Ann Inst Stat Math (2010) 62:439–463
DOI 10.1007/s10463-009-0219-3

Estimation in nonparametric location-scale regression models with censored data

Cédric Heuchenne · Ingrid Van Keilegom

Received: 14 August 2006 / Revised: 19 March 2008 / Published online: 19 March 2009

Abstract Consider the random vector (X, Y), where X is completely observed and Y is subject to random right censoring. It is well known that the completely nonparametric kernel estimator of the conditional distribution $F(\cdot|x)$ of Y given $X = x$ suffers from inconsistency problems in the right tail (Beran 1981, Technical Report, University of California, Berkeley), and hence any location function $m(x)$ that involves the right tail of $F(\cdot|x)$ (like the conditional mean) cannot be estimated consistently in a completely nonparametric way. In this paper, we propose an alternative estimator of $m(x)$, that, under certain conditions, does not share the above inconsistency problems. The estimator is constructed under the model $Y = m(X) + \sigma(X)\varepsilon$, where $\sigma(\cdot)$ is an unknown scale function and ε (with location zero and scale one) is independent of X. We obtain the asymptotic properties of the proposed estimator of $m(x)$, we compare it with the completely nonparametric estimator via simulations and apply it to a study of quasars in astronomy.

Keywords Bandwidth · Bootstrap · Kernel estimation · Nonparametric regression · Right censoring · Survival analysis

This research was supported by Belgian IAP research network grant nr. P5/24.

C. Heuchenne (✉)
HEC-Management School of ULg, University of Liège,
Boulevard du Rectorat, 7, 4000 Liège, Belgium
e-mail: heuchenne@stat.ucl.ac.be

I. Van Keilegom
Institut de Statistique, Université catholique de Louvain,
Voie du Roman Pays, 20, 1348 Louvain-la-Neuve, Belgium

1 Introduction

Consider a random vector (X, Y), where X is a one-dimensional covariate and Y represents the response. We suppose that Y is subject to random right censoring, i.e., instead of observing Y we only observe (Z, Δ), where $Z = \min(Y, C)$, $\Delta = I(Y \leq C)$ and C represents the censoring time, which is supposed to be independent of Y conditionally on X. Let $(Y_i, C_i, X_i, Z_i, \Delta_i)(i = 1, \ldots, n)$ be n independent copies of (Y, C, X, Z, Δ).

It is well known that any location function $m(x)$ that involves the right tail of the conditional distribution $F(\cdot|x) = P(Y \leq \cdot|X = x)$ of Y given $X = x$ (like the conditional mean $E(Y|X = x) = \int y\, dF(y|x)$) often cannot be estimated in a consistent way in a completely nonparametric model, due to the presence of right censoring. In fact, the completely nonparametric (kernel) estimator of $F(\cdot|x)$ is not consistent in the right tail (see Beran 1981) if the conditional distribution of Y has a strictly larger support than the conditional distribution of C. In this paper, we present a way to overcome this problem by imposing the following weak model assumption : we assume that the relation between X and Y is given by

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1}$$

where $m(X)$ and $\sigma(X)$ are some unknown but smooth location and scale functions and the error term ε is independent of X, has location zero and scale one. So, we assume that the conditional distribution of Y given X depends on X only via $m(X)$ and $\sigma(X)$. Under this weak model assumption, we will show that the inconsistency problems can be much reduced. Model (1) has been studied extensively in the literature on censored data; see e.g., Fan and Gijbels (1994), Van Keilegom and Akritas (1999), Einmahl and Van Keilegom (2007), Neumeyer et al. (2006), and Chen et al. (2005).

The method we propose applies to any L-functional of the type (see e.g., Serfling 1980, p. 265):

$$m(x) = a_0 \int_0^1 F^{-1}(s|x)J(s)\,\mathrm{d}s + \sum_{j=1}^{k} a_j F^{-1}(s_j|x), \tag{2}$$

where $F^{-1}(s|x) = \inf\{y : F(y|x) \geq s\}$ is the quantile function of Y given x, $J(s)$ is a given weight function satisfying $\int_0^1 J(s)\mathrm{d}s = 1, k \geq 0, a_0, \ldots, a_k$ are real numbers such that $\sum_{j=0}^{k} a_j = 1$, and $0 \leq s_1, \ldots, s_k \leq 1$. This definition of $m(x)$ includes a very broad class of common location functions. For example, when $J \equiv 1, a_0 = 1$ and $k = 0$, $m(x)$ equals the conditional mean and when $a_0 = 0, k = 1, a_1 = 1$ and $s_1 = 1/2$, we obtain the conditional median.

The method proposed in this paper consists in first estimating the conditional distribution $F(y|x)$ under model (1), and then to plug-in the obtained estimator in (2). To estimate $F(y|x)$, note that under model (1), $\varepsilon^0 = (Y - m^0(X))/\sigma^0(X)$ is independent of X for any location function $m^0(X)$ and scale function $\sigma^0(X)$ (for a formal definition of location and scale functions, see Sect. 2). Hence,

$$F(y|x) = P\left(\varepsilon^0 \le \frac{y - m^0(x)}{\sigma^0(x)}\middle| X = x\right) = F_\varepsilon^0\left(\frac{y - m^0(x)}{\sigma^0(x)}\right), \tag{3}$$

where F_ε^0 is the distribution of ε^0 (for some chosen location and scale functions $m^0(\cdot)$ and $\sigma^0(\cdot)$). The idea is now to identify m^0 and σ^0 in such a way that they can be estimated consistently, i.e., choose location and scale functions that do not make use of the right tail of the distribution of Y given X (like truncated mean and variance). We then estimate $F(y|x)$ by replacing $m^0(\cdot)$, $\sigma^0(\cdot)$ and $F_\varepsilon^0(\cdot)$ by appropriate estimators. It is easy to see that, provided there is a region of the covariate space where censoring is light, the so-obtained estimator of $F(\cdot|x)$ behaves well in the right tail (see Van Keilegom and Akritas 1999). Hence, the estimator of $m(x)$ based on the latter estimator of $F(\cdot|x)$ will outperform the completely nonparametric estimator. This fact is explained in more detail and in a more formal way at the end of Sect. 2.

The estimation of the conditional quantile or mean function with censored data has been studied extensively in the literature. Dabrowska (1987, 1992), Van Keilegom and Veraverbeke (1998), Chen et al. (2005), among others, studied the nonparametric estimation of the conditional quantile function, whereas Powell (1986), Buchinski and Hahn (1998) and Portnoy (2003) estimated this function under the assumption of a parametric model. For the estimation of the conditional mean function, Doksum and Yandell (1982), Dabrowska (1987), Fan and Gijbels (1994), Kim and Truong (1998) and Cai and Hong (2003) used a nonparametric approach, whereas a large number of other papers, including e.g., Buckley and James (1979), Akritas (1994, 1996), Heuchenne and Van Keilegom (2007) assumed a polynomial model for the regression function.

This paper is organized as follows. In the next section, we introduce some notations and describe the estimation procedure in detail. In Sect. 3, we state the asymptotic properties of the estimator obtained in Sect. 2. Section 4 contains a simulation study, in which the new estimator is compared with the corresponding completely nonparametric estimator, while in Sect. 5 a data set on spectral energy distributions of quasars is analyzed by means of the two methods. Finally, the Appendix contains the proofs of the main results of Sect. 3.

2 Notations and description of the method

We assume throughout that regression model (1) holds. Define $F(y|x) = P(Y \le y|x)$, $G(y|x) = P(C \le y|x)$, $H(y|x) = P(Z \le y|x)$, $H_\delta(y|x) = P(Z \le y, \Delta = \delta|x)$, and $F_X(x) = P(X \le x)$. The probability density functions of the distributions defined above will be denoted with lower case letters, and R_X denotes the support of the variable X.

Let $m^0(\cdot)$ be any location function and $\sigma^0(\cdot)$ be any scale function, meaning that $m^0(x) = T(F(\cdot|x))$ and $\sigma^0(x) = S(F(\cdot|x))$ for some functionals T and S that satisfy $T(F_{aY+b}(\cdot|x)) = aT(F_Y(\cdot|x)) + b$ and $S(F_{aY+b}(\cdot|x)) = aS(F_Y(\cdot|x))$, for all $a \ge 0$ and $b \in I\!R$ as in Huber (1981, p. 59 and 202) (here $F_{aY+b}(\cdot|x)$ denotes the conditional distribution of $aY + b$ given $X = x$). Then, it can be easily seen that if model (1) holds, the model $Y = m^0(X) + \sigma^0(X)\varepsilon^0$ with ε^0 independent of X, is also valid.

The estimator of $m(\cdot)$ described below applies this idea to the following choices for $m^0(\cdot)$ and $\sigma^0(\cdot)$:

$$m^0(x) = \int_0^1 F^{-1}(s|x)L(s)\,\mathrm{d}s, \quad \sigma^{02}(x) = \int_0^1 F^{-1}(s|x)^2 L(s)\,\mathrm{d}s - m^{02}(x), \tag{4}$$

where $L(s)$ is a given score function satisfying $\int_0^1 L(s)\,\mathrm{d}s = 1$ and $L(s) \geq 0$ for all $0 \leq s \leq 1$. The key idea will be to choose L in such a way that $m^0(x)$ and $\sigma^0(x)$ can be estimated in a consistent way (a data-driven choice of L is given in Remark 4) and then to use these estimators of $m^0(x)$ and $\sigma^0(x)$ in the construction of an estimator of $m(x)$.

Before explaining the method in detail, let us introduce some more notations. Let $F_\varepsilon(y) = P(\varepsilon \leq y)$ and $S_\varepsilon(y) = 1 - F_\varepsilon(y)$ denote the distribution and survival function of $\varepsilon = (Y - m(X))/\sigma(X)$, where m and σ are the location and scale functions of interest. Likewise, define F_ε^0 and S_ε^0 for the distribution and survival function of $\varepsilon^0 = (Y - m^0(X))/\sigma^0(X)$, where m^0 and σ^0 are defined in (4). Next, for $E^0 = (Z - m^0(X))/\sigma^0(X)$, we denote $H_\varepsilon^0(y) = P(E^0 \leq y)$, $H_{\varepsilon\delta}^0(y) = P(E^0 \leq y, \Delta = \delta)$, $H_\varepsilon^0(y|x) = P(E^0 \leq y|x)$, $H_{\varepsilon\delta}^0(y|x) = P(E^0 \leq y, \Delta = \delta|x)$ ($\delta = 0, 1$) and for $C^0 = (C - m^0(X))/\sigma^0(X)$, we denote $G_\varepsilon^0(y) = P(C^0 \leq y)$.

