The Buckley–James estimator and induced smoothing

Wang, You-Gan, Zhao, Yudong and Fu, Liya

 doi: 10.1111/j.1467-842X.XXX

THE BUCKLEY-JAMES ESTIMATOR AND INDUCED SMOOTHING

YOU-GAN WANG[1]*, YUDONG ZHAO[2] AND LIYA FU[3]

[1]*Queensland University of Technology, Australia.*
[2]*Danone Nutricia Research, Singapore.*
[3]*Xi'an Jiaotong University, China.*

Summary

The Buckley-James (BJ) estimator is known to be consistent and efficient for a linear regression model with censored data. However, its application in practice is handicapped by the lack of a reliable numerical algorithm for finding the solution. For a given data set, the iterative approach may yield multiple solutions or no solution at all. To alleviate this problem, we modify the induced smoothing approach originally proposed by Brown & Wang (2005, *Biometrika*). The resulting estimating functions become smooth, thus eliminating the tendency of the iterative procedure to oscillate between different parameter values. In addition to facilitating point estimation the smoothing approach enables easy evaluation of the projection matrix, thus providing a means of calculating standard errors. Extensive simulation studies were carried out to evaluate the performance of different estimators. In general, smoothing greatly alleviates numerical issues that arise in the estimation process. In particular, the one-step smoothing estimator eliminates non-convergence problems and performs similarly to full iteration until convergence. The proposed estimation procedure is illustrated using a dataset from a multiple-myeloma study.

Key words: censored data; covariance estimates; induced smoothing; rank regression; standard errors.

1. Introduction

The accelerated failure time model, as an alternative to the proportional hazards model of Cox (1972), has attracted considerable attention in recent years because its linear formulation makes it the model of choice for biomedical and environmental studies. Recent important contributions to the literature on the accelerated failure time model include work by Li & Yin (2009) and Fu & Wang (2011). A distinguishing feature of survival data is that they are

* Author to whom correspondence should be addressed.
[1]School of Mathematical Sciences, Queensland University of Technology, Australia.
Email: you-gan.wang@qut.edu.au
[2]Danone Nutricia Research, Singapore.
Email: yudong.zhao@gmail.com
[3]School of Mathematics and Statistics, Xi'an Jiaotong University, China.
Acknowledgment. We wish to thank the Associate Editor and two anonymous referees for their constructive comments, which have led to a much improved paper. This research is partially funded by Australian Research Council Discovery Project DP130100766, DP160104292, the National Science Foundation of China (No.11201365 and No.11301408), and the Ph.D. Programs Foundation of the Ministry of Education of China (2012020112005).

often subject to censoring, which introduces extra complexity in modeling and estimation. The statistical literature on this topic is abundant.

We now describe the basic structure of accelerated failure time models. Assume that T_i $(1 \leq i \leq n)$ is the failure time for subject i and that $\boldsymbol{x_i}$ is the corresponding covariate vector of dimension p. Under the usual accelerated failure time model

$$y_i = \log(T_i) = \alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta} + \epsilon_i,$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients, and ϵ_i is the error term with mean 0. Usually T_i is subject to censoring at C_i, a random variable independent of T_i. We assume that, conditional on $\{\boldsymbol{x}_i\}$, $\{C_i\}$ and $\{\epsilon_i\}$ are independent. Note that the conditional distribution of $\{\epsilon_i\}$ may depend on $\{\boldsymbol{x}_i\}$. We will regard the covariates $\{\boldsymbol{x}_i\}$ as realizations of a multivariate random variable.

The observed data may be written as the triplet (T_i^*, δ_i, x_i), where $T_i^* = \min(T_i, C_i)$, and δ_i is the value of an indicator function, $I(T_i \leq C_i)$. Write $y_i^* = \log(T_i^*)$ and $c_i = \log(C_i)$. Then $e_i(\beta) = y_i^* - \boldsymbol{x}_i^\top \boldsymbol{\beta}$ defines an error term.

In order to carry out the least-squares estimation, one will need to impute the censored observations in calculating the residuals. Buckley & James (1979) proposed a semiparametric iterative algorithm that alternates between imputation of censored failure times and least-squares estimation. However, the Buckley-James (BJ) estimator is a root of a discontinuous estimating function which may have multiple roots (Jin, Lin, Wei et al. 2006). Jin et al. used a linear approximation to obtain the solution of the BJ estimating equations and a new resampling approach to calculate the standard errors of the resulting estimates. Their approach using a linear approximation is elegant, but it does not completely solve the non-convergence problem due to the lack of smoothness in the Kaplan-Meier (KM) estimator used in the iteration. We will provide more details in the next section.

For convenience, when there is no confusion, we will suppress the dependence of e_i on $\boldsymbol{\beta}$. Recent research has been focusing on two main issues in respect of estimating $\boldsymbol{\beta}$: (i) obtaining unique estimates of $\boldsymbol{\beta}$, which includes addressing the issues of convergence and choice of initial values, and (ii) estimation of the covariance matrix of $\hat{\boldsymbol{\beta}}$.

In this paper we will follow the methodology proposed by Brown & Wang (2005), who developed an induced smoothing approach for easy calculation of standard errors. Johnson & Strawderman (2009) and Wang & Fu (2011) further investigated this approach for analysis of clustered failure time data. Both asymptotic and simulation results demonstrate that these smoothed estimates perform well. However, this smoothing idea is not directly applicable to the BJ estimator. Our aim is to apply a generalised approach of induced smoothing so that the Kaplan-Meier function becomes smooth, which leads to smooth projection matrices in the iterative approach and thus alleviates the problem that the iterations may oscillate.

The induced smoothing approach results in smooth estimating functions for which the derivatives (i.e., the projection matrix) and solutions can be easily evaluated numerically. Another advantage is that the standard errors can be obtained simultaneously because they are updated in the course of the iteration, consequently reducing the computational burden relative to the resampling approach.

We briefly introduce the BJ estimator in Section 2, and apply the modified induced smoothing procedure in Section 3. We compare the performance of various estimators of $\boldsymbol{\beta}$ in Section 4 via extensive simulation studies, and provide a brief illustration using the myeloma study in Section 5.

2. The Buckley-James Estimator

Suppose that $F(t)$ is the cumulative distribution function of ϵ_i. Buckley & James (1979) proposed the imputation of the censored observations by means of the equation

$\hat{y}_i = \delta_i y_i + (1-\delta_i)\hat{y}_{iC}$, where

$$\hat{y}_{iC} = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \frac{\int_{e_i}^\infty u dF(u)}{1 - F(e_i)}.$$

Since $F(u)$ is unknown, Buckley & James (1979) suggested replacing it by its Kaplan-Meier estimator which is given by

$$\hat{F}_\beta(t) = 1 - \prod_{i: e_i < t} \left\{ 1 - \frac{\delta_i}{\sum_{j=1}^n I(e_j \geq e_i)} \right\}.$$

Denote the mean of $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ by $\bar{\boldsymbol{x}}$. After the censored observations have been imputed, $\boldsymbol{\beta}$ may be estimated via the usual least-squares approach by minimising the objective function

$$n^{-1} \sum_{i=1}^n (\hat{y}_i - \alpha - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2,$$

The corresponding 'score' functions for $\boldsymbol{\beta}$ (after profiling out α) are

$$U(\beta) = n^{-1} \sum_{i=1}^n (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\hat{y}_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}). \tag{1}$$

The function $U(\boldsymbol{\beta})$ is nonlinear in $\boldsymbol{\beta}$ because $\hat{y}_i$ depends on $\boldsymbol{\beta}$ via $F(e_i)$. The estimator can be obtained by an iterative algorithm based on the equation

$$\hat{\boldsymbol{\beta}} = g(\hat{\boldsymbol{\beta}}) = \mathcal{X}^{-1} \sum_{i=1}^n (\boldsymbol{x}_i - \bar{\boldsymbol{x}})\hat{y}_i,$$

in which $\mathcal{X}$ is the square matrix $\sum_{i=1}^n \{(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top\}$, and $\hat{y}_i$ is evaluated at the previous $\hat{\boldsymbol{\beta}}$ values. However, there are two major problems here: convergence and the choice of initial values for $\hat{\boldsymbol{\beta}}$. There is no guarantee that this iteration converges, and the iteration may lapse into a state in which it oscillates between two or more values. Furthermore, it is not clear how the initial values will affect the final estimate, especially when the algorithm does not converge and one may have to choose a "final" estimate from the oscillating values. This iterative approach and the non-smooth nature of the problem make statistical inference on $\hat{\boldsymbol{\beta}}$ difficult. Estimation of the covariance matrix is also problematical. We aim to modify the estimating functions and their derivatives (the latter being known as the projection matrix) so that they become smooth in $\boldsymbol{\beta}$ which facilitates calculating the standard errors. The estimates of $\boldsymbol{\beta}$ will also become easier to obtain after smoothing.

