SUPPLEMENT TO "GENERALIZATION ERROR BOUNDS OF DYNAMIC TREATMENT REGIMES IN PENALIZED REGRESSION-BASED LEARNING"

BY EUN JEONG OH[1,*], MIN QIAN[1,†] AND KEN CHEUNG[1,‡]

[1]Department of Biostatistics, Columbia University, [*]eo2366@caa.columbia.edu; [†]mq2158@cumc.columbia.edu; [‡]yc632@cumc.columbia.edu

S.1. Proofs of Lemmas. **Proof of Lemma 1.**
We use induction to prove the results. At the last stage T, note that the l_1-PLS estimator $\hat{\boldsymbol{\theta}}_T$ satisfies the following first order condition:

$$-2\mathbb{E}_n[(Y_T - \Phi_T^\intercal \hat{\boldsymbol{\theta}}_T)\phi_{Tj}] + \lambda_T w_{Tj}\mathrm{sgn}(\hat{\theta}_{Tj}) = 0 \text{ for } j = 1, \ldots, J_T,$$

where $\mathrm{sgn}(x) = 1$ if $x > 0$, $\mathrm{sgn}(x) = -1$ if $x < 0$ and $\mathrm{sgn}(x) \in [-1, 1]$ if $x = 0$ for any $x \in \mathbb{R}$. This implies

$$-2\mathbb{E}_n[(Y_T - \Phi_T^\intercal \hat{\boldsymbol{\theta}}_T)\Phi_T^\intercal \boldsymbol{\theta}_T] + \lambda_T \sum_{j=1}^{J_T} w_{Tj}\mathrm{sgn}(\hat{\theta}_{Tj})\theta_{Tj} = 0$$

for any $\boldsymbol{\theta}_T \in \mathbb{R}^{J_T}$. In particular, $-2\mathbb{E}_n[(Y_T - \Phi_T^\intercal \hat{\boldsymbol{\theta}}_T)\Phi_T^\intercal \hat{\boldsymbol{\theta}}_T] + \lambda_T \sum_{j=1}^{J_T} w_{Tj}|\hat{\theta}_{Tj}| = 0$. Therefore, for any $\boldsymbol{\theta}_T \in \mathbb{R}^{J_T}$, we have

$$\begin{aligned}0 &= 2\mathbb{E}_n[(Y_T - \Phi_T^\intercal \hat{\boldsymbol{\theta}}_T)\Phi_T^\intercal(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_T)] + \lambda_T \sum_{j=1}^{J_T} w_{Tj}\mathrm{sgn}(\hat{\theta}_{Tj})\theta_{Tj} - \lambda_T \sum_{j=1}^{J_T} w_{Tj}|\hat{\theta}_{Tj}| \\ &\le 2\mathbb{E}_n[(Y_T - \Phi_T^\intercal \hat{\boldsymbol{\theta}}_T)\Phi_T^\intercal(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_T)] + \lambda_T \sum_{j=1}^{J_T} w_{Tj}|\theta_{Tj}| - \lambda_T \sum_{j=1}^{J_T} w_{Tj}|\hat{\theta}_{Tj}|. \qquad \text{(S.1)}\end{aligned}$$

Fix n. Following (S.1), on the event $\Omega_{T,2}(\boldsymbol{\theta}_T)$, we have

$$\begin{aligned}0 &\le 2 \max_{j\in\{1,\ldots,J_T\}} \left|\mathbb{E}_n\left[(Y_T - \Phi_T^\intercal \boldsymbol{\theta}_T)\frac{\phi_{Tj}}{\bar{w}_{Tj}}\right]\right| \Big(\sum_{j=1}^{J_T} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}|\Big) - 2\mathbb{E}_n[\Phi_T^\intercal(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_T)]^2 \\ &\quad + \lambda_T \sum_{j\in I_T(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}| - \lambda_T \sum_{j\in I_T^c(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj}| \\ &\le \frac{4(\gamma+1)}{3}\lambda_T \Big(\sum_{j\in I_T(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}|\Big) - \frac{2(1-2\gamma)}{3}\lambda_T \Big(\sum_{j\in I_T^c(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj}|\Big) \\ &\quad - 2\mathbb{E}_n[\Phi_T^\intercal(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_T)]^2. \qquad \text{(S.2)}\end{aligned}$$

This implies

$$\sum_{j\in I_T^c(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj}| \le \frac{2(\gamma+1)}{1-2\gamma}\Big(\sum_{j\in I_T(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}|\Big) \qquad \text{(S.3)}$$

$$\text{and } \mathbb{E}_n[\Phi_T^\intercal(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_T)]^2 \le \frac{2(\gamma+1)}{3}\lambda_T\Big(\sum_{j\in I_T(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}|\Big). \qquad \text{(S.4)}$$

Thus, under condition (3.9) on the event $\Omega_{T,1}(\boldsymbol{\theta}_T)$, we have

$$
\begin{aligned}
&-\mathbb{E}_n[\Phi_T^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_T)]^2\\
&\le \max_{j,k\in\{1,\dots,J_T\}}\Big|(E-\mathbb{E}_n)\Big(\frac{\phi_{Tj}\phi_{Tk}}{\bar{w}_{Tj}\bar{w}_{Tk}}\Big)\Big|\Big(\sum_{j=1}^{J_T}\bar{w}_{Tj}|\hat{\theta}_{Tj}-\theta_{Tj}|\Big)^2-E[\Phi_T^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_T)]^2\\
&\le \frac{\tau_T}{16|I_T(\boldsymbol{\theta}_T)|}\Big(\sum_{j\in I_T(\boldsymbol{\theta}_T)}\bar{w}_{Tj}|\hat{\theta}_{Tj}-\theta_{Tj}|\Big)^2-\frac{\tau_T}{|I_T(\boldsymbol{\theta}_T)|}\Big(\sum_{j\in I_T(\boldsymbol{\theta}_T)}\bar{w}_{Tj}|\hat{\theta}_{Tj}-\theta_{Tj}|\Big)^2
\end{aligned}
$$

$$
= -\frac{15\tau_T}{16|I_T(\boldsymbol{\theta}_T)|}\Big(\sum_{j\in I_T(\boldsymbol{\theta}_T)}\bar{w}_{Tj}|\hat{\theta}_{Tj}-\theta_{Tj}|\Big)^2. \tag{S.5}
$$

Plugging (S.5) into (S.2) yields

$$
\begin{aligned}
0\le \frac{4(\gamma+1)}{3}\lambda_T\Big(\sum_{j\in I_T(\boldsymbol{\theta}_T)}\bar{w}_{Tj}|\hat{\theta}_{Tj}-\theta_{Tj}|\Big)-\frac{2(1-2\gamma)}{3}\lambda_T\Big(\sum_{j\in I_T^c(\boldsymbol{\theta}_T)}\bar{w}_{Tj}|\hat{\theta}_{Tj}|\Big)&\\
-\frac{15\tau_T}{8|I_T(\boldsymbol{\theta}_T)|}\Big(\sum_{j\in I_T(\boldsymbol{\theta}_T)}\bar{w}_{Tj}|\hat{\theta}_{Tj}-\theta_{Tj}|\Big)^2&.
\end{aligned}
$$

Rearranging the terms, we obtain

$$
\sum_{j\in I_T(\boldsymbol{\theta}_T)}\bar{w}_{Tj}|\hat{\theta}_{Tj}-\theta_{Tj}|\le\frac{32(\gamma+1)|I_T(\boldsymbol{\theta}_T)|\lambda_T}{45\tau_T}. \tag{S.6}
$$

Plugging (S.6) into (S.3) and (S.4) yields

$$
\begin{aligned}
\sum_{j=1}^{J_T}\bar{w}_{Tj}|\hat{\theta}_{Tj}-\theta_{Tj}|&\le\Big[\frac{32(\gamma+1)}{15(1-2\gamma)}\Big]\frac{|I_T(\boldsymbol{\theta}_T)|\lambda_T}{\tau_T}\le\Big[\frac{16(2\gamma+5)}{3(1-2\gamma)}\Big]\frac{|I_T(\boldsymbol{\theta}_T)|\lambda_T}{\tau_T}\\
\text{and } \mathbb{E}_n[\Phi_T^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_T)]^2&\le\Big[\frac{64(\gamma+1)^2}{135}\Big]\frac{|I_T(\boldsymbol{\theta}_T)|\lambda_T^2}{\tau_T}\le\Big[\frac{16(2\gamma+5)^2}{27}\Big]\frac{|I_T(\boldsymbol{\theta}_T)|\lambda_T^2}{\tau_T},
\end{aligned}
$$

on the event $\Omega_{T,1}(\boldsymbol{\theta}_T)\cap\Omega_{T,2}(\boldsymbol{\theta}_T)$. Thus,

$$
\begin{aligned}
E[\Phi_T^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_T)]^2&\le\frac{(1-2\gamma)^2\tau_T}{144|I_T(\boldsymbol{\theta}_T)|}\Big(\sum_{j=1}^{J_T}\bar{w}_{Tj}|\hat{\theta}_{Tj}-\theta_{Tj}|\Big)^2+\mathbb{E}_n[\Phi_T^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_T)]^2\\
&\le\Big[\frac{64(2\gamma+5)^2}{81}\Big]\frac{|I_T(\boldsymbol{\theta}_T)|\lambda_T^2}{\tau_T}.
\end{aligned}
$$

Hence, (C.8) and (C.9) hold for $t=T$ on the event $\Omega_{T,1}(\boldsymbol{\theta}_T)\cap\Omega_{T,2}(\boldsymbol{\theta}_T)$.

Now we prove the results for $t<T$. Assume we have

$$
\sum_{j=1}^{J_s}\bar{w}_{sj}|\hat{\theta}_{sj}-\theta_{sj}|\le\frac{16(2\gamma+5)}{3(1-2\gamma)\lambda_s}\max_{s'\in\{s,\dots,T\}}\Big\{c_{s,s'}\frac{|I_{s'}(\boldsymbol{\theta}_{s'})|\lambda_{s'}^2}{\tau_{s'}}\Big\}, \tag{S.7}
$$

$$
E[\Phi_s^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s)]^2\le\frac{64(2\gamma+5)^2}{81}\max_{s'\in\{s,\dots,T\}}\Big\{c_{s,s'}\frac{|I_{s'}(\boldsymbol{\theta}_{s'})|\lambda_{s'}^2}{\tau_{s'}}\Big\}, \tag{S.8}
$$

$$
\text{and } \mathbb{E}_n[\Phi_s^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s)]^2\le\frac{16(2\gamma+5)^2}{27}\max_{s'\in\{s,\dots,T\}}\Big\{c_{s,s'}\frac{|I_{s'}(\boldsymbol{\theta}_{s'})|\lambda_{s'}^2}{\tau_{s'}}\Big\}, \tag{S.9}
$$

where $c_{s,s}=1, c_{s,s'}=2(2\gamma+5)(5S+3)(T-s)^2c_{s+1,s'}/9$ for $s'=s+1,\dots,T$ and $s=t+1,\dots,T$, on the event $\cap_{s=t+1}^{T}\Big\{\Omega_{s,1}(\boldsymbol{\theta}_s,\dots,\boldsymbol{\theta}_T)\cap\Omega_{s,2}(\boldsymbol{\theta}_s,\dots,\boldsymbol{\theta}_T)\cap\Omega_{s+1,3}(\boldsymbol{\theta}_{s+1},\dots,\boldsymbol{\theta}_T)\Big\}$.

Using similar arguments as above, $\hat{\boldsymbol{\theta}}_t$ satisfies the first order condition:

$$-2\mathbb{E}_n[(\tilde{Y}_t(\hat{\boldsymbol{\theta}}_{t+1},\dots,\hat{\boldsymbol{\theta}}_T)-\Phi_t^{\mathsf{T}}\hat{\boldsymbol{\theta}}_t)\phi_{tj}]+\lambda_t w_{tj}\mathrm{sgn}(\hat{\theta}_{tj})=0 \text{ for } j=1,\dots,J_t.$$

Thus

$$-2\mathbb{E}_n[(\tilde{Y}_t(\hat{\boldsymbol{\theta}}_{t+1},\dots,\hat{\boldsymbol{\theta}}_T)-\Phi_t^{\mathsf{T}}\hat{\boldsymbol{\theta}}_t)\Phi_t^{\mathsf{T}}\boldsymbol{\theta}_t]+\lambda_t\sum_{j=1}^{J_t}w_{tj}\mathrm{sgn}(\hat{\theta}_{tj})\theta_{tj}=0$$

for any $\boldsymbol{\theta}_t\in\mathbb{R}^{J_t}$. In particular, $-2\mathbb{E}_n[(\tilde{Y}_t(\hat{\boldsymbol{\theta}}_{t+1},\dots,\hat{\boldsymbol{\theta}}_T)-\Phi_t^{\mathsf{T}}\hat{\boldsymbol{\theta}}_t)\Phi_t^{\mathsf{T}}\hat{\boldsymbol{\theta}}_t]+\lambda_t\sum_{j=1}^{J_t}w_{tj}|\hat{\theta}_{tj}|=0$. Hence, for any $\boldsymbol{\theta}_t\in\mathbb{R}^{J_t}$ on the event $\Omega_{t,2}(\boldsymbol{\theta}_t,\dots,\boldsymbol{\theta}_T)$, we have

(S.10)

$$\begin{aligned}
0 &\le 2\mathbb{E}_n[(\tilde{Y}_t(\hat{\boldsymbol{\theta}}_{t+1},\dots,\hat{\boldsymbol{\theta}}_T)-\Phi_t^{\mathsf{T}}\hat{\boldsymbol{\theta}}_t)\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t)]+\lambda_t\sum_{j=1}^{J_t}w_{tj}|\theta_{tj}|-\lambda_t\sum_{j=1}^{J_t}w_{tj}|\hat{\theta}_{tj}|\\
&=2\mathbb{E}_n[(\tilde{Y}_t(\boldsymbol{\theta}_{t+1},\dots,\boldsymbol{\theta}_T)-\Phi_t^{\mathsf{T}}\boldsymbol{\theta}_t)\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t)]+\lambda_t\sum_{j=1}^{J_t}w_{tj}|\theta_{tj}|-\lambda_t\sum_{j=1}^{J_t}w_{tj}|\hat{\theta}_{tj}|\\
&\quad+2\mathbb{E}_n[(\tilde{Y}_t(\hat{\boldsymbol{\theta}}_{t+1},\dots,\hat{\boldsymbol{\theta}}_T)-\tilde{Y}_t(\boldsymbol{\theta}_{t+1},\dots,\boldsymbol{\theta}_T))\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t)]-2\mathbb{E}_n[\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t)]^2\\
&\le\frac{4(\gamma+1)}{3}\lambda_t\Big(\sum_{j\in I_t(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\hat{\theta}_{tj}-\theta_{tj}|\Big)-\frac{2(1-2\gamma)}{3}\lambda_t\Big(\sum_{j\in I_t^c(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\hat{\theta}_{tj}|\Big)
\end{aligned}$$

(S.11)

$$\quad-\mathbb{E}_n[\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t)]^2+\mathbb{E}_n\big[(\tilde{Y}_t(\hat{\boldsymbol{\theta}}_{t+1},\dots,\hat{\boldsymbol{\theta}}_T)-\tilde{Y}_t(\boldsymbol{\theta}_{t+1},\dots,\boldsymbol{\theta}_T))\big]^2.$$

