Statistica Sinica

Efficient and Robust Estimation of τ-year Risk Prediction Models Leveraging Time Varying Intermediate Outcomes

Yu Zheng*, Tian Lu†, Tianxi Cai*

* *Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115*

† *Department of Biomedical Data Science, Stanford University, Palo Alto, CA 94305*

Supplementary Material

This supplementary material includes proof for the proposed method (Part 1) and R code for simulation studies (Part 2).

Part 1: Proof

Throughout, we let $\mathbb{B} = E\{I(T_i^\dagger > t_s)\boldsymbol{\Psi}_i^{\otimes 2}\dot{g}(\bar{\boldsymbol{\gamma}}_{\tau|t_s}^{\mathsf{T}}\boldsymbol{\Psi}_i)$, $\mathbf{L}_i = \mathbb{J}^{-1}\mathbf{X}_i^{\mathsf{T}} g(\bar{\boldsymbol{\gamma}}_{\tau|t_s}^{\mathsf{T}}\boldsymbol{\Psi}_i)I(T_i^\dagger > t_s)$,

$$\mathbb{J} = E\{\mathbf{X}_i^{\otimes 2}\dot{g}(\bar{\boldsymbol{\beta}}^{\mathsf{T}}\mathbf{X}_i)\}, \quad \mathbf{U}_i^X = \mathbf{X}_i\{Y_{\tau i} - g(\bar{\boldsymbol{\beta}}^{\mathsf{T}}\mathbf{X}_i)\}, \qquad \mathbf{F}_{1i} = \mathbb{J}^{-1}\mathbf{X}_i\{Y_{\tau i} - g(\bar{\boldsymbol{\beta}}^{\mathsf{T}}\mathbf{X}_i)\},$$

$$\mathbb{A} = E\{\boldsymbol{\Phi}_i^{\otimes 2}\dot{g}(\bar{\boldsymbol{\theta}}^{\mathsf{T}}\boldsymbol{\Phi}_i)\}, \quad \mathbf{U}_i^\Phi = \boldsymbol{\Phi}_i\{Y_{t_s i} - g(\bar{\boldsymbol{\theta}}_{t_s}^{\mathsf{T}}\boldsymbol{\Phi}_i)\}, \qquad \mathbf{F}_{2i} = \mathbb{J}^{-1}\mathbf{X}_i\{Y_{t_s i} - g(\bar{\boldsymbol{\theta}}_{t_s}^{\mathsf{T}}\boldsymbol{\Phi}_i)\},$$

$$\mathbf{U}_i^\Psi = \boldsymbol{\Psi}_i\{Y_{\tau i} - g(\bar{\boldsymbol{\gamma}}_{\tau|t_s}^{\mathsf{T}}\boldsymbol{\Psi}_i)\}, \quad \mathbf{F}_{3i} = \mathbb{J}^{-1}\mathbf{X}_i(Y_{\tau i} - g(\bar{\boldsymbol{\gamma}}_{\tau|t_s}^{\mathsf{T}}\boldsymbol{\Psi}_i).$$

For any a vector $\boldsymbol{a}$, $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^{\mathsf{T}}$. We use $\approx$ to denote equivalence up to $o_p(1)$.

A Consistency of $\widehat{\boldsymbol{\beta}}$ obtained with S at a single visit t_s

Under a mild condition that there does not exist a $\boldsymbol{\beta}$ such that $P(\boldsymbol{\beta}^\top \mathbf{X}_1 > \boldsymbol{\beta}^\top \mathbf{X}_2 | T_1^\dagger \le t \le T_2^\dagger) = 1$, using a similar argument given by Tian et al. (2007), we can show that $\mathbf{U}_0(\boldsymbol{\beta}) = 0$ has a unique solution. To show $\widehat{\boldsymbol{\beta}}$ converges to $\bar{\boldsymbol{\beta}}$ in probability, from Newey and McFadden (1994), it suffices to show that $\sup_{\boldsymbol{\beta}} |\widehat{\mathbf{U}}_n(\boldsymbol{\beta}) - \mathbf{U}_0(\boldsymbol{\beta})| = o_p(1)$. To this end, we recall that

$$\widehat{\mathbf{U}}_n(\boldsymbol{\beta}) \equiv n^{-1} \sum_{i=1}^n \mathbf{X}_i \left\{ \widehat{Y}_{\tau i}^{t_s} - g(\boldsymbol{\beta}^\top \mathbf{X}_i) \right\} = 0.$$

with

$$\widehat{Y}_{\tau i}^{t_s} = g(\widehat{\boldsymbol{\theta}}_{t_s}^\top \boldsymbol{\Phi}_i) + \widehat{\varpi}_{t_s i}\ g(\widehat{\boldsymbol{\gamma}}_{\tau|t_s}^\top \boldsymbol{\Psi}_i).$$

