The least squares method in heteroscedastic censored regression models

TU/e technische universiteit eindhoven
/department of mathematics and computing science

Memorandum COSOR 99-18

The Least Squares Method in Heteroscedastic Censored Regresssion Models

I. Van Keilegom and M.G. Akritas

Eindhoven, November 1999
The Netherlands

The Least Squares Method in Heteroscedastic Censored Regression Models

Ingrid Van Keilegom and Michael G. Akritas

Eindhoven University of Technology and Penn State University

October 6, 1999

Abstract

Consider the heteroscedastic polynomial regression model $Y = \beta_0 + \beta_1 X + \cdots + \beta_p X^p + \sqrt{\mathrm{Var}(Y|X)}\varepsilon$, where ε is independent of X, and Y is subject to random censoring. Provided that the censoring on Y is 'light' in some region of X, we construct a least squares estimator for the regression parameters whose asymptotic bias is shown to be as small as desired. The least squares estimator is defined as a functional of the Van Keilegom and Akritas (1999) estimator of the bivariate distribution $P(X \leq x, Y \leq y)$, and its asymptotic normality is obtained.

KEY WORDS: Asymptotic normality; Asymptotic representation; Bivariate distribution; Least squares estimator; Nonparametric regression residuals; Polynomial regression; Right censoring.

1 Introduction

Fitting regression models with censored data has received considerable attention in the statistical literature. For the more recent contributions see Tsiatis (1990), Ritov (1990), Lai and Ying (1991a, 1991b, 1994), Zhou (1992a,b), Fygenson and Zhou (1994), Fygenson and Ritov (1994), Akritas (1994), Stute (1993, 1996), Akritas, Murphy and LaValley (1995), LeBlanc and Crowley (1995), Ying, Jung and Wei (1995), Akritas (1996) and references therein. As outlined in Fygenson and Ritov (1994), the rank estimators of Tsiatis (1990) and M-estimators of Ritov (1990) suffer from certain computational as well as theoretical difficulties. In addition, the variance of the resulting estimator depends on the density of the error distribution which is not estimated well with incomplete data. These difficulties motivated Wei, Ying and Lin (1990), and Lin and Wei (1992a,b) to devise inference procedures, similar to those derived from the likelihood ratio statistic in the parametric case, in an effort to circumvent the difficulty of estimating the variance-covariance matrix. The recent papers by Fygenson and Ritov (1994), Akritas, Murphy and LaValley (1995), and Akritas (1996) propose estimators which are generally applicable in the framework of the homoscedastic accelerated failure time model, are easy to compute, and obtain variance estimators that bypass the need to estimate the density of the error distribution.

Though the technology for fitting regression models with censored data is at present quite advanced, all methods mentioned above pertain to the homoscedastic regression model. (Zhou 1992a claims that his result also holds for heteroscedastic models; even so, his derivations require the (very often) unrealistic assumption that the censoring distribution does not depend on the covariate.) Since with uncensored data the least squares estimator is consistent and asymptotically normal even in heteroscedastic regression models, we turn our attention to least squares estimation. As remarked in Akritas (1994), existing methods for extending the least squares estimator to censored data give biased results when there is heavy censoring at the upper tails of the conditional distribution of the response given the covariate; see also the expressions of α^*, β^* in Theorem 2.1 of Fygenson and Zhou (1994), as well as the simulations (b) reported in their Table 1. The reason for this is quite simple : the least squares estimator is a moments estimator and estimation of moments cannot be achieved without information about the tails of the distribution.

The least squares estimator we propose in this paper does not restrict to homoscedastic regression models, but can be used under any heteroscedastic model

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

where $E(\varepsilon) = 0, \mathrm{Var}(\varepsilon) = \sigma_\varepsilon^2$ and ε is independent of X. The estimators also have the advantage that their asymptotic bias can be made as small as desirable, provided there is a region of X where the censoring of Y is 'light'. The basic idea that leads to this nice feature is that tail information can be transferred from such a region of light censoring to other regions where censoring is 'heavy'. In this way we obtain enough information about the tails of the distribution in order to estimate moments in an accurate way. This idea was proposed in Akritas (1994).

