Available online at www.sciencedirect.com

ScienceDirect

Journal of Statistical Planning and Inference 137 (2007) 3894–3903

journal of statistical planning and inference

www.elsevier.com/locate/jspi

M-estimation in regression models for censored data

Zhezhen Jin*

Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 W 168th Street, New York, NY 10032, USA

Received 30 December 2005; received in revised form 7 September 2006; accepted 6 April 2007
Available online 4 May 2007

Abstract

In this paper, we study M-estimators of regression parameters in semiparametric linear models for censored data. A class of consistent and asymptotically normal M-estimators is constructed. A resampling method is developed for the estimation of the asymptotic covariance matrix of the estimators.

Keywords: Accelerated failure time model; M-estimator; Rank estimator; Resampling; Semiparametric model; Survival data; Variance estimation

1. Introduction

A way to study the covariate effects on the response is to use the semiparametric linear regression model in which the response is sum of a linear combination of covariates and error term whose distribution is completely unspecified.

For uncensored data, it is well known that the use of M-estimation methods for the model can overcome some of the robustness limitations of the least squares approach (Huber, 1981; Hampel et al., 1986). For right censored data, Ritov (1990), Zhou (1992) and Lai and Ying (1994) developed M-estimators for the model. The approach of Zhou (1992) is based on a weighted data approach similar to that in Koul et al. (1981), which requires that the censoring time is independent of covariates. The approaches of Ritov (1990) and Lai and Ying (1994) are extensions of the estimator of Buckley and James (1979) to general M-estimators. Under certain conditions, both Ritov (1990) and Lai and Ying (1994) rigorously proved that their estimating equation has a solution that is $\sqrt{n}$-consistent and asymptotically normal. But these results do not guarantee consistency of all solutions. They were not able to show that the estimating equation has a unique solution, nor were they able to devise an algorithm that could be used in finding a consistent estimator. Also, the variance of resulting estimator is difficult to estimate well because it involves the unknown density function of error term and its derivative.

In this note, we develop a new M-estimation procedure for the semiparametric linear model with right censored data. The method is easy to implement and yields a class of estimators which are consistent and asymptotically normal. We also develop a resampling method for the estimation of the limiting covariance matrix.

In the next section, the semiparametric linear regression model is introduced and the new M-estimation method is presented. Simulation studies are presented in Section 3 and some remark is given in Section 4. Proofs are outlined in Appendix A.

* Tel.: +1 212 305 9404; fax: +1 212 305 9408.
E-mail addresses: zj7@columbia.edu, zjin@biostat.columbia.edu (Z. Jin).

0378-3758/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2007.04.008

2. Model and *M*-estimation

Let (Y_i, X_i, δ_i), $i = 1, \ldots, n$, be independent and identically distributed observed data, where X_i is a $p \times 1$ covariate, $Y_i = \min\{T_i, C_i\}$ with T_i being failure time and C_i being the censoring time, $\delta_i = 1\{T_i \leqslant C_i\}$ is the failure indicator. Throughout the paper, it is assumed that T_i and C_i are independent conditional on X_i.

2.1. Accelerated failure time model

The semiparametric linear regression model is of form

$$T_i = X_i^{\mathrm{T}} \beta_0 + \varepsilon_i, \tag{1}$$

where β_0 is the unknown true $p \times 1$ parameter of interest and ε_i $(i = 1, \ldots, n)$ are unobservable independent random errors with a common but completely unspecified distribution function $F(\cdot)$. (Thus, the mean of ε is not necessarily 0). In survival analysis in which the response is the logarithm transformation of the survival time, the model (1) is the accelerated failure time (AFT) model.

2.2. M-estimation

Ritov (1990), Lai and Ying (1994) considered the following estimating equation $\Psi(\beta; s) = 0$ with $s(\cdot)$ being a known function and

$$\Psi(\beta; s) = \sum_{i=1}^{n} (X_i - \bar{X}) \left\{ \delta_i s(Y_i - X_i^{\mathrm{T}} \beta) + (1 - \delta_i) \frac{\int_{Y_i - X_i^{\mathrm{T}} \beta}^{\infty} s(u) \, \mathrm{d}\hat{F}_{n,\beta}(u)}{1 - \hat{F}_{n,\beta}(Y_i - X_i^{\mathrm{T}} \beta)} \right\}, \tag{2}$$

where $\bar{X} = (1/n) \sum_{i=1}^{n} X_i$, and $\hat{F}_{n,b}(t)$ is the Kaplan–Meier estimator based on data $\{(Y_i - X_i^{\mathrm{T}} b, \delta_i), i = 1, \ldots, n\}$. Specifically,

$$\hat{F}_{n,b}(t) = 1 - \prod_{i: e_i(b) < t} \left[1 - \frac{\delta_i}{\sum_{j=1}^{n} 1\{e_j(b) \geqslant e_i(b)\}} \right], \tag{3}$$

where $e_i(b) = Y_i - X_i^{\mathrm{T}} b$, $i = 1, \ldots, n$.

Notice that, if $s(u) = u$, then (2) becomes the Buckley–James estimating function based on least squares principle (Buckley and James, 1979); if $s(u) = |u|^{r-1}$ with $r > 1$, then (2) becomes estimating function of L_r regression; if $s(u) = \mathrm{sign}(u)$, then (2) becomes estimating function of least absolute deviation; and (2) becomes the maximum likelihood estimating function when $s(u) = -\dot{f}(u)/f(u)$, where $f(u)$ is the density function of ε and $\dot{f}(u)$ is its derivative.