We first establish convergence properties for $\widehat{\boldsymbol{\theta}}_{t_s}$ and $\widehat{\boldsymbol{\gamma}}_{\tau|t_s}$. Using similar arguments as given in Uno et al. (2007) for the convergence of $\widetilde{\boldsymbol{\beta}}$ to $\bar{\boldsymbol{\beta}}$ and the fact that $\max(\lambda_1, \lambda_2) = o(n^{-\frac{1}{2}})$, we may show that $\sup_{\boldsymbol{\theta}} |\widehat{\mathbf{Q}}_n(\boldsymbol{\theta}) - \mathbf{Q}_0(\boldsymbol{\theta})| + \sup_{\boldsymbol{\gamma}} |\widehat{\mathbf{D}}_n(\boldsymbol{\gamma}) - \mathbf{D}_0(\boldsymbol{\gamma})| \to 0$ in probability, where

$$\mathbf{Q}_0(\boldsymbol{\theta}) = E[\boldsymbol{\Phi}_i \{Y_{t_s i} - g(\boldsymbol{\theta}^\top \boldsymbol{\Phi}_i)\}], \quad \text{and} \quad \mathbf{D}_0(\boldsymbol{\gamma}) = E[I(T_i^\dagger > t_s) \boldsymbol{\Psi}_i (Y_i - g(\boldsymbol{\gamma}^\top \boldsymbol{\Psi}_i))].$$

This, together with the fact that $\mathbf{Q}_0(\boldsymbol{\theta}) = 0$ and $\mathbf{D}_i(\boldsymbol{\gamma}) = 0$ respectively have unique solutions at $\bar{\boldsymbol{\theta}}_{t_s}$ and $\bar{\boldsymbol{\gamma}}_{\tau|t_s}$, implies that $\widehat{\boldsymbol{\theta}}_{t_s} \to \bar{\boldsymbol{\theta}}_{t_s}$ and $\widehat{\boldsymbol{\gamma}}_{\tau|t_s} \to \bar{\boldsymbol{\gamma}}_{\tau|t_s}$ in probability.

Next, let $\mathbf{U}_n^*(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n \mathbf{X}_i \{Y_{\tau i}^{t_s} - g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}$ and

$$Y_{\tau i}^{t_s} = g(\bar{\boldsymbol{\theta}}_{t_s}^\top \boldsymbol{\Phi}_i) + \varpi_{t_s i}\ g(\bar{\boldsymbol{\gamma}}_{\tau|t_s}^\top \boldsymbol{\Psi}_i).$$

The consistency of $\widehat{\boldsymbol{\theta}}_{t_s}$ and $\widehat{\boldsymbol{\gamma}}_{\tau|t_s}$ together with the uniform consistency of $\widehat{G}(\cdot)$ implies that $\sup_{\boldsymbol{\beta}} |\mathbf{U}_n^*(\boldsymbol{\beta}) - \widehat{\mathbf{U}}_n(\boldsymbol{\beta})| \to 0$ in probability. This, coupled with the uniform law of large numbers (Pollard, 1990), implies that $\sup_{\boldsymbol{\beta}} |\widehat{\mathbf{U}}_n(\boldsymbol{\beta}) - E\{\mathbf{U}_n^*(\boldsymbol{\beta})\}| \to 0$ in probability. It is not difficult to see that $E\{\mathbf{U}_n^*(\boldsymbol{\beta}) = \mathbf{U}_0(\boldsymbol{\beta})$ since

