Efficient Estimation for the Accelerated Failure Time Model

Donglin ZENG and D. Y. LIN

The accelerated failure time model provides a natural formulation of the effects of covariates on potentially censored response variable. The existing semiparametric estimators are computationally intractable and statistically inefficient. In this article we propose an approximate nonparametric maximum likelihood method for the accelerated failure time model with possibly time-dependent covariates. We estimate the regression parameters by maximizing a kernel-smoothed profile likelihood function. The maximization can be achieved through conventional gradient-based search algorithms. The resulting estimators are consistent and asymptotically normal. The limiting covariance matrix attains the semiparametric efficiency bound and can be consistently estimated. We also provide a consistent estimator for the error distribution. Extensive simulation studies demonstrate that the asymptotic approximations are accurate in practical situations and the new estimators are considerably more efficient than the existing ones. Illustrations with clinical and epidemiologic studies are provided.

KEY WORDS: Censoring; Kernel smoothing; Linear regression; Profile likelihood; Semiparametric efficiency; Survival data.

1. INTRODUCTION

The proportional hazards model (Cox 1972) and the accelerated failure time model are the two major approaches to the regression analysis of censored data (Cox and Oakes 1984, chap. 5; Kalbfleisch and Prentice 2002, chap. 2). Due to the availability of efficient inference procedures that are implemented in all statistical software packages, the proportional hazards model is used almost exclusively in practice. As noted by D. R. Cox (Reid 1994, p. 450), however, the accelerated failure time model (i.e., the log-linear model) is "in many ways more appealing because of its quite direct physical interpretation," especially when the response variable does not pertain to failure time. This model may provide more accurate or more concise summarization of the data than the proportional hazards model in certain applications.

The presence of censoring poses major challenges in the semiparametric analysis of the accelerated failure time model. Rank estimators have been studied by Prentice (1978), Tsiatis (1990), Wei, Ying, and Lin (1990), Lai and Ying (1991a), Robins and Tsiatis (1992), Ying (1993), Lin and Ying (1995), Jones (1997), Yang (1997), and Zhou (2005), and least squares estimators have been studied by Buckley and James (1979), Ritov (1990), and Lai and Ying (1991b). Because the estimating functions are discrete with potentially multiple roots, it is difficult to calculate these estimators and even more difficult to estimate their variances. Recently, Jin, Lin, Wei, and Ying (2003) and Jin, Lin, and Ying (2006) developed approximations to these estimators that can be obtained through linear programming, together with resampling procedures for variance estimation. However, their methods cannot handle time-dependent covariates and are not computationally feasible for large data sets. Furthermore, none of the existing estimators achieves the semiparametric efficiency bound.

The efficient inference for the proportional hazards model is based on the partial likelihood, which is a special case of the profile likelihood. The profile likelihood fails for the accelerated failure time model because the function is very discrete. In this article we use kernel smoothing to construct a smooth approximation to the profile likelihood function for the regression parameters of the accelerated failure time model with possibly time-dependent covariates. The kernel-smoothed profile likelihood function has a local expansion similar to that of the partial likelihood function. The estimators that maximize this function can be easily calculated by the Newton–Raphson algorithm or any other optimization algorithms for smooth objective functions. In addition, the estimators are consistent and asymptotically normal, with a limiting covariance matrix that attains the semiparametric efficiency bound and can be readily estimated. Furthermore, we provide an explicit estimator for the error distribution.

The rest of the article is organized as follows. In Section 2 we describe the proposed estimation approach. In Section 3 we present the asymptotic properties of the new estimators. In Section 4 we report the results of our simulation studies. In Section 5 we provide applications to two major medical studies. We provide some concluding remarks in Section 6, and outline the proofs of the asymptotic results in the Appendix.

2. ESTIMATION

Let T and $\mathbf{X}(\cdot)$ denote the failure time and a d-vector of possibly time-dependent covariates. When the covariates are all time-independent, the accelerated failure time model takes the log-linear form

$$\log T = -\beta^{\mathrm{T}}\mathbf{X} + \epsilon, \tag{1}$$

where ϵ is a measurement error independent of $\mathbf{X}$ (Cox and Oakes 1984, pp. 64–65; Kalbfleisch and Prentice 2002, p. 44). To accommodate time-dependent covariates, we consider the following extension:

$$e^{\epsilon} = \int_0^T e^{\beta^{\mathrm{T}}\mathbf{X}(t)}\,dt \tag{2}$$

(Cox and Oakes 1984, p. 67). For theoretical developments, it is helpful to express model (2) in terms of hazard function. Let $\lambda(t)$ and $\Lambda(t)$ denote the hazard function and cumulative hazard function of e^{ϵ}, and let $\lambda_{T|\mathbf{X}}(t)$ and $\Lambda_{T|\mathbf{X}}(t)$ denote the conditional hazard and cumulative hazard functions of T given $\mathbf{X}$. Then

$$\Lambda_{T|\mathbf{X}}(t) = -\log \mathrm{P}\left(\epsilon > \log \int_0^t e^{\beta^{\mathrm{T}}\mathbf{X}(s)}\,ds\right)$$

Donglin Zeng is Associate Professor (E-mail: *dzeng@bios.unc.edu*) and D. Y. Lin is Dennis Gillings Distinguished Professor (E-mail: *lin@bios.unc.edu*), Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599. This research was supported by the National Institutes of Health.

Journal of the American Statistical Association
December 2007, Vol. 102, No. 480, Theory and Methods
DOI 10.1198/016214507000001085

$$= \Lambda\left(\int_0^t e^{\beta^T \mathbf{X}(s)}\, ds\right)$$

or, equivalently,

$$\lambda_{T|\mathbf{X}}(t) = \lambda\left(\int_0^t e^{\beta^T \mathbf{X}(s)}\, ds\right) e^{\beta^T \mathbf{X}(t)}. \tag{3}$$

Let C denote the censoring time. For a random sample of size n, the data consist of $(Y_i, \Delta_i, \mathbf{X}_i(\cdot))$, $i = 1, \ldots, n$, where $Y_i = \min(T_i, C_i)$, $\Delta_i = I(T_i \le C_i)$, and $I(\cdot)$ is the indicator function. We make the standard assumptions that C is independent of T conditional on $\mathbf{X}$ and that the distribution of C does not depend functionally on β. Then the log-likelihood function concerning β and λ is

$$n^{-1} \sum_{i=1}^n \left[\Delta_i \beta^T \mathbf{X}_i(Y_i) + \Delta_i \log \lambda\left(e^{R_i(\beta)}\right) - \Lambda\left(e^{R_i(\beta)}\right)\right], \tag{4}$$

where $R_i(\beta) = \log \int_0^{Y_i} e^{\beta^T \mathbf{X}_i(s)}\, ds$.

The maximization of (4) with respect to β and Λ would yield the nonparametric maximum likelihood estimators. Unfortunately, the maximum of (4) does not exist, as we now explain. In nonparametric maximum likelihood estimation, Λ is considered a right-continuous function, so the objective function (4) becomes

$$n^{-1} \sum_{i=1}^n \left[\Delta_i \beta^T \mathbf{X}_i(Y_i) + \Delta_i \log \Lambda\left\{e^{R_i(\beta)}\right\} - \Lambda\left(e^{R_i(\beta)}\right)\right],$$

where $\Lambda\{y\}$ is the jump size of $\Lambda(t)$ at $t = y$. Simple algebraic manipulations yield that for fixed β, the estimator of Λ has jumps at the $e^{R_i(\beta)}$, and the jump size is equal to

$$\frac{\Delta_i}{\sum_{j=1}^n I(R_j(\beta) \ge R_i(\beta))}.$$

Plugging this expression into the objective function and discarding constant terms, we obtain the profile log-likelihood function for β as

$$n^{-1} \sum_{i=1}^n \left[\Delta_i \beta^T \mathbf{X}_i(Y_i) - \Delta_i \log\left\{\sum_{j=1}^n I\left(R_j(\beta) \ge R_i(\beta)\right)\right\}\right].$$

Because the second term depends only on the ranks of the $R_i(\beta)$ and these ranks are stable as β becomes extreme, the objective function cannot achieve its maximum for finite β.