Recently, Jin et al. (2006a) studied the least squares approach with $s(u) = u$ for the model (2) along the line of Buckley and James (1979). Below we extend their approach for general function $s(u)$. Specifically, with an initial estimator $\hat{\beta}_{(0)}$, which is $\sqrt{n}$ consistent, obtain a new estimator $\hat{\beta}_{(1)}$ by the one-step Newton–Raphson approximation:

$$\hat{\beta}_{(1)} = \hat{\beta}_{(0)} - \left[\left. \frac{\partial \mathrm{U}(\beta; \hat{\beta}_{(0)}, s)}{\partial \beta} \right|_{\beta = \hat{\beta}_{(0)}} \right]^{-1} \mathrm{U}(\hat{\beta}_{(0)}; \hat{\beta}_{(0)}, s), \tag{4}$$

where

$$\mathrm{U}(\beta; b, s) = \sum_{i=1}^{n} (X_i - \bar{X}) \left\{ s(e_i(\beta)) + (1 - \delta_i) \left[\frac{\int_{e_i(b)}^{\infty} s(u) \, \mathrm{d}\hat{F}_{n,b}(u)}{1 - \hat{F}_{n,b}(e_i(b))} - s(e_i(b)) \right] \right\}. \tag{5}$$

Continuing this process yields a class of estimators $\hat{\beta}_{(m)}$:

$$\hat{\beta}_{(m)} = \hat{\beta}_{(m-1)} - \left[\left. \frac{\partial \mathrm{U}(\beta; \hat{\beta}_{(m-1)}, s)}{\partial \beta} \right|_{\beta = \hat{\beta}_{(m-1)}} \right]^{-1} \mathrm{U}(\hat{\beta}_{(m-1)}; \hat{\beta}_{(m-1)}, s) \tag{6}$$

for $m \geqslant 1$.

Under regularity conditions, it can be shown that, if a consistent estimator of β_0 is chosen as the initial value, then, for any fixed m, $\hat{\beta}_{(m)}$ should also be consistent. In addition, $\hat{\beta}_{(m)}$ is asymptotically normal if the initial estimator is asymptotically normal. Actually, it can be shown that $\hat{\beta}_{(m)}$ is a linear combination of initial estimator and targeted estimator of the estimating equation $\mathrm{U}(\beta; \beta, s) = 0$, see Appendix A.

A consistent and asymptotically normal initial estimator of β_0 can be obtained by the rank-based method of Jin et al. (2003) or by the least squares based method of Jin et al. (2006a). In particular, the Gehan-type rank estimator $\hat{\beta}_{\mathrm{G}}$ can be used as the initial estimator $\hat{\beta}_{(0)}$, which can be calculated by minimizing the following convex function:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i [e_i(\beta) - e_j(\beta)]^-,$$

where $a^- = 1\{a < 0\}|a|$. The minimization can be done with linear programming method (Jin et al., 2003). Throughout the sequel, we assume that the initial estimator is $\hat{\beta}_{\mathrm{G}}$.

The following assumptions will be made:

(C1) $s(\cdot)$ is twice continuously differentiable and $\int_{-\infty}^{\infty} s^2(t)\,\mathrm{d}F(t) < \infty$.
(C2) $\|X_i\| \leqslant B$ for all i for some nonrandom constant B.
(C3) $F(\cdot)$ has a twice continuously differentiable densify function $f(\cdot)$ such that

$$\int_{-\infty}^{\infty} t^2\,\mathrm{d}F(t) < \infty, \quad \int_{-\infty}^{\infty} \left[\frac{\dot{f}(t)}{f(t)} \right]^2 \mathrm{d}F(t) < \infty,$$
$$\int_{-\infty}^{\infty} \sup_{|h| \leqslant h_0} \{|\dot{f}(t+h)| + |\ddot{f}(t+h)|\}\,\mathrm{d}t < \infty \quad \text{for some } h_0 > 0.$$

(C4) $D_s = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} (X_i - \overline{X}) X_i^{\mathrm{T}} \dot{s}(Y_i - X_i^{\mathrm{T}} \beta_0)$ exists and is nonsingular, where $\dot{s}(\cdot)$ is the first derivative of $s(\cdot)$.
(C5) $\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} (X_i - \overline{X})^{\otimes j} \mathrm{P}(C_i - X_i^{\mathrm{T}} \beta_0 \geqslant t | X_i) = \Gamma_j(t)$ exists for $j = 0, 1, 2$, where $a^{\otimes 0}$ is the identity matrix, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa^{\mathrm{T}}$. In addition, the matrix A_s is nonsingular, where

$$A_s = \int_{-\infty}^{\infty} \left[\left\{ \Gamma_2(t) - \Gamma_0^{-1}(t) \Gamma_1^{\otimes 2}(t) \right\} \int_t^{\infty} \{1 - F(v)\} \dot{s}(v)\,\mathrm{d}v \right] \mathrm{d}\lambda(t).$$

(C6) The matrix A_{G} is nonsingular, where

$$A_{\mathrm{G}} = \lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \int_{-\infty}^{\infty} 1\{e_j(\beta_0) \geqslant t\} (X_i - X_j)^{\otimes 2} \{\dot{\lambda}(t)/\lambda(t)\}\,\mathrm{d}N_i(\beta_0, t),$$

$$\lambda(t) = f(t)/(1 - F(t)) \quad \text{and} \quad N_i(\beta_0, t) = \delta_i 1\{e_j(\beta_0) \geqslant t\}.$$

The condition (C2) is a technical assumption which may be relaxed to $\max_i \|X_i\| = \mathrm{O}(n^{\alpha})$ for any $\alpha > 0$. The condition (C3) is assumed to obtain linear expansion of $\mathrm{U}(\beta; b, s)$. The conditions (C4) and (C5) hold if the vector of covariates does not lie in a lower-dimensional hyperplane. The condition (C6) is assumed to obtain the asymptotic results of the initial estimate $\hat{\beta}_{\mathrm{G}}$.

