Inference in Cluster Randomized Trials with Matched Pairs *

Yuehao Bai
Department of Economics
University of Michigan
yuehaob@umich.edu

Jizhou Liu
Booth School of Business
University of Chicago
jliu32@chicagobooth.edu

Azeem M. Shaikh
Department of Economics
University of Chicago
amshaikh@uchicago.edu

Max Tabord-Meehan
Department of Economics
University of Chicago
maxtm@uchicago.edu

March 27, 2023

Abstract

This paper considers the problem of inference in cluster randomized trials where treatment status is determined according to a "matched pairs" design. Here, by a cluster randomized experiment, we mean one in which treatment is assigned at the level of the cluster; by a "matched pairs" design we mean that a sample of clusters is paired according to baseline, cluster-level covariates and, within each pair, one cluster is selected at random for treatment. We study the large sample behavior of a weighted difference-in-means estimator and derive two distinct sets of results depending on if the matching procedure does or does not match on cluster size. We then propose a variance estimator which is consistent in either case. Combining these results establishes the asymptotic exactness of tests based on these estimators. Next, we consider the properties of two common testing procedures based on t-tests constructed from linear regressions, and argue that both are generally conservative in our framework. Finally, we study the behavior of a randomization test which permutes the treatment status for clusters within pairs, and establish its finite sample and asymptotic validity for testing specific null hypotheses. A simulation study confirms the practical relevance of our theoretical results.

KEYWORDS: Experiment, matched pairs, cluster-level randomization, randomized controlled trial, treatment assignment

JEL classification codes: C12, C14

*We thank Xun Huang for providing excellent research assistance.

1 Introduction

This paper studies the problem of inference in cluster randomized experiments where treatment status is determined according to a "matched pairs" design. Here, by a cluster randomized experiment, we mean one in which treatment is assigned at the level of the cluster; by a "matched pairs" design we mean that the sample of clusters is paired according to baseline, cluster-level covariates and, within each pair, one cluster is selected at random for treatment. Cluster matched pair designs feature prominently in all parts of the sciences: examples in economics include Banerjee et al. (2015) and Crépon et al. (2015).

Following recent work in Bugni et al. (2022), we develop our results in a sampling framework where clusters are realized as a random sample from a population of clusters. Importantly, in this framework cluster sizes are modeled as random and "non-ignorable", meaning that "large" clusters and "small" clusters may be heterogeneous, and, in particular, the effects of the treatment may vary across clusters of differing sizes. The framework additionally allows for the possibility of two-stage sampling, in which a subset of units is sampled from the set of units within each sampled cluster.

We first study the large sample behavior of a weighted difference-in-means estimator under two distinct sets of assumptions on the matching procedure. Specifically, we distinguish between settings where the matching procedure does or does not match on a function of cluster size. For both cases we establish conditions under which our estimator is asymptotically normal and derive simple closed form expressions for the asymptotic variance. Using these results we establish formally that employing cluster size as a matching variable delivers a weak improvement in asymptotic efficiency relative to only matching on baseline covariates. We then propose a variance estimator which is consistent for either asymptotic variance depending on the nature of the matching procedure. Combining these results establishes the asymptotic exactness of tests based on our estimators. We then consider the asymptotic properties of two commonly recommended inference procedures based on linear regressions of the individual-level outcomes on a constant and cluster-level treatment. The first inference procedure clusters at the level of treatment assignment. The second inference procedure clusters at the level of assignment pairs, as recently recommended in de Chaisemartin and Ramirez-Cuellar (2019). We establish that both procedures are generally conservative in our framework.

Next, we study the behavior of a randomization test which permutes the treatment status for clusters within pairs. We establish the finite-sample validity of such a test for testing a certain null hypothesis related to the equality of outcome distributions between treatment and control clusters, and then establish asymptotic validity for testing null hypotheses about the size-weighted average treatment effect. We emphasize, however, that the later result relies heavily on our choice of test-statistic, which is studentized using our novel variance estimator. In simulations, we find that this randomization test controls size more reliably than any of the other inference procedures we consider in the paper, while delivering comparable power.

The analysis of data from cluster randomized experiments and data from experiments with matched pairs has received considerable attention, but most work has focused on only one of these two features at a time. Recent work on the analysis of cluster randomized experiments includes Athey and Imbens (2017), Hayes and Moulton (2017), and Su and Ding (2021) (see Bugni et al., 2022, for a general discussion of this literature as well as further references). Recent work on the analysis of matched pairs experiments includes Jiang et al. (2020), Cytrynbaum (2021), Bai et al. (2023), and Bai (2022) (see Bai et al., 2022, for a general discussion of this literature as well as further references). Two papers which focus specifically on the analysis of cluster randomized experiments with matched pairs are Imai et al. (2009) and de Chaisemartin and Ramirez-Cuellar (2019). Both papers maintain a finite-population perspective, where the primary source of uncertainty is "design-based", stemming from the randomness in treatment assignment. In such a framework, both papers study the finite and large-sample behavior of difference-in-means type estimators and propose corresponding variance estimators which are shown to be conservative. In contrast, our paper maintains a "super-population" sampling framework and proposes a novel variance estimator which is shown to be asymptotically exact in our setting.

The remainder of the paper is organized as follows. In Section 2 we describe our setup and notation. Section 3 presents our main results. Section 4 studies the finite sample behavior of our proposed tests via a simulation study.

2 Setup and Notation

In this section we introduce the notation and assumptions which are common to both matching procedures considered in Section 3. We broadly follow the setup and notation developed in Bugni et al. (2022). Let $Y_{i,g} \in \mathbf{R}$ denote the (observed) outcome of interest for the ith unit in the gth cluster, $D_g \in \{0,1\}$ denote the treatment received by the gth cluster, $X_g \in \mathbf{R}^k$ the observed, baseline covariates for the gth cluster, and $N_g \in \mathbf{Z}_+$ the size of the gth cluster. In what follows we sometimes refer to the vector (X_g, N_g) as W_g. Further denote by $Y_{i,g}(d)$ the potential outcome of the ith unit in cluster g, when all units in the gth cluster receive treatment $d \in \{0,1\}$. As usual, the observed outcome and potential outcomes are related to treatment assignment by the relationship

$$Y_{i,g} = Y_{i,g}(1)D_g + Y_{i,g}(0)(1 - D_g) \ . \quad (1)$$

In addition, define $\mathcal{M}_g$ to be the (possibly random) subset of $\{1, 2, \ldots, N_g\}$ corresponding to the observations within the gth cluster that are sampled by the researcher. We emphasize that a realization of $\mathcal{M}_g$ is a *set* whose cardinality we denote by $|\mathcal{M}_g|$, whereas a realization of N_g is a positive integer. For example, in the event that all observations in a cluster are sampled, $\mathcal{M}_g = \{1, \ldots, N_g\}$ and $|\mathcal{M}_g| = N_g$. We assume throughout that our sample consists of $2G$ clusters and denote by P_G the distribution of the observed data

$$Z^{(G)} := (((Y_{i,g} : i \in \mathcal{M}_g), D_g, X_g, N_g) : 1 \le g \le 2G) \ ,$$

and by Q_G the distribution of

$$(((Y_{i,g}(1), Y_{i,g}(0) : 1 \le i \le N_g), \mathcal{M}_g, X_g, N_g) : 1 \le g \le 2G) \ .$$

Note that P_G is determined jointly by (1) together with the distribution of $D^{(G)} := (D_g : 1 \le g \le 2G)$ and Q_G, so we will state our assumptions below in terms of these two quantities.

We now describe some preliminary assumptions on Q_G that we maintain throughout the paper. In order to do so, it is useful to introduce some further notation. To this end, for $d \in \{0, 1\}$, define

$$\bar{Y}_g(d) := \frac{1}{|\mathcal{M}_g|} \sum_{i \in \mathcal{M}_g} Y_{i,g}(d) \ .$$

Further define $R_G(\mathcal{M}_g^{(G)}, X^{(G)}, N^{(G)})$ to be the distribution of

$$((Y_{i,g}(1), Y_{i,g}(0) : 1 \leq i \leq N_g) : 1 \leq g \leq 2G) \ \Big| \ \mathcal{M}_g^{(G)}, X^{(G)}, N^{(G)} \ ,$$

where $\mathcal{M}_g^{(G)} := (\mathcal{M}_g : 1 \leq g \leq 2G)$, $X^{(G)} := (X_g : 1 \leq g \leq 2G)$ and $N^{(G)} := (N_g : 1 \leq g \leq 2G)$. Note that Q_G is completely determined by $R_G(\mathcal{M}_g^{(G)}, X^{(G)}, N^{(G)})$ and the distribution of $(\mathcal{M}_g^{(G)}, X^{(G)}, N^{(G)})$. The following assumption states our main requirements on Q_G using this notation.

Assumption 2.1. The distribution Q_G is such that

(a) $\{(\mathcal{M}_g, X_g, N_g), 1 \leq g \leq 2G\}$ is an i.i.d. sequence of random variables.

(b) For some family of distributions $\{R(m, x, n) : (m, x, n) \in \text{supp}(\mathcal{M}_g, X_g, N_g)\}$,

$$R_G(\mathcal{M}_g^{(G)}, X^{(G)}, N^{(G)}) = \prod_{1 \leq g \leq 2G} R(\mathcal{M}_g, X_g, N_g) \ .$$

(c) $P\{|\mathcal{M}_g| \geq 1\} = 1$ and $E[N_g^2] < \infty$.

(d) For some $C < \infty$, $P\{E[Y_{i,g}^2(d)|X_g, N_g] \leq C$ for all $1 \leq i \leq N_g\} = 1$ for all $d \in \{0, 1\}$ and $1 \leq g \leq 2G$.

(e) $\mathcal{M}_g \perp\!\!\!\perp (Y_{i,g}(1), Y_{i,g}(0) : 1 \leq i \leq N_g) \ \big| \ X_g, N_g$ for all $1 \leq g \leq 2G$.

(f) For $d \in \{0, 1\}$ and $1 \leq g \leq 2G$,

$$E[\bar{Y}_g(d)|N_g] = E\left[\frac{1}{N_g} \sum_{1 \leq i \leq N_g} Y_{i,g}(d) \Big| N_g\right] \text{ w.p.1} \ .$$

As shown in Bugni et al. (2022), an important implication of Assumptions 2.1(a)-(b) for our purposes is that

$$\left\{\left(\bar{Y}_g(1), \bar{Y}_g(0), |\mathcal{M}_g|, X_g, N_g\right), 1 \leq g \leq 2G\right\} \ , \tag{2}$$

is an i.i.d. sequence of random variables. Assumptions 2.1 (c)-(d) impose some mild regularity on the (conditional) moments of the distribution of cluster sizes and potential outcomes, Assumptions 2.1 (e)-(f) impose high-level restrictions on the two-stage sampling procedure. Bugni et al. (2022) provide a detailed discussion of the implications of each of these assumptions.

Our object of interest is the size-weighted cluster-level average treatment effect, which may be expressed in our notation as

$$\Delta(Q_G) = E\left[\frac{N_g}{E[N_g]}\left(\frac{1}{N_g}\sum_{i=1}^{N_g}(Y_{i,g}(1) - Y_{i,g}(0))\right)\right] = E\left[\frac{1}{E[N_g]}\sum_{i=1}^{N_g}(Y_{i,g}(1) - Y_{i,g}(0))\right] .$$

This parameter, which weights the cluster-level average treatment effects proportional to cluster size, can be thought of as the average treatment effect where individuals are the unit of interest. Note that Assumptions 2.1 (a)–(b) imply that we may express $\Delta(Q_G)$ as a function of R and the common distribution of $(\mathcal{M}_g, X_g, N_g)$. In particular, this implies that $\Delta(Q_G)$ does not depend on G. Accordingly, in what follows we simply denote $\Delta = \Delta(Q_G)$.

