"Communication in Statistics-Simulation and Computation" 32 p.147-164 (2003)

SEMI-PARAMETRIC MLE IN SIMPLE LINEAR REGRESSION ANALYSIS WITH INTERVAL-CENSORED DATA

Qiqing Yu * and George Y. C. Wong *

Mathematics Department, SUNY at Binghamton, NY 13902, USA
E-mail: qyu@math.binghamton.edu
and
Strang Cancer Prevention Center, 428 E 72nd Street, NY 10021, USA

Running title: SMLE in linear regression.

Current version 3/12/2002
AMS 1991 subject classification: Primary 62 J05; Secondary 62 G05.
Key words: bootstrap, generalized likelihood, iterative algorithm, consistency.

ABSTRACT

Consider the model $Y = \beta X + \epsilon$ with interval-censored data, where ϵ has an unknown cdf F_o. The semi-parametric MLE (SMLE) of β is well defined, but cannot be obtained by algorithms for M-estimators, or by the Newton-Raphson method or the Monte-Carlo method. Thus it has not been studied in the literature even in the case of complete data. We propose a feasible algorithm to obtain all solutions of the SMLE. Simulation suggests that the SMLE is consistent and the bootstrap estimator of the variance of the SMLE matches the sample variance. We compare the SMLE to the Buckley-James estimator (BJE) in four data sets with sample sizes up to 374. The results show that the SMLE is more robust and more reliable than the BJE.

1. INTRODUCTION

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 consider the simple linear regression problem with interval-censored data. In particular, we assume that

A1. $Y = \beta X + \epsilon$, where Y is a random variable, X a covariate taking at least two values, β an (unknown) regression coefficient and ϵ a random variable with an unknown c.d.f. F_o.

Since no further assumption is made on (β, F_o), this is a semi-parametric problem.

There are several estimators for the simple linear regression model with complete data. They are the least squares estimator (LSE), Theil's estimator (Theil (1)), various M-estimators (Huber (2)), adaptive estimators (Bickel (3)), and Bayes estimators.

Sometimes Y may be right censored. There have been extensive studies on linear regression models with right-censored data, see, *e.g.*, Buckley and James (4), Miller (5) and Ritov (6), among others.

Interval censoring is another type of censoring. For example, in medical follow-up studies, a patient may be inspected (or interviewed) total of K times ($K \geq 1$), at $C_{K,1} < \cdots < C_{K,K}$. Time to event of interest, Y, is often not observable but is only known to lie within two consecutive inspection times $C_{K,j}$ and $C_{K,j+1}$, where $C_{K,0} = -\infty$, $C_{K,K+1} = \infty$, and $1 \leq j \leq K$. In such a case we are only able to observe (L, R, X), where $-\infty \leq L < Y \leq R \leq +\infty$. We say such an observation is *interval censored*. Hereafter, we assume that

A2. the observable random vector is (L, R, X), where (L, R) is an extended random vector (*e.g.*, $R = \infty$ under right censoring) such that either $L < Y \leq R$ or $L = Y = R$.

---

* Partially supported by DAMD17-99-I-9390 and DAMD17-00-1-0448.

Odell, *et al.* ((7)) considered the parametric maximum likelihood estimation for interval -censored data based on a Weibull distribution. Self and Grossman (8) considered the linear regression problem with interval-censored data for a given distribution F_o with location-scale parameter and proposed a marginal likelihood approach.

Several authors have proposed estimators for β under linear regression models with interval-censored data, assuming F_o is unknown. Rabinowitz, *et al.* (9) proposed a class of score statistics to estimate β. Their approach parallels the construction of the Buckley & James estimator (BJE) for right-censored data. Li and Pu (10) considered generalizations of Miller's estimator and the BJE for interval-censored data that contain exact observations. Zhang and Li (11) and Li and Zhang (12), among others, studied M-estimators with doubly-censored data and Case 1 interval-censored data. These approaches can be viewed as a modification of the Semi-parametric MLE (SMLE). But they are not an SMLE (see Example 1.1 below).

Under the semi-parametric set-up, it is well known that the LSE is not efficient unless F_o is a normal distribution. The BJE is an extension of the LSE under censoring (see Li and Zhang (12)), thus the BJE is not efficient. Consequently, since the BJE is also an M-estimator (see Zhang and Li ((11), p. 2723)), an M-estimator may not be efficient. Li and Zhang proposed efficient M-estimators for Case 1 interval-censored data. However, it is not clear how to obtain an efficient estimator for arbitrary interval-censored data.

The SMLE of (β, F_o) has long been ignored and there is no algorithm in the literature for obtaining the SMLE with interval-censored data. Recently, Li and Zhang (12) mentioned without a proof that the SMLE (they called the profile MLE) should be consistent with Case 1 interval-censored data. However, there is no hint on how to compute it. Unlike the MLE in other cases, the SMLE cannot be computed by standard numerical methods, *e.g.*, the Monte Carlo method, the Newton-Raphson method, the M-estimation methods and the finite algorithms discussed in Osborne (13). For the sake of simplicity, we illustrate with 4 artificial complete observations as follows.

