To appear in *Journal of Statistical Computation and Simulation*, (2003).

The Semi-Parametric MLE In Linear Regression With Right-Censored Data

By Qiqing Yu [1] and George Y. C. Wong [1]

Department of Mathematical Sciences, SUNY, Binghamton, NY 13902, USA
email address: qyu@math.binghamton.edu
and
Strang Cancer Prevention Center, 428 E 72nd Street, NY 10021, USA
email address: gwong@strang.org

AMS 1991 subject classification: Primary 62 J05; Secondary 62 G05.

Short Title: SMLE

Key Words: Least squares estimator, Buckley-James estimator, product-limit-estimator, non-iterative algorithm.

Abstract: Consider the linear regression model, $Y_i = \beta'\mathbf{X}_i + \epsilon_i$, $i = 1, ..., n$, where Y_i maybe right censored and ϵ_i has an unknown cdf F_o. The semi-parametric maximum likelihood estimator (SMLE) of β, denoted by $\tilde{\beta}_n$, has not been studied, and there is no known method in the literature for finding $\tilde{\beta}_n$. $\tilde{\beta}_n$ cannot be obtained by standard numerical methods, including the Newton-Raphson method and the Monte Carlo method. We propose a feasible non-iterative algorithm for $\tilde{\beta}_n$. Simulation results suggest that $\tilde{\beta}_n$ is consistent if F_o is continuous and $\tilde{\beta}_n$ is efficient if F_o has discontinuity points. A consistency and efficiency result is established under certain regularity conditions. We apply the SMLE to three data sets of sizes 10, 69 and 374, respectively. The results indicate that the SMLE may have advantages over the Buckley-James estimator in applications.

[1] Partially supported by DOD Grant DAMD17-99-1-9390 and DAMD17-00-1-0448.

1. Introduction. We study the semi-parametric maximum likelihood estimator (SMLE) in the linear regression model with right-censored data. Regression analysis is one of the most widely used statistical techniques. Its applications occur in almost every field, including economics, engineering, the physical sciences, management, life and biological sciences and the social sciences. We assume: $Y = \beta'\mathbf{X} + \epsilon$, where $\mathbf{X}$ is a $p \times 1$ covariate vector, β' the transpose of a regression coefficient vector β, and ϵ a random variable with an unknown cdf F_o ($F_o(t) = P(\epsilon \leq t)$). There is no additional assumption on F_o.

This is a semi-parametric set-up with β and F_o unknown. There is a rich literature on this semi-parametric problem and there are several estimators available, such as, the least squares estimator (LSE) without censoring, and its various modifications under right censoring, *e.g.*, Miller's (1981) estimator, the Buckley-James (1979) estimator (BJE), and the modifications proposed by Chatterjee and McLeish (1986) and Leurgans (1987); Theil's estimator (Theil (1950)) and the Theil-Sen estimator (Sen, 1968), and their extensions to the right-censored data (see, *e.g.*, Akritas *et al.* (1995)); various M-estimators (Huber (1964), Ritov (1990) and Hillis (1991)); Bayes estimators (Koul, Susarla and Van Ryzin (1981)); L-estimators and R-estimators (see, e.g., Tsiatis (1990) and Wei, *et al.* (1990)). Furthermore, there are many other estimators, see for example, Manski (1985), Powell (1986), Duncan (1986), Fernandez (1986), Horowitz (1986), Ruud (1986), Ichimura (1987) and Moon (1989), among others.

Cosslett (1987) derives the efficient lower bound for the estimators under the semi-parametric set-ups with smoothness conditions. One may suspect that the SMLE would attain this efficient lower bound. However, in the literature, the SMLE has not been studied and there is no known way of computing the SMLE. More recently, Yu and Wong (2003) show that the SMLE cannot be obtained by standard numerical methods such as the Newton-Raphson algorithm, the Monte Carlo method and the M-estimation approach, because the derivative of the log likelihood equals to zero almost everywhere. When data are not censored, Yu and Wong propose an algorithm for obtaining the SMLE and establish some of its asymptotic properties.

In this paper, we extend Yu and Wong's result to right-censored data. In particular, we propose a non-iterative algorithm for obtaining the SMLE $\tilde{\beta}_n$ with right-censored data and study its properties. Simulation results reported in this paper suggest that

(1) if F_o is continuous then the SMLE is consistent, and

(2) if F_o has a discontinuous point then the SMLE is super efficient, in the sense that

$$\mathrm{P}\{\tilde{\beta}_n \neq \beta \text{ for infinitely many } n\} = 0 \ (i.e.\ \mathrm{P}\{\tilde{\beta}_n \neq \beta \text{ i.o.}) = 0), \tag{1.1}$$

$$\sqrt{n}(\tilde{\beta}_n - \beta) \xrightarrow{D} N(0,0) \text{ and } \lim_{n\to\infty} n\sigma^2_{\tilde{\beta}_n} = 0, \text{ where } n \text{ is the sample size.} \tag{1.2}$$

The paper is organized as follows. In Section 2, we define the SMLE, denoted by $\tilde{\beta}_n$ or $\tilde{\beta}$ and propose a non-iterative algorithm to obtain all SMLE's. In Section 3, we establish a strong consistency result and a super efficiency result under certain regularity conditions. In Section 4, we apply our method to three real data sets. The SMLE gives estimates of β with (significantly) correct signs in the three examples. However, the BJE gives a value of β with a significantly wrong sign in a cancer research data set, gives several values of β with both correct and incorrect signs in the Stanford heart transplant data, and gives a value of β with significantly correct sign, but the BJE does not appear as good as the SMLE. It suggests that the SMLE may have advantage over the BJE or other estimators in applications. In Section 5, we study the asymptotic properties of the SMLE by means of simulation. In Appendix, we present details of some proofs in Section 3.

2. The SMLE. We shall introduce the SMLE and present its algorithms in this section.

2.1. Existence of the SMLE. Under the right censorship model, the observable random vector is $(M, \delta, \mathbf{X})$, where $M = \min(Y, C)$, C is a right censoring random variable, $\delta = \mathbf{1}_{(Y \leq C)}$ and $\mathbf{1}_A$ is the indicator function of an event A. Denote $S_o = 1 - F_o$. Let $(M_i, \delta_i, \mathbf{X}_i)$ be i.i.d. copies of $(M, \delta, \mathbf{X})$. Denote $T(\mathbf{b}) = M - \mathbf{b}'\mathbf{x}$ and $T_i = T_i(\mathbf{b}) = M_i - \mathbf{b}'\mathbf{X}_i$. Since $(T_i(\beta), \delta_i)$, $i = 1, ..., n$, are i.i.d. copies of $T(\beta)$, the non-parametric likelihood function is

$$Ł(S, \mathbf{b}) = \prod_{i=1}^{n} [(f(T_i(\mathbf{b})))^{\delta_i} (S(T_i(\mathbf{b})))^{1-\delta_i}], \ \mathbf{b} \in \mathcal{R}^p, \ S \in \mathcal{F}, \ f(x) = S(x-) - S(x), \tag{2.1}$$

where $\mathcal{F} = \{H: \ H$ is a non-increasing function on $[-\infty, \infty]$, $H(-\infty) = 1, H(\infty) = 0\}$, and $S(x-) = \lim_{s\uparrow x} S(s)$. The SMLE of (S_o, β) is a value of $(S, \mathbf{b})$ that maximizes $\mathrm{L}(S, \mathbf{b})$.

Remark 2.1. Given $\mathbf{b}$, the non-parametric MLE (NPMLE) of a survival function maximizes $\mathrm{L}(S, \mathbf{b})$ over $S \in \mathcal{F}$. It is well known that the NPMLE is the product-limit-estimator (PLE) $\hat{S}_{\mathbf{b}}$ based on $(T_i(\mathbf{b}), \delta_i)$, $i = 1, ..., n$ (see, *e.g.*, Miller (1981)). Denote $\hat{f}_{\mathbf{b}}(t) = \hat{S}_{\mathbf{b}}(t-) - \hat{S}_{\mathbf{b}}(t)$ and $\hat{F}_{\mathbf{b}} = 1 - \hat{S}_{\mathbf{b}}$. Let $\mathcal{B}$ be the set of all possible values of $\mathbf{b}$ that maximize $\mathrm{L}(\hat{S}_{\mathbf{b}}, \mathbf{b})$ over all $\mathbf{b} \in \mathcal{R}^p$. The SMLE of (S_o, β) is then $(\hat{S}_{\mathbf{b}}, \mathbf{b})$, where $\mathbf{b} \in \mathcal{B}$. Under our assumptions, $\alpha = E(\epsilon)$ may not exist. If α exists, an estimator of α based on the least squares principle is $\hat{\alpha} = a(\tilde{\beta})$, where $\tilde{\beta} \in \mathcal{B}$ and $a(\mathbf{b}) = \frac{\int_{x<\infty} x d\hat{F}_{\mathbf{b}}(x)}{\int_{x<\infty} 1 d\hat{F}_{\mathbf{b}}(x)}$, where $\int_{x<\infty} 1 d\hat{F}_{\mathbf{b}}(x)$ may not equal 1. Denote $\mathcal{L}(\mathbf{b}) = \mathrm{L}(\hat{S}_{\mathbf{b}}, \mathbf{b})$.