$$
\begin{aligned}
\mathbf{U}_0(\boldsymbol{\beta}) - E\{\mathbf{U}_n^*(\boldsymbol{\beta})\} &= E[\mathbf{X}_i\{Y_{t_s i} + I(T_i^\dagger > t_s)Y_{\tau i} - g(\boldsymbol{\beta}^\mathsf{T}\mathbf{X}_i)\}] \\
&\quad - E\left[\mathbf{X}_i\{g(\bar{\boldsymbol{\theta}}_{t_s}^\mathsf{T}\boldsymbol{\Phi}_i) + \varpi_{t_s i}\ g(\bar{\boldsymbol{\gamma}}_\tau^\mathsf{T}\boldsymbol{\Psi}_i) - g(\boldsymbol{\beta}^\mathsf{T}\mathbf{X}_i)\}\right] \\
&= E[\mathbf{X}_i\{Y_{t_s i} - g(\bar{\boldsymbol{\theta}}_{t_s}^\mathsf{T}\boldsymbol{\Phi}_i)\}] + \left[E\{\mathbf{X}_i Y_{\tau i}|T_i^\dagger > t_s\} - E\{\mathbf{X}_i g(\bar{\boldsymbol{\gamma}}_\tau^\mathsf{T}\boldsymbol{\Psi}_i)|T_i^\dagger > t_s\}\right] P(T_i^\dagger > t_s) \\
&= E[\mathbf{X}_i\{Y_{t_s i} - g(\bar{\boldsymbol{\theta}}_{t_s}^\mathsf{T}\boldsymbol{\Phi}_i)\}] + E[\mathbf{X}_i\{Y_{\tau i} - g(\bar{\boldsymbol{\gamma}}_\tau^\mathsf{T}\boldsymbol{\Psi}_i)\}|T_i > t_s]P(T_i^\dagger > t_s)
\end{aligned}
$$

The last equality holds because $E(\mathbf{X}_i Y_{\tau i}|T_i^\dagger > t_s) = E(\mathbf{X}_i Y_{\tau i}|T_i > t_s)$ due to the independence between C_i and $(\mathbf{X}_i, T_i^\dagger)$. Both terms in the above quantity are 0 because $\mathbf{Q}_0(\bar{\boldsymbol{\theta}}_\tau) = \mathbf{D}_0(\bar{\boldsymbol{\gamma}}) = 0$ and $\mathbf{X}$ is part of $\boldsymbol{\Phi}(\mathbf{X})$ and $\boldsymbol{\Phi}(\mathbf{Z})$. Therefore, $\sup_{\boldsymbol{\beta}} |\widehat{\mathbf{U}}_n(\boldsymbol{\beta}) - \mathbf{U}_0(\boldsymbol{\beta})| \to 0$ in probability and hence $\widehat{\boldsymbol{\beta}} \to \bar{\boldsymbol{\beta}}$ in probability.

B Asymptotic Distribution of $\widehat{\beta}$ Obtained with S measured at t_s

Uno et al. (2007) has shown that:

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}) \approx n^{-\frac{1}{2}} \sum_{i=1}^n \mathbb{J}^{-1}\left[w_{\tau i}\mathbf{U}_i^X + \int_0^\tau \psi_i(s) dE\{\mathbf{U}_i^X I(T_i^\dagger \wedge \tau \le s)\}\right]$$

where $\psi_i(s) = \int_0^s \frac{dM_{ci}(u)}{\pi(u)}$, $\pi(u) = \Pr(T_i > u)$, $M_{ci}(u) = I(T_i \le u, \delta_i = 0) - \int_0^u I(T_i > v)d\Lambda_c(v)$ and $\Lambda_c(\cdot)$ is the cumulative hazard function of C. It can be further shown that

$$
\begin{aligned}
\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}) &\approx n^{-\frac{1}{2}} \sum_{i=1}^n \mathbb{J}^{-1}\left[w_{\tau i}\mathbf{U}_i^X + \int_0^\tau \psi_i(s) dE\{\mathbf{U}_i^X I(T_i^\dagger \le s)\} + \psi(\tau)E\{\mathbf{U}_i^X I(T_i^\dagger > \tau)\}\right] \\
&\approx n^{-\frac{1}{2}} \sum_{i=1}^n \mathbb{J}^{-1}\left[w_{\tau i}\mathbf{U}_i^X - \int_0^\tau E\{\mathbf{U}_i^X I(T_i^\dagger \le s) d\psi_i(s)\}\right] \\
&\approx n^{-\frac{1}{2}} \sum_{i=1}^n \left[w_{\tau i}\mathbf{F}_{1i} + \int_0^\tau E\{\mathbf{F}_{1i} I(T_i^\dagger > s)\}\frac{dM_{ci}(s)}{\pi(s)}\right].
\end{aligned}
$$

