Journal of the Korean
Statistical Society
Vol. 26, No. 3, 1997

Robust Regression and Stratified Residuals for Left-Truncated and Right-Censored Data

Chul-Ki Kim[1]

Abstract

Computational algorithms to calculate M-estimators and rank estimators of regression parameters from left-truncated and right-censored data are developed herein. In the case of M-estimators, new statistical methods are also introduced to incorporate leverage assessments and concomitant scale estimation in the presence of left truncation and right censoring on the observed responses. Furthermore, graphical methods to examine the residuals from these data are presented. Two real data sets are used for illustration.

Key Words : EM algorithm; M-estimators; Metrically Winsorized residuals; Rank regression; Survival analysis.

[1]Department of Statistics, Ewha Womans University, Seoul, 120-750, Korea.

1. INTRODUCTION

Consider the linear regression model

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

where the ϵ_j are i.i.d. random variables and the x_j are independent $p \times 1$ random vectors independent of $\{\epsilon_n\}$. Huber's (1973) M-estimators $\hat{\alpha}$, $\hat{\beta}$ of α, β are defined as a solution vector to the minimization problem

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

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

$$\sum_{j=1}^{n} \rho'(y_j - a - b^T x_j) = 0, \quad \sum_{j=1}^{n} x_j \rho'(y_j - a - b^T x_j) = 0. \tag{1.3}$$

For Huber's score function, (1.2) is tantamount to applying the method of least-squares to "metrically Winsorized residuals", cf. Huber (1981, p.180).

Another class of robust estimators of α, β is defined by the ranks of the residuals, cf. Jurečkova (1969, 1971). The idea is to first estimate β using linear rank statistics based on $y_j(b)$. Specifically, let $R_j(b)$ be the rank of $y_j(b)$ in the set $\{y_1(b), \cdots, y_n(b)\}$ and define the linear rank statistic

$$L(b) = \sum_{j=1}^{n} x_j a_n(R_j(b)), \tag{1.4}$$

where the scores $a_n(k)$ are generated from a score function $\phi : [0,1] \to (-\infty, \infty)$ that satisfies $\int_0^1 \phi(u)du = 0$ and $\int_0^1 \phi^2(u)du < \infty$ via

$$a_n(k) = \phi(k/n), \text{ or } a_n(k) = \phi(k/(n+1)), \text{ or } a_n(k) = E\phi(U^n_{(k)}), \tag{1.5}$$

$U^n_{(1)} \le \cdots \le U^n_{(n)}$ being the order statistics of a sample of size n from the uniform distribution on [0,1]. The rank estimator $\hat{\beta}$ is defined by the estimating equation $L(b) = 0$.

Suppose that the responses y_j in (1.1) are not completely observable due to left truncation and right censoring by random variables t_j and c_j such that $\infty > t_j \ge -\infty$ and $-\infty < c_j \le \infty$. Let $\tilde{y}_j = y_j \wedge c_j$ and $\delta_j = I(y_j \le c_j)$, where we use $\wedge$ and $\vee$ to denote minimum and maximum, respectively. In addition to right censorship of the responses y_j by c_j, we shall also assume left truncation in the sense that $(x_j, \tilde{y}_j, \delta_j)$ can be observed only when $\tilde{y}_j \geq t_j$. The data, therefore, consist of n observations

$$(x_i^o, \tilde{y}_i^o, \delta_i^o, t_i^o) \quad \text{with} \quad \tilde{y}_i^o \geq t_i^o, \quad i = 1, \cdots, n. \tag{1.6}$$

Unless otherwise stated, it will be assumed that (t_j, c_j, x_j) are independent of the sequence $\{\epsilon_n\}$.

Generalizations and asymptotic theories of the rank approach and M-estimators for regression analysis with left-truncated and right-censored (l.t. r.c.) data were recently provided by Lai and Ying (1991b, 1994). Lin and Geyer (1992) developed computational methods to implement these rank regression procedures. For censored and/or truncated data in the case $p > 1$, the rank estimators "require minimizing discrete objective functions with multiple local minima" and "conventional optimization algorithms cannot be used to solve such minimization problems", as noted by Lin and Geyer (1992) who used probabilistic algorithms (simulated annealing) to circumvent these difficulties. In Section 4, we discuss further this issue of multiple local minima and show that the difficulties can sometimes be resolved by using a good preliminary estimate, which we describe in Section 2.

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 Section 3, we augment and modify these algorithms for computing M-estimators of regression parameters from l.t.r.c. data. In the case $p = 1$ 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 l.t.r.c. 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. Not only do we address this important issue in Section 3, but we also robustify the estimating equations by removing the effect of bad leverage points. Furthermore, by making use of the simple preliminary estimate of the regression parameters in Section 2, we develop a computationally simpler version of the adjustments for small risk set sizes proposed by Lai and Ying (1994). In Section 5, using the general concept of residuals introduced by Cox and Snell (1968), we show how residuals computed from l.t.r.c. data can be plotted and used for regression diagnostics. In Section 6, we illustrate the computational and graphical methods developed here for the implementation of rank and M-estimators on two real data sets from biomedical studies.

2. INITIALIZING WITH A SIMPLE PRELIMINARY ESTIMATOR

Both the M-estimator in Section 3 and the rank estimator in Section 4 involve 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 (1.6), partition the range of x_i^o-values into $m(\geq 1)$ subsets $\mathcal{X}_1, \cdots, \mathcal{X}_m$, thereby stratifying the data, with the kth stratum corresponding to x_i^o-values in $\mathcal{X}_k$. From the n_k quadruples $(x_i^o, \tilde{y}_i^o, \delta_i^o, t_i^o)$ of observed data within the kth stratum (i.e., $x_i^o \in \mathcal{X}_k$ and $\sum_{k=1}^m n_k = n$), define

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

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

$$\sum_{k=1}^m \sum_{i: x_i^o \in \mathcal{X}_k} \delta_i^o (\tilde{y}_i^o - a - b^T x_i^o) n_k \hat{S}_k(\tilde{y}_i^o) I(\#_k(\tilde{y}_i^o) \geq s)/\#_k(\tilde{y}_i^o) = 0, \quad (2.2)$$
$$\sum_{k=1}^m \sum_{i: x_i^o \in \mathcal{X}_k} \delta_i^o x_i^o (\tilde{y}_i^o - a - b^T x_i^o) n_k \hat{S}_k(\tilde{y}_i^o) I(\#_k(\tilde{y}_i^o) \geq s)/\#_k(\tilde{y}_i^o) = 0 \quad (2.3)$$

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

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

The weights are introduced to adjust for the bias caused by censoring and truncation in the usual (unadjusted) least-squares estimates from $\{(x_i^o, \tilde{y}_i^o) :$

$1 \le i \le n\}$. The rationale behind these weights is explained in Gross and Lai (1995). Briefly, suppose that the complete data of which $\{(x_i^o, \tilde{y}_i^o, \delta_i^o, t_i^o) : 1 \le i \le n\}$ is an observable subsample are (x_j, y_j, t_j, c_j), $1 \le j \le \nu_n$. By choosing the strata suitably so that the x-values do not change much within $\mathcal{X}_k$, we can assume, at least approximately, that (t_j, c_j) is conditionally independent of (x_j, y_j) given $x_j \in \mathcal{X}_k$, and that all the y_j with $x_j \in \mathcal{X}_k$ have the same conditional survival function (given the covariate belongs to the kth stratum) 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, LEVERAGE ADJUSTMENTS AND CONCOMITANT SCALE ESTIMATION