3. The Modified Buckley-James Estimator with Smoothing

Suppose that $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically normal, $\mathrm{N}(0, \boldsymbol{\Gamma})$. We can then express $\hat{\boldsymbol{\beta}}$ as, $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{\Gamma}/n)^{1/2}\boldsymbol{Z}$, where the random vector $\boldsymbol{Z} \sim \mathrm{N}(0, \boldsymbol{I}_p)$ and $\boldsymbol{I}_p$ is the identity matrix of dimension p. This perturbation of $\hat{\boldsymbol{\beta}}$ results in a smoothed version of the function $U(\boldsymbol{\beta})$ given by (1). The smoothed version is equal to $\bar{U}(\boldsymbol{\beta}) = \mathrm{E}_Z(U(\boldsymbol{\beta} + h\boldsymbol{\Gamma}^{1/2}\boldsymbol{Z}))$, where the expectation is over $\boldsymbol{Z}$ and $h = 1/\sqrt{n}$. The idea of induced smoothing is to obtain $\hat{\boldsymbol{\beta}}$ by solving $\bar{U}(\boldsymbol{\beta}) = 0$ for a given $\boldsymbol{\Gamma}$. It can be seen that $\bar{U}(\boldsymbol{\beta}) - U(\boldsymbol{\beta})$ is in general close to 0 for any fixed $\boldsymbol{\Gamma}$, so the smoothed version provides almost the same numerical solution. Brown & Wang (2005) proposed that the asymptotic covariance of $\hat{\boldsymbol{\beta}}$ be used for $\boldsymbol{\Gamma}$ so their approach is referred to as naturally induced smoothing. The identity matrix $\boldsymbol{I}_p$ is found to work well as an initial value of the matrix of $\boldsymbol{\Gamma}$ (Wang & Fu, 2011). Unlike the situation in resampling approaches, we do not rely on the randomness of $\bar{U}$ to obtain the var($\hat{\boldsymbol{\beta}}$). So a fixed $\boldsymbol{\Gamma}$ such as $\boldsymbol{I}_p$ will not lead to an incorrect covariance for $\hat{\boldsymbol{\beta}}$, because both the projection matrix and V (covariance matrix of U or $\bar{U}$) are asymptotically unchanged and can be estimated via the smoothed functions.

In this paper, we aim to obtain analytical results by modifying the induced smoothing approach that produces asymptotically equivalent estimators. Note that the predicted value, $\hat{y}_i = \delta_i y_i + (1 - \delta_i)\hat{y}_{iC}$, where $\hat{y}_{iC}$ is a rather complicated and unsmooth function of $\boldsymbol{\beta}$. We suggest replacing $\hat{y}_{iC}$ with a modified version that is smooth and computationally convenient:

$$\tilde{y}_{iC} = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \frac{\tilde{\xi}_i}{\tilde{S}(e_i)}$$

where $\tilde{\xi}_i = \int_{e_i}^{\infty} u d\tilde{F}_\beta(u)$ (a smooth estimator), and $\tilde{F}(\cdot)$ is a smoothed estimator of $F(\cdot)$.

To obtain a smoothed estimate of $F(e)$ or $S(e) = 1 - F(e)$, first consider $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}} + h\boldsymbol{\Gamma}^{1/2}Z$, where the random vector $Z \sim \mathrm{N}(0, \boldsymbol{I}_p)$, with $\boldsymbol{I}_p$ being the identity matrix of dimension p. This motivates us to consider a smoothed estimate of $S(e)$ in $\boldsymbol{\beta}$ starting from the Nelson-Aalen estimator for the cumulative hazard function for any $e = y - \boldsymbol{x}^\top\boldsymbol{\beta}$,

$$\Lambda(e) = \sum_{j=1}^{n} \frac{\delta_j \mathbf{1}\{e_j < e\}}{n - \sum_{k=1}^{n} \mathbf{1}\{e_k < e_j\}}$$

where $\mathbf{1}\{E\}$ is the "indicator" of the event E, equal to 1 if E occurs and to 0 otherwise. By applying induced smoothing to the $\boldsymbol{\beta}$ used in forming the residuals, we obtain the following smoothed Nelson-Aalen estimator,

$$\begin{aligned}
\tilde{\Lambda}(e) &= \sum_{j=1}^{n} \frac{\delta_j \mathrm{E}_Z(\mathbf{1}\{y_j - y - (\boldsymbol{x}_j - \boldsymbol{x})^\top\hat{\boldsymbol{\beta}} + h(\boldsymbol{x}_j - \boldsymbol{x})^\top\boldsymbol{\Gamma}^{1/2}\boldsymbol{Z} < 0\})}{n - \sum_{k=1}^{n} \mathrm{E}_Z(\mathbf{1}\{y_k - y_j - (\boldsymbol{x}_k - \boldsymbol{x}_j)^\top\hat{\boldsymbol{\beta}} + h(\boldsymbol{x}_k - \boldsymbol{x}_j)^\top\boldsymbol{\Gamma}^{1/2}\boldsymbol{Z} < 0\})} \\
&= \sum_{j=1}^{n} \frac{\delta_j \mathrm{E}_Z(\mathbf{1}\{h(\boldsymbol{x}_j - \boldsymbol{x})^\top\boldsymbol{\Gamma}^{1/2}\boldsymbol{Z} < e - e_j\})}{n - \sum_{k=1}^{n} \mathrm{E}_Z(\mathbf{1}\{h(\boldsymbol{x}_k - \boldsymbol{x}_j)^\top\boldsymbol{\Gamma}^{1/2}\boldsymbol{Z} < e_j - e_k\})} \\
&= \sum_{j=1}^{n} \frac{\delta_j \Phi((e - e_j)/\sqrt{\boldsymbol{d}_{\cdot j}\boldsymbol{\Gamma}\boldsymbol{d}_{\cdot j}^\top h^2})}{n - \sum_{k=1}^{n} \Phi(r_{jk})},
\end{aligned} \tag{2}$$

in which $\boldsymbol{d}_{\cdot j} = (\boldsymbol{x} - \boldsymbol{x}_j)^\top$, $r_{jk} = (e_j - e_k)/\sqrt{u_{jk}}$, $u_{jk} = \boldsymbol{d}_{jk}\boldsymbol{\Gamma}\boldsymbol{d}_{jk}^\top h^2$, and $\boldsymbol{d}_{jk} = (\boldsymbol{x}_j - \boldsymbol{x}_k)^\top$. Here, $\boldsymbol{x}$ is the covariate vector corresponding to e, and $\Phi(\cdot)$ is the cumulative distribution function of the $\mathrm{N}(0,1)$ distribution. Assuming that $h \to 0$ as $n \to \infty$ and that for any given e there exists an η such that $\inf_{w \in [e-\eta, e+\eta]} h \sum_{k=1}^{n} \mathbf{1}\{e_k \geq w\} \xrightarrow{P} \infty$ as $n \to \infty$, we can show the pointwise consistency of $\tilde{\Lambda}(e)$ with respect to $\Lambda(e)$. Subsequently we can form the induced-smooth survival function $\tilde{S}(e) = \exp\{-\tilde{\Lambda}(e)\}$.

