LEAST ABSOLUTE DEVIATIONS ESTIMATION FOR THE ACCELERATED FAILURE TIME MODEL

Jian Huang[1], Shuangge Ma[2] and Huiliang Xie[1]

[1]*University of Iowa and* [2]*Yale University*

Abstract: The accelerated failure time (AFT) model assumes a linear relationship between event time and covariates. We propose a robust weighted least-absolute-deviations (LAD) method for estimation in the AFT model with right-censored data. This method uses the Kaplan-Meier weights in the LAD objective function to account for censoring. We show that the proposed estimator is root-n consistent and asymptotically normal under appropriate assumptions. It can also be easily computed using existing software, which makes it especially useful for data with medium to high dimensional covariates. The proposed method is evaluated using simulations and demonstrated on two clinical data sets.

Key words and phrases: Asymptotic normality, Kaplan-Meier weights, least absolute deviations, right censored data, robust regression.

1. Introduction

The accelerated failure time (AFT) model is a linear regression model in which the response variable is the logarithm or a known monotone transformation of a failure time (Kalbfleisch and Prentice (1980)). As a useful alternative to the Cox model (Cox (1972)), this model has a more intuitive linear regression interpretation, see Wei (1992) for a lucid discussion. Semiparametric estimation in the AFT model with an unspecified error distribution has been considered by many authors. Two methods have received special attention. One method is the Buckley-James estimator which adjusts censored observations using the Kaplan-Meier estimator in the least squares regression. The other is the rank-based estimator which is motivated by the score function of the partial likelihood, see for example, Prentice (1978), Buckley and James (1979), Ritov (1990), Tsiatis (1990), Wei, Ying and Lin (1990), Ying (1993) and Jin, Lin, Wei and Ying (2003), among others.

For uncensored data, the least-absolute-deviations (LAD) method has received much attention due to its robustness property (Bassett and Koenker (1978) and Koenker and Bassett (1978)). Powell (1984) and Newey and Powell (1990) proposed LAD estimators in regression models with censored response when the censoring variables are always observable. Ying, Jung and Wei (1995) proposed

a median regression estimator in the AFT model with right-censored response variable. They pointed out that, in addition to its robustness property, the LAD method is particularly attractive for the AFT model due to the simple fact that the median is well defined for censored data as long as censoring is not too heavy. Yang (1999) and Subramanian (2002) also considered median based regression methods for censored data. These estimators have rigorous theoretical justification under appropriate conditions. However, they are difficult to compute since the estimating equations for these estimators involve estimation of survival or hazard functions, which in turn involve the unknown regression parameters. Most of these approaches either demand a brutal searching procedure in a high dimensional coefficient space, or need to use stochastic algorithms such as simulated annealing (Lin and Geyer (1992)). However, brutal search or simulated annealing can be slow and difficult to implement. There appear to be no computer programs readily available for computing these estimators. This makes application of these methods difficult in practice.

In this paper, we propose a weighted LAD estimator in the AFT model using Kaplan-Meier weights. For simplicity of notation, we call this estimator the KMW-LAD estimator. Use of Kaplan-Meier weights to account for censoring was proposed by Stute (1993, 1996, 1999) in the context of least squares estimation in the AFT model. A seemingly different but equivalent formulation of the Kaplan-Meier weights was proposed by Zhou (1992). Zhou (1992) proved asymptotic results of his weighted least squares estimator. His proof used the explicit expression of the least squares estimator and is not applicable to the LAD estimator. Bang and Tsiatis (2002) also used this method in analyzing censored cost data. In a recent paper by Zhou (2005), motivated by the work of Ying, Jung and Wei (1995), a weighting method was also proposed in the context of censored median regression. This weighting method is exactly the same as that of Zhou (1992). Zhou (2005) also proposed using an artificial censoring in connection with the weights of Zhou (1992).

An important advantage of the proposed KMW-LAD estimator is that it can be computed using existing software. The computational simplicity is especially valuable for data with medium to high dimensional covariates. The KMW-LAD estimator also has rigorous theoretical justification under appropriate conditions. In the following, we first define the KMW-LAD estimator in the AFT model. In Section 3, we state the consistency and asymptotic normality results for the KMW-LAD estimator and discuss the assumptions needed for these results. These assumptions are different from but comparable to those for the existing estimators of the AFT model. In Section 4, we use simulations to evaluate the KMW-LAD estimator in finite samples and illustrate it using two clinical trial data sets. Some concluding remarks are given in Section 5.

2. The LAD Regression for Censored Data

Let T be the logarithm of the failure time and $X = (X_1, \ldots, X_d)'$ be a d-dimensional covariate vector. The AFT model assumes

$$T = \beta_0 + X_1\beta_1 + \cdots + X_d\beta_d + \epsilon, \tag{2.1}$$

where β_0 is the intercept, $\beta_1, \ldots, \beta_d$ are the regression coefficients and ϵ is the error term with an unknown distribution function. When T is subject to right censoring, we can only observe (Y, δ, X) with $Y = \min\{T, C\}$, where C is the logarithm of the censoring time and $\delta = 1\{T \leq C\}$ is the censoring indicator. Suppose that a random sample (Y_i, δ_i, X_i), $i = 1, \ldots, n$, with the same distribution as (Y, δ, X) is available.