Throughout the sequel we shall use the following notation for the l.t.r.c. data (1.6). Let $\tilde{y}_i^o(b) = \tilde{y}_i^o - b^T x_i^o$, $t_i^o(b) = t_i^o - b^T x_i^o$ and

$$N(b,u) = \sum_{i=1}^n I(t_i^o(b) \le u \le \tilde{y}_i^o(b)), \quad \Delta(b,u) = \sum_{i=1}^n I(\tilde{y}_i^o(b) = u, \delta_i^o = 1),$$

$$\hat{F}_b(u|v) = 1 - \prod_{i: v < \tilde{y}_i^o(b) \le u, \delta_i^o = 1} \{1 - \Delta(b, \tilde{y}_i^o(b))/N(b, \tilde{y}_i^o(b))\}. \tag{3.1}$$

The notation $\hat{F}_b(u|v-)$ will be used to denote (3.1) in which "$v < \tilde{y}_i^o(b)$" is replaced by "$v \le \tilde{y}_i^o(b)$." The function $\hat{F}_b(u|-\infty)$ is the product-limit estimate of the common distribution function $F(u)$ of the $\epsilon_j + \alpha$ in (1.1).

Put $\psi = \rho'$ in (1.3). To extend (1.3) to the l.t.r.c. data (1.6), Lai and Ying (1994) applied a "missing information principle" which leads to replacing (1.3) by the estimating equations

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

where

$$\begin{aligned}\psi_i^*(a,b) &= \delta_i^o \psi(\tilde{y}_i^o(b) - a) + (1 - \delta_i^o) \int_{u > \tilde{y}_i^o(b)} \psi(u-a) d\hat{F}_b(u|\tilde{y}_i^o(b)) \\ &- \int_{u \ge t_i^o(b)} \psi(u-a) d\hat{F}_b(u|t_i^o(b)-),\end{aligned} \tag{3.3}$$

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 $\int_{-\infty}^{\infty} \psi(u-a)d\hat{F}_b(u|-\infty) = 0$. Hence for the censored regression model ($t_i \equiv -\infty$), the last term in (3.3) vanishes and 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 Incorporating Scale and Leverage and Dampening Instability Due to Small Risk Set Sizes

For complete data, a robust choice of $\psi(=\rho')$ is Huber's score function which involves some scale parameter. Moreover, the influence of extreme design points on the linear fit can be down-weighted via the leverage values $h_j = z_j^T(\sum_1^n z_t z_t^T)z_j$, where $z_j = (1, x_j^T)^T$. Accordingly estimating equations of robust M-estimators modify (1.3) as

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

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 and leverage. Moreover, the product-limit estimate $\hat{F}_\beta(u|v)$ may be quite unstable when v is near $\min_{i\le n} t_i^o(\beta)$ or near $\max_{i\le n} \tilde{y}_i^o(\beta)$, cf. Lai and Ying (1991a). Hence it is desirable to down-weight such v in (3.2). This is done in Section 4.1 of 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 l.t.r.c 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 with concomitant scale estimation can not be an EM procedure for l.t.r.c. data and may be numerically unstable. We can avoid these difficulties by using a separate scale estimate $\widetilde{\sigma}$ based on $\hat{F}_{\widetilde{\beta}}$, where $\widetilde{\beta}$ is the preliminary estimate of β introduced in Section 2. Moreover, instead of down-weighting the extreme order statistics of the $t_k^o(b)$ and $\tilde{y}_k^o(b)$ via an elaborate smoothing scheme as in Lai and Ying (1994), we propose to trim away the extreme order statistics of the $t_k^o(\widetilde{\beta})$ and $\tilde{y}_k^o(\widetilde{\beta})$, which do not involve b, so that we do not need the elaborate scheme to ensure smooth trimming in b.

Ordering the $\tilde{y}_k^o(\widetilde{\beta})$ as $\tilde{y}_{[1]}^o(\widetilde{\beta}) \geq \cdots \geq \tilde{y}_{[n]}^o(\widetilde{\beta})$ and the $t_k^o(\widetilde{\beta})$ as $t_{(1)}^o(\widetilde{\beta}) \leq \cdots \leq t_{(n)}^o(\widetilde{\beta})$, let

$$y_{[r]} = \tilde{y}_{[r]}^o(\widetilde{\beta}), \quad t_{(r)} = t_{(r)}^o(\widetilde{\beta}), \quad G_{\widetilde{\beta}}(y) = \hat{F}_{\widetilde{\beta}}(y|t_{(r)})/\hat{F}_{\widetilde{\beta}}(y_{[r]}|t_{(r)}), \tag{3.6}$$

for $t_{(r)} \leq y \leq y_{[r]}$. To extend the scale estimate based on the median absolute deviation of residuals to l.t.r.c. data, we use the product-limit curve $G_{\widetilde{\beta}}$ of the residuals $\tilde{y}_i^o(\widetilde{\beta})$ to estimate σ by

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

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 $\text{med}(G)$ given by the median of the histogram of G. Let X denote the $n \times (p+1)$ matrix whose ith row is $I(\tilde{y}_i^o(\widetilde{\beta}) \leq y_{[r]}, t_i^o(\widetilde{\beta}) \geq t_{(r)})(1, x_i^{oT})$. Let

$$H = (h_{ik})_{1 \leq i,k \leq n} = X(X^T X)^{-1} X^T. \tag{3.8}$$

Replacing the σ in (3.4) by $\widetilde{\sigma}$, we can apply the same arguments as in Section 2 of Lai and Ying (1994) to modify (3.4) for l.t.r.c. data as

$$\sum_{i=1}^n \sqrt{1 - h_{ii}}\, I(\tilde{y}_i^o(\widetilde{\beta}) \leq y_{[r]}, t_i^o(\widetilde{\beta}) \geq t_{(r)})\, \psi_i(a,b) \begin{pmatrix} 1 \\ x_i^o \end{pmatrix} = 0, \tag{3.9}$$

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

$$\begin{aligned} \psi_i(a,b) &= \delta_i^o \psi((\tilde{y}_i^o(b) - a)/(\widetilde{\sigma}\sqrt{1 - h_{ii}})) \\ &+ (1 - \delta_i^o) \int_{u > \tilde{y}_i^o(b)} \psi\left(\frac{u - a}{\widetilde{\sigma}\sqrt{1 - h_{ii}}}\right) d\hat{F}_b(u|\tilde{y}_i^o(b)) \\ &- \int_{u \geq t_i^o(b)} \psi\left(\frac{u - a}{\widetilde{\sigma}\sqrt{1 - h_{ii}}}\right) d\hat{F}_b(u|t_i^o(b)-). \end{aligned} \tag{3.10}$$

Experience has shown that taking r as small as 2 or 3 in (3.9) goes a long way towards avoiding potential instabilities in the unadjusted version ($r = 1$).

3.2 Computation of M-estimators

We describe an iterative algorithm for computing the M-estimator defined by (3.9)-(3.10). 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 the following steps.