It can be seen that $\tilde{e}_i = \tilde{\xi}_i/\tilde{S}(e_i)$ is smooth in $\boldsymbol{\beta}$ and $\tilde{S}(e) - S(e) = o_p(n^{-1/2})$ (the proof is given in the Appendix). The final estimator can be obtained via the following iterative algorithm:

$$\hat{\boldsymbol{\beta}}_{(m+1)} = \tilde{g}(\hat{\boldsymbol{\beta}}_{(m)}) = \mathcal{X}^{-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})\tilde{y}_i, \tag{3}$$

where $\tilde{y}_i = \delta_i y_i + (1 - \delta_i)\tilde{y}_{iC}$, and $\tilde{y}_{iC}$ is evaluated at $\hat{\boldsymbol{\beta}}_{(m)}$.

The final solution $\hat{\boldsymbol{\beta}}$ satisfies $U(\hat{\boldsymbol{\beta}}) = n^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\tilde{y}_i - \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}) = 0$ or $\tilde{g}(\hat{\boldsymbol{\beta}}) - \hat{\boldsymbol{\beta}} = 0$. Let $U_0(\boldsymbol{\beta}_0) = n^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})[\delta_i e_i + (1-\delta_i)\mathrm{E}(\epsilon_i|\epsilon_i > e_i)]$. Because $\mathrm{E}(U_0(\boldsymbol{\beta}_0)) = 0$ and $\delta_i e_i + (1-\delta_i)\mathrm{E}(\epsilon_i|\epsilon_i > e_i)$ $(i = 1, 2, ..., n)$ are independent, $\sqrt{n}U_0(\boldsymbol{\beta}_0)$ asymptotically follows a normal distribution $\mathrm{N}(0, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top s_i^2,$$

and $s_i^2 = \mathrm{var}\{\delta_i e_i + (1-\delta_i)\mathrm{E}(\epsilon_i|\epsilon_i > e_i)\}$. By noting that $\sqrt{n}||U(\boldsymbol{\beta}_0) - U_0(\boldsymbol{\beta}_0)|| = o_p(1)$ and using the Chebyshev inequality, we know the distribution of $\sqrt{n}U(\boldsymbol{\beta}_0)$ is also asymptotically $\mathrm{N}(0, \boldsymbol{\Sigma})$. Since $U(\boldsymbol{\beta})$ is continuous in $\boldsymbol{\beta}$, the consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}$ can be easily established by using the Taylor series expansion of $U(\hat{\boldsymbol{\beta}})$.

We now consider the estimation of $s_i^2 = \mathrm{var}\{\delta_i e_i + (1-\delta_i)\tilde{e_i}\}$, which is equal to $\mathrm{E}(\delta_i e_i^2 + (1-\delta_i)\tilde{e_i}^2)$. This leads to an estimate of s_i^2 equal to

$$\hat{s}_i^2 = p_0\bar{e}^2 + (1-p_0)\bar{\bar{e}}^2$$

where p_0 is the proportion of uncensored data, $\bar{e}^2$ is the mean of e_i^2 for the uncensored data, and $\bar{\bar{e}}^2$ is the mean of $\tilde{e}_i^2$ for the censored data. Replacing s_i^2 by $\hat{s}_i^2$ in $\boldsymbol{\Sigma}$ gives an estimate of $\hat{\boldsymbol{\Sigma}}$ for $\boldsymbol{\Sigma}$.

We now work out $\mathrm{var}(\hat{\boldsymbol{\beta}})$. Denote the estimating function at the mth step of iteration by $U(\hat{\boldsymbol{\beta}}_{(m)}, \hat{\boldsymbol{\beta}}_{(m-1)}) = n^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\tilde{y}_i - \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{(m)})$, where $\tilde{y}_i$ is evaluated at $\hat{\boldsymbol{\beta}}_{(m-1)}$. Since $U(\boldsymbol{\beta})$ is smooth in $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}_{(m-1)}$ is consistent (starting from a consistent estimator), we now have

$$U(\hat{\boldsymbol{\beta}}_{(m)}, \hat{\boldsymbol{\beta}}_{(m-1)}) = U(\boldsymbol{\beta}_0) - \boldsymbol{B}(\hat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}_0) + (\boldsymbol{D} + \boldsymbol{B})(\hat{\boldsymbol{\beta}}_{(m-1)} - \boldsymbol{\beta}_0) + o_p(n^{-1/2}),$$

where $U(\boldsymbol{\beta}_0)$ is the estimating function $U(\hat{\boldsymbol{\beta}})$ evaluated at $\boldsymbol{\beta}_0$, $\boldsymbol{B} = n^{-1}\mathcal{X}$, and $\boldsymbol{D} = \mathrm{E}(U'(\boldsymbol{\beta}))$. This implies that

$$\hat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}_0 = (\boldsymbol{I} + \boldsymbol{B}^{-1}\boldsymbol{D})(\hat{\boldsymbol{\beta}}_{(m-1)} - \boldsymbol{\beta}_0) + \boldsymbol{B}^{-1}U(\boldsymbol{\beta}_0) + o_p(n^{-1/2}), \tag{4}$$

which leads to

$$\hat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}_0 = (\boldsymbol{I} + \boldsymbol{B}^{-1}\boldsymbol{D})^m(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_0) + (\boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{B}^{-1}\boldsymbol{D})^m)\boldsymbol{D}^{-1}U(\boldsymbol{\beta}_0) + o_p(n^{-1/2}).$$

Here, $\hat{\boldsymbol{\beta}}_G$ is a consistent initial estimator such as the Gehan estimator. Since $\boldsymbol{B} + \boldsymbol{D}$ is non-negative and therefore $(\boldsymbol{I} + \boldsymbol{B}^{-1}\boldsymbol{D})^m$ approaches a zero matrix as m becomes sufficiently large, the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_{(m)}$ is approximately equal to $\boldsymbol{\Gamma}/n = \boldsymbol{D}^{-1}\mathrm{cov}\{U(\boldsymbol{\beta}_0)\}\{\boldsymbol{D}^\top\}^{-1}$, and $\boldsymbol{\Gamma}$ can be estimated by $\hat{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\Sigma}}\{\hat{\boldsymbol{D}}^\top\}^{-1}$, where

$$\hat{\boldsymbol{D}} = \{U'(\boldsymbol{\beta})\}|_{\hat{\boldsymbol{\beta}}} = n^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{v}_i - \boldsymbol{x}_i)^\top.$$

Here the expression for $\boldsymbol{v}_i$ is a bit tedious, but easy enough to calculate:

$$\boldsymbol{v}_i = \frac{\partial(1-\delta_i)(\tilde{e}_i - e_i)}{\partial\boldsymbol{\beta}}|_{\hat{\boldsymbol{\beta}}} = (1-\delta_i)\left\{\tilde{e}_i\frac{\partial\tilde{\Lambda}(e_i)}{\partial\boldsymbol{\beta}}|_{\hat{\boldsymbol{\beta}}} - \tilde{S}^{-1}\frac{\partial\tilde{\xi}_i}{\partial\boldsymbol{\beta}}|_{\hat{\boldsymbol{\beta}}} + x_i\right\},$$

and

$$\frac{\partial\tilde{\Lambda}(e_i)}{\partial\boldsymbol{\beta}}|_{\hat{\boldsymbol{\beta}}} = -\sum_{j=1}^{n}\frac{\delta_j\left[\{n - \sum_{k=1}^{n}\Phi(\frac{e_{jk}}{h_2})\}\phi(\frac{e_{ij}}{h_1})\boldsymbol{d}_{ij}h_1^{-1} + \Phi(\frac{e_{ij}}{h_1})\{\sum_{k=1}^{n}\phi(\frac{e_{jk}}{h_2})\boldsymbol{d}_{jk}h_2^{-1}\}\right]}{\{n - \sum_{k=1}^{n}\Phi(\frac{e_{jk}}{h_2})\}^2}. \tag{5}$$

