

# Estimating Mean Survival Time: When is it Possible?

YING DING

*Department of Biostatistics, University of Pittsburgh*

BIN NAN

*Department of Biostatistics, University of Michigan*

**ABSTRACT.** For right-censored survival data, it is well-known that the mean survival time can be consistently estimated when the support of the censoring time contains the support of the survival time. In practice, however, this condition can be easily violated because the follow-up of a study is usually within a finite window. In this article, we show that the mean survival time is still estimable from a linear model when the support of some covariate(s) with non-zero coefficient(s) is unbounded regardless of the length of follow-up. This implies that the mean survival time can be well estimated when the support of linear predictor is wide in practice. The theoretical finding is further verified for finite samples by simulation studies. Simulations also show that, when both models are correctly specified, the linear model yields reasonable mean square prediction errors and outperforms the Cox model, particularly with heavy censoring and short follow-up time.

*Key words:* censored linear regression, empirical process theory, Gehan weights, mean survival time, unbounded covariate support

## 1. Introduction

Estimating mean survival time becomes increasingly important, especially in oncology studies. For example, Zhao *et al.* (2011) proposed to use patient's mean survival time as the clinical outcome for the evaluation of optimal personalized treatment in lung cancer clinical trials. In some circumstances, the restricted mean survival time has been used to either compare the treatment effects or predict the individual patient's survival outcomes. However, it depends on the choice for the upper limit time  $T^*$ , and it is often difficult to explain the restriction part to clinicians or patients, whereas the unrestricted mean survival time provides an intuitive and straightforward interpretation. Another motivation for the mean survival time estimation comes from health economics studies, where the analysis of cost-effectiveness usually requires the estimation of mean lifetimes under different treatments. See, for example, Etzioni *et al.* (1999) and Paltiel *et al.* (2009).

The linear regression model for censored survival data, as an important alternative to the Cox model (Cox, 1972), has been extensively studied in recent years, see, for example, Ritov (1990), Tsiatis (1990), Wei *et al.* (1990), Ying (1993), Jin *et al.* (2003), & Ding & Nan (2011), among many others. This type of model appeals in many ways (Cox & Oakes, 1984): it models the failure time directly and thus has a more intuitive interpretation, it may provide more accurate summary of the data when Cox's proportional hazards assumption is violated, and more importantly, it can be used to predict the failure time in a straightforward way.

Obviously, a good mean survival time estimation/prediction requires a good estimator for the intercept parameter in a linear model. The study of such a linear model has primarily focused on the slope parameter estimation. Commonly used estimating methods for slopes include the Buckley–James method (Buckley & James, 1979) that imputes the censored failure time using mean residual given covariates and the rank-based method (Prentice, 1978; Tsiatis, 1990; Ying, 1993, among many others), that is derived by using linear rank tests for the right

censored data. Ritov (1990) showed that the class of weighted rank-based estimating functions of Tsiatis (1990) is asymptotically equivalent to the class of Buckley–James estimating functions on transformed residuals.

The estimation of intercept parameter when the error distribution is unspecified, however, has not been thoroughly studied mostly because of the finite follow-up time in practice so that the intercept is usually believed to be underestimated. Buckley & James (1979) claimed that in general, the intercept can not be consistently estimated. In some of their simulations, however, Heller & Simonoff (1990) and Schneider & Weissfeld (1986) found that the intercept can sometimes be estimated quite well. Without the presence of covariates, using an integration by parts argument with a truncation technique, Susarla & Van Ryzin (1980) showed that when the support of censoring time distribution contains the support of failure time distribution together with appropriate assumptions for the tail probability, the mean failure time estimation based on a Kaplan–Meier-type estimator is consistent almost surely under random censoring. Using the reverse martingale approach, Stute & Wang (1993) established more general strong consistency results including the mean failure time estimation. On the basis of the work of Susarla *et al.* (1984) and Susarla & Van Ryzin (1980), Wang *et al.* (2008) conjectured that the intercept can be consistently estimated when the supports of some covariates are not restricted to finite intervals. In this article, we confirm such a conjecture by formally establishing the consistency result for the intercept estimator. This result makes the estimation of mean survival time possible under a linear regression model when some covariate multiplied by its non-zero regression coefficient, or in general, the linear predictor ( $\beta_0 X$ ), has a wide support in a practical setting. This wide support condition excludes the case when all the regression covariates are categorical with finite number of categories.

In a linear model, the intercept estimation is equivalent to the mean failure time estimation on the residual scale if true values of the slope parameters are given. In reality, however, the slope parameters need to be estimated, which dramatically complicates the study of asymptotic properties of the intercept estimation. For the consistency of intercept estimation when slopes are estimated, we are only aware of Lai & Ying (1991) who assumed bounded covariates, bounded support of the failure time distribution and wider support of the censoring time distribution. The latter assumption, however, is often violated in practice because of the nature of limited follow-up time in, for example, most of the human disease studies. Instead of assuming wider support of the censoring time distribution, we consider the setting that the supports of some covariates with non-zero coefficients are not restricted to finite intervals, which requires additional consideration on the slope estimation because its theoretical developments to date have been primarily under the assumption of bounded covariates. The unbounded covariate support is a technical condition and approximates the practical situation where the ranges of the explanatory covariates are wide.

This article is organized as follows. In Section 2, we present the strong consistency property of an intercept estimator under the assumption of unbounded covariates. In Section 3, we present both in probability and almost sure consistency as well as asymptotic normality results for the Gehan-weighted rank-based slope estimators without assuming bounded covariates. In Section 4, we conduct simulation studies by varying the covariate support and the truncation time under different error distributions with different sample sizes. We also compare the failure time prediction performance with the Cox model under the standard extreme value error distribution for which both models fit the data correctly. In Section 5, we provide an application to the Mayo primary biliary cirrhosis (PBC) study for illustration. We provide some concluding remarks in Section 6. Proofs of the technical results rely on modern empirical process theory and are included in the Supporting information.

**2. Intercept estimation**

Consider the linear regression model:

$$T_i = \alpha_0 + X_i' \beta_0 + \zeta_i, \quad i = 1, \dots, n, \tag{2.1}$$

where  $\zeta_i, i = 1, \dots, n$  are independent and identically distributed (i.i.d.) with zero mean. The response variable  $T_i$  for the  $i$ th subject is the failure time (possibly transformed by a known monotone function). When  $T_i$  is subject to right censoring, we only observe  $(Y_i, \Delta_i, X_i)$ , where  $Y_i = \min(T_i, C_i)$ ,  $C_i$  is the censoring time (possibly transformed by the same function that yields  $T_i$ ) and  $\Delta_i = 1(T_i \leq C_i)$ . Here, we assume that  $(X_i, C_i), i = 1, \dots, n$  are i.i.d. and independent of  $\zeta_i$ .

Throughout the sequel, we consider one-dimensional  $\beta_0$  just for notational simplicity. All the results in this article hold for multiple-dimensional  $\beta_0$ . Denote the parameter space for  $\beta$  by  $\mathcal{B}$ , an arbitrary subset of the real line. For any  $\beta \in \mathcal{B}$ , we denote

$$e_{\beta,i} = T_i - \beta X_i, \quad e_{0,i} = T_i - \beta_0 X_i = \alpha_0 + \zeta_i,$$

and

$$\epsilon_{\beta,i} = Y_i - \beta X_i, \quad \epsilon_{0,i} = Y_i - \beta_0 X_i.$$

Here,  $e_{\beta,i}$  are the failure times in the residual scale with  $\beta_0$  replaced by  $\beta$ ,  $\epsilon_{\beta,i}$  are the corresponding observed times in the residual scale for a fixed  $\beta$ , and  $e_{0,i}$  are the error terms that have absorbed the intercept in model (2.1). We use  $F$  and  $G$  to denote the distribution functions of  $e_{0,i}$  and  $C_i$ , and  $f$  and  $g$  to denote their density functions, respectively. Now we adopt the empirical process notation of van der Vaart & Wellner (1996). In particular, for a function  $f$  of a random variable  $U$  that follows distribution  $P$ , we denote

$$Pf = \int f(u) dP(u), \quad \mathbb{P}_n f = n^{-1} \sum_{i=1}^n f(U_i), \quad \mathbb{G}_n f = n^{1/2}(\mathbb{P}_n - P)f,$$

and refer all the details to the reference. Set  $\epsilon_\beta = Y - \beta X$  and  $\epsilon_0 = Y - \beta_0 X$ . Define

$$H_n^{(0)}(\beta, s) = \mathbb{P}_n\{1(\epsilon_\beta \leq s, \Delta = 1)\}, \quad h^{(0)}(\beta, s) = P\{1(\epsilon_\beta \leq s, \Delta = 1)\}; \tag{2.2}$$

and

$$H_n^{(1)}(\beta, s) = \mathbb{P}_n\{1(\epsilon_\beta \geq s)\}, \quad h^{(1)}(\beta, s) = P\{1(\epsilon_\beta \geq s)\}. \tag{2.3}$$

