# Online Supplement to "Covariate Adjustment in Experiments with Matched Pairs"*

Yuehao Bai[a], Liang Jiang[†b], Joseph P. Romano[c], Azeem M. Shaikh[d], and Yichong Zhang[e]

[a]Department of Economics, University of Southern California, USA. Email: yuehao.bai@usc.edu

[b]Fanhai International School of Finance, Fudan University, China. Email: jiangliang@fudan.edu.cn

[c]Departments of Economics & Statistics, Stanford University, USA. Email: romano@stanford.edu

[d]Department of Economics, University of Chicago, USA. Email: amshaikh@uchicago.edu

[e]School of Economics, Singapore Management University, Singapore. Email: yczhang@smu.edu.sg

February 19, 2024

**Abstract**

This document collects supplementary materials for the main paper. Section A contains the proof of the main results. Section B gives auxiliary lemmas and their proof. Section C provides the details of the LASSO regressors in the simulations. Section D provides the details of the regressors for regression adjustments used in the empirical application.

KEYWORDS: Experiment, matched pairs, covariate adjustment, randomized controlled trial, treatment assignment, LASSO

JEL classification codes: C12, C14

*We thank the Editor, Elie Tamer, the Associate Editor and two anonymous referees for many useful comments that helped to improve this paper. Yichong Zhang acknowledges the financial support from the NSFC under the grant No. 72133002 and a Lee Kong Chian fellowship. Any and all errors are our own.

†Corresponding author.

# A Proofs of Main Results

In the appendix, we use $a_n \lesssim b_n$ to denote there exists $c > 0$ such that $a_n \le c b_n$.

## A.1 Proof of Theorem 3.1

Step 1: Decomposition by recursive conditioning

To begin, note

$$\begin{aligned}
\hat{\mu}_n(1) &= \frac{1}{2n} \sum_{1 \le i \le 2n} (2D_i(Y_i(1) - \hat{m}_{1,n}(X_i, W_i)) + \hat{m}_{1,n}(X_i, W_i)) \\
&= \frac{1}{2n} \sum_{1 \le i \le 2n} (2D_i Y_i(1) - (2D_i - 1)\hat{m}_{1,n}(X_i, W_i)) \\
&= \frac{1}{2n} \sum_{1 \le i \le 2n} (2D_i Y_i(1) - (2D_i - 1)m_{1,n}(X_i, W_i)) + o_P(n^{-1/2}) \\
&= \frac{1}{2n} \sum_{1 \le i \le 2n} (2D_i Y_i(1) - D_i m_{1,n}(X_i, W_i) - (1 - D_i)m_{1,n}(X_i, W_i)) + o_P(n^{-1/2}) \ , \qquad \text{(A.1)}
\end{aligned}$$

where the third equality follows from (9). Similarly,

$$\hat{\mu}_n(0) = \frac{1}{2n} \sum_{1 \le i \le 2n} (2(1 - D_i)Y_i(0) - D_i m_{0,n}(X_i, W_i) - (1 - D_i)m_{0,n}(X_i, W_i)) + o_P(n^{-1/2}) \ . \qquad \text{(A.2)}$$

It follows from (A.1)–(A.2) that

$$\hat{\Delta}_n = \frac{1}{n} \sum_{1 \le i \le 2n} D_i \phi_{1,n,i} - \frac{1}{n} \sum_{1 \le i \le 2n} (1 - D_i)\phi_{0,n,i} + o_P(n^{-1/2}) \ , \qquad \text{(A.3)}$$

where

$$\begin{aligned}
\phi_{1,n,i} &= Y_i(1) - \frac{1}{2}(m_{1,n}(X_i, W_i) + m_{0,n}(X_i, W_i)) \\
\phi_{0,n,i} &= Y_i(0) - \frac{1}{2}(m_{1,n}(X_i, W_i) + m_{0,n}(X_i, W_i)) \ .
\end{aligned}$$

Next, consider

$$\mathbb{L}_n = \frac{1}{2\sqrt{n}} \sum_{1 \le i \le 2n} (2D_i - 1)E[m_{1,n}(X_i, W_i) + m_{0,n}(X_i, W_i)|X_i] \ .$$

For simplicity, define $M_{d,n}(X_i) = E[m_{d,n}(X_i, W_i)|X_i]$ for $d \in \{0, 1\}$. It follows from Assumption 2.2 that $E[\mathbb{L}_n|X^{(n)}] = 0$. On the other hand,

$$\begin{aligned}
\mathrm{Var}[\mathbb{L}_n|X^{(n)}] &= \frac{1}{4n} \sum_{1 \le j \le n} \big(M_{1,n}(X_{\pi(2j-1)}) + M_{0,n}(X_{\pi(2j-1)}) - (M_{1,n}(X_{\pi(2j)}) + M_{0,n}(X_{\pi(2j)}))\big)^2 \\
&\lesssim \frac{1}{n} \sum_{1 \le j \le n} |M_{1,n}(X_{\pi(2j-1)}) - M_{1,n}(X_{\pi(2j)})|^2 + \frac{1}{n} \sum_{1 \le j \le n} |M_{0,n}(X_{\pi(2j-1)}) - M_{0,n}(X_{\pi(2j)})|^2
\end{aligned}$$

$$\xrightarrow{P} 0 \ ,$$

where the inequality follows from $(a+b)^2 \le 2(a^2+b^2)$ and the convergence follows from Assumptions 2.3 and 3.1(c). By Markov's inequality and the fact that $E[\mathbb{L}_n | X^{(n)}] = 0$, for any $\epsilon > 0$,

$$P\{|\mathbb{L}_n| > \epsilon | X^{(n)}\} \le \frac{\text{Var}[\mathbb{L}_n | X^{(n)}]}{\epsilon^2} \xrightarrow{P} 0 \ .$$

Since probabilities are bounded, we have $\mathbb{L}_n = o_P(1)$. This fact, together with (A.3), imply

$$\sqrt{n}(\hat{\Delta}_n - \Delta(Q)) = A_n - B_n + C_n - D_n \ ,$$

where

$$\begin{aligned}
A_n &= \frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} \Big( D_i \phi_{1,n,i} - E[D_i \phi_{1,n,i} | X^{(n)}, D^{(n)}] \Big) \\
B_n &= \frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} \Big( (1-D_i) \phi_{0,n,i} - E[(1-D_i) \phi_{0,n,i} | X^{(n)}, D^{(n)}] \Big) \\
C_n &= \frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} D_i (E[Y_i(1)|X_i] - E[Y_i(1)]) \\
D_n &= \frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} (1-D_i)(E[Y_i(0)|X_i] - E[Y_i(0)]) \ .
\end{aligned}$$

Note that conditional on $X^{(n)}$ and $D^{(n)}$, $A_n$ and $B_n$ are independent while $C_n$ and $D_n$ are constants.

Step 2: Conditional central limit theorems

We first analyze the limiting behavior of $A_n$. Define

$$s_n^2 = \sum_{1 \le i \le 2n} D_i \, \text{Var}[\phi_{1,n,i} | X_i] \ .$$

Note by Assumption 2.2 that $s_n^2 = n \, \text{Var}[A_n | X^{(n)}, D^{(n)}]$. We proceed verify the Lindeberg condition for $A_n$ conditional on $X^{(n)}$ and $D^{(n)}$, i.e., we show that for every $\epsilon > 0$,

$$\frac{1}{s_n^2} \sum_{1 \le i \le 2n} E[|D_i(\phi_{1,n,i} - E[\phi_{1,n,i}|X_i])|^2 I\{|D_i(\phi_{1,n,i} - E[\phi_{1,n,i}|X_i])| > \epsilon s_n\} | X^{(n)}, D^{(n)}] \xrightarrow{P} 0 \ . \tag{A.4}$$

To that end, first note Lemma B.2 implies

$$\frac{s_n^2}{nE[\text{Var}[\phi_{1,n,i}|X_i]]} \xrightarrow{P} 1 \ . \tag{A.5}$$

(A.5) and Assumption 3.1(a) imply that for all $\lambda > 0$,

$$P\{\epsilon s_n > \lambda\} \xrightarrow{P} 1 \ . \tag{A.6}$$

Furthermore, for some $c > 0$,

$$P\left\{\frac{s_n^2}{n} > c\right\} \to 1 \ . \tag{A.7}$$

Next, note for any $\lambda > 0$ and $\delta_1 > 0$, the left-hand side of (A.4) can be written as

$$\begin{aligned}
&\frac{1}{s_n^2/n}\frac{1}{n}\sum_{1\le i\le 2n: D_i=1} E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]| > \epsilon s_n\}|X^{(n)}, D^{(n)}] \\
&\le \frac{1}{s_n^2/n}\frac{1}{n}\sum_{1\le i\le 2n} E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]| > \epsilon s_n\}|X^{(n)}, D^{(n)}] \\
&\le \frac{1}{c}\frac{1}{n}\sum_{1\le i\le 2n} E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]| > \lambda\}|X^{(n)}, D^{(n)}] + o_P(1) \\
&\le \frac{2}{c}\frac{1}{2n}\sum_{1\le i\le 2n} E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]| > \lambda\}|X_i] + o_P(1) \ ,
\end{aligned} \tag{A.8}$$

where the first inequality follows by inspection, the second follows from (A.6)–(A.7), and the last follows from Assumption 2.2. We then argue

$$\begin{aligned}
&\frac{1}{2n}\sum_{1\le i\le 2n} E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]| > \lambda\}|X_i] \\
&\qquad = E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]| > \lambda\}] + o_P(1) \ .
\end{aligned} \tag{A.9}$$

To this end, we once again verify the Lindeberg condition in Lemma 11.4.2 of Lehmann and Romano (2005). Note

$$|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]| > \lambda\} \le |\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 \ .$$

Therefore, in light of Lemma B.1, we only need to verify

$$\lim_{\gamma\to\infty}\limsup_{n\to\infty} E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 > \gamma\}] = 0 \ , \tag{A.10}$$

which follows immediately from Lemma B.3.

Another application of (A.10) implies (A.4). Lindeberg's central limit theorem and (A.5) then imply that

$$\sup_{t\in\mathbf{R}} |P\{A_n/\sqrt{E[\mathrm{Var}[\phi_{1,n,i}|X_i]]} \le t|X^{(n)}, D^{(n)}\} - \Phi(t)| \xrightarrow{P} 0 \ .$$

Similar arguments lead to

$$\sup_{t\in\mathbf{R}} |P\{B_n/\sqrt{E[\mathrm{Var}[\phi_{0,n,i}|X_i]]} \le t|X^{(n)}, D^{(n)}\} - \Phi(t)| \xrightarrow{P} 0 \ .$$

Step 3: Combining conditional and unconditional components

Meanwhile, it follows from the same arguments as those in (S.22)–(S.25) of Bai et al. (2022) that

$$C_n - D_n \xrightarrow{d} N\left(0, \frac{1}{2}E\left[(E[Y_i(1)|X_i] - E[Y_i(1)] - (E[Y_i(0)|X_i] - E[Y_i(0)]))^2\right]\right) \ .$$

To establish (10), define $\nu_n^2 = \nu_{1,n}^2 + \nu_{0,n}^2 + \nu_2^2$, where

$$\begin{aligned}
\nu_{1,n}^2 &= E[\mathrm{Var}[\phi_{1,n,i}|X_i]] \\
\nu_{0,n}^2 &= E[\mathrm{Var}[\phi_{0,n,i}|X_i]] \\
\nu^2 &= \frac{1}{2}E\left[(E[Y_i(1)|X_i] - E[Y_i(1)] - (E[Y_i(0)|X_i] - E[Y_i(0)]))^2\right]
\end{aligned}$$

Note

$$\frac{\sqrt{n}(\hat{\Delta}_n - \Delta(Q))}{\nu_n} = \frac{A_n}{\nu_{1,n}}\frac{\nu_{1,n}}{\nu_n} - \frac{B_n}{\nu_{0,n}}\frac{\nu_{0,n}}{\nu_n} + \frac{C_n - D_n}{\nu_2}\frac{\nu_2}{\nu_n} .$$

Further note $\nu_n, \nu_{1,n}, \nu_{0,n}, \nu_2$ are all constants conditional on $X^{(n)}$ and $D^{(n)}$. Suppose by contradiction that $\frac{\sqrt{n}(\hat{\Delta}_n - \Delta(Q))}{\nu_n}$ does not converge in distribution to $N(0,1)$. Then, there exists $\epsilon > 0$ and a subsequence $\{n_k\}$ such that

$$\sup_{t \in \mathbf{R}} |P\{\sqrt{n}_k(\hat{\Delta}_{n_k} - \Delta(Q))/\nu_{n_k} \leq t\} - \Phi(t)| \to \epsilon . \tag{A.11}$$

Because the sequence $\nu_{1,n_k}$ and $\nu_{0,n_k}$ are bounded by Assumptions 3.1(b), there is a further subsequence, which with some abuse of notation we still denote by $\{n_k\}$, along which $\nu_{1,n_k} \to \nu_1^*$ and $\nu_{0,n_k} \to \nu_0^*$ for some $\nu_1^*, \nu_0^* \geq 0$. Then, $\nu_{1,n_k}/\nu_{n_k}, \nu_{0,n_k}/\nu_{n_k}, \nu_2/\nu_{n_k}$ all converge to constants. Therefore, it follows from Lemma S.1.2 of Bai et al. (2022) that

$$\sqrt{n}_k(\hat{\Delta}_{n_k} - \Delta(Q))/\nu_{n_k} \xrightarrow{d} N(0,1) ,$$

a contradiction to (A.11). Therefore, the desired convergence in Theorem 3.1 follows.

Step 4: Rearranging the variance formula

To conclude the proof with the the variance formula as stated in the theorem, note

$$\begin{aligned}
&\mathrm{Var}\left[Y_i(0) - \frac{1}{2}(m_{1,n}(X_i,W_i) + m_{0,n}(X_i,W_i))\Big|X_i\right] \\
&= \mathrm{Var}\left[E\Big[Y_i(0) - \frac{1}{2}(m_{1,n}(X_i,W_i) + m_{0,n}(X_i,W_i))\Big|X_i,W_i\Big]\Big|X_i\right] \\
&\quad + E\Big[\mathrm{Var}\left[Y_i(0) - \frac{1}{2}(m_{1,n}(X_i,W_i) + m_{0,n}(X_i,W_i))\Big|X_i,W_i\right]\Big|X_i\Big] \\
&= \mathrm{Var}\left[E\Big[\frac{Y_i(1)+Y_i(0)}{2}\Big|X_i,W_i\Big] - \frac{1}{2}(m_{1,n}(X_i,W_i) + m_{0,n}(X_i,W_i)) - E\Big[\frac{Y_i(1)-Y_i(0)}{2}\Big|X_i,W_i\Big]\Big|X_i\right] \\
&\quad + E[\mathrm{Var}[Y_i(0)|X_i,W_i]|X_i] \\
&= \mathrm{Var}\left[E\Big[\frac{Y_i(1)+Y_i(0)}{2}\Big|X_i,W_i\Big] - \frac{1}{2}(m_{1,n}(X_i,W_i) + m_{0,n}(X_i,W_i))\Big|X_i\right] \\
&\quad + \mathrm{Var}\left[E\Big[\frac{Y_i(1)-Y_i(0)}{2}\Big|X_i,W_i\Big]\Big|X_i\right] \\
&\quad - 2\mathrm{Cov}\Big[E\Big[\frac{Y_i(1)+Y_i(0)}{2}\Big|X_i,W_i\Big] - \frac{1}{2}(m_{1,n}(X_i,W_i) + m_{0,n}(X_i,W_i)), E\Big[\frac{Y_i(1)-Y_i(0)}{2}\Big|X_i,W_i\Big]\Big|X_i\Big] \\
&\quad + E[\mathrm{Var}[Y_i(0)|X_i,W_i]|X_i] ,
\end{aligned} \tag{A.12}$$

where the first equality follows from the law of total variance, the second one follows by direct calculation, and the last one follows by expanding the variance of the sum. Similarly,

$$\mathrm{Var}\left[Y_i(1) - \frac{1}{2}(m_{1,n}(X_i,W_i) + m_{0,n}(X_i,W_i))\Big|X_i\right]$$

$$
\begin{aligned}
&= \text{Var}\left[E\Big[\frac{Y_i(1)+Y_i(0)}{2}\Big|X_i,W_i\Big] - \frac{1}{2}(m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i))\Big|X_i\right] \\
&\quad + \text{Var}\left[E\Big[\frac{Y_i(1)-Y_i(0)}{2}\Big|X_i,W_i\Big]\Big|X_i\right] \\
&\quad + 2\text{Cov}\Big[E\Big[\frac{Y_i(1)+Y_i(0)}{2}\Big|X_i,W_i\Big] - \frac{1}{2}(m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i)), E\Big[\frac{Y_i(1)-Y_i(0)}{2}\Big|X_i,W_i\Big]\Big|X_i\Big] \\
&\quad + E[\text{Var}[Y_i(1)|X_i,W_i]|X_i] \ . \qquad (A.13)
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\sigma_n^2(Q) &= \frac{1}{2}E[\text{Var}[E[Y_i(1)+Y_i(0)|X_i,W_i] - (m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i))|X_i]] \\
&\quad + \frac{1}{2}E[\text{Var}[E[Y_i(1)-Y_i(0)|X_i,W_i]|X_i]] + \frac{1}{2}\text{Var}[E[Y_i(1)-Y_i(0)|X_i]] \\
&\quad + E[\text{Var}[Y_i(0)|X_i,W_i]|X_i] + E[\text{Var}[Y_i(1)|X_i,W_i]|X_i] \\
&= \frac{1}{2}E[\text{Var}[E[Y_i(1)+Y_i(0)|X_i,W_i] - (m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i))|X_i]] \\
&\quad + \frac{1}{2}E[(E[Y_i(1)-Y_i(0)|X_i,W_i] - E[Y_i(1)-Y_i(0)|X_i])^2] \\
&\quad + \frac{1}{2}E[(E[Y_i(1)-Y_i(0)|X_i] - E[Y_i(1)-Y_i(0)])^2] \\
&\quad + E[(Y_i(0)-E[Y_i(0)|X_i,W_i])^2] + E[(Y_i(1)-E[Y_i(1)|X_i,W_i])^2] \\
&= \frac{1}{2}E[\text{Var}[E[Y_i(1)+Y_i(0)|X_i,W_i] - (m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i))|X_i]] \\
&\quad + \frac{1}{2}\text{Var}[E[Y_i(1)-Y_i(0)|X_i,W_i]] + E[\text{Var}[Y_i(0)|X_i,W_i]] + E[\text{Var}[Y_i(1)|X_i,W_i]] \ ,
\end{aligned}
$$

where the first equality follows by definition, the second one follows from (A.12)–(A.13), the third one again follows by definition, and the last one follows because by the law of iterated expectations,

$$
E[(E[Y_i(1)-Y_i(0)|X_i,W_i] - E[Y_i(1)-Y_i(0)|X_i])(E[Y_i(1)-Y_i(0)|X_i] - E[Y_i(1)-Y_i(0)])] = 0 \ .
$$

