



## A new lack-of-fit test for quantile regression with censored data

M. Conde-Amboage, I. Van Keilegom and W. González-Manteiga

Version: Author's Accepted Manuscript

This is a post-peer-review, pre-copyedit version of an article published in Scandinavian Journal of Statistics. The final authenticated version is available online at <https://doi.org/10.1111/sjos.12512>.

### HOW TO CITE

Conde-Amboage, M., Van Keilegom, I., & González-Manteiga, W. (2021). A new lack-of-fit test for quantile regression with censored data. *Scandinavian Journal of Statistics*, 48(2), 655-688. DOI <https://doi.org/10.1111/sjos.12512>.

### FUNDING

Research has been funded by project INNPAR2D (reference code MTM2016-76969-P), from the Ministry of Economy and Competitiveness and the European Regional Development Fund (ERDF) and Grant Number: 694409 from the H2020 European Research Council.

---

## ORIGINAL ARTICLE

Field of the paper

# A new lack-of-fit test for quantile regression with censored data

Mercedes Conde-Amboage<sup>1,2</sup> | Ingrid Van Keilegom<sup>2</sup> |  
Wenceslao González-Manteiga<sup>1</sup>

<sup>1</sup>Models of Optimization, Decision, Statistics and Applications Research Group (MODESTYA), Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Spain.

<sup>2</sup>Research Centre for Operations Research and Statistics (ORSTAT), KU Leuven, Belgium.

### Correspondence

Ingrid Van Keilegom, Research Centre for Operations Research and Statistics (ORSTAT), KU Leuven, Belgium.  
Email: [ingrid.vankeilegom@kuleuven.be](mailto:ingrid.vankeilegom@kuleuven.be)

### Funding information

Project MTM2016-76969-P (Spanish State Research Agency, AEI) co-funded by the European Regional Development Fund (ERDF).

Post-doctoral grant from the Ministry of Culture, Education and University Planning and the Ministry of Economy, Employment and Industry of the Galician Government. European Research Council (2016-2021, Horizon 2020/ERC grant No. 694409).

A new lack-of-fit test for quantile regression models will be presented for the case where the response variable is right-censored. The test is based on the cumulative sum of residuals, and it extends the ideas of He and Zhu (2003) to censored quantile regression. It will be shown that the empirical process associated with the test statistic converges to a Gaussian process under the null hypothesis and is consistent. To approximate the critical values of the test, a bootstrap mechanism will be used. A simulation study will be carried out to study the performance of the new test in comparison with other tests available in the literature. Finally, a real data application will be presented to show the good properties of the new lack-of-fit test in practice.

**Keywords** — bootstrap calibration, censored data, lack-of-fit tests, quantile regression.

## 1 | INTRODUCTION

Although mean regression is still a traditional benchmark in regression studies, the quantile approach is receiving increasing attention, because it allows a more complete description of the conditional distribution of the response given the covariate, and it is more robust to deviations from error normality. That is, while classical regression gives only information on the conditional expectation, quantile regression extends the viewpoint on the whole conditional

distribution of the response variable. Moreover, this kind of methods allows us to deal with regression scenarios with complex data (such as censored data), in many cases, under better conditions than a classical mean regression. For all these reasons, quantile regression is a very useful statistical technology for a large diversity of disciplines.

We consider the following parametric quantile regression model:

$$Y = q_\tau(X, \theta_\tau) + \varepsilon, \quad (1)$$

where  $Y$  is the survival time (or some transformation of it),  $X$  is a  $d$ -dimensional vector of covariates with compact support  $R_X$ ,  $q_\tau$  represents the conditional quantile function of  $Y$  given  $X$ , which is known up to a finite-dimensional parameter  $\theta_\tau \in \mathbb{R}^q$ , and the error  $\varepsilon$  verifies  $\mathbb{P}(\varepsilon \leq 0|X) = \tau$ , for some specified quantile of interest  $0 < \tau < 1$ . Due to the presence of random right censoring, we do not always observe  $Y$ . Instead we observe  $(Z, \delta)$ , where  $Z = \min\{Y, C\}$ ,  $\delta = \mathbb{I}(Y \leq C)$ , the random variable  $C$  represents the censoring time, and  $\mathbb{I}$  is the indicator function. Finally, consider a sample of independent and identically distributed (i.i.d.) observations  $(Z_i, \delta_i, X_i)$  for  $i = 1, \dots, n$ , having the same distribution as  $(Z, \delta, X)$ .

In the literature, we can find several proposals to estimate the parameter  $\theta$  under random censoring, see for instance, Powell (1986), Portnoy (2003), Peng and Huang (2008) or Leng and Tong (2013) (extending Ying et al. (1995)'s median regression model) based on parametric ideas, De Backer et al. (2017) based on semiparametric ideas, or Gannoun et al. (2005), Wang and Wang (2009), De Backer et al. (2019) and De Backer et al. (2020) based on nonparametric ideas. Even in a high-dimensional scenario, there exist estimation proposals such as Shows et al. (2010) and Wu and Yin (2015).

In this paper we will focus on testing a parametric censored regression model versus nonparametric alternatives. The main goal will be to realize the following lack-of-fit test:

$$H_0 : q_\tau \in Q_\theta = \{q_\tau(\cdot, \theta) : \theta \in \Theta\}, \quad (2)$$

for some  $\Theta \subset \mathbb{R}^q$  compact, versus a nonparametric alternative. Several authors have addressed the problem of testing a parametric quantile regression model with complete data, such as He and Zhu (2003), Zheng (1998), Conde-Amboage et al. (2015), Horowitz and Spokoiny (2002), Birke et al. (2017), Maistre et al. (2017) or Wang et al. (2018), among others. However, the literature for the case where the response is subject to random right censoring is much scarcer. To the best of our knowledge, there is only the paper of Wang (2008), in which a nonparametric approach is proposed to test the lack-of-fit of a quantile regression model under random right censoring. The test is an adaptation of the previous test proposed by Zheng (1998) to censored data.

In the mean regression context with censored data, lack-of-fit tests for the regression function have been developed by several authors, such as Lopez and Patilea (2009) based on kernel smoothing ideas (extending the ideas of Zheng (1996)), Stute et al. (2000) based on empirical processes marked by residuals (extending the ideas of Stute (1997)), Pardo-Fernández et al. (2007) based on the comparison between the error distribution of parametric and nonparametric residuals, or Peng and Taylor (2017) in the context of mixture cure models.

Taking into account the state of the art, we propose and study two lack-of-fit tests for parametric quantile regression models with censored data based on the integrated regression function introduced by Stute (1997) to avoid the use of bandwidth parameters in the definition of the test statistic. In Section 2 we develop a characterization of the null hypothesis, and we present the two new test statistics. The asymptotic properties of these tests are presented in Section 3, together with a bootstrap method to approximate the critical values of the tests. Section 4 contains a simulation study, in which the performance of the test is studied under homo- and heteroscedastic models, with

different error distributions and different quantile levels of interest. We also compare the proposed test with the test proposed by Wang (2008). A real data application is presented in Section 5 to illustrate how the tests can be used in practice. Section 6 contains some conclusions and possible extensions of this work. Finally, the Appendix contains the proofs of the asymptotic results.

## 2 | THE NEW TESTS

As mentioned before, our main goal is to test the hypothesis:

$$\begin{cases} H_0 : q_\tau \in Q_\theta = \{q_\tau(\cdot, \theta) : \theta \in \Theta \subset \mathbb{R}^q\} \\ H_a : q_\tau \notin Q_\theta. \end{cases}$$

Let us denote by  $\theta_\tau$  the true but unknown parameter vector  $\theta$  under the null hypothesis. Denote  $F_A$  and  $f_A$  for the cumulative distribution function and density function of a random vector  $A$  (for instance  $Y, C, \varepsilon, (X, Y)$  or conditional variables such as  $\varepsilon|X$ ), respectively. Finally, for vectors  $x, t \in \mathbb{R}^d$ , we use the notation  $x \leq t$  to indicate that each component of  $x$  is less than or equal to the corresponding component of  $t$ .

In what follows, we assume that the functions in  $Q_\theta$  are differentiable with respect to the components of  $\theta$ , and we let

$$q_\tau^{(1)}(x, \theta) = \frac{\partial}{\partial \theta} q_\tau(x, \theta) = \left( \frac{\partial}{\partial \theta_1} q_\tau(x, \theta), \dots, \frac{\partial}{\partial \theta_q} q_\tau(x, \theta) \right)'$$

for any  $\theta \in \Theta$ . Also assume that  $\mathbb{P}(q_\tau^{(1)}(X, \theta) = 0) = 0$  for any  $\theta \in \Theta$ .

In order to present the new lack-of-fit tests the following characterization of the null hypothesis will be crucial:

**Lemma 1** (*Characterization of the null hypothesis*) *The null hypothesis  $H_0 : q_\tau \in Q_\theta$  holds if and only if for some  $\theta_\tau \in \Theta \subset \mathbb{R}^q$ , and for any  $t \in \mathbb{R}$ ,*

$$\int \psi_\tau(y - q_\tau(x, \theta_\tau)) q_\tau^{(1)}(x, \theta_\tau) \mathbb{1}(x \leq t) dF_{X,Y}(x, y) = 0, \quad (3)$$

where  $\psi_\tau(r) = \tau \mathbb{1}(r > 0) + (\tau - 1) \mathbb{1}(r < 0)$ .

Note that it is possible to establish a different characterization of the null hypothesis without including the derivative of the quantile regression function, following the original idea of Stute (1997). That is, the null hypothesis  $H_0 : q_\tau \in Q_\theta$  holds if and only if for some  $\theta_\tau \in \Theta \subset \mathbb{R}^q$ , and for any  $t \in \mathbb{R}$ ,

$$\int \psi_\tau(y - q_\tau(x, \theta_\tau)) \mathbb{1}(x \leq t) dF_{X,Y}(x, y) = 0.$$

This can be easily derived following the arguments in the proof of Lemma 1.

Based on Lemma 1, we will consider consistent estimators of (3) to define the test statistics. If the true parameter  $\theta_\tau$  were known, following the ideas of Stute et al. (2000) (for the mean regression context) and He and Zhu (2003) (for the quantile regression context), the tests could be based on the processes

$$R_{n,1}^0(t) = n^{1/2} \sum_{i=1}^n W_i^{\text{KM}} \psi_\tau(Z_i - q_\tau(X_i, \theta_\tau)) q_\tau^{(1)}(X_i, \theta_\tau) \mathbb{1}(X_i \leq t)$$

$$R_{n,2}^0(t) = n^{1/2} \sum_{i=1}^n W_i^{\text{KM}} \psi_{\tau}(Z_i - q_{\tau}(X_i, \theta_{\tau})) \mathbb{I}(X_i \leq t), \quad (4)$$

where  $W_i^{\text{KM}}$  denote the Kaplan-Meier weights (Kaplan and Meier (1958)) that are given by

$$W_i^{\text{KM}} = \frac{\delta_{[i]}}{n-i+1} \prod_{j=1}^{i-1} \left( \frac{n-j}{n-j+1} \right)^{\delta_{[j]}}.$$

Here,  $\delta_{[i]}$  represents the censoring indicator associated with  $Z_{[i]}$ , where  $Z_{[1]} \leq Z_{[2]} \leq \dots \leq Z_{[n]}$  are the order statistics. For notational simplicity, we will assume that  $Z_{[i]} = Z_i$ ,  $X_{[i]} = X_i$  and  $\delta_{[i]} = \delta_i$ .

He and Zhu (2003) established that in regression problems, residuals are just one of the components in the gradient equation and by using all components in the gradient equation, the procedure can be made more sensitive to departures from the model. In order to test his statement we will consider the two possible empirical processes  $R_{n,1}^0$  and  $R_{n,2}^0$ .

Since the true  $\theta_{\tau}$  is not known, we need to replace  $\theta_{\tau}$  in (4) by an appropriate estimator. We will adapt the ideas of Stute (1999) to the quantile regression setup, and will consider the following estimator:

$$\widehat{\theta}_{\tau} = \arg \min_{\theta_{\tau} \in \mathbb{R}^q} \sum_{i=1}^n W_i^{\text{KM}} \rho_{\tau}(Z_i - q_{\tau}(X_i, \theta_{\tau})), \quad (5)$$

where  $\rho_{\tau}(u) = u(\tau - \mathbb{I}(u < 0))$  is the well-known quantile loss function.

This leads to the following processes useful for lack-of-fit testing of the parametric model:

$$\begin{aligned} R_{n,1}^1(t) &= n^{1/2} \sum_{i=1}^n W_i^{\text{KM}} \psi_{\tau}(Z_i - q_{\tau}(X_i, \widehat{\theta}_{\tau})) q_{\tau}^{(1)}(X_i, \widehat{\theta}_{\tau}) \mathbb{I}(X_i \leq t) \\ R_{n,2}^1(t) &= n^{1/2} \sum_{i=1}^n W_i^{\text{KM}} \psi_{\tau}(Z_i - q_{\tau}(X_i, \widehat{\theta}_{\tau})) \mathbb{I}(X_i \leq t). \end{aligned} \quad (6)$$

Then, we could define our test statistics using any continuous functional of the empirical processes defined in (6), such as a supremum or a Cramer-von Mises norm. Following the ideas used by several authors in the literature, see for instance Stute (1997), He and Zhu (2003) or Stute et al. (2000), our test statistics are defined as follows:

$$\begin{aligned} T_{n,1} &= \text{largest eigenvalue of } \frac{1}{n} \sum_{i=1}^n R_{n,1}^1(X_i) [R_{n,1}^1(X_i)]' \\ T_{n,2} &= \frac{1}{n} \sum_{i=1}^n R_{n,2}^1(X_i)^2 \end{aligned} \quad (7)$$

### 3 | ASYMPTOTIC PROPERTIES

In this section the asymptotic behaviour of the new tests will be studied under the null and the alternative hypothesis. We focus on the empirical processes  $R_{n,1}^k$  with  $k = 0, 1$ , because the analysis of  $R_{n,2}^k$  is analogous. Let us introduce the following notation:

$$H(y) = \mathbb{P}(Z \leq y)$$

$$\begin{aligned}\tilde{H}^0(y) &= \mathbb{P}(Z \leq y, \delta = 0) = \int_{-\infty}^y (1 - F(u))G(du) \\ \tilde{H}^{11}(x, y) &= \mathbb{P}(X \leq x, Z \leq y, \delta = 1) \\ \gamma_0(y) &= \exp\left(\int_{-\infty}^{y-} \frac{\tilde{H}^0(dz)}{1 - H(z)}\right).\end{aligned}$$

Moreover, for any real-valued function  $\varphi$  defined on  $\mathbb{R}^{d+1}$ , consider

$$\begin{aligned}\gamma_1^\varphi(y) &= \frac{1}{1 - H(y)} \int \mathbb{I}(y \leq w) \varphi(x, w) \gamma_0(w) \tilde{H}^{11}(dx, dw), \\ \gamma_2^\varphi(y) &= \int \int \frac{\mathbb{I}(v < y, v < w) \varphi(x, w) \gamma_0(w)}{(1 - H(v))^2} \tilde{H}^0(dv) \tilde{H}^{11}(dx, dw), \quad \text{and} \\ \eta^\varphi(X_i, Z_i, \delta_i) &= \varphi(X_i, Z_i) \gamma_0(Z_i) \delta_i + \gamma_1^\varphi(Z_i) (1 - \delta_i) - \gamma_2^\varphi(Z_i).\end{aligned}$$

In our particular case, we consider the function

$$\varphi_{x_0}(x, y; \theta) = \psi_\tau(y - q_\tau(x, \theta)) q_\tau^{(1)}(x, \theta) \mathbb{I}(x \leq x_0)$$

and we use the notation  $\eta(X_i, Z_i, \delta_i; x_0, \theta) \equiv \eta^{\varphi_{x_0}(\cdot; \theta)}(X_i, Z_i, \delta_i)$ .