Below, we derive an upper bound for the last term in (S.11). Note that

$$\begin{aligned}
&\mathbb{E}_n\big[\tilde{Y}_t(\hat{\boldsymbol{\theta}}_{t+1},\dots,\hat{\boldsymbol{\theta}}_T)-\tilde{Y}_t(\boldsymbol{\theta}_{t+1},\dots,\boldsymbol{\theta}_T)\big]^2\\
&=\mathbb{E}_n\Bigg[\sum_{s=t+1}^{T}\Big[\max_{a_s}\Phi_s^{\mathsf{T}}(H_s,a_s)\hat{\boldsymbol{\theta}}_s-\max_{a_s}\Phi_s^{\mathsf{T}}(H_s,a_s)\boldsymbol{\theta}_s-\Phi_s^{\mathsf{T}}(H_s,A_s)(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s)\Big]\Bigg]^2
\end{aligned}$$

(S.12)

$$\le 2(T-t)\sum_{s=t+1}^{T}\Big\{\mathbb{E}_n\Big[\max_{a_s}\big|\Phi_s^{\mathsf{T}}(H_s,a_s)(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s)\big|^2\Big]+\mathbb{E}_n\Big[\Phi_s^{\mathsf{T}}(H_s,A_s)(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s)\Big]^2\Big\}.$$

We can further show that

$$\begin{aligned}
&\mathbb{E}_n\Big[\max_{a_s}\big|\Phi_s^{\mathsf{T}}(H_s,a_s)(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s)\big|^2\Big]\\
&\le(\mathbb{E}_n-E)\Big[\textstyle\sum_{a_s}\big|\Phi_s^{\mathsf{T}}(H_s,a_s)(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s)\big|^2\Big]+E\Big[\textstyle\sum_{a_s}\big|\Phi_s^{\mathsf{T}}(H_s,a_s)(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s)\big|^2\Big]\\
&\le\max_{j,k\in\{1,\dots,J_s\}}\Big|(\mathbb{E}_n-E)\Big(\sum_{a_s\in\mathcal{A}_s}\frac{\phi_{sj}(H_s,a_s)\phi_{sk}(H_s,a_s)}{\bar{w}_{sj}\bar{w}_{sk}}\Big)\Big|\Big(\sum_{j=1}^{J_s}\bar{w}_{sj}|\hat{\theta}_{sj}-\theta_{sj}|\Big)^2\\
&\qquad+E\Big[\sum_{a_s\in\mathcal{A}_s}p_s(a_s|H_s)S\big[\Phi_s^{\mathsf{T}}(H_s,a_s)(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s)\big]^2\Big]
\end{aligned}$$

$$\leq \frac{(1-2\gamma)^2|\mathcal{A}_s|}{144\max_{s'\in\{s,\dots,T\}}\{|I_{s'}(\boldsymbol{\theta}_{s'})|/\tau_{s'}\}}\bigg(\frac{16(2\gamma+5)}{3(1-2\gamma)\lambda_s}\max_{s'\in\{s,\dots,T\}}\bigg\{c_{s,s'}\frac{|I_{s'}(\boldsymbol{\theta}_{s'})|\lambda_{s'}^2}{\tau_{s'}}\bigg\}\bigg)^2$$
$$+ SE\big[\Phi_s^{\mathsf{T}}(H_s,A_s)(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s)\big]^2$$
$$\leq \frac{16(2\gamma+5)^2|\mathcal{A}_s|}{81}\bigg[\max_{s'\in\{s,\dots,T\}}\bigg\{c_{s,s'}\frac{|I_{s'}(\boldsymbol{\theta}_{s'})|\lambda_{s'}^2}{\tau_{s'}}\bigg\}\bigg]$$
$$+\frac{64(2\gamma+5)^2S}{81}\bigg[\max_{s'\in\{s,\dots,T\}}\bigg\{c_{s,s'}\frac{|I_{s'}(\boldsymbol{\theta}_{s'})|\lambda_{s'}^2}{\tau_{s'}}\bigg\}\bigg]$$

(S.13)
$$\leq \frac{80(2\gamma+5)^2S}{81}\bigg[\max_{s'\in\{s,\dots,T\}}\bigg\{c_{s,s'}\frac{|I_{s'}(\boldsymbol{\theta}_{s'})|\lambda_{s'}^2}{\tau_{s'}}\bigg\}\bigg]$$

where the second inequality follows from the assumption that $p_s(a_s|h_s)\geq S^{-1}$ for all (h_s,a_s) pairs, the third inequality follows from the definition of $\Omega_{s,3}(\boldsymbol{\theta}_s,\dots,\boldsymbol{\theta}_T)$ and (S.7), the fourth inequality follows from (S.8) and condition (C.4), and the last inequality follows from the fact that $|\mathcal{A}_s|\leq S$. Plugging (S.9) and (S.13) into (S.12) and noticing that $c_{s,s'}\leq c_{t+1,s'}$ for any $s\geq t+1$ and $s'\geq s$, we have

$$\text{(S.14)}\quad \mathbb{E}_n\big[\tilde{Y}_t(\hat{\boldsymbol{\theta}}_{t+1},\dots,\hat{\boldsymbol{\theta}}_T)-\tilde{Y}_t(\boldsymbol{\theta}_{t+1},\dots,\boldsymbol{\theta}_T)\big]^2\leq C_t\max_{s\in\{t+1,\dots,T\}}\bigg\{c_{t+1,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\bigg\},$$

where $C_t=32(2\gamma+5)^2(5S+3)(T-t)^2/81$. This, together with (S.11), implies that, on the event $\cap_{s=t+1}^T\Big\{\Omega_{s,1}(\boldsymbol{\theta}_s,\dots,\boldsymbol{\theta}_T)\cap\Omega_{s,2}(\boldsymbol{\theta}_s,\dots,\boldsymbol{\theta}_T)\cap\Omega_{s,3}(\boldsymbol{\theta}_s,\dots,\boldsymbol{\theta}_T)\Big\}\cap\Omega_{t,2}(\boldsymbol{\theta}_t,\dots,\boldsymbol{\theta}_T)$,

$$0\leq\frac{4(\gamma+1)}{3}\lambda_t\sum_{j\in I_t(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\hat{\theta}_{tj}-\theta_{tj}|-\frac{2(1-2\gamma)}{3}\lambda_t\sum_{j\in I_t^c(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\hat{\theta}_{tj}|$$

(S.15)
$$-\mathbb{E}_n[\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t)]^2+C_t\max_{s\in\{t+1,\dots,T\}}\bigg\{c_{t+1,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\bigg\}.$$

Thus

$$\sum_{j\in I_t^c(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\hat{\theta}_{tj}|$$
$$\leq\frac{2(\gamma+1)}{1-2\gamma}\Big(\sum_{j\in I_t(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\hat{\theta}_{tj}-\theta_{tj}|\Big)+\frac{3C_t}{2(1-2\gamma)\lambda_t}\max_{s\in\{t+1,\dots,T\}}\bigg\{c_{t+1,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\bigg\}$$

and

(S.16)
$$\mathbb{E}_n[\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t)]^2\leq\frac{4(\gamma+1)}{3}\lambda_t\Big(\sum_{j\in I_t(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\hat{\theta}_{tj}-\theta_{tj}|\Big)+C_t\max_{s\in\{t+1,\dots,T\}}\bigg\{c_{t+1,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\bigg\}.$$

If $I_t(\boldsymbol{\theta}_t)$ is empty (i.e., $\boldsymbol{\theta}_t=\mathbf{0}$), then it is easy to verify that (C.8) and (C.9) hold. If $I_t(\boldsymbol{\theta}_t)$ is non-empty, define the sets

$$\Theta_{t,1}(\boldsymbol{\theta}_t)=\bigg\{\tilde{\boldsymbol{\theta}}_t\in\mathbb{R}^{J_t}:\sum_{j\in I_t^c(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\tilde{\theta}_{tj}|$$

$$\leq \frac{2(\gamma+1)}{1-2\gamma}\Big(\sum_{j\in I_t(\boldsymbol{\theta}_t)} \bar{w}_{tj}|\tilde{\theta}_{tj}-\theta_{tj}|\Big) + \frac{3C_t}{2(1-2\gamma)\lambda_t}\max_{s\in\{t+1,\dots,T\}}\left\{c_{t+1,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\right\}\Bigg\},$$

$$\Theta_{t,2}(\boldsymbol{\theta}_t) = \Bigg\{\tilde{\boldsymbol{\theta}}_t \in \mathbb{R}^{J_t} : \sum_{j\in I_t(\boldsymbol{\theta}_t)} \bar{w}_{tj}|\tilde{\theta}_{tj}-\theta_{tj}| > \max\left\{\frac{8(2\gamma+5)|I_t(\boldsymbol{\theta}_t)|\lambda_t}{9\tau_t}, \frac{C_t}{2\lambda_t}\max_{s\in\{t+1,\dots,T\}}\left\{c_{t+1,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\right\}\right\}\Bigg\}.$$

On the event $\cap_{s=t}^T\Big\{\Omega_{s,1}(\boldsymbol{\theta}_s,\dots,\boldsymbol{\theta}_T)\cap\Omega_{s,2}(\boldsymbol{\theta}_s,\dots,\boldsymbol{\theta}_T)\cap\Omega_{s+1,3}(\boldsymbol{\theta}_{s+1},\dots,\boldsymbol{\theta}_T)\Big\}$, we have $\hat{\boldsymbol{\theta}}_t\in\Theta_{t,1}(\boldsymbol{\theta}_t)$. Thus, $\hat{\boldsymbol{\theta}}_t\in\Theta_{t,1}(\boldsymbol{\theta}_t)$ on the event $\Omega_{t,2}(\boldsymbol{\theta}_t,\dots,\boldsymbol{\theta}_T)$. Note that condition (3.9) by Assumption (A2) implies that

$$E[\Phi_t^{\mathsf{T}}(\tilde{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t)]^2 \geq \frac{\tau_t(\sum_{j\in I_t(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\tilde{\theta}_{tj}-\theta_{tj}|)^2}{|I_t(\boldsymbol{\theta}_t)|} \tag{S.17}$$

for any $\boldsymbol{\theta}_t$ and $\tilde{\boldsymbol{\theta}}_t$. In addition,

$$\sup_{\tilde{\boldsymbol{\theta}}_t\in\Theta_{t,1}(\boldsymbol{\theta}_t)\cap\Theta_{t,2}(\boldsymbol{\theta}_t)}\Bigg\{\frac{4(\gamma+1)}{3}\lambda_t\Big(\sum_{j\in I_t(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\tilde{\theta}_{tj}-\theta_{tj}|\Big) - \frac{2(1-2\gamma)}{3}\lambda_t\Big(\sum_{j\in I_t^c(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\tilde{\theta}_{tj}|\Big)$$
$$- \mathbb{E}_n[\Phi_t^{\mathsf{T}}(\tilde{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t)]^2 + C_t\max_{s\in\{t+1,\dots,T\}}\left\{c_{t+1,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\right\}\Bigg\}$$
$$\leq \sup_{\tilde{\boldsymbol{\theta}}_t\in\Theta_{t,1}(\boldsymbol{\theta}_t)\cap\Theta_{t,2}(\boldsymbol{\theta}_t)}\Bigg\{\frac{4(\gamma+1)}{3}\lambda_t\Big(\sum_{j\in I_t(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\tilde{\theta}_{tj}-\theta_{tj}|\Big) - \frac{2(1-2\gamma)}{3}\lambda_t\Big(\sum_{j\in I_t^c(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\tilde{\theta}_{tj}|\Big)$$
$$+ \max_{j,k\in\{1,\dots,J_t\}}\Big|(E-\mathbb{E}_n)\Big(\frac{\phi_{tj}\phi_{tk}}{\bar{w}_{tj}\bar{w}_{tk}}\Big)\Big|\Big(\sum_{j=1}^{J_t}\bar{w}_{tj}|\tilde{\theta}_{tj}-\theta_{tj}|\Big)^2$$
$$- E[\Phi_t^{\mathsf{T}}(\tilde{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t)]^2 + C_t\max_{s\in\{t+1,\dots,T\}}\left\{c_{t+1,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\right\}\Bigg\}$$
$$\leq \sup_{\tilde{\boldsymbol{\theta}}_t\in\Theta_{t,1}(\boldsymbol{\theta}_t)\cap\Theta_{t,2}(\boldsymbol{\theta}_t)}\Bigg\{\frac{4(\gamma+1)}{3}\lambda_t\Big(\sum_{j\in I_t(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\tilde{\theta}_{tj}-\theta_{tj}|\Big)$$
$$+ \frac{(1-2\gamma)^2}{144\max_{s\in\{t,\dots,T\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}}\times$$
$$\left[\frac{3}{1-2\gamma}\Big(\sum_{j\in I_t(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\tilde{\theta}_{tj}-\theta_{tj}| + \frac{C_t}{2\lambda_t}\max_{s\in\{t+1,\dots,T\}}\left\{c_{t+1,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\right\}\Big)\right]^2$$
$$- E[\Phi_t^{\mathsf{T}}(\tilde{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t]^2 + C_t\max_{s\in\{t+1,\dots,T\}}\left\{c_{t+1,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\right\}\Bigg\}$$
$$\leq \sup_{\tilde{\boldsymbol{\theta}}_t\in\Theta_{t,1}(\boldsymbol{\theta}_t)\cap\Theta_{t,2}(\boldsymbol{\theta}_t)}\Bigg\{\frac{4(\gamma+1)}{3}\lambda_t\Big(\sum_{j\in I_t(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\tilde{\theta}_{tj}-\theta_{tj}|\Big) + \frac{\tau_t}{4|I_t(\boldsymbol{\theta}_t)|}\Big(\sum_{j\in I_t(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\tilde{\theta}_{tj}-\theta_{tj}|\Big)^2$$
$$- \frac{\tau_t}{|I_t(\boldsymbol{\theta}_t)|}\Big(\sum_{j\in I_t(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\tilde{\theta}_{tj}-\theta_{tj}|\Big)^2 + 2\lambda_t\Big(\sum_{j\in I_t(\boldsymbol{\theta}_t)}\bar{w}_{tj}|\tilde{\theta}_{tj}-\theta_{tj}|\Big)\Bigg\}$$

$$< 0,$$

where the second inequality follows from the definition of $\Omega_{t,1}(\boldsymbol{\theta}_t,\dots,\boldsymbol{\theta}_T)$ and $\Theta_{t,1}(\boldsymbol{\theta}_t)$, the third inequality follows from the definition of $\Theta_{t,2}(\boldsymbol{\theta}_t)$ and (S.17), and the last inequality follows from the definition of $\Theta_{t,2}(\boldsymbol{\theta}_t)$.

Since $\hat{\boldsymbol{\theta}}_t$ satisfies inequality (S.15), we have $\hat{\boldsymbol{\theta}}_t \in \Theta_{t,1}(\boldsymbol{\theta}_t)\cap\Theta_{t,2}(\boldsymbol{\theta}_t)^C$ on the event $\cap_{s=t}^T\Big\{\Omega_{s,1}(\boldsymbol{\theta}_s,\dots,\boldsymbol{\theta}_T)\cap\Omega_{s,2}(\boldsymbol{\theta}_s,\dots,\boldsymbol{\theta}_T)\cap\Omega_{s+1,3}(\boldsymbol{\theta}_{s+1},\dots,\boldsymbol{\theta}_T)\Big\}$. This, together with (S.16) and (C.4), implies that

$$
\begin{aligned}
&\sum_{j=1}^{J_t}\bar{w}_{tj}|\hat{\theta}_{tj}-\theta_{tj}|\\
&\qquad\le \max\left\{\frac{16(2\gamma+5)|I_t(\boldsymbol{\theta}_t)|\lambda_t}{3(1-2\gamma)\tau_t},\ \frac{3C_t}{(1-2\gamma)\lambda_t}\max_{s\in\{t+1,\dots,T\}}\left\{c_{t+1,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\right\}\right\}\\
&\qquad= \frac{16(2\gamma+5)}{3(1-2\gamma)\lambda_t}\max_{s\in\{t,\dots,T\}}\left\{c_{t,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}_n[\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t)]^2\\
&\qquad\le \max\left\{\frac{16(2\gamma+5)^2|I_t(\boldsymbol{\theta}_t)|\lambda_t^2}{27\tau_t},\ \frac{(2\gamma+5)C_t}{3}\max_{s\in\{t+1,\dots,T\}}\left\{c_{t+1,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\right\}\right\}\\
&\qquad= \frac{16(2\gamma+5)^2}{27}\max_{s\in\{t,\dots,T\}}\left\{c_{t,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\right\},
\end{aligned}
$$

and

$$
\begin{aligned}
&E[\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t)]^2\\
&\qquad\le \frac{(1-2\gamma)^2}{144\max_{s\in\{t,\dots,T\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}}\Big(\sum_{j=1}^{J_t}\bar{w}_{tj}|\hat{\theta}_{tj}-\theta_{tj}|\Big)^2+\mathbb{E}_n[\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t)]^2\\
&\qquad\le \frac{64(2\gamma+5)^2}{81}\max_{s\in\{t,\dots,T\}}\left\{c_{t,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\right\}.
\end{aligned}
$$

This completes the proof. □

Proof of Lemma 2.