Throughout the paper we study the asymptotic behavior of a size-weighted difference-in-means estimator:

$$\hat{\Delta}_G := \hat{\mu}_G(1) - \hat{\mu}_G(0) ,$$

where

$$\hat{\mu}_G(d) := \frac{1}{N(d)}\sum_{g=1}^{2G} I\{D_g = d\}\frac{N_g}{|\mathcal{M}_g|}\sum_{i\in\mathcal{M}_g} Y_{i,g} ,$$

with

$$N(d) := \sum_{g=1}^{2G} N_g I\{D_g = d\} .$$

Note that this estimator may be obtained as the estimator of the coefficient of D_g in a weighted least squares regression of $Y_{i,g}$ on a constant and D_g with weights equal to $\sqrt{N_g/|\mathcal{M}_g|}$. In the special case that all observations in each cluster are sampled, so that $\mathcal{M}_g = \{1, 2, \ldots, N_g\}$ for all $1 \le g \le G$ with probability one, this estimator collapses to the standard difference-in-means estimator.

Remark 2.1. Following the recommendations in Bruhn and McKenzie (2009) and Glennerster and Takavarasha (2013), it is common practice to conduct inference in matched pair experiments using the standard errors obtained from a regression of individual level outcomes on treatment and a collection of pair-level fixed effects. We do not analyze the asymptotic properties of such an approach for two reasons. First, in the context of individual-level randomized experiments, Bai et al. (2022) and Bai et al. (2023) argue that such a regression estimator is in fact numerically equivalent to the simple difference-in-means estimator, but that the resulting standard errors are generally conservative. This result generalizes immediately to the clustered setting in the special case where all clusters are the same size and $\mathcal{M}_g = \{1, 2, \ldots, N_g\}$. Second, when cluster sizes vary, this numerical equivalence no longer holds, and in such cases de Chaisemartin and Ramirez-Cuellar (2019) argue (in an alternative inferential framework) that the corresponding regression estimator may no longer be consistent for the average treatment effect of interest. ■

Remark 2.2. Bugni et al. (2022) also define an alternative treatment effect parameter given by

$$\Delta^{\text{eq}}(Q_G) = E\left[\frac{1}{N_g}\sum_{i=1}^{N_g}(Y_{i,g}(1) - Y_{i,g}(0))\right] .$$

This parameter, which weights the cluster-level average treatment effects equally regardless of cluster size, can be thought of as the average treatment effect where the clusters themselves are the units of interest. It can be shown under appropriate assumptions that for this parameter, the analysis of matched-pair designs for individual-level treatments developed in Bai et al. (2022) applies directly to the data obtained from the cluster-level averages $\{(\bar{Y}_g, D_g, X_g, N_g) : 1 \le g \le 2G\}$, where $\bar{Y}_g = \frac{1}{|\mathcal{M}_g|}\sum_{i \in \mathcal{M}_g} Y_{i,g}$. As a result, we do not pursue a detailed description of inference for this parameter in the paper. ■

3 Main Results

3.1 Asymptotic Behavior of $\hat{\Delta}_G$ for Cluster-Matched Pair Designs

In this section we consider the asymptotic behavior of $\hat{\Delta}_G$ for two distinct types of cluster-matched pair designs. Section 3.1.1 studies a setting where cluster size is *not* used as a matching variable when forming pairs. Section 3.1.2 considers the setting where we do allow for pairs to be matched based on cluster size in an appropriate sense made formal below.

3.1.1 Not Matching on Cluster Size

In this section we consider a setting where cluster size is not used as a matching variable. First we describe our formal assumptions on the mechanism determining treatment assignment. The G pairs of clusters may be represented by the sets

$$\{\pi(2g-1), \pi(2g)\} \text{ for } g = 1, ..., G \ ,$$

where $\pi = \pi_G(X^{(G)})$ is a permutation of $2G$ elements. Given such a π, we assume that treatment status is assigned as follows:

Assumption 3.1. Treatment status is assigned so that

$$\{((Y_{i,g}(1), Y_{i,g}(0) : 1 \le i \le N_g), N_g, \mathcal{M}_g)\}_{g=1}^{2G} \perp\!\!\!\perp D^{(G)} | X^{(G)} \ .$$

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

We further require that the clusters in each pair be "close" in terms of their baseline covariates in the following sense:

Assumption 3.2. The pairs used in determining treatment assignment satisfy

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

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

Bai et al. (2022) provide results which facilitate the construction of pairs which satisfy Assumption 3.2. For instance, if $\dim(X_g) = 1$ and we order clusters from smallest to largest according to X_g and then pair adjacent units, it follows from Theorem 4.1 in Bai et al. (2022) that Assumption 3.2 is satisfied if $E[X_g^2] < \infty$. Next, we state the additional assumptions on Q_G we require beyond those stated in Assumption 2.1:

Assumption 3.3. The distribution Q_G is such that

(a) $E[\bar{Y}_g^r(d)N_g^\ell | X_g = x]$, are Lipschitz for $d \in \{0,1\}$, $r, \ell \in \{0,1,2\}$,

(b) For some $C < \infty$, $P\{E[N_g|X_g] \le C\} = 1$.

Assumption 3.3(a) is a smoothness requirement analogous to Assumption 2.1(c) in Bai et al. (2022) that ensures that units within clusters which are "close" in terms of their baseline covariates are suitably comparable. Assumption 3.3(b) imposes an additional restriction on the distribution of cluster sizes beyond what is stated in Assumption 2.1(c). Under these assumptions we obtain the following result:

Theorem 3.1. *Under Assumptions 2.1, 3.1, 3.2, 3.3,*

$$\sqrt{G}(\hat{\Delta}_G - \Delta) \xrightarrow{d} N(0, \omega^2) ,$$

as $G \to \infty$, where

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

with

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

Note that the asymptotic variance we obtain in Theorem 3.1 corresponds exactly to the asymptotic variance of the difference-in-means estimator for matched pairs designs with individual-level assignment (as derived in Bai et al., 2022), but with transformed cluster-level potential outcomes given by $\tilde{Y}_g(d)$. Accordingly, our result collapses exactly to theirs when $P\{N_g = 1\} = 1$. Theorem 3.1 also quantifies the gain in precision obtained from using a matched pairs design versus complete randomization (i.e. assigning half of the clusters to treatment at random): it can be shown that the limiting distribution of $\hat{\Delta}_G$ under complete randomization is given by

$$\sqrt{G}(\hat{\Delta}_G - \Delta) \xrightarrow{d} N(0, \omega_0^2) ,$$

where $\omega_0^2 = E[\tilde{Y}_g^2(1)] + E[\tilde{Y}_g^2(0)]$. We thus immediately obtain that $\omega^2 \le \omega_0^2$. Moreover, this inequality is strict unless $E[\tilde{Y}_g(1) + \tilde{Y}_g(0)|X_g] = 0$.

3.1.2 Matching on Cluster Size

In this section we repeat the exercise in Section 3.1.1 in a setting where the assignment mechanism matches on baseline characteristics *and* (some function of) cluster size in an appropriate sense to be made formal below. First, we describe how to modify our assumptions on the mechanism determining treatment assignment.

The G pairs of clusters are still represented by the sets

$$\{\pi(2g-1), \pi(2g)\} \text{ for } g = 1, ..., G ,$$

however, now we allow the permutation $\pi = \pi_G(X^{(G)}, N^{(G)}) = \pi_G(W^{(G)})$ to be a function of cluster size. Given such a π, we assume that treatment status is assigned as follows:

Assumption 3.4. Treatment status is assigned so that

$$\{((Y_{i,g}(1), Y_{i,g}(0) : 1 \le i \le N_g), \mathcal{M}_g)\}_{g=1}^{2G} \perp\!\!\!\perp D^{(G)} | (X^{(G)}, N^{(G)}) .$$

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

We also modify the assumption on how pairs are formed:

Assumption 3.5. The pairs used in determining treatment assignment satisfy

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

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

Unlike for Assumption 3.2, the discussion in Bai et al. (2022) does not provide conditions for matching algorithms which guarantee that Assumption 3.5 holds. Accordingly, in Proposition 3.1 we provide lower level sufficient conditions for Assumption 3.5 which are easier to verify, at the expense of imposing stronger assumptions on the distribution of cluster sizes:

Proposition 3.1. *Suppose $E[N_g^4] < \infty$ and*

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

for $r \in \{1, 2, 3, 4\}$, then Assumption 3.5 holds.

We also modify the smoothness requirement as follows:

Assumption 3.6. The distribution Q_G is such that $E[\bar{Y}_g^r(d) | W_g = w]$ are Lipschitz for $d \in \{0, 1\}$, $r \in \{1, 2\}$.

We then obtain the following analogue to Theorem 3.1:

Theorem 3.2. *Under Assumptions 2.1, 3.4, 3.5, 3.6,*

$$\sqrt{G}(\hat{\Delta}_G - \Delta) \xrightarrow{d} N(0, \nu^2) ,$$

as $G \to \infty$, where

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

with

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

Note that the asymptotic variance ν^2 has exactly the same form as ω^2 from Section 3.1.1, with the only difference being that the final term of the expression conditions on both cluster characteristics X_g and cluster size N_g. From this result it then follows that matching on cluster size in addition to cluster characteristics leads to a weakly lower asymptotic variance. To see this, note that by comparing ω^2 and ν^2 we obtain that

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

It then follows by the law of iterated expectations and Jensen's inequality that $\omega^2 \geq \nu^2$.

3.2 Variance Estimation

In this section we construct variance estimators for the asymptotic variances ω^2 and ν^2 obtained in Section 3.1. In fact, we propose a *single* variance estimator that is consistent for *both* ω^2 and ν^2 depending on the nature of the matching procedure. As noted in the discussion following Theorem 3.1, the expressions for ω^2 and ν^2 correspond exactly to the asymptotic variance obtained in Bai et al. (2022) with the individual-level outcome replaced by a cluster-level transformed outcome. We thus follow the variance construction from Bai et al. (2022), but replace the individual outcomes with feasible versions of these transformed outcomes. To that end, consider the observed adjusted outcome defined as:

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

where

$$\bar{Y}_g = \frac{1}{|\mathcal{M}_g|}\sum_{i\in\mathcal{M}_g} Y_{i,g} ,$$

We then propose the following variance estimator:

$$\hat{v}_G^2 = \hat{\tau}_G^2 - \frac{1}{2}\hat{\lambda}_G^2 , \tag{3}$$

where

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

Note that the construction of $\hat{v}_G^2$ can be motivated using the same intuition as the variance estimators studied in Bai et al. (2022) and Bai et al. (2023): to consistently estimate quantities like (for instance)

$E[E[\tilde{Y}_g(1)|X_g]E[\tilde{Y}_g(0)|X_g]]$ which appear in ω^2, we average across "pairs of pairs" of clusters. As a consequence, we will additionally require that the matching algorithm satisfy the condition that "pairs of pairs" of clusters are sufficiently close in terms of their baseline covariates/cluster size, as formalized in the following two assumptions:

Assumption 3.7. The pairs used in determining treatment status satisfy

$$\frac{1}{G} \sum_{1 \leq j \leq \lfloor \frac{G}{2} \rfloor} \left| X_{\pi(4j-k)} - X_{\pi(4j-\ell)} \right|^2 \xrightarrow{P} 0$$

for any $k \in \{2, 3\}$ and $l \in \{0, 1\}$.