Theorem 2.1. *$\mathcal{L}(\mathbf{b})$, $\mathbf{b} \in \mathcal{R}^p$, only has finitely many values. Thus the SMLE of β exists.*

Proof. Let $T_{(1)}(\mathbf{b}) \leq \cdots \leq T_{(n)}(\mathbf{b})$ be the order statistics of $T_i(\mathbf{b})$'s and $\delta_{(i)}(\mathbf{b})$ denote the δ_j associated with $T_{(i)}(\mathbf{b})$. Denote $r(T_i, \delta_i)$ the rank of (T_i, δ_i) among $(T_1, \delta_1), ..., (T_n, \delta_n)$, which is defined as follows:

1. If there is no tie among T_i's, $r(T_i, \delta_i) = j$, where $T_{(j)} = T_i$.
2. If $T_{(k)} < T_i = T_j < T_{(k+3)}$ and $\delta_i = 0$ and $\delta_j = 1$, then $r(T_j, \delta_j) = k + 1$ and $r(T_i, \delta_i) = k + 2$, as the right-censored observation occurs after T_j.
3. If there are exactly k ties at (T_i, δ_i), namely, $(T_{h_j}, \delta_{h_j}) = (T_i, \delta_i)$, for k different h_j's, $T_{(m)} \leq T_i \leq T_{(m+k)}$, $(T_{(m)}, \delta_{(m)}) \neq (T_i, \delta_i)$ and $(T_{(m+k)}, \delta_{(m+k)}) \neq (T_i, \delta_i)$ then $r(T_{h_j}, \delta_{h_j}) = m + \frac{k+1}{2}$ for all k h_j's.

As argued in Remark 2.1, in order to find the SMLE, it suffices to maximize

$$\mathcal{L}(\mathbf{b}) = \mathrm{L}(\hat{S}_{\mathbf{b}}, \mathbf{b}) = \prod_{i=1}^{n} [\hat{S}_{\mathbf{b}}(T_{(i)}(\mathbf{b})-) - \hat{S}_{\mathbf{b}}(T_{(i)}(\mathbf{b}))]^{\delta_{(i)}(\mathbf{b})} (\hat{S}_{\mathbf{b}}(T_{(i)}(\mathbf{b})))^{1-\delta_{(i)}(\mathbf{b})}. \tag{2.2}$$

Note that $\hat{S}_{\mathbf{b}}(T_{(j)}(\mathbf{b})) = \prod_{T(i) \leq T_{(j)}} (1 - \frac{\delta_{(i)}(\mathbf{b})}{n-i+1})$. Given a sample of size n, there are at most finitely many distinct values of $\hat{S}_{\mathbf{b}}(T_{(j)}(\mathbf{b}))$, $\mathbf{b} \in \mathcal{R}^p$, as there are at most 2^n distinct values of the vector $(\delta_{(1)}(\mathbf{b}), ..., \delta_{(n)}(\mathbf{b}))$. In view of (2.2),

$$\mathcal{L}(\mathbf{b}_1) = \mathcal{L}(\mathbf{b}) \text{ if } (\delta_{(1)}(\mathbf{b}_1), ..., \delta_{(n)}(\mathbf{b}_1)) = (\delta_{(1)}(\mathbf{b}), ..., \delta_{(n)}(\mathbf{b})) \ . \tag{2.3}$$

Thus, $\mathcal{L}(\mathbf{b})$ has finitely many values. Consequently, the SMLE of β exists. □

2.2 Non-iterative algorithms for the SMLE. The proof of the Theorem 2.1 actually suggests an algorithm to find all solutions of the SMLE of β. Let $\mathcal{V}$ be the collection of all possible values of $(\delta_{(1)}(\mathbf{b}), ..., \delta_{(n)}(\mathbf{b}))$, then $||\mathcal{V}|| \leq 2^n$. For each $\mathbf{v} \in \mathcal{V}$, denote $\mathcal{H}_{\mathbf{v}} = \{\mathbf{b}: \ \mathbf{b} \in \mathcal{R}^p, (\delta_{(1)}(\mathbf{b}), ..., \delta_{(n)}(\mathbf{b})) = \mathbf{v}\}$, then $\{\mathcal{H}_{\mathbf{v}}, \mathbf{v} \in \mathcal{V}\}$ is a partition of $\mathcal{R}^p$.

Algorithm 1 (for obtaining all possible SMLE's with $p \geq 1$):

1. Obtain the set $\mathcal{V}$.
2. For each $\mathbf{v} \in \mathcal{V}$, select an element $\mathbf{b}_{\mathbf{v}} \in \mathcal{H}_{\mathbf{v}}$, then compute $\hat{S}_{\mathbf{b}_{\mathbf{v}}}$ and $\mathcal{L}(\mathbf{b}_{\mathbf{v}})$.
3. Find the maximizers of $\mathcal{L}(\mathbf{b}_{\mathbf{v}})$, $\mathbf{v} \in \mathcal{V}$. Given each maximizer $\mathbf{b}_{\mathbf{v}_o}$, each element in $\mathcal{H}_{\mathbf{v}_o}$ is an SMLE of β.

The algorithm is non-iterative, as $\mathcal{V}$ is finite and the PLE has a closed-form expression. One may want more details on how to choose $\mathbf{b}_{\mathbf{v}} \in \mathcal{H}_{\mathbf{b}}$. For simplicity, we first explain for the case $p = 1$ as follows.

Algorithm 2 (for obtaining all possible solutions of the SMLE with $p = 1$):

1. Compute $b = \frac{M_i - M_j}{X_i - X_j}$, where $X_i \neq X_j$ and $i < j$. Order the resulting distinct slopes b's as $a_1 < \cdots < a_{m_a}$ and denote $\mathcal{A}_1 = \{a_1, ..., a_{m_a}\}$.
2. Let $\mathcal{A}_2$ be the set consisting of $a_1 - 1$, $a_{m_a} + 1$ and points $\frac{a_{i-1} + a_i}{2}$, $i = 2, ..., m_a$.
3. For each $b \in \mathcal{A}_1 \cup \mathcal{A}_2$, compute $\hat{S}_b$ and $\mathcal{L}(b)$.
4. Each maximum point b of $\mathcal{L}(b)$ over $b \in \mathcal{A}_1 \cup \mathcal{A}_2$ is an SMLE of β. Moreover, if an SMLE $\tilde{\beta} \in (a_{i-1}, a_i)$, then each point $b \in (a_{i-1}, a_i)$ is also an SMLE.

The above algorithm can be justified as follows. In view of (2.3), $\mathcal{L}(b) = \mathcal{L}(b_1)$ if $r(T_i(\mathbf{b}), \delta_i) = r(T_i(\mathbf{b}_1), \delta_i)$ for each i. Note that the solution to $T_i(b) = T_j(b)$ is the point at which $T_i(b)$ and $T_j(b)$ change their ranks. Since $T_i(b) = M_i - bX_i$, $T_i(b) = T_j(b)$ yields $b = \frac{M_i - M_j}{X_i - X_j}$, etc. This motivates us to consider the set $\mathcal{A}_1$. By the definition of a_i's, each $T_k(b)$, $k = 1, ..., n$, will not change its rank if $b \in (a_i, a_{i+1})$, $i = 0, ..., m_a$, where $a_0 = -\infty$ and $a_{m_a+1} = \infty$. Thus $\hat{S}_b(T_k(b))$, $k = 1, ..., n$, will not change their values if $b \in (a_i, a_{i+1})$. Consequently, $\mathcal{L}(\cdot)$ is constant in b on the open intervals (a_i, a_{i+1}). It follows that in order to find the maximizer of $\mathcal{L}(b)$ for $b \in (-\infty, \infty)$, it suffices to compare $\mathcal{L}(b)$ for $b = a_i$ and a point in the interval (a_i, a_{i+1}), for all possible i. This results in the set $\mathcal{A}_1 \cup \mathcal{A}_2$. Therefore, the algorithm results in all possible solutions of the SMLE of β.

