A Modified Semi-Parametric MLE
In Linear Regression Analysis
With Complete Data Or Right-Censored Data

By Qiqing Yu* and George Y. C. Wong *

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

Current version: 8/27/2004

First version: 4/2/2000

Short title: Modified SMLE

AMS 1991 subject classification: Primary 62 J 05; Secondary 62 G05.

Key Words: Simulation studies, non-iterative algorithm, semi-parametric MLE, asymptotic normality, arbitrary error distribution.

Abstract: Consider a linear regression model where the response variable may be right censored. The standard MLE-based parametric approach to estimation of regression coefficients requires that the parametric form of the error distribution is known. Given a data set, we may not be able to find a valid parametric form for the error distribution. In such a case, the error distribution is unknown and arbitrary, and a semi-parametric approach is plausible. A special modified semi-parametric MLE of the regression coefficients is proposed in this paper. Simulation suggests that the MSMLE is consistent, asymptotically normally distributed and maybe efficient. We apply the new procedure to engineering data.

---

* partially supported by DOD Grant DAMD17-99-1-9390 and DAMD17-00-1-0448.

1. Introduction. We consider the linear regression problem with right-censored data. Regression analysis is one of the most widely used statistical techniques. Its applications occur in almost every field, including engineering, economics, the physical sciences, management, life and biological sciences and the social sciences. In particular, one desires to estimate the relationship between one variable Y and an independent variable (or vector) $\mathbf{X}$. For instance, we may want to estimate the relationship between time to rupture and stress for a certain alloy, or between shear strength and age of a batch of sustainer propellant in rocket testing, or between time to failure for an insulation for electric motors and temperature. One relationship between Y and $\mathbf{X}$ is $Y = \boldsymbol{\beta}'\mathbf{X} + \epsilon$, where Y is a random variable, $\mathbf{X}$ a $p \times 1$ covariate vector, $\boldsymbol{\beta}'$ the transpose of a regression coefficient vector $\boldsymbol{\beta}$, and ϵ a random variable with an unknown distribution function F_o ($F_o(t) = P(\epsilon \leq t)$). Note that $\alpha = E(\epsilon)$ may not be zero so that the intercept term that one finds in the standard textbook regression model is incorporated into ϵ.

Sometimes Y may not be observed. For instance, some insulation for electric motors did not fail by the termination of a temperature-accelerated life test (see the Insulation data (Nelson (1973)), and only the time interval, C, of the test was recorded. In this case, Y is the failure time of the insulation material and may be right censored by C.

This is a semi-parametric estimation problem, as $\boldsymbol{\beta}$ and F_o are unknown. Several estimators of $\boldsymbol{\beta}$ are available, including

(1) the least squares estimator (LSE) and its various modifications under right censoring *e.g.*, Miller's (1981, p.146) estimator, the Buckley-James (1979) estimator (BJE), and the modifications proposed by Chatterjee and Mcleish (1986) and Leurgans (1987);

(2) the Theil-Sen estimator with complete-data (Theil (1950) and Sen (1968)), and its modifications with right-censored data (Ireson and Rao (1985) and Akritas, Murphy and LaValley (1995));

(3) various M-estimators (Huber (1964), Ritov (1990) and Zhang and Li (1996));

(4) adaptive estimators (Bickel (1982) and Koul and Susarla (1983));

(5) the semi-parametric MLE (Yu and Wong (2003)).

The standard MLE-based parametric alternatives require the assumption that the form of the error distribution is known, which is not required for the semi-parametric approach. It is well known that under certain regularity conditions, the parametric MLE of the regression coefficient vector is consistent and efficient. Zhang and Li (1996) show that under the semi-parametric set-up certain M-estimators of the regression coefficients are also consistent and efficient. The drawback in the M-estimation approach, like most numerical algorithms for solving optimization problems, is that the solution to the M-estimator may not be unique and there is no general algorithm that guarantees that one obtains the "efficient M-estimate", as demonstrated in Yu and Wong (2001).

In this paper, we propose a modified semi-parametric MLE (MSMLE) of $\boldsymbol{\beta}$**.** The estimator maximizes a modified generalized likelihood function, to be specified in the next section. Recall that in parametric MLE-based analysis, if the MLE does not have a closed form solution, then in general, a numerical algorithm only guarantees that it leads to a zero-point of the score function, instead of the MLE. We propose an algorithm that **guarantees** finding the MSMLE. Our simulation studies suggest that the MSMLE is likely to be consistent and asymptotically normally distributed, and maybe efficient. In a technical report (see Yu and Wong (2004)), we actually prove these properties under a discrete assumption on the error distribution.

The paper is organized as follows. In Section 2, we introduce the MSMLE. In Section 3, we illustrate how to compute the MSMLE and its variance by a simple right-censored engineering data set. In Section 4, we present our simulation results on the asymptotic properties of the estimator and compare them to the corresponding parametric MLE or BJE. In Section 5, we apply our procedure to engineering data. Section 6 is a concluding remark. Some detailed proofs of the statements in Sections 2 and 3 are given in the Appendix.

2. The MSMLE. Let Y_i and $\mathbf{X}_i$ satisfy $Y_i = \boldsymbol{\beta}'\mathbf{X}_i + \epsilon_i$, $i = 1, ..., n$, where $(Y_i, \mathbf{X}_i, \epsilon_i, C_i)$ are n i.i.d. copies of $(Y, \mathbf{X}, \epsilon, C)$ which are explained in Section 1. Our observations are $(M_i, \delta_i, \mathbf{X}_i)$, where M_i equals Y_i if it is not censored, and is a censored value otherwise; δ_i is the indicator that is 0 if Y_i is censored and 1 otherwise. Let $S_o = 1 - F_o$ and denote by f_o the pdf of F_o. Let F be a cdf, f its pdf and $S = 1 - F$. For a $p \times 1$ vector $\mathbf{b}$, denote

$$T_i(\mathbf{b}) = M_i - \mathbf{b}'\mathbf{X}_i.$$

The generalized likelihood function of $(\boldsymbol{\beta}, S_o)$ defined by Kiefer and Wolfowitz (1956), is

$$Ł(S, \mathbf{b}, f) = \prod_{i=1}^{n} [(f(T_i(\mathbf{b})))^{\delta_i}(S(T_i(\mathbf{b})))^{1-\delta_i}].$$

Both S and f in the above expression are unspecified. One can estimate S by the product-limit-estimator (PLE), $\hat{S}_{\mathbf{b}}$ (see (A.1) in the Appendix). Moreover, motivated by the fact that ϵ is often continuous, it is desirable to estimate f by a kernel estimator,

$$f_{\hat{S}_{\mathbf{b}}} = \frac{\hat{S}_{\mathbf{b}}((t-h_n)-) - \hat{S}_{\mathbf{b}}(t+h_n)}{2h_n}, \text{ where } h_n > 0, \lim_{n\to\infty} h_n = 0 \quad (2.1)$$

