Polynomial Regression with Censored Data based on Preliminary Nonparametric Estimation

Cédric Heuchenne[1,2]

Institut de Statistique

Université catholique de Louvain

Ingrid Van Keilegom[1]

Institut de Statistique

Université catholique de Louvain

December 13, 2006

Abstract

Consider the polynomial regression model $Y = \beta_0 + \beta_1 X + \ldots + \beta_p X^p + \sigma(X)\varepsilon$, where $\sigma^2(X) = \mathrm{Var}(Y|X)$ is unknown, and ε is independent of X and has zero mean. Suppose that Y is subject to random right censoring. A new estimation procedure for the parameters $\beta_0, \ldots, \beta_p$ is proposed, which extends the classical least squares procedure to censored data. The proposed method is inspired by the method of Buckley and James (1979), but is, unlike the latter method, a non-iterative procedure due to nonparametric preliminary estimation of the conditional regression function. The asymptotic normality of the estimators is established. Simulations are carried out for both methods and they show that the proposed estimators have usually smaller variance and smaller mean squared error than the Buckley-James estimators. The two estimation procedures are also applied to a medical and an astronomical data set.

KEY WORDS: Bandwidth; Bootstrap; Kernel estimation; Least squares estimation; Linear regression; Nonparametric regression; Right censoring; Survival analysis.

[1] This research was supported by the contract 'Projet d'Actions de Recherche Concertées' nr. 98/03–217, and by Belgian IAP research network grant nr. P5/24.

[2] The first author is now at the University of Liège, HEC-Managment School of ULg.

1 Introduction

Suppose the random vector (X, Y) satisfies the polynomial regression model

$$Y = \beta_0 + \beta_1 X + \ldots + \beta_p X^p + \sigma(X)\varepsilon, \tag{1.1}$$

where $\sigma^2(X) = \mathrm{Var}(Y|X)$, and the error term ε (with unknown distribution F_ε) is independent of X and has zero mean. 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 the random variable C represents the censoring time, which is independent of Y, conditionally on X. Usually, Y is some known monotone transformation of the survival time. In case this transformation is the logarithmic transformation, model (1.1) is called the accelerated failure time model. 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) and let $V = (X, Z, \Delta)$ denote the vector of observed random variables.

A number of extensions to censored data of the least squares procedure for estimating $\beta_0, \ldots, \beta_p$ have been studied in the literature. The list of 'first-generation' estimators includes e.g. Miller (1976), Buckley and James (1979), Koul, Susarla and Van Ryzin (1981), and Leurgans (1987), while more recent contributions have been made by Zhou (1992), Stute (1993), Fygenson and Zhou (1994), Akritas (1994,1996) and Van Keilegom and Akritas (2000). The idea of the estimator of Buckley and James (1979) is as follows. Consider for simplicity the case where $p = 1$, and suppose that $\sigma(X) \equiv 1$. Then,

$$E(Y_i^*|X_i) = \beta_0 + \beta_1 X_i,$$

where $Y_i^* = Y_i\Delta_i + E(Y_i|Y_i > C_i, X_i)(1 - \Delta_i)$. The idea of Buckley and James (1979) is to write

$$E(Y_i|Y_i > C_i, X_i) = \beta_1 X_i + \frac{1}{1 - F_{\beta_1}(Z_i - \beta_1 X_i)} \int_{Z_i - \beta_1 X_i}^{\infty} y \, dF_{\beta_1}(y)$$

and next to estimate Y_i^* by the 'synthetic' data points

$$\hat{Y}_i^*(\beta_1) = Y_i\Delta_i + \Big\{\beta_1 X_i + \frac{1}{1 - \hat{F}_{\beta_1}(Z_i - \beta_1 X_i)} \int_{Z_i - \beta_1 X_i}^{\infty} y \, d\hat{F}_{\beta_1}(y)\Big\}(1 - \Delta_i),$$

where $F_{\beta_1}(y)$ is the distribution of $Y - \beta_1 X$ and $\hat{F}_{\beta_1}(y)$ is the Kaplan-Meier (1958) estimator of $F_{\beta_1}(y)$ based on $(Z_i - \beta_1 X_i, \Delta_i)$ $(i = 1, \ldots, n)$. Next, Buckley and James (1979) estimate the parameters (β_0, β_1) from the normal equations :

$$\begin{cases} \sum_{i=1}^{n} (\hat{Y}_i^*(\beta_1) - \beta_0 - \beta_1 X_i) = 0, \\ \sum_{i=1}^{n} (\hat{Y}_i^*(\beta_1) - \beta_0 - \beta_1 X_i) X_i = 0. \end{cases} \tag{1.2}$$

A solution to these equations can be found in an iterative way. Ritov (1990) and Lai and Ying (1991) obtained the asymptotic properties of a (slightly modified) version of this estimator.

Although this estimator behaves usually well in practice, there are a number of disadvantages : (1) the iterative procedure suffers in certain cases from convergence problems which lead to unstable solutions or no solution at all; and (2) the estimation method restricts to homoscedastic models, while in practice the data often follow a heteroscedastic model. In light of these drawbacks, we propose in this paper a variant of the Buckley-James procedure, which does not suffer from the above disadvantages. The idea is to estimate $E(Y_i|Y_i > C_i, X_i)$ (and hence Y_i^*) in a nonparametric way. This is done by using kernel smoothing with an adaptively chosen bandwidth parameter. The advantage of this is that, contrary to the Buckley-James procedure, the so-obtained 'synthetic' data points do not depend on the unknown β-vector and hence the normal equations have an explicit (non-iterative) solution. As will be seen in the simulations, this leads to more stable solutions and hence to a smaller variance. Moreover, contrary to other methods which construct 'synthetic' data points (e.g. Koul, Susarla and Van Ryzin (1981), Leurgans (1987), Akritas (1996)), the 'synthetic' data points of the new method use information from the whole model. The details of the proposed method are given in the next section.

This paper is organized as follows. In the next section, we introduce some notations and describe the estimation procedure in detail. In Section 3 we state the asymptotic normality result of the regression parameter estimators. Section 4 contains a simulation study, in which the new procedure is compared with the Buckley-James method, while in Section 5 two data sets on cancer of the larynx and on spectral energy distributions of quasars are analyzed by means of the two methods. Finally, the Appendix contains the proofs of the main results of Section 3.

2 Notations and description of the method

We assume throughout that regression model (1.1) holds. Let $m(\cdot)$ be any location function and $\sigma(\cdot)$ be any scale function, meaning that $m(x) = T(F(\cdot|x))$ and $\sigma(x) = S(F(\cdot|x))$ 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 $Y = m(X) + \sigma(X)\varepsilon$ with ε independent of X, is also valid. So from now on, m and σ can denote any location and scale function, and are not restricted to the conditional mean and variance. Also, we use the notation $\varepsilon = (Y - m(X))/\sigma(X)$ for any location function m and scale function σ. Define $F(y|x) = P(Y \leq y|x)$, $G(y|x) =$

$P(C \leq y|x)$, $H(y|x) = P(Z \leq y|x)$, $H_\delta(y|x) = P(Z \leq y, \Delta = \delta|x)$, $H(y) = P(Z \leq y)$, $F_X(x) = P(X \leq x)$, $F_e(y) = P(\varepsilon \leq y)$, $S_e(y) = 1 - F_e(y)$, and for $E = (Z - m(X))/\sigma(X)$ we denote $H_e(y) = P(E \leq y)$, $H_{e\delta}(y) = P(E \leq y, \Delta = \delta)$, $H_e(y|x) = P(E \leq y|x)$ and $H_{e\delta}(y|x) = P(E \leq y, \Delta = \delta|x)$ $(\delta = 0, 1)$. The probability density functions of the distributions defined above will be denoted with lower case letters, and let R_X denote the support of the variable X.

As already outlined in Section 1, the idea of the proposed method is to estimate $E(Y_i|Y_i > C_i, X_i)$ in a nonparametric way, in order to obtain a direct non-iterative estimator for the β-coefficients. One can write

$$E(Y_i|Y_i > C_i, X_i) = m(X_i) + \frac{\sigma(X_i)}{1 - F_e(E_i)} \int_{E_i}^{\infty} y\, dF_e(y). \tag{2.1}$$

The main idea is now to estimate $m(\cdot), \sigma(\cdot)$ and $F_e(\cdot)$ in a nonparametric way and to plug-in the so-obtained estimator of $E(Y_i|Y_i > C_i, X_i)$ into the formula of Y_i^*. Since these new Y_i^*'s do not depend on the β-coefficients, the resulting minimization problem and normal equations (similar to equation (1.2)) yield explicit solutions for β. However, due to the censoring mechanism, it is in general impossible to obtain consistent, nonparametric estimators of the conditional mean and variance. We will therefore use location and scale functions $m(\cdot)$ and $\sigma(\cdot)$, that can be estimated in a consistent way under censoring (and change $F_e(\cdot)$ accordingly). Since equation (2.1) remains valid when m and σ are any location and scale function respectively, we can choose for them the following L-functions:

$$m(x) = \int_0^1 F^{-1}(s|x) J(s)\, ds, \quad \sigma^2(x) = \int_0^1 F^{-1}(s|x)^2 J(s)\, ds - m^2(x), \tag{2.2}$$

where $F^{-1}(s|x) = \inf\{y; F(y|x) \geq s\}$ is the quantile function of Y given x and $J(s)$ is a given score function satisfying $\int_0^1 J(s)\, ds = 1$. When $J(s)$ is chosen appropriately (namely put to zero in the right tail, there where the quantile function cannot be estimated in a consistent way due to the right censoring), $m(x)$ and $\sigma(x)$ can be estimated consistently. Now, replace the distribution $F(y|x)$ in (2.2) by the Beran (1981) estimator, defined by :

$$\hat{F}(y|x) = 1 - \prod_{Z_i \leq y, \Delta_i = 1} \left\{1 - \frac{W_i(x, a_n)}{\sum_{j=1}^n I(Z_j \geq Z_i) W_j(x, a_n)}\right\} \tag{2.3}$$

