Supplemental Appendix: For Online Publication

A Sufficient Conditions for Assumptions 3.2 and 3.5

We only lay out the argument for Assumption 3.2 and an identical argument applies to Assumption 3.5. Let $k_x = \dim(X_g)$. Note

$$\frac{1}{G}\sum_{1\le j\le G}\|X_{\pi(2j)}-X_{\pi(2j-1)}\|^r \le \Big(1\vee \max_{1\le g\le 2G}\|X_g\|^r\Big)\frac{1}{G}\sum_{1\le j\le G}\left\|\frac{X_{\pi(2j)}-X_{\pi(2j-1)}}{1\vee\max_{1\le g\le 2G}\|X_g\|}\right\|^r . \quad (15)$$

Consider a non-bipartite matching algorithm that minimizes the left-hand side of (15) for $r=2$ for Assumption 3.2 (or $r=4$ for Assumption 3.5). Because

$$X_g/\max_{1\le g\le 2G}\|X_g\|\in[0,1]^{k_x} ,$$

to study

$$\frac{1}{G}\sum_{1\le j\le G}\left\|\frac{X_{\pi(2j)}-X_{\pi(2j-1)}}{1\vee\max_{1\le g\le 2G}\|X_g\|}\right\|^r , \quad (16)$$

we can assume without loss of generality that $X_g\in[0,1]^{k_x}$ for $1\le g\le 2G$. Consider as an auxiliary proof device the block-path algorithm in the proof of Theorem 4.2 in Bai et al. (2022) with blocks of side lengths $1/m$. Using the inequality $c^r\le c$ if $r\ge 1$ and $c\in[0,1]$, note if $x_1,x_2\in[0,1]^{k_x}$, then

$$\|x_1-x_2\|^r = k_x^{2/r}(\|x_1-x_2\|/\sqrt{k_x})^r \le k_x^{2/r}\|x_1-x_2\|/\sqrt{k_x} = k_x^{2/r-1/2}\|x_1-x_2\| .$$

Therefore, following the proof of Theorem 4.2 in Bai et al. (2022) or Lemma A.1 in Cytrynbaum (2021),

$$\frac{1}{n}\sum_{1\le j\le G}\|X_{\pi(2j)}-X_{\pi(2j-1)}\|^r \le \left(\frac{\sqrt{k_x}}{m}\right)^r + \frac{2}{n}k_x^{2/r}m^{k_x-1} .$$

Taking $m\asymp n^{1/(r+k_x-1)}$, (16) is of order $n^{-r/(r+k_x-1)}$. On the other hand, if $E[\|X_g\|^d]<\infty$, Lemma S.1.1 in Bai et al. (2022) implies $\max_{1\le g\le 2G}\|X_g\|^r = o_P(n^{r/d})$. Therefore, as long as $d\ge k_x$, the left-hand side of (15) converges to zero in probability.

Note further that, when verifying Assumption 3.5, if $\|W_g\|$ is bounded, then

$$\frac{1}{G}\sum_{1\le j\le G}\|W_{\pi(2j)}-W_{\pi(2j-1)}\|^4 \lesssim \frac{1}{G}\sum_{1\le j\le G}\|W_{\pi(2j)}-W_{\pi(2j-1)}\|^2 ,$$

and therefore any algorithm that minimizes the right-hand of the above display will satisfy Assumption 3.5.

B Proofs of Main Results

Please note that in what follows we will use the notation $a \lesssim b$ to denote $a \le cb$ for some constant c.

B.1 Proof of Theorem 3.1

Proof. We have that

$$\hat{\Delta}_G = \frac{\frac{1}{G}\sum_{1\le g\le 2G}\bar{Y}_g(1)N_gD_g}{\frac{1}{G}\sum_{1\le g\le 2G}N_gD_g} - \frac{\frac{1}{G}\sum_{1\le g\le 2G}\bar{Y}_g(0)N_g(1-D_g)}{\frac{1}{G}\sum_{1\le g\le 2G}N_g(1-D_g)} .$$

In particular, for $h(x,y,z,w) = \frac{x}{y} - \frac{z}{w}$, observe that

$$\hat{\Delta}_G = h\left(\frac{1}{G}\sum_{1\le g\le 2G}\bar{Y}_g(1)N_gD_g, \frac{1}{G}\sum_{1\le g\le 2G}N_gD_g, \frac{1}{G}\sum_{1\le g\le 2G}\bar{Y}_g(0)N_g(1-D_g), \frac{1}{G}\sum_{1\le g\le 2G}N_g(1-D_g)\right) ,$$

and by Assumption 3.1,

$$\Delta = h\Bigg(\frac{1}{G}\sum_{1\le g\le 2G}E[\bar{Y}_g(1)N_g]D_g, \frac{1}{G}\sum_{1\le g\le 2G}E[N_g]D_g,$$
$$\frac{1}{G}\sum_{1\le g\le 2G}E[\bar{Y}_g(0)N_g](1-D_g), \frac{1}{G}\sum_{1\le g\le 2G}E[N_g](1-D_g)\Bigg) .$$

The Jacobian of $h(\cdot)$ is

$$D_h(x,y,z,w) = \begin{pmatrix}\frac{1}{y} & -\frac{x}{y^2} & -\frac{1}{w} & \frac{z}{w^2}\end{pmatrix} .$$

By Lemma C.1 and the Delta method,

$$\sqrt{G}(\hat{\Delta}_G - \Delta) \xrightarrow{d} N(0, D_{h0}\mathbb{V}D_{h0}') ,$$

where

$$D_{h0} = \begin{pmatrix}\frac{1}{E[N_g]} & -\frac{E[\bar{Y}_g(1)N_g]}{E[N_g]^2} & -\frac{1}{E[N_g]} & \frac{E[\bar{Y}_g(0)N_g]}{E[N_g]^2}\end{pmatrix}$$

and $\mathbb{V}$ is defined in Lemma C.1. It then follows from Lemma C.2 that

$$D_{h0}\mathbb{V}D_{h0}' = \omega^2 ,$$

as desired. ■

B.2 Proof of Theorem 3.2

Proof. This proof follows from an identical argument to Theorem 3.1, but this time invoking Lemmas C.3 and C.4. ■

B.3 Proof of Theorem 3.3

The desired conclusion follows immediately from Lemmas C.5-C.7 and the continuous mapping theorem. ■

B.4 Proof of Theorem 3.4

By the first result in Theorem 3.6 in Bugni et al. (2024),

$$\hat{\omega}^2_{\mathrm{CR,G}} = \frac{1}{2}\left(\hat{\omega}^2_{\mathrm{CR,G}}(1) + \hat{\omega}^2_{\mathrm{CR,G}}(0)\right) , \tag{17}$$

(where we note that the factor of 1/2 appears since we are normalizing by the number of *pairs*), and

$$\hat{\omega}^2_{\mathrm{CR,G}}(d) := \frac{1}{\left(\frac{1}{2G}\sum_{1\le g\le 2G} N_g I\{D_g = d\}\right)^2} \frac{1}{2G} \sum_{1\le g\le 2G} \left[\left(\frac{N_g}{|\mathcal{M}_g|}\right)^2 I\{D_g = d\} \left(\sum_{i\in\mathcal{M}_g} \hat{\epsilon}_{i,g}(d)\right)^2\right] ,$$

with

$$\hat{\epsilon}_{i,g}(d) := Y_{i,g} - \frac{1}{\sum_{1\le g\le 2G} N_g I\{D_g = d\}} \sum_{1\le g\le 2G} N_g \bar{Y}_g I\{D_g = d\} .$$

Fix $d \in \{0,1\}$, $r \in \{0,1,2\}$, $\ell \in \{1,2\}$ arbitrarily. Then by Lemmas C.12 and C.15,

$$\frac{1}{2G}\sum_{1\le g\le 2G} N_g^\ell \bar{Y}_g^r(d) I\{D_g = d\} \xrightarrow{P} \frac{E[N^l \bar{Y}_g^r(d)]}{2} .$$

The result then follows from additional algebra and repeated applications of the continuous mapping theorem; an identical derivation appears as the second result in Theorem 3.6 of Bugni et al. (2024). ■

B.5 Proof of Theorem 3.5

Let $\mathbf{1}_K$ denote a column of ones of length K. Then consider the following cluster-robust variance estimator where clusters are defined at the level of the *pair*:

$$\left(\frac{1}{G}\sum_{1\le j\le G}\sum_{g\in\lambda_j} X_g' X_g\right)^{-1}\left(\frac{1}{G}\sum_{1\le j\le G}\left(\sum_{g\in\lambda_j} X_g'\hat{\epsilon}_g\right)\left(\sum_{g\in\lambda_j} X_g'\hat{\epsilon}_g\right)'\right)\left(\frac{1}{G}\sum_{1\le g\le G}\sum_{g\in\lambda_j} X_g' X_g\right)^{-1}, \tag{18}$$

where $\lambda_j := \{\pi(2j-1), \pi(2j)\}$, and

$$\begin{aligned} X_g &:= \left(\mathbf{1}_{|\mathcal{M}_g|}\cdot\sqrt{\tfrac{N_g}{|\mathcal{M}_g|}}, \qquad \mathbf{1}_{|\mathcal{M}_g|}\cdot\sqrt{\tfrac{N_g}{|\mathcal{M}_g|}} D_g \right) \\ \hat{\epsilon}_g &:= \sqrt{\frac{N_g}{|\mathcal{M}_g|}}\left(Y_{i,g} - (\hat{\mu}_G(1) - \hat{\mu}_G(0))D_g - \hat{\mu}_G(0) \;:\; i\in\mathcal{M}_g\right)' . \end{aligned}$$

Imposing the condition that $N_g = n$ are equal and fixed and $|\mathcal{M}_g| = N_g$, and then following the algebra in, for instance, the proof of Theorem 3.4 in Bai et al. (2024c), it can be shown that

$$\hat{\omega}^2_{\text{PCVE,G}} = \frac{1}{G}\sum_{1\le j\le G}\left(\sum_{g\in\lambda_j}\bar{Y}_g I\{D_g = 1\} - \sum_{g\in\lambda_j}\bar{Y}_g I\{D_g = 0\}\right)^2 - (\hat{\mu}_G(1) - \hat{\mu}_G(0))^2 .$$

By some additional algebra and repeated applications of Lemmas C.15, C.16, and the continuous mapping theorem we thus obtain that

$$\begin{aligned}\hat{\omega}^2_{\text{PCVE,G}} \xrightarrow{P} & E[\text{Var}[\bar{Y}_g(1)|X_g]] + E[\text{Var}[\bar{Y}_g(1)|X_g]] \\ & + E[\big((E[\bar{Y}_g(1)|X_g] - E[\bar{Y}_g(1)]) - (E[\bar{Y}_g(0)|X_g] - E[\bar{Y}_g(0)])\big)^2] .\end{aligned}$$

Simplifying using the law of total variance and the fact that $\tilde{Y}_g(d) = \bar{Y}_g(d) - E[\bar{Y}_g(d)]$ once we impose that $N_g = n$, we then obtain

$$\hat{\omega}^2_{\text{PCVE,G}} \xrightarrow{P} E[\tilde{Y}_g^2(1)] + E[\tilde{Y}_g^2(0)] - \frac{1}{2}E[(E[\tilde{Y}_g(1) + \tilde{Y}_g(0)|X_g])^2] + \frac{1}{2}E\left[(E[\tilde{Y}_g(1) - \tilde{Y}_g(0)|X_g])^2\right] .$$

The conclusion then follows. ■

B.6 Proof of Theorem 3.6

Proof. Note that the null hypothesis (9) combined with Assumption 2.1(e) implies that

$$\bar{Y}_g(1)|(X_g, N_g) \stackrel{d}{=} \bar{Y}_g(0)|(X_g, N_g) . \tag{19}$$

If the assignment mechanism satisfies Assumption 3.4, the result then follows by applying Theorem 3.4 in Bai et al. (2022) to the cluster-level outcomes $\{(\bar{Y}_g, D_g, X_g, N_g) : 1 \le g \le 2G\}$. If instead the assignment mechanism satisfies Assumption 3.1, then note that (19) is in fact equivalent to the statement

$$(\bar{Y}_g(1), N_g)|X_g \stackrel{d}{=} (\bar{Y}_g(0), N_g)|X_g . \tag{20}$$

The result then follows by applying Theorem 3.4 in Bai et al. (2022) using (20) as the null hypothesis. To establish this equivalence, we first begin with (19) and verify that for any Borel sets A and B,

$$P\{\bar{Y}_g(1) \in A, N_g \in B|X_g\} = P\{\bar{Y}_g(0) \in A, N_g \in B|X_g\} \text{ a.s.}$$

By the definition of a conditional expectation, note we only need to verify for all Borel sets C,

$$E[P\{\bar{Y}_g(1) \in A, N_g \in B|X_g\}I\{X_g \in C\}] = P\{\bar{Y}_g(0) \in A, N_g \in B, X_g \in C\} .$$

We have

$$
\begin{aligned}
&E[P\{\bar{Y}_g(1) \in A, N_g \in B | X_g\} I\{X_g \in C\}] \\
&= P\{\bar{Y}_g(1) \in A, N_g \in B, X_g \in C\} \\
&= E[P\{\bar{Y}_g(1) \in A | X_g, N_g\} I\{N_g \in B\} I\{X_g \in C\}] \\
&= E[P\{\bar{Y}_g(0) \in A | X_g, N_g\} I\{N_g \in B\} I\{X_g \in C\}] \\
&= P\{\bar{Y}_g(0) \in A, N_g \in B, X_g \in C\} \ ,
\end{aligned}
$$

where the first and second equalities follow from the definition of conditional expectations, the the third follows from (19), and the last follows again from the definition of a conditional expectation. The opposite implication follows from a similar argument and is thus omitted. ■

B.7 Proof of Theorem 3.7

Note that

$$
\begin{aligned}
\sqrt{G}\hat{\Delta}_G =& \sqrt{G}\left(\frac{1}{N(1)}\sum_{1\le g\le 2G} D_g N_g \bar{Y}_g - \frac{1}{N(0)}\sum_{1\le g\le 2G}(1-D_g)N_g\bar{Y}_g\right) \\
&= \frac{1}{N(1)}\sqrt{G}\sum_{1\le g\le 2G}\left(D_g N_g\bar{Y}_g - (1-D_g)N_g\bar{Y}_g\right) + \left(\frac{1}{N(1)} - \frac{1}{N(0)}\right)\sqrt{G}\sum_{1\le g\le 2G}(1-D_g)N_g\bar{Y}_g \\
&= \frac{1}{N(1)/G}\frac{1}{\sqrt{G}}\sum_{1\le j\le G}\left(N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)}\right)\left(D_{\pi(2j)} - D_{\pi(2j-1)}\right) \\
&\qquad + \frac{\frac{1}{\sqrt{G}}(N(0)-N(1))}{\frac{N(1)}{G}\frac{N(0)}{G}}\frac{1}{G}\sum_{1\le g\le 2G}(1-D_g)N_g\bar{Y}_g \\
&= \frac{1}{N(1)/G}\frac{1}{\sqrt{G}}\sum_{1\le j\le G}\left(N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)}\right)\left(D_{\pi(2j)} - D_{\pi(2j-1)}\right) \\
&\qquad - \frac{\frac{1}{\sqrt{G}}\sum_{1\le j\le G}(N_{\pi(2j)} - N_{\pi(2j-1)})(D_{\pi(2j)} - D_{\pi(2j-1)})}{\frac{N(1)}{G}\frac{N(0)}{G}}\frac{1}{G}\sum_{1\le g\le 2G}(1-D_g)N_g\bar{Y}_g \ .
\end{aligned}
$$

Hence the randomization distribution of $\sqrt{G}\hat{\Delta}_G$ is given by

$$
\tilde{R}_G(t) := P\left\{\sqrt{G}\check{\Delta}(\epsilon_1,\dots,\epsilon_G) \le t \,\middle|\, Z^{(G)}\right\} \ , \tag{21}
$$

where

$$
\begin{aligned}
\sqrt{G}\check{\Delta}(\epsilon_1,\dots,\epsilon_G) = & \frac{1}{\tilde{N}(1)/G}\frac{1}{\sqrt{G}}\sum_{1\le j\le G}\epsilon_j\left(N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)}\right)\left(D_{\pi(2j)} - D_{\pi(2j-1)}\right) \\
&- \frac{\frac{1}{\sqrt{G}}\sum_{1\le j\le G}\epsilon_j(N_{\pi(2j)} - N_{\pi(2j-1)})(D_{\pi(2j)} - D_{\pi(2j-1)})}{\frac{\tilde{N}(1)}{G}\frac{\tilde{N}(0)}{G}}\frac{1}{G}\sum_{1\le g\le 2G}(1-\tilde{D}_g)N_g\bar{Y}_g \ ,
\end{aligned}
$$

ϵ_j, $j = 1, \ldots, G$ are i.i.d. Rademacher random variables generated independently of $Z^{(G)}$, $\{\tilde{D}_g : 1 \leq g \leq 2G\}$ denotes the assignment of cluster g after applying the transformation implied by $\{\epsilon_j : 1 \leq j \leq G\}$, and

$$\tilde{N}(d) = \sum_{1 \leq g \leq 2G} N_g I\{\tilde{D}_g = d\} .$$

By construction, $\hat{v}_G^2$ evaluated at the transformation of the data implied by $\{\epsilon_j : 1 \leq j \leq G\}$ is given by

$$\check{v}_G^2(\epsilon_1, \ldots, \epsilon_G) = \hat{\tau}_G^2 - \frac{1}{2}\check{\lambda}_G^2(\epsilon_1, \ldots, \epsilon_G) \tag{22}$$

where $\hat{\tau}_G^2$ is defined in (6), and

$$\check{\lambda}_G^2(\epsilon_1, \ldots, \epsilon_G) = \frac{2}{G} \sum_{1 \leq j \leq \lfloor G/2 \rfloor} \epsilon_{2j-1}\epsilon_{2j} \left(\hat{Y}_{\pi(4j-3)} - \hat{Y}_{\pi(4j-2)}\right)\left(\hat{Y}_{\pi(4j-1)} - \hat{Y}_{\pi(4j)}\right) \\ \times \left(D_{\pi(4j-3)} - D_{\pi(4j-2)}\right)\left(D_{\pi(4j-1)} - D_{\pi(4j)}\right) .$$

The desired conclusion then follows from Lemmas C.8 and C.9, along with Theorem 5.2 in Chung and Romano (2013). ■

B.8 Proof of Theorem 3.8

Step 1: Limit of $\hat{\beta}_G$

We first establish that $\hat{\beta}_G \xrightarrow{P} \beta^*$ for β^* in (13). Recall that $\hat{\beta}_G$ is the OLS estimator of the slope coefficient in the linear regression of $(\hat{Y}_{\pi(2g-1)}\bar{N}_G - \hat{Y}_{\pi(2g)}\bar{N}_G)(D_{\pi(2g-1)} - D_{\pi(2g)})$ on a constant and $(\psi_{\pi(2g-1)} - \psi_{\pi(2g)})(D_{\pi(2g-1)} - D_{\pi(2g)})$, where $\bar{N}_G = \frac{1}{2G}\sum_{1 \leq g \leq 2G} N_g$. Equivalently, we have $\hat{\beta}_G$ as the OLS estimator of the slope coefficient in the linear regression of $\hat{\mu}_{1,j} - \hat{\mu}_{0,j}$ on a constant and $\hat{\psi}_{1,j} - \hat{\psi}_{0,j}$, where

$$\begin{aligned}
\hat{\mu}_{1,j} &= \left(\bar{Y}_{\pi(2j-1)}(1) - \frac{\frac{1}{G}\sum_{1 \leq g \leq 2G} \bar{Y}_g D_g N_g}{\frac{1}{G}\sum_{1 \leq g \leq 2G} D_g N_g}\right) N_{\pi(2j-1)} D_{\pi(2j-1)} \\
&\quad + \left(\bar{Y}_{\pi(2j)}(1) - \frac{\frac{1}{G}\sum_{1 \leq g \leq 2G} \bar{Y}_g D_g N_g}{\frac{1}{G}\sum_{1 \leq g \leq 2G} D_g N_g}\right) N_{\pi(2j)} D_{\pi(2j)} \\
\hat{\mu}_{0,j} &= \left(\bar{Y}_{\pi(2j-1)}(0) - \frac{\frac{1}{G}\sum_{1 \leq g \leq 2G} \bar{Y}_g (1-D_g) N_g}{\frac{1}{G}\sum_{1 \leq g \leq 2G} (1-D_g) N_g}\right) N_{\pi(2j-1)} (1-D_{\pi(2j-1)}) \\
&\quad + \left(\bar{Y}_{\pi(2j)}(0) - \frac{\frac{1}{G}\sum_{1 \leq g \leq 2G} \bar{Y}_g (1-D_g) N_g}{\frac{1}{G}\sum_{1 \leq g \leq 2G} (1-D_g) N_g}\right) N_{\pi(2j)} (1-D_{\pi(2j)}) . \\
\hat{\psi}_{1,j} &= \psi_{\pi(2j-1)} D_{\pi(2j-1)} + \psi_{\pi(2j)} D_{\pi(2j)} \\
\hat{\psi}_{0,j} &= \psi_{\pi(2j-1)} (1-D_{\pi(2j-1)}) + \psi_{\pi(2j)} (1-D_{\pi(2j)}) .
\end{aligned}$$

We start by studying an infeasible version of $\hat{\beta}_G$. Let $\tilde{\beta}_G$ denote the OLS estimator of the slope coefficient in the linear regression of $\tilde{\mu}_{1,j} - \tilde{\mu}_{0,j}$ on a constant and $\hat{\psi}_{1,j} - \hat{\psi}_{0,j}$ with j denoting the pair, where

$$\begin{aligned}\tilde{\mu}_{1,j} &= \left(\bar{Y}_{\pi(2j-1)}(1) - \frac{E[\bar{Y}_g(1)N_g]}{E[N_g]}\right) N_{\pi(2j-1)} D_{\pi(2j-1)} \\ &\quad + \left(\bar{Y}_{\pi(2j)}(1) - \frac{E[\bar{Y}_g(1)N_g]}{E[N_g]}\right) N_{\pi(2j)} D_{\pi(2j)} \\ \tilde{\mu}_{0,j} &= \left(\bar{Y}_{\pi(2j-1)}(0) - \frac{E[\bar{Y}_g(0)N_g]}{E[N_g]}\right) N_{\pi(2j-1)} (1 - D_{\pi(2j-1)}) \\ &\quad + \left(\bar{Y}_{\pi(2j)}(0) - \frac{E[\bar{Y}_g(0)N_g]}{E[N_g]}\right) N_{\pi(2j)} (1 - D_{\pi(2j)}) \ .\end{aligned}$$

Lemma C.10 then implies $\tilde{\beta}_G \xrightarrow{P} \beta^*$ for β^* in (13). Lemma C.11 shows $\tilde{\beta}_G - \hat{\beta}_G \xrightarrow{P} 0$. Therefore, $\hat{\beta}_G \xrightarrow{P} \beta^*$.

