Statistical Inference on Quantile Treatment Effect under Covariate-Adaptive Randomization

Yuhang Tao[1] and Li-Xin Zhang[2,1]

[1]*Zhejiang University and* [2]*Zhejiang Gongshang University*

Supplementary Material

S1 Proofs

S1.1 Proofs of Main Theorems

Proof of Theorem 1: Let $\boldsymbol{\delta} = (\delta_1, \delta_0)^\mathsf{T} \in \mathbb{R}^2$, $\boldsymbol{q^*(\tau)} = (q_1(\tau), q_0(\tau))^\mathsf{T}$, $\widehat{\boldsymbol{q}}^*(\boldsymbol{\tau}) = (\widehat{q}_1(\tau), \widehat{q}_0(\tau))^\mathsf{T}$, $\widetilde{\boldsymbol{T}}_{\boldsymbol{i}} = (T_i, 1 - T_i)^\mathsf{T}$, $i = 1, \cdots, n$, and

$$U_n(\boldsymbol{\delta}, \tau) = \sum_{i=1}^{n} \left[\rho_\tau \left\{ Y_i - \widetilde{\boldsymbol{T}}_{\boldsymbol{i}}^{\mathsf{T}} \boldsymbol{q}^*(\boldsymbol{\tau}) - \widetilde{\boldsymbol{T}}_{\boldsymbol{i}}^{\mathsf{T}} \boldsymbol{\delta}/\sqrt{n} \right\} - \rho_\tau \left\{ Y_i - \widetilde{\boldsymbol{T}}_{\boldsymbol{i}}^{\mathsf{T}} \boldsymbol{q}^*(\boldsymbol{\tau}) \right\} \right]. \tag{S1.1}$$

Then, by the change of variable, we have that

$$\sqrt{n} \left\{ \widehat{\boldsymbol{q}}^*(\boldsymbol{\tau}) - \boldsymbol{q}^*(\boldsymbol{\tau}) \right\} = \arg\min_{\boldsymbol{\delta}} U_n(\boldsymbol{\delta}, \tau). \tag{S1.2}$$

Using Knight's identity (Knight (1998)), we have

$$U_n(\boldsymbol{\delta},\tau) = -\boldsymbol{\delta}^{\mathsf{T}} U_{1n}(\tau) + U_{2n}(\boldsymbol{\delta},\tau),$$

where

$$\begin{aligned} U_{1n}(\tau) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{\boldsymbol{T}}_{\boldsymbol{i}} \left\{\tau - I(v_i \le 0)\right\}, \\ U_{2n}(\boldsymbol{\delta},\tau) &= \sum_{i=1}^{n} \int_0^{\widetilde{\boldsymbol{T}}_{\boldsymbol{i}}^{\mathsf{T}} \boldsymbol{\delta}/\sqrt{n}} \left\{I(v_i \le s) - I(v_i \le 0)\right\} ds, \end{aligned} \tag{S1.3}$$

and $v_i = Y_i - \widetilde{\boldsymbol{T}}_{\boldsymbol{i}}^{\mathsf{T}} \boldsymbol{q^*(\tau)}$.

First, we turn to $U_{2n}(\boldsymbol{\delta},\tau)$. In order to derive the asymptotic behavior of $U_{2n}(\boldsymbol{\delta},\tau)$, we can write

$$\begin{aligned} U_{2n}(\boldsymbol{\delta},\tau) =& \sum_{i=1}^{n} \int_0^{\widetilde{\boldsymbol{T}}_{\boldsymbol{i}}^{\mathsf{T}} \boldsymbol{\delta}/\sqrt{n}} \{I(v_i \le s) - I(v_i \le 0)\} ds \\ =& \sum_{i=1}^{n} T_i \int_0^{\delta_1/\sqrt{n}} \left[I\{Y_i(1) - q_1(\tau) \le s\} - I\{Y_i(1) - q_1(\tau) \le 0\}\right] ds \\ &+ \sum_{i=1}^{n} (1-T_i) \int_0^{\delta_0/\sqrt{n}} \left[I\{Y_i(0) - q_0(\tau) \le s\} - I\{Y_i(0) - q_0(\tau) \le 0\}\right] ds \\ =& \frac{1}{n} \sum_{i=1}^{n} T_i \int_0^{\delta_1} \sqrt{n} \left[I\left\{Y_i(1) - q_1(\tau) \le \frac{s}{\sqrt{n}}\right\} - I\left\{Y_i(1) - q_1(\tau) \le 0\right\}\right] ds \\ &+ \frac{1}{n} \sum_{i=1}^{n} (1-T_i) \int_0^{\delta_0} \sqrt{n} \left[I\left\{Y_i(0) - q_0(\tau) \le \frac{s}{\sqrt{n}}\right\} - I\left\{Y_i(0) - q_0(\tau) \le 0\right\}\right] ds \\ :=& U_{2n,1}(\delta_1,\tau) + U_{2n,0}(\delta_0,\tau). \end{aligned}$$

A simple calculation shows that

$$\begin{aligned}
&\mathbb{E}\left[\int_0^{\delta_1}\sqrt{n}\left[I\left\{Y_i(1)-q_1(\tau)\le\frac{s}{\sqrt{n}}\right\}-I\left\{Y_i(1)-q_1(\tau)\le 0\right\}\right]ds\right]\\
=&\int_0^{\delta_1}\sqrt{n}\left[F_1\left\{q_1(\tau)+\frac{s}{\sqrt{n}}\right\}-F_1\left\{q_1(\tau)\right\}\right]ds\\
=&\int_0^{\delta_1}f_1\left\{q_1(\tau)\right\}sds+o(1)\\
=&\frac{1}{2}f_1\left\{q_1(\tau)\right\}\delta_1^2+o(1).
\end{aligned}$$

Using the same method, we can derive that

$$\begin{aligned}
&\mathbb{E}\left[\int_0^{\delta_1}\sqrt{n}\left[I\left\{Y_i(1)-q_1(\tau)\le\frac{s}{\sqrt{n}}\right\}-I\{Y_i(1)-q_1(\tau)\le 0\}\right]ds\mid X_i\right]\\
=&\frac{1}{2}f_1\left\{q_1(\tau)|X\right\}\delta_1^2+o_P(1).
\end{aligned}$$

Besides, it is easy to note that

$$\mathbb{E}\left[\int_0^{\delta_1}\sqrt{n}\left[I\left\{Y_i(1)-q_1(\tau)\le\frac{s}{\sqrt{n}}\right\}-I\left\{Y_i(1)-q_1(\tau)\le 0\right\}\right]ds\mid X_i\right]$$

and $f_1\left\{q_1(\tau)\right\}$ are bounded, which implies that

$$\mathbb{E}\left[\int_0^{\delta_1}\sqrt{n}\left[I\left\{Y_i(1)-q_1(\tau)\le\frac{s}{\sqrt{n}}\right\}-I\left\{Y_i(1)-q_1(\tau)\le 0\right\}\right]ds\right]\xrightarrow{L^2}\frac{1}{2}f_1\left\{q_1(\tau)\right\}.$$

