1 Derivation of Semiparametric Efficiency Bound

From the result of Theorem 4.2 of Robins and Rotnitzky (1992) or Lemma A.2 of Robins and Rotnitzky (1995), we know that the observed-data efficient score $\boldsymbol{S}_{eff}$ satisfies

$$\boldsymbol{S}_{eff} = \boldsymbol{Q}_{eff} + \sum_{k=1}^{m} (R_k - \pi_k R_{k-1}) \overline{\pi}_k^{-1} \{\boldsymbol{Q}_{eff} - E[\boldsymbol{Q}_{eff} | \overline{\boldsymbol{W}}_{k-1}]\}, \quad (1)$$

where $\boldsymbol{Q}_{eff}$ is the unique $\boldsymbol{Q}$ in $\Lambda^{full,\perp}$ satisfying

$$\boldsymbol{S}_{eff}^{full} = \boldsymbol{Q} + \Pi[\nu(\boldsymbol{Q}) | \Lambda^{full,\perp}], \quad (2)$$

with $\boldsymbol{S}_{eff}^{full}$ to be the full-data efficient score and $\nu(\boldsymbol{Q}) = \sum_{k=1}^{m} (1-\pi_k) \overline{\pi}_k^{-1} \{\boldsymbol{Q} - E[\boldsymbol{Q} | \overline{\boldsymbol{W}}_{k-1}]\}$. It is easy to show that the orthocomplement of the nuisance tangent space with full data under multivariate normal assumption is

$$\Lambda^{full,\perp} = \{\boldsymbol{a}(\boldsymbol{X}, \boldsymbol{T}) \boldsymbol{\varepsilon} : E[\boldsymbol{a}(\boldsymbol{X}, \boldsymbol{T}) | \boldsymbol{T}] = 0\}. \quad (3)$$

Since $\boldsymbol{Q}_{eff} \in \Lambda^{full,\perp}$, there exist $\boldsymbol{a}_Q(\boldsymbol{X}, \boldsymbol{T})$ such that $\boldsymbol{Q}_{eff} = \boldsymbol{a}_Q(\boldsymbol{X}, \boldsymbol{T}) \boldsymbol{\epsilon}$. Substitute $\boldsymbol{a}_Q(\boldsymbol{X}, \boldsymbol{T}) \boldsymbol{\varepsilon}$ for $\boldsymbol{Q}_{eff}$ in (1), we obtain $\boldsymbol{S}_{eff} = \boldsymbol{a}_Q(\boldsymbol{X}, \boldsymbol{T}) \boldsymbol{\varepsilon}^*$. Thus, we remain to show that $\boldsymbol{a}_Q(\boldsymbol{X}, \boldsymbol{T}) = \{\boldsymbol{X} - \boldsymbol{\varphi}_{eff}(\boldsymbol{t})\}^T \boldsymbol{\Sigma}^{*,-1}$, which satisfies $E[\boldsymbol{a}_Q(\boldsymbol{X}, \boldsymbol{T}) | \boldsymbol{T}] = 0$ from (3).

Since $\Pi[\nu(\boldsymbol{Q}_{eff}) | \Lambda^{full,\perp}] \in \Lambda^{full,\perp}$, there exist $\boldsymbol{a}_\nu(\boldsymbol{X}, \boldsymbol{T})$ such that $\Pi[\nu(\boldsymbol{Q}_{eff}) | \Lambda^{full,\perp}] =$

$\boldsymbol{a}_\nu(\boldsymbol{X},\boldsymbol{T})\boldsymbol{\epsilon}$. It follows that $\nu(\boldsymbol{Q}_{eff}) - \boldsymbol{a}_\nu(\boldsymbol{X},\boldsymbol{T})\boldsymbol{\epsilon} \in \Lambda^{full}$ and satisfies

$$E\left[\{\nu(\boldsymbol{Q}_{eff}) - \boldsymbol{a}_\nu(\boldsymbol{X},\boldsymbol{T})\boldsymbol{\epsilon}\}\,\boldsymbol{\varepsilon}^T\boldsymbol{a}(\boldsymbol{X},\boldsymbol{T})^T\right] = 0 \tag{4}$$

for any $\boldsymbol{a}(\boldsymbol{X},\boldsymbol{T})$. Recall that $\boldsymbol{Q}_{eff}$ must satisfy the restriction (2), we obtain that

$$\boldsymbol{a}_\nu(\boldsymbol{X},\boldsymbol{T})\boldsymbol{\epsilon} = \boldsymbol{a}_Q(\boldsymbol{X},\boldsymbol{T})\boldsymbol{\epsilon} - \boldsymbol{S}_{eff}^{full}. \tag{5}$$

In fact, Wang et al. (2005) derived the full-data semiparametric efficient score under the multivariate normal assumption,

$$\boldsymbol{S}_{eff}^{full} = \{\boldsymbol{X} - \boldsymbol{\varphi}_{eff}(\boldsymbol{t})\}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}, \tag{6}$$

where $\boldsymbol{\Sigma}$ is the conditional variance of $\boldsymbol{\varepsilon}$. Plugging (5) into (4) and after some calculation, we obtain that

$$\boldsymbol{a}_Q(\boldsymbol{X},\boldsymbol{T})E\left[\sum_{k=1}^{m}(1-\pi_k)\overline{\pi}_k^{-1}\{\boldsymbol{\varepsilon} - E[\boldsymbol{\varepsilon}|\overline{\boldsymbol{W}}_{k-1}]\}\boldsymbol{\varepsilon}^T + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T \middle| \boldsymbol{X},\boldsymbol{T}\right] = \{\boldsymbol{X} - \boldsymbol{\varphi}_{eff}(\boldsymbol{t})\}^T.$$

Define

$$\boldsymbol{M}(t) = \sum_{k=1}^{t}(R_k - \pi_k R_{k-1})\overline{\pi}_k^{-1}\{\boldsymbol{\varepsilon}^T - E[\boldsymbol{\varepsilon}|\overline{\boldsymbol{W}}_{k-1}]\}$$

and

$$\boldsymbol{K}(\boldsymbol{X},\boldsymbol{T}) = E\left[\sum_{k=1}^{m}(1-\pi_k)\overline{\pi}_k^{-1}\{\boldsymbol{\varepsilon} - E[\boldsymbol{\varepsilon}|\overline{\boldsymbol{W}}_{k-1}]\}\boldsymbol{\varepsilon}^T \middle| \boldsymbol{X},\boldsymbol{T}\right].$$

Then, $\boldsymbol{a}_Q(\boldsymbol{X},\boldsymbol{T}) = \{\boldsymbol{X} - \boldsymbol{\varphi}_{eff}(\boldsymbol{t})\}^T\boldsymbol{\Sigma}^{*,-1}$ if and only if $\boldsymbol{K}(\boldsymbol{X},\boldsymbol{T}) + E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T|\boldsymbol{X},\boldsymbol{T}] = \boldsymbol{\Sigma}^*$. From the fact (Robins et al., 1995) that $\boldsymbol{M}(t)$ is a discrete time mean zero martingale process with respect to the filtration $\sigma\{\overline{\boldsymbol{W}}_{t-1}, R_1, \ldots, R_{t-1}, \boldsymbol{\varepsilon}\}$, we have $E[\boldsymbol{\varepsilon}\boldsymbol{M}(T)^T] = 0$ and $Var[\boldsymbol{M}(T)|\boldsymbol{X}, \boldsymbol{T}] = \boldsymbol{K}(\boldsymbol{X}, \boldsymbol{T})$ if (2.4) is true. Thus,

$$\begin{aligned}\boldsymbol{\Sigma}^* &= E\left[(\boldsymbol{M}(T) - \boldsymbol{\varepsilon})(\boldsymbol{M}(T) - \boldsymbol{\varepsilon})^T|\boldsymbol{X}, \boldsymbol{T}\right] \\ &= \boldsymbol{K}(\boldsymbol{X}, \boldsymbol{T}) + E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T|\boldsymbol{X}, \boldsymbol{T}].\end{aligned}$$

