T E C H N I C A L

R E P O R T

0537

MEAN PRESERVATION IN NONPARAMETRIC REGRESSION WITH CENSORED DATA

HEUCHENNE C., and I. VAN KEILEGOM

I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Mean Preservation in Nonparametric Regression with Censored Data

Cédric Heuchenne[1]

Institut de Statistique

Université catholique de Louvain

Ingrid Van Keilegom[1]

Institut de Statistique

Université catholique de Louvain

August 17, 2005

Abstract

Consider the heteroscedastic model $Y = m(X) + \sigma(X)\varepsilon$, where ε and X are independent, Y is subject to right censoring, $m(\cdot)$ is an unknown but smooth location function (like e.g. conditional mean, median, trimmed mean...) and $\sigma(\cdot)$ an unknown but smooth scale function. In this paper we consider the estimation of $m(\cdot)$ under this model. The estimator we propose is a Nadaraya-Watson type estimator, for which the censored observations are replaced by 'synthetic' data points estimated under the above model. The estimator offers an alternative for the completely nonparametric estimator of $m(\cdot)$, which cannot be estimated consistently in a completely nonparametric way, whenever high quantiles of the conditional distribution of Y given $X = x$ are involved.

We obtain the asymptotic properties of the proposed estimator of $m(x)$ and study its finite sample behavior in a simulation study. The method is also applied to a study of quasars in astronomy.

KEY WORDS: Bandwidth; Censored regression; Kernel estimation; Location-scale model; Nonparametric regression; Survival analysis.

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

1 Introduction

Let (X, Y) be a random vector, 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). We assume that the relation between X and Y is given by

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1.1}$$

where $m(X)$ and $\sigma(X)$ are some unknown but smooth location and scale functions and the error term ε is independent of X. So, we assume that the conditional distribution of Y given X depends on X only via its first and second conditional moment.

In this paper we study the estimation of the function $m(\cdot)$ under model (1.1). We do not restrict this function to be the conditional mean, but allow it to be any L-functional (see e.g. Serfling, 1980, p. 265) :

$$m(x) = a_0 \int_0^1 F^{-1}(s|x)L(s)\,ds + \sum_{j=1}^{k} a_j F^{-1}(s_j|x), \tag{1.2}$$

where $F^{-1}(s|x) = \inf\{y : F(y|x) \geq s\}$ is the quantile function of Y given x, $L(s)$ is a given weight function satisfying $\int_0^1 L(s)ds = 1$, $L(s) \geq 0$ for all $0 \leq s \leq 1$, $k \geq 0$, $a_0, \ldots, a_k$ 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 $L \equiv 1$, $a_0 = 1$ 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.

It is well known that the conditional mean $E(Y|X)$ (and any other location function that involves high quantiles of $F(\cdot|x)$) cannot be consistently estimated in a completely nonparametric way, due to the presence of right censoring. The estimator we propose below attempts to solve this problem, by making use of model (1.1). In fact, when ε is independent of X, the right tail of the distribution $F(\cdot|x)$ can be estimated well provided there is a region in the support of the covariate where censoring is 'light' (this is because we can estimate this right tail from the right tail of the error distribution, which is a global distribution, and hence it can be better estimated than the local distribution $F(\cdot|x)$). In this way we are able to estimate relatively well the right tail of $F(\cdot|x)$ for any x, also for those that belong to regions where censoring is heavy.

The method we propose consists in first consistently estimating the conditional distribution $F(y|x)$ under model (1.1), and second to plug-in the obtained estimator in (1.2). To estimate $F(\cdot|x)$, we replace the censored observations by new 'synthetic' data points that are obtained under model (1.1), and we then estimate the distribution $F(\cdot|x)$ by using a weighted empirical distribution function on the new data points. The method uses model (1.1) only in the construction of synthetic data points, and does not use the model in the construction of the estimator itself. So, in a sense, it is little sensitive to the validity of model (1.1), and it can be expected that the estimator works well, even in situations where model (1.1) does not hold.

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 (1997, 1998), Chen, Dahl and Kahn (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), Heuchenne and Van Keilegom (2004) 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 procedures in detail. In Section 3 we state the asymptotic properties of the estimator of $m(\cdot)$ obtained in Section 2. As a byproduct we also obtain the asymptotic properties of the estimator of $F(\cdot|x)$. Section 4 contains a simulation study, in which the new estimator is compared with the completely nonparametric estimator, and with an estimator proposed in Heuchenne and Van Keilegom (2005). In Section 5, a data set on spectral energy distributions of quasars is analyzed by means of the three methods. Finally, the Appendix contains the proofs of the asymptotic results of Section 3.

2 Description of the method

We start with some notations and definitions. Let $m^0(\cdot)$ be any location function and $\sigma^0(\cdot)$ be any scale function, meaning that $m^0(x) = T(F(\cdot|x))$ and $\sigma^0(x) = S(F(\cdot|x))$ for some functionals T and S that satisfy $T(F_{aY+b}(\cdot|x)) = aT(F_Y(\cdot|x)) + b$ and $S(F_{aY+b}(\cdot|x)) = aS(F_Y(\cdot|x))$, for all $a \geq 0$ and $b \in I\!R$ (here $F_{aY+b}(\cdot|x)$ denotes the conditional distribution of $aY + b$ given $X = x$). Then, it can be easily seen that if model (1.1) holds, the model $Y = m^0(X) + \sigma^0(X)\varepsilon^0$ with ε^0 independent of X, is also valid. Define

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

where $J(s)$ is a given weight function satisfying $\int_0^1 J(s)\,ds = 1$ and $J(s) \geq 0$ for all $0 \leq s \leq 1$. We will choose J in such a way that $m^0(x)$ and $\sigma^0(x)$ can be estimated in a consistent way (i.e. choose J in such a way that the right tail of $F(\cdot|x)$ does not need to be estimated) and we will then use these estimators of $m^0(x)$ and $\sigma^0(x)$ in the construction of an estimator of $m(x)$.

Before explaining the estimator, let us introduce some notations. Define $F(y|x) = P(Y \leq y|x)$, $G(y|x) = P(C \leq y|x)$, $H(y|x) = P(Z \leq y|x)$, $H_\delta(y|x) = P(Z \leq y, \Delta = \delta|x)$, and $F_X(x) = P(X \leq x)$. 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 (2.1). Next, for $E = (Z - m(X))/\sigma(X)$ define $H_\varepsilon(y) = P(E \leq y)$, $H_{\varepsilon\delta}(y) = P(E \leq y, \Delta = \delta)$, $H_\varepsilon(y|x) = P(E \leq y|x)$ and $H_{\varepsilon\delta}(y|x) = P(E \leq y, \Delta = \delta|x)$ $(\delta = 0, 1)$. Define analogous functions $H_\varepsilon^0(y)$, $H_{\varepsilon\delta}^0(y)$, $H_\varepsilon^0(y|x)$ and $H_{\varepsilon\delta}^0(y|x)$ for $E^0 = (Z - m^0(X))/\sigma^0(X)$ and $G_\varepsilon^0(y) = P(C^0 \leq y)$ for $C^0 = (C - m^0(X))/\sigma^0(X)$. 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.

The idea of the proposed method is first to estimate the true unknown survival time of censored observations by making use of model (1.1), and then to estimate $m(x)$ by using a kernel type estimator based on these new data. Replacing censored observations by 'synthetic' (or estimated) survival times, has been widely used in parametric regression with censored data. See e.g. Buckley and James (1979), Koul, Susarla and Van Ryzin (1981), Leurgans (1987) and Heuchenne and Van Keilegom (2004).

The extension of this idea to nonparametric estimation of any L-functional of the type (1.2) is as follows. First, note that $m(x)$ can be written as

$$m(x) = a_0 E[YL(F(Y|x))|x] + \sum_{j=1}^{k} a_j F^{-1}(s_j|x),$$

and that $F(y|x) = E[I(Y \leq y)|x]$. Let $\phi_1(y|x) = yL(F(y|x))$ and $\phi_{2t}(y|x) = \phi_2(y|x) = I(y \leq t)$ for fixed t. The idea is now to replace $E[\phi_j(Y|x)|x]$ $(j = 1, 2)$ by a kernel estimator of the type $\sum_{i=1}^n W_i(x, a_n)\hat{\phi}_j^*(Z_i, \Delta_i|x)$, where $W_i(x, a_n)$ are local weights defined below, and $\hat{\phi}_j^*(Z_i, \Delta_i|x)$ $(i = 1, \ldots, n)$ are estimators of $\phi_j^*(Z_i, \Delta_i|x)$, which are chosen in such a way that $E[\phi_j^*(Z_i, \Delta_i|x)|x] = E[\phi_j(Y_i|x)|x]$. It is easy to check that this preservation of means is obtained for

$$\begin{aligned}\phi_j^*(z, \delta|x) &= \phi_j(z|x)\delta + E[\phi_j(Y|x)|Y > z, x](1 - \delta) \\ &= \phi_j(z|x)\delta + \frac{1}{1 - F(z|x)} \int_z^{+\infty} \phi_j(y|x)dF(y|x)(1 - \delta) \qquad (2.2)\end{aligned}$$

(see also Fan and Gijbels (1994), where a similar idea has been used in a completely nonparametric context).

To estimate the function $\phi_j^*(z,\delta|x)$, we need an estimator of $F(\cdot|x)$. Note that

$$F(y|x) = F_\varepsilon^0\Big(\frac{y-m^0(x)}{\sigma^0(x)}\Big),$$

and hence we need to estimate F_ε^0, m^0 and σ^0. The functions m^0 and σ^0 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) :

$$\tilde{F}(y|x) = 1 - \prod_{Z_i \le y, \Delta_i = 1} \Big\{1 - \frac{W_i(x,a_n)}{\sum_{j=1}^n I(Z_j \ge Z_i) W_j(x,a_n)}\Big\}, \tag{2.3}$$

where

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

$(i=1,\dots,n)$ 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-Meier (1958) estimator when all weights $W_i(x,a_n)$ equal n^{-1}. This yields

$$\hat{m}^0(x) = \int_0^1 \tilde{F}^{-1}(s|x)J(s)\,ds, \quad \hat{\sigma}^{02}(x) = \int_0^1 \tilde{F}^{-1}(s|x)^2 J(s)\,ds - \hat{m}^{02}(x) \tag{2.4}$$

as estimators for $m^0(x)$ and $\sigma^{02}(x)$. In practice, the score function J will be chosen in such a way that $\tilde{F}(\cdot|x)$ is consistent on the support of J. Next, estimate the residual distribution F_ε^0 (suppose no ties) :

$$\hat{F}_\varepsilon^0(y) = 1 - \prod_{\hat{E}^0_{(i)} \le y, \Delta_{(i)}=1} \Big(1 - \frac{1}{n-i+1}\Big) \;, \tag{2.5}$$

where $\hat{E}_i^0 = (Z_i - \hat{m}^0(X_i))/\hat{\sigma}^0(X_i)$, $\hat{E}^0_{(i)}$ is the i-th order statistic of $\hat{E}_1^0,\dots,\hat{E}_n^0$ and $\Delta_{(i)}$ is the corresponding censoring indicator. This estimator has been studied in detail by Van Keilegom and Akritas (1999). Next, define

