 Data analysis

In order to analyse the data set hereunder, the distributions of the statistics $T_{KSI,k}$ and $T_{CMI,k}$, $k = 0, 1, 2$, under the null hypothesis are needed. Unfortunately, the asymptotic distributions obtained in Theorem 3.1 are too complicated and contain too many unknow quantities. We therefore propose a bootstrap procedure to estimate the critical values of the tests in practical situations. This is based on a smoothed version of the 'naive bootstrap' described in Efron (1981) and on the method of Pardo Fernández, Van Keilegom and González Manteiga (2007).

First, define $\tilde{E}_1^{0k}, \ldots, \tilde{E}_n^{0k}$ the standardized versions of the residuals $\hat{E}_1^0, \ldots, \hat{E}_n^0$. Note that this standardization depends on the testing procedure. Indeed, define $\lambda_1 = \int e d\hat{F}_\varepsilon^0(e)$, $\lambda_2^2 = \int (e-\lambda_1)^2 J(\hat{F}_\varepsilon^0(e)) d\hat{F}_\varepsilon^0(e)$ and $\lambda_3^2 = \int (e-\lambda_1)^2 d\hat{F}_\varepsilon^0(e)$. We will have $\tilde{E}_i^{00} = (\hat{E}_i^0 - \lambda_1)/\lambda_2$

and $\tilde{E}_i^{01} = \tilde{E}_i^{02} = (\hat{E}_i^0 - \lambda_1)/\lambda_3$, $i = 1, \ldots, n$. The boostrap procedure consists of the following steps. For fixed B and $b = 1, \ldots, B$,

1. For $i = 1, \ldots, n$:
   · Let

$$Y_{i,b,0}^{**} = m_{\vartheta_{n0}^T}(X_i) + \hat{\sigma}^0_{\vartheta_{n0}^T}(X_i)\varepsilon_{i,b}^{*0} \quad \text{for } k = 0,$$
$$Y_{i,b,1}^{**} = m_{\vartheta_{n0}^T}(X_i) + \sigma_{\vartheta_{n1}^T}(X_i)\varepsilon_{i,b}^{*1} \quad \text{for } k = 1,$$
$$Y_{i,b,2}^{**} = \tilde{m}_T(X_i) + \sigma_{\vartheta_{n2}^T}(X_i)\varepsilon_{i,b}^{*2} \quad \text{for } k = 2,$$

   where $\hat{\sigma}^0_{\vartheta_{n0}^T}(X_i) = \int (y - m_{\vartheta_{n0}^T}(X_i))^2 J(\hat{F}(y|X_i))d\hat{F}(y|X_i)$, $\varepsilon_{i,b}^{*k} = V_{i,b}^k + aS_{i,b}$, $V_{i,b}^k$ is drawn from $\tilde{F}_\varepsilon^{0k}$, $k = 0, 1, 2$, (the Kaplan-Meier estimator based on the standardized residuals) and $S_{i,b}$ is a random variable with mean 0 and variance 1 which introduces a small perturbation in the residuals (controlled by the constant a).
   · Select $C_{i,b}^{**}$ from a smoothed version of $\hat{G}(\cdot|X_i)$, the Beran (1981) estimator of the distribution $G(\cdot|X_i) = P(C \leq \cdot|X_i)$ obtained by replacing Δ_i by $1 - \Delta_i$ in the expression of $\hat{F}(\cdot|X_i)$.
   · Let $Z_{i,b,k}^{**} = \min(Y_{i,b,k}^{**}, C_{i,b}^{**})$ and $\Delta_{i,b,k}^{**} = I(Y_{i,b,k}^{**} \leq C_{i,b}^{**})$.

2. The bootstrap sample is $\{(X_i, Z_{i,b,k}^{**}, \Delta_{i,b,k}^{**}),\ i = 1, \ldots, n\}$ for $k = 0, 1$ or 2.

3. Let $T_{KSI,b,k}^{**}$ and $T_{CMI,b,k}^{**}$ be the test statistics calculated with the corresponding bootstrap sample ($k = 0, 1$ or 2).

Let $T_{KSI,(b),k}^{**}$ be the $b-$th order statistic of $T_{KSI,1,k}^{**}, \ldots, T_{KSI,B,k}^{**}$, $k = 0, 1, 2$, and analogously for $T_{CMI,(b),k}^{**}$. Then $T_{KSI,([(1-\alpha)B]+1),k}^{**}$ and $T_{CMI,([(1-\alpha)B]+1),k}^{**}$ (where $[\cdot]$ denotes the integer part) approximate the $(1-\alpha)-$quantiles of the distributions of $T_{KSI,k}$ and $T_{CMI,k}$, $k = 0, 1, 2$.