Theorem 1. *Under the conditions* (C1)–(C6), $\hat{\beta}_{(m)}$ *is consistent and* $\sqrt{n}(\hat{\beta}_{(m)} - \beta_0)$ *is asymptotically normally distributed as* $n \to \infty$ *for each* $m \geqslant 1$.

The proof of Theorem 1 is given in Appendix A.

The limiting covariance matrices of both $\hat{\beta}_{\mathrm{G}}$ and $\hat{\beta}_{(m)}$, however, involve the unknown hazard function $\lambda(\cdot)$ of error term ε. To estimate the limiting covariance matrices, we develop a resampling procedure to approximate the distribution of $\hat{\beta}_{(m)}$.

2.3. Variance estimation

Let $\hat{\beta}^*_{\mathrm{G}}$ be a minimizer of

$$\sum_{i=1}^{n}\sum_{j=1}^{n} Z_i Z_j \delta_i [e_i(\beta) - e_j(\beta)]^-,$$

where Z_i $(i = 1, \ldots, n)$ satisfy following condition (C7):

(C7) Z_i $(i = 1, \ldots, n)$ are independent and identically distributed positive random variables with a known distribution satisfying $E(Z_i) = \mathrm{var}(Z_i) = 1$.

Furthermore, define

$$\mathrm{U}^*(\beta; b, s) = \sum_{i=1}^{n} Z_i (X_i - \bar{X})[s(e_i(\beta)) + (1 - \delta_i)\xi_{is}(b) - \bar{\xi}_s(b)], \tag{7}$$

where

$$\xi_{is}(b) = \frac{\int_{e_i(b)}^{\infty} s(u)\,\mathrm{d}\hat{F}^*_{n,b}(u)}{1 - \hat{F}^*_{n,b}(e_i(b))} - s(e_i(b))$$

with

$$\hat{F}^*_{n,b}(t) = 1 - \prod_{i:e_i(b)<t}\left[1 - \frac{Z_i\delta_i}{\sum_{j=1}^{n} Z_j 1\{e_j(b) \geqslant e_i(b)\}}\right], \tag{8}$$

and $\bar{\xi}_s(b) = n^{-1}\sum_{i=1}^{n}[s(e_i(b) + (1 - \delta_i)\xi_{is}(b)]$. Then define the iterative sequence: $\hat{\beta}^*_{(0)} = \hat{\beta}^*_{\mathrm{G}}$ and

$$\hat{\beta}^*_{(m)} = \hat{\beta}^*_{(m-1)} - \left[\left.\frac{\partial \mathrm{U}^*(\beta; \hat{\beta}^*_{(m-1)}, s)}{\partial \beta}\right|_{\beta=\hat{\beta}^*_{(m-1)}}\right]^{-1} \mathrm{U}^*(\hat{\beta}^*_{(m-1)}; \hat{\beta}^*_{(m-1)}, s) \tag{9}$$

for $m \geqslant 1$.

Let $\mathscr{F}_n$ denote the σ-field generated by the original data (Y_i, δ_i, X_i) $(i = 1, \ldots, n)$.

Theorem 2. *Under the conditions* (C1)–(C7), $\hat{\beta}^*_{(m)}$ *is consistent and*

$$P(\sqrt{n}(\hat{\beta}^*_{(m)} - \hat{\beta}_{(m)}) \in G | \mathscr{F}_n) - P(\sqrt{n}(\hat{\beta}_{(m)} - \beta_0) \in G) \to 0$$

in probability as $n \to \infty$ *for any finite p- dimensional rectangular G.*

The proof of Theorem 2 is given in Appendix A. Theorem 2 shows that the conditional distribution of $\sqrt{n}(\hat{\beta}^*_{(m)} - \hat{\beta}_{(m)})$ given the data (Y_i, δ_i, X_i) $(i = 1, \ldots, n)$ is asymptotically equivalent to the asymptotic distribution of $\sqrt{n}(\hat{\beta}_{(m)} - \beta_0)$. Consequently, the distribution of $\hat{\beta}_{(m)}$ can be approximated by a large number of realizations of $\hat{\beta}^*_{(m)}$ after repeatedly

Table 1
Summary statistics for the simulation studies

Error		Censoring (%)	L_r estimators							
			$r=1.5$				$r=2.5$			
			Bias	SE	SEE	CP	Bias	SE	SEE	CP
Normal	β_1	0	0.009	0.207	0.203	0.938	0.011	0.206	0.203	0.938
		25	0.011	0.222	0.222	0.944	0.012	0.221	0.223	0.955
		50	0.011	0.260	0.256	0.939	0.009	0.256	0.269	0.964
	β_2	0	−0.001	0.218	0.199	0.920	−0.001	0.221	0.199	0.922
		25	0.001	0.230	0.219	0.939	0.001	0.234	0.221	0.932
		50	0.006	0.265	0.253	0.932	0.005	0.264	0.265	0.937
Extremevalue	β_1	0	−0.003	0.268	0.272	0.946	0.000	0.248	0.246	0.950
		25	−0.013	0.289	0.295	0.950	−0.005	0.289	0.293	0.951
		50	−0.003	0.348	0.366	0.954	0.009	0.379	0.406	0.962
	β_2	0	0.003	0.277	0.265	0.929	0.007	0.259	0.244	0.941
		25	0.005	0.296	0.298	0.935	0.009	0.288	0.291	0.963
		50	0.019	0.357	0.371	0.941	0.022	0.369	0.396	0.964

Bias, bias of the parameter estimator; SE, standard error of the parameter estimator; SEE, mean of the standard error estimator; CP, coverage probability of the 95% Wald-type confidence interval.

generating the random variables $(Z_1, \ldots, Z_n)$. The empirical distribution of $\hat{\beta}^*_{(m)}$ can then be used to approximate the distribution of $\hat{\beta}_{(m)}$. The empirical percentiles of $\hat{\beta}^*_{(m)}$ or Wald method can be used to obtain confidence intervals for individual components of β_0. A typical way of generating Z_i is to use the standard exponential distribution which has mean 1 and variance 1.