Let $\widehat{F}_n$ be the Kaplan-Meier estimator of the distribution function F of T (Kaplan and Meier (1958)). Following Stute and Wang (1993), $\widehat{F}_n$ can be written as $\widehat{F}_n(y) = \sum_{i=1}^n w_{ni}1\{Y_{(i)} \leq y\}$, where the w_{ni}'s are the Kaplan-Meier weights

$$w_{n1} = \frac{\delta_{(1)}}{n}, \text{ and } w_{ni} = \frac{\delta_{(i)}}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1}\right)^{\delta_{(j)}}, \; i = 2, \ldots, n. \tag{2.2}$$

Here $Y_{(1)} \leq \cdots \leq Y_{(n)}$ are the order statistics of Y_i's and $\delta_{(1)}, \ldots, \delta_{(n)}$ are the associated censoring indicators. Similarly, let $X_{(1)}, \ldots, X_{(n)}$ be the associated covariates of the ordered Y_i's. Let $\beta = (\beta_0, \beta_1, \ldots, \beta_d)$. The KMW-LAD estimator $\widehat{\beta}_n$ is the minimizer of

$$L_n(\beta) = \sum_{i=1}^n w_{ni}|Y_{(i)} - \beta_0 - X_{(i1)}\beta_1 - \cdots - X_{(id)}\beta_d|. \tag{2.3}$$

Robustness is gained by using the LAD objective function. $\widehat{\beta}_n$ can be computed using the R function *rq* in the R library *quantreg*. The LAD regression program is also available in many other packages, such as the LAV command in the IML library, Proc Quantreg in SAS, and the quantile regression (qreg) procedure in STATA.

The weights w_{ni}'s are the jumps of the Kaplan-Meier estimator $\widehat{F}_n$. Another weighting method uses $w_{ni} = \delta_i/(1 - \widehat{G}_n(Y_i))$ (Zhou (1992), Bang and Tsiatis (2002) and Zhou (2005)), where $\widehat{G}_n$ is the Kaplan-Meier estimator of G, the distribution function of C_i. These two seemingly different weighting methods are equivalent. This can be shown as follows. By the property of the Kaplan-Meier estimator,

$$[1 - \widehat{F}_n(t)][1 - \widehat{G}_n(t)] = \frac{1}{n}\sum_{i=1}^n 1\{Y_i > t\}. \tag{2.4}$$

Let s be a jump point of $\widehat{F}_n$. Let $\nabla f(x) = f(x) - f(x-)$ for any right continuous function f. Because $\widehat{F}_n$ and $\widehat{G}_n$ do not jump at the same time points, (2.4) implies that

$$\nabla (1 - \widehat{F}_n)(1 - \widehat{G}_n)(s) = -[1 - \widehat{G}_n(s)]\nabla \widehat{F}_n(s) = -\frac{1}{n}.$$

It follows that,

$$\nabla \widehat{F}_n(s) = \frac{1}{n}\frac{1}{1 - \widehat{G}_n(s)}. \tag{2.5}$$

For general s, to avoid $\widehat{G}_n(s) = 1$, we use $\widehat{G}_n(s-)$. It follows from (2.5) that $\delta_i/[n(1 - \widehat{G}_n(Y_i-))] \leq 1$ for all $i = 1, \ldots, n$, since the jumps of $\widehat{F}_n$ are bounded by 1. Therefore, these weights are stable and there is no need to artificially censor the observed times Y_i's as suggested by Zhou (2005). We note that with the least squares approach, no artificial censoring was suggested by Zhou (1992) or Stute (1993, 1996, 1999). The simulation study of Stute (1999) in the context of the least squares estimation, and our simulation study in Section 4 below, also show good finite sample performance.

As shown in Theorems 1 and 2 below, the KMW-LAD estimator is root-n consistent and asymptotically normal. However, the asymptotic variance does not have a simple form. In particular, the conditional density function of the error term is involved in the asymptotic variance in the term $E[ZZ'f_\varepsilon(0|Z)]$. Although in principle we can estimate $E[ZZ'f_\varepsilon(0|Z)]$ using a kernel estimator, this is not straightforward due to censoring. Therefore, we propose the following nonparametric bootstrap (Efron and Tibshirani (1993)) for inference.

Sample $m \approx 0.632n$ from the n observations without replacement. The bootstrap sample is estimated following the same procedure as for the complete sample. The bootstrap procedure is then repeated B times. After proper scale adjustment, the sample variance of the bootstrap estimates provides an estimate of the variance of $\widehat{\beta}_n$. We use $m = 0.632n$ since the expected number of distinct bootstrap observations is about $0.632n$. Simulation studies in Section 4 are used to investigate finite sample performance of this bootstrap procedure (and see Ma and Kosorok (2005)).

