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*

QuantOM,†

HEC-Management School of University of Liège

Université de Liège

Cesar Sánchez Sellero

Departamento de Estatística e I.O.

Universidad de Santiago de Compostela

June 1, 2010

Abstract

Suppose the random vector (X, Y) satisfies the regression model $Y = m(X) + \sigma(X)\varepsilon$, where $m(\cdot) = E(Y|\cdot)$ (for instance a linear function of the exogenous variable X), $\sigma^2(\cdot) = \mathrm{Var}(Y|\cdot)$ (by example a constant in the homoscedastic case) and ε is independent of X. The response Y is subject to random right censoring and the covariate X is completely observed. New goodness-of-fit testing procedures for m and $\sigma^2(\cdot)$ are proposed. They are based on an integrated regression function technique which uses artificial data points constructed with the method of Heuchenne and Van Keilegom (2007b). Weak convergence of the resulting processes is obtained and their finite sample behaviour is compared via simulations with the method of Stute, González Manteiga and Sánchez Sellero (2000).

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

---

*Financial support from IAP research network P6/03 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged.

†Centre for Quantitative Methods and Operations Management

1 Introduction

For a long time, the econometric literature devotes an increasing interest to modeling censored data. This is displayed in a number of papers provided, e.g., by Chamberlain (1988), Kiefer (1988) or Lewbel and Linton (2002). The main motivation for this interest comes from the fact that in many econometric settings, duration variables can be subject to random right censoring. Indeed, durations are possibly not completely observed since their evolution can be interrupted for several reasons, by example, simply the limits of a survey. Since lots of econometric studies use these durations as endogenous variables, it is necessary to achieve statistical inference for censored data in the regression context. In this paper, we therefore 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 Van Keilegom (2007a, 2007b) 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. 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 Appendix 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 $m^0(x)=T(F(\cdot|x))$ and $\sigma^0(x)=S(F(\cdot|x))$ for some functionals T and S that satisfy $T(F_{aY+b}(\cdot|x))=aT(F_Y(\cdot|x))+b$ 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. Define $F(y|x)=P(Y\leq y|x)$, $G(y|x)=P(C\leq y|x)$, $H(y|x)=P(Z\leq y|x)$, $H(y)=P(Z\leq y)$, $H_\delta(y|x)=P(Z\leq y,\Delta=\delta|x)$, $F_X(x)=P(X\leq x)$, $F^0_\varepsilon(y)=P(\varepsilon^0\leq y)$, $S^0_\varepsilon(y)=1-F^0_\varepsilon(y)$, for $E^0=(Z-m^0(X))/\sigma^0(X)$, we denote $H^0_\varepsilon(y)=P(E^0\leq y)$, $H^0_{\varepsilon\delta}(y)=P(E^0\leq y,\Delta=\delta)$, $H^0_\varepsilon(y|x)=P(E^0\leq y|x)$, $H^0_{\varepsilon\delta}(y|x)=P(E^0\leq y,\Delta=\delta|x)$ ($\delta=0,1$) and for $C^0=(C-m^0(X))/\sigma^0(X)$, we denote $G^0_\varepsilon(y)=P(C^0\leq y)$. The probability density functions of the distributions defined above will be denoted with lower case letters and $R_X = [x_e, x_s]$ denotes the compact support of the variable X.

We have

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

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

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

where $F^{-1}(s|x) = \inf\{y; F(y|x) \geq s\}$ 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) :

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

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,

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

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

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

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 $\hat{E}_1^0, \dots, \hat{E}_n^0$ 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 $\psi^k(X_i, Y_i)(k = 0, 1)$:

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

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

where $m_{\theta_0}(\cdot)$ in (2.6) is replaced by $m(\cdot)$ to obtain the expression of $\hat{\psi}_T^{2*}(X_i, Z_i, \Delta_i)$, $T < \tau_{H_\varepsilon^0}$ and $\tau_F = \inf\{y : F(y) = 1\}$ for any distribution F. Truncations by T in the above integrals and denominators are due to right censoring (however, when $\tau_{F_\varepsilon^0} \le \tau_{G_\varepsilon^0}$, T can be chosen arbitrarily close to $\tau_{F_\varepsilon^0}$). In $\hat{\psi}_T^{1*}(X_i, Z_i, \Delta_i)$, θ_0 can be replaced by its estimator obtained by the method of Heuchenne and Van Keilegom (2007b), while $m(\cdot)$ in $\hat{\psi}_T^{2*}(X_i, Z_i, \Delta_i)$ can be replaced by a nonparametric estimator, say $\tilde{m}_T(X_i)$, developed, by example, in Heuchenne and Van Keilegom (2007c, 2007d).

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

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

as estimators for the parameters describing $\mathcal{M}_k = \{\Psi_{\vartheta_k} : \vartheta_k \in \Theta_k\}$ (Θ_k is a compact subset of $I\!R^{D_k}$, 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, $k = 0, 1, 2$. 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 $\hat{\psi}_T^{1*}(X_i, Z_i, \Delta_i) = \hat{\psi}_T^{1*}(X_i, Z_i, \Delta_i, \vartheta_{n0}^T) = \widehat{(Y_i - m_{\vartheta_{n0}^T}(X_i))_T^{2*}}$, $i = 1, \ldots, n$, 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

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

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.

As it is clear from the definitions of $\hat{\psi}^{k*}_{Ti}(\vartheta^T_{n0})$, $\Psi_{\vartheta^T_{nk}}(X_i)$ and ϑ^T_{nk} for $k = 0, 1, 2$, expression (2.8) is actually estimating

$$n^{-1/2} \sum_{i=1}^{n} (\psi^{k*}_{Ti}(\theta^T_0) - \Psi_{\theta^T_k}(X_i)) I(X_i \le x), \;\; k = 0, 1, 2, \tag{2.9}$$

where

$$\psi^{0*}_{Ti} = Y^*_{Ti} = Y_i \Delta_i + \Big\{ m^0(X_i) + \frac{\sigma^0(X_i)}{1 - F^0_\varepsilon(E^0_i \wedge T)} \int_{E^0_i \wedge T}^{T} y \, dF^0_\varepsilon(y) \Big\} (1 - \Delta_i), \tag{2.10}$$

$$\begin{aligned} \psi^{1*}_{Ti}(\theta^T_0) = (Y_i - m_{\theta^T_0}(X_i))^{2*}_T &= (Y_i - m_{\theta^T_0}(X_i))^2 \Delta_i + \Big\{ (m^0(X_i) - m_{\theta^T_0}(X_i))^2 \\ &+ \frac{\int_{E^0_i \wedge T}^{T} \sigma^0(X_i) y [\sigma^0(X_i) y + 2(m^0(X_i) - m_{\theta^T_0}(X_i))] \, dF^0_\varepsilon(y)}{1 - F^0_\varepsilon(E^0_i \wedge T)} \Big\} (1 - \Delta_i), \end{aligned} \tag{2.11}$$

where ψ^{2*}_{Ti} is obtained by replacing $m_{\theta^T_0}(X_i)$ in $\psi^{1*}_{Ti}(\theta^T_0)$ by $m_T(X_i)$, $i = 1, \ldots, n$ ($m_T(\cdot)$ is the stochastic limit of $\tilde{m}_T(\cdot)$ when $n \to \infty$), $\Psi_{\theta^T_0}(\cdot) = m_{\theta^T_0}(\cdot)$, $\Psi_{\theta^T_p}(\cdot) = \sigma^2_{\theta^T_p}(\cdot)$, $p = 1, 2$, and $\theta^T_k = (\theta^T_{k1}, \ldots, \theta^T_{kD_k})$, $k = 0, 1, 2$, are the unique parameters which minimize

$$E[\{E(\psi^{k*}_T(\theta^T_0)|X) - \Psi_{\vartheta_k}(X)\}^2]$$

(see hypothesis (A10) in the Appendix). However, $\psi^{k*}_T(\theta^T_0)$, $\Psi_{\theta^T_k}(X)$, θ^T_k can be made arbitrarily close to $\psi^{k*}(X, Z, \Delta, \theta_0)$, $\Psi_{\theta_k}(X)$, θ_k, $k = 0, 1, 2$, provided $\tau_{F^0_\varepsilon} \le \tau_{G^0_\varepsilon}$.

Remark 2.1 (Test with known 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 Appendix.

Theorem 3.1 *Assume (A1)-(A10) (in the Appendix). 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* $\sup\{|R_n(x)|; x \in R_X\} = o_P(n^{-1/2})$ *and* $\chi_k^x(X_i, Z_i, \Delta_i, \theta_0^T, \theta_k^T)$ *is defined in the Appendix.*

Theorem 3.2 *Assume (A1)-(A10). 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*

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

Corollary 3.3 *Under the null hypothesis* H_0 *and the assumptions of Theorem 3.2,*

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

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

k=0,1,2.

Remark 3.1 (Non zero mean asymptotic representation) In the asymptotic representation of Theorem 3.1, the expression $n^{-1}\sum_{i=1}^{n}(\psi_{Ti}^{k*}(\theta_0^T) - \Psi_{\theta_k^T}(X_i))I(X_i \leq x)$ has in fact 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 (2007d), the inconsistent region of (2.4) is smaller than inconsistent regions of other distribution estimators for censored data (like,e.g., the Beran estimator). Indeed, as mentioned in Section 2, $\psi_T^{k*}(\theta_0^T) - \Psi_{\theta_k^T}(X)$, $k = 0, 1, 2$, can be made arbitrarily close to a random variable with zero conditional mean provided $\tau_{F_\varepsilon^0} \leq \tau_{G_\varepsilon^0}$. Moreover, even if this last inequality is not true, the results of Theorem 3.2 and Corollary 3.3 could remain valid if $\psi_T^{k*}(\theta_0^T)$ would satisfy a model for which its conditional mean would be $\Psi_{\theta_k^T}(X) + R_n(X)$, for some $R_n(y)$

such that $\sup_{y \in R_X} |R_n(y)| = o_P(n^{-1/2})$.