Similarly as in the proof of Lemma 1, we prove the results using induction. Consider fixed n and fixed $\boldsymbol{\theta}_T\in\Theta_T$. Since $E[\Phi_{T2}^{\mathsf{T}}(H_T,A_T)|H_T]=\mathbf{0}$ a.s., we have $E(\Phi_{T1}\Phi_{T2}^{\mathsf{T}})=\mathbf{0}_{J_{T1}\times J_{T2}}$. On the event $\Omega_{T,1}(\boldsymbol{\theta}_T)$, we have

$$
\begin{aligned}
&\mathbb{E}_n\big[\Phi_T^{\mathsf{T}}(\boldsymbol{\theta}_T-\hat{\boldsymbol{\theta}}_T)\Phi_{T2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{T2}-\boldsymbol{\theta}_{T2})\big]\\
&\quad=(E-\mathbb{E}_n)[\Phi_T^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_T)\Phi_{T2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{T2}-\boldsymbol{\theta}_{T2})]-E[\Phi_{T2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{T2}-\boldsymbol{\theta}_{T2})]^2\\
&\quad\le \max_{j,k\in\{1,\dots,J_T\}}\Big|(E-\mathbb{E}_n)\Big(\frac{\phi_{Tj}\phi_{Tk}}{\bar{w}_{Tj}\bar{w}_{Tk}}\Big)\Big|\Big(\sum_{j=1}^{J_T}\bar{w}_{Tj}|\hat{\theta}_{Tj}-\theta_{Tj}|\Big)\Big(\sum_{j=J_{T1}+1}^{J_T}\bar{w}_{Tj}|\hat{\theta}_{Tj}-\theta_{Tj}|\Big)\\
&\qquad-E[\Phi_{T2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{T2}-\boldsymbol{\theta}_{T2})]^2\\
&\quad\le \frac{(1-2\gamma)(2\gamma+5)}{27}\lambda_T\Big(\sum_{j=J_{T1}+1}^{J_T}\bar{w}_{Tj}|\hat{\theta}_{Tj}-\theta_{Tj}|\Big)-E[\Phi_{T2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{T2}-\boldsymbol{\theta}_{T2})]^2,
\end{aligned}
$$

where the last inequality follows from the definition of $\Omega_{T,1}(\boldsymbol{\theta}_T)$ and (C.8). Note that (S.1) holds for any $\boldsymbol{\theta}_T \in \mathbb{R}^{J_T}$. In particular, with $(\hat{\boldsymbol{\theta}}_{T1}^\mathsf{T}, \boldsymbol{\theta}_{T2}^\mathsf{T})^\mathsf{T}$, on the event $\Omega_{T,1}(\boldsymbol{\theta}_T) \cap \Omega_{T,2}(\boldsymbol{\theta}_T)$, we have

$$
\begin{aligned}
0 &\le 2\mathbb{E}_n[(Y_T - \Phi_T^\mathsf{T}\hat{\boldsymbol{\theta}}_T)\Phi_{T2}^\mathsf{T}(\hat{\boldsymbol{\theta}}_{T2} - \boldsymbol{\theta}_{T2})] + \lambda_T \sum_{j=J_{T1}+1}^{J_T} w_{Tj}|\theta_{Tj}| - \lambda_T \sum_{j=J_{T1}+1}^{J_T} w_{Tj}|\hat{\theta}_{Tj}| \\
&\le \frac{4\gamma+1}{3}\lambda_T\Big(\sum_{j=J_{T1}+1}^{J_T} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}|\Big) + \lambda_T \sum_{j=J_{T1}+1}^{J_T} w_{Tj}|\theta_{Tj}| - \lambda_T \sum_{j=J_{T1}+1}^{J_T} w_{Tj}|\hat{\theta}_{Tj}| \\
&\quad + \frac{2(1-2\gamma)(2\gamma+5)}{27}\lambda_T\Big(\sum_{j=J_{T1}+1}^{J_T} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}|\Big) - 2E[\Phi_{T2}^\mathsf{T}(\hat{\boldsymbol{\theta}}_{T2} - \boldsymbol{\theta}_{T2})]^2 \\
&\le \frac{4(2\gamma+3)(4-\gamma)}{27}\lambda_T\Big(\sum_{j\in I_{T2}(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}|\Big) \\
&\quad - \frac{4(1-2\gamma)(2-\gamma)}{27}\lambda_T\Big(\sum_{j\in I_{T2}^c(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj}|\Big) - 2E[\Phi_{T2}^\mathsf{T}(\hat{\boldsymbol{\theta}}_{T2} - \boldsymbol{\theta}_{T2})]^2. \qquad \text{(S.18)}
\end{aligned}
$$

This implies

$$
\sum_{j\in I_{T2}^c(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj}| \le \frac{(2\gamma+3)(4-\gamma)}{(1-2\gamma)(2-\gamma)}\Big(\sum_{j\in I_{T2}(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}|\Big) \qquad \text{(S.19)}
$$

$$
\text{and } E[\Phi_{T2}^\mathsf{T}(\hat{\boldsymbol{\theta}}_{T2} - \boldsymbol{\theta}_{T2})]^2 \le \frac{2(2\gamma+3)(4-\gamma)}{27}\lambda_T\Big(\sum_{j\in I_{T2}(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}|\Big). \qquad \text{(S.20)}
$$

Note that Assumption (A2) implies that

$$
E[\Phi_{T2}^\mathsf{T}(\hat{\boldsymbol{\theta}}_{T2} - \boldsymbol{\theta}_{T2})]^2 \ge \frac{\tau_T(\sum_{j\in I_{T2}(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}|)^2}{|I_{T2}(\boldsymbol{\theta}_T)|}. \qquad \text{(S.21)}
$$

Plugging (S.21) into (S.18) yields

$$
\begin{aligned}
0 &\le \frac{4(2\gamma+3)(4-\gamma)}{27}\lambda_T\Big(\sum_{j\in I_{T2}(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}|\Big) \\
&\quad - \frac{4(1-2\gamma)(2-\gamma)}{27}\lambda_T\Big(\sum_{j\in I_{T2}^c(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj}|\Big) - \frac{2\tau_T(\sum_{j\in I_{T2}(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}|)^2}{|I_{T2}(\boldsymbol{\theta}_T)|} \\
&\le \Big(\sum_{j\in I_{T2}(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}|\Big)\times \\
&\qquad \Big[\frac{4(2\gamma+3)(4-\gamma)}{27}\lambda_T - \frac{2\tau_T}{|I_{T2}(\boldsymbol{\theta}_T)|}\Big(\sum_{j\in I_{T2}(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}|\Big)\Big].
\end{aligned}
$$

Thus

$$
\sum_{j\in I_{T2}(\boldsymbol{\theta}_T)} \bar{w}_{Tj}|\hat{\theta}_{Tj} - \theta_{Tj}| \le \frac{2(2\gamma+3)(4-\gamma)|I_{T2}(\boldsymbol{\theta}_T)|\lambda_T}{27\tau_T}.
$$

This, together with (S.19) and (S.20), implies that

$$\sum_{j=J_{T1}+1}^{J_T} \bar{w}_{Tj}|\hat{\theta}_{Tj}| \le \Big[\frac{28(2\gamma+3)(4-\gamma)}{27(1-2\gamma)(2-\gamma)}\Big] \frac{|I_{T2}(\boldsymbol{\theta}_T)|\lambda_T}{\tau_T}$$

$$\text{and } E[\Phi_{T2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{T2}-\boldsymbol{\theta}_{T2})]^2 \le \Big[\frac{2(2\gamma+3)(4-\gamma)}{27}\Big]^2 \frac{|I_{T2}(\boldsymbol{\theta}_T)|\lambda_T^2}{\tau_T}.$$

Algebra suffices to show (C.10) and (C.11) for $t=T$. This also implies that

$$\begin{aligned}
&\mathbb{E}_n[\Phi_{T2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{T2}-\boldsymbol{\theta}_{T2})]^2\\
&\le \max_{j,k\in\{1,\dots,J_t\}} \Big|(\mathbb{E}_n - E)\Big(\frac{\phi_{tj}\phi_{tk}}{\bar{w}_{tj}\bar{w}_{tk}}\Big)\Big|\Big(\sum_{j=J_{t1}+1}^{J_t} \bar{w}_{tj}|\hat{\theta}_{tj}-\theta_{tj}|\Big)^2 + E[\Phi_{t2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{t2}-\boldsymbol{\theta}_{t2})]^2\\
&\le \frac{(1-2\gamma)^2\tau_T}{144|I_T(\boldsymbol{\theta}_T)|}\Big(\Big[\frac{28(2\gamma+3)(4-\gamma)}{27(1-2\gamma)(2-\gamma)}\Big]\frac{|I_{T2}(\boldsymbol{\theta}_T)|\lambda_T}{\tau_T}\Big)^2 + \Big[\frac{2(2\gamma+3)(4-\gamma)}{27}\Big]^2\frac{|I_{T2}(\boldsymbol{\theta}_T)|\lambda_T^2}{\tau_T}\\
&\le 2\Big[\frac{2(2\gamma+3)(4-\gamma)}{27}\Big]^2\frac{|I_{T2}(\boldsymbol{\theta}_T)|\lambda_T^2}{\tau_T}.
\end{aligned}$$

Now we prove the results for $t<T$. Suppose

$$\sum_{j=J_{s1}+1}^{J_s} \bar{w}_{sj}|\hat{\theta}_{sj}-\theta_{sj}| \le \Big[\frac{81}{(1-2\gamma)^2}-3\Big]\lambda_s^{-1}\max_{s'\in\{s,\dots,T\}}\Big\{\bar{c}_{s,s'}\frac{|I_{s'2}(\boldsymbol{\theta}_{s'})|\lambda_{s'}^2}{\tau_{s'}}\Big\}$$

$$E[\Phi_{s2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{s2}-\boldsymbol{\theta}_{s2})]^2 \le \Big[3-\frac{(1-2\gamma)^2}{9}\Big]^2\max_{s'\in\{s,\dots,T\}}\Big\{\bar{c}_{s,s'}\frac{|I_{s'2}(\boldsymbol{\theta}_{s'})|\lambda_{s'}^2}{\tau_{s'}}\Big\},$$

and

$$\begin{aligned}
&\mathbb{E}_n[\Phi_{s2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{s2}-\boldsymbol{\theta}_{s2})]^2\\
&\quad\le \Big[\frac{81\max_{s'\in\{s,\dots,T\}}\{\bar{c}_{s,s'}/c_{s,s'}\}}{16(1-2\gamma)^2}+1\Big]\Big[3-\frac{(1-2\gamma)^2}{9}\Big]^2\max_{s'\in\{s,\dots,T\}}\Big\{\bar{c}_{s,s'}\frac{|I_{s'2}(\boldsymbol{\theta}_{s'})|\lambda_{s'}^2}{\tau_{s'}}\Big\},
\end{aligned}$$

for $s'=s+1,\dots,T$ and $s=T,\dots,t+1$, on the event $\cap_{s=t+1}^T\Big\{\Omega_{s,1}(\boldsymbol{\theta}_s,\dots,\boldsymbol{\theta}_T)\cap \Omega_{s,2}(\boldsymbol{\theta}_s,\dots,\boldsymbol{\theta}_T)\cap\Omega_{s+1,3}(\boldsymbol{\theta}_{s+1},\dots,\boldsymbol{\theta}_T)\Big\}$.

Note that (S.10) holds for any $\boldsymbol{\theta}_t\in\mathbb{R}^{J_t}$. In particular, with $\boldsymbol{\theta}_t=(\hat{\boldsymbol{\theta}}_{t1}^{\mathsf{T}},\boldsymbol{\theta}_{t2}^{\mathsf{T}})^{\mathsf{T}}$, we have on the event $\cap_{s=t}^T\{\Omega_{s,1}(\boldsymbol{\theta}_s,\dots,\boldsymbol{\theta}_T)\cap\Omega_{s,2}(\boldsymbol{\theta}_s,\dots,\boldsymbol{\theta}_T)\cap\Omega_{s+1,3}(\boldsymbol{\theta}_{s+1},\dots,\boldsymbol{\theta}_T)\}$,

$$\begin{aligned}
0 &\le 2\mathbb{E}_n[(\tilde{Y}_t(\hat{\boldsymbol{\theta}}_{t+1},\dots,\hat{\boldsymbol{\theta}}_T)-\Phi_t^{\mathsf{T}}\hat{\boldsymbol{\theta}}_t)\Phi_{t2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{t2}-\boldsymbol{\theta}_{t2})] + \lambda_t\sum_{j=J_{t1}+1}^{J_t} w_{tj}|\theta_{tj}| - \lambda_t\sum_{j=J_{t1}+1}^{J_t} w_{tj}|\hat{\theta}_{tj}|\\
&\le 2\mathbb{E}_n[(\tilde{Y}_t(\boldsymbol{\theta}_{t+1},\dots,\boldsymbol{\theta}_T)-\Phi_t^{\mathsf{T}}\boldsymbol{\theta}_t)\Phi_{t2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{t2}-\boldsymbol{\theta}_{t2})] + \lambda_t\sum_{j=J_{t1}+1}^{J_t} w_{tj}|\theta_{tj}| - \lambda_t\sum_{j=J_{t1}+1}^{J_t} w_{tj}|\hat{\theta}_{tj}|\\
&\quad - E[\Phi_{t2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{t2}-\boldsymbol{\theta}_{t2})]^2 + \mathbb{E}_n[\tilde{Y}_t(\hat{\boldsymbol{\theta}}_{t+1},\dots,\hat{\boldsymbol{\theta}}_T)-\tilde{Y}_t(\boldsymbol{\theta}_{t+1},\dots,\boldsymbol{\theta}_T)]^2\\
&\quad + \{(\mathbb{E}_n-E)[\Phi_{t2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{t2}-\boldsymbol{\theta}_{t2})]^2 + 2(E-\mathbb{E}_n)[\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_t)\Phi_{t2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{t2}-\boldsymbol{\theta}_{t2})]\},
\end{aligned} \tag{S.22}$$

where the last inequality follows from AM-GM inequality and the condition that $E(\Phi_{t1}\Phi_{t2}^{\mathsf{T}}) = \mathbf{0}$. Below, we derive upper bounds for the last two terms of (S.22). Since $\Phi_{s1}(H_s)$ does not involve treatment A_s, it is easy to see that $Y_t(\boldsymbol{\theta}_{t+1}, \ldots, \boldsymbol{\theta}_T)$ defined in (C.7) can be re-written as

$$\tilde{Y}_t(\boldsymbol{\theta}_{t+1}, \ldots, \boldsymbol{\theta}_T) = Y_t + \sum_{s=t+1}^{T} \Big[Y_s + \max_{a_s} \Phi_{s2}^{\mathsf{T}}(H_s, a_s)\boldsymbol{\theta}_{s2} - \Phi_{s2}^{\mathsf{T}}(H_s, A_s)\boldsymbol{\theta}_{s2} \Big].$$

