The Korean Communications
in Statistics Vol. 3, No. 3, 1996
pp. 113-123

Efficient Score Estimation and Adaptive Rank and M-estimators from Left-Truncated and Right-Censored Data

Chul-Ki Kim[1)]

Abstract

Data-dependent (adaptive) choice of asymptotically efficient score functions for rank estimators and M-estimators of regression parameters in a linear regression model with left-truncated and right-censored data are developed herein. The locally adaptive smoothing techniques of Müller and Wang (1990) and Uzunoğullari and Wang (1992) provide good estimates of the hazard function h and its derivative h' from left-truncated and right-censored data. However, since we need to estimate h'/h for the asymptotically optimal choice of score functions, the naive estimator, which is just a ratio of estimated h' and h, turns out to have a few drawbacks. An alternative method to overcome these shortcomings and also to speed up the algorithms is developed. In particular, we use a subroutine of the PPR (Projection Pursuit Regression) method coded by Friedman and Stuetzle (1981) to find the nonparametric derivative of log(h) for the problem of estimating h'/h.

1. Introduction

Consider the linear regression model

$$y_i = \beta^T x_i + \varepsilon_i \quad (i=1,2,\cdots), \tag{1.1}$$

where the εi are i.i.d. random variables representing unobservable disturbances and having a common continuously differentiable distribution function F with density f, β is a d×1 vector of unknown parameters and the xi are either nonrandom or independent d×1 random vectors independent of {εn}. Suppose that the responses yi in (1.1) are not completely observable due to left truncation and right censoring by random variables t_i and c_i such that $\infty > t_i \geq -\infty$ and $-\infty < c_i \leq \infty$. Let $\widetilde{y_i} = y_i \wedge c_i$ and $\delta_i = I_{(y_i \leq c_i)}$, where we use $\wedge$ and $\vee$ to denote minimum and maximum, respectively. In addition to right censorship of the responses y_i by c_i, we shall also assume left truncation in the sense that $(\widetilde{y_i}, \delta_i, x_i)$ can be observed only when $\widetilde{y_i} \geq t_i$. The data, therefore, consist of n observations $(\widetilde{y_i^o}, t_i^o, \delta_i^o, x_i^o)$ with $\widetilde{y_i^o} \geq t_i^o$, $i=1,\cdots,n$.

---

1) Assistant Professor, Department of Statistics, Ewha Womans University, Seoul, 120-750, Korea.

It is usually assumed that (ti,ci,xi) are independent of the sequence $\{\varepsilon n\}$. The special case $ti \equiv -\infty$ corresponds to the "censored regression model" which is of basic importance in statistical modelling and analysis of failure time data (cf. Kalbfleisch and Prentice, 1980; Lawless, 1982). The special case $ci \equiv \infty$ corresponds to the "truncated regression model" in econometrics (cf. Tobin, 1958; Goldberger, 1981; Amemiya, 1985) and in astronomy (cf. Segal, 1975; Nicoll and Segal, 1980), which assumes the presence of truncation variables τi so that (xi,yi) can be observed only when $yi \le \tau i$ (or equivalently, when $-y_i \ge -\tau_i = t_i$).

Instead of assuming the (t_i, c_i, x_i) to be independent so that the sample $\{(\widetilde{y_i^o}, t_i^o, c_i^o, x_i^o): 1 \le i \le n\}$ can be regarded as having been generated by a larger, randomly stopped sample of independent random vectors (y_i, t_i, c_i, x_i), $1 \le i \le m(n) \triangleq \inf\{m: \sum_{i=1}^{m} I(t_i \le y_i \wedge c_i) = n\}$, an alternative setting proposed by Turnbull (1976) is to assume that (t_i^o, c_i^o, x_i^o) are independent random vectors that are independent of $\{\varepsilon_n\}$ and such that $c_i^o \ge t_i^o$ and

$$\tau_0 = 0, \quad \tau_j = \inf\{i > \tau_{j-1}: y_i \ge t_j^o\}, \quad \widetilde{y_j^o} = y_{\tau_j} \wedge c_j^o, \tag{1.2}$$
$$(t_i, c_i, x_i) = (t_j^o, c_j^o, x_j^o) \text{ for } \tau_{j-1} < i \le \tau_j.$$

