

Efficient Estimation of Censored Linear Regression Model

BY KANI CHEN

*Department of Mathematics, Hong Kong University of Science and Technology, Kowloon,
 Hong Kong*

makchen@ust.hk

AND YUANYUAN LIN

*Department of Mathematics, Hong Kong University of Science and Technology, Kowloon,
 Hong Kong*

Corresponding author: linyy@ust.hk

SUMMARY

In linear regression or accelerated failure time model, the method of efficient estimation, with or without censoring, has long been overlooked. The main reason is that complications arise from multiple roots of the efficient score and density estimation. In particular, when smoothing is involved, uncertainty in the choice of bandwidth is inevitable. Zeng and Lin (2007) provided a novel efficient estimation method for the accelerated failure time model by maximizing a kernel-smoothed profile likelihood function. This paper proposes a one-step efficient estimation method based on counting process martingale, which has several advantages: it avoids the multiple root problem, the initial estimator is easily available, and it is easy to implement numerically with a built-in inference procedure. The requirement on bandwidth is rather loose and less restrictive than that imposed in Zeng and Lin (2007). A simple and effective data-driven bandwidth selection method is provided. The resulting estimator is proved to be semiparametric efficient with the same asymptotic variance as the efficient estimator when the error distribution is assumed to be known up to a location shift. The asymptotic properties of the proposed method are justified and the asymptotic variance matrix of the regression coefficients is provided in a closed form. Numerical studies with supportive evidence are presented. Applications are illustrated with the well-known PBC data and the Colorado Plateau uranium miners data.

Some key words: Linear regression model; Accelerated failure time model; Estimating equation; One-step efficient estimation; Counting process martingale.

1. INTRODUCTION

The linear regression model is one of the most classical and widely used statistical models. For complete data, the regression parameters are usually estimated by the least squares, rank estimation and M-estimation. Likewise, in the presence of censoring, the linear model or accelerated failure time model, owing to its straightforward interpretation, is an important alternative to the Cox model for regression analysis. The popular estimation approaches are rank methods (e.g. logrank and Gehan's method) and the least squares method (e.g. Buckley-James), see Gehan (1965) and Buckley and James (1979). Other existing methods, such as Tarone-Ware,

Kalbfleisch-Prentice, Peto-Peto and Fleming-Harrington also have clear interpretations and are easy to compute, see Kalbfleisch and Prentice (1980), Peto and Peto (1972), Fleming and Harrington (1991). An M-estimation method for censored linear regression can be found in Zhou (1992). The co-existence of the aforementioned competing methods reflect the fact that it has not proven which method is the best. They are asymptotically efficient only under certain specific error distributions. None of them achieves the semiparametric efficiency bound. Even though some general theory for rank methods has been developed in Tsiatis (1990), Lai and Ying (1991b) and Ying (1993), a specific construction of the semiparametric efficient estimator is not available in the literature.

Recently, Zeng and Lin (2007) developed an elegant efficient estimation method for the accelerated failure time model by maximizing a kernel-smoothed profile likelihood function. This approach is intuitively appealing. It leads to a semiparametric efficient estimator under certain restrictive choice of kernel function and bandwidth.

An easy-to-implement semiparametric efficient estimation, with obvious advantage over the above methods in terms of estimation accuracy, is much desired. The semiparametric efficient estimation, however, is generally associated with the problem of density estimation and computational difficulties. The efficient score with estimated density and its derivative is likely to have multiple roots. Moreover, when smoothing is involved, uncertainty in the choice of bandwidth is inevitable. These and other technical/numerical complications constitute the limitations of efficient estimation.

In spite of the difficulties, efficient estimation is worth pursuing. The most appealing factor is that, when censoring is absent or independent of covariate, the semiparametric efficient estimator is asymptotically as accurate as the parametric efficient estimator when the error distribution is assumed to be known up to a location shift. Consider a linear regression model

$$y_i = \alpha + \beta'x_i + \epsilon_i, \quad 1 \leq i \leq n,$$

where (y_i, x_i, ϵ_i) are *i.i.d* copies of (y, x, ϵ) and x is independent of ϵ . Assume the density of ϵ_i , $f(\cdot)$, is known and let $\psi(\cdot) = \dot{f}(\cdot)/f(\cdot)$. The score $(1, x)' \psi(y - \alpha - \beta'x) = (1, x)' \psi(\epsilon)$ and the estimating equation $\sum_{i=1}^n (1, x_i)' \psi(y_i - \alpha - \beta'x_i) = 0$ provide the efficient estimation. The asymptotic variance of the estimator of β is $\{var(x)\mathcal{I}\}^{-1}$, where $\mathcal{I} = E\{\psi^2(\epsilon)\}$ is the Fisher information for the location parameter. When f is entirely unknown, α is not identifiable and is better absorbed into the error term for proper formulation. The score function for β is $(x - Ex)\psi(y - \beta'x) = (x - Ex)\psi(\epsilon)$. Heuristically, the efficient estimator of β can be obtained by solving

$$\sum_{i=1}^n (x_i - \bar{x}) \hat{\psi}(y_i - \beta'x_i) = 0,$$

where $\hat{\psi}(\cdot)$ is a proper estimator of $\psi(\cdot)$. However, numerical complications arise here. First, the curve estimation itself can be difficult. Second, the above estimating function is sensitive to the curve estimation and deviation can result in bias of the estimator of β . Moreover, the estimating equation may have multiple roots, depending on the shape of the curve $\hat{\psi}$. As a result, efficient estimation becomes unfavorable compared with least squares or rank estimation. We note that the asymptotic variance of the semiparametric efficient estimator $\{var(x)\mathcal{I}\}^{-1}$ is the same as that of the parametric efficient estimator.

With the ample theoretical and numerical development in smoothing techniques as well as in survival analysis, it is now timely to tackle the technical difficulties and find a more reliable numerical procedure for efficient estimation. This paper proposes a competitive alternative statis-

tical procedure: a one-step semiparametric efficient estimation. There are several advantages to this method. The high order accuracy of curve estimation, such as that imposed in Zeng and Lin (2007), is not required. Moreover, even if the curve estimation is not consistent, the estimator of the regression parameter is still consistent and asymptotically normal. The choice of bandwidth, which can be rather delicate in the context of curve estimation, is not restrictive. Moreover, the curve estimation is for the hazard function rather than the density function and can be naturally handled using counting process martingale. A simple and effective data-driven bandwidth selection method is also provided. Although the one-step estimation requires an appropriate initial estimator, this is not a problem as many ideal candidates exist. Using the one-step estimation, we avoid solving the estimating equation and therefore avoid the multiple root problem. This method also has a built-in variance estimation. The computation is straightforward. In summary, the proposed method provides a competitive alternative to the existing methods.

It is worth noting that even though Buckley-James, the log-rank and Gehan's methods do not involve smoothing, the estimation of asymptotic variances does if the plug-in rules are used. An alternative way is by the resampling procedures proposed recently in the literature, see Jin et al. (2000).

In the next section, we describe the one-step efficient estimation procedure with theoretical justification. Numerical studies with supportive evidence are presented in Section 3. Section 4 describes the applications to the well-known PBC data and the Colorado Plateau uranium miners data. A few closing remarks are given in Section 5. All proofs are deferred to the Appendix.

