Rank-based inference for the accelerated failure time model with partially interval-censored data

Taehwa Choi, Sangbum Choi and Dipankar Bandyopadhyay

Sungshin Women's University, Korea University and Virginia Commonwealth University

Supplementary Material

This supplementary material provides the technical details of asymptotic results (Section S1), the details of the rank regression framework for the DC endpoint (Section S2), and additional simulation results relegated from the main paper (Section S3).

S1 Asymptotic Results

This section provides theoretical justification of the proposed methods. We assume the following regularity conditions for the asymptotic results.

(C1) The true value of β, denoted by β_0, lies in the interior of a known compact set $\mathbb{B} \subset \mathbb{R}^p$. The covariate X is uniformly bounded, i.e., $\sup_i \|X_i\| < \infty$ for $i = 1, \ldots, n$.

(C2) The residual distribution $F_0 \in \mathcal{F}$ is uniformly bounded away from 0, and has a density with continuous derivative bounded away from 0 on their support.

(C3) The distribution of Δ depends only on the observed data $\{\Delta, \Delta T, (1-\Delta)U, (1-\Delta)V, X\}$. There exists a positive constant c_0 such that $P(\Delta = 1|X) > c_0$ with probability 1.

(C4) The joint density of the examination times $(W_1, \ldots, W_K)$ given $\Delta = 0$ is continuous and differentiable in their support with respect to some dominating measure. There exists a positive constant τ_0 such that $P(\min_{0 \le k \le K-1}(W_{k+1} - W_k) > \tau_0 | X, K, \Delta = 0) = 1$.

Some necessary notations are as follows. Let $\mathbb{B}$ be the parameter space in $\mathbb{R}^p$, and $\|\cdot\|$ be the supremum norm in the metric space $\mathbb{R}^p$. Define $\mathbb{P}_n$ and P_0 as the empirical and true probability measures, respectively, along with the corresponding empirical process $\mathbb{G}_n = n^{1/2}(\mathbb{P}_n - \mathrm{P}_0)$. Denote $\mathbb{S}_n^{(k)}(\beta, t) = \mathbb{P}_n\{\eta_1 I(u_\beta \ge t) X^k\}$ and $\mathrm{S}_0^{(k)}(\beta, t) = \mathrm{P}_0\{\eta_1 I(u_\beta \ge t) X^k\}$ for $k = 0, 1$, where $u_\beta \equiv u(\beta) = \log \tilde{U} - \beta' X$.

Lemma 1. The classes of functions $\{\eta_2[\mathbb{S}_n^{(0)}(\beta, t)X - \mathbb{S}_n^{(1)}(\beta, t)] : \beta \in \mathbb{R}^p, t \in \mathbb{R}\}$ and $\{\eta_2[\mathrm{S}_0^{(0)}(\beta, t)X - \mathrm{S}_0^{(1)}(\beta, t)] : \beta \in \mathbb{R}^p, t \in \mathbb{R}\}$ are Donsker.

Proof. From Exercise 9 and 14 of Section 2.6 in van der Vaart and Wellner (1996), the class of indicator function of half space is a VC-class, and thus a Donsker class. Hence, both $\mathcal{F}_0 = \{\eta_1 I(u_\beta \ge t) : \beta \in \mathbb{R}^p, t \in \mathbb{R}\}$ and $\mathcal{F}_1 = \{\eta_2 I(u_\beta \ge t)X : \beta \in \mathbb{R}^p, t \in \mathbb{R}\}$ are Donsker. Let $\bar{\mathcal{F}}_k$ be the closure of $\mathcal{F}_k$ $(k = 0, 1)$. By Theorems 2.10.2 and 2.10.3 of van der Vaart and Wellner (1996), $\mathbb{S}_n^{(k)}(\beta, u)$ and $\mathrm{S}_0^{(k)}(\beta, u)$ belong to the convex hull of $\bar{\mathcal{F}}_k$, so they are also in Donsker classes. By their Theorem 2.10.6, the set of bounded functions $\{\eta_2[\mathbb{S}_n^{(0)}(\beta, t)X - \mathbb{S}_n^{(1)}(\beta, t)] : \beta \in \mathbb{R}^p, t \in \mathbb{R}\}$ and $\{\eta_2[\mathrm{S}_0^{(0)}(\beta, t)X - \mathrm{S}_0^{(1)}(\beta, t)] : \beta \in \mathbb{R}^p, t \in \mathbb{R}\}$ are Donsker, and thus Glivenko-Cantelli. □

Theorem 1. Under conditions (C1)–(C4), the proposed regression estimator $\hat{\beta}$ is strongly consistent for β_0, and $n^{1/2}(\hat{\beta}-\beta_0)$ converges in distribution to a zero-mean normal distribution with covariance matrix $\Gamma = A^{-1}\Omega(A^{-1})'$.

