Supplementary Materials of "Causal Proportional Hazards Estimation with a Binary Instrumental Variable"

Behzad Kianian[1], Jung In Kim[2], Jason P. Fine [1], and Limin Peng[1]

[1]*Department of Biostatistics and Bioinformatics, Emory University*

[2]*Departments of Biostatistics, University of North Carolina at Chapel Hill*

S1. Justifications for κ_v

Recall $\boldsymbol{U} = (W, \delta, D, \boldsymbol{X})$. It is easy to see that

$$E(D(1-V)|\boldsymbol{U}) = P(D(1-V)=1|\boldsymbol{U}) = P(D_1 = D_0 = 1|\boldsymbol{U})P(V=0|D_1 = D_0 = 1, \boldsymbol{U})$$

$$= P(D_1 = D_0 = 1|\boldsymbol{U})P(V=0|D_1 = D_0 = 1, W_1 = min(T_1, C), \delta_1 = I(T_1 \le C), \boldsymbol{X})$$

$$= P(D_1 = D_0 = 1|\boldsymbol{U})P(V=0|\boldsymbol{X}).$$

The last equality uses the aforementioned assumption that censoring is independent of the instrumental variable V conditional on $\boldsymbol{X}$ and the assumption of joint independence of (D_1, D_0, T_1, T_0) and V conditional on $\boldsymbol{X}$.

Similarly, $E((1-D)V|\boldsymbol{U}) = P(D_1 = D_0 = 0|U)P(V=1|\boldsymbol{X})$. It then follows that

$$\kappa_v \quad = \quad E\left\{1 - \frac{D(1-V)}{P(V=0|\boldsymbol{X})} - \frac{(1-D)V}{P(V=1|\boldsymbol{X})}|\boldsymbol{U}\right\}$$

$$= 1 - P(D_1 = D_0 = 1|\boldsymbol{U}) - P(D_1 = D_0 = 0|\boldsymbol{U}) = P(D_1 > D_0|\boldsymbol{U}).$$

The result above indicates that κ_v is always nonnegative. The justification for using the projected weight $E(\kappa|\boldsymbol{U})$ follows from the arguments in Abadie et al. (2002)

S2. Regularity conditions and proofs of Theorems 1–2

We first introduce some new notation. Define

$$\bar{\boldsymbol{U}}_{n,\kappa}(\boldsymbol{\beta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\infty \kappa_i \left[\mathbf{Z}_i - \left\{ \frac{\mathbf{s}_\mathbf{c}^{(\mathbf{1})}(\boldsymbol{\beta}, \mathbf{s})}{s_c^{(0)}(\boldsymbol{\beta}, s)} \right\} \right] dM_i(s),$$

$\boldsymbol{E}_{n,\kappa}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{S}^{(1)}_{n,\kappa}(\boldsymbol{\beta},t)}{S^{(0)}_{n,\kappa}(\boldsymbol{\beta},t)}$, $\boldsymbol{E}_{n,\hat{\kappa}}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{S}^{(1)}_{n,\hat{\kappa}}(\boldsymbol{\beta},t)}{S^{(0)}_{n,\hat{\kappa}}(\boldsymbol{\beta},t)}$, $\boldsymbol{e}_c(\boldsymbol{\beta}, t) = \frac{\boldsymbol{s}^{(1)}_{c}(\boldsymbol{\beta},t)}{s^{(0)}_{c}(\boldsymbol{\beta},t)}$, $\boldsymbol{V}_{n,\kappa}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{S}^{(2)}_{n,\kappa}(\boldsymbol{\beta},t)}{S^{(0)}_{n,\kappa}(\boldsymbol{\beta},t)} - \boldsymbol{E}_{n,\kappa}(\boldsymbol{\beta}, t)^{\otimes 2}$, $\boldsymbol{V}_{n,\hat{\kappa}}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{S}^{(2)}_{n,\hat{\kappa}}(\boldsymbol{\beta},t)}{S^{(0)}_{n,\hat{\kappa}}(\boldsymbol{\beta},t)} - \boldsymbol{E}_{n,\hat{\kappa}}(\boldsymbol{\beta}, t)^{\otimes 2}$, and $\boldsymbol{v}_c(\boldsymbol{\beta}, t) = \frac{\boldsymbol{s}^{(2)}_{c}(\boldsymbol{\beta},t)}{s^{(0)}_{c}(\boldsymbol{\beta},t)} - \boldsymbol{e}_c(\boldsymbol{\beta}, t)^{\otimes 2}$. Let

$$\boldsymbol{\Sigma}_0 = \int_0^\infty \boldsymbol{v}_c(\boldsymbol{\beta}_0, t) s_c^{(0)}(\boldsymbol{\beta}_0, t) h_0(t) dt. \tag{B.1}$$

We let $\|\cdot\|$ denote Euclidean norm.