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

Now, let $\hat{\phi}_1(y|x) = yL(\hat{F}(y\wedge T_x|x))$ and $\hat{\phi}_{2t}(y|x) = \hat{\phi}_2(y|x) = I(y \le t)$, and let

$$\hat{\phi}_j^*(z,\delta|x) = \hat{\phi}_j(z|x)\delta + \frac{1}{1-\hat{F}(z\wedge T_x|x)} \int_{z\wedge T_x}^{T_x} \hat{\phi}_j(y|x)d\hat{F}(y|x)(1-\delta), \tag{2.6}$$

where $T_x = T\sigma^0(x) + m^0(x)$, $T < \tau_{H_\varepsilon^0(\cdot)}$ and $\tau_{F(\cdot)} = \inf\{y : F(y) = 1\}$ for any distribution F. Note that we have to truncate the integral at T_x in the above definition. However, when $\tau_{F^0_\varepsilon(\cdot)} \le \tau_{G^0_\varepsilon(\cdot)}$, T can be chosen arbitrarily close to $\tau_{F^0_\varepsilon(\cdot)}$. The estimator of $m(x)$ is now defined by

$$
\begin{aligned}
&\hat{m}^T(x) \\
&= a_0 \sum_{i=1}^n W_i(x, a_n) \hat{\phi}_1^*(Z_i, \Delta_i|x) + \sum_{j=1}^k a_j [\hat{F}_{\phi_2}^{-1}(s_j|x) \wedge T_x] \\
&= a_0 \sum_{i=1}^n W_i(x, a_n) \Big[Y_i L(\hat{F}(Y_i \wedge T_x|x)) \Delta_i \qquad (2.7) \\
&\quad + \frac{1}{1 - \hat{F}(C_i \wedge T_x|x)} \int_{C_i \wedge T_x}^{T_x} y L(\hat{F}(y|x)) d\hat{F}(y|x)(1 - \Delta_i) \Big] + \sum_{j=1}^k a_j [\hat{F}_{\phi_2}^{-1}(s_j|x) \wedge T_x],
\end{aligned}
$$

where

$$
\begin{aligned}
&\hat{F}_{\phi_2}(t|x) \\
&= \sum_{i=1}^n W_i(x, a_n) \hat{\phi}_{2t}^*(Z_i, \Delta_i|x) \qquad (2.8) \\
&= \sum_{i=1}^n W_i(x, a_n) \left[I(Y_i \le t)\Delta_i + \frac{1}{1 - \hat{F}(C_i \wedge T_x|x)} \int_{C_i \wedge T_x}^{T_x} I(y \le t) d\hat{F}(y|x)(1 - \Delta_i) \right].
\end{aligned}
$$

Note that $\hat{m}^T(x)$ is actually estimating

$$m^T(x) = a_0 E[\tilde{\phi}_1^*(Z, \Delta|x)|x] + \sum_{j=1}^k a_j [F_{\phi_2}^{-1}(s_j|x) \wedge T_x],$$

where

$$\tilde{\phi}_j^*(z, \delta|x) = \tilde{\phi}_j(z|x)\delta + \frac{1}{1 - F(z \wedge T_x|x)} \int_{z \wedge T_x}^{T_x} \tilde{\phi}_j(y|x) dF(y|x)(1 - \delta),$$

$\tilde{\phi}_1(y|x) = yL(F(y \wedge T_x|x))$, $\tilde{\phi}_{2t}(y|x) = \phi_{2t}(y|x)$ and $F_{\phi_2}(t|x) = E[\tilde{\phi}_{2t}^*(Z, \Delta|x)|x]$. As before, $m^T(x)$ and $F_{\phi_2}(t|x)$ can be made arbitrarily close to $m(x)$ and $F(t|x)$ respectively, provided $\tau_{F^0_\varepsilon(\cdot)} \le \tau_{G^0_\varepsilon(\cdot)}$.

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

$$\tilde{m}^T(x) = a_0 \int_{-\infty}^{\tilde{T}_x} y L(\tilde{F}(y|x))\, d\tilde{F}(y|x) + \sum_{j=1}^k a_j [\tilde{F}^{-1}(s_j|x) \wedge \tilde{T}_x], \qquad (2.9)$$

where $\tilde{T}_x < \tau_{H(\cdot|x)}$. Note that we truncate at $\tilde{T}_x$, because of the 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 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 $\tilde{m}^T(x)$, and the difference between the two truncation points can be substantial, especially when the censoring proportion is not uniform over x. Indeed, 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 up to 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.

Finally, note that in Heuchenne and Van Keilegom (2005) an alternative estimator of $m(x)$ has been studied, which also makes use of model (1.1). The estimator is defined by

$$\hat{m}^T_{\text{alt}}(x) = a_0 \int_{-\infty}^{\hat{T}_x} yL(\hat{F}(y|x))\, d\hat{F}(y|x) + \sum_{j=1}^{k} a_j [\hat{F}^{-1}(s_j|x) \wedge \hat{T}_x], \tag{2.10}$$

where $\hat{T}_x = T\hat{\sigma}^0(x) + \hat{m}^0(x)$ and $T < \tau_{H^0_\varepsilon(\cdot)}$. We will compare the here proposed estimator $\hat{m}^T(x)$ with the estimators $\tilde{m}^T(x)$ and $\hat{m}^T_{\text{alt}}(x)$ in a simulation study (see Section 4).

3 Asymptotic results

We first give some asymptotic results for the estimator $\hat{m}^T(x)$ proposed in Section 2. We then state, as a by-product, some asymptotic results for the estimator $\hat{F}_{\phi_2}(t|x)$ defined in (2.8). The proofs of the results below, as well as the assumptions under which they are valid, can be found in the Appendix.

3.1 Results for $\hat{m}^T(x)$

Theorem 3.1 *Assume (A1)–(A5), (A6) (i), (A7), L is continuously differentiable, $\int_0^1 L(s)ds = 1$ and $L(s) \geq 0$ 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 3.2 *Assume (A1)–(A7). Then, for any $x \in R_X$,*

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

where $\sup\{|R_n(x)|; x \in R_X\} = o_P((na_n)^{-1/2})$ and the function $B(z, \delta|x)$ is given in the Appendix.

Theorem 3.3 *Under the assumptions of Theorem 3.2, for any $x \in R_X$,*

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

where

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

Proof. The result is obtained by using Lyapounov's Theorem. It's easy to check that the Lyapounov ratio is $O((na_n)^{-1/2})$ since $E|Z|^\lambda < \infty$ (λ is given in assumption (A3) (iii) in the Appendix).

3.2 Results for $\hat{F}_{\phi_2}(t|x)$

Theorem 3.4 *Assume (A1), (A2), (A3) (i), (ii),(A4), (A5) and (A7). Then,*

$$\sup_{x\in R_X} \sup_{-\infty<t<\infty} |\hat{F}_{\phi_2}(t|x) - F_{\phi_2}(t|x)| = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).$$

Theorem 3.5 *Assume (A1), (A2), (A3) (i), (ii), (A4), (A5) and (A7). Then, for any* $x \in R_X$ *and* $-\infty < t < \infty$*,*

$$\hat{F}_{\phi_2}(t|x) - F_{\phi_2}(t|x) = \frac{1}{na_n}\sum_{i=1}^{n} K\Big(\frac{x-X_i}{a_n}\Big) A(t, Z_i, \Delta_i|x) + R_n(t|x),$$

where $\sup\{|R_n(t|x)|; x \in R_X\} = o_P((na_n)^{-1/2})$ *and the function* $A(t,z,\delta|x)$ *is given in the Appendix.*

Theorem 3.6 *Under the assumptions of Theorem 3.5, for any* $x \in R_X$ *and* $-\infty < t < \infty$*,*

$$(na_n)^{1/2}(\hat{F}_{\phi_2}(t|x) - F_{\phi_2}(t|x)) \xrightarrow{d} N(0, s^2(t|x)),$$

where

$$s^2(t|x) = f_X(x) \int K^2(u)\,du \sum_{\delta=0,1} \int A^2(t,z,\delta|x)\,dH_\delta(z|x).$$

Proof. The result is obtained by using Lyapounov's Theorem.

Theorem 3.7 *Assume (A1), (A2), (A3) (i)-(ii), (A4), (A5) and (A7). Then,*

$$\sup_{x\in R_X,\ |t-s|\le d_n} |\hat{F}_{\phi_2}(t|x) - F_{\phi_2}(t|x) - \hat{F}_{\phi_2}(s|x) + F_{\phi_2}(s|x)| = o_P((na_n)^{-1/2}),$$

where $d_n \sim (na_n)^{-1/2}(\log a_n^{-1})^{1/2}$.

Remark 3.8 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 $\Delta_i^* = I(Y_i^* \le C_i^*)$. 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

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

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 3.9 A similar idea as the one developed above to estimate $m(x)$, can be used to better estimate any scale function $\sigma(x)$. We propose

$$\hat{\sigma}^{T2}(x) = a_0^2 \left\{ \sum_{i=1}^{n} W_i(x, a_n) \hat{\phi}_3^*(Z_i, \Delta_i|x) - \hat{m}^{T2}(x) \right\}$$
$$+ \sum_{j=1}^{k} a_j^2 \left\{ \sum_{i=1}^{n} W_i(x, a_n) \hat{\phi}_4^{j*}(Z_i, \Delta_i|x) \right\}^2,$$

where $\hat{\phi}_3(y|x) = y^2 L(\hat{F}(y \wedge T_x|x))$, $\hat{\phi}_4^j(y|x) = \rho_j(y - \hat{F}_{\phi_2}^{-1}(s_j|x) \wedge T_x)$, $\hat{\phi}_4^{j*}(z, \delta|x)$ is defined in the same way as (2.6) and $\rho_j(u) = s_j u I(u \ge 0) + (s_j - 1) u I(u < 0)$. The asymptotic results for $\hat{\sigma}^{T2}(x)$ can be obtained along the same lines as for the estimator $\hat{m}^T(x)$.

4 Simulations

In this section, we compare the finite sample behavior of the estimators $\hat{m}^T(x)$, $\tilde{m}^T(x)$ and $\hat{m}_{\text{alt}}^T(x)$. We are interested in the behavior of the integrated mean squared error, defined by $IMSE = \int E[(\hat{m}(x) - m(x))^2]\, dx$ for any estimator of $m(x)$. The simulations are carried out for samples of size $n = 100$ and the results are obtained by using 250 simulations. We compare the three methods for four different locations : the conditional mean, the conditional truncated mean ($L(s) = (1/0.9)I(0.05 < s \le 0.95)$), the conditional median and the conditional third quartile.

For the weights that appear in the Beran estimator $\tilde{F}(y|x)$, we choose a biquadratic kernel function $K(x) = (15/16)(1 - x^2)^2 I(|x| \le 1)$. The bandwidth sequence a_n is selected for each estimator as the minimizer of an approximated $IMSE$ among a grid of 20 possible values of a_n. The weight function $J(s)$ equals $J(s) = I(s \leq b)/b$, where $b = \min_{1\leq i\leq n} \tilde{F}(+\infty|X_i)$. The point $(T_x - \hat{m}^0(x))/\hat{\sigma}^0(x)$, respectively $\tilde{T}_x$, is chosen larger than (or equal to) $\hat{E}^0_{(n)}$, respectively $Z_{(n)}$ in order to consider all the jumps of $\hat{F}(y|x)$ and $\tilde{F}(y|x)$.

The first model we consider is

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \sigma\varepsilon, \tag{4.1}$$

for various choices of $\beta_0, \beta_1, \beta_2, \beta_3$ and σ, where X has a uniform distribution on the interval $[0,1]$ or $[0,3]$, 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 α_0, α_1, α_2, and α_3, where ε^* has a normal distribution with zero

β_0	β_1	β_2	β_3	CP	$IMSE$			
α_0	α_1	α_2	α_3	σ^2	mean	trunc. mean	median	3^{rd} quartile
0	1	-1	1	30.1	0.089	0.090	0.098	0.110
4.1	-14	0	19.8	0.5	0.088	0.088	0.093	0.097
					0.084	0.086	0.090	0.098
0	1	0	0	35.3	0.124	0.126	0.141	0.168
3.6	-10.5	0	12	1	0.123	0.124	0.133	0.144
					0.115	0.120	0.126	0.142
0	1	-1	1	50.2	0.093	0.095	0.104	0.147
1.3	-6	4	3.2	0.5	0.091	0.092	0.098	0.110
					0.085	0.087	0.093	0.109
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
					0.322	0.325	0.341	0.377
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
					0.393	0.394	0.412	0.508
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
					0.385	0.389	0.463	0.581

Table 1: *Results for* $\tilde{m}^T(x)$ *(first line),* $\hat{m}^T_{\text{alt}}(x)$ *(second line) and* $\hat{m}^T(x)$ *(third line) for model (4.1) with large optimal bandwidth* a_n. R_X *is* $[0,1]$ $([0,3])$ *for the three first (last) models.*

β_0	β_1	β_2	β_3	CP	$IMSE$			
α_0	α_1	α_2	α_3	σ^2	mean	trunc. mean	median	3^{rd} 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
					1.766	1.759	1.777	1.849
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
					1.305	1.318	1.351	1.438
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
					1.512	1.516	1.596	1.766
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
					1.616	1.623	1.745	1.853

Table 2: *Results for $\tilde{m}^T(x)$ (first line), $\hat{m}^T_{\text{alt}}(x)$ (second line) and $\hat{m}^T(x)$ (third line) for model (4.1) with moderately large optimal bandwidth a_n. R_X is* $[0, 3]$.

