# On least-squares regression with censored data

By ZHEZHEN JIN

*Department of Biostatistics, Columbia University, New York, New York 10032, U.S.A.*

zjin@biostat.cpmc.columbia.edu

D. Y. LIN

*Department of Biostatistics, CB 7420, University of North Carolina, Chapel Hill,
North Carolina 27599, U.S.A.*

lin@bios.unc.edu

AND ZHILIANG YING

*Department of Statistics, Columbia University, New York, New York 10027, U.S.A.*

zying@stat.columbia.edu

SUMMARY

The semiparametric accelerated failure time model relates the logarithm of the failure time linearly to the covariates while leaving the error distribution unspecified. The present paper describes simple and reliable inference procedures based on the least-squares principle for this model with right-censored data. The proposed estimator of the vector-valued regression parameter is an iterative solution to the Buckley–James estimating equation with a preliminary consistent estimator as the starting value. The estimator is shown to be consistent and asymptotically normal. A novel resampling procedure is developed for the estimation of the limiting covariance matrix. Extensions to marginal models for multivariate failure time data are considered. The performance of the new inference procedures is assessed through simulation studies. Illustrations with medical studies are provided.

*Some key words*: Accelerated failure time model; Buckley-James estimator; Gehan statistic; Linear model; Linear programming; Rank estimator; Resampling; Semiparametric model; Survival data; Variance estimation.

## 1. INTRODUCTION

The linear regression model, together with the least-squares estimator, plays a fundamental role in data analysis. For potentially censored failure time data, the least-squares estimator cannot be calculated because the failure times are unknown for censored observations. A number of authors (Miller, 1976; Buckley & James, 1979; Koul et al., 1981) extended the least-squares principle so as to accommodate censoring. The estimator of Miller (1976) requires that the censoring time satisfy the same regression model as the failure time, while the estimator of Koul et al. (1981) requires that the censoring time be independent of covariates. Miller & Halpern (1982) found that the Buckley–James estimator is more reliable than those of Miller and Koul et al.

The asymptotic properties of the Buckley–James estimator were studied rigorously by Ritov (1990) and Lai & Ying (1991). They showed that, with a slight modification to the tail, any consistent root of the Buckley–James estimating function must be asymptotically normal and that the estimator is semiparametrically efficient when the underlying error distribution is normal, which is a well-known property of the least-squares estimator for uncensored data. The limiting covariance matrix, however, involves the unknown hazard function of the error term and it is therefore difficult to estimate it directly.

The Buckley–James estimator is a root of an estimating function which is discontinuous and may have multiple roots even in the limit. To facilitate computation, Buckley & James (1979) proposed a semiparametric em algorithm which iterates between imputation of censored failure times and least-squares estimation. The convergence of the algorithm is not guaranteed. In fact, it is known that the iterative sequence may become trapped in a loop, oscillating between two or more points.

As a result of its weak requirements on the censoring mechanism and its comparable efficiency with the classical least-squares estimator, the Buckley–James estimator is a natural choice for the accelerated failure time model. The developments of this estimator have been largely academic for two major reasons. First, there does not exist a computationally efficient algorithm that guarantees a consistent solution. Secondly, there is no reliable method for estimating the sampling distribution of the estimator or the standard error thereof.

A key step in the Buckley–James iterative algorithm is the initial estimator. As shown in Ritov (1990) and Lai & Ying (1991), the estimating function is locally asymptotically linear. Using this result, we can show that, if the initial estimator is consistent, then, for each fixed $m$, an $m$-step estimator must also be consistent. In addition, if the initial estimator is asymptotically normal, then so is the $m$-step estimator.

In the light of the foregoing findings, we propose to approximate the consistent root of the Buckley–James estimating equation by using a consistent estimator as the initial value in the Buckley–James iteration. To be specific, the initial value is chosen to be the rank estimator with the Gehan (1965) weight function (Prentice, 1978; Tsiatis, 1990), which can be easily calculated via the linear programming technique (Jin et al., 2003). To estimate the limiting covariance matrix, we develop a resampling scheme which involves similar iterations. The resampling approach shares the spirit of that of Jin et al. (2003). However, the actual procedure is very different because the Buckley–James estimating function is not related to the Gehan statistic and involves the Kaplan–Meier estimator of the error distribution in a complex manner. We provide rigorous justifications for the proposed parameter estimator and resampling procedure, and demonstrate their usefulness through simulated and real data.

## 2. Accelerated failure time model and Buckley–James estimator

Suppose that there is a random sample of $n$ subjects. For $i = 1, \ldots, n$, let $T_i$ and $C_i$ be, respectively, the failure time and censoring time for the $i$th subject, and let $X_i$ be the corresponding $p$-vector of covariates. As usual, assume that $T_i$ and $C_i$ are independent conditional on $X_i$. The data consist of $(\tilde{T}_i, \delta_i, X_i)$ $(i = 1, \ldots, n)$, where $\tilde{T}_i = \min(T_i, C_i)$, $\delta_i = 1_{\{T_i \leqslant C_i\}}$, and $1_{\{.\}}$ is the indicator function.

Write $Y_i = \log T_i$. The semiparametric linear regression model takes the form

$$Y_i = X_i' \beta_0 + \varepsilon_i, \tag{1}$$

where $\beta_0$ is a $p$-vector of unknown regression parameters, and $\varepsilon_i$ $(i = 1, \ldots, n)$ are independent error terms with a common but completely unspecified distribution function $F$. Model (1) if often referred to as the accelerated failure time or accelerated life model (Cox & Oakes, 1984, pp. 64–5; Kalbfleisch & Prentice, 2002, pp. 218–9). This model is intuitively appealing as it provides a direct characterisation of the effects of covariates on the failure time. One may replace the log-transformation of the failure time in (1) by a different transformation.

For uncensored data, the classical least-squares estimator is obtained by minimising the objective function

$$n^{-1} \sum_{i=1}^{n} (Y_i - \alpha - X_i'\beta)^2 \tag{2}$$

with respect to $\alpha$ and $\beta$, where $\alpha$ pertains to the mean of the error distribution. The minimisation of (2) yields the following estimating equation for $\beta_0$:

$$\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - X_i'\beta) = 0, \tag{3}$$

where $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$. Of course, the resulting estimator has a simple closed-form expression and its covariance matrix can be easily estimated.

In the presence of censoring, the values of the $T_i$ associated with $\delta_i = 0$ are unknown, so that (3) cannot be used directly to estimate $\beta_0$. Buckley & James (1979) modified (3) by replacing each $Y_i$ with $E(Y_i | \tilde{T}_i, \delta_i, X_i)$, which is approximated by

$$\hat{Y}_i(\beta) = \delta_i \tilde{Y}_i + (1 - \delta_i) \left[ \frac{\int_{e_i(\beta)}^{\infty} u \, d\hat{F}_\beta(u)}{1 - \hat{F}_\beta\{e_i(\beta)\}} + X_i'\beta \right],$$

where $\tilde{Y}_i = \log \tilde{T}_i$, $e_i(\beta) = \tilde{Y}_i - X_i'\beta$ and $\hat{F}_\beta$ is the Kaplan–Meier estimator of $F$ based on the transformed data $\{e_i(\beta), \delta_i\}$ $(i = 1, \ldots, n)$, that is

$$\hat{F}_\beta(t) = 1 - \prod_{i:e_i(\beta)<t} \left( 1 - \frac{\delta_i}{\sum_{j=1}^{n} 1_{\{e_j(\beta) \geqslant e_i(\beta)\}}} \right). \tag{4}$$

Define

$$U(\beta, b) = \sum_{i=1}^{n} (X_i - \bar{X})\{\hat{Y}_i(b) - X_i'\beta\},$$

or

$$U(\beta, b) = \sum_{i=1}^{n} (X_i - \bar{X})\{\hat{Y}_i(b) - \bar{Y}(b) - (X_i - \bar{X})'\beta\},$$

where $\bar{Y}(b) = n^{-1} \sum_{i=1}^{n} \hat{Y}_i(b)$. Then the Buckley–James estimator $\hat{\beta}_{\mathrm{BJ}}$ is the root of $U(\beta, \beta) = 0$. It is easy to see that $U(\beta, \beta)$ is neither continuous nor monotone in $\beta$. Thus, it is difficult to calculate the estimator, especially when $\beta$ is multi-dimensional.

