SUPPLEMENTARY MATERIAL TO ROBUST Q-LEARNING

Let $Z \sim P$ for some probability measure P and suppose $f(\cdot)$ is any real-valued, $P-$measurable function; then, we define the $L^2(P)$ norm of $f(\cdot)$ as $\|f(Z)\|_{P,2} = \{\int f(\omega)^2 dP(\omega)\}^{1/2}$. In addition, let $\|\boldsymbol{x}\|_q$ denote the usual $q-$ norm of a vector $\boldsymbol{x}$ for $q = 1, 2, \infty$. The following general lemmas will be helpful in our proofs.

Lemma 1. *Let A_n and B_n be sequences of random vectors, $n \geq 1$. Let $\epsilon > 0$ be arbitrary and, for any vector norm, suppose that $\lim_{n\to\infty} P(\|A_n\| > \epsilon | B_n) = 0$. Then, $\lim_{n\to\infty} P(\|A_n\| > \epsilon) = 0$. By Chebyshev's inequality, a sufficient condition for proving that $\lim_{n\to\infty} P(\|A_n\| > \epsilon) = 0$ is that $\lim_{n\to\infty} E(\|A_n\|^q | B_n) = 0$ from some $q \geq 1$.*

The above lemma essentially repeats Lemma 6.1 in Chernozhukov et al. (2018) and will not be proved here. The following lemma is a direct consequence of a well-known result and also has an easy proof; see, for example, Stewart (1969).

Lemma 2. *Let $\boldsymbol{M}_n$ and $\hat{\boldsymbol{M}}_n$ be two sequences of square matrices and let $\|\cdot\|$ be any proper matrix norm. Suppose there exists $n_0 < \infty$ such that (i) $\boldsymbol{M}_n^{-1}$ and $\boldsymbol{M}_n$ exist for $n \geq n_0$, with $0 < C_1 \leq \|\boldsymbol{M}_n^{-1}\| \leq C_2 < \infty$; and, (ii) $\|\hat{\boldsymbol{M}}_n - \boldsymbol{M}_n\| \leq (2\|\boldsymbol{M}_n^{-1}\|)^{-1}$. Then,*

$$\|\hat{\boldsymbol{M}}_n^{-1} - \boldsymbol{M}_n^{-1}\| \leq 2C_2^2 \|\hat{\boldsymbol{M}}_n - \boldsymbol{M}_n\|.$$

We will also have need of the following lemma.

Lemma 3. *Let $\boldsymbol{B}_1, \ldots, \boldsymbol{B}_N$ be independent, identically distributed vectors from P_0, where $\boldsymbol{B}_i \in \mathcal{B} \subset \mathbb{R}^d$. Let $\boldsymbol{I}_n$ be a randomly chosen subset of the integers $1, \ldots, N$ of length $n = O(N)$ and let its complement $\boldsymbol{I}_n^c$ have $N-n = O(N)$ elements. Let $\boldsymbol{F}_{I_n}$ and $\boldsymbol{F}_{I_n^c}$ be the corresponding disjoint subsets of $\boldsymbol{B}_1, \ldots, \boldsymbol{B}_N$. Let $\gamma_j : \mathcal{B} \to \mathbb{R}, j = 1, 2$ and let $\hat{\gamma}_j(\cdot; \boldsymbol{F}_{I_n^c})$ be an estimator of $\gamma_j(\cdot)$ derived from the data $\boldsymbol{F}_{I_n^c}$. Finally, define*

$$\boldsymbol{G}_{n,N} = \frac{1}{n} \sum_{i \in \boldsymbol{I}_n} \left\{\hat{\gamma}_1(\boldsymbol{B}_i; \boldsymbol{F}_{I_n^c}) - \gamma_1(\boldsymbol{B}_i)\right\} \left\{\hat{\gamma}_2(\boldsymbol{B}_i; \boldsymbol{F}_{I_n^c}) - \gamma_2(\boldsymbol{B}_i)\right\} \boldsymbol{h}(\boldsymbol{B}_i) \tag{1}$$

where $\boldsymbol{h}(\boldsymbol{B}_i)$ *is any finite dimensional vector- or matrix-valued function of* $\boldsymbol{B}_i$ *such that* $P(\|\boldsymbol{h}(\boldsymbol{B}_i)\|_\infty \leq C) = 1$ *for some* $C < \infty$*. Then,*

$$\|\boldsymbol{G}_{n,N}\|_\infty \leq C \left\|\hat{\gamma}_1(\boldsymbol{B}; \boldsymbol{F}_{I_n^c}) - \gamma_1(\boldsymbol{B})\right\|_{\mathbb{P}_n,2} \left\|\hat{\gamma}_2(\boldsymbol{B}; \boldsymbol{F}_{I_n^c}) - \gamma_2(\boldsymbol{B})\right\|_{\mathbb{P}_n,2}, \tag{2}$$

where $\mathbb{P}_n$ *is the empirical measure on* $\boldsymbol{F}_{I_n}$*. Moreover, for* $j = 1, 2$ *define*

$$\left\|\hat{\gamma}_j(\boldsymbol{B}; \boldsymbol{F}_{I_n^c}) - \gamma_j(\boldsymbol{B})\right\|_{P_0,2}^2 = \mathbb{E}\left\{\left\|\hat{\gamma}_j(\boldsymbol{B}; \boldsymbol{F}_{I_n^c}) - \gamma_j(\boldsymbol{B})\right\|_{\mathbb{P}_n,2}^2 \middle| \boldsymbol{F}_{I_n^c}\right\} \tag{3}$$

and suppose (3) *is* $o_p(N^{-a_j})$*, where* $a_j \geq 0$*. Then,* $\|\boldsymbol{G}_{n,N}\|_\infty = o_p(N^{-(a_1+a_2)/2})$*.*

Proof. Let $r_{ij}(\boldsymbol{F}_{I_n^c}) = \hat{\gamma}_j(\boldsymbol{B}_i; \boldsymbol{F}_{I_n^c}) - \gamma_j(\boldsymbol{B}_i)$ for $i = 1, \ldots, n$ and $j = 1, 2$. Under the assumption that $P(\|\boldsymbol{h}(\boldsymbol{B}_i)\|_\infty \leq C) = 1$, the triangle and Cauchy-Schwarz equalities imply

$$\begin{aligned}
\|\boldsymbol{G}_{n,N}\|_\infty &\leq \frac{1}{n}\sum_{i\in\boldsymbol{I}_n} \left|r_{i1}(\boldsymbol{F}_{I_n^c}) r_{i2}(\boldsymbol{F}_{I_n^c})\right| \|\boldsymbol{h}(\boldsymbol{B}_i)\|_\infty \\
&\leq C\left[\frac{1}{n}\sum_{i\in\boldsymbol{I}_n}\left\{r_{i1}(\boldsymbol{F}_{I_n^c})\right\}^2\right]^{1/2}\left[\frac{1}{n}\sum_{i\in\boldsymbol{I}_n}\left\{r_{i2}(\boldsymbol{F}_{I_n^c})\right\}^2\right]^{1/2},
\end{aligned}$$

the representation (2) now following immediately from the definition of the $L^2(Q)$ norm given earlier using $Q = \mathbb{P}_n$. To establish that $\|\boldsymbol{G}_{n,N}\|_\infty = o_p(N^{-(a_1+a_2)})$, we first use Markov's inequality: for any $\epsilon > 0$,

$$P\left(\|\boldsymbol{G}_{n,N}\|_\infty > \epsilon | \boldsymbol{F}_{I_n^c}\right) \leq \epsilon^{-1}\mathbb{E}\left(\|\boldsymbol{G}_{n,N}\|_\infty | \boldsymbol{F}_{I_n^c}\right).$$

Using (2) and the Cauchy-Schwarz inequality again, it follows that

$$\begin{aligned}
P\left(\|\boldsymbol{G}_{n,N}\|_\infty > \epsilon | \boldsymbol{F}_{I_n^c}\right) &\leq \frac{C}{\epsilon}\mathbb{E}\left\{\prod_{j=1}^2 \left\|\hat{\gamma}_j(\boldsymbol{B}; \boldsymbol{F}_{I_n^c}) - \gamma_j(\boldsymbol{B})\right\|_{\mathbb{P}_n,2} \middle| \boldsymbol{F}_{I_n^c}\right\} \\
&\leq \frac{C}{\epsilon}\prod_{j=1}^2 \mathbb{E}\left\{\left\|\hat{\gamma}_j(\boldsymbol{B}; \boldsymbol{F}_{I_n^c}) - \gamma_j(\boldsymbol{B})\right\|_{\mathbb{P}_n,2}^2 \middle| \boldsymbol{F}_{I_n^c}\right\}^{1/2} \\
&= \frac{C}{\epsilon}\prod_{j=1}^2 \left\|\hat{\gamma}_j(\boldsymbol{B}; \boldsymbol{F}_{I_n^c}) - \gamma_j(\boldsymbol{B})\right\|_{P_0,2}
\end{aligned}$$

the last result following directly from (3). Under the stated assumptions, the right-hand side is now seen to be $o_p(N^{-(a_1+a_2)/2})$, as desired. □

The statement and proof of Lemma 3 employs a simple form of sample splitting in which the unknown function $\gamma_j(\cdot)$ is estimated by $\hat{\gamma}_j(\cdot; \boldsymbol{F}_{I_n^c})$ from a sample $\boldsymbol{F}_{I_n^c}$ that is independent of the $\boldsymbol{B}_i$s (i.e., $\boldsymbol{F}_{I_n}$) appearing in the calculation of (1). Importantly, Lemma 3 does not preclude the possibility that $\gamma_1(\cdot) = \gamma_2(\cdot)$ and $\hat{\gamma}_1(\cdot) = \hat{\gamma}_2(\cdot)$; in this case, (1) reduces to

$$\boldsymbol{L}_{n,m} = \frac{1}{n} \sum_{i \in \boldsymbol{I}_n} \left\{ \hat{\gamma}(\boldsymbol{B}_i; \boldsymbol{F}_{I_n^c}) - \gamma(\boldsymbol{B}_i) \right\}^2 \boldsymbol{h}(\boldsymbol{B}_i) \tag{4}$$

and $\|\boldsymbol{G}_{n,N}\|_\infty = o_p(N^{-a})$ for some $a \geq 0$ provided that

$$\left\| \hat{\gamma}(\boldsymbol{B}; \boldsymbol{F}_{I_n^c}) - \gamma(\boldsymbol{B}) \right\|_{P_0,2}^2 = o_p(N^{-a}). \tag{5}$$

A related lemma now follows.

Lemma 4. *Let $(R_1, \boldsymbol{B}_1), \ldots, (R_n, \boldsymbol{B}_N)$ be independent, identically distributed vectors from P_0, where $\boldsymbol{B}_i \in \mathcal{B} \subset \mathbb{R}^d$ and $E_i \in \mathbb{R}$. Suppose $\mathbb{E}(R_i|\boldsymbol{B}_i) = 0$ and $var(R_i|\boldsymbol{B}_i) = \theta_i^2 \in (0, C_1]$ for $i = 1 \ldots N$ and a constant $C_1 < \infty$. Let $\boldsymbol{I}_n$ be a randomly chosen subset of the integers $1, \ldots, N$ of length $n = O(N)$ and let its complement $\boldsymbol{I}_n^c$ have $N - n$ elements. Let $\boldsymbol{F}_{I_n}$ and $\boldsymbol{F}_{I_n^c}$ be the corresponding disjoint subsets of $(R_1, \boldsymbol{B}_1), \ldots, (R_N, \boldsymbol{B}_N)$. Let $\gamma : \mathcal{B} \to \mathbb{R}$ and let $\hat{\gamma}(\cdot; \boldsymbol{F}_{I_n^c})$ be an estimator of $\gamma(\cdot)$ derived from the data $\boldsymbol{F}_{I_n^c}$. Finally, define*

$$\boldsymbol{L}_{n,N} = \frac{1}{n} \sum_{i \in \boldsymbol{I}_n} R_i \left\{ \hat{\gamma}(\boldsymbol{B}_i; \boldsymbol{F}_{I_n^c}) - \gamma(\boldsymbol{B}_i) \right\} \boldsymbol{h}(\boldsymbol{B}_i) \tag{6}$$

where $\boldsymbol{h}(\boldsymbol{B}_i)$ is any finite dimensional vector-valued function of $\boldsymbol{B}_i$ such that $P(\|\boldsymbol{h}(\boldsymbol{B}_i)\|_2 \leq C_2) = 1$ for $C_2 < \infty$. Suppose that

$$\left\| \hat{\gamma}(\boldsymbol{B}; \boldsymbol{F}_{I_n^c}) - \gamma(\boldsymbol{B}) \right\|_{P_0,2}^2 = \mathbb{E}\left\{ \left\| \hat{\gamma}(\boldsymbol{B}; \boldsymbol{F}_{I_n^c}) - \gamma(\boldsymbol{B}) \right\|_{\mathbb{P}_n,2}^2 \middle| \boldsymbol{F}_{I_n^c} \right\} \tag{7}$$

is $o_p(N^{-a})$ where $a \geq 0$. Then, $\|\boldsymbol{L}_{n,N}\|_\infty = o_p(N^{-(1+a)/2})$.

