Statistica Sinica **19** (2009), 1621-1640

LOCAL LINEAR QUANTILE REGRESSION WITH DEPENDENT CENSORED DATA

Anouar El Ghouch and Ingrid Van Keilegom

Université catholique de Louvain

Abstract: We consider the problem of nonparametrically estimating the conditional quantile function from censored dependent data. The method proposed here is based on a local linear fit using the check function approach. The asymptotic properties of the proposed estimator are established. Since the estimator is defined as a solution of a minimization problem, we also propose a numerical algorithm. We investigate the performance of the estimator for small samples through a simulation study, and we also discuss the optimal choice of the bandwidth parameters.

Key words and phrases: Censoring, kernel smoothing, local linear smoothing, mixing sequences, nonparametric regression, quantile regression, strong mixing, survival analysis.

1. Introduction

Quantile regression (QR) is a common way to investigate the possible relationships between a covariate X and a response variable Y. Unlike the mean regression method that relies only on the central tendency of the data, the quantile regression approach allows the analyst to estimate the functional dependence between variables for all portions of the conditional distribution of the response variable. In other words, quantile regression extends the framework of estimating only the behavior of the central part of a cloud of data points onto all parts of the conditional distribution. In that sense QR provides a more complete view of relationships between variables of interest. Since it was introduced by Koenker and Bassett (1978) as a robust (to outliers) and flexible (to error distribution) linear regression method based on minimizing asymmetrically weighted absolute residuals, QR has received considerable interest in the literature of theoretical and applied statistics.

In survival (duration for economists) analysis, QR becomes attractive as an alternative to popular regression techniques like the Cox proportional hazards model or the accelerated failure time model; see Koenker and Bilias (2001) and Koenker and Geling (2001). This is mainly due to the transformation equivalence of the quantile operator; see Powell (1986). Thus, one can use any monotone transformation of the response variable to estimate the QR curve and then back transform the estimates to the original scale without loss of information. For a review of recent investigations and development involving QR see for example Yu, Lu, and Stander (2003) or the book by Koenker (2005).

A frequent problem in survival data analysis is censoring, which may be due to different causes: in econometrics censoring can be due to the loss of some subjects under study; in clinical trials censoring can be caused by the end of the follow-up period; in ecology or environmental studies, single or multiple detection limits lead to censored observations. In recent years parametric and semiparametric quantile regression with fixed (type I censoring, namely the Tobit model) or random censoring has begun to receive more attention. See for example Chernozhukov and Hong (2002), Bang and Tsiatis (2002), Honoré, Khan, and Powell (2002), Portnoy (2003), and the references given therein.

As an alternative to restrictions imposed by (semi)-parametric estimators, a vast literature has also been devoted to the nonparametric QR method. With completely observed data, this includes Fan, Hu, and Truong (1994), Yu and Jones (1998), Cai (2002), Gannoun, Saracco, and Yu (2003) among many others. However, under random censoring, the available studies are few. To estimate the conditional quantile function, Dabrowska (1992) and Van Keilegom and Veraverbeke (1998), among others, follow the classical approach of inverting the conditional survival function estimator. The latter is obtained by smoothing with respect to the covariate using either Nadaraya-Watson or Gasser-Müller type weights. Strong asymptotic representations and asymptotic normality have been shown. Following the same idea, Leconte, Poiraud-Casanova, and Thomas-Agnan (2002) proposes to estimate the conditional survival curve and so, by inversion, the quantile function, via a double smoothing technique using Nadaraya-Watson type weights. That is, the resulting estimator is smooth with respect to both the response variable and the covariate. Gannoun, Saracco, Yuan, and Bonney (2005) suggested another approach based on minimizing a weighted integral of the check function (see (2.1) below) over the joint distribution function estimator of Stute (2005). Under strong assumptions on the data generating procedure, they prove the consistency and the asymptotic normality of the proposed estimator.

The literature mentioned above focuses on the i.i.d. case. However in many real applications the data are collected sequentially in time or space, and the assumption of independence does not hold. Here we only give some typical examples from the literature involving correlated data that are subject to censoring. In the clinical trials domain it frequently happens that the patients from the same hospital have correlated survival times due to unmeasured variables like the quality of the hospital equipment; an example of such data can be found in Lipsitz and Ibrahim (2000). Clustering can also be naturally imposed by the experiment, as for example the data analyzed by Yin and Cai (2005) that involve children with inflammation of the middle ear. Censored correlated data are also a common problem in the domain of environmental and spatial (geographical) statistics. In fact, due to the process being used in the data sampling procedure, e.g., the analytical equipment, only the measurements that exceed some thresholds, for example the method detection limits or the instrumental detection limits, can be included in the data analysis; examples of such data can be found in Alavi and Thavaneswaran (2002), Zhao and Frey (2004) and Eastoe, Halsall, Heffernan, and Hung (2006).

Another common assumption in the analysis of censored data is the independence between the covariate and the censoring variable. This assumption is required to make the estimation of the censoring distribution easier (without smoothing); however, it is reasonable only when the censoring is not associated to the characteristic of the individuals under study.

We propose a new nonparametric estimation procedure for quantile regression curves based on the local linear (LL) smoother. This smoothing method was chosen for its many attractive properties: as no boundary effect, design adaptation, and mathematical efficiency. See Fan and Gijbels (1996). In the context of dependent uncensored data the LL approach has been successfully applied to the quantile regression problem by many authors; see for example Yu and Jones (1998), Honda (2000), Cai (2002) and Gannoun et al. (2003).

The estimator proposed in this work is shown to have good properties even if the censored data are correlated, or if the censoring distribution depends on the explanatory variable. The proofs provided for consistency and asymptotic normality are stated under weak conditions. Whenever such conditions are fulfilled, our estimator enjoys similar properties to those of the 'classical' LL estimator for uncensored independent data. Furthermore, to solve the known computational complexity related to QR estimators (like the non-differentiability of the objective function) we adapt the Majorize-Minimize (MM) algorithm as proposed by Hunter and Lange (2000) to our censored case. The resulting algorithm is simple to implement and rapidly converges to the solution.

The paper is organized as follows. In the next section we describe the estimation methodology. In Section 3 we study some asymptotic results for the proposed approach. Section 4 presents the MM algorithm and shows how it can be applied to our situation. In Section 5 we analyze the finite sample performance of the proposed estimator via a simulation study. In Section 6 we discuss the problem of the choice of the smoothing parameters, and we suggest a data-driven procedure based on cross validation. We also study this procedure via a simulation analysis. The proofs of the asymptotic results are provided as a supplementary material available at the following web site `http://www.stat.sinica.edu.tw/statistica`.

2. Methodology

To motivate our approach, we first start with the case where there is no censoring. Let (X_i, Y_i), $i = 1 \ldots, n$ denote the available (uncensored) data points. We denote by $F_x(t)$ the unknown common conditional distribution function (CDF) of Y given $X = x$. Given a (sub)distribution function $L_x(t)$, $\bar{L}_x(t)$ will denote the corresponding survival function , i.e., $\bar{L}_x(t) = 1 - L_x(t)$, and $\dot{L}_x(t)$ its partial derivative with respect to x. We also use $\mathbb{E}_x(.)$ as shorthand for $\mathbb{E}(.|X = x)$. For any $\pi \in (0, 1)$, $Q_\pi(x)$ will denote the conditional quantile function (CQF) of Y given $X = x$. That is, $Q_\pi(x) = \inf\{t : F_x(t) \geq \pi\}$ or, equivalently,

$$Q_\pi(x) = \arg\min_a \mathbb{E}_x\Big(\varphi_\pi(Y - a)\Big), \tag{2.1}$$

where $\varphi_\pi(s) = s(\pi - I(s < 0))$ is the 'check' function and $I(.)$ is the indicator function. As a special case, by taking $\pi = 0.5$ we obtain $med(x) = \arg\min_a \mathbb{E}_x(|Y - a|)$, the conditional median regression function.