## 3. New inference procedures

Following Buckley & James (1979), we can 'linearise' the estimating function by first fixing an initial value $b$ and then solving the equation $U(\beta, b) = 0$ for $\beta$. This operation

leads to $\beta = L(b)$, where

$$L(b) = \left\{ \sum_{i=1}^{n} (X_i - \bar{X})^{\otimes 2} \right\}^{-1} \left[ \sum_{i=1}^{n} (X_i - \bar{X})\{\hat{Y}_i(b) - \bar{Y}(b)\} \right].$$

Here and in the sequel $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa'$. Continuing this process yields a simple iterative algorithm,

$$\hat{\beta}_{(m)} = L(\hat{\beta}_{(m-1)}) \quad (m \geqslant 1). \tag{5}$$

It can be shown through the arguments of Lai & Ying (1991) that $L(b)$ is asymptotically linear in $b$. Thus, if a consistent estimator of $\beta_0$ is chosen as the initial value in (5), then, for any fixed $m$, $\hat{\beta}_{(m)}$ should also be consistent. In addition, $\hat{\beta}_{(m)}$ is expected to be asymptotically normal if the initial estimator is asymptotically normal.

A consistent and asymptotically normal initial estimator of $\beta_0$ can be obtained by the rank-based method of Jin et al. (2003). We set the initial estimator $\hat{\beta}_{(0)}$ to the Gehan-type rank estimator $\hat{\beta}_G$, which can be calculated by minimising the convex function

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \{e_i(\beta) - e_j(\beta)\}^-,$$

where $a^- = 1_{\{a < 0\}}|a|$. This minimisation is a simple linear programming problem (Jin et al., 2003). Given $\hat{\beta}_{(0)}$, the iteration in (5) involves trivial calculations of the least-squares estimators.

We show in the Appendix that, for each fixed $m$, $\hat{\beta}_{(m)}$ is consistent and asymptotically normal. In addition, $\hat{\beta}_{(m)}$ is asymptotically a linear combination of the Gehan estimator $\hat{\beta}_G$ and the Buckley–James estimator $\hat{\beta}_{BJ}$ in that

$$\hat{\beta}_{(m)} = (I - D^{-1}A)^m \hat{\beta}_G + \{I - (I - D^{-1}A)^m\}\hat{\beta}_{BJ} + o_p(n^{-\frac{1}{2}}), \tag{6}$$

where $I$ is the identity matrix, $D := \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^{\otimes 2}$ is the usual slope matrix of the least-squares estimating function for uncensored data, and $A$ is the slope matrix of the Buckley–James estimating function defined in the Appendix.

When the level of censorship shrinks to zero, the matrix $A$ approaches $D$. Then the first term on the right-hand side of (6) becomes negligible and every $\hat{\beta}_{(m)}$ approaches the usual least-squares estimator. If the iterative algorithm given in (5) converges, then the limit solves exactly the original Buckley–James estimating equation. Even if the iterative sequence does not converge, the estimators are still consistent and asymptotically normal. In terms of the large-sample behaviour characterised by (6), it can be shown that, if the hazard function $\lambda(y)$ of the error distribution is nondecreasing in $y$, as is the case in particular with the normal, logistic and double-exponential distributions, then the matrix $D - A$ is nonnegative definite, which implies that $(I - D^{-1}A)^m$ approaches 0 or $\hat{\beta}_{(m)}$ approaches $\hat{\beta}_{BJ}$ as $m$ tends to $\infty$.

It follows from (6) that $\hat{\beta}_{(m)}$ is asymptotically normal, as shown formally in the Appendix. Since the limiting covariance matrices of both $\hat{\beta}_G$ and $\hat{\beta}_{BJ}$ involve the unknown hazard function $\lambda(.)$, the limiting covariance matrix of $\hat{\beta}_{(m)}$ does too. Thus, we develop a resampling procedure to approximate the distribution of $\hat{\beta}_{(m)}$.

Let $\hat{\beta}_G^*$ be a minimiser of

$$\sum_{i=1}^{n} \sum_{j=1}^{n} Z_i Z_j \delta_i \{e_i(\beta) - e_j(\beta)\}^-,$$

where $Z_i$ $(i=1,\ldots,n)$ are independent positive random variables with $E(Z_i)=\text{var}(Z_i)=1$. This is a slight modification of the one given in Jin et al. (2003). We further define

$$L^*(b) = \left\{ \sum_{i=1}^{n} Z_i(X_i - \bar{X})^{\otimes 2} \right\}^{-1} \left[ \sum_{i=1}^{n} Z_i(X_i - \bar{X})\{\hat{Y}_i^*(b) - \bar{Y}^*(b)\} \right],$$

where

$$\hat{Y}_i^*(b) = \delta_i \tilde{Y}_i + (1 - \delta_i) \left[ \frac{\int_{e_i(b)}^{\infty} u d\hat{F}_b^*(u)}{1 - \hat{F}_b^*\{e_i(b)\}} + X_i'b \right],$$

$$\hat{F}_b^*(t) = 1 - \prod_{i:e_i(b) < t} \left( 1 - \frac{Z_i \delta_i}{\sum_{j=1}^{n} Z_j 1_{\{e_j(b) \geq e_i(b)\}}} \right),$$

and $\bar{Y}^*(b) = n^{-1} \sum_{i=1}^{n} \hat{Y}_i^*(b)$. Finally, we define the iterative sequence $\hat{\beta}_{(0)}^* = \hat{\beta}_G^*$ and $\hat{\beta}_{(m)}^* = L^*(\hat{\beta}_{(m-1)}^*)$ $(m \geq 1)$.

In the Appendix, we show that the conditional distribution of $n^{\frac{1}{2}}(\hat{\beta}_{(m)}^* - \hat{\beta}_{(m)})$ given the data $(\tilde{T}_i, \delta_i, X_i)$ $(i=1,\ldots,n)$ converges almost surely to the asymptotic distribution of $n^{\frac{1}{2}}(\hat{\beta}_{(m)} - \beta_0)$. To approximate the distribution of $\hat{\beta}_{(m)}$, we obtain a large number of realisations of $\hat{\beta}_{(m)}^*$ by repeatedly generating the random sample $(Z_1, \ldots, Z_n)$ while fixing the data $(\tilde{T}_i, \delta_i, X_i)$ $(i=1,\ldots,n)$ at their observed values. The empirical distribution of $\hat{\beta}_{(m)}^*$ can then be used to approximate the distribution of $\hat{\beta}_{(m)}$. Confidence intervals for individual components of $\beta_0$ can be constructed by the Wald method or from the empirical percentiles of $\hat{\beta}_{(m)}^*$.