Assumption 3.8. The pairs used in determining treatment status satisfy

$$\frac{1}{G} \sum_{1 \leq j \leq \lfloor \frac{G}{2} \rfloor} N^2_{\pi(4j-k)} \left| W_{\pi(4j-k)} - W_{\pi(4j-\ell)} \right|^2 \xrightarrow{P} 0$$

for any $k \in \{2, 3\}$ and $l \in \{0, 1\}$.

We then obtain the following two consistency results for the estimator $\hat{v}^2_G$:

Theorem 3.3. *Suppose Assumption 2.1 holds. If additionally Assumptions 3.1–3.3, 3.7 hold, then*

$$\hat{v}^2_G \xrightarrow{P} \omega^2 \ .$$

Alternatively, if Assumptions 3.4–3.6, 3.8 hold, then

$$\hat{v}^2_G \xrightarrow{P} \nu^2 \ .$$

Next, we derive the probability limits of two commonly recommended variance estimators obtained from a (weighted) linear regression of the individual-level outcomes $Y_{i,g}$ on a constant and cluster-level treatment D_g. The first variance estimator we consider, which we denote by $\hat{\omega}^2_{\text{CR,G}}$, is simply the cluster-robust variance estimator of the coefficient of D_g as defined in equation (15) in the appendix. Theorem 3.4 derives the probability limit of $\hat{\omega}^2_{\text{CR,G}}$ under a matched pair design which matches on baseline covariates as defined in Section 3.1.1, and shows that it is generally too large relative to ω^2.

Theorem 3.4. *Under Assumptions 2.1, 3.1–3.3,*

$$\hat{\omega}^2_{\text{CR,G}} \xrightarrow{P} E[\tilde{Y}_g(1)^2] + E[\tilde{Y}_g(0)^2] \geq \omega^2 \ ,$$

with equality if and only if

$$E[\tilde{Y}_g(1) + \tilde{Y}_g(0)|X_g] = 0 \ . \tag{4}$$

The next variance estimator we consider, which we denote by $\hat{\omega}^2_{\text{PCVE,G}}$, is the variance estimator of the coefficient of D_g obtained from clustering on the assignment *pairs* of clusters as defined in equation (16) in the appendix. de Chaisemartin and Ramirez-Cuellar (2019) call this the pair-cluster variance estimator (PCVE). Theorem 3.5 derives the probability limit of $\hat{\omega}^2_{\mathrm{PCVE,G}}$ in the special case where $N_g = k$ for $g = 1, \dots, 2G$ for some fixed k and $|\mathcal{M}_g| = N_g$, and shows that it is generally too large relative to ω^2.

Theorem 3.5. *Suppose Assumptions 2.1, 3.1–3.3 hold. If in addition we impose that $N_g = k$ for $g = 1, \dots, 2G$ for some fixed positive integer k and that $|\mathcal{M}_g| = N_g$, then*

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

with equality if and only if

$$E[\tilde{Y}_g(1) - \tilde{Y}_g(0) | X_g] = 0 \ . \tag{5}$$

Although we do not derive the probability limit of $\hat{\omega}^2_{\mathrm{PCVE,G}}$ in the general case, our simulation evidence in Section 4 suggests that the limit of $\hat{\omega}^2_{\mathrm{PCVE,G}}$ remains conservative, and that the conditions under which it is consistent for ω^2 are the same as those in equation (5). From Theorems 3.4 and 3.5 we obtain that neither cluster-robust standard error is consistent for ω^2 unless the baseline covariates are irrelevant for potential outcomes in a certain sense.[1] However, equation (5) holds when the average treatment difference for the sampled units in a cluster are homogeneous, in the sense that $\bar{Y}_g(1) - \bar{Y}_g(0)$ is constant, which is not the case for equation (4).

3.3 Randomization Tests

In this section we study the properties of a randomization test based on the idea of permuting the treatment assignments for clusters within pairs. In Section 3.3.1 we present some finite-samples properties of our proposed test, in Section 3.3.2 we establish its large sample validity for testing the null hypothesis $H_0 : \Delta(Q_G) = 0$.

First we construct the test. Denote by $\mathbf{H}_G$ the group of all permutations on $2G$ elements and by $\mathbf{H}_G(\pi)$ the subgroup that only permutes elements within pairs defined by π:

$$\mathbf{H}_G(\pi) = \{h \in \mathbf{H}_G : \{\pi(2g-1), \pi(2g)\} = \{h(\pi(2j-1)), h(\pi(2j))\} \text{ for } 1 \leq g \leq G\} \ .$$

Define the action of $h \in \mathbf{H}_G(\pi)$ on $Z^{(G)}$ as follows:

$$hZ^{(G)} = \{((Y_{i,g} : i \in \mathcal{M}_g), D_{h(g)}, X_g, N_g) : 1 \leq g \leq 2G\} \ .$$

The randomization test we consider is then given by

$$\phi^{\mathrm{rand}}_G(Z^{(G)}) = I\{T_G(Z^{(G)}) > \hat{R}^{-1}_G(1-\alpha)\} \ ,$$

[1] We note that the conditions under which $\hat{\omega}^2_{\mathrm{CR,G}}$ and $\hat{\omega}^2_{\mathrm{PCVE,G}}$ are consistent for ω^2 are exactly analogous to the conditions under which Bai et al. (2022) derive (in the setting of an individual-level matched pairs experiment) that the two-sample t-test and matched pairs t-test are asymptotically exact, respectively.

where

$$\hat{R}_G(t) = \frac{1}{|\mathbf{H}_G(\pi)|} \sum_{h \in \mathbf{H}_G(\pi)} I\{T_G(hZ^{(G)}) \le t\} ,$$

with

$$T_G(Z^{(G)}) = \left| \frac{\sqrt{G}\hat{\Delta}_G}{\hat{v}_G} \right| .$$

Remark 3.1. As is often the case for randomization tests, $\hat{R}_G(t)$ may be difficult to compute in situations where $|\mathbf{H}_G(\pi)| = 2^G$ is large. In such cases, we may replace $\mathbf{H}_G(\pi)$ with a stochastic approximation $\hat{\mathbf{H}}_G = \{h_1, h_2, \ldots, h_B\}$, where h_1 is the identity transformation and $h_2, \ldots, h_B$ are i.i.d uniform draws from $\mathbf{H}_G(\pi)$. The results in Section 3.3.1 continue to hold with such an approximation; the results in Section 3.3.2 continue to hold provided $B \to \infty$ as $G \to \infty$. ■

3.3.1 Finite-Sample Results

In this section we explore some finite-sample properties of the proposed test. Consider testing the null hypothesis that the distribution of potential outcomes within a cluster are equal across treatment and control conditional on observable characteristics and cluster size:

$$H_0^{X,N} : (Y_{i,g}(1) : 1 \le i \le N_g)|(X_g, N_g) \stackrel{d}{=} (Y_{i,g}(0) : 1 \le i \le N_g)|(X_g, N_g) . \quad (6)$$

We then establish the following result:

Theorem 3.6. *Suppose Assumption 2.1 holds and that the treatment assignment mechanism satisfies Assumption 3.1 or 3.4. Then for the problem of testing (6) at level $\alpha \in (0,1)$, $\phi_G^{\text{rand}}(Z^{(G)})$ satisfies*

$$E[\phi_G^{\text{rand}}(Z^{(G)})] \le \alpha ,$$

under the null hypothesis.

Remark 3.2. The proof of Theorem 3.6 follows classical arguments that underlie the finite sample validity of randomization tests more generally. Accordingly, as in those arguments, the result continues to hold if the test statistic T_G is replaced by any other test-statistic which is a function of $\{(\bar{Y}_g, D_g, X_g, N_g) : 1 \le g \le 2G\}$.
■

3.3.2 Large-Sample Results

In this section we establish the large sample validity of the randomization test ϕ_G^{rand} for testing the null hypothesis

$$H_0 : \Delta(Q_G) = 0 . \quad (7)$$

In Remark 3.3 we describe how to modify the test for testing non-zero null hypotheses.

Theorem 3.7. *Suppose Q_G satisfies Assumption 2.1, and either*

- *Assumption 3.3 with treatment assignment mechanism satisfying Assumption 3.1 and 3.7 ,*
- *Assumption 3.6 with treatment assignment mechanism satisfying Assumptions 3.4 and 3.8 .*

Further, suppose that the probability limit of $\hat{v}_G^2$ is positive, then

$$\sup_{t \in \mathbf{R}} |\hat{R}_n(t) - (\Phi(t) - \Phi(-t))| \xrightarrow{P} 0 ,$$

where $\Phi(\cdot)$ is the standard normal CDF. Thus for the problem of testing (7) at level $\alpha \in (0,1)$, $\phi_G^{\text{rand}}(Z^{(G)})$ with $T_G(Z^{(G)})$ satisfies

$$\lim_{G \to \infty} E[\phi_G^{\text{rand}}(Z^{(G)})] = \alpha ,$$

under the null hypothesis.

Here we note that, unlike for the null hypothesis considered in Section 3.3.1, the choice of test-statistic is *crucial* for establishing the large sample validity of the test. Similar observations have been made in related contexts in Janssen (1997), Chung and Romano (2013), Bugni et al. (2018), Bai et al. (2022).

Remark 3.3. We briefly describe how to modify the test ϕ_G^{rand} for testing general null hypotheses of the form

$$H_0 : \Delta(Q_G) = \Delta_0 .$$

To this end, let

$$\tilde{Z}^{(G)} := (((Y_{i,g} - D_g \Delta_0 : i \in \mathcal{M}_g), D_g, X_g, N_g) : 1 \leq g \leq 2G) ,$$

then it can be shown that under the assumptions given in Theorem 3.7, the test $\phi_G^{\text{rand}}(\tilde{Z}^{(G)})$ obtained by replacing $Z^{(G)}$ with $\tilde{Z}^{(G)}$ satisfies

$$\lim_{G \to \infty} E[\phi_G^{\text{rand}}(\tilde{Z}^{(G)})] = \alpha ,$$

under the null hypothesis. ■

4 Simulations

In this section we examine the finite sample behavior of the inference procedures presented in Section 3. We further compare these procedures to tests and confidence intervals constructed using the standard cluster-robust variance estimator (CR) and the pair cluster variance estimator (PCVE) proposed in de Chaisemartin and Ramirez-Cuellar (2019). For $d \in \{0,1\}$, $1 \leq g \leq 2G$, the potential outcomes are generated according to the equation

$$Y_{i,g}(d) = \mu_d(X_g, X_g^{(N)}) + 2\epsilon_{d,i,g} .$$

Where, in each specification, $(X_g, X_g^{(N)})$, $g = 1, \ldots, 2G$ are i.i.d with $X_g, X_g^{(N)} \sim \ Beta(2,4)$, and $(\epsilon_{0,i,g}, \epsilon_{1,i,g})$, $g = 1, \ldots, 2G$, $i = 1, \ldots, N_g$ are i.i.d with $\epsilon_{0,i,g}, \epsilon_{1,i,g} \sim N(0,1)$. We consider the following two specifications for μ_d:

Model 1: $\mu_1(X_g, X_g^{(N)}) = \mu_0(X_g, X_g^{(N)}) = 10(X_g - 1/3) + 6(X_g^{(N)} - 1/3) + 2$.

Model 2: $\mu_1(X_g, X_g^{(N)}) = 10(X_g^2 - 1/7) + 6(X_g^{(N)} - 1/3) + 2$ and $\mu_0(X_g, X_g^{(N)}) = 0$.

