T E C H N I C A L
R E P O R T

0520

NONLINEAR 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

Nonlinear 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

April 21, 2005

Abstract

Suppose the random vector (X, Y) satisfies the regression model $Y = m(X) + \sigma(X)\varepsilon$, where $m(\cdot) = E(Y|\cdot)$ belongs to some parametric class $\{m_\theta(\cdot) : \theta \in \Theta\}$ of regression functions, $\sigma^2(\cdot) = \mathrm{Var}(Y|\cdot)$ is unknown, and ε is independent of X. The response Y is subject to random right censoring, and the covariate X is completely observed. A new estimation procedure for the true, unknown parameter vector θ_0 is proposed, that extends the classical least squares procedure for nonlinear regression to the case where the response is subject to censoring. The consistency and asymptotic normality of the proposed estimator are established. The estimator is compared via simulations with an estimator proposed by Stute in 1999, and both methods are also applied to a fatigue life data set of strain-controlled materials.

KEY WORDS: Bandwidth selection; Bootstrap; Fatigue life data; Kernel method; Least squares estimation; Nonparametric regression; Right censoring; Survival analysis.

[1] This research was supported by IAP research network grant nr. P5/24 of the Belgian government (Belgian Science Policy).

1 Introduction

Consider the heteroscedastic regression model

$$Y = m_{\theta_0}(X) + \sigma(X)\varepsilon, \tag{1.1}$$

where $\sigma^2(\cdot) = \mathrm{Var}(Y|\cdot)$, $m_{\theta_0}(\cdot) = E(Y|\cdot)$ is the regression curve, known upto a parameter vector $\theta \in \Theta$ with true unknown value θ_0, Θ is a compact subset of $I\!R^d$, and the error term ε is independent of the (one-dimensional) covariate X. Suppose that Y is subject to random right censoring, i.e. instead of observing Y, we only observe (Z, Δ), where $Z = \min(Y, C)$, $\Delta = I(Y \le C)$ and the random variable C represents the censoring time, which is 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 are interested in the estimation of the parameter vector θ_0 by means of an extension to censored data of the classical least squares procedure for nonlinear regression. When the regression curve is polynomial, this estimation problem has been studied extensively in the literature, see e.g. Heuchenne and Van Keilegom (2004) for a literature overview. When the regression curve belongs however to some parametric, but nonlinear family of regression functions, much less research has been devoted to this problem. Stute (1999) proposed the following estimation procedure, which is an extension of his earlier paper (Stute (1993)) on linear regression : minimize

$$\sum_{i=1}^{n} W_{in}\{Z_{(i)} - m_\theta(X_{(i)})\}^2 \tag{1.2}$$

with respect to θ, where

$$W_{in} = \frac{\Delta_{(i)}}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\Delta_{(j)}}$$

is the jump size of the bivariate empirical distribution function $\hat{F}(x, y)$ proposed by Stute (1993) :

$$\hat{F}(x, y) = \sum_{i=1}^{n} W_{in} I(X_{(i)} \le x, Z_{(i)} \le y),$$

$Z_{(1)}, \ldots, Z_{(n)}$ are the order statistics of $Z_1, \ldots, Z_n$, and $X_{(i)}$ and $\Delta_{(i)}$ are the corresponding covariate and censoring indicator (note that for simplicity we have considered untied observations). The method is very easy to implement in practice and, unlike many of the estimation procedures for polynomial regression, it is not based on a, sometimes delicate, tuning or bandwidth parameter. A major drawback of this estimation procedure however, is that it assumes that (1) Y and C are independent (unconditionally on X) and that

(2) $P(Y \leq C|X, Y) = P(Y \leq C|Y)$, which is satisfied when e.g. C is independent of X. Both assumptions are often violated in practice.

In this paper we propose a new estimation method for θ_0, which does not require the above two assumptions. The idea of the method is as follows :

1. Estimate the true unknown survival time of a censored observation $(X_i, Z_i, \Delta_i = 0)$ by a nonparametric estimator of $E(Y_i|X_i, Y_i > Z_i)$.
2. Estimate θ_0 by minimizing the least squares criterium for completely observed data, applied to the 'synthetic' data obtained in step (1).

For polynomial regression models a similar estimation procedure has been studied in Heuchenne and Van Keilegom (2004). The estimation in step (1) is done by using kernel smoothing with an adaptively chosen bandwidth parameter. The details of the proposed method are given in the next section.

The paper is organized as follows. In the next section, the estimation procedure is described in detail. Section 3 summarizes the main asymptotic results, including the asymptotic normality of the estimator. In Section 4 we present the results of a simulation study, in which the new procedure is compared with the method of Stute (1999). Section 5 is devoted to the analysis of data from a study on the relationship between fatigue life of metal and applied stress. The Appendix contains the proofs of the main results of Section 3.

2 Notations and description of the method

As outlined in the introduction, the idea of the proposed method consists of first estimating the unknown survival times of the censored observations, and second to apply a standard nonlinear least squares procedure on the so-obtained artificial data points. Define

$$Y_i^* = Y_i \Delta_i + E[Y_i | Y_i > C_i, X_i](1 - \Delta_i)$$

and note that $E(Y_i|X_i) = E(Y_i^*|X_i) = m_{\theta_0}(X_i)$. Hence, we can work in the sequel with the variable Y_i^* instead of with Y_i. In order to estimate Y_i^* for a censored observation, we first need to introduce a number of notations.

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 $X = x$). Let $\varepsilon^0 = (Y - m^0(X))/\sigma^0(X)$. Then, it can be easily seen that if model (1.1) holds (i.e. ε is independent of X), then ε^0 is also independent of X.

Define $F(y|x) = P(Y \le y|x)$, $G(y|x) = P(C \le y|x)$, $H(y|x) = P(Z \le y|x)$, $H_\delta(y|x) = P(Z \le y, \Delta = \delta|x)$, $F_X(x) = P(X \le x)$, $F_\varepsilon^0(y) = P(\varepsilon^0 \le y)$, $S_\varepsilon^0(y) = 1 - F_\varepsilon^0(y)$, and for $E^0 = (Z - m^0(X))/\sigma^0(X)$ we denote $H_\varepsilon^0(y) = P(E^0 \le y)$, $H_{\varepsilon\delta}^0(y) = P(E^0 \le y, \Delta = \delta)$, $H_\varepsilon^0(y|x) = P(E^0 \le y|x)$ and $H_{\varepsilon\delta}^0(y|x) = P(E^0 \le y, \Delta = \delta|x)$ ($\delta = 0, 1$). 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.