Similarly, it can be shown that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{t_s} - \bar{\boldsymbol{\theta}}_{t_s}) \approx n^{-\frac{1}{2}} \sum_{i=1}^n \mathbb{A}^{-1}\left[w_{t_s i}\mathbf{U}_i^\Phi + \int_0^{t_s} E\{\mathbf{U}_i^\Phi I(T_i^\dagger > s)\}\frac{dM_{ci}(s)}{\pi(s)}\right], \quad \text{and} \tag{B.1}$$

$$\sqrt{n}(\widehat{\boldsymbol{\gamma}}_{\tau|t_s} - \bar{\boldsymbol{\gamma}}_{\tau|t_s}) \approx n^{-\frac{1}{2}} \sum_{i=1}^n \mathbb{B}^{-1}\left[I(T_i^\dagger > t_s)w_{\tau i}\mathbf{U}_i^\Psi + \int_{t_s}^\tau E\{\mathbf{U}_i^\Psi I(T_i^\dagger > s)\}\frac{dM_{ci}(s)}{\pi(s)}\right]. \tag{B.2}$$

Here, (B.2) is true because $I(T_i > t_s)w_{\tau i} = I(T_i^\dagger > t_s)w_{\tau i}$.

Now, from a Taylor series expansion of $\widehat{\mathbf{U}}_n(\cdot)$ and a law of large numbers,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}) \approx \mathbb{J}^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathbf{X}_i \Bigg[\{g(\widehat{\boldsymbol{\theta}}_{t_s}^{\mathsf{T}} \boldsymbol{\Phi}_i) - g(\bar{\boldsymbol{\theta}}_{t_s}^{\mathsf{T}} \boldsymbol{\Phi}_i)\} + \widehat{\varpi}_{t_s i}\{g(\widehat{\boldsymbol{\gamma}}_{\tau|t_s}^{\mathsf{T}} \boldsymbol{\Psi}_i) - g(\bar{\boldsymbol{\gamma}}_{\tau|t_s}^{\mathsf{T}} \boldsymbol{\Psi}_i)\}$$

$$+ (\widehat{\varpi}_{t_s i} - \varpi_{t_s i}) g(\bar{\boldsymbol{\gamma}}_{\tau|t_s}^{\mathsf{T}} \boldsymbol{\Psi}_i) + Y_{\tau i}^{t_s} - g(\bar{\boldsymbol{\beta}}^{\mathsf{T}} \mathbf{X}_i) \Bigg]$$

$$\approx \mathbb{J}^{-1} n^{-1} \sum_{i=1}^{n} \mathbf{X}_i \boldsymbol{\Phi}_i^{\mathsf{T}} \dot{g}(\bar{\boldsymbol{\theta}}_{t_s}^{\mathsf{T}} \boldsymbol{\Phi}_i) \sqrt{n}(\widehat{\boldsymbol{\theta}}_{t_s} - \bar{\boldsymbol{\theta}}_{t_s}) + \mathbb{J}^{-1} n^{-1} \sum_{i=1}^{n} \varpi_{t_s i} \mathbf{X}_i \boldsymbol{\Psi}_i^{\mathsf{T}} \dot{g}(\bar{\boldsymbol{\gamma}}_{\tau|t_s}^{\mathsf{T}} \boldsymbol{\Psi}_i) \sqrt{n}(\widehat{\boldsymbol{\gamma}}_{\tau|t_s} - \bar{\boldsymbol{\gamma}}_{\tau|t_s})$$

$$+ \mathbb{J}^{-1} n^{-1} \sum_{i=1}^{n} \varpi_{t_s i} \mathbf{X}_i g(\bar{\boldsymbol{\gamma}}_{\tau|t_s}^{\mathsf{T}} \boldsymbol{\Psi}_i) n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi_i(t_s) + \mathbb{J}^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathbf{X}_i \{Y_{\tau i}^{t_s} - g(\bar{\boldsymbol{\beta}}^{\mathsf{T}} \mathbf{X}_i)\}$$

$$\approx \mathbb{J}^{-1} \mathbb{K} \sqrt{n}(\widehat{\boldsymbol{\theta}}_{t_s} - \bar{\boldsymbol{\theta}}_{t_s}) + \mathbb{J}^{-1} \mathbb{L} \sqrt{n}(\widehat{\boldsymbol{\gamma}}_{\tau|t_s} - \bar{\boldsymbol{\gamma}}_{\tau|t_s}) + n^{-\frac{1}{2}} \sum_{i=1}^{n} \Big[E(\mathbf{L}_i)\psi_i(t_s) + \mathbb{J}^{-1} \mathbf{X}_i \{Y_{\tau i}^{t_s} - g(\bar{\boldsymbol{\beta}}^{\mathsf{T}} \mathbf{X}_i)\} \Big],$$