Step 2: Improvement in Efficiency

We first establish the limiting distribution of $\hat{\Delta}_G^{\text{adj}}$. Define

$$\bar{\psi}_{d,G} = \frac{1}{G} \sum_{1 \le g \le 2G} \psi_g I\{D_g = d\}$$

for $d \in \{0, 1\}$. Note that

$$\begin{aligned}&\frac{1}{G} \sum_{1 \le g \le 2G} (\bar{Y}_g(1)N_g - (\psi_g - \bar{\psi}_G)'\hat{\beta}_G)D_g \\ &= \frac{1}{G} \sum_{1 \le g \le 2G} (\bar{Y}_g(1)N_g - (\psi_g - \bar{\psi}_G)'\beta^*)D_g - \frac{1}{G} \sum_{1 \le g \le 2G} (\psi_g - \bar{\psi}_{1,G})'(\hat{\beta}_G - \beta^*)D_g - (\bar{\psi}_{1,G} - \bar{\psi}_G)'(\hat{\beta}_G - \beta^*) \\ &= \frac{1}{G} \sum_{1 \le g \le 2G} (\bar{Y}_g(1)N_g - (\psi_g - \bar{\psi}_G)'\beta^*)D_g - O_P(G^{-1/2})o_P(1) \\ &= \frac{1}{G} \sum_{1 \le g \le 2G} (\bar{Y}_g(1)N_g - (\psi_g - \bar{\psi}_G)'\beta^*)D_g + o_P(G^{-1/2}) \\ &= \frac{1}{G} \sum_{1 \le g \le 2G} (\bar{Y}_g(1)N_g - (\psi_g - E[\psi_g])'\beta^*)D_g - (\bar{\psi}_G - E[\psi_g])'\beta^* + o_P(G^{-1/2}) \ .\end{aligned}$$

where the second equality follows because $\hat{\beta}_G - \beta^* = o_P(1)$,

$$\frac{1}{G} \sum_{1 \le g \le 2G} (\psi_g - \bar{\psi}_{1,G})D_g = 0 \ ,$$

and

$$\sqrt{G}(\bar{\psi}_{1,G} - \bar{\psi}_G) = O_P(1) \ .$$

The last equality follows from the arguments that establish (A.24) in Bai et al. (2024a). Define

$$\begin{aligned}\tilde{\Delta}_G^{\text{adj}} &= \frac{\frac{1}{G} \sum_{1 \le g \le 2G} (\bar{Y}_g(1)N_g - (\psi_g - E[\psi_g])'\beta^*)D_g}{\frac{1}{G} \sum_{1 \le g \le 2G} N_g D_g} \\ &\quad - \frac{\frac{1}{G} \sum_{1 \le g \le 2G} (\bar{Y}_g(0)N_g - (\psi_g - E[\psi_g])'\beta^*)(1 - D_g)}{\frac{1}{G} \sum_{1 \le g \le 2G} N_g(1 - D_g)} \ .\end{aligned}$$

It follows from previous arguments that

$$\begin{aligned}
&\sqrt{G}(\hat{\Delta}_G^{\text{adj}} - \Delta) - \sqrt{G}(\tilde{\Delta}_G^{\text{adj}} - \Delta) \\
&= \sqrt{G}(\bar{\psi}_G - E[\psi_g])'\beta^* \left(\frac{1}{\frac{1}{G}\sum_{1\le g\le 2G} N_g D_g} - \frac{1}{\frac{1}{G}\sum_{1\le g\le 2G} N_g(1-D_g)} \right) + o_P(1) \\
&= o_P(1) \ .
\end{aligned}$$

It follows from the proof of Theorem 3.2 applied to $\bar{Y}_g(d)N_g - (\psi_g - E[\psi_g])'\beta^*$ instead of $\bar{Y}_g(d)N_g$ and Assumptions 2.1, 3.5, 3.6, 3.9, and 3.10 that $\sqrt{G}(\tilde{\Delta}_G^{\text{adj}} - \Delta) \xrightarrow{d} N(0, \varsigma^2)$ for ς^2 in (12).

Finally, we show that $\varsigma^2 \le \nu^2$. First note that by definition it follows immediately that

$$E[(E[Y_g^*(1) - Y_g^*(0)|W_g] - \Delta)^2] = E[(E[\tilde{Y}_g(1) - \tilde{Y}_g(0)|W_g] - \Delta)^2] \ .$$

It thus remains to show that

$$E[\text{Var}[Y_g^*(1)|W_g]] + E[\text{Var}[Y_g^*(0)|W_g]] \le E[\text{Var}[\tilde{Y}_g(1)|W_g]] + E[\text{Var}[\tilde{Y}_g(0)|W_g]] \ .$$

To that end,

$$\begin{aligned}
&E[\text{Var}[Y_g^*(1)|W_g]] + E[\text{Var}[Y_g^*(0)|W_g]] \\
&= E\left[\text{Var}\left[\tilde{Y}_g(1) - \frac{(\psi_g - E[\psi_g])'\beta^*}{E[N_g]}\Big|W_g\right]\right] + E\left[\text{Var}\left[\tilde{Y}_g(0) - \frac{(\psi_g - E[\psi_g])'\beta^*}{E[N_g]}\Big|W_g\right]\right] \\
&= E[\text{Var}[\tilde{Y}_g(1)|W_g]] + E[\text{Var}[\tilde{Y}_g(0)|W_g]] + 2E\left[\text{Var}\left[\frac{(\psi_g - E[\psi_g])'\beta^*}{E[N_g]}\Big|W_g\right]\right] \\
&\qquad - 2E\left[\text{Cov}\left[\tilde{Y}_g(1) + \tilde{Y}_g(0), \frac{(\psi_g - E[\psi_g])'\beta^*}{E[N_g]}\Big|W_g\right]\right] \\
&= E[\text{Var}[\tilde{Y}_g(1)|W_g]] + E[\text{Var}[\tilde{Y}_g(0)|W_g]] + \frac{2}{E[N_g]^2}E\left[\text{Var}\left[\psi_g'\beta^*|W_g\right]\right] \\
&\qquad - \frac{2}{E[N_g]}E\left[\text{Cov}\left[\tilde{Y}_g(1) + \tilde{Y}_g(0), \psi_g'\beta^*\Big|W_g\right]\right] \\
&= E[\text{Var}[\tilde{Y}_g(1)|W_g]] + E[\text{Var}[\tilde{Y}_g(0)|W_g]] - \frac{2}{E[N_g]^2}E\left[\text{Var}\left[\psi_g'\beta^*|W_g\right]\right]
\end{aligned}$$

where the first equality follows by definition, the last equality by noting that β^* is the projection coefficient of $\frac{E[N_g]}{2}(\tilde{Y}_g(1) + \tilde{Y}_g(0) - E[\tilde{Y}_g(1) + \tilde{Y}_g(0)|W_g])$ on $\psi_g - E[\psi_g|W_g]$,

$$E[N_g]E[(\tilde{Y}_g(1) + \tilde{Y}_g(0) - E[\tilde{Y}_g(1) + \tilde{Y}_g(0)|W_g])(\psi_g - E[\psi_g|W_g])'\beta^*] = 2E[((\psi_g - E[\psi_g|W_g])'\beta^*)^2] \ ,$$

or equivalently,

$$E[N_g]E[\text{Cov}[\tilde{Y}_g(1) + \tilde{Y}_g(0), \psi_g'\beta^*|W_g]] = 2E[\text{Var}[\psi_g'\beta^*|W_g]] \ . \tag{23}$$

We thus obtain

$$\varsigma^2 = \nu^2 - \kappa^2 \ ,$$

where

$$\kappa^2 = \frac{2}{E[N_g]^2} E\left[\operatorname{Var}\left[\psi_g' \beta^* | W_g\right]\right] ,$$

and the desired result follows. ■

B.9 Proof of Theorem 3.9

The desired result follows from combining the arguments used to establish Theorem 3.3 and those used to establish Theorem 3.2 in Bai et al. (2024a). ■

C Auxiliary Lemmas

Lemma C.1. *Suppose Q satisfies Assumptions 2.1 and 3.3 and the treatment assignment mechanism satisfies Assumptions 3.1–3.2. Define*

$$\begin{aligned}
\mathbb{L}_G^{\text{YN1}} &= \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} (\bar{Y}_g(1) N_g D_g - E[\bar{Y}_g(1) N_g] D_g) \\
\mathbb{L}_G^{\text{N1}} &= \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} (N_g D_g - E[N_g] D_g) \\
\mathbb{L}_G^{\text{YN0}} &= \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} (\bar{Y}_g(0) N_g (1 - D_g) - E[\bar{Y}_g(0) N_g](1 - D_g)) \\
\mathbb{L}_G^{\text{N0}} &= \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} (N_g (1 - D_g) - E[N_g](1 - D_g)) .
\end{aligned}$$

Then, as $G \to \infty$,

$$(\mathbb{L}_G^{\text{YN1}}, \mathbb{L}_G^{\text{N1}}, \mathbb{L}_G^{\text{YN0}}, \mathbb{L}_G^{\text{N0}})' \xrightarrow{d} N(0, \mathbb{V}) ,$$

where

$$\mathbb{V} = \mathbb{V}_1 + \mathbb{V}_2$$

for

$$\mathbb{V}_1 = \begin{pmatrix} \mathbb{V}_1^1 & 0 \\ 0 & \mathbb{V}_1^0 \end{pmatrix}$$

$$\begin{aligned}
\mathbb{V}_1^1 &= \begin{pmatrix} E[\operatorname{Var}[\bar{Y}_g(1) N_g | X_g]] & E[\operatorname{Cov}[\bar{Y}_g(1) N_g, N_g | X_g]] \\ E[\operatorname{Cov}[\bar{Y}_g(1) N_g, N_g | X_g]] & E[\operatorname{Var}[N_g | X_g]] \end{pmatrix} \\
\mathbb{V}_1^0 &= \begin{pmatrix} E[\operatorname{Var}[\bar{Y}_g(0) N_g | X_g]] & E[\operatorname{Cov}[\bar{Y}_g(0) N_g, N_g | X_g]] \\ E[\operatorname{Cov}[\bar{Y}_g(0) N_g, N_g | X_g]] & E[\operatorname{Var}[N_g | X_g]] \end{pmatrix}
\end{aligned}$$

$$\mathbb{V}_2 = \frac{1}{2} \operatorname{Var}[(E[\bar{Y}_g(1) N_g | X_g], E[N_g | X_g], E[\bar{Y}_g(0) N_g | X_g], E[N_g | X_g])'] .$$

Proof. We break the proof into the following steps:

Step 1: Decomposition by conditioning on $X^{(G)}$ and $D^{(G)}$

Note

$$(\mathbb{L}_G^{\mathrm{YN1}}, \mathbb{L}_G^{\mathrm{N1}}, \mathbb{L}_G^{\mathrm{YN0}}, \mathbb{L}_G^{\mathrm{N0}}) = (\mathbb{L}_{1,G}^{\mathrm{YN1}}, \mathbb{L}_{1,G}^{\mathrm{N1}}, \mathbb{L}_{1,G}^{\mathrm{YN0}}, \mathbb{L}_{1,G}^{\mathrm{N0}}) + (\mathbb{L}_{2,G}^{\mathrm{YN1}}, \mathbb{L}_{2,G}^{\mathrm{N1}}, \mathbb{L}_{2,G}^{\mathrm{YN0}}, \mathbb{L}_{2,G}^{\mathrm{N0}}) ,$$

where

$$\begin{aligned}
\mathbb{L}_{1,G}^{\mathrm{YN1}} &= \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} (\bar{Y}_g(1) N_g D_g - E[\bar{Y}_g(1) N_g D_g | X^{(G)}, D^{(G)}]) \\
\mathbb{L}_{2,G}^{\mathrm{YN1}} &= \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} (E[\bar{Y}_g(1) N_g D_g | X^{(G)}, D^{(G)}] - E[\bar{Y}_g(1) N_g] D_g)
\end{aligned}$$

and similarly for the rest. Next, note $(\mathbb{L}_{1,G}^{\mathrm{YN1}}, \mathbb{L}_{1,G}^{\mathrm{N1}}, \mathbb{L}_{1,G}^{\mathrm{YN0}}, \mathbb{L}_{1,G}^{\mathrm{N0}}), G \ge 1$ is a triangular array of mean-zero random vectors. Conditional on $X^{(G)}, D^{(G)}$, $(\mathbb{L}_{1,G}^{\mathrm{YN1}}, \mathbb{L}_{1,G}^{\mathrm{N1}}) \perp\!\!\!\perp (\mathbb{L}_{1,G}^{\mathrm{YN0}}, \mathbb{L}_{1,G}^{\mathrm{N0}})$. Moreover, it follows from $Q_G = Q^{2G}$ and Assumption 3.1 that

$$\begin{aligned}
&\mathrm{Var}\left[\begin{pmatrix} \mathbb{L}_{1,G}^{\mathrm{YN1}} \\ \mathbb{L}_{1,G}^{\mathrm{N1}} \end{pmatrix} \middle| X^{(G)}, D^{(G)} \right] \\
&= \begin{pmatrix} \frac{1}{G}\sum_{1 \le g \le 2G} \mathrm{Var}[\bar{Y}_g(1) N_g | X_g] D_g & \frac{1}{G}\sum_{1 \le g \le 2G} \mathrm{Cov}[\bar{Y}_g(1) N_g, N_g | X_g] D_g \\ \frac{1}{G}\sum_{1 \le g \le 2G} \mathrm{Cov}[\bar{Y}_g(1) N_g, N_g | X_g] D_g & \frac{1}{G}\sum_{1 \le g \le 2G} \mathrm{Var}[N_g | X_g] D_g \end{pmatrix} .
\end{aligned}$$

Step 2: Limits of conditional variances

For the upper left component, we have

$$\frac{1}{G} \sum_{1 \le g \le 2G} \mathrm{Var}[\bar{Y}_g(1) N_g | X_g] D_g = \frac{1}{G} \sum_{1 \le g \le 2G} E[\bar{Y}_g^2(1) N_g^2 | X_g] D_g - \frac{1}{G} \sum_{1 \le g \le 2G} E[\bar{Y}_g(1) N_g | X_g]^2 D_g . \tag{24}$$

Note

$$\begin{aligned}
&\frac{1}{G} \sum_{1 \le g \le 2G} E[\bar{Y}_g^2(1) N_g^2 | X_g] D_g \\
&= \frac{1}{2G} \sum_{1 \le g \le 2G} E[\bar{Y}_g^2(1) N_g^2 | X_g] + \frac{1}{2}\Big(\frac{1}{G} \sum_{1 \le g \le 2G : D_g = 1} E[\bar{Y}_g^2(1) N_g^2 | X_g] - \frac{1}{G} \sum_{1 \le g \le 2G : D_g = 0} E[\bar{Y}_g^2(1) N_g^2 | X_g] \Big) .
\end{aligned}$$

It follows from the weak law of large numbers, the application of which is permitted by Lemma C.12, that

$$\frac{1}{2G} \sum_{1 \le g \le 2G} E[\bar{Y}_g^2(1) N_g^2 | X_g] \xrightarrow{P} E[\bar{Y}_g^2(1) N_g^2] .$$

On the other hand, it follows from Assumptions 3.2 and 3.3(a) that

$$\begin{aligned}
&\Big| \frac{1}{G} \sum_{1 \le g \le 2G : D_g = 1} E[\bar{Y}_g^2(1) N_g^2 | X_g] - \frac{1}{G} \sum_{1 \le g \le 2G : D_g = 0} E[\bar{Y}_g^2(1) N_g^2 | X_g] \Big| \\
&\le \frac{1}{G} \sum_{1 \le j \le G} |E[\bar{Y}_{\pi(2j-1)}^2(1) N_{\pi(2j-1)}^2 | X_{\pi(2j-1)}] - E[\bar{Y}_{\pi(2j)}^2(1) N_{\pi(2j)}^2 | X_{\pi(2j)}]|
\end{aligned}$$

$$\lesssim \frac{1}{G} \sum_{1 \le j \le G} \|X_{\pi(2j-1)} - X_{\pi(2j)}\| \xrightarrow{P} 0 \ .$$

Therefore,

$$\frac{1}{G} \sum_{1 \le g \le 2G} E[\bar{Y}_g^2(1) N_g^2 | X_g] D_g \xrightarrow{P} E[\bar{Y}_g^2(1) N_g^2] \ .$$

Meanwhile,

$$\begin{aligned} &\frac{1}{G} \sum_{1 \le g \le 2G} E[\bar{Y}_g(1) N_g | X_g]^2 D_g \\ &= \frac{1}{2G} \sum_{1 \le g \le 2G} E[\bar{Y}_g(1) N_g | X_g]^2 + \frac{1}{2} \Big(\frac{1}{G} \sum_{1 \le g \le 2G : D_g = 1} E[\bar{Y}_g(1) N_g | X_g]^2 - \frac{1}{G} \sum_{1 \le g \le 2G : D_g = 0} E[\bar{Y}_g(1) N_g | X_g]^2 \Big) \ . \end{aligned}$$

It follows from the weak law of large numbers, the application of which is permitted by Lemma C.12, that

$$\frac{1}{2G} \sum_{1 \le g \le 2G} E[\bar{Y}_g(1) N_g | X_g]^2 \xrightarrow{P} E[E[\bar{Y}_g(1) N_g | X_g]^2] \ .$$

Next,

$$\begin{aligned} &\Big| \frac{1}{G} \sum_{1 \le g \le 2G : D_g = 1} E[\bar{Y}_g(1) N_g | X_g]^2 - \frac{1}{G} \sum_{1 \le g \le 2G : D_g = 0} E[\bar{Y}_g(1) N_g | X_g]^2 \Big| \\ &\le \frac{1}{G} \sum_{1 \le j \le G} |E[\bar{Y}_{\pi(2j-1)}(1) N_{\pi(2j-1)} | X_{\pi(2j-1)}] - E[\bar{Y}_{\pi(2j)}(1) N_{\pi(2j)} | X_{\pi(2j)}]| \\ &\qquad \times |E[\bar{Y}_{\pi(2j-1)}(1) N_{\pi(2j-1)} | X_{\pi(2j-1)}] + E[\bar{Y}_{\pi(2j)}(1) N_{\pi(2j)} | X_{\pi(2j)}]| \\ &\lesssim \Big(\frac{1}{G} \sum_{1 \le j \le G} \|X_{\pi(2j-1)} - X_{\pi(2j)}\|^2 \Big)^{1/2} \\ &\qquad \times \Big(\frac{1}{G} \sum_{1 \le j \le G} (|E[\bar{Y}_{\pi(2j-1)}(1) N_{\pi(2j-1)} | X_{\pi(2j-1)}] + E[\bar{Y}_{\pi(2j)}(1) N_{\pi(2j)} | X_{\pi(2j)}]|)^2 \Big)^{1/2} \\ &\lesssim \Big(\frac{1}{G} \sum_{1 \le j \le G} \|X_{\pi(2j-1)} - X_{\pi(2j)}\|^2 \Big)^{1/2} \\ &\qquad \times \Big(\frac{1}{G} \sum_{1 \le j \le G} (|E[\bar{Y}_{\pi(2j-1)}(1) N_{\pi(2j-1)} | X_{\pi(2j-1)}]|^2 + |E[\bar{Y}_{\pi(2j)}(1) N_{\pi(2j)} | X_{\pi(2j)}]|^2) \Big)^{1/2} \\ &\le \Big(\frac{1}{G} \sum_{1 \le j \le G} \|X_{\pi(2j-1)} - X_{\pi(2j)}\|^2 \Big)^{1/2} \Big(\frac{1}{G} \sum_{1 \le g \le 2G} E[\bar{Y}_g(1) N_g | X_g]^2 \Big)^{1/2} \xrightarrow{P} 0 \ , \end{aligned}$$

where the first inequality follows by inspection, the second follows from Assumption 3.3(a) and the Cauchy-Schwarz inequality, the third follows from $(a+b)^2 \le 2a^2 + 2b^2$, the last follows by inspection again and the convergence in probability follows from Assumption 3.2 and the law of large numbers. Therefore,

$$\frac{1}{G} \sum_{1 \le g \le 2G} E[\bar{Y}_g(1) N_g | X_g]^2 D_g \xrightarrow{P} E\left[E[\bar{Y}_g(1) N_g | X_g]^2\right] \ ,$$

and hence it follows from (24) that

$$\frac{1}{G}\sum_{1\leq g\leq 2G}\text{Var}[\bar{Y}_g(1)N_g|X_g]D_g \xrightarrow{P} E[\text{Var}[\bar{Y}_g(1)N_g|X_g]] \ .$$

An identical argument establishes that

$$\frac{1}{G}\sum_{1\leq g\leq 2G}\text{Var}[N_g|X_g]D_g \xrightarrow{P} E[\text{Var}[N_g|X_g]] \ .$$

To study the off-diagonal components, note that

$$\begin{aligned}&\frac{1}{G}\sum_{1\leq g\leq 2G}\text{Cov}[\bar{Y}_g(1)N_g, N_g|X_g]D_g \\ &\qquad = \frac{1}{G}\sum_{1\leq g\leq 2G}E[\bar{Y}_g(1)N_g^2|X_g]D_g - \frac{1}{G}\sum_{1\leq g\leq 2G}E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g]D_g \ . \qquad (25)\end{aligned}$$

By a similar argument to that used above, it can be shown that

$$\frac{1}{G}\sum_{1\leq g\leq 2G}E[\bar{Y}_g(1)N_g^2|X_g]D_g \xrightarrow{P} E[\bar{Y}_g(1)N_g^2] \ .$$

Meanwhile,

$$\begin{aligned}&\frac{1}{G}\sum_{1\leq g\leq 2G}E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g]D_g \\ &= \frac{1}{2G}\sum_{1\leq g\leq 2G}E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g] \\ &\quad + \frac{1}{2}\Big(\frac{1}{G}\sum_{1\leq g\leq 2G:D_g=1}E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g] - \frac{1}{G}\sum_{1\leq g\leq 2G:D_g=0}E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g]\Big) \ .\end{aligned}$$

Note that

$$E[E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g]] = E[[N_gE[\bar{Y}_g(1)|W_g]|X_g]E[N_g|X_g]] \lesssim E[N_g^2] < \infty \ ,$$

where the equality follows by the law of iterated expectations and the inequality by Lemma C.12 and Jensen's inequality, and the law of iterated expectations. Thus by the weak law of large numbers,

$$\frac{1}{2G}\sum_{1\leq g\leq 2G}E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g] \xrightarrow{P} E[E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g]] \ .$$

