

Estimation of a general parametric location in censored regression

Cédric HEUCHENNE *

QuantOM,[†]

HEC-Management School of University of Liège

and Institute of Statistics

Université catholique de Louvain

Ingrid VAN KEILEGOM ‡

Institute of Statistics

Université catholique de Louvain

June 2, 2011

Abstract

Consider the random vector (X, Y) , where Y represents a response variable and X an explanatory variable. The response Y is subject to random right censoring, whereas X is completely observed. Let $m(x)$ be a conditional location function of Y given $X = x$. In this paper we assume that $m(\cdot)$ belongs to some parametric class $\mathcal{M} = \{m_\theta : \theta \in \Theta\}$ and we propose a new method for estimating the true unknown value θ_0 . The method is based on nonparametric imputation for the censored observations. The consistency and asymptotic normality of the proposed estimator are established.

Key Words: Least-squares estimator; L-functionals; Nadaraya-Watson estimator; Non-linear regression; Nonparametric regression; Right censoring; Survival analysis.

*Financial support from IAP research network P6/03 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged.

[†]Centre for Quantitative Methods and Operations Management

[‡]Financial support from IAP research network P6/03 of the Belgian Government (Belgian Science Policy), and from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650 is gratefully acknowledged.

1 Introduction

Consider the random vector (X, Y) , where Y represents a (possible transformation of a) response variable and X an explanatory variable. This paper is concerned with the estimation of a location functional of Y given X , when Y is subject to random right censoring and X is completely observed. We suppose that this location functional belongs to some parametric family.

This problem has been widely studied in the literature when the location functional is the conditional mean (see e.g. Stute (1993), Fyngenson and Zhou (1994), Van Keilegom and Akritas (2000) among many others) or the conditional median (see e.g. Buchinsky and Hahn (1998), Portnoy (2003), Yin *et al* (2008), Wang and Wang (2009) and the references therein). Here we focus attention on L -functionals, given by

$$m(x) = \int_0^1 F^{-1}(s|x)J(s)ds,$$

where $F(y|x) = P(Y \leq y|X = x)$ is the conditional distribution of Y given $X = x$, $F^{-1}(s|x) = \inf\{y : F(y|x) \geq s\}$ is the conditional quantile of order s , and $J(s)$ is a weight function satisfying $J(s) \geq 0$ for all $0 \leq s \leq 1$ and $\int_0^1 J(s)ds = 1$. This type of location functionals includes as special cases the conditional mean, trimmed mean, or any other kind of weighted mean. The conditional median can be regarded as a limiting special case, obtained when $J(s)$ puts all its mass on $s = 1/2$.

Another interesting special case is obtained for $J(s) = I(1 - \delta < s \leq 1)/\delta$ for some $0 < \delta < 1$. Suppose the upper bound of the support of Y given $X = x$ is finite (say equal to τ_x) and one is interested in the estimation of the support curve $x \rightarrow \tau_x$. The above choice of J yields a robust estimator of this curve for small values of δ , and is an interesting alternative to the so-called m -frontiers or α -frontiers, which are based on order statistics of order m or quantiles of order $0 < \alpha < 1$ (see e.g. Cazals *et al* (2002) and Aragon *et al* (2005)).

We suppose in this paper that Y is subject to random right censoring, i.e. instead of observing Y we only observe (Z, Δ) , where $Z = \min(Y, C)$ is the observed survival time, $\Delta = I(Y \leq C)$ is the censoring indicator, and the random variable C represents the censoring time, which is independent of Y conditionally on X . Let (X_i, Z_i, Δ_i) ($i = 1, \dots, n$) be n independent copies of (X, Z, Δ) .

In the context of regression with right censored responses it is well known that the nonparametric kernel estimator of the conditional distribution $F(\cdot|x)$ is inconsistent in the right tail. The modeling of the above location functional $m(x)$ is therefore especially

attractive in this framework, since particular choices of J -functions enable us to get rid of these inconsistent parts (see Section 2 for more details).