According to our simulation studies, the algorithm is feasible if the sample size n is around 1000. There is still room to accelerate the algorithm. For example, we find from our simulation experience that it is often that we can skip Step 2 of Algorithm 2 and replace $\mathcal{A}_1 \cup \mathcal{A}_2$ in Steps 3 and 4 by $\mathcal{A}_1$. Recall that Algorithm 1 is for multiple regression. The following is an algorithm for finding an SMLE in multiple regression, which is faster than Algorithm 1 and often works:

Algorithm 3.

1. For each set of p linear independent equations of the form $M_{i_j} - \mathbf{b}'\mathbf{X}_{i_j} = M_{k_j} - \mathbf{b}'\mathbf{X}_{k_j}$, $j = 1, ..., p$, find the solution $\mathbf{b}$. Let $\mathbf{b}_1$, ..., $\mathbf{b}_h$ be all such (distinct) solutions.
2. Compute $\mathcal{L}(\mathbf{b}_i)$, $i = 1, ..., h$.
3. Treat the value of $\mathbf{b}_i$ that maximizes $\mathcal{L}(\mathbf{b}_j)$'s as an SMLE of β.

Note that for multiple regression the number of solutions obtained in Step 1 of Algorithm 3 is often large, and the algorithm may be very slow if n or p is large. For example, if $n = 500$ and $p = 3$, simulation suggests that it may take more than one day to compute the SMLE. A feasible approach is to select randomly a subset among all possible systems of p linearly independent equations; then to compute the solutions $\mathbf{b}_i$'s and finally to find the $\mathbf{b}_i$ that maximizes $\mathcal{L}(\mathbf{b}_j)$ among those $\mathbf{b}_i$'s. This is similar to the Monte Carlo approach.

Remark 2.2. The SMLE of β may not be unique (see Example 4.2 in Section 4). There are several ways to select an estimate if the SMLE is not unique. One can select the SMLE that minimizes $\int_{x<\infty}(x - a(\mathbf{b}))^2 d\hat{F}_{\mathbf{b}}(x)$, $\mathbf{b} \in \mathcal{B}$, where $a(\cdot)$ is as in Remark 2.1; or select one that is the closest one to the BJE or to the center of $\mathcal{B}$, where the center of a set of vectors is a vector with each of its coordinates being the median of the collection of the corresponding coordinates of the vectors in the set.

3. Super efficiency of the SMLE. To-date, we are only able to prove consistency and efficiency of the SMLE when ϵ is a discrete random variable, though simulations indicate that the SMLE is also consistent if ϵ is a continuous random variable and if censoring does exist. To study properties of the SMLE, we make use of the following assumptions:

A1 $\mathrm{P}\left(rank \begin{pmatrix} 1 & \cdots & 1 \\ \mathbf{X}_1 & \cdots & \mathbf{X}_{p+1} \end{pmatrix} = p+1, \delta > 0 \right) > 0$, where $\delta = \mathbf{1}_{(Y \le C)}$.

A2 $E(\ln \mathrm{L}(S_o, \beta))$ is finite.

A3 ϵ and $(\mathbf{X}, C)$ are independent.

A1 is an identifiability condition. It weakens the condition $E(|(\mathbf{X}_1, ..., \mathbf{X}_n)(\mathbf{X}_1, ..., \mathbf{X}_n)'|) \neq 0$ in Bickel (1982). Let f_o be the pdf of F_o. In the complete-data case with discrete ϵ, A2 reduces to $|E(\ln f_o(\epsilon))| < \infty$, which is satisfied by most distributions in the literature. For instance, it corresponds to the mean of a Poisson distribution is finite and the variance of a normal distribution is finite. A3 is an assumption used by several authors under regression models (see for example, Akritas *et al.* (1995)). Some authors assume that $C - \beta'\mathbf{X}$ and ϵ are independent and the censoring variable C also follows a regression model, that is, the distribution of $C - \beta'\mathbf{X}$ does not depend on β (see Sen (1968)).

Theorem 3.1. *Suppose that ϵ is discrete and Assumptions A1, A2 and A3 hold. Then $P\{\tilde{\beta}_n \neq \beta \text{ i.o.}\} = 0$ and $\lim_{n\to\infty} \sup_x |\hat{S}_{\tilde{\beta}_n}(x) - S_o(x)| = 0$ a.s..*

Proof. Denote $\Lambda(S, b) = E(\delta \ln f(T(\mathbf{b})) + (1-\delta)\ln S(T(\mathbf{b})))$, where $S \in \mathcal{F}$ and $f(t) = S(t-) - S(t+)$. Then $\Lambda(S, \mathbf{b}) = \frac{1}{n}E(\ln \mathrm{L}(S, \mathbf{b}))$ and with probability one,

$$\begin{aligned} \underline{\lim}_{n\to\infty} \frac{1}{n}\ln \mathrm{L}(\hat{S}_{\tilde{\beta}_n}, \tilde{\beta}_n) \geq & \lim_{n\to\infty} \frac{1}{n}\ln \mathrm{L}(S_o, \beta) \quad (\text{as } \tilde{\beta}_n \text{ is an SMLE}) \\ = & \Lambda(S_o, \beta) \text{ (by the strong law of large number (SLLN))} . \end{aligned} \tag{3.1}$$

We shall prove the theorem in two cases: (1) $\mathbf{X}$ is discrete; (2) otherwise.

Case (1). Assume $\mathbf{X}$ is discrete. Let Ω_0 be the event such that (3.1) holds. It is proved in Lemma 6.1 (see Appendix) that

$$\mathrm{P}\{\omega \in \Omega_0 \cap \{\tilde{\beta}_n \neq \beta \text{ i.o.}\}\} = 0. \tag{3.2}$$

As a consequence, $P\{\tilde{\beta}_n \neq \beta \text{ i.o.}\} \leq 1 - P(\Omega_0) = 0$, as $\mathrm{P}\{\Omega_0\} = 1$ by (3.1). This completes the proof in Case (1).

Case (2). We shall show that if $\mathrm{P}\{\tilde{\beta}_n \neq \beta \text{ i.o.}\} > 0$ then it leads to a contradiction.

Write $\mathbf{X} = (X_1, ..., X_p)'$, $\tilde{\beta}_n = (\tilde{\beta}_{n,1}, ..., \tilde{\beta}_{n,p})'$ and $\beta = (\beta_1, ..., \beta_p)'$. If $\tilde{\beta}_n \neq \beta$ i.o. then $\tilde{\beta}_{n,k} \neq \beta_k$ i.o. for some $k \in \{1, ..., p\}$. By assumption, $\mathbf{X}$ is not discrete. If it is true that $\tilde{\beta}_{n,k} \neq \beta_k$ i.o. implies X_k is discrete, then it is conceivable that by ignoring the coordinates $\beta_{n,i}$ $(i \neq k)$ satisfying that $\tilde{\beta}_{n,i} = \beta_i$ for all large n, the proof of the theorem would be the same as in Case (1). Without loss of generality (WLOG), we can assume that X_k is not discrete. It is proved in Lemma 6.2 (see Appendix) that w.p.1,

$$\tfrac{1}{n}\ln\mathcal{L}(\mathbf{b}) \to -\infty \text{ uniformly on the set } \{\mathbf{b}: \ b_k \neq \beta_k\}. \tag{3.3}$$

If $\mathrm{P}\{\tilde{\beta}_n \neq \beta \text{ i.o.}\} > 0$, then by (3.3), $P\{\lim_{k\to\infty} \sup_{n \geq k, \tilde{\beta}_{n,k} \neq \beta_k} \frac{1}{n}\ln\mathrm{L}(\hat{S}_{\tilde{\beta}_n}, \tilde{\beta}_n) = -\infty\} > 0$, which contradicts (3.1), namely, with probability one, $\varliminf_{n\to\infty} \frac{1}{n}\ln\mathrm{L}(\hat{S}_{\tilde{\beta}_n}, \tilde{\beta}_n) \geq \Lambda(S_o, \beta) > -\infty$ (by A2). This concludes the proof in Case (2). □

Remark 3.1. Theorem 3.1 is a special case of (1.1), in which F_o only needs to be discontinuous. (1.1) implies strong consistency of $\tilde{\beta}$. When F_o is continuous, (1.1) is no longer true, as pointed out in Yu and Wong (2003). However, we are working on a proof that $\tilde{\beta}$ is still consistent under certain regularity conditions when F_o is continuous.