The conclusion then follows. ■

## A.2 Proof of Theorem 3.2

Theorem 3.1 implies $\hat{\Delta}_n \xrightarrow{P} \Delta(Q)$. Next, we show

$$
\hat{\tau}_n^2 - E[\text{Var}[\phi_{1,n,i}|X_i]] + E[\text{Var}[\phi_{0,n,i}|X_i]] + E[(E[Y_i(1)|X_i] - E[Y_i(0)|X_i])^2] \xrightarrow{P} 0 \ . \qquad (A.14)
$$

To that end, define

$$
\mathring{Y}_i = Y_i - \frac{1}{2}(m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i)) \ .
$$

Note

$$
\hat{\tau}_n^2 = \frac{1}{n}\sum_{1\le j\le n}\Big(\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)} + (\tilde{Y}_{\pi(2j-1)} - \tilde{Y}_{\pi(2j)} - (\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)}))\Big)^2
$$

$$= \frac{1}{n} \sum_{1 \le j \le n} (\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)})^2 + \frac{1}{n} \sum_{1 \le j \le n} (\tilde{Y}_{\pi(2j-1)} - \tilde{Y}_{\pi(2j)} - (\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)}))^2$$
$$+ \frac{2}{n} \sum_{1 \le j \le n} (\tilde{Y}_{\pi(2j-1)} - \tilde{Y}_{\pi(2j)} - (\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)}))(\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)}) \ .$$

Therefore, to establish (A.14), we first show

$$\frac{1}{n} \sum_{1 \le j \le n} (\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)})^2 - E[\text{Var}[\phi_{1,n,i}|X_i]] + E[\text{Var}[\phi_{0,n,i}|X_i]] + E[(E[Y_i(1)|X_i] - E[Y_i(0)|X_i])^2] \xrightarrow{P} 0 \tag{A.15}$$

and

$$\frac{1}{n} \sum_{1 \le j \le n} (\tilde{Y}_{\pi(2j-1)} - \tilde{Y}_{\pi(2j)} - (\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)}))^2 \xrightarrow{P} 0 \ . \tag{A.16}$$

(A.16) immediately follows from repeated applications of the inequality $(a-b)^2 \le 2(a^2+b^2)$ and (12). To verify (A.15), note

$$\frac{1}{n} \sum_{1 \le j \le n} (\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)})^2 = \frac{1}{n} \sum_{1 \le i \le 2n} \mathring{Y}_i^2 - \frac{2}{n} \sum_{1 \le j \le n} \mathring{Y}_{\pi(2j-1)} \mathring{Y}_{\pi(2j)} \ .$$

It follows from similar arguments to those in the proof of Lemma B.2 below that

$$\frac{1}{n} \sum_{1 \le i \le 2n} \mathring{Y}_i^2 - E[\phi_{1,n,i}^2] + E[\phi_{0,n,i}^2] \xrightarrow{P} 0 \ .$$

Similarly, it follows from the proof of the same lemma that

$$\frac{2}{n} \sum_{1 \le j \le n} \mathring{Y}_{\pi(2j-1)} \mathring{Y}_{\pi(2j)} - 2E[E[\phi_{1,n,i}|X_i]E[\phi_{0,n,i}|X_i]] \xrightarrow{P} 0 \ .$$

To establish (A.15), note

$$\begin{aligned}
&E[\phi_{1,n,i}^2] + E[\phi_{0,n,i}^2] - 2E[E[\phi_{1,n,i}|X_i]E[\phi_{0,n,i}|X_i]] \\
&= E[\text{Var}[\phi_{1,n,i}|X_i]] + E[\text{Var}[\phi_{0,n,i}|X_i]] + E[E[\phi_{1,n,i}|X_i]^2] + E[E[\phi_{0,n,i}|X_i]^2] - 2E[E[\phi_{1,n,i}|X_i]E[\phi_{0,n,i}|X_i]] \\
&= E[\text{Var}[\phi_{1,n,i}|X_i]] + E[\text{Var}[\phi_{0,n,i}|X_i]] + E[(E[\phi_{1,n,i}|X_i] - E[\phi_{0,n,i}|X_i])^2] \\
&= E[\text{Var}[\phi_{1,n,i}|X_i]] + E[\text{Var}[\phi_{0,n,i}|X_i]] + E[(E[Y_i(1)|X_i] - E[Y_i(0)|X_i])^2] \ ,
\end{aligned}$$

where the last equality follows from the definition of $\phi_{1,n,i}$ and $\phi_{0,n,i}$. It then follows from the Cauchy-Schwarz inequality that

$$\begin{aligned}
&\left| \frac{1}{n} \sum_{1 \le j \le n} (\tilde{Y}_{\pi(2j-1)} - \tilde{Y}_{\pi(2j)} - (\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)}))(\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)}) \right| \\
&\qquad \le \left( \frac{1}{n} \sum_{1 \le j \le n} (\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)})^2 \right) \left( \frac{1}{n} \sum_{1 \le j \le n} (\tilde{Y}_{\pi(2j-1)} - \tilde{Y}_{\pi(2j)} - (\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)}))^2 \right) \xrightarrow{P} 0 \ ,
\end{aligned}$$

which, together with (A.15)–(A.16) as well as Assumptions 2.1(b) and 3.1(b), imply (A.14).

Next, we show

$$\hat{\lambda}_n \xrightarrow{P} E[(E[Y_i(1)|X_i] - E[Y_i(0)|X_i])^2] \ . \tag{A.17}$$

Note

$$\begin{aligned}
&\hat{\lambda}_n - \frac{2}{n} \sum_{1 \le j \le \lfloor \frac{n}{2} \rfloor} (\mathring{Y}_{\pi(4j-3)} - \mathring{Y}_{\pi(4j-2)})(\mathring{Y}_{\pi(4j-1)} - \mathring{Y}_{\pi(4j)})(D_{\pi(4j-3)} - D_{\pi(4j-2)})(D_{\pi(4j-1)} - D_{\pi(4j)}) \quad \text{(A.18)} \\
&= \frac{2}{n} \sum_{1 \le j \le \lfloor \frac{n}{2} \rfloor} (\tilde{Y}_{\pi(4j-3)} - \mathring{Y}_{\pi(4j-3)} - (\tilde{Y}_{\pi(4j-2)} - \mathring{Y}_{\pi(4j-2)}))(\mathring{Y}_{\pi(4j-1)} - \mathring{Y}_{\pi(4j)}) \\
&\qquad \times (D_{\pi(4j-3)} - D_{\pi(4j-2)})(D_{\pi(4j-1)} - D_{\pi(4j)}) \\
&\quad + \frac{2}{n} \sum_{1 \le j \le \lfloor \frac{n}{2} \rfloor} (\mathring{Y}_{\pi(4j-3)} - \mathring{Y}_{\pi(4j-2)})(\tilde{Y}_{\pi(4j-1)} - \mathring{Y}_{\pi(4j-1)}) - (\tilde{Y}_{\pi(4j)} - \mathring{Y}_{\pi(4j)})) \\
&\qquad \times (D_{\pi(4j-3)} - D_{\pi(4j-2)})(D_{\pi(4j-1)} - D_{\pi(4j)}) \\
&\quad + \frac{2}{n} \sum_{1 \le j \le \lfloor \frac{n}{2} \rfloor} (\tilde{Y}_{\pi(4j-3)} - \mathring{Y}_{\pi(4j-3)} - (\tilde{Y}_{\pi(4j-2)} - \mathring{Y}_{\pi(4j-2)})) \\
&\qquad \times (\tilde{Y}_{\pi(4j-1)} - \mathring{Y}_{\pi(4j-1)}) - (\tilde{Y}_{\pi(4j)} - \mathring{Y}_{\pi(4j)}))(D_{\pi(4j-3)} - D_{\pi(4j-2)})(D_{\pi(4j-1)} - D_{\pi(4j)}) \ .
\end{aligned}$$

In what follows, we show

$$\frac{2}{n} \sum_{1 \le j \le \lfloor \frac{n}{2} \rfloor} (\mathring{Y}_{\pi(4j-3)} - \mathring{Y}_{\pi(4j-2)})^2 = O_P(1) \tag{A.19}$$

$$\frac{2}{n} \sum_{1 \le j \le \lfloor \frac{n}{2} \rfloor} (\mathring{Y}_{\pi(4j-1)} - \mathring{Y}_{\pi(4j)})^2 = O_P(1) \tag{A.20}$$

$$\frac{2}{n} \sum_{1 \le j \le \lfloor \frac{n}{2} \rfloor} (\tilde{Y}_{\pi(4j-3)} - \mathring{Y}_{\pi(4j-3)} - (\tilde{Y}_{\pi(4j-2)} - \mathring{Y}_{\pi(4j-2)}))^2 = o_P(1) \tag{A.21}$$

$$\frac{2}{n} \sum_{1 \le j \le \lfloor \frac{n}{2} \rfloor} (\tilde{Y}_{\pi(4j-1)} - \mathring{Y}_{\pi(4j-1)}) - (\tilde{Y}_{\pi(4j)} - \mathring{Y}_{\pi(4j)}))^2 = o_P(1) \tag{A.22}$$

$$\begin{aligned}
&\frac{2}{n} \sum_{1 \le j \le \lfloor \frac{n}{2} \rfloor} (\mathring{Y}_{\pi(4j-3)} - \mathring{Y}_{\pi(4j-2)})(\mathring{Y}_{\pi(4j-1)} - \mathring{Y}_{\pi(4j)})(D_{\pi(4j-3)} - D_{\pi(4j-2)})(D_{\pi(4j-1)} - D_{\pi(4j)}) \\
&\qquad \xrightarrow{P} E[(E[Y_i(1)|X_i] - E[Y_i(0)|X_i])^2] \ .
\end{aligned} \tag{A.23}$$

To establish (A.19)–(A.20), note they follow directly from (A.15) and Assumptions 2.1(b) and 3.1(b). Next, note (A.21) follows from repeated applications of the inequality $(a+b)^2 \le 2(a^2+b^2)$ and (12). (A.22) can be established by similar arguments. (A.23) follows from similar arguments to those in the proof of Lemma S.1.7 of Bai et al. (2022), with the uniform integrability arguments replaced by arguments similar to those in the proof of Lemma B.2, together with Assumptions 2.1–2.4 and 3.1. (A.18)–(A.23) imply (A.17) immediately.

Finally, note we have shown

$$\hat{\sigma}_n^2 - \sigma_n^2 \xrightarrow{P} 0 \ .$$

Assumption 3.1(a) implies $\sigma_n^2$ is bounded away from zero, so

$$\frac{\hat{\sigma}_n}{\sigma_n} \xrightarrow{P} 1 \ .$$

The conclusion of the theorem then follows. ■

## A.3 Proof of Theorem 4.1

We will apply the Frisch-Waugh-Lovell theorem to obtain an expression for $\hat{\beta}_n^{\text{naive}}$. Consider the linear regression of $\psi_i$ on 1 and $D_i$. Define

$$\hat{\mu}_{\psi,n}(d) = \frac{1}{n} \sum_{1 \le i \le 2n} \psi_i I\{D_i = d\}$$

for $d \in \{0, 1\}$ and

$$\hat{\Delta}_{\psi,n} = \hat{\mu}_{\psi,n}(1) - \hat{\mu}_{\psi,n}(0) \ .$$

The $i$th residual based on the OLS estimation of this linear regression model is given by

$$\tilde{\psi}_i = \psi_i - \hat{\mu}_{\psi,n}(0) - \hat{\Delta}_{\psi,n} D_i \ .$$

$\hat{\beta}_n^{\text{naive}}$ is then given by the OLS estimator of the coefficient in the linear regression of $Y_i$ on $\tilde{\psi}_i$. Note

$$\begin{aligned}\frac{1}{2n} \sum_{1 \le i \le 2n} \tilde{\psi}_i \tilde{\psi}_i' &= \frac{1}{2n} \sum_{1 \le i \le 2n} (\psi_i - \hat{\mu}_{\psi,n}(1))(\psi_i - \hat{\mu}_{\psi,n}(1))' D_i + \frac{1}{2n} \sum_{1 \le i \le 2n} (\psi_i - \hat{\mu}_{\psi,n}(0))(\psi_i - \hat{\mu}_{\psi,n}(0))'(1 - D_i) \\ &= \frac{1}{2n} \sum_{1 \le i \le 2n} \psi_i \psi_i' - \frac{1}{2} \hat{\mu}_{\psi,n}(1) \hat{\mu}_{\psi,n}(1)' - \frac{1}{2} \hat{\mu}_{\psi,n}(0) \hat{\mu}_{\psi,n}(0)' \ .\end{aligned}$$

It follows from Assumption 4.1(b) and the weak law of large number that

$$\frac{1}{2n} \sum_{1 \le i \le 2n} \psi_i \psi_i' \xrightarrow{P} E[\psi_i \psi_i'] \ .$$

On the other hand, it follows from Assumptions 2.2–2.3 and 4.1(b)–(c) as well as similar arguments to those in the proof of Lemma S.1.5 of Bai et al. (2022) that

$$\hat{\mu}_{\psi,n}(d) \xrightarrow{P} E[\psi_i]$$

for $d \in \{0, 1\}$. Therefore,

$$\frac{1}{2n} \sum_{1 \le i \le 2n} \tilde{\psi}_i \tilde{\psi}_i' \xrightarrow{P} \text{Var}[\psi_i] \ .$$

Next,

$$\frac{1}{2n} \sum_{1 \le i \le 2n} \tilde{\psi}_i Y_i = \frac{1}{2n} \sum_{1 \le i \le 2n} (\psi_i - \hat{\mu}_{\psi,n}(1)) Y_i(1) D_i + \frac{1}{2n} \sum_{1 \le i \le 2n} (\psi_i - \hat{\mu}_{\psi,n}(0)) Y_i(0)(1 - D_i)$$

It follows from similar arguments as above as well as Assumptions 2.1(b), 2.2–2.3, and 4.1(b)–(c) that

$$\frac{1}{2n}\sum_{1\le i\le 2n}\tilde{\psi}_i Y_i \xrightarrow{P} \mathrm{Cov}[\psi_i, Y_i(1)+Y_i(0)]\ .$$

The convergence of $\hat{\beta}_n^{\text{naive}}$ therefore follows from the continuous mapping theorem and Assumption 4.1(a).