*Remark* 1. Jin et al. (2003) presented a resampling approach to the approximation of the distributions of general rank estimators by perturbing $L_1$ loss functions. Their approach is not applicable to the present setting because the function $L$ defined in (5) pertains to a least-squares estimating function with imputed failure times and cannot be expressed as a weighted Gehan estimating function. In contrast to the resampling procedure of Jin et al. (2003), $L^*$ is defined directly through perturbing the $X_i$, $Y_i$ and $\hat{F}_b(.)$, and does not have an equivalent minimisation expression. The resampling scheme needs to account properly for the sampling variations due to both the least-squares estimation and the imputation of censored failure times. The perturbation of $\hat{F}_b(.)$ is particularly intriguing and delicate. For example, if the $Z_j$ in the denominator of $\hat{F}_b^*(.)$ were omitted, then $n^{\frac{1}{2}}\{\hat{F}_{\beta_0}^*(.) - \hat{F}_{\beta_0}(.)\}$ would still provide a valid approximation to the distribution of $n^{\frac{1}{2}}\{\hat{F}_{\beta_0}(.) - F(.)\}$, but the resampling distribution for $\hat{\beta}$ would no longer be valid.

*Remark* 2. Instead of least-squares estimation, one can employ $M$-estimation, replacing the square loss in (2) by a general loss function. In fact, extensions of the Buckley–James estimator to general $M$-estimators were studied by Ritov (1990) and Lai & Ying (1994). We can again use the Gehan estimator as the initial value and develop a similar iterative algorithm and resampling scheme. All the theoretical results continued to hold.

## 4. Extensions to multivariate failure time data

### 4·1. *Multiple events data*

Multiple events data arise when a subject can potentially experience several types of event or failure. For $k = 1, \ldots, K$ and $i = 1, \ldots, n$, let $T_{ki}$ be the time to the $k$th failure of the $i$th subject, let $C_{ki}$ be the censoring time on $T_{ki}$, and let $X_{ki}$ be the corresponding

$p_k$-vector of covariates. We assume that $(T_{1i}, \ldots, T_{Ki})$ is independent of $(C_{1i}, \ldots, C_{Ki})$ conditional on $(X_{1i}, \ldots, X_{Ki})$. The data consist of $(\tilde{T}_{ki}, \delta_{ki}, X_{ki})$ $(k = 1, \ldots, K; i = 1, \ldots, n)$, where $\tilde{T}_{ki} = \min(T_{ki}, C_{ki})$ and $\delta_{ki} = I_{\{T_{ki} \leqslant C_{ki}\}}$.

The marginal accelerated failure time models take the form

$$\log T_{ki} = X'_{ki}\beta_k + \varepsilon_{ki} \quad (k = 1, \ldots, K; i = 1, \ldots, n),$$

where $\beta_k$ is a $p_k$-vector of unknown regression parameters, and $(\varepsilon_{1i}, \ldots, \varepsilon_{Ki})$ $(i = 1, \ldots, n)$ are independent random vectors from an unspecified joint distribution with marginal distribution functions $F_1, \ldots, F_K$.

Define $\tilde{Y}_{ki} = \log \tilde{T}_{ki}$, $e_{ki}(\beta) = \tilde{Y}_{ki} - X'_{ki}\beta$ and

$$\hat{Y}_{ki}(\beta) = \delta_{ki}\tilde{Y}_{ki} + (1 - \delta_{ki})\left[\frac{\int_{e_{ki}(\beta)}^{\infty} u\, d\hat{F}_{k,\beta}(u)}{1 - \hat{F}_{k,\beta}\{e_{ki}(\beta)\}} + X'_{ki}\beta\right],$$

where $1 - \hat{F}_{k,\beta}$ is the left-continuous version of the Kaplan–Meier estimator of $1 - F_k$ based on the transformed data $\{e_{ki}(\beta), \delta_{ki}\}$ $(i = 1, \ldots, n)$, which is in the form of (4). We estimate $\beta_k$ through the iterative procedure $\hat{\beta}_{k(m)} = L_k(\hat{\beta}_{k(m-1)})$ $(m \geqslant 1)$, where

$$L_k(b) = \left\{\sum_{i=1}^{n}(X_{ki} - \bar{X}_k)^{\otimes 2}\right\}^{-1}\left\{\sum_{i=1}^{n}(X_{ki} - \bar{X}_k)\hat{Y}_{ki}(b)\right\},$$

$\bar{X}_k = n^{-1}\sum_{i=1}^{n} X_{ki}$, and $\hat{\beta}_{k(0)}$ is a minimiser of

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{ki}\{e_{ki}(\beta) - e_{kj}(\beta)\}^{-}.$$

Writing $B = (\beta_1, \ldots, \beta_K)$ and $\hat{B}_{(m)} = (\hat{\beta}_{1(m)}, \ldots, \hat{\beta}_{K(m)})$, we show in the Appendix that $n^{\frac{1}{2}}(\hat{B}_{(m)} - B)$ is asymptotically zero-mean normal.

Let $\hat{\beta}^*_{k(0)}$ be a minimiser of

$$\sum_{i=1}^{n}\sum_{j=1}^{n} Z_i Z_j \delta_{ki}\{e_{ki}(\beta) - e_{kj}(\beta)\}^{-},$$

where $(Z_1, \ldots, Z_n)$ are defined in §3. Also, define $\hat{\beta}^*_{k(m)} = L^*_k(\hat{\beta}^*_{k(m-1)})$ $(m \geqslant 1)$, where

$$L^*_k(b) = \left\{\sum_{i=1}^{n} Z_i(X_{ki} - \bar{X}_k)^{\otimes 2}\right\}^{-1}\left[\sum_{i=1}^{n} Z_i(X_{ki} - \bar{X}_k)\{\hat{Y}^*_{ki}(b) - \bar{Y}^*_k(b)\}\right],$$

$$\hat{Y}^*_{ki}(b) = \delta_{ki}\tilde{Y}_{ki} + (1 - \delta_{ki})\left[\frac{\int_{e_{ki}(b)}^{\infty} u\, d\hat{F}^*_{k,b}(u)}{1 - \hat{F}^*_{k,b}\{e_{ki}(b)\}} + X'_{ki}b\right],$$

$$\hat{F}^*_{k,b}(t) = 1 - \prod_{i: e_{ki}(b) < t}\left(1 - \frac{Z_i \delta_{ki}}{\sum_{j=1}^{n} Z_j 1_{\{e_{kj}(b) \geqslant e_{ki}(b)\}}}\right),$$

and $\bar{Y}^*_k(b) = n^{-1}\sum_{i=1}^{n}\hat{Y}^*_{ki}(b)$. Writing $\hat{B}^*_{(m)} = (\hat{\beta}^*_{1(m)}, \ldots, \hat{\beta}^*_{K(m)})$, we show in the Appendix that the conditional distribution of $n^{\frac{1}{2}}(\hat{B}^*_{(m)} - \hat{B}_{(m)})$, given the data $(\tilde{T}_{ki}, \delta_{ki}, X_{ki})$ $(k = 1, \ldots, K; i = 1, \ldots, n)$, is asymptotically the same as the distribution of $n^{\frac{1}{2}}(\hat{B}_{(m)} - B)$. Thus, simultaneous inference about $B$ can be carried out on the basis of the empirical joint distribution of $\hat{B}^*_{(m)}$, which is obtained by repeatedly generating the random sample $(Z_1, \ldots, Z_n)$.

