# SINGLE-INDEX QUANTILE REGRESSION MODELS FOR CENSORED DATA

Axel Bücher, Anouar El Ghouch, Ingrid Van Keilegom

REPRINT | 2021 / 26

# Single-Index Quantile Regression Models for Censored Data

**Axel Bücher, Anouar El Ghouch, and Ingrid Van Keilegom**

**Abstract** When the dimension of the covariate space is high, semiparametric regression models become indispensable to gain flexibility while avoiding the curse of dimensionality. These considerations become even more important for incomplete data. In this work, we consider the estimation of a semiparametric single-index model for conditional quantiles with right-censored data. Iteratively applying the local-linear smoothing approach, we simultaneously estimate the linear coefficients and the link function. We show that our estimating procedure is consistent and we study its asymptotic distribution. Numerical results are used to show the validity of our procedure and to illustrate the finite-sample performance of the proposed estimators.

## 1 Introduction

Quantile regression is a very attractive alternative to the classical mean-regression model based on the quadratic loss. While the latter provides only information about the central behavior of the data, by varying the quantile level, the former provides a more complete picture, both in the center and in the tails. At the same time, one

**Electronic supplementary material** The online version of this chapter (https://doi.org/10.1007/978-3-030-73249-3_10) contains supplementary material, which is available to authorized users.

A. Bücher
Mathematisches Institut, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany
e-mail: axel.buecher@hhu.de

A. El Ghouch
ISBA, UCLouvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium
e-mail: anouar.elghouch@uclouvain.be

I. Van Keilegom (✉)
ORSTAT, KU Leuven, Naamsestraat 69, box 3500, 3000 Leuven, Belgium
e-mail: ingrid.vankeilegom@kuleuven.be



A. Daouia and A. Ruiz-Gazen (eds.), *Advances in Contemporary Statistics and Econometrics*, https://doi.org/10.1007/978-3-030-73249-3_10

does not need to impose restrictive assumptions about the unknown data generating process. There are many cases where studying the conditional mean is uninformative compared to the conditional upper or lower quantiles representing more extreme situations. A nice illustration can be found in Elsner et al. (2008), where the interest lies in the lifetime-maximum wind speeds of tropical cyclones. The authors found that trends are near zero for the mean and lower quantiles (median and below), but are upward for higher quantiles.

With the objective of providing a robust yet easily computable alternative to linear mean models, Koenker and Bassett (1978) propose a method to estimate a linear quantile model using the so-called check loss function. This seminal work inspired many researchers from different fields and the method has been generalized and adapted to a wide range of statistical applications including fully nonparametric methods like local-polynomial or spline smoothing; see, e.g., Yu and Jones (1998) and Koenker et al. (1994). Although a completely nonparametric approach is flexible, its application requires a large amount of data in order to overcome the curse of dimensionality. While retaining much flexibility, semiparametric models avoid the curse of dimensionality by imposing some structure on the model. One such structure is the single-index model in which one assumes that the objective function depends linearly on the covariates through an unknown link function. Many widely used parametric models can be seen as particular cases of the single-index model. Examples are the linear regression model and the generalized linear model. In a single-index model, no matter the number of covariates, the curse of dimensionality is avoided because the nonparametric part (link function) is of dimension one. This model was investigated and successfully applied to many objective functions, including the conditional mean and conditional quantiles. For some related papers, see, for example, Ichimura (1993), Klein and Spady (1993), Härdle et al. (1993), Carroll et al. (1997), Delecroix et al. (2003), Wu et al. (2010), and Kong and Xia (2012) to cite just some of the relevant papers.

The majority of the available literature is devoted to the case where the variable of interest, say $Y$, is completely observed. This is not the case in many interesting applications including survival analysis where censoring prevents the direct application of "classical" semiparametric methods because instead of observing $Y$, one only observes the minimum of $Y$ and a censoring variable. For general results on (linear) quantile regression within such a setting, see, e.g., Portnoy (2003), Wang and Wang (2009), and references therein. Compared to the uncensored case, the literature on single-index models dealing with censoring is very sparse. To the best of our knowledge, the only paper so far is the one of Christou and Akritas (2019) who studied a non-iterative approach based on a combination of four local smoothing estimators: the local Kaplan–Meier estimator for estimating the conditional distribution function of the censoring variable, the nonparametric estimator of Kong et al. (2013), a Nadaraya–Watson-type estimator for estimating the link function, and a local-linear estimator for estimating the desired conditional quantile. For the case of the conditional mean, we refer to Lopez et al. (2013) and the references therein.

In this paper, we study the single-index model for the conditional quantile function when the data are right-censored. We estimate the parameters of interest by constructing a weighted check function in a way similar to the method of El Ghouch and Van Keilegom (2009). The main difficulties here are the non-differentiability of the check loss function and the fact that the weight function depends on the censoring distribution, which is unknown and needs to be estimated and then plugged-in in the estimating equation. Our proposed local-linear estimation method is based on an iterative procedure involving a $\sqrt{n}$-consistent estimator of the single-index parameters. In every iteration, we need to maximize a large number of local equations. We derive the asymptotic properties of the resulting quantile regression function under some suitable sufficient conditions. The practical performance of the proposed method is examined via Monte Carlo experiments. The estimator is shown to perform very well for data of moderate size, even when the percentage of censoring is relatively high.

The remainder of the paper is organized as follows. Section 2 describes the estimation procedure. The asymptotic properties such as the consistency and the asymptotic normality of our semiparametric estimator are obtained in Sect. 3. The problem of selecting the bandwidth parameter is tackled in Sect. 4. Simulation studies are presented in Sects. 5, and 6 highlights a brief application to real data. Proofs and technical lemmas are deferred to an online supplement.

## 2 Model and Estimation

Suppose that $Y$ is a non-negative response depending on a $d$-dimensional covariate $X$. The object of interest in this paper is the $\tau$th conditional quantile of $Y$ given $X = x$, $\tau \in (0, 1)$, which we denote by $Q_\tau(x)$. We impose a single-index structure on $Q_\tau$, i.e., we suppose that

$$Q_\tau(x) = m_\tau(x^T \beta_{0,\tau}), \tag{1}$$

where $m_\tau : \mathbb{R} \to \mathbb{R}$ is an unknown smooth link function and where $\beta_{0,\tau}$ is a vector of unknown coefficients in the unit sphere $S^{d-1} = \{\beta \in \mathbb{R}^d : \|\beta\| = 1\}$, where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^d$. For identifiability reasons, we suppose that the first coordinate of $\beta_{0,\tau}$ is positive. As long as it will not cause any ambiguity, we suppress the index $\tau$ and write $m = m_\tau$ and $\beta_0 = \beta_{0,\tau}$. In model (1), estimating $Q_\tau$ boils down to estimating $m$ and $\beta_0$.

For $u \in \mathbb{R}$, let $\rho_\tau(u) = u\{\tau - \mathbb{1}(u < 0)\}$ denote the check function. Then, it is well known that $\beta_0$ is given by

$$\begin{aligned}\beta_0 &= \operatorname{argmin}_{\beta \in \mathbb{R}^d} \mathbb{E}[\rho_\tau\{Y - m(X^T\beta)\}] \\ &= \operatorname{argmin}_{\beta \in \mathbb{R}^d} \mathbb{E}\left[\mathbb{E}[\rho_\tau\{Y - m(X^T\beta)\}|X^T\beta]\right]. \end{aligned} \tag{2}$$

The expressions $\mathbb{E}[\rho_\tau\{Y - m(X^T\beta)\}]$ and $\mathbb{E}[\rho_\tau\{Y - m(X^T\beta)\}|X^T\beta]$ can be interpreted as the expected and the conditional expected loss, respectively.