(*e.g.*, $h_n = O(n^{-1/5})$, as suggested in Härdle (1990, p.59)), and $S(t-) = \lim_{x\uparrow t} S(x)$. It is obvious that $f_{\hat{S}_{\mathbf{b}}}$ is a function of h_n. Then verify that

$$Ł(\hat{S}_{\mathbf{b}}, \mathbf{b}, f_{\hat{S}_{\mathbf{b}}}) = \prod_{i=1}^{n} [(\frac{1}{2h_n}[\hat{S}_{\mathbf{b}}((T_i(\mathbf{b})-h_n)-) - \hat{S}_{\mathbf{b}}(T_i(\mathbf{b})+h_n)])^{\delta_i}(\hat{S}_{\mathbf{b}}(T_i(\mathbf{b})))^{1-\delta_i}]. \quad (2.2)$$

We propose to estimate $\boldsymbol{\beta}$ by the value of $\mathbf{b}$, denoted by $\hat{\beta}$ that maximizes $Ł(\hat{S}_{\mathbf{b}}, \mathbf{b}, f_{\hat{S}_{\mathbf{b}}})$ over all possible $\mathbf{b}$ values. We call $\hat{\beta}$ the MSMLE of $\boldsymbol{\beta}$, because $Ł(\hat{S}_{\mathbf{b}}, \mathbf{b}, f_{\hat{S}_{\mathbf{b}}})$ is a modified semi-parametric likelihood function. We prove in the Appendix that *the likelihood function (2.2) takes on finitely many values. Thus the MSMLE exists.*

Remark 2.1. Since F_o is arbitrary, $E(\epsilon)$ may not exist. Even if it does, it is well known (see Buckley and James (1979)) that in general there is no consistent estimator of $E(\epsilon)$ $(=\alpha)$ under right censoring. Thus people formulate the censored regression model as $Y = \boldsymbol{\beta}'\mathbf{X} + \epsilon$ (see, *e.g.*, Lai and Ying (1991) and Zhang and Li (1996)) to emphasize that the main interest of the model is $\boldsymbol{\beta}$, which has a consistent estimator. One can estimate α by $\hat{\alpha}$ given in (3.1) of Section 3. $\hat{\alpha}$ is the conditional expectation based on $\hat{F}_{\hat{\beta}}$, as $\lim_{t\to\infty} \hat{F}_{\hat{\beta}}(t) < 1$ very often under right censoring.

To study properties of the MSMLE, we need the commonly-used assumption that ϵ and $(\mathbf{X}, C)$ are independent (see, *e.g.*, Zhang and Li (1996, p.2721)). We also need the following identifiability condition: $\mathrm{P}\left(rank \begin{pmatrix} 1 & \cdots & 1 \\ \mathbf{X}_1 & \cdots & \mathbf{X}_{p+1} \end{pmatrix} = p+1,\ \delta_1 = \cdots = \delta_{p+1} = 1 \right) > 0$ (see Yu and Wong (2002, p.409)). In the case of simple linear regression with complete-data it becomes $P\{\mathbf{X}_1 \neq \mathbf{X}_2\} > 0$. The assumption is weaker than the condition $E(|(\mathbf{X}_1, ..., \mathbf{X}_n)(\mathbf{X}_1, ..., \mathbf{X}_n)')|) \neq 0$ with complete-data used in Bickel (1982), where $|A|$ is the determinant of a matrix A.

3. Computing the MSMLE and its covariance matrix. Since the PLE can only take on finitely many values, so does the likelihood function. Hence, finding the MSMLE can be done by exhaustive search. A formal presentation of the algorithm is given in Appendix. In this section, we shall illustrate the algorithm via a simple data set, and discuss how to estimate the standard deviation (SD) of the MSMLE.

3.1. Covariance of the MSMLE. Simulation results suggest that the MSMLE is asymptotically normally distributed (see §4.2). If the MSMLE is also efficient, then one expects that a consistent estimate of the covariance matrix of the MSMLE is the empirical Fisher information matrix, with all the unknown variables being replaced by their consistent estimates. Our simulation results suggest that it is indeed a consistent estimator of the Fisher information matrix, which is the efficient lower bound of the covariance matrix of an estimator of the regression coefficient vector. However, our simulation results also indicate that it often underestimates the variance of the MSMLE for the moderate sample sizes in our simulation studies. It may be due to the fact that sample sizes are not large enough.

Our simulation results suggest that for moderate sample sizes the MSMLE is at least as good as the BJE in general (see §4.5). Thus we can use the estimate of the covariance matrix of the BJE, which is

$$\hat{\Sigma} = \hat{\sigma}^2(\mathbf{C}_u'\mathbf{C}_u)^{-1}, \tag{3.1}$$

where $\hat{\sigma}^2 = \frac{\int_{t \le T_{(n)}(\hat{\beta})}(t-\hat{\alpha})^2 d\hat{F}_{\hat{\beta}}(t)}{\hat{F}_{\hat{\beta}}(T_{(n)}(\hat{\beta}))}$, $\mathbf{C}_u = \begin{pmatrix} \delta_1 & \delta_1\mathbf{X}_1' \\ \cdot & \cdots \\ \delta_n & \delta_n\mathbf{X}_n' \end{pmatrix}$, $\hat{\alpha} = \frac{\int_{x \le T_{(n)}(\hat{\beta})} x d\hat{F}_{\hat{\beta}}(x)}{\int_{x \le T_{(n)}(\hat{\beta})} 1 d\hat{F}_{\hat{\beta}}(x)}$, $T_{(1)} \le \cdots \le T_{(n)}$ are order statistics of the T_i's and $\hat{F}_{\mathbf{b}}$ is the PLE of F (see Miller (1981, p. 152)). Simulation indicates that it is a conservative estimate for $n \ge 100$ unless normality holds.

An approximate confidence interval for the i-coordinate of $\boldsymbol{\beta}$, say β_i, is $\hat{\beta}_i \pm z_{\alpha/2}\hat{\sigma}_i$, where $z_{\alpha/2}$ is the $1 - \alpha/2$ percentile of $N(0,1)$ and $\hat{\sigma}_i^2$ is the i-diagonal element of the estimated covariance matrix (3.1) of the MSMLE.

Under certain regularity conditions, it can be shown that the efficient lower bound (ELB) of the covariance matrix of a semi-parametric estimator of $\boldsymbol{\beta}$ is the inverse of the minus Fisher information matrix

$$\Sigma = -\left(\frac{\partial \ln Ł}{\partial \mathbf{b} \partial \mathbf{b}'}\right)^{-1},$$

which can be shown that, under further regularity conditions,

$$\Sigma = \left(E([-\delta\frac{f_o'}{f_o}(T(\boldsymbol{\beta})) + (1-\delta)\frac{f_o}{S_o}(T(\boldsymbol{\beta}))]^2\mathbf{X}\mathbf{X}')\right)^{-1}. \tag{3.2}$$

A consistent estimator $\tilde{\Sigma}$ of Σ can be obtained by choosing appropriate kernel estimators for f_o and f_o', and replacing S_o by its PLE in expression (3.2), and approximating the expectation operation by an average. We conjecture that the MSMLE is efficient, thus $\tilde{\Sigma}$ is another estimate of the covariance matrix of the MSMLE of $\boldsymbol{\beta}$.

