Least Absolute Deviations Estimation for the Accelerated Failure Time Model

Jian Huang[1,2], Shuangge Ma[3], and Huiliang Xie[1]

[1]Department of Statistics and Actuarial Science, and [2]Program in Public Health Genetics,

University of Iowa

[3]Department of Biostatistics, University of Washington–Seattle

* *e-mail:* jian@stat.uiowa.edu

June 2005

The University of Iowa

Department of Statistics and Actuarial Science

Technical Report No. 350

Summary The accelerated failure time (AFT) model assumes a linear relationship between the event time and the 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 mild assumptions. The proposed estimator can 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 with two clinical data sets.

KEYWORDS: 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 from 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-deviation (LAD) method has received much attention due to its robustness property with respect to the response variable in the regression (Bassett and Koenker 1978; 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 the 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 in practice difficult.

In this paper, we propose a weighted LAD estimator in the AFT model using the Kaplan-Meier weights. For simplicity of notations, we call this estimator the KMW-LAD estimator. The use of the Kaplan-Meier weights to account for censoring was first proposed by Stute (1993, 1996, 1999) in least squares estimation of the AFT model. 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 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_i be the logarithm of the failure time and $X_i = (X_{i1}, \ldots, X_{id})'$ a d-dimensional covariate vector for the ith subject in a random sample of size n. The AFT model assumes

$$T_i = \beta_0 + X_{i1}\beta_1 + \cdots X_{id}\beta_d + \varepsilon_i,\ i = 1, \ldots, n, \qquad (1)$$

where β_0 is the intercept, $\beta_1, \ldots, \beta_d$ are the regression coefficients and ε_i is the error term with an unknown distribution function. When T_i is subject to right censoring, we can only observe

(Y_i, δ_i, X_i) with $Y_i = \min\{T_i, C_i\}$, where C_i is logarithm of the censoring time and $\delta_i = 1\{\{T_i \leq C_i\}$ is the censoring indicator. Suppose that a random sample $(Y_i, \delta_i, X_i), i = 1, \ldots, n$ 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 w_{ni}'s are the Kaplan-Meier weights and can be expressed as,

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

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 $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_d)$. The KMW-LAD estimator $\widehat{\boldsymbol{\beta}}_n$ is the minimizer of

$$L_n(\boldsymbol{\beta}) = \sum_{i=1}^n w_{ni} |Y_{(i)} - \beta_0 - X_{(i1)}\beta_1 - \cdots - X_{(id)}\beta_d|. \tag{2}$$

Robustness is gained by using the LAD objective function. $\widehat{\boldsymbol{\beta}}_n$ 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 in SAS, and the quantile regression (qreg) procedure in STATA.

As shown in Theorems 1 and 2 below, the KMW-LAD estimator is consistent and 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 $\widehat{\boldsymbol{\beta}}_n$. We use $m = 0.632n$ since the expected number of distinct bootstrap observations is about $0.632n$. Computationally, it may be more efficient to use a smaller bootstrap sample size. Simulation studies in Section 4 are used to investigate finite sample performance of this bootstrap procedure.

3. Large Sample Properties

We now state the consistency and asymptotic normality results for $\widehat{\boldsymbol{\beta}}_n$. Denote $\boldsymbol{\beta}_0$ as the unknown true value of $\boldsymbol{\beta}$. We first introduce notations 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). Denote

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

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

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

Let sign$(x) = -1, 0, 1$ if $x < 0, = 0, > 0$, respectively. For $j = 0, \ldots, d$, denote

$$\begin{aligned} \gamma_0(y) &= \exp\left\{ \int_0^{y-} \frac{\widetilde{H}^0(dw)}{1 - H(w)} \right\}, \\ \gamma_{1j}(y; \boldsymbol{\beta}) &= \frac{1}{1 - H(y)} \int 1_{\{w > y\}} \text{sign}(w - \mathbf{z}'\boldsymbol{\beta}) z_j \gamma_0(w) \widetilde{H}^{11}(dz, dw), \\ \gamma_{2j}(y; \boldsymbol{\beta}) &= \iint \frac{1_{\{v<y, v<w\}} \text{sign}(w - z'\boldsymbol{\beta}) z_j \gamma_0(w)}{[1 - H(v)]^2} \widetilde{H}^0(dv) \widetilde{H}^{11}(dz, dw). \end{aligned}$$

