Web Appendix for "Clustered restricted mean survival time regression"

Xinyuan Chen[*], Michael O. Harhay, and Fan Li

A1 Unbiasedness of the Estimating Equation

Here we show the estimation equation $\mathbf{\Phi}^*(\tilde{\boldsymbol{\beta}})$ is unbiased at the true value of $\tilde{\boldsymbol{\beta}}_T$ for a given value of L under regularity conditions (i)-(vii). That is, $\mathbb{E}\{\boldsymbol{\epsilon}_{ijk}(\tilde{\boldsymbol{\beta}}_T)\} = \mathbf{0}$. Let

$$\boldsymbol{\epsilon}_{ijk}(\tilde{\boldsymbol{\beta}}_T) = \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\Delta_{ijk}w_{ij}^C(Y_{ijk})[Y_{ijk} - g^{-1}\{\tilde{\boldsymbol{\beta}}_T'\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\}].$$

We then have

$$\begin{aligned}
\mathbb{E}\{\boldsymbol{\epsilon}_{ijk}(\tilde{\boldsymbol{\beta}}_T)|\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\} =& \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\mathbb{E}\left\{\Delta_{ijk}w_{ij}^C(Y_{ijk})Y_{ijk}|\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\} \\
&- \tilde{\boldsymbol{Z}}_{ij}^T(L_k)g^{-1}\left\{\tilde{\boldsymbol{\beta}}_T'\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\}\mathbb{E}\left\{\Delta_{ijk}w_{ij}^C(Y_{ijk})|\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\} \\
=& \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\mathbb{E}\left[\mathbb{E}\left\{\Delta_{ijk}w_{ij}^C(Y_{ijk})Y_{ijk}|T_{ij}\right\}|\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right] \\
&- \tilde{\boldsymbol{Z}}_{ij}^T(L_k)g^{-1}\left\{\tilde{\boldsymbol{\beta}}_T'\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\}\mathbb{E}\left[\mathbb{E}\left\{\Delta_{ijk}w_{ij}^C(Y_{ijk})|T_{ij}\right\}|\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right]
\end{aligned}$$

Plugging in the definition of the IPCW, we obtain

$$\begin{aligned}
\mathbb{E}\{\boldsymbol{\epsilon}_{ijk}(\tilde{\boldsymbol{\beta}}_T)|\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\} =& \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\mathbb{E}\left[\mathbb{E}\left\{\frac{\mathbb{I}(T_{ij}\wedge L_k \le C_{ij})}{\mathbb{P}(T_{ij}\wedge L_k \le C_{ij})}(C_{ij}\wedge T_{ij}\wedge L_k)|T_{ij}\right\}|\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right] \\
&- \tilde{\boldsymbol{Z}}_{ij}^T(L_k)g^{-1}\left\{\tilde{\boldsymbol{\beta}}_T'\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\}\mathbb{E}\left[\mathbb{E}\left\{\frac{\mathbb{I}(T_{ij}\wedge L_k \le C_{ij})}{\mathbb{P}(T_{ij}\wedge L_k \le C_{ij})}|T_{ij}\right\}|\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right] \\
=& \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\mathbb{E}\left[\mathbb{E}\left\{\frac{\mathbb{I}(T_{ij}\wedge L_k \le C_{ij})}{\mathbb{P}(T_{ij}\wedge L_k \le C_{ij})}(T_{ij}\wedge L_k)|T_{ij}\right\}|\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right] \\
&- \tilde{\boldsymbol{Z}}_{ij}^T(L_k)g^{-1}\left\{\tilde{\boldsymbol{\beta}}_T'\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\}\mathbb{E}\left[\mathbb{E}\left\{\frac{\mathbb{I}(T_{ij}\wedge L_k \le C_{ij})}{\mathbb{P}(T_{ij}\wedge L_k \le C_{ij})}|T_{ij}\right\}|\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right] \\
=& \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\mathbb{E}\left\{T_{ij}\wedge L_k|\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\} - \tilde{\boldsymbol{Z}}_{ij}^T(L_k)g^{-1}\left\{\tilde{\boldsymbol{\beta}}_T'\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\} \\
=& \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\left[\mathbb{E}\left\{T_{ij}\wedge L_k|\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\} - g^{-1}\left\{\tilde{\boldsymbol{\beta}}_T'\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\}\right] \\
=& \mathbf{0}.
\end{aligned}$$

The unconditional expectation is equal to $\mathbf{0}$ since it is the average over conditional terms whose expectations are $\mathbf{0}$.

[*]xchen@math.msstate.edu

A2 Proof of Theorem 2.1

We first re-introduce the regularity conditions and the theorem before start the proof:

(i) $\{\boldsymbol{X}_i, \boldsymbol{\Delta}_i, \boldsymbol{Z}_i, m_i\}$, $i = 1, \ldots, n$, are independent.

(ii) $\mathbb{P}\{R_{ij}(t) = 1\} > 0$ for $t \in (0, \tau)$, $i = 1, \ldots, n$.

(iii) There exists a finite constant M_Z such that $|Z_{ijl}| < M_Z < \infty$ for $j = 1, \ldots, m_i$ and $i = 1, \ldots, n$, where Z_{ijl} is the lth component of $\boldsymbol{Z}_{ij}$.