3.2. Computing the MSMLE. Hereafter we shall illustrate through a simple engineering data example how to find the MSMLE of a simple linear regression model and a multiple linear regression model.

Example 3.1. (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 in 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 the 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 log of stress (in KSI), M stand for log of hours to the first failure in a test pair and Y stand for log of hours to failure. Thus Y is only observed for one specimen in a pair and is censored for the other. The resulting values of (X_i, M_i, δ_i)'s are: $(\ln 44, \ln 1350, 1)$, $(\ln 37, \ln 2435, 1)$, $(\ln 34, \ln 5578, 1)$, $(\ln 32, \ln 8322, 1)$, $(\ln 27, \ln 11495, 1)$, $(\ln 44, \ln 1350, 0)$, $(\ln 37, \ln 2435, 0)$, $(\ln 34, \ln 5578, 0)$, $(\ln 32, \ln 8322, 0)$, $(\ln 27, \ln 11495, 0)$. Here, we actually fitted the original data with a log-linear model rather than a linear model, following a referee's suggestion.

Since $\frac{1}{2h_n}$ in (2.2) does not depend on $\mathbf{b}$, it suffices to maximize a simpler expression, denoted by $l(\mathbf{b})$, which is the right hand side of (2.2) with the factor $\frac{1}{2h_n}$ dropped. Note that the bandwidth h_n is a coefficient to be determined. Various suggestions have been made on the bandwidth in the literature. They are all associated with the standard deviation of the distribution of the observable random variables. Under right censoring, based on our simulation results (see §4.4), we suggest the following choice of the bandwidth h_n:

$$h_n = cn^{-1/5}, \text{ where } c \text{ equals } 2\times \text{ the sample SD of } M_i\text{'s.}$$

In this example, the sample SD of M_i's is 0.83. Thus we take $h_n = cn^{-1/5}$ in (2.1), with $n = 10$ and $c = 2 \times 0.83 = 1.66$. That is, $h_n \approx 1.05$.

Simple linear regression $Y = \beta X + \epsilon$. Now $\mathbf{b} = b$. $\hat\beta$ can be obtained in 3 steps.

Step 1. The PLE and the likelihood function (2.2) or $l(\cdot)$, as functions of b, will change values only at the solutions to the equations $T_i(b) = T_j(b) + kh_n$, $k = 0, \pm 1, \pm 2$, where $T_i(b) = M_i - bX_i$. The solutions are of the form $b_{ijk} = \frac{M_i - M_j + kh_n}{X_i - X_j}$, where $X_i \neq X_j$ for

$i < j$, $k = 0, \pm 1, \pm 2$. *e.g.*, $b_{120} = b_{670} = \frac{\ln 1350 - \ln 2435}{\ln 44 - \ln 37} \approx -3.4$ (corresponding to $T_1(b) = T_2(b)$), $b_{121} = b_{671} = \frac{\ln 1350 - \ln 2435 - 1.05}{\ln 44 - \ln 37} \approx -9.2$ (corresponding to $T_1(b) = T_2(b) - h_n$), etc. Among them there are only 50 distinct values $(= 5 \times 5 \times (5-1)/2)$. Now order these 50 b_{ijk} as $a_1 < \cdots < a_{50}$. In particular, $a_1 = -41.2$ and $a_{50} = 28.0$.

Step 2. These 50 points partition the real line into 51 open intervals. Add to the set $\{a_1, ..., a_{50}\}$ the "midpoints" of these intervals, where the "midpoints" of $(-\infty, a_1)$ and (a_{50}, ∞) are defined as $a_1 - 1$ and $a_{50} + 1$, respectively. The augmented set, denoted by B, contains 101 distinct b values. The set $\{l(b) : b \in B\}$ contains all possible values of the "likelihood" $l(\cdot)$.

Step 3. By exhaustive search, it is found that the maximum value of $l(b)$ is $l = 0.051$. There are three b values that attain $l(b) = 0.051$: $b = -6.6$, -5.5 and -4.5. We take the median $\hat\beta = -5.5$.

By formula (3.1), the estimate of the SD of the MSMLE is 0.5.

Remark 3.1. (Applying a simpler algorithm). From our simulation studies, we found that the MSMLE always falls within the set obtained in Step 1. In this example, all the three maximum points belong to $\{a_1, ..., a_{50}\}$. Thus in practice, one may skip Step 2 and treat the solutions as the MSMLE, though they may only be very close to the MSMLE.

Multiple linear regression. For illustration purposes, we also fitted to the data to the model $Y = \beta_1 X + \beta_2 (X)^2 + \epsilon$. Now $T_i(\mathbf{b}) = M_i - b_1 X_i - b_2 (X_i)^2$, where $\mathbf{b}' = (b_1, b_2)$.

Step 1. Find solutions $\mathbf{b} = (b_1, b_2)'$ of equations of the form

$$\begin{cases} T_i(\mathbf{b}) \pm h m_1 = T_j(\mathbf{b}) \\ T_k(\mathbf{b}) \pm h m_2 = T_l(\mathbf{b}), \end{cases} \text{ where } h = cn^{-1/5} \text{ and } m_1, m_2 = -2, -1, 0, 1, 2. \tag{3.3}$$

For example, for $(i, j, k, l, m_1, m_2) = (1, 2, 1, 3, 0, 1)$, we have the system of linear equations

$$\begin{cases} \ln 1350 - b_1 \ln 44 - b_2 (\ln 44)^2 = \ln 2435 - b_1 \ln 37 - b_2 (\ln 37)^2 \\ \ln 1350 - b_1 \ln 44 - b_2 (\ln 44)^2 - 1.05 = \ln 5578 - b_1 \ln 34 - b_2 (\ln 34)^2. \end{cases}$$

In particular, there are 50 $(= 5 \times \binom{5}{2})$ distinct linear equations of the form $T_i(\mathbf{b}) \pm h m_2 = T_j(\mathbf{b})$, and at most 1225 $(= \binom{50}{2})$ systems of linearly independent equations of form (3.3). Thus there are at most 1225 solutions of b.

However, the number of calculation can be reduced if we notice that there are 10 $(= \binom{5}{3})$ systems of two linearly independent equations of the form $T_i(\mathbf{b}) = T_j(\mathbf{b}) = T_k(\mathbf{b})$, where i, j and k are all distinct, and 5 $(= \binom{5}{4})$ systems of two linearly independent equations $T_i(\mathbf{b}) = T_j(\mathbf{b})$ and $T_l(\mathbf{b}) = T_k(\mathbf{b})$, where i, j, l and k are all distinct. Thus there are at most $5^2 \times (10 + 5) = 375$ distinct equations of form (3.3), and at most 375 solutions of $\mathbf{b}$.

Step 2. (Skipped, see Remark 3.1).

Step 3. Using formula (2.2), we can derive $l(\mathbf{b})$, for each $\mathbf{b}$ in the set obtained in the previous step. By exhaustive search, it is found that the maximum value of $l(\mathbf{b})$ is $l \approx e^{-4.29}$. There are two $\mathbf{b}$ values that maximize $l(\mathbf{b})$ over all possible $\mathbf{b}$ values. They are $\mathbf{b}' = (2.59, -0.87), (-0.04, -0.48)$. The median of them is not an MSMLE. We take the first one $\hat{\beta}' = (2.56, -0.87)$. The covariance matrix of $\hat{\beta}$ can be obtained by (3.1).