4. Applications. We apply the SMLE to two medical follow-up data sets and one engineering data set and compare it to the BJE. Note that both the SMLE and the BJE may not be unique. Algorithm 1 in §2.2 gives all solutions to the SMLE of β. Yu and Wong (2002) propose an algorithm that can find all possible solutions to the BJE. In Example 4.1, both the SMLE and the BJE are unique and are significantly different from zero. The SMLE gives a correct sign to β, but the BJE gives a wrong sign. In Example 4.2, there are 2 SMLE's of β and 6 BJE's. However, the SMLE's of β are significantly different from zero with correct signs, but the BJE's are not significant, with 2 negative signs and 4 positive signs. In Example 4.3, both the SMLE and the BJE are significantly different from zero with correct signs. However, the SMLE appears better than the BJE. Of course, there is no guarantee that the data are indeed from linear regression models. However, in these cases, the SMLE estimates look reasonable but the BJE estimates are not as good as the SMLE's.

Example 4.1. The following is a regression analysis on a breast cancer relapse follow-up study based on data obtained from 374 women with stages I - III unilateral invasive breast cancer surgically treated at Memorial Sloan-Kettering Cancer Center between 1985 and 1990. The median follow-up duration was 46 months. Relapse time was given by the time interval between surgery and the initial relapse. If a patient did not relapse toward the end of the study, then her relapse time was right censored. Of the 374 observations, 302 were right censored (no relapse). Maximum tumor size was determined for each woman at the time of surgery. It was an interesting question how remission duration is related to a covariate of the extent of initial tumor size. We fit the data to a model $E(\log Y_i|X_i) = \alpha + \beta X_i$, where Y_i is the relapse time and X_i is the tumor size at the time of surgery.

There is only one BJE and only one SMLE for this data set. The standard deviation (SE) of the SMLE is obtained by the bootstrap method and the SE of the BJE is computed by formula (4.1). For this data set, it took altogether 10 seconds to compute $\tilde{\beta}$ and the BJE.

The a priori guess would be that patients with larger tumor size had remission earlier, that is, β should be negative. However, the BJE is significantly greater than 1 ($\hat{\beta}_{BJE} = 2.029$ with an SE 0.383), which is against our hypothesis. One may argue that the data may not fit the accelerated failure time model. However, $\tilde{\beta} = -2.837$ with an SE 0.415. Thus our SMLE is significantly less than 0, which confirms our a priori guess. In this example, $\tilde{\beta}$ gives better analysis than the BJE. We use the following formula to compute the SE of the BJE:

$$\hat{\Sigma}^2 = \hat{\sigma}^2(\mathbf{C}_u'\mathbf{C}_u)^{-1}, \text{ where } \hat{\sigma}^2 = \frac{\int_{t<\infty}(t - a(\hat{\beta}))^2 d\hat{F}_{\hat{\beta}}(t)}{\int_{t<\infty} 1 d\hat{F}_{\hat{\beta}}(t)}, \ \mathbf{C}_u = \begin{pmatrix} \delta_1 & \delta_1\mathbf{X}_1' \\ \cdot & \cdots \\ \delta_n & \delta_n\mathbf{X}_n' \end{pmatrix}, \tag{4.1}$$

where $a(\cdot)$ is given in Remark 2.1 and $\hat{\beta}$ is an estimate of β (see Miller (1981, p. 152)).

Example 4.2 (the Stanford heart transplant data). The data and detailed description can be found in Miller (1981, p.156) and other textbooks. In this data, right-censored survival time, indicator of death, and age of the recipient at time of transplant were recorded. $n = 69$. For the purpose of illustration, several methods are compared using the data of all death against age. The a-priori guess H_0 would be that younger patients fare better, that is $H_0: \beta < 0$. As in Miller's paper, the logarithm to base 10 of the survival times were used in the regression analysis. In Table 1, we compare the Miller estimator, the BJE and the Cox procedure to the SMLE. The Cox model is $P(Y > t|X = x) = (S(t))^{e^{bx}}$, where S is a baseline survival function. Thus we expect $H_0: b > 0$ rather than $b < 0$ as in the simple linear regression model. The entries in Table 1 related to the Miller estimator and the Cox procedure as well as their standard errors (SE) are copied from Miller (1981, p.156). For the current data, there are altogether 6 BJE's, one of them, $\hat{\beta}_{BJE} = -0.028$ with SE 0.015, was reported in Buckley and James (1979). As commented by Miller (1981, p.162), "The three procedures give conflicting results for the regression on age. The Cox method indicates there is a highly significant age effect. The Miller method says there is no effect due to age, and the Buckley-James approach gives borderline significance to age".

Table 1. Regression analysis on the Stanford heart transplant data
age at transplant v.s. all death

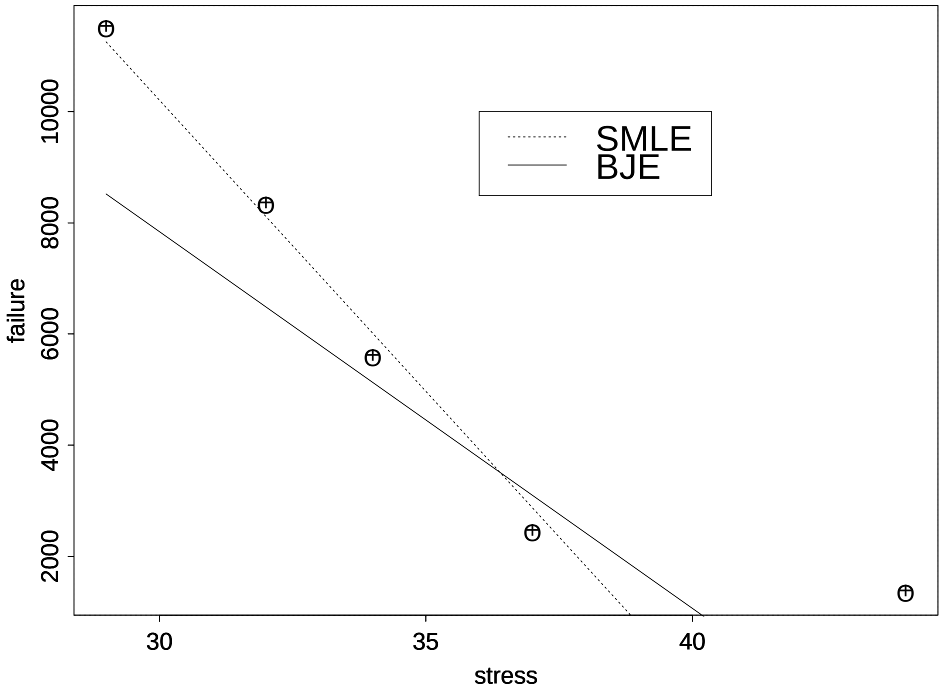
	Miller (SE)	BJE (SE)	SMLE (SE)	Cox (SE)
H_0:	< 0	< 0	< 0	> 0
	-0.006	-0.0278 (0.0149)	-9.771 (2.869)	0.058 (0.023)
	0.004 (0.017)	0.0141(0.0143)	-10.088(2.949)	
	0.002 (0.016)	0.0158(0.0143)		
		0.0157(0.0143)		
		0.0143(0.0143)		
		-0.0110(0.0140)		

In addition, as seen from Table 1, there are one more negative BJE estimate and 4 more positive BJE estimates. The six BJE's basically give the same conclusion as the Miller estimates. On the other hand, there are two SMLE's $\tilde{\beta}$ and both are significantly negative and confirm with both the a priori guess and the Cox procedure. We used the bootstrap method to estimate the standard error of $\tilde{\beta}$. It took less than a second to compute $\tilde{\beta}$.

We also compute the SMLE and the BJE of the regression coefficient to the other regressors in this data set. The results are similar. Though there is no guarantee that the log-linear regression model is correct for these data sets, the fact that the SMLE accords with the Cox procedure and gives a more reasonable estimate than the BJE and the Miller estimator suggests that the SMLE and the Cox procedure are more useful in applications.