The present least squares estimator will be a functional of the Van Keilegom and Akritas (1999) estimator of the bivariate distribution function, which is valid in the context of any heteroscedastic nonparametric regression model

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1.2}$$

where the functions m and σ are 'smooth' and ε is independent of X. As such it generalizes the least squares estimators of Akritas (1996), Akritas (1994) and of Cristóbal Cristóbal et al. (1987) which include both the ordinary least squares estimator as well as the ordinary and generalized ridge regression estimators as special cases. See also Remark 2.1 below. (It is assumed that a transformation, e.g. the logarithmic transformation, has been applied on the typically positive censored response variable. Thus, (1.2) can be viewed as a nonparametric and heteroscedastic accelerated failure time model.)

To set the notation, consider a bivariate random vector (X, Y) having distribution function $F(x, y)$. Let C be a random variable censoring Y. It will be assumed that C is conditionally independent of Y given X, and also that X is always observable. Thus the available data are realizations of the i.i.d. random vectors (X_i, Z_i, Δ_i), $i = 1, \ldots, n$, where $Z_i = Y_i \wedge C_i$, and $\Delta_i = I(Y_i = Z_i)$.

We describe first the Van Keilegom and Akritas (1999) estimator of $F(x, y)$. Like the Akritas (1994) estimator, it is based on the relationship

$$F(x, y) = \int_{-\infty}^{x} F(y|t)\, dF_X(t), \tag{1.3}$$

where F_X is the marginal distribution function of X and $F(y|t)$ is the conditional distribution $P(Y \leq y|X = t)$. Relationship (1.3) suggests the estimator

$$\hat{F}(x,y) = \int_{-\infty}^{x} \hat{F}(y|t)\, d\hat{F}_X(t), \tag{1.4}$$

where $\hat{F}(y|t)$ is a suitable estimator of $F(y|t)$ and $\hat{F}_X(t)$ is the empirical distribution function of the X_i. Noting that model (1.2) implies $F(y|x) = F_e\left(\frac{y-m(x)}{\sigma(x)}\right)$, where F_e is the distribution function of ε, Van Keilegom and Akritas (1999) used the estimator

$$\hat{F}(y|x) = \hat{F}_e\left(\frac{y - \hat{m}(x)}{\hat{\sigma}(x)}\right), \tag{1.5}$$

where $\hat{m}(x), \hat{\sigma}(x)$ are nonparametric regression estimators of m and σ and $\hat{F}_e$ is the usual Kaplan-Meier estimator evaluated from $(Z_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$. The estimator (1.5) for the conditional distribution function and relation (1.3) lead to

$$\hat{F}(x,y) = \int_{-\infty}^{x} \hat{F}_e\left(\frac{y - \hat{m}(t)}{\hat{\sigma}(t)}\right) d\hat{F}_X(t) \tag{1.6}$$

as an estimator for the bivariate distribution. Note that (1.5) will estimate well the tails of $F(y|x)$, even if there is heavy censoring at $X = x$, provided there is a region of x-values where the censoring is 'light'. This statement is made more precise in the next section.

The idea for estimating the parameters in a polynomial heteroscedastic regression model with censored data is to express the ordinary least squares estimator as a functional of the (bivariate) empirical distribution function of (X, Y), and replace the uncensored data empirical distribution function with $\hat{F}(x, y)$. A similar idea was also used in Akritas (1994), and in Stute (1993, 1996), but these estimators suffer from the disadvantage that their asymptotic bias increases as the censoring in the upper tails increases. In addition, the estimator in Stute (1993, 1996) uses the often unrealistic assumption that the censoring variable is independent from the response variable. As noted above, the present estimator minimizes the undesirable effects of heavy censoring in the upper tails.