Proof. The proof relies on a variant of Chebyshev's inequality. Let

$$\boldsymbol{L}_{n,N,j} = \frac{1}{n} \sum_{i \in \boldsymbol{I}_n} R_i \left\{ \hat{\gamma}(\boldsymbol{B}_i; \boldsymbol{F}_{I_n^c}) - \gamma(\boldsymbol{B}_i) \right\} \boldsymbol{h}_j(\boldsymbol{B}_i)$$

be the jth element of $\boldsymbol{L}_{n,N}$. Let $\mathcal{B}_n = \{\boldsymbol{I}_n, (\boldsymbol{B}_k, k \in \boldsymbol{I}_n)\}$. Then, it is easy to show that

$$\mathbb{E}\left(\boldsymbol{L}_{n,N,j} \middle| \boldsymbol{F}_{I_n^c} \right) \;=\; \mathbb{E}\left\{ \mathbb{E}\left(\boldsymbol{L}_{n,N,j} \middle| \boldsymbol{F}_{I_n^c}, \mathcal{B}_n \right) \middle| \boldsymbol{F}_{I_n^c} \right\} \;=\; 0;$$

this follows from calculating the inner expectation on the right-hand-side and using the assumption that $\mathbb{E}(R_i|\boldsymbol{B}_i) = 0$ for every i. Using a similar conditioning argument,

$$\text{var}\left(\boldsymbol{L}_{n,N,j}\middle|\boldsymbol{F}_{I_n^c}\right) = \mathbb{E}\left\{\text{var}\left(\boldsymbol{L}_{n,N,j}\middle|\boldsymbol{F}_{I_n^c},\mathcal{B}_n\right)\middle|\boldsymbol{F}_{I_n^c}\right\}.$$

Straightforward calculations now show

$$\text{var}\left(\boldsymbol{L}_{n,N,j}\middle|\boldsymbol{F}_{I_n^c},\mathcal{B}_n\right) = \frac{1}{n^2}\sum_{i\in\boldsymbol{I}_n}\left\{\hat{\gamma}(\boldsymbol{B}_i;\boldsymbol{F}_{I_n^c}) - \gamma(\boldsymbol{B}_i)\right\}^2\boldsymbol{h}_j^2(\boldsymbol{B}_i)\text{var}\left(R_i|\boldsymbol{B}_i\right),$$

implying that

$$\begin{aligned}
\text{var}\left(\boldsymbol{L}_{n,N,j}\middle|\boldsymbol{F}_{I_n^c}\right) &= \mathbb{E}\left[\frac{1}{n^2}\sum_{i\in\boldsymbol{I}_n}\left\{\hat{\gamma}(\boldsymbol{B}_i;\boldsymbol{F}_{I_n^c}) - \gamma(\boldsymbol{B}_i)\right\}^2\boldsymbol{h}_j^2(\boldsymbol{B}_i)\theta_i^2\middle|\boldsymbol{F}_{I_n^c}\right] \\
&\leq \frac{C_1C_2}{n}\mathbb{E}\left[\frac{1}{n}\sum_{i\in\boldsymbol{I}_n}\left\{\hat{\gamma}(\boldsymbol{B}_i;\boldsymbol{F}_{I_n^c}) - \gamma(\boldsymbol{B}_i)\right\}^2\middle|\boldsymbol{F}_{I_n^c}\right] \\
&= \frac{C_1C_2}{n}\left\|\hat{\gamma}(\boldsymbol{B};\boldsymbol{F}_{I_n^c}) - \gamma(\boldsymbol{B})\right\|_{P_0,2}^2 \\
&= o_p(N^{-(1+a)}),
\end{aligned}$$

the last step following from the assumptions on (7) made in the statement of the lemma and the fact that $n = O(N)$. Using a vector form of Chebyshev's inequality, it can then be shown that $\|\boldsymbol{L}_{n,N}\|_2 = o_p(N^{-(1+a)/2})$; since $\|\boldsymbol{L}_{n,N}\|_\infty \leq \|\boldsymbol{L}_{n,N}\|_2$, the stated result follows. □

1. Proof of Theorem 1

To review our main assumptions, we assume that we observe N independently identically distributed trajectories of $(\boldsymbol{X}_1, A_1, \boldsymbol{X}_2, A_2, Y) \sim P_0$. The vector $\boldsymbol{X}_1 \in \mathcal{X}_1 \subset \mathbb{R}^{p_1}$ consists of baseline covariates measured before treatment at the first decision point $A_1 \in \{0,1\}$ and the vector $\boldsymbol{X}_2 \in \mathcal{X}_2 \subset \mathbb{R}^{p_2}$ consists of intermediate covariates measured before treatment at the second decision point $A_2 \in \{0,1\}$. For notational convenience we define $\boldsymbol{S}_i^0 = (\boldsymbol{X}_{1i}^\top, A_{1i}, \boldsymbol{X}_{2i}^\top)^\top \in \mathcal{S} \subset \mathbb{R}^{p_1+p_2+1}$ and $\boldsymbol{W}_i^0 = \boldsymbol{X}_{1i} \in \mathcal{X}_1 \subset \mathbb{R}^{p_1}$. We will also have need to define the variables $\boldsymbol{S}_i$ and $\boldsymbol{W}_i, i = 1,\ldots,N$; respectively, each represents some finite dimensional function of the variables in $\boldsymbol{S}_i^0$ and $\boldsymbol{W}_i^0$. We note that knowledge of $\boldsymbol{S}_i^0$ and $\boldsymbol{W}_i^0$ respectively implies knowledge of $\boldsymbol{S}_i$ and $\boldsymbol{W}_i$; however, the reverse may not hold. The

observed outcome $Y \in \mathbb{R}$ (measured after A_2) is assumed continuous, with a larger value of Y indicating a better clinical outcome.

The developments below assume that the original sample, with elements independently and identically distributed as P_0, has been split into two independent samples, say $\boldsymbol{D}_{I_n}$ and $\boldsymbol{D}_{I_n^c}$, being respectively of sizes $n = O(N)$ and $N - n = O(N)$. The nuisance parameters $\hat{\mu}_{2Y}(\cdot)$, $\hat{\mu}_{2A}(\cdot)$, $\hat{\mu}_{1Y}(\cdot)$, and $\hat{\mu}_{1A}(\cdot)$ are estimated using the data in $\boldsymbol{D}_{I_n^c}$; the finite dimensional parameters of interest are then estimated using the data $\boldsymbol{D}_{I_n}$, treating $\hat{\mu}_{2Y}(\cdot)$, $\hat{\mu}_{2A}(\cdot)$, $\hat{\mu}_{1Y}(\cdot)$, and $\hat{\mu}_{1A}(\cdot)$ as if they were known functions. As developed here, our use of sample-splitting is a simple form of cross-fitting and can be generalized easily to make better use of the full sample (Chernozhukov et al., 2018); the simpler form used here suffices to establish the main ideas of the proofs. Lemmas 3 and 4 play an important role in several of the proofs; since $n = O(N)$, statements of the form $o_p(N^{-a})$ and $o_p(n^{-1})$ are equivalent, we use the latter to emphasize that the technical arguments rely on sample splitting, where a sample $\boldsymbol{D}_{I_n}$ of size n is used to estimate the finite dimensional parameters of interest.

To simplify notation, where needed all calculations implicitly condition on the set of selected indices $\boldsymbol{I}_n$. Using notation from the main paper, let $\Delta_{2i} = \Delta_2(\boldsymbol{S}_i^0)$, $\Delta_{1i} = \Delta_1(\boldsymbol{W}_i^0)$, $\mu_{2Ai} = \mu_{2A}(\boldsymbol{S}_i^0)$, $\hat{\mu}_{2Ai} = \hat{\mu}_{2A}(\boldsymbol{S}_i^0)$, $\mu_{1Ai} = \mu_{1A}(\boldsymbol{W}_i^0)$, and $\hat{\mu}_{1Ai} = \hat{\mu}_{1A}(\boldsymbol{W}_i^0)$. In addition, as in the main paper, we define the matrices

$$\mathbf{V}_{2n} = \frac{1}{n}\sum_{i\in\boldsymbol{I}_n}(A_{2i} - \mu_{2Ai})^2\boldsymbol{S}_i^{\otimes 2} \quad \text{and} \quad \hat{\mathbf{V}}_{2n} = \frac{1}{n}\sum_{i\in\boldsymbol{I}_n}(A_{2i} - \hat{\mu}_{2Ai})^2\boldsymbol{S}_i^{\otimes 2}$$

$$\mathbf{V}_{1n} = \frac{1}{n}\sum_{i\in\boldsymbol{I}_n}(A_{1i} - \hat{\mu}_{1Ai})^2\boldsymbol{W}_i^{\otimes 2} \quad \text{and} \quad \hat{\mathbf{V}}_{1n} = \frac{1}{n}\sum_{i\in\boldsymbol{I}_n}(A_{1i} - \hat{\mu}_{1Ai})^2\boldsymbol{W}_i^{\otimes 2}$$

where $\boldsymbol{x}^{\otimes 2} = \boldsymbol{x}\boldsymbol{x}^\top$ for any vector $\boldsymbol{x}$.

We make the following assumptions.

Assumption 1. *(i) The support of $\boldsymbol{W}^0$ and the conditional treatment effect $\Delta_1(\boldsymbol{W}^0)$ are uniformly bounded; (ii) the support of $\boldsymbol{S}^0$ and the conditional treatment effect $\Delta_2(\boldsymbol{S}^0)$ are uniformly bounded; and, the supports of $\boldsymbol{S}$ and $\boldsymbol{W}$ are uniformly bounded.*

Assumption 2. *(i) $\left\|\hat{\mu}_{1A}(\boldsymbol{W}^0; \boldsymbol{D}_{I_n^c}) - \mu_{1A}(\boldsymbol{W}^0)\right\|_{P_0,2}^2 = o_p(n^{-1/2})$; (ii) $\left\|\hat{\mu}_{2A}(\boldsymbol{S}^0; \boldsymbol{D}_{I_n^c}) - \mu_{2A}(\boldsymbol{S}^0)\right\|_{P_0,2}^2 = o_p(n^{-1/2})$.*

Assumption 3. *(i)* $\left\|\hat{\mu}_{1Y}(\boldsymbol{W}^0;\boldsymbol{D}_{I_n^c})-\mu_{1Y}(\boldsymbol{W}^0)\right\|_{P_0,2}^2=o_p(1)$; *(ii)* $\left\|\hat{\mu}_{2Y}(\boldsymbol{S}^0;\boldsymbol{D}_{I_n^c})-\mu_{2Y}(\boldsymbol{S}^0)\right\|_{P_0,2}^2=o_p(1)$.

Assumption 4. *(i)* $\left\|\hat{\mu}_{1Y}(\boldsymbol{W}^0;\boldsymbol{D}_{I_n^c})-\mu_{1Y}(\boldsymbol{W}^0)\right\|_{P_0,2}\left\|\hat{\mu}_{1A}(\boldsymbol{W}^0;\boldsymbol{D}_{I_n^c})-\mu_{1A}(\boldsymbol{W}^0)\right\|_{P_0,2}=o_p(n^{-1/2})$; *(ii)* $\left\|\hat{\mu}_{2Y}(\boldsymbol{S}^0;\boldsymbol{D}_{I_n^c})-\mu_{2Y}(\boldsymbol{S}^0)\right\|_{P_0,2}\left\|\hat{\mu}_{2A}(\boldsymbol{S}^0;\boldsymbol{D}_{I_n^c})-\mu_{2A}(\boldsymbol{S}^0)\right\|_{P_0,2}=o_p(n^{-1/2})$

Assumption 5. *There exists* $1\le n_0<\infty$ *such that* $\mathbf{V}_{jn}$ *and* $\hat{\mathbf{V}}_{jn}, j=1,2$ *are positive definite for* $n\ge n_0$.

Assumption 6. $P\big(|\boldsymbol{S}_1^\top\boldsymbol{\beta}_2^*|=0\big)=0$.

We prove this Theorem with help from the following lemma.