3. Simulation studies

We conducted simulation studies to examine the finite sample performance of the proposed methods. The failure times were generated from the model

$$\log T = 2 + X_1 + X_2 + \varepsilon,$$

where X_1 was generated from the Bernoulli distribution with 0.5 success probability, X_2 was generated from normal distribution with mean 0 and standard deviation 0.5, and ε had the standard normal, extreme-value distribution. The censoring times were generated from the uniform distribution $\mathrm{U}[0, c]$, where c was chosen to yield a desired level of censoring. We considered a sample size of 100. The L_r norm with $r = 1.5$ and 2.5 was considered for the M-estimation. We estimated β_1 and β_2 with three iterations with the resampling being based on 1000 realizations of standard exponential random variables.

The results based on 1000 simulated data sets were summarized in Table 1. The parameter estimator is virtually unbiased. The resampling procedure estimated standard errors accurately. Most of the coverage probabilities of confidence intervals were close to the 0.95 confidence coefficient. The results also showed that the proposed methods perform slightly better for the extreme-value error case than the standard normal error case.

4. Remark

In this note, we propose a new class of M-estimation procedure for the semiparametric linear regression model for censored data. The M-estimator $\hat{\beta}$ satisfying $\Psi(\hat{\beta}; s) = 0$ can be obtained if the convergence of $\hat{\beta}_{(m)}$ is reached as $m \to \infty$. It is also noted that the proposed approach can be extended to the analysis of multiple events data including clustered failure time data along the line of Jin et al. (2006a, b).

Acknowledgments

I thank Dr. Melas and two anonymous referees for their helpful constructive comments. This research was supported in part by the National Science Foundation career award DMS-0134431.

Appendix A. Proofs of asymptotic results

We first prove the following Lemma. It will be used in the proof of Theorems 1 and 2.

Lemma 1. *Under the conditions* (C1) *and* (C3),

$$\delta_i s(e_i(\beta_0)) + (1-\delta_i)\frac{\int_{e_i(\beta_0)}^{\infty} s(u)\,\mathrm{d}F(u)}{1-F(e_i(\beta_0))} = Es(\varepsilon_i) + \int_{-\infty}^{\infty}\left\{s(t) - \frac{\int_t^{\infty} s(u)\,\mathrm{d}F(u)}{1-F(t)}\right\}\mathrm{d}M_i(t), \tag{A.1}$$

where $M_i(t) = N_i(\beta_0, t) - \int_0^t I\{e_i(\beta_0) \geqslant u\}\lambda(u)\,\mathrm{d}u$.

Proof. Notice that

$$\begin{aligned}&\int_{-\infty}^{\infty}\left\{s(t) - \frac{\int_t^{\infty} s(u)\,\mathrm{d}F(u)}{1-F(t)}\right\}\mathrm{d}M_i(t)\\ &\quad = \delta_i s(e_i(\beta_0)) - \frac{\delta_i \int_{e_i(\beta_0)}^{\infty} s(u)\,\mathrm{d}F(u)}{1-F(e_i(\beta_0))} - \int_{-\infty}^{e_i(\beta_0)}\left\{s(t) - \frac{\int_t^{\infty} s(u)\,\mathrm{d}F(u)}{1-F(t)}\right\}\lambda(t)\,\mathrm{d}t.\end{aligned}$$

By integration by part,

$$\begin{aligned}\int_{-\infty}^{e_i(\beta_0)}\left\{s(t) - \frac{\int_t^{\infty} s(u)\,\mathrm{d}F(u)}{1-F(t)}\right\}\lambda(t)\,\mathrm{d}t &= -\frac{\int_{e_i(\beta_0)}^{\infty} s(u)\,\mathrm{d}F(u)}{1-F(e_i(\beta_0))} + \int_{-\infty}^{\infty} s(u)\,\mathrm{d}F(u)\\ &= -\frac{\int_{e_i(\beta_0)}^{\infty} s(u)\,\mathrm{d}F(u)}{1-F(e_i(\beta_0))} + Es(\varepsilon_i).\end{aligned}$$

Consequently, (A.1) holds. □

Proof of Theorem 1. It was shown in Jin et al. (2003) that $\hat\beta_{(0)} = \hat\beta_{\mathrm{G}}$ is consistent and $\sqrt{n}(\hat\beta_{\mathrm{G}} - \beta_0)$ is asymptotically normally distributed. We first prove the theorem when $m=1$, then prove when $m>1$ by the induction method.

(i) *Consistency of* $\hat\beta_{(1)}$. Observe that

$$\left.\frac{\partial \mathrm{U}(\beta;\hat\beta_{(0)},s)}{\partial\beta}\right|_{\beta=\hat\beta_{(0)}} = -\sum_{i=1}^{n}(X_i-\bar X)X_i^{\mathrm{T}}\dot s(e_i(\hat\beta_{(0)})).$$

By assumptions (C1)–(C4) and Slutsky's Theorem,

$$-\frac{1}{n}\left.\frac{\partial \mathrm{U}(\beta;\hat\beta_{(0)},s)}{\partial\beta}\right|_{\beta=\hat\beta_{(0)}} \to D_s \tag{A.2}$$

as $n\to\infty$.

