Smoothed rank-based procedure for censored data

Yudong Zhao

Lundbeck, Singapore
e-mail: yudong.zhao@gmail.com

Bruce M. Brown

School of Mathematics and Statistics, University of New South Wales, Sydney 2052,
Australia
e-mail: bruce.brown@unsw.edu.au

and

You-Gan Wang[*]

Centre for Applications in Natural Resource Mathematics (CARM), School of Mathematics
and Physics, The University of Queensland, Queensland 4072, Australia
e-mail: you-gan.wang@uq.edu.au

Abstract: A smoothed rank-based procedure is developed for the accelerated failure time model to overcome computational issues. The proposed estimator is based on an EM-type procedure coupled with the induced smoothing. The proposed iterative approach converges provided the initial value is based on a consistent estimator, and the limiting covariance matrix can be obtained from a sandwich-type formula. The consistency and asymptotic normality of the proposed estimator are also established. Extensive simulations show that the new estimator is not only computationally less demanding but also more reliable than the other existing estimators.

Keywords and phrases: Accelerated failure time model, Buckley-James estimator, induced smoothing, rank-based procedure.

Contents

[*]Corresponding author.

1. Introduction

The accelerated failure time (AFT) model is a useful alternative to the popular Cox regression model for survival data analysis, especially when the key assumptions of the Cox model, for example, proportional hazards, are not guaranteed to hold in practice. The direct interpretability of the AFT model in terms of life expectancy is more desirable for practitioners, compared to the indirect interpretation based on hazards in the Cox model. For more discussion, see Wei (1992), among others.

Buckley and James (1979) proposed solving normal equations with censored responses replaced by their expected life length based on the Kaplan-Meier estimate. Simulation studies conducted by Miller and Halpern (1982) showed that the Buckley-James estimator is reliable. However, its asymptotic properties are difficult to establish; see Lai and Ying (1991). In addition, the iteration involved can become trapped in a loop and thus may never converge. Non-convergence occurs more frequently when censoring occurs more severely. Stare et al. (2000) suggested that it is safe to use the Buckley-James estimator only when censoring is less than 20%.

Ritov (1990) and Tsiatis (1990), among others, studied rank estimators based on the linear rank statistics for censored data proposed by Prentice (1978). A Gehan-type weight function is commonly used, and is a starting point in considering rank estimators, as the function leads to a monotonic estimating function. Although other weighting schemes could yield more efficient estimates, the monotonicity of their associated estimating functions is not ensured, and the uniqueness of the estimates is not guaranteed. Jin et al. (2003) developed an iterative procedure for approximating possibly non-monotone weighted log-rank estimating functions around the true values of regression parameters. The iteration algorithm starts with a Gehan-type estimator, and solves a monotone estimating function at each step with updated weights. The covariance matrix of the estimated regression coefficients are then estimated by using a resampling approach to approximate their distributions. However, although this numerical procedure can be conveniently implemented, for large datasets the high computational burden resulting from resampling techniques is undesirable.

To overcome difficulties in estimating the covariance matrix, usually caused by non-smooth estimating functions, Brown and Wang (2005) developed an induced smoothing technique, resulting in continuously differentiable objective or estimating functions, which makes the standard numerical methods applicable. This induced smoothing technique provides a natural method for determining kernel bandwidths when replacing discontinuous functions in estimating equations with smoothed distribution kernels. Recently, Chiou et al. (2013) developed a fast and efficient rank-based inference procedure for the AFT model in the context of case-cohort studies by adopting the induced smoothing approach.

Consider the linear regression model,

$$Y_i = X_i^T \beta_0 + \varepsilon_i; \quad \varepsilon_i, i = 1 \dots n, \text{ are independent and identically distributed.}$$

Denote $X_i - X_j$ by d_{ij}. The corresponding rank estimating function is,

$$
\begin{aligned}
S_n(\beta) &= -\frac{1}{n(n-1)} \sum_{i<j}^{n} \mathrm{sgn}\{Y_i - Y_j - d_{ij}^T \beta\} d_{ij} \\
&= -\frac{1}{n(n-1)} \sum_{i<j}^{n} \mathrm{sgn}\{\varepsilon_i - \varepsilon_j\} d_{ij}.
\end{aligned}
\tag{1}
$$

The smoothed rank estimating function derived by Brown and Wang (2005) is,

$$
\widetilde{S}_n(\beta) = -\frac{1}{n(n-1)} \sum_{i \neq j}^{n} \left\{ \Phi\left(\frac{Y_i - Y_j - d_{ij}^T \beta}{\sqrt{d_{ij}^T \Gamma d_{ij}}} \right) - \frac{1}{2} \right\} d_{ij},
$$

in which Γ is a $p \times p$ covariance matrix of $\hat{\beta}$, the root of the smoothed rank estimating function. Γ can be estimated with a sandwich formula $\hat{\Gamma} = D^{-1} \mathrm{cov}\{\widetilde{S}_n(\beta_0)\}\{D^T\}^{-1}$, where $D = E\{\widetilde{S}_n'(\beta)\}_{\beta_0}$, and $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. Here, D and $\mathrm{cov}\{\widetilde{S}(\beta_0)\}$ can be replaced by their estimates $\hat{D} = \widetilde{S}_n'(\hat{\beta})$ and $\widehat{\mathrm{cov}}\{\widetilde{S}_n(\beta_0)\}$, respectively. By starting from an initial matrix of $\Gamma_{(0)}$ based on an initial estimate of $\beta_{(0)}$, we can obtain $\beta_{(k+1)}$ by iterating between solving $\widetilde{S}_{(k)}(\beta)$ using $\Gamma_{(k)}$, and updating $\Gamma_{(k+1)}$. Thus, the unique estimator of β_0 and its estimated covariance matrix $\hat{\Gamma}$ are the limits of the sequence of $\beta_{(k)}$ and $\Gamma_{(k)}, k \geq 1$. The smoothed rank estimating function with variable kernel bandwidths has an appealing rationale. The consistent estimator $\hat{\beta}$ to be obtained can be stochastically expressed as $\beta_0 + \Gamma^{\frac{1}{2}} N(0, I_p)$, which is a random perturbation to the true parameter values β_0. This leads us to consider $\widetilde{S}(\beta_0) = E\{S(\beta_0 + \Gamma^{\frac{1}{2}} N(0, I_p))\}$, an average of random perturbations of S. Since monotonicity of the estimating equations is needed for the induced smoothing method, application to the AFT models is made by Brown and Wang (2006) only for the Gehan estimating equations.

In this article, we investigate the induced smoothing approach for the AFT model together with the EM algorithm, i.e, iterating between imputing the censored lifetime using their expected conditional survival time and solving the smoothed rank estimating equation for "complete data." To overcome non-smoothness and non-monotonicity due to imputation for the censored lifetime based on the Kaplan-Meier estimate of survival functions, we first propose in Section 2 an estimate of the conditional expectation of residuals using kernel smoothing, which is smooth in the residuals, but also more importantly the regression coefficient β. The choice of the kernel function and bandwidths are discussed in Section 3. Using the consistent Gehan estimator as the initial value, we develop an iteration procedure consisting of imputing censored residuals with

their smoothed estimates and solving smoothed monotone rank estimation equations established by using the induced smoothing approach in Section 4. In light of the smoothness of the estimating functions in each iteration, we can obtain the limiting covariance matrix by the standard procedure. Theoretical justifications for the locally asymptotic linearity of the resulting estimating functions and asymptotic normality of the proposed estimators are provided. Simulation results and two real examples are also presented in Section 5 and Section 6, to demonstrate the usefulness of the proposed method.

2. Smoothed estimates of conditional mean residuals

Suppose that there is a random sample of n subjects. Let T_i and C_i be the failure time and censoring time, respectively, for the ith subject and X_i be the associated $p \times 1$ vector of covariates. It is usually assumed that T_i and C_i are independent, conditional on X_i. The observed data may be written as a triplet $(\tilde{T}_i, \delta_i, X_i); 1 \leq i \leq n$, where $\tilde{T}_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$ is the censoring indicator. The usual accelerated failure time model takes the form,

$$\log(T_i) = X_i^T \beta_0 + \varepsilon_i,$$

where β_0 is the true value of a $p \times 1$ vector of regression coefficients, and $\{\varepsilon_i; 1 \leq i \leq n\}$ are independent error terms following a common distribution F.

Denote by $\{Y_i = \log(\tilde{T}_i), 1 \leq i \leq n\}$ the observed responses. For fixed β, in the presence of censoring, the observed residual, $\tilde{e}_i = Y_i - X_i^T \beta$ is not the actual residual any more for the ith subject. Note that $\tilde{e}_i = \min(e_i, c_i)$, where $e_i = \log T_i - X_i^T \beta$ and $c_i = \log C_i - X_i^T \beta$, and thus, whenever the observed lifetime is censored, the corresponding residual is also censored. Buckley and James (1979) suggested replacing the censored residual $\tilde{e}_i$ with the conditional expectation of e_i, which can be expressed as

$$m(\tilde{e}_i; \beta) = E(e_i | e_i > \tilde{e}_i) = \frac{\int_{\tilde{e}_i(\beta)}^{\infty} u dF_\beta(u)}{1 - F_\beta\{\tilde{e}_i(\beta)\}}. \tag{2}$$

In the Buckley-James estimator for the AFT model, F is replaced by its Kaplan-Meier estimator $\hat{F}$ and the censored observation Y_i by

$$X_i^T \beta + \frac{\int_{\tilde{e}_i(\beta)}^{\infty} u d\hat{F}_\beta(u)}{1 - \hat{F}_\beta\{\tilde{e}_i(\beta)\}}. \tag{3}$$

The estimated conditional mean lifetime in (3) is piecewise linear in β with jumps where there is a change in the ranks of $\{\tilde{e}_i(\beta), 1 \leq i \leq n\}$. When imputing in the subsequent iteration procedures for updating β, the discontinuous estimate of $m(\tilde{e}_i; \beta)$ can lead to non-uniqueness of β estimates, which is indicated by iterates oscillating between two values.