In the next section we give the precise definition of the least squares estimator, and we state the assumptions. The main result on the asymptotic normality of the least squares estimator is presented in Section 3, while the proofs are given in Section 4. This work is part of Van Keilegom (1998) where more details can be found.

2 Definitions and Assumptions

Let $(X_i, Y_i, C_i, Z_i, \Delta_i)$, $i = 1, \ldots, n$, be n independent random vectors as defined in Section 1 and let (X, Y, C, Z, Δ) have the joint distribution of each $(X_i, Y_i, C_i, Z_i, \Delta_i)$. We assume that regression model (1.2) is valid, where m and σ are assumed to be a location and scale functional, respectively. Note that if in model (1.2) m or σ is replaced by another location functional $m^\star$ or scale functional $\sigma^\star$, the resulting error variables $\varepsilon_i^\star$, $i = 1, \ldots, n$, are still i.i.d. and each $\varepsilon_i^\star$ is independent of X_i. (The terms 'location' and 'scale' functional are used here in the sense defined, e.g., in Huber 1981, p. 59, 202). We do not restrict Y to be positive, but allow Y to represent any monotone transformation (such as e.g. the logarithm) of the survival time of a patient involved in a clinical study. We will use the abbreviation E for $\frac{Z-m(X)}{\sigma(X)}$ and the notations $F(x,y) = P(X \leq x, Y \leq y)$, $F_X(x) = P(X \leq x)$, $F(y|x) = P(Y \leq y|x)$, $G(y|x) = P(C \leq y|x)$, $H(y|x) = P(Z \leq y|x)$ and $H_1(y|x) = P(Z \leq y, \Delta = 1|x)$. Assuming conditional independence of Y and C given X, entails that $1 - H(y|x) = (1 - F(y|x))(1 - G(y|x))$. Furthermore, set $F_e(y) = P(\varepsilon \leq y), G_e(y) = P(\frac{C-m(X)}{\sigma(X)} \leq y), H_e(y) = P(E \leq y)$, $H_{e1}(y) = P(E \leq y, \Delta = 1)$, $H_e(y|x) = P(E \leq y|x)$ and $H_{e1}(y|x) = P(E \leq y, \Delta = 1|x)$. Finally, we use lower case letters for the probability density functions of the distribution functions defined above.

Throughout the paper we will use the following location and scale functionals, which are a special case of the general functional for L-statistics (see e.g. Serfling 1980)

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

where $J(s)$ is a given score function satisfying $\int J(s)ds = 1$, and $F^{-1}(s|x) = \inf\{t; F(t|x) \geq s\}$ is the quantile function of Y given x. The motivation behind this choice of location functional is that by proper choice of J, the right tails of $F(y|x)$ (which may be poorly estimated due to the censoring mechanism) are not involved. Natural estimators for $m(x)$ and $\sigma(x)$ are

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

where $\tilde{F}$ is the conditional Kaplan-Meier estimator introduced by Beran (1981)

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

and

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

where K is a known probability density function (kernel) and $\{a_n\}$ is a sequence of positive constants tending to zero as $n \to \infty$ (bandwidth).

Let $\hat{E}_i = \frac{Z_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)}$. We estimate the distributions H_e and F_e by, respectively, the usual empirical distribution function based on the $\hat{E}_i$'s and the Kaplan-Meier (1958) estimator based on the $(\hat{E}_i, \Delta_i)$'s. Note that $\hat{E}_1, \ldots, \hat{E}_n$ are not independent. Thus,

$$\hat{H}_e(y) = n^{-1}\sum_{i=1}^n I(\hat{E}_i \leq y), \quad \hat{F}_e(y) = 1 - \prod_{\hat{E}_{(i)} \leq y, \Delta_{(i)}=1} \left(\frac{n-i}{n-i+1}\right),$$

where $\hat{E}_{(i)}$ is the i-th order statistic of $\hat{E}_1, \ldots, \hat{E}_n$ and $\Delta_{(i)}$ is the corresponding indicator. This leads to the estimator in (1.6). A sufficient condition under which $\hat{F}(x, y)$ has no regions of undefined mass, is that there exists a subset R of the support of X such that $\tau_{F(\cdot|x)} \leq \tau_{G(\cdot|x)}$, for all x in R, where τ_F denotes the upper bound of the support of any distribution function F. Indeed, this condition guarantees that $\tau_{F_e} \leq \tau_{G_e}$ and hence $F_e(y)$ will be well estimated for all y.

