Biometrika (2003), **90**, 2, *pp.* 341–353

Printed in Great Britain

Rank-based inference for the accelerated failure time model

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, University of North Carolina, Chapel Hill, North Carolina 27599-7420, U.S.A.

lin@bios.unc.edu

L. J. WEI

Department of Biostatistics, Harvard University, Boston, Massachusetts 02115, U.S.A.

wei@sdac.harvard.edu

and ZHILIANG YING

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

zying@stat.columbia.edu

Summary

A broad class of rank-based monotone estimating functions is developed for the semiparametric accelerated failure time model with censored observations. The corresponding estimators can be obtained via linear programming, and are shown to be consistent and asymptotically normal. The limiting covariance matrices can be estimated by a resampling technique, which does not involve nonparametric density estimation or numerical derivatives. The new estimators represent consistent roots of the non-monotone estimating equations based on the familiar weighted log-rank statistics. Simulation studies demonstrate that the proposed methods perform well in practical settings. Two real examples are provided.

Some key words: Accelerated life model; Censoring; Gehan statistic; Linear programming; Rank estimator; Survival data; Weighted log-rank statistic.

1. Introduction

The accelerated failure time model or accelerated life model relates the logarithm of the failure time linearly to the covariates (Kalbfleisch & Prentice, 1980, pp. 32–4; Cox & Oakes, 1984, pp. 64–5). As a result of its direct physical interpretation, this model provides an attractive alternative to the popular Cox (1972) proportional hazards model for the regression analysis of censored failure time data.

The presence of censoring in failure time data creates a serious challenge in the semiparametric analysis of the accelerated failure time model. Several semiparametric estimators were proposed around 1980, Buckley & James (1979) providing a modification of the least-squares estimator to accommodate censoring and Prentice (1978) proposing rank estimators based on the well-known weighted log-rank statistics. The asymptotic properties of the Buckley–James and rank estimators were rigorously studied by Ritov (1990), Tsiatis (1990), Lai & Ying (1991a, b) and Ying (1993) among others.

Despite the theoretical advances, semiparametric methods for the accelerated failure time model have rarely been used in applications, mainly because of the lack of efficient and reliable computational methods. The existing semiparametric estimating functions are step functions of the regression parameters with potentially multiple roots, and the corresponding estimators may not be well defined. Furthermore, analytical evaluations of the covariance matrices of the estimators would require nonparametric density function estimation. To bypass the variance-covariance estimation, Wei et al. (1990) developed an inference procedure based on the so-called minimum-dispersion statistic. Calculation of the minimum-dispersion statistics and of the semiparametric estimators involves minimisations of discrete objective functions with potentially multiple local minima. Such minimisation problems cannot be solved by conventional optimisation algorithms. Lin & Geyer (1992) suggested a computational method based on simulated annealing (Kirkpatrick et al., 1983), but their algorithm is not guaranteed to yield the true minimum.

In the present paper, we provide simple and reliable methods for implementing the aforementioned rank estimators. We first show that the rank estimator with the Gehan (1965)-type weight function can be readily obtained by minimising a convex objective function through a standard linear programming technique. We then introduce a class of monotone estimating functions to approximate the possibly non-monotone weighted log-rank estimating functions around the true values of the regression parameters. These estimating equations are solved through an iterative algorithm with the Gehan-type estimator as the initial value. Each iteration can also be executed via linear programming. This procedure yields a consistent root of the weighted log-rank estimating equation, which potentially contains inconsistent roots. The covariance matrices for both the Gehan-type and iterative estimators can be easily estimated by a resampling technique, which does not involve density estimation.

2. Proposed methods

2·1. *Accelerated failure time model and rank estimators*

For $i=1, \ldots, n$, let T_i be the failure time for the ith subject and let X_i be the associated p-vector of covariates. The accelerated failure time model specifies that

$$\log T_i = \beta_0' X_i + \varepsilon_i \quad (i=1, \ldots, n), \tag{2·1}$$

where β_0 is a p-vector of unknown regression parameters and ε_i $(i=1, \ldots, n)$ are independent error terms with a common, but completely unspecified, distribution.

Let C_i be the censoring time for T_i. Assume that T_i and C_i are independent conditionally on X_i. We impose Conditions 1–4 of Ying (1993, p. 80). The data consist of $(\tilde{T}_i, \Delta_i, X_i)$ $(i=1, \ldots, n)$, where $\tilde{T}_i = T_i \wedge C_i$ and $\Delta_i = 1_{\{T_i \leqslant C_i\}}$. Here and in the sequel, $a \wedge b = \min(a, b)$, and $1_{\{.\}}$ is the indicator function.

Define $e_i(\beta) = \log \tilde{T}_i - \beta' X_i$, $N_i(\beta; t) = \Delta_i 1_{\{e_i(\beta) \leqslant t\}}$ and $Y_i(\beta; t) = 1_{\{e_i(\beta) \geqslant t\}}$. Note that N_i and Y_i are the counting process and at-risk process on the time scale of the residual. Write

$$S^{(0)}(\beta; t) = n^{-1} \sum_{i=1}^{n} Y_i(\beta; t), \quad S^{(1)}(\beta; t) = n^{-1} \sum_{i=1}^{n} Y_i(\beta; t) X_i.$$

The weighted log-rank estimating function for β_0 takes the form

$$U_\phi(\beta) = \sum_{i=1}^{n} \Delta_i \phi\{\beta;\, e_i(\beta)\}[X_i - \bar{X}\{\beta;\, e_i(\beta)\}],$$

or

$$U_\phi(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(\beta;\, t)\{X_i - \bar{X}(\beta;\, t)\}\, dN_i(\beta;\, t), \tag{2·2}$$

where $\bar{X}(\beta;\, t) = S^{(1)}(\beta;\, t)/S^{(0)}(\beta;\, t)$, and ϕ is a possibly data-dependent weight function satisfying Condition 5 of Ying (1993, p. 90). The choices of $\phi = 1$ and $\phi = S^{(0)}$ correspond to the log-rank (Mantel, 1966) and Gehan (1965) statistics, respectively.