To deal with the non-smoothness caused by applying the Kaplan-Meier estimate $\hat{F}$, smoothed estimates of F can be used instead in (2). Numerous

smoothed estimates for lifetime survival functions have been proposed and studied in the literature. They could be adopted here to obtain a smoothed estimate of the residual distribution. However, care is needed because the smoothness property is required not only in the residuals but, more importantly, also, in the regression coefficients β. Bearing this in mind, we start by considering the kernel estimator for the hazard rate function $\alpha(t)$, developed by Ramlau-Hansen (1983).

Let $E_1 < E_2 < \cdots$ be the ordered successive jump residuals for a given β. This kernel estimator of hazard rate was derived by smoothing the increments of the Nelson-Aalen estimator and may be written as, in terms of residuals,

$$\hat{\alpha}(e) = h_1^{-1} \sum_j k\left(\frac{e - E_j}{h_1}\right) (Y(E_j))^{-1}, \tag{4}$$

where $Y(e) = n - \sum_{i=1}^{n} I(\tilde{e}_i < e) = n(1 - F_n(e))$, and $F_n(e) = n^{-1} \sum_{i=1}^{n} I(\tilde{e}_i < e)$. Here, the kernel function k is a bounded function that vanishes outside $[-1, 1]$ and has integral 1. The bandwidth h_1 is a positive parameter. By assuming a continuous underlying distribution of $\tilde{e}_i$, $i = 1 \ldots n$, further kernel smoothing applied to the at-risk process $Y(t)$ gives $\hat{Y}(e) = n(1 - \hat{F}_n(e))$, where $\hat{F}_n(e) = n^{-1} \sum_{i=1}^{n} K\{(e - \tilde{e}_i)/h_2\}$ and $K(t) = \int_{-\infty}^{t} k(u)du$; see Azzalini (1981) among others. Here, the bandwidth parameters h_1 and h_2 are to be determined among uncensored subjects, and among all subjects, respectively. Substituting $Y(e)$ with $\hat{Y}(e)$ in $\hat{\alpha}(e)$, we may further derive

$$\tilde{\alpha}(e) = (nh_1)^{-1} \sum_j \frac{k(\frac{e - E_j}{h_1})}{1 - n^{-1} \sum_{i=1}^{n} K(\frac{E_j - \tilde{e}_i}{h_2})}. \tag{5}$$

Then a kernel-based estimator $\tilde{A}(e)$ for the cumulative hazard function $A(e) = \int_{-\infty}^{e} \alpha(u)du$ is

$$\tilde{A}(e) = \sum_{j=1}^{n} \frac{\delta_j K(\frac{e - \tilde{e}_j}{h_1})}{n - \sum_{i=1}^{n} K(\frac{\tilde{e}_j - \tilde{e}_i}{h_2})}. \tag{6}$$

Note that $\tilde{A}(e)$ is continuous in e, and thus the conditional survival distribution of the residual errors can be subsequently estimated by $\tilde{S}(e) = \exp^{-\tilde{A}(e)}$ for a given β and convariates x. Let $\{\tilde{e}_{(i)}; 1 \leq i \leq n\}$ be the ordered residuals of $\{\tilde{e}_i; 1 \leq i \leq n\}$. Substituting the kernel estimator $\tilde{S}(e)$ into (2) for a censored residual $\tilde{e}_i$ leads to a smoothed estimator for $m(\tilde{e}_i, \beta)$,

$$\tilde{m}(\tilde{e}_i; \beta) = -\frac{\int_{\tilde{e}_i}^{\infty} e d\tilde{S}(e)}{\tilde{S}(\tilde{e}_i)} = \frac{\sum_{k=j+1}^{n+1} \{\tilde{S}(\tilde{e}_{(k-1)}) - \tilde{S}(\tilde{e}_{(k)})\}\{\tilde{e}_{(k-1)} + \tilde{e}_{(k)}\}}{2\tilde{S}(\tilde{e}_i)}, \tag{7}$$

where j is the rank of $\tilde{e}_i$ $(j = 1, 2, \ldots, n)$, i.e., $\tilde{e}_i = \tilde{e}_{(j)}$. For convenience, we define $\tilde{e}_{(n+1)} = 0$, and $\tilde{S}(\tilde{e}_{(n+1)}) = 0$. It is obvious that $\tilde{m}(\cdot)$ is smooth in the residuals and the regression coefficients β. Under the conditions required in

Ramlau-Hansen ([1983](#)) for showing the consistency of $\hat{\alpha}(t)$ in ([4](#)), we have the following consistency results in Theorem 1.

Theorem 1. *Denote the support of F by $\mathbb{E} = (-\infty, T]$, where $T = \sup\{u : F(u) < 1\} < \infty$. Let k be a bounded density kernel of order q and the bandwidths $h_1 \to 0$ and $h_2 \to 0$ satisfying $h_2 > \log(n)/n$, as $n \to \infty$. For any given interior point $e \in \mathbb{E}$, if there exists an $\varepsilon > 0$ such that $\inf_{s \in [e-\varepsilon, e+\varepsilon]} h_1 Y(s) \xrightarrow{P} \infty$ and $\max(\sqrt{h_2 \log(1/h_2)/n}, h_2^q) = o(h_1)$, the kernel smoothed estimators $\tilde{\alpha}(e)$ and $\tilde{A}(e)$ are pointwise consistent, and $\sup_{e \in \mathbb{E}} |\tilde{m}(e; \beta) - m(e; \beta)| \xrightarrow{P} 0$.*

3. Choice of bandwidths and kernel functions

If smoothing the at-risk process is regarded as smoothing a survival distribution for complete data, the optimal choice for h_2 is of order $n^{-1/(2q-1)}$; while the optimal choice for h_1 is of order $N_1^{-1/(2q+1)}$ for estimating the kernel density in $\tilde{\alpha}(e)$, where N_1 is the number of uncensored subjects. See Azzalini ([1981](#)) for a detailed discussion. In these cases of h_1 and h_2, from Giné and Nickl ([2009](#)), it can be shown that for $q \geqslant 1$, $\max(\sqrt{h_2 \log(1/h_2)/n}, h_2^q) = O(n^{-q/(2q-1)}\sqrt{\log n}) = o(n^{-1/2})$, and thus, that the convergence in Theorem 1 follows.

As defined in estimating the conditional expectation of regression residuals, the kernel function k is required to be a bounded function that vanishes outside $[-1, 1]$ and integrates to 1. Although any kernel functions satisfying this requirement can be used and we expect that the choice of kernel function is a secondary matter compared with selecting the bandwidths as shown in Andersen et al. ([1992](#)), we suggest using the Epanechnikov kernel for convenience. This also makes K in the denominator of $\tilde{\alpha}(e)$ the cumulative Epanechnikov kernel, which is

$$
K(t) = \begin{cases} 0, & t \leq -1 \\ -t^3/4 + 3t/4 + 1/2, & |t| < 1 \\ 1, & t \geqslant 1. \end{cases}
$$

Corresponding to the Epanechnikov kernel, in all the simulation studies performed in Section [5](#) and the examples shown in Section [6](#), the optimal bandwidth $(40\sqrt{\pi})^{1/5}\sigma_1 N_1^{-1/5}$ for h_1 was used, and h_2 was set to be $1.3\sigma_2 N_2^{-1/3}$, where $N_2 = n$. Here σ_1 and σ_2 are estimates of the standard deviations of $e = \log T - \hat{\beta}^T X$, respectively, among the uncensored subjects and among all subjects. In consideration of heavy-tailed and skewed errors, they can be estimated, following Silverman ([1986](#)), by

$$
\hat{\sigma} = \min\left\{\hat{s}, \frac{Q_3 - Q_1}{1.34}\right\},
$$

where $\hat{s}$, Q_1 and Q_3 are the sample standard deviation, the first and third sample quartiles, among N_1 uncensored subjects when estimating σ_1, or among all N_2 subjects when estimating σ_2, respectively. The scale parameter $(40\sqrt{\pi})^{1/5}\sigma_1$ in h_1 is calculated by referring to a normal underlying distribution and using

the Epanechnikov kernel in the optimal smoothing parameter suggested in Silverman (1986), which minimizes the approximate mean integrated square error (AMISE) for the kernel density. When determining the scale parameter for h_2, we recommend using $1.3\sigma_2$ for general use for ease of computation, which has been shown in Azzalini (1981) as a reliable choice against various underlying distributions for the Epanechnikov kernel. Although plug-in methods or even cross-validation methods may provide a more accurate approximation to the bandwidth minimizing AMISE, see Altman and Leger (1995) among others, the high computational burden required makes them practically difficult to implement in the estimating procedure proposed later.