Note that on event $\Omega_{s,3}(\boldsymbol{\theta}_s, \ldots, \boldsymbol{\theta}_T)$

$$\begin{aligned}
&\max_{j,k\in\{1,\ldots,J_t\}} \Big| (E - \mathbb{E}_n)\Big(\sum_{a_s\in\mathcal{A}_s} \frac{\phi_{sj}(H_s,a_s)\phi_{sk}(H_s,a_s)}{\bar{w}_{sj}\bar{w}_{sk}} \Big) \Big| \\
&\qquad\qquad \leq \frac{(1-2\gamma)^2 |\mathcal{A}_s| \max_{s\in\{t,\ldots,T\}}\{\bar{c}_{t,s}/c_{t,s}\}\lambda_t^2}{144 \max_{s\in\{t,\ldots,T\}} \big\{ \big[\max_{s\in\{t,\ldots,T\}}\{\bar{c}_{t,s}\lambda_t^2/c_{t,s}\} \big] I_{s2}(\boldsymbol{\theta}_s)/\tau_s \big\}} \\
&\qquad\qquad \leq \frac{(1-2\gamma)^2 |\mathcal{A}_s| \max_{s'\in\{s,\ldots,T\}}\{\bar{c}_{s,s'}/c_{s,s'}\}\lambda_s^2}{144 \max_{s'\in\{s,\ldots,T\}} \big\{ \bar{c}_{s,s'} I_{s'2}(\boldsymbol{\theta}_{s'})\lambda_{s'}^2/\tau_{s'} \big\}}.
\end{aligned}$$

Using the similar arguments as those in the proof of Lemma 1, we can show that

$$\begin{aligned}
&\mathbb{E}_n[\tilde{Y}_t(\hat{\boldsymbol{\theta}}_{t+1}, \ldots, \hat{\boldsymbol{\theta}}_T) - \tilde{Y}_t(\boldsymbol{\theta}_{t+1}, \ldots, \boldsymbol{\theta}_T)]^2 \\
&= \mathbb{E}_n\Bigg[\sum_{s=t+1}^{T} \Big[\max_{a_s} \Phi_{s2}^{\mathsf{T}}(H_s,a_s)\hat{\boldsymbol{\theta}}_{s2} - \max_{a_s} \Phi_{s2}^{\mathsf{T}}(H_s,a_s)\boldsymbol{\theta}_{s2} - \Phi_{s2}^{\mathsf{T}}(H_s,A_s)(\hat{\boldsymbol{\theta}}_{s2} - \boldsymbol{\theta}_{s2}) \Big] \Bigg]^2 \\
&\leq 2(T-t) \sum_{s=t+1}^{T} \Big\{ \mathbb{E}_n\Big[\max_{a_s} \big| \Phi_{s2}^{\mathsf{T}}(H_s,a_s)(\hat{\boldsymbol{\theta}}_{s2} - \boldsymbol{\theta}_{s2}) \big|^2 \Big] + \mathbb{E}_n\Big[\Phi_{s2}^{\mathsf{T}}(H_s,A_s)(\hat{\boldsymbol{\theta}}_{s2} - \boldsymbol{\theta}_{s2}) \Big]^2 \Big\} \\
&\leq C_{t2} \max_{s\in\{t+1,\ldots,T\}} \Big\{ \bar{c}_{t+1,s} \frac{|I_{s2}(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s} \Big\}, \qquad \text{(S.23)}
\end{aligned}$$

where $C_{t2} = 2(T-t)^2(S+1)\Big[\frac{81 \max_{s\in\{t+1,\ldots,T\}}\{\bar{c}_{t+1,s}/c_{t+1,s}\}}{16(1-2\gamma)^2} + 1\Big]\Big[3 - \frac{(1-2\gamma)^2}{9}\Big]^2$. In addition,

$$\begin{aligned}
&(\mathbb{E}_n - E)[\Phi_{t2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{t2} - \boldsymbol{\theta}_{t2})]^2 + 2(E - \mathbb{E}_n)[\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t)\Phi_{t2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{t2} - \boldsymbol{\theta}_{t2})] \\
&\qquad \leq 3 \max_{j,k\in\{1,\ldots,J_t\}} \Big| (E - \mathbb{E}_n)\Big(\frac{\phi_{tj}\phi_{tk}}{\bar{w}_{tj}\bar{w}_{tk}} \Big) \Big| \Big(\sum_{j=1}^{J_t} \bar{w}_{tj}|\hat{\theta}_{tj} - \theta_{tj}| \Big) \Big(\sum_{j=J_{t1}+1}^{J_t} \bar{w}_{tj}|\hat{\theta}_{tj} - \theta_{tj}| \Big) \\
&\qquad \leq \frac{(1-2\gamma)(2\gamma+5)}{9} \lambda_t \Big(\sum_{j=J_{t1}+1}^{J_t} \bar{w}_{tj}|\hat{\theta}_{tj} - \theta_{tj}| \Big), \qquad \text{(S.24)}
\end{aligned}$$

where the last inequality follows from the definition of $\Omega_{t,1}(\boldsymbol{\theta}_t, \ldots, \boldsymbol{\theta}_T)$, (C.8), and (C.4). Plugging (S.23) and (S.24) into (S.22) yields

$$\begin{aligned}
0 \leq \frac{4\gamma+1}{3} \lambda_t \Big(\sum_{j=J_{t1}+1}^{J_t} \bar{w}_{tj}|\hat{\theta}_{tj} - \theta_{tj}| \Big) + \lambda_t \sum_{j=J_{t1}+1}^{J_t} w_{tj}|\theta_{tj}| - \lambda_t \sum_{j=J_{t1}+1}^{J_t} w_{tj}|\hat{\theta}_{tj}| \\
- E[\Phi_{t2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{t2} - \boldsymbol{\theta}_{t2})]^2
\end{aligned}$$

$$+ C_{t2} \max_{s\in\{t+1,\ldots,T\}} \Big\{ \bar{c}_{t+1,s} \frac{|I_{s2}(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s} \Big\} + \frac{(1-2\gamma)(2\gamma+5)}{9} \lambda_t \Big(\sum_{j=J_{t1}+1}^{J_t} \bar{w}_{tj} |\hat{\theta}_{tj} - \theta_{tj}| \Big)$$

$$\leq \Big[2 - \frac{(1-2\gamma)^2}{9} \Big] \lambda_t \Big(\sum_{j\in I_{t2}(\boldsymbol{\theta}_t)} \bar{w}_{tj} |\hat{\theta}_{tj} - \theta_{tj}| \Big) - \frac{(1-2\gamma)^2}{9} \lambda_t \Big(\sum_{j\in I_{t2}^c(\boldsymbol{\theta}_t)} \bar{w}_{tj} |\hat{\theta}_{tj}| \Big)$$

(S.25)

$$- E[\Phi_{t2}^{\mathsf{T}} (\hat{\boldsymbol{\theta}}_{t2} - \boldsymbol{\theta}_{t2})]^2 + C_{t2} \max_{s\in\{t+1,\ldots,T\}} \Big\{ \bar{c}_{t+1,s} \frac{|I_{s2}(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s} \Big\}.$$

This implies

$$\sum_{j\in I_{t2}^c(\boldsymbol{\theta}_t)} \bar{w}_{tj} |\hat{\theta}_{tj}| \leq \Big[\frac{18}{(1-2\gamma)^2} - 1 \Big] \Big(\sum_{j\in I_{t2}(\boldsymbol{\theta}_t)} \bar{w}_{tj} |\hat{\theta}_{tj} - \theta_{tj}| \Big) + \frac{9C_{t2}}{(1-2\gamma)^2 \lambda_t} \max_{s\in\{t+1,\ldots,T\}} \Big\{ \bar{c}_{t+1,s} \frac{|I_{s2}(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s} \Big\}$$

and

$$E[\Phi_{t2}^{\mathsf{T}} (\hat{\boldsymbol{\theta}}_{t2} - \boldsymbol{\theta}_{t2})]^2 \leq \Big[2 - \frac{(1-2\gamma)^2}{9} \Big] \lambda_t \Big(\sum_{j\in I_{t2}(\boldsymbol{\theta}_t)} \bar{w}_{tj} |\hat{\theta}_{tj} - \theta_{tj}| \Big) + C_{t2} \max_{s\in\{t+1,\ldots,T\}} \Big\{ \bar{c}_{t+1,s} \frac{|I_{s2}(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s} \Big\}.$$

If $I_{t2}(\boldsymbol{\theta}_t)$ is empty (i.e., $\boldsymbol{\theta}_{t2} = \mathbf{0}$), then (C.10) and (C.11) hold. If $I_{t2}(\boldsymbol{\theta}_t)$ is non-empty, define the sets

$$\acute{\Theta}_1(\boldsymbol{\theta}_t) = \Big\{ \tilde{\boldsymbol{\theta}}_t \in \mathbb{R}^{J_{t2}} : \sum_{j\in I_{t2}^c(\boldsymbol{\theta}_t)} \bar{w}_{tj} |\tilde{\theta}_{tj}| \leq \Big[\frac{18}{(1-2\gamma)^2} - 1 \Big] \Big(\sum_{j\in I_{t2}(\boldsymbol{\theta}_t)} \bar{w}_{tj} |\tilde{\theta}_{tj} - \theta_{tj}| \Big) + \frac{9C_{t2}}{(1-2\gamma)^2 \lambda_t} \max_{s\in\{t+1,\ldots,T\}} \Big\{ \bar{c}_{t+1,s} \frac{|I_{s2}(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s} \Big\} \Big\},$$

$$\acute{\Theta}_2(\boldsymbol{\theta}_t) = \Big\{ \tilde{\boldsymbol{\theta}}_t \in \mathbb{R}^{J_{t2}} : \sum_{j\in I_{t2}(\boldsymbol{\theta}_t)} w_{tj} |\tilde{\theta}_{tj} - \theta_{tj}| > \max \Big\{ \Big[3 - \frac{(1-2\gamma)^2}{9} \Big] \frac{|I_{t2}(\boldsymbol{\theta}_t)|\lambda_t}{\tau_t}, \frac{C_{t2}}{\lambda_t} \max_{s\in\{t+1,\ldots,T\}} \Big\{ \bar{c}_{t+1,s} \frac{|I_{s2}(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s} \Big\} \Big\} \Big\}.$$

Thus, $\hat{\boldsymbol{\theta}}_{t2} \in \acute{\Theta}_1(\boldsymbol{\theta}_t)$ on the event $\Omega_{t,2}(\boldsymbol{\theta}_t, \ldots, \boldsymbol{\theta}_T)$. Next, for any $\boldsymbol{\theta}_{t2} \in \Theta_t$ and $\tilde{\boldsymbol{\theta}}_{t2} \in \acute{\Theta}_1(\boldsymbol{\theta}_t)$, note that condition (3.9) by Assumption (A2) implies that

$$E[\Phi_{t2}^{\mathsf{T}} (\tilde{\boldsymbol{\theta}}_{t2} - \boldsymbol{\theta}_{t2})]^2 \geq \frac{\tau_t (\sum_{j\in I_{t2}(\boldsymbol{\theta}_t)} \bar{w}_{tj} |\tilde{\theta}_{tj} - \theta_{tj}|)^2}{|I_{t2}(\boldsymbol{\theta}_t)|}. \tag{S.26}$$

In addition, on the event $\cap_{s=t}^T \{ \Omega_{s,1}(\boldsymbol{\theta}_s, \ldots, \boldsymbol{\theta}_T) \cap \Omega_{s,2}(\boldsymbol{\theta}_s, \ldots, \boldsymbol{\theta}_T) \cap \Omega_{s+1,3}(\boldsymbol{\theta}_{s+1}, \ldots, \boldsymbol{\theta}_T) \}$,

$$\sup_{\tilde{\boldsymbol{\theta}}_t \in \acute{\Theta}_1(\boldsymbol{\theta}_t) \cap \acute{\Theta}_2(\boldsymbol{\theta}_t)} \Big\{ \Big[2 - \frac{(1-2\gamma)^2}{9} \Big] \lambda_t \Big(\sum_{j\in I_{t2}(\boldsymbol{\theta}_t)} \bar{w}_{tj} |\tilde{\theta}_{tj} - \theta_{tj}| \Big) - \frac{(1-2\gamma)^2}{9} \lambda_t \Big(\sum_{j\in I_{t2}^c(\boldsymbol{\theta}_t)} \bar{w}_{tj} |\tilde{\theta}_{tj}| \Big) - E[\Phi_{t2}^{\mathsf{T}} (\tilde{\boldsymbol{\theta}}_{t2} - \boldsymbol{\theta}_{t2})]^2$$

$$
\begin{aligned}
&\qquad\qquad + C_{t2} \max_{s\in\{t+1,\ldots,T\}} \Big\{ \bar{c}_{t+1,s} \frac{|I_{s2}(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s} \Big\} \Big\} \\
\leq & \sup_{\tilde{\boldsymbol{\theta}}_t \in \acute{\Theta}_1(\boldsymbol{\theta}_t)\cap\acute{\Theta}_2(\boldsymbol{\theta}_t)} \Big\{ \Big[2 - \frac{(1-2\gamma)^2}{9}\Big] \lambda_t \Big(\sum_{j\in I_{t2}(\boldsymbol{\theta}_t)} \bar{w}_{tj} |\tilde{\theta}_{tj} - \theta_{tj}| \Big) \\
&\quad - \frac{\tau_t}{|I_{t2}(\boldsymbol{\theta}_t)|} \Big(\sum_{j\in I_{t2}(\boldsymbol{\theta}_t)} \bar{w}_{tj} |\tilde{\theta}_{tj} - \theta_{tj}| \Big)^2 + C_{t2} \max_{s\in\{t+1,\ldots,T\}} \Big\{ \bar{c}_{t+1,s} \frac{|I_{s2}(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s} \Big\} \Big\} \\
\leq & \sup_{\tilde{\boldsymbol{\theta}}_t \in \acute{\Theta}_1(\boldsymbol{\theta}_t)\cap\acute{\Theta}_2(\boldsymbol{\theta}_t)} \Big\{ \Big(\sum_{j\in I_{t2}(\boldsymbol{\theta}_t)} \bar{w}_{tj} |\tilde{\theta}_{tj} - \theta_{tj}| \Big) \times \\
&\qquad \Big[\Big[3 - \frac{(1-2\gamma)^2}{9}\Big] \lambda_t - \frac{\tau_t}{|I_{t2}(\boldsymbol{\theta}_t)|} \Big(\sum_{j\in I_{t2}(\boldsymbol{\theta}_t)} \bar{w}_{tj} |\tilde{\theta}_{tj} - \theta_{tj}| \Big) \Big] \Big\} \\
< & \, 0,
\end{aligned}
$$

where the first inequality follows from (S.26), and the last two inequalities follow from the definition of $\acute{\Theta}_2(\boldsymbol{\theta}_t)$.