The goal of this paper is to propose a new estimation method for $m(x)$, when it is believed that $m(\cdot)$ belongs to the class

$$\mathcal{M} = \{m_\theta : \theta \in \Theta\},$$

consisting of location functionals determined by a finite-dimensional parameter vector $\theta \in \Theta$, where Θ is a compact subset of \mathbb{R}^d . The class \mathcal{M} can be taken equal to the class of polynomial functions of order $d - 1$, but any other parametric class of “smooth” functions (in θ) can be chosen as well.

The paper is organized as follows. In the next section we introduce some notation and explain in detail the proposed estimation procedure for θ_0 , the true unknown value of θ . Section 3 gives the main asymptotic properties of the proposed estimator. In Section 4 we summarize the results of the paper and give ideas for future research, whereas the Appendix contains the proofs of the main asymptotic results.

2 Description of the method

The estimator θ_n is defined as follows. First, we note that θ_0 can be written as

$$\begin{aligned} \theta_0 &= \operatorname{argmin}_\theta E \left[\left(\int_0^1 F^{-1}(s|X) J(s) ds - m_\theta(X) \right)^2 \right] \\ &= \operatorname{argmin}_\theta E \left[(Y - m_\theta(X))^2 J(F(Y|X)) \right] \\ &= \operatorname{argmin}_\theta E \left[(Z - m_\theta(X))^2 J(F(Z|X)) \Delta \right. \\ &\quad \left. + \frac{\int_Z^\infty (y - m_\theta(X))^2 J(F(y|X)) dF(y|X)}{1 - F(Z|X)} (1 - \Delta) \right]. \end{aligned}$$

The idea is now to estimate θ_0 by a minimizer θ_n of an empirical version of the above quantity, namely

$$\begin{aligned} \theta_n &= \operatorname{argmin}_{\theta \in \Theta} n^{-1} \sum_{i=1}^n \left[(Z_i - m_\theta(X_i))^2 J(\hat{F}(Z_i|X_i)) \Delta_i \right. \\ &\quad \left. + \frac{\int_{Z_i}^\infty (y - m_\theta(X_i))^2 J(\hat{F}(y|X_i)) d\hat{F}(y|X_i)}{1 - \hat{F}(Z_i|X_i)} (1 - \Delta_i) \right]. \end{aligned} \quad (2.1)$$

Here, $\hat{F}(y|x)$ is the nonparametric kernel estimator of the conditional distribution $F(y|x)$ proposed by Beran (1981) :

$$\hat{F}(y|x) = 1 - \prod_{Z_i \leq y, \Delta_i = 1} \left\{ 1 - \frac{W_i(x, a_n)}{\sum_{k=1}^n I(Z_k \geq Z_i) W_k(x, a_n)} \right\}, \quad (2.2)$$

(when no ties are present), where

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

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

Hence, the estimation procedure for θ_0 can be summarized as follows :

1. First, for fixed θ , estimate the weighted squared error $(Z_i - m_\theta(X_i))^2 J(F(Z_i|X_i))$ of an uncensored observation $(X_i, Y_i, \Delta_i = 1)$ by $(Y_i - m_\theta(X_i))^2 J(\hat{F}(Y_i|X_i))$, where \hat{F} is defined in (2.2), and of a censored observation $(X_i, C_i, \Delta_i = 0)$ by a nonparametric estimator of $E[(Y - m_\theta(X))^2 J(\hat{F}(Y|X)) | X = X_i, Y > C_i]$.
2. Then, estimate θ_0 by minimizing the average of the weighted squared errors obtained under the previous step.

Although the above idea of estimating θ_0 has never been considered in the present context of nonlinear parametric estimation of a general location functional, similar versions of this idea have been applied in other contexts. See e.g. Akritas (1996), who, in the context of polynomial regression, first replaced all observations Z_i (censored and uncensored ones) by a nonparametric estimator $\hat{m}(X_i)$ of $m(X_i)$, and then applied a classical least squares procedure on the so-obtained ‘synthetic’ data $(X_i, \hat{m}(X_i))$. His method has the disadvantage that it is quite sensitive to the choice of the bandwidth, as the bandwidth is playing an important role for both the censored and the uncensored data. Another related methodology is given in Heuchenne and Van Keilegom (2007), who consider the estimation of the conditional mean of Y given X when the relation between Y and X is given by a nonparametric location-scale model. They also replace the censored observations by some kind of synthetic data estimated under the assumed location-scale model. Also see Pardo-Fernández *et al* (2007) for a goodness-of-fit test in parametric censored regression.