2. ONE-STEP EFFICIENT ESTIMATION

Let (Y^*, X) be the response-covariate pair of random variables, satisfying

$$Y^* = \beta'X + \epsilon, \quad (1)$$

where X is a d -dimensional covariate, β is the d -vector regression parameter to be estimated and ϵ is the unobservable error term independent of X . Y^* in model (1) could be some transformation of survival times. When the log transformation is used, the resulting model is often regarded as an accelerated failure time model. Let C be the censoring variable. As usual, we assume that C is conditionally independent of Y^* given X . Set $Y = \min(Y^*, C)$, $\delta = I(Y^* \leq C)$, $N(t) = \delta I(Y \leq t)$, and $Y(t) = I(Y \geq t)$. $I(\cdot)$ is the indicator function throughout this paper. Then,

$$M(t) \equiv N(t) - \int_{-\infty}^t Y(s) \lambda(s - \beta'X) ds$$

is a counting process martingale, where $\lambda(\cdot)$ is the hazard function of ϵ . Let $\Lambda(\cdot)$ be the cumulative hazard function of ϵ . For ease of presentation, we also define $N(t, \beta) = \delta I(Y - \beta'X \leq t)$ and $Y(t, \beta) = I(Y - \beta'X \geq t)$. Then,

$$M(t, \beta_0) \equiv N(t, \beta_0) - \int_{-\infty}^t Y(s, \beta_0) \lambda(s) ds$$

is also a counting process martingale based on ϵ .

In the presence of censoring, the observations are (Y_i, δ_i, X_i) , $1 \leq i \leq n$, which are iid copies of (Y, δ, X) . A linear regression model without censoring is a special case with $P(C = \infty) = 1$.

Consider the following estimating function

$$U_\phi(\beta) = \sum_{i=1}^n \int_0^\tau \phi(t) \left\{ X_i - \frac{\sum_{j=1}^n X_j Y_j(t, \beta)}{\sum_{j=1}^n Y_j(t, \beta)} \right\} dN_i(t, \beta)$$

$$= \sum_{i=1}^n \int_0^\tau \phi(t - \beta' X_i) \left\{ X_i - \frac{\sum_{j=1}^n X_j Y_j(t - \beta' X_i + \beta' X_j)}{\sum_{j=1}^n Y_j(t - \beta' X_i + \beta' X_j)} \right\} dN_i(t), \quad (2)$$

where $\tau = \inf\{t : P(\epsilon < t, C - \beta'_0 X < t) = 1\}$. All the popular estimation procedures can be viewed as special cases with particular choices of $\phi(t)$. Choosing $\phi(t) = 1$ yields the well-known log-rank estimation and choosing $\phi(t) = \sum_{j=1}^n Y_j(t, \beta)$ leads to Gehan's estimation.

It will be shown later that, under certain regularity conditions, choosing $\phi(t) = \dot{\lambda}(t)/\lambda(t)$ leads to the efficient estimation if $\lambda(\cdot)$ satisfies certain smoothness conditions. When the distribution of ϵ is unknown, it is natural to plug in a proper estimator of $\dot{\lambda}(\cdot)/\lambda(\cdot)$. Such estimation shall inevitably involve smoothing techniques. Let $K(\cdot)$ be an infinitely differentiable symmetric kernel function with support in $[-1, 1]$ and h be the bandwidth. Let $K_h(t) = (1/h)K(t/h)$. Define

$$\hat{\lambda}(t, \beta) = \int_0^\tau K_h(s - t) d\hat{\Lambda}(s, \beta), \quad \hat{\dot{\lambda}}(t, \beta) = \frac{1}{h^2 \kappa_2} \int_0^\tau (s - t) K_h(s - t) d\hat{\Lambda}(s, \beta)$$

and $\hat{\psi}(t, \beta) = \hat{\dot{\lambda}}(t, \beta)/\hat{\lambda}(t, \beta)$, where $\kappa_2 = \int_{-1}^1 u^2 K(u) du$. $\hat{\Lambda}(\cdot)$ is the nonparametric estimator of the cumulative hazard function $\Lambda(\cdot)$ introduced by Nelson (1972) and generalized by Aalen (1978). The kernel smoothing estimator of $\lambda(\cdot)$ is essentially a way of smoothing the increments of $\hat{\Lambda}(\cdot)$. Moreover, $\hat{\lambda}(\cdot)$ and $\hat{\dot{\lambda}}(\cdot)$ are consistent estimates for $\lambda(\cdot)$ and $\dot{\lambda}(\cdot)$, see Ramlau-Hansen (1983) and Jiang and Doksum (2003). A straightforward way of computing an estimator is to directly replace $\phi(t, \beta)$ with $\hat{\psi}(t, \beta)$ in estimating equation (2) and solve the equation for β . However, such a direct procedure may not be numerically feasible. In general, this equation is likely to have multiple solutions. The conventional numerical tools, such as the Newton-Raphson method, become unreliable.

To avoid numerical complications, we propose the following one-step efficient estimation procedure. Here and throughout the paper, $v^{\otimes 2} = vv'$ for any vector v of dimension d . $\|\cdot\|$ is the Euclidean norm and β_0 be the true value of β . Let β_1 be certain initial estimator of β . Denote $\sum_{j=1}^n X_j Y_j(t, \beta) / \sum_{j=1}^n Y_j(t, \beta)$ in estimating equation (2) as $\bar{X}(t, \beta)$. Let

$$V(\beta_1) = \int_0^\tau \hat{\psi}^2(t, \beta_1) \sum_{j=1}^n \{X_j - \bar{X}(t, \beta_1)\}^{\otimes 2} Y_j(t, \beta_1) \sum_{i=1}^n dN_i(t, \beta_1) / \sum_{j=1}^n Y_j(t, \beta_1).$$

Define the proposed estimator,

$$\hat{\beta} = \beta_1 + V^{-1}(\beta_1) U_{\hat{\psi}}(\beta_1). \quad (3)$$

Some regularity conditions are assumed here.

- (i) $P(\epsilon > \tau) > 0$ and τ is assumed to be finite;
- (ii) The covariate X is bounded, i.e, $P(\|X\| \leq K) = 1$ for some $0 < K < \infty$;
- (iii) The hazard function $\lambda(t)$ is bounded away from 0 for $0 \leq t \leq \tau$ and is continuously differentiable up to the second order;
- (iv) The bandwidth $h = n^r$ with $r \in (-1/3, 0)$;
- (v) The initial estimator $\beta_1 = \beta_0 + O_p(n^{-1/2})$.

THEOREM 1. *Assume the above conditions (i)-(v) hold. In fact,*

$$\hat{\beta} = \beta_0 + \Sigma^{-1} \sum_{i=1}^n \xi_i / n + r_n, \quad (4)$$

where the remainder $r_n = O_p(1/\sqrt{n^2 h^3} + h^2/\sqrt{n})$,

$$\xi_i = \int_0^\tau \{\dot{\lambda}(t)/\lambda(t)\} \{X_i - \bar{X}(t, \beta_0)\} dN_i(t, \beta_0)$$

and

$$\Sigma = \int_0^\tau \left\{ \dot{\lambda}(t)/\lambda(t) \right\}^2 E \{X - E(X|Y \geq t + \beta'_0 X)\}^{\otimes 2} d\Lambda(t),$$

which can be consistently estimated by $-V(\beta_1)/n$. It follows that $\hat{\beta}$ is consistent for β_0 as $n \rightarrow \infty$, and

$$\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow N(0, \Sigma^{-1})$$

in distribution as $n \rightarrow \infty$.

Remark 1. Conditions (i)-(ii) are regularity conditions. Condition (i) implies that at the end of the study, there is a positive chance that an individual will have survived without being censored. Such restriction frequently appears in the literature and is used to avoid the usual technical difficulties of a possible tail instability. Condition (iii) is the smoothness condition for the underlying density function. $V(\beta)$ and Σ are nonsingular under conditions (i)-(iii). This is basically in the same spirit as Theorem 8.4.1 of Fleming and Harrington (1991). Condition (iv) is needed for curve estimation. Condition (v) implies that the initial estimator must be close to β_0 to ensure convergence of the desired semiparametric efficient estimator. Condition (v) is satisfied by various existing estimators such as, for example, the logrank or Gehan's estimator.