3. Large Sample Properties

We now state the consistency and asymptotic normality results for $\widehat{\beta}_n$. Denote β_0 as the unknown true value of β. We first introduce the notation needed for stating these results. Let H denote the distribution function of Y. Let τ_Y, τ_T and τ_C be the end points of the support of Y, T and C, respectively. Let $Z = (1, X_1, \ldots, X_d)' = (Z_0, Z_1, \ldots, Z_d)'$, and F^0 be the joint distribution of

(Z, T). Write

$$\widetilde{F}^0(z, t) = \begin{cases} F^0(z, t), & t < \tau_Y \\ F^0(z, \tau_Y-) + F^0(z, \tau_Y)1\{\tau_Y \in A\}, & t \geqslant \tau_Y \end{cases},$$

where A denotes the set of atoms of H. Define sub-distribution functions

$$\widetilde{H}^{11}(z, y) = P(Z \leq z, Y \leq y, \delta = 1), \quad \widetilde{H}^0(y) = P(Y \leq y, \delta = 0).$$

Let $\text{sgn}(x) = -1, 0, 1$ if $x < 0, = 0, > 0$, respectively. For $j = 0, \ldots, d$, let

$$\gamma_0(y) = \exp\left\{ \int_0^{y-} \frac{\widetilde{H}^0(dw)}{1 - H(w)} \right\},$$

$$\gamma_{1j}(y; \beta) = \frac{1}{1 - H(y)} \int 1\{w > y\}\text{sgn}(w - z'\beta)z_j\gamma_0(w)\widetilde{H}^{11}(dz, dw),$$

$$\gamma_{2j}(y; \beta) = \iint \frac{1\{v < y, v < w\}\text{sgn}(w - z'\beta)z_j\gamma_0(w)}{[1 - H(v)]^2} \widetilde{H}^0(dv)\widetilde{H}^{11}(dz, dw).$$

For $j = 0, 1, \ldots, d$, let $\psi_j = Z_j\text{sgn}(Y - Z'\beta_0)\gamma_0(Y)\delta + \gamma_{1j}(Y; \beta_0)(1-\delta) - \gamma_{2j}(Y; \beta_0)$, and $\sigma_{ij} = \text{Cov}(\psi_i, \psi_j)$. Write $\Sigma = (\sigma_{ij})_{0 \leq i, j \leq d}$. We assume the following conditions.

A1: Let $F_\varepsilon(\cdot|z)$ be the conditional distribution function of ε given $Z = z$ and $f_\varepsilon(\cdot|z)$ be its conditional density function. Then $F_\varepsilon(0|z) = 0.5$, and $f_\varepsilon(e|z)$ is continuous in e in a neighborhood of 0 for almost all z.

A2: T and C are independent and $P(T \leq C|T, Z) = P(T \leq C|T)$.

A3: $\tau_T < \tau_C$ or $\tau_T = \tau_C = \infty$.

A4: $E[ZZ'f_\varepsilon(0|Z)]$ is finite and nonsingular.

A5: (a) The covariate Z is bounded and the right end point of the support of $Z'\beta_0$ is strictly less than τ_Y; (b) $E\left[\|Z\|^2\gamma_0^2(Y)\delta\right] < \infty$ and $\int |z_j|D^{1/2}(w)\widetilde{F}^0$ $(dz, dw) < \infty$, for $j = 0, \ldots, d$, where $D(y) = \int_0^{y-} [(1 - H(w))(1 - G(w))]^{-1}$ $G(dw)$. Here G is the distribution function of the censoring time C.

In (A1), we only need that $\text{median}(\epsilon|X = x) = 0$. The distribution of ϵ can depend on covariates; this allows heteroscedastic error terms. For example, the results below hold for $\epsilon_i = \sigma(X_i)\epsilon_{0i}$, where the ϵ_{0i}'s are independent and identically distributed with median 0. This is weaker than the corresponding assumption in the Buckley-James method (Buckley and James (1979)) and the rank-based method (Jin et al. (2003)), where the error terms ϵ_i's are assumed to have a common distribution and to be independent of the X_i's. (A2) assumes that δ is conditionally independent of the covariate X given the failure time Y. It also assumes that T and C are independent, which is the same as that for the Kaplan-Meier estimator. However, we note that (A2) does allow the

censoring variable to be dependent on the covariates. In comparison, in the Buckley-James and rank-based estimators, it is assumed that $T - \beta_0 - X'\beta$ and $C - \beta_0 - X'\beta$ are conditionally independent given X. (A3) ensures that the distribution of T can be estimated over its support. It is part of the conditions for the identification of β_0 in the model. (A4) is the same as the assumption for the consistency and asymptotic normality of the LAD estimator in linear regression models. (A1)$-$(A4) ensure identifiability of β_0 and consistency of the KMW-LAD estimator. (A5a) and (A5b) are technical assumptions for proving asymptotic normality. (A5b) together with (A4) entail finite asymptotic variance of the KMW-LAD estimator. (A5c) guarantees that the bias of Kaplan-Meier integral is on the order of $o(n^{-1/2})$. It is related to the degree of censoring and the tail behavior of the Kaplan-Meier estimator. Therefore, the assumptions needed for theoretical justification of the KMW-LAD estimator are quite reasonable and comparable to those of the Buckley-James and rank-based estimators.

Theorem 1. (Consistency). *Suppose that* (A1)$-$(A4) *hold. Then* $\widehat{\beta}_n \to_P \beta_0$ *as* $n \to \infty$.