For the moment, let us suppose that there is no censoring and that we observe an i.i.d. sample $(X_i, Y_i)_{i=1}^n$ from $(X, Y)$. The following procedure for estimating $\beta_0$ and $m(v)$, where $v \in \mathbb{R}$ is arbitrary, stems from Wu et al. (2010). The main idea is to define an empirical analog of the expected loss in (2), which can be minimized subsequently. Let $\beta \in S^{d-1}$ be given. Then, assuming that $m$ is sufficiently smooth and that $X_i^T\beta$ is close to $v$, a Taylor expansion yields

$$m(X_i^T\beta) \approx m(v) + m'(v)(X_i^T\beta - v) = a + b(X_i^T\beta - v),$$

where $a = m(v)$ and $b = m'(v)$. Thus,

$$\sum_{i=1}^{n} \rho_\tau \left\{Y_i - a - b(X_i^T\beta - v)\right\} K\{(X_i^T\beta - v)/h\} \tag{3}$$

with some kernel function $K$ and a bandwidth $h$ represents an empirical analog of the conditional expected loss in (2). Note that, for given $\beta = \beta_0$, minimizing (3) with respect to $a$ and $b$ yields oracle estimators for $m(v)$ and $m'(v)$, respectively. To get an empirical analog of $\mathbb{E}[\rho_\tau\{Y - m(X^T\beta)\}]$, we need to average (3) over $v$. Hence, setting $v = v_j = X_j^T\beta$, we obtain

$$\sum_{j=1}^{n}\sum_{i=1}^{n} \rho_\tau \left\{Y_i - a_j - b_j(X_{ij}^T\beta)\right\} w_{ij}(\beta), \tag{4}$$

where $X_{ij} = X_i - X_j$ and where

$$w_{ij}(\beta) = \left\{\sum_{i=1}^{n} K\left(\frac{X_{ij}^T\beta}{h}\right)\right\}^{-1} K\left(\frac{X_{ij}^T\beta}{h}\right).$$

By minimizing the expression in (4) with respect to $(a_j, b_j)_{j=1}^n$ and $\beta$, we obtain estimators of $(m(v_j), m'(v_j))_{j=1}^n$ and $\beta_0$. To simplify this minimization problem, Wu et al. (2010) proposed an iterative procedure based on successive estimation of $\beta_0$ and $(m(v), m'(v))$, for any given $v \in \mathbb{R}$. In the present paper, we adapt their approach to the case where the observations of the response variable may be censored.

In the presence of censoring, we do not fully observe the response variables $Y_i$. Instead, we observe a sequence of i.i.d. triplets $(X_i, Z_i, \Delta_i)_{i=1}^n$ from $(X, Z, \Delta)$, where $Z = \min(Y, C)$, $\Delta = \mathbb{1}(Y \le C)$, and $C \ge 0$ denotes a censoring variable.

Assume for the moment that $C$ is independent of $Y$ given $X^T\beta$ and let $F_{C|X^T\beta}(z|x^T\beta) = \Pr(C \le z|X^T\beta = x^T\beta)$ denote the conditional distribution of $C$ given $X^T\beta = x^T\beta$. Then, some simple calculations based on the tower property of conditional expectations show that, for any measurable function $h : \mathbb{R}^2 \to \mathbb{R}$,

$$\mathbb{E}[h(Y, X^T\beta) \mid X^T\beta] = \mathbb{E}\left[\frac{h(Z, X^T\beta)\Delta}{1 - F_{C|X^T\beta}(Z - |X^T\beta)} \,\middle|\, X^T\beta\right]. \tag{5}$$

Therefore, we can write $\mathbb{E}\left[\rho_\tau\{Y - a - b(X^T\beta - v)\} \mid X^T\beta\right]$ as

$$
\begin{aligned}
&\mathbb{E}\left[Q(\beta)\rho_\tau\{Z-a-b(X^T\beta-v)\}\,|\,X^T\beta\right]\\
&=\tau\mathbb{E}\left[Y-Z|X^T\beta\right]+\mathbb{E}\Big[\{Z-a-b(X^T\beta-v)\}\big[\tau\\
&\qquad\qquad -Q(\beta)\mathbb{1}\{Z<a+b(X^T\beta-v)\}\big]\Big|X^T\beta\Big],
\end{aligned}
$$

where $Q(\beta)=\Delta/\{1-F_{C|X^T\beta}(Z-|X^T\beta)\}$. This suggests to replace (3) by either

$$
\sum_{i=1}^n \hat{Q}_i(\beta)\rho_\tau\{Z_i-a-b(X_i^T\beta-v)\}K\left(\frac{X_i^T\beta-v}{h}\right), \tag{6}
$$

or

$$
\sum_{i=1}^n\{Z_i-a-b(X_i^T\beta-v)\}\left[\tau-\hat{Q}_i(\beta)\mathbb{1}\{Z_i<a+b(X_i^T\beta-v)\}\right]K\left(\frac{X_i^T\beta-v}{h}\right), \tag{7}
$$

with $\hat{Q}_i(\beta)=\Delta_i/\{1-\hat{F}_{C|X^T\beta}(Z_i-|X_i^T\beta)\}$, where $\hat{F}_{C|X^T\beta}$ is a suitable estimator of $F_{C|X^T\beta}$. For instance, one may use the local Kaplan–Meier estimator given by

$$
\hat{F}_{C|X^T\beta}(z|x^T\beta)=1-\prod_{Z_i\le z}\left(1-\frac{B_i(\beta,x)}{\sum_{Z_j\ge Z_i}B_i(\beta,x)}\right)^{1-\Delta_i},
$$

with

$$
B_i(\beta,x)=\frac{K\left(\frac{\beta^TX_i-\beta^Tx}{a_n}\right)}{\sum_{j=1}^n K\left(\frac{\beta^TX_j-\beta^Tx}{a_n}\right)},
$$

and where $a_n$ is a bandwidth sequence converging to zero as $n$ tends to infinity. When $B_i=n^{-1}$ for all $i$, $\hat{F}_{C|X^T\beta}$ reduces to the classical (unconditional) Kaplan–Meier estimator, subsequently simply denoted by $\hat{F}_C$. Note that, for any given $\beta$, both (6) and (7) are convex functions. Although the numerical minimization of (6) may be easier than that of (7), in this work we opt for the latter because, as is well known, the Kaplan–Meier estimator is very unstable at the right tail and this problem can be adequately and automatically dealt with through (7). In fact, in (6), the Kaplan–Meier estimator needs to be calculated for every $Z_i$ whereas in (7), using the fact that $\hat{Q}_i(\beta)\mathbb{1}\{Z_i<a+b(X_i^T\beta-v)\}=0$ if $Z_i\ge a+b(X_i^T\beta-v)$, the observations beyond $m(x^T\beta)$ would have no or a very small impact (depending on the bandwidth) on the resulting estimator. A very similar approach was used in El Ghouch and Van Keilegom (2009) for the case of one covariate. An approach based on minimizing a quantity closely related to (7) can be found in He et al. (2013) for analyzing high-dimensional survival data.

For simplicity, and to avoid some technical difficulties, in the present paper, we assume that

(C1) $C$ is independent of $Y$ given $X$ and $C$ are independent of $X$

(a different assumption, also used for instance by Bouaziz and Lopez (2010) recently, under which the asymptotic results in this paper remain valid is given in Remark 1 below). In such a case, $Y$ and $C$ are independent given $X^T\beta$, and $F_{C|X^T\beta}(z|x^T\beta) = \Pr(C \leq z) = F_C(z)$ so that the unconditional Kaplan–Meier estimator can be used. To sum up, we estimate $m(v)$ and $m'(v)$ by $\hat{m}(v,\beta) = \hat{a}(v,\beta)$ and $\hat{m}'(v,\beta) = \hat{b}(v,\beta)$, respectively, where

$$(\hat{a}(v,\beta), \hat{b}(v,\beta)) = \operatorname{argmin}_{a,b\in\mathbb{R}} \sum_{i=1}^{n} \{Z_i - a - b(X_i^T\beta - v)\}\left[\tau - \hat{Q}_i \mathbb{1}\{Z_i < a + b(X_i^T\beta - v)\}\right] K\left(\frac{X_i^T\beta - v}{h}\right), \quad (8)$$

and where $\hat{Q}_i = \Delta_i/\{1 - \hat{F}_C(Z_i-)\}$ with the unconditional Kaplan–Meier estimator $\hat{F}_C$. Still, it remains to construct an estimator for $\beta_0$. To do so, we proceed as in the uncensored case and define the following empirical analog of (4):

$$\sum_{j=1}^{n}\sum_{i=1}^{n} \{Z_i - a_j - b_j(X_{ij}^T\beta)\}\left[\tau - \hat{Q}_i \mathbb{1}\{Z_i < a_j + b_j(X_{ij}^T\beta)\}\right] w_{ij}(\beta).$$

The joint minimization of the resulting expression with respect to $(a_j, b_j)_{j=1}^n$ and $\beta$ is complicated and likely to lead to unstable estimates, hence we propose the following iterative procedure adapted from Wu et al. (2010).

Step 1. Start with an initial estimator $\hat{\beta}^{(0)}$ of $\beta_0$ and set $\beta_{iter} = \hat{\beta}^{(0)}$ (see below for a suitable example on how to obtain $\hat{\beta}^{(0)}$).

Step 2. For $j = 1, \ldots, n$, let

$$(\hat{a}_j, \hat{b}_j) = \operatorname{argmin}_{a,b\in\mathbb{R}} \sum_{i=1}^{n} \{Z_i - a - b(X_{ij}^T\beta_{iter})\}\big[\tau - \hat{Q}_i \mathbb{1}\{Z_i < a + b(X_{ij}^T\beta_{iter})\}\big] w_{ij}(\beta_{iter}).$$

Step 3. Using the estimates $(\hat{a}_j, \hat{b}_j)_{j=1}^n$, set

$$\beta^\star = \operatorname{argmin}_{\beta\in\mathbb{R}^d} \sum_{j=1}^{n}\sum_{i=1}^{n} \{Z_i - \hat{a}_j - \hat{b}_j(X_{ij}^T\beta)\}\big[\tau - \hat{Q}_i \mathbb{1}\{Z_i < \hat{a}_j + \hat{b}_j(X_{ij}^T\beta)\}\big] w_{ij}(\beta_{iter})$$

and update $\beta_{iter}$ by setting $\beta_{iter} = \operatorname{sgn}(\beta_1^\star)\beta^\star/\|\beta^\star\|$.