For $j = 0, 1, \ldots, d$, let

$$\psi_j = Z_j \text{sign}(Y - Z'\boldsymbol{\beta}_0) \gamma_0(Y) \delta + \gamma_{1j}(Y; \boldsymbol{\beta}_0)(1 - \delta) - \gamma_{2j}(Y; \boldsymbol{\beta}_0),$$

and $\sigma_{ij} = \text{Cov}(\psi_i, \psi_j)$. Denote $\Sigma = (\sigma_{ij})_{0 \le i,j \le d}$. We assume the following conditions:

A1: Let $F_\varepsilon(\cdot|z)$ be the conditional distribution function of ε given $Z = z$ and $f_\varepsilon(\cdot|z)$ 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 $P(T \le C|T, Z) = P(T \le C|T)$.

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 $Z'\boldsymbol{\beta}_0$ 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 median$(\varepsilon|X = x) = 0$. The distribution of ε can depend on covariates. This allows heteroscedastic error terms. For example, the results below hold for $\varepsilon_i =$

$\sigma(X_i)\varepsilon_{0i}$, where 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 X_i's. (A2) assumes that δ is conditionally independent of the covariate X given the failure time Y. It also assumes that Y 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 $\boldsymbol{\beta}_0$ 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)ensures identifiability of $\boldsymbol{\beta}_0$ and consistency of the KMW-LAD estimator. (A5a) and (A5b) are technical assumptions for proving asymptotic normality. (A5b) together with (A4) entails finite asymptotic variance of the KMW-LAD estimator. (A5c) guarantees that the bias of Kaplan-Meier integral is in 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 mild and comparable to those of the Buckley-James and rank based estimators.

Theorem 1. (Consistency) Suppose assumptions (A1) – (A3) and (A4a) hold, then $\widehat{\boldsymbol{\beta}}_n \rightarrow_P \boldsymbol{\beta}_0$ as $n \rightarrow \infty$.

Theorem 2. (Asymptotic Normality) Suppose that assumptions (A1) – (A5) hold. Let $A = 2E(ZZ'f_\varepsilon(0|Z))$. Then

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = A^{-1}\sqrt{n}\sum_{i=1}^{n} w_{ni}X_i \text{sign}(Y_i - Z_i'\boldsymbol{\beta}_0) + o_p(1).$$

In particular, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \to_D N(0, A^{-1}\Sigma A^{-1})$.

Theorem 1 establishes the asymptotic validity of the KMW-LAD estimator in the sense it is consistent. Theorem 2 shows that the KMW-LAD estimator has the standard and desired asymptotic properties of root-n rate of convergence and asymptotically normality. However, as indicated above, the asymptotic variance matrix involves the conditional density $f_\varepsilon(\cdot|z)$ at 0 in the expression of A. So a simple plug-in variance estimator is not available. We suggested using the $0.632n$ nonparametric bootstrap for variance estimation. Consistency of this bootstrap procedure can be proved in a similar way as in Ma and Kosorok (2005).

4. Simulation Studies and Examples

In this section, we use simulations to evaluate the finite sample performance of the KMW-LAD estimator. We also use two real data sets to illustrate the use of the KMW-LAD estimator.

4.1 Simulation studies

We first compare the KMW-LAD estimator with the median regression estimator of Ying et al. (1995). Consider the AFT model with a single covariate. We set the sample size 100 and $(\beta_0, \beta_1) = (0, 1)$. The covariates X 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 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 example 1–3, respectively. The only difference is that the errors for examples 4–6 have a Cauchy distribution, which has heavier tails than the normal distribution. The simulation settings here are similar to those in Ying et al. (1995). Summary statistics based on 200 replicates are given in Table 1. It can be seen that both approaches behave well under all simulated scenarios. The biases of the proposed estimator are negligible. The sample standard deviations and the mean squared errors of the proposed estimator are comparable with or slightly smaller than the counterparts from the estimator of Ying et al. (1995). The accuracy of the proposed approach decreases as the censoring rate increases, as expected. The proposed estimator is more stable for errors with less variations.