Theorem 2. (Asymptotic Normality). *Suppose that* (A1)$-$(A5) *hold. Let* $Z_{(i)} = (1, X'_{(i)})'$ *and* $A = 2E[ZZ'f_\varepsilon(0|Z)]$. *We have*

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) = A^{-1}\sqrt{n}\sum_{i=1}^n w_{ni}Z_{(i)}sgn(Y_{(i)} - Z'_{(i)}\beta_0) + o_p(1).$$

In particular, $\sqrt{n}(\widehat{\beta}_n - \beta_0) \to_D N(0, A^{-1}\Sigma A^{-1})$.

4. Numerical Studies

4.1. Simulation studies

We first compare the KMW-LAD estimator with the median regression estimator of Ying et al. (1995) using simulation. Consider the AFT model with a single covariate. We set the sample size at 100 and take $(\beta_0, \beta_1) = (0, 1)$. The covariates X for Examples 1$-$6 are generated from the $U(0, 1)$ distribution. In Examples 1$-$3, the errors are normally distributed with mean 0 and standard deviation 0.5. Censoring variables are generated uniformly distributed and independent of the covariates and the event. The censoring rates for Examples 1$-$3 are 0%, 30% and 70%, respectively. Examples 4$-$6 are similar to Examples 1$-$3, respectively. The only difference is that the errors for Examples 4$-$6 have a $Cauchy(0, 0.5)$ distribution. In Examples 7 and 8, X has truncated $N(0, 0.5)$ and $N(0, 0.25)$ distributions, respectively. The errors are $Cauchy(0, 1)$ distributed and the censoring variables are generated from $N(1, 1)$ and independent of X. The censoring rates for Examples 7 and 8 are both around 30%. The simulation

settings here are similar to those in Ying et al. (1995). Summary statistics based on 1,000 replicates are given in Table 1.

Table 1. Simulation study: comparison of the proposed approach with Ying's approach. $(\beta_0, \beta_1) = (0, 1)$. The values below are sample means (standard deviations) and median of mean squared errors.

	Ying			KMW-LAD		
	β_0	β_1	mse	β_0	β_1	mse
1	-0.011 (0.128)	1.019 (0.221)	0.031	-0.001 (0.126)	1.002 (0.221)	0.031
2	-0.006 (0.159)	1.035 (0.320)	0.061	0.010 (0.152)	0.943 (0.286)	0.050
3	-0.003 (0.256)	1.029 (0.496)	0.171	-0.009 (0.245)	1.003 (0.479)	0.123
4	-0.011 (0.166)	1.029 (0.294)	0.054	0.005 (0.163)	1.002 (0.282)	0.047
5	-0.024 (0.228)	1.023 (0.408)	0.104	-0.014 (0.197)	0.997 (0.348)	0.072
6	0.009 (0.438)	0.977 (0.858)	0.287	-0.091 (0.344)	0.992 (0.610)	0.221
7	0.046 (0.156)	1.051 (0.456)	0.135	-0.117 (0.202)	0.965 (0.495)	0.194
8	0.009 (0.187)	1.092 (0.798)	0.321	-0.073 (0.207)	0.933 (0.978)	0.389

It can be seen that both approaches perform well under all simulated scenarios. The biases of the proposed estimator are small. The accuracy of the proposed approach decreases as the censoring rate increases, as can be seen from Examples 1−6. With regard to relative efficiency for estimating β_1, the two approaches are equally efficient in Example 1; the KWM-LAD is more efficient in Examples 2−6; while Ying's approach is more efficient in Examples 7 and 8. Extensive simulation studies show that in general the relative efficiency is model and data dependent.

We use the following simulation study to assess the bootstrap inference procedure. Consider the AFT model with a three dimensional covariate vector. We set sample size at 100 and take $(\beta_0, \beta_1, \beta_2, \beta_3) = (0, 1, 1, 1)$. Covariates are marginally uniformly distributed. In Example 9, the covariates are generated in a manner such that the pairwise correlation coefficient between the ith and the jth components is $0.5^{|i-j|}$. Errors are generated as normally distributed with mean 0 and standard deviation 0.5. In Example 10, covariates are independently generated. For both examples, censoring variables are generated independent of the covariates and the event. The censoring rates are $\sim 30\%$. Confidence intervals are constructed using the nonparametric bootstrap, based on the asymptotic normality results in Theorem 2. The marginal empirical coverage rates of 95% confidence intervals are $(0.942, 0.948, 0.968, 0.954)$ in Example 9 and $(0.958, 0.954, 0.968, 0.956)$ in Example 10, based on 1,000 replicates and 1,000 bootstrap for each sample. Extensive simulation studies under different simulated scenarios all yield similar, satisfactory empirical coverage rates.

We also conduct simulation studies to examine the sensitivity of the proposed approach to violation of assumption (A2). Set the sample size at 100 and the generating parameter value $(\beta_0, \beta_1) = (0, 1)$. Let X be $U(0, 1)$ distributed. In Examples 11 and 12, T is normally distributed with mean X and variance 0.025. In Examples 13 and 14, T is normally distributed with mean X and variance $0.025X$; the censoring variables are normally distributed with mean $v + X$ and variance 0.025, where v is chosen with certain pre-specified censoring rate. The censoring rates are 25% for Examples 11 and 13, and 50% for Examples 12 and 14. For Examples 11−14, the mean correlation coefficients between the event time and the censoring are 0.769, 0.768, 0.818, 0.818, respectively. Simulation based on 1,000 replicates shows that the sample means of the estimates are (-0.038, 1.002), (-0.121, 1.099), (-0.034, 1.083) and (0.001, 0.970) for Examples 11−14, respectively. The empirical coverage rates of the 95% confidence intervals based on the nonparametric bootstrap are 0.983, 0.981, 0.984 and 0.985 for β_1 in Examples 11−14, respectively. So even if assumption A2 is seriously violated, the proposed estimation procedure still performs well, although the nonparametric bootstrap over-estimates the variances.