2 Lemma

Lemma 1. *If* (2.1), (2.2), *and* (2.3) *are true and either* (2.4) *or* (2.8) *is true,*

$$\begin{aligned}E\Bigg\{\tilde{\boldsymbol{X}}^T\boldsymbol{\Delta}\boldsymbol{V}^{-1}\Bigg[&R_m\overline{\pi}_m^{-1}(\boldsymbol{\tau}^*)\{\boldsymbol{Y} - \boldsymbol{\mu}\} \\ &- \sum_{k=1}^{m}(R_k - \pi_k(\boldsymbol{\tau}^*)R_{k-1})\overline{\pi}_k^{-1}(\boldsymbol{\tau}^*)\{\boldsymbol{\phi}_k(\overline{\boldsymbol{W}}_{k-1}, \boldsymbol{\eta}^*) - \boldsymbol{\mu}\}\Bigg]\Bigg\} = 0.\end{aligned}$$

Proof. We write

$$\begin{aligned}E\Bigg\{\tilde{\boldsymbol{X}}^T\boldsymbol{\Delta}\boldsymbol{V}^{-1}\Bigg[&R_m\overline{\pi}_m^{-1}(\boldsymbol{\tau}^*)\{\boldsymbol{Y} - \boldsymbol{\mu}\} \\ &- \sum_{k=1}^{m}(R_k - \pi_k(\boldsymbol{\tau}^*)R_{k-1})\overline{\pi}_k^{-1}(\boldsymbol{\tau}^*)\{\boldsymbol{\phi}_k(\overline{\boldsymbol{W}}_{k-1}, \boldsymbol{\eta}^*) - \boldsymbol{\mu}\}\Bigg]\Bigg\} = C_1 - C_2.\end{aligned}$$

Under assumption of (2.2), $E[R_k|R_{k-1} = 1, \overline{\boldsymbol{W}}_{k-1}, \boldsymbol{Y}] = E[R_k|R_{k-1} = 1, \overline{\boldsymbol{W}}_{k-1}] = \pi_k$. After some simple calculation, we have $E[R_k|\overline{\boldsymbol{W}}_{k-1}] = \overline{\pi}_k$.

Furthermore, if (2.4) is true, then $\pi_k(\boldsymbol{\tau}^*) = \pi_k$. Thus, we have

$$
\begin{aligned}
C_1 &= E\left\{E\left[\tilde{\boldsymbol{X}}^T\boldsymbol{\Delta}\boldsymbol{V}^{-1}R_m\overline{\pi}_m^{-1}(\boldsymbol{\tau}^*)(\boldsymbol{Y}-\boldsymbol{\mu})|\overline{\boldsymbol{W}}_{m-1},\boldsymbol{Y}\right]\right\} \\
&= E\left[\tilde{\boldsymbol{X}}^T\boldsymbol{\Delta}\boldsymbol{V}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})\right] \\
&= 0
\end{aligned}
$$

and

$$
\begin{aligned}
C_2 = {} & E\left[\sum_{k=1}^{m-1}\tilde{\boldsymbol{X}}^T\boldsymbol{\Delta}\boldsymbol{V}^{-1}(R_k-\pi_k(\boldsymbol{\tau}^*)R_{k-1})\overline{\pi}_k^{-1}(\boldsymbol{\tau}^*)\{\boldsymbol{\phi}_k(\overline{\boldsymbol{W}}_{k-1},\boldsymbol{\eta}^*)-\boldsymbol{\mu}\}\right] \\
&+ E\left[\tilde{\boldsymbol{X}}^T\boldsymbol{\Delta}\boldsymbol{V}^{-1}E[R_m\pi_m^{-1}(\boldsymbol{\tau}^*)-R_{m-1}|\overline{\boldsymbol{W}}_{m-1},R_{m-1}=1]\overline{\pi}_{m-1}^{-1}(\boldsymbol{\tau}^*)\{\boldsymbol{\phi}_k(\overline{\boldsymbol{W}}_{m-1},\boldsymbol{\eta}^*)-\boldsymbol{\mu}\}\right] \\
&= E\left[\sum_{k=1}^{m-1}\tilde{\boldsymbol{X}}^T\boldsymbol{\Delta}\boldsymbol{V}^{-1}(R_k-\pi_k(\boldsymbol{\tau}^*)R_{k-1})\overline{\pi}_k^{-1}(\boldsymbol{\tau}^*)\{\boldsymbol{\phi}_k(\overline{\boldsymbol{W}}_{k-1},\boldsymbol{\eta}^*)-\boldsymbol{\mu}\}\right] \\
&\cdots \\
&= 0.
\end{aligned}
$$

If (2.8) is true, $\boldsymbol{\phi}_k(\overline{\boldsymbol{W}}_{k-1};\boldsymbol{\eta}^*) = E[\boldsymbol{Y}|\overline{\boldsymbol{W}}_{k-1}]$. Noting that we can rewrite the target equations as

$$
E\left\{\tilde{\boldsymbol{X}}^T\boldsymbol{\Delta}\boldsymbol{V}^{-1}\left[\{\boldsymbol{Y}-\boldsymbol{\mu}\} + \sum_{k=1}^{m}(R_k-\pi_k(\boldsymbol{\tau}^*)R_{k-1})\overline{\pi}_k^{-1}(\boldsymbol{\tau}^*)\{\boldsymbol{Y}-\boldsymbol{\phi}_k(\overline{\boldsymbol{W}}_{k-1},\boldsymbol{\eta}^*)\}\right]\right\} = C_1 + C_2.
$$

Obviously, $C_1 = 0$. On the other hand,

$$
\begin{aligned}
C_2 &= E\left[\sum_{k=1}^{m-1} \tilde{\boldsymbol{X}}^T \boldsymbol{\Delta} \boldsymbol{V}^{-1}(R_k - \pi_k(\boldsymbol{\tau}^*)R_{k-1})\overline{\pi}_k^{-1}(\boldsymbol{\tau}^*)\{\boldsymbol{Y} - \boldsymbol{\phi}_k(\overline{\boldsymbol{W}}_{k-1}, \boldsymbol{\eta}^*)\}\right] \\
&\quad + \tilde{\boldsymbol{X}}^T \boldsymbol{\Delta} \boldsymbol{V}^{-1}\left\{\pi_m \pi_m^{-1}(\boldsymbol{\tau}^*) - 1\right\}\overline{\pi}_{m-1}^{-1}(\boldsymbol{\tau}^*) E\left[\{\boldsymbol{Y} - \boldsymbol{\phi}_m(\overline{\boldsymbol{W}}_{m-1}, \boldsymbol{\eta}^*)|\overline{\boldsymbol{W}}_{k-1}\}\right] \\
&= E\left[\sum_{k=1}^{m-1} \tilde{\boldsymbol{X}}^T \boldsymbol{\Delta} \boldsymbol{V}^{-1}(R_k - \pi_k(\boldsymbol{\tau}^*)R_{k-1})\overline{\pi}_k^{-1}(\boldsymbol{\tau}^*)\{\boldsymbol{Y} - \boldsymbol{\phi}_k(\overline{\boldsymbol{W}}_{m-1}, \boldsymbol{\eta}^*)\}\right] \\
&\quad \cdots \\
&= 0.
\end{aligned}
$$