Proof. Let $\psi_\beta(\mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}) = \eta_2\{\mathbb{S}_n^{(0)}(\beta, v_\beta)X - \mathbb{S}_n^{(1)}(\beta, v_\beta)\}$ and $\psi_\beta(\mathrm{S}_0^{(0)}, \mathrm{S}_0^{(1)}) = \eta_2\{\mathrm{S}_0^{(0)}(\beta, v_\beta)X - \mathrm{S}_0^{(1)}(\beta, v_\beta)\}$, where $v_\beta \equiv v(\beta) = \log\tilde{V} - \beta' X$. Define

$$\Psi_n(\beta, \mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}) = \mathbb{P}_n\psi_\beta(\mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}), \quad \Psi_0(\beta, \mathrm{S}_0^{(0)}, \mathrm{S}_0^{(1)}) = \mathrm{P}_0\psi_\beta(\mathrm{S}_0^{(0)}, \mathrm{S}_0^{(1)}). \tag{S1.1}$$

Notice that Ψ_n and Ψ_0 represent random and deterministic maps, respectively, corresponding to the estimating equation $S_n(\beta)$ and its limiting function $S_0(\beta)$. The Gehan estimator $\hat{\beta}$ can be defined as a Z-estimator of the map $\Psi_n(\hat{\beta}, \mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}) = o_p(n^{-1/2})$, while the true parameter β_0 satisfies $\Psi_0(\beta_0, \mathrm{S}_0^{(0)}, \mathrm{S}_0^{(1)}) = 0$, because

$$\begin{aligned}\Psi_0(\beta_0, \mathrm{S}_0^{(0)}, \mathrm{S}_0^{(1)}) &= E\left[\eta_2[E\{\eta_1 I(u_{\beta_0} \geq v_{\beta_0})\}X - E\{\eta_1 I(u_{\beta_0} \geq v_{\beta_0})X\}]\right] \\ &= E\left[\eta_2\{E(\eta_1)X - \eta_1 X\}\right] = 0.\end{aligned}$$

To check the convergence of the estimating function Ψ_n to Ψ_0, we use the triangle inequality to induce the following result,

$$\begin{aligned}&\|\Psi_n(\beta, \mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}) - \Psi_0(\beta, \mathrm{S}_0^{(0)}, \mathrm{S}_0^{(1)})\| \\ &\quad = \|\Psi_n(\beta, \mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}) - \mathrm{P}_0\psi_\beta(\mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}) + \mathrm{P}_0\psi_\beta(\mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}) - \Psi_0(\beta, \mathrm{S}_0^{(0)}, \mathrm{S}_0^{(1)})\| \\ &\quad \leq \|(\mathbb{P}_n - \mathrm{P}_0)[\eta_2\{\mathbb{S}_n^{(0)}X - \mathbb{S}_n^{(1)}\}]\| + \|\mathrm{P}_0[\eta_2\{\mathbb{S}_n^{(0)} - \mathrm{S}_0^{(0)}\}X]\| + \|\mathrm{P}_0[\eta_2\{\mathbb{S}_n^{(1)} - \mathrm{S}_0^{(1)}\}]\|.\end{aligned}$$