Next, by the triangle inequality

$$\begin{aligned}&\Big|\frac{1}{G}\sum_{1\leq g\leq 2G:D_g=1}E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g] - \frac{1}{G}\sum_{1\leq g\leq 2G:D_g=0}E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g]\Big| \\ &\leq \frac{1}{G}\sum_{1\leq j\leq G}\big|E[\bar{Y}_{\pi(2j-1)}(1)N_{\pi(2j-1)}|X_{\pi(2j-1)}]E[N_{\pi(2j-1)}|X_{\pi(2j-1)}]\end{aligned}$$

$$-E[\bar{Y}_{\pi(2j)}(1)N_{\pi(2j)}|X_{\pi(2j)}]E[N_{\pi(2j)}|X_{\pi(2j)}]\Big| \ ,$$

and for each j,

$$\begin{aligned}
&\big|E[\bar{Y}_{\pi(2j-1)}(1)N_{\pi(2j-1)}|X_{\pi(2j-1)}]E[N_{\pi(2j-1)}|X_{\pi(2j-1)}] - E[\bar{Y}_{\pi(2j)}(1)N_{\pi(2j)}|X_{\pi(2j)}]E[N_{\pi(2j)}|X_{\pi(2j)}]\big| \\
&= \Big|(E[\bar{Y}_{\pi(2j-1)}(1)N_{\pi(2j-1)}|X_{\pi(2j-1)}] - E[\bar{Y}_{\pi(2j)}(1)N_{\pi(2j)}|X_{\pi(2j)}])E[N_{\pi(2j)}|X_{\pi(2j)}] \\
&\qquad + (E[N_{\pi(2j-1)}|X_{\pi(2j-1)}] - E[N_{\pi(2j)}|X_{\pi(2j)}])E[\bar{Y}_{\pi(2j-1)}(1)N_{\pi(2j-1)}|X_{\pi(2j-1)}]\Big| \\
&\lesssim \big|E[\bar{Y}_{\pi(2j-1)}(1)N_{\pi(2j-1)}|X_{\pi(2j-1)}] - E[\bar{Y}_{\pi(2j)}(1)N_{\pi(2j)}|X_{\pi(2j)}]\big| \\
&\qquad + \big|E[N_{\pi(2j-1)}|X_{\pi(2j-1)}] - E[N_{\pi(2j)}|X_{\pi(2j)}]\big| \ ,
\end{aligned}$$

where the final inequality follows from the triangle inequality, Assumption 3.3(b) and Lemma C.12. Therefore,

$$\begin{aligned}
&\Big|\frac{1}{G}\sum_{1\le g\le 2G:D_g=1} E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g] - \frac{1}{G}\sum_{1\le g\le 2G:D_g=0} E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g]\Big| \\
&\lesssim \frac{1}{G}\sum_{1\le j\le G}\big(\big|E[\bar{Y}_{\pi(2j-1)}(1)N_{\pi(2j-1)}|X_{\pi(2j-1)}] - E[\bar{Y}_{\pi(2j)}(1)N_{\pi(2j)}|X_{\pi(2j)}]\big| \\
&\qquad\qquad + \big|E[N_{\pi(2j-1)}|X_{\pi(2j-1)}] - E[N_{\pi(2j)}|X_{\pi(2j)}]\big|\big) \\
&\lesssim \frac{1}{G}\sum_{1\le j\le G}\|X_{\pi(2j-1)} - X_{\pi(2j)}\| \xrightarrow{P} 0 \ ,
\end{aligned}$$

where the final inequality follows from Assumptions 3.3 and the convergence in probability follows from Assumption 3.1. Proceeding as in the case of the upper left component, we obtain that

$$\frac{1}{G}\sum_{1\le g\le 2G}\mathrm{Cov}[\bar{Y}_g(1)N_g, N_g|X_g]D_g \xrightarrow{P} E[\mathrm{Cov}[\bar{Y}_g(1)N_g, N_g|X_g]] \ .$$

Thus we have established that

$$\mathrm{Var}\left[\begin{pmatrix}\mathbb{L}^{\mathrm{YN1}}_{1,G}\\ \mathbb{L}^{\mathrm{N1}}_{1,G}\end{pmatrix}\middle| X^{(G)}, D^{(G)}\right] \xrightarrow{P} \mathbb{V}^1_1 \ .$$

Similarly,

$$\mathrm{Var}\left[\begin{pmatrix}\mathbb{L}^{\mathrm{YN0}}_{1,G}\\ \mathbb{L}^{\mathrm{N0}}_{1,G}\end{pmatrix}\middle| X^{(G)}, D^{(G)}\right] \xrightarrow{P} \mathbb{V}^0_1 \ .$$

Step 3: Conditional CLT

We now establish

$$\rho(\mathcal{L}((\mathbb{L}^{\mathrm{YN1}}_{1,G}, \mathbb{L}^{\mathrm{N1}}_{1,G}, \mathbb{L}^{\mathrm{YN0}}_{1,G}, \mathbb{L}^{\mathrm{N0}}_{1,G})'|X^{(G)}, D^{(G)}), N(0,\mathbb{V}_1)) \xrightarrow{P} 0 \ , \tag{26}$$

where $\mathcal{L}(\cdot)$ is used to denote the law of a random variable and ρ is any metric that metrizes weak convergence. For that purpose, note that we only need to show that for any subsequence $\{G_k\}$ there exists a further subsequence $\{G_{k_l}\}$ along which

$$\rho(\mathcal{L}((\mathbb{L}^{\text{YN1}}_{1,G_{k_l}}, \mathbb{L}^{\text{N1}}_{1,G_{k_l}}, \mathbb{L}^{\text{YN0}}_{1,G_{k_l}}, L^{\text{N0}}_{1,G_{k_l}})|X^{(G_{k_l})}, D^{(G^{k_l})}, N(0, \mathbb{V}_1)) \to 0 \text{ with probability one} . \tag{27}$$

In order to extract such a subsequence, we verify the conditions in the Lindeberg central limit theorem in Proposition 2.27 of van der Vaart (1998) are satisfied in probability for the original sequence, because then we can extract a subsequence along which the conditions in that proposition hold almost surely. The second condition in that proposition is satisfied because we have shown

$$\text{Var}[(\mathbb{L}^{\text{YN1}}_{1,G}, \mathbb{L}^{\text{N1}}_{1,G}, \mathbb{L}^{\text{YN0}}_{1,G}, \mathbb{L}^{\text{N0}}_{1,G})'|X^{(G)}, D^{(G)}] \xrightarrow{P} \mathbb{V}_1 .$$

The first condition in that proposition can be verified component wise because of the following inequality:

$$\left|\sum_{1\le j\le k} a_j\right| I\left\{\left|\sum_{1\le j\le k} a_j\right| > \epsilon\right\} \le \sum_{1\le j\le k} k|a_j| I\left\{|a_j| > \frac{\epsilon}{k}\right\} . \tag{28}$$

Therefore, we will only verify that

$$\begin{aligned}\frac{1}{G}\sum_{1\le g\le 2G} &E[(D_g(\bar{Y}_g(1)N_g - E[\bar{Y}_g(1)N_g|X_g]))^2 \\ &\times I\{(D_g(\bar{Y}_g(1)N_g - E[\bar{Y}_g(1)N_g|X_g]))^2 > \epsilon^2 G\}|X^{(G)}, D^{(G)}] \xrightarrow{P} 0\end{aligned} \tag{29}$$

To verify (29), note it follows from (28) that

$$\begin{aligned}&\frac{1}{G}\sum_{1\le g\le 2G} E[(D_g(\bar{Y}_g(1)N_g - E[\bar{Y}_g(1)N_g|X_g]))^2 I\{(D_g(\bar{Y}_g(1)N_g - E[\bar{Y}_g(1)N_g|X_g]))^2 > \epsilon^2 G\}|X^{(G)}, D^{(G)}] \\ &\lesssim \frac{1}{G}\sum_{1\le g\le 2G} E[D_g(\bar{Y}_g(1)N_g - E[\bar{Y}_g(1)N_g|X_g])^2 I\{D_g(\bar{Y}_g(1)N_g - E[\bar{Y}_g(1)N_g|X_g])^2 > \epsilon^2 G/2\}|X^{(G)}, D^{(G)}] \\ &\le \frac{1}{G}\sum_{1\le g\le 2G} E[(\bar{Y}_g(1)N_g - E[\bar{Y}_g(1)N_g|X_g])^2 I\{|\bar{Y}_g(1)N_g - E[\bar{Y}_g(1)N_g|X_g]| > \epsilon\sqrt{G}/\sqrt{2}\}|X_g] .\end{aligned}$$

Fix any $m > 0$. For G large enough, the previous line

$$\begin{aligned}&\le \frac{1}{G}\sum_{1\le g\le 2G} E[(\bar{Y}_g(1)N_g - E[\bar{Y}_g(1)N_g|X_g])^2 I\{|\bar{Y}_g(1)N_g - E[\bar{Y}_g(1)N_g|X_g]| > m\}|X_g] \\ &\xrightarrow{P} 2E[(\bar{Y}_g(1)N_g - E[\bar{Y}_g(1)N_g|X_g])^2 I\{|\bar{Y}_g(1)N_g - E[\bar{Y}_g(1)N_g|X_g]| > m\}]\end{aligned}$$

because $E[(\bar{Y}_g(1)N_g - E[\bar{Y}_g(1)N_g|X_g])^2] < \infty$. As $m \to \infty$, the last expression goes to 0. Therefore, it follows from a similar diagonalization argument to that in the proof of Lemma B.3 of Bai (2022) that both conditions in Proposition 2.27 of van der Vaart (1998) hold in probability, and therefore there must be a subsequence along which they hold almost surely, so (27) and hence (26) holds.

Step 4: Unconditional components

Next, we study $(\mathbb{L}_{2,G}^{\rm YN1}, \mathbb{L}_{2,G}^{\rm N1}, \mathbb{L}_{2,G}^{\rm YN0}, \mathbb{L}_{2,G}^{\rm N0})$. It follows from $Q_G = Q^{2G}$ and Assumption 3.1 that

$$\begin{pmatrix} \mathbb{L}_{2,G}^{\rm YN1} \\ \mathbb{L}_{2,G}^{\rm N1} \\ \mathbb{L}_{2,G}^{\rm YN0} \\ \mathbb{L}_{2,G}^{\rm N0} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} D_g (E[\bar{Y}_g(1) N_g | X_g] - E[\bar{Y}_g(1) N_g]) \\ \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} D_g (E[N_g | X_g] - E[N_g]) \\ \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} (1 - D_g)(E[\bar{Y}_g(0) N_g | X_g] - E[\bar{Y}_g(0) N_g]) \\ \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} (1 - D_g)(E[N_g | X_g] - E[N_g]) \end{pmatrix} .$$

For $\mathbb{L}_{2,G}^{\rm YN1}$, note it follows from Assumption 3.1 that

$$\begin{aligned} \text{Var}[\mathbb{L}_{2,G}^{\rm YN1} | X^{(G)}] &= \frac{1}{4G} \sum_{1 \le j \le G} (E[\bar{Y}_{\pi(2j-1)}(1) N_{\pi(2j-1)} | X_{\pi(2j-1)}] - E[\bar{Y}_{\pi(2j)}(1) N_{\pi(2j)} | X_{\pi(2j)}])^2 \\ &\lesssim \frac{1}{G} \sum_{1 \le j \le G} \| X_{\pi(2j-1)} - X_{\pi(2j)} \|^2 \xrightarrow{P} 0 . \end{aligned}$$

Therefore, it follows from Markov's inequality conditional on $X^{(G)}$ and $D^{(G)}$, and the fact that probabilities are bounded and hence uniformly integrable, that

$$\mathbb{L}_{2,G}^{\rm YN1} = E[\mathbb{L}_{2,G}^{\rm YN1} | X^{(G)}] + o_P(1) .$$

Applying a similar argument to each of $L_{2,G}^{\rm N1}$, $L_{2,G}^{\rm YN0}$, $L_{2,G}^{\rm N0}$ allows us to conclude that

$$\begin{pmatrix} \mathbb{L}_{2,G}^{\rm YN1} \\ \mathbb{L}_{2,G}^{\rm N1} \\ \mathbb{L}_{2,G}^{\rm YN0} \\ \mathbb{L}_{2,G}^{\rm N0} \end{pmatrix} = \begin{pmatrix} \frac{1}{2\sqrt{G}} \sum_{1 \le g \le 2G} (E[\bar{Y}_g(1) N_g | X_g] - E[\bar{Y}_g(1) N_g]) \\ \frac{1}{2\sqrt{G}} \sum_{1 \le g \le 2G} (E[N_g | X_g] - E[N_g]) \\ \frac{1}{2\sqrt{G}} \sum_{1 \le g \le 2G} (E[\bar{Y}_g(0) N_g | X_g] - E[\bar{Y}_g(0) N_g]) \\ \frac{1}{2\sqrt{G}} \sum_{1 \le g \le 2G} (E[N_g | X_g] - E[N_g]) \end{pmatrix} + o_P(1) .$$

It thus follows from the central limit theorem, the application of which is justified by Jensen's inequality combined with Assumption 2.1(b) and Lemma C.12, that

$$(\mathbb{L}_{2,G}^{\rm YN1}, \mathbb{L}_{2,G}^{\rm N1}, \mathbb{L}_{2,G}^{\rm YN0}, \mathbb{L}_{2,G}^{\rm N0})' \xrightarrow{d} N(0, \mathbb{V}_2) .$$

<u>Step 5: Combining unconditional and conditional components</u>

Because (26) holds and $(\mathbb{L}_{2,G}^{\rm YN1}, \mathbb{L}_{2,G}^{\rm N1}, \mathbb{L}_{2,G}^{\rm YN0}, \mathbb{L}_{2,G}^{\rm N0})$ is deterministic conditional on $X^{(G)}, D^{(G)}$, the conclusion of the theorem follows from Lemma S.1.3 in Bai et al. (2022). ■

Lemma C.2. *Let $\mathbb{V}$ be defined as in Lemma C.1, and D_{h0} be defined as in the proof of Theorem 3.1, then*

$$D_{h0} \mathbb{V} D_{h0}' = \omega^2 ,$$

where

$$\omega^2 = E[\tilde{Y}_g^2(1)] + E[\tilde{Y}_g^2(0)] - \frac{1}{2} E[(E[\tilde{Y}_g(1) + \tilde{Y}_g(0) | X_g])^2] .$$

Proof. To see this, note by the laws of total variance and total covariance that $\mathbb{V}$ in Lemma C.1 is symmetric with entries

$$
\begin{aligned}
\mathbb{V}_{11} &= \mathrm{Var}[\bar{Y}_g(1)N_g] - \frac{1}{2}\,\mathrm{Var}[E[\bar{Y}_g(1)N_g|X_g]] \\
\mathbb{V}_{12} &= \mathrm{Cov}[\bar{Y}_g(1)N_g, N_g] - \frac{1}{2}\,\mathrm{Cov}[E[\bar{Y}_g(1)N_g|X_g], E[N_g|X_g]] \\
\mathbb{V}_{13} &= \frac{1}{2}\,\mathrm{Cov}[E[\bar{Y}_g(1)N_g|X_g], E[\bar{Y}_g(0)N_g|X_g]] \\
\mathbb{V}_{14} &= \frac{1}{2}\,\mathrm{Cov}[E[\bar{Y}_g(1)N_g|X_g], E[N_g|X_g]] \\
\mathbb{V}_{22} &= \mathrm{Var}[N_g] - \frac{1}{2}\,\mathrm{Var}[E[N_g|X_g]] \\
\mathbb{V}_{23} &= \frac{1}{2}\,\mathrm{Cov}[E[N_g|X_g], E[\bar{Y}_g(0)N_g|X_g]] \\
\mathbb{V}_{24} &= \frac{1}{2}\,\mathrm{Cov}[E[N_g|X_g], E[N_g|X_g]] \\
\mathbb{V}_{33} &= \mathrm{Var}[\bar{Y}_g(0)N_g] - \frac{1}{2}\,\mathrm{Var}[E[\bar{Y}_g(0)N_g|X_g]] \\
\mathbb{V}_{34} &= \mathrm{Cov}[\bar{Y}_g(0)N_g, N_g] - \frac{1}{2}\,\mathrm{Cov}[E[\bar{Y}_g(0)N_g|X_g], E[N_g|X_g]] \\
\mathbb{V}_{44} &= \mathrm{Var}[N_g] - \frac{1}{2}\,\mathrm{Var}[E[N_g|X_g]]\ .
\end{aligned}
$$

We separately calculate the variance terms involving conditional expectations and those that don't. The terms not involving conditional expectations are

$$
\begin{aligned}
&\frac{\mathrm{Var}[\bar{Y}_g(1)N_g]}{E[N_g]^2} + \frac{\mathrm{Var}[N_g]E[\bar{Y}_g(1)N_g]^2}{E[N_g]^4} + \frac{\mathrm{Var}[\bar{Y}_g(0)N_g]}{E[N_g]^2} + \frac{\mathrm{Var}[N_g]E[\bar{Y}_g(0)N_g]^2}{E[N_g]^4} \\
&\qquad - \frac{2\,\mathrm{Cov}[\bar{Y}_g(1)N_g, N_g]E[\bar{Y}_g(1)N_g]}{E[N_g]^3} - \frac{2\,\mathrm{Cov}[\bar{Y}_g(0)N_g, N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]^3} \\
&= \frac{E[\bar{Y}_g^2(1)N_g^2] - E[\bar{Y}_g(1)N_g]^2}{E[N_g]^2} + \frac{E[N_g^2]E[\bar{Y}_g(1)N_g]^2 - E[N_g]^2E[\bar{Y}_g(1)N_g]^2}{E[N_g]^4} \\
&\qquad + \frac{E[\bar{Y}_g^2(0)N_g^2] - E[\bar{Y}_g(0)N_g]^2}{E[N_g]^2} + \frac{E[N_g^2]E[\bar{Y}_g(0)N_g]^2 - E[N_g]^2E[\bar{Y}_g(0)N_g]^2}{E[N_g]^4} \\
&\qquad - \frac{2E[\bar{Y}_g(1)N_g^2]E[\bar{Y}_g(1)N_g]}{E[N_g]^3} + \frac{2E[\bar{Y}_g(1)N_g]E[N_g]E[\bar{Y}_g(1)N_g]}{E[N_g]^3} \\
&\qquad - \frac{2E[\bar{Y}_g(0)N_g^2]E[\bar{Y}_g(0)N_g]}{E[N_g]^3} + \frac{2E[\bar{Y}_g(0)N_g]E[N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]^3} \\
&= \frac{E[\bar{Y}_g^2(1)N_g^2]}{E[N_g]^2} + \frac{E[\bar{Y}_g^2(0)N_g^2]}{E[N_g]^2} + \frac{E[N_g^2]E[\bar{Y}_g(1)N_g]^2}{E[N_g]^4} + \frac{E[N_g^2]E[\bar{Y}_g(0)N_g]^2}{E[N_g]^4} \\
&\qquad - \frac{2E[\bar{Y}_g(1)N_g^2]E[\bar{Y}_g(1)N_g]}{E[N_g]^3} - \frac{2E[\bar{Y}_g(0)N_g^2]E[\bar{Y}_g(0)N_g]}{E[N_g]^3} \\
&= E[\tilde{Y}_g^2(1)] + E[\tilde{Y}_g^2(0)]\ ,
\end{aligned}
$$

where

$$
\tilde{Y}_g(d) = \frac{N_g}{E[N_g]}\left(\bar{Y}_g(d) - \frac{E[\bar{Y}_g(d)N_g]}{E[N_g]}\right)
$$

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

Next, the terms involving conditional expectations are

$$
\begin{aligned}
&-\frac{\text{Var}[E[\bar{Y}_g(1)N_g|X_g]]}{2E[N_g]^2}-\frac{\text{Var}[E[N_g|X_g]]E[\bar{Y}_g(1)N_g]^2}{2E[N_g]^4} \\
&\quad-\frac{\text{Var}[E[\bar{Y}_g(0)N_g|X_g]]}{2E[N_g]^2}-\frac{\text{Var}[E[N_g|X_g]]E[\bar{Y}_g(0)N_g]^2}{2E[N_g]^4} \\
&\quad+\frac{\text{Cov}[E[\bar{Y}_g(1)N_g|X_g],E[N_g|X_g]]E[\bar{Y}_g(1)N_g]}{E[N_g]^3}+\frac{\text{Cov}[E[\bar{Y}_g(0)N_g|X_g],E[N_g|X_g]]E[\bar{Y}_g(0)N_g]}{E[N_g]^3} \\
&\quad-\frac{\text{Cov}[E[\bar{Y}_g(1)N_g|X_g],E[\bar{Y}_g(0)N_g|X_g]]}{E[N_g]^2}+\frac{\text{Cov}[E[\bar{Y}_g(1)N_g|X_g],E[N_g|X_g]]E[\bar{Y}_g(0)N_g]}{E[N_g]E[N_g]^2} \\
&\quad+\frac{\text{Cov}[E[N_g|X_g],E[\bar{Y}_g(0)N_g|X_g]]E[\bar{Y}_g(1)N_g]}{E[N_g]^2E[N_g]}-\frac{\text{Cov}[E[N_g|X_g],E[N_g|X_g]]E[\bar{Y}_g(1)N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]^2E[N_g]^2} \\
&=-\frac{E[E[\bar{Y}_g(1)N_g|X_g]^2]-E[\bar{Y}_g(1)N_g]^2}{2E[N_g]^2}-\frac{(E[E[N_g|X_g]^2]-E[N_g]^2)E[\bar{Y}_g(1)N_g]^2}{2E[N_g]^4} \\
&\quad-\frac{E[E[\bar{Y}_g(0)N_g|X_g]^2]-E[\bar{Y}_g(0)N_g]^2}{2E[N_g]^2}-\frac{(E[E[N_g|X_g]^2]-E[N_g]^2)E[\bar{Y}_g(0)N_g]^2}{2E[N_g]^4} \\
&\quad+\frac{(E[E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g]]-E[\bar{Y}_g(1)N_g]E[N_g])E[\bar{Y}_g(1)N_g]}{E[N_g]^3} \\
&\quad+\frac{(E[E[\bar{Y}_g(0)N_g|X_g]E[N_g|X_g]]-E[\bar{Y}_g(0)N_g]E[N_g])E[\bar{Y}_g(0)N_g]}{E[N_g]^3} \\
&\quad-\frac{E[E[\bar{Y}_g(1)N_g|X_g]E[\bar{Y}_g(0)N_g|X_g]]-E[\bar{Y}_g(1)N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]E[N_g]} \\
&\quad+\frac{(E[E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g]]-E[\bar{Y}_g(1)N_g]E[N_g])E[\bar{Y}_g(0)N_g]}{E[N_g]E[N_g]^2} \\
&\quad+\frac{(E[E[\bar{Y}_g(0)N_g|X_g]E[N_g|X_g]]-E[\bar{Y}_g(0)N_g]E[N_g])E[\bar{Y}_g(1)N_g]}{E[N_g]^2E[N_g]} \\
&\quad-\frac{(E[E[N_g|X_g]E[N_g|X_g]]-E[N_g]E[N_g])E[\bar{Y}_g(1)N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]^2E[N_g]^2} \\
&=-\frac{E[E[\bar{Y}_g(1)N_g|X_g]^2]}{2E[N_g]^2}-\frac{E[E[N_g|X_g]^2]E[\bar{Y}_g(1)N_g]^2}{2E[N_g]^4}-\frac{E[E[\bar{Y}_g(0)N_g|X_g]^2]}{2E[N_g]^2}-\frac{E[E[N_g|X_g]^2]E[\bar{Y}_g(0)N_g]^2}{2E[N_g]^4} \\
&\quad+\frac{E[E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g]]E[\bar{Y}_g(1)N_g]}{E[N_g]^3}+\frac{E[E[\bar{Y}_g(0)N_g|X_g]E[N_g|X_g]]E[\bar{Y}_g(0)N_g]}{E[N_g]^3} \\
&\quad-\frac{E[E[\bar{Y}_g(1)N_g|X_g]E[\bar{Y}_g(0)N_g|X_g]]}{E[N_g]^2}+\frac{E[E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g]]E[\bar{Y}_g(0)N_g]}{E[N_g]^3} \\
&\quad+\frac{E[E[\bar{Y}_g(0)N_g|X_g]E[N_g|X_g]]E[\bar{Y}_g(1)N_g]}{E[N_g]^3}-\frac{E[E[N_g|X_g]^2]E[\bar{Y}_g(1)N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]^4} \\
&=-\frac{1}{2}E[E[\tilde{Y}_g(1)|X_g]^2]-\frac{1}{2}E[E[\tilde{Y}_g(0)|X_g]^2]-E[E[\tilde{Y}_g(1)|X_g]E[\tilde{Y}_g(0)|X_g]] \\
&=-\frac{1}{2}E[(E[\tilde{Y}_g(1)+\tilde{Y}_g(0)|X_g])^2]\ ,
\end{aligned}
$$

as desired. ■

Lemma C.3. *Suppose Q satisfies Assumptions 2.1 and 3.6 and the treatment assignment mechanism satisfies*