Because the asymptotic theory for (1.6) is based on an i.i.d. representation for $\hat{F}_e$ which is valid up to any point smaller than $\tau_{F_e} \wedge \tau_{G_e}$, Van Keilegom and Akritas (1999) worked with the slightly modified version of (1.6),

$$\hat{F}_T(x, y) \;=\; \int_{-\infty}^{x} \hat{F}_e\left(\frac{y \wedge T_t - \hat{m}(t)}{\hat{\sigma}(t)}\right) d\hat{F}_X(t), \tag{2.3}$$

where the $T_t \leq T\sigma(t) + m(t)$ for $t \in R_X$ and where $T < \tau_{H_e}$. This is actually an estimator of $F_T(x, y) = \int_{-\infty}^{x} F_e\left(\frac{y \wedge T_t - m(t)}{\sigma(t)}\right) dF_X(t)$, which can become arbitrarily close to $F(x, y)$ if $\tau_{F_e} \leq \tau_{G_e}$ and T_t, respectively T, is chosen sufficiently close to $T\sigma(t) + m(t)$, respectively τ_{H_e}, for all t. The asymptotic properties of $\hat{F}_T(x, y)$, as well as of $\hat{F}_e(y)$ and $\hat{F}(y|x)$, have been studied in Van Keilegom and Akritas (1999). We refer to that paper for details.

Consider now the heteroscedastic polynomial regression model (1.1). Thus we assume that $E(\mathbf{Y}|\mathcal{X}) = \mathcal{X}\beta$, where $\mathbf{Y}$ is the $n \times 1$ response vector and $\mathcal{X}$ denotes the $n \times (p+1)$ design matrix whose k-th column contains the elements $X_i^k, i = 1, \ldots, n$. The least squares estimator is defined as the value of β which minimizes

$$\int (y - \beta_0 - \beta_1 x - \ldots - \beta_p x^p)^2 d\hat{F}_T(x, y) \;. \tag{2.4}$$

Note that $\hat{F}_T(x,y)$ does not use the particular structure of $m(x)$ implied by the model (1.1). Also, since (1.1) holds, model (1.2) holds for any location functional m and scale functional σ. Provided $\hat{F}_e(\infty)=1$, the solution to the minimization problem in (2.4) is

$$\hat{\beta}_T = n(\mathcal{X}'\mathcal{X})^{-1}\begin{pmatrix} \int x^0 y\, d\hat{F}_T(x,y) \\ \vdots \\ \int x^p y\, d\hat{F}_T(x,y) \end{pmatrix}. \tag{2.5}$$

This is the same as the uncensored data least squares estimator with $\hat{F}_T(x,y)$ replacing the usual bivariate empirical distribution function corresponding to (X_i, Y_i). Similarly, define

$$\beta_T = n[E(\mathcal{X}'\mathcal{X})]^{-1}\begin{pmatrix} \int x^0 y\, dF_T(x,y) \\ \vdots \\ \int x^p y\, dF_T(x,y). \end{pmatrix}.$$

Remark 2.1. Direct calculations reveal that in the homoscedastic linear regression model with uncensored data, the proposed slope estimator is the same as that obtained by least squares on the pairs $(X_i, \hat{m}(X_i))$, $i=1,\ldots,n$. Compare with Akritas (1996) and Cristóbal Cristóbal et al. (1987). In particular, the latter paper shows that a similar class of estimators in the uncensored homoscedastic case includes the ordinary least squares estimator as well as the ordinary and generalized ridge regression estimators as special cases by appropriate choices for the kernel function and the window.