Lemma 5. *Suppose Assumptions 1, 2, and 5 hold. Let* $d_2=dim(\boldsymbol{S})$ *and* $d_1=dim(\boldsymbol{W})$. *Then* $\|\tilde{\boldsymbol{\beta}}_{2n}^*-\boldsymbol{\beta}_{2n}^*\|_\infty=o_p(n^{-1/2})$ *and* $\|\tilde{\boldsymbol{\beta}}_{1n}^*-\boldsymbol{\beta}_{1n}^*\|_\infty=o_p(n^{-1/2})$, *where*

$$\tilde{\boldsymbol{\beta}}_{2n}^*=\underset{\boldsymbol{\beta}_2\in\mathbb{R}^{d_2}}{\operatorname{argmin}}\sum_{i\in\boldsymbol{I}_n}\left\{A_{2i}-\hat{\mu}_{2A}(\boldsymbol{S}_i^0)\right\}^2\left\{\Delta_2(\boldsymbol{S}_i^0)-\boldsymbol{S}_i^\top\boldsymbol{\beta}_2\right\}^2,$$

$$\boldsymbol{\beta}_{2n}^*=\underset{\boldsymbol{\beta}_2\in\mathbb{R}^{d_2}}{\operatorname{argmin}}\sum_{i\in\boldsymbol{I}_n}\left\{A_{2i}-\mu_{2A}(\boldsymbol{S}_i^0)\right\}^2\left\{\Delta_2(\boldsymbol{S}_i^0)-\boldsymbol{S}_i^\top\boldsymbol{\beta}_2\right\}^2,$$

$$\tilde{\boldsymbol{\beta}}_{1n}^*=\underset{\boldsymbol{\beta}_1\in\mathbb{R}^{d_1}}{\operatorname{argmin}}\sum_{i\in\boldsymbol{I}_n}\left\{A_{1i}-\hat{\mu}_{1A}(\boldsymbol{W}_i)\right\}^2\left\{\Delta_1(\boldsymbol{W}_i^0)-\boldsymbol{W}_i^\top\boldsymbol{\beta}_1\right\}^2,$$

$$\boldsymbol{\beta}_{1n}^*=\underset{\boldsymbol{\beta}_1\in\mathbb{R}^{d_1}}{\operatorname{argmin}}\sum_{i\in\boldsymbol{I}_n}\left\{A_{1i}-\mu_{1A}(\boldsymbol{W}_i)\right\}^2\left\{\Delta_1(\boldsymbol{W}_i^0)-\boldsymbol{W}_i^\top\boldsymbol{\beta}_1\right\}^2.$$

Proof. Below, we will prove that $\|\tilde{\boldsymbol{\beta}}_{2n}^*-\boldsymbol{\beta}_{2n}^*\|_\infty=o_p(n^{-1/2})$; the result that $\|\tilde{\boldsymbol{\beta}}_{1n}^*-\boldsymbol{\beta}_{1n}^*\|_\infty=o_p(n^{-1/2})$ follows from essentially identical arguments. Using the definitions of $\tilde{\boldsymbol{\beta}}_{2n}^*$ and $\boldsymbol{\beta}_{2n}^*$ and assuming n is large enough so that Assumption 5 holds, straightforward algebra shows

$$\begin{aligned}\tilde{\boldsymbol{\beta}}_{2n}^*-\boldsymbol{\beta}_{2n}^*=&(\hat{\mathbf{V}}_{2n}^{-1}-\mathbf{V}_{2n}^{-1})\left\{\frac{1}{n}\sum_{i\in\boldsymbol{I}_n}(A_{2i}-\mu_{2Ai})^2\boldsymbol{S}_i\Delta_{2i}\right\}\\&+\hat{\mathbf{V}}_{2n}^{-1}\left\{\frac{1}{n}\sum_{i\in\boldsymbol{I}_n}(\mu_{2Ai}-\hat{\mu}_{2Ai})^2\boldsymbol{S}_i\Delta_{2i}\right\}.\end{aligned}$$

Taking norms and using the triangle inequality, it can be shown that

$$\sqrt{n}\|\tilde{\boldsymbol{\beta}}_{2n}^*-\boldsymbol{\beta}_{2n}^*\|_\infty\le\sqrt{n}\left\|\hat{\mathbf{V}}_{2n}^{-1}-\mathbf{V}_{2n}^{-1}\right\|_\infty(A_n+B_n)+\sqrt{n}\left\|\mathbf{V}_{2n}^{-1}\right\|_\infty B_n \tag{8}$$

where

$$A_n = \left\| \frac{1}{n} \sum_{i \in \boldsymbol{I}_n} (A_{2i} - \mu_{2Ai})^2 \boldsymbol{S}_i \Delta_{2i} \right\|_\infty$$
$$B_n = \left\| \frac{1}{n} \sum_{i \in \boldsymbol{I}_n} (\hat{\mu}_{2Ai} - \mu_{2Ai})^2 \boldsymbol{S}_i \Delta_{2i} \right\|_\infty .$$

Suppose that $\|\hat{\mathbf{V}}_{2n} - \mathbf{V}_{2n}\|_\infty = o_p(n^{-1/2})$. Then, by Lemma 2 and Assumption 5, we have for n sufficiently large that

$$\|\hat{\mathbf{V}}_{2n}^{-1} - \mathbf{V}_{2n}^{-1}\|_\infty \leq K_1 \|\hat{\mathbf{V}}_{2n} - \mathbf{V}_{2n}\|_\infty \tag{9}$$

for any constant K_1 such that $2\|\mathbf{V}_{2n}^{-1}\|_\infty^2 \leq K_1$. It can be seen that

$$A_n \xrightarrow{p} \left\| \mathbb{E}\left[\{A_2 - \mu_{2A}(\boldsymbol{S}^0)\}^2 \boldsymbol{S} \Delta_2(\boldsymbol{S}^0) \right] \right\|_\infty = \left\| \mathbb{E}\left[\boldsymbol{S} \Delta_2(\boldsymbol{S}^0) \mathrm{var}(A_2 | \boldsymbol{S}^0) \right] \right\|_\infty < \infty.$$

Since cross-fitting is used to estimate $\hat{\mu}_{2A}(\cdot)$, we also see that B_n is an example of (4); hence, using Lemma 3 and Assumption 2, it follows that $B_n = o_p(n^{-1/2})$. It follows from these results and (9) that (8) is $o_p(1)$.

In order to prove that $\|\hat{\mathbf{V}}_{2n} - \mathbf{V}_{2n}\|_\infty = o_p(n^{-1/2})$, we begin by writing $\|\hat{\mathbf{V}}_{2n} - \mathbf{V}_{2n}\|_\infty \leq C_n + 2D_n$, where

$$C_n = \left\| \frac{1}{n} \sum_{i \in \boldsymbol{I}_n} (\mu_{2Ai} - \hat{\mu}_{2Ai})^2 \boldsymbol{S}_i^{\otimes 2} \right\|_\infty$$
$$D_n = \left\| \frac{1}{n} \sum_{i \in \boldsymbol{I}_n} (A_{2i} - \mu_{2Ai})(\hat{\mu}_{2Ai} - \mu_{2Ai}) \boldsymbol{S}_i^{\otimes 2} \right\|_\infty .$$

Again, because cross-fitting is used to estimate $\hat{\mu}_{2A}(\cdot)$, we can see that C_n is also an example of (4) and it follows by previously stated arguments that $C_n = o_p(n^{-1/2})$. In order to establish the behavior of D_n, we first note that $D_n = \max_{j=1,\ldots,d_2} \sum_{k=1}^{d_2} |H_{njk}|$ where d_2 is finite and

$$H_{njk} = \frac{1}{n} \sum_{i \in \boldsymbol{I}_n} (A_{2i} - \mu_{2Ai})(\hat{\mu}_{2Ai} - \mu_{2Ai}) S_{ik} S_{ij}.$$

It suffices to establish the behavior of H_{njk}. First, using the definition of $\mu_{2Ai} = \mu_{2A}(\boldsymbol{S}_i^0) = \mathbb{E}(A_{2i} | \boldsymbol{S}_i^0)$ and the fact that $\hat{\mu}_{2Ai} = \hat{\mu}_{2A}(\boldsymbol{S}_i^0)$ where $\hat{\mu}_{2A}(\cdot)$ is estimated from data $\boldsymbol{D}_{I_n^c}$ that is independent of $\boldsymbol{S}_i^0 \in \boldsymbol{D}_{I_n}$ for each i, it is easy to see that $\mathbb{E}(H_{njk} | \boldsymbol{S}_1^0, \ldots, \boldsymbol{S}_n^0, \boldsymbol{D}_{I_n^c}) = 0$ and

hence that $\mathbb{E}(H_{njk}|\boldsymbol{D}_{I_n^c}) = 0$. Using these same properties, it is also easily shown that

$$\mathrm{var}(H_{njk}|\boldsymbol{D}_{I_n^c}) = \frac{1}{n^2}\sum_{i\in\boldsymbol{I}_n}\mathbb{E}\left\{(\hat{\mu}_{2Ai}-\mu_{2Ai})^2\mathrm{var}(A_{2i}|\boldsymbol{S}_i^0)(S_{ik}S_{ij})^2|\boldsymbol{D}_{I_n^c}\right\}.$$

Under Assumption 1, we can find a constant $K_2 < \infty$ such that

$$\begin{aligned}\mathrm{var}(H_{njk}|\boldsymbol{D}_{I_n^c}) &\leq \frac{K_2}{n}\mathbb{E}\left[\frac{1}{n}\sum_{i\in\boldsymbol{I}_n}\{\hat{\mu}_{2A}(\boldsymbol{S}_i^0)-\mu_{2A}(\boldsymbol{S}_i^0)\}^2\Big|\boldsymbol{D}_{I_n^c}\right]\\ &= \frac{K_2}{n}\mathbb{E}\left\{\|\hat{\mu}_{2A}(\boldsymbol{S}^0;\boldsymbol{D}_{I_n^c})-\mu_{2A}(\boldsymbol{S}^0)\|^2_{\mathbb{P}_n,2}\Big|\boldsymbol{D}_{I_n^c}\right\}\\ &= \frac{K_2}{n}\|\hat{\mu}_{2A}(\boldsymbol{S}^0;\boldsymbol{D}_{I_n^c})-\mu_{2A}(\boldsymbol{S}^0)\|^2_{P_0,2},\end{aligned}$$

By Chebyshev's inequality, for all $\epsilon > 0$ we then have

$$P\left(n^{1/2}|H_{njk}| > \epsilon\Big|\boldsymbol{D}_{I_n^c}\right) \leq \frac{K_2}{\epsilon^2}\|\hat{\mu}_{2A}(\boldsymbol{S}^0;\boldsymbol{D}_{I_n^c})-\mu_{2A}(\boldsymbol{S}^0)\|^2_{P_0,2},$$

where the right-hand side is $o_p(n^{-1/2})$ by Assumption 2. Lemma 1 now implies that $H_{njk} = o_p(n^{-1/2})$ and hence that $D_n = o_p(n^{-1/2})$. Therefore, $\|\hat{\mathbf{V}}_{2n}-\mathbf{V}_{2n}\|_\infty \leq C_n + 2D_n = o_p(n^{-1/2})$, proving the desired result. A similar argument shows $\sqrt{n}\|\tilde{\boldsymbol{\beta}}_{1n}^* - \boldsymbol{\beta}_{1n}^*\|_\infty = o_p(1)$. □

As in the main paper, we define

$$\boldsymbol{\beta}_2^* = \underset{\boldsymbol{\beta}_2\in\mathbb{R}^{d_2}}{\mathrm{argmin}}\,\mathbb{E}\left[\mathrm{var}\left(A_2|\boldsymbol{S}^0\right)\left\{\Delta_2(\boldsymbol{S}^0)-\boldsymbol{S}^\top\boldsymbol{\beta}_2\right\}^2\right], \tag{10}$$

$$\mathbf{V}_2 = \mathbb{E}\left\{\mathrm{var}\left(A_2|\boldsymbol{S}^0\right)\boldsymbol{S}^{\otimes 2}\right\} \tag{11}$$

$$\boldsymbol{\beta}_1^* = \underset{\boldsymbol{\beta}_1\in\mathbb{R}^{d_1}}{\mathrm{argmin}}\,\mathbb{E}\left[\mathrm{var}\left(A_1|\boldsymbol{W}^0\right)\left\{\Delta_1(\boldsymbol{W}^0)-\boldsymbol{W}^\top\boldsymbol{\beta}_1\right\}^2\right], \tag{12}$$

$$\mathbf{V}_1 = \mathbb{E}\left\{\mathrm{var}\left(A_1|\boldsymbol{W}^0\right)\boldsymbol{W}^{\otimes 2}\right\}. \tag{13}$$

As $n\to\infty$, it is easy to see that $\boldsymbol{\beta}_{2n}^*$ converges to

$$\boldsymbol{\beta}_2^* = \mathbf{V}_2^{-1}\mathbb{E}\left\{\mathrm{var}\left(A_2|\boldsymbol{S}^0\right)\boldsymbol{S}\Delta_2(\boldsymbol{S}^0)\right\};$$

similarly, $\boldsymbol{\beta}_{1n}^*$ converges to

$$\boldsymbol{\beta}_1^* = \mathbf{V}_1^{-1}\mathbb{E}\left\{\mathrm{var}\left(A_1|\boldsymbol{W}^0\right)\boldsymbol{W}\Delta_1(\boldsymbol{W}^0)\right\}.$$

With these preliminaries in place, we can now prove the main result.

Proof of Theorem 1, part (a). We desire to show that $\hat{\boldsymbol{\beta}}_{2n}$ is an asymptotically linear estimator of $\boldsymbol{\beta}_2^*$ with the claimed influence function, where $\boldsymbol{\beta}_2^*$ is defined in (10) and

$$\hat{\boldsymbol{\beta}}_{2n} = \underset{\boldsymbol{\beta}_2 \in \mathbb{R}^{d_2}}{\operatorname{argmin}} \sum_{i \in \boldsymbol{I}_n} \left[Y_i - \hat{\mu}_{2Y}(\boldsymbol{S}_i^0) - \{A_{2i} - \hat{\mu}_{2A}(\boldsymbol{S}_i^0)\} \cdot \boldsymbol{S}_i^\top \boldsymbol{\beta}_2 \right]^2 .$$

In view of Lemma 5, we can proceed by establishing asymptotically linear representations for both $\sqrt{n}(\hat{\boldsymbol{\beta}}_{2n} - \tilde{\boldsymbol{\beta}}_{2n}^*)$ and $\sqrt{n}(\boldsymbol{\beta}_{2n}^* - \boldsymbol{\beta}_2^*)$; combined, these will lead to that for $\sqrt{n}(\hat{\boldsymbol{\beta}}_{2n} - \boldsymbol{\beta}_2^*)$.