(iv) $\Lambda_{ij}^C(t) < \infty$ and $\Lambda_{ij}^C(t)$ is absolutely continuous for $t \in (0, \tau]$.

(v) There exists neighborhood $\mathcal{B}_C$ of $\boldsymbol{\beta}_C$ such that for $\vartheta = 0, 1, 2$,

$$\sup_{t\in(0,\tau],\ \in\mathcal{B}_C} \left\| n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \exp\left(\boldsymbol{\beta}_C' \boldsymbol{Z}_{ij}^C\right) R_{ij}(t) \boldsymbol{Z}_{ij}^{C\otimes\vartheta} - \boldsymbol{r}_C^{(\vartheta)}(t; \boldsymbol{\beta}) \right\| \xrightarrow{p} 0,$$

where $\boldsymbol{v}^{\otimes 0} = 1$, $\boldsymbol{v}^{\otimes 1} = \boldsymbol{v}$, $\boldsymbol{v}^{\otimes 2} = \boldsymbol{v}\boldsymbol{v}'$, and $\boldsymbol{r}_C^{(\vartheta)}(t; \boldsymbol{\beta}) = \mathbb{E}\left\{\exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_{ij}^C\right) R_{ij}(t) \boldsymbol{Z}_{ij}^{C\otimes\vartheta}\right\}$.

(vi) Define $h(\boldsymbol{x}) = \partial g^{-1}(\boldsymbol{x})/\partial \boldsymbol{x}$, where $h(\cdot)$ exists and is continuous in an open neighborhood $\tilde{\mathcal{B}}_T$ of $\tilde{\boldsymbol{\beta}}_T$.

(vii) Matrices $\boldsymbol{\Omega}_T(\tilde{\boldsymbol{\beta}}_T)$ and $\boldsymbol{\Omega}_C(\boldsymbol{\beta}_C)$ are both positive definite, and are defined as:

$$\boldsymbol{\Omega}_T(\tilde{\boldsymbol{\beta}}_T) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{K} \tilde{\boldsymbol{Z}}_{ij}^T(L_k)^{\otimes 2} h\left\{\tilde{\boldsymbol{\beta}}_T' \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\}$$

$$\boldsymbol{\Omega}_C(\boldsymbol{\beta}_C) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^{\tau} \left\{ \frac{\boldsymbol{r}_C^{(2)}(t; \boldsymbol{\beta}_C)}{r^{(0)}(t; \boldsymbol{\beta}_C)} - \bar{\boldsymbol{z}}_C(t; \boldsymbol{\beta}_C)^{\otimes 2} \right\} \mathrm{d}N_{ij}^C(t),$$

where $\bar{\boldsymbol{z}}_C(t; \boldsymbol{\beta}_C) = \boldsymbol{r}_C^{(1)}(t; \boldsymbol{\beta}_C)/r^{(0)}(t; \boldsymbol{\beta}_C)$ and $N_{ij}^C(t) = \mathbb{I}(X_{ij} \le t, \Delta_{ij} = 0)$ is the counting process for censoring time.

Theorem 1 (Theorem 2.1)**.** *Under regularity conditions (i)-(vii), as* $n \to \infty$, $\sqrt{n}\boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}}_T) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}(\tilde{\boldsymbol{\beta}}_T))$, *where* $\boldsymbol{\Sigma}(\tilde{\boldsymbol{\beta}}_T) = \mathbb{E}\{\boldsymbol{B}_i(\tilde{\boldsymbol{\beta}}_T)^{\otimes 2}\}$. *Speci cally,* $\boldsymbol{B}_i(\tilde{\boldsymbol{\beta}}_T) = \sum_{j=1}^{m_i} \boldsymbol{B}_{ij}(\tilde{\boldsymbol{\beta}}_T)$, *and* $\boldsymbol{B}_{ij}(\tilde{\boldsymbol{\beta}}_T) = \sum_{k=1}^{K}\{\boldsymbol{\epsilon}_{ijk}(\tilde{\boldsymbol{\beta}}_T) + \boldsymbol{\Omega}_C^{-1}(\boldsymbol{\beta}_C)\boldsymbol{U}_{ij}^C(\boldsymbol{\beta}_C)\boldsymbol{K}_k^C(\tilde{\boldsymbol{\beta}}_T) + \boldsymbol{J}_{ijk}^C(\tilde{\boldsymbol{\beta}}_T)\}$, *where we de ne:*

$$\boldsymbol{\epsilon}_{ijk}(\tilde{\boldsymbol{\beta}}_T) = \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\Delta_{ijk} w_{ij}^C(Y_{ijk}) \left[Y_{ijk} - g^{-1}\left\{\tilde{\boldsymbol{\beta}}_T' \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\}\right]$$

$$\boldsymbol{U}_{ij}^C = \int_0^{\tau} \left\{\boldsymbol{Z}_{ij}^C - \bar{\boldsymbol{z}}_C(t; \boldsymbol{\beta}_C)\right\} \mathrm{d}M_{ij}^C(t)$$