For a fixed point x_0 in the support of X, according to Fan et al. (1994) and Yu and Jones (1998), we define the local linear estimators of $Q_\pi(x_0)$ and its derivative, i.e., $\dot{Q}_\pi(x_0) := \partial Q_\pi(x_0)/\partial x$, through:

$$\arg\min_{(\alpha_0,\alpha_1)} \sum_{i=1}^{n} \varphi_\pi\Big(Y_i - \alpha_0 - \alpha_1(X_i - x_0)\Big) K_{h_1}(X_i - x_0), \tag{2.2}$$

where $K_{h_1}(.) = h_1^{-1} K_1(./h_1) \geq 0$, K_1 is a bounded kernel function with bounded support, say $[-1, 1]$, and $0 < h_1 \equiv h_{1n} \to 0$ is a bandwidth parameter satisfying $nh_1 \to \infty$. The key idea behind this procedure is to locally approximate the quantile function in the neighborhood of x_0 via Taylor's formula $Q_\pi(x) \approx \alpha_0 + \alpha_1(x - x_0)$. The kernel K_1 and the smoothing parameter h_1 determine the shape and the width of the local neighborhood.

Unfortunately, the estimation equation (2.2) cannot be used with censored data. In fact, in the presence of censoring, we do not observe Y_i but only $Z_i = \min(Y_i, C_i)$ and $\delta_i = I(Y_i \leq C_i)$, where C_i is the censoring variable, supposed to be independent of Y_i given X_i. However, note that for any a and x,

$$\mathbb{E}_x[I(Y < a)] = \mathbb{E}_x\left[\frac{\delta I(Z < a)}{\bar{G}_X(Z)}\right] \tag{2.3}$$

where, for any x, $\bar{G}_x(.)$ denotes the conditional survival function of C. This, in connection with (2.1), suggests a natural way to extend (2.2) to the censoring case by substituting Y and $I(Y < a)$ by Z and $\delta I(Z < a)\bar{G}_X^{-1}(Z)$, respectively.

Of course, in data analysis, G_x is unknown and needs to be estimated. This can be done via Beran's estimator (see Beran (1981))

$$\bar{\hat{G}}_x(t) := 1 - \hat{G}_x(t) = \prod_{i=1}^{n} \left(1 - \frac{(1-\delta_i)I(Z_i \le t)w_{0i}(x)}{\sum_{j=1}^{n} I(Z_j \ge Z_i)w_{0j}(x)}\right), \tag{2.4}$$

where $w_{0i}(x) = [K_0(X_i - x)/h_0)]/[\sum_{j=1}^{n} K_0((X_j - x)/h_0)]$ are Nadaraya-Watson (NW) weights, K_0 is a kernel function, and $0 < h_0 \equiv h_{0n} \to 0$ is a bandwidth sequence satisfying $nh_0 \to \infty$.

So, for censored data, as an LL estimator for $\beta_0 := Q_\pi(x_0)$ and $\beta_1 := \dot{Q}_\pi(x_0)$ we propose $\hat{\beta} := (\hat{\beta}_0, \hat{\beta}_1)^T \equiv (\hat{Q}_\pi(x_0), \hat{\dot{Q}}_\pi(x_0))^T$ to be the minimizer of $\Gamma_{1,n}(\alpha, x_0)$ over $\alpha := (\alpha_0, \alpha_1)^T$, where $\Gamma_{1,n}(\alpha, x)$ is given by

$$\sum_{i=1}^{n} \Big[Z_i - \alpha_0 - \alpha_1(X_i - x)\Big]\Big[\pi - \frac{\delta_i}{\bar{\hat{G}}_{X_i}(Z_i)} I(Z_i < \alpha_0 + \alpha_1(X_i - x))\Big] K_{h_1}(X_i - x). \tag{2.5}$$

3. Asymptotic Theory

Unlike the mean regression estimator procedure which leads to an explicit solution, there is no closed formula for the proposed estimators. So, to make the asymptotic analysis easier, we start by giving an asymptotic expression for the estimators. To do so, we need to introduce some notations and assumptions that are useful in what follows.

Fix x_0 in the interior of the support of X. We suppose throughout that the survival time Y and the censoring time C are nonnegative random variables with continuous marginal distribution functions, and they are independent given X. We also assume that the distributions of X and of Y given X are absolutely continuous. Denote, respectively, by $f_0(x)$ and $f_x(y)$ the marginal density of X and the conditional density of Y given $X = x$. Let $f(x, y) = f_0(x)f_x(y)$ be the joint density of (X, Y) and assume that $f(x_0, \beta_0) > 0$. The process (X_t, Y_t, C_t), $t = 0, \pm 1, \ldots, \pm\infty$, has the same distribution as (X, Y, C) and is assumed to be stationary α-mixing. By this we mean that if $\mathcal{F}_J^L$ $(-\infty \le J, L \le \infty)$ denotes the σ-field generated by the family $\{(X_t, Y_t, C_t),\ J \le t \le L\}$, then the mixing coefficients

$$\alpha(t) = \sup_{A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_t^{\infty}} \Big|P(A \cap B) - P(A)P(B)\Big|$$

satisfy $\lim_{t\to\infty} \alpha(t) = 0$. For the properties of this and other mixing conditions we refer to Bradley (1986) and Doukhan (1994). In this work, the mixing coefficient $\alpha(t)$ is assumed to be $O(t^{-\nu})$ for some $\nu > 3.5$. We also suppose that the function

$x \to Q_\pi(x)$ is twice differentiable at $x = x_0$. Put $\ddot{Q}_\pi(x_0) = \partial^2 Q_\pi(x_0)/\partial x^2$, $u_j = \int u^j K_1(u)du$, $v_j = \int v^j K_1^2(v)dv$,

$$\Lambda_u = \begin{pmatrix} u_0 & u_1 \\ u_1 & u_2 \end{pmatrix} \quad \text{and} \quad \Omega_v = \begin{pmatrix} v_0 & v_1 \\ v_1 & v_2 \end{pmatrix},$$

and suppose that $|u_1^2 - u_0 u_2| > 0$.

The extra assumptions needed to prove the asymptotic results are listed below. For a given x, define $H_x^0(t) = P(Z \le t, \delta = 0 | X = x) = \int_0^t \bar{F}_x(s)dG_x(s)$, the sub-CDF of censored observations, $H_x(t) = P(Z \le t | X = x) = 1 - \bar{F}_x(t)\bar{G}_x(t)$, the CDF of the observed survival times, and $\mathcal{T}_x = \sup\{t : H_x(t) < 1\}$, the right endpoint of the support of H_x. For any $t < \mathcal{T}_x$, let

$$\zeta_\pi(x,t) = \int_0^t \frac{dF_x(s)}{\bar{G}_x(s)} - \pi^2 \quad \text{and} \quad \sigma_\pi^2(x) = \frac{\zeta_\pi(x, Q_\pi(x))}{f_x^2(Q_\pi(x)) f_0(x)}.$$

(A1) π is such that $Q_\pi(x_0) < \mathcal{T}_{x_0}$.

(A2) (a) $f_0(x)$ and $\ddot{Q}_\pi(x)$ are continuous at $x = x_0$.

(b) $G_x(t)$ and $f_x(t)$ are continuous at $(x,t) = (x_0, \beta_0)$.

(A3) There exists a neighborhood J of x_0 such that:

(a) f_0' exists and is Lipschitz on J.

(b) $x \to \dot{H}_x(t)$ and $x \to \dot{H}_x^0(t)$ exist and are Lipschitz on J uniformly in $t \ge 0$.

(c) $\sup_{j \ge j_*} \sup_{u,v \in J} f_j(u,v) \le M_*$, for some $j_* \ge 1$ and $0 < M_* < \infty$, where $f_j(u,v)$, $j = 1, 2, \ldots$, denotes the density of (X_1, X_{j+1}).