Because  $\alpha_0 = Ee_{0,i} = \int_{-\infty}^{\infty} t dF(t)$ , if the slope  $\beta_0$  is known, then a natural estimator of  $\alpha_0$  is given by

$$\hat{\alpha}_n = \int_{-\infty}^{\infty} t d\hat{F}_n(t), \tag{2.4}$$

where  $\hat{F}_n(t)$  is the Kaplan–Meier estimator of the distribution function  $F(t)$  of  $e_0 = T - \beta_0 X$ . In a regression setting, however,  $\beta_0$  is unknown and hence needs to be estimated. Let  $\hat{\beta}_n$  be an estimator of  $\beta_0$ , a direct extension of (2.4) yields the estimator of interest:

$$\hat{\alpha}_{n,\hat{\beta}_n} = \int_{-\infty}^{\infty} t d\hat{F}_{n,\hat{\beta}_n}(t), \tag{2.5}$$

where  $\hat{F}_{n,\beta}(t)$  is the Kaplan–Meier estimator of the distribution function  $F_\beta(t)$  of  $e_\beta = T - \beta X$ , which is given by

$$\hat{F}_{n,\beta}(t) = 1 - \prod_{i:\epsilon_{\beta,i} \leq t} \left\{ 1 - \frac{\Delta_i/n}{H_n^{(1)}(\beta, \epsilon_{\beta,i})} \right\}. \tag{2.6}$$

Note that the aforementioned estimator does not automatically provide a consistent estimator of  $F_\beta(t)$  because  $T - \beta X$  and  $C - \beta X$  are not independent except when  $\beta = \beta_0$ . We will follow the method of Lai & Ying (1991) to show that  $\hat{F}_{n,\hat{\beta}_n}(t)$  converges to  $F(t)$  when  $\hat{\beta}_n$  converges to  $\beta_0$  with certain polynomial rate.

Susarla *et al.* (1984) showed that the aforementioned  $\hat{\alpha}_{n,\hat{\beta}_n}$  is identical to the Buckley–James estimator of  $\alpha_0$  for a fixed  $\hat{\beta}_n$ . When there is no covariates (equivalently  $\beta_0 = 0$ ) or  $\beta_0$  is given, Stute & Wang (1993) and Susarla & Van Ryzin (1980) studied the asymptotic properties of the mean survival time estimator (2.4). They provided the following key sufficient condition:

$$\{t : t \in \text{the support of } T - \beta_0 X\} \subseteq \{t : t \in \text{the support of } C - \beta_0 X\} \tag{2.7}$$

for the consistency of (2.4). Now we replace  $\beta_0$  by its estimator  $\hat{\beta}_n$  and want to show the consistency of (2.5). The proof of Stute & Wang (1993) for the consistence of the mean survival time estimation makes use of the martingale theory that cannot be directly adopted here because of the dependence between  $T - \beta X$  and  $C - \beta X$  when  $\beta \neq \beta_0$ . We shall use the empirical process theory as well as the properties of stochastic integrals with censored data in Lai & Ying (1988) together with the delicate controlling of the tail fluctuations used by Lai & Ying (1991) and Ying (1993) to show the desirable result.

First, we introduce the following regularity conditions:

*Condition 1.* The covariates  $X_i$  are i.i.d. random variables with a finite second moment.

*Condition 2.* The error  $e_0$ 's density  $f$  and its derivative  $\dot{f}$  are bounded and satisfy

$$\int_{-\infty}^{\infty} \left( \dot{f}(t)/f(t) \right)^2 f(t) dt < \infty.$$

*Condition 3.* The conditional density of  $C$  given  $X$  is uniformly bounded for all possible values of  $X$ , that is,

$$\sup_{x \in \mathcal{X}, t \in \mathcal{C}} g_{C|X}(t | X = x) < \infty,$$

where  $\mathcal{X}$  and  $\mathcal{C}$  denote the support of  $X$  and  $C$ , respectively.

*Condition 4.* The error  $e_0$  has a finite second moment, that is,  $Ee_0^2 < \infty$ .

Condition 1 is different to the common assumption of bounded covariates in Lai & Ying (1991), Tsiatis (1990), & Ying (1993), and many others. Here we only assume finite second moment. Hence, even with a short follow-up time, the support of the censoring time in the residual scale can be extended to infinity provided that the support of  $X$  is the real line and  $\beta_0 \neq 0$ , which yields that the supports of  $e_0$  and  $C - \beta_0 X$  are equivalent, and thus, the sufficient condition (2.7) is satisfied. This requirement is for theoretical justification, whereas in practice, wide support for  $X$  works reasonably well. Condition 2 is exactly the same as condition 2 in Ying (1993). Condition 3 implies condition 3 in Ying (1993) as well as condition (3.5) in Lai & Ying (1991). Condition 4 implies condition 4 in Ying (1993) where  $\theta_0 = 2$ .

In the following theorems 2.1 and 2.2, we omit the constants in front of the rate expressions to further simplify the notation.

**Theorem 2.1.** *Suppose conditions 1–3 hold and define*

$$F(\beta, t) = 1 - \exp \left\{ - \int_{u \leq t} \frac{dh^{(0)}(\beta, u)}{h^{(1)}(\beta, u)} \right\}, \tag{2.8}$$

where  $h^{(0)}(\beta, u)$  and  $h^{(1)}(\beta, u)$  are given in (2.2) and (2.3), respectively. Then for every  $\varepsilon > 0$ , with probability 1, we have

$$\sup \left\{ |\hat{F}_{n,\beta}(t) - F(\beta, t)| : \beta \in \mathcal{B}, H_n^{(1)}(\beta, t) \geq n^{-\varepsilon} \right\} = o \left( n^{-\frac{1}{2} + 3\varepsilon} \right), \tag{2.9}$$

and

$$\sup \left\{ |F(\beta, t) - F(t)| : |\beta - \beta_0| \leq n^{-3\varepsilon}, h^{(1)}(\beta, t) \geq n^{-\varepsilon} \right\} = O(n^{-\varepsilon}), \tag{2.10}$$

where  $\hat{F}_{n,\beta}(t)$  is given in (2.6). Consequently, for every  $0 < \varepsilon \leq \frac{1}{8}$ , with probability 1, we have

$$\sup \left\{ |\hat{F}_{n,\beta}(t) - F(t)| : |\beta - \beta_0| \leq n^{-3\varepsilon}, H_n^{(1)}(\beta, t) \geq n^{-\varepsilon} \right\} = O(n^{-\varepsilon}). \tag{2.11}$$

Introduced by Lai & Ying (1991),  $F(\beta, t)$  defined in (2.8) is an important intermediate quantity. On the one hand, it is the limit of the Kaplan–Meier estimator  $\hat{F}_{n,\beta}(t)$  for a fixed  $\beta$ ; on the other hand, it equals to  $F(t)$ , the true distribution function of  $e_0$ , when  $\beta$  is replaced by the true slope  $\beta_0$  in (2.8). The biggest difference between theorem 2.1 and lemma 2 of Lai & Ying (1991) is that we do not require bounded covariate support.