Then, it follows from Lemma S1.2, we have

$$\frac{1}{n}\sum_{i=1}^{n}(2T_i-1)\int_0^{\delta_1}\sqrt{n}\left[I\left\{Y_i(1)-q_1(\tau)\le\frac{s}{\sqrt{n}}\right\}-I\{Y_i(1)-q_1(\tau)\le 0\}\right]ds\xrightarrow{P}0.$$

Furthermore, by the law of large numbers for triangular arrays, we have

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^{\delta_1}\sqrt{n}\left[I\left\{Y_i(1)-q_1(\tau)\le\frac{s}{\sqrt{n}}\right\}-I\left\{Y_i(1)-q_1(\tau)\le 0\right\}\right]ds\xrightarrow{P}\frac{1}{2}f_1\left\{q_1(\tau)\right\}\delta_1^2.$$

Thus, we can derive that

$$\begin{aligned}U_{2n,1}(\delta_1,\tau) =&\frac{1}{2n}\sum_{i=1}^{n}(2T_i-1)\int_0^{\delta_1}\sqrt{n}\left[I\left\{Y_i(1)-q_1(\tau)\le\frac{s}{\sqrt{n}}\right\}-I\{Y_i(1)-q_1(\tau)\le 0\}\right]ds\\ &+\frac{1}{2n}\sum_{i=1}^{n}\int_0^{\delta_1}\sqrt{n}\left[I\left\{Y_i(1)-q_1(\tau)\le\frac{s}{\sqrt{n}}\right\}-I\{Y_i(1)-q_1(\tau)\le 0\}\right]ds\\ &\xrightarrow{P}\frac{1}{4}f_1\left\{q_1(\tau)\right\}\delta_1^2.\end{aligned}$$

Using the same method, we can derive that

$$U_{2n,0}(\delta_0,\tau)\xrightarrow{P}\frac{1}{4}f_0\left\{q_0(\tau)\right\}\delta_0^2.$$

Therefore, we obtain that

$$U_{2n}(\boldsymbol{\delta},\tau)\xrightarrow{P}\frac{1}{4}f_1\left\{q_1(\tau)\right\}\delta_1^2+\frac{1}{4}f_0\left\{q_0(\tau)\right\}\delta_0^2=\frac{1}{4}\boldsymbol{\delta}^{\mathsf{T}}\begin{pmatrix}f_1\{q_1(\tau)\} & 0\\ 0 & f_0\{q_0(\tau)\}\end{pmatrix}\boldsymbol{\delta}=\frac{1}{2}\boldsymbol{\delta}^{\mathsf{T}}\widetilde{Q}(\tau)\boldsymbol{\delta},$$

where

$$\widetilde{Q}(\tau)=\frac{1}{2}\begin{pmatrix}f_1\{q_1(\tau)\} & 0\\ 0 & f_0\{q_0(\tau)\}\end{pmatrix}.$$

Applying lemma S1.1, we have

$$\begin{aligned}&\sqrt{n}\left\{\widehat{\boldsymbol{q}}^*(\boldsymbol{\tau})-\boldsymbol{q}^*(\boldsymbol{\tau})\right\}\\ =&\{\widetilde{Q}(\tau)\}^{-1}U_{1n}(\tau)+o_P(1)\\ =&\frac{2}{\sqrt{n}}\begin{pmatrix}\dfrac{1}{f_1\{q_1(\tau)\}} & 0\\ 0 & \dfrac{1}{f_0\{q_0(\tau)\}}\end{pmatrix}\sum_{i=1}^{n}\begin{pmatrix}T_i\left[\tau-I\{Y_i(1)\le q_1(\tau)\}\right]\\ (1-T_i)\left[\tau-I\{Y_i(0)\le q_0(\tau)\}\right]\end{pmatrix}+o_P(1).\end{aligned}$$

Thus, we obtain that

$$\sqrt{n}\left\{\widehat{q}(\tau)-q(\tau)\right\}=\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{1}{f_1\{q_1(\tau)\}}T_i\left(\tau-I\{Y_i(1)\le q_1(\tau)\}\right)\right.$$
$$\left.-\frac{1}{f_0\{q_0(\tau)\}}(1-T_i)\left(\tau-I\{Y_i(0)\le q_0(\tau)\}\right)\right]+o_P(1). \quad \text{(S1.4)}$$

By simple decomposition, we have

$$\begin{aligned}&\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{1}{f_1\{q_1(\tau)\}}T_i\left(\tau-I\{Y_i(1)\le q_1(\tau)\}\right)\right.\\&\left.-\frac{1}{f_0\{q_0(\tau)\}}(1-T_i)\left(\tau-I\{Y_i(0)\le q_0(\tau)\}\right)\right]\\=&\widetilde{\xi}_{n,1}(\tau)+\widetilde{\xi}_{n,2}(\tau)+\widetilde{\xi}_{n,3}(\tau), \quad \text{(S1.5)}\end{aligned}$$

where

$$\begin{aligned}\widetilde{\xi}_{n,1}(\tau)=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(2T_i-1)\left[\frac{\tau-\mathbb{P}\{Y_i(1)\le q_1(\tau)|X_i\}}{f_1\{q_1(\tau)\}}\right.\\&\left.+\frac{\tau-\mathbb{P}\{Y_i(0)\le q_0(\tau)|X_i\}}{f_0\{q_0(\tau)\}}\right], \quad \text{(S1.6)}\\\widetilde{\xi}_{n,2}(\tau)=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\tau-\mathbb{P}\{Y_i(1)\le q_1(\tau)|X_i\}}{f_1\{q_1(\tau)\}}-\frac{\tau-\mathbb{P}\{Y_i(0)\le q_0(\tau)|X_i\}}{f_0\{q_0(\tau)\}}\right], \quad \text{(S1.7)}\\\widetilde{\xi}_{n,3}(\tau)=&\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left[T_i\frac{\mathbb{P}\{Y_i(1)\le q_1(\tau)\mid X_i\}-I\{Y_i(1)\le q_1(\tau)\}}{f_1\{q_1(\tau)\}}\right.\\&\left.-(1-T_i)\frac{\mathbb{P}\{Y_i(0)\le q_0(\tau)\mid X_i\}-I\{Y_i(0)\le q_0(\tau)\}}{f_0\{q_0(\tau)\}}\right]. \quad \text{(S1.8)}\end{aligned}$$

It follows from Assumption 2 that there exists a $\sigma_1^2(\tau)$ such that

$$\widetilde{\xi}_{n,1}(\tau)\xrightarrow{d}\mathcal{N}(0,\sigma_1^2(\tau)).$$

Notice that

$$\mathbb{E}\left[\frac{\tau-\mathbb{P}\{Y_i(1)\le q_1(\tau)\mid X_i\}}{f_1\{q_1(\tau)\}}-\frac{\tau-\mathbb{P}\{Y_i(0)\le q_0(\tau)\mid X_i\}}{f_0\{q_0(\tau)\}}\right]=0,$$

and

$$\begin{aligned}\mathbb{E}\left[\frac{\mathbb{P}\{Y_i(1)\le q_1(\tau)\mid X_i\}-I\{Y_i(1)\le q_1(\tau)\}}{f_1\{q_1(\tau)\}}\middle| X_i\right]&=0,\\ \mathbb{E}\left[\frac{\mathbb{P}\{Y_i(0)\le q_0(\tau)\mid X_i\}-I\{Y_i(0)\le q_0(\tau)\}}{f_0\{q_0(\tau)\}}\middle| X_i\right]&=0.\end{aligned}$$