(A4) K_0 is a symmetric density that has a bounded support, say $[-1,1]$, with a first derivative K_0' satisfying $|K_0'(x)| \le \Lambda |x|^\kappa$ for some $\kappa \ge 0$ and $\Lambda > 0$.

(A5) $\zeta_\pi(x,t)$ is continuous at $(x,t) = (x_0, \beta_0)$.

Theorem 1. *Let* (A1)$-$(A4) *hold. If* (i) $nh_1^5 = O(1)$, (ii) $\log(n)/(nh_0^5) = O(1)$, (iii) $nh_1h_0^4 = O(1)$, *and* (iv) $n^{-2\nu+7}(\log n)^{2\nu-3}h_0^{-4(2\nu+1)+12\kappa} = o(1)$ *with* $\kappa < (2\nu-1)/4$ (*see Assumption* (A4)), *then,*

$$\begin{pmatrix} \hat{\beta}_0 - \beta_0 \\ h_1(\hat{\beta}_1 - \beta_1) \end{pmatrix} - \frac{h_1^2}{2} \Lambda_u^{-1} \begin{pmatrix} u_2 \\ u_3 \end{pmatrix} \ddot{Q}_\pi(x_0) = \frac{a_n^2}{f(x_0, \beta_0)} \Lambda_u^{-1} \sum_{t=1}^n e_t \tilde{X}_{ht} K_1(X_{ht}) + r_n,$$

where $r_n = o_p(a_n) + o_p(h_1^2) + O_p(h_0^2)$, $a_n^{-1} = \sqrt{nh_1}$, $e_t = \pi - I(Z_t < Q_\pi(X_t))[\delta_t/(\bar{G}_{X_t}(Z_t))]$, *and* $\tilde{X}_{ht} = (1, X_{ht})^T$ *with* $X_{ht} = h_1^{-1}(X_t - x_0)$.

As a consequence of this theorem, we obtain the asymptotic normality of $(\hat{\beta}_0, \hat{\beta}_1)$.

Theorem 2. *Let* (A1)$-$(A5) *hold with* $j_* = 1$. *If* h_0 *and* h_1 *are such that* (i) $h_1 = C_1 n^{-\gamma_1}$ *for some* $C_1 > 0$ *and* $1/5 < \gamma_1 < 1$, *and conditions* (ii)$-$(iv) *of Theorem* 1 *hold, then*

$$\sqrt{nh_1}\left(\begin{pmatrix}\hat{\beta}_0 - \beta_0 \\ h_1(\hat{\beta}_1 - \beta_1)\end{pmatrix} - \frac{h_1^2}{2}\Lambda_u^{-1}\begin{pmatrix}u_2 \\ u_3\end{pmatrix}\ddot{Q}_\pi(x_0) + O_p(h_0^2) + o_p(h_1^2)\right)$$
$$\xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \sigma_\pi^2(x_0)\Sigma\Big),$$

where $\Sigma = \Lambda_u^{-1}\Omega_v\Lambda_u^{-1}$.

In particular, the following corollary is valid.

Corollary 1. *Under the assumptions of Theorem,* 2 *if* $u_1 = v_1 = 0$, *then* $\hat{\beta}_0$ *and* $\hat{\beta}_1$ *are asymptotically independent and*

$$\sqrt{nh_1}\Big(h_1^i(\hat{\beta}_i - \beta_i) - \frac{u_{2+i}}{2u_{2i}}h_1^2\ddot{Q}_\pi(x_0) + O_p(h_0^2) + o_p(h_1^2)\Big) \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \frac{v_{2i}}{u_{2i}^2}\sigma_\pi^2(x_0)\Big),$$

for $i = 0, 1$.

The result states that, under weak conditions, the LL estimator $\hat{Q}_\pi(x_0)$ proposed here converges to the true quantile parameter $Q_\pi(x_0)$ at the expected rate $\sqrt{nh_1}$. The asymptotic bias and variance of this estimator are given by ($u_0 = 1$),

$$\mathbb{B}ias(\hat{Q}_\pi(x_0)) = \frac{u_2 h_1^2 \ddot{Q}_\pi(x_0)}{2} + O(h_0^2), \text{ and}$$
$$\mathbb{V}ar(\hat{Q}_\pi(x_0)) = \frac{v_0\sigma_\pi^2(x_0)}{nh_1}.$$

These formulas are similar to the 'classical' ones obtained in the independent uncensored case. However, due to the approximation of G_x by $\hat{G}_x$, we can see that there is an extra bias term, $O(h_0^2)$, which may dominate the mean-squared error. If the censoring time C is independent of the covariate X then, instead of (2.4), one may use the Kaplan-Meier estimator and, in that case, the extra bias term becomes $O(\log\log n/n)$. For the asymptotic variance, we can easily verify that $\sigma_\pi^2(x_0)$ is larger than $\pi(1-\pi)/(f_{x_0}^2(Q_\pi(x_0))f_0(x_0))$, which is the expression we obtain when there is no censoring. We can also see that $\sigma_\pi^2(x_0)$ becomes larger as the proportion of censoring in the data increases. The assumption $j_* = 1$ that is made in Theorem 2 is required in order to derive a simple expression for the asymptotic variance. In fact if $j_* > 1$, then we can easily check from the proof given in the Appendix that $nh_1\mathbb{V}ar(\hat{Q}_\pi(x_0)) \rightarrow v_0\sigma_\pi^2(x_0) + 2\rho_*/f^2(x_0, \beta_0)$, where $\rho_* = \lim_{n\to\infty} h_1^{-1}\sum_{j=1}^{j_*-1} C_j^+ < \infty$, with $C_j^+ =$

$\mathbb{C}ov(e_1K_1(X_{h1}), e_{j+1}K_1(X_{h(j+1)}))$. This also shows that the dependence of the observations influences the variance of the estimator. The coefficient ρ_* can be seen as a quantification of such an effect.

As mentioned in the Introduction it is known that the local polynomial regression smoothers automatically correct for boundary effect. A natural question arises whether the estimator $\hat{Q}_\pi$ proposed here still has the same asymptotic properties near the endpoints. Suppose that the support of f_0 is $[0,\infty)$. Take $x_0 = ch_1$ for some $0 < c < 1$ and set $u_{j,c} = \int_{-c}^{1} u^j K_1(u)du$, $v_{j,c} = \int_{-c}^{1} v^j K_1^2(v)dv$, $u_c = (u_{2,c}^2 - u_{1,c}u_{3,c})/(u_{0,c}u_{2,c} - u_{1,c}^2)$, and $v_c = (u_{2,c}^2 v_{0,c} - 2u_{1,c}u_{2,c}v_{1,c} + u_{1,c}^2 v_{2,c})/(u_{0,c}u_{2,c} - u_{1,c}^2)^2$. Under conditions similar to those of Theorem 2, following exactly the same procedure as in the proof of Theorem 1 and 2, it can be shown that $\hat{Q}_\pi(ch_1)$ is asymptotically normal with asymptotic bias and variance given by

$$\mathbb{B}ias(\hat{Q}_\pi(ch_1)) = \frac{u_c h_1^2 \ddot{Q}_\pi(0+)}{2} + O(h_0), \text{ and}$$
$$\mathbb{V}ar(\hat{Q}_\pi(ch_1)) = \frac{v_c \sigma_\pi^2(0+)}{nh_1}.$$

Comparing this result with the one obtained in the interior of the domain of X, we see that the bias is larger. In fact, the bias term due to $|\hat{G}_x - G_x|$ becomes of order h_0 instead of the optimal order h_0^2 available in the interior of the domain. This is clearly due to the fact that in our estimation of G_x we have used the local constant approach which, in contrast to the LL method, suffers from a boundary effect. A boundary kernel or a linear correction may be used to ensure a better behavior of $\hat{G}_x$ near the endpoints. However, such a procedure needs careful investigation since one needs to control the error induced by estimating $G_x(t)$ uniformly in x and t (see the proof of Theorem 1). This is clearly more involved, and thus more difficult, than the current approach based on the local constant smoother. Moreover, the local linear estimator of $G_x(t)$ is not necessarily monotone in t, whereas the local constant procedure produces a monotone increasing estimator on $[0,1]$.

