Parametric conditional variance estimation in location-scale models with censored data

Cédric Heuchenne[†] **and Géraldine Laurent**[‡]

Rue Louvrex, 14, B-4000 Liège, Belgium

e-mail: C.Heuchenne@ulg.ac.be; g.laurent@ulg.ac.be

Abstract: Suppose the random vector (X, Y) satisfies the regression model $Y = m(X)+\sigma(X)\varepsilon$, where $m(\cdot) = E(Y|\cdot)$, $\sigma^2(\cdot) = \text{Var}(Y|\cdot)$ belongs to some parametric class $\{\sigma_\theta(\cdot) : \theta \in \Theta\}$ 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 is proposed for $\sigma_\theta(\cdot)$ when $m(\cdot)$ is unknown. It is based on nonlinear least squares estimation extended to conditional variance in the censored case. The consistency and asymptotic normality of the proposed estimator are established. The estimator is studied via simulations and an important application is devoted to fatigue life data analysis.

MSC 2010 subject classifications: Primary 62N01, 62N02; secondary 62N05.

Keywords and phrases: bandwidth, bootstrap, kernel method, least squares estimation, nonparametric regression, right censoring, survival analysis.

---

[*]This research was supported by IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy) and by the Belgian national fund for scientific research F.R.S.-FNRS, grant nr. PDR T.0080.16.

[†]HEC Liège, University of Liège and Institute of Statistics, Biostatistics and Actuarial Sciences, Université catholique de Louvain

[‡]HEC Liège, University of Liège

Contents

1. Introduction

Study of the conditional variance with censored data involves an increasing interest among scientists. Indeed, domains like Medicine, Economics, Astronomy or Finance are closely concerned by this topic. In financial time series for instance, volatility (conditionally on time) often represents the quantity of interest and in this context, censoring can appear, by example in [20], when limitations are imposed on asset prices to mitigate their fluctuations. Therefore, although the methodology proposed in this paper enlarges beyond the following topic, we are here interested in the relationship between fatigue life of metal, ceramic or composite materials and applied stress. This important input to design-for-reliability processes is motivated by the need to develop and present quantitative fatigue-life information used in the design of jet engines. Indeed, 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. Moreover, fatigue life may be censored since failures may result from impurities or vacuums in the studied materials, or no failure may occur at all due to time constraints of the experiments. In particular, a frequently asked question in this context is to know whether or not the variability of fatigue life depends on the applied stress. Furthermore, in case of heteroscedasticity, a parametric shape for this (conditional) variability should be provided. We therefore consider the general heteroscedastic regression model

$$Y = m(X) + \sigma_{\theta_0}(X)\varepsilon, \tag{1.1}$$

where $m(\cdot) = E(Y|\cdot)$ is the regression curve, $\sigma^2_{\theta_0}(\cdot) = \mathrm{Var}(Y|\cdot)$, known upto a parameter vector $\theta \in \Theta$ with true unknown value θ_0, Θ is a compact subset of $I\!R^d$, and ε is independent of the (one-dimensional) covariate X. In the context displayed above, a discussion can therefore be lead about the constancy of $\sigma_{\theta_0}(\cdot)$ ($\sigma_{\theta_0}(\cdot) = \theta_0$ for a one-dimensional θ_0) and its parametric refinements to be possibly brought to fit available information. We do not consider any parametric form for $m(\cdot)$ and the distribution of ε. Indeed, since our objective is to estimate a parametric form for $\sigma_{\theta_0}(X)$ (keeping in mind that a further step is to develop goodness-of-fit tests), we want the procedure to be free of any misspecification of other quantities; these could indeed seriously influence the quality of the results on the variance itself. Suppose also 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. 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). Since X is complete, we thus observe (X_i, Z_i, Δ_i) (i=1,. . . ,n).

The aim of this paper is more specifically to extend classical least squares procedures to take into account censored data when estimating $\sigma_{\theta_0}(\cdot)$. If a lot of work has been devoted to polynomial estimation of the regression function for censored data (see e.g. [5] for a literature overview), much less work is achieved for the estimation of the conditional variance. In fact, model (1.1) was already considered in fatigue curve analysis ([13, 15]) but with parametric forms for

$m(\cdot)$ and the distribution of ε. As explained above, we want to avoid using such parametric influences (see also Section 5). In the same idea, [6] developed a methodology to estimate a parametric curve for $m(\cdot)$ without any assumed parametric shape for the conditional standard deviation and the residuals distribution. [3] proposed a goodness-of-fit test for any scale function but only adapted to a subfamily of tested parametric functions.

We thus propose a new estimation method for θ_0. The idea of the method is as follows. First, we construct for each observation a new square of the multiplicative error term that is nonparametrically estimated. Then, θ_0 is estimated by minimizing the least squares criterion for completely observed data (and parametric conditional variance estimation), applied to the so-obtained new squared errors. The procedure involves different choices of bandwidth parameters for kernel smoothing.

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 and Section 5 is devoted to a deep 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 unknown squares of multiplicative error terms of the type $\tilde{\varepsilon}^2(X) = \sigma^2_{\theta_0}(X)\varepsilon^2$, and second of applying a standard least squares procedure on the so-obtained artificial squared errors.

Define

$$\tilde{\varepsilon}^{2*}(X_i, Z_i, \Delta_i) = \tilde{\varepsilon}_i^{2*} = (Y_i - m(X_i))^2\Delta_i + E[(Y_i - m(X_i))^2 | Y_i > C_i, X_i, C_i](1 - \Delta_i).$$

It follows that for continuous distributions $F(y|x) = P(Y \leq y|x)$ and $G(y|x) = P(C \leq y|x)$,

$$\begin{aligned} E[\tilde{\varepsilon}^{2*}(X,Z,\Delta)|X=x] &= \int_{-\infty}^{+\infty} (y-m(x))^2(1-G(y|x))dF(y|x) \\ &+ \int_{-\infty}^{+\infty}\int_{c}^{+\infty} \frac{\int_c^{+\infty}(z-m(x))^2 dF(z|x)}{1-F(c|x)} dF(y|x)dG(c|x) \\ &= \int_{-\infty}^{+\infty} (y-m(x))^2(1-G(y|x))dF(y|x) \\ &\quad + \int_{-\infty}^{+\infty}\int_{c}^{+\infty} (z-m(x))^2 dF(z|x)dG(c|x) \\ &= E((Y-m(X))^2|X=x). \end{aligned}$$

Hence, we can work in the sequel with the variable $\tilde{\varepsilon}_i^{2*}$ instead of with $\tilde{\varepsilon}_i^2$. In order to estimate $\tilde{\varepsilon}_i^{2*}$, 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 $m^0(x) = T^0(F(\cdot|x))$ and $\sigma^0(x) = S^0(F(\cdot|x))$ for some functionals T and S that satisfy $T(F_{aY+b}(\cdot|x)) = aT(F(\cdot|x)) + b$ and $S(F_{aY+b}(\cdot|x)) = aS(F(\cdot|x))$, 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. Indeed, we can write $\varepsilon = (m^0(X) - m(X))/\sigma(X) + \sigma^0(X)/\sigma(X)\varepsilon^0$. Then, $S^0(F_\varepsilon(\cdot)) = (\sigma^0(X)/\sigma(X))S^0(F_\varepsilon^0(\cdot))$, where for $F_\varepsilon(t) = P(\varepsilon \leq t)$ and $F_\varepsilon^0(y) = P(\varepsilon^0 \leq y)$, $S^0(F_\varepsilon^0(\cdot)) = 1$ (for identification purpose in the model $Y = m^0(X) + \sigma^0(X)\varepsilon^0$) and $S^0(F_\varepsilon(\cdot))$ does not depend on X. It ensues that $\sigma^0(X)/\sigma(X)$ does not depend on X as well. In a similar way, calculating $T^0(F_\varepsilon(\cdot))$ leads to show that $(m^0(X) - m(X))/\sigma(X)$ is independent of X and conclude that ε^0 is also independent of X.

Define

$$H(y|x) = P(Z \leq y|x) \ (H(y) = P(Z \leq y)),$$

the observable (un)conditional distribution,

$$H_\delta(y|x) = P(Z \le y, \Delta = \delta|x),$$

the observable conditional subdistributions for $\delta = 0, 1$,

$$S^0_\varepsilon(y) = 1 - F^0_\varepsilon(y), \text{ the survival functions of } \varepsilon^0,$$

and $F_X(x) = P(X \le x)$. For $E^0 = (Z - m^0(X))/\sigma^0(X)$, we also denote $H^0_\varepsilon(y) = P(E^0 \le y)$, $H^0_{\varepsilon\delta}(y) = P(E^0 \le y, \Delta = \delta)$, $H^0_\varepsilon(y|x) = P(E^0 \le y|x)$ and $H^0_{\varepsilon\delta}(y|x) = P(E^0 \le y, \Delta = \delta|x)$ ($\delta = 0, 1$). 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

$$\tilde{\varepsilon}_i^{2*} = (Y_i - m(X_i))^2\Delta_i + \frac{\int_{E_i^0}^{\infty} [m^0(X_i) + \sigma^0(X_i)y - m(X_i)]^2\, dF^0_\varepsilon(y)}{1 - F^0_\varepsilon(E^0_i)}(1 - \Delta_i)$$

for any location function $m^0(\cdot)$ and scale function $\sigma^0(\cdot)$ and where $E^0_i = (Z_i - m^0(X_i))/\sigma^0(X_i)$ $(i = 1, \ldots, n)$. m^0 and σ^0 are now chosen in such a way that they can be estimated consistently. As is well known (see by example [19]), 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:

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

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 estimator ([1]), defined by (in the case of no ties):

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

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

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

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

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

denote the Kaplan-Meier-type estimator ([9]) of F^0_ε (in the case of no ties), where $\hat{E}^0_i = (Z_i - \hat{m}^0(X_i))/\hat{\sigma}^0(X_i)$, $\hat{E}^0_{(i)}$ is the i-th order statistic of $\hat{E}^0_1, \ldots, \hat{E}^0_n$ and $\Delta_{(i)}$ is the corresponding censoring indicator. This estimator has been studied in detail by [18]. Finally, $m(x)$ is estimated by the method of [7] applied to the estimation of a conditional mean:

$$\hat{m}^T(x) = \hat{m}^0(x) + \hat{\sigma}^0(x) \int_{-\infty}^{T} y d\hat{F}^0_\varepsilon(y), \tag{2.5}$$

where $T < \tau_{H^0_\varepsilon}$ ($\tau_F = \inf\{y : F(y) = 1\}$ for any distribution F) is a truncation point that has to be introduced to avoid any inconsistent part of $\hat{F}^0_\varepsilon(y)$. However, when $\tau_{F^0_\varepsilon} \le \tau_{G^0_\varepsilon}$, the bound T can be chosen arbitrarily close to $\tau_{F^0_\varepsilon}$.

This leads to the following estimator of $\tilde{\varepsilon}^{2*}_i$:

$$\widehat{\tilde{\varepsilon}^{2*}_{Ti}} = (Y_i - \hat{m}^T(X_i))^2 \Delta_i + \Big\{ \frac{\hat{\sigma}^{02}(X_i)}{1 - \hat{F}^0_\varepsilon(\hat{E}^{0T}_i)} \int_{\hat{E}^{0T}_i}^{T} (y^2 - 2y \int_{-\infty}^{T} e d\hat{F}^0_\varepsilon(e))\, d\hat{F}^0_\varepsilon(y)$$

$$+\hat{\sigma}^{02}(X_i)\{\int_{-\infty}^{T} yd\hat{F}_{\varepsilon}^{0}(y)\}^2\Big\}(1-\Delta_i), \quad (2.6)$$

where $\hat{E}_i^{0T} = \hat{E}_i^0 \wedge T$.