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.

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, censoring probability curves that do or do not wiggle a lot,...). The proportion of censoring (in % 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.

First, we compare $\hat{m}^T(x)$ with $\tilde{m}^T(x)$. 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 L, the worse the estimation, but $\hat{m}^T(x)$ resists better than $\tilde{m}^T(x)$. This is due to the locality of the Beran estimator and its inconsistency problems. On the other hand, $\hat{m}^T(x)$ is a global estimator and its inconsistency problems are considerably less important than for the Beran estimator. As a consequence, a more wiggly curve or an increase of the proportion of censoring affects more $\tilde{m}^T(x)$ than $\hat{m}^T(x)$.

We next compare the new estimator $\hat{m}^T(x)$ with its competitor $\hat{m}^T_{\text{alt}}(x)$. The main motivation for $\hat{m}^T(x)$ with respect to $\hat{m}^T_{\text{alt}}(x)$ is as follows. The use of global information sometimes penalizes the estimation procedure since the amount of local information is decreased by reducing the support of the score function J. These instability problems especially arise when the censoring probability curve contains some peaks in certain regions of the covariate space. Although both estimators $\hat{m}^T_{\text{alt}}(x)$ and $\hat{m}^T(x)$ are based on $\hat{m}^0(\cdot)$ and $\hat{\sigma}^0(\cdot)$ (and can thus suffer from small supports of J), $\hat{m}^T(x)$ only uses them in the estimation of censored synthetic data points and not in the construction of the estimator itself. Hence, it preserves uncensored local information and is less sensitive to these instability problems. This explains why $\hat{m}^T(x)$ often outperforms $\hat{m}^T_{\text{alt}}(x)$ for the estimation of the mean and truncated mean. Note that the instability problems also depend on the shape of the curve to estimate and the amount of censoring. Therefore, Table 1 shows approximately the same results for both estimators since the chosen models are relatively flat. In Table 2, $\hat{m}^T(x)$ outperforms $\hat{m}^T_{\text{alt}}(x)$ for large proportions of censoring and more wiggly models, while $\hat{m}^T(x)$ is the best one in Table 3 at all censoring levels. On the other hand, the estimator $\hat{m}^T_{\text{alt}}(x)$ behaves better for quantile estimation. This is because $\hat{m}^T(x)$ is highly based on $\hat{F}_{\phi_2}(\cdot|x)$ which uses less 'homogeneous' information

β_0	β_1	β_2	β_3	CP	*IMSE*			
α_0	α_1	α_2	α_3	σ^2	mean	trunc. mean	median	3^{rd} 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
					1.059	1.067	1.125	1.276
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
					1.025	1.038	1.086	1.209
4	-7.5	6	-1.3	38.3	1.262	1.286	1.371	1.628
3.4	-7.45	7	-1.6	1	1.239	1.248	1.283	1.373
					1.231	1.248	1.313	1.450
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
					1.112	1.118	1.212	1.389
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
					1.283	1.308	1.423	1.665
4	-7.5	6	-1.3	74.7	1.493	1.590	2.412	2.119
3	-7.6	6	-1.3	1	1.512	1.576	2.005	2.176
					1.493	1.544	2.006	2.176

Table 3: *Results for $\tilde{m}^T(x)$ (first line), $\hat{m}^T_{\text{alt}}(x)$ (second line) and $\hat{m}^T(x)$ (third line) for model (4.1) with small optimal bandwidth a_n. R_X is $[0,3]$.*

(in the sense that the true data points are mixed with data estimated by means of the general heteroscedastic model) than the global $\hat{F}^0_\varepsilon(\cdot)$ used by $\hat{m}^T_{\text{alt}}(x)$.

Next, we consider the case where model (1.1) is not satisfied. For this, we generate random response and censoring variables from Weibull distributions with the following parameters

$$\begin{aligned} Y|X = x &\sim \text{Weibull}(x, d_x), \\ C|X = x &\sim \text{Weibull}((0.3 + x)/\xi, d_x), \end{aligned} \tag{4.2}$$

where X has a uniform distribution on the interval $[0, 3]$ and $d_x = 2 + d_1 x \geq 0$ and ξ are chosen such that $(0.3 + x)/\xi \geq 0$ for all $0 \leq x \leq 3$. From the conditional independence between Y and C for given X, it follows that the censoring probability curve is given by $P(\Delta = 0|X = x) = (0.3 + x)/((\xi + 1)x + 0.3)$. For $\varepsilon = (Y - m(X))/\sigma(X)$, with $m(X) = E(Y|X)$ and $\sigma^2(X) = \text{Var}(Y|X)$, we have

$$P(\varepsilon \leq t|x) = 1 - \exp(-\{t[\Gamma(1 + 2d_x^{-1}) - \Gamma^2(1 + d_x^{-1})]^{1/2} + \Gamma(1 + d_x^{-1})\}^{d_x}).$$

d_1	CP	*IMSE*			
ξ		mean	trunc. mean	median	3^{rd} quartile
0	36.58	1.488	1.672	1.507	3.882
1		1.479	0.999	1.412	3.271
		1.314	0.993	1.308	2.909
1	38.27	0.576	1.205	0.794	1.473
0.71		0.566	0.737	0.648	1.076
		0.517	0.605	0.622	1.022
2	38.03	0.362	1.069	0.515	0.769
0.55		0.348	0.754	0.388	0.602
		0.318	0.538	0.390	0.603
3	38.13	0.270	1.006	0.386	0.536
0.42		0.258	0.804	0.282	0.422
		0.241	0.541	0.288	0.448
4	38.28	0.235	0.953	0.319	0.377
0.32		0.218	0.843	0.221	0.346
		0.195	0.574	0.229	0.361

Table 4: *Results for $\tilde{m}^T(x)$ (first line), $\hat{m}^T_{\text{alt}}(x)$ (second line) and $\hat{m}^T(x)$ (third line) for model (4.2).*

This shows that model (1.1) is satisfied only when d_x does not depend on x. Since $d_x = 2 + d_1 x$, model (1.1) is satisfied only when $d_1 = 0$. Also, the larger the value of d_1, the larger the discrepancy between model (1.1) and model (4.2).

Table 4 shows the simulation results for model (4.2) with different values of d_1 and ξ. ξ is chosen such that the censoring probability curve is approximately the same for each d_1. $\hat{m}^T_{\text{alt}}(x)$ and $\hat{m}^T(x)$ do not seem to be very sensitive to model assumption (1.1), since Beran's method obtains the largest $IMSE$ for all values of d_1. When the value of d_1 increases, $\hat{m}^T(x)$ seems to resist better than $\hat{m}^T_{\text{alt}}(x)$ for conditional mean or truncated mean estimation, whereas $\hat{m}^T_{\text{alt}}(x)$ continues to outperform $\hat{m}^T(x)$ in quantile estimation. This can also be explained by the fact that the data set on which $\hat{F}_{\phi_2}(\cdot|x)$ is constructed becomes more and more heterogeneous as d_1 increases.

The final setting we consider is a normal heteroscedastic regression model

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + (\gamma X + 0.1)\varepsilon, \tag{4.3}$$

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 5. Similar conclusions as above hold in this heteroscedastic case.

β_0	β_1	β_2	β_3	CP	$IMSE$			
α_0	α_1	α_2	α_3	γ^2	mean	trunc. mean	median	3^{rd} 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
					0.346	0.345	0.358	0.990
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
					0.582	0.563	0.566	0.922
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
					1.010	1.039	1.061	2.196

Table 5: *Results for $\tilde{m}^T(x)$ (first line), $\hat{m}^T_{\text{alt}}(x)$ (second line) and $\hat{m}^T(x)$ (third line) for model (4.3). R_X is $[0, 3]$ for the first model and $[0, 1]$ for the two other ones.*

5 Data analysis

In this section, we illustrate the proposed method on a data set, which comes from a study of quasars in astronomy. The data have been considered by Vignali, Brandt and Schneider (2003) and have been previously analyzed by Heuchenne and Van Keilegom (2005). To date, many studies have focused on the dependence on luminosity and redshift of quasar ultraviolet-to-X-ray spectral energy distributions (characterized by means of the spectral index $\alpha_{ox} = 0.384 \log(L_{2\ keV}/L_{2500\ \mathring{A}})$, where $l_{uv} = \log L_{2500\ \mathring{A}}$ and $l_x = \log L_{2\ keV}$ denote the rest-frame 2500 $\mathring{A}$ and 2 keV luminosity densities). 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 $Z_i = (\max_{j:j=1,\ldots,137}(l_{x,j}) - l_{x,i}),\ i = 1,\ldots,137$.

The choice of the bandwidth is achieved with the bootstrap procedure of Remark 3.8 (adapted to each estimator). The selected bandwidth is approximately the same for the three methods (around 0.75). The results for the estimation of the conditional mean, truncated mean, median and first quartile are given in Figures 1 to 4. The estimator $\hat{m}^T(x)$ suggests to use linear functions for the four proposed locations. As expected, $\hat{m}^T(x)$ is less smooth than $\hat{m}^T_{\text{alt}}(x)$, especially for the first quartile.

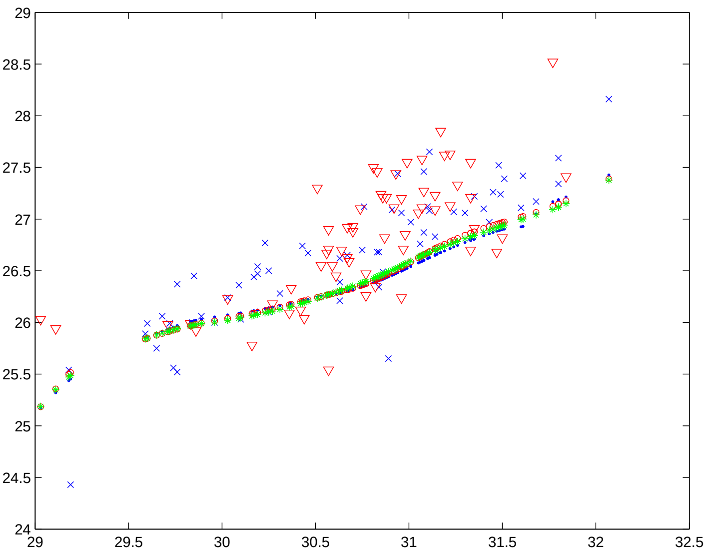

Figure 1: *Estimated conditional mean for the quasars data. The estimators $\tilde{m}^T(x)$, $\hat{m}^T_{\text{alt}}(x)$ and $\hat{m}^T(x)$ are indicated by $\cdot$, $\circ$ and $*$ respectively. Uncensored data points are represented by $\times$, and (left) censored observations by $\bigtriangledown$.*

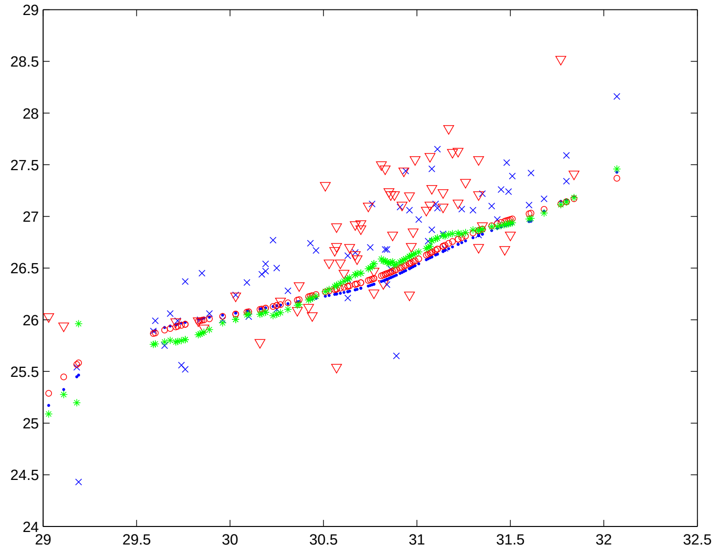

Figure 2: *Estimated conditional truncated mean for the quasars data (5 percent of truncation at both sides). The estimators* $\tilde{m}^T(x)$, $\hat{m}^T_{\text{alt}}(x)$ *and* $\hat{m}^T(x)$ *are indicated by* $\cdot$, $\circ$ *and* $*$ *respectively. Uncensored data points are represented by* $\times$, *and (left) censored observations by* $\bigtriangledown$.