$\square$

3 Proof of Theorem 1

To simplify notation, we define μ_{ij} to be the function $\mu(\cdot)$ evaluated at the true parameter value and we can similarly define $\mu_{ij,t}^{(1)}$.

$$
\begin{aligned}
\sqrt{nh}\{\hat{\theta}(t,\hat{\boldsymbol{\beta}}) - \theta_0(t)\} &= \sqrt{nh}\{\hat{\theta}(t,\hat{\boldsymbol{\beta}}) - \hat{\theta}(t,\boldsymbol{\beta}_0)\} + \sqrt{nh}\{\hat{\theta}(t,\boldsymbol{\beta}_0) - \theta_0(t)\} \\
&= \sqrt{h}\left\{\frac{\partial\hat{\theta}(t,\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}\right\}\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \sqrt{nh}\{\hat{\theta}(t,\boldsymbol{\beta}_0) - \theta_0(t)\} + o_p(1).
\end{aligned}
$$

The asymptotic property of $\hat{\theta}(t,\hat{\boldsymbol{\beta}})$ is equivalent to the asymptotic property of $\hat{\theta}(t,\boldsymbol{\beta}_0)\}$ as long as $\hat{\beta}$ is $\sqrt{n}$-consistent and $\hat{\boldsymbol{\theta}}(\boldsymbol{t},\cdot)$ is continuous differentiable and its derivative is bounded in a neighborhood of $\boldsymbol{\beta}_0$. Let $\hat{\theta}_{[k]}(t,\boldsymbol{\beta}_0)$ be the estimator in the kth iteration. At initial, we ignore within-cluster correlation. Then, by expanding the first component of equations (2.6), we have

$$\hat{\theta}_{[0]}(t,\boldsymbol{\beta}_0)-\theta_0(t)=\frac{1}{2}b_{[0]}(t)h^2+W_2^{-1}(t)n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}K_h(T_{ij}-t)\mu_{ij,t}^{(1)}v_i^{jj}$$
$$\times\left[R_{im}\overline{\pi}_{im}^{*,-1}\{Y_{ij}-\mu_{ij}\}-\sum_{k=1}^{m}(R_{ik}-\pi_{ik}^{*}R_{i(k-1)})\overline{\pi}_{ik}^{-1}\{\boldsymbol{\phi}_k^{*}(\overline{\boldsymbol{W}}_{i(k-1)})-\mu_{ij}\}\right]$$
$$+o_p(h^2+\{\log(n)/nh\}^{1/2}+n^{-1/2}),$$

where $b_{[0]}(t)=\theta_0^{(2)}(t)$.

At the kth iteration, where $k\geq 1$, $\tilde{\theta}(\cdot,\boldsymbol{\beta}_0)$ is replaced by $\hat{\theta}_{[k-1]}(\cdot,\boldsymbol{\beta}_0)$. Define

$$b_{[k]}(t)=b_{[0]}(t)-\sum_{j=1}^{m}\sum_{l\neq j}^{m}E[\Delta_{jj}v^{jl}\Delta_{ll}b_{[k-1]}(T_l)|T_j=t]f_j(t),$$
$$Q(t,s)=\sum_{j=1}^{m}\sum_{l\neq j}^{m}E[\Delta_{jj}v^{jl}\Delta_{ll}W_2^{-1}(T_l)|T_j=t,T_l=s]f_{jl}(t,s),$$
$$\ddot{\mathcal{A}}(B;t,s)=-\sum_{j=1}^{m}\sum_{l\neq j}^{m}E[\Delta_{jj}v^{jl}\Delta_{ll}W_2^{-1}(T_l)B(T_l,s)|T_j=t]f_j(t),$$
$$Q_{1,[k]}(t,s)=-Q(t,s)+\ddot{\mathcal{A}}(Q_{1,[k-1]};t,s),$$

and

$$Q_{2,[k]}(t,s)=\ddot{\mathcal{A}}(Q_{2,[k-1]};t,s),$$

with $Q_{1,[1]}(t,s)=0$ and $Q_{2,[1]}(t,s)=-Q(t,s)$. After some calculation, we have

$$
\begin{aligned}
\hat{\theta}_{[k]}(t, \boldsymbol{\beta}_0) - \theta_0(t) = {} & \frac{1}{2} b_{[k]}(t) h^2 + W_2^{-1}(t) n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{l=1}^{m} K_h(T_{ij} - t) \mu_{ij,t}^{(1)} v_i^{jl} \\
& \times \left[R_{im} \overline{\pi}_{im}^{*,-1} \{Y_{il} - \mu_{il}\} - \sum_{k=1}^{m} (R_{ik} - \pi_{ik}^{*} R_{i(k-1)}) \overline{\pi}_{ik}^{*,-1} \{\phi_{kl}^{*}(\overline{\boldsymbol{W}}_{i(k-1)}) - \mu_{il}\} \right] \\
& + W_2^{-1}(t) n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{l=1}^{m} \mu_{ij,t}^{(1)} v_i^{jl} Q_{1,[k]}(t, T_{ij}) \\
& \times \left[R_{im} \overline{\pi}_{im}^{*,-1} \{Y_{il} - \mu_{il}\} - \sum_{k=1}^{m} (R_{ik} - \pi_{ik}^{*} R_{i(k-1)}) \overline{\pi}_{ik}^{*,-1} \{\phi_{kl}^{*}(\overline{\boldsymbol{W}}_{i(k-1)}) - \mu_{il}\} \right] \\
& - W_2^{-1}(t) n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \mu_{ij,t}^{(1)} v_i^{jj} Q_{2,[k]}(t, T_{ij}) \\
& \times \left[R_{im} \overline{\pi}_{im}^{*,-1} \{Y_{ij} - \mu_{ij}\} - \sum_{k=1}^{m} (R_{ik} - \pi_{ik}^{*} R_{i(k-1)}) \overline{\pi}_{ik}^{*,-1} \{\phi_{kj}^{*}(\overline{\boldsymbol{W}}_{i(k-1)}) - \mu_{ij}\} \right] \\
& + o_p(h^2 + \{\log(n)/nh\}^{1/2} + n^{-1/2}).
\end{aligned}
$$

At the convergence, we can obtain the asymptotic expansion of $\hat{\theta}(t, \boldsymbol{\beta}_0) - \theta_0(t)$ by replacing $b_{[k]}$, $Q_{1,[k]}$, and $Q_{2,[k]}$ with b^*, Q_1^*, and Q_2^*, where b^*, Q_1^*, and Q_2^* satisfy the following equations

$$
\begin{aligned}
b^*(t) &= b_{[0]}(t) - \sum_{j=1}^{m} \sum_{l \neq j}^{m} E[\Delta_{jj} v^{jl} \Delta_{ll} b^*(T_l) | T_j = t] f_j(t), \\
Q_1^*(t, s) &= -Q(t, s) + \ddot{\mathcal{A}}(Q_1^*; t, s),
\end{aligned}
$$

and

$$
Q_2^*(t, s) = \ddot{\mathcal{A}}(Q_2^*; t, s).
$$

Noting that all terms except $b^*(t)h^2/2$ are asymptotically 0 if either (2.4) or (2.8) is true.