Remark 2. When assuming that the error distribution in model (1) is known up to a location shift, the efficient score for β is

$$S(\beta) = \sum_{i=1}^n \int_0^\tau \{\dot{\lambda}(t)/\lambda(t)\} \{X_i dN_i(t, \beta) - X_i Y_i(t, \beta) \lambda(t) dt\}.$$

And the asymptotic variance of the efficient estimator of β is Σ^{-1} , which is the same as the asymptotic variance of the semiparametric efficient estimator. The proposed estimator is semiparametric-efficient. Most importantly, it is as efficient as the parametric efficient estimator, when the error distribution is known up to a location shift.

Remark 3. Zeng and Lin (2007) provided a novel method to compute an efficient estimator for the accelerated failure time model. There are certain theoretical restrictive conditions on the kernel function and bandwidth. In particular, the asymptotic normality relies on the assumptions that the first $(m-1)$ moments of the kernel function are 0 for some $m > 3$ and the bandwidth $h = n^\nu$ with $\nu \in (-1/(2m), -1/6)$. Such assumptions could be too stringent in practice. The proposed efficient estimation procedure in this paper is valid under much looser requirements on the kernel function and bandwidth. The requirement on the bandwidth $h = n^r$ with $r \in (-1/3, 0)$ is only a necessary condition for the consistency of the estimation of $\dot{\lambda}(\cdot)$, as the variance of $\hat{\lambda}(\cdot)$ is at the order of $1/(nh^3)$. And the positivity and symmetry of the kernel function is standard in the literature.

Remark 4. Although the estimation of $\lambda(\cdot)$ and $\dot{\lambda}(\cdot)$ may theoretically have different orders of optimal bandwidth, using different choices of bandwidth to estimate them separately only results in marginal improvement. In this paper, choosing the same bandwidth to estimate $\lambda(\cdot)$ and $\dot{\lambda}(\cdot)$, due to its analytical and computational simplicity, yields desirable and favorable results.

Remark 5. The remainder term in (4) is a degenerate U-statistics which is zero correlated with the main iid term plus negligible terms by a careful evaluation in the Appendix. It then induces

a data-driven bandwidth selection method by minimizing the remainder term r_n over h . As the main iid terms do not depend on h , it is equivalent to minimize

$$\left\| \frac{1}{n} \Sigma^{-1} \sum_{i=1}^n \xi_i + r_n \right\|$$

over h . The proposed data-driven bandwidth selection method is simple and effective. Moreover, in view of (3), one can write

$$V^{-1}(\beta_1)U_{\hat{\psi}}(\beta_1) = -(\beta_1 - \beta_0) + \frac{1}{n} \Sigma^{-1} \sum_{i=1}^n \xi_i + r_n. \quad (5)$$

Under condition (v), $\beta_1 - \beta_0 = (1/n) \sum_{i=1}^n Z_i + O_p(1/n)$ where Z_i are iid mean zero terms. Then,

$$V^{-1}(\beta_1)U_{\hat{\psi}}(\beta_1) = -\frac{1}{n} \sum_{i=1}^n Z_i + \frac{1}{n} \Sigma^{-1} \sum_{i=1}^n \xi_i + r_n + O_p\left(\frac{1}{n}\right). \quad (6)$$

Most importantly, it can be verified that Z_i is also zero correlated with the degenerate U-statistics leading term of r_n which is at the order of $(1/\sqrt{n^2 h^3} + h^2/\sqrt{n})$. Then the mean squared error of $nV^{-1}(\beta_1)U_{\hat{\psi}}(\beta_1)/\sqrt{n}$

$$\text{MSE}(h) = \sigma^2 + O\left(\frac{1}{nh^3} + h^4\right) + o\left(\frac{1}{nh^3} + h^4\right) + O\left(\frac{1}{n}\right) \quad (7)$$

decreases till h is at the order of $n^{-1/3}$, where σ^2 is the variance of $-\sum_{i=1}^n Z_i/\sqrt{n} + \Sigma^{-1} \sum_{i=1}^n \xi_i/\sqrt{n}$ which is independent of h . As a result, it is intuitively appealing to select the bandwidth by minimizing $V^{-1}(\beta_1)U_{\hat{\psi}}(\beta_1)$ in real implementation. The computation is straightforward. Supportive simulation results can be found in section 3.

Remark 6. Consider a generalized linear model

$$Y = g(\beta' X) + \epsilon,$$

where $g(\cdot)$ is a known differentiable link function, Y is the response variable, X , β and ϵ are given in model (1). Under certain regularity conditions, the regression parameter β is identifiable. Our proposed method can be applied to such a generalized linear model in a straightforward fashion.

3. NUMERICAL STUDIES

Simulation studies are conducted to examine the finite sample performance of our proposed one-step efficient estimation method. Similar to Jin et al. (2006) and Zeng and Lin (2007), the studies are based on the following model

$$\log(T) = 2 + \beta_1 X_1 + \beta_2 X_2 + \epsilon, \quad (8)$$

where X_1 is Bernoulli with probability of success 0.5, X_2 is normal with mean 0 and standard deviation 0.5 and $(\beta_1, \beta_2) = (-1, -1)$. X_1 and X_2 are independent. We consider six error distributions: ϵ follows the standard normal distribution; ϵ follows (standard) extreme-value distribution; ϵ follows the distribution corresponding to the proportional odds model (the standard logistic distribution); and e^ϵ follows Weibull distribution with scale parameter 1 and shape parameter 1, denoted by Weibull(1,1); ϵ follows mixtures of $N(0, 1)$ and $N(0, 9)$ with mixing probabilities (0.5, 0.5) and (0.95, 0.05), denoted by $0.5N(0, 1) + 0.5N(0, 9)$ and

0.95 $N(0, 1) + 0.05N(0, 9)$. Censoring times are generated from the Uniform $[0, \tau]$, where τ was chosen to produce a 25% censoring rate. We set the sample size $n = 200$. We use the kernel function $K(\cdot)$ to be the standard normal density function (Gaussian kernel) in our simulation. We let the logrank estimator and Gehan's estimator to be the initial estimator in our procedure. We used the recommended optimal bandwidths for the profile-likelihood approach in Zeng and Lin (2007). The simulation results are based on 1000 replications and can be summarized in the following tables.

(Insert Tables 1, 2, 3 here)

It can be seen that the proposed estimator is asymptotically as efficient as that of Zeng and Lin (2007). And the proposed estimator is generally stable and not sensitive to the choice of bandwidth. One can see from Tables 1 and 2 that the proposed method is not sensitive to the initial estimator. Both logrank and Gehan's initial estimators provide favorable results. Table 3 reports that the proposed bandwidth selection method works well compared with the existing methods. Although for logistic and $0.5N(0, 1) + 0.5N(0, 9)$ error distributions the variance is slightly underestimated with small bandwidths, the proposed data-driven bandwidth selection method gives accurate estimation. Further simulation shows that the proposed estimator is not sensitive to the choice of kernel function. Indeed, our proposed procedure performs well in finite sample studies.