Note that Model 1 satisfies the homogeneity condition in (5) whereas Model 2 does not. In both cases, N_g, $g = 1, \dots, 2G$ are i.i.d with $N_g \sim \ Binomial(R, X_g^{(N)}) + (500 - R)$, where R determines the difference in maximum and minimum cluster sizes. In particular R satisfies the property that $N_g \in [N_{min}, N_{max}]$ with $N_{max} - N_{min} = R$ and we consider $R \in \{49, 149, 249, 349, 449\}$ with $N_{max} = 500$ fixed. For each model and distribution of cluster sizes, we consider two alternative pair-matching procedures. First, we consider a design which matches clusters using X_g only. To construct these pairs, we sort the clusters according to X_g and pair adjacent clusters. Next, we consider a design which matches clusters using both X_g and N_g. To construct these pairs, we match the clusters according to their Mahalanobis distance using the non-bipartite matching algorithm from the R package `nbpMatching`.

Tables 1–4 report the coverage and average length of 95% confidence intervals constructed using our variance estimator as well as the CR and PCVE estimators. For Model 1 in Table 1 we find that, in accordance with Theorems 3.3–3.5, the CR variance estimator is extremely conservative, whereas our proposed variance estimator (denoted $\hat{v}_G^2$) and the PCVE variance estimator have exact coverage asymptotically. This translates to significantly smaller confidence intervals: on average the confidence intervals constructed using $\hat{v}_G^2$ or PCVE are almost half the length of those constructed using CR when $G \geq 50$. However, the confidence intervals constructed using $\hat{v}_G^2$ or PCVE undercover when $G < 50$. We find similar results when matching on both X_g and N_g in Table 2. Comparing across Tables 1 and 2 we find that, in line with the discussions following Theorems 3.1 and 3.2, matching on N_g in addition to X_g results in a large reduction in the average length of confidence intervals constructed using $\hat{v}_G^2$ (or PCVE), but no change to the average length of confidence intervals constructed using CR.

Moving to Model 2 in Tables 3 and 4, here we find that confidence intervals constructed using CR continue to be conservative, but now the confidence intervals constructed using PCVE are *also* conservative, and numerically very similar to those constructed using CR. In contrast, the confidence intervals constructed using $\hat{v}_G^2$ remain exact asymptotically. Once again this translates to smaller confidence intervals for $\hat{v}_G^2$: on average the confidence intervals constructed using $\hat{v}_G^2$ are approximately 25% smaller than those constructed using CR or PCVE when $G \geq 50$. However once again we find that the confidence intervals constructed using $\hat{v}_G^2$ can undercover when $G < 50$, with the size of the distortion growing as a function of the cluster size heterogeneity.

Next, to further address the small-sample coverage distortions observed in Tables 1-4, we study the size and power of 0.05-level hypothesis tests conducted using our proposed randomization test, as well as standard t-tests constructed using the CR and PCVE estimators, in Tables 5–6 below.[2] In Table 5 we find that tests based on the CR variance estimator are extremely conservative, and this translates to having essentially no power against our chosen alternative. Tests based on the PCVE estimator produce non-trivial power, but

[2]Here we move to studying the properties of hypothesis tests instead of confidence intervals to avoid having to perform test-inversion for our randomization test, but we expect that similar results would continue to hold for confidence intervals as well.

also size-distortions in small samples. In contrast, since Model 1 satisfies the null hypothesis considered in (6), our randomization test is finite-sample valid by construction, and displays comparable power to the PCVE-based test even when the later does not control size. When moving to Model 2 in Table 6 we are only guaranteed that the randomization test is asymptotically valid, but we find that the test is still able to control size in small samples as long as cluster-size heterogeneity is not too large (moreover, in such cases both the CR and PCVE-based tests also fail to control size). Finally, the randomization test displays favorable power relative to both the CR and PCVE-based tests throughout Table 6 except for some cases when $G = 12$.

Table 1: Model 1 - Matching on X_g[1]

N_{max}/N_{min}		$G=12$	$G=26$	$G=50$	$G=100$	$G=150$	$G=200$	$G=250$
Coverage								
	$\hat{v}_G^2$	0.9185	0.9290	0.9420	0.9465	0.9375	0.9460	0.9515
1.11	CR	0.9985	0.9990	0.9995	1	1	1	1
	PCVE	0.9230	0.9310	0.9385	0.9405	0.9395	0.9480	0.9520
	$\hat{v}_G^2$	0.9005	0.9345	0.9345	0.9480	0.9490	0.9545	0.9615
1.42	CR	0.9980	0.9995	0.9985	0.9995	0.9995	1	1
	PCVE	0.9035	0.9380	0.9375	0.9490	0.9495	0.9550	0.9595
	$\hat{v}_G^2$	0.9130	0.9330	0.9380	0.9385	0.9490	0.9455	0.9365
1.99	CR	0.9985	0.9985	1	1	1	1	0.9995
	PCVE	0.9095	0.9230	0.9420	0.9420	0.9495	0.9460	0.9350
	$\hat{v}_G^2$	0.9065	0.9180	0.9340	0.9415	0.9470	0.9450	0.9520
3.31	CR	0.9950	0.9980	0.9980	0.9985	1	0.9985	0.9995
	PCVE	0.8980	0.9155	0.9330	0.9380	0.9465	0.9470	0.9500
	$\hat{v}_G^2$	0.9035	0.9230	0.9420	0.9340	0.9440	0.9415	0.9495
9.80	CR	0.9925	0.9940	0.9970	0.9985	0.9975	0.9995	0.9990
	PCVE	0.8925	0.9100	0.9365	0.9330	0.9425	0.9385	0.9475
Average Length								
	$\hat{v}_G^2$	1.72150	1.16078	0.84582	0.59830	0.48784	0.42466	0.37936
1.11	CR	3.20593	2.21689	1.61886	1.15015	0.94053	0.81591	0.73010
	PCVE	1.69494	1.15171	0.84119	0.59746	0.48744	0.42415	0.37895
	$\hat{v}_G^2$	1.75019	1.18859	0.86476	0.61378	0.50112	0.43567	0.38917
1.42	CR	3.21821	2.22957	1.62982	1.15829	0.94732	0.82180	0.73543
	PCVE	1.72075	1.17840	0.86140	0.61286	0.50024	0.43527	0.38897
	$\hat{v}_G^2$	1.80502	1.23175	0.89937	0.63958	0.52250	0.45322	0.40566
1.99	CR	3.24165	2.25077	1.64811	1.17207	0.95862	0.83166	0.74408
	PCVE	1.77287	1.21936	0.89602	0.63843	0.52133	0.45352	0.40524
	$\hat{v}_G^2$	1.90111	1.30589	0.96060	0.68446	0.55910	0.48664	0.43505
3.31	CR	3.27892	2.28895	1.68064	1.19654	0.97928	0.84959	0.76030
	PCVE	1.85679	1.29128	0.95566	0.68299	0.55824	0.48568	0.43437
	$\hat{v}_G^2$	2.09510	1.45719	1.08057	0.77340	0.63320	0.55071	0.49226
9.80	CR	3.35580	2.36729	1.75068	1.24963	1.02275	0.88759	0.79443
	PCVE	2.03228	1.43576	1.07565	0.77259	0.63171	0.54976	0.49203

[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 2: Model 1 - Matching on X_g and N_g[1]

N_{max}/N_{min}		$G=12$	$G=26$	$G=50$	$G=100$	$G=150$	$G=200$	$G=250$
				Coverage				
	$\hat{v}_G^2$	0.9105	0.9285	0.9345	0.9430	0.9470	0.9495	0.9565
1.11	CR	1	1	1	1	1	1	1
	PCVE	0.9100	0.9260	0.9360	0.9460	0.9460	0.9480	0.9555
	$\hat{v}_G^2$	0.9210	0.9410	0.9400	0.9510	0.9490	0.9300	0.9445
1.42	CR	1	1	1	1	1	1	1
	PCVE	0.9215	0.9405	0.9425	0.9555	0.9465	0.9325	0.9425
	$\hat{v}_G^2$	0.9170	0.9460	0.9420	0.9505	0.9485	0.9495	0.9570
1.99	CR	1	1	1	1	1	1	1
	PCVE	0.9110	0.9440	0.9395	0.9520	0.9490	0.9510	0.9555
	$\hat{v}_G^2$	0.9220	0.9280	0.9295	0.9430	0.9440	0.9480	0.9390
3.31	CR	1	1	1	1	1	1	1
	PCVE	0.9150	0.9290	0.9325	0.9470	0.9435	0.9510	0.9405
	$\hat{v}_G^2$	0.9015	0.9260	0.9320	0.9505	0.9485	0.9405	0.9435
9.80	CR	1	1	1	1	1	1	1
	PCVE	0.8860	0.9225	0.9380	0.9495	0.9485	0.9420	0.9475
				Average Length				
	$\hat{v}_G^2$	1.20496	0.64428	0.39514	0.24765	0.19157	0.16045	0.14069
1.11	CR	3.21594	2.22170	1.62079	1.15081	0.94092	0.81621	0.73031
	PCVE	1.18192	0.63873	0.39376	0.24689	0.19111	0.16028	0.14062
	$\hat{v}_G^2$	1.16805	0.58866	0.34117	0.19821	0.14670	0.12020	0.10335
1.42	CR	3.23229	2.23499	1.63182	1.15901	0.94776	0.82214	0.73561
	PCVE	1.14574	0.58388	0.34065	0.19783	0.14622	0.12000	0.10327
	$\hat{v}_G^2$	1.18988	0.60685	0.34699	0.19474	0.14244	0.11466	0.09729
1.99	CR	3.25786	2.25761	1.65083	1.17312	0.95917	0.83201	0.74440
	PCVE	1.16373	0.59889	0.34582	0.19426	0.14229	0.11456	0.09728
	$\hat{v}_G^2$	1.27089	0.64963	0.37337	0.20857	0.15167	0.12110	0.10157
3.31	CR	3.29929	2.29885	1.68464	1.19841	0.98016	0.85013	0.76067
	PCVE	1.23316	0.64188	0.37129	0.20767	0.15108	0.12084	0.10134
	$\hat{v}_G^2$	1.41981	0.75053	0.43329	0.24285	0.17464	0.13851	0.11558
9.80	CR	3.38816	2.38329	1.75642	1.25248	1.02442	0.88868	0.79508
	PCVE	1.36449	0.73612	0.42992	0.24197	0.17401	0.13826	0.11549

[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 3: Model 2 - Matching on X_g[1]