(in the case of no ties), where $W_i(x, a_n)$ $(i = 1, \ldots, n)$ are the Nadaraya-Watson weights

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

K is a kernel function 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

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

as estimators for $m(x)$ and $\sigma^2(x)$. Let

$$\hat{F}_e(y) = 1 - \prod_{\hat{E}_{(i)} \le y, \Delta_{(i)}=1} \left(1 - \frac{1}{n-i+1}\right) \ , \tag{2.5}$$

denote the proposed Kaplan-Meier (1958) estimator of F_e (in the case of no ties), where $\hat{E}_i = (Z_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$, $\hat{E}_{(i)}$ is the i-th order statistic of $\hat{E}_1, \ldots, \hat{E}_n$ and $\Delta_{(i)}$ is the corresponding censoring indicator. This estimator has been studied in detail by Van Keilegom and Akritas (1999). This leads to

$$\hat{Y}^*_{Ti} = Y_i \Delta_i + \Big\{\hat{m}(X_i) + \frac{\hat{\sigma}(X_i)}{1 - \hat{F}_e(\hat{E}^T_i)} \int_{\hat{E}^T_i}^{\hat{S}_i} y\, d\hat{F}_e(y)\Big\}(1 - \Delta_i) \tag{2.6}$$

as an estimator of Y^*_i, where $\hat{S}_i = (T_{X_i} - \hat{m}(X_i))/\hat{\sigma}(X_i)$, $\hat{E}^T_i = \hat{E}_i \wedge \hat{S}_i$, and for any x, $T_x \le T\sigma(x) + m(x)$, where $T < \tau_{H_e}$ and $\tau_F = \inf\{y : F(y) = 1\}$ for any distribution F. Note that due to the right censoring, we have to truncate the integral in the definition of $\hat{Y}^*_{Ti}$ (however, when $\tau_{F_e} \le \tau_{G_e}$, the bound $\hat{S}_i$ can be chosen arbitrarily close to τ_{F_e} for n sufficiently large). Finally, define the estimator of $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)$ by the usual least squares estimator based on the pairs $(X_i, \hat{Y}^*_{Ti})$ $(i = 1, \ldots, n)$ and denote these estimators by $\hat{\boldsymbol{\beta}}_T = (\hat{\beta}_{T0}, \ldots, \hat{\beta}_{Tp})$. As it is clear from the definition of $\hat{Y}^*_{Ti}$, $\hat{\beta}_{T0}, \ldots, \hat{\beta}_{Tp}$ are actually estimating $\boldsymbol{\beta}_T = (\beta_{T0}, \ldots, \beta_{Tp})' = (\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}' E(Y^*_{Ti}|X_i)^n_{i=1}$ (conditionally on $X_1, \ldots, X_n$), where the element (i, j) of the matrix $\mathcal{X}$ equals X_i^{j-1} $(i = 1, \ldots, n; j = 1, \ldots, p + 1)$,

$$Y^*_{Ti} = Y_i \Delta_i + \Big\{m(X_i) + \frac{\sigma(X_i)}{1 - F_e(E^T_i)} \int_{E^T_i}^{S_i} y\, dF_e(y)\Big\}(1 - \Delta_i),$$

$S_i = (T_{X_i} - m(X_i))/\sigma(X_i)$ and $E^T_i = (Z_i \wedge T_{X_i} - m(X_i))/\sigma(X_i) = E_i \wedge S_i$. As before, these coefficients $\beta_{T0}, \ldots, \beta_{Tp}$ can be made arbitrarily close to $\beta_0, \ldots, \beta_p$, provided $\tau_{F_e} \le \tau_{G_e}$.

Another way to construct new data points should be to replace each data point Y_i by an estimation of its conditional location function $m(X_i)$. This alternative estimation method has been studied by Akritas (1996) (Biometrics). The method of Akritas offers the advantage of being more robust to outliers, since all observations are transformed, whereas in our method we only change the censored observations. On the other hand, our method has the advantage of making use of the model $Y = m(X) + \sigma(X)\epsilon$ in the construction of the synthetic data points, and so it uses the model in a more efficient way. In particular, this leads to an estimator that is less sensible to regions with heavy censoring.

3 Asymptotic results

We start with developing an asymptotic representation for $\hat{\beta}_{Tj} - \beta_{Tj}$ $(j = 0, \ldots, p)$. This representation is useful to obtain afterwards the asymptotic normality of the estimators. The assumptions and notations used in the results below, as well as the proof of the first result, are given in the Appendix.

Theorem 3.1 *Assume (A1)-(A8), Then,*

$$\begin{pmatrix} \hat{\beta}_{T0} - \beta_{T0} \\ \vdots \\ \hat{\beta}_{Tp} - \beta_{Tp} \end{pmatrix} = M^{-1} n^{-1} \sum_{i=1}^{n} \rho(X_i, Z_i, \Delta_i) + \begin{pmatrix} o_P(n^{-1/2}) \\ \vdots \\ o_P(n^{-1/2}) \end{pmatrix},$$

where $M = (M_{jk})$ $(j, k = 1, \ldots, p+1)$, $M_{jk} = E(X^{j+k-2})$, $\rho = (\rho_0, \ldots, \rho_p)'$,

$$\begin{aligned} \rho_j(X_i, Z_i, \Delta_i) \quad = \quad & \int_{R_X} x^j \sigma(x) \int_{-\infty}^{+\infty} \left\{ \frac{\varphi(X_i, Z_i, \Delta_i, e_x^T(z))}{(1 - F_e(e_x^T(z)))^2} \int_{e_x^T(z)}^{S_x} u \, dF_e(u) \right. \\ & \left. + \frac{1}{1 - F_e(e_x^T(z))} \int_{e_x^T(z)}^{S_x} u \, d\varphi(X_i, Z_i, \Delta_i, u) \right\} dH_0(z|x) dF_X(x) \\ & + f_X(X_i) \int B_j(z, Z_i, \Delta_i | X_i) \, dH_0(z|X_i) + X_i^j (Y_{Ti}^* - E[Y_{Ti}^* | X_i]) \end{aligned}$$

$(j = 0, \ldots, p; i = 1, \ldots, n)$.

Theorem 3.2 *Under the assumptions of Theorem 3.1,* $n^{1/2}(\hat{\beta}_{T0} - \beta_{T0}, \ldots, \hat{\beta}_{Tp} - \beta_{Tp})' \xrightarrow{d} N(0, \Sigma)$*, where*

$$\Sigma = M^{-1} E[\rho(X, Z, \Delta) \rho'(X, Z, \Delta)] M^{-1}.$$

The proof of this result follows readily from Theorem 3.1.

Remark 3.3 (Homoscedastic model) Note that when model (1.1) is homoscedastic (i.e. $\sigma \equiv 1$), the representation in Theorem 3.1 simplifies. In fact, it is easily seen that the function ζ equals zero in that case.

Remark 3.4 (Bandwidth choice) The choice of the bandwidth parameter can be carried out through the minimization of the function

$$\min_{a_n} \sum_{i=1}^{n} (\hat{Y}_{Ti}^*(a_n) - \hat{\beta}_{T0}(a_n) - \ldots - \hat{\beta}_{Tp}(a_n) X_i^p)^2, \tag{3.1}$$

over a specific grid of values of the smoothing parameter a_n. The rationale of this bandwidth rule is to minimize the least squares criterium function, not only with respect to the parameters $\boldsymbol{\beta}_T$, but also with respect to the bandwidth a_n. This idea has been used in other contexts as well, see e.g. Härdle, Hall and Ichimura (1993) where a similar principle is used in the context of single index models. Note that the argument a_n is added to $\hat{Y}^*_{Ti}$ and $\hat{\beta}_{Tj}$ $(i = 1, \ldots, n; j = 0, \ldots, p)$ in order to highlight the dependence on a_n of these quantities. This procedure to select the bandwidth is illustrated in Section 4 on some finite sample simulations.

Remark 3.5 (Bootstrap approximation) For the computation of the variance of the estimator $\hat{\beta}_T$ 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 $\hat{F}(\cdot|X^*_i)$, and a C^*_i from $\hat{G}(\cdot|X^*_i)$ (which is the Beran (1981) estimator of $G(\cdot|X^*_i)$ obtained by replacing Δ_i by $1-\Delta_i$ in the expression for $\hat{F}(\cdot|X^*_i)$). Finally, let $Z^*_i = \min(Y^*_i, C^*_i)$ and $\Delta^*_i = I(Y^*_i \leq C^*_i)$. For each so-obtained resample $\{(X^*_i, Z^*_i, \Delta^*_i) : i = 1, \ldots, n\}$, calculate a bootstrap estimator of the regression parameters. Repeat this for a large number of bootstrap samples (say B). The variance of these B bootstrap estimates is then an approximation of the variance of the estimator $\hat{\beta}_T$. In a similar way, the bootstrap can also be used to approximate the full distribution of $\hat{\beta}_T$.