To see (12) is satisfied, note

$$\frac{1}{2n}\sum_{1\le i\le 2n}(\hat{m}_{d,n}(X_i,W_i)-m_{d,n}(X_i,W_i))^2 = (\hat{\beta}_n^{\text{naive}}-\beta^{\text{naive}})'\left(\frac{1}{2n}\sum_{1\le i\le 2n}\psi_i\psi_i'\right)(\hat{\beta}_n^{\text{naive}}-\beta^{\text{naive}})\ .$$

(12) then follows from the fact that $\hat{\beta}_n^{\text{naive}} \xrightarrow{P} \beta$, Assumption 4.1(b), and the weak law of large numbers. To establish (9), first note

$$\frac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i-1)(\hat{m}_{d,n}(X_i,W_i)-m_{d,n}(X_i,W_i)) = \frac{1}{\sqrt{2}}\sqrt{n}\hat{\Delta}_{\psi,n}'(\hat{\beta}_n^{\text{naive}}-\beta^{\text{naive}})\ .$$

In what follows, we establish

$$\sqrt{n}\hat{\Delta}_{\psi,n} = O_P(1)\ , \tag{A.24}$$

from which (9) follows immediately because $\hat{\beta}_n^{\text{naive}}-\beta^{\text{naive}} = o_P(1)$. Note by Assumption 2.2 that $E[\sqrt{n}\hat{\Delta}_{\psi,n}|X^{(n)}] = 0$. Also note

$$\sqrt{n}\hat{\Delta}_{\psi,n} = F_n - G_n + H_n\ ,$$

where

$$\begin{aligned}
F_n &= \frac{1}{\sqrt{n}}\sum_{1\le i\le 2n}(\psi_i - E[\psi_i|X_i])D_i\ ,\\
G_n &= \frac{1}{\sqrt{n}}\sum_{1\le i\le 2n}(\psi_i - E[\psi_i|X_i])(1-D_i)\ , \quad \text{and}\\
H_n &= \frac{1}{\sqrt{n}}\sum_{1\le j\le n}(E[\psi_{\pi(2j-1)}|X_{\pi(2j-1)}]-E[\psi_{\pi(2j)}|X_{\pi(2j)}])(D_{\pi(2j-1)}-D_{\pi(2j)})\ .
\end{aligned}$$

We will argue $F_n, G_n, H_n$ are all $O_P(1)$. Since this could be carried out separately for each entry of $F_n$ and $G_n$, we assume without loss of generality that $k_\psi = 1$. First, it follows from Assumptions 2.2–2.3 and 4.1(c) as well as similar arguments to those in the proof of Lemma S.1.4 of Bai et al. (2022) that

$$\mathrm{Var}[F_n|X^{(n)},D^{(n)}] = \frac{1}{n}\sum_{1\le i\le 2n}\mathrm{Var}[\psi_i|X_i]D_i \xrightarrow{P} E[\mathrm{Var}[\psi_i|X_i]] > 0\ .$$

It then follows from similar arguments using the Lindeberg central limit theorem as in the proof of Lemma S.1.4 of Bai et al. (2022) that $F_n = O_P(1)$. Similar arguments establish $G_n = O_P(1)$. Finally, we show $H_n = O_P(1)$. Note that $E[H_n|X^{(n)}] = 0$ and by Assumptions 2.2–2.3 and 4.1(c),

$$\mathrm{Var}[H_n|X^{(n)}] = \frac{1}{n}\sum_{1\le j\le n}(E[\psi_{\pi(2j-1)}|X_{\pi(2j-1)}]-E[\psi_{\pi(2j)}|X_{\pi(2j)}])^2 \xrightarrow{P} 0\ .$$

Therefore, for any fixed $\epsilon > 0$, Markov's inequality implies

$$P\{|H_n - E[H_n|X^{(n)}]| > \epsilon | X^{(n)}\} \leq \frac{\text{Var}[H_n|X^{(n)}]}{\epsilon^2} \xrightarrow{P} 0 \ .$$

Since probabilities are bounded and therefore uniformly integrable, we have that

$$P\{|H_n - E[H_n|X^{(n)}]| > \epsilon\} \to 0 \ .$$

Therefore, (A.24) follows. Finally, it is straightforward to see Assumption 3.1 is implied by Assumption 4.1. ■

## A.4 Proof of Theorem 4.2

By the Frisch-Waugh-Lovell theorem, $\hat{\beta}_n^{\text{pfe}}$ is equal to the OLS estimator in the linear regression of $\{(Y_{\pi(2j-1)} - Y_{\pi(2j)}, Y_{\pi(2j)} - Y_{\pi(2j-1)}) : 1 \leq j \leq n\}$ on $\{(2D_{\pi(2j-1)} - 1, 2D_{\pi(2j)} - 1) : 1 \leq j \leq n\}$ and $\{(\psi_{\pi(2j-1)} - \psi_{\pi(2j)}, \psi_{\pi(2j)} - \psi_{\pi(2j-1)}) : 1 \leq j \leq n\}$. To apply the Frisch-Waugh-Lovell theorem again, we study the linear regression of $\{(\psi_{\pi(2j-1)} - \psi_{\pi(2j)}, \psi_{\pi(2j)} - \psi_{\pi(2j-1)}) : 1 \leq j \leq n\}$ on $\{(2D_{\pi(2j-1)} - 1, 2D_{\pi(2j)} - 1) : 1 \leq j \leq n\}$. The OLS estimator of the regression coefficient in such a regression equals

$$\hat{\Delta}_{\psi,n} = \frac{1}{n} \sum_{1 \leq j \leq n} (D_{\pi(2j-1)} - D_{\pi(2j)})(\psi_{\pi(2j-1)} - \psi_{\pi(2j)}) \ .$$

The residual is therefore $\{(\psi_{\pi(2j-1)} - \psi_{\pi(2j)} - (2D_{\pi(2j-1)} - 1)\hat{\Delta}_{\psi,n}, \psi_{\pi(2j)} - \psi_{\pi(2j-1)} - (2D_{\pi(2j)} - 1)\hat{\Delta}_{\psi,n}) : 1 \leq j \leq n\}$. $\hat{\beta}_n^{\text{pfe}}$ equals the OLS estimator of the coefficient in the linear regression of $\{(Y_{\pi(2j-1)} - Y_{\pi(2j)}, Y_{\pi(2j)} - Y_{\pi(2j-1)}) : 1 \leq j \leq n\}$ on those residuals. Define

$$\begin{aligned} \delta_{Y,j} &= (D_{\pi(2j-1)} - D_{\pi(2j)})(Y_{\pi(2j-1)} - Y_{\pi(2j)}) \quad \text{and} \\ \delta_{\psi,j} &= (D_{\pi(2j-1)} - D_{\pi(2j)})(\psi_{\pi(2j-1)} - \psi_{\pi(2j)}) \end{aligned}$$

Apparently $\hat{\Delta}_{\psi,n} = \frac{1}{n}\sum_{1 \leq j \leq n} \delta_{\psi,j}$. A moment's thought reveals that $\hat{\beta}_n^{\text{pfe}}$ further equals the coefficient estimate using least squares in the linear regression of $\delta_{Y,j}$ on $\delta_{\psi,j} - \hat{\Delta}_{\psi,n}$ for $1 \leq j \leq n$. It follows from Assumptions 2.1(b)–(c), 2.2–2.3, and 4.1(b)–(c) as well as similar arguments to those in the proof of Lemma S.1.5 of Bai et al. (2022) that

$$\begin{aligned} \hat{\Delta}_{\psi,n} &\xrightarrow{P} 0 \quad \text{and} \\ \frac{1}{n} \sum_{1 \leq j \leq n} \delta_{Y,j} &\xrightarrow{P} \Delta(Q) \ . \end{aligned} \tag{A.25}$$

Next, note that

$$\begin{aligned} &\frac{1}{n} \sum_{1 \leq j \leq n} \delta_{\psi,j}\delta_{\psi,j}' \\ &= \frac{1}{n} \sum_{1 \leq j \leq n} (\psi_{\pi(2j-1)} - \psi_{\pi(2j)})(\psi_{\pi(2j-1)} - \psi_{\pi(2j)})' \end{aligned}$$

$$= \frac{1}{n} \sum_{1 \leq i \leq 2n} \psi_i \psi_i' - \frac{1}{n} \sum_{1 \leq j \leq n} (\psi_{\pi(2j-1)} \psi_{\pi(2j)}' + \psi_{\pi(2j)} \psi_{\pi(2j-1)}') \ . \tag{A.26}$$

For convenience, we introduce the following notation:

$$\begin{aligned} \mu_d(X_i) &= E[Y_i(d)|X_i] \\ \Psi(X_i) &= E[\psi_i|X_i] \\ \xi_d(X_i) &= E[\psi_i Y_i(d)|X_i] \ . \end{aligned}$$

The first term in (A.26) converges in probability to $2E[\psi_i \psi_i']$ by the weak law of large numbers. For the second term, we have that

$$\begin{aligned} & E\Big[\frac{1}{n} \sum_{1 \leq j \leq n} (\psi_{\pi(2j-1)} \psi_{\pi(2j)}' + \psi_{\pi(2j)} \psi_{\pi(2j-1)}') \Big| X^{(n)} \Big] \\ & = \frac{1}{n} \sum_{1 \leq i \leq 2n} \Psi(X_i) \Psi(X_i)' - \frac{1}{n} \sum_{1 \leq j \leq n} (\Psi(X_{\pi(2j-1)}) - \Psi(X_{\pi(2j)}))(\Psi(X_{\pi(2j-1)}) - \Psi(X_{\pi(2j)}))' \\ & \xrightarrow{P} 2E[\Psi(X_i)\Psi(X_i)'] \ , \end{aligned}$$

where the convergence in probability holds because of Assumptions 2.2–2.3 and 4.1(c). It follows from Assumptions 2.2–2.3 and 4.1(b)–(c) as well as similar arguments to those in the proof of Lemma S.1.6 of Bai et al. (2022) that

$$\Big| \frac{1}{n} \sum_{1 \leq j \leq n} (\psi_{\pi(2j-1)} \psi_{\pi(2j)}' + \psi_{\pi(2j)} \psi_{\pi(2j-1)}') - E\Big[\frac{1}{n} \sum_{1 \leq j \leq n} (\psi_{\pi(2j-1)} \psi_{\pi(2j)}' + \psi_{\pi(2j)} \psi_{\pi(2j-1)}') \Big| X^{(n)} \Big] \Big| \xrightarrow{P} 0 \ .$$

Therefore,

$$\frac{1}{n} \sum_{1 \leq j \leq n} \delta_{\psi,j} \delta_{\psi,j}' \xrightarrow{P} 2E[\operatorname{Var}[\psi_i|X_i]] \ .$$

We now turn to

$$\frac{1}{n} \sum_{1 \leq j \leq n} \delta_{\psi,j} \delta_{Y,j} = \frac{1}{n} \sum_{1 \leq j \leq n} (\psi_{\pi(2j-1)} - \psi_{\pi(2j)})(Y_{\pi(2j-1)} - Y_{\pi(2j)}) \ .$$

Note that

$$\begin{aligned} E[\psi_{\pi(2j-1)} Y_{\pi(2j-1)}|X^{(n)}] &= \frac{1}{2}\xi_1(X_{\pi(2j-1)}) + \frac{1}{2}\xi_0(X_{\pi(2j-1)}) \\ E[\psi_{\pi(2j-1)} Y_{\pi(2j)}|X^{(n)}] &= \frac{1}{2}\Psi(X_{\pi(2j-1)})(\mu_1(X_{\pi(2j)}) + \mu_0(X_{\pi(2j)})) \ . \end{aligned}$$

It follows from Assumptions 2.1(b)–(c), 2.2–2.3, 4.1(b)–(c) as well as similar arguments to those in the proof of Lemma S.1.6 of Bai et al. (2022) that

$$\frac{1}{n} \sum_{1 \leq j \leq n} \delta_{\psi,j} \delta_{Y,j} \xrightarrow{P} E[\psi_i (Y_i(1) + Y_i(0))] - E[\Psi(X_i)(\mu_1(X_i) + \mu_0(X_i))] \ .$$

The convergence in probability of $\hat{\beta}_n^{\text{pfe}}$ now follows from Assumption 4.1(a) and the continuous mapping theorem. (9)–(12) can be established using similar arguments to those in the proof of Theorem 4.1. Finally, it is straightforward to see Assumption 3.1 is implied by Assumption 4.1. ■

## A.5 Proof of Theorem 5.1

We divide the proof into three steps. In the first step, we show

$$|\hat{\alpha}^{\mathrm{r}}_{d,n} - \alpha^{\mathrm{r}}_{d,n}| + \|\hat{\beta}^{\mathrm{r}}_{d,n} - \beta^{\mathrm{r}}_{d,n}\|_1 = O_P\left(s_n \lambda^{\mathrm{r}}_n\right) \ . \tag{A.27}$$

In the second step, we show (9), (12), and Assumption 3.1 hold. In the third step, we show the asymptotic variance achieves the minimum under the approximately correct specification condition in Theorem 5.1.

Step 1: Proof of (A.27)

Note that

$$\begin{aligned}
&\frac{1}{n}\sum_{1\le i\le 2n} I\{D_i = d\}(Y_i(d) - \hat{\alpha}^{\mathrm{r}}_{d,n} - \psi'_{n,i}\hat{\beta}^{\mathrm{r}}_{d,n})^2 + \lambda^{\mathrm{r}}_{d,n}\|\hat{\Omega}_n(d)\hat{\beta}^{\mathrm{r}}_{d,n}\|_1 \\
&\le \frac{1}{n}\sum_{1\le i\le 2n} I\{D_i = d\}(Y_i(d) - \alpha^{\mathrm{r}}_{d,n} - \psi'_{n,i}\beta^{\mathrm{r}}_{d,n})^2 + \lambda^{\mathrm{r}}_{d,n}\|\hat{\Omega}_n(d)\beta^{\mathrm{r}}_{d,n}\|_1 \ .
\end{aligned}$$

Rearranging the terms, we then have

$$\begin{aligned}
&\frac{1}{n}\sum_{1\le i\le 2n} I\{D_i = d\}(\hat{\alpha}^{\mathrm{r}}_{d,n} - \alpha^{\mathrm{r}}_{d,n} + \psi'_{n,i}(\hat{\beta}^{\mathrm{r}}_{d,n} - \beta^{\mathrm{r}}_{d,n}))^2 + \lambda^{\mathrm{r}}_{d,n}\|\hat{\Omega}_n(d)\hat{\beta}^{\mathrm{r}}_{d,n}\|_1 \\
&\le \left(\frac{2}{n}\sum_{1\le i\le 2n} I\{D_i = d\}\epsilon_{n,i}(d)\psi'_{n,i}\right)(\hat{\beta}^{\mathrm{r}}_{d,n} - \beta^{\mathrm{r}}_{d,n}) + \left(\frac{2}{n}\sum_{1\le i\le 2n} I\{D_i = d\}\epsilon_{n,i}(d)\right)(\hat{\alpha}^{\mathrm{r}}_{d,n} - \alpha^{\mathrm{r}}_{d,n}) \\
&+ \lambda^{\mathrm{r}}_{d,n}\|\hat{\Omega}_n(d)\beta^{\mathrm{r}}_{d,n}\|_1
\end{aligned} \tag{A.28}$$

Next, define

$$\mathbb{U}_n(d) = \Omega_n^{-1}(d)\frac{1}{n}\sum_{1\le i\le 2n} I\{D_i = d\}(\psi_{n,i}\epsilon_{n,i}(d) - E[\psi_{n,i}\epsilon_{n,i}(d)])$$

and

$$\mathcal{E}_n(d) = \left\{\|\mathbb{U}_n(d)\|_\infty \le \frac{6\bar{\sigma}}{\underline{\sigma}}\sqrt{\frac{\log(2np_n)}{n}},\ \left|\frac{1}{n}\sum_{i\in[2n]} I\{D_i = d\}\epsilon_{n,i}(d) - E[\epsilon_{n,i}(d)]\right| \le \sqrt{\frac{\log(2np_n)}{n}}\right\} \ .$$

Lemma B.4 implies $P\{\mathcal{E}_n(d)\} \to 1$ for $d \in \{0,1\}$.