### 3.1 | Behaviour of the test under the simple null hypothesis

We start by showing the convergence of the test statistic  $T_{n,1}$  under the simple null hypothesis where  $\theta_\tau$  is known. In a first step, we need to obtain the limit of the process  $R_{n,1}^0$ , for which we need to introduce the following conditions:

**(A1)** Assume that

- (i)  $Y$  and  $C$  are independent.
- (ii)  $\mathbb{P}(\delta = 1|X, Y) = \mathbb{P}(\delta = 1|Y)$ , that is, the probability of having a non-censored observation only depends on the response variable  $Y$ .
- (iii) The upper bound of the support of  $Y$  is strictly smaller than the upper bound of the support of  $C$ . Moreover, the distribution functions of  $Y$  and  $C$  are continuous.
- (iv) The conditional distribution of the error,  $F_{\varepsilon|X}(y|x)$ , is continuously differentiable with respect to  $y$  uniformly in  $x$  and for all  $x \in R_X$ ,  $f_{\varepsilon|X=x}(0) > 0$ .

**(A2)**  $\mathbb{E}[\gamma_0(Z)^2 \delta] < \infty$ .

**(A3)**  $\int \frac{1}{(1-H(y))^3} d\tilde{H}^0(y) < \infty$ .

**(A4)** All partial derivatives of  $q_\tau(x, \theta)$  with respect to the components of  $\theta$  of order 0, 1 and 2 exist and are continuous in  $(x, \theta)$  for all  $x$  and  $\theta$ , and  $\mathbb{P}(q_\tau^{(1)}(X, \theta) = 0) = 0$  for all  $\theta \in \Theta$ . Moreover,  $q_\tau^{(1)}(x, \theta)$  is Lipschitz in  $\theta$  uniformly in  $x$ , and

$$\sup_{x, \theta} \left\| q_\tau^{(1)}(x, \theta) \right\| < \infty,$$

where  $\|\cdot\|$  denotes the Euclidean norm.

Assumptions (A1)(i) - (A1)(iii) are quite common in the survival analysis literature when covariates are present, because they are needed to get the convergence of the Kaplan-Meier integrals involved in the definition of the empirical processes defined in (6). This kind of assumptions has been introduced by Stute (1993) and used by several authors such as Stute et al. (2000), Lopez and Patilea (2009) or Sánchez-Sellero et al. (2005), among others. Note that conditions (A1)(i) and (A1)(iii) are identification conditions needed when working with Kaplan-Meier estimators. Moreover, assumption (A1)(ii) allows some kind of dependency between  $C$  and  $X$ , and Lopez and Patilea (2009) pointed out that this condition is suitable for randomized clinical trials. Finally, assumption (A1)(iv) about the conditional error distribution is essential in the quantile regression setup (see for instance, He and Zhu (2003), Maistre et al. (2017) or Conde-Amboage et al. (2015)), because it guarantees that there exist data points around the conditional quantile that we want to estimate. Conditions (A2) and (A3) allow us to obtain the IID representation of the empirical process  $R_{n,1}^0$  given in Theorem 2. More specifically, these assumption are needed to apply Theorem 3.11 in Sánchez-Sellero (2001), which shows the uniform representation of Kaplan-Meier integrals. Moreover, these assumptions are also used by Stute et al. (2000). Finally, condition (A4) impose smoothness conditions on the quantile regression function. For instance, the assumption related to  $q_\tau^{(1)}$  is used to obtain the asymptotic behaviour of the empirical process  $R_{n,1}^{0,1}$  (similarly to He and Zhu (2003)) while the conditions about  $q_\tau^{(2)}$  are useful to get the IID representation of the estimator (5) (similarly to Stute et al. (2000)).

We start by presenting an iid representation for the empirical process  $R_{n,1}^0$ . Note that all results presented in this section can be easily adapted to the empirical process  $R_{n,2}^0$  by considering

$$\tilde{\varphi}_{x_0}(x, y; \theta) = \psi_\tau(y - q_\tau(x, \theta)) \mathbb{I}(x \leq x_0).$$

**Theorem 2 (IID representation of  $R_{n,1}^0$ )** *Under  $H_0$ , if conditions (A1) - (A4) are verified, then the empirical process  $R_{n,1}^0$  admits the following representation:*

$$R_{n,1}^0(x_0) = n^{-1/2} \sum_{i=1}^n \eta(X_i, Z_i, \delta_i; x_0, \theta_\tau) + Q_n(x_0), \quad (8)$$

where

$$\sup_{x_0 \in R_X} |Q_n(x_0)| = O\left(\frac{(\log n)^3}{\sqrt{n}}\right) \text{ almost surely.}$$

**Remark 1.** Note that Theorem 2 can be easily extended to hold uniformly in  $\theta \in \Theta$ . This is because it can be seen from the proof of Theorem 2 that the remainder term is negligible uniformly in  $\theta$ , thanks to assumption (A4). This will be useful in order to study the asymptotic behaviour of the estimator introduced in (5).

Theorem 2 will be crucial to obtain the asymptotic limit under  $H_0$  of the process  $R_{n,1}^0$  in the space  $\ell^\infty(R_X)$  of bounded functions defined on  $R_X$ . We also establish the limiting distribution of the test statistic  $T_{n,1}$  when evaluated at the true  $\theta_\tau$ . We denote this test statistic by  $T_{n,1}^0$ .

**Corollary 3 (Simple null hypothesis)** *Under  $H_0$ , if conditions (A1) - (A4) are verified, then the process  $R_{n,1}^0$  converges weakly in  $\ell^\infty(R_X)$  to a centred Gaussian process  $R_\infty^0$  with covariance function*

$$\text{Cov}(R_\infty^0(x_0), R_\infty^0(x_1)) = \mathbb{E}[\eta(X, Z, \delta; x_0, \theta_\tau) \eta(X, Z, \delta; x_1, \theta_\tau)].$$

Moreover, the test statistic  $T_{n,1}^0$  converges to the largest eigenvalue of the Cramer-von Mises norm  $\mathbb{E}[\mathcal{R}_\infty^0(X)\mathcal{R}_\infty^0(X)']$  of the process  $\mathcal{R}_\infty^0$ .

Next, we will study the behaviour of the test statistic  $T_{n,1}$  under a composite null hypothesis and under local alternatives of the form

$$H_{1n} : q_\tau(\cdot) = q_\tau(\cdot, \theta_\tau) + \frac{1}{\sqrt{n}} h(\cdot),$$

where  $\theta_\tau$  is the true value under  $H_0$ , and  $h(\cdot)$  is a real-valued function with  $\mathbb{E}|h(X)| < \infty$ . Since these hypotheses involve the estimator  $\hat{\theta}_\tau$  given in (5), we first need to study the asymptotic behavior of the latter estimator.

### 3.2 | Behaviour of the estimator $\hat{\theta}_\tau$

We will need the following notation:

$$\bar{\theta}_\tau = \arg \min_{\theta} \bar{S}(\theta), \quad (9)$$

where  $\bar{S}(\theta) = \mathbb{E}_{H_{1n}}[\rho_\tau(Y - q_\tau(X, \theta))]$  under  $H_{1n}$ , and we will use the following decomposition to study the estimator  $\hat{\theta}_\tau$ :

$$\hat{\theta}_\tau - \theta_\tau = [\hat{\theta}_\tau - \bar{\theta}_\tau] + [\bar{\theta}_\tau - \theta_\tau],$$

which is the typical decomposition in a first term contributing to the variance of the estimator, and a second term leading to the bias of the estimator.

For the asymptotic study of these two terms, we need the following additional conditions:

(A5) For all  $\epsilon > 0$ ,

$$\sup_{\theta: \|\theta - \theta_\tau\| \geq \epsilon} \mathbb{E}[\rho_\tau(Y - q_\tau(X, \theta))] > \mathbb{E}[\rho_\tau(Y - q_\tau(X, \theta_\tau))]$$

and  $\mathbb{E}[\sup_{\theta \in \Theta} |Y - q_\tau(X, \theta)|] < \infty$ .

(A6) The matrix  $\Omega = \mathbb{E}[f_{\epsilon|X}(0) q_\tau^{(1)}(X, \theta_\tau) q_\tau^{(1)}(X, \theta_\tau)']$  is non-singular. Moreover, all components of the vector  $\mathbb{E}[f_{\epsilon|X}(0) h(X) q_\tau^{(1)}(X, \theta_\tau)]$  and of the matrix  $\mathbb{E}[f_{\epsilon|X}(0) h(X) q_\tau^{(2)}(X, \theta_\tau)]$  are finite, where  $q_\tau^{(2)}(X, \theta) = \frac{\partial^2}{\partial \theta \partial \theta'} q_\tau(X, \theta)$ .

(A7)  $\mathbb{E}[\gamma_0^4(Z)\delta] < \infty$  and  $\mathbb{E}[H^2(Z)/(1-H(Z))^2\gamma_0^4(Z)\delta] < \infty$ .

(A8)  $F_{Z|X=x}(z)$  is Lipschitz in  $z$  uniformly in  $x$ .

Assumptions (A5)-(A8) are needed to obtain the asymptotic behaviour of the proposed estimator (5) given in Theorem 4. The IID representation given in Theorem 4 is used to obtain the representation of the empirical process  $\mathcal{R}_{n,1}^1$  under local alternatives (see Theorem 6). Note that in order to obtain the asymptotic distribution of the test statistics we could use other estimators, with a different IID representation, for which these conditions could change. In fact, there are other estimators in the framework of censored quantile regression, see e.g. Portnoy (2003), Peng and Huang (2008) or Wang and Wang (2009).

We are now ready to state an asymptotic representation for  $\widehat{\theta}_\tau - \theta_\tau$ .

**Theorem 4 (IID representation of  $\widehat{\theta}_\tau$ )** Under  $H_{1,n}$ , if conditions (A1)-(A7) are verified, then  $\widehat{\theta}_\tau - \theta_\tau = o(1)$ , and

$$\widehat{\theta}_\tau - \theta_\tau = \frac{1}{n} \Omega^{-1} \sum_{i=1}^n \xi(X_i, Z_i, \delta_i; \widetilde{\theta}_\tau) + \frac{1}{\sqrt{n}} \Omega^{-1} \mathbb{E} \left[ f_{\varepsilon|X}(0) h(X) q_\tau^{(1)}(X, \theta_\tau) \right] + o\left(\frac{1}{\sqrt{n}}\right),$$

where

$$\begin{aligned} \xi(X_i, Z_i, \delta_i; \theta) &= \eta(X_i, Z_i, \delta_i; \infty, \theta) \\ &= \psi_\tau(Z_i - q_\tau(X_i, \theta)) q_\tau^{(1)}(X_i, \theta) \gamma_0(Z_i) \delta_i + \gamma_1^{\varphi_\infty(\cdot, \cdot; \theta)}(Z_i) (1 - \delta_i) - \gamma_2^{\varphi_\infty(\cdot, \cdot; \theta)}(Z_i). \end{aligned}$$

Note that under the null hypothesis, the function  $h$  equals zero and  $\widetilde{\theta}_\tau = \theta_\tau$ , and hence

$$\widehat{\theta}_\tau - \theta_\tau = \frac{1}{n} \Omega^{-1} \sum_{i=1}^n \xi(X_i, Z_i, \delta_i; \theta_\tau) + o\left(\frac{1}{\sqrt{n}}\right).$$

In particular, the estimator  $\widehat{\theta}_\tau$  is asymptotically unbiased in this case, since  $\mathbb{E}(\xi(X, Z, \delta; \theta_\tau)) = 0$ .

The iid representation of the estimator  $\widehat{\theta}_\tau$  will be necessary to show the behaviour of the new tests under the composite null or alternative hypothesis. Moreover, in view of Theorem 4, we can derive the asymptotic normality of the estimator  $\widehat{\theta}_\tau$ . This result is stated in the following corollary:

**Corollary 5 (Asymptotic normality of  $\widehat{\theta}_\tau$ )** Under  $H_{1,n}$ , if conditions (A1)-(A7) are verified, then

$$\sqrt{n}(\widehat{\theta}_\tau - \theta_\tau) \xrightarrow{d} N(b, \sigma^2),$$

where  $\sigma^2 = \Omega^{-1} \text{Var}(\xi(X, Z, \delta; \theta_\tau)) \Omega^{-1}$  and  $b = \Omega^{-1} \mathbb{E}[f_{\varepsilon|X}(0) h(X) q_\tau^{(1)}(X, \theta_\tau)]$ .

### 3.3 | Behaviour of the test under $H_{1,n}$

We can now derive the asymptotic representation for the empirical process  $R_{n,1}^1$  under local alternatives. This result is presented in the following theorem:

**Theorem 6 (IID representation of  $R_{n,1}^1$ )** Under  $H_{1,n}$ , if conditions (A1) - (A8) are verified, then the process  $R_{n,1}^1$  allows the following representation:

$$\begin{aligned} R_{n,1}^1(x_0) &= R_{n,1}^0(x_0) - \frac{1}{\sqrt{n}} \Omega(x_0) \Omega^{-1} \sum_{i=1}^n \xi(X_i, Z_i, \delta_i; \widetilde{\theta}_\tau) \\ &\quad - \Omega(x_0) \Omega^{-1} \mathbb{E} \left[ f_{\varepsilon|X}(0) h(X) q_\tau^{(1)}(X, \theta_\tau) \right] \\ &\quad + \mathbb{E} \left[ f_{\varepsilon|X}(0) h(X) q_\tau^{(1)}(X, \theta_\tau) \mathbb{I}(X \leq x_0) \right] + o_p(1), \end{aligned}$$

uniformly in  $x_0 \in \mathcal{R}_X$ , where  $\Omega(x_0) = \mathbb{E}[f_{\varepsilon|X}(0) q_\tau^{(1)}(X, \theta_\tau) q_\tau^{(1)}(X, \theta_\tau)' \mathbb{I}(X \leq x_0)]$ .

Note that the second term on the right hand side reflects the effect of the estimation of the parameter  $\theta_\tau$ , while the third and fourth terms are constants reflecting the deviation from the null hypothesis. Moreover, these terms related to the deviation are the same ones that He and Zhu (2003) obtained for complete data.

Also note that if we consider a model under the composite null hypothesis, the empirical process  $R_{n,1}^1$  allows the following representation:

$$R_{n,1}^1(x_0) = R_{n,1}^0(x_0) - \frac{1}{\sqrt{n}} \Omega(x_0) \Omega^{-1} \sum_{i=1}^n \xi(X_i, Z_i, \delta_i; \theta_\tau) + o_p(1),$$

where the second term represents the estimation effect, and there is no bias in this case.