## 4·2. *Clustered failure time data*

Clustered failure time data arise when the subjects are sampled in clusters such that the failure times within the same cluster tend to be correlated. Suppose that there are $n$ clusters, the $i$th cluster having $K_i$ members. The $K_i$ are assumed to be relatively small compared to $n$. For the $k$th member of the $i$th cluster, let $T_{ik}$ and $C_{ik}$ denote the failure and censoring times, and $X_{ik}$ the corresponding $p$-vector of covariates. It is assumed that $(T_{i1}, \ldots, T_{iK_i})$ and $(C_{i1}, \ldots, C_{iK_i})$ are independent conditional on $(X_{i1}, \ldots, X_{iK_i})$. The data consist of $(\tilde{T}_{ik}, \delta_{ik}, X_{ik})$ $(k = 1, \ldots, K_i; i = 1, \ldots, n)$, where $\tilde{T}_{ik} = \min(T_{ik}, C_{ik})$ and $\delta_{ik} = I_{\{T_{ik} \leqslant C_{ik}\}}$.

Suppose that the marginal distributions of the $T_{ik}$ satisfy the accelerated failure time model

$$\log T_{ik} = X_{ik}'\beta_0 + \varepsilon_{ik} \quad (k = 1, \ldots, K_i; i = 1, \ldots, n),$$

where $\beta_0$ is a $p$-vector of unknown regression parameters, and $(\varepsilon_{i1}, \ldots, \varepsilon_{iK_i})$ $(i = 1, \ldots, n)$ are independent random vectors. For each $i$, the error terms $\varepsilon_{i1}, \ldots, \varepsilon_{iK_i}$ are assumed to be exchangeable with a common marginal distribution $F$. It is also assumed that, for any $i$ and $j$, and $K \leqslant \min(K_i, K_j)$, the vectors $(\varepsilon_{i1}, \ldots, \varepsilon_{iK})$ and $(\varepsilon_{j1}, \ldots, \varepsilon_{jK})$ have the same distribution.

Define $\tilde{Y}_{ik} = \log \tilde{T}_{ik}$, $e_{ik}(\beta) = \tilde{Y}_{ik} - X_{ik}'\beta$ and

$$\hat{Y}_{ik}(\beta) = \delta_{ik}\tilde{Y}_{ik} + (1 - \delta_{ik})\left[\frac{\int_{e_{ik}(\beta)}^{\infty} u\,d\hat{F}_\beta(u)}{1 - \hat{F}_\beta\{e_{ik}(\beta)\}} + X_{ik}'\beta\right],$$

where $1 - \hat{F}_\beta$ is the left-continuous version of the Kaplan–Meier estimator of $1 - F$ based on the transformed data $\{e_{ik}(\beta), \delta_{ik}\}$ $(k = 1, \ldots, K_i; i = 1, \ldots, n)$. We estimate $\beta_0$ iteratively using $\hat{\beta}_{(m)} = L(\hat{\beta}_{(m-1)})$ $(m \geqslant 1)$, where

$$L(b) = \left\{\sum_{i=1}^{n} \sum_{k=1}^{K_i} (X_{ik} - \bar{X})^{\otimes 2}\right\}^{-1} \left\{\sum_{i=1}^{n} \sum_{k=1}^{K_i} (X_{ik} - \bar{X})\hat{Y}_{ik}(b)\right\},$$

$\bar{X} = \sum_{i=1}^{n} \sum_{k=1}^{K_i} X_{ik} / \sum_{i=1}^{n} K_i$, and $\hat{\beta}_{(0)}$ is a minimiser of

$$\sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{j=1}^{n} \sum_{l=1}^{K_j} \delta_{ik}\{e_{ik}(\beta) - e_{jl}(\beta)\}^-,$$

which can again be obtained by linear programming.

Let $\hat{\beta}_{(0)}^*$ be a minimiser of

$$\sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{j=1}^{n} \sum_{l=1}^{K_j} Z_i Z_j \delta_{ik}\{e_{ik}(\beta) - e_{jl}(\beta)\}^-,$$

and $\hat{\beta}_{(m)}^* = L^*(\hat{\beta}_{(m-1)}^*)$ $(m \geqslant 1)$, where

$$L^*(b) = \left\{\sum_{i=1}^{n} \sum_{k=1}^{K_i} Z_i(X_{ik} - \bar{X})^{\otimes 2}\right\}^{-1} \left[\sum_{i=1}^{n} \sum_{k=1}^{K_i} Z_i(X_{ik} - \bar{X})\{\hat{Y}_{ik}^*(b) - \bar{Y}^*(b)\}\right],$$

$$\hat{Y}_{ik}^*(b) = \delta_{ik}\tilde{Y}_{ik} + (1 - \delta_{ik})\left[\frac{\int_{e_{ik}(b)}^{\infty} u\,d\hat{F}_b^*(u)}{1 - \hat{F}_b^*\{e_{ik}(b)\}} + X_{ik}'b\right],$$

$$\hat{F}_b^*(t) = 1 - \prod_{1 \leqslant i \leqslant n, 1 \leqslant k \leqslant K_i : e_{ik}(b) < t} \left(1 - \frac{Z_i \delta_{ik}}{\sum_{j=1}^{n} \sum_{l=1}^{K_j} Z_j 1_{\{e_{jl}(b) \geqslant e_{ik}(b)\}}}\right)$$

and $\bar{Y}*(b) = \sum_{i=1}^{n} \sum_{k=1}^{K_i} \hat{Y}_{ik}^*(b) / \sum_{i=1}^{n} K_i$. We show in the Appendix that, for any $m$, the random vector $n^{\frac{1}{2}}(\hat{\beta}_{(m)} - \beta_0)$ converges in distribution to a zero-mean normal random vector, whose distribution can be approximated by the conditional distribution of $n^{\frac{1}{2}}(\hat{\beta}_{(m)}^* - \hat{\beta}_{(m)})$ given the data $(\tilde{T}_{ik}, \delta_{ik}, X_{ik})$ $(k = 1, \ldots, K_i; i = 1, \ldots, n)$. Thus, we can use the empirical distribution of $\hat{\beta}_{(m)}^*$ to make inference about $\beta_0$.

## 5. Simulation studies

We conducted simulation studies to assess the performance of the proposed methods. For efficiency comparisons, we also calculated the log-rank estimator and the rank estimator with normal scores. The failure times were generated from the model

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

where $X_1$ is Bernoulli with success probability $0 \cdot 5$, $X_2$ is normal with mean 0 and standard deviation $0 \cdot 5$, and $\varepsilon$ has the standard normal, extreme-value or logistic distribution. The censoring times were generated from the $\text{Un}[0, c]$ distribution, where $c$ was chosen to yield a desired level of censoring. We considered random samples of 50, 100 and 200 subjects. We estimated $\beta_1$ and $\beta_2$ with three iterations. The resampling was based on 1000 realisations of standard exponential random variables.