In this formulation, $(\widetilde{y_i^o}, t_i^o, c_i^o, x_i^o)$ are independent random vectors such that the conditional distribution of $\widetilde{y_i^o}$ given (t_i^o, c_i^o, x_i^o) is

$$P\{\widetilde{y_i^o} \le y | t_i^o, c_i^o, x_i^o\} = \{F(y - \beta^T x_i^o) - F(t_i^o - \beta^T x_i^o)\} / \{1 - F(t_i^o - \beta^T x_i^o)\}, \quad y \ge t_i^o \tag{1.3}$$

Suppose that (t_i^o, c_i^o, x_i^o) are i.i.d. random vectors whose distributions do not depend on β and that conditional distribution of $\widetilde{y_i^o}$ given (t_i^o, c_i^o, x_i^o) is determined by (1.3). Under the assumption that the density function f of F is known, the maximum likelihood estimator of β is asymptotically normal with mean 0 and covariance $n^{-1}V_f$, where V_f^{-1} is the Fisher information matrix. Without assuming f to be known, it is shown in Lai and Ying (1992b) that adaptive estimators can nevertheless be constructed so that they are asymptotically normal with mean 0 and covariance $n^{-1}V_f$ when x_i has mean 0 and is independent of $(t_i - \beta^T x_i, c_i - \beta^T x_i)$. In general, these estimators may have larger asymptotic covariance matrices (in the sense of nonnegative definite differences) than $n^{-1}V_f$, but can still be shown to attain the asymptotically minimal covariance matrix for the asymptotic distributions of regular estimators. Such optimality results follow from the generalization of the Hájek convolution theorem and asymptotic minimax bounds to semiparametric models by Begun, Hall, Huang and Wellner (1983), since the (t_i^o, c_i^o, x_i^o) are assumed to be i.i.d. random vectors. Lai and Ying (1992b) have further removed the restrictive assumption that the (t_i^o, c_i^o, x_i^o) be identically distributed, which excludes the important case of nonrandom t_i^o, c_i^o and x_i^o, and have also developed asymptotic lower bounds for minimax risks in the general setting where (t_i^o, c_i^o, x_i^o) are only assumed to be independent. Moreover, they also consider the setting of independent (t_i, c_i, x_i) instead of independent (t_i^o, c_i^o, x_i^o) and develop asymptotic lower bounds in this alternative setting.

Lai and Ying (1991b,1992b,1994) have shown how to construct estimators that are asymptotically efficient in either settting, in the sense that the covariance matrix of the asymptotically normal distribution coincides with that given by the asymptotic lower bound. The basic idea involves estimating the score function h'/h for rank estimators and the score function $h'/h-h$ for M-estimators with l.t.r.c. data, where h and h' are the hazard function and its derivative, respectively. We shall make use of an alternative method that uses extensions of locally adaptive hazard smoothing to l.t.r.c. data (cf. Müller and Wang, 1990; Uzunoğullari and Wang, 1992) and the idea of estimating h'/h from $\log(h)$, as mentioned in Section 2. Section 2 shows how to estimate the hazard function h, its derivative h' and their ratio h'/h in practice. In Section 3 we apply these techniques in finding asymptotically efficient score functions for rank estimators and M-estimators in a linear regression model with l.t.r.c. data.

2. Locally Adaptive Hazard Smoothing with L.T.R.C. Data

Müller and Wang (1990) considered the problem of local bandwidth choice for nonparametric kernel estimation of a hazard function and its derivative under censoring. When estimating the hazard function or its derivative, they proposed that bandwidths should be chosen locally by adapting to the local Mean-Squared Error (MSE). They also showed that such locally adaptive bandwidth choice is indeed feasible and proposed asymptotically efficient methods to achieve it. To estimate the hazard function or its derivative, the convolution of the Nelson (1972) estimator with a kernel function was considered. The convolution can be carried out numerically by using the Fast Fourier Transformation (FFT). Later Uzunoğullari and Wang (1992) studied the estimation of the hazard rate function for left-truncated and right-censored data based on kernel-smoothing methods and presented its asymptotic properties including consistency, asymptotic normality and asymptotic formulas for the MSE to faciliate locally adaptive bandwidth choice.