Example 1.1. Suppose the observations (X_i, Y_i)'s are: $(-1,0)$, $(0,1)$, $((1,2)$ and $(4,0)$. The generalized likelihood function (Kiefer and Wolfowitz (14)) is

$$Ł(F,b) = \prod_{i=1}^{4} f(Y_i - bX_i), \text{ where } f(t) = F(t) - F(t-) \text{ and } F \text{ is a cdf.} \tag{1.1}$$

Then, given b, Ł is maximized by the empirical cdf $\hat{F}_b$. where $\hat{F}_b(t) = \frac{1}{4}\sum_{i=1}^{4} \mathbf{1}_{(Y_i - bX_i \leq t)}$ and $\mathbf{1}_A$ is the indicator function of the event A. That is,

$$Ł(F,b) \leq Ł(\hat{F}_b, b) \text{ and } Ł(\hat{F}_b, b) = \begin{cases} (\frac{3}{4})^3\frac{1}{4} & \text{if } b = 1, \\ (\frac{1}{2})^2(\frac{1}{4})^2 & \text{if } b = 0,\ -1/4 \text{ or } -2/3, \\ (\frac{1}{4})^4 & \text{otherwise.} \end{cases} \tag{1.2}$$

Thus the SMLE of β is 1.

A key requirement in the Newton-Raphson method and the Monte Carlo method (see, *e.g.*, Ingber (15)), as well as in the finite algorithms discussed in Osborne (13), is that Ł is continuous at a neighborhood of the maximum point or Ł is convex. It follows from (1.2) that $l(b) = Ł(\hat{F}_b, b) = \frac{1}{128}$ a.e., which is not the maximum of $Ł(F,b)$. That is, $l(\cdot)$ is neither continuous at the SMLE nor is convex in b. Consequently, the two standard methods, as well as the finite algorithms discussed in Osborne, do not help to find the SMLE.

The M-estimate considered by Zhang and Li (12) is a solution to $\frac{\partial \ln Ł(F,b)}{\partial b} = 0$, where F is properly defined. Since $\frac{d \ln l}{b} = 0$ a.e. by (1.2), the SMLE is somewhat an M-estimator. But this M-estimation approach is non-informative, as every value of b is an M-estimate. Theil's estimator is the median of the collection of slopes of the line segments connecting (X_i, Y_i) and (X_j, Y_j), where $1 \leq i < j \leq n$ $(= 4)$. Thus, Theil's estimator is not an SMLE. It is easy to verify that the LSE is not an SMLE neither. □

Remark 1.1. It is worth mentioning that in Example 1.1, one may modify the likelihood function in (1.1) as follows:

$$Ł = \prod_{i=1}^{n} f(Y_i - bX_i),\ \ f = F' \text{ and } F \in \Theta_0, \tag{1.3}$$

where Θ_0 is a <u>subset</u> of certain absolutely continuous cdf's. Then the values of (F, b) that maximizes Ł over all b and over all $F \in \Theta_0$ can be called a modified SMLE or repaired SMLE. But it has not been called the

SMLE in the literature. This modification is the motivation of the M-estimation approach and the approach based on score functions.

More recently, Yu and Wong (16) studied the SMLE with right-censored data. Simulation suggests that the SMLE $\hat{\beta}$ is consistent and $\lim_{n\to\infty} nVar(\hat{\beta}) = 0$ if F_o is discontinuous. The property was proved under a discrete assumption that (X, Y) takes on finitely many values. In contrast, all the existing estimators do not have the second property. Under continuous assumptions, the SMLE with interval-censored data may attain the efficient lower bound (see Cosslett (17)).

In this paper, we shall study how to derive an SMLE with interval-censored data. In Section 2, we define the SMLE. In Section 3, we propose feasible algorithms to obtain all possible SMLE's. In Section 4, we present simulation results. The simulation results indicate that the SMLE is computationally feasible, is consistent and its standard error can be estimated by the bootstrap method. In Section 5, we apply our procedure to four data sets and compare to the LSE or the BJE. Several comments are made in Section 6.

2. THE SMLE

Let $(Y_i, X_i, \epsilon_i, L_i, R_i)$, $i = 1, ..., n$, be i.i.d. copies of (Y, X, ϵ, L, R). Denote random intervals

$$I_i = I_i(b) = \begin{cases} (L_i - bX_i, R_i - bX_i] & \text{if } L_i < R_i, \\ [L_i - bX_i, R_i - bX_i] & \text{if } L_i = R_i. \end{cases}$$

Note that I_i is a singleton if $L_i = R_i$. Denote μ_F the measure induced by $F \in \mathcal{F}$, where $\mathcal{F}$ is the class of all distribution functions. In other words,

$$\mu_F(I) = \begin{cases} F(v) - F(u) & \text{if } I = (u, v], \\ F(v) - F(u-) & \text{if } I = [u, v]. \end{cases}$$