4. Smoothed rank procedure for the AFT model

In this section, we propose an EM-type procedure for the AFT model, iterating between estimating the kernel smoothed conditional expectations of the censored residuals developed in Section 2 and solving the ordinary rank-based estimating equations for complete data, with censored residuals being imputed by their estimates obtained in the expectation step. The induced smoothing technique is also introduced into this framework to facilitate estimating the covariance matrices required in the iteration procedure.

With the censored residuals being imputed by their smoothed estimates of their conditional expectations and replaced in (1) for the ordinary rank-based estimating function for complete data, an estimating function can be defined as

$$
U(\beta) = -\frac{1}{n(n-1)} \sum_{i<j}^{n} \mathrm{sgn}\big\{\delta_i \tilde{e}_i + (1-\delta_i)\tilde{m}(\tilde{e}_i; \beta) - \delta_j \tilde{e}_j - (1-\delta_j)\tilde{m}(\tilde{e}_j; \beta)\big\} d_{ij}
$$

$$
= -\frac{1}{n(n-1)} \sum_{i<j}^{n} \mathrm{sgn}\big\{Y_i - Y_j - d_{ij}^T\beta + V_i - V_j\big\} d_{ij}, \tag{8}
$$

where $V_k = (1-\delta_k)\{\tilde{m}(\tilde{e}_k; \beta) - \tilde{e}_k\}, k = i, j$. Conditional on $\{V_i; 1 \le i \le n\}$ for censored residuals, we may approximate $U(\beta)$ with the smoothed $\tilde{U}(\beta)$ using the induced smoothing technique as follows:

$$
\tilde{U}(\beta) = -\frac{1}{n(n-1)} \sum_{i \ne j}^{n} \left\{ \Phi\left(\frac{Y_i - Y_j - d_{ij}^T\beta + V_i - V_j}{\sqrt{d_{ij}^T \Gamma d_{ij}}}\right) - \frac{1}{2} \right\} d_{ij}, \tag{9}
$$

where $\Gamma = D^{-1}\mathrm{cov}\{\tilde{U}(\beta_0)\}\{D^T\}^{-1}$ and $D = E\{\widetilde{U}'(\beta)\}_{\beta_0}$. In the sequel, we assume $n^{-1}(n-1)^{-1}\sum_{i \ne j}^{n} d_{ij}d_{ij}^T \le M$ for a constant value M, which is reasonable as long as $||X_i||$ is bounded almost surely by a non-random constant for $1 \le i \le n$. Note that $\tilde{U}(\beta)$ is asymptotically equivalent to $U(\beta)$ as presented in Lemma 1 although unconditionally it is not established by directly taking an overall average on $U(\beta)$ after perturbing β everywhere whenever it appears inside the estimating function. In the iteration procedure developed later where $\{V_i; 1 \le i \le n\}$ have been pre-fixed from the previous estimation step, $\tilde{U}(\beta)$ is naturally obtained as if there is no censored data.

Lemma 1. *Under the conditions*

1. *The covariates X are uniformly bounded and*
2. *The probability density functions for $\tilde{E}_{ij} = \delta_i \tilde{e}_i + (1 - \delta_i)m(\tilde{e}_i; \beta_0) - \delta_j \tilde{e}_j - (1 - \delta_j)m(\tilde{e}_j; \beta_0)$ for all $i \neq j$ are bounded,*

the induced estimating functions $\tilde{U}(\beta)$ at $\beta = \beta_0$ are asymptotically equivalent to $U(\beta_0)$ in probability, i.e., $\sqrt{n}||\tilde{U}(\beta_0) - U(\beta_0)|| = o_p(1)$.

Since it is reasonable to assume that $\delta_i \tilde{e}_i + (1 - \delta_i)m(\tilde{e}_i, \beta_0)$, $i = 1 \ldots n$, are identically distributed, we have that $\Pr\{\delta_i \tilde{e}_i + (1 - \delta_i)m(\tilde{e}_i; \beta_0) > \delta_j \tilde{e}_j + (1 - \delta_j)m(\tilde{e}_j; \beta_0)\} = 1/2$. From Lemma 1, this indicates that $E\{U(\beta_0)\} \to 0$ and thus, $E\{\tilde{U}(\beta_0)\} \to 0$ as $n \to \infty$. In light of the continuity of $\tilde{U}(\beta)$ everywhere in β, $\tilde{U}(\beta)$ is locally linear at β_0 by applying the Taylor expansion.

Suppose the root of $\tilde{U}(\beta)$ is $\hat{\beta}_{DSR}$. We show in the Appendix that $\hat{\beta}_{DSR}$ is a consistent estimator of β_0 and $n^{1/2}(\hat{\beta}_{DSR} - \beta_0)$ is asymptotically normally distributed as $N(0, n\Gamma)$. Denote the covariate vectors corresponding to the ordered $\{\tilde{e}_{(i)}, 1 \leq i \leq n\}$ by $\{X_{(i)}, 1 \leq i \leq n\}$. Because $\{\tilde{m}(\tilde{e}_i, \beta), 1 \leq i \leq n\}$ for censored residuals are continuous in β, asymptotically we may estimate D by $\hat{D} = \{\tilde{U}'(\beta)\}_{\hat{\beta}}$, i.e.,

$$\hat{D} = \frac{1}{n(n-1)} \sum_{i \leq j} \frac{d_{ij}(d_{ij} - v_{ij})^T}{\sqrt{d_{ij}^T \Gamma d_{ij}}} \phi \left(\frac{Y_i - Y_j - d_{ij}^T \hat{\beta} + V_i - V_j}{\sqrt{d_{ij}^T \Gamma d_{ij}}} \right), \qquad (10)$$

where $\hat{\beta}$ is an estimate of β_0 obtained from a consistent estimator, $v_{ij} = v_i - v_j$ and for $k = i, j$

$$v_k = \frac{\partial V_k}{\partial \beta}\Big|_{\hat{\beta}} = (1 - \delta_k) \left\{ \tilde{m}(\tilde{e}_k, \beta) \frac{\partial \tilde{A}(\tilde{e}_k)}{\partial \beta} - \tilde{S}^{-1} \frac{\partial \int_{\tilde{e}_k}^\infty u d\tilde{S}(u)}{\partial \beta} + X_k \right\} \qquad (11)$$

in which

$$\frac{\partial \tilde{A}(\tilde{e}_k)}{\partial \beta}\Big|_{\hat{\beta}}$$

$$= -\sum_{j=1}^n \frac{\delta_j \left[\{n - \sum_{i=1}^n K(\frac{\tilde{e}_j - \tilde{e}_i}{h_2})\} k(\frac{\tilde{e}_k - \tilde{e}_j}{h_1}) d_{kj} h_1^{-1} + K(\frac{\tilde{e}_k - \tilde{e}_j}{h_1}) \{\sum_{i=1}^n k(\frac{\tilde{e}_j - \tilde{e}_i}{h_2}) d_{ji} h_2^{-1}\} \right]}{\{n - \sum_{i=1}^n K(\frac{\tilde{e}_j - \tilde{e}_i}{h_2})\}^2}, \qquad (12)$$

and for $\tilde{e}_{(j)} = \tilde{e}_k$, $\tilde{e}_{(n+1)} = 0$, $X_{(n+1)} = 0_{p \times p}$ and $S(\tilde{e}_{(n+1)}) = 0$

$$\frac{\partial \int_{e_k}^\infty u d\tilde{S}(u)}{\partial \beta}\Big|_{\hat{\beta}}$$

$$= \frac{1}{2} \left[\sum_{i=j+1}^{n+1} \left\{ \tilde{S}(\tilde{e}_{(i-1)}) \frac{\partial \tilde{A}(\tilde{e}_{(i-1)})}{\partial \beta} - \tilde{S}(\tilde{e}_{(i)}) \frac{\partial \tilde{A}(\tilde{e}_{(i)})}{\partial \beta} \right\} \{\tilde{e}_{(i-1)} + \tilde{e}_{(i)}\} \right.$$

$$\left. + \left\{ \tilde{S}(\tilde{e}_{(i-1)}) - \tilde{S}(\tilde{e}_{(i)}) \right\} \{X_{(i-1)} + X_{(i)}\} \right]. \qquad (13)$$

Because of the asymptotic equivalence of $\tilde{U}(\beta_0)$ and $U(\beta_0)$, the asymptotic co-variance matrix $\text{cov}\{U(\beta_0)\}$ needs to be estimated next. This can be approached by applying the Hajek projection method to

$$\hat{U}_2(0;\beta_0) = -\frac{1}{n(n-1)}\sum_{i<j}\text{sgn}\{\delta_i\tilde{e}_i+(1-\delta_i)m(\tilde{e}_i,\beta_0)-\delta_j\tilde{e}_j-(1-\delta_j)m(\tilde{e}_j,\beta_0)\}d_{ij},$$

which is asymptotically equivalent to $U(\beta_0)$ as obtained from Lemma 1. It follows estimating $\text{cov}\{U(\beta_0)\}$ by

$$\widehat{\text{cov}}\{U(\beta_0)\} = \frac{4}{n^2(n-1)^2}\sum_{i=1}^{n}\hat{\xi}_i\hat{\xi}_i^T \tag{14}$$

where

$$\hat{\xi}_i = \sum_{j\neq i}^{n}\left[I\left\{\delta_i\tilde{e}_i + (1-\delta_i)\tilde{m}(\tilde{e}_i;\hat{\beta}) > \delta_j\tilde{e}_j + (1-\delta_j)\tilde{m}(\tilde{e}_j;\hat{\beta})\right\} - \frac{1}{2}\right]d_{ij}.$$

The proposed estimating function $\tilde{U}(\beta)$ is smooth but not necessarily mono-tone in β, i.e., $\hat{\beta}_{DSR}$ is not necessarily a unique root of $\tilde{U}(\beta)$. However, in light of the local linearity of $\tilde{U}(\beta)$, we propose to approximate $\hat{\beta}_{DSR}$ with the following iterative algorithm.