Finally, the new squared errors (2.6) are introduced into the least squares problem

$$\min_{\theta\in\Theta}\sum_{i=1}^{n}[\widehat{\tilde{\varepsilon}_{Ti}^{2*}} - \sigma_\theta^2(X_i)]^2. \quad (2.7)$$

In order to focus on the primary issues, we assume the existence of a well-defined minimizer of (2.7). 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 (2.7) by $\hat{\theta}_n^T = (\hat{\theta}_{n1}^T, \ldots, \hat{\theta}_{nd}^T)$. As it is clear from the definition of $\widehat{\tilde{\varepsilon}_{Ti}^{2*}}$, $\hat{\theta}_{n1}^T, \ldots, \hat{\theta}_{nd}^T$ are actually estimating the unique $\theta_0^T = (\theta_{01}^T, \ldots, \theta_{0d}^T)$ which minimizes $E[\{E(\tilde{\varepsilon}_T^{2*}|X) - \sigma_\theta^2(X)\}^2]$ (see hypothesis (A19), where

$$\tilde{\varepsilon}_T^{2*} = (Y - m^T(X))^2\Delta_i + \Big\{\frac{\sigma^{02}(X)}{1-F_\varepsilon^0(E^{0T})}\int_{E^{0T}}^{T}(y^2 - 2y\int_{-\infty}^{T} edF_\varepsilon^0(e))\,dF_\varepsilon^0(y)$$
$$+\sigma^{02}(X)\{\int_{-\infty}^{T} ydF_\varepsilon^0(y)\}^2\Big\}(1-\Delta_i),$$

$$m^T(X) = m^0(X) + \sigma^0(X)\int_{-\infty}^{T} ydF_\varepsilon^0(y)$$

and $E^{0T} = E^0 \wedge T$. As before, these coefficients $\theta_{01}^T, \ldots, \theta_{0d}^T$ can be made arbitrarily close to $\theta_{01}, \ldots, \theta_{0d}$, provided $\tau_{F_\varepsilon^0} \le \tau_{G_\varepsilon^0}$.

Remark 2.1 (Truncation T) The advantage of using (2.5) in (2.6) is double. On one side, it enables to use model (1.1) in a very simple way simplifying the censored part of (2.6) and on the other side, it reduces inconsistencies of those estimated squared errors. Indeed, suppose a local estimator for $m(x)$ based on (2.2) is chosen instead of (2.5): it is consistent up to a point $\tilde{T}_x < \tau_{H(\cdot|x)}$ depending on x. In this case, it can be shown that $m^0(x) + \sigma^0(x)\tau_{H^0_\varepsilon} \geq \tau_{H(\cdot|x)}$ for any value of x such that consistent areas of (2.5) can be substantially larger than for local estimators (see [7] for a complete discussion).

Remark 2.2 (Conditional scale function) Some researchers in (nonparametric) survival analysis often criticize the idea of estimating quantities which use the whole support of Y given $X = x$. The ultimate consequence in this paper is that we need to define the vector of pseudo-parameters θ_0^T instead of θ_0. A possible solution to avoid that problem is to only estimate parts of $F(\cdot|X = x)$ by truncating inconsistent areas (as it is the case for $\hat{m}^0(x)$ and $\hat{\sigma}^0(x)$) or by simply considering quantiles of this conditional distribution. The methodology proposed in this paper can be easily extended to this type of estimation. The idea is thus to reduce the area where $F(\cdot|X = x)$ is estimated by defining a conditional scale function. For the sake of accuracy, let's take the simple example of the estimation of the conditional quantile of order s $(0 < s < 1)$ of $W = |Y - F^{-1}_{Y|X}(s|X)|$. This can be defined as

$$F^{-1}_{W|X}(s|X = x) = \inf \left\{ y : F^0_\varepsilon\left(\frac{y + F^{-1}_{Y|X}(s|x) - m^0(x)}{\sigma^0(x)}\right) \right.$$
$$\left. - F^0_\varepsilon\left(\frac{-y + F^{-1}_{Y|X}(s|x) - m^0(x)}{\sigma^0(x)}\right) \geq s \right\},$$

where $F_{W|X}(\cdot|X)$ denotes the conditional distribution of W given X. $F^{-1}_{Y|X}(s|X = x)$ can be estimated with the method proposed in [7] (similarly to $\hat{m}^T(\cdot)$ in this paper and with the same objective of enabling to estimate consistently -with $\hat{F}^0_\varepsilon(\cdot)$- the conditional quantile for each value of x). This leads to the estimator (to make the following formula more readable, we omit here the theoretical bound T which 'cuts' inconsistent parts of $\hat{F}^0_\varepsilon(\cdot)$)

$$\hat{F}^{-1}_{W|X}(s|X = x) = \inf \left\{ y : \hat{F}^0_\varepsilon(y/\hat{\sigma}^0(x) + (\hat{F}^0_\varepsilon)^{-1}(s)) - \hat{F}^0_\varepsilon(-y/\hat{\sigma}^0(x) \right.$$

$$+(\hat{F}_{\varepsilon}^{0})^{-1}(s)) \geq s\Big\},$$

where $(\hat{F}_{\varepsilon}^{0})^{-1}(s) = \inf\{y : \hat{F}_{\varepsilon}^{0}(y) \geq s\}$. Finally, the resulting estimated quantiles $\hat{F}_{W|X}^{-1}(s|X_i)$, $i = 1, \ldots, n$, can be introduced in a least squares problem of the type (2.7):

$$\min_{\theta \in \Theta} \sum_{i=1}^{n} [\hat{F}_{W|X}^{-1}(s|X_i) - \sigma_{\theta}(X_i)]^2.$$

where $\sigma_{\theta}(\cdot)$ now denotes the corresponding parametric quantile function. Since the concept of variance is more widely used in other domains (see Section 5) and often preferred to this type of scale function, and since we also want to highlight the benefit of using $\hat{F}_{\varepsilon}^{0}(\cdot)$ to improve consistency (see Remark 2.1), we decided to study the variance function. However, as it can be seen here, it is easy to extend the methodology to another scale function.

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 proofs of the two first results, are given in the Appendix.

Theorem 3.1 *Assume (A1) (i)–(iii), (A2) (i), (ii), (A3), (A4) (i), (A6),* Θ *is compact,* θ_0^T *is an interior point of* Θ, $\sigma_{\theta}^2(x)$ *is continuous in* (x, θ) *for all* x *and* θ *and (A9). Let*

$$S_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} (\widehat{\tilde{\varepsilon}_{Ti}^{2*}} - \sigma_{\theta}^2(X_i))^2.$$

Then,

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

and

$$S_n(\hat\theta_n^T) = E[Var[\tilde\varepsilon_T^{2*}|X]] + E[(E[\tilde\varepsilon_T^{2*}|X] - \sigma^2_{\theta_0^T}(X))^2] + o_P(1).$$

Theorem 3.2 *Assume (A1)-(A9). 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 $\Omega = (\Omega_{jk})$ $(j, k = 1, \dots, d)$,

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

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

$$\rho_j(X_i, Z_i, \Delta_i) = \chi_j(X_i, Z_i, \Delta_i) + \frac{\partial\sigma^2_{\theta_0^T}(X_i)}{\partial\theta_j}(\tilde\varepsilon_{Ti}^{2*} - \sigma^2_{\theta_0^T}(X_i))$$

and $\chi_j(X_i, Z_i, \Delta_i)$ is defined in the Appendix $(j = 1, \dots, d; i = 1, \dots, n)$.

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.

4. Practical implementation and simulations

4.1. Practical implementation

The estimator $\hat{\theta}_n^T$ depends on a number of parameters: the score function J, the bandwidth a_n and the cut off point T that can be chosen in a data driven way. The function J is computed as in [5], i.e., $J(s) = b^{-1}I(0 \leq s \leq b)$, $(0 \leq s \leq 1)$, where $b = \min_{1\leq i\leq n} \hat{F}(+\infty|X_i)$ (the region where the Beran estimators $\hat{F}(\cdot|X_1), \ldots, \hat{F}(\cdot|X_n)$ are inconsistent is not used and we exploit to a maximum the 'consistent' region), while the point T can be chosen equal to the last order statistic $\hat{E}^0_{(n)}$ of the estimated residuals $\hat{E}^0_1, \ldots, \hat{E}^0_n$ (in this way, all the Kaplan-Meier jumps of the integral (2.6) are considered). When $\hat{E}^0_{(n)}$ is censored, it is redefined as uncensored.

To choose the bandwidth parameter, we could minimize (with respect to a_n) an asymptotic expression of

$$IMSE(a_n) = E\left[\int \{\sigma^2_{\hat{\theta}_n^T(a_n)}(x) - \sigma^2_{\theta_0^T}(x)\}^2 dx\right] \tag{4.1}$$

where $\hat{\theta}_n^T(a_n)$ denotes $\hat{\theta}_n^T$ determined with bandwidth parameter a_n. However, that would involve complicated expressions with too many unknown quantities. We therefore prefer to use the following bootstrap procedure. This is based on the method proposed by [11].

For $b = 1, \ldots, B$,

1. for $i = 1, \ldots, n$,

   Step 1. Generate $X^*_{i,b}$ from the empirical distribution of $X_1, \ldots, X_n$.

   Step 2. Select at random $Y^*_{i,b}$ from the distribution $\hat{F}(\cdot|X^*_{i,b})$ calculated with a pilot bandwidth g_n asymptotically larger than the original a_n.

   Step 3. Generate $C^*_{i,b}$ from $\hat{G}(\cdot|X^*_{i,b})$ calculated with g_n (which is the Beran (1981) estimator of $G(\cdot|X^*_{i,b})$ obtained by replacing Δ_i by $1 - \Delta_i$ in the expression of $\hat{F}(\cdot|X^*_{i,b})$).

Step 4. Define $Z^*_{i,b} = \min(Y^*_{i,b}, C^*_{i,b})$ and $\Delta^*_{i,b} = I(Y^*_{i,b} \leq C^*_{i,b})$.

2. Compute $\hat{\theta}^{T*}_{bn}(a_n)$ the estimator of the variance vector of parameters based on the bandwidth parameter a_n and the obtained resample

$$\{(X^*_{1,b}, Z^*_{1,b}, \Delta^*_{1,b}), \dots, (X^*_{n,b}, Z^*_{n,b}, \Delta^*_{n,b})\}.$$

From this, (4.1) can be approximated by

$$IMSE^*(a_n) = B^{-1} \sum_{b=1}^{B} \int [\sigma^2_{\hat{\theta}^{T*}_{bn}(a_n)}(x) - \sigma^2_{\hat{\theta}^{T}_{n}(g_n)}(x)]^2 dx, \qquad (4.2)$$

where $\widehat{\theta}^T_n(g_n)$ is the estimator of θ^T_0 based on the initial sample and the bandwidth g_n. We now select the value of a_n that minimizes $IMSE^*(a_n)$. The same bootstrap procedure is also used to approximate the distribution of $\hat{\theta}^T_n$, instead of using the asymptotic distribution of Theorem 3.3, which is hard to estimate in practice. Bootstrap confidence intervals illustrate this in Section 5.

