⊠연구논문

Robust Regression for Right-Censored Data

Chul-Ki Kim
Dept. of Statistics, Ewha Womans University

Abstract

In this paper we develop computational algorithms to calculate M-estimators of regression parameters from right-censored data that are naturally collected in quality control. In the case of M-estimators, a new statistical method is also introduced to incorporate concomitant scale estimation in the presence of right censoring on the observed responses. Furthermore, we illustrate this by simulations.

Key Words : EM algorithm, M-estimators, Metrically Winsorized residual, Survival analysis

1. Introduction

Consider the linear regression model

$$y_j = \alpha + \beta^T x_j + \varepsilon_j \quad (j = 1, 2, \cdots), \tag{1.1}$$

where the ε_j are i.i.d. random variables and the x_j are independent $p \times 1$ random vectors independent of $\{\varepsilon_n\}$. Taking the location parameter α in (1.1) to be a minimizer of the function $R(a) = E\rho(y_j - \beta x_j - a)$, Huber's (1973) M-estimators $\widehat{\alpha}$, $\widehat{\beta}$ of α, β based on $(x_1, y_1), \cdots, (x_n, y_n)$ are defined as a solution vector to the minimization problem

$$\sum_{j=1}^{n} \rho(y_j - a - b^T x_j) \Big(= \int \rho(y-a) dF^*_{n,b}(y) \Big) = \min!, \tag{1.2}$$

where $F^*_{n,b}$ is the empirical distribution constructed from $y_j(b)=y_j-b^Tx_j$, $j=1,\cdots,n$. When ρ is differentiable, the M-estimators $\widehat{\alpha}$ and $\widehat{\beta}$ can also be defined as a solution of the system of estimating equations

$$\sum_{j=1}^{n}\rho'(y_j-a-b^Tx_j)=0, \quad \sum_{j=1}^{n}x_j\rho'(y_j-a-b^Tx_j)=0. \tag{1.3}$$

Choosing ρ suitably gives estimators that have desirable robustness properties. A well-known robust choice of ρ is Huber's score function

$$\rho'(u)=u \text{ if } |u|\le c \text{ and } \rho'=\pm c \text{ if } |u|>c, \tag{1.4}$$

where c represents some measure of the dispersion of F. Using (1.4) in (1.2) is tantamount to applying the method of least-squares to "metrically Winsorized residuals", cf. Huber (1981, p.180).

Suppose that the responses y_j in (1.1) are not completely observable due to right censoring by random variable c_j such that $-\infty < c_j < \infty$. Let $\widetilde{y}_j=y_j\wedge c_j$ and $\delta_j=I(y_j\le c_j)$, where we use $\wedge$ and $\vee$ to denote minimum and maximum respectively. The data, therefore, consist of n observations

$$(x_i,\widetilde{y}_i,\delta_i), \quad i=1,\cdots,n. \tag{1.5}$$

Unless otherwise stated, it will be assumed that (c_j,x_j) are independent of the sequence $\{\varepsilon_n\}$. The "censored regression model" is widely studied in quality control, biostatistics and economics, for which Buckley and James (1979) proposed a method to modify the estimating equations (1.3) in the case $\rho(u)=u^2/2$ and Lai and Ying (1991b, 1994) recently provided generalization and asymtotic theories of the rank approach and M-estimators for regression analysis with right-censored data.

For complete data, M-estimators of α and β involve much lower computational complexity than rank estimators and efficient algorithms for computing them have been developed (cf. Huber 1977, 1981). In chapter 3, we augment and modify these algorithms for computing M-estimators of regression parameters from right-censored data. In the cas $p=1$ and $\rho(u)=u^2/2$, our computational method is an improvement of that of Buckley and James (1979) whose iterative scheme often settles down to oscillating between two values. Our somewhat more thorough search procedure enables us to get around these difficulties. As indicated by Lai and Ying (1994, p.1239), the M-estimators based on right-censored data developed in their paper and by previous authors have not considered concomitant scale estimation which is basic to the idea of metrically Winsorized residuals underlying Huber's score function (1.4). We address this important issue in chapter 3. Furthermore, by making use of the simple preliminary estimate of the regression parameters in Chapter 2, we develop a computationally simpler version of the adjustments for small risk set sizes proposed by Lai and Ying (1994).