By assumptions A1 and A2, $I_i(\beta)$ are i.i.d. random intervals and $\epsilon_i \in I_i(\beta)$. Thus the generalized likelihood function defined by Kiefer and Wolfowitz (14) is

$$Ł(F, b) = \prod_{i=1}^{n} \mu_F(I_i(b)), \ F \in \mathcal{F}, \ b \in \mathcal{R} \ \ (= (-\infty, \infty)). \tag{2.1}$$

A semi-parametric MLE of (F_o, β) maximizes Ł. Given b, the GMLE of F_o based on observations $I_i(b)$'s maximizes $Ł(\cdot, b)$ over $F \in \mathcal{F}$. Thus, in order to find the SMLE of (F_o, β), it suffices to maximize

$$l(b) = Ł(\hat{F}_b, b), \quad b \in \mathcal{R}. \tag{2.2}$$

The SMLE of β may not be unique (see Example 3.1). Denote the set of all solutions of the SMLE of β by $\mathcal{B}$. Then each $\hat{F}_b$, $b \in \mathcal{B}$, is an SMLE of F_o.

Under assumption A1, $\alpha = E(\epsilon)$ may not exist. If it does exist, α can be estimated by $\hat{\alpha} = \int t\hat{F}_b(t)$, where $b \in \mathcal{B}$. However, even if $b = \beta$, $\hat{\alpha}$ is not consistent unless Y is observable (with positive probability) everywhere on its range, or P(Y is not censored$|Y = t) > 0$ for all possible t. Thus, in general, α is not identifiable under censoring (see, *e.g.*, Buckley and James (4)). We shall ignore $\hat{\alpha}$ in our study.

3. METHODS

In this section, we shall first introduce an algorithm which guarantees to obtain all solutions of the SMLE and then introduce another algorithm which is faster, but offer no proof.

In view of (2.2), in order to find the SMLE, it suffices to compare $l(b)$, $b \in \mathcal{R}$. Let $u_i(b)$ and $v_i(b)$ be the endpoints of the interval $I_i(b)$, *i.e.*, $u_i(b) = L_i - bX_i$ and $v_i(b) = R_i - bX_i$. Let $T_{2i-1}(b) = u_i(b)$ and $T_{2i}(b) = v_i(b)$, $i = 1, ..., n$. Then we can determine the ranks of the 2n extended random variables. Given b, it is well known that $Ł(\hat{F}_b, b)$ only depends on the ranks of the 2n $T_i(b)$'s (see Turnbull (18)). That is,

$$l(b_1) = l(b_2) \text{ if the rank of } T_i(b_1) \text{ is the same as the rank of } T_i(b_2) \text{ for each } i. \tag{3.1}$$

Note that the ranks of these $T_i(b)$'s will change only at the solutions of the equations $T_i(b) = T_j(b)$, $i \neq j$. The latter equations yield equations of forms

$$L_i - bX_i = L_j - bX_j,\ L_i - bX_i = R_j - bX_j \text{ or } R_i - bX_i = R_j - bX_j, \tag{3.2}$$

where L_i, R_i, L_j and R_j are finite, and $X_i \neq X_j$. Since there are at most $4n^2$ equations of forms in (3.2), there are at most $4n^2$ distinct solutions to these equations, denoted by $b_1 < \cdots < b_m$. Let $b_0 = -\infty$ and $b_{m+1} = \infty$. By construction, if $b \in (b_k, b_{k+1})$, then $T_i(b)$'s will not change their ranks. Consequently, it follows from (3.1) that

$$\text{for each } k,\ l(b)\ (= \mathrm{L}(\hat{F}_b, b)) \text{ is constant on the open interval } (b_k, b_{k+1}). \tag{3.3}$$

There are $m+1$ disjoint open intervals of form (b_k, b_{k+1}) and m disjoint closed intervals of form $[b_k, b_k]$. As a consequence, there are at most $2m+1$ distinct values of $l(b)$, which can be represented by $l(b_j^o)$, $j = 1$, ..., $2m+1$, where $b_1^o = b_1 - 1$, $b_{2m+1}^o = b_m + 1$, $b_{2i}^o = b_i$, $i = 1$, ..., m, and $b_{2i+1}^o = (b_i + b_{i+1})/2$, $i = 1$, ..., $m-1$. Denote

$$\mathcal{A}_1 = \{b_1^o, ..., b_{2m+1}^o\}. \tag{3.4}$$

In order to find an SMLE of β, it suffices to compare $l(b)$ for $b \in \mathcal{A}_1$. Let $\mathcal{B}^o$ be the set of all points in $\mathcal{A}_1$ that maximize $l(b)$, $b \in \mathcal{A}_1$. The $\mathcal{B}^o \subset \mathcal{B}$. Moreover, if $b_{2j+1}^o \in \mathcal{B}^o$, then $(b_{2j}^o, b_{2j+2}^o) \subset \mathcal{B}$ by (3.3). To summarize, we have the following algorithm:

Algorithm 3.1 (for obtaining all solutions of the SMLE of β):

(1) Derive the set $\mathcal{A}_1$ (see (3.4)).