It is easily seen that

$$Y_i^* = Y_i\Delta_i + [m^0(X_i) + \frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^0)} \int_{E_i^0}^\infty y\, dF_\varepsilon^0(y)](1 - \Delta_i)$$

for any location function $m^0(\cdot)$ and scale function $\sigma^0(\cdot)$. The idea is now to choose m^0 and σ^0 in such a way that they can be estimated consistently. As is well known, the right tail of the distribution $F(y|\cdot)$ cannot be estimated in a consistent way due to the presence of right censoring. Therefore, we work with the following choices of m^0 and σ^0 :

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

where $F^{-1}(s|x) = \inf\{y; F(y|x) \ge s\}$ 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), $m^0(x)$ and $\sigma^0(x)$ can be estimated consistently. Now, replace the distribution $F(y|x)$ in (2.1) by the Beran (1981) estimator, defined by (in the case of no ties) :

$$\hat{F}(y|x) = 1 - \prod_{Z_i \le y, \Delta_i = 1} \Big\{1 - \frac{W_i(x, a_n)}{\sum_{j=1}^n I(Z_j \ge Z_i) W_j(x, a_n)}\Big\}, \qquad (2.2)$$

where

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

K is a kernel function and $\{a_n\}$ a bandwidth sequence, and define

$$\hat{m}^0(x) = \int_0^1 \hat{F}^{-1}(s|x)J(s)\, ds, \quad \hat{\sigma}^{02}(x) = \int_0^1 \hat{F}^{-1}(s|x)^2 J(s)\, ds - \hat{m}^{02}(x) \qquad (2.3)$$

as estimators for $m^0(x)$ and $\sigma^{02}(x)$. Next, let

$$\hat{F}_\varepsilon^0(y) = 1 - \prod_{\hat{E}_{(i)}^0 \le y, \Delta_{(i)} = 1} \left(1 - \frac{1}{n - i + 1}\right), \qquad (2.4)$$

denote the Kaplan-Meier (1958)-type estimator of F_ε^0 (in the case of no ties), where $\hat{E}_i^0 = (Z_i - \hat{m}^0(X_i))/\hat{\sigma}^0(X_i)$, $\hat{E}_{(i)}^0$ is the i-th order statistic of $\hat{E}_1^0, \ldots, \hat{E}_n^0$ and $\Delta_{(i)}$ is the corresponding censoring indicator. This estimator has been studied in detail by Van Keilegom and Akritas (1999). This leads to the following estimator of Y_i^* :

$$\hat{Y}_{Ti}^* = Y_i\Delta_i + \Big\{\hat{m}^0(X_i) + \frac{\hat{\sigma}^0(X_i)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})} \int_{\hat{E}_i^{0T}}^{\hat{S}_i} y \, d\hat{F}_\varepsilon^0(y)\Big\}(1 - \Delta_i), \tag{2.5}$$

where $\hat{S}_i = (T_{X_i} - \hat{m}^0(X_i))/\hat{\sigma}^0(X_i)$, $\hat{E}_i^{0T} = \hat{E}_i^0 \wedge \hat{S}_i$, and for any x, $T_x \leq T\sigma^0(x) + m^0(x)$, where $T < \tau_{H_\varepsilon^0}$ 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 $\tau_{F_\varepsilon^0} \leq \tau_{G_\varepsilon^0}$, the bound $\hat{S}_i$ can be chosen arbitrarily close to $\tau_{F_\varepsilon^0}$ for n sufficiently large). Finally, the new data points (2.5) are introduced into the least squares problem

$$\min_{\theta \in \Theta} \sum_{i=1}^n [\hat{Y}_{Ti}^* - m_\theta(X_i)]^2. \tag{2.6}$$

In order to focus on the primary issues, we assume the existence of a well-defined minimizer of (2.6). The solution of this problem can be obtained using an (iterative) procedure for nonlinear minimization problems, like e.g. a Newton-Raphson procedure. Denote a minimizer of (2.6) by $\hat{\theta}_n^T = (\hat{\theta}_{n1}^T, \ldots, \hat{\theta}_{nd}^T)$. As it is clear from the definition of $\hat{Y}_{Ti}^*$, $\hat{\theta}_{n1}^T, \ldots, \hat{\theta}_{nd}^T$ are actually estimating the unique $\theta_0^T = (\theta_{01}^T, \ldots, \theta_{0d}^T)$ which minimizes $E[\{E(Y_T^*|X) - m_\theta(X)\}^2]$ (see hypothesis (A10) in the Appendix), where

$$Y_T^* = Y\Delta + \Big\{m^0(X) + \frac{\sigma^0(X)}{1 - F_\varepsilon^0(E^{0T})} \int_{E^{0T}}^{S_X} y \, dF_\varepsilon^0(y)\Big\}(1 - \Delta),$$

$S_X = (T_X - m^0(X))/\sigma^0(X)$ and $E^{0T} = (Z \wedge T_X - m^0(X))/\sigma^0(X) = E^0 \wedge S_X$. As before, these coefficients $\theta_{01}^T, \ldots, \theta_{0d}^T$ can be made arbitrarily close to $\theta_{01}, \ldots, \theta_{0d}$, provided $\tau_{F_\varepsilon^0} \leq \tau_{G_\varepsilon^0}$.

3 Asymptotic results

We start by showing the convergence in probability of $\hat{\theta}_n^T$ and of the least squares criterion function. This will allow us to develop an asymptotic representation for $\hat{\theta}_{nj}^T - \theta_{0j}^T$ ($j = 1, \ldots, d$), which in turn will give rise to the asymptotic normality of these estimators. The assumptions and notations used in the results below, as well as the proof of the two first results, are given in the Appendix.

Theorem 3.1 *Assume (A1) (i)-(iii), (A2) (i), (ii), (A3) (i), (ii), (A4) (i), (A6), (A10) and $m_\theta(x)$ is continuous in (x, θ). Let $S_n(\theta) = \frac{1}{n}\sum_{i=1}^n [\hat{Y}_{Ti}^* - m_\theta(X_i)]^2$. Then,*

$$\hat{\theta}_n^T - \theta_0^T = o_P(1),$$

and

$$S_n(\hat{\theta}_n^T) = E[\sigma^{02}(X)Var(\varepsilon_T^{0*}|X)] + E[\{E(Y_T^*|X) - m_{\theta_0^T}(X)\}^2] + o_P(1),$$

where

$$\varepsilon_T^{0*} = \varepsilon^0\Delta + \frac{1}{1 - F_\varepsilon^0(E^{0T})}\int_{E^{0T}}^{S_X} u dF_\varepsilon^0(u)(1-\Delta).$$

Theorem 3.2 *Assume (A1)-(A10). Then,*

$$\hat{\theta}_n^T - \theta_0^T \;= \Omega^{-1}n^{-1}\sum_{i=1}^n \rho(X_i, Z_i, \Delta_i) + \begin{pmatrix} o_P(n^{-1/2}) \\ \vdots \\ o_P(n^{-1/2}) \end{pmatrix},$$

where $\Omega = (\Omega_{jk})$ $(j, k = 1, \ldots, d)$,

$$\Omega_{jk} = E\left[\frac{\partial m_{\theta_0^T}(X)}{\partial\theta_j}\frac{\partial m_{\theta_0^T}(X)}{\partial\theta_k} - \{Y_T^* - m_{\theta_0^T}(X)\}\frac{\partial^2 m_{\theta_0^T}(X)}{\partial\theta_j\partial\theta_k}\right],$$