Recalling notation introduced earlier, it is easy to show that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2n} - \tilde{\boldsymbol{\beta}}_{2n}^*) = \hat{\mathbf{V}}_{2n}^{-1} \left[\frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} (A_{2i} - \hat{\mu}_{2Ai}) \boldsymbol{S}_i \{ Y_i - \hat{\mu}_{2Yi} - (A_{2i} - \hat{\mu}_{2Ai}) \Delta_{2i} \} \right] \tag{14}$$

By adding and subtracting terms and using the model assumptions on Y_i, one can write

$$Y_i - \hat{\mu}_{2Yi} - (A_{2i} - \hat{\mu}_{2Ai})\Delta_{2i} = \epsilon_{2i} + (\mu_{2Yi} - \hat{\mu}_{2Yi}) + (\hat{\mu}_{2Ai} - \mu_{2Ai})\Delta_{2i}. \tag{15}$$

The decomposition (15) implies that the term in the square brackets on the right-hand side of (14) can be decomposed into six terms:

$$\frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} (A_{2i} - \hat{\mu}_{2Ai}) \boldsymbol{S}_i \{ Y_i - \hat{\mu}_{2Yi} - (A_{2i} - \mu_{2Ai}) \Delta_{2i} \} \tag{16}$$

$$= \frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} \epsilon_{2i} (A_{2i} - \mu_{2Ai}) \boldsymbol{S}_i \tag{17}$$

$$- \frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} (A_{2i} - \mu_{2Ai})(\hat{\mu}_{2Yi} - \mu_{2Yi}) \boldsymbol{S}_i \tag{18}$$

$$+ \frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} (A_{2i} - \mu_{2Ai})(\hat{\mu}_{2Ai} - \mu_{2Ai}) \Delta_{2i} \boldsymbol{S}_i \tag{19}$$

$$- \frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} \epsilon_{2i} (\hat{\mu}_{2Ai} - \mu_{2Ai}) \boldsymbol{S}_i \tag{20}$$

$$+ \frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} (\hat{\mu}_{2Ai} - \mu_{2Ai})(\hat{\mu}_{2Yi} - \mu_{2Yi}) \boldsymbol{S}_i \tag{21}$$

$$- \frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} (\hat{\mu}_{2Ai} - \hat{\mu}_{2Ai})^2 \Delta_{2i} \boldsymbol{S}_i. \tag{22}$$

Under the assumptions of this theorem, the central limit theorem establishes the asymptotic normality of (17), which is $O_p(1)$. The terms (18)-(20) are each seen to be examples to which Lemma 4 applies; under Assumptions 1-5, it follows that each term is $o_p(1)$. The terms (21) and (22) are both seen to be examples to which Lemma 3 applies; again, under Assumptions 1-5, each term is $o_p(1)$. Because $\|\hat{\mathbf{V}}_{2n}^{-1} - \mathbf{V}_{2n}^{-1}\|_\infty = o_p(1)$ it follows that $\sqrt{n}(\hat{\boldsymbol{\beta}}_{2n} - \tilde{\boldsymbol{\beta}}_{2n}^*)$ can be written

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2n} - \tilde{\boldsymbol{\beta}}_{2n}^*) = \frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} \mathbf{V}_{2n}^{-1} \epsilon_{2i}(A_{2i} - \mu_{2Ai}) \boldsymbol{S}_i + o_p(1). \tag{23}$$

Turning to $\sqrt{n}(\boldsymbol{\beta}_{2n}^* - \boldsymbol{\beta}_2^*)$, we can write

$$\begin{aligned} \sqrt{n}(\boldsymbol{\beta}_{2n}^* - \boldsymbol{\beta}_2^*) &= \sqrt{n}\left[\left\{n^{-1} \sum_{i \in \boldsymbol{I}_n} \mathbf{V}_{2n}^{-1}(A_{2i} - \mu_{2Ai})^2 \Delta_{2i} \boldsymbol{S}_i\right\} - \boldsymbol{\beta}_2^*\right] \\ &= \frac{1}{\sqrt{n}} \mathbf{V}_{2n}^{-1} \sum_{i \in \boldsymbol{I}_n} (A_{2i} - \mu_{2Ai})^2 \boldsymbol{S}_i \big(\Delta_{2i} - \boldsymbol{S}_i^\top \boldsymbol{\beta}_2^*\big). \end{aligned} \tag{24}$$

Hence, using (23) and (24) and collecting terms,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2n} - \boldsymbol{\beta}_2^*) = \frac{1}{\sqrt{n}} \mathbf{V}_{2n}^{-1} \sum_{i \in \boldsymbol{I}_n} (A_{2i} - \mu_{2Ai}) \boldsymbol{S}_i H_{2i} + o_p(1)$$

where $H_{2i} = \epsilon_{2i} + (A_{2i} - \mu_{2Ai})\big(\Delta_{2i} - \boldsymbol{S}_i^\top \boldsymbol{\beta}_2^*\big)$. Using the fact that $Y_i - \mu_{2Yi} = \epsilon_{2i} + (A_{2i} - \mu_{2Ai})\Delta_{2i}$, we may write

$$\begin{aligned} H_{2i} &= Y_i - \mu_{2Yi} - (A_{2i} - \mu_{2Ai})\Delta_{2i} + (A_{2i} - \mu_{2Ai})\big(\Delta_{2i} - \boldsymbol{S}_i^\top \boldsymbol{\beta}_2^*\big) \\ &= Y_i - \mu_{2Yi} - (A_{2i} - \mu_{2Ai}) \boldsymbol{S}_i^\top \boldsymbol{\beta}_2^*. \end{aligned} \tag{25}$$

Consequently,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2n} - \boldsymbol{\beta}_2^*) = \frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} \mathrm{Inf}_{2in} + o_p(1)$$

where

$$\mathrm{Inf}_{2in} = \mathbf{V}_{2n}^{-1}(A_{2i} - \mu_{2Ai}) \boldsymbol{S}_i \big\{Y_i - \mu_{2Yi} - (A_{2i} - \mu_{2Ai}) \boldsymbol{S}_i^\top \boldsymbol{\beta}_2^*\big\}.$$

Since $\|\mathbf{V}_{2n}^{-1} - \mathbf{V}_2^{-1}\|_\infty = o_p(1)$, it now follows that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2n} - \boldsymbol{\beta}_2^*) = \frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} \mathrm{Inf}_{2i} + o_p(1),$$

where

$$\mathrm{Inf}_{2i} = \mathbf{V}_2^{-1}(A_{2i} - \mu_{2Ai})\boldsymbol{S}_i\big\{Y_i - \mu_{2Yi} - (A_{2i} - \mu_{2Ai})\boldsymbol{S}_i^\top\boldsymbol{\beta}_2^*\big\} \tag{26}$$

has mean zero and variance $\mathbf{V}_2^{-1}\boldsymbol{Q}_2\mathbf{V}_2^{-1}$ where $\mathbf{Q}_2 = \mathbb{E}(\boldsymbol{J}_2^{\otimes 2})$ and

$$\boldsymbol{J}_2 = \{A_2 - \mu_{2A}(\boldsymbol{S}^0)\}\boldsymbol{S}\left[Y - \mu_{2Y}(\boldsymbol{S}^0) - \{A_2 - \mu_{2A}(\boldsymbol{S}^0)\}\boldsymbol{S}^\top\boldsymbol{\beta}_2^*\right]. \tag{27}$$

□

Proof of Theorem 1, part (b). As in part (a), we need to show that $\hat{\boldsymbol{\beta}}_{1n}$ is an asymptotically linear estimator of $\boldsymbol{\beta}_1^*$ with a certain influence function, where $\boldsymbol{\beta}_1^*$ is defined in (12) and

$$\hat{\boldsymbol{\beta}}_{1n} = \underset{\boldsymbol{\beta}_1\in\mathbb{R}^{d_1}}{\operatorname{argmin}} \sum_{i\in\boldsymbol{I}_n}\left[\hat{\tilde{Y}}_i - \hat{\mu}_{1Y}(\boldsymbol{W}_i^0) - \{A_{1i} - \hat{\mu}_{1A}(\boldsymbol{W}_i^0)\}\cdot\boldsymbol{W}_i^\top\boldsymbol{\beta}_1\right]^2,$$

where $\hat{\tilde{Y}}_i$ is calculated as

$$\hat{\tilde{Y}}_i = Y_i + \boldsymbol{S}_i^\top\hat{\boldsymbol{\beta}}_{2n}\big\{I(\boldsymbol{S}_i^\top\hat{\boldsymbol{\beta}}_{2n} > 0) - A_{2i}\big\}. \tag{28}$$

Proceeding similarly to the proof of part (a), we will establish asymptotically linear representations for both $\sqrt{n}(\hat{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{1n}^*)$ and $\sqrt{n}(\boldsymbol{\beta}_{1n}^* - \boldsymbol{\beta}_1^*)$; combining these will provide the claimed influence function for $\sqrt{n}(\hat{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_1^*)$.

We begin with $\sqrt{n}(\hat{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{1n}^*)$. Define $\epsilon_{1i} = \tilde{Y}_i - \mathbb{E}(\tilde{Y}_i \mid \boldsymbol{W}_i^0, A_{1i})$, where $\tilde{Y}_i$ is given by

$$\tilde{Y}_i = Y_i + \boldsymbol{S}_i^\top\boldsymbol{\beta}_2^*\big\{I(\boldsymbol{S}_i^\top\boldsymbol{\beta}_2^* > 0) - A_{2i}\big\}; \tag{29}$$

by construction, $\epsilon_{1i} = \tilde{Y}_i - \mu_{1Ai} - (A_{1i} - \mu_{1Ai})\Delta_{1i}$ and $\mathbb{E}(\epsilon_{1i} \mid \boldsymbol{W}_i^0, A_{1i}) = 0$. In addition, let $\hat{\delta}_i = \hat{\tilde{Y}}_i - \tilde{Y}_i$. Similarly to the proof in part (a), we can decompose $\sqrt{n}(\hat{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{1n}^*)$ into the sum of several terms:

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{1n}-\tilde{\boldsymbol{\beta}}^*_{1n}) =$$

$$\hat{\mathbf{V}}_{1n}^{-1}\Big\{\frac{1}{\sqrt{n}}\sum_{i\in\boldsymbol{I}_n}(A_{1i}-\mu_{1Ai})\big(\epsilon_{1i}+\hat{\delta}_i\big)\boldsymbol{W}_i\Big\} \tag{30}$$

$$-\hat{\mathbf{V}}_{1n}^{-1}\Big\{\frac{1}{\sqrt{n}}\sum_{i\in\boldsymbol{I}_n}(A_{1i}-\mu_{1Ai})(\hat{\mu}_{1Yi}-\mu_{1Yi})\boldsymbol{W}_i\Big\} \tag{31}$$

$$+\hat{\mathbf{V}}_{1n}^{-1}\Big\{\frac{1}{\sqrt{n}}\sum_{i\in\boldsymbol{I}_n}(A_{1i}-\mu_{1Ai})(\hat{\mu}_{1Ai}-\mu_{1Ai})\Delta_{1i}\boldsymbol{W}_i\Big\} \tag{32}$$

$$-\hat{\mathbf{V}}_{1n}^{-1}\Big\{\frac{1}{\sqrt{n}}\sum_{i\in\boldsymbol{I}_n}(\hat{\mu}_{1Ai}-\mu_{1Ai})\epsilon_{1i}\boldsymbol{W}_i\Big\} \tag{33}$$

$$-\hat{\mathbf{V}}_{1n}^{-1}\Big\{\frac{1}{\sqrt{n}}\sum_{i\in\boldsymbol{I}_n}(\hat{\mu}_{1Ai}-\mu_{1Ai})\hat{\delta}_i\boldsymbol{W}_i\Big\} \tag{34}$$

$$-\hat{\mathbf{V}}_{1n}^{-1}\Big\{\frac{1}{\sqrt{n}}\sum_{i\in\boldsymbol{I}_n}(\hat{\mu}_{1Ai}-\mu_{1Ai})(\hat{\mu}_{1Yi}-\mu_{1Yi})\boldsymbol{W}_i\Big\} \tag{35}$$

$$+\hat{\mathbf{V}}_{1n}^{-1}\Big\{\frac{1}{\sqrt{n}}\sum_{i\in\boldsymbol{I}_n}(\hat{\mu}_{1Ai}-\mu_{1Ai})^2\Delta_{1i}\boldsymbol{W}_i\Big\} \tag{36}$$

Assuming that $\hat{\mu}_{1A}(\cdot)$ and $\hat{\mu}_{1Y}(\cdot)$ are estimated similarly to $\hat{\mu}_{2A}(\cdot)$ and $\hat{\mu}_{2Y}(\cdot)$ (i.e., meaning, sample splitting has been used) and in view of the fact that $\hat{\mathbf{V}}_{1n}$ is a consistent estimator of $\mathbf{V}_1$, Lemma 4 implies that the terms (31), (32), and (33) are all $o_p(1)$ under Assumptions 1 – 5; similarly, Lemma 3 implies that the terms (35) and (36) are also $o_p(1)$ under these same assumptions. To complete this part of the proof, we must therefore establish the asymptotic behavior of (30) and (34), both of which depend on the asymptotic behavior of $\hat{\delta}_i = \hat{\tilde{Y}}_i - \tilde{Y}_i$. The terms (30) and (34) isolate the potential for non-regular behavior; however, as we will see, Assumption 6 is only needed for controlling such behavior in (30).

To determine the asymptotic behavior of (30), let $D_i = I(\boldsymbol{S}_i^\top\boldsymbol{\beta}_2^* > 0) - A_{2i}$ and

$$\hat{R}_{ni} = I(\boldsymbol{S}_i^\top\hat{\boldsymbol{\beta}}_{2n} > 0) - I(\boldsymbol{S}_i^\top\boldsymbol{\beta}_2^* > 0)$$

for $i \in \boldsymbol{I}_n$.