The quantities $h_1 = \sqrt{\boldsymbol{d}_{ij}\boldsymbol{\Gamma}\boldsymbol{d}_{ij}^\top h^2}$, $\quad h_2 = \sqrt{\boldsymbol{d}_{jk}\boldsymbol{\Gamma}\boldsymbol{d}_{jk}^\top h^2}$, $\quad e_{ij} = e_i - e_j$, and $\partial\tilde{\xi}_i/\partial\boldsymbol{\beta}$ can be obtained numerically or via some approximation approach.

The smoothed estimates and their covariance matrix estimator obtained by the sandwich approach are not greatly affected by the choice of h. In fact, the following result holds; the proof is given in the Appendix.

Proposition. Under regularity conditions, for $m \geq 1$, we have $\hat{\boldsymbol{\beta}}_{(m)} = \boldsymbol{\beta}_0 + o_p(n^{-1/2})$ and $\mathrm{var}(\hat{\boldsymbol{\beta}}_{(m)}) = \boldsymbol{D}^{-1}\mathrm{cov}\{U(\beta_0)\}\{\boldsymbol{D}^\top\}^{-1} + o_p(n^{-1})$.

After a single step of the iterative procedure, we have

$$\boldsymbol{B}(\hat{\boldsymbol{\beta}}_{(2)} - \boldsymbol{\beta}_0) = (\boldsymbol{B} + \boldsymbol{D})(\hat{\boldsymbol{\beta}}_{(1)} - \boldsymbol{\beta}_0) + U(\boldsymbol{\beta}_0) + o_p(n^{-1/2}), \tag{6}$$

which implies that $\hat{\boldsymbol{\beta}}_{(2)}$ and $\hat{\boldsymbol{\beta}}_{(1)}$ are consistent provided that the initial $\hat{\boldsymbol{\beta}}_{(0)}$ is consistent. It is also apparent that the asymptotic variance of $\hat{\boldsymbol{\beta}}_{(2)}$ is $\boldsymbol{D}^{-1}\mathrm{cov}\{U(\boldsymbol{\beta}_0)\}\{\boldsymbol{D}^\top\}^{-1}$, which is

$\mathbf{\Gamma}/n$. Thus it may suffice in practice to use just one iteration, producing the one-step smoothed Buckley-James estimator.

4. Simulation Studies

To examine the finite sample performance of the proposed one-step smoothed Buckley-James procedure, we now consider the same setup as in Jin, Lin, Wei et al. (2006). Specifically, the failure times are generated from the AFT model: $\log T = 2 + X_1 + X_2 + \epsilon$, where X_1 follows a Bernoulli distribution with a success probability of 0.5, and X_2 is normally distributed as $\mathrm{N}(0, 0.5^2)$. We investigate five different error distributions for ϵ: the standard normal; the standard logistic; the type I extreme-value distribution (location =0 and scale=1); a mixture of normals, $0.9\mathrm{N}(0,1) + 0.1\mathrm{N}(0,9)$ and a Weibull distribution with hazard rate $1/(2\sqrt{t})$ for $\exp(\epsilon)$, denoted by Weibull(0.5, 1). The censoring mechanism is the same as in Jin, Lin, Wei et al. (2006). We first generate an error vector e from the assumed distribution and hence obtain T. Censoring times (C) are generated from a uniform distribution on $(0, c)$ and the value c is so chosen that $\Pr(T > C) = p$, where p is the desired censoring proportion.

In addition to the proposed estimator (SBJ1), we also include, for efficiency comparisons, the profile likelihood estimator (PLE) proposed by Zeng & Lin (2008), the log-rank estimator (Jin, Lin, Wei et al. 2003), the least squares (LSQ) estimators proposed by Jin, Lin, Wei et al. (2006), the Gehan estimator, and the smoothed Buckley-James estimator after convergence (SBJ). The log-rank and least squares estimators are asymptotically efficient under the extreme-value and normal error distributions, respectively.

The results based on the 1000 simulations for $n = 100, 200$, and 400 are summarized in Table 1. The maximum number of iterations was set at 30 for all iterative estimators and the estimates at convergence or the last iteration were used. The convergence criterion was for the relative error to be ≤ 0.001 for both of the β_1 and β_2 estimates. The bandwidth h used was $1.3\sigma_2 n^{-1/3}$ (see Azzalini, 1981 and Giné & Nickl, 2009).

As can be seen from Table 1, the proposed estimators of β_1 and β_2 are virtually unbiased. The induced smoothing procedure for estimating variances appears to reflect the true variations, and the confidence intervals have proper coverage probabilities. We also evaluated the ratio (denoted by RE) of the mean squared errors of SBJ1 to those of the estimator of interest. Here the smaller values of RE indicate that SBJ1 is more efficient. The simulation results in Table 1 showed that the proposed SBJ1 estimator appeare to perform similarly to the LSQ and SBJ estimators in terms of mean squared errors. This is as expected because all three are aimed at improving only the computational aspect of the BJ estimator.

As one might expect, the performance of any estimators depends on the underlying distribution. When the errors are normally distributed, the new estimators (SBJ1 and SBJ) appear to be more efficient than the Gehan, the Log-rank, and the PLE estimators. Even for the logistic distribution, SBJ1 and SBJ appear to perform similarly to the Gehan estimator (all three appear to be more efficient than the Log-rank and PLE estimators). However, when the errors follow the extreme-value, Weibull, or mixture normal distributions, the SBJ estimator appear to become less efficient than the Gehan, the Log-rank, and the PLE. In addition, the Log-rank estimator seems to be most efficient for the extreme value distribution and the Gehan estimator seems to perform well for the logistic distribution. The smoothed BJ estimator seemed to perform similarly to the Log-rank estimator for the logistic and mixture normal errors.

Figure 1 displays comparisons of the proposed 1-step SBJ estimates of β_1 with the SBJ estimates after convergence and with the initial Gehan estimates. The comparisons are based on 1000 simulated datasets generated with standard normal errors and with a sample size of 100 and a 25% censoring rate. Ninety-nine point nine percent of the 1000 simulated datasets produced relative differences between the 1-step SBJ estimates and SBJ estimates that were less than or equal to 1%. On the other hand, the 1-step estimates were considerably different from the Gehan estimators, with roughly 40% of the relative differences larger than 1%. Similar phenomena were also observed for β_2 estimators.

We also conducted an additional set of simulations to compare the stability of the SBJ estimator with that of the LSQ estimator proposed in Jin, Lin, Wei et al. (2006) in terms of the number of simulations for which convergence is not achieved within 30 iterations (for both SBJ and LSQ estimators). Note that non-convergence typically takes the form of oscillation between multiple parameter values, and that maximal number of iterations permitted is not a contributing factor to non-convergence. The simulation results are summarized in Table 2 for $n = 50$ and $n = 100$. Table 2 indicates that when the sample size is relatively small ($n \leq 100$) the original Buckley-James and LSQ estimator may not exist. Table 2 also indicates that non-convergence rates are greatly reduced by our proposed smoothing method. This is consistent with the findings of Stare, Heinzl & Harrell (2000), who concluded that it is safe to use the Buckley-James estimator only when censoring is less than 20%.