Write $\mathrm{U}(b) = \mathrm{U}(b;b,s)$. With the arguments similar to those in Lai and Ying (1994) we can show that, uniformly in $\|b-\beta_0\| \leqslant n^{-1/3}$,

$$\mathrm{U}(b) = \mathrm{U}(\beta_0) - nA_s(b-\beta_0) + \mathrm{o}(n^{1/2} + n\|b-\beta_0\|) \tag{A.3}$$

almost surely. In particular, since $\sqrt{n}(\hat{\beta}_{(0)}-\beta_0)=\mathrm{O_P}(1)$,

$$\mathrm{U}(\hat{\beta}_{(0)})=\mathrm{U}(\beta_0)-nA_s(\hat{\beta}_{(0)}-\beta_0)+\mathrm{o_P}(n^{1/2}).$$

Therefore, $(1/n)\mathrm{U}(\hat{\beta}_{(0)})=\mathrm{o_P}(1)$. This together with (A.2) and expression (4), the consistency of $\hat{\beta}_{(1)}$ follows.

(ii) *Convergence of* $\sqrt{n}(\hat{\beta}_{(1)}-\beta_0)$. By the expression (4) and (A.2)–(A.3),

$$\begin{aligned}\sqrt{n}(\hat{\beta}_{(1)}-\beta_0)&=\sqrt{n}(\hat{\beta}_{(0)}-\beta_0)+D_s^{-1}\left[\frac{1}{\sqrt{n}}\mathrm{U}(\beta_{(0)})-A_s\sqrt{n}(\hat{\beta}_{(0)}-\beta_0)\right]+\mathrm{o_P}(1)\\&=(I-D_s^{-1}A_s)\sqrt{n}(\widehat{\beta}_{\mathrm{G}}-\beta_0)+D_s^{-1}\frac{1}{\sqrt{n}}\mathrm{U}(\beta_0)+\mathrm{o_P}(1).\end{aligned}\tag{A.4}$$

Note that

$$\begin{aligned}\mathrm{U}(\beta_0)&=\sum_{i=1}^{n}(X_i-\overline{X})\left\{\delta_i s(e_i(\beta_0))+(1-\delta_i)\frac{\int_{e_i(\beta_0)}^{\infty}s(u)\,\mathrm{d}\widehat{F}_{\beta_0}(u)}{1-\widehat{F}_{\beta_0}(e_i(\beta_0))}\right\}\\&=\sum_{i=1}^{n}(X_i-\overline{X})\left\{\delta_i s(e_i(\beta_0))+(1-\delta_i)\frac{\int_{e_i(\beta_0)}^{\infty}s(u)\,\mathrm{d}F(u)}{1-F(e_i(\beta_0))}\right\}\\&\quad+\sum_{i=1}^{n}(X_i-\overline{X})(1-\delta_i)\left\{\frac{\int_{e_i(\beta_0)}^{\infty}s(u)\,\mathrm{d}\widehat{F}_{\beta_0}(u)}{1-\widehat{F}_{\beta_0}(e_i(\beta_0))}-\frac{\int_{e_i(\beta_0)}^{\infty}s(u)\,\mathrm{d}F(u)}{1-F(e_i(\beta_0))}\right\}.\end{aligned}\tag{A.5}$$

It follows from the martingale representation of the Kaplan–Meier estimator (Fleming and Harrington, 1991, p. 98) that, almost surely,

$$\frac{\int_{e_i(\beta_0)}^{\infty}s(u)\,\mathrm{d}\widehat{F}_{\beta_0}(u)}{1-\widehat{F}_{\beta_0}(e_i(\beta_0))}-\frac{\int_{e_i(\beta_0)}^{\infty}s(u)\,\mathrm{d}F(u)}{1-F(e_i(\beta_0))}=n^{-1}\sum_{j=1}^{n}\int_{-\infty}^{\infty}\phi_i(t)\,\mathrm{d}M_j(t)+\mathrm{o}(n^{-1/2})\tag{A.6}$$

for some random process $\phi_i(t)$. By (A.1) in Lemma 1 and (A.5)–(A.6),

$$\mathrm{U}(\beta_0)=\sum_{i=1}^{n}\int_{-\infty}^{\infty}\left[(X_i-\overline{X})\left\{s(t)-\frac{\int_t^{\infty}s(u)\,\mathrm{d}F(u)}{1-F(t)}\right\}+\overline{\phi}^x(t)\right]\mathrm{d}M_i(t)+\mathrm{o}(n^{1/2}),\tag{A.7}$$

where $\overline{\phi}^x(t)$ is the limit of $n^{-1}\sum_{i=1}^{n}(X_i-\overline{X})(1-\delta_i)\phi_i(t)$.

As shown by Jin et al. (2003, 2006b),

$$\sqrt{n}(\widehat{\beta}_{\mathrm{G}}-\beta_0)=A_{\mathrm{G}}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\eta_i(t)\,\mathrm{d}M_i(t)+\mathrm{o_P}(1),\tag{A.8}$$

where the $\eta_i(t)$'s are nonrandom functions.