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 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 Appendix by $E[\{E(Y_{0T}^*|X) - m_{\vartheta_0}(X)\}^2]$, where $Y_{0T}^* = Y_T^* - n^{-1/2} r(X)$. Theoretically, the use of C_0, Y_0, Z_0 and Y_{0T}^* enables in fact to make the main parts of the asymptotic representations under H_{1n} independent of n and equal to the asymptotic representations obtained under the null hypothesis. Then, it can be shown that the term $n^{-1/2}\Omega^{-1} E[r(X)\frac{\partial m_{\theta_0^T}(X)}{\partial \vartheta_0}]$ is added to the asymptotic representation of Theorem 3.2 in Heuchenne and Van Keilegom (2007b), 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, $j = 1, \ldots, D_0$. That leads to add the term

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

to the asymptotic representation of Theorem 3.1 (in the conditional mean case), 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

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

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

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

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 $\hat{E}_{(n)}^0$ of the estimated residuals $\hat{E}_i^0$, $i = 1, \ldots, n$. 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.

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 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 (2007b), Stute's method suffers from restrictive conditions leading to increases of biases and variances of resulting estimators. On the other side,

$a(x)$	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%).*

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, 2007d). 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 Corollary 3.3 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

$$Y_{i,b,0}^{**} = m_{\vartheta_{n0}^T}(X_i) + \hat{\sigma}_{\vartheta_{n0}^T}^0(X_i)\varepsilon_{i,b}^{*0} \quad \text{for } k = 0,$$

$$Y_{i,b,1}^{**} = m_{\vartheta_{n0}^T}(X_i) + \sigma_{\vartheta_{n1}^T}(X_i)\varepsilon_{i,b}^{*1} \quad \text{for } k = 1,$$

$$Y_{i,b,2}^{**} = \tilde{m}_T(X_i) + \sigma_{\vartheta_{n2}^T}(X_i)\varepsilon_{i,b}^{*2} \quad \text{for } k = 2,$$

   where $\hat{\sigma}_{\vartheta_{n0}^T}^0(X_i) = \int (y - m_{\vartheta_{n0}^T}(X_i))^2 J(\hat{F}(y|X_i)) d\hat{F}(y|X_i)$, $\varepsilon_{i,b}^{*k} = V_{i,b}^k + aS_{i,b}$, $V_{i,b}^k$ 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 $G(\cdot|X_i)$ 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 $T_{CMI,k}$, $k = 0, 1, 2$.

This bootstrap procedure will be applied to approximate the critical values in the following practical situation. The survey *Encuesta de Poblacin 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.

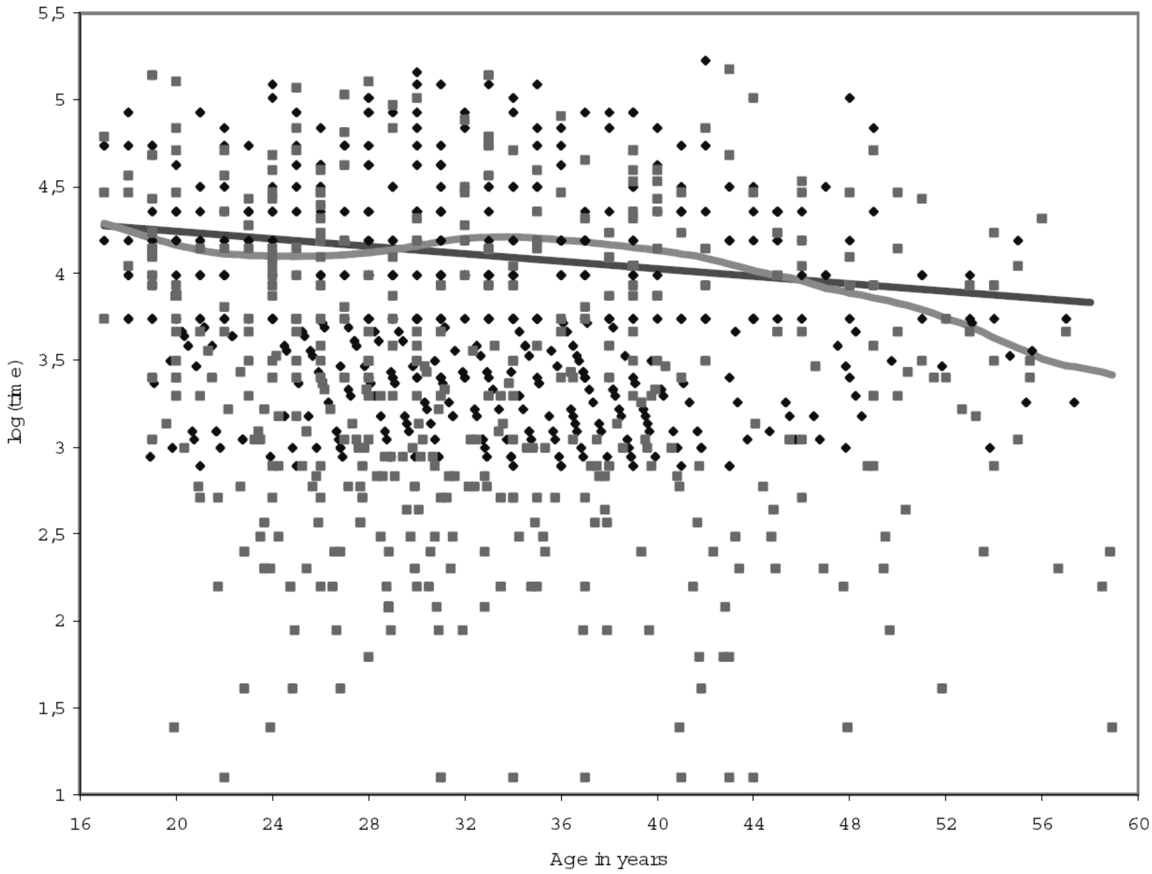

Figure 1: *Logarithm of unemployment time of married women in Galicia against age: parametric and nonparametric estimations.*

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 (2007b), and the curve is a nonparametric estimation of the regression function as described in Heuchenne and Van Keilegom (2007d). 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.

Appendix

The following notations are needed in the statement of the asymptotic results given Section 3.

$$\xi_\varepsilon(z,\delta,y) = (1 - F_\varepsilon^0(y)) \left\{ - \int_{-\infty}^{y\wedge z} \frac{dH_{\varepsilon 1}^0(s)}{(1 - H_\varepsilon^0(s))^2} + \frac{I(z \leq y, \delta = 1)}{1 - H_\varepsilon^0(z)} \right\},$$

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

$$\eta(z,\delta|x) = \int_{-\infty}^{+\infty} \xi(z,\delta,v|x) J(F(v|x))\, dv\, \sigma^0(x)^{-1},$$

$$\zeta(z,\delta|x) = \int_{-\infty}^{+\infty} \xi(z,\delta,v|x) J(F(v|x)) \frac{v - m^0(x)}{\sigma^0(x)}\, dv\, \sigma^0(x)^{-1},$$

$$\gamma_1(y|x) = \int_{-\infty}^{y} \frac{h^0_\varepsilon(s|x)}{(1-H^0_\varepsilon(s))^2}\, dH^0_{\varepsilon 1}(s) + \int_{-\infty}^{y} \frac{d\, h^0_{\varepsilon 1}(s|x)}{1-H^0_\varepsilon(s)},$$

$$\gamma_2(y|x) = \int_{-\infty}^{y} \frac{s h^0_\varepsilon(s|x)}{(1-H^0_\varepsilon(s))^2}\, dH^0_{\varepsilon 1}(s) + \int_{-\infty}^{y} \frac{d\,(s h^0_{\varepsilon 1}(s|x))}{1-H^0_\varepsilon(s)},$$

$$\varphi(x,z,\delta,y) = \xi_\varepsilon\left(\frac{z-m^0(x)}{\sigma^0(x)}, \delta, y\right) - S^0_\varepsilon(y)\eta(z,\delta|x)\gamma_1(y|x) - S^0_\varepsilon(y)\zeta(z,\delta|x)\gamma_2(y|x),$$

$$\begin{aligned}
\chi_{10}(v_1,z_2,\delta_2,p) = I(\delta_1=0) f_X^{-1}(x_1)\sigma^{0^p}(x_1) \Bigg\{ \Bigg[&-\frac{(e^{0T}_{x_1}(z_1))^p f^0_\varepsilon(e^{0T}_{x_1}(z_1))}{1-F^0_\varepsilon(e^{0T}_{x_1}(z_1))} I(z_{1_{x_1}} \le T) \\
&+ \frac{f^0_\varepsilon(e^{0T}_{x_1}(z_1)) \int_{e^{0T}_{x_1}(z_1)}^T u^p\, dF^0_\varepsilon(u)}{(1-F^0_\varepsilon(e^{0T}_{x_1}(z_1)))^2} - I(p=1) \Bigg] \eta(z_2,\delta_2|x_1) \\
&+ \Bigg[\frac{e^{0T}_{x_1}(z_1) f^0_\varepsilon(e^{0T}_{x_1}(z_1)) \int_{e^{0T}_{x_1}(z_1)}^T u^p\, dF^0_\varepsilon(u)}{(1-F^0_\varepsilon(e^{0T}_{x_1}(z_1)))^2} - p\frac{\int_{e^{0T}_{x_1}(z_1)}^T u^p\, dF^0_\varepsilon(u)}{1-F^0_\varepsilon(e^{0T}_{x_1}(z_1))} \\
&- \frac{(e^{0T}_{x_1}(z_1))^{p+1} f^0_\varepsilon(e^{0T}_{x_1}(z_1))}{1-F^0_\varepsilon(e^{0T}_{x_1}(z_1))} I(z_{1_{x_1}} \le T) \Bigg] \zeta(z_2,\delta_2|x_1) \Bigg\}, \quad p = 1,2,
\end{aligned}$$

$$\begin{aligned}
\chi_{20}(v_1,v_2,p) = \frac{\sigma^{0^p}(x_1) I(\delta_1=0)}{1-F^0_\varepsilon(e^{0T}_{x_1}(z_1))} \Bigg\{ \Bigg[&\frac{\int_{e^{0T}_{x_1}(z_1)}^T u^p\, dF^0_\varepsilon(u)}{1-F^0_\varepsilon(e^{0T}_{x_1}(z_1))} - (e^{0T}_{x_1}(z_1))^p \Bigg] \\
&\times \varphi(x_2,z_2,\delta_2,e^{0T}_{x_1}(z_1)) + T^p \varphi(x_2,z_2,\delta_2,T) \\
&- p\int_{e^{0T}_{x_1}(z_1)}^T u^{p-1}\varphi(x_2,z_2,\delta_2,u)du \Bigg\}, \quad p = 1,2,
\end{aligned}$$