4.2. Simulations

We now study the finite sample behavior of the newly proposed estimator compared to a similar methodology but replacing $\hat{F}^0_\varepsilon(\frac{\cdot - \hat{m}^0(X_i)}{\hat{\sigma}^0(X_i)})$ by $\hat{F}(\cdot|X_i)$. More precisely, the new squared errors are in this case replaced by

$$\widetilde{\hat{\varepsilon}^{2*}_i} = (Y_i - \tilde{m}(X_i))^2 \Delta_i + \frac{1}{1 - \hat{F}(C_i \wedge \tilde{T}_{X_i}|X_i)} \int_{C_i \wedge \tilde{T}_{X_i}}^{\tilde{T}_{X_i}} (y - \tilde{m}(X_i))^2 d\hat{F}(y|X_i)(1 - \Delta_i)$$

and $\tilde{m}(x) = \displaystyle\int_{-\infty}^{\tilde{T}_x} y d\hat{F}(y|x)$. Practically, the point $\tilde{T}_x$ is chosen as the largest data point in the window defined by the bandwidth parameter. When this data point is censored, it is redefined as uncensored. Then, the resulting estimator

$\tilde{\theta}_n$ is obtained by minimization over $\theta \in \Theta$ of the expression

$$\sum_{i=1}^{n} \left(\widetilde{\hat{\varepsilon}_i^{2*}} - \sigma_\theta^2(X_i) \right)^2 .$$

We are primarily interested in the behavior of the estimator bias, variance and mean squared error (MSE). The simulations are carried out for samples of size n = 100 and/or 200, $B = 500$ and the results are obtained by using $R = 1000$ simulations.

We work with a biquadratic kernel function $K(x) = (15/16)(1 - x^2)^2 I(|x| \leq 1)$. The bandwidth a_n is selected by minimizing the expression (4.2) over a grid of 21 possible bandwidths depending on the covariate support. The step of the grid is added to its largest value to obtain the pilot bandwidth g_n. For small values of a_n, the window $[x - a_n, x + a_n]$ at a point x might not contain any X_i $(i = 1, \dots, n)$ 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 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 endpoints of the interval. In such cases, the bandwidth is redefined as the maximum of these two distances. We did not consider the boundary issue in this paper. The estimator of [18], $\hat{F}_\varepsilon^0(\cdot)$, indeed involves $\hat{m}^0(\cdot)$ and $\hat{\sigma}^0(\cdot)$, and these can suffer from bias increasing close to the boundaries of the support of X. In the complete data case, many methods have been developed to handle this problem (see [2] for an overview of existing methods, including a new one). However, if these methods often enable to obtain a smaller bias of the studied estimators, the resulting variance is also larger, which often does not markedly lead to better mean squared errors (see also [5] for an application of boundary kernels in a similar context). Since our final objective is a parametric estimator (not a nonparametric one) based a least squares procedure using estimated (artificial) squared errors, we deemed the influence of boundary corrections too weak to be applied in this context.

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

$$Y = \beta_0 \exp(\beta_1 X + \beta_2 X^2) + (\gamma_0 + \gamma_1 X)\varepsilon, \tag{4.3}$$

where $\beta_0 = 1.25$, $\beta_1 = 0.8$, $\beta_2 = 1$, $\gamma_0 = 1$ and $\gamma_1 = 0.1, 0.25, 0.5$ or 0.75. 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 $C = \alpha_0 \exp(\alpha_1 X + \alpha_2 X^2) + \gamma_1 \varepsilon^*$ for certain choices of $\alpha_0, \alpha_1, \alpha_2$ and where ε^* has a standard normal distribution. It is easy to see that, under this model,

$$P(\Delta = 0|X = x) = 1 - \Phi\Big(\frac{\alpha_0 \exp(\alpha_1 x + \alpha_2 x^2) - \theta_0 \exp(\theta_1 x + \theta_2 x^2)}{\sqrt{(\gamma_1 x + \gamma_0)^2 + \gamma_1^2}}\Big).$$

Table 4.2 summarizes the simulation results for different values of α_0, α_1, α_2 and γ_1. For fixed value of γ_1, α_0, α_1 and α_2 are chosen in such a way that some variation in the different percentages of censoring (in % and denoted CP in the tables) is obtained. The censoring proportion is computed as the average of $P(\Delta = 0|x)$ for an equispaced grid of values of x. The smoothing parameter is chosen as the minimizer of (4.2) among a grid of values varying between 0.15 and 0.30 by step of 0.0075, and the value of the pilot bandwidth is 0.3075. Main observations from Table 4.2 are listed in the next paragraph.

For a fixed sample size, the censoring percentage increase seems to induce a deterioration of the results, whatever the considered estimation method. Moreover, a more important heteroscedasticity (through the choice of γ_1) implies worse results. Regardless of the sample size and the censoring percentage, the MSE values obtained with $\hat{\theta}_n^T = (\hat{\gamma}_{0n}^T, \hat{\gamma}_{1n}^T)$ are often smaller than the corresponding values obtained with $\tilde{\theta}_n = (\tilde{\gamma}_{0n}, \tilde{\gamma}_{1n})$ (in fact always smaller for a sample size $n = 200$). More clearly, the bias for $\hat{\theta}_n^T$ appears to be always smaller

Table 1
Results for $\tilde{\theta}_n$ (first line) and $\tilde{\theta}_n^T$ (second line) for model (4.3)

n	α_0 / γ_0	α_1 / γ_1	α_2 / CP	$\hat{\gamma}_0$ Bias	$\hat{\gamma}_0$ Var	$\hat{\gamma}_0$ MSE	$\hat{\gamma}_1$ Bias	$\hat{\gamma}_1$ Var	$\hat{\gamma}_1$ MSE
100	1.25	1.53	0.9	−0.2751	0.0254	0.1011	0.3298	0.0784	0.1872
	1	0.1	20.05	−0.0497	0.0774	0.0799	0.1063	0.1881	0.1994
100	1.25	1.58	0.9	−0.2515	0.0378	0.1011	0.2864	0.1120	0.1941
	1	0.25	20.06	−0.0562	0.0895	0.0926	0.1092	0.1970	0.2089
100	1.25	1.68	0.9	−0.2108	0.0620	0.1065	0.2158	0.1770	0.2236
	1	0.5	20.02	−0.0644	0.1147	0.1188	0.1048	0.2619	0.2729
100	1.25	1.787	0.9	−0.1852	0.0964	0.1307	0.1673	0.2651	0.2930
	1	0.75	20.03	−0.0771	0.1248	0.1308	0.0983	0.3090	0.3187
100	1.25	1.12	1	−0.3567	0.0339	0.1611	0.4477	0.1035	0.3039
	1	0.1	29.83	−0.0517	0.1009	0.1036	0.1923	0.1728	0.2098
100	1.25	1.155	1	−0.3194	0.0419	0.1439	0.3908	0.1304	0.2832
	1	0.25	29.69	−0.0522	0.1137	0.1164	0.1826	0.1951	0.2285
100	1.25	1.212	1	−0.2638	0.0792	0.1488	0.3040	0.2217	0.3142
	1	0.5	29.74	−0.0601	0.1338	0.1375	0.1587	0.3207	0.3459
100	1.25	1.276	1	−0.2247	0.1257	0.1762	0.2455	0.3315	0.3919
	1	0.75	29.68	−0.0797	0.1417	0.1481	0.1614	0.3270	0.3530
200	1.25	1.53	0.9	−0.2335	0.0095	0.0640	0.3093	0.0304	0.1261
	1	0.1	20.05	−0.0227	0.0342	0.0347	0.0497	0.0703	0.0728
200	1.25	1.58	0.9	−0.1977	0.0129	0.0520	0.2534	0.0414	0.1057
	1	0.25	20.06	−0.0243	0.0327	0.0333	0.0480	0.0697	0.0720
200	1.25	1.68	0.9	−0.1422	0.0207	0.0409	0.1711	0.0642	0.0935
	1	0.5	20.02	−0.0226	0.0355	0.0360	0.0363	0.0836	0.0849
200	1.25	1.787	0.9	−0.1084	0.0285	0.0403	0.1170	0.0906	0.1043
	1	0.75	20.03	−0.0260	0.0380	0.0387	0.0263	0.1030	0.1036
200	1.25	1.12	1	−0.3137	0.0100	0.1084	0.4270	0.0366	0.2189
	1	0.1	29.83	−0.0395	0.0398	0.0414	0.1201	0.0713	0.0858
200	1.25	1.155	1	−0.2680	0.0134	0.0852	0.3593	0.0479	0.1770
	1	0.25	29.69	−0.0383	0.0403	0.0418	0.1191	0.0761	0.0902
200	1.25	1.212	1	−0.1936	0.0227	0.0602	0.2537	0.0747	0.1391
	1	0.5	29.74	−0.0332	0.0426	0.0437	0.1044	0.0934	0.1043
200	1.25	1.276	1	−0.1448	0.0379	0.0588	0.1808	0.1120	0.1447
	1	0.75	29.68	−0.0344	0.0517	0.0529	0.0867	0.1215	0.1290

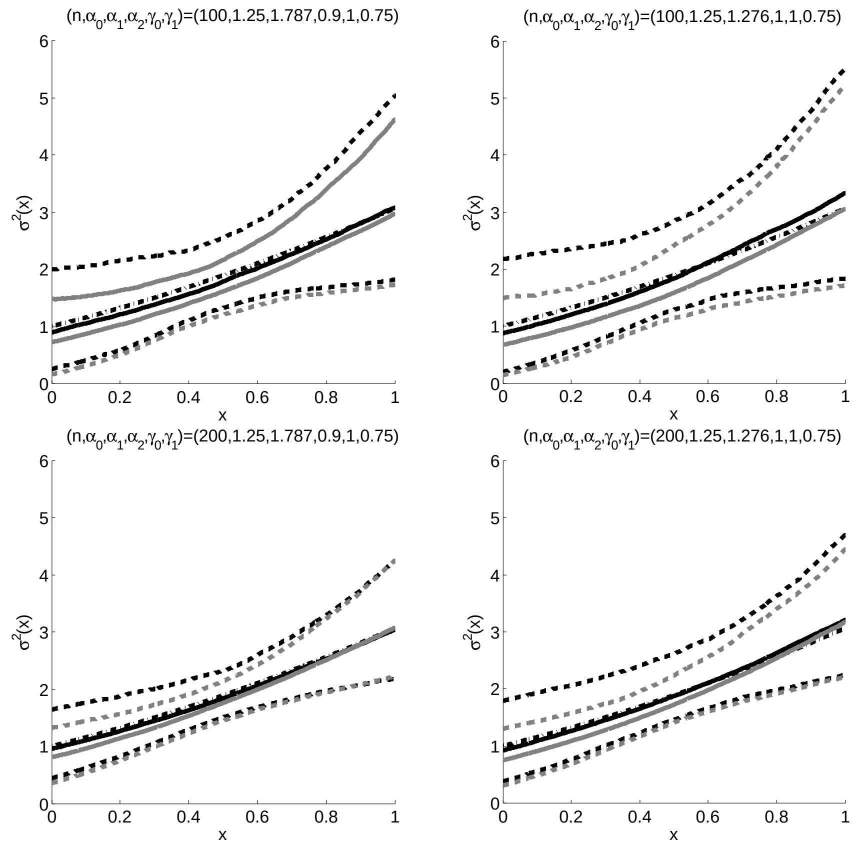

FIG 1. *Comparison of the estimated conditional variances obtained with* $\tilde{\theta}_n$ *(grey curves) and* $\hat{\theta}_n^T$ *(black curves) for model (4.3). Dashed curves :* 5% *and* 95% *quantiles of the empirical distribution based on the estimated conditional variances; solid curve: median of this empirical distribution; dash-dotted curve: true conditional variance curve.*

than for $\tilde{\theta}_n$. Obviously, the sample size increase is accompanied by an improvement of the results (bias, variance and MSE) independently of the considered estimator. This improvement seems however more important for $\hat{\theta}_n^T$.