As we have seen from this example, the MSMLE may not be unique. There are several ways to select an estimate when the MSMLE is not unique. Under the simple linear regression model, *i.e.*, $p = 1$, we choose the MSMLE that is closest to the median of the set of all MSMLEs. The method can easily be extended to multiple regression. In particular, we choose an MSMLE that is closest to the center of the set of all MSMLEs, where the $k-$th coordinate of the center is the median of the set consisting of all the $k-$th coordinates of the MSMLEs.

4. Performance of the MSMLE. In this section, we shall present simulation results in which we investigated consistency and efficiency of the MSMLE when F_o is not discrete, as it is proved in Yu and Wong (2004) that the MSMLE is super efficient under a discrete assumption. In our simulation studies, we considered the following distribution assumptions:

(1) the case that F_o has a normal, logistic or exponential distribution;

(2) the case of complete-data or right-censored data.

We also investigated the influence of the bandwidth on the MSMLE, and the relative efficiency of the MSMLE with respect to (w.r.t.) the parametric MLE or w.r.t. the BJE under various distribution assumptions.

In our simulation, for convenience, we assumed that ϵ, X and C are independent. In each simulation study, we had 1000 replications and computed the sample mean and sample standard error (SE) of the 1000 estimates. Hereafter $Exp(\mu, \sigma)$ denotes an exponential distribution with the pdf $f(x) = \frac{1}{\sigma} e^{-[\frac{x-\mu}{\sigma}+1]} \mathbf{1}_{(x>\mu-\sigma)}$, where $\mathbf{1}_{(A)}$ is the indicator of the set A. We considered 6 different cases.

Case 1 (complete-data under a normal distribution). Suppose $\epsilon \sim N(0, 0.09)$, $X \sim U(0, 9)$ (uniform distribution) and $\beta = 1$. Computing times for sample sizes less than 200 are less than 12 minutes.

Case 2 (censored-data under a normal distribution). Suppose $\epsilon \sim N(3, 0.09)$, $X \sim U(0, 9)$ and C equals 0.5 and 39 w.p. 0.5 and 0.5, respectively. $\beta = 1$. Computing times for sample sizes less than 200 are less than 2.5 minutes.

Case 3 (censored-data under a normal distribution). Suppose $\epsilon \sim N(3, 1)$, $X \sim Exp(0, 1)$ and $C \sim Exp(1, 1)$. $\beta = 1$. Computing times for sample sizes less than 200 are less than 14 seconds.

Case 4 (complete-data under a logistic distribution). Suppose ϵ has a logistic distribution with mean 0 and variance 1, $X \sim Exp(0, 1)$ and $\beta = 1$. Computing times for sample sizes less than 200 are less than 53 seconds.

Case 5 (censored-data under a logistic distribution). Suppose ϵ has a logistic distribution with mean 0 and variance 1, $X \sim Exp(0, 1)$ and $C \sim Exp(1, 1)$. $\beta = 1$. Computing times for sample sizes less than 200 are less than 9 seconds.

Case 6 (censored-data under an exponential distribution). Suppose ϵ, C and $\mathbf{X}$ have distributions $Exp(5, 2)$, $Exp(3, 4)$ and $Exp(2, 2)$, respectively. $\beta = 1$. Computing times for sample sizes less than 200 are less than 10 seconds.

The results of these cases are summarized in the Tables 1-6. In order to investigate the influence of the bandwidth on estimation we carried out simulation studies with various bandwidths h_n in Cases 1, 3, 4 and 5. The results are summarized in Tables 1, 2 and 3. Entries in the column corresponding to $\hat{\sigma}_e$ are the sample means of the estimates of the SD using formula (3.2). Table 4 presents the relative efficiency (RE) of the MSMLE with respect to the parametric MLE (written as "MLE" in the table), where RE= SE_{MLE}/SE_{MSMLE}. The entries in the 3rd and 4th columns of the table are the sample mean (SE) of the MSMLE and the parametric MLE, respectively, in the 1000 replications. In Case 1, the parametric MLE has an explicit expression, which is the LSE. In the other cases, the MLE can only be obtained by numerical methods. In particular, we used the Newton-Raphson method to derive the MLE. In Table 2, we compare the BJE to the MSMLE under normal distributions. In Table 5, we compare the sample variance of the MSMLE to the ELB under the exponential distribution (Case 6).

The following are main observations from our simulation.

4.1. Consistency. All the 6 cases suggest that the MSMLE $\hat{\beta}$ is consistent, as the values of β are all within 2 SE's from the sample means and the SE's are decreasing in n, (see Tables 1-4).

Table 1. Simulation results on bandwidth effect under $N(\mu, \sigma^2)$.
Case 1, $\beta = 1$ and $\sigma_Y = \sqrt{4.5 + 0.09}$

n	c in h_n	MLE (SE)	MSMLE (SE)	$\hat{\sigma}_e$
32	0.5	1.000 (0.022)	1.000 (0.030)	0.020
	1		0.999 (0.028)	0.020
	2		0.998 (0.032)	0.021
	4		0.960 (0.044)	0.022
	10		0.927 (0.052)	0.025
200	0.5	1.000 (0.008)	1.000 (0.012)	0.008
	1		1.000 (0.010)	0.008
	2		1.000 (0.010)	0.008
	4		0.998 (0.019)	0.008
	10		0.958 (0.096)	0.011
300	1	1.000 (0.007)	1.000 (0.008)	0.007
	2		1.000 (0.008)	0.007
	4		1.000 (0.012)	0.007
	10		0.932 (0.038)	0.008

Table 2. Simulation results on comparison to BJE under $N(\mu, \sigma^2)$.
Case 3, $\beta = 1$ and $\sigma_Y = \sqrt{2}$

n	c in h_n	BJE (SE)	MSMLE (SE)	$\hat{\sigma}_e$
200	1	0.909 (1.140)	1.146 (0.520)	0.624
	5		1.034 (0.429)	0.617
	10		1.058 (0.697)	0.658
300	1	0.952 (0.876)	1.104 (0.407)	0.494
	5		1.013 (0.332)	0.494
	10		1.021 (0.561)	0.515