Let $Y_1, Y_2, \cdots$ be independent random variables having a common distribution function F. Let (T_i, C_i) be i.i.d. random vectors that are independent of the Y_i. Let $\widetilde{Y}_i = Y_i \wedge C_i$ and $\delta_i = I_{(Y_i \le C_i)}$. In the presence of the right-censoring and left-truncation variables C_i and T_i, the Y_i are not completely observable; one only observes $(\widetilde{Y}_i, \delta_i)$ when $\widetilde{Y}_i \ge T_i$. Thus, the data consist of n observations $(\widetilde{Y}_i^o, \delta_i^o, T_i^o)_{1 \le i \le n}$ with $\widetilde{Y}_i^o \ge T_i^o$ for each i obtained from a larger sample $(Y_i, T_i, C_i)_{1 \le i \le m(n)}$. In the context of survival analysis, the Y_i are survival times or functions of survival times (such as their logarithms). Our aim is to estimate $h(x) = H'(x) = f(x)/(1-F(x))$, the hazard function (where H is a cumulative hazard function of the Y_i and $f=F'$ is assumed to exist). Let $\overline{L}_1(x) = P(T_i^o \le x \le \widetilde{Y}_i^o, \delta_i^o = 1 \mid \widetilde{Y}_i^o \ge T_i^o)$ be the conditional subsurvival function for the uncensored observations and let

$$\bar{L}_{1n}(x)=(n+1)^{-1}\sum_{i=1}^{n}I(T_i^o\le x\le \widetilde{Y}_i^o,\ \delta_i^o=1) \tag{2.1}$$

be the corresponding modified empirical subsurvival function. Similarly, denote the modified empirical survival function of $\bar{L}(x)=P(T_i^o\le x\le \widetilde{Y}_i^o \mid \widetilde{Y}_i^o\ge T_i^o)$ by

$$\bar{L}_n(x)=(n+1)^{-1}\sum_{i=1}^{n}I(T_i^o\le x\le \widetilde{Y}_i^o). \tag{2.2}$$

Using the fact that

$$\frac{dL_1(x)}{\bar{L}(x)}=h(t), \tag{2.3}$$

where for any distribution function G, $\bar{G}$=1−G under l.t.r.c., the analog of the Nelson (1972) estimator of $H(x)$ is

$$H_n(x)=\int_o^x[1-L_n(y)]^{-1}dL_{1n}(y). \tag{2.4}$$

To estimate $h^{(\nu)}(x)$, we consider the following kernel estimate, which is a convolution of the Nelson estimator H_n with an appropriate kernel function K_ν:

$$\hat{h}^{(\nu)}=\frac{1}{b^{\nu+1}}\int K_\nu\left(\frac{x-u}{b}\right)dH_n(u)=\frac{1}{b^{\nu+1}}\sum_{i=1}^{n}K_\nu\left(\frac{x-\widetilde{y}^o_{(i)}}{b}\right)\delta^o_{(i)}/n_{(i)}. \tag{2.5}$$

Here $\widetilde{y}^o_{(i)}$ are the order statistics of $\widetilde{Y}_i^o$; if $\widetilde{y}^o_{(i)}$ is uncensored, $\delta^o_{(i)}=1$, otherwise 0; and $n_{(i)}$ is the size of the risk set at $\widetilde{y}^o_{(i)}$, i.e. $n_{(i)}=\sum_{j=1}^{n}I(t_j^o\le \widetilde{y}^o_{(i)}\le \widetilde{y}^o_j)$. Further, $b=b(n)$ is a sequence of bandwidths for which we require

$$b\to 0,\ nb^{2\nu+1}\to\infty,\ \frac{nb}{(\log n)^2}\to\infty,\quad as\ \ n\to\infty,$$

and K_ν is a kernel function of bounded variation and is of order (ν,k) with support [-1,1], i.e. satisfying

$$K_\nu\in M_{\nu,k}=\left\{f\in L^2([-1,1]):\int f(x)x^jdx=\begin{cases}(-1)^\nu\nu! & j=\nu\\ 0 & 0\le j< k,\ j\ne\nu\\ \ne 0 & j=k\end{cases}\right\}$$