Remark 5.1 (Parametric bootstrap) When a parametric form is assumed for the distribution $F_\varepsilon(\cdot)$ of ε with $E[\varepsilon] = 0$ and $Var[\varepsilon] = 1$, we can avoid the smoothed bootstrap part characterized by $\varepsilon_{i,b}^{*k}$ above and directly generate residuals from a parametrically estimated version of $F_\varepsilon(\cdot)$. A classical maximum likelihood procedure involving the standardized residuals $\tilde{E}_i^{01}$ can provide estimates for the parameters of $F_\varepsilon(\cdot)$. We can use those generated residuals to construct bootstrap samples when $k = 1$ or 2 while we need to divide them by the estimate of $(\int e^2 J(F_\varepsilon(e))dF_\varepsilon(e))^{1/2}$ when $k = 0$.

This bootstrap procedure will be applied to approximate the critical values in the following practical situation. The survey *Encuesta de Población Activa* (Labour Force

Survey) is carried out by the Spanish Institute for Statistics to collect information about employment. About 60,000 homes in Spain are surveyed each three months. Each home is followed for the next 18 months. Here the available information corresponds to unemployment spells of married women in the region of Galicia. It is 1,009 spells in total, but three of them were deleted after outlier detection, so the sample size will be 1,006.

If a woman is still unemployed when the follow-up ends, then a censored observation appears. In this data set, 563 out of 1,006 observations were censored. Here a regression model of the time of unemployment over the age when entering the unemployment stock is studied. In particular, the goodness-of-fit of a linear model of the logarithm of the time of unemployment over the age is tested by using the techniques proposed in this paper.

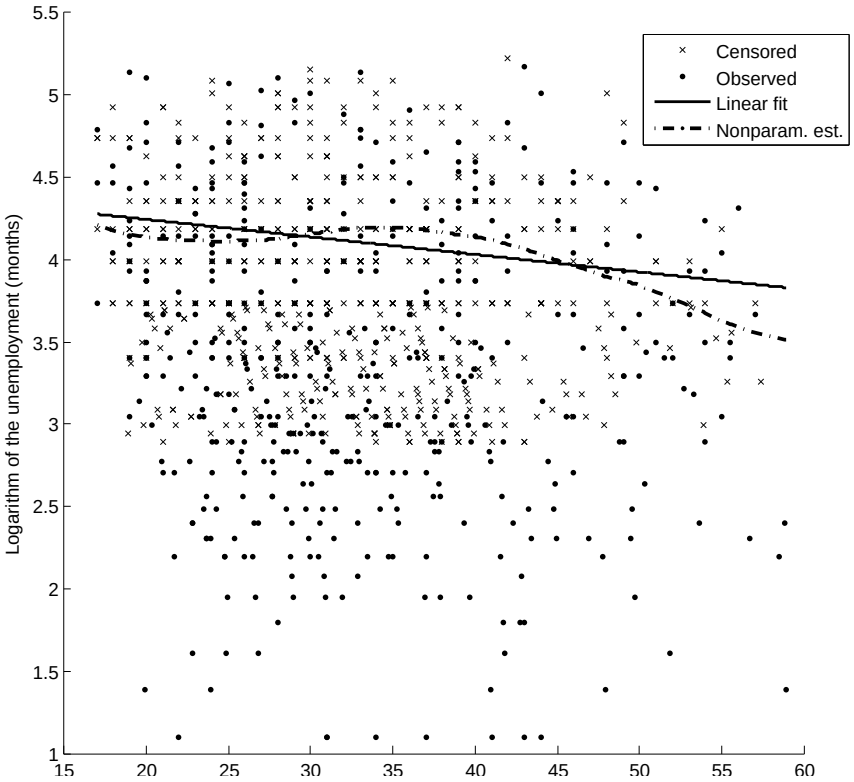


Figure 1: Logarithm of unemployment time of married women in Galicia against age: parametric and nonparametric estimations.