The results for a sample size of 100 based on 10 000 simulated datasets are summarised in Table 1. The proposed parameter estimator is virtually unbiased. The resampling

Table 1. *Summary statistics for the simulation studies*

| | | | | | | | Rank estimators | | | | |
| | Censoring | Proposed estimator | | | | log-rank | | | normal-scores | | |
| | | Bias | SE | SEE | CP | Bias | SE | RE | Bias | SE | RE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Normal error | | | | | | |
| $\beta_1$ | 0% | 0·001 | 0·203 | 0·196 | 0·941 | 0·001 | 0·223 | 0·835 | 0·001 | 0·205 | 0·981 |
| | 25% | −0·002 | 0·223 | 0·214 | 0·938 | −0·003 | 0·244 | 0·838 | −0·002 | 0·225 | 0·982 |
| | 50% | −0·003 | 0·256 | 0·248 | 0·938 | −0·004 | 0·275 | 0·865 | −0·003 | 0·257 | 0·986 |
| $\beta_2$ | 0% | 0·001 | 0·205 | 0·194 | 0·932 | 0·000 | 0·222 | 0·847 | 0·001 | 0·206 | 0·984 |
| | 25% | 0·001 | 0·225 | 0·213 | 0·930 | 0·001 | 0·244 | 0·856 | 0·001 | 0·227 | 0·984 |
| | 50% | 0·003 | 0·257 | 0·247 | 0·937 | 0·004 | 0·273 | 0·882 | 0·004 | 0·258 | 0·991 |
| | | | | | Extreme-value error | | | | | | |
| $\beta_1$ | 0% | −0·002 | 0·262 | 0·251 | 0·942 | 0·000 | 0·206 | 1·621 | −0·001 | 0·268 | 0·956 |
| | 25% | 0·000 | 0·301 | 0·288 | 0·940 | 0·001 | 0·241 | 1·569 | 0·002 | 0·316 | 0·912 |
| | 50% | 0·009 | 0·372 | 0·365 | 0·950 | 0·008 | 0·311 | 1·428 | 0·016 | 0·404 | 0·846 |
| $\beta_2$ | 0% | −0·001 | 0·263 | 0·247 | 0·934 | −0·003 | 0·211 | 1·552 | −0·002 | 0·270 | 0·950 |
| | 25% | 0·001 | 0·302 | 0·285 | 0·935 | −0·001 | 0·245 | 1·518 | 0·001 | 0·318 | 0·905 |
| | 50% | 0·012 | 0·374 | 0·360 | 0·944 | 0·012 | 0·319 | 1·371 | 0·019 | 0·408 | 0·839 |
| | | | | | Logistic error | | | | | | |
| $\beta_1$ | 0% | 0·002 | 0·369 | 0·356 | 0·942 | 0·002 | 0·397 | 0·863 | −0·185 | 0·544 | 0·413 |
| | 25% | −0·003 | 0·391 | 0·381 | 0·944 | −0·002 | 0·412 | 0·900 | −0·024 | 0·551 | 0·501 |
| | 50% | −0·001 | 0·432 | 0·431 | 0·949 | −0·001 | 0·436 | 0·981 | 0·004 | 0·491 | 0·772 |
| $\beta_2$ | 0% | 0·002 | 0·371 | 0·351 | 0·934 | 0·000 | 0·397 | 0·875 | −0·191 | 0·542 | 0·417 |
| | 25% | 0·003 | 0·395 | 0·378 | 0·937 | 0·002 | 0·414 | 0·911 | −0·028 | 0·540 | 0·534 |
| | 50% | 0·009 | 0·435 | 0·429 | 0·948 | 0·010 | 0·437 | 0·991 | 0·017 | 0·493 | 0·778 |

Bias, bias of the parameter estimator; SE, standard error of the parameter estimator; SEE, mean of the standard error estimator; CP, coverage probability of the 95% confidence interval; RE, mean squared error of the proposed estimator over that of the rank estimator.

procedure accurately captures the variability of the parameter estimator, and the confidence intervals have proper coverage probabilities. The proposed estimator is slightly more efficient than the normal-scores rank estimator under normal error, appreciably more efficient under extreme-value error and substantially more efficient under logistic error. The proposed estimator is less efficient than the log-rank estimator under extreme-value error, but more efficient under normal and logistic errors.

Figure 1 compares the proposed estimates of $\beta_1$ after convergence with the estimates after three iterations and with the initial values based on 1000 simulated datasets under normal error with sample size of 100 and 25% censoring. In the 1000 datasets, 91% of the estimates after three iterations differ by less than 0·001 from the estimates after convergence, although the estimates are considerably different from the initial values. Similar phenomena are observed in other settings. Thus, it suffices to use a small number of iterations, three say, in practice.
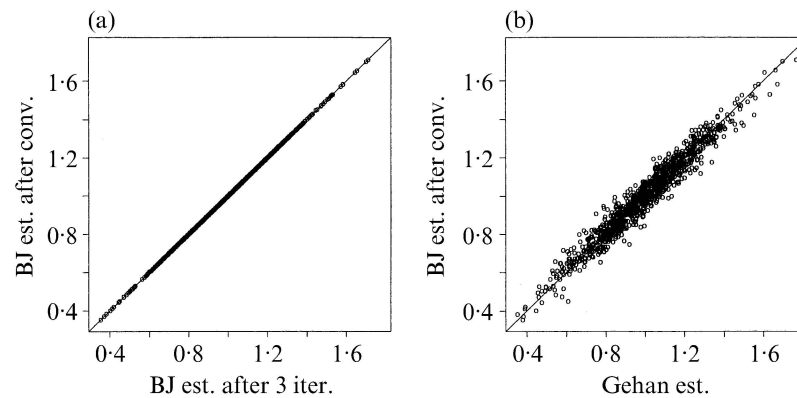


Fig. 1: Simulation studies. Comparisons of different parameter estimates: (a) Buckley–James-type estimates after convergence versus Buckley–James-type estimates after three iterations; (b) Buckley–James-type estimates after convergence versus the Gehan-type estimates.

## 6. EXAMPLES

We first present a reanalysis of the Stanford heart transplantation data (Miller & Halpern, 1982). Following Miller & Halpern (1982), we consider two models, one regressing the base-10 logarithm of the survival time on the patient's age and T5 mismatch core for the 157 patients with complete records on the T5 mismatch score, and one regressing the base-10 logarithm of the survival time on age and age$^2$ for the 152 patients who survived for at least 10 days after transplantation. We use standard exponential random variables in the resampling. The results of the analysis are shown in Table 2. The estimates after three iterations are almost identical to the final estimates. The point estimates under the proposed method are similar to those of Miller & Halpern (1982), but the estimated standard errors are considerably larger. The simple variance estimator used by Miller & Halpern is not compatible with the asymptotic variance formula given in Ritov (1990) and Lai & Ying (1991). Indeed, none of the existing variance estimators for the Buckley–James estimator has been shown to be consistent. Incidentally, the second model fits the data well whereas the first one does not (Miller & Halpern, 1982; Wei et al., 1990).