Proof of Theorem 1: Define $\phi(\boldsymbol{\alpha}, \boldsymbol{O}) \equiv 1 - \frac{D(1-V)}{1-\psi(\boldsymbol{\alpha},\boldsymbol{X})} - \frac{(1-D)V}{\psi(\boldsymbol{\alpha},\boldsymbol{X})}$, and $\boldsymbol{Q}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) = n^{-1/2}\boldsymbol{U}_{n,\phi(\boldsymbol{\alpha},\boldsymbol{O})}(\boldsymbol{\beta})$. Then $\boldsymbol{Q}_n(\boldsymbol{\alpha}_0, \boldsymbol{\beta}) = n^{-1/2}\boldsymbol{U}_{n,\kappa}(\boldsymbol{\beta})$ and $\boldsymbol{Q}_n(\hat{\boldsymbol{\alpha}}, \boldsymbol{\beta}) = n^{-1/2}\boldsymbol{U}_{n,\hat{\kappa}}(\boldsymbol{\beta})$. Under conditions (C1)-(C3), we have $\|\partial \boldsymbol{Q}_n(\boldsymbol{\alpha}, \boldsymbol{\beta})/\partial \boldsymbol{\alpha}\|$ is bounded in a neighborhood of $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$. Given $\hat{\boldsymbol{\alpha}}$ is a consistent estimator of $\boldsymbol{\alpha}_0$ (i.e condition (C5)), applying Taylor expansion to $\boldsymbol{Q}_n(\hat{\boldsymbol{\alpha}}, \boldsymbol{\beta})$ around $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ implies that

$$\sup_{\boldsymbol{\beta} \in \boldsymbol{\mathcal{B}}} \|n^{-1/2}\{\boldsymbol{U}_{n,\hat{\kappa}}(\boldsymbol{\beta}) - \boldsymbol{U}_{n,\kappa}(\boldsymbol{\beta})\}\| \to_{a.s.} 0. \tag{B.2}$$

By the Glivenko-Cantelli Theorem (van der Vaart and Wellner, 1996), we can show under conditions (C1)-(C2) that

$$\sup_{\boldsymbol{\beta}\in\boldsymbol{\mathcal{B}},t} \|n^{-1}S_{n,\kappa}^{(j)}(\boldsymbol{\beta},t) - s_c^{(j)}(\boldsymbol{\beta},t)\| \rightarrow_{a.s.} 0, \quad j=0,1.$$

Given condition (C3), this implies $\sup_{\boldsymbol{\beta}\in\boldsymbol{\mathcal{B}},t} \|E_{n,\kappa}(\boldsymbol{\beta},t) - e_c(\boldsymbol{\beta},t)\| \rightarrow_{a.s.} 0$. Then,

$$\begin{aligned}\sup_{\boldsymbol{\beta}\in\boldsymbol{\mathcal{B}}} \|\boldsymbol{U}_{n,\kappa}(\boldsymbol{\beta}) - \bar{\boldsymbol{U}}_{n,\kappa}(\boldsymbol{\beta})\| &\leq n^{-1/2}\sum_{i=1}^{n}\int_0^\infty \|E_{n,\kappa}(\boldsymbol{\beta},s) - e_c(\boldsymbol{\beta},s)\| dM_i(s) \\ &\leq \sup_{\boldsymbol{\beta}\in\boldsymbol{\mathcal{B}},t} \|E_{n,\kappa}(\boldsymbol{\beta},t) - e_c(\boldsymbol{\beta},t)\| \cdot \{n^{-1/2}\sum_{i=1}^{n}\int_0^\infty dM_i(s)\} = o(1), \ a.s. \end{aligned} \tag{B.3}$$

By the results in Abadie (2003) and an application of the Glivenko-Cantelli Theorem (van der Vaart and Wellner, 1996), we get

$$\sup_{\boldsymbol{\beta}\in\boldsymbol{\mathcal{B}}} \|\bar{\boldsymbol{U}}_{n,\kappa}(\boldsymbol{\beta}) - \boldsymbol{\mu}_c(\boldsymbol{\beta})\| = o(1), \ a.s. \tag{B.4}$$

It follows from (B.2),(B.3), and (B.4) that

$$\sup_{\boldsymbol{\beta}\in\boldsymbol{\mathcal{B}}} \|n^{-1/2}\boldsymbol{U}_{n,\hat{\kappa}}(\boldsymbol{\beta}) - \boldsymbol{\mu}_c(\boldsymbol{\beta})\| = o(1), \ a.s. \tag{B.5}$$