4.3. Small cell lung cancer data

We use the lung cancer study data in Ying et al. (1995) as the first example to demonstrate the proposed method. For patients with small cell lung cancer (SCLC), the standard therapy is to use a combination of etoposide and cisplatin. However, the optimal sequencing and administration schedule have not been established. The data are from a clinical study which was designed to evaluate two regimens: Arm A: cisplatin followed by etoposide; and Arm B: etoposide followed by cisplatin. In this study, 121 patients with limited-stage SCLC were randomly assigned to these two groups, with 62 patients to A and 59 patients to B. At the time of the analysis, there was no loss to follow-up. Each death time was either observed or administratively censored. Therefore, the censoring variable does not depend on the covariates, which are the treatment indicator and patients' entry age. Let T be the base 10 logarithm of the patients' failure time. Let $X_1 = 0$ if the patient is in Group A and 1 otherwise. Let X_2 denote the patients' entry age. We assume the AFT model $T = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$. The data and model settings are the same as in Ying et al. (1995).

The proposed approach gives $\hat{\beta}_0 = 2.693(0.164)$, $\hat{\beta}_1 = -0.146(0.050)$ and $\hat{\beta}_2 = 0.001(0.003)$, where the numbers in parentheses are the estimated standard errors obtained using the nonparametric bootstrap with 1,000 bootstrap samples. The median regression estimates of Ying et al. (1995) (reproduced from their paper) are $\hat{\beta}_0 = 3.028$, $\hat{\beta}_1 = -0.163(0.090)$ and $\hat{\beta}_2 = -0.004(0.005)$. (No standard error

was given for $\hat{\beta}_0$ in Ying et al. (1995)). Estimates of the covariates effects from the two methods are reasonably close. The effect of X_1 is modestly significant, which indicates that Arm A tends to give better results than Arm B. The effect of entry age is not significant. As pointed in Ying et al. (1995), one advantage of the AFT model is that the median survival time for a prospective patient can be predicted based on the above estimates.

4.4. PBC data

Between 1974 and 1984, the Mayo Clinic conducted a double-blinded randomized clinical trial in primary cirrhosis of the liver (PBC). 312 patients participated in the trial and there were 18 covariates. We focus on the 276 patients with complete records only. Descriptions of this data set can be found in Fleming and Harrington (1991), where the Cox model is used in the analysis. As an alternative, we apply the AFT model using the proposed KMW-LAD estimator as an illustration. *log* transformations of the covariates *alkphos*, *bili*, *chol*, *copper*, *platelet*, *protime*, *sgot* and *trig* are first made, so that the marginal distributions of those covariates are closer to normal. We also apply the logarithm transformation to the observed time.

The KMW-LAD estimates and their estimated standard errors are shown in Table 2. The estimated standard errors are obtained with 1,000 bootstrap samples. As a comparison, we also include the estimates obtained based on the Cox model in Table 2. Although estimates from two different models are not directly comparable, we can see that the biological conclusions, in terms of positive or negative association with survival, are similar. Here we note that because the Cox model models the conditional hazard function, while the AFT model models the failure time directly, opposite signs of the corresponding regression coefficients in the two models indicate qualitative agreement. Because the dimension of the covariates is relatively high, we are not able to apply the existing censored median regression estimator for the AFT model, due to computational difficulties.

5. Concluding Remarks

The KMW-LAD estimator does not require independence between the censoring variable and covariates. However, it does require independence between the censoring time and event time. In many studies, such as the small cell lung cancer study, this assumption is reasonable, since censoring was done administratively. Simulation studies reported in Section 4 show that although coverage rates of the confidence intervals are slightly higher than the nominal 95% rate

when the censoring variable and the event are strongly correlated (correlation coefficients ranging from 0.77 to 0.82), the estimation results are satisfactory. This suggests that the KMW-LAD estimator is quite robust to departure from the independence assumption. However in general, if the independence assumption is not satisfied, caution is needed in applying the proposed KMW-LAD estimator.