Figure 1 also illustrates the previous results. For each value of the covariable, the dashed curves represent the 5% and 95% quantiles of the empirical distribution of the estimated conditional variance (constructed with the R estimations at this value). The solid curve represents the median of this empirical distribution while the true curve is dash-dotted. The grey (respectively black) color represents these curves for $\tilde{\theta}_n$ (respectively $\hat{\theta}_n^T$). From Figure 1, we remark that the median curve obtained with $\hat{\theta}_n^T$ is globally closer to the true curve than with $\tilde{\theta}_n$. However, quantile curves are more distant for $\sigma_{\hat{\theta}_n^T}(\cdot)$ than for $\sigma_{\tilde{\theta}_n}(\cdot)$. These effects obviously decrease when n increases.

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

$$Y = \beta_0 + \beta_1 \log(X - \beta_2) + \exp(\gamma_0 + \gamma_1 \log X)\varepsilon, \tag{4.4}$$

where $\beta_0 = 10$, $\beta_1 = -2$, $\beta_2 = 70$, $\gamma_0 = 1$ and $\gamma_1 = -0.23$. X has a uniform distribution on $[75, 150]$ and ε is a standard normal random variable. These digits are initially motivated by the type of models met in fatigue life data analysis (see Section 5 in [6]). The censoring variable C satisfies $C = \alpha_0 + \alpha_1 \log(X - \alpha_2) + \exp(\eta_0 + \eta_1 \log X)\varepsilon^*$ for certain choices of α_0, α_1, α_2, η_0, η_1 and where ε^* has a standard normal distribution. α_0 is chosen in such a way that some increasing in the different censoring percentages is obtained. Moreover, conditional mean of the censoring variable is differently sloping while its conditional variance varies as well. The smoothing parameter is chosen as the minimizer of (4.2) among a grid of values varying between 11.25 and 22.50 by step of 0.5625, and the value of the pilot bandwidth is 23.0625. Other simulations (not reported here) showed that the final results were not very sensible to the choice of these digits.

The great advantage of $\hat{\theta}_n^T = (\hat{\gamma}_{0n}^T, \hat{\gamma}_{1n}^T)$ (with respect to $\tilde{\theta}_n = (\tilde{\gamma}_{0n}, \tilde{\gamma}_{1n})$) is its apparently small bias. If its variance is often larger, the impact on MSE is relatively moderate. In Table 2, these important characteristics are observed as well. Moreover, when the censoring percentage increases, $\hat{\theta}_n^T$ seems to deteriorate less than $\tilde{\theta}_n$ (feature also observed on a small scale in Table 4.2). For this range of censoring percentages and these models, the inconsistency in the right tail of the Beran estimator ([1]) combined with its local property (especially when constructing $\widetilde{\tilde{\varepsilon}_i^{2*}}$ for censored data) has a large impact on the estimation of

Table 2
Results for $\tilde{\theta}_n$ (first line) and $\hat{\theta}_n^T$ (second line) for model (4.4)

				$\hat{\gamma}_0$			$\hat{\gamma}_1$		
n	α_0	α_1	α_2	Bias	Var	MSE	Bias	Var	MSE
	η_0	η_1	CP						
200	10.75	−2	75	0.9300	2.2928	3.1577	−0.2034	0.1036	0.1452
	0.5	−0.25	14.69	0.6997	2.5481	3.0376	−0.1515	0.1162	0.1391
200	10.50	−2	75	1.1489	2.4206	3.7406	−0.2509	0.1096	0.1726
	0.5	−0.25	20.68	0.9098	2.6177	3.4456	−0.1946	0.1196	0.1575
200	10.25	−2	75	1.4640	2.4293	4.5724	−0.3191	0.1101	0.2120
	0.5	−0.25	27.87	1.2047	2.6691	4.1203	−0.2546	0.1225	0.1873
200	10	−2	75	1.8686	2.5638	6.0555	−0.4069	0.1165	0.2821
	0.5	−0.25	36.02	1.5088	2.5720	4.8487	−0.3151	0.1189	0.2182
200	9.601	−2	75	2.7332	2.7522	10.2224	−0.5964	0.1261	0.4818
	0.5	−0.25	50.06	1.8004	2.4759	5.7175	−0.3667	0.1188	0.2532
200	9.315	−2	75	3.5586	2.9622	15.6257	−0.7783	0.1365	0.7422
	0.5	−0.25	60.02	1.8482	2.8809	6.2968	−0.3681	0.1406	0.2761
200	11.77	−2.25	72.5	1.0284	2.2208	3.2785	−0.2260	0.1007	0.1517
	−2	0.2	14.70	0.7335	2.6073	3.1454	−0.1583	0.1190	0.1441
200	11.535	−2.25	72.5	1.3114	2.2708	3.9905	−0.2873	0.1026	0.1851
	−2	0.2	20.70	1.0102	3.0111	4.0315	−0.2150	0.1382	0.1844
200	11.3	−2.25	72.5	1.6844	2.3584	5.1955	−0.3684	0.1064	0.2422
	−2	0.2	27.90	1.2617	2.7886	4.3804	−0.2653	0.1285	0.1989
200	11.067	−2.25	72.5	2.1462	2.3639	6.9701	−0.4692	0.1068	0.3270
	−2	0.2	36.01	1.4868	2.5639	4.7743	−0.3085	0.1200	0.2152
200	10.695	−2.25	72.5	3.0570	2.3013	11.6463	−0.6681	0.1047	0.5511
	−2	0.2	50.06	1.5360	2.8365	5.1957	−0.3081	0.1368	0.2317
200	10.43	−2.25	72.5	3.8283	2.3556	17.0113	−0.8369	0.1079	0.8083
	−2	0.2	60.04	1.3659	3.4927	5.3584	−0.2624	0.1708	0.2397

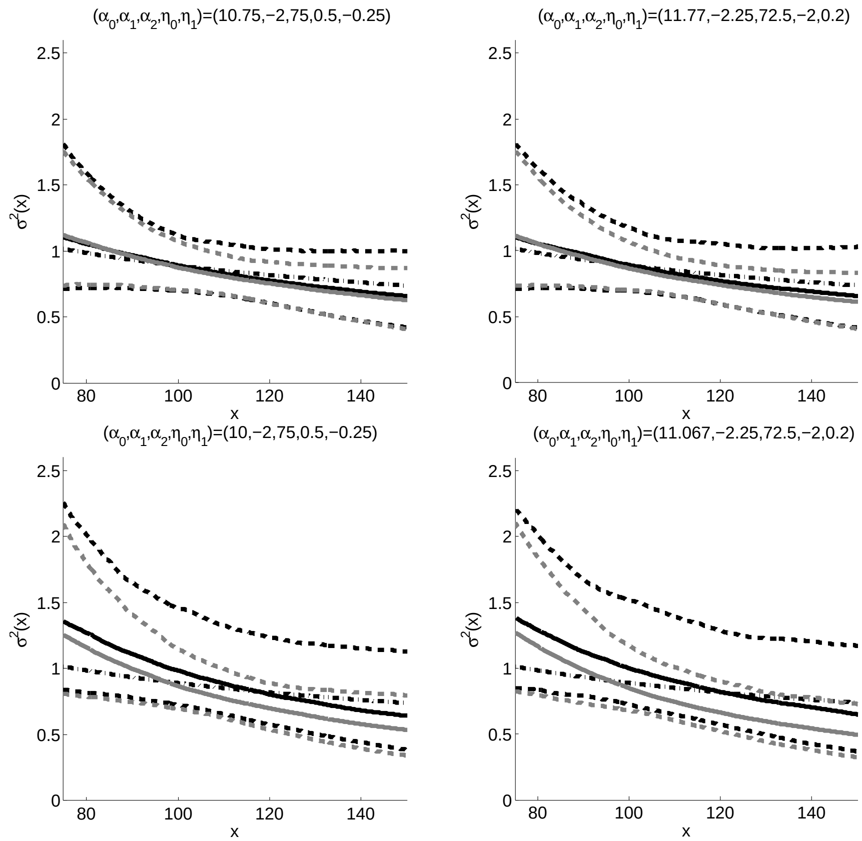

Fig 2. *Comparison of the estimated conditional variances obtained with* $\tilde{\theta}_n$ *(grey curves) and* $\hat{\theta}_n^T$ *(black curves) for model (4.4). Dashed curves :* 5% *and* 95% *quantiles of the empirical distribution based on the estimated conditional variances; solid curve: median of this empirical distribution; dash-dotted curve: true conditional variance curve.*

$\sigma_{\tilde{\theta}_n}(\cdot)$. Obviously, $\widehat{\tilde{\varepsilon}_i^{2*}}$ is also deteriorated, in particular, by the decrease of the upper bound in the integrals of $\hat{m}^0(x)$ and $\hat{\sigma}^0(x)$. This effect seems to be however slighter. The above characteristics also appear on Figure 2 which is constructed similarly to Figure 1.

5. Data analysis

As mentionned in Section 1, we are here interested in the relationship between fatigue life of metal, ceramic or composite materials and applied stress. From

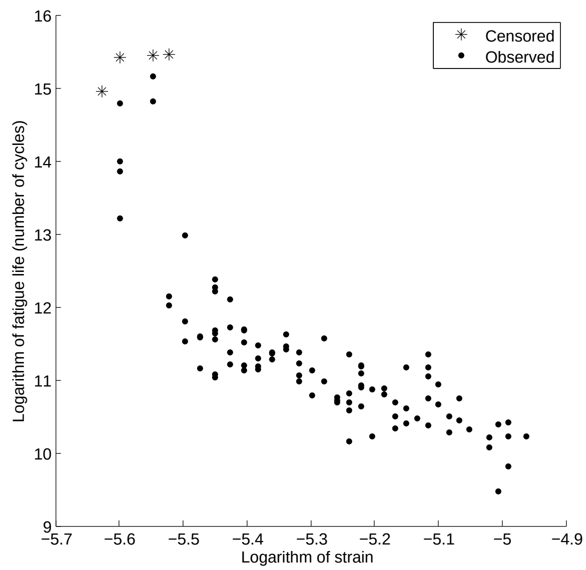

FIG 3. *Fatigue life data. Scatter plot of the logarithm of fatigue life versus the logarithm of strain for specimens of a nickel-base superalloy.*

a long time, an important question in fatigue analysis is to understand how the variability of fatigue life given the stress (or the strain) depends on stress (or strain). Several authors addressed this problem, among others, [13, 14, 15] who studied the number of cycles before failure of nickel-base superalloys as functions of the strain or the pseudostress (Young's modulus times strain). By example, [14] considered model (1.1) with the following form for the conditional standard deviation of the logarithm of the number of cycles before failure:

$$\sigma_{\theta_0}(X) = \exp(\gamma_0 + \gamma_1 \log X). \tag{5.1}$$

However, those authors assumed parametric forms for both $m(X)$ and the error distribution.

We present, in this section, a data set of $n = 246$ specimens of a nickel-base superalloy provided by [16] and studied by [15]. For these data, we consider model (1.1) where Y is the logarithm of the number of cycles before failure and X is the corresponding strain (see Figure 3). A quick graphical check enables to easily observe that the above variance model does not correctly fit the data. As suggested by [15], we only fitted the model (5.1) on the 115 observations for which strain is below 0.007. In addition, since we are interested by the conditional variance shape (whether constant or not), it seems appealing to consider the left part of the data.

The bandwidth is chosen by the bootstrap procedure explained in Section 4.1 among a grid varying from 4.5×10^{-4} to 10.5×10^{-4} by step of 0.3×10^{-4} with a pilot bandwidth equal to 10.8×10^{-4} ($B = 1000$). The estimators for $\theta_0 = (\gamma_0, \gamma_1)$ are respectively $\tilde{\theta}_n = (-28.1264, -5.1639)$ and $\hat{\theta}_n^T = (-29.8660, -5.4759)$.