Furthermore, we denote that

$$\sigma_2^2(\tau)=\mathrm{Var}\left[\frac{\tau-\mathbb{P}\left\{Y_i(1)\le q_1(\tau)\mid X_i\right\}}{f_1\{q_1(\tau)\}}-\frac{\tau-\mathbb{P}\left\{Y_i(0)\le q_0(\tau)\mid X_i\right\}}{f_0\{q_0(\tau)\}}\right],$$

and

$$\begin{aligned}\sigma_3^2(\tau)=&\ \mathrm{Var}\left[\frac{\mathbb{P}\left\{Y_i(1)\le q_1(\tau)\mid X_i\right\}-I\{Y_i(1)\le q_1(\tau)\}}{f_1\{q_1(\tau)\}}\right]\\ &+\mathrm{Var}\left[\frac{\mathbb{P}\left\{Y_i(0)\le q_0(\tau)\mid X_i\right\}-I\{Y_i(0)\le q_0(\tau)\}}{f_0\{q_0(\tau)\}}\right]\\ =&\ \frac{1}{f_1\{q_1(\tau)\}^2}\mathbb{E}\left[\mathbb{P}\left\{Y_i(1)\le q_1(\tau)\mid X_i\right\}\left(1-\mathbb{P}\left\{Y_i(1)\le q_1(\tau)\mid X_i\right\}\right)\right]\\ &+\frac{1}{f_0\{q_0(\tau)\}^2}\mathbb{E}\left[\mathbb{P}\left\{Y_i(0)\le q_0(\tau)\mid X_i\right\}\left(1-\mathbb{P}\left\{Y_i(0)\le q_0(\tau)\mid X_i\right\}\right)\right].\end{aligned}$$

Therefore, it follows from Assumption 2 that

$$\widetilde{\xi}_{n,1}(\tau)+\widetilde{\xi}_{n,2}(\tau)+\widetilde{\xi}_{n,3}(\tau)\xrightarrow{d}\mathcal{N}(0,\sigma_1^2(\tau)+\sigma_2^2(\tau)+\sigma_3^2(\tau)),$$

which implies that

$$\sqrt{n}\left\{\widehat{q}(\tau)-q(\tau)\right\}\xrightarrow{d}\mathcal{N}(0,\sigma_1^2(\tau)+\sigma_2^2(\tau)+\sigma_3^2(\tau)).$$

□

Proof of Theorem 2: Let

$$L_{1,n}(\delta_1,\tau,\boldsymbol{\psi})=\sum_{i=1}^{n}T_i\left[\rho_\tau\left\{Y_i-q_1(\tau)-\frac{\delta_1}{\sqrt{n}}\right\}-\rho_\tau\left\{Y_i-q_1(\tau)\right\}+\frac{\delta_1}{\sqrt{n}}\boldsymbol{\lambda}_1^{\mathsf{T}}\boldsymbol{\psi}(X_i)\right].$$

Then, by the change of variable, we have that

$$\sqrt{n}\left\{\widehat{q}^{\star}_{1,adj}(\tau,\boldsymbol{\lambda}_1,\boldsymbol{\psi})-q_1(\tau)\right\}=\arg\min_{\delta_1}\ L_{1,n}(\delta_1,\tau,\boldsymbol{\psi}).$$

Using Knight's identity (Knight (1998)), we have

$$L_{1,n}(\delta_1,\tau,\boldsymbol{\psi})=-\delta_1 L_{1,n,1}(\tau,\boldsymbol{\psi})+L_{1,n,2}(\delta_1,\tau),$$

where

$$L_{1,n,1}(\tau,\boldsymbol{\psi})=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}T_i\left[\tau-I\{Y_i(1)\le q_1(\tau)\}-\boldsymbol{\lambda}_1^{\mathsf{T}}\boldsymbol{\psi}(X_i)\right],$$

and

$$L_{1,n,2}(\delta_1,\tau)=\sum_{i=1}^{n}T_i\int_0^{\delta_1/\sqrt{n}}\left[I\{Y_i(1)\le q_1(\tau)+s\}-I\{Y_i(1)\le q_1(\tau)\}\right]ds.$$

It has been proved in proof of Theorem 1 that $L_{1,n,2}(\delta_1,\tau)=\frac{1}{4}f_1\{q_1(\tau)\}\delta_1^2+o_P(1)$. Then it follows from lemma S1.1 that

$$\begin{aligned}&\sqrt{n}\left\{\widehat{q}^{\star}_{1,adj}(\tau,\boldsymbol{\lambda}_1,\boldsymbol{\psi})-q_1(\tau)\right\}\\&=\frac{2}{f_1\{q_1(\tau)\}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}T_i\left[\tau-I\{Y_i(1)\le q_1(\tau)\}-\boldsymbol{\lambda}_1^{\mathsf{T}}\boldsymbol{\psi}(X_i)\right]+o_P(1).\end{aligned}$$

Similarly, we can obtain that

$$\begin{aligned}&\sqrt{n}\left\{\widehat{q}^{\star}_{0,adj}(\tau,\boldsymbol{\lambda}_0,\boldsymbol{\psi})-q_0(\tau)\right\}\\&=\frac{2}{f_0\{q_0(\tau)\}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-T_i)\left[\tau-I\{Y_i(0)\le q_0(\tau)\}-\boldsymbol{\lambda}_0^{\mathsf{T}}\boldsymbol{\psi}(X_i)\right]+o_P(1).\end{aligned}$$

Then we have

$$
\begin{aligned}
&\sqrt{n}\left\{\widehat{q}_{adj}(\tau,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_0,\boldsymbol{\psi})-q(\tau)\right\}\\
&=\widetilde{\xi}_{n,1}(\tau)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(2T_i-1)\left[\frac{\boldsymbol{\lambda}_1^{\mathsf{T}}\boldsymbol{\psi}(X_i)}{f_1\{q_1(\tau)\}}+\frac{\boldsymbol{\lambda}_0^{\mathsf{T}}\boldsymbol{\psi}(X_i)}{f_0\{q_0(\tau)\}}\right]+\widetilde{\xi}_{n,2}(\tau)+\widetilde{\xi}_{n,3}(\tau)\\
&\quad-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\boldsymbol{\lambda}_1^{\mathsf{T}}\boldsymbol{\psi}(X_i)}{f_1\{q_1(\tau)\}}-\frac{\boldsymbol{\lambda}_0^{\mathsf{T}}\boldsymbol{\psi}(X_i)}{f_0\{q_0(\tau)\}}\right]\\
&:=\widetilde{\xi}_{n,3}(\tau)+\widetilde{\xi}_{n,4}(\tau,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_0,\boldsymbol{\psi})+\widetilde{\xi}_{n,5}(\tau,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_0,\boldsymbol{\psi}),
\end{aligned}
$$

where

$$
\widetilde{\xi}_{n,4}(\tau,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_0,\boldsymbol{\psi})=\widetilde{\xi}_{n,2}(\tau)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\boldsymbol{\lambda}_1^{\mathsf{T}}\boldsymbol{\psi}(X_i)}{f_1\{q_1(\tau)\}}-\frac{\boldsymbol{\lambda}_0^{\mathsf{T}}\boldsymbol{\psi}(X_i)}{f_0\{q_0(\tau)\}}\right],
$$

and

$$
\begin{aligned}
\widetilde{\xi}_{n,5}(\tau,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_0,\boldsymbol{\psi})&=\widetilde{\xi}_{n,1}(\tau)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(2T_i-1)\left[\frac{\boldsymbol{\lambda}_1^{\mathsf{T}}\boldsymbol{\psi}(X_i)}{f_1\{q_1(\tau)\}}+\frac{\boldsymbol{\lambda}_0^{\mathsf{T}}\boldsymbol{\psi}(X_i)}{f_0\{q_0(\tau)\}}\right]\\
&=\widetilde{\xi}_{n,1}(\tau)+o_P(1).
\end{aligned}
$$