The first term on the right side of the above inequality converges to zero in probability by the Glivenko-Cantelli property, as shown in Lemma 1. The second and third terms also converge to zero in probability, since all components in $\mathbb{S}_n^{(k)}$ and $\mathrm{S}_0^{(k)}$ for $k = 0, 1$ are Glivenko-Cantelli, and $\|\mathrm{P}_0[\eta_2 X\{\mathbb{S}_n^{(0)} - \mathrm{S}_0^{(0)}\}]\| \leq \|\mathbb{S}_n^{(0)} - \mathrm{S}_0^{(0)}\| \cdot \mathrm{P}_0|\eta_2 X| \to_p 0$ and $\|\mathrm{P}_0[\eta_2\{\mathbb{S}_n^{(1)} - \mathrm{S}_0^{(1)}\}]\| \leq \|\mathbb{S}_n^{(1)} - \mathrm{S}_0^{(1)}\| \cdot \mathrm{P}_0|\eta_2| \to_p 0$. Therefore, $\|\Psi_n(\beta, \mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}) - \Psi_0(\beta, \mathrm{S}_0^{(0)}, \mathrm{S}_0^{(1)})\| \to_p 0$. By Lemma 1, this result can be further improved to give

$$\|n^{1/2}\{\Psi_n(\beta, \mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}) - \Psi_0(\beta, \mathrm{S}_0^{(0)}, \mathrm{S}_0^{(1)})\}\| = O_p(1). \tag{S1.2}$$

To prove the strong consistency of $\hat{\beta}$, we first note that β_0 is the unique solution to $\Psi_0(\beta_0, \mathrm{S}_0^{(0)}, \mathrm{S}_0^{(1)}) = 0$ by the condition (C3). Then, for any fixed $\epsilon > 0$, there exists a $\delta > 0$, such that $P(|\hat{\beta} - \beta_0| > \epsilon) \leq P(|\Psi_0(\hat{\beta}, \mathrm{S}_0^{(0)}, \mathrm{S}_0^{(1)})| > \delta)$. By the definition of almost sure convergence, it suffices to show that $|\Psi_0(\hat{\beta}, \mathrm{S}_0^{(0)}, \mathrm{S}_0^{(1)})| \to_p 0$. The standard property of Glivenko-Cantelli implies $\mathbb{S}_n^{(0)} \to \mathrm{S}_0^{(0)}$ and $\mathbb{S}_n^{(1)} \to \mathrm{S}_0^{(1)}$, with probability tending to one. Hence, from (S1.2), we have the desired inequalities

$$|\Psi_0(\hat{\beta}, \mathrm{S}_0^{(0)}, \mathrm{S}_0^{(1)})| \leq |\Psi_n(\hat{\beta}, \mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)})| + |\Psi_n(\hat{\beta}, \mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}) - \Psi_0(\hat{\beta}, \mathrm{S}_0^{(0)}, \mathrm{S}_0^{(1)})| = o_p(1),$$

which implies $P(|\hat{\beta} - \beta_0| > \epsilon) \to 0$, and hence, $\hat{\beta}$ is consistent. It can be easily shown that (S1.2) further implies $n^{1/2}(\hat{\beta} - \beta_0) = O_p(1)$.

Next, we prove the asymptotic normality of $\hat{\beta}$. Let $|\beta - \beta_0| \leq K n^{-1/2}$ with $K < \infty$. Then, we have

$$\begin{aligned} n^{1/2}\{\Psi_n(\beta, \mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}) - \Psi_n(\beta_0, \mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)})\} &= n^{1/2}\mathbb{P}_n\{\psi_\beta(\mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}) - \psi_{\beta_0}(\mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)})\} \\ &= \mathbb{G}_n\{\psi_\beta(\mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}) - \psi_{\beta_0}(\mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)})\} + n^{1/2}\mathrm{P}_0\{\psi_\beta(\mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}) - \psi_{\beta_0}(\mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)})\}. \end{aligned} \tag{S1.3}$$

The first term of the right side in (S1.3) equals $\mathbb{G}_n[\eta_2\{\mathbb{S}_n^{(0)}(\beta,\cdot)-\mathbb{S}_n^{(0)}(\beta_0,\cdot)\}X]-\mathbb{G}_n[\eta_2\{\mathbb{S}_n^{(1)}(\beta,\cdot)-\mathbb{S}_n^{(1)}(\beta_0,\cdot)\}]$, which converges to zero in probability, since $\mathbb{S}_n^{(k)}(\beta,\cdot)-\mathbb{S}_n^{(k)}(\beta_0,\cdot)$ $(k=0,1)$ converge to zero in quadratic mean, while the second term is asymptotically equivalent to $n^{1/2}A(\beta-\beta_0)+o_p(1)$, where $A=E\{\partial\psi_\beta(\mathrm{S}_0^{(0)},\mathrm{S}_0^{(1)})/(\partial\beta)|_{\beta=\beta_0}\}$. Therefore, equation (S1.3), together with $\Psi_n(\hat\beta,\mathbb{S}_n^{(0)},\mathbb{S}_n^{(1)})=o_p(n^{-1/2})$, yields the asymptotic linearity when β is replaced by $\hat\beta$, such that

$$n^{1/2}\Psi_n(\beta_0,\mathbb{S}_n^{(0)},\mathbb{S}_n^{(1)})=-n^{1/2}A(\hat\beta-\beta_0)+o_p(1). \tag{S1.4}$$