On the event $\mathcal{E}_n(d)$, we have

$$\begin{aligned}
&\left|\left(\frac{2}{n}\sum_{1\le i\le 2n} I\{D_i = d\}\epsilon_{n,i}(d)\psi'_{n,i}\right)(\hat{\beta}^{\mathrm{r}}_{d,n} - \beta^{\mathrm{r}}_{d,n})\right| \\
&\le \left\|\Omega_n^{-1}(d)\frac{2}{n}\sum_{1\le i\le 2n} I\{D_i = d\}\epsilon_{n,i}(d)\psi_{n,i}\right\|_\infty \|\Omega_n(d)(\hat{\beta}^{\mathrm{r}}_{d,n} - \beta^{\mathrm{r}}_{d,n})\|_1
\end{aligned}$$

$$\leq 2\|\mathbb{U}_n(d)\|_\infty \|\Omega_n(d)(\hat\beta^{\rm r}_{d,n} - \beta^{\rm r}_{d,n})\|_1 + \left\|\Omega_n^{-1}(d)\frac{2}{n}\sum_{1\leq i\leq 2n} I\{D_i = d\}E[\epsilon_{n,i}(d)\psi_{n,i}]\right\|_\infty \|\Omega_n(d)(\hat\beta^{\rm r}_{d,n} - \beta^{\rm r}_{d,n})\|_1$$
$$\leq 2\|\mathbb{U}_n(d)\|_\infty \|\Omega_n(d)(\hat\beta^{\rm r}_{d,n} - \beta^{\rm r}_{d,n})\|_1 + \left\|\Omega_n^{-1}(d)2E[\epsilon_{n,i}(d)\psi_{n,i}]\right\|_\infty \|\Omega_n(d)(\hat\beta^{\rm r}_{d,n} - \beta^{\rm r}_{d,n})\|_1$$
$$\leq \left(\frac{12\bar\sigma}{\underline{\sigma}\ell\ell_n} + d_n\right)\lambda^{\rm r}_{d,n}\|\Omega_n(d)(\hat\beta^{\rm r}_{d,n} - \beta^{\rm r}_{d,n})\|_1 \ ,$$

where $d_n = o(1)$ and the last inequality follows from (18) and the fact that

$$\lambda^{\rm r}_{d,n} \geq \ell\ell_n\sqrt{\frac{\log(2np_n)}{n}} \ .$$

Next, define

$$\hat\delta_{d,n} = \hat\beta^{\rm r}_{d,n} - \beta^{\rm r}_{d,n}$$

and let $S_{d,n}$ be the support of $\beta^{\rm r}_{d,n}$. Then, we have

$$\|\hat\Omega_n(d)\hat\beta^{\rm r}_{d,n}\|_1 = \|(\hat\Omega_n(d)\hat\beta^{\rm r}_{d,n})_{S_{d,n}}\|_1 + \|(\hat\Omega_n(d)\hat\beta^{\rm r}_{d,n})_{S^c_{d,n}}\|_1 = \|(\hat\Omega_n(d)\hat\beta^{\rm r}_{d,n})_{S_{d,n}}\|_1 + \|(\hat\Omega_n(d)\hat\delta_{d,n})_{S^c_{d,n}}\|_1 \ ,$$
$$\|\hat\Omega_n(d)\beta^{\rm r}_{d,n}\|_1 = \|(\hat\Omega_n(d)\beta^{\rm r}_{d,n})_{S_{d,n}}\|_1 \leq \|(\hat\Omega_n(d)\hat\beta^{\rm r}_{d,n})_{S_{d,n}}\|_1 + \|(\hat\Omega_n(d)\hat\delta^{\rm r}_{d,n})_{S_{d,n}}\|_1 \ ,$$

and thus,

$$\left(\frac{12\bar\sigma}{\underline{\sigma}\ell\ell_n} + d_n\right)\|\Omega_n(d)\hat\delta_{d,n}\|_1 + \|\hat\Omega_n(d)\beta^{\rm r}_{d,n}\|_1 - \|\hat\Omega_n(d)\hat\beta^{\rm r}_{d,n}\|_1$$
$$= \left(\frac{12\bar\sigma}{\underline{\sigma}\ell\ell_n} + d_n\right)\|(\Omega_n(d)\hat\delta_{d,n})_{S_{d,n}}\|_1 + \left(\frac{12\bar\sigma}{\underline{\sigma}\ell\ell_n} + d_n\right)\|(\Omega_n(d)\hat\delta_{d,n})_{S^c_{d,n}}\|_1 + \|\hat\Omega_n(d)\beta^{\rm r}_{d,n}\|_1 - \|\hat\Omega_n(d)\hat\beta^{\rm r}_{d,n}\|_1$$
$$\leq \left(\frac{12\bar\sigma}{\underline{\sigma}\ell\ell_n} + d_n\right)\|(\Omega_n(d)\hat\delta_{d,n})_{S_{d,n}}\|_1 + \left(\frac{12\bar\sigma}{\underline{\sigma}\ell\ell_n} + d_n\right)\|(\Omega_n(d)\hat\delta_{d,n})_{S^c_{d,n}}\|_1 + \|(\hat\Omega_n(d)\hat\delta^{\rm r}_{d,n})_{S_{d,n}}\|_1 - \|(\hat\Omega_n(d)\hat\delta^{\rm r}_{d,n})_{S^c_{d,n}}\|_1 \ .$$

Further define $\breve\delta_{d,n} = (\hat\alpha^{\rm r}_{d,n} - \alpha^{\rm r}_{d,n}, \hat\delta'_{d,n})'$ and $\breve S_{d,n} = \{1, S_{d,n}+1\}$[1] and recall $\breve\psi_{n,i} = (1, \psi'_{n,i})'$. Then, together with (A.28), we have

$$\begin{aligned}
0 &\leq \frac{1}{n}\sum_{1\leq i\leq 2n} I\{D_i = d\}(\breve\psi'_{n,i}\breve\delta_{d,n})^2 \\
&\leq \lambda^{\rm r}_{d,n}\left[\left(\frac{12\bar\sigma}{\underline{\sigma}\ell\ell_n} + d_n + \bar c\right)\|(\Omega_n(d)\hat\delta_{d,n})_{S_{d,n}}\|_1 - \left(\underline{c} - \frac{12\bar\sigma}{\underline{\sigma}\ell\ell_n} - d_n\right)\|(\Omega_n(d)\hat\delta_{d,n})_{S^c_{d,n}}\|_1\right] \\
&\quad + \lambda^{\rm r}_{d,n}(1/\ell_n + d_n)|\hat\alpha^{\rm r}_{d,n} - \alpha^{\rm r}_{d,n}| \\
&\leq \lambda^{\rm r}_{d,n}\left[\left(\frac{12\bar\sigma}{\underline{\sigma}\ell\ell_n} + d_n + \bar c\right)\bar\sigma\|(\hat\delta_{d,n})_{S_{d,n}}\|_1 - \left(\underline{c} - \frac{12\bar\sigma}{\underline{\sigma}\ell\ell_n} - d_n\right)\underline{\sigma}\|(\hat\delta_{d,n})_{S^c_{d,n}}\|_1\right] \\
&\quad + \lambda^{\rm r}_{d,n}(1/\ell_n + d_n)|\hat\alpha^{\rm r}_{d,n} - \alpha^{\rm r}_{d,n}| \\
&\leq \lambda^{\rm r}_{d,n}\left[\left(\frac{12\bar\sigma}{\underline{\sigma}\ell\ell_n} + d_n + \bar c\right)\bar\sigma\|(\breve\delta_{d,n})_{\breve S_{d,n}}\|_1 - \left(\underline{c} - \frac{12\bar\sigma}{\underline{\sigma}\ell\ell_n} - d_n\right)\underline{\sigma}\|(\breve\delta_{d,n})_{\breve S^c_{d,n}}\|_1\right] \ . 
\end{aligned} \tag{A.29}$$

Define

$$\mathcal{C}_n = \left\{u \in \mathbf{R}^{p_n+1} : \|u_{\breve S^c_{d,n}}\|_1 \leq \frac{2\bar\sigma\bar c}{\underline{\sigma}\underline{c}}\|u_{\breve S_{d,n}}\|_1\right\} \ .$$

[1] For example, if $S_{d,n} = \{2,4,9\}$, we have $\breve S_{d,n} = \{1,3,5,10\}$.

For sufficiently large $n$, we have $\breve{\delta}_{d,n} \in \mathcal{C}_n$. It follows from Bickel et al. (2009) and Assumption 5.4 that

$$\inf_{u \in \mathcal{C}_n} (\|u_{\breve{S}_{d,n}}\|_1)^{-2}(s_n+1)u' \left( \frac{1}{n} \sum_{1\le i\le 2n} I\{D_i = d\} \breve{\psi}_{n,i} \breve{\psi}'_{n,i} \right) u \ge 0.25\kappa_1^2 \ .$$

Therefore, we have

$$\begin{aligned} 0.25\kappa_1^2 \|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1^2 &\le \frac{1}{n} \sum_{1\le i\le 2n} I\{D_i = d\}(\breve{\psi}'_{n,i}\breve{\delta}_{d,n})^2 \\ &\le \left( \frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n + \bar{c} \right) \lambda^{\mathrm{r}}_{d,n}(s_n+1)\|(\breve{\delta}_{d,n})_{S_{d,n}}\|_1 \ , \end{aligned}$$

which implies

$$\|(\breve{\delta}_{d,n})_{S_{d,n}}\|_1 \le 4 \left( \frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n + \bar{c} \right) (s_n+1)\lambda^{\mathrm{r}}_{d,n}/\kappa_1^2 \ .$$

We then have

$$\begin{aligned} |\hat{\alpha}^{\mathrm{r}}_{d,n} - \alpha^{\mathrm{r}}_{d,n}| + \|\hat{\beta}^{\mathrm{r}}_{d,n} - \beta^{\mathrm{r}}_{d,n}\|_1 &\le \|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1 + \|(\breve{\delta}_{d,n})_{\breve{S}^c_{d,n}}\|_1 \\ &\le \left(1 + \frac{2\bar{\sigma}\bar{c}}{\underline{\sigma}\underline{c}}\right) \|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1 \\ &\le 4\left(1 + \frac{2\bar{\sigma}\bar{c}}{\underline{\sigma}\underline{c}}\right) \left( \frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n + \bar{c} \right) (s_n+1)\lambda^{\mathrm{r}}_{d,n}/\kappa_1^2 \ . \end{aligned}$$

Then, we have (A.27) holds because $P\{\mathcal{E}_n(d)\} \to 1$.

Step 2: Verifying (9), (12), and Assumption 3.1

By (A.29), on $\mathcal{E}_n(d)$ we have

$$\begin{aligned} \frac{1}{n} \sum_{1\le i\le 2n} I\{D_i = d\}(\hat{\alpha}^{\mathrm{r}}_{d,n} - \alpha^{\mathrm{r}}_{d,n} + \psi'_{n,i}\hat{\delta}_{d,n})^2 &\le \lambda^{\mathrm{r}}_{d,n} \left( \frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n + \bar{c} \right) \bar{\sigma}\|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1 \\ &\le 4 \left( \frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n + \bar{c} \right)^2 \bar{\sigma}(s_n+1)\lambda^{\mathrm{r},2}_{d,n}/\kappa_1^2 \ . \end{aligned}$$

Because $P\{\mathcal{E}_n(d)\} \to 1$, we have

$$\frac{1}{n} \sum_{1\le i\le 2n} I\{D_i = d\}(\hat{\alpha}^{\mathrm{r}}_{d,n} - \alpha^{\mathrm{r}}_{d,n} + \psi'_{n,i}(\hat{\beta}_{d,n} - \beta_{d,n}))^2 = O_P\left(s_n(\lambda^{\mathrm{r}}_n)^2\right) = o_P(1) \ ,$$

which implies (12) holds.

Next, we show (9) for $\hat{\beta}^{\mathrm{r}}_{d,n}$. First note

$$\left| \frac{1}{\sqrt{2n}} \sum_{1\le i\le 2n} (2D_i - 1)(\hat{m}_{d,n}(X_i, W_{n,i}) - m_{d,n}(X_i, W_{n,i})) \right|$$

$$
\begin{aligned}
&\left|\frac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i-1)\psi'_{n,i}(\hat\beta^{\mathrm{r}}_{d,n}-\beta^{\mathrm{r}}_{d,n})\right| \\
&\le \left\|\frac{1}{\sqrt{2n}}\sum_{1\le i\le n}(2D_i-1)\psi_{n,i}\right\|_\infty \|\hat\beta^{\mathrm{r}}_{d,n}-\beta^{\mathrm{r}}_{d,n}\|_1 \ .
\end{aligned}
$$

Next, note that it follows from Assumption 2.2 that conditional on $X^{(n)}$ and $W_n^{(n)}$,

$$
\{D_{\pi(2j-1)}-D_{\pi(2j)} : 1\le j\le n\}
$$

is a sequence of independent Rademacher random variables. Therefore, Hoeffding's inequality implies

$$
\begin{aligned}
&P\left\{\left\|\frac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i-1)\psi_{n,i}\right\|_\infty > t \,\middle|\, X^{(n)},W_n^{(n)}\right\} \\
&\le \sum_{1\le l\le p_n} P\left\{\left|\frac{1}{\sqrt{2n}}\sum_{1\le j\le n}(\psi_{n,\pi(2j-1)}-\psi_{n,\pi(2j)})(D_{\pi(2j-1)}-D_{\pi(2j)})\right| > t \,\middle|\, X^{(n)},W_n^{(n)}\right\} \\
&\le \sum_{1\le l\le p_n} 2\exp\left(-\frac{t^2}{\frac{1}{n}\sum_{1\le j\le n}(\psi_{n,\pi(2j-1)}-\psi_{n,\pi(2j)})^2}\right) \ .
\end{aligned}
$$

Define

$$
\nu_n^2 = \max_{1\le l\le p_n}\frac{1}{n}\sum_{1\le i\le 2n}\psi^2_{n,i,l} \ .
$$

We then have

$$
P\left\{\left\|\frac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i-1)\psi_{n,i}\right\|_\infty > \nu_n\sqrt{2\log(p_n\vee n)} \,\middle|\, X^{(n)},W_n^{(n)}\right\} \le (p_n\vee n)^{-1} \ . \tag{A.30}
$$

Next, we determine the order of $\nu_n^2$. Note

$$
\begin{aligned}
E[\nu_n^2] &\le \max_{1\le l\le p_n} 2E[\psi^2_{n,i,l}] + 2E\left[\frac{1}{2n}\sum_{1\le i\le 2n}(\psi^2_{n,i,l}-E[\psi^2_{n,i,l}])\right] \\
&\lesssim 1 + E\left[\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}e_i\psi^2_{n,i,l}\right|\right] \\
&\lesssim 1 + \Xi_n E\left[\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}e_i\psi_{n,i,l}\right|\right] \\
&\lesssim 1 + \Xi_n E\left[\sup_{f\in\mathcal{F}_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}f(e_i,\psi_{n,i,l})\right|\right]
\end{aligned}
$$

where $\{e_i : 1\le i\le n\}$ is an i.i.d. sequence of Rademacher random variables,

$$
\mathcal{F}_n = \{f:\mathbf{R}\times\mathbf{R}^{p_n}\mapsto\mathbf{R}, f(e,\psi)=e\psi_l, 1\le l\le p_n\} \ ,
$$

and $\psi_l$ is the $l$th element of $\psi$. Note the second inequality follows from Lemma 2.3.1 of van der Vaart and Wellner (1996), the third inequality follows from Theorem 4.12 of Ledoux and Talagrand (1991) and the definition of $\Xi_n$, and the last follows from Assumption 5.1. Note also $\mathcal{F}_n$ has an envelope $F = \Xi_n$ and

$$\sup_{n \geq 1} \sup_{f \in \mathcal{F}_n} E[f^2] < \infty$$

because of Assumption 5.1. Because the cardinality of $\mathcal{F}_n$ is $p_n$, for any $\epsilon < 1$ we have that

$$\sup_{Q: Q \text{ is a discrete distribution with finite support}} \mathcal{N}(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \leq \frac{p_n}{\epsilon} ,$$

where $\mathcal{N}(\epsilon, \mathcal{F}, L_2(Q))$ is the covering number for class $\mathcal{F}$ under the metric $L_2(Q)$ using balls of radius $\epsilon$. Therefore, Corollary 5.1 of Chernozhukov et al. (2014) implies

$$E\left[\sup_{f \in \mathcal{F}_n} \left| \frac{1}{2n} \sum_{1 \leq i \leq 2n} e_i \psi_{n,i,l} \right| \right] \lesssim \sqrt{\frac{\log p_n}{n}} + \frac{\Xi_n \log p_n}{n} = o(\Xi_n^{-1}) .$$

Therefore, $\nu_n = O_p(1)$. Together with (A.30), they imply

$$\left\| \frac{1}{\sqrt{2n}} \sum_{1 \leq i \leq 2n} (2D_i - 1)\psi_{n,i} \right\|_\infty = O_P\left(\sqrt{\log(p_n \vee n)}\right) .$$

In light of (A.27) and Assumption 5.3, we have

$$\left\| \frac{1}{\sqrt{2n}} \sum_{1 \leq i \leq 2n} (2D_i - 1)\psi_{n,i} \right\|_\infty \|\hat{\beta}^{\mathrm{r}}_{d,n} - \beta^{\mathrm{r}}_{d,n}\|_1 = O_P\left(\frac{s_n \ell\ell_n \log(p_n \vee n)}{\sqrt{n}}\right) = o_P(1) .$$

Next, note that Assumption 3.1(a) and 3.1(b) follow Assumption 5.1, and Assumption 3.1(c) follows Assumptions 5.1 and 5.2.

