Conditional variance estimation in censored regression models

Cédric Heuchenne*

QuantOM,†

HEC-Management School of the University of Liège

University of Liège

June 1, 2010

Abstract

Suppose the random vector (X, Y) satisfies the regression model $Y = m(X) + \sigma(X)\varepsilon$, where $m(\cdot) = E(Y|\cdot)$, $\sigma^2(\cdot) = \mathrm{Var}(Y|\cdot)$ 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.

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

*This research was supported by IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy)

†Centre for Quantitative Methods and Operations Management

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 Wei (2002), 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. Suppose also that Y is subject to random right censoring, i.e. instead of observing Y, we only observe (Z, Δ), where $Z = \min(Y, C)$, $\Delta = I(Y \le C)$ 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).

The objective is to extend classical least squares procedures to censored data in order to estimate $\sigma_{\theta_0}(\cdot)$. If a lot of work was devoted to polynomial estimation of the regression function for censored data (see e.g. Heuchenne and Van Keilegom (2007a) for a literature overview), much less work was achieved for the estimation of the conditional variance. In fact, model (1.1) was already considered in fatigue curve analysis (Nelson, 1984, Pascual and Meeker, 1997) but with parametric forms for $m(\cdot)$ and the distribution of ε. Since choices for those forms can considerably influence inference results on $\sigma_{\theta_0}(\cdot)$, it can be important to consider its estimation without any parametric constraint on the other quantities of model (1.1). In the same idea, Heuchenne and Van Keilegom (2007b) 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.

In this paper we 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 and 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

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

and note that $E((Y_i - m(X_i))^2|X_i) = E(\tilde{\varepsilon}_i^{2*}|X_i) = \sigma^2_{\theta_0}(X_i)$. 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 $m^0(x) = T(F(\cdot|x))$ 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

$F(y|x) = P(Y \leq y|x)$, the response conditional distribution,

$G(y|x) = P(C \leq y|x)$, the censoring conditional distribution,

$$H(y|x) = P(Z \leq y|x)\ (H(y) = P(Z \leq y)),\ \text{the observable (un)conditional distribution,}$$

$$H_\delta(y|x) = P(Z \leq y, \Delta = \delta|x),\ \text{the observable conditional subdistributions for } \delta = 0, 1,$$

$$F_\varepsilon^0(y) = P(\varepsilon^0 \leq y),\ S_\varepsilon^0(y) = 1 - F_\varepsilon^0(y),\ \text{the distribution and survival functions of } \varepsilon^0,$$

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

It is easily seen that

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

for any location function $m^0(\cdot)$ and scale function $\sigma^0(\cdot)$. 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 Van Keilegom and Veraverbeke (1997)), 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 $F^{-1}(s|x) = \inf\{y; F(y|x) \geq s\}$ is the quantile function of Y given x and $J(s)$ is a given score function satisfying $\int_0^1 J(s) \, ds = 1$. When $J(s)$ is chosen appropriately (namely put to zero in the right tail, there where the quantile function cannot be estimated in a consistent way due to the right censoring), $m^0(x)$ 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):

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

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

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

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). Finally, $m(x)$ is estimated by the method of Heuchenne and Van Keilegom (2008) applied to the estimation of a conditional mean:

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

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

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

$$\begin{aligned}\widehat{\tilde{\varepsilon}_{Ti}^{2*}} = (Y_i - \hat{m}^T(X_i))^2 \Delta_i + \Big\{ \frac{\hat{\sigma}^{02}(X_i)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})} \int_{\hat{E}_i^{0T}}^T (y^2 - 2y \int_{-\infty}^T e d\hat{F}_\varepsilon^0(e))\, d\hat{F}_\varepsilon^0(y) \\ + \hat{\sigma}^{02}(X_i) \{ \int_{-\infty}^T y d\hat{F}_\varepsilon^0(y) \}^2 \Big\} (1 - \Delta_i),\end{aligned} \tag{2.6}$$

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

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

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 (A9), where

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

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

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 $\tau_{F_\varepsilon^0} \leq \tau_{G_\varepsilon^0}$.

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 $m^0(x) + \sigma^0(x)\tau_{H_\varepsilon^0} \geq \tau_{H(\cdot|x)}$ for any value of x such that consistent areas of (2.5) can be substantially larger than for local estimators (see Heuchenne and Van Keilegom, 2008, for a complete discussion).