Step 1: Initial calculation. Use the induced smoothing method in Brown and Wang (2006) for the AFT model to get the Gehan rank estimator $\hat{\beta}_G$ and let $\hat{\beta}_{(0)} = \hat{\beta}_G$.

Step 2: Updating conditional mean residuals. Estimate $\{\tilde{m}(\tilde{e}_i;\hat{\beta}_{(m)}), 1 \leq i \leq n\}$ for censored residuals using the kernel method proposed in Section 2.

Step 3: Updating regression estimates. Solving $\hat{\beta}_{(m+1)}$ from $\tilde{U}(\beta) = 0$ as if $\{Y_i + V_i; 1 \leq i \leq n\}$ are completely observed lifetimes. To solve the estimat-ing equation, the numerical algorithm proposed in Brown and Wang (2005) is embedded in Step 3, consisting of

1) Using $\hat{\beta}_{(m)}$ and $\Gamma_{(m)}$ as initial values
2) Updating $\hat{\beta}_{(m)}$ by solving $\tilde{U}(\beta) = 0$ with Γ being fixed at the updated $\Gamma_{(m)}$
3) Updating (10) and (14) by using the updated $\hat{\beta}_{(m)}$ and hence $\Gamma_{(m)}$
4) Repeating 2) and 3) until convergence. The updated $\hat{\beta}_{(m)}$ and $\Gamma_{(m)}$ after convergence are denoted as $\hat{\beta}_{(m+1)}$ and $\Gamma_{(m+1)}$

Step 4: Continuation. Repeat Steps 2 and 3 until convergence.

By starting from an initial value obtained from a consistent estimator, such as the Gehan rank estimator $\hat{\beta}_G$, we show in the Appendix that the mth-step estimator

$$\hat{\beta}_{(m)} = \beta_0 + (I - B^{-1}D)^m(\hat{\beta}_G - \beta_0) - (I - (I - B^{-1}D)^m)D^{-1}\tilde{U}(\beta_0)$$
$$+ o_p(n^{-1/2}) \tag{15}$$

where

$$
B = \frac{1}{n(n-1)} E \sum_{i \neq j} \frac{d_{ij} d_{ij}^T}{\sqrt{d_{ij}^T \Gamma d_{ij}}} \phi \left(\frac{Y_i - Y_j - d_{ij}^T \beta + V_i - V_j}{\sqrt{d_{ij}^T \Gamma d_{ij}}} \right),
$$

is a consistent estimator of β_0 and asymptotically normally distributed. Provided that $\hat{\beta}_{(m-1)}$ at the $(m-1)$th-step of iteration is a fixed and consistent estimator of β_0 so that the imputed lifetime $Y_i + V_i, i = 1 \ldots n$ and $\Gamma_{(m-1)}$ are given, it can be seen that $\beta_{(m)}$ is a unique root of $\tilde{U}\{\hat{\beta}, \hat{\beta}_{(m-1)}\}$, the estimating function in (9) at the mth-step, since now this conditional function, as an ordinary rank-based estimating function for complete data, is smooth and monotone. By starting from a fixed initial value obtained from a consistent estimator, by induction $\hat{\beta}_{(m)}$ is a unique estimator at the mth-step of iteration.

Apparently, $\hat{\beta}_{(m)}$ approaches the usual smoothed rank estimator when the censoring vanishes. When the hazard $\alpha(e)$ is nondecreasing in the residual e, we further show that the matrix $B - D$ is nonnegative, which implies from (15) that when m tends to infinity, $(I - B^{-1}D)^m$ tends to a zero matrix, and $\hat{\beta}_{(m)}$ uniquely estimates β_0 as lomg as $\hat{\beta}_{(0)}$ is consistent. The covariance matrix of $\hat{\beta}_{(m)}$ approaches $\hat{\Gamma} = D^{-1} \text{cov}\{\tilde{U}_n(\beta_0)\}\{D^T\}^{-1}$ after sufficiently many steps of iteration m, which in turn approaches to Γ as $n \to \infty$.

5. Simulations

To examine the finite sample performance of the proposed rank procedure, we considered the same model simulated by Jin et al. (2006a). Specifically, the failure times were generated from the following model:

$$
\log T = 2 + X_1 + X_2 + \epsilon,
$$

where X_1 follows Bernoulli with .5 success probability and X_2 has the independent normal distribution $N(0, 0.5^2)$. Four different error distributions for ϵ: the standard normal distribution; the standard logistic distribution; the extreme-value distribution; and a mixture of $N(0, 1)$ and $N(0, 9)$ with mixing probabilities (0.9, 0.1), denoted by $0.9N(0, 1) + 0.1N(0, 9)$, and one Weibull distribution with hazard rate $1/(2\sqrt{t})$ for $\exp(\epsilon)$, denoted by Weibull(0.5, 1), were investigated. We considered random samples of 100 and 200 subjects. The censoring times were generated from the uniform distribution $\text{Un}[0, c]$, where c was chosen to yield separately a 25% and a 50% censoring rate for each scenario.

In addition to the proposed estimator (DSR), we included the profile likelihood estimator (PLE) proposed by Zeng and Lin (2007), the log-rank (Logrank), the least squares (LSQ) estimators proposed by Jin et al. (2003, 2006a), and the Gehan estimator for efficiency comparisons. The log-rank and least squares estimators are asymptotically efficient under the extreme-value and normal error distributions, respectively. The profile likelihood method demonstrated substantial gain in efficacy for ϵ following Weibull(2, 1) in the simulations conducted by Zeng and Lin (2007). We did not include this case as it

might be more informative for comparisons after further transformations to the LSQ, Log-rank, and DSR methods for the Weibull(2, 1) error of L-shape than comparing them directly. Note that the results for the Gehan estimator asymptotically also represented the performance of its smoothed version proposed by Brown and Wang (2006), as well as the one proposed by Heller (2007), which is essentially the same as Brown and Wang (2006) as commented in Johnson and Strawderman (2009).

The results of the simulations based on 1000 repetitions for $n = 100$ and $n = 200$ are summarized in Table 1. In Table 1, we find that the proposed parameter estimators of β_1 and β_2 are virtually unbiased. The standard errors of the PLE, Log-rank, and LSQ estimators are very similar to those obtained for the same scenarios in Zeng and Lin (2007) and Jin et al. (2003, 2006a), respectively, though not presented here. The induced smoothing procedure for estimating variances appears to reflect true variations, and the confidence intervals had proper coverage probabilities. The Epanechnikov kernel and the bandwidths used worked well, and the approximation of the conditional mean residual seemed appropriate in producing reasonable estimates.

With the normal error and the 25% censoring rate, the proposed estimator was slightly less efficient than the least squares estimator, but more efficient than the other two estimators. For the 50% censoring rate, it outperformed all the other estimators, even including the least squares estimator. Under the extreme-value error, the proposed estimator was less efficient than the log-rank estimator, but much more efficient than the other two estimators. Under the logistic and mixture normal error distributions, the proposed estimators were more efficient than the other three estimators. For the Weibull case, the relative performance of all the methods was, as expected, consistent with those for the extreme value distribution, because a log-Weibull distribution of ϵ is theoretically equivalent to a scaled extreme value distribution. In addition, the least squares method also performed well for the logistic apart from normal, but poorly for the other error distributions. The Log-rank regression is most efficacious for the distributions scaled to the extreme value. However, its performance was generally the worst for all the other scenarios. In general, the PLE estimator performed well when the censoring rate was 25%. However, when the censoring rate was increased to 50%, it showed no better performance than the least squares and log-rank estimators. The Gehan estimator performed well for normal and logistic distributions for the 25% censoring rate. However, the performance was pretty much poor for the 50% censoring rate. In summary, the doubly smoothed estimator appeared to retain relatively high efficiency against different underlying distributions. The efficiency gains were particularly substantial for severe censoring.