Since $\hat{\boldsymbol{\theta}}_{t2}$ satisfies inequality (S.25), we have $\hat{\boldsymbol{\theta}}_{t2} \in \acute{\Theta}_1(\boldsymbol{\theta}_t) \cap \acute{\Theta}_2(\boldsymbol{\theta}_t)^C$ on the event $\cap_{s=t}^T \big\{ \Omega_{s,1}(\boldsymbol{\theta}_s, \ldots, \boldsymbol{\theta}_T) \cap \Omega_{s,2}(\boldsymbol{\theta}_s, \ldots, \boldsymbol{\theta}_T) \cap \Omega_{s+1,3}(\boldsymbol{\theta}_{s+1}, \ldots, \boldsymbol{\theta}_T) \big\}$. Algebra suffices to show

$$
\begin{aligned}
&\sum_{j=J_{t1}+1}^{J_s} \bar{w}_{tj} |\hat{\theta}_{tj} - \theta_{tj}| \\
&\quad \leq \max \Big\{ \Big[\frac{81}{(1-2\gamma)^2} - 3 \Big] \frac{|I_{t2}(\boldsymbol{\theta}_t)|\lambda_t}{\tau_t}, \frac{27 C_{t2}}{(1-2\gamma)^2 \lambda_t} \max_{s\in\{t+1,\ldots,T\}} \Big\{ \bar{c}_{t+1,s} \frac{|I_{s2}(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s} \Big\} \Big\} \\
&\quad = \Big[\frac{81}{(1-2\gamma)^2} - 3 \Big] \lambda_t^{-1} \max_{s\in\{t,\ldots,T\}} \Big\{ \bar{c}_{t,s} \frac{|I_{s2}(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}, \Big\},
\end{aligned}
$$

and

$$
\begin{aligned}
&E\big[\Phi_{t2}^\intercal (\hat{\boldsymbol{\theta}}_{t2} - \boldsymbol{\theta}_{t2})\big]^2 \\
&\leq \max \Big\{ \Big[3 - \frac{(1-2\gamma)^2}{9}\Big]^2 \frac{|I_{t2}(\boldsymbol{\theta}_t)|\lambda_t^2}{\tau_t}, \Big[3 - \frac{(1-2\gamma)^2}{9}\Big] C_{t2} \max_{s\in\{t+1,\ldots,T\}} \Big\{ \bar{c}_{t+1,s} \frac{|I_{s2}(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s} \Big\} \Big\} \\
&= \Big[3 - \frac{(1-2\gamma)^2}{9}\Big]^2 \max_{s\in\{t,\ldots,T\}} \Big\{ \bar{c}_{t,s} \frac{|I_{s2}(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s} \Big\},
\end{aligned}
$$

where $\bar{c}_{t,t} = 1$ and

$$
\begin{aligned}
\bar{c}_{t,s} &= 9 C_{t2} \bar{c}_{t+1,s} / [27 - (1-2\gamma)^2] \\
&= 2(T-t)^2 (S+1) \Big[\frac{81 \max_{s\in\{t+1,\ldots,T\}} \{\bar{c}_{t+1,s}/c_{t+1,s}\}}{16(1-2\gamma)^2} + 1 \Big] \Big[3 - \frac{(1-2\gamma)^2}{9}\Big] \bar{c}_{t+1,s},
\end{aligned}
$$

for $s = t+1, \ldots, T$. In addition, since

$$
\begin{aligned}
&\max_{j,k\in\{1,\ldots,J_t\}} \Big| (\mathbb{E}_n - E) \Big(\frac{\phi_{tj}\phi_{tk}}{\bar{w}_{tj}\bar{w}_{tk}} \Big) \Big| \\
&\qquad\qquad \leq \frac{(1-2\gamma)^2 \max_{s\in\{t,\ldots,T\}} \{\bar{c}_{t,s}/c_{t,s}\} \lambda_t^2}{144 \max_{s\in\{t,\ldots,T\}} \big\{ \big[\max_{s\in\{t,\ldots,T\}} \{\bar{c}_{t,s}\lambda_t^2/c_{t,s}\} \big] I_{s2}(\boldsymbol{\theta}_s)/\tau_s \big\}}
\end{aligned}
$$

$$\leq \frac{(1-2\gamma)^2 \max_{s\in\{t,\ldots,T\}}\{\bar{c}_{t,s}/c_{t,s}\}\lambda_t^2}{144\max_{s\in\{t,\ldots,T\}}\left\{\bar{c}_{t,s}I_{s2}(\boldsymbol{\theta}_s)\lambda_s^2/\tau_s\right\}},$$

it is easy to verify that

$$\begin{aligned}
&\mathbb{E}_n[\Phi_{t2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{t2}-\boldsymbol{\theta}_{t2})]^2\\
&\leq \max_{j,k\in\{1,\ldots,J_t\}}\left|(\mathbb{E}_n-E)\Big(\frac{\phi_{tj}\phi_{tk}}{\bar{w}_{tj}\bar{w}_{tk}}\Big)\right|\Big(\sum_{j=J_{t1}+1}^{J_t}\bar{w}_{tj}|\hat{\theta}_{tj}-\theta_{tj}|\Big)^2+E[\Phi_{t2}^{\mathsf{T}}(\hat{\boldsymbol{\theta}}_{t2}-\boldsymbol{\theta}_{t2})]^2\\
&\leq \left[\frac{81\max_{s\in\{t,\ldots,T\}}\{\bar{c}_{t,s}/c_{t,s}\}}{16(1-2\gamma)^2}+1\right]\left[3-\frac{(1-2\gamma)^2}{9}\right]^2\max_{s\in\{t,\ldots,T\}}\left\{\bar{c}_{t,s}\frac{|I_{s2}(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\right\}.
\end{aligned}$$

This completes the proof. □

Proof of Lemma 3.

Note that $||\phi_{tj}\phi_{tk}/(\bar{w}_{tj}\bar{w}_{tk})-E[\phi_{tj}\phi_{tk}/(\bar{w}_{tj}\bar{w}_{tk})]||_\infty \leq 2u^2$ and $E[\phi_{tj}\phi_{tk}/(\bar{w}_{tj}\bar{w}_{tk})]^2 \leq b^2u^2$ for all $j,k\in\{1,\ldots,J_t\}$ by Assumptions (A3) and (A4). Next we apply Bernstein's inequality in Lemma S.1(a) with $\zeta_i=\pm[\phi_{tj}\phi_{tk}/(\bar{w}_{tj}\bar{w}_{tk})-E(\phi_{tj}\phi_{tk}/(\bar{w}_{tj}\bar{w}_{tk})]$ and $\kappa=(1-2\gamma)^2n/[144\max_{s\in\{t,\ldots,T\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}]$. Using the union bound argument, we obtain

$$\begin{aligned}
&\mathbf{P}(\{\Omega_{t,1}(\boldsymbol{\theta}_t,\ldots,\boldsymbol{\theta}_T)\}^C)\\
&\leq J_t(J_t+1)\times\\
&\exp\left(-\frac{(1-2\gamma)^4n}{2u^2\max_{s\in\{t,\ldots,T\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}[144^2b^2\max_{s\in\{t,\ldots,T\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}+96(1-2\gamma)^2]}\right)\\
&\leq \exp(-\varphi)/3,
\end{aligned}$$

where the second inequality follows from the definition of Θ in (C.1). □

Proof of Lemma 4.

We first show the result for $t=T$. For any $\boldsymbol{\theta}_T\in\Theta_T$, $\max_j\left|E\Big[\Phi_T^{\mathsf{T}}(\boldsymbol{\theta}_T-\boldsymbol{\theta}_T^*)\frac{\phi_{Tj}}{\bar{w}_{Tj}}\Big]\right|\leq \gamma b\lambda_T$ under Assumption (A4). Since $\boldsymbol{\theta}_T^*$ minimizes $E[Y_T-\Phi_T^{\mathsf{T}}\boldsymbol{\theta}_T]^2$, we have $E[(Y_T-\Phi_T^{\mathsf{T}}\boldsymbol{\theta}_T^*)\phi_{Tj}/\bar{w}_{Tj}]=0$ for $j=\{1,\ldots,J_T\}$. Thus,

$$\max_{j\in\{1,\ldots,J_T\}}\left|E\Big[\Big(Y_T-\Phi_T^{\mathsf{T}}\boldsymbol{\theta}_T\Big)\frac{\phi_{Tj}}{\bar{w}_{Tj}}\Big]\right|=\max_{j\in\{1,\ldots,J_T\}}\left|E\Big[\Phi_T^{\mathsf{T}}(\boldsymbol{\theta}_T-\boldsymbol{\theta}_T^*)\frac{\phi_{Tj}}{\bar{w}_{Tj}}\Big]\right|\leq\gamma\lambda_T b.$$

This implies

$$\begin{aligned}
&\max_{j\in\{1,\ldots,J_T\}}\left|\mathbb{E}_n\Big[\Big(Y_T-\Phi_T^{\mathsf{T}}\boldsymbol{\theta}_T\Big)\frac{\phi_{Tj}}{\bar{w}_{Tj}}\Big]\right|\\
&\leq \max_{j\in\{1,\ldots,J_T\}}\left|(\mathbb{E}_n-E)\Big[\epsilon_T\frac{\phi_{Tj}}{\bar{w}_{Tj}}\Big]\right|+\max_{j\in\{1,\ldots,J_T\}}\left|(\mathbb{E}_n-E)\Big[(Q_T^o-\Phi_T^{\mathsf{T}}\boldsymbol{\theta}_T)\frac{\phi_{Tj}}{\bar{w}_{Tj}}\Big]\right|+\gamma\lambda_T b.
\end{aligned}$$

Under Assumptions (A1) and (A3), it can be shown that $E(\epsilon_{Ti}\phi_{Tj}/\bar{w}_{Tj})=0$ and $\sum_{i=1}^n E|(\epsilon_{Ti}\phi_{Tj}/\bar{w}_{Tj})^l|\leq l!n\sigma^2b^2(cu)^{l-2}/2$ for $j\in\{1,\ldots,J_T\}$ and all integers $l\geq 2$. Applying Bernstein's inequality in Lemma S.1(b), we obtain

$$\begin{aligned}
&\mathbf{P}\Big(\left|(\mathbb{E}_n-E)\Big[\epsilon_T\frac{\phi_{Tj}}{\bar{w}_{Tj}}\Big]\right|>\frac{1-2\gamma(3b-2)}{12}\lambda_T\Big)\\
&\qquad\leq 2\exp\Big(-\frac{[1-2\gamma(3b-2)]^2\lambda_T^2n}{288\sigma^2b^2+24c[1-2\gamma(3b-2)]u\lambda_T}\Big).
\end{aligned}$$

Similarly, under Assumption (A3), for any $\boldsymbol{\theta}_T \in \Theta_T^*$ and $j \in \{1,\ldots,J_T\}$, $\|(Q_T^o - \Phi_T^\mathsf{T}\boldsymbol{\theta}_T)\phi_{Tj}/\bar{w}_{Tj} - E((Q_T^o - \Phi_T^\mathsf{T}\boldsymbol{\theta}_T)\phi_{Tj}/\bar{w}_{Tj})\|_\infty \le 4\eta u$ and $E[(Q_T^o - \Phi_T^\mathsf{T}\boldsymbol{\theta}_T)\phi_{Tj}/\bar{w}_{Tj}]^2 \le 4\eta^2 b^2$. Then we have

$$\mathbf{P}\Big(\Big|(\mathbb{E}_n - E)\Big[(Q_T^o - \Phi_T^\mathsf{T}\boldsymbol{\theta}_T)\frac{\phi_{Tj}}{\bar{w}_{Tj}}\Big]\Big| > \frac{1-2\gamma(3b-2)}{12}\lambda_T\Big) \le 2\exp\Big(-\frac{[1-2\gamma(3b-2)]^2\lambda_T^2 n}{288(2\eta b)^2 + 32[1-2\gamma(3b-2)]u\eta\lambda_T}\Big).$$

The result follows from the union bound argument and condition (C.2).

Next, we show the results for $t < T$. For any $(\boldsymbol{\theta}_1^\mathsf{T},\ldots,\boldsymbol{\theta}_T^\mathsf{T})^\mathsf{T} \in \Theta$, note that

$$\begin{aligned}
&E\Big[\tilde{Y}_t(\boldsymbol{\theta}_{t+1},\ldots,\boldsymbol{\theta}_T) - \tilde{Y}_t(\boldsymbol{\theta}_{t+1}^*,\ldots,\boldsymbol{\theta}_T^*)\Big]^2 \\
&= E\Big\{\sum_{s=t+1}^T \Big[\max_{a_s}\Phi_s^\mathsf{T}(H_s,a_s)\boldsymbol{\theta}_s - \max_{a_s}\Phi_s^\mathsf{T}(H_s,a_s)\boldsymbol{\theta}_s^* - \Phi_s^\mathsf{T}(H_s,A_s)(\boldsymbol{\theta}_s - \boldsymbol{\theta}_s^*)\Big]\Big\}^2 \\
&\le 2(T-t)\sum_{s=t+1}^T \Big\{E\Big[\max_{a_s}\big[\Phi_s^\mathsf{T}(H_s,a_s)(\boldsymbol{\theta}_s - \boldsymbol{\theta}_s^*)\big]^2\Big] + E\Big[\Phi_s^\mathsf{T}(H_s,A_s)(\boldsymbol{\theta}_s - \boldsymbol{\theta}_s^*)\Big]^2\Big\} \\
&\le 2(T-t)\sum_{s=t+1}^T \Big\{E\Big[\textstyle\sum_{a_s}\big[\Phi_s^\mathsf{T}(H_s,a_s)(\boldsymbol{\theta}_s - \boldsymbol{\theta}_s^*)\big]^2\Big] + E\Big[\Phi_s^\mathsf{T}(H_s,A_s)(\boldsymbol{\theta}_s - \boldsymbol{\theta}_s^*)\Big]^2\Big\} \\
&\le 2(T-t)\sum_{s=t+1}^T \Big\{E\Big[\textstyle\sum_{a_s} p(a_s|H_s)S\big[\Phi_s^\mathsf{T}(H_s,a_s)(\boldsymbol{\theta}_s - \boldsymbol{\theta}_s^*)\big]^2\Big] \\
&\qquad\qquad + E\Big[\Phi_s^\mathsf{T}(H_s,A_s)(\boldsymbol{\theta}_s - \boldsymbol{\theta}_s^*)\Big]^2\Big\} \\
&\le 2(T-t)\sum_{s=t+1}^T \Big\{SE\Big[\Phi_s^\mathsf{T}(H_s,A_s)(\boldsymbol{\theta}_s - \boldsymbol{\theta}_s^*)\Big]^2 + E\Big[\Phi_s^\mathsf{T}(H_s,A_s)(\boldsymbol{\theta}_s - \boldsymbol{\theta}_s^*)\Big]^2\Big\} \\
&\le 2(T-t)(S+1)\gamma^2\sum_{s=t+1}^T \lambda_s^2 \\
&\le \frac{9}{16}\gamma^2\lambda_t^2,
\end{aligned}$$

where the last inequality holds under condition (C.4) and the fact that $c_{t,s} \ge 32(S+1)(T-t)^2/9$. Since $\boldsymbol{\theta}_t^*$ minimizes $E\big[\tilde{Y}_t(\boldsymbol{\theta}_{t+1}^*,\ldots,\boldsymbol{\theta}_T^*) - \Phi_t^\mathsf{T}\boldsymbol{\theta}_t\big]^2$, we have $E\big[\big(\tilde{Y}_t(\boldsymbol{\theta}_{t+1}^*,\ldots,\boldsymbol{\theta}_T^*) - \Phi_t^\mathsf{T}\boldsymbol{\theta}_t^*\big)\phi_{tj}\big] = 0$. Thus, for $j = 1,\ldots,J_t$, we have