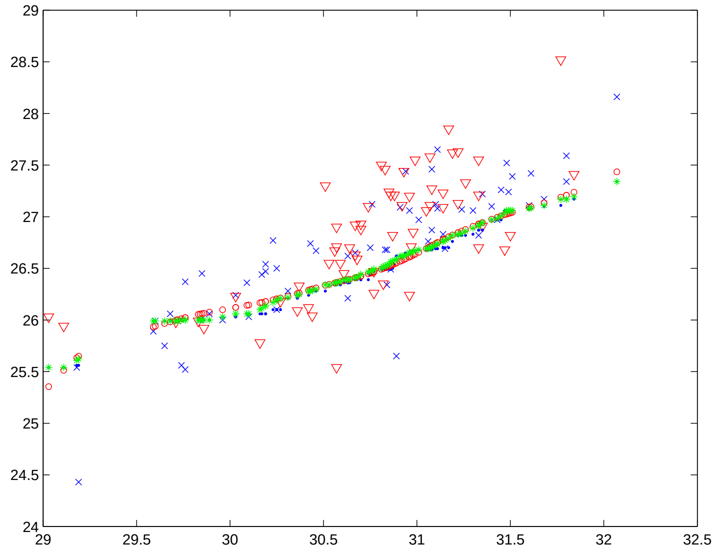

Figure 3: *Estimated conditional median for the quasars data. The estimators* $\tilde{m}^T(x)$, $\hat{m}^T_{\text{alt}}(x)$ *and* $\hat{m}^T(x)$ *are indicated by* $\cdot$, $\circ$ *and* $*$ *respectively. Uncensored data points are represented by* $\times$, *and (left) censored observations by* $\bigtriangledown$.

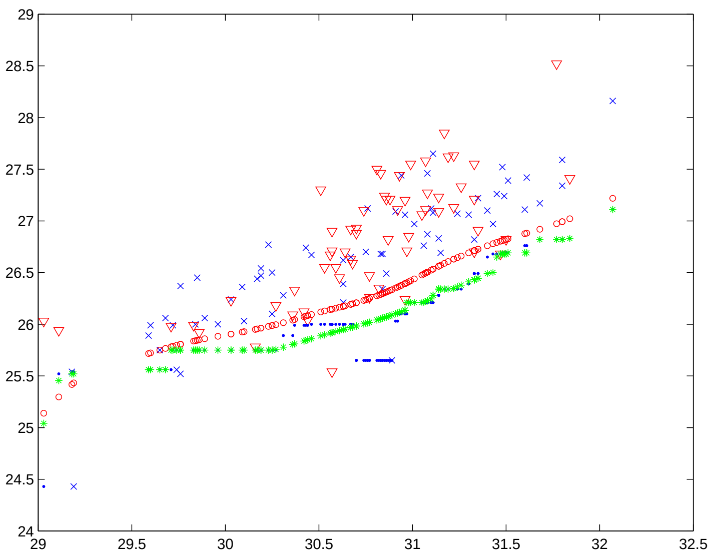

Figure 4: *Estimated conditional first quartile for the quasars data. The estimators* $\tilde{m}^T(x)$, $\hat{m}^T_{\text{alt}}(x)$ *and* $\hat{m}^T(x)$ *are indicated by* $\cdot$, $\circ$ *and* $*$ *respectively. Uncensored data points are represented by* $\times$, *and (left) censored observations by* $\bigtriangledown$.

Appendix : Proofs of main results

The following functions enter the asymptotic representation of $\hat{m}^T(x) - m^T(x)$ which we established in Section 3.

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

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

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

$$h_{x,y}(z,\delta) = \Big[\eta(z,\delta|x) + \zeta(z,\delta|x)\frac{y - m^0(x)}{\sigma^0(x)}\Big] f_\varepsilon^0\Big(\frac{y - m^0(x)}{\sigma^0(x)}\Big) f_X^{-1}(x),$$

$$A(t,z,\delta|x)$$
$$= E\left[I(\Delta = 0)\left\{h_{x,Z_x^T}(z,\delta)\frac{F(T_x \wedge t|x) - F(Z_x^T \wedge t|x)}{(1 - F(Z_x^T|x))^2} + \frac{h_{x,T_x\wedge t}(z,\delta) + h_{x,Z_x^T\wedge t}(z,\delta)}{1 - F(Z_x^T|x)}\right\}\right]$$
$$+ f_X^{-1}(x)[\tilde{\phi}^*_{2t}(z,\delta|x) - E\{\tilde{\phi}^*_{2t}(Z,\Delta|x)|x\}],$$

$$B(z,\delta|x) = a_0 E\left[I(\Delta = 1) Z L'(F(Z_x^T|x)) h_{x,Z_x^T}(z,\delta)\right.$$

$$+I(\Delta=0)h_{x,Z_x^T}(z,\delta)\Big\{\frac{\int_{Z_x^T}^{T_x} M(y|x)\,dF(y|x)}{(1-F(Z_x^T|x))^2}-\frac{M(Z_x^T|x)}{1-F(Z_x^T|x)}\Big\}$$

$$+I(\Delta=0)h_{x,T_x}(z,\delta)\frac{M(T_x|x)}{1-F(Z_x^T|x)}-I(\Delta=0)\frac{\int_{Z_x^T}^{T_x} h_{x,y}(z,\delta)L(F(y|x))\,dy}{1-F(Z_x^T|x)}\Bigg]$$

$$+a_0 f_X^{-1}(x)[\tilde{\phi}_1^*(z,\delta|x)-E\{\tilde{\phi}_1^*(Z,\Delta|x)|x\}]$$

$$-\sum_{j=1}^{k}\frac{a_j A(F^{-1}(s_j|x),z,\delta|x)I(s_j\le F(T_x|x))}{f(F^{-1}(s_j|x)|x)},$$

where $Z_x^T=Z\wedge T_x$ and $M(y|x)=yL(F(y|x))$.

For a (sub)distribution function $L(y|x)$ we will use the notations $l(y|x)=L'(y|x)=(\partial/\partial y)L(y|x)$, $\dot{L}(y|x)=(\partial/\partial x)L(y|x)$ and similar notations will be used for higher order derivatives.

The assumptions needed for the results of Section 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 is a symmetric density with compact support, and K is twice continuously differentiable.

$(A2)(i)$ There exist $0\le s_a\le s_b\le 1$ such that $s_b\le\inf_x F(\tilde{T}_x|x)$, $s_a\le\inf\{s\in[0,1];J(s)\neq 0\}$, $s_b\ge\sup\{s\in[0,1];J(s)\neq 0\}$ and $\inf_{x\in R_X}\inf_{s_a\le s\le s_b} f(F^{-1}(s|x)|x)>0$.
(ii) J is twice continuously differentiable, $\int_0^1 J(s)ds=1$ and $J(s)\ge 0$ 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 twice continuously differentiable and $\inf_{x\in R_X}\sigma^0(x)>0$.
(iii) $E|Z|^\lambda<\infty$, with $\lambda\ge(12+8\delta)/(1+4\delta)$ and δ chosen as in $(A1)(i)$.

$(A4)$ $\eta(z,\delta|x)$ 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 $\sup_{x,y}|y^2L'(y|x)|<\infty$, 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) L is twice continuously differentiable, $\int_0^1 L(s)ds=1$, $L(s)\ge 0$ for all $0\le s\le 1$.

$(A7)$ For the density $f_{X|Z,\Delta}(x|z,\delta)$ of X given (Z,Δ), $\sup_{x,z}|f_{X|Z,\Delta}(x|z,\delta)|<\infty$,

$\sup_{x,z} |\dot{f}_{X|Z,\Delta}(x|z,\delta)| < \infty$, $\sup_{x,z} |\ddot{f}_{X|Z,\Delta}(x|z,\delta)| < \infty$ $(\delta = 0, 1)$.

For the proofs below, we will use throughout the abbreviated notations $\hat{T}^x = (T_x - \hat{m}^0(x))/\hat{\sigma}^0(x)$, $T^x = (T_x - m^0(x))/\sigma^0(x) = T$, $\hat{E}^0_{ix} = (Z_i - \hat{m}^0(x))/\hat{\sigma}^0(x)$, $E^0_{ix} = (Z_i - m^0(x))/\sigma^0(x)$, $\hat{E}^{0T}_{ix} = \hat{E}^0_{ix} \wedge \hat{T}^x$, $E^{0T}_{ix} = E^0_{ix} \wedge T$, $\hat{E}^{0T_t}_{ix} = (Z_i \wedge T_x \wedge t - \hat{m}^0(x))/\hat{\sigma}^0(x)$, $E^{0T_t}_{ix} = (Z_i \wedge T_x \wedge t - m^0(x))/\sigma^0(x)$, $\hat{T}^x_t = (T_x \wedge t - \hat{m}^0(x))/\hat{\sigma}^0(x)$ and $T^x_t = (T_x \wedge t - m^0(x))/\sigma^0(x)$.