For the presentation of the asymptotic results in the next section, we need to introduce the following notation. Let $H(y|x) = P(Z \leq y | X = x)$, $H_\delta(y|x) = P(Z \leq y, \Delta = \delta | X =$

$x)$ ($\delta = 0, 1$), $F_\varepsilon(y|x) = P(\varepsilon \leq y|X = x)$ and $F_X(x) = P(X \leq x)$. The probability density functions of the above distribution functions will be denoted by lower case letters. Also, let

$$\epsilon_\theta(x, z, \delta, F) = \left\{ \delta(z - m_\theta(x))^2 J(F(z|x)) + (1 - \delta) \frac{\int_z^{+\infty} (y - m_\theta(x))^2 J(F(y|x)) dF(y|x)}{1 - F(z|x)} \right\}$$

(where $E[\epsilon_\theta(X, Z, \Delta, F)] = E[(Y - m_\theta(X))^2 J(F(Y|X))]$), which implies that the estimator θ_n can be written as

$$\theta_n = \operatorname{argmin}_{\theta \in \Theta} \sum_{i=1}^n \epsilon_\theta(X_i, Z_i, \Delta_i, \hat{F}),$$

where \hat{F} is the Beran estimator defined in (2.2).

3 Asymptotic results

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

Theorem 3.1 *Assume (A1), (A2) (i), (A4) (i), (iv), (v) and (A7). Moreover, assume that J is continuously differentiable, $\int_0^1 J(s) ds = 1$, $J(s) \geq 0$ for all $0 \leq s \leq 1$, F_X is two times continuously differentiable, $\inf_{x \in R_X} f_X(x) > 0$, Θ is compact, θ_0 is an interior point of Θ , and $m_\theta(x)$ is continuous in (x, θ) for all x and θ . Let*

$$S_n(\theta) = \frac{1}{n} \sum_{i=1}^n \epsilon_\theta(X_i, Z_i, \Delta_i, \hat{F}).$$

Then

$$\theta_n - \theta_0 = o_P(1),$$

and

$$S_n(\theta_n) = E[\epsilon_{\theta_0}(X, Z, \Delta, F)] + o_P(1).$$

The next result decomposes the difference $\theta_n - \theta_0$ into a sum of i.i.d. terms and a negligible term of lower order. This decomposition will be crucial for obtaining the asymptotic normality of θ_n .

Theorem 3.2 Assume (A1)-(A7). Then,

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

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

$$\Omega_{jk} = E \left[\frac{\partial m_{\theta_0}(X)}{\partial \theta_j} \frac{\partial m_{\theta_0}(X)}{\partial \theta_k} \right],$$

$\rho = (\rho_1, \dots, \rho_d)^T$, and for any $j = 1, \dots, d$ and $i = 1, \dots, n$,

$$\begin{aligned} \rho_j(X_i, Z_i, \Delta_i) &= \frac{\partial m_{\theta_0}(X_i)}{\partial \theta_j} \left\{ \Delta_i (Z_i - m_{\theta_0}(X_i)) J(F(Z_i|X_i)) \right. \\ &\quad \left. + (1 - \Delta_i) \frac{\int_{Z_i}^{+\infty} (y - m_{\theta_0}(X_i)) J(F(y|X_i)) dF(y|X_i)}{1 - F(Z_i|X_i)} \right\} \\ &\quad + f_X(X_i) \sum_{\delta=0,1} \int \chi_j((X_i, z, \delta), (Z_i, \Delta_i)) dH_\delta(z|X_i), \end{aligned}$$

where the function χ_j is defined in the Appendix.

We are now ready to state the asymptotic normality of θ_n .

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

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

The proof of this result follows readily from Theorem 3.2.

4 Summary and future research

In this paper we have proposed a new method to estimate the coefficients of a parametric conditional location function, when the response is subject to random right censoring. The proposed estimator is a least squares type estimator, for which the censored observations are replaced by nonparametrically imputed values. The consistency and asymptotic normality of the estimator are established.