$$\begin{aligned}
\chi_{11}(v_1,z_2,\delta_2,\theta_0^T) = {}& 2[m^0(x_1) - m_{\theta_0^T}(x_1)]\chi_{10}(v_1,z_2,\delta_2,1) \\
&- I(\delta_1=0)\frac{2\sigma^0(x_1)\int_{e^{0T}_{x_1}(z_1)}^T u\, dF^0_\varepsilon(u)}{1-F^0_\varepsilon(e^{0T}_{x_1}(z_1)} f_X(x_1)\sigma^0(x_1)\eta(z_2,\delta_2|x_1) \\
&+ \chi_{10}(v_1,z_2,\delta_2,2),
\end{aligned}$$

$$\chi_3(v_1,m) = I(\delta_1=1)(m(x_1)-z_1)$$

$$-I(\delta_1=0)(m^0(x_1)+\frac{\sigma^0(x_1)\int_{e^{0T}_{x_1}(z_1)}^T u\,dF^0_\varepsilon(u)}{1-F^0_\varepsilon(e^{0T}_{x_1}(z_1))}-m(x_1)),$$

$$\chi_{21}(v_1,v_2,\theta_0^T)=2[m(x_1)-m_{\theta_0^T}(x_1)]\chi_{20}(v_1,v_2,1)+\chi_{20}(v_1,v_2,2)$$

$$+2\sum_{d=1}^{D_0}\frac{\partial m_{\theta_0^T}(x_1)}{\partial\vartheta_{0d}}\chi_3(v_1,m_{\theta_0^T})\kappa_{0d}(x_2,z_2,\delta_2),$$

$$\chi_k^x(v_1,\theta_0^T,\theta_k^T)=\sum_{\delta=0,1}\Big\{f_X(x_1)\int\chi_{1k}(x_1,z,\delta,z_1,\delta_1,\theta_0^T)dH_\delta(z|x_1)I(x_1\le x)$$

$$+\int_{x_e\wedge x}^{x\wedge x_s}\int\chi_{2k}(y,z,\delta,v_1,\theta_0^T)h_\delta(z|y)f_X(y)dzdy\Big\}$$

$$-\sum_{d=1}^{D_k}\kappa_{kd}(v_1)\int_{x_e\wedge x}^{x\wedge x_s}\frac{\partial\Psi_{\theta_k^T}(y)}{\partial\vartheta_{kd}}dF_X(y)$$

$$+(\psi_{T1}^{k*}(\theta_0^T)-\Psi_{\theta_k^T}(x_1))I(x_1\le x),\quad k=0,1,2,$$

where $v_q=(x_q,z_q,\delta_q)$ for all $x_q\in R_X,\ z_q\in I\!R$, $\delta_q=0,1$, $q=1,2$. $T=(T_x-m^0(x))/\sigma^0(x)$, $z_x=(z-m^0(x))/\sigma^0(x)$ and $e_x^{0T}(z)=z_x\wedge T$, for any $x\in R_X$, $z\in I\!R$. ϑ_{kd} (ϑ_{nkd}^T, θ_{kd}^T) is the d^{th} component ($d=1,\ldots,D_k$) of ϑ_k (ϑ_{nk}^T, θ_k^T), $\vartheta_{nkd}^T-\theta_{kd}^T$ has an asymptotic representation given by $n^{-1}\sum_{i=1}^n\kappa_{kd}(V_i)+o_P(n^{-1/2})$, $\kappa_{0d}(X_i,Z_i,\Delta_i)$ is the d^{th} component of the vector $\Omega^{-1}\rho(X_i,Z_i,\Delta_i)$, $i=1,\ldots,n$, in the representation of Theorem 3.2 in HVK (2007b) (κ_{kd}, for $k=1,2$, is obtained by a straightforward extension of this theorem to the conditional variance case). Finally, in order to work with general functions, we denote (in the proofs below and the functions $\chi_k^x(v_1,\theta_0^T,\theta_k^T)$ above) $\chi_{10}(v_1,z_2,\delta_2,\theta_0^T)=\chi_{10}(v_1,z_2,\delta_2,1)$, $\chi_{20}(v_1,v_2,\theta_0^T)=\chi_{20}(v_1,v_2,1)$ and define $\chi_{12}(v_1,z_2,\delta_2,\theta_0^T)$ and $\chi_{22}(v_1,v_2,\theta_0^T)$ in Lemma A.1 as functions corresponding to the third test described in Section 2.

Let $\tilde T_x$ be any value less than the upper bound of the support of $H(\cdot|x)$ such that

$\inf_{x\in R_X}(1-H(\tilde{T}_x|x))>0$. For a (sub)distribution function $L(y|x)$ we will use the notations $l(y|x)=L'(y|x)=(\partial/\partial y)L(y|x)$, $\dot{L}(y|x)=(\partial/\partial x)L(y|x)$ and similar notations will be used for higher order derivatives.

The assumptions needed for the asymptotic results are listed below.

(A1)(i) $na_n^4\to 0$ and $na_n^{3+2\delta}(\log a_n^{-1})^{-1}\to\infty$ for some $\delta<1/2$.

(ii) $R_X=[x_e,x_s]$ is a compact interval of length L_X.

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

(iv) Ω is non-singular.

(A2)(i) There exist $0\le s_0\le s_1\le 1$ such that $s_1\le\inf_x F(\tilde{T}_x|x)$, $s_0\le\inf\{s\in[0,1];J(s)\ne 0\}$, $s_1\ge\sup\{s\in[0,1];J(s)\ne 0\}$ and $\inf_{x\in R_X}\inf_{s_0\le s\le s_1} f(F^{-1}(s|x)|x)>0$.

(ii) J is twice continuously differentiable, $\int_0^1 J(s)ds=1$ and $J(s)\ge 0$ for all $0\le s\le 1$.

(iii) The function $x\to T_x$ $(x\in R_X)$ is twice continuously differentiable.

(A3)(i) F_X is three times continuously differentiable and $\inf_{x\in R_X} f_X(x)>0$.

(ii) m^0 and σ^0 are twice continuously differentiable and $\inf_{x\in R_X}\sigma^0(x)>0$.

(iii) $E[\varepsilon^{02}]<\infty$ and $E[|Z|^{4(1+v)}]<\infty$ for some $v>0$.

(A4)(i) $\eta(z,\delta|x)$ and $\zeta(z,\delta|x)$ are twice continuously differentiable with respect to x and their first and second derivatives (with respect to x) are bounded, uniformly in $x\in R_X$, $z<\tilde{T}_x$ and δ.

(ii) The first derivatives of $\eta(z,\delta|x)$ and $\zeta(z,\delta|x)$ with respect to z are of bounded variation and the variation norms are uniformly bounded over all x.

$(A5)$ The function $y \to P(m^0(X) + e\sigma^0(X) \le y)$ $(y \in I\!R)$ is differentiable for all $e \in I\!R$ and the derivative is uniformly bounded over all $e \in I\!R$.

$(A6)$ For $L(y|x) = H(y|x), H_1(y|x), H^0_\varepsilon(y|x)$ or $H^0_{\varepsilon 1}(y|x)$: $L'(y|x)$ is continuous in (x,y) and $\sup_{x,y}|y^2 L'(y|x)| < \infty$. The same holds for all other partial derivatives of $L(y|x)$ with respect to x and y up to order three and $\sup_{x,y}|y^3 L'''(y|x)| < \infty$.

$(A7)$ (i) $\sup_{x_2,z_2}\sum_{\delta_1=0,1}\int|\chi'_{1k}(x_2,z_1,\delta_1,z_2,\delta_2,\theta_0^T)|h(z_1)\,dz_1 < \infty$ $(\delta_2 = 0,1,\ k = 0,1,2)$,

(ii) $\sup_{z_2}\sum_{\delta_1=0,1}\int\sup_{x_2}|\chi''_{1k}(x_2,z_1,\delta_1,z_2,\delta_2,\theta_0^T)|h(z_1)\,dz_1 < \infty$ $(\delta_2 = 0,1,\ k = 0,1,2)$,

where $\chi'^{(\prime)}_{1k}(x_2,z_1,\delta_1,z_2,\delta_2,\theta_0^T)$ equals the first (second) derivative of $\chi_{1k}(x_2,z_1,\delta_1,z_2,\delta_2,\theta_0^T)$ with respect to x_2 when $z_1 \ne T_{x_2}$ and equals 0 otherwise.

$(A8)$ For the density $f_{X|Z,\Delta}(x|z,\delta)$ of X given (Z,Δ), $\sup_{x,z}|f_{X|Z,\Delta}(x|z,\delta)| < \infty$, $\sup_{x,z}|\dot f_{X|Z,\Delta}(x|z,\delta)| < \infty$ and $\sup_{x,z}|\ddot f_{X|Z,\Delta}(x|z,\delta)| < \infty$ $(\delta = 0,1)$.

$(A9)$ Θ_k is compact and θ_k^T is an interior point of Θ_k, $k = 0,1,2$. All partial derivatives of $\Psi_{\vartheta_k}(x)$ with respect to the components of ϑ_k and x up to order three exist and are continuous in (x,ϑ_k) for all x and ϑ_k $(k = 0,1,2)$.

$(A10)$ The function $E[\{E(\psi_T^{k*}(\theta_0^T)|X) - \Psi_{\vartheta_k}(X)\}^2]$ has a unique minimum in $\vartheta_k = \theta_k^T$ for each k, $k = 0,1,2$.