Proof of Theorem 3.1. Consider the expression $\hat{m}^T(x) - m^T(x) = \Omega_1(x) + \Omega_2(x)$, where

$$\begin{aligned}
\Omega_1(x) &= a_0 \sum_{i=1}^n W_i(x, a_n)[\hat{\phi}^*_1(Z_i, \Delta_i|x) - \tilde{\phi}^*_1(Z_i, \Delta_i|x)] \\
&\quad + a_0 \sum_{i=1}^n W_i(x, a_n)[\tilde{\phi}^*_1(Z_i, \Delta_i|x) - E\{\tilde{\phi}^*_1(Z, \Delta|x)|x\}] \\
&= \Omega_{11}(x) + \Omega_{12}(x),
\end{aligned}$$

and

$$\Omega_2(x) = \sum_{j=1}^k a_j(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x) - \sum_{j=1}^k a_j(F^{-1}_{\phi 2}(s_j|x) \wedge T_x).$$

First, we treat $\Omega_{11}(x)$.

$$\begin{aligned}
\Omega_{11}(x) &= a_0 \sum_{i=1}^n W_i(x, a_n) \Big\{ I(\Delta_i = 1) Y_i [L(\hat{F}^0_\varepsilon(\hat{E}^{0T}_{ix})) - L(F^0_\varepsilon(E^{0T}_{ix}))] \\
&\quad + I(\Delta_i = 0) \Big[\frac{\int_{\hat{E}^{0T}_{ix}}^{\hat{T}^x} (\hat{m}^0(x) + \hat{\sigma}^0(x) e) L(\hat{F}^0_\varepsilon(e)) d\hat{F}^0_\varepsilon(e)}{1 - \hat{F}^0_\varepsilon(\hat{E}^{0T}_{ix})} \\
&\quad - \frac{\int_{E^{0T}_{ix}}^{T} (m^0(x) + \sigma^0(x) e) L(F^0_\varepsilon(e)) dF^0_\varepsilon(e)}{1 - F^0_\varepsilon(E^{0T}_{ix})} \Big] \Big\} \\
&= a_0 \sum_{i=1}^n W_i(x, a_n) \left\{ A_{1i}(x) + A_{2i}(x) \right\}.
\end{aligned}$$

We have $\sup_{x,z} \Big| \hat{F}^0_\varepsilon \Big\{ \frac{z \wedge T_x - \hat{m}^0(x)}{\hat{\sigma}^0(x)} \Big\} - F^0_\varepsilon \Big\{ \frac{z \wedge T_x - m^0(x)}{\sigma^0(x)} \Big\} \Big| = O_P((na_n)^{-1/2} (\log a_n^{-1})^{1/2})$. This is shown as follows. Write

$$\begin{aligned}
& \hat{F}^0_\varepsilon(\frac{z \wedge T_x - \hat{m}^0(x)}{\hat{\sigma}^0(x)}) - F^0_\varepsilon(\frac{z \wedge T_x - m^0(x)}{\sigma^0(x)}) \\
&= \hat{F}^0_\varepsilon(\frac{z \wedge T_x - \hat{m}^0(x)}{\hat{\sigma}^0(x)}) - F^0_\varepsilon(\frac{z \wedge T_x - \hat{m}^0(x)}{\hat{\sigma}^0(x)}) \\
&\quad + F^0_\varepsilon(\frac{z \wedge T_x - \hat{m}^0(x)}{\hat{\sigma}^0(x)}) - F^0_\varepsilon(\frac{z \wedge T_x - m^0(x)}{\hat{\sigma}^0(x)})
\end{aligned}$$

$$+F_\varepsilon^0(\frac{z\wedge T_x - m^0(x)}{\hat\sigma^0(x)}) - F_\varepsilon^0(\frac{z\wedge T_x - m^0(x)}{\sigma^0(x)})$$
$$= \alpha_n^1(z,x) + \alpha_n^2(z,x) + \alpha_n^3(z,x). \qquad \text{(A.1)}$$

Using Corollary 3.2 in Van Keilegom and Akritas (1999) (hereafter abbreviated by VKA), $\sup_{x,z}|\alpha_n^1(z,x)|$ is $O_p(n^{-1/2})$. For the two other terms, we use two first order Taylor developments :

$$\alpha_n^2(z,x) + \alpha_n^3(z,x) \ = \ -\frac{\hat m^0(x) - m^0(x)}{\hat\sigma^0(x)} f_\varepsilon^0(A_x) - \frac{\hat\sigma^0(x) - \sigma^0(x)}{\hat\sigma^0(x)} \frac{z\wedge T_x - m^0(x)}{\sigma^0(x)} f_\varepsilon^0(B_x),$$

for some A_x (B_x) between $\frac{z\wedge T_x - m^0(x)}{\hat\sigma^0(x)}$ and $\frac{z\wedge T_x - \hat m^0(x)}{\hat\sigma^0(x)}$ ($\frac{z\wedge T_x - m^0(x)}{\sigma^0(x)}$ and $\frac{z\wedge T_x - m^0(x)}{\hat\sigma^0(x)}$). Using Proposition 4.5 of VKA (1999) and the fact that $\sup_e |e f_\varepsilon^0(e)| < +\infty$, $\alpha_n^2(z,x)+\alpha_n^3(z,x) = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ *a.s.* Therefore, since $E[|Z|] < \infty$,

$$\sup_x |\sum_{i=1}^n W_i(x,a_n)A_{1i}(x)| = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).$$

Next, write

$$\begin{aligned}
A_{2i}(x) \ &= I(\Delta_i = 0)\Bigg\{ \frac{(\hat F_\varepsilon^0(\hat E_{ix}^{0T}) - F_\varepsilon^0(E_{ix}^{0T}))\int_{\hat E_{ix}^{0T}}^{\hat T^x}(\hat m^0(x) + \hat\sigma^0(x)e)L(\hat F_\varepsilon^0(e))d\hat F_\varepsilon^0(e)}{(1-\hat F_\varepsilon^0(\hat E_{ix}^{0T}))(1-F_\varepsilon^0(E_{ix}^{0T}))} \\
&+ \frac{\int_{\hat E_{ix}^{0T}}^{E_{ix}^{0T}}(\hat m^0(x) + \hat\sigma^0(x)e)L(\hat F_\varepsilon^0(e))d\hat F_\varepsilon^0(e)}{1-F_\varepsilon^0(E_{ix}^{0T})} + \frac{\int_T^{\hat T^x}(\hat m^0(x) + \hat\sigma^0(x)e)L(\hat F_\varepsilon^0(e))d\hat F_\varepsilon^0(e)}{1-F_\varepsilon^0(E_{ix}^{0T})} \\
&+ \frac{1}{1-F_\varepsilon^0(E_{ix}^{0T})}\int_{E_{ix}^{0T}}^T [(\hat m^0(x) - m^0(x)) + (\hat\sigma^0(x) - \sigma^0(x))e] \\
&\qquad \times [L(\hat F_\varepsilon^0(e)) - L(F_\varepsilon^0(e))]d\hat F_\varepsilon^0(e) \\
&+ \frac{1}{1-F_\varepsilon^0(E_{ix}^{0T})}\int_{E_{ix}^{0T}}^T [(\hat m^0(x) - m^0(x)) + (\hat\sigma^0(x) - \sigma^0(x))e]L(F_\varepsilon^0(e))dF_\varepsilon^0(e) \\
&+ \frac{1}{1-F_\varepsilon^0(E_{ix}^{0T})}\int_{E_{ix}^{0T}}^T (m^0(x) + \sigma^0(x)e)[L(\hat F_\varepsilon^0(e)) - L(F_\varepsilon^0(e))]d\hat F_\varepsilon^0(e) \\
&+ \frac{1}{1-F_\varepsilon^0(E_{ix}^{0T})}\int_{E_{ix}^{0T}}^T (m^0(x) + \sigma^0(x)e)L(F_\varepsilon^0(e))d(\hat F_\varepsilon^0(e) - F_\varepsilon^0(e))\Bigg\} \\
&= I(\Delta_i = 0)\sum_{j=1}^7 B_{ji}. \qquad \text{(A.2)}
\end{aligned}$$

Using Corollary 3.2 and Proposition 4.5 of VKA (1999), the above-mentioned uniform consistency of $\hat F_\varepsilon^0(\cdot)$, the continuous differentiability of L and the fact that $\sup_e |e f_\varepsilon^0(e)| < \infty$, it is easy to check that $A_{2i}(x) = |E_{ix}^{0T}| O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ such that

$$\sup_x |\sum_{i=1}^n W_i(x,a_n)A_{2i}(x)| = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).$$

We also have by Theorem 3.3 of Heuchenne (2005),

$$\sup_{x\in R_X} |\Omega_{12}(x)| = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2})\ a.s.,$$

since $E[|Z|^\lambda] < \infty$. Next, we treat $\Omega_2(x)$. First, we show that $\sup_{x\in R_X} |\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x| = O_P(1)$. Define

$$\xi_\alpha = \inf_{x\in R_X} (m^0(x) + \sigma^0(x)(F_\varepsilon^0)^{-1}(s_\alpha)) = \inf_{x\in R_X} F_{\phi 2}^{-1}(s_\alpha|x).$$

We have

$$\begin{aligned}
&P(\inf_{x\in R_X} (\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x) < \xi_\alpha) \qquad \text{(A.3)}\\
&\le P(\sup_{x\in R_X} |F_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) - s_j \wedge F_{\phi 2}(T_x|x)| \ge s_j \wedge F_{\phi 2}(T_x|x) - s_\alpha)\\
&\le P(\sup_{x\in R_X} |F_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) - \hat{F}_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)| \ge (s_j \wedge F_{\phi 2}(T_x|x) - s_\alpha)/2)\\
&\quad +P(\sup_{x\in R_X} |\hat{F}_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) - s_j \wedge F_{\phi 2}(T_x|x)| \ge (s_j \wedge F_{\phi 2}(T_x|x) - s_\alpha)/2).
\end{aligned}$$

Using Theorem 3.4 (consistency of $\hat{F}_{\phi_2}$), the first term on the right hand side of (A.3) tends to zero. For the second term on the right hand side of (A.3), write

$$\begin{aligned}
&P(\sup_{x\in R_X} |D_{1j}(x)| \ge \varepsilon_{j\alpha})\\
&\le P(D \sup_{x\in R_X} I(s_j \le \hat{F}_{\phi 2}(T_x|x), s_j > F_{\phi 2}(T_x|x)) \ge \varepsilon_{j\alpha}/4)\\
&\quad +P(D \sup_{x\in R_X} I(s_j > \hat{F}_{\phi 2}(T_x|x), s_j \le F_{\phi 2}(T_x|x)) \ge \varepsilon_{j\alpha}/4)\\
&\quad +P(\sup_{x\in R_X} \{|\hat{F}_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x)|x) - s_j| I(s_j \le \hat{F}_{\phi 2}(T_x|x), s_j \le F_{\phi 2}(T_x|x))\} \ge \varepsilon_{j\alpha}/4)\\
&\quad +P(\sup_{x\in R_X} \{|\hat{F}_{\phi 2}(T_x|x) - F_{\phi 2}(T_x|x)| I(s_j > \hat{F}_{\phi 2}(T_x|x), s_j > F_{\phi 2}(T_x|x))\} \ge \varepsilon_{j\alpha}/4)\\
&\quad = D_2 + D_3 + D_4 + D_5,
\end{aligned}$$

where $D_{1j}(x) = \hat{F}_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) - s_j \wedge F_{\phi 2}(T_x|x)$, $\varepsilon_{j\alpha} = (s_j \wedge F_{\phi 2}(T_x|x) - s_\alpha)/2$ and $D = \max_j(\sup_{x\in R_X} |D_{1j}(x)|)$. D_2, D_3 and D_5 tend to zero using Theorem 3.4. D_4 is bounded by

$$P(\sup_{x\in R_X} \sup_{-\infty<y<\infty} |\hat{F}_{\phi 2}(y|x) - \hat{F}_{\phi 2}(y-|x)| \ge \varepsilon_{j\alpha}/4),$$

for which Theorem 3.7 is used. Since $\inf_{x\in R_X}(F_{\phi 2}^{-1}(s_j|x) \wedge T_x) \ge \inf_{x\in R_X}(F_{\phi 2}^{-1}(s_\alpha|x)) = \xi_\alpha$, we have

$$\sup_{x\in R_X} |\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x - F_{\phi 2}^{-1}(s_j|x) \wedge T_x| := \sup_{x\in R_X} |D_{6j}(x)| = O_P(1).$$