In the future, it would be interesting to compare the proposed method with other estimators that have been proposed in the literature, for instance, when the conditional location is the conditional mean. The least squares estimators obtained in this paper can

be introduced in a test statistic to test the validity of the assumed parametric model, and it would be interesting to work out the asymptotic theory for that test statistic. Finally, extensions of the current work to semiparametric models (like the partial linear or single index model) can also be worked out based on the results in this paper.

Appendix

We first introduce the following functions, which are needed in the statement of the asymptotic results given in Section 3 :

$$\begin{aligned} \xi(z, \delta, y|x) &= (1 - F(y|x)) \left\{ - \int_{-\infty}^{y \wedge z} \frac{dH_1(s|x)}{(1 - H(s|x))^2} + \frac{I(z \leq y, \delta = 1)}{1 - H(z|x)} \right\}, \\ \chi_j(v^1, z^2, \delta^2) &= \frac{\partial m_{\theta_0}(x^1)}{\partial \theta_j} \left\{ \left[\delta^1(z^1 - m_{\theta_0}(x^1))J'(F(z^1|x^1)) \right. \right. \\ &\quad \left. \left. + (1 - \delta^1) \frac{\int_{z^1}^{+\infty} (y - m_{\theta_0}(x^1))J(F(y|x^1))dF(y|x^1)}{(1 - F(z^1|x^1))^2} \right] \xi(z^2, \delta^2, z^1|x^1) \right. \\ &\quad \left. + (1 - \delta^1) \left[\frac{\int_{z^1}^{+\infty} (y - m_{\theta_0}(x^1))J(F(y|x^1))d\xi(z^2, \delta^2, y|x^1)}{1 - F(z^1|x^1)} \right. \right. \\ &\quad \left. \left. + \frac{\int_{z^1}^{+\infty} (y - m_{\theta_0}(x^1))J'(F(y|x^1))\xi(z^2, \delta^2, y|x^1)dF(y|x^1)}{1 - F(z^1|x^1)} \right] \right\}, \end{aligned}$$

$j = 1, \dots, d$, where $v^1 = (x^1, z^1, \delta^1)$.

Let T_x be any value less than the upper bound of the support of $H(\cdot|x)$ such that $\inf_{x \in R_X} (1 - H(T_x|x)) > 0$. For a (sub)distribution function $L(y|x)$ we will use the notations $l(y|x) = L'(y|x) = (\partial/\partial y)L(y|x)$, $\dot{L}(y|x) = (\partial/\partial x)L(y|x)$ and similar notations will be used for higher order derivatives.

The assumptions needed for the results of Section 3 are listed below.

(A1)(i) $na_n^3(\log n)^{-3} \rightarrow \infty$ and $na_n^4 \rightarrow 0$.

(ii) The support R_X of X is a compact interval.

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

(iv) Ω is non-singular.

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

(A3) F_X is three times continuously differentiable and $\inf_{x \in R_X} f_X(x) > 0$.

(A4)(i) $L(y|x)$ is continuous,

(ii) $L'(y|x) = l(y|x)$ exists, is continuous in (x, y) and $\sup_{x,y} |yL'(y|x)| < \infty$,

(iii) $L''(y|x)$ exists, is continuous in (x, y) and $\sup_{x,y} |y^2L''(y|x)| < \infty$,

(iv) $\dot{L}(y|x)$ exists, is continuous in (x, y) and $\sup_{x,y} |y\dot{L}(y|x)| < \infty$,

(v) $\ddot{L}(y|x)$ exists, is continuous in (x, y) and $\sup_{x,y} |y^2\ddot{L}(y|x)| < \infty$,

(vi) $\ddot{L}'(y|x)$ exists, is continuous in (x, y) and $\sup_{x,y} |y\ddot{L}'(y|x)| < \infty$,

for $L(y|x) = H(y|x)$ and $H_1(y|x)$.

(A5) For the density $f_{X|Z,\Delta}(x|z, \delta)$ of X given (Z, Δ) , $\sup_{x,z} |f_{X|Z,\Delta}(x|z, \delta)| < \infty$, $\sup_{x,z} |\dot{f}_{X|Z,\Delta}(x|z, \delta)| < \infty$, $\sup_{x,z} |\ddot{f}_{X|Z,\Delta}(x|z, \delta)| < \infty$ ($\delta = 0, 1$).