This now leads to the following result, regarding the asymptotic distribution of the process  $R_{n,1}^1$  and the test statistic  $T_{n,1}^1$  under  $H_{1,n}$ :

**Corollary 7 (Composite and alternative hypotheses)** *Under  $H_{1,n}$ , if conditions (A1)-(A8) are verified, then the process  $R_{n,1}^1$  converges weakly in  $\ell^\infty(R_X)$  to the Gaussian process  $R_\infty^1$  with covariance function*

$$\begin{aligned} & \text{Cov}(R_\infty^1(x_0), R_\infty^1(x_1)) \\ &= \mathbb{E} \left[ \left\{ \eta(X, Z, \delta; x_0, \theta_\tau) - \Omega(x_0) \Omega^{-1} \xi(X, Z, \delta; \theta_\tau) \right\} \left\{ \eta(X, Z, \delta; x_1, \theta_\tau) - \Omega(x_1) \Omega^{-1} \xi(X, Z, \delta; \theta_\tau) \right\} \right], \end{aligned}$$

and with bias function given by

$$\begin{aligned} \mathbb{E}(R_\infty^1(x_0)) &= -\Omega(x_0) \Omega^{-1} \mathbb{E} \left[ f_{\varepsilon|X}(0) h(X) q_\tau^{(1)}(X, \theta_\tau) \right] \\ &\quad + \mathbb{E} \left[ f_{\varepsilon|X}(0) h(X) q_\tau^{(1)}(X, \theta_\tau) \mathbb{1}(X \leq x_0) \right]. \end{aligned}$$

Moreover, the test statistic  $T_{n,1}^1$  converges to the largest eigenvalue of the Cramer-von Mises norm  $\mathbb{E}[R_\infty^1(X) R_\infty^1(X)']$  of the process  $R_\infty^1$ .

**Remark 2.** Note that we could consider a more general alternative hypothesis as follows

$$H_{1a_n} : q_\tau(\cdot) = q_\tau(\cdot, \theta_\tau) + a_n h(\cdot),$$

where  $\theta_\tau$  is the true value under  $H_0$ ,  $h(\cdot)$  is a real-valued function with  $\mathbb{E}|h(X)| < \infty$  and  $a_n$  is a constant different from zero. Then, if  $a_n = 1/\sqrt{n}$  we have the local alternative, if  $a_n \sqrt{n}$  tends to infinity we have local alternatives that converge to the null at a slower rate than  $1/\sqrt{n}$ , and when  $a_n = 1$  we have global alternatives. Theorem 6 can be adapted to this new scenario. If we focus on the empirical process  $R_{n,1}^{1b}$  defined along the proof of Theorem 6, we could see that while the sum of IID terms (see proof of Theorem 4) will be the same under  $H_{1a_n}$  or  $H_{1,n}$ , the second and third terms of ((18)) will be quantities of the order  $O(a_n \sqrt{n})$ , which go to infinity if we have local alternatives that converge to the null hypothesis at a slower rate than  $1/\sqrt{n}$  or when we have global alternatives.

### 3.4 | Bootstrap approximation

In order to calibrate the distribution of the test statistics given in (7), we have several options. First, we could use a multiplier bootstrap procedure using the asymptotic behaviour of the empirical process under the composite null

hypothesis, similarly to Stute et al. (2000) or He and Zhu (2003). This would however entail the estimation of unknown quantities such as the conditional density of the error and the derivatives of the quantile regression function, which is not easy. Another option could be to use the limiting distribution of the test statistics, but the convergence to this limit could be slow, and moreover the estimation of the asymptotic variance is not easy.

In the framework of censored regression, several authors have presented bootstrap procedures. See e.g. De Uña-Álvarez and Roca-Pardiñas (2009) who present a bootstrap resampling for additive mean regression models assuming that  $C$  and  $(X, Y)$  are independent, Li and Datta (2001) who develop a bootstrap approach to nonparametric regression where the response variable  $Y$  and the censoring variable  $C$  are conditionally independent given the covariate  $X$ , or Pardo-Fernández and Van Keilegom (2006) who test the equality of regression functions under the assumption that  $C$  is independent of  $Y$  given  $X$ , among others. All of these proposals are based on conditions that need not be verified under Stute's conditions (A1) (i) and (ii).

For these reasons, we decide to use a bootstrap procedure following the ideas presented in Orbe and Núñez Antón (2013), which is based on Stute's conditions in the context of classical mean regression. For simplicity, we will explain the procedure for the test statistic  $T_{n,1}$ . The procedure is analogous for  $T_{n,2}$ . For a sample  $(X_i, Z_i, \delta_i)$  with  $i = 1, \dots, n$ , an estimator  $\widehat{\theta}_\tau$  and the test statistic  $T_{n,1}$ , we follow the following steps:

**Step 1:** For a given covariate  $X_i = (X_{i1}, X_{i2}, \dots, X_{id})$ , define the set  $\mathcal{A}_i = \prod_{j=1}^d \mathcal{A}_{ij}$ , where

$$\mathcal{A}_{ij} = \{k \in \{1, \dots, n\} : X_{ij} - g_n^j \leq X_{kj} \leq X_{ij} + g_n^j\},$$

for a given  $d$ -dimensional smoothing parameter  $g_n = (g_n^1, \dots, g_n^d)^1$ .

Draw a bootstrap error  $\varepsilon_i^*$  from the conditional Kaplan-Meier estimators (or the estimators of Beran (1981)) of the survival function of  $\varepsilon$ , constructed based on the set  $\mathcal{A}_i$ :

$$\widehat{F}_{\varepsilon|X=X_i}(t) = 1 - \prod_{j \in \mathcal{A}_i, r_j \leq t; \delta_j=1} \left(1 - \frac{1}{\sum_l \mathbb{1}(r_l \geq r_j)}\right),$$

where  $r_j = Z_j - q_\tau(X_j, \widehat{\theta}_\tau)$  with  $j = 1, \dots, n$ .

**Step 2:** Define bootstrap data  $(X_1, Y_1^*), \dots, (X_n, Y_n^*)$ , where

$$Y_i^* = q_\tau(X_i, \widehat{\theta}_\tau) + \varepsilon_i^*, \quad i = 1, \dots, n.$$

Then, the bootstrap responses  $Y_i^*$  follow the null hypothesis by construction.

**Step 3:** Generate bootstrap resamples of the censoring indicator taking into account condition (A1)-(ii) as

$$\mathbb{P}(\delta_i^* = 1 | Y_i^*, X_i) = \mathbb{P}(\delta_i^* = 1 | Y_i^*) = 1 - \widehat{F}_C(Y_i^*),$$

where  $\widehat{F}_C$  represents the Kaplan-Meier estimator of the distribution function of the censoring variable  $C$ .

Finally, we can build the bootstrap sample of the censoring variables  $\{C_1^*, \dots, C_n^*\}$  as follows:

- If  $\delta_i^* = 1$  then  $C_i^*$  is taken from  $\widehat{F}_C$  restricted to the interval  $[Y_i^*, \infty)$ .
- If  $\delta_i^* = 0$  then  $C_i^*$  is taken from  $\widehat{F}_C$  restricted to the interval  $[-\infty, Y_i^*)$ .

<sup>1</sup>This smoothing parameter will be selected using the procedure proposed by Li and Datta (2001).

Let  $Z_i^* = \min\{Y_i^*, C_i^*\}$ .

**Step 4:** Given the bootstrap sample  $\{(X_1, Z_1^*, \delta_1^*), \dots, (X_n, Z_n^*, \delta_n^*)\}$ , fit the parametric model (1) and denote by  $\widehat{\theta}_\tau^*$  the bootstrap estimator of the parameter vector  $\widehat{\theta}_\tau$ . Then, we compute the bootstrap test statistic

$$T_{n,1}^* = \text{largest eigenvalue of } \frac{1}{n} \sum_{i=1}^n R_{n,1}^{1,*}(X_i) [R_{n,1}^{1,*}(X_i)]',$$

where

$$R_{n,1}^{1,*}(t) = n^{1/2} \sum_{i=1}^n W_i^{\text{KM},*} \psi_\tau(Z_i^* - q_\tau(X_i, \widehat{\theta}_\tau^*)) q_\tau^{(1)}(X_i, \widehat{\theta}_\tau^*) \mathbb{I}(X_i \leq t).$$

**Step 5:** Repeat steps 1, 2, 3 and 4 many times.

If a number,  $B$ , of bootstrap samples are generated, then the p-value of the test may be approximated by the proportion of bootstrap test statistics, denoted by  $T_{n,1,b}^*$  for  $b = 1, \dots, B$ , that are not smaller than the original test statistic  $T_{n,1}$ , that is,

$$\frac{1}{B} \sum_{b=1}^B \mathbb{I}(T_{n,1} \leq T_{n,1,b}^*).$$

To have an idea about the validity of the previous bootstrap procedure, we have to take into account two arguments. On the one hand, the convergence of the Kaplan-Meier and Beran estimators, used in Steps (3) and (1) respectively, have already been studied in the literature, see for instance Lo and Singh (1986) or Gonzalez-Manteiga and Cadarso-Suarez (1994). Moreover, bootstrapped estimators of these quantities have also been studied by, for example, Akritas (1986) and Van Keilegom and Veraverbeke (1997).

On the other hand, note that to obtain the bootstrap censored indicators (Step 3) we are using assumption (A1)-(ii), that is, the probability of obtaining a censored or non-censored observation given the response variable does not depend on the covariate. That is, we are just replicating the properties of the original data. Moreover, the main difference between our bootstrap procedure and Orbe and Núñez Antón (2013)'s proposal is the fact that we are not assuming the independence between the covariates and the error distribution, that is, we adapt their proposal for heteroscedastic scenarios in the quantile regression context.

Moreover, in order to formally prove the consistency of our bootstrap procedure we could use arguments similar to Sant'Anna (2016), who prove the validity of a nonparametric bootstrap procedure to approximate the limiting variance function associated with the estimators of Kaplan-Meier integrals.

## 4 | SIMULATIONS

We will study the performance of our proposed tests under the null and the alternative hypotheses using a Monte Carlo simulation study. More precisely, we will verify how well our tests respect the nominal level, and we will compare the power of our tests with that of the test proposed by Wang (2008) under different model scenarios. The latter test

is defined as follows:

$$T_{n,W} = \frac{1}{n(n-1)h^d} \sum_{i=1}^n \sum_{j \neq i}^n K\left(\frac{X_i - X_j}{h}\right) \widehat{e}_i \widehat{e}_j,$$

where  $\widehat{e}_i = \mathbb{I}(Z_i \geq q_\tau(X_i, \widehat{\theta}_\tau)) - (1 - \tau) \widehat{F}_C(q_\tau(X_i, \widehat{\theta}_\tau))$ ,  $i = 1, \dots, n$ ,  $\widehat{F}_C$  is the Kaplan-Meier estimator of the censoring survival function  $F_C$ ,  $K$  is a  $d$ -dimensional kernel and  $h$  is a bandwidth parameter.

In all experiments, the number of simulated samples is 2000 and the number of bootstrap replications is  $B = 500$ . To perform the bootstrap procedure we need to select the bandwidth parameter  $g_n$  introduced in Step 1 of the procedure. To this end, we generalize Li and Datta (2001)'s proposal (see their Remark 2.1) to dimensions  $d \geq 1$ , i.e. we propose to work with  $g_n^j = c_j n^{-0.55/(d+4)}$  for some  $c_j > 0$  with  $j = 1, \dots, d$ , which reduces to their proposal  $g_n^1 = c_1 n^{-0.11}$  when  $d = 1$ . By working with  $g_n^j = c_j n^{-0.55/(d+4)}$  we respect Li and Datta (2001)'s underlying idea to oversmooth with respect to the optimal bandwidth for estimation, which is proportional to  $n^{-1/(d+4)}$  in  $d$  dimensions. For the constant  $c_j$  we use a simple rule used by Groeneboom and Jongbloed (2015), Lopuhaä and Musta (2017) and Lopuhaä and Musta (2018), among others, that consists in taking each constant equal to the range of the  $j$ -th covariate for  $j = 1, \dots, d$ .

In order to optimize the adjustment of the nominal level, we use presmoothed Kaplan-Meier weights instead of classical ones to compute the empirical processes in (6). Presmoothing has the advantage that it reduces the asymptotic mean squared error variability of classical Kaplan-Meier estimators. See Cao et al. (2005) or Dikta (1998) for more details. To compute the presmoothed Kaplan-Meier weights we use the R package `condSURV` (see Meira-Machado and Sestelo (2016)). Moreover, to compute the proposed estimator  $\widehat{\theta}_\tau$  given in (5) the function `rq` available in the R package `quantreg` is used.

We first focus on the behaviour under the null hypothesis, in order to check the adjustment of the significance level. We simulate data from the following quantile regression model studied by Wang (2008):

$$\text{Model 1: } Y = -0.7 + X + (\varepsilon - c_\tau),$$

where  $X$  follows an uniform distribution on the interval  $(0, 1)$  and the error  $\varepsilon$  follows a standard normal distribution, which is drawn independently of the covariate. The constant  $c_\tau$  represents the  $\tau$ -quantile of a standard normal distribution. The null hypothesis is the hypothesis of a linear model. Moreover, the censoring variable  $C$  is independent of  $X$  and follows a normal distribution with variance 1. The mean of  $C$  is chosen so as to obtain a certain prespecified censoring rate. Note that we could allow the censoring variable to depend on  $X$ , but since we will compare our tests with the test of Wang (2008) and since this test requires  $C$  and  $X$  to be independent, we pay attention to the independent case.

Table 1 shows the proportion of rejections associated with the two test statistics  $T_{n,1}$  and  $T_{n,2}$  for different sample sizes,  $n$ , different quantiles of interest,  $\tau$ , different nominal levels,  $\alpha$ , and different censoring rates. From Table 1 we can conclude that the tests show a good adjustment of the significance level for all considered scenarios when  $\tau = 0.25$  or  $\tau = 0.50$ . Note that for  $\tau = 0.75$  and a high censoring rate the level is not so good, as can be expected since this is a difficult situation. But also in this case, the proportion of rejections gets closer to the nominal level when the sample size increases. Finally, there are no notable differences between the two test procedures.

Next, we study the behaviour under the null hypothesis of a linear model for three other models, involving other

Cen. rate	$\tau$	$n$	$T_{n,1}$			$T_{n,2}$		
			$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
20%	0.25	100	0.093	0.044	0.006	0.094	0.048	0.006
		500	0.110	0.051	0.010	0.108	0.052	0.008
		1000	0.113	0.054	0.016	0.108	0.057	0.016
	0.50	100	0.086	0.040	0.006	0.090	0.042	0.008
		500	0.088	0.043	0.011	0.084	0.044	0.010
		1000	0.095	0.045	0.011	0.097	0.048	0.013
	0.75	100	0.100	0.050	0.010	0.100	0.044	0.008
		500	0.088	0.040	0.007	0.092	0.040	0.008
		1000	0.090	0.039	0.008	0.090	0.041	0.010
45%	0.25	100	0.112	0.050	0.011	0.106	0.050	0.010
		500	0.105	0.054	0.013	0.103	0.048	0.011
		1000	0.100	0.055	0.012	0.100	0.050	0.013
	0.50	100	0.090	0.041	0.009	0.089	0.038	0.010
		500	0.094	0.046	0.009	0.094	0.046	0.008
		1000	0.100	0.052	0.013	0.100	0.052	0.014
	0.75	100	0.080	0.035	0.010	0.074	0.034	0.007
		500	0.094	0.044	0.010	0.099	0.045	0.010
		1000	0.102	0.052	0.008	0.104	0.052	0.008

**TABLE 1** Proportion of rejections associated with our proposed lack-of-fit tests for Model 1.

error distributions, heteroscedastic errors or more than one covariate.

$$\text{Model 2: } Y = -0.7 + X_1 + \frac{1}{3} (\varepsilon_2 - c_{\tau,2}),$$

$$\text{Model 3: } Y = -0.7 + X_1 + (X_1 + 0.5) (\varepsilon_3 - c_{\tau,3}),$$

$$\text{Model 4: } Y = -1.2 + X_1 + X_2 + (\varepsilon_4 - c_{\tau,4}),$$

where  $X_1$  and  $X_2$  are independent and follow an uniform distribution on the interval  $(0, 1)$ , the errors  $\varepsilon_2$ ,  $\varepsilon_3$  and  $\varepsilon_4$  are drawn independently of the covariate(s),  $\varepsilon_2$  follows a Student- $t$  distribution with 6 degrees of freedom, and  $\varepsilon_3$  and  $\varepsilon_4$  follow a standard normal distribution. The constants  $c_{\tau,2}$ ,  $c_{\tau,3}$  and  $c_{\tau,4}$  represent the  $\tau$ -quantiles of the corresponding error distributions. Moreover, the censoring variable  $C$  is independent of  $X$  and follows a normal distribution with variance 1, and where the mean is chosen so as to obtain a certain prespecified censoring rate.