4 Proof of Lemma 1

At the convergence, the first component of equations (2.6) is

$$
\begin{aligned}
0 = & \, n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} K_h(T_{ij} - t) \mu_{ij,t}^{(1)}(\hat{\boldsymbol{\alpha}}) \\
& \times \Bigg[v_i^{jj} R_{im} \overline{\pi}_{im}^{*,-1} \{ Y_{ij} - \mu[\boldsymbol{X}_{ij}^T \boldsymbol{\beta} + \hat{\theta}(t, \boldsymbol{\beta}) + \hat{\alpha}_1(t, \boldsymbol{\beta})(T_{ij} - t)/h] \} \\
& + \sum_{l \neq j}^{m} v_i^{jl} R_{im} \overline{\pi}_{im}^{*,-1} \{ Y_{il} - \mu[\boldsymbol{X}_{il}^T \boldsymbol{\beta} + \hat{\theta}(T_{il}, \boldsymbol{\beta})] \} \\
& - \sum_{k=1}^{m} v_i^{jj} (R_{ik} - \pi_{ik}^* R_{i(k-1)}) \overline{\pi}_{ik}^{*,-1} \{ \phi_{kj}^*(\overline{\boldsymbol{W}}_{i(k-1)}) - \mu[\boldsymbol{X}_{ij}^T \boldsymbol{\beta} + \hat{\theta}(t, \boldsymbol{\beta}) + \hat{\alpha}_1(t, \boldsymbol{\beta})(T_{ij} - t)/h] \} \\
& - \sum_{l \neq j}^{m} \sum_{k=1}^{m} v_i^{jl} (R_{ik} - \pi_{ik}^* R_{i(k-1)}) \overline{\pi}_{ik}^{*,-1} \{ \phi_{kl}^*(\overline{\boldsymbol{W}}_{i(k-1)}) - \mu[\boldsymbol{X}_{il}^T \boldsymbol{\beta} + \hat{\theta}(T_{il}, \boldsymbol{\beta})] \} \Bigg].
\end{aligned}
$$

Taking derivative with respect to $\boldsymbol{\beta}$ and evaluated at $\boldsymbol{\beta}_0$ on both sides. After some simple calculation, we obtain

$$
\begin{aligned}
0 &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}K_h(T_{ij}-t)\left\{\frac{\partial\mu_{ij,t}^{(1)}(\hat{\boldsymbol{\alpha}})}{\partial\boldsymbol{\beta}_0^T}\right\}\\
&\times\Bigg[v_i^{jj}R_{im}\overline{\pi}_{im}^{*,-1}\{Y_{ij}-\mu[\boldsymbol{X}_{ij}^T\boldsymbol{\beta}_0+\hat{\theta}(t,\boldsymbol{\beta}_0)+\hat{\alpha}_1(t,\boldsymbol{\beta}_0)(T_{ij}-t)/h]\}\\
&+\sum_{l\neq j}^{m}v_i^{jl}R_{im}\overline{\pi}_{im}^{*,-1}\{Y_{il}-\mu[\boldsymbol{X}_{il}^T\boldsymbol{\beta}_0+\hat{\theta}(T_{il},\boldsymbol{\beta}_0)]\}\\
&-\sum_{k=1}^{m}v_i^{jj}(R_{ik}-\pi_{ik}^{*}R_{i(k-1)})\overline{\pi}_{ik}^{*,-1}\{\phi_{kj}^{*}(\overline{\boldsymbol{W}}_{i(k-1)})-\mu[\boldsymbol{X}_{ij}^T\boldsymbol{\beta}_0+\hat{\theta}(t,\boldsymbol{\beta}_0)+\hat{\alpha}_1(t,\boldsymbol{\beta}_0)(T_{ij}-t)/h]\}\\
&-\sum_{l\neq j}^{m}\sum_{k=1}^{m}v_i^{jl}(R_{ik}-\pi_{ik}^{*}R_{i(k-1)})\overline{\pi}_{ik}^{*,-1}\{\phi_{kl}^{*}(\overline{\boldsymbol{W}}_{i(k-1)})-\mu[\boldsymbol{X}_{il}^T\boldsymbol{\beta}_0+\hat{\theta}(T_{il},\boldsymbol{\beta}_0)]\}\Bigg]\\
&-n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{m}K_h(T_{ij}-t)\mu_{ij,t}^{(1)}(\hat{\boldsymbol{\alpha}})v_i^{jl}\mu_{il,t}^{(1)}(\hat{\boldsymbol{\alpha}})\boldsymbol{X}_{il}\\
&-n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}K_h(T_{ij}-t)\mu_{ij,t}^{(1)}(\hat{\boldsymbol{\alpha}})v_i^{jj}\mu_{ij,t}^{(1)}\left\{\frac{\partial\hat{\alpha}_1(t,\boldsymbol{\beta}_0)}{\partial\beta^T}\right\}(T_{ij}-t)/h\\
&+n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}K_h(T_{ij}-t)\mu_{ij,t}^{(1)}(\hat{\boldsymbol{\alpha}})v_i^{jj}\mu_{ij,t}^{(1)}(\hat{\boldsymbol{\alpha}})\hat{\boldsymbol{\varphi}}(t)\\
&+n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l\neq j}^{m}K_h(T_{ij}-t)\mu_{ij,t}^{(1)}(\hat{\boldsymbol{\alpha}})v_i^{jl}\mu_{il,t}^{(1)}(\hat{\boldsymbol{\alpha}})\hat{\boldsymbol{\varphi}}(T_{il})\\
&= C_{1n}-C_{2n}-C_{3n}+C_{4n}+C_{5n}.
\end{aligned}
$$

When either (2.4) or (2.8) is true, C_{1n} is asymptotically 0. In addition, if $\hat{\theta}(\cdot)$ is consistent, we will have

$$C_{2n} = \sum_{j=1}^{m}\sum_{l=1}^{m} E[\Delta_{jj} v^{jl} \Delta_{ll} \boldsymbol{X}_l | T_j = t] f_j(t) + o_p(1),$$

$$C_{3n} = o_p(1),$$

$$C_{4n} = \sum_{j=1}^{m} E[\Delta_{jj} v^{jj} \Delta_{jj} | T_j = t] f_j(t) \hat{\boldsymbol{\varphi}}(t) + o_p(1),$$

and

$$C_{5n} = \sum_{j=1}^{m}\sum_{l \neq j}^{m} \int E[\Delta_{jj} v^{jl} \Delta_{ll} | T_j = t] \hat{\boldsymbol{\varphi}}(T_l) f_{lj}(T_l, t) dT_l + o_p(1).$$

Thus, we have

$$\sum_{j=1}^{m} E[\Delta_{jj} v^{jj} \Delta_{jj} | T_j = t] f_j(t) \hat{\boldsymbol{\varphi}}(t) - \sum_{j=1}^{m}\sum_{l=1}^{m} E[\Delta_{jj} v^{jl} \Delta_{ll} \boldsymbol{X}_l | T_j = t] f_j(t)$$
$$+ \sum_{j=1}^{m}\sum_{l \neq j}^{m} \int E[\Delta_{jj} v^{jl} \Delta_{ll} | T_j = t] \hat{\boldsymbol{\varphi}}(T_l) f_{lj}(T_l, t) dT_l = o_p(1).$$

Since $\hat{\boldsymbol{\varphi}}(t)$ uniformly converge to $\boldsymbol{\varphi}(t)$, we can show that these equations is equivalent to

$$\sum_{j=1}^{m}\sum_{l=1}^{m} E\left[\Delta_{jj} v^{jl} \Delta_{ll} \{\boldsymbol{X}_l - \boldsymbol{\varphi}(T_l)\} | T_j = t\right] f_j(t) = 0,$$

5 Proof of Theorem 2

It is easy to show that

$$n^{-1/2} \sum_{i=1}^{n} \boldsymbol{D}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}}) \boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}, \hat{\theta}) = n^{-1/2} \sum_{i=1}^{n} \boldsymbol{D}_i(\hat{\boldsymbol{\beta}}, \boldsymbol{\zeta}^*) \boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}, \hat{\theta}) + o_p(1).$$