$$\boldsymbol{D}_{ij}^C(t) = \int_0^{t} \left\{\boldsymbol{Z}_{ij}^C - \bar{\boldsymbol{z}}_C(u; \boldsymbol{\beta}_C)\right\} \mathrm{d}\Lambda_{ij}^C(u)$$

$$\boldsymbol{K}_k^C(\tilde{\boldsymbol{\beta}}_T) = \lim_{n\to\infty} n^{-1} \sum_{l=1}^{n} \sum_{h=1}^{m_l} \boldsymbol{\epsilon}_{lhk}(\tilde{\boldsymbol{\beta}}_T) \boldsymbol{D}_{lh}^C(Y_{lhk})'$$

$$J_{ijlh}^C(t) = \int_0^{t} e^{\,{}'_{C} Z^{C}_{ij}} R_{ij}(u) \frac{\mathrm{d}M_{lh}^C(u)}{r_C^{(0)}(u; \boldsymbol{\beta}_C)}$$

$$\boldsymbol{J}_{ijk}^C(\tilde{\boldsymbol{\beta}}_T) = \lim_{n\to\infty} n^{-1} \sum_{l=1}^{n} \sum_{h=1}^{m_l} \boldsymbol{\epsilon}_{lhk}(\tilde{\boldsymbol{\beta}}_T) J_{ijlh}^C(Y_{lhk}).$$

Here $\mathrm{d}M_{ij}^C(t) = \mathrm{d}N_{ij}^C(t) - R_{ij}(t)\mathrm{d}\Lambda_{ij}^C(t)$ *is the martingale increment for the censoring time.*

Proof. Theorem 2.1 states the asymptotic normality of the estimation equation, where the estimation consistency of the IPCW is required. The estimation consistency of the IPCW is further determined by that of the marginal hazard function of censoring times. For clustered survival outcomes, the estimation consistency of the marginal hazard function of censoring timer has been established by Spiekerman and Lin (1998). Then using the continuous mapping theorem, the consistency of the IPCW can be obtained. Thus, for censoring times C_{ij}, we can obtain the following for the weights using results in Zhang and Schaubel (2011):

$$\sqrt{n}\left\{\hat{w}_{ij}^C(t) - w_{ij}^C(t)\right\} = n^{-1/2} w_{ij}^C(t)\left\{\boldsymbol{D}_{ij}^C(t)'\boldsymbol{\Omega}_C^{-1}(\boldsymbol{\beta}_C)\sum_{l=1}^{n}\sum_{h=1}^{m_l}\boldsymbol{U}_{lh}^C(\boldsymbol{\beta}_C) + \sum_{l=1}^{n}\sum_{h=1}^{m_l} J_{ijlh}^C(t)\right\} + o_p(1).$$

The normalized estimating equation can be expressed as:

$$\begin{aligned}
&\sqrt{n}\boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}}_T) \\
&= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\Delta_{ijk}\left[Y_{ijk} - g^{-1}\left\{\tilde{\boldsymbol{\beta}}_T'\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\}\right]\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\left[w_{ij}^C(Y_{ijk}) + \left\{\hat{w}_{ij}^C(Y_{ijk}) - w_{ij}^C(Y_{ijk})\right\}\right] \\
&= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\Delta_{ijk}\left[Y_{ijk} - g^{-1}\left\{\tilde{\boldsymbol{\beta}}_T'\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\}\right]\tilde{\boldsymbol{Z}}_{ij}^T(L_k) w_{ij}^C(Y_{ijk}) \qquad \text{(A2.1)} \\
&\quad + n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\Delta_{ijk}\left[Y_{ijk} - g^{-1}\left\{\tilde{\boldsymbol{\beta}}_T'\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\}\right]\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\left\{\hat{w}_{ij}^C(Y_{ijk}) - w_{ij}^C(Y_{ijk})\right\}. \qquad \text{(A2.2)}
\end{aligned}$$

By definition, (A2.1) is $n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\boldsymbol{\epsilon}_{ijk}(\tilde{\boldsymbol{\beta}}_T)$. And (A2.2) can be rewritten as