Figure 1 shows a scatter plot of the unemployment data, together with a linear regression fit and a nonparametric estimation of the regression function. The horizontal axis represents the age in years when becoming unemployed, the vertical axis represents the natural logarithm of the time of unemployment in months, squares are used for uncensored

observations and diamonds for censored ones, the straigth line is a linear fit obtained by the method of Heuchenne and Van Keilegom (2008a), and the curve is a nonparametric estimation of the regression function as described in Heuchenne and Van Keilegom (2010). A bandwidth of nine years was used in these estimations. The goodness-of-fit of the linearity was tested by using the methods proposed in this paper. The p-values were 0.005 when using the Kolmogorov-Smirnov type statistic and 0.007 when using the Cramer-von Mises statistic. Other values of the bandwidth parameter led to similar p-values. Then, some evidence is found against a linear model, due to a different evolution of the logarithm of unemployment time in relation with the age. In particular, the unemployment time doesn't look monotone as function of the age.

6 Conclusion

In this paper, we propose new goodness-of-fit testing procedures for the conditional mean and variance functions in a location-scale regression model when the response is possibly right-censored. We obtain weak convergence of the corresponding processes under the null hypothesis and local (Pitman) alternatives. We compare our new methods with the one of Stute et al. (2000) and show its superiority in many situations. We finally provide a bootstrap algorithm for the distributions of the test statistics under the null hypothesis and apply the resulting procedures on a data set about unemployment in Galicia.

An important issue for further research stays the treatment of the right tails of nonparametric estimators with censored data. In this paper, this problem is alleviated by the use of the error distribution estimator $\hat{F}_{\varepsilon}^{0}(y)$, which can be consistent far in the right tail with respect to a local Beran (1981) estimator, see also Heuchenne and Van Keilegom (2010), pages 443, 444 for a full detailed explanation of this problem. Another way to partially circumvent this problem is to consider truncated versions of the conditional mean and variance functions. This objective, different from this paper, is however less practical since it leads to final estimators and tests that correspond to different location and scale functions more difficult to interpret, compare... in applications. Such a work can be achieved using ideas developed for example in Dette and Heuchenne (2012) and Van Keilegom and Heuchenne (2012).

References

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical Report, Univ. California, Berkeley.

Billingsley, P. (1968). Convergence of probability measures. Wiley, New York.

Dette, H. and Heuchenne, C. (2012). Scale checks in censored regression. *Scandinavian Journal of Statistics.*, **39**, 323–339.

Efron, B. (1981). Censored data and the bootstrap. *Journal of the American Statistical Association*, **76**, 312–319.

Heuchenne, C. and Laurent, G. (2017). Parametric conditional variance estimation in location-scale models with censored data. *Electronic Journal of Statistics*, **11**, 148-176.

Heuchenne, C. and Van Keilegom, I. (2007). Polynomial regression with censored data based on preliminary nonparametric estimation. *Annals of the Institute of Statistical Mathematics*, **59**, 273–298.

Heuchenne, C. and Van Keilegom, I. (2008a). Nonlinear regression with censored data. *Technometrics*, **49**, 34–44.

Heuchenne, C. and Van Keilegom, I. (2008b). Location estimation in nonparametric regression with censored data. *Journal of Multivariate Analysis*, **98**, 1558–1582.

Heuchenne, C. and Van Keilegom, I. (2010). Estimation in nonparametric location-scale regression models with censored data. *Annals of the Institute of Statistical Mathematics*, **62**, 439–463.

Heuchenne, C. and Van Keilegom, I. Estimation of a general parametric location in censored regression. In: Van Keilegom, Ingrid; Wilson, Paul W., Eds. Exploring research frontiers in contemporary statistics and econometrics - A Festschrift for Léopold Simar. Berlin, Heidelberg: Springer-Verlag, 2012, 177-188.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481.

Levenberg, K. (1944). A method for the solution of certain problems in least squares. *Quarterly of Applied Mathematics*, **2**, 164–168.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics*, **11**, 431–441.

Pardo-Fernández, J. C., Van Keilegom, I. and González-Manteiga, W. (2007). Goodness-of-fit tests for parametric models in censored regression. *Canadian Journal of Statistics*, **35**, 249–264.

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. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis*, **45**, 89–103.

Stute, W. (1997). Nonparametric model checks for censored regression. *Annals of Statistics*, **25**, 613–641.

Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica*, **9**, 1089–1102.

Stute, W., González Manteiga, W. and Sánchez Sellero, C. (2000). Nonparametric Model Checks In Censored Regression. *Communications in Statistics. Theory and Methods*, **29**, 1611–1629.

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