$$\begin{aligned}
&\Big|E\Big[\Big(\tilde{Y}_t(\boldsymbol{\theta}_{t+1},\ldots,\boldsymbol{\theta}_T) - \Phi_t^\mathsf{T}\boldsymbol{\theta}_t\Big)\frac{\phi_{tj}}{\bar{w}_{tj}}\Big]\Big| \\
&\quad\le \Big|E\Big[\Big(\tilde{Y}_t(\boldsymbol{\theta}_{t+1},\ldots,\boldsymbol{\theta}_T) - \tilde{Y}_t(\boldsymbol{\theta}_{t+1}^*,\ldots,\boldsymbol{\theta}_T^*)\Big)\frac{\phi_{tj}}{\bar{w}_{tj}}\Big]\Big| + \Big|E\Big[\Phi_t^\mathsf{T}(\boldsymbol{\theta}_t - \boldsymbol{\theta}_t^*)\frac{\phi_{tj}}{\bar{w}_{tj}}\Big]\Big| \\
&\quad\le \frac{7}{4}\gamma b\lambda_t,
\end{aligned}$$

where the last inequality holds from Assumption (A4). Hence,

$$\begin{aligned}
&\max_{j\in\{1,\ldots,J_t\}}\Big|\mathbb{E}_n\Big[\Big(\tilde{Y}_t(\boldsymbol{\theta}_{t+1},\ldots,\boldsymbol{\theta}_T)-\Phi_t^{\mathsf{T}}\boldsymbol{\theta}_t\Big)\frac{\phi_{tj}}{\bar{w}_{tj}}\Big]\Big| \\
&\le \max_{j\in\{1,\ldots,J_t\}}\Big|(\mathbb{E}_n-E)\Big[\Big(\tilde{Y}_t(\boldsymbol{\theta}_{t+1},\ldots,\boldsymbol{\theta}_T)-\Phi_t^{\mathsf{T}}\boldsymbol{\theta}_t\Big)\frac{\phi_{tj}}{\bar{w}_{tj}}\Big]\Big|+\frac{7}{4}\gamma b\lambda_t \\
&\le \max_{j\in\{1,\ldots,J_t\}}\Big|(\mathbb{E}_n-E)\Big[\epsilon_t\frac{\phi_{tj}}{\bar{w}_{tj}}\Big]\Big|+\max_{j\in\{1,\ldots,J_t\}}\Big|(\mathbb{E}_n-E)\Big[f(\boldsymbol{\theta}_t,\ldots,\boldsymbol{\theta}_T)\frac{\phi_{tj}}{\bar{w}_{tj}}\Big]\Big|+\frac{7}{4}\gamma b\lambda_t,
\end{aligned}$$

where $f(\boldsymbol{\theta}_t,\ldots,\boldsymbol{\theta}_T)=Q_t^o(H_t,A_t)-\Phi_t^{\mathsf{T}}\boldsymbol{\theta}_t+\sum_{s=t+1}^T\big[-\max_{a_s}Q_s^o(H_s,a_s)+Q_s^o(H_s,A_s)-\Phi_s^{\mathsf{T}}\boldsymbol{\theta}_s+\max_{a_s}\Phi_s^{\mathsf{T}}\boldsymbol{\theta}_s\big]$.

Under Assumptions (A1) and (A3), it can be shown that $E(\epsilon_{ti}\phi_{tj}/\bar{w}_{tj})=0$ and $\sum_{i=1}^n E|(\epsilon_{ti}\phi_{tj}/\bar{w}_{tj})^l|\le l!n\sigma^2b^2(cu)^{l-2}/2$ for all integers $l\ge 2$. Therefore, for $\delta=(4\gamma+1)/12-7b\gamma/8$, if we apply Bernstein's inequality in Lemma S.1(b), we have

$$\mathbf{P}\Big(\Big|(\mathbb{E}_n-E)\Big[\epsilon_{ti}\frac{\phi_{tj}}{\bar{w}_{tj}}\Big]\Big|>\delta\lambda_t\Big)\le 2\exp\Big(-\frac{\delta^2\lambda_t^2n}{2[\sigma^2b^2+cu\delta\lambda_t]}\Big).$$

Similarly, we can show that $\|f(\boldsymbol{\theta}_t,\ldots,\boldsymbol{\theta}_T)\phi_{tj}/\bar{w}_{tj}-E(f(\boldsymbol{\theta}_t,\ldots,\boldsymbol{\theta}_T)\phi_{tj}/\bar{w}_{tj})\|_\infty\le 4[1+2(T-t)]\eta u$ and $E[f(\boldsymbol{\theta}_t,\ldots,\boldsymbol{\theta}_T)\phi_{tj}/\bar{w}_{tj}]^2\le 4[1+2(T-t)]^2\eta^2b^2$ using Assumptions (A3) and (A4). Applying Bernstein's inequality in Lemma S.1(a) yields

$$\begin{aligned}
&\mathbf{P}\Big(\Big|(\mathbb{E}_n-E)f(\boldsymbol{\theta}_t,\ldots,\boldsymbol{\theta}_T)\frac{\phi_{tj}}{\bar{w}_{tj}}\Big|>\delta\lambda_t\Big) \\
&\qquad\le 2\exp\left(-\frac{\delta^2\lambda_t^2n}{2\big[4[1+2(T-t)]^2\eta^2b^2+4[1+2(T-t)]\eta u\delta\lambda_t/3\big]}\right).
\end{aligned}$$

The result follows from the union bound argument and condition (C.3). □

Proof of Lemma 5.

We first note that

$$\begin{aligned}
&||\textstyle\sum_{a_t}\phi_{tj}(H_t,a_t)\phi_{tk}(H_t,a_t)/(\bar{w}_{tj}\bar{w}_{tk})-E[\sum_{a_t}\phi_{tj}(H_t,a_t)\phi_{tk}(H_t,a_t)/(\bar{w}_{tj}\bar{w}_{tk})]||_\infty \\
&\qquad\qquad\le 2|\mathcal{A}_t|u^2
\end{aligned}$$

and $E[\sum_{a_t}\phi_{tj}(H_t,a_t)\phi_{tk}(H_t,a_t)/(\bar{w}_{tj}\bar{w}_{tk})]^2\le|\mathcal{A}_t|^2b^2u^2$ for all $j,k\in\{1,\ldots,J_t\}$ by Assumptions (A3) and (A4). Next we apply Bernstein's inequality in Lemma S.1(a) with $\zeta_i=\pm[\sum_{a_{ti}}\phi_{tj}(H_{ti},a_{ti})\phi_{tk}(H_{ti},a_{ti})/(\bar{w}_{tj}\bar{w}_{tk})-E(\sum_{a_{ti}}\phi_{tj}(H_{ti},a_{ti})\phi_{tk}(H_{ti},a_{ti})/(\bar{w}_{tj}\bar{w}_{tk})]$ and $\kappa=(1-2\gamma)^2|\mathcal{A}_t|n/[144\max_{s\in\{t,\ldots,T\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}]$ such that $\zeta_i\le 2|\mathcal{A}_t|u^2$ and $\sum_{i=1}^n E\zeta_i^2\le n|\mathcal{A}_t|^2b^2u^2$. Using the union bound argument, we obtain

$$\begin{aligned}
&\mathbf{P}(\{\Omega_{t,3}(\boldsymbol{\theta}_t,\ldots,\boldsymbol{\theta}_T)\}^C) \\
&\le J_t(J_t+1)\times \\
&\exp\left(-\frac{(1-2\gamma)^4n}{2u^2\max_{s\in\{t,\ldots,T\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}[144^2b^2\max_{s\in\{t,\ldots,T\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}+96(1-2\gamma)^2]}\right) \\
&\le\exp(-\varphi)/3,
\end{aligned}$$

where the second inequality follows from the definition of Θ in (C.1). □

LEMMA S.1. ***(Bernstein's inequalities)*** *Let $\zeta_1,\dots,\zeta_n$ be independent and square integrable random variables such that $E(\zeta_i)=0$ for all $i=1,\dots,n$.*

(a) *Assume there exists some positive constants q and ν such that $\zeta_i\le q$ a.s. for all $i=1,\dots,n$ and $\sum_{i=1}^n E\zeta_i^2\le\nu$. Then for any $\kappa>0$,*

$$\mathbf{P}\Big(\sum_{i=1}^n\zeta_i>\kappa\Big)\le\exp\Big(-\frac{\kappa^2}{2(\nu+q\kappa/3)}\Big).$$

(b) *Assume there exists some positive constants q and ν such that $\sum_{i=1}^n E[(\zeta_i^l)_+]\le l!\nu q^{l-2}/2$ for all $l\ge 2$. Then for any $\kappa>0$,*

$$\mathbf{P}\Big(\sum_{i=1}^n\zeta_i>\kappa\Big)\le\exp\Big(-\frac{\kappa^2}{2(\nu+q\kappa)}\Big).$$

S.2. Additional Results and Proofs for Penalized Q-learning.

For any $\varphi>0$, $0\le\gamma<2/(21b-8)$ and tuning parameter λ_t, define

$$\Theta_t^{*Q}=\left\{\boldsymbol{\theta}_t\in\mathbb{R}^{J_t}:\|\Phi_t^\mathsf{T}(\boldsymbol{\theta}_t-\boldsymbol{\theta}_t^{*Q})\|_\infty\le\eta\quad\text{and}\quad E[\Phi_t^\mathsf{T}(\boldsymbol{\theta}_t-\boldsymbol{\theta}_t^{*Q})]^2\le\gamma^2\lambda_t^2\right\}$$

for $t=1,\dots,T$. Denote $J=\max_{\{t=1,\dots,T\}}J_t$, and

$$\Theta^Q=\Bigg\{(\boldsymbol{\theta}_1^\mathsf{T},\dots,\boldsymbol{\theta}_T^\mathsf{T})^\mathsf{T}\in\Pi_{t=1}^T\Theta_t^{*Q}:\max_{t\in\{1,\dots,T\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}$$

$$\le\frac{(1-2\gamma)^2}{144b}\left[\sqrt{\frac{1}{9b^2}+\frac{n}{2u^2[\log(3J(J+1))+\varphi]}}-\frac{1}{3b}\right]\Bigg\}. \tag{S.27}$$

THEOREM 5. *(Q-learning) Suppose there exists a constant $S\ge1$ such that $p_t(a_t|h_t)\ge S^{-1}$ for all (h_t,a_t) pairs for $t=1,\dots,T$. Assume assumptions (B1)–(B4) hold. For any given $0\le\gamma<2/(21b-8)$ and $\varphi>0$, suppose the tuning parameters $\lambda_t^Q,t=1,\dots,T$, satisfy*

$$\lambda_T^Q\ge\frac{8\max\{3c,4\eta\}u[\log(12J_T)+\varphi]}{[1-2\gamma(3b-2)]n}+\frac{12\max\{\sigma,2\eta\}b}{[1-2\gamma(3b-2)]}\sqrt{\frac{2[\log(12J_T)+\varphi]}{n}}, \tag{S.28}$$

$$\lambda_t^Q\ge\frac{16\max\{3c,4\big[1+2(T-t)\big]\eta\}u[\log(12J_t)+\varphi]}{[2-(21b-8)\gamma]n}$$

$$+\frac{24\max\{\sigma,2\big[1+2(T-t)\big]\eta\}b}{2-(21b-8)\gamma}\sqrt{\frac{2[\log(12J_t)+\varphi]}{n}}, \tag{S.29}$$

(S.30)

$$\text{and }(\lambda_t^Q)^2\ge\tilde c_{t,s}(\lambda_s^Q)^2\ \text{ with }\tilde c_{t,t}=1,\ \tilde c_{t,s}=\frac19(2\gamma+5)5S\tilde c_{t+1,s},$$

for $t=1,\dots,T$, $s=t+1$. Let Θ^Q be the set defined in (S.27) and assume Θ^Q is nonempty. Then for any $(\boldsymbol{\theta}_1^\mathsf{T},\dots,\boldsymbol{\theta}_T^\mathsf{T})^\mathsf{T}\in\Theta$, we have

$$\mathbf{P}\left(\bigcap_{t=1}^T\left\{E[\Phi_t^\mathsf{T}\hat{\boldsymbol{\theta}}_t^Q-Q_t^o]^2\le E[\Phi_t^\mathsf{T}\boldsymbol{\theta}_t-Q_t^o]^2+K_{t1}\max_{s\in\{t,t+1\}}\left(\tilde c_{t,s}\frac{|I_s(\boldsymbol{\theta}_s)|\lambda_s^2}{\tau_s}\right)\right\}\right)$$

$$\ge1-T\exp(-\varphi), \tag{S.31}$$

where $K_{t1}=[64(2\gamma+5)^2]/81+[32\gamma b(2\gamma+5)]/[3(1-2\gamma)]$.

Proof of Theorem 5.
For any $(\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_T^\top)^\top \in \Theta^Q$, denote the Q-learning pseudo-outcome

$$\tilde{Y}_t^Q(\boldsymbol{\theta}_{t+1}) = Y_t + \max_{a_{t+1}} \Phi_{t+1}^\top(H_{t+1}, a_{t+1})\boldsymbol{\theta}_{t+1} \tag{S.32}$$

when $t = T-1, \ldots, 1$, and $\tilde{Y}_t^Q(\boldsymbol{\theta}_{t+1}) \equiv Y_T$ when $t = T$ for the convenience of notation. For $t = 1, \ldots, T$, Let $|\mathcal{A}_t|$ be the cardinality of $\mathcal{A}_t$. Define the events

$$\Omega_{t,1}^Q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1}) = \left\{ \max_{j,k \in \{1,\ldots,J_t\}} \left|(E - \mathbb{E}_n)\left(\frac{\phi_{tj}\phi_{tk}}{\bar{w}_{tj}\bar{w}_{tk}}\right)\right| \le \frac{(1-2\gamma)^2}{144 \max_{s \in \{t,t+1\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}} \right\},$$

$$\Omega_{t,2}^Q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1}) = \left\{ \max_{j \in \{1,\ldots,J_t\}} \left|\mathbb{E}_n\left[\left(\tilde{Y}_t^Q(\boldsymbol{\theta}_{t+1}) - \Phi_t^\top \boldsymbol{\theta}_t\right)\frac{\phi_{tj}}{\bar{w}_{tj}}\right]\right| \le \frac{4\gamma+1}{6}\lambda_t^Q \right\},$$

$$\begin{aligned}\Omega_{t,3}^Q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1}) = \Bigg\{ \max_{j,k \in \{1,\ldots,J_t\}} &\left|(E - \mathbb{E}_n)\left(\sum_{a_t \in \mathcal{A}_t} \frac{\phi_{tj}(H_t, a_t)\phi_{tk}(H_t, a_t)}{\bar{w}_{tj}\bar{w}_{tk}}\right)\right| \\ &\le \frac{(1-2\gamma)^2|\mathcal{A}_t|}{144 \max_{s \in \{t,t+1\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}} \Bigg\}.\end{aligned}$$

We can show that

$$\begin{aligned}
&E[\Phi_t^\top \hat{\boldsymbol{\theta}}_t^Q - Q_t^o]^2 \\
&\quad = E[\Phi_t^\top \boldsymbol{\theta}_t - Q_t^o]^2 + E[\Phi_t^\top(\hat{\boldsymbol{\theta}}_t^Q - \boldsymbol{\theta}_t)]^2 + 2E[\Phi_t^\top(\boldsymbol{\theta}_t - \boldsymbol{\theta}_t^{*Q})][\Phi_t^\top(\hat{\boldsymbol{\theta}}_t^Q - \boldsymbol{\theta}_t)] \\
&\quad \le E[\Phi_t^\top \boldsymbol{\theta}_t - Q_t^o]^2 + E[\Phi_t^\top(\hat{\boldsymbol{\theta}}_t^Q - \boldsymbol{\theta}_t)]^2 \\
&\qquad\qquad + 2 \max_{j \in \{1,\ldots,J_t\}} \left|E\left[\Phi_t^\top(\boldsymbol{\theta}_t - \boldsymbol{\theta}_t^{*Q})\frac{\phi_{tj}}{\bar{w}_{tj}}\right]\right| \left(\sum_{j=1}^{J_t} \bar{w}_{tj}|\hat{\theta}_{tj}^Q - \theta_{tj}|\right) \\
&\quad \le E[\Phi_t^\top \boldsymbol{\theta}_t - Q_t^o]^2 + E[\Phi_t^\top(\hat{\boldsymbol{\theta}}_t^Q - \boldsymbol{\theta}_t)]^2 + 2\gamma b \lambda_t^Q \left(\sum_{j=1}^{J_t} \bar{w}_{tj}|\hat{\theta}_{tj}^Q - \theta_{tj}|\right).
\end{aligned}$$

By Lemma 6 presented below, on the event $\cap_{t=1}^T \left\{\Omega_{t,1}^Q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1}) \cap \Omega_{t,2}^Q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1}) \cap \Omega_{t+1,3}^Q(\boldsymbol{\theta}_{t+1}, \boldsymbol{\theta}_{t+2})\right\}$, we have

$$E[\Phi_t^\top \hat{\boldsymbol{\theta}}_t^Q - Q_t^o]^2 \le E[\Phi_t^\top \boldsymbol{\theta}_t - Q_t^o]^2 + K_{t1} \max_{s \in \{t,t+1\}} \left\{\tilde{c}_{t,s}\frac{|I_s(\boldsymbol{\theta}_s)|(\lambda_s^Q)^2}{\tau_s}\right\},$$

for $t = 1, \ldots, T$, where $\Omega_{T,1}^Q(\boldsymbol{\theta}_T, \boldsymbol{\theta}_{T+1})$, $\Omega_{T,2}^Q(\boldsymbol{\theta}_T, \boldsymbol{\theta}_{T+1})$, $\Omega_{T+1,3}^Q(\boldsymbol{\theta}_{T+1}, \boldsymbol{\theta}_{T+2})$, and $\Omega_{T,3}^Q(\boldsymbol{\theta}_T, \boldsymbol{\theta}_{T+1})$ are defined as the universe for the convenience of notation.