To estimate the bias and the variance (as functions of b) for the case ν=0, one can use the following analogues of the Müller and Wang (1990) estimates:

$$\hat{\beta}(x,b)=\int\hat{h}(x-by)K_0(y)\,dy-\hat{h}(x) \tag{2.6}$$

for the bias, and

$$\hat{v}(x,b)=\frac{1}{nb}\int K_0(y)^2\frac{\hat{h}(x-by)}{\bar{L}_n(x-by)}\,dy \tag{2.7}$$

for the variance. Note that $\hat{\beta}$ and $\hat{v}$ can be evaluated numerically by discrete approximations to these convolution integrals. We applied the FFT for computation in the simulation study. Consequently, the proposed local bandwidth estimate for the case ν=0 can be obtained as a minimizer of

$$\widehat{MSE}(x, b) = \hat{v}(x, b) + \hat{\beta}^2(x, b) \tag{2.8}$$

with respect to b, which we denote by $\hat{b}_0$. The algorithmic implementation of the proposed method requires a number of choices to be made. These concern $\hat{b}$, the pilot bandwidth to compute $\hat{h}$ via (2.5) for (2.6) and (2.7), and $b^{(1)}, b^{(2)}$, the bounds between which a minimizer is sought. Once $\hat{b}$ is initialized by any available option, it is recommended to choose $b^{(1)} = 0.5\,\hat{b}, b^{(2)} = 2.0\,\hat{b}$ (cf. Müller and Wang, 1990; Uzunoḡullari and Wang, 1992).

Similarly, one can estimate h', the derivative of a hazard function, as discussed in Müller and Wang (1990) and Uzunoḡullari and Wang (1992). Figure 1 illustrates the preceding algorithm. As we see in the plots, the estimates of $h^{(\nu)}(x)$ for ν=0, 1 have a somewhat poor performance at small risk set size near the edges of the interval.

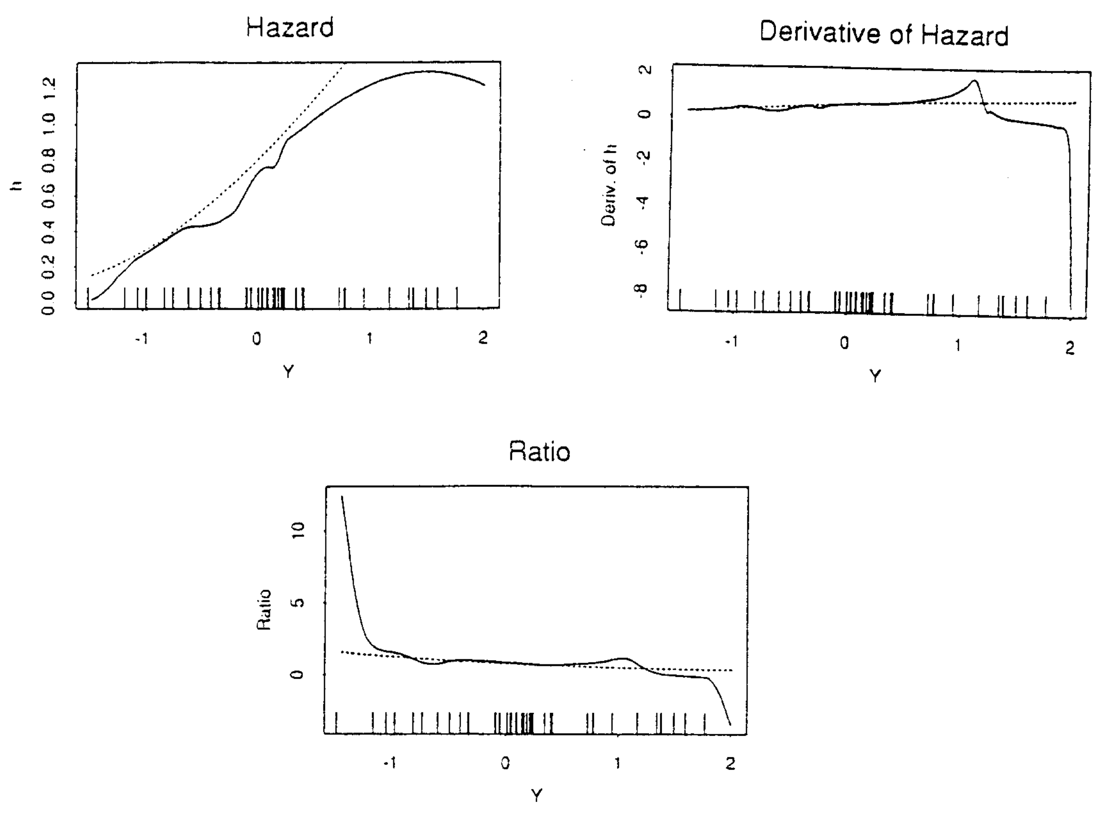

Figure 1. Estimates of hazard function h, its derivative h' and their ratio h'/h are represented by the solid lines; the actual ones, by the dotted lines. The uncensored data points over the interval are shown by the vertical strokes.