Algebra now shows

$$\frac{1}{\sqrt{n}}\sum_{i\in\boldsymbol{I}_n}(A_{1i}-\mu_{1Ai})\boldsymbol{W}_i\big(\epsilon_{1i}+\hat{\delta}_i\big) =$$

$$\frac{1}{\sqrt{n}}\sum_{i\in\boldsymbol{I}_n}(A_{1i}-\mu_{1Ai})\boldsymbol{W}_i\{\epsilon_{1i}+\boldsymbol{S}_i^\top(\hat{\boldsymbol{\beta}}_{2n}-\boldsymbol{\beta}_2^*)D_i\} \tag{37}$$

$$+\frac{1}{\sqrt{n}}\sum_{i\in\boldsymbol{I}_n}(A_{1i}-\mu_{1Ai})\boldsymbol{W}_i\boldsymbol{S}_i^\top\boldsymbol{\beta}_2^*\hat{R}_{ni} \tag{38}$$

$$+\frac{1}{\sqrt{n}}\sum_{i\in\boldsymbol{I}_n}(A_{1i}-\mu_{1Ai})\boldsymbol{W}_i\boldsymbol{S}_i^\top(\hat{\boldsymbol{\beta}}_{2n}-\boldsymbol{\beta}_2^*)\hat{R}_{ni}. \tag{39}$$

Although (38) and (39) can be easily combined, treating these two terms separately turns out to be advantageous. We first consider (38). Note that $\hat{R}_{ni}\in\{-1,0,1\}$ and, importantly, that $|\hat{R}_{ni}|\le R_{ni}$, where $R_{ni}=I\{0\le|\boldsymbol{S}_i^\top\boldsymbol{\beta}_2^*|\le|\boldsymbol{S}_i^\top(\hat{\boldsymbol{\beta}}_{2n}-\boldsymbol{\beta}_2^*)|\}$. It follows that

$$|\boldsymbol{S}_i^\top\boldsymbol{\beta}_2^*|R_{ni}\le|\boldsymbol{S}_i^\top(\hat{\boldsymbol{\beta}}_{2n}-\boldsymbol{\beta}_2^*)|R_{ni},$$

an inequality that is trivially true when $R_{ni}=0$ and true for $R_{ni}=1$ in view of its definition. Consequently, considering the jth element of the vector $\boldsymbol{W}_i$, we have

$$\begin{aligned}
\Big|\sum_{i\in\boldsymbol{I}_n}(A_{1i}-\mu_{1Ai})W_{ij}\boldsymbol{S}_i^\top\boldsymbol{\beta}_2^*\hat{R}_{ni}\Big| &\le \sum_{i\in\boldsymbol{I}_n}\big|(A_{1i}-\mu_{1Ai})W_{ij}\big|\big|\boldsymbol{S}_i^\top\boldsymbol{\beta}_2^*\big|\big|\hat{R}_{ni}\big| \\
&\le \sum_{i\in\boldsymbol{I}_n}\big|(A_{1i}-\mu_{1Ai})W_{ij}\big|\big|\boldsymbol{S}_i^\top\boldsymbol{\beta}_2^*\big|R_{ni} \\
&\le \sum_{i\in\boldsymbol{I}_n}\big|A_{1i}-\mu_{1Ai}\big|\big|W_{ij}\big|\big|\boldsymbol{S}_i^\top(\hat{\boldsymbol{\beta}}_{2n}-\boldsymbol{\beta}_2^*)\big|R_{ni} \\
&\le 4C\sum_{i\in\boldsymbol{I}_n}\big|\boldsymbol{S}_i^\top(\hat{\boldsymbol{\beta}}_{2n}-\boldsymbol{\beta}_2^*)\big|R_{ni},
\end{aligned}$$

the last step following from the fact that A_{1i} is binary, $\hat{\mu}_{1A}(\cdot)\in[0,1]$, and $|W_{ij}|$ is bounded, say, by a finite constant C. Considering (39), a similar calculation shows that

$$\Big|\sum_{i\in\boldsymbol{I}_n}(A_{1i}-\mu_{1Ai})\boldsymbol{W}_i\boldsymbol{S}_i^\top(\hat{\boldsymbol{\beta}}_{2n}-\boldsymbol{\beta}_2^*)\hat{R}_{ni}\Big| \le 4C\sum_{i\in\boldsymbol{I}_n}\big|\boldsymbol{S}_i^\top(\hat{\boldsymbol{\beta}}_{2n}-\boldsymbol{\beta}_2^*)\big|R_{ni}.$$

Therefore,

$$\begin{aligned}(38)+(39) &\leq 8C\frac{1}{\sqrt{n}}\sum_{i\in\boldsymbol{I}_n}\left|\boldsymbol{S}_i^\top(\hat{\boldsymbol{\beta}}_{2n}-\boldsymbol{\beta}_2^*)\right|R_{ni}\\ &\leq 8C\sqrt{n}\left[\frac{1}{n}\sum_{i\in\boldsymbol{I}_n}\left\{\boldsymbol{S}_i^\top(\hat{\boldsymbol{\beta}}_{2n}-\boldsymbol{\beta}_2^*)\right\}^2\right]^{1/2}\left[\frac{1}{n}\sum_{i\in\boldsymbol{I}_n}R_{ni}^2\right]^{1/2}\\ &= 8C\lambda_{1n}\sqrt{n}\|\hat{\boldsymbol{\beta}}_{2n}-\boldsymbol{\beta}_2^*\|_2\,\bar{R}_n^{1/2},\end{aligned}$$

where $\|\cdot\|_2$ is the usual Euclidean vector norm and $\lambda_{1n}>0$ is the square root of the maximum eigenvalue of $n^{-1}\sum_{i\in\boldsymbol{I}_n}\boldsymbol{S}_i^{\otimes 2}$. Because $\sqrt{n}\|\hat{\boldsymbol{\beta}}_{2n}-\boldsymbol{\beta}_2^*\|_2=O_p(1)$ and λ_{1n} converges to a finite constant as $n\to\infty$ under our assumptions, it follows that (38) + (39) is $o_p(1)$ if $\bar{R}_n=o_p(1)$. However, Markov's inequality implies that

$$P(\bar{R}_n>\alpha\mid\boldsymbol{S}_1^0,\ldots,\boldsymbol{S}_n^0)\leq(n\alpha)^{-1}\sum_{i\in\boldsymbol{I}_n}\mathbb{E}(R_{ni}\mid\boldsymbol{S}_1^0,\ldots,\boldsymbol{S}_n^0)$$

for any $\alpha>0$, where $\mathbb{E}(R_{ni}\mid\boldsymbol{S}_1^0,\ldots,\boldsymbol{S}_n^0)=P\{R_{ni}=1\mid\boldsymbol{S}_1^0,\ldots,\boldsymbol{S}_n^0\}$. Letting

$$I_i^*=I\big(|\boldsymbol{S}_i^\top\boldsymbol{\beta}_2^*|=0\big),$$

an easy conditioning argument shows $\mathbb{E}(R_{ni}\mid\boldsymbol{S}_1^0,\ldots,\boldsymbol{S}_n^0)=I_i^*+(1-I_i^*)P_{ni}$ where

$$P_{ni}=P\{|\boldsymbol{S}_i^\top(\hat{\boldsymbol{\beta}}_{2n}-\boldsymbol{\beta}_2^*)|\geq k_i\mid(\boldsymbol{S}_1^0,\ldots,\boldsymbol{S}_n^0),\,k_i>0\}$$

for $k_i=|\boldsymbol{S}_i^\top\boldsymbol{\beta}_2^*|$. Letting $n\to\infty$, the fact that $\hat{\boldsymbol{\beta}}_{2n}\xrightarrow{p}\boldsymbol{\beta}_2^*$ as $n\to\infty$ implies $P_{ni}\to 0$ for each i; hence,

$$\lim_{n\to\infty}P(\bar{R}_n>\alpha\mid\boldsymbol{S}_1^0,\ldots,\boldsymbol{S}_n^0)\leq\frac{1}{n\alpha}\lim_{n\to\infty}\sum_{i\in\boldsymbol{I}_n}I_i^*.$$

However, under our assumptions,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i\in\boldsymbol{I}_n}I_i^*\xrightarrow{p}P\big(|\boldsymbol{S}_1^\top\boldsymbol{\beta}_2^*|=0\big)$$

and it follows from Assumption 6 that $\bar{R}_n\xrightarrow{p}0$.

Because $\|\hat{\mathbf{V}}_{1n}^{-1}-\mathbf{V}_{1n}^{-1}\|_\infty=o_p(1)$, we have now shown that

$$(30)=\mathbf{V}_{1n}^{-1}\Big[\frac{1}{\sqrt{n}}\sum_{i\in\boldsymbol{I}_n}(A_{1i}-\mu_{1Ai})\boldsymbol{W}_i\{\epsilon_{1i}+\boldsymbol{S}_i^\top(\hat{\boldsymbol{\beta}}_{2n}-\boldsymbol{\beta}_2^*)D_i\}+o_p(1)\Big]\tag{40}$$

under Assumptions 1 – 6, where we recall the notation $D_i = I(\boldsymbol{S}_i^\top \boldsymbol{\beta}_2^* > 0) - A_{2i}$.

To establish (34), observe that we may similarly decompose it as above, leading to

$$\frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} (\hat{\mu}_{1Ai} - \mu_{1Ai}) \boldsymbol{W}_i \hat{\delta}_i =$$

$$\frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} (\hat{\mu}_{1Ai} - \mu_{1Ai}) \boldsymbol{W}_i \epsilon_{1i} \tag{41}$$

$$+ \frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} (\hat{\mu}_{1Ai} - \mu_{1Ai}) \boldsymbol{W}_i \boldsymbol{S}_i^\top \boldsymbol{\beta}_2^* \hat{R}_{ni} \tag{42}$$

$$+ \frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} (\hat{\mu}_{1Ai} - \mu_{1Ai}) \boldsymbol{W}_i \boldsymbol{S}_i^\top (\hat{\boldsymbol{\beta}}_{2n} - \boldsymbol{\beta}_2^*) \hat{R}_{ni} \tag{43}$$

$$+ \frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} (\hat{\mu}_{1Ai} - \mu_{1Ai}) \boldsymbol{W}_i \boldsymbol{S}_i^\top (\hat{\boldsymbol{\beta}}_{2n} - \boldsymbol{\beta}_2^*) D_i. \tag{44}$$

The term (41) can be handled using Lemma 4. The remaining terms can be handled similarly to (38) and (39); however, the required decomposition of terms differs some and, importantly, can make use of Assumption 2. In particular, establishing the behavior of (41)-(44) can be done under Assumptions 1 – 5, without additionally imposing Assumption 6, showing that any effect of non-regularity is limited to the behavior of (38) and (39), or equivalently,

$$\frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} (A_{1i} - \mu_{1Ai}) \boldsymbol{W}_i \boldsymbol{S}_i^\top \hat{\boldsymbol{\beta}}_{2n} \hat{R}_{ni}.$$

The above proof establishes an asymptotic linear representation for $\sqrt{n}(\hat{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{1n}^*)$. Turning to $\sqrt{n}(\boldsymbol{\beta}_{1n}^* - \boldsymbol{\beta}_1^*)$, we can write

$$\sqrt{n}(\boldsymbol{\beta}_{1n}^* - \boldsymbol{\beta}_1^*) = \frac{1}{\sqrt{n}} \mathbf{V}_{1n}^{-1} \sum_{i \in \boldsymbol{I}_n} (A_{1i} - \mu_{1Ai})^2 \boldsymbol{W}_i \big(\Delta_{1i} - \boldsymbol{W}_i^\top \boldsymbol{\beta}_1^*\big). \tag{45}$$

Hence, using (40) and (45) and collecting terms, it follows that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_1^*) = \frac{1}{\sqrt{n}} \mathbf{V}_{1n}^{-1} \sum_{i \in \boldsymbol{I}_n} (A_{1i} - \mu_{1Ai}) \boldsymbol{W}_i H_{1i} + o_p(1)$$

where $H_{1i} = \epsilon_{1i} + (A_{1i} - \mu_{1Ai})\big(\Delta_{1i} - \boldsymbol{W}_i^\top \boldsymbol{\beta}_1^*\big) + \boldsymbol{S}_i^\top (\hat{\boldsymbol{\beta}}_{2n} - \boldsymbol{\beta}_2^*) D_i$. Because

$$\epsilon_{1i} = \tilde{Y}_i - \mu_{1Ai} - (A_{1i} - \mu_{1Ai}) \Delta_{1i}$$

we have

$$H_{1i} = \tilde{Y}_i - \mu_{1Ai} - (A_{1i} - \mu_{1Ai})\boldsymbol{W}_i^\top \boldsymbol{\beta}_1^* + \boldsymbol{S}_i^\top (\hat{\boldsymbol{\beta}}_{2n} - \boldsymbol{\beta}_2^*)\{I(\boldsymbol{S}_i^\top \boldsymbol{\beta}_2^* > 0) - A_{2i}\}.$$

Using the fact that $\|\mathbf{V}_{1n}^{-1} - \mathbf{V}_1^{-1}\|_\infty = o_p(1)$, it follows that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_1^*) = \frac{1}{\sqrt{n}} \mathbf{V}_1^{-1} \sum_{i \in \boldsymbol{I}_n} (A_{1i} - \mu_{1Ai}) \boldsymbol{W}_i H_{1i} + o_p(1).$$