As explained in the introduction, we first estimate $F(\cdot|x)$ under model (1) using equation (3). The functions m^0 and σ^0 in (3) depend themselves also on $F(\cdot|x)$, which we estimate by means of the completely nonparametric kernel estimator of Beran (1981) (in the case of no ties):

$$\tilde{F}(y|x) = 1 - \prod_{Z_i \leq y, \Delta_i = 1} \left\{ 1 - \frac{W_i(x, a_n)}{\sum_{j=1}^n I(Z_j \geq Z_i) W_j(x, a_n)} \right\}, \tag{5}$$

where

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

($i = 1, \ldots, n$) are Nadaraya–Watson weights, $K_a(\cdot) = a_n^{-1} K(\cdot/a_n)$, K is a density function (kernel) and $\{a_n\}$ a bandwidth sequence. Note that this estimator reduces to the Kaplan and Meier (1958) estimator when all weights $W_i(x, a_n)$ equal n^{-1}. This yields

$$\hat{m}^0(x) = \int_0^1 \tilde{F}^{-1}(s|x)L(s)\,\mathrm{d}s, \quad \hat{\sigma}^{02}(x) = \int_0^1 \tilde{F}^{-1}(s|x)^2 L(s)\,\mathrm{d}s - \hat{m}^{02}(x) \tag{6}$$

as estimators for $m^0(x)$ and $\sigma^{02}(x)$. In practice (see Remark 4 and Sect. 4), the support of the score function L will be chosen in such a way that it estimates a large part of the consistent region of $\tilde{F}(\cdot|x)$, for any x. Next, estimate the residual distribution F_ε^0 (suppose no ties):

$$\hat{F}_\varepsilon^0(y) = 1 - \prod_{\hat{E}_{(i)}^0 \le y, \Delta_{(i)}=1} \left(1 - \frac{1}{n-i+1}\right), \tag{7}$$

where $\hat{E}_i^0 = (Z_i - \hat{m}^0(X_i))/\hat{\sigma}^0(X_i)$, $\hat{E}_{(i)}^0$ is the ith order statistic of $\hat{E}_1^0, \ldots, \hat{E}_n^0$ and $\Delta_{(i)}$ is the corresponding censoring indicator. This estimator has been studied in detail by Van Keilegom and Akritas (1999). This leads to the following estimator of $F(y|x)$:

$$\hat{F}(y|x) = \hat{F}_\varepsilon^0\left(\frac{y - \hat{m}^0(x)}{\hat{\sigma}^0(x)}\right). \tag{8}$$

Finally, we define

$$\hat{m}^T(x) = a_0 \int_{-\infty}^{\hat{T}_x} yJ(\hat{F}(y|x))\, \mathrm{d}\hat{F}(y|x) + \sum_{j=1}^k a_j[\hat{F}^{-1}(s_j|x) \wedge \hat{T}_x], \tag{9}$$

where $\hat{T}_x = T\hat{\sigma}^0(x) + \hat{m}^0(x)$, $T < \tau_{H_\varepsilon^0}$ and $\tau_F = \inf\{y : F(y) = 1\}$ for any distribution F. Like in Van Keilegom and Akritas (1999), T differs from $\tau_{H_\varepsilon^0}$ for technical reasons and should be chosen in practice as close as possible to $\tau_{H_\varepsilon^0}$ (see also Remark 4). As it is clear from (9), $\hat{m}^T(x)$ is actually estimating

$$m^T(x) = a_0 \int_{-\infty}^{T_x} yJ(F(y|x))\, \mathrm{d}F(y|x) + \sum_{j=1}^k a_j[F^{-1}(s_j|x) \wedge T_x], \tag{10}$$

where $T_x = T\sigma^0(x) + m^0(x)$, which can be made arbitrarily close to $m(x)$, provided $\tau_{F_\varepsilon^0} \le \tau_{G_\varepsilon^0}$.

For sake of comparison, the completely nonparametric estimator of $m(x)$ is given by

$$\tilde{m}^T(x) = a_0 \int_{-\infty}^{\tilde{T}_x} yJ(\tilde{F}(y|x))\, \mathrm{d}\tilde{F}(y|x) + \sum_{j=1}^k a_j[\tilde{F}^{-1}(s_j|x) \wedge \tilde{T}_x], \tag{11}$$

where $\tilde{T}_x < \tau_{H(\cdot|x)}$ such that $\inf_{x \in R_X}(1 - H(\tilde{T}_x|x)) > 0$. Note that we truncate at $\tilde{T}_x$, because of the possible inconsistency of $\tilde{F}(y|x)$ for $y > \tilde{T}_x$ (see e.g., Van Keilegom and Veraverbeke 1997).

Note that in the definition of $\hat{m}^T(x)$ we have to truncate at the point $\hat{T}_x$ due to the presence of right censoring. However, T_x is always greater than or equal to the truncation point $\tilde{T}_x$ used in the definition of $\tilde{m}^T(x)$. Indeed,

$$\begin{aligned}\tau_{G^0_\varepsilon} &= \inf\left\{e : \int_{R_X} G^0_\varepsilon(e|x)\mathrm{d}F_X(x) = 1\right\}\\ &= \sup_{x\in R_X} \inf\{e : G^0_\varepsilon(e|x) = 1\} = \sup_{x\in R_X} \inf\left\{\frac{t - m^0(x)}{\sigma^0(x)} : G(t|x) = 1\right\}\\ &= \sup_{x\in R_X} \frac{\tau_{G(\cdot|x)} - m^0(x)}{\sigma^0(x)}.\end{aligned}$$

Since $\tau_{F^0_\varepsilon} = (\tau_{F(\cdot|x)} - m^0(x))/\sigma^0(x)$ for any $x \in R_X$, $\tau_{H^0_\varepsilon} = \tau_{F^0_\varepsilon} \wedge \tau_{G^0_\varepsilon}$ and $\tau_{H(\cdot|x)} = \tau_{F(\cdot|x)} \wedge \tau_{G(\cdot|x)}$, it is clear that $m^0(x) + \sigma^0(x)\tau_{H^0_\varepsilon} \geq \tau_{H(\cdot|x)}$ for any value of x. Moreover, the difference between the two truncation points can be substantial, especially when the censoring proportion is not uniform over x. In fact, if $\tau_{F(\cdot|x)} \leq \tau_{G(\cdot|x)}$ for a small subset of R_X, it is obvious that $\tau_{F^0_\varepsilon} \leq \tau_{G^0_\varepsilon}$. Practically, that means that when there exists a region in the interval R_X of 'light' censoring, then the estimator $\hat{F}^0_\varepsilon$ of the error distribution remains consistent upto far in the right tail (and hence T_x will be large), whereas $\tilde{T}_x$ completely depends on the censoring proportion at the point x. In heavy censored regions $\tilde{T}_x$ can therefore be quite small. This is the main motivation for using $\hat{m}^T(x)$ instead of the completely nonparametric estimator $\tilde{m}^T(x)$.

The following functions enter the asymptotic representation of $\hat{m}^T(x) - m^T(x)$, which we establish in Sect. 3. For a (sub)distribution function $L(y|x)$ we will use the notations $l(y|x) = L'(y|x) = (\partial/\partial y)L(y|x)$, $\dot{L}(y|x) = (\partial/\partial x)L(y|x)$ and similar notations will be used for higher order derivatives. Also, let $\hat{T}_i = \frac{T_{X_i} - \hat{m}^0(X_i)}{\hat{\sigma}^0(X_i)}$, $E^{0T}_i = E^0_i \wedge T$ and $\hat{E}^{0T}_i = \hat{E}^0_i \wedge \hat{T}_i$, $i = 1, \ldots, n$.

$$\xi(z, \delta, y|x) = (1 - F(y|x))\left\{-\int_{-\infty}^{y\wedge z} \frac{\mathrm{d}H_1(s|x)}{(1 - H(s|x))^2} + \frac{I(z \leq y, \delta = 1)}{1 - H(z|x)}\right\},$$

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

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

$$\begin{aligned}B(z, \delta|x) = -f_X^{-1}(x)\sigma^0(x)\Bigg\{&\left[a_0\int_0^{F^0_\varepsilon(T)} J(s)\mathrm{d}s + \sum_{j=1}^k a_j\right]\eta(z, \delta|x)\\ &+ \left[a_0\int_0^{F^0_\varepsilon(T)} (F^0_\varepsilon)^{-1}(s)J(s)\mathrm{d}s + \sum_{j=1}^k a_j((F^0_\varepsilon)^{-1}(s_j) \wedge T)\right]\zeta(z, \delta|x)\Bigg\}.\end{aligned}$$

The assumptions needed for the results of Sect. 3 are listed below.

(A1) (i) $na_n^4 \to 0$ and $na_n^{3+2\delta}(\log a_n^{-1})^{-1} \to \infty$ for some $\delta < 1/2$.
(ii) R_X is a compact interval.
(iii) K has compact support, $\int uK(u)du = 0$ and K is twice continuously differentiable.

(A2) (i) There exist $0 \le s_a \le s_b \le 1$ such that $s_b \le \inf_x F(\tilde{T}_x|x)$, $s_a \le \inf\{s \in [0,1]; L(s) \neq 0\}$, $s_b \ge \sup\{s \in [0,1]; L(s) \neq 0\}$ and $\inf_{x \in R_X} \inf_{s_a \le s \le s_b} f(F^{-1}(s|x)|x) > 0$.
(ii) L is twice continuously differentiable, $\int_0^1 L(s)ds = 1$ and $L(s) \ge 0$ for all $0 \le s \le 1$.

(A3) (i) F_X is three times continuously differentiable and $\inf_{x \in R_X} f_X(x) > 0$.
(ii) m^0 and σ^0 are three times continuously differentiable and $\inf_{x \in R_X} \sigma^0(x) > 0$.
(iii) $E[\varepsilon^{02}] < \infty$ and $E|E^0| < \infty$.

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

(A5) For $L(y|x) = H(y|x)$, $H_1(y|x)$, $H_\varepsilon^0(y|x)$ or $H_{\varepsilon 1}^0(y|x)$: $L'(y|x)$ is continuous in (x, y) and $\sup_{x,y} |y^2 L'(y|x)| < \infty$, and the same holds for all other partial derivatives of $L(y|x)$ with respect to x and y up to order three.

(A6) (i) Let $s_\alpha < F_\varepsilon^0(T)$ and s_β be such that $0 < s_\alpha < s_j < s_\beta < 1$ for all $j = 1, \ldots, k$ and let $Q = [s_\alpha, s_\beta \wedge F_\varepsilon^0(T)]$. Then, $\inf_{s \in Q} f_\varepsilon^0((F_\varepsilon^0)^{-1}(s)) > 0$.
(ii) J is three times continuously differentiable, $\int_0^1 J(s)ds = 1$, $J(s) \ge 0$ for all $0 \le s \le 1$.

(A7) (i) For the density $f_{X|Z,\Delta}(x|z,\delta)$ of X given (Z, Δ), $\sup_{x,z} |f_{X|Z,\Delta}(x|z,\delta)| < \infty$, $\sup_{x,z} |\dot{f}_{X|Z,\Delta}(x|z,\delta)| < \infty$, $\sup_{x,z} |\ddot{f}_{X|Z,\Delta}(x|z,\delta)| < \infty (\delta = 0, 1)$.