(2) Compute $\hat{F}_b$ and $l(b)$, where $b \in \mathcal{A}_1$. For computing $\hat{F}_b$, we refer to Turnbull (18) or Wellner and Zhan (19). Then

$$b \in \mathcal{B} \text{ iff either (1) } b \in \mathcal{B}^o \text{ or (2) } b \in (b_{2j}^o, b_{2j+2}^o) \text{ for a } b_{2j+1}^o \in \mathcal{B}^o.$$

Each $(\hat{\beta}, \hat{F}_{\hat{\beta}})$, $\hat{\beta} \in \mathcal{B}$, is an SMLE of (β, F_o).

If the SMLE is not unique, one can make the following choices: (1) choosing an SMLE that is closest to the median of $\mathcal{B}$; (2) choosing an SMLE that is closest to the median of $\mathcal{B}^o$; (3) choosing an SMLE that is closest to the BJE. The second choice is the easiest one to implement.

It is often time-consuming to compute the GMLE $\hat{F}_b$. Thus it is desirable to reduce the distinct values of b involved in Step 2 of Algorithm 3.1. The following algorithm reduces the number of b involved by a factor of 4.

Algorithm 3.2.

1. Find the solution b to the equation of form $L_i - bX_i = L_j - bX_j$ or $R_i - bX_i = R_j - bX_j$, where L_i, L_j, R_i and R_j are finite, and $X_i \neq X_j$. Let $\mathcal{A}_2$ be the set of all the distinct elements of these solutions.
2. Compute $\hat{F}_b$ and $l(b)$, $b \in \mathcal{A}_2$. Then identify those points in $\mathcal{A}_2$ which maximize $l(b)$, $b \in \mathcal{A}_2$ and denote $\mathcal{B}^*$ the collection of these points.
3. Choose $b \in \mathcal{B}^*$ that is closest to the median of $\mathcal{B}^*$. Treat it as the SMLE of β.

This algorithm is faster and our simulation results suggest that if n is large then it yields an SMLE. To accelerate the algorithm, we may further make the following modifications:

(i) Randomly select a subset J of the set $\{(i,j) : 1 \leq i < j \leq n\}$. Let $\mathcal{A}_2^*$ be the collection of the solutions b to the equation of form $L_i - bX_i = L_j - bX_j$ or $R_i - bX_i = R_j - bX_j$, where $(i,j) \in J$.

(ii) Modify $\mathcal{A}_2$ as $\mathcal{A}_3 = \mathcal{A}_2^* \cap [a, b]$, where $[a, b]$ is an interval to which we suspect that β belongs. For example, we may let $[a, b] = [\tilde{\beta} - 3\tilde{\sigma}_{\tilde{\beta}}, \tilde{\beta} + 3\tilde{\sigma}_{\tilde{\beta}}]$, where $\tilde{\beta}$ is the BJE and $\tilde{\sigma}_{\tilde{\beta}}$ is the standard error (SE) of $\tilde{\beta}$.

Note that Step (i) reduces the amount of (i,j) in calculation, and Step (ii) further reduces the cost in computing the GMLE of F for selected $b \in \mathcal{A}_3$.

The following example illustrates the difference between the two algorithms.

Example 3.1. Suppose that there are 3 observations (L_i, R_i, X_i)'s. They are $(1,4,1)$, $(3,6,-1)$, $(1,2,1)$. We first consider Algorithm 3.1. Step 1 results in $m = 5$, $(b_1, b_2, b_3, b_4, b_5) = (-2.5, -2, -1, -0.5, 0.5)$ and

$\mathcal{A}_1 = \{b_1^o, \ldots, b_{11}^o\} = \{-3.5, -2.5, -2.25, -2, -1.75, -1, -0.75, -0.5, 0, 0.5, 1.5\}$.

In Step 2, we need to compute $\hat{F}_b$ and $l(b)$, $b \in \mathcal{A}_1$. For simplicity, we only demonstrate for b_1^o and b_3^o.

(1) $b_1^o = -3.5$. Then the I_i's are $(4.5, 7.5]$, $(-0.5, 2.5]$, $(4.5, 5.5]$. $\hat{F}_{-3.5}(t) = \frac{1}{3}\mathbf{1}_{(t\geq 1)} + \frac{2}{3}\mathbf{1}_{(t\geq 5)}$. $l(b_1^o) = \frac{1}{3}(\frac{2}{3})^2$.

(2) $b_3^o = -2.25$. Then the I_i's are $(2.25, 6.25]$, $(0.75, 3.75]$ and $(2.25, 4.25]$. $\hat{F}_{-2.25}(t) = \mathbf{1}_{(t\geq 3.5)}$ and $l(b_3^o) = 1$.