By condition (C4), $\boldsymbol{\mu}_c(\boldsymbol{\beta})$ is a concave function with a unique maximizer $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Suppose $\hat{\boldsymbol{\beta}}$ does not converge to $\boldsymbol{\beta}_0$, a.s. Then $P(\mathcal{E}) > 0$, where $\mathcal{E}$=$\{\exists$ a subsequence n_k such that $\hat{\boldsymbol{\beta}}_{n_k} \rightarrow \boldsymbol{\beta}^* \neq \boldsymbol{\beta}_0\}$. By the definition of $\hat{\boldsymbol{\beta}}$, we have $n^{-1/2}\boldsymbol{U}_{n,\hat{\kappa}}(\hat{\boldsymbol{\beta}}_{n_k}) \geq n^{-1/2}\boldsymbol{U}_{n,\hat{\kappa}}(\boldsymbol{\beta}_0)$ in $\mathcal{E}$, which implies $\boldsymbol{\mu}_c(\boldsymbol{\beta}^*) \geq \boldsymbol{\mu}_c(\boldsymbol{\beta}_0)$ given (B.5). This contradicts the fact that $\boldsymbol{\beta}_0$ is the unique maximizer of $\boldsymbol{\mu}_c(\boldsymbol{\beta})$. Therefore, we have $\hat{\boldsymbol{\beta}} \rightarrow_{a.s.} \boldsymbol{\beta}_0$.

Proof of Theorem 2: Define $\boldsymbol{A}_i(\boldsymbol{\beta}) = \int_0^\infty \kappa_i\{\boldsymbol{Z}_i - E_{n,\kappa}(\boldsymbol{\beta}, s)\}dN_i(s)$, $\hat{\boldsymbol{A}}_i(\boldsymbol{\beta}) = \int_0^\infty \hat{\kappa}_i\{\boldsymbol{Z}_i - E_{n,\hat{\kappa}}(\boldsymbol{\beta}, s)\}dN_i(s)$. Then

$$0 = \boldsymbol{U}_{n,\hat{\kappa}}(\hat{\boldsymbol{\beta}}) = \boldsymbol{U}_{n,\kappa}(\hat{\boldsymbol{\beta}}) + n^{-1/2}\sum_{i=1}^{n}\{\hat{\boldsymbol{A}}_i(\hat{\boldsymbol{\beta}}) - \boldsymbol{A}_i(\hat{\boldsymbol{\beta}})\}. \tag{B.6}$$

Given the consistency of $\hat{\boldsymbol{\beta}}$, the Taylor expansion of $\boldsymbol{U}_{n,\kappa}(\boldsymbol{\beta})$ around $\boldsymbol{\beta} = \boldsymbol{\beta}_{\mathbf{0}}$ gives

$$\boldsymbol{U}_{n,\kappa}(\hat{\boldsymbol{\beta}}) \approx \boldsymbol{U}_{n,\kappa}(\boldsymbol{\beta}_{\mathbf{0}}) - \boldsymbol{\varphi}_n(\boldsymbol{\beta}_{\mathbf{0}})\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathbf{0}}) + o(1), \tag{B.7}$$

where $\boldsymbol{\varphi}_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\infty \kappa_i \boldsymbol{V}_{n,\kappa}(\boldsymbol{\beta}, s)dN_i(s)$, and $\approx$ means the difference is $o(1)$, *a.s.*

On the other hand, we can write

$$\begin{aligned} n^{-1/2}\sum_{i=1}^{n}\{\hat{\boldsymbol{A}}_i(\boldsymbol{\beta}) - \boldsymbol{A}_i(\boldsymbol{\beta})\} &= n^{-1/2}\sum_{i=1}^{n}(\hat{\kappa}_i - \kappa_i)\{\boldsymbol{Z}_i - \boldsymbol{E}_{n,\kappa}(\boldsymbol{\beta}, s)\}dM_i(s) \\ &\quad - n^{-1/2}\sum_{i=1}^{n}\hat{\kappa}_i\{\boldsymbol{E}_{n,\hat{\kappa}}(\boldsymbol{\beta}, s) - \boldsymbol{E}_{n,\kappa}(\boldsymbol{\beta}, s)\}dM_i(s). \end{aligned}$$