Note that Model 2 will allow to verify the effect that the error distribution has on the level, whereas Models 3 and 4 will show the behaviour of the new tests under heteroscedasticity and in multivariate scenarios, respectively. Table 2 shows the proportion of rejections associated with the two test statistics  $T_{n,1}$  and  $T_{n,2}$  for different sample sizes,  $n$ , different quantiles of interest,  $\tau$ , and different nominal significance levels,  $\alpha$ , for Models 2, 3 and 4. From Tables 1 and 2 we can conclude that the tests show a good adjustment of the significance level, independently of the error distribution, the dimension of the covariate, the quantile of interest and the censoring rate, and both in homo- and heteroscedastic scenarios.

The test proposed by Wang (2008) requires  $C$  and  $X$  to be independent, but this is a severe restriction in practice. Therefore, we will check the adjustment of the nominal level of the new tests when  $C$  can depend on  $X$ . Let us consider the following regression model that has been used by Stute (1993) in the context of classical mean regression:

$$\text{Model 5: } Y = 3 + X_1 + X_2 + (\varepsilon - c_\tau),$$

where  $X_1$  and  $X_2$  are independent and follow an uniform distribution on the interval  $(0, 1)$  and the error  $\varepsilon$  follows a standard normal distribution, which is drawn independently of the covariates. The constant  $c_\tau$  represents the  $\tau$ -quantile of a standard normal distribution. Moreover, the censoring variable is defined as follows:

$$C = 3 + X_1 + X_2 + \bar{\varepsilon},$$

where  $\bar{\varepsilon}$  follows a normal distribution with variance 1 and with different means to control the censoring rate.

Table 3 shows the proportion of rejections associated with the two test statistics  $T_{n,1}$  and  $T_{n,2}$  for different sample sizes,  $n$ , different quantiles of interest,  $\tau$ , and different nominal significance levels,  $\alpha$ , for Model 5. We can conclude that the new tests show a good adjustment of the significance level, when the censoring variable  $C$  depends on the covariate  $X$ .

Next, let us compare our proposed tests with the lack-of-fit test in Wang (2008). We simulate data for the following quantile regression models that have been considered in Wang (2008):

$$\text{Model 6: } Y = -0.7 + X_1 - c h(X_1) + (\varepsilon - c_\tau),$$

$$\text{Model 7: } Y = -1.2 + X_1 + X_2 - c h(X_1, X_2) + (\varepsilon - c_\tau),$$

where  $X_1$  and  $X_2$  are independent and follow an uniform distribution on the interval  $(0, 1)$ , and  $c_\tau$  denotes the  $\tau$ -quantile of the error  $\varepsilon$ . We consider two possible error distributions, namely a normal distribution and a log-normal

Model	CR	$\tau$	$n$	$T_{n,1}$			$T_{n,2}$		
				$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
2	35%	0.25	100	0.097	0.054	0.010	0.094	0.050	0.009
			500	0.110	0.060	0.018	0.116	0.058	0.019
			1000	0.120	0.064	0.014	0.120	0.064	0.018
	0.50	100	0.095	0.049	0.011	0.089	0.046	0.009	
		500	0.111	0.062	0.013	0.108	0.061	0.013	
		1000	0.114	0.066	0.022	0.116	0.068	0.020	
	0.75	100	0.082	0.035	0.006	0.077	0.036	0.006	
		500	0.114	0.061	0.016	0.109	0.060	0.016	
		1000	0.100	0.059	0.014	0.102	0.057	0.014	
3	30%	0.25	100	0.107	0.058	0.014	0.103	0.058	0.016
			500	0.107	0.056	0.014	0.102	0.054	0.014
			1000	0.098	0.050	0.010	0.097	0.050	0.008
	0.50	100	0.087	0.043	0.010	0.088	0.042	0.010	
		500	0.098	0.053	0.012	0.102	0.050	0.012	
		1000	0.102	0.048	0.015	0.099	0.050	0.012	
	0.75	100	0.100	0.048	0.012	0.099	0.051	0.012	
		500	0.098	0.050	0.014	0.098	0.051	0.014	
		1000	0.106	0.048	0.010	0.110	0.048	0.008	
4	45%	0.25	100	0.094	0.050	0.010	0.099	0.050	0.010
			500	0.108	0.054	0.014	0.105	0.053	0.012
			1000	0.101	0.050	0.008	0.108	0.048	0.010
	0.50	100	0.081	0.044	0.006	0.081	0.040	0.006	
		500	0.088	0.042	0.012	0.086	0.039	0.012	
		1000	0.090	0.043	0.010	0.094	0.040	0.011	
	0.75	100	0.064	0.024	0.003	0.060	0.018	0.002	
		500	0.094	0.050	0.009	0.094	0.050	0.008	
		1000	0.098	0.045	0.012	0.086	0.048	0.011	

**TABLE 2** Proportion of rejections associated with our proposed lack-of-fit tests for Models 2, 3 and 4 (CR = censoring rate).

Cen. rate	$\tau$	$n$	$T_{n,1}$			$T_{n,2}$		
			$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
25%	0.25	100	0.096	0.046	0.011	0.098	0.044	0.010
		500	0.087	0.040	0.008	0.091	0.044	0.008
		1000	0.100	0.050	0.011	0.104	0.050	0.012
	0.50	100	0.085	0.046	0.008	0.100	0.048	0.010
		500	0.090	0.038	0.008	0.098	0.046	0.011
		1000	0.108	0.060	0.014	0.114	0.062	0.016
	0.75	100	0.081	0.034	0.008	0.080	0.038	0.007
		500	0.100	0.045	0.010	0.098	0.051	0.015
		1000	0.094	0.048	0.010	0.096	0.058	0.014

**TABLE 3** Proportion of rejections associated with our proposed lack-of-fit tests for Model 5.

distribution, to check the effect of the error distribution on the power. The null hypothesis is the hypothesis of a linear model in  $X_1$  (Model 6) or in  $X_1$  and  $X_2$  (Model 7). Note that the function  $ch(\cdot)$  represents the deviation of the model from the null hypothesis, and the constant  $c$  will control these deviations.

In order to make a fair comparison between our proposals and Wang (2008)'s test, we will first analyse the adjustment of the nominal level under Model 6 and Model 7. That is, we will consider the case  $c = 0$  and the results are shown in Table 4. According to Table 4 we can conclude that the proportion of rejections associated with the new lack-of-fit tests is around the nominal level in the different considered scenarios while the test proposed by Wang (2008) is a bit conservative in some situations.

Table 5 shows the proportion of rejections associated with the proposed tests and the test proposed by Wang (2008) for Model 6. Note that to choose the smoothing parameter needed to compute Wang (2008)'s test a generalized cross-validation (GCV) procedure is used (see page 323 of Wang (2008)). From Table 5 we can conclude that the proportion of rejections goes to one when the sample size and/or the constant  $c$  increases, showing the consistency of the new tests, independently of the quantile of interest. Moreover, the power of the new lack-of-fit tests is significantly higher than that of the test of Wang (2008) in most of the scenarios that are considered. Similar conclusions can be drawn from Table 6 for Model 7.

## 5 | REAL DATA APPLICATION

In this section we will show the performance of the new lack-of-fit tests in practice. To this aim, two different real data sets will be considered. Firstly, we will analyse data coming from the University of Massachusetts AIDS Research Unit (UMARU) IMPACT Study. This is a 5-year collaborative research project about drug abuse. A detailed description of the study can be found in Hosmer et al. (2008). The data set has been analysed by several authors such as Yang et al. (2014) and Geerdens et al. (2020), among others. Moreover, the data are available in the R package `quantreg` under the name `uis`. For each of the 517 patients in this data set, the time to return to drug abuse (measured in days, and denoted by `TIME`) is registered, as well as the age of the variable (`AGE`, measured in years), a compliance variable (denoted by `FRAC`), and the `RACE` (where 0 represents white and 1 other). The variable `FRAC` allows us to compare the

Model	$\tau$	$n$	$\varepsilon \sim N(0, 1)$			$\varepsilon \sim \exp(N(0, 0.5))$		
			$T_{n,1}$	$T_{n,2}$	$T_{n,W}$	$T_{n,1}$	$T_{n,2}$	$T_{n,W}$
6	0.25	100	0.050	0.049	0.030	0.054	0.050	0.031
		500	0.048	0.048	0.035	0.042	0.042	0.032
		1000	0.049	0.049	0.033	0.052	0.052	0.038
	0.50	100	0.056	0.055	0.028	0.052	0.047	0.024
		500	0.053	0.054	0.032	0.057	0.056	0.032
		1000	0.061	0.060	0.040	0.058	0.059	0.028
	0.75	100	0.043	0.041	0.020	0.050	0.044	0.024
		500	0.050	0.052	0.030	0.052	0.050	0.030
		1000	0.054	0.052	0.032	0.050	0.047	0.030
7	0.25	100	0.050	0.048	0.028	0.050	0.052	0.03
		500	0.048	0.052	0.021	0.054	0.064	0.02
		1000	0.054	0.049	0.019	0.062	0.057	0.02
	0.50	100	0.042	0.042	0.032	0.044	0.039	0.031
		500	0.042	0.042	0.016	0.052	0.054	0.019
		1000	0.050	0.046	0.019	0.051	0.057	0.016
	0.75	100	0.034	0.039	0.130	0.042	0.042	0.084
		500	0.040	0.039	0.014	0.048	0.049	0.016
		1000	0.047	0.042	0.024	0.062	0.055	0.016

**TABLE 4** Proportion of rejections associated with our proposed lack-of-fit tests and the test proposed by Wang (2008) for Models 6 and 7 when  $c = 0$  for  $\alpha = 0.05$ , where the censoring rate is between 30% – 35% for all scenarios.

$h(x)$	$c$	$\tau$	$n$	$\varepsilon \sim N(0, 1)$			$\varepsilon \sim \exp(N(0, 0.5))$			
				$T_{n,1}$	$T_{n,2}$	$T_{n,W}$	$T_{n,1}$	$T_{n,2}$	$T_{n,W}$	
$(x - 0.5)^2$	4	0.25	100	0.448	0.456	0.145	0.909	0.920	0.710	
			500	0.989	0.990	0.841	1.000	1.000	1.000	
			1000	1.000	1.000	0.994	1.000	1.000	1.000	
		0.50	100	0.355	0.362	0.124	0.654	0.676	0.453	
			500	0.946	0.952	0.770	1.000	0.999	1.000	
			1000	0.998	0.998	0.988	1.000	1.000	1.000	
	0.75	100	0.171	0.176	0.050	0.214	0.218	0.103		
		500	0.659	0.690	0.402	0.814	0.828	0.818		
		1000	0.906	0.922	0.808	0.981	0.985	0.994		
		6	0.25	100	0.775	0.786	0.414	0.994	0.996	0.976
				500	1.000	1.000	0.999	1.000	1.000	1.000
				1000	1.000	1.000	1.000	1.000	1.000	1.000
0.50	100		0.692	0.714	0.308	0.934	0.942	0.846		
	500		1.000	1.000	0.998	1.000	1.000	1.000		
	1000		1.000	1.000	1.000	1.000	1.000	1.000		
0.75	100	0.372	0.388	0.108	0.458	0.462	0.272			
	500	0.965	0.976	0.862	0.992	0.992	0.996			
	1000	1.000	1.000	0.996	1.000	1.000	1.000			
	$e^x$	4	0.25	100	0.513	0.514	0.079	0.898	0.902	0.301
				500	0.990	0.987	0.525	1.000	1.000	0.988
				1000	1.000	1.000	0.896	1.000	1.000	1.000
0.50			100	0.463	0.434	0.084	0.528	0.526	0.084	
			500	0.970	0.958	0.491	0.984	0.982	0.622	
			1000	0.999	0.997	0.878	1.000	1.000	0.946	
0.75		100	0.216	0.168	0.044	0.104	0.090	0.038		
		500	0.716	0.648	0.272	0.404	0.377	0.119		
		1000	0.922	0.871	0.628	0.539	0.508	0.240		
		6	0.25	100	0.792	0.800	0.148	0.980	0.985	0.556
				500	1.000	1.000	0.846	1.000	1.000	1.000
				1000	1.000	1.000	0.994	1.000	1.000	1.000
0.50	100		0.729	0.711	0.110	0.789	0.786	0.138		
	500		1.000	1.000	0.718	1.000	1.000	0.836		
	1000		1.000	1.000	0.980	1.000	1.000	0.996		
0.75	100	0.461	0.418	0.064	0.260	0.240	0.032			
	500	0.988	0.978	0.478	0.774	0.742	0.210			
	1000	1.000	0.998	0.876	0.924	0.900	0.505			

**TABLE 5** Proportion of rejections under Model 6 for our proposed lack-of-fit tests and for the test proposed by Wang (2008). The nominal level is  $\alpha = 0.05$  and the censoring rate is between 30% and 35% for all scenarios.

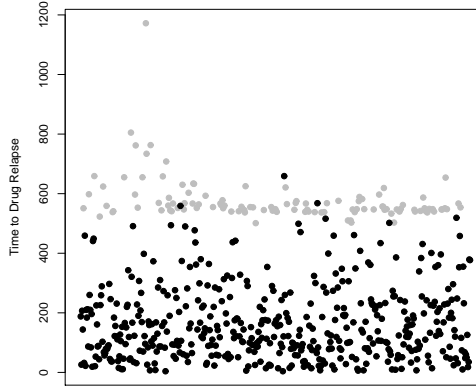
$h(x_1, x_2)$	$c$	$\tau$	$n$	$\varepsilon \sim N(0, 1)$			$\varepsilon \sim \exp(N(0, 0.5))$			
				$T_{n,1}$	$T_{n,2}$	$T_{n,W}$	$T_{n,1}$	$T_{n,2}$	$T_{n,W}$	
$(x_1 + x_2)^2$	2	0.25	100	0.398	0.430	0.126	0.722	0.716	0.528	
			500	0.994	0.994	0.821	1.000	1.000	1.000	
			1000	1.000	1.000	0.996	1.000	1.000	1.000	
		0.50	100	0.216	0.231	0.066	0.372	0.376	0.200	
			500	0.918	0.920	0.612	0.994	0.992	0.992	
			1000	0.996	0.995	0.978	1.000	1.000	1.000	
	0.75	100	0.060	0.055	0.097	0.080	0.070	0.086		
		500	0.362	0.343	0.262	0.560	0.526	0.509		
		1000	0.612	0.601	0.688	0.819	0.797	0.950		
	4	0.25	100	100	0.764	0.773	0.281	0.847	0.832	0.740
				500	1.000	1.000	0.994	1.000	1.000	1.000
				1000	1.000	1.000	1.000	1.000	1.000	1.000
0.50			100	0.510	0.544	0.142	0.663	0.662	0.380	
			500	0.997	0.998	0.952	1.000	1.000	1.000	
			1000	1.000	1.000	1.000	1.000	1.000	1.000	
0.75		100	0.148	0.183	0.072	0.227	0.268	0.097		
		500	0.868	0.846	0.570	0.961	0.942	0.842		
		1000	0.981	0.971	0.974	0.994	0.990	1.000		
$e^{x_1+x_2}$		2	0.25	100	0.649	0.672	0.258	0.808	0.774	0.758
				500	1.000	1.000	0.984	1.000	1.000	1.000
				1000	1.000	1.000	1.000	1.000	1.000	1.000
	0.50		100	0.382	0.403	0.157	0.553	0.554	0.384	
			500	0.988	0.984	0.947	1.000	1.000	1.000	
			1000	1.000	0.999	1.000	1.000	1.000	1.000	
	0.75	100	0.060	0.055	0.097	0.152	0.156	0.094		
		500	0.362	0.343	0.262	0.824	0.791	0.830		
		1000	0.612	0.601	0.688	0.955	0.936	0.998		
	4	0.25	100	100	0.840	0.842	0.424	0.876	0.868	0.848
				500	1.000	1.000	1.000	1.000	1.000	1.000
				1000	1.000	1.000	1.000	1.000	1.000	1.000
0.50		100	100	0.522	0.556	0.284	0.627	0.638	0.693	
			500	0.999	0.999	1.000	1.000	1.000	1.000	
			1000	1.000	1.000	1.000	1.000	1.000	1.000	
0.75	100	100	0.154	0.186	0.094	0.216	0.266	0.152		
		500	0.924	0.902	0.932	0.978	0.968	0.994		
		1000	0.998	0.998	1.000	0.999	0.999	1.000		

**TABLE 6** Proportion of rejections under Model 7 for our proposed lack-of-fit tests and for the test proposed by Wang (2008). The nominal level is  $\alpha = 0.05$  and the censoring rate is between 30% and 35% for all scenarios.

length of stay in two treatments with different durations and it is defined as

$$\text{FRAC} = \begin{cases} \frac{\text{length of stay in treatment}}{90} & \text{for the short treatment} \\ \frac{\text{length of stay in treatment}}{180} & \text{for the long treatment} \end{cases}$$

The time to drug abuse is subject to random right censoring, and is shown in Figure 1.