(A6) Θ is compact and θ_0 is an interior point of Θ . All partial derivatives of $m_\theta(x)$ with respect to the components of θ up to order three exist and are continuous in (x, θ) for all x and θ .

(A7) The function $E[(Y - m_\theta(X))^2 J(F(Y|X))]$ has a unique minimum in $\theta = \theta_0$.

Proof of Theorem 3.1. We prove the consistency of θ_n by verifying the conditions of Theorem 5.7 in van der Vaart (1998, p. 45). From the definition of θ_n and condition (A7), it follows that it suffices to show that

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

where $S_0(\theta) = E[(Y - m_\theta(X))^2 J(F(Y|X))] = E[\varepsilon^2 J(F_\varepsilon(\varepsilon|X))]$. The second statement of Theorem 3.1 then follows immediately from (A.1) together with the consistency of θ_n . (A.1) is obtained by using (A2), the uniform consistency of the Beran (1981) estimator (given in Proposition A.3 of Van Keilegom and Akritas (1999)) and Theorem 2 of Jennrich (1969). \square

Proof of Theorem 3.2. For some θ_{1n} between θ_n and θ_0 ,

$$\theta_n - \theta_0 = - \left\{ \frac{\partial^2 S_n(\theta_{1n})}{\partial \theta \partial \theta^T} \right\}^{-1} \frac{\partial S_n(\theta_0)}{\partial \theta} = -R_{1n}^{-1} R_{2n}.$$

First, we treat R_{2n} .

$$\begin{aligned} R_{2nk} &= -\frac{2}{n} \sum_{i=1}^n \frac{\partial m_{\theta_0}(X_i)}{\partial \theta_k} \left\{ \Delta_i (Z_i - m_{\theta_0}(X_i)) J'(F(Z_i|X_i)) (\hat{F}(Z_i|X_i) - F(Z_i|X_i)) \right. \\ &\quad + (1 - \Delta_i) \left[\frac{\int_{Z_i}^{+\infty} (y - m_{\theta_0}(X_i)) J'(F(y|X_i)) (\hat{F}(y|X_i) - F(y|X_i)) dF(y|X_i)}{1 - F(Z_i|X_i)} \right. \\ &\quad + \frac{\int_{Z_i}^{+\infty} (y - m_{\theta_0}(X_i)) J(F(y|X_i)) dF(y|X_i)}{(1 - F(Z_i|X_i))^2} (\hat{F}(Z_i|X_i) - F(Z_i|X_i)) \\ &\quad \left. \left. + \frac{\int_{Z_i}^{+\infty} (y - m_{\theta_0}(X_i)) J(F(y|X_i)) d(\hat{F}(y|X_i) - F(y|X_i))}{1 - F(Z_i|X_i)} \right] \right\} \\ &\quad - \frac{2}{n} \sum_{i=1}^n \frac{\partial m_{\theta_0}(X_i)}{\partial \theta_k} \left\{ \Delta_i (Z_i - m_{\theta_0}(X_i)) J(F(Z_i|X_i)) \right. \\ &\quad \left. + (1 - \Delta_i) \frac{\int_{Z_i}^{+\infty} (y - m_{\theta_0}(X_i)) J(F(y|X_i)) dF(y|X_i)}{1 - F(Z_i|X_i)} \right\} + o_P(n^{-1/2}) \\ &= R_{21nk} + R_{22nk} + o_P(n^{-1/2}), \end{aligned}$$

$k = 1, \dots, d$. Developing R_{21nk} leads to

$$R_{21nk} = -\frac{2}{n^2 a_n} \sum_{i \neq j} K\left(\frac{X_i - X_j}{a_n}\right) \chi_k(V_i, Z_j, \Delta_j) + o_P(n^{-1/2}),$$

where $V_i = (X_i, Z_i, \Delta_i)$. Next, we rewrite R_{21nk} as

$$\begin{aligned} R_{21nk} &= \frac{-2}{n^2 a_n} \sum_{i \neq j} \{A_k^*(V_i, V_j) + E[A_k(V_i, V_j)|V_i] + E[A_k(V_i, V_j)|V_j] - E[A_k(V_i, V_j)]\} \\ &\quad + o_P(n^{-1/2}) \\ &= T_{1,k}^n + T_{2,k}^n + T_{3,k}^n + T_{4,k}^n + o_P(n^{-1/2}), \end{aligned}$$

where

$$A_k(V_i, V_j) = K\left(\frac{X_i - X_j}{a_n}\right) \chi_k(V_i, Z_j, \Delta_j)$$