These numerical results were obtained from the following simulation setting with sample size 200:

$Y_i \overset{iid}{\sim} N(0,1)$ and independently $T_i \overset{iid}{\sim} N(0,1)$,
$C_i = T_i + (\exp(-T_i) \vee 0.5) U_i$ with $U_i \overset{iid}{\sim}$ Unif[0,5],
42% are left truncated; 22% of these are right censored,
and a kernel function $K_0(y) = \begin{cases} (15/16)(1-y^2)^2 & \text{if } |y| \le 1 \\ 0 & \text{if } |y| > 1 \end{cases}$ was used.

The estimate of the ratio h'/h in the third panel of the figure will be used for the adaptive estimation problem of the slope β in a linear regression model with l.t.r.c. data. Alternatively, we can compute the nonparametric derivative of $\log(h)$, which is h'/h, by using the subroutine *der* of PPR (Friedman and Stuetzle, 1981).

3. Efficient Score Functions for Rank and M-estimators

Consider the problem of estimating the slope β in the regression model (1.1) with a single parameter. Recalling the notation described in Section 1, a starting point of the development in Lai and Ying (1991b) is the following general class of rank statistics formed from the residuals $e_i(b) = \widetilde{y_i^o} - bx_i^o$. Let $e_{(1)}(b) \le \cdots \le e_{(k)}(b)$ denote all the ordered uncensored residuals. For $i=1,\cdots,k$, let

$$J(i,b) = \{j \le n: t_j^o - bx_j^o \le e_{(i)}(b) \le \widetilde{y_j^o} - bx_j^o\}, \quad n_i(b) = \#J(i,b), \tag{3.1}$$
$$\bar{x}(i,b) = (\sum_{j \in J(i,b)} x_j^o)/n_i(b),$$

where the notation #A denotes the number of elements of set A. Let ψ be a twice continously differentiable function on (0,1) such that $\sup_{0<x<1} |\psi''(x)| < \infty$. Let p be a nondecreasing and twice continuously differentiable function on the real line such that

$$p(y) = 0 \text{ for } y \le 0 \text{ and } p(y) = 1 \text{ for } y \ge 1. \tag{3.2}$$

Take $0 < \lambda < 1/18$ and define $p_n(z) = p(n^\lambda(z - cn^{-\lambda}))$ for $0 \le z \le 1$. Define the product-limit estimator $\hat{F}_{n,b}$ and the rank statistic associated with ψ by

$$1 - \hat{F}_{n,b}(u) = \prod_{i: e_{(i)}(b) < u,\, \delta^o_{(i)} = 1} \{1 - p_n(n^{-1} n_i(b))/n_i(b)\} \tag{3.3}$$

and

$$S_n(b) = \sum_{i=1}^{k} \psi(\hat{F}_{n,b}(e_{(i)}(b))) p_n(n^{-1} n_i(b)) \{x_{(i)} - \bar{x}(i,b)\} \tag{3.4}$$

Lai and Ying (1991b) defined in the present setting, in which the responses yi are subject to left truncation and right censoring, a rank estimator $\hat{\beta}_n$ of β as a zero-crossing of $S_n(b)$. Thus $S_n(b)$ is a step function of b, but there is no guarantee that it is monotonic. Therefore there might be multiple solutions of $S_n(b)=0$.

By Theorem 2 in Lai and Ying (1991b), the rank estimator $\hat{\beta}_n$, defined as a zero-crossing of the linear rank statistic (3.4), is asymptotically normal $N(\beta, v/(A^2K^2n))$ as $n \to \infty$. Letting $h=f/(1-F)$ denote the hazard function of F, we express A as

$$A=\int_{\tau_0}^{\tau}\psi(\tilde{F}(s))[h'(s)/h(s)][G_2(s)-G_1^2(s)/G_0(s)]dF(s),$$

where for r=0,1,2 and with $F(x) < 1$

$$\lim_{m\to\infty} m^{-1}\sum_{i=1}^{m} E\{x_i^r P[t_i-\beta x_i \le s \le c_i-\beta x_i | x_i]\} = G_r(s),$$

$$\lim_{m\to\infty} m^{-1}\sum_{i=1}^{m} P\{t_i-\beta x_i \le c_i-\beta x_i < s\} = \overline{G}(s),$$

$$\tau_0=\inf\{s: G_0(s)>0\},\quad \tau=\inf\{s>\tau_0: (1-F(s))G_0(s)=0\},$$

$$\text{and}\quad \tilde{F}(s)=(F(s)-F(\tau_0))/(1-F(\tau_0))=P\{\varepsilon_1\le s|\varepsilon_1\ge\tau_0\}$$