4. Minimization Algorithm

In the previous section we have shown that the QR method combined with LL smoothing has some attractive theoretical features in the context of censored dependent data. However, the application of this method may be restrained by its computational complexity. In fact, for simulations or practical applications with the QR method, especially with large data sets, an efficient optimization routine to solve the mathematical minimization problem imposed by the definition of the QR estimator (see (2.5)) becomes essential. Obviously, classical optimization techniques based on the differentiability of the objective function cannot be used here. With uncensored data, much work has been done to develop an efficient computational tool for QR, especially in the parametric linear case (Simplex algorithm, Interior point algorithm, Smoothing algorithm, MM algorithm, etc). Unfortunately, in our context those 'standard' optimization techniques cannot be used without adaptation. In general, modifying the existing methods is difficult and the performance of the resulting algorithm may not be satisfactory, see for example Fitzenberger (1997) for more about this subject.

Due to its simplicity and numerical stability, we investigate the MM (majorize-minimize) algorithm as explained by Hunter and Lange (2000). First, note that $\Gamma_{1,n}(\alpha, x_0)$, given in (2.5), can be written as $\Gamma_{1,n}(\alpha, x_0) = \sum_{i=1}^n \varphi_\pi(r_i(\alpha), a_i)$, where $\varphi_\pi(r, a) = r[\pi - aI(r < 0)]$, $a_i = \delta_i/\hat{\tilde{G}}_{X_i}(Z_i) \geq 0$, $r_i(\alpha) = \tilde{Z}_i - \alpha^T \mathbf{X}_i$, with $\tilde{Z}_i = K_1((X_i - x_0)/h_1)Z_i$, $\mathbf{X}_i^T = (K_1((X_i - x_0)/h_1), (X_i - x_0)K_1((X_i - x_0)/h_1))$. This linear reparametrization shows that the LL quantile estimators for $(Q_\pi(x_0), \dot{Q}_\pi(x_0))^T$ can be obtained from the parametric quantile regression of $\tilde{Z}_i$ on $\mathbf{X}_i$. Nevertheless, note that the check function, φ_π, depends not only on the residuals $r_i(\alpha)$ but also on the random 'weights' a_i. This makes the optimization problem more difficult than in the classical uncensored case in which $a_i \equiv 1$. Let $\alpha_{(k)}$ denote the kth iterate in finding the minimum point. For notational convenience we omit the parameter α in the expression of $r_i(\alpha)$, that is $r_i \equiv r_i(\alpha)$ and $r_{i(k)} \equiv r_i(\alpha_{(k)})$. The idea behind the MM algorithm is to majorize the underlying function $\varphi_\pi(., a)$ by a surrogate function, say ξ_a, such that at a given iteration k, $\xi_a(r|r_{(k)}) \geq \varphi_\pi(r, a)$, for all r, and $\xi_a(r_{(k)}|r_{(k)}) = \varphi_\pi(r_{(k)}, a)$. Following the idea of Hunter and Lange (2000), taking into account censoring, we propose as a majorizer function

$$\xi_a(r|r_{(k)}) = \frac{1}{4}\left(\frac{ar^2}{\epsilon + |r_{(k)}|} + (4\pi - 2a)r + c_k\right).$$

Here the constant c_k has to be chosen such that $\xi_a(r_{(k)}|r_{(k)}) = \varphi_\pi(r_{(k)}, a)$, and $0 < \epsilon \leq 1$ is a small smoothing parameter to be selected by the analyst. The next iterate $\alpha_{(k+1)}$ is the minimizer of $\sum_{i=1}^n \xi_{a_i}(r_i|r_{i(k)})$ with respect to α. By doing so, it can be shown that $\Gamma_{1,n}(\alpha_{(k+1)}, x_0) \leq \Gamma_{1,n}(\alpha_{(k)}, x_0)$. Arguments similar to those in Hunter and Lange (2000) lead to the iterative algorithm described below. Put $w_{i(k)} = a_i/(\epsilon + |r_{i(k)}|)$ and $v_{i(k)} = a_i - 2\pi - w_{i(k)}r_{i(k)}$. Define $\mathcal{V}_k^T = (v_{1(k)}, \ldots, v_{n(k)})$, $\mathcal{W}_k = diag(w_{1(k)}, \ldots, w_{n(k)})$, $\mathcal{X}^T = [\mathbf{X}_1, \ldots, \mathbf{X}_n]$ and $\mathcal{D}_k = -[\mathcal{X}^T \mathcal{W}_k \mathcal{X}]^{-1}[\mathcal{X}^T \mathcal{V}_k]$.

Algorithm

(0) Choose a small tolerance value, say $\tau = 10^{-6}$. Choose an ϵ such that $\epsilon \ln \epsilon \approx -\tau/n$. Set $k = 0$ and initialize $\alpha_{(0)}$.

(1) Let $i = 0$. Calculate $\mathcal{D}_k$ and set $\tilde{\alpha}_{(k)} = \alpha_{(k)} + \mathcal{D}_k$.

(2) While $\Gamma_{1,n}(\alpha_{(k)}, x_0) \leq \Gamma_{1,n}(\tilde{\alpha}_{(k)}, x_0)$, set $i = i + 1$ and $\tilde{\alpha}_{(k)} = \alpha_{(k)} + 2^{-i}\mathcal{D}_k$.

(3) Set $\alpha_{(k+1)} = \tilde{\alpha}_{(k)}$. If the stopping conditions are not satisfied, replace k by $k + 1$ and go to step (1).

The stopping criteria for this algorithm are satisfied when $\|\alpha_{(k+1)} - \alpha_{(k)}\| < \tau$ and $|\Gamma_{1,n}(\alpha_{(k+1)}, x_0) - \Gamma_{1,n}(\alpha_{(k)}, x_0)| < \tau$. Numerical instability caused, for example, by a bad choice of the smoothing parameters (h_0 and h_1) may lead to divergence of the algorithm, so it is necessary to include a maximum number of iterations that the method is allowed to run. Also, in order to be sure that the resulting optimum point does not correspond to a local minimum it is preferable to re-start the algorithm at least one time with a different starting point.

5. Simulation Study

To evaluate the quality of the proposed method we performed several simulations. The same data generating procedures as those considered in El Ghouch and Van Keilegom (2008) were used. That is, we simulated $n = 300$ observations from the following model

$$Y_t = r(X_t) + \sigma(X_t)\epsilon_t, \quad \text{and} \quad C_t = \tilde{r}(X_t) + \sigma(X_t)\tilde{\epsilon}_t,$$

where $r(x) = 12.5 + 3x - 4x^2 + x^3$ and $\tilde{r}(x) = r(x) + \beta(x)\sigma(x)$, with $\sigma(x) = (x - 1.5)^2 a_0 + a_1$ and $\beta(x) = (x - 1.5)^2 b_0 + b_1$. We took ϵ_t and $\tilde{\epsilon}_t \sim \mathcal{N}(0, 1)$ and X_t to have a uniform distribution on $[0, 3]$, with X_t, ϵ_t, and $\tilde{\epsilon}_t$ mutually independent. By varying b_0 and b_1 we controlled the shape and the amount of censoring, while by varying a_0 and a_1 we changed the variation in the sampled data. Under this model the percentage of censoring (PC, hereafter) was given by $PC(x) = 1 - \Phi(\beta(x)/\sqrt{2})$, where Φ is the distribution function of a standard normal random variable. Four cases were studied.