With a convergence criterion of 0.001, the proposed numerical procedures converged very rapidly with a 25% censoring rate, usually within 5 iterative steps. For the scenarios with a 50% censoring rate, convergence was always achieved between 10 to 50 steps, depending on the sample sizes. Table 2 demonstrates the computing time for the rank-based procedures in seconds for each replicate averaged among all the simulations, when running on a personal computer with an Intel® 2.33GHz CPU and 2G RAM. For the 25% censoring rate,

TABLE 1
Summary of simulation studies

N	Censoring		DSR				PLE		LSQ		Log-rank		Gehan	
			Bias	SE	SEE	CP	Bias	RE	Bias	RE	Bias	RE	Bias	RE
							Normal error							
100	25%	β_1	0.010	0.222	0.227	0.952	-0.001	0.962	0.011	1.051	0.014	0.916	0.011	0.975
		β_2	0.002	0.227	0.227	0.938	0.002	0.901	0.003	1.011	0.001	0.879	0.004	0.964
	50%	β_1	0.006	0.250	0.243	0.954	0.012	0.741	0.009	0.885	0.008	0.832	0.010	0.807
		β_2	-0.005	0.254	0.246	0.936	0.002	0.742	-0.007	0.915	-0.006	0.884	-0.008	0.807
200	25%	β_1	0.002	0.158	0.160	0.947	-0.003	0.955	0.002	1.015	0.004	0.959	0.002	0.937
		β_2	-0.005	0.155	0.160	0.952	0.004	0.939	-0.004	1.034	-0.003	0.903	-0.003	0.973
	50%	β_1	0.007	0.166	0.170	0.958	0.009	0.834	0.006	0.847	0.004	0.848	0.007	0.752
		β_2	-0.002	0.175	0.171	0.948	0.003	0.837	-0.005	0.858	-0.006	0.804	-0.006	0.798
							Logistic error							
100	25%	β_1	-0.0002	0.395	0.395	0.944	-0.015	0.875	0.005	0.963	0.002	0.866	0.001	0.991
		β_2	-0.002	0.394	0.398	0.943	-0.007	0.872	-0.003	0.949	0.011	0.872	-0.002	0.972
	50%	β_1	-0.016	0.430	0.423	0.944	-0.015	0.679	-0.012	0.839	-0.016	0.805	-0.014	0.838
		β_2	0.005	0.421	0.429	0.947	-0.012	0.667	-0.003	0.811	-0.002	0.780	-0.001	0.798
200	25%	β_1	-0.012	0.276	0.275	0.953	0.014	0.950	-0.015	0.941	-0.014	0.863	-0.009	1.006
		β_2	-0.004	0.281	0.276	0.938	-0.005	0.866	-0.007	0.961	-0.004	0.860	-0.004	1.009
	50%	β_1	0.005	0.298	0.297	0.953	0.006	0.768	0.003	0.839	0.008	0.774	0.005	0.842
		β_2	0.014	0.289	0.298	0.956	0.008	0.756	0.013	0.838	0.009	0.792	0.017	0.875
							Extreme value error							
100	25%	β_1	0.002	0.251	0.251	0.948	0.007	0.781	0.004	0.791	-0.003	1.232	0.001	0.892
		β_2	0.019	0.257	0.256	0.948	-0.011	0.783	0.023	0.742	0.015	1.232	0.017	0.886
	50%	β_1	0.003	0.274	0.275	0.950	-0.010	0.661	0.006	0.645	0.005	1.102	0.008	0.727
		β_2	-0.016	0.298	0.283	0.924	-0.008	0.631	-0.021	0.717	-0.015	1.218	-0.017	0.792
200	25%	β_1	0.010	0.172	0.173	0.949	0.014	0.916	-0.001	0.741	0.005	1.192	0.001	0.861
		β_2	-0.007	0.177	0.174	0.952	-0.003	0.912	0.010	0.716	0.004	1.158	0.006	0.860
	50%	β_1	0.001	0.186	0.186	0.951	0.001	0.722	0.004	0.645	-0.001	1.129	0.001	0.714
		β_2	-0.001	0.193	0.187	0.944	0.008	0.758	0.002	0.665	-0.002	1.091	0.001	0.722
							Mixture Normal error							
100	25%	β_1	0.0001	0.267	0.263	0.949	0.010	0.932	0.003	0.817	0.003	0.847	-0.002	1.018
		β_2	-0.007	0.263	0.266	0.949	0.001	0.953	-0.009	0.779	-0.007	0.831	-0.007	0.998
	50%	β_1	0.008	0.282	0.289	0.954	-0.009	0.769	0.012	0.744	0.007	0.776	0.009	0.920
		β_2	-0.005	0.306	0.289	0.952	-0.013	0.745	-0.007	0.760	-0.004	0.777	-0.006	0.884
200	25%	β_1	0.002	0.179	0.184	0.958	0.008	0.956	0.006	0.790	0.005	0.816	0.003	1.024
		β_2	0.004	0.185	0.185	0.954	0.008	0.991	0.003	0.793	-0.001	0.834	0.004	1.015
	50%	β_1	-0.007	0.212	0.201	0.944	0.012	0.816	-0.006	0.754	-0.011	0.814	-0.001	0.854
		β_2	0.005	0.209	0.203	0.958	0.005	0.821	0.003	0.760	0.003	0.789	0.003	0.859
							Log-Weibull error							
100	25%	β_1	0.005	0.484	0.500	0.951	0.011	0.814	0.005	0.730	0.009	1.225	0.008	0.862
		β_2	-0.036	0.511	0.505	0.940	-0.029	0.808	-0.032	0.735	-0.024	1.179	-0.032	0.866
	50%	β_1	-0.008	0.557	0.537	0.943	-0.013	0.621	-0.006	0.649	-0.001	1.064	-0.010	0.699
		β_2	-0.018	0.544	0.545	0.934	-0.008	0.614	-0.023	0.630	-0.017	1.029	-0.025	0.680
200	25%	β_1	0.011	0.356	0.345	0.943	0.012	0.942	0.010	0.737	0.014	1.244	0.009	0.854
		β_2	0.009	0.342	0.346	0.953	0.015	0.885	0.011	0.731	0.007	1.173	0.010	0.858
	50%	β_1	0.004	0.361	0.367	0.951	0.004	0.762	0.008	0.637	-0.002	1.059	0.007	0.701
		β_2	-0.006	0.372	0.369	0.950	-0.002	0.790	-0.018	0.622	-0.004	1.072	-0.016	0.685

SE: standard error of the parameter estimator; SEE: mean of the standard error estimator; CP: coverage probability of the 95% confidence interval; RE: ratio mean squared error of the proposed estimator to that of the other estimator

the computational speed of the newly proposed estimator was almost the same as the profile likelihood estimator. Both were much faster than the other two estimators based on resampling techniques. When the censoring rate was 50%, the DSR method was slower than the profile likelihood and the Gehan esti-

TABLE 2
Average computing time (Sec.)

N	Censoring	DSR	PLE	Log-rank	Gehan
100	25%	23.9	23.2	138.7	31.2
100	50%	53.7	22.9	119.1	27.1
200	25%	87.4	90.9	804.7	167.5
200	50%	199.6	92.1	605.5	126.4

mator. However, some of the other procedures except for the DSR converged even faster compared with their computing time for the 25% censoring rate when n was fixed. This is because in all the estimating functions of the other estimators, the censoring indicator is a multiplier, which means that the computational complexity for these estimators decreases along with the increased censoring rate. However, the DSR method can always manage to recover the censored information, which implies a more complex computational requirement when the censoring rate is higher. This perhaps, from a different perspective, explains why the performance of the DSR method for the high censoring rate was better than the others for most of the scenarios and even very close to those supposed to be asymptotically efficient for the underlying distributions.

6. Two examples

We first applied the proposed doubly smoothed method to the Mayo primary biliary cirrhosis (PBC) study described in Fleming and Harrington (1991), which has been widely investigated, such as in Jin et al. (2006a) and Zeng and Lin (2007). The data contain information about the survival time and prognostic factors for 418 patients. Jin et al. (2006a) fitted the accelerated failure time model with five covariates – age, log(albumin), log(bilirubin), edema, and log(protime) – using rank and least squares estimators. We fitted the same model using the proposed method, for which the Epanechnikov kernel was used in estimating conditional mean residuals, separately with the bandwidths $a_1 = (40\sqrt{\pi})^{1/5}\sigma_1 N_1^{-1/5}$ and $a_2 = \sigma_1 N_1^{-1/7}$ for h_1. The results are presented in Table 3, together with the estimates from other methods, cited directly from Zeng and Lin (2007) and Jin et al. (2003, 2006a), for ease of comparison.

TABLE 3
Analysis results of the Mayo PBC data

Parameter	DSR^{a_1} Est.	SE	DSR^{a_2} Est.	SE	PLE Est.	SE	LSQ Est.	SE	Log-rank Est.	SE
Age	-0.0167	0.0044	-0.0191	0.0047	-0.0286	0.0061	-0.0256	0.0063	-0.0265	0.0042
log(Albumin)	1.3847	0.5532	1.3968	0.6068	1.6212	0.4761	1.6174	0.5409	1.6558	0.3683
log(Bilirubin)	-0.4093	0.0642	-0.4472	0.0846	-0.6175	0.0669	-0.5885	0.0752	-0.5849	0.0455
Edema	-0.8506	0.2977	-0.8572	0.2643	-0.7985	0.3179	-0.8430	0.2604	-0.7338	0.1781
log(Protime)	-2.6064	1.1089	-2.2957	1.1664	-2.4095	0.8050	-2.3331	0.8543	-1.9439	0.4622

Est.: parameter estimate; SE: standard error of the parameter estimator

The parameter estimates and the variance estimates produced from the proposed method appear similar for the two different choices of bandwidths. The new estimating method led to numerically smaller absolute effects in age, log(Albumin) and log(Bilirubin) than those from the other methods. For Edema and log(Protime), our estimates are comparable with those from the Gehan and least squares regressions. As the censoring rate in this dataset is approximately as high as 61.5%, the new estimates could provide a useful alternative explanation for the model in view of the relatively good performance of the method for severe censoring, shown in the simulations.