The reason why nonparametric maximum likelihood estimation fails for the accelerated failure time model is that the estimator of Λ is very nonsmooth. Thus we are motivated to seek a smooth version. It is not obvious, however, what kind of smooth estimator should be used for Λ or, equivalently, λ. To find such an estimator, we start with the simplest case of a piecewise constant λ. To be specific, we partition an interval containing all $e^{R_i(\beta)}$'s into J_n equally spaced intervals, $0 \equiv t_0 < t_1 < \cdots < t_{J_n} \equiv M$, where M denotes an upper bound for the $e^{R_i(\beta)}$ over all possible β's in a bounded set. A piecewise constant λ takes the form

$$\lambda(t) = \sum_{k=1}^{J_n} c_k I(t \in [t_{k-1}, t_k)).$$

Because for any t,

$$\Lambda(t) = \sum_{k=1}^{J_n} c_k (t - t_k) I(t_{k-1} \le t < t_k) + \frac{M}{J_n} \sum_{k=1}^{J_n} c_k I(t \ge t_k),$$

the log-likelihood function (4) can be rewritten as

$$\begin{aligned} &n^{-1} \sum_{i=1}^n \Delta_i \beta^T \mathbf{X}_i(Y_i) \\ &\quad + n^{-1} \sum_{k=1}^{J_n} \log c_k \left\{\sum_{i=1}^n \Delta_i I\left(e^{R_i(\beta)} \in [t_{k-1}, t_k)\right)\right\} \\ &\quad - n^{-1} \sum_{k=1}^{J_n} c_k \left\{\sum_{i=1}^n \left(e^{R_i(\beta)} - t_k\right) I\left(t_{k-1} \le e^{R_i(\beta)} < t_k\right) \right. \\ &\quad \left. + \frac{M}{J_n} \sum_{i=1}^n I\left(e^{R_i(\beta)} \ge t_k\right)\right\}. \end{aligned} \tag{5}$$

By differentiating with respect to c_k, we see that the solution to the score equation of c_k is

$$\begin{aligned} c_k = &\left(\sum_{j=1}^n \Delta_j I\left(e^{R_j(\beta)} \in [t_{k-1}, t_k)\right)\right) \\ &\bigg/ \left(\sum_{j=1}^n \left(e^{R_j(\beta)} - t_k\right) I\left(t_{k-1} \le e^{R_j(\beta)} < t_k\right) \right. \\ &\left. + \sum_{j=1}^n I\left(e^{R_j(\beta)} \ge t_k\right) \frac{M}{J_n}\right), \end{aligned}$$

where $0/0 = 0$. After plugging the equations for the c_k into (5) and discarding an additive component irrelevant to β, we obtain the following sieve profile log-likelihood function:

$$\begin{aligned} l_n^p(\beta) = &\frac{1}{n} \sum_{i=1}^n \Delta_i \beta^T \mathbf{X}_i(Y_i) \\ &+ \sum_{k=1}^{J_n} \left\{\frac{1}{n} \sum_{i=1}^n \Delta_i I\left(e^{R_i(\beta)} \in [t_{k-1}, t_k)\right)\right\} \\ &\times \log\left\{\frac{J_n}{nM} \sum_{j=1}^n \Delta_j I\left(e^{R_j(\beta)} \in [t_{k-1}, t_k)\right)\right\} \\ &- \sum_{k=1}^{J_n} \left[\frac{1}{n} \left\{\sum_{i=1}^n \Delta_i I\left(e^{R_i(\beta)} \in [t_{k-1}, t_k\right)\right\} \right. \\ &\times \log\left\{\frac{J_n}{nM} \sum_{j=1}^n \left(e^{R_j(\beta)} - t_k\right) I\left(e^{R_j(\beta)} \in [t_{k-1}, t_k)\right) \right. \\ &\left.\left. + \frac{1}{n} \sum_{j=1}^n I\left(e^{R_j(\beta)} \ge t_k\right)\right\}\right]. \end{aligned}$$

Even the function $l_n^p(\beta)$ is not smooth and may have multiple local maxima, as illustrated in Figure 1(a). Thus we need

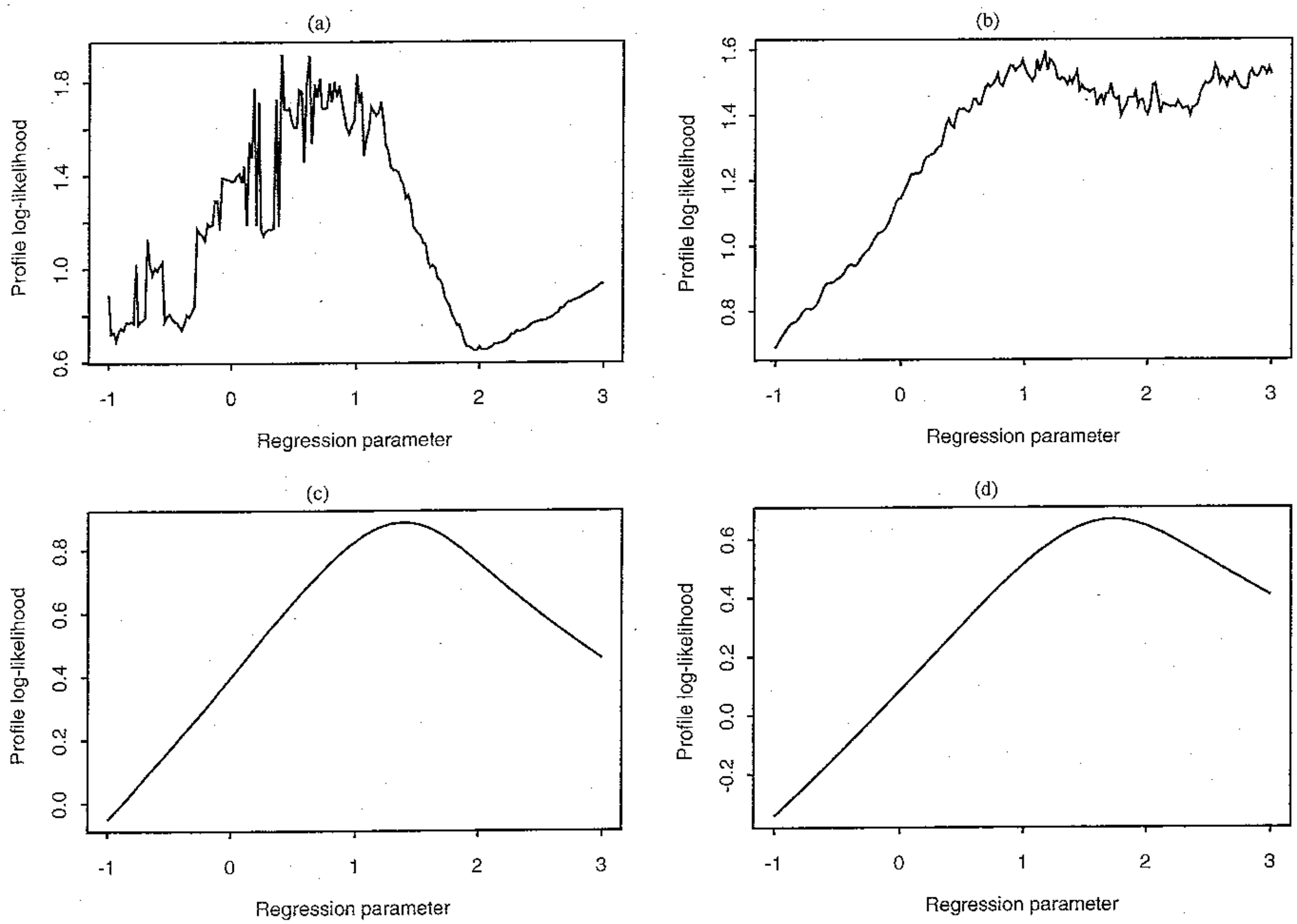

Figure 1. Profile log-likelihood functions for a simulated data set with $n = 200$ and a single uniform covariate. (a) Profile log-likelihood function with a piece-wise constant hazard function. (b)–(d) Kernel-smoothed profile log-likelihood functions with bandwidths n^{-1}, $n^{-1/5}$, and $n^{-1/9}$.

further smoothing on this function. We show in the supplementary technical report that as $n \to \infty$, $J_n \to \infty$, and $J_n/n \to 0$, $l_n^p(\beta)$ converges uniformly in a compact set of β to

$$l(\beta) = E\Bigg[\Delta\beta^{\mathrm{T}}\mathbf{X}(Y) + \Delta \log\Bigg(\frac{d\mathrm{P}(\Delta = 1, \int_0^Y e^{\beta^{\mathrm{T}}\mathbf{X}(s)}\,ds \le t)}{dt} \Big/ \mathrm{P}\Big(\int_0^Y e^{\beta^{\mathrm{T}}\mathbf{X}(s)}\,ds \ge t\Big)\Bigg)\Bigg|_{t=\int_0^Y e^{\beta^{\mathrm{T}}\mathbf{X}(s)}\,ds}\Bigg].$$