Let $\hat{\beta}_\phi$ be a root of the estimating function $U_\phi(\beta)$. It has been established that the random vector $n^{\frac{1}{2}}(\hat{\beta}_\phi - \beta_0)$ is asymptotically zero-mean normal with covariance matrix $A_\phi^{-1} B_\phi A_\phi^{-1}$, where

$$A_\phi = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(\beta_0;\, t)\{X_i - \bar{X}(\beta_0;\, t)\}^{\otimes 2}\{\dot{\lambda}(t)/\lambda(t)\}\, dN_i(\beta_0;\, t),$$

$$B_\phi = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi^2(\beta_0;\, t)\{X_i - \bar{X}(\beta_0;\, t)\}^{\otimes 2}\, dN_i(\beta_0;\, t),$$

$\lambda(.)$ is the common hazard function of the error terms, and $\dot{\lambda}(t) = d\lambda(t)/dt$ (Tsiatis, 1990; Lai & Ying, 1991b; Ying, 1993). The foregoing result requires that A_ϕ be nonsingular. This condition is assumed to hold in the present paper.

In general, $U_\phi(\beta)$ is neither continuous nor componentwise monotone in β. Thus, it is difficult to solve the equation $U_\phi(\beta) = 0$, especially when β is high-dimensional. In fact, there are potentially multiple solutions to the equation, some of which are inconsistent. Although one may define the estimator to be a minimiser of the norm of $U_\phi(\beta)$, it is not easy to locate a minimiser either, because of the discontinuity and non-monotonicity. Another difficulty lies in the variance estimation. Since A_ϕ involves the derivative of the hazard function, direct evaluation of the covariance matrix would require nonparametric estimation of the density function, which cannot be done reliably in practical samples with censored observations. One might estimate A_ϕ by the numerical derivative of $n^{-1}U_\phi$; however, this kind of estimator can be highly unstable.

2·2. *Gehan-type weight function*

Considerable simplification arises in the special case of $\phi(\beta;\, t) = S^{(0)}(\beta;\, t)$, which is referred to as the Gehan-type weight function. In this case, U_ϕ can be written as

$$U_G(\beta) = \sum_{i=1}^{n} \Delta_i S^{(0)}\{\beta;\, e_i(\beta)\}[X_i - \bar{X}\{\beta;\, e_i(\beta)\}],$$

or

$$U_G(\beta) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_i (X_i - X_j) 1_{\{e_i(\beta) \leqslant e_j(\beta)\}}, \tag{2·3}$$

which is monotone in each component of β (Fygenson & Ritov, 1994). It is not difficult to see that $(X_i - X_j)1_{\{e_i(\beta)\leqslant e_j(\beta)\}}$ is the gradient in β of $\{e_i(\beta) - e_j(\beta)\}^-$, where

$a^- = |a| 1_{\{a<0\}}$. Thus, the right-hand side of (2·3) is the gradient of the convex function

$$L_G(\beta) := n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_i \{e_i(\beta) - e_j(\beta)\}^-.$$

We denote a minimiser of $L_G(\beta)$ by $\hat{\beta}_G$. Although $\hat{\beta}_G$ is not necessarily unique, all the minimisers of $L_G(\beta)$ are asymptotically equivalent.

The minimisation of $L_G(\beta)$ can be carried out by linear programming: we minimise the linear function $\sum_{i=1}^n \sum_{j=1}^n \Delta_i u_{ij}$ subject to the linear constraints $u_{ij} \geqslant 0$ and

$$u_{ij} \geqslant -\{e_i(\beta) - e_j(\beta)\} \quad (i, j = 1, \ldots, n).$$

This type of linear programming has been used extensively in the econometrics literature, especially in connection with the regression quantiles introduced by Koenker & Bassett (1978), and was previously considered by Lin et al. (1998) for the rank estimation of the accelerated time model for counting processes. The minimisation of $L_G(\beta)$ is equivalent to the minimisation of

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_i |e_i(\beta) - e_j(\beta)| + \left| M - \beta' \sum_{k=1}^{n} \sum_{l=1}^{n} \Delta_k (X_l - X_k) \right|,$$

where M is an extremely large number. The latter type of minimisation can be implemented with the algorithm of Koenker & D'Orey (1987), which is available in S-Plus and other software.

According to the general asymptotic theory for the rank estimators stated in § 2·1, the random vector $n^{\frac{1}{2}}(\hat{\beta}_G - \beta_0)$ is asymptotically zero-mean normal with covariance matrix $A_G^{-1} B_G A_G^{-1}$, where A_G and B_G are A_ϕ and B_ϕ evaluated at $\phi = G$. As discussed before, A_G involves the unknown hazard function, so that it is difficult to estimate the covariance matrix analytically. We will develop a resampling scheme similar to those of Rao & Zhao (1992), Parzen et al. (1994) and Jin et al. (2001) to approximate the distribution of $\hat{\beta}_G$, particularly its variance-covariance matrix.

To be specific, we define a new loss function

$$L_G^*(\beta) := n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_i \{e_i(\beta) - e_j(\beta)\}^- Z_i,$$

where Z_i $(i = 1, \ldots, n)$ are independent positive random variables with $E(Z_i) = \mathrm{var}(Z_i) = 1$, and are independent of the data $(\tilde{T}_i, \Delta_i, X_i)$ $(i = 1, \ldots, n)$. This loss function is essentially a perturbed version of the original $L_G(\beta)$, perturbed by random variables Z_i $(i = 1, \ldots, n)$. Minimisation of L_G^* is also a linear programming problem. Let $\hat{\beta}_G^*$ be a minimiser of L_G^*, that is a root of the estimating function

$$U_G^*(\beta) := \sum_{i=1}^{n} \Delta_i S^{(0)}\{\beta; e_i(\beta)\}[X_i - \bar{X}\{\beta; e_i(\beta)\}] Z_i.$$

Note that U_G^* has the same mean and approximately the same variance as U_G. We show in the Appendix that the asymptotic distribution of $n^{\frac{1}{2}}(\hat{\beta}_G - \beta_0)$ can be approximated by the conditional distribution of $n^{\frac{1}{2}}(\hat{\beta}_G^* - \hat{\beta}_G)$ given the data $(\tilde{T}_i, \Delta_i, X_i)$ $(i = 1, \ldots, n)$.

Conditional on the data $(\tilde{T}_i, \Delta_i, X_i)$ $(i = 1, \ldots, n)$, the only random elements in $L_G^*(\beta)$ are the Z_i's. To approximate the distribution of $\hat{\beta}_G$, we produce a large number of realisations of $\hat{\beta}_G^*$ by repeatedly generating the random sample $(Z_1, \ldots, Z_n)$ while holding the data $(\tilde{T}_i, \Delta_i, X_i)$ $(i = 1, \ldots, n)$ at their observed values. The covariance matrix of $\hat{\beta}_G$ can then be approximated by the empirical covariance matrix of $\hat{\beta}_G^*$. Confidence intervals for individual components of β_0 can be obtained from the percentiles of the empirical distribution of $\hat{\beta}_G^*$ or by the Wald method.