4. APPLICATIONS

We first present an analysis of the well-known Mayo primary biliary cirrhosis (PBC) data (Fleming and Harrington, 1991, app. D.1). The dataset contains information on the survival time T and prognostic factors for 418 patients. Analogous to Jin et al. (2003, 2006) and Zeng and Lin (2007), we fit the accelerated failure time model with five covariates, age, log(albumin), log(bilirubin), edema and log(protime). The estimates of the regression parameters can be obtained using the proposed method with the Gaussian kernel function and the bandwidths $\sigma n^{-1/5}$, $\sigma n^{-1/7}$ and the bandwidth selected by the proposed data-driven selection method in Remark 5, denoted by h_{dd} . In these two examples, σ is the sample standard deviation of $\log(T)$. The results are displayed in Table 4. The parameter estimates are generally not sensitive to the choice of bandwidth. And the variance estimation with bandwidth h_{dd} is similar to that of Zeng and Lin (2007).

Table 4: Accelerated failure time regression for the Mayo PBC data.

Covariate	$h = \sigma n^{-1/5}$		$h = \sigma n^{-1/7}$		h_{dd}	
	Est	SE	Est	SE	Est	SE
Age	0.0282	0.0056	0.0290	0.0060	0.0293	0.0064
log(albumin)	-1.8637	0.4574	-1.8035	0.4890	-1.7214	0.5230
log(bilirubin)	0.5861	0.0591	0.5847	0.0635	0.5872	0.0679
Edema	0.7415	0.2075	0.7415	0.2248	0.7458	0.2430
log(protime)	2.2743	0.5998	2.4225	0.6460	2.4548	0.6941

For further illustration, we apply the proposed method to analyze the Colorado Plateau uranium miners data. The data set is collected to study the effects of radon exposure and smoking on the rates of lung cancer. The detailed description of the data set can be found in Langholz and Goldstein (1996). The study consists of 3347 Caucasian male miners who worked underground for at least one month in the uranium mines of the four-state Colorado Plateau area. For each

subject, the information includes the age at entry to the study, the cumulative radon exposure, the cumulative smoking in number of packs and the death information. In this study, a total of 258 miners died of lung cancer. Subjects who died of lung cancer are taken to be failures and all others were censored at their exit times. Let X denote the cumulative radon exposure in 100 working level months (WLMs), Z be the cumulative smoking in 1000 packs and W be the age at entry to the study. We fit the following model

$$\log(T) = \beta_1 X + \beta_2 W + \beta_3 Z + \epsilon$$

using the proposed method with the Gaussian kernel function. Table 5 presents the estimation results which are generally not sensitive to the choice of bandwidth. It is seen that the cumulative radon exposure and age have significant negative effect on the survival time and smoking is hazardous.

Table 5: Analysis of Colorado Plateau Uranium Miners Data.

Covariate	$h = \sigma n^{-1/5}$		$h = \sigma n^{-1/7}$		h_{dd}	
	Est	SE	Est	SE	Est	SE
X	-0.0214	0.0008	-0.0223	0.0013	-0.0218	0.0010
W	-0.0261	0.0013	-0.0268	0.0020	-0.0266	0.0016
Z	-0.0077	0.0012	-0.0066	0.0019	-0.0070	0.0016

NOTE: Est denotes the parameter estimate; SE is the estimated standard error.

5. CONCLUDING REMARKS

This paper proposes a one-step efficient estimation and inference approach for linear regression models or accelerated failure time models with censored data. The goal of the paper is to advocate such a numerical procedure of efficient estimation that is both reliable and easy to implement, as shown in the simulation studies and real examples. We understand that efficient estimation has its limitation in terms of, for example, robustness. However, a truncated version of the curve estimation could be considered. For example, when the estimator of $\lambda(\cdot)$ is too close to 0, we could replace the weight function by 1. Such modified versions may result in more stable, albeit less efficient, estimators.

We aim to develop a mature program in S-plus and SAS for practitioners or non-experts to compute the semiparametric efficient estimators. Our future work will consider further extensions of the method to partial linear models or general transformation models.

APPENDIX

The proof relies on counting process martingale techniques, see for example, Andersen and Gill (1982) and Chen (2004). More notation is needed. $|\cdot|$ for a vector or matrix means the sum of the absolute values of all elements. Define

$$\begin{aligned}
 U_{tr}(\beta) &= \sum_{i=1}^n \int_0^\tau \{\dot{\lambda}(t)/\lambda(t)\} \{X_i - \bar{X}(t, \beta)\} dN_i(t, \beta) \\
 &= \sum_{i=1}^n \int_0^\tau \{\dot{\lambda}(t)/\lambda(t)\} \{X_i - \bar{X}(t, \beta)\} dM_i(t, \beta), \\
 V_{tr}(\beta) &= \int_0^\tau \left\{ \dot{\lambda}(t)/\lambda(t) \right\}^2 \sum_{j=1}^n \{X_j - \bar{X}(t, \beta)\}^{\otimes 2} Y_j(t, \beta) \sum_{i=1}^n dN_i(t, \beta) / \sum_{j=1}^n Y_j(t, \beta).
 \end{aligned}$$

Let $U_{est}(\beta)$ be defined the same as $U_{tr}(\beta)$ except with $\dot{\lambda}(t)/\lambda(t)$ replaced by $\hat{\lambda}(t)/\hat{\lambda}(t)$ defined in Section 2. Set $\bar{N}(t, \beta) = (1/n) \sum_{i=1}^n N_i(t, \beta)$. The following lemma provides the approximations that will be used later.

LEMMA 1. Assume conditions (i)-(iv) hold. Then, there exists a sequence $C_n \rightarrow \infty$ such that,

$$\sup_{\|\beta - \beta_0\| \leq C_n n^{-1/2}} \left| \frac{1}{n} U_{est}(\beta) - \frac{1}{n} U_{tr}(\beta) \right| = O_p\left(\frac{1}{\sqrt{n^2 h^3}} + \frac{h^2}{\sqrt{n}}\right); \quad (\text{A.1})$$

$$\sup_{\|\beta - \beta_0\| \leq C_n n^{-1/2}} |nV^{-1}(\beta) - nV_{tr}^{-1}(\beta)| = o_p(1). \quad (\text{A.2})$$

Proof. Denote $\tilde{Y} = Y - \beta'X$. By the boundedness of $(\tilde{Y} - t)/h$ and $(\tilde{Y} - t)K_h(\tilde{Y} - t)$, $\text{var}\{\delta K_h(\tilde{Y} - t)\} = O(1/h)$ and $\text{var}\{\delta(\tilde{Y} - t)K_h(\tilde{Y} - t)\} = O(h)$ for any fixed t . The empirical approximation analogous to the proof of Lemma 1 of Chen (2004, p. 1523) can be applied to show, for every sequence B_n with $B_n \rightarrow 0$, that

$$\sup_{\|\beta - \beta_0\| \leq B_n, 0 \leq t \leq \tau} \left| \int_0^\tau K_h(s - t) d\bar{N}(s, \beta) - E \left\{ \int_0^\tau K_h(s - t) d\bar{N}(s, \beta) \right\} \right| = O_p\left(\frac{1}{\sqrt{nh}}\right); \quad (\text{A.3})$$

$$\begin{aligned} \sup_{\|\beta - \beta_0\| \leq B_n, 0 \leq t \leq \tau} \left| \int_0^\tau (s - t) K_h(s - t) d\bar{N}(s, \beta) \right. \\ \left. - E \left\{ \int_0^\tau (s - t) K_h(s - t) d\bar{N}(s, \beta) \right\} \right| = O_p(\sqrt{h/n}). \end{aligned} \quad (\text{A.4})$$

Observe that

$$\hat{\lambda}(t) = \frac{1}{h^2 \kappa_2} \int_0^\tau \frac{(s - t) K_h(s - t)}{(1/n) \sum_{j=1}^n I(\tilde{Y}_j \geq s)} d\bar{N}(s, \beta); \quad (\text{A.5})$$

$$\hat{\lambda}(t) = \int_0^\tau \frac{K_h(s - t)}{(1/n) \sum_{j=1}^n I(\tilde{Y}_j \geq s)} d\bar{N}(s, \beta). \quad (\text{A.6})$$

Using (A.3) and (A.4), it can be shown that, for every sequence B_n with $B_n \rightarrow 0$,

$$\sup_{\|\beta - \beta_0\| \leq B_n, 0 \leq t \leq \tau} |\hat{\lambda}(t) - E\hat{\lambda}(t)| = O_p\left(\frac{1}{\sqrt{nh^3}}\right); \quad (\text{A.7})$$

$$\sup_{\|\beta - \beta_0\| \leq B_n, 0 \leq t \leq \tau} |\hat{\lambda}(t) - E\hat{\lambda}(t)| = O_p\left(\frac{1}{\sqrt{nh}}\right) \quad (\text{A.8})$$

by appealing to Theorem 2.4.3 of van der Vaart and Wellner (1996, p.123).