We use the following simulation study to assess the bootstrap inference procedure. Consider the AFT model with a three-dimensional covariate. We set sample size 100 and $(\beta_0, \beta_1, \beta_2, \beta_3) = (0, 1, 1, 1)$. In example 7, the covariates are generated in a manner such that the pairwise correlation coefficients between the i^{th} and the j^{th} components are $0.5^{|i-j|}$. Errors are generated as normally distributed with mean 0 and standard deviation 0.5. In example 8, 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 $(0.965, 0.940, 0.950, 0.950)$ for example 7 and $(0.945, 0.960, 0.955, 0.950)$ for example 8, based on 200 replicates and 100 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). Let the sample size be 100 and the generating parameter value $(\beta_0, \beta_1) = (0, 1)$. Let X be $U(0, 1)$ distributed. In examples 9 and 10, T is normally distributed with mean X and variance 0.025. In examples 11 and 12, T is normally distributed with mean X and variance $0.025X$. The censoring variables are normally distributed with mean $C + X$ and variance 0.025. The censoring rates are 0.25 for example 9 and 11, and 0.50 for examples 10 and 12. For examples 9–12, the mean correlation coefficients between the event time and the censoring are 0.768, 0.765, 0.814, 0.817, respectively, which indicate strong correlations. Simulation based on 200 replicates shows that the sample means of the replicates are very close to the true values (results not shown here). The empirical coverage rates of the 95% confidence intervals based on the nonparametric bootstrap are 0.915, 0.915, 0.920 and 0.925 for β_1 of examples 9–12, respectively.

So even if assumption A2 is seriously violated, the proposed approach may still behave reasonably well.

4.2 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 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 regimen: 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 yield the following estimates:

$$\hat{\beta}_0 = 2.693(0.165), \hat{\beta}_1 = -0.146(0.049) \;\; and \;\; \hat{\beta}_2 = 0.001(0.003),$$

where the numbers in parentheses are the estimated standard errors obtained using the nonparametric bootstrap. The median regression estimates of Ying et al. (1995) (reproduced from their paper) are

$$\hat{\beta}_0 = 3.028, \hat{\beta}_1 = -0.163(0.090) \;\; and \;\; \hat{\beta}_2 = -0.004(0.005).$$

Estimates of the covariates effects from the two methods are reasonably close. The KMW-LAD estimate has smaller estimated standard errors. 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 above estimates.

4.3 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. There are 18 covariates in this data set. 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 observe time.

The KMW-LAD estimates and corresponding estimated standard errors are shown in Table 2. As a comparison, we also include the estimates obtained using 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 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

We have proposed the KMW-LAD estimator for the AFT model, following the weighted least squares estimator of Stute (1993, 1996). We have shown that, under appropriate conditions, the KMW-LAD estimator is root-n consistent and asymptotically normal. As the median regression estimator of Ying et al. (1995), the KMW-LAD estimator does not require the errors to be identically distributed. Simulation studies suggest that the KWM-LAD estimator and the proposed nonparametric bootstrap work well with small sample sizes when the degree of censoring ranges from mild to heavy. The main advantage of the KMW-LAD estimator over many of the estimators for the AFT model, including the Buckley-James estimator, the rank based estimators, and the existing censored median regression estimators, is that it can be easily computed by many existing softwares. This makes it easier to apply the KMW-LAD estimator to the analysis of censored data in practice, especially with medium to high dimensional covariates.

The KMW-LAD estimator does not require independence between the the censoring variable and covariates. However, it does require independence between the censoring time and the 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 the coverage rate of the confidence intervals is slightly lower 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 same simulation studies also show that actual coverage rate of 92% is close to the nominal rate of 95%. This suggests that the KMW-LAD estimator is quite robust to the departure of independence assumption. However, in general if the independence assumption is not satisfied, caution is needed in applying the proposed KMW-LAD estimator.

ACKNOWLEDGMENTS

The work of Huang is supported in part by the NIH grant HL72288-01. The work of Ma is partly supported by N01-HC-95159 from the National Heart, Lung, and Blood Institute.