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

$$S_n(\hat{\theta}_n^T) = E[Var[\tilde{\varepsilon}_T^{2*}|X]] + E[(E[\tilde{\varepsilon}_T^{2*}|X] - \sigma_\theta^2(X))^2] + o_P(1).$$

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

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

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

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

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

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.

Appendix

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

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

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

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

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

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

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

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

$$\pi_0(v_1,v_2) = \frac{I(\delta_1=0)}{1-F^0_\varepsilon(e^{0T}_{x_1}(z_1))}\left\{\left[\frac{\int_{e^{0T}_{x_1}(z_1)}^T e\,dF^0_\varepsilon(e)}{1-F^0_\varepsilon(e^{0T}_{x_1}(z_1))} - e^{0T}_{x_1}(z_1)\right]\right.$$

$$\times\ \varphi(v_2, e^{0T}_{x_1}(z_1)) + T\varphi(v_2,T)$$

$$\left. - \int_{e^{0T}_{x_1}(z_1)}^T \varphi(v_2,e)de\right\},$$

$$\pi_{00}(v_1,z_2,\delta_2) = I(\delta_1=1, z_1\le T_{x_1})(\eta(z_2,\delta_2|x_1) + z_{1_{x_1}}\zeta(z_2,\delta_2|x_1))$$

$$+I(\delta_1=0)\left[-\frac{e^{0T}_{x_1}(z_1)f^0_\varepsilon(e^{0T}_{x_1}(z_1))}{1-F^0_\varepsilon(e^{0T}_{x_1}(z_1))}I(z_{1_{x_1}}\le T)\right.$$

$$\left.+\frac{f^0_\varepsilon(e^{0T}_{x_1}(z_1))\int_{e^{0T}_{x_1}(z_1)}^T e\,dF^0_\varepsilon(e)}{(1-F^0_\varepsilon(e^{0T}_{x_1}(z_1)))^2}\right]$$

$$\times\left[\eta(z_2,\delta_2|x_1) + e^{0T}_{x_1}(z_1)\zeta(z_2,\delta_2|x_1)\right],$$

$$\pi(v_1) = -[\eta(z_1,\delta_1|x_1) + T\zeta(z_1,\delta_1|x_1)]Th^0_{\varepsilon 1}(T|x_1)$$

$$+\int_{R_X}\int_{-\infty}^{+\infty}\sum_{\delta=0,1}\pi_0(y,z,\delta,x_1,z_1,\delta_1)dH_\delta(z|y)dF_X(y)$$

$$+\int_{-\infty}^{+\infty}\sum_{\delta=0,1}\pi_{00}(x_1,z,\delta,z_1,\delta_1)dH_\delta(z|x_1)$$

$$+[z_{1_{x_1}}I(z_{1_{x_1}}\le T,\delta_1=1) + \frac{\int_{e^{0T}_{x_1}(z_1)}^T e dF^0_\varepsilon(e)}{1-F^0_\varepsilon(e^{0T}_{x_1}(z_1))}I(\delta_1=0)] - \int_{-\infty}^T e dF^0_\varepsilon(e),$$

$$\mathcal{A}_c(x,z) = \int_{e^{0T}_x(z)}^T (y^2 - 2y\int_{-\infty}^T e dF^0_\varepsilon(e))\,dF^0_\varepsilon(y) + (1-F^0_\varepsilon(e^{0T}_x(z)))\{\int_{-\infty}^T y dF^0_\varepsilon(y)\}^2,$$

$$\chi_k(v_1) = -\frac{\partial\sigma^2_\theta(x_1)}{\partial\theta_k}\sigma^0(x_1)\times$$

$$\sum_{\delta=0,1}\int_{-\infty}^{+\infty}\left\{2\delta(m^T(x_1)-z)[\eta(z_1,\delta_1|x_1) + \int_{-\infty}^T e dF^0_\varepsilon(e)\zeta(z_1,\delta_1|x_1)]\right.$$

$$-(1-\delta)\frac{\sigma^0(x_1)}{1-F^0_\varepsilon(e^{0T}_{x_1}(z))}\left\{\left\{\frac{f^0_\varepsilon(e^{0T}_{x_1}(z))\int_{e^{0T}_{x_1}(z)}^T(y^2-2y\int_{-\infty}^T e dF^0_\varepsilon(e))\,dF^0_\varepsilon(y)}{1-F^0_\varepsilon(e^{0T}_{x_1}(z))}\right.\right.$$