Since $1/K=\int_{-\infty}^{\infty}(G_0+\overline{G})dF$ and $v=K\int_{\tau_0}^{\tau}\psi^2(\tilde{F}(s))[G_2(s)-G_1^2(s)/G_0(s)]dF(s)$, it then follows from the Schwarz inequality that

$$(AK)^{-2}v \ge \{\int_{-\infty}^{\infty}(G_0+\overline{G})dF\}\{\int_{\tau_0}^{\tau}(h'/h)^2(G_2-G_1^2/G_0)dF\}^{-1}, \tag{3.5}$$

and that equality of (3.5) holds also in the case

$$\psi(\tilde{F}(s))=h'(s)/h(s). \tag{3.6}$$

Since h is usually unknown in practice, Lai and Ying (1991b,1992b) studied how to use the observed data $(\widetilde{y_i^o}, t_i^o, \delta_i^o, x_i^o)$, $i=1,\cdots,n$, to estimate the asymptotically optimal score function (3.6) for the linear rank statistic (3.4) from which we obtain an asymptotically normal rank estimator $\hat{\beta}_n$ that achieves the lower bound in (3.5).

The basic idea of Lai and Ying (1991b,1992b) is to divide the sample into two disjoint subsets. One might randomly split it into two sets of data. From the first subsample, define the residuals $e_i(b)=\widetilde{y_i^o}-bx_i^o$ $(i\le n/2)$ and order the uncensored ones among them as $e_{(1)}(b)\le\cdots\le e_{(k_1)}(b)$. Let $n_1=[n/2]$, i.e. the largest integer$\le n/2$, and define $J(i,b)$, $n_i(b)$, $\bar{x}(i,b)$ as in (3.1) but with n_1 replacing n (i.e. on the basis only of the first sample). Analogous to (3.4), define

$$S_{n,1}(b)=\sum_{i=1}^{k_1}\psi_{n,2}(e_{(i)}(b))p_n(n^{-1}n_i(b))[x^o_{(i)}-\bar{x}(i,b)], \tag{3.7}$$

where $\psi_{n,2}(s)$ is an estimate of $h'(s)/h(s)$ from the second subsample of $n_2=n-n_1$ observations $(\widetilde{y_r^o}, t_r^o, \delta_r^o, x_r^o)$, $n_1<r\le n$. Likewise from the second subsample, define the residuals $e_i^*(b)=\widetilde{y^o}_{n_1+i}-bx^o_{n_1+i}$ $(i\le n_2)$ and order the uncensored ones among them as $e^*_{[1]}(b)\le\cdots\le e^*_{[k_2]}(b)$. As in (3.1), let $J^*(i,b)=\{n_1<r\le n: t_r^o-bx_r^o\le e^*_{[i]}\le \widetilde{y^o}_r-bx_r^o\}$ $n_i^*(b)=\# J^*(i,b)$, $\bar{x}^*(i,b)=(\sum_{r\in J^*(i,b)} x_r^o)/n_i^*(b)$. Define

$$S_{n,2}(b)=\sum_{i=1}^{k_2}\psi_{n,1}(e^*_{[i]}(b))p_n(n^{-1}n_i^*(b))[x^o_{n_1+[i]}-\bar{x}^*(i,b)], \tag{3.8}$$

where $\psi_{n,1}(s)$ is an estimate of $h'(s)/h(s)$ based on the first subsample. Combining the two subsample statistics (3.7) and (3.8) gives the linear rank statistic

$$S_n^*(b) = S_{n,1}(b) + S_{n,2}(b) \quad (3.9)$$

A crucial part of this adaptive method in applications is how to construct the estimator ψ_{nj} of h'/h based on the jth subsample (j=1,2). As discussed in the previous section, the locally adaptive hazard smoothing technique is used for the estimation of the score function. Replacing $\widetilde{Y^o}_i$ and T_i^o in the section by $\widetilde{y^o}_i - bx_i^o$ and $t_i^o - bx_i^o$ of the first subsample, we can estimate the hazard function h and its derivative h' by the locally adaptive choice of the bandwidth on an interval for which the lower and upper endpoints are the minimum and maximum of the residuals of the second subsample, and vice versa. Choosing k to be large relative to the length of interval, the estimator ψ_{nj} is evaluated at consecutive points that are at a distance of 2^{-k} apart; and we interpolate the associated value for each uncensored residual that lies between two such grid points.