Of course, $\hat{\beta}_T$ estimates β_T which becomes arbitrarily close to the true β by suitable choice of T, provided $\tau_{F_e} \le \tau_{G_e}$. The results listed below will be shown under the following conditions. Let $\tilde{T}_x (x \in R_X)$ be such that $\tilde{T}_x < \tau_{H(\cdot|x)}$ and $\inf_{x\in R_X}(1 - H(\tilde{T}_x|x)) > 0$.

(A1)(i) The sequence a_n satisfies $na_n^4 \to 0$ and $na_n^{3+2\delta}(\log a_n^{-1})^{-1} \to \infty$ for some $\delta > 0$.
(ii) The support R_X of X is bounded, convex and its interior is not empty.
(iii) The probability density function K has compact support, $\int uK(u)\,du = 0$ and K is twice continuously differentiable.

(A2)(i) The function $x \to T_x$ $(x \in R_X)$ is twice continuously differentiable and 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) The function J is twice continuously differentiable, $\int_0^1 J(s)\,ds = 1$ and $J(s) \geq 0$ for all $0 \leq s \leq 1$.

(A3)(i) The distribution F_X is thrice continuously differentiable and $\inf_{x \in R_X} f_X(x) > 0$.
(ii) The functions m and σ are twice continuously differentiable and $\inf_{x \in R_X} \sigma(x) > 0$.
(iii) The error variable ε has finite expectation.

(A4) The functions

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

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

are twice continuously differentiable with respect to x, their first and second derivatives (with respect to x) are bounded, uniformly in $x \in R_X, z < \tilde{T}_x$ and δ, and for any $\delta = 0, 1$, the first derivatives (considered as functions in z) are of bounded variation and the variation norms are uniformly bounded over all x.

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

Let $L(y|x)$ stand for either $H(y|x)$, $H_1(y|x)$, $H_e(y|x)$ or $H_{e1}(y|x)$. We will use the notations $l(y|x) = L'(y|x) = \frac{\partial}{\partial y} L(y|x), \dot{L}(y|x) = \frac{\partial}{\partial x} L(y|x)$ and similar notations will be used for higher order derivatives.
(A6)(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 $\sup_{x,y} |y^2 L''(y|x)| < \infty$.
(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) $\dot{L}'(y|x)$ exists, is continuous in (x, y) and $\sup_{x,y} |y\dot{L}'(y|x)| < \infty$.
(vii) $\ddot{L}'(y|x)$ exists, is continuous in (x, y) and $\sup_{x,y} |y\ddot{L}'(y|x)| < \infty$.
(viii) $L'''(y|x)$ exists, is continuous in (x, y) and $\sup_{x,y} |y^3 L'''(y|x)| < \infty$.

Throughout the paper, the symbol K will be used for any constant, whose value may differ from line to line.

3 Main Results

This section presents the asymptotic theory of the least squares (LS) estimator. The asymptotic representation for the estimator involves the functions $\eta(z,\delta|x)$ and $\zeta(z,\delta|x)$ defined in assumption (A4) and

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

where $S_e(y) = 1 - F_e(y)$,

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

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

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

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

We use the abbreviated notations $n^{-1}(\mathcal{X}'\mathcal{X}) = A_n = (\hat{a}_{rs})$ and $n^{-1}E(\mathcal{X}'\mathcal{X}) = A = (a_{rs})$. The $(k+1,\ell+1)$-th element of A^{-1} and A_n^{-1} will be denoted by, respectively, $g_{k\ell}(A)$ and $g_{k\ell}(A_n)$. Also, let $\tilde{G}_{rs}(A) = (\tilde{g}_{k\ell}^{rs}(A))$ be the matrix of the partial derivative of the elements $g_{k\ell}(A)$, $k,\ell = 0,\dots,p$, with respect to a_{rs}.