Table 2. PBC data: comparison of the KMW-LAD estimator and the maximum partial likelihood estimator for the Cox model. s.e.: standard error.

covariate	Cox estimate	s.e.	KMW-LAD estimate	s.e.
$intercept$	NA	NA	3.974	6.091
age	0.031	0.011	-0.012	0.015
alb	-0.612	0.300	0.464	0.378
$log(alkphos)$	0.039	0.147	0.246	0.170
$ascites$	0.211	0.377	-0.543	0.847
$log(bili)$	0.632	0.178	-0.153	0.236
$log(chol)$	0.162	0.292	-0.032	0.347
$edtrt$	0.918	0.380	-0.854	0.982
$hepmeg$	-0.087	0.257	0.037	0.299
$log(platelet)$	0.132	0.285	-0.261	0.321
$log(protime)$	2.482	1.348	1.815	1.759
sex	-0.182	0.318	0.035	0.386
$log(sgot)$	0.406	0.309	-0.040	0.374
$spiders$	0.049	0.241	-0.123	0.322
$stage$	0.366	0.178	-0.071	0.153
trt	-0.006	0.212	0.127	0.259
$log(trig)$	-0.144	0.251	-0.192	0.375
$log(copper)$	0.284	0.176	-0.129	0.198

Acknowledgements

Appendix

We now prove Theorems 1 and 2. The proofs use the convergence results for

Kaplan-Meier weighted statistics of Stute (1993, 1996). They also use the empirical process theory and the methods of Hjort and Pollard (1993) for asymptotic analysis in convex minimization problems.

Proof of Theorem 1.(Consistency) Let $M_n(\beta) = \sum_{i=1}^{n} w_{ni}[|Y_{(i)} - Z'_{(i)}\beta| - |Y_{(i)} - Z'_{(i)}\beta_0|]$. Then the minimizers of M_n are identical to those of L_n in (2.3), since M_n is a shift of L_n by a constant term independent of β. Because the L_1 norm is convex, M_n is a convex function of β.

In Stute (1993), under (A2) and (A3), it was proved that for any measurable function φ,

$$S_n^{\varphi} \equiv \sum_{i=1}^{n} w_{ni}\varphi(z_{(i)}, y_{(i)}) \to S^{\varphi} \equiv \int \varphi d\tilde{F}^0, \quad a.s.$$

provided that $\int |\varphi| dF^0$ is finite. Applying this result to $\varphi_\beta(z,t) = |t - z'\beta| - |t - z'\beta_0|$, when $\tau_T < \tau_C$ or $\tau_Y = \infty$, we obtain

$$M_n(\beta) \longrightarrow M(\beta), \quad a.s., \quad \text{for any fixed } \beta \in R^{d+1}, \tag{6.1}$$

where the limit

$$M(\beta) \equiv E\left\{|T - Z'\beta| - |T - Z'\beta_0|\right\}$$
$$= E\left[\int_0^{Z'(\beta - \beta_0)} [2F_\varepsilon(e|Z) - 1]de\right].$$

By (A1), $\partial M(\beta)/\partial\beta|_{\beta=\beta_0} = 0$, and by (A1) and (A4), $\partial^2 M(\beta)/\partial\beta^2 = 2E\left[ZZ'f_\varepsilon(Z'(\beta - \beta_0)|Z)\right] \geq 0$ and strict inequality holds for $\beta \neq \beta_0$ in a neighborhood of β_0. Thus,

$$h(\delta) \equiv \inf_{\|\beta-\beta_0\|=\delta} M(\beta) > 0, \quad \text{for any } \delta > 0. \tag{6.2}$$

By the convexity lemma of Pollard (1991), for any compact set K in a convex open subset of R^{d+1},

$$\sup_{\beta \in K} |M_n(\beta) - M(\beta)| \to_P 0 \tag{6.3}$$

follows from the convexity of $\varphi_\beta(z,t)$ as a function of β and (6.1). By Lemma 2 of Hjort and Pollard (1993), $\widehat{\beta}_n \to_P \beta_0$. This completes the proof of Theorem 1.

Proof of Theorem 2.(Asymptotic Normality) Let

$$M_n(s) = n \sum_{i=1}^{n} w_{ni}\left[|Y_{(i)} - Z'_{(i)}(\beta_0 + n^{-\frac{1}{2}}s)| - |Y_{(i)} - Z'_{(i)}\beta_0|\right],$$

$$R_n(z, y; s) = |y - z'(\beta_0 + n^{-\frac{1}{2}}s)| - |y - z'\beta_0| + n^{-\frac{1}{2}}\text{sgn}(y - z'\beta_0)z's.$$

We can express $M_n(s)$ as

$$M_n(s) = n \sum_{i=1}^{n} w_{ni} R_n(Z_{(i)}, Y_{(i)}; s) - n^{\frac{1}{2}} \sum_{i=1}^{n} w_{ni} \operatorname{sgn}(Y_{(i)} - Z'_{(i)} \beta_0) Z'_{(i)} s. \qquad (6.4)$$

Denote the first term on right-hand side of (6.4) by $Q_n(s) = n \sum_{i=1}^{n} w_{ni} R_n(Z_{(i)}, Y_{(i)}; s)$. We first show that, for any fixed s,

$$Q_n(s) \to_P \frac{1}{2} s' A s, \qquad (6.5)$$

where A is defined in Theorem 2. Let the empirical counterparts of $H(y), \widetilde{H}^0(y)$ and $\widetilde{H}^{11}(z, y)$ be:

$$H_n(y) = n^{-1} \sum_{i=1}^{n} 1\{Y_{(i)} \le y\}$$

$$\widetilde{H}_n^0(y) = n^{-1} \sum_{i=1}^{n} 1\{Y_{(i)} \le y, \delta_{(i)} = 0\}$$

$$\widetilde{H}_n^{11}(z, y) = n^{-1} \sum_{i=1}^{n} 1\{Z_{(i)} \le z, Y_{(i)} \le y, \delta_{(i)} = 1\}.$$