Step 3: Asymptotic variance

Suppose the true specification is approximately sparse as specified in Theorem 5.1. Let $\tilde{Y}_i(d) = Y_i(d) - \mu_d(X_i)$, $\tilde{\psi}_{n,i} = \psi_{n,i} - E[\psi_{n,i}|X_i]$, and $\tilde{R}_{n,i}(d) = R_{n,i}(d) - E[R_{n,i}(d)|X_i]$. Then, we have

$$E\left[(E[\tilde{Y}_i(1) + \tilde{Y}_i(0)|W_{n,i}, X_i] - \tilde{\psi}'_{n,i}(\beta^{\mathrm{r}}_{1,n} + \beta^{\mathrm{r}}_{0,n}))^2\right] = E[(\tilde{R}_{n,i}(1) + \tilde{R}_{n,i}(0))^2] = o(1) .$$

This concludes the proof. ■

## A.6 Proof of Theorem 5.2

We divide the proof into three steps. In the first step, we show

$$\hat{\beta}^{\text{refit}}_n - \beta^{\text{refit}}_n = o_P(1) . \tag{A.31}$$

In the second step, we show (9), (12), and Assumption 3.1 hold. In the third step, we show that $\sigma_n^{\mathrm{na},2} \geq \sigma_n^{\mathrm{refit},2}$ and $\sigma_n^{\mathrm{r},2} \geq \sigma_n^{\mathrm{refit},2}$.

Step 1: Proof of (A.31)

Let

$$\begin{aligned}
\hat{\Delta}_{\Gamma,n} &= \frac{1}{n} \sum_{1 \leq j \leq n} (D_{\pi(2j-1)} - D_{\pi(2j)})(\Gamma_{n,\pi(2j-1)} - \Gamma_{n,\pi(2j)}) \ , \\
\hat{\Delta}_{\hat{\Gamma},n} &= \frac{1}{n} \sum_{1 \leq j \leq n} (D_{\pi(2j-1)} - D_{\pi(2j)})(\hat{\Gamma}_{n,\pi(2j-1)} - \hat{\Gamma}_{n,\pi(2j)}) \ , \\
\delta_{\Gamma,j} &= (D_{\pi(2j-1)} - D_{\pi(2j)})(\Gamma_{n,\pi(2j-1)} - \Gamma_{n,\pi(2j)}) \ , \\
\delta_{\hat{\Gamma},j} &= (D_{\pi(2j-1)} - D_{\pi(2j)})(\hat{\Gamma}_{n,\pi(2j-1)} - \hat{\Gamma}_{n,\pi(2j)}) \ .
\end{aligned}$$

Then, by the proof of Theorem 4.2, we have $\hat{\beta}_n^{\mathrm{refit}}$ equals the coefficient estimate using least squares in the linear regression of $\delta_{Y,j}$ on $\delta_{\hat{\Gamma},j} - \hat{\Delta}_{\hat{\Gamma},n}$. Then, for any $u \in \mathbf{R}^2$ such that $||u||_2 = 1$, we have

$$\begin{aligned}
&\left| \left( \frac{1}{n} \sum_{1 \leq j \leq n} ((\delta_{\hat{\Gamma},j} - \hat{\Delta}_{\hat{\Gamma},n})' u)^2 \right)^{1/2} - \left( \frac{1}{n} \sum_{1 \leq j \leq n} ((\delta_{\Gamma,j} - \hat{\Delta}_{\Gamma,n})' u)^2 \right)^{1/2} \right| \\
&\leq \left( \frac{1}{n} \sum_{1 \leq j \leq n} ((\delta_{\hat{\Gamma},j} - \delta_{\Gamma,j})' u)^2 - ((\hat{\Delta}_{\hat{\Gamma},n} - \hat{\Delta}_{\Gamma,n})' u)^2 \right)^{1/2} \\
&\leq \left( \frac{2}{n} \sum_{1 \leq i \leq 2n} \left\| \hat{\Gamma}_{n,i} + (\hat{\alpha}_{1,n}^{\mathrm{r}}, \hat{\alpha}_{0,n}^{\mathrm{r}})' - \Gamma_{n,i} - (\alpha_{1,n}^{\mathrm{r}}, \alpha_{0,n}^{\mathrm{r}})' \right\|_2^2 \right)^{1/2} \\
&\lesssim \sum_{d \in \{0,1\}} \frac{1}{2n} \sum_{1 \leq i \leq 2n} (\hat{\alpha}_{d,n}^{\mathrm{r}} - \alpha_{d,n}^{\mathrm{r}} + \psi_{n,i}' (\hat{\beta}_{d,n}^{\mathrm{r}} - \beta_{d,n}^{\mathrm{r}}))^2 = o_P(1) \ ,
\end{aligned}$$

where the second inequality is by the fact that

$$\begin{aligned}
\delta_{\Gamma,j} &= (D_{\pi(2j-1)} - D_{\pi(2j)})(\Gamma_{n,\pi(2j-1)} + (\alpha_{1,n}^{\mathrm{r}}, \alpha_{0,n}^{\mathrm{r}})' - \Gamma_{n,\pi(2j)} - (\alpha_{1,n}^{\mathrm{r}}, \alpha_{0,n}^{\mathrm{r}})') \ , \\
\delta_{\hat{\Gamma},j} &= (D_{\pi(2j-1)} - D_{\pi(2j)})(\hat{\Gamma}_{n,\pi(2j-1)} + (\hat{\alpha}_{1,n}^{\mathrm{r}}, \hat{\alpha}_{0,n}^{\mathrm{r}})' - \hat{\Gamma}_{n,\pi(2j)} - (\hat{\alpha}_{1,n}^{\mathrm{r}}, \hat{\alpha}_{0,n}^{\mathrm{r}})') \ ,
\end{aligned}$$

and the last equality is by the proof of Theorem 5.1. This implies

$$\begin{aligned}
&\frac{1}{n} \sum_{1 \leq j \leq n} (\delta_{\hat{\Gamma},j} - \hat{\Delta}_{\hat{\Gamma},n})(\delta_{\hat{\Gamma},j} - \hat{\Delta}_{\hat{\Gamma},n})' - 2E[\mathrm{Var}[\Gamma_{n,i} | X_i]] \\
&= \frac{1}{n} \sum_{1 \leq j \leq n} (\delta_{\hat{\Gamma},j} - \hat{\Delta}_{\hat{\Gamma},n})(\delta_{\hat{\Gamma},j} - \hat{\Delta}_{\hat{\Gamma},n})' - \frac{1}{n} \sum_{1 \leq j \leq n} (\delta_{\Gamma,j} - \hat{\Delta}_{\Gamma,n})(\delta_{\Gamma,j} - \hat{\Delta}_{\Gamma,n})' \\
&+ \frac{1}{n} \sum_{1 \leq j \leq n} (\delta_{\Gamma,j} - \hat{\Delta}_{\Gamma,n})(\delta_{\Gamma,j} - \hat{\Delta}_{\Gamma,n})' - 2E[\mathrm{Var}[\Gamma_{n,i} | X_i]] = o_P(1) \ ,
\end{aligned}$$

where the last equality holds due to the same argument as used in the proof of Theorem 4.2. Similarly, we can show that

$$\frac{1}{n}\sum_{1\le j\le n}\delta_{Y,j}(\delta_{\hat{\Gamma},j}-\hat{\Delta}_{\hat{\Gamma},n})-E[\mathrm{Cov}[\Gamma_{n,i},Y_i(1)+Y_i(0)|X_i]]=o_P(1)\ ,$$

which leads to (A.31).

Step 2: Verifying (9), (12), and Assumption 3.1

We first show (9). We have

$$\begin{aligned}
&\frac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i-1)(\hat{m}_{d,n}(X_i,W_{n,i})-m_{d,n}(X_i,W_{n,i}))\\
&=\frac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i-1)(\hat{\Gamma}_{n,i}-\Gamma_{n,i})'\hat{\beta}_n^{\mathrm{refit}}+\frac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i-1)\Gamma_{n,i}'(\hat{\beta}_n^{\mathrm{refit}}-\beta_n^{\mathrm{refit}})\\
&=\left(\tfrac{1}{\sqrt{2n}}\textstyle\sum_{1\le i\le 2n}(2D_i-1)\psi_{n,i}'(\hat{\beta}_{1,n}^{\mathrm{r}}-\beta_{1,n}^{\mathrm{r}}),\tfrac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i-1)\psi_{n,i}'(\hat{\beta}_{0,n}^{\mathrm{r}}-\beta_{0,n}^{\mathrm{r}})\right)\hat{\beta}_n^{\mathrm{refit}}\\
&+\left(\tfrac{1}{\sqrt{2n}}\textstyle\sum_{1\le i\le 2n}(2D_i-1)\psi_{n,i}'\beta_{1,n}^{\mathrm{r}},\tfrac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i-1)\psi_{n,i}'\beta_{0,n}^{\mathrm{r}}\right)(\hat{\beta}_{1,n}^{\mathrm{refit}}-\beta_{1,n}^{\mathrm{refit}})\\
&=o_P(1)\ ,
\end{aligned}$$

where the last equality holds by (A.31) and the facts that

$$\left(\tfrac{1}{\sqrt{2n}}\textstyle\sum_{1\le i\le 2n}(2D_i-1)\psi_{n,i}'(\hat{\beta}_{1,n}^{\mathrm{r}}-\beta_{1,n}^{\mathrm{r}}),\tfrac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i-1)\psi_{n,i}'(\hat{\beta}_{0,n}^{\mathrm{r}}-\beta_{0,n}^{\mathrm{r}})\right)=o_P(1)$$

as shown in Theorem 5.1 and

$$\left(\tfrac{1}{\sqrt{2n}}\textstyle\sum_{1\le i\le 2n}(2D_i-1)\psi_{n,i}'\beta_{1,n}^{\mathrm{r}},\tfrac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i-1)\psi_{n,i}'\beta_{0,n}^{\mathrm{r}}\right)=O_P(1)\ .$$

Next, we show (12). We note that

$$\begin{aligned}
&\frac{1}{2n}\sum_{1\le i\le 2n}(\hat{m}_{d,n}(X_i,W_{n,i})-m_{d,n}(X_i,W_{n,i}))^2\\
&\lesssim\frac{1}{2n}\sum_{1\le i\le 2n}((\hat{\Gamma}_{n,i}-\Gamma_{n,i})'\hat{\beta}_n^{\mathrm{refit}})^2+\frac{1}{2n}\sum_{1\le i\le 2n}(\Gamma_{n,i}'(\hat{\beta}_n^{\mathrm{refit}}-\beta_n^{\mathrm{refit}}))^2\\
&\lesssim\frac{1}{2n}\sum_{1\le i\le 2n}((\psi_{n,i}'(\beta_{1,n}^{\mathrm{r}}-\hat{\beta}_{1,n}^{\mathrm{r}}))^2+(\psi_{n,i}'(\beta_{0,n}^{\mathrm{r}}-\hat{\beta}_{0,n}^{\mathrm{r}}))^2)||\hat{\beta}_n^{\mathrm{refit}}||_2^2\\
&+\frac{1}{2n}\sum_{1\le i\le 2n}[(\psi_{n,i}'\beta_{1,n}^{\mathrm{r}})^2+(\psi_{n,i}'\beta_{0,n}^{\mathrm{r}})^2]||\hat{\beta}_n^{\mathrm{refit}}-\beta_n^{\mathrm{refit}}||_2^2\\
&\lesssim\sum_{d=0,1}\frac{1}{2n}\sum_{1\le i\le 2n}\left[(\alpha_{d,n}^{\mathrm{r}}-\hat{\alpha}_{d,n}^{\mathrm{r}}+\psi_{n,i}'(\beta_{d,n}^{\mathrm{r}}-\hat{\beta}_{d,n}^{\mathrm{r}}))^2+(\alpha_{d,n}^{\mathrm{r}}-\hat{\alpha}_{d,n}^{\mathrm{r}})^2\right]||\hat{\beta}_n^{\mathrm{refit}}||_2^2+o_P(1)\\
&=o_P(1)\ .
\end{aligned}$$

Last, Assumption 3.1(1) can be verified in the same manner as we did in the proof of Theorem 5.1.

Step 3: Asymptotic variance

Recall $\sigma_2^2(Q)$ and $\sigma_3^2(Q)$ defined in Theorem 3.1. As we have already verified (9) for $\hat{m}_{d,n}(X_i, W_{n,i}) = \hat{\Gamma}_{n,i}\hat{\beta}_n^{\text{refit}}$ and $m_{d,n}(X_i, W_{n,i}) = \Gamma_{n,i}\beta_n^{\text{refit}}$, we have, for $b \in \{\text{unadj}, \text{r}, \text{refit}\}$, that

$$\sigma_n^{\text{b},2} - \sigma_2^2(Q) - \sigma_3^2(Q) = \frac{1}{2} E\left[\text{Var}[E[Y_i(1) + Y_i(0)|X_i, W_{n,i}] - \Gamma'_{n,i}\gamma^b|X_i]\right]$$

with

$$\gamma^{\text{unadj}} = (0,0)' \ , \quad \gamma^{\text{r}} = (1,1)' \ , \quad \text{and} \quad \gamma^{\text{refit}} = \beta_n^{\text{refit}} \ .$$

In addition, we note that

$$\frac{1}{2} E\left[\text{Var}[E[Y_i(1) + Y_i(0)|X_i, W_{n,i}] - \Gamma'_{n,i}\gamma|X_i]\right]$$

is minimized at $\gamma = \beta_n^{\text{refit}}$, which leads to the desired result. ■

# B Auxiliary Lemmas

**Lemma B.1.** *Suppose $\phi_n, n \geq 1$ is a sequence of random variables satisfying*

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} E[|\phi_n| I\{|\phi_n| > \lambda\}] = 0 \ . \tag{B.1}$$

*Suppose $X$ is another random variable defined on the same probability space with $\phi_n, n \geq 1$. Then,*

$$\lim_{\gamma \to \infty} \limsup_{n \to \infty} E[E[|\phi_n||X] I\{E[|\phi_n||X] > \gamma\}] = 0 \ . \tag{B.2}$$

Proof. Fix $\epsilon > 0$. We will show there exists $\gamma > 0$ so that

$$\limsup_{n \to \infty} E[E[|\phi_n||X] I\{E[|\phi_n||X] > \gamma\}] < \epsilon \ . \tag{B.3}$$

First note the event $\{E[|\phi_n||X] > \gamma\}$ is measurable with respect to the $\sigma$-algebra generated by $X$, and therefore

$$E[E[|\phi_n||X] I\{E[|\phi_n||X] > \gamma\}] = E[|\phi_n| I\{E[|\phi_n||X] > \gamma\}] \ . \tag{B.4}$$

Next, by Theorem 10.3.5 of Dudley (1989), (B.1) implies that there exists a $\delta > 0$ such that for any sequence of events $A_n$ such that $\limsup_{n \to \infty} P\{A_n\} < \delta$, we have

$$\limsup_{n \to \infty} E[|\phi_n| I\{A_n\}] < \epsilon \ . \tag{B.5}$$

In light of the previous result, note

$$P\{E[|\phi_n||X] > \gamma\} \leq \frac{E[E[|\phi_n||X]]}{\gamma} = \frac{E[|\phi_n|]}{\gamma}$$

By Theorem 10.3.5 of Dudley (1989) again, (B.1) implies $\limsup_{n \to \infty} E[|\phi_n|] < \infty$, so by choosing $\gamma$ large enough, we can make sure

$$\limsup_{n\to\infty} P\{E[|\phi_n||X] > \gamma\} < \delta \text{ for all } n \ .$$