Example 4.3. (Creep Rupture Test Data (Nelson and Hahn (1973)). It was desired to estimate the relationship between time to rupture and stress for a certain alloy. The test involved running paired creep-rupture specimens tandem. That is, pairs of specimens were linked end to end and put under a constant test stress. When one member of a pair of specimens failed, the other was taken off test. Tandem specimens were used to hasten the completion of the test and to provide more information on the lower tail of the distribution. The specimens were tested at 5 different stress conditions with one pair of specimens at each condition. Let X stand for stress (in KSI), M for time in hours to the first failure in a test pair and Y for time in hours to failure. Thus Y is only observed for one specimen between a pair and is censored for the other.

Fig. 1. SMLE v.s. BJE For Stress Data



"+" stands for a right-censored observation and "O" an exact observation.

The 10 data are plotted in Figure 1. In Figure 1, we also fit Buckley-James regression line to the data. The SMLE of β is $\tilde{\beta}_n = -1048$, and the BJE is $\hat{\beta}_{BJE} = -678$. The eight points on the left of Figure 1 (4 exact and 4 censored observations) are almost on a straight line whose slope is $\tilde{\beta}_n$. Again, the SMLE seems doing better than the BJE.

5. Simulation studies. In this section, we shall present some of the results in our simulation studies. The main purpose is to study the properties of the SMLE when F_o is arbitrary, *i.e.*, continuous, or discontinuous but not necessarily discrete, as the discrete case is considered in Theorem 3.1. We assume, for convenience, that X, ϵ and C are independent. We consider several cases in our simulation studies:

(1) F_o is continuous (Examples 5.3, 5.4 and 5.5), or discrete (Example 5.1), or discontinuous but not discrete (Examples 5.2).

(2) All the underlying distributions belong to the exponential family (Examples 5.3 and 5.5) or F_o does not belong to the exponential family (other examples).

Example 5.1. Suppose ϵ equals 0.5 and 25.5, with probabilities 0.5 and 0.5, respectively, $X \sim U(2,3)$, $C \sim U(24, 24.2)$, and $(\alpha, \beta) = (13, 1)$. We computed the estimates with sample sizes 32 and 200. The computing times for the SMLE are 1 second and 3.6 seconds, respectively. There is only one SMLE each time.

Example 5.2. Suppose ϵ is a mixture of $U(0, 0.5)$ and a unit point mass concentrated at 0.25, with probabilities 0.5 and 0.5, respectively, $X \sim U(1,2)$, $C \sim U(0,4)$, and $(\alpha, \beta) = (0.25, 1)$. We computed the estimates with sample sizes 32 and 200. The computing times are less than a second and 20 seconds, respectively. There is only one SMLE each time.

Hereafter denote $Exp(\mu, \sigma)$ a distribution with the pdf $f(x) = \frac{1}{\sigma} e^{-[\frac{x-\mu}{\sigma}+1]} \mathbf{1}_{(x>\mu-\sigma)}$.

Example 5.3. Suppose ϵ, C and $\mathbf{X}$ have distributions $Exp(5,2)$, $Exp(3,4)$ and $Exp(2,2)$, respectively. $(\alpha, \beta) = (5,1)$. We compute the estimates with sample sizes 32 and 200. The computing times are 1 second and 4 seconds, respectively. When $n = 200$, there are less than 9 SMLE's each time.

Example 5.4. Suppose $\epsilon \sim U(-1,1)$, $X \sim Exp(0,19)$, $C \sim Exp(0,25)$, $(\alpha, \beta) = (0,2)$. We computed the estimates with sample sizes 32 and 200. The computing times for the SMLE are 1 second and 17 seconds, respectively. When $n = 200$, there are less than 9 SMLE's most of the time.

Example 5.5. Suppose $\epsilon \sim N(0,1)$, $C \sim N(0,6)$, $X \sim N(2,1)$, $(\alpha,\beta) = (0,1)$. We computed the estimates with sample sizes 32, 100 and 200. The computing times for the SMLE are 1 second, 3.5 seconds and 7.6 seconds, respectively.

Table 2. Simulation Results on estimating (α, β)

		(α,β)	SMLE $(\hat{\alpha},\tilde{\beta})$	BJE
Example 5.1 (discrete F_o).				
n=32	average	(13, 1)	(0.941, 1.000)	(-5.388, 1.749)
	SE		(0.044, 0.000)	(22.368, 6.702)
n=200	average	(13, 1)	(0.360, 1.000)	(1.388, 0.301)
	SE		(0.000, 0.000)	(1.075, 0.255)
Example 5.2 (discontinuous F_o).				
n=32	average	(0.25, 1)	(0.248, 1.001)	(0.258, 0.997)
	SE		(0.038, 0.021)	(0.142, 0.961)
n=200	average	(0.25, 1)	(0.250, 1.000)	(0.247, 1.002)
	SE		(0.010, 0.000)	(0.050, 0.033)
Example 5.3 (continuous F_o).				
n=32	average	(5, 1)	(3.693, 1.415)	(4.035, 0.884)
	SE		(3.784, 3.227)	(2.106, 1.708)
n=200	average	(5, 1)	(4.701, 0.957)	(4.701, 0.959)
	SE		(1.294, 0.693)	(0.798, 0.753)
Example 5.4 (continuous F_o).				
n=32	average	(0, 2)	(1.446, 2.129)	(0.008, 2.001)
	SE		(5.411, 0.464)	(0.291, 0.022)
n=200	average	(0, 2)	(0.821, 2.077)	(-0.002, 2.000)
	SE		(0.876, 0.080)	(0.093, 0.007)
Example 5.5 (continuous F_o).				
n=32	average	(0, 1)	(-0.2093, 1.1651)	(-0.0110, 1.0145)
	SE		(1.5933, 0.8971)	(0.7867, 0.3814)
n=200	average	(0, 1)	(-0.2598, 1.1516)	(0.0048, 0.9942)
	SE		(0.7068, 0.3831)	(0.2350, 0.1116)

The results of the above examples are summarized in Table 2 and and some results in Example 5 are summarized in Figures 2-7. In the two rows of each block of Table 2, we present the sample averages and sample standard errors (SE) of the two pairs of estimates in 1000 replications. In Figures 2 and 5, we present the histograms of the numbers of SMLE's of β when the sample size n is 100 and 200, respectively; in Figures 3 and 6, the histograms of the medians of SMLE's of β; in Figures 4 and 7, the histograms of the ranges of SMLE's.

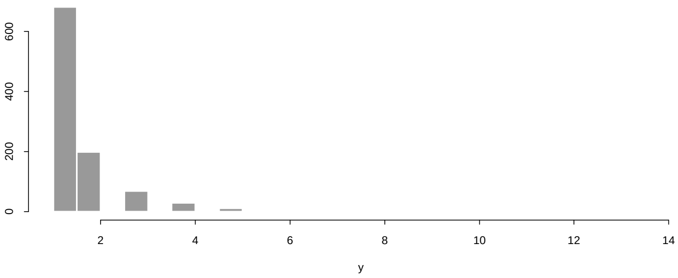


Fig. 2. Histogram of # of SMLE (n=100)

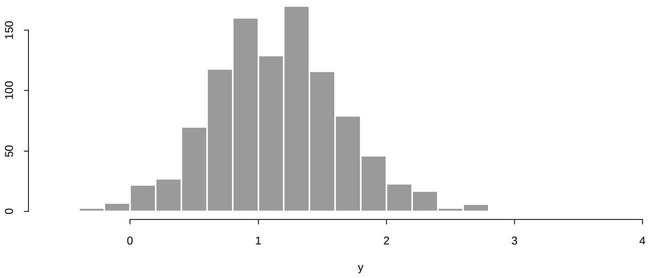


Fig. 3. Histogram of median of SMLE (n=100)

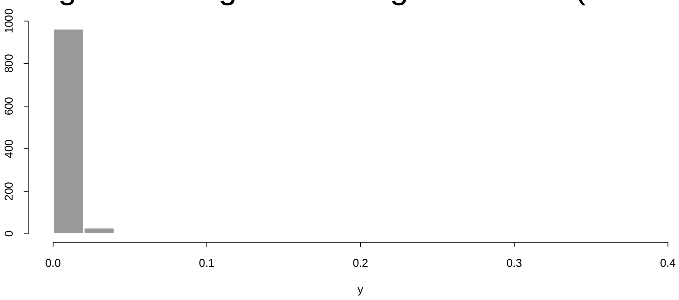


Fig. 4. Histogram of range of SMLE (n=100)

The following are main observations from these simulation results.

1. The SMLE of β is consistent when F_o is continuous as well as when F_o is discontinuous, as in each case the SE is decreasing as n increases and the difference between the true value and the SMLE is with two SE's.