Step 4. Repeat Steps 2 and 3 until the difference between two consecutive estimations of $\beta$ is smaller than a given threshold and define the final estimate $\hat{\beta}$ by setting $\hat{\beta} = \beta_{iter}$.

Step 5. For any desired index value $v \in \mathbb{R}$, estimate $m(v)$ and $m'(v)$ by $\hat{m}(v, \hat{\beta}) = \hat{a}(\hat{\beta})$ and $\hat{m}'(v, \hat{\beta}) = \hat{b}(\hat{\beta})$, the latter estimators being defined in (8). For any desired index value $x \in \mathbb{R}^d$, estimate $Q_\tau(x)$ by $\hat{m}(x^T\hat{\beta}, \hat{\beta})$.

Step 1 requires an initial estimator for $\beta_0$. We propose to use an estimator adapted from the OPG (outer product of gradients) method in the mean-regression context in Xia et al. (2002). The method requires that $X$ has a density, and the underlying idea is as follows: For any $x \in \mathbb{R}^d$, we have $\partial m(x^T\beta_0)/\partial x = m'(x^T\beta_0)\beta_0$. Hence, the partial derivatives of $m(x^T\beta_0)$ with respect to $x$ are parallel to $\beta_0$. For $j = 1, \ldots, n$, let $b_j = m'(X_j^T\beta_0)\beta_0$. One can easily see that the (standardized) eigenvector corresponding to the largest eigenvalue of $V_n = n^{-1}\sum_{i=1}^n b_j b_j^T$ is given by $\beta_0$, which suggests to estimate $\beta_0$ by replacing $b_j$ in the definition of $V_n$ by suitable estimators $\hat{b}_j$, that is, we define $\hat{\beta}_0$ as the (standardized) eigenvector corresponding to the largest eigenvalue of $\hat{V}_n = n^{-1}\sum_{j=1}^n \hat{b}_j\hat{b}_j^T$. For the estimation of $b_j$, we propose to use the local-polynomial estimators

$$(\hat{a}_j, \hat{b}_j^T) = \text{argmin}_{(a,b^T)\in\mathbb{R}^{d+1}} \sum_{i=1}^n \{Z_i - a - b^T X_{ij}\} \left[\tau - \hat{Q}_i \mathbb{1}\{Z_i < a + b^T X_{ij})\}\right] K(X_{ij}/h),$$

where $K$ denotes a $d$-dimensional kernel.

## 3 Asymptotic Results

In this section, we present asymptotic results for the final estimator $\hat{m} = \hat{m}(\hat{\beta})$ arising from Step 5 of the procedure described in the preceding section. In particular, we show that the estimator for $m$ does not depend on the specific form (or asymptotic distribution) of the parametric estimator $\hat{\beta}$, as long as it is $\sqrt{n}$-consistent for $\beta_0$ . In a non-censored case, the latter assumption has for instance been shown for a similar recursively defined estimator in Kong and Xia (2012). In a censored case, it is satisfied for the maximum likelihood estimator proposed by Strzalkowska-Kominiak and Cao (2013) and for the regression-like semiparametric estimator of Bouaziz and Lopez (2010).

We begin by describing technical conditions. For fixed $v \in \mathbb{R}$, suppose that there exist neighborhoods $U_{\beta_0}$, $U_{m(v)}$, and $U_v$ of $\beta_0$, $m(v)$ and $v$, respectively, such that:

(A1) The kernel $K$ is a density function on $\mathbb{R}$ which is symmetric around 0, has a compact support denoted by $\mathrm{supp}(K)$, and is differentiable with a bounded derivative.
(A2) The function $m$ is twice continuously differentiable on $U_v$ with bounded derivatives.
(A3) (i) The support of $X$, denoted by $\mathrm{supp}(X)$, is contained in a compact subset $D_X$ of $\mathbb{R}^d$.
(ii) For any $\beta \in U_{\beta_0}$, the random variable $X^T\beta$ has a density $f_{X^T\beta}$. The function $U_{\beta_0} \times U_v \to \mathbb{R}$, $(\beta, u) \mapsto f_{X^T\beta}(u)$ is bounded and Lipschitz-continuous at $(\beta_0, v)$. In addition, $f_{X^T\beta_0}(v) > 0$.
(A4) (i) The conditional distribution $F_{Y|X}$ of $Y$ given $X$ has a conditional density $f_{Y|X}(\cdot|\cdot)$ that is bounded on $U_{m(v)} \times \mathrm{supp}(X)$.
(ii) For any $\beta \in U_{\beta_0}$, the conditional distribution of $Y$ given $X^T\beta$ has a conditional density $f_{Y|X^T\beta}(\cdot|\cdot)$. The function $U_{\beta_0} \times U_{m(v)} \times U_v \to \mathbb{R}$, $(\beta, y, u) \mapsto f_{Y|X^T\beta}(y \mid u)$ is bounded and Lipschitz-continuous at $(\beta_0, m(v), v)$. In addition, $f_{Y|X^T\beta_0}(m(v) \mid v) > 0$.
(iii) $U_{\beta_0} \times U_{m(v)} \times U_v \to \mathbb{R}$, $(\beta, y, u) \mapsto f_{Y|X^T\beta}(y|u)$ is partially differentiable with respect to $y$ and the derivative, denoted by $f'_{Y|X^T\beta}(y|u)$, is bounded.
(A5) The point $v \in \mathbb{R}$ satisfies $F_Z\{m(v)\} < 1$, where $F_Z$ denotes the c.d.f. of $Z$.

Before we formulate the main results, let us introduce some additional notations. For $\beta \in \mathbb{R}^d$ and $u \in \mathbb{R}$, let $\mathcal{X}_i(\beta, u) = \big(1, (X_i^T\beta - u)/h\big)^T$, $\mathcal{Z}_i(\beta, u) = Z_i - m(u) - m'(u)(X_i^T\beta - u)$, and $\mathcal{K}_i(\beta, u) = K\{(X_i^T\beta - u)/h\}$. Moreover, set $\bar{K}_j = \int_{\mathbb{R}} u^j K(u)\,du$ and $\bar{K}'_j = \int_{\mathbb{R}} u^j K^2(u)\,du$ for $j \in \{0, 1, 2, 3\}$ and let

$$\bar{K} = \begin{pmatrix} \bar{K}_0 & \bar{K}_1 \\ \bar{K}_1 & \bar{K}_2 \end{pmatrix}, \qquad \bar{K}' = \begin{pmatrix} \bar{K}'_0 & \bar{K}'_1 \\ \bar{K}'_1 & \bar{K}'_2 \end{pmatrix}.$$

For some constant $M > 0$, let $U_M$ denote the closed $d$-dimensional ball of radius $M$ with center 0, i.e., $U_M = \{\gamma \in \mathbb{R}^d : \|\gamma\| \le M\}$. Finally, for $\beta \in \mathbb{R}^d$ and $u \in \mathbb{R}$ (usually considered to be close to $\beta_0$ and $v$), let

$$\mathbb{M}_n(u, \beta) = \sqrt{nh}\left\{\begin{pmatrix} \hat{m}(u, \beta) - m(v) \\ h\{\hat{m}'(u, \beta) - m'(v)\} \end{pmatrix} - \frac{h^2}{2}\bar{K}^{-1}\begin{pmatrix} \bar{K}_2 \\ \bar{K}_3 \end{pmatrix} m''(v)\right\}$$

with $\hat{m}(u, \beta)$ and $\hat{m}'(u, \beta)$ as defined in (8).

**Theorem 1** *Suppose that (C1) is met and that* $h = h(n) \to 0$ *satisfies* $\lim_{n\to\infty} nh^3 = \infty$ *and* $nh^5 = O(1)$ *as* $n \to \infty$*. Then, for any* $v \in \mathbb{R}$ *that satisfies conditions (A1)–(A5) and for any* $M > 0$,

$$\sup_{(\gamma,\kappa)\in U_M\times[-M,M]}\Bigg\|\mathbb{M}_n(v_n^\kappa, \beta_n^\gamma) - V^{-1}\frac{1}{\sqrt{nh}}\sum_{i=1}^n\Big[\tau - Q_i\mathbb{1}\left\{Z_i < m(X_i^T\beta_0)\right\}\Big] \times \mathcal{X}_i(\beta_0, v)\mathcal{K}_i(\beta_0, v)\Bigg\| = o_P(1),$$

*where* $v_n^\kappa = v + \kappa/\sqrt{n}$ *and* $\beta_n^\gamma = \beta_0 + \gamma/\sqrt{n}$, *where* $Q_i = \Delta_i/\{1 - F_C(Z_i-)\}$ *and where* $V = \big[ f_{Y|X^T\beta_0}\{m(v) \mid v\}\ f_{X^T\beta_0}(v)\big]\bar{K}$.