As we expected, when the sample size is large ($n = 400$), non-convergence is no longer a concern for either LSQ or SBJ. Note also that SBJ1 is based on a one-step iteration whence it has no convergence issues even for small sample sizes, but nevertheless its performance is nearly as good as the full iteration estimator (see Table 1). Simulation code in R that was used to produce Tables 1 and 2 is available from the authors upon request.

5. Myeloma Study

We first apply the method to the data from a study on multiple myeloma (Krall, Uthoff & Harley 1975). The dataset is available as an example in the documentation for PROC PHREG in the SAS/STAT 12.1 user guide (pp. 493-494). We investigate two standardized covariates, LOGBUN and HGB. There are a total of 65 subjects. The Gehan rank estimates of (β_1, β_2)

are $\hat{\boldsymbol{\beta}} = (-0.532, 0.292)$. The PLE estimates are $(-0.606, 0.330)$ when starting from $(0, 0)$. However, the estimates become (-0.640, 0.337) when starting from the Gehan estimates as the initial values. The linear approximation method proposed by Jin, Lin, Wei et al. (2006) produces $\hat{\boldsymbol{\beta}} = (-0.5197, 0.2813)$.

We now consider the newly proposed procedure based on the BJ estimator. Using $\boldsymbol{\Gamma}_0 = \boldsymbol{I}_p/n$, we obtain convergence after 5 iterations to $\hat{\boldsymbol{\beta}} = (-0.5176, 0.2632)$. The intercept is estimated to be 3.967. When using very different initial values for β_1 and β_2, explicitly the least squares estimate and (0, 0), we obtain the same final estimates. This indicates that the smoothed iteration approach in general is capable of provide unique final estimates and that the initial values are not critical. Note that the asymptotic variance for $\hat{\beta}_1$ is the same as that for $\hat{\beta}_2$ if X_1 and X_2 are standardized. The corresponding non-asymptotic standard errors for $\hat{\boldsymbol{\beta}}$ are $(0.114, 0.113)$, which are about 25% smaller than those obtained from the Gehan's rank estimation (see Jin, Lin, Wei et al. 2003).

6. Discussion

In this paper, we have introduced a perturbation method for the BJ estimation method for censored data. The perturbation approach makes the projection matrix smooth and hence the iterations do not tend to oscillate between different values. Alternating between estimating the parameters and estimating the asymptotic covariance matrix $\boldsymbol{\Gamma} = n\text{var}(\hat{\boldsymbol{\beta}})$ leads to smoothed versions of both estimates. Use of the smoothed Kaplan-Meier function makes the iterative approach stable and induces reliable convergence.

The smoothing procedure is based on solving $\mathrm{E}_Z(U(\boldsymbol{\theta} + h\boldsymbol{\Gamma}^{1/2}\boldsymbol{Z})) = 0$ instead of $U(\boldsymbol{\theta}) = 0$, where the subscript $\boldsymbol{Z}$ indicates the expectation is taken with respect to the standard normal variable $\boldsymbol{Z}$. Zeng & Lin (2008) proposed using resampling methods for obtaining this expectation using $\boldsymbol{\Gamma} = \boldsymbol{I}_p$, which is also used as an initial covariance matrix in Brown & Wang (2005). In some cases, explicit expressions are not available but numerical evaluation is always possible through simulating many $\boldsymbol{Z}$ realizations (see Jin, Shao & Ying 2015). Generally speaking, if $U(\boldsymbol{\theta}+h\boldsymbol{\Gamma}^{1/2}\boldsymbol{Z})$ takes the form $G(H(\boldsymbol{\theta}),\boldsymbol{\theta})$, we can consider two-step induced smoothing, $\mathrm{E}_Z G(\bar{H}(\boldsymbol{\theta}),\boldsymbol{\theta}+h\boldsymbol{\Gamma}^{1/2}\ \boldsymbol{Z})$, where $\bar{H}(\boldsymbol{\theta}) = \mathrm{E}_Z(H(\boldsymbol{\theta}+h\boldsymbol{\Gamma}^{1/2}\boldsymbol{Z}))$. Such an approach becomes useful when both expectations have explicit expressions as is the case in the setting that we consider. Our example using the myeloma study clearly shows the advantages of smoothing. It would be of interest to see how it performs in other cases and to investigate how to apply such smoothing procedures more generally.

Our application of induced smoothing to the BJ estimator solves the "non-smoothness curse" in survival analysis. The statistical merits of this new estimator are not limited to the numerical studies presented here. For example, our work has made it possible to further improve the finite sample performance of the estimator in terms of bias, mean squared error and the achieved coverage probability of confidence intervals (the latter being related to estimation of standard errors of the estimators).

Appendices

A1. Proof of $\tilde{S}(e) - S(e) = o_p(n^{-1/2})$

We first show that the denominator $n - \sum_{k=1}^{n}\Phi\{r_{jk}\}$ converges to the jump process $Y(e_j) = \#\{i : e_i \geq e_j\}$, where $r_{jk} = (e_j - e_k)/\sqrt{u_{jk}}$, $u_{jk} = \boldsymbol{d}_{jk}\boldsymbol{\Gamma}\boldsymbol{d}_{jk}^{\top}h^2$ and $\boldsymbol{d}_{jk} = (\boldsymbol{x}_j - \boldsymbol{x}_k)^{\top}$. Denote $\sqrt{u_{jk}}$ by $h_{j,k}(\hat{\boldsymbol{\beta}})$. We re-write

$$\begin{aligned} n - \sum_{k=1}^{n}\Phi\{r_{jk}\} &= n\left(1 - \int_{-\infty}^{e_j} n^{-1}\sum_{k=1}^{n}\frac{1}{\sqrt{u_{jk}}}\phi\left\{\frac{t-e_k}{\sqrt{u_{jk}}}\right\}\mathrm{d}t\right) \\ &= n\left(1 - \int_{-\infty}^{e_j}\sum_{k=1}^{n}\frac{1}{nh_{j,k}(\hat{\boldsymbol{\beta}})}\phi\left\{\frac{t-e_k}{h_{j,k}(\hat{\boldsymbol{\beta}})}\right\}\mathrm{d}t\right). \end{aligned}$$

Note that the smoothing bandwidth function $h_{\cdot,\cdot}(\hat{\boldsymbol{\beta}})$ is of the order $O(h)$. Following the approach used in the proof of Theorem 2 in Terrell & Scott (1992), we can show that the integrand that appears in the right hand side is a consistent estimator of $f_e(t)$, the density function of e, and thus that the right hand side is a consistent estimator of $n\{1 - F(e_j)\}$. The point-wise asymptotic equivalence of $Y(e_j)$ and $n\{1 - F(e_j)\}$ follows by using the inequality presented in Theorem 1 in Giné & Nickl (2009).

We now demonstrate the point-wise convergence of $\tilde{\Lambda}(e)$. Set $\tilde{\Lambda}(s)$ equal to $\int_{-\infty}^{s} \tilde{\alpha}(e)\mathrm{d}e$, where

$$\tilde{\alpha}(e) = \sum_{j=1}^{n} \frac{n^{-1} h_{j,\cdot}^{-1}(\hat{\boldsymbol{\beta}})\phi((e - e_j)/h_{j,\cdot}(\hat{\boldsymbol{\beta}}))}{1 - n^{-1}\sum_{k=1}^{n} \Phi\{r_{jk}\}}.$$

For a given e, let $\hat{\alpha}(e) = \sum_{j=1}^{n} h_{j,\cdot}^{-1}(\hat{\boldsymbol{\beta}})\phi((e - e_j)/h_{j,\cdot}(\hat{\boldsymbol{\beta}}))Y^{-1}(e_j)$. It can be shown that $|\tilde{\alpha}(e) - \hat{\alpha}(e)| \xrightarrow{P} 0$, since ϕ is a bounded kernel, under the assumption $\inf_{w \in [e-\eta, e+\eta]} h \sum_{k=1}^{n} \mathbf{1}\{e_k \geq w\} \xrightarrow{P} \infty$ as $n \to \infty$.