Theorem 3.1 *Assume (A1)-(A5), $H(y|x)$ and $H_1(y|x)$ satisfy (A6)(i)-(vi), $H_e(y|x)$ and $H_{e1}(y|x)$ satisfy (A6)(ii,iii,vi,vii,viii) and the p-variate distribution of $(X,X^2,\dots,X^p)$ is nonsingular. Then,*

$$\begin{pmatrix}\hat\beta_{T,0}-\beta_{T,0}\\ \vdots\\ \hat\beta_{T,p}-\beta_{T,p}\end{pmatrix} = n[E(\mathcal{X}'\mathcal{X})]^{-1}\begin{pmatrix}n^{-1}\sum_{i=1}^n\Psi_0(X_i,Z_i,\Delta_i)\\ \vdots\\ n^{-1}\sum_{i=1}^n\Psi_p(X_i,Z_i,\Delta_i)\end{pmatrix} \tag{3.1}$$

$$+ n[E(\mathcal{X}'\mathcal{X})]^{-1} \begin{pmatrix} \int x^0 \int y \, dF_e(\frac{y \wedge T_x - m(x)}{\sigma(x)}) \, d(\hat{F}_X(x) - F_X(x)) \\ \vdots \\ \int x^p \int y \, dF_e(\frac{y \wedge T_x - m(x)}{\sigma(x)}) \, d(\hat{F}_X(x) - F_X(x)) \end{pmatrix}$$

$$+ \sum_{r,s=0}^{p} \tilde{G}_{rs}(A) \begin{pmatrix} \int x^0 y \, dF_T(x,y) \\ \vdots \\ \int x^p y \, dF_T(x,y) \end{pmatrix} \int x^{r+s} \, d(\hat{F}_X(x) - F_X(x)) + \begin{pmatrix} o_P(n^{-1/2}) \\ \vdots \\ o_P(n^{-1/2}) \end{pmatrix},$$

where for $k = 0, \ldots, p$,

$$\begin{aligned} \Psi_k(t,z,\delta) &= \int_{R_X} x^k \int_{-\infty}^{T_x} y \, d\varphi\left(t, z, \delta, \frac{y - m(x)}{\sigma(x)}\right) dF_X(x) \\ &+ t^k \int_{-\infty}^{T_t} y \, d\left[\left(\eta(z,\delta|t) + \zeta(z,\delta|t)\frac{y-m(t)}{\sigma(t)}\right) f_e\left(\frac{y-m(t)}{\sigma(t)}\right)\right]. \end{aligned}$$

Theorem 3.2 *Let the assumptions imposed in Theorem 3.1 hold. Then,* $n^{1/2}(\hat{\beta}_T - \beta_T) \xrightarrow{d} N(\mathbf{0}; \mathbf{\Sigma})$, *where* $\mathbf{\Sigma} = (\sigma_{k\ell})$ *and*

$$\begin{aligned} \sigma_{k\ell} &= Cov\left\{\sum_{i=0}^{p} g_{ki}(A)\Psi_i(X,T,\Delta), \sum_{i=0}^{p} g_{\ell i}(A)\Psi_i(X,T,\Delta)\right\} \\ &+ Cov\left\{\sum_{i=0}^{p} g_{ki}(A)X^i \int t dF_e\left(\frac{t \wedge \bar{T}_X - m(X)}{\sigma(X)}\right) + \sum_{i,r,s=0}^{p} \tilde{g}_{ki}^{rs}(A) \int x^i t dF_T(x,t) X^{r+s},\right. \\ &\left. \sum_{i=0}^{p} g_{\ell i}(A)X^i \int t dF_e\left(\frac{t \wedge \bar{T}_X - m(X)}{\sigma(X)}\right) + \sum_{i,r,s=0}^{p} \tilde{g}_{\ell i}^{rs}(A) \int x^i t dF_T(x,t) X^{r+s}\right\}. \end{aligned}$$

The proof of this result is straightforward.