Remark 3.6 (Practical implementation) The proposed estimator can be easily implemented in practice. In fact, the parameters on which the estimator depends, can all be chosen in an adaptive way. The finite sample performance of $\hat{\beta}_T$ for these adaptively chosen parameters is illustrated in the next section. Programs (written in Matlab) of the estimator $\hat{\beta}_T$ can be obtained by simple request to the authors. First of all, for the score function J, we recommend the choice

$$J(s) = b^{-1} I(0 \leq s \leq b) \quad (0 \leq s \leq 1),$$

where $b = \min_{1 \leq i \leq n} \hat{F}(+\infty|X_i)$. In this way, the region where the Beran estimators $\hat{F}(\cdot|X_1), \ldots, \hat{F}(\cdot|X_n)$ are inconsistent is not used, and on the other hand, we exploit to a maximum the "consistent" region. For the bandwidth, the procedure explained in Remark 3.4 is completely data-driven and easy to implement whereas the choice of the kernel K is of minor importance. Finally, $\hat{S}_i$ $(i = 1, \ldots, n)$ can be chosen larger (or equal) than the last order statistic $\hat{E}_{(n)}$ of the estimated residuals. In this way, all the Kaplan-Meier jumps of the integral (2.6) are considered.

Remark 3.7 (Extensions) The estimation procedure and the methodology used to obtain the results of this section could be used as a basis for a number of more general models. For instance, it could be studied how the proposed estimation method can be adapted to any (non)linear parametric regression model with censored data. Also, the extension to situations where the covariate is subject to censoring could be considered (in that case the Beran estimator will need to be replaced by e.g. the estimator proposed in Van Keilegom (2003)). Finally, it would be interesting to extend the obtained results to semiparametric regression models, like partial linear or single index models.

4 Simulations

In this section we compare the finite sample behavior of the Buckley-James (1979) estimator with the estimator proposed in this paper by means of Monte Carlo simulations. We are primarily interested in the behavior of the bias and variance of the two estimators. The simulations are carried out for samples of size $n = 100$ and the results are obtained by using 500 simulations.

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

$$Y = \beta_0 + \beta_1 X + \sigma\varepsilon, \tag{4.1}$$

for various choices of β_0, β_1 and σ, where X has a uniform distribution on the unit interval and the error term ε is a standard normal random variable. The censoring variable C satisfies $C = \alpha_0 + \alpha_1 X + \sigma\varepsilon^*$, for certain choices of α_0 and α_1 and where ε^* has a standard normal distribution. We further assume that ε and ε^* are independent of X, and that ε is independent of ε^*. It is easy to see that, under this model,

$$P(\Delta = 0|X = x) = 1 - \Phi\Big(\frac{\alpha_0 - \beta_0 + (\alpha_1 - \beta_1)x}{\sqrt{2}\sigma}\Big).$$

For the weights that appear in the Beran estimator $\hat{F}(y|x)$, we choose a biquadratic kernel function $K(x) = (15/16)(1 - x^2)^2 I(|x| \leq 1)$. In order to improve the behavior near the boundaries of the covariate space, we work with the boundary corrected kernels proposed by Müller and Wang (1994). As a consequence of the fact that these boundary corrected kernels can become negative, the Beran estimator decreases at certain time points. In these cases, the estimator is redefined as being constant until it starts increasing again.

For the bandwidth sequence a_n, we select the minimizer of (3.1) among a grid of 20 possible values between 0 and 1. For small values of a_n, the window $[x - a_n, x + a_n]$ at a point x does sometimes not contain any X_i $(i = 1, \dots, n)$ 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.

In a number of situations, the iterative Buckley-James method does not converge, but oscillates around two or more values. In such cases, the estimator is defined as the average of these values.

Table 1 summarizes the simulation results for different values of $\alpha_0, \alpha_1, \beta_0, \beta_1$ and σ. For fixed values of β_0, β_1 and σ, the values of α_0 and α_1 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 table shows that, in general, the Buckley-James estimator has a larger variance but a smaller bias than the newly proposed estimator. In most cases the effect of the bias on the mean squared error is however small (relative to the variance). As a consequence, the new estimator has in most cases a smaller mean squared error than the Buckley-James estimator. These facts can be explained in the following way. First, that the new estimator has a larger bias than the Buckley-James estimator is due to the use of smoothing methods. They imply a certain inherent bias, but the contribution of this bias to the mean squared error is in most cases small. Second, the smoothing parameter a_n gives an additional possibility to fine-tune the new estimation procedure. The dependence on a bandwidth a_n can thus be considered as an advantage for the new method, since it allows to optimize the estimation procedure. Third, the Buckley-James estimator suffers in certain cases from instability problems that are inherent to this method, as explained in Section 1.

Next, suppose that Y and C are distributed according to

$$\begin{aligned} Y|X = x &\sim \mathit{Weibull}(\exp[-d(\gamma_0 + \gamma_1 x + \gamma_2 x^2)], d), \\ C|X = x &\sim \mathit{Weibull}(\exp[-d(\alpha_0 + \alpha_1 x + \alpha_2 x^2)], d) \end{aligned}$$

and are independent conditionally on X. The covariate X is uniformly distributed on $[0, 1]$. It is easy to check that

$$\begin{aligned} \log Y|X = x &\sim F(y|x) = 1 - \exp(-\exp[d(y - \gamma_0 - \gamma_1 x - \gamma_2 x^2)]), \qquad (4.2) \\ \log C|X = x &\sim G(y|x) = 1 - \exp(-\exp[d(y - \alpha_0 - \alpha_1 x - \alpha_2 x^2)]). \end{aligned}$$

It follows that $\log Y$ has, conditionally on $X = x$, an extreme value distribution and hence $E(\log Y|X = x) = -D/d + \gamma_0 + \gamma_1 x + \gamma_2 x^2 = \beta_0 + \beta_1 + \beta_2 x^2$ and $Var(\log Y|X = x) = \pi^2/(6d^2)$, where $\beta_0 = -D/d + \gamma_0$, $\beta_1 = \gamma_1$, $\beta_2 = \gamma_2$ and $D = 0.5772$ is the Euler

β_0	β_1		$\hat{\beta}_0$			$\hat{\beta}_1$		
α_0	α_1	σ^2	Bias	Var	MSE	Bias	Var	MSE
0	1		-.004	.022	.022	-.009	.068	.069
0.6	0.85	0.5	.005	.021	.021	-.019	.065	.066
0	1		-.013	.026	.026	-.011	.084	.084
0.27	0.45	0.5	-.009	.024	.024	-.043	.075	.077
0	1		-.006	.041	.041	-.015	.141	.141
1.5	-0.5	1	.002	.040	.040	-.052	.135	.137
0	1		-.018	.050	.050	-.013	.169	.169
0.6	-0.2	1	-.008	.047	.047	-.074	.153	.158
0	5		-.004	.021	.021	-.011	.069	.069
1	4.1	0.5	.008	.021	.021	-.050	.067	.069
0	5		-.013	.025	.025	-.006	.088	.088
0.5	4	0.5	.011	.025	.025	-.079	.086	.092
0	5		-.006	.042	.042	-.014	.138	.138
1.3	3.9	1	.009	.041	.041	-.067	.130	.135
0	5		-.015	.047	.047	-.004	.186	.186
1	3	1	.033	.047	.048	-.170	.171	.200

Table 1: *Results for the Buckley-James estimator (first line) and the new estimator (second line) for model (4.1).*

constant. It easily follows that if $m(x) = E(\log Y|x)$ and $\sigma^2(x) = \mathrm{Var}(\log Y|x)$, then $P(\varepsilon \leq y|x) = P((\log Y - m(x))/\sigma(x) \leq y|x) = 1 - \exp(-\exp(y\pi/\sqrt{6} - D))$. Since this is independent of x, model (1.1) holds. Further, with $a_x = \exp(-d(\gamma_0 + \gamma_1 x + \gamma_2 x^2))$ and $b_x = \exp(-d(\alpha_0 + \alpha_1 x + \alpha_2 x^2))$, the conditional censoring probability curve is given by $P(\Delta = 0|X = x) = b_x/(a_x + b_x)$.

The bias, variance and mean squared error of the new and the Buckley-James estimator for 16 sets of parameters are given in Table 2. The results are similar (but even more pronouncing) than in Table 1 : in most cases, the new estimator has a slightly larger bias, but a much smaller variance, which leads to a substantial smaller mean squared error compared to the Buckley-James estimator. Other choices of the parameters lead to similar results.

The final setting we consider is a normal heteroscedastic regression model

$$Y = \beta_0 + \beta_1 X + \gamma X \varepsilon, \tag{4.3}$$

with $\beta_0 = 0$, $\beta_1 = 10$, X has a uniform distribution on $[0, 1]$, ε has a standard normal distribution, and γ equals $1, 2, 3$ or 5. The censoring variable is given by $C = \alpha_0 + \alpha_1 X + \varrho\varepsilon^*$, where ε^* has a standard normal distribution. We further assume that ε and ε^* are independent of X, and that ε is independent of ε^*. As the Buckley-James estimator is limited to homoscedastic models, we continue using the same estimator as before, while the new estimator is now taking the heteroscedasticity into account. Therefore, we expect the Buckley-James estimator to behave poorly when there is much heteroscedasticity in the model. This is indeed confirmed by the results in Table 3, which show deteriorating results for the Buckley-James estimator for increasing values of γ.

A final remark on the choice of the bandwidth : simulations have shown that the estimator proposed in this paper is not very sensitive (relatively to other situations where kernel smoothing is used) to the choice of the bandwidth. This is because the estimators of the regression parameters are obtained by taking a weighted average of the artificial data points $\hat{Y}_i^*$ $(i = 1, \ldots, n)$. In this way, the effect of the choice of the bandwidth is in some way averaged out. This is a typical phenomenon in situations where kernel smoothing is used in the construction of a root-n consistent estimator.