Proof of Theorem 3.1. The expression to develop to obtain an asymptotic representation is decomposed according to

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n}(\hat{\psi}_{Ti}^{k*}(\vartheta_{n0}^{T}) - \Psi_{\vartheta_{nk}^{T}}(X_i))I(X_i \le x) &= n^{-1}\sum_{i=1}^{n}(\hat{\psi}_{Ti}^{k*}(\vartheta_{n0}^{T}) - \psi_{Ti}^{k*}(\theta_0^{T}))I(X_i \le x) \\
&\quad + n^{-1}\sum_{i=1}^{n}(\psi_{Ti}^{k*}(\theta_0^{T}) - \Psi_{\vartheta_{nk}^{T}}(X_i))I(X_i \le x) \\
&= \Omega_{1n}^{x} + \Omega_{2n}^{x}. \qquad \text{(A.1)}
\end{aligned}
$$

First, we treat Ω_{1n}^{x}. By Lemma A.1, this term is decomposed into two parts for which the first one is rewritten

$$
n^{-2}a_n^{-1}\sum_{j\neq i}K\Big(\frac{X_i - X_j}{a_n}\Big)\chi_{1k}(V_i, Z_j, \Delta_j, \theta_0^{T})I(X_i \le x) + o_P(n^{-1/2}). \qquad \text{(A.2)}
$$

We can easily show that (A.2) can be written as

$$
\begin{aligned}
&(n^2 a_n)^{-1}\sum_{j\neq i}\{\chi_{1kK}^{x*}(V_i, V_j, \theta_0^{T}) + E[\chi_{1kK}^{x}(V_i, V_j, \theta_0^{T})|V_i] \\
&\quad + E[\chi_{1kK}^{x}(V_i, V_j, \theta_0^{T})|V_j] - E[\chi_{1kK}^{x}(V_i, V_j, \theta_0^{T})]\} + o_P(n^{-1/2}) \qquad \text{(A.3)} \\
&= T_1 + T_2 + T_3 + T_4 + o_P(n^{-1/2}),
\end{aligned}
$$

where

$$
\chi_{1kK}^{x}(V_i, V_j, \theta_0^{T}) = K(\frac{X_i - X_j}{a_n})\chi_{1k}(V_i, Z_j, \Delta_j, \theta_0^{T})I(X_i \le x),
$$

and

$$
\begin{aligned}
\chi_{1kK}^{x*}(V_i, V_j, \theta_0^{T}) &= \chi_{1kK}^{x}(V_i, V_j, \theta_0^{T}) - E[\chi_{1kK}^{x}(V_i, V_j, \theta_0^{T})|V_i] \\
&\quad - E[\chi_{1kK}^{x}(V_i, V_j, \theta_0^{T})|V_j] + E[\chi_{1kK}^{x}(V_i, V_j, \theta_0^{T})].
\end{aligned}
$$

Consider

$$
\begin{aligned}
&(na_n)^{-1}\sum_{j=1}^{n} E[\chi_{1kK}^{x}(V_i,V_j,\theta_0^T)|V_j] \\
&= (na_n)^{-1}\sum_{j=1}^{n}\sum_{\delta=0,1}\int\int \chi_{1k}(y,z,\delta,Z_j,\Delta_j,\theta_0^T)K(\frac{y-X_j}{a_n})I(y\le x)h_\delta(z|y)f_X(y)\,dz\,dy \\
&= n^{-1}\sum_{j=1}^{n}\sum_{\delta=0,1}\int\int K(u)(\chi_{1k}(X_j,z,\delta,Z_j,\Delta_j,\theta_0^T)+a_nu\frac{\partial\chi_{1k}(X_j,z,\delta,Z_j,\Delta_j,\theta_0^T)}{\partial y} \\
&\qquad +(a_nu)^2\frac{\partial^2\chi_{1k}(y',z,\delta,Z_j,\Delta_j,\theta_0^T)}{\partial y^2})(h_\delta(z|X_j)+a_nu\dot{h}_\delta(z|X_j)+(a_nu)^2\ddot{h}_\delta(z|y'')) \\
&\qquad \times(f_X(X_j)+a_nuf_X'(X_j)+(a_nu)^2f_X''(y'''))I(X_j\le x-a_nu)\,dz\,du \\
&= n^{-1}\sum_{j=1}^{n}\left[f_X(X_j)\sum_{\delta=0,1}\int\chi_{1k}(X_j,z,\delta,Z_j,\Delta_j,\theta_0^T)\,dH_\delta(z|X_j)\int I(X_j\le x-a_nu)K(u)du\right] \\
&\quad +\int n^{-1}\sum_{j=1}^{n}O(a_n|Z_j|)I(X_j\le x-a_nu)uK(u)du+o_P(n^{-1/2}).
\end{aligned}
$$

where y', y'' and y''' lie between X_j and X_j+a_nu. In a similar way, using two Taylor expansions of order 2, we get

$$(n^2a_n)^{-1}\sum_{i\ne j}E[\chi_{1kK}^{x}(V_i,V_j,\theta_0^T)|V_i]=o_P(n^{-1/2}).$$

It follows, using Lemma A.2, that

$$
\begin{aligned}
T_2+T_3+T_4 \;&=\; n^{-1}\sum_{i=1}^{n}f_X(X_i)\sum_{\delta=0,1}\int\chi_{1k}(X_i,z,\delta,Z_i,\Delta_i,\theta_0^T)\,dH_\delta(z|X_i)I(X_i\le x) \\
&\quad +o_P(n^{-1/2}).
\end{aligned}
$$

Now, let

$$\chi_{1k}^{+}(V_i,Z_j,\Delta_j,\theta_0^T)=\max(\chi_{1k}(V_i,Z_j,\Delta_j,\theta_0^T),0),$$

$$\chi_{1k}^{-}(V_i,Z_j,\Delta_j,\theta_0^T)=-\min(\chi_{1k}(V_i,Z_j,\Delta_j,\theta_0^T),0)$$

and similar definitions for $\chi^{+x}_{1kK}(V_i, V_j, \theta_0^T)$ and $\chi^{-x}_{1kK}(V_i, V_j, \theta_0^T)$. So,

$$\chi^{+x*}_{1kK}(V_i, V_j, \theta_0^T) = \chi^{+x}_{1kK}(V_i, V_j, \theta_0^T) - E[\chi^{+x}_{1kK}(V_i, V_j, \theta_0^T)|V_i]$$

$$-E[\chi^{+x}_{1kK}(V_i, V_j, \theta_0^T)|V_j] + E[\chi^{+x}_{1kK}(V_i, V_j, \theta_0^T)]$$

and similarly for $\chi^{-x*}_{1kK}(V_i, V_j, \theta_0^T)$. It is clear that $\chi^{x*}_{1kK}(V_i, V_j, \theta_0^T) = \chi^{+x*}_{1kK}(V_i, V_j, \theta_0^T) - \chi^{-x*}_{1kK}(V_i, V_j, \theta_0^T)$. Partition R_X into $q = [\frac{L_X}{n^{-1/2-\varepsilon_1}}]$ intervals ($[\cdot]$ denotes the integer part) $(x_0, x_1), \ldots, (x_l, x_{l+1}), \ldots, (x_{q-1}, x_q)$ ($l = 0, \ldots, q-1$, $\varepsilon_1 > 0$, $x_0 = x_e$ and $x_q = x_s$) of length $C_1 n^{-1/2-\varepsilon_1}$, where $1 \leq C_1 \leq 2$. Using the monotonicity of the functions $\chi^{+x}_{1kK}(V_i, V_j, \theta_0^T) + E[\chi^{+x}_{1kK}(V_i, V_j, \theta_0^T)]$ and $E[\chi^{+x}_{1kK}(V_i, V_j, \theta_0^T)|V_i] + E[\chi^{+x}_{1kK}(V_i, V_j, \theta_0^T)|V_j]$ with respect to x, we have

$$\sup_x |(n^2 a_n)^{-1} \sum_{j \neq i} \chi^{+x^*}_{1kK}(V_i, V_j, \theta_0^T)| \tag{A.4}$$

$$\leq \max_{0 \leq l \leq q} |(n^2 a_n)^{-1} \sum_{j \neq i} \chi^{+x_l^*}_{1kK}(V_i, V_j, \theta_0^T)|$$

$$+2 \max_{0 \leq l \leq q-1} |(n^2 a_n)^{-1} \sum_{j \neq i} \Big\{ E[K(\frac{X_i - X_j}{a_n}) \chi^{+}_{1k}(V_i, Z_j, \Delta_j, \theta_0^T) I(x_l \leq X_i \leq x_{l+1})|V_i]$$

$$+E[K(\frac{X_i - X_j}{a_n}) \chi^{+}_{1k}(V_i, Z_j, \Delta_j, \theta_0^T) I(x_l \leq X_i \leq x_{l+1})|V_j] \Big\} |.$$

First, we treat the first term on the right hand side of the inequality (A.4). By Chebichev inequality and for some $C_2 > 0$,

$$P(|(n^2 a_n)^{-1} \sum_{j \neq i} \chi^{+x_l^*}_{1kK}(V_i, V_j, \theta_0^T)| > C_2 n^{-1/2-\varepsilon_1}) \tag{A.5}$$

$$\leq \frac{1}{C_2^2 n^{3-2\varepsilon_1} a_n^2} E[\sum_{j \neq i} \sum_{r \neq s} \chi^{+x_l^*}_{1kK}(V_i, V_j, \theta_0^T) \chi^{+x_l^*}_{1kK}(V_r, V_s, \theta_0^T)].$$

Since $E[\chi^{+x_l^*}_{1kK}(V_i, V_j, \theta_0^T)] = 0$, the terms for which $i, j \neq r, s$ are zero. The terms for which either i or j equals r or s and the other differs from r and s, are also zero, because, for example when $i = r$ and $j \neq s$,

$$E[\chi_{1kK}^{+x_l^*}(V_i, V_j, \theta_0^T) E[\chi_{1kK}^{+x_l^*}(V_i, V_s, \theta_0^T) | V_i, V_j]] = 0.$$