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\Delta_{ijk}\left[Y_{ijk} - g^{-1}\left\{\tilde{\boldsymbol{\beta}}_T'\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\}\right]\tilde{\boldsymbol{Z}}_{ij}^T(L_k)\left\{\hat{w}_{ij}^C(Y_{ijk}) - w_{ij}^C(Y_{ijk})\right\} \\
&= n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\boldsymbol{\epsilon}_{ijk}(\tilde{\boldsymbol{\beta}}_T)\left\{\boldsymbol{D}_{ij}^C(Y_{ijk})'\boldsymbol{\Omega}_C^{-1}(\boldsymbol{\beta}_C)\sum_{l=1}^{n}\sum_{h=1}^{m_l}\boldsymbol{U}_{lh}^C(\boldsymbol{\beta}_C) + \sum_{l=1}^{n}\sum_{h=1}^{m_l}J_{ijlh}^C(Y_{ijk})\right\} + \boldsymbol{o_p}(\mathbf{1}) \\
&= n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\sum_{l=1}^{n}\sum_{h=1}^{m_l}\boldsymbol{\epsilon}_{ijk}(\tilde{\boldsymbol{\beta}}_T)\left\{\boldsymbol{D}_{ij}^C(Y_{ijk})'\boldsymbol{\Omega}_C^{-1}(\boldsymbol{\beta}_C)\boldsymbol{U}_{lh}^C(\boldsymbol{\beta}_C) + J_{ijlh}^C(Y_{ijk})\right\} + \boldsymbol{o_p}(\mathbf{1}) \\
&= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\left[\left\{n^{-1}\sum_{l=1}^{n}\sum_{h=1}^{m_l}\boldsymbol{\epsilon}_{lhk}(\tilde{\boldsymbol{\beta}}_T)\boldsymbol{D}_{lh}^C(Y_{lhk})'\right\}\boldsymbol{\Omega}_C^{-1}(\boldsymbol{\beta}_C)\boldsymbol{U}_{ij}^C(\boldsymbol{\beta}_C)\right. \\
&\qquad\qquad \left. + \left\{n^{-1}\sum_{l=1}^{n}\sum_{h=1}^{m_l}\boldsymbol{\epsilon}_{lhk}(\tilde{\boldsymbol{\beta}}_T)J_{ijlh}^C(Y_{lhk})\right\}\right] + \boldsymbol{o_p}(\mathbf{1}).
\end{aligned}$$

We then define the following:

$$\begin{aligned}
\boldsymbol{K}_k^C(\tilde{\boldsymbol{\beta}}_T) &= \lim_{n\to\infty} n^{-1}\sum_{l=1}^{n}\sum_{h=1}^{m_l}\boldsymbol{\epsilon}_{lhk}(\tilde{\boldsymbol{\beta}}_T)\boldsymbol{D}_{lh}^C(Y_{lhk})', \\
\boldsymbol{J}_{ijk}^C(\tilde{\boldsymbol{\beta}}_T) &= \lim_{n\to\infty} n^{-1}\sum_{l=1}^{n}\sum_{h=1}^{m_l}\boldsymbol{\epsilon}_{lhk}(\tilde{\boldsymbol{\beta}}_T)J_{ijlh}^C(Y_{lhk}),
\end{aligned}$$

which are bounded because of conditions (iii)-(v). Then, (A2.2) further becomes

$$n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\left\{\boldsymbol{K}_k^C(\tilde{\boldsymbol{\beta}}_T)\boldsymbol{\Omega}_C^{-1}(\boldsymbol{\beta}_C)\boldsymbol{U}_{ij}^C(\boldsymbol{\beta}_C) + \boldsymbol{J}_{ijk}^C(\tilde{\boldsymbol{\beta}}_T)\right\} + \boldsymbol{o_p}(\mathbf{1}). \qquad \text{(A2.3)}$$

Putting (A2.1) and (A2.3) together, we obtain

$$\begin{aligned}
&\sqrt{n}\boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}}_T) \\
&= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\boldsymbol{\epsilon}_{ijk}(\tilde{\boldsymbol{\beta}}_T) + n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\left\{\boldsymbol{K}_k^C(\tilde{\boldsymbol{\beta}}_T)\boldsymbol{\Omega}_C^{-1}(\boldsymbol{\beta}_C)\boldsymbol{U}_{ij}^C(\boldsymbol{\beta}_C) + \boldsymbol{J}_{ijk}^C(\tilde{\boldsymbol{\beta}}_T)\right\} + \boldsymbol{o_p}(\mathbf{1}) \\
&= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\left\{\boldsymbol{\epsilon}_{ijk}(\tilde{\boldsymbol{\beta}}_T) + \boldsymbol{K}_k^C(\tilde{\boldsymbol{\beta}}_T)\boldsymbol{\Omega}_C^{-1}(\boldsymbol{\beta}_C)\boldsymbol{U}_{ij}^C(\boldsymbol{\beta}_C) + \boldsymbol{J}_{ijk}^C(\tilde{\boldsymbol{\beta}}_T)\right\} + \boldsymbol{o_p}(\mathbf{1}).
\end{aligned}$$

Now write

$$\boldsymbol{B}_{ij}(\tilde{\boldsymbol{\beta}}_T) = \sum_{k=1}^{K} \left\{ \boldsymbol{\epsilon}_{ijk}(\tilde{\boldsymbol{\beta}}_T) + \boldsymbol{K}_k^C(\tilde{\boldsymbol{\beta}}_T)\boldsymbol{\Omega}_C^{-1}(\boldsymbol{\beta}_C)\boldsymbol{U}_{ij}^C(\boldsymbol{\beta}_C) + \boldsymbol{J}_{ijk}^C(\tilde{\boldsymbol{\beta}}_T) \right\},$$

and $\boldsymbol{B}_i(\tilde{\boldsymbol{\beta}}_T) = \sum_{j=1}^{m_i} \boldsymbol{B}_{ij}(\tilde{\boldsymbol{\beta}}_T)$. We then rewrite the normalized estimating equation as

$$\sqrt{n}\boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}}_T) = \sqrt{n}\sum_{i=1}^{n} \boldsymbol{B}_i(\tilde{\boldsymbol{\beta}}_T) + \boldsymbol{o_p}(\mathbf{1}),$$

and by the central limit theorem

$$\sqrt{n}\boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}}_T) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}(\tilde{\boldsymbol{\beta}}_T)\right),$$

with $\boldsymbol{\Sigma}(\tilde{\boldsymbol{\beta}}_T) = \mathbb{E}\{\boldsymbol{B}_i(\tilde{\boldsymbol{\beta}}_T)^{\otimes 2}\}$. □