Note that some of the above assumptions (assumptions on derivatives, positivity of infima) are needed for technical reasons, but are not inherent to the proposed method. Other assumptions are unavoidable, e.g., the use of kernel smoothing requires continuity, and implies assumptions on the bandwidth and the kernel (see (A1) (i), (A1) (iii), (A3) (ii)). Note that the assumption $na_n^4 \to 0$ is required by Theorem 3.1 of Van Keilegom and Akritas (1999). Since this paper deals with estimators of the Kaplan–Meier type, consistency results cannot be obtained by considering the whole support of distributions of right censored variables (see assumptions (A2)(i) and (A6) (i)). As a consequence, assumption (A2) (i) enables to reduce those supports using truncation points $\tilde{T}_x$, $x \in R_X$, and to identify $m^0(x)$ and $\sigma^0(x)$ in such a way that they can be consistently estimated at any value of x (if $\inf_x F(\tilde{T}_x|x)$ is known).

3 Asymptotic results

In this section, we show the consistency of $\hat{m}^T(x)$ uniformly over x. We also develop an asymptotic representation for $\hat{m}^T(x) - m^T(x)$, which is useful for obtaining afterwards the asymptotic normality.

Theorem 1 *Assume* (A1), (A2), (A3) (i), m^0 *and* σ^0 *are twice continuously differentiable and* $\inf_{x \in R_X} \sigma^0(x) > 0$, (A3) (iii), (A4), (A5), (A6) (i), J *is continuously differentiable,* $\int_0^1 J(s)ds = 1$ *and* $J(s) \geq 0$ *for all* $0 \leq s \leq 1$. *Then,*

$$\sup_{x \in R_X} |\hat{m}^T(x) - m^T(x)| = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).$$

Theorem 2 *Assume* (A1)–(A7) *and* $\sup_e |e^3 (f_\varepsilon^0)''(e)| < \infty$. *Then, for any* $x \in R_X$,

$$\hat{m}^T(x) - m^T(x) = n^{-1} \sum_{i=1}^n K_a(x - X_i) B(Z_i, \Delta_i | x) + R_n(x),$$

where $\sup\{|R_n(x)|; x \in R_X\} = o_P((na_n)^{-1/2})$.

Theorem 3 *Under the assumptions of Theorem 2,*

$$(na_n)^{1/2}(\hat{m}^T(x) - m^T(x)) \xrightarrow{d} N(0, s^2(x)),$$

where

$$s^2(x) = \int K^2(u) \mathrm{d}u \sum_{\delta=0,1} \int B^2(z, \delta | x) f_X(x) \, \mathrm{d}H_\delta(z|x).$$

Remark 1 In order to select an appropriate bandwidth sequence a_n, the bootstrap procedure proposed by Li and Datta (2001) can be used. First, generate $X_1^*, \ldots, X_n^*$ i.i.d. from the empirical distribution of $X_1, \ldots, X_n$. Next, for each $i = 1, \ldots, n$, select at random a Y_i^* from the distribution $\tilde{F}(\cdot|X_i^*)$, and a C_i^* from $\tilde{G}(\cdot|X_i^*)$ (which is the Beran (1981) estimator of $G(\cdot|X_i^*)$ obtained by replacing Δ_i by $1 - \Delta_i$ in the expression of $\tilde{F}(\cdot|X_i^*)$). For the generation of these bootstrap data we use a pilot bandwidth g_n asymptotically larger than the original a_n. Next, let $Z_i^* = \min(Y_i^*, C_i^*)$ and $\Delta_i^* = I(Y_i^* \leq C_i^*)$. For each resample $\{(X_i^{j*}, Z_i^{j*}, \Delta_i^{j*}) : i = 1, \ldots, n\}$, $j = 1, \ldots, B$ for some large B, let $\hat{m}_{a_n}^{*jT}(x)$ be the estimator of $m^T(x)$ obtained by using bandwidth a_n. From this, the integrated mean squared error $\int E[\hat{m}^T(x) - m^T(x)]^2 \, dx$ can be approximated by

$$IMSE^*(a_n) = B^{-1} \sum_{j=1}^B \int [\hat{m}_{a_n}^{*jT}(x) - \hat{m}_{g_n}^T(x)]^2 \, \mathrm{d}x.$$

We now select the value of a_n that minimizes $IMSE^*(a_n)$. The same bootstrap procedure can also be used to approximate the distribution of $\hat{m}^T(x)$, instead of using the above asymptotic distribution, which might be hard to estimate in practice.

Remark 2 A similar idea as the one developed above to estimate $m(x)$, can be used to better estimate any scale function $\sigma(x)$. Indeed, the principle of using Eq. 3 in order to better estimate the right tail of the distribution $F(y|x)$ can also be applied in the construction of an estimator of $\sigma(x)$. Define

$$\hat{\sigma}^{T2}(x) = a_0^\sigma \int_0^{\hat{F}(\hat{T}_x|x)} (\hat{F}^{-1}(s|x) - \hat{m}_0^T(x))^2 J^\sigma(s)\mathrm{d}s + \sum_{j=1}^{k^\sigma} a_j^\sigma \left\{ \int_0^{\hat{F}(\hat{T}_x|x)} \rho_j(\hat{F}^{-1}(s|x) - \hat{m}_j^T(x))\mathrm{d}s \right\}^2,$$

where $\hat{m}_0^T, \ldots, \hat{m}_{k^\sigma}^T$ are general estimators of location functions of the type

$$\hat{m}_p^T(x) = a_{p0}^m \int_0^{\hat{F}(\hat{T}_x|x)} \hat{F}^{-1}(s|x) J_p^m(s)\,\mathrm{d}s + \sum_{j=1}^{k_p^m} a_{pj}^m [\hat{F}^{-1}(s_{pj}^m|x) \wedge \hat{T}_x],$$

$p = 0, \ldots, k^\sigma$, $J^\sigma(s)$ and $J_p^m(s)$ are given weight functions satisfying $\int_0^1 J^\sigma(s)\mathrm{d}s = 1$ and $\int_0^1 J_p^m(s)\mathrm{d}s = 1$, $p = 0, \ldots, k^\sigma$, $k^\sigma \geq 0$, $k_p^m \geq 0$, $a_0^\sigma, \ldots, a_{k^\sigma}^\sigma$ are positive real numbers (a_0^σ can be zero if $k^\sigma > 0$), $a_{p0}^m, \ldots, a_{pk_p^m}^m$ are real numbers such that $\sum_{j=0}^{k_p^m} a_{pj}^m = 1$, $p = 0, \ldots, k^\sigma$, $\rho_j(u) = s_j^\sigma u I(u \geq 0) + (s_j^\sigma - 1)u I(u < 0)$, $j = 1, \ldots, k^\sigma$, and $0 < s_1^\sigma, \ldots, s_{k^\sigma}^\sigma, s_{p1}^m, \ldots, s_{pk_p^m}^m < 1$, $p = 0, \ldots, k^\sigma$. The asymptotic results for $\hat{\sigma}^{T2}(x)$ can be obtained along the same lines as for the estimator $\hat{m}^T(x)$.

Remark 3 Note that when model (1) is homoscedastic (i.e., $\sigma \equiv c$ for some $c > 0$) and we estimate σ^0 by a global estimator $\hat{\sigma}^0$, the representation in Theorem 2 simplifies. In fact, the function $\zeta(z, \delta|x)$ in the definition of $B(z, \delta|x)$ disappears in that case, since this function is coming from the local estimator $\hat{\sigma}^0(x)$ (see proposition 4.9 in Van Keilegom and Akritas 1999).

Remark 4 The estimator $\hat{m}^T(x)$ is easy to implement in practice, and the parameters on which it depends (namely the truncation point T, the bandwidth a_n and the score function L) can be chosen in a data driven way. In Remark 1 we explained already how to choose the bandwidth a_n by means of a bootstrap procedure. The truncation point T can be taken equal to the largest residual $\hat{E}_{(n)}^0$. Finally, for the weight function L in the definition of m^0 and σ^0 we recommend the following function: $L(s) = I(0 \leq s \leq b)/b$, where $b = \min_{1 \leq i \leq n} \tilde{F}(+\infty|X_i)$. In this way, we avoid the values of s for which $\tilde{F}^{-1}(s|X_i)$ is inconsistent, and on the other hand we exploit to a maximum the consistent region.

4 Simulations

In this section, we compare the finite sample behavior of the completely nonparametric location estimator $\tilde{m}^T(x)$ with the location estimator $\hat{m}^T(x)$ proposed in this paper by means of Monte Carlo simulations. We are interested in the behavior of the integrated mean squared error of the estimators, defined by $IMSE = \int E[\hat{m}^T(x) - m(x)]^2\,dx$ for $\hat{m}^T(x)$ and similarly for $\tilde{m}^T(x)$. The simulations are carried out for samples of size $n = 100$ and the results are obtained by using 250 simulations.

In the first setting, we generate i.i.d. data from the normal homoscedastic regression model

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \sigma\varepsilon, \tag{12}$$

for various choices of $\beta_0, \beta_1, \beta_2, \beta_3$ and σ, where X has a uniform distribution on the interval [0, 3], and the error term ε is a normal random variable with zero mean and variance 1. The censoring variable C satisfies $C = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \alpha_3 X^3 + \sigma\varepsilon^*$, for certain choices of $\alpha_0, \alpha_1, \alpha_2$, and α_3, where ε^* has a normal distribution with zero mean and variance 1. We further assume that ε and ε^* are independent of X, that ε is independent of ε^*, and that σ is known. It is easy to see that, under this model,

$$P(\Delta = 0|X = x) = 1 - \Phi\left(\frac{\alpha_0 - \beta_0 + (\alpha_1 - \beta_1)x + (\alpha_2 - \beta_2)x^2 + (\alpha_3 - \beta_3)x^3}{\sqrt{2}\sigma}\right).$$

For the weights that appear in the Beran estimator $\tilde{F}(y|x)$, we choose a biquadratic kernel function $K(x) = (15/16)(1 - x^2)^2 I(|x| \leq 1)$.

For the bandwidth sequence a_n, we select for each estimator the minimizer of an approximated *IMSE* among a grid of 20 possible values of a_n between 0 and 3. This *IMSE* is computed as follows. For each a_n and each simulation, we compute an integrated squared error (*ISE*) using the true parameters of the model (12) and we obtain the approximated *IMSE* for each a_n by averaging those *ISE* over the 250 simulations. A bootstrap technique for computing the smoothing parameter is proposed in Sect. 3, but for simulations it is too computationally intensive. For small values of a_n, it sometimes happens that the window $[x - a_n, x + a_n]$ at a point x does not contain any X_i $(i = 1, \ldots, n)$ for which the corresponding Y_i is uncensored (and in that case estimation of $F(\cdot|x)$ is impossible). We enlarge the window in that case such that it contains at least one uncensored data point in its interior. It also happens sometimes that the bandwidth a_n at a point x is larger than the distance from x to both the left and right endpoint of the interval. In such cases, the bandwidth is redefined as the maximum of these two distances. Finally, we work with $L(s) = I(s \leq b)/b$, where $b = \min_{1 \leq i \leq n} \tilde{F}(+\infty|X_i)$, as recommended in Remark 4.