Note that the sum between the norm signs in Theorem 1 consists of centered summands as a consequence of (5). The uniformity in $\gamma$ and $\kappa$ in Theorem 1 is essential for the next corollary which can be regarded as the main result of this paper: it states that the final estimator for $Q_\tau(x)$ in Step 5 is asymptotically normally distributed.

**Corollary 1** *Let* $\hat{\beta}_n \in S^{d-1}$ *be an estimator for* $\beta_0$ *such that* $\hat{\gamma}_n = \sqrt{n}(\hat{\beta}_n - \beta_0) = O_P(1)$. *Suppose that (C1) and the conditions on the bandwidth of Theorem 1 are met. Then, for any* $v \in \mathbb{R}$ *that satisfies conditions (A1)–(A5) and for any* $x \in \mathbb{R}^d$ *such that* $v = x^T\beta_0$ *satisfies conditions (A1)–(A5),*

$$\mathbb{M}_n(v, \hat{\beta}_n) \rightsquigarrow \mathcal{N}_2\big(0, \sigma^2(v)\bar{K}^{-1}\bar{K}'\bar{K}^{-1}\big), \text{ and}$$
$$\mathbb{M}_n(x^T\hat{\beta}_n, \hat{\beta}_n) \rightsquigarrow \mathcal{N}_2\big(0, \sigma^2(x^T\beta_0)\bar{K}^{-1}\bar{K}'\bar{K}^{-1}\big),$$

*where, for any* $v \in \mathbb{R}$,

$$\sigma^2(v) = \frac{\Phi_{\beta_0}\{m(v) \mid v\} - \tau^2}{f^2_{Y|X^T\beta_0}\{m(v) \mid v\}\ f_{X^T\beta_0}(v)}$$

*and where, for any* $u, v \in \mathbb{R}$,

$$\Phi_{\beta_0}(u \mid v) = \mathbb{E}\left[\frac{\mathbb{1}(Y < u)}{1 - F_C(Y-)}\,\middle|\, X^T\beta_0 = v\right].$$

**Remark 1** The results of Theorem 1 and Corollary 1 remain valid provided we replace Condition (C1) by the following Condition (C2) originating from Stute (1993). Note that it is also imposed in Bouaziz and Lopez (2010).

(C2) $\Delta$ is independent of $X$ given $Y$ and $C$ are independent of $Y$.

We also refer to Lopez et al. (2013), where assumption (C1) is replaced by a weaker assumption involving independence between $C$ and $Y$ conditional on $g(X)$ for some function $g$. For the sake of brevity, we omit further details.

## 4 Bandwidth Selection

The practical performance of any nonparametric regression technique depends crucially on the choice of smoothing parameters. A (theoretical) local optimal bandwidth can be derived from the result in Corollary 1 by minimizing the asymptotic mean squared error of $\hat{m}(v, \hat{\beta})$ with respect to $h$, yielding

$$h_n^{opt} = h_n^{opt}(v) = \left\{ \frac{\sigma^2(v)\bar{K}_0}{\{m''(v)\}^2 \bar{K}_2^2} \right\}^{1/5} n^{-1/5}.$$

Unfortunately, this expression is not directly applicable in practice, since it depends on several unknown quantities. Even in the simpler non-censored case, the derivation of reliable estimators for the respective quantities is delicate. For that reason, alternative procedures for the bandwidth selection have been proposed, see, e.g., Yu and Jones (1998) or Kong and Xia (2012) for procedures relying on the mean-regression case. However, these procedures are not directly applicable in the presence of censoring. For that reason, we propose to use the following leave-one out cross-validation (CV) procedure (see also Zheng and Yang 1998; Leung 2005; El Ghouch and Van Keilegom 2009):

(CV1) For a given $h$, estimate $\hat{\beta} = \hat{\beta}(h)$ as in Steps 1–4.

(CV2) For any $j = 1, \ldots, n$, set $\hat{m}_{-j,h}(X_j^T \hat{\beta}) = \hat{a}_{-j}(X_j^T \hat{\beta}, \hat{\beta})$, where, for any $v \in \mathbb{R}$ and $\beta \in S^{d-1}$,

$$\big(\hat{a}_{-j}(v,\beta), \hat{b}_{-j}(v,\beta)\big) = \operatorname{argmin}_{a,b\in\mathbb{R}} \sum_{\substack{i=1,\ldots,n \\ i\neq j}} \{Z_i - a - b(X_i^T\beta - v)\}$$
$$\times \hat{Q}_{i,-j}\left[\tau - \mathbb{1}\{Z_i < a + b(X_i^T\beta - v)\}\right] K\left(\frac{X_i^T\beta - v}{h}\right)$$

denotes the estimator based on all observations except the $j$th.

(CV3) For $j \in \{1, \ldots, n\}$ such that $\Delta_j = 1$, set $\widehat{cv}_{-j,h} = |\hat{m}_{-j,h}(X_j^T \hat{\beta}) - Z_j|$. Let $CV(h)$ denote either the median or the mean or the $m$%-trimmed mean of that sample (referred to as MAE, MSE, or trimmed MSE in the following).

(CV4) Repeat the first three steps for several bandwidths and set $h_n^{CV} = \operatorname{argmin}_h CV(h)$.

We consider 10%-trimmed MSE, which, together with the MSE and the MAE, yields three different criteria.

## 5 Numerical Results

In this section, we assess the finite-sample performance of the 5-step estimator for $m(v)$. For reasons of numerical stability, we constrain all minimizations to a compact set $[-M, M]^p$, with $M = 10$. Additionally, we stop the algorithm in Step 4 after atmost 25 iterations, if convergence has not occurred until then. We perform 500 repetitions for two different models, two sample sizes ($n = 200, 400$), two levels of censoring (on average 25% and 50%), three values of $\tau \in \{0.3, 0.5, 0.7\}$, two dimensions $d \in \{3, 6\}$ and 61 values for $v \in \{0.05, 0.075, 0.1, \ldots, 1.525, 1.55\}$. We consider 15 different bandwidths $h \in \{0.1, 0.15, \ldots, 0.75, 0.8\}$. Additionally, we

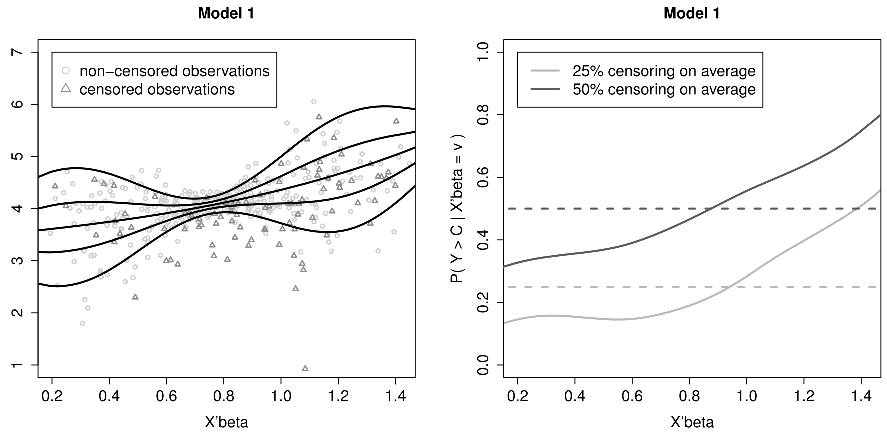


**Fig. 1** Left: True quantile curves for $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$ (black curves, in increasing order) and a simulated sample of size $n = 400$ (for $d = 3$, with 25% censoring on average). Right: Probability of censoring $v \mapsto \Pr(Y > C \mid X^T\beta_0 = v)$ for Model 1. The average probability of censoring $\Pr(Y > C)$ is 25% for the black curve and 50% for the gray curve

investigate the performance of the cross-validation method described in Sect. 4. The considered models are as follows.