References

Bassett, G. and Koenker, R. (1978). Asymptotic theory of least absolute error regression. Journal of American Statistical Association, **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). Journal of Royal Statistical Society, Series 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*. New York: John Wiley.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of American Statistical Associations, **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. Journal of Computational and Graphical Statistics, **1**, 77-90.

Ma, S. and Kosorok, M.R. (2005). Robust semiparametric M-estimation and weighted bootstrap. Journal of Multivariate Analysis. In press.

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. Journal of 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. Annals of Statistics, **18**, 303-328.

Subramanian, S. (2002). Median regression using nonparametric kernel estimation. Journal of Nonparametric Statistics, **14**, 583-605.

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

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. Scandinavia Journal of Statistics, **23**, 461-471.

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

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

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

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. Statistics in Medicine, **11**, 1871-1879.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard functions. Journal of American Statistical Association, **94**, 137-145.

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

Ying, Z. L., Jung, S. H. and Wei, L. J. (1995). Survival analysis with median regression models. Journal of American Statistical Association, **90**, 178-184.

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 heavily rely on the empirical process theory and the methods of Hjort and Pollard (1993) for asymptotics in convex minimization problems.

Proof of Theorem 1 (Consistency): Let $M_n(\boldsymbol{\beta}) = \sum_{i=1}^n w_{ni}[|Y_{(i)} - \mathbf{Z}'_{(i)}\boldsymbol{\beta}| - |Y_{(i)} - \mathbf{Z}'_{(i)}\boldsymbol{\beta}_0|]$. Then the minimizers of M_n are identical to those of L_n in (2), since M_n is a shift of L_n by a constant term independent of $\boldsymbol{\beta}$. Because the L_1 norm is convex, M_n is a convex function of $\boldsymbol{\beta}$.

In Stute (1993), it was proved that for any measurable function φ,

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

provided that $\int |\varphi| dF^0$ is finite. Applying this result to $\varphi_{\boldsymbol{\beta}}(\mathbf{z}, t) = |t - \mathbf{z}'\boldsymbol{\beta}| - |t - \mathbf{z}'\boldsymbol{\beta}_0|$, when $\tau_T < \tau_C$ or $\tau_Y = \infty$, we obtain

$$M_n(\boldsymbol{\beta}) \longrightarrow M(\boldsymbol{\beta}), \quad a.s., \quad \text{for any fixed } \boldsymbol{\beta} \in R^{d+1}, \tag{3}$$

where the limit

$$\begin{aligned} M(\boldsymbol{\beta}) &\equiv E\left[|T - \mathbf{Z}'\boldsymbol{\beta}| - |T - \mathbf{Z}'\boldsymbol{\beta}_0|\right] \\ &= E\left[\int_0^{\mathbf{Z}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)} 2F_\varepsilon(e|\mathbf{Z}) - 1 \, de\right]. \end{aligned}$$

By (A1),

$$\frac{\partial M(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = \mathbf{0}.$$

By (A1) and (A4),

$$\frac{\partial M(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 2E\left[\mathbf{Z}\mathbf{Z}' f_\varepsilon\big(\mathbf{Z}'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|\mathbf{Z}\big)\right] \geq 0$$

and strict inequality holds for $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ in a neighborhood of $\boldsymbol{\beta}_0$. Thus,

$$h(\delta) \equiv \inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|=\delta} M(\boldsymbol{\beta}) > 0, \quad \text{for any } \delta > 0. \tag{4}$$

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

$$\sup_{\boldsymbol{\beta}\in K} |M_n(\boldsymbol{\beta}) - M(\boldsymbol{\beta})| \to_P 0 \tag{5}$$

follows the convexity of $\varphi_{\boldsymbol{\beta}}(\mathbf{z}, t)$ as a function of $\boldsymbol{\beta}$ and (3). By Lemma 2 of Hjort and Pollard (1993), $\widehat{\boldsymbol{\beta}}_n \to_P \boldsymbol{\beta}_0$.