If A is continuously invertible, the above equation yields $n^{1/2}(\hat\beta-\beta_0)=-A^{-1}\mathbb{G}_n\psi_{\beta_0}(\mathbb{S}_n^{(0)},\mathbb{S}_n^{(1)})+o_p(1)$. Then, by the Theorem 3.3.1 of van der Vaart and Wellner (1996) and the consistency result of $\hat\beta$, $n^{1/2}(\hat\beta-\beta_0)$ converges in distribution to the zero-mean normal distribution with covariance matrix $\Gamma=A^{-1}\Omega(A^{-1})'$, where $\Omega=\mathrm{cov}\{\psi_{\beta_0}(\mathrm{S}_0^{(0)},\mathrm{S}_0^{(1)})\}=E\{\psi_{\beta_0}(\mathrm{S}_0^{(0)},\mathrm{S}_0^{(1)})'\psi_{\beta_0}(\mathrm{S}_0^{(0)},\mathrm{S}_0^{(1)})\}$. □

Remark 1. Using a similar argument of the proof of Theorem 1, we can prove the strong consistency of $\hat\beta_\phi$ for β_0. Note that $\psi_\phi(\mathrm{S}_0^{(0)},\mathrm{S}_0^{(1)})=\eta_2\{X-\mathrm{S}_0^{(1)}(\beta,v_\beta)/\mathrm{S}_0^{(0)}(\beta,v_\beta)\}$ and $\psi_{\phi_0}\equiv\psi_{\phi_0}(\mathrm{S}_0^{(0)},\mathrm{S}_0^{(1)})=\eta_2\{X-\mathrm{S}_0^{(1)}(\beta,v_\beta)/\mathrm{S}_0^{(0)}(\beta,v_{\beta_0})\}$. Furthermore, asymptotic normality of $\hat\beta_\phi$ can be derived by using Theorem 3.3.1 of van der Vaart and Wellner (1996), such that $n^{1/2}(\hat\beta_\phi-\beta_0)\to N(0,\Gamma_\phi)$, $\Gamma_\phi=A_\phi^{-1}\Omega_\phi(A_\phi^{-1})'$, where $A_\phi=E\{\partial\psi_\phi(S_0^{(0)},S_0^{(1)})/(\partial\beta)|_{\beta=\beta_0}\}$ and $\Omega_\phi=E\{\psi_{\phi_0}(S_0^{(0)},S_0^{(1)})'\,\psi_{\phi_0}(S_0^{(0)},S_0^{(1)})\}$ with $\psi_\phi(\mathrm{S}_0^{(0)},\mathrm{S}_0^{(1)})=\eta_2\{X-\mathrm{S}_0^{(1)}(\beta,v_\beta)/\mathrm{S}_0^{(0)}(\beta,v_\beta)\}$.

Theorem 2. Under conditions (C1)–(C4), for any $k\geq 1$, an iterative estimator $\hat\beta_{(k)}$

is strongly consistent for β_0, and $n^{1/2}(\hat{\beta}_{(k)} - \beta_0)$ converges to the same distribution of $n^{1/2}(\hat{\beta}_\phi - \beta_0)$ as $k \to \infty$.

Proof. From Theorem 1, an initial estimator $\hat{\beta}_{(0)} = \hat{\beta}$ is strongly consistent for β_0, implying $n^{-1}L_\phi(\beta; \hat{\beta}_{(0)})$ converges almost surely to $\lim_{n\to\infty} n^{-1}L_\phi(\beta; \beta_0)$, with its minimizer at $\beta = \beta_0$. Because $n^{-1}L_\phi(\beta, \hat{\beta}_{(0)})$ is convex, $\hat{\beta}_{(1)}$ converges to β_0 almost surely. This argument can be successively applied for $k \geq 2$, and we argue that $\hat{\beta}_{(k)} \to_p \beta_0$ as $n \to \infty$ for all k.