2. If F_o is not continuous, then (1.1) and (1.2) hold (see Examples 5.1 and 5.2). The SMLE of β dominates the BJE of β.

3. If F_o is continuous, then the SMLE does not have much advantage over the BJE (see Examples 5.3 and 5.4). The SMLE of β is better than the BJE in Example 5.3 but worse in Examples 5.4 and 5.5. Note that the BJE is not efficient in Examples 5.1- 5.4.

Fig. 5. Histogram of # of SMLE (n=200)

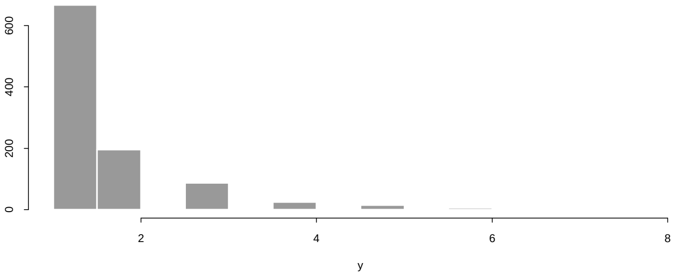


Fig. 6. Histogram of median of SMLE (n=200)

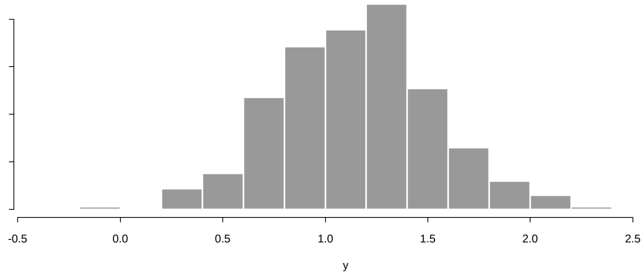
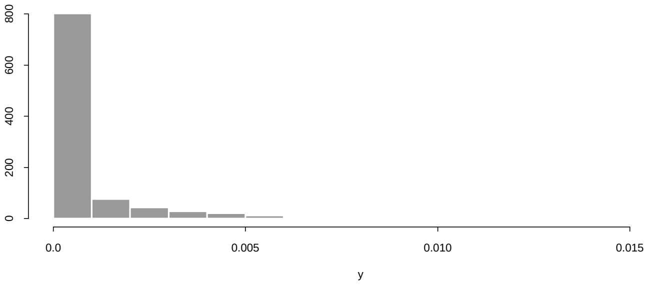


Fig. 7. Histogram of range of SMLE (n=200)



4. As we notice from the above examples that the SMLE is not unique. In particular, if (a) $p = 1$, (b) there is no censoring and (c) both $\mathbf{X}$ and ϵ are continuous, then (1) there are exactly $n(n-1)/2$ solutions to the SMLE of β, and (2) the range of the SMLE's, *i.e.*, the difference between the largest and the smallest SMLE's tends to infinitely, as n tends to infinity. However, if right censoring does exist, simulation results suggest that even though the SMLE may still be non-unique, the range of the SMLE's tends to zero. It is seen from Figures 2-7 that more than 65% of the time the SMLE is unique, the median is within the interval between the lower quartile and upper quartile, and the largest range is 0.4 and 0.015, when the sample size is 100 and 200, respectively. The range decreases significantly as n increases from 100 to 200. It is also worth mentioning that the sample standard deviation of the BJE, which is the efficient estimator in this case, is only 0.11 when $n = 200$. Thus 0.015 is sufficiently small. Moreover, it is seen from Figure 7 that 99% of the time, the range is actually 0.005 while $\beta = 1$.

Variance estimation is an important issue in the SMLE procedure. In view of the properties of the SMLE in the discontinuous cases, we suggest to use the bootstrap method. According to our simulation studies which are not presented here, it seems adequate.

6. Appendix. We shall prove the lemmas needed in the proof of Theorem 3.1. Hereafter, we assume that the assumptions in Theorem 3.1 hold.

Lemma 6.1. *If* $\mathbf{X}$ *is discrete, then (3.2) holds.*

Proof. By assumption, both ϵ and $\mathbf{X}$ are discrete. Denote the range of $\mathbf{X}$ by $\mathcal{R}_{\mathbf{X}} = \{\mathbf{x}_i : \ i \geq 1\}$ and the range of ϵ by $\mathcal{D} = \{d_j : j \geq 1\}$. Then by SLLN, w.p.1,

$$\hat{P}(\delta = h, \epsilon = d_j, \mathbf{X} = \mathbf{x}_i) \to P(\delta = h, \epsilon = d_j, \mathbf{X} = \mathbf{x}_i) \text{ for all } i \geq 1, \, j \geq 1, \, h = 0, \, 1, \qquad (6.1.1)$$

where $\hat{P}$ is the empirical version of P. Let Ω_1 be the subset of Ω_0 that satisfies (6.1.1). Then $\mathrm{P}(\Omega_1) = 1$ and it suffices to show that $\{\tilde{\beta}_n \neq \beta \text{ i.o.}\} \cap \Omega_1 = \emptyset$.

Now fix an $\omega \in \{\tilde{\beta}_n \neq \beta \text{ i.o.}\}$. By taking a subsequence, WLOG, we can assume that $\tilde{\beta}_n \neq \beta$ for all n and $\tilde{\beta}_n \to \mathbf{b}_*$, where $\mathbf{b}_*$ may not be finite. By Helly's selection theorem and by taking a subsequence, WLOG, we can assume that $\hat{S}_{\tilde{\beta}_n}(d_j + (\beta - \tilde{\beta}_n)'\mathbf{x}_i)$ converges to $S_*(d_j + (\beta - \mathbf{b}_*)'\mathbf{x}_i)$ for a function S_* ($\in \mathcal{F}$) and $\hat{f}_{\tilde{\beta}_n}(d_j + (\beta - \tilde{\beta}_n)'\mathbf{x}_i)$ converges to $f_*(d_j + (\beta - \mathbf{b}_*)'\mathbf{x}_i)$ for a function f_*, for all d_j and $\mathbf{x}_i$, where $f_* \geq 0$ and $f_*(t) \leq S_*(t-) - S_*(t+)$ for each t. Then

$$\begin{aligned}
&\overline{\lim_{n\to\infty}} \frac{1}{n} \ln \mathrm{L}(\hat{S}_{\tilde{\beta}_n}, \tilde{\beta}_n) \\
=& \overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{i=1}^{n} \Big[\delta_i \ln \hat{f}_{\tilde{\beta}_n}(T_i(\tilde{\beta}_n)) + (1-\delta_i) \ln \hat{S}_{\tilde{\beta}_n}(T_i(\tilde{\beta}_n)) \Big] \\
=& \overline{\lim_{n\to\infty}} \sum_{i,j} \Big[\hat{P}(\delta = 1, \epsilon = d_j, \mathbf{X} = \mathbf{x}_i) \ln \hat{f}_{\tilde{\beta}_n}(d_j + (\beta - \tilde{\beta}_n)'\mathbf{x}_i) \\
&\qquad + \hat{P}(\delta = 0, \epsilon = d_j, \mathbf{X} = \mathbf{x}_i) \ln \hat{S}_{\tilde{\beta}_n}(d_j + (\beta - \tilde{\beta}_n)'\mathbf{x}_i) \Big] \\
\leq& E(\delta \ln f_*(\epsilon + (\beta - \mathbf{b}_*)'\mathbf{X}) + (1-\delta) \ln S_*(\epsilon + (\beta - \mathbf{b}_*)'\mathbf{X})) \quad \text{(by Fatou's Lemma)}),
\end{aligned} \tag{6.1.2}$$

provided that $\omega \in \Omega_1$. It follows from (3.1) and (6.1.2) that

$$E(\delta \ln f_*(\epsilon + (\beta - \mathbf{b}_*)'\mathbf{X}) + (1-\delta) \ln S_*(\epsilon + (\beta - \mathbf{b}_*)'\mathbf{X})) \geq \Lambda(S_o, \beta). \tag{6.1.3}$$

By the Shannon-Kolmogorov inequality, it can be shown that if $S \in \mathcal{F}$, $f(t) \leq S(t-) - S(t+)$ and $\mathrm{P}\{\epsilon : (f(\epsilon), S(\epsilon)) = (f_o(\epsilon), S_o(\epsilon))\} \neq 1$, then

$$E(\delta \ln f(\epsilon) + (1-\delta) \ln S(\epsilon)) < E(\delta \ln f_o(\epsilon) + (1-\delta) \ln S_o(\epsilon)) \; (= \Lambda(S_o, \beta)),$$