**FIGURE 1** Representation of the variable TIME of the data set *uis*. The black, respectively grey, points represent the uncensored, respectively censored, observations in the sample.

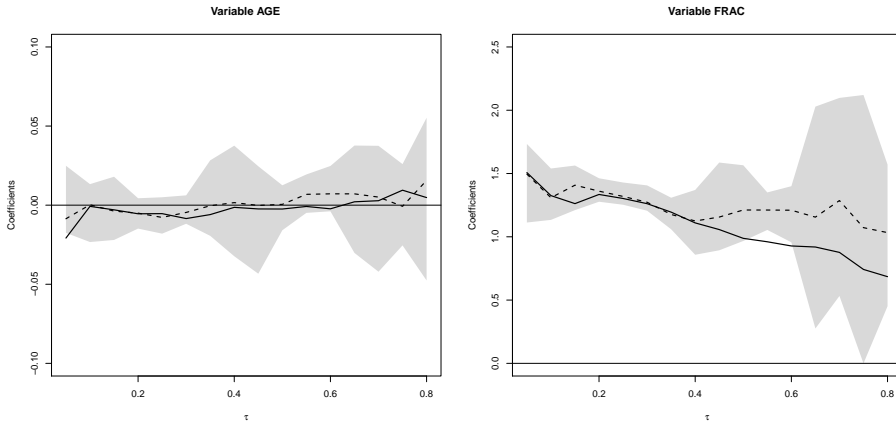
The main goal will be to test if the behaviour of the accelerated failure time (AFT) model can be explained using a linear quantile regression model. We fit a separate AFT model for  $\text{RACE}=0$  and 1. Hence, under the null hypothesis the model will be

$$\log(\text{TIME}) = \theta_{\tau,0,k} + \theta_{\tau,1,k} \text{FRAC} + \theta_{\tau,2,k} \text{AGE} + \varepsilon_k, \quad (10)$$

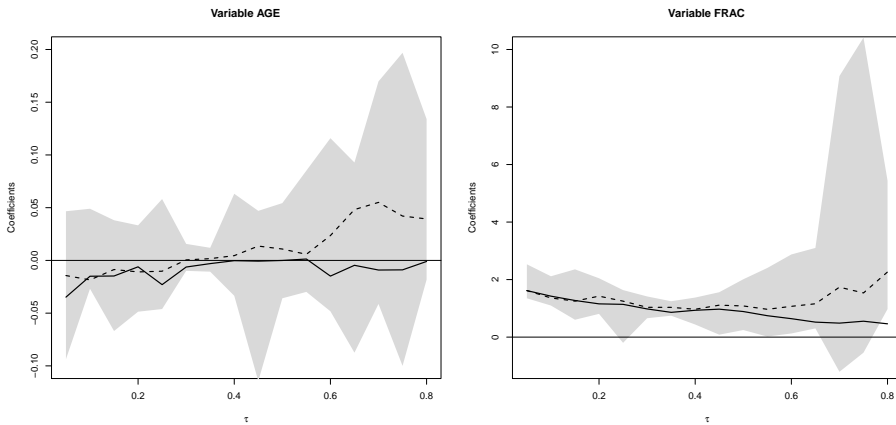
where  $\varepsilon_k$  denotes the unknown error having  $\tau$ -quantile equal to zero, and  $k = 0, 1$  depending on the value of RACE.

Figure 2 represents two estimators of the parameters  $\theta_{\tau,j,k}$  ( $j = 1, 2; k = 0, 1$ ) for different values of  $\tau$ , namely the estimator of Portnoy (2003) and the one proposed in (5). In most cases, the new estimator is inside the confidence intervals of Portnoy (2003)'s estimator, and the main differences between the two estimators are for high quantile orders.

Table 7 contains the p-values obtained from the application of the proposed lack-of-fit tests for different values of  $\tau$ . Although the simulations in Section 4 showed that the behaviour of the new tests for the proposed smoothing parameter  $g_n$  is really good, in practice we recommend to compute the tests for a reasonable range of bandwidths, to have a more global view on the power of the tests for a range of bandwidth values. Therefore, we consider bandwidths  $g_n$  of the form  $g_n \sim n^{-\beta}$  for different values of the parameter  $\beta$ , in addition to the bandwidth  $g_n = cn^{-0.55/6} = cn^{-0.0917}$  used in the simulation section. Moreover, as in the Monte Carlo simulation study, the constant  $c$  is equal to the range of the covariate, following the ideas of Groeneboom and Jongbloed (2015). The number of bootstrap replications is  $B = 500$ .



(a) RACE=0



(b) RACE=1

**FIGURE 2** Estimation of the regression coefficients  $\theta_{\tau,j,k}$  ( $j = 1, 2; k = 0, 1$ ) in model (10) for different values of  $\tau$ . The dashed curve is the estimator of Portnoy (2003), while the solid curve is the new estimator proposed in (5). The grey area represents confidence intervals of Portnoy (2003)'s estimator.

RACE	$\tau$	$\beta = 0.02$		$\beta = 0.05$		Li-Datta's bw $\beta = 0.0917$		$\beta = 0.15$		Wang (2008)
		$T_{n,1}$	$T_{n,2}$	$T_{n,1}$	$T_{n,2}$	$T_{n,1}$	$T_{n,2}$	$T_{n,1}$	$T_{n,2}$	
0	0.20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	5.7 e-08
	0.40	0.000	0.002	0.004	0.006	0.002	0.000	0.018	0.026	0.001
	0.60	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.9 e-4
	0.80	0.016	0.036	0.024	0.044	0.026	0.042	0.044	0.076	0.003
1	0.20	0.166	0.084	0.182	0.100	0.178	0.108	0.210	0.102	0.444
	0.40	0.346	0.276	0.326	0.262	0.328	0.260	0.360	0.264	0.045
	0.60	0.148	0.048	0.156	0.062	0.152	0.048	0.154	0.052	0.551
	0.80	0.446	0.454	0.472	0.514	0.466	0.494	0.478	0.526	0.149

**TABLE 7** P-values associated with our proposed lack-of-fit tests and with the test of Wang (2008) for model (10). The proposed tests are calculated using the bandwidth parameter  $g_n = cn^{-\beta}$  for a range of values of  $\beta$  and using the bandwidth proposed by Li and Datta (2001).

It follows from Table 7 that for all considered bandwidths we can reject the null hypothesis for the different values of  $\tau$  when RACE=0, i.e. we reject the linear quantile regression model (10) when RACE=0 except for  $T_{n,2}$  with  $\beta = 0.15$  and  $\tau = 0.8$ . On the other hand, the linear model is not rejected when RACE=1. Note that the parameter  $g_n$  does not seem to have an important effect on the p-value of the tests. The same conclusions are reached with the test of Wang (2008), except when  $\tau = 0.4$  and RACE=1.

Secondly, we will use a dataset about Pancreatic Cancer (more information at <https://seer.cancer.gov/statfacts/html/pancreas.html>). Pancreatic Cancer is relatively rare and was estimated to be the 12<sup>th</sup> most common type of cancer in 2014 representing less than 3% of new diagnosed cases. It is found to be more common with increasing ages. However it is the forth leading cause of cancer death in the United States (7% of cancer deaths). We will explain the survival (in months) after a pancreatic cancer diagnostic (denoted by  $Y$ ), that is a right censored variable, based on the age at diagnosis (denoted by  $X_1$ ) and the tumour size (denoted by  $X_2$ ) using the following median regression model:

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_1 * X_2 + \varepsilon, \quad (11)$$

where  $X_1 * X_2$  denotes the interaction between the variables  $X_1$  and  $X_2$ . The covariates  $X_1$  and  $X_2$  are standardised.

Moreover, as 8 out of 10 patients are diagnosed after the cancer has spread outside of the pancreas, we will focus on patients whose cancer has spread to regional lymph nodes. In addition, we will consider white patients and five different states of the United States: Hawaii (denoted by HI), Iowa (denoted by IA), New Jersey (denoted by NJ), Utah (denoted by UT) and Washington (denoted by WA). Furthermore, model (11) will be fitted for males and females.

Table 8 contains the p-values obtained from the application of the proposed lack-of-fit tests for model (11) and for the test proposed by Wang (2008). Furthermore, as in the Monte Carlo simulation study,  $g_n = cn^{-0.55/(d+4)}$  where the constant  $c$  is equal to the range of the covariate, following the ideas of Groeneboom and Jongbloed (2015), and the number of bootstrap replications is  $B = 500$ . According to Table 8, independently of the considered lack-of-fit test, model (11) can be accepted in some states (as Male and Female in Hawaii) or can be rejected in other states (as Female in Washington). Moreover, our test can detect deviations from model (11) in some situations (as Male and

Female in New Jersey) while the test proposed by Wang (2008) would accept the parametric model.

STATE	Male			Female		
	$T_{n,1}$	$T_{n,2}$	Wang (2008)	$T_{n,1}$	$T_{n,2}$	Wang (2008)
HI	0.592	0.716	0.878	0.882	0.652	0.824
IA	0.036	0.028	0.617	0.010	0.010	0.235
NJ	0.000	0.004	0.156	0.000	0.004	0.150
UT	0.038	0.036	0.483	0.230	0.748	0.903
WA	0.014	0.048	0.354	0.002	0.000	0.007

**TABLE 8** P-values associated with our proposed lack-of-fit tests and with the test proposed by Wang (2008) for model (11).

## 6 | CONCLUSIONS

Two new lack-of-fit tests for quantile regression are proposed, in the case where the response variable is subject to random right censoring. The tests are based on the cumulative sum of residuals, and extend the ideas of He and Zhu (2003) to censored quantile regression. We showed that the empirical processes associated with the test statistics converge to a Gaussian process under the null hypothesis. The limiting distribution under local alternative hypotheses is also established. Moreover, to approximate the critical values of the lack-of-fit tests, a bootstrap procedure is used.

A Monte Carlo simulation study is conducted to study the behaviour of the new tests under homo- and heteroscedastic error distributions for different quantiles levels and for different censoring rates. It is found that the new tests are generally more powerful than the test of Wang (2008). Moreover, the proposed lack-of-fit tests are more general than Wang (2008)'s proposal because they do not need to assume that the covariates are continuous or independent of the censoring variable. Finally, two real data applications are presented in order to illustrate how the tests can be used in practice.

Note that we can extend the proposed tests using different weighting functions instead of the indicator function considered in this paper, following the ideas of Colling and Van Keilegom (2017) for semiparametric transformation models. In particular, the new tests can be adapted to a high-dimensional context using unidimensional linear projections of the covariates following the ideas of Conde-Amboage et al. (2015) for quantile regression models with complete data. This is currently under investigation. It would also be useful to extend the tests so that the null hypothesis can be tested for several (or for a range of) quantile levels simultaneously.

## Acknowledgements

The authors gratefully acknowledge the support of Project MTM2016-76969-P (Spanish State Research Agency, AEI) co-funded by the European Regional Development Fund (ERDF). The work of M. Conde-Amboage has been supported by a post-doctoral grant from the Ministry of Culture, Education and University Planning and the Ministry of Economy, Employment and Industry of the Galician Government. I. Van Keilegom gratefully acknowledges support

by the European Research Council (2016-2021, Horizon 2020/ERC grant No. 694409).

## Appendix

**Proof of Lemma 1.** Given the function

$$\varphi_{x_0}(x, y; \theta_\tau) = \psi_\tau(y - q_\tau(x, \theta_\tau)) q_\tau^{(1)}(x, \theta_\tau) \mathbb{1}(x \leq x_0),$$

it follows that  $\int \varphi_{x_0}(x, y; \theta_\tau) dF_{X,Y}(x, y) = \int \mathbb{E}[\varphi_{x_0}(X, Y; \theta_\tau) | X = x] dF_X(x)$ , where

$$\begin{aligned} \mathbb{E}[\varphi_{x_0}(X, Y; \theta_\tau) | X = x] &= \mathbb{E}[\psi_\tau(Y - q_\tau(X, \theta_\tau)) q_\tau^{(1)}(X, \theta_\tau) \mathbb{1}(X \leq t) | X = x] \\ &= \mathbb{E}[\psi_\tau(Y - q_\tau(X, \theta_\tau)) | X = x] q_\tau^{(1)}(x, \theta_\tau) \mathbb{1}(x \leq t). \end{aligned}$$

Then, the implication “ $\Rightarrow$ ” is clear because of the definition of the function  $\psi_\tau$ , i.e.

$$\begin{aligned} \text{The null hypothesis } H_0 : q_\tau \in Q_\theta \text{ holds} &\implies \mathbb{P}[\varepsilon \leq 0 | X] = \tau \\ &\implies \mathbb{E}[\psi_\tau(Y - q_\tau(X, \theta_\tau)) | X] = 0 \\ &\implies \mathbb{E}[\varphi_{x_0}(X, Y; \theta_\tau) | X] = 0 \\ &\implies \int \varphi_{x_0}(x, y; \theta_\tau) dF_{X,Y}(x, y) = 0. \end{aligned}$$

Moreover, the implication “ $\Leftarrow$ ” is a consequence of the following development:

$$\begin{aligned} \int \varphi_{x_0}(x, y; \theta_\tau) dF_{X,Y}(x, y) = 0 &\implies \int \mathbb{E}[\varphi_{x_0}(X, Y; \theta_\tau) | X = x] dF_X(x) = 0 \\ &\implies \int \mathbb{E}[\psi_\tau(Y - q_\tau(X, \theta_\tau)) | X = x] q_\tau^{(1)}(x, \theta_\tau) \mathbb{1}(x \leq t) dF_X(x) = 0, \forall t \in \mathbb{R} \\ &\implies \int_{-\infty}^t \mathbb{E}[\psi_\tau(Y - q_\tau(X, \theta_\tau)) | X = x] q_\tau^{(1)}(x, \theta_\tau) dF_X(x) = 0, \forall t \in \mathbb{R} \\ &\implies \mathbb{E}[\psi_\tau(Y - q_\tau(X, \theta_\tau)) | X] q_\tau^{(1)}(X, \theta_\tau) = 0 \\ &\implies \mathbb{E}[\psi_\tau(Y - q_\tau(X, \theta_\tau)) | X] = 0 \implies \mathbb{P}[\varepsilon \leq 0 | X] = \tau \\ &\implies \text{The null hypothesis } H_0 : q_\tau \in Q_\theta \text{ holds,} \end{aligned}$$

where we have taken into account that  $\mathbb{P}(q_\tau^{(1)}(X, \theta) = 0) = 0$  for any  $\theta \in \Theta$ . □

**Proof of Theorem 2.** The empirical process  $R_{n,1}^0$  can be written as a Kaplan-Meier integral

$$R_{n,1}^0(x_0) = n^{1/2} \int \varphi_{x_0}(x, y; \theta_\tau) d\widehat{F}_{X,Y}(x, y),$$

where  $\widehat{F}_{X,Y}(x, y) = \sum_{i=1}^n W_i^{\text{KM}} \mathbb{1}(X_i \leq x, Z_i \leq y)$ . The i.i.d. representation given in (8) relies on the results of Theorem 3.11 in Sánchez-Sellero (2001) (see page 123). Therefore, we will prove the conditions established in the latter theorem.