**Theorem 2.2.** *Suppose conditions 1–4 hold and additionally assume  $\beta_0 \neq 0$  and that the support of  $X$  is the whole real line, that is,  $f_X(x) > 0$  for all  $-\infty < x < \infty$ . Define*

$$T_n = \sup \left\{ t : H_n^{(1)}(\beta, t) \geq n^{-\varepsilon}, |\beta - \beta_0| \leq n^{-3\varepsilon} \right\}, \tag{2.12}$$

and let  $\hat{F}_{n,\beta}(t) = 1$  for  $t > T_n$ . Then for every  $0 < \varepsilon \leq \frac{1}{8}$ , with probability 1, we have

$$\sup \left\{ \left| \int_{-\infty}^{\infty} t d\hat{F}_{n,\beta}(t) - \alpha_0 \right| : |\beta - \beta_0| \leq n^{-3\varepsilon} \right\} = o(1). \tag{2.13}$$

It is clearly seen from theorem 2.2 that  $\hat{\alpha}_{n,\hat{\beta}_n}$  given in (2.5) is a consistent estimator of the intercept  $\alpha_0$  when  $\hat{\beta}_n$  is consistent with a polynomial convergence rate. This requires a good estimator of the slope parameter  $\beta_0 \neq 0$  under conditions 1–4 as well as the assumption of unbounded support for  $X$ . In the next section, we show that such an estimator can be obtained by the Gehan-weighted rank-based estimating method.

### 3. Slope estimation with unbounded covariate support

Define

$$H_n^{(2)}(\beta, s) = \mathbb{P}_n \{ 1(\varepsilon_\beta \geq s)X \} \quad \text{and} \quad h^{(2)}(\beta, s) = P \{ 1(\varepsilon_\beta \geq s)X \}. \tag{3.1}$$

Then the general rank-based estimating function of Tsiatis (1990) is given by

$$\mathbb{P}_n \left\{ \omega_n(\beta, \epsilon_\beta) \left[ X - \frac{H_n^{(2)}(\beta, \epsilon_\beta)}{H_n^{(1)}(\beta, \epsilon_\beta)} \right] \Delta \right\}, \tag{3.2}$$

where  $\omega_n(\beta, s)$  is a weight function and  $H_n^{(1)}(\beta, s) = \mathbb{P}_n\{1(\epsilon_\beta \geq s)\}$  is defined in (2.3). We consider the Gehan weight function  $\omega_n(\beta, s) = H_n^{(1)}(\beta, s)$  in (3.2), which yields the following estimating function:

$$\Psi_n \left( \beta, H_n^{(1)}, H_n^{(2)} \right) = \mathbb{P}_n \left\{ \left[ H_n^{(1)}(\beta, \epsilon_\beta) X - H_n^{(2)}(\beta, \epsilon_\beta) \right] \Delta \right\}. \tag{3.3}$$

It is well-known that the aforementioned estimating function is a discrete (Kalbfleisch & Prentice, 2002) and monotone function (Fyngenson & Ritov, 1994) of  $\beta$ , and can be solved by linear programming (Jin *et al.*, 2003; Lin *et al.*, 1998) or by a Newton-type algorithm (Yu & Nan, 2006).

3.1. Convergence in probability and asymptotic normality

The reason of assuming bounded covariates and/or truncated residual time in the current literature is to bound the denominator  $H_n^{(1)}(\beta, \epsilon_\beta)$  in (3.2) away from zero. Such an issue disappears in (3.3) when the Gehan weight function is used. Without concerning bounding the denominator away from zero, we can follow the same proofs in Nan *et al.* (2009) to obtain the following results. Details are thus omitted here but referred to Ding (2010).

**Proposition 3.1.** *Suppose conditions 1–3 hold. Assume that  $\beta_0 \in \mathcal{B}$  is the unique root of  $\Psi(\beta, h^{(1)}, h^{(2)}) = P \{ [h^{(1)}(\beta, \epsilon_\beta) X - h^{(2)}(\beta, \epsilon_\beta)] \Delta \}$ .*

(1) *The approximate root  $\hat{\beta}_n$  satisfying*

$$\Psi_n \left( \hat{\beta}_n, H_n^{(1)}(\hat{\beta}_n, \cdot), H_n^{(2)}(\hat{\beta}_n, \cdot) \right) = o_p(1)$$

*is a consistent estimator of  $\beta_0$ .*

(2) *Suppose that  $\Psi(\beta, h^{(1)}(\beta, \cdot), h^{(2)}(\beta, \cdot))$  is differentiable with bounded and continuous derivative  $\dot{\Psi}_\beta(\beta, h^{(1)}(\beta, \cdot), h^{(2)}(\beta, \cdot))$  in a neighbourhood of  $\beta_0$ , and that  $\dot{\Psi}_\beta(\beta_0, h^{(1)}(\beta_0, \cdot), h^{(2)}(\beta_0, \cdot))$  is non-singular. Then for an approximate root  $\hat{\beta}_n$  satisfying*

$$\Psi_n \left( \hat{\beta}_n, H_n^{(1)}(\hat{\beta}_n, \cdot), H_n^{(2)}(\hat{\beta}_n, \cdot) \right) = o_p \left( n^{-1/2} \right),$$

*we have that  $n^{1/2}(\hat{\beta}_n - \beta_0)$  is asymptotically normal with the following asymptotic representation:*

$$n^{1/2} \left( \hat{\beta}_n - \beta_0 \right) = \mathbb{G}_n \{ m(\beta_0, \epsilon_0; \Delta, X) \} + o_p(1),$$

*where*

$$\begin{aligned} & m(\beta_0, \epsilon_0; \Delta, X) \\ &= \left\{ -\dot{\Psi}_\beta \left( \beta_0, h^{(1)}(\beta_0, \cdot), h^{(2)}(\beta_0, \cdot) \right) \right\}^{-1} \left\{ \left[ h^{(1)}(\beta_0, \cdot) X - h^{(2)}(\beta_0, \cdot) \right] \Delta \right. \\ & \quad \left. - \int [1(\epsilon_0 \geq t) X] dP_{\epsilon_0, \Delta}(t, 1) + \int [1(\epsilon_0 \geq t) X] dP_{\epsilon_0, \Delta, X}(t, 1, X) \right\}. \end{aligned} \tag{3.4}$$

Proposition 3.1 implies that  $|\hat{\beta}_n - \beta_0| = O_p(n^{-3\varepsilon})$  for any  $0 < \varepsilon \leq \frac{1}{8}$  with probability approaching 1. Hence, we have that  $\hat{\alpha}_{n, \hat{\beta}_n}$  converges to  $\alpha_0$  in probability by theorem 2.2.

3.2. *Almost sure convergence with polynomial rate*

Following theorem 5 in Ying (1993), the almost sure consistency of the Gehan-weighted rank-based slope estimator  $\hat{\beta}_n$  with a polynomial rate can also be achieved under the unbounded covariate support assumption, which leads to the strong convergence of the intercept estimator  $\hat{\alpha}_{n, \hat{\beta}_n}$  from theorem 2.2.

**Proposition 3.2.** *Suppose all the assumptions in theorem 2.2 hold, and additionally, we assume that the tail probability of  $X$  satisfies*

$$P(|X| > t) \leq Mt^\theta \exp(-\eta t^\gamma), \tag{3.5}$$

for some constants  $M > 0$ ,  $|\theta| < \infty$ ,  $\eta > 0$ , and  $\gamma > 0$ . Then with probability 1, the estimator  $\hat{\beta}_n$  satisfying  $\Psi_n(\hat{\beta}_n, H_n^{(1)}(\hat{\beta}_n, \cdot), H_n^{(2)}(\hat{\beta}_n, \cdot)) = o(n^{-1/2})$  almost surely converges to  $\beta_0$  with a polynomial rate, that is,  $|\hat{\beta}_n - \beta_0| = o(n^{-1/2+\varepsilon})$  almost surely for every  $\varepsilon > 0$ .