A3 Proof of Theorem 2.2

Theorem 2 (Theorem 2.2). *Under regularity conditions (i)-(vii), as $n \to \infty$, $\hat{\boldsymbol{\beta}}_T$ is consistent and asymptotically normal, that is $\hat{\boldsymbol{\beta}}_T \xrightarrow{p} \tilde{\boldsymbol{\beta}}_T$, and $\sqrt{n}(\hat{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_T) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_T^{-1}(\tilde{\boldsymbol{\beta}}_T)\boldsymbol{\Sigma}(\tilde{\boldsymbol{\beta}}_T)\boldsymbol{\Omega}_T^{-1}(\tilde{\boldsymbol{\beta}}_T))$.*

Proof. Theorem 2.2 first establishes the consistency of $\hat{\boldsymbol{\beta}}_T$ and then specifies its asymptotic distribution. With Theorem 2.1 established, the consistency of $\hat{\boldsymbol{\beta}}_T$ can be established using the inverse function theorem (Foutz, 1977), and its asymptotic distribution can be obtained via Taylor's expansion. We first show the consistency by verifying the following conditions needed by the inverse mapping theorem:

1. $\partial \boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}})/\partial \boldsymbol{\beta}$ exists and is continuous in an open neighborhood $\tilde{\mathcal{B}}_T$ of $\tilde{\boldsymbol{\beta}}_T$.
   Verification:
$$\frac{\partial \boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} = -n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{K} \tilde{\boldsymbol{Z}}_{ij}^T(L_k)^{\otimes 2} \Delta_{ijk} w_{ijk}^C(Y_{ijk}) h\left\{\tilde{\boldsymbol{\beta}}' \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\},$$
   where $h(\boldsymbol{x}) = \partial g^{-1}(\boldsymbol{x})/\partial \boldsymbol{x}$. This condition holds because $h(\boldsymbol{x})$ is assumed to exist and be continuous in an open neighborhood $\tilde{\mathcal{B}}_T$ of $\tilde{\boldsymbol{\beta}}_T$;

2. $-\partial \boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}})/\partial \boldsymbol{\beta}|_{\tilde{\boldsymbol{\beta}}=\tilde{\boldsymbol{\beta}}_T}$ is positive definite with probability 1 as $n \to \infty$.
   Verification:
$$\begin{aligned}
&- \left.\frac{\partial \boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}}\right|_{\tilde{\boldsymbol{\beta}}=\tilde{\boldsymbol{\beta}}_T} \\
&= \mathbb{E}\left[\sum_{j=1}^{m_i}\sum_{k=1}^{K} \Delta_{ijk} w_{ijk}^C(Y_{ijk}) h\left\{\tilde{\boldsymbol{\beta}}' \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\} \tilde{\boldsymbol{Z}}_{ij}^T(L_k)^{\otimes 2}\right] + \boldsymbol{o_p}(\mathbf{1}) \\
&= \mathbb{E}\left(\mathbb{E}\left[\sum_{j=1}^{m_i}\sum_{k=1}^{K} \Delta_{ijk} w_{ijk}^C(Y_{ijk}) h\left\{\tilde{\boldsymbol{\beta}}' \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\} \tilde{\boldsymbol{Z}}_{ij}^T(L_k)^{\otimes 2} | T_{ij}, \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right]\right) + \boldsymbol{o_p}(\mathbf{1}) \\
&= \mathbb{E}\left[\sum_{j=1}^{m_i}\sum_{k=1}^{K} \mathbb{E}\left\{\Delta_{ijk} w_{ijk}^C(Y_{ijk}) | T_{ij}, \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\} h\left\{\tilde{\boldsymbol{\beta}}' \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\} \tilde{\boldsymbol{Z}}_{ij}^T(L_k)^{\otimes 2}\right] + \boldsymbol{o_p}(\mathbf{1}) \\
&= \mathbb{E}\left[\sum_{j=1}^{m_i}\sum_{k=1}^{K} \mathbb{E}\left\{\frac{\mathbb{I}(C_{ij} \geq T_{ij} \wedge L_k)}{\mathbb{P}(C_{ij} \geq T_{ij} \wedge L_k)} | T_{ij}, \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\} h\left\{\tilde{\boldsymbol{\beta}}' \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\} \tilde{\boldsymbol{Z}}_{ij}^T(L_k)^{\otimes 2}\right] + \boldsymbol{o_p}(\mathbf{1}) \\
&= \mathbb{E}\left[\sum_{j=1}^{m_i}\sum_{k=1}^{K} h\left\{\tilde{\boldsymbol{\beta}}' \tilde{\boldsymbol{Z}}_{ij}^T(L_k)\right\} \tilde{\boldsymbol{Z}}_{ij}^T(L_k)^{\otimes 2}\right] + \boldsymbol{o_p}(\mathbf{1}) \\
&= \boldsymbol{\Omega}_T(\tilde{\boldsymbol{\beta}}_T) + \boldsymbol{o_p}(\mathbf{1}).
\end{aligned}$$
   This condition holds as $n \to \infty$ because $\boldsymbol{\Omega}_T(\tilde{\boldsymbol{\beta}}_T)$ is previously assumed to be positive definite;