4 Proofs

Proof of Theorem 3.1. Write

$$\begin{pmatrix} \hat{\beta}_{T,0} - \beta_{T,0} \\ \vdots \\ \hat{\beta}_{T,p} - \beta_{T,p} \end{pmatrix} = n(\mathcal{X}'\mathcal{X})^{-1} \begin{pmatrix} \int x^0 y \, d(\hat{F}_T(x,y) - F_T(x,y)) \\ \vdots \\ \int x^p y \, d(\hat{F}_T(x,y) - F_T(x,y)) \end{pmatrix} \quad (4.1)$$

$$+ \quad n((\mathcal{X}'\mathcal{X})^{-1} - [E(\mathcal{X}'\mathcal{X})]^{-1}) \begin{pmatrix} \int x^0 y \, dF_T(x,y) \\ \vdots \\ \int x^p y \, dF_T(x,y) \end{pmatrix}.$$

We start with the first term on the right hand side of (4.1).

$$\begin{aligned} & \int x^k y \, d(\hat{F}_T(x,y) - F_T(x,y)) \qquad (4.2) \\ = & \int x^k \int y \, d\left[\hat{F}_e\left(\frac{y \wedge T_x - \hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{y \wedge T_x - m(x)}{\sigma(x)}\right)\right] d\hat{F}_X(x) \\ + & \int x^k \int y \, dF_e\left(\frac{y \wedge T_x - m(x)}{\sigma(x)}\right) d(\hat{F}_X(x) - F_X(x)). \end{aligned}$$

Using integration by parts, the first term on the right hand side of (4.2) can be written as

$$\begin{aligned} - & \int\int_{R_X} x^k I(y \le T_x) \left[\hat{F}_e\left(\frac{y - \hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{y - m(x)}{\sigma(x)}\right)\right] d\hat{F}_X(x) \, dy \qquad (4.3) \\ + & \int x^k T_x \left[\hat{F}_e\left(\frac{T_x - \hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{T_x - m(x)}{\sigma(x)}\right)\right] d\hat{F}_X(x). \end{aligned}$$

This is somewhat similar to the first term of equation (5.4) in the proof of Theorem 3.5 in Van Keilegom and Akritas (1999), and a similar derivation shows that (4.3) equals

$$\begin{aligned} - & \; n^{-1} \sum_{i=1}^n \int\int_{R_X} x^k I(y \le T_x) \varphi\left(X_i, Z_i, \Delta_i, \frac{y - m(x)}{\sigma(x)}\right) dF_X(x) \, dy \\ - & \; n^{-1} \sum_{i=1}^n \int_{-\infty}^{T_{X_i}} X_i^k \left[\eta(Z_i, \Delta_i | X_i) + \zeta(Z_i, \Delta_i | X_i) \frac{y - m(X_i)}{\sigma(X_i)}\right] f_e\left(\frac{y - m(X_i)}{\sigma(X_i)}\right) dy \\ + & \; n^{-1} \sum_{i=1}^n \int x^k T_x \varphi\left(X_i, Z_i, \Delta_i, \frac{T_x - m(x)}{\sigma(x)}\right) dF_X(x) \\ + & \; n^{-1} \sum_{i=1}^n X_i^k T_{X_i} \left[\eta(Z_i, \Delta_i | X_i) + \zeta(Z_i, \Delta_i | X_i) \frac{T_{X_i} - m(X_i)}{\sigma(X_i)}\right] f_e\left(\frac{T_{X_i} - m(X_i)}{\sigma(X_i)}\right) \\ + & \; o_P(n^{-1/2}) \end{aligned}$$

and this equals $n^{-1} \sum_{i=1}^n \Psi_k(X_i, Z_i, \Delta_i) + o_P(n^{-1/2})$. For the second term on the right hand side of (4.1) we have that

$$n((\mathcal{X}'\mathcal{X})^{-1} - [E(\mathcal{X}'\mathcal{X})]^{-1}) = A_n^{-1} - A^{-1} = (g_{kl}(A_n) - g_{kl}(A)).$$