Table 3. Simulation results on various bandwidths under logistic distribution. $\beta = 1$ and $\sigma_Y = \sqrt{2}$				
n	c in h_n	MLE (SE)	MSMLE (SE)	$\hat{\sigma}_e$
Case 4. Complete-data.				
32	1	0.996(0.345)	0.988(0.612)	0.346
	2		0.964(0.494)	0.342
	4		0.973(0.431)	0.339
	10		0.989(0.496)	0.340
200	1	0.998(0.125)	0.988(0.207)	0.129
	2		1.003(0.180)	0.129
	4		0.995(0.157)	0.129
	10		0.999(0.152)	0.129
300	1	0.991(0.101)	0.989(0.193)	0.105
	2		0.998(0.151)	0.105
	4		0.998(0.116)	0.105
	10		0.986(0.111)	0.105
Case 5. Censored-data.				
32	1	1.045 (0.472)	1.564 (1.527)	0.802
	2		1.274 (1.149)	0.747
	4		1.155 (0.993)	0.733
	10		1.598 (1.535)	0.837
100	1	1.010 (0.264)	1.128 (0.561)	0.387
	2		1.053 (0.450)	0.382
	4		0.991 (0.432)	0.381
	10		1.092 (0.731)	0.401
200	1	1.002 (0.186)	1.044 (0.336)	0.269
	2		1.022 (0.283)	0.267
	4		0.979 (0.274)	0.267
	10		1.027 (0.450)	0.273
400	1	1.005 (0.131)	1.029 (0.232)	0.190
	2		1.011 (0.194)	0.190
	4		0.991 (0.184)	0.190
	10		1.009(0.267)	0.188

4.2. Asymptotic normality. Our simulation results suggest that the MSMLE is asymptotically normally distributed. In Figure 4.1, we plot the QQ-plot of the MSMLEs versus $N(0, 1)$. In each of the six QQ-plots, there are 1000 MSMLE estimates. The error distributions are normal, logistic and exponential, respectively, the sample sizes of the top three QQ-plots are 50, 100 and 100, respectively, and the sample sizes of the bottom three QQ-plots are 200, 400 and 400, respectively. It is seen that the QQ-plot of the MSMLEs from the normal error distribution appears quite linear, even with a sample size of 200, as we expected. The other two QQ-plots also become linear as n becomes large.

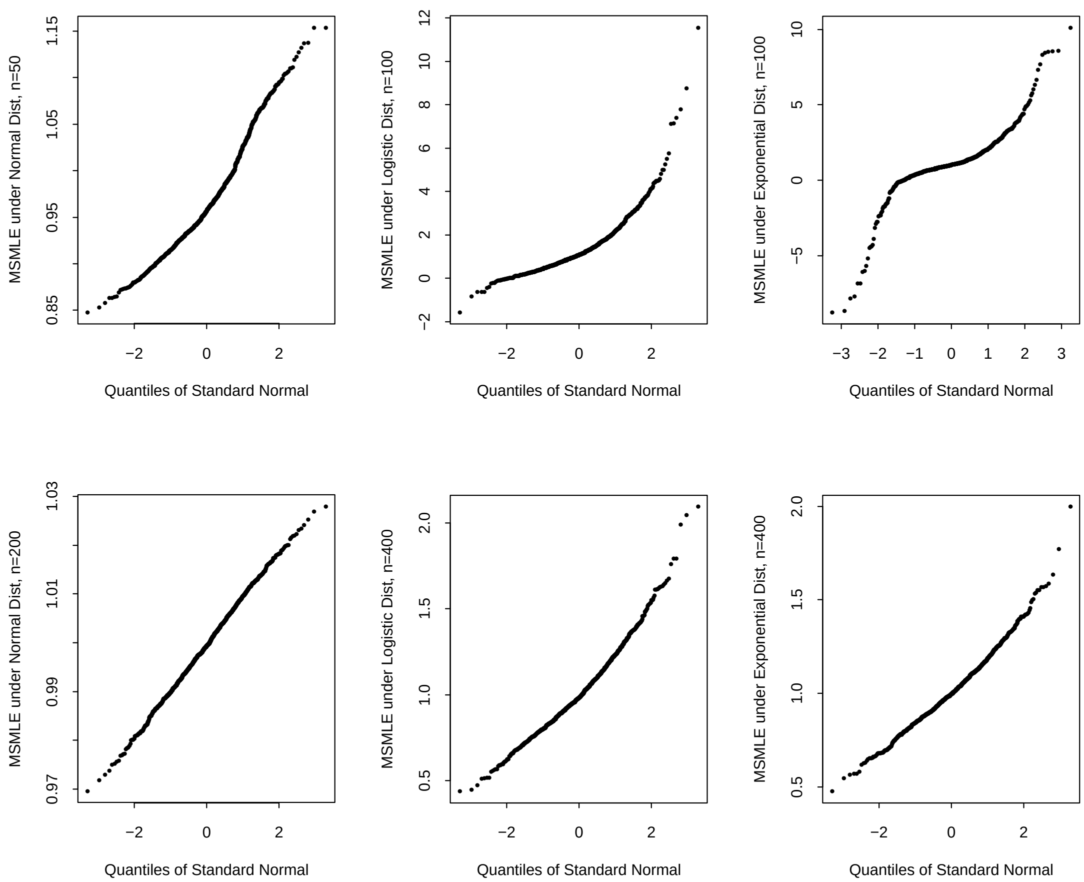


Fig.4.1. QQ-plot of the MSMLE v.s. $N(0,1)$

4.3. Efficiency. Simulation results in Table 4 are inconclusive on whether the MSMLE is asymptotically efficient. None of the relative efficiencies displayed in Table 4 are 1 for the largest sample sizes displayed. However, they all appear to be on the increase. In Case 6, the efficient lower bound (ELB) of the estimator of β is $Var(\epsilon)/(n \cdot Var(X)) = 2.5^2/n$. It is seen from Table 5 that when $n = 800$, the MSMLE practically attains the ELB. Note that in Table 5, $\hat{\sigma}^2_{\hat{\beta}}$ stands for the sample variance in the simulation.

Table 4. Simulation results on relative efficiency of MSMLE w.r.t MLE. $\beta = 1$

case	sample size	MLE (SE)	MSMLE (SE)	RE
1	32	1.000 (0.0220)	0.998 (0.0300)	0.67
	200	1.000 (0.0083)	1.000 (0.0094)	0.88
	300	1.000 (0.0072)	1.000 (0.0081)	0.89
2	32	1.000 (0.030)	0.995 (0.042)	0.71
	200	1.000 (0.011)	0.994 (0.013)	0.85
	300	1.000 (0.009)	0.999 (0.010)	0.90
4	32	0.996 (0.345)	0.973 (0.431)	0.80
	200	0.998 (0.124)	0.995 (0.152)	0.82
	400	1.022 (0.090)	1.017 (0.103)	0.87
5	32	1.045 (0.472)	1.155 (0.993)	0.48
	200	1.002 (0.186)	0.979 (0.274)	0.68
	400	1.005 (0.131)	0.991 (0.184)	0.71

Table 5. Comparison between the SE of the MSMLE and the ELB

$$
\begin{array}{lccccccc}
n: & 32 & 100 & 200 & 400 & 800 & & \sqrt{n}\cdot ELB \\
\sqrt{n}\hat{\sigma}_{\hat{\beta}}: & \Big(10.363 & 5.707 & 3.875 & 3.132 & 2.503 \Big) & \cdots & \Big(\quad 2.5 \quad \Big)
\end{array}
$$

It suggests that the MSMLE may be efficient in the exponential distribution case. Also the simulation results indicate that the convergence rate of the MSMLE is $\sqrt{n}$.