To summarize, Step 2 gives the following results:

$$\begin{array}{c} i: \\ b_i^o: \\ l(b_i^o): \end{array} \begin{array}{c} \begin{array}{ccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \end{array} \\ \begin{pmatrix} -3.5 & -2.5 & -2.25 & -2 & -1.75 & -1 & -0.75 & -0.5 & 0 & 0.5 & 1.5 \\ \frac{4}{27} & \frac{4}{27} & 1 & 1 & 1 & 1 & 1 & 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix} \end{array}$$

Thus $\mathcal{B} = (-2.5, -0.5)$ and $\mathcal{B}^o = \{-2.25, -2, -1.75, -1, -0.75\}$. In this example there are infinitely many SMLE's. We choose the median of $\mathcal{B}^o$, that is, $\hat{\beta} = -1.75$.

If we use Algorithm 3.2, Step 1 results in $\mathcal{A}_2 = \{-2, -1\}$ and Step 2 results in an estimate $\hat{\beta} = -1$. Both algorithms lead to an SMLE and the second algorithm is obviously faster.

Remark 3.1. Variance estimation is important for making inferences. Since we do not make any specific assumption on F_o, there are two possibilities for the variance of $\hat{\beta}$.

1. Under interval censoring without exact observations, if the regularity conditions stated for the Cramer-Rao lower bound hold and F_o is differentiable, then a consistent estimator of the efficient lower bound for the regression problem is

$$\hat{\sigma}^2_{\hat{\beta}} = 1/\sum_{i=1}^{n}\left(\frac{f(R_i - \beta X_i) - f(L_i - \beta X_i)}{\mu_F(I_i(\beta))}(X_i - \overline{X})\right)^2\Big|_{\beta=\hat{\beta}, f(t)=\frac{F(t+h_n)-F(t-h_n)}{2h_n}, F=\hat{F}_{\hat{\beta}}}, \quad (3.5)$$

where $h_n \approx 0$, *e.g.*, $h_n = n^{-1/5}$.

2. If F_o is differentiable but $\frac{\partial E(\ln l(\beta))}{\partial \beta} \neq E(\frac{\partial \ln l(\beta)}{\partial \beta})$, or if F_o is not continuous, then the regularity conditions stated for the Cramer-Rao lower bound do not hold. In such situations, $\hat{\sigma}_{\hat{\beta}}$ is not a good estimator.

In view of the above discussion, we suggest to use the bootstrap method (see, *e.g.*, Davison and Hinkley (20)) to estimate the variance of $\hat{\beta}$. It seems to us from simulation that the bootstrap estimator is consistent.

4. SIMULATION RESULTS

Hereafter, we denote $\hat{\beta}$ the SMLE. The simulation results suggest that $\hat{\beta}$ is consistent and the bootstrap estimates of the variances of $\hat{\beta}$ match the sample variances. The simulation results also indicate that the SMLE is a feasible procedure computationally.

In our simulations, we make use of a Case 1 or Case 2 interval censorship model. Under the Case 2 interval censorship model, the observable random vector satisfies

$$(L, R) = \begin{cases} (-\infty, U) & \text{if } Y \leq U \\ (U, U+V) & \text{if } U < Y \leq U+V \\ (U+V, \infty) & \text{if } Y > U+V, \end{cases}$$

$V > 0$, and Y and (U, V) are independent. Under the Case 1 interval censorship model, the observable random vector satisfies

$$(L, R) = \begin{cases} (-\infty, U) & \text{if } Y \leq U \\ (U, \infty) & \text{if } Y > U, \end{cases}$$

and Y and U are independent.

In our simulation, we further assume that the underlying distribution of Y and the censoring vector satisfy the following conditions: (1) U is a continuous nonnegative random variables; (2) V is a nonnegative continuous random variable; (3) X, ϵ, U and V are independent.

Note that in our assumptions, ϵ and X may be continuous or discontinuous. We present simulation results under three different sets of distributions. Under each set of assumptions, we compute $\hat{\beta}$ based on sample sizes 30 and 100, respectively, with 500 simulations in each case. The program was written in C language and the simulation was carried out on a Pentium III PC.

Case 1. (Case 2 interval censoring). Assume that ϵ has the uniform distribution on the interval $(0, 2)$; U, V and X have exponential distributions; and $\beta = 1$.

Case 2. (Case 2 interval censoring). Assume that ϵ takes values 0.8 and 4 w.p. 0.8 and 0.2, respectively; X is a discrete random variable which takes values i w.p. $i/15$, $i = 1, 2, 3, 4, 5$; $\beta = 1$; and U and V have uniform distributions on the intervals $(0, 4)$ and $(0.5, 5)$, respectively.

Case 3. (Case 1 interval censoring). Assume that ϵ has a uniform distribution on the union of intervals $(0, 0.5) \cup (50.5, 51)$; the inspection time U has a uniform distribution in the interval (0,3); X takes values 0.5 and 1 w.p. 0.5 and 0.5, respectively; and $\beta = 1$.