For convenience, we write

$$
\widetilde{\xi}_{n,4}(\tau,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_0,\boldsymbol{\psi})=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{\xi}_{4,i}(\tau,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_0,\boldsymbol{\psi}),
$$

where for $i=1,\cdots,n$,

$$
\begin{aligned}
\widetilde{\xi}_{4,i}(\tau,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_0,\boldsymbol{\psi})=&\frac{\tau-\mathbb{P}\{Y_i(1)\le q_1(\tau)\mid X_i\}}{f_1\{q_1(\tau)\}}-\frac{\tau-\mathbb{P}\{Y_i(0)\le q_0(\tau)\mid X_i\}}{f_0\{q_0(\tau)\}}\\
&-\left[\frac{\boldsymbol{\lambda}_1^{\mathsf{T}}\boldsymbol{\psi}(X_i)}{f_1\{q_1(\tau)\}}-\frac{\boldsymbol{\lambda}_0^{\mathsf{T}}\boldsymbol{\psi}(X_i)}{f_0\{q_0(\tau)\}}\right].
\end{aligned}
$$

Note that $\mathbb{E}\{\widetilde{\xi}_{4,i}(\tau, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_0, \boldsymbol{\psi})\} = 0$ and

$$
\begin{aligned}
&\mathrm{Var}\left[\widetilde{\xi}_{4,i}(\tau, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_0, \boldsymbol{\psi})\right] \\
&= \sigma_2^2(\tau) + \mathrm{Var}\left[\frac{\boldsymbol{\lambda}_1^{\mathsf{T}}\boldsymbol{\psi}(X_i)}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{\lambda}_0^{\mathsf{T}}\boldsymbol{\psi}(X_i)}{f_0\{q_0(\tau)\}}\right] \\
&\quad - \frac{2}{f_1\{q_1(\tau)\}}\left[\frac{\boldsymbol{\lambda}_1}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{\lambda}_0}{f_0\{q_0(\tau)\}}\right]^{\mathsf{T}} \mathrm{Cov}\left[\tau - \mathbb{P}\{Y_i(1) \le q_1(\tau)|X_i\}, \boldsymbol{\psi}(X_i)\right] \\
&\quad + \frac{2}{f_0\{q_0(\tau)\}}\left[\frac{\boldsymbol{\lambda}_1}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{\lambda}_0}{f_0\{q_0(\tau)\}}\right]^{\mathsf{T}} \mathrm{Cov}\left[\tau - \mathbb{P}\{Y_i(0) \le q_0(\tau)|X_i\}, \boldsymbol{\psi}(X_i)\right] \\
&= \sigma_2^2(\tau) + \left[\frac{\boldsymbol{\lambda}_1}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{\lambda}_0}{f_0\{q_0(\tau)\}}\right]^{\mathsf{T}} \boldsymbol{\Sigma}_{\boldsymbol{\psi}\boldsymbol{\psi}} \left[\frac{\boldsymbol{\lambda}_1}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{\lambda}_0}{f_0\{q_0(\tau)\}}\right] \\
&\quad - 2\left[\frac{\boldsymbol{\lambda}_1}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{\lambda}_0}{f_0\{q_0(\tau)\}}\right]^{\mathsf{T}} \left[\frac{\boldsymbol{b}_1(\tau, \boldsymbol{\psi})}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{b}_0(\tau, \boldsymbol{\psi})}{f_0\{q_0(\tau)\}}\right], \\
&= \sigma_2^2(\tau) + V(\tau, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_0, \boldsymbol{\psi}),
\end{aligned}
$$

where $\boldsymbol{b}_a(\tau, \boldsymbol{\psi}) = \mathbb{E}[(\tau - I\{Y_i(a) \le q_a(\tau)\})\boldsymbol{\psi}(X_i)]$ for $a = 0, 1$.

Therefore, it follows from Assumption 2 that

$$
\widetilde{\xi}_{n,3}(\tau) + \widetilde{\xi}_{n,4}(\tau, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_0, \boldsymbol{\psi}) + \widetilde{\xi}_{n,5}(\tau, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_0, \boldsymbol{\psi}) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\tau) + V(\tau, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_0, \boldsymbol{\psi})),
$$

which implies that

$$
\sqrt{n}\left\{\widehat{q}_{adj}(\tau, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_0, \boldsymbol{\psi}) - q(\tau)\right\} \xrightarrow{d} \mathcal{N}(0, \sigma^2(\tau) + V(\tau, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_0, \boldsymbol{\psi})).
$$

Next, we investigate for which parameter values of $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_0$, $V(\tau, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_0, \boldsymbol{\psi})$ reaches its minimum. Let

$$
\boldsymbol{\omega} = \frac{\boldsymbol{\lambda}_1}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{\lambda}_0}{f_0\{q_0(\tau)\}},
$$

then we have

$$\frac{\partial V}{\partial \boldsymbol{\omega}} = 2\boldsymbol{\Sigma}_{\boldsymbol{\psi}\boldsymbol{\psi}}\boldsymbol{\omega} - 2\left[\frac{\boldsymbol{b}_1(\tau,\boldsymbol{\psi})}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{b}_0(\tau,\boldsymbol{\psi})}{f_0\{q_0(\tau)\}}\right].$$

Therefore, $V(\tau, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_0, \boldsymbol{\psi})$ attains its minimal value when the following equation holds.

$$\boldsymbol{\Sigma}_{\boldsymbol{\psi}\boldsymbol{\psi}}\left[\frac{\boldsymbol{\lambda}_1}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{\lambda}_0}{f_0\{q_0(\tau)\}}\right] = \frac{\boldsymbol{b}_1(\tau,\boldsymbol{\psi})}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{b}_0(\tau,\boldsymbol{\psi})}{f_0\{q_0(\tau)\}}. \tag{S1.9}$$

Under this situation

$$V_{min}(\tau,\boldsymbol{\psi}) = -\left[\frac{\boldsymbol{b}_1(\tau,\boldsymbol{\psi})}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{b}_0(\tau,\boldsymbol{\psi})}{f_0\{q_0(\tau)\}}\right]^{\mathsf{T}} \boldsymbol{\Sigma}^{+}_{\boldsymbol{\psi}\boldsymbol{\psi}}\left[\frac{\boldsymbol{b}_1(\tau,\boldsymbol{\psi})}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{b}_0(\tau,\boldsymbol{\psi})}{f_0\{q_0(\tau)\}}\right],$$

and this minimum is achieved when

$$\frac{\boldsymbol{\lambda}_1}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{\lambda}_0}{f_0\{q_0(\tau)\}} = \boldsymbol{\Sigma}^{+}_{\boldsymbol{\psi}\boldsymbol{\psi}}\left[\frac{\boldsymbol{b}_1(\tau,\boldsymbol{\psi})}{f_1\{q_1(\tau)\}} - \frac{\boldsymbol{b}_0(\tau,\boldsymbol{\psi})}{f_0\{q_0(\tau)\}}\right].$$