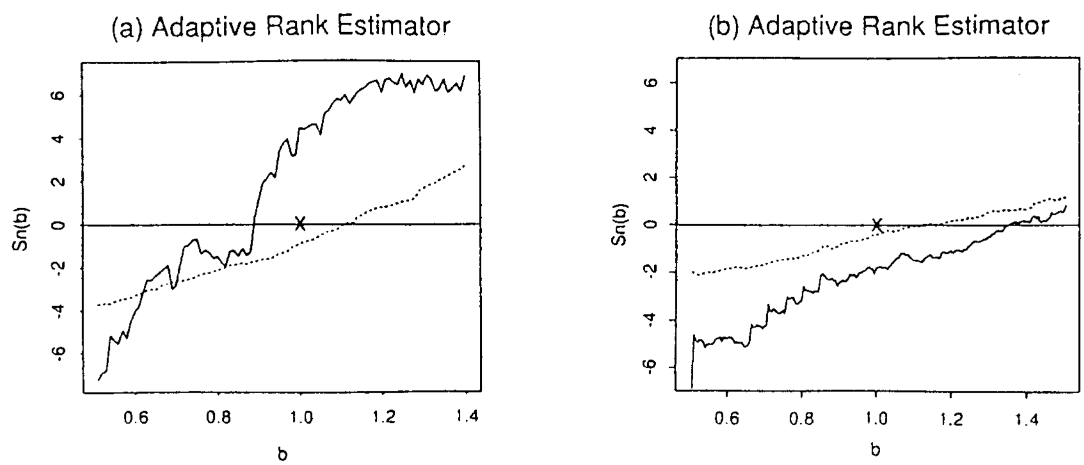

Figure 2. Plots of the adaptive rank statistics $S_n(b)$ versus b; the solid line represents the adaptive rank statistics $S_n(b)$ with the naive estimate (a) and the nonparametric estimate (b) of score function h'/h; the dotted line, the rank statistics $S_n(b)$ with the optimal score function of the known hazard function (see (3.4)). The true value of the slope β indicated by "X" is 1 and the zero-crossings are found near 1.

Figure 2 gives plots of the adaptive rank statistics $S_n(b)$ over an interval of b, based on the data obtained from the following simulation experiment:

$$y_i = \beta x_i + \varepsilon_i \quad (i = 1, \cdots, 100)$$

where

$$\varepsilon_i \overset{iid}{\sim} N(0,1),\ x_i \overset{iid}{\sim} \text{Uniform}[0,1],\ t_i \overset{iid}{\sim} N(0,1),$$
$$c_i = t_i + (\exp(-t_i) \vee 0.5) u_i \text{ with } u_i \overset{iid}{\sim} \text{Uniform}[0,5],$$

13% are left truncated; 24% of these are right censored.

A few problems in numerical computing are considered at this point: first, there is no closed form for $S_n(b)$; second, there might be multiple zeros of $S_n(b)$. A preliminary estimator of β is therefore used as a guide to find a proper interval of the slope b by using some other methods such as the nonparametric regression techniques introduced in Kim and Lai (1995) and the weighted M-estimators in regression analysis by Gross and Lai (1994a).

However, we observe a few drawbacks of the *naive* estimate of score function h'/h, which is the estimated h' to estimated h ratio: since the score function is sensitive to the estimated values of h as a denominator, we can see a jittery behavior of the adaptive rank statistic $S_n(b)$ over an interval of b, as shown in Figure 2 (a); furthermore, the estimation problem of h and h' involves FFT at every value of residuals and bandwidths, which slows down the whole computing procedure. Instead of using the naive estimate, we can run the subroutine *der* of PPR to nonparametrically differentiate $\log(h)$ and thus to obtain an adaptive estimate of score function h'/h. This resultant estimate of h'/h plays a role of troubleshooting for the adaptive estimation of the score function h'/h: with the alternative adaptive score function, the rank statistic $S_n(b)$ is less jittery over the interval of b, as shown in Figure 2 (b). Therefore, once h is obtained by using the locally adaptive smoothing technique, one might prefer the latter nonparametric estimate of h'/h to the naive one.