where $\mathbb{K} = E\{\mathbf{X}_i \boldsymbol{\Phi}_i^{\mathsf{T}} \dot{g}(\bar{\boldsymbol{\theta}}_{t_s}^{\mathsf{T}} \boldsymbol{\Phi}_i)\}$, and $\mathbb{L} = E\{I(T_i^\dagger > t_s) \mathbf{X}_i \boldsymbol{\Phi}_i^{\mathsf{T}} \dot{g}(\bar{\boldsymbol{\gamma}}_{\tau|t_s}^{\mathsf{T}} \boldsymbol{\Psi}_i)]$. Following from the expansions of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\gamma}}$ given in (B.1) and (B.2), we have

$$\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}) \approx\ & n^{-\frac{1}{2}} \sum_{i=1}^{n} w_{t_s i} \mathbb{J}^{-1} \mathbb{K} \mathbb{A}^{-1} \mathbf{U}_i^{\Phi} + n^{-\frac{1}{2}} \sum_{i=1}^{n} I(T_i > t_s) w_{\tau i} \mathbb{J}^{-1} \mathbb{L} \mathbb{B}^{-1} \mathbf{U}_i^{\Psi} \\
& + n^{-\frac{1}{2}} \sum_{i=1}^{n} [E(\mathbf{L}_i)\psi_i(t_s) + \mathbb{J}^{-1} \mathbf{X}_i \{Y_{\tau i}^{t_s} - g(\bar{\boldsymbol{\beta}}^{\mathsf{T}} \mathbf{X}_i)\}] \\
& + n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{t_s} E\{\mathbb{J}^{-1} \mathbb{K} \mathbb{A}^{-1} \mathbf{U}_i^{\Phi} I(T_i^\dagger > s)\} \frac{dM_{ci}(s)}{\pi(s)} \\
& + n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_{t_s}^{\tau} E\{\mathbb{J}^{-1} \mathbb{L} \mathbb{B}^{-1} \mathbf{U}_i^{\Psi} I(T_i^\dagger > s)\} \frac{dM_{ci}(s)}{\pi(s)}]
\end{aligned}$$

To further simplify, we note that for $j = 1, ..., p$,

$$[\mathbb{K}\mathbb{A}^{-1}]_{j,} = E[X_j \boldsymbol{\Phi}^{\mathsf{T}} \dot{g}(\bar{\boldsymbol{\theta}}_{t_s}^{\mathsf{T}} \boldsymbol{\Phi})] E[\boldsymbol{\Phi}^{\otimes 2} \dot{g}(\bar{\boldsymbol{\theta}}_{t_s}^{\mathsf{T}} \boldsymbol{\Phi})]^{-1} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\, E[\dot{g}(\bar{\boldsymbol{\theta}}_{t_s}^{\mathsf{T}} \boldsymbol{\Phi}) \{X_j - \boldsymbol{\alpha}^{\mathsf{T}} \boldsymbol{\Phi}\}^2].$$

Since each X_j is an element of $\boldsymbol{\Phi}(\mathbf{X})$, $\min_{\boldsymbol{\alpha}} E[\dot{g}(\bar{\boldsymbol{\theta}}_{t_s}^{\mathsf{T}} \boldsymbol{\Phi}(\mathbf{X}))\{X_j - \boldsymbol{\alpha}^{\mathsf{T}} \boldsymbol{\Phi}(\mathbf{X})\}^2] = 0$. Thus we have

$$\mathbf{X}_i = \mathbb{K}\mathbb{A}^{-1} \boldsymbol{\Phi}_i, \qquad \text{and} \qquad \mathbb{J}^{-1} \mathbb{K} \mathbb{A}^{-1} \mathbf{U}_i^{\Phi} = \mathbf{F}_{2i}.$$