$\rho = (\rho_1, \ldots, \rho_d)'$,

$$\begin{aligned} \rho_j(X_i, Z_i, \Delta_i) \quad = \quad & \int_{R_X} \frac{\partial m_{\theta_0^T}(x)}{\partial\theta_j}\sigma(x)\int\left\{\frac{\varphi\{X_i, Z_i, \Delta_i, e_x^{0T}(z)\}}{[1 - F_\varepsilon^0\{e_x^{0T}(z)\}]^2}\int_{e_x^{0T}(z)}^{S_x} u\, dF_\varepsilon^0(u)\right. \\ & \left. + \frac{1}{1 - F_\varepsilon^0\{e_x^{0T}(z)\}}\int_{e_x^{0T}(z)}^{S_x} u\, d\varphi(X_i, Z_i, \Delta_i, u)\right\} dH_0(z|x)\, dF_X(x) \\ & + f_X(X_i)\int B_j(z, Z_i, \Delta_i|X_i)\, dH_0(z|X_i) + \frac{\partial m_{\theta_0^T}(X_i)}{\partial\theta_j}(Y_{Ti}^* - m_{\theta_0^T}(X_i)) \end{aligned}$$

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

Theorem 3.3 *Under the assumptions of Theorem 3.2,* $n^{1/2}(\hat{\theta}_n^T - \theta_0^T) \xrightarrow{d} N(0, \Sigma)$*, where*

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

The proof of this result follows readily from Theorem 3.2.

Remark 3.4 Note that when model (1.1) is polynomial, the representation in Theorem 3.2 reduces to the one obtained by Heuchenne and Van Keilegom (2004). Note that when the model is polynomial, $E(Y_{Ti}^*|X_i) = m_{\theta_0^T}(X_i)$.

4 Practical implementation and simulations

4.1 Practical implementation

The estimator $\hat{\theta}_n^T$ depends on a number of parameters, namely on the bandwidth a_n, the score function J, and the cut off point T. All of them can be chosen in a data driven way. First, for the score function J, we recommend the choice $J(s) = b^{-1} I(0 \leq s \leq b)$ $(0 \leq s \leq 1)$, 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.

A number of adaptive procedures can be used to select the bandwidth a_n, depending on the criterion function one has in mind. If the goal would be to optimize the estimation of m^0 and σ^0, one could use e.g. a bootstrap approach that selects the bandwidth that minimizes the estimated MSE of their estimators. On the other hand, it seems more appropriate here to choose the bandwidth in function of the end goal, namely in order to optimize the estimation of θ. We therefore prefer to choose the bandwidth by minimizing the least squares criterium function

$$\sum_{i=1}^{n} \{\hat{Y}_{Ti}^*(a_n) - m_{\hat{\theta}_n^T(a_n)}(X_i)\}^2 \tag{4.1}$$

over a grid of a_n-values. Note that we have added the argument a_n to $\hat{Y}_{Ti}^*$ and $\hat{\theta}_n^T$ $(i = 1, \ldots, n)$ in order to emphasize the dependence on a_n of these quantities. We illustrate this procedure to select the bandwidth in the next subsection.

Finally, $\hat{S}_i$ $(i = 1, \ldots, n)$ can be chosen larger (or equal) than the last order statistic $\hat{E}_{(n)}^0$ of the estimated residuals. In this way, all the Kaplan-Meier jumps of the integral (2.5) are considered.

As an alternative to the normal approximation obtained in the previous section, a bootstrap procedure can be used to approximate the distribution of $\hat{\theta}_n^T$. The procedure proposed by Li and Datta (2001) can be used for this. It extends Efron's (1981) procedure developed for censored data to the nonparametric regression context. If in addition one wants to take advantage of the validity of model (1.1), a more elaborate procedure can be used, in which the survival times are drawn under model (1.1) (instead of from a nonparametric estimator of their conditional distribution).

4.2 Simulations

We compare now the finite sample behavior of Stute's (1999) estimator with the newly proposed estimator. 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 250 simulations.

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

$$Y = \frac{\theta_0}{11} \sin(\theta_1 X^2) + \sigma\varepsilon \tag{4.2}$$

where $\theta_0 = \theta_1 = 11$, $\sigma^2 = 0.5$ or 1, 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 $C = (\alpha_0/11)\sin(\alpha_1 X^2) + \sigma\varepsilon^*$, 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,

$$P(\Delta = 0|X = x) = 1 - \Phi\Big(\frac{(\alpha_0/11)\sin(\alpha_1 x^2) - (\theta_0/11)\sin(\theta_1 x^2)}{\sqrt{2}\sigma}\Big).$$

We compare here the new method for homoscedastic errors, with Stute's method. Note that the conditions of the latter method, which are outlined in Section 1 (i.e. Y and C are independent and $P(Y \le C|X, Y) = P(Y \le C|Y)$) are not satisfied for this model.

We work with a biquadratic kernel function $K(x) = (15/16)(1 - x^2)^2 I(|x| \le 1)$. 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). Since these kernels can become negative, it may happen that 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 two methods, the Levenberg-Marquardt algorithm (Levenberg (1944) and Marquardt (1963)) is used to solve equations (1.2) and (2.6) (for a fixed value of the bandwidth parameter).

The bandwidth a_n is selected by minimizing expression (4.1) over a grid of 20 possible bandwidths between 0 and 1. For small values of a_n, the window $[x - a_n, x + a_n]$ at a point x might not contain any X_i $(i = 1, \ldots, n)$ for which the corresponding Y_i is uncensored (and in that case estimation of $F(\cdot|x)$ is impossible). We enlarge the window in that case such that it contains at least one uncensored data point in its interior. It might also happen 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.

Table 1 summarizes the simulation results for different values of α_0, α_1 and σ. For fixed value of 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 proportion of censoring (in % and denoted CP in the tables) is computed as the average of $P(\Delta = 0|x)$ for an equispaced grid of values of x.

α_0	α_1	$\hat{\theta}_0$			$\hat{\theta}_1$		
σ^2	CP	Bias	Var	MSE	Bias	Var	MSE
24	1.6	-0.56	4.340	4.652	-0.06	0.209	0.212
1	34.37	-0.36	2.808	2.940	0.034	0.116	0.117
11	11	1.271	3.669	5.285	0.160	0.240	0.266
1	50	-0.58	3.046	3.377	0.043	0.169	0.171
11	12	1.258	3.909	5.493	0.350	0.284	0.406
1	51.12	-0.61	3.175	3.544	0.023	0.198	0.199
24	1.6	-0.49	2.043	2.285	-0.04	0.081	0.083
0.5	32.73	-0.16	1.660	1.686	0.021	0.052	0.052
11	11	0.701	1.799	2.290	0.136	0.120	0.138
0.5	50	-0.44	1.641	1.836	0.027	0.088	0.089
11	12	0.637	2.078	2.484	0.300	0.152	0.242
0.5	51.52	-0.46	1.581	1.794	-0.01	0.113	0.113

Table 1: *Results for the Stute estimator (first line) and the new estimator (second line) for model (4.2).*

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

$$Y = \frac{5}{4}\exp(\theta_0 X + \theta_1 X^2) + \sigma\varepsilon \tag{4.3}$$

where $\theta_0 = 0.8$ and $\theta_1 = 1$. The other quantities in model (4.3) are chosen as in (4.2). The censoring variable C satisfies $C = \alpha_0 \exp(\alpha_1 X + \alpha_2 X^2) + \sigma\varepsilon^*$, with the same characteristics as in the first setting. Table 2 summarizes the simulation results for different values of $\alpha_0, \alpha_1, \alpha_2$ and σ chosen as for Table 1. Again we compare the new method for homoscedastic errors with Stute's estimator, whose assumptions on the survival and censoring variables are not satisfied here.