4.4. Bandwidth impact. The bandwidth in (2.1) may make a difference for moderate sample sizes. As suggested in Härdle (1990, p.59 or p.91), the optimal bandwidth with complete-data (in terms of the mean squared error of the estimator) is $h_n = cn^{-1/5}$, where $c \approx \min\{\hat{\sigma}, iqr\}$, $\hat{\sigma}$ is the SE of the observations and "iqr" is the inter-quartile-range. Under the linear regression model, our observations are the (Y_i, X_i)'s. It may be appropriate to replace $\hat{\sigma}$ and "iqr" by the SD of the Y_i's. For Cases 1 and 2, the SD of Y is $\sqrt{\frac{9^2}{12} + 0.09}$ and we selected $h_n = cn^{-1/5}$, with $c = 0.5, 1, 2, 4, 10$. For Cases 4 and 5, the SD of Y is

$\sqrt{2}$ and we selected $h_n = cn^{-1/5}$, with $c = 1, 2, 4, 5, 10$. The influence of the bandwidth on the MSMLE is displayed in Table 1, 2 and 3. From these tables we notice that (1) in general, the MSMLE is consistent for each h_n selected in the tables; (2) as far as efficiency is concerned, if n is large enough, $h_n = c_o \sigma_\epsilon n^{-5}$ with c_o between 1 and 3 (which corresponds to c around 2 and 5) may have some optimality (empirically). We suggest choosing $c_o = 2$.

4.5. Justification of the covariance estimator in (3.1). If the MSMLE is efficient, then a naive estimator of $\sigma^2_{\hat{\beta}}$ is the estimate of the ELB of the variance given by Expression (3.2). The performance of this estimate is studied via $\hat{\sigma}_e$ in Tables 1, 2 and 3. The estimator maybe appropriate for large sample sizes, but often underestimates for sample sizes ≤ 100. Thus one may consider a conservative estimator.

The BJE is a semi-parametric estimator of $\boldsymbol{\beta}$ in the linear regression model. It is well known that if $\epsilon \sim N(\mu, \sigma^2)$ such as in Cases 1, 2 and 3, the BJE is efficient. From Table 1, we note that the BJE, which is the LSE and hence the MLE, is close to the MSMLE when sample sizes are large. However, in Case 3, which is also under a normal distribution, the SE of the BJE is twice the SE of the MSMLE for the sample sizes displayed in the table. We also compared the relative efficiency of the MSMLE w.r.t. the BJE in three uniform distribution (unif) cases and one exponential distributions (expon.) case (see Table 6). The results suggest that the MSMLE is as good as the BJE. Since Expression (3.1) is the estimate of the covariance matrix of the BJE, one expects that it is a conservative estimate of the covariance matrix of the MSMLE. This is confirmed by our simulation results (but were not displayed).

Table 6. Estimates of the relative efficiency of MSMLE $\hat{\beta}$ to the BJE.

case :	1	3	5	6	*others*[2]
$\frac{\hat{\sigma}^2_{\tilde{\beta}_{BJE}}}{\hat{\sigma}^2_{\hat{\beta}}}$:	?[1]	1.0	1.1	3.7	> 1
F_o :	*normal*	*normal*	*logistic.*	*expon.*	*unif.*

Note:

[1] In Case 1, *i.e.*, under the normal distribution with complete-data, the parametric MLE is the LSE, which is also the BJE. We only carried out simulation up to sample size $n = 300$. The relative efficiency is 0.89 and is still increasing quite significantly (see Table 4). Thus it is inconclusive in this case.

[2] We carried out simulation with several other distributions including uniform distributions to compare the MSMLE with the BJE. The results are summarized in the column "others".

5. Applications. We compare the MSMLE to the BJE using two real data sets in this section. The BJE is based on the least squares principle (see Buckley and James (1979)) and is also an M-estimator (see Zhang and Li (1996)). We do not compare the MSMLE to other approaches for the following reasons.

1. The least squares approach is the common approach in linear regression with complete data, though it is not efficient under the semi-parametric set-up.
2. There exist efficient M-estimators under the semi-parametric set-up (see Zhang and Li (1996)). However, the M-estimation approach depends on the kernel selection and its applications with real data sets have not been seen in the literature except for the special case of the BJE as far as we know.
3. The other estimators discussed in Section 1 are not efficient.

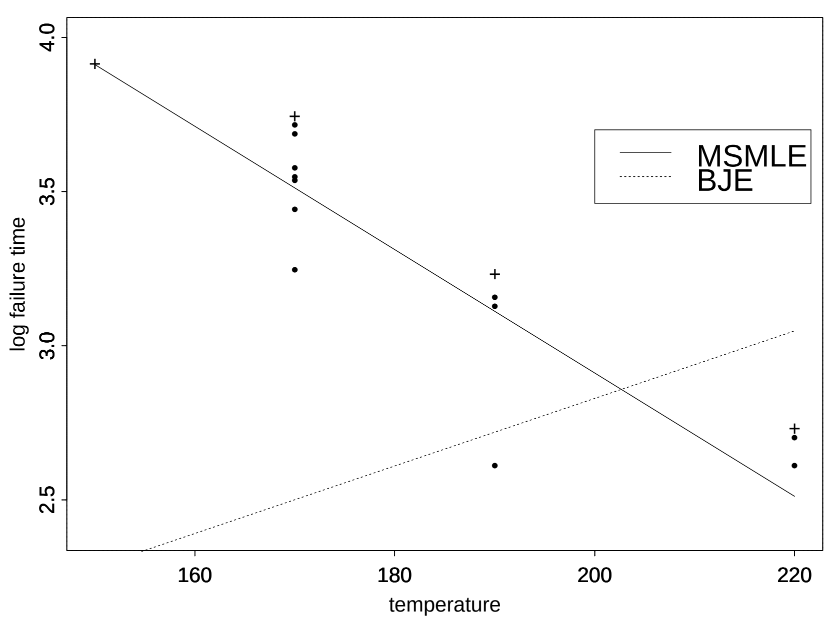
In the first data set of size 40, the MSMLE looks reasonable, whereas the BJE appears pretty bad, the sign of the BJE of the slope is not even right in view of the scatter plot of the data set. In the second data set of size 99, the normality maybe valid and the two estimates do not differ greatly. In order to remind readers that the normality of the error distribution is not always valid even under the complete-data case, we present a simulation studies in which the LSE is worse than the MSMLE in the case of complete-data with both small and large sample sizes.

Example 5.1. (Insulation data (Nelson 1973)). To evaluate a new Class-B insulation for electric motors, temperature-accelerated life testing was conducted on 40 motorettes. The main purpose was to estimate the distribution of insulation at the design temperature of 130^oC. Ten motorettes were put on test at each of four temperatures (150^oC, 170^oC, 190^oC, and 220^oC). Let X be the temperature (in ^{o}C) and Y the logarithm of hours to failure of an insulation at temperature X. The data are plotted in Figure 5.1.

The MSMLE of (α, β) is $(0.923, -0.022)$, with an SE $(0.515, 0.003)$. Thus it is significant. The standard procedure for censored regression in the literature is the BJE. For this data set, the BJE of (α, β) derived using the algorithm given in Buckley and James (1979) is $(8.951, 0.011)$ with their standard errors $(1.629, 0.008)$ estimated by (3.1).