3. $-\partial \boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}})/\partial \boldsymbol{\beta}$ converges in probability to a fixed function uniformly in an open neighborhood $\tilde{\mathcal{B}}_T$ of $\tilde{\boldsymbol{\beta}}_T$.
   Verification:
   This condition holds by the law of large numbers;

4. The estimating function is asymptotically unbiansed, i.e., $\boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}}_T) \xrightarrow{p} \mathbf{0}$.
   Verification:
   Theorem 2.1 states $\sqrt{n}\boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}}_T) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(\tilde{\boldsymbol{\beta}}_T))$, and thus this condition holds by Chebyshev's inequality.

Thus, by the inverse mapping theorem, $\hat{\boldsymbol{\beta}}_T \xrightarrow{p} \tilde{\boldsymbol{\beta}}_T$.

We then show the asymptotic distribution of $\hat{\boldsymbol{\beta}}_T$. By Taylor's expansion of the estimating equation $\boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}})$ around $\tilde{\boldsymbol{\beta}}_T$, we have

$$\mathbf{0} = \boldsymbol{\Phi}(\hat{\boldsymbol{\beta}}_T) = \boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}}_T) + \left.\frac{\partial \boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}}\right|_{\tilde{\boldsymbol{\beta}}=\tilde{\boldsymbol{\beta}}^*} \left(\hat{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_T\right),$$

where $\tilde{\boldsymbol{\beta}}^*$ lies between $\hat{\boldsymbol{\beta}}_T$ and $\tilde{\boldsymbol{\beta}}_T$. Then write

$$
\begin{aligned}
&-\boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}}_T) = \left\{ \left. \frac{\partial \boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \right|_{\tilde{}=\tilde{}} \left(\hat{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_T \right) \right\} \\
\Longrightarrow\ &-\boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}}_T) \left\{ \left. \frac{\partial \boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \right|_{\tilde{}=\tilde{}} \right\}^{-1} = \hat{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_T \\
\Longrightarrow\ &\sqrt{n}\boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}}_T) \left\{ -\left. \frac{\partial \boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \right|_{\tilde{}=\tilde{}} \right\}^{-1} = \sqrt{n} \left(\hat{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_T \right) \\
\Longrightarrow\ &\sqrt{n}\boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}}_T) \boldsymbol{\Omega}_T^{-1}(\tilde{\boldsymbol{\beta}}^*) = \sqrt{n} \left(\hat{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_T \right) \\
\Longrightarrow\ &\sqrt{n} \left(\hat{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_T \right) = \boldsymbol{\Omega}_T(\tilde{\boldsymbol{\beta}}_T)\sqrt{n}\boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}}_T) + \boldsymbol{o_p}(\mathbf{1}).
\end{aligned}
$$

Following Theorem 2.1, we have

$$
\sqrt{n} \left(\hat{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_T \right) \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \boldsymbol{\Omega}_T^{-1}(\tilde{\boldsymbol{\beta}}_T) \boldsymbol{\Sigma}(\tilde{\boldsymbol{\beta}}_T) \boldsymbol{\Omega}_T^{-1}(\tilde{\boldsymbol{\beta}}_T) \right).
$$

□

References

Foutz, R. V. (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association*, 72(357):147–148.

Spiekerman, C. F. and Lin, D. (1998). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*, 93(443):1164–1175.

Zhang, M. and Schaubel, D. E. (2011). Estimating differences in restricted mean lifetime using observational data subject to dependent censoring. *Biometrics*, 67(3):740–749.

A4 Web Tables

Table 1: Relative bias (BIAS), empirical standard errors (ESE), average estimated standard errors (ASE), and coverage percentages of 95% confidence intervals (Cover) for the regression coefficients using the unclustered ($\hat{\boldsymbol{V}}_{ZS}$) and clustered ($\hat{\boldsymbol{V}}$) sandwich variance estimators. Large sample scenario (number of clusters $n = 100$, average cluster size $\bar{m} = 20$) with the identity link. Data were generated using $\boldsymbol{\alpha} = (4, 2, -2)'$. True $\boldsymbol{\beta}_T = (3.242, 0.574, -0.571)'$ for $L = 4$, $(4.296, 0.947, -0.943)'$ for $L = 6$, and $(5.136, 1.276, -1.270)'$ for $L = 8$.