□

Proof of Theorem 3: First, we prove conclusion (i). For convenience we write $\boldsymbol{\beta} = \boldsymbol{b}_1(\tau,\boldsymbol{\psi}_{DB})/f_1\{q_1(\tau)\} - \boldsymbol{b}_0(\tau,\boldsymbol{\psi}_{DB})/f_0\{q_0(\tau)\}$. Note that if $\Sigma_{\boldsymbol{\psi}\boldsymbol{\psi},DB}$ is nonsingular, then

$$\begin{aligned} V_{min}(\tau,\boldsymbol{\psi}^*) =& -\boldsymbol{\beta}^{\mathsf{T}}H^{\mathsf{T}}\left(H\boldsymbol{\Sigma}_{\boldsymbol{\psi}\boldsymbol{\psi},DB}H^{\mathsf{T}}\right)^{-1}H\boldsymbol{\beta} = -\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi},DB}\boldsymbol{\beta} \\ &= V_{min}(\tau,\boldsymbol{\psi}_{DB}). \end{aligned}$$

Next, we turn to conclusion (ii). Let $\boldsymbol{\Sigma}^{1/2}_{\boldsymbol{\psi}\boldsymbol{\psi},DB}$ be the square root of $\boldsymbol{\Sigma}_{\boldsymbol{\psi}\boldsymbol{\psi},DB}$. Define $B = H\boldsymbol{\Sigma}^{1/2}_{\boldsymbol{\psi}\boldsymbol{\psi},DB}$ and $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\Sigma}^{-1/2}_{\boldsymbol{\psi}\boldsymbol{\psi},DB}\boldsymbol{\beta}$, where $\boldsymbol{\Sigma}^{-1/2}_{\boldsymbol{\psi}\boldsymbol{\psi},DB}$ denotes the

square root of $\boldsymbol{\Sigma}^{+}_{\boldsymbol{\psi}\boldsymbol{\psi},DB}$. Thus,

$$\begin{aligned} V_{min}(\tau,\boldsymbol{\psi}^*) =& -\widetilde{\boldsymbol{\beta}}^{\mathsf{T}}(B^{\mathsf{T}}(BB^{\mathsf{T}})^{-1}B)\widetilde{\boldsymbol{\beta}} \geq -\left\|\widetilde{\boldsymbol{\beta}}\right\|_2^2 = -\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Sigma}^{+}_{\boldsymbol{\psi}\boldsymbol{\psi},DB}\boldsymbol{\beta} \\ &= V_{min}(\tau,\boldsymbol{\psi}_{DB}). \end{aligned}$$

The above inequality is satified for any $\boldsymbol{\psi}^* \in \Psi$, thus we prove the conclusion (ii). □

Proof of Theorem 4: By equation (S1.4), we obtain that

$$\begin{aligned} \widehat{\sigma}^2_{mb}(\tau) &= \frac{\ell}{n-\ell+1}\sum_{i=0}^{n-\ell}\left\{\widehat{q}_{i,\ell,mb}(\tau) - \bar{q}_{mb}(\tau)\right\}^2 \\ &= \frac{\ell}{n-\ell+1}\sum_{i=0}^{n-\ell}\widehat{q}^2_{i,\ell,mb}(\tau) - \ell\bar{q}^2_{mb}(\tau) \\ &= \frac{\ell}{n-\ell+1}\sum_{i=0}^{n-\ell}\left\{\widehat{q}_{i,\ell,mb}(\tau) - q(\tau)\right\}^2 - \ell\left\{\bar{q}_{mb}(\tau) - q(\tau)\right\}^2 \\ &= \frac{\ell}{n-\ell+1}\sum_{i=0}^{n-\ell}\left\{\frac{1}{\ell}\sum_{j=i+1}^{i+\ell}\breve{\xi}_j + o_P(\ell^{-1/2})\right\}^2 - \ell\left\{\bar{q}_{mb}(\tau) - q(\tau)\right\}^2 \\ &= \frac{1}{(n-\ell+1)\ell}\sum_{i=0}^{n-\ell}\left(\sum_{j=i+1}^{i+\ell}\breve{\xi}_j\right)^2 - \ell\left\{\bar{q}_{mb}(\tau) - q(\tau)\right\}^2 + o_P(1), \end{aligned}$$

where

$$\begin{aligned} \breve{\xi}_j =& \frac{2}{f_1\{q_1(\tau)\}}T_j\left[\tau - I\{Y_j(1) \leq q_1(\tau)\}\right] \\ &- \frac{2}{f_0\{q_0(\tau)\}}(1-T_j)\left[\tau - I\{Y_j(0) \leq q_0(\tau)\}\right]. \end{aligned}$$

According to Assumption 5 and equation (S1.5),

$$\frac{1}{(n-\ell+1)\ell}\sum_{i=0}^{n-\ell}\left(\sum_{j=i+1}^{i+\ell}\breve{\xi}_j\right)^2 = \frac{n}{n-\ell+1}\frac{1}{n\ell}\sum_{i=0}^{n-\ell}\left(\sum_{j=i+1}^{i+\ell}\breve{\xi}_j\right)^2 \xrightarrow{P} \sigma^2(\tau).$$

Moreover, it follows from equation (S1.4) and Assumption 2 that

$$\begin{aligned}\sqrt{\ell}\left\{\bar{q}_{mb}(\tau)-q(\tau)\right\} &= \frac{\sqrt{\ell}}{n-\ell+1}\sum_{i=0}^{n-\ell}\left\{\widehat{q}_{i,\ell,mb}(\tau)-q(\tau)\right\} \\ &= \frac{1}{(n-\ell+1)\sqrt{\ell}}\sum_{i=0}^{n-\ell}\sum_{j=i+1}^{i+\ell}\breve{\xi}_j + o_P(1) \\ &= \frac{\sqrt{\ell}}{n-\ell+1}\sum_{j=1}^{n}\breve{\xi}_j + o_P(1) \\ &= \frac{\sqrt{n\ell}}{n-\ell+1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\breve{\xi}_j + o_P(1) \xrightarrow{P} 0,\end{aligned}$$

which implies that

$$\ell\left\{\bar{q}_{mb}(\tau)-q(\tau)\right\}^2 \xrightarrow{P} 0.$$

Thus, we obtain that

$$\widehat{\sigma}^2_{mb}(\tau) \xrightarrow{P} \sigma^2(\tau).$$

□

S1.2 Lemmas With Proofs

Lemma S1.1. *Suppose $f_n(x,\tau)$ $(n \geq 1)$ are convex in x for each τ and bounded in τ for each x. Let $g_n(x,\tau) = -x^{\mathsf{T}}W_n(\tau) + \frac{1}{2}x^{\mathsf{T}}Q(\tau)x$, where $\{W_n(\tau)\}$ is a sequence of bounded random vectors and $Q(\tau)$ is a $d\times d$ non-stochastic symmetric positive definite matrix for each τ. Furthermore, we assume that the maximum eigenvalue of $Q(\tau)$ is bounded from above and*

the minimum eigenvalue of $Q(\tau)$ *is bounded away from* 0 *for each* $\tau \in (0,1)$. *If for each* $\tau \in (0,1)$ *and each* x

$$f_n(x,\tau) - g_n(x,\tau) \xrightarrow{P} 0,$$

then if we denote $x_n(\tau)$ *as the argmin of* $f_n(x,\tau)$, *we have*

$$x_n(\tau) = \{Q(\tau)\}^{-1} W_n(\tau) + o_P(1).$$

Proof of Lemma S1.1: The proof is similar to that of Lemma 2 and Basic Corollary in Hjort and Pollard (2011) and we omit the details. □