Table 1. Simulation Results on estimating β with interval-censored data.

cases		$n = 30$	$n = 100$	β
Case 1.	$\hat{\beta}_n$ average (SE)	1.020 (0.710)	1.060 (0.401)	1
continuous F_o	average $\hat{se}_B$	0.489	0.319	
	SE of $\hat{se}_B$	(0.264)	(0.203)	
Case 2.	$\hat{\beta}_n$ average (SE)	1.015 (0.279)	1.007 (0.073)	1
discrete F_o	average $\hat{se}_B$	0.286	0.098	
	SE of $\hat{se}_B$	(0.067)	(0.013)	
Case 3.	$\hat{\beta}_n$ average (SE)	0.824 (1.355)	0.986 (0.391)	1
continuous F_o	average $\hat{se}_B$	0.430	0.275	
	SE of $\hat{se}_B$	(0.336)	(0.098)	

In Table 1, the entries in the column corresponding to n stand for the results with sample size n. The results in Table 1 suggest that $\hat{\beta}$ is consistent in all the three cases. Verify that $\frac{\partial E(\ln l(\beta))}{\partial \beta} \neq E(\frac{\partial \ln l(\beta)}{\partial \beta})$ in all the three cases, thus we cannot use $\hat{\sigma}^2_{\hat{\beta}}$ (see (3.5)) to estimate the variance of $\hat{\beta}$. In each simulation, using the bootstrap method described in Efron and Tibshirani (21) or Davison and Hinkley (20), we resampled (with replacement) B times. Each resample size is n. The bootstrap estimate of $\sigma_{\hat{\beta}}$, denoted by $\hat{se}_B$, is the sample standard error (SE) of the B estimates $\hat{\beta}$ based on the B resamples. As suggested by Efron and Tibshirani, we set $B = 30$. The entries corresponding to the row of "average $\hat{se}_B$" and the row of "SE of $\hat{se}_B$" are the sample means and the sample standard errors of these $\hat{se}_B$ in the 500 simulations, respectively. The differences between the sample SE's of $\hat{\beta}$ and the bootstrap estimates $\hat{se}_B$ are within 2 standard errors for the three cases, except for Case 3 when $n = 30$. This suggests that the bootstrap estimator $\hat{se}_B$ of the $\sigma_{\hat{\beta}}$ is appropriate.

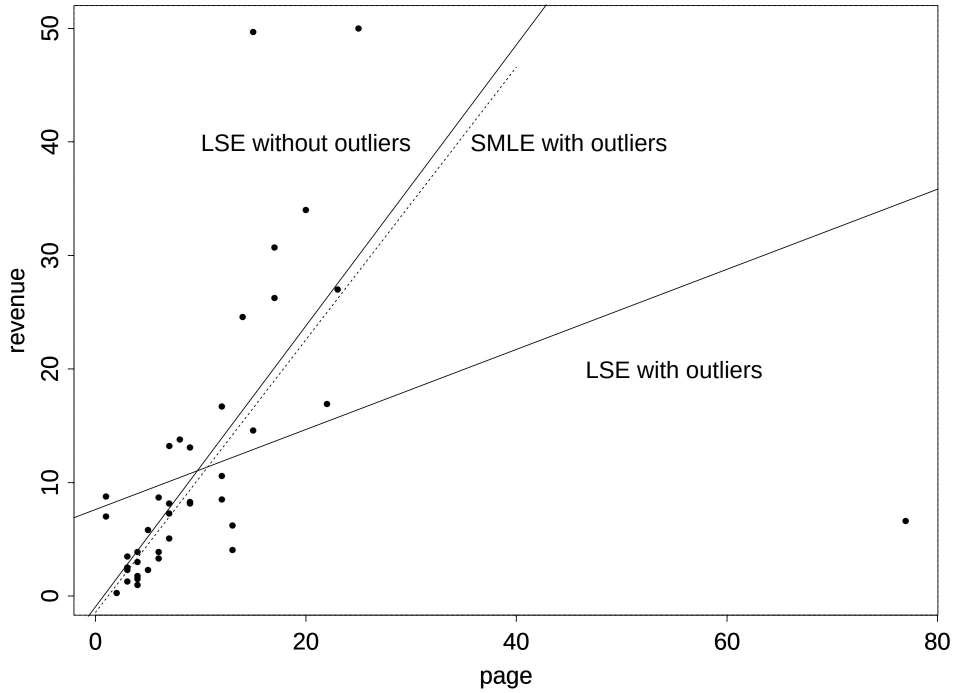
We also carried out simulation under the assumption that ϵ has a normal distribution or an exponential distribution. The results are similar and are not reported here.

5. APPLICATIONS

In this section, we apply our method to several data sets, including complete data ($L = R$ w.p.1), right-censored data and interval-censored data. We also compare the SMLE to the BJE, which is also an M-estimator (see Li and Zhang (12)).

Example 5.1. (Magazine advertising (Chatterjee and Price ((22), p. 257)). In a study of revenue from advertising, data were collected for 41 magazines in 1986. There was no censoring. Let X denote the number of pages of advertising and Y the advertising revenue.