Figure 4 represents the conditional means estimates $\hat{m}^T(x)$ and $\tilde{m}(x)$ together with their corresponding standard deviations $\sigma_{\hat{\theta}_n^T}(x)$ and $\sigma_{\tilde{\theta}_n}(x)$. Solid curves correspond to $\hat{m}^T(x)$ (black) and $\tilde{m}(x)$ (grey) while dashed curves correspond to $\hat{m}^T(x) \pm \sigma_{\hat{\theta}_n^T}(x)$ (black) and $\tilde{m}(x) \pm \sigma_{\tilde{\theta}_n}(x)$ (grey). Note that the smoothing parameters for $\hat{m}^T(x)$ and $\tilde{m}(x)$ are chosen by the bootstrap method proposed by [7] (with the same pilot bandwidth as above). This graph clearly exhibits a decreasing standard deviation.

Next, confidence intervals are provided for both methods in Table 5. To correctly approximate the distribution of our estimators, a double bootstrap procedure is proposed: in each resample ($B = 1000$), an optimal value for the bandwidth parameter is obtained by a second bootstrap stage (using $B' = 1000$ bootstrap samples for each resample). The first stage resamples are generated with the procedure of Section 4.1 and are also used to obtain the optimal bandwidth for the initial sample. Next, this procedure together with expression (4.2) are adapted to compute each (first stage) resample optimal bandwidth (the pilot bandwidth is kept constant throughout the entire methodology). The confidence intervals are two-sided and their level is 0.95. Both basic and percentile boot-

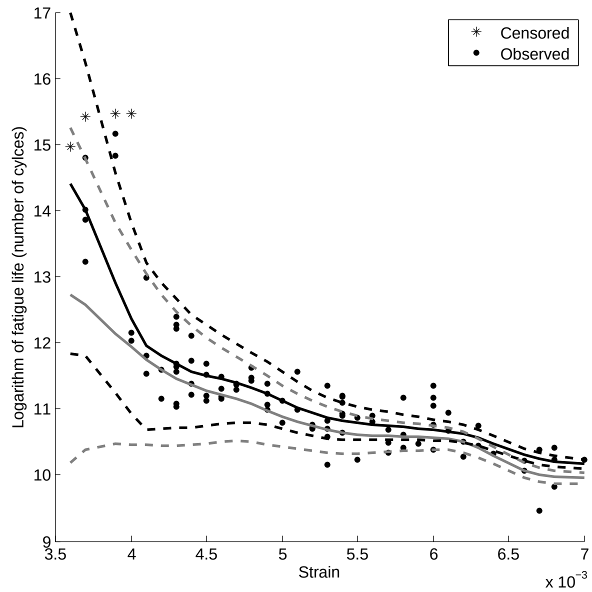

FIG 4. *Fatigue life data. Scatter plot of the fatigue life logarithm versus the strain for specimens of a nickel-base superalloy where the strain domain is restricted to values below* 0.007. *Solid black and grey curves correspond to* $\hat{m}^T(x)$ *and* $\tilde{m}(x)$ *respectively; black and grey dashed curves to* $(\hat{m}^T(x) - \sigma_{\hat{\theta}_n^T}(x); \hat{m}^T(x) + \sigma_{\hat{\theta}_n^T}(x))$ *and* $(\tilde{m}(x) - \sigma_{\tilde{\theta}_n}(x); \tilde{m}(x) + \sigma_{\tilde{\theta}_n}(x))$ *respectively, for* $x \in [0.0036; 0.0070]$.

strap methods are developed. We can observe that the intervals lengths using bootstrap for $\hat{\gamma}_{0,n}^T$ and $\hat{\gamma}_{1,n}^T$ are larger than the ones using bootstrap for $\tilde{\gamma}_{0,n}$ and $\tilde{\gamma}_{1,n}$. 0 is never included in the confidence interval for γ_1 except when using the estimated distribution of $\hat{\gamma}_{1,n}^T$ with the basic bootstrap procedure. Again, that suggests a non constant variance.

In Figures 5 and 6, we construct for a grid of values of x, basic and percentile bootstrap confidence intervals (two-sided, 95%) for $\sigma^2_{\theta_0^T}(x)$, i.e., by using the distributions of $\sigma^2_{\tilde{\theta}_n}(x)$ and $\sigma^2_{\hat{\theta}_n^T}(x)$, each estimated with the B resamples. On the left panel, confidence intervals do not always contain $\sigma^2_{\tilde{\theta}_n}(x)$ (especially for small values of x) while $\sigma^2_{\hat{\theta}_n^T}(x)$ is always included in the corresponding bounds.

Figure 7 compares $\sigma_{\hat{\theta}_n^T}(x)$ and $\sigma_{\tilde{\theta}_n}(x)$ with nonparametric versions $\hat{\sigma}^T(x)$ and

$\tilde{\sigma}(x)$ defined by the square roots of

$$\hat{\sigma}^{T2}(x) = \hat{\sigma}^{02}(x)\left[\int_{-\infty}^{T} y^2 d\hat{F}_{\varepsilon}^0(y) - \left(\int_{-\infty}^{T} y d\hat{F}_{\varepsilon}^0(y)\right)^2\right]$$

and

$$\tilde{\sigma}^2(x) = \int_{-\infty}^{\tilde{T}_x} y^2 d\hat{F}(y|x) - \left(\int_{-\infty}^{\tilde{T}_x} y d\hat{F}(y|x)\right)^2$$

respectively. For these two estimators, the optimal bandwidth is obtained with the method proposed by [7] adapted to the variance case (with the same pilot bandwidth as above). Under the assumed model, once again, the standard deviation seems to be far from a constant, whether for parametric or nonparametric estimations. The proposed parametric model however better fits the data. Since (5.1) is of the form $\gamma_0 x^{\gamma_1}$, goodness-of-fit tests for a conditional standard deviation or for any other scale function can be considered as equivalent (see [3]). The test proposed in this last paper could thus be applied. However, other testing procedures specifically adapted to the conditional standard deviation can be studied as well, for example, using the artificial data points (2.6). The method proposed by [3] to test a constant conditional variance and the form (5.1) leads to $p-$values equal to 0.000 and 0.018 respectively. The null hypothesis with the model (5.1) cannot thus be rejected at the 1 percent level even though the fit is not perfect.

Appendix

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

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

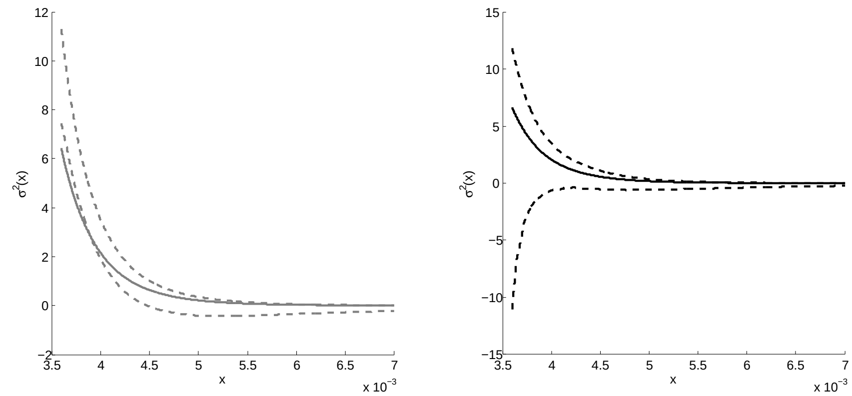

Fig 5. *Fatigue life data. Confidence intervals for* $\sigma^2_{\theta_0^T}(x)$ *based on the basic bootstrap where strain domain is restricted to values below* 0.007*; on the left, constructed with the estimated distribution of* $\sigma^2_{\tilde{\theta}_n}(x)$ *and, on the right, with the estimated distribution of* $\sigma^2_{\hat{\theta}_n^T}(x)$ *(for each value of* x*).*

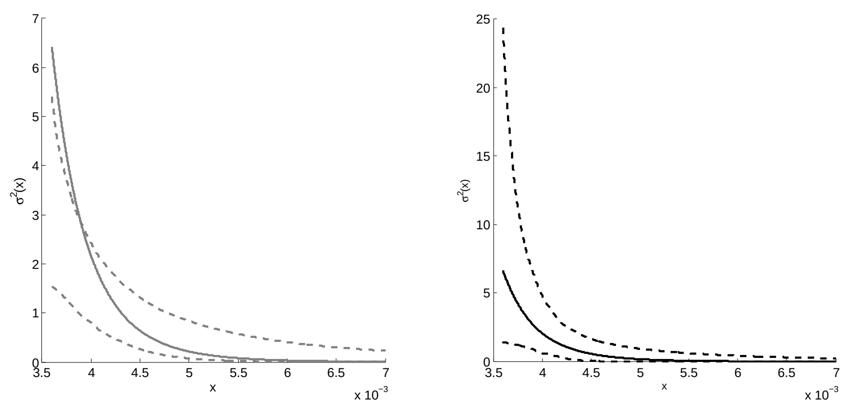

Fig 6. *Fatigue life data. Confidence intervals for* $\sigma^2_{\theta_0^T}(x)$ *based on the percentile bootstrap where strain domain is restricted to values below* 0.007*; on the left, constructed with the estimated distribution of* $\sigma^2_{\tilde{\theta}_n}(x)$ *and, on the right, with the estimated distribution of* $\sigma^2_{\hat{\theta}_n^T}(x)$ *(for each value of* x*).*

Table 3
Fatigue life data. Confidence intervals for γ_0 and γ_1 where strain domain is restricted to values below 0.007. The first line (for γ_0 and γ_1) is obtained with the estimated $\tilde{\theta}_n$ distribution and the second line with the estimated $\hat{\theta}_n^T$ distribution.

	Confidence interval Basic Boot.	Confidence interval Percentile
γ_0	$[-46.8696; -22.4944]$ $[-9.2987; 2.4264]$	$[-33.7584; -9.3832]$ $[-13.3782; -1.6531]$
	Confidence interval Basic Boot.	Confidence interval Percentile
γ_1	$[-8.5922; -4.2020]$ $[-9.2987; 2.4264]$	$[-6.1258; -1.7356]$ $[-13.3782; -1.6531]$

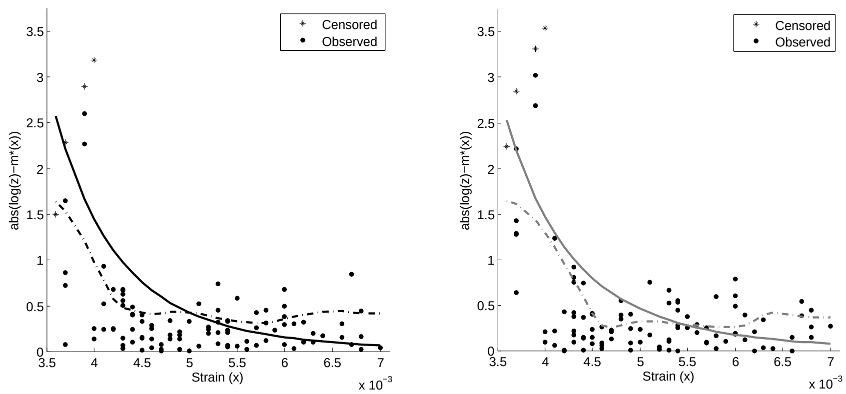

Fig 7. *Fatigue life data. Scatter plot of the residuals of a nonparametric fit versus the strain with $m^*(x) = \hat{m}^T(x)$ for the left panel and $m^*(x) = \tilde{m}(x)$ for the right panel. Solid black and grey curves correspond to $\sigma_{\hat{\theta}_n^T}(x)$ and $\sigma_{\tilde{\theta}_n}(x)$ respectively; dash-dotted black and grey curves represent the nonparametric $\hat{\sigma}^T(x)$ and $\tilde{\sigma}(x)$ respectively for $x \in [0.0036; 0.0070]$.*