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

$$\begin{aligned}
& E[A_k(V_i, V_j)|V_i] \\
&= \sum_{\delta=0,1} \int \int \chi_k(V_i, z, \delta) K\left(\frac{X_i - x}{a_n}\right) h_\delta(z|x) f_X(x) dz dx \\
&= a_n \sum_{\delta=0,1} \int \int \chi_k(V_i, z, \delta) K(u) (h_\delta(z|X_i) - a_n u \dot{h}_\delta(z|X_i) + O(a_n^2)) \\
&\quad \times (f_X(X_i) - a_n u f'_X(X_i) + O(a_n^2)) dz du \\
&= a_n f_X(X_i) \sum_{\delta=0,1} \int \chi_k(V_i, z, \delta) h_\delta(z|X_i) dz + O(a_n^3) = O(a_n^3) \tag{A.2}
\end{aligned}$$

for $i = 1, \dots, n$, since

$$\sum_{\delta=0,1} \int \xi(z, \delta, y|x) h_\delta(z|x) dz = 0$$

for all $x \in R_X$ and $y \leq T_x$. Hence, we also have $E[A_k(V_i, V_j)] = O(a_n^3)$. In a similar way, using three Taylor expansions of order 2, we get

$$\begin{aligned}
E[A_k(V_i, V_j)|V_j] &= a_n f_X(X_j) \sum_{\delta=0,1} \int \chi_k((X_j, z, \delta), (Z_j, \Delta_j)) dH_\delta(z|X_j) \\
&\quad + O(a_n^3). \tag{A.3}
\end{aligned}$$

Note that for $T_{1,k}^n$, $E[T_{1,k}^n] = 0$, resulting, by Chebyshev's inequality, in

$$\begin{aligned}
P(|T_{1,k}^n| > K(na_n)^{-1}) &\leq K^{-2}(na_n)^2 E[(T_{1,k}^n)^2] \\
&= 4K^{-2}n^{-2} \sum_{j \neq i} \sum_{m \neq l} E[A_k^*(V_i, V_j)A_k^*(V_l, V_m)],
\end{aligned}$$

for any $K > 0$. Since $E[A_k^*(V_i, V_j)] = 0$, the terms for which $i, j \neq l, m$ are zero. The terms for which either i or j equals l or m and the other differs from l and m , are also zero, because, for example when $i = l$ and $j \neq m$,

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

Thus, only the $2n(n-1)$ terms for which (i, j) equals (l, m) or (m, l) remain. Since $A_k^*(V_i, V_j)$ is bounded by $CK\left(\frac{X_i - X_j}{a_n}\right) + O(a_n)$ for some constant $C > 0$, we have (in the case (i, j) equals (l, m)) that

$$E[A_k^*(V_i, V_j)] \leq C^2 a_n \int f_X^2(x) dx \int K^2(u) du + O(a_n^2) = O(a_n).$$

The case (i, j) equals (m, l) is treated similarly. It now follows that

$$T_{1,k}^n = o_P(n^{-1}a_n^{-1}), \quad (\text{A.4})$$

which is $o_P(n^{-1/2})$. By (A.2), (A.3), (A.4), we finally obtain

$$\begin{aligned} R_{21nk} &= -\frac{2}{n} \sum_{i=1}^n \sum_{\delta=0,1} \int \chi_k((X_i, z, \delta), (Z_i, \Delta_i)) f_X(X_i) dH_\delta(z|X_i) \\ &\quad + o_P(n^{-1/2}), \quad k = 1, \dots, d. \end{aligned}$$

Finally, we treat the term R_{1n} .