Assume $\phi(t) = 0$ for t near the endpoint to avoid possible tail instabilities. We further assume that for any β_n and ϵ_n converging to β_0 and 0 in probability, respectively,

$$w(\beta_n, t + \epsilon_n) \equiv \frac{\phi(\beta_n)}{\sum_{j=1}^n \eta_{1j} I(t_i + \epsilon_n \leq t_j + \epsilon_n)} = w(\beta_n, t) + \dot{w}_0(t)\epsilon_n + o(n^{-1/2} + \epsilon_n),$$

where, $\dot{w}_0(t)$ is the derivative of the limit of $w(\beta_0, t)$ as $n \to \infty$. Then, we expand

$$\begin{aligned} S_\phi(\beta; \hat{\beta}_{(k-1)}) = {} & n^{-1} \sum_{i=1}^n w_i\{\hat{\beta}_{(k-1)}, v_i(\hat{\beta}_{(k-1)})\}\eta_{2i}\eta_{1j}(X_i - X_j) I\{u_i(\beta) \leq v_j(\beta)\} \\ & + n^{-1} \sum_{i=1}^n \dot{w}_{0i}(t)\eta_{2i}\eta_{1j}(X_i - X_j) I\{u_i(\beta) \leq v_j(\beta)\}(\beta - \hat{\beta}_{(k-1)}) \\ & + o(n^{1/2} + n\|\beta - \hat{\beta}_{(k-1)}\|). \end{aligned}$$

By Remark 1, the first term on the right-hand side of above equation has a linear expansion $\sum_{i=1}^n \psi_{\phi_0,i} + n(\beta - \beta_0)A_\phi + o(n^{1/2} + n\|\beta - \beta_0\|)$. Thus,

$$\begin{aligned} S_\phi(\hat{\beta}_{(k)}; \hat{\beta}_{(k-1)}) = {} & \sum_{i=1}^n \psi_{\phi_0,i} + n(A_\phi + B_\phi)(\hat{\beta}_{(k)} - \beta_0) - nB_\phi(\hat{\beta}_{(k-1)} - \beta_0) \\ & + o(n^{1/2} + n\|\hat{\beta}_{(k)} - \beta_0\| + n\|\hat{\beta}_{(k-1)} - \beta_0\|), \end{aligned} \tag{S1.5}$$

with

$$B_\phi = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n}\sum_{j=1}^{n} \dot{w}_{0i}\eta_{2i}\eta_{1j} I(u_{0i} \le v_{0j}) \left\{ X_i - \frac{\sum_{j=1}^{n} \eta_{1j} X_j I(u_{0i} \le v_{0j})}{\sum_{j=1}^{n} \eta_{1j} I(u_{0i} \le v_{0j})} \right\}^{\otimes 2},$$

where, B_ϕ is assumed to be a nonsingular limiting slope matrix of $S_\phi(\beta; b)$ for β_0, $u_{0i} \equiv u_i(\beta_0)$ and $v_{0i} \equiv v_i(\beta_0)$. Assume that A, A_ϕ and $A_\phi + B_\phi$ are nonsingular matrix. Then, from equation (S1.5) and the asymptotic linearity of $S_n(\hat{\beta})$ for $k = 1$, we observe

$$n^{1/2}(\hat{\beta}_{(1)} - \beta_0) = -n^{-1/2}(A_\phi + B_\phi)^{-1}\left\{ \sum_{i=1}^{n} \psi_{\phi_0,i} + B_\phi A^{-1} \sum_{i=1}^{n} \psi_{\beta_0,i} \right\} + o(1 + n^{1/2}\|\hat{\beta}_{(1)} - \beta_0\| + n^{1/2}\|\hat{\beta} - \beta_0\|).$$