It suffices to seek a smooth approximation to $l(\beta)$ using the empirical observations. Toward this end, we choose a kernel function $K(\cdot)$ with bandwidth a_n. The theory of kernel estimation indicates that under suitable regularity conditions,

$$\frac{1}{na_n}\sum_{i=1}^n \Delta_i K\left(\frac{R_i(\beta) - \log t}{a_n}\right) \to \frac{d\mathrm{P}(\Delta = 1, R(\beta) \le s)}{ds}\Bigg|_{s=\log t} = \frac{d\mathrm{P}(\Delta = 1, e^{R(\beta)} \le t)}{dt}t$$

and

$$\frac{1}{na_n}\sum_{i=1}^n \int_{-\infty}^{\log t} K\left(\frac{R_i(\beta) - s}{a_n}\right) ds \to \mathrm{P}(R(\beta) \le \log t) = \mathrm{P}\big(e^{R(\beta)} \le t\big).$$

Thus we approximate

$$\frac{d\mathrm{P}(\Delta = 1, \int_0^Y e^{\beta^{\mathrm{T}}\mathbf{X}(s)}\,ds \le t)/dt}{\mathrm{P}(\int_0^Y e^{\beta^{\mathrm{T}}\mathbf{X}(s)}\,ds \ge t)}$$

by

$$\frac{1}{t}\,\frac{(na_n)^{-1}\sum_{i=1}^n \Delta_i K(\frac{R_i(\beta) - \log t}{a_n})}{\int_{\log t}^{\infty}(na_n)^{-1}\sum_{i=1}^n K(\frac{R_i(\beta) - s}{a_n})\,ds}.$$

Because the expectation in $l(\beta)$ can be approximated by the empirical measure, we obtain a kernel-smoothed approximation of $l(\beta)$,

$$l_n^s(\beta) = \frac{1}{n}\sum_{i=1}^n \Delta_i\beta^{\mathrm{T}}\mathbf{X}_i(Y_i) - \frac{1}{n}\sum_{i=1}^n \Delta_i R_i(\beta) + \frac{1}{n}\sum_{i=1}^n \Delta_i \log\left\{\frac{1}{na_n}\sum_{j=1}^n \Delta_j K\left(\frac{R_j(\beta) - R_i(\beta)}{a_n}\right)\right\}$$

$$-\frac{1}{n}\sum_{i=1}^{n}\Delta_i \log\left\{\frac{1}{n}\sum_{j=1}^{n}\int_{-\infty}^{(R_j(\beta)-R_i(\beta))/a_n} K(s)\,ds\right\}.$$

We propose to maximize $l_n^s(\beta)$ over β and denote the resulting estimator as $\widehat{\beta}_n$. Because $K(\cdot)$ is a smooth kernel function, we can use the Newton–Raphson algorithm or other gradient-based search algorithms to calculate $\widehat{\beta}_n$. Figure 1(b)–1(d) shows the values of $l_n^s(\beta)$ with various choices of bandwidth for a simulated data set. It is evident that an appropriate choice of bandwidth will lead to a smooth profile likelihood function. We discuss the specific choices of the kernel function and bandwidth in Section 4.

Given $\widehat{\beta}_n$, we estimate $\lambda(t)$ by the following kernel-smoothed estimator:

$$\widehat{\lambda}_n(t) = \frac{(na_n t)^{-1}\sum_{i=1}^{n}\Delta_i K(\frac{R_i(\widehat{\beta}_n)-\log t}{a_n})}{n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{(R_i(\widehat{\beta}_n)-\log t)/a_n} K(u)\,du}.$$

The corresponding estimator of $\Lambda(t)$ is

$$\widehat{\Lambda}_n(t) = \int_{-\infty}^{\log t}\frac{(na_n)^{-1}\sum_{i=1}^{n}\Delta_i K(\frac{R_i(\widehat{\beta}_n)-s}{a_n})}{n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{(R_i(\widehat{\beta}_n)-s)/a_n} K(u)\,du}\,ds. \qquad (6)$$

3. ASYMPTOTIC RESULTS

Let β_0 and λ_0 denote the true values of β and λ. For any function $g(\cdot)$, let $g'(\cdot)$ denote the first derivative of $g(\cdot)$ and let $g^{(r)}(\cdot)$ denote the rth derivative of $g(\cdot)$ ($r \geq 0$). We impose the following regularity conditions:

(C.1) The true parameter value β_0 belongs to a known compact set $\mathcal{B}$ in $\mathcal{R}^d$.
(C.2) If there exists a constant vector η and a deterministic function $g(\cdot)$ such that $\eta^{\mathrm{T}}\mathbf{X}(T) = g(\epsilon)$ with probability 1, then $\eta = \mathbf{0}$ and $g = 0$.
(C.3) For $t \geq 0$, $\lambda_0(t)$ is positive and thrice-continuously differentiable with $\lambda_0'(0) > 0$.
(C.4) The censoring time C has a positive and twice-continuously differentiable density in $[0, \tau)$, and there exists a positive constant δ_0 such that $\mathrm{P}(C \geq \tau | X(s), s \leq \tau) > \delta_0$ with probability 1.
(C.5) The kernel function $K(\cdot)$ is thrice-continuously differentiable, and $K^{(r)}(\cdot)$, $r = 0, 1, 2, 3$, have bounded variations in $(-\infty, \infty)$.

Remark 1. When $\mathbf{X}$ is time-independent, condition (C.2) is equivalent to the condition where the column vectors of $[1, \mathbf{X}^{\mathrm{T}}]$ are linearly independent with positive probability. This condition ensures the identifiability of the regression parameters. Conditions (C.3) and (C.4) are the smoothness conditions imposed on the underlying density functions. The second part of condition (C.4) states that at least some subjects are censored at the end of the study. Condition (C.5) ensures that the kernel-smoothed estimators used in $l_n^s(\beta)$ and their derivatives are consistent approximations; condition (C.5) is satisfied by various kernel functions, including Gaussian kernels and smooth kernels with bounded support.

The consistency and asymptotic normality of $\widehat{\beta}_n$ are stated next.

Theorem 1. Suppose that conditions (C.1)–(C.5) hold and that $a_n = n^{\nu}$ with $\nu \in (0, -1/2)$. Then $\widehat{\beta}_n$ is strongly consistent for β_0 as $n \to \infty$.

Theorem 2. Suppose that conditions (C.1)–(C.5) hold and that the first $(m-1)$ moments of $K(\cdot)$ are 0 for some $m > 3$ and $a_n = n^{\nu}$ with $\nu \in (-1/2m, -1/6)$. Then, as $n \to \infty$, $\sqrt{n}(\widehat{\beta}_n - \beta_0)$ converges in distribution to a mean-0 normal random vector with covariance matrix equal to the semiparametric efficiency bound of β_0.

Remark 2. The proof of Theorem 1 is based on the uniform approximation of $l_n^s(\beta)$ to a function with a unique maximum at β_0 and makes use of the proof of theorem 5.7 of van der Vaart (1998). The proof of Theorem 2 is based on the expansion of the score function for $\widehat{\beta}_n$ and relies on the modern empirical process theory. In the proof we show that the inverse negative second derivative of $l_n^s(\widehat{\beta}_n)$ can be used to estimate the asymptotic covariance matrix of $\sqrt{n}(\widehat{\beta}_n - \beta_0)$. Theorem 2 imposes more stringent requirements on the choice of the kernel function and the bandwidth than Theorem 1, because consistency of the first and second derivatives of the kernel density function is needed in establishing the asymptotic distribution of $\widehat{\beta}_n$. These assumptions have been commonly made in the nonparametric literature (e.g., Schuster 1969; Rao 1983, p. 238; Hart 1997, p. 64); they are sufficient conditions and may not be necessary.

To describe the asymptotic properties of $\widehat{\Lambda}_n$, we introduce some notation. Let $\mathbf{l}_\beta^*$ denote the efficient score function for β_0, which is defined in (A.2) of the Appendix. Write $R_0 = R(\beta_0)$, $\mathbf{R}^{(1)}(\beta) = \partial R(\beta)/\partial\beta$, $\mathbf{R}_0^{(1)} = \mathbf{R}^{(1)}(\beta_0)$, and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^{\mathrm{T}}$. In addition, let $f_{Z_1|\mathbf{Z}_2}$ denote the conditional density of Z_1 given $\mathbf{Z}_2$, and let $f'_{Z_1|\mathbf{Z}_2}$ denote the derivative of $f_{Z_1|\mathbf{Z}_2}$ with respect to Z_1. Finally, let $\widetilde{\tau}$ denote any constant less than $\sup_{\mathbf{X}}\int_0^{\tau} e^{\beta_0^{\mathrm{T}}\mathbf{X}(t)}\,dt$.