					$\hat{\boldsymbol{V}}_{ZS}$		$\hat{\boldsymbol{V}}$	
L	Censoring %	Parameter	BIAS	ESE	ASE	Cover	ASE	Cover
4	25	$\beta_{T,0}$	-0.001	0.030	0.024	0.880	0.030	0.946
		$\beta_{T,1}$	0.001	0.057	0.042	0.860	0.054	0.930
		$\beta_{T,2}$	0.001	0.051	0.042	0.886	0.052	0.960
	50	$\beta_{T,0}$	-0.001	0.031	0.026	0.912	0.032	0.956
		$\beta_{T,1}$	-0.001	0.061	0.046	0.866	0.058	0.926
		$\beta_{T,2}$	0.001	0.054	0.046	0.898	0.056	0.962
6	25	$\beta_{T,0}$	-0.004	0.051	0.042	0.899	0.050	0.950
		$\beta_{T,1}$	0.001	0.091	0.072	0.896	0.088	0.935
		$\beta_{T,2}$	0.001	0.081	0.072	0.929	0.082	0.956
	50	$\beta_{T,0}$	-0.001	0.054	0.050	0.932	0.057	0.966
		$\beta_{T,1}$	0.003	0.102	0.087	0.898	0.099	0.933
		$\beta_{T,2}$	0.002	0.092	0.086	0.939	0.095	0.961
8	25	$\beta_{T,0}$	-0.001	0.071	0.061	0.906	0.070	0.946
		$\beta_{T,1}$	0.001	0.126	0.104	0.889	0.122	0.928
		$\beta_{T,2}$	0.001	0.110	0.104	0.942	0.113	0.957
	50	$\beta_{T,0}$	-0.001	0.078	0.080	0.953	0.086	0.972
		$\beta_{T,1}$	-0.001	0.150	0.137	0.932	0.149	0.948
		$\beta_{T,2}$	0.004	0.135	0.136	0.955	0.143	0.962

Table 2: Empirical standard errors (ESE), relative bias (BIAS), average estimated standard errors (ASE), and coverage percentages of 95% confidence intervals (Cover) for the regression coefficients using the unclustered ($\hat{\boldsymbol{V}}_{ZS}$) and clustered ($\hat{\boldsymbol{V}}$) sandwich variance estimators. Large sample scenario (number of clusters $n = 100$, average cluster size $\bar{m} = 20$) with the log link. Data were generated using $\boldsymbol{\alpha} = (1.25, \log(2), -\log(2))'$. True $\boldsymbol{\beta}_T = (1.147, 0.169, -0.142)'$ for $L = 4$, $(1.425, 0.224, -0.191)'$ for $L = 6$, and $(1.603, 0.263, -0.227)'$ for $L = 8$.

					$\hat{\boldsymbol{V}}_{ZS}$		$\hat{\boldsymbol{V}}$	
L	Censoring %	Parameter	BIAS	ESE	ASE	Cover	ASE	Cover
4	25	$\beta_{T,0}$	-0.001	0.015	0.009	0.741	0.016	0.940
		$\beta_{T,1}$	0.001	0.025	0.014	0.721	0.025	0.955
		$\beta_{T,2}$	-0.004	0.017	0.014	0.894	0.018	0.954
	50	$\beta_{T,0}$	-0.001	0.016	0.001	0.768	0.016	0.944
		$\beta_{T,1}$	0.002	0.026	0.015	0.752	0.026	0.954
		$\beta_{T,2}$	-0.005	0.019	0.016	0.899	0.019	0.954
6	25	$\beta_{T,0}$	-0.001	0.020	0.011	0.739	0.020	0.943
		$\beta_{T,1}$	0.002	0.032	0.018	0.729	0.033	0.954
		$\beta_{T,2}$	-0.004	0.021	0.018	0.911	0.022	0.958
	50	$\beta_{T,0}$	-0.001	0.021	0.013	0.800	0.021	0.952
		$\beta_{T,1}$	0.002	0.034	0.022	0.779	0.034	0.947
		$\beta_{T,2}$	-0.003	0.024	0.022	0.927	0.025	0.963
8	25	$\beta_{T,0}$	-0.001	0.024	0.014	0.750	0.024	0.939
		$\beta_{T,1}$	0.002	0.038	0.022	0.738	0.038	0.948
		$\beta_{T,2}$	-0.002	0.024	0.022	0.927	0.026	0.955
	50	$\beta_{T,0}$	-0.001	0.025	0.017	0.845	0.026	0.956
		$\beta_{T,1}$	0.002	0.042	0.028	0.803	0.042	0.943
		$\beta_{T,2}$	-0.003	0.030	0.028	0.945	0.031	0.968

Table 3: Relative bias (BIAS), empirical standard errors (ESEs), average estimated standard errors (ASEs), and coverage percentages of 95% confidence intervals (Cover) for the regression coefficients using using the uncorrected sandwich estimator ($\hat{\boldsymbol{V}}_{UC}$), and four bias-corrected variance estimators ($\hat{\boldsymbol{V}}_{MD}$, $\hat{\boldsymbol{V}}_{KC}$, $\hat{\boldsymbol{V}}_{FG}$, $\hat{\boldsymbol{V}}_{MBN}$). Small sample scenarios (number of clusters $n \in \{20, 30, 40\}$) under the log link. Data were generated using $\boldsymbol{\alpha} = (1.25, \log(2), -\log(2))'$. True $\boldsymbol{\beta}_T = (1.147, 0.169, -0.142)'$ for $L = 4$, $(1.425, 0.224, -0.191)'$ for $L = 6$, and $(1.603, 0.263, -0.227)'$ for $L = 8$. With 25% censoring, $\lambda_0^C = 0.025$ and $\boldsymbol{\beta}_C = (\log(1.5), -\log(1.5))'$.