Lemma S1.2. *Assume that Assumptions 1 and 3 hold. If* $\{f_n\}$ *is a sequence of measurable functions satisfies that for* $a = 0, 1$, *there is a measurable function* $\tilde{f}_a$ *such that*

$$\mathbb{E}\left[f_n\{Y(a)\}|X\right] \xrightarrow{L^2} \tilde{f}_a(X).$$

Then, we have

$$\frac{1}{n}\sum_{i=1}^{n}(2T_i - 1) f_n\{Y_i(1)\} \xrightarrow{P} 0,$$

and

$$\frac{1}{n}\sum_{i=1}^{n}(2T_i - 1) f_n\{Y_i(0)\} \xrightarrow{P} 0.$$

Proof of Lemma S1.2: For $k = 1, \cdots, n$, let $\eta_{k,n} = (2T_k - 1)[f_n\{Y_k(1)\} - \mathbb{E}\{f_n(Y_k(1))|X_k\}]$ and $\mathcal{G}_k$ be the history σ-field generated by the covariates

$\{X_i\}_{i=1}^{k+1}$, the assignments $\{T_i\}_{i=1}^{k}$, and the potential outcome $\{Y_i(1), Y_i(0)\}_{i=1}^{k}$.

Let $\mathcal{G}_{k,n} = \mathcal{G}_k$, $k = 1, \cdots, n$, then we have

$$\begin{aligned}\mathbb{E}\left(\eta_{k,n} \mid \mathcal{G}_{k,n}\right) &= \mathbb{E}\left(2T_k - 1 \mid \mathcal{G}_{k,n}\right) \\ &\quad \times \mathbb{E}\left[f_n\{Y_k(1)\} - \mathbb{E}\left\{f_n(Y_k(1)) \mid X_k\right\} \mid \mathcal{G}_{k,n}\right] = 0,\end{aligned}$$

which implies that $\{\eta_{k,n}\}$ is a martingale difference triangular array with respect to $\{\mathcal{G}_{k,n}\}$. It follows from law of large numbers for martingale difference triangular array that

$$\frac{1}{n}\sum_{i=1}^{n}\eta_{i,n} \xrightarrow{P} 0.$$

Besides, it follows from Assumption 2 that

$$\frac{1}{n}\sum_{i=1}^{n}(2T_i - 1)\tilde{f}_1(X_i) \xrightarrow{P} 0.$$

It follows from Markov's inequality that for any $\varepsilon > 0$,

$$\begin{aligned}&\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}(2T_i - 1)\left\{\mathbb{E}\left(f_n\{Y_i(1)\} \mid X_i\right) - \tilde{f}_1(X_i)\right\} > \varepsilon\right] \\ &\leq \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}\left|\mathbb{E}\left(f_n\{Y_i(1)\} \mid X_i\right) - \tilde{f}_1(X_i)\right| > \varepsilon\right] \\ &\leq \frac{1}{\varepsilon^2}\mathbb{E}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\left|\mathbb{E}\left(f_n\{Y_i(1)\} \mid X_i\right) - \tilde{f}_1(X_i)\right|\right\}^2\right] \\ &\leq \frac{1}{n\varepsilon^2}\sum_{i=1}^{n}\mathbb{E}\left[\left\{\mathbb{E}\left(f_n\{Y_i(1)\} \mid X_i\right) - \tilde{f}_1(X_i)\right\}^2\right] \\ &= \frac{1}{\varepsilon^2}\mathbb{E}\left[\left\{\mathbb{E}\left(f_n\{Y(1)\} \mid X\right) - \tilde{f}_1(X)\right\}^2\right] \to 0,\end{aligned}$$

which implies that

$$\frac{1}{n}\sum_{i=1}^{n}(2T_i-1)\left[\mathbb{E}\left(f_n\{Y_i(1)\}\mid X_i\right)-\tilde{f}_1(X_i)\right]\xrightarrow{P}0.$$

Therefore, we obtain that

$$\frac{1}{n}\sum_{i=1}^{n}(2T_i-1)f_n\{Y_i(1)\}\xrightarrow{P}0.$$

Similarly, it can be shown that

$$\frac{1}{n}\sum_{i=1}^{n}(2T_i-1)f_n\{Y_i(0)\}\xrightarrow{P}0.$$

□

S1.3 Properties of Adaptive Randomization via Mahalanobis Distance

In this section, we prove that Assumption 2 holds for adaptive randomization via Mahalanobis distance and we begin with the following assumption.

Assumption S1.1. *(i) Suppose that $\mathbf{\Sigma}_{XX}$ is covariance matrix of X_i and $\mathbf{\Sigma}_{XX}^{-1/2}$ is the Cholesky square root of $\mathbf{\Sigma}_{XX}^{-1}$. There exists a constant $\gamma>2$ such that $\mathbb{E}[\|\mathbf{\Sigma}_{XX}^{-1/2}\{X_i-\mathbb{E}(X_i)\}\|^{\gamma}]<\infty$.*

(ii) Suppose that the distribution of X_i is Γ_X and there exists an $k_c\in\mathbb{Z}^+$ and $0<c_v\leq 1$ such that

$$\Gamma_X^{k_c*}(A)\geq c_v\int_A v^{k_c*}(y)dy\quad \textit{for any Borel set } A,$$

where Γ_X^{k*} *is the* k*-th convolution of* Γ_X*,* $v(y)$ *is a density function with* $\inf_{y\in O} v(y) > 0$ *for an open set* O *and* v^{k*} *is the* k*-th convolution of* v*.*

Proposition S1.1. *Suppose that Assumption S1.1 holds. Let*

$$\Lambda_k = \sum_{i=1}^{k} (2T_{2i-1} - 1)\boldsymbol{\Sigma}_{XX}^{-1/2}(X_{2i-1} - X_{2i}).$$

Then, $\{\Lambda_k\}_{k=1}^{\infty}$ *is a positive Harris recurrent Markov chain.*

Proof. The proof is similar to that of Lemma A.3 in Liu et al. (2024), and we omit the details. □

Proposition S1.2. *Suppose that Assumption S1.1 holds. Then, Assumption 2 holds for adaptive randomization via Mahalanobis distance.*