**Model 1 (location-scale model, censoring independent of the covariate)**

For $i = 1, \ldots, n$, we consider

$$Y_i = 3 + \tfrac{1}{2}\exp(X_i^T\beta_0) + \{1 + \tfrac{3}{4}\sin(2\pi X_i^T\beta_0)\}\,\varepsilon_i, \qquad X_i = (X_{i,1}, \ldots, X_{i,d}),$$

where $X_{i,j}$ is i.i.d. uniform on $(0, 1)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, d$, and where $\varepsilon_i$ is i.i.d. normal with mean 0 and variance 0.25. During the simulation study, we consider the vector $\beta_0 = \|(d, d-1, \ldots, 1)\|_2^{-1} \times (d, d-1, \ldots, 1)$. Note that the support of $X^T\beta_0$ is the interval $[0, \|\beta_0\|_1]$, with $\|\beta_0\|_1 = 1.60$ for $d = 3$ and $\|\beta_0\|_1 = 2.20$ for $d = 6$. The $\tau$th conditional quantile of $Y_i$ given $X_i = x$ is given by

$$Q_\tau(x) = q_\tau\left(\tfrac{1}{2}\exp(x^T\beta_0), \tfrac{1}{2}\{1 + \tfrac{3}{4}\sin(2\pi x^T\beta_0)\}\right), \tag{9}$$

where $q_\tau(\mu, \sigma)$ denotes the $\tau$th-quantile of the normal distribution with mean $\mu$ and standard deviation $\sigma$. The curves are depicted in the left panel of Fig. 1, for $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

The censoring variables are i.i.d. normal with mean $\mu_C$ and variance $\sigma_C^2 = 1$, independent of $X_i$ and $\varepsilon_i$. We consider two choices for the mean $\mu_C$, which result in either a proportion of censoring of about 50% or of about 25% (for instance, for $d = 3$ the choices are $\mu_C = 4.2$ to obtain a proportion of censoring of about 50%, and $\mu_C = 5$ for proportion of censoring of about 25%). A sample of size $n = 400$ with $d = 3$ and 25% censoring is depicted in the left panel of Fig. 1.

Note that the probability of censoring given $X = x$ is given by

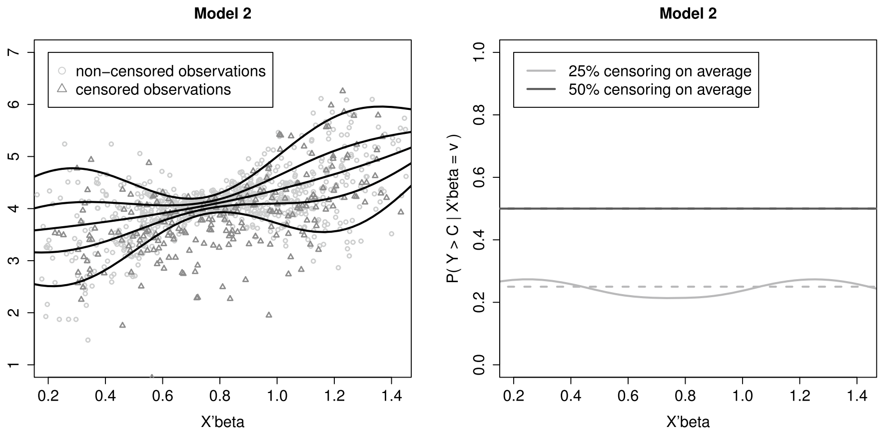


**Fig. 2** Left: True quantile curves for $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$ (black curves, in increasing order) and a simulated sample of size $n = 400$ (for $d = 3$, with 25% censoring on average). Right: Probability of censoring $v \mapsto \Pr(Y > C \mid X^T\beta_0 = v)$ for Model 2. The average probability of censoring $\Pr(Y > C)$ is 25% for the black curve and 50% for the gray curve

$$\Pr(Y > C \mid X = x) = \Phi\left(\frac{3 + \frac{1}{2}\exp(x^T\beta_0) - \mu_C}{\sqrt{1 + \frac{1}{4}\{1 + \frac{3}{4}\sin(2\pi x^T\beta_0)\}^2}}\right),$$

where $\Phi$ is the standard normal cumulative distribution function. The corresponding curves $v \mapsto \Pr(Y > C \mid X^T\beta_0 = v)$ are depicted in the right panel of Fig. 1 for $\mu_C \in \{4.2, 5\}$ (which, for $d = 3$, yields a proportion of censoring of about 50% and 25%, respectively). From these graphs, we expect the estimator $\hat{m}(v, \hat{\beta})$ to have worse performance for large values of $v$.

**Model 2 (location-scale model, censoring depending on the covariate)**

We consider the same data generating mechanism for $Y_i$ as for Model 1. In particular, the conditional quantile curves are given by (9).

The censoring variables are i.i.d. normal with mean $\mu_C + \frac{1}{2}\exp(X^T\beta_0)$ and variance $\sigma_C^2 = 1$, independent of $\varepsilon_i$. We consider two choices for the mean $\mu_C$, which result in either a proportion of censoring of about 50% or of about 25% (for instance, for $d = 3$ the choices are $\mu_C = 3$ to obtain a proportion of censoring of about 50%, and $\mu_C = 3.8$ for proportion of censoring of about 25%). A sample of size $n = 400$ with $d = 3$ and 25% censoring is depicted in the left panel of Fig. 2.

The probability of censoring given $X = x$ is given by

$$\Pr(Y > C \mid X = x) = \Phi\left(\frac{3 - \mu_C}{\sqrt{1 + \frac{1}{4}\{1 + \frac{3}{4}\sin(2\pi x^T\beta_0)\}^2}}\right).$$

The corresponding curves $v \mapsto \Pr(Y > C \mid X^T\beta_0 = v)$ are depicted in the right panel of Fig. 2 for $\mu_C \in \{3, 3.8\}$ (which, for $d = 3$, yields a proportion of censoring of about 50% and 25%, respectively). The curves are much flatter than in Model 1, whence we may expect the estimator to perform similarly throughout the support of $X^T\beta_0$.

The results of our simulation study for the fixed bandwidth case are reported in Table 1 and Figs. 3 and 4. The results in Table 1 concern both the performance of the estimator of $\beta$ and the estimator of $m(v)$ for various values of $v$. We state the minimal MSE (for $\hat{\beta}$: the minimal summed MSE over the coordinates of $\beta$), over all 15 bandwidth choices $h \in \{0.1, 0.15, \ldots, 0.8\}$, alongside with the value realizing that minimum. The results in Figs. 3 and 4 illustrate the performance of the estimator $\hat{m}(v)$ in dependence of the bandwidth parameter $h$, for a fixed value of $v = 0.85$. The reported boxplots concern the empirical squared estimation error over $N = 500$ simulation runs, and are only reported for $d = 3$ (the results for $d = 6$ look very similar and are not presented here for the sake of brevity).

Overall, the results are as to be expected: for both models, they (greatly) improve with larger sample sizes and a smaller proportion of censoring. Concerning the quantile level, the results are in most cases best for $\tau = 0.5$, closely followed by $\tau = 0.3$ and then $\tau = 0.7$. Despite the fact that the estimator for Model 2 (lower half of Table 1 and Fig. 4) is more complicated (being based on the local Kaplan–Meier estimator for the censoring distribution), the performance of the estimator is often better than for the Model 1, in particular for the parametric estimator $\hat{\beta}$.

Finally, Table 2 shows simulation results on the cross-validation method based on the 10%-trimmed MSE for choosing the optimal bandwidth as described in Sect. 4. For the sake of brevity, we only consider Model 1 with $d = 6$. We measure the quality of the cross-validation method in terms of the relative efficiency:

$$RE = \frac{MSE(\hat{t}, h^{gl.opt})}{MSE(\hat{t}, h_n^{CV})},$$

where $h^{gl.opt} = \min_{h\in\{0.1,\ldots,0.8\}} \big\{MSE(\hat{\beta}, h) + MSE(\hat{m}(0.7), h) + MSE(\hat{m}(1), h) + MSE(\hat{m}(1.3), h)\big\}$ and where $\hat{t} \in \{\hat{\beta}, \hat{m}(0.7), \hat{m}(1), \hat{m}(1.3)\}$.

The results in Table 2 show that, overall, the cross-validation method works reasonably well but we also noticed that in some cases, the method may lead to unsatisfactory results. Therefore more work is needed to develop a better solution for this challenging problem of bandwidth selection.