Next, in order to show (A.1), one can write

$$\begin{aligned} \frac{1}{n} U_{est}(\beta) - \frac{1}{n} U_{tr}(\beta) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left[\{\hat{\lambda}(t) - \dot{\lambda}(t)\} / \lambda(t) \right] \{X_i - \bar{X}(t, \beta)\} dM_i(t, \beta) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \int_0^\tau \{\dot{\lambda}(t) / \lambda^2(t)\} \{\hat{\lambda}(t) - \lambda(t)\} \{X_i - \bar{X}(t, \beta)\} dM_i(t, \beta) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \int_0^\tau \{\hat{\lambda}(t) - \dot{\lambda}(t)\} \{\hat{\lambda}(t) - \lambda(t)\} / \lambda^2(t) \{X_i - \bar{X}(t, \beta)\} dM_i(t, \beta) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \int_0^\tau \hat{\lambda}(t) \{\hat{\lambda}(t) - \lambda(t)\}^2 / \{\lambda^2(t) \hat{\lambda}(t)\} \{X_i - \bar{X}(t, \beta)\} dM_i(t, \beta) \\ &= \Pi_1 - \Pi_2 - \Pi_3 + \Pi_4, \quad \text{say.} \end{aligned} \quad (\text{A.9})$$

We first show for any sequence B_n with $B_n \rightarrow 0$,

$$\sup_{\|\beta - \beta_0\| \leq B_n} |\Pi_1 - E(\Pi_1)| = O_p\left(\frac{1}{\sqrt{n^2 h^3}} + \frac{h^2}{\sqrt{n}}\right). \quad (\text{A}\cdot 10)$$

Let R_n be the bias of $\hat{\lambda}(t)$. Using the delta method and plugging in (A.5), one can write

$$\begin{aligned} \Pi_1 &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \frac{\hat{\lambda}(t) - E\hat{\lambda}(t) + E\hat{\lambda}(t) - \dot{\lambda}(t)}{\lambda(t)} \right\} \{X_i - \bar{X}(t, \beta)\} dM_i(t, \beta) \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \sum_{j=1}^n \left[\frac{\delta_j(\tilde{Y}_j - t) K_h(\tilde{Y}_j - t)}{P(\tilde{Y}_1 \geq \tilde{Y}_j | \tilde{Y}_j)} - E \left\{ \frac{\delta_j(\tilde{Y}_j - t) K_h(\tilde{Y}_j - t)}{P(\tilde{Y}_1 \geq \tilde{Y}_j | \tilde{Y}_j)} \right\} \right] \\ &\quad \frac{X_i - \bar{X}(t, \beta)}{nh^2 \kappa_2 \lambda(t)} dM_i(t, \beta) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \int_0^\tau \sum_{j=1}^n \left(\frac{\delta_j(\tilde{Y}_j - t) K_h(\tilde{Y}_j - t)}{\{P(\tilde{Y}_1 \geq \tilde{Y}_j | \tilde{Y}_j)\}^2} \left\{ \frac{1}{n} \sum_{k=1}^n I(\tilde{Y}_k \geq \tilde{Y}_j) - P(\tilde{Y}_1 \geq \tilde{Y}_j | \tilde{Y}_j) \right\} \right. \\ &\quad \left. - E \left[\frac{\delta_j(\tilde{Y}_j - t) K_h(\tilde{Y}_j - t)}{\{P(\tilde{Y}_1 \geq \tilde{Y}_j | \tilde{Y}_j)\}^2} \left\{ \frac{1}{n} \sum_{k=1}^n I(\tilde{Y}_k \geq \tilde{Y}_j) - P(\tilde{Y}_1 \geq \tilde{Y}_j | \tilde{Y}_j) \right\} \right] \right) \\ &\quad \frac{X_i - \bar{X}(t, \beta)}{nh^2 \kappa_2 \lambda(t)} dM_i(t, \beta) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \int_0^\tau \sum_{j=1}^n \left(\frac{\delta_j(\tilde{Y}_j - t) K_h(\tilde{Y}_j - t) \left\{ (1/n) \sum_{k=1}^n I(\tilde{Y}_k \geq \tilde{Y}_j) - P(\tilde{Y}_1 \geq \tilde{Y}_j | \tilde{Y}_j) \right\}^2}{\{P(\tilde{Y}_1 \geq \tilde{Y}_j | \tilde{Y}_j)\}^2 (1/n) \sum_{k=1}^n I(\tilde{Y}_k \geq \tilde{Y}_j)} \right. \\ &\quad \left. - E \left[\frac{\delta_j(\tilde{Y}_j - t) K_h(\tilde{Y}_j - t) \left\{ (1/n) \sum_{k=1}^n I(\tilde{Y}_k \geq \tilde{Y}_j) - P(\tilde{Y}_1 \geq \tilde{Y}_j | \tilde{Y}_j) \right\}^2}{\{P(\tilde{Y}_1 \geq \tilde{Y}_j | \tilde{Y}_j)\}^2 (1/n) \sum_{k=1}^n I(\tilde{Y}_k \geq \tilde{Y}_j)} \right] \right) \\ &\quad \frac{X_i - \bar{X}(t, \beta)}{nh^2 \kappa_2 \lambda(t)} dM_i(t, \beta) + \frac{R_n}{n} \sum_{i=1}^n \int_0^\tau \frac{X_i - \bar{X}(t, \beta)}{\lambda(t)} dM_i(t, \beta) \\ &= \Theta_1 - \Theta_2 + \Theta_3 + \Theta_4, \quad \text{say,} \end{aligned} \quad (\text{A}\cdot 11)$$

where $R_n = O(h^2)$. By expressing Θ_1 in the form of a first order degenerate U-statistic, one can write

$$\Theta_1 = \frac{1}{n} \sum_{1 \leq i < j \leq n} g_{ij} + \frac{1}{n} \sum_{i=1}^n g_{ii}, \quad (\text{A}\cdot 12)$$

where

$$g_{ij} = \frac{\delta_i \{h_{ij} - E(h_{ij} | X_i, Y_i, \delta_i)\} \{X_i - \bar{X}(\tilde{Y}_i, \beta)\}}{nh^2 \kappa_2 \lambda(\tilde{Y}_i)}, \quad h_{ij} = \frac{\delta_j(\tilde{Y}_j - \tilde{Y}_i) K_h(\tilde{Y}_j - \tilde{Y}_i)}{P(\tilde{Y}_1 \geq \tilde{Y}_j | \tilde{Y}_j)}.$$

Here $g_{ii} = 0$ almost surely for $i = 1, \dots, n$. By conditions (i)-(iv) and the boundedness of $(\tilde{Y} - t)K_h(\tilde{Y} - t)$ for all t , $\text{var}(\Theta_1) = \text{var}(g_{ij}) = O(1/(n^2 h^3))$. It can be easily checked that $\text{var}(\Theta_4) = O(h^4/n)$. Therefore, the uniform convergence

$$\sup_{\|\beta - \beta_0\| \leq B_n} |\Theta_1 - E(\Theta_1)| = O_p\left(\frac{1}{\sqrt{n^2 h^3}}\right), \quad \sup_{\|\beta - \beta_0\| \leq B_n} |\Theta_4 - E(\Theta_4)| = O_p\left(\frac{h^2}{\sqrt{n}}\right) \quad (\text{A}\cdot 13)$$

can be shown by appealing to Theorem 2.4.3 of van der Vaart and Wellner (1996, p.123).