Assumptions 3.4–3.5. Define

$$
\begin{aligned}
\mathbb{L}_G^{\mathrm{YN1}} &= \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} (\bar{Y}_g(1) N_g D_g - E[\bar{Y}_g(1) N_g] D_g) \\
\mathbb{L}_G^{\mathrm{N1}} &= \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} (N_g D_g - E[N_g] D_g) \\
\mathbb{L}_G^{\mathrm{YN0}} &= \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} (\bar{Y}_g(0) N_g (1 - D_g) - E[\bar{Y}_g(0) N_g](1 - D_g)) \\
\mathbb{L}_G^{\mathrm{N0}} &= \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} (N_g (1 - D_g) - E[N_g](1 - D_g)) \ .
\end{aligned}
$$

Then, as $G \to \infty$,

$$
(\mathbb{L}_G^{\mathrm{YN1}}, \mathbb{L}_G^{\mathrm{N1}}, \mathbb{L}_G^{\mathrm{YN0}}, \mathbb{L}_G^{\mathrm{N0}})' \xrightarrow{d} N(0, \mathbb{V}) \ ,
$$

where

$$
\mathbb{V} = \mathbb{V}_1 + \mathbb{V}_2
$$

for

$$
\mathbb{V}_1 = \begin{pmatrix} \mathbb{V}_1^1 & 0 \\ 0 & \mathbb{V}_1^0 \end{pmatrix}
$$

$$
\begin{aligned}
\mathbb{V}_1^1 &= \begin{pmatrix} E[\mathrm{Var}[\bar{Y}_g(1) N_g | W_g]] & 0 \\ 0 & 0 \end{pmatrix} \\
\mathbb{V}_1^0 &= \begin{pmatrix} E[\mathrm{Var}[\bar{Y}_g(0) N_g | W_g]] & 0 \\ 0 & 0 \end{pmatrix}
\end{aligned}
$$

$$
\mathbb{V}_2 = \frac{1}{2} \mathrm{Var}[(E[\bar{Y}_g(1) N_g | W_g], N_g, E[\bar{Y}_g(0) N_g | W_g], N_g)'] \ .
$$

PROOF. We will only verify Steps 1 and 2 in the proof of Lemma C.1 because Steps 3–5 are identical. Note

$$
(\mathbb{L}_G^{\mathrm{YN1}}, \mathbb{L}_G^{\mathrm{N1}}, \mathbb{L}_G^{\mathrm{YN0}}, \mathbb{L}_G^{\mathrm{N0}}) = (\mathbb{L}_{1,G}^{\mathrm{YN1}}, 0, \mathbb{L}_{1,G}^{\mathrm{YN0}}, 0) + (\mathbb{L}_{2,G}^{\mathrm{YN1}}, \mathbb{L}_G^{\mathrm{N1}}, \mathbb{L}_{2,G}^{\mathrm{YN0}}, \mathbb{L}_G^{\mathrm{N0}}) \ ,
$$

where

$$
\begin{aligned}
\mathbb{L}_{1,G}^{\mathrm{YN1}} &= \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} (\bar{Y}_g(1) N_g D_g - E[\bar{Y}_g(1) N_g D_g | N^{(G)}, X^{(G)}, D^{(G)}]) \\
\mathbb{L}_{2,G}^{\mathrm{YN1}} &= \frac{1}{\sqrt{G}} \sum_{1 \le g \le 2G} (E[\bar{Y}_g(1) N_g D_g | N^{(G)}, X^{(G)}, D^{(G)}] - E[\bar{Y}_g(1) N_g] D_g)
\end{aligned}
$$

and similarly for $\mathbb{L}_G^{\mathrm{YN0}}$. Next, note $(\mathbb{L}_{1,G}^{\mathrm{YN1}}, 0, \mathbb{L}_{1,G}^{\mathrm{YN0}}, 0), G \ge 1$ is a triangular array of normalized sums of random vectors. Conditional on $N^{(G)}, X^{(G)}, D^{(G)}$, $\mathbb{L}_{1,G}^{\mathrm{YN1}} \perp\!\!\!\perp \mathbb{L}_{1,G}^{\mathrm{YN0}}$. Moreover, it follows from $Q_G = Q^{2G}$ and

Assumption 3.4 that

$$\mathrm{Var}\left[\mathbb{L}_{1,G}^{\mathrm{YN1}}\middle|N^{(G)},X^{(G)},D^{(G)}\right]=\mathrm{Var}[\bar{Y}_g(1)N_g|W_g]D_g\ .$$

We have

$$\frac{1}{G}\sum_{1\le g\le 2G}\mathrm{Var}[\bar{Y}_g(1)N_g|W_g]D_g=\frac{1}{G}\sum_{1\le g\le 2G}E[\bar{Y}_g^2(1)N_g^2|W_g]D_g-\frac{1}{G}\sum_{1\le g\le 2G}E[\bar{Y}_g(1)N_g|W_g]^2D_g\ . \tag{30}$$

Note

$$\begin{aligned}&\frac{1}{G}\sum_{1\le g\le 2G}E[\bar{Y}_g^2(1)N_g^2|W_g]D_g\\&=\frac{1}{2G}\sum_{1\le g\le 2G}E[\bar{Y}_g^2(1)N_g^2|W_g]+\frac{1}{2}\Big(\frac{1}{G}\sum_{1\le g\le 2G:D_g=1}E[\bar{Y}_g^2(1)N_g^2|W_g]-\frac{1}{G}\sum_{1\le g\le 2G:D_g=0}E[\bar{Y}_g^2(1)N_g^2|W_g]\Big)\ .\end{aligned}$$

It follows from the weak law of large numbers, the application of which is permitted by Lemma C.12,

$$\frac{1}{2G}\sum_{1\le g\le 2G}E[\bar{Y}_g^2(1)N_g^2|W_g]\xrightarrow{P}E[\bar{Y}_g^2(1)N_g^2]\ .$$

On the other hand,

$$\begin{aligned}&\Big|\frac{1}{G}\sum_{1\le g\le 2G:D_g=1}E[\bar{Y}_g^2(1)N_g^2|W_g]-\frac{1}{G}\sum_{1\le g\le 2G:D_g=0}E[\bar{Y}_g^2(1)N_g^2|W_g]\Big|\\&\le\frac{1}{G}\sum_{1\le j\le G}|N_{\pi(2j-1)}^2E[\bar{Y}_{\pi(2j-1)}^2(1)|W_{\pi(2j-1)}]-N_{\pi(2j)}^2E[\bar{Y}_{\pi(2j)}^2(1)|W_{\pi(2j)}]|\\&\le\frac{1}{G}\sum_{1\le j\le G}N_{\pi(2j)}^2|E[\bar{Y}_{\pi(2j-1)}^2(1)|W_{\pi(2j-1)}]-E[\bar{Y}_{\pi(2j)}^2(1)|W_{\pi(2j)}]|\\&\qquad+\frac{1}{G}\sum_{1\le j\le G}|N_{\pi(2j)}^2-N_{\pi(2j-1)}^2||E[\bar{Y}_{\pi(2j-1)}^2(1)|W_{\pi(2j-1)}]|\\&\lesssim\frac{1}{G}\sum_{1\le j\le G}N_{\pi(2j)}^2\|W_{\pi(2j-1)}-W_{\pi(2j)}\|+\frac{1}{G}\sum_{1\le j\le G}|N_{\pi(2j)}^2-N_{\pi(2j-1)}^2|\xrightarrow{P}0\ ,\end{aligned}$$

where the first inequality follows from Assumption 3.4 and the triangle inequality, the second inequality by some algebraic manipulations, the final inequality by Assumption 3.6 and Lemma C.12, and the convergence in probability follows from Assumption 3.5 and Lemmas C.13 and C.14. Therefore,

$$\frac{1}{G}\sum_{1\le g\le 2G}E[\bar{Y}_g^2(1)N_g^2|W_g]D_g\xrightarrow{P}E[\bar{Y}_g^2(1)N_g^2]\ .$$

Meanwhile,

$$\frac{1}{G}\sum_{1\le g\le 2G}E[\bar{Y}_g(1)N_g|W_g]^2D_g$$

$$= \frac{1}{2G} \sum_{1 \le g \le 2G} E[\bar{Y}_g(1)N_g|W_g]^2 + \frac{1}{2}\Big(\frac{1}{G} \sum_{1 \le g \le 2G : D_g = 1} E[\bar{Y}_g(1)N_g|W_g]^2 - \frac{1}{G} \sum_{1 \le g \le 2G : D_g = 0} E[\bar{Y}_g(1)N_g|W_g]^2\Big) .$$

It follows from the weak law of large numbers, the application of which is permitted by Lemma C.12 and Assumption 2.1(c) that

$$\frac{1}{2G} \sum_{1 \le g \le 2G} E[\bar{Y}_g(1)N_g|W_g]^2 \xrightarrow{P} E[E[\bar{Y}_g(1)N_g|W_g]^2] .$$

Next,

$$\begin{aligned}
&\Big|\frac{1}{G} \sum_{1 \le g \le 2G : D_g = 1} E[\bar{Y}_g(1)N_g|W_g]^2 - \frac{1}{G} \sum_{1 \le g \le 2G : D_g = 0} E[\bar{Y}_g(1)N_g|W_g]^2\Big| \\
&\le \frac{1}{G} \sum_{1 \le j \le G} |E[\bar{Y}_{\pi(2j-1)}(1)N_{\pi(2j-1)}|W_{\pi(2j-1)}] - E[\bar{Y}_{\pi(2j)}(1)N_{\pi(2j)}|W_{\pi(2j)}]| \\
&\qquad\qquad\qquad \times |E[\bar{Y}_{\pi(2j-1)}(1)N_{\pi(2j-1)}|W_{\pi(2j-1)}] + E[\bar{Y}_{\pi(2j)}(1)N_{\pi(2j)}|W_{\pi(2j)}]| \\
&\le \Big(\frac{1}{G} \sum_{1 \le j \le G} |E[\bar{Y}_{\pi(2j-1)}(1)N_{\pi(2j-1)}|W_{\pi(2j-1)}] - E[\bar{Y}_{\pi(2j)}(1)N_{\pi(2j)}|W_{\pi(2j)}]|^2\Big)^{1/2} \\
&\qquad \times \Big(\frac{1}{G} \sum_{1 \le j \le G} |E[\bar{Y}_{\pi(2j-1)}(1)N_{\pi(2j-1)}|W_{\pi(2j-1)}] + E[\bar{Y}_{\pi(2j)}(1)N_{\pi(2j)}|W_{\pi(2j)}]|^2\Big)^{1/2} \\
&\lesssim \Big(\frac{1}{G} \sum_{1 \le j \le G} |E[\bar{Y}_{\pi(2j-1)}(1)N_{\pi(2j-1)}|W_{\pi(2j-1)}] - E[\bar{Y}_{\pi(2j)}(1)N_{\pi(2j)}|W_{\pi(2j)}]|^2\Big)^{1/2} \\
&\qquad \times \Big(\frac{1}{G} \sum_{1 \le g \le 2G} E[\bar{Y}_g(1)N_g|W_g]^2\Big)^{1/2} \xrightarrow{P} 0 ,
\end{aligned}$$

where the first inequality follows by inspection, the second follows from Cauchy-Schwarz, the third follows from $(a+b)^2 \le 2a^2 + 2b^2$, and the convergence in probability follows from Assumptions 3.5–3.6, Lemma C.13, and the weak law of large numbers. Therefore,

$$\frac{1}{G} \sum_{1 \le g \le 2G} E[\bar{Y}_g(1)N_g|W_g]^2 D_g \xrightarrow{P} E\left[E[\bar{Y}_g(1)N_g|W_g]^2\right] ,$$

and hence it follows from (30) that

$$\frac{1}{G} \sum_{1 \le g \le 2G} \text{Var}[\bar{Y}_g(1)N_g|W_g] D_g \xrightarrow{P} E[\text{Var}[\bar{Y}_g(1)N_g|W_g]] .$$

Similarly,

$$\frac{1}{G} \sum_{1 \le g \le 2G} \text{Var}[\bar{Y}_g(0)N_g|W_g] D_g \xrightarrow{P} E[\text{Var}[\bar{Y}_g(0)N_g|W_g]] .$$

Putting these results together, we obtain

$$\text{Var}[(\mathbb{L}^{\text{YN1}}_{1,G}, 0, \mathbb{L}^{\text{YN0}}_{1,G}, 0)'|W^{(G)}, D^{(G)}] \xrightarrow{P} \mathbb{V}_1 .$$

The rest of the proof is identical to Steps 3–5 in the proof of Lemma C.1 and is omitted. ■

Lemma C.4. *Let* $\mathbb{V}$ *be defined as in Lemma C.3, and* D_{h0} *be defined as in the proof of Theorem 3.1, then*

$$D_{h0}\mathbb{V}D'_{h0} = \nu^2 \ ,$$

where

$$\nu^2 = E[\tilde{Y}_g^2(1)] + E[\tilde{Y}_g^2(0)] - \frac{1}{2}E[(E[\tilde{Y}_g(1) + \tilde{Y}_g(0)|W_g])^2] \ .$$

Proof. $\mathbb{V}$ in Lemma C.3 is symmetric with entries

$$\begin{aligned}
\mathbb{V}_{11} &= \text{Var}[\bar{Y}_g(1)N_g] - \frac{1}{2}\,\text{Var}[E[\bar{Y}_g(1)N_g|W_g]] \\
\mathbb{V}_{12} &= \text{Cov}[E[\bar{Y}_g(1)N_g|W_g], N_g] - \frac{1}{2}\,\text{Cov}[E[\bar{Y}_g(1)N_g|W_g], N_g] \\
\mathbb{V}_{13} &= \frac{1}{2}\,\text{Cov}[E[\bar{Y}_g(1)N_g|W_g], E[\bar{Y}_g(0)N_g|W_g]] \\
\mathbb{V}_{14} &= \frac{1}{2}\,\text{Cov}[E[\bar{Y}_g(1)N_g|W_g], N_g] \\
\mathbb{V}_{22} &= \text{Var}[N_g] - \frac{1}{2}\,\text{Var}[N_g] \\
\mathbb{V}_{23} &= \frac{1}{2}\,\text{Cov}[N_g, E[\bar{Y}_g(0)N_g|X_g]] \\
\mathbb{V}_{24} &= \frac{1}{2}\,\text{Var}[N_g] \\
\mathbb{V}_{33} &= \text{Var}[\bar{Y}_g(0)N_g] - \frac{1}{2}\,\text{Var}[E[\bar{Y}_g(0)N_g|W_g]] \\
\mathbb{V}_{34} &= \text{Cov}[E[\bar{Y}_g(0)N_g|W_g], N_g] - \frac{1}{2}\,\text{Cov}[E[\bar{Y}_g(0)N_g|W_g], N_g] \\
\mathbb{V}_{44} &= \text{Var}[N_g] - \frac{1}{2}\,\text{Var}[N_g] \ .
\end{aligned}$$

We proceed by mirroring the algebra in Lemma C.2. Expanding and simplifying the first half of the expression:

$$\begin{aligned}
&\frac{\text{Var}[\bar{Y}_g(1)N_g]}{E[N_g]^2} + \frac{\text{Var}[N_g]E[\bar{Y}_g(1)N_g]^2}{E[N_g]^4} + \frac{\text{Var}[\bar{Y}_g(0)N_g]}{E[N_g]^2} + \frac{\text{Var}[N_g]E[\bar{Y}_g(0)N_g]^2}{E[N_g]^4} \\
&\qquad - \frac{2\,\text{Cov}[E[\bar{Y}_g(1)N_g|W_g], N_g]E[\bar{Y}_g(1)N_g]}{E[N_g]^3} - \frac{2\,\text{Cov}[E[\bar{Y}_g(0)N_g|W_g], N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]^3} \\
&= \frac{E[\bar{Y}_g^2(1)N_g^2] - E[\bar{Y}_g(1)N_g]^2}{E[N_g]^2} + \frac{E[N_g^2]E[\bar{Y}_g(1)N_g]^2 - E[N_g]^2E[\bar{Y}_g(1)N_g]^2}{E[N_g]^4} \\
&\qquad + \frac{E[\bar{Y}_g^2(0)N_g^2] - E[\bar{Y}_g(0)N_g]^2}{E[N_g]^2} + \frac{E[N_g^2]E[\bar{Y}_g(0)N_g]^2 - E[N_g]^2E[\bar{Y}_g(0)N_g]^2}{E[N_g]^4} \\
&\qquad - \frac{2E[\bar{Y}_g(1)N_g^2]E[\bar{Y}_g(1)N_g]}{E[N_g]^3} + \frac{2E[\bar{Y}_g(1)N_g]E[N_g]E[\bar{Y}_g(1)N_g]}{E[N_g]^3} \\
&\qquad - \frac{2E[\bar{Y}_g(0)N_g^2]E[\bar{Y}_g(0)N_g]}{E[N_g]^3} + \frac{2E[\bar{Y}_g(0)N_g]E[N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]^3} \\
&= \frac{E[\bar{Y}_g^2(1)N_g^2]}{E[N_g]^2} + \frac{E[\bar{Y}_g^2(0)N_g^2]}{E[N_g]^2} + \frac{E[N_g^2]E[\bar{Y}_g(1)N_g]^2}{E[N_g]^4} + \frac{E[N_g^2]E[\bar{Y}_g(0)N_g]^2}{E[N_g]^4} \\
&\qquad - \frac{2E[\bar{Y}_g(1)N_g^2]E[\bar{Y}_g(1)N_g]}{E[N_g]^3} - \frac{2E[\bar{Y}_g(0)N_g^2]E[\bar{Y}_g(0)N_g]}{E[N_g]^3}
\end{aligned}$$

$$= E[\tilde{Y}_g^2(1)] + E[\tilde{Y}_g^2(0)] \ ,$$

where

$$\tilde{Y}_g(d) = \frac{N_g}{E[N_g]}\left(\bar{Y}_g(d) - \frac{E[\bar{Y}_g(d)N_g]}{E[N_g]}\right)$$

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

Expanding the second half of the expression:

$$\begin{aligned}
&- \frac{\text{Var}[E[\bar{Y}_g(1)N_g|W_g]]}{2E[N_g]^2} - \frac{\text{Var}[N_g]E[\bar{Y}_g(1)N_g]^2}{2E[N_g]^4} \\
&\quad - \frac{\text{Var}[E[\bar{Y}_g(0)N_g|W_g]]}{2E[N_g]^2} - \frac{\text{Var}[N_g]E[\bar{Y}_g(0)N_g]^2}{2E[N_g]^4} \\
&\quad + \frac{\text{Cov}[E[\bar{Y}_g(1)N_g|W_g], N_g]E[\bar{Y}_g(1)N_g]}{E[N_g]^3} + \frac{\text{Cov}[E[\bar{Y}_g(0)N_g|W_g], N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]^3} \\
&\quad - \frac{\text{Cov}[E[\bar{Y}_g(1)N_g|W_g], E[\bar{Y}_g(0)N_g|W_g]]}{E[N_g]^2} + \frac{\text{Cov}[E[\bar{Y}_g(1)N_g|W_g], N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]E[N_g]^2} \\
&\quad + \frac{\text{Cov}[N_g, E[\bar{Y}_g(0)N_g|W_g]]E[\bar{Y}_g(1)N_g]}{E[N_g]^2E[N_g]} - \frac{\text{Cov}[N_g, N_g]E[\bar{Y}_g(1)N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]^2E[N_g]^2} \\
&= -\frac{E[E[\bar{Y}_g(1)N_g|W_g]^2] - E[\bar{Y}_g(1)N_g]^2}{2E[N_g]^2} - \frac{(E[N_g^2] - E[N_g]^2)E[\bar{Y}_g(1)N_g]^2}{2E[N_g]^4} \\
&\quad - \frac{E[E[\bar{Y}_g(0)N_g|W_g]^2] - E[\bar{Y}_g(0)N_g]^2}{2E[N_g]^2} - \frac{(E[N_g^2] - E[N_g]^2)E[\bar{Y}_g(0)N_g]^2}{2E[N_g]^4} \\
&\quad + \frac{(E[E[\bar{Y}_g(1)N_g|W_g]N_g] - E[\bar{Y}_g(1)N_g]E[N_g])E[\bar{Y}_g(1)N_g]}{E[N_g]^3} \\
&\quad + \frac{(E[E[\bar{Y}_g(0)N_g|W_g]N_g] - E[\bar{Y}_g(0)N_g]E[N_g])E[\bar{Y}_g(0)N_g]}{E[N_g]^3} \\
&\quad - \frac{E[E[\bar{Y}_g(1)N_g|W_g]E[\bar{Y}_g(0)N_g|W_g]] - E[\bar{Y}_g(1)N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]E[N_g]} \\
&\quad + \frac{(E[E[\bar{Y}_g(1)N_g|W_g]N_g] - E[\bar{Y}_g(1)N_g]E[N_g])E[\bar{Y}_g(0)N_g]}{E[N_g]E[N_g]^2} \\
&\quad + \frac{(E[E[\bar{Y}_g(0)N_g|W_g]N_g] - E[\bar{Y}_g(0)N_g]E[N_g])E[\bar{Y}_g(1)N_g]}{E[N_g]^2E[N_g]} \\
&\quad - \frac{(E[N_g^2] - E[N_g]^2)E[\bar{Y}_g(1)N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]^2E[N_g]^2} \\
&= -\frac{E[E[\bar{Y}_g(1)N_g|W_g]^2]}{2E[N_g]^2} - \frac{E[N_g^2]E[\bar{Y}_g(1)N_g]^2}{2E[N_g]^4} - \frac{E[E[\bar{Y}_g(0)N_g|W_g]^2]}{2E[N_g]^2} - \frac{E[N_g^2]E[\bar{Y}_g(0)N_g]^2}{2E[N_g]^4} \\
&\quad + \frac{E[E[\bar{Y}_g(1)N_g|W_g]N_g]E[\bar{Y}_g(1)N_g]}{E[N_g]^3} + \frac{E[E[\bar{Y}_g(0)N_g|W_g]N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]^3} \\
&\quad - \frac{E[E[\bar{Y}_g(1)N_g|W_g]E[\bar{Y}_g(0)N_g|W_g]]}{E[N_g]^2} + \frac{E[E[\bar{Y}_g(1)N_g|W_g]N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]^3} \\
&\quad + \frac{E[E[\bar{Y}_g(0)N_g|W_g]N_g]E[\bar{Y}_g(1)N_g]}{E[N_g]^3} - \frac{E[N_g^2]E[\bar{Y}_g(1)N_g]E[\bar{Y}_g(0)N_g]}{E[N_g]^4} \\
&= -\frac{1}{2}E[E[\tilde{Y}_g(1)|W_g]^2] - \frac{1}{2}E[E[\tilde{Y}_g(0)|W_g]^2] - E[E[\tilde{Y}_g(1)|W_g]E[\tilde{Y}_g(0)|W_g]]
\end{aligned}$$

$$= -\frac{1}{2}E[(E[\tilde{Y}_g(1) + \tilde{Y}_g(0)|W_g])^2] ,$$

as desired. ■

Lemma C.5. *Consider the following adjusted potential outcomes:*

$$\hat{Y}_g(d) = \frac{N_g}{\frac{1}{2G}\sum_{1\le j\le 2G} N_j}\left(\bar{Y}_g(d) - \frac{\frac{1}{G}\sum_{1\le j\le 2G}\bar{Y}_j(d)I\{D_j = d\}N_j}{\frac{1}{G}\sum_{1\le j\le 2G} I\{D_j = d\}N_j}\right) .$$

Note the usual relationship still holds for adjusted outcomes, i.e. $\hat{Y}_g = D_g\hat{Y}_g(1) + (1 - D_g)\hat{Y}_g(0)$. *If Assumptions 2.1 holds, and additionally Assumptions 3.2–3.3 (or Assumptions 3.5–3.6) hold, then*

$$\hat{\mu}_G(d) = \frac{1}{G}\sum_{1\le g\le 2G}\hat{Y}_g(d)I\{D_g = d\} \xrightarrow{P} 0$$

$$\hat{\sigma}^2_G(d) = \frac{1}{G}\sum_{1<g<2G}\left(\hat{Y}_g - \hat{\mu}_G(d)\right)^2 I\{D_g = d\} \xrightarrow{P} \text{Var}\left[\tilde{Y}_g(d)\right] .$$