The conclusion of the theorem follows from the union probability bounds of the events $\Omega_{t,1}^Q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1})$, $\Omega_{t,2}^Q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1})$, and $\Omega_{t,3}^Q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1})$ for $t = 1, \ldots, T$, provided in Lemmas 7, 8, and 9. $\square$

LEMMA 6. *(Q-learning) Assume there exists a constant $S \ge 1$ such that $p_t(a_t|h_t) \ge S^{-1}$ for all (h_t, a_t) pairs. Suppose Assumption (B2) and condition (S.30) hold. Then, for any $(\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_T^\top)^\top \in \Theta^Q$, on the event $\cap_{t=1}^T \{\Omega_{t,1}^Q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1}) \cap \Omega_{t,2}^Q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1}) \cap$*

$\Omega^Q_{t+1,3}(\boldsymbol{\theta}_{t+1},\boldsymbol{\theta}_{t+2})\Big\}$, *we have*

$$\sum_{j=1}^{J_t}\bar{w}_{tj}|\hat{\theta}^Q_{tj}-\theta_{tj}|\leq\frac{16(2\gamma+5)}{3(1-2\gamma)\lambda^Q_t}\max_{s\in\{t,t+1\}}\left\{\tilde{c}_{t,s}\frac{|I_s(\boldsymbol{\theta}_s)|(\lambda^Q_s)^2}{\tau_s}\right\}\tag{S.33}$$

$$E[\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}^Q_t-\boldsymbol{\theta}_t)]^2\leq\frac{64(2\gamma+5)^2}{81}\max_{s\in\{t,t+1\}}\left\{\tilde{c}_{t,s}\frac{|I_s(\boldsymbol{\theta}_s)|(\lambda^Q_s)^2}{\tau_s}\right\},\tag{S.34}$$

for $t=1,\dots,T$, *where* $\Omega^Q_{T,1}(\boldsymbol{\theta}_T,\boldsymbol{\theta}_{T+1})$, $\Omega^Q_{T,2}(\boldsymbol{\theta}_T,\boldsymbol{\theta}_{T+1})$, $\Omega^Q_{T+1,3}(\boldsymbol{\theta}_{T+1},\boldsymbol{\theta}_{T+2})$, *and* $\Omega^Q_{T,3}(\boldsymbol{\theta}_T,\boldsymbol{\theta}_{T+1})$ *are defined as the universe for the convenience of notation.*

LEMMA 7. *(Q-learning) Suppose Assumptions (B3) and (B4) hold. Then for any* $\varphi>0$ *and* $(\boldsymbol{\theta}_1^{\mathsf{T}},\dots,\boldsymbol{\theta}_T^{\mathsf{T}})^{\mathsf{T}}\in\Theta^Q$, $\mathbf{P}(\{\Omega^Q_{t,1}(\boldsymbol{\theta}_t,\boldsymbol{\theta}_{t+1})\}^C)\leq\exp(-\varphi)/3$ *for* $t=1,\dots,T$.

LEMMA 8. *(Q-learning) Suppose Assumptions (B1), (B3), and (B4) hold. Then for any* $\varphi>0$, *if* λ^Q_t *satisfies conditions (S.28), (S.29) and (S.30), then for* $(\boldsymbol{\theta}_1^{\mathsf{T}},\dots,\boldsymbol{\theta}_T^{\mathsf{T}})^{\mathsf{T}}\in\Theta^Q$, $\mathbf{P}(\{\Omega^Q_{t,2}(\boldsymbol{\theta}_t,\boldsymbol{\theta}_{t+1})\}^C)\leq\exp(-\varphi)/3$ *for* $t=1,\dots,T$.

LEMMA 9. *(Q-learning) Suppose Assumptions (B3) and (B4) hold. Then for any* $\varphi>0$ *and* $(\boldsymbol{\theta}_1^{\mathsf{T}},\dots,\boldsymbol{\theta}_T^{\mathsf{T}})^{\mathsf{T}}\in\Theta^Q$, $\mathbf{P}(\{\Omega^Q_{t,3}(\boldsymbol{\theta}_t,\boldsymbol{\theta}_{t+1})\}^C)\leq\exp(-\varphi)/3$ *for* $t=1,\dots,T$.

Proof of Lemma 6.

At the last stage T, the proof is identical to Lemma 1, and we have

$$\sum_{j=1}^{J_T}\bar{w}_{Tj}|\hat{\theta}^Q_{Tj}-\theta_{Tj}|\leq\Big[\frac{16(2\gamma+5)}{3(1-2\gamma)}\Big]\frac{|I_T(\boldsymbol{\theta}_T)|\lambda_T}{\tau_T},$$

$$E[\Phi_T^{\mathsf{T}}(\hat{\boldsymbol{\theta}}^Q_T-\boldsymbol{\theta}_T)]^2\leq\Big[\frac{64(2\gamma+5)^2}{81}\Big]\frac{|I_T(\boldsymbol{\theta}_T)|\lambda_T^2}{\tau_T},$$

$$\text{and }\mathbb{E}_n[\Phi_T^{\mathsf{T}}(\hat{\boldsymbol{\theta}}^Q_T-\boldsymbol{\theta}_T)]^2\leq\Big[\frac{16(2\gamma+5)^2}{27}\Big]\frac{|I_T(\boldsymbol{\theta}_T)|\lambda_T^2}{\tau_T}.$$

on the event $\Omega^Q_{T,1}(\boldsymbol{\theta}_T)\cap\Omega^Q_{T,2}(\boldsymbol{\theta}_T)$.

Now we prove the results for $t<T$ using mathematical induction. Assume we have

$$\sum_{j=1}^{J_s}\bar{w}_{sj}|\hat{\theta}^Q_{sj}-\theta_{sj}|\leq\frac{16(2\gamma+5)}{3(1-2\gamma)\lambda^Q_s}\max_{s'\in\{s,s+1\}}\left\{\tilde{c}_{s,s'}\frac{|I_{s'}(\boldsymbol{\theta}_{s'})|(\lambda^Q_{s'})^2}{\tau_{s'}}\right\},\tag{S.35}$$

$$E[\Phi_s^{\mathsf{T}}(\hat{\boldsymbol{\theta}}^Q_s-\boldsymbol{\theta}_s)]^2\leq\frac{64(2\gamma+5)^2}{81}\max_{s'\in\{s,s+1\}}\left\{\tilde{c}_{s,s'}\frac{|I_{s'}(\boldsymbol{\theta}_{s'})|(\lambda^Q_{s'})^2}{\tau_{s'}}\right\},\tag{S.36}$$

$$\text{and }\mathbb{E}_n[\Phi_s^{\mathsf{T}}(\hat{\boldsymbol{\theta}}^Q_s-\boldsymbol{\theta}_s)]^2\leq\frac{16(2\gamma+5)^2}{27}\max_{s'\in\{s,s+1\}}\left\{\tilde{c}_{s,s'}\frac{|I_{s'}(\boldsymbol{\theta}_{s'})|(\lambda^Q_{s'})^2}{\tau_{s'}}\right\},\tag{S.37}$$

for $s=t+1,\dots,T$, on the event $\cap_{s=t+1}^{T}\Big\{\Omega^Q_{t,1}(\boldsymbol{\theta}_t,\boldsymbol{\theta}_{t+1})\cap\Omega^Q_{t,2}(\boldsymbol{\theta}_t,\boldsymbol{\theta}_{t+1})\cap\Omega^Q_{t+1,3}(\boldsymbol{\theta}_{t+1},\boldsymbol{\theta}_{t+2})\Big\}$, where $\tilde{c}_{s,s}=1$ and $\tilde{c}_{s,s+1}=5S(2\gamma+5)\tilde{c}_{s+1,s+1}/9$.

Using similar arguments in Lemma 1, for any $\boldsymbol{\theta}_t \in \mathbb{R}^{J_t}$ on the event $\Omega^Q_{t,2}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1})$, we have

$$
\begin{aligned}
0 &\le 2\mathbb{E}_n[(\tilde{Y}^Q_t(\boldsymbol{\theta}_{t+1}) - \Phi_t^{\mathsf{T}}\boldsymbol{\theta}_t)\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}^Q_t - \boldsymbol{\theta}_t)] + \lambda^Q_t \sum_{j=1}^{J_t} w_{tj}|\theta_{tj}| - \lambda^Q_t \sum_{j=1}^{J_t} w_{tj}|\hat{\theta}^Q_{tj}| \\
&\quad + 2\mathbb{E}_n[(\tilde{Y}^Q_t(\hat{\boldsymbol{\theta}}^Q_{t+1}) - \tilde{Y}^Q_t(\boldsymbol{\theta}_{t+1}))\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}^Q_t - \boldsymbol{\theta}_t)] - 2\mathbb{E}_n[\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}^Q_t - \boldsymbol{\theta}_t)]^2 \\
&\le \frac{4(\gamma+1)}{3}\lambda^Q_t\Big(\sum_{j\in I_t(\boldsymbol{\theta}_t)} \bar{w}_{tj}|\hat{\theta}^Q_{tj} - \theta_{tj}|\Big) - \frac{2(1-2\gamma)}{3}\lambda^Q_t\Big(\sum_{j\in I^c_t(\boldsymbol{\theta}_t)} \bar{w}_{tj}|\hat{\theta}^Q_{tj}|\Big) \\
&\quad - \mathbb{E}_n[\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}^Q_t - \boldsymbol{\theta}_t)]^2 + \mathbb{E}_n\big[(\tilde{Y}^Q_t(\hat{\boldsymbol{\theta}}^Q_{t+1}) - \tilde{Y}^Q_t(\boldsymbol{\theta}_{t+1}))\big]^2. \qquad \text{(S.38)}
\end{aligned}
$$

The difference to Lemma 1 (A-learning) is the last term in (S.38). We derive an upper bound for this term. Following similar arguments in Lemma 1, we can show that

$$
\begin{aligned}
&\mathbb{E}_n\big[\tilde{Y}^Q_t(\hat{\boldsymbol{\theta}}^Q_{t+1}) - \tilde{Y}^Q_t(\boldsymbol{\theta}_{t+1})\big]^2 \\
&\quad = \mathbb{E}_n\Big[\max_{a_{t+1}} \Phi_{t+1}^{\mathsf{T}}(H_{t+1}, a_{t+1})\hat{\boldsymbol{\theta}}^Q_{t+1} - \max_{a_{t+1}} \Phi_{t+1}^{\mathsf{T}}(H_{t+1}, a_{t+1})\boldsymbol{\theta}_{t+1}\Big]^2 \\
&\quad \le \mathbb{E}_n\Big[\max_{a_{t+1}} \big|\Phi_{t+1}^{\mathsf{T}}(H_{t+1}, a_{t+1})(\hat{\boldsymbol{\theta}}^Q_{t+1} - \boldsymbol{\theta}_{t+1})\big|^2\Big] \\
&\quad \le \frac{16(2\gamma+5)^2|\mathcal{A}_s|}{81}\Big[\max_{s'\in\{t+1,t+2\}}\Big\{\tilde{c}_{t+1,s'}\frac{|I_{s'}(\boldsymbol{\theta}_{s'})|(\lambda^Q_{s'})^2}{\tau_{s'}}\Big\}\Big] \\
&\qquad\qquad + \frac{64(2\gamma+5)^2 S}{81}\Big[\max_{s'\in\{t+1,t+2\}}\Big\{\tilde{c}_{t+1,s'}\frac{|I_{s'}(\boldsymbol{\theta}_{s'})|(\lambda^Q_{s'})^2}{\tau_{s'}}\Big\}\Big] \\
&\quad \le \frac{80(2\gamma+5)^2 S}{81}\Big[\max_{s'\in\{t+1,t+2\}}\Big\{\tilde{c}_{t+1,s'}\frac{|I_{s'}(\boldsymbol{\theta}_{s'})|(\lambda^Q_{s'})^2}{\tau_{s'}}\Big\}\Big] \\
&\quad \le C_t \max_{s\in\{t+1,t+2\}}\Big\{\tilde{c}_{t+1,s}\frac{|I_s(\boldsymbol{\theta}_s)|(\lambda^Q_s)^2}{\tau_s}\Big\},
\end{aligned}
$$

where $C_t = 80(2\gamma+5)^2 S/81$. Using similar proof techniques as in Lemma 1, we have

$$
\begin{aligned}
&\sum_{j=1}^{J_t} \bar{w}_{tj}|\hat{\theta}^Q_{tj} - \theta_{tj}| \\
&\qquad \le \max\Big\{\frac{16(2\gamma+5)|I_t(\boldsymbol{\theta}_t)|\lambda^Q_t}{3(1-2\gamma)\tau_t}, \frac{3C_t}{(1-2\gamma)\lambda^Q_t}\max_{s\in\{t+1,t+2\}}\Big\{\tilde{c}_{t+1,s}\frac{|I_s(\boldsymbol{\theta}_s)|(\lambda^Q_s)^2}{\tau_s}\Big\}\Big\} \\
&\qquad = \frac{16(2\gamma+5)}{3(1-2\gamma)\lambda^Q_t}\max_{s\in\{t,t+1\}}\Big\{\tilde{c}_{t,s}\frac{|I_s(\boldsymbol{\theta}_s)|(\lambda^Q_s)^2}{\tau_s}\Big\},
\end{aligned}
$$

$$
\mathbb{E}_n[\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}^Q_t - \boldsymbol{\theta}_t)]^2 \le \frac{16(2\gamma+5)^2}{27}\max_{s\in\{t,t+1\}}\Big\{\tilde{c}_{t,s}\frac{|I_s(\boldsymbol{\theta}_s)|(\lambda^Q_s)^2}{\tau_s}\Big\},
$$

and

$$
E[\Phi_t^{\mathsf{T}}(\hat{\boldsymbol{\theta}}^Q_t - \boldsymbol{\theta}_t)]^2 \le \frac{64(2\gamma+5)^2}{81}\max_{s\in\{t,t+1\}}\Big\{\tilde{c}_{t,s}\frac{|I_s(\boldsymbol{\theta}_s)|(\lambda^Q_s)^2}{\tau_s}\Big\}.
$$

This completes the proof. $\square$

Proof of Lemma 7.