2·3. *General weight functions*

The efficiency of the Gehan statistic and the corresponding rank estimator $\hat{\beta}_G$ depends on the censoring distribution. The variance of $\hat{\beta}_\phi$ is minimised if the limit of $\phi(t)$ is proportional to $\dot{\lambda}(t)/\lambda(t)$. Thus, it is desirable to use general weight functions. The simple conversion of the Gehan estimating function U_G to an L_1-type loss function does not extend directly to the general weighted log-rank estimating functions. On the basis of $\hat{\beta}_G$, however, it is possible to construct weighted Gehan-type loss functions, whose minimisers can be readily obtained by the linear programming algorithm.

We consider the following modification of (2·2):

$$\tilde{U}_\phi(\beta; \hat{\beta}) := \sum_{i=1}^{n} \int_{-\infty}^{\infty} \psi\{\hat{\beta}; t + (\beta - \hat{\beta})'X_i\} S^{(0)}(\beta; t)\{X_i - \bar{X}(\beta; t)\}\, dN_i(\beta; t), \qquad (2{\cdot}4)$$

where $\psi(b; x) = \phi(b; x)/S^{(0)}(b; x)$, and $\hat{\beta}$ is a preliminary consistent estimator of β_0, $\hat{\beta}_G$ say. Clearly,

$$\begin{aligned}\tilde{U}_\phi(\beta; \hat{\beta}) &= \sum_{i=1}^{n} \Delta_i \psi\{\hat{\beta}; e_i(\beta) + (\beta - \hat{\beta})'X_i\} S^{(0)}\{\beta; e_i(\beta)\}[X_i - \bar{X}\{\beta; e_i(\beta)\}] \\ &= n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi\{\hat{\beta}; e_i(\hat{\beta})\} \Delta_i (X_i - X_j) 1_{\{e_i(\beta) \leqslant e_j(\beta)\}}.\end{aligned}$$

This is the same as (2·3) except for the weights $\psi\{\hat{\beta}; e_i(\hat{\beta})\}$, which are free of β. Thus, $\tilde{U}_\phi(\beta; \hat{\beta})$ is monotone in each component of β and is the gradient of the following extension of $L_G(\beta)$:

$$L_\phi(\beta; \hat{\beta}) := n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi\{\hat{\beta}; e_i(\hat{\beta})\} \Delta_i \{e_i(\beta) - e_j(\beta)\}^-.$$

As in the case of L_G, minimisation of $L_\phi(.; \hat{\beta})$ can be implemented via linear programming.

We propose the following iterative algorithm for estimating β_0:

$$\hat{\beta}_{(0)} = \hat{\beta}_G, \quad \hat{\beta}_{(k)} = \arg\min_\beta L_\phi(\beta; \hat{\beta}_{(k-1)}) \quad (k \geqslant 1).$$

By comparing (2·2) and (2·4), we see that, if $\hat{\beta}_{(k)}$ converges to a limit as the number of iterations $k \to \infty$, then the limit must satisfy $U_\phi(\beta) = 0$. In such situations, the proposed algorithm constitutes a numerical method for solving the original estimating equation $U_\phi(\beta) = 0$. The resulting estimator is consistent although the original estimating equation may contain inconsistent roots. In all the simulated and real datasets we have tested, $\hat{\beta}_{(k)}$ converges as k increases and, for k as small as 3, $\hat{\beta}_{(k)}$ is already close to the limit. It is important to point out that $\hat{\beta}_{(k)}$ itself is a legitimate estimator for any k and the convergence of the algorithm is not essential to the validity of the proposed method.

We show in the Appendix that, for any k, the estimator $\hat{\beta}_{(k)}$ is consistent and asymptotically normal. In fact, it follows from (A·5) in the Appendix that

$$\hat{\beta}_{(k)} = \{(A_\phi + D_\phi)^{-1} D_\phi\}^k \hat{\beta}_G + [I - \{(A_\phi + D_\phi)^{-1} D_\phi\}^k] \hat{\beta}_\phi + o_p(n^{-\frac{1}{2}}),$$

where I is the $p \times p$ identity matrix,

$$D_\phi = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int \dot{\psi}_0(t) S^{(0)}(\beta_0; t)\{X_i - \bar{X}(\beta_0; t)\}^{\otimes 2}\, dN_i(\beta_0; t), \tag{2·5}$$

and $\dot{\psi}_0(t)$ is the derivative of $\psi_0(t)$, the limit of $\psi(\beta_0; t)$ as $n\to\infty$. Thus, $\hat{\beta}_{(k)}$ is asymptotically a weighted average of $\hat{\beta}_\phi$ and $\hat{\beta}_G$, and can be regarded as an approximation to $\hat{\beta}_\phi$. If $\{(A_\phi + D_\phi)^{-1}D_\phi\}^k \to 0$ as $k\to\infty$, then the limit of $\hat{\beta}_{(k)}$ as $k\to\infty$, denoted by $\hat{\beta}_{(\infty)}$, is asymptotically equivalent to $\hat{\beta}_\phi$. Note that $\{(A_\phi + D_\phi)^{-1}D_\phi\}^k \to 0$ as $k\to\infty$ is equivalent to all the eigenvalues of $(A_\phi + D_\phi)^{-1}D_\phi$ being less than 1 in absolute value. This condition holds if A_ϕ is positive definite and D_ϕ is nonnegative definite.

The most commonly used weight functions, including the log-rank, Prentice–Wilcoxon (Prentice, 1978) and the more general G^ρ class of Harrington & Fleming (1982), are all nonnegative and nonincreasing. For such weight functions, A_ϕ is indeed positive definite provided that the distribution of X does not concentrate on a $(p-1)$-dimensional hyperplane (Ying, 1993, pp. 81–2). It is evident from (2·5) that D_ϕ will be nonnegative definite if $\dot{\psi}_0(.) \geqslant 0$, which is true of all commonly used weight functions. Therefore, $\hat{\beta}_\phi$ is asymptotically equivalent to $\hat{\beta}_{(\infty)}$, at least for most important weight functions.