Proof. It suffices to show that

$$\frac{1}{G}\sum_{1\le g\le 2G}\hat{Y}^r_g(d)I\{D_g = d\} \xrightarrow{P} E\left[\tilde{Y}^r_g(d)\right]$$

for $r \in \{1, 2\}$. We prove this result only for $r = 1$ and $d = 1$; the other cases can be proven similarly. To this end, write

$$\frac{1}{G}\sum_{1\le g\le 2G}\hat{Y}_g(1)I\{D_g = 1\} = \frac{1}{G}\sum_{1\le g\le 2G}\hat{Y}_g(1)D_g = \frac{1}{G}\sum_{1\le g\le 2G}\tilde{Y}_g(1)D_g + \frac{1}{G}\sum_{1\le g\le 2G}\left(\hat{Y}_g(1) - \tilde{Y}_g(1)\right)D_g .$$

Note that

$$\frac{1}{G}\sum_{1\le g\le 2G}\left(\hat{Y}_g(1) - \tilde{Y}_g(1)\right)D_g = \left(\frac{1}{\frac{1}{2G}\sum_{1\le g\le 2G} N_g} - \frac{1}{E[N_g]}\right)\left(\frac{1}{G}\sum_{1\le g\le 2G}\bar{Y}_g(1)N_gD_g\right)$$
$$-\left(\frac{\frac{1}{G}\sum_{1\le g\le 2G}\bar{Y}_g(d)I\{D_g = d\}N_g}{\left(\frac{1}{2G}\sum_{1\le g\le 2G} N_g\right)^2} - \frac{E[\bar{Y}_g(d)N_g]}{E[N_g]^2}\right)\left(\frac{1}{G}\sum_{1\le g\le 2G} N_gD_g\right)$$

By the weak law of large numbers, Lemma C.15 and Slutsky's theorem, we have

$$\frac{1}{G}\sum_{1\le g\le 2G}\left(\hat{Y}_g(1) - \tilde{Y}_g(1)\right)D_g \xrightarrow{P} 0 .$$

Lemma C.15 implies

$$\frac{1}{G}\sum_{1\le g\le 2G}\tilde{Y}_g(d)D_g \xrightarrow{P} E\left[\tilde{Y}_g(d)\right] = 0 .$$

Thus, the result follows. ■

Lemma C.6. *If Assumptions 2.1 holds, and Assumptions 3.2-3.3 hold, then*

$$\hat{\tau}_G^2 \xrightarrow{P} E\left[\text{Var}\left[\tilde{Y}_g(1)\middle|X_g\right]\right] + E\left[\text{Var}\left[\tilde{Y}_g(0)\middle|X_g\right]\right] + E\left[\left(E\left[\tilde{Y}_g(1)\middle|X_g\right] - E\left[\tilde{Y}_g(0)\middle|X_g\right]\right)^2\right]$$

in the case where we match on cluster size. Instead, if Assumptions 2.1 and 3.5-3.6 hold, then

$$\hat{\tau}_G^2 \xrightarrow{P} E\left[\text{Var}\left[\tilde{Y}_g(1)\middle|W_g\right]\right] + E\left[\text{Var}\left[\tilde{Y}_g(0)\middle|W_g\right]\right] + E\left[\left(E\left[\tilde{Y}_g(1)\middle|W_g\right] - E\left[\tilde{Y}_g(0)\middle|W_g\right]\right)^2\right]$$

in the case where we do not match on cluster size.

PROOF. Note that

$$\hat{\tau}_G^2 = \frac{1}{G}\sum_{1\le j\le G}\left(\hat{Y}_{\pi(2j)} - \hat{Y}_{\pi(2j-1)}\right)^2 = \frac{1}{G}\sum_{1\le g\le 2G}\hat{Y}_g^2 - \frac{2}{G}\sum_{1\le j\le G}\hat{Y}_{\pi(2j)}\hat{Y}_{\pi(2j-1)}.$$

Since

$$\frac{1}{G}\sum_{1\le g\le 2G}\hat{Y}_g^2 = \hat{\sigma}_G^2(1) - \hat{\mu}_G^2(1) + \hat{\sigma}_G^2(0) - \hat{\mu}_G^2(0)$$

It follows from Lemma C.5 that

$$\frac{1}{G}\sum_{1\le g\le 2G}\hat{Y}_g^2 \xrightarrow{P} E[\tilde{Y}_g^2(1)] + E[\tilde{Y}_g^2(0)]$$

Next, we argue that

$$\frac{2}{G}\sum_{1\le j\le G}\hat{Y}_{\pi(2j)}\hat{Y}_{\pi(2j-1)} \xrightarrow{P} 2E[\mu_1(W_g)\mu_0(W_g)]\ ,$$

where we use the notation $\mu_d(W_g)$ to denote $E[\tilde{Y}_g(d)|W_g]$. To this end, first note that

$$\frac{2}{G}\sum_{1\le j\le G}\hat{Y}_{\pi(2j)}\hat{Y}_{\pi(2j-1)} = \frac{2}{G}\sum_{1\le j\le G}\tilde{Y}_{\pi(2j)}\tilde{Y}_{\pi(2j-1)} + \frac{2}{G}\sum_{1\le j\le G}\left(\hat{Y}_{\pi(2j)}\hat{Y}_{\pi(2j-1)} - \tilde{Y}_{\pi(2j)}\tilde{Y}_{\pi(2j-1)}\right)\ .$$

Note that

$$\begin{aligned}
&\frac{2}{G}\sum_{1\le j\le G}\left(\hat{Y}_{\pi(2j)}(1)\hat{Y}_{\pi(2j-1)}(0) - \tilde{Y}_{\pi(2j)}(1)\tilde{Y}_{\pi(2j-1)}(0)\right)D_{\pi(2j)} \\
&= \frac{2}{G}\sum_{1\le j\le G}\left(\left(\hat{Y}_{\pi(2j)}(1) - \tilde{Y}_{\pi(2j)}(1)\right)\hat{Y}_{\pi(2j-1)}(0)D_{\pi(2j)} + \left(\hat{Y}_{\pi(2j-1)}(0) - \tilde{Y}_{\pi(2j-1)}(0)\right)\tilde{Y}_{\pi(2j)}(1)D_{\pi(2j)}\right) \\
&= \frac{2}{G}\sum_{1\le j\le G}\Bigg(\left(\hat{Y}_{\pi(2j)}(1) - \tilde{Y}_{\pi(2j)}(1)\right)\tilde{Y}_{\pi(2j-1)}(0)D_{\pi(2j)} \\
&\qquad + \left(\hat{Y}_{\pi(2j)}(1) - \tilde{Y}_{\pi(2j)}(1)\right)\left(\hat{Y}_{\pi(2j-1)}(0) - \tilde{Y}_{\pi(2j-1)}(0)\right)D_{\pi(2j)} \\
&\qquad + \left(\hat{Y}_{\pi(2j-1)}(0) - \tilde{Y}_{\pi(2j-1)}(0)\right)\tilde{Y}_{\pi(2j)}(1)D_{\pi(2j)}\Bigg)\ ,
\end{aligned}$$

for which the first term is given as follows:

$$\frac{2}{G}\sum_{1\le j\le G}\left(\hat{Y}_{\pi(2j)}(1)-\tilde{Y}_{\pi(2j)}(1)\right)\tilde{Y}_{\pi(2j-1)}(0)D_{\pi(2j)}$$

$$=\left(\frac{1}{\frac{1}{2G}\sum_{1\le g\le 2G}N_g}-\frac{1}{E[N_g]}\right)\left(\frac{2}{G}\sum_{1\le j\le G}N_{\pi(2j)}\bar{Y}_{\pi(2j)}(1)\tilde{Y}_{\pi(2j-1)}(0)D_{\pi(2j)}\right)$$

$$-\left(\frac{\frac{1}{2G}\sum_{1\le g\le 2G}\bar{Y}_g(1)I\{D_g=1\}N_g}{\left(\frac{1}{2G}\sum_{1\le g\le 2G}N_g\right)^2}-\frac{E[\bar{Y}_g(1)N_g]}{E[N_g]^2}\right)\left(\frac{2}{G}\sum_{1\le j\le G}N_{\pi(2j)}\tilde{Y}_{\pi(2j-1)}(0)D_{\pi(2j)}\right) .$$

Lemma C.16 implies

$$\frac{2}{G}\sum_{1\le j\le G}N_{\pi(2j)}\bar{Y}_{\pi(2j)}(1)\tilde{Y}_{\pi(2j-1)}(0)D_{\pi(2j)}\xrightarrow{P}E[E[N_g\bar{Y}_g(1)|X_g]E[\tilde{Y}_g(0)|X_g]]$$
$$\frac{2}{G}\sum_{1\le j\le G}N_{\pi(2j)}\tilde{Y}_{\pi(2j-1)}(0)D_{\pi(2j)}\xrightarrow{P}E[E[N_g|X_g]E[\tilde{Y}_g(0)|X_g]]$$

for the case of not matching on cluster sizes. For the case where we match on cluster sizes,

$$\frac{2}{G}\sum_{1\le j\le G}N_{\pi(2j)}\bar{Y}_{\pi(2j)}(1)\tilde{Y}_{\pi(2j-1)}(0)D_{\pi(2j)}\xrightarrow{P}E[N_gE[\bar{Y}_g(1)|W_g]E[\tilde{Y}_g(0)|W_g]]$$
$$\frac{2}{G}\sum_{1\le j\le G}N_{\pi(2j)}\tilde{Y}_{\pi(2j-1)}(0)D_{\pi(2j)}\xrightarrow{P}E[N_gE[\tilde{Y}_g(0)|W_g]]$$

Then, by the weak law of large numbers, Lemma C.15, and the continuous mapping theorem, we have

$$\frac{2}{G}\sum_{1\le j\le G}\left(\hat{Y}_{\pi(2j)}(1)-\tilde{Y}_{\pi(2j)}(1)\right)\tilde{Y}_{\pi(2j-1)}(0)D_{\pi(2j)}\xrightarrow{P}0 .$$

By repeating the same arguments for the other two terms, we conclude that

$$\frac{2}{G}\sum_{1\le j\le G}\left(\hat{Y}_{\pi(2j)}(1)\hat{Y}_{\pi(2j-1)}(0)-\tilde{Y}_{\pi(2j)}(1)\tilde{Y}_{\pi(2j-1)}(0)\right)D_{\pi(2j)}\xrightarrow{P}0 ,$$

which immediately implies

$$\frac{2}{G}\sum_{1\le j\le G}\hat{Y}_{\pi(2j)}\hat{Y}_{\pi(2j-1)}-\tilde{Y}_{\pi(2j)}\tilde{Y}_{\pi(2j-1)}\xrightarrow{P}0 .$$

Thus, it is left to show that

$$\frac{2}{G}\sum_{1\le j\le G}\tilde{Y}_{\pi(2j)}\tilde{Y}_{\pi(2j-1)}\xrightarrow{P}2E[\mu_1(W_g)\mu_0(W_g)] ,$$

for the case of matching on cluster sizes, and for the case of not matching on cluster size,

$$\frac{2}{G}\sum_{1\le j\le G}\tilde{Y}_{\pi(2j)}\tilde{Y}_{\pi(2j-1)}\xrightarrow{P}2E[\mu_1(X_g)\mu_0(X_g)]\ ,$$

both of which follow from Lemmas C.16 and C.17. Hence, in the case where we match on cluster size,

$$\begin{aligned}\hat{\tau}_n^2 &\xrightarrow{P} E\left[\tilde{Y}_g^2(1)\right]+E\left[\tilde{Y}_g^2(0)\right]-2E\left[\mu_1\left(W_g\right)\mu_0\left(W_g\right)\right]\\ &=E\left[\operatorname{Var}\left[\tilde{Y}_g(1)|W_g\right]\right]+E\left[\operatorname{Var}\left[\tilde{Y}_g(0)|W_g\right]\right]+E\left[\left(\mu_1\left(W_g\right)-\mu_0\left(W_g\right)\right)^2\right]\\ &=E\left[\operatorname{Var}\left[\tilde{Y}_g(1)|W_g\right]\right]+E\left[\operatorname{Var}\left[\tilde{Y}_g(0)|W_g\right]\right]+E\left[\left(E\left[\tilde{Y}_g(1)|X_i\right]-E\left[\tilde{Y}_g(0)|W_g\right]\right)^2\right]\ .\end{aligned}$$

And the corresponding result holds in the case where we do not match on cluster size. ■

Lemma C.7. *If Assumptions 2.1 holds, and Assumptions 3.2-3.3, 3.7 hold, then*

$$\hat{\lambda}_G^2\xrightarrow{P}E\left[\left(E\left[\tilde{Y}_g(1)|X_g\right]-E\left[\tilde{Y}_g(0)|X_g\right]\right)^2\right]$$

in the case where we do not match on cluster size. Instead, if Assumptions 3.5-3.6, 3.8 hold, then

$$\hat{\lambda}_G^2\xrightarrow{P}E\left[\left(E\left[\tilde{Y}_g(1)|W_g\right]-E\left[\tilde{Y}_g(0)|W_g\right]\right)^2\right]$$

in the case where we match on cluster size.

Proof. Note that

$$\begin{aligned}\hat{\lambda}_G^2 =&\frac{2}{G}\sum_{1\le j\le\lfloor G/2\rfloor}\left(\left(\hat{Y}_{\pi(4j-3)}-\hat{Y}_{\pi(4j-2)}\right)\left(\hat{Y}_{\pi(4j-1)}-\hat{Y}_{\pi(4j)}\right)\left(D_{\pi(4j-3)}-D_{\pi(4j-2)}\right)\left(D_{\pi(4j-1)}-D_{\pi(4j)}\right)\right)\\ =&\underbrace{\frac{2}{G}\sum_{1\le j\le\lfloor G/2\rfloor}\left(\left(\tilde{Y}_{\pi(4j-3)}-\tilde{Y}_{\pi(4j-2)}\right)\left(\tilde{Y}_{\pi(4j-1)}-\tilde{Y}_{\pi(4j)}\right)\left(D_{\pi(4j-3)}-D_{\pi(4j-2)}\right)\left(D_{\pi(4j-1)}-D_{\pi(4j)}\right)\right)}_{:=\tilde{\lambda}_G^2}\\ &+\frac{2}{G}\sum_{1\le j\le\lfloor G/2\rfloor}\left(\left(\left(\hat{Y}_{\pi(4j-3)}-\hat{Y}_{\pi(4j-2)}\right)\left(\hat{Y}_{\pi(4j-1)}-\hat{Y}_{\pi(4j)}\right)-\left(\tilde{Y}_{\pi(4j-3)}-\tilde{Y}_{\pi(4j-2)}\right)\left(\tilde{Y}_{\pi(4j-1)}-\tilde{Y}_{\pi(4j)}\right)\right)\right.\\ &\left.\times\left(D_{\pi(4j-3)}-D_{\pi(4j-2)}\right)\left(D_{\pi(4j-1)}-D_{\pi(4j)}\right)\right)\end{aligned}$$

Note that

$$\begin{aligned}&\left(\hat{Y}_{\pi(4j-3)}(1)-\hat{Y}_{\pi(4j-2)}(0)\right)\left(\hat{Y}_{\pi(4j-1)}(1)-\hat{Y}_{\pi(4j)}(0)\right)D_{\pi(4j-3)}D_{\pi(4j-1)}\\ &\quad-\left(\tilde{Y}_{\pi(4j-3)}(1)-\tilde{Y}_{\pi(4j-2)}(0)\right)\left(\tilde{Y}_{\pi(4j-1)}(1)-\tilde{Y}_{\pi(4j)}(0)\right)D_{\pi(4j-3)}D_{\pi(4j-1)}\\ &=\left(\hat{Y}_{\pi(4j-3)}(1)-\hat{Y}_{\pi(4j-2)}(0)-\left(\tilde{Y}_{\pi(4j-3)}(1)-\tilde{Y}_{\pi(4j-2)}(0)\right)\right)\left(\tilde{Y}_{\pi(4j-1)}(1)-\tilde{Y}_{\pi(4j)}(0)\right)D_{\pi(4j-3)}D_{\pi(4j-1)}\\ &+\left(\hat{Y}_{\pi(4j-3)}(1)-\hat{Y}_{\pi(4j-2)}(0)-\left(\tilde{Y}_{\pi(4j-3)}(1)-\tilde{Y}_{\pi(4j-2)}(0)\right)\right)\end{aligned}$$

$$\times \left(\hat{Y}_{\pi(4j-1)}(1) - \hat{Y}_{\pi(4j)}(0) - \left(\tilde{Y}_{\pi(4j-1)}(1) - \tilde{Y}_{\pi(4j)}(0)\right)\right) D_{\pi(4j-3)} D_{\pi(4j-1)}$$

$$+ \left(\hat{Y}_{\pi(4j-1)}(1) - \hat{Y}_{\pi(4j)}(0) - \left(\tilde{Y}_{\pi(4j-1)}(1) - \tilde{Y}_{\pi(4j)}(0)\right)\right) \left(\tilde{Y}_{\pi(4j-3)}(1) - \tilde{Y}_{\pi(4j-2)}(0)\right) D_{\pi(4j-3)} D_{\pi(4j-1)} .$$

Then we can show that each term converges to zero in probability by repeating the arguments in Lemma C.6. Similar arguments imply the same result holds for other cross products, which implies $\hat{\lambda}_G^2 - \tilde{\lambda}_G^2 \xrightarrow{P} 0$. Finally, by Lemma S.1.7 of Bai et al. (2022) and Lemma C.17, we have

$$\hat{\lambda}_G^2 = \tilde{\lambda}_G^2 + o_P(1) \xrightarrow{P} E\left[\left(E\left[\tilde{Y}_g(1)\middle|W_g\right] - E\left[\tilde{Y}_g(0)\middle|W_g\right]\right)^2\right]$$

in the case where we match on cluster size, and

$$\hat{\lambda}_G^2 = \tilde{\lambda}_G^2 + o_P(1) \xrightarrow{P} E\left[\left(E\left[\tilde{Y}_g(1)\middle|X_g\right] - E\left[\tilde{Y}_g(0)\middle|X_g\right]\right)^2\right]$$

in the case where we do not match on cluster size. ■

Lemma C.8. *Let $\tilde{R}_G(t)$ denote the randomization distribution of $\sqrt{G}\hat{\Delta}_G$ (see equation (21)). Then under the null hypothesis (10), we have that*

$$\sup_{t \in \mathbf{R}} |\tilde{R}_G(t) - \Phi(t/\tau)| \xrightarrow{P} 0 ,$$

where, in the case where we match on cluster size,

$$\tau^2 = E[\text{Var}[\tilde{Y}_g(1)|W_g]] + E[\text{Var}[\tilde{Y}_g(0)|W_g]] + E\left[(E[\tilde{Y}_g(1)|W_g] - E[\tilde{Y}_g(0)|X_g])^2\right] ,$$

and in the case where we do not match on cluster size,

$$\tau^2 = E[\text{Var}[\tilde{Y}_g(1)|X_g]] + E[\text{Var}[\tilde{Y}_g(0)|X_g]] + E\left[(E[\tilde{Y}_g(1)|X_g] - E[\tilde{Y}_g(0)|X_g])^2\right] .$$

PROOF. For a random transformation of the data, it follows as a consequence of Lemma C.15 that

$$\frac{1}{G} \sum_{1 \le g \le 2G} I\{\tilde{D}_g = d\} N_g \xrightarrow{P} E[N_g] ,$$
$$\frac{1}{G} \sum_{1 \le g \le 2G} (1 - \tilde{D}_g) N_g \bar{Y}_g \xrightarrow{P} E[N_g \bar{Y}_g(0)] .$$

Combining this with Lemma C.18 and a straightforward modification of Lemma A.3. in Chung and Romano (2013) to two dimensional distributions, we obtain that

$$\sup_{t \in \mathbf{R}} |\tilde{R}_G(t) - \Phi(t/\tau)| \xrightarrow{P} 0 ,$$

where when we match on cluster size

$$\tau^2 = \frac{1}{E[N_g]^2} \left(E[\text{Var}(N_g \bar{Y}_g(1)|W_g)] + E[\text{Var}(N_g \bar{Y}_g(0)|W_g)] + E\left[(E[N_g \bar{Y}_g(1)|W_g] - E[N_g \bar{Y}_g(0)|W_g])^2\right]\right) ,$$

and when we do *not* match on cluster size

$$\tau^2 = \frac{1}{E[N_g]^2}\Big(E[\mathrm{Var}(N_g\bar{Y}_g(1)|X_g)] + E[\mathrm{Var}(N_g\bar{Y}_g(0)|X_g)] + E\left[(E[N_g\bar{Y}_g(1)|X_g] - E[N_g\bar{Y}_g(0)|X_g])^2\right] +$$
$$- 2\frac{E[N_g\bar{Y}_g(0)]}{E[N_g]}\left(E[N_g^2\bar{Y}_g(1)] + E[N_g^2\bar{Y}_g(0)]\right.$$
$$\left.- \left(E\left[E[N_g\bar{Y}_g(1)|X_g]E[N_g|X_g]\right] + E\left[E[N_g\bar{Y}_g(0)|X_g]E[N_g|X_g]\right]\right)\right) + \left(\frac{E[N_g\bar{Y}_g(0)]}{E[N_g]}\right)^2 2E[\mathrm{Var}(N_g|X_g)]\Big) .$$

Note than, since under the null, $E[N_g\bar{Y}_g(1)] = E[N_g\bar{Y}_g(0)]$, we obtain

$$\begin{aligned}
&E[\mathrm{Var}[\tilde{Y}_g(1)|X_g]] + E[\mathrm{Var}[\tilde{Y}_g(0)|X_g]] + E[(E[\tilde{Y}_g(1)|X_g] - E[\tilde{Y}_g(0)|X_g])^2] \\
&= \frac{E[\mathrm{Var}[N_g\bar{Y}_g(1)|X_g]]}{E[N_g]^2} + \frac{E[\mathrm{Var}[N_g\bar{Y}_g(0)|X_g]]}{E[N_g]^2} + \frac{2E[\mathrm{Var}[N_g|X_g]]E[N_g\bar{Y}_g(d)]^2}{E[N_g]^4} \\
&\quad + \frac{E[(E[N_g\bar{Y}_g(1)|X_g] - E[N_g\bar{Y}_g(0)|X_g])^2}{E[N_g]^2} \\
&\quad - 2\frac{E[N_g\bar{Y}_g(1)](E[N_g^2\bar{Y}_g(1)] - E[E[N_g\bar{Y}_g(1)|X_g]E[N_g|X_g]])}{E[N_g]^3} \\
&\quad - 2\frac{E[N_g\bar{Y}_g(0)](E[N_g^2\bar{Y}_g(0)] - E[E[N_g\bar{Y}_g(0)|X_g]E[N_g|X_g]])}{E[N_g]^3} .
\end{aligned}$$

The result then follows immediately. ■

Lemma C.9. *Let $\check{v}_G^2(\epsilon_1, \ldots, \epsilon_G)$ be defined as in equation (22). If Assumption 2.1 holds, and Assumptions 3.6-3.5 (or Assumptions 3.3-3.2) hold,*

$$\check{v}_G^2(\epsilon_1, \ldots, \epsilon_G) \xrightarrow{P} \tau^2 ,$$

where τ^2 is defined in (C.8).