Proof of Theorem 2 (Asymptotic Normality): Let

$$M_n(\mathbf{s}) = n \sum_{i=1}^{n} w_{ni} \left[|Y_{(i)} - Z'_{(i)}(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{s})| - |Y_{(i)} - Z'_{(i)}\boldsymbol{\beta}_0| \right],$$

and

$$R_n(\mathbf{z}, y; \mathbf{s}) = |y - \mathbf{z}'(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{s})| - |y - \mathbf{z}'\boldsymbol{\beta}_0| + n^{-1/2}\text{sign}(y - \mathbf{z}'\boldsymbol{\beta}_0)\mathbf{z}'\mathbf{s}.$$

Write

$$M_n(\mathbf{s}) = n \sum_{i=1}^{n} w_{ni} R_n(Z_{(i)}, Y_{(i)}; \mathbf{s}) - n^{1/2} \sum_{i=1}^{n} w_{ni}\text{sign}(y_{(i)} - Z'_{(i)}\boldsymbol{\beta}_0) Z'_{(i)}\mathbf{s}.$$

Let $Q_n(\mathbf{s}) = n \sum_{i=1}^{n} w_{ni} R_n(Z_{(i)}, Y_{(i)}; \mathbf{s})$. We first show that, for any fixed s,

$$Q_n(\mathbf{s}) \to_P \frac{1}{2}\mathbf{s}' A \mathbf{s}.$$

Introduce the empirical counterparts of $H(y)$, $\widetilde{H}^0(y)$ and $\widetilde{H}^{11}(\mathbf{z}, y)$:

$$\begin{aligned} H_n(y) &= n^{-1} \sum_{i=1}^{n} 1\{Y_i \le y\} \\ \widetilde{H}_n^0(y) &= n^{-1} \sum_{i=1}^{n} 1\{Y_i \le y, \delta_i = 0\} \\ \widetilde{H}_n^{11}(\mathbf{z}, y) &= n^{-1} \sum_{i=1}^{n} 1\{Z_i \le \mathbf{z}, Y_i \le y, \delta_i = 1\}. \end{aligned}$$

By Lemma 5.1 of Stute (1996),

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

Decompose $Q_n(\mathbf{s})$ into 4 parts:

$$Q_n(\mathbf{s}) = I_{n1} + I_{n2} + I_{n3} + I_{n4},$$

where

$$\begin{aligned}
I_{n1} &= \int nR_n(\mathbf{z}, y; \mathbf{s}) \exp\left\{\int_{-\infty}^{y-} \frac{\widetilde{H}^0(du)}{1 - H(u)}\right\} \widetilde{H}^{11}(d\mathbf{z}, dy), \\
I_{n2} &= \int nR_n(\mathbf{z}, y; \mathbf{s}) \exp\left\{\int_{-\infty}^{y-} \frac{\widetilde{H}^0(du)}{1 - H(u)}\right\} (\widetilde{H}_n^{11} - \widetilde{H}^{11})(d\mathbf{z}, dy), \\
I_{n3} &= \int nR_n(\mathbf{z}, y; \mathbf{s}) \left(\exp\left\{\int_{-\infty}^{y-} \frac{\widetilde{H}_n^0(du)}{1 - H_n(u)}\right\} - \exp\left\{\int_{-\infty}^{y-} \frac{\widetilde{H}^0(du)}{1 - H(u)}\right\}\right) d\widetilde{H}_n^{11}(\mathbf{z}, y), \\
I_{n4} &= \int \left(\exp\left\{\int_{-\infty}^{y-} n \ln(1 + \frac{1}{n(1 - H_n(u))}) \widetilde{H}_n^0(du)\right\} - \exp\left\{\int_{-\infty}^{y-} \frac{\widetilde{H}_n^0(du)}{1 - H_n(u)}\right\}\right) \\
&\quad \times nR_n(\mathbf{z}, y; \mathbf{s}) \, \widetilde{H}_n^{11}(d\mathbf{z}, dy).
\end{aligned}$$

Under (A3) and (A4), the first term

$$\begin{aligned}
I_{n1} &= \int nR_n(\mathbf{z}, y; \mathbf{s}) \exp\left\{\int_{-\infty}^{y-} \frac{\widetilde{H}^0(du)}{1 - H(u)}\right\} \widetilde{H}^{11}(d\mathbf{z}, dy), \\
&= \int nR_n(\mathbf{z}, y; \mathbf{s}) \widetilde{F}^0(d\mathbf{z}, dy) \\
&= E[nR_n(Z, T; \mathbf{s})] = \mathbf{s}' E[ZZ' f_\varepsilon(0|Z)]\mathbf{s} + o(1).
\end{aligned}$$