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

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

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

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

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

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

$$\pi_0(v_1,v_2) = \frac{I(\delta_1=0)}{1-F_\varepsilon^0(e_{x_1}^{0T}(z_1))}\left\{\left[\frac{\int_{e_{x_1}^{0T}(z_1)}^{T} e\,dF_\varepsilon^0(e)}{1-F_\varepsilon^0(e_{x_1}^{0T}(z_1))} - e_{x_1}^{0T}(z_1)\right]\right.$$
$$\times\ \varphi(v_2, e_{x_1}^{0T}(z_1)) + T\varphi(v_2,T)$$
$$\left. - \int_{e_{x_1}^{0T}(z_1)}^{T} \varphi(v_2,e)de\right\},$$

$$\pi_{00}(v_1,z_2,\delta_2) = I(\delta_1=1, z_1\le T_{x_1})(\eta(z_2,\delta_2|x_1) + z_{1_{x_1}}\zeta(z_2,\delta_2|x_1))$$
$$+I(\delta_1=0)\left[-\frac{e_{x_1}^{0T}(z_1)f_\varepsilon^0(e_{x_1}^{0T}(z_1))}{1-F_\varepsilon^0(e_{x_1}^{0T}(z_1))}I(z_{1_{x_1}}\le T)\right.$$
$$\left.+\frac{f_\varepsilon^0(e_{x_1}^{0T}(z_1))\int_{e_{x_1}^{0T}(z_1)}^{T} e\,dF_\varepsilon^0(e)}{(1-F_\varepsilon^0(e_{x_1}^{0T}(z_1)))^2}\right]$$
$$\times\left[\eta(z_2,\delta_2|x_1) + e_{x_1}^{0T}(z_1)\zeta(z_2,\delta_2|x_1)\right],$$

$$\pi(v_1) = -[\eta(z_1,\delta_1|x_1) + T\zeta(z_1,\delta_1|x_1)]Th_{\varepsilon 1}^0(T|x_1)$$
$$+\int_{R_X}\int_{-\infty}^{+\infty}\sum_{\delta=0,1}\pi_0(y,z,\delta,x_1,z_1,\delta_1)dH_\delta(z|y)dF_X(y)$$
$$+\int_{-\infty}^{+\infty}\sum_{\delta=0,1}\pi_{00}(x_1,z,\delta,z_1,\delta_1)dH_\delta(z|x_1)$$
$$+[z_{1_{x_1}}I(z_{1_{x_1}}\le T,\delta_1=1) + \frac{\int_{e_{x_1}^{0T}(z_1)}^{T} edF_\varepsilon^0(e)}{1-F_\varepsilon^0(e_{x_1}^{0T}(z_1))}I(\delta_1=0)]$$
$$-\int_{-\infty}^{T} edF_\varepsilon^0(e),$$

$$\mathcal{A}_c(x,z) = \int_{e_x^{0T}(z)}^{T}(y^2 - 2y\int_{-\infty}^{T} edF_\varepsilon^0(e))\,dF_\varepsilon^0(y)$$
$$+(1-F_\varepsilon^0(e_x^{0T}(z)))\{\int_{-\infty}^{T} ydF_\varepsilon^0(y)\}^2,$$

$$\chi_k(v_1) = -\frac{\partial\sigma_\theta^2(x_1)}{\partial\theta_k}\sigma^0(x_1)\times$$
$$\sum_{\delta=0,1}\int_{-\infty}^{+\infty}\left\{2\delta(m^T(x_1)-z)[\eta(z_1,\delta_1|x_1) + \int_{-\infty}^{T} edF_\varepsilon^0(e)\zeta(z_1,\delta_1|x_1)]\right.$$

$$
\begin{aligned}
&-(1-\delta)\frac{\sigma^0(x_1)}{1-F_\varepsilon^0(e_{x_1}^{0T}(z))}\left\{\left\{\frac{f_\varepsilon^0(e_{x_1}^{0T}(z))\int_{e_{x_1}^{0T}(z)}^T(y^2-2y\int_{-\infty}^T edF_\varepsilon^0(e))\,dF_\varepsilon^0(y)}{1-F_\varepsilon^0(e_{x_1}^{0T}(z))}\right.\right.\\
&\qquad\left.-[e_{x_1}^{0T}(z)-2\int_{-\infty}^T edF_\varepsilon^0(e)]e_{x_1}^{0T}(z)f_\varepsilon^0(e_{x_1}^{0T}(z))I(z_{x_1}\le T)\right\}\eta(z_1,\delta_1|x_1)\\
&\quad+\left\{\frac{e_{x_1}^{0T}(z)f_\varepsilon^0(e_{x_1}^{0T}(z))\int_{e_{x_1}^{0T}(z)}^T(y^2-2y\int_{-\infty}^T edF_\varepsilon^0(e))\,dF_\varepsilon^0(y)}{1-F_\varepsilon^0(e_{x_1}^{0T}(z))}-2\mathcal{A}_c(x_1,z)\right.\\
&\quad\left.\left.\left.-[e_{x_1}^{0T}(z)-2\int_{-\infty}^T edF_\varepsilon^0(e)](e_{x_1}^{0T}(z))^2f_\varepsilon^0(e_{x_1}^{0T}(z))I(z_{x_1}\le T)\right\}\zeta(z_1,\delta_1|x_1)\right\}\right\}\\
&\hspace{28em}dH_\delta(z|x_1)\\
&+\sum_{\delta=0,1}\int_{R_X}\int_{-\infty}^{+\infty}\frac{\partial\sigma_\theta^2(x)}{\partial\theta_k}\left\{2\delta(m^T(x)-z)\sigma^0(x)\pi(v_1)\right.\\
&\quad+(1-\delta)\frac{\sigma^{02}(x)}{1-F_\varepsilon^0(e_x^{0T}(z))}\left\{\left[\frac{\int_{e_x^{0T}(z)}^T(y^2-2y\int_{-\infty}^T edF_\varepsilon^0(e))\,dF_\varepsilon^0(y)}{1-F_\varepsilon^0(e_x^{0T}(z))}\right.\right.\\
&\qquad\qquad\left.-(e_x^{0T}(z))^2+2e_x^{0T}(z)\int_{-\infty}^T eF_\varepsilon^0(e)\right]\varphi(v_1,e_x^{0T}(z))\\
&\qquad\quad+\left[T^2-2T\int_{-\infty}^T edF_\varepsilon^0(e)\right]\varphi(v_1,T)\\
&\qquad\quad-\int_{e_x^{0T}(z)}^T(2y-2\int_{-\infty}^T edF_\varepsilon^0(e))\varphi(v_1,y)dy\\
&\qquad\quad+2\left[(1-F_\varepsilon^0(e_x^{0T}(z)))\int_{-\infty}^T edF_\varepsilon^0(e)-\int_{e_x^{0T}(z)}^T ydF_\varepsilon^0(y)\right]\\
&\qquad\qquad\left.\left.\times\pi(v_1)\right\}\right\}dH_\delta(z|x)dF_X(x)+o_P(n^{1/2}),
\end{aligned}
$$

where $v_q = (x_q, z_q, \delta_q)$ for all $x_q \in R_X,\ z_q \in I\!R,\ \delta_q = 0,1,\ q = 1,2$. $T = (T_x - m^0(x))/\sigma^0(x)$, $z_x = (z - m^0(x))/\sigma^0(x)$, $e_x^{0T}(z) = z_x \wedge T$, for any $x \in R_X$, $z \in I\!R$ and θ_k is the k^{th} component of θ, $k = 1, \ldots, d$.

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 asymptotic results 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) $R_X = [x_e, x_s]$ is a compact interval of length L_X.

(iii) K is a symmetric density with compact support and K is twice continuously differentiable.

(iv) Ω is non-singular.

$(A2)(i)$ There exist $0 \le s_0 \le s_1 \le 1$ so that $\inf_{x\in R_X} \inf_{s_0\le s\le s_1} f(F^{-1}(s|x)|x) > 0$, where $s_1 \le \inf_x F(\tilde{T}_x|x)$, $s_0 \le \inf\{s \in [0,1]; J(s) \ne 0\}$ and $s_1 \ge \sup\{s \in [0,1]; J(s) \ne 0\}$.

(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 $E[Z^4] < \infty$.

$(A4)(i)$ $\eta(z,\delta|x)$ 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 $y \to P(m^0(X) + e\sigma^0(X) \le y)$ $(y \in I\!\!R)$ 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 $\sup_{x,y} |y^3 L'''(y|x)| < \infty$.

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

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

$(A8)$ Θ is compact and θ_0^T is an interior point of Θ. All partial derivatives of $\sigma_\theta^2(x)$ with respect to the components of θ and x up to order three exist and are continuous in (x,θ) for all x and θ.

$(A9)$ The function $E[\{E(\tilde{\varepsilon}_T^{2*}|X) - \sigma_\theta^2(X)\}^2]$ 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 [17], page 45. From the definition of $\hat{\theta}_n^T$ and condition $(A9)$, it follows that it suffices to show that

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

where

$$S_0(\theta) = E[Var[\tilde{\varepsilon}_T^{2*}|X]] + E[(E[\tilde{\varepsilon}_T^{2*}|X] - \sigma_\theta^2(X))^2].$$

The second statement of Theorem 3.1 then follows immediately from (A.1) together with the consistency of $\hat{\theta}_n^T$. First,

$$\begin{aligned} S_n(\theta) &= \frac{1}{n}\sum_{i=1}^n (\widehat{\tilde{\varepsilon}_{Ti}^{2*}} - \tilde{\varepsilon}_{Ti}^{2*})^2 + \frac{1}{n}\sum_{i=1}^n (\tilde{\varepsilon}_{Ti}^{2*} - \sigma_\theta^2(X_i))^2 \\ &\quad + \frac{2}{n}\sum_{i=1}^n (\widehat{\tilde{\varepsilon}_{Ti}^{2*}} - \tilde{\varepsilon}_{Ti}^{2*})(\tilde{\varepsilon}_{Ti}^{2*} - \sigma_\theta^2(X_i)) \\ &= S_{n1} + S_{n2}(\theta) + S_{n3}(\theta). \end{aligned}$$

S_{n1} and $\sup_\theta |S_{n3}(\theta)|$ are treated by Lemma A.1 while S_{n2} is rewritten as

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

Since $E[Z^4] < \infty$,

$$S_{n21} = E[Var[\tilde{\varepsilon}_T^{2*}|X]] + o(1) \text{ a.s..}$$

Using the fact that $E[\varepsilon^{02}] < \infty$ together with two applications of Theorem 2 of [8] (for $S_{n22}(\theta)$ and $S_{n23}(\theta)$) finishes the proof.