In the third setting we consider a normal heteroscedastic regression model

$$Y = \theta_0 X + \sin(\theta_1 X) + \gamma X\varepsilon, \tag{4.4}$$

with $\theta_0 = 1$, $\theta_1 = 10$, X has a uniform distribution on $[0, 1]$, ε has a standard normal distribution, and γ equals $1, 2, 3$ or 4. The censoring variable is given by $C = \alpha_0 X + \sin(\alpha_1 X) + \gamma\varepsilon^*$, where ε^* has a standard normal distribution. We further assume that ε

α_0	α_1	α_2	$\hat{\theta}_0$			$\hat{\theta}_1$		
σ^2	CP		Bias	Var	MSE	Bias	Var	MSE
1.25	1.1	1	-0.52	0.081	0.349	0.549	0.104	0.406
1	33.98		0.002	0.077	0.077	-0.01	0.098	0.098
1.25	0.8	1	-0.71	0.142	0.646	0.711	0.173	0.678
1	50		0.026	0.088	0.088	-0.05	0.111	0.113
1.25	0.2	1.65	-0.91	0.173	0.995	0.912	0.208	1.040
1	55.22		0.033	0.108	0.109	-0.05	0.132	0.135
1.25	1.05	1	-0.36	0.038	0.164	0.376	0.049	0.190
0.5	32.27		0.011	0.039	0.039	-0.02	0.049	0.050
1.25	0.8	1	-0.53	0.064	0.345	0.524	0.081	0.355
0.5	50		0.030	0.045	0.046	-0.04	0.057	0.059
1.25	0.25	1.65	-0.66	0.077	0.516	0.672	0.094	0.546
0.5	53.62		0.046	0.056	0.058	-0.06	0.069	0.072

Table 2: *Results for the Stute estimator (first line) and the new estimator (second line) for model (4.3).*

and ε^* are independent of X, and that ε is independent of ε^*. The assumptions of Stute's method outlined in the introduction are again not satisfied here, and moreover the model is heteroscedastic, whereas Stute works under a homoscedastic model. The assumptions of the new method are satisfied. Since the model is heteroscedastic, we estimate here the scale function $\sigma(\cdot)$. Table 3 summarizes the simulation results for increasing values of γ, approximately constant proportion of censoring and approximately the same shape of censoring probability curve.

Tables 1 till 3 show that the new method outperforms Stute's estimator when the restrictive conditions of Stute's procedure are not satisfied. This is especially reflected in the bias, which is in most cases quite large in comparison with the new method.

Let us now consider a homoscedastic model, in which C and X are independent :

$$Y = \frac{5}{4}\exp(\theta_0 X + \theta_0 X^2) + \sigma\varepsilon \tag{4.5}$$

where $\theta_0 = 0.8$, $\theta_1 = 1$, and X, ε and σ are as in model (4.2). The censoring variable C satisfies $C = \alpha_0 + \rho\varepsilon^*$ for some α_0 and ρ, where ε^* is the same as before. This model equals model (4.3) except that the censoring variable is independent of X here. It is easily seen that Stute's assumptions are satisfied in that case. The results in Table 4 now indicate that Stute's estimator behaves better than the new one, which is not surprising, since the assumptions made by Stute (which are satisfied here) are stronger than our assumptions.

α_0	α_1	$\hat{\theta}_0$			$\hat{\theta}_1$		
γ	CP	Bias	Var	MSE	Bias	Var	MSE
1.2	11	-0.12	0.044	0.059	0.279	0.090	0.168
1	49.54	-0.05	0.037	0.039	0.038	0.066	0.068
1.2	12	-0.17	0.171	0.200	0.278	1.091	1.168
2	50.84	-0.17	0.124	0.154	0.022	0.694	0.694
1.2	13.7	-0.15	0.416	0.438	-0.72	4.166	4.678
3	50.33	-0.27	0.286	0.358	-0.16	1.560	1.589
1.2	14.3	-0.16	0.669	0.695	-1.14	5.227	6.518
4	49.99	-0.31	0.478	0.572	-0.42	3.025	3.205

Table 3: *Results for the Stute estimator (first line) and the new estimator (second line) for model (4.4).*

α_0	ρ	$\hat{\theta}_1$			$\hat{\theta}_2$		
σ^2	CP	Bias	Var	MSE	Bias	Var	MSE
7.5	10	0.003	0.098	0.098	-0.01	0.131	0.131
1	33.04	0.086	0.095	0.103	-0.12	0.130	0.144
2.7	15	-0.01	0.139	0.139	-0.00	0.186	0.186
1	50.87	0.162	0.137	0.164	-0.22	0.192	0.240
7	17	0.009	0.055	0.055	-0.02	0.073	0.073
0.5	40.82	0.087	0.056	0.063	-0.12	0.077	0.091
0	13	0.012	0.084	0.085	-0.02	0.118	0.119
0.5	59.12	0.173	0.081	0.111	-0.24	0.119	0.175

Table 4: *Results for the Stute estimator (first line) and the new estimator (second line) for model (4.3).*

Table 5 shows the relative performance of the new estimator and the one of Stute for small samples. Let $n = 30$ and consider the model

$$Y = \theta_0 \exp(\theta_1 X + \theta_2 X^2) + (\gamma X + 0.1)\varepsilon, \tag{4.6}$$

where $\theta_0 = 1.25$, $\theta_1 = 0.8$ $\theta_2 = 1$, and $\gamma = 1, 2, 3$ or 4. The other quantities of (4.6) are chosen as in (4.2) and the censoring variable C satisfies $C = \alpha_0 \exp(\alpha_1 X + \alpha_2 X^2) + \gamma\varepsilon^*$ for some $\alpha_0, \alpha_1, \alpha_2$ and ε^* chosen as in the other models. Table 5 shows that also for small samples the new method performs well.

α_0	α_1	α_2	$\hat{\theta}_0$			$\hat{\theta}_1$			$\hat{\theta}_2$		
σ^2	γ	CP	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
1.25	1.1	1	0.544	0.231	0.526	-1.51	1.233	3.513	1.244	0.923	2.471
1	1	32.77	0.232	0.148	0.202	-0.49	1.307	1.544	0.359	1.043	1.172
1.25	1.4	0.9	0.516	0.351	0.617	-1.62	2.560	5.175	1.400	2.137	4.096
1	2	33	0.242	0.270	0.328	-0.47	2.665	2.882	0.340	2.227	2.343
1.25	1.6	0.9	0.457	0.690	0.899	-1.49	6.397	8.628	1.384	5.343	7.259
1	3	33.06	0.266	0.450	0.521	-0.39	5.736	5.887	0.268	4.860	4.932
1.25	1.8	0.9	0.427	1.013	1.196	-1.47	8.124	10.28	1.489	7.377	9.593
1	4	32.65	0.276	0.749	0.825	-0.18	11.20	11.23	0.118	9.176	9.190

Table 5: *Results for the Stute estimator (first line) and the new estimator (second line) for model (4.6) and* $n = 30$.