Thus, at the convergence, equations (2.7) are equivalent to

$$n^{-1/2}\sum_{i=1}^{n}\boldsymbol{D}_i(\hat{\boldsymbol{\beta}},\boldsymbol{\zeta}^*)[\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\hat{\boldsymbol{\tau}},\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\beta}},\theta_0)+\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\hat{\boldsymbol{\tau}},\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\beta}},\hat{\theta})-\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\hat{\boldsymbol{\tau}},\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\beta}},\theta_0)]+o_p(1)=0.$$

We make Taylor expansion on the first term in the square bracket,

$$\begin{aligned}
0 &= n^{-1/2}\sum_{i=1}^{n}\boldsymbol{D}_i(\boldsymbol{\beta}_0,\boldsymbol{\zeta}^*)\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\boldsymbol{\tau}^*,\boldsymbol{\eta}^*,\boldsymbol{\beta}_0,\theta_0)\\
&+\frac{\partial}{\partial\boldsymbol{\beta}}\left\{n^{-1}\sum_{i=1}^{n}\boldsymbol{D}_i(\boldsymbol{\beta}_0,\boldsymbol{\zeta}^*)\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\boldsymbol{\tau}^*,\boldsymbol{\eta}^*,\boldsymbol{\beta}_0,\theta_0)\right\}\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta_0})\\
&+\frac{\partial}{\partial\boldsymbol{\tau}}\left\{n^{-1}\sum_{i=1}^{n}\boldsymbol{D}_i(\boldsymbol{\beta}_0,\boldsymbol{\zeta}^*)\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\boldsymbol{\tau}^*,\boldsymbol{\eta}^*,\boldsymbol{\beta}_0,\theta_0)\right\}\sqrt{n}(\hat{\boldsymbol{\tau}}-\boldsymbol{\tau^*})\\
&+\frac{\partial}{\partial\boldsymbol{\eta}}\left\{n^{-1}\sum_{i=1}^{n}\boldsymbol{D}_i(\boldsymbol{\beta}_0,\boldsymbol{\zeta}^*)\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\boldsymbol{\tau}^*,\boldsymbol{\eta}^*,\boldsymbol{\beta}_0,\theta_0)\right\}\sqrt{n}(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta^*})\\
&+n^{-1/2}\sum_{i=1}^{n}\boldsymbol{D}_i(\hat{\boldsymbol{\beta}},\boldsymbol{\zeta}^*)[\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\hat{\boldsymbol{\tau}},\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\beta}},\hat{\theta})-\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\hat{\boldsymbol{\tau}},\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\beta}},\theta_0)]+o_p(1).
\end{aligned}$$

Therefore, $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)=\boldsymbol{A}^{-1}(\boldsymbol{V})[C_{1n}+C_{2n}+C_{3n}+C_{4n}]$, where

$$\begin{aligned}
C_{1n} &= n^{-1/2}\sum_{i=1}^{n}\boldsymbol{D}_i(\boldsymbol{\beta}_0,\boldsymbol{\zeta}^*)\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\boldsymbol{\tau}^*,\boldsymbol{\eta}^*,\boldsymbol{\beta}_0,\theta_0)\\
C_{2n} &= E\left[\boldsymbol{D}(\boldsymbol{\beta}_0,\boldsymbol{\zeta}^*)\frac{\partial}{\partial\boldsymbol{\tau}}\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\boldsymbol{\tau}^*,\boldsymbol{\eta}^*,\boldsymbol{\beta}_0,\theta_0)\right]E\left[\frac{\partial}{\partial\boldsymbol{\tau}}\boldsymbol{S}(\boldsymbol{R},\boldsymbol{W}_{obs};\boldsymbol{\tau}^*)\right]^{-1}n^{-1/2}\sum_{i=1}^{n}\boldsymbol{S}(\boldsymbol{R}_i,\boldsymbol{W}_{obs,i};\boldsymbol{\tau}^*)\\
C_{3n} &= E\left[\boldsymbol{D}(\boldsymbol{\beta}_0,\boldsymbol{\zeta}^*)\frac{\partial}{\partial\boldsymbol{\eta}}\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\boldsymbol{\tau}^*,\boldsymbol{\eta}^*,\boldsymbol{\beta}_0,\theta_0)\right]E\left[\frac{\partial}{\partial\boldsymbol{\eta}}\boldsymbol{l}(\boldsymbol{W}_{obs};\boldsymbol{\eta}^*)\right]^{-1}n^{-1/2}\sum_{i=1}^{n}\boldsymbol{l}(\boldsymbol{W}_{obs,i};\boldsymbol{\eta}^*)\\
C_{4n} &= n^{-1/2}\sum_{i=1}^{n}\boldsymbol{D}_i(\hat{\boldsymbol{\beta}},\boldsymbol{\zeta}^*)[\mu\{\boldsymbol{X}_i\hat{\boldsymbol{\beta}}+\boldsymbol{\theta}_0(\boldsymbol{T}_i)\}-\mu\{\boldsymbol{X}_i\hat{\boldsymbol{\beta}}+\hat{\boldsymbol{\theta}}(\boldsymbol{T}_i,\hat{\boldsymbol{\beta}})\}]
\end{aligned}$$

By Lemma 1, if either (2.4) or (2.8) is true, $E[C_{1n}]=0$. $E[C_{2n}]=E[C_{3n}]=0$ follows from the fact that $E[\boldsymbol{S}(\boldsymbol{R},\boldsymbol{W}_{obs};\boldsymbol{\tau}^*)=E[\boldsymbol{l}(\boldsymbol{W}_{obs};\boldsymbol{\eta}^*)]=0$.

We remains to show that C_{4n} is order of $o_p(1)$. From the result of

Theorem 1, after some calculation, we can write $C_{4n} = -C_{41n} - C_{42n} -$

$C_{43n} - C_{44n} + o_p(1)$, where

$$C_{41n} = \frac{1}{2}(n^{1/2}h^2)n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{m}\tilde{\boldsymbol{X}}_{ij}\mu_{ij}^{(1)}v_i^{jl}\mu_{il}^{(1)}b^*(T_{il}) + O_p(n^{1/2}h^4),$$

$$\begin{aligned}C_{42n} &= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{m}\tilde{\boldsymbol{X}}_{ij}\mu_{ij}^{(1)}v_i^{jl}\mu_{il}^{(1)} \\ &\times\left\{W_2^{-1}(T_{il})n^{-1}\sum_{i'=1}^{n}\sum_{j'=1}^{m}\sum_{l'=1}^{m}K_h(T_{i'j'} - T_{il})\mu_{i'j',T_{il}}^{(1)}v_{i'}^{j'l'}\varepsilon_{i'l'}^*(\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}, \theta_0)\right\} \\ &= n^{-1/2}\sum_{i'=1}^{n}\sum_{j'=1}^{m}\sum_{l'=1}^{m}\mu_{i'j',t}^{(1)}v_{i'}^{j'l'}v_{i'}^{j'l'}\varepsilon_{i'l'}^*(\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}, \theta_0) \\ &\times\left\{\sum_{j=1}^{m}\sum_{l=1}^{m}E[\tilde{\boldsymbol{X}}_j\Delta_{jj}v_i^{jl}\Delta_{ll}W_2^{-1}(T_l)|T_l = t]f_l(t)|_{t=T_{i'j'}}\right\}\end{aligned}$$