We show in the Appendix that, under mild conditions, $\hat{\beta}_\phi - \hat{\beta}_{(k)} = o_p(n^{-\frac{1}{2}})$ for any k of order $\log(n)$ provided that $\{(A_\phi + D_\phi)^{-1}D_\phi\}^k \to 0$ as $k\to\infty$. This implies that, with a finite number of iterations, $\hat{\beta}_{(k)}$ is asymptotically equivalent to the consistent roots of the original estimating function $U_\phi(\beta)$. In practice, one would terminate the iteration process when the difference between successive estimates is less than a certain bound. This convergence criterion may be set on the basis of the observed $\hat{\beta}_G$ and its estimated standard error, as demonstrated in the next section.

Although $\hat{\beta}_{(k)}$ can be easily obtained, it is again difficult to estimate the limiting covariance matrix analytically. As in the case of $\hat{\beta}_G$, we appeal to the resampling approach. To be specific, we construct the perturbed objective function

$$L^*_\phi(\beta; b) := n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi\{b; e_i(b)\}\Delta_i\{e_i(\beta) - e_j(\beta)\}^- Z_i,$$

and define $\hat{\beta}^*_{(0)} = \hat{\beta}^*_G$ and $\hat{\beta}^*_{(k)} = \arg\min_\beta L^*_\phi(\beta; \hat{\beta}^*_{(k-1)})$ $(k \geqslant 1)$. We show in the Appendix that the asymptotic distribution of $n^{\frac{1}{2}}(\hat{\beta}_{(k)} - \beta_0)$ can be approximated by the conditional distribution of $n^{\frac{1}{2}}(\hat{\beta}^*_{(k)} - \hat{\beta}_{(k)})$ given the data $(\tilde{T}_i, \Delta_i, X_i)$ $(i = 1, \ldots, n)$. Inference about β_0 can then be carried out on the basis of the empirical distribution of $\hat{\beta}^*_{(k)}$.

3. Numerical results

3·1. *Real examples*

The results reported here and in § 3·2 are based on $\psi = 1$ and $\psi = 1/S^{(0)}$, which are referred to as the Gehan and log-rank weight functions. For the resampling procedure, we simulated the Z_i from the unit exponential distribution and generated 500 realisations. The results were similar when other distributions of the Z_i were used. The Wald method was used to construct the confidence intervals.

We first applied the proposed methods to a study on multiple myeloma reported by Krall et al. (1975). Out of a total of 65 patients who were treated with alkylating agents, 48 died during the study and 17 survived. This is the main example in the PROC PHREG of the SAS/STAT User's Guide (1999, pp. 2608–17, 2536–641), which focused on the Cox model with haemoglobin, HGB, and the logarithm of blood urea nitrogen, BUN, as the covariates. To improve numerical efficiencies, we standardised the two covariates log(BUN) and HGB in our analysis. The parameter estimates for standardised log(BUN) and HGB based on the Gehan weight function are $-0{\cdot}532$ and $0{\cdot}292$ with estimated standard errors of $0{\cdot}146$ and $0{\cdot}169$. The corresponding 95% confidence intervals are $(-0{\cdot}818, -0{\cdot}246)$ and $(-0{\cdot}039, 0{\cdot}623)$.

We used the iterative algorithm to obtain the estimates based on the log-rank weight function. The results from the Gehan weight function suggested that it is not necessary to have accurate numbers beyond the second decimal point. Thus, we set the convergence criterion for the iteration to be $0{\cdot}01$ between successive estimates. Based on this criterion, convergence was achieved at the third iteration. The resulting estimates are $-0{\cdot}505$ and $0{\cdot}268$ with 95% confidence intervals $(-0{\cdot}880, -0{\cdot}130)$ and $(-0{\cdot}017, 0{\cdot}553)$. The estimates are identical up to the eighth decimal point after 5 iterations.

For further illustration, we considered the well-known Mayo primary biliary cirrhosis data (Fleming & Harrington, 1991, Appendix D.1). The database contains information about the survival time and prognostic factors for 418 patients. The data were used by Dickson et al. (1989) to build a Cox proportional hazards model for the natural history of the disease with five covariates, age, log(albumin), log(bilirubin), oedema and log(protime); see Fleming & Harrington (1991, Table 4.6.3). We considered the same set of covariates in our analysis. We again standardised the covariates. The estimates based on the Gehan weight function are $-0{\cdot}270$, $0{\cdot}204$, $-0{\cdot}593$, $-0{\cdot}223$ and $-0{\cdot}244$ with estimated standard errors of $0{\cdot}062$, $0{\cdot}069$, $0{\cdot}071$, $0{\cdot}070$ and $0{\cdot}080$. These results also suggested that the convergence criterion of $0{\cdot}01$ can be used for calculating the log-rank-type estimates. This criterion was again met at the third iteration. The estimates were identical up to the eighth decimal point after 12 iterations. We display in Table 1 the results for the unstandardised covariates. Although the conclusions are not qualitatively different, the analysis under the accelerated failure time model provides an alternative and more direct interpretation of the effects of covariates on the survival time, as compared to the Cox regression analysis. The results based on the two weight functions are fairly similar. The point estimates for the log-rank weight function are similar to those of Lin & Geyer (1992), which were obtained by a simulated annealing algorithm.

Table 1. *Accelerated failure time regression for the Mayo primary biliary cirrhosis data*

	Gehan weight function			Log-rank weight function		
Parameter	Est	SE	95% CI	Est	SE	95% CI
Age	−0·0258	0·0059	(−0·037, −0·014)	−0·0265	0·0042	(−0·035, −0·018)
log(Albumin)	1·5906	0·5352	(0·542, 2·639)	1·6558	0·3683	(0·934, 2·378)
log(Bilirubin)	−0·5789	0·0698	(−0·716, −0·442)	−0·5849	0·0455	(−0·674, −0·496)
Oedema	−0·8781	0·2768	(−1·421, −0·336)	−0·7338	0·1781	(−1·083, −0·385)
log(Protime)	−2·7680	0·9085	(−4·549, −0·987)	−1·9439	0·4622	(−2·850, −1·038)

Est, estimate; SE, standard error; CI, confidence interval.