The second example is the well-known Stanford heart transplant data analyzed by Miller and Halpern (1982) using the Cox and Buckley-James regression methods. Survival time for a total of 184 heart-transplanted patients was collected with their ages at the time of the first transplant and T5 mismatch scores. Following Miller and Halpern (1982), the 27 patients who did not have T5 mismatch scores and the 5 patients with survival time of less than 10 days were excluded from our analysis. Out of the remaining 152 patients, 55 were censored as of February 1980. The dataset is available and named "stanford2" in the "survival" package in R. A quadratic model of base 10 logarithm of survival time with regard to age was also considered here as reported in Miller and Halpern (1982), rather than a linear model for a better fit.

The AFT model is usually used when the proportional hazards assumption required by the Cox model is questionable. In this regard, many existing methods for assessing the adequacy of the proportional hazards assumption can be first used here. Our perspective is that even though the Cox model is valid, the AFT model can be a reliable alternative by providing a different interpretation to the covariate effect. In light of this, instead of checking the proportional odds assumption, we explored graphically the appropriateness of the log-linear specification assumed in the AFT model. Kaplan-Meier estimates of the log survival time conditional on age at quantiles $\tau = 0.01, 0.25, 0.5$, and 0.75 were plotted against age. Larger quantiles were not included because there are relatively few ages at which these quantiles are available around observed data points. Meanwhile, at each age, the convention to redefine the largest log survival time as uncensored whenever it is not was adopted. Fitted quadratic models of these quantiles with regard to age using least squares estimate were also plotted. Figure 1 demonstrates that the quadratic age model assumed fits the data well, and the patterns on age effect at different quantiles appear similar to each other, indicating that a homogeneous AFT model is adequate for the data.

To better interpret the age effect to avoid the possible multicollinearity between age and age^2, we further centralized age at its mean value 42, following Wei et al. (1990). The parameter estimates for centralized age and age^2 using the proposed DSR method and a convergence criteria of 10^{-6} are -0.033 and -0.0014 with the corresponding 95% confidence intervals $(-0.051, -0.016)$ and $(-0.0028, -0.00014)$, respectively. Compared with the results from the Gehan $(-0.036$ and $-0.0017)$ and log-rank $(-0.038$ and $-0.0016)$ methods reported in Wei et al. (1990), the DSR estimators demonstrate also negative linear and quadratic, but both age effects are smaller in magnitude.
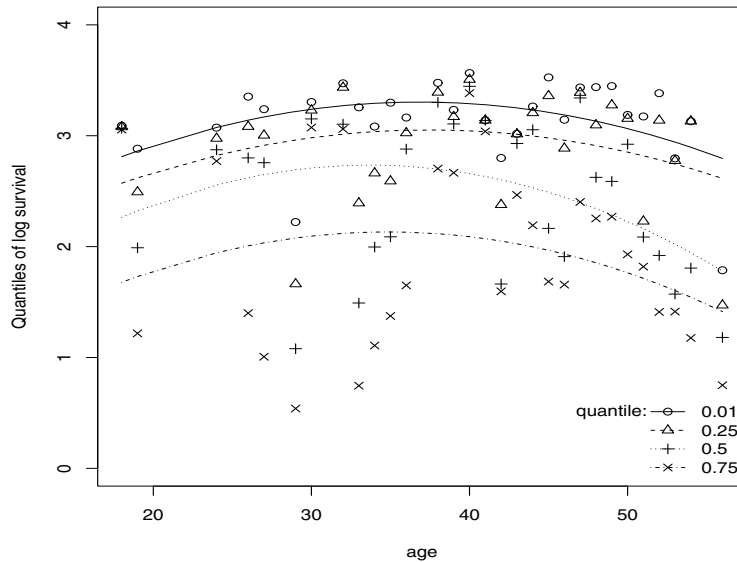
FIG 1. *Estimated quantiles of log survival time and fitted curves from least squares estimates of age quadratic models.*

7. Discussion

In this article, we have proposed an induced smoothing rank regression method for censored data, where censored observations are imputed by using kernel smoothed estimates for conditional expectations of lifetimes. The original purpose is to recover the efficiency of rank procedure with censored data following the logistic distribution, as in the common Wilcoxon type procedures. The simulation results also showed that the performance of the proposed estimator generally outperformed the other existing estimators, in particular, for severe censoring. The kernel estimation used is based on the convolution formulation for the hazard function in Ramlau-Hansen (1983). Other smoothed estimators, such as those proposed in Chaubey and Sen (2008), could also be adopted here. We could even consider smoothing the mean residual lifetime directly as in Abdous and Berred (2005), among others. However, the kernel estimator we proposed is convenient for demonstrating smoothness in the regression coefficients, and for showing asymptotic properties for the final regression estimators.

The proposed method could also be extended to clustered data settings. To this end, we can consider induced smoothing the weighted rank regression suggested by Wang and Zhao (2008) with censored observations being imputed by certain smoothed estimates of conditional expectations of lifetimes, adjusted for correlations. Another potential application can be developing a smoothed estimating equation in a quantile regression (Wang et al. (2009)), when both smoothing and imputing are needed for censored data. This is of particular interest in presence of heteroscedasiticity.

Appendix

1. Proof of Theorem 1

Denote by $\mathbb{E}$ the support of F. We have

$$|\tilde{\alpha}(e) - \hat{\alpha}(e)| \leq n^{-1} h_1^{-1} \sum_j \left| \frac{k(\frac{e-E_j}{h_1})}{1 - \hat{F}_n(E_j)} - \frac{k(\frac{e-E_j}{h_1})}{1 - F_n(E_j)} \right|.$$

Suppose that the distribution function of $\tilde{e}_i = \min(e_i, c_i)$, $i = 1 \ldots n$, is continuous. From Theorem 1 in Giné and Nickl (2009) and the conditions assumed on h_1 and h_2, it can be shown that $\sup_{E \in \mathbb{E}} |\hat{F}_n(E) - F_n(E)| = O_{a.s.}(\max(\sqrt{h_2 \log(1/h_2)}, \sqrt{n} h_2^q)) = o_{a.s.}(h_1)$. This means also that $\sup_{E \in \mathbb{E}} |(1 - \hat{F}_n(E))^{-1} - (1 - F_n(E))^{-1}| = o_{a.s.}(h_1)$ from the continuous mapping theorem. It follows that $|\tilde{\alpha}(e) - \hat{\alpha}(e)| \xrightarrow{P} 0$ by noting that k is a bounded kernel.

Observing that the counting process arguments are still applicable here in term of the residuals even though arbitrarily large negative values may be observed, we further obtain that $\hat{\alpha}(e) \xrightarrow{P} \alpha(e)$ as $n \to \infty$, under those conditions required for showing consistency of $\hat{\alpha}(e)$ in Ramlau-Hansen (1983). It implies that $\tilde{\alpha}(e) \xrightarrow{P} \alpha(e)$. By dominated convergence again, we have $|\tilde{A}(e) - A(e)| \xrightarrow{P} 0$ as $n \to 0$. The consistency of $\tilde{S}(e)$ follows immediately. Following the derivation of Lemma 1 in James and Smith (1984), now we can show that $\sup_{e \in \mathbb{E}} |\tilde{m}(e; \beta) - m(e; \beta)| \xrightarrow{P} 0$.

2. Proof of Lemma 1

Assume that z is a random vector, following a standard multivariate normal distribution $N(0, I_p)$, where I_p is a $p \times p$ identity matrix, and Γ is a covariance matrix of order $O(n^{-1})$. From the definition of $\tilde{U}(\beta)$, we can write $\tilde{U}(\beta_0) = E_z\{\hat{U}_1(z; \beta_0)\}$ and $U(\beta_0) = \hat{U}_1(0; \beta_0)$, where

$$\hat{U}_1(z; \beta_0) = -\frac{1}{n(n-1)} \sum_{i<j} \text{sgn}\{\delta_i \tilde{e}_i + (1 - \delta_i)\tilde{m}(\tilde{e}_i) - \delta_j \tilde{e}_j - (1 - \delta_j)\tilde{m}(\tilde{e}_j) - d_{ij}^T \Gamma^{1/2} z\} d_{ij},$$

and further

$$\tilde{U}(\beta_0) - U(\beta_0) = \int_{\mathbb{R}^p} \{\hat{U}_1(z; \beta_0) - \hat{U}_1(0; \beta_0)\} d\mathbf{\Phi}(\mathbf{z}). \tag{A1}$$

We first prove that $\sqrt{n} ||\hat{U}_1(z; \beta_0) - \hat{U}_2(z; \beta_0)|| \to 0$ in probability, where

$$\hat{U}_2(z; \beta_0) = -\frac{1}{n(n-1)} \sum_{i<j} \text{sgn}\{\delta_i \tilde{e}_i + (1 - \delta_i)m(\tilde{e}_i) - \delta_j \tilde{e}_j - (1 - \delta_j)m(\tilde{e}_j) - d_{ij}^T \Gamma^{1/2} z\} d_{ij}.$$