Similarly,

$$[\mathbb{LB}^{-1}]_{j,} = E[I(T^\dagger > t_s)X_j\boldsymbol{\Psi}^\intercal\dot{g}(\bar{\boldsymbol{\gamma}}_{\tau|t_s}^\intercal\boldsymbol{\Psi})]E[I(T^\dagger > t_s)\boldsymbol{\Psi}^{\otimes 2}\dot{g}(\bar{\boldsymbol{\gamma}}_{\tau|t_s}^\intercal\boldsymbol{\Psi})]^{-1}$$

$$= \operatorname*{argmin}_{\boldsymbol{\alpha}} E[\dot{g}(\bar{\boldsymbol{\theta}}_{t_s}^\intercal\boldsymbol{\Psi}(\mathbf{X}))\{X_j - \boldsymbol{\alpha}^\intercal\boldsymbol{\Psi}(\mathbf{X})\}^2], \qquad \text{for } j = 1, ..., p.$$

$$\mathbf{X}_i = \mathbb{LB}^{-1}\boldsymbol{\Psi}_i, \qquad \text{and} \qquad \mathbb{J}^{-1}\mathbb{LB}^{-1}\mathbf{U}_i^{\Psi} = \mathbf{F}_{3i}.$$

The above equation can be written as:

$$\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}) \approx{}& n^{-\frac{1}{2}}\sum_{i=1}^{n}[w_{t_s i}\mathbf{F}_{2i} + I(T_i^\dagger > t_s)w_{\tau i}\mathbf{F}_{3i} + E(\mathbf{L}_i)\psi_i(t_s) + \mathbb{J}^{-1}\mathbf{X}_i\{Y_{\tau i}^{t_s} - g(\bar{\boldsymbol{\beta}}^\intercal\mathbf{X}_i)\} \\
&+ E\{\mathbf{F}_{2i}I(T_i^\dagger > s)\}\frac{dM_{ci}(s)}{\pi(s)} + E\{\mathbf{F}_{3i}I(T_i^\dagger > s)\}\frac{dM_{ci}(s)}{\pi(s)}] \\
\approx{}& n^{-\frac{1}{2}}\sum_{i=1}^{n}[\mathbf{F}_{1i} + \int_0^{t_s} E(\mathbf{L}_i)\frac{dM_{ci}(s)}{\pi(s)} + \{\frac{I(C_i > t_s)}{G(t_s)} - 1\}\mathbf{L}_i \\
&+ (w_{t_s i} - 1)\mathbf{F}_{2i} + \int_0^{t_s} E[\mathbf{F}_{2i}I(T_i^\dagger > s)]\frac{dM_{ci}(s)}{\pi(s)} \\
&+ (w_{\tau i} - 1)I(T_i > t_s)\mathbf{F}_{3i} + \int_{t_s}^{\tau} E\{\mathbf{F}_{3i}I(T_i^\dagger > s)\}\frac{dM_{ci}(s)}{\pi(s)}]
\end{aligned}$$

Note that

$$\begin{aligned}
(w_{t_s i} - 1)\mathbf{F}_{2i} &= \left\{\frac{I(T_i \le t_s)\delta_i}{G(T_i)} + \frac{I(T_i > t_s)}{G(t_s)} - 1\right\}\mathbf{F}_{2i} \\
&= \left\{\frac{I(T_i \le t_s)}{G(T_i)} - \frac{I(T_i \le t_s)(1-\delta_i)}{G(T_i)} + \frac{1}{G(t_s)} - \frac{I(T_i \le t_s)}{G(t_s)} - 1\right\}\mathbf{F}_{2i} \\
&= [-\int_0^{t_s}\frac{dI(T_i \le s)(1-\delta_i)}{G(s)} + \frac{1}{G(t_s)} - 1 + \{\frac{1}{G(T_i)} - \frac{1}{G(t_s)}\}I(T_i \le t_s)]\mathbf{F}_{1i} \\
&= [-\int_0^{t_s}\frac{dI(T_i \le s)(1-\delta_i)}{G(s)} + \int_0^{t_s} d\frac{1}{G(s)} - \int_0^{t_s} I(T_i \le s)d\frac{1}{G(s)}]\mathbf{F}_{1i} \\
&= \left\{-\int_0^{t_s}\frac{dI(T_i \le s)(1-\delta_i)}{G(s)} + \int_0^{t_s} I(T_i > s)d\frac{1}{G(s)}\right\}\mathbf{F}_{2i} \\
&= -\int_0^{t_s}\mathbf{F}_{2i}\frac{S(s)dM_{ci}(s)}{\pi(s)}
\end{aligned}$$

and similarly, $(w_{\tau i} - 1)I(T_i^\dagger > t_s)\mathbf{F}_{3i} = -\int_{t_s}^{\tau} I(T_i^\dagger > t_s)\mathbf{F}_{3i}\frac{dM_{ci}(s)}{\pi(s)}$. Together with the fact that