Fig. 1. SMLE v.s. LSE



The 41 data are plotted in Figure 1. Roughly speaking, there are three outliers in the data set. They are (25, 50), (15, 49.7), (77, 6.6). The SMLE of β is unique for this data set. The SMLE and the LSE are significantly different (see the first block of Table 2). The entries in the second block of Table 2 are results after deleting the three outliers. From Table 2, it is seen that the SMLE of β does not change after deleting outliers, though the estimate of α changes.

In Figure 1, we also plot the fitted straight lines with and without deleting those three outliers. We further plot the fitted line by the SMLE method without deleting the outliers. From Figure 1, it is seen that the fitted line by the SMLE approach using the original data is very close to the least squares fitted line after deleting outliers. This suggests that the SMLE is robust while the LSE is not.

Table 2. Results on estimating (α, β)

		SMLE (SE)	LSE (SE)
with outliers	β	1.200 (0.196)	0.353 (0.1449)
	α	-1.427 (3.178)	7.604 (2.3466)
without outliers	β	1.200 (0.1379)	1.238 (0.138)
	α	-0.642 (1.410)	-0.962 (1.409)

Example 5.2. (Application to the Stanford heart transplant data). The data and detailed description can be found in Miller ((5), p. 156). In this data, right-censored survival time, indicator of death, indicator of rejection, 2 covariates including T5 mismatch and age of the recipient at time of transplant were recorded for 69 patients. For illustrative purposes, Buckley and James (4) compared their method to Miller's estimator and Cox method by fitting a simple linear regression for all death against age. We compare our method to the BJE using the same data under the log linear regression model.

Our algorithm results in an SMLE $\hat{\beta} = -9.77$ with a standard error (SE) 1.56. The SMLE is significantly negative and is consistent with our a priori guess, namely, younger patients fare better in the surgery.

There are 3 BJE's of β: -0.028, 0.014 and 0.016 with the SE 0.015 (only the first estimate and its SE were reported in Buckley and James (4)). They are all not significant at level 0.05. This means that there is no effect due to age.

In this example, the SMLE make more practical sense than the BJE, and thus is more reliable than the BJE.

Example 5.3. (Application to a cancer research data). Our data analysis is applied to a standard breast cancer relapse follow-up study based on data 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. A relapse that took place between two successive follow-up visits was regarded as interval censored. If a patient did not relapse toward the end of the study, then her relapse time was right censored. Of the 374 observations, 300 were right censored (no relapse), 21 were left censored and 53 were strictly interval censored. Bone marrow micrometastasis (BMM) was determined for each woman at the time of surgery. An important question is whether remission duration is related to the extent of initial tumor burden defined as number of BMM cells detected. We compute the BJE and $\hat{\beta}$ under the log linear regression model.

One expects that the larger the BMM was, the shorter the patient survived. Thus $\beta < 0$. The BJE of β is -0.012 and our estimate is $\hat{\beta} = -0.059$ with a (bootstrap) standard error 0.02. Note that the BJE does not fall in the interval $(\hat{\beta} - 2SE, \hat{\beta} + 2SE)$, so they are significantly different. Moreover, the SMLE leans more toward the correct direction than the BJE.

Example 5.4. (Application to a breast cosmesis data). We applied our procedure to the interval-censored data set published in Finkelstein and Wolfe (23). The data is a result of a retrospective study to compare early breast cancer patients who have been treated with primary radiation therapy and a adjuvant chemotherapy to those treated with radiotherapy alone with respect to the cosmetic effects of their treatment. There is only one covariate, the group status, in this data set. There are 94 patients.

The BJE of β is -0.29 and our estimate $\hat{\beta}$ is -0.67 with a (bootstrap) standard error 0.336. The BJE falls in the interval $(\hat{\beta} - 2SE, \hat{\beta} + 2SE)$. Thus they are not significantly different.

6. CONCLUSION

In this paper, we propose algorithms for the SMLE of β. The procedure is actually applicable to complete data or right-censored data, as demonstrated in Examples 5.1 and 5.2.

We believe that each SMLE of β is consistent under interval censoring. The outline of the proof is as follows.

(1) By the definition of the SMLE and the strong law of large number, $\underline{\lim}_{n\to\infty} (\ln \mathrm{L}(\hat{F}_{\hat{\beta}}, \hat{\beta}))/n \geq E(\ln \mathrm{L}(F_o, \beta))/n$ a.s..

(2) By Fatou's Lemma, $\overline{\lim}_{n\to\infty} (\ln \mathrm{L}(\hat{F}_{\hat{\beta}}, \hat{\beta}))/n \leq E(\ln \mathrm{L}(F_*, b_*))/n$ a.s., where (F_*, b_*) is the limit of a convergent subsequence of the SMLE $(\hat{F}_{\hat{\beta}}, \hat{\beta})$.

(3) $E(\ln \mathrm{L}(F, b))/n \leq E(\ln \mathrm{L}(F_o, \beta))/n$ for each (F, b), and the equality holds only if $b = \beta$ by the Shannon-Kolmogorov inequality and the following assumption:

A3. $F \in \mathcal{F}$ and $\mathrm{P}\{F(W - bX) = F_o(W - \beta X)$ for $W = L$ or $R\} = 1$ imply $b = \beta$.