Let $\hat{\Lambda}(s) = \int_{-\infty}^{s} \hat{\alpha}(e)\mathrm{d}e$ be the Nelson-Aalen estimator of $\Lambda(s) = \int_{-\infty}^{s} \alpha(e)\mathrm{d}e$, where $\alpha(e)$ is the hazard rate function defined based on residuals. We notice that $\hat{\alpha}(e)$ can be equivalently written as

$$\hat{\alpha}(e) = \int_{\mathbb{E}} h_{j,\cdot}^{-1}(\hat{\boldsymbol{\beta}})\phi((s - e_j)/h_{j,\cdot}(\hat{\boldsymbol{\beta}}))\mathrm{d}\hat{\Lambda}(s).$$

Following an approach similar to that used in Ramlau-Hansen (1983), we further obtain that $\hat{\alpha}(e) \xrightarrow{P} \alpha(e)$ as $n \to \infty$ by observing that the counting process arguments are still applicable here in terms of the residuals even though arbitrarily large negative values may be observed. This result implies that $\tilde{\alpha}(e) \xrightarrow{P} \alpha(e)$ as $n \to \infty$. By the dominated convergence theorem, it follows that $|\tilde{A}(e) - A(e)| \xrightarrow{P} 0$ as $n \to \infty$. The consistency of $\tilde{S}(e) = \exp\{-\tilde{\Lambda}(e)\}$ follows immediately.

A2. Proof of Proposition

Denote the estimating function $U(\beta)$ at the mth step of iteration by

$$U(\hat{\boldsymbol{\beta}}_{(\boldsymbol{m})}, \hat{\boldsymbol{\beta}}_{(m-1)}) = n^{-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\tilde{y}_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{(m)}),$$

where $\tilde{y}_i$ is evaluated at $\hat{\boldsymbol{\beta}}_{(m-1)}$. Since $U(\boldsymbol{\beta})$ is smooth in $\boldsymbol{\beta}$, if we suppose that $\hat{\boldsymbol{\beta}}_{(m-1)}$ is consistent when starting from a consistent estimator initially, we now have

$$U(\hat{\boldsymbol{\beta}}_{(m)}, \hat{\boldsymbol{\beta}}_{(m-1)}) = U(\boldsymbol{\beta}_0) - \boldsymbol{B}(\hat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}_0) + (\boldsymbol{D} + \boldsymbol{B})(\hat{\boldsymbol{\beta}}_{(m-1)} - \boldsymbol{\beta}_0) + o_p(n^{-1/2})$$

where $U(\boldsymbol{\beta}_0)$ is the estimating function $U(\hat{\boldsymbol{\beta}})$ evaluated at $\boldsymbol{\beta}_0$, $\boldsymbol{B} = n^{-1}\mathcal{X}$, and $\boldsymbol{D} = \mathrm{E}(U'(\boldsymbol{\beta}))_{\boldsymbol{\beta}_0}$. Because $U(\hat{\boldsymbol{\beta}}_{(m)}, \hat{\boldsymbol{\beta}}_{(m-1)}) = 0$, we obtain

$$\hat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}_0 = (\boldsymbol{I} + \boldsymbol{B}^{-1}\boldsymbol{D})(\hat{\boldsymbol{\beta}}_{(m-1)} - \boldsymbol{\beta}_0) + \boldsymbol{B}^{-1}U(\boldsymbol{\beta}_0) + o_p(n^{-1/2}).$$

For a given $m \geq 1$, we have

$$\hat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}_0 = (\boldsymbol{I} + \boldsymbol{B}^{-1}\boldsymbol{D})^m(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_0) + (\boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{B}^{-1}\boldsymbol{D})^m)\boldsymbol{D}^{-1}U(\boldsymbol{\beta}_0) + o_p(n^{-1/2})$$

Therefore,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}_0) = (\boldsymbol{I} + \boldsymbol{B}^{-1}\boldsymbol{D})^m\sqrt{n}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_0) + (\boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{B}^{-1}\boldsymbol{D})^m)\boldsymbol{D}^{-1}\sqrt{n}U(\boldsymbol{\beta}_0) + o_p(1)$$

where $\hat{\boldsymbol{\beta}}_G$ is a consistent initial estimator. According to Jin, Lin, Wei et al. (2003), $\sqrt{n}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_0)$ is asymptotically normal as $n \to +\infty$, and when the level of censorship shrinks to zero, the matrix $-\boldsymbol{D}$ approaches $\boldsymbol{B}$, $\boldsymbol{B} + \boldsymbol{D}$ is non-negative, and therefore $(\boldsymbol{I} + \boldsymbol{B}^{-1}\boldsymbol{D})^m$ approaches a zero matrix as m becomes sufficiently large (Jin, Lin, Wei et al. 2006). Consequenctly $\sqrt{n}(\hat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}_0) = o_p(1)$, and the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_{(m)}$ is approximately equal to $\boldsymbol{\Gamma}/n = \boldsymbol{D}^{-1}\mathrm{cov}\{U(\boldsymbol{\beta}_0)\}\{\boldsymbol{D}^\top\}^{-1}$.

References

AZZALINI, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika* **68**, 326–28.

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

BUCKLEY, J. & JAME, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429–36.

COX, D.R. (1972). Regression models and life-tables (with Discussion). *J. R. Statist. Soc. B* **34**, 187–220.

FU, L. & WANG, Y-G. (2011). Nonparametric rank regression for analyzing water quality concentration data with multiple detection limits. *Environmental Science and Technology* **45**, 1481–1489.

GINÉ, E. & 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**, 569–96.

JAECKEL, L. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **42**, 1540–1552.

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

JIN, Z., LIN, D.Y., WEI, L.J. & YING, Z. (2006). On least squares regression with censored data. *Biometrika* **93**, 147–162.

JIN Z., SHAO, Y. & YING, Z. (2015). A Monte Carlo method for variance estimation for estimators based on induced smoothing. *Biostatistics* **16** : 179–88.

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

KALBFLEISCH, J.D. & PRENTICE, R.L. (2002). *The Statistical Analysis of Failure Time Data.*, 2nd ed. New York: Wiley.

KRALL, J.M., UTHOFF, V.A. & HARLEY, J.B. (1975). A step-up procedure for selecting variables associated with survival. *Biometrics* **31**, 49–57.

LAROCQUE, D. & NEVALAINEN, J. (2007). A weighted multivariate sign test for cluster-correlated data. *Biometrika* **94**, 267–283.

LAWLESS, J.F. (2003). *Statistical Models and methods for Lifetime Data*, 2nd ed. New York: Wiley.

LI, H. & YIN G.S. (2009). Generalized method of moments estimation for linear regression with clustered failure time data. *Biometrika* **96**, 293–306.

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

STARE, J., HEINZL, H. & 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, 125–34.

TERRELL, G.R. & SCOTT, D.W. (1992). Variable Kernel Density Estimation. *Ann. Statist.* **20**, 1236–1265.

WANG, Y-G. & FU, L. (2011). Rank regression for the accelerated failure time model with clustered and censored data. *Comput. Statist. Data Anal.* **55**, 2334–2343.

ZENG, D. & LIN, D.Y. (2008). Efficient resampling methods for nonsmooth estimating functions. *Biostatistics* **9**, 355–363.

TABLE 1

Summary of simulation studies. SE: standard error of the parameter estimator; SEE: mean of the standard error estimator; CP: coverage probability of the 95% confidence interval; RE: ratio of the mean squared errors of SBJ1 to the estimator of interest.