2. Initializing with Simple Preliminary Estimator

The M-estimator in Chapter 3 involves iterative search which has to be initialized somewhere. We propose to initialize with a simple preliminary estimator $(\tilde{\alpha}, \tilde{\beta})$ recently introduced by Gross and Lai (1995). Given the observed data (1.5), partition the range of x_i-values into $m(\geq 1)$ subsets $X_1, \cdots, X_m$, thereby stratifying the data, with the kth stratum corresponding to x_i-values in X_k. From the n_k 3-tuples $(x_i, \tilde{y}_i, \delta_i)$ of observed data within the kth stratum (i.e. $x_i \in X_k$ and $\sum_{k=1}^{m} n_k = n$), define

$$\#_k(y) = \sum_{i=1}^{n} I(x_i \in X_k, y \leq \tilde{y}_i), \quad d_k(y) = \sum_{i=1}^{n} I(x_i \in X_k, \tilde{y}_i = y, \delta_i = 1),$$
$$\hat{S}_k(y) = \prod_{i: x_i \in X_k, \tilde{y}_i \leq y, \delta_i = 1} \{1 - d_k(\tilde{y}_i) I(\#_k(\tilde{y}_i) \geq s) / \#_k(\tilde{y}_i)\}, \tag{2.1}$$

in which $s \geq 2$ is some prescribed lower bound on the risk set size $\#_k(\cdot)$ to avoid instabilities in the product-limit curve $\hat{S}_k(\cdot)$, using an idea of Lai and Ying (1991a). The preliminary estimator $(\tilde{\alpha}, \tilde{\beta})$ of (α, β) is defined by the linear equations

$$\sum_{k=1}^{m} \sum_{i: x_i \in X_k} \delta_i (\tilde{y}_i - a - b^T x_i) n_k \hat{S}_k(\tilde{y}_i) I(\#_k(\tilde{y}_i) \geq s) / \#_k(\tilde{y}_i) \doteq 0 \tag{2.2}$$

$$\sum_{k=1}^{m} \sum_{i:x_i \in X_k} \delta_i x_i (\tilde{y}_i - a - b^T x_i) n_k \hat{S}_k(\tilde{y}_i) I(\#_k(\tilde{y}_i) \geq s) / \#_k(\tilde{y}_i) = 0. \tag{2.3}$$

Note that (2.2) and (2.3) are the usual normal equations of weighted least-squares estimates of α and β based on $\{(x_i, \tilde{y}_i): 1 \leq i \leq n\}$ with weights

$$w_i = \delta_i n_k \hat{S}_k(\tilde{y}_i) I(\#_k(\tilde{y}_i) \geq s) / \#_k(\tilde{y}_i) \quad (x_i \in X_k). \tag{2.4}$$

The weights are introduced to adjust for the bias caused by censoring in the usual (unadjusted) least-squares estimates from $\{(x_i, \tilde{y}_i): 1 \leq i \leq n\}$. The rationale behind these weights is explained in Gross and Lai (1995). Briefly, suppose that $\{(x_i, \tilde{y}_i, \delta_i) : 1 \leq i \leq n\}$ are observable right-censored data. By choosing the strata suitably so that the x_i-values do not change much within X_k, we can assume, at least approximately, that C_j is conditionally independent of (x_j, y_j), given $x_j \in X_k$, and that all the y_j with $x_j \in X_k$ have the same conditional survival function which is estimated by $\hat{S}_k$. It then follows that the weights w_i in (2.4) have the effect of making the preliminary estimate $(\tilde{\alpha}, \tilde{\beta})$ consistent under reasonable assumptions, cf. Gross and Lai (1995).