Thus, only the $2n(n-1)$ terms for which (i, j) equals (r, s) or (s, r) remain. Under (A1)-(A10), it is easy to check that those terms are $O(a_n)$ uniformly in x such that (A.5) is $O(n^{-1+2\varepsilon_1} a_n^{-1})$ and

$$\begin{aligned}
&P(\max_{0 \leq l \leq q} |(n^2 a_n)^{-1} \sum_{j \neq i} \chi_{1kK}^{+x_l^*}(V_i, V_j, \theta_0^T)| > C_2 n^{-1/2-\varepsilon_1}) \\
&\leq \sum_{l=0}^{q} P(|(n^2 a_n)^{-1} \sum_{j \neq i} \chi_{1kK}^{+x_l^*}(V_i, V_j, \theta_0^T)| > C_2 n^{-1/2-\varepsilon_1}) \\
&= O(n^{-1/2+3\varepsilon_1} a_n^{-1}),
\end{aligned}$$

which tends to zero for ε_1 sufficiently small, for instance, $\varepsilon_1 = 1/18$. Next, we treat the second term on the right hand side of the inequality (A.4). Write

$$\begin{aligned}
&(na_n)^{-1} \sum_{j=1}^{n} E[K(\frac{X - X_j}{a_n}) \chi_{1k}^{+}(V, Z_j, \Delta_j, \theta_0^T) I(x_l \leq X \leq x_{l+1}) | V_j] \\
&\leq n^{-1} \sum_{j=1}^{n} \int \sum_{\delta=0,1} \int |\chi_{1k}(X_j + a_n u, z, \delta, Z_j, \Delta_j, \theta_0^T)| \\
&\qquad f_X(X_j + a_n u) h_\delta(z | X_j + a_n u) I(x_l \leq X_j + a_n u \leq x_{l+1}) K(u) dz du \\
&\leq \int n^{-1} \sum_{j=1}^{n} \Bigg[f_X(X_j) \sum_{\delta=0,1} \int |\chi_{1k}(X_j, z, \delta, Z_j, \Delta_j, \theta_0^T)| \, dH_\delta(z | X_j) \\
&\qquad \times I(x_l - a_n u \leq X_j \leq x_{l+1} - a_n u) K(u) du] \\
&\quad + \int n^{-1} \sum_{j=1}^{n} O(a_n |Z_j|) I(x_l - a_n u \leq X_j \leq x_{l+1} - a_n u) |u| K(u) du + O(a_n^2).
\end{aligned}$$

$$\leq \int n^{-1} \sum_{j=1}^{n} \Bigg[f_X(X_j) \sum_{\delta=0,1} \int |\chi_{1k}(X_j, z, \delta, Z_j, \Delta_j, \theta_0^T)| \, dH_\delta(z|X_j)$$

$$\times I(x_l \leq X_j \leq x_{l+1}) K(u) du \Big]$$

$$+ \int n^{-1} \sum_{j=1}^{n} O(a_n |Z_j|) I(x_l \leq X_j \leq x_{l+1}) |u| K(u) du + o_P(n^{-1/2})$$

$$= A_{1nl} + A_{2nl} + o_P(n^{-1/2}),$$

uniformly in x_l and using Lemma A.2. In this way,

$$P(2 \max_{0 \leq l \leq q-1} |A_{1nl} + A_{2nl}| > C_3 n^{-1/2-\varepsilon_1}) = o(1), \tag{A.6}$$

using Lemma A.3 with $\nu_1 = \varepsilon_1$. With similar arguments, it is easy to check that

$$2 \max_{0 \leq l \leq q-1} |(na_n)^{-1} \sum_{i=1}^{n} E[K(\frac{X_i - X_j}{a_n}) \chi_{1k}^{+}(V_i, Z_j, \Delta_j, \theta_0^T) | V_i] I(x_l \leq X_i \leq x_{l+1})| = o_P(n^{-1/2}).$$

The second term in the expression of Lemma A.1 is also written

$$n^{-2} \sum_{i \neq j} \chi_{2k}(V_i, V_j, \theta_0^T) I(X_i \leq x) + o_P(n^{-1/2}). \tag{A.7}$$

As in (A.3), (A.7) can be decomposed in

$$n^{-2} \sum_{j \neq i} \{ \chi_{2k}^{x*}(V_i, V_j, \theta_0^T) + E[\chi_{2k}(V_i, V_j, \theta_0^T) | V_i] I(X_i \leq x) \tag{A.8}$$

$$+ E[\chi_{2k}(V_i, V_j, \theta_0^T) I(X_i \leq x) | V_j] - E[\chi_{2k}(V_i, V_j, \theta_0^T) I(X_i \leq x)] \} + o_P(n^{-1/2}),$$

where

$$\chi_{2k}^{x*}(V_i, V_j, \theta_0^T) = \chi_{2k}(V_i, V_j, \theta_0^T) I(X_i \leq x) - E[\chi_{2k}(V_i, V_j, \theta_0^T) | V_i] I(X_i \leq x)$$

$$- E[\chi_{2k}(V_i, V_j, \theta_0^T) I(X_i \leq x) | V_j] + E[\chi_{2k}(V_i, V_j, \theta_0^T) I(X_i \leq x)].$$

The second and fourth terms of (A.8) equal zero by definition of $\chi_{2k}(V_i, V_j, \theta_0^T)$. For the third term of (A.8), we obtain

$$n^{-1}\sum_{i=1}^{n}\int_{x_e\wedge x}^{x\wedge x_s}\sum_{\delta=0,1}\int \chi_{2k}(y,z,\delta,V_i,\theta_0^T)h_\delta(z|y)f_X(y)dzdy + o_P(n^{-1/2}). \tag{A.9}$$

The first term of (A.8) is treated similarly to T_1 in (A.3) but in an easier way.

Now, we treat Ω_{2n}^x. It is given by

$$\begin{aligned}\Omega_{2n}^x &= n^{-1}\sum_{i=1}^{n}(\psi_{Ti}^{k*}(\theta_0^T) - \Psi_{\theta_k^T}(X_i))I(X_i \le x)\\ &+n^{-1}\sum_{i=1}^{n}(\Psi_{\theta_k^T}(X_i) - \Psi_{\vartheta_{nk}^T}(X_i))I(X_i \le x).\end{aligned} \tag{A.10}$$

The first term of (A.10) enters the asymptotic representation and easy calculations for the second term show that it is given by

$$\begin{aligned}&-n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{d=1}^{D_k}\frac{\partial\Psi_{\theta_k^T}(X_i)}{\partial\vartheta_{kd}}\kappa_{kd}(X_j,Z_j,\Delta_j)I(X_i\le x) + o_P(n^{-1/2})\\ &= -n^{-1}\sum_{i=1}^{n}\sum_{d=1}^{D_k}\kappa_{kd}(X_i,Z_i,\Delta_i)\int_{x_e\wedge x}^{x\wedge x_s}\frac{\partial\Psi_{\theta_k^T}(y)}{\partial\vartheta_{kd}}dF_X(y) + o_P(n^{-1/2}),\end{aligned}$$

where the last equality is obtained using weak convergence of the above empirical process. This finishes the proof.

Proof of Theorem 3.2. First, let rewrite the asymptotic representation of Theorem 3.1 as

$$\begin{aligned}&n^{-1/2}\sum_{i=1}^{n}\Bigg\{\sum_{\delta=0,1}\Big\{f_X(X_i)\int\chi_{1k}(X_i,z,\delta,Z_i,\Delta_i,\theta_0^T)dH_\delta(z|X_i)I(X_i\le x)\\ &+\int_{x_e\wedge x}^{x\wedge x_s}\int\chi_{2k}(y,z,\delta,V_i,\theta_0^T)h_\delta(z|y)f_X(y)dzdy\Big\}\end{aligned}$$

$$-\sum_{d=1}^{D_k}\kappa_{kd}(X_i,Z_i,\Delta_i)\int_{x_e\wedge x}^{x\wedge x_s}\frac{\partial\Psi_{\theta_k^T}(y)}{\partial\vartheta_{kd}}dF_X(y)$$

$$+(\psi_{Ti}^{k*}(\theta_0^T)-\Psi_{\theta_k^T}(X_i))I(X_i\le x)\Big\}+o_P(n^{-1/2})$$

$$=n^{-1/2}\sum_{i=1}^{n}\Big\{\chi_{4k}(V_i)I(X_i\le x)+\int_{x_e\wedge x}^{x\wedge x_s}\chi_{5k}(y,V_i)dF_X(y) \tag{A.11}$$

$$-\sum_{d=1}^{D_k}\kappa_{kd}(V_i)\int_{x_e\wedge x}^{x\wedge x_s}\frac{\partial\Psi_{\theta_k^T}(y)}{\partial\vartheta_{kd}}dF_X(y)+(\psi_{Ti}^{k*}(\theta_0^T)-\Psi_{\theta_k^T}(X_i))I(X_i\le x)\Bigg\}+o_P(n^{-1/2}),$$

where the last term is uniform in x. Since $n^{-1/2}\sum_{i=1}^{n}\kappa_{kd}(V_i)=O_P(1),\ d=1\ldots,D_k$, using Theorem 3.3 of Heuchenne and Van Keilegom (2007b) and the term $\int_{x_e\wedge x}^{x\wedge x_s}\frac{\partial\Psi_{\theta_k^T}(y)}{\partial\vartheta_{kd}}dF_X(y)$ is continuous with respect to x, the third term of the above expression is tight by Stone's condition. For $\chi_{6k}^n(x)=n^{-1/2}\sum_{i=1}^{n}\int_{x_e\wedge x}^{x\wedge x_s}\chi_{5k}(y,V_i)dF_X(y)$ and $x_1\le x\le x_2$, we compute

$$E[(\chi_{6k}^n(x_2)-\chi_{6k}^n(x))^2(\chi_{6k}^n(x)-\chi_{6k}^n(x_1))^2]=n^{-2}E[(\sum_{i=1}^{n}\alpha_i)^2(\sum_{i=1}^{n}\beta_i)^2], \tag{A.12}$$

according to the notations α_i and β_i of Lemma 5.1 of Stute (1997). In this case, $\alpha_i=\int_{(x\vee x_e)\wedge x_s}^{(x_2\wedge x_s)\vee x_e}\chi_{5k}(y,V_i)dF_X(y)$ and $\beta_i=\int_{(x_1\vee x_e)\wedge x_s}^{(x\wedge x_s)\vee x_e}\chi_{5k}(y,V_i)dF_X(y)$ are i.i.d. square integrable random variables with zero mean. Therefore, using this lemma, the term on the right hand side of the above expression is bounded by

$$n^{-1}(E[\alpha_1^4])^{1/2}(E[\beta_1^4])^{1/2}+\frac{3n(n-1)}{n^2}(E[\alpha_1^4])^{1/2}(E[\beta_1^4])^{1/2}\le 3(E[\alpha_1^4])^{1/2}(E[\beta_1^4])^{1/2},$$

which is $O((\int_{(x_1\vee x_e)\wedge x_s}^{(x_2\wedge x_s)\vee x_e}dF_X(y))^2)$ and where $\int_{x_e}^{(x\wedge x_s)\vee x_e}dF_X(y)$ is a continuous nondecreasing function on R_X. For the first (respectively fourth) term on the right hand side of (A.11), we also use Lemma 5.1 of Stute (1997) with $\chi_{4k}(V_i)I(x<X_i\le x_2)$ (resp. $(\psi_{Ti}^{k*}(\theta_0^T)-\Psi_{\theta_k^T}(X_i))I(x<X_i\le x_2)$) for α_i and $\chi_{4k}(V_i)I(x_1<X_i\le x)$ (resp.