Using the similar arguments used in Lemma 3, we obtain

$$\begin{aligned}
&\mathbf{P}(\{\Omega^Q_{t,3}(\boldsymbol{\theta}_t,\boldsymbol{\theta}_{t+1})\}^C)\\
&\leq J_t(J_t+1)\times\\
&\exp\left(-\frac{(1-2\gamma)^4 n}{2u^2\max_{s\in\{t,t+1\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}[144^2 b^2 \max_{s\in\{t,t+1\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}+96(1-2\gamma)^2]}\right)\\
&\leq \exp(-\varphi)/3,
\end{aligned}$$

where the second inequality follows from the definition of Θ^Q in (S.27). $\square$

Proof of Lemma 8.

At the last stage T, the proofs are identical to Lemma 4. Now we prove the results for $t<T$. For any $(\boldsymbol{\theta}_1^\mathsf{T},\ldots,\boldsymbol{\theta}_T^\mathsf{T})^\mathsf{T}\in\Theta^Q$, note that

$$\begin{aligned}
E\big[\tilde{Y}^Q_t(\boldsymbol{\theta}_{t+1})-\tilde{Y}^Q_t(\boldsymbol{\theta}^{*Q}_{t+1})\big]^2 &= E\Big[\max_{a_{t+1}}\Phi^\mathsf{T}_{t+1}(H_{t+1},a_{t+1})\boldsymbol{\theta}_{t+1}-\max_{a_{t+1}}\Phi^\mathsf{T}_{t+1}(H_{t+1},a_{t+1})\boldsymbol{\theta}^{*Q}_{t+1}\Big]^2\\
&\leq E\Big[\max_{a_{t+1}}\big|\Phi^\mathsf{T}_{t+1}(H_{t+1},a_{t+1})(\boldsymbol{\theta}_{t+1}-\boldsymbol{\theta}^{*Q}_{t+1})\big|^2\Big]\\
&\leq E\Big[\textstyle\sum_{a_{t+1}}\big[\Phi^\mathsf{T}_{t+1}(H_{t+1},a_{t+1})(\boldsymbol{\theta}_{t+1}-\boldsymbol{\theta}^{*Q}_{t+1})\big]^2\Big]\\
&\leq E\Big[\textstyle\sum_{a_{t+1}}p(a_{t+1}|H_{t+1})S\big[\Phi^\mathsf{T}_{t+1}(H_{t+1},a_{t+1})(\boldsymbol{\theta}_{t+1}-\boldsymbol{\theta}^{*Q}_{t+1})\big]^2\Big]\\
&\leq SE\Big[\Phi^\mathsf{T}_{t+1}(H_{t+1},A_{t+1})(\boldsymbol{\theta}_{t+1}-\boldsymbol{\theta}^{*Q}_{t+1})\Big]^2\\
&\leq S\gamma^2(\lambda^Q_{t+1})^2\\
&\leq \frac{9}{16}\gamma^2(\lambda^Q_t)^2
\end{aligned}$$

where the last inequality holds under condition (S.30) and the fact that $\tilde{c}_{t,s}\geq 16S/9$. The rest of the proofs are similar to Lemma 4. This completes the proof. $\square$

Proof of Lemma 9.

Using the similar arguments used in Lemma 5, we obtain

$$\begin{aligned}
&\mathbf{P}(\{\Omega^Q_{t,3}(\boldsymbol{\theta}_t,\boldsymbol{\theta}_{t+1})\}^C)\\
&\leq J_t(J_t+1)\times\\
&\exp\left(-\frac{(1-2\gamma)^4 n}{2u^2\max_{s\in\{t,t+1\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}[144^2 b^2 \max_{s\in\{t,t+1\}}\{|I_s(\boldsymbol{\theta}_s)|/\tau_s\}+96(1-2\gamma)^2]}\right)\\
&\leq \exp(-\varphi)/3,
\end{aligned}$$

where the second inequality follows from the definition of Θ^Q in (S.27). $\square$

S.3. Additional Simulation Results. Additional simulation results for Scenarios 1-5 with $p=200$ based on $1,000$ replications are summarized in Table 3 below. The mean of values with its standard deviation in parentheses is also reported. The table also shows the median number of inactive variables incorrectly selected in the model, denoted by FP, and the median number of active variables left out of the model, denoted by FN, which are both recorded along with the mean absolute deviation in parentheses. The median of contrast function root-mean-square error (cRMSE) is also calculated for the Alearn- and Qlearn-PROaL as well as for the PAL method. Table 3 shows that both Alearn- and Qlearn-PROaL outperform other methods in all scenarios except Scenario 5 due to not penalizing a few key variables. Overall patterns are similar to Table 1 in terms of which methods are superior with higher value, better selection performance, and lower cRMSE.

We also consider an extra scenario with three decision points $T = 3$. In this scenario, the treatments at all stages, A_1, A_2, and A_3, are randomly generated from $Bernoulli(0.5)$. Each of the p-dimensional baseline covariates O_1 is independently generated from $N(45, 15^2)$. The stage-2 covariate is $O_2 \sim N(1.5O_{11}, 10^2)$, where O_{11} is the first component of O_1. The stage-3 covariate is generated from $O_3 \sim N(0.5O_2, 10^2)$. The outcomes are generated as follows: $Y_t = 0$, $t = 1, 2$, and $Y_3 \sim N(20 - |0.6O_{11} - 40|\{I(A_1 > 0) - I(O_{11} > 30)\}^2 - |0.8O_2 - 60|\{I(A_2 > 0) - I(O_2 > 40)\}^2 - |1.4O_3 - 40|\{I(A_3 > 0) - I(O_3 > 40)\}^2, 1^2)$. The results with three decision points $T = 3$ and $p = 60$ based on $1,000$ replications are provided in Table 4. The Alearn-PROaL method has the highest value estimate among all methods. Furthermore, the value estimation by Alearn-PROaL improves as the sample size grows, whereas the estimated value of all other methods remains very similar or even decreases. In terms of selection performance, there is not much difference between Alearn- and Qlearn-PROaL. However, Qlearn-PROaL has a smaller FP at the terminal stage (stage 3) when $n = 50$ and a larger FN at the initial stage (stage 1) when $n = 150$ compared to Alearn-PROaL.

TABLE 3
Simulation results for $p = 200$. The mean value is reported with the standard deviation in parentheses. The median FP, FN, and cRMSE are recorded with the mean absolute deviation in parentheses. The best results are highlighted in boldface.

			Stage 2			Stage 1		
n	Method	Value	FP	FN	cRMSE	FP	FN	cRMSE
Scenario 1								
50	Optimal	2.29						
	Alearn-PROaL	1.88 (0.34)	5 (7.41)	2 (1.48)	2.25 (0.50)	1 (1.48)	3 (0)	1.45 (0.32)
	Qlearn-PROaL	**2.06** (0.25)	3.5 (3.71)	1 (1.48)	**1.81** (0.60)	3 (4.45)	2 (0)	**1.14** (0.35)
	PAL	1.46 (0.45)	1 (1.48)	3 (0)	1.94 (0.65)	2 (1.48)	3 (0)	2.18 (0.49)
	BOWL-linear	0.96 (0.30)	198 (0)	0 (0)	-	197 (0)	0 (0)	-
	BOWL-radial	1.53 (0.84)	-	-	-	-	-	-
150	Optimal	2.29						
	Alearn-PROaL	2.25 (0.04)	2 (2.97)	0 (0)	0.87 (0.30)	0 (0)	2 (0)	0.76 (0.20)
	Qlearn-PROaL	**2.27** (0.02)	0 (0)	0 (0)	0.75 (0.22)	0 (0)	1 (1.48)	**0.56** (0.09)
	PAL	2.16 (0.11)	0 (0)	1 (0)	**0.66** (0.33)	1 (1.48)	2 (0)	1.05 (0.44)
	BOWL-linear	0.91 (0.14)	198 (0)	0 (0)	-	197 (0)	0 (0)	-
	BOWL-radial	1.92 (0.38)	-	-	-	-	-	-
Scenario 2								
50	Optimal	2.48						
	Alearn-PROaL	1.80 (0.40)	5 (7.41)	2 (1.48)	2.46 (0.42)	1 (1.48)	5 (0)	2.03 (0.25)
	Qlearn-PROaL	**1.97** (0.33)	4 (4.45)	2 (1.48)	**2.18** (0.51)	3 (4.45)	4 (1.48)	**1.87** (0.26)
	PAL	1.34 (0.44)	2 (1.48)	3 (1.48)	2.46 (0.84)	2 (1.48)	5 (0)	2.71 (0.53)
	BOWL-linear	0.96 (0.30)	198 (0)	0 (0)	-	195 (0)	0 (0)	-
	BOWL-radial	1.46 (0.89)	-	-	-	-	-	-
150	Optimal	2.48						
	Alearn-PROaL	**2.31** (0.09)	3 (2.97)	0 (0)	1.17 (0.42)	2 (2.97)	2 (1.48)	**1.21** (0.38)
	Qlearn-PROaL	2.27 (0.03)	1 (1.48)	0 (0)	**0.81** (0.28)	1 (1.48)	2 (1.48)	1.51 (0.05)
	PAL	2.22 (0.16)	0 (0)	1 (1.48)	1.13 (0.53)	1 (1.48)	3 (1.48)	1.34 (0.51)
	BOWL-linear	0.93 (0.14)	198 (0)	0 (0)	-	195 (0)	0 (0)	-
	BOWL-radial	1.82 (0.57)	-	-	-	-	-	-
Scenario 3								
50	Optimal	2.29						
	Alearn-PROaL	1.91 (0.33)	4 (5.93)	2 (1.48)	2.30 (0.59)	1 (1.48)	3 (0)	1.43 (0.30)
	Qlearn-PROaL	**2.10** (0.13)	3.5 (5.19)	1 (1.48)	**1.69** (0.60)	3 (4.45)	2 (1.48)	**1.06** (0.37)
	PAL	1.46 (0.40)	1 (1.48)	3 (1.48)	2.35 (0.69)	2 (1.48)	3 (0)	2.11 (0.58)
	BOWL linear	1.20 (0.36)	198 (0)	0 (0)	-	197 (0)	0 (0)	-
	BOWL radial	1.93 (0.36)	-	-	-	-	-	-
150	Optimal	2.29						
	Alearn-PROaL	2.20 (0.10)	1 (1.48)	0 (0)	1.39 (0.49)	0 (0)	2 (0)	0.84 (0.31)
	Qlearn-PROaL	**2.27** (0.02)	0 (0)	0 (0)	**0.70** (0.20)	0 (0)	1 (1.48)	**0.56** (0.09)
	PAL	1.73 (0.33)	1 (1.48)	3 (1.48)	2.18 (0.48)	2 (1.48)	2 (0)	1.14 (0.51)
	BOWL-linear	1.12 (0.13)	198 (0)	0 (0)	-	197 (0)	0 (0)	-
	BOWL-radial	2.00 (0.00)	-	-	-	-	-	-
Scenario 4								
50	Optimal	2.48						
	Alearn-PROaL	1.83 (0.38)	3 (4.45)	2 (1.48)	2.55 (0.54)	1 (1.48)	5 (0)	2.05 (0.29)
	Qlearn-PROaL	**2.04** (0.14)	4 (5.93)	2 (1.48)	**2.12** (0.44)	3 (4.45)	4 (1.48)	**1.86** (0.28)
	PAL	1.34 (0.41)	1 (1.48)	3 (1.48)	2.65 (0.60)	2 (1.48)	5 (0)	2.70 (0.59)
	BOWL linear	1.18 (0.30)	198 (0)	0 (0)	-	195 (0)	0 (0)	-
	BOWL radial	1.87 (0.50)	-	-	-	-	-	-
150	Optimal	2.48						
	Alearn-PROaL	2.19 (0.17)	1 (1.48)	1 (1.48)	1.73 (0.60)	2 (2.97)	2 (1.48)	**1.36** (0.42)
	Qlearn-PROaL	**2.26** (0.03)	2 (1.48)	0 (0)	**0.84** (0.26)	1 (1.48)	2 (1.48)	1.50 (0.05)
	PAL	1.71 (0.37)	1 (1.48)	3 (1.48)	2.34 (0.49)	2 (1.48)	3 (1.48)	1.53 (0.56)
	BOWL-linear	1.12 (0.13)	198 (0)	0 (0)	-	195 (0)	0 (0)	-
	BOWL-radial	2.00 (0.09)	-	-	-	-	-	-
Scenario 5								
50	Optimal	7.19						
	Alearn-PROaL	6.27 (1.28)	0 (0)	3 (0)	18.47 (1.93)	0 (0)	1 (0)	3.67 (3.54)
	Qlearn-PROaL	6.26 (1.32)	0 (0)	3 (0)	**16.97** (0.28)	0 (0)	1 (0)	**0.52** (0.06)
	PAL	2.99 (1.73)	2 (1.48)	4 (0)	22.38 (4.69)	4 (1.48)	1 (0)	16.89 (8.56)
	BOWL-linear	4.92 (1.05)	198 (0)	0 (0)	-	199 (0)	0 (0)	-
	BOWL-radial	**6.76** (0.00)	-	-	-	-	-	-
150	Optimal	7.19						
	Alearn-PROaL	6.76 (0.17)	0 (0)	3 (0)	18.05 (1.06)	0 (0)	1 (0)	1.93 (1.57)
	Qlearn-PROaL	**6.78** (0.15)	0 (0)	3 (0)	**16.78** (0.08)	0 (0)	0 (0)	**0.33** (0.15)
	PAL	4.64 (1.66)	1 (1.48)	4 (0)	19.69 (2.00)	7 (2.97)	1 (0)	15.73 (4.25)
	BOWL-linear	3.57 (0.65)	198 (0)	0 (0)	-	199 (0)	0 (0)	-
	BOWL-radial	6.76 (0.00)	-	-	-	-	-	-

TABLE 4

Simulation results for $T = 3$ extra scenario with $p = 60$. The mean value is reported with the standard deviation in parentheses. The median FP and FN are recorded with the mean absolute deviation in parentheses. The best results are highlighted in boldface.

			Stage 3		Stage 2		Stage 1	
n	Method	Value	FP	FN	FP	FN	FP	FN
50	Optimal	20	0	0	0	0	0	0
	Alearn-PROaL	**15.03** (1.48)	3 (1.48)	1 (0)	1 (0)	1 (0)	0 (0)	1 (0)
	Qlearn-PROaL	14.36 (2.02)	1 (0)	1 (0)	1 (0)	1 (0)	0 (0)	1 (0)
	BOWL-linear	13.09 (2.36)	63 (0)	0 (0)	61 (0)	0 (0)	59 (0)	0 (0)
	BOWL-radial	13.08 (2.39)	-	-	-	-	-	-
150	Optimal	20	0	0	0	0	0	0
	Alearn-PROaL	**16.65** (1.50)	2 (0)	1 (0)	1 (0)	1 (0)	0 (0)	0 (0)
	Qlearn-PROaL	14.74 (1.79)	2 (0)	1 (0)	1 (0)	1 (0)	0 (0)	1 (0)
	BOWL-linear	13.41 (0.82)	63 (0)	0 (0)	61 (0)	0 (0)	59 (0)	0 (0)
	BOWL-radial	11.85 (2.01)	-	-	-	-	-	-