Defining

$$\boldsymbol{K}_n = \frac{1}{n} \sum_{i \in \boldsymbol{I}_n} (A_{1i} - \mu_{1Ai})\{I(\boldsymbol{S}_i^\top \boldsymbol{\beta}_2^* > 0) - A_{2i}\} \boldsymbol{W}_i \boldsymbol{S}_i^\top$$

and letting $\boldsymbol{K}$ denote its limit in probability, the results from part (a), in particular (26), now imply that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_1^*) = \frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} \mathrm{Inf}_{1i} + o_p(1).$$

for

$$\mathrm{Inf}_{1i} = \mathbf{V}_1^{-1}\big[(A_{1i} - \mu_{1Ai})\boldsymbol{W}_i\big\{\tilde{Y}_i - \mu_{1Ai} - (A_{1i} - \mu_{1Ai})\boldsymbol{W}_i^\top \boldsymbol{\beta}_1^*\big\} + \boldsymbol{K}\,\mathrm{Inf}_{2i}\big]. \tag{46}$$

This representation result implies $\sqrt{n}(\hat{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_1^*) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_1^{-1}\mathbf{Q}_1\mathbf{V}_1^{-1})$ where we define the matrix $\mathbf{Q}_1 = \mathbb{E}\{(\boldsymbol{J}_1 + \boldsymbol{K}\mathbf{V}_2^{-1}\boldsymbol{J}_2)^{\otimes 2}\}$, $\boldsymbol{J}_2$ is given in (27),

$$\boldsymbol{J}_1 = \{A_1 - \mu_{1A}(\boldsymbol{W}^0)\}\boldsymbol{W}\left[\tilde{Y} - \mu_{1A}(\boldsymbol{W}^0) - \{A_1 - \mu_{1A}(\boldsymbol{W}^0)\}\boldsymbol{W}^\top \boldsymbol{\beta}_1^*\right]$$

and

$$\boldsymbol{K} = \mathbb{E}\big[\{A_1 - \mu_{1A}(\boldsymbol{W}^0)\}\{I(\boldsymbol{S}^\top \boldsymbol{\beta}_2^* > 0) - A_2\}\boldsymbol{W}\boldsymbol{S}^\top\big].$$

□

Proof of Corollary to Theorem 1. As established in the proof of Theorem 1, the regularity Assumption 6 is imposed only to control the potentially non-regular behavior of the terms (38) and (39). The origin of this non-regular behavior is the dependence of each term on

$$\hat{R}_{ni} = I(\boldsymbol{S}_i^\top \hat{\boldsymbol{\beta}}_{2n} > 0) - I(\boldsymbol{S}_i^\top \boldsymbol{\beta}_2^* > 0), i \in \boldsymbol{I}_n.$$

In view of the proof of Theorem 1, establishing that each of (38) and (39) is $o_p(1)$ is sufficient to prove the corollary as stated.

To simply the proof of these results, let $\hat{\boldsymbol{\beta}}_{2n}^{(-i)}$ be the least squares estimator based on the subset of subjects that excludes subject i, and define

$$\hat{\tilde{R}}_{ni} = I(\boldsymbol{S}_i^\top \hat{\boldsymbol{\beta}}_{2n}^{(-i)} > 0) - I(\boldsymbol{S}_i^\top \boldsymbol{\beta}_2^* > 0), i \in \boldsymbol{I}_n$$

and also

$$\tilde{R}_{ni} = I\{0 \leq |\boldsymbol{S}_i^\top \boldsymbol{\beta}_2^*| \leq |\boldsymbol{S}_i^\top (\hat{\boldsymbol{\beta}}_{2n}^{(-i)} - \boldsymbol{\beta}_2^*)|\};$$

similarly to before, $|\hat{\tilde{R}}_{ni}| \leq \tilde{R}_{ni}$.

Then, considering (38) with $\hat{R}_{ni}$ replaced by $\hat{\tilde{R}}_{ni}$, we may write

$$E\left[(38)\right] = \frac{1}{\sqrt{n}} \sum_{i \in \boldsymbol{I}_n} E\left(\boldsymbol{W}_i \boldsymbol{S}_i^\top \boldsymbol{\beta}_2^* \hat{\tilde{R}}_{ni}\, E\left[A_{1i} - \mu_{1Ai} \mid \boldsymbol{W}_i, \boldsymbol{S}_i, \boldsymbol{I}_n, \hat{\boldsymbol{\beta}}_{2n}^{(-i)}\right] \mid \boldsymbol{I}_n\right).$$

In view of the definition of $\hat{\boldsymbol{\beta}}_{2n}^{(-i)}$, we have

$$E\left[A_{1i} - \mu_{1Ai} \mid \boldsymbol{W}_i, \boldsymbol{S}_i, \boldsymbol{I}_n, \hat{\boldsymbol{\beta}}_{2n}^{(-i)}\right] = E\left[A_{1i} - \mu_{1Ai} \mid \boldsymbol{W}_i, \boldsymbol{S}_i, \boldsymbol{I}_n\right] = 0,$$

the last equality following by assumption. Therefore, it follows that $E\left[(38) \mid \boldsymbol{I}_n\right] = 0$. Arguing similarly and using the conditional variance formula,

$$var\left[(38) \mid \boldsymbol{I}_n\right] = \frac{1}{n} \sum_{i \in \boldsymbol{I}_n} E\left(\boldsymbol{W}_i^{\otimes 2} \left(\boldsymbol{S}_i^\top \boldsymbol{\beta}_2^*\right)^2 \left[A_{1i} - \mu_{1Ai}\right]^2 \hat{\tilde{R}}_{ni}^2 \mid \boldsymbol{I}_n\right).$$

Let V_{kj} denote the (k, j) element of the matrix on the right-hand side of this last expression. Then, under Assumptions 1 – 5, it can be shown that

$$V_{kj} \leq \frac{C}{n} \sum_{i \in \boldsymbol{I}_n} E\left((\boldsymbol{S}_i^\top \boldsymbol{\beta}_2^*)^2\, \tilde{R}_{ni} \mid \boldsymbol{I}_n\right)$$

for some finite constant $C > 0$. Similarly to the proof of Theorem 1, the fact that

$$|\boldsymbol{S}_i^\top \boldsymbol{\beta}_2^*| \tilde{R}_{ni} \leq |\boldsymbol{S}_i^\top (\hat{\boldsymbol{\beta}}_{2n}^{(-i)} - \boldsymbol{\beta}_2^*)| \tilde{R}_{ni}$$

now implies the inequality

$$V_{kj} \leq \frac{C}{n} \sum_{i \in \boldsymbol{I}_n} E\left(\left[\boldsymbol{S}_i^\top (\hat{\boldsymbol{\beta}}_{2n}^{(-i)} - \boldsymbol{\beta}_2^*)\right]^2\, \tilde{R}_{ni} \mid \boldsymbol{I}_n\right),$$

from which it follows that

$$
\begin{aligned}
V_{kj} &\leq \frac{C}{n}\sum_{i\in \boldsymbol{I}_n} E\left((\hat{\boldsymbol{\beta}}_{2n}^{(-i)}-\boldsymbol{\beta}_2^*)^\top \boldsymbol{S}_i^{\otimes 2}(\hat{\boldsymbol{\beta}}_{2n}^{(-i)}-\boldsymbol{\beta}_2^*)\tilde{R}_{ni} \mid \boldsymbol{I}_n\right), \\
&\leq C\lambda_{1n}^2 E\left(\|\hat{\boldsymbol{\beta}}_{2n}^{(-i)}-\boldsymbol{\beta}_2^*\|_2^2 \mid \boldsymbol{I}_n\right), \qquad (47)
\end{aligned}
$$

where $\lambda_{1n}^2 > 0$ is the maximum eigenvalue of $n^{-1}\sum_{i\in \boldsymbol{I}_n} \boldsymbol{S}_i^{\otimes 2}$. Because the right-hand side of (47) goes to zero as $n \to \infty$ for all (k, j), it follows that (38) is $o_p(1)$. A similar argument establishes that (39) is $o_p(1)$.

□

2. Additional simulation studies

2.1. **Various sample sizes.** In this section, we complement the main simulation study by examining the performance of our proposed method under various sample sizes. We generate 500 datasets with sample sizes N of 1000, 500, and 250 using the same generative model as in Section 5.1 in the paper, and except in Section 2.2, use the same methods to estimate $\mu_{1Y}(\cdot)$, $\mu_{2Y}(\cdot)$, $\mu_{1A}(\cdot)$, and $\mu_{2A}(\cdot)$ as in the main simulation study.

Tables S1-S6 in this document respectively summarize the results for sample sizes N of 1000, 500, and 250 sample sizes. Overall, our method outperforms both $\mathrm{Q}_{N,N}$ and $\mathrm{dWOLS}_{N,N}$ for all sample sizes, particularly when the underlying treatment assignment and the outcome models are nonlinear. More specifically, when the postulated parametric models for the nuisance parameters are correctly specified, the bias of the $\mathrm{dWOLS}_{N,N}$ estimators are comparable to the estimators obtained by our proposed method, but the latter usually has a substantially smaller standard error. However, when the parametric models for the outcome and the treatment assignment models are both misspecified, the $\mathrm{dWOLS}_{N,N}$ estimators exhibit large biases. When $N = 250$, under the linear treatment assignment model and FGS^R outcome model, the proposed method shows larger bias in estimating $\boldsymbol{\beta}_2$ than the $\mathrm{dWOLS}_{N,N}$ (Table S5). We conjecture that this occurs because there are only a relatively small number of units that are rerandomized at stage 2 (i.e., 50%), and that this subsequently affects the performance of Super Learner.

TABLE S1. Performance of the proposed Q-learning method in estimating the second stage parameters under different model complexities (N =1000). The true parameter values for the linear and FGS outcome models are $\beta^*_{2,1} = 1$, $\beta^*_{2,2} = 1$ and $\beta^*_{2,1} \approx 0$, $\beta^*_{2,2} \approx -2$, respectively.

	$\beta^*_{2,1}$						$\beta^*_{2,2}$					
	$Q_{N,N}$		Proposed		$dWOLS_{N,N}$		$Q_{N,N}$		Proposed		$dWOLS_{N,N}$	
Models	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.
Randomized Treatment Assignment Model												
Linear^R	0.007	0.055	0.041	0.112	0.007	0.113	0.003	0.056	0.008	0.112	0.003	0.114
FGS^R	0.002	0.581	0.015	0.787	0.010	1.136	0.055	0.370	0.055	0.424	0.049	0.667
Linear Treatment Assignment Model												
Linear^R	0.000	0.060	0.005	0.145	0.004	0.147	0.011	0.063	0.026	0.144	0.020	0.149
FGS^R	2.235	2.490	0.039	1.010	0.035	2.224	2.478	0.809	0.067	0.529	0.016	0.968
Quadratic Treatment Assignment Model												
Linear^R	0.004	0.054	0.006	0.117	0.003	0.108	0.002	0.059	0.015	0.116	0.005	0.121
FGS^R	0.872	0.630	0.166	0.754	0.873	1.284	0.012	0.332	0.050	0.410	0.021	0.570
InterQuad Treatment Assignment Model												
Linear^R	0.008	0.056	0.009	0.118	0.007	0.113	0.004	0.057	0.024	0.117	0.005	0.116
FGS^R	0.718	0.665	0.077	0.817	0.703	1.253	0.442	0.333	0.068	0.416	0.437	0.560

TABLE S2. Performance of the proposed Q-learning method in estimating the first stage parameters under different model complexities (N =1000). The true parameter values are $\beta^*_{1,1} = \beta^*_{1,2} = 0$.

	$\beta^*_{1,1}$						$\beta^*_{1,2}$					
	$Q_{N,N}$		Proposed		$dWOLS_{N,N}$		$Q_{N,N}$		Proposed		$dWOLS_{N,N}$	
Models	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.
Randomized Treatment Assignment Model												
Linear^R	0.001	0.140	0.004	0.150	0.006	0.150	0.013	0.138	0.008	0.140	0.005	0.149
FGS^R	0.024	0.823	0.034	0.858	0.015	1.072	0.035	0.584	0.027	0.638	0.011	0.866
Linear Treatment Assignment Model												
Linear^R	0.156	0.152	0.003	0.173	0.016	0.177	0.153	0.131	0.011	0.163	0.008	0.172
FGS^R	2.206	3.284	0.192	1.132	0.197	4.588	1.859	2.836	0.148	0.855	0.167	4.377
Quadratic Treatment Assignment Model												
Linear^R	2.395	0.138	0.023	0.171	0.008	0.173	0.679	0.123	0.013	0.162	0.004	0.164
FGS^R	7.259	0.889	0.043	0.963	0.286	1.213	1.656	0.827	0.059	0.735	0.041	1.201
InterQuad Treatment Assignment Model												
Linear^R	2.297	0.143	0.012	0.171	0.012	0.178	0.420	0.128	0.017	0.161	0.010	0.166
FGS^R	7.580	0.829	0.098	0.975	0.279	1.203	2.348	0.705	0.057	0.715	0.275	0.908

TABLE S3. Performance of the proposed Q-learning method in estimating the second stage parameters under different model complexities (N =500). The true parameter values for the linear and FGS outcome models are $\beta^*_{2,1} = 1$, $\beta^*_{2,2} = 1$ and $\beta^*_{2,1} \approx 0$, $\beta^*_{2,2} \approx -2$, respectively.