By Lemma 5.1 of Stute (1996),

$$Q_n(s) = \int n R_n(z, y; s) \exp\left\{ \int_{-\infty}^{y-} n \ln(1 + \frac{1}{n(1 - H_n(u))}) \widetilde{H}_n^0(du) \right\} \widetilde{H}_n^{11}(dz, dy).$$

Write $Q_n(s) = I_{1n} + I_{2n} + I_{3n} + I_{4n}$, where

$$I_{1n} = \int n R_n(z, y; s) \exp\left\{ \int_{-\infty}^{y-} \frac{\widetilde{H}^0(du)}{1 - H(u)} \right\} \widetilde{H}^{11}(dz, dy),$$

$$I_{2n} = \int n R_n(z, y; s) \exp\left\{ \int_{-\infty}^{y-} \frac{\widetilde{H}^0(du)}{1 - H(u)} \right\} (\widetilde{H}_n^{11} - \widetilde{H}^{11})(dz, dy),$$

$$I_{3n} = \int n R_n(z, y; s) \left(\exp\left\{ \int_{-\infty}^{y-} \frac{\widetilde{H}_n^0(du)}{1 - H_n(u)} \right\} \right.$$
$$\left. - \exp\left\{ \int_{-\infty}^{y-} \frac{\widetilde{H}^0(du)}{1 - H(u)} \right\} \right) d\widetilde{H}_n^{11}(z, y),$$

$$I_{4n} = \int \left(\exp\left\{ \int_{-\infty}^{y-} n \ln(1 + \frac{1}{n(1 - H_n(u))}) \widetilde{H}_n^0(du) \right\} \right.$$
$$\left. - \exp\left\{ \int_{-\infty}^{y-} \frac{\widetilde{H}_n^0(du)}{1 - H_n(u)} \right\} \right) \times n R_n(z, y; s) \widetilde{H}_n^{11}(dz, dy).$$

Under (A3) and (A4), and recall $A = 2E[ZZ'f_\varepsilon(0|Z)]$, the first term

$$I_{1n} = \int nR_n(z,y;s)\exp\left\{\int_{-\infty}^{y-}\frac{\widetilde{H}^0(du)}{1-H(u)}\right\}\widetilde{H}^{11}(dz,dy)$$

$$= \int nR_n(z,y;s)\widetilde{F}^0(dz,dy) = E[nR_n(Z,T;s)]$$

$$= s'E[ZZ'f_\varepsilon(0|Z)]s + o(1) = \frac{1}{2}s'As + o(1). \qquad (6.6)$$

Under (A5b), and by Lemma 19.31 of Van der Vaart (1998), the second term

$$I_{2n} = \sum_{i=1}^{n} nR_n(Z_{(i)},Y_{(i)};s)\delta_i\gamma_0(Y_{(i)}) - E[nR_n(Z,T\wedge C;s)\delta\gamma_0(T\wedge C)]$$

$$= \mathbb{G}_n\left(\sqrt{n}R_n(z,t\wedge c;s)\delta\gamma_0(t\wedge c)\right) = o_P(1). \qquad (6.7)$$

For I_{3n}, we first note that, for $\eta < \tau_Y$,

$$\sup_{y\leq\eta}\left|\exp\left\{\int_{-\infty}^{y-}\frac{\widetilde{H}_n^0(du)}{1-H_n(u)} - \int_{-\infty}^{y-}\frac{\widetilde{H}^0(du)}{1-H(u)}\right\} - 1\right|$$

$$= \sup_{y\leq\eta}\left|\exp\left\{\int_{-\infty}^{y-}\frac{\widetilde{H}_n^0(du)}{1-H_n(u)} - \int_{-\infty}^{y-}\frac{\widetilde{H}_n^0(du)}{1-H(u)} + \int_{-\infty}^{y-}\frac{\widetilde{H}_n^0(du)}{1-H(u)}\right.\right.$$

$$\left.\left. - \int_{-\infty}^{y-}\frac{\widetilde{H}^0(du)}{1-H(u)}\right\} - 1\right|$$

$$= |\exp\{o_P(1) + o_P(1)\} - 1| = o_P(1),$$

where the second equality follows the generalized Glivenko-Cantelli Theorem (Van der Vaart and Wellner (1996)). We also have

$$R_n(z,y;s) = \begin{cases} 2\left(n^{-\frac{1}{2}}z's - (y-z'\beta_0)\right)1\{z'\beta_0 < y < z'(\beta_0 + n^{-\frac{1}{2}}s)\} & z's > 0 \\ -2\left(n^{-\frac{1}{2}}z's - (y-z'\beta_0)\right)1\{z'(\beta_0 + n^{-\frac{1}{2}}s) < y < z'\beta_0\} & z's < 0. \end{cases}$$