N_{max}/N_{min}		$G=12$	$G=26$	$G=50$	$G=100$	$G=150$	$G=200$	$G=250$
				Coverage				
	$\hat{v}_G^2$	0.9260	0.9375	0.9420	0.9420	0.9460	0.9465	0.9510
1.11	CR	0.9570	0.9635	0.9755	0.9790	0.9825	0.9835	0.9800
	PCVE	0.9560	0.9645	0.9750	0.9785	0.9825	0.9835	0.9805
	$\hat{v}_G^2$	0.9280	0.9395	0.9455	0.9405	0.9490	0.9495	0.9490
1.42	CR	0.9525	0.9705	0.9705	0.9715	0.9795	0.9860	0.9820
	PCVE	0.9535	0.9710	0.9705	0.9735	0.9795	0.9860	0.9820
	$\hat{v}_G^2$	0.9180	0.9325	0.9385	0.9455	0.9480	0.9420	0.9465
1.99	CR	0.9415	0.9595	0.9680	0.9765	0.9770	0.9805	0.9800
	PCVE	0.9415	0.9605	0.9675	0.9770	0.9780	0.9800	0.9805
	$\hat{v}_G^2$	0.8965	0.9290	0.9390	0.9480	0.9440	0.9400	0.9495
3.31	CR	0.9325	0.9615	0.9700	0.9750	0.9775	0.9750	0.9765
	PCVE	0.9315	0.9615	0.9685	0.9755	0.9780	0.9745	0.9770
	$\hat{v}_G^2$	0.8850	0.9085	0.9295	0.9380	0.9360	0.9375	0.9445
9.80	CR	0.9155	0.9460	0.9640	0.9660	0.9660	0.9685	0.9755
	PCVE	0.9175	0.9450	0.9635	0.9660	0.9665	0.9680	0.9755
				Average Length				
	$\hat{v}_G^2$	1.64579	1.11414	0.80852	0.57317	0.46677	0.40525	0.36269
1.11	CR	1.88285	1.31397	0.96438	0.68747	0.56044	0.48713	0.43634
	PCVE	1.88367	1.31373	0.96432	0.68752	0.56044	0.48718	0.43636
	$\hat{v}_G^2$	1.67055	1.13171	0.81934	0.58015	0.47436	0.41154	0.36739
1.42	CR	1.90602	1.32885	0.97303	0.69262	0.56755	0.49258	0.44032
	PCVE	1.90579	1.32897	0.97283	0.69257	0.56751	0.49262	0.44026
	$\hat{v}_G^2$	1.67377	1.14094	0.83413	0.59068	0.48377	0.41909	0.37493
1.99	CR	1.90337	1.33455	0.98635	0.70162	0.57506	0.49879	0.44584
	PCVE	1.90395	1.33471	0.98606	0.70146	0.57506	0.49874	0.44586
	$\hat{v}_G^2$	1.69386	1.16940	0.85636	0.61062	0.49954	0.43424	0.38770
3.31	CR	1.91395	1.35515	1.00133	0.71846	0.58755	0.51145	0.45702
	PCVE	1.91241	1.35461	1.00137	0.71861	0.58755	0.51149	0.45699
	$\hat{v}_G^2$	1.74999	1.23124	0.90607	0.64424	0.52971	0.45990	0.41091
9.80	CR	1.95803	1.40591	1.04446	0.74668	0.61421	0.53318	0.47665
	PCVE	1.95767	1.40633	1.04420	0.74671	0.61422	0.53315	0.47665

[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 4: Model 2 - Matching on X_g and N_g[1]

N_{max}/N_{min}		$G=12$	$G=26$	$G=50$	$G=100$	$G=150$	$G=200$	$G=250$
				Coverage				
	$\hat{v}_G^2$	0.9420	0.9480	0.9545	0.9495	0.9455	0.9530	0.9530
1.11	CR	0.9670	0.9845	0.9875	0.9900	0.9915	0.9950	0.9935
	PCVE	0.9680	0.9850	0.9865	0.9900	0.9910	0.9950	0.9935
	$\hat{v}_G^2$	0.9315	0.9475	0.9515	0.9530	0.9515	0.9580	0.9510
1.42	CR	0.9665	0.9850	0.9850	0.9895	0.9915	0.9955	0.9955
	PCVE	0.9660	0.9850	0.9845	0.9900	0.9915	0.9960	0.9955
	$\hat{v}_G^2$	0.9270	0.9430	0.9510	0.9520	0.9480	0.9575	0.9520
1.99	CR	0.9650	0.9825	0.9885	0.9905	0.9930	0.9970	0.9945
	PCVE	0.9670	0.9815	0.9880	0.9900	0.9930	0.9970	0.9945
	$\hat{v}_G^2$	0.9160	0.9365	0.9525	0.9480	0.9510	0.9525	0.9485
3.31	CR	0.9580	0.9795	0.9890	0.9885	0.9930	0.9955	0.9940
	PCVE	0.9580	0.9800	0.9890	0.9890	0.9930	0.9955	0.9940
	$\hat{v}_G^2$	0.9065	0.9330	0.9430	0.9510	0.9515	0.9495	0.9510
9.80	CR	0.9410	0.9765	0.9845	0.9890	0.9880	0.9955	0.9915
	PCVE	0.9430	0.9755	0.9830	0.9890	0.9875	0.9955	0.9915
				Average Length				
	$\hat{v}_G^2$	1.57502	1.02869	0.73036	0.51031	0.41388	0.35765	0.31902
1.11	CR	1.89796	1.31976	0.96665	0.68810	0.56233	0.48793	0.43636
	PCVE	1.89800	1.31982	0.96657	0.68813	0.56236	0.48790	0.43634
	$\hat{v}_G^2$	1.58361	1.03237	0.73193	0.50975	0.41335	0.35758	0.31856
1.42	CR	1.91602	1.33100	0.97594	0.69418	0.56753	0.49302	0.44052
	PCVE	1.91549	1.33128	0.97597	0.69423	0.56756	0.49301	0.44049
	$\hat{v}_G^2$	1.61080	1.04567	0.74313	0.51722	0.41903	0.36217	0.32297
1.99	CR	1.93406	1.34395	0.98875	0.70392	0.57534	0.49967	0.44684
	PCVE	1.93403	1.34409	0.98881	0.70388	0.57529	0.49964	0.44680
	$\hat{v}_G^2$	1.63660	1.07550	0.76774	0.53170	0.43114	0.37227	0.33175
3.31	CR	1.94629	1.37114	1.01341	0.72038	0.58976	0.51183	0.45771
	PCVE	1.94802	1.37098	1.01337	0.72047	0.58984	0.51198	0.45771
	$\hat{v}_G^2$	1.70687	1.13039	0.80947	0.55966	0.45337	0.39151	0.34801
9.80	CR	1.98400	1.41410	1.05392	0.75111	0.61528	0.53484	0.47768
	PCVE	1.98403	1.41488	1.05356	0.75103	0.61532	0.53482	0.47769

[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 5: Model 1 - Randomization Test (RT) vs. CR/PCVE [1]

		Size under H_0			Power under $H_1 : \Delta_0 + 1/4$		
N_{max}/N_{min}		$G = 12$	$G = 26$	$G = 50$	$G = 12$	$G = 26$	$G = 50$
				Matching on X_g			
	RT	0.0395	0.0560	0.0505	0.0755	0.1220	0.2030
1.11	CR	0.0015	0.0010	0.0005	0.0095	0.0105	0.0160
	PCVE	0.0770	0.0690	0.0615	0.1195	0.1410	0.1995
	RT	0.0610	0.0445	0.0540	0.0935	0.1055	0.1970
1.42	CR	0.0020	0.0005	0.0015	0.0105	0.0105	0.0210
	PCVE	0.0965	0.0620	0.0625	0.1365	0.1220	0.1955
	RT	0.0505	0.0505	0.0505	0.0770	0.1130	0.1820
1.99	CR	0.0015	0.0015	0	0.0130	0.0100	0.0195
	PCVE	0.0905	0.0770	0.0580	0.1195	0.1260	0.1825
	RT	0.0570	0.0595	0.0555	0.0745	0.1130	0.1670
3.31	CR	0.0050	0.0020	0.0020	0.0145	0.0190	0.0270
	PCVE	0.1020	0.0845	0.0670	0.1220	0.1340	0.1760
	RT	0.0455	0.0500	0.0475	0.0715	0.1105	0.1410
9.80	CR	0.0075	0.0060	0.0030	0.0280	0.0230	0.0305
	PCVE	0.1075	0.0900	0.0635	0.1335	0.1380	0.1605
				Matching on X_g and N_g			
	RT	0.0490	0.0535	0.0585	0.1165	0.3050	0.6760
1.11	CR	0	0	0	0	0	0
	PCVE	0.0900	0.0740	0.0640	0.1540	0.2395	0.5015
	RT	0.0440	0.0475	0.0480	0.1290	0.3595	0.7820
1.42	CR	0	0	0	0	0	0
	PCVE	0.0785	0.0595	0.0575	0.1635	0.2810	0.5705
	RT	0.0510	0.0400	0.0480	0.1255	0.3380	0.7795
1.99	CR	0	0	0	0	0	0
	PCVE	0.0890	0.0560	0.0605	0.1580	0.2630	0.5785
	RT	0.0440	0.0500	0.0555	0.1185	0.3370	0.7075
3.31	CR	0	0	0	0	0	0
	PCVE	0.0850	0.0710	0.0675	0.1590	0.2825	0.5220
	RT	0.0525	0.0550	0.0500	0.1180	0.2780	0.5965
9.80	CR	0	0	0	0.0005	0	0
	PCVE	0.1140	0.0775	0.0620	0.1750	0.2540	0.4625

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

Table 6: Model 2 - Randomization Test (RT) vs. CR/PCVE[1]

		Size under H_0			Power under $H_1 : \Delta_0 + 1/4$		
N_{max}/N_{min}		$G = 12$	$G = 26$	$G = 50$	$G = 12$	$G = 26$	$G = 50$
Matching on X_g							
	RT	0.0345	0.0425	0.0480	0.0305	0.0790	0.1650
1.11	CR	0.0430	0.0365	0.0245	0.0540	0.0645	0.1120
	PCVE	0.0440	0.0355	0.0250	0.0550	0.0655	0.1115
	RT	0.0370	0.0365	0.0445	0.0370	0.0675	0.1685
1.42	CR	0.0475	0.0295	0.0295	0.0575	0.0560	0.1125
	PCVE	0.0465	0.0290	0.0295	0.0560	0.0540	0.1145
	RT	0.0465	0.0445	0.0490	0.0385	0.0785	0.1485
1.99	CR	0.0585	0.0405	0.0320	0.0620	0.0675	0.1005
	PCVE	0.0585	0.0395	0.0325	0.0615	0.0675	0.1005
	RT	0.0565	0.0495	0.0520	0.0390	0.0660	0.1360
3.31	CR	0.0675	0.0385	0.0300	0.0610	0.0620	0.1010
	PCVE	0.0685	0.0385	0.0315	0.0595	0.0625	0.1025
	RT	0.0700	0.0660	0.0600	0.0405	0.0550	0.1140
9.80	CR	0.0845	0.0540	0.0360	0.0585	0.0600	0.0895
	PCVE	0.0825	0.0550	0.0365	0.0595	0.0580	0.0895
Matching on X_g and N_g							
	RT	0.0250	0.0310	0.0370	0.0195	0.0735	0.1800
1.11	CR	0.0330	0.0155	0.0125	0.0240	0.0365	0.0765
	PCVE	0.0320	0.0150	0.0135	0.0235	0.0360	0.0790
	RT	0.0295	0.0290	0.0345	0.0205	0.0730	0.1740
1.42	CR	0.0335	0.0150	0.0150	0.0245	0.0385	0.0640
	PCVE	0.0340	0.0150	0.0155	0.0250	0.0365	0.0675
	RT	0.0345	0.0325	0.0415	0.0200	0.0665	0.1655
1.99	CR	0.0350	0.0175	0.0115	0.0225	0.0310	0.0600
	PCVE	0.0330	0.0185	0.0120	0.0230	0.0320	0.0610
	RT	0.0390	0.0390	0.0340	0.0150	0.0590	0.1415
3.31	CR	0.0420	0.0205	0.0110	0.0220	0.0295	0.0610
	PCVE	0.0420	0.0200	0.0110	0.0210	0.0310	0.0595
	RT	0.0555	0.0445	0.0415	0.0260	0.0405	0.1180
9.80	CR	0.0590	0.0235	0.0155	0.0295	0.0270	0.0505
	PCVE	0.0570	0.0245	0.0170	0.0295	0.0265	0.0510

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

A Proofs of Main Results

A.1 Proof of Proposition 3.1

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. ■

A.2 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 the Jacobian is

$$D_h(x,y,z,w) = \left(\frac{1}{y}, -\frac{x}{y^2}, -\frac{1}{w}, \frac{z}{w^2}\right).$$