5 Data analysis

We illustrate the proposed method on two data sets. The first one is about 90 male larynx cancer patients, diagnosed and treated during the period 1970-1978 in a peripheral hospital in the Netherlands (see Kardaun (1983) for more details). The variable of interest is the time interval (in years) between first treatment and death of the patient. At the

γ_0	γ_1	γ_2		$\hat{\beta}_0$			$\hat{\beta}_1$			$\hat{\beta}_2$		
α_0	α_1	α_2	d	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
7.6	1	1		.013	.069	.069	-.053	1.66	1.66	.043	1.62	1.62
8.7	-0.2	1	5/3	-.014	.064	.064	.172	1.45	1.48	-.248	1.35	1.41
7.6	1	1		.013	.085	.085	-.064	2.32	2.32	.067	2.41	2.41
8.2	-0.2	1	5/3	-.032	.073	.074	.323	1.72	1.82	-.452	1.60	1.81
7.6	1	1		.022	.200	.201	-.088	4.70	4.71	.061	4.50	4.50
9	-0.2	1	1	-.014	.182	.183	.174	4.07	4.10	-.257	3.78	3.85
7.6	1	1		.027	.255	.255	-.117	6.43	6.45	.100	6.36	6.37
8.2	-0.2	1	1	-.046	.205	.207	.381	4.60	4.75	-.507	4.27	4.53
7.6	5	1		.013	.070	.070	-.054	1.66	1.66	.040	1.60	1.60
8.6	4	1	5/3	-.004	.069	.069	.121	1.56	1.57	-.185	1.44	1.48
7.6	5	1		.014	.087	.087	-.072	2.31	2.31	.069	2.35	2.35
8.1	4	1	5/3	-.006	.090	.090	.212	2.17	2.22	-.316	2.05	2.15
7.6	5	1		.020	.202	.203	-.083	4.73	4.73	.058	4.50	4.50
8.9	4	1	1	-.007	.190	.190	.162	4.27	4.29	-.252	3.94	4.01
7.6	5	1		.027	.264	.265	-.115	6.49	6.51	.099	6.32	6.33
8.1	4	1	1	-.025	.236	.237	.357	5.28	5.41	-.513	4.79	5.06
6.7	5	5		.017	.127	.127	-.064	2.97	2.97	.044	2.84	2.84
7.9	4	5	5/4	-.001	.126	.126	.131	2.84	2.86	-.188	2.63	2.66
6.7	5	5		.021	.161	.161	-.093	4.09	4.10	.082	4.06	4.06
7.2	4	5	5/4	-.007	.166	.167	.288	3.94	4.02	-.370	3.70	3.84
6.7	5	5		.044	.840	.842	-.170	19.0	19.1	.120	17.8	17.8
8.9	4	5	0.5	-.022	.789	.789	.333	17.4	17.5	-.463	15.9	16.1
6.7	5	5		.076	1.14	1.15	-.303	25.9	25.9	.246	24.2	24.2
7.2	4	5	0.5	-.069	.971	.976	.782	21.2	21.8	-1.00	19.2	20.2
6.7	1	5		.016	.085	.086	-.064	1.83	1.84	.047	1.68	1.68
7	2	4	5/3	-.002	.081	.081	.103	1.69	1.70	-.142	1.51	1.53
6.7	1	5		.031	.127	.128	-.128	2.62	2.63	.104	2.37	2.38
6.5	2	4	5/3	-.006	.110	.110	.236	2.16	2.21	-.307	1.90	1.99
6.7	1	5		.027	.332	.332	-.103	7.48	7.49	.073	7.04	7.04
7.9	2	3	0.8	-.033	.305	.306	.336	6.63	6.75	-.441	6.04	6.24
6.7	1	5		.054	.463	.466	-.241	10.8	10.8	.212	10.5	10.5
6.8	2	3	0.8	-.077	.377	.383	.727	8.05	8.58	-.928	7.28	8.14

Table 2: *Results for the Buckley-James estimator (first line) and the new estimator (second line) for model (4.2).*

				$\hat{\beta}_0$			$\hat{\beta}_1$		
α_0	α_1	ϱ	γ	Bias	Var	MSE	Bias	Var	MSE
0.7	9.85	1	1	.085	.007	.014	.181	.049	.082
				.063	.006	.010	.135	.048	.066
1.5	9.5	2	2	.166	.027	.054	.366	.197	.331
				.100	.023	.033	-.266	.190	.260
2.4	10	4	3	.230	.061	.114	.475	.466	.692
				.119	.052	.066	-.326	.444	.550
2.6	10	4	5	.465	.172	.388	.967	1.20	2.13
				.201	.136	.177	-.569	1.16	1.48

Table 3: *Results for the Buckley-James estimator (first line) and the new estimator (second line) for model (4.3).*

end of the study (1 March 1981) 40 patients were alive, and their survival time was therefore censored to the right. We are interested in studying the relationship between $Y = \log(\text{survival time})$ and $X = \log(\text{age of the patient at diagnosis (in years)})$. The data shown in Figure 1 suggest that a linear model might be appropriate :

$$Y = \beta_0 + \beta_1 X + \varepsilon, \tag{5.1}$$

where ε and X are independent and $E(\varepsilon) = 0$. The Buckley-James (1979) algorithm and the new method yield respectively the values -1.03 and -0.97 for the slope parameter and 5.64 and 5.39 for the intercept parameter. It was observed that the Buckley-James method does not converge to a single value of the slope parameter, but oscillates between three values. The estimator is defined as the average of these values. For the new method, boundary corrected kernels are used. The bandwidth is selected from a grid of 16 bandwidths, according to the method described in Remark 3.4. From Figure 1 it is clear that the regression lines (and also the new data points) obtained from the Buckley-James method and the new method are very close to each other. By using the bootstrap method explained in Remark 3.5, the variance of the slope respectively intercept of the new method is given by 1.05 respectively 18.32. Confidence intervals obtained from the percentile bootstrap method are $[-3.10, 0.92]$ for the slope and $[-2.65, 14.00]$ for the intercept. The intervals obtained from the normal approximation are very similar, which suggests that the asymptotic normality result is accurate here.

The second data set 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 by means of the spectral index $\alpha_{ox} =$

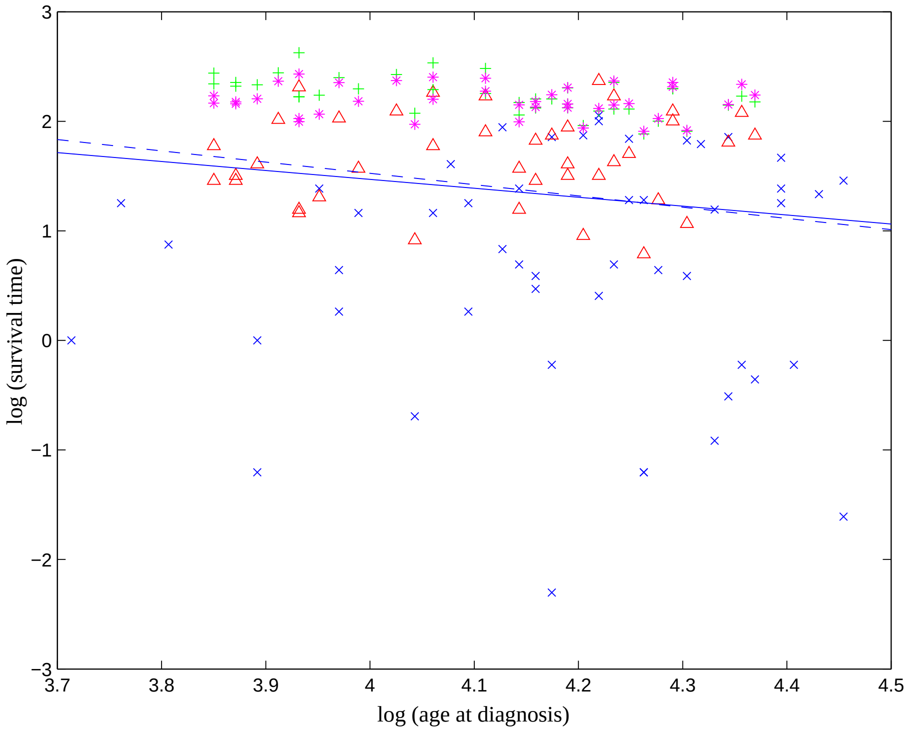

Figure 1: *Linear regression for the larynx cancer data. The solid, respectively, dashed line represents the estimated regression line for the new, respectively, Buckley-James method. Uncensored data points are given by* $\times$*, and censored observations by* $\triangle$*. The new data points obtained from the new, respectively, Buckley-James method are represented by* $*$*, respectively* $+$*.*

$0.384 \log(L_{2\ keV}/L_{2500\ \mathring{A}})$, 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) (see Vignali, Brandt and Schneider (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 Figure 2 the results of the regression of l_x on l_{uv} for both the new and the Buckley-James algorithm, assuming that model (5.1) is valid (where the latter is again obtained by taking the average of the values around which it oscillates). We observe a big similarity between the two regression lines. For both methods there is a strong correlation between the two variables. The slope and intercept are respectively 0.75 and 3.48 for the new method and 0.74 and 3.76 for the Buckley-James method. The variance of the slope and intercept for the new method equal 0.006 and 5.68 respectively, while the percentile