3·2. *Simulation studies*

Extensive simulation studies were conducted to assess the operating characteristics of the proposed methods in practical settings. One series of the experiments mimicked the foregoing multiple myeloma study. The failure times were generated from the model

$$\log T = 2{\cdot}9413 - 0{\cdot}5196 \times \log(\text{BUN}) + 0{\cdot}2775 \times \text{HGB} + \varepsilon, \tag{3·1}$$

where ε is a normal random variable with mean 0 and standard deviation 1·027. The regression coefficients and the standard deviation of ε in model (3·1) were estimated from the parametric accelerated failure time model. The censoring times were generated from the Un$(0, \tau)$ distribution, where τ was chosen to yield a desired level of censoring.

We first examined the difference between log-rank-type estimates obtained after a fixed number of iterations versus those obtained after convergence. We generated 1000 datasets with sample size of 65 from model (3·1). The standardised covariate values from the original myeloma study were used. The algorithm was considered convergent if the difference between two consecutive estimates was less than 0·0001; convergence was usually achieved within 10 iterations. Figure 1(a) plots the estimates obtained after convergence versus those obtained after three iterations for the estimation of the first regression coefficient under 20% censoring; the two sets of estimates are clearly very similar. Figure 1(b) displays the corresponding plot for the log-rank estimates versus the Gehan estimates, the two of which are noticeably different. Thus, the phenomenon in Fig.1(a) is not an artifact of the closeness of the Gehan and log-rank estimates. The plots for the estimation of the second regression coefficient and for other censoring levels revealed similar results. The values of the log-rank estimating functions at convergence versus after three iterations were similar: under 20% censoring, the medians of the L_1-norms of $n^{-1}\tilde{U}_\phi$ among the 1000 datasets were 0·003 at convergence and 0·009 after three iterations; the corresponding maxima were 0·07 and 0·09. These results demonstrated that a small number of iterations is sufficient.

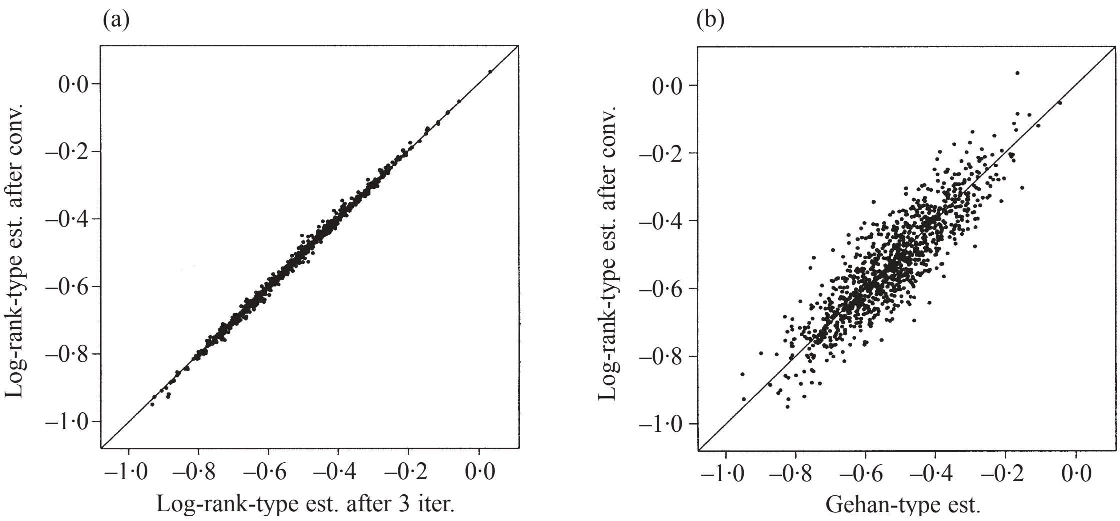

Fig. 1. Comparisons of different estimates of β_{01} under 20% censoring: (a) log-rank-type estimates after convergence versus log-rank-type estimates after three iterations; (b) log-rank-type estimates after convergence versus Gehan-type estimates.

Most of the simulation studies were concerned with the performance of the proposed inference procedures based on the Gehan-type estimator and the log-rank-type estimator with three iterations. We considered sample sizes of 65 and 130. For $n = 65$, the standardised covariate values from the multiple myeloma dataset were used; for $n = 130$, each covariate value in the original dataset was duplicated. For each configuration of the simulation parameters, we generated 500 datasets with the same set of covariate values. The results of the simulation studies are summarised in Table 2. The parameter estimators

Table 2. *Summary statistics for the simulation studies: regression coefficients β_{01} and β_{02} for* log(BUN) *and* HGB, *respectively*

n	Censoring	Weight	Parameter	Bias	SE	SEE	90% CP	95% CP
65	0%	Gehan	β_{01}	0·004	0·127	0·128	0·89	0·93
			β_{02}	0·001	0·127	0·129	0·89	0·94
		Log-rank	β_{01}	−0·017	0·135	0·138	0·90	0·94
			β_{02}	−0·001	0·140	0·138	0·88	0·93
	20%	Gehan	β_{01}	0·003	0·137	0·134	0·87	0·93
			β_{02}	0·001	0·137	0·139	0·88	0·94
		Log-rank	β_{01}	−0·006	0·138	0·148	0·91	0·95
			β_{02}	0·004	0·151	0·149	0·89	0·93
	30%	Gehan	β_{01}	0·0005	0·143	0·138	0·87	0·93
			β_{02}	0·0002	0·144	0·144	0·88	0·93
		Log-rank	β_{01}	−0·010	0·145	0·152	0·90	0·94
			β_{02}	0·004	0·158	0·156	0·88	0·93
130	0%	Gehan	β_{01}	−0·003	0·091	0·091	0·90	0·95
			β_{02}	−0·004	0·091	0·091	0·88	0·95
		Log-rank	β_{01}	−0·008	0·094	0·097	0·92	0·94
			β_{02}	0·003	0·093	0·097	0·91	0·97
	20%	Gehan	β_{01}	−0·0001	0·097	0·096	0·89	0·95
			β_{02}	−0·005	0·098	0·098	0·89	0·95
		Log-rank	β_{01}	0·005	0·108	0·105	0·88	0·94
			β_{02}	−0·001	0·106	0·107	0·90	0·96
	30%	Gehan	β_{01}	0·0004	0·101	0·099	0·89	0·94
			β_{02}	−0·004	0·103	0·102	0·89	0·95
		Log-rank	β_{01}	−0·0001	0·111	0·107	0·87	0·93
			β_{02}	0·010	0·116	0·110	0·88	0·93

Bias, bias of the estimator; SE, standard error of the estimator; SEE, mean of the standard error estimator; 90% CP, coverage probability of the 90% Wald-type confidence interval; 95% CP, coverage probability of the 95% Wald-type confidence interval

appear to be virtually unbiased. The standard error estimators reflect well the true variabilities of the parameter estimators, and the corresponding confidence intervals have satisfactory coverage probabilities.