This can be proceeded by showing that any component k, $k = 1 \ldots p$, of $\sqrt{n}\{\hat{U}_1(z; \beta_0) - \hat{U}_2(z; \beta_0)\}$, in term of $\tilde{e}$, is convergent to zero in probability. Due to the assumption that $\tilde{e}_i$, $i = 1 \ldots n$, are independent and identically distributed and uniformly censored, it is reasonable to assume that $\tilde{E}_{ij} = \delta_i \tilde{e}_i + (1 - \delta_i)m(\tilde{e}_i) - \delta_j \tilde{e}_j - (1 - \delta_j)m(\tilde{e}_j)$ for all $i \neq j$ follows a same continuous distribution G with $G(0) = 1/2$. Denote the pdf of G by g and $\tilde{m}(\tilde{e}_i) - m(\tilde{e}_i)$ by M_i for $i = 1 \ldots n$. After some simplifications, we obtain

$$E\|\sqrt{n}(\hat{U}_1(z; \beta_0) - \hat{U}_2(z; \beta_0))\|^2$$

$$\leq \frac{1}{n(n-1)^2} E\left[\sum_{i \neq j} \|d_{ij}\| \left[I\{\tilde{E}_{ij} > d_{ij}^T \Gamma^{1/2} z - (1 - \delta_i)M_i + (1 - \delta_j)M_j\} \right.\right.$$

$$\left.\left. - I\{\tilde{E}_{ij} > d_{ij}^T \Gamma^{1/2} z\} \right] \right]^2$$

$$\leq \frac{1}{n(n-1)} E \sum_{i \neq j} \|d_{ij}\|^2 |I\{\tilde{E}_{ij} > d_{ij}^T \Gamma^{1/2} z - (1 - \delta_i)M_i + (1 - \delta_j)M_j\}$$

$$- I\{\tilde{E}_{ij} > d_{ij}^T \Gamma^{1/2} z\}|$$

$$= \frac{1}{n(n-1)} E \sum_{i \neq j} \|d_{ij}\|^2 g(\zeta_{ij})\{(1 - \delta_i)M_i - (1 - \delta_j)M_j\}$$

where, by the mean value theorem, ζ_{ij} is between $d_{ij}^T \Gamma^{1/2} z$ and $d_{ij}^T \Gamma^{1/2} z - (1 - \delta_i)M_i + (1 - \delta_j)M_j$. The last inequality holds because the expectation of each of those individual items who have different subscripts $i \neq i' \neq j \neq j'$ equals 0. From Theorem 1, the right-hand side of the above inequality tends to zero. We therefore have $E\|\sqrt{n}(\hat{U}_1(z; \beta_0) - \hat{U}_2(z; \beta_0))\|^2 \to 0$ as $n \to 0$. It indicates that $\sqrt{n}\|\hat{U}_1(z; \beta_0) - \hat{U}_2(z; \beta_0)\| = o_p(1)$ by the Chebyshev inequality.

The asymptotic equivalence of $\hat{U}_1(z; \beta_0)$ and $\hat{U}_2(z; \beta_0)$ and (A1) imply that

$$\sqrt{n}\|\tilde{U}(\beta_0) - U(\beta_0)\| \leq \int_{\mathbb{R}^p} \sqrt{n}\|\hat{U}_2(z; \beta_0) - \hat{U}_2(0; \beta_0)\| d\mathbf{\Phi}(\mathbf{z}) + o_p(1) \quad \text{(A2)}$$

If We can show that $\sqrt{n}\|\hat{U}_2(z; \beta_0) - \hat{U}_2(0; \beta_0)\|$ converges in probability to 0 by showing its componentwise convergence, then we complete the proof of Lemma 1. Denote the kth component of $\hat{U}_2(z; \beta_0) - \hat{U}_2(0; \beta_0)$ by $\hat{U}_{2k}(z) - \hat{U}_{2k}(0)$. Since $E\{\hat{U}_2(0; \beta_0)\} = 0$, we have

$$\sqrt{n}E\{\hat{U}_{2k}(z) - \hat{U}_{2k}(0)\} = \sqrt{n}E[E\{\hat{U}_{2k}(z)|z\}]$$

$$= \frac{1}{\sqrt{n}(n-1)} E \sum_{i \neq j} d_{ijk} \left[G(d_{ij}^T \Gamma^{1/2} z) - G(0) \right]$$

$$= \frac{\Gamma^{1/2}}{\sqrt{n}(n-1)} E \sum_{i \neq j} d_{ijk} g(\tau_{ij}) d_{ij}^T z \quad \text{(A3)}$$

where, by the mean value theorem, τ_{ij} is between 0 and $|d_{ij}^T \Gamma^{1/2} z|$. As Γ is matrix of order $O(n^{-1})$ and z is of standard normal, $\|\Gamma^{1/2} z\|$ tends to zero in probability

and $g(\tau_{ij}) \to g(0)$ in probability for all $i \neq j$. Hence, under regularity conditions on g, it follows that (A3) tends to zero. Meanwhile, simplification leads to

$$n\mathrm{var}\{\hat{U}_{2k}(z) - \hat{U}_{2k}(0)\} \leq \frac{1}{n(n-1)} E \sum_{i \neq j}^{n} d_{ijk}^2 |I\{\tilde{E}_{ij} > d_{ij}^T \Gamma^{1/2} z\} - I\{\tilde{E}_{ij} > 0\}|$$

Following again the line in (A3), $n\mathrm{var}\{\hat{U}_{2k}(z) - \hat{U}_{2k}(0)\}$ tends to zero as n tends to infinity. This shows

$$\sqrt{n}||\hat{U}_2(z; \beta_0) - \hat{U}_2(0; \beta_0)|| = o_p(1)$$

and hence the right-hand side in (A2) converges in probability to zero, concluding the proof.

3. Proof of the consistency and asymptotic normality of $\hat{\beta}_{DSR}$ and $\hat{\beta}_{(m)}$

To show first the consistency and asymptotic normality of $\hat{\beta}_{DSR}$, we note that routine application of the projection method gives that $\sqrt{n}\hat{U}_2(0; \beta_0)$ converges in distribution to a normal distribution $N(0, n\Sigma)$ as $n \to \infty$, where

$$\Sigma = \frac{4}{n^2(n-1)^2} \sum_{i=1}^{n} E(\xi_i \xi_i^T) \tag{A4}$$

and

$$\xi_i = \sum_{j \neq i}^{n} \left[H\{\delta_j \tilde{e}_j + (1 - \delta_j)m(\tilde{e}_j)\} - \frac{1}{2} \right] d_{ij},$$

where H is the distribution function of $\delta_i \tilde{e}_i + (1 - \delta_i)m(\tilde{e}_i)$, $i = 1 \ldots n$. This implies that $||\hat{U}_2(0; \beta_0)|| = o_p(1)$. Since $\sqrt{n}||\tilde{U}(\beta_0) - \hat{U}_2(0; \beta_0)|| \leq \sqrt{n}||\tilde{U}(\beta_0) - U(\beta_0)|| + \sqrt{n}||U(\beta_0) - \hat{U}_2(0; \beta_0)|| = o_p(1)$ from Lemma 1, $||\tilde{U}(\beta_0)|| = o_p(1)$ by the triangle inequality.

Let $h = \hat{\beta}_{DSR} - \beta_0$ and denote $\hat{\beta}_{DSR}$ by $\beta_0 + h$. Due to the continuity of $\tilde{U}(\beta)$ in β, generalizations of the mean value theorem give that almost surely

$$||\{\tilde{U}(\hat{\beta}_{DSR}) - \tilde{U}(\beta_0)\} - D_h(\hat{\beta}_{DSR} - \beta_0)|| = 0, \tag{A5}$$

where $D_h = \int_0^1 J\tilde{U}(\beta_0 + th)dt$, $0 \leq t \leq 1$, and $J\tilde{U}(\cdot)$ is the Jacobian matrix of $\tilde{U}(\cdot)$. The triangle inequality and (A5) further give that

$$\begin{aligned} & || - D_h(\hat{\beta}_{DSR} - \beta_0)|| \\ \leq & ||\{\tilde{U}(\beta_0) - \tilde{U}(\hat{\beta}_{DSR})\}|| + ||\{\tilde{U}(\hat{\beta}_{DSR}) - \tilde{U}(\beta_0)\} - D_h(\hat{\beta}_{DSR} - \beta_0)|| \\ = & o_p(1) \end{aligned} \tag{A6}$$

By assuming the Jacobian matrix $J\tilde{U}(\cdot)$ is non-singular within $0 \leq t \leq 1$, it shows that $\hat{\beta}_{DSR}$ is a consistent estimator of β_0. Note that this consistency result states nothing about the uniqueness of $\hat{\beta}_{DSR}$.

Application of the Taylor expansion to the continuous estimating function $\tilde{U}(b)$ at $||b - \beta_0|| \leq n^{-1/2}$ gives

$$\sup_{||b - \beta_0|| \leq n^{-1/2}} \frac{||\sqrt{n}\{\tilde{U}(b) - \tilde{U}(\beta_0)\} - \sqrt{n}D^T(b - \beta_0)||}{1 + \sqrt{n}||b - \beta_0||} = o_p(1).$$

From $\sqrt{n}||\tilde{U}(\beta_0) - \hat{U}_2(0; \beta_0)|| = o_p(1)$, it leads to

$$\sqrt{n}\{\tilde{U}(b) - \hat{U}_2(0; \beta_0)\} = \sqrt{n}D(b - \beta_0) + o_p(1 + \sqrt{n}||b - \beta_0||). \tag{A7}$$

The consistency of $\hat{\beta}_{DSR}$ and (A7) give that

$$\sqrt{n}(\hat{\beta}_{DSR} - \beta_0) = -\sqrt{n}D^{-1}\hat{U}_2(0; \beta_0) + o_p(1). \tag{A8}$$

The asymptotic normality of $\hat{\beta}_{DSR}$ immediately follows from the asymptotic normality of $\sqrt{n}\hat{U}_2(0; \beta_0)$ as shown above.