			SBJ1				LSQ		Gehan		Log-rank		PLE		SBJ	
N	Censoring		Bias	SE	SEE	CP	Bias	RE	Bias	RE	Bias	RE	Bias	RE	Bias	RE
Normal error																
100	25%	β_1	-0.002	0.222	0.221	0.942	-0.003	1.005	-0.003	0.957	0.004	0.890	-0.003	0.899	-0.003	1.000
		β_2	0.001	0.221	0.226	0.950	0.001	1.007	-0.001	0.946	0.009	0.906	-0.003	0.884	0.001	1.000
	50%	β_1	-0.001	0.258	0.267	0.942	-0.002	1.020	-0.002	0.934	-0.006	0.957	-0.006	0.853	-0.002	1.003
		β_2	-0.011	0.280	0.272	0.925	-0.012	1.025	-0.009	0.920	-0.013	0.975	-0.012	0.823	-0.011	1.006
200	25%	β_1	-0.005	0.152	0.157	0.956	-0.005	1.003	-0.007	0.945	-0.002	0.881	-0.006	0.932	-0.005	1.000
		β_2	-0.001	0.158	0.158	0.941	-0.001	1.001	-0.002	0.946	0.002	0.872	-0.002	0.922	-0.001	1.000
	50%	β_1	0.012	0.192	0.187	0.940	0.011	1.014	0.012	0.907	0.010	0.965	0.010	0.943	0.011	1.008
		β_2	0.006	0.193	0.190	0.935	0.006	1.020	0.007	0.893	0.005	1.000	0.006	0.921	0.006	1.011
400	25%	β_1	0.005	0.106	0.110	0.962	0.005	1.002	0.006	0.947	0.003	0.898	0.006	0.962	0.005	1.000
		β_2	0.008	0.116	0.110	0.928	0.008	1.002	0.008	0.963	0.008	0.962	0.006	0.950	0.008	0.999
	50%	β_1	0.005	0.127	0.128	0.954	0.006	1.004	0.005	0.919	0.005	0.966	0.008	0.940	0.006	1.004
		β_2	-0.003	0.131	0.129	0.944	-0.003	1.010	-0.004	0.935	-0.001	1.002	-0.003	0.978	-0.003	1.003
Logistic error																
100	25%	β_1	0.003	0.408	0.398	0.948	0.003	0.999	0.001	1.044	0.005	0.908	-0.001	0.952	0.003	0.997
		β_2	-0.021	0.412	0.403	0.945	-0.021	1.005	-0.014	1.032	-0.020	0.904	-0.021	0.906	-0.021	0.998
	50%	β_1	0.016	0.457	0.469	0.954	0.017	1.000	0.014	0.980	0.013	0.944	0.015	0.804	0.016	0.991
		β_2	0.010	0.480	0.477	0.938	0.010	1.008	0.006	1.010	0.003	0.985	0.006	0.816	0.011	0.990
200	25%	β_1	0.001	0.277	0.281	0.949	0.001	1.002	-0.0005	1.046	0.008	0.927	-0.0004	0.956	0.001	0.997
		β_2	0.008	0.280	0.282	0.951	0.008	0.998	0.006	1.050	0.003	0.920	0.007	0.943	0.008	0.997
	50%	β_1	0.002	0.315	0.332	0.959	0.001	0.995	0.002	1.007	-0.006	0.965	0.005	0.914	0.003	0.992
		β_2	0.009	0.342	0.335	0.932	0.009	1.001	0.009	0.994	0.013	0.932	0.013	0.908	0.009	0.994
400	25%	β_1	0.001	0.202	0.197	0.948	0.002	0.993	-0.001	1.118	0.001	0.875	-0.003	1.024	0.001	0.997
		β_2	-0.005	0.197	0.197	0.958	-0.005	0.999	-0.005	1.062	-0.003	0.927	-0.004	1.012	-0.005	0.998
	50%	β_1	-0.005	0.225	0.228	0.944	-0.005	0.995	-0.004	0.977	-0.001	0.919	-0.003	0.894	-0.005	0.995
		β_2	-0.003	0.234	0.229	0.940	-0.003	0.998	-0.006	1.004	0.003	1.006	-0.006	0.959	-0.003	0.995
Extreme value error																
100	25%	β_1	0.010	0.287	0.293	0.944	0.011	1.007	0.007	1.168	0.001	1.596	0.009	1.072	0.010	0.993
		β_2	0.023	0.302	0.302	0.931	0.022	1.005	0.020	1.205	0.011	1.635	0.017	1.090	0.023	0.991
	50%	β_1	0.001	0.350	0.343	0.930	0.002	1.017	0.001	1.109	0.004	1.650	0.0004	0.881	0.001	0.976
		β_2	-0.012	0.336	0.347	0.952	-0.011	1.016	-0.016	1.101	-0.011	1.591	-0.015	0.937	-0.011	0.974
200	25%	β_1	0.006	0.204	0.202	0.937	0.005	1.004	0.004	1.165	0.005	1.722	0.010	1.265	0.006	0.994
		β_2	-0.006	0.197	0.203	0.952	-0.006	1.004	-0.004	1.177	-0.003	1.617	-0.001	1.210	-0.006	0.992
	50%	β_1	0.009	0.231	0.240	0.957	0.009	1.001	0.010	1.129	0.009	1.728	0.012	1.251	0.009	0.971
		β_2	-0.007	0.235	0.241	0.943	-0.007	1.011	-0.010	1.088	-0.007	1.703	-0.010	1.154	-0.007	0.979
400	25%	β_1	-0.006	0.146	0.142	0.952	-0.006	1.000	-0.006	1.138	-0.008	1.654	-0.006	1.270	-0.006	0.995
		β_2	0.004	0.144	0.143	0.962	0.004	0.998	0.009	1.211	0.006	1.808	0.009	1.450	0.004	0.993
	50%	β_1	0.003	0.174	0.167	0.944	0.003	0.991	0.002	1.097	-0.002	1.727	0.001	1.361	0.003	0.979
		β_2	-0.001	0.164	0.168	0.964	-0.001	1.002	-0.001	1.042	-0.002	1.690	-0.001	1.363	-0.001	0.985
Mixture Normal error																
100	25%	β_1	-0.012	0.299	0.289	0.943	-0.012	0.984	-0.016	1.301	-0.017	1.040	-0.017	1.223	-0.012	0.993
		β_2	-0.0003	0.295	0.289	0.949	0.0003	0.987	0.002	1.305	0.004	1.070	0.0006	1.249	-0.0003	0.992
	50%	β_1	-0.002	0.323	0.335	0.945	-0.002	0.970	0.002	1.130	0.002	1.032	0.002	0.959	-0.002	0.974
		β_2	-0.002	0.333	0.338	0.949	-0.001	0.988	-0.005	1.121	-0.003	1.047	-0.007	0.998	-0.002	0.972
200	25%	β_1	-0.003	0.203	0.204	0.953	-0.002	0.995	-0.004	1.250	-0.008	1.021	-0.005	1.190	-0.003	0.995
		β_2	0.007	0.202	0.207	0.959	0.007	0.988	0.008	1.265	0.005	1.035	0.011	1.184	0.008	0.994
	50%	β_1	-0.0003	0.228	0.237	0.960	-0.001	0.965	0.003	1.086	-0.005	0.977	0.003	1.019	0.0003	0.982
		β_2	-0.004	0.234	0.240	0.957	-0.004	0.983	-0.009	1.121	-0.005	1.046	-0.008	1.104	-0.004	0.979
400	25%	β_1	0.010	0.142	0.143	0.948	0.010	0.993	0.005	1.264	0.007	1.061	0.005	1.238	0.010	0.996
		β_2	0.015	0.144	0.143	0.942	0.014	0.993	0.014	1.251	0.014	1.041	0.015	1.213	0.015	0.996
	50%	β_1	-0.006	0.156	0.163	0.956	-0.006	0.975	-0.006	1.130	-0.005	1.059	-0.006	1.157	-0.006	0.980
		β_2	0.008	0.163	0.162	0.950	0.008	0.977	0.005	1.143	0.012	1.030	0.007	1.080	0.008	0.981
Log-Weibull error																
100	25%	β_1	0.011	0.601	0.588	0.936	0.010	1.005	0.008	1.187	-0.007	1.698	0.008	1.132	0.011	0.992
		β_2	-0.011	0.585	0.621	0.942	-0.011	1.011	-0.008	1.170	0.002	1.515	0.003	1.088	-0.011	0.991
	50%	β_1	-0.030	0.677	0.700	0.949	-0.030	1.010	-0.032	1.139	-0.021	1.765	-0.039	0.986	-0.029	0.968
		β_2	0.017	0.701	0.710	0.947	0.016	1.016	0.014	1.124	-0.016	1.761	-0.002	0.999	0.017	0.970
200	25%	β_1	0.006	0.392	0.403	0.961	0.006	1.001	0.006	1.142	-0.006	1.648	0.002	1.214	0.006	0.994
		β_2	0.012	0.395	0.407	0.958	0.012	1.003	0.012	1.141	0.003	1.615	0.012	1.216	0.012	0.994
	50%	β_1	0.006	0.488	0.476	0.945	0.005	0.997	0.004	1.137	-0.001	1.797	-0.001	1.286	0.006	0.973
		β_2	0.006	0.467	0.485	0.943	0.005	0.999	0.002	1.109	-0.007	1.693	-0.005	1.217	0.006	0.972
400	25%	β_1	-0.019	0.291	0.284	0.950	-0.019	1.002	-0.021	1.201	-0.016	1.810	-0.014	1.499	-0.019	0.994
		β_2	-0.006	0.278	0.284	0.952	-0.006	1.000	-0.008	1.229	-0.007	1.744	-0.009	1.454	-0.006	0.993
	50%	β_1	-0.010	0.335	0.332	0.956	-0.010	0.993	-0.004	1.107	0.004	1.827	0.006	1.418	-0.010	0.977
		β_2	-0.023	0.330	0.333	0.954	-0.024	0.998	-0.021	1.079	-0.010	1.864	-0.009	1.503	-0.024	0.981