1. For $k = 0$, set $\theta^{(0)} = (\tilde{\alpha}, \tilde{\beta}^T)^T$, which is the preliminary estimate given in Section 2, and compute $y_{[r]}$, $t_{(r)}$ and $\tilde{\sigma}$ using $\tilde{\alpha}$, $\tilde{\beta}$ and (3.6)-(3.7). Also evaluate X^TX and the diagonal elements of (3.8).

2. Compute $\tilde{y}_i^o(\beta^{(k)})$ and $t_i^o(\beta^{(k)})$ for $i = 1, \cdots, n$.

3. Evaluate $\hat{F}_{\beta^{(k)}}(u|v)$ or $\hat{F}_{\beta^{(k)}}(u|v-)$ by (3.1) at $u \in \{\tilde{y}_i^o(\beta^{(k)}) : \delta_i^o = 1, i \leq n\}$, $v \in \{\tilde{y}_i^o(\beta^{(k)}), t_i^o(\beta^{(k)})\}_{i \leq n}$ with $u \geq v$.

4. Compute the $n \times 1$ vector $\Psi^{(k)}$ whose ith component is $\sqrt{1 - h_{ii}}\ \psi_i(\alpha^{(k)}, \beta^{(k)})\ I(\tilde{y}_i^o(\tilde{\beta}) \leq y_{[r]}, t_i^o(\tilde{\beta}) \geq t_{(r)})$.

5. Solve the linear equation $X^TXz = X^T\Psi^{(k)}$ 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

$$C(\theta) = \tilde{\sigma} \| \sum_{i=1}^{n} \sqrt{1 - h_{ii}}\ \psi_i(\theta)(1, x_i^{oT})^T \|. \tag{3.11}$$

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, the estimating equation (1.3) is equivalent to (1.2) and a Gauss-Newton procedure to solve (1.2) with $0 \leq \rho''(= d^2\rho/du^2) \leq 1$ 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 l.t.r.c. 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 $E_{\alpha^{(k)},\beta^{(k)},\hat{F}_{\beta^{(k)}}}\{\sum_j \rho(y_j - a - b^T x_j)|\text{observed data}\}$, where $\rho' = \psi$ and $\sum_j$ denotes the sum of the residuals in the complete sample that yields the observed data set as its observable subset, 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 $q_k \approx 1/E\rho''(\epsilon_1) \geq 1$ 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 $\hat{F}_b$ in (3.10). This oscillatory behavior was noted by Buckley and James (1979) in their successive substitution algorithm to solve (3.2)-(3.3) for the case $\psi(u) = u$, $p = 1$ and $t_i = -\infty$ (no truncation). 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 $\psi_i(\alpha_l^{(k+1)}, \beta_l^{(k+1)})$, we can compute the criterion $C(\theta_l^{(k+1)})$ defined in (3.11). 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

(3.11) 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 following example.

Example 1. Consider a simple linear regression model $y_j = \beta x_j + \epsilon_j$, where the ϵ_j are i.i.d. random variables whose common distribution function F is contaminated normal of the form $F = 0.7N(0,1) + 0.3N(0,8^2)$. The x_j are independent, uniformly distributed on $[-1,1]$ and independent of the ϵ_j. The y_j are subject to right censoring by i.i.d. N(6,5^2) random variables c_j that are independent of the (x_j, ϵ_j), and also to left truncation by $t_j = \min(c_j, u_j)$ in which the u_j are i.i.d. N(−6,5^2) and independent of (x_j, ϵ_j, t_j). This corresponds to 28% censoring rate and 24% truncation rate. A sample of 50 l.t.r.c. data $(x_i^o, \tilde{y}_i^o, \delta_i^o, t_i^o)$, $i = 1, \cdots, 50$, was generated from the model with $\beta = 1$. The sample values can be found in Table 1 of Section 5. Since the (x_j, y_j) are i.i.d. and independent of the (t_j, c_j), no stratification is needed $(m = 1)$ for the preliminary estimator in Section 2, which assumes unknown $\alpha (= 0)$. Taking $s = 2$ in (2.1)-(2.3), the preliminary estimator was found to be $\tilde{\alpha} = -0.42$, $\tilde{\beta} = 3.14$. Initializing with this preliminary estimate, the scale estimate (3.7) was found to be $\tilde{\sigma} = 3.69$ and the Huber-type M-estimator defined by (3.9) with $r = 2$, was found to be $(\hat{\alpha}^H, \hat{\beta}^H) = (-0.80, 0.83)$ after 8 iterations. Using a similar algorithm to solve (3.2)-(3.3) with $\rho'(u) = u$ (the Buckley-James-type score function) gave $(\hat{\alpha}^{BJ}, \hat{\beta}^{BJ}) = (-0.77, 2.00)$ after 10 iterations. Note that $\hat{\beta}^H$ is much closer to the true value $\beta = 1$ than $\hat{\beta}^{BJ}$. Moreover, both $\hat{\alpha}^H$ and $\hat{\alpha}^{BJ}$ underestimate $\alpha = 0$ since they are actually consistent estimates of certain trimmed means of the ϵ_j as shown in Section 4. 3 of Lai and Ying (1994).

4. RANK REGRESSION WITH L.T.R.C. DATA

To extend the linear rank statistics (1.4) to the l.t.r.c. data (1.6) and thereby to derive rank estimators of β from l.t.r.c. data, Lai and Ying (1991b) first introduced the following modification of the product-limit estimator:

$$\tilde{F}_b(u) = 1 - \prod_{i:\tilde{y}_i^o(b) < u, \delta_i^o = 1} \{1 - p_n(N(b, \tilde{y}_i^o(b))/n)\ \Delta(b, \tilde{y}_i^o(b))/N(b, \tilde{y}_i^o(b))\}, \quad (4.1)$$

where the factor $p_n(\cdot)$ is used to trim out the $\tilde{y}_i^o(b)$ with relatively small risk set sizes $N(b, \tilde{y}_i^o(b))$. Letting $\psi = \phi - \Phi$, in which ϕ is the score function in (1.5) and $\Phi(u) = (1-u)^{-1} \int_u^1 \phi(t)dt$, $0 \le u < 1$, they defined

$$R(b) = \sum_{i=1}^{n} \delta_i^o \psi(\tilde{F}_b^o(\tilde{y}_i^o(b)))\, p_n\left(\frac{N(b, \tilde{y}_i^o(b))}{n}\right) \left\{x_i^o - \frac{\sum_{k=1}^n I(\tilde{y}_k^o(b) \geq \tilde{y}_i^o(b) \geq t_k^o(b)) x_k^o}{\sum_{k=1}^n I(\tilde{y}_k^o(b) \geq \tilde{y}_i^o(b) \geq t_k^o(b))}\right\}. \tag{4.2}$$

The rank estimator $\hat{\beta}$ of β is defined by the estimating equation $R(b) = 0$, or more precisely, as a zero-crossing of $R(b)$ in the case $p = 1$ and as a minimizer of $\|R(b)\|$ in the case $p > 1$, where $\|\cdot\|$ denotes some norm of p-dimensional vectors.