Table 2. *Accelerated failure time regression for the Stanford heart transplant data*

|  | | Gehan estimator | | Proposed estimator | | | |
|  | | | | Three iterations | | At convergence | |
|  | Covariate | Est | SE | Est | SE | Est | SE |
|---|---|---|---|---|---|---|---|
| Model 1 | Age | −0·0211 | 0·0106 | −0·0149 | 0·0098 | −0·0148 | 0·0098 |
|  | T5 | −0·0265 | 0·1507 | −0·0027 | 0·1477 | −0·0028 | 0·1477 |
| Model 2 | Age | 0·1046 | 0·0474 | 0·1070 | 0·0474 | 0·1070 | 0·0473 |
|  | $Age^2$ | −0·0017 | 0·0006 | −0·0017 | 0·0006 | −0·0017 | 0·0006 |

Est, estimate of regression parameter; SE, standard error estimate based on 10 000 resamples.

As a second example, we consider the Mayo primary biliary cirrhosis data (Fleming & Harrington, 1991, pp. 153–4). We regress the natural logarithm of the survival time on five covariates for 418 patients. The estimates of the regression parameters at three iterations are −0·0256, 1·6174, −0·5885, −0·8430 and −2·3331 for age, log(albumin), log(bilirubin), oedema and log(protime), respectively. The corresponding standard error estimates based on 10 000 resamples are 0·0063, 0·5409, 0·0752, 0·2604 and 0·8543. These results are comparable to the Gehan and log-rank estimates of Jin et al. (2003).

Finally, we consider the litter-matched tumourigenesis data reported in Mantel et al. (1977). There are 50 female litters, each with 3 rats. We regress the natural logarithm of the time to tumour appearance on the treatment indicator, which takes values 0 and 1 for the treated and untreated rats, respectively. The point estimate at three iterations is 0·1565. With 10 000 resamples, the standard error is estimated at 0·1008. The corresponding 95% Wald confidence interval is (−0·0411, 0·3541). These results differ slightly from those of Lee et al. (1993).

## 7. REMARKS

As a result of its direct physical interpretation, the accelerated failure time model provides an attractive alternative to the popular proportional hazards model (Cox, 1972) for the regression analysis of censored data, especially when the response variable does not pertain to a failure time. The least-squares estimation is a natural approach to the analysis of this model, but is hindered by the presence of censoring. The inference procedures developed in the present paper represent a practical way of implementing the least-squares principle with censored data.

For uncensored data, the rank test with normal scores is as efficient as the $t$-test at the normal distribution in the sense of Pitman efficiency and the asymptotic relative efficiency is no less than 1 at any symmetric distribution, e.g. Hettmansperger (1991, pp. 110–2). These asymptotic results may not apply to the estimation of regression parameters with censored data in small samples. In our simulation studies, we have always found the proposed estimator to be more efficient than the rank estimator with normal scores. It would be worthwhile to conduct further investigations.

Jin et al. (2003) described inference procedures based on the rank estimators. Computationally, it is less time-consuming to calculate the proposed estimator than the rank estimators except for the Gehan estimator because the former involves linear programming at the initial step only whereas the latter require linear programming at

each step of the iteration. One could use a rank estimator other than the Gehan estimator as the initial value in the proposed iterative scheme, although the computation would be more intensive.

Lin & Wei (1992a, b) proposed inference procedures for the Buckley–James estimator based on minimum-dispersion statistics (Wei et al., 1990). The implementation of these statistics requires minimisation of discrete objective functions with possibly multiple minima, for which no reliable algorithm is available. Furthermore, this approach does not provide variance estimates.

In §4, the correlation structure of multivariate failure times is taken into account in the covariance estimation, but not in the construction of the estimators. One possible approach to incorporating the correlation structure into the parameter estimation is to mimic the weighted least-squares estimators for uncensored multivariate normal responses. It would be worthwhile to investigate the efficiency gain of such procedures.

## Appendix
### *Proofs of asymptotic results*

We first consider univariate failure time data. Let $\mathscr{F}$ denote the $\sigma$-field generated by the original data $(\tilde{T}_i, \delta_i, X_i)$ $(i = 1, \ldots, n)$. Assume that the regularity conditions described in Lai & Ying (1991, p. 1376) hold and that their tail modification is used in the construction of the estimating function. Define

$$A = \int_{-\infty}^{\infty} \left[ \{\Gamma_2(t) - \Gamma_0^{-1}(t)\Gamma_1^{\otimes 2}(t)\} \int_t^{\infty} \{1 - F(s)\} ds \right] d\lambda(t),$$

where

$$\Gamma_j(t) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^{\otimes j} \, \mathrm{pr}(C_i - X_i'\beta \geqslant t | X_i) \quad (j = 0, 1, 2).$$

Write $U(b) = U(b, b)$. For $m \geqslant 1$,

$$\hat{\beta}_{(m)} = \hat{\beta}_{(m-1)} + \left\{ \sum_{i=1}^{n} (X_i - \bar{X})^{\otimes 2} \right\}^{-1} U(\hat{\beta}_{(m-1)}). \tag{A1}$$

It follows from the arguments of Lai & Ying (1991) that, uniformly in $\|b - \beta_0\| \leqslant n^{-1/3}$,

$$U(b) = U(\beta_0) - nA(b - \beta_0) + o(n^{1/2} + n\|b - \beta_0\|) \tag{A2}$$

almost surely. Thus,

$$\hat{\beta}_{(m)} - \beta_0 = (I - D^{-1}A)(\hat{\beta}_{(m-1)} - \beta_0) + n^{-1} D^{-1} U(\beta_0) + o(n^{-1/2} + \|\hat{\beta}_{(m-1)} - \beta_0\|).$$

Since $\hat{\beta}_{(0)} = \hat{\beta}_G$,

$$\hat{\beta}_{(1)} - \beta_0 = (I - D^{-1}A)(\hat{\beta}_G - \beta_0) + n^{-1} D^{-1} U(\beta_0) + o(n^{-1/2} + \|\hat{\beta}_G - \beta_0\|).$$

By induction,

$$\hat{\beta}_{(m)} - \beta_0 = (I - D^{-1}A)^m (\hat{\beta}_G - \beta_0) + n^{-1}\{I - (I - D^{-1}A)^m\}A^{-1}U(\beta_0)$$

$$+ o\left(n^{-1/2} + \sum_{j=0}^{m-1} \|\hat{\beta}_{(j)} - \beta_0\|\right) \quad (m \geqslant 1). \tag{A3}$$

It follows from (A3) that $\hat{\beta}_{(m)}$ is consistent for every fixed $m$.