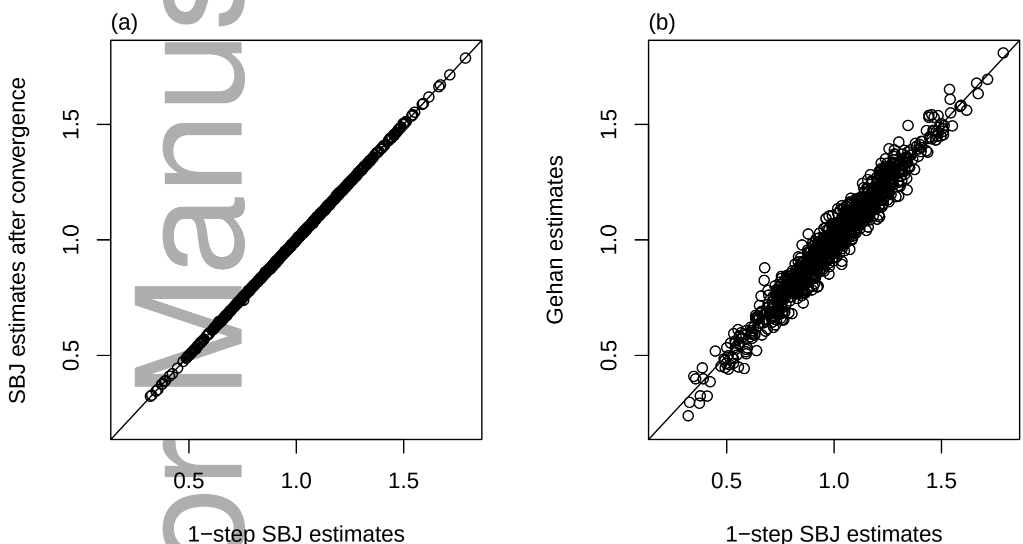
TABLE 2

Counts of instances of non-convergence within 30 iterations over 1000 simulations. RE: ratio of the mean squared error of SBJ to that of LSQ; # of NC: counts of non-convergence

			LSQ			SBJ		
N	Censoring		Bias	SE	# of NC	Bias	RE	# of NC
				Normal error				
50	25%	β_1	0.008	0.329	144	0.009	1.006	0
		β_2	0.010	0.325		0.010	1.011	
	50%	β_1	-0.006	0.378	255	-0.006	1.017	2
		β_2	0.015	0.390		0.014	1.036	
100	25%	β_1	-0.001	0.218	25	-0.001	1.005	0
		β_2	0.006	0.230		0.007	1.002	
	50%	β_1	0.017	0.265	57	0.016	1.027	0
		β_2	-0.008	0.260		-0.008	1.014	
				Logistic error				
50	25%	β_1	0.001	0.578	204	0.010	1.014	1
		β_2	0.003	0.597		-0.007	1.013	
	50%	β_1	0.014	0.644	386	0.010	1.023	2
		β_2	-0.009	0.687		-0.007	1.029	
100	25%	β_1	0.009	0.396	80	0.008	1.005	1
		β_2	-0.006	0.410		-0.006	1.003	
	50%	β_1	-0.015	0.465	193	-0.014	1.009	1
		β_2	0.009	0.459		0.008	1.011	
				Extreme value error				
50	25%	β_1	0.013	0.423	135	0.013	1.032	1
		β_2	-0.026	0.419		-0.026	1.042	
	50%	β_1	-0.006	0.473	261	0.930	1.075	0
		β_2	0.002	0.523		-0.004	1.068	
100	25%	β_1	-0.004	0.285	18	-0.005	1.021	0
		β_2	0.003	0.297		0.003	1.020	
	50%	β_1	-0.003	0.338	43	-0.004	1.045	0
		β_2	-0.003	0.347		-0.003	1.049	
				Mixture Normal error				
50	25%	β_1	-0.013	0.420	185	-0.013	1.005	1
		β_2	0.014	0.413		0.015	0.999	
	50%	β_1	0.012	0.323	328	0.011	1.024	2
		β_2	0.019	0.333		0.020	1.033	
100	25%	β_1	0.001	0.283	72	-0.001	1.000	1
		β_2	0.015	0.304		0.015	0.992	
	50%	β_1	-0.013	0.339	126	-0.012	0.994	1
		β_2	-0.009	0.343		-0.008	1.007	
				Log-Weibull error				
50	25%	β_1	0.015	0.850	180	0.016	1.034	2
		β_2	0.038	0.826		0.035	1.034	
	50%	β_1	-0.013	1.018	322	-0.014	1.071	2
		β_2	0.016	1.000		0.015	1.078	
100	25%	β_1	0.013	0.565	89	0.013	1.018	0
		β_2	-0.006	0.578		-0.006	1.016	
	50%	β_1	-0.011	0.680	162	-0.012	1.045	0
		β_2	-0.013	0.675		-0.012	1.043	

Figure Caption

Figure 1: Simulation results assuming normal errors and a 25% censoring rate. (a) Smoothed Buckley-James estimates versus one-step smoothed Buckley-James estimates; (b) Gehan estimates versus one-step smoothed Buckley-James estimates


(a)
SBJ estimates after convergence
0.5
1.0
1.5
1−step SBJ estimates
(b)
Gehan estimates
0.5
1.0
1.5
1−step SBJ estimates

anzs_12155_f1.pdf