Theorem 3. Under the conditions of Theorem 2, $\sqrt{n}(\widehat{\Lambda}_n - \Lambda_0)$ converges weakly in $l^{\infty}([0, \widetilde{\tau}])$ to a mean-0 Gaussian process with covariance function $E\{Q(Y, \Delta, \mathbf{X}; t)Q(Y, \Delta, \mathbf{X}; s)\}$ at (t, s) as $n \to \infty$, where

$$\begin{aligned}
&Q(Y, \Delta, \mathbf{X}; t)\\
&\quad = \int_{-\infty}^{\log t}\frac{\Delta f_{R_0|\Delta}(s)}{\mathrm{P}(R_0 > s)}\,ds\\
&\qquad - \int_{-\infty}^{\log t}\frac{E[\Delta f_{R_0|\Delta}(s)]}{\mathrm{P}(R_0 > s)^2} I(R_0 > s)\,ds\\
&\qquad + \int_{-\infty}^{\log t}\left\{\frac{E[\Delta \mathbf{R}_0^{(1)^{\mathrm{T}}} f'_{R_0|\Delta,\mathbf{R}_0^{(1)}}(s)]}{\mathrm{P}(R_0 > s)}\right.\\
&\qquad \left. - \frac{E[\Delta f_{R_0|\Delta}(s)]E[\Delta \mathbf{R}_0^{(1)^{\mathrm{T}}} f_{R_0|\Delta,\mathbf{R}_0^{(1)}}(s)]}{\mathrm{P}(R_0 > s)^2}\right\}ds\\
&\quad \times E\big[\mathbf{l}_\beta^{*\otimes 2}\big]^{-1}\mathbf{l}_\beta^*.
\end{aligned}$$

4. NUMERICAL STUDIES

We conducted numerous simulation studies to examine the small-sample performance of the proposed inference procedures. We generated failure times from the following model:

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

where X_1 is Bernoulli with .5 success probability and X_2 is independent normal with mean 0 and standard derivation .5. This is the same model used by Jin et al. (2006). We considered six error distributions: standard normal distribution; (standard) extreme-value distribution; Weibull distributions with hazard rates $2t$ and $1/(2\sqrt{t})$, denoted by Weibull(2, 1) and Weibull(.5, 1); and mixtures of N(0, 1) and N(0, 9) with mixing probabilities (.5, .5) and (.95, .05), denoted by .5N(0, 1) + .5N(0, 9) and .95N(0, 1) + .05N(0, 9). (Weibull distributions pertain to e^{ϵ} rather than to ϵ.) We generated censoring times from the uniform $[0, \tau]$ distribution, where τ was chosen to produce a 25% censoring rate. We set n to 100, 200, and 400.

We chose the kernel function $K(\cdot)$ to be the standard normal density for convenience and tractability. The smoothed profile likelihood function involves the kernel density of $(\log Y + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$ for uncensored subjects and the cumulative kernel density of $(\log Y + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$ for all subjects. We used the optimal bandwidths (Jones 1990; Jones and Sheather 1991), $(8\sqrt{2}/3)^{1/5}\sigma_1 n^{-1/5}$ and $4^{1/3}\sigma_2 n^{-1/3}$, where σ_1 and σ_2 are the sample standard deviations of $(\log Y + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$ (with $\boldsymbol{\beta}$ being the initial value in the estimation) among uncensored subjects and among all subjects. Our experience indicated that the variance estimator may be sensitive to the choice of bandwidth, especially for heavy-tailed error distributions. Thus in the variance estimation, we replaced σ_1 and σ_2 by the minimum of the sample standard deviation and the interquartile range divided by 1.34, as suggested by the Silverman (1986) rule of thumb for choosing bandwidth in kernel estimation.

We obtained the estimates of the regression parameters using the quasi-Newton search algorithm in MATLAB, which conducts search within the trusted region to avoid local maxima. We set the initial values to 0. We terminated the iterations when the change in the function value or the gradient length was less than 10^{-20}. We used the curvatures of $l_n^s(\widehat{\boldsymbol{\beta}}_n)$ to estimate the variances.

For efficiency comparisons, we also included the log-rank and least squares estimators as implemented by Jin et al. (2003, 2006). The log-rank and least squares estimators are asymptotically efficient under the extreme-value and normal error distributions, respectively. We did not evaluate the variance estimators because of heavy computational burdens, especially for $n = 400$.

Table 1 summarizes the results of these studies. The proposed estimators of β_1 and β_2 are virtually unbiased. The variance estimators accurately reflect the true variations and the confidence intervals have proper coverage probabilities. Although it does not satisfy the moment condition of Theorem 2, the Gaussian kernel seems to provide desirable results. Under the normal error, the proposed estimators are slightly less efficient than the least squares estimators and more efficient than the log-rank estimators. Under the extreme-value error, the proposed estimators are slightly less efficient than the log-rank estimators and more efficient than the least squares estimators. Under all other error distributions, the proposed estimators are more efficient than the log-rank and least squares estimators. The efficiency gains are particularly substantial under the Weibull(2, 1) distribution.

The quasi-Newton algorithm converged very rapidly in all of the simulation runs. Completing the simulations for the proposed inference procedures took approximately 2 hours on an IBM BladeCenter HS40 machine, compared with 12 hours for the Gehan estimator and 24 hours for the log-rank and least squares estimators (without variance estimation).

5. EXAMPLES

We first consider the well-known Mayo primary biliary cirrhosis (PBC) study (Fleming and Harrington 1991, app. D.1). The data contain information about the survival time and prognostic factors for 418 patients. Jin et al. (2003, 2006) fitted the accelerated failure time model with five covariates—age, log(albumin), log(bilirubin), edema, and log(protime)—using the rank and least squares estimators. We fit the same model using the proposed method with normal kernel function and with bandwidths of $\sigma n^{-1/5}, \sigma n^{-1/7}$, and $\sigma n^{-1/9}$, where σ is the sample standard derivation of $\log Y$, as well as the optimal bandwidths described in the previous section, denoted by a_n^{opt}. The results are given in Table 2. The parameter estimates as well as the variance estimates are robust to the choice of bandwidth. Our parameter estimates are similar to the Gehan estimates of Jin et al. (2003) and the least squares estimates of Jin et al. (2006), whereas our standard error estimates tend to be smaller.

It is natural to estimate the conditional survival function at time t given covariates $\mathbf{X}$ by $\exp\{-\widehat{\Lambda}_n(e^{\widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{X}}t)\}$. We can estimate the marginal survival function for a subgroup by averaging the conditional survival function estimates. Figure 2 displays the estimated survival curves for the PBC patients in two age groups. The model-based estimates agree well with the Kaplan–Meier estimates except at the right tails.

Our second example pertains to the Cardiovascular Health Study (CHS), a major epidemiologic cohort study involving 5,888 men and women age 65 years and older from 4 U.S. field centers (Fried et al. 1991). Primary endpoints of this study include myocardial infraction, stroke, and cardiovascular disease mortality. The investigators are particularly interested in assessing the effects of baseline risk factors on the time to the first occurrence of a primary endpoint among Caucasian subjects. The total number of Caucasian subjects is 3,907, and about 27% of them have reached the primary endpoints. A total of 10 baseline covariates are considered: age, sex, an ordinary hypertension scale, body mass index, systolic blood pressure, smoking status (1, smoker; 0, nonsmoker), diabetes status (1, yes; 0, no), and 3 dummy variables comparing the 4 field centers. We fit the accelerated failure time model using the proposed method with the normal kernel function and optimal bandwidth. Table 3 displays the results of our analysis, along with those of the proportional hazards regression. Using either model, significant effects can be concluded due to age, sex, hypertension, systolic blood pressure, smoking, diabetes, and field centers.

6. DISCUSSION

The proposed estimators are much easier to calculate than the existing ones. Indeed, the existing methods are computationally intractable for large studies, such as the CHS study. A second advantage of the proposed estimators is that they achieve the semiparametric efficiency bound. (The existing estimators are asymptotically efficient only under specific error distributions.) Finally, the kernel-smoothed profile likelihood function allows