From the scatter plot in Figure 5.1, it is seen that the BJE is absurd, though the BJE is not significant. In this data set of moderate sample size, our new approach has a compelling advantage over the least squares principle and over the M-estimation approach.

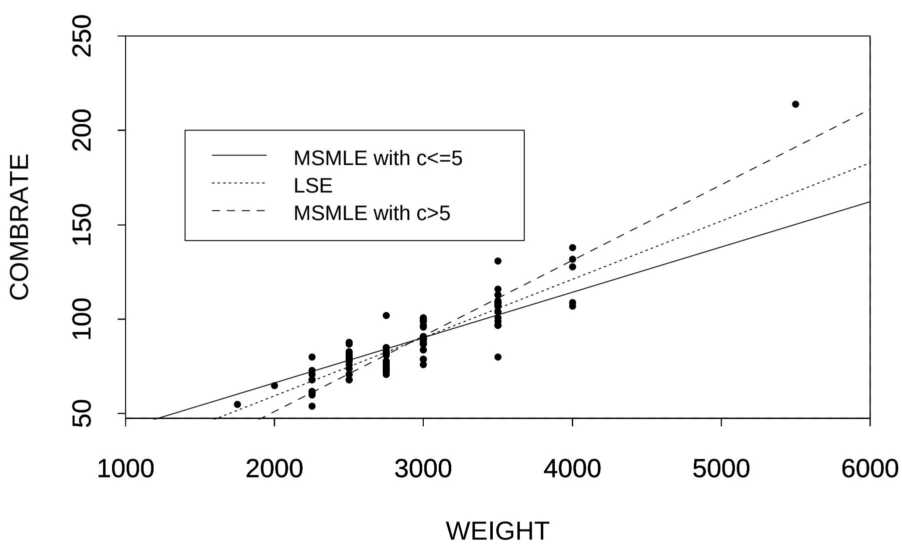
Fig. 5.1. MSMLE vs. BJE For Insulation Data



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

Example 5.2. (Auto Data). This data set consists of 99 observations taken from a larger data set of automobile fuel consumption data published by Transport Canada in 1985. One can also find it in Jobson (1991). Data are uncensored. Y is the fuel consumption or combustion rate (COMBRATE) and X is the weight of the car. For this data set, the sample standard deviation of Y_i's is 21.1. The MSMLE of β is 0.024 with a standard error 0.002 if the coefficient c in the bandwidth $h = cn^{-1/5}$ is between 0.1 and 5, and is 0.040 with a standard error 0.002 if the coefficient c is between 6 and 100. According to our simulation results, we choose $c = 42$ ($\approx 2\times$ sample SD of Y) and thus we choose the estimate 0.040 with a standard error 0.002. From Figure 5.2, it is seen that the choice of $c = 42$ gives a better estimate than the other choice. The MSMLE of α is -28.857 with a standard error 5.162. The standard procedure for this data set is the LSE. The LSE of (α, β) for this data set is $(-13.770, 0.035)$ with standard errors $(4.926, 0.002)$. Notice that these two estimates are not as drastically different as in Example 5.1. A Q-Q residual plot reveals that the error distribution is very close to a normal distribution and thus one expects that the LSE is efficient and so is the MSMLE.

Fig. 5.2. Regression Line for Auto Data



Note that in the case of complete data, the LSE is not efficient in general. The following is a simulation study to remind readers that even for a small sample size, the variance of the LSE maybe much larger than the variance of the MSMLE.

Simulation results. Case 7. Suppose ϵ and $\mathbf{X}$ have distributions $Exp(3,1)$ and $Exp(0,1)$, respectively. $\beta = 1$. The simulation results of the LSE and the MSMLE with sample sizes 10, 32, 100 and 200 are summarized in Table 7. It is seen from Table 7 that the relative efficiency of the LSE w.r.t. the MSMLE is getting to 0.64 and the standard error of the LSE is already larger than the standard error of the MSMLE under a small sample size of 10, if the error distribution is not normal.

Table 7. Simulation results on relative efficiency of LSE w.r.t MSMLE.
$\beta = 1$

sample size	LSE (SE)	MSMLE (SE)	RE ($=SE_{MSMLE}/SE_{LSE}$)
10	0.972 (0.455)	1.017 (0.454)	<1
32	1.004 (0.201)	1.006 (0.152)	0.76
100	0.986 (0.099)	0.998 (0.068)	0.68
200	0.999 (0.067)	0.995 (0.043)	0.64

6. Concluding Remarks. We propose a new estimator, the MSMLE, using the likelihood principle. Simulation results suggest that the MSMLE is consistent and asymptotically normally distributed. Our simulation results suggest that the MSMLE is more efficient than the BJE, a common procedure in this problem. In particular, simulation studies suggest that (1) the MSMLE is efficient if the error distribution is exponential and (2) the relative efficiency of the MSMLE to the parametric MLE ranges from 0.71 to 0.9 under the logistic distribution and normal distributions for the sample sizes up to 300-400.

The MSMLE depends on the bandwidth h_n. Selection of optimal bandwidth relies on what criterion is adopted. Possible criteria may include minimizing the mean squared error in f_o, minimizing the variance of $\hat{\beta}$ or maximizing $l(\mathbf{b})$ in (2.2) over all h_n and all $\mathbf{b}$.

As one expects, there does not exist a unique choice of bandwidth that is always optimal. Simulation studies also suggest that the bandwidth $h_n = 2\times$ sample SD of $Y \times n^{-1/5}$ may be appropriate for large sample sizes and the practical examples in Section 5 suggest that this choice works well in practice.

Appendix. We shall prove the existence of the MSMLE here. Moreover, we shall present the algorithm for the MSMLE formally and prove that it will find the MSMLE.

Let $T_{(1)}(\mathbf{b}) \le \cdots \le T_{(n)}(\mathbf{b})$ be order statistics of the $T_i(\mathbf{b})$'s and let $\delta_{(i)}(\mathbf{b})$ be the δ_j associated with $T_{(i)}(\mathbf{b})$. Note that

$$\hat{S}_{\mathbf{b}}(t) = \prod_{T_{(i)}(\mathbf{b})\le t} (1 - \frac{\delta_{(i)}(\mathbf{b})}{n-i+1}), \qquad (A.1)$$

and thus given t, $\hat{S}_{\mathbf{b}}(t)$ is constant on the set on which the vector $(\delta_{(1)}(\mathbf{b}), ..., \delta_{(n)}(\mathbf{b}))$ is a constant. Given a sample and the h_n, there are at most finitely many distinct values of $\hat{S}_{\mathbf{b}}(T_{(j)}(\mathbf{b}))$, 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), the likelihood

$$Ł(\mathbf{b}) \propto \prod_{i=1}^{n} \left[[\hat{S}_b((T_i(b) - h_n)-) - \hat{S}_b(T_i(b) + h_n)]^{\delta_i} (\hat{S}_b(T_i(b)))^{1-\delta_i}\right], \qquad (A.2)$$

thus the likelihood takes on finitely many values. Consequently, the MSMLE exists. □