Define $\boldsymbol{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{O}) = \partial\phi(\boldsymbol{\alpha}, \boldsymbol{O})/\partial\boldsymbol{\alpha}^T$. Note that $\boldsymbol{D}_\phi(\boldsymbol{\alpha}_0, \boldsymbol{O})$ is bounded under conditions (C2). This implies $\sup_i |\hat{\kappa}_i - \kappa_i| = o(1)$, *a.s.* given condition (C5). By the Taylor expansion of $\phi(\boldsymbol{\alpha}, \boldsymbol{O}_i)$ around $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ and condition (C6), we have

$$\begin{aligned} n^{1/2}(\hat{\kappa}_i - \kappa_i) &= \phi(\hat{\boldsymbol{\alpha}}, \boldsymbol{O}_i) - \phi(\boldsymbol{\alpha}_0, \boldsymbol{O}_i) \approx \boldsymbol{D}_\phi(\boldsymbol{\alpha}_0, \boldsymbol{O}_i)n^{1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \\ &= n^{-1/2}\sum_{j=1}^{n}\boldsymbol{D}_\phi(\boldsymbol{\alpha}_0, \boldsymbol{O}_i)\boldsymbol{I}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_0, \boldsymbol{O}_j) \equiv n^{-1/2}\sum_{j=1}^{n}\boldsymbol{I}_\kappa(\boldsymbol{\alpha}_0, O_j, O_i) \end{aligned} \tag{B.8}$$

Given these results, we can further approximate $n^{-1/2}\sum_{i=1}^{n}\{\hat{\boldsymbol{A}}_i(\boldsymbol{\beta}) - \boldsymbol{A}_i(\boldsymbol{\beta})\}$ as follows.

First, using the fact that $\sup_i |\hat{\kappa}_i - \kappa_i| = o(1)$, *a.s.* and applying the Glivenko-Cantelli Theorem to $\boldsymbol{E}_{n,\kappa}$ and $\boldsymbol{E}_{n,\hat{\kappa}}$, we get

$$n^{-1/2}\sum_{i=1}^{n}\{\hat{\boldsymbol{A}}_i(\boldsymbol{\beta}) - \boldsymbol{A}_i(\boldsymbol{\beta})\} \approx n^{-1/2}\sum_{i=1}^{n}(\hat{\kappa}_i - \kappa_i)\{\boldsymbol{Z}_i - \bar{\boldsymbol{e}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, s)\}dM_i(s)$$

$$-n^{-1/2}\sum_{i=1}^{n}\kappa_i\{\bar{\boldsymbol{e}}(\hat{\boldsymbol{\alpha}},\boldsymbol{\beta},s)-\bar{\boldsymbol{e}}(\boldsymbol{\alpha}_0,\boldsymbol{\beta},s)\}dM_i(s)\equiv(I)-(II) \quad \text{(B.9)}$$

where $\bar{\boldsymbol{e}}(\boldsymbol{\alpha},\boldsymbol{\beta},t) = \bar{s}^{(1)}(\boldsymbol{\alpha},\boldsymbol{\beta},t)/\bar{\boldsymbol{s}}^{(0)}(\boldsymbol{\alpha},\boldsymbol{\beta},t)$ and $\bar{\boldsymbol{s}}^{(j)}(\boldsymbol{\alpha},\boldsymbol{\beta},t) = E[\phi(\boldsymbol{\alpha},\boldsymbol{O})Y(t)\boldsymbol{Z}^{\otimes j}$ $\exp\{\boldsymbol{\beta}^T\boldsymbol{Z}\}|D_1 > D_0]$, $j=0,1,2$.

Secondly, plugging in (B.8) into (I), coupled with standard manipulations, leads to

$$\begin{aligned}(I) &= n^{-1/2}\sum_{i=1}^{n}\left(n^{-1}\sum_{j=1}^{n}\left[\boldsymbol{I}_\kappa(\boldsymbol{\alpha}_0,\boldsymbol{O}_i,\boldsymbol{O}_j)\int_0^\infty\{\boldsymbol{Z}_j-\boldsymbol{e}_c(\boldsymbol{\beta},s)\}dM_j(s)\right]\right)\\ &\approx n^{-1/2}\sum_{i=1}^{n}\boldsymbol{I}_{\boldsymbol{A},(I)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta},\boldsymbol{O}_i), \qquad \text{(B.10)}\end{aligned}$$

where $\boldsymbol{I}_{\boldsymbol{A},(I)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta},\boldsymbol{O}_i) = E_{\boldsymbol{O}}\left[\boldsymbol{I}_\kappa(\boldsymbol{\alpha}_0,\boldsymbol{O}_i,\boldsymbol{O})\int_0^\infty\{\boldsymbol{Z}-\boldsymbol{e}_c(\boldsymbol{\beta},s)\}dM(s)\right]$ and $E_{\boldsymbol{O}}$ stands for expectation with respect to $\boldsymbol{O}$.