**Condition 1:** The family of functions  $\varphi_{x_0}$  should be a VC-class. In order to prove this statement, we will focus on checking that the family of indicator functions of the form  $x \rightarrow \mathbb{1}(x \leq x_0)$  is a VC-class because the quantities  $\psi_\tau(y - q_\tau(x, \theta_\tau))$  and  $q_\tau^{(1)}(x, \theta_\tau)$  are bounded, thanks to condition (A4). The family of indicator functions is a VC-class as a consequence of the two following arguments:

- Van Der Vaart and Wellner (1996) established that the set of cells in  $\mathbb{R}^d$  is a VC-class (see Example 2.6.1 on page 135).
- Van der Vaart (1998) proved that a collection of sets is a VC-class of sets if and only if the collection of corresponding indicator functions is a VC-class of functions (see page 275).

**Condition 2:** The family of functions  $\varphi_{x_0}$  should admit an envelope  $\Phi$  that verifies

$$\mathbb{E} \left[ \Phi(X, Z)^2 \gamma_0(Z)^2 \delta \right] < \infty.$$

Bearing in mind that  $\varphi_{x_0}(x, z; \theta_\tau) = \psi_\tau(z - q_\tau(x, \theta_\tau)) q_\tau^{(1)}(x, \theta_\tau) \mathbb{1}(x \leq x_0)$ , we can consider

$$\Phi(x, z) = |\psi_\tau(z - q_\tau(x, \theta_\tau)) q_\tau^{(1)}(x, \theta_\tau)|,$$

that is uniformly bounded because of assumption (A4). So, this second condition is a consequence of (A2).

**Condition 3:** This condition holds directly by (A3). □

**Proof of Corollary 3.** To obtain the weak convergence of the empirical process  $R_{n,1}^0$ , we can focus on the dominant terms in the i.i.d. representation given in (8). The result will be proved by showing that the class

$$\mathcal{F} = \{(x, z, \delta) \rightarrow \eta(x, z, \delta; x_0, \theta_\tau) : x_0 \in R_X\}$$

is Donsker. But, before showing the Donsker property we will first show that  $\mathbb{E}[\eta(X, Z, \delta; x_0, \theta_\tau)] = 0$ . Indeed,

$$\begin{aligned} \mathbb{E}[\eta(X, Z, \delta; x_0, \theta_\tau)] &= \mathbb{E} \left[ \psi_\tau(Z - q_\tau(X, \theta_\tau)) q_\tau^{(1)}(X, \theta_\tau) \mathbb{1}(X \leq x_0) \gamma_0(Z) \delta \right] \\ &\quad + \mathbb{E}[\gamma_1^{\varphi_{x_0}(\cdot; \theta_\tau)}(Z) (1 - \delta)] - \mathbb{E}[\gamma_2^{\varphi_{x_0}(\cdot; \theta_\tau)}(Z)]. \end{aligned}$$

Following the results presented in Stute (1995) (without covariates) and Stute (1996) (with covariates), it follows that  $\mathbb{E}[\gamma_1^{\varphi_{x_0}(\cdot; \theta_\tau)}(Z) (1 - \delta)] = \mathbb{E}[\gamma_2^{\varphi_{x_0}(\cdot; \theta_\tau)}(Z)]$ . Under conditions (A1) and (A2), we can apply Lemma 3.17 in Sánchez-Sellero (2001), and conclude that

$$\begin{aligned} \mathbb{E}[\eta(X, Z, \delta; x_0, \theta_\tau)] &= \mathbb{E}[\psi_\tau(Y - q_\tau(X, \theta_\tau)) q_\tau^{(1)}(X, \theta_\tau) \mathbb{1}(X \leq x_0)] \\ &= \mathbb{E} \left[ \mathbb{E}[\psi_\tau(Y - q_\tau(X, \theta_\tau)) | X] q_\tau^{(1)}(X, \theta_\tau) \mathbb{1}(X \leq x_0) \right] = 0 \end{aligned} \quad (12)$$

for each  $x_0 \in \mathbb{R}^d$ , since  $\mathbb{E}[\psi_\tau(\varepsilon) | X] = 0$ .

For showing the Donsker property, the proof parallels the proof of Lemma 8 below, in which it is shown that the functions  $(x, z, \delta) \rightarrow \eta(x, z, \delta; x_0, \theta)$ , indexed by  $x_0 \in R_X$  and  $\theta \in \Theta$ , are Donsker. The proof here is analogous but simpler, since we fix  $\theta$  at its true value  $\theta_\tau$ . As a consequence, we refer to the latter proof for more details. Note that since  $\theta = \theta_\tau$ , it is easily seen that assumption (A8) is not needed here.

Finally, to show the second statement in Corollary 3, it suffices to notice that  $T_{n,1}^0$  is a continuous function of  $R_{n,1}^0$ ,

and hence its convergence follows from the continuous mapping theorem.  $\square$

**Proof of Theorem 4.** We first focus on  $\tilde{\theta}_\tau - \theta_\tau$  and will start by showing that  $\tilde{\theta}_\tau - \theta_\tau = o_p(1)$ . We prove this result by verifying the conditions of Theorem 5.7 in Van der Vaart (1998) (see page 45), where in our case  $M_n = -\tilde{S}$  and  $M = -S$  (similar to Lemma 4.2 of Van Keilegom et al. (2008)), with  $S(\theta) = \mathbb{E}_{H_0} [\rho_\tau(Y - q_\tau(X, \theta))]$  and with  $\tilde{S}$  as defined in (9). Note that

$$\begin{aligned} & |\tilde{S}(\theta) - S(\theta)| \\ & \leq \mathbb{E} \left[ \left| \rho_\tau \left( q_\tau(X, \theta_\tau) - q_\tau(X, \theta) + \varepsilon + \frac{1}{\sqrt{n}} h(X) \right) - \rho_\tau \left( q_\tau(X, \theta_\tau) - q_\tau(X, \theta) + \varepsilon \right) \right| \right] \\ & \leq \frac{1}{\sqrt{n}} \mathbb{E} |h(X)|, \end{aligned}$$

where the last inequality follows since  $|\rho_\tau(a + b) - \rho_\tau(a)| \leq |b| \max(\tau, 1 - \tau) \leq |b|$  for any  $a$  and  $b$ . Hence, since  $\mathbb{E}|h(X)| < \infty$ , the first condition of Theorem 5.7 in Van der Vaart (1998) follows. Moreover, the second condition is a consequence of Assumption (A5).

On the other hand, taking into account the definition of  $\tilde{\theta}_\tau$  and using a Taylor expansion we can write

$$0 = \frac{\partial}{\partial \theta} \tilde{S}(\tilde{\theta}_\tau) = \frac{\partial}{\partial \theta} \tilde{S}(\theta_\tau) + \frac{\partial^2}{\partial \theta \partial \theta'} \tilde{S}(\theta^*) (\tilde{\theta}_\tau - \theta_\tau),$$

where  $\theta^*$  is an element on the line segment between  $\tilde{\theta}_\tau$  and  $\theta_\tau$ . For any  $\theta$ , we have

$$\begin{aligned} \frac{\partial}{\partial \theta} \tilde{S}(\theta) &= \mathbb{E} \left[ \frac{\partial}{\partial \theta} \rho_\tau(Y - q_\tau(X, \theta)) \right] = -\mathbb{E} \left[ \psi_\tau(Y - q_\tau(X, \theta)) q_\tau^{(1)}(X, \theta) \right] \\ &= -\mathbb{E} \left[ (\tau - F_{Y|X}(q_\tau(X, \theta))) q_\tau^{(1)}(X, \theta) \right], \end{aligned}$$

and evaluated at  $\theta_\tau$  we have

$$\begin{aligned} \frac{\partial}{\partial \theta} \tilde{S}(\theta_\tau) &= -\mathbb{E} \left[ \left( \tau - F_{\varepsilon|X} \left( -\frac{1}{\sqrt{n}} h(X) \right) \right) q_\tau^{(1)}(X, \theta_\tau) \right] \\ &= -\mathbb{E} \left[ \left( F_{\varepsilon|X}(0) - F_{\varepsilon|X} \left( -\frac{1}{\sqrt{n}} h(X) \right) \right) q_\tau^{(1)}(X, \theta_\tau) \right] \\ &= -\frac{1}{\sqrt{n}} \mathbb{E} \left[ f_{\varepsilon|X}(0) h(X) q_\tau^{(1)}(X, \theta_\tau) \right] + o \left( \frac{1}{\sqrt{n}} \right). \end{aligned}$$

Next, the matrix of second order derivatives equals

$$\begin{aligned} \frac{\partial^2}{\partial \theta \partial \theta'} \tilde{S}(\theta) &= -\frac{\partial}{\partial \theta} \mathbb{E} \left[ (\tau - F_{Y|X}(q_\tau(X, \theta))) q_\tau^{(1)}(X, \theta) \right] \\ &= -\mathbb{E} \left[ \frac{\partial}{\partial \theta} \left( (\tau - F_{Y|X}(q_\tau(X, \theta))) q_\tau^{(1)}(X, \theta) \right) \right] \\ &= -\mathbb{E} \left[ -f_{Y|X}(q_\tau(X, \theta)) q_\tau^{(1)}(X, \theta) q_\tau^{(1)}(X, \theta)' + (\tau - F_{Y|X}(q_\tau(X, \theta))) q_\tau^{(2)}(X, \theta) \right], \end{aligned}$$

where  $q_\tau^{(2)}(X, \theta) = \frac{\partial^2}{\partial \theta \partial \theta'} q_\tau(X, \theta)$  is the Hessian matrix of the regression function. As  $\tilde{\theta}_\tau - \theta_\tau = o(1)$  and  $\theta^*$  is an

element between  $\tilde{\theta}_\tau$  and  $\theta_\tau$ , it follows from conditions (A4) and (A7) that we can write

$$\begin{aligned}
& \frac{\partial^2}{\partial \theta \partial \theta'} \tilde{S}(\theta^*) \\
&= \frac{\partial^2}{\partial \theta \partial \theta'} \tilde{S}(\theta_\tau) + o(1) \\
&= -\mathbb{E} \left[ -\hat{r}_{Y|X}(q_\tau(X, \theta_\tau)) q_\tau^{(1)}(X, \theta_\tau) q_\tau^{(1)}(X, \theta_\tau)' + (\tau - F_{Y|X}(q_\tau(X, \theta_\tau))) q_\tau^{(2)}(X, \theta_\tau) \right] + o(1) \\
&= -\mathbb{E} \left[ -\hat{f}_{\varepsilon|X} \left( -\frac{1}{\sqrt{n}} h(X) \right) q_\tau^{(1)}(X, \theta_\tau) q_\tau^{(1)}(X, \theta_\tau)' \right. \\
&\quad \left. + \left( F_{\varepsilon|X}(0) - F_{\varepsilon|X} \left( -\frac{1}{\sqrt{n}} h(X) \right) \right) q_\tau^{(2)}(X, \theta_\tau) \right] + o(1) \\
&= -\mathbb{E} \left[ -\hat{f}_{\varepsilon|X}(0) q_\tau^{(1)}(X, \theta_\tau) q_\tau^{(1)}(X, \theta_\tau)' + \frac{1}{\sqrt{n}} h(X) \hat{f}_{\varepsilon|X}(0) q_\tau^{(2)}(X, \theta_\tau) \right] + o(1) \\
&= \mathbb{E} \left[ \hat{f}_{\varepsilon|X}(0) q_\tau^{(1)}(X, \theta_\tau) q_\tau^{(1)}(X, \theta_\tau)' \right] + o(1). \tag{13}
\end{aligned}$$

Finally, we can conclude that

$$\tilde{\theta}_\tau - \theta_\tau = \frac{1}{\sqrt{n}} \mathbb{E} \left[ \hat{f}_{\varepsilon|X}(0) q_\tau^{(1)}(X, \theta_\tau) q_\tau^{(1)}(X, \theta_\tau)' \right]^{-1} \mathbb{E} \left[ \hat{f}_{\varepsilon|X}(0) h(X) q_\tau^{(1)}(X, \theta_\tau) \right] + o \left( \frac{1}{\sqrt{n}} \right).$$

Next, we consider the term  $\hat{\theta}_\tau - \tilde{\theta}_\tau$ . The proof is rather similar to that of  $\tilde{\theta}_\tau - \theta_\tau$ . To show that  $\hat{\theta}_\tau - \tilde{\theta}_\tau = o_p(1)$ , we check the conditions of Theorem 5.7 in Van der Vaart (1998), with  $M_n = -\tilde{S}_n$  and  $M = -\tilde{S}$ , and with  $\tilde{S}_n(\theta) = \sum_{i=1}^n W_i^{KM} \rho_\tau(Z_i - q_\tau(X_i, \theta)) = \int \rho_\tau(y - q_\tau(x, \theta)) d\tilde{F}_{X,Y}(x, y)$ . To show that  $\sup_{\theta \in \Theta} |\tilde{S}_n(\theta) - \tilde{S}(\theta)| = o_p(1)$ , we will use Theorem 3.10 in Sánchez-Sellero (2001) where the uniform consistency of Kaplan-Meier integrals is shown. Theorem 1 in Sánchez-Sellero et al. (2005) can be seen as an extension of the result of Sánchez-Sellero (2001) for censored and truncated data. However, note that Theorem 1 in Sánchez-Sellero et al. (2005) shows the almost sure representation while we only need the uniform consistency, so our conditions are less strong. In order to apply Theorem 3.10 in Sánchez-Sellero (2001), we have to check the following three conditions:

**Condition 1:** The family of functions  $\mathcal{F}_\rho = \{(x, y) \rightarrow \rho_\tau(y - q_\tau(x, \theta)) : \theta \in \Theta\}$  should be a VC-class. Careful verification of the proof of Theorem 3.10 in Sánchez-Sellero (2001) shows that this condition is required to show that the class is Glivenko-Cantelli. Instead of showing that the class is VC, we can equally well calculate the bracketing number  $N_{[\cdot]}(\varepsilon, \mathcal{F}_\rho, L_1(P))$  of the class, and show that it is finite for all  $\varepsilon > 0$ . Theorem 2.4.1 in Van Der Vaart and Wellner (1996) shows that the class is then Glivenko-Cantelli. To calculate the bracketing number, note that the functions in  $\mathcal{F}_\rho$  are the composition of two functions, namely

$$(x, y) \rightarrow y - q_\tau(x, \theta),$$

and

$$u \rightarrow \rho_\tau(u) = u\tau \mathbb{1}(u \geq 0) - u(1 - \tau)\mathbb{1}(u < 0) = \rho_\tau^1(u) - \rho_\tau^2(u),$$

which is the difference of two monotone increasing functions. The functions in the class  $\{x \rightarrow q_\tau(x, \theta) : \theta \in \Theta\}$  are Lipschitz in  $\theta$  uniformly in  $x$ , thanks to (A4). Hence, it follows from Theorem 2.7.11 in Van Der Vaart

and Wellner (1996) that the  $\epsilon^2$ -bracketing number of this class is bounded by the  $\epsilon^2$ -covering number of the class  $\Theta$ , which is of the order of  $M_\epsilon = O(\epsilon^{-2q})$ . Denote the brackets of the class  $\{(x, y) \rightarrow y - q_\tau(x, \theta)\}$  by  $[L_k, U_k]$ ,  $k = 1, \dots, M_\epsilon$ , where  $L_k(x, y) = y - U_k(x)$ , with  $U_k(x)$  one of the upper brackets of the class  $\{x \rightarrow q_\tau(x, \theta) : \theta \in \Theta\}$ , and similarly for  $U_k(x, y)$ . Then, it is easy to see that the  $\epsilon$ -brackets for  $\mathcal{F}_\rho$  are given by  $(x, y) \rightarrow [\rho_\tau^1(L_k(x, y)) - \rho_\tau^2(U_k(x, y)), \rho_\tau^1(U_k(x, y)) - \rho_\tau^2(L_k(x, y))]$ . This shows that the class is Glivenko-Cantelli.