4. Remarks

Fygenson & Ritov (1994) characterised the class of weight functions in (2·2) which yields monotone estimating functions. The class is quite restrictive and involves the censoring time distribution. By contrast, the class of monotone estimating functions developed in the present paper is very broad and yields consistent roots of the original weighted log-rank estimating equations.

On the basis of (2·4), we can obtain a one-step approximation to $\hat{\beta}_\phi$ for any weight function ϕ, as follows. Define

$$\tilde{\psi}(\beta;\, x) = S^{(0)}(\beta;\, x)^{-1}\phi(\beta;\, x)I - D_\phi A_G^{-1},$$

and replace $\psi\{\hat{\beta};\, t + (\beta - \hat{\beta})'X_i\}$ in (2·4) by $\tilde{\psi}\{\hat{\beta}_G;\, t + (\beta - \hat{\beta}_G)'X_i\}$. It then follows from (A·4) in the Appendix that the resulting estimator is asymptotically equivalent to $\hat{\beta}_\phi$. To implement this one-step estimator, one would need to estimate A_G and D_ϕ. A consistent estimator of A_G can be easily obtained because $A_G^{-1}B_GA_G^{-1}$ and B_G can be consistently estimated from the data. The estimation of D_ϕ would require an estimator of $\lambda(.)$, which can be obtained either nonparametrically or parametrically.

The proposed estimators do not achieve the semiparametric efficiency bound. By adopting the approach of Lai & Ying (1991b, § 3), one can construct data-dependent weight functions which yield asymptotically efficient estimators provided that the iterative procedure converges.

Acknowledgement

This research was supported by the National Institutes of Health, the National Science Foundation and the New York City Council Speaker's Fund for Public Health Research. The authors thank the reviewers for their helpful comments. Software implementing the proposed methods is available from the authors.

Appendix

Derivation of asymptotic results

Asymptotic properties of $\hat\beta_G^*$. Both $n^{-1}L_G$ and $n^{-1}L_G^*$ are convex functions and, by the strong law of large numbers, converge almost surely to the same limiting function. The first derivative of the limiting function is zero at β_0 and its second derivative at β_0 is A_G, which is nonsingular. Thus, this function has a unique minimiser β_0. It then follows from convex analysis that $\hat\beta_G \to \beta_0$ and $\hat\beta_G^* \to \beta_0$ almost surely. The latter convergence is with respect to the joint probability space of $(\tilde T_i, \Delta_i, X_i)$ $(i = 1, \ldots, n)$ and Z_i $(i = 1, \ldots, n)$. We shall use $\mathscr{F}$ to denote the σ-field generated by the original data.

It follows from Theorem 2 of Ying (1993) that almost surely

$$U_G(\hat\beta_G) = U_G(\beta_0) + nA_G(\hat\beta_G - \beta_0) + o(n^{\frac{1}{2}} + n\|\hat\beta_G - \beta_0\|). \tag{A·1}$$

By similar arguments,

$$U_G^*(\hat\beta_G^*) = U_G^*(\hat\beta_G) + nA_G(\hat\beta_G^* - \hat\beta_G) + o(n^{\frac{1}{2}} + n\|\hat\beta_G^* - \hat\beta_G\|) \tag{A·2}$$

almost surely. The functions U_G^* and U_G have the same asymptotic slope matrix A_G in (A·1) and (A·2) because $E\{U_G^*(\beta)|\mathscr{F}\} = U_G(\beta)$ and $\hat\beta_G$ and $\hat\beta_G^*$ are consistent.

We have that $U_G^*(\hat\beta_G) = U_G^*(\hat\beta_G) - U_G(\hat\beta_G) + o(n^{\frac{1}{2}})$ since $\hat\beta_G$ is a root of $U_G(\beta)$. Thus

$$n^{-\frac{1}{2}}U_G^*(\hat\beta_G) = n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_{-\infty}^{\infty} S^{(0)}(\hat\beta_G; t)\{X_i - \bar X(\hat\beta_G; t)\}\, dN_i(\hat\beta_G; t)(Z_i - 1) + o(1).$$

Conditionally on $\mathscr{F}$, the right-hand side of the above display is a normalised sum of independent zero-mean random vectors. Since its conditional covariance matrix converges almost surely to B_G, the multivariate central limit theorem implies that $n^{-\frac{1}{2}}U_G^*(\hat\beta_G)$ converges in distribution to $\mathscr{N}(0, B_G)$ almost surely. It then follows from (A·2) that the conditional distribution of $n^{\frac{1}{2}}(\hat\beta_G^* - \hat\beta_G)$ given $\mathscr{F}$ converges almost surely to $\mathscr{N}(0, A_G^{-1}B_GA_G^{-1})$, which is the limiting distribution of $n^{\frac{1}{2}}(\hat\beta_G - \beta_0)$.

Asymptotic properties of $\hat\beta_{(k)}$. We first show that $\hat\beta_{(k)}$ is strongly consistent for any given k. Recall that $\hat\beta_{(0)} = \hat\beta_G$. As a result of the strong consistency of $\hat\beta_{(0)}$, the function $n^{-1}L_\phi(\beta; \hat\beta_{(0)})$ converges almost surely to $\lim n^{-1}L_\phi(\beta; \beta_0)$, which reaches its minimum at $\beta = \beta_0$. Since $n^{-1}L_\phi(\beta; \hat\beta_{(0)})$ is convex, $\hat\beta_{(1)}$ must converge to β_0 almost surely. Applying these arguments successively for $k = 2, 3, \ldots$, we see that $\hat\beta_{(k)} \to \beta_0$ almost surely as $n \to \infty$ for all k.