$(\psi_{Ti}^{k*}(\theta_0^T) - \Psi_{\theta_k^T}(X_i))I(x_1 < X_i \le x))$ for β_i in (A.12). Note that in this case $\alpha_i\beta_i = 0$. We also refer to Remark 3.1 for the calculation of the conditional mean of $(\psi_{Ti}^{k*}(\theta_0^T) - \Psi_{\theta_k^T}(X_i))$. Thus, applications of Lemma 5.1 of Stute (1997) lead to establish that the right hand side of (A.12) applied to the first and fourth terms of the right hand side of (A.11) is also $O((\int_{(x_1\vee x_e)\wedge x_s}^{(x_2\wedge x_s)\vee x_e} dF_X(y))^2)$. Finally, applications of Theorem 15.7 in Billingsey (1968) to the first, second and fourth terms on the right hand side of (A.11) finish the proof.

Proof of Corollary 3.3. The convergence of $T_{KSI,k}$, $k = 0, 1, 2$, follows directly from the weak convergence of the process $ICP_k(x)$ and the continuous mapping Theorem. For $T_{CMI,k}$, write

$$\int_{R_X} ICP_k^2(x)d\hat{F}_X(x) - \int_{R_X} W_k^2(x)dF_X(x)$$
$$\le \int_{R_X} (ICP_k^2(x) - W_k^2(x))d\hat{F}_X(x) + \int_{R_X} W_k^2(x)d(\hat{F}_X(x) - F_X(x)).$$

For the first term on the right hand side of the above inequality, we apply the Skorohod construction (see Serfling, 1980) to the process $ICP_k(x)$ such that $\sup_{x\in R_X} |ICP_k(x) - W_k(x)| \to 0,\ a.s.$ The second term is jointly treated by the almost sure uniform consistency of usual empirical processes and the Helly-Bray Theorem (see p. 97 in Rao, 1965) applied to each of the trajectories of $W_k(x)$.

Lemma A.1 *Under (A1)-(A10), the expression* $n^{-1}\sum_{i=1}^n(\hat{\psi}_{Ti}^{k*}(\vartheta_{n0}^T) - \psi_{Ti}^{k*}(\theta_0^T))I(X_i \le x)$,

$k = 0, 1, 2$, *has the following structure:*

$$n^{-2}a_n^{-1}\sum_{i=1}^n\sum_{j=1}^n K\Big(\frac{X_i - X_j}{a_n}\Big)\chi_{1k}(V_i, Z_j, \Delta_j, \theta_0^T)I(X_i \le x)$$
$$+n^{-2}\sum_{i=1}^n\sum_{j=1}^n \chi_{2k}(V_i, V_j, \theta_0^T)I(X_i \le x) + o_P(n^{-1/2}),$$

where $V_i = (X_i, Z_i, \Delta_i)$.

Proof. First, we treat the case $k = 0$. Following the lines of the proof of Theorems 3.1 (the term $A_{1i} + A_{2i} + A_{3i}$) of Heuchenne and Van Keilegom (2007a) (hereafter abbreviated by HVK), we obtain

$$n^{-1}\sum_{i=1}^n(\hat{Y}_{Ti}^* - Y_{Ti}^*)I(X_i \le x)$$
$$= n^{-2}a_n^{-1}\sum_{i=1}^n\sum_{j=1}^n K\Big(\frac{X_i - X_j}{a_n}\Big)\chi_{10}(V_i, Z_j, \Delta_j, 1)I(X_i \le x)$$
$$+n^{-2}\sum_{i=1}^n\sum_{j=1}^n \chi_{20}(V_i, V_j, 1)I(X_i \le x) + o_P(n^{-1/2}). \tag{A.13}$$

Note that the representation in Theorem 3.1 of HVK (2007a) (equation (A.5)) adapted to the integrated regression function is here complemented by the term

$$n^{-1}\sum_{i=1}^n I(\Delta_i = 0)I(X_i \le x)\hat{\sigma}^0(X_i)\left[\frac{\int_{\hat{E}_i^0 \wedge T}^T u\,d\hat{F}_\varepsilon^0(u)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^0 \wedge T)} - \frac{\int_{\hat{E}_i^{0T}}^{\hat{T}_i} u\,d\hat{F}_\varepsilon^0(u)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})}\right],$$

for which $\hat{T}_i = \frac{T_{X_i} - \hat{m}^0(X_i)}{\hat{\sigma}^0(X_i)}$ and $\hat{E}_i^{0T} = \hat{E}_i^0 \wedge \hat{T}_i$. The above term is rewritten

$$n^{-1}\sum_{i=1}^n I(\Delta_i = 0)I(X_i \le x)\hat{\sigma}^0(X_i)\frac{\int_{\hat{T}_i}^T u\,d\hat{F}_\varepsilon^0(u)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})}I(E_i^0 \le T, \hat{E}_i^0 \le T)$$
$$+n^{-1}\sum_{i=1}^n O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})I(\hat{E}_i^0 > T, \hat{E}_i^0 \le \hat{T}_i)$$
$$+n^{-1}\sum_{i=1}^n O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})I(\hat{E}_i^0 \le T, \hat{E}_i^0 > \hat{T}_i)$$
$$= W_1 + W_2 + W_3,$$

using Proposition 4.5 and Corollary 3.2 of Van Keilegom and Akritas (1999) (hereafter abbreviated by VKA) and the fact that $\sup_e |ef_\varepsilon^0(e)| < \infty$ (more details about the developments above can be obtained in the proof of Theorem 3.1 of HVK (2007a) since very similar expressions are handled). When $\hat{E}_i^0 \le T$, it holds that $E_i^0 \le T\hat{\sigma}^0(X_i)/\sigma^0(X_i) + [\hat{m}^0(X_i) - m^0(X_i)]/\sigma^0(X_i) \le T + V$, where

$$V = [\inf_x \sigma^0(x)]^{-1}[\sup_x |\hat{m}^0(x) - m^0(x)| + T \sup_x |\hat{\sigma}^0(x) - \sigma^0(x)|] = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$$

and hence, by Proposition 4.5 of VKA (1999), $W_2 + W_3$ is bounded by

$$\begin{aligned}
&O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})\, n^{-1}\sum_{i=1}^n \{I(T - V < E_i^0 \le T) + I(T < E_i^0 \le T + V)\} \\
&= O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})\, \{[\tilde{H}_\varepsilon^0(T) - \tilde{H}_\varepsilon^0(T - V)] + [\tilde{H}_\varepsilon^0(T + V) - \tilde{H}_\varepsilon^0(T)]\},
\end{aligned}$$

where $\tilde{H}_\varepsilon^0(\cdot)$ is the empirical distribution of E_i^0, $i = 1, \ldots, n$. Using the fact that $\tilde{H}_\varepsilon^0(y) - H_\varepsilon^0(y) = O_P(n^{-1/2})$ uniformly in y, the above term is $o_P(n^{-1/2})$. In the same way,

$$\begin{aligned}
W_1 &= n^{-1}\sum_{i=1}^n I(\Delta_i = 0)I(X_i \le x)\hat{\sigma}^0(X_i)\frac{\int_{\hat{T}_i}^T u\, d\hat{F}_\varepsilon^0(u)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})} I(E_i^0 \le T) + o_P(n^{-1/2}) \\
&= n^{-1}\sum_{i=1}^n I(\Delta_i = 0)I(X_i \le x)\sigma^0(X_i)\frac{\int_{\hat{T}_i}^T u\ d\hat{F}_\varepsilon^0(u)}{1 - F_\varepsilon^0(E_i^{0T})} I(E_i^0 \le T) + o_P(n^{-1/2}) \\
&= n^{-1}\sum_{i=1}^n I(\Delta_i = 0)I(X_i \le x)B_{1i}I(E_i^0 \le T) + o_P(n^{-1/2}),
\end{aligned}$$

where $B_{1i} = -B_{5i}$ in Theorems 3.1 of HVK (2007a). It is treated similarly such that

$$\begin{aligned}
W_1 = n^{-1}\sum_{i=1}^n \frac{T f_\varepsilon^0(T)}{1 - F_\varepsilon^0(E_i^{0T})}\ &[\hat{m}^0(X_i) - m^0(X_i) + T(\hat{\sigma}^0(X_i) - \sigma^0(X_i))] \\
&\times I(\Delta_i = 0, X_i \le x, E_i^0 \le T) + o_P(n^{-1/2}),
\end{aligned}$$

where use is made of asymptotic representations of Propositions 4.8 and 4.9 of VKA (1999). The resulting representation is then added to the asymptotic expression (A.5) with double sums of Theorem 3.1 of HVK (2007a) (adapted to the present problem of integrated regression function) to finally obtain (A.13).

Next, we treat the cases $k = 1$ and 2. Straightforward calculations lead to

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}[(Y_i - \widehat{m_{\theta_0^T}}(X_i))_T^{2*} - (Y_i - m_{\theta_0^T}(X_i))_T^{2*}]I(X_i \le x) \\
&= n^{-2}a_n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} K\Big(\frac{X_i - X_j}{a_n}\Big)\chi_{11}(V_i, Z_j, \Delta_j, \theta_0^T)I(X_i \le x) \\
&\quad + n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big\{2[m^0(X_i) - m_{\theta_0^T}(X_i)]\chi_{20}(V_i, V_j, 1) + \chi_{20}(V_i, V_j, 2)\Big\}\, I(X_i \le x) \\
&\quad + o_P(n^{-1/2}). \qquad (A.14)
\end{aligned}
$$

Replacing θ_0^T by ϑ_{n0}^T in $(Y_i - \widehat{m_{\theta_0^T}}(X_i))_T^{2*}$ simply adds the following term to (A.14)

$$
n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{2\sum_{d=1}^{D_0}\frac{\partial m_{\theta_0^T}(X_i)}{\partial \vartheta_{0d}}\chi_3(V_i, m_{\theta_0^T})\kappa_{0d}(X_j, Z_j, \Delta_j)\}I(X_i \le x) + o_P(n^{-1/2}). \quad (A.15)
$$

