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

May 10, 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 single variance estimator which is consistent in either regime. 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. We additionally 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. Finally, we propose a covariate-adjusted estimator which adjusts for additional baseline covariates not used for treatment assignment, and establish conditions under which such an estimator leads to improvements in precision. 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 would like to thank seminar and conference participants at Indiana University and CIREQ for helpful comments on this paper. 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 in addition to baseline covariates delivers a weak (and often strict) improvement in asymptotic efficiency relative to matching on baseline covariates alone. 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 potential outcome distributions under treatment and control, and then establish asymptotic validity for testing null hypotheses about the size-weighted average treatment effect. We emphasize, however, that the latter 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.

Finally, we derive large-sample results for a covariate-adjusted version of our estimator, which is designed to improve precision by exploiting additional baseline covariates which were not used for treatment assignment. As discussed in Bai et al. (2023a) and Cytrynbaum (2023), standard covariate adjustments based on a regression using treatment-covariate interactions (see, for instance,am Negi and Wooldridge, 2021, for a succinct treatment) are not guaranteed to improve efficiency when treatment assignment is not completely randomized. For this reason, we consider a modified version of the estimator developed in Bai et al. (2023a) for individual-level matched pair experiments. Our results show that our covariate-adjusted estimator is guaranteed to improve asymptotic efficiency relative to the unadjusted estimator, whenever the matching procedure matches on cluster size. Interestingly, we also find that this improvement in efficiency is *not* guaranteed when cluster size is excluded as a matching variable, and document in a simulation study that in fact such covariate adjustments may increase variance.

The analysis of data from cluster randomized experiments and data from experiments with matched pairs has received considerable attention (see Donner and Klar, 2000; Athey and Imbens, 2017; Hayes and Moulton, 2017, for general overviews), but most recent work has focused on only one of these two features at a time. Recent work on the analysis of cluster randomized experiments includes Middleton and Aronow (2015), Su and Ding (2021), Schochet et al. (2021), and Wang et al. (2022) (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. (2023b), 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. We conclude with recommendations for empirical practice in Section 5.

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

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

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

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

where $\mathcal{M}_g^{(G)} := (\mathcal{M}_g : 1 \le g \le 2G)$, $X^{(G)} := (X_g : 1 \le g \le 2G)$ and $N^{(G)} := (N_g : 1 \le g \le 2G)$. 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) $\{(\mathcal{M}_g, X_g, N_g), 1 \le g \le 2G\}$ is an i.i.d. sequence of random variables.

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

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

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

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

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

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

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

For completeness, we reproduce some of the observations from Bugni et al. (2022) regarding these assumptions. As shown in Bugni et al. (2022), an important implication of Assumptions 2.1(a)–(b) for our purposes is that

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

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, in order to permit the application of relevant laws of large numbers and central limit theorems. Note that Assumption 2.1.(c) does not rule out the possibility of observing arbitrarily large clusters, but does place restrictions on the heterogeneity of cluster sizes. For instance, two consequences of Assumptions 2.1.(a) and (c) are that

$$\frac{\sum_{1\leq g\leq G} N_g^2}{\sum_{1\leq g\leq G} N_g} = O_P(1)\ ,$$

and

$$\frac{\max_{1\leq g\leq G} N_g^2}{\sum_{1\leq g\leq G} N_g} \xrightarrow{P} 0\ ,$$

which mirror heterogeneity restrictions imposed in the analysis of clustered data when cluster sizes are modeled as non-random (see for example Assumption 2 in Hansen and Lee, 2019). Assumptions 2.1(e)–(f) impose high-level restrictions on the two-stage sampling procedure. Assumption 2.1(e) allows the subset of observations sampled by the experimenter to depend on X_g and N_g, but rules out dependence on the potential outcomes within the cluster itself. Assumption 2.1(f) is a high-level assumption which guarantees that we can extrapolate from the observations that are sampled to the observations that are not sampled. It can be shown that Assumptions 2.1(e)–(f) are satisfied if $\mathcal{M}_g$ is drawn as a random sample without replacement from $\{1, 2, \ldots, N_g\}$ in an appropriate sense (see Lemma 2.1 in Bugni et al., 2022).

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 $\Delta = \Delta(Q_G)$.

In Sections 3.1–3.3, we study the asymptotic behavior of the following size-weighted difference-in-means estimator:

$$\hat{\Delta}_G := \hat{\mu}_G(1) - \hat{\mu}_G(0) \ , \tag{3}$$

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 $1 \leq g \leq G$ with probability one, this estimator collapses to the standard difference-in-means estimator. In Section 3.4 we consider a covariate-adjusted modification of $\hat{\Delta}_G$ which is designed to incorporate additional baseline covariates which were not used for treatment assignment.

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. (2023b) 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 (and in some cases possibly invalid). 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. 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 $\{(\bar{Y}_g, D_g, X_g, N_g) : 1 \leq g \leq 2G\}$, 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. ■

Remark 2.3. In Appendix C, we consider a generalization of our main results to settings with multiple treatments (i.e. “matched-tuples” designs) as considered in Bai et al. (2023b). ■

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

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

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 $C < \infty$, $P\{E[N_g|X_g] \leq C\} = 1$.

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 and 3.1–3.3,*

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

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 $\omega^2 \leq \omega_0^2$. 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

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

Conditional on $W^{(G)}$, $(D_{\pi(2g-1)}, D_{\pi(2g)})$, $g = 1, ..., G$ 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 can be verified using the results in Bai et al. (2