(B.3) then follows from (B.4)–(B.5). ■

**Lemma B.2.** *Suppose Assumptions 2.1–2.3 and 3.1 hold. Then,*

$$\frac{s_n^2}{nE[\mathrm{Var}[\phi_{1,n,i}|X_i]]} \xrightarrow{P} 1 \ .$$

Proof. To begin, note it follows from Assumption 2.2 and $Q_n = Q^{2n}$ that

$$\frac{1}{n}\sum_{1\le i\le 2n} D_i \,\mathrm{Var}[\phi_{1,n,i}|X_i] = \frac{1}{2n}\sum_{1\le i\le 2n}\mathrm{Var}[\phi_{1,n,i}|X_i] + \frac{1}{2n}\sum_{1\le i\le 2n: D_i=1}\mathrm{Var}[\phi_{1,n,i}|X_i] - \frac{1}{2n}\sum_{1\le i\le 2n: D_i=0}\mathrm{Var}[\phi_{1,n,i}|X_i] \ . \quad \text{(B.6)}$$

Next,

$$\left|\frac{1}{2n}\sum_{1\le i\le 2n: D_i=1}\mathrm{Var}[\phi_{1,n,i}|X_i] - \frac{1}{2n}\sum_{1\le i\le 2n: D_i=0}\mathrm{Var}[\phi_{1,n,i}|X_i]\right| \le \frac{1}{2n}\sum_{1\le j\le n}|\,\mathrm{Var}[\phi_{1,n,\pi(2j-1)}|X_{\pi(2j-1)}] - \mathrm{Var}[\phi_{1,n,\pi(2j)}|X_{\pi(2j)}]| \ . \quad \text{(B.7)}$$

In what follows, we will show

$$\frac{1}{n}\sum_{1\le j\le n}|\,\mathrm{Cov}[Y_{\pi(2j-1)}(1), m_{1,n}(X_{\pi(2j-1)}, W_{\pi(2j-1)})|X_{\pi(2j-1)}]] - \mathrm{Cov}[Y_{\pi(2j)}(1), m_{1,n}(X_{\pi(2j)}, W_{\pi(2j)})|X_{\pi(2j)}]]| \xrightarrow{P} 0 \ .$$

To that end, first note from Assumptions 2.3 and 3.1(c) that

$$\frac{1}{n}\sum_{1\le j\le n}|E[Y_{\pi(2j-1)}(1)m_{1,n}(X_{\pi(2j-1)}, W_{\pi(2j-1)})|X_{\pi(2j-1)}] - E[Y_{\pi(2j)}(1)m_{1,n}(X_{\pi(2j)}, W_{\pi(2j)})|X_{\pi(2j)}]| \lesssim \frac{1}{n}\sum_{1\le j\le n}|X_{\pi(2j-1)} - X_{\pi(2j)}| \xrightarrow{P} 0 \ .$$

Next, note

$$\begin{aligned}
&\frac{1}{n}\sum_{1\le j\le n}|E[Y_{\pi(2j-1)}(1)|X_{\pi(2j-1)}]E[m_{1,n}(X_{\pi(2j-1)}, W_{\pi(2j-1)})|X_{\pi(2j-1)}] \\
&\qquad - E[Y_{\pi(2j)}(1)|X_{\pi(2j)}]E[m_{1,n}(X_{\pi(2j)}, W_{\pi(2j)})|X_{\pi(2j)}]| \\
&\le \frac{1}{n}\sum_{1\le j\le n}|E[Y_{\pi(2j-1)}(1)|X_{\pi(2j-1)}]||E[m_{1,n}(X_{\pi(2j-1)}, W_{\pi(2j-1)})|X_{\pi(2j-1)}] - E[m_{1,n}(X_{\pi(2j)}, W_{\pi(2j)})|X_{\pi(2j)}]| \\
&\qquad + \frac{1}{n}\sum_{1\le j\le n}|E[Y_{\pi(2j-1)}(1)|X_{\pi(2j-1)}] - E[Y_{\pi(2j)}(1)|X_{\pi(2j)}]||E[m_{1,n}(X_{\pi(2j)}, W_{\pi(2j)})|X_{\pi(2j)}]|
\end{aligned}$$

$$
\begin{aligned}
&\leq \left( \frac{1}{n} \sum_{1 \leq j \leq n} |E[Y_{\pi(2j-1)}(1)|X_{\pi(2j-1)}]|^2 \right)^{1/2} \\
&\qquad \times \left( \frac{1}{n} \sum_{1 \leq j \leq n} |E[m_{1,n}(X_{\pi(2j-1)}, W_{\pi(2j-1)})|X_{\pi(2j-1)}] - E[m_{1,n}(X_{\pi(2j)}, W_{\pi(2j)})|X_{\pi(2j)}]|^2 \right)^{1/2} \\
&\quad + \left( \frac{1}{n} \sum_{1 \leq j \leq n} |E[m_{1,n}(X_{\pi(2j)}, W_{\pi(2j)})|X_{\pi(2j)}]|^2 \right)^{1/2} \\
&\qquad \times \left( \frac{1}{n} \sum_{1 \leq j \leq n} |E[Y_{\pi(2j-1)}(1)|X_{\pi(2j-1)}] - E[Y_{\pi(2j)}(1)|X_{\pi(2j)}]|^2 \right)^{1/2} \\
&\lesssim \left( \frac{1}{n} \sum_{1 \leq i \leq 2n} |E[Y_i(1)|X_i]|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{1 \leq j \leq n} |X_{\pi(2j-1)} - X_{\pi(2j)}|^2 \right)^{1/2} \\
&\quad + \left( \frac{1}{n} \sum_{1 \leq i \leq 2n} |E[m_{1,n}(X_i, W_i)|X_i]|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{1 \leq j \leq n} |X_{\pi(2j-1)} - X_{\pi(2j)}|^2 \right)^{1/2} \xrightarrow{P} 0 ,
\end{aligned}
$$

where the first inequality follows from the triangle inequality, the second follows from the Cauchy-Schwarz inequality, the last follows from Assumptions 2.1(c) and 3.1(c). To see the convergence holds, first note because

$$E[|E[Y_i(1)|X_i]|^2] \leq E[E[Y_i^2(1)|X_i]] = E[Y_i^2(1)] < \infty ,$$

the weak law of large numbers implies

$$\frac{1}{n} \sum_{1 \leq i \leq 2n} |E[Y_i(1)|X_i]|^2 \xrightarrow{P} 2E[|E[Y_i(1)|X_i]|^2] < \infty .$$

On the other hand,

$$\frac{1}{2n} \sum_{1 \leq i \leq 2n} |E[m_{1,n}(X_i, W_i)|X_i]|^2 \leq \frac{1}{2n} \sum_{1 \leq i \leq 2n} E[m_{1,n}^2(X_i, W_i)|X_i] .$$

Assumption 3.1(b) and Lemma B.1 imply

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} E[E[m_{1,n}^2(X_i, W_i)|X_i] I\{E[m_{1,n}^2(X_i, W_i)|X_i] > \lambda\}] = 0 .$$

Therefore, Lemma 11.4.2 of Lehmann and Romano (2005) implies

$$\frac{1}{2n} \sum_{1 \leq i \leq 2n} E[m_{1,n}^2(X_i, W_i)|X_i] - E[E[m_{1,n}^2(X_i, W_i)|X_i]] \xrightarrow{P} 0 .$$

Finally, note $E[E[m_{1,n}^2(X_i, W_i)|X_i]] = E[m_{1,n}^2(X_i, W_i)]$ is bounded for $n \geq 1$ by Assumption 3.1(b), so

$$\frac{1}{n} \sum_{1 \leq i \leq 2n} |E[m_{1,n}(X_i, W_i)|X_i]|^2 = O_P(1) .$$

The desired convergence therefore follows.

Similar arguments applied termwise imply the right-hand side of (B.7) is $o_P(1)$. (B.6)–(B.7) then imply

$$\frac{s_n^2}{n} - \frac{1}{2n} \sum_{1 \le i \le 2n} \operatorname{Var}[\phi_{1,n,i}|X_i] \to 0 \ . \tag{B.8}$$

Next, we argue

$$\frac{1}{2n} \sum_{1 \le i \le 2n} \operatorname{Var}[\phi_{1,n,i}|X_i] - E[\operatorname{Var}[\phi_{1,n,i}|X_i]] \to 0 \ . \tag{B.9}$$

To establish (B.9), we verify the uniform integrability condition in Lemma 11.4.2 of Lehmann and Romano (2005). To that end, we will repeatedly use the inequality

$$\left| \sum_{1 \le j \le k} a_j \right| I \left\{ \left| \sum_{1 \le j \le k} a_j \right| > \lambda \right\} \le \sum_{1 \le j \le k} k |a_j| I \left\{ |a_j| > \frac{\lambda}{k} \right\} \tag{B.10}$$

$$|ab| I\{|ab| > \lambda\} \le |a|^2 I\{|a| > \sqrt{\lambda}\} + |b|^2 I\{|b| > \sqrt{\lambda}\} \ . \tag{B.11}$$

Note

$$\begin{aligned}
&E[|\operatorname{Var}[\phi_{1,n,i}|X_i] - E[\operatorname{Var}[\phi_{1,n,i}|X_i]]| I\{|\operatorname{Var}[\phi_{1,n,i}|X_i] - E[\operatorname{Var}[\phi_{1,n,i}|X_i]]| > \lambda\}] \\
&\lesssim E\left[ |\operatorname{Var}[\phi_{1,n,i}|X_i]| I\left\{ |\operatorname{Var}[\phi_{1,n,i}|X_i]| > \frac{\lambda}{2} \right\} \right] + E[\operatorname{Var}[\phi_{1,n,i}|X_i]] I\left\{ E[\operatorname{Var}[\phi_{1,n,i}|X_i]] > \frac{\lambda}{2} \right\} \\
&\le E\left[ E[\phi_{1,n,i}^2|X_i] I\left\{ E[\phi_{1,n,i}^2|X_i] > \frac{\lambda}{2} \right\} \right] + E[\phi_{1,n,i}^2] I\left\{ E[\phi_{1,n,i}^2] > \frac{\lambda}{2} \right\} \ ,
\end{aligned}$$

where in the second inequality we use the fact that the variance of a random variable is bounded by its second moment. Note Assumption 3.1 implies $E[\phi_{1,n,i}^2]$ is bounded for $n \ge 1$, and therefore

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} E[\phi_{1,n,i}^2] I\left\{ E[\phi_{1,n,i}^2] > \frac{\lambda}{2} \right\} = 0 \ .$$

On the other hand

$$\begin{aligned}
&E\left[ E[\phi_{1,n,i}^2|X_i] I\left\{ E[\phi_{1,n,i}^2|X_i] > \frac{\lambda}{2} \right\} \right] \qquad \text{(B.12)} \\
&\lesssim E\left[ E[Y_i^2(1)|X_i] I\left\{ E[Y_i^2(1)|X_i] > \frac{\lambda}{12} \right\} \right] + E\left[ E[m_{1,n}^2(X_i, W_i)|X_i] I\left\{ E[m_{1,n}^2(X_i, W_i)|X_i] > \frac{\lambda}{3} \right\} \right] \\
&\quad + E\left[ E[m_{0,n}^2(X_i, W_i)|X_i] I\left\{ E[m_{0,n}^2(X_i, W_i)|X_i] > \frac{\lambda}{3} \right\} \right] \\
&\quad + E\left[ |E[Y_i(1) m_{1,n}(X_i, W_i)|X_i]| I\left\{ |E[Y_i(1) m_{1,n}(X_i, W_i)|X_i]| > \frac{\lambda}{12} \right\} \right] \\
&\quad + E\left[ |E[Y_i(1) m_{0,n}(X_i, W_i)|X_i]| I\left\{ |E[Y_i(1) m_{0,n}(X_i, W_i)|X_i]| > \frac{\lambda}{12} \right\} \right] \\
&\quad + E\left[ |E[m_{1,n}(X_i, W_i) m_{0,n}(X_i, W_i)|X_i]| I\left\{ |E[m_{1,n}(X_i, W_i) m_{0,n}(X_i, W_i)|X_i]| > \frac{\lambda}{6} \right\} \right] \ .
\end{aligned}$$

It follows from Assumptions 2.1(b) and 3.1(b) together with Lemma B.1 that

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty} E\left[E[Y_i^2(1)|X_i] I\left\{E[Y_i^2(1)|X_i] > \frac{\lambda}{12}\right\}\right] = 0$$
$$\lim_{\lambda\to\infty}\limsup_{n\to\infty} E\left[E[m_{1,n}^2(X_i,W_i)|X_i] I\left\{E[m_{1,n}^2(X_i,W_i)|X_i] > \frac{\lambda}{3}\right\}\right] = 0$$
$$\lim_{\lambda\to\infty}\limsup_{n\to\infty} E\left[E[m_{0,n}^2(X_i,W_i)|X_i] I\left\{E[m_{0,n}^2(X_i,W_i)|X_i] > \frac{\lambda}{3}\right\}\right] = 0 \ .$$

For the last term in (B.12), note

$$\begin{aligned} & E\left[|E[m_{1,n}(X_i,W_i)m_{0,n}(X_i,W_i)|X_i]| I\left\{|E[m_{1,n}(X_i,W_i)m_{0,n}(X_i,W_i)|X_i]| > \frac{\lambda}{6}\right\}\right] \\ & \leq E\left[E[|m_{1,n}(X_i,W_i)m_{0,n}(X_i,W_i)||X_i] I\left\{E[|m_{1,n}(X_i,W_i)m_{0,n}(X_i,W_i)||X_i] > \frac{\lambda}{6}\right\}\right] \ . \end{aligned}$$

Meanwhile,

$$\begin{aligned} & E\left[E[|m_{1,n}(X_i,W_i)m_{0,n}(X_i,W_i)| I\left\{|m_{1,n}(X_i,W_i)m_{0,n}(X_i,W_i)| > \lambda\right\}\right] \\ & \leq E[m_{1,n}^2(X_i,W_i) I\{|m_{1,n}(X_i,W_i)| > \sqrt{\lambda}\}] + E[m_{0,n}^2(X_i,W_i) I\{|m_{0,n}(X_i,W_i)| > \sqrt{\lambda}\}] \ . \end{aligned}$$

It then follows from the previous two inqualities, Assumption 3.1(b), and Lemma B.1 that

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty} E\left[|E[m_{1,n}(X_i,W_i)m_{0,n}(X_i,W_i)|X_i]| I\left\{|E[m_{1,n}(X_i,W_i)m_{0,n}(X_i,W_i)|X_i]| > \frac{\lambda}{6}\right\}\right] = 0 \ .$$

Similar arguments establish

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty} E\left[|E[Y_i(1)m_{1,n}(X_i,W_i)|X_i]| I\left\{|E[Y_i(1)m_{1,n}(X_i,W_i)|X_i]| > \frac{\lambda}{12}\right\}\right] = 0$$
$$\lim_{\lambda\to\infty}\limsup_{n\to\infty} E\left[|E[Y_i(1)m_{0,n}(X_i,W_i)|X_i]| I\left\{|E[Y_i(1)m_{0,n}(X_i,W_i)|X_i]| > \frac{\lambda}{12}\right\}\right] = 0 \ .$$

Therefore, (B.9) follows. The conclusion then follows from (B.8)–(B.9) and Assumption 3.1(a). ■

**Lemma B.3.** *Suppose Assumptions 2.1–2.3 and 3.1 hold. Then,*

$$\lim_{\gamma\to\infty}\limsup_{n\to\infty} E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 > \gamma\}] = 0 \ .$$

Proof. Note

$$\begin{aligned} & E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 > \gamma\}] \\ & \lesssim E\left[(\phi_{1,n,i}^2 + E[\phi_{1,n,i}|X_i]^2) I\left\{\phi_{1,n,i}^2 + E[\phi_{1,n,i}|X_i]^2 > \frac{\gamma}{2}\right\}\right] \\ & \lesssim E\left[\phi_{1,n,i}^2 I\left\{\phi_{1,n,i}^2 > \frac{\gamma}{4}\right\}\right] + E\left[E[\phi_{1,n,i}|X_i]^2 I\left\{E[\phi_{1,n,i}|X_i]^2 > \frac{\gamma}{4}\right\}\right] \ . \end{aligned}$$

where the first inequality follows from $(a+b)^2 \leq 2(a^2+b^2)$ and the second inequality follows from (B.10).