By Assumption 3.1,

$$\sqrt{G}\Big(\frac{1}{G}\sum_{1\le g\le 2G}\bar{Y}_gN_gD_g - E[\bar{Y}_g(1)N_g]\Big) = \frac{1}{\sqrt{G}}\sum_{1\le g\le 2G}(\bar{Y}_g(1)N_gD_g - E[\bar{Y}_g(1)N_g]D_g)$$

and similarly for the other three terms. The desired conclusion then follows from Lemma A.1 together with an application of the delta method. To see this, note by the laws of total variance and total covariance that $\mathbb{V}$ in Lemma A.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

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

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

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

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

■

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

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

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

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

$$
\begin{aligned}
\mathbb{V}_1^1 &= \begin{pmatrix} E[\mathrm{Var}[\bar{Y}_g(1)N_g|X_g]] & E[\mathrm{Cov}[\bar{Y}_g(1)N_g, N_g|X_g]] \\ E[\mathrm{Cov}[\bar{Y}_g(1)N_g, N_g|X_g]] & E[\mathrm{Var}[N_g|X_g]] \end{pmatrix} \\
\mathbb{V}_1^0 &= \begin{pmatrix} E[\mathrm{Var}[\bar{Y}_g(0)N_g|X_g]] & E[\mathrm{Cov}[\bar{Y}_g(0)N_g, N_g|X_g]] \\ E[\mathrm{Cov}[\bar{Y}_g(0)N_g, N_g|X_g]] & E[\mathrm{Var}[N_g|X_g]] \end{pmatrix} \\
\mathbb{V}_2 &= \frac{1}{2}\mathrm{Var}[(E[\bar{Y}_g(1)N_g|X_g], E[N_g|X_g], E[\bar{Y}_g(0)N_g|X_g], E[N_g|X_g])'] \ .
\end{aligned}
$$

Proof of Lemma A.1. 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 $(\mathbb{L}^{\mathrm{YN1}}_{1,G}, \mathbb{L}^{\mathrm{N1}}_{1,G}, \mathbb{L}^{\mathrm{YN0}}_{1,G}, \mathbb{L}^{\mathrm{N0}}_{1,G}), G \geq 1$ is a triangular array of normalized sums of random vectors. Conditional on $X^{(G)}, D^{(G)}$, $(\mathbb{L}^{\mathrm{YN1}}_{1,G}, \mathbb{L}^{\mathrm{N1}}_{1,G}) \perp\!\!\!\perp (\mathbb{L}^{\mathrm{YN0}}_{1,G}, \mathbb{L}^{\mathrm{N0}}_{1,G})$. Moreover, it follows from $Q_G = Q^{2G}$ and Assumption 3.1 that

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

For the upper left component, we have

$$\frac{1}{G}\sum_{1\leq g\leq 2G}\mathrm{Var}[\bar{Y}_g(1)N_g|X_g]D_g = \frac{1}{G}\sum_{1\leq g\leq 2G}E[\bar{Y}^2_g(1)N^2_g|X_g]D_g - \frac{1}{G}\sum_{1\leq g\leq 2G}E[\bar{Y}_g(1)N_g|X_g]^2D_g . \tag{8}$$

Note

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

It follows from the weak law of large numbers, the application of which is permitted by Lemma B.1, that

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

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

$$\begin{aligned}&\Big|\frac{1}{G}\sum_{1\leq g\leq 2G:D_g=1}E[\bar{Y}^2_g(1)N^2_g|X_g] - \frac{1}{G}\sum_{1\leq g\leq 2G:D_g=0}E[\bar{Y}^2_g(1)N^2_g|X_g]\Big|\\ &\leq \frac{1}{G}\sum_{1\leq j\leq G}|E[\bar{Y}^2_{\pi(2j-1)}(1)N^2_{\pi(2j-1)}|X_{\pi(2j-1)}] - E[\bar{Y}^2_{\pi(2j)}(1)N^2_{\pi(2j)}|X_{\pi(2j)}]|\\ &\lesssim \frac{1}{G}\sum_{1\leq j\leq G}|X_{\pi(2j-1)} - X_{\pi(2j)}| \xrightarrow{P} 0 .\end{aligned}$$

Therefore,

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

Meanwhile,

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

It follows from the weak law of large numbers (the application of which is permitted by Lemma B.1) that

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

Next,

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

$$\leq \Big(\frac{1}{G}\sum_{1\leq j\leq G}|X_{\pi(2j-1)}-X_{\pi(2j)}|^2\Big)^{1/2}\Big(\frac{1}{G}\sum_{1\leq g\leq 2G}E[\bar{Y}_g(1)N_g|X_g]^2\Big)^{1/2}\xrightarrow{P}0\ ,$$

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

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

and hence it follows from (8) that

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

An identical argument establishes that

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

To study the off-diagonal components, note that

$$\frac{1}{G}\sum_{1\leq g\leq 2G}\mathrm{Cov}[\bar{Y}_g(1)N_g,N_g|X_g]D_g = \frac{1}{G}\sum_{1\leq g\leq 2G}E[\bar{Y}_g(1)N_g^2|X_g]D_g - \frac{1}{G}\sum_{1\leq g\leq 2G}E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g]D_g\ . \tag{9}$$

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,

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

Note that

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

where the equality follows by the law of iterated expectations and the inequality by Lemma B.1 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

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

and for each j,

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

where the final inequality follows from the triangle inequality, Assumption 3.3(b) and Lemma B.1.

Thus we have that

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

$$\lesssim \frac{1}{G}\sum_{1\le j\le G}\big|E[\bar{Y}_{\pi(2j-1)}(1)N_{\pi(2j-1)}|X_{\pi(2j-1)}] - E[\bar{Y}_{\pi(2j)}(1)N_{\pi(2j)}|X_{\pi(2j)}]\big| + \big|E[N_{\pi(2j-1)}|X_{\pi(2j-1)}] - E[N_{\pi(2j)}|X_{\pi(2j)}]\big|$$

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

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

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

Thus we have established that

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

Similarly,

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

It thus follows from similar arguments to those used in Lemma A.2 that

$$\rho(\mathcal{L}((\mathbb{L}^{\text{YN1}}_{1,G}, \mathbb{L}^{\text{N1}}_{1,G}, \mathbb{L}^{\text{YN0}}_{1,G}, \mathbb{L}^{\text{N0}}_{1,G})'|X^{(G)}, D^{(G)}), N(0,\mathbb{V}_1)) \xrightarrow{P} 0\ , \tag{10}$$

where $\mathcal{L}(\cdot)$ denotes the law of a random variable and ρ is any metric that metrizes weak convergence.

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

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

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

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

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

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

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

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

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 B.1) that

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

Because (10) holds and $(\mathbb{L}_{2,G}^{\text{YN1}}, \mathbb{L}_{2,G}^{\text{N1}}, \mathbb{L}_{2,G}^{\text{YN0}}, \mathbb{L}_{2,G}^{\text{N0}})$ is deterministic conditional on $X^{(G)}, D^{(G)}$, the conclusion of the theorem follows from Lemma S.1.3 in Bai et al. (2022). ■

A.3 Proof of Theorem 3.2

Proof. We have that

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

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

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

and the Jacobian is

$$D_h(x,y,z,w) = \left(\frac{1}{y}, -\frac{x}{y^2}, -\frac{1}{w}, \frac{z}{w^2}\right) .$$

By Assumption 3.4,

$$\sqrt{G}\Big(\frac{1}{G}\sum_{1\le g\le 2G}\bar{Y}_g N_g D_g - E[\bar{Y}_g(1)N_g]\Big) = \frac{1}{\sqrt{G}}\sum_{1\le g\le 2G}(\bar{Y}_g(1)N_g D_g - E[\bar{Y}_g(1)N_g]D_g)$$

and similarly for the other three terms. The desired conclusion then follows from Lemma A.2 together with an application of the Delta method. To see this, note by the laws of total variance and total covariance that $\mathbb{V}$ in Lemma A.2 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 Theorem 3.1. Expanding and simplifying the first half of the expression:

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

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

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

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

■

Lemma A.2. *Suppose Q satisfies Assumptions 2.1 and 3.6 and the treatment assignment mechanism satisfies Assumptions 3.4–3.5. Define*

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

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

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

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

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

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

Proof of Lemma A.2. 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

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

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

$$\frac{1}{G}\sum_{1\le g\le 2G}\mathrm{Var}[\bar{Y}_g(1)N_g|W_g]D_g = \frac{1}{G}\sum_{1\le g\le 2G}E[\bar{Y}_g^2(1)N_g^2|W_g]D_g - \frac{1}{G}\sum_{1\le g\le 2G}E[\bar{Y}_g(1)N_g|W_g]^2D_g\ . \tag{11}$$

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 B.1,

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

On the other hand,

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

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 B.1, and the convergence in probability follows from Assumption 3.5 and Lemma B.2. 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,

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

It follows from the weak law of large numbers, the application of which is permitted by Lemma B.1 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,

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

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 3.6, 3.5 and the 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 (11) that

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

Similarly,

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

We now establish

$$\rho(\mathcal{L}((\mathbb{L}_{1,G}^{\mathrm{YN1}},0,\mathbb{L}_{1,G}^{\mathrm{YN0}},0)|W^{(G)},D^{(G)}),N(0,\mathbb{V}_1))\xrightarrow{P}0\ , \tag{12}$$

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

$$\rho(\mathcal{L}((\mathbb{L}_{1,G_{k_l}}^{\mathrm{YN1}},0,\mathbb{L}_{1,G_{k_l}}^{\mathrm{YN0}},0)|W^{(G_{k_l})},D^{(G^{k_l})},N(0,\mathbb{V}_1))\to 0 \text{ with probability one}\ . \tag{13}$$

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). First note that by the results proved so far,

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

Next, We will use the inequality

$$\left|\sum_{1\le j\le k}a_j\right|I\left\{\left|\sum_{1\le j\le k}a_j\right|>\epsilon\right\}\le\sum_{1\le j\le k}k|a_j|I\left\{|a_j|>\frac{\epsilon}{k}\right\}\ . \tag{14}$$

It follows from (14) that

$$\begin{aligned}
&\frac{1}{G}\sum_{1\le g\le 2G}E[(D_g(\bar{Y}_g(1)N_g-E[\bar{Y}_g(1)N_g|W_g]))^2+((1-D_g)(\bar{Y}_g(0)N_g-E[\bar{Y}_g(0)N_g|W_g]))^2\\
&\qquad\times I\{(D_g(\bar{Y}_g(1)N_g-E[\bar{Y}_g(1)N_g|W_g]))^2+((1-D_g)(\bar{Y}_g(0)N_g-E[\bar{Y}_g(0)N_g|W_g]))^2>\epsilon^2G\}|W^{(G)},D^{(G)}]\\
&\lesssim\frac{1}{G}\sum_{1\le g\le 2G}E[D_g(\bar{Y}_g(1)N_g-E[\bar{Y}_g(1)N_g|W_g])^2I\{D_g(\bar{Y}_g(1)N_g-E[\bar{Y}_g(1)N_g|W_g])^2>\epsilon^2G/2\}|W^{(G)},D^{(G)}]\\
&\qquad+\frac{1}{G}\sum_{1\le g\le 2G}E[(1-D_g)(\bar{Y}_g(0)N_g-E[\bar{Y}_g(0)N_g|W_g])^2I\{(1-D_g)(\bar{Y}_g(0)N_g-E[\bar{Y}_g(0)N_g|W_g])^2>\epsilon^2G/2\}|W^{(G)},D^{(G)}]\\
&\le\frac{1}{G}\sum_{1\le g\le 2G}E[(\bar{Y}_g(1)N_g-E[\bar{Y}_g(1)N_g|W_g])^2I\{|\bar{Y}_g(1)N_g-E[\bar{Y}_g(1)N_g|W_g]|>\epsilon\sqrt{G}/\sqrt{2}\}|W_g]\\
&\qquad+\frac{1}{G}\sum_{1\le g\le 2G}E[(\bar{Y}_g(0)N_g-E[\bar{Y}_g(0)N_g|W_g])^2I\{|\bar{Y}_g(0)N_g-E[\bar{Y}_g(0)N_g|W_g]|>\epsilon\sqrt{G}/\sqrt{2}\}|W_g]\ .
\end{aligned}$$