We compare the two methods for four different locations : the conditional mean, a conditional truncated mean (the trimmed mean corresponding to $J(s) = (1/0.9)$ $I(0.05 < s \leq 0.95)$), the conditional median and conditional third quartile. For the estimators $\tilde{F}(y|x)$ in (11) and $\hat{\tilde{F}}^0_\varepsilon(y)$, the last data point or the last residual is often censored. In this case, this point is redefined as uncensored.

Tables 1, 2 and 3 summarize the simulation results for different values of $\alpha_0, \alpha_1, \alpha_2$, $\alpha_3, \beta_0, \beta_1, \beta_2, \beta_3$ and σ. For fixed values of $\beta_0, \beta_1, \beta_2, \beta_3$ and σ, the values of $\alpha_0, \alpha_1, \alpha_2$ and α_3 are chosen in such a way that some variation in the censoring probability curves is obtained (different proportions of censoring, different degrees of smoothness of the censoring probability curve,...). The proportion of censoring (in %

Table 1 Results for $\tilde{m}^T(x)$ (first line) and $\hat{m}^T(x)$ (second line) for model (12) with large optimal bandwidth a_n

β_0	β_1	β_2	β_3	CP	*IMSE*			
α_0	α_1	α_2	α_3	σ^2	Mean	Trim. mean	Median	3rd quartile
0	0.4	0	0	37.1	0.326	0.331	0.349	0.404
−0.4	1	−0.05	0	0.5	0.320	0.322	0.336	0.365
0	0.4	0	0	38.2	0.357	0.361	0.381	0.429
0.3	0.4	0	0	0.5	0.355	0.356	0.369	0.395
0	0.4	0	0	58.8	0.390	0.396	0.454	0.569
0.24	0	0	0.02	0.5	0.390	0.388	0.408	0.507
0	0.4	0	0	71.1	0.394	0.414	0.507	0.718
−0.3	0	0	0.05	0.5	0.384	0.390	0.445	0.586

Table 2 Results for $\tilde{m}^T(x)$ (first line) and $\hat{m}^T(x)$ (second line) for model (12) with moderately large optimal bandwidth a_n

β_0	β_1	β_2	β_3	CP	*IMSE*			
α_0	α_1	α_2	α_3	σ^2	Mean	Trim. mean	Median	3rd quartile
0	1	0	0	35.5	1.759	1.765	1.802	2.148
2	0	−0.2	0.09	0.5	1.749	1.747	1.762	1.772
0	1	0	0	38.2	1.333	1.347	1.392	1.604
0.3	1	0	0	0.5	1.299	1.303	1.319	1.354
0	1	0	0	58.0	1.631	1.681	1.862	1.926
0.5	0.13	0.2	0	0.5	1.517	1.525	1.547	1.676
0	1	0	0	72.0	1.760	1.832	2.091	2.015
0	0.4	0.1	0	0.5	1.618	1.626	1.698	1.824

Table 3 Results for $\tilde{m}^T(x)$ (first line) and $\hat{m}^T(x)$ (second line) for model (12) with small optimal bandwidth a_n

β_0	β_1	β_2	β_3	CP	*IMSE*			
α_0	α_1	α_2	α_3	σ^2	Mean	Trim. mean	Median	3rd quartile
4	−7.5	6	−1.3	31.7	1.139	1.159	1.260	1.570
3.5	−7.45	7	−1.6	0.5	1.081	1.085	1.100	1.165
4	−7.5	6	−1.3	38.2	1.047	1.066	1.161	1.513
4.3	−7.5	6	−1.3	0.5	1.030	1.034	1.043	1.111
4	−7.5	6	−1.3	51.3	1.251	1.314	1.508	1.559
3.2	−7.6	7	−1.6	0.5	1.142	1.158	1.188	1.315
4	−7.5	6	−1.3	56.4	1.336	1.392	1.553	2.043
3	−7.6	7	−1.6	1	1.296	1.321	1.391	1.620

and denoted by CP in the tables) is computed as the average of $P(\Delta = 0|x)$ for an equispaced grid of values of x.

The tables show that, in general, $\hat{m}^T(x)$ has smaller *IMSE* than $\tilde{m}^T(x)$ for each of the four considered location functions. The higher the quantile, or the smaller the support of J, the worse the estimation. The new method resists however better. The simulations can be explained as follows. The most important problem of the Beran estimator is its consistency in the right tail : this is mainly due to the fact that it is a local estimator. In regions with a large proportion of censored data, the Beran estimator therefore behaves badly. The other estimator also has this problem but at a lower degree : it uses a global estimator of the distribution of the residuals. The inconsistency problems arise thus in the right tail of a global distribution. On the other hand, the new approach is based on the estimation of $m^0(\cdot)$ and $\sigma^0(\cdot)$. The score function L in these functions is determined by $\min_{1\le i\le n} \tilde{F}(+\infty|X_i)$. When censoring is heavy, this value can be small. In that case, the estimators $\hat{m}^0(\cdot)$ and $\hat{\sigma}^0(\cdot)$ will be quite variable and unstable.

The results of Tables 1, 2 and 3 show that the relative performance of the two methods depends on the shape of the regression function and the amount of censoring. In fact, when the regression function is relatively flat, the optimal bandwidth will be quite large. Hence, there will be little difference between the local and global estimators. Table 1 summarizes the results for this kind of regression functions. When the regression function becomes more and more wiggly, the merits of the proposed method become clearer (see Tables 2 and 3). In Tables 2 and 3, the models are more wiggly, leading to smaller bandwidth parameters and hence the advantages of the new estimator in comparison with the completely nonparametric estimator become more and more transparent.

The final setting we consider is a normal heteroscedastic regression model

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + (\gamma X + 0.1)\varepsilon, \tag{13}$$

where X has a uniform distribution on [0, 1] or on [0, 3], and ε has a normal distribution with zero mean and variance equal to one. The censoring variable is given by $C = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \alpha_3 X^3 + \gamma\varepsilon^*$, where ε^* has a normal distribution with zero mean and variance equal to one. We further assume that ε and ε^* are independent of X, and that ε is independent of ε^*. The variance of Y given X is now supposed to be unknown. The results are in Table 4. Not surprisingly, one can show that when the degree of heteroscedasticity is small, the gain in precision of $\hat{m}^T(x)$ with respect to $\tilde{m}^T(x)$ is relatively small, since $\hat{m}^T(x)$ looses some precision due to the estimation of the scale function $\sigma^0(x)$. However, the estimator $\hat{m}^T(x)$ still outperforms $\tilde{m}^T(x)$ for all models and all location functions considered.

5 Data analysis

We illustrate the proposed method on a data set which comes from a study of quasars in astronomy. To date, many studies have focused on the dependence on luminosity and redshift of quasar ultraviolet-to-X-ray spectral energy distributions (characterized

Table 4 Results for $\tilde{m}^T(x)$ (first line) and $\hat{m}^T(x)$ (second line) for model (13). R_X is [0, 3] for the first model and [0, 1] for the two other ones

β_0	β_1	β_2	β_3	CP	*IMSE*			
α_0	α_1	α_2	α_3	σ^2	Mean	Trim. mean	Median	3rd quartile
0	0.4	0	0	58.2	0.365	0.377	0.425	0.957
−0.1	0	0	0.1	0.1	0.338	0.347	0.335	0.943
0	1	6	−4	48.9	0.621	0.631	0.638	0.950
0.5	1	−5	9	1	0.570	0.566	0.557	0.866
0	1	6	−4	56.8	1.040	1.066	1.152	2.546
0.5	0.8	−6	8.5	5	1.032	1.032	1.069	2.161

by means of the spectral index $\alpha_{ox} = 0.384 \log(L_{2\,\mathrm{keV}}/L_{2500\,\text{Å}})$, where $l_{uv} = \log L_{2500\,\text{Å}}$ and $l_x = \log L_{2\,\mathrm{keV}}$ denote the rest-frame 2500 Å and 2 keV luminosity densities) (see Vignali et al. 2003). This allows to obtain information and to validate the proposed mechanism driving quasar broad-band emission (accretion disk onto a super-massive black hole). Due to technical constraints of the used instruments, only upper bounds on 69 of the 137 values of l_x are observed, leading thus to left censoring. Right-censored data points are next obtained by replacing the left-censored $l_{x,i}$ by $Z_i = (\max_{j:j=1,\ldots,137}(l_{x,j}) - l_{x,i}),\ i = 1, \ldots, 137$.

We show in Figs. 1 and 2 the results of regression of l_x on l_{uv} for the new estimator $\hat{m}^T(x)$ and the completely nonparametric estimator $\tilde{m}^T(x)$. The bandwidth is selected from a grid of 18 bandwidths, according to the method described in Remark 1. The selected bandwidth parameter is approximately the same for each method (around 0.75). For the conditional mean, trimmed mean (defined in Sect. 4) and median, the relation we observe between the two variables for both methods suggests to fit a linear model to these data (as made in Heuchenne and Van Keilegom (2007)). For the first quartile, this relation is not so obvious for $\tilde{m}^T(x)$, while the new estimator again suggests to choose a linear model. Note that, contrary to the simulation section, we focus here on the first and not the third quartile. This is because for left censored data, the first quartile is harder to estimate, and hence it interests us more.

Appendix : Proofs of main results

We start with three lemmas, that are needed in the proofs of Theorems 1 and 2. Note that uniform consistency results and asymptotic representations for the local estimators $\hat{m}^0(x)$ and $\hat{\sigma}^0(x)$ are given by Propositions 4.5, 4.8 and 4.9 of Van Keilegom and Akritas (1999). As a consequence, the proofs below mainly develop properties for global quantities (namely location properties for the residuals treated in the three lemmas below, and a result about the quantiles of their distribution treated in the proof of Theorem 1).