$$\left.-[e^{0T}_{x_1}(z) - 2\int_{-\infty}^T e dF^0_\varepsilon(e)]e^{0T}_{x_1}(z)f^0_\varepsilon(e^{0T}_{x_1}(z))I(z_{x_1}\le T)\right\}\eta(z_1,\delta_1|x_1)$$

$$+\left\{\frac{e^{0T}_{x_1}(z)f^0_\varepsilon(e^{0T}_{x_1}(z))\int_{e^{0T}_{x_1}(z)}^T(y^2-2y\int_{-\infty}^T e dF^0_\varepsilon(e))\,dF^0_\varepsilon(y)}{1-F^0_\varepsilon(e^{0T}_{x_1}(z))} - 2\mathcal{A}_c(x_1,z)\right.$$

$$\left.\left.\left.-[e^{0T}_{x_1}(z) - 2\int_{-\infty}^T e dF^0_\varepsilon(e)](e^{0T}_{x_1}(z))^2 f^0_\varepsilon(e^{0T}_{x_1}(z))I(z_{x_1}\le T)\right\}\zeta(z_1,\delta_1|x_1)\right\}\right\}dH_\delta(z|x_1)$$

$$
\begin{aligned}
&+ \sum_{\delta=0,1} \int_{R_X} \int_{-\infty}^{+\infty} \frac{\partial \sigma^2_\theta(x)}{\partial \theta_k} \Big\{ 2\delta (m^T(x) - z)\sigma^0(x)\pi(v_1) \\
&\quad +(1-\delta) \frac{\sigma^{02}(x)}{1 - F^0_\varepsilon(e^{0T}_x(z))} \Bigg\{ \Bigg[\frac{\int_{e^{0T}_x(z)}^T (y^2 - 2y \int_{-\infty}^T e dF^0_\varepsilon(e))\, dF^0_\varepsilon(y)}{1 - F^0_\varepsilon(e^{0T}_x(z))} \\
&\qquad\qquad - (e^{0T}_x(z))^2 + 2 e^{0T}_x(z) \int_{-\infty}^T e F^0_\varepsilon(e) \Bigg] \varphi(v_1, e^{0T}_x(z)) \\
&\qquad + \left[T^2 - 2T \int_{-\infty}^T e dF^0_\varepsilon(e) \right] \varphi(v_1, T) \\
&\qquad - \int_{e^{0T}_x(z)}^T (2y - 2 \int_{-\infty}^T e dF^0_\varepsilon(e)) \varphi(v_1, y) dy \\
&\qquad + 2 \left[(1 - F^0_\varepsilon(e^{0T}_x(z))) \int_{-\infty}^T e dF^0_\varepsilon(e) - \int_{e^{0T}_x(z)}^T y dF^0_\varepsilon(y) \right] \\
&\qquad\qquad \times \pi(v_1) \Bigg\} \Big\}\, dH_\delta(z|x) dF_X(x) + o_P(n^{1/2}),
\end{aligned}
$$

where $v_q = (x_q, z_q, \delta_q)$ for all $x_q \in R_X$, $z_q \in I\!R$, $\delta_q = 0, 1$, $q = 1, 2$. $T = (T_x - m^0(x))/\sigma^0(x)$, $z_x = (z - m^0(x))/\sigma^0(x)$, $e^{0T}_x(z) = z_x \wedge T$, 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 $0 \le s_0 \le s_1 \le 1$ 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 $E[Z^4] < \infty$.

$(A4)(i)$ $\eta(z,\delta|x)$ 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 $L(y|x) = H(y|x), H_1(y|x), H_\varepsilon^0(y|x)$ or $H_{\varepsilon 1}^0(y|x)$: $L'(y|x)$ is continuous in (x,y) and $\sup_{x,y} |y^2 L'(y|x)| < \infty$. 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 van der Vaart (1998, p. 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

$$S_{n2}(\theta) = \frac{1}{n}\sum_{i=1}^n (\tilde{\varepsilon}_{Ti}^{2*} - E[\tilde{\varepsilon}_{Ti}^{2*}|X_i])^2 + \frac{1}{n}\sum_{i=1}^n (E[\tilde{\varepsilon}_{Ti}^{2*}|X_i] - \sigma_\theta^2(X_i))^2$$