Next, we show that $\hat{\beta}_{(m)}$ is also a consistent estimator of β_0 and asymptotically normally distributed. Denote the estimating function $\tilde{U}(\beta)$ in the mth iteration by $\tilde{U}\{\hat{\beta}_{(m)}, \hat{\beta}_{(m-1)}\}$, where $\hat{\beta}_{(m-1)}$ comes from the estimates of the conditional mean residuals in the $(m-1)$th iteration. Suppose that $\hat{\beta}_{(m-1)}$ is a consistent estimator of β_0 when starting iterations from a consistent estimator $\hat{\beta}_{(0)}$. Because of the continuity of $\tilde{U}(\beta)$ everywhere in β, the Taylor expansion of $\tilde{U}\{\hat{\beta}, \hat{\beta}_{(m-1)}\}$ regarding $\hat{\beta}_{(m-1)}$ in $||\hat{\beta}_{(m-1)} - \beta_0|| \leq n^{-1/2}$ gives

$$\tilde{U}\{\hat{\beta}, \hat{\beta}_{(m-1)}\} = \tilde{U}\{\hat{\beta}, \beta_0\} - (B - D)(\hat{\beta}_{(m-1)} - \beta_0) + o_p(1). \tag{A9}$$

Since $||\tilde{U}(\beta_0)|| = o_p(1)$, it is evident that $||\tilde{U}\{\beta_0, \hat{\beta}_{(m-1)}\}|| = o_p(1)$. Following the line in showing the consistency of $\hat{\beta}_{DSR}$, we can obtain that the root of the estimating function $\tilde{U}\{\hat{\beta}, \hat{\beta}_{(m-1)}\}$, i.e. $\hat{\beta}_{(m)}$, is a consistent estimator of β_0. Furthermore, given $\hat{\beta}_{(m-1)}$, it can be shown that $\hat{\beta}_{(m)}$ is a unique root of $\tilde{U}\{\hat{\beta}, \hat{\beta}_{(m-1)}\}$ by regarding it as a smoothed rank estimating function for complete data and following the line similar to Lemma 1 and Lemma 2 in Johnson and Strawderman (2009).

From the consistency of $\hat{\beta}_{(m)}$ and Taylor expansion, we have

$$\hat{\beta}_{(m)} - \beta_0 = (I - B^{-1}D)\{\hat{\beta}_{(m-1)} - \beta_0\} - B^{-1}\tilde{U}(\beta_0) + o_p(n^{-1/2})$$

By induction, we have for a fixed $m \geqslant 1$

$$\begin{aligned}
\hat{\beta}_{(m)} - \beta_0 &= (I - B^{-1}D)^m(\hat{\beta}_G - \beta_0) \\
&\quad - (I - (I - B^{-1}D)^m)D^{-1}\tilde{U}(\beta_0) + o_p(n^{-1/2})
\end{aligned}$$

Since $\sqrt{n}\tilde{U}(\beta_0)$ converges in distribution to $N(0, n\Sigma)$, as shown in (A4), as $n \to \infty$ and $\sqrt{n}(\hat{\beta}_G - \beta_0)$ is also asymptotically normal as shown in Jin et al. (2006b), it follows that $\hat{\beta}_{(m)}$ is asymptotically normally distributed as $n \to \infty$.

The asymptotic covariance matrix of $\hat{\beta}_{(m)}$ is determined by the variations of both $\hat{\beta}_G$ and $\tilde{U}(\beta_0)$. However, it approximately equals $\Gamma = D^{-1}\text{cov}\{\tilde{U}(\beta_0)\}\{D^T\}^{-1}$ when m is sufficiently large.

This can be seen by first showing that $B - D$ is nonnegative under the condition that the hazard $\alpha(e)$ is nondecreasing in e. We find that for any $i = 1 \ldots n$, $V_i = (1 - \delta_i)\{m(e_i; \beta) - e_i\} = (1 - \delta_i)\int_{e_i}^{\infty} S(t)dt/S(e_i)$ is nonincreasing in e_i, shown as follows,

$$\frac{\partial V_i}{\partial e_i} = (1 - \delta_i)\frac{\alpha(e_i)\int_{e_i}^{\infty} S(t)dt - S(e_i)}{S(e_i)} = (1 - \delta_i)\frac{\int_{e_i}^{\infty}\{\alpha(e_i) - \alpha(t)\}S(t)dt}{S(e_i)} \leq 0$$

It follows that for any vector $\gamma_{p\times 1}$ and $v_i = \partial V_i/\partial\beta$,

$$v_i^T\gamma = \frac{\partial V_i}{\partial e_i}\left\{\frac{\partial e_i}{\partial\beta}\right\}^T\gamma = -\frac{\partial V}{\partial e}x_i^T\gamma$$

is nondecreasing in $\gamma^T x_i$. It implies that $B - D$ is nonnegative and therefore $(I - B^{-1}D)^m$ approaches a zero matrix as m is sufficiently large.

Acknowledgements

References

ALTMAN, N. AND LEGER, C. (1995). Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference* **46** 195–214. MR1354087

ANDERSEN, P. K., BORGAN, O., GILL, R. D. AND KEIDING, N. (1992). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York. MR1198884

ABDOUS, B. AND BERRED, A. (1998). Mean residual life estimation. *Journal of Statistical Planning and Inference* **132** 3–19. MR2163677

AZZALINI, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika* **68** 326–328. MR0614972

BROWN, B. M. AND WANG, Y.-G. (2005). Standard errors and covariance matrices for smoothed rank estimators. *Biometrika* **92** 149–158. MR2158616

BROWN, B. M. AND WANG, Y.-G. (2006). Induced smoothing for rank regression with censored survival times. *Statistics in Medicine* **26** 828–836. MR2339177

BUCKLEY, J. AND JAMES, I. (1979). Linear regression with censored data. *Biometrika* **66** 429–436.

CHAUBEY, Y. P. AND SEN, A. (2008). Smooth estimation of mean residual life under random censoring. In *Beyond Parametrics in Interdisciplinary Re-*

search: Festschrift in Honor of Professor Pranab K. Sen. Beachwood, OH: Inst. Math. Statist. **1**, *Inst. Math. Stat. Collect.* 35–49. MR2459252

CHIOU, S. H., KANG, S. AND YAN, J. (2013). Fast accelerated failure time modeling for case-cohort data. *Statistics and Computing* DOI 10.1007/s11222-013-9388-2. MR3223541

FLEMING, T. R. AND HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York. MR1100924

GINÉ, E. AND NICKL, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probability Theory and Related Fields* **143** 569–596. MR2475673

HELLER, G. (2007). Smoothed rank regression with censored data. *Journal of the American Statistical Association* **102** 552–559. MR2370852

JAMES, I. R. AND SMITH, P. J. (1984). Consistency results for linear regression with censored data. *Annals of Statistics* **12** 590–600. MR0740913

JIN, Z., LIN, D. Y., WEI, L. J. AND YING, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90** 341–353. MR1986651

JIN, Z., LIN, D. Y., AND YING, Z. (2006a). On least-squares regression with censored data. *Biometrika* **93** 147–161. MR2277747

JIN, Z., LIN, D. Y. AND YING, Z. (2006b). Rank regression analysis of multivariate failure time data based on marginal linear models. *Scandinavian Journal of Statistics* **33** 1–23. MR2255106

JOHNSON, L. M. AND STRAWDERMAN, R. L. (2009). Induced smoothing for the semiparametric accelerated failure time model: Asymptotics and extensions to clustered data. *Biometrika* **96** 577–590. MR2538758

LAI, T. L. AND YING, Z. (1991). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Annals of Statistics* **10** 1370–1402. MR1126329

MILLER, R. G. AND HALPERN, J. (1982). Regression with censored data. *Biometrika* **69** 521–531. MR0695199

PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65** 167–179. MR0497517

RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Annals of Statistics* **11** 453–466. MR0696058

RITOV, Y. (1990). Estimation in a linear regression model with censored data. *Annals of Statistics* **18** 303–328. MR1041395

SILVERMAN, B. W. (1986). *Density Estimation.* Chapman and Hall, London. MR0848134

STARE, J., HEINZL, H. AND HARRELL, F. (2000). On the use of Buckley and James least squares regression for survival data. In *New Approaches in Applied Statistics.* Ferligoj, A. and Mrvar, A. (Editors), Metodološki zvezki 16, Ljubljana: FDV, 125–134.

TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Annals of Statistics* **18** 354–372. MR1041397

WANG, Y.-G., SHAO, Q. AND ZHU, M. (2009). Quantile regression without the curse of unsmoothness. *Computational Statistics and data Analysis* **53** 3696–3705. MR2751958

Wang, Y.-G. and Zhao, Y. (2008). Weighted rank regression for clustered data analysis. *Biometrics* **64** 39–45. MR2422817

Wei, L. J., Ying, Z. and Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77** 845–851. MR1086694

Wei, L. J. (1992). The accelerated failure time model: A useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine* **11** 1871–1880.

Zeng, D. and Lin, D. Y. (2007). Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association* **102** 1387–1396. MR2412556