The exponential tail probability bound (3.5) implies condition 1 in Ying (1993), which is  $\max_{i \leq n} |X_i| = o(n^\varepsilon)$  almost surely for every  $\varepsilon > 0$ . This is because for every  $t > 0$ , we have

$$\begin{aligned} P\left(\max_{i \leq n} |X_i| > t\right) &= 1 - P\left(\max_{i \leq n} |X_i| \leq t\right) \\ &= 1 - [1 - P(|X| > t)]^n \\ &\leq 1 - [1 - Mt^\theta \exp(-\eta t^\gamma)]^n \\ &\leq nMt^\theta \exp(-\eta t^\gamma), \end{aligned}$$

where the last inequality holds because  $(1 - s)^n \geq 1 - ns$  for  $0 \leq s \leq 1$ . Therefore, for every fixed  $t > 0$  and  $\varepsilon > 0$ ,

$$\sum_{n=1}^{\infty} P\left(\max_{i \leq n} |X_i| > n^\varepsilon t\right) \leq \sum_{n=1}^{\infty} nM(n^\varepsilon t)^\theta \exp\{-\eta(n^\varepsilon t)^\gamma\} < \infty.$$

Then by the Borel–Cantelli lemma,  $P(\lim_{n \rightarrow \infty} n^{-\varepsilon} \max_{i \leq n} |X_i| = 0) = 1$ , that is,  $\max_{i \leq n} |X_i| = o(n^\varepsilon)$  almost surely. As we mentioned earlier, our conditions 2–4 imply conditions 2–4 in Ying (1993). Furthermore, Ying (1993) pointed out that Gehan weights satisfy his condition 5 and his equation (4.7). Hence, the conclusion in proposition 3.2 follows directly from his equation (4.8) in his theorem 5. The detailed argument is thus omitted. The exponential tail probability condition holds for many commonly used distributions, for example, normal, Weibull, and extreme value distributions.

4. Simulations

4.1. *Intercept estimation*

We conduct extensive simulations to investigate the finite sample performance of the intercept estimation under different scenarios. Failure times are generated from the following model:

$$T = 2 + X_1 + X_2 + \zeta.$$

This is a submodel of Jin *et al.* (2006), Zeng & Lin (2007), & Ding & Nan (2011) in their simulations. Five different error distributions are considered, which are (i)  $\zeta \sim N(0, 0.5^2)$ ; (ii)  $\zeta \sim \text{Gumbel}(-0.5\gamma, 0.5)$  that has mean zero, where  $\gamma$  is the Euler constant; (iii)  $\zeta \sim \text{Laplace}(0, 0.5)$ ; (iv)  $\zeta \sim \text{Logistic}(0, 0.5)$ ; and (v)  $\zeta \sim t(0, df = 30)$ . In each scenario,  $X_1$  is Bernoulli with  $p = 0.5$ , and  $X_2$  is continuous. We investigate three different distributions of  $X_2$ : (i)  $X_2 \sim N(0, 1)$ ; (ii)  $X_2 \sim U(-2, 2)$ ; and (iii)  $X_2 \sim U(-0.5, 0)$ . The censoring distribution is  $C \sim U(0, 5) \wedge \tau$ , here  $\tau$  is a truncation time that reflects the length of follow-up time. We choose  $\tau = 1.5$  and  $\tau = 4$  to yield censoring rate ranges (76%, 88%) and (45%, 52%), respectively. We simulate 1000 runs for each setting and report the simulation results in Table 1 for two different sample sizes: 100 and 400. Under each scenario, we also compute the asymptotic bias of the intercept estimate in the following way. The intercept is the mean residual time, which equals to the area under the residual survival function  $S(e_0)$ . When the support of the residual censoring time ( $C - \beta_0 X$ ) is shorter than the support of the residual survival time ( $e_0 = T - \beta_0 X$ ), that is,

$$a \equiv \sup(C - \beta_0 X) < b \equiv \sup(T - \beta_0 X),$$

the survival function of the residual time is estimable up to  $a$ , with  $S(a) > 0$ . Then to calculate the intercept, the survival function will be forced to drop down to zero beyond  $a$ . Hence, the asymptotic bias of the intercept is the area under  $S(e_0)$  between  $a$  and  $b$ , that is,  $\int_a^b S(e_0) de_0$ .

The first covariate setting corresponds to the unbounded covariate support. The asymptotic bias of the intercept estimator, denoted by  $\delta_\alpha$ , is zero in this setting, because  $a = b = \infty$  and  $S(e_0)$  is an integrable function (thus,  $\int_a^b S(e_0) de_0 = 0$ ). It is clearly seen that the bias of the intercept estimator is minimal even with the shorter truncation time  $\tau = 1.5$  for all error distributions. The bias is also very small in the second covariate setting, where the support of  $X_2$  is bounded, but wide. The asymptotic bias  $\delta_\alpha$  is also very close to zero for each case under this wide support setting. The bias becomes noticeable when the support of  $X_2$  gets narrower in the third setting with truncation time  $\tau = 1.5$ , which is consistent with the asymptotic bias  $\delta_\alpha$ . With the longer truncation time  $\tau = 4$ , which is close to the setting of Lai & Ying (1991) who assumed wider censoring time support, the bias of the intercept estimator is negligible for all error distributions and covariate supports. The bias for the slope estimators is minimal for most of the simulation settings except for the binary covariate  $X_1$  under the third setting ( $X_2 \sim U(-0.5, 0)$ ) with Normal and Gumbel errors when the follow-up time is short. This is possibly because when the censoring rate is very high ( $\geq 85\%$ ), the probability to observe a non-zero value of  $X_1$  under the uncensored case is very small (about 7.5% or less); therefore, the estimation for  $\beta_1$  did not perform well in this case.

For the shorter follow-up setting with  $\tau = 1.5$ , Fig. 1 displays the Kaplan–Meier curves of the estimated residual survival time  $T_i - \hat{\beta}_{n,1} X_{i,1} - \hat{\beta}_{n,2} X_{i,2}$  under five error distributions, where each curve is obtained from a sample with size  $n = 400$ . From left to right, the three panels correspond to  $X_2 \sim N(0, 1)$ ,  $X_2 \sim U(-2, 2)$ , and  $X_2 \sim U(-0.5, 0)$ , respectively. It is clearly seen that whenever the survival curve is close to zero at the right tail, the corresponding intercept estimator in Table 1 has minimal bias. We notice from our intensive simulations that a satisfactory intercept estimator (bias  $< 5\%$ ) can be obtained when the right tail of the Kaplan–Meier curve goes below 0.15. This provides a practical rule of thumb for having a sense of adequacy of the intercept estimation.

In Section 3, we have shown the asymptotic normality of the Gehan-weighted rank-based method (for the slope estimator) with unbounded covariate support. Current theory for other estimating methods, for example, the standard Buckley–James estimator (Buckley & James, 1979) or the modified Buckley–James estimator (Ying, 1993), requires bounded covariate