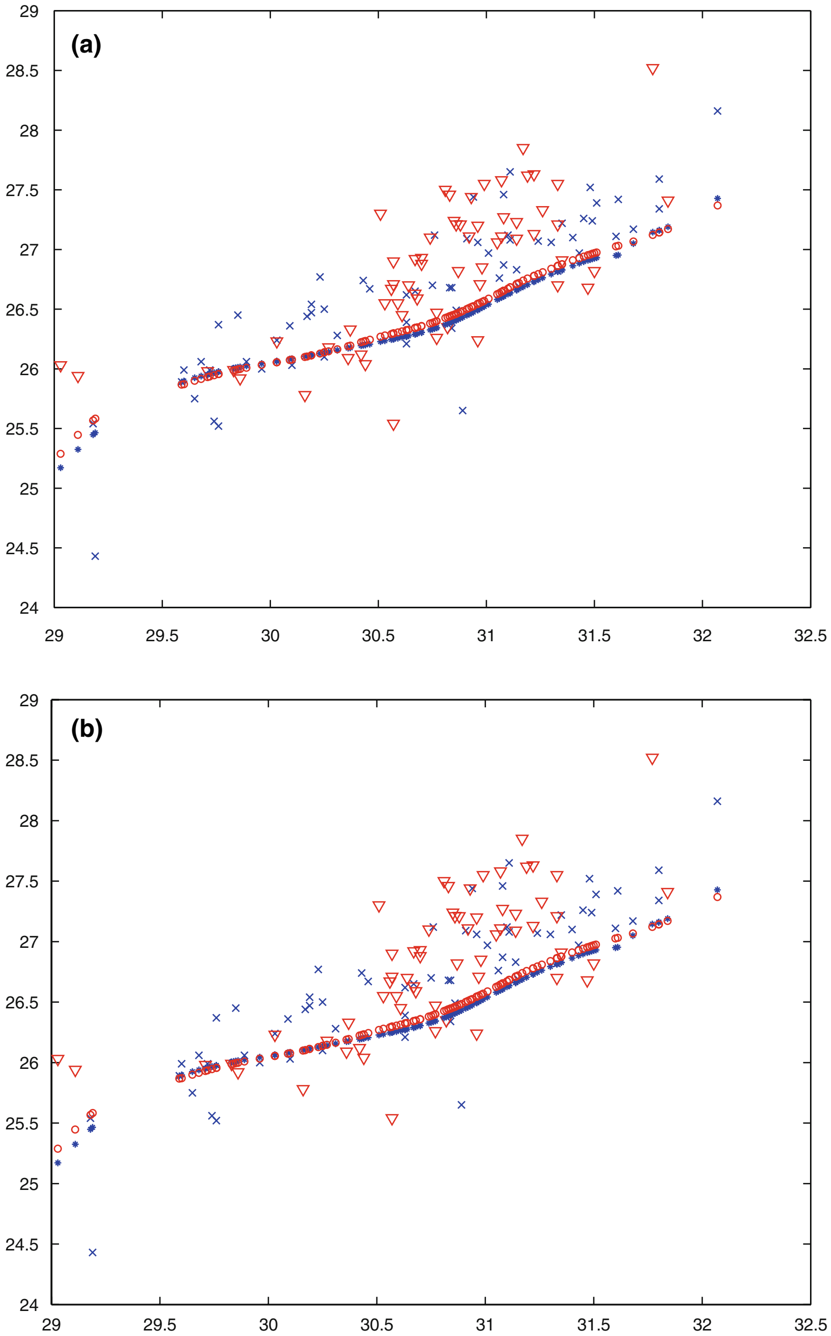
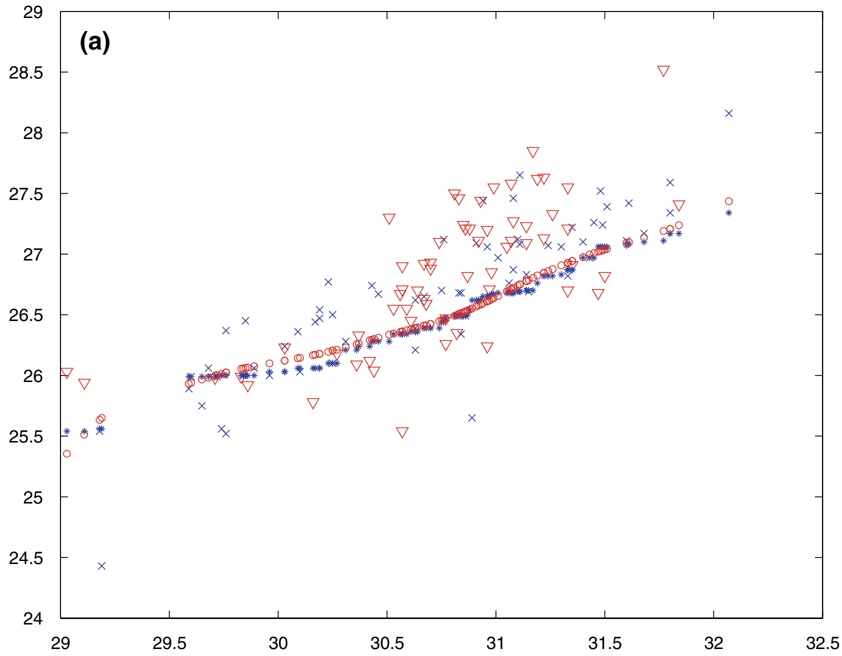


Fig. 1 Regression curve estimation for the quasar data. The estimators $\tilde{m}^T(x)$ and $\hat{m}^T(x)$ are indicated by $*$ and $\circ$ respectively. Uncensored data points are represented by $\times$, and (*left*) censored observations by $\triangledown$. **a** Conditional mean; **b** Conditional truncated mean (5% of truncation at both sides)



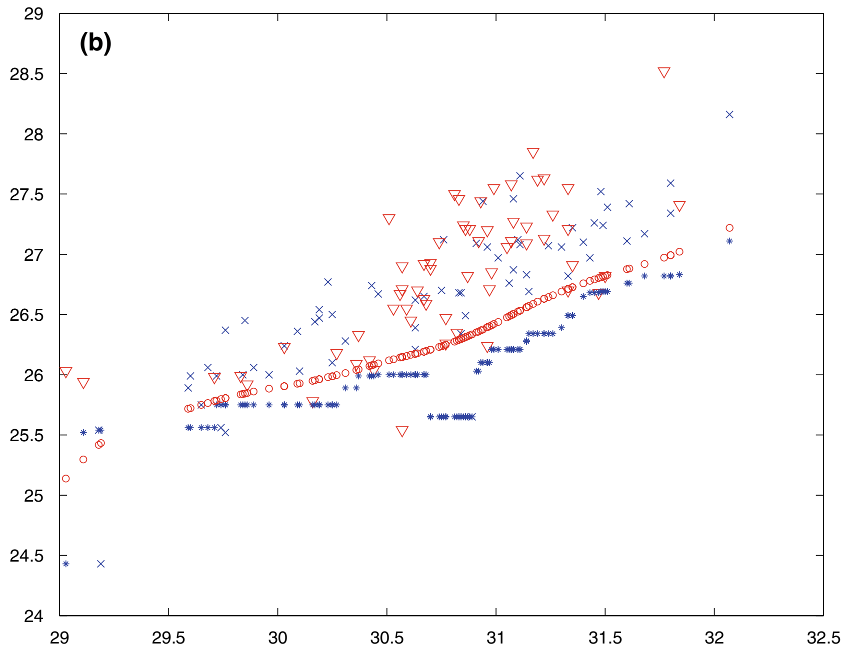


Fig. 2 Regression curve estimation for the quasar data. The estimators $\tilde{m}^T(x)$ and $\hat{m}^T(x)$ are indicated by $*$ and $\circ$ respectively. Uncensored data points are represented by $\times$, and (*left*) censored observations by $\triangledown$. **a** Conditional median; **b** Conditional first quantile

Lemma 1 *Assume* (A1)–(A5), (A6)(ii), (A7) *and* $\sup_e |e^3 (f_\varepsilon^0)''(e)| < \infty$. *Then,*

$$n^{-1}\sum_{i=1}^{n}\left\{\hat{E}_i^0 J(\hat{F}_\varepsilon^0(\hat{E}_i^0)) I(\hat{E}_i^0 \le \hat{T}_i) I(\Delta_i = 1) + \frac{\int_{\hat{E}_i^{0T}}^{\hat{T}_i} e J(\hat{F}_\varepsilon^0(e))\, \mathrm{d}\hat{F}_\varepsilon^0(e)}{1-\hat{F}_\varepsilon^0(\hat{E}_i^{0T})} I(\Delta_i = 0)\right\}$$
$$-\int_{-\infty}^{T} e J(F_\varepsilon^0(e))\, \mathrm{d}F_\varepsilon^0(e) = o_P((na_n)^{-1/2}).$$

Proof The proof is divided into two parts. In a first step, we introduce some properties of the quantities in the statement of the lemma. This leads to double sums, that we treat in a second step. Consider

$$n^{-1}\sum_{i=1}^{n}\{\hat{E}_i^0 J(\hat{F}_\varepsilon^0(\hat{E}_i^0)) I(\hat{E}_i^0 \le \hat{T}_i) - E_i^0 J(F_\varepsilon^0(E_i^0)) I(E_i^0 \le T)\} I(\Delta_i = 1)$$
$$+ n^{-1}\sum_{i=1}^{n}\left\{\frac{\int_{\hat{E}_i^{0T}}^{\hat{T}_i} e J(\hat{F}_\varepsilon^0(e))\, d\hat{F}_\varepsilon^0(e)}{1-\hat{F}_\varepsilon^0(\hat{E}_i^{0T})} - \frac{\int_{E_i^{0T}}^{T} e J(F_\varepsilon^0(e))\, \mathrm{d}F_\varepsilon^0(e)}{1-F_\varepsilon^0(E_i^{0T})}\right\}$$
$$\times I(\Delta_i = 0) = A_1 + A_2. \tag{14}$$

Using Corollary 3.2 and Proposition 4.5 in Van Keilegom and Akritas (1999) (hereafter abbreviated by VKA), the differentiability of J and the fact that $E|\varepsilon^0| < \infty$, we have

$$A_1 = n^{-1}\sum_{i=1}^{n} I(Z_i \le T_{X_i}) I(\Delta_i = 1)\{(\hat{E}_i^0 - E_i^0) J(F_\varepsilon^0(\hat{E}_i^0))$$
$$+ E_i^0[J(F_\varepsilon^0(\hat{E}_i^0)) - J(F_\varepsilon^0(E_i^0))]\} + O_P(n^{-1/2}).$$

Next, using Proposition 4.5 in VKA, and the fact that $\sup_y |y^2 f_\varepsilon^{0'}(y)| < \infty$ and $\sup_y |y f_\varepsilon^0(y)| < \infty$,

$$F_\varepsilon^0(\hat{E}_i^0) - F_\varepsilon^0(E_i^0) = (\hat{E}_i^0 - E_i^0) f_\varepsilon^0(E_i^0) + o_P((na_n)^{-1/2})$$
$$= -\frac{\hat{m}^0(X_i) - m^0(X_i)}{\sigma^0(X_i)} f_\varepsilon^0(E_i^0)$$
$$-\frac{\hat{\sigma}^0(X_i) - \sigma^0(X_i)}{\sigma^0(X_i)} E_i^0 f_\varepsilon^0(E_i^0) + o_P((na_n)^{-1/2}). \tag{15}$$