Thirdly, assessing $\bar{\boldsymbol{e}}(\hat{\boldsymbol{\alpha}},\boldsymbol{\beta},s)-\bar{\boldsymbol{e}}(\boldsymbol{\alpha}_0,\boldsymbol{\beta},s)$ through the Taylor expansion and using condition (C6), we derive that

$$\begin{aligned}(II) &\approx n^{-1/2}\sum_{i=1}^{n}\left\{n^{-1}\sum_{j=1}^{n}\int_0^\infty\phi(\boldsymbol{\alpha}_0,\boldsymbol{O}_j)\boldsymbol{D}_{\bar{\boldsymbol{e}}}(\boldsymbol{\alpha}_0,\boldsymbol{\beta},s)dM_j(s)\right\}\boldsymbol{I}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_0,\boldsymbol{O}_i)\\ &\approx n^{-1/2}\sum_{i=1}^{n}\boldsymbol{I}_{\boldsymbol{A},(II)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta},\boldsymbol{O}_i) \qquad \text{(B.11)}\end{aligned}$$

where $\boldsymbol{D}_{\bar{\boldsymbol{e}}}(\boldsymbol{\alpha},\boldsymbol{\beta},t) = \partial\bar{\boldsymbol{e}}(\boldsymbol{\alpha},\boldsymbol{\beta},t)/\partial\boldsymbol{\alpha}$, and $\boldsymbol{I}_{\boldsymbol{A},(II)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta},\boldsymbol{O}_i) = E_{\boldsymbol{O}}\{\int_0^\infty\phi(\boldsymbol{\alpha}_0,\boldsymbol{O})\boldsymbol{D}_{\bar{\boldsymbol{e}}}(\boldsymbol{\alpha}_0,\boldsymbol{\beta},s)\ dM(s)\}\cdot\boldsymbol{I}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_0,\boldsymbol{O}_i)$.

Define $\boldsymbol{I}_{\boldsymbol{A}}(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{O}_i) = \boldsymbol{I}_{\boldsymbol{A},(I)}(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{O}_i) - \boldsymbol{I}_{\boldsymbol{A},(II)}(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{O}_i)$ and $\boldsymbol{a}_i(\boldsymbol{\beta}) = \int_0^\infty\kappa_i\{\boldsymbol{Z}_i-\boldsymbol{e}_c(\boldsymbol{\beta},s)\}\ dM_i(s)$. It follows from (B.6), (B.7), (B.9), (B.10), and (B.11) that

$$n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\approx\{\varphi_n(\boldsymbol{\beta}_0)\}^{-1}\{\bar{\boldsymbol{U}}_{n,\kappa}(\boldsymbol{\beta}_0)+n^{-1/2}\sum_{i=1}^{n}\boldsymbol{I}_{\boldsymbol{A}}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\boldsymbol{O}_i)\}$$

$$= n^{-1/2}\sum_{i=1}^{n}\{\varphi_n(\boldsymbol{\beta}_0)\}^{-1}\{\boldsymbol{a}_i(\boldsymbol{\beta}_0)+\boldsymbol{I}_{\boldsymbol{A}}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\boldsymbol{O}_i)\}$$

By the Central Limit Theory, we have

$$n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\rightarrow_d N(0,\Omega),$$

where

$$\Omega = E([\{\varphi_n(\boldsymbol{\beta}_0)\}^{-1}\{\boldsymbol{a}_i(\boldsymbol{\beta}_0)+\boldsymbol{I}_{\boldsymbol{A}}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\boldsymbol{O})\}]^{\otimes 2}). \tag{B.12}$$

S3. Generalizations to complex survival settings

Survival data are often subject to complications other than random right censoring, for example, left truncation, competing risks, and recurrent events. The proposed weighting scheme can be readily adapted to accommodate these additional data complexities.

Left truncation: Suppose the survival time T is subject to left truncation by L. We observe $\tilde{\boldsymbol{O}} \equiv (\tilde{W},\tilde{\delta},\tilde{L},\tilde{D},\tilde{X},\tilde{V})$, where $\tilde{O}$ follows the conditional distribution of (W,δ,L,D,X,V) given $L<W$. Let $\tilde{O}_i \equiv (\tilde{W}_i,\tilde{\delta}_i,\tilde{L}_i,\tilde{D}_i,\tilde{X}_i,\tilde{V}_i)$ be the sample analogue of $\tilde{O}$. Assume that (L,C) is independent of T given (V,D,X), and (L,C) is independent of V given X.