## 6 Case Study

In this section, we fit the single-index quantile regression model to a subset of the data from the University of Massachusetts AIDS Research Unit IMPACT Study (called `UIS`-dataset), available online at the John Wiley & Sons website, ftp://ftp.wiley.com/public/sci_tech_med/survival. This dataset has been extensively studied

**Table 1** Minimal summed MSE of $\beta$ and minimal MSE of $\hat{m}$ for four values of $\upsilon$ in Model 1 (upper half) and Model 2 (lower half), multiplied by $10^3$, over all bandwidths $h \in \{0.1, 0.15, \ldots, 0.75, 0.8\}$, alongside with the bandwidth realizing that minimum. The first and third quarter are for $d = 3$, while the the second and fourth quarter are for dimension $d = 6$

| $n$ | Cens. | $\tau$ | $\hat{\beta}$ | $h_{opt}$ | $\hat{m}(0.4)$ | $h_{opt}$ | $\hat{m}(0.7)$ | $h_{opt}$ | $\hat{m}(1)$ | $h_{opt}$ | $\hat{m}(1.3)$ | $h_{opt}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 200 | 0.25 | 0.3 | 20.3 | 0.50 | 24.4 | 0.45 | 3.0 | 0.25 | 8.5 | 0.50 | 34.5 | 0.80 |
| 200 | 0.50 | 0.3 | 38.1 | 0.55 | 32.7 | 0.55 | 5.4 | 0.35 | 12.3 | 0.70 | 60.4 | 0.80 |
| 200 | 0.25 | 0.5 | 17.3 | 0.75 | 10.0 | 0.80 | 2.5 | 0.55 | 6.2 | 0.80 | 39.2 | 0.80 |
| 200 | 0.50 | 0.5 | 37.5 | 0.75 | 16.0 | 0.80 | 4.3 | 0.70 | 9.5 | 0.80 | 77.8 | 0.80 |
| 200 | 0.25 | 0.7 | 23.3 | 0.55 | 22.9 | 0.50 | 7.2 | 0.30 | 12.1 | 0.80 | 75.3 | 0.75 |
| 200 | 0.50 | 0.7 | 68.0 | 0.55 | 36.3 | 0.60 | 20.0 | 0.45 | 24.9 | 0.80 | 149.1 | 0.80 |
| 400 | 0.25 | 0.3 | 8.3 | 0.45 | 13.0 | 0.45 | 1.8 | 0.30 | 3.6 | 0.45 | 15.7 | 0.80 |
| 400 | 0.50 | 0.3 | 13.0 | 0.50 | 17.8 | 0.50 | 2.1 | 0.30 | 5.1 | 0.45 | 28.5 | 0.80 |
| 400 | 0.25 | 0.5 | 7.8 | 0.75 | 5.5 | 0.80 | 0.9 | 0.50 | 2.7 | 0.80 | 21.3 | 0.80 |
| 400 | 0.50 | 0.5 | 13.9 | 0.75 | 8.8 | 0.80 | 1.6 | 0.50 | 4.7 | 0.80 | 44.1 | 0.80 |
| 400 | 0.25 | 0.7 | 9.6 | 0.55 | 14.0 | 0.45 | 2.5 | 0.20 | 6.0 | 0.55 | 35.2 | 0.70 |
| 400 | 0.50 | 0.7 | 23.6 | 0.55 | 23.3 | 0.50 | 7.1 | 0.30 | 12.9 | 0.75 | 83.5 | 0.80 |
| 200 | 0.25 | 0.3 | 109.6 | 0.80 | 137.1 | 0.50 | 5.9 | 0.55 | 26.2 | 0.30 | 126.0 | 0.80 |
| 200 | 0.50 | 0.3 | 189.5 | 0.80 | 174.3 | 0.80 | 11.4 | 0.70 | 59.1 | 0.80 | 170.9 | 0.10 |
| 200 | 0.25 | 0.5 | 60.8 | 0.80 | 18.7 | 0.80 | 7.9 | 0.80 | 22.1 | 0.75 | 40.4 | 0.80 |
| 200 | 0.50 | 0.5 | 132.4 | 0.80 | 30.0 | 0.80 | 21.7 | 0.80 | 51.7 | 0.80 | 108.4 | 0.80 |
| 200 | 0.25 | 0.7 | 67.4 | 0.80 | 73.1 | 0.55 | 21.2 | 0.30 | 27.9 | 0.80 | 33.0 | 0.80 |
| 200 | 0.50 | 0.7 | 163.5 | 0.75 | 66.2 | 0.55 | 63.0 | 0.40 | 87.2 | 0.80 | 98.3 | 0.45 |
| 400 | 0.25 | 0.3 | 45.4 | 0.30 | 75.6 | 0.50 | 3.0 | 0.60 | 4.7 | 0.25 | 47.4 | 0.25 |
| 400 | 0.50 | 0.3 | 94.3 | 0.80 | 111.7 | 0.45 | 3.9 | 0.60 | 17.0 | 0.40 | 106.0 | 0.80 |
| 400 | 0.25 | 0.5 | 28.2 | 0.80 | 11.6 | 0.80 | 3.7 | 0.70 | 9.0 | 0.45 | 18.3 | 0.80 |
| 400 | 0.50 | 0.5 | 61.4 | 0.80 | 14.4 | 0.80 | 7.0 | 0.80 | 18.7 | 0.35 | 39.1 | 0.80 |
| 400 | 0.25 | 0.7 | 32.0 | 0.80 | 47.2 | 0.45 | 7.7 | 0.20 | 10.7 | 0.80 | 16.8 | 0.80 |
| 400 | 0.50 | 0.7 | 82.9 | 0.80 | 48.9 | 0.50 | 18.7 | 0.25 | 31.0 | 0.55 | 41.8 | 0.80 |
| 200 | 0.25 | 0.3 | 17.1 | 0.45 | 20.9 | 0.50 | 3.0 | 0.30 | 7.6 | 0.45 | 28.6 | 0.80 |
| 200 | 0.50 | 0.3 | 26.2 | 0.50 | 26.7 | 0.60 | 4.0 | 0.35 | 10.5 | 0.50 | 43.9 | 0.80 |
| 200 | 0.25 | 0.5 | 13.7 | 0.80 | 12.4 | 0.80 | 2.1 | 0.60 | 5.6 | 0.80 | 24.9 | 0.80 |
| 200 | 0.50 | 0.5 | 23.7 | 0.75 | 21.0 | 0.80 | 3.5 | 0.65 | 8.4 | 0.80 | 46.4 | 0.80 |
| 200 | 0.25 | 0.7 | 16.1 | 0.55 | 32.1 | 0.45 | 5.7 | 0.30 | 8.2 | 0.80 | 42.5 | 0.80 |
| 200 | 0.50 | 0.7 | 32.2 | 0.60 | 57.0 | 0.50 | 10.2 | 0.45 | 14.7 | 0.75 | 87.1 | 0.80 |
| 400 | 0.25 | 0.3 | 7.4 | 0.40 | 11.2 | 0.50 | 1.5 | 0.20 | 3.1 | 0.40 | 11.6 | 0.80 |
| 400 | 0.50 | 0.3 | 11.0 | 0.40 | 15.1 | 0.55 | 2.3 | 0.25 | 4.1 | 0.40 | 17.3 | 0.80 |
| 400 | 0.25 | 0.5 | 5.9 | 0.80 | 8.0 | 0.70 | 0.7 | 0.60 | 2.3 | 0.80 | 13.5 | 0.80 |
| 400 | 0.50 | 0.5 | 10.0 | 0.70 | 14.4 | 0.80 | 1.3 | 0.55 | 3.8 | 0.80 | 21.1 | 0.80 |
| 400 | 0.25 | 0.7 | 6.7 | 0.55 | 16.8 | 0.40 | 2.0 | 0.20 | 4.2 | 0.65 | 22.3 | 0.70 |
| 400 | 0.50 | 0.7 | 12.8 | 0.55 | 36.6 | 0.40 | 3.7 | 0.25 | 7.7 | 0.70 | 49.5 | 0.75 |
| 200 | 0.25 | 0.3 | 74.5 | 0.30 | 118.8 | 0.55 | 4.8 | 0.60 | 11.5 | 0.30 | 90.3 | 0.30 |
| 200 | 0.50 | 0.3 | 109.0 | 0.40 | 157.9 | 0.80 | 7.1 | 0.45 | 19.3 | 0.40 | 141.8 | 0.30 |
| 200 | 0.25 | 0.5 | 48.4 | 0.80 | 20.6 | 0.80 | 5.2 | 0.65 | 11.5 | 0.40 | 27.5 | 0.80 |

(continued)

**Table 1** (continued)

| $n$ | Cens. | $\tau$ | $\hat{\beta}$ | $h_{opt}$ | $\hat{m}(0.4)$ | $h_{opt}$ | $\hat{m}(0.7)$ | $h_{opt}$ | $\hat{m}(1)$ | $h_{opt}$ | $\hat{m}(1.3)$ | $h_{opt}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 200 | 0.50 | 0.5 | 81.3 | 0.80 | 31.6 | 0.80 | 10.0 | 0.70 | 21.4 | 0.55 | 44.5 | 0.80 |
| 200 | 0.25 | 0.7 | 47.6 | 0.80 | 103.5 | 0.55 | 11.9 | 0.25 | 13.5 | 0.80 | 22.0 | 0.80 |
| 200 | 0.50 | 0.7 | 89.0 | 0.80 | 120.5 | 0.65 | 26.3 | 0.80 | 28.9 | 0.80 | 41.7 | 0.80 |
| 400 | 0.25 | 0.3 | 27.5 | 0.25 | 61.7 | 0.50 | 2.9 | 0.55 | 2.9 | 0.30 | 26.6 | 0.20 |
| 400 | 0.50 | 0.3 | 39.8 | 0.30 | 85.3 | 0.50 | 3.3 | 0.60 | 4.3 | 0.35 | 37.2 | 0.25 |
| 400 | 0.25 | 0.5 | 23.0 | 0.80 | 13.8 | 0.80 | 2.9 | 0.70 | 5.3 | 0.35 | 12.8 | 0.45 |
| 400 | 0.50 | 0.5 | 38.3 | 0.80 | 17.5 | 0.80 | 4.2 | 0.70 | 7.1 | 0.40 | 18.6 | 0.50 |
| 400 | 0.25 | 0.7 | 23.4 | 0.80 | 68.4 | 0.50 | 4.5 | 0.20 | 6.1 | 0.80 | 13.6 | 0.35 |
| 400 | 0.50 | 0.7 | 41.9 | 0.80 | 91.0 | 0.55 | 10.0 | 0.80 | 11.3 | 0.80 | 25.3 | 0.35 |