Proof. From Lemma C.6, we see that $\hat{\tau}_G^2 \xrightarrow{P} \tau^2$. It therefore suffices to show that $\check{\lambda}_G^2(\epsilon_1, \ldots, \epsilon_G) \xrightarrow{P} 0$. In order to do so, note that $\check{\lambda}_G^2(\epsilon_1, \ldots, \epsilon_G)$ may be decomposed into sums of the form

$$\frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}\hat{Y}_{\pi(4j-k)}\hat{Y}_{\pi(4j-\ell)}D_{\pi(4j-k')}D_{\pi(4j-\ell')} ,$$

where $(k, k') \in \{2, 3\}^2$ and $(l, l') \in \{0, 1\}^2$. Note that

$$\begin{aligned}
&\frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}\hat{Y}_{\pi(4j-k)}\hat{Y}_{\pi(4j-\ell)}D_{\pi(4j-k')}D_{\pi(4j-\ell')} \\
&= \frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}\tilde{Y}_{\pi(4j-k)}\tilde{Y}_{\pi(4j-\ell)}D_{\pi(4j-k')}D_{\pi(4j-\ell')} \\
&\quad + \frac{G}{n}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}\left(\hat{Y}_{\pi(4j-k)}\hat{Y}_{\pi(4j-\ell)} - \tilde{Y}_{\pi(4j-k)}\tilde{Y}_{\pi(4j-\ell)}\right)D_{\pi(4j-k')}D_{\pi(4j-\ell')} .
\end{aligned}$$

By following the arguments in Lemma S.1.9 of Bai et al. (2022) and Lemma C.17, we have that

$$\frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}\tilde{Y}_{\pi(4j-k)}\tilde{Y}_{\pi(4j-\ell)}D_{\pi(4j-k')}D_{\pi(4j-\ell')} \xrightarrow{P} 0 \ .$$

As for the second term, we show that it convergences to zero in probability in the case where $k = k' = 3$ and $\ell = \ell' = 1$. And the other cases should hold by repeating the same arguments.

$$\begin{aligned}
&\frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}\left(\hat{Y}_{\pi(4j-3)}\hat{Y}_{\pi(4j-1)} - \tilde{Y}_{\pi(4j-3)}\tilde{Y}_{\pi(4j-1)}\right)D_{\pi(4j-3)}D_{\pi(4j-1')}\\
&= \frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}\left(\hat{Y}_{\pi(4j-3)}(1)\hat{Y}_{\pi(4j-1)}(1) - \tilde{Y}_{\pi(4j-3)}(1)\tilde{Y}_{\pi(4j-1)}(1)\right)D_{\pi(4j-3)}D_{\pi(4j-1')}\\
&= \frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}\left(\hat{Y}_{\pi(4j-3)}(1) - \tilde{Y}_{\pi(4j-3)}(1)\right)\tilde{Y}_{\pi(4j-1)}(1)D_{\pi(4j-3)}D_{\pi(4j-1')}\\
&\quad + \frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}\left(\hat{Y}_{\pi(4j-3)}(1) - \tilde{Y}_{\pi(4j-3)}(1)\right)\left(\hat{Y}_{\pi(4j-1)}(1) - \tilde{Y}_{\pi(4j-1)}(1)\right)D_{\pi(4j-3)}D_{\pi(4j-1')}\\
&\quad + \frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}\left(\hat{Y}_{\pi(4j-1)}(1) - \tilde{Y}_{\pi(4j-1)}(1)\right)\tilde{Y}_{\pi(4j-3)}(1)D_{\pi(4j-3)}D_{\pi(4j-1')} \ ,
\end{aligned}$$

for which the first term is given as follows:

$$\begin{aligned}
&\frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}\left(\hat{Y}_{\pi(4j-3)}(1) - \tilde{Y}_{\pi(4j-3)}(1)\right)\tilde{Y}_{\pi(4j-1)}(1)D_{\pi(4j-3)}D_{\pi(4j-1')}\\
&= \left(\frac{1}{\frac{1}{2G}\sum_{1\le g\le 2G} N_g} - \frac{1}{E[N_g]}\right)\Bigg(\frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}N_{\pi(4j-3)}\bar{Y}_{\pi(4j-3)}(1)\\
&\qquad \times\tilde{Y}_{\pi(4j-1)}(1)D_{\pi(4j-3)}D_{\pi(4j-1')}\Bigg)\\
&\quad - \left(\frac{\frac{1}{2G}\sum_{1\le g\le 2G}\bar{Y}_g(1)I\{D_g = 1\}N_g}{\left(\frac{1}{2G}\sum_{1\le g\le 2G} N_g\right)^2} - \frac{E[\bar{Y}_g(1)N_g]}{E[N_g]^2}\right)\Bigg(\frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}N_{\pi(4j-3)}\\
&\qquad \times\tilde{Y}_{\pi(4j-1)}(1)D_{\pi(4j-3)}D_{\pi(4j-1')}\Bigg) \ .
\end{aligned}$$

by following the same argument in Lemma S.1.7 from Bai et al. (2022) and Lemma C.17, we have

$$\begin{aligned}
&\frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}N_{\pi(4j-3)}\bar{Y}_{\pi(4j-3)}(1)\tilde{Y}_{\pi(4j-1)}(1)D_{\pi(4j-3)}D_{\pi(4j-1')} \xrightarrow{P} 0\\
&\frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor} \epsilon_{2j-1}\epsilon_{2j}N_{\pi(4j-3)}\tilde{Y}_{\pi(4j-1)}(1)D_{\pi(4j-3)}D_{\pi(4j-1')} \xrightarrow{P} 0 \ .
\end{aligned}$$

Then, by the weak law of large numbers, Lemma C.15 and the continuous mapping theorem, we have

$$\frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor}\epsilon_{2j-1}\epsilon_{2j}\left(\hat{Y}_{\pi(4j-3)}(1)-\tilde{Y}_{\pi(4j-3)}(1)\right)\tilde{Y}_{\pi(4j-1)}(1)D_{\pi(4j-3)}D_{\pi(4j-1')}\xrightarrow{P}0 \ .$$

By repeating the same arguments for the other two terms, we conclude that

$$\frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor}\epsilon_{2j-1}\epsilon_{2j}\left(\hat{Y}_{\pi(4j-3)}\hat{Y}_{\pi(4j-1)}-\tilde{Y}_{\pi(4j-3)}\tilde{Y}_{\pi(4j-1)}\right)D_{\pi(4j-3)}D_{\pi(4j-1')}\xrightarrow{P}0 \ .$$

Therefore, for $(k,k')\in\{2,3\}^2$ and $(l,l')\in\{0,1\}^2$,

$$\frac{2}{G}\sum_{1\le j\le \lfloor \frac{G}{2}\rfloor}\epsilon_{2j-1}\epsilon_{2j}\hat{Y}_{\pi(4j-k)}\hat{Y}_{\pi(4j-\ell)}D_{\pi(4j-k')}D_{\pi(4j-\ell')}\xrightarrow{P}0 \ ,$$

which implies $\check{\lambda}^2_G(\epsilon_1,\ldots,\epsilon_G)\xrightarrow{P}0$, and thus $\check{\nu}^2_G(\epsilon_1,\ldots,\epsilon_G)\xrightarrow{P}\tau^2$. ■

Lemma C.10. *Suppose all assumptions in Theorem 3.8 hold. Then,*

$$\frac{1}{G}\sum_{1\le j\le G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})'\xrightarrow{P}2E[\psi_g\psi_g']-2E[E[\psi_g|W_g][\psi_g|W_g]']=2E[\mathrm{Var}[\psi_g|W_g]]$$
$$\frac{1}{G}\sum_{1\le j\le G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})(\tilde{\mu}_{1,j}-\tilde{\mu}_{0,j})\xrightarrow{P}E\left[\mathrm{Cov}\left[\tilde{Y}_g(1)+\tilde{Y}_g(0),\psi_g\middle|W_g\right]\right]E[N_g]$$

Proof. Note that

$$\begin{aligned}
&\frac{1}{G}\sum_{1\le j\le G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})'\\
&=\frac{1}{G}\sum_{1\le j\le G}\hat{\psi}_{1,j}\hat{\psi}_{1,j}'+\hat{\psi}_{0,j}\hat{\psi}_{0,j}'-\hat{\psi}_{1,j}\hat{\psi}_{0,j}'-\hat{\psi}_{0,j}\hat{\psi}_{1,j}'\\
&=\frac{1}{G}\sum_{1\le g\le 2G}\psi_g\psi_g'D_g+\frac{1}{G}\sum_{1\le g\le 2G}\psi_g\psi_g'(1-D_g)\\
&\quad-\frac{1}{G}\sum_{1\le j\le G}\psi_{\pi(2j)}\psi_{\pi(2j-1)}'D_{\pi(2j)}-\frac{1}{G}\sum_{1\le j\le G}\psi_{\pi(2j-1)}\psi_{\pi(2j)}'D_{\pi(2j-1)}\\
&\quad-\frac{1}{G}\sum_{1\le j\le G}\psi_{\pi(2j)}\psi_{\pi(2j-1)}'D_{\pi(2j-1)}-\frac{1}{G}\sum_{1\le j\le G}\psi_{\pi(2j-1)}\psi_{\pi(2j)}'D_{\pi(2j)}\\
&=\frac{1}{G}\sum_{1\le g\le 2G}\psi_g\psi_g'-\frac{1}{G}\sum_{1\le j\le G}(\psi_{\pi(2j)}\psi_{\pi(2j-1)}'+\psi_{\pi(2j-1)}\psi_{\pi(2j)}') \ .
\end{aligned}$$

Assumptions 2.1, 3.5, 3.6, 3.9, 3.10 and Lemma C.16 imply

$$\frac{1}{G}\sum_{1\le j\le G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})'\xrightarrow{P}2E[\psi_g\psi_g']-2E[E[\psi_g|W_g][\psi_g|W_g]']=2E[\mathrm{Var}[\psi_g|W_g]] \ .$$

On the other hand,

$$
\begin{aligned}
&\frac{1}{G}\sum_{1\le j\le G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})(\tilde{\mu}_{1,j}-\tilde{\mu}_{0,j})\\
&=\frac{1}{G}\sum_{1\le j\le G}\hat{\psi}_{1,j}\tilde{\mu}_{1,j}+\hat{\psi}_{0,j}\tilde{\mu}_{0,j}-\tilde{\mu}_{1,j}\hat{\psi}_{0,j}-\tilde{\mu}_{0,j}\hat{\psi}_{1,j}\\
&=\frac{1}{G}\sum_{1\le g\le 2G}\left(\bar{Y}_g(1)-\frac{E[\bar{Y}_g(1)N_g]}{E[N_g]}\right)N_g\psi_g D_g+\frac{1}{G}\sum_{1\le g\le 2G}\left(\bar{Y}_g(0)-\frac{E[\bar{Y}_g(0)N_g]}{E[N_g]}\right)N_g\psi_g(1-D_g)\\
&\quad-\frac{1}{G}\sum_{1\le j\le G}\left(\bar{Y}_{\pi(2j-1)}(1)-\frac{E[\bar{Y}_g(1)N_g]}{E[N_g]}\right)N_{\pi(2j-1)}\psi_{\pi(2j)}D_{\pi(2j-1)}\\
&\quad-\frac{1}{G}\sum_{1\le j\le G}\left(\bar{Y}_{\pi(2j)}(1)-\frac{E[\bar{Y}_g(1)N_g]}{E[N_g]}\right)N_{\pi(2j)}\psi_{\pi(2j-1)}D_{\pi(2j)}\\
&\quad-\frac{1}{G}\sum_{1\le j\le G}\left(\bar{Y}_{\pi(2j-1)}(0)-\frac{E[\bar{Y}_g(0)N_g]}{E[N_g]}\right)N_{\pi(2j-1)}\psi_{\pi(2j)}(1-D_{\pi(2j-1)})\\
&\quad-\frac{1}{G}\sum_{1\le j\le G}\left(\bar{Y}_{\pi(2j)}(0)-\frac{E[\bar{Y}_g(0)N_g]}{E[N_g]}\right)N_{\pi(2j)}\psi_{\pi(2j-1)}(1-D_{\pi(2j)})\ .
\end{aligned}
$$

Lemma C.16 implies that under Assumptions 2.1, 3.5, 3.6, 3.9, and 3.10, we have

$$
\begin{aligned}
&\frac{1}{G}\sum_{1\le g\le 2G}\left(\bar{Y}_g(1)-\frac{E[\bar{Y}_g(1)N_g]}{E[N_g]}\right)N_g\psi_g D_g\xrightarrow{P}E[\bar{Y}_g(1)N_g\psi_g]-\frac{E[\bar{Y}_g(1)N_g]}{E[N_g]}E[N_g\psi_g]\\
&\frac{1}{G}\sum_{1\le g\le 2G}\left(\bar{Y}_g(0)-\frac{E[\bar{Y}_g(0)N_g]}{E[N_g]}\right)N_g\psi_g(1-D_g)\xrightarrow{P}E[\bar{Y}_g(0)N_g\psi_g]-\frac{E[\bar{Y}_g(0)N_g]}{E[N_g]}E[N_g\psi_g]\\
&\frac{1}{G}\sum_{1\le j\le G}\left(\bar{Y}_{\pi(2j-1)}(1)-\frac{E[\bar{Y}_g(1)N_g]}{E[N_g]}\right)N_{\pi(2j-1)}\psi_{\pi(2j)}D_{\pi(2j-1)}\\
&\qquad\xrightarrow{P}\frac{1}{2}E\left[E\left[\bar{Y}_g(1)N_g|W_g\right]E\left[\psi_g|W_g\right]\right]-\frac{1}{2}\frac{E[\bar{Y}_g(1)N_g]}{E[N_g]}E\left[E\left[N_g|W_g\right]E\left[\psi_g|W_g\right]\right]\ .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\frac{1}{G}\sum_{1\le j\le G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})(\tilde{\mu}_{1,j}-\tilde{\mu}_{0,j})\\
&\xrightarrow{P}E[(\bar{Y}_g(1)+\bar{Y}_g(0))N_g\psi_g]-E\left[E\left[(\bar{Y}_g(1)+\bar{Y}_g(0))N_g|W_g\right]E\left[\psi_g|W_g\right]\right]\\
&\qquad-\frac{E[(\bar{Y}_g(1)+\bar{Y}_g(0))N_g]}{E[N_g]}E[N_g\psi_g]+\frac{E[(\bar{Y}_g(1)+\bar{Y}_g(0))N_g]}{E[N_g]}E\left[E\left[N_g|W_g\right]E\left[\psi_g|W_g\right]\right]\\
&=E\left[\mathrm{Cov}\left[(\bar{Y}_g(1)+\bar{Y}_g(0))N_g-\frac{E[(\bar{Y}_g(1)+\bar{Y}_g(0))N_g]}{E[N_g]}N_g,\psi_g\middle|W_g\right]\right]\\
&=E\left[\mathrm{Cov}\left[\tilde{Y}_g(1)+\tilde{Y}_g(0),\psi_g\middle|W_g\right]\right]E[N_g]\ ,
\end{aligned}
$$

as desired. ■

Lemma C.11. *Suppose all assumptions in Theorem 3.8 hold. Then,* $\tilde{\beta}_G-\hat{\beta}_G\xrightarrow{P}0$.

PROOF. Note that

$$\tilde{\beta}_G - \hat{\beta}_G = \left(\frac{1}{G}\sum_{1\leq j\leq G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})'\right)^{-1}\left(\frac{1}{G}\sum_{1\leq j\leq G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})(\tilde{\mu}_{1,j}-\tilde{\mu}_{0,j}-(\hat{\mu}_{1,j}-\hat{\mu}_{0,j}))\right) .$$

We want to show that the following term converges to zero:

$$\begin{aligned}
&\frac{1}{G}\sum_{1\leq j\leq G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})(\tilde{\mu}_{1,j}-\hat{\mu}_{1,j})\\
&=\frac{1}{G}\sum_{1\leq j\leq G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})\left(\frac{\frac{1}{G}\sum_{1\leq g\leq 2G}\bar{Y}_g D_g N_g}{\frac{1}{G}\sum_{1\leq g\leq 2G}D_g N_g}-\frac{E[\bar{Y}_g(1)N_g]}{E[N_g]}\right)N_{\pi(2j-1)}D_{\pi(2j-1)}\\
&\qquad+\frac{1}{G}\sum_{1\leq j\leq G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})\left(\frac{\frac{1}{G}\sum_{1\leq g\leq 2G}\bar{Y}_g D_g N_g}{\frac{1}{G}\sum_{1\leq g\leq 2G}D_g N_g}-\frac{E[\bar{Y}_g(1)N_g]}{E[N_g]}\right)N_{\pi(2j)}D_{\pi(2j)}\\
&=\left(\frac{\frac{1}{G}\sum_{1\leq g\leq 2G}\bar{Y}_g D_g N_g}{\frac{1}{G}\sum_{1\leq g\leq 2G}D_g N_g}-\frac{E[\bar{Y}_g(1)N_g]}{E[N_g]}\right)\frac{1}{G}\sum_{1\leq j\leq G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})(N_{\pi(2j-1)}D_{\pi(2j-1)}+N_{\pi(2j)}D_{\pi(2j)}) .
\end{aligned}$$

Lemma C.16 implies

$$\begin{aligned}
&\frac{1}{G}\sum_{1\leq j\leq G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})(N_{\pi(2j-1)}D_{\pi(2j-1)}+N_{\pi(2j)}D_{\pi(2j)})\\
&=\frac{1}{G}\sum_{1\leq g\leq 2G}\psi_g N_g D_g-\frac{1}{G}\sum_{1\leq j\leq G}\psi_{\pi(2j)}N_{\pi(2j-1)}D_{\pi(2j-1)}-\frac{1}{G}\sum_{1\leq j\leq G}\psi_{\pi(2j-1)}N_{\pi(2j)}D_{\pi(2j)}\\
&\xrightarrow{P} E[\psi_g N_g]-E[E[\psi_g|W_g]E[N_g|W_g]]\\
&=E\left[\mathrm{Cov}\left[\psi_g,N_g|W_g\right]\right] .
\end{aligned}$$

By Lemma C.15 and the continuous mapping theorem,

$$\left(\frac{\frac{1}{G}\sum_{1\leq g\leq 2G}\bar{Y}_g D_g N_g}{\frac{1}{G}\sum_{1\leq g\leq 2G}D_g N_g}-\frac{E[\bar{Y}_g(1)N_g]}{E[N_g]}\right)\xrightarrow{P}0 ,$$

which implies that

$$\frac{1}{G}\sum_{1\leq j\leq G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})(\tilde{\mu}_{1,j}-\hat{\mu}_{1,j})\xrightarrow{P}0 .$$

Similarly,

$$\frac{1}{G}\sum_{1\leq j\leq G}(\hat{\psi}_{1,j}-\hat{\psi}_{0,j})(\tilde{\mu}_{0,j}-\hat{\mu}_{0,j})\xrightarrow{P}0 .$$

The result then follows. ■

Lemma C.12. *If Assumption 2.1 holds, then*

$$\left|E[\bar{Y}_g^r(d)|X_g,N_g]\right|\leq C \quad a.s. ,$$

for $r \in \{1,2\}$ for some constant $C > 0$,

$$E\left[\bar{Y}_g^r(d) N_g^\ell\right] < \infty \ ,$$

for $r \in \{1,2\}, \ell \in \{0,1,2\}$, and

$$E\left[E[\bar{Y}_g(d) N_g | X_g]^2\right] < \infty \ .$$

Proof. We show the first statement for $r = 2$, since the case $r = 1$ follows similarly. By the Cauchy-Schwarz inequality,

$$\bar{Y}_g(d)^2 = \left(\frac{1}{|\mathcal{M}_g|}\sum_{i\in\mathcal{M}_g} Y_{i,g}(d)\right)^2 \le \frac{1}{|\mathcal{M}_g|}\sum_{i\in\mathcal{M}_g} Y_{i,g}(d)^2 \ ,$$

and hence

$$\left|E[\bar{Y}_g(d)^2 | X_g, N_g]\right| \le E\left[\frac{1}{|\mathcal{M}_g|}\sum_{i\in\mathcal{M}_g} E[Y_{i,g}(d)^2|X_g,N_g]\middle| X_g, N_g\right] \le C \ ,$$

where the first inequality follows from the above derivation, Assumption 2.1(e) and the law of iterated expectations, and final inequality follows from Assumption 2.1(d). We show the next statement for $r = \ell = 2$, since the other cases follow similarly. By the law of iterated expectations,

$$\begin{aligned} E\left[\bar{Y}_g^2(d) N_g^2\right] &= E\left[N_g^2 E[\bar{Y}_g^2(d)|X_g,N_g]\right] \\ &\lesssim E\left[N_g^2\right] < \infty \ , \end{aligned}$$

where the final line follows by Assumption 2.1(c). Finally,

$$\begin{aligned} E\left[E[\bar{Y}_g(d)N_g|X_g]^2\right] &= E\left[E[N_g E[\bar{Y}_g(d)|X_g,N_g]|X_g]^2\right] \\ &\lesssim E\left[E[N_g|X_g]^2\right] < \infty \ , \end{aligned}$$

where the final line follows from Jensen's inequality and Assumption 2.1(c). ■

Lemma C.13. *Suppose Assumption 3.5 holds. Then,*

$$\frac{1}{G}\sum_{g=1}^{G} N_{\pi(2g)}^\ell \left\|W_{\pi(2g)} - W_{\pi(2g-1)}\right\|^r \xrightarrow{P} 0 \ ,$$

for $\ell \in \{0,1,2\}$, $r \in \{1,2\}$.