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

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 Jennrich (1969) (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

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

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

$$R_{1n}=-\frac{2}{n}\left\{\sum_{i=1}^{n}(\widehat{\tilde{\varepsilon}_{Ti}^{2*}}-\tilde{\varepsilon}_{Ti}^{2*})\frac{\partial^2\sigma_{\theta_{1n}}(X_i)}{\partial\theta\partial\theta'}+\sum_{i=1}^{n}(\tilde{\varepsilon}_{Ti}^{2*}-\sigma_{\theta_{1n}}(X_i))\frac{\partial^2\sigma_{\theta_{1n}}(X_i)}{\partial\theta\partial\theta'}\right.$$
$$\left.-\sum_{i=1}^{n}\Big(\frac{\partial\sigma_{\theta_{1n}}(X_i)}{\partial\theta}\Big)\Big(\frac{\partial\sigma_{\theta_{1n}}(X_i)}{\partial\theta'}\Big)\right\}=R_{11n}+R_{12n}+R_{13n}.$$

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

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

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,*

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

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

$$
\begin{aligned}
\widehat{\tilde\varepsilon_T^{2*}} - \tilde\varepsilon_T^{2*} &= \{(Y-\hat m^T(X))^2 - (Y-m^T(X))^2\}\Delta + (1-\Delta) \\
&\quad \times \left\{ \frac{\hat\sigma^{02}(X)}{1-\hat F_\varepsilon^0(\hat E^{0T})} \int_{\hat E^{0T}}^T (y^2 - 2y \int_{-\infty}^T e d\hat F_\varepsilon^0(e))\, d\hat F_\varepsilon^0(y) + \hat\sigma^{02}(X)\{\int_{-\infty}^T y d\hat F_\varepsilon^0(y)\}^2 \right. \\
&\quad \left. - \frac{\sigma^{02}(X)}{1-F_\varepsilon^0(E^{0T})} \int_{E^{0T}}^T (y^2 - 2y \int_{-\infty}^T e dF_\varepsilon^0(e))\, dF_\varepsilon^0(y) - \sigma^{02}(X)\{\int_{-\infty}^T y dF_\varepsilon^0(y)\}^2 \right\} \\
&= A_u(X,Z,\Delta) + A_c(X,Z,\Delta).
\end{aligned}
$$

Using Theorem 3.1 of Heuchenne and Van Keilegom (2009),

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

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

$$
\begin{aligned}
A_c(X,Z,\Delta) &= \left\{ \frac{\hat\sigma^{02}(X) - \sigma^{02}(X)}{1-\hat F_\varepsilon^0(\hat E^{0T})} \hat{\mathcal{A}}_c(X,Z) \right. \\
&\quad + \frac{\sigma^{02}(X)(\hat F_\varepsilon^0(\hat E^{0T}) - F_\varepsilon^0(E^{0T}))}{(1-\hat F_\varepsilon^0(\hat E^{0T}))(1-F_\varepsilon^0(E^{0T}))} \hat{\mathcal{A}}_c(X,Z) \\
&\quad \left. + \frac{\sigma^{02}(X)}{1-F_\varepsilon^0(E^{0T})} (\hat{\mathcal{A}}_c(X,Z) - \mathcal{A}_c(X,Z)) \right\} (1-\Delta), \qquad \text{(A.2)}
\end{aligned}
$$

where

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

Using Proposition 4.5 and Corollary 3.2 of Van Keilegom and Akritas (1999) together with an order one Taylor development and the fact that $\sup_y |y f_\varepsilon^0(y)| < \infty$, 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 Heuchenne and Van Keilegom (2010) and Lemma A.1 of Heuchenne and Van Keilegom (2007a),

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

such that

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

Using integration by parts, Corollary 3.2 and Proposition 4.5 of Van Keilegom and Akritas (1999) makes the first term of the right hand side of the above expression bounded by

$$(E^{0^2} + |E^0| + 1)O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}), \qquad \text{(A.3)}$$