$$\begin{aligned}C_{43n} &= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{m}\tilde{\boldsymbol{X}}_{ij}\mu_{ij}^{(1)}v_i^{jl}\mu_{il}^{(1)} \\ &\times\left\{W_2^{-1}(T_{il})n^{-1}\sum_{i'=1}^{n}\sum_{j'=1}^{m}\sum_{l'=1}^{m}\mu_{i'j',T_{il}}^{(1)}v_{i'}^{j'l'}Q_1^*(T_{il}, T_{i'j'})\varepsilon_{i'l'}^*(\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}, \theta_0)\right\} \\ &= n^{-1/2}\sum_{i'=1}^{n}\sum_{j'=1}^{m}\sum_{l'=1}^{m}\mu_{i'j',t}^{(1)}v_{i'}^{j'l'}v_{i'}^{j'l'}\varepsilon_{i'l'}^*(\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}, \theta_0) \\ &\times\left\{\sum_{j=1}^{m}\sum_{l=1}^{m}E[\tilde{\boldsymbol{X}}_j\Delta_{jj}v_i^{jl}\Delta_{ll}W_2^{-1}(T_l)Q_1^*(T_l, t)]|_{t=T_{i'j'}}\right\}f_l(t)\end{aligned}$$

$$\begin{aligned}C_{44n} &= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{m}\tilde{\boldsymbol{X}}_{ij}\mu_{ij}^{(1)}v_i^{jl}\mu_{il}^{(1)} \\ &\times\left\{W_2^{-1}(T_{il})n^{-1}\sum_{i'=1}^{n}\sum_{j'=1}^{m}\mu_{i'j',T_{il}}^{(1)}v_{i'}^{j'j'}Q_2^*(T_{il}, T_{i'j'})\varepsilon_{i'j'}^*(\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}, \theta_0)\right\} \\ &= n^{-1/2}\sum_{i'=1}^{n}\sum_{j'=1}^{m}\mu_{i'j',t}^{(1)}v_{i'}^{j'j'}v_{i'}^{j'l'}\varepsilon_{i'j'}^*(\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}, \theta_0) \\ &\times\left\{\sum_{j=1}^{m}\sum_{l=1}^{m}E[\tilde{\boldsymbol{X}}_j\Delta_{jj}v_i^{jl}\Delta_{ll}W_2^{-1}(T_l)Q_2^*(T_l, t)]|_{t=T_{i'j'}}\right\}f_l(t)\end{aligned}$$

As long as $\hat{\theta}$ is consistent and $nh^8 \to 0$, we will have $C_{41n} \to 0$. In addition, from the result of Lemma 1, it is easy to show that $C_{42n} - C_{44n}$ are all order of $o_p(1)$.

6 Proof of Corollary 1

Let $\boldsymbol{D}^*$, $\boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^*$, $\boldsymbol{S}_{\boldsymbol{\tau}}$, and $\boldsymbol{S}_{\boldsymbol{\eta}}$ be the shorts for $\boldsymbol{D}(\boldsymbol{\beta}_0, \boldsymbol{\zeta}^*)$, $\boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^*(\boldsymbol{\tau}^*, \boldsymbol{\eta}^*, \boldsymbol{\beta}_0, \theta_0)$, $\boldsymbol{S}(\boldsymbol{R}, \boldsymbol{W}_{obs}; \boldsymbol{\tau}^*)$, and $\boldsymbol{l}(\boldsymbol{W}_{obs}; \boldsymbol{\eta}^*)$. Proof of (b) and (d) are straightforward. We only provide proof of (a) and (c).

(a) It is easy to show that if (2.4) is true, $E[\boldsymbol{D}^* \partial \boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^* / \partial \boldsymbol{\tau}] = -E[\boldsymbol{D}^* \boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^* \boldsymbol{S}_{\boldsymbol{\tau}}]$, $E[\partial \boldsymbol{S}_{\boldsymbol{\tau}} / \partial \boldsymbol{\tau}] = -E[\boldsymbol{S}_{\boldsymbol{\tau}} \boldsymbol{S}_{\boldsymbol{\tau}}^T]$, and $\pi(\boldsymbol{\tau}^*) = \pi$. Thus $\boldsymbol{B}_{\boldsymbol{\phi}}(\boldsymbol{V})$ reduces to

$$\boldsymbol{B}_{\boldsymbol{\phi}}(\boldsymbol{V}) = Var\left\{ \boldsymbol{D}^* \boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^* - E[\boldsymbol{D}^* \boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^* \boldsymbol{S}_{\boldsymbol{\tau}}] E[\boldsymbol{S}_{\boldsymbol{\tau}} \boldsymbol{S}_{\boldsymbol{\tau}}^T]^{-1} \boldsymbol{S}_{\boldsymbol{\tau}} \right\}.$$

Instead, if true π_{ij} are known and are used in LAIPW method, the asymptotic variance of $\hat{\boldsymbol{\beta}}$ is $\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\phi}}(\boldsymbol{V}) = \boldsymbol{A}^{-1}(\boldsymbol{V}) \tilde{\boldsymbol{B}}_{\boldsymbol{\phi}}(\boldsymbol{V}) \boldsymbol{A}^{-1}(\boldsymbol{V})$ with $\tilde{\boldsymbol{B}}_{\boldsymbol{\phi}}(\boldsymbol{V}) = Var\{\boldsymbol{D}^* \boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^*\}$. We conclude that $\boldsymbol{B}_{\boldsymbol{\phi}}(\boldsymbol{V}) \leq \tilde{\boldsymbol{B}}_{\boldsymbol{\phi}}(\boldsymbol{V})$ since $\boldsymbol{D}^* \boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^* - E[\boldsymbol{D}^* \boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^* \boldsymbol{S}_{\boldsymbol{\tau}}] E[\boldsymbol{S}_{\boldsymbol{\tau}} \boldsymbol{S}_{\boldsymbol{\tau}}^T]^{-1} \boldsymbol{S}_{\boldsymbol{\tau}}$ is the residual from the population regression of $\boldsymbol{D}^* \boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^*$ on $\boldsymbol{S}_{\boldsymbol{\tau}}$.

(c) Noting that $\boldsymbol{A}(\boldsymbol{V})$ is not a function of $\boldsymbol{\phi}$, minimizing $\boldsymbol{\Omega}_{\boldsymbol{\phi}}(\boldsymbol{V})$ is equivalent to minimize $\boldsymbol{B}_{\boldsymbol{\phi}}(\boldsymbol{V})$.

Since (2.4) is true, we rewrite

$$E\left[\boldsymbol{D}(\boldsymbol{\beta}_0, \boldsymbol{\zeta}^*) \frac{\partial}{\partial \boldsymbol{\tau}} \boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^*(\boldsymbol{\tau}^*, \boldsymbol{\eta}^*, \boldsymbol{\beta}_0, \theta_0) \right] E\left[\frac{\partial}{\partial \boldsymbol{\tau}} \boldsymbol{S}(\boldsymbol{R}, \boldsymbol{W}_{obs}; \boldsymbol{\tau}^*) \right]^{-1} \boldsymbol{S}(\boldsymbol{R}, \boldsymbol{W}_{obs}; \boldsymbol{\tau}^*)$$

as $E[\boldsymbol{D}^*\boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^*\boldsymbol{S_\tau}]E[\boldsymbol{S_\tau}\boldsymbol{S_\tau}^T]^{-1}\boldsymbol{S_\tau}$, which is a projection of $\boldsymbol{D}^*\boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^*$ on $\boldsymbol{S_\tau}$. Thus, there exist a series of function $\boldsymbol{a}_k^*(\cdot)$ such that

$$E[\boldsymbol{D}^*\boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^*\boldsymbol{S_\tau}]E[\boldsymbol{S_\tau}\boldsymbol{S_\tau}^T]^{-1}\boldsymbol{S_\tau} = \sum_{k=1}^{m}(R_k - \pi_k R_{k-1})\boldsymbol{a}_k^*(\overline{\boldsymbol{W}}_{k-1}).$$