Letting $S(d_j) = S_*(d_j + (\beta - \mathbf{b}_*)'\mathbf{x})$ and $f(t) = S(t-) - S(t+)$, it follows from the discrete assumption on $(\epsilon, \mathbf{X})$ and the above inequality that

$$\begin{aligned}
&E(\delta \ln f_*(\epsilon + (\beta - \mathbf{b}_*)'\mathbf{X}) + (1-\delta) \ln S_*(\epsilon + (\beta - \mathbf{b}_*)'\mathbf{X}) | \mathbf{X}) < \Lambda(S_o, \beta)), \\
&\text{if } |f_*(d_j + (\beta - \mathbf{b}_*)'\mathbf{x}) - f_o(d_j)| > 0 \text{ for a } (j, \mathbf{x}).
\end{aligned} \tag{6.1.4}$$

(6.1.4), (6.1.3) and the definition of f_* imply that

$$\lim_{n\to\infty} \hat{f}_{\tilde{\beta}_n}(d_j + (\beta - \tilde{\beta}_n)'\mathbf{x}_i) = f_o(d_j) \text{ for each } i \text{ and } j, \text{ where } \textstyle\sum_t \hat{f}_{\tilde{\beta}_n}(t) \leq 1 \tag{6.1.5}$$

(as $(\hat{f}_{\tilde{\beta}_n}, \hat{S}_{\tilde{\beta}_n}) \in \mathcal{F}$ by (2.1)).

Since $f_o(d_j) > 0$ and $\sum_{j\geq 1} f_o(d_j) = 1$, WLOG, we can assume that the sequence $\{f_o(d_j)\}_{j\geq 1}$ is decreasing in j. There are finitely many $f_o(d_j) = f_o(d_1)$. Similarly, there are finitely many $f_o(d_j)$'s that equal the i-th largest value of f_o, $i = 2, 3, \ldots$.

For a better presentation, we first assume that $\{f_o(d_j)\}_{j\geq 1}$ is a strictly decreasing infinite sequence. By A1, there are at least two distinct values of $\mathbf{X}$ such that

$$d_j + (\beta - \tilde{\beta}_n)'\mathbf{x}_1 \neq d_j + (\beta - \tilde{\beta}_n)'\mathbf{x}_2, \; \forall\, j, \text{ if } \tilde{\beta}_n \neq \beta, \tag{6.1.6}$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ may depend on n. Since $\tilde{\beta}_n \neq \beta$, (6.1.5) implies

$$\hat{f}_{\tilde{\beta}_n}(d_j + (\beta - \tilde{\beta}_n)'\mathbf{x}_i) \approx f_o(d_j), \; i = 1, 2, \; \forall\, j, \text{ if } n \text{ is large enough}, \tag{6.1.7}$$

$$\hat{f}_{\tilde{\beta}_n}(d_j + (\beta - \tilde{\beta}_n)'\mathbf{x}_i) > f_o(d_{j-1}), \; i = 1, 2, \text{ for } j \leq j_n, \text{ where } \lim_{n\to\infty} j_n = \infty. \tag{6.1.8}$$

If $\{\tilde\beta_n \neq \beta \text{ i.o.}\} \cap \Omega_1$ is not empty, then we arrive the following contradiction:

$$
\begin{aligned}
1 &\geq \lim_{n\to\infty} \sum_t \hat f_{\tilde\beta_n}(t) && \text{(see (6.1.5))}\\
&\geq \lim_{n\to\infty} \sum_{i=1}^{2} \sum_j \hat f_{\tilde\beta_n}(d_j + (\beta - \tilde\beta_n)'\mathbf{x}_i) && \text{(by (6.1.6) and (6.1.8))}\\
&= \lim_{n\to\infty} 2\sum_j f_o(d_j) = 2 && \text{(by (6.1.7))}.
\end{aligned}
$$

This concludes the proof of the lemma under the additional assumption at the beginning of this paragraph.

To prove the lemma without that assumption, WLOG, we can assume that there are exactly m points at which f_o is the same. So they are $d_1 > d_2 > \cdots > d_m$. WLOG, we can further assume that $(\beta - \tilde\beta_n)'\mathbf{x}_1 > (\beta - \tilde\beta_n)'\mathbf{x}_2$ if $\tilde\beta_n \neq \beta$. Then (6.1.6) becomes

$$
d_j + (\beta - \tilde\beta_n)'\mathbf{x}_1 > d_j + (\beta - \tilde\beta_n)'\mathbf{x}_2 \text{ for each } j, \text{ if } \tilde\beta_n \neq \beta. \tag{6.1.9}
$$

It is then obvious that the set $\{d_j + (\beta - \tilde\beta_n)'\mathbf{x}_i :\ i = 1, 2, j = 1, ..., m\}$ contains at least $m+1$ distinct values. WLOG, we can assume that $d_1 + (\beta - \tilde\beta_n)'\mathbf{x}_1$, $d_j + (\beta - \tilde\beta_n)'\mathbf{x}_2$, $j = 1$, ..., m, are distinct. Consequently, the contradiction obtained after (6.1.8) becomes

$$
\begin{aligned}
1 &\geq \lim_{n\to\infty} \sum_t \hat f_{\tilde\beta_n}(t) \qquad \text{(see (6.1.5))}\\
&\geq \lim_{n\to\infty} \Big[\hat f_{\tilde\beta_n}(d_1 + (\beta - \tilde\beta_n)'\mathbf{x}_1) - \hat f_{\tilde\beta_n}(d_{m+1} + (\beta - \tilde\beta_n)'\mathbf{x}_1) + \sum_{j\geq 1} \hat f_{\tilde\beta_n}(d_j + (\beta - \tilde\beta_n)'\mathbf{x}_2)\Big]\\
&\qquad \text{(by (6.1.6) and (6.1.8))}\\
&= f_o(d_1) + \sum_{j\geq 1} f_o(d_j) \qquad \text{(by (6.1.7))}\\
&= f_o(d_1) - f_o(d_{m+1}) + 1 > 1. \ \square
\end{aligned}
$$

Lemma 6.2. *Suppose that* $\mathbf{X} = (X_1, ..., X_p)'$ *and* X_k *is not discrete for a* $k \in \{1, ..., p\}$. *Then w.p.1,* $\frac{1}{n}\ln\mathcal{L}(\mathbf{b}) \to -\infty$ *uniformly on the set* $\{\mathbf{b} :\ b_k \neq \beta_k\}$, *where* $\mathbf{b} = (b_1, ..., b_p)'$. *That is, (3.3) holds.*

Proof. Hereafter, we suppose that $\mathbf{b}$ satisfies $b_k \neq \beta_k$. Let $\mathcal{H}_n(\mathbf{b})$ be the collection of distinct values of $T_1(\mathbf{b})$, ..., $T_n(\mathbf{b})$, then

$$
\hat S_{\mathbf{b}}(s) = \prod_{t\in\mathcal{H}_n, t\leq s} \Big(1 - \frac{N_{t,\mathbf{b}}}{N^+_{t,\mathbf{b}}}\Big), \tag{6.2.1}
$$

where $N^+_{t,\mathbf{b}} = \sum_{k=1}^n \mathbf{1}_{(T_k(\mathbf{b})\geq t)}$ and $N_{t,\mathbf{b}} = \sum_{k=1}^n \mathbf{1}_{(T_k(\mathbf{b})=t,\delta_k=1)}$. Let $\mathcal{D}_{X_k}$ be the set of discontinuous points of F_{X_k}, the cdf of X_k. Let $\mathcal{Q}_{\mathbf{b}}$ be the set of all continuity points of $S_{\mathbf{b}}(t)$, where $S_{\mathbf{b}}$ is a limiting function of $\hat S_{\mathbf{b}}$ based on $(T_i(\mathbf{b}), \delta_i)$'s. Note that $T(\mathbf{b}) = \epsilon + (\beta - \mathbf{b})'\mathbf{X}$. Thus $P(T(\mathbf{b}) \in \mathcal{Q}_{\mathbf{b}}) \geq p_k = P\{X_k \notin \mathcal{D}_{X_k}\}$ uniformly on the set $\{\mathbf{b} : b_k \neq \beta_k\}$. Note that if $t \in \mathcal{Q}_{\mathbf{b}}$ and $b_k \neq \beta_k$, then w.p.1, $N_{t,\mathbf{b}} \leq 1$. Thus

$$
\hat f_{\mathbf{b}}(t) = \hat S_{\mathbf{b}}(t-) - \hat S_{\mathbf{b}}(t) \leq \frac{1}{N^+_{t,\mathbf{b}}} \tag{6.2.2}
$$

in view of (6.2.1). By A1 $\mathrm{P}\{T(\mathbf{b}) \in \mathcal{Q}_{\mathbf{b}}, \delta = 1\} > 0$. Then $\exists\ \eta > 0$ and $t_\eta = t_\eta(\mathbf{b})$ such that

$$
P\{T(\mathbf{b}) \in \mathcal{Q}_{\mathbf{b}}, T(\mathbf{b}) \geq t_\eta\} \geq \eta \text{ and } \eta = P\{T(\mathbf{b}) \in \mathcal{Q}_{\mathbf{b}}, \delta = 1, T(\mathbf{b}) \leq t_\eta\}. \tag{6.2.3}
$$