	$\beta^*_{2,1}$						$\beta^*_{2,2}$					
	$Q_{N,N}$		Proposed		$dWOLS_{N,N}$		$Q_{N,N}$		Proposed		$dWOLS_{N,N}$	
Models	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.
Randomized Treatment Assignment Model												
Linear^R	0.004	0.083	0.013	0.168	0.004	0.169	0.009	0.079	0.023	0.166	0.010	0.161
FGS^R	0.111	0.972	0.113	1.143	0.097	1.906	0.029	0.464	0.015	0.582	0.006	0.815
Linear Treatment Assignment Model												
Linear^R	0.014	0.087	0.021	0.207	0.007	0.214	0.017	0.087	0.006	0.205	0.019	0.211
FGS^R	2.409	0.909	0.117	1.311	0.136	2.250	2.448	0.575	0.185	0.761	0.061	2.111
Quadratic Treatment Assignment Model												
Linear^R	0.003	0.083	0.027	0.171	0.007	0.169	0.004	0.079	0.007	0.169	0.004	0.163
FGS^R	0.738	0.805	0.005	1.101	0.695	1.533	0.008	0.523	0.049	0.658	0.002	0.932
InterQuad Treatment Assignment Model												
Linear^R	0.005	0.086	0.014	0.171	0.010	0.172	0.003	0.083	0.007	0.169	0.005	0.170
FGS^R	0.653	1.085	0.008	1.229	0.627	2.162	0.468	0.551	0.046	0.668	0.436	0.977

TABLE S4. Performance of the proposed Q-learning method in estimating the first stage parameters under different model complexities (N =500). The true parameter values are $\beta^*_{1,1} = \beta^*_{1,2} = 0$.

	$\beta^*_{1,1}$						$\beta^*_{1,2}$					
	$Q_{N,N}$		Proposed		$dWOLS_{N,N}$		$Q_{N,N}$		Proposed		$dWOLS_{N,N}$	
Models	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.
Randomized Treatment Assignment Model												
Linear^R	0.007	0.208	0.017	0.217	0.020	0.205	0.007	0.189	0.008	0.204	0.011	0.206
FGS^R	0.107	1.286	0.061	1.181	0.036	1.556	0.006	0.871	0.024	0.870	0.004	1.078
Linear Treatment Assignment Model												
Linear^R	0.147	0.227	0.015	0.241	0.016	0.238	0.167	0.183	0.025	0.228	0.009	0.225
FGS^R	1.882	1.131	0.038	1.242	0.003	1.475	1.615	0.801	0.016	0.925	0.035	1.117
Quadratic Treatment Assignment Model												
Linear^R	2.392	0.187	0.003	0.243	0.006	0.242	0.676	0.180	0.004	0.230	0.007	0.241
FGS^R	7.441	1.172	0.028	1.292	0.406	1.580	1.457	0.786	0.018	0.962	0.032	1.183
InterQuad Treatment Assignment Model												
Linear^R	2.325	0.199	0.009	0.245	0.001	0.250	0.414	0.170	0.006	0.231	0.009	0.236
FGS^R	7.412	1.249	0.199	1.394	0.176	1.998	2.368	0.938	0.015	1.026	0.136	1.353

TABLE S5. Performance of the proposed Q-learning method in estimating the second stage parameters under different model complexities (N =250). The true parameter values for the linear and FGS outcome models are $\beta^*_{2,1} = 1$, $\beta^*_{2,2} = 1$ and $\beta^*_{2,1} \approx 0$, $\beta^*_{2,2} \approx -2$, respectively.

	$\beta^*_{2,1}$						$\beta^*_{2,2}$					
	$Q_{N,N}$		Proposed		$dWOLS_{N,N}$		$Q_{N,N}$		Proposed		$dWOLS_{N,N}$	
Models	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.
Randomized Treatment Assignment Model												
Linear^R	0.035	0.120	0.038	0.261	0.031	0.250	0.008	0.116	0.012	0.252	0.008	0.244
FGS^R	0.032	1.138	0.053	1.541	0.070	2.229	0.058	0.694	0.089	0.959	0.045	1.340
Linear Treatment Assignment Model												
Linear^R	0.008	0.132	0.027	0.302	0.007	0.336	0.009	0.127	0.004	0.291	0.007	0.315
FGS^R	2.315	1.237	0.488	1.755	0.286	2.616	2.519	0.717	0.682	1.111	0.274	1.545
Quadratic Treatment Assignment Model												
Linear^R	0.014	0.120	0.023	0.265	0.020	0.257	0.034	0.127	0.032	0.257	0.028	0.268
FGS^R	0.738	2.374	0.061	1.937	0.522	5.242	0.051	0.736	0.004	1.167	0.078	1.400
InterQuad Treatment Assignment Model												
Linear^R	0.010	0.127	0.042	0.256	0.007	0.262	0.009	0.120	0.018	0.250	0.012	0.247
FGS^R	0.794	1.066	0.265	1.655	0.804	2.139	0.428	0.691	0.000	1.093	0.457	1.244

TABLE S6. Performance of the proposed Q-learning method in estimating the first stage parameters under different model complexities (N =250). The true parameter values are $\beta^*_{1,1} = \beta^*_{1,2} = 0$.

	$\beta^*_{1,1}$						$\beta^*_{1,2}$					
	$Q_{N,N}$		Proposed		$dWOLS_{N,N}$		$Q_{N,N}$		Proposed		$dWOLS_{N,N}$	
Models	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.
Randomized Treatment Assignment Model												
Linear^R	0.020	0.284	0.005	0.308	0.008	0.303	0.031	0.244	0.008	0.291	0.014	0.299
FGS^R	0.166	1.389	0.019	1.661	0.092	2.477	0.159	1.200	0.019	1.274	0.041	1.698
Linear Treatment Assignment Model												
Linear^R	0.194	0.284	0.050	0.338	0.012	0.330	0.174	0.262	0.016	0.320	0.002	0.349
FGS^R	1.960	1.575	0.153	1.704	0.089	1.919	1.641	1.261	0.055	1.286	0.126	1.540
Quadratic Treatment Assignment Model												
Linear^R	2.413	0.289	0.016	0.351	0.001	0.339	0.664	0.257	0.016	0.336	0.026	0.367
FGS^R	7.840	2.066	0.106	1.855	0.398	2.460	1.373	1.273	0.093	1.457	0.096	2.040
InterQuad Treatment Assignment Model												
Linear^R	2.329	0.294	0.052	0.349	0.007	0.358	0.451	0.251	0.006	0.334	0.024	0.369
FGS^R	7.423	1.469	0.212	1.837	0.107	2.335	2.288	1.188	0.079	1.461	0.238	1.721

2.2. **Performance under alternative nonparametric estimation methods.** In this section, we assess the performance of the proposed method when the nuisance parameters are instead estimated using either random forests or generalized additive models, recalling that both were included as part of the library used by Super Learner. As in the main simulation study, we generated 500 datasets of size N =2000 using the same generative model as in Section 2.1. In Table S7 and S8, columns RF-RF and GAM-GAM respectively represent cases where a random forest and a generalized additive model are used for both the treatment assignment (i.e., $\mu_{1A}(\cdot)$ and $\mu_{2A}(\cdot)$) and the outcome models (i.e., $\mu_{1Y}(\cdot)$ and $\mu_{2Y}(\cdot)$). The column RF-GAM represents a case where a random forest is used for the outcome model and a generalized additive model is used for the treatment assignment model. Overall, as the complexity of the outcome or treatment assignment model increases, modeling the nuisance parameters using a random forest results in estimators that are less biased compared with those obtained by the generalized additive model. However, comparing these results with those presented in Tables 1 and 2, we observe that Super Learner typically performs better than these other data adaptive methods.

TABLE S7. Performance of the proposed Q-learning method in estimating the second stage parameters using machine learning methods under different model complexities (N =2000). The true parameter values for the linear and FGS outcome models are $\beta^*_{2,1} = 1$, $\beta^*_{2,2} = 1$ and $\beta^*_{2,1} \approx 0$, $\beta^*_{2,2} \approx -2$, respectively.

	$\beta^*_{2,1}$						$\beta^*_{2,2}$					
	RF-RF		GAM-GAM		RF-GAM		RF-RF		GAM-GAM		RF-GAM	
Models	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.
Randomized Treatment Assignment Model												
Linear^R	0.021	0.088	0.051	0.078	0.001	0.086	0.021	0.084	0.048	0.078	0.021	0.084
FGS^R	0.096	0.521	0.236	0.684	0.049	0.499	0.047	0.282	0.091	0.436	0.091	0.275
Linear Treatment Assignment Model												
Linear^R	0.005	0.107	0.009	0.103	0.006	0.110	0.002	0.103	0.012	0.103	0.011	0.107
FGS^R	0.447	0.687	0.023	0.823	0.037	0.712	0.522	0.405	0.074	0.553	0.064	0.419
Quadratic Treatment Assignment Model												
Linear^R	0.013	0.091	0.033	0.080	0.065	0.085	0.014	0.088	0.042	0.080	0.002	0.083
FGS^R	0.194	0.628	0.764	0.737	0.301	0.616	0.039	0.337	0.091	0.501	0.022	0.352
InterQuad Treatment Assignment Model												
Linear^R	0.035	0.091	0.076	0.081	0.044	0.085	0.038	0.087	0.076	0.081	0.010	0.083
FGS^R	0.192	0.651	0.654	0.761	0.219	0.648	0.071	0.329	0.304	0.499	0.065	0.344

TABLE S8. Performance of the proposed Q-learning method in estimating the first stage parameters using machine learning methods under different model complexities (N =2000). The true parameter values are $\beta^*_{1,1} = \beta^*_{1,2} = 0$.

	$\beta^*_{1,1}$						$\beta^*_{1,2}$					
	RF-RF		GAM-GAM		RF-GAM		RF-RF		GAM-GAM		RF-GAM	
Models	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.	Bias	S.D.
Randomized Treatment Assignment Model												
Linear^R	0.016	0.104	0.017	0.108	0.003	0.107	0.029	0.099	0.004	0.100	0.003	0.102
FGS^R	0.029	0.537	0.200	0.725	0.014	0.612	0.020	0.413	0.063	0.514	0.012	0.444
Linear Treatment Assignment Model												
Linear^R	0.008	0.116	0.000	0.127	0.003	0.122	0.008	0.110	0.005	0.119	0.008	0.116
FGS^R	0.020	0.677	0.039	0.775	0.081	0.700	0.086	0.491	0.021	0.575	0.009	0.508
Quadratic Treatment Assignment Model												
Linear^R	0.033	0.118	0.002	0.121	0.027	0.122	0.008	0.113	0.002	0.113	0.006	0.116
FGS^R	0.035	0.691	0.296	0.794	0.101	0.707	0.007	0.508	0.035	0.600	0.037	0.519
InterQuad Treatment Assignment Model												
Linear^R	0.020	0.119	0.015	0.122	0.045	0.122	0.009	0.114	0.003	0.114	0.031	0.116
FGS^R	0.083	0.681	0.344	0.786	0.145	0.704	0.031	0.509	0.164	0.610	0.032	0.527

2.3. **Additional simulation results for the non-regular setting.** In Section 5.2 of the main paper, we showed that the proposed method can provide valid inferences in non-regular settings when the assumptions of Corollary 1 are satisfied. Table S9 supplements the results of Table 3 for the main simulation study by providing additional comparisons with $\mathrm{Q}_{N,N}$ and $\mathrm{dWOLS}_{N,N}$.

Below, we also consider a simulation that corresponds to a non-regular setting when the assumptions of Corollary 1 and Assumption 6 are both violated. We consider the same general set-up as in Section 5.2, but modify the outcome models as described below:

- Linear$^{NR,\varpi}$: $Y = \mathbf{X}_1^\top \alpha_1 + \breve{\mathbf{X}}_2^\top \alpha_2 + A_1 \mathbf{X}_1^\top \boldsymbol{\theta}_1 + A_2 R(\theta_{21} A_1 + \theta_{22} \breve{X}_{21}) + \epsilon$ where $\alpha_1 = (1, 0.1, 0.1, 0.1, 0.1)^\top$, $\alpha_2 = (1, 0.1, 0.1, 0.1)^\top$, $\boldsymbol{\theta}_1 = (0, 0, 0, 0, 0)^\top$, and $\boldsymbol{\theta}_2 = 2(\varpi, \varpi)^\top$;
- Non-linear$^{NR,\varpi}$: $Y = f(\mathbf{X}_1) + A_1 \mathbf{X}_1^\top \boldsymbol{\theta}_1 + A_2 R(\theta_{21} A_1 + \theta_{22} \breve{X}_{21}) + \epsilon$ where $\boldsymbol{\theta}_1 = (0, 0, 0, 0, 0)^\top$, $\boldsymbol{\theta}_2 = 2(\varpi, \varpi)^\top$, and for $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5)^\top$, we set

$$f(\boldsymbol{x}) = \sin(\pi x_1 x_2) + 2(x_3 - 0.5)^2 + x_4 + 1.5 \frac{x_1}{|x_2| + |x_3|} + 2x_1(x_2 + x_3).$$

Here, $\breve{\boldsymbol{X}}_2 = (\breve{X}_{21}, \breve{X}_{22}, X_{23}, X_{24})^\top$ where $\breve{X}_{21}$ is generated from a Bernoulli distribution with success probability $[1 + \exp\{-(2X_{11} - 2A_1 - 1)\}]^{-1}$, $X_{22} = U_{11}$, $X_{23} = U_{12}$, $X_{24} = 0.35X_{15} + U_{13}$, and $X_{25} = U_{14}$, where $U_{1l}, l = 1, \ldots, 3$ are independent and uniformly distributed on [0,1]; the noise variable ϵ is again generated from $\mathrm{N}(0, \sigma = 0.5)$.