**Condition 2:** The family of functions  $\mathcal{F}_\rho$  should admit an envelope  $\Phi$  that satisfies  $\mathbb{E}[\Phi(X, Y)] < \infty$ . Note that

$$\begin{aligned} |\rho_\tau(y - q_\tau(x, \theta))| &= |(y - q_\tau(x, \theta)) (\tau - \mathbb{1}(y - q_\tau(x, \theta) < 0))| \\ &\leq \max\{\tau, 1 - \tau\} |y - q_\tau(x, \theta)|. \end{aligned}$$

This shows that we can take  $\Phi(x, y) = \sup_{\theta \in \Theta} |y - q_\tau(x, \theta)|$ , and hence the condition is satisfied thanks to (A5).

**Condition 3:** The integral  $\int \frac{1}{(1-H(y))^2} d\bar{H}^0(y)$  should be finite, which holds directly by (A3).

Next, to obtain an i.i.d. representation for  $\widehat{\theta}_\tau - \bar{\theta}_\tau$ , we again follow the proof for the bias term  $\widehat{\theta}_\tau - \theta_\tau$ , and write

$$0 = \frac{\partial}{\partial \theta} \bar{S}_n(\widehat{\theta}_\tau) = \frac{\partial}{\partial \theta} \bar{S}_n(\bar{\theta}_\tau) + \frac{\partial^2}{\partial \theta \partial \theta'} \bar{S}_n(\theta^*)(\widehat{\theta}_\tau - \bar{\theta}_\tau),$$

where  $\theta^*$  is an element on the line segment between  $\widehat{\theta}_\tau$  and  $\bar{\theta}_\tau$ . For any  $\theta$ , we have

$$\begin{aligned} \frac{\partial}{\partial \theta} \bar{S}_n(\theta) &= \int \frac{\partial}{\partial \theta} \rho_\tau(y - q_\tau(x, \theta)) d\widehat{F}_{X,Y}(x, y) \\ &= - \int \psi_\tau(y - q_\tau(x, \theta)) q_\tau^{(1)}(x, \theta) d\widehat{F}_{X,Y}(x, y), \end{aligned}$$

and evaluated at  $\bar{\theta}_\tau$  we have

$$\begin{aligned} \frac{\partial}{\partial \theta} \bar{S}_n(\bar{\theta}_\tau) &= - \int \psi_\tau(y - q_\tau(x, \bar{\theta}_\tau)) q_\tau^{(1)}(x, \bar{\theta}_\tau) d\widehat{F}_{X,Y}(x, y) \\ &= - \sum_{i=1}^n W_i^{\text{KM}} \psi_\tau(Z_i - q_\tau(X_i, \bar{\theta}_\tau)) q_\tau^{(1)}(X_i, \bar{\theta}_\tau) \\ &= -n^{-1} \sum_{i=1}^n \eta(X_i, Z_i, \delta_i; \infty, \bar{\theta}_\tau) + o_p(n^{-1/2}) = -n^{-1} \sum_{i=1}^n \xi(X_i, Z_i, \delta_i; \bar{\theta}_\tau) + o_p(n^{-1/2}) \end{aligned}$$

as a consequence of Theorem 2 and Remark 1. Moreover, using the convergence of Kaplan-Meier integrals (see the arguments presented in the proof of Theorem 2) it follows that

$$\begin{aligned} \frac{\partial^2}{\partial \theta \partial \theta'} \bar{S}_n(\theta_\tau) &= - \int \psi_\tau(y - q_\tau(x, \theta_\tau)) q_\tau^{(1)}(x, \theta_\tau) q_\tau^{(1)}(x, \theta_\tau)' d\widehat{F}_{X,Y}(x, y) \\ &= - \int \psi_\tau(y - q_\tau(x, \theta_\tau)) q_\tau^{(1)}(x, \theta_\tau) q_\tau^{(1)}(x, \theta_\tau)' dF_{X,Y}(x, y) + o_p(1) \\ &= \frac{\partial^2}{\partial \theta \partial \theta'} \bar{S}(\theta_\tau) + o_p(1). \end{aligned}$$

Then, as  $\widehat{\theta}_\tau - \theta_\tau = o_p(1)$ ,  $\widehat{\theta}_\tau - \bar{\theta}_\tau = o(1)$  and  $\theta^*$  is an element between  $\widehat{\theta}_\tau$  and  $\bar{\theta}_\tau$ , it follows from equation (13) that we

can write

$$\frac{\partial^2}{\partial \theta \partial \theta'} \tilde{S}_n(\theta^*) = \frac{\partial^2}{\partial \theta \partial \theta'} \tilde{S}(\theta_\tau) + o_p(1) = \mathbb{E} \left[ f_{\varepsilon|X}(0) q_\tau^{(1)}(X, \theta_\tau) q_\tau^{(1)}(X, \theta_\tau)' \right] + o_p(1).$$

So

$$\hat{\theta}_\tau - \tilde{\theta}_\tau = \frac{1}{n} \Omega^{-1} \sum_{i=1}^n \xi(X_i, Z_i, \delta_i; \tilde{\theta}_\tau) + o_p(1),$$

and the asymptotic representation of the estimator in (5) is derived.  $\square$

The proof of Theorem 6 uses the following lemma, which we prove first.

**Lemma 8** *Under assumptions (A4), (A7)-(A8), we have*

$$\begin{aligned} \sup_{x_0 \in R_X} \sup_{\|\theta - \theta_\tau\| \leq \delta_n} \left\| \frac{1}{n} \sum_{i=1}^n \eta(X_i, Z_i, \delta_i; x_0, \theta) - \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n \eta(X_i, Z_i, \delta_i; x_0, \theta) \right] \right. \\ \left. - \frac{1}{n} \sum_{i=1}^n \eta(X_i, Z_i, \delta_i; x_0, \theta_\tau) \right\| = o_p(n^{-1/2}), \end{aligned} \quad (14)$$

for all positive sequences  $\delta_n = o(1)$ .

**Proof.** We will show the statement by proving that the class

$$\mathcal{F} = \{(x, z, \delta) \rightarrow \eta(x, z, \delta; x_0, \theta) : \theta \in \Theta, x_0 \in R_X\} \quad (15)$$

is Donsker. We will do this by showing that

$$\int \left\{ \log N_{[\cdot]}(\varepsilon, \mathcal{F}, L_2(P)) \right\}^{1/2} d\varepsilon < \infty,$$

where  $N_{[\cdot]}(\varepsilon, \mathcal{F}, L_2(P))$  is the  $\varepsilon$ -bracketing number of the class  $\mathcal{F}$  with respect to the  $L_2$  metric, i.e. the smallest number of  $\varepsilon$ -brackets needed to cover the space  $\mathcal{F}$ .

We will first construct brackets for the functions

$$\varphi_{x_0}(x, z; \theta) = \psi_\tau(z - q_\tau(x, \theta)) q_\tau^{(1)}(x, \theta) \mathbb{1}(x \leq x_0),$$

indexed by  $\theta$  and  $x_0$ . Fix  $\varepsilon > 0$ . Since it follows from (A4) that the functions in the class  $\{x \rightarrow q_\tau(x, \theta) : \theta \in \Theta\}$  and in the class  $\{x \rightarrow q_\tau^{(1)}(x, \theta) : \theta \in \Theta\}$  are Lipschitz in  $\theta$  uniformly in  $x$ , we know from Theorem 2.7.11 in Van Der Vaart and Wellner (1996) that there exist  $K_n = O(\varepsilon^{-2q})$   $\varepsilon^2$ -brackets  $[\ell_k, u_k]$  and  $K_n$   $\varepsilon^2$ -brackets  $[\bar{\ell}_k, \bar{u}_k]$ ,  $k = 1, \dots, K_n$ , such that  $\mathbb{E}[(u_k(X) - \ell_k(X))^2] < \varepsilon^4$  and  $\mathbb{E}[(\bar{u}_k(X) - \bar{\ell}_k(X))^2] < \varepsilon^4$ , and such that for each  $\theta \in \Theta$ , there exist  $j, k$  with  $\ell_k(x) \leq q_\tau(x, \theta) \leq u_k(x)$  and  $\bar{\ell}_j(x) \leq q_\tau^{(1)}(x, \theta) \leq \bar{u}_j(x)$ . Then,

$$\tau - \mathbb{1}(z \leq u_k(x)) \leq \psi_\tau(z - q_\tau(x, \theta)) \leq \tau - \mathbb{1}(z \leq \ell_k(x)). \quad (16)$$

Moreover, since  $R_X$  is compact, there exist  $I_n = O(\varepsilon^{-2d})$   $\varepsilon^2$ -brackets  $[\tilde{\ell}_i, \tilde{u}_i]$ ,  $i = 1, \dots, I_n$  such that for all  $x_0 \in R_X$

there is a  $i$  such that  $\tilde{\ell}_i(x) \leq \mathbb{I}(x \leq x_0) \leq \bar{u}_i(x)$ . Hence, combining these brackets we have  $M_n = K_n^2 I_n = O(\epsilon^{-4q-2d})$  brackets for the functions  $(x, z) \rightarrow \varphi_{x_0}(x, z; \theta)$  indexed by  $\theta$  and  $x_0$ , which we denote by  $[L_\ell, U_\ell]$ ,  $\ell = 1, \dots, M_n$ . By construction, for all  $\theta$  and  $x_0$  there exists a  $\ell$  with the property that  $L_\ell(x, z) \leq \varphi_{x_0}(x, z; \theta) \leq U_\ell(x, z)$ . We need to show that

$$\mathbb{E}[(U_\ell(X, Z) - L_\ell(X, Z))^2] \leq \epsilon^2.$$

The expected value at the left hand side can be split into several terms, since  $L_\ell$  and  $U_\ell$  are the product of three brackets. The most difficult term is the one involving the brackets for  $\psi_\tau(z - q_\tau(x, \theta))$ , which can be bounded by

$$\begin{aligned} & K \mathbb{E} \left\{ \int \left[ \mathbb{I}(z \leq u_k(X)) - \mathbb{I}(z \leq \ell_k(X)) \right]^2 dF_{Z|X}(z) \right\} \\ & \leq K \mathbb{E} [F_{Z|X}(u_k(X)) - F_{Z|X}(\ell_k(X))] \\ & \leq \bar{K} \mathbb{E}[u_k(X) - \ell_k(X)] \leq \bar{K} \left( \mathbb{E}[u_k(X) - \ell_k(X)]^2 \right)^{1/2} \leq \bar{K} \epsilon^2, \end{aligned}$$

since  $F_{Z|X=x}(z)$  is assumed to be Lipschitz in  $z$  uniformly in  $x$ .

Next, consider the class  $\mathcal{F}$  given in (15). The function  $\eta$  consists of three terms, depending on  $\gamma_0, \gamma_1$  and  $\gamma_2$ , which we denote by  $A, B$  and  $C$ . Hence,  $\mathcal{F} = \mathcal{F}_A + \mathcal{F}_B + \mathcal{F}_C$ . First, note that  $A(x, z, \delta; x_0, \theta) = \varphi_{x_0}(x, z; \theta) \gamma_0(z) \delta$ . Hence, the brackets for  $A(x, z, \delta; x_0, \theta)$  are given by  $L_\ell(x, z) \gamma_0(z) \delta \leq U_\ell(x, z) \gamma_0(z) \delta$ ,  $\ell = 1, \dots, M_n$ . To show that  $\mathbb{E}[(U_\ell(X, Z) - L_\ell(X, Z))^2 \gamma_0^2(Z) \delta] \leq \epsilon^2$ , we can as before decompose this expected value in several terms. The most difficult term is again the one involving the brackets for  $\psi_\tau(z - q_\tau(x, \theta))$ , which can be bounded by

$$K \mathbb{E} \left[ \int_{\ell_k(X)}^{u_k(X)} \gamma_0^2(z) dH^1(z|X) \right],$$

where  $H^1(z|x) = \mathbb{P}(Z \leq z, \delta = 1|X = x)$ . Using Cauchy-Schwarz inequality this is bounded by  $\bar{K} \epsilon^2 \mathbb{E}[\gamma_0^4(Z) \delta]$ , which is proportional to  $\epsilon^2$  thanks to assumption (A8). Hence,  $\mathcal{F}_A$  is Donsker, since

$$\int \{ \log N_{[]}(\epsilon, \mathcal{F}_A, L_2(P)) \}^{1/2} d\epsilon \leq K \int_0^{2M} \{ \log(\epsilon^{-4q-2d}) \}^{1/2} d\epsilon < \infty,$$

where  $\sup_{x, z, x_0, \theta} |\varphi_{x_0}(x, z; \theta)| \leq M < \infty$ , since for  $\epsilon > 2M$ , one bracket suffices, namely  $-M \gamma_0(z) \delta \leq M \gamma_0(z) \delta$ .

The term  $B(z, \delta; x_0, \theta)$  can be dealt with in a somewhat similar way: we will show that the brackets are given by

$$\begin{aligned} \tilde{L}_\ell(z, \delta) &= \frac{1 - \delta}{1 - H(z)} \int \mathbb{I}(z \leq w) L_\ell(x, w) \gamma_0(w) \tilde{H}^{11}(dx, dw) \\ &\leq \tilde{U}_\ell(z, \delta) = \frac{1 - \delta}{1 - H(z)} \int \mathbb{I}(z \leq w) U_\ell(x, w) \gamma_0(w) \tilde{H}^{11}(dx, dw). \end{aligned}$$

Indeed,

$$\begin{aligned} & \mathbb{E}[(\tilde{U}_\ell(X, Z) - \tilde{L}_\ell(X, Z))^2] \\ & \leq \mathbb{E} \left[ \frac{1 - \delta}{(1 - H(Z))^2} \left( \int \mathbb{I}(Z \leq w) \{U_\ell(x, w) - L_\ell(x, w)\} \gamma_0(w) \tilde{H}^{11}(dx, dw) \right)^2 \right] \\ & \leq \mathbb{E} \left[ \frac{1 - \delta}{(1 - H(Z))^2} \int \mathbb{I}(Z \leq w) \{U_\ell(x, w) - L_\ell(x, w)\}^2 \gamma_0^2(w) \tilde{H}^{11}(dx, dw) \right]. \end{aligned}$$

Since  $\int \mathbb{I}(z \leq w)(1 - H(z))^{-2} H(dz) = H(w)/(1 - H(w))$ , the latter equals

$$\mathbb{E} \left[ \frac{H(Z)}{1 - H(Z)} (U_\ell(X, Z) - L_\ell(X, Z))^2 \gamma_0^2(Z) \delta \right].$$

Using a similar decomposition as above, we can conclude that the latter is bounded by  $K\epsilon^2$  for some  $K < \infty$ , provided  $\mathbb{E} [H^2(Z)/(1 - H(Z))^2 \gamma_0^4(Z) \delta] < \infty$ . Hence, the class  $\mathcal{F}_B$  is Donsker.