$\Omega_2(x)$ is therefore rewritten as

$$\begin{aligned}\Omega_2(x) \quad &= \sum_{j=1}^{k} a_j D_{6j}(x) I(s_j \le \hat{F}_{\phi 2}(T_x|x), s_j > F_{\phi 2}(T_x|x)) \\ &+ \sum_{j=1}^{k} a_j D_{6j}(x) I(s_j \le \hat{F}_{\phi 2}(T_x|x), s_j \le F_{\phi 2}(T_x|x)) \\ &+ \sum_{j=1}^{k} a_j D_{6j}(x) I(s_j > \hat{F}_{\phi 2}(T_x|x), s_j \le F_{\phi 2}(T_x|x)), \end{aligned} \tag{A.4}$$

where the suprema of the first and third terms are negligible, using the same arguments as for D_2 and D_3. Note that when $s_j > F_\varepsilon^0(T)$ for all $j,\ j = 1\ldots, k$, only the first term of (A.4) is considered and treated with Theorem 3.4. Next, $\sup_{x\in R_X} |\Omega_2(x)|$ is now bounded by

$$\begin{aligned} &\sum_{j=1}^{k} |a_j| \sup_{x\in R_X} \{|D_{6j}(x)| I(s_j \le \hat{F}_{\phi 2}(T_x|x), s_j \le F_{\phi 2}(T_x|x), F_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) \in Q)\} \\ &+ \sum_{j=1}^{k} |a_j| \sup_{x\in R_X} \{|D_{6j}(x)| I(s_j \le \hat{F}_{\phi 2}(T_x|x), s_j \le F_{\phi 2}(T_x|x), F_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) \notin Q)\} \\ &+ O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}) \\ &= D_7 + D_8 + O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}), \end{aligned}$$

where D_8 is negligible by Theorems 3.4 and 3.7. Now, we define

$$\begin{aligned} D_{9j}(x) \ &= F^{-1}(\max(F(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x), s_\alpha) \wedge (s_\beta \wedge F_\varepsilon^0(T))|x) \\ &-F^{-1}(F(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)|x), \end{aligned}$$

such that

$$D_7 = \sum_{j=1}^{k} |a_j| \sup_{x\in R_X} \{|D_{9j}(x)| I(s_j \le \hat{F}_{\phi 2}(T_x|x), s_j \le F_{\phi 2}(T_x|x), F_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) \in Q)\}.$$

Therefore, using a Taylor development

$$D_{9j}(x) \le \frac{1}{f(F^{-1}(\theta_{jx}|x)|x)} |F_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) - F_{\phi 2}(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)|,$$

where θ_{jx} is between $\max(F(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x), s_\alpha) \wedge (s_\beta \wedge F_\varepsilon^0(T))$ and $F(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)$. Finally, the desired order is obtained with a successive application of Theorems 3.4 and 3.7.

Proof of Theorem 3.2. We continue to use the same notations as in the proof of Theorem 3.1. First, consider $\Omega_{11}(x)$.

$$\begin{aligned}
&\sum_{i=1}^n W_i(x,a_n)A_{1i}(x) \qquad \text{(A.5)}\\
&= \sum_{i=1}^n W_i(x,a_n)I(\Delta_i=1)Y_iL'(F_\varepsilon^0(E_{ix}^{0T}))(\hat F_\varepsilon^0(\hat E_{ix}^{0T}) - F_\varepsilon^0(E_{ix}^{0T})) + o_P((na_n)^{-1/2}),
\end{aligned}$$

using the uniform consistency of $\hat F_\varepsilon^0(\cdot)$ as in (A.1) and a second order Taylor expansion. Next, using Proposition 4.5 in VKA (1999), and the fact that $\sup_y |y^2 f_\varepsilon^{0'}(y)| < \infty$ and $\sup_y |y f_\varepsilon^0(y)| < \infty$,

$$\begin{aligned}
&F_\varepsilon^0(\hat E_{ix}^{0T}) - F_\varepsilon^0(E_{ix}^{0T})\\
&= (\hat E_{ix}^{0T} - E_{ix}^{0T})f_\varepsilon^0(E_{ix}^{0T}) + o_P((na_n)^{-1/2})\\
&= -\frac{\hat m^0(X_i) - m^0(X_i)}{\sigma^0(X_i)} f_\varepsilon^0(E_{ix}^{0T}) - \frac{\hat\sigma^0(X_i) - \sigma^0(X_i)}{\sigma^0(X_i)} E_{ix}^{0T} f_\varepsilon^0(E_{ix}^{0T}) + o_P((na_n)^{-1/2}) \qquad \text{(A.6)}
\end{aligned}$$

The asymptotic representation for $\hat F_\varepsilon^0(\hat E_{ix}^{0T}) - F_\varepsilon^0(E_{ix}^{0T})$ is therefore given by

$$(na_n)^{-1}\sum_{j=1}^n K(\frac{x-X_j}{a_n}) h_{x,Z_i\wedge T_x}(Z_j,\Delta_j) + o_P((na_n)^{-1/2}), \qquad \text{(A.7)}$$

where use is made of Propositions 4.8, 4.9 and Corollary 3.2 of VKA (1999). Next, consider the expression $\int_{\hat E_{ix}^{0T}}^{\hat T^x} (\hat m^0(x) + \hat\sigma^0(x)e)L(\hat F_\varepsilon^0(e))d\hat F_\varepsilon^0(e)$ which appears in the term B_{1i} of (A.2). We have

$$\begin{aligned}
&\int_{\hat E_{ix}^{0T}}^{\hat T^x} \hat m^0(x)L(\hat F_\varepsilon^0(e))d\hat F_\varepsilon^0(e) \qquad \text{(A.8)}\\
&= m^0(x)\left\{\int_{E_{ix}^{0T}}^{T} L(F_\varepsilon^0(e))dF_\varepsilon^0(e) + \int_{E_{ix}^{0T}}^{T} L(F_\varepsilon^0(e))d(\hat F_\varepsilon^0(e) - F_\varepsilon^0(e))\right\}\\
&\quad + O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}),
\end{aligned}$$

using the uniform consistency of $\hat m^0(\cdot)$ and $\hat F_\varepsilon^0(\cdot)$. By using integration by parts, the second term is $|E_{ix}^{0T}|O_P(n^{-1/2})$. In the same way,

$$\begin{aligned}
&\int_{\hat E_{ix}^{0T}}^{\hat T^x} \hat\sigma^0(x)eL(\hat F_\varepsilon^0(e))d\hat F_\varepsilon^0(e) \qquad \text{(A.9)}\\
&= \sigma^0(x)\left\{\int_{E_{ix}^{0T}}^{T} eL(F_\varepsilon^0(e))dF_\varepsilon^0(e) + \int_{E_{ix}^{0T}}^{T} eL(F_\varepsilon^0(e))d(\hat F_\varepsilon^0(e) - F_\varepsilon^0(e))\right.\\
&\quad \left. + \int_{\hat E_{ix}^{0T}}^{E_{ix}^{0T}} eL(\hat F_\varepsilon^0(e))d\hat F_\varepsilon^0(e) + \int_{T}^{\hat T^x} eL(\hat F_\varepsilon^0(e))d\hat F_\varepsilon^0(e)\right\} + |E_{ix}^{0T}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).
\end{aligned}$$

Using the fact that $\sup_e |ef_\varepsilon^0(e)| < \infty$, it is easily shown that the second, third and fourth terms are $|E_{ix}^{0T}| O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$. From this, we conclude

$$\sum_{i=1}^n W_i(x,a_n) I(\Delta_i = 0) B_{1i} \tag{A.10}$$
$$= \sum_{i=1}^n W_i(x,a_n) I(\Delta_i = 0) \times$$
$$\left\{ \frac{(\hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T}) - F_\varepsilon^0(E_{ix}^{0T}))[m^0(x) \int_{E_{ix}^{0T}}^T L(F_\varepsilon^0(e)) dF_\varepsilon^0(e) + \sigma^0(x) \int_{E_{ix}^{0T}}^T eL(F_\varepsilon^0(e)) dF_\varepsilon^0(e)]}{(1 - F_\varepsilon^0(E_{ix}^{0T}))^2} \right\}$$
$$+ o_P((na_n)^{-1/2}),$$

where the representation $(A.7)$ will be used for $\hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T}) - F_\varepsilon^0(E_{ix}^{0T})$. Next, consider the expression B_{2i}. Easy calculations show that

$$B_{2i} = \frac{\int_{\hat{E}_{ix}^{0T}}^{E_{ix}^{0T}} (m^0(x) + \sigma^0(x)e) L(F_\varepsilon^0(e)) d\hat{F}_\varepsilon^0(e)}{1 - F_\varepsilon^0(E_{ix}^{0T})} + |E_{ix}^{0T}| o_P((na_n)^{-1/2}).$$

We have

$$\int_{\hat{E}_{ix}^{0T}}^{E_{ix}^{0T}} (m^0(x) + \sigma^0(x)e) L(F_\varepsilon^0(e)) d\hat{F}_\varepsilon^0(e)$$
$$= \int_{\hat{E}_{ix}^{0T}}^{E_{ix}^{0T}} (m^0(x) + \sigma^0(x)e) L(F_\varepsilon^0(e)) dF_\varepsilon^0(e)$$
$$+ \int_{\hat{E}_{ix}^{0T}}^{E_{ix}^{0T}} (m^0(x) + \sigma^0(x)e) L(F_\varepsilon^0(e)) d(\hat{F}_\varepsilon^0(e) - F_\varepsilon^0(e)).$$

The second term on the right hand side of the equation above is $|E_{ix}^{0T}| O_P(n^{-1/2})$, which follows, using integration by parts, from Corollary 3.2 and Proposition 4.5 of VKA (1999) and the fact that $\sup_e |ef_\varepsilon^0(e)| < \infty$. Hence,

$$\begin{aligned} B_{2i} &= \frac{\int_{-\infty}^{E_{ix}^{0T}} (m^0(x) + \sigma^0(x)e) L(F_\varepsilon^0(e)) dF_\varepsilon^0(e) - \int_{-\infty}^{\hat{E}_{ix}^{0T}} (m^0(x) + \sigma^0(x)e) L(F_\varepsilon^0(e)) dF_\varepsilon^0(e)}{1 - F_\varepsilon^0(E_{ix}^{0T})} \\ &\quad + |E_{ix}^{0T}| o_P((na_n)^{-1/2}) \\ &= \frac{[\hat{m}^0(x) - m^0(x) + E_{ix}^{0T}(\hat{\sigma}^0(x) - \sigma^0(x))]}{\sigma^0(x)(1 - F_\varepsilon^0(E_{ix}^{0T}))} (m^0(x) + \sigma^0(x) E_{ix}^{0T}) L(F_\varepsilon^0(E_{ix}^{0T})) f_\varepsilon^0(E_{ix}^{0T}) \\ &\quad + |E_{ix}^{0T}| o_P((na_n)^{-1/2}), \end{aligned}$$

using a Taylor expansion. Note that the term $|E_{ix}^{0T}| o_P((na_n)^{-1/2})$ in the expression above is obtained from the fact that $\sup_e |ef_\varepsilon^0(e)| < \infty$ and $\sup_e |e^2 f_\varepsilon^{0\prime}(e)| < \infty$. A similar expression for B_{3i} is obtained such that

$$\sum_{i=1}^n W_i(x,a_n)I(\Delta_i=0)(B_{2i}+B_{3i}) \tag{A.11}$$
$$= \sum_{i=1}^n W_i(x,a_n)I(\Delta_i=0)\times$$
$$\Bigg\{\frac{[\hat{m}^0(x)-m^0(x)+E_{ix}^{0T}(\hat{\sigma}^0(x)-\sigma^0(x))]}{\sigma^0(x)(1-F_\varepsilon^0(E_{ix}^{0T}))}(m^0(x)+\sigma^0(x)E_{ix}^{0T})L(F_\varepsilon^0(E_{ix}^{0T}))f_\varepsilon^0(E_{ix}^{0T})$$
$$+\frac{[m^0(x)-\hat{m}^0(x)+T(\sigma^0(x)-\hat{\sigma}^0(x))]}{\sigma^0(x)(1-F_\varepsilon^0(E_{ix}^{0T}))}(m^0(x)+\sigma^0(x)T)L(F_\varepsilon^0(T))f_\varepsilon^0(T)\Bigg\}$$
$$+o_P((na_n)^{-1/2}).$$