Consequently, by (A.4) and (A.7)–(A.8),

$$\sqrt{n}(\widehat{\beta}_{(1)}-\beta_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_{i1}+\mathrm{o_P}(1),\tag{A.9}$$

where

$$H_{i1}=\int_{-\infty}^{\infty}\left((I-D_s^{-1}A_s)A_{\mathrm{G}}^{-1}\eta_i(t)+D_s^{-1}\left[(X_i-\overline{X})\left\{s(t)-\frac{\int_t^{\infty}s(u)\,\mathrm{d}F(u)}{1-F(t)}\right\}+\overline{\phi}^x(t)\right]\right)\mathrm{d}M_i(t).$$

The convergence of $\sqrt{n}(\widehat{\beta}_{(1)}-\beta_0)$ follows from the multivariate central limit theorem. Its asymptotic variance is the limit of $(1/n)\sum_{i=1}^{n}H_{i1}^{\otimes 2}$ as $n\to\infty$.

(iii) *Consistency of* $\hat{\beta}_{(m)}$ *and the convergence of* $\sqrt{n}(\hat{\beta}_{(m)}-\beta_0)$ *for* $m>1$. The proof can be obtained by the induction method with the use of the similar arguments in (i) and (ii) repeatedly. In fact, it can be shown that

$$\sqrt{n}(\widehat{\beta}_{(m)}-\beta_0)=(I-D_s^{-1}A_s)\sqrt{n}(\widehat{\beta}_{(m-1)}-\beta_0)+D_s^{-1}\frac{1}{\sqrt{n}}\mathrm{U}(\beta_0)+\mathrm{o_P}(1). \tag{A.10}$$

Consequently, by the induction method, the consistency of $\hat{\beta}_{(m)}$ follows and

$$\sqrt{n}(\widehat{\beta}_{(m)}-\beta_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_{im}+\mathrm{o_P}(1), \tag{A.11}$$

where

$$\begin{aligned}H_{im}=\int_{-\infty}^{\infty}\Bigg(&(I-D_s^{-1}A_s)^m A_{\mathrm{G}}^{-1}\eta_i(t)\\&+\left\{I-(I-D_s^{-1}A_s)^m\right\}A_s^{-1}\left[(X_i-\overline{X})\left\{s(t)-\frac{\int_t^{\infty}s(u)\,\mathrm{d}F(u)}{1-F(t)}\right\}+\overline{\phi}^x(t)\right]\Bigg)\,\mathrm{d}M_i(t)\end{aligned}$$

for each fixed $m>1$. Thus, $\sqrt{n}(\widehat{\beta}_{(m)}-\beta_0)$ is asymptotically normal as $n\to\infty$ with the variance as the limit of $(1/n)\sum_{i=1}^{n}H_{im}^{\otimes 2}$ as $n\to\infty$. This completes the proof of Theorem 1. □

Proof of Theorem 2. The consistency of $\hat{\beta}^*_{(0)}=\hat{\beta}^*_{\mathrm{G}}$ and the convergence of $\sqrt{n}(\hat{\beta}^*_{(0)}-\hat{\beta}_{(0)})$ were shown in Jin et al. (2006b). We first prove the theorem when $m=1$, then prove when $m>1$ by the induction method.

(i) *Consistency of* $\hat{\beta}^*_{(1)}$ *and convergence of the conditional distribution of* $\sqrt{n}(\hat{\beta}^*_{(1)}-\hat{\beta}_{(1)})$ *given* $\mathscr{F}_n$. Notice that

$$\frac{\partial \mathrm{U}^*(\beta;\hat{\beta}^*_{(m-1)},s)}{\partial\beta}=-\sum_{i=1}^{n}Z_i(X_i-\bar{X})X_i^{\mathrm{T}}\dot{s}(e_i(\beta)).$$

Since $EZ_i=1$, by the strong law of large number, as $n\to\infty$,

$$\frac{1}{n}\sum_{i=1}^{n}(Z_i-1)(X_i-\bar{X})X_i^{\mathrm{T}}\dot{s}(e_i(\beta))\to 0.$$

If $\hat{\beta}^*_{(m-1)}$ is consistent, by Slutsky's theorem, as $n\to\infty$,

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})X_i^{\mathrm{T}}\dot{s}(e_i(\hat{\beta}^*_{(m-1)}))\to D_s,$$

which implies that, as $n\to\infty$,

$$\frac{1}{n}\sum_{i=1}^{n}Z_i(X_i-\bar{X})X_i^{\mathrm{T}}\dot{s}(e_i(\hat{\beta}^*_{(m-1)}))\to D_s. \tag{A.12}$$

Since $\hat{\beta}^*_{(0)}$ is consistent, (A.12) holds for $\hat{\beta}^*_{(0)}$. Thus, as $n\to\infty$,

$$-\frac{1}{n}\frac{\partial \mathrm{U}^*(\beta;\hat{\beta}^*_{(0)},s)}{\partial\beta}\Bigg|_{\beta=\hat{\beta}^*_{(0)}}\to D_s.$$

Write $\mathrm{U}^*(b)=\mathrm{U}^*(b;b,s)$. By incorporating the random weights Z_i into the derivation of the asymptotic linearity in Lai and Ying (1994), we can establish the following linear approximation analogous to (A.3):

$$\mathrm{U}^*(\widehat{\beta}^*_{(0)})=\mathrm{U}^*(\widehat{\beta}_{(0)})-nA_s(\widehat{\beta}^*_{(0)}-\widehat{\beta}_{(0)})+\mathrm{o}(n^{1/2}+n\|\widehat{\beta}^*_{(0)}-\beta_0\|+n\|\widehat{\beta}_{(0)}-\beta_0\|) \tag{A.13}$$

almost surely. The slope matrix A_s remains the same as in (A.2) since $EZ_i = 1$ and A_s is the limit of the gradient of $n^{-1}E\mathrm{U}^*(b)$ at $b = \beta_0$. Observe that

$$\mathrm{U}^*(\widehat{\beta}_{(0)}) - \mathrm{U}(\widehat{\beta}_{(0)}) = I + II, \tag{A.14}$$

where

$$I = \sum_{i=1}^{n}(Z_i - 1)(X_i - \overline{X})[s(e_i(\widehat{\beta}_{(0)})) + (1-\delta_i)\xi_{is}(\hat{\beta}_{(0)}) - \bar{\xi}_s(\hat{\beta}_{(0)})],$$

$$II = \sum_{i=1}^{n}(X_i - \overline{X})(1-\delta_i)\left[\xi_{is}(\hat{\beta}_{(0)}) - \frac{\int_{e_i(\widehat{\beta}_{(0)})}^{\infty} s(u)\,\mathrm{d}\widehat{F}_{\widehat{\beta}_{(0)}}(u)}{1 - \widehat{F}_{\widehat{\beta}_{(0)}}(e_i(\widehat{\beta}_{(0)}))}\right].$$