Finally, other simulations (not reported here) show that models different from the ones considered here, lead to similar simulation results : whenever the restrictive conditions of Stute's method are satisfied, his method outperforms the new one, whereas the new method behaves considerable better than Stute's estimator in situations where these conditions are not satisfied.

5 Data analysis

The relationship between fatigue life of metal, ceramic and composite materials and applied stress is an important input to design-for-reliability processes. This is motivated by the need to develop and present quantitative fatigue-life information used in the design of jet engines. For example, according to the air speed that enters an aircraft engine, the fan, the compressor and the turbine rotate at different speeds and therefore are submitted to different stresses.

We present, in this section, a set of low-cycle fatigue life data for a strain-controlled test on 26 cylindrical specimens of a nickel-base superalloy. The data were originally described and analyzed in Nelson (1984), and can also be found in Meeker and Escobar (1998). Figure 1 shows the log of the number of cycles before failure against the pseudostress (Young's modulus times strain). Four censored data are observed; a data point is censored if failure occurs in the radius, weld or threads (censoring coming from impurities or vacuums) or if no failure occurs at all. Therefore, it may be reasonable to think that censoring depends on pseudostress. So, the assumptions of Stute's (1999) procedure (i.e. Y independent of C and $P(Y \leq C|X,Y) = P(Y \leq C|Y)$) are possibly not satisfied. Moreover, the data seem to follow a heteroscedastic model.
A model often used in the literature (see Pascual and Meeker (1997)) is given by

$$\log Y = \beta_0 + \beta_1 \log(X - \gamma) + \sigma(X)\varepsilon \qquad (X > \gamma), \tag{5.1}$$

where ε is independent of X and $E[\varepsilon] = 0$. Unlike Pascual and Meeker (1997), we do not

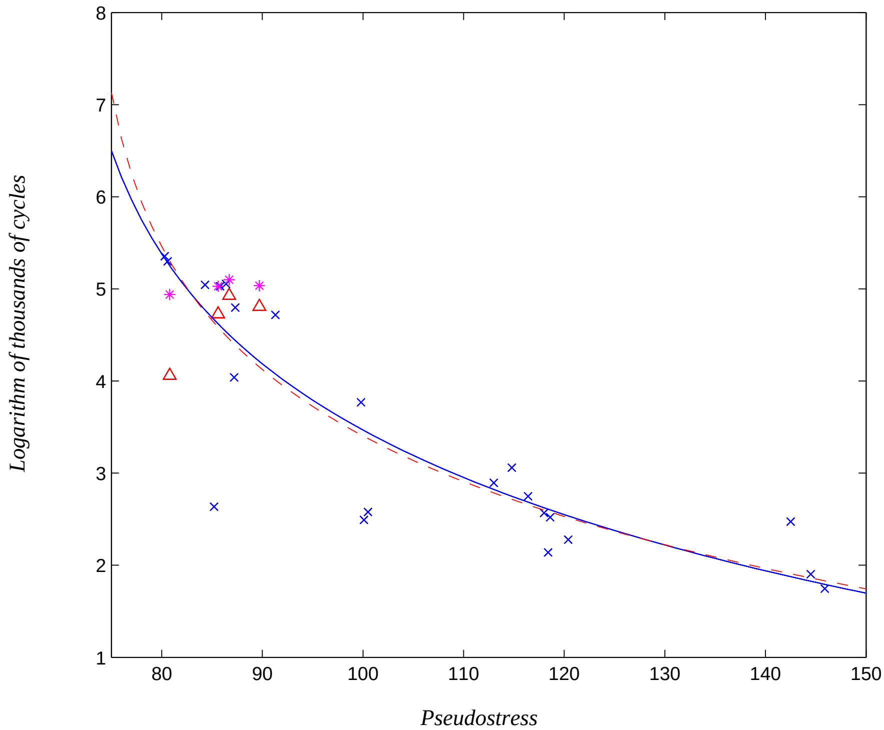

Figure 1: *Nonlinear regression for the fatigue data. The solid, respectively, dashed line represents the estimated regression curve for the new method, respectively, Stute's method. Uncensored data points are given by* $\times$*, and censored observations by* $\triangle$*. The new data points obtained from the new method are represented by* $*$*.*

impose any parametric form for $\sigma(\cdot)$. This has obviously the advantage of being more robust and flexible, although Pascual and Meeker's parametric model for $\sigma(\cdot)$ will be more efficient in some particular situations.

Note that our method does not require the estimation of the variance function $\sigma^2(\cdot)$ (since we work with $\sigma^{02}(\cdot)$ in the estimation procedure). It would however be possible to estimate $\sigma^2(\cdot)$ by using the following idea, in analogy with the idea developed in Heuchenne and Van Keilegom (2005) for the nonparametric estimation of the mean function : create artificial data points Y^* and $(Y^2)^*$ such that $E(Y^*|X) = E(Y|X)$ and $E[(Y^2)^*|X] = E(Y^2|X)$ (Y^* is as defined in Section 2, $(Y^2)^*$ can be defined similarly), and estimate the variance based on these new synthetic data. This estimator of $\sigma^2(\cdot)$ can then also be used to validate the parametric form of $\sigma^2(\cdot)$ used in Pascual and Meeker (1997).

The parameter γ in model (5.1) can be interpreted as a fatigue limit parameter i.e. specimens tested below this fatigue limit level of stress will never fail. Therefore, this parameter is constrained to be positive. The data set is then analyzed by means of Stute's procedure and the new method. For the computation of the variance, an asymptotic formula is used for Stute's method while for the new method we make use of a bootstrap approximation (see Subsection 4.1 for more details). Finally, confidence intervals are constructed for each parameter.

The results obtained by Stute's method are $\hat{\beta}_{0s} = 11.0474$, $\hat{\beta}_{1s} = -2.1315$, $\hat{\gamma}_s = 65.2730$ with estimated variances 13.6611; 0.6587; 120.5972 respectively, while the new method provides $\hat{\beta}_{0n} = 9.2432$, $\hat{\beta}_{1n} = -1.7221$, $\hat{\gamma}_n = 71.1797$ with estimated variances 9.3389; 0.4100; 116.4890 respectively. Confidence intervals for Stute's method are $\beta_0 \in [3.8031,\ 18.2917]$, $\beta_1 \in [-3.7222,\ -0.5408]$, and $\gamma \in [43.7489,\ 86.7971]$. For the new method, confidence intervals based on the normal approximation are $\beta_0 \in [3.6009;\ 15.5802]$, $\beta_1 \in [-3.0564;\ -0.5463]$ and $\gamma \in [48.9448;\ 91.2534]$, and intervals obtained with the bootstrap percentile method are $\beta_0 \in [6.7985;\ 19.0904]$, $\beta_1 \in [-3.7492;\ -1.1587]$ and $\gamma \in [33.4215;\ 77.9497]$. Graphs of the estimated curves are given in Figure 1.