Table 1. Summary statistics for the simulation studies

Error distribution	n		Profile likelihood				Log-rank		Least squares	
			Bias	SE	SEE	CP	Bias	SE	Bias	SE
N(0, 1)	100	β_1	.007	.234	.239	.950	.007	.237	.004	.219
		β_2	.006	.236	.244	.948	−.004	.242	.001	.225
	200	β_1	−.003	.161	.163	.955	−.008	.166	−.007	.155
		β_2	.000	.163	.164	.953	−.003	.168	−.005	.157
	400	β_1	.003	.113	.113	.941	−.001	.116	.001	.108
		β_2	−.003	.113	.113	.949	−.007	.121	−.006	.112
Extreme-value	100	β_1	.021	.284	.285	.940	.004	.233	.008	.283
		β_2	.023	.285	.291	.953	.001	.237	.005	.290
	200	β_1	.003	.194	.191	.935	−.009	.167	−.005	.209
		β_2	.015	.203	.193	.933	−.005	.171	−.006	.213
	400	β_1	.006	.133	.131	.933	−.003	.119	−.002	.152
		β_2	.009	.128	.131	.950	.003	.115	.004	.144
Weibull(2, 1)	100	β_1	.011	.133	.133	.955	.016	.332	.019	.530
		β_2	.012	.138	.136	.960	.009	.303	−.005	.529
	200	β_1	.001	.084	.080	.933	−.009	.194	−.025	.390
		β_2	.002	.089	.081	.935	.002	.197	−.004	.385
	400	β_1	.001	.053	.050	.931	−.002	.121	−.011	.284
		β_2	.003	.053	.051	.940	.004	.116	.011	.275
Weibull(.5, 1)	100	β_1	.005	.096	.104	.958	.004	.120	.004	.101
		β_2	.002	.097	.105	.965	−.002	.119	−.002	.101
	200	β_1	−.003	.067	.069	.947	−.007	.086	−.004	.074
		β_2	−.001	.071	.070	.955	−.003	.088	−.002	.075
	400	β_1	.001	.049	.047	.936	−.001	.062	−.001	.053
		β_2	.001	.047	.047	.961	.001	.060	.001	.051
.5N(0, 1) + .5N(0, 9)	100	β_1	.022	.392	.422	.947	.006	.448	.018	.471
		β_2	−.004	.417	.433	.940	−.026	.453	−.026	.465
	200	β_1	.003	.289	.282	.938	−.008	.329	−.003	.341
		β_2	.003	.278	.284	.945	−.012	.317	−.008	.328
	400	β_1	.003	.193	.191	.941	.003	.226	.002	.235
		β_2	−.001	.190	.194	.942	−.007	.221	−.007	.234
.95N(0, 1) + .05N(0, 9)	100	β_1	.006	.248	.252	.946	.004	.255	.002	.259
		β_2	.006	.249	.255	.946	−.002	.259	−.002	.265
	200	β_1	−.005	.169	.173	.959	−.009	.175	−.008	.178
		β_2	−.002	.175	.174	.943	−.009	.189	−.010	.187
	400	β_1	.002	.118	.120	.953	.000	.125	.001	.125
		β_2	−.002	.121	.121	.946	−.006	.133	−.006	.131

NOTE: SE is the standard error of the parameter estimator, SEE is the mean of the standard error estimator, and CP is the coverage probability of the 95% confidence interval. Each entry is based on 1,000 replicates.

one to perform likelihood ratio test and to conduct likelihood-based model selection and variable selection.

For practical sample sizes, oversmoothing the kernel density function, as suggested by the asymptotic theory, may induce bias, and the choice of the kernel function with zero moments may yield nonnegative values for the estimated hazard function. The Gaussian kernel function and the bandwidth that we suggest perform well under a wide variety of settings. The

Table 2. Accelerated failure time regression for the Mayo PBC data

	$a_n = \sigma n^{-1/5}$		$a_n = \sigma n^{-1/7}$		$a_n = \sigma n^{-1/9}$		a_n^{opt}	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Age	.0263	.0061	.0287	.0065	.0299	.0068	.0286	.0061
log(albumin)	−1.5138	.5251	−1.6267	.5284	−1.5761	.5613	−1.6212	.4761
log(bilirubin)	.5959	.0606	.6272	.0795	.6500	.0815	.6175	.0669
Edema	.9588	.3075	.8167	.2633	.7943	.2665	.7985	.3179
log(protime)	2.4228	.7391	2.7811	.8834	2.9989	.9242	2.4095	.8050

NOTE: Estimate denotes the parameter estimate, SE, the (estimated) standard error.

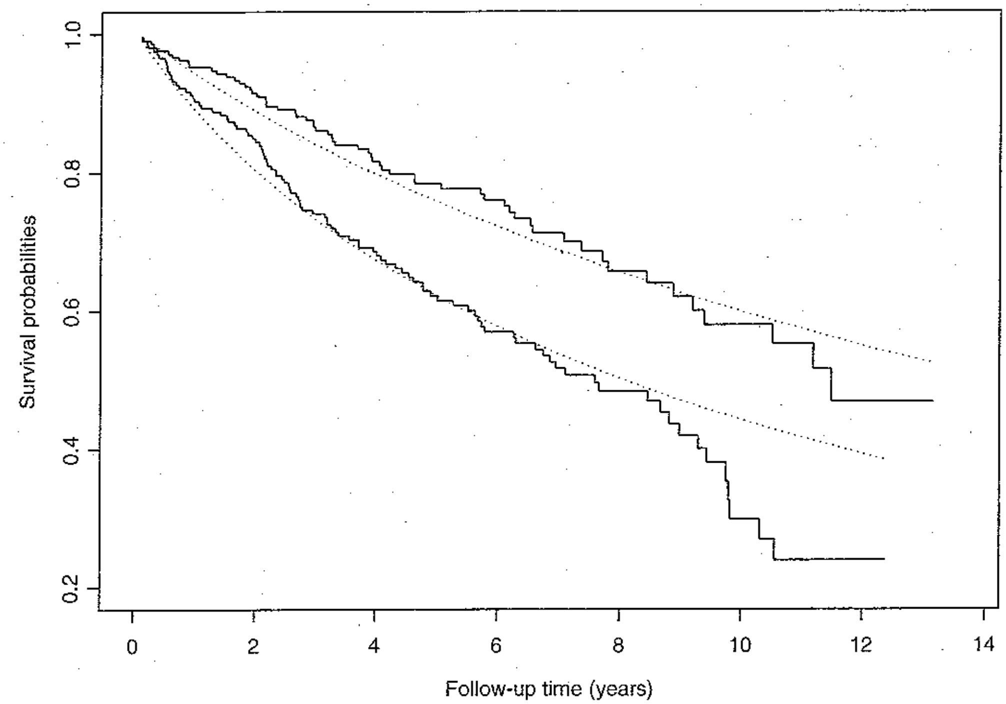

Figure 2. Estimated survival functions for the PBC patients. The lower and upper solid curves are the Kaplan–Meier estimates for the patients with ages > 50 and ≤50; the dotted curves are the corresponding model-based estimates.

Gaussian kernel function is a standard choice in nonparametric estimation due to its tractability. The optimal bandwidth for density estimation may not be optimal for estimating its derivatives. It would be worthwhile to develop data-adaptive strategies for selecting the kernel function and the bandwidth.

By providing computationally feasible and statistically efficient inference procedures, our work makes the accelerated failure time model a more viable alternative to the proportional hazards model. Goodness-of-fit tests as implemented in S–PLUS revealed that the proportional hazards assumption is questionable for the PBC and CHS data. It would be worthwhile to develop appropriate techniques to check the accelerated failure time model and to determine which of the two models fits the data better. In principle, by including appropriate time-dependent covariates, both the proportional hazards and accelerated failure time models can fit any data reasonably well.

Table 3. Analysis of the CHS data under the proportional hazards and accelerated failure time models

	PH model		AFT model	
Parameter	Estimate	SE	Estimate	SE
Age	.0863	.0055	.0656	.0054
Gender	.4002	.0620	.3460	.0506
Hypertension	.1573	.0428	.1090	.0337
Body mass index	.0099	.0073	.0040	.0059
Systolic blood pressure	.0078	.0017	.0063	.0016
Smoking	.4865	.0976	.3306	.0831
Diabetes	.5271	.0808	.4365	.0705
Center 2 versus center 1	−.0522	.0870	−.0282	.0667
Center 3 versus center 1	.0446	.0854	.0468	.0659
Center 4 versus center 1	−.2108	.0933	−.1639	.0743

NOTE: Estimate denotes the parameter estimate; SE, the (estimated) standard error.

It is sensible to fit both models to the same data, because they provide different measures of regression effects.

In some applications, censoring arises when the assay cannot detect values below or above certain thresholds. This is the case with, for instance, the coronary artery calcification data. For such censored data, the accelerated failure time model is particularly appealing, because the concept of hazard is irrelevant in this context. When the assay cannot detect values below certain thresholds, the resulting data are left-censored. We can turn left-censored data to right-censored data by regarding $-T$ and $-C$ as the response and censoring variables.

The accelerated failure time model and the proposed kernel-smoothed profile likelihood approach can be extended to multivariate failure time data, which arise when each subject can potentially experience several events or when the study subjects are sampled in clusters such that the failure times within the same cluster are correlated. Extension to joint modeling for event times and repeated measures is also possible. These extensions are currently under investigation.

APPENDIX: PROOFS OF ASYMPTOTIC RESULTS

Here we sketch the proofs of Theorems 1–3. The details are provided in the supplementary technical report.