(1) $b_1 = 0.95$ and $b_0 = 0$, the PC constant and equal to 25%.

(2) $b_1 = 0.95$ and $b_0 = -0.27$, the PC convex with minimum, 25%, at $x = 1.5$.

(3) $b_1 = 0$ and $b_0 = 0$, the PC constant and equal to 50%.

(4) $b_1 = 0$ and $b_0 = -0.238$, the PC convex with minimum, 50%, at $x = 1.5$.

Three values for a_0 were investigated: $a_0 = 0$, $a_0 = -0.25$ and $a_0 = 0.25$. The first one corresponds to a homoscedastic regression model in the second (third) case, $\sigma(x)$ is concave (convex) with maximum (minimum) at $x = 1.5$. Our objective was to study the LL estimator of the median conditional regression function $med(x_0) \equiv Q_{0.5}(x_0)$ at $x_0 = 1.5$. Note that, with Y conditionally distributed as a normal, $med(x)$ is actually the conditional mean function $\mathbb{E}[Y|X = x]$. To study the effect of dependency, we generated our data from an autoregressive process, $AR(1)$ defined by $\mathcal{E}_t = \gamma\mathcal{E}_{t-1} + \upsilon_t$ for any arbitrary sequence $\mathcal{E}_t$, where the υ_t are i.i.d. $\mathcal{N}(0, 1)$. To get the desired uniform distribution for X_t we used the probability integral transform method, see El Ghouch and Van Keilegom (2008) for more details about this procedure. Three stationary strong mixing processes were considered in this study:

Model 1: X_t is generated from an $AR(1)$, with $\gamma = 0.5$, ϵ_t and $\tilde{\epsilon_t}$ are i.i.d.

Model 2: X_t is generated from an $AR(1)$, with $\gamma = -0.5$, ϵ_t and $\tilde{\epsilon_t}$ are i.i.d.

Model 3: X_t, ϵ_t and $\tilde{\epsilon_t}$ are from $AR(1)$, with γ equal to 0.8, 0.5 and 0.5, respectively.

Remark. As noted by a referee, the normal distribution is not the best choice to model survival data, since they are usually non-negative. This was done to allow Y to have infinite support and to avoid unnatural assumptions about the data generating process. Note also that in many survival applications, one assumes that the response variable T is related to the covariate X via a transformation model. This includes the Box-Cox model, i.e., $g(T) = \beta X + \epsilon$, see Cai, Tian, and Wei (2005), which itself includes the well-known accelerated failure time model, with $g \equiv \log$. In our case, Y may be seen as a pre-transformed response variable, i.e., $Y = g(T)$.

The bandwidth parameters h_0 and h_1 needed in the estimation procedure ranged from 0.2 to 3 by steps of 0.1 and 0.05, respectively. The estimated regression function was evaluated using the $1,653$ possible combinations of these two bandwidths. We worked with the Epanechnikov kernel, $K(x) = (3/4)(1 - x^2)I(-1 \leq x \leq 1)$, for both the Beran estimator of G_x and for the LL smoother of $med(x)$. The MM algorithm described in Section 4 was used. As a starting value for this iterative procedure, we chose the standard LL estimator of the mean regression function based on all the observed data (both censored and uncensored part). The simulations showed that this initial approximation was good enough for a quick convergence. To evaluate the finite sample performance of our estimator at each scenario, $N = 500$ replications were used. Two distance measures were approximated, the first one the mean absolute deviation error (MADE) given by $N^{-1}\sum_{i=1}^{N}|\hat{Q}_{0.5}^{(i)}(x_0) - r(x_0)|$, and the second one the mean squared error (MSE) defined as $N^{-1}\sum_{i=1}^{N}[\hat{Q}_{0.5}^{(i)}(x_0) - r(x_0)]^2$. Tables 1−3 summarize the results of this simulation study. Each entry in those tables represents the best result, in terms of MADE, obtained over all the tested pairs (h_0, h_1). The minimum obtained value of MSE (denoted below by mse^*) appears in the seventh column within parentheses.

Table 1. Optimal results for Model 1 ($\pi = 0.5$ and $a_1 = 0.5$).

a_0	b_1	b_0	h_0	h_1	MADE $\times 10^{-2}$	MSE(mse^*) $\times 10^{-2}$	BIAS $\times 10^{-2}$	VAR $\times 10^{-2}$
-0.25	0.95	0	1.0	0.65	5.402	0.482(0.481)	2.447	0.423
		-0.27	1.1	0.65	5.221	0.455(0.455)	2.011	0.415
	0	0	0.2	0.70	7.742	1.009(1.000)	5.105	0.748
		-0.238	0.2	0.70	7.571	0.946(0.944)	4.588	0.735
0	0.95	0	1.0	0.65	5.510	0.513(0.510)	2.289	0.461
		-0.27	2.5	0.70	5.349	0.475(0.474)	2.016	0.435
	0	0	0.2	0.65	7.858	1.040(0.987)	5.158	0.773
		-0.238	0.2	0.70	7.798	0.978(0.962)	4.857	0.742
0.25	0.95	0	1.0	0.65	5.660	0.543(0.535)	2.057	0.501
		-0.27	2.1	0.70	5.384	0.478(0.473)	1.240	0.463
	0	0	0.2	0.75	7.869	1.106(1.104)	4.858	0.870
		-0.238	0.2	0.65	7.902	1.111(1.111)	4.807	0.880

Table 2. Optimal results for Model 2 ($\pi = 0.5$ and $a_1 = 0.5$).

a_0	b_1	b_0	h_0	h_1	MADE $\times 10^{-2}$	MSE(mse^*) $\times 10^{-2}$	BIAS $\times 10^{-2}$	VAR $\times 10^{-2}$
-0.25	0.95	0	1.3	0.65	5.660	0.497(0.496)	2.900	0.413
		-0.27	1.4	0.65	5.503	0.468(0.468)	2.490	0.406
	0	0	0.2	0.60	7.992	1.023(1.023)	5.567	0.713
		-0.238	0.2	0.65	7.871	1.000(1.002)	5.013	0.748
0	0.95	0	1.4	0.65	5.788	0.522(0.520)	2.718	0.449
		-0.27	2.9	0.65	5.589	0.489(0.485)	1.841	0.455
	0	0	0.2	0.60	8.108	1.052(1.052)	5.389	0.762
		-0.238	0.2	0.60	8.018	0.986(0.986)	5.122	0.723
0.25	0.95	0	1.5	0.65	5.932	0.547(0.547)	2.510	0.484
		-0.27	2.8	0.75	5.506	0.470(0.470)	2.394	0.413
	0	0	0.2	0.60	8.296	1.101(1.101)	5.289	0.821
		-0.238	0.2	0.60	8.282	1.113(1.113)	5.273	0.835

Table 3. Optimal results for Model 3 ($\pi = 0.5$ and $a_1 = 0.5$).

a_0	b_1	b_0	h_0	h_1	MADE $\times 10^{-2}$	MSE(mse^*) $\times 10^{-2}$	BIAS $\times 10^{-2}$	VAR $\times 10^{-2}$
-0.25	0.95	0	0.7	0.65	6.600	0.700(0.699)	3.417	0.583
		-0.27	0.8	0.65	6.506	0.684(0.683)	3.016	0.593
	0	0	0.2	0.70	9.003	1.310(1.310)	7.040	0.814
		-0.238	0.2	0.70	9.000	1.420(1.402)	7.310	0.885
0	0.95	0	0.9	0.65	6.770	0.742(0.742)	2.966	0.654
		-0.27	0.9	0.60	6.667	0.714(0.714)	1.992	0.674
	0	0	0.2	0.70	9.201	1.430(1.400)	6.890	0.955
		-0.238	0.2	0.70	9.163	1.369(1.369)	6.788	0.908
0.25	0.95	0	0.8	0.60	6.925	0.773(0.773)	2.279	0.721
		-0.27	0.9	0.60	6.826	0.752(0.752)	1.890	0.717
	0	0	0.2	0.70	9.327	1.441(1.441)	6.688	0.993
		-0.238	0.2	0.70	9.012	1.457(1.457)	6.813	0.992