It is proved in Lemma 6.3 that

$$
N^+_{t,\mathbf{b}}/n \text{ converges uniformly in } t \text{ and } \mathbf{b}. \tag{6.2.4}
$$

Thus uniformly for each $\mathbf{b}$ with $b_k \neq \beta_k$,

$$\begin{aligned}
\frac{1}{n}\ln\mathcal{L}(\mathbf{b}) \leq& \frac{1}{n}\sum_{i:T_i(\mathbf{b})\in\mathcal{Q}_{\mathbf{b}},T_i(\mathbf{b})\leq t_\eta(\mathbf{b})} \delta_i \ln\hat{f}_{\mathbf{b}}(T_i(\mathbf{b})) \\
\leq& \left(\frac{1}{n}\sum_{i:T_i(\mathbf{b})\in\mathcal{Q}_{\mathbf{b}},T_i(\mathbf{b})\leq t_\eta(\mathbf{b})} \delta_i\right)\ln\frac{1}{N^+_{t_\eta \mathbf{b}}} && \text{(by (6.2.2))} \\
\to& \lim_{n\to\infty} \eta\ln\frac{1}{\eta n} && \text{(uniformly in } \mathbf{b} \text{ by (6.2.3) and (6.2.4))} \\
=& -\infty. \quad \square
\end{aligned}$$

Lemma 6.3. *Statement (6.2.4) holds.*

Proof. Let $f_1(d_j, \mathbf{x}_i) = P\{\epsilon = d_j, \mathbf{X} = \mathbf{x}_i\}$ and $\hat{f}_1(d_j, \mathbf{x}_i) = \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{(\epsilon_k = d_j, \mathbf{X}_k = \mathbf{x}_i)}$. Then

$$\begin{aligned}
P\{T(\mathbf{b}) \geq t\} =& \sum_{i\geq 1}\sum_{j\geq 1} \mathbf{1}_{(d_j + (\beta - \mathbf{b})'\mathbf{x}_i \geq t)} f_1(d_j, \mathbf{x}_i); \\
N^+_{t,\mathbf{b}}/n =& \sum_{i\geq 1}\sum_{j\geq 1} \mathbf{1}_{(d_j + (\beta - \mathbf{b})'\mathbf{x}_i \geq t)} \hat{f}_1(d_j, \mathbf{x}_i),
\end{aligned}$$

Since $\sum_{i,j} f_1(d_j, \mathbf{x}_i) = 1$ for each $\eta > 0$, there is an k_η such that

$$\sum_{i,j\leq k_\eta} f_1(d_j, \mathbf{x}_i) \geq 1 - \eta/2. \tag{6.3.1}$$

Since $\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{(\epsilon_k = d_j, \mathbf{X}_k = \mathbf{x}_i)}$ converges to $f_1(d_j, \mathbf{x}_i)$ uniformly in i and j by the Glivenko-Cantelli theorem and k_η is finite, there is an n_η such that

$$\left|\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{(\epsilon_k = d_j, \mathbf{X}_k = \mathbf{x}_i)} - f_1(d_j, \mathbf{x}_i)\right| \leq \eta/k_\eta^2, \text{ if } i, j \leq k_\eta \text{ and } n \geq n_\eta; \tag{6.3.2}$$

$$\sum_{i,j\leq k_\eta} \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{(\epsilon_k = d_j, \mathbf{X}_k = \mathbf{x}_i)} \geq 1 - \eta, \text{ if } n \geq n_\eta. \tag{6.3.3}$$

Note that η, k_η and n_η do not depend on $\mathbf{b}$ and t. Moreover,

$$\begin{aligned}
\left|N^+_{t,\mathbf{b}}/n - P\{T(\mathbf{b}) \geq t\}\right| \leq& \Big|\sum_{i,j\leq k_\eta} \mathbf{1}_{(d_j + (\beta - \mathbf{b})'\mathbf{x}_i \geq t)}\big(\hat{f}_1(d_j, \mathbf{x}_i) - f_1(d_j, \mathbf{x}_i)\big)\Big| \\
&+\Big|\sum_{i \text{ or } j > k_\eta} \mathbf{1}_{(d_j + (\beta - \mathbf{b})'\mathbf{x}_i \geq t)}\big(\hat{f}_1(d_j, \mathbf{x}_i) - f_1(d_j, \mathbf{x}_i)\big)\Big| \\
\leq& k^2 \cdot \eta/k^2 + \eta = 2\eta, \text{ if } n \geq n_\eta, && \text{(by (6.3.1), (6.3.2) and (6.3.3))}
\end{aligned}$$

uniformly in t and $\mathbf{b}$. Since η is arbitrary, it completes the proof of the lemma. $\square$

Acknowledgment: The authors thank the Editor, an Associate Editor and the referee for their invaluable comments and suggestions.

References.

* Bickel, P.J. (1982). On adaptive estimation. *Ann. Statist.* 10 647-671.
* Akritas, M. G., Murphy, S. A., and LaValley, M. P. (1995). The Theil-Sen Estimator with Doubly Censored Data and Applications to Astronomy. *JASA* 90 170 -177.
* Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* 66 429-436.
* Chatterjee, S. and McLeish, D. L. (1986). Fitting linear regression models to censored data by least squares and maximum likelihood methods. *Comm. Stat.* A 15 3227-3243.
* Chatterjee S. and Price B. (1991). Regression analysis by example. 2nd ed. Wiley. New York.
* Cosslett, S.R. (1987). Efficient bounds for distribution-free estimator of the binary choice and the censored regression models. *Econometrica* 55 559-585.
* Duncan, G.M. (1986). A semi-parametric censored regression estimator. *Journal of Econmetrics* 32 5-34.
* Fernandez, L. (1986). Non-parametric maximum likelihood estimation of censored regression models. *Journal of Econmetrics* 32 35-57.
* Hillis, S.L. (1991). Extending M-estimation to include censored data via James's method. *Comm. Stat.* B 20 121-128.
* Horowitz, J.L. (1986). A distribution-free least squares estimator for censored linear regression models. *Journal of Econmetrics* 32 59-84.
* Huber, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* 35 73-101.
* Ichimura, H. (1987). Estimation of single index models. Ph.D thesis, Department of Economics, Massachusetts Institute of Technology.
* Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* 9 1276-1288.
* Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* 74 301-309.
* Manski, C.F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics* 3 205-228.
* Miller, R.G. (1981). Survival Analysis. *Wiley.*
* Moon, C.G. (1989). A Monte Carlo comparison of semiparametric Tobit estimators. *Journal of applied econometics* 4 361-382. Power, J.L. (1986). Symmetrically trimmed least squares estimation for Tobit models. *Econometrica* 54 1434-1460.
* Nelson, W. and Hahn, G. J. (1973). Linear estimation of a regression relationship from censored data-Part II best linear unbiased estimation and theory. *Technometrics* 15 133-150.
* Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* 18 303-328.
* Ruud, P.A. (1986). Consistent estimation of limited dependent variable models despite misspecification of distribution. *Journal of Econmetrics* 32 157-187.
* Sen, P.K. (1968). Estimates of the regression coefficient based on Kendall' tau. *JASA* 63 1379-1389.
* Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis, I. Proc. Kon. Ned. Akad. v. Wetensch. A 53 386-392.
* Tsiatis, A.A (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* 18 354-372.
* Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *JASA* 69 169-173.
* Wei, L.J., Ying, Z.L. and Lin, D.Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* 77 845-851.
* Yu, Q.Q. and Wong, G.Y.C. (2002). How to find all Buckley-James estimates instead of just one ? *JSCS* 72 451-460.
* Yu, Q.Q. and Wong, G.Y.C. (2003). Asymptotic properties of the generalized semi-parametric MLE in linear regression. *Statistica Sinica* (In press).