B_{4i} and B_{6i} are $|E_{ix}^{0T}|o_P((na_n)^{-1/2})$. For B_{5i}, $\hat{F}_\varepsilon^0(e)$ is replaced by $F_\varepsilon^0(e)$ and the remaining terms are $|E_{ix}^{0T}|o_P((na_n)^{-1/2})$ using integration by parts, the uniform consistency of $\hat{m}^0(\cdot)$, $\hat{\sigma}^0(\cdot)$ and $\hat{F}_\varepsilon^0(\cdot)$ and the fact that $\sup_e |ef_\varepsilon^0(e)|<\infty$. Then, use is made of the asymptotic representations of Propositions 4.8 and 4.9 of VKA (1999) such that B_{5i} is given by

$$-\frac{(na_n)^{-1}f_X^{-1}(x)\sigma(x)}{1-F_\varepsilon^0(E_{ix}^{0T})}\Bigg\{\int_{E_{ix}^{0T}}^T L(F_\varepsilon^0(e))dF_\varepsilon^0(e)\sum_{j=1}^n K(\frac{x-X_j}{a_n})\eta(Z_j,\Delta_j|x)$$
$$+\int_{E_{ix}^{0T}}^T eL(F_\varepsilon^0(e))dF_\varepsilon^0(e)\sum_{j=1}^n K(\frac{x-X_j}{a_n})\zeta(Z_j,\Delta_j|x)\Bigg\}$$
$$+|E_{ix}^{0T}|o_P((na_n)^{-1/2}). \tag{A.12}$$

Finally, B_{7i} is $|E_{ix}^{0T}|O_P(n^{-1/2})$ using integration by parts. From those developments, we can write

$$\begin{aligned}\Omega_{11}(x) &= a_0\sum_{i=1}^n W_i(x,a_n)\tilde{B}_1(Z_i,\Delta_i|x)\times\frac{1}{na_n}\sum_{j=1}^n K(\frac{x-X_j}{a_n})\eta(Z_j,\Delta_j|x)\\ &\quad +a_0\sum_{i=1}^n W_i(x,a_n)\tilde{B}_2(Z_i,\Delta_i|x)\times\frac{1}{na_n}\sum_{j=1}^n K(\frac{x-X_j}{a_n})\zeta(Z_j,\Delta_j|x)\\ &\quad +o_P((na_n)^{-1/2}),\end{aligned} \tag{A.13}$$

where

$$\tilde{B}_1(Z_i,\Delta_i|x) = f_X^{-1}(x)\ \Big\{I(\Delta_i=1)Z_iL'(F(Z_{ix}^T|x))f_\varepsilon^0(E_{ix}^{0T})$$
$$+I(\Delta_i=0)\Bigg[f_\varepsilon^0(E_{ix}^{0T})\Bigg\{\frac{\int_{Z_{ix}^T}^{T_x} yL(F(y|x))dF(y|x)}{(1-F(Z_{ix}^T|x))^2}$$

$$-\frac{Z_{ix}^T L(F(Z_{ix}^T|x))}{1-F(Z_{ix}^T|x)}\Bigg\} + f_\varepsilon^0(T)\frac{T_x L(F(T_x|x))}{1-F(Z_{ix}^T|x)}$$

$$-\sigma^0(x)\frac{\int_{E_{ix}^{0T}}^T L(F_\varepsilon^0(e))dF_\varepsilon^0(e)}{1-F_\varepsilon^0(E_{ix}^{0T})}\Bigg]\Bigg\},$$

$$\tilde{B}_2(Z_i,\Delta_i|x) = f_X^{-1}(x)\ \Big\{I(\Delta_i=1)Z_iL'(F(Z_{ix}^T|x))f_\varepsilon^0(E_{ix}^{0T})E_{ix}^{0T}$$

$$+I(\Delta_i=0)\Bigg[f_\varepsilon^0(E_{ix}^{0T})E_{ix}^{0T}\Bigg\{\frac{\int_{Z_{ix}^T}^{T_x} yL(F(y|x))dF(y|x)}{(1-F(Z_{ix}^T|x))^2}$$

$$-\frac{Z_{ix}^T L(F(Z_{ix}^T|x))}{1-F(Z_{ix}^T|x)}\Bigg\} + f_\varepsilon^0(T)T\frac{T_x L(F(T_x|x))}{1-F(Z_{ix}^T|x)}$$

$$-\sigma^0(x)\frac{\int_{E_{ix}^{0T}}^T eL(F_\varepsilon^0(e))dF_\varepsilon^0(e)}{1-F_\varepsilon^0(E_{ix}^{0T})}\Bigg]\Bigg\},$$

and $Z_{ix}^T = Z_i \wedge T_x$, $i=1,\ldots,n$. Using Theorem 3.3 of Heuchenne (2005) for the new data points $\tilde{B}_1(Z,\Delta|x)$, $\tilde{B}_2(Z,\Delta|x)$, $\eta(Z,\Delta|x)$ and $\zeta(Z,\Delta|x)$, the asymptotic representation of $\Omega_{11}(x)$ is

$$\frac{1}{na_n}\sum_{j=1}^n K(\frac{x-X_j}{a_n})\tilde{B}_3(Z_j,\Delta_j|x) + R_{n1}(x), \tag{A.14}$$

where

$$\tilde{B}_3(Z,\Delta|x) = a_0(E[\tilde{B}_1(Z,\Delta|x)|x]\eta(Z,\Delta|x) + E[\tilde{B}_2(Z,\Delta|x)|x]\zeta(Z,\Delta|x)),$$

and $\sup\{|R_{n1}(x)|; x \in R_X\} = o_P((na_n)^{-1/2})$. Note that this rate can be obtained since $E[\eta(Z,\Delta|x)|x] = E[\zeta(Z,\Delta|x)|x] = 0$. For $\Omega_{12}(x)$, we readily obtain, using Theorem 3.3 of Heuchenne (2005) with new data points equal to 1 and $\tilde{\phi}_1^*(Z_i,\Delta_i|x) - E[\tilde{\phi}_1^*(Z,\Delta|x)|x]$,

$$\frac{a_0}{na_nf_X(x)}\sum_{i=1}^n K(\frac{x-X_i}{a_n})(\tilde{\phi}_1^*(Z_i,\Delta_i|x) - E[\tilde{\phi}_1^*(Z,\Delta|x)|x]) + R_{n2}(x), \tag{A.15}$$

where $\sup\{|R_{n2}(x)|; x \in R_X\} = o_P((na_n)^{-1/2})$.
Next, rewrite the second term on the right hand side of (A.4) as

$$\sum_{j=1}^k a_j(D_{6j}(x) - D_{10j}(x))I(s_j \le F_{\phi 2}(T_x|x), s_j \le \hat{F}_{\phi 2}(T_x|x))$$

$$+\sum_{j=1}^k a_jD_{10j}(x)I(s_j \le F_{\phi 2}(T_x|x), s_j \le \hat{F}_{\phi 2}(T_x|x)) = \Omega_{21}(x) + \Omega_{22}(x),$$

where

$$D_{10j}(x) = \frac{s_j \wedge F_{\phi 2}(T_x|x) - \hat{F}_{\phi 2}(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)}{f(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)}.$$

Using Theorem 3.7, $s_j \wedge F_{\phi 2}(T_x|x)$ can be replaced by $\hat{F}_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)$ in $D_{10j}(x)$ of $\Omega_{21}(x)$. We next rewrite $\Omega_{21}(x)$ as

$$\sum_{j=1}^{k} \frac{a_j(D_{11j}(x) + D_{12j}(x))}{f(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)} I(s_j \leq F_{\phi 2}(T_x|x), s_j \leq \hat{F}_{\phi 2}(T_x|x)), \tag{A.16}$$

where

$$D_{11j}(x) = f(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) D_{6j}(x) - (F(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) - F(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)),$$

and

$$\begin{aligned} D_{12j}(x) &= F_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) - F_{\phi 2}(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) \\ &\quad - \hat{F}_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) + \hat{F}_{\phi 2}(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x). \end{aligned}$$

Using a second order Taylor expansion, we get

$$D_{11j}(x) = -\frac{f'(\theta_{jx}|x)}{2} D_{6j}(x)^2,$$

where θ_{jx} is between $\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x$ and $F_{\phi 2}^{-1}(s_j|x) \wedge T_x$. Thus, using the proof of Theorem 3.1, the first term of (A.16) is $O_P((na_n)^{-1} \log a_n^{-1})$ since $\sup_{x,y} |f'(y|x)| < \infty$ and $\inf_x f(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) > 0$. Next, we treat the second term of (A.16). First, define

$$D_{13j}(x) = \frac{D_{12j}(x)}{f(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)} I(s_j \leq F_{\phi 2}(T_x|x), s_j \leq \hat{F}_{\phi 2}(T_x|x)).$$

The second term of (A.16) can then be rewritten as

$$\sum_{j=1}^{k} a_j D_{13j}(x) I(|D_{6j}(x)| > d_n) + \sum_{j=1}^{k} a_j D_{13j}(x) I(|D_{6j}(x)| \leq d_n),$$

where $d_n \sim (na_n)^{-1/2} (\log a_n^{-1})^{1/2}$. The first term of this expression is negligible using Theorem 3.1 and the second one is $o_P((na_n)^{-1/2})$ using Theorem 3.7. Finally, $\Omega_{22}(x)$ can be written as

$$\sum_{j=1}^{k} a_j D_{10j}(x) I(s_j \leq F_{\phi 2}(T_x|x)) + o_P((na_n)^{-1/2}),$$

where use is made of Theorems 3.4 and 3.5.

Proof of Theorem 3.4. Write

$$\hat{F}_{\phi 2}(t|x) - F_{\phi 2}(t|x) = \sum_{i=1}^{n} W_i(x,a_n) I(\Delta_i = 0) \left\{ \frac{\int_{\hat{E}_{ix}^{0T_t}}^{\hat{T}_t^x} d\hat{F}_\varepsilon^0(e)}{1 - \hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T})} - \frac{\int_{E_{ix}^{0T_t}}^{T_t^x} dF_\varepsilon^0(e)}{1 - F_\varepsilon^0(E_{ix}^{0T})} \right\}$$

$$+ \sum_{i=1}^{n} W_i(x,a_n) \{ \tilde{\phi}_{2t}^*(Z_i, \Delta_i|x) - E[\tilde{\phi}_{2t}^*(Z,\Delta|x)|x] \}$$

$$= \sum_{i=1}^{n} W_i(x,a_n) \{ I(\Delta_i = 0) \Omega_{3t}(Z_i,\Delta_i|x) + \Omega_{4t}(Z_i,\Delta_i|x) \}.$$

First, we treat $\Omega_{3t}(Z_i, \Delta_i|x)$. We have

$$\Omega_{3t}(Z_i,\Delta_i|x) = (\hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T}) - F_\varepsilon^0(E_{ix}^{0T})) \frac{\int_{\hat{E}_{ix}^{0T_t}}^{\hat{T}_t^x} d\hat{F}_\varepsilon^0(e)}{(1 - \hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T}))(1 - F_\varepsilon^0(E_{ix}^{0T}))}$$

$$+ \frac{\hat{F}_\varepsilon^0(\hat{T}_t^x) - F_\varepsilon^0(T_t^x) + F_\varepsilon^0(E_{ix}^{0T_t}) - \hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T_t})}{1 - F_\varepsilon^0(E_{ix}^{0T})}$$

$$= \Omega_{31t}(Z_i,\Delta_i|x) + \Omega_{32t}(Z_i,\Delta_i|x).$$

Since by (A.1), we showed

$$\sup_{x,t,z} |\hat{F}_\varepsilon^0(\frac{z \wedge T_x \wedge t - \hat{m}^0(x)}{\hat{\sigma}^0(x)}) - F_\varepsilon^0(\frac{z \wedge T_x \wedge t - m^0(x)}{\sigma^0(x)})| = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}),$$

we have

$$\sup_{x,t} |\sum_{i=1}^{n} W_i(x,a_n) \{ I(\Delta_i = 0)(\Omega_{31t}(Z_i,\Delta_i|x) + \Omega_{32t}(Z_i,\Delta_i|x)) \}|$$

$$= O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).$$

For $\Omega_{4t}(Z_i, \Delta_i|x)$, we use Theorem 3.3 of Heuchenne (2005) with new data points $\tilde{\phi}_{2t}^*(Z_i, \Delta_i|x)$ and we obtain the result.