The second term

$$\begin{aligned}
I_{n2} &= \sum_{i=1}^{n} nR_n(\mathbf{z}_i, y_i; \mathbf{s}) \delta_i \gamma_0(y_i) - E[nR_n(Z, T \wedge C; \mathbf{s}) \delta \gamma_0(T \wedge C)] \\
&= \mathbb{G}_n \left(\sqrt{n} R_n(\mathbf{z}, t \wedge c; \mathbf{s}) \delta \gamma_0(t \wedge c)\right)
\end{aligned}$$

With condition (A5b), I_{n2} converges to 0 in probability by Lemma 19.31 of Van der Vaart (1998).

For the third term I_{n3}, we first note that, for $\eta < \tau_Y$,

$$
\begin{aligned}
& \sup_{y\le\eta}\left|\exp\left\{\int_{-\infty}^{y-}\frac{\widetilde{H}_n^0(du)}{1-H_n(u)}-\int_{-\infty}^{y-}\frac{\widetilde{H}^0(du)}{1-H(u)}\right\}-1\right| \\
= & \sup_{y\le\eta}\left|\exp\left\{\int_{-\infty}^{y-}\frac{\widetilde{H}_n^0(du)}{1-H_n(u)}-\int_{-\infty}^{y-}\frac{\widetilde{H}_n^0(du)}{1-H(u)}+\int_{-\infty}^{y-}\frac{\widetilde{H}_n^0(du)}{1-H(u)}-\int_{-\infty}^{y-}\frac{\widetilde{H}^0(du)}{1-H(u)}\right\}-1\right| \\
= & \left|\exp\{o_P(1)+o_P(1)\}-1\right| = o_P(1),
\end{aligned}
$$

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

$$
R_n(\mathbf{z},y;\mathbf{s}) = \begin{cases} 2\left(n^{-1/2}\mathbf{z}'\mathbf{s}-(y-\mathbf{z}'\boldsymbol{\beta}_0)\right)1\{\mathbf{z}'\boldsymbol{\beta}_0<y<\mathbf{z}'(\boldsymbol{\beta}_0+n^{-1/2}\mathbf{s})\}, & \mathbf{z}'\mathbf{s}>0 \\ -2\left(n^{-1/2}\mathbf{z}'\mathbf{s}-(y-\mathbf{z}'\boldsymbol{\beta}_0)\right)1\{\mathbf{z}'(\boldsymbol{\beta}_0+n^{-1/2}\mathbf{s})<y<\mathbf{z}'\boldsymbol{\beta}_0\}, & \mathbf{z}'\mathbf{s}<0 \end{cases},
$$

Thus under (A5a),

$$
\begin{aligned}
I_{n3} &= \int nR_n(\mathbf{z},y;\mathbf{s})\gamma_0(y)\left(\exp\left\{\int_{-\infty}^{y-}\frac{\widetilde{H}_n^0(du)}{1-H_n(u)}-\int_{-\infty}^{y-}\frac{\widetilde{H}^0(du)}{1-H(u)}\right\}-1\right)\widetilde{H}_n^{11}(d\mathbf{z},dy) \\
&= o_P(1)\int nR_n(\mathbf{z},y;\mathbf{s})\gamma_0(y)\widetilde{H}_n^{11}(d\mathbf{z},dy) \\
&= o_P(1)(I_{n1}+I_{n2}) \\
&= o_P(1).
\end{aligned}
$$

For the last term I_{n4},

$$
\begin{aligned}
I_{n4} = & \int nR_n(\mathbf{z},y;\mathbf{s})\exp\left\{\int_{-\infty}^{y-}\frac{\widetilde{H}_n^0(du)}{1-H_n(u)}\right\} \\
& \left[\exp\left\{\int_{-\infty}^{y-} n\ln(1+\frac{1}{n(1-H_n(u))})-\frac{1}{1-H_n(u)}\widetilde{H}^0(du)\right\}-1\right]\widetilde{H}_n^{11}(d\mathbf{z},dy)
\end{aligned}
$$