Define $\tilde{N}(t)=I(\tilde{L}<\tilde{W}\le t,\delta=1)$, $\tilde{Y}(t)=I(\tilde{W}\ge t>\tilde{L})$, and $\tilde{M}(t)\equiv\tilde{N}(t)-\int_0^t\tilde{Y}(s)\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})h_0(s)ds$. The partial likelihood score equation under left truncation (Andersen et al., 2012) suggests that $\tilde{\boldsymbol{\mu}}_c(\boldsymbol{\beta}_0)=0$, where

$$\tilde{\boldsymbol{\mu}}_c(\boldsymbol{\beta}) = E\left[\int_0^\infty\left\{\tilde{\boldsymbol{Z}}-\frac{\tilde{\mathbf{s}}_c^{(1)}(\boldsymbol{\beta},s)}{\tilde{s}_c^{(0)}(\boldsymbol{\beta},s)}\right\}d\tilde{M}(s)\middle| D_1>D_0, L<W\right],$$

where $\tilde{\boldsymbol{s}}_c^{(j)}(\boldsymbol{\beta}, s) = E(\tilde{Y}(s)\tilde{\mathbf{Z}}^{\otimes j}e^{\boldsymbol{\beta}^T\tilde{\mathbf{Z}}}|D_1 > D_0, L < W)$ $(j = 0, 1, 2)$. Applying the same technique shown in (3), one can establish that

$$\tilde{\boldsymbol{\mu}}_c(\boldsymbol{\beta}) = \frac{1}{\Pr(D_1 > D_0|L < W)} E\left[\tilde{\kappa}\int_0^\infty \left\{\tilde{\boldsymbol{Z}} - \frac{\tilde{\boldsymbol{s}}_c^{(1)}(\boldsymbol{\beta}, s)}{\tilde{s}_c^{(0)}(\boldsymbol{\beta}, s)}\right\} d\tilde{M}(s)\middle| L < W\right], \qquad \text{(C.1)}$$

where

$$\tilde{\kappa} = 1 - \frac{\tilde{D}(1 - \tilde{V})}{\Pr(\tilde{V} = 0|\tilde{\boldsymbol{X}}, L < W)} - \frac{(1 - \tilde{D})\tilde{V}}{\Pr(\tilde{V} = 1|\tilde{\boldsymbol{X}}, L < W)}$$

and

$$\boldsymbol{s}_c^{(j)}(\boldsymbol{\beta}, s) = \frac{E(\kappa_l\tilde{Y}(s)\tilde{\mathbf{Z}}^{\otimes j}e^{\boldsymbol{\beta}^T\tilde{\mathbf{Z}}}|L < W)}{\Pr(D_1 > D_0, L < W)}, \quad j = 0, 1, 2.$$

The result in (C.1) suggests a simple adaptation of the proposed method to the case with random left truncation, where the main modification is to replace $\boldsymbol{Z}_i$, $Y_i(t)$, $N_i(t)$ in $\boldsymbol{U}_{n,\kappa}(\boldsymbol{\beta})$ by $\tilde{\boldsymbol{Z}}_i$, $\tilde{Y}_i(t)$, $\tilde{N}_i(t)$ respectively. The weights $\hat{\kappa}$ or $\hat{\kappa}_v$ can be calculated in the same way as in Section 2.3 and 2.5 based on $\tilde{D}_i, \tilde{V}_i, \tilde{\boldsymbol{X}}_i, \tilde{W}_i$ observed under left truncation.

Competing risks: Consider a typical competing risks setting with K types of competing failures. Let $T = \min(T_1, \ldots, T_K)$, where T_k denotes the latent event time to failure type k $(k = 1, \ldots, K)$. Let C denote time to random censoring for T, which satisfies the same censoring assumptions stated in Section 2.3. Let $W = \min(T, C)$ and define η as 0 if $T > C$ and the type of failure otherwise. We observe $(T, \eta, V, D, \boldsymbol{X})$.

When the interest lies in the minimal event time T, one can simply apply the procedures in Section 2.3-2.5 to the observed data on $(T, I(\eta \neq 0), V, D, \boldsymbol{X})$. This is appropriate because T, when treated as a survival outcome of interest, is only subject to random censoring by C, and $I(\eta \neq 0)$ indicates whether T is observed or not.

When the interest pertains to a specific type of failure, say type k, we propose to consider the following variant of model (1) to define the causal treatment effect of interest:

$$h_k(t; D, \boldsymbol{X}) = h_{k,0}(t) \exp\{\beta_{d,k} D + \boldsymbol{\beta}_{x,k}^T \boldsymbol{X}\}, \tag{C.2}$$

where $h_k(t; D, \boldsymbol{X})$ is the type-k cause-specific hazard function for compliers defined as

$$h_k(t; D, \boldsymbol{X}) = \lim_{\Delta t \to 0} \frac{\Pr(t \le T \le t + \Delta t | T \ge t, \delta = k, D_1 > D_0, D, \boldsymbol{X})}{\Delta t},$$

and $h_{k,0}(t)$ is an unspecified baseline cause-specific hazard at time t for type k. Under model (C.2), $\boldsymbol{\beta}_{d,k}$ represents the causal treatment effect on the type-k cause-specific hazard for compliers after adjusting for covariates in $\boldsymbol{X}$. When all subjects are compliers, one can estimate model (C.2) using a slightly modified partial likelihood score equation, which is (2) with $I(\eta = k)$ replacing δ (Kalbfleisch and Prentice, 2011). Following the same arguments for justifying the weighting technique presented in Section 2.3, we can show that incorporating $\hat{\kappa}$ or $\hat{\kappa}_v$ into this modified score equation yields an unbiased estimating equation for $\boldsymbol{\beta}_{x,k}$. In other words, naively treating the competing risks for type-k failure as independent censoring events and applying the proposed IV method for randomly censored data lead to legitimate estimation and inference for the causal treatment effect on the type-k cause-specific hazard.