Since the distribution of $(X, X^2, \ldots, X^p)$ is nonsingular, it follows that A has full rank and using similar techniques as in Theorem 17.8 in Arnold (1981), we also have that A_n has full rank with probability 1. Hence, a first order Taylor expansion applied to each $g_{kl}(A_n) - g_{kl}(A)$ yields that $A_n^{-1} - A^{-1}$ is asymptotically equivalent to

$$\sum_{r=0}^{p}\sum_{s=0}^{p} \tilde{G}_{rs}(A)(\hat{a}_{rs} - a_{rs}) = \sum_{r=0}^{p}\sum_{s=0}^{p} \tilde{G}_{rs}(A) \int x^{r+s}\, d(\hat{F}_X(x) - F_X(x)). \tag{4.4}$$

Finally, by the use of Bernstein's inequality (Uspensky 1937), it is seen that $n^{-1}\sum_{i=1}^{n} \Psi_k(X_i, Z_i, \Delta_i)$ is $O(n^{-1/2}(\log n)^{1/2})$ a.s. Therefore, since the second term on the right hand side of (4.2) and the integral in (4.4) are $O(n^{-1/2}(\log n)^{1/2})$ a.s., we can replace the factor $n(\mathcal{X}'\mathcal{X})^{-1}$ in the first term on the right hand side of (4.1) by $n[E(\mathcal{X}'\mathcal{X})]^{-1}$. This finishes the proof.

References

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

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

Akritas, M. G., Murphy, S. and LaValley, M. (1995). The Theil-Sen estimator with doubly censored data and applications to astronomy. *J. Amer. Statist. Assoc.* **90** 170-177.

Arnold, S. F. (1981). *The theory of linear models and multivariate analysis.* Wiley, New York.

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

Cristóbal Cristóbal, J. A., Faraldo Roca, P. and González Manteiga, W. (1987). A class of linear regression parameter estimators constructed by nonparametric estimation. *Ann. Statist.* **15** 603-609.

Fygenson, M. and Ritov, Y. (1994). Monotone estimating equations for censored data. *Ann. Statist.* **22** 732-746.

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

Huber, P. J. (1981). *Robust statistics.* Wiley, New York.

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

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

Lai, T. and Ying, Z. (1991b). Rank regression methods for left truncated and right censored data. *Ann. Statist.* **19** 531-554.

Lai, T. and Ying, Z. (1994). A missing information principle and M-estimators in regression analysis with censored or truncated data. *Ann. Statist.* **22** 1222-1255.

LeBlanc, M. and Crowley, J. (1995). Semiparametric regression functionals. *J. Amer. Statist. Assoc.* **90** 95-105.

Lin, D.Y. and Wei, L.J. (1992a). Linear regression analysis based on Buckley-James estimating equations. *Biometrics* **48** 679-681.

Lin, D.Y. and Wei, L.J. (1992b). Linear regression analysis of multivariate failure time observations. *J. Amer. Statist. Assoc.* **87** 1091-1097.

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

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

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

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.* **23** 461-471.

Tsiatis, A.A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18** 354-372.

Uspensky, J. V. (1937). *Introduction to Mathematical Probability.* McGraw-Hill, New York.

Van Keilegom, I. (1998). *Nonparametric estimation of the conditional distribution in regression with censored data.* Ph.D. thesis, Limburgs Universitair Centrum.

Van Keilegom, I. and Akritas, M. G. (1999). Transfer of tail information in censored regression models. *Ann. Statist.* (to appear).

Wei, L.J., Ying, Z. and Lin, D. (1990). Linear regression analysis of censored survival data based on linear rank tests. *Biometrika* **77** 845-851.

Ying, Z., Jung, S.H. and Wei, L.J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90** 178-184.

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

Zhou, M. (1992b). M-estimation in censored linear models. *Biometrika* **79** 837-841.