To motivate an algorithm for the MSMLE, we make use of the following notation. Since we have $T_i(\mathbf{b}) \pm h_n$ and $T_i(\mathbf{b})$ in the likelihood (2.2) or (A.2), for convenience, we write $v_{3i-2}(\mathbf{b}) = T_i(\mathbf{b}) - h_n$, $v_{3i-1}(\mathbf{b}) = T_i(\mathbf{b})$, $v_{3i}(\mathbf{b}) = T_i(\mathbf{b}) + h_n$, $i = 1, ..., n$. Let $r(v_i(\mathbf{b}))$ be the rank of $v_i(\mathbf{b})$ among $v_1(\mathbf{b}), ..., v_{3n}(\mathbf{b})$. Since the PLE only depends on the ranks of $T_i(\mathbf{b})$'s (see (A.1)), by (A.2), likelihood (2.2) is constant on the set $B_{\mathbf{r}}$, where $B_{\mathbf{r}} = \{\mathbf{b} : (r(v_1(\mathbf{b})), ..., r(v_{3n}(\mathbf{b}))) = \mathbf{r})\}$, for each of the possible values $\mathbf{r} = (r_1, ..., r_{3n})$ of $(r(v_1(\mathbf{b})), ..., r(v_{3n}(\mathbf{b})))$, the ranks of the $3n$ $v_i(\mathbf{b})$'s. The above discussion suggests the algorithm for the MSMLE:

Algorithm 1. For each value $\mathbf{r}$ of the ranks of the $3n$ $v_i(\mathbf{b})$'s, one can find a point $\mathbf{b_r}$ in $B_\mathbf{r}$. There are at most finitely many possible different values of $\mathbf{r}$, and thus there are at most finitely many different values of the likelihood function, say $\L(\mathbf{b_r})$ for all $\mathbf{b_r}$ we have. Hence, finding the MSMLE can be done by exhaustive search. □

The key to the algorithm is to identify these $B_\mathbf{r}$'s. For different $\mathbf{r}_1$ and $\mathbf{r}_2$, the ranks of $\{v_k(\mathbf{b}_{\mathbf{r}_1}), i = 1, ..., 3n\}$ and $\{v_k(\mathbf{b}_{\mathbf{r}_2}), i = 1, ..., 3n\}$ are different by definition. It is obvious that $v_k(\mathbf{b})$'s change their ranks only at the solutions to equations $v_i(\mathbf{b}) = v_j(\mathbf{b})$. By the definition of $v_i(\mathbf{b})$, these equations yield the following 3 types.

(1) $T_i(\mathbf{b}) + kh_n = T_j(\mathbf{b}) + kh_n$, $k = 0, \pm 1$ (*i.e.* $M_i - \mathbf{b}X_i = M_j - \mathbf{b}X_j$);

(2) $T_i(\mathbf{b}) - h_n = T_j(\mathbf{b}) + kh_n$, $k = 0, 1$;

(3) $T_i(\mathbf{b}) + h_n = T_j(\mathbf{b}) + kh_n$, $k = 0, -1$.

These are the equations of the form $T_i(\mathbf{b}) = T_j(\mathbf{b}) + kh_n$, $k = 0, \pm 1, \pm 2$. Finding the solutions to these equations will help to identify the sets $B_\mathbf{r}$.

We first explain how to find these $B_\mathbf{r}$ in the case of simple linear regression. Now the space for b is the whole real line. The solution to the equation $T_i(b) = T_j(b) + kh_n$ (i.e., $M_i - bX_i = M_j - bX_j + kh_n$), if exists, is either a singleton point (if $X_i \neq X_j$) or the whole real line (if $(X_i, M_i) = (X_j, M_j)$). Let $a_1 < \cdots < a_m$ be all the distinct solutions to the equations that have a unique solution. These points partition the real line into $m + 1$ open intervals and m singleton sets. That is, there are $2m + 1$ $B_\mathbf{r}$'s. The likelihood function (2.2) is constant on each of these subsets. The MSMLE can be found in three steps.

1. Compute the points $b = \frac{M_i - M_j + kh_n}{X_i - X_j}$, for all $X_i \neq X_j$, $i < j$ and $k = 0, \pm 1, \pm 2$. Order the distinct elements of these b values as $a_1 < \cdots < a_m$;
2. Add to the above set the points $a_1 - 1$, $a_m + 1$ and points $\frac{a_{i-1} + a_i}{2}$, $i = 2, ..., m$.
3. The maximum value of likelihood is obtained within the set of b values obtained in Steps 1 and 2. Moreover, if b is an MSMLE and $b \in (a_i, a_{i+1})$, then each point in

(a_i, a_{i+1}) is also an MSMLE. □

From our simulation studies we noticed that the MSMLE always falls in the set obtained in Step 1. Thus one may skip Step 2 in practice. This simplifies the search for $\mathbf{b_r}$. In multiple regression, in order to find the points $\mathbf{b}$ in Step 1, one can solve one by one the $\binom{3n}{p}$ systems of p linear equations $v_{i_k}(\mathbf{b}) = v_{j_k}(\mathbf{b})$, $k = 1, ..., p$, if they are linearly independent. This can be implemented easily in computer, though the computation load of this approach is large. We expect that for $p \leq 3$ and on a Pentiem 3 computer, it is feasible to compute the MSMLE with a sample of size around 500.

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

References.

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.

Bickel, P.J. (1982). On adaptive estimation. *Ann. Statist.* 10 647-671.

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.

Härdle W. (1990). Smoothing techniques, with implementation in S. *Springer-Verlag*, N.Y.

Jobson J.D. (1991). Applied multivariate data analysis. Vol. I: regression and experimental design. Springer. p.589.

Huber, P.J. (1964). Robust estimation of a location parameter". *Ann. Math. Statist.* 35 73-101.

Ireson, M. J. and Rao, P. V. (1985). Interval Estimation of Slope with Right-Censored Data. *Biometrika* 72 601-608.

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.

Koul, H. and Susarla, V. (1983). Adaptive estimation in linear regression. *Statist. and Decisions* 11 379-400.

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

Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* 74 301-309.

Miller, R.G. (1981). Survival Analysis. *Wiley.*

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.

Nelson, W. (1973). Analysis of residuals from censored data. *Technometrics* 15 697-715.

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

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.

Yu, Q.Q. and Wong, G.Y.C. (2001). How to find all Buckley-James estimates instead of just one ? *Journal of Statistical Computation and Simulation*, 72 451-460.

Yu, Q.Q. and Wong, G.Y.C. (2002). Asymptotic properties of a modified semi- parametric MLE in linear regression analysis with right-censored data. *Acta Mathematica Sinica*, 18 405-416.

Yu, Q.Q. and Wong, G.Y.C. (2003). Asymptotic properties of the generalized semi- parametric MLE in linear regression. *Statistica Sinica*, 13 311-326.

Yu, Q.Q. and Wong, G. Y. C. (2004). Asymptotic properties of a modified semi-parametric MLE in linear regression with right-censored data under a discrete assumption. (Technical report).

Zhang, C.H. and Li, X. (1996). Linear regression with doubly censored data. *Ann. Statist.* 24 2720-2743.