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

$$\begin{aligned}
&\le\frac{1}{G}\sum_{1\le g\le 2G}E[(\bar{Y}_g(1)N_g-E[\bar{Y}_g(1)N_g|W_g])^2I\{|\bar{Y}_g(1)N_g-E[\bar{Y}_g(1)N_g|W_g]|>m\}|W_g]\\
&\qquad+\frac{1}{G}\sum_{1\le g\le 2G}E[(\bar{Y}_g(0)N_g-E[\bar{Y}_g(0)N_g|W_g])^2I\{|\bar{Y}_g(0)N_g-E[\bar{Y}_g(0)N_g|W_g]|>m|W_g]\\
&\xrightarrow{P}2E[(\bar{Y}_g(1)N_g-E[\bar{Y}_g(1)N_g|W_g])^2I\{|\bar{Y}_g(1)N_g-E[\bar{Y}_g(1)N_g|W_g]|>m\}]\\
&\qquad+E[(\bar{Y}_g(1)N_g-E[\bar{Y}_g(1)N_g|W_g])^2I\{|\bar{Y}_g(1)N_g-E[\bar{Y}_g(1)N_g|W_g]|>m\}]\ .
\end{aligned}$$

because $E[(\bar{Y}_g(1)N_g-E[\bar{Y}_g(1)N_g|W_g])^2]<\infty$ and $E[(\bar{Y}_g(0)N_g-E[\bar{Y}_g(0)N_g|W_g])^2]<\infty$. As $m\to\infty$, the last expression goes to 0. Therefore, it follows from similar arguments to those 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 (13) and hence (12) holds.

Next, we study $(\mathbb{L}_{2,G}^{\mathrm{YN1}}, \mathbb{L}_{G}^{\mathrm{N1}}, \mathbb{L}_{2,G}^{\mathrm{YN0}}, \mathbb{L}_{G}^{\mathrm{N0}})$. It follows from $Q_G = Q^{2G}$ and Assumption 3.4 that

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

For $\mathbb{L}_{2,G}^{\mathrm{YN1}}$, it follows from similar arguments to those used above that $\mathrm{Var}[\mathbb{L}_{2,G}^{\mathrm{YN1}} | W^{(G)}] \xrightarrow{P} 0$. Therefore, it follows from Markov's inequality conditional on $W^{(G)}$ and $D^{(G)}$, and the fact that probabilities are bounded and hence uniformly integrable, that

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

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

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

It thus follows from the central limit theorem (the application of which is justified by Assumption 2.1(c) and Lemma B.1) that

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

Because (10) holds and $(\mathbb{L}_{2,G}^{\mathrm{YN1}}, \mathbb{L}_{G}^{\mathrm{N1}}, \mathbb{L}_{2,G}^{\mathrm{YN0}}, \mathbb{L}_{G}^{\mathrm{N0}})$ is deterministic conditional on $N^{(G)}, X^{(G)}, D^{(G)}$, the conclusion of the theorem follows from Lemma S.1.3 in Bai et al. (2022). ■

A.4 Proof of Theorem 3.3

The desired conclusion follows immediately from Lemmas B.4-B.6.

A.5 Proof of Theorem 3.4

By the derivation in Theorem 3.6 in Bugni et al. (2022),

$$\hat{\omega}_{\mathrm{CR,G}}^2 = \frac{1}{2} \left(\hat{\omega}_{\mathrm{CR,G}}^2(1) + \hat{\omega}_{\mathrm{CR,G}}^2(0) \right) , \tag{15}$$

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

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

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 Lemma S.1.5 in Bai et al. (2022) applied to the observations $(N_g^\ell \bar{Y}_g^r(d) : 1 \le g \le 2G)$,

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

The result then follows by an identical derivation to that of Theorem 3.6 in Bugni et al. (2022).

A.6 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*:

$$\left(\frac{1}{G}\sum_{1\le j\le G}\sum_{g\in\lambda_j} X_g' X_g\right)^{-1}\left(\frac{1}{G}\sum_{1\le j\le G}\left(\sum_{g\in\lambda_j} X_g'\hat{\epsilon}_g\right)\left(\sum_{g\in\lambda_j} X_g'\hat{\epsilon}_g\right)'\right)\left(\frac{1}{G}\sum_{1\le g\le G}\sum_{g\in\lambda_j} X_g' X_g\right)^{-1}, \tag{16}$$

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

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

Imposing the condition that $N_g = k$ 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. (2023), 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 Lemmas S.1.5 and S.1.6 of Bai et al. (2022) applied to the observations $(\bar{Y}_g(d) : 1\le g\le 2G)$, and the continuous mapping theorem, we thus obtain that

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

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 $N_g = k$, we then obtain

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

The conclusion then follows.

A.7 Proof of Theorem 3.6

Proof. Note that the null hypothesis (6) combined with Assumption 2.1(e) implies that

$$\bar{Y}_g(1)|(X_g,N_g) \stackrel{d}{=} \bar{Y}_g(0)|(X_g,N_g) \ . \tag{17}$$

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 (17) is in fact equivalent to the statement

$$(\bar{Y}_g(1),N_g)|X_g \stackrel{d}{=} (\bar{Y}_g(0),N_g)|X_g \ . \tag{18}$$

The result then follows by applying Theorem 3.4 in Bai et al. (2022) using (18) as the null hypothesis. To establish this equivalence, we first begin with (17) 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

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

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

where the first and second equalities follow from the definition of conditional expectations, the the third follows from (17), and the last follows again from the definition of a conditional expectation. The opposite implication follows from a similar argument and is thus omitted. ■

A.8 Proof of Theorem 3.7

The desired conclusion follows from Lemmas B.7 and B.9, along with a straightforward modification of Lemma A.3. in Chung and Romano (2013).

B Auxiliary Lemmas

Lemma B.1. *If Assumption 2.1 holds,*

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

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,

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

and hence

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

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 B.2. *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.

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

where the inequality follows by Cauchy-Schwarz. It follows from an argument similar to the proof of Proposition 3.1 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 B.3. *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. $((\tilde{Y}_g(1), \tilde{Y}_g(0)) : 1 \leq g \leq 2G) \perp D^{(G)} \mid X^{(G)}$ *(or* $((\tilde{Y}_g(1), \tilde{Y}_g(0)) : 1 \leq g \leq 2G) \perp D^{(G)} \mid W^{(G)}$*)*
3. $\frac{1}{G}\sum_{j=1}^{G}\left|\mu_d(X_{\pi(2j)}) - \mu_d(X_{\pi(2j-1)})\right| \xrightarrow{P} 0$, *where we use* $\mu_d(X_g)$ *to denote* $E[\tilde{Y}_g(d) \mid X_g]$ *for* $d \in \{0,1\}$.
   (or $\frac{1}{G}\sum_{j=1}^{G}\left|\mu_d(W_{\pi(2j)}) - \mu_d(W_{\pi(2j-1)})\right| \xrightarrow{P} 0$*)*
4. $\frac{1}{G}\sum_{j=1}^{G}\left|\left(\mu_1(X_{\pi(2j)}) - \mu_1(X_{\pi(2j-1)})\right)\left(\mu_0(X_{\pi(2j)}) - \mu_0(X_{\pi(2j-1)})\right)\right| \xrightarrow{P} 0$.
   (or $\frac{1}{G}\sum_{j=1}^{G}\left|\left(\mu_1(W_{\pi(2j)}) - \mu_1(W_{\pi(2j-1)})\right)\left(\mu_0(W_{\pi(2j)}) - \mu_0(W_{\pi(2j-1)})\right)\right| \xrightarrow{P} 0$*)*
5. $\frac{1}{4G}\sum_{k\in\{2,3\},\ell\in\{0,1\}}\sum_{1\leq j\leq \frac{G}{2}}\left(\mu_d\left(X_{\pi(4j-\ell)}\right) - \mu_d\left(X_{\pi(4j-k)}\right)\right)^2 \xrightarrow{P} 0$.
   (or $\frac{1}{4G}\sum_{k\in\{2,3\},\ell\in\{0,1\}}\sum_{1\leq j\leq \frac{G}{2}}\left(\mu_d\left(W_{\pi(4j-\ell)}\right) - \mu_d\left(W_{\pi(4j-k)}\right)\right)^2 \xrightarrow{P} 0$*)*

Proof. Note that

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

where the inequality follows by Lemma B.1. The second result follows directly by inspection and Assumption 3.4 (or Assumption 3.1). In terms of the third result, by Assumption 3.3 and 3.2,

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

Meanwhile,

$$\frac{1}{G}\sum_{j=1}^{G}\left|\mu_1(W_{\pi(2j)}) - \mu_1(W_{\pi(2j-1)})\right| \lesssim \frac{1}{G}\sum_{j=1}^{G}\left|E[N_{\pi(2j)}\bar{Y}_{\pi(2j)}(d) \mid W_{\pi(2j)}] - E[N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)}(d) \mid W_{\pi(2j-1)}]\right|$$
$$+ \frac{1}{G}\sum_{j=1}^{G}\left|E[N_{\pi(2j)} \mid W_{\pi(2j)}] - E[N_{\pi(2j-1)} \mid W_{\pi(2j-1)}]\right|$$
$$\lesssim \frac{1}{G}\sum_{j=1}^{G}\left|N_{\pi(2j)}\left(E[\bar{Y}_{\pi(2j)}(d) \mid W_{\pi(2j)}] - E[\bar{Y}_{\pi(2j-1)}(d) \mid W_{\pi(2j-1)}]\right)\right| + \frac{1}{G}\sum_{j=1}^{G}\left|N_{\pi(2j)} - N_{\pi(2j-1)}\right|$$
$$+ \frac{1}{G}\sum_{j=1}^{G}\left|(N_{\pi(2j)} - N_{\pi(2j-1)})E[\bar{Y}_{\pi(2j-1)}(d) \mid W_{\pi(2j-1)}]\right|$$
$$\lesssim \frac{1}{G}\sum_{j=1}^{G}N_{\pi(2j)}\left|W_{\pi(2j)} - W_{\pi(2j-1)}\right| ,$$

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

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

Similarly,

$$\frac{1}{G}\sum_{j=1}^{G}\left|\left(\mu_1(W_{\pi(2j)})-\mu_1(W_{\pi(2j-1)})\right)\left(\mu_0(W_{\pi(2j)})-\mu_0(W_{\pi(2j-1)})\right)\right|$$
$$\leq \frac{1}{G}\sum_{j=1}^{G}\left|\mu_1(W_{\pi(2j)})-\mu_1(W_{\pi(2j-1)})\right|\left|\mu_0(W_{\pi(2j)})-\mu_0(W_{\pi(2j-1)})\right|$$
$$\lesssim \frac{1}{G}\sum_{j=1}^{G} N_{\pi(2j)}^2\left|W_{\pi(2j)}-W_{\pi(2j-1)}\right|^2 \xrightarrow{P} 0 ,$$