From this, the fact that J is twice continuously differentiable and that $E|\varepsilon^0| < \infty$, A_1 can be rewritten as

$$\begin{aligned}
A_1 &= n^{-1}\sum_{i=1}^{n} I(Z_i \le T_{X_i}) I(\Delta_i = 1)(\hat{E}_i^0 - E_i^0)[J(F_\varepsilon^0(E_i^0)) \\
&\quad + J'(F_\varepsilon^0(E_i^0))E_i^0 f_\varepsilon^0(E_i^0)] + o_P((na_n)^{-1/2}) \\
&= (n^2 a_n)^{-1}\sum_{i=1}^{n} I(Z_i \le T_{X_i}) I(\Delta_i = 1) f_X^{-1}(X_i)[J(F_\varepsilon^0(E_i^0)) \\
&\quad + J'(F_\varepsilon^0(E_i^0))E_i^0 f_\varepsilon^0(E_i^0)] \\
&\quad \times \left\{\sum_{j=1}^{n} K\left(\frac{X_i - X_j}{a_n}\right)[\eta(Z_j, \Delta_j|X_i) + \zeta(Z_j, \Delta_j|X_i)E_i^0]\right\},
\end{aligned} \tag{16}$$

where the last equality follows from Propositions 4.8 and 4.9 in VKA. Next, we treat the term A_2. Using Corollary 3.2 in VKA, Lemma A1 in Heuchenne and Van Keilegom (2007) and the uniform consistency of $\hat{m}^0$ and $\hat{\sigma}^0$ in (15) (see Proposition 4.5 in VKA), we have

$$\begin{aligned}
A_2 &= n^{-1}\sum_{i=1}^{n} I(\Delta_i = 0)\left\{\frac{\hat{F}_\varepsilon^0(\hat{E}_i^{0T}) - F_\varepsilon^0(E_i^{0T})}{(1-\hat{F}_\varepsilon^0(\hat{E}_i^{0T}))(1-F_\varepsilon^0(E_i^{0T}))}\int_{\hat{E}_i^{0T}}^{\hat{T}_i} eJ(F_\varepsilon^0(e))\,\mathrm{d}\hat{F}_\varepsilon^0(e)\right. \\
&\quad \left. + \frac{1}{1-F_\varepsilon^0(E_i^{0T})}\left[\int_{\hat{E}_i^{0T}}^{\hat{T}_i} eJ(F_\varepsilon^0(e))\,d\hat{F}_\varepsilon^0(e) - \int_{E_i^{0T}}^{T} eJ(F_\varepsilon^0(e))\,\mathrm{d}F_\varepsilon^0(e)\right]\right\} \\
&\quad + o_P((na_n)^{-1/2}) \\
&= n^{-1}\sum_{i=1}^{n} I(\Delta_i = 0)\{A_{21i} + A_{22i} + A_{23i}\} + o_P((na_n)^{-1/2}).
\end{aligned} \tag{17}$$

For A_{21i}, we write

$$\begin{aligned}
\int_{\hat{E}_i^{0T}}^{\hat{T}_i} eJ(F_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e) &= \int_{E_i^{0T}}^{T} eJ(F_\varepsilon^0(e))\mathrm{d}F_\varepsilon^0(e) + \int_{T}^{\hat{T}_i} eJ(F_\varepsilon^0(e))\mathrm{d}\hat{F}_\varepsilon^0(e) \\
&\quad + \int_{\hat{E}_i^{0T}}^{E_i^{0T}} eJ(F_\varepsilon^0(e))\mathrm{d}\hat{F}_\varepsilon^0(e) \\
&\quad + \int_{E_i^{0T}}^{T} eJ(F_\varepsilon^0(e))\mathrm{d}(\hat{F}_\varepsilon^0(e) - F_\varepsilon^0(e)) \\
&= B_{1i} + B_{2i} + B_{3i} + B_{4i}.
\end{aligned}$$

Easy calculations show that the three last terms of this expression are $|E_i^{0T}|$ $O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ uniformly in i, such that

$$n^{-1}\sum_{i=1}^{n} I(\Delta_i=0)A_{21i} = n^{-1}\sum_{i=1}^{n} I(\Delta_i=0)\frac{F_\varepsilon^0(\hat{E}_i^{0T})-F_\varepsilon^0(E_i^{0T})}{(1-F_\varepsilon^0(E_i^{0T}))^2} \times \int_{E_i^{0T}}^{T} eJ(F_\varepsilon^0(e))\,\mathrm{d}F_\varepsilon^0(e) + o_P((na_n)^{-1/2}), \quad (18)$$

using the fact that $E|E^{0T}|<\infty$. Next,

$$n^{-1}\sum_{i=1}^{n} I(\Delta_i=0)\{A_{22i}+A_{23i}\} = n^{-1}\sum_{i=1}^{n} I(\Delta_i=0)\frac{(B_{2i}+B_{3i}+B_{4i})}{1-F_\varepsilon^0(E_i^{0T})}.$$

B_{4i} is $|E_i^{0T}|o_P((na_n)^{-1/2})$ uniformly in i. For B_{3i}, we write

$$\begin{aligned}
&\int_{\hat{E}_i^{0T}}^{E_i^{0T}} eJ(F_\varepsilon^0(e))\mathrm{d}F_\varepsilon^0(e) + \int_{\hat{E}_i^{0T}}^{E_i^{0T}} eJ(F_\varepsilon^0(e))\mathrm{d}(\hat{F}_\varepsilon^0(e)-F_\varepsilon^0(e))\\
&\quad = -\left\{\int_0^{\hat{E}_i^{0T}} eJ(F_\varepsilon^0(e))\mathrm{d}F_\varepsilon^0(e) - \int_0^{E_i^{0T}} eJ(F_\varepsilon^0(e))\mathrm{d}F_\varepsilon^0(e)\right\} + |E_i^{0T}|o_P((na_n)^{-1/2})\\
&\quad = -E_i^{0T}J(F_\varepsilon^0(E_i^{0T}))f_\varepsilon^0(E_i^{0T})[\hat{E}_i^{0T}-E_i^{0T}] + |E_i^{0T}|o_P((na_n)^{-1/2}).
\end{aligned}$$

The last equality is obtained using Proposition 4.5 in VKA, the fact that J is continuously differentiable, that $\sup_e|ef_\varepsilon^0(e)|<\infty$ and that $\sup_e|e^2 f_\varepsilon^{0'}(e)|<\infty$. A similar expression is found for B_{2i}. This together with (18), (17), (16) and (14) leads to

$$A_1+A_2 = (n^2a_n)^{-1}\sum_{i\neq j} B_0(X_i,Z_i,\Delta_i,Z_j,\Delta_j)K\left(\frac{X_i-X_j}{a_n}\right) + o_P((na_n)^{-1/2}), \quad (19)$$

where

$$\begin{aligned}
B_0(X_i,Z_i,\Delta_i,Z_j,\Delta_j) = f_X^{-1}(X_i)\Bigg[& I(\Delta_i=1, Z_i\le T_{X_i})M'(E_i^0)\gamma_{ij}(E_i^0)\\
& + I(\Delta_i=0)\left\{\frac{\int_{E_i^{0T}}^T M(e)\mathrm{d}F_\varepsilon^0(e)}{(1-F_\varepsilon^0(E_i^{0T}))^2} - \frac{M(E_i^{0T})}{1-F_\varepsilon^0(E_i^{0T})}\right\}\\
& \times f_\varepsilon^0(E_i^{0T})\gamma_{ij}(E_i^{0T})\\
& + I(\Delta_i=0)\frac{M(T)}{1-F_\varepsilon^0(E_i^{0T})}f_\varepsilon^0(T)\gamma_{ij}(T)\Bigg],
\end{aligned}$$

$M(e) = eJ(F_\varepsilon^0(e))$ and $\gamma_{ij}(e) = \eta(Z_j, \Delta_j|X_i) + e\zeta(Z_j, \Delta_j|X_i)$.

Next, let $V_k = (X_k, Z_k, \Delta_k)$, $A(V_i, V_j) = B_0(X_i, Z_i, \Delta_i, Z_j, \Delta_j)K(\frac{X_i - X_j}{a_n})$ and $A^*(V_i, V_j) = A(V_i, V_j) - E[A(V_i, V_j)|V_i] - E[A(V_i, V_j)|V_j] + E[A(V_i, V_j)]$. Then, the main term on the right hand side of (19) can be written as

$$
\begin{aligned}
&(n^2a_n)^{-1} \sum_{i \neq j} \left\{A^*(V_i, V_j) + E[A(V_i, V_j)|V_i] + E[A(V_i, V_j)|V_j] - E[A(V_i, V_j)]\right\} \\
&= C_1 + C_2 + C_3 + C_4.
\end{aligned}
$$

First, consider

$$
\begin{aligned}
&(n^2a_n)^{-1} \sum_{i \neq j} E[A(V_i, V_j)|V_i] \\
&= \frac{n-1}{n^2a_n} \sum_{i=1}^{n} \int \sum_{\delta=0,1} \int B_0(X_i, Z_i, \Delta_i, z, \delta)K\left(\frac{X_i - x}{a_n}\right) h_\delta(z|x) f_X(x) \mathrm{d}z\mathrm{d}x \\
&= \frac{n-1}{n^2} \sum_{i=1}^{n} \left\{ \int \sum_{\delta=0,1} \int B_0(X_i, Z_i, \Delta_i, z, \delta)K(u)[h_\delta(z|X_i) - ua_n\dot{h}_\delta(z|X_i) \right. \\
&\quad \left. + O(a_n^2)][f_X(X_i) - a_nuf_X'(X_i) + O(a_n^2)]\ \mathrm{d}z\,\mathrm{d}u \right\} \\
&= \frac{n-1}{n^2} \sum_{i=1}^{n} f_X(X_i) \int \sum_{\delta=0,1} B_0(X_i, Z_i, \Delta_i, z, \delta)h_\delta(z|X_i)\,\mathrm{d}z + O(a_n^2) = O(a_n^2),
\end{aligned}
$$

since $E[\eta(Z, \Delta|X)|X] = E[\zeta(Z, \Delta|X)|X] = 0$. Hence, we also have that $E[A(V_i, V_j)] = O(a_n^2)$. In a similar way, we have for $E[A(V_i, V_j)|V_j]$, using three Taylor developments of order two, that

$$
\begin{aligned}
&(n^2a_n)^{-1} \sum_{i \neq j} E[A(V_i, V_j)|V_j] \\
&\quad = n^{-1} \sum_{j=1}^{n} f_X(X_j) \int \sum_{\delta=0,1} B_0(X_j, z, \delta, Z_j, \Delta_j)\,dH_\delta(z|X_j) + O(a_n^2) \\
&\quad = O_P(n^{-1/2}).
\end{aligned}
$$