Proof of Theorem 3.2. For some θ_{1n} between $\hat{\theta}_n^T$ and θ_0^T

$$\hat{\theta}_n^T - \theta_0^T = -\left\{\frac{\partial^2 S_n(\theta_{1n})}{\partial\theta\partial\theta'}\right\}^{-1}\frac{\partial S_n(\theta_0^T)}{\partial\theta} = -R_{1n}^{-1}R_{2n}.$$

We have

$$\begin{aligned} R_{2n} &= -\frac{2}{n}\sum_{i=1}^{n}(\widehat{\tilde{\varepsilon}_{Ti}^{2*}} - \tilde{\varepsilon}_{Ti}^{2*})\frac{\partial\sigma^2_{\theta_0^T}(X_i)}{\partial\theta} - \frac{2}{n}\sum_{i=1}^{n}\{\tilde{\varepsilon}_{Ti}^{2*} - \sigma^2_{\theta_0^T}(X_i)\}\frac{\partial\sigma^2_{\theta_0^T}(X_i)}{\partial\theta} \\ &= R_{21n} + R_{22n}, \end{aligned}$$

such that R_{22n} is a sum of i.i.d. random vectors with zero mean (by definition of θ_0^T). For each component j of R_{21n}, we use Lemma A.2 while for R_{1n}, we write

$$\begin{aligned} R_{1n} = -\frac{2}{n}\Bigg\{&\sum_{i=1}^{n}(\widehat{\tilde{\varepsilon}_{Ti}^{2*}} - \tilde{\varepsilon}_{Ti}^{2*})\frac{\partial^2\sigma^2_{\theta_{1n}}(X_i)}{\partial\theta\partial\theta'} + \sum_{i=1}^{n}(\tilde{\varepsilon}_{Ti}^{2*} - \sigma^2_{\theta_{1n}}(X_i))\frac{\partial^2\sigma^2_{\theta_{1n}}(X_i)}{\partial\theta\partial\theta'} \\ &-\sum_{i=1}^{n}\Big(\frac{\partial\sigma^2_{\theta_{1n}}(X_i)}{\partial\theta}\Big)\Big(\frac{\partial\sigma^2_{\theta_{1n}}(X_i)}{\partial\theta'}\Big)\Bigg\} = R_{11n} + R_{12n} + R_{13n}. \end{aligned}$$

Using assumption (A8) and Lemme A.1, we have that each component of R_{11n} is $o_P(1)$. Again using condition (A8),

$$\begin{aligned} R_{1n} &= \frac{2}{n}\sum_{i=1}^{n}\frac{\partial\sigma^2_{\theta_0^T}(X_i)}{\partial\theta}\Big(\frac{\partial\sigma^2_{\theta_0^T}(X_i)}{\partial\theta}\Big)' - \frac{2}{n}\sum_{i=1}^{n}\{\tilde{\varepsilon}^{2*}_{Ti} - \sigma^2_{\theta_0^T}(X_i)\}\frac{\partial^2\sigma^2_{\theta_0^T}(X_i)}{\partial\theta\partial\theta'} + o_P(1) \\ &= 2E\Big[\frac{\partial\sigma^2_{\theta_0^T}(X)}{\partial\theta}\Big(\frac{\partial\sigma^2_{\theta_0^T}(X)}{\partial\theta}\Big)' - \{\tilde{\varepsilon}^{2*}_{T} - \sigma^2_{\theta_0^T}(X)\}\frac{\partial^2\sigma^2_{\theta_0^T}(X)}{\partial\theta\partial\theta'}\Big] + o_P(1) \\ &= 2\Omega + o_P(1). \end{aligned}$$

The result now follows.

Lemma A.1 *Assume (A1) (i)–(iii), (A2) (i), (ii), (A3) (i), (ii),* $E[\varepsilon^{02}] < \infty$, $E[|Z|] < \infty$, *(A4) (i) and (A6). Then,*

$$|\widehat{\tilde{\varepsilon}^{2*}_T} - \tilde{\varepsilon}^{2*}_T| \le (Z^2 + |Z| + 1)O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).$$

where $O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ *is uniform in* X *and* Z, *for* $\Delta = 0, 1$.

Proof. We have

$$\begin{aligned} \widehat{\tilde{\varepsilon}^{2*}_T} - \tilde{\varepsilon}^{2*}_T &= \{(Y - \hat{m}^T(X))^2 - (Y - m^T(X))^2\}\Delta + (1-\Delta) \\ &\quad\times\Bigg\{\frac{\hat{\sigma}^{02}(X)}{1 - \hat{F}^0_\varepsilon(\hat{E}^{0T})}\int_{\hat{E}^{0T}}^{T}(y^2 - 2y\int_{-\infty}^{T} ed\hat{F}^0_\varepsilon(e))\,d\hat{F}^0_\varepsilon(y) \\ &\quad-\frac{\sigma^{02}(X)}{1 - F^0_\varepsilon(E^{0T})}\int_{E^{0T}}^{T}(y^2 - 2y\int_{-\infty}^{T} edF^0_\varepsilon(e))\,dF^0_\varepsilon(y) \\ &\quad+\hat{\sigma}^{02}(X)\{\int_{-\infty}^{T} yd\hat{F}^0_\varepsilon(y)\}^2 - \sigma^{02}(X)\{\int_{-\infty}^{T} ydF^0_\varepsilon(y)\}^2\Bigg\} \\ &= A_u(X,Z,\Delta) + A_c(X,Z,\Delta). \end{aligned}$$

Using Theorem 3.1 of [7],

$$|A_u(X,Z,\Delta)| \le |Y|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}) + O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).$$

For $A_c(X,Z,\Delta)$, write

$$\begin{aligned} A_c(X,Z,\Delta) &= \left\{ \frac{\hat\sigma^{02}(X)-\sigma^{02}(X)}{1-\hat F_\varepsilon^0(\hat E^{0T})}\hat{\mathcal{A}}_c(X,Z) \right. \\ &\quad + \frac{\sigma^{02}(X)(\hat F_\varepsilon^0(\hat E^{0T})-F_\varepsilon^0(E^{0T}))}{(1-\hat F_\varepsilon^0(\hat E^{0T}))(1-F_\varepsilon^0(E^{0T}))}\hat{\mathcal{A}}_c(X,Z) \\ &\quad \left. + \frac{\sigma^{02}(X)}{1-F_\varepsilon^0(E^{0T})}(\hat{\mathcal{A}}_c(X,Z)-\mathcal{A}_c(X,Z)) \right\}(1-\Delta), \end{aligned}$$

where

$$\hat{\mathcal{A}}_c(X,Z) = \int_{\hat E^{0T}}^{T}(y^2-2y\int_{-\infty}^{T} ed\hat F_\varepsilon^0(e))\,d\hat F_\varepsilon^0(y)+(1-\hat F_\varepsilon^0(\hat E^{0T}))\{\int_{-\infty}^{T} yd\hat F_\varepsilon^0(y)\}^2.$$

Using Proposition 4.5 and Corollary 3.2 of [18] together with an order one Taylor development and the fact that $\sup_{y\le T}|yf_\varepsilon^0(y)|<\infty$, coefficients of $\hat{\mathcal{A}}_c(X,Z)$ in the two first terms of $A_c(X,Z,\Delta)$ are

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

Now, using Lemma A.2 of [7] and Lemma A.1 of [5],

$$\begin{aligned} \hat{\mathcal{A}}_c(X,Z) &= \int_{\hat E^{0T}}^{T}(y^2-2y\int_{-\infty}^{T} edF_\varepsilon^0(e))\,d\hat F_\varepsilon^0(y) \\ &\quad +(1-F_\varepsilon^0(E^{0T}))\{\int_{-\infty}^{T} ydF_\varepsilon^0(y)\}^2 \\ &\quad +O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}), \end{aligned}$$

so that

$$\begin{aligned} \hat{\mathcal{A}}_c(X,Z)-\mathcal{A}_c(X,Z) &= \int_{\hat E^{0T}}^{T}(y^2-2y\int_{-\infty}^{T} edF_\varepsilon^0(e))\,d(\hat F_\varepsilon^0(y)-F_\varepsilon^0(y)) \\ &\quad +\int_{\hat E^{0T}}^{E^{0T}}(y^2-2y\int_{-\infty}^{T} edF_\varepsilon^0(e))\,dF_\varepsilon^0(y) \\ &\quad +O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}). \end{aligned}$$

Using integration by parts, Corollary 3.2 and Proposition 4.5 of [18] makes the first term of the right hand side of the above expression bounded by

$$((E^0)^2 + |E^0| + 1)O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}), \tag{A.2}$$

while the second term is rewritten using an order one Taylor development

$$-(\kappa_n^2 - 2\kappa_n \int_{-\infty}^{T} edF_\varepsilon^0(e))f_\varepsilon^0(\kappa_n)(\hat{E}^{0T} - E^{0T}),$$

for κ_n between E^{0T} and $\hat{E}^{0T}$, which can be shown to be bounded by (A.2) using similar calculations. This finishes the proof.

Lemma A.2 *Assume (A1) (i)-(iii), (A2)–(A8). Then,*

$$(1/n)\sum_{i=1}^{n}(\widehat{\tilde{\varepsilon}_{Ti}^{2*}} - \tilde{\varepsilon}_{Ti}^{2*})\frac{\partial\sigma_\theta^2(X_i)}{\partial\theta_k} = (1/n)\sum_{i=1}^{n}\chi_k(V_i) + o_P(n^{-1/2}),\ k = 1, \ldots, d.$$

Proof. Using similar arguments as in Lemma A.1,

$$\begin{aligned} A_u(X, Z, \Delta) \;=\; & 2\Delta(Y - m^T(X))\left\{(m^0(X) - \hat{m}^0(X))\right. \\ & +(\sigma^0(X) - \hat{\sigma}^0(X))\int_{-\infty}^{T} ydF_\varepsilon^0(y) \\ & \left.-\sigma^0(X)(\int_{-\infty}^{T} yd\hat{F}_\varepsilon^0(y) - \int_{-\infty}^{T} ydF_\varepsilon^0(y))\right\} + o_P(n^{-1/2}), \end{aligned} \tag{A.3}$$

and

$$\begin{aligned} A_c(X, Z, \Delta) \;=\; & \left\{\frac{2\sigma^0(X)(\hat{\sigma}^0(X) - \sigma^0(X))}{1 - F_\varepsilon^0(E^{0T})}\mathcal{A}_c(X, Z)\right. \\ & +\frac{\sigma^{02}(X)(\hat{F}_\varepsilon^0(\hat{E}^{0T}) - F_\varepsilon^0(E^{0T}))}{(1 - F_\varepsilon^0(E^{0T}))^2}\mathcal{A}_c(X, Z) \\ & \left.+\frac{\sigma^{02}(X)}{1 - F_\varepsilon^0(E^{0T})}(\hat{\mathcal{A}}_c(X, Z) - \mathcal{A}_c(X, Z))\right\}(1 - \Delta) \\ & +R_{n1}(X, Z, \Delta), \end{aligned} \tag{A.4}$$

where $R_{n1}(X,Z,\Delta)$ is bounded by

$$((E^0)^2+|E^0|+1)o_P(n^{-1/2}). \tag{A.5}$$

Next,

$$\begin{aligned}
\hat{\mathcal{A}}_c(X,Z)-\mathcal{A}_c(X,Z) &= -2[\int_{-\infty}^{T} ed(\hat{F}_\varepsilon^0(e)-F_\varepsilon^0(e))]\int_{E^{0T}}^{T} ydF_\varepsilon^0(y) \\
&\quad +\int_{E^{0T}}^{T}(y^2-2y\int_{-\infty}^{T} edF_\varepsilon^0(e))d(\hat{F}_\varepsilon^0(y)-F_\varepsilon^0(y)) \\
&\quad +[E^{0T}-2\int_{-\infty}^{T} edF_\varepsilon^0(e)]E^{0T} \\
&\qquad \times[\hat{F}_\varepsilon^0(E^{0T})-F_\varepsilon^0(E^{0T})-\hat{F}_\varepsilon^0(\hat{E}^{0T})+F_\varepsilon^0(\hat{E}^{0T})] \\
&\quad -[E^{0T}-2\int_{-\infty}^{T} edF_\varepsilon^0(e)]E^{0T}f_\varepsilon^0(E^{0T})(\hat{E}^{0T}-E^{0T}) \\
&\quad +(F_\varepsilon^0(E^{0T})-\hat{F}_\varepsilon^0(\hat{E}^{0T}))\{\int_{-\infty}^{T} edF_\varepsilon^0(e)\}^2 \\
&\quad +2(1-F_\varepsilon^0(E^{0T}))\int_{-\infty}^{T} edF_\varepsilon^0(e) \\
&\qquad \times\{\int_{-\infty}^{T} yd(\hat{F}_\varepsilon^0(y)-F_\varepsilon^0(y))\} \\
&\quad +R_{n2}(X,Z,\Delta),
\end{aligned} \tag{A.6}$$