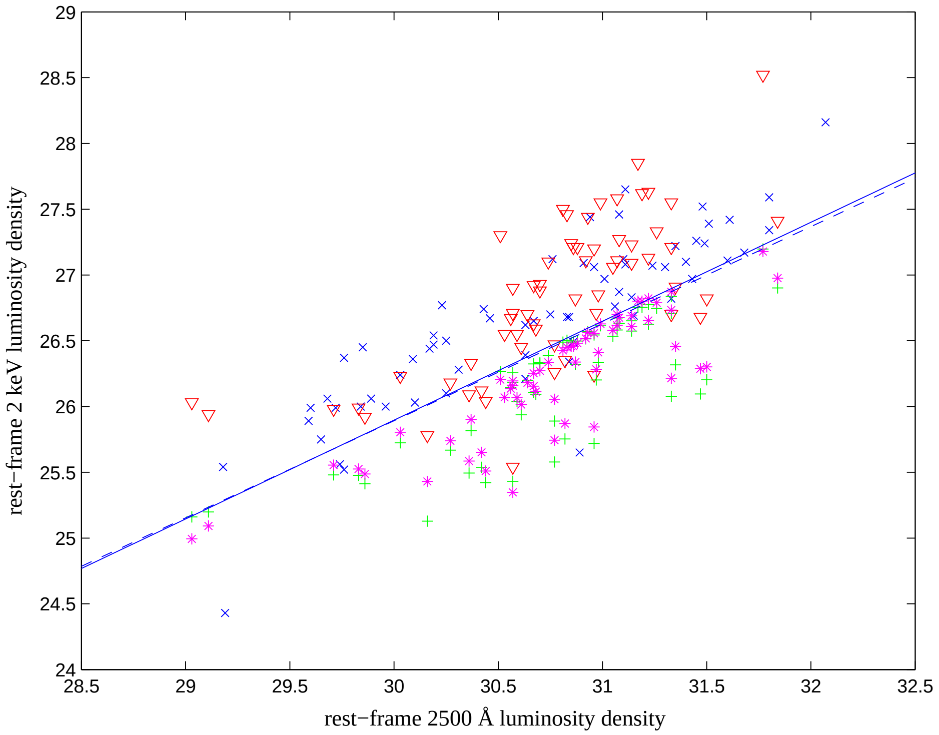

Figure 2: *Linear regression for the quasar data. The solid, respectively, dashed line represents the estimated regression line for the new, respectively, Buckley-James method. Uncensored data points are given by* $\times$*, and censored observations by* $\bigtriangledown$*. The new data points obtained from the new, respectively, Buckley-James method are represented by* $*$*, respectively* $+$*.*

bootstrap confidence intervals are given by $[0.52, 0.83]$ and $[0.98, 10.57]$ respectively.

Finally, note that direct comparison of the parametric estimator with the nonparametric estimator $\hat{m}(x)$ is not possible, since the latter function estimates $m(x)$ defined in (2.2) and the former estimates the conditional mean function. It would be interesting to compare the parametric estimator with a nonparametric estimator of the conditional mean. This can be done by means of the Beran estimator defined in (2.3). However, since the Beran estimator is inconsistent in the right tail, the so-obtained estimator of the conditional mean will be inconsistent. Alternatively, a more elaborated estimator can be used which makes use of the independence between ε and X to overcome these inconsistency problems.

Appendix : Proofs of main results

The following functions enter the asymptotic representation of $\hat{\beta}_{Tj} - \beta_{Tj}$ $(j = 0, \ldots, p)$, which we established in Section 3.

$$\xi_e(z, \delta, y) = (1 - F_e(y)) \left\{ - \int_{-\infty}^{y \wedge z} \frac{dH_{e1}(s)}{(1 - H_e(s))^2} + \frac{I(z \leq y, \delta = 1)}{1 - H_e(z)} \right\},$$

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

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

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

$$\gamma_1(y|x) = \int_{-\infty}^{y} \frac{h_e(s|x)}{(1 - H_e(s))^2}\, dH_{e1}(s) + \int_{-\infty}^{y} \frac{d\, h_{e1}(s|x)}{1 - H_e(s)},$$

$$\gamma_2(y|x) = \int_{-\infty}^{y} \frac{s h_e(s|x)}{(1 - H_e(s))^2}\, dH_{e1}(s) + \int_{-\infty}^{y} \frac{d\, (s h_{e1}(s|x))}{1 - H_e(s)},$$

$$\varphi(x, z, \delta, y) = \xi_e\left(\frac{z - m(x)}{\sigma(x)}, \delta, y\right) - S_e(y)\eta(z, \delta|x)\gamma_1(y|x) - S_e(y)\zeta(z, \delta|x)\gamma_2(y|x),$$

$$\alpha_i(v) = \frac{\int_v^{S_i} u\, dF_e(u)}{1 - F_e(v)},$$

$$\begin{aligned} B_k(z, Z_j, \Delta_j|X_i) = X_i^k f_X^{-1}(X_i)\sigma(X_i) \Bigg\{ & \left[\alpha_i'(e_i^T(z)) - 1 + \frac{S_i f_e(S_i)}{1 - F_e(e_i^T(z))} \right] \eta(Z_j, \Delta_j|X_i) \\ & + \left[e_i^T(z)\alpha_i'(e_i^T(z)) - \alpha_i(e_i^T(z)) + \frac{S_i^2 f_e(S_i)}{1 - F_e(e_i^T(z))} \right] \zeta(Z_j, \Delta_j|X_i) \Bigg\}, \end{aligned}$$

$(k = 0, \ldots, p; i, j = 1, \ldots, n)$ where $S_i = S_{X_i}$, $e_i^T(z) = e_{X_i}^T(z)$ and for any $x \in R_X$, $S_x = (T_x - m(x))/\sigma(x)$ and $e_x^T(z) = (z \wedge T_x - m(x))/\sigma(x)$.

Let $\tilde{T}_x$ be any value less than the upper bound of the support of $H(\cdot|x)$ such that $\inf_{x \in R_X}(1 - H(\tilde{T}_x|x)) > 0$. 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 compact, convex and its interior is not empty.

(iii) K is a density with compact support, $\int uK(u)du = 0$ and K is twice continuously differentiable.
(iv) $det(M) \neq 0$.

$(A2)(i)$ There exist $0 \leq s_0 \leq s_1 \leq 1$ such that $s_1 \leq \inf_x F(\tilde{T}_x|x)$, $s_0 \leq \inf\{s \in [0,1]; J(s) \neq 0\}$, $s_1 \geq \sup\{s \in [0,1]; J(s) \neq 0\}$ and $\inf_{x \in R_X} \inf_{s_0 \leq s \leq s_1} f(F^{-1}(s|x)|x) > 0$.
(ii) J is twice continuously differentiable, $\int_0^1 J(s)ds = 1$ and $J(s) \geq 0$ for all $0 \leq s \leq 1$.
(iii) The function $x \to T_x$ $(x \in R_X)$ is twice continuously differentiable.

$(A3)(i)$ F_X is three times continuously differentiable and $\inf_{x \in R_X} f_X(x) > 0$.
(ii) m and σ are twice continuously differentiable and $\inf_{x \in R_X} \sigma(x) > 0$.
(iii) In the model $Y = m(X) + \sigma(X)\varepsilon$, $E[\varepsilon^2] < \infty$ and $E[E^4] < \infty$.

$(A4)(i)$ $\eta(z, \delta|x)$ 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.
(ii) The first derivatives of $\eta(z, \delta|x)$ and $\zeta(z, \delta|x)$ with respect to z are of bounded variation and the variation norms are uniformly bounded over all x.

$(A5)$ The function $y \to P(m(X) + e\sigma(X) \leq y)$ $(y \in I\!R)$ is differentiable for all $e \in I\!R$ and the derivative is uniformly bounded over all $e \in I\!R$.

$(A6)$ For $L(y|x) = H(y|x), H_1(y|x), H_e(y|x)$ or $H_{e1}(y|x)$: $L'(y|x)$ is continuous in (x, y) and $\sup_{x,y} |y^2 L'(y|x)| < \infty$, the same holds for all other partial derivatives of $L(y|x)$ with respect to x and y up to order three, and $\sup_{x,y} |y^3 L'''(y|x)| < \infty$.

$(A7)$ (i) $\sup_{x,z} \int |B'_k(t, z, \delta|x)| h(t)dt < \infty$ $(k = 0, \ldots, p;\ \delta = 0, 1)$.
(ii) $\sup_z \int \sup_x |B''_k(t, z, \delta|x)| h(t)dt < \infty$ $(k = 0, \ldots, p;\ \delta = 0, 1)$, where $B_k^{'(')}(t, z, \delta|x)$ equals the first (second) derivative of $B_k(t, z, \delta|x)$ with respect to x when $t \neq T_x$ and equals 0 otherwise.

$(A8)$ For the density $f_{X|Z,\Delta}(x|z, \delta)$ of X given (Z, Δ), $\sup_{x,z} |f_{X|Z,\Delta}(x|z, \delta)| < \infty$, $\sup_{x,z} |\dot{f}_{X|Z,\Delta}(x|z, \delta)| < \infty$, $\sup_{x,z} |\ddot{f}_{X|Z,\Delta}(x|z, \delta)| < \infty$ $(\delta = 0, 1)$.