For technical reasons in the derivation of an asymptotic theory for such rank estimators, the trimming factor $p_n(w)$ in Lai and Ying (1991b) is taken to be of the form $f(n^\lambda(w - cn^\lambda))$ for some $c > 0$, $0 < \lambda < 1$ and for some smooth function $f(t)$ that vanishes for $t \leq 0$ and assumes the value 1 for $t \geq 1$. However, empirical studies have shown that the simple trimming $p_n(w) = I(w \geq r/n)$, or equivalently

$$p_n(N(b, \tilde{y}_i^o(b))/n) = I(N(b, \tilde{y}_i^o(b)) \geq r), \tag{4.3}$$

works well for r as small as 2 or 3 in moderately sized samples.

In the case $p = 1$, the function $R(b)$ may have multiple zero-crossings, while in the case $p \geq 1$, $\|R(b)\|$ typically has multiple local minima. The preliminary estimate $\tilde{\beta}$ in Section 2 can be used to resolve which of these zero-crossings (or local minima of $\|R(b)\|$) should be chosen as the rank estimator. In practice, one can restrict the search for zero-crossings or minima within some neighborhood of $\tilde{\beta}$. For $p = 1$ or 2, a grid search within this region can be carried out without much difficulty. For larger p, we can use the simulated annealing algorithm in Lin and Geyer (1992) to find the minimizer of $\|R(b)\|$ within this region.

Example 2. Consider the simple regression model $y_j = \beta x_j + \epsilon_j$, in which the ϵ_j are i.i.d. $N(0,1)$ and x_j are i.i.d. uniformly distributed on $[0,2]$ and independent of the ϵ_j. The y_j are subject to left truncation by i.i.d. $N(0,1)$ random variables t_j that are independent of the (x_j, ϵ_j), and to right censoring by $c_j = t_j + \max(e^{-t_j}, 0.5)u_j$, in which the u_j are i.i.d. uniformly distributed on $[0,5]$ and independent of the (x_j, ϵ_j, t_j). A sample of 200 l.t.r.c. data $(x_i^o, \tilde{y}_i^o, \delta_i^o, t_i^o)$, $i = 1, \cdots, 200$, was generated from this model with $\beta = 1$ and the raw data are plotted in Figure 1 (a). Taking $m = 1$, $s = r = 2$ in (2.1)-(2.3) and (4.3), the preliminary estimator was found to be $\tilde{\alpha} = -0.0727$ and

$\widetilde{\beta} = 0.962$. Figure 1 (b) plots the function $R(b)$ for $-25 \leq b \leq 25$. Although the function looks continuous because of the plotter and the relatively small jumps within the wide range of $R(b)$, $R(b)$ is actually a step function with a zero-crossing at $b = 0.952$ (where $R(b+) = 0.007$, $R(b-) = -0.008$). The function $R(b)$ approaches 0 as $|b| \to \infty$. In fact, for $b = -25$ or $b = 35$, $R(b)$ is equal to 0 to 6 decimal places. This illustrates the point, noted by Lin and Geyer (1992), that $|R(b)|$ may have multiple local minima.

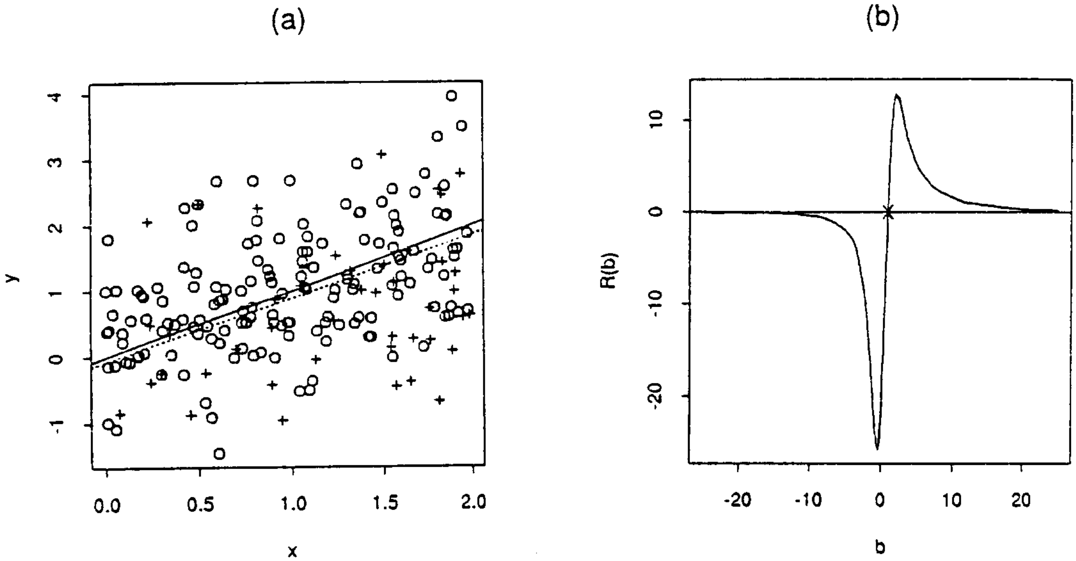

Figure 1. (a) Plot of raw data(o : uncensored data; + : censored). The solid and the dotted lines represent the true and fitted(using the rank method) regression lines, respectively ; (b) Plot of linear rank statistic $R(b)$ versus b.

For complete data, after evaluating the rank estimator $\hat{\beta}$ as the zero-crossing (or minimizer of the norm) of (1.4), one can use the Hodges-Lehmann or other rank estimators of location based on the residuals $y_j - \hat{\beta}^T x_j$ to estimate the location parameter α in (1.1). Alternatively one can also estimate α by the median, or some trimmed mean, or other linear combinations of the order statistics of the residuals $y_j - \hat{\beta}^T x_j$. For l.t.r.c. data, a convenient estimate of α after the rank estimator $\hat{\beta}$ has been computed is the trimmed mean

$$\hat{\alpha} = \{\int u d\widetilde{F}_{\hat{\beta}}(u)\}/\widetilde{F}_{\hat{\beta}}(\max\{\tilde{y}_i^o(\hat{\beta}) : \delta_i^o = 1, N(\hat{\beta}, \tilde{y}_i^o(\hat{\beta})) \geq r\}). \quad (4.4)$$

Note that with $p_n(\cdot)$ given by (4.3), the modified product-limit estimator $\widetilde{F}_b(u)$ in (4.1) has no jumps at u with $N(b, u) < r$. For the simulated data in

Example 2, the estimate of $\alpha(=0)$ given by (4.4) is $\hat{\alpha} = -0.0686$, which is quite close to the preliminary estimate $\widetilde{\alpha} = -0.0727$.

5. STRATIFICATION OF RESIDUALS AND REGRESSION DIAGNOSTICS

Let $(\hat{\alpha}, \hat{\beta})$ be an estimate of the parameter (α, β) of the assumed regression model (1.1). For complete data, under (1.1) and suitable regularity conditions, the residuals $r_j = y_j - \hat{\alpha} - \hat{\beta}^T x_j$ provide approximations of the unobservable i.i.d. random variables ϵ_j with an $O_p(1/\sqrt{n})$ error. Therefore substantial deviations of these residuals from an i.i.d. pattern (e.g. conspicuous trends with respect components of the covariate vector x_j) suggest inadequacies and possible improvements of the assumed regression model. For the l.t.r.c. data (1.6), we can likewise compute

$$r_i^o = \tilde{y}_i^o - \hat{\alpha} - \hat{\beta}^T x_i^o \ (= \tilde{y}_i^o(\hat{\beta}) - \hat{\alpha}), \ i = 1, \cdots, n, \tag{5.1}$$

and regard them as "residuals" in the general sense of Cox and Snell (1968), namely, as approximations of the unobservable $\epsilon_i^o = \tilde{y}_i^o - \alpha - \beta^T x_i^o$, which comprise the observable subset of the complete sample of i.i.d. ϵ_j that are subject to left truncation by $t_j - \alpha - \beta^T x_j$ and right censoring by $c_j - \alpha - \beta^T x_j$. Substantial deviations of (5.1) from the observable part of an i.i.d. sample subject to truncation and censoring, therefore, again suggest inadequacies of the assumed regression model (1.1), if we can quantify and detect these deviations.