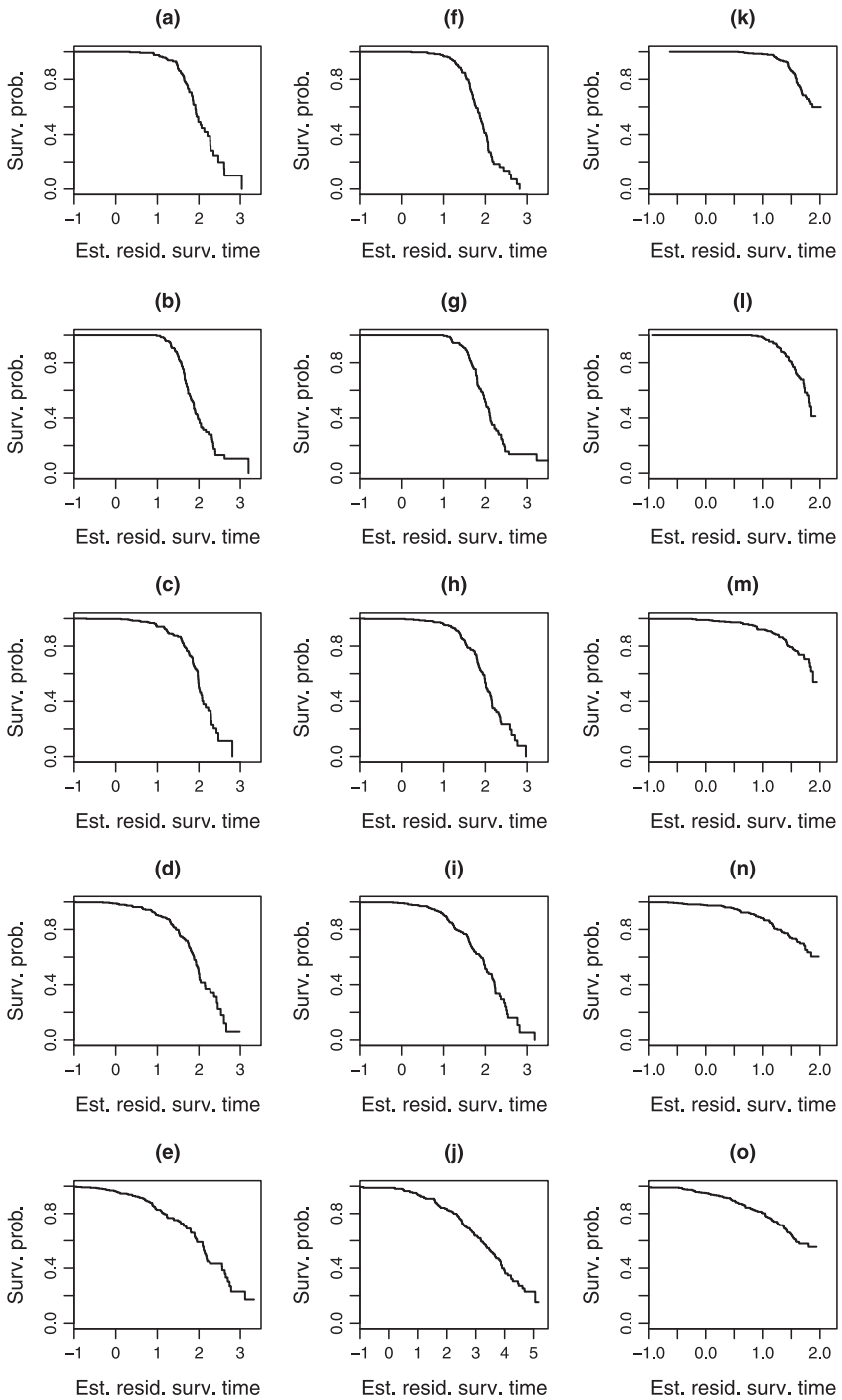


Fig. 1. K–M plots of the estimated residual survival time  $(T - \hat{\beta}_{n,1}X_1 - \hat{\beta}_{n,2}X_2)$  under  $\tau = 1.5$ . Each column corresponds to one of the five error distributions given in Table 1. (a)–(e):  $X_2 \sim N(0, 1)$ ; (f)–(j):  $X_2 \sim U(-2, 2)$ ; (k)–(o):  $X_2 \sim U(-0.5, 0)$ .

Table 1. Summary of the simulation statistics

Err. dist	Cen. rate	n = 100			n = 400			$\delta_\alpha$
		$\alpha$	$\beta_1$	$\beta_2$	$\alpha$	$\beta_1$	$\beta_2$	
$X_2 \sim N(0, 1)$								
(a)	0.83 <sup>†</sup>	2.00 (0.22)	1.02 (0.27)	1.01 (0.18)	2.00 (0.10)	1.00 (0.12)	1.00 (0.08)	0
	0.51 <sup>‡</sup>	2.00 (0.09)	1.00 (0.14)	1.00 (0.08)	2.00 (0.04)	1.00 (0.07)	1.00 (0.04)	0
(b)	0.82 <sup>†</sup>	1.96 (0.19)	1.02 (0.24)	1.01 (0.15)	1.98 (0.10)	1.00 (0.10)	1.00 (0.07)	0
	0.51 <sup>‡</sup>	2.00 (0.11)	1.01 (0.14)	1.00 (0.08)	2.00 (0.05)	1.00 (0.07)	1.00 (0.04)	0
(c)	0.82 <sup>†</sup>	1.99 (0.28)	1.04 (0.39)	1.02 (0.25)	1.99 (0.13)	1.00 (0.18)	1.00 (0.11)	0
	0.51 <sup>‡</sup>	2.00 (0.12)	1.00 (0.17)	1.00 (0.09)	2.00 (0.06)	1.00 (0.08)	1.00 (0.04)	0
(d)	0.80 <sup>†</sup>	1.97 (0.30)	1.04 (0.40)	1.02 (0.26)	1.97 (0.14)	1.01 (0.18)	0.99 (0.11)	0
	0.51 <sup>‡</sup>	1.99 (0.16)	1.00 (0.23)	1.00 (0.12)	2.00 (0.07)	1.00 (0.11)	1.00 (0.06)	0
(e)	0.78 <sup>†</sup>	1.95 (0.30)	1.03 (0.39)	1.03 (0.23)	1.96 (0.15)	1.02 (0.19)	1.00 (0.11)	0
	0.51 <sup>‡</sup>	1.99 (0.11)	1.00 (0.26)	1.01 (0.14)	1.99 (0.05)	1.00 (0.13)	1.00 (0.07)	0
$X_2 \sim U(-2, 2)$								
(a)	0.79 <sup>†</sup>	2.03 (0.24)	1.01 (0.24)	1.02 (0.18)	1.99 (0.10)	1.00 (0.11)	1.00 (0.08)	0*
	0.52 <sup>‡</sup>	2.01 (0.09)	1.00 (0.14)	1.00 (0.07)	2.00 (0.04)	1.00 (0.07)	1.00 (0.03)	0*
(b)	0.78 <sup>†</sup>	1.98 (0.21)	1.01 (0.21)	1.02 (0.16)	1.98 (0.10)	0.99 (0.10)	0.99 (0.07)	-0.01
	0.51 <sup>‡</sup>	2.00 (0.11)	1.00 (0.14)	1.00 (0.07)	2.00 (0.05)	1.00 (0.07)	1.00 (0.03)	0*
(c)	0.78 <sup>†</sup>	2.02 (0.30)	1.02 (0.32)	1.04 (0.25)	1.99 (0.14)	1.00 (0.15)	1.00 (0.11)	-0.01
	0.51 <sup>‡</sup>	2.00 (0.12)	1.00 (0.17)	1.00 (0.09)	2.00 (0.06)	1.00 (0.08)	1.00 (0.04)	0*
(d)	0.77 <sup>†</sup>	1.99 (0.30)	1.02 (0.36)	1.03 (0.25)	1.98 (0.15)	1.00 (0.17)	1.00 (0.11)	-0.02
	0.51 <sup>‡</sup>	2.00 (0.16)	1.00 (0.23)	1.00 (0.11)	2.00 (0.08)	1.00 (0.11)	1.00 (0.05)	0*
(e)	0.76 <sup>†</sup>	1.96 (0.29)	1.02 (0.36)	1.03 (0.23)	1.95 (0.14)	1.01 (0.18)	1.00 (0.11)	-0.04
	0.52 <sup>‡</sup>	1.99 (0.18)	1.00 (0.27)	1.00 (0.13)	2.00 (0.09)	1.00 (0.13)	1.00 (0.06)	0*
$X_2 \sim U(-0.5, 0)$								
(a)	0.88 <sup>†</sup>	1.81 (0.28)	0.74 (0.28)	1.06 (0.74)	1.80 (0.28)	1.11 (0.29)	1.00 (0.34)	-0.20
	0.45 <sup>‡</sup>	2.00 (0.14)	1.00 (0.13)	0.99 (0.44)	2.00 (0.07)	1.00 (0.06)	1.00 (0.22)	0*
(b)	0.85 <sup>†</sup>	1.77 (0.21)	0.78 (0.36)	1.07 (0.58)	1.75 (0.11)	1.20 (0.42)	1.01 (0.29)	-0.25
	0.45 <sup>‡</sup>	1.99 (0.16)	1.00 (0.13)	1.00 (0.48)	2.00 (0.07)	1.00 (0.06)	1.00 (0.22)	0*
(c)	0.86 <sup>†</sup>	1.80 (0.40)	0.95 (0.40)	1.10 (1.18)	1.75 (0.21)	1.06 (0.30)	1.01 (0.52)	-0.25
	0.45 <sup>‡</sup>	1.99 (0.18)	1.00 (0.15)	0.99 (0.56)	2.00 (0.08)	1.00 (0.06)	1.00 (0.26)	0*
(d)	0.81 <sup>†</sup>	1.68 (0.42)	1.04 (0.39)	1.04 (1.18)	1.66 (0.19)	1.02 (0.21)	1.01 (0.51)	-0.35
	0.46 <sup>‡</sup>	1.99 (0.24)	1.00 (0.21)	0.99 (0.76)	2.00 (0.11)	1.00 (0.11)	1.00 (0.35)	0*
(e)	0.79 <sup>†</sup>	1.61 (0.40)	1.03 (0.38)	1.03 (1.15)	1.59 (0.20)	1.01 (0.18)	1.01 (0.52)	-0.41
	0.46 <sup>‡</sup>	1.98 (0.28)	1.00 (0.25)	0.98 (0.87)	2.00 (0.14)	0.99 (0.13)	1.01 (0.42)	0*

The empirical mean (standard deviation) for both the intercept and slope parameters are provided. The asymptotic biases of the intercept estimator ( $\delta_\alpha$ ) are also presented. (i)  $\zeta \sim N(0, 0.5^2)$ ; (ii)  $\zeta \sim Gumbel(-0.5\gamma, 0.5)$ ; (iii)  $\zeta \sim Laplace(0, 0.5)$ ; (iv)  $\zeta \sim Logistic(0, 0.5)$ ; and (v)  $\zeta \sim T(0, df = 30)$ .