Now, recursively using (S1.5), we obtain following equation in general

$$\begin{aligned} n^{1/2}(\hat{\beta}_{(k)} - \beta_0) = & -n^{-1/2} \sum_{j=1}^{k} \{(A_\phi + B_\phi)^{-1} B_\phi\}^{j-1} (A_\phi + B_\phi)^{-1} \sum_{i=1}^{n} \psi_{\phi_0,i} \\ & - n^{-1/2}\{(A_\phi + B_\phi)^{-1} B_\phi\}^{k} A^{-1} \sum_{i=1}^{n} \psi_{\beta_0,i} + o\left(1 + n^{1/2} \sum_{j=0}^{k} \|\hat{\beta}_{(j)} - \beta_0\| \right) \end{aligned}$$

or equivalently,

$$\begin{aligned} n^{1/2}(\hat{\beta}_{(k)} - \beta_0) = & -n^{-1/2}[I - \{(A_\phi + B_\phi)^{-1} B_\phi\}^{k}] A_\phi^{-1} \sum_{i=1}^{n} \psi_{\phi_0,i} \\ & - n^{-1/2}\{(A_\phi + B_\phi)^{-1} B_\phi\}^{k} A^{-1} \sum_{i=1}^{n} \psi_{\beta_0,i} + o\left(1 + n^{1/2} \sum_{j=0}^{k} \|\hat{\beta}_{(j)} - \beta_0\| \right). \end{aligned} \quad \text{(S1.6)}$$

This implies that asymptotic normality of $n^{1/2}(\hat{\beta}_{(k)} - \beta_0)$ follows from that of $n^{-1/2}\sum_{i=1}^{n} \psi_{\phi_0,i}$ and $n^{-1/2}\sum_{i=1}^{n} \psi_{\beta_0,i}$. Now, $\hat{\beta}_{(k)}$ can be expressed by the weighted average of $\hat{\beta}$ and $\hat{\beta}_\phi$

$$\hat{\beta}_{(k)} = \{(A_\phi + B_\phi)^{-1} B_\phi\}^{k} \hat{\beta} + [I - \{(A_\phi + B_\phi)^{-1} B_\phi\}^{k}] \hat{\beta}_\phi + o_p(n^{-1/2}).$$

Furthermore, $n^{-1/2}\sum_{i=1}^{n}\psi_{\beta_0,i}$ vanishes because $\{(A_\phi + B_\phi)^{-1}B_\phi\}^k \to 0$ as $k \to \infty$, implying $n^{1/2}(\hat{\beta}_{(k)} - \beta_0) = n^{1/2}(\hat{\beta}_\phi - \beta_0) + o_p(1)$ for $k \to \infty$. □

S2 Doubly-censored rank regression

Now, we describe our proposed rank-based method for the doubly-censored (DC) data. Let $e_i(\beta) = \log \tilde{T}_i - \beta' X_i$ denote the residual under model (1.1). The rank estimating equation for DC data can be constructed by examining the rank between $e_i(\beta)$ and $e_j(\beta)$ for all $1 \le i < j \le n$. To perform the rank regression with DC data, we solve the following estimating function

$$\tilde{S}_n(\beta) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_{3i}\Delta_{2j}(X_i - X_j)I\{e_i(\beta) \le e_j(\beta)\}, \tag{S2.7}$$

or equivalently,

$$\tilde{L}_n(\beta) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_{3i}\Delta_{2j}\{e_i(\beta) - e_j(\beta)\}_-, \tag{S2.8}$$

where $\Delta_{2i} = \delta_{1i} + \delta_{2i}$ and $\Delta_{3i} = \delta_{1i} + \delta_{3i}$. We define the Gehan estimator as $\tilde{\beta} = \arg\min_{\beta\in\mathbb{B}} \tilde{L}_n(\beta)$, which can be obtained via linear programming as before. By introducing the general weight $\tilde{\phi}_i(\beta)$, we can also consider the generalized log-rank estimating function

$$\tilde{S}_\phi(\beta) = n^{-1}\sum_{i=1}^{n}\tilde{\phi}_i(\beta)\Delta_{3i}\left\{X_i - \frac{\sum_{j=1}^{n}\Delta_{2j}X_jI\{e_i(\beta) \le e_j(\beta)\}}{\sum_{j=1}^{n}\Delta_{2j}I\{e_i(\beta) \le e_j(\beta)\}}\right\}, \tag{S2.9}$$

where the choice of $\tilde{\phi}_i(\beta) = \sum_{j=1}^{n}\Delta_{2j}I\{e_j(\beta) \ge e_i(\beta)\}$ and 1 leads the log-rank and the Gehan estimators, respectively. Instead of directly working with (S2.9), we again formulate a monotone weighted estimating function, with its objective function as