Next, note

$$
\begin{aligned}
&E\left[E[\phi_{1,n,i}|X_i]^2 I\left\{E[\phi_{1,n,i}|X_i]^2 > \frac{\gamma}{4}\right\}\right] \\
&\lesssim E\left[E[Y_i(1)|X_i]^2 I\left\{E[Y_i(1)|X_i]^2 > \frac{\gamma}{24}\right\}\right] + E\left[E[m_{1,n}(X_i,W_i)|X_i]^2 I\left\{E[m_{1,n}(X_i,W_i)|X_i]^2 > \frac{\gamma}{6}\right\}\right] \\
&\qquad + E\left[E[m_{0,n}(X_i,W_i)|X_i]^2 I\left\{E[m_{0,n}(X_i,W_i)|X_i]^2 > \frac{\gamma}{6}\right\}\right] \\
&\qquad + E\left[|E[Y_i(1)|X_i]E[m_{1,n}(X_i,W_i)|X_i]| I\left\{|E[Y_i(1)|X_i]E[m_{1,n}(X_i,W_i)|X_i]| > \frac{\gamma}{24}\right\}\right] \\
&\qquad + E\left[|E[Y_i(1)|X_i]E[m_{0,n}(X_i,W_i)|X_i]| I\left\{|E[Y_i(1)|X_i]E[m_{0,n}(X_i,W_i)|X_i]| > \frac{\gamma}{24}\right\}\right] \\
&\qquad + E\left[|E[m_{1,n}(X_i,W_i)|X_i]E[m_{0,n}(X_i,W_i)|X_i]| I\left\{|E[m_{1,n}(X_i,W_i)|X_i]E[m_{0,n}(X_i,W_i)|X_i]| > \frac{\gamma}{12}\right\}\right] \\
&\lesssim E\left[E[Y_i^2(1)|X_i] I\left\{E[Y_i^2(1)|X_i] > \frac{\gamma}{24}\right\}\right] + E\left[E[m_{1,n}^2(X_i,W_i)|X_i] I\left\{E[m_{1,n}^2(X_i,W_i)|X_i] > \frac{\gamma}{6}\right\}\right] \\
&\qquad + E\left[E[m_{0,n}^2(X_i,W_i)|X_i] I\left\{E[m_{0,n}^2(X_i,W_i)|X_i] > \frac{\gamma}{6}\right\}\right] \\
&\qquad + E\left[|E[Y_i(1)|X_i]| I\left\{|E[Y_i(1)|X_i]| > \sqrt{\frac{\gamma}{24}}\right\}\right] \\
&\qquad + E\left[|E[m_{1,n}(X_i,W_i)|X_i]| I\left\{|E[m_{1,n}(X_i,W_i)|X_i]| > \sqrt{\frac{\gamma}{24}}\right\}\right] \\
&\qquad + E\left[|E[m_{0,n}(X_i,W_i)|X_i]| I\left\{|E[m_{0,n}(X_i,W_i)|X_i]| > \sqrt{\frac{\gamma}{24}}\right\}\right] \\
&\qquad + E\left[|E[m_{1,n}(X_i,W_i)|X_i]| I\left\{|E[m_{1,n}(X_i,W_i)|X_i]| > \sqrt{\frac{\gamma}{12}}\right\}\right] \\
&\qquad + E\left[|E[m_{0,n}(X_i,W_i)|X_i]| I\left\{|E[m_{0,n}(X_i,W_i)|X_i]| > \sqrt{\frac{\gamma}{12}}\right\}\right] \\
&\le E\left[E[Y_i^2(1)|X_i] I\left\{E[Y_i^2(1)|X_i] > \frac{\gamma}{24}\right\}\right] + E\left[E[m_{1,n}^2(X_i,W_i)|X_i] I\left\{E[m_{1,n}^2(X_i,W_i)|X_i] > \frac{\gamma}{6}\right\}\right] \\
&\qquad + E\left[E[m_{0,n}^2(X_i,W_i)|X_i] I\left\{E[m_{0,n}^2(X_i,W_i)|X_i] > \frac{\gamma}{6}\right\}\right] \\
&\qquad + E\left[E[Y_i^2(1)|X_i] I\left\{E[Y_i^2(1)|X_i] > \frac{\gamma}{24}\right\}\right] \\
&\qquad + E\left[E[m_{1,n}^2(X_i,W_i)|X_i] I\left\{E[m_{1,n}^2(X_i,W_i)|X_i] > \frac{\gamma}{24}\right\}\right] \\
&\qquad + E\left[E[m_{0,n}^2(X_i,W_i)|X_i] I\left\{E[m_{0,n}^2(X_i,W_i)|X_i] > \sqrt{\frac{\gamma}{24}}\right\}\right] \\
&\qquad + E\left[E[m_{1,n}^2(X_i,W_i)|X_i] I\left\{E[m_{1,n}^2(X_i,W_i)|X_i] > \sqrt{\frac{\gamma}{12}}\right\}\right] \\
&\qquad + E\left[E[m_{0,n}^2(X_i,W_i)|X_i] I\left\{E[m_{0,n}^2(X_i,W_i)|X_i] > \sqrt{\frac{\gamma}{12}}\right\}\right] ,
\end{aligned}
$$

where the first inequality follows from (B.10), the second one follows from the conditional Jensen's inequality and (B.11), and the third one follows again from the conditional Jensen's inequality. It then follows from Lemma B.1 together with Assumptions 2.1(b) and 3.1(b) that

$$
\lim_{\gamma\to\infty}\limsup_{n\to\infty} E\left[E[\phi_{1,n,i}|X_i]^2 I\left\{E[\phi_{1,n,i}|X_i]^2 > \frac{\gamma}{4}\right\}\right] = 0 .
$$

Similar arguments lead to

$$\lim_{\gamma\to\infty}\limsup_{n\to\infty} E\left[\phi_{1,n,i}^2 I\left\{\phi_{1,n,i}^2 > \frac{\gamma}{4}\right\}\right] = 0 \ .$$

The conclusion then follows. ■

**Lemma B.4.** *Suppose Assumptions in Theorem 5.1 hold. Then,*

$$P\left\{\left|\frac{1}{n}\sum_{i\in[2n]} I\{D_i = d\}\epsilon_{n,i}(d) - E[\epsilon_{n,i}(d)]\right| \le \sqrt{\frac{\log(2np_n)}{n}}\right\} \to 1$$

*and*

$$P\left\{\left\|\Omega_n^{-1}(d)\frac{1}{n}\sum_{1\le i\le 2n} I\{D_i = d\}(\psi_{n,i}\epsilon_{n,i}(d) - E[\psi_{n,i}\epsilon_{n,i}(d)])\right\|_\infty \le \frac{6\bar{\sigma}}{\underline{\sigma}}\sqrt{\frac{\log(2np_n)}{n}}\right\} \to 1 \ .$$

Proof. For the first result, we note that

$$\begin{aligned}\left|\frac{1}{n}\sum_{i\in[2n]} I\{D_i = d\}\epsilon_{n,i}(d) - E[\epsilon_{n,i}(d)]\right| &\le \left|\frac{1}{n}\sum_{i\in[2n]} I\{D_i = d\}(\epsilon_{n,i}(d) - E[\epsilon_{n,i}(d)|X_i])\right| \\ &+ \left|\frac{1}{n}\sum_{i\in[2n]} (I\{D_i = d\} - 1/2)(E[\epsilon_{n,i}(d)|X_i] - E[\epsilon_{n,i}(d)])\right| \\ &+ \left|\frac{1}{2n}\sum_{i\in[2n]} (E[\epsilon_{n,i}(d)|X_i] - E[\epsilon_{n,i}(d)])\right| .\end{aligned}$$

The first two terms on the RHS of the above display are $O_P(1/\sqrt{n})$. The last term on the RHS is also $O_P(1/\sqrt{n})$ by Chebyshev's inequality. This implies the desired result.

For the second result, define

$$\mathcal{E}_{n,0}(d) = \begin{pmatrix} \max_{d\in\{0,1\}} \frac{1}{2n}\sum_{1\le i\le 2n} E[\epsilon_{n,i}^4(d)|X_i] \le c_0 < \infty \ , \\ \min_{1\le l\le p_n} \frac{1}{n}\sum_{1\le i\le 2n} I\{D_i = d\}\operatorname{Var}[\psi_{n,i,l}\epsilon_{n,i}(d)|X_i] \ge \underline{\sigma}^2 > 0 \ , \end{pmatrix}$$

$$\mathcal{E}_{n,1}(d) = \left\{\left\|\frac{1}{n}\sum_{1\le i\le 2n} I\{D_i = d\}(\psi_{n,i}\epsilon_{n,i}(d) - E[\psi_{n,i}\epsilon_{n,i}(d)|X_i])\right\|_\infty \le 2.04\bar{\sigma}\sqrt{\log(2np_n)/n}\right\} \ ,$$

$$\mathcal{E}_{n,2}(d) = \left\{\left\|\frac{1}{n}\sum_{1\le i\le 2n} I\{D_i = d\}(E[\psi_{n,i}\epsilon_{n,i}(d)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(d)])\right\|_\infty \le 3.96\bar{\sigma}\sqrt{\log(2np_n)/n}\right\} \ ,$$

$$\mathcal{E}_{n,3}(d) = \left\{\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n} (E[\psi_{n,i,l}^2\epsilon_{n,i}^2(d)|X_i] - E[\psi_{n,i,l}^2\epsilon_{n,i}^2(d)])\right|^{1/2} \le 0.01\bar{\sigma}\right\} \ ,$$

and

$$\mathcal{E}_{n,4}(d) = \left\{ \max_{1\le l\le p_n} \left| \frac{1}{2n} \sum_{1\le i\le 2n} (2I\{D_i = d\} - 1)(E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}|X_i] - E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}]) \right|^{1/2} \le 0.01\bar{\sigma} \right\} .$$

We aim to show that $P\{\mathcal{E}_{n,1}(d)\} \to 1$ and $P\{\mathcal{E}_{n,2}(d)\} \to 1$. Then, by letting $C = 6\bar{\sigma}/\underline{\sigma}$ which implies

$$\begin{aligned}
P\{\mathcal{E}_n(d)\} &= 1 - P\{\mathcal{E}^c_n(d)\} \\
&\ge 1 - P\left\{ \left\| \frac{1}{n} \sum_{1\le i\le 2n} I\{D_i = d\}(\psi_{n,i}\epsilon_{n,i}(d) - E[\psi_{n,i}\epsilon_{n,i}(d)]) \right\|_\infty \ge C\underline{\sigma}\sqrt{\frac{\log(2np_n)}{n}} \right\} \\
&= 1 - P\left\{ \left\| \frac{1}{n} \sum_{1\le i\le 2n} I\{D_i = d\}(\psi_{n,i}\epsilon_{n,i}(d) - E[\psi_{n,i}\epsilon_{n,i}(d)]) \right\|_\infty \ge 6\bar{\sigma}\sqrt{\frac{\log(2np_n)}{n}} \right\} \\
&\ge 1 - P\{\mathcal{E}^c_{n,1}(d)\} - P\{\mathcal{E}^c_{n,2}(d)\} \to 1 .
\end{aligned}$$

First, we show $P\{\mathcal{E}_{n,3}(d)\} \to 1$. Let

$$t_n = C\sqrt{\frac{\log(np_n)\Xi^2_n}{n}} \to 0$$

for some sufficiently large constant $C > 0$ and $\{e_i\}_{1\le i\le 2n}$ be a sequence of i.i.d. Rademacher random variables independent of everything else. Then, for any fixed $t > 0$, we have

$$\begin{aligned}
&\left(1 - \frac{4\max_{1\le l\le p_n} \mathrm{Var}[E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i]]}{2nt^2}\right) P\left\{ \max_{1\le l\le p_n} \left| \frac{1}{2n} \sum_{1\le i\le 2n} \left[E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i] - E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)]\right] \right| \ge t \right\} \\
&\le 2P\left\{ \max_{1\le l\le p_n} \left| \frac{1}{2n} \sum_{1\le i\le 2n} 4e_i E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i] \right| \ge t \right\} \\
&= o(1) + 2E\left[ P\left\{ \max_{1\le l\le p_n} \left| \frac{1}{2n} \sum_{1\le i\le 2n} 4e_i E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i] \right| \ge t \,\middle|\, X^{(n)} \right\} I\{\mathcal{E}_{n,0}(d)\} \right] \\
&\lesssim o(1) + p_n \exp\left(-\frac{nt^2}{\Xi^2_n c}\right) = o(1),
\end{aligned}$$

where the first inequality is by van der Vaart and Wellner (1996, Lemma 2.3.7), the second inequality is by the Hoeffding's inequality conditional on $X^{(n)}$ and the fact that, on $\mathcal{E}_{n,0}(d)$,

$$\begin{aligned}
\frac{1}{2n} \sum_{1\le i\le 2n} (E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i])^2 &\le \frac{1}{2n} \sum_{1\le i\le 2n} E[\psi^4_{n,i,l}|X_i]E[\epsilon^4_{n,i}(d)|X_i] \\
&\le \frac{\Xi^2_n}{2n} \sum_{1\le i\le 2n} E[\psi^2_{n,i,l}|X_i]E[\epsilon^4_{n,i}(d)|X_i] \le \Xi^2_n C c_0,
\end{aligned}$$

where $C$ is a fixed constant, and the last equality is by the fact that $\log(p_n)\Xi^2_n = o(n)$. Furthermore, we note that

$$
\begin{aligned}
&\frac{4\max_{1\le l\le p_n}\mathrm{Var}[E[\psi_{n,i,l}^2\epsilon_{n,i}^2(d)|X_i]]}{2n} \\
&\lesssim \frac{\max_{1\le l\le p_n} E\left[E[\psi_{n,i,l}^4|X_i]E[\epsilon_{n,i}^4(d)|X_i]\right]}{n} \\
&\lesssim \frac{\Xi_n^2\max_{1\le l\le p_n} E\left[E[\psi_{n,i,l}^2|X_i]E[\epsilon_{n,i}^4(d)|X_i]\right]}{n} \\
&\lesssim \frac{\Xi_n^2 E[\epsilon_{n,i}^4(d)]}{n} \\
&= o(1).
\end{aligned}
$$

Therefore, we have

$$
P\left\{\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}\left[E[\psi_{n,i,l}^2\epsilon_{n,i}^2(d)|X_i]-E[\psi_{n,i,l}^2\epsilon_{n,i}^2(d)]\right]\right|\ge t\right\}=o(1)
$$

for any fixed $t>0$, which is the desired result.

Next, we show $P\{\mathcal{E}_{n,4}(d)\}\to 1$. Define $a_{n,i,l}=E[\epsilon_{n,i}^2(d)\psi_{n,i,l}^2|X_i]-E[\epsilon_{n,i}^2(d)\psi_{n,i,l}^2]$. Then, we have

$$
\begin{aligned}
&P\left\{\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}(2I\{D_i=d\}-1)(E[\epsilon_{n,i}^2(d)\psi_{n,i,l}^2|X_i]-E[\epsilon_{n,i}^2(d)\psi_{n,i,l}^2])\right|>t\,\middle|\,X^{(n)}\right\}I\{\mathcal{E}_{n,0}(d)\} \\
&\le \sum_{1\le l\le p_n}P\left\{\left|\frac{1}{2n}\sum_{1\le j\le n}(I\{D_{\pi(2j-1)}=d\}-I\{D_{\pi(2j)}=d\})(a_{n,\pi(2j-1),l}-a_{n,\pi(2j),l})\right|>t\,\middle|\,X^{(n)}\right\}I\{\mathcal{E}_{n,0}(d)\} \\
&\le \sum_{1\le l\le p_n}\exp\left(-\frac{2nt^2}{\frac{1}{n}\sum_{1\le j\le n}(a_{n,\pi(2j-1),l}-a_{n,\pi(2j),l})^2}\right)I\{\mathcal{E}_{n,0}(d)\} \\
&\le \exp\left(\log(p_n)-\frac{2nt^2}{\Xi_n^2c^2}\right)\ ,
\end{aligned}
$$

where, conditional on $X^{(n)}$, $\{I\{D_{\pi(2j-1)}=d\}-I\{D_{\pi(2j)}=d\}\}_{1\le j\le n}$ is a sequence of i.i.d. Rademacher random variables, the second last inequality is by Hoeffding's inequality, and the last inequality is by that, on $\mathcal{E}_{n,0}(d)$,

$$
\begin{aligned}
&\left(\frac{1}{n}\sum_{1\le j\le n}(a_{n,\pi(2j-1),l}-a_{n,\pi(2j),l})^2\right)^{1/2} \\
&\le \left(\frac{1}{n}\sum_{1\le j\le n}(E[\psi_{n,\pi(2j-1),l}^2\epsilon_{n,i}^2(d)|X_{\pi(2j-1)}])^2\right)^{1/2}+\left(\frac{1}{n}\sum_{1\le j\le n}(E[\psi_{n,\pi(2j),l}^2\epsilon_{n,i}^2(d)|X_{\pi(2j)}])^2\right)^{1/2} \\
&\le \left(\frac{2}{n}\sum_{1\le i\le 2n}(E[\psi_{n,i,l}^2\epsilon_{n,i}^2(d)|X_i])^2\right)^{1/2} \\
&\le \Xi_n c\ .
\end{aligned}
$$

By letting $t = C\sqrt{\frac{\log(p_n)\Xi_n^2}{n}}$ for some sufficiently large $C$ and noting that $P\{\mathcal{E}_{n,0}(d)\} \to 1$, we have

$$\max_{1\le l\le p_n}\left|\sum_{1\le i\le 2n}\frac{(2I\{D_i=d\}-1)(E[\epsilon_{n,i}^2(d)\psi_{n,i,l}^2|X_i]-E[\epsilon_{n,i}^2(d)\psi_{n,i,l}^2])}{2n}\right| = O_p\left(\sqrt{\frac{\log(p_n)\Xi_n^2}{n}}\right) ,$$

and thus, $P\{\mathcal{E}_{n,4}(d)\} \to 1$.