<sup>†</sup> $\tau = 1.5$ .

<sup>‡</sup> $\tau = 4$ .

\*Value  $\in (-0.01, 0)$ .

support. Thus, we conduct a simulation study to compare these three methods under different settings. We choose the smooth weight function  $p(n^{\lambda t} - c)$  in the modified Buckley–James estimator (Ying, 1993) as

$$p(n^{\lambda t} - c) = \begin{cases} 0, & \text{if } n^{\lambda t} - c \leq 0 \\ \{\cos(\pi(n^{\lambda t} - c - 1)) + 1\}/2, & \text{if } 0 < n^{\lambda t} - c < 1 \\ 1, & \text{if } n^{\lambda t} - c \geq 1 \end{cases}$$

with  $c = 2$  and  $\lambda = 0.9$ , which satisfies the twice differentiability condition. Failure times are generated from the model  $T = 2 + X + \zeta$  with  $\zeta \sim N(0, 0.5^2)$ . We consider two covariate settings: (i) unbounded with  $X \sim N(0, 1)$  and (ii) bounded with  $X \sim U(-1, 1)$ . The censoring distribution is  $C \sim U(0, 4) \wedge \tau$ . Similar to the first set of simulations, we choose  $\tau = 1$  and  $\tau = 3$  to yield high and moderate censoring rates. We simulate 1000 runs for each setting and report the simulation results in Table 2 for two different sample sizes: 100 and 400.

For short follow-up time ( $\tau = 1$ ) and unbounded covariate support ( $X \sim N(0, 1)$ ), the Gehan-weighted slope estimators have negligible bias under both sample sizes, while the standard and modified Buckley–James slope estimators have noticeable biases under both sample sizes, with smaller biases from the modified Buckley–James method. The corresponding intercept estimators under three methods have the same trend in terms of bias. With bounded covariate, the Gehan-weighted slope estimator and the modified Buckley–James slope estimator have the similar biases (negligible under bigger sample size), which are smaller than the bias from the standard Buckley–James slope estimator. For the corresponding intercept estimators, the Gehan-weighted method yields the smallest biases, whereas the standard Buckley–James method yields the largest biases. When the follow-up time is long ( $\tau = 3$ ), with either unbounded or bounded covariate, all three methods yield virtually unbiased estimators for both intercept and slope.

4.2. Failure time prediction

We also compare the survival time prediction accuracy of the linear model with the Cox model via simulations. In order to have a fair comparison, we generate data from the following model:

$$T = X + e_0, \tag{4.1}$$

where  $e_0$  follows the standard extreme value distribution with  $F(t) = 1 - e^{-e^t}$ . It is well-known that such a model setting fits both the linear regression model and the Cox model. In model

Table 2. Comparison of three methods for estimating both the slope and the intercept parameters

Cen. rate	$\alpha_{GH}$	$\beta_{GH}$	$\alpha_{BJ}$	$\beta_{BJ}$	$\alpha_{MBJ}$	$\beta_{MBJ}$
<i>n</i> = 100						
(a) $X \sim N(0, 1)$						
0.84 <sup>†</sup>	2.00 (0.27)	1.01 (0.19)	1.79 (0.20)	0.85 (0.15)	1.83 (0.20)	0.88 (0.15)
0.52 <sup>‡</sup>	2.00 (0.08)	1.01 (0.08)	1.97 (0.07)	0.97 (0.07)	1.98 (0.07)	0.99 (0.08)
(b) $X \sim U(-1, 1)$						
0.92 <sup>†</sup>	1.87 (0.45)	1.10 (0.54)	1.65 (0.29)	0.85 (0.35)	1.68 (0.28)	0.89 (0.34)
0.52 <sup>‡</sup>	2.00 (0.07)	1.00 (0.12)	1.99 (0.07)	0.98 (0.11)	1.99 (0.07)	0.99 (0.12)
<i>n</i> = 400						
(a) $X \sim N(0, 1)$						
0.84 <sup>†</sup>	1.98 (0.14)	0.99 (0.08)	1.91 (0.11)	0.93 (0.07)	1.93 (0.11)	0.96 (0.08)
0.52 <sup>‡</sup>	2.00 (0.04)	1.00 (0.05)	1.99 (0.04)	0.99 (0.04)	1.99 (0.04)	1.00 (0.04)
(b) $X \sim U(-1, 1)$						
0.92 <sup>†</sup>	1.81 (0.18)	1.01 (0.22)	1.75 (0.15)	0.95 (0.19)	1.79 (0.17)	0.99 (0.20)
0.52 <sup>‡</sup>	2.00 (0.03)	1.00 (0.06)	2.00 (0.03)	0.99 (0.06)	2.00 (0.03)	1.00 (0.06)

The true regression function is  $T = 2 + X + \zeta$  with  $\zeta \sim N(0, 0.5^2)$ . Sample sizes are  $n = 100$  and  $n = 400$ .

*GH*, Gehan-weighted estimator; *BJ*, Buckley–James estimator; *MBJ*, modified Buckley–James estimator.

<sup>†</sup> $\tau = 1.0$ .

<sup>‡</sup> $\tau = 3.0$ .

(4.1), we have  $\alpha_0 = Ee_0 = -\gamma$  and  $\beta_1 = 1$ , where  $\gamma$  is the Euler constant. Note that we only use a single covariate for illustrative purpose.

We generate censoring time from  $C \sim U(-3, 3) \wedge \tau$ , where  $\tau$  is a fixed truncation time taking different values in order to generate different censoring rates. As in the first simulation study, covariate  $X$  is generated from three distributions:  $N(0, 1)$ ,  $U(-2, 2)$ , and  $U(-1, 1)$  to represent three scenarios of the covariate support (unbounded, bounded but wide, bounded but narrow). For each simulation setting, two independent data sets are generated, namely the training set and the test set, at each simulation run. Both the linear model and the Cox model are fitted using the training set, and survival times are predicted for the test set using the fitted models.

For the linear model, the predicted survival time for subject  $i$  in the test set with covariate  $X_i^*$  is calculated as  $\hat{T}_i^{LR} = \hat{E}(T_i|X_i^*) = \hat{\alpha}_n + \hat{\beta}_n^{LR} X_i^*$ , where  $\hat{\beta}_n^{LR}$  is solved by the Gehan-weighted rank-based estimating equation and  $\hat{\alpha}_n$  is estimated from (2.5). For the Cox model, the predicted survival time is calculated by

$$\hat{T}_i^{Cox} = \int t d \left[ 1 - \exp \left\{ -\hat{\Lambda}_{0,n}(t) e^{\hat{\beta}_n^{Cox} X_i^*} \right\} \right],$$

where  $\hat{\Lambda}_{0,n}(t)$  is the Breslow estimator of the baseline cumulative hazard function  $\Lambda_0(t)$ , whereas  $\hat{\beta}_n^{Cox}$  is the partial likelihood estimator. We use the following measure to determine the prediction accuracy:

Table 3. Comparison of prediction accuracy

X	$\tau$	Cen. rate	Sample size			
			200		2000	
			Linear	Cox	Linear	Cox
(a)	-2	0.83	0.86 (1.93)	0.33 (4.96)	0.98 (1.67)	0.33 (4.96)
	-1	0.69	0.96 (1.72)	0.58 (2.86)	1.00 (1.65)	0.58 (2.85)
	0	0.55	0.99 (1.68)	0.84 (1.97)	1.00 (1.65)	0.84 (1.96)
	1	0.45	0.99 (1.67)	0.97 (1.71)	1.00 (1.65)	0.97 (1.70)
	3	0.42	1.00 (1.67)	1.00 (1.67)	1.00 (1.65)	1.00 (1.65)
		OLS		(1.66)		(1.65)
(b)	-2	0.82	0.85 (1.95)	0.32 (5.27)	0.96 (1.72)	0.31 (5.27)
	-1	0.67	0.96 (1.73)	0.54 (3.08)	1.00 (1.65)	0.54 (3.07)
	0	0.54	0.99 (1.69)	0.81 (2.06)	1.00 (1.65)	0.80 (2.05)
	1	0.46	0.99 (1.67)	0.96 (1.72)	1.00 (1.65)	0.96 (1.71)
	3	0.42	0.99 (1.67)	1.00 (1.67)	1.00 (1.65)	1.00 (1.65)
		OLS		(1.66)		(1.65)
(c)	-2	0.86	0.69 (2.41)	0.38 (4.37)	0.74 (2.24)	0.38 (4.37)
	-1	0.72	0.95 (1.76)	0.68 (2.45)	0.97 (1.70)	0.67 (2.44)
	0	0.55	0.99 (1.68)	0.93 (1.78)	1.00 (1.65)	0.93 (1.77)
	1	0.44	1.00 (1.67)	1.00 (1.67)	1.00 (1.65)	1.00 (1.65)
	3	0.41	1.00 (1.67)	1.00 (1.66)	1.00 (1.65)	1.00 (1.65)
		OLS		(1.66)		(1.65)

Relative mean prediction accuracy to the case without censoring is listed. Empirical mean of  $MSE_p$  are given in parentheses.  $MSE_p$  obtained from ordinary least squares (OLS) is also listed.

- (a):  $X \sim N(0, 1)$ .
- (b):  $X \sim U(-2, 2)$ .
- (c):  $X \sim U(-1, 1)$ .

$$MSE_p = \frac{1}{n} \sum_{i=1}^n (T_i^* - \hat{T}_i)^2, \tag{4.2}$$

where  $\hat{T}_i$  is either  $\hat{T}_i^{LR}$  or  $\hat{T}_i^{Cox}$  depending on which model is used and  $T_i^*$  is the true survival time for the  $i$ th subject in the test set. Two sample sizes for the training data set are considered:  $n = 200$  and  $n = 2000$ , and 1000 runs are conducted for each simulation setting. We consider a sample size of 10,000 for the test data set in all scenarios. The results are summarized in Table 3. For each scenario, we calculate the relative prediction accuracy to the case without censoring, that is, the ratio of the empirical mean  $MSE_p$  under no censoring to that under each corresponding censored case, in addition to reporting the empirical mean of  $MSE_p$  (given in parentheses). Note that  $\tau \geq 3$  implies no truncation. The  $MSE_p$  obtained from ordinary least squares is also listed for each no-censoring scenario.

From Table 3, we see that the linear model is much less sensitive to the truncation time, especially for wide covariate support, for example,  $X \sim N(0, 1)$  and  $X \sim U(-2, 2)$ , where the linear model yields almost perfect prediction error regardless of truncation time. The linear model performs uniformly better than the Cox model. The Cox model does extremely poorly in cases with heavy truncation. This is not surprising because the baseline hazard function in the Cox model is not estimable after the last observation time ( $\leq \tau$ ) in the training set. The

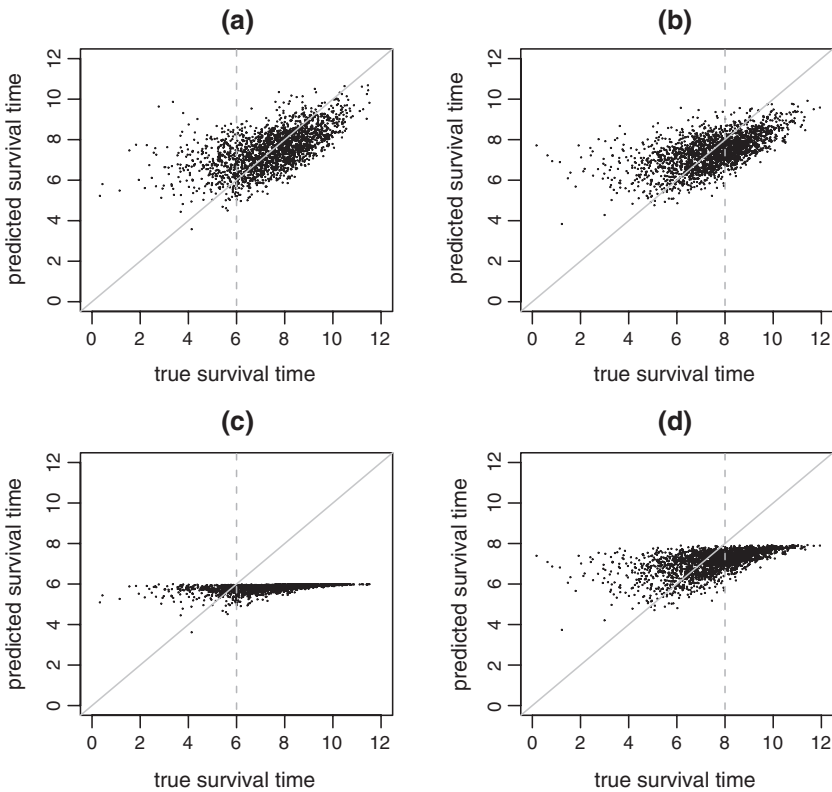


Fig. 2. The predicted survival time versus the true survival time for data generated from model (4.1) with  $X \sim N(0, 1)$  and  $n = 2000$ . (a): linear model with  $\tau = -2$ ; (b): linear model with  $\tau = 0$ ; (c): Cox model with  $\tau = -2$ ; and (d): Cox model with  $\tau = 0$ . Constant 8 was added to shift all the simulated survival times to positive values.

convention is to set the failure time distribution function to be 1 after that time point. This introduces bias when predicting the survival time, and obviously, the bias becomes more severe when the follow-up time is shorter. The difference between the two models is clearly seen from Fig. 2. The two models perform equally well when there is no censoring (Table 3).

**5. A real data example**

We consider the well-known Mayo PBC study as an illustrative example (Fleming & Harrington, 1991, app. D.1). The data contain information about the survival time and prognostic factors for 418 patients. Jin *et al.* (2003) and Jin *et al.* (2006) fitted the accelerated failure time model with five covariates, namely age, log(albumin), log(bilirubin), edema, and log(protime). They estimated slope parameters for those covariates using rank-based and least

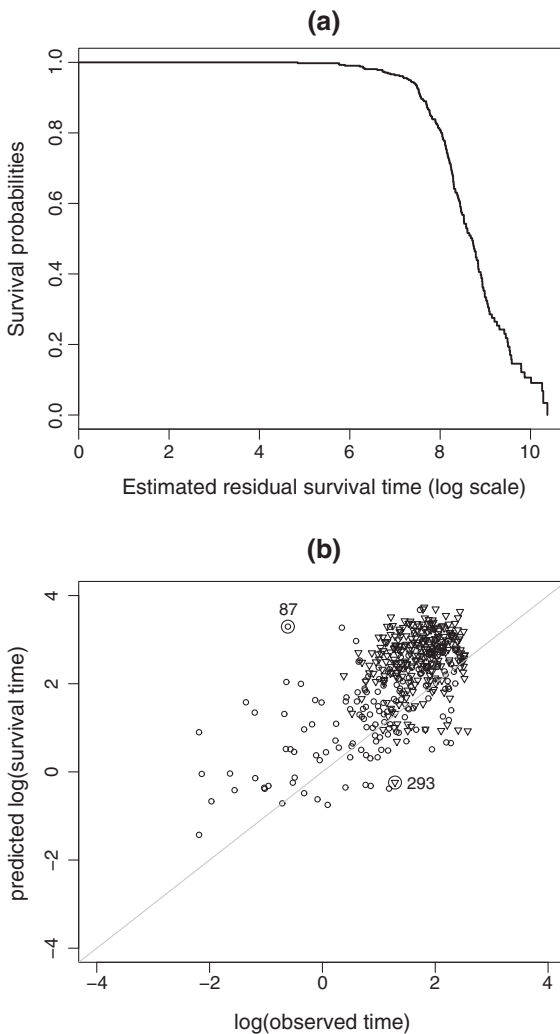


Fig. 3. (a): K–M plot of the estimated residual survival time for the PBC data. (b): Predicted survival time versus the observed time points (in the logarithm scale) for the PBC data. Circle: individual who failed; triangle: individual who was censored. Subjects 87 and 293 are two potential outliers.