As in the case of complete data, one can plot the residuals (5.1), with different symbols to represent the uncensored ($\delta_i^o = 1$) and censored ($\delta_i^o = 0$) residuals, versus each component of the covariate vector x_i^o. Although such plots are easy to do, direct examination of them is often not revealing because of the uncertainties in the censored residuals and because the corresponding ϵ_i^o need not be i.i.d. themselves due to truncation and censoring. We propose to (i) partition each covariate axis into subsets and thereby stratify the residuals, (ii) compute from the l.t.r.c. residuals for each stratum a trimmed mean of the kind described in Gross and Lai (1995), and (iii) include these trimmed means as horizontal line segments over the partitions in the covariate axis.

To be more specific, noting that the kth stratum in the above stratification of the residuals can be described by a subset I_k of the covariate space, define $N_k(b, u)$, $\Delta_k(b, u)$ and $\hat{F}_{b,k}(u|v)$ as in (3.1) but with the sum or product over

i restricted to $x_i^o \in I_k$. Take $\nu \geq 2$, let R' be the smallest order statistic of $\tilde{y}_1^o(\hat{\beta}), \cdots, \tilde{y}_n^o(\hat{\beta})$ such that $\min_k N_k(\hat{\beta}, R') \geq \nu$, and let R'' be the largest order statistic of the $\tilde{y}_i^o(\hat{\beta})$ such that $\min_k N_k(\hat{\beta}, R'') \geq \nu$. Define the trimmed mean

$$\mu_k(R', R'') = \{\int_{R'}^{R''} ud\hat{F}_{\hat{\beta},k}(u|R')\}/\hat{F}_{\hat{\beta},k}(R''|R') - \hat{\alpha}. \tag{5.2}$$

As shown by Gross and Lai (1995), under (1.1) and certain regularity conditions, (5.2) is a consistent estimate of $E(\epsilon_j|c < \epsilon_j + \alpha < d)$ when ν is some fraction of n, where c and d are certain quantiles of the common distribution of the $\epsilon_j + \alpha$. Hence substantial differences among strata in their trimmed means (5.2) suggest possible departures from the regression model (1.1).

We illustrate this idea in the example below, where we also show how the influential-data diagnostics for linear regression in Belseley, Kuh and Welsch (1980) can be extended to l.t.r.c. data. In Section 6, this idea is applied again to truncated/censored data sets previously analyzed in the literature and reveals certain inadequacies of previously assumed linear regression models for these survival data.

Example 1 (continued). Taking $\nu = 2$ to define R' and R'' for (5.2), we found $R' = -8.96$ and $R'' = 4.78$ for the residuals associated with $(\hat{\alpha}^H, \hat{\beta}^H)$ in Example 1, and $R' = -9.54$, $R'' = 5.55$ for those associated with $(\hat{\alpha}^{BJ}, \hat{\beta}^{BJ})$. Figure 2 plots the residuals associated with $(\hat{\alpha}^H, \hat{\beta}^H)$ and their trimmed means over three subintervals of the range of observed x-values. These trimmed means, represented by solid line segments are close to 0 and show no trends of departure from the model. Also given for comparison are the trimmed means (dotted line segments) of the $(\hat{\alpha}^{BJ}, \hat{\beta}^{BJ})$-induced residuals, which are not shown in the figure. The dotted line segments are further from 0 than the solid line segments but again show no conspicuous departure from the regression model.

The contaminated normal errors ϵ_j in this example are likely to yield influential observations, and the presence of censoring and truncation requires adjustments/reconstructions such as in (3.3) which may give certain observations even more influence than in the complete data case. For the present data set, what are the influential observations if we use $(\hat{\alpha}^{BJ}, \hat{\beta}^{BJ})$? Is Huber's score function robust enough against these influential observations? To address this question, we can extend standard influential-data diagnostics for linear regression with complete data to l.t.r.c. data. The idea is to recompute, for each observation $(x_i^o, y_i^o, \delta_i^o, t_i^o)$, the M-estimator $(\hat{\alpha}_{(-i)}, \hat{\beta}_{(-i)})$ obtained by deleting this observation from the sample. Because of the computational efficiency of the algorithm in Section 3.2, the task of computing $(\hat{\alpha}_{(-i)}, \hat{\beta}_{(-i)})$ $i = 1, \cdots, n$ is quite manageable. Table 1 gives the values $\hat{\alpha} - \hat{\alpha}_{(-i)}$ and $\hat{\beta} - \hat{\beta}_{(-i)}$ for the M-estimators using Huber's score function and the Buckley-James score function. In addition, it also gives the difference in the fitted values $\hat{y}_i - \hat{y}_{(-i)}$, where $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i^o$ and $\hat{y}_{(-i)} = \hat{\alpha}_{(-i)} + \hat{\beta}_{(-i)} x_i^o$.

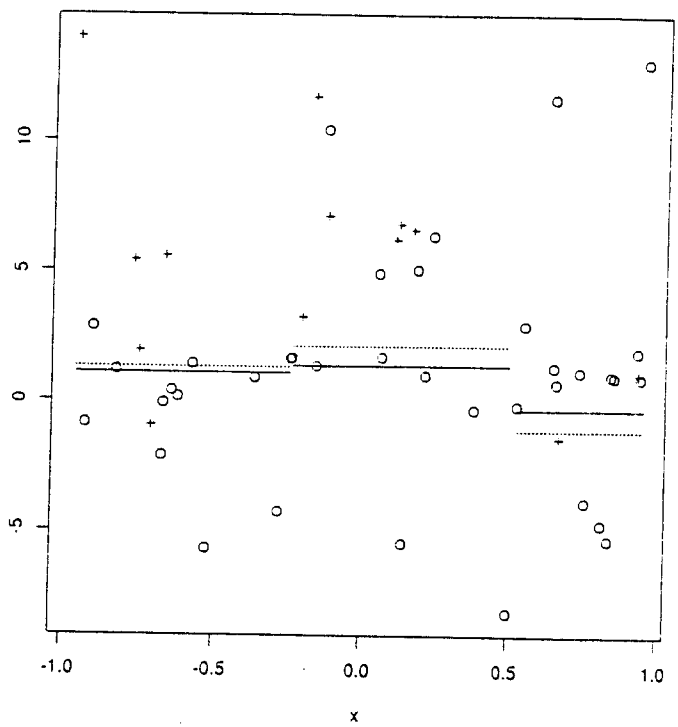

Figure 2. Plot of residuals based on Huber-type M-estimator versus x. The horizontal line segments represent the trimmed means of the residuals for three strata, with the solid lines corresponding to the M-estimator $(\hat{\alpha}^H, \hat{\beta}^H)$ and the dotted lines corresponding to $(\hat{\alpha}^{BJ}, \hat{\beta}^{BJ})$. o denotes an uncensored residual and + denotes a censored residual.