Note that

$$U(\beta_0) = \sum_{i=1}^{n} (X_i - \bar{X})\left[\delta_i e_i(\beta_0) + (1 - \delta_i)\frac{\int_{e_i(\beta_0)}^{\infty} u d\hat{F}_{\beta_0}(u)}{1 - \hat{F}_{\beta_0}\{e_i(\beta_0)\}}\right]$$

$$= \sum_{i=1}^{n} (X_i - \bar{X})\left[\delta_i e_i(\beta_0) + (1 - \delta_i)\frac{\int_{e_i(\beta_0)}^{\infty} u dF(u)}{1 - F\{e_i(\beta_0)\}}\right]$$

$$+ \sum_{i=1}^{n} (X_i - \bar{X})(1 - \delta_i)\left[\frac{\int_{e_i(\beta_0)}^{\infty} u d\hat{F}_{\beta_0}(u)}{1 - \hat{F}_{\beta_0}\{e_i(\beta_0)\}} - \frac{\int_{e_i(\beta_0)}^{\infty} u dF(u)}{1 - F\{e_i(\beta_0)\}}\right].$$

Through integration by parts, we can establish the equality

$$\delta_i e_i(\beta_0) + (1 - \delta_i)\frac{\int_{e_i(\beta_0)}^{\infty} u dF(u)}{1 - F\{e_i(\beta_0)\}} = E\varepsilon_i + \int_{-\infty}^{\infty}\left\{t - \frac{\int_t^{\infty} u dF(u)}{1 - F(t)}\right\}dM_i(t), \tag{A4}$$

where

$$M_i(t) = \delta_i I\{e_i(\beta_0) \leqslant t\} - \int_0^t I\{e_i(\beta_0) \geqslant u\}\lambda(u)du.$$

In addition, it follows from the martingale representation of the Kaplan–Meier estimator (Fleming & Harrington, 1991, p. 98) that, almost surely,

$$\frac{\int_{e_i(\beta_0)}^{\infty} u d\hat{F}_{\beta_0}(u)}{1 - \hat{F}_{\beta_0}\{e_i(\beta_0)\}} - \frac{\int_{e_i(\beta_0)}^{\infty} u dF(u)}{1 - F\{e_i(\beta_0)\}} = n^{-1}\sum_{j=1}^{n}\int_{-\infty}^{\infty}\xi_i(t)dM_j(t) + o(n^{-1/2})$$

for some random process $\xi_i(t)$. Thus,

$$U(\beta_0) = \sum_{i=1}^{n}\int_{-\infty}^{\infty}\left[(X_i - \bar{X})\left\{t - \frac{\int_t^{\infty} u dF(u)}{1 - F(t)}\right\} + \bar{\xi}^x(t)\right]dM_i(t) + o(n^{1/2}), \tag{A5}$$

where $\bar{\xi}^x(t)$ is the limit of $n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})(1 - \delta_i)\xi_i(t)$.

As shown by Jin et al. (2005),

$$\hat{\beta}_G - \beta_0 = (nA_G)^{-1}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\eta_i(t)dM_i(t) + o(n^{-1/2} + \|\hat{\beta}_G - \beta_0\|), \tag{A6}$$

where $A_G$ is a positive definite matrix and the $\eta_i(t)$ are nonrandom functions. We conclude from (A3), (A5) and (A6) that, for each fixed $m$,

$$\hat{\beta}_{(m)} - \beta_0 = n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\left((I - D^{-1}A)^m A_G^{-1}\eta_i(t) + \{I - (I - D^{-1}A)^m\}A^{-1}\right.$$

$$\left. \times \left[(X_i - \bar{X})\left\{t - \frac{\int_t^{\infty} u dF(u)}{1 - F(t)}\right\} + \bar{\xi}^x(t)\right]\right)dM_i(t)$$

$$+ o\left(n^{-1/2} + \sum_{j=0}^{m-1}\|\hat{\beta}_{(j)} - \beta_0\|\right) \tag{A7}$$

almost surely. Thus, $\hat{\beta}_{(m)}$ is asymptotically normal as $n \to \infty$.

Our next task is to show that the conditional distribution of $n^{1/2}(\hat{\beta}^*_{(m)} - \hat{\beta}_{(m)})$ given $\mathscr{F}$ converges almost surely to the limiting distribution of $n^{1/2}(\hat{\beta}_{(m)} - \beta_0)$. For $m \geq 1$,

$$\hat{\beta}^*_{(m)} = \hat{\beta}^*_{(m-1)} + \left\{ \sum_{i=1}^{n} Z_i (X_i - \bar{X})^{\otimes 2} \right\}^{-1} U^*(\hat{\beta}^*_{(m-1)}), \qquad (A8)$$

where

$$U^*(b) = \sum_{i=1}^{n} Z_i (X_i - \bar{X}) \{ \hat{Y}^*_i(b) - \bar{Y}^*(b) - (X_i - \bar{X})'b \}.$$

It follows from (A8) that $\hat{\beta}^*_{(m)}$ is consistent for every $m$. By incorporating the random weights $Z_i$ into the derivation of the asymptotic linearity in Lai & Ying (1991), we can establish the following linear approximation analogous to (A2):

$$U^*(\hat{\beta}^*_{(m)}) = U^*(\hat{\beta}_{(m)}) - nA(\hat{\beta}^*_{(m)} - \hat{\beta}_{(m)}) + o(n^{1/2} + n\|\hat{\beta}^*_{(m)} - \beta_0\| + n\|\hat{\beta}_{(m)} - \beta_0\|) \qquad (A9)$$

almost surely. Note that the slope matrix $A$ remains the same as in (A2), since $EZ_i = 1$ and $A$ is the limit of the gradient of $n^{-1}EU^*(b)$ at $b = \beta_0$. Clearly,

$$U^*(\hat{\beta}_{(m)}) - U(\hat{\beta}_{(m)}) = \sum_{i=1}^{n} (Z_i - 1)(X_i - \bar{X}) \{ \hat{Y}^*_i(\hat{\beta}_{(m)}) - \bar{Y}^*(\hat{\beta}_{(m)}) - (X_i - \bar{X})'\hat{\beta}_{(m)} \}$$

$$+ \sum_{i=1}^{n} (X_i - \bar{X}) \{ \hat{Y}^*_i(\hat{\beta}_{(m)}) - X_i'\hat{\beta}_{(m)} \} - U(\hat{\beta}_{(m)}). \qquad (A10)$$

Since $E(Z_i - 1|\mathscr{F}) = 0$ and $\hat{Y}^*_i(\hat{\beta}_{(m)}) - X_i'\hat{\beta}_{(m)}$ can be approximated by the left-hand side of (A4), we have

$$\sum_{i=1}^{n} (Z_i - 1)(X_i - \bar{X}) \{ \hat{Y}^*_i(\hat{\beta}_{(m)}) - \bar{Y}^*(\hat{\beta}_{(m)}) - (X_i - \bar{X})'\hat{\beta}_{(m)} \}$$

$$= \sum_{i=1}^{n} (Z_i - 1) \int_{-\infty}^{\infty} (X_i - \bar{X}) \left\{ t - \frac{\int_t^{\infty} u\, dF(u)}{1 - F(t)} \right\} dM_i(t) + o(n^{1/2}). \quad (A11)$$

By approximating $\hat{F}^*_{\hat{\beta}_{(m)}} - \hat{F}_{\hat{\beta}_{(m)}}$ with a weighted sum of $Z_i - 1$, we can show that

$$\sum_{i=1}^{n} (X_i - \bar{X}) \{ \hat{Y}^*_i(\hat{\beta}_{(m)}) - X_i'\hat{\beta}_{(m)} \} - U(\hat{\beta}_{(m)})$$

$$= \sum_{i=1}^{n} (X_i - \bar{X})(1 - \delta_i) \left[ \frac{\int_{e_i(\hat{\beta}_{(m)})}^{\infty} u\, d\hat{F}^*_{\hat{\beta}_{(m)}}(u)}{1 - \hat{F}^*_{\hat{\beta}_{(m)}} \{e_i(\hat{\beta}_{(m)})\}} - \frac{\int_{e_i(\hat{\beta}_{(m)})}^{\infty} u\, d\hat{F}_{\hat{\beta}_{(m)}}(u)}{1 - \hat{F}_{\hat{\beta}_{(m)}} \{e_i(\hat{\beta}_{(m)})\}} \right]$$

$$= \sum_{i=1}^{n} (Z_i - 1) \int_{-\infty}^{\infty} \bar{\xi}^x(t)\, dM_i(t) + o(n^{1/2}). \qquad (A12)$$