When $k = 2$, we have to replace $m_{\theta_0^T}(X_i)$ in (A.14) by $m_T(X_i)$, the limit of a nonparametric estimator $\tilde{m}_T(X_i)$. In this case, if we replace $m_T(X_i)$ by $\tilde{m}_T(X_i)$ in $(Y_i - \widehat{m_T}(X_i))_T^{2*}$, it is easy to check for the nonparametric estimator of HVK (2007d) that, under (A1)-(A10), it only introduces terms which don't affect the structure of the resulting expression (A.14) and therefore defines specific forms for $\chi_{12}(V_i, Z_j, \Delta_j, \theta_0^T)$ and $\chi_{22}(V_i, V_j, \theta_0^T)$. Indeed, if we denote $m_1^T(\cdot)$, the limit of the nonparametric estimator obtained by HVK (2007d), the added terms are

$$
-n^{-2}a_n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} K(\frac{X_i - X_j}{a_n})\Big\{2\chi_3(V_i, m_1^T)f_X^{-1}(X_i)\sigma^0(X_i)
$$

$$\times \ [F_\varepsilon^0(T)\eta(Z_j,\Delta_j|X_i) + \int_{-\infty}^T edF_\varepsilon^0(e)\zeta(Z_j,\Delta_j|X_i)]\Big\} I(X_i \le x)$$

$$-n^{-2}\sum_{i=1}^n\sum_{j=1}^n 2\chi_3(V_i,m_1^T)\Big\{\sigma^0(X_i)\Big\{\gamma_j(T)Th_{\varepsilon 1}^0(T|X_j)$$

$$-\int_{R_X}\int_{-\infty}^{+\infty}\sum_{\delta=0,1}(1/\sigma^0(y))\chi_{20}(y,z,\delta,X_j,Z_j,\Delta_j,1)dH_\delta(z|y)dF_X(y)$$

$$-\int_{-\infty}^{+\infty}\sum_{\delta=0,1}[I(\delta=1,z\le T_{X_j})\gamma_j(z_{X_j}) + (1/\sigma^0(X_j))f_X(X_j)\chi_{10}(X_j,z,\delta,Z_j,\Delta_j,1)$$

$$+I(\delta=0)\gamma_j(\frac{\int_{e_{X_j}^{0T}(z)}^T edF_\varepsilon^0(e)}{1-F_\varepsilon^0(e_{X_j}^{0T}(z))})]dH_\delta(z|X_j)$$

$$-[E_j^0 I(E_j^0\le T,\Delta_j=1) + \frac{\int_{E_j^0\wedge T}^T edF_\varepsilon^0(e)}{1-F_\varepsilon^0(E_j^0\wedge T)}I(\Delta_j=0)] + \int_{-\infty}^T edF_\varepsilon^0(e)\Big\}$$

$$-m^0(X_i)\varphi(X_j,Z_j,\Delta_j,T)\Big\} I(X_i\le x) + o_P(n^{-1/2}), \qquad \text{(A.16)}$$

where $\gamma_j(t) = \eta(Z_j,\Delta_j|X_j) + t\zeta(Z_j,\Delta_j|X_j)$ and $e_j^{0T}(z) = e_{X_j}^{0T}(z)$, $j=1,\ldots,n$.

For the nonparametric estimator of HVK (2007c) (with limit $m_2^T(\cdot)$), the added terms are

$$n^{-2}a_n^{-1}\sum_{i=1}^n\sum_{j=1}^n K(\frac{X_i-X_j}{a_n})\Big\{2\chi_3(V_i,m_2^T)[f_X^{-1}(X_i)(\psi_T^{**}(Z_j,\Delta_j|X_i) - E[\psi_T^{**}(Z,\Delta|X_i)|X_i])$$

$$+B_3(Z_j,\Delta_j|X_i)]\} I(X_i\le x)$$

$$+n^{-2}\sum_{i=1}^n\sum_{j=1}^n\left\{2\chi_3(V_i,m_2^T)\int_{-\infty}^{\infty}\frac{1}{1-F(z_{X_i}^T|X_i)}\left[(\frac{\int_{z_{X_i}^T}^{T_{X_i}} zdF(z|X_i)}{1-F(z_{X_i}^T|X_i)} - z_{X_i}^T)\right.\right.$$

$$\times \ \varphi(X_j,Z_j,\Delta_j,e_i^{0T}(z)) + T_{X_i}\varphi(X_j,Z_j,\Delta_j,T)$$

$$\left.\left.-\sigma^0(X_i)\int_{e_i^{0T}(z)}^T\varphi(X_j,Z_j,\Delta_j,e)de\right]dH_0(z|X_i)\right\} I(X_i\le x)$$

$$+o_P(n^{-1/2}), \qquad \text{(A.17)}$$

where

$$\psi_T^{**}(Z_j,\Delta_j|X_i) = Z_j\Delta_j + \{\frac{\int_{Z_j\wedge T_{X_i}}^{T_{X_i}} ydF(y|X_i)}{1-F(Z_j\wedge T_{X_i}|X_i)}\}(1-\Delta_j),$$

$$B_3(Z_j,\Delta_j|X_i) = E[B_1(Z,\Delta|X)|X=X_i]\eta(Z_j,\Delta_j|X_i)$$

$$+E[B_2(Z,\Delta|X)|X=X_i]\zeta(Z_j,\Delta_j|X_i),$$

$$B_1(Z,\Delta|X=X_i) = f_X^{-1}(X_i)\left\{\frac{I(\Delta=0)}{1-F(Z_{X_i}^T|X_i)}\left[f_\varepsilon^0(E_{X_i}^{0T})(\frac{\int_{Z_{X_i}^T}^{T_{X_i}} ydF(y|X_i)}{1-F(Z_{X_i}^T|X_i)} - Z_{X_i}^T)\right.\right.$$

$$\left.\left.+f_\varepsilon^0(T)T_{X_i} - \sigma^0(X_i)(F(T_{X_i}|X_i) - F(Z_{X_i}^T|X_i))\right]\right\},$$

$$B_2(Z,\Delta|X_i) = f_X^{-1}(X_i)\left\{\frac{I(\Delta=0)}{1-F(Z_{X_i}^T|X_i)}\left[f_\varepsilon^0(E_{X_i}^{0T})E_{X_i}^{0T}(\frac{\int_{Z_{X_i}^T}^{T_{X_i}} ydF(y|X_i)}{1-F(Z_{X_i}^T|X_i)} - Z_{X_i}^T)\right.\right.$$

$$\left.\left.+f_\varepsilon^0(T)TT_{X_i} - \sigma^0(X_i)\int_{Z_{X_i}^T}^{T_{X_i}} ydF(y|X_i)\right]\right\},$$

$Z_{X_i}^T = Z \wedge T_{X_i}$, $E_{X_i}^0 = (Z - m^0(X_i))/\sigma^0(X_i)$ and $E_{X_i}^{0T} = E_{X_i}^0 \wedge T$, $i = 1, \ldots, n$.

Lemma A.2 *Let $\chi(V)$ a general function of $V = (X, Z, \Delta)$ satisfying $E[\chi(V)|X] = 0$, $\sup_{x\in R_X} E[\chi^4(V)|x] < \infty$ and the fact that $|\chi(V)|$ is bounded by a polynom of order 1 in $|Z|$. Also assume that $na_n^3 \to \infty$, $na_n^4 \to 0$, $E[Z^4] < \infty$ and $\sup_{x\in R_X} |f_X(x)| < \infty$. We have*

$$\sup_{x,u} |(1/n)\sum_{i=1}^n \chi(V_i)[I(X_i \le x - ua_n) - I(X_i \le x)]| = o_P(n^{-1/2}),$$

for $u \in [-R, R]$, $R > 0$, $V_1, \ldots, V_n$ a set of i.i.d. r.v. with the same law as V and $x \in R_X = (x_e, x_s)$, a bounded set.

Proof. Let $\chi^+(V) = \max(\chi(V), 0)$ and $\chi^-(V) = -\min(\chi(V), 0)$. It is clear that $\chi(V) = \chi^+(V) - \chi^-(V)$. The expression in Lemma A.2 is bounded by

$$\sup_{x,u} |(1/n)\sum_{i=1}^n \chi^+(V_i)[I(X_i \le x - ua_n) - I(X_i \le x)]$$

$$-E[\chi^+(V)[I(X \leq x - ua_n) - I(X \leq x)]]|$$

$$+\sup_{x,u} |(1/n) \sum_{i=1}^{n} \chi^-(V_i)[I(X_i \leq x - ua_n) - I(X_i \leq x)]$$

$$-E[\chi^-(V)[I(X \leq x - ua_n) - I(X \leq x)]]|, \quad (A.18)$$

where both above terms are treated similarly. Partition R_X into $m = [\frac{L_X}{Ra_n}]$ intervals (x_0, x_1), $(x_1, x_2), \ldots,$ $(x_{l-1}, x_l), \ldots,$ (x_{m-1}, x_m) ($l = 1, \ldots, m$, $x_0 = x_e$ and $x_m = x_s$) of length $C_1 a_n$, where $R \leq C_1 \leq 2R$. Let define intervals $I_\alpha = (x_{\alpha-1}, x_{\alpha+1})$, for $\alpha = 1, \ldots, m-1$. Since the distance between x and $x - ua_n$ is smaller than Ra_n, there always exists an α such that $x, x - ua_n \in I_\alpha$. Partition each I_α by a grid $x_{\alpha,\beta} = x_\alpha + \beta \frac{C_1 a_n}{p_n}$, $p_n = [a_n^{1-\delta_1} n^{1/2} + 1]$, $\beta = -p_n, \ldots, p_n$ and $\delta_1 > 0$. Therefore, the first term of (A.18) is bounded by

$$\sup_{|t-s| \leq Ra_n,\ t,s \in R_X} |(1/n) \sum_{i=1}^{n} \chi^+(V_i)[I(X_i \leq t) - I(X_i \leq s)]$$

$$-E[\chi^+(V)[I(X \leq t) - I(X \leq s)]]|$$

$$\leq \max_{\alpha,\beta,\zeta} |(1/n) \sum_{i=1}^{n} \chi^+(V_i)[I(X_i \leq x_{\alpha,\beta}) - I(X_i \leq x_{\alpha,\zeta})]$$

$$-E[\chi^+(V)[I(X \leq x_{\alpha,\beta}) - I(X \leq x_{\alpha,\zeta})]]|$$

$$+4 \max_{1 \leq \alpha \leq m-1,\ -p_n \leq \beta \leq p_{n-1}} |E[\chi^+(V)[I(X \leq x_{\alpha,\beta+1}) - I(X \leq x_{\alpha,\beta})]]|,$$

due to the monotonicity of the functions $\chi^+(V)I(X \leq x)$ and $E[\chi^+(V)I(X \leq x)]$ with respect to x. The second term of the above expression is $O(n^{-1/2} a_n^{\delta_1})$. Next, denote, for