$$\left\{\frac{I(C_i > t_s)}{G(t_s)} - 1\right\}\mathbf{L}_i = (w_{t_s i} - 1)\mathbf{L}_i = -\int_0^{t_s}\mathbf{L}_i\frac{S(s)dM_{ci}(s)}{\pi(s)}$$

we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}) \approx n^{-\frac{1}{2}} \sum_{i=1}^{n} \left(\mathbf{F}_{1i} + \int_0^{t_s} \Big[E(\mathbf{L}_i) - \{\mathbf{L}_i + \mathbf{F}_{2i} - E(\mathbf{F}_{2i}|T_i^\dagger > s)\} S(s) \Big] \frac{dM_{ci}(s)}{\pi(s)} \right.$$
$$\left. - \int_{t_s}^{\tau} \{\mathbf{F}_{3i} - E(\mathbf{F}_{3i}|T_i^\dagger > s\} S(s) \frac{dM_{ci}(s)}{\pi(s)} \right)$$
$$= n^{-\frac{1}{2}} \sum_{i=1}^{n} \left(\mathbf{F}_{1i} - \int_0^{t_s} \{\mathbf{F}_{2i} + \mathbf{L}_i - E(\mathbf{F}_{2i} + \mathbf{L}_i|T_i^\dagger > s)\} S(s) \frac{dM_{ci}(s)}{\pi(s)} \right.$$
$$\left. - \int_{t_s}^{\tau} \{\mathbf{F}_{3i} - E(\mathbf{F}_{3i}|T_i^\dagger > s)\} S(s) \frac{dM_{ci}(s)}{\pi(s)} \right).$$

By the Central Limit Theory, we have $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}) \to N(0, \boldsymbol{\Sigma}_{t_s})$ where

$$\boldsymbol{\Sigma}_{t_s} = \mathrm{var}(\mathbf{F}_{1i}) + \int_0^{t_s} \mathrm{var}(\mathbf{F}_{2i} + \mathbf{L}_i|T_i^\dagger > s) \frac{S(s)^2 d\Lambda_c(s)}{\pi(s)} + \int_{t_s}^{\tau} \mathrm{var}(\mathbf{F}_{3i}|T_i^\dagger > s) \frac{S(s)^2 d\Lambda_c(s)}{\pi(s)}.$$

References

Newey, W.K. and McFadden, D. (1994). *The Handbook of Econometrics. Vol. 4: Estimation in large samples*. D. McFadden and R. Engle.

Pollard, David. (1990). NSF-CBMS regional Conference series in probability and statistics: Empirical Processes: theory and applications. *JSTOR 94*, pp. i–86.

Tian, L. and Cai, T. and Geotghebeur, E. and Wei, L.J. (2007). Model evaluation based on the distribution of estimated absolute prediction error. *Biometrika 94*, pp. 297–311.

Uno, H. and Cai, T. and Tian, L. and Wei, L.J. (2007). Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association 102*, pp. 527–537.

Part 2: Simulation Code in R

The following code is for low dimension setting. For the setting with moderate number of covariates, the code is very similar except for changing the code for generating T_i, S_i, Z_i based on the formula in Section 3.2.

The code for the functions Est.Surv.C.FUN and PTB.Surv.C.FUN can be requested for downloading from the author (ezheng@sdac.harvard.edu).

############### Begin Code ###################

library(glmpath);library(glmnet);library(survival) # The following code is used to simulate data, generate estimators and resampling for inference, and find the optimal linear combination of estimators; # The code for the analysis specific functions is provided following the program;

t0=0.8;#this is the τ

allts=c(0.05,0.1,0.15,0.2); #these are the t_ks

n.beta=4; #length of $\boldsymbol{\beta}$

nts=length(allts); # number of time points for intermediate outcome

nn=500 ; # sample size

Tc=12; Ct=0.5; # Rare events/heavy censoring setting

Tc=6; Ct=1; # Moderate events/censoring setting

set.seed(seed) # setting seed from 1 to 500 for the 500 runs

The following code is to simulate datasets.