Proof of Theorem 3.1. Let $\beta^*_T = (\beta^*_{T0}, \ldots, \beta^*_{Tp})$ be the least squares estimator obtained from the pairs (X_i, Y^*_{Ti}) $(i = 1, \ldots, n)$. We will first consider

$$\hat{\beta}_T - \beta^*_T = (n^{-1}\mathcal{X}'\mathcal{X})^{-1} n^{-1} \mathcal{X}'(\hat{\mathcal{Y}}^* - \mathcal{Y}^*),$$

where $\mathcal{Y}^* = (Y^*_{T1}, \ldots, Y^*_{Tn})'$, and $\hat{\mathcal{Y}}^* = (\hat{Y}^*_{T1}, \ldots, \hat{Y}^*_{Tn})'$. The $(k+1)^{th}$ element $(k = 0, \ldots, p)$ of the vector $n^{-1}\mathcal{X}'(\hat{\mathcal{Y}}^* - \mathcal{Y}^*)$ equals

$$n^{-1} \sum_{\Delta_i = 0} X_i^k \left\{ [\hat{m}(X_i) - m(X_i)] + \frac{\hat{\sigma}(X_i)}{1 - \hat{F}_e(\hat{E}_i^T)} \int_{\hat{E}_i^T}^{\hat{S}_i} u\, d\hat{F}_e(u) - \frac{\sigma(X_i)}{1 - F_e(E_i^T)} \int_{E_i^T}^{S_i} u\, dF_e(u) \right\}$$

$$= n^{-1} \sum_{\Delta_i=0} X_i^k \{A_{1i} + A_{2i} + A_{3i}\}.$$

The asymptotic representation given in Proposition 4.8 of Van Keilegom and Akritas (1999) (hereafter abbreviated by VKA) yields

$$A_{1i} = -(na_n)^{-1} f_X^{-1}(X_i)\sigma(X_i) \sum_{j=1}^n K(\frac{X_i - X_j}{a_n})\eta(Z_j, \Delta_j | X_i) + o_P(n^{-1/2}),$$

uniformly in $i = 1, \ldots, n$. Next, write

$$\begin{aligned} & n^{-1} \sum_{\Delta_i=0} X_i^k \{A_{1i} + A_{2i} + A_{3i}\} \qquad \text{(A.1)} \\ &= n^{-1} \sum_{\Delta_i=0} X_i^k \{A_{1i} + A_{2i} + A_{3i}\} I(E_i \le U_n) + n^{-1} \sum_{\Delta_i=0} X_i^k \{A_{1i} + A_{2i} + A_{3i}\} I(E_i > U_n), \end{aligned}$$

where $U_n < 0$ is defined by $U_n = -n^{1/2} a_n^{1+\gamma}$ for some $\gamma > 0$ to be determined later. First, let us show that the first sum of this expression is asymptotically negligible. Let V_n be the number of residuals E_i that are less than or equal to U_n. Then, by the law of the iterated logarithm (see e.g. Serfling (1980), page 35),

$$V_n - nH_e(U_n) \le 2[H_e(U_n)(1 - H_e(U_n))n \log\log n]^{1/2} \quad a.s..$$

Since $|U_n|^4 H_e(U_n) \le \int_{-\infty}^{U_n} |y|^4 \, dH_e(y) \to 0$, it follows that $H_e(U_n) \le C_n |U_n|^{-4}$ for some sequence $C_n \to 0$. From this, we have that $V_n = o(n|U_n|^{-4} + |U_n|^{-2} n^{1/2} (\log\log n)^{1/2})$ a.s. Next, $A_{1i} + A_{2i} + A_{3i}$ is bounded in probability, which follows from Lemma A.1, the fact that $E|\varepsilon| < \infty$, the uniform consistency of $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$ given by Proposition 4.5 in VKA (1999) and the consistency of $\sup_{x,z} |\hat{F}_e(\frac{z \wedge T_x - \hat{m}(x)}{\hat{\sigma}(x)}) - F_e(\frac{z \wedge T_x - m(x)}{\sigma(x)})|$ which is obtained as follows.

$$\begin{aligned} \hat{F}_e(\frac{z \wedge T_x - \hat{m}(x)}{\hat{\sigma}(x)}) - F_e(\frac{z \wedge T_x - m(x)}{\sigma(x)}) &= \hat{F}_e(\frac{z \wedge T_x - \hat{m}(x)}{\hat{\sigma}(x)}) - F_e(\frac{z \wedge T_x - \hat{m}(x)}{\hat{\sigma}(x)}) \\ &+ F_e(\frac{z \wedge T_x - \hat{m}(x)}{\hat{\sigma}(x)}) - F_e(\frac{z \wedge T_x - m(x)}{\hat{\sigma}(x)}) \\ &+ F_e(\frac{z \wedge T_x - m(x)}{\hat{\sigma}(x)}) - F_e(\frac{z \wedge T_x - m(x)}{\sigma(x)}) \\ &= \alpha_n^1(z, x) + \alpha_n^2(z, x) + \alpha_n^3(z, x). \end{aligned}$$

Using Corollary 3.2 of VKA (1999), $\sup_{x,z} |\alpha_n^1(z, x)|$ is $O_p(n^{-1/2})$. For the two other terms, we use two first order Taylor developments

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

for some A_x (B_x) between $\frac{z\wedge T_x-m(x)}{\hat{\sigma}(x)}$ and $\frac{z\wedge T_x-\hat{m}(x)}{\hat{\sigma}(x)}$ ($\frac{z\wedge T_x-m(x)}{\sigma(x)}$ and $\frac{z\wedge T_x-m(x)}{\hat{\sigma}(x)}$). Using Proposition 4.5 of VKA (1999) and the fact that $\sup_e |ef_e(e)| < +\infty$, $\alpha_n^2(z,x)+\alpha_n^3(z,x) = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ *a.s.* Therefore, the first term on the right hand side of (A.1) is $o_P(|U_n|^{-4}) = o_P(n^{-1/2})$ for γ small enough. We next consider the second term on the right hand side of (A.1). Write

$$\begin{aligned}
A_{2i}+A_{3i} &= \frac{\hat{\sigma}(X_i)-\sigma(X_i)}{1-\hat{F}_e(\hat{E}_i^T)}\int_{\hat{E}_i^T}^{\hat{S}_i} u\,d\hat{F}_e(u) + \sigma(X_i)\frac{\hat{F}_e(\hat{E}_i^T)-F_e(E_i^T)}{(1-\hat{F}_e(\hat{E}_i^T))(1-F_e(E_i^T))}\int_{\hat{E}_i^T}^{\hat{S}_i} u\,d\hat{F}_e(u)\\
&\quad+\frac{\sigma(X_i)}{1-F_e(E_i^T)}\int_{\hat{E}_i^T}^{E_i^T} u\,d\hat{F}_e(u) + \frac{\sigma(X_i)}{1-F_e(E_i^T)}\int_{E_i^T}^{S_i} u\,d(\hat{F}_e(u)-F_e(u))\\
&\quad+\frac{\sigma(X_i)}{1-F_e(E_i^T)}\int_{S_i}^{\hat{S}_i} u\,d\hat{F}_e(u)\\
&=\sum_{j=1}^{5} B_{ji}
\end{aligned}$$

First consider

$$\begin{aligned}
\int_{\hat{E}_i^T}^{\hat{S}_i} u\,d\hat{F}_e(u) &= \int_{E_i^T}^{S_i} u\,dF_e(u) + \int_{E_i^T}^{S_i} u\,d(\hat{F}_e(u)-F_e(u)) \qquad \text{(A.2)}\\
&\quad+\int_{\hat{E}_i^T}^{E_i^T} u\,d\hat{F}_e(u) + \int_{S_i}^{\hat{S}_i} u\,d\hat{F}_e(u),
\end{aligned}$$

which appears in B_{1i} and B_{2i}. By using integration by parts, the third term of (A.2) can be rewritten as

$$\begin{aligned}
&[E_i^T(\hat{F}_e(E_i^T)-F_e(E_i^T))] + [E_i^T F_e(E_i^T) - (\hat{E}_i^T)F_e(E_i^T)]\\
&+[(\hat{E}_i^T)(F_e(E_i^T)-\hat{F}_e(\hat{E}_i^T))] - \int_{\hat{E}_i^T}^{E_i^T} \hat{F}_e(u)du. \qquad \text{(A.3)}
\end{aligned}$$

By Corollary 3.2 in VKA (1999) and the order of U_n, the first term of (A.3) is $O_P(a_n^{1+\gamma})$, while from Proposition 4.5 in VKA (1999) it follows that the second and fourth term are $O_P(a_n^{1/2+\gamma}(\log a_n^{-1})^{1/2})$. Using the fact that $\sup_{x,z}|\hat{F}_e(\frac{z\wedge T_x-\hat{m}(x)}{\hat{\sigma}(x)}) - F_e(\frac{z\wedge T_x-m(x)}{\sigma(x)})| = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ yields that the order of the third term is $O_P(a_n^{1/2+\gamma}(\log a_n^{-1})^{1/2})$. Hence, the third term of (A.2) is $O_P(a_n^{1/2+\gamma}(\log a_n^{-1})^{1/2})$, uniformly in $i = 1,\dots,n$. In a similar way it can be shown that the second and fourth term of (A.2) are of this order, which implies that

$$B_{1i}+B_{2i} = \frac{\hat{\sigma}(X_i)-\sigma(X_i)}{1-F_e(E_i^T)}\int_{E_i^T}^{S_i} u\,dF_e(u) + \sigma(X_i)\frac{\hat{F}_e(\hat{E}_i^T)-F_e(E_i^T)}{(1-F_e(E_i^T))^2}\int_{E_i^T}^{S_i} u\,dF_e(u) + o_P(n^{-1/2}).$$