It follows from Lemma 1 of Ying (1993) that

$$\sup_{\|\beta - \beta_0\| \leq B_n, 0 \leq t \leq \tau} \left| (1/n) \sum_{j=1}^n I(\tilde{Y}_j \geq t) - P(\tilde{Y}_1 \geq t) \right| = O_p(n^{-1/2}). \quad (\text{A}\cdot 14)$$

Using (A·14), it can be shown in a similar fashion that

$$\sup_{\|\beta - \beta_0\| \leq B_n} |\Theta_2 - E(\Theta_2)| = o_p\left(\frac{1}{\sqrt{n^2 h^3}}\right). \quad (\text{A}\cdot 15)$$

Consider Θ_3 , under condition (i), $\inf_{0 \leq t \leq \tau} P(Y \geq t) > 0$. Then there exists a constant $c > 0$ such that $P(Y \geq t) > c$ for $0 \leq t \leq \tau$. In view of (A·14), $(1/n) \sum_{j=1}^n I(\tilde{Y}_j \geq t) > C$ for some constant $C > 0$ and all $0 \leq t \leq \tau$. Together with (A·7),

$$\begin{aligned} & \sup_{\|\beta - \beta_0\| \leq B_n} |\Theta_3| \\ & \leq \frac{C}{\sqrt{n}} \sup_{\|\beta - \beta_0\| \leq B_n, 0 \leq t \leq \tau} \left| \sum_{j=1}^n \left[\frac{\delta_j(\tilde{Y}_j - t) K_h(\tilde{Y}_j - t)}{n h^2 \kappa_2 \{P(\tilde{Y}_1 \geq \tilde{Y}_j | \tilde{Y}_j)\}^2} - E \left\{ \frac{\delta_j(\tilde{Y}_j - t) K_h(\tilde{Y}_j - t)}{n h^2 \kappa_2 \{P(\tilde{Y}_1 \geq \tilde{Y}_j | \tilde{Y}_j)\}^2} \right\} \right] \right| \\ & \sup_{\|\beta - \beta_0\| \leq B_n, 1 \leq j \leq n} \left| \frac{1}{n} \sum_{k=1}^n I(\tilde{Y}_k \geq \tilde{Y}_j) - P(\tilde{Y}_1 \geq \tilde{Y}_j | \tilde{Y}_j) \right|^2 = o_p\left(\frac{1}{\sqrt{n^3 h^3}}\right) \quad (\text{A}\cdot 16) \end{aligned}$$

for some constant $C > 0$. Thus, (A·10) is proved.

It can be shown in a similar fashion that

$$\begin{aligned} & \sup_{\|\beta - \beta_0\| \leq B_n} |\Pi_2 - E(\Pi_2)| = O_p\left(\frac{1}{\sqrt{n^2 h}} + \frac{1}{\sqrt{n^3 h^2}} + \frac{h^2}{\sqrt{n}}\right); \\ & \sup_{\|\beta - \beta_0\| \leq B_n} |\Pi_i - E(\Pi_i)| = o_p\left(\frac{1}{\sqrt{n^2 h^3}} + \frac{h^2}{\sqrt{n}}\right) \end{aligned}$$

for $i = 3, 4$. Finally, let $m_1(\beta)$ be the expectation of Π_1 . Since $m_1(\beta)$ is continuous in β and $m_1(\beta_0) = 0$, by the Taylor expansion, for every sequence B_n with $B_n \rightarrow 0$ and any $0 < \delta < 1$,

$$\sup_{\|\beta - \beta_0\| \leq B_n} \|m_1(\beta) - G_n(\beta_0)(\beta - \beta_0)\| = o(B_n^{1+\delta})$$

almost surely, where

$$\begin{aligned} G_n(\beta) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \frac{\partial}{\partial \beta} \left[\left\{ \hat{\lambda}(t) - \dot{\lambda}(t) \right\} / \lambda(t) \{X_i - \bar{X}(t, \beta)\} \right] dM_i(t, \beta) \\ &+ \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \hat{\lambda}(t) - \dot{\lambda}(t) \right\} / \lambda(t) \{X_i - \bar{X}(t, \beta)\} \frac{\partial}{\partial \beta} dM_i(t, \beta) \end{aligned}$$

is the gradient of Π_1 by formal differentiation. Thus, $G_n(\beta_0) = O_p(n^{-1/2} + 1/\sqrt{n^2 h^3} + h^2/\sqrt{n})$. Then there exists a sequence $C_n \rightarrow \infty$ such that

$$\sup_{\|\beta - \beta_0\| \leq C_n n^{-1/2}} |m_1(\beta)| = o\left(\frac{1}{\sqrt{n^2 h^3}} + \frac{h^2}{\sqrt{n}}\right). \quad (\text{A}\cdot 17)$$

Combining (A·10) and (A·17), we have

$$\sup_{\|\beta - \beta_0\| \leq C_n n^{-1/2}} |\Pi_1| = O_p\left(\frac{1}{\sqrt{n^2 h^3}} + \frac{h^2}{\sqrt{n}}\right).$$

Likewise, it can be verified that Π_i is $o_p(1/\sqrt{n^2 h^3} + h^2/\sqrt{n})$ uniformly over $\|\beta - \beta_0\| \leq C_n n^{-1/2}$ for $i = 2, 3, 4$. Therefore, (A.1) is proved. The proof of (A.2) follows along the same line as the proof of (A.1) and is omitted here. The proof of Lemma 1 is complete.

Proof of Theorem 1

Firstly, we prove (4) in Theorem 1. Observe that

$$\begin{aligned} \hat{\beta} = & \beta_0 + (\beta_1 - \beta_0) + n\{V^{-1}(\beta_1) - V_{tr}^{-1}(\beta_1)\} \frac{1}{n} \{U_{est}(\beta_1) - U_{tr}(\beta_1) + U_{tr}(\beta_1) \\ & - U_{tr}(\beta_0) - nA_n(\beta_1 - \beta_0) + nA_n(\beta_1 - \beta_0) + U_{tr}(\beta_0)\} \\ & + nV_{tr}^{-1}(\beta_1) \frac{1}{n} \{U_{est}(\beta_1) - U_{tr}(\beta_1)\} \\ & + nV_{tr}^{-1}(\beta_1) \frac{1}{n} \{U_{tr}(\beta_1) - U_{tr}(\beta_0) - nA_n(\beta_1 - \beta_0)\} \\ & + nV_{tr}^{-1}(\beta_1) A_n(\beta_1 - \beta_0) + nV_{tr}^{-1}(\beta_1) \frac{1}{n} U_{tr}(\beta_0). \end{aligned} \quad (\text{A.18})$$