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

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

for κ_n between E^{0T} and $\hat{E}^{0T}$, which can be shown to be bounded by (A.3) 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,

$$\begin{aligned} A_u(X, Z, \Delta) \ &= \ 2\Delta(Y - m^T(X)) \Big\{ (m^0(X) - \hat{m}^0(X)) + (\sigma^0(X) - \hat{\sigma}^0(X)) \int_{-\infty}^T y dF_\varepsilon^0(y) \\ &\quad -\sigma^0(X)(\int_{-\infty}^T y d\hat{F}_\varepsilon^0(y) - \int_{-\infty}^T y dF_\varepsilon^0(y)) \Big\} + o_P(n^{-1/2}), \qquad \text{(A.4)} \end{aligned}$$

and

$$\begin{aligned} A_c(X, Z, \Delta) \ &= \ \Big\{ \frac{2\sigma^0(X)(\hat{\sigma}^0(X) - \sigma^0(X))}{1 - F_\varepsilon^0(E^{0T})} \mathcal{A}_c(X, Z) \\ &\quad + \frac{\sigma^{02}(X)(\hat{F}_\varepsilon^0(\hat{E}^{0T}) - F_\varepsilon^0(E^{0T}))}{(1 - F_\varepsilon^0(E^{0T}))^2} \mathcal{A}_c(X, Z) \\ &\quad + \frac{\sigma^{02}(X)}{1 - F_\varepsilon^0(E^{0T})} (\hat{\mathcal{A}}_c(X, Z) - \mathcal{A}_c(X, Z)) \Big\} (1 - \Delta) + R_{n1}(X, Z, \Delta), \text{(A.5)} \end{aligned}$$

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

$$(E^{0^2} + |E^0| + 1)o_P(n^{-1/2}). \qquad \text{(A.6)}$$

Next,

$$\begin{aligned} \hat{\mathcal{A}}_c(X, Z) - \mathcal{A}_c(X, Z) \ &= \ -2[\int_{-\infty}^T e d(\hat{F}_\varepsilon^0(e) - F_\varepsilon^0(e))] \int_{E^{0T}}^T y dF_\varepsilon^0(y) \\ &\quad + \int_{E^{0T}}^T (y^2 - 2y \int_{-\infty}^T e dF_\varepsilon^0(e)) d(\hat{F}_\varepsilon^0(y) - F_\varepsilon^0(y)) \end{aligned}$$

$$+[E^{0T} - 2\int_{-\infty}^{T} e dF_\varepsilon^0(e)]E^{0T}$$

$$\times[\hat{F}_\varepsilon^0(E^{0T}) - F_\varepsilon^0(E^{0T}) - \hat{F}_\varepsilon^0(\hat{E}^{0T}) + F_\varepsilon^0(\hat{E}^{0T})]$$

$$-[E^{0T} - 2\int_{-\infty}^{T} e dF_\varepsilon^0(e)]E^{0T} f_\varepsilon^0(E^{0T})(\hat{E}^{0T} - E^{0T})$$

$$+(F_\varepsilon^0(E^{0T}) - \hat{F}_\varepsilon^0(\hat{E}^{0T}))\{\int_{-\infty}^{T} e dF_\varepsilon^0(e)\}^2$$

$$+2(1 - F_\varepsilon^0(E^{0T}))\int_{-\infty}^{T} e dF_\varepsilon^0(e)\{\int_{-\infty}^{T} y d(\hat{F}_\varepsilon^0(y) - F_\varepsilon^0(y))\}$$

$$+R_{n2}(X, Z, \Delta), \tag{A.7}$$

where $R_{n2}(X, Z, \Delta)$ is bounded by (A.6). 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.5) and the third, fourth and fifth terms on the right hand side of (A.7)), we need to introduce the sum used in the statement of Lemma A.2. More precisely, for the second term of (A.5), we have

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

$$= \frac{1}{n}\sum_{i=1}^{n}(1 - \Delta_i)\frac{\sigma^{02}(X_i)(\hat{F}_\varepsilon^0(\hat{E}_i^{0\hat{T}}) - F_\varepsilon^0(E_i^{0T}))}{(1 - F_\varepsilon^0(E_i^{0T}))^2}$$

$$\times\int_{E_i^{0T}}^{T}(y^2 - 2y\int_{-\infty}^{T} e dF_\varepsilon^0(e))\, dF_\varepsilon^0(y)\frac{\partial\sigma_\theta^2(X_i)}{\partial\theta_k}$$