where $R_{n2}(X,Z,\Delta)$ is bounded by (A.5). To treat the terms where both $\hat{E}^{0T}$ and E^{0T} are involved (i.e. the second term on the right hand side of (A.4) and the third, fourth and fifth terms on the right hand side of (A.6)), we need to introduce the sum used in the statement of Lemma A.2. More precisely, for the second term of (A.4), we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\sigma^{02}(X_i)(\hat{F}_\varepsilon^0(\hat{E}_i^{0T})-F_\varepsilon^0(E_i^{0T}))}{(1-F_\varepsilon^0(E_i^{0T}))^2}\mathcal{A}_c(X_i,Z_i)\frac{\partial\sigma_\theta^2(X_i)}{\partial\theta_k} \\
=\ &\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\sigma^{02}(X_i)(\hat{F}_\varepsilon^0(\hat{E}_i^{0\hat{T}})-F_\varepsilon^0(E_i^{0T}))}{(1-F_\varepsilon^0(E_i^{0T}))^2}
\end{aligned}$$

$$\times \int_{E_i^{0T}}^{T} (y^2 - 2y\int_{-\infty}^{T} edF_\varepsilon^0(e))\, dF_\varepsilon^0(y) \frac{\partial \sigma_\theta^2(X_i)}{\partial \theta_k}$$
$$+\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\sigma^{02}(X_i)(\hat{F}_\varepsilon^0(\hat{E}_i^{0T}) - F_\varepsilon^0(E_i^{0T}))}{1-F_\varepsilon^0(E_i^{0T})}\{\int_{-\infty}^{T} edF_\varepsilon^0(e)\}^2 \frac{\partial \sigma_\theta^2(X_i)}{\partial \theta_k}$$
$$+R_{n3}, \qquad \text{(A.7)}$$

$k = 1, \dots, d$, and where $\hat{T}_i = \frac{T_{X_i} - \hat{m}^0(X_i)}{\hat{\sigma}^0(X_i)}$ and $\hat{E}_i^{0\hat{T}} = \hat{E}_i^0 \wedge \hat{T}_i$, $i = 1, \dots, n$. It is easily shown that

$$\begin{aligned} R_{n3} &\leq \frac{C}{n}\sum_{i=1}^{n} |\hat{F}_\varepsilon^0(\hat{T}_i) - \hat{F}_\varepsilon^0(T_i)| I(E_i^0 \leq T < \hat{E}_i^0) \\ &\leq O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}) \frac{C}{n}\sum_{i=1}^{n} I(E_i^0 \leq T < \hat{E}_i^0) \end{aligned}$$

for some $C > 0$. When $\hat{E}_i^0 > T$, it holds that $E_i^0 > T\hat{\sigma}^0(X_i)/\sigma^0(X_i) + [\hat{m}^0(X_i) - m^0(X_i)]/\sigma^0(X_i) \geq T - V$, where $V = [\inf_x \sigma^0(x)]^{-1}[\sup_x |\hat{m}^0(x) - m^0(x)| + T\sup_x |\hat{\sigma}^0(x) - \sigma^0(x)|] = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ and hence the above expression is bounded by

$$\begin{aligned} &O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})\, n^{-1}\sum_{i=1}^{n} I(T - V < E_i^0 \leq T) \\ &= O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})\, \{\tilde{H}_\varepsilon^0(T) - \tilde{H}_\varepsilon^0(T - V)\}, \end{aligned}$$

where $\tilde{H}_\varepsilon^0(\cdot)$ is the empirical distribution of E_i^0, $i = 1, \dots, n$. Using the fact that $\tilde{H}_\varepsilon^0(y) - H_\varepsilon^0(y) = O_P(n^{-1/2})$ uniformly in y, the above term is $o_P(n^{-1/2})$. Using similar arguments together with Lemma B.1 of [18], the third and fourth terms on the right hand side of (A.6) are treated as

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\sigma^{02}(X_i)}{1-F_\varepsilon^0(E_i^{0T})}\Bigg\{[E_i^{0T} - 2\int_{-\infty}^{T} edF_\varepsilon^0(e)]E_i^{0T} \\ &\quad\times[\hat{F}_\varepsilon^0(\hat{E}_i^{0\hat{T}}) - F_\varepsilon^0(\hat{E}_i^{0\hat{T}}) - \hat{F}_\varepsilon^0(\hat{E}_i^{0T}) + F_\varepsilon^0(\hat{E}_i^{0T})] \\ &-[E_i^{0T} - 2\int_{-\infty}^{T} edF_\varepsilon^0(e)]E_i^{0T} f_\varepsilon^0(E_i^{0T})(\hat{E}_i^{0T} - E_i^{0T})\Bigg\}\frac{\partial \sigma_\theta^2(X_i)}{\partial \theta_k} + o_P(n^{-1/2}) \end{aligned}$$

$$= -\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\sigma^{02}(X_i)}{1-F_\varepsilon^0(E_i^{0T})}[E_i^{0T}-2\int_{-\infty}^{T}edF_\varepsilon^0(e)]E_i^{0T}f_\varepsilon^0(E_i^{0T})(\hat{E}_i^0-E_i^0)$$
$$\times(E_i^0\le T)\frac{\partial\sigma_\theta^2(X_i)}{\partial\theta_k}+o_P(n^{-1/2}),\quad k=1,\ldots,d. \tag{A.8}$$

Finally, together with (A.3), (A.4), (A.6), (A.7) and (A.8), and Lemma A.3, we obtain

$$(1/n)\sum_{i=1}^{n}(\widehat{\tilde{\varepsilon}_{Ti}^{2*}}-\tilde{\varepsilon}_{Ti}^{2*})\frac{\partial\sigma_\theta^2(X_i)}{\partial\theta_k}$$
$$=\frac{-1}{n^2a_n}\sum_{i=1}^{n}\sum_{j=1}^{n}K(\frac{X_i-X_j}{a_n})\frac{\partial\sigma_\theta^2(X_i)}{\partial\theta_k}f_X^{-1}(X_i)\sigma^0(X_i)$$
$$\times\Bigg\{2\Delta_i(m^T(X_i)-Z_i)[\eta(Z_j,\Delta_j|X_i)+\int_{-\infty}^{T}edF_\varepsilon^0(e)\zeta(Z_j,\Delta_j|X_i)]$$
$$-(1-\Delta_i)\frac{\sigma^0(X_i)}{1-F_\varepsilon^0(E_i^{0T})}\Bigg\{\Bigg\{\frac{f_\varepsilon^0(E_i^{0T})\int_{E_i^{0T}}^{T}(y^2-2y\int_{-\infty}^{T}edF_\varepsilon^0(e))\,dF_\varepsilon^0(y)}{1-F_\varepsilon^0(E_i^{0T})}$$
$$-[E_i^{0T}-2\int_{-\infty}^{T}edF_\varepsilon^0(e)]E_i^{0T}f_\varepsilon^0(E_i^{0T})I(E_i^0\le T)\Bigg\}\eta(Z_j,\Delta_j|X_i)$$
$$+\Bigg\{\frac{E_i^{0T}f_\varepsilon^0(E_i^{0T})\int_{E_i^{0T}}^{T}(y^2-2y\int_{-\infty}^{T}edF_\varepsilon^0(e))\,dF_\varepsilon^0(y)}{1-F_\varepsilon^0(E_i^{0T})}-2\mathcal{A}_c(X_i,Z_i)$$
$$-[E_i^{0T}-2\int_{-\infty}^{T}edF_\varepsilon^0(e)](E_i^{0T})^2f_\varepsilon^0(E_i^{0T})I(E_i^0\le T)\Bigg\}\zeta(Z_j,\Delta_j|X_i)\Bigg\}\Bigg\}$$
$$+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial\sigma_\theta^2(X_i)}{\partial\theta_k}\big\{2\Delta_i(m^T(X_i)-Z_i)\sigma^0(X_i)\pi(V_j)$$
$$+(1-\Delta_i)\frac{\sigma^{02}(X_i)}{1-F_\varepsilon^0(E_i^{0T})}\Bigg\{\Bigg[\frac{\int_{E_i^{0T}}^{T}(y^2-2y\int_{-\infty}^{T}edF_\varepsilon^0(e))\,dF_\varepsilon^0(y)}{1-F_\varepsilon^0(E_i^{0T})}$$
$$-(E_i^{0T})^2+2E_i^{0T}\int_{-\infty}^{T}eF_\varepsilon^0(e)\Bigg]\varphi(X_j,Z_j,\Delta_j,E_i^{0T})$$
$$+\Bigg[T^2-2T\int_{-\infty}^{T}edF_\varepsilon^0(e)\Bigg]\varphi(X_j,Z_j,\Delta_j,T)$$
$$-\int_{E_i^{0T}}^{T}(2y-2\int_{-\infty}^{T}edF_\varepsilon^0(e))\varphi(X_j,Z_j,\Delta_j,y)dy$$

$$+2\left[(1-F_\varepsilon^0(E_i^{0T}))\int_{-\infty}^T edF_\varepsilon^0(e)-\int_{E_i^{0T}}^T ydF_\varepsilon^0(y)\right]$$
$$\times\pi(V_j)\}\}+o_P(n^{1/2}).$$

Finally, usual calculations on U-statistics (see by example [5]) finish the proof.

Lemma A.3 *Assume (A1) (i)-(iii), (A2), (A3) (i)-(ii),* $E[\varepsilon^{02}]<\infty$, $E[Z^2]<\infty$, *(A4)–(A7). Then*

$$\int_{-\infty}^T ed\hat{F}_\varepsilon^0(e)-\int_{-\infty}^T edF_\varepsilon^0(e)=(1/n)\sum_{i=1}^n\pi(V_i)+o_P(n^{-(1/2)}).$$

Proof. This result is easily obtained by using the proofs of Lemma A.1 to A.3 of [7], the asymptotic representation of the residuals distribution given in Theorem 3.1 of [18] and simple calculations on U-statistics.

References

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

[2] Dai, J. and Sperlich, S. (2010). Simple and effective boundary correction for kernel densities and regression with an application to the world income and Engel curve estimation. *Comput. Statist. Data Anal.*, **54**, 2487–2497.

[3] Dette, H. and Heuchenne, C. (2012). Scale checks in censored regression. *Scand. J. Statist.*, **39**, 323–339.

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

[5] Heuchenne, C. and Van Keilegom, I. (2007). Polynomial regression with censored data based on preliminary nonparametric estimation. *Ann. Instit. Statist. Math.*, **59**, 273–298.

[6] Heuchenne, C. and Van Keilegom, I. (2008). Nonlinear regression with censored data. *Technometrics*, **49**, 34–44.

[7] Heuchenne, C. and Van Keilegom, I. (2010). Estimation in nonparametric location-scale regression models with censored data. *Ann. Inst. Statist. Math.*, **62**, 439–463.

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

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

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

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

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

[13] Nelson, W. (1984). Fitting of fatigue curves with nonconstant standard deviation to data with runouts. *J. Testing Eval.*, **12**,

[14] 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.

[15] Pascual, F.G. and Meeker, W.Q. (1999). Estimating fatigue curves with the random fatigue-limit model. *Technometrics*, **4**, 277–290.

[16] Shen, C.L. (1994). *Statistical analysis of fatigue data.* Unpublished Ph.D. dissertation, University of Arizona, Department of Aerospace and Mechanical Engineering.

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

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

[19] Van Keilegom, I. and Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst.*

Statist. Math., **49**, 467–491.

[20] Wei, S. X. (2002). A censored-GARCH model of asset returns with price limits. *Journal of Empirical Finance*, **9**, 197–223.