Note that,

$$\begin{aligned} U_{tr}(\beta_0) = & \sum_{i=1}^n \int_0^\tau \{\dot{\lambda}(t)/\lambda(t)\} \{X_i - E(X_i|Y_i \geq t + \beta'_0 X_i)\} dM_i(t, \beta_0) \\ & + \sum_{i=1}^n \int_0^\tau \{\dot{\lambda}(t)/\lambda(t)\} \{E(X_i|Y_i \geq t + \beta'_0 X_i) - \bar{X}(t, \beta_0)\} dM_i(t, \beta_0) \\ = & \Pi_1 + \Pi_2, \quad \text{say.} \end{aligned}$$

Using the approximation established in Ying (1993), $E(X|Y \geq t + \beta'_0 X) - \bar{X}(t, \beta_0)$ is $O_p(n^{-1/2})$. It follows from the fact that $M(t, \beta_0)$ is a martingale that $\text{var}(\Pi_2/\sqrt{n}) \rightarrow 0$ as $n \rightarrow \infty$. Therefore Π_2 is $o_p(\sqrt{n})$. Under conditions (i)-(iii), Π_1 is the sum of i.i.d bounded random variables with mean 0 and variance

$$\Sigma = \text{var} \left[\int_0^\tau \{\dot{\lambda}(t)/\lambda(t)\} \{X - E(X|Y \geq t + \beta'_0 X)\} dM(t, \beta_0) \right].$$

By the property of the counting process martingale,

$$\Sigma = \int_0^\tau \left\{ \dot{\lambda}(t)/\lambda(t) \right\}^2 E \{X - E(X|Y \geq t + \beta'_0 X)\}^{\otimes 2} d\Lambda(t).$$

Therefore by the central limit theorem, $U_{tr}(\beta_0)/\sqrt{n} \rightarrow N(0, \Sigma)$ in distribution as $n \rightarrow \infty$.

Observe that $U_{tr}(\beta)$ is a weighted logrank estimating function with weight $\dot{\lambda}(\cdot)/\lambda(\cdot)$. It can be shown that, for every sequence B_n with $B_n \rightarrow 0$ and any $0 < \delta < 1$,

$$\sup_{\|\beta - \beta_0\| \leq B_n} \{|U_{tr}(\beta) - U_{tr}(\beta_0) - A_n n(\beta - \beta_0)|/(\sqrt{n} + n\|\beta - \beta_0\|^{1+\delta})\} = o_p(1), \quad (\text{A.19})$$

where

$$A_n = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \frac{\dot{\lambda}(t)}{\lambda(t)} \right\}^2 \{X_i - E(X_i|Y_i \geq t + \beta'_0 X_i)\}^{\otimes 2} dM_i(t, \beta_0) = \frac{1}{n} \left\{ \frac{\partial}{\partial \beta} U_{tr}(\beta_0) + o_p(\sqrt{n}) \right\}$$

which is similar to Theorem 2 of Ying (1993, p. 91). The proof given there essentially uses empirical approximation. Together with condition (v),

$$|U_{tr}(\beta_1) - U_{tr}(\beta_0) - A_n n(\beta_1 - \beta_0)| = o_p(\sqrt{n}). \quad (\text{A.20})$$

By the law of large numbers, $A_n \rightarrow \Sigma$ in probability as $n \rightarrow \infty$ under condition (ii). By (A.2) and condition (v), $nV^{-1}(\beta_1) - nV_{tr}^{-1}(\beta_1) = o_p(1)$ and $A_n \sqrt{n}(\beta_1 - \beta_0) = O_p(1)$. It can be shown similarly to the second part of Theorem 3.2 of Anderson and Gill (1982, p. 1108) that

$$-V_{tr}(\beta_1)/n \rightarrow \Sigma \quad (\text{A.21})$$

in probability as $n \rightarrow \infty$. Therefore,

$$\begin{aligned} (\beta_1 - \beta_0) + nV_{tr}^{-1}(\beta_1)A_n(\beta_1 - \beta_0) &= o_p(1/\sqrt{n}), \\ V_{tr}^{-1}(\beta_1)\{U_{tr}(\beta_1) - U_{tr}(\beta_0) - nA_n(\beta_1 - \beta_0)\} &= o_p(1/\sqrt{n}). \end{aligned}$$

By Lemma 1, the leading term of $\{U_{est}(\beta_1) - U_{tr}(\beta_1)\}/n$ is $O_p(1/\sqrt{n^2h^3} + h^2/\sqrt{n})$. Hence,

$$\hat{\beta} = \beta_0 + V_{tr}^{-1}(\beta_1)U_{tr}(\beta_0) + r_n, \quad (\text{A}\cdot 22)$$

where the remainder $r_n = O_p(1/\sqrt{n^2h^3} + h^2/\sqrt{n})$. In view of (A.21), (4) is proved.

Thus, under condition (iv), the consistency of $\hat{\beta}$ is proved. Using (A.21) and the fact that $U_{tr}(\beta_0)/\sqrt{n} \rightarrow N(0, \Sigma)$ in distribution as $n \rightarrow \infty$, we have

$$\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow N(0, \Sigma^{-1})$$

in distribution as $n \rightarrow \infty$. The proof is complete.

REFERENCES

- AALEN, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701-726.
- ANDERSON, P. K. & GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100-1120.
- BUCKLEY, J. & JAMES, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429-436.
- CHEN, K. (2004). Statistical estimation in the proportional hazards model with risk set sampling. *Ann. Statist.* **32**, 1513-1532.
- CHEN, K., JIN, Z. & YING, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89**, 659-668.
- FLEMING, T. R. & HARRINGTON, D. P. (1991). Counting processes and survival analysis. New York: Wiley.
- GEHAN, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* **52**, 203-223.
- JIANG, J. & DOKSUM, K. A. (2003). On local polynomial estimation of hazard functions and their derivatives under random censoring. In *Mathematical Statistics and Applications: Festschrift for Constance van Eeden, IMS Lecture Notes and Monograph Series* **42**, 463-482.
- JIN, Z., LIN, D. Y., WEI, L. J. & YING, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341-353.
- JIN, Z., LIN, D. Y. & YING, Z. (2006). On least-squares regression with censored data. *Biometrika* **93**, 147-161.
- KALBFLEISCH, J. D. & PRENTICE, R. L. (1980). Statistical analysis of failure time data. New York: Wiley.
- LAI, T. L. & YING, Z. (1991). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19**, 531-556.
- LANGHOLZ, B. & GOLDSTEIN, L. (1996). Risk set sampling in epidemiologic cohort studies. *Statist. Sci.* **11**, 35-53.
- MILLER, R. G. & HALPERN, J. (1982). Regression with censored data. *Biometrika* **69**, 521-531.
- NELSON, W. (1972). Theory and application of hazard plotting for censored failure data. *Technometrics* **14**, 945-965.
- PETO, R. & PETO J. (1972). Asymptotically efficient rank invariant test procedures. *J. Roy. Statist. Soc. Ser. A* **135**, 185-207.
- RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel function. *Ann. Statist.* **11**, 453-466.
- RITOV, Y. (1990). Estimation in a linear model with censored data. *Ann. Statist.* **18**, 303-328.

- TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354-372.
- VAN DER VAART, A. W. & WELLNER, J. A. (1996). Weak convergence of empirical processes. New York: Springer.
- YING, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21**, 76-99.
- ZENG, D. & LIN, D. Y. (2007). Efficient estimation for the accelerated failure time model. *J. Amer. Statist. Assoc.* **102**, 1387-1396.
- ZHOU, M. (1992). M-estimation in censored linear models. *Biometrika* **79**, 837-841.

Table 1: Simulation results for the proposed method with logrank initial estimator.