Proof. We follow the proof strategy of Theorem 3.6 and Lemma A.3 in Ma et al. (2024) to establish the proposition. Without loss of generality, we assume that n is even. Let $\mathcal{H}_k$ denotes the σ-field generated by $\{X_i, T_i\}_{i=1}^{2k}$. Note that

$$\mathbb{P}(T_{2i-1} = 1, T_{2i} = -1) = \frac{1}{2} - \frac{2\rho - 1}{2} \operatorname{sgn}\left\{\Lambda_{i-1}^{\intercal}(X_{2i-1} - X_{2i})\right\},$$

which follow that

$$\begin{aligned}
&\mathbb{E}\left\{(2T_{2i-1} - 1)m(X_{2i-1}) + (2T_{2i} - 1)m(X_{2i}) | \mathcal{H}_{i-1}, X_{2i-1}, X_{2i}\right\} \\
=& -(2\rho - 1)\mathbb{E}\left[\left\{m(X_{2i-1}) - m(X_{2i})\right\} \operatorname{sgn}\left\{\Lambda_{i-1}^{\intercal}(X_{2i-1} - X_{2i})\right\} \mid \Lambda_{i-1}\right] \\
:=& \varphi(\Lambda_{i-1}).
\end{aligned}$$

It follows from Proposition S1.1 that $\{\Lambda_k\}$ admits an invariant probability measure, denoted by π. Then, we consider the following Poisson equation

$$\widehat{\varphi}(\Lambda) - P_\lambda \widehat{\varphi}(\Lambda) = \varphi(\Lambda) - \pi\{\varphi(\Lambda)\}, \tag{S1.10}$$

where P_λ denotes the transition probability measure of $\{\Lambda_k\}$. The Poisson equation (S1.10) admits a solution satisfying the bound $|\widehat{\varphi}(\Lambda)| \leq R(\|\Lambda\|+1)$ for some constant $R > 0$. Without loss of generality, we assume that $\widehat{\varphi}(\Lambda) = 0$. Besides, notice that $\{\Lambda_k\}$ and $\{-\Lambda_k\}$ have the same distribution. Then, $\pi\{\varphi(\Lambda)\} = 0$, which follows that $\widehat{\varphi}(-\Lambda) = -\widehat{\varphi}(\Lambda)$ and

$$\widehat{\varphi}(\Lambda) - P_\lambda \widehat{\varphi}(\Lambda) = \varphi(\Lambda).$$

Let $\xi_i^* = \sum_{k=2i-1}^{2i} \xi_k$, $\Delta M_{i,1} = \xi_i^* - \mathbb{E}(\xi_i^* | \mathcal{H}_{i-1})$, $\Delta M_{i,2} = \widehat{\varphi}(\Lambda_i) - \mathbb{E}\{\widehat{\varphi}(\Lambda_i)|\mathcal{H}_{i-1}\}$ and $\Delta M_i = \Delta M_{i,1} + \Delta M_{i,2}$. Then, $\{\Delta M_i\}_{i=1}^{n/2}$ is a sequence of martingale difference with respect to $\mathcal{H}_i$ and

$$\sum_{i=1}^{n} \xi_i = \sum_{i=1}^{n/2} \Delta M_i + \widehat{\varphi}(\Lambda_0) - \widehat{\varphi}(\Lambda_{n/2}).$$

Note that $\{\Lambda_k\}$ is positive recurrent and $\max_k \mathbb{E}[\|\Lambda_k\|^4] < \infty$. Then, $\widehat{\varphi}(\Lambda_k) = O(\|\Lambda_k\|^2) = O(1)$ in L_2, which follows that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n/2} \Delta M_i + o_P(1).$$

For the martingale difference sequence $\{\Delta M_i\}_{i=1}^{n/2}$, a simple calculation yields

that

$$
\begin{aligned}
&\mathbb{E}\left\{(\Delta M_i)^2|\mathcal{H}_{i-1}\right\} \\
&= 2\,\mathrm{Var}\left\{m(X)\right\} + 2\mathbb{E}\left[(2T_{2i-1}-1)\left\{m(X_{2i-1})-m(X_{2i})\right\}\widehat{\varphi}(\Lambda_i)|\mathcal{H}_{i-1}\right] \\
&\quad + P_\lambda\widehat{\varphi}^2(\Lambda_{i-1}) - \widehat{\varphi}^2(\Lambda_{i-1}) \\
&\quad + \mathbb{E}[\{(2T_{2i-1}-1)m(X_{2i-1}) + (2T_{2i}-1)m(X_{2i}) + \widehat{\varphi}(\Lambda_i)\}\{f(X_{2i-1})+f(X_{2i})\}|\mathcal{H}_{i-1}] \\
&\quad + 2\sigma_f^2 + \mathrm{Var}\{Z(1)\} + \mathrm{Var}\{Z(0)\} \\
&= \varsigma_m(\Lambda_{i-1}) + \varsigma_{mf}(\Lambda_{i-1}) + 2\sigma_f^2 + \mathrm{Var}\{Z(1)\} + \mathrm{Var}\{Z(0)\},
\end{aligned}
$$

where

$$
\begin{aligned}
\varsigma_m(\Lambda_{i-1}) &= 2\,\mathrm{Var}\left\{m(X)\right\} + 2\mathbb{E}\left[(2T_{2i-1}-1)\left\{m(X_{2i-1})-m(X_{2i})\right\}\widehat{\varphi}(\Lambda_i)|\Lambda_{i-1}\right] \\
&\quad + P_\lambda\widehat{\varphi}^2(\Lambda_{i-1}) - \widehat{\varphi}^2(\Lambda_{i-1}),
\end{aligned}
$$

and

$$
\varsigma_{mf}(\Lambda_{i-1}) = \mathbb{E}[\{(2T_{2i-1}-1)m(X_{2i-1}) + (2T_{2i}-1)m(X_{2i}) + \widehat{\varphi}(\Lambda_i)\}\{f(X_{2i-1})+f(X_{2i})\}|\Lambda_{i-1}].
$$

It follows from the law of large numbers for positive Harris chains (Meyn and Tweedie, 2009, Theorem 17.1.7) that

$$
\frac{2}{n}\sum_{i=1}^{n/2}\mathbb{E}\left\{(\Delta M_i)^2|\mathcal{H}_{i-1}\right\} \xrightarrow{P} \pi\{\varsigma_m(\Lambda)\} + \pi\{\varsigma_{mf}(\Lambda)\} + 2\sigma_f^2 + \mathrm{Var}\{Z(1)\} + \mathrm{Var}\{Z(0)\}.
$$

Note that

$$
\begin{aligned}
&\pi\{\varsigma_{mf}(\Lambda)\} \\
&= \mathbb{E}_\pi\left[\{(2T_{2i-1}-1)m(X_{2i-1}) + (2T_{2i}-1)m(X_{2i}) + \widehat{\varphi}(\Lambda_i)\}\{f(X_{2i-1}) + f(X_{2i})\}\right] \\
&= \mathbb{E}_\pi\left[\{-(2T_{2i-1}-1)m(X_{2i-1}) - (2T_{2i}-1)m(X_{2i}) + \widehat{\varphi}(-\Lambda_i)\}\{f(X_{2i-1}) + f(X_{2i})\}\right] \\
&= -\pi\{\varsigma_{mf}(\Lambda)\},
\end{aligned}
$$

which follows that $\pi\{\varsigma_{mf}(\Lambda)\} = 0$.