Proof of Theorem 1

In view of the proof of theorem 5.7 of van der Vaart (1998), it suffices to show that $\sup_{\beta\in\mathcal{B}} |l_n^s(\beta) - l(\beta)| \to_{\text{a.s.}} 0$ as $n \to \infty$ and that β_0 is the unique maximizer of $l(\beta)$. The uniform convergence follows from the following results:

$$\sup_{\beta\in\mathcal{B},s}\left|\frac{1}{na_n}\sum_{j=1}^{n}\Delta_j K\left(\frac{R_j(\beta)-s}{a_n}\right)-\frac{d\mathrm{P}(\Delta=1,R(\beta)\le s)}{ds}\right|\xrightarrow{\text{a.s.}}0$$

and

$$\sup_{\beta\in\mathcal{B},s}\left|\frac{1}{n}\sum_{j=1}^{n}\int_{-\infty}^{(R_j(\beta)-s)/a_n}K(u)\,du-\mathrm{P}(R(\beta)>s)\right|\xrightarrow{\text{a.s.}}0.$$

We can verify these results by appealing to lemma 2.4 of Schuster (1969) and theorem 2.4.3 of van der Vaart and Wellner (1996). To show that β_0 is the unique maximizer of $l(\beta)$, we define

$$\Lambda(t;\beta)=\int_{-\infty}^{\log t}\frac{d\mathrm{P}(\Delta=1,R(\beta)\le s)/ds}{\mathrm{P}(R(\beta)>s)}\,ds.$$

Note that $\Lambda(t;\beta_0)=\Lambda_0(t)$ and $E[\Lambda(\int_0^Y e^{\beta^{\mathrm T}\mathbf{X}(s)}\,ds;\beta)]=\mathrm{P}(\Delta=1)$. Thus if β maximizes $l(\beta)$, then

$$\begin{aligned}E\bigg[&\Delta\beta^{\mathrm T}\mathbf{X}(Y)+\Delta\log\lambda\bigg(\int_0^Y e^{\beta^{\mathrm T}\mathbf{X}(s)}\,ds;\beta\bigg)\\&-\Lambda\bigg(\int_0^Y e^{\beta^{\mathrm T}\mathbf{X}(s)}\,ds;\beta\bigg)\bigg]\\&\ge E\bigg[\Delta\beta_0^{\mathrm T}\mathbf{X}(Y)+\Delta\log\lambda_0\bigg(\int_0^Y e^{\beta_0^{\mathrm T}\mathbf{X}(s)}\,ds\bigg)\\&-\Lambda_0\bigg(\int_0^Y e^{\beta_0^{\mathrm T}\mathbf{X}(s)}\,ds\bigg)\bigg],\end{aligned}$$

where $\lambda(t;\beta)=\partial\Lambda(t;\beta)/\partial t$. It follows from the nonnegativity of the Kullback–Leibler information that

$$\begin{aligned}\exp\bigg\{&\Delta\beta^{\mathrm T}\mathbf{X}(Y)+\Delta\log\lambda\bigg(\int_0^Y e^{\beta^{\mathrm T}\mathbf{X}(s)}\,ds;\beta\bigg)\\&-\Lambda\bigg(\int_0^Y e^{\beta^{\mathrm T}\mathbf{X}(s)}\,ds;\beta\bigg)\bigg\}\\&=\exp\bigg\{\Delta\beta_0^{\mathrm T}\mathbf{X}(Y)+\Delta\log\lambda_0\bigg(\int_0^Y e^{\beta_0^{\mathrm T}\mathbf{X}(s)}\,ds\bigg)\\&-\Lambda_0\bigg(\int_0^Y e^{\beta_0^{\mathrm T}\mathbf{X}(s)}\,ds\bigg)\bigg\}\end{aligned}$$

with probability 1. We choose $\Delta=0$ and $Y=\tau$. The resulting equation subtracted from the equation obtained from choosing $\Delta=1$ and integrating from $Y=y$ to $Y=\tau$ yields $\Lambda(\int_0^y e^{\beta^{\mathrm T}\mathbf{X}(s)}\,ds;\beta)=\Lambda_0(\int_0^y e^{\beta_0^{\mathrm T}\mathbf{X}(s)}\,ds)$. Thus there exists an increasing and differentiable function G such that $\int_0^T e^{\beta^{\mathrm T}\mathbf{X}(s)}\,ds=G(\int_0^T e^{\beta_0^{\mathrm T}\mathbf{X}(s)}\,ds)$. Differentiating both sides with respect to T yields $(\beta-\beta_0)^{\mathrm T}\mathbf{X}(T)=\log G'(e^{\epsilon})$. It then follows from condition (C.2) that $\beta=\beta_0$, and thus $\Lambda=\Lambda_0$.

Proof of Theorem 2

Let P and $\mathbb{P}_n$ denote the probability measure and empirical measure. Because $\partial l_n^s(\widehat\beta_n)/\partial\beta=\mathbf{0}$, we have

$$\mathbf{0}=\mathbb{P}_n\bigg[\Delta\mathbf{X}(Y)-\Delta\mathbf{R}^{(1)}(\widehat\beta_n)+\Delta\frac{\mathbf{g}_{1n}(Y,\mathbf{X};\widehat\beta_n)}{g_{2n}(Y,\mathbf{X};\widehat\beta_n)}-\Delta\frac{\mathbf{g}_{3n}(Y,\mathbf{X};\widehat\beta_n)}{g_{4n}(Y,\mathbf{X};\widehat\beta_n)}\bigg],$$

where, if we define $\psi(y,\mathbf{x};\beta)=\log\int_0^y\exp\{\beta^{\mathrm T}\mathbf{x}(s)\}\,ds$ and $\psi^{(1)}(y,\mathbf{x};\beta)=\partial\psi(y,\mathbf{x};\beta)/\partial\beta$, then

$$\begin{aligned}\mathbf{g}_{1n}(y,\mathbf{x};\beta)=\mathbb{P}_n\bigg\{&\Delta K^{(1)}\bigg(\frac{R(\beta)-\psi(y,\mathbf{x};\beta)}{a_n}\bigg)\\&\times\frac{\mathbf{R}^{(1)}(\beta)-\psi^{(1)}(y,\mathbf{x};\beta)}{a_n^2}\bigg\},\end{aligned}$$

$$g_{2n}(y,\mathbf{x};\beta)=\mathbb{P}_n\bigg\{\Delta\frac{1}{a_n}K\bigg(\frac{R(\beta)-\psi(y,\mathbf{x};\beta)}{a_n}\bigg)\bigg\},$$

$$\begin{aligned}\mathbf{g}_{3n}(y,\mathbf{x};\beta)=\mathbb{P}_n\bigg\{&\frac{\mathbf{R}^{(1)}(\beta)-\psi^{(1)}(y,\mathbf{x};\beta)}{a_n}\\&\times K\bigg(\frac{R(\beta)-\psi(y,\mathbf{x};\beta)}{a_n}\bigg)\bigg\},\end{aligned}$$

and

$$g_{4n}(y,\mathbf{x};\beta)=\mathbb{P}_n\bigg\{\int_{-\infty}^{(R(\beta)-\psi(y,\mathbf{x};\beta))/a_n}K(s)\,ds\bigg\}.$$

We denote the expectations of $\mathbf{g}_{kn}(y,\mathbf{x};\beta)$ $(k=1,\dots,4)$ as $\mathbf{g}_{k0}(y,\mathbf{x};\beta)$.

Write $\mathbb{G}_n=\sqrt{n}(\mathbb{P}_n-P)$, $\widehat R=R(\widehat\beta_n)$, and $\widehat{\mathbf{R}}^{(1)}=\mathbf{R}^{(1)}(\widehat\beta_n)$. In addition, let $(\widetilde Y,\widetilde\Delta,\widetilde{\mathbf{X}})$ be an independent copy of $(Y,\Delta,\mathbf{X})$, and let $\widetilde R(\beta)$ and $\widetilde{\mathbf{R}}^{(1)}(\beta)$ be the corresponding copies of $R(\beta)$ and $\mathbf{R}^{(1)}(\beta)$. Furthermore, write $\widetilde R_0=\widetilde R(\beta_0)$, $\widetilde{\mathbf{R}}_0^{(1)}=\widetilde{\mathbf{R}}^{(1)}(\beta_0)$, $\check R=\widetilde R(\widehat\beta_n)$, and $\check{\mathbf{R}}^{(1)}=\widetilde{\mathbf{R}}^{(1)}(\widehat\beta_n)$. Then we can express the score equation as