Notice that

$$s(e_i(\widehat{\beta}_{(0)})) + (1-\delta_i)\xi_{is}(\widehat{\beta}_{(0)}) = \delta_i s(e_i(\widehat{\beta}_{(0)})) + (1-\delta_i)\left\{\frac{\int_{e_i(\widehat{\beta}_{(0)})}^{\infty} s(u)\,\mathrm{d}\widehat{F}^*_{\widehat{\beta}_{(0)}}(u)}{1 - \widehat{F}^*_{\widehat{\beta}_{(0)}}(e_i(\widehat{\beta}_{(0)}))}\right\}.$$

With expansion at β_0, it can be shown that

$$\begin{aligned} s(e_i(\widehat{\beta}_{(0)})) + (1-\delta_i)\xi_{is}(\hat{\beta}_{(0)}) - \bar{\xi}_s(\hat{\beta}_{(0)}) &= \delta_i s(e_i(\beta_0)) + (1-\delta_i)\left\{\frac{\int_{e_i(\beta_0)}^{\infty} s(u)\,\mathrm{d}F(u)}{1 - F(e_i(\beta_0))}\right\} \\ &\quad - Es(\varepsilon_i) + o_{\mathrm{P}}(n^{-1/2}). \end{aligned}$$

Since $E(Z_i - 1|\mathscr{F}) = 0$, we have

$$I = \sum_{i=1}^{n}(Z_i - 1)\int_{-\infty}^{\infty}(X_i - \overline{X})\left\{s(t) - \frac{\int_t^{\infty} s(u)\,\mathrm{d}F(u)}{1 - F(t)}\right\}\mathrm{d}M_i(t) + o_{\mathrm{P}}(n^{1/2}). \tag{A.15}$$

As in Jin et al. (2006a), by approximating $\widehat{F}^*_{\widehat{\beta}_{(0)}} - \widehat{F}_{\widehat{\beta}_{(0)}}$ with a weighted sum of $Z_i - 1$, we can show that

$$\begin{aligned} II &= \sum_{i=1}^{n}(X_i - \overline{X})(1-\delta_i)\left\{\frac{\int_{e_i(\widehat{\beta}_{(0)})}^{\infty} s(u)\,\mathrm{d}\widehat{F}^*_{\widehat{\beta}_{(0)}}(u)}{1 - \widehat{F}^*_{\widehat{\beta}_{(0)}}(e_i(\widehat{\beta}_{(0)}))} - \frac{\int_{e_i(\widehat{\beta}_{(0)})}^{\infty} s(u)\,\mathrm{d}\widehat{F}_{\widehat{\beta}_{(0)}}(u)}{1 - \widehat{F}_{\widehat{\beta}_{(0)}}(e_i(\widehat{\beta}_{(0)}))}\right\} \\ &= \sum_{i=1}^{n}(Z_i - 1)\int_{-\infty}^{\infty}\overline{\phi}^x(t)\,\mathrm{d}M_i(t) + o_{\mathrm{P}}(n^{1/2}). \end{aligned} \tag{A.16}$$

By (A.13)–(A.16), we obtain

$$\begin{aligned} \mathrm{U}^*(\widehat{\beta}^*_{(0)}) &= \mathrm{U}(\widehat{\beta}_{(0)}) + \sum_{i=1}^{n}(Z_i - 1)\int_{-\infty}^{\infty}\left[(X_i - \overline{X})\left\{s(t) - \frac{\int_t^{\infty} s(u)\,\mathrm{d}F(u)}{1 - F(t)}\right\} + \overline{\phi}^x(t)\right]\mathrm{d}M_i(t) \\ &\quad - nA_s(\widehat{\beta}^*_{(0)} - \widehat{\beta}_{(0)}) + o(n^{1/2} + n\|\widehat{\beta}^*_{(0)} - \beta_0\| + n\|\widehat{\beta}_{(0)} - \beta_0\|). \end{aligned} \tag{A.17}$$

Because of (A.12) and (A.17), the difference between the expressions (9) and (6) for $m = 1$ leads to

$$\begin{aligned} \sqrt{n}(\widehat{\beta}^*_{(1)} - \widehat{\beta}_{(1)}) &= \frac{1}{\sqrt{n}}D_s^{-1}\sum_{i=1}^{n}(Z_i - 1)\int_{-\infty}^{\infty}\left[(X_i - \overline{X})\left\{s(t) - \frac{\int_t^{\infty} s(u)\,\mathrm{d}F(u)}{1 - F(t)}\right\} + \overline{\phi}^x(t)\right]\mathrm{d}M_i(t) \\ &\quad + (I - D_s^{-1}A_s)\sqrt{n}(\widehat{\beta}^*_{(0)} - \widehat{\beta}_{(0)}) + o(1 + \sqrt{n}\|\widehat{\beta}^*_{(0)} - \beta_0\| + \sqrt{n}\|\widehat{\beta}_{(0)} - \beta_0\|). \end{aligned}$$