To avoid possible tail instabilities, we assume that, for any β_n and η_n which converge almost surely to β_0 and 0, respectively,

$$\psi(\beta_n;\, t+\eta_n) = \psi(\beta_n;\, t) + \dot{\psi}_0(t)\eta_n + o(n^{-\frac{1}{2}} + \eta_n)$$

uniformly in t. This condition holds for virtually all the existing weight functions for t away from the 'endpoint'. For it to hold for all t, we typically would have to use ϕ which is truncated at the tail, that is $\phi(t)=0$ for t near the endpoint. Under this condition and Conditions 1–5 of Ying (1993), we can expand $\tilde{U}_\phi(\beta;\, \hat{\beta}_{(k-1)})$ to yield

$$\begin{aligned}\tilde{U}_\phi(\beta;\, \hat{\beta}_{(k-1)}) = {} & \sum_{i=1}^{n} \int \psi(\hat{\beta}_{(k-1)};\, t) S^{(0)}(\beta;\, t)\{X_i - \bar{X}(\beta;\, t)\}\, dN_i(\beta;\, t) \\ & + \sum_{i=1}^{n} \int \dot{\psi}_0(t) S^{(0)}(\beta;\, t)\{X_i - \bar{X}(\beta;\, t)\} X_i'\, dN_i(\beta;\, t)(\beta - \hat{\beta}_{(k-1)}) \\ & + o(n^{\frac{1}{2}} + n\|\beta - \hat{\beta}_{(k-1)}\|)\end{aligned}$$

as $n\to\infty$ and $\beta\to\beta_0$. By Theorem 2 of Ying (1993), the first term on the right-hand side of the above display has the linear expansion $U_\phi(\beta_0) + nA_\phi(\beta-\beta_0) + o(n^{\frac{1}{2}} + n\|\beta-\beta_0\|)$. Thus,

$$\begin{aligned}\tilde{U}_\phi(\hat{\beta}_{(k)};\, \hat{\beta}_{(k-1)}) = {} & U_\phi(\beta_0) + n(A_\phi + D_\phi)(\hat{\beta}_{(k)} - \beta_0) - nD_\phi(\hat{\beta}_{(k-1)} - \beta_0) \\ & + o(n^{\frac{1}{2}} + n\|\hat{\beta}_{(k)} - \beta_0\| + n\|\hat{\beta}_{(k-1)} - \beta_0\|).\end{aligned} \qquad (\text{A}\cdot 3)$$

Suppose that A_G, A_ϕ and $A_\phi + D_\phi$ are nonsingular. By combining (A·1) and (A·3) with $k=1$, we obtain

$$\begin{aligned}n^{\frac{1}{2}}(\hat{\beta}_{(1)} - \beta_0) = {} & -n^{-\frac{1}{2}}(A_\phi + D_\phi)^{-1}\{U_\phi(\beta_0) + D_\phi A_G^{-1} U_G(\beta_0)\} \\ & + o(1 + n^{\frac{1}{2}}\|\hat{\beta}_{(1)} - \beta_0\| + n^{\frac{1}{2}}\|\hat{\beta}_G - \beta_0\|).\end{aligned} \qquad (\text{A}\cdot 4)$$

It then follows from recursive use of (A·3) that in general

$$\begin{aligned}n^{\frac{1}{2}}(\hat{\beta}_{(k)} - \beta_0) = {} & -n^{-\frac{1}{2}} \sum_{j=1}^{k} \{(A_\phi + D_\phi)^{-1} D_\phi\}^{j-1} (A_\phi + D_\phi)^{-1} U_\phi(\beta_0) \\ & - n^{-\frac{1}{2}}\{(A_\phi + D_\phi)^{-1} D_\phi\}^k A_G^{-1} U_G(\beta_0) + o\left(1 + n^{\frac{1}{2}} \sum_{j=0}^{k} \|\hat{\beta}_{(j)} - \beta_0\|\right)\end{aligned}$$

or

$$\begin{aligned}n^{\frac{1}{2}}(\hat{\beta}_{(k)} - \beta_0) = {} & -n^{-\frac{1}{2}}[I - \{(A_\phi + D_\phi)^{-1} D_\phi\}^k] A_\phi^{-1} U_\phi(\beta_0) \\ & - n^{-\frac{1}{2}}\{(A_\phi + D_\phi)^{-1} D_\phi\}^k A_G^{-1} U_G(\beta_0) + o\left(1 + n^{\frac{1}{2}} \sum_{j=0}^{k} \|\hat{\beta}_{(j)} - \beta_0\|\right).\end{aligned} \qquad (\text{A}\cdot 5)$$

The asymptotic normality of $n^{\frac{1}{2}}(\hat{\beta}_{(k)} - \beta_0)$ then follows from that of $n^{-\frac{1}{2}}U_\phi(\beta_0)$ and $n^{-\frac{1}{2}}U_G(\beta_0)$.

Suppose that ϕ is truncated at the tail and the error density is sufficiently smooth. Then it follows from equation (4.5) of Lai & Ying (1991b) that, for some $\eta>0$ and $\gamma>0$, (A·3) holds with $o(n^{\frac{1}{2}-\eta})$ as the remainder term uniformly for $\hat{\beta}_{(k-1)}$ and $\hat{\beta}_{(k)}$ in an $O(n^{-\frac{1}{2}+\gamma})$ neighbourhood of β_0. Thus, if $\{(A_\phi + D_\phi)^{-1}D_\phi\}^k \to 0$ as $k\to\infty$, then all the $\hat{\beta}_{(j)}$ $(j \leqslant k)$ stay in an $O(n^{-\frac{1}{2}+\gamma})$ neighbourhood of β_0 so that the remainder term in (A·5) becomes $o(kn^{-\eta}) = o(1)$ provided that $k\to\infty$ and $k = O(\log n)$. Hence, $n^{\frac{1}{2}}(\hat{\beta}_{(k)} - \beta_0) = n^{\frac{1}{2}}(\hat{\beta}_\phi - \beta_0) + o_p(1)$ for $k\to\infty$ and $k = O(\log n)$.