Proof. By the Cauchy-Schwarz inequality

$$\frac{1}{G}\sum_{g=1}^{G} N_{\pi(2g)}^\ell |W_{\pi(2g)} - W_{\pi(2g-1)}|^r \le \left[\left(\frac{1}{G}\sum_{g=1}^{G} N_{\pi(2g)}^{2\ell}\right)\left(\frac{1}{G}\sum_{g=1}^{G}|W_{\pi(2g)} - W_{\pi(2g-1)}|^{2r}\right)\right]^{1/2} \ ,$$

$\frac{1}{G}\sum_{g=1}^{G} N_{\pi(2g)}^{2\ell} \le \frac{1}{G}\sum_{g=1}^{2G} N_g^{2\ell} = O_P(1)$ by the law of large numbers, $\frac{1}{G}\sum_g \|W_{\pi(2g)} - W_{\pi(2g-1)}\|^{2r} \xrightarrow{P} 0$ by assumption, hence the result follows. ■

Lemma C.14. *If Assumptions 2.1 and 3.5 hold,*

$$\frac{1}{G}\sum_{g=1}^{G}\left|N_{\pi(2g)}^2 - N_{\pi(2g-1)}^2\right| \xrightarrow{P} 0 \ .$$

Proof.

$$\begin{aligned}\frac{1}{G}\sum_{g=1}^{G}\left|N_{\pi(2g)}^2 - N_{\pi(2g-1)}^2\right| &= \frac{1}{G}\sum_{g=1}^{G}\left|N_{\pi(2g)} - N_{\pi(2g-1)}\right|\left|N_{\pi(2g)} + N_{\pi(2g-1)}\right| \\ &\le \left[\left(\frac{1}{G}\sum_{g=1}^{G}\left|N_{\pi(2g)} - N_{\pi(2g-1)}\right|^2\right)\left(\frac{1}{G}\sum_{g=1}^{G}\left|N_{\pi(2g)} + N_{\pi(2g-1)}\right|^2\right)\right]^{1/2} ,\end{aligned}$$

where the inequality follows by Cauchy-Schwarz. It follows from an argument similar to the proof of Lemma C.13 that $\frac{1}{G}\sum_{g=1}^{G}\left|N_{\pi(2g)} + N_{\pi(2g-1)}\right|^2 = O_P(1)$. By Assumption 3.5, $\frac{1}{G}\sum_{g=1}^{G}\left|N_{\pi(2g)} - N_{\pi(2g-1)}\right|^2 \xrightarrow{P} 0$. Hence the result follows. ■

Lemma C.15. *Let $Z_1, Z_2, \ldots, Z_G$ be i.i.d random variables. Then,*

(a) *Suppose $E[|Z_g|] < \infty$, $E[Z_g|X_g = x]$ is Lipschitz,*

$$Z^{(G)} \perp\!\!\!\perp D^{(G)}|X^{(G)} ,$$

and conditional on $X^{(G)}$, $(D_{\pi(2j-1)}, D_{\pi(2j)})$, $j = 1, ..., G$ are i.i.d. and each uniformly distributed over $\{(0,1),(1,0)\}$, and

$$\frac{1}{G}\sum_{1\le j\le G}\|X_{\pi(2j-1)} - X_{\pi(2j)}\| \xrightarrow{P} 0 \ .$$

Then, as $G \to \infty$,

$$\frac{1}{G}\sum_{1\le g\le 2G} Z_g D_g \xrightarrow{P} E[Z_g] \ .$$

(b) *Suppose $E[Z_g^2] < \infty$, $E[Z_g|W_g = w]$ is Lipschitz, $E[N_g^{2\ell}] < \infty$,*

$$Z^{(G)} \perp\!\!\!\perp D^{(G)}|W^{(G)} ,$$

and conditional on $W^{(G)}$, $(D_{\pi(2j-1)}, D_{\pi(2j)})$, $j = 1, ..., G$ are i.i.d. and each uniformly distributed over $\{(0,1),(1,0)\}$, and

$$\frac{1}{G}\sum_{1\le j\le G}\|W_{\pi(2j-1)} - W_{\pi(2j)}\|^2 \xrightarrow{P} 0 \ .$$

Then, as $G \to \infty$,

$$\frac{1}{G}\sum_{1\le g\le 2G} Z_g N_g^{\ell} D_g \xrightarrow{P} E[Z_g N_g^{\ell}] \ .$$

Proof. (a) follows from Lemma S.1.5 in Bai et al. (2022). (b) follows by combining the arguments in the proofs of that lemma and the proof of Lemma C.3. ■

Lemma C.16. *Let $(Z_1, \tilde{Z}_1), \ldots, (Z_G, \tilde{Z}_G)$ be i.i.d random vectors. Suppose Assumption 2.1 holds, $E[|Z_g|] < \infty$, $E[Z_g|X_g = x]$ and $E[\tilde{Z}_g|X_g = x]$ are Lipschitz,*

$$(Z^{(G)}, \tilde{Z}^{(G)}) \perp\!\!\!\perp D^{(G)}|X^{(G)} ,$$

and conditional on $X^{(G)}$, $(D_{\pi(2j-1)}, D_{\pi(2j)})$, $j = 1, ..., G$ are i.i.d. and each uniformly distributed over $\{(0,1),(1,0)\}$, and

$$\frac{1}{G}\sum_{1\le j\le G}\|X_{\pi(2j-1)} - X_{\pi(2j)}\|^2 \xrightarrow{P} 0 ,$$

Then,

$$\frac{1}{n}\sum_{1\le j\le G} Z_{\pi(2j-1)}\tilde{Z}_{\pi(2j)} \xrightarrow{P} E[E[Z_g|X_g]E[\tilde{Z}_g|X_g]]$$

$$\frac{1}{n}\sum_{1\le j\le G} Z_{\pi(2j-1)}\tilde{Z}_{\pi(2j)}D_{\pi(2j-1)} \xrightarrow{P} \frac{1}{2}E[E[Z_g|X_g]E[\tilde{Z}_g|X_g]] .$$

Proof. The proof is identical to the proof of Lemma S.1.6 in Bai et al. (2022) and is therefore omitted. ■

Lemma C.17. *If Assumptions 2.1 holds, and additionally Assumptions 3.2-3.3, 3.7 (or Assumptions 3.5-3.6, 3.8) hold, then*

1. $E\left[\tilde{Y}_g^2(d)\right] < \infty$ *for* $d \in \{0,1\}$.
2. $((\tilde{Y}_g(1), \tilde{Y}_g(0)) : 1 \le g \le 2G) \perp\!\!\!\perp D^{(G)}|X^{(G)}$ *or* $((\tilde{Y}_g(1), \tilde{Y}_g(0)) : 1 \le g \le 2G) \perp\!\!\!\perp D^{(G)}|W^{(G)}$.
3. *When not matching on cluster size,* $\frac{1}{G}\sum_{1\le j\le G}\left|\mu_d(X_{\pi(2j)}) - \mu_d(X_{\pi(2j-1)})\right| \xrightarrow{P} 0$, *where we use* $\mu_d(X_g)$ *to denote* $E[\tilde{Y}_g(d)|X_g]$ *for* $d \in \{0,1\}$ *or when matching on cluster size*

   $$\frac{1}{G}\sum_{1\le j\le G}\left|\mu_d(W_{\pi(2j)}) - \mu_d(W_{\pi(2j-1)})\right| \xrightarrow{P} 0 .$$

4. *When not matching on cluster size,*

   $$\frac{1}{G}\sum_{1\le j\le G}\left|\left(\mu_1(X_{\pi(2j)}) - \mu_1(X_{\pi(2j-1)})\right)\left(\mu_0(X_{\pi(2j)}) - \mu_0(X_{\pi(2j-1)})\right)\right| \xrightarrow{P} 0 ,$$

   or when matching on cluster size

   $$\frac{1}{G}\sum_{1\le j\le G}\left|\left(\mu_1(W_{\pi(2j)}) - \mu_1(W_{\pi(2j-1)})\right)\left(\mu_0(W_{\pi(2j)}) - \mu_0(W_{\pi(2j-1)})\right)\right| \xrightarrow{P} 0 .$$

5. *When not matching on cluster size*

   $$\frac{1}{4G}\sum_{k\in\{2,3\},\ell\in\{0,1\}}\sum_{1\le j\le \frac{G}{2}}\left(\mu_d\left(X_{\pi(4j-\ell)}\right) - \mu_d\left(X_{\pi(4j-k)}\right)\right)^2 \xrightarrow{P} 0 ,$$

or when matching on cluster size

$$\frac{1}{4G}\sum_{k\in\{2,3\},\ell\in\{0,1\}}\sum_{1\le j\le\frac{G}{2}}\left(\mu_d\left(W_{\pi(4j-\ell)}\right)-\mu_d\left(W_{\pi(4j-k)}\right)\right)^2\xrightarrow{P}0\ .$$

Proof. Note that

$$\begin{aligned}E\left[\tilde{Y}_g^2(d)\right]&\le E\left[N_g^2\left(\bar{Y}_g(d)-\frac{E\left[\bar{Y}_g(d)N_g\right]}{E\left[N_g\right]}\right)^2\right]\\&\lesssim E\left[N_g^2\bar{Y}_g^2(d)\right]+\left(\frac{E\left[\bar{Y}_g(d)N_g\right]}{E\left[N_g\right]}\right)^2E[N_g^2]<\infty\end{aligned}$$

where the inequality follows by Lemma C.12. The second result follows directly by inspection and Assumption 3.1 (or Assumption 3.4). In terms of the third result, by Assumption 3.2 and 3.3,

$$\frac{1}{G}\sum_{1\le j\le G}\left|\mu_1(X_{\pi(2j)})-\mu_1(X_{\pi(2j-1)})\right|\lesssim\frac{1}{G}\sum_{1\le j\le G}\left\|X_{\pi(2j)}-X_{\pi(2j-1)}\right\|\xrightarrow{P}0\ .$$

Meanwhile,

$$\begin{aligned}&\frac{1}{G}\sum_{1\le j\le G}\left|\mu_1(W_{\pi(2j)})-\mu_1(W_{\pi(2j-1)})\right|\\&\lesssim\frac{1}{G}\sum_{1\le j\le G}\left|E[N_{\pi(2j)}\bar{Y}_{\pi(2j)}(d)|W_{\pi(2j)}]-E[N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)}(d)|W_{\pi(2j-1)}]\right|\\&\qquad+\frac{1}{G}\sum_{1\le j\le G}\left|E[N_{\pi(2j)}|W_{\pi(2j)}]-E[N_{\pi(2j-1)}|W_{\pi(2j-1)}]\right|\\&\lesssim\frac{1}{G}\sum_{1\le j\le G}\left|N_{\pi(2j)}\left(E[\bar{Y}_{\pi(2j)}(d)|W_{\pi(2j)}]-E[\bar{Y}_{\pi(2j-1)}(d)|W_{\pi(2j-1)}]\right)\right|+\frac{1}{G}\sum_{1\le j\le G}\left|N_{\pi(2j)}-N_{\pi(2j-1)}\right|\\&\qquad+\frac{1}{G}\sum_{1\le j\le G}\left|(N_{\pi(2j)}-N_{\pi(2j-1)})E[\bar{Y}_{\pi(2j-1)}(d)|W_{\pi(2j-1)}]\right|\\&\lesssim\frac{1}{G}\sum_{1\le j\le G}N_{\pi(2j)}\left\|W_{\pi(2j)}-W_{\pi(2j-1)}\right\|\ ,\end{aligned}$$

which converges to zero in probability by Assumption 3.5 and Lemma C.13. To prove the fourth result, by Assumption 3.2 and 3.3,

$$\frac{1}{G}\sum_{1\le j\le G}\left|\left(\mu_1(X_{\pi(2j)})-\mu_1(X_{\pi(2j-1)})\right)\left(\mu_0(X_{\pi(2j)})-\mu_0(X_{\pi(2j-1)})\right)\right|\lesssim\frac{1}{G}\sum_{1\le j\le G}\left\|X_{\pi(2j)}-X_{\pi(2j-1)}\right\|^2\xrightarrow{P}0\,.$$

Similarly,

$$\begin{aligned}&\frac{1}{G}\sum_{1\le j\le G}\left|\left(\mu_1(W_{\pi(2j)})-\mu_1(W_{\pi(2j-1)})\right)\left(\mu_0(W_{\pi(2j)})-\mu_0(W_{\pi(2j-1)})\right)\right|\\&\le\frac{1}{G}\sum_{1\le j\le G}\left|\mu_1(W_{\pi(2j)})-\mu_1(W_{\pi(2j-1)})\right|\left|\mu_0(W_{\pi(2j)})-\mu_0(W_{\pi(2j-1)})\right|\end{aligned}$$

$$\lesssim \frac{1}{G} \sum_{1 \le j \le G} N_{\pi(2j)}^2 \left\| W_{\pi(2j)} - W_{\pi(2j-1)} \right\|^2 \xrightarrow{P} 0 \ ,$$

where the last step follows by Assumption 3.5 and Lemma C.13. Finally, the fifth result follows the same argument by Assumption 3.7 (or Assumption 3.8). ■

Lemma C.18.

$$\rho \left(\mathcal{L} \left((\mathbb{K}_G^{YN}, \mathbb{K}_G^N)' | Z^{(G)} \right), N(0, \mathbb{V}_R) \right) \xrightarrow{P} 0 \ ,$$

where

$$\begin{pmatrix} \mathbb{K}_G^{YN} \\ \mathbb{K}_G^N \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{G}} \sum_{1 \le j \le G} \epsilon_j \left(N_{\pi(2j)} \bar{Y}_{\pi(2j)} - N_{\pi(2j-1)} \bar{Y}_{\pi(2j-1)} \right) (D_{\pi(2j)} - D_{\pi(2j-1)}) \\ \frac{1}{\sqrt{G}} \sum_{1 \le j \le G} \epsilon_j (N_{\pi(2j)} - N_{\pi(2j-1)})(D_{\pi(2j)} - D_{\pi(2j-1)}) \end{pmatrix} \ ,$$

and where, in the case where we match on cluster size,

$$\mathbb{V}_R = \begin{pmatrix} \mathbb{V}_R^1 & 0 \\ 0 & 0 \end{pmatrix} \ ,$$

with

$$\mathbb{V}_R^1 = E[\text{Var}(N_g \bar{Y}_g(1)|W_g)] + E[\text{Var}(N_g \bar{Y}_g(0)|W_g)] + E\left[(E[N_g \bar{Y}_g(1)|W_g] - E[N_g \bar{Y}_g(0)|W_g])^2 \right] \ ,$$

and when we do not match on cluster size,

$$\mathbb{V}_R = \begin{pmatrix} \mathbb{V}_R^{1,1} & \mathbb{V}_R^{1,2} \\ \mathbb{V}_R^{1,2} & \mathbb{V}_R^{2,2} \end{pmatrix} \ ,$$

with

$$\begin{aligned} \mathbb{V}_R^{1,1} &= E[\text{Var}(N_g \bar{Y}_g(1)|X_g)] + E[\text{Var}(N_g \bar{Y}_g(0)|X_g)] + E\left[(E[N_g \bar{Y}_g(1)|X_g] - E[N_g \bar{Y}_g(0)|X_g])^2 \right] \\ \mathbb{V}_R^{1,2} &= E[N_g^2 \bar{Y}_g(1)] + E[N_g^2 \bar{Y}_g(0)] - \left(E\left[E[N_g \bar{Y}_g(1)|X_g] E[N_g|X_g] \right] + E\left[E[N_g \bar{Y}_g(0)|X_g] E[N_g|X_g] \right] \right) \\ \mathbb{V}_R^{2,2} &= 2E[\text{Var}(N_g|X_g)] \ . \end{aligned}$$

Proof. Using the fact that ϵ_j, $j = 1, \ldots, G$ and $\epsilon_j(D_{\pi(2j)} - D_{\pi(2j-1)})$, $j = 1, \ldots, G$ have the same distribution conditional on $Z^{(G)}$, it suffices to study the limiting distribution of $(\tilde{\mathbb{K}}_G^{YN}, \tilde{\mathbb{K}}_G^N)'$ conditional on $Z^{(G)}$, where

$$\begin{aligned} \tilde{\mathbb{K}}_G^{YN} &:= \frac{1}{\sqrt{G}} \sum_{1 \le j \le G} \epsilon_j \left(N_{\pi(2j)} \bar{Y}_{\pi(2j)} - N_{\pi(2j-1)} \bar{Y}_{\pi(2j-1)} \right) \ , \\ \tilde{\mathbb{K}}_G^N &:= \frac{1}{\sqrt{G}} \sum_{1 \le j \le G} \epsilon_j \left(N_{\pi(2j)} - N_{\pi(2j-1)} \right) \ . \end{aligned}$$

We will show

$$\rho \left(\mathcal{L} \left((\tilde{\mathbb{K}}_G^{YN}, \tilde{\mathbb{K}}_G^N)' | Z^{(G)} \right), N(0, \mathbb{V}_R) \right) \xrightarrow{P} 0 \ , \tag{31}$$

where $\mathcal{L}(\cdot)$ denote the law and ρ is any metric that metrizes weak convergence. To that end, we will employ the Lindeberg central limit theorem in Proposition 2.27 of van der Vaart (1998) and a subsequencing argument. Indeed, to verify (31), note we need only show that for any subsequence $\{G_k\}$ there exists a further subsequence $\{G_{k_l}\}$ such that

$$\rho\left(\mathcal{L}\left((\tilde{\mathbb{K}}^{YN}_{G_{k_l}}, \tilde{\mathbb{K}}^{N}_{G_{k_l}})'|Z^{(G_{k_l})}\right), N(0, \mathbb{V}_R)\right) \to 0 \text{ with probability one }. \tag{32}$$

To that end, define

$$\mathbb{V}_{R,n} = \begin{pmatrix} \mathbb{V}^{1,1}_{R,n} & \mathbb{V}^{1,2}_{R,n} \\ \mathbb{V}^{1,2}_{R,n} & \mathbb{V}^{2,2}_{R,n} \end{pmatrix} = \text{Var}[(\tilde{\mathbb{K}}^{YN}_G, \tilde{\mathbb{K}}^{N}_G)'|Z^{(G)}] ,$$

where

$$\mathbb{V}^{1,1}_{R,n} = \frac{1}{G}\sum_{1\le j\le G}(N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)})^2$$

$$\mathbb{V}^{1,2}_{R,n} = \frac{1}{G}\sum_{1\le j\le G}(N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)})(N_{\pi(2j)} - N_{\pi(2j-1)})$$

$$\mathbb{V}^{2,2}_{R,n} = \frac{1}{G}\sum_{1\le j\le G}(N_{\pi(2j)} - N_{\pi(2j-1)})^2 .$$

We first show that

$$\mathbb{V}_{R,n} \xrightarrow{P} \mathbb{V}_R . \tag{33}$$

Consider the case where we match on cluster size. The weak law of large numbers and Lemma C.16 imply

$$\mathbb{V}^{1,1}_{R,n} \xrightarrow{P} E[\text{Var}[N_g\bar{Y}_g(1)]|W_g] + E[\text{Var}[N_g\bar{Y}_g(0)]|W_g] + E\left[(E[N_g\bar{Y}_g(1)|W_g] - E[N_g\bar{Y}_g(0)|W_g])^2\right] .$$

Next, we show that in this case $\mathbb{V}^{1,2}_{R,n}$ and $\mathbb{V}^{2,2}_{R,n}$ are $o_P(1)$. For $\mathbb{V}^{2,2}_{R,n}$ this follows immediately from Assumption 3.5. For $\mathbb{V}^{1,2}_{R,n}$ note that by the Cauchy-Schwarz inequality,

$$\begin{aligned}&\frac{1}{G}\sum_{1\le j\le G}\left(\left(N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)}\right)\left(N_{\pi(2j)} - N_{\pi(2j-1)}\right)\right)\\ &\le \left(\left(\frac{1}{G}\sum_{1\le j\le G}\left(N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)}\right)^2\right)\left(\frac{1}{G}\sum_{1\le j\le G}\left(N_{\pi(2j)} - N_{\pi(2j-1)}\right)^2\right)\right)^{1/2} .\end{aligned}$$

The second term of the product on the RHS is $o_P(1)$ by Assumption 3.5. The first term is $O_P(1)$ since

$$\frac{1}{G}\sum_{1\le j\le G}\left(N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)}\right)^2 \lesssim \frac{1}{G}\sum_{1\le g\le 2G}N_g^2\bar{Y}_g(1)^2 + \frac{1}{G}\sum_{1\le g\le 2G}N_g^2\bar{Y}_g(0)^2 = O_P(1) ,$$

where the first inequality follows from exploiting the fact that $|a-b|^2 \le 2(a^2+b^2)$ and the definition of $\bar{Y}_g$, and the final equality follows from Lemma C.12 and the law of large numbers. We can thus conclude that $\mathbb{V}^{1,2}_{R,n} = o_P(1)$ when matching on cluster size.

In the case where we do *not* match on cluster size, again by the weak law of large numbers and Lemma C.16, it can be shown that (33) holds. Next, we verify the Lindeberg condition in Proposition 2.27 of van der

Vaart (1998). Note that for an arbitrary $\delta > 0$,

$$
\begin{aligned}
&\frac{1}{G}\sum_{1\le j\le G} E[((\epsilon_j(N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)}))^2 + (\epsilon_j(N_{\pi(2j)} - N_{\pi(2j-1)}))^2) \\
&\qquad \times I\{((\epsilon_j(N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)}))^2 + (\epsilon_j(N_{\pi(2j)} - N_{\pi(2j-1)}))^2) > \delta^2 G\}|Z^{(G)}] \\
&= \frac{1}{G}\sum_{1\le j\le G} E[((N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)})^2 + (N_{\pi(2j)} - N_{\pi(2j-1)})^2) \\
&\qquad \times I\{((N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)})^2 + (N_{\pi(2j)} - N_{\pi(2j-1)})^2) > \delta^2 G\}|Z^{(G)}] \\
&\lesssim \frac{1}{G}\sum_{1\le j\le G} (N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)})^2 I\{(N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)})^2 > \delta^2 G/2\} \\
&\qquad + \frac{1}{G}\sum_{1\le j\le G} (N_{\pi(2j)} - N_{\pi(2j-1)})^2 I\{(N_{\pi(2j)} - N_{\pi(2j-1)})^2 > \delta^2 G/2\}\ .
\end{aligned}
$$

where the inequality follows from (28) and the fact that $(N_g, \bar{Y}_g), 1 \le g \le 2G$ are all constants conditional on $Z^{(G)}$. The last line converges in probability to zero as long as we can show

$$
\begin{aligned}
&\frac{1}{G}\max_{1\le j\le G}(N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)})^2 \xrightarrow{P} 0 \\
&\frac{1}{G}\max_{1\le j\le G}(N_{\pi(2j)} - N_{\pi(2j-1)})^2 \xrightarrow{P} 0\ .
\end{aligned}
$$

Note

$$
\begin{aligned}
\frac{1}{G}\max_{1\le j\le G}(N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)})^2 &\lesssim \frac{1}{G}\max_{1\le j\le G}\left(N^2_{\pi_{(2j-1)}}\bar{Y}^2_{\pi(2j-1)} + N^2_{\pi_{(2j)}}\bar{Y}^2_{\pi(2j)}\right) \\
&\lesssim \frac{1}{G}\max_{1\le g\le 2G}\left(N_g^2\bar{Y}_g^2(1) + N_g^2\bar{Y}_g^2(0)\right) \xrightarrow{P} 0
\end{aligned}
$$

Where the first inequality follows from the fact that $|a-b|^2 \le 2(a^2+b^2)$, the second by inspection, and the convergence by Lemma S.1.1 in Bai et al. (2022) along with Assumption 2.1(c) and Lemma C.12. The second statement follows similarly. Therefore, we have verified both conditions in Proposition 2.27 of van der Vaart (1998) hold in probability, and therefore for each subsequence there must exists a further subsequence along which both conditions hold with probability one, so (32) holds, and the conclusion of the lemma follows. ■

D Addtional Simulations

D.1 Simulation Results in Finite Populations

In this section, we compare the finite population design-based coverage properties of confidence intervals constructed using our proposed variance estimator $\hat{v}_G^2$ versus the estimators $\hat{\omega}^2_{\mathrm{CR,G}}$ and $\hat{\omega}^2_{\mathrm{PCVE,G}}$ introduced in Section 3.2. We revisit the simulation setting considered in Tables 1–4 in Section 4.1, but now use each DGP to generate the covariates and outcomes only *once*, and then fix these in repeated samples.

Tables 9–12 present our results. From Tables 9 and 10, we see that both $\hat{\nu}_G^2$ and $\hat{\omega}^2_{\mathrm{PCVE,G}}$ are consistent in large populations when there is sufficient "homogeneity" in treatment effects, but undercover in small populations. This behavior is not surprising given that asymptotically exact inference is often feasible even in the design-based paradigm as long as treatment effects are sufficiently homogeneous; see for instance Bai et al. (2024d) for a discussion in the context of completely randomized experiments. On the other hand, Tables 11 and 12 illustrate that when there is treatment effect heterogeneity, all three estimators are conservative, leading to a coverage probability of 1 for all population sizes. However, although all three estimators over-cover, our proposed variance estimator $\hat{v}_G^2$ produces confidence intervals with the shortest average length in all cases.

D.2 Simulation Results for Different Choices of $|\mathcal{M}_g|$

In this section, we repeat the simulation exercise from Section 4.1 for different choices of the second stage sample size $|\mathcal{M}_g| = \lfloor \rho \cdot N_g \rfloor$ for $\rho \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$. In each case, we generate samples as in Section 4.1, but sample a fraction $\rho \cdot N_g$ of each cluster without replacement when computing $\hat{\Delta}_G$ and $\hat{v}_G^2$. Results for $G = 50$ and $G = 250$ are presented in Tables 13–16. In each table, the results stay roughly the same across different values of ρ, with the average lengths of the confidence intervals slightly decreasing when ρ increases. The stability across ρ is not surprising in our model given the heavy dependence across the units within the same cluster.