Proof of Theorem 3.5. We continue to use the same notations as in the proof of Theorem 3.4. An asymptotic representation for the numerator of $\Omega_{32t}(Z_i, \Delta_i|x)$ is given by

$$\frac{1}{f_X(x) n a_n} \sum_{j=1}^{n} K(\frac{x - X_j}{a_n}) \Big\{ [f_\varepsilon^0(T_t^x) + f_\varepsilon^0(E_{ix}^{0T_t})] \eta(Z_j, \Delta_j|x) \qquad \text{(A.17)}$$

$$[f_\varepsilon^0(T_t^x) T_t^x + f_\varepsilon^0(E_{ix}^{0T_t}) E_{ix}^{0T_t}] \zeta(Z_j, \delta_j|x) \Big\} + o_P((na_n)^{-1/2}).$$

For $\Omega_{31t}(Z_i, \Delta_i|x)$, it is straightforward that

$$\Omega_{31t}(Z_i,\Delta_i|x) = (\hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T}) - F_\varepsilon^0(E_{ix}^{0T})) \frac{\int_{E_{ix}^{0T_t}}^{T_t^x} dF_\varepsilon^0(e)}{(1 - F_\varepsilon^0(E_{ix}^{0T}))^2} + o_P((na_n)^{-1/2}). \qquad \text{(A.18)}$$

Therefore, with (A.17) and (A.18), we obtain

$$\sum_{i=1}^{n} W_i(x,a_n)I(\Delta_i=0)\Omega_{3t}(Z_i,\Delta_i|x)$$
$$=\sum_{i=1}^{n} W_i(x,a_n)\tilde{B}_{4t}(Z_i,\Delta_i|x)\times\frac{1}{na_n}\sum_{j=1}^{n}K(\frac{x-X_j}{a_n})\eta(Z_j,\Delta_j|x)$$
$$+\sum_{i=1}^{n} W_i(x,a_n)\tilde{B}_{5t}(Z_i,\Delta_i|x)\times\frac{1}{na_n}\sum_{j=1}^{n}K(\frac{x-X_j}{a_n})\zeta(Z_j,\Delta_j|x)$$
$$+o_P((na_n)^{-1/2}), \tag{A.19}$$

where

$$\tilde{B}_{4t}(Z_i,\Delta_i|x)=f_X(x)^{-1}I(\Delta_i=0)\ \left\{\frac{f_\varepsilon^0(T_t^x)+f_\varepsilon^0(E_{ix}^{0T_t})}{1-F_\varepsilon^0(E_{ix}^{0T})}\right.$$
$$\left.+\frac{F_\varepsilon^0(T_t^x)-F_\varepsilon^0(E_{ix}^{0T_t})}{(1-F_\varepsilon^0(E_{ix}^{0T}))^2}f_\varepsilon^0(E_{ix}^{0T})\right\},$$

and

$$\tilde{B}_{5t}(Z_i,\Delta_i|x)=f_X(x)^{-1}I(\Delta_i=0)\ \left\{\frac{f_\varepsilon^0(T_t^x)T_t^x+f_\varepsilon^0(E_{ix}^{0T_t})E_{ix}^{0T_t}}{1-F_\varepsilon^0(E_{ix}^{0T})}\right.$$
$$\left.+\frac{F_\varepsilon^0(T_t^x)-F_\varepsilon^0(E_{ix}^{0T_t})}{(1-F_\varepsilon^0(E_{ix}^{0T}))^2}E_{ix}^{0T}f_\varepsilon^0(E_{ix}^{0T})\right\}.$$

Using Theorem 3.3 of Heuchenne (2005) for the new data points $\tilde{B}_{4t}(Z,\Delta|x)$, $\tilde{B}_{5t}(Z,\Delta|x)$, $\eta(Z,\Delta|x)$ and $\zeta(Z,\Delta|x)$, the asymptotic representation for (A.19) is

$$\frac{1}{na_n}\sum_{j=1}^{n}K(\frac{x-X_j}{a_n})\tilde{B}_{6t}(Z_j,\Delta_j|x)+R_{n1}(t|x), \tag{A.20}$$

where

$$\tilde{B}_{6t}(Z,\Delta|x)=E[\tilde{B}_{4t}(Z,\Delta|x)|x]\eta(Z,\Delta|x)+E[\tilde{B}_{5t}(Z,\Delta|x)|x]\zeta(Z,\Delta|x), \tag{A.21}$$

and $\sup\{|R_{n1}(t|x)|; x\in R_X\}=o_P((na_n)^{-1/2})$. For $\Omega_{4t}(Z_i,\Delta_i|x)$, we use Theorem 3.3 of Heuchenne (2005) with new data points equal to 1 and $\tilde{\phi}_{2t}^*(Z_i,\Delta_i|x)-E[\tilde{\phi}_{2t}^*(Z,\Delta|x)|x]$ such that we obtain

$$\sum_{i=1}^{n} W_i(x,a_n)\Omega_{4t}(Z_i,\Delta_i|x)$$
$$=\frac{1}{na_nf_X(x)}\sum_{i=1}^{n}K(\frac{x-X_i}{a_n})(\tilde{\phi}_{2t}^*(Z_i,\Delta_i|x)-E[\tilde{\phi}_{2t}^*(Z,\Delta|x)|x])+R_{n2}(t|x), \tag{A.22}$$

where $\sup\{|R_{n2}(t|x)|; -\infty < t < \infty, x \in R_X\} = o_P((na_n)^{-1/2})$.

Proof of Theorem 3.7. First, the expression in the theorem is bounded by the following sum, which is written in terms of new data points :

$$\sup_{x\in R_X,\ |t-s|\leq d_n} |\sum_{i=1}^n W_i(x,a_n)\Big\{\hat{\phi}^*_{2t}(Z_i,\Delta_i|x) - \hat{\phi}^*_{2s}(Z_i,\Delta_i|x)$$

$$-\tilde{\phi}^*_{2t}(Z_i,\Delta_i|x) + \tilde{\phi}^*_{2s}(Z_i,\Delta_i|x)\Big\}|$$

$$+ \sup_{x\in R_X,\ |t-s|\leq d_n} |\sum_{i=1}^n W_i(x,a_n)\Big\{\tilde{\phi}^*_{2t}(Z_i,\Delta_i|x) - \tilde{\phi}^*_{2s}(Z_i,\Delta_i|x)$$

$$-E[\tilde{\phi}^*_{2t}(Z,\Delta|x)|x] + E[\tilde{\phi}^*_{2s}(Z,\Delta|x)|x]\Big\}|$$

$$= D_1 + D_2.$$

$D_2 = o_P((na_n)^{-1/2})$ using Theorem 4.3 of Heuchenne (2005). Using classical arguments, we obtain

$$\begin{aligned}
&\hat{\phi}^*_{2t}(Z_i,\Delta_i|x) - \hat{\phi}^*_{2s}(Z_i,\Delta_i|x) - \tilde{\phi}^*_{2t}(Z_i,\Delta_i|x) + \tilde{\phi}^*_{2s}(Z_i,\Delta_i|x)\\
&= I(\Delta_i = 0)\Bigg\{(\hat{F}^0_\varepsilon(\hat{E}^{0T}_{ix}) - F^0_\varepsilon(E^{0T}_{ix}))\frac{F^0_\varepsilon(T^x_t) - F^0_\varepsilon(T^x_s) - F^0_\varepsilon(E^{0T_t}_{ix}) + F^0_\varepsilon(E^{0T_s}_{ix})}{(1-\hat{F}^0_\varepsilon(\hat{E}^{0T}_{ix}))(1-F^0_\varepsilon(E^{0T}_{ix}))}\\
&+\frac{\hat{F}^0_\varepsilon(\hat{T}^x_t) - F^0_\varepsilon(T^x_t) - \hat{F}^0_\varepsilon(\hat{T}^x_s) + F^0_\varepsilon(T^x_s)}{1-F^0_\varepsilon(E^{0T}_{ix})}\\
&+\frac{F^0_\varepsilon(E^{0T_t}_{ix}) - \hat{F}^0_\varepsilon(\hat{E}^{0T_t}_{ix}) + \hat{F}^0_\varepsilon(\hat{E}^{0T_s}_{ix}) - F^0_\varepsilon(E^{0T_s}_{ix})}{1-F^0_\varepsilon(E^{0T}_{ix})}\Bigg\} + O_P((na_n)^{-1}\log a_n^{-1})\\
&= D_{11} + D_{12} + D_{13} + O_P((na_n)^{-1}\log a_n^{-1}).
\end{aligned}$$

It is easily seen that $\sup_{x\in R_X,\ |t-s|\leq d_n} |D_{11}| = O_P((na_n)^{-1}\log a_n^{-1})$ using two Taylor developments and the fact that $\sup_e |f^0_\varepsilon(e)| < \infty$. Easy calculations show that

$$\begin{aligned}
D_{12} &= \frac{I(\Delta_i=0)}{1-F^0_\varepsilon(E^{0T}_{ix})}\Bigg\{\frac{\hat{m}^0(x) - m^0(x)}{\hat{\sigma}^0(x)}(f^0_\varepsilon(T^x_s) - f^0_\varepsilon(T^x_t))\\
&\quad+\frac{\hat{\sigma}^0(x) - \sigma^0(x)}{\hat{\sigma}^0(x)}[(T^x_s - T^x_t)f^0_\varepsilon(T^x_s) - T^x_t(f^0_\varepsilon(T^x_t) - f^0_\varepsilon(T^x_s))]\Bigg\} + o_P((na_n)^{-1/2}),
\end{aligned}$$

such that $\sup_{x\in R_X,\ |t-s|\leq d_n} |D_{12}| = o_P((na_n)^{-1/2})$. D_{13} is treated in a similar way and this finishes the proof.

References

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

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

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

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

Cai, Z. and Hong, Y. (2003). Weighted local linear approach to censored nonparametric regression. *Recent Advances and Trends in Nonparametric Statistics* (M. G. Akritas and D. M. Politis, eds.), 217-231.

Chen, S., Dahl, G. and Kahn, S. (2005). Nonparametric identification and estimation of a censored location-scale regression model. *J. Amer. Statist. Assoc.*, **100**, 212–221.

Dabrowska, D. M. (1987). Nonparametric regression with censored survival time data. *Scand. J. Statist.*, **14**, 181–197.

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

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

Fan, J. and Gijbels, I. (1994). Censored regression : local linear approximations and their applications. *J. Amer. Statist. Assoc.*, **89**, 560–570.

Heuchenne, C. (2005). Strong uniform consistency results of the weighted average of conditional artificial data points (submitted).

Heuchenne, C. and Van Keilegom, I. (2004). Polynomial regression with censored data based on preliminary nonparametric estimation (conditionally accepted by *Ann. Inst. Statist. Math.*).

Heuchenne, C. and Van Keilegom, I. (2005). Estimation in nonparametric location-scale regression models with censored data (submitted).

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457–481.

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

Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.*, **9**, 1276–1288.

Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika*, **74**, 301–309.

Li, G. and Datta, S. (2001). A bootstrap approach to non-parametric regression for right censored data. *Ann. Inst. Statist. Math.*, **53**, 708–729.

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

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

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

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

Van Keilegom, I. and Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Statist. Math.*, **49**, 467-491.

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

Vignali, C., Brandt, W.N. and Schneider, D.P. (2003). X-ray emission from radio-quiet quasars in the SDSS early data release. The $\alpha_{\rm OX}$ dependence upon luminosity. *Astronom. J.*, **125**, 433–443.

Address for correspondence :

Institut de Statistique
Université catholique de Louvain
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve
Belgium
Email: heuchenne@stat.ucl.ac.be