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

Lemma B.4. *Consider the following adjusted potential outcomes:*

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

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*

$$\hat{\mu}_G(d)=\frac{1}{G}\sum_{1\leq g\leq 2G}\hat{Y}_g(d)I\{D_g=d\}\xrightarrow{P}0$$
$$\hat{\sigma}_G^2(d)=\frac{1}{G}\sum_{1<g<2G}\left(\hat{Y}_g-\hat{\mu}_G(d)\right)^2 I\{D_g=d\}\xrightarrow{P}\text{Var}\left[\tilde{Y}_g(d)\right] .$$

Proof. It suffices to show that

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

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

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

Note that

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

By weak law of large number, Lemma A.2 (or Lemma A.1) and Slutsky's theorem, we have

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

By applying Lemma S.1.5 from Bai et al. (2022) and Lemma B.3, we have

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

Thus, the result follows. ■

Lemma B.5. *If Assumptions 2.1 holds, and Assumptions 3.2-3.3 hold, then*

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

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

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

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 B.4 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 $E[\tilde{Y}_g(d) \mid W_g]$. To this end, first note that

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

Note that

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

for which the first term is given as follows:

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

By following the same argument in Lemma S.1.6 from Bai et al. (2022) and Lemma B.3, we have

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

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

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

Then, by weak law of large number, Lemma A.2 (or Lemma A.1) and Slutsky's 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 can be proved by applying Lemma S.1.6 from [BAI-Inference] and Lemma B.3. Hence, in the case where we match on cluster size,

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

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

Lemma B.6. *If Assumptions 2.1 holds, and Assumptions 2.1 and 3.2-3.3, 3.7 hold, then*

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

in the case where we match on cluster size. Instead, if Assumptions 3.5-3.6, 3.8 hold, then

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

in the case where we do not 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

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

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

Then we can show that each term converges to zero in probability by repeating the arguments in Lemma B.5. The results should hold for any treatment combination, 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 B.3, we have

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

in the case where we match on cluster size, and

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

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

Lemma B.7. *Let $\tilde{R}_g(t)$ denote the randomization distribution of $\sqrt{G}\hat{\Delta}_G$ (see equation (19)). Then under the null hypothesis (7), 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,

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

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

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

with (in both cases)

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

Proof. 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

$$
\tilde{R}_G(t) := P\Bigg\{\frac{1}{\tilde{N}(1)/G}\frac{1}{\sqrt{G}}\sum_{1\le j\le G}\epsilon_j\left(N_{\pi(2j)}\bar{Y}_{\pi(2j)} - N_{\pi(2j-1)}\bar{Y}_{\pi(2j-1)}\right)\left(D_{\pi(2j)} - D_{\pi(2j-1)}\right)
$$

$$- \frac{\frac{1}{\sqrt{G}} \sum_{1 \leq j \leq G} \epsilon_j (N_{\pi(2j)} - N_{\pi(2j-1)})(D_{\pi(2j)} - D_{\pi(2j-1)})}{\frac{\tilde{N}(1)}{G} \frac{\tilde{N}(0)}{G}} \frac{1}{G} \sum_{1 \leq g \leq 2G} (1 - \tilde{D}_g) N_g \bar{Y}_g \leq t \Bigg| Z^{(G)} \Bigg\} , \quad (19)$$

where ϵ_j j = 1,...,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

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

For a random transformation of the data, it follows as a consequence of Lemmas A.1 and A.2 that

$$\frac{1}{G} \sum_{1 \leq g \leq 2G} I\{\tilde{D}_g = d\} N_g \xrightarrow{P} E[N_g] ,$$

$$\frac{1}{G} \sum_{1 \leq g \leq 2G} (1 - \tilde{D}_g) N_g \bar{Y}_g \xrightarrow{p} E[N_g \bar{Y}_g(0)] .$$

Combining this with Lemma B.8 and a straightforward modification of Lemma A.3. in Chung and Romano (2013), 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

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

and when we do *not* match on cluster size

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

The result then follows from further algebraic manipulations to simplify τ in each case (see for instance Lemma B.10). ■

Lemma B.8.

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

where

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

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

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

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

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

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

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

$$\tilde{\mathbb{K}}_G^{N} := \frac{1}{\sqrt{G}} \sum_{1 \le j \le G} \epsilon_j \left(N_{\pi(2j)} - N_{\pi(2j-1)} \right) .$$

We will show

$$\rho \left(\mathcal{L} \left((\tilde{\mathbb{K}}_G^{YN}, \tilde{\mathbb{K}}_G^{N})' | Z^{(G)} \right), N(0, \mathbb{V}_R) \right) \xrightarrow{P} 0 , \tag{20}$$

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 (20), note we need only show that for any subsequence $\{G_k\}$ there exists a further subsequence $\{G_{k_l}\}$ such that

$$\rho \left(\mathcal{L} \left((\tilde{\mathbb{K}}_{G_{k_l}}^{YN}, \tilde{\mathbb{K}}_{G_{k_l}}^{N})' | Z^{(G_{k_l})} \right), N(0, \mathbb{V}_R) \right) \to 0 \text{ with probability one} . \tag{21}$$

To that end, define

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

where

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

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

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

First consider the case where we match on cluster size. By arguing as in Lemma S.1.6 of Bai et al. (2022), it can be shown that

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

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

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

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 B.1 and the law of large numbers. We can thus conclude that $\mathbb{V}_{R,n}^{1,2} = o_P(1)$ when matching on cluster size.

$$\mathbb{V}_{R,n} \xrightarrow{P} \mathbb{V}_R . \tag{22}$$

In the case where we do *not* match on cluster size, again by arguing as in Lemma S.1.6 of Bai et al. (2022), it can be shown that (22) holds. Next, we verify the Lindeberg condition in Proposition 2.27 of van der Vaart (1998). Note that

$$\frac{1}{G} \sum_{1 \le j \le G} E[((\epsilon_j (N_{\pi(2j)} \bar{Y}_{\pi(2j)} - N_{\pi(2j-1)} \bar{Y}_{\pi(2j-1)}))^2 + (\epsilon_j (N_{\pi(2j)} - N_{\pi(2j-1)}))^2)$$
$$\times I\{((\epsilon_j (N_{\pi(2j)} \bar{Y}_{\pi(2j)} - N_{\pi(2j-1)} \bar{Y}_{\pi(2j-1)}))^2 + (\epsilon_j (N_{\pi(2j)} - N_{\pi(2j-1)}))^2) > \epsilon^2 G\} | Z^{(G)}]$$

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

where the inequality follows from (14) 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

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

Note

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

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 B.1. 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 (21) holds, and the conclusion of the lemma follows. ■

Lemma B.9. *Let*

$$\breve{\nu}_G^2(\epsilon_1, \dots, \epsilon_G) = \hat{\tau}_G^2 - \frac{1}{2} \check{\lambda}_G^2 (\epsilon_1, \dots, \epsilon_G) \tag{23}$$

where $\hat{\tau}_G^2$ is defined in (3) and $\check{\Delta}_G(\epsilon_1, \dots, \epsilon_G)$ is defined in (19), and, independent of Z^G, ϵ_j, $j = 1, \dots, G$ are i.i.d. Rademacher random variables. If Assumptions 2.1 holds, and Assumptions 3.6-3.5 (or Assumptions 3.3-3.2) hold,

$$\breve{\nu}_G^2(\epsilon_1, \dots, \epsilon_G) \xrightarrow{P} \tau^2 \ ,$$

where τ^2 is defined in (B.7).

Proof. From Lemma B.5, we see that $\hat{\tau}_G^2 \xrightarrow{P} \tau^2$. It therefore suffices to show that $\check{\lambda}_G^2(\epsilon_1, \dots, \epsilon_G) \xrightarrow{P} 0$. In order to do so, note that $\check{\lambda}_G^2(\epsilon_1, \dots, \epsilon_G)$ 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

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

By following the arguments in Lemma S.1.9 of Bai et al. (2022) and Lemma B.3, 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:

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

by following the same argument in Lemma S.1.6 from Bai et al. (2022) and Lemma B.3, 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 weak law of large number, Lemma A.2 (or Lemma A.1) and Slutsky's 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 B.10. *If $E[N_g\bar{Y}_g(1)] = E[N_g\bar{Y}_g(0)]$, then for τ defined in Lemma B.7 (when not matching on cluster size),*

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

Proof. Note if $E[N_g\bar{Y}_g(1)] = E[N_g\bar{Y}_g(0)]$, then

$$E[\mathrm{Var}[\tilde{Y}_g(1)|X_g]] + E[\mathrm{Var}[\tilde{Y}_g(0)|X_g]] + E[(E[\tilde{Y}_g(1)|X_g] - E[\tilde{Y}_g(0)|X_g])^2]$$

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

The result then follows immediately. ■

References

Athey, S. and Imbens, G. W. (2017). The Econometrics of Randomized Experiments. In *Handbook of Economic Field Experiments*, vol. 1. Elsevier, 73–140.

Bai, Y. (2022). Optimality of Matched-Pair Designs in Randomized Controlled Trials. *American Economic Review*, **112** 3911–3940.

Bai, Y., Liu, J. and Tabord-Meehan, M. (2023). Inference for Matched Tuples and Fully Blocked Factorial Designs. ArXiv:2206.04157 [econ, math, stat], URL http://arxiv.org/abs/2206.04157.

Bai, Y., Romano, J. P. and Shaikh, A. M. (2022). Inference in Experiments With Matched Pairs. *Journal of the American Statistical Association*, **117** 1726–1737.

Banerjee, A., Duflo, E., Glennerster, R. and Kinnan, C. (2015). The miracle of microfinance? evidence from a randomized evaluation. *American economic journal: Applied economics*, **7** 22–53.

Bruhn, M. and McKenzie, D. (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics*, **1** 200–232.

Bugni, F., Canay, I., Shaikh, A. and Tabord-Meehan, M. (2022). Inference for Cluster Randomized Experiments with Non-ignorable Cluster Sizes. ArXiv:2204.08356 [econ, stat], URL http://arxiv.org/abs/2204.08356.

Bugni, F. A., Canay, I. A. and Shaikh, A. M. (2018). Inference Under Covariate-Adaptive Randomization. *Journal of the American Statistical Association*, **113** 1784–1796.

Chung, E. and Romano, J. P. (2013). Exact and asymptotically robust permutation tests. *Annals of Statistics*, **41** 484–507.

Crépon, B., Devoto, F., Duflo, E. and Parienté, W. (2015). Estimating the Impact of Microcredit on Those Who Take It Up: Evidence from a Randomized Experiment in Morocco. *American Economic Journal: Applied Economics*, **7** 123–150.

Cytrynbaum, M. (2021). Designing representative and balanced experiments by local randomization. *arXiv preprint arXiv:2111.08157*.

de Chaisemartin, C. and Ramirez-Cuellar, J. (2019). At what level should one cluster standard errors in paired and small-strata experiments? *arXiv preprint arXiv:1906.00288*.

Glennerster, R. and Takavarasha, K. (2013). *Running Randomized Evaluations: A Practical Guide*. Princeton University Press.

Hayes, R. J. and Moulton, L. H. (2017). *Cluster randomised trials*. Chapman and Hall/CRC.

Imai, K., King, G. and Nall, C. (2009). The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation. *Statistical Science*, **24** 29–53.

Janssen, A. (1997). Studentized permutation tests for non-i.i.d. hypotheses and the generalized Behrens-Fisher problem. *Statistics & Probability Letters*, **36** 9–21.

Jiang, L., Liu, X., Phillips, P. C. and Zhang, Y. (2020). Bootstrap inference for quantile treatment effects in randomized experiments with matched pairs. *The Review of Economics and Statistics* 1–47.

Su, F. and Ding, P. (2021). Model-assisted analyses of cluster-randomized experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

van der Vaart, A. W. (1998). *Asymptotic statistics*, vol. 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.