$$\tilde{S}_\phi(\beta, b) = n^{-1}\sum_{i=1}^{n}\tilde{w}_i\{b, e_i(b)\}\Delta_{3i}\Delta_{2j}(X_i - X_j)I\{e_i(\beta) \leq e_j(\beta)\}, \tag{S2.10}$$

and

$$\tilde{L}_\phi(\beta, b) = n^{-1}\sum_{i=1}^{n}\tilde{w}_i\{b, e_i(b)\}\Delta_{3i}\Delta_{2j}\{e_i(\beta) - e_j(\beta)\}_-, \tag{S2.11}$$

where, $\tilde{w}_i(b,t) = \tilde{\phi}_i(b)/\sum_{j=1}^{n}\Delta_{2j}I\{e_j(b) \geq t\}$. Let $\tilde{\beta}_{(k)} = \arg\min_{\beta\in\mathbb{B}}\tilde{L}_\phi(\beta, \tilde{\beta}_{(k-1)})$. Then, we can finally obtain the log-rank estimator from $\tilde{\beta}_\phi = \lim_{k\to\infty}\tilde{\beta}_{(k)}$. With a suitable choice of k that depends on n, $\tilde{\beta}_{(k)}$ is asymptotically equivalent to the consistent roots of $\tilde{S}_\phi(\beta)$. Whether the algorithm converges or not, $\tilde{\beta}_{(k)}$ is consistent and asymptotically normal.

S3 Additional Simulations

This section provides further simulation results relegated from the main paper.

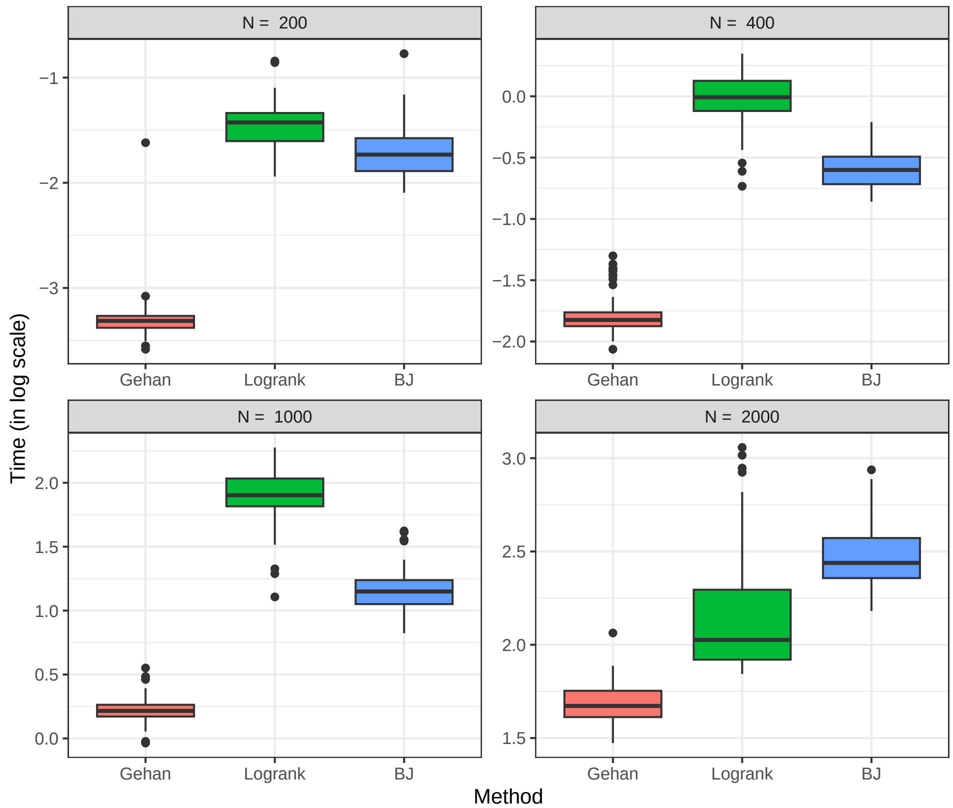

Figure S1: Computing times measured in log scale across various sample sizes. Red, green and blue colored boxes represent Gehan, log-rank and Buckley-James estimators, respectively.