For C_1, note that $E[C_1] = 0$ and hence, by Chebyshev' s inequality,

$$
\begin{aligned}
P(|C_1| &> K(na_n)^{-1}E[A^*(V_1, V_2)^2]^{1/2}) \\
&\leq K^{-2}(na_n)^2E[A^*(V_1, V_2)^2]^{-1}E[C_1^2] \\
&= K^{-2}n^{-2}E[A^*(V_1, V_2)^2]^{-1} \sum_{j \neq i} \sum_{m \neq l} E[A^*(V_i, V_j)A^*(V_l, V_m)]. \qquad (20)
\end{aligned}
$$

Since $E[A^*(V_i, V_j)] = 0$, the terms for which $i, j \neq l, m$ are zero. The terms for which either i or j equals l or m and the other differs from l and m, are also zero, because, for example when $i = l$ and $j \neq m$,

$$E[A^*(V_i, V_j)E[A^*(V_i, V_m)|V_i, V_j]] = 0.$$

Thus, only the $2n(n-1)$ terms for which (i, j) equals (l, m) or (m, l) stay such that, (20) is bounded by $2K^{-2}$, which can be made arbitrarily small for K large enough. It now follows that $C_1 = O_P((na_n)^{-1})$ and hence (14) is $o_P((na_n)^{-1/2})$. The result now follows since it is easily seen that (using $E[\varepsilon^{02}] < \infty$)

$$n^{-1}\sum_{i=1}^{n}\left\{E_i^0 J(F_\varepsilon^0(E_i^0))I(E_i^0 \le T)I(\Delta_i = 1) + \frac{\int_{E_i^{0T}}^T eJ(F_\varepsilon^0(e))\mathrm{d}F_\varepsilon^0(e)}{1 - F_\varepsilon^0(E_i^{0T})}I(\Delta_i = 0)\right\}$$
$$-\int_{-\infty}^{T} eJ(F_\varepsilon^0(e))dF_\varepsilon^0(e)| = O_P(n^{-1/2}).$$

□

Remark 5 A weaker version of Lemma 1 can be obtained under less restrictive conditions. In fact, it can be easily seen that if (A1), (A2), (A3) (i) hold, if m^0 and σ^0 are twice continuously differentiable and $\inf_{x \in R_X} \sigma^0(x) > 0$, if (A3) (iii), (A4), (A5) hold and J is continuously differentiable, $\int_0^1 J(s)ds = 1$ and $J(s) \ge 0$ for all $0 \le s \le 1$, then the expression at the left hand side in Lemma 1 is $O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$.

Lemma 2 *Assume* (A1), (A2), (A3) (i), m^0 *and* σ^0 *are twice continuously differentiable and* $\inf_{x \in R_X} \sigma^0(x) > 0$, (A3) (iii), (A4), (A5), J *is continuously differentiable,* $\int_0^1 J(s)\mathrm{d}s = 1$ *and* $J(s) \ge 0$ *for all* $0 \le s \le 1$. *Then,*

$$\int_{-\infty}^{T} eJ(\hat{F}_\varepsilon^0(e))\mathrm{d}\hat{F}_\varepsilon^0(e) - \int_{-\infty}^{T} eJ(F_\varepsilon^0(e))\mathrm{d}F_\varepsilon^0(e) = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).$$

Proof By Lemma 1, it suffices to prove that

$$n^{-1}\sum_{i=1}^{n}\left\{\hat{E}_i^0 J(\hat{F}_\varepsilon^0(\hat{E}_i^0))I(\hat{E}_i^0 \le T) - \hat{E}_i^0 J(\hat{F}_\varepsilon^0(\hat{E}_i^0))I(\hat{E}_i^0 \le \hat{T}_i)\right\} I(\Delta_i = 1)$$
$$+ n^{-1}\sum_{i=1}^{n}\left\{\frac{\int_{\hat{E}_i^0 \wedge T}^T eJ(\hat{F}_\varepsilon^0(e))\mathrm{d}\hat{F}_\varepsilon^0(e)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^0 \wedge T)} - \frac{\int_{\hat{E}_i^{0T}}^{\hat{T}_i} eJ(\hat{F}_\varepsilon^0(e))\mathrm{d}\hat{F}_\varepsilon^0(e)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})}\right\} I(\Delta_i = 0)$$
$$= O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}). \tag{21}$$

The left hand side of (21) can be written as

$$n^{-1}\sum_{i=1}^{n}\left\{\hat{E}_i^0 J(\hat{F}_\varepsilon^0(\hat{E}_i^0))[I(\hat{E}_i^0 \le T) - I(\hat{E}_i^0 \le \hat{T}_i)]\right\} I(\Delta_i = 1)$$
$$+ n^{-1}\sum_{i=1}^{n}\left\{\frac{\int_{\hat{E}_i^0 \wedge T}^T eJ(\hat{F}_\varepsilon^0(e))\mathrm{d}\hat{F}_\varepsilon^0(e)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^0 \wedge T)} I(\hat{E}_i^0 \le T, \hat{E}_i^0 > \hat{T}_i)\right.$$

$$- \frac{\int_{\hat{E}_i^{0T}}^{\hat{T}_i} eJ(\hat{F}_\varepsilon^0(e))\mathrm{d}\hat{F}_\varepsilon^0(e)}{1-\hat{F}_\varepsilon^0(\hat{E}_i^{0T})} I(\hat{E}_i^0 > T, \hat{E}_i^0 \leq \hat{T}_i)$$
$$\left. + \frac{\int_{\hat{T}_i}^{T} eJ(\hat{F}_\varepsilon^0(e))\mathrm{d}\hat{F}_\varepsilon^0(e)}{1-\hat{F}_\varepsilon^0(\hat{E}_i^{0T})} I(\hat{E}_i^0 \leq T, \hat{E}_i^0 \leq \hat{T}_i)\right\} I(\Delta_i = 0). \tag{22}$$

Using classical arguments, the three last terms in the above expression are $O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ and the first one can be rewritten as

$$n^{-1}\sum_{i=1}^n E_i^0 J(F_\varepsilon^0(E_i^0))[I(\hat{E}_i^0 \leq T) - I(\hat{E}_i^0 \leq \hat{T}_i)]I(\Delta_i = 1)$$
$$+ O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}),$$

since $E|E^0| < \infty$. Using arguments similar to those used in Lemma A.1 in VKA, we find that

$$n^{-1}\sum_{i=1}^n \Big\{E_i^0 J(F_\varepsilon^0(E_i^0))\Delta_i[I(\hat{E}_i^0 \leq T) - I(E_i^0 \leq T)]$$
$$- E[E^0 J(F_\varepsilon^0(E^0))\Delta I(\hat{E}^0 \leq T)|\mathcal{X}_n]$$
$$+ E[E^0 J(F_\varepsilon^0(E^0))\Delta I(E^0 \leq T)]\Big\} = o_P(n^{-1/2}), \tag{23}$$

where $E[\cdot|\mathcal{X}_n]$ is the mean conditional on the data (X_j, Z_j, Δ_j), $j = 1, \ldots, n$. Finally, since

$$E[E^0 J(F_\varepsilon^0(E^0))\Delta I(\hat{E}^0 \leq T)|\mathcal{X}_n] - E[E^0 J(F_\varepsilon^0(E^0))\Delta I(E^0 \leq T)]$$
$$= \int_{R_X}\int_T^{\frac{T\hat{\sigma}^0(x)+\hat{m}^0(x)-m^0(x)}{\sigma^0(x)}} eJ(F_\varepsilon^0(e))h_{\varepsilon 1}^0(e|x)f_X(x)\,\mathrm{d}e\,\mathrm{d}x$$
$$= O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}), \tag{24}$$

it follows that the first term of (22) is also $O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$.

The next lemma is a refinement of Lemma 2, obtained under somewhat stronger conditions. □

Lemma 3 *Assume* (A1)–(A5), (A6)(ii), (A7) *and* $\sup_e |e^3(f_\varepsilon^0)''(e)| < \infty$. *Then,*

$$\int_{-\infty}^T eJ(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e) - \int_{-\infty}^T eJ(F_\varepsilon^0(e))\mathrm{d}F_\varepsilon^0(e) = o_P((na_n)^{-1/2}).$$

Proof Similarly as in the proof of Lemma 2, we will prove the lemma by showing that the four terms of (22) are of the stated order. First, we treat the first term of (22).

It can be written as

$$n^{-1}\sum_{i=1}^{n} I(\Delta_i=1)\Big\{E_i^0 J(F_\varepsilon^0(E_i^0))[I(\hat{E}_i^0\le T)-I(\hat{E}_i^0\le \hat{T}_i)]$$
$$+\,E_i^0[J(\hat{F}_\varepsilon^0(\hat{E}_i^0))-J(F_\varepsilon^0(E_i^0))][I(\hat{T}_i<\hat{E}_i^0\le T)-I(T<\hat{E}_i^0\le \hat{T}_i)]$$
$$+\,(\hat{E}_i^0-E_i^0)J(\hat{F}_\varepsilon^0(\hat{E}_i^0))[I(\hat{T}_i<\hat{E}_i^0\le T)-I(T<\hat{E}_i^0\le \hat{T}_i)]\Big\}. \tag{25}$$

Note that $|\hat{E}_i^0-E_i^0|I(\hat{T}_i<\hat{E}_i^0\le T)=O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ uniformly in i by Proposition 4.5 in VKA. When $\hat{E}_i^0\le T$ it holds that $E_i^0\le T\hat{\sigma}^0(X_i)/\sigma^0(X_i)+[\hat{m}^0(X_i)-m^0(X_i)]/\sigma^0(X_i)\le T+V$, where $V=[\inf_x\sigma^0(x)]^{-1}[\sup_x|\hat{m}^0(x)-m^0(x)|+\sup_x|\hat{\sigma}^0(x)-\sigma^0(x)|]=O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ and hence the third term of (25) is bounded by

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

where $\tilde{H}_\varepsilon^0(\cdot)$ is the empirical distribution of $E_i^0, i=1,\ldots,n$. Using the fact that $\tilde{H}_\varepsilon^0(y)-H_\varepsilon^0(y)=O_P(n^{-1/2})$ uniformly in y, the above term is $o_P(n^{-1/2})$. The second term of (25) and the second and third terms of (22) are treated similarly. In the same way, the last term of (22) becomes

$$n^{-1}\sum_{i=1}^{n} I(\Delta_i=0)\frac{\int_{\hat{T}_i}^{T} eJ(\hat{F}_\varepsilon^0(e))\mathrm{d}\hat{F}_\varepsilon^0(e)}{1-\hat{F}_\varepsilon^0(\hat{E}_i^{0T})}I(E_i^0\le T)+o_P((na_n)^{-1/2}). \tag{26}$$

Next, using classical arguments, (26) is written

$$n^{-1}\sum_{i=1}^{n} I(\Delta_i=0)\frac{\int_{\hat{T}_i}^{T} eJ(F_\varepsilon^0(e))\mathrm{d}F_\varepsilon^0(e)}{1-F_\varepsilon^0(E_i^{0T})}I(E_i^0\le T)+O_P(n^{-1/2})$$
$$=-(n^2a_n)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} I(E_i^0\le T)B_{01}(X_i,Z_i,\Delta_i,Z_j,\Delta_j)K\left(\frac{X_i-X_j}{a_n}\right)$$
$$+\,O_P(n^{-1/2}),$$

where

$$B_{01}(X_i,Z_i,\Delta_i,Z_j,\Delta_j)=I(\Delta_i=0)f_X^{-1}(X_i)\frac{TJ(F_\varepsilon^0(T))f_\varepsilon^0(T)}{1-F_\varepsilon^0(E_i^{0T})}$$
$$\times\,[\eta(Z_j,\Delta_j|X_i)+T\zeta(Z_j,\Delta_j|X_i)].$$

Treating the function B_{01} in a similar way as the function B_0 in Lemma 1, we find that the above expression equals

$$-n^{-1}\sum_{i=1}^{n} f_X(X_i)\int_{-\infty}^{T_{X_i}} B_{01}(X_i, z, 0, Z_i, \Delta_i)\,\mathrm{d}H_0(z|X_i) + O(a_n^2) = O_P(n^{-1/2}),$$

since it is a sum of i.i.d. random variables with zero mean.