$$\begin{aligned}\mathbf{0}=\;&\mathbb{G}_n\bigg\{\Delta\mathbf{X}(Y)-\Delta\widehat{\mathbf{R}}^{(1)}+\Delta\frac{\mathbf{g}_{1n}(Y,\mathbf{X};\widehat\beta_n)}{g_{2n}(Y,\mathbf{X};\widehat\beta_n)}-\Delta\frac{\mathbf{g}_{3n}(Y,\mathbf{X};\widehat\beta_n)}{g_{4n}(Y,\mathbf{X};\widehat\beta_n)}\bigg\}\\&+\mathbb{G}_n\bigg\{E_{\widetilde Y,\widetilde\Delta,\widetilde{\mathbf{X}}}\bigg[\widetilde\Delta\frac{\Delta K^{(1)}(\frac{\widehat R-\check R}{a_n})\frac{\widehat{\mathbf{R}}^{(1)}-\check{\mathbf{R}}^{(1)}}{a_n^2}}{g_{2n}(\widetilde Y,\widetilde{\mathbf{X}};\widehat\beta_n)}\bigg]\bigg\}\\&-\mathbb{G}_n\bigg\{E_{\widetilde Y,\widetilde\Delta,\widetilde{\mathbf{X}}}\bigg[\widetilde\Delta\frac{\Delta\frac{1}{a_n}K(\frac{\widehat R-\check R}{a_n})\mathbf{g}_{10}(\widetilde Y,\widetilde{\mathbf{X}};\widehat\beta_n)}{g_{2n}(\widetilde Y,\widetilde{\mathbf{X}};\widehat\beta_n)g_{20}(\widetilde Y,\widetilde{\mathbf{X}};\widehat\beta_n)}\bigg]\bigg\}\\&-\mathbb{G}_n\bigg\{E_{\widetilde Y,\widetilde\Delta,\widetilde{\mathbf{X}}}\bigg[\widetilde\Delta\frac{\frac{\widehat{\mathbf{R}}^{(1)}-\check{\mathbf{R}}^{(1)}}{a_n}K(\frac{\widehat R-\check R}{a_n})}{g_{4n}(\widetilde Y,\widetilde{\mathbf{X}};\widehat\beta_n)}\bigg]\bigg\}\\&+\mathbb{G}_n\bigg\{E_{\widetilde Y,\widetilde\Delta,\widetilde{\mathbf{X}}}\bigg[\widetilde\Delta\frac{\int_{-\infty}^{(\widehat R-\check R)/a_n}K(s)\,ds\,\mathbf{g}_{30}(\widetilde Y,\widetilde{\mathbf{X}};\widehat\beta_n)}{g_{4n}(\widetilde Y,\widetilde{\mathbf{X}};\widehat\beta_n)g_{40}(\widetilde Y,\widetilde{\mathbf{X}};\widehat\beta_n)}\bigg]\bigg\}\\&+\sqrt{n}E\bigg[\Delta\mathbf{X}(Y)-\Delta\widehat{\mathbf{R}}^{(1)}+\Delta\frac{\mathbf{g}_{10}(Y,\mathbf{X};\widehat\beta_n)}{g_{20}(Y,\mathbf{X};\widehat\beta_n)}\\&-\Delta\frac{\mathbf{g}_{30}(Y,\mathbf{X};\widehat\beta_n)}{g_{40}(Y,\mathbf{X};\widehat\beta_n)}\bigg].\end{aligned}\tag{A.1}$$

By evaluating the limits of $\mathbf{g}_{kn}$ $(k=1,\dots,4)$ and applying theorem 2.11.23 of van der Vaart and Wellner (1996), we can simplify (A.1) as

$$\begin{aligned}\mathbf{0}=\mathbb{G}_n\mathbf{l}_\beta^*(Y,\Delta,\mathbf{X})+\sqrt{n}E\bigg[&\Delta\mathbf{X}(Y)-\Delta\mathbf{R}^{(1)}(\widehat\beta_n)\\&+\Delta\frac{\mathbf{g}_{10}(Y,\mathbf{X};\widehat\beta_n)}{g_{20}(Y,\mathbf{X};\widehat\beta_n)}-\Delta\frac{\mathbf{g}_{30}(Y,\mathbf{X};\widehat\beta_n)}{g_{40}(Y,\mathbf{X};\widehat\beta_n)}\bigg],\end{aligned}$$

where

$$\begin{aligned}&\mathbf{l}_\beta^*(Y,\Delta,\mathbf{X})\\&\quad=\int\bigg\{\mathbf{X}(s)\\&\qquad-E_{\widetilde Y,\widetilde\Delta,\widetilde{\mathbf{X}}}\bigg[\widetilde{\mathbf{X}}(\widetilde Y)\,|\,\widetilde R_0=\log\int_0^s e^{\beta_0^{\mathrm T}\mathbf{X}(u)}\,du,\widetilde\Delta=1\bigg]\bigg\}\,dM_{\mathbf{X}}(s)\\&\qquad+\int\bigg\{\mathbf{R}_0^{(1)}-E_{\widetilde Y,\widetilde\Delta,\widetilde{\mathbf{X}}}\bigg[\widetilde{\mathbf{R}}_0^{(1)}\bigg|\widetilde R_0=\log\int_0^s e^{\beta_0^{\mathrm T}\mathbf{X}(u)}\,du,\widetilde\Delta=1\bigg]\bigg\}\\&\qquad\times\frac{\lambda_0'(\int_0^s e^{\beta_0^{\mathrm T}\mathbf{X}(u)}\,du)\int_0^s e^{\beta_0^{\mathrm T}\mathbf{X}(u)}\,du}{\lambda_0(\int_0^s e^{\beta_0^{\mathrm T}\mathbf{X}(u)}\,du)}\,dM_{\mathbf{X}}(s)\end{aligned}\tag{A.2}$$

and $M_{\mathbf{X}}(s)=\Delta I(Y\le s)-I(Y>s)\,d\Lambda(\int_0^s e^{\beta_0^{\mathrm T}\mathbf{X}(u)}\,du)$.

When $\mathbf{X}(\cdot)$ is time-independent, $\mathbf{l}^*_\beta$ is easily recognized as the efficient score function for β_0 given by Bickel, Klaassen, Ritov, and Wellner (1993, p. 149). For general $\mathbf{X}(\cdot)$, by expanding $\mathbf{l}^*_\beta$, we obtain

$$\begin{aligned}\mathbf{l}^*_\beta(Y,\Delta,\mathbf{X}) &= \Delta\mathbf{X}(Y) + \left\{\Delta\frac{\lambda_0'(\int_0^Y e^{\beta_0^{\mathrm{T}}\mathbf{X}(s)}\,ds)}{\lambda_0(\int_0^Y e^{\beta_0^{\mathrm{T}}\mathbf{X}(s)}\,ds)}\right.\\ &\quad \left. - \lambda_0\left(\int_0^Y e^{\beta_0^{\mathrm{T}}\mathbf{X}(s)}\,ds\right)\right\}\int_0^Y \mathbf{X}(s)e^{\beta_0^{\mathrm{T}}\mathbf{X}(s)}\,ds\\ &\quad + \Delta\mathbf{S}\left(\int_0^Y e^{\beta_0^{\mathrm{T}}\mathbf{X}(s)}\,ds\right) - \int_0^Y \mathbf{S}\left(\int_0^t e^{\beta_0^{\mathrm{T}}\mathbf{X}(s)}\,ds\right)\\ &\quad \times\lambda_0\left(\int_0^t e^{\beta_0^{\mathrm{T}}\mathbf{X}(s)}\,ds\right)e^{\beta_0^{\mathrm{T}}\mathbf{X}(t)}\,dt,\end{aligned}$$

where

$$\begin{aligned}\mathbf{S}(t) &= -E_{\widetilde{Y},\widetilde{\Delta},\widetilde{\mathbf{X}}}[\widetilde{\mathbf{X}}(\widetilde{Y})\mid\widetilde{R}_0=\log t,\widetilde{\Delta}=1]\\ &\quad - E_{\widetilde{Y},\widetilde{\Delta},\widetilde{\mathbf{X}}}[\widetilde{\mathbf{R}}_0^{(1)}\mid\widetilde{R}_0=\log t,\widetilde{\Delta}=1]\lambda_0'(t)t/\lambda_0(t).\end{aligned}$$

It is easy to verify that $\mathbf{l}^*_\beta$ is the score along a submodel passing (β_0,Λ_0) with tangent direction $\int \mathbf{S}(t)\,d\Lambda_0(t)$ for Λ_0. By the kernel approximation and the zero-moments condition of $K(\cdot)$, we obtain

$$\sqrt{n}E\left[\Delta\mathbf{X}(Y) - \Delta\mathbf{R}_0^{(1)} + \Delta\frac{\mathbf{g}_{10}(Y,\mathbf{X};\beta_0)}{g_{20}(Y,\mathbf{X};\beta_0)} - \Delta\frac{\mathbf{g}_{30}(Y,\mathbf{X};\beta_0)}{g_{40}(Y,\mathbf{X};\beta_0)}\right] = o_p(1).$$