squares methods. We consider the same model. The slope parameters are estimated by Gehan-weighted rank-based approach, and the intercept estimator is obtained by (2.5). The estimated coefficients for the five prognostic factors are  $(-0.025, 1.498, -0.554, -0.904, -2.822)$  with estimated standard errors  $(0.005, 0.479, 0.052, 0.234, 0.923)$ , which are similar to those reported in Jin *et al.* (2003). The intercept estimator is 8.692. The right tail of the Kaplan–Meier curve of the residual survival time almost touches zero (Fig. 3A), which indicates a valid intercept estimation from (2.5) for the PBC data.

We then perform the leave-one-out cross-validation to check the prediction performance of the model. Figure 3B shows the predicted survival time against the observed time in the logarithm scale. The circles correspond to the patients who failed, and the triangles correspond to the patients who were censored. The figure suggests that the accelerated failure time model provides a reasonable prediction of the survival time for this data set with most of the censored subjects having predicted survival times above the 45-degree line, except for a few subjects who might be outliers. For example, subject 87 (circled in Fig. 3B) was a 37-year-old woman with quite good prognostic status: no edema, good albumin (4.4), low bilirubin (1.1), and moderate protime (10.7). Yet she survived for no longer than roughly half a year. Subject 293, on the other hand, was a 57-year-old woman with poor prognostic status. In spite of low albumin (2.98), high bilirubin (8.5), and protime (12.3), and edema resistant to diuretics, she remains alive after more than 3.5 years. This same subject was also detected as an outlier in the residual plot for the covariate edema from a Cox model for the same data (Fleming & Harrington, 1991, p. 184).

## 6. Concluding remarks

In practice, the mean survival time can be well estimated when the follow-up time is long or the covariate range is wide (even with a short follow-up time). The first situation corresponds to the well-known technical condition that the support of the censoring time contains the support of the survival time and the second situation corresponds to the technical condition of unbounded covariate support that we have established in this article.

Model checking is very important in data analysis. For the linear model, one can follow the method developed for the Cox model by visualizing the cumulative sums of the martingale-based residuals to assess how unusual the observed residual patterns would be, see, for example, Lin *et al.* (1996, 1993).

Bias and variance trade-off plays an important role in assessing prediction errors, which requires knowing the asymptotic joint distribution of both the intercept and slope parameter estimators. We do not pursue it here. We also want to point out that any prediction beyond the follow-up time needs to be interpreted cautiously because it lacks empirical verification without obtaining new data with extended follow-up.

The asymptotic distribution of the intercept estimator is still unknown. However, a trimmed mean can be estimated at  $n^{1/2}$ -rate with an asymptotically normal distribution. We refer to Ding (2010) for details.

## Acknowledgements

The authors would like to thank two referees and an associate editor for their very helpful comments.

Ding's research is supported in part by NIH grants P30CA047904 and R01-EY024226. Nan's research is supported in part by NSF grant DMS-10-07590 and NIH grant R01-AG036802.

## References

- Buckley, J. & James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429–436.
- Cox, D. R. (1972). Regression models and lifetables. *J. R. Stat. Soc. Ser. B* **34**, 187–220.
- Cox, D. R. & Oakes, D. (1984). *Analysis of survival data*, CRC Press.
- Ding, Y. (2010). *Some new insights about the accelerated failure time model*, Ph.D. Thesis, Biostatistics, University of Michigan.
- Ding, Y. & Nan, B. (2011). A sieve M-theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data. *Ann. Stat.* **39**, 3032–3061.
- Etzioni, R. D., Feuer, E. J., Sullivan, S. D., Lin, D. Y., Hu, C. & Ramsey, S. D. (1999). On the use of survival analysis techniques to estimate medical care costs. *J. Health. Econ.* **18**, 365–380.
- Fleming, T. R. & Harrington, D. P. (1991). *Counting processes and survival analysis*, Wiley, New York.
- Fyngenson, M. & Ritov, Y. (1994). Monotone estimating equations for censored data. *Ann. Stat.* **22**, 732–746.
- Heller, G. & Simonoff, J. S. (1990). A comparison of estimators for regression with a censored response variable. *Biometrika* **77**, 515–520.
- 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. (2002). *The statistical analysis of failure time data*, (2nd edn), Wiley, Hoboken, NJ.
- Lai, T. L. & Ying, Z. (1988). Stochastic integrals of empirical-type processes with applications to censored regression. *J. Multivariate Anal.* **27**, 334–358.
- Lai, T. L. & Ying, Z. (1991). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Ann. Stat.* **10**, 1370–1402.
- Lin, D. Y., Robins, J. M. & Wei, L. J. (1996). Comparing two failure time distributions in the presence of dependent censoring. *Biometrika* **83**, 381–393.
- Lin, D. Y., Wei, L. J. & Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–572.
- Lin, D. Y., Wei, L. J. & Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85**, 605–618.
- Nan, B., Kalbfleisch, J. D. & Yu, M. (2009). Asymptotic theory for the semiparametric accelerated failure time model with missing data. *Ann. Stat.* **37**, 2351–2376.
- Paltiel, A. D., Freedberg, K. A., Scott, C. A., Schackman, B. R., Losina, E., Wang, B., Seage, I. G. R., Sloan, C. E., Sax, P. E. & Walensky, R. P. (2009). HIV preexposure prophylaxis in the United States: impact on lifetime infection risk, clinical outcomes, and cost-effectiveness. *Clin. Infect. Dis.* **48**, 806–815.
- Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65**, 167–179.
- Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Stat.* **18**, 303–328.
- Schneider, H. & Weissfeld, L. (1986). Estimation in linear models with censored data. *Biometrika* **73**, 741–745.
- Stute, W. & Wang, J. L. (1993). The strong law under random censorship. *Ann. Stat.* **21**, 1591–1607.
- Susarla, V., Tsai, W. Y. & Van Ryzin, J. (1984). A Buckley-James-type estimator for the mean with censored data. *Biometrika* **71**, 624–625.
- Susarla, V. & Van Ryzin, J. (1980). Large sample theory for an estimator of the mean survival time from censored samples. *Ann. Stat.* **8**, 1002–1016.
- Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Stat.* **18**, 354–372.
- van der Vaart, A. W. & Wellner, J. A. (1996). *Weak convergence and empirical processes*, Springer, New York.
- Wang, S., Nan, B., Zhu, J. & Beer, D. G. (2008). Doubly penalized Buckley-James method for survival data with high-dimensional covariates. *Biometrics* **64**, 132–140.
- Wei, L. J., Ying, Z. & Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845–851.
- Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Stat.* **21**, 76–99.
- Yu, M. G. & Nan, B. (2006). A hybrid Newton-type method for censored survival data using double weights in linear models. *Lifetime Data Anal.* **12**, 345–364.
- Zeng, D. & Lin, D. Y. (2007). Efficient estimation for the accelerated failure time model. *J. Am. Stat. Assoc.* **102**, 1387–1396.



Zhao, Y., Zeng, D., Socinski, M. A. & Kosorok, M. R. (2011). Reinforcement learning strategies for clinical trials in non-small cell lung cancer. *Biometrics* **67**, 1422–1433.

*Received July 2013, in final form June 2014*

Ying Ding, 130 DeSoto Street, 318C Parran Hall, Pittsburgh, 15261, USA.

E-mail: yingding@pitt.edu

### **Supporting information**

Additional supporting information may be found in the online version of this article at the publisher's web site including:

Proofs of Technical Lemmas.

Proofs of Theorems 2.1 and 2.2.