Then, it follows from central limit theorem for martingale difference sequence (Hall and Heyde, 1980, Corollary 3.1) that

$$
\sqrt{\frac{2}{n}}\sum_{i=1}^{n/2}\Delta M_i \xrightarrow{d} \mathcal{N}(0, \pi\{\varsigma_m(\Lambda)\} + 2\sigma_f^2 + \mathrm{Var}\{Z(1)\} + \mathrm{Var}\{Z(0)\}),
$$

which follows that

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{2}\pi\{\varsigma_m(\Lambda)\} + \sigma_f^2 + \frac{1}{2}\mathrm{Var}\{Z(1)\} + \frac{1}{2}\mathrm{Var}\{Z(0)\}\right).
$$

Let $f(X_i) = Z_i(1) = Z_i(0) = 0$, we have

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(2T_i - 1)m(X_i) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{2}\pi\{\varsigma_m(\Lambda)\}\right).
$$

This completes the proof of the proposition.

□

S2 Additional Simulation Results

S2.1 Additional Simulation Results for Section 6

Additional simulation results are provided for point estimation for $\tau = 0.25$ and $\tau = 0.75$ under sample sizes $n = 1,000$. The results are presented in Tables 1 and 2. The simulation settings are the same as the point estimation in Section 6.

Table 1: Simulated bias, SD, RMSE, SE and CP of the QTE estimates under different models and randomization procedures ($n = 1000$, $\tau = 0.25$)

						SE		CP	
M.	Ran.	Estimator	Bias	SD	RMSE	CAB	MB	CAB	MB
1	CR	$\widehat{q}(\tau)$	0.003	0.445	0.445	0.442	0.437	0.938	0.946
	HH	$\widehat{q}(\tau)$	0.005	0.397	0.397	0.405	0.421	0.942	0.951
		$\widehat{q}_{adj}(\tau, \boldsymbol{\psi}_{SB})$	0.000	0.383	0.383	0.379	0.396	0.938	0.950
	COV	$\widehat{q}(\tau)$	0.004	0.405	0.405	0.411	0.408	0.943	0.946
		$\widehat{q}_{adj}(\tau, \boldsymbol{\psi}_{COV})$	-0.005	0.385	0.385	0.384	0.382	0.936	0.945
	ARM	$\widehat{q}(\tau)$	0.024	0.413	0.414	0.414	0.406	0.939	0.946
		$\widehat{q}_{adj}(\tau, \boldsymbol{\psi}_{ARM})$	0.013	0.388	0.388	0.396	0.388	0.942	0.946
2	CR	$\widehat{q}(\tau)$	-0.026	0.612	0.612	0.622	0.603	0.946	0.946
	HH	$\widehat{q}(\tau)$	0.004	0.549	0.549	0.568	0.588	0.949	0.961
		$\widehat{q}_{adj}(\tau, \boldsymbol{\psi}_{SB})$	0.022	0.531	0.531	0.541	0.562	0.939	0.955
	COV	$\widehat{q}(\tau)$	0.010	0.557	0.556	0.567	0.597	0.948	0.964
		$\widehat{q}_{adj}(\tau, \boldsymbol{\psi}_{COV})$	0.015	0.505	0.505	0.521	0.555	0.946	0.966
	ARM	$\widehat{q}(\tau)$	-0.043	0.561	0.563	0.566	0.544	0.942	0.941
		$\widehat{q}_{adj}(\tau, \boldsymbol{\psi}_{ARM})$	-0.017	0.546	0.546	0.544	0.521	0.941	0.939

Abbreviations: M, Model; Ran, Randomization.

Table 2: Simulated bias, SD, RMSE, SE and CP of the QTE estimates under different models and randomization procedures ($n = 1000$, $\tau = 0.75$)

						SE		CP	
M.	Ran.	Estimator	Bias	SD	RMSE	CAB	MB	CAB	MB
1	CR	$\widehat{q}(\tau)$	-0.002	0.515	0.515	0.518	0.514	0.945	0.946
	HH	$\widehat{q}(\tau)$	-0.023	0.432	0.433	0.445	0.484	0.951	0.967
		$\widehat{q}_{adj}(\tau, \boldsymbol{\psi}_{SB})$	-0.020	0.424	0.425	0.436	0.476	0.944	0.967
	COV	$\widehat{q}(\tau)$	-0.012	0.447	0.447	0.450	0.456	0.943	0.952
		$\widehat{q}_{adj}(\tau, \boldsymbol{\psi}_{COV})$	-0.007	0.436	0.436	0.439	0.445	0.943	0.948
	ARM	$\widehat{q}(\tau)$	-0.016	0.469	0.469	0.480	0.472	0.948	0.953
		$\widehat{q}_{adj}(\tau, \boldsymbol{\psi}_{ARM})$	-0.003	0.463	0.463	0.474	0.466	0.949	0.953
2	CR	$\widehat{q}(\tau)$	-0.045	1.210	1.211	1.241	1.200	0.948	0.944
	HH	$\widehat{q}(\tau)$	-0.005	0.716	0.716	0.719	0.758	0.945	0.960
		$\widehat{q}_{adj}(\tau, \boldsymbol{\psi}_{SB})$	0.010	0.680	0.680	0.673	0.715	0.933	0.958
	COV	$\widehat{q}(\tau)$	-0.025	1.102	1.102	1.121	1.175	0.938	0.959
		$\widehat{q}_{adj}(\tau, \boldsymbol{\psi}_{COV})$	-0.014	1.031	1.031	1.041	1.099	0.940	0.960
	ARM	$\widehat{q}(\tau)$	0.028	1.176	1.176	1.155	1.186	0.941	0.952
		$\widehat{q}_{adj}(\tau, \boldsymbol{\psi}_{ARM})$	-0.035	1.148	1.148	1.104	1.136	0.934	0.946

Abbreviations: M, Model; Ran, Randomization.

Bibliography

Hall, P. and C. C. Heyde (1980). *Martingale limit theory and its application*. Academic press.

Hjort, N. L. and D. Pollard (2011). Asymptotics for minimisers of convex processes. *arXiv preprint arXiv:1107.3806*.

Knight, K. (1998). Limiting distributions for l1 regression estimators under general conditions. *Annals of statistics 26*(2), 755–770.

Liu, Y., Y. Zhou, P. Li, and F. Hu (2024). Cluster-adaptive network a/b testing: From randomization to estimation. *Journal of Machine Learning Research 25*(170), 1–48.

Ma, W., P. Li, L.-X. Zhang, and F. Hu (2024). A new and unified family of covariate adaptive randomization procedures and their properties. *Journal of the American Statistical Association 119*(545), 151–162.

Meyn, S. and R. L. Tweedie (2009). *Markov Chains and Stochastic Stability*. Cambridge University Press.

Qin, Y., Y. Li, W. Ma, H. Yang, and F. Hu (2022). Adaptive randomization via mahalanobis distance. *Statistica Sinica 1*, 1–40.

Zhang, L.-X. (2023). Asymptotic properties of multi-treatment covariate adaptive randomization procedures for balancing observed and unobserved covariates. *arXiv preprint arXiv:2305.13842*.