By plugging (A11) and (A12) into the right-hand side of (A10) and then plugging the resulting expression into (A9), we obtain

$$U^*(\hat{\beta}^*_{(m)}) = U(\hat{\beta}_{(m)}) + \sum_{i=1}^{n} (Z_i - 1) \int_{-\infty}^{\infty} \left[ (X_i - \bar{X}) \left\{ t - \frac{\int_t^{\infty} u\, dF(u)}{1 - F(t)} \right\} + \bar{\xi}^x(t) \right] dM_i(t)$$

$$- nA(\hat{\beta}^*_{(m)} - \hat{\beta}_{(m)}) + o(n^{1/2} + n\|\hat{\beta}^*_{(m)} - \beta_0\| + n\|\hat{\beta}_{(m)} - \beta_0\|). \qquad (A13)$$

Taking the difference of (A1) and (A8) and making use of (A13) for $U^*(\hat{\beta}^*_{(m-1)})$, we can show that

$$\hat{\beta}^*_{(m)} - \hat{\beta}_{(m)} = n^{-1}D^{-1}\sum_{i=1}^{n}(Z_i - 1)\int_{-\infty}^{\infty}\left[(X_i - \bar{X})\left\{t - \frac{\int_t^{\infty} udF(u)}{1 - F(t)}\right\} + \bar{\xi}^x(t)\right]dM_i(t)$$

$$+ (I - D^{-1}A)(\hat{\beta}^*_{(m-1)} - \hat{\beta}_{(m-1)})$$

$$+ o(n^{-1/2} + \|\hat{\beta}^*_{(m-1)} - \beta_0\| + \|\hat{\beta}_{(m-1)} - \beta_0\|).$$

By induction,

$$\hat{\beta}^*_{(m)} - \hat{\beta}_{(m)} = (I - D^{-1}A)^m(\hat{\beta}^*_G - \hat{\beta}_G) + n^{-1}\{I - (I - D^{-1}A)^m\}A^{-1}$$

$$\times \sum_{i=1}^{n}(Z_i - 1)\int_{-\infty}^{\infty}\left[(X_i - \bar{X})\left\{t - \frac{\int_t^{\infty} udF(u)}{1 - F(t)}\right\} + \bar{\zeta}^x(t)\right]dM_i(t)$$

$$+ o\left(n^{-1/2} + \sum_{j=0}^{m-1}\|\hat{\beta}^*_{(j)} - \beta_0\| + \sum_{j=0}^{m-1}\|\hat{\beta}_{(j)} - \beta_0\|\right).$$

As shown by Jin et al. (2005),

$$\hat{\beta}^*_G - \hat{\beta}_G = (nA_G)^{-1}\sum_{i=1}^{n}(Z_i - 1)\int_{-\infty}^{\infty}\eta_i(t)dM_i(t) + o(n^{-1/2} + \|\hat{\beta}^*_G - \beta_0\| + \|\hat{\beta}_G - \beta_0\|).$$

Thus,

$$\hat{\beta}^*_{(m)} - \hat{\beta}_{(m)} = n^{-1}\sum_{i=1}^{n}(Z_i - 1)\int_{-\infty}^{\infty}\left((I - D^{-1}A)^m A_G^{-1}\eta_i(t) + \{I - (I - D^{-1}A)^m\}A^{-1}\right.$$

$$\left. \times \left[(X_i - \bar{X})\left\{t - \frac{\int_t^{\infty} udF(u)}{1 - F(t)}\right\} + \bar{\xi}^x(t)\right]\right)dM_i(t)$$

$$+ o\left(n^{-1/2} + \sum_{j=0}^{m-1}\|\hat{\beta}^*_{(j)} - \beta_0\| + \sum_{j=0}^{m-1}\|\hat{\beta}_{(j)} - \beta_0\|\right). \tag{A14}$$

Comparing (A14) with (A7), we see that the conditional distribution of $n^{1/2}(\hat{\beta}^*_{(m)} - \hat{\beta}_{(m)})$ given $\mathscr{F}$ converges almost surely to the limiting distribution of $n^{1/2}(\hat{\beta}_{(m)} - \beta_0)$.

For multiple events data, equations of the forms of (A7) and (A14) hold for each of the $K$ types of failure. It then follows from the multivariate central limit theorem that $n^{1/2}(\hat{B}_{(m)} - B)$ is asymptotically zero-mean normal and that the conditional distribution of $n^{1/2}(\hat{B}^*_{(m)} - \hat{B}_{(m)})$ given the data $(\tilde{T}_{ki}, \delta_{ki}, X_{ki})$ $(k = 1, \ldots, K; i = 1, \ldots, n)$ converges almost surely to the limiting distribution of $n^{1/2}(\hat{B}_{(m)} - B)$. For clustered failure time data, the proofs are essentially the same as those of univariate failure time data except that we need to use the properties of the Kaplan–Meier estimator for dependent observations (Ying & Wei, 1994) and to replace the martingale central limit theorem by the central limit theorem for empirical processes. The details are omitted.

## References

Buckley, J. & James, 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.

Cox, D. R. & Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.

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

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily single-censored samples. *Biometrika* **52**, 203–23.

Hettmansperger, T. P. (1991). *Statistical Inference Based on Ranks*. Malabar: Krieger.

Jin, Z., Lin, D. Y., Wei, L. J. & Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–53.

JIN, Z, LIN, D. Y. & YING, Z. (2005). Rank regression analysis of multivariate failure time data based on marginal linear models. *Scand. J. Statist.* To appear.

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

KOUL, H., SUSARLA, V. & VAN RYZIN, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9**, 1276–88.

LAI, T. L. & YING, Z. (1991). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Ann. Statist.* **10**, 1370–402.

LAI, T. L. & YING, Z. (1994). A missing information principle and M-estimators in regression analysis with censored and truncated data. *Ann. Statist.* **21**, 1222–55.

LEE, E. W., WEI, L. J. & YING, Z. (1993). Linear regression analysis for highly stratified failure time data. *J. Am. Statist. Assoc.* **88**, 557–65.

LIN, J. S. & WEI, L. J. (1992a). Linear regression analysis based on Buckley-James estimating equations. *Biometrics* **48**, 679–81.

LIN, J. S. & WEI, L. J. (1992b). Linear regression analysis for multivariate failure time observations. *J. Am. Statist. Assoc.* **87**, 1091–7.

MANTEL, N., BOHIDAR, N. R. & CIMINERA, J. L. (1977). A Mantel-Haenszel analysis of litter-matched time-to-response data, with modifications for recovery of interlitter information. *Cancer Res.* **37**, 3863–8.

MILLER, R. G. (1976). Least squares regression with censored data. *Biometrika* **63**, 449–64.

MILLER, R. G. & HALPERN, J. (1982). Regression with censored data. *Biometrika* **69**, 521–31.

PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65**, 167–79.

RITOV, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18**, 303–28.

TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354–72.

WEI, L. J., YING, Z. & LIN, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845–51.

YING, Z. & WEI, L. J. (1994). The Kaplan-Meier estimate for dependent failure time observations. *J. Mult. Anal.* **50**, 17–29.