After analyzing and comparing, we have the following remarks. As expected, the MADE and the MSE increased with censoring. On average, the increase in MADE(MSE) was around 41% (105%) as the PC increased by 100%. The justification for this difference between MSE and MADE comes from the fact that the first one, based on the L_2 norm, is more sensitive to extreme values than the second one based on the L_1 norm. Note also that the bias term was more affected by the increase of the PC than the variance term. Also, it seems that changing the shape of the PC curve, from constant to convex, was less critical. This can be explained by the fact that in the linear approximation, the effectively used data are those close to the investigated point $x_0 = 1.5$ ($h_1 \approx 0.65$). Regarding the dependence in the data, first, by comparing Table 1 and 2, we can see that varying the value of γ from 0.5 to -0.5 has relatively little effect on the results, although it seems that under the positive dependence, the LL median estimator behaves better than in the negative dependence case. Second, and more importantly, Table 3, which reports the 'strong' dependence case, indicates relatively large error measures (MADE and MSE). This is essentially due to the increase of the variance of the estimator. The latter remains the most important element of the MSE in all our simulations. Obviously, the resulting variance is also affected by the variation in the simulated data. However, the effect of heteroscedasticity is not clear. This may be explained by the relative robustness of median regression, or more generally quantile regression, to variations in $\sigma^2(.)$. Now, concerning the behavior of the bandwidth parameters, we can remark that h_1 remained almost unchanged in all our results with a small tendency to be large with heavy censoring. This attitude can be attributed to the increase in the variation due to censoring. However, h_0 behaved differently. In fact, as the PC in the data increased, the optimal value of h_0 became smaller. For example, in Model 1 with $a_0 = -0.25$ and $b_0 = 0$, the value of h_0 changed by 80% as PC went from 25% to 50%. This is due to the fact that this bandwidth is only used in the estimation of the censoring distribution, a task that becomes easier as the PC becomes larger. Another interesting remark related to censoring is the fact that with small PC (25%), h_0 was always smaller than h_1. The opposite happened when the PC was large (50%). In addition, we can see that the way h_0 behaved also depends on the degree of dependence in the data and on the shape of the PC curve. To gain further understanding, Figure 1 displays the boxplots of the MADE with respect to $h_0 = 0.2, 0.3, \ldots, 3$ (see Figures (a1), (b1) and (c1)), and with respect to $h_1 = 0.2, 0.3, \ldots, 3$ (see Figures (a2), (b2) and (c2)). We can clearly see that whatever the values are of h_0, the MADE tended to be smaller when h_1 was chosen near the optimal value (between 0.6 and 0.7). We can also see that the variations of MADE were larger as h_1 moved away from its optimal value. This was especially the case for high percentages of

Table 4. Optimal results for Model 1 ($\pi = 0.25$ and $a_1 = 0.5$).

a_0	b_1	b_0	h_0	h_1	MADE $\times 10^{-2}$	MSE(mse^*) $\times 10^{-2}$	BIAS $\times 10^{-2}$	VAR $\times 10^{-2}$
-0.25	0.95	0	1.7	0.60	5.943	0.550(0.550)	3.230	0.445
		-0.27	1.7	0.60	5.819	0.533(0.533)	2.869	0.451
	0	0	0.8	0.60	6.601	0.682(0.682)	3.004	0.592
		-0.238	0.8	0.60	6.540	0.682(0.679)	2.917	0.597
0	0.95	0	1.7	0.65	5.662	0.498(0.498)	2.611	0.430
		-0.27	2.3	0.65	5.555	0.474(0.473)	2.092	0.431
	0	0	1.0	0.70	6.209	0.620(0.618)	2.618	0.549
		-0.238	1.0	0.70	6.150	0.604(0.602)	2.891	0.525
0.25	0.95	0	2.7	0.85	5.209	0.432(0.428)	1.783	0.394
		-0.27	2.7	0.85	5.064	0.402(0.402)	1.533	0.378
	0	0	2.1	1.0	5.851	0.538(0.528)	2.254	0.487
		-0.238	1.8	1.0	5.751	0.508(0.508)	1.409	0.488

censoring and/or high degrees of dependence. Practically, this means that one can have a good idea about the optimal value of h_1 without necessarily having any knowledge of h_0. These remarks remain true in all other studied cases (not shown here). By contrast, it is not always clear which bandwidth h_0 one can choose just by inspecting the left part of Figure 1. In particular, the boxplots (a1), (b1) and (c1) indicate a large instability of the MADE which is only caused by changes in the value of h_1. In general, this instability increases as h_0 becomes larger. So, we can conclude that the main variation in the quality of the proposed estimator is associated with the bandwidth h_1, h_0 captures just a small part of this variability.

We have also investigated the performance of $\hat{Q}_\pi(x_0)$ for $\pi = 0.25$ and $\pi = 0.75$. The results for $\pi = 0.25$ are shown in Table 4. Compared to Table 1 we can see that for moderate percentages of censoring (25%), except the case where $a_0 = 0.25$, the MSE was larger. This is similar to the uncensored data situation and, in our case, is mainly due to the bias component of the estimator. However, the opposite happened when the percentage of censoring increased. In fact, when the PC was 50%, the estimator behaved better for small values of π. To understand why, note that the censoring probability given $Y = Q_\pi(X)$ equals $P(Y > C|Y = Q_\pi(X)) = \int G_x(Q_\pi(x))f_0(x)dx$, and this increases as π becomes larger and so both bias and variance increase. Compared to the case of $\pi = 0.5$, the result that we obtained for $\pi = 0.75$ (not shown here) was, globally, not very satisfactory, especially for high percentages of censoring. In general, due to the censoring mechanism, π has to be chosen such that $Q_\pi(x_0) < \mathcal{T}_{x_0}$ (see (A1)). The difficulty here is that, in practice, the true value of $\mathcal{T}_{x_0}$ is unknown and so one should select a reasonable (not too large) value of π in order to get a consistent estimator for $Q_\pi(x)$.