It was shown by Jin et al. (2003, 2006b) that

$$\sqrt{n}(\widehat{\beta}^*_{(0)} - \widehat{\beta}_{(0)}) = \frac{1}{\sqrt{n}}A_{\mathrm{G}}^{-1}\sum_{i=1}^{n}(Z_i - 1)\int_{-\infty}^{\infty}\eta_i(t)\,\mathrm{d}M_i(t) + o_{\mathrm{P}}(1). \tag{A.18}$$

Since $\sqrt{n}(\widehat{\beta}^*_{(0)} - \beta_0) = \mathrm{O_P}(1)$ and $\sqrt{n}(\widehat{\beta}_{(0)} - \beta_0) = \mathrm{O_P}(1)$, we have

$$\sqrt{n}(\widehat{\beta}^*_{(1)} - \widehat{\beta}_{(1)}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_i - 1) H_{i1} + \mathrm{o_P}(1), \tag{A.19}$$

where H_{i1} is the same as that in (A.9). Note that conditional on $\mathscr{F}_n$, we have that $(Z_i - 1)H_{i1}$ are independent with mean 0 and covariance $H_{i1}^{\otimes 2}$. Consequently, $\hat{\beta}^*_{(1)}$ is consistent and the conditional distribution of $\sqrt{n}(\widehat{\beta}^*_{(1)} - \widehat{\beta}_{(1)})$ given $\mathscr{F}_n$ converges to the asymptotic distribution of $\sqrt{n}(\widehat{\beta}_{(1)} - \beta_0)$ in probability.

(ii) *Consistency of $\hat{\beta}^*_{(m)}$ and convergence of the conditional distribution of $\sqrt{n}(\hat{\beta}^*_{(m)} - \hat{\beta}_{(m)})$ given $\mathscr{F}_n$ for $m > 1$.* By replacing $\hat{\beta}^*_{(0)}$ and $\hat{\beta}_{(0)}$ with $\hat{\beta}^*_{(m-1)}$ and $\hat{\beta}_{(m-1)}$, respectively, in the part (1) and with similar arguments therein, we can show that

$$\begin{aligned}\widehat{\beta}^*_{(m)} - \widehat{\beta}_{(m)} = {} & n^{-1} D_s^{-1} \sum_{i=1}^{n} (Z_i - 1) \int_{-\infty}^{\infty} \left[(X_i - \overline{X}) \left\{ s(t) - \frac{\int_t^{\infty} s(u)\,\mathrm{d}F(u)}{1 - F(t)} \right\} + \overline{\phi}^x(t) \right] \mathrm{d}M_i(t) \\ & + (I - D_s^{-1} A_s)(\widehat{\beta}^*_{(m-1)} - \widehat{\beta}_{(m-1)}) + \mathrm{o_P}(n^{-1/2}).\end{aligned}$$

By induction,

$$\begin{aligned}\widehat{\beta}^*_{(m)} - \widehat{\beta}_{(m)} = {} & (I - D_s^{-1} A_s)^m (\widehat{\beta}^*_{(0)} - \widehat{\beta}_{(0)}) + n^{-1} \{ I - (I - D_s^{-1} A_s)^m \} A_s^{-1} \\ & \times \sum_{i=1}^{n} (Z_i - 1) \int_{-\infty}^{\infty} \left[(X_i - \overline{X}) \left\{ s(t) - \frac{\int_t^{\infty} s(u)\,\mathrm{d}F(u)}{1 - F(t)} \right\} + \overline{\phi}^x(t) \right] \mathrm{d}M_i(t) + \mathrm{o_P}(n^{-1/2}).\end{aligned}$$

Combined with (A.18), we have

$$\sqrt{n}(\widehat{\beta}^*_{(m)} - \widehat{\beta}_{(m)}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_i - 1) H_{im} + \mathrm{o_P}(1). \tag{A.20}$$

Consequently, $\widehat{\beta}^*_{(m)}$ is consistent and the conditional distribution of $\sqrt{n}(\widehat{\beta}^*_{(m)} - \widehat{\beta}_{(m)})$ given $\mathscr{F}_n$ converges to the limiting distribution of $\sqrt{n}(\widehat{\beta}_{(m)} - \beta_0)$ in probability. This completes the proof of Theorem 2. $\square$

References

Buckley, J., James, I., 1979. Linear regression with censored data. Biometrika 66, 429–436.

Fleming, T.R., Harrington, D.P., 1991. Counting Processes and Survival Analysis. Wiley, New York.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.J., 1986. Robust Statistics, the Approach Based on Influence Functions. Wiley, New York.

Huber, P.J., 1981. Robust Statistics. Wiley, New York.

Jin, Z., Lin, D.Y., Wei, L.J., Ying, Z., 2003. Rank-based inference for the accelerated failure time model. Biometrika 90, 341–353.

Jin, Z., Lin, D.Y., Ying, Z., 2006a. On least-squares regression with censored data. Biometrika 93, 147–161.

Jin, Z., Lin, D.Y., Ying, Z., 2006b. Rank regression analysis of multivariate failure time data based on marginal linear models. Scand. J. Statist. 33, 1–23.

Koul, H., Susarla, V., Van Ryzin, J., 1981. Regression analysis with randomly right-censored data. Ann. Statist. 9, 1276–1288.

Lai, T.L., Ying, Z., 1994. A missing information principle and *M*-estimators in regression analysis with censored and truncated data. Ann. Statist. 21, 1222–1255.

Ritov, Y., 1990. Estimation in a linear regression model with censored data. Ann. Statist. 18, 303–328.

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