Next, we show $P\{\mathcal{E}_{n,1}(d)\} \to 1$. We note that, for $d \in \{0,1\}$, conditional on $(D^{(n)}, X^{(n)})$, $\{\psi_{n,i}\epsilon_{n,i}(d)\}_{1\le i\le 2n}$ are independent. In what follows, we couple

$$\mathbb{U}_n = \frac{1}{n}\sum_{1\le i\le 2n} I\{D_i=d\}(\psi_{n,i}\epsilon_{n,i}(d) - E[\psi_{n,i}\epsilon_{n,i}(d)|X_i])$$

with a centered Gaussian random vector as in Theorem 2.1 in Chernozhukov et al. (2017). Let $Z = (Z_1,\dots,Z_{p_n})$ be a Gaussian random vector with $E[Z_l]=0$ for $1\le l\le p_n$ and $\text{Var}[Z] = \text{Var}[\mathbb{U}_n|X^{(n)},D^{(n)}]$ that additionally satisfies the conditions of that theorem. Specifically, $Z = (Z_1,\cdots,Z_{p_n})$ is a centered Gaussian random vector in $R^{p_n}$ such that on $\mathcal{E}_{n,0}(d)\cap\mathcal{E}_{n,3}(d)\cap\mathcal{E}_{n,4}(d)$,

$$\begin{aligned}
E[ZZ'] &= \frac{1}{n^2}\sum_{1\le i\le 2n} I\{D_i=d\}E[\epsilon_{n,i}^2(d)\psi_{n,i}\psi_{n,i}'|X_i] \\
&\quad -\frac{1}{n}\left(\frac{1}{n}\sum_{1\le i\le 2n} I\{D_i=d\}E[\epsilon_{n,i}(d)\psi_{n,i}|X_i]\right)\left(\frac{1}{n}\sum_{1\le i\le 2n} I\{D_i=d\}E[\epsilon_{n,i}(d)\psi_{n,i}|X_i]\right)'
\end{aligned}$$

and

$$\begin{aligned}
\max_{1\le l\le p_n} E[Z_l^2] &\le \frac{\max_{1\le l\le p_n}\sum_{1\le i\le 2n} I\{D_i=d\}E[\epsilon_{n,i}^2(d)\psi_{n,i,l}^2|X_i]}{n^2} \\
&\le \frac{\overline{\sigma}^2}{n} + \frac{\max_{1\le l\le p_n}\sum_{1\le i\le 2n} I\{D_i=d\}(E[\epsilon_{n,i}^2(d)\psi_{n,i,l}^2|X_i]-E[\epsilon_{n,i}^2(d)\psi_{n,i,l}^2])}{n^2} \\
&\le \frac{\overline{\sigma}^2}{n} + \frac{\max_{1\le l\le p_n}\sum_{1\le i\le 2n}(2I\{D_i=d\}-1)(E[\epsilon_{n,i}^2(d)\psi_{n,i,l}^2|X_i]-E[\epsilon_{n,i}^2(d)\psi_{n,i,l}^2])}{2n^2} \\
&\quad + \frac{\max_{1\le l\le p_n}\sum_{1\le i\le 2n}(E[\epsilon_{n,i}^2(d)\psi_{n,i,l}^2|X_i]-E[\epsilon_{n,i}^2(d)\psi_{n,i,l}^2])}{2n^2} \\
&\le \frac{1.02\overline{\sigma}^2}{n} .
\end{aligned}$$

Further define $q(1-\alpha)$ as the $(1-\alpha)$ quantile of $||Z||_\infty$. Then, we have

$$q(1-1/n) \le \frac{1.02\overline{\sigma}(\sqrt{2\log(2p_n)}+\sqrt{2\log(n)})}{\sqrt{n}} \le 2.04\overline{\sigma}\sqrt{\log(2np_n)/n},$$

where the first inequality is by the last display in the proof of Lemma E.2 in Chetverikov and Sørensen (2022) and the second inequality is by the fact that $\sqrt{a}+\sqrt{b} \le \sqrt{2(a+b)}$ for $a,b>0$. Therefore, we have

$$\begin{aligned}
P\{\mathcal{E}_{n,1}^c(d)) &\le P\{\mathcal{E}_{n,1}^c(d), \mathcal{E}_{n,0}(d), \mathcal{E}_{n,3}(d), \mathcal{E}_{n,4}(d)\} + o(1) \\
&= EP\{\mathcal{E}_{n,1}^c(d)|D^{(n)},X^{(n)}\}I\{\mathcal{E}_{n,0}(d),\mathcal{E}_{n,3}(d),\mathcal{E}_{n,4}(d)\} + o(1)
\end{aligned}$$

$$\leq E[P\{||Z||_\infty \geq 2.04\bar{\sigma}\sqrt{\log(2np_n)/n}|D^{(n)}, X^{(n)})\}I\{\mathcal{E}_{n,0}(d), \mathcal{E}_{n,3}(d), \mathcal{E}_{n,4}(d)\}] + o(1)$$
$$\leq E[P\{||Z||_\infty \geq q(1-1/n)|D^{(n)}, X^{(n)}\}] = o(1),$$

where the second inequality is by Theorem 2.1 in Chernozhukov et al. (2017).

Finally, we turn to $\mathcal{E}_{n,2}(d)$ with $d = 1$. We have

$$\begin{aligned}
&\frac{1}{n}\sum_{1\leq i\leq 2n} I\{D_i = 1\}(E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)]) \\
&= \frac{1}{2n}\sum_{1\leq i\leq 2n}(E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)]) + \frac{1}{2n}\sum_{1\leq i\leq 2n}(2D_i - 1)(E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)]).
\end{aligned} \tag{B.13}$$

Note $\{E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)]\}_{1\leq i\leq 2n}$ is a sequence of independent centered random variables and

$$\max_{1\leq l\leq p_n} E[(E[\psi_{n,i,l}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i,l}\epsilon_{n,i}(1)])^2] \leq \bar{\sigma}^2.$$

Following Theorem 2.1 in Chernozhukov et al. (2017), Lemma E.2 in Chetverikov and Sørensen (2022), and similar arguments to the ones above, we have

$$P\left\{\left\|\frac{1}{2n}\sum_{1\leq i\leq 2n}(E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)])\right\|_\infty \leq \bar{\sigma}\sqrt{2\log(2np_n)/n}\right) \to 1. \tag{B.14}$$

For the second term on the RHS of (B.13), we define $g_{n,i,l} = E[\psi_{n,i,l}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i,l}\epsilon_{n,i}(1)]$. We have

$$\begin{aligned}
&P\left\{\left\|\frac{1}{2n}\sum_{1\leq i\leq 2n}(2D_i - 1)E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)]\right\|_\infty > t\,\middle|\, X^{(n)}\right\} \\
&\leq \sum_{1\leq l\leq p_n} P\left\{\left|\frac{1}{2n}\sum_{1\leq j\leq n}(D_{\pi(2j-1)} - D_{\pi(2j)})(g_{n,\pi(2j-1),l} - g_{n,\pi(2j),l})\right| > t\,\middle|\, X^{(n)}\right\} \\
&\leq \sum_{1\leq l\leq p_n} \exp\left(-\frac{2nt^2}{\frac{1}{n}\sum_{1\leq j\leq n}(g_{n,\pi(2j-1),l} - g_{n,\pi(2j),l})^2}\right),
\end{aligned}$$

where, conditional on $X^{(n)}$, $\{(D_{\pi(2j-1)} - D_{\pi(2j)})\}_{1\leq j\leq n}$ is a sequence of i.i.d. Rademacher random variables and the last inequality is by Hoeffding's inequality. In addition, on $\mathcal{E}_{n,3}(1)$, we have

$$\begin{aligned}
&\left(\frac{1}{n}\sum_{1\leq j\leq n}(g_{n,\pi(2j-1),l} - g_{n,\pi(2j),l})^2\right)^{1/2} \\
&\leq \left(\frac{1}{n}\sum_{1\leq j\leq n}(E[\psi_{n,\pi(2j-1),l}\epsilon_{n,i}(1)|X_{\pi(2j-1)}])^2\right)^{1/2} + \left(\frac{1}{n}\sum_{1\leq j\leq n}(E[\psi_{n,\pi(2j),l}\epsilon_{n,i}(1)|X_{\pi(2j)}])^2\right)^{1/2} \\
&\leq \left(\frac{2}{n}\sum_{1\leq i\leq 2n}(E[\psi_{n,i,l}\epsilon_{n,i}(1)|X_i])^2\right)^{1/2}
\end{aligned}$$

$$
\leq \left( \frac{2}{n} \sum_{1 \leq i \leq 2n} E[\psi_{n,i,l}^2 \epsilon_{n,i}^2(1)|X_i] \right)^{1/2}
$$

$$
\leq \left( \frac{2}{n} \sum_{1 \leq i \leq 2n} \left[ E[\psi_{n,i,l}^2 \epsilon_{n,i}^2(1)|X_i] - E[\psi_{n,i,l}^2 \epsilon_{n,i}^2(1)] \right] \right)^{1/2} + 2\overline{\sigma}
$$

$$
\leq 2.02\overline{\sigma}.
$$

Therefore, we have

$$
\begin{aligned}
& P\left\{ \left\| \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2D_i - 1)E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)] \right\|_\infty > 2.02\sqrt{\frac{\log(np_n)\overline{\sigma}^2}{n}} \right\} \\
& \leq P\left\{ \left\| \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2D_i - 1)E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)] \right\|_\infty > 2.02\sqrt{\frac{\log(np_n)\overline{\sigma}^2}{n}}, \mathcal{E}_{n,3}(1) \right\} + o(1) \\
& \leq E\left[ P\left\{ \left\| \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2D_i - 1)E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)] \right\|_\infty > 2.02\sqrt{\frac{\log(np_n)\overline{\sigma}^2}{n}} \middle| X^{(n)} \right\} I\{\mathcal{E}_{n,3}(1)\} \right] + o(1) \\
& = o(1) \ . \qquad \text{(B.15)}
\end{aligned}
$$

Combining (B.13), (B.14), (B.15), and the fact that $\sqrt{2} + 2.02 \leq 3.98$, we have $P\{\mathcal{E}_{n,2}(1)\} \to 1$. The same result holds for $\mathcal{E}_{n,2}(0)$. ■

# C Details for Simulations

The regressors in the LASSO-based adjustment are as follows.

(i) For Models 1-6, we use $\{1, X_i, W_i, X_i^2, W_i^2, X_iW_i, (X_i - \tilde{X})I\{X_i > \tilde{X}\}, (W_i - \tilde{W})I\{W_i > \tilde{W}\}, (X_i - \tilde{X})^2 I\{X_i > \tilde{X}\}, (W_i - \tilde{W})^2 I\{W_i > \tilde{W}\}\}$ where $\tilde{X}$ and $\tilde{W}$ are the sample medians of $\{X_i\}_{i\in[2n]}$ and $\{W_i\}_{i\in[2n]}$, respectively.

(ii) For Models 7-9, we use $\{1, X_i, W_i, X_i^2, W_i^2, X_{i1}W_{i1}, X_{i2}W_{i1}, X_{i1}W_{i2}, X_{i2}W_{i2}, (X_{ij} - \tilde{X}_j)I\{X_{ij} > \tilde{X}_j\}, (X_{ij} - \tilde{X}_j)^2 I\{X_{ij} > \tilde{X}_j\}, (W_{ij} - \tilde{W}_1)I\{W_{ij} > \tilde{W}_j\}, (W_{ij} - \tilde{W}_j)^2 I\{W_{ij} > \tilde{W}_j\}\}$ where $\tilde{X}_j$ and $\tilde{W}_j$, for $j = 1, 2$, are the sample medians of $\{X_{ij}\}_{i\in[2n]}$ and $\{W_{ij}\}_{i\in[2n]}$, respectively.

(iii) For Models 10-11, we use $\{1, X_i, W_i, X_i^2, W_i^2, X_{i1}W_{i1}, X_{i2}W_{i2}, X_{i3}W_{i1}, X_{i4}W_{i2}, (X_{ij} - \tilde{X}_j)I\{X_{ij} > \tilde{X}_j\}, (X_{ij} - \tilde{X}_j)^2 I\{X_{ij} > \tilde{X}_j\}, (W_{ij} - \tilde{W}_j)I\{W_{ij} > \tilde{W}_j\}, (W_{ij} - \tilde{W}_j)^2 I\{W_{ij} > \tilde{W}_j\}\}$ where $\tilde{X}_j$,for $j = 1, 2, 3, 4$, and $\tilde{W}_j$, for $j = 1, 2$, are the sample medians of $\{X_{ij}\}_{i\in[2n]}$ and $\{W_{ij}\}_{i\in[2n]}$, respectively.

(iv) Models 12-15 already contain high-dimensional covariates. We just use $X_i$ and $W_i$ as the LASSO regressors.

## D Details for Empirical Application

"GM" corresponds to the method used in Groh and McKenzie (2016). Groh and McKenzie (2016) estimated the effect by regression with regressors including some baseline variables, a dummy for missing observations, and dummies for the pairs. Specifically, for profits and revenues, the regressors are the baseline value for the outcome of interest, a dummy for missing observations, and pair dummies; for investment, the regressors only include pair dummies. The standard errors for the "GM" ATE estimate are calculated by the usual heteroskedastity-consistent estimator. The "GM" results in Table 5 were obtained by applying the Stata code provided by Groh and McKenzie (2016).

The description of other methods is similar to that in Section 6.2. Specifically:

(i) $X_i$ includes gender and 13 additional matching variables for all adjustments. Three of the matching variables are continuous, and the others are dummies.

(ii) To maintain comparability, we keep $X_i$ and $W_i$ consistent across all adjustments except for "refit" for each outcome variable. For profits and revenue, $W_i$ includes the baseline value for the outcome of interest, a dummy for whether the firm is above the 95th percentile of the control firms' distributions of the outcome variable, and a dummy for missing observations. For investment, $W_i$ includes all the covariates used for the first two outcome variables.

(iii) For "refit", we intentionally expand the dimensions of $W_i$. In addition to the baseline values used in the other adjustments and the dummy variables for missing observations, the $W_i$ used in "refit" also includes the interaction of the continuous original $W_i$ variables with three continuous variables and the first three discrete variables in $X_i$.

(iv) All the continuous variables in $X_i$ and $W_i$ are standardized initially when the regression-adjusted estimators are employed.

## References


Bai, Y., Romano, J. P. and Shaikh, A. M. (2022). Inference in Experiments With Matched Pairs. *Journal of the American Statistical Association*, **117** 1726–1737. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2021.1883437, URL https://doi.org/10.1080/01621459.2021.1883437.

Bickel, P. J., Ritov, Y. and Tsybakov, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, **37** 1705–1732. Publisher: Institute of Mathematical Statistics, URL https://projecteuclid.org/journals/annals-of-statistics/volume-37/issue-4/Simultaneous-analysis-of-Lasso-and-Dantzig-selector/10.1214/08-AOS620.full.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics*, **42** 1564–1597.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, **45** 2309–2352. Publisher: Institute of Mathematical Statistics, URL https://projecteuclid.org/journals/annals-of-probability/volume-45/issue-4/Central-limit-theorems-and-bootstrap-in-high-dimensions/10.1214/16-AOP1113.full.

Chetverikov, D. and Sørensen, J. R.-V. (2022). Analytic and Bootstrap-after-Cross-Validation Methods for Selecting Penalty Parameters of High-Dimensional M-Estimators. Tech. Rep. arXiv:2104.04716, arXiv. ArXiv:2104.04716 [econ, math, stat] type: article, URL http://arxiv.org/abs/2104.04716.

Dudley, R. M. (1989). *Real Analysis and Probability*. Wadsworth and Brook/Cole.

Groh, M. and McKenzie, D. (2016). Macroinsurance for microenterprises: A randomized experiment in post-revolution egypt. *Journal of Development Economics*, **118** 13–25.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Classics in Mathematics, Springer-Verlag, Berlin Heidelberg. URL https://www.springer.com/gp/book/9783642202117.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. 3rd ed. Springer, New York.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer-Verlag, New York.