$$+\frac{1}{n}\sum_{i=1}^{n}(1 - \Delta_i)\frac{\sigma^{02}(X_i)(\hat{F}_\varepsilon^0(\hat{E}_i^{0T}) - F_\varepsilon^0(E_i^{0T}))}{1 - F_\varepsilon^0(E_i^{0T})}\{\int_{-\infty}^{T} e dF_\varepsilon^0(e)\}^2\frac{\partial\sigma_\theta^2(X_i)}{\partial\theta_k}$$

$$+R_{n3}, \tag{A.8}$$

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

$$R_{n3} \le \frac{C}{n}\sum_{i=1}^{n}|\hat{F}_\varepsilon^0(\hat{T}_i) - \hat{F}_\varepsilon^0(T_i)|I(E_i^0 \le T < \hat{E}_i^0)$$

$$\le O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})\frac{C}{n}\sum_{i=1}^{n}I(E_i^0 \le T < \hat{E}_i^0)$$

for some $C > 0$. When $\hat{E}_i^0 > T$, it holds that $E_i^0 > T\hat{\sigma}^0(X_i)/\sigma^0(X_i) + [\hat{m}^0(X_i) - m^0(X_i)]/\sigma^0(X_i) \ge T - V$, 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

$$O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})\, n^{-1}\sum_{i=1}^{n}I(T - V < E_i^0 \le T)$$

$$= O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})\,\{\tilde{H}_\varepsilon^0(T) - \tilde{H}_\varepsilon^0(T - V)\},$$

where $\tilde{H}_\varepsilon^0(\cdot)$ is the empirical distribution of E_i^0, $i = 1, \ldots, n$. Using the fact that $\tilde{H}_\varepsilon^0(y) - H_\varepsilon^0(y) = O_P(n^{-1/2})$ uniformly in y, the above term is $o_P(n^{-1/2})$. Using similar arguments together with Lemma B.1 of Van Keilegom and Akritas (1999), the third and fourth terms on the right hand side of (A.7) are treated as

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n (1-\Delta_i)\frac{\sigma^{02}(X_i)}{1-F_\varepsilon^0(E_i^{0T})}\Big\{[E_i^{0T} - 2\int_{-\infty}^T edF_\varepsilon^0(e)]E_i^{0T} \\
&\qquad \times[\hat{F}_\varepsilon^0(\hat{E}_i^{0\hat{T}}) - F_\varepsilon^0(\hat{E}_i^{0\hat{T}}) - \hat{F}_\varepsilon^0(\hat{E}_i^{0T}) + F_\varepsilon^0(\hat{E}_i^{0T})] \\
&\quad -[E_i^{0T} - 2\int_{-\infty}^T edF_\varepsilon^0(e)]E_i^{0T} f_\varepsilon^0(E_i^{0T})(\hat{E}_i^{0T} - E_i^{0T})\Big\}\frac{\partial\sigma_\theta^2(X_i)}{\partial\theta_k} + o_P(n^{-1/2}) \\
= & -\frac{1}{n}\sum_{i=1}^n (1-\Delta_i)\frac{\sigma^{02}(X_i)}{1-F_\varepsilon^0(E_i^{0T})}[E_i^{0T} - 2\int_{-\infty}^T edF_\varepsilon^0(e)]E_i^{0T} f_\varepsilon^0(E_i^{0T})(\hat{E}_i^0 - E_i^0)I(E_i^0 \le T) \\
&\quad \times\frac{\partial\sigma_\theta^2(X_i)}{\partial\theta_k} + o_P(n^{-1/2}), \quad k = 1, \ldots, d. \qquad \text{(A.9)}
\end{aligned}$$

Finally, together with (A.4), (A.5), (A.7), (A.8) and (A.9), and Lemma A.3, we obtain