Recurrent events: In survival settings, the event of interest may occur repeatedly over time. The proportional hazards model can be naturally extended to a proportional intensity model to accommodate recurrent events (Andersen and Gill, 1982). Let $T^{(j)}$

denote the j-th recurrent event. Define $N^*(t) = \sum_{j=1}^{\infty} I(T^{(j)} \leq t)$ and $N^r(t) = \sum_{j=1}^{\infty} I(L < T^{(j)} \leq R)$, which respectively represent the underlying and the observed counting processes of recurrent events. Here $(L, R]$ denotes the time window in which recurrent events are observed. We assume L and R are independent of V given $\boldsymbol{X}$ and are independent of $T^{(j)}$'s conditional on $(V, D, \boldsymbol{X})$. Let $Y^r(t) = I(L < t \leq R)$, which denotes the at-risk process. A causal proportional intensity model is defined similarly to the Cox's proportional hazards model (1):

$$\lambda(t) = \lambda_0(t) \exp\{\beta_{r,d} D + \boldsymbol{\beta}_{r,x}^T \boldsymbol{X}\}, \tag{C.3}$$

where $\lambda(t)$ denotes the intensity function associated with $N^*(\cdot)$ given compliers (i.e. $D_1 > D_0$), and $\lambda_0(t)$ is an unspecified baseline intensity function. The causal treatment effect on the recurrent events for compliers is captured by $\boldsymbol{\beta}_{r,d}$. As shown by Andersen and Gill (1982), in the setting where all subjects are compliers, $\boldsymbol{\beta}_{r,d}$ can be estimated by equation (2) with $N^r(\cdot)$ in place of $N(\cdot)$ and $Y^r(\cdot)$ in place of $Y(\cdot)$. Adapting the weighting technique developed in Section 2.3 and 2.5, we can similarly modify the estimating equation for $\boldsymbol{\beta}_{r,d}$ by incorporating weights $\hat{\kappa}$ or $\hat{\kappa}_v$. That is, we can obtain an unbiased estimate for $\boldsymbol{\beta}_{r,d}$ by solving the equation (7) with $N^r(\cdot)$ and $Y^r(\cdot)$ in place of $N(\cdot)$ and $Y(\cdot)$.

S4. Supplemental Figures and Tables

Below we provide two figures showing the objective function and estimating function surfaces for the three proposed estimators. These were selected to demonstrate the numerical issues present in the approach and why convergence sometimes fails, and the benefits of the modified and truncated weight $\hat{\kappa}_{v,tr}$. The figures also demonstrate why we prefer to utilize the objective function over the estimating equation.

References

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of econometrics 113*(2), 231–263.

Abadie, A., J. Angrist, and G. Imbens (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica 70*(1), 91–117.

Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (2012). *Statistical models based on counting processes.* Springer Science & Business Media.

Andersen, P. K. and R. D. Gill (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics 10*(4), 1100–1120.

Kalbfleisch, J. D. and R. L. Prentice (2011). *The statistical analysis of failure time data*, Volume 360. John Wiley & Sons.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer Science & Business Media.