Appendix : Proofs of main results

The following notations are needed in the statement of the asymptotic results given in Section 3.

$$\xi_\varepsilon(z,\delta,y) = (1 - F_\varepsilon^0(y)) \left\{ -\int_{-\infty}^{y\wedge z} \frac{dH_{\varepsilon 1}^0(s)}{(1-H_\varepsilon^0(s))^2} + \frac{I(z \le y, \delta = 1)}{1 - H_\varepsilon^0(z)} \right\},$$

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

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

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

$$\gamma_1(y|x) = \int_{-\infty}^{y} \frac{h_\varepsilon^0(s|x)}{(1-H_\varepsilon^0(s))^2}\,dH_{\varepsilon 1}^0(s) + \int_{-\infty}^{y} \frac{d\,h_{\varepsilon 1}^0(s|x)}{1-H_\varepsilon^0(s)},$$

$$\gamma_2(y|x) = \int_{-\infty}^{y} \frac{sh_\varepsilon^0(s|x)}{(1-H_\varepsilon^0(s))^2}\,dH_{\varepsilon 1}^0(s) + \int_{-\infty}^{y} \frac{d\,(sh_{\varepsilon 1}^0(s|x))}{1-H_\varepsilon^0(s)},$$

$$\varphi(x,z,\delta,y) = \xi_\varepsilon\left(\frac{z-m^0(x)}{\sigma^0(x)},\delta,y\right) - S_\varepsilon^0(y)\eta(z,\delta|x)\gamma_1(y|x) - S_\varepsilon^0(y)\zeta(z,\delta|x)\gamma_2(y|x),$$

$$\alpha_i(v) = \frac{\int_v^{S_i} u\,dF_\varepsilon^0(u)}{1-F_\varepsilon^0(v)},$$

$$\begin{aligned} &B_k(z,Z_j,\Delta_j|X_i) \\ &\quad= \frac{\partial m_{\theta_0^T}(X_i)}{\partial\theta_k} f_X^{-1}(X_i)\sigma^0(X_i)\left\{\left[\alpha_i'(e_i^{0T}(z)) - 1 + \frac{S_i f_\varepsilon^0(S_i)}{1-F_\varepsilon^0(e_i^{0T}(z))}\right]\eta(Z_j,\Delta_j|X_i)\right. \\ &\qquad\left. + \left[e_i^{0T}(z)\alpha_i'(e_i^{0T}(z)) - \alpha_i(e_i^{0T}(z)) + \frac{S_i^2 f_\varepsilon^0(S_i)}{1-F_\varepsilon^0(e_i^{0T}(z))}\right]\zeta(Z_j,\Delta_j|X_i)\right\}, \end{aligned}$$

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

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 a compact interval.
(iii) K is a density with compact support, $\int uK(u)du = 0$ and K is twice continuously differentiable.
(iv) Ω is non-singular.

$(A2)(i)$ There exist $0 \le s_0 \le s_1 \le 1$ such that $s_1 \le \inf_x F(\tilde{T}_x|x)$, $s_0 \le \inf\{s \in [0,1]; J(s) \ne 0\}$, $s_1 \ge \sup\{s \in [0,1]; J(s) \ne 0\}$ and $\inf_{x\in R_X}\inf_{s_0\le s\le s_1} f(F^{-1}(s|x)|x) > 0$.
(ii) J is twice continuously differentiable, $\int_0^1 J(s)ds = 1$ and $J(s) \ge 0$ for all $0 \le s \le 1$.
(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^0 and σ^0 are twice continuously differentiable and $\inf_{x\in R_X} \sigma^0(x) > 0$.
(iii) $E[\varepsilon^{02}] < \infty$ and $E[E^{04}] < \infty$.

$(A4)(i)$ $\eta(z,\delta|x)$ 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 $y \to P(m^0(X) + e\sigma^0(X) \le y)$ $(y \in I\!R)$ is differentiable for all $e \in I\!R$ and the derivative is uniformly bounded over all $e \in I\!R$.

$(A6)$ For $L(y|x) = H(y|x), H_1(y|x), H^0_\varepsilon(y|x)$ or $H^0_{\varepsilon 1}(y|x)$:$L'(y|x)$ 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 $\sup_{x,y} |y^3 L'''(y|x)| < \infty$.

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

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

$(A9)$ Θ is compact and θ_0^T is an interior point of Θ. All partial derivatives of $m_\theta(x)$ with respect to the components of θ up to order three exist and are continuous in (x,θ) for all x and θ. Moreover, the matrix Ω defined in Theorem 3.2 is non-singular.

$(A10)$ The function $E[\{E(Y_T^*|X) - m_\theta(X)\}^2]$ has a unique minimum in $\theta = \theta_0^T$.

Proof of Theorem 3.1. We prove the consistency of $\hat{\theta}_n^T$ by verifying the conditions of Theorem 5.7 in van der Vaart (1998, p. 45). From the definition of $\hat{\theta}_n^T$ and condition $(A10)$, it follows that it suffices to show that

$$\sup_\theta |S_n(\theta) - S_0(\theta)| \to_P 0, \tag{A.1}$$

where $S_0(\theta) = E[\sigma^{02}(X) Var(\varepsilon_T^{0*}|X)] + E[\{E(Y_T^*|X) - m_\theta(X)\}^2]$. The second statement of Theorem 3.1 then follows immediately from (A.1) together with the consistency of $\hat{\theta}_n^T$.

To prove (A.1) we write

$$S_n(\theta) = \frac{1}{n}\sum_{i=1}^n (\hat{Y}_{Ti}^* - Y_{Ti}^*)^2 + \frac{1}{n}\sum_{i=1}^n (Y_{Ti}^* - m_\theta(X_i))^2 + \frac{2}{n}\sum_{i=1}^n (\hat{Y}_{Ti}^* - Y_{Ti}^*)(Y_{Ti}^* - m_\theta(X_i))$$
$$= S_{n1} + S_{n2}(\theta) + S_{n3}(\theta).$$

In order to treat S_{n1} and $S_{n3}(\theta)$, we first consider the difference

$$\hat{Y}_{Ti}^* - Y_{Ti}^* = \Bigg\{[\hat{m}^0(X_i) - m^0(X_i)] + \frac{\hat{\sigma}^0(X_i)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})}\int_{\hat{E}_i^{0T}}^{\hat{S}_i} u\, d\hat{F}_\varepsilon^0(u)$$
$$- \frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^{0T})}\int_{E_i^{0T}}^{S_i} u\, dF_\varepsilon^0(u)\Bigg\}(1 - \Delta_i)$$
$$= \{A_{1i} + A_{2i} + A_{3i}\}(1 - \Delta_i).$$