6. REGRESSION ANALYSIS OF SURVIVAL DATA

In this section we apply the methods in Sections 3 and 4 to two real data sets that have been previously analyzed in the literature using the regression model (1.1). In the first data set, only truncation is present, while only censoring is present in the second data set.

Table 1. Comparison of Huber-type M-estimator and Berkely-James estimator

i	x_i^o	$\hat{y}_i^o$	t_i^o	δ_i^o	$\hat{\alpha}^H - \hat{\alpha}^H_{(-i)}$	$\hat{\beta}^H - \hat{\beta}^H_{(-i)}$	$\hat{\alpha}^{BJ} - \hat{\alpha}^{BJ}_{(-i)}$	$\hat{\beta}^{BJ} - \hat{\beta}^{BJ}_{(-i)}$	$\hat{y}_i^H - \hat{y}^H_{(-i)}$	$\hat{y}_i^{BJ} - \hat{y}^{BJ}_{(-i)}$
1	-0.959	12.4	-7.52	0	0.184	-0.186	-0.001	-0.001	0.362	0.000
2	-0.925	-2.43	-5.72	1	-0.027	0.378	-0.092	-0.045	-0.377	-0.050
3	-0.902	1.31	-2.50	1	-0.240	-1.170	0.173	0.000	0.815	0.173
4	-0.823	-0.287	-4.69	1	-0.005	0.009	0.032	-0.046	-0.012	0.070
5	-0.766	3.96	-9.62	0	0.256	-0.395	1.260	-0.425	0.559	1.586
6	-0.746	0.517	-5.54	0	0.051	-0.192	1.030	-0.044	0.007	1.063
7	-0.706	-2.32	-4.31	0	-0.094	-0.048	0.014	-0.048	-0.060	0.048
8	-0.671	-3.48	-4.71	1	-0.043	0.554	1.190	1.120	-0.415	0.438
9	-0.666	-1.43	-18.6	1	0.279	0.110	0.227	-0.045	0.206	0.257
10	-0.664	4.22	-1.16	0	-0.013	-0.185	0.183	-0.002	0.110	0.184
11	-0.640	-0.925	-6.01	1	0.084	0.114	1.360	0.675	0.011	0.928
12	-0.619	-1.14	-5.98	1	-0.007	0.110	1.350	0.770	-0.075	0.873
13	-0.573	0.161	-1.64	1	-0.035	0.022	0.574	0.653	-0.048	0.200
14	-0.524	-6.93	-7.63	1	-0.563	0.324	1.600	1.350	-0.733	0.893
15	-0.366	-0.199	-6.16	1	0.106	0.018	0.048	0.000	0.099	0.048
16	-0.283	-5.29	-13.1	1	-0.078	0.289	1.670	0.875	-0.160	1.422
17	-0.248	0.630	-7.85	1	0.087	-0.084	0.247	0.000	0.108	0.247
18	-0.241	0.668	-6.41	1	-0.097	-0.185	0.067	-0.049	-0.052	0.079
19	-0.206	2.28	2.28	0	-0.003	0.001	-0.001	-0.001	-0.003	-0.001
20	-0.171	10.8	-8.54	0	0.264	-0.091	1.140	0.417	0.280	1.069
21	-0.157	0.424	-4.52	1	0.000	0.019	0.009	-0.001	-0.003	0.009
22*	-0.128	9.54	4.59	1	-0.132	-1.170	2.370	-0.317	0.018	2.411
23	-0.124	6.23	-10.6	0	0.168	-0.185	0.370	0.000	0.191	0.370
24	0.053	4.17	-1.05	1	-0.035	-0.046	0.062	-0.001	-0.037	0.062
25	0.065	0.948	-6.51	1	-0.017	-0.084	0.051	-0.045	-0.022	0.048
26	0.110	5.51	-21.9	0	0.295	0.019	0.242	-0.005	0.297	0.241
27	0.122	6.13	2.69	0	0.050	0.022	0.042	0.000	0.053	0.042
28	0.140	-6.16	-8.22	1	-0.316	0.036	0.952	0.770	-0.311	1.060
29	0.170	5.94	2.63	0	0.052	0.022	0.034	-0.006	0.056	0.033
30	0.184	4.43	-5.03	1	0.046	0.035	1.170	0.598	0.052	1.280
31	0.215	0.373	-3.33	1	0.071	0.020	-0.028	-0.045	0.075	-0.038
32	0.237	5.76	-4.50	1	0.027	0.021	1.540	0.667	0.032	1.698
33	0.381	-0.800	-4.24	1	0.012	-0.067	0.080	0.000	-0.014	0.080
34	0.500	-8.54	-6.20	1	-0.882	-0.181	-0.731	-0.001	-0.973	-0.732
35	0.528	-0.538	-9.53	1	0.201	0.018	0.266	0.000	0.211	0.266
36	0.551	2.58	-6.37	1	-0.106	0.021	1.050	0.630	-0.094	1.397
37*	0.641	11.4	-5.45	1	0.002	0.161	1.570	1.180	0.105	2.326
38	0.649	1.06	-10.0	1	0.186	0.115	0.187	0.000	0.261	0.187
39	0.658	0.440	-2.28	1	-0.012	-0.087	0.140	0.000	-0.069	0.140
40	0.667	-1.66	-3.69	0	0.041	0.020	1.580	0.487	0.054	1.905
41	0.735	0.953	-4.02	1	-0.011	-0.073	0.172	-0.002	-0.065	0.171
42	0.754	-4.02	-14.5	1	-0.037	-0.327	2.360	-0.316	-0.284	2.122
43	0.809	-4.84	-11.6	1	-0.063	-0.298	2.080	-0.317	-0.304	1.824
44	0.833	-5.42	-5.81	1	-0.243	-0.308	1.830	-0.316	-0.500	1.567
45	0.841	0.897	-0.185	1	-0.074	-0.185	0.349	0.000	-0.230	0.349
46	0.851	0.855	-10.7	1	0.211	0.115	0.202	-0.045	0.309	0.164
47	0.929	1.89	-6.82	1	-0.020	0.115	0.029	-0.045	0.087	-0.013
48	0.932	1.03	-8.28	0	0.197	0.324	1.970	0.876	0.499	2.786
49	0.940	0.889	-4.22	1	0.023	-0.054	0.169	0.000	-0.028	0.169
50*	0.951	13.0	-8.09	1	0.417	0.465	2.940	1.510	0.859	4.376

6.1 CDC Data on Truncation-Related AIDS

Table 1 of Kalbfleisch and Lawless (1989) reports the data from the centers of Disease Control (CDC) on 295 blood transfusion patients who were diagnosed with AIDS prior to July 1, 1986. It gives the month of infection (INF), with month 1 being January 1978, the age in years at the time of blood transfusion (AGE=age+1), and the duration in months (DIAG) between infection with the human immunodeficiency virus by blood transfusion and the clinical manifestation of acquired immune deficiency syndrome (AIDS). Kalbfleisch and Lawless (1989) suggested using INF–0.5 as the time of infection and excluding two patients with INF exceeding 90 months from the study, leaving 293 patients with DIAG that exceeds 101–(INF–0.5), where 101 is the total number of months in the study.