$$\begin{aligned} R_{1n} &= -\frac{2}{n} \left\{ \sum_{i=1}^n \left[\Delta_i (Z_i - m_{\theta_{1n}}(X_i)) J(\hat{F}(Z_i|X_i)) \right. \right. \\ &\quad \left. \left. + (1 - \Delta_i) \frac{\int_{Z_i}^{+\infty} (y - m_{\theta_{1n}}(X_i)) J(\hat{F}(y|X_i)) d\hat{F}(y|X_i)}{1 - \hat{F}(Z_i|X_i)} \right] \frac{\partial^2 m_{\theta_{1n}}(X_i)}{\partial \theta \partial \theta^T} \right. \\ &\quad \left. - \sum_{i=1}^n \left[\Delta_i J(\hat{F}(Z_i|X_i)) + (1 - \Delta_i) \frac{\int_{Z_i}^{+\infty} J(\hat{F}(y|X_i)) d\hat{F}(y|X_i)}{1 - \hat{F}(Z_i|X_i)} \right] \right. \\ &\quad \left. \times \left(\frac{\partial m_{\theta_{1n}}(X_i)}{\partial \theta} \right) \left(\frac{\partial m_{\theta_{1n}}(X_i)}{\partial \theta^T} \right) \right\} \\ &= R_{11n}(\theta_{1n}, \hat{F}) + R_{12n}(\theta_{1n}, \hat{F}). \end{aligned}$$

Using the uniform consistency of the Beran (1981) estimator together with (A6), it is clear that

$$R_{11n}(\theta_{1n}, \hat{F}) + R_{12n}(\theta_{1n}, \hat{F}) = R_{11n}(\theta_0, F) + R_{12n}(\theta_0, F) + o_P(1).$$

Since $E[R_{11n}(\theta_0, F)] = 0$ and

$$E[R_{12n}(\theta_0, F)] = 2E \left[\left(\frac{\partial m_{\theta_0}(X)}{\partial \theta} \right) \left(\frac{\partial m_{\theta_0}(X)}{\partial \theta^T} \right) \right],$$

we obtain that

$$R_{1n} = 2E \left[\left(\frac{\partial m_{\theta_0}(X)}{\partial \theta} \right) \left(\frac{\partial m_{\theta_0}(X)}{\partial \theta^T} \right) \right] + o_P(1).$$

This finishes the proof. □

References

- Akritis, M.G. (1996). On the use of nonparametric regression techniques for fitting parametric regression models. *Biometrics*, **52**, 1342-1362.
- Aragon, Y., Daouia, A. and Thomas-Agnan, C. (2005). Nonparametric frontier estimation: a conditional quantile-based approach. *Econometric Theory*, **21**, 358-389.
- Beran, R. (1981). *Nonparametric regression with randomly censored survival data*. Technical Report, Univ. California, Berkeley.
- Buchinsky, M. and Hahn, J. (1998). An alternative estimator for the censored quantile regression model. *Econometrica*, **66**, 653-671.
- Cazals, C., Florens, J.P. and Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, **106**, 1-25.
- Fygenon, M. and Zhou, M. (1994). On using stratification in the analysis of linear regression models with right censoring. *Annals of Statistics*, **22**, 747-762.
- Heuchenne, C. and Van Keilegom, I. (2007). Nonlinear regression with censored data. *Technometrics*, **49**, 34-44.
- Jennrich, R.I. (1969). Asymptotic properties of nonlinear least squares estimators. *Annals of Mathematical Statistics*, **40**, 633-643.
- Pardo-Fernández, J.C., Van Keilegom, I. and González-Manteiga, W. (2007). Goodness-of-fit tests for parametric models in censored regression. *Canadian Journal of Statistics*, **35**, 249-264.
- Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, **98**, 1001-1012.
- Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis*, **45**, 89-103.
- Van der Vaart, A.W. (1998). *Asymptotic statistics*. Cambridge University Press, Cambridge.
- Van Keilegom, I. and Akritis, M. G. (1999). Transfer of tail information in censored regression models. *Annals of Statistics*, **27**, 1745-1784.
- Van Keilegom, I. and Akritis, M. G. (2000). The least squares method in heteroscedastic censored regression models. In : *Asymptotics in Statistics and Probability*, Ed. M.L. Puri, VSP, 379-391.
- Yin, G., Zeng, D. and Li, H. (2008). Power-transformed linear quantile regression with censored data. *Journal of the American Statistical Association*, **103**, 1214-1224.

Wang, J. and Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association*, **104**, 1117-1128.