Thus the asymptotic normality and efficiency of $\sqrt{n}(\widehat{\beta}_n-\beta_0)$ hold if we can show that

$$\frac{\partial}{\partial\beta}E\left[\Delta\mathbf{X}(Y) - \Delta\mathbf{R}_0^{(1)} + \Delta\frac{\mathbf{g}_{10}(Y,\mathbf{X};\beta_0)}{g_{20}(Y,\mathbf{X};\beta_0)} - \Delta\frac{\mathbf{g}_{30}(Y,\mathbf{X};\beta_0)}{g_{40}(Y,\mathbf{X};\beta_0)}\right] = -E[\mathbf{l}^*_\beta(Y,\Delta,\mathbf{X})^{\otimes 2}] < 0.$$

Suppose that $E[\mathbf{l}^*_\beta(Y,\Delta,\mathbf{X})^{\otimes 2}]$ is singular. Then there exists a constant vector $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}^{\mathrm{T}}E[\mathbf{l}^*_\beta(Y,\Delta,\mathbf{X})^{\otimes 2}]\boldsymbol{\alpha}=0$. Thus $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{l}^*_\beta=0$ almost surely. We multiply this equation by $d\exp\{-\Lambda_0(\int_0^y e^{\beta_0^{\mathrm{T}}\mathbf{X}(u)}du)\}$ and integrate y from 0 to t. Then $\int_0^t \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{X}(u)e^{\beta_0^{\mathrm{T}}\mathbf{X}(u)}\,du = \chi(\int_0^t e^{\beta_0^{\mathrm{T}}\mathbf{X}(u)}du)$ for some differentiable function χ. We differentiate both sides with respect to $t=T$ to obtain $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{X}(T) = \chi'(\int_0^T e^{\beta_0^{\mathrm{T}}\mathbf{X}(u)}\,du) = \chi'(\epsilon)$. By condition (C.2), $\boldsymbol{\alpha}=\mathbf{0}$; therefore, $\widehat{\beta}_n$ is an asymptotically efficient estimator of β_0.

Proof of Theorem 3

Define $S_{R_0}(t) = \mathrm{P}(R_0>t)$. It follows from the proofs for the asymptotic properties of $\widehat{\beta}_n$ that

$$\begin{aligned}&\sqrt{n}\widehat{\Lambda}_n(t)\\ &= \mathbb{G}_n\left[\int_{-\infty}^{\log t}\frac{\Delta f_{R_0|\Delta}(s)}{S_{R_0}(s)}\,ds\right]\\ &\quad - \mathbb{G}_n\left[\int_{-\infty}^{\log t}\frac{E_{\widetilde{Y},\widetilde{\Delta},\widetilde{\mathbf{X}}}[\widetilde{\Delta}f_{\widetilde{R}_0|\widetilde{\Delta}}(s)]}{S_{R_0}(s)^2}I(R_0>s)\,ds\right]\\ &\quad + \sqrt{n}\int_{-\infty}^{\log t}\left\{\frac{E_{\widetilde{Y},\widetilde{\Delta},\widetilde{\mathbf{X}}}[\widetilde{\Delta}\widetilde{\mathbf{R}}_0^{(1)\mathrm{T}}f'_{\widetilde{R}_0|\widetilde{\Delta},\widetilde{\mathbf{R}}_0^{(1)}}(s)]}{S_{R_0}(s)}\right.\\ &\quad \left. - \frac{E_{\widetilde{Y},\widetilde{\Delta},\widetilde{\mathbf{X}}}[\widetilde{\Delta}f_{\widetilde{R}_0|\widetilde{\Delta}}(s)]E_{\widetilde{Y},\widetilde{\Delta},\widetilde{\mathbf{X}}}[\widetilde{\Delta}\widetilde{\mathbf{R}}_0^{(1)\mathrm{T}}f_{\widetilde{R}_0|\widetilde{\Delta},\widetilde{\mathbf{R}}_0^{(1)}}(s)]}{S_{R_0}(s)^2}\right\}ds\\ &\quad \times(\widehat{\beta}_n-\beta_0)\\ &\quad + \sqrt{n}\int_{-\infty}^{\log t}\frac{E_{\widetilde{Y},\widetilde{\Delta},\widetilde{\mathbf{X}}}[\widetilde{\Delta}f_{\widetilde{R}_0|\widetilde{\Delta}}(s)]}{S_{R_0}(s)}\,ds + o_p(1).\end{aligned}$$

On noting that

$$\int_{-\infty}^{\log t}\frac{E_{\widetilde{Y},\widetilde{\Delta},\widetilde{\mathbf{X}}}[\widetilde{\Delta}f_{\widetilde{R}_0|\widetilde{\Delta}}(s)]}{S_{R_0}(s)}\,ds = \Lambda_0(t),$$

we reach the conclusion of Theorem 3.

[Received August 2006. Revised July 2007.]

REFERENCES

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993), *Efficient and Adaptive Estimation for Semiparametric Models*, New York: Springer.

Buckley, J., and James, I. (1979), "Linear Regression With Censored Data," *Biometrika*, 66, 429–436.

Cox, D. R. (1972), "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, 187–220.

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

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

Fried, L. P., Borhani, N. O., Enright, P., Furberg, C. D., Gardin, J. M., Kronmal, R. A., Kuller, L. H., Manolio, T. A., Mittelmark, M. B., Newman, A., O'Leary, D., Psaty, B., Rautaharju, P., and Tracy, R. (1991), "The Cardiovascular Health Study: Design and Rationale," *The Annals of Epidemiology*, 1, 263–276.

Hart, J. D. (1997), *Nonparametric Smoothing and Lack-of-Fit Tests*, New York: Springer.

Jin, Z., Lin, D. Y., Wei, L. J., and Ying, Z. (2003), "Rank-Based Inference for the Accelerated Failure Time Model," *Biometrika*, 90, 341–353.

Jin, Z., Lin, D. Y., and Ying, Z. (2006), "On Least-Squares Regression With Censored Data," *Biometrika*, 93, 147–161.

Jones, M. C. (1990), "The Performance of Kernel Density Functions in Kernel Distribution Function Estimation," *Statistics and Probability Letters*, 9, 129–132.

Jones, M. C., and Sheather, S. J. (1991), "Using Non-Stochastic Terms to Advantage in Kernel-Based Estimation of Integrated Squared Density Derivatives," *Statistics and Probability Letters*, 11, 511–514.

Jones, M. P. (1997), "A Class of Semiparametric Regressions for the Accelerated Failure Time Model," *Biometrika*, 84, 73–84.

Kalbfleisch, J. D., and Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data* (2nd ed.), Hoboken, NJ: Wiley.

Lai, T. L., and Ying, Z. (1991a), "Rank Regression Methods for Left-Truncated and Right-Censored Data," *The Annals of Statistics*, 19, 531–556.

——— (1991b), "Large-Sample Theory of a Modified Buckley–James Estimator for Regression Analysis With Censored Data," *The Annals of Statistics*, 19, 1370–1402.

Lin, D. Y., and Ying, Z. (1995), "Semiparametric Inference for the Accelerated Life Model With Time-Dependent Covariates," *Journal of Statistical Planning and Inference*, 44, 47–63.

Prentice, R. L. (1978), "Linear Rank Tests With Right-Censored Data," *Biometrika*, 65, 167–179.

Rao, P. (1983), *Nonparametric Functional Estimation*, Orlando, FL: Academic Press.

Reid, N. (1994), "A Conversation With Sir David Cox," *Statistical Science*, 9, 439–455.

Ritov, Y. (1990), "Estimation in a Linear Regression Model With Censored Data," *The Annals of Statistics*, 18, 303–328.

Robins, J. M., and Tsiatis, A. A. (1992), "Semiparametric Estimation of an Accelerated Failure Time Model With Time-Dependent Covariates," *Biometrika*, 79, 311–319.

Schuster, E. F. (1969), "Estimation of a Probability Density Function and Its Derivatives," *The Annals of Mathematical Statistics*, 40, 1187–1195.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall.

Tsiatis, A. A. (1990), "Estimating Regression Parameters Using Linear Rank Tests for Censored Data," *The Annals of Statistics*, 18, 354–372.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, London: Cambridge University Press.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer.

Wei, L. J., Ying, Z., and Lin, D. Y. (1990), "Linear Regression Analysis of Censored Survival Data Based on Rank Tests," *Biometrika*, 77, 845–851.

Yang, S. (1997), "Extended Weighted Log-Rank Estimating Functions in Censored Regression," *Journal of the American Statistical Association*, 92, 977–984.

Ying, Z. (1993), "A Large-Sample Study of Rank Estimation for Censored Regression Data," *The Annals of Statistics*, 21, 76–99.

Zhou, M. (2005), "Empirical Likelihood Analysis of the Rank Estimator for the Censored Accelerated Failure Time Model," *Biometrika*, 92, 492–498.