Zi=mvrnorm(nn,rep(0,3),matrix(c(1,0.3,0.3,0.3,1,0.3,0.3,0.3,1),3,3))

ginv.logit = function(pp)log(pp/(1-pp))

uni=ginv.logit(runif(nn))

$Ti = exp(Zi\% * \%c(0.5, 0.5, 0.5) + 0.5 * Zi[, 1]^2 + Zi[, 2]^2 + 0.5 * Zi[, 3]^2 - 3 + uni) * Tc$

Si=matrix(,ncol=length(allts),nrow=length(Ti))

for (i in 1:length(allts)){

    tsi=allts[i] # let tsi=allts[4] for constant correlation

    $Si[, i] = uni + 0.1 * (Zi[, 1] + Zi[, 2]) + rnorm(nn, mean = 0, sd = 1)/(10 * tsi\hat{}1.5);$

    }

Xi = pmin(Ti,Ci); Deltai = 1*(Ti ¡= Ci);

mydata=cbind(Ti,Xi,Deltai,Si,Zi)

Pdelta=mean((1-mydata[,3])*(mydata[,2]¡t0)) # censor rate;

Pevent=mean((mydata[,2]¡t0)*mydata[,3]) # event rate

PScensor=CorSlogT=NULL

for (aa in 1:nts){

    ts0=allts[aa]

    Si=mydata[,3+aa]

    PScensor=c(PScensor,mean(mydata[,2]¡ts0))

    CorSlogT=c(CorSlogT,cor(Si,log(mydata[,1])))

####### censor intermediate covariate;

```
mydata[mydata[,2]¡allts[aa],3+aa]=NA
}
```

The following code is to generate different estimators for $\boldsymbol{\beta}$.

beta.all includes the estimator from non-censored data, IPW_{KM}, $\text{IPW}_{Cox,\mathbf{X}}$, AIPW_{KM},

$\text{AUG}^{t_s}_{KM,\mathbf{X}}$, and $\text{AUG}^{t_s}_{KM,\mathbf{Z}}$ in the manuscript.

```
Modelfit = Est.Surv.C.FUN(mydata,lambda=NULL,nknot=3,tsi=allts,
AIPW=1,CoxT=F,NSselection=”ridge”,selection=”none”))
beta.all=c(Modelfit$beta);
```

The following code is to do resampling for the estimators;

```
PTB.Surv=PTB.Surv.C.FUN(mydata,n.ptb=500,nknot=3,tsi=allts,NSselection=”ridge”,
AIPW=1,ipwcox=1,CoxT=F, selection=”none”)
PTBipw=PTB.Surv$betaIPW; PTBipwcox.x=PTB.Surv$bhatIPWcox.x
PTBAIPWkm.x=PTB.Surv$bhatAIPWkm.x; PTBts=PTB.Surv$beta.ts
se.betaIPW=apply(PTBipw,2,sd); se.betaIPWcox.x=apply(PTBipwcox.x,2,sd)
se.betaAIPWkm.x=apply(PTBAIPWkm.x,2,sd); se.beta.ts=apply(PTBts,2,sd)
```

The following code used to generate the optimal linear combination of the estimators;

comb_a is $\text{AUG}^{KM,\mathbf{Z}}_{CMB}$ in the manuscript;

comb_x is $\text{AUG}^{KM,\mathbf{X}}_{CMB}$ in the manuscript;

```
comb_a=comb_sd_a=comb_x=comb_sd_x=comb_xs=comb_sd_xs=NULL
inv.beta=Modelfit$beta[-c(1,3,4),]
```

```
for (i in 1:n.beta){

    Y=PTBipw[,i]

    PTBtsi=PTBts[,seq(i,ncol(PTBts),by=n.beta)]

    XS=matrix(Y,ncol=ncol(PTBtsi),nrow=length(Y))-PTBtsi

    X.x=XS[,seq(1,ncol(XS),by=2)]

    newXS=inv.beta[,i]

    newx=newXS[c(1,seq(2,length(newXS),by=2))]

    lassoModel.XS=try(cv.glmnet(XS,y=Y,alpha=1))

    lassoModel.x=try(cv.glmnet(X.x,y=Y,alpha=1))

    beta_a=coef(lassoModel.XS,s="lambda.min")

    comb_a=c(comb_a,t(newXS[-1])%*%beta_a[-1]+newXS[1]*(1-sum(beta_a[-1])))

    comb_sd_a=c(comb_sd_a,sd(Y-XS%*%beta_a[-1]))

    beta_x=coef(lassoModel.x,s="lambda.min")

    comb_x=c(comb_x,t(newx[-1])%*%beta_x[-1]+newx[1]*(1-sum(beta_x[-1])))

    comb_sd_x=c(comb_sd_x,sd(Y-X.x%*%beta_x[-1]))

}

############### End Code ###################
```