$B_{1i}+B_{2i}$ can now be written as a sum of i.i.d. terms (up to the $o_P(n^{-1/2})$ remainder term), by applying the representation for $\hat{\sigma}(X_i)-\sigma(X_i)$ given by Proposition 4.9 in VKA

(1999) and using the fact that

$$\hat{F}_e(\hat{E}_i^T) - F_e(E_i^T) = (na_n)^{-1} \sum_{j=1}^n K(\frac{X_i - X_j}{a_n}) f_X^{-1}(X_i)[\eta(Z_j, \Delta_j|X_i) + \zeta(Z_j, \Delta_j|X_i) E_i^T] f_e(E_i^T)$$

$$+n^{-1} \sum_{j=1}^n \varphi(X_j, Z_j, \Delta_j, E_i^T) + o_P(n^{-1/2}), \qquad \text{(A.4)}$$

where this development is obtained after two Taylor expansions and by applying Theorem 3.1, Lemma B.1 and Propositions 4.8 and 4.9 in VKA (1999). For B_{3i} write

$$\int_{\hat{E}_i^T}^{E_i^T} u \, d\hat{F}_e(u) = \int_{\hat{E}_i^T}^{E_i^T} u \, dF_e(u) + \int_{\hat{E}_i^T}^{E_i^T} u \, d(\hat{F}_e(u) - F_e(u)).$$

Integrating by parts the second term of the expression above and using Corollary 3.2 in VKA (1999) and the fact that $|\hat{E}_i^T - E_i^T| = O_P(a_n^{1/2+\gamma}(\log a_n^{-1})^{1/2})$, we obtain

$$E_i^T[\hat{F}_e(E_i^T) - F_e(E_i^T) - \hat{F}_e(\hat{E}_i^T) + F_e(\hat{E}_i^T)] - \int_{\hat{E}_i^T}^{E_i^T} (\hat{F}_e(u) - F_e(u)) \, du + o_P(n^{-1/2}).$$

It is easy to see that the integral in this expression is also $o_P(n^{-1/2})$. As a consequence of Theorem 3.1 and Lemma B.1 in VKA (1999), the first term is $o_P(|E_i^T|n^{-1/2})$. Hence,

$$B_{3i} = -\frac{\sigma(X_i)}{1 - F_e(E_i^T)} \Big[\int_0^{\hat{E}_i^T} u \, dF_e(u) - \int_0^{E_i^T} u \, dF_e(u) \Big] + o_P(|E_i^T|n^{-1/2})$$

$$= [\hat{m}(X_i) - m(X_i) + E_i^T(\hat{\sigma}(X_i) - \sigma(X_i))] \frac{E_i^T f_e(E_i^T)}{1 - F_e(E_i^T)} + o_P(|E_i^T|n^{-1/2}) + o_P(n^{-1/2})$$

using a Taylor expansion. Note that the term $o_P(n^{-1/2})$ in the expression above is obtained from the fact that $\sup_z |zf_e(z)| < \infty$ and $\sup_z |z^2 f_e'(z)| < \infty$. Next, the term B_{4i} is given by

$$B_{4i} = \frac{\sigma(X_i)}{1 - F_e(E_i^T)} \left\{ S_i[\hat{F}_e(S_i) - F_e(S_i)] - E_i^T[\hat{F}_e(E_i^T) - F_e(E_i^T)] - \int_{E_i^T}^{S_i} (\hat{F}_e(u) - F_e(u)) \, du \right\}.$$

Finally, the term B_{5i} is treated in the same way as the term B_{3i}, leading to

$$B_{5i} = -[\hat{m}(X_i) - m(X_i) + S_i(\hat{\sigma}(X_i) - \sigma(X_i))] \frac{S_i f_e(S_i)}{1 - F_e(E_i^T)} + o_P(|S_i|n^{-1/2}) + o_P(n^{-1/2}).$$

It now follows that the complete asymptotic representation for the $(k+1)^{th}$ component of $n^{-1}\mathcal{X}'(\hat{\mathcal{Y}}^* - \mathcal{Y}^*)$ can be written as

$$n^{-1} \sum_{\Delta_i = 0} X_i^k I(E_i > U_n) \left\{ \left[-\frac{E_i^T f_e(E_i^T)}{1 - F_e(E_i^T)} + \frac{f_e(E_i^T) \int_{E_i^T}^{S_i} u \, dF_e(u)}{(1 - F_e(E_i^T))^2} - 1 + \frac{S_i f_e(S_i)}{1 - F_e(E_i^T)} \right] \right.$$

$$
\begin{aligned}
&\times (na_n)^{-1} f_X^{-1}(X_i)\sigma(X_i)\sum_{j=1}^n K\Big(\frac{X_i - X_j}{a_n}\Big)\eta(Z_j, \Delta_j|X_i) \\
&+ \left[\frac{E_i^T f_e(E_i^T)\int_{E_i^T}^{S_i} u\, dF_e(u)}{(1 - F_e(E_i^T))^2} - \frac{\int_{E_i^T}^{S_i} u\, dF_e(u)}{1 - F_e(E_i^T)} - \frac{(E_i^T)^2 f_e(E_i^T)}{1 - F_e(E_i^T)} + \frac{S_i^2 f_e(S_i)}{1 - F_e(E_i^T)}\right] \\
&\quad \times (na_n)^{-1} f_X^{-1}(X_i)\sigma(X_i)\sum_{j=1}^n K\Big(\frac{X_i - X_j}{a_n}\Big)\zeta(Z_j, \Delta_j|X_i) \\
&+\sigma(X_i)\left[\frac{\int_{E_i^T}^{S_i} u\, dF_e(u)}{(1 - F_e(E_i^T))^2} - \frac{E_i^T}{1 - F_e(E_i^T)}\right] n^{-1}\sum_{j=1}^n \varphi(X_j, Z_j, \Delta_j, E_i^T) \\
&+\frac{\sigma(X_i)S_i}{1 - F_e(E_i^T)} n^{-1}\sum_{j=1}^n \varphi(X_j, Z_j, \Delta_j, S_i) \\
&- \frac{\sigma(X_i)}{1 - F_e(E_i^T)} n^{-1}\sum_{j=1}^n \int_{E_i^T}^{S_i} \varphi(X_j, Z_j, \Delta_j, u)du \Bigg\} + o_P(n^{-1/2}), \qquad \text{(A.5)}
\end{aligned}
$$

where use is made of the representations for $\hat{F}_e(\cdot), \hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$ given by Theorem 3.1 and Propositions 4.8 and 4.9 in VKA (1999) respectively, and of the representation for $\hat{F}_e(\hat{E}_i^T) - F_e(E_i^T)$ given in (A.4).
We can rewrite the sum of the first two terms of equation (A.5) as

$$
(n^2 a_n)^{-1}\sum_{j\neq i}(1 - \Delta_i)I(E_i > U_n)B_k(Z_i, Z_j, \Delta_j|X_i)K(\frac{X_i - X_j}{a_n}) + o_P(n^{-1/2}). \qquad \text{(A.6)}
$$

Using a similar development as for the first term of (A.1) it is easily shown that (A.6) can be written as

$$
\begin{aligned}
&(n^2 a_n)^{-1}\sum_{j\neq i}(1 - \Delta_i)B_k(Z_i, Z_j, \Delta_j|X_i)K(\frac{X_i - X_j}{a_n}) + o_P(n^{-1/2}) \\
&= (n^2 a_n)^{-1}\sum_{j\neq i}\{A_k^*(V_i, V_j) + E[A_k(V_i, V_j)|V_i] + E[A_k(V_i, V_j)|V_j] - E[A_k(V_i, V_j)]\} \\
&\quad + o_P(n^{-1/2}) \\
&= T_1 + T_2 + T_3 + T_4 + o_P(n^{-1/2}),
\end{aligned}
$$

where

$$
A_k(V_i, V_j) = (1 - \Delta_i)B_k(Z_i, Z_j, \Delta_j|X_i)K(\frac{X_i - X_j}{a_n}),
$$

$A_k^*(V_i, V_j) = A_k(V_i, V_j) - E[A_k(V_i, V_j)|V_i] - E[A_k(V_i, V_j)|V_j] + E[A_k(V_i, V_j)]$ and $V_i = (X_i, Z_i, \Delta_i)$. Consider

$$
E[A_k(V_i, V_j)|V_i]
$$

$$
\begin{aligned}
&= (1-\Delta_i)\sum_{\delta=0,1}\int\int B_k(Z_i,z,\delta|X_i)K(\frac{X_i-x}{a_n})h_\delta(z|x)f_X(x)\,dz\,dx\\
&= a_n(1-\Delta_i)\sum_{\delta=0,1}\int\int B_k(Z_i,z,\delta|X_i)K(u)(h_\delta(z|X_i)-a_nu\dot{h}_\delta(z|X_i)+O(a_n^2))\\
&\qquad \times(f_X(X_i)-a_nuf'_X(X_i)+O(a_n^2))\,dz\,du\\
&= a_n(1-\Delta_i)f_X(X_i)\sum_{\delta=0,1}\int B_k(Z_i,z,\delta|X_i)h_\delta(z|X_i)\,dz+O(a_n^3)=O(a_n^3),
\end{aligned}
$$

since $E[\eta(Z,\Delta|X)|X]=E[\zeta(Z,\Delta|X)|X]=0$, where $\dot{h}_\delta(z|x)$ denotes the derivative of $h_\delta(z|x)$ with respect to x. Hence, we also have that $E[A_k(V_i,V_j)]=O(a_n^3)$. In a similar way we have for $E[A_k(V_i,V_j)|V_j]$, using three Taylor expansions of order 2,

$$
E[A_k(V_i,V_j)|V_j]=a_nf_X(X_j)\sum_{\delta=0,1}(1-\delta)\int B_k(z,Z_j,\Delta_j|X_j)\,dH_\delta(z|X_j)+O(a_n^3).
$$