Gross (1994) fitted the linear regression model (1.1) to these data, taking $y = -\log(\text{DIAG})$ (with the modification that sets DIAG to be .5 if DIAG =0), $x = -1$ if age ≤ 4 and $x = 1$ if age > 4. Since $y = -\log(\text{DIAG})$ is left-truncated by $t = -\log(101.5-\text{INF})$, Gross used the Tsui-Jewell-Wu (1988) method and the rank regression method of Lai and Ying (1991) to estimate the slope β. The Tsui-Jewell-Wu method referred to in Section 1 is a special case of (3.2)-(3.3) with $\psi(u) = u$ (which we denoted by BJ in Example 1) and $c_i \equiv \infty$ (no censoring). Gross used a robustified modification, which she denotes by TJW, of this method. The rank regression method she used corresponds to (4.2) with the Wilcoxon-type score function $\psi(w) = 1 - w$. To get rid of ties to 0 and simplify the computation, Gross added independent random variables that are uniform on $(-0.5, 0.5)$ to all the DIAG values and used extensive search to find the TJW and rank estimators of β.

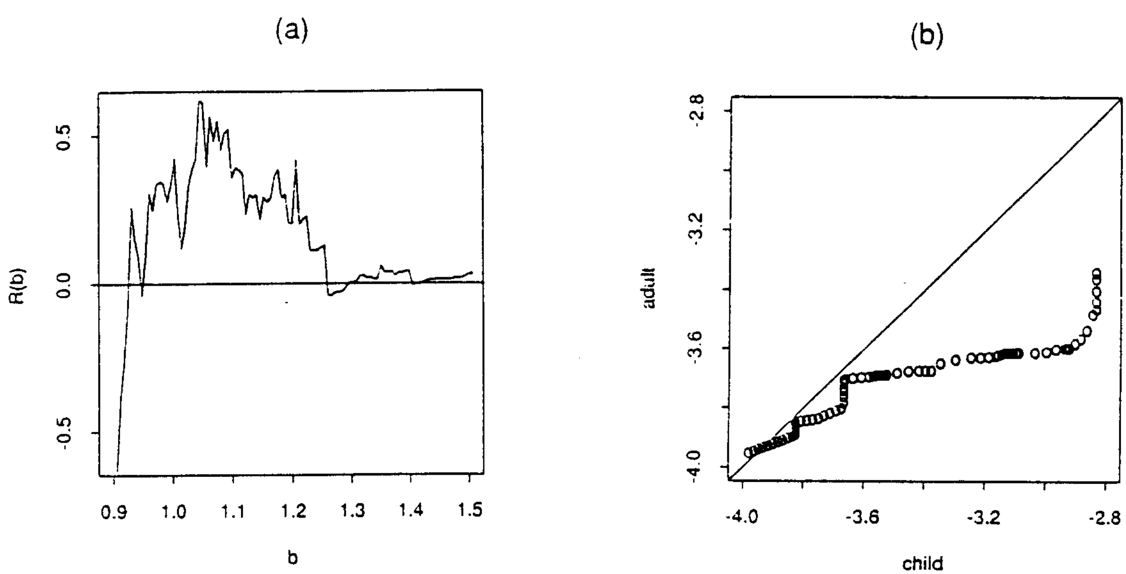

Figure 3. (a) Zero-crossings of rank estimator ; (b) Q-Q plot of residuals of two age groups : child (age ≤ 4) and adult (age > 4).

Without altering the data via this tie-breaking device, we recomputed the rank and the M-estimates. For the rank estimator, Gross found three zero-crossings of the function $R(b)$, evaluated from the altered data, at $b = -0.89$, -0.49, 0.92. She pointed out that since negative values of b do not make much sense, one should take 0.92 as the rank estimator. Using the original data instead, we evaluated $R(b)$ over a fine grid in $[0, 1.5]$. Figure 3 (a) plots the function for $0.9 \leq b \leq 1.5$, showing zero-crossings at $b = 0.92$, 0.95, 1.25, 1.30, 1.40. Note that $R(b)$ is essentially zero for $b \geq 1.25$. Taking $m = 1$ and $s = 2$ in (2.1)-(2.4), we found the preliminary estimator to be $\tilde{\alpha} = -3.58$ and $\tilde{\beta} = 0.53$ (which is in good agreement with the value $\tilde{\beta} = 0.59$ from her altered data reported by Gross). Using this preliminary estimate, we used the algorithm in Section 3.2 and found $(\hat{\alpha}^{BJ}, \hat{\beta}^{BJ})=(-3.71, 0.93)$ and $(\hat{\alpha}^{H}, \hat{\beta}^{H})=(-2.86, 1.58)$, using $r = 2$ and the same notation as in Example 1. In evaluating the Huber-type M-estimator, we simply set $h_{ii} = 0$ since the covariate only takes the values -1 and 1.

Gross (1994) noted that the rank estimate $\hat{\beta} = 0.92$ (from both her computation and ours) appears to be too high in comparison with another estimate of around 0.55 obtained from the parametric modeling of Kalbfleish and Lawless (1988), while her weighted least-squares estimate $\tilde{\beta} = 0.59$ (or $\tilde{\beta} = 0.53$ from the unaltered data) appears to be much closer. We use the method of stratified residuals in Section 5 to examine the adequacy of the assumed model. There are two natural strata, corresponding to $x = -1$ (or age ≤ 4) and $x = 1$ (or age > 4). Taking $R' = -3.95$ in (5.2) which corresponds to $\nu = 3$ in the notation of Section 5, we computed the product-limit curves $\hat{F}_{\hat{\beta},x=-1}(u|R')$ and $\hat{F}_{\hat{\beta},x=1}(u|R')$. Replacing these two discrete distribution by their histograms, we computed various quantiles of the distribution up to $u \leq R''$ (as defined in Section 5). Figure 3 (b), which gives the Q-Q plot of these quantiles, shows marked departure from the diagonal line, showing that $\hat{F}_{\hat{\beta},x=-1}(\cdot|R')$ tends to be stochastically smaller than $\hat{F}_{\hat{\beta},x=1}(\cdot|R')$.

The assumption of i.i.d. ϵ_j in (1.1) does not seem to be valid. Since there are 33 patients with age ≤ 4 and 260 patients with age > 4, the sample sizes appear to be adequate for treating the two groups separately, as in Gross and Lai (1995), without making the questionable assumption in (1.1) that the two groups have the same distribution after adjusting for a mean difference between the two groups.