TABLE S9. Performance of the proposed Q-learning method under different degrees of non-regularity.

Models	$\beta^*_{1,1}$			$\beta^*_{1,2}$		
	$Q_{N,N}$	Proposed	$dWOLS_{N,N}$	$Q_{N,N}$	Proposed	$dWOLS_{N,N}$
Randomized Treatment Assignment Model						
$\text{Linear}^{NR,0}$	0.939(0.26)	0.956(0.42)	0.940(0.41)	0.932(0.26)†	0.959(0.40)	0.933(0.41)†
$\text{Non-linear}^{NR,0}$	0.949(0.95)	0.965(1.40)†	0.933(1.40)†	0.949(0.44)	0.952(0.56)	0.946(0.55)
$\text{Linear}^{NR,1}$	0.937(0.45)	0.962(0.46)	0.938(0.70)	0.936(0.45)	0.966(0.45)†	0.933(0.70)†
$\text{Non-linear}^{NR,1}$	0.953(0.99)	0.966(1.41)	0.924(1.46)†	0.936(0.51)†	0.946(0.60)	0.929(0.69)†
Linear Treatment Assignment Model						
$\text{Linear}^{NR,0}$	0.913(0.26)†	0.965(0.45)†	0.940(0.45)	0.898(0.26)†	0.948(0.44)	0.940(0.45)
$\text{Non-linear}^{NR,0}$	0.396(0.96)†	0.949(1.58)	0.921(1.56)†	0.102(0.48)†	0.955(0.65)	0.936(0.65)
$\text{Linear}^{NR,1}$	0.856(0.45)†	0.953(0.51)	0.939(0.76)	0.845(0.45)†	0.955(0.50)	0.924(0.76)†
$\text{Non-linear}^{NR,1}$	0.343(1.10)†	0.952(1.61)	0.923(1.63)†	0.101(0.55)†	0.948(0.70)	0.942(0.79)
InterQuad Treatment Assignment Model						
$\text{Linear}^{NR,0}$	0.949(0.29)	0.966(0.46)†	0.956(0.45)	0.939(0.29)	0.957(0.45)	0.934(0.45)†
$\text{Non-linear}^{NR,0}$	0.858(1.10)†	0.959(1.45)	0.786(1.45)†	0.897(0.50)†	0.962(0.63)	0.789(0.62)†
$\text{Linear}^{NR,1}$	0.951(0.49)	0.964(0.52)	0.948(0.77)	0.937(0.50)†	0.950(0.51)	0.939(0.77)
$\text{Non-linear}^{NR,1}$	0.861(1.05)†	0.964(1.47)	0.804(1.53)†	0.910(0.58)†	0.967(0.68)	0.850(0.78)†

Numbers in parentheses correspond to average confidence interval length.

In the above generative models, $\boldsymbol{S}^0 = (\boldsymbol{X}_1^\top, A_1, \breve{\boldsymbol{X}}_2^\top)^\top$ and $\boldsymbol{W}^0 = \boldsymbol{X}_1$; the vectors $\boldsymbol{S} = R(1, A_1, \breve{X}_{21}, \breve{X}_{22})^\top$ and $\boldsymbol{W} = (1, X_{11}, X_{12})^\top$ are respectively used to model the second and first stage Q-functions. The second stage blip functions are linear and correctly specified in both models, leading to $\boldsymbol{\beta}_2^* = (0, \boldsymbol{\theta}_2^\top, 0)^\top$. In both scenarios, for each subject i, we define the first-stage pseudo outcome as

$$\tilde{Y}_i = Y_i + \varpi R \boldsymbol{S}_i^\top \boldsymbol{\beta}_2^* \{I(\boldsymbol{S}_i^\top \boldsymbol{\beta}_2^* > 0) - A_{2i}\},$$

its estimate $\hat{\tilde{Y}}_i$ being calculated by substituting $\hat{\boldsymbol{\beta}}_{2n}$ in for $\boldsymbol{\beta}_2^*$. The construction of the pseudo-outcome, specifically the projection $\boldsymbol{S}^\top \boldsymbol{\beta}_2^*$, results in the violation of Assumption 6 for both generative models. In particular, $\varpi = 0$ corresponds to no second-stage effect modifier because $\boldsymbol{\beta}_2^* = \mathbf{0}$; hence, we have $P(|\boldsymbol{S}^\top \boldsymbol{\beta}_2^*| = 0) = 1$. Setting $\varpi = 1$ instead implies that there is no second-stage treatment effect when $R(A_1 + \breve{X}_{21}) = 0$ and a reasonably strong effect when $R(A_1 + \breve{X}_{21}) = 1$; in this case, we have $\|\boldsymbol{\beta}_2^*\| > 0$ and $0 < P(|\boldsymbol{S}^\top \boldsymbol{\beta}_2^*| = 0) < 1$. In addition to violating Assumption 6, both generative models also violate the key assumption of Corollary 1 (i.e., $E\left(A_{1i} - \mu_{1Ai} \mid \boldsymbol{W}_i, \boldsymbol{S}_i, \boldsymbol{I}_n\right) = 0, i \in \boldsymbol{I}_n$) because the Stage 2 decision rule depends on A_1.

TABLE S10. Performance of proposed Q-learning method under different degrees of non-regularity when Corollary 1 and Assumption 6 both fail.

	$\beta^*_{1,1}$		$\beta^*_{1,2}$		$\beta^*_{1,3}$	
Models	Proposed	dWOLS$^{\kappa=0.05}_{m,N}$	Proposed	dWOLS$^{\kappa=0.05}_{m,N}$	Proposed	dWOLS$^{\kappa=0.05}_{m,N}$
Randomized Treatment Assignment Model						
Linear^{NR,0}	0.938(0.12)	0.981(0.14)†	0.949(0.40)	0.968(0.46)†	0.965(0.40)†	0.984(0.46)†
Non-linear^{NR,0}	0.940(0.29)	0.967(0.43)†	0.965(1.29)†	0.977(1.69)†	0.949(0.39)	0.973(0.68)†
Linear^{NR,1}	0.951(0.22)	0.976(0.26)†	0.948(0.77)	0.973(0.89)†	0.950(0.77)	0.983(0.89)†
Non-linear^{NR,1}	0.928(0.34)†	0.957(0.48)	0.967(1.43)†	0.980(1.82)†	0.941(0.69)	0.980(0.93)†
Linear Treatment Assignment Model						
Linear^{NR,0}	0.908(0.12)†	0.980(0.17)†	0.954(0.43)	0.968(0.50)†	0.939(0.43)	0.976(0.50)†
Non-linear^{NR,0}	0.909(0.29)†	0.984(0.47)†	0.975(1.32)†	0.972(1.77)†	0.942(0.46)	0.981(0.78)†
Linear^{NR,1}	0.951(0.24)	0.972(0.28)†	0.943(0.85)	0.970(0.97)†	0.949(0.86)	0.969(0.98)
Non-linear^{NR,1}	0.955(0.35)	0.962(0.47)	0.971(1.50)†	0.976(1.93)†	0.942(0.80)	0.978(1.06)†
InterQuad Treatment Assignment Model						
Linear^{NR,0}	0.934(0.12)†	0.983(0.15)†	0.952(0.41)	0.975(0.48)†	0.948(0.41)	0.975(0.48)†
Non-linear^{NR,0}	0.939(0.30)	0.914(0.45)†	0.944(1.29)	0.884(1.75)†	0.958(0.45)	0.885(0.76)†
Linear^{NR,1}	0.950(0.26)	0.972(0.31)†	0.962(0.89)	0.983(1.02)†	0.941(0.90)	0.963(1.03)
Non-linear^{NR,1}	0.907(0.38)†	0.971(0.55)†	0.968(1.51)†	0.923(1.95)†	0.962(0.87)	0.945(1.12)

Numbers in parentheses correspond to average confidence interval length.

We compare the coverages of our estimators with those obtained for dWOLS using the m-out-of-n bootstrap (i.e., dWOLS$^{\kappa=0.05}_{m,N}$). Table S10 presents the results based on 1000 datasets of size 2000. In the table, $\beta^*_{1,1}$, $\beta^*_{1,2}$ and $\beta^*_{1,3}$ respectively the true parameter values corresponding to variables A_1 (i.e., we assume $\boldsymbol{W}$ has an intercept), A_1X_{11} and A_1X_{12}. When $\varpi = 1$, $(\beta^*_{1,1}, \beta^*_{1,2}, \beta^*_{1,3}) \approx (0.55, 0.00, -0.57)$ and $(\beta^*_{1,1}, \beta^*_{1,2}, \beta^*_{1,3}) \approx (0.77, 0.00, -0.29)$ for the linear and non-linear outcome models, respectively (note: these are approximated by simulation). Also, when $\varpi = 0$, $(\beta^*_{1,1}, \beta^*_{1,2}, \beta^*_{1,3}) \approx (-0.23, 0.00, -0.29)$ and $(\beta^*_{1,1}, \beta^*_{1,2}, \beta^*_{1,3}) = (0, 0, 0)$ for the linear and non-linear outcome models, respectively. Despite violation of the regularity assumption, the proposed method continues to provide valid confidence intervals for the parameters $\beta_{1,2}$, and $\beta_{1,3}$. However, the confidence intervals for $\beta_{1,1}$ exhibit less-than-nominal coverage in a majority of cases. As expected, when either of the outcome or the treatment assignment models are correctly specified the dWOLS$^{\kappa=0.05}_{m,N}$ provides valid, if typically conservative, confidence intervals for all parameters. However, when both of these models are misspecified, the coverage rates are more substantially compromised.

2.4. **Value Function Estimates for Regular Case.** In this section we plot the value functions corresponding to the estimated decision rules obtained in the regular setting (see Section 5.1 of the main paper). All methods perform essentially identically as far as the

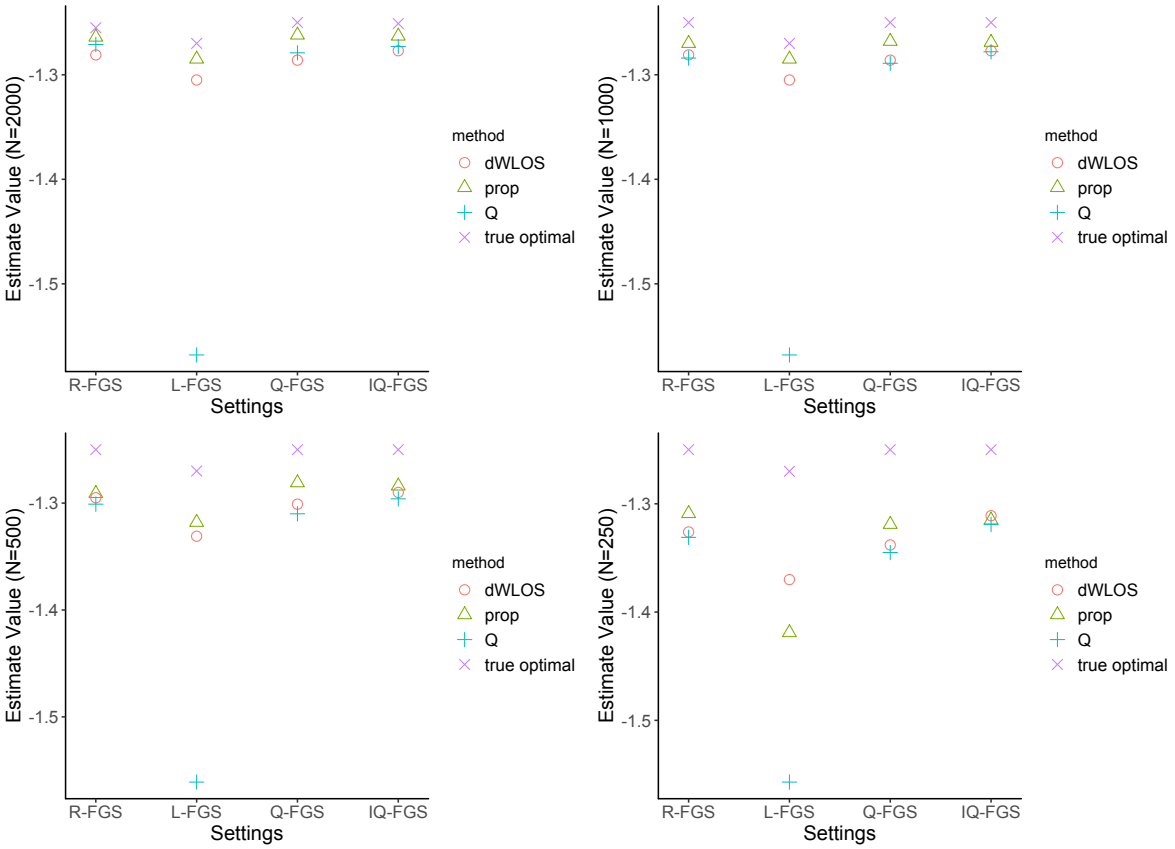

FIGURE S1. Value functions for estimated decision rules for regular case and different sample sizes.

underlying value function when the outcome model is linear, even for $N = 250$; these results are not included. In the case where the outcome model is not linear, we see that the proposed method typically results in a value function that is closest to the true optimal value, followed by dWOLS and then standard Q-learning. However, the overall degree of discrepancy is generally small, particularly between dWOLS and the proposed approach.

REFERENCES

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. & ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* **21**, c1 – c68.

STEWART, G. W. (1969). On the continuity of the generalized inverse. *SIAM Journal on Applied Mathematics* **17**, 33–45.