					$\hat{\boldsymbol{V}}_{UC}$		$\hat{\boldsymbol{V}}_{MD}$		$\hat{\boldsymbol{V}}_{KC}$		$\hat{\boldsymbol{V}}_{FG}$		$\hat{\boldsymbol{V}}_{MBN}$	
L	n	Parameter	BIAS	ESE	ASE	Cover	ASE	Cover	ASE	Cover	ASE	Cover	ASE	Cover
4	20	$\beta_{T,0}$	0.001	0.036	0.033	0.909	0.035	0.926	0.034	0.913	0.034	0.913	0.047	0.977
		$\beta_{T,1}$	0.011	0.059	0.053	0.897	0.056	0.918	0.054	0.905	0.054	0.905	0.078	0.979
		$\beta_{T,2}$	0.021	0.037	0.037	0.948	0.039	0.958	0.038	0.952	0.038	0.954	0.067	1.000
	30	$\beta_{T,0}$	-0.001	0.029	0.027	0.913	0.028	0.929	0.028	0.924	0.028	0.924	0.037	0.981
		$\beta_{T,1}$	-0.013	0.047	0.044	0.912	0.046	0.923	0.045	0.918	0.045	0.918	0.062	0.988
		$\beta_{T,2}$	0.016	0.032	0.031	0.945	0.032	0.954	0.032	0.950	0.032	0.950	0.052	0.995
	40	$\beta_{T,0}$	-0.001	0.025	0.024	0.923	0.025	0.929	0.024	0.927	0.024	0.926	0.030	0.980
		$\beta_{T,1}$	0.002	0.040	0.039	0.933	0.040	0.940	0.039	0.937	0.039	0.936	0.050	0.983
		$\beta_{T,2}$	0.011	0.027	0.027	0.942	0.028	0.948	0.028	0.945	0.028	0.944	0.041	0.993
6	20	$\beta_{T,0}$	-0.001	0.048	0.043	0.908	0.046	0.924	0.044	0.915	0.044	0.914	0.053	0.957
		$\beta_{T,1}$	0.012	0.078	0.069	0.896	0.073	0.917	0.071	0.903	0.071	0.903	0.086	0.957
		$\beta_{T,2}$	0.017	0.046	0.046	0.948	0.049	0.953	0.047	0.951	0.047	0.951	0.067	0.998
	30	$\beta_{T,0}$	-0.002	0.038	0.036	0.919	0.037	0.928	0.037	0.927	0.037	0.927	0.043	0.961
		$\beta_{T,1}$	-0.013	0.062	0.058	0.920	0.060	0.931	0.059	0.928	0.059	0.928	0.069	0.965
		$\beta_{T,2}$	0.013	0.039	0.039	0.942	0.040	0.950	0.039	0.946	0.039	0.946	0.054	0.989
	40	$\beta_{T,0}$	-0.001	0.033	0.031	0.930	0.032	0.936	0.032	0.933	0.032	0.932	0.035	0.957
		$\beta_{T,1}$	-0.002	0.052	0.051	0.934	0.052	0.940	0.051	0.939	0.051	0.938	0.058	0.964
		$\beta_{T,2}$	0.009	0.033	0.034	0.950	0.035	0.953	0.034	0.950	0.034	0.950	0.043	0.981
8	20	$\beta_{T,0}$	-0.001	0.056	0.051	0.908	0.054	0.924	0.052	0.914	0.052	0.914	0.058	0.945
		$\beta_{T,1}$	0.010	0.092	0.081	0.899	0.086	0.917	0.083	0.906	0.083	0.907	0.095	0.941
		$\beta_{T,2}$	0.017	0.053	0.053	0.944	0.056	0.952	0.055	0.946	0.055	0.948	0.070	0.991
	30	$\beta_{T,0}$	-0.002	0.044	0.042	0.923	0.044	0.929	0.043	0.928	0.043	0.928	0.047	0.957
		$\beta_{T,1}$	-0.013	0.072	0.068	0.926	0.070	0.936	0.069	0.930	0.069	0.930	0.077	0.955
		$\beta_{T,2}$	0.011	0.044	0.045	0.946	0.046	0.957	0.045	0.951	0.045	0.950	0.056	0.985
	40	$\beta_{T,0}$	-0.001	0.039	0.037	0.935	0.038	0.941	0.037	0.938	0.037	0.938	0.040	0.946
		$\beta_{T,1}$	-0.003	0.062	0.059	0.933	0.061	0.940	0.060	0.938	0.060	0.938	0.065	0.950
		$\beta_{T,2}$	0.007	0.038	0.039	0.954	0.040	0.956	0.040	0.955	0.040	0.955	0.046	0.975