**Table 2** Relative Efficiency of $\hat{\beta}$ and of $\hat{m}$ in Model 1 ($d = 6$) based on the 10% trimmed MSE cross-validation criterion

| $n$ | Cens. | $\tau$ | $\hat{\beta}$ | $\hat{m}(0.7)$ | $\hat{m}(1)$ | $\hat{m}(1.3)$ |
|---|---|---|---|---|---|---|
| 200 | 0.25 | 0.3 | 0.88 | 0.91 | 0.88 | 0.65 |
| 200 | 0.50 | 0.3 | 0.61 | 0.45 | 0.55 | 0.63 |
| 200 | 0.25 | 0.7 | 0.79 | 0.98 | 0.72 | 0.61 |
| 200 | 0.50 | 0.7 | 0.72 | 0.83 | 0.88 | 0.64 |
| 400 | 0.25 | 0.3 | 0.91 | 1.04 | 0.46 | 0.78 |
| 400 | 0.50 | 0.3 | 0.82 | 0.79 | 0.99 | 0.72 |
| 400 | 0.25 | 0.7 | 0.81 | 0.99 | 0.90 | 0.76 |
| 400 | 0.50 | 0.7 | 0.74 | 0.87 | 0.76 | 0.68 |

in the textbook Hosmer et al. (2008), see in particular Section 1.3 and the references therein.

The censored, dependent variable of interest $Y$ is the number of days from admission of a drug abusing patient until his/her self-reported return to drug use. While the entire UIS-dataset from the above website consists of (incomplete) data on 628 subjects, we only consider a subsample of size $n = 202$, consisting of patients receiving one particular treatment (long term) and stemming from one particular treatment site (site A). The proportion of censoring, i.e., the proportion of patients that did not return to drug use, is about 21%. We are interested in the effects of 4 (approximately continuous) covariates on the dependent variable: length of treatment in days ($X_1$), age at enrollment ($X_2$), Beck Depression Score at admission ($X_3$), and number of prior drug treatments ($X_4$).

To preprocess the data, we take logarithms of the number of days to return to drug use. The four covariates are standardized to have mean 0 and variance 1. Denote the estimated values of the single-index parameter by $\hat{\beta}(\tau) = (\hat{\beta}_1(\tau), \ldots, \hat{\beta}_4(\tau))' \in S^3$, where $\tau \in \{0.1, 0.3, 0.5, 0.7\}$. Note that due to the proportion of censoring of about 21%, higher quantiles cannot be expected to give any insight into the relationship

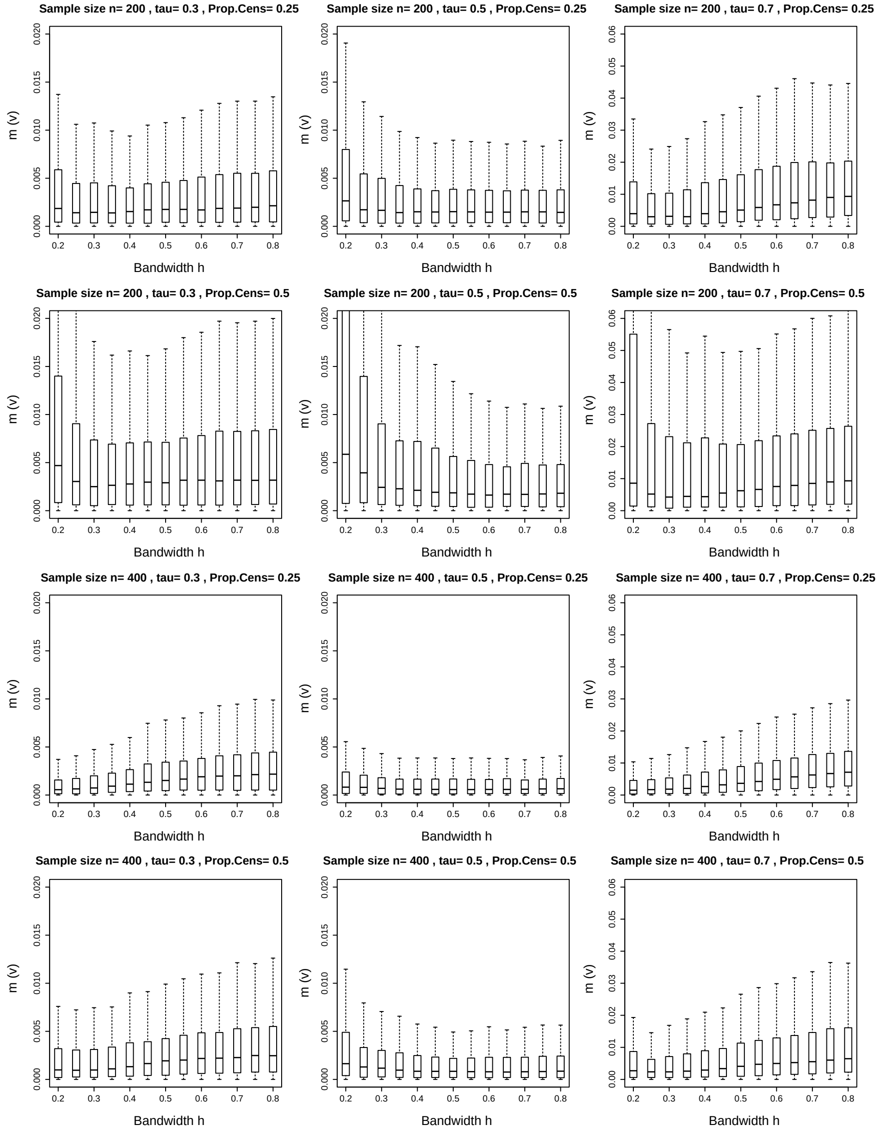


**Fig. 3** Squared estimation error of $\hat{m}(v)$ for $v = 0.85$ against the bandwidth $h$ in Model 1 for $d = 3$. Upper six pictures: $n = 200$, lower six pictures: $n = 400$. Note the different scale in the last column (corresponding to $\tau = 0.7$)

between the dependent variable and the covariates (see also the plot of the observations in Fig. 5). The bandwidth parameters are chosen based on the 10%-trimmed MSE-criterion.