Finally, for the term  $C(z; x_0, \theta)$  a similar decomposition as for  $B(z, \delta; x_0, \theta)$  shows that the class  $\mathcal{F}_C$  is also Donsker.

□

**Proof of Theorem 6.** As a consequence of Theorem 2 and Remark 1, we can write  $R_{n,1}^1$  as follows:

$$\begin{aligned} R_{n,1}^1(x_0) &= n^{1/2} \sum_{i=1}^n W_i^{\text{KM}} \psi_\tau(Z_i - q_\tau(X_i, \widehat{\theta}_\tau)) q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) \mathbb{I}(X_i \leq x_0) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta(X_i, Z_i, \delta_i; x_0, \widehat{\theta}_\tau) + o_p(1) \\ &= \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta(X_i, Z_i, \delta_i; x_0, \widehat{\theta}_\tau) - \mathbb{E} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta(X_i, Z_i, \delta_i; x_0, \widehat{\theta}_\tau) \right] \right\} \\ &\quad + \mathbb{E} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta(X_i, Z_i, \delta_i; x_0, \widehat{\theta}_\tau) \right] + o_p(1) \\ &= R_{n,1}^{1a}(x_0) + R_{n,1}^{1b}(x_0) + o_p(1). \end{aligned} \tag{17}$$

From Lemma 8 we can conclude that  $R_{n,1}^{1a}(x_0) = R_{n,1}^0(x_0) + o_p(1)$  uniformly in  $x_0$ . On the other hand, following the argument used to obtain (12), we can write

$$\begin{aligned} R_{n,1}^{1b}(x_0) &= \sqrt{n} \mathbb{E} \left[ \eta(X, Z, \delta; x_0, \widehat{\theta}_\tau) \right] \\ &= \sqrt{n} \mathbb{E} \left[ \psi_\tau(Y - q_\tau(X, \widehat{\theta}_\tau)) q_\tau^{(1)}(X, \widehat{\theta}_\tau) \mathbb{I}(X \leq x_0) \right] \\ &= \sqrt{n} \mathbb{E} \left[ \mathbb{E} \left\{ \psi_\tau(Y - q_\tau(X, \widehat{\theta}_\tau)) \middle| X \right\} q_\tau^{(1)}(X, \widehat{\theta}_\tau) \mathbb{I}(X \leq x_0) \right]. \end{aligned}$$

For brevity, let us denote

$$\ell(x, \theta) = q_\tau(x, \theta) - q_\tau(x, \theta_\tau) - \frac{1}{\sqrt{n}} h(x).$$

Then,

$$\begin{aligned} \mathbb{E} \left\{ \psi_\tau(Y - q_\tau(X, \widehat{\theta}_\tau)) \middle| X \right\} &= \mathbb{E} \left\{ \tau - \mathbb{I}(\varepsilon < \ell(X, \widehat{\theta}_\tau)) \middle| X \right\} \\ &= \tau - F_{\varepsilon|X}(\ell(X, \widehat{\theta}_\tau)) \\ &= F_{\varepsilon|X}(0) - F_{\varepsilon|X}(\ell(X, \widehat{\theta}_\tau)) \\ &= -f_{\varepsilon|X}(\xi_1) \left( q_\tau(X, \widehat{\theta}_\tau) - q_\tau(X, \theta_\tau) - \frac{1}{\sqrt{n}} h(X) \right) \\ &= -f_{\varepsilon|X}(\xi_1) \left( q_\tau^{(1)}(X, \xi_2)' (\widehat{\theta}_\tau - \theta_\tau) - \frac{1}{\sqrt{n}} h(X) \right) \end{aligned}$$

$$= -f_{\varepsilon|X}(0) \left( q_{\tau}^{(1)}(X, \theta_{\tau})' (\widehat{\theta}_{\tau} - \theta_{\tau}) - \frac{1}{\sqrt{n}} h(X) \right) + o_p \left( \frac{1}{\sqrt{n}} \right),$$

where  $\xi_1$  is an element between  $q_{\tau}(X, \widehat{\theta}_{\tau}) - q_{\tau}(X, \theta_{\tau}) - h(X)/\sqrt{n}$  and 0, and  $\xi_2$  is on the line segment between  $\widehat{\theta}_{\tau}$  and  $\theta_{\tau}$ . Then, it follows from Theorem 4 that

$$\begin{aligned} R_{n,1}^b(x_0) &= -\mathbb{E} \left[ f_{\varepsilon|X}(0) q_{\tau}^{(1)}(X, \theta_{\tau}) q_{\tau}^{(1)}(X, \theta_{\tau})' \mathbb{1}(X \leq x_0) \right] \sqrt{n} (\widehat{\theta}_{\tau} - \theta_{\tau}) \\ &\quad + \mathbb{E} \left[ f_{\varepsilon|X}(0) h(X) q_{\tau}^{(1)}(X, \theta_{\tau}) \mathbb{1}(X \leq x_0) \right] + o_p(1) \\ &= -\mathbb{E} \left[ f_{\varepsilon}(0|X) q_{\tau}^{(1)}(X, \theta_{\tau}) q_{\tau}^{(1)}(X, \theta_{\tau})' \mathbb{1}(X \leq x_0) \right] \\ &\quad \times \left\{ \frac{1}{\sqrt{n}} \Omega^{-1} \sum_{i=1}^n \xi(X_i, Z_i, \delta_i; \widetilde{\theta}_{\tau}) + \Omega^{-1} \mathbb{E} \left[ f_{\varepsilon|X}(0) h(X) q_{\tau}^{(1)}(X, \theta_{\tau}) \right] + o_p(1) \right\} \\ &\quad + \mathbb{E} \left[ f_{\varepsilon|X}(0) h(X) q_{\tau}^{(1)}(X, \theta_{\tau}) \mathbb{1}(X \leq x_0) \right] + o_p(1) \\ &= -\frac{1}{\sqrt{n}} \Omega(x_0) \Omega^{-1} \sum_{i=1}^n \xi(X_i, Z_i, \delta_i; \widetilde{\theta}_{\tau}) - \Omega(x_0) \Omega^{-1} \mathbb{E} \left[ f_{\varepsilon|X}(0) h(X) q_{\tau}^{(1)}(X, \theta_{\tau}) \right] \\ &\quad + \mathbb{E} \left[ f_{\varepsilon|X}(0) h(X) q_{\tau}^{(1)}(X, \theta_{\tau}) \mathbb{1}(X \leq x_0) \right] + o_p(1). \end{aligned} \tag{18}$$

The result now follows from the above representation together with (17).  $\square$

## References

- M. G. Akritas. Bootstrapping the kaplan-meier estimator. *J. Amer. Statist. Assoc.*, 81:1032–1038, 1986.
- R. Beran. Nonparametric regression with randomly censored survival data. *Technical Report. University of California, Berkeley.*, 1981.
- M. Birke, N. Neumeyer, and S. Volgushev. The independence process in conditional quantile location-scale models and an application to testing for monotonicity. *Statist. Sinica*, 27:1815–1839, 2017.
- R. Cao, I. López-de Ullibarri, P. Janssen, and N. Veraverbeke. Presmoothed Kaplan-Meier and Nelson-Aalen estimators. *J. Nonparametr. Stat.*, 17:31–56, 2005.
- B Colling and I. Van Keilegom. Goodness-of-fit tests in semiparametric transformation models using the integrated regression function. *J. Multivariate Anal.*, 160:10–30, 2017.
- M. Conde-Amboage, C. Sánchez-Sellero, and W. González-Manteiga. A lack-of-fit test for quantile regression models with high-dimensional covariates. *Comput. Statist. Data Anal.*, 88:128–138, 2015.
- M. De Backer, A. El Ghouch, and I. Van Keilegom. Semiparametric copula quantile regression for complete or censored data. *Electron. J. Stat.*, 11:1660–1698, 2017.
- M. De Backer, A. El Ghouch, and I. Van Keilegom. An adapted loss function for censored quantile regression. *J. Amer. Statist. Assoc.*, 114:1126–1137, 2019.
- M. De Backer, A. El Ghouch, and I. Van Keilegom. Linear censored quantile regression: a novel minimum-distance approach. *Scand. J. Stat.*, 47:1275–1306, 2020.
- J. De Uña-Álvarez and J. Roca-Pardiñas. Additive models in censored regression. *Comput. Statist. Data Anal.*, 2009.
- G. Dikta. On semiparametric random censorship models. *J. Statist. Plann. Inference*, 66:253–279, 1998.

- A. Gannoun, J. Saracco, A. Yuan, and G. E. Bonney. Non-parametric quantile regression with censored data. *Scand. J. Stat.*, 32: 527–550, 2005.
- C. Geerdens, P. Janssen, and I. Van Keilegom. Goodness-of-fit test for a parametric survival function with cure fraction. *TEST*, 29:768–792, 2020.
- W. Gonzalez-Manteiga and C. Cadarso-Suarez. Asymptotic properties of a generalized kaplan-meier estimator with some applications. *J. Nonparametr. Stat.*, 4:65–78, 1994.
- P. Groeneboom and G. Jongbloed. Nonparametric confidence intervals for monotone functions. *Ann. Statist.*, 43:2019–2054, 2015.
- X. He and L.-X. Zhu. A lack-of-fit test for quantile regression. *J. Amer. Statist. Assoc.*, 98:1013–1022, 2003.
- J. L. Horowitz and V. G. Spokoiny. An adaptive, rate-optimal test of linearity for median regression models. *J. Amer. Statist. Assoc.*, 97:822–835, 2002.
- D. W. Hosmer, S. Lemeshow, and S. May. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. John Wiley & Sons, 2008.
- E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, 53:457–481, 1958.
- C. Leng and X. Tong. A quantile regression estimator for censored data. *Bernoulli*, 19:344–361, 2013.
- G. Li and S. Datta. A bootstrap approach to nonparametric regression for right censored data. *Ann. Inst. Statist. Math.*, 53: 708–729, 2001.
- S. H. Lo and K. Singh. The product-limit estimator and the bootstrap: some asymptotic representations. *Probab. Theory Related Fields*, 71:455–465, 1986.
- O. Lopez and V. Patilea. Nonparametric lack-of-fit tests for parametric mean-regression models with censored data. *J. Multivariate Anal.*, 100:210–230, 2009.
- H. P. Lopuhaä and E. Musta. Isotonized smooth estimators of a monotone baseline hazard in the Cox model. *J. Statist. Plann. Inference*, 191:43–67, 2017.
- H. P. Lopuhaä and E. Musta. Smoothed isotonic estimators of a monotone baseline hazard in the Cox model. *Scand. J. Stat.*, 45:753–791, 2018.
- S. Maistre, P. Lavergne, and V. Patilea. Powerful nonparametric checks for quantile regression. *J. Statist. Plann. Inference*, 180: 13–29, 2017.
- L. Meira-Machado and M. Sestelo. condSURV: An R package for the estimation of the conditional survival function for ordered multivariate failure time data. *The R Journal*, 8:460–473, 2016.
- J. Orbe and V. Núñez Antón. Confidence intervals on regression models with censored data. *Comm. Statist. Simulation Comput.*, 42:2140–2159, 2013.
- J. C. Pardo-Fernández, I. Van Keilegom, and W. González-Manteiga. Goodness-of-fit tests for parametric models in censored regression. *Canad. J. Statist.*, 35:249–264, 2007.
- J. C. Pardo-Fernández and I. Van Keilegom. Comparison of regression curves with censored responses. *Scand. J. Stat.*, 33: 409–434, 2006.
- L. Peng and Y. Huang. Survival analysis with quantile regression models. *J. Amer. Statist. Assoc.*, 103:637–649, 2008.
- Y. Peng and J. M. Taylor. Residual-based model diagnosis methods for mixture cure models. *Biometrics*, 73:495–505, 2017.

- S. Portnoy. Censored quantile regression. *J. Amer. Statist. Assoc.*, 98:1001–1012, 2003.
- J. Powell. Censored regression quantiles. *Journal of Econometrics*, 32:143–155, 1986.
- C. Sánchez-Sellero. *Inferencia Estadística en datos con censura y/o truncamiento*. University of Santiago de Compostela (PhD dissertation), 2001.
- C. Sánchez-Sellero, W. González-Manteiga, and I. Van Keilegom. Uniform representation of product-limit integrals with applications. *Scand. J. Stat.*, 32:563–581, 2005.
- P. H. C. Sant’Anna. Program evaluation with right-censored data. Available at SSRN: <https://ssrn.com/abstract=2752760>, 2016.
- J. H. Shows, W. Lu, and H. H. Zhang. Sparse estimation and inference for censored median regression. *J. Statist. Plann. Inference*, 140:1903–1917, 2010.
- W. Stute. Consistent estimation under random censorship when covariables are present. *J. Multivariate Anal.*, 45:89–103, 1993.
- W. Stute. The central limit theorem under random censorship. *Ann. Statist.*, 23:422–439, 1995.
- W. Stute. Distributional convergence under random censorship when covariables are present. *Scand. J. Stat.*, 23:461–471, 1996.
- W. Stute. Nonparametric model checks for regression. *Ann. Statist.*, 25:613–641, 1997.
- W. Stute. Nonlinear censored regression. *Statist. Sinica*, 9:1089–1102, 1999.
- W. Stute, W. González-Manteiga, and C. Sánchez-Sellero. Nonparametric model checks in censored regression. *Comm. Statist. Theory Methods*, 29:1611–1629, 2000.
- A. W. Van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.
- A. W. Van Der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer, New York, 1996.
- I. Van Keilegom and N. Veraverbeke. Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Statist. Math.*, 49:467–491, 1997. doi: 10.1023/A:1003166728321. URL <https://doi.org/10.1023/A:1003166728321>.
- I. Van Keilegom, W. González-Manteiga, and C. Sánchez-Sellero. Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST*, 17:401–415, 2008.
- H. J. Wang and L. Wang. Locally weighted censored quantile regression. *J. Amer. Statist. Assoc.*, 104:1117–1128, 2009.
- H. J. Wang, I. W. McKeague, and M. Qian. Testing for marginal linear effects in quantile regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 80:433–452, 2018.
- L. Wang. Nonparametric test for checking lack of fit of the quantile regression model under random censoring. *Canad. J. Statist.*, 36:321–336, 2008.
- Y. Wu and G. Yin. Conditional quantile screening in ultrahigh-dimensional heterogeneous data. *Biometrika*, 102:65–76, 2015.
- S.J. Yang, A. El Ghouch, and I. Van Keilegom. Varying coefficient models having different smoothing variables with randomly censored data. *Electron. J. Stat.*, 8:226–252, 2014.
- Z. Ying, S.H. Jung, and L. J. Wei. Survival analysis with median regression models. *J. Amer. Statist. Assoc.*, 103:1214–1224, 1995.
- J. X. Zheng. A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics*, 75:263–289, 1996.
- J. X. Zheng. A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric Theory*, 14:123–138, 1998.