Define

$$\Lambda = \left\{\sum_{k=1}^{m}(R_k - \pi_k R_{k-1})\boldsymbol{a}_k(\overline{\boldsymbol{W}}_{k-1}) : \boldsymbol{a}_k(\cdot) \text{ are a series of arbitrary functions}\right\}.$$

Then, we claim that

$$\boldsymbol{D}^*\boldsymbol{\varepsilon}_{\boldsymbol{\phi}_{opt}}^* = R_m\overline{\pi}_m^{-1}\boldsymbol{D}^*\boldsymbol{\varepsilon} - \Pi\left[R_m\overline{\pi}_m^{-1}\boldsymbol{D}^*\boldsymbol{\varepsilon}|\Lambda\right]$$

because (i) $\boldsymbol{D}^*\boldsymbol{\varepsilon}_{\boldsymbol{\phi}_{opt}}^* - R_m\overline{\pi}_m^{-1}\boldsymbol{D}(\boldsymbol{\beta}_0, \boldsymbol{\zeta}^*)\boldsymbol{\varepsilon} \in \Lambda$ and (ii)

$$E\left[\left\{\boldsymbol{D}^*\boldsymbol{\varepsilon}_{\boldsymbol{\phi}_{opt}}^*\right\}^T\left\{\sum_{k=1}^{m}(R_k - \pi_k R_{k-1})\boldsymbol{a}_k(\overline{\boldsymbol{W}}_{k-1})\right\}\right] = 0$$

for any $\boldsymbol{a}_k(\cdot)$. (i) can be verified by definition. From a martingale covariance calculation, (ii) will hold if $\phi_{kl}(\overline{\boldsymbol{W}}_{i(k-1)};\boldsymbol{\eta}) = E[Y_{il}|\overline{\boldsymbol{W}}_{i(k-1)}]$ is correctly specified.

Since $E[\boldsymbol{D}^*\boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^*\boldsymbol{S_\tau}]E[\boldsymbol{S_\tau}\boldsymbol{S_\tau}^T]^{-1}\boldsymbol{S_\tau} \in \Lambda$, we have

$$\begin{aligned} Var\left\{\boldsymbol{D}^*\boldsymbol{\varepsilon}_{\boldsymbol{\phi}_{opt}}^*\right\} &= Var\left\{R_m\overline{\pi}_m^{-1}\boldsymbol{D}^*\boldsymbol{\varepsilon} - \Pi\left[R_m\overline{\pi}_m^{-1}\boldsymbol{D}^*\boldsymbol{\varepsilon}|\Lambda\right]\right\} \\ &\leq Var\left\{\boldsymbol{D}^*\boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^* - E[\boldsymbol{D}^*\boldsymbol{\varepsilon}_{\boldsymbol{\phi}}^*\boldsymbol{S_\tau}]E[\boldsymbol{S_\tau}\boldsymbol{S_\tau}^T]^{-1}\boldsymbol{S_\tau}\right\}. \end{aligned}$$

Table A1: Comparison of computational time (in seconds) among naive, LIPW-KPEE, and LAIPW-KPEE estimators based on 100 replications of the simulation with $n = 500$.

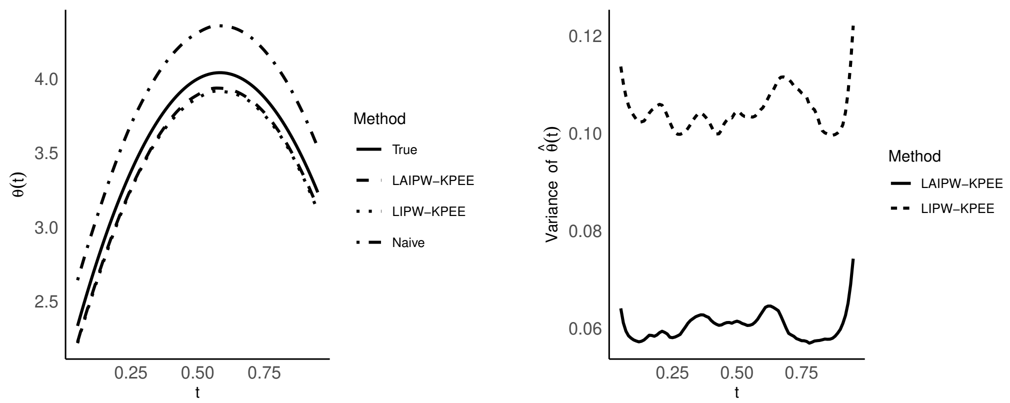
	Computational Time
Naive	55.13
LIPW-KPEE	
True π	21.26
Consistent $\hat{\pi}$	21.28
Wrong $\hat{\pi}$	21.85
LAIPW-KPEE	
True π and ϕ	25.79
Consistent $\hat{\pi}$ and $\hat{\phi}$	25.78
Wrong $\hat{\pi}$	25.46
Wrong $\hat{\phi}$	24.83
Both wrong	25.28

7 Additional Results for Simulation

7.1 Computational Time Comparison under $n = 500$

7.2 Performance Evaluation under Small Sample Size $n = 100$

We further conduct a simulation in the setting of small sample size with $n = 100$. As expected, a smaller sample size leads to increased variability, particularly in the estimation of $\theta(\cdot)$. Nevertheless, our proposed methods



(a) (b)

Figure A1: (a) Comparison of the true $\theta(t)$ and point-wise empirical mean of the nonparametric functions $\hat{\theta}(t)$ with $n = 100$. (b) Comparison of point-wise empirical variance of the nonparametric functions $\hat{\theta}(t)$ with $n = 100$.

still demonstrate reasonable bias control and maintain robustness across different model specifications. The results in Table A2, Figure A1, and Table A3 confirm that, even with a small sample size, the overall patterns observed remain consistent with the results for $n = 500$.

8 Additional Results for Application

The missing data model is constructed as follows:

$$\text{logit}\{\pi_{ij}\} = \boldsymbol{X}_{ij}^T \boldsymbol{\tau},$$

Table A2: Comparison of naive, LIPW-KPEE, and LAIPW-KPEE estimators in terms of bias, estimated standard error (EST S.E.), empirical standard error (EMP S.E.), and empirical mean squared error (EMP MSE) of $\hat{\boldsymbol{\beta}}$ and empirical mean integrated mean squared error (EMP MISE) of $\hat{\theta}(\cdot)$ based on 100 replications of the simulation with $n = 100$.