Finally, together with (23) and (24), the first term of (25) becomes using a Taylor development and Propositions 4.8 and 4.9 in VKA,

$$\begin{aligned}&\int_{R_X} TJ(F_\varepsilon^0(T))h_{\varepsilon 1}^0(T|x)\left\{T\frac{\hat{\sigma}^0(x)-\sigma^0(x)}{\sigma^0(x)}+\frac{\hat{m}^0(x)-m^0(x)}{\sigma^0(x)}\right\}\\&\quad\times f_X(x)dx + o_P(n^{-1/2})\\&\quad= (na_n)^{-1}\sum_{j=1}^{n}\int_{R_X} W(Z_j, \Delta_j|x)K\left(\frac{x-X_j}{a_n}\right)\mathrm{d}x + o_P(n^{-1/2}),\end{aligned}\tag{27}$$

where $W(Z_j, \Delta_j|x) = -TJ(F_\varepsilon^0(T))h_{\varepsilon 1}^0(T|x)\{T\zeta(Z_j, \Delta_j|x)+\eta(Z_j, \Delta_j|x)\}$. Using three Taylor developments of order two for $\zeta(Z_j, \Delta_j|x)$, $\eta(Z_j, \Delta_j|x)$ and $h_{e1}(T|x)$ around X_j, we obtain using condition $(A4)$, that (27) equals

$$n^{-1}\sum_{j=1}^{n} W(Z_j, \Delta_j|X_j) + o_P(n^{-1/2}),\tag{28}$$

which is a sum of i.i.d. random variables with zero mean and hence it is $O_P(n^{-1/2})$. This finishes the proof.

We are now ready to prove the main results of the paper. □

Proof of Theorem 1 Write for any $x \in R_X$,

$$\begin{aligned}&\hat{m}^T(x) - m^T(x)\\&= a_0\hat{m}^0(x)\left\{\int_{-\infty}^{T} J(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e) - \int_{-\infty}^{T} J(F_\varepsilon^0(e))\mathrm{d}F_\varepsilon^0(e)\right\}\\&\quad+\{\hat{m}^0(x)-m^0(x)\}\left\{a_0\int_{-\infty}^{T} J(F_\varepsilon^0(e))dF_\varepsilon^0(e)+\sum_{j=1}^{k} a_j\right\}\\&\quad+a_0\hat{\sigma}^0(x)\left\{\int_{-\infty}^{T} eJ(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e) - \int_{-\infty}^{T} eJ(F_\varepsilon^0(e))\mathrm{d}F_\varepsilon^0(e)\right\}\\&\quad+\{\hat{\sigma}^0(x)-\sigma^0(x)\}\left\{a_0\int_{-\infty}^{T} eJ(F_\varepsilon^0(e))\mathrm{d}F_\varepsilon^0(e)+\sum_{j=1}^{k} a_j((F_\varepsilon^0)^{-1}(s_j)\wedge T)\right\}\end{aligned}$$

$$
\begin{aligned}
&+\hat{\sigma}^0(x)\left[\sum_{j=1}^k a_j\left\{(\hat{F}_\varepsilon^0)^{-1}(s_j)\wedge T-(F_\varepsilon^0)^{-1}(s_j)\right\} I(s_j\le \hat{F}_\varepsilon^0(T), s_j\le F_\varepsilon^0(T))\right.\\
&+\sum_{j=1}^k a_j\left\{T-(F_\varepsilon^0)^{-1}(s_j)\right\} I(\hat{F}_\varepsilon^0(T)<s_j\le F_\varepsilon^0(T))\\
&\left.+\sum_{j=1}^k a_j\left\{(\hat{F}_\varepsilon^0)^{-1}(s_j)\wedge T-T\right\} I(F_\varepsilon^0(T)<s_j\le \hat{F}_\varepsilon^0(T))\right]\\
=&\sum_{\ell=1}^7 A_\ell(x).
\end{aligned}
$$

Since $E|\varepsilon^0|<\infty$, $\sup_x|A_2(x)|$ and $\sup_x|A_4(x)|$ are $O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ using Proposition 4.5 in VKA. From Corollary 3.2 in VKA and Theorem 1 in Doss and Gill (1992) we obtain that $\sup_{s\in Q}|(\hat{F}_\varepsilon^0)^{-1}(s)-(F_\varepsilon^0)^{-1}(s)|=O_P(n^{-1/2})$ and hence $\sup_x|A_5(x)|=O_P(n^{-1/2})$. For $\sup_x|A_3(x)|$, we use Lemma 2. In a similar way, it can be shown that $\sup_x|A_1(x)|$ is of negligible order. Finally, $A_6(x)$ and $A_7(x)$ are uniformly negligible using Corollary 3.2 in VKA. □

Proof of Theorem 2 We use the same decomposition of $\hat{m}^T(x)-m^T(x)$ as in the proof of Theorem 1. Using Propositions 4.8, 4.9 in VKA and the fact that $E|\varepsilon^0|<\infty$, we obtain that

$$
\begin{aligned}
A_2(x)=&-\left[a_0\int_0^{F_\varepsilon^0(T)} J(s)\mathrm{d}s+\sum_{j=1}^k a_j\right](na_n)^{-1}f_X^{-1}(x)\sigma^0(x)\\
&\times\sum_{i=1}^n K\left(\frac{x-X_i}{a_n}\right)\eta(Z_i,\Delta_i|x)+R_n(x),
\end{aligned}
$$

and

$$
\begin{aligned}
A_4(x)=&-(na_n)^{-1}f_X^{-1}(x)\sigma^0(x)\left\{a_0\int_0^{F_\varepsilon^0(T)}(F_\varepsilon^0)^{-1}(s)J(s)\mathrm{d}s\right.\\
&\left.+\sum_{j=1}^k a_j((F_\varepsilon^0)^{-1}(s_j)\wedge T)\right\}\sum_{i=1}^n K\left(\frac{x-X_i}{a_n}\right)\zeta(Z_i,\Delta_i|x)+R_n(x),
\end{aligned}
$$

where $R_n(x)=O_P((na_n)^{-3/4}(\log n)^{3/4})$. For $A_3(x)$ (and similarly for $A_1(x)$) we use Lemma 3. The remaining terms $A_5(x)$, $A_6(x)$ and $A_7(x)$ are $o_P((na_n)^{-1/2})$, as shown in the proof of Theorem 1. Therefore,

$$
\hat{m}^T(x)-m^T(x)=(na_n)^{-1}\sum_{i=1}^n K\left(\frac{x-X_i}{a_n}\right)B(Z_i,\Delta_i|x)+o_P((na_n)^{-1/2}).
$$

□

Proof of Theorem 3 The result follows immediately from Theorem 2 and the central limit theorem for triangular arrays (see e.g., Serfling 1980). □

References

Akritas, M. G. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *Annals of Statistics*, *22*, 1299–1327.

Akritas, M. G. (1996). On the use of nonparametric regression techniques for fitting parametric regression models. *Biometrics*, *52*, 1341–1362.

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

Buchinski, M., Hahn, J. (1998). An alternative estimator for censored quantile regression. *Econometrica*, *66*, 653–671.

Buckley, J., James, I. (1979). Linear regression with censored data. *Biometrika*, *66*, 429–436.

Cai, Z., Hong, Y. (2003). Weighted local linear approach to censored nonparametric regression. In M. G. Akritas, D. M. Politis (Eds.), *Recent advances and trends in nonparametric statistics* (pp. 217–231).

Chen, S., Dahl, G., Kahn, S. (2005). Nonparametric identification and estimation of a censored location-scale regression model. *Journal of the American Statistical Association*, *100*, 212–221.

Dabrowska, D. M. (1987). Nonparametric regression with censored survival time data. *Scandinavian Journal of Statistics*, *14*, 181–197.

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

Doksum, K., Yandell, B. S. (1982). Properties of regression estimates based on censored survival data. In P. J. Bickel, K. A. Doksum, J. L. Hodges (Eds.), *Festschrift for E. L. Lehmann* (pp. 140–156). Wadsworth.

Doss, H., Gill, D. (1992). An elementary approach to weak convergence for quantile processes, with applications to censored survival data. *Journal of the American Statistical Association*, *87*, 869–877.

Einmahl, J., Van Keilegom, I. (2007). Tests for independence in nonparametric regression. *Statistica Sinica* (to appear).

Fan, J., Gijbels, I. (1994). Censored regression : local linear approximations and their applications. *Journal of the American Statistical Association*, *89*, 560–570.

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

Huber, P. J. (1981). *Robust statistics*. New York: Wiley.

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

Kim, H. T., Truong, Y. K. (1998). Nonparametric regression estimates with censored data: local linear smoothers and their applications. *Biometrics*, *54*, 1434–1444.

Li, G., Datta, S. (2001). A bootstrap approach to non-parametric regression for right censored data. *Annals of the Institute of Statistical Mathematics*, *53*, 708–729.

Neumeyer, N., Dette, H., Nagel, E.-R. (2006). Bootstrap tests for the error distribution in linear and nonparametric regression models. *Australian and New Zealand Journal of Statistics*, *48*, 129–156.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, *98*, 1001–1012.

Powell, J. (1986). Censored regression quantiles. *Journal of Econometrics*, *32*, 143–155.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.

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

Van Keilegom, I., Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Annals of the Institute of Statistical Mathematics*, *49*, 467–491.

Van Keilegom, I., Veraverbeke, N. (1998). Bootstrapping quantiles in a fixed design regression model with censored data. *Journal of Statistical Planning and Inference*, *69*, 115–131.

Vignali, C., Brandt, W. N., Schneider, D. P. (2003). X-ray emission from radio-quiet quasars in the SDSS early data release. The α_{OX} dependence upon luminosity. *The Astronomical Journal*, *125*, 433–443.