(The assumptions made in Section 4 satisfy A3.)

Statements (1) and (2) imply that $E(\ln \mathrm{L}(F_*, b_*))/n = E(\ln \mathrm{L}(F_o, \beta))/n$, thus statement (3) implies that $b_* = \beta$. Since b_* is an arbitrary limiting point of $\hat{\beta}$, $\hat{\beta}$ is consistent.

In general, the BJE is not efficient and is not robust, as the BJE reduces to the LSE in the case of complete data and the LSE is not efficient and is not robust (see Draper and Smith ((24), p. 342)). We apply both the BJE procedure and the SMLE to complete data, right-censored data and interval-censored data. The SMLE seems quite robust and always give reasonable estimates. The BJE may give unreasonable estimates.

For the semi-parametric set-up, there are several estimation procedures, *e.g.*, the BJE and the M-estimators. They are all obtained by iterative algorithms. The current procedure is also obtained by an iterative algorithm, unless the data are Case 1 interval-censored data. In the latter case, the SMLE can be obtained by a non-iterative algorithm, as the GMLE has closed-form solution (see Ayer *et al.* (25)). The reason is as follows. There are two steps in obtaining all SMLE's: (1) finding all the discontinuity points b_1, ..., b_m ($m \leq 4n^2$, see (3.4)), and (2) computing the GMLE $\hat{F}_b$ and comparing $l(b)$, $b = b_i$'s (see algorithm 3.1). Since the GMLE with Case 1 interval-censored data has a closed-form expression, all the SMLE's of β can be obtained in finitely many steps. However, with Case 1 interval-censored data, the BJE and the M-estimates cannot be obtained by a non-iterative algorithm.

We only discuss the simple linear regression model in this paper. The method can be extended to the multiple linear regression model. However, the computation can only performed on high-speed computers. Also, only extension of Algorithm 3.2 together with Steps (i) and (ii) is feasible due to the heavy computation cost.

BIGLIOGRAPHY

(1) Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis, I. Proc. Kon. Ned. Akad. v. Wetensch. A 53 386-392.
(2) Huber, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* 35 73-101.
(3) Bickel, P.J. (1982). On adaptive estimation. *Ann. Statist.* 10 647-671.
(4) Buckley, J and James, L. (1979). Linear regression with censored data. *Biometrika* 66 429-436.
(5) Miller, R.G. (1981). *Survival Analysis. Wiley.* N.Y.
(6) Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* 18 303-328.
(7) Odell, P.M., Anderson, K.M. and D'Agostino, R.B. (1992). Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. *Biometrics* 48 951-959.
(8) Self, S.G. and Grossman, E.A. (1986). Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers. *Biometrics* 42 521-530.
(9) Rabinowitz, D. Tsiatis, A. and Aragon, J. (1995). Regression with interval-censored data. *Biometrika* 82 501-513.
(10) Li, L.X. and Pu, Z.W. (1999). Regression models with arbitrarily interval-censored observations. *Comm. in Statist., Theory and Methods* 28 1547-1563.
(11) Zhang, C.H. and Li, X. (1996). Linear regression with doubly censored data. *Ann. Statist.* 24 2720-2743.
(12) Li, G. and Zhang, C.H. (1998). Linear regression with interval censored data. *Ann. Statist.* 26 1306-1327.
(13) Osborne M.R. (1985). *Finite algoriths in optimization and data analysis.* Wiley N.Y.
(14) Kiefer, J and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* 27 887-906.
(15) L. Ingber (1989). Very fast simulation re-annealing. *Mathematical and Computer Modelling* 12 967-973.
(16) Yu, Q.Q. and Wong, G.Y.C. (1999). The semi-parametric MLE in linear regression with right-censored data. (Preprint).
(17) Cosslett, S.R. (1987). Efficient bounds for distribution-free estimator of the binary choice and the censored regression models. *Econometrica* 55 559-585.
(18) Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *JASA* 69 169-173.
(19) Wellner, J.A. and Zhan, Y. (1997). A hybrid algorithm for computation of the NPMLE from censored data. *JASA* **92** 945-959.
(20) Davison, A.C. and Hinkley, D.V. (1997) (1997). *Bootstrap methods and their application.* Cambridge Press. N.Y.
(21) Efron, B and Tibshirani, R.J. (1993). *An introduction to the bootstrap.* Chapman & Hall. N.Y.
(22) Chatterjee S. and Price B. (1991). *Regression analysis by example.* Wiley, N.Y.
(23) Finkelstein, D. M. and Wolfe, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* 41 933-945.
(24) Draper, N.R. and Smith, H. (1981). *Applied Regression analysis. Wiley.* N.Y.
(25) Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T. and Silverman, E. (1955). An empirical distribution function for sampling incomplete information. *Ann. Math. Statist.* 26 641-647.