Using Proposition 4.5 of Van Keilegom and Akritas (1999) (hereafter abbreviated by VKA), $A_{1i} = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ a.s. uniformly in i. Next, write

$$A_{2i} + A_{3i}$$
$$= \frac{\hat{\sigma}^0(X_i) - \sigma^0(X_i)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})}\int_{\hat{E}_i^{0T}}^{\hat{S}_i} u\, d\hat{F}_\varepsilon^0(u) + \sigma^0(X_i)\frac{\hat{F}_\varepsilon^0(\hat{E}_i^{0T}) - F_\varepsilon^0(E_i^{0T})}{(1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T}))(1 - F_\varepsilon^0(E_i^{0T}))}\int_{\hat{E}_i^{0T}}^{\hat{S}_i} u\, d\hat{F}_\varepsilon^0(u)$$
$$+ \frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^{0T})}\int_{\hat{E}_i^{0T}}^{E_i^{0T}} u\, d\hat{F}_\varepsilon^0(u) + \frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^{0T})}\int_{E_i^{0T}}^{S_i} u\, d(\hat{F}_\varepsilon^0(u) - F_\varepsilon^0(u))$$
$$+ \frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^{0T})}\int_{S_i}^{\hat{S}_i} u\, d\hat{F}_\varepsilon^0(u)$$
$$= \sum_{j=1}^5 B_{ji}.$$

We will now prove the convergence to zero of each of these five terms. First, by using integration by parts, we can write

$$\int_{\hat{E}_i^{0T}}^{E_i^{0T}} u\, d\hat{F}_\varepsilon^0(u) = E_i^{0T}[\hat{F}_\varepsilon^0(E_i^{0T}) - F_\varepsilon^0(E_i^{0T})] + [E_i^{0T}F_\varepsilon^0(E_i^{0T}) - \hat{E}_i^{0T}F_\varepsilon^0(E_i^{0T})]$$
$$+ \hat{E}_i^{0T}[F_\varepsilon^0(E_i^{0T}) - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})] - \int_{\hat{E}_i^{0T}}^{E_i^{0T}} \hat{F}_\varepsilon^0(u)du. \tag{A.2}$$

By Corollary 3.2 in VKA (1999), the first term of (A.2) is $|E_i^{0T}|O_P(n^{-1/2})$, while from Proposition 4.5 in VKA (1999), it follows that the second and fourth term are $|E_i^{0T}|O((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ a.s. The third term is $|E_i^{0T}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ using Proposition 4.5 of VKA (1999) and the fact that

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

This can be obtained as follows.

$$
\begin{aligned}
&\hat{F}^0_\varepsilon\Big\{\frac{z\wedge T_x-\hat{m}^0(x)}{\hat{\sigma}^0(x)}\Big\}-F^0_\varepsilon\Big\{\frac{z\wedge T_x-m^0(x)}{\sigma^0(x)}\Big\}\\
&=\hat{F}^0_\varepsilon\Big\{\frac{z\wedge T_x-\hat{m}^0(x)}{\hat{\sigma}^0(x)}\Big\}-F^0_\varepsilon\Big\{\frac{z\wedge T_x-\hat{m}^0(x)}{\hat{\sigma}^0(x)}\Big\}\\
&\quad+F^0_\varepsilon\Big\{\frac{z\wedge T_x-\hat{m}^0(x)}{\hat{\sigma}^0(x)}\Big\}-F^0_\varepsilon\Big\{\frac{z\wedge T_x-m^0(x)}{\sigma^0(x)}\Big\}\\
&=\alpha^1_n(z,x)+\alpha^2_n(z,x).
\end{aligned}
$$

Using Corollary 3.2 of VKA (1999), $\sup_{x,z}|\alpha^1_n(z,x)|$ is $O_P(n^{-1/2})$. For $\alpha^2_n(z,x)$, we use two first order Taylor developments :

$$
\alpha^2_n(z,x) \;=\; -\frac{\hat{m}^0(x)-m^0(x)}{\hat{\sigma}^0(x)}f^0_\varepsilon(A_x)-\frac{\hat{\sigma}^0(x)-\sigma^0(x)}{\hat{\sigma}^0(x)}\frac{z\wedge T_x-m^0(x)}{\sigma^0(x)}f^0_\varepsilon(B_x),
$$

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 $\sup_e|ef^0_\varepsilon(e)|<+\infty$, $\alpha^2_n(z,x)=O((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ $a.s.$ Hence, $B_{3i}=|E_i^{0T}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$. In a similar way it can be shown that $B_{5i}=O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$. Using Lemma A.1 of Heuchenne and Van Keilegom (2004), it follows that $\int_{E_i^{0T}}^{S_i}ud\hat{F}^0_\varepsilon(u)=O_P(1)$. Hence, combining this with the rates obtained for B_{3i} and B_{5i} we get that $\int_{\hat{E}_i^{0T}}^{\hat{S}_i}ud\hat{F}^0_\varepsilon(u)=|E_i^{0T}|O_P(1)$. Therefore, using Proposition 4.5 in VKA (1999) and the uniform consistency of $\hat{F}^0_\varepsilon(\hat{E}_i^{0T})$, B_{1i} and B_{2i} are $|E_i^{0T}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$. Finally,

$$
\begin{aligned}
B_{4i}&=\frac{\sigma^0(X_i)}{1-F^0_\varepsilon(E_i^{0T})}\Big\{S_i[\hat{F}^0_\varepsilon(S_i)-F^0_\varepsilon(S_i)]-E_i^{0T}[\hat{F}^0_\varepsilon(E_i^{0T})-F^0_\varepsilon(E_i^{0T})]\\
&\quad-\int_{E_i^{0T}}^{S_i}(\hat{F}^0_\varepsilon(u)-F^0_\varepsilon(u))\,du\Big\}\\
&=|E_i^{0T}|O_P(n^{-1/2}),
\end{aligned}
$$

such that

$$
\hat{Y}^*_{Ti}-Y^*_{Ti}=|E_i^{0T}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}). \tag{A.3}
$$

Therefore, $|S_{n1}|=O_P((na_n)^{-1}\log a_n^{-1})$. In the same way, using (A.3) and the continuity of $m_\theta(x)$ on $R_X\times\Theta$, $\sup_\theta|S_{n3}(\theta)|=O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$. For $S_{n2}(\theta)$, write

$$
\begin{aligned}
S_{n2}(\theta)&=\frac{1}{n}\sum_{i=1}^n\{Y^*_{Ti}-E(Y^*_{Ti}|X_i)\}^2+\frac{1}{n}\sum_{i=1}^n\{E(Y^*_{Ti}|X_i)-m_\theta(X_i)\}^2\\
&\quad+\frac{2}{n}\sum_{i=1}^n\{Y^*_{Ti}-E(Y^*_{Ti}|X_i)\}\{E(Y^*_{Ti}|X_i)-m_\theta(X_i)\}=S_{n21}+S_{n22}(\theta)+S_{n23}(\theta).
\end{aligned}
$$