Thus under (A5a),

$$I_{3n} = \int nR_n(z,y;s)\gamma_0(y)\left(\exp\left\{\int_{-\infty}^{y-}\frac{\widetilde{H}_n^0(du)}{1-H_n(u)} - \int_{-\infty}^{y-}\frac{\widetilde{H}^0(du)}{1-H(u)}\right\} - 1\right)\widetilde{H}_n^{11}(dz,dy)$$

$$= o_P(1)\int nR_n(z,y;s)\gamma_0(y)\widetilde{H}_n^{11}(dz,dy)$$

$$= o_P(1)(I_{1n} + I_{2n}) = o_P(1). \qquad (6.8)$$

Finally,

$$I_{4n} = \int n R_n(z, y; s) \exp\left\{ \int_{-\infty}^{y-} \frac{\widetilde{H}_n^0(du)}{1 - H_n(u)} \right\}$$
$$\left[\exp\left\{ \int_{-\infty}^{y-} n \ln\left(1 + \frac{1}{n(1 - H_n(u))} \right) - \frac{1}{1 - H_n(u)} \widetilde{H}^0(du) \right\} - 1 \right] \widetilde{H}_n^{11}(dz, dy).$$

The expression in the square brackets is bounded between $[-2n(1 - H_n(y-))]^{-1}$ and 0. Because $nR_n(z, y; s)$ vanishes when y goes beyond $z'\beta_0$ or $z'(\beta_0 + n^{-1/2}s)$, by the generalized Glivenko-Cantelli Theorem and (A5a),

$$I_{4n} = O_P(n^{-1}) \int n R_n(z, y; s) \exp\left\{ \int_{-\infty}^{y-} \frac{\widetilde{H}_n^0(du)}{1 - H_n(u)} \right\} \widetilde{H}_n^{11}(dz, dy)$$
$$= O_P(n^{-1})(I_{1n} + I_{2n} + I_{3n}) = o_P(1). \tag{6.9}$$

Combining (6.6) to (6.9), (6.5) follows. Therefore, $M_n(s) = s'As/2 - s'n^{1/2} \sum_{i=1}^{n} w_{ni} Z_{(i)} \operatorname{sgn}(Y_{(i)} - Z_{(i)}'\beta_0) + o_P(1)$. Under (A2)-(A5), by Theorem 3.1 of Stute (1996), $n^{1/2} \sum_{i=1}^{n} w_{ni} Z_{(i)} \operatorname{sgn}(Y_{(i)} - Z_{(i)}'\beta_0) \to_D N(0, \Sigma)$. By the Basic Corollary of Hjort and Pollard (1993), we have $\sqrt{n}(\widehat{\beta}_n - \beta_0) = A^{-1}\sqrt{n} \sum_{i=1}^{n} w_{ni} Z_{(i)} \operatorname{sgn}(Y_{(i)} - Z_{(i)}'\beta_0) + o_p(1)$. This completes the proof of Theorem 2.

References

Bang, H. and Tsiatis, A. A. (2002). Median regression with censored cost data. *Biometrics* **58**, 643-649.

Bassett, G. and Koenker, R. (1978). Asymptotic theory of least absolute error regression. *J. Amer. Statist. Assoc.* **73**, 618-622.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429-436.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, New York.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.

Hjort, N. L. and Pollard, D. (1993). Asymptotics for minimizers of convex processes. Preprint, Department of Statistics, Yale University.

Jin, Z., Lin, D. Y., Wei, L. J. and Ying, Z. L. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341-353.

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

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

Lin, D. Y. and Geyer, C. J. (1992). Computational methods for semiparametric linear regression with censored data. *J. Comput. Graph. Statist.* **1**, 77-90.

Ma, S. and Kosorok, M. R. (2005). Robust semiparametric M-estimation and weighted bootstrap. *J. Multivariate Anal.* **96**, 190-217.

Newey, W. K. and Powell, J. L. (1990). Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. *Econometric Theory* **6**, 295-317.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, **7**, 186-199.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *J. Econometrics* **25**, 303-325.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65**, 167-179.

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

Subramanian, S. (2002). Median regression using nonparametric kernel estimation. *J. Nonparametr. Stat.* **14**, 583-605.

Stute, W. (1993). Consistent estimation under random censorship when covariables are available. *J. Multivariate Anal.* **45**, 89-103.

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.* **23**, 461-471.

Stute, W. (1999). Nonlinear censored regression. *Statist. Sinica* **9**, 1089-1102

Stute, W. and Wang, J. L. (1993). The strong law under random censorship. *Ann. Statist.* **14**, 1351-1365.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 303-328.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Van der Vaart, A. W., and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.

Wei, L. J., Ying, Z. L. and Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845-851.

Wei, L. J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statist. Medicine* **11**, 1871-1879.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard functions. *J. Amer. Statist. Assoc.* **94**, 137-145.

Ying, Z. L. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21**, 76-99.

Ying, Z. L., Jung, S. H. and Wei, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90**, 178-184.

Zhou, L. Z. (2005). A simple censored median regression estimator.

Zhou, M. (1992). M-estimation in censored linear models. *Biometrika* **79**, 837-841. *Accepted for publication by Statist. Sinica.*

Department of Statistics and Actuarial Science, and Program in Public Health Genetics, University of Iowa, 241 SH, Iowa City, IA 52242, U.S.A.

E-mail: jian-huang@uiowa.edu

Division of Biostatistics, Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College Street, New Haven, CT 06520, U.S.A.

E-mail: shuangge.ma@yale.edu

Department of Statistics and Actuarial Science, University of Iowa, 241 SH, Iowa City, IA 52242, U.S.A.

E-mail: huiliang-xie@uiowa.edu