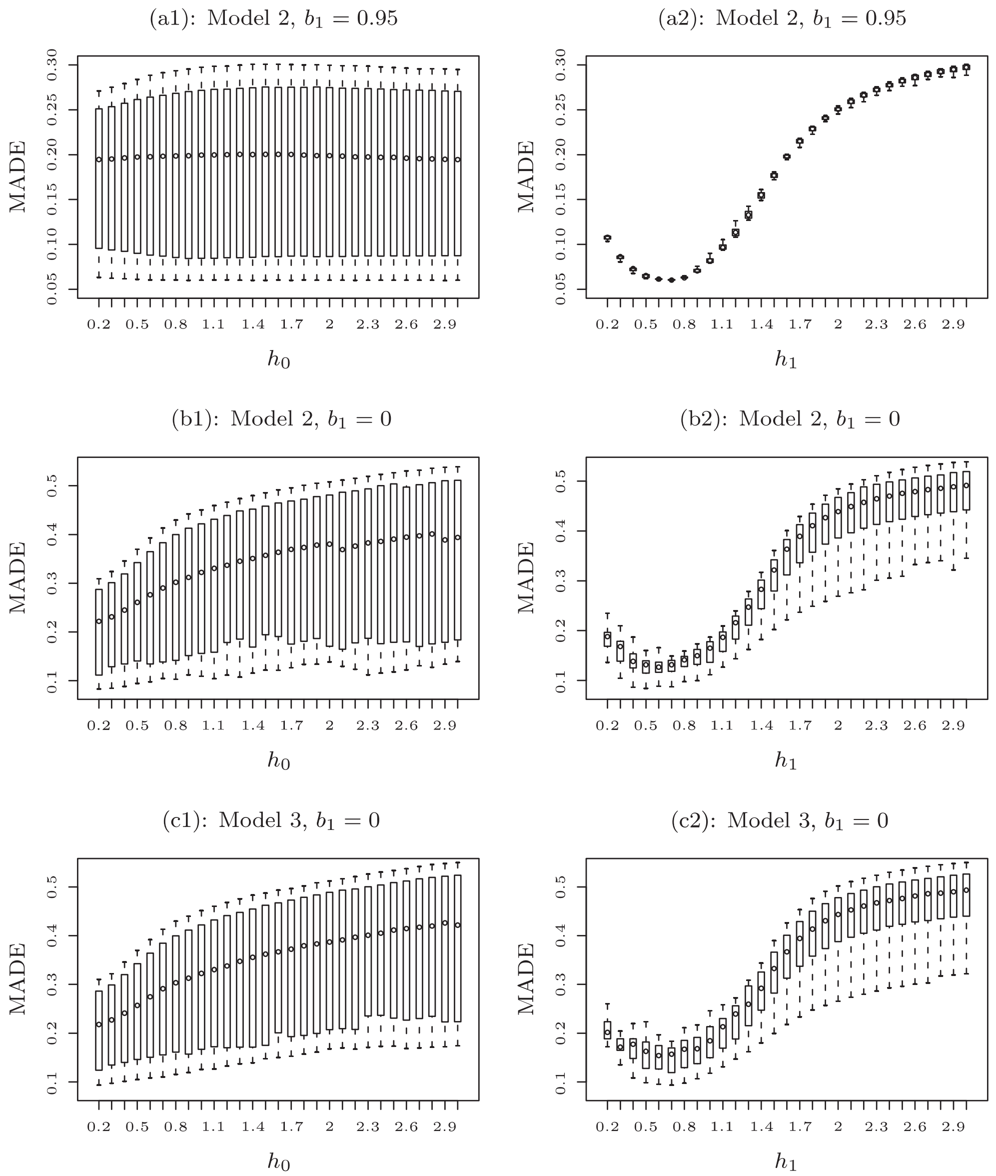

Figure 1. Boxplots of mean absolute deviation error (MADE) for Model 2 and Model 3, with $a_0 = 0.25$, $a_1 = 0.5$, and $b_0 = 0$.

6. Bandwidth Selection

The practical performance of any nonparametric regression technique depends strongly on the smoothing parameters. Choosing an optimal bandwidth is often problematic. In this section we discuss this problem from a practical point of view in the framework of censored QR with dependent data. Much research has been carried out in the area of mean regression with uncensored data. However, when the observations are subject to censoring, the bandwidth selection question is still unsolved and, even in the independent case, there is no consistent method available in the literature. There is also a limited investigation about bandwidth selection in the context of nonparametric (uncensored) quantile regression. See, for example, Yu and Jones (1998), Zheng and Yang (1998) and Leung (2005) for more about this subject. One of the data-driven methods mostly used in the literature is the cross-validation (CV) technique. The CV criterion approximates the prediction error by removing some observations from the process. To be precise, let's focus on the median case and suppose for the moment that the data are uncensored. In such a situation, one may use the following **local leave-block-out** CV statistic :

$$CV_{x_0}(h_1) = n_k^{-1} \sum_{j \in J_k} \phi\Big(\hat{med}_r(X_j) - Y_j\Big), \tag{6.1}$$

where ϕ is a given positive function, J_k (for some $0 < k \leq 1$) is the set of the $n_k = \lfloor nk \rfloor$ nearest neighbor points to x_0 and $\hat{med}_r$ is the LL median estimator defined as in (2.2) but without the observations (X_i, Y_i), $i = 1, \ldots, n$, for which $|i - j| \leq r$. In this study we investigate two choices of ϕ: (1) $\phi(u) = |u|$ and (2) $\phi(u) = u^2$. These choices correspond to the L_1 and the L_2 cross-validation, respectively. The CV rule given by (6.1) can be seen as a generalization of the conventional global L_2-leave-one-out CV ($\phi(u) = u^2$, $k = 1$ and $r = 0$). By leaving out more than one observation ($r > 0$) we omit the data points that may be highly correlated with (X_j, Y_j). On the other hand, with the local adaptation we try to capture the local behavior of the underlying process. Of course, in the case of censoring this procedure cannot be used unless the conditional censoring distribution is known, which is not the case in most practical situations. As an adaptation of this method to the censored situation, we propose the following procedure:

Algorithm

(0) Choose a small value for r and k. Let's say $r = 3$ and $k = 0.25$.

(1) For each j in J_k, do the following: (a) denote by I_j the index set of all the data (X_t, Z_t, δ_t) for which $|t - j| > r$. (b) For each $i \in I_j$, compute $\hat{G}_{X_i}(Z_i)$ as given in (2.4) but with only the observations (Z_t, δ_t, X_t), $t = 1, \ldots, n$, for which $t \in I_j$. (c) Calculate $\hat{med}_r(X_j)$ as given by (2.5) but with only the observations (Z_t, δ_t, X_t), $t = 1, \ldots, n$, for which $t \in I_j$.

(2) Calculate the censored CV criterion $CCV_{x_0}(h_0, h_1) = n_k^{-1} \sum_{j \in J_k} \phi(\hat{med}_r(X_j) - Z_j)$. CCV_{x_0} has to be evaluated several times with different values for h_0 and h_1. A natural selection procedure is to choose (h_0, h_1) that simultaneously minimize $CCV_{x_0}(h_0, h_1)$. Here after we will call this approach '**Method I**'. The second method that we propose is based on the following idea. From the results of the simulation study given in the previous section, it is clear that a consistent choice of the bandwidth h_1 must lead to an estimator with a relative small error term even if the value of h_0 used to estimate G_x is not the optimal one. Also, whenever h_1 is 'good', the error terms, as a function of h_0 should be relatively stable. As a consequence, we propose the following modification:

Method II

(0) Compute $CCV_{x_0}(h_0, h_1)$ for all possible combinations of h_0 and h_1 from some preselected set $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively.

(1) Pick $\hat{h}_1$ for which the values in $\{CCV_{x_0}(h_0, h_1), h_0 \in \mathcal{H}_0\}$, tend (globally) to be small **and** do not change very much (small variation).

(2) Choose $\hat{h}_0$ that minimizes $CCV_{x_0}(h_0, \hat{h}_1)$.
It is better to perform the step (1) of this algorithm via a visual inspection. However, in order to get an automatic approach we propose here to do the following :

- For each value of $h_1 \in \mathcal{H}_1$, let $MC(h_1)$ and $SC(h_1)$ be the mean and the standard deviation of $\{CVV_{x_0}(h_0, h_1),\ h_0 \in \mathcal{H}_0\}$.
- Select the bandwidth h_1 that corresponds to the minimum of $MC(h_1) + \lambda SC(h_1)$.

The parameter $\lambda \geq 0$ determines the trade-off between the mean and the variance. Choosing a big value for λ means that we penalize those values of h_1 that are more affected by changes in h_0.

Due to the high computational cost needed by the cross-validation method, we run a small simulation study based on 100 replications. Our objective here is to compare Method I and Method II. For each simulated dataset, we vary the value of (h_0, h_1) in $\{0.20, 0.45, \ldots, 2.95\} \times \{0.2, 0.7, \ldots, 2.7\}$. From those pairs we select the best one, then we use the latter to compute the LL median estimator that we denote by $\tilde{med}(x_0)$. Let $\hat{med}(x_0)$ be the LL median estimator based on the optimal (fixed) value of (h_0, h_1) as obtained in the last section. As measure of the performance we calculate the empirical mean of $|\tilde{med}(x_0) - \hat{med}(x_0)|$ evaluated over all the simulated data. Almost in all our simulations we have obtained better results using the L_2- cross-validation, that is why we will not show the results corresponding to the L_1 norm. We will also report only

Table 5. The estimated error for Method I and Method II.