The expression in the square brackets is bounded between $[-2n(1-H_n(y-))]^{-1}$ and 0. Because

$nR_n(\mathbf{z}, y; \mathbf{s})$ vanishes when y goes beyond $\mathbf{z}'\boldsymbol{\beta}_0$ or $\mathbf{z}'(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{s})$, by the generalized Glivenko-Cantelli theorem and condition(A5a),

$$\begin{aligned} I_{n4} &= O_P(n^{-1}) \int nR_n(\mathbf{z}, y; \mathbf{s}) \exp\left\{ \int_{-\infty}^{y-} \frac{\widetilde{H}_n^0(du)}{1 - H_n(u)} \right\} \widetilde{H}_n^{11}(d\mathbf{z}, dy) \\ &= O_P(n^{-1})(I_{n1} + I_{n2} + I_{n3}) \\ &= o_P(1) \end{aligned}$$

Therefore

$$M_n(s) = \frac{1}{2}\mathbf{s}' A\mathbf{s} - \mathbf{s}' n^{1/2} \sum_{i=1}^{n} w_{ni} \text{sign}(y_{(i)} - \mathbf{z}'_{(i)}\boldsymbol{\beta}_0)\mathbf{z}_{(i)} + o_P(1).$$

Under (A2)–(A5), by Theorem 3.1 in Stute (1996),

$$n^{1/2} \sum_{i=1}^{n} w_{ni} \text{sign}(Y_{(i)} - \mathbf{Z}'_{(i)}\boldsymbol{\beta}_0)\mathbf{Z}_{(i)} \rightarrow_D N(0, \Sigma).$$

By the Basic Corollary of Hjort and Pollard (1993), we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = A^{-1}\sqrt{n} \sum_{i=1}^{n} w_{ni} Z_i \text{sign}(Y_i - Z_i'\boldsymbol{\beta}_0) + o_p(1).$$

This completes the proof.

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		
Example	β_0	β_1	mse	β_0	β_1	mse
1	-0.004 (0.128)	1.004 (0.224)	0.034	0.007 (0.125)	0.991 (0.218)	0.031
2	0.011 (0.146)	1.005 (0.305)	0.054	-0.029 (0.132)	0.860 (0.247)	0.054
3	0.009 (0.253)	0.999 (0.463)	0.154	-0.101 (0.224)	1.005 (0.455)	0.117
4	-0.021 (0.162)	1.042 (0.283)	0.051	-0.007 (0.166)	1.021(0.292)	0.058
5	-0.032 (0.232)	1.039 (0.412)	0.114	-0.050 (0.214)	1.001 (0.363)	0.083
6	-0.024 (0.321)	1.058 (0.558)	0.264	-0.068 (0.300)	0.970 (0.506)	0.179

Table 2. PBC data: comparison of the KMW-LAD estimate with the Cox model estimate. s.e.: standard error.

	Cox		KMW-LAD	
covariate	estimate	s.e.	estimate	s.e.
$intercept$	NA	NA	3.974	6.190
age	0.031	0.011	-0.012	0.014
alb	-0.612	0.300	0.464	0.385
$log(alkphos)$	0.039	0.147	0.246	0.155
$ascites$	0.211	0.377	-0.543	0.826
$log(bili)$	0.632	0.178	-0.153	0.245
$log(chol)$	0.162	0.292	-0.032	0.378
$edtrt$	0.918	0.380	-0.854	0.926
$hepmeg$	-0.087	0.257	0.037	0.347
$log(platelet)$	0.132	0.285	-0.261	0.304
$log(protime)$	2.482	1.348	1.815	1.799
sex	-0.182	0.318	0.035	0.383
$log(sgot)$	0.406	0.309	-0.040	0.388
$spiders$	0.049	0.241	-0.123	0.318
$stage$	0.366	0.178	-0.071	0.140
trt	-0.006	0.212	0.127	0.243
$log(trig)$	-0.144	0.251	-0.192	0.381
$log(copper)$	0.284	0.176	-0.129	0.207