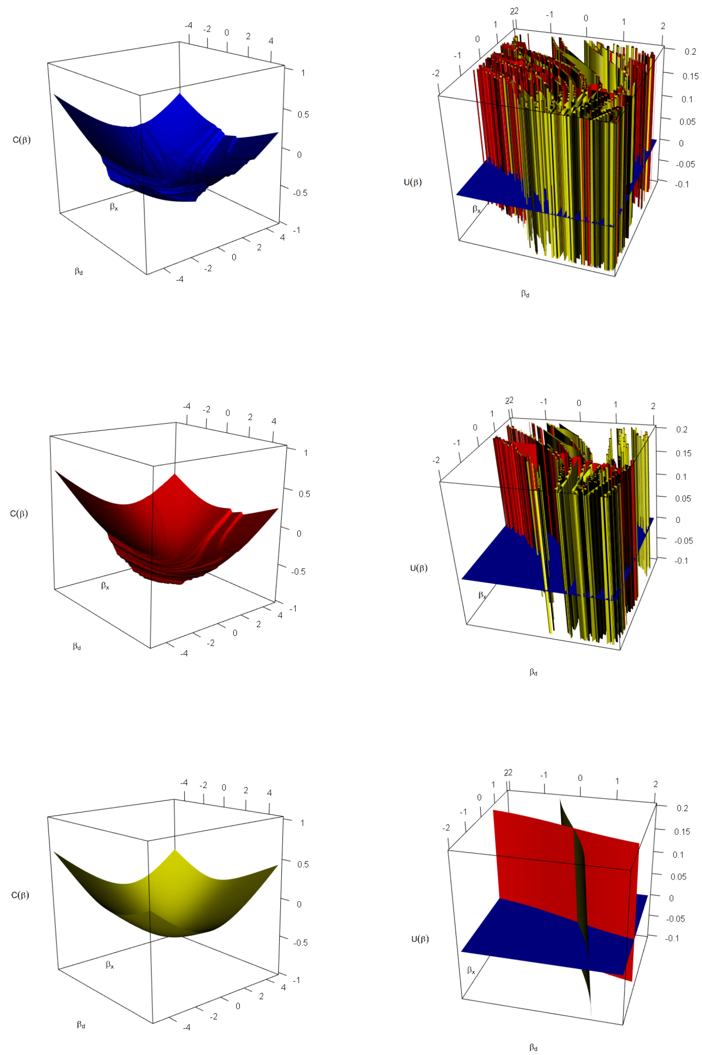

Figure 1: Non-converged dataset from simulation 2 case 1. The left column shows the objective function for κ, κ_v, $\kappa_{v,tr}$, respectively. The right column shows the estimating function plots for methods, κ, κ_v, $\kappa_{v,tr}$, respectively

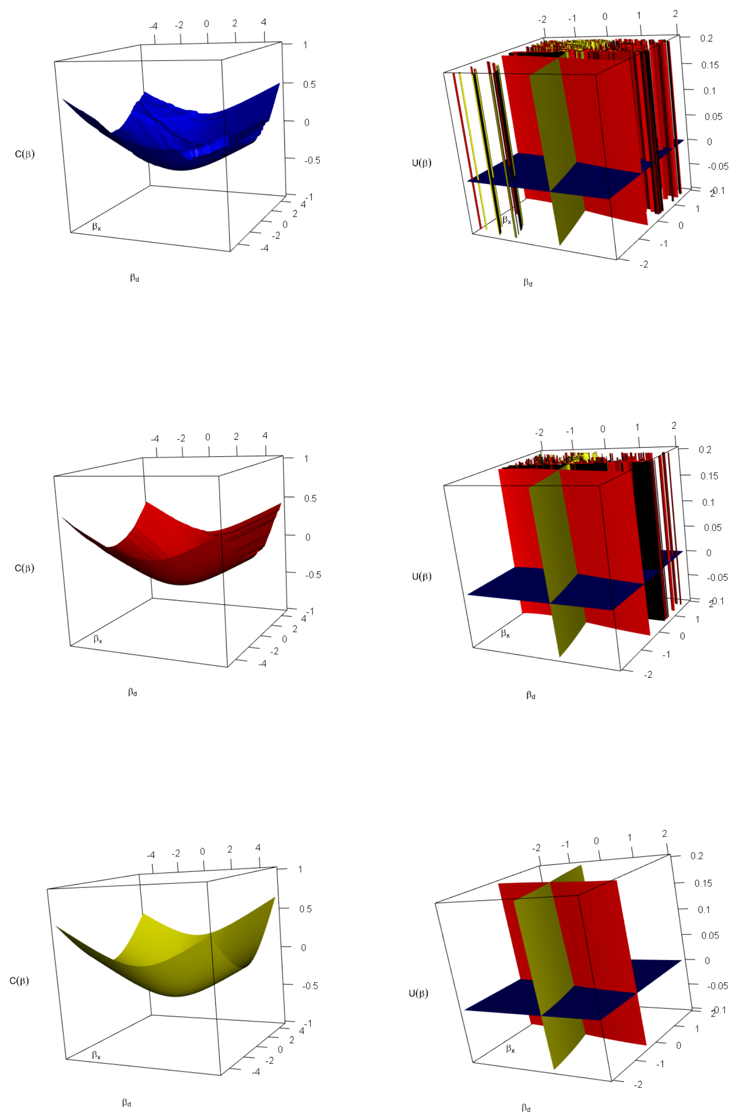

Figure 2: Dataset from scenario 2 case 1 with converged estimate. The left column shows the objective function for κ, κ_v, $\kappa_{v,tr}$, respectively. The right column shows the estimating function plots for κ, κ_v, $\kappa_{v,tr}$, respectively