We first compare the computation time of competing methods, which are summarized in Figure S1. Times are measured in seconds and transformed into the log scale. Sample sizes of n = 200, 400, 1000, and 2000 were considered. Among the three methods, the Gehan approach appears to be the fastest. When the sample sizes are either of $n = 200, 400, 1000$, the log-rank method is slower than the Buckley-James method. However, for a larger sample size ($n = 2000$), the Buckley-James method does not seems to be efficient as more computing time is required than the proposed method. This is partly because the Buckley-James method involves nonparametric function estimation, which requires additional time to arrive at the solution under larger sample sizes.

Next, we simulate the PIC data with larger sample sizes of $n = 1000$ and 2000. Various scenarios are considered, under the censoring rates of 30% and 60% and error distributions following Normal(0,1) denoted as $N(0,1)$, Extreme Value (EV), and Exponential(1), denoted as Exp(1). The simulation results are summarized in Table S1. For larger sample sizes, the proposed methods still perform very well. Biases are much smaller (nearly unbiased), than those observed with moderate sample sizes. Furthermore, the estimated standard errors are smaller, and the target coverage probabilities are well achieved.

Table S1: Simulation results for the PIC data. Table entries are the average bias (Bias), empirical standard error (ESE), asymptotic standard error (ASE), and coverage probability (CP) of the 95% Wald-type confidence intervals for the parameter estimates obtained from the Gehan and log-rank methods, under $n = 1000$ and 2000, censoring rates of 30% and 60%, and error distributions that follow Normal(0,1) denoted as $N(0,1)$, Extreme Value (EV), and Exponential(1), denoted as Exp(1).

				Gehan				Log-rank			
Censoring	Error	n	Par	Bias	ESE	ASE	CP	Bias	ESE	ASE	CP
30%	$N(0,1)$	1000	β_1	–0.002	0.034	0.032	0.944	0.003	0.034	0.035	0.956
			β_2	–0.004	0.059	0.064	0.958	0.001	0.062	0.070	0.975
		2000	β_1	0.000	0.023	0.023	0.957	0.005	0.024	0.025	0.965
			β_2	0.003	0.048	0.045	0.939	0.008	0.048	0.049	0.961
	EV	1000	β_1	–0.003	0.037	0.036	0.950	0.000	0.043	0.049	0.983
			β_2	–0.002	0.073	0.072	0.950	0.000	0.088	0.096	0.965
		2000	β_1	–0.002	0.026	0.025	0.946	0.007	0.032	0.035	0.959
			β_2	–0.003	0.050	0.051	0.952	0.003	0.064	0.068	0.956
	Exp(1)	1000	β_1	0.001	0.019	0.019	0.961	0.005	0.026	0.033	0.986
			β_2	0.001	0.038	0.038	0.945	0.003	0.055	0.064	0.971
		2000	β_1	0.000	0.013	0.013	0.951	0.007	0.024	0.023	0.942
			β_2	0.001	0.026	0.026	0.959	0.007	0.045	0.045	0.952
60%	$N(0,1)$	1000	β_1	–0.003	0.038	0.037	0.943	0.000	0.039	0.040	0.952
			β_2	–0.004	0.072	0.076	0.958	–0.002	0.075	0.081	0.972
		2000	β_1	0.000	0.028	0.027	0.937	0.004	0.028	0.028	0.948
			β_2	0.003	0.052	0.053	0.955	0.005	0.053	0.057	0.969
	EV	1000	β_1	–0.003	0.042	0.042	0.944	–0.001	0.048	0.052	0.971
			β_2	–0.002	0.082	0.085	0.952	0.000	0.094	0.104	0.973
		2000	β_1	–0.001	0.030	0.030	0.946	0.001	0.034	0.037	0.968
			β_2	–0.003	0.059	0.060	0.954	0.000	0.067	0.074	0.967
	Exp(1)	1000	β_1	0.001	0.023	0.023	0.946	0.003	0.030	0.034	0.974
			β_2	0.000	0.044	0.045	0.963	0.001	0.058	0.067	0.981
		2000	β_1	0.000	0.016	0.016	0.952	0.002	0.021	0.024	0.974
			β_2	0.002	0.030	0.031	0.959	0.004	0.040	0.047	0.984

Bibliography

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