Since $E[\varepsilon^{02}] < \infty$, it is easily seen that

$$S_{n21} = E[\sigma^{02}(X)\{\varepsilon^*_{0T} - E(\varepsilon^*_{0T}|X)\}^2] + o(1) \quad a.s.$$

For $S_{n22}(\theta)$ and $S_{n23}(\theta)$ we use Theorem 2 of Jennrich (1969). The function g in this theorem is given by $g_\theta(x) = \{E(Y^*_T|x) - m_\theta(x)\}^2$ for $S_{n22}(\theta)$ and

$$g_\theta(z,\delta,x) = \left[z\delta + (1-\delta)\left\{m^0(x) + \frac{\sigma^0(x)}{1 - F^0_\varepsilon(\frac{z\wedge T_x - m^0(x)}{\sigma^0(x)})}\int_{\frac{z\wedge T_x - m^0(x)}{\sigma^0(x)}}^{\frac{T_x - m^0(x)}{\sigma^0(x)}} u dF^0_\varepsilon(u)\right\} - E(Y^*_T|x)\right]$$
$$\times \{E(Y^*_T|x) - m_\theta(x)\}$$

for $S_{n23}(\theta)$. Since $E|\varepsilon^0| < \infty$, $|g_\theta(x)| \le C_1$ and$|g_\theta(z,\delta,x)| \le h(z,\delta) = C_2 z\delta + C_3$ for some $C_1, C_2, C_3 > 0$, and for all (z,δ,x) and θ, where h is integrable with respect to the joint distribution of (z,δ,x). From this,

$$\sup_{\theta\in\Theta} |S_n(\theta) - E[\sigma^{02}(X)Var(\varepsilon^*_{0T}|X)] - E[\{E(Y^*_T|X) - m_\theta(X)\}^2]| = o_P(1). \tag{A.4}$$

This finishes the proof.

Proof of Theorem 3.2. For some θ_{1n} between $\hat\theta^T_n$ and θ^T_0

$$\hat\theta^T_n - \theta^T_0 = -\left\{\frac{\partial^2 S_n(\theta_{1n})}{\partial\theta\partial\theta'}\right\}^{-1}\frac{\partial S_n(\theta^T_0)}{\partial\theta} = -R_1^{-1}R_2.$$

We have

$$R_2 = -\frac{2}{n}\sum_{i=1}^n(\hat Y^*_{T_i} - Y^*_{T_i})\frac{\partial m_{\theta^T_0}(X_i)}{\partial\theta} - \frac{2}{n}\sum_{i=1}^n\{Y^*_{T_i} - m_{\theta^T_0}(X_i)\}\frac{\partial m_{\theta^T_0}(X_i)}{\partial\theta} = R_{21} + R_{22},$$

such that R_{22} is a sum of i.i.d. random variables with zero mean (by definition of θ^T_0). For each component j of R_{21}, we use a technique similar to Theorem 1 of Heuchenne and Van Keilegom (2004) to obtain an asymptotic representation. So we obtain

$$R_{21j} = -\frac{2}{n}\sum_{i=1}^n\int_{R_X}\frac{\partial m_{\theta^T_0}(x)}{\partial\theta_j}\sigma(x)\int_{-\infty}^{+\infty}\left\{\frac{\varphi(X_i,Z_i,\Delta_i,e^{0T}_x(z))}{\{1 - F^0_\varepsilon(e^{0T}_x(z))\}^2}\int_{e^{0T}_x(z)}^{S_x} u\, dF^0_\varepsilon(u)\right.$$
$$\left. + \frac{1}{1 - F^0_\varepsilon(e^{0T}_x(z))}\int_{e^{0T}_x(z)}^{S_x} u\, d\varphi(X_i,Z_i,\Delta_i,u)\right\} dH_0(z|x)dF_X(x)$$
$$+ f_X(X_i)\int_{-\infty}^{\infty} B_j(z,Z_i,\Delta_i|X_i)\, dH_0(z|X_i) + o_P(n^{-1/2}), \tag{A.5}$$

$(j = 1,\dots,d; i = 1,\dots,n)$. For R_1, we write

$$R_1 = -\frac{2}{n}\left\{\sum_{i=1}^n(\hat Y^*_{T_i} - Y^*_{T_i})\frac{\partial^2 m_{\theta_{1n}}(X_i)}{\partial\theta\partial\theta'} + \sum_{i=1}^n(Y^*_{T_i} - m_{\theta_{1n}}(X_i))\frac{\partial^2 m_{\theta_{1n}}(X_i)}{\partial\theta\partial\theta'}\right.$$
$$\left. - \sum_{i=1}^n\Big(\frac{\partial m_{\theta_{1n}}(X_i)}{\partial\theta}\Big)\Big(\frac{\partial m_{\theta_{1n}}(X_i)}{\partial\theta'}\Big)\right\} = R_{11} + R_{12} + R_{13}.$$

Using assumption (A9), the fact that $|\hat{Y}_{T_i}^* - Y_{T_i}^*| = |E_i^{0T}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ (see the proof of Theorem 3.1), we have that $R_{11} = o_P(1)$. Again using condition (A9),

$$
\begin{aligned}
R_1 &= \frac{2}{n}\sum_{i=1}^{n} \frac{\partial m_{\theta_0^T}(X_i)}{\partial \theta}\Big(\frac{\partial m_{\theta_0^T}(X_i)}{\partial \theta}\Big)' - \frac{2}{n}\sum_{i=1}^{n}\{Y_{T_i}^* - m_{\theta_0^T}(X_i)\}\frac{\partial^2 m_{\theta_0^T}(X_i)}{\partial\theta\partial\theta'} + o_P(1) \\
&= 2E\Big[\frac{\partial m_{\theta_0^T}(X)}{\partial \theta}\Big(\frac{\partial m_{\theta_0^T}(X)}{\partial \theta}\Big)' - \{Y_T^* - m_{\theta_0^T}(X)\}\frac{\partial^2 m_{\theta_0^T}(X)}{\partial\theta\partial\theta'}\Big] + o_P(1) \\
&= 2\Omega + o_P(1).
\end{aligned}
$$

The result now follows.

References

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

Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.*, **76**, 312–319.

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 (in preparation).

Jennrich, R.I. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.*, **40**, 633–643.

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

Levenberg, K. (1944). A method for the solution of certain problems in least squares. *Quart. Appl. Math.*, **2**, 164–168.

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

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, **11**, 431–441.

Meeker, W.Q. and Escobar, L.A. (1998). *Statistical methods for reliability data.* Wiley, New-York.

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

Nelson, W. (1984). Fitting of fatigue curves with nonconstant standard deviation to data with runouts. *J. Testing Eval.*, **12**, 69–77.

Pascual, F.G. and Meeker, W.Q. (1997). Analysis of fatigue data with runouts based on a model with nonconstant standard deviation and a fatigue limit parameter. *J. Testing Eval.*, **25**, 292–301.

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

Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica*, **9**, 1089–1102.

Van der Vaart, A.W. (1998). *Asymptotic statistics.* Cambridge University Press, Cambridge.

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

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