Asymptotic properties of $\hat{\beta}^*_{(k)}$. We impose the same regularity conditions as in the previous section. The convexity arguments and the strong law of large numbers can again be employed to establish the strong consistency of $\hat{\beta}^*_{(k)}$. By definition, $\hat{\beta}^*_{(k)}$ is a root of

$$\tilde{U}^*_\phi(\beta;\, \hat{\beta}^*_{(k-1)}) := \sum_{i=1}^{n} \int \psi\{\hat{\beta}^*_{(k-1)};\, t + (\beta - \hat{\beta}^*_{(k-1)})' X_i\} S^{(0)}(\beta;\, t)\{X_i - \bar{X}(\beta;\, t)\}\, dN_i(\beta;\, t) Z_i.$$

By asymptotic linear expansions similar to those used in establishing (A·3), we have

$$\tilde{U}^*_\phi(\hat\beta^*_{(k)};\,\hat\beta^*_{(k-1)}) = \sum_{i=1}^n \int \psi\{\hat\beta_{(k-1)};\, t + (\hat\beta_{(k)} - \hat\beta_{(k-1)})' X_i\} S^{(0)}(\hat\beta_{(k)};\, t)\{X_i - \bar{X}(\hat\beta_{(k)};\, t)\}\, dN_i(\hat\beta_{(k)};\, t) Z_i$$
$$+ n(A_\phi + D_\phi)(\hat\beta^*_{(k)} - \hat\beta_{(k)}) - nD_\phi(\hat\beta^*_{(k-1)} - \hat\beta_{(k-1)}) + d^*_n, \qquad (A\cdot 6)$$

where $d^*_n = o(n^{\frac{1}{2}} + n\sum_{j=0}^k \|\hat\beta_{(j)} - \beta_0\| + n\sum_{j=0}^k \|\hat\beta^*_{(j)} - \beta_0\|)$.

We can replace the Z_i in the first term on the right-hand side of (A·6) by $Z_i - 1$ since $\hat\beta_{(k)}$ is a root of $\tilde{U}_\phi(\beta;\, \hat\beta_{(k-1)})$. With this replacement, the first term has mean 0 conditionally on $\mathscr{F}$, so that the integral can be approximated by $\int \phi(\beta_0; t)\{X_i - \bar{X}(\beta_0;\, t)\}\, dN_i(\beta_0;\, t)$. Thus,

$$\tilde{U}^*_\phi(\hat\beta^*_{(k)};\,\hat\beta^*_{(k-1)}) = \sum_{i=1}^n \int \phi(\beta_0;\, t)\{X_i - \bar{X}(\beta_0;\, t)\}\, dN_i(\beta_0;\, t)(Z_i - 1)$$
$$+ n(A_\phi + D_\phi)(\hat\beta^*_{(k)} - \hat\beta_{(k)}) - nD_\phi(\hat\beta^*_{(k-1)} - \hat\beta_{(k-1)}) + d^*_n,$$

which yields

$$n^{\frac{1}{2}}(\hat\beta^*_{(k)} - \hat\beta_{(k)}) = -n^{-\frac{1}{2}}(A_\phi + D_\phi)^{-1} \sum_{i=1}^n \int \phi(\beta_0;\, t)\{X_i - \bar{X}(\beta_0;\, t)\}\, dN_i(\beta_0;\, t)(Z_i - 1)$$
$$+ (A_\phi + D_\phi)^{-1} D_\phi n^{\frac{1}{2}}(\hat\beta^*_{(k-1)} - \hat\beta_{(k-1)}) + n^{-\frac{1}{2}} d^*_n.$$

Therefore,

$$\begin{aligned} n^{\frac{1}{2}}(\hat\beta^*_{(k)} - \hat\beta_{(k)}) = &-n^{-\frac{1}{2}}[I - \{(A_\phi + D_\phi)^{-1} D_\phi\}^k] A_\phi^{-1} \\ &\times \sum_{i=1}^n \int \phi(\beta_0;\, t)\{X_i - \bar{X}(\beta_0;\, t)\}\, dN_i(\beta_0;\, t)(Z_i - 1) \\ &- n^{-\frac{1}{2}}\{(A_\phi + D_\phi)^{-1} D_\phi\}^k A_G^{-1} \\ &\times \sum_{i=1}^n \int S^{(0)}(\beta_0;\, t)\{X_i - \bar{X}(\beta_0;\, t)\}\, dN_i(\beta_0;\, t)(Z_i - 1) + n^{-\frac{1}{2}} d^*_n. \end{aligned}$$

Comparing the above equation with (A·5), we see that the conditional distribution of $n^{\frac{1}{2}}(\hat\beta^*_{(k)} - \hat\beta_{(k)})$ given $\mathscr{F}$ converges almost surely to the limiting distribution of $n^{\frac{1}{2}}(\hat\beta_{(k)} - \beta_0)$.

Under the additional conditions stated in the last paragraph of the previous section, (A·6) can be strengthened with $d^*_n = o(n^{-\xi})$ for some $\xi > 0$. Hence, for $k \to \infty$ and $k = O(\log n)$, $n^{\frac{1}{2}}(\hat\beta^*_{(k)} - \hat\beta_{(k)})$ has the same asymptotic distribution as $n^{\frac{1}{2}}(\hat\beta_{(k)} - \beta_0)$.

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.
Dickson, E. R., Grambsch, P. M., Fleming, T. R., Fisher, L. D. & Langworthy, A. (1989). Prognosis in primary biliary cirrhosis: model for decision making. *Hepatology* **10**, 1–7.
Fleming, T. R. & Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* New York: Wiley.
Fygenson, M. & Ritov, Y. (1994). Monotone estimating equations for censored data. *Ann. Statist.* **22**, 732–46.
Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily single-censored samples. *Biometrika* **52**, 203–23.
Harrington, D. P. & Fleming, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika* **69**, 133–43.
Jin, Z., Ying, Z. & Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381–90.
Kalbfleisch, J. D. & Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.
Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. (1983). Optimisation by simulated annealing. *Science* **220**, 671–80.

KOENKER, R. & BASSETT, G. S. (1978). Regression quantiles. *Econometrica* **46**, 33–50.

KOENKER, R. & D'OREY, V. (1987). Computing regression quantiles. *Appl. Statist.* **36**, 383–93.

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

LAI, T. L. & YING, Z. (1991a). Large sample theory of a modified Buckley–James estimator for regression analysis with censored data. *Ann. Statist.* **19**, 1370–402.

LAI, T. L. & YING, Z. (1991b). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19**, 531–56.

LIN, D. Y. & GEYER, C. J. (1992). Computational methods for semiparametric linear regression with censored data. *J. Comp. Graph. Statist.* **1**, 77–90.

LIN, D. Y., WEI, L. J. & YING, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85**, 605–18.

MANTEL, N. (1966). Evaluation of survival data and two new rank order statistics arising in its considerations. *Cancer Chemo. Rep.* **50**, 163–70.

PARZEN, M. I., WEI, L. J. & YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–50.

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

RAO, C. R. & ZHAO, L. C. (1992). Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap. *Sankhyā* A **54**, 323–31.

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

SAS INSTITUTE, INC. (1999). *SAS/STAT User's Guide*, Version 8. Cary, NC: SAS Institute Inc.

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

[*Received August* 2001. *Revised November* 2002]