Method			Model 2				Model 3			
			I	II			I	II		
a_0	b_1	b_0		λ=0	λ=1	λ=3.5		λ=0	λ=1	λ=3.5
-0.25	0.95	0	0.059	0.060	0.051	0.035	0.060	0.056	0.054	0.043
		-0.27	0.065	0.058	0.050	0.042	0.065	0.057	0.054	0.048
	0	0	0.142	0.083	0.078	0.058	0.143	0.089	0.077	0.063
		-0.238	0.150	0.079	0.078	0.057	0.145	0.100	0.076	0.066
0	0.95	0	0.059	0.054	0.049	0.037	0.057	0.053	0.052	0.041
		-0.27	0.064	0.058	0.052	0.038	0.055	0.051	0.051	0.043
	0	0	0.282	0.133	0.061	0.061	0.278	0.138	0.067	0.070
		-0.238	0.287	0.143	0.066	0.067	0.275	0.135	0.069	0.071
0.25	0.95	0	0.056	0.051	0.046	0.031	0.053	0.048	0.050	0.048
		-0.27	0.063	0.059	0.050	0.042	0.049	0.047	0.045	0.040
	0	0	0.241	0.138	0.081	0.071	0.202	0.101	0.084	0.078
		-0.238	0.253	0.122	0.083	0.074	0.201	0.105	0.075	0.076

the result for $\lambda = 0$, $\lambda = 1$ and $\lambda = 3.5$, this latter was, in general, the best one among a large set of values that we have tested in our simulation study. However, we have remarked that as the percentage of censoring increases, one needs a larger value of λ to get better results. This is clearly due to the fact that heavy censoring leads to more variation (instability) in the resulting estimator which affects the choice of λ. Table 5 displays the results of this study for Model 2 and Model 3. We can see that Method II produces smaller error term than Method I. This is especially clear for highly censored data. In fact with 50% of censoring, we see that Method I fails while Method II still works whenever the used λ is not 'too bad'. With a good value of λ we can notice that Method II enjoys a performance close to the optimal one. This is encouraging, since the data-driven method is feasible while the optimal bandwidths are unknown in real data analysis. We can also remark that there is no significant difference in the performance of Method II between Model 2 and Model 3. Overall, in this simulation study, we have shown the importance of correcting the 'naive' cross-validation approach especially when the censored observations are a large fraction of the available data. Moreover we have demonstrated the accuracy of using Method II to get reasonable bandwidth parameters.

Acknowledgement

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

References

Alavi, A. and Thavaneswaran, A. (2002). Nonparametric estimators for censored correlated data. *Comm. Statist. Theory Methods* **31**, 977-985.

Bang, H. and Tsiatis, A. A. (2002). Median regression with censored cost data. *Biometrics* **58**, 643-649.

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

Bradley, R. C. (1986). Basic properties of strong mixing conditions. In *Dependence in Probability and Statistics* (Edited by E. Eberlein and M. S. Taqqu), 165-192. Birkhäuser Boston Inc.

Cai, T., Tian, L. and Wei, L. J. (2005). Semiparametric Box-Cox power transformation models for censored survival observations. *Biometrika* **92**, 619-632.

Cai, Z. (2002). Regression quantiles for time series data. *Econom. Theory* **18**, 169-192.

Chernozhukov, V. and Hong, H. (2002). Three-step censored quantile regression and extramarital affairs. *J. Amer. Statist. Assoc.* **97**, 872-882.

Dabrowska, D. M. (1992). Nonparametric quantile regression with censored data. *Sankhyā Ser. A* **54**, 252-259.

Doukhan, P. (1994). *Mixing: Properties and Examples.* Lecture Notes in Statistics. Springer.

Eastoe, E. F., Halsall, C. J., Heffernan, J. E. and Hung, H. (2006). A statistical comparison of survival and replacement analyses for the use of censored data in a contaminant air database: A case study from the Canadian Arctic. *Atmospheric Environment* **40**, 6528-6540.

El Ghouch, A. and Van Keilegom, I. (2008). Nonparametric regression with dependent censored data. *Scand. J. Statist.* **35**, 228-247.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*, Volume 66 of *Monographs on Statistics and Applied Probability.* Chapman & Hall, London.

Fan, J., Hu, T. C. and Truong, Y. K. (1994). Robust non-parametric function estimation. *Scand. J. Statist.* **21**, 433-446.

Fitzenberger, B. (1997). A guide to censored quantile regressions. In*Robust inference*, Volume 15 of *Handbook of Statistics*, pp.405-437. North-Holland, Amsterdam.

Gannoun, A., Saracco, J. and Yu, K. (2003). Nonparametric prediction by conditional median and quantiles. *J. Statist. Plann. Inference* **117**, 207-223.

Gannoun, A., Saracco, J., Yuan, A. and Bonney, G. E. (2005). Non-parametric quantile regression with censored data. *Scand. J. Statist.* **32**, 527-550.

Honda, T. (2000). Nonparametric estimation of a conditional quantile for α-mixing processes. *Ann. Inst. Statist. Math.* **52**, 459-470.

Honoré, B., Khan, S. and Powell, J. L. (2002). Quantile regression under random censoring. *J. Econometrics* **109**, 67-105.

Hunter, D. R. and Lange, K. (2000). Quantile regression via an MM algorithm. *J. Comput. Graph. Statist.* **9**, 60-77.

Koenker, R. (2005). *Quantile regression*, Volume **38** of *Econometric Society Monographs.* Cambridge University Press.

Koenker, R. and Bassett, G. Jr. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

Koenker, R. and Bilias, Y. (2001). Quantile regression for duration data: A reappraisal of the Pennsylvania reemployment bonus experiments. *Empirical Economics* **26**, 199-220.

Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: a quantile regression survival analysis. *J. Amer. Statist. Assoc.* **96**, 458-468.

Leconte, E., Poiraud-Casanova, S. and Thomas-Agnan, C. (2002). Smooth conditional distribution function and quantiles under random censorship. *Lifetime Data Anal.* **8**, 229-246.

Leung, D. H.-Y. (2005). Cross-validation in nonparametric regression with outliers. *Ann. Statist.* **33**, 2291-2310.

Lipsitz, S. R. and Ibrahim, J. G. (2000). Estimation with correlated censored survival data with missing covariates. *Biostatistics* **1**, 315-327.

Portnoy, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.* **98**, 1001-1012.

Powell, J. L. (1986). Censored regression quantiles. *J. Econometrics* **32**, 143-155.

Stute, W. (2005). Consistent estimation under random censorship when covariates are present. *Scand. J. Statist.* **32**, 527-550.

Van Keilegom, I. and Veraverbeke, N. (1998). Bootstrapping quantiles in a fixed design regression model with censored data. *J. Statist. Plann. Inference* **69**, 115-131.

Yin, G. and Cai, J. (2005). Quantile regression models with multivariate failure time data. *Biometrics* **61**, 151-161.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93**, 228-237.

Yu, K., Lu, Z. and Stander, J. (2003). Quantile regression: applications and current research areas. *The Statistician* **52**, 331-350.

Zhao, Y. and Frey, H. C. (2004). Quantification of variability and uncertainty for censored data sets and application to air toxic emission factors. *Risk Analysis* **24**, 1019-1034.

Zheng, Z. G. and Yang, Y. (1998). Cross-validation and median criterion. *Statist. Sinica* **8**, 907-921.

Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium.

E-mail: Anouar.ElGhouch@metri.unige.ch

Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium.

E-mail: ingrid.vankeilegom@uclouvain.be

(Received July 2007; accepted June 2008)