$$\begin{aligned}
&(1/n)\sum_{i=1}^n (\widehat{\tilde{\varepsilon}_{Ti}^{2*}} - \tilde{\varepsilon}_{Ti}^{2*})\frac{\partial\sigma_\theta^2(X_i)}{\partial\theta_k} \qquad \text{(A.10)} \\
= & \frac{-1}{n^2 a_n}\sum_{i=1}^n\sum_{j=1}^n K(\frac{X_i - X_j}{a_n})\frac{\partial\sigma_\theta^2(X_i)}{\partial\theta_k} f_X^{-1}(X_i)\sigma^0(X_i) \\
&\quad \times\Big\{2\Delta_i(m^T(X_i) - Z_i)[\eta(Z_j, \Delta_j|X_i) + \int_{-\infty}^T edF_\varepsilon^0(e)\zeta(Z_j, \Delta_j|X_i)] \\
&\quad -(1-\Delta_i)\frac{\sigma^0(X_i)}{1-F_\varepsilon^0(E_i^{0T})}\Big\{\Big\{\frac{f_\varepsilon^0(E_i^{0T})\int_{E_i^{0T}}^T (y^2 - 2y\int_{-\infty}^T edF_\varepsilon^0(e))\,dF_\varepsilon^0(y)}{1-F_\varepsilon^0(E_i^{0T})} \\
&\qquad -[E_i^{0T} - 2\int_{-\infty}^T edF_\varepsilon^0(e)]E_i^{0T} f_\varepsilon^0(E_i^{0T})I(E_i^0 \le T)\Big\}\eta(Z_j, \Delta_j|X_i) \\
&\quad +\Big\{\frac{E_i^{0T} f_\varepsilon^0(E_i^{0T})\int_{E_i^{0T}}^T (y^2 - 2y\int_{-\infty}^T edF_\varepsilon^0(e))\,dF_\varepsilon^0(y)}{1-F_\varepsilon^0(E_i^{0T})} - 2\mathcal{A}_c(X_i, Z_i) \\
&\qquad -[E_i^{0T} - 2\int_{-\infty}^T edF_\varepsilon^0(e)](E_i^{0T})^2 f_\varepsilon^0(E_i^{0T})I(E_i^0 \le T)\Big\}\zeta(Z_j, \Delta_j|X_i)\Big\}\Big\} \\
&\quad +\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \frac{\partial\sigma_\theta^2(X_i)}{\partial\theta_k}\Big\{2\Delta_i(m^T(X_i) - Z_i)\sigma^0(X_i)\pi(V_j) \\
&\quad +(1-\Delta_i)\frac{\sigma^{02}(X_i)}{1-F_\varepsilon^0(E_i^{0T})}\Big\{\Big[\frac{\int_{E_i^{0T}}^T (y^2 - 2y\int_{-\infty}^T edF_\varepsilon^0(e))\,dF_\varepsilon^0(y)}{1-F_\varepsilon^0(E_i^{0T})}
\end{aligned}$$

$$
-(E_i^{0T})^2 + 2E_i^{0T}\int_{-\infty}^{T} eF_\varepsilon^0(e)\Big]\varphi(X_j, Z_j, \Delta_j, E_i^{0T})
$$

$$
+\left[T^2 - 2T\int_{-\infty}^{T} edF_\varepsilon^0(e)\right]\varphi(X_j, Z_j, \Delta_j, T)
$$

$$
-\int_{E_i^{0T}}^{T}(2y - 2\int_{-\infty}^{T} edF_\varepsilon^0(e))\varphi(X_j, Z_j, \Delta_j, y)dy
$$

$$
+2\left[(1 - F_\varepsilon^0(E_i^{0T}))\int_{-\infty}^{T} edF_\varepsilon^0(e) - \int_{E_i^{0T}}^{T} ydF_\varepsilon^0(y)\right]
$$

$$
\times\pi(V_j)\}\} + o_P(n^{1/2}). \qquad (A.11)
$$

Finally, usual calculations on U-statistics (see by example Heuchenne and Van Keilgom 2007a) 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 Heuchenne and Van Keilegom (2010), the asymptotic representation of the residuals distribution given in Theorem 3.1 of Van Keilegom and Akritas (1999) and simple calculations on U-statistics.

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. (2007a). Polynomial regression with censored data based on preliminary nonparametric estimation. *Ann. Instit. Statist. Math.*, **59**, 273–298.

Heuchenne, C. and Van Keilegom, I. (2007b). Nonlinear regression with censored data. *Technometrics*, **49**, 34–44.

Heuchenne, C. and Van Keilegom, I. (2010). Estimation in nonparametric location-scale regression models with censored data. To appear in *Ann. Inst. Statist. Math.*

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.

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

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

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.

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.

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.

Wei, S. X. (2002). A censoredGARCH model of asset returns with price limits. *Journal of Empirical Finance*, **9**, 197–223.