Analogous to the asymptotically efficient rank estimators described above, we can construct asymptotically efficient M-estimators. In fact, Lai and Ying (1994) showed consistency and asymptotic normality of a class of M-estimators in the left-truncated and right-censored regression model with known density of the errors and extended the idea of Lai and Ying (1991b) on adaptive choice of score functions in constructing asymptotically efficient rank estimators of β to M-estimators.

Since density f of the errors in model (1.1) is usually unknown, the optimal score function for M-estimators $\psi=(h'/h)-h$, which is f'/f, is not available to form the asymptotically efficient M-estimators. Therefore, we need to extend the idea described above for rank estimators to M-estimators: the sample is divided into two disjoint subsets, the first of which is $\{(\widetilde{y^o_i}, t^o_i, \delta^o_i, x^o_i): i \leq n/2\}$; from the first subsample, the root-finding statistics are constructed (cf. Section 5 of Lai and Ying, 1994) with $\hat{\psi}_{n,1}$ which is an estimate of $(h'/h)-h$ based on the second subsample of $n_2=n-n_1$ observations, and vice versa; analogous to (3.9), combining the two subsample statistics gives the statistic to find asymptotically optimal M-estimators. To adaptively find smooth consistent estimates $\hat{\psi}_{n,1}$ and $\hat{\psi}_{n,2}$ of $\psi=(h'/h)-h$

based on the two subsamples, we also use the locally adaptive smoothing techniques, as described in Section 2. Replacing $\widetilde{Y^o}_i$ and T_i^o in the section by $\widetilde{y^o}_i - bx_i^o$ and $t_i^o - bx_i^o$ of each subsample leads to the same problem of finding asymptotically optimal M-estimators as that of rank estimators described earlier in this section.

4. Conclusion

In addition to the nonparametric estimation of the score function h'/h, we can improve the root-finding scheme of (3.9) by taking an average of the zero-crossings obtained from a number of replications of the cross-validation method described in Section 3. However, for this heavy computing, there needs a much less expensive way to adaptively estimate the score functions for the rank and M-estimators.

References

[1] Amemiya, T. (1985). *Advanced Econometrics*. Harvard Univ. Press, Cambridge.

[2] Begun, J.M., Hall, W.J., Huang, W.M. and Wellner, J.A. (1983). Information and asymptotic efficiency in parametric nonparametric models. *Ann. Statist.* 11, 432-452.

[3] Friedman, J.H. and Stuetzle, W. (1981). Projection pursuit regression. *J. Am. Statist. Assoc.* 76, 817-823.

[4] Goldberger, A.S. (1981). Linear regression after selection. *J. Econometrics* 15, 357-366.

[5] Gross, S. and Lai, T.L. (1994a). Nonparametric estimation and regression analysis with left-truncated and right-censored data. *Tech. Report. L94-2*, Dept. of Statistics, Stanford University.

[6] Kalbfleisch, J.D. and Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

[7] Kim, C.K. and Lai, T.L. (1995). Generalized Additive Models for Left-Truncated and Right-Censored Data. *Tech. Report. L95-6*, Dept. of Statistics, Stanford University.

[8] Lai, T.L. and Ying, Z. (1991b). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* 19, 531-556.

[9] Lai, T.L. and Ying, Z. (1992b). Asymptotically efficient estimation in censored and truncated regression models. *Statistica Sinica* 2, 17-46.

[10] Lai, T.L. and Ying, Z. (1994). A missing information principle and M-estimators in regression analysis with censored and truncated data. *Ann. Statist.*, 22, 1222-1255.

[11] Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. Wiley, New York.

[12] Müller, H.G. and Wang, J.L. (1990). Locally adaptive hazard smoothing. *Probab. Theory Related Fields* 85, 523-538.

[13] Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* 14, 945-966.

[14] Nicoll, J.F. and Segal, I.E. (1980). Nonparametric estimation of the observational cutoff bias. *Astron. Astrophy.* 82, L3-L6.

[15] Segal, I.E. (1975). Observational validation of the chronometric cosmology. I. Preliminaries and the redshift magnitude relation. *Proc. Natl. Acad. Sci. USA* 72, 2437-2477.

[16] Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* 26, 24-36.

[17] Uzunoğullari, Ü and Wang, J.L. (1992). A comparison of hazard rate estimators for left-truncated and right-censored data. *Biometrika* 79, 297-310