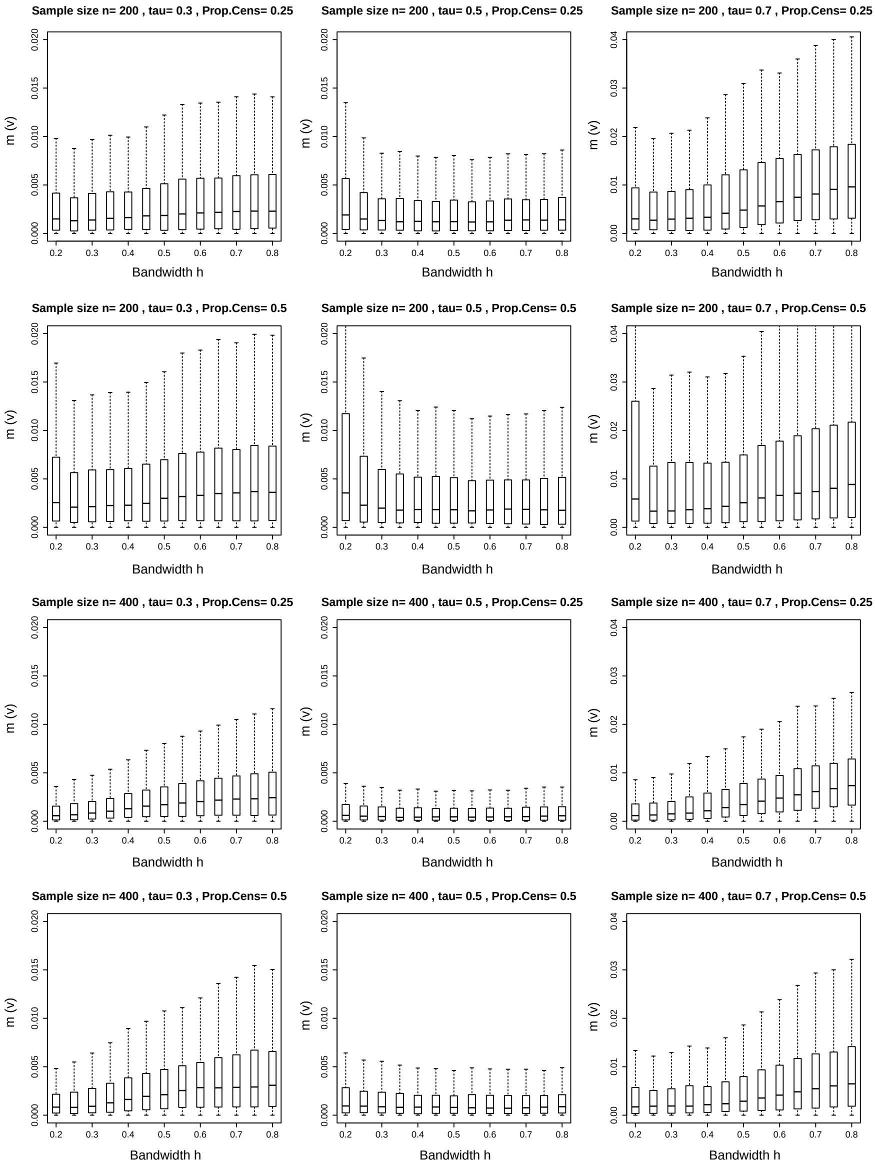


**Fig. 4** Squared estimation error of $\hat{m}(v)$ for $v = 0.85$ against the bandwidth $h$ in Model 2 for $d = 3$. Upper six pictures: $n = 200$, lower six pictures: $n = 400$. Note the different scale in the last column (corresponding to $\tau = 0.7$)

The estimated link functions, based on the 10%-trimmed-mean criterion, are shown in Fig. 5, whereas the estimated single-index parameters are given in Table 3.

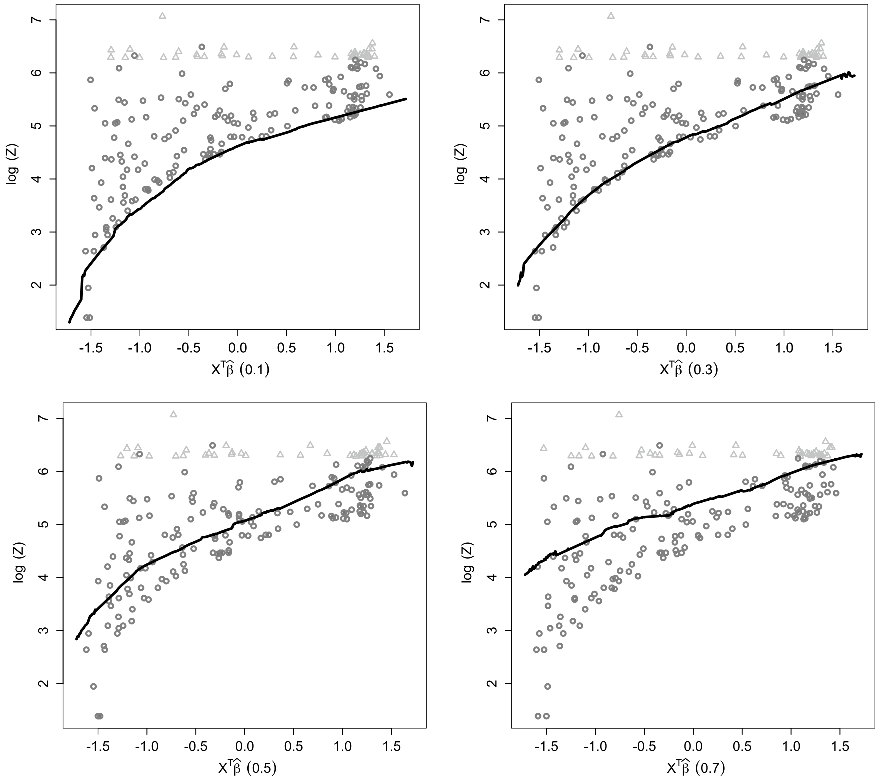


**Fig. 5** Estimated link function $x^T\hat{\beta} \mapsto \hat{m}(x^T\hat{\beta})$, for $\tau \in \{0.1, 0.3, 0.5, 0.7\}$ (from upper left to lower right)

**Table 3** Estimated single-index parameter for the `UIS`-dataset

| $\tau$ | $\hat{\beta}_1(\tau)$ | $\hat{\beta}_2(\tau)$ | $\hat{\beta}_3(\tau)$ | $\hat{\beta}_4(\tau)$ |
|---|---|---|---|---|
| 0.1 | 0.999 | 0.005 | −0.040 | −0.001 |
| 0.3 | 0.999 | 0.007 | −0.041 | 0.004 |
| 0.5 | 0.996 | 0.045 | −0.034 | −0.073 |
| 0.7 | 0.994 | −0.052 | −0.095 | 0.021 |

The triangles and circles in Fig. 5 are the censored and uncensored observations, respectively.

The results reveal some interesting features about the effects of the covariates on the response. First of all, we observe that for all quantile levels under consideration, the covariate "length of treatment in days'' seems to have a more important impact than the three other covariates, since the coefficients of the standardized variables are very different in size, as can be seen from Table 3. As a general conclusion, a longer treatment period results in a longer time until drug abusers return to drug use. The estimated link function is strictly increasing for all quantile levels and non-linear and strictly concave for $\tau \in \{0.1, 0.3, 0.5\}$. Furthermore, it is interesting to note that the strength of concavity increases with decreasing quantile. Hence, the marginal utility of an increase of $X_1$ in its left tail is largest for those patients which generally tend to return to drug abuse rather quickly (i.e., small quantiles of the response—these may be considered as the most interesting group of patients).

**Acknowledgements** The third author acknowledges financial support from the European Research Council (2016–2021, Horizon 2020, and grant agreement 694409). Computational resources have been provided by the supercomputing facilities of the UCLouvain (CISM/UCL) and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CÉCI) funded by the Fonds de la Recherche Scientifique de Belgique under convention 2.5020.11.

## References

Bouaziz, O., & Lopez, O. (2010). Conditional density estimation in a censored single-index regression model. *Bernoulli*, *16*(2), 514–542.

Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*(438), 477–489.

Christou, E., & Akritas, M. G. (2019). Single index quantile regression for censored data. *Statistical Methods & Applications*, *28*, 655–678.

Delecroix, M., Härdle, W., & Hristache, M. (2003). Efficient estimation in conditional single-index regression. *Journal of Multivariate Analysis*, *86*(2), 213–226.

El Ghouch, A., & Van Keilegom, I. (2009). Local linear quantile regression with dependent censored data. *Statistica Sinica*, *19*(4), 1621–1640.

Elsner, J. B., Kossin, J. P., & Jagger, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature*, *455*(7209), 92–95.

Härdle, W., Hall, P., & Ichimura, H. (1993). Optimal smoothing in single-index models. *Annals of Statistics*, *21*(1), 157–178.

He, X., Wang, L., & Hong, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Annals of Statistics*, *41*(1), 342–369.

Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis* (Second ed.). Wiley series in probability and statistics. Hoboken: Wiley-Interscience [Wiley]. Regression modeling of time-to-event data.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics*, *58*(1–2), 71–120.

Klein, R. W., & Spady, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica*, *61*(2), 387–421.

Koenker, R., & Bassett, G, Jr. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50.

Koenker, R., Ng, P., & Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, *81*(4), 673–680.

Kong, E., Linton, O., & Xia, Y. (2013). Global bahadur representation for nonparametric censored regression quantiles and its applications. *Econometric Theory*, 941–968.

Kong, E., & Xia, Y. (2012). A single-index quantile regression model and its estimation. *Econometric Theory*, *28*(4), 730–768.

Leung, D. H.-Y. (2005). Cross-validation in nonparametric regression with outliers. *Annals of Statistics*, *33*(5), 2291–2310.

Lopez, O., Patilea, V., & Van Keilegom, I. (2013). Single index regression models in the presence of censoring depending on the covariates. *Bernoulli*, *19*(3), 721–747.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, *98*(464), 1001–1012.
Strzalkowska-Kominiak, E., & Cao, R. (2013). Maximum likelihood estimation for conditional distribution single-index models under censoring. *Journal of Multivariate Analysis*, *114*, 74–98.
Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis*, *45*(1), 89–103.
Wang, H. J., & Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association 104*(487), 1117–1128.
Wu, T. Z., Yu, K., & Yu, Y. (2010). Single-index quantile regression. *Journal of Multivariate Analysis*, *101*(7), 1607–1621.
Xia, Y., Tong, H., Li, W. K., & Zhu, L.-X. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society Series B 64*(3), 363–410.
Yu, K., & Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, *93*(441), 228–237.
Zheng, Z. G., & Yang, Y. (1998). Cross-validation and median criterion. *Statistica Sinica*, *8*(3), 907–921.