	$h = n^{-1/5}$				$h = n^{-1/7}$				$h = n^{-1/9}$				ideal estimator	
	BIAS	SE	SEE	CP	BIAS	SE	SEE	CP	BIAS	SE	SEE	CP	BIAS	SE
$\epsilon \sim \text{standard normal}$														
β_1	0.003	0.162	0.161	0.946	0.006	0.162	0.167	0.952	0.006	0.162	0.171	0.965	0.008	0.156
β_2	0.001	0.162	0.163	0.950	0.001	0.166	0.169	0.954	0.001	0.165	0.172	0.964	0.001	0.157
$\epsilon \sim \text{Extreme-value}$														
β_1	0.003	0.169	0.157	0.931	0.004	0.168	0.165	0.946	0.003	0.166	0.168	0.962	0.001	0.169
β_2	0.001	0.165	0.158	0.933	0.008	0.169	0.165	0.944	0.002	0.165	0.169	0.955	0.004	0.164
$\epsilon \sim \text{the standard logistic distribution}$														
β_1	0.005	0.284	0.252	0.925	0.000	0.279	0.260	0.936	0.005	0.275	0.273	0.945	0.008	0.278
β_2	0.001	0.280	0.256	0.921	0.003	0.280	0.268	0.932	0.003	0.280	0.276	0.949	0.002	0.280
$e^\epsilon \sim \text{Weibull}(1,1)$														
β_1	0.009	0.181	0.174	0.944	0.001	0.182	0.179	0.945	0.002	0.181	0.184	0.952	0.006	0.180
β_2	0.009	0.183	0.179	0.946	0.006	0.185	0.184	0.953	0.001	0.185	0.185	0.949	0.006	0.181
$\epsilon \sim 0.5N(0,1)+0.5N(0,9)$														
β_1	0.003	0.279	0.240	0.922	0.008	0.271	0.264	0.939	0.006	0.279	0.268	0.941	0.007	0.274
β_2	0.003	0.284	0.244	0.920	0.006	0.275	0.267	0.942	0.006	0.275	0.268	0.948	0.004	0.275
$\epsilon \sim 0.95N(0,1)+0.05N(0,9)$														
β_1	0.002	0.167	0.167	0.949	0.005	0.168	0.171	0.952	0.009	0.162	0.174	0.959	0.000	0.164
β_2	0.006	0.169	0.169	0.951	0.001	0.166	0.170	0.954	0.001	0.166	0.175	0.961	0.004	0.169

NOTE: SE is the standard error of the parameter estimates; SEE is the estimated standard error; CP is the coverage probability; ideal estimator is the proposed efficient estimator obtained by plugging in the true λ and $\hat{\lambda}$.

Table 2: Simulation results for the proposed method with Gehan’s initial estimator.

	$h = n^{-1/5}$				$h = n^{-1/7}$				$h = n^{-1/9}$				ideal estimator	
	BIAS	SE	SEE	CP	BIAS	SE	SEE	CP	BIAS	SE	SEE	CP	BIAS	SE
	$\epsilon \sim \text{standard normal}$													
β_1	0.001	0.162	0.161	0.951	0.006	0.161	0.166	0.956	0.001	0.160	0.170	0.960	0.001	0.161
β_2	0.003	0.165	0.164	0.950	0.001	0.162	0.168	0.953	0.001	0.159	0.171	0.964	0.003	0.159
	$\epsilon \sim \text{Extreme-value}$													
β_1	0.003	0.178	0.172	0.940	0.003	0.173	0.178	0.947	0.001	0.173	0.180	0.954	0.002	0.167
β_2	0.006	0.177	0.176	0.941	0.004	0.178	0.175	0.948	0.006	0.175	0.181	0.958	0.004	0.169
	$\epsilon \sim \text{the standard logistic distribution}$													
β_1	0.006	0.284	0.253	0.926	0.002	0.274	0.269	0.938	0.000	0.275	0.273	0.948	0.002	0.259
β_2	0.008	0.282	0.255	0.924	0.009	0.277	0.268	0.936	0.001	0.279	0.276	0.946	0.002	0.259
	$e^\epsilon \sim \text{Weibull}(1,1)$													
β_1	0.002	0.181	0.172	0.936	0.007	0.178	0.178	0.952	0.008	0.178	0.183	0.959	0.009	0.178
β_2	0.008	0.183	0.176	0.935	0.003	0.180	0.183	0.953	0.004	0.182	0.187	0.957	0.006	0.179
	$\epsilon \sim 0.5N(0,1)+0.5N(0,9)$													
β_1	0.003	0.278	0.248	0.925	0.005	0.270	0.253	0.938	0.005	0.272	0.270	0.950	0.001	0.279
β_2	0.001	0.278	0.251	0.926	0.003	0.275	0.256	0.936	0.005	0.279	0.273	0.946	0.002	0.279
	$\epsilon \sim 0.95N(0,1)+0.05N(0,9)$													
β_1	0.004	0.168	0.167	0.948	0.006	0.167	0.171	0.951	0.002	0.169	0.177	0.959	0.002	0.168
β_2	0.009	0.167	0.168	0.949	0.008	0.167	0.171	0.952	0.002	0.169	0.177	0.960	0.004	0.169

Table 3: Simulation results with the proposed data-driven bandwidth selection method.

	Proposed								Z&L		Logrank		Gehan	
	Logrank initial estimator				Gehan's initial estimator				BIAS	SE	BIAS	SE	BIAS	SE
	BIAS	SE	SEE	CP	BIAS	SE	SEE	CP						
	$\epsilon \sim \text{standard normal}$													
β_1	0.001	0.164	0.167	0.955	0.004	0.166	0.164	0.943	0.009	0.164	0.003	0.167	0.002	0.167
β_2	0.004	0.164	0.168	0.957	0.003	0.165	0.167	0.958	0.007	0.164	0.005	0.168	0.007	0.165
	$\epsilon \sim \text{Extreme-value}$													
β_1	0.004	0.175	0.181	0.956	0.005	0.174	0.178	0.957	0.009	0.196	0.006	0.172	0.006	0.195
β_2	0.001	0.176	0.183	0.957	0.004	0.176	0.180	0.956	0.004	0.199	0.009	0.171	0.003	0.201
	$\epsilon \sim \text{the standard logistic distribution}$													
β_1	0.008	0.264	0.260	0.942	0.007	0.267	0.269	0.950	0.005	0.274	0.010	0.266	0.007	0.260
β_2	0.002	0.265	0.263	0.946	0.002	0.270	0.271	0.951	0.003	0.278	0.008	0.279	0.005	0.267
	$e^\epsilon \sim \text{Weibull}(1,1)$													
β_1	0.005	0.174	0.180	0.960	0.006	0.178	0.187	0.956	0.005	0.187	0.008	0.180	0.008	0.191
β_2	0.004	0.180	0.185	0.951	0.003	0.180	0.188	0.957	0.008	0.186	0.002	0.181	0.007	0.194
	$\epsilon \sim 0.5N(0,1)+0.5N(0,9)$													
β_1	0.007	0.274	0.265	0.944	0.003	0.272	0.279	0.958	0.009	0.274	0.001	0.298	0.001	0.263
β_2	0.002	0.278	0.268	0.943	0.002	0.270	0.272	0.951	0.008	0.277	0.010	0.306	0.016	0.275
	$\epsilon \sim 0.95N(0,1)+0.05N(0,9)$													
β_1	0.007	0.165	0.172	0.953	0.006	0.165	0.167	0.951	0.010	0.171	0.006	0.172	0.003	0.164
β_2	0.005	0.167	0.174	0.954	0.000	0.169	0.171	0.954	0.004	0.172	0.009	0.182	0.007	0.167

NOTE: Z&L is the profile-likelihood method in Zeng and Lin (2007).