Table 9: Model 1 - Finite Population - Matching on X_g[1]

N_{max}/N_{min}	VCE	$G=12$	$G=26$	$G=50$	$G=100$	$G=150$	$G=200$	$G=250$
				Coverage				
	$\hat{v}^2$	0.8990	0.9295	0.9460	0.9380	0.9470	0.9340	0.9505
1.11	CR	1	1	0.9990	1	1	1	0.9995
	PCVE	0.9095	0.9270	0.9450	0.9365	0.9470	0.9325	0.9480
	$\hat{v}^2$	0.9060	0.9315	0.9475	0.9375	0.9515	0.9330	0.9465
1.42	CR	1	1	0.9990	1	1	1	0.9990
	PCVE	0.9085	0.9305	0.9450	0.9370	0.9530	0.9320	0.9480
	$\hat{v}^2$	0.9030	0.9260	0.9450	0.9370	0.9480	0.9375	0.9495
1.99	CR	1	1	1	1	1	1	0.9980
	PCVE	0.9170	0.9250	0.9450	0.9360	0.9485	0.9330	0.9480
	$\hat{v}^2$	0.8775	0.9190	0.9395	0.9430	0.9425	0.9385	0.9485
3.31	CR	1	1	1	1	1	0.9995	0.9965
	PCVE	0.9075	0.9175	0.9435	0.9395	0.9435	0.9360	0.9470
	$\hat{v}^2$	0.8880	0.9085	0.9440	0.9390	0.9415	0.9455	0.9405
9.80	CR	1	1	1	0.9995	1	0.9965	0.9925
	PCVE	0.9075	0.9100	0.9465	0.9400	0.9420	0.9455	0.9410
				Average Length				
	$\hat{v}^2$	1.12824	1.05815	0.84888	0.59101	0.44808	0.41502	0.38434
1.11	CR	2.93266	2.25955	1.56492	1.20447	0.90146	0.79447	0.72000
	PCVE	1.11395	1.04746	0.84517	0.58917	0.44726	0.41469	0.38418
	$\hat{v}^2$	1.07152	1.06921	0.84835	0.60402	0.45275	0.42419	0.40010
1.42	CR	2.98019	2.30454	1.56619	1.21866	0.90291	0.79774	0.72714
	PCVE	1.06215	1.05823	0.84533	0.60213	0.45189	0.42370	0.39987
	$\hat{v}^2$	1.05214	1.08426	0.82321	0.62589	0.46226	0.44162	0.42537
1.99	CR	3.02136	2.38754	1.56696	1.24393	0.90557	0.80431	0.73828
	PCVE	1.04815	1.07367	0.82097	0.62399	0.46142	0.44114	0.42500
	$\hat{v}^2$	1.04528	1.11925	0.82767	0.64119	0.47469	0.47427	0.46200
3.31	CR	3.09726	2.42478	1.56226	1.29434	0.91920	0.82070	0.75534
	PCVE	1.04739	1.11017	0.82627	0.63952	0.47380	0.47367	0.46149
	$\hat{v}^2$	1.19775	1.19395	0.82358	0.70239	0.51101	0.53635	0.53192
9.80	CR	3.19729	2.59330	1.55023	1.39250	0.94697	0.85953	0.79422
	PCVE	1.20833	1.18286	0.82301	0.70132	0.51043	0.53549	0.53114

[1] Number of clusters = $2G$ with $G = 12, 26, 50, 100, 150, 200, 250$. Number of replications for each G is 2000. $N_{max} = 500$.

Table 10: Model 1 - Finite Population - Matching on X_g and N_g[1]

N_{max}/N_{min}	VCE	$G = 12$	$G = 26$	$G = 50$	$G = 100$	$G = 150$	$G = 200$	$G = 250$
				Coverage				
	$\hat{v}^2$	0.9225	0.8930	0.9365	0.9475	0.9500	0.9560	0.9505
1.11	CR	1	1	1	1	1	1	1
	PCVE	0.9055	0.9405	0.9360	0.9470	0.9465	0.9590	0.9510
	$\hat{v}^2$	0.9245	0.9220	0.9410	0.9480	0.9455	0.9530	0.9475
1.42	CR	1	1	1	1	1	1	1
	PCVE	0.9115	0.9230	0.9390	0.9460	0.9540	0.9545	0.9485
	$\hat{v}^2$	0.9370	0.8555	0.9490	0.9455	0.9515	0.9480	0.9540
1.99	CR	1	1	1	1	1	1	1
	PCVE	0.9225	0.9290	0.9505	0.9465	0.9490	0.9495	0.9555
	$\hat{v}^2$	0.9070	0.8475	0.9515	0.9610	0.9625	0.9665	0.9545
3.31	CR	1	1	1	1	1	1	1
	PCVE	0.9035	0.9425	0.9515	0.9595	0.9610	0.9615	0.9550
	$\hat{v}^2$	0.9020	0.8175	0.9415	0.9580	0.9665	0.9635	0.9645
9.80	CR	1	1	1	1	1	1	1
	PCVE	0.8980	0.9155	0.9475	0.9580	0.9635	0.9655	0.9640
				Average Length				
	$\hat{v}^2$	1.06353	0.54293	0.39449	0.26347	0.20698	0.14455	0.13742
1.11	CR	2.93374	2.26348	1.56660	1.20512	0.90170	0.79480	0.72016
	PCVE	1.04531	0.54223	0.39200	0.26298	0.20627	0.14448	0.13730
	$\hat{v}^2$	1.03963	0.80633	0.29493	0.19849	0.16039	0.12190	0.09622
1.42	CR	2.98061	2.30678	1.56787	1.21943	0.90334	0.79804	0.72736
	PCVE	1.01824	0.80046	0.29340	0.19762	0.16023	0.12191	0.09628
	$\hat{v}^2$	1.09840	0.63621	0.25458	0.16747	0.14000	0.12914	0.09993
1.99	CR	3.01789	2.38973	1.56826	1.24480	0.90602	0.80477	0.73865
	PCVE	1.08265	0.63690	0.25379	0.16716	0.13959	0.12888	0.09985
	$\hat{v}^2$	1.02165	0.71836	0.26920	0.21766	0.17826	0.13358	0.09376
3.31	CR	3.09474	2.42593	1.56316	1.29498	0.91953	0.82124	0.75591
	PCVE	1.00793	0.71943	0.26743	0.21631	0.17711	0.13257	0.09323
	$\hat{v}^2$	1.13033	0.88192	0.28810	0.26255	0.18366	0.12748	0.10254
9.80	CR	3.19046	2.59270	1.55106	1.39307	0.94746	0.86046	0.79523
	PCVE	1.11007	0.87854	0.28778	0.26048	0.18279	0.12726	0.10232

[1] Number of clusters $= 2G$ with $G = 12, 26, 50, 100, 150, 200, 250$. Number of replications for each G is 2000. $N_{max} = 500$.

Table 11: Model 2 - Finite Population - Matching on X_g[1]

N_{max}/N_{min}	VCE	$G=12$	$G=26$	$G=50$	$G=100$	$G=150$	$G=200$	$G=250$
				Coverage				
	$\hat{v}^2$	1	1	0.9995	1	1	1	0.9990
1.11	CR	1	1	1	1	1	1	1
	PCVE	1	1	1	1	1	1	1
	$\hat{v}^2$	1	1	0.9990	1	1	1	0.9990
1.42	CR	1	1	1	1	1	1	1
	PCVE	1	1	1	1	1	1	1
	$\hat{v}^2$	1	1	0.9995	1	1	1	0.9985
1.99	CR	1	1	1	1	1	1	0.9995
	PCVE	1	1	1	1	1	1	0.9995
	$\hat{v}^2$	1	1	0.9990	1	0.9990	0.9985	0.9970
3.31	CR	1	1	1	1	1	1	0.9995
	PCVE	1	1	1	1	1	1	0.9995
	$\hat{v}^2$	1	1	1	0.9995	0.9990	0.9965	0.9960
9.80	CR	1	1	1	1	1	0.9995	0.9985
	PCVE	1	1	1	1	1	0.9995	0.9985
				Average Length				
	$\hat{v}^2$	1.51070	1.10752	0.81935	0.63852	0.44747	0.39393	0.35735
1.11	CR	1.66339	1.31058	0.92939	0.76490	0.52908	0.47240	0.42471
	PCVE	1.67962	1.31421	0.93901	0.76591	0.53029	0.47223	0.42367
	$\hat{v}^2$	1.53829	1.15173	0.81013	0.64981	0.45659	0.39764	0.36359
1.42	CR	1.72251	1.36383	0.92401	0.77462	0.53466	0.47511	0.43120
	PCVE	1.73073	1.37272	0.92403	0.77627	0.53808	0.47500	0.42954
	$\hat{v}^2$	1.45130	1.16632	0.79474	0.67449	0.45573	0.40764	0.37492
1.99	CR	1.69166	1.40618	0.92103	0.79970	0.53349	0.48243	0.44014
	PCVE	1.64456	1.39143	0.90836	0.80309	0.53384	0.48525	0.43991
	$\hat{v}^2$	1.51039	1.23004	0.82204	0.71496	0.47173	0.42133	0.38757
3.31	CR	1.73747	1.46680	0.92359	0.84163	0.54618	0.49257	0.45049
	PCVE	1.72595	1.47169	0.92085	0.84367	0.54881	0.49426	0.45014
	$\hat{v}^2$	1.71776	1.31631	0.80818	0.79366	0.48406	0.44659	0.41584
9.80	CR	1.86668	1.60387	0.90901	0.92440	0.55159	0.51513	0.47517
	PCVE	1.93059	1.59637	0.89610	0.92753	0.54784	0.51592	0.47486

[1] Number of clusters $= 2G$ with $G = 12, 26, 50, 100, 150, 200, 250$. Number of replications for each G is 2000. $N_{max} = 500$.

Table 12: Model 2 - Finite Population - Matching on X_g and N_g[1]

N_{max}/N_{min}	VCE	$G=12$	$G=26$	$G=50$	$G=100$	$G=150$	$G=200$	$G=250$
				Coverage				
	$\hat{\upsilon}^2$	1	1	1	1	1	1	1
1.11	CR	1	1	1	1	1	1	1
	PCVE	1	1	1	1	1	1	1
	$\hat{\upsilon}^2$	1	1	1	1	1	1	1
1.42	CR	1	1	1	1	1	1	1
	PCVE	1	1	1	1	1	1	1
	$\hat{\upsilon}^2$	1	1	1	1	1	1	1
1.99	CR	1	1	1	1	1	1	1
	PCVE	1	1	1	1	1	1	1
	$\hat{\upsilon}^2$	1	1	1	1	1	1	1
3.31	CR	1	1	1	1	1	1	1
	PCVE	1	1	1	1	1	1	1
	$\hat{\upsilon}^2$	1	1	1	1	1	1	1
9.80	CR	1	1	1	1	1	1	1
	PCVE	1	1	1	1	1	1	1
				Average Length				
	$\hat{\upsilon}^2$	1.43001	0.98768	0.71225	0.57552	0.38898	0.34712	0.30917
1.11	CR	1.66632	1.31199	0.93037	0.76575	0.52947	0.47207	0.42466
	PCVE	1.66130	1.30434	0.94045	0.76739	0.52682	0.47164	0.42340
	$\hat{\upsilon}^2$	1.35210	1.08903	0.68790	0.58216	0.39551	0.34579	0.31063
1.42	CR	1.71907	1.36641	0.92554	0.77532	0.53521	0.47431	0.43152
	PCVE	1.71252	1.36754	0.92406	0.77703	0.53714	0.47354	0.43086
	$\hat{\upsilon}^2$	1.36855	1.04579	0.68163	0.60169	0.38793	0.35447	0.31701
1.99	CR	1.68436	1.40552	0.92186	0.80133	0.53444	0.48159	0.44058
	PCVE	1.64990	1.37699	0.91400	0.80336	0.53179	0.48544	0.43940
	$\hat{\upsilon}^2$	1.43146	1.11080	0.69613	0.64046	0.40438	0.36571	0.32568
3.31	CR	1.73042	1.46136	0.92487	0.84401	0.54673	0.49209	0.45137
	PCVE	1.71754	1.45545	0.92046	0.84452	0.55122	0.49459	0.44999
	$\hat{\upsilon}^2$	1.62023	1.24723	0.68231	0.71972	0.41039	0.37921	0.34731
9.80	CR	1.85014	1.59673	0.91020	0.92797	0.55260	0.51529	0.47639
	PCVE	1.92935	1.60166	0.90340	0.93148	0.54945	0.51515	0.47589

[1] Number of clusters $= 2G$ with $G = 12, 26, 50, 100, 150, 200, 250$. Number of replications for each G is 2000. $N_{max} = 500$.

Table 13: Model 1 - $|\mathcal{M}_g| = \rho \cdot N_g$ with $G = 50$ - Matching on X_g and N_g [1]

N_{max}/N_{min}	VCE	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
		Coverage				
	$\hat{v}^2$	0.9435	0.9315	0.9335	0.9335	0.9405
1.11	CR	1	1	1	1	1
	PCVE	0.9440	0.9335	0.9330	0.9360	0.9420
	$\hat{v}^2$	0.9455	0.9320	0.9455	0.9385	0.9405
1.42	CR	1	1	1	1	1
	PCVE	0.9485	0.9325	0.9465	0.9365	0.9405
	$\hat{v}^2$	0.9345	0.9350	0.9450	0.9400	0.9380
1.99	CR	1	1	1	1	1
	PCVE	0.9380	0.9400	0.9460	0.9430	0.9410
	$\hat{v}^2$	0.9395	0.9370	0.9345	0.9420	0.9395
3.31	CR	1	1	1	1	1
	PCVE	0.9400	0.9405	0.9380	0.9480	0.9380
	$\hat{v}^2$	0.9425	0.9410	0.9495	0.9385	0.9270
9.80	CR	1	1	1	1	1
	PCVE	0.9435	0.9445	0.9505	0.9370	0.9325
		Average Length				
	$\hat{v}^2$	0.40207	0.39959	0.39833	0.39692	0.39552
1.11	CR	1.62141	1.62149	1.62140	1.62101	1.62086
	PCVE	0.39996	0.39815	0.39623	0.39516	0.39415
	$\hat{v}^2$	0.35158	0.34891	0.34777	0.34562	0.34375
1.42	CR	1.63392	1.63325	1.63252	1.63225	1.63232
	PCVE	0.35029	0.34731	0.34535	0.34384	0.34229
	$\hat{v}^2$	0.35889	0.35386	0.35342	0.35057	0.34797
1.99	CR	1.65320	1.65233	1.65205	1.65185	1.65086
	PCVE	0.35634	0.35293	0.35167	0.34858	0.34715
	$\hat{v}^2$	0.38701	0.38306	0.37956	0.37682	0.37493
3.31	CR	1.68841	1.68715	1.68610	1.68575	1.68542
	PCVE	0.38437	0.37985	0.37746	0.37493	0.37263
	$\hat{v}^2$	0.44908	0.44416	0.44082	0.43789	0.43528
9.80	CR	1.75885	1.75848	1.75757	1.75719	1.75705
	PCVE	0.44459	0.43984	0.43769	0.43398	0.43209

[1] Number of clusters = $2G$ with $G = 50$ throughout. Number of replications for each ρ is 2000. $N_{max} = 500$.

Table 14: Model 2 - $|\mathcal{M}_g| = \rho \cdot N_g$ with $G = 50$ - Matching on X_g and N_g[1]

N_{max}/N_{min}	VCE	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
				Coverage		
	$\hat{v}^2$	0.9540	0.9540	0.9455	0.9530	0.9515
1.11	CR	0.9870	0.9880	0.9890	0.9895	0.9890
	PCVE	0.9870	0.9875	0.9890	0.9895	0.9890
	$\hat{v}^2$	0.9530	0.9525	0.9525	0.9560	0.9565
1.42	CR	0.9865	0.9900	0.9880	0.9865	0.9890
	PCVE	0.9870	0.9900	0.9875	0.9870	0.9890
	$\hat{v}^2$	0.9500	0.9485	0.9475	0.9455	0.9520
1.99	CR	0.9860	0.9885	0.9870	0.9890	0.9880
	PCVE	0.9860	0.9895	0.9870	0.9885	0.9880
	$\hat{v}^2$	0.9460	0.9470	0.9475	0.9480	0.9470
3.31	CR	0.9850	0.9890	0.9875	0.9840	0.9845
	PCVE	0.9845	0.9905	0.9870	0.9845	0.9850
	$\hat{v}^2$	0.9475	0.9420	0.9450	0.9475	0.9455
9.80	CR	0.9790	0.9850	0.9820	0.9860	0.9835
	PCVE	0.9785	0.9855	0.9820	0.9865	0.9835
				Average Length		
	$\hat{v}^2$	0.73376	0.73321	0.73231	0.73105	0.72948
1.11	CR	0.96896	0.96879	0.96889	0.96688	0.96575
	PCVE	0.96936	0.96884	0.96858	0.96676	0.96545
	$\hat{v}^2$	0.73555	0.73355	0.73364	0.73220	0.73197
1.42	CR	0.97830	0.97690	0.97677	0.97497	0.97623
	PCVE	0.97814	0.97681	0.97712	0.97551	0.97590
	$\hat{v}^2$	0.74875	0.74732	0.74460	0.74303	0.74426
1.99	CR	0.99345	0.99257	0.98995	0.98866	0.99003
	PCVE	0.99326	0.99258	0.99005	0.98826	0.99013
	$\hat{v}^2$	0.77167	0.77033	0.76704	0.76421	0.76631
3.31	CR	1.01607	1.01609	1.01194	1.00929	1.01166
	PCVE	1.01571	1.01559	1.01192	1.00913	1.01135
	$\hat{v}^2$	0.81196	0.81261	0.81153	0.80961	0.80766
9.80	CR	1.05338	1.05499	1.05399	1.05304	1.05132
	PCVE	1.05429	1.05492	1.05480	1.05326	1.05128

[1] Number of clusters = $2G$ with $G = 50$ throughout. Number of replications for each ρ is 2000. $N_{max} = 500$.

Table 15: Model 1 - $|\mathcal{M}_g| = \rho \cdot N_g$ with $G = 250$ - Matching on X_g and N_g[1]

N_{max}/N_{min}	VCE	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
			Coverage			
	$\hat{v}^2$	0.9460	0.9385	0.9540	0.9550	0.9535
1.11	CR	1	1	1	1	1
	PCVE	0.9460	0.9395	0.9530	0.9540	0.9510
	$\hat{v}^2$	0.9505	0.9455	0.9570	0.9425	0.9555
1.42	CR	1	1	1	1	1
	PCVE	0.9530	0.9470	0.9570	0.9400	0.9555
	$\hat{v}^2$	0.9505	0.9470	0.9530	0.9565	0.9365
1.99	CR	1	1	1	1	1
	PCVE	0.9495	0.9500	0.9555	0.9575	0.9370
	$\hat{v}^2$	0.9410	0.9475	0.9400	0.9450	0.9455
3.31	CR	1	1	1	1	1
	PCVE	0.9425	0.9465	0.9395	0.9440	0.9465
	$\hat{v}^2$	0.9510	0.9485	0.9455	0.9495	0.9405
9.80	CR	1	1	1	1	1
	PCVE	0.9470	0.9480	0.9500	0.9510	0.9430
			Average Length			
	$\hat{v}^2$	0.14449	0.14312	0.14249	0.14173	0.14127
1.11	CR	0.73103	0.73070	0.73057	0.73044	0.73034
	PCVE	0.14444	0.14309	0.14229	0.14165	0.14116
	$\hat{v}^2$	0.10899	0.10714	0.10574	0.10481	0.10393
1.42	CR	0.73644	0.73611	0.73590	0.73575	0.73559
	PCVE	0.10897	0.10709	0.10560	0.10464	0.10387
	$\hat{v}^2$	0.10480	0.10230	0.10073	0.09930	0.09825
1.99	CR	0.74537	0.74501	0.74487	0.74471	0.74447
	PCVE	0.10477	0.10234	0.10059	0.09919	0.09814
	$\hat{v}^2$	0.11023	0.10740	0.10511	0.10385	0.10256
3.31	CR	0.76179	0.76141	0.76113	0.76098	0.76078
	PCVE	0.11014	0.10734	0.10523	0.10372	0.10248
	$\hat{v}^2$	0.12613	0.12277	0.12007	0.11823	0.11673
9.80	CR	0.79667	0.79620	0.79573	0.79560	0.79545
	PCVE	0.12599	0.12262	0.12008	0.11836	0.11675

[1] Number of clusters = $2G$ with $G = 250$ throughout. Number of replications for each ρ is 2000. $N_{max} = 500$.

Table 16: Model 2 - $|\mathcal{M}_g| = \rho \cdot N_g$ with $G = 250$ - Matching on X_g and N_g[1]

N_{max}/N_{min}	VCE	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
		Coverage				
	$\hat{v}^2$	0.9525	0.9500	0.9515	0.9545	0.9490
1.11	CR	0.9935	0.9935	0.9940	0.9940	0.9955
	PCVE	0.9930	0.9930	0.9940	0.9940	0.9955
	$\hat{v}^2$	0.9525	0.9520	0.9505	0.9545	0.9515
1.42	CR	0.9935	0.9945	0.9965	0.9935	0.9950
	PCVE	0.9935	0.9945	0.9970	0.9935	0.9960
	$\hat{v}^2$	0.9490	0.9480	0.9535	0.9555	0.9515
1.99	CR	0.9950	0.9940	0.9945	0.9925	0.9950
	PCVE	0.9945	0.9940	0.9945	0.9925	0.9940
	$\hat{v}^2$	0.9470	0.9510	0.9480	0.9480	0.9465
3.31	CR	0.9930	0.9925	0.9940	0.9950	0.9935
	PCVE	0.9925	0.9925	0.9935	0.9950	0.9935
	$\hat{v}^2$	0.9505	0.9510	0.9520	0.9550	0.9470
9.80	CR	0.9935	0.9915	0.9935	0.9935	0.9935
	PCVE	0.9925	0.9915	0.9935	0.9930	0.9940
		Average Length				
	$\hat{v}^2$	0.32094	0.32029	0.31989	0.31952	0.31931
1.11	CR	0.43789	0.43732	0.43698	0.43672	0.43658
	PCVE	0.43788	0.43732	0.43700	0.43676	0.43657
	$\hat{v}^2$	0.32054	0.32012	0.31967	0.31917	0.31898
1.42	CR	0.44196	0.44168	0.44144	0.44098	0.44075
	PCVE	0.44193	0.44176	0.44142	0.44099	0.44083
	$\hat{v}^2$	0.32540	0.32455	0.32406	0.32367	0.32335
1.99	CR	0.44862	0.44792	0.44768	0.44744	0.44705
	PCVE	0.44870	0.44802	0.44771	0.44746	0.44718
	$\hat{v}^2$	0.33416	0.33324	0.33299	0.33244	0.33192
3.31	CR	0.45933	0.45865	0.45869	0.45818	0.45777
	PCVE	0.45940	0.45876	0.45880	0.45823	0.45785
	$\hat{v}^2$	0.35255	0.35044	0.34980	0.34945	0.34852
9.80	CR	0.48124	0.47943	0.47896	0.47883	0.47811
	PCVE	0.48147	0.47949	0.47922	0.47898	0.47819

[1] Number of clusters = $2G$ with $G = 250$ throughout. Number of replications for each ρ is 2000. $N_{max} = 500$.