3. M-Estimators and Concominant Scale Estimation

Throughout the sequel we shall use the following notation for the right-censored data (1.5). Let $\tilde{y}_i(b) = \tilde{y}_i - b^T x_i$ and

$$\begin{aligned} N(b, u) &= \sum_{i=1}^{n} I(\tilde{y}_i(b) \geq u), \quad \triangle(b, u) = I(\tilde{y}_i(b) = u, \delta_i = 1), \\ \hat{F}_b(u \mid v) &= 1 - \prod_{i:v < \tilde{y}_i(b) \leq u, \delta_i = 1} \{1 - \triangle(b, \tilde{y}_i(b)) / N(b, \tilde{y}_i(b))\}. \end{aligned} \tag{3.1}$$

The notation $\hat{F}_b(u \mid v-)$ will be used to denote (3.1) in which " $v < \tilde{y}_i(b)$" is replaced by " $v \leq \tilde{y}_i(b)$". The function $\hat{F}_b(u \mid -\infty)$ is the product-limit estimate of the common distribution function $F(u)$ of the $\varepsilon_j + a$ in (1.1). Put $\psi = \rho'$ in (1.3). To extend (1.3) to the right-censored data (1.5), Lai and Ying (1994) applied "missing information principle" which leads to replacing (1.3) by the estimating equations

$$\sum_{i=1}^{n} \psi_i^*(a, b) = 0, \quad \sum_{i=1}^{n} x_i \psi_i^*(a, b) = 0, \tag{3.2}$$

where

$$\psi_i^*(a, b) = \delta_i \psi(\widetilde{y}_i(b) - a) + (1 - \delta_i) \int_{u > \widetilde{y}_i(b)} \psi(u - a)\, d\widehat{F}_b(u \mid \widetilde{y}_i(b)) \tag{3.3}$$

cf. (2.24) and (2.26) of Lai and Ying (1994), where it is also noted that the first equation in (2.24) there gives $\int_{-\infty}^{\infty} \psi(u-a)\, d\widehat{F}_b(u \mid -\infty) = 0$ Hence for the censored regression model the estimating equations (3.2) reduce to those of Buckley-James (1979) in the case $\psi(u) = u$ and to those of Ritov (1990) for general ψ.

3.1 Incorporation Scale and Dampening Instability due to Small Risk Set Sizes

For complete data, a robust choice of $\psi(=\rho')$ is Huber's score function (1.4) which involves some scale parameter c. Accordingly estimating equations of robust M-estimators modefy (1.3) as

$$\sum_{j=1}^{n} \sigma \psi\left(\frac{y_j - a - b^T x_j}{\sigma}\right)\begin{pmatrix} 1 \\ x_j \end{pmatrix} = 0, \tag{3.4}$$

in which σ is an unknown scale parameter to be estimated from the estimating equation