$C_2 > 0$,

$$G_{1n\alpha\beta\zeta\delta_2} = |(1/n)\sum_{i=1}^{n}\chi^+(V_i)I(\chi^+(V_i) > C_2 n^{\delta_2})[I(X_i \leq x_{\alpha,\beta}) - I(X_i \leq x_{\alpha,\zeta})]$$

$$-E[\chi^+(V)I(\chi^+(V) > C_2 n^{\delta_2})[I(X \leq x_{\alpha,\beta}) - I(X \leq x_{\alpha,\zeta})]]|$$

and

$$G_{2n\alpha\beta\zeta\delta_2} = |(1/n)\sum_{i=1}^{n}\chi^+(V_i)I(\chi^+(V_i) \leq C_2 n^{\delta_2})[I(X_i \leq x_{\alpha,\beta}) - I(X_i \leq x_{\alpha,\zeta})]$$

$$-E[\chi^+(V)I(\chi^+(V) \leq C_2 n^{\delta_2})[I(X \leq x_{\alpha,\beta}) - I(X \leq x_{\alpha,\zeta})]]|,$$

for some $\delta_2 > 0$. By Chebichev inequality, we have, for some $C_3 > 0$,

$$P(|G_{1n\alpha\beta\zeta\delta_2}| > C_3 n^{-1/2} a_n^{\delta_1}) \leq \frac{E[\omega^2_{1n\alpha\beta\zeta\delta_2}(V)]}{C_3^2 a_n^{2\delta_1}},$$

where

$$\omega_{1n\alpha\beta\zeta\delta_2}(V) = \chi^+(V)I(\chi^+(V) > C_2 n^{\delta_2})[I(X \leq x_{\alpha,\beta}) - I(X \leq x_{\alpha,\zeta})]$$

and by Bernstein inequality,

$$P(|G_{2n\alpha\beta\zeta\delta_2}| > C_4 n^{-1/2} a_n^{\delta_1}) \leq 2\exp(-C_4^2 \nu_{n\alpha\beta\zeta\delta_1\delta_2}),$$

for some $C_4 > 0$ and where

$$\nu_{n\alpha\beta\zeta\delta_1\delta_2} = \frac{a_n^{2\delta_1}}{2Var[\omega_{2n\alpha\beta\zeta\delta_2}(V)] + (2/3)n^{-1/2+\delta_2} a_n^{\delta_1}}$$

and

$$\omega_{2n\alpha\beta\zeta\delta_2}(V) = \chi^+(V)I(\chi^+(V) \leq C_2 n^{\delta_2})[I(X \leq x_{\alpha,\beta}) - I(X \leq x_{\alpha,\zeta})].$$

We thus have, for some $C_5 > 0$,

$$P(\max_{\alpha,\beta,\zeta} |G_{1n\alpha\beta\zeta\delta_2}| + \max_{\alpha,\beta,\zeta} |G_{2n\alpha\beta\zeta\delta_2}| > C_5 n^{-1/2} a_n^{\delta_1})$$
$$\leq \sum_{\alpha=1}^{m-1} \sum_{\beta=-p_n}^{p_n} \sum_{\zeta=-p_n}^{p_n} \left\{ \frac{4E[\omega^2_{1n\alpha\beta\zeta\delta_2}(V)]}{C_5^2 a_n^{2\delta_1}} + 2\exp(-(C_5^2/4)\nu_{n\alpha\beta\zeta\delta_1\delta_2}) \right\}. \quad \text{(A.19)}$$

The first term on the right hand side of (A.19) is bounded by

$$C_7 a_n^{1/2-2\delta_1} (E[I(|Z| > C_6 n^{\delta_2})])^{1/2} \leq C_7 n^{-1/8+(2\delta_1/3)} (1 - H(C_6 n^{\delta_2}) + H(-C_6 n^{\delta_2}))^{1/2},$$

for some $C_6, C_7 > 0$. Using the fact that $E[|Z^4|] < \infty$, it is easy to check that $1 - H(C_6 n^{\delta_2})$ and $H(-C_6 n^{\delta_2})$ are $O(n^{-4\delta_2})$. Therefore, the first term on the right hand side of (A.19) is $O(n^{-1/8+(2\delta_1/3)-2\delta_2})$. In the second term of (A.19),

$$\nu_{n\alpha\beta\zeta\delta_1\delta_2} \geq \frac{n^{-2\delta_1/3}}{O(n^{-1/4}) + O(n^{-1/2+\delta_2-\delta_1/4})}.$$

The number of terms in (A.19) is $O(n^{3/4+(2\delta_1/3)})$ such that that $\delta_2 = 1/2 - \delta_1/2$ and $\delta_1 < 9/56$ is a choice such that the term on the right hand side of (A.19) tends to zero.

Lemma A.3 *Let $\chi(V)$ a positive function of $V = (X, Z, \Delta)$ satisfying $\sup_x E[\chi(V)|x] < \infty$ and the fact that $\chi(V)$ is bounded by a polynom of order 2 in $|Z|$. Also assume that $\sup_x |f_X(x)| < \infty$, $\inf_x |f(x)| > 0$, $\sup_{x,z} |f_{X|Z,\Delta}(x|z,\delta)| < \infty$ and $E[|Z|^{4(1+\upsilon)}] < \infty$ for some $\upsilon > 0$. We have*

$$P(|(1/n)\sum_{i=1}^{n} \chi(V_i) I(x < X_i \leq x + d)| > C_1 n^{-1/2-\nu_1}) = c_n n^{-1/2-\nu_1},$$

where $d = C_2 n^{-1/2-\nu_1}$ for some $C_1, C_2, \nu_1 > 0$, $V_1, \ldots, V_n$ is a set of i.i.d. r.v. with the same law as V and c_n independent of x tends to zero when $n \to \infty$.

Proof. Write, for some $C_3, \nu_2 > 0$,

$$
\begin{aligned}
&(1/n)\sum_{i=1}^{n}\chi(V_i)I(x < X_i \le x+d)\\
&= (1/n)\sum_{i=1}^{n}\chi(V_i)I(\chi(V_i) > C_3 n^{\nu_2})I(x < X_i \le x+d)\\
&= (1/n)\sum_{i=1}^{n}\{\chi(V_i)I(\chi(V_i) \le C_3 n^{\nu_2})I(x < X_i \le x+d)\\
&\qquad\qquad -E[\chi(V)I(\chi(V) \le C_3 n^{\nu_2})I(x < X \le x+d)]\}\\
&\quad +E[\chi(V)I(\chi(V) \le C_3 n^{\nu_2})I(x < X \le x+d)]\\
&= R_{1nd\nu_2}(x) + R_{2nd\nu_2}(x) + R_{3d\nu_2}(x).
\end{aligned}
$$

$R_{3d\nu_2}(x)$ is clearly bounded by $C_4 n^{-1/2-\nu_1}$ for some $C_4 > 0$ since $\sup_x E[\chi(V)|x] < \infty$ and $\sup_x f_X(x) < \infty$. For $R_{1nd\nu_2}(x)$, we use Markov inequality.

$$
\begin{aligned}
&P(|R_{1nd\nu_2}(x)| > (C_1/3)n^{-1/2-\nu_1})\\
&\le \frac{3n^{1/2+\nu_1}}{C_1}E[I(x < X \le x+d)\chi(V)I(\chi(V) > C_3 n^{\nu_2})]\\
&\le \frac{3n^{1/2+\nu_1}}{C_1}E[I(x < X \le x+d)E[\chi(V)I(Z^2 > C_5 n^{\nu_2})|X]], \qquad \text{(A.20)}
\end{aligned}
$$

for some $C_5 > 0$. Since $\inf_x |f(x)| > 0$, $\sup_{x,z} |f_{X|Z,\Delta}(x|z,\delta)| < \infty$ and $E[|Z|^{4(1+\upsilon)}] < \infty$,

$$
\begin{aligned}
E[\chi(V)I(Z^2 > C_5 n^{\nu_2})|X] &\le C_6[1 - H(C_5^{1/2}n^{\nu_2/2}) + H(-C_5^{1/2}n^{\nu_2/2})]^{1/2}\\
&\le C_7 n^{-\nu_2(1+\upsilon)},
\end{aligned}
$$

for some constants $C_6, C_7 > 0$. Therefore, the term on the right hand side of the inequality (A.20) is bounded by $C_8 n^{-\nu_2(1+\upsilon)}$, where $C_8 = 3\sup_x |f(x)| C_7 C_2/C_1$. Next, we use

Bernstein inequality for $R_{2nd\nu_2}(x)$.

$$P(|R_{2nd\nu_2}(x)| > (C_1/3)n^{-1/2-\nu_1}) \leq 2\exp(-\nu_{n\nu_1\nu_2 x}),$$

where

$$\nu_{n\nu_1\nu_2 x} = \frac{(C_1^2/9)n^{-2\nu_1}}{2Var[\omega_{n\nu_1\nu_2 x}(V)] + (2/9)C_1C_3n^{-1/2-\nu_1+\nu_2}}$$

and

$$\omega_{n\nu_1\nu_2 x}(V) = \chi(V)I(\chi(V) \leq C_3n^{\nu_2})I(x < X \leq x+d).$$

Therefore, for well-chosen ν_1 and υ, we can always find ν_2 such that $(1/2+\nu_1)/(1+\upsilon) < \nu_2 < 1/2 - \nu_1$. This finishes the proof.

Acknowledgements. Thanks to Gema lvarez-Llorente, M. Soledad Otero-Girldez, and J. de Ua-lvarez (University of Vigo, Spain) for providing the Galician unemployment data.

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.

Chamberlain, G. (1986). Efficiency in semiparametric models with censoring. *Journal of Econometrics*, **32**, 189–218.

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

Heuchenne, C. and Van Keilegom, I. (2007a). 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. (2007b). Nonlinear regression with censored data. *Technometrics*, **49**, 34–44.

Heuchenne, C. and Van Keilegom, I. (2007c). Location estimation in nonparametric regression with censored data. *Journal of Multivariate Analysis*, **98**, 1558–1582.

Heuchenne, C. and Van Keilegom, I. (2007d). Estimation in nonparametric location-scale regression models with censored data. To appear in *Annals of the Institute of Statistical Mathematics*

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

Kiefer, N.M. (1988). Economic duration data and hazard functions. *Journal of Economic Literature*, **26**, 646–679.

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

Lewbel, A. and Linton, O. (2002). Nonparametric Censored and Truncated Regression. *Econometrica*, **70**, 765–779.

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 Gonzlez-Manteiga, W. (2007). Goodness-of-fit tests for parametric models in censored regression. To appear in *Canadian Journal of Statistics*

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.