6.2 Stanford Heart Transplant Data

Table 1 of Miller and Halpern (1982) gives the failure times in days, as of February 1980, of 152 patients who survived at least 10 days after entering the Stanford heart transplant study. Also given are each patient's age at the time of the transplant and T5 mismatch score. Miller and Halpern (1982), Leurgans (1987) and Lin and Geyer (1992) fitted the linear regression model (1.1) with $y = \log_{10}$(failure time) and $x =$ age or $x =$ (age, mismatch score) or $x =$ (age, age^2) to these data (which are not subject to truncation). Focusing on the case $x =$ (age, mismatch score), we took $s = 2$ and followed Leurgans' (1987) stratification of the data into $m = 4$ groups, obtaining the preliminary estimate $\tilde{\alpha} = 2.21$, $\tilde{\beta}_1 = 0.00085$ and $\tilde{\beta}_2 = 0.092$. We then used the algorithm in Section 3.2 to compute M-estimators and obtained the results $(\hat{\alpha}^{BJ}, \hat{\beta}_1^{BJ}, \hat{\beta}_2^{BJ}) = (3.20, -0.014, 0.00045)$, which is in close agreement with the values obtained with a somewhat less accurate scheme by Miller and Halpern (1982) except for $\hat{\beta}_2^{BJ}$, and $(\hat{\alpha}^{H}, \hat{\beta}_1^{H}, \hat{\beta}_2^{H}) = (3.92, -0.028, 0.014)$. Examination of the residuals using the method in Section 5 suggests possible nonlinearities in the relationship between log-survival and age. Since there is no truncation, we do not have to condition on R' in 5.2), which is tantamount

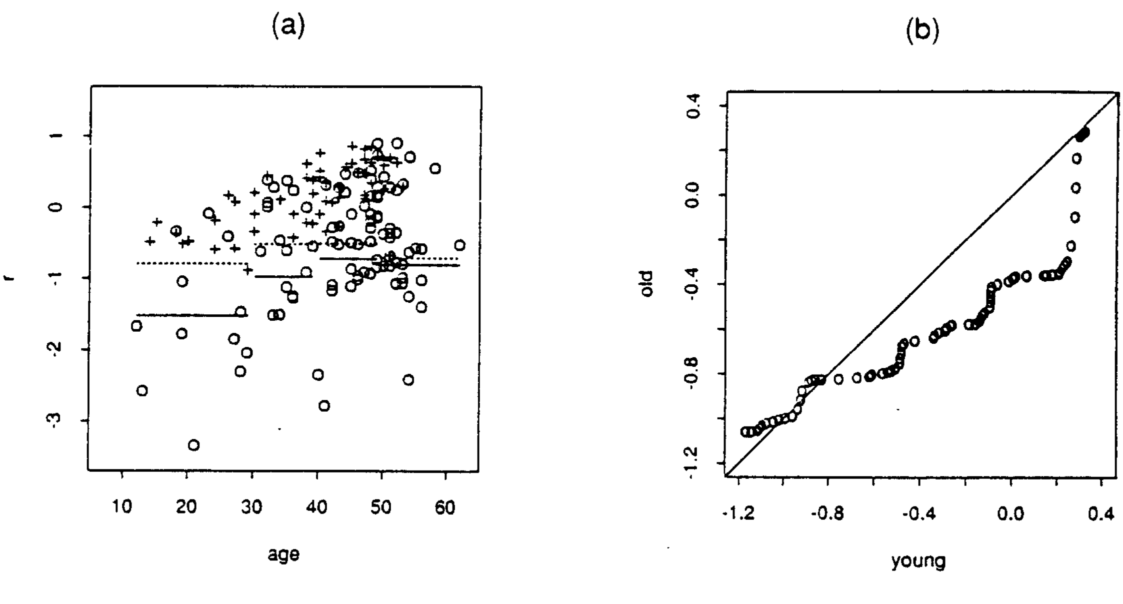

Figure 4. (a) Plot of residuals based on Huber-type M-estimator versus age. The horizontal line segments represent the trimmed means of the residuals for four strata, with the solid lines corresponding to the M-estimator $(\hat{\alpha}^{H}, \hat{\beta}_1^{H}, \hat{\beta}_2^{H})$ and the dotted lines corresponding to $(\hat{\alpha}^{BJ}, \hat{\beta}_1^{BJ}, \hat{\beta}_2^{BJ})$. o denotes an uncensored residual and + denotes a censored residual ; (b) Q-Q plot of residuals of two age groups : child (age < 50) and adult (age $\geq$ 50).

to taking $R' = -\infty$. For Huber's M-estimator, taking $\nu = 10$ gives $R'' = 3.55$ in (5.2). Figure 4 (a) plots these residuals versus age and the trimmed means (5.2) over four subintervals of the age range used by Leurgans (1987). It shows an upward trend followed by a downward trend. Dividing the data into two groups which correspond to age < 50 and to age ≥ 50, Figure 4 (b) gives the Q-Q plot of the quantiles of the Kaplan-Meier curves of these two groups of residuals. The Q-Q plot lies substantially below the diagonal line, revealing inadequacies of the linear regression model of log-survival on age and mismatch score.

7. CONCLUSION

In this paper we have introduced a relatively simple method to compute M-estimators of regression parameters from l.t.r.c. 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 a 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_j. Starting with a good preliminary estimate described in Section 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.

Even though the set of residuals based on l.t.r.c. data can only be regarded as approximations of the observable subset of an i.i.d. sample that is subject to left truncation and right censoring, suitable stratification of these residuals and associated graphical summaries can reveal departures from the assumed regression model. The usefulness of this idea is illustrated on two real data sets to which regression models have been previously fitted.

REFERENCES

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

(2) Buckley, J. and James, I. (1979), Linear regression with censored data, *Biometrika* **66**, 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*, **30**, 248–275.

(4) Dempster, 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*, **39**, 1–38.

(5) Gross, S. (1994), Weighted estimation in linear regression for truncated survival data, To appear in *Scandinavian Journal of Statistics.*

(6) 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.*

(7) Huber, P.J. (1973), Robust regression: Asymptotics, conjectures and Monte Carlo, *Annals of Statistics*, **1**, 799–821.

(8) Huber, P.J. (1977), *Robust Statistical Procedures*, Philadelphia: Society of Industrial and Applied Mathematics.

(9) Huber, P.J. (1981), *Robust Statistics*, New York: Wiley.

(10) Jurečkova, J. (1969), Asymptotic linearity of a rank statistic in regression parameter, *Annals of Mathematical Statistics*, **40**, 1889-1900.

(11) Jurečkova, J. (1971), Nonparametric estimates of regression coefficients, *Annals of Mathematical Statistics*, **42**, 1328–1338.

(12) Kalbfleisch, J.D. and Lawless, J.F. (1989), Inference based on retrospective ascertainment: An analysis of data on transfusion-related AIDS, *Journal of the American Statistical Association*, **84**, 360–372.

(13) Kay, R. (1977), Proportional hazard regression models and the analysis of censored survival data, *Applied Statistics*, **26**, 227–237.

(14) Lai, T.L. and Ying, Z. (1991a),Estimating a distribution function with truncated and censored data, *Annals of Statistics*, **19**, 417–442.

(15) Lai, T.L. and Ying, Z. (1991b), Rank regression methods for left-truncated and right-censored data, *Annals of Statistics*, **19**, 531–556.

(16) 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*, **22**, 1222–1255.

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

(18) Lin, D.Y. and Geyer, C.J. (1992), Computational methods for semiparametric linear regression with censored data, *Journal of Computational and Graphical Statistics*, **1**, 77–90.

(19) Miller, R.G. and Halpern, J. (1982), Regression with censored data, *Biometrika*, **69**, 521–531.

(20) Ritov, Y. (1990), Estimation in a linear regression model with censored data, *Annals of Statistics*, **18**, 303–328.

(21) Tsui, K.L., Jewell, N.P. and Wu, C.F.J. (1988), A nonparametric approach to the truncated regression problem, *Journal of the American Statistics Association*, **83**, 785–792.