$$\sum_{j=1}^{n} \chi(\sigma^{-1}(y_j - a - b^T x_j)) = 0, \tag{3.5}$$

cf. Sections 7.7 and 7.9 of Huber (1981). In particular the choice $\chi(u) = \text{sign}(|u|-1)$ in (3.5) estimates σ by the median of absolute residuals, i.e., $\text{med}_{j\le n} |\ y_j - a - b^T x_j\ |$ .With this standardization, Huber's score function takes the form $\psi(x) = \min\{1, |x|\}$. Unlike (3.4)-(3.5), the estimating equations (3.2)-(3.3) do not incorporate scale. Moreover, the product-limit estimate $\widehat{F}_\beta(u \mid v)$ may be quite unstable when v is near $\max_{i\le n} \tilde{y}_i(\beta)$, cf. Lai and Ying (1991a). Hence it is desirable to down-weight such v in (3.3). This is done in Section 4.1 Lai and Ying (1994) by introducing smoothing kernels to dampen the instability due to small risk set sizes. These smoothing kernels substantially increase the computational complexity of the modified estimating equations although they are important in ensuring such modifications to be "smooth" (in some average sense) in b. Although in principle one can apply the "missing information principle" in Section 2 of Lai and Ying (1994) to extend (3.4) and (3.5) to right-censored data in the same way as Lai and Ying (1994) extended (1.3) to the form (3.2)-(3.3), this approach leads to simultaneous equations which are difficult to solve numerically because standard iterative schemes may be numerically unstable. We can avoid this difficulty by using a separate scale estimate $\breve{\sigma}$ based on $\widehat{F}_{\breve{\beta}}$, where $\breve{\beta}$ is the preliminary estimate of β introduced in Chapter 2. Moreover, instead of down-weighting the extreme order statistics of $\tilde{y}_k(b)$ via an elaborate smoothing scheme as Lai and Ying (1994), we propose to trim away the extreme order statistics of $\tilde{y}_k(\breve{\beta})$, which do not involve b, so that we do not need the elaborate scheme to ensure trimming smoothly in b.

Ordering the $\tilde{y}_k(\breve{\beta})$ as $\tilde{y}_{[1]}(\breve{\beta}) \ge \cdots \ge \tilde{y}_{[n]}(\breve{\beta})$ and let

$$y_{[r]} = \tilde{y}_{[r]}(\breve{\beta}), \qquad G_{\breve{\beta}}(y) = \widehat{F}_{\breve{\beta}}(y \mid -\infty) / \widehat{F}_{\breve{\beta}}(y_{[r]} \mid -\infty), \tag{3.6}$$

for $y \le y_{[r]}$. To extend the scale estimate based on the median absolute deviation of residuals to right-censored data, we use the product-limit curve $G_{\breve{\beta}}$ of the residuals $\tilde{y}_i(\breve{\beta})$ to estimate σ by

$$\breve{\sigma} = \text{median absolute deviation of } G_{\breve{\beta}}, \tag{3.7}$$

where the median absolute deviation of a discrete probability distribution G with atoms a_k $(k=1,\cdots,K)$ is defined as the median of the discrete distribution that assigns mass $G(a_k)$ to $|a_k-\text{med}(G)|$ for $k=1,\cdots,K$, with med(G) given by the median of the histogram of G. Let X denote the $n\times(p+1)$ matrix whose ith row is $I(\tilde{y}_i(\tilde{\beta})\leq y_{[r]})(1,x_i^T)$.

Replacing the σ in (3.4) by $\tilde{\sigma}$, we can apply the same arguments as in Section 2 of Lai and Ying (1994) to modify (3.4) for right-censored data as

$$\sum_{i=1}^{n} I(\tilde{y}_i(\tilde{\beta})\leq y_{[r]})\psi_i(a,b)\binom{1}{x_i}=0, \tag{3.8}$$

where we take $r\geq 2$ to dampen potential instability due to small risk set sizes and

$$\psi_i(a,b)=\delta_i\psi((\tilde{y}_i(b)-a)/\tilde{\sigma})+(1-\delta_i)\int_{u>\tilde{y}_i(b)}\psi\left(\frac{u-a}{\tilde{\sigma}}\right)d\hat{F}_b(u\mid\tilde{y}_i(b)). \tag{3.9}$$

In fact, Lai and Ying (1991a) proposed a class of smooth functions to dampen the potential instability. However, most of our simulation studies have shown that taking r as small as 2 or 3 in (3.8) goes a long way towards avoiding potential instabilities in the unadjusted version (r=1) and that the smoothing method proposed by them is not far better than this one.

3.2 Computation of M-Estimators

We describe an iterative algorithm for computing the M-estimator defined by (3.8)-(3.9). Let $\theta=(\alpha,\beta^T)^T$ and let $\theta^{(k)}=(\alpha^{(k)},\beta^{(k)T})^T$ denote the result after the kth iteration to compute the M-estimator of θ. The algorithm consists of following steps.

1. For k=0, set $\theta^{(0)}=(\tilde{\alpha},\tilde{\beta}^T)^T$, which is the preliminary estimate given in Chapter 2, and compute $y_{[r]}$ and $\tilde{\sigma}$ using $\tilde{\alpha}$, $\tilde{\beta}$ and (3.6)-(3.7).
2. Compute $\tilde{y}_i(\beta^{(k)})$ for $i=1,\cdots,n$.

3. Evaluate $\widehat{F}_{\beta^{(k)}}(u \mid v)$ or $\widehat{F}_{\beta^{(k)}}(u \mid v-)$ by (3.3) at $u \in \{ \tilde{y}_i(\beta^{(k)}) : \delta_i = 1,\ i \le n \}$, $v \in \{ \tilde{y}_i(\beta^{(k)}) \}_{i \le n}$, with $u \ge v$.
4. Compute the $n \times 1$ vector $\Psi^{(k)}$ whose ith component is $\psi_i(\alpha^{(k)}, \beta^{(k)})$ $I(\tilde{y}_i(\tilde{\beta})) \le y_{[r]})$.
5. Solve the linear equation $X^T X z = X^T \Psi^{(k)}$ to find $z = z^{(k)}$.
6. Put $\theta^{(k+1)} = \theta^{(k)} + q_k z^{(k)}$, where $0 < q_k < 2$ is a relaxation factor.
7. Increase counter from k to $k+1$ and go to step 2.

Let $||u||$ denote some norm of $(p+1) \times 1$ vectors u, e.g. the Euclidean norm or the maximum of the absolute values of the components of u. Define

$$C(\theta) = \tilde{\sigma} || \sum_{i=1}^{n} \psi_i(\theta) (1, x_i^T)^T || \tag{3.10}$$

The above iterative scheme terminates if $||C(\theta^{(k)})|| < \eta$ or if $||\theta^{(k+1)} - \theta^{(k)}|| < \eta'$ for some prescribed thresholds η and η', or at $k = K$ if such numerical convergence has not occurred within K iterations. The M-estimator is given by the terminal value of these iterations.

This algorithm is similar to that used to compute M-estimators for complete data, cf. Huber (1981, pp. 181-182). For complete data, estimating equation (1.3) is equivalent to (1.2) and a Gauss-Newton procedure to solve (1.2) with $0 \le \rho''(= d^2\rho/du^2) \le 1$ consists of least-squares iterations of the type in steps 5 and 6. The estimating equations (3.2)-(3.3) can in fact be interpreted as an EM procedure (cf. Dempster, Laird and Rubin, 1977) to modify (1.2) for the incomplete right-censored data, in which the E-step is given by (3.3) and the M-step computes $\alpha^{(k+1)}$ and $\beta^{(k+1)}$ as solutions to (3.2), or equivalently, as minimizers of $E_{\alpha^{(k)}, \beta^{(k)}, \hat{F}_{\beta^{(k)}}} \{ \sum_j \rho(y_j - a - b^T x_j) \mid \text{observed data} \}$, where $\rho' = \psi$ and $\sum_j$ denotes the sum of the residuals in the complete sample that yields the observed data set, cf. Lai and Ying (1994). Hence step 4 in the above algorithm can be regarded as an E-step to reconstruct the scores of the residuals in the unobservable complete sample, while steps 5 and 6 can be regarded as an M-step, applying the usual least-squares iterations to compute M-estimators for complete data but with the reconstructed scores in place of the complete data scores.

How should we choose the relaxation factor q_k in step 6? First note that q_k=1 corresponds to the usual least-squares iteration. In the case of complete data, the relaxation factor was originally introduced because theoretical considerations indicated that $q_k \approx 1/E\rho''(\varepsilon_1) \geq 1$ would give faster convergence than q_k=1, but "empirical experience shows hardly any difference", as noted by Huber (1981, p.183). We use here the relaxation factor for another purpose, namely, to avoid the possibility of the iterations oscillating between two (or more) values because of the discreteness of $\widehat{F}_b$ in (3.9). This oscillatory behavior was noted by Buckley and James (1979) in their successive substitution algorithm to solve (3.2)-(3.3) for the case $\psi(u)=u, p=1$. It is clear that this oscillatory behavior can be avoided by supplementing their successive substitution algorithm with a more thorough search at intermediate points between the two oscillatory values. Our use of relaxation factors in step 6 is targeted towards avoiding such oscillatory behavior. Specifically, we choose $q_k \in \{2^{-l}: l=0,1,\cdots,L-1\}$ that minimizes $C(\theta^{(k)}+q_k z^{(k)})$ over these L choices of q_k. To implement this, steps 6 and 7 are broken into L parallel computations, giving $\theta_l^{(k+1)}=\theta^{(k)}+2^{-l}z^{(k)}$ in step 6 so that step 7 goes back to steps 2-4 with $\beta_l^{(k+1)}$ and $\alpha_l^{(k+1)}$ in place of $\beta^{(k+1)}$ and $\alpha^{(k+1)}$. After computing $\psi_l(\alpha_l^{(k+1)}, \beta_l^{(k+1)})$, we can compute the criterion $C(\theta_l^{(k+1)})$ defined in (3.10). We then choose the l^* that minimizes $C(\theta_l^{(k+1)})$ and set $\theta^{(k+1)}=\theta_{l^*}^{(k+1)}$. The above method to compute M-estimators, which is based on EM ideas, is also applicable to the unadjusted M-estimators defined by (3.2)-(3.3) and to the more complicated refinement thereof defined by (4.34) of Lai and Ying (1994) using smoothing kernels to dampen the instability due to small risk set sizes. It is much simpler than the simulated annealing algorithm to minimize (3.10) directly, which Lin and Geyer (1992, Section 6) proposed to use to compute the Buckley-James and similar regression estimators when $p>1$. Moreover, our algorithm typically converges after a few iterations, as will be illustrated in the simulation study of Section 3.3

3.3 Simulations

Since the above algorithm is easily applicable to a multiple linear regression model, for simplicity, we consider a simple linear regression model $y_j=\beta x_j+\varepsilon_j$, where the ε_j are i.i.d. random variables whose common distribution function

F is contaminated normal of the form $F = 0.9N(0,1) + 0.1N(0,3)$. The x_j are independent, uniformly distributed on [-2, 2] and independent of the ε_j. The y_j are subject to right censoring by i.i.d. $N(2,3)$ random variables c_j that are independent of the (x_j, ε_j). This corresponds to about 30% censoring rate. Samples with different sample sizes, (n=30,n=50,n=100), which satisfy the above conditions, are generated from the model with $\beta = 1$ and these samplings are replicated by 50, 100, 150 and 200 times. For each sample size, the numerical results are tabulated in Table 1-2. These tables have means and standard deviations of $\hat{\alpha}^H$, $\hat{\beta}^H$ (the Huber's M-estimators) and $\hat{\alpha}^{BJ}$, $\hat{\beta}^{BJ}$ (the Buckley-James esitmators) respectively. No stratification is needed for the preliminary estimator in Chpter 2.

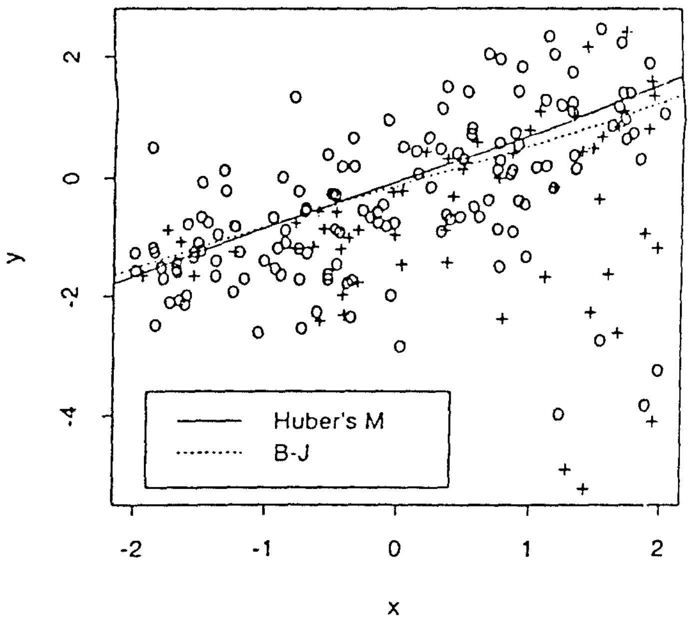
As shown in <Figure 1>, the Huber's M-estimator is less affected by outliers, which are either censored or uncensored, than the Buckley-James estimator. Figure 2 shows how the criterion $C(\theta)$ behaves over a region of $\theta = (\alpha, \beta)$ given a set of censored data. And Table 1-2 tell us that most of the standard deviations of the Buckley-James estimator are relatively larger than those of the M-estimator. That is, the M-estimator is less variable than the Buckley-James estimator.

< Table 1 > Comparison of Huber's M-estimator and Buckley-James estimator, by sampling with replication=50,100 and n=30,50,100. (The values of the first line in each row represent means and the values in the parentheses show standard deviations)

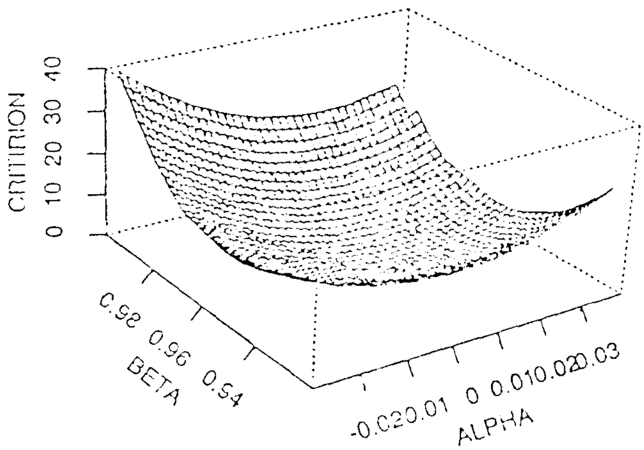
	R=50				R=100			
n	$\hat{\alpha}^{BJ}$	$\hat{\beta}^{BJ}$	$\hat{\alpha}^{H}$	$\hat{\beta}^{H}$	$\hat{\alpha}^{BJ}$	$\hat{\beta}^{BJ}$	$\hat{\alpha}^{H}$	$\hat{\beta}^{H}$
30	-0.1540 (0.2983)	0.9734 (0.2871)	-0.1163 (0.2824)	0.9854 (0.2376)	-0.0924 (0.2985)	1.0069 (0.2735)	-0.0348 (0.2653)	0.9959 (0.2529)
50	-0.1360 (0.2618)	0.9863 (0.1974)	-0.0838 (0.2194)	0.9851 (0.1693)	-0.0796 (0.2225)	1.0182 (0.2037)	-0.0235 (0.2182)	1.0148 (0.1973)
100	-0.0932 (0.1466)	1.0094 (0.1209)	-0.0529 (0.1392)	0.9973 (0.1125)	-0.0519 (0.1604)	0.9996 (0.1414)	-0.0138 (0.1418)	0.9983 (0.1310)

< Table 2 > Comparison of Huber's M-estimator and Buckley-James estimator, by sampleing with replication=150, 200 and n=100, 150, 200. (The values of first line in each row represent means and the values in the parentheses show standard deviations)

	R=150				R=200			
n	$\hat{\alpha}^{BJ}$	$\hat{\beta}^{BJ}$	$\hat{\alpha}^{H}$	$\hat{\beta}^{H}$	$\hat{\alpha}^{BJ}$	$\hat{\beta}^{BJ}$	$\hat{\alpha}^{H}$	$\hat{\beta}^{H}$
100	-0.1055 (0.2871)	0.9982 (0.2775)	-0.0544 (0.2709)	1.0025 (0.2442)	-0.1395 (0.2497)	1.0068 (0.2899)	-0.0841 (0.2483)	0.9918 (0.2466)
150	-0.0899 (0.2025)	1.0128 (0.1931)	-0.0454 (0.1846)	1.0057 (0.1833)	-0.0648 (0.2019)	1.0021 (0.1888)	-0.0042 (0.1958)	1.0076 (0.1719)
200	-0.0573 (0.1321)	0.9906 (0.1283)	-0.0236 (0.1163)	0.9902 (0.1153)	-0.0422 (0.1495)	0.9968 (0.1261)	-0.0134 (0.1282)	0.9934 (0.1189)



< Figure 1 > The lines represent fitted regression lines is Buckley-James estimator and Huber's M-estimator, respectively. (The dotted line stands for Buckley-James and, the solid line for Huber's. Also in raw data, o denotes uncensored data and + denotes censored data)



< Figure 2 > Criterion defined by (3.10) versus α and β

4. Conclusion

In this paper we have introduced a relatively simple method to compute M-estimators of regression parameters from right-censored data. The computational method essentially uses the same algorithm for computing M-estimators from complete data to iteratively reconstructed data. The reconstruction is based on the missing information principle and involves at each iteration the updated value of the M-estimator and the product-limit estimate of the underlying distribution of the ε_i. Starting with a good preliminary estimate described in Chapter 2, the iterative scheme typically converges after a few iterations. This computational method makes M-estimators much more attractive in practice than rank estimators that have similar robustness properties but much higher computational complexity.

References

[1] Belseley, D.A., Kuh, E. and Welsch, R.E.(1980), *Regression Diagnostics,* Wiley, New York.

[2] Buckley, J. and James, I.(1979), "Linear regression with censored data." *Biometrica,* Vol. 66, pp. 429-436.

[3] Cox, D.R. and Snell, E.J.(1968), "A general definition of residuals (with discussion)," *Journal of Royal Statistical Society, Series B,* Vol. 30, pp. 248-275.

[4] Demsper, A.P., Laird, N.M. and Rubin, D.B.(1977), "Maximum likelihood from incomplete data via the EM algorithm(with discussion)." *Journal of Royal Statistical Society, Series B,* Vol. 39, pp. 1-38.

[5] Gross, S. and Lai, T.L.(1995), "Nonparametric estimation and regression analysis with left truncated and right censored data," To appear in *Journal of the American Statistical Association.*

[6] Huber, P.J.(1973), "Robust regression: Asymtotics, conjectures and Monte Carlo," *Annals of Statistics,* Vol. 1, pp. 799-821.

[7] Huber, P.J.(1977), *Robust Statistical procedures,* Philadelphia: Society of Industrial and Applied Mathematics.

[8] Huber, P.J.(1981), *Robust regression,* Wiley, New York.

[9] Lai, T.L. and Ying, Z.(1991a), "Estimating a distribution function with truncated and censored data," *Annals Statistics,* Vol. 19, pp. 417-442.

[10] Lai, T.L. and Ying, Z.(1991b), "Rank regression methods for left-truncated and right-censored data," *Annals Statistics,* Vol. 19, pp. 531-556.

[11] Lai, T.L. and Ying, Z.(1994), "A missing information principle and M-estimators in regression analysis with censored and truncated data," *Annals of Statistics,* Vol. 22, pp. 1222-1255.

[12] Lin, D.Y and Geyer, C.J.(1992), "Computational methods for semiparametric linear regression with censored data," *Journal of Computational and Graphical Statistics,* Vol. 1, pp. 77-90.

[13] Ritov, Y.(1990), "Estimation in a linear regssion model with censored data," *Annals Statistics,* Vol. 18, pp. 303-328.