	$\boldsymbol{\beta_1 = 1}$				$\boldsymbol{\beta_2 = 4}$				$\boldsymbol{\theta(\cdot)}$
	Bias of $\hat{\beta}_1$	EST S.E.	EMP S.E.	EMP MSE	Bias of $\hat{\beta}_2$	EST S.E.	EMP S.E.	EMP MSE	EMP MISE
Naive	0.046	0.037	0.055	0.003	0.414	0.234	0.290	0.237	0.153
LIPW-KPEE									
True π	0.051	0.053	0.064	0.004	0.296	0.375	0.363	0.130	0.135
Consistent $\hat{\pi}$	0.050	0.052	0.063	0.004	0.284	0.337	0.355	0.125	0.119
Wrong $\hat{\pi}$	0.048	0.054	0.058	0.003	0.475	0.339	0.328	0.302	0.215
LAIPW-KPEE									
True π and ϕ	0.035	0.043	0.045	0.002	0.241	0.284	0.304	0.092	0.064
Consistent $\hat{\pi}$ and $\hat{\phi}$	0.051	0.056	0.065	0.004	0.246	0.297	0.306	0.092	0.072
Wrong $\hat{\pi}$	0.049	0.057	0.059	0.004	0.249	0.304	0.309	0.096	0.067
Wrong $\hat{\phi}$	0.092	0.112	0.126	0.016	0.265	0.322	0.323	0.104	0.122
Both wrong	0.061	0.072	0.078	0.006	0.407	0.309	0.303	0.224	0.247

Table A3: Comparison of computational time (in seconds) among naive, LIPW-KPEE, and LAIPW-KPEE estimators based on 100 replications of the simulation with $n = 100$.

	Computational Time
Naive	2.46
LIPW-KPEE	
True π	1.03
Consistent $\hat{\pi}$	1.04
Wrong $\hat{\pi}$	1.07
LAIPW-KPEE	
True π and ϕ	1.19
Consistent $\hat{\pi}$ and $\hat{\phi}$	1.19
Wrong $\hat{\pi}$	1.18
Wrong $\hat{\phi}$	1.16
Both wrong	1.17

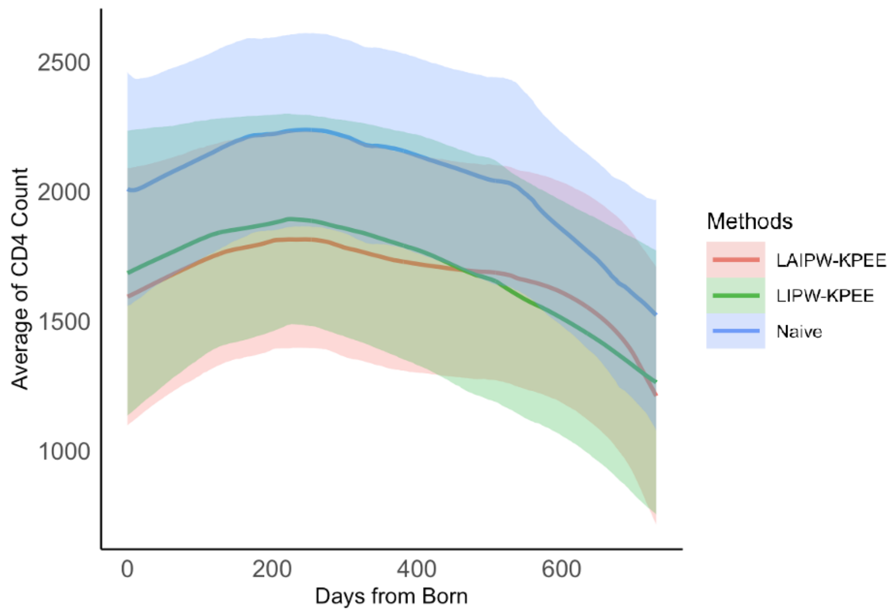
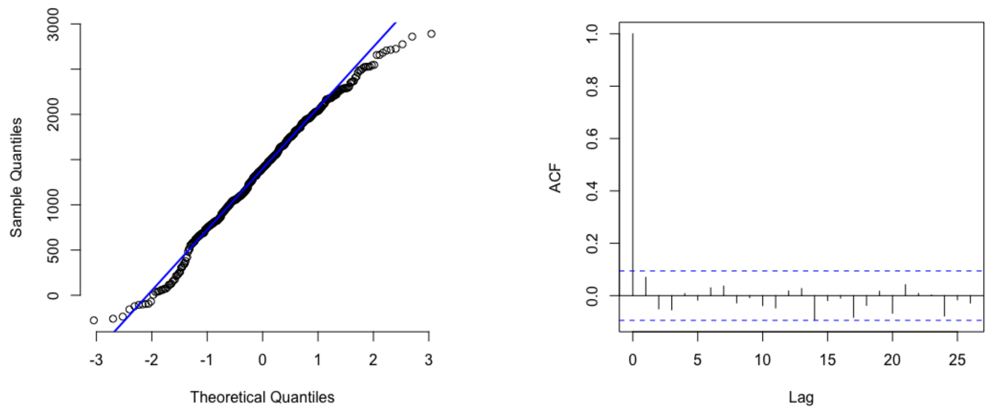


Figure A2: Average CD4 count with 95% confidence interval over time, adjusting for other covaraites, in LEOPARD study derived from naive method, LIPW-KPEE method, and LAIPW-KPEE method.

where $\boldsymbol{X}_{ij}$ consist of the function of sex, mean-centered birth weight, preterm, delivery mode, maternal prenatal ART history, mother's CD4 count, and breastfeeding status. For the conditional mean model, we adopt a similar approach but utilize a quadratic form of age. The curve of $\theta(\text{age})$ with 95% confidence interval is displayed in Figure A2.

In longitudinal studies, transforming the data during model diagnostics is essential to eliminate within-cluster correlation in the standard residu-

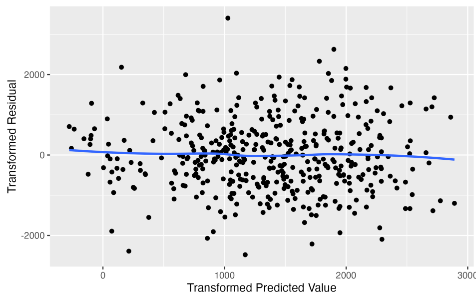


(a) (b)

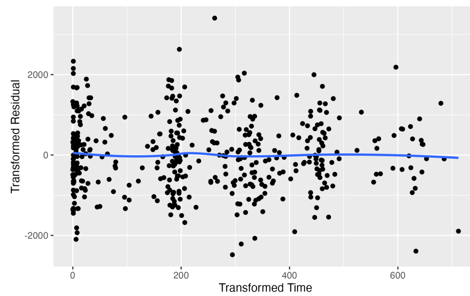
Figure A3: (a) Normal quantile plot of the transformed residuals. (b) Autocorrelation plot of the transformed residuals.

als. We achieve this with Cholesky decomposition, $\boldsymbol{V}_i = \boldsymbol{L}_i \boldsymbol{L} i^T$. Denote the residuals of the conditional mean model as $\boldsymbol{r}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i \hat{\boldsymbol{\eta}}$. we apply the transformation $\boldsymbol{r}_i^* = \boldsymbol{L}_i^{-1} \boldsymbol{r}_i$. Figure A3 demonstrates that the residuals closely approximate a normal distribution, with no autocorrelation. Additionally, Figure 1 shows no obvious systematic trend, and the variance remains homogeneous over time.

We assess the goodness of fit for the missing data model using the le Cessie–van Houwelingen–Copas–Hosmer unweighted sum of squares test (Hosmer et al., 1997). The p-value obtained from this test indicates that the model fits the data well.



(a)



(b)

Figure A4: Scatterplot of (a) the transformed residuals versus transformed predicted values, and (b) the transformed residuals versus transformed age.

References

Hosmer, D. W., T. Hosmer, S. Le Cessie, and S. Lemeshow (1997). A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in medicine 16*(9), 965–980.

Robins, J. M. and A. Rotnitzky (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS epidemiology*, pp. 297–331. Springer.

Robins, J. M. and A. Rotnitzky (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association 90*(429), 122–129.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the american statistical association 90*(429), 106–121.

Wang, N., R. J. Carroll, and X. Lin (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association 100*(469), 147–157.