It follows that

$$
T_2+T_3+T_4=n^{-1}\sum_{i=1}^n f_X(X_i)\int B_k(z,Z_i,\Delta_i|X_i)\,dH_0(z|X_i)+O(a_n^2).
$$

For T_1, note that $E[T_1]=0$ and hence, by Chebyshev's inequality,

$$
\begin{aligned}
&P(|T_1|>K(na_n)^{-1}E[A_k^*(V_1,V_2)^2]^{1/2})\\
&\le K^{-2}(na_n)^2E[A_k^*(V_1,V_2)^2]^{-1}E[T_1^2]\\
&= K^{-2}n^{-2}E[A_k^*(V_1,V_2)^2]^{-1}\sum_{j\neq i}\sum_{m\neq l}E[A_k^*(V_i,V_j)A_k^*(V_l,V_m)]. \qquad \text{(A.7)}
\end{aligned}
$$

Since $E[A_k^*(V_i,V_j)]=0$, 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 $j\neq m$,

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

Thus, only the $2n(n-1)$ terms for which (i,j) equals (l,m) or (m,l) stay such that, (A.7) is bounded by $2K^{-2}$, which can be made arbitrarily small for K large enough. Since $A_k^*(V_1,V_2)$ is bounded by $K(\frac{X_1-X_2}{a_n})C+O(a_n)$ for some constant $C>0$, independent of X_1 and X_2, we have that $E[A_k^*(V_1,V_2)^2]\le C^2a_n\int f_X^2(x)\,dx\int K^2(u)\,du+O(a_n^2)=O(a_n)$ (and similarly for $E[A_k^*(V_1,V_2)A_k^*(V_2,V_1)]$). It now follows that $T_1=O_P(n^{-1}a_n^{-1/2})=o_P(n^{-1/2})$.

We next consider the third, fourth and fifth term of (A.5). Their sum equals

$$
n^{-1}\sum_{\Delta_i=0}I(E_i>U_n)X_i^k\left\{\frac{\int_{E_i^T}^{S_i}u\,dF_e(u)}{(1-F_e(E_i^T))^2}n^{-1}\sigma(X_i)\sum_{j=1}^n\varphi(X_j,Z_j,\Delta_j,E_i^T)\right.
$$

$$+ \frac{\sigma(X_i)}{1-F_e(E_i^T)} \int_{E_i^T}^{S_i} u\, n^{-1} \sum_{j=1}^n d\varphi(X_j, Z_j, \Delta_j, u) \Bigg\}$$

$$= n^{-2} \sum_{j\neq i} h_k(V_i, V_j) + o_P(n^{-1/2}), \tag{A.8}$$

where

$$h_k(V_i, V_j) = (1-\Delta_i) X_i^k \left\{ \frac{\int_{E_i^T}^{S_i} u\, dF_e(u)}{(1-F_e(E_i^T))^2} \sigma(X_i) \varphi(X_j, Z_j, \Delta_j, E_i^T) \right.$$

$$\left. + \frac{\sigma(X_i)}{1-F_e(E_i^T)} \int_{E_i^T}^{S_i} u\, d\varphi(X_j, Z_j, \Delta_j, u) \right\},$$

using arguments similar as before. Defining $h_k^*(V_i, V_j) = h_k(V_i, V_j) + h_k(V_j, V_i)$, (A.8) can be written as

$$\frac{n-1}{2n} \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{j>i} h_k^*(V_i, V_j) + o_P(n^{-1/2}).$$

Using the Hájek-projection of a U-statistic on its conditional expectations (see e.g. Serfling (1980), page 189), this expression equals

$$n^{-1} \sum_{i=1}^n E[h_k^*(V_i, V_j)|V_i] + o_P(n^{-1/2})$$

$$= n^{-1} \sum_{i=1}^n \int x^k \sigma(x) \int \left\{ \frac{\varphi(X_i, Z_i, \Delta_i, e_x^T(z))}{(1-F_e(e_x^T(z)))^2} \int_{e_x^T(z)}^{S_x} u\, dF_e(u) \right.$$

$$\left. + \frac{1}{1-F_e(e_x^T(z))} \int_{e_x^T(z)}^{S_x} u\, d\varphi(X_i, Z_i, \Delta_i, u) \right\} dH_0(z|x) dF_X(x) + o_P(n^{-1/2}).$$

It remains to consider $\beta_T^* - \beta_T$, which equals

$$M^{-1} \begin{pmatrix} n^{-1} \sum_{i=1}^n (Y_{Ti}^* - E[Y_{Ti}^*|X_i]) \\ \vdots \\ n^{-1} \sum_{i=1}^n X_i^p (Y_{Ti}^* - E[Y_{Ti}^*|X_i]) \end{pmatrix} + \begin{pmatrix} o_P(n^{-1/2}) \\ \vdots \\ o_P(n^{-1/2}) \end{pmatrix},$$

using standart arguments. This finishes the proof.

Lemma A.1 *Assume (A1)(i) − (iii), (A2)(i), (ii), (A3)(ii), F_X is twice continuously differentiable , $\inf_{x\in R_X} f_X(x) > 0$, for $L(y|x) = H(y|x)$ or $H_1(y|x)$, $L(y|x)$ is continuous, $\dot{L}(y|x)$ and $\ddot{L}(y|x)$ exist, are continuous in (x, y), $\sup_{x,y} |y\dot{L}(y|x)| < \infty$ and $\sup_{x,y} |y^2\ddot{L}(y|x)| < \infty$, $H_e'(y|x)$ exists, is continuous in (x, y), $\sup_{x,y} |yH_e'(y|x)| < \infty$ and $E[|\varepsilon|] < \infty$. Then, for any $T < \tau_{H_e}$,*

$$\int_{-\infty}^T |u|\, d\hat{F}_e(u)$$

is bounded in probability.

Proof. We have for $T < \tau_{H_e}$,

$$\int_{-\infty}^{T} |u| d\hat{F}_e(u) = \sum_{\Delta_i=1} |\frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)}| \, W_i \, I(\frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)} \leq T),$$

where W_i are the Kaplan-Meier jumps of $\hat{F}_e$. First, let us show that the jumps of $\hat{F}_e$ are uniformly $O_P(n^{-1})$. It is easily seen that the jump of $\hat{F}_e$ at the j-th order statistic of $\hat{E}_i = (Y_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$ $(i = 1, \ldots, n)$ is bounded by $(n - j + 1)^{-1} \leq (n - a + 1)^{-1}$, where a is the number of $\hat{E}_i$'s smaller than or equal to T. From Proposition A.3 in VKA (1999) we know that

$$\hat{H}_e(T) = H_e(T) + o_P(1),$$

where $\hat{H}_e$ is the empirical distribution function of the $\hat{E}_i$'s. Thus, $a = nH_e(T) + o_P(n)$ and $(n - a + 1)^{-1} = O_P(n^{-1})$. It follows that, for n sufficiently large,

$$\begin{aligned}\sum_{\Delta_i=1} |\frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)}| W_i I(\frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)} \leq T) &\leq O_P(n^{-1}) \sum_{\Delta_i=1} |\frac{Y_i - m(X_i)}{\sigma(X_i)}| + o_P(1) \\ &= O_P(1) \int |y| d\tilde{H}_{e1}(y) + o_P(1) \\ &= O_P(1)[\int |y| dH_{e1}(y) + o_P(1)] + o_P(1) \\ &= O_P(1),\end{aligned}$$

where $\tilde{H}_{e1}(y) = n^{-1} \sum_{i=1}^{n} I(\frac{Y_i - m(X_i)}{\sigma(X_i)} \leq y, \Delta_i = 1)$, and provided that $\int |y| dH_{e1} = E[|\varepsilon| I(\Delta = 1)] \leq E[|\varepsilon|] < \infty$.

Acknowledgments. The authors like to thank Michael Akritas for helpful discussions and valuable input in this work.

References

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

Akritas, M. G. (1996). On the use of nonparametric regression techniques for fitting parametric regression models. *Biometrics*, **52**, 1342–1362.

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

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

Fygenson, M. and Zhou, M. (1994). On using stratification in the analysis of linear regression models with right censoring. *Ann. Statist.*, **22**, 747–762.

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

Kardaun, O. (1983). Statistical survival analysis of male larynx-cancer patients - a case study. *Statist. Neerlandica*, **37**, 103–125.

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

Lai, T. and Ying, Z. (1991). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Ann. Statist.*, **19**, 1370–1402.

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.

Miller, R.G. (1976). Least squares regression with censored data. *Biometrika*, **63**, 449–464.

Müller, H.-G. and Wang, J.-L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, **50**, 61–76.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.*, **18**, 303–328.

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

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multiv. Analysis*, **45**, 89–103.

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 Akritas, M. G. (2000). The least squares method in heteroscedastic censored regression models. In : *Asymptotics in Statistics and Probability*, Ed. M.L.

Puri, VSP, 379–391.

Van Keilegom, I. (2003). A note on the nonparametric estimation of the bivariate distribution under dependent censoring. *J. Nonpar. Statist.*, **16**, 659-670.

Vignali, C., Brandt, W.N. and 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.

Zhou, M. (1992). Asymptotic normality of the 'synthetic data' regression estimator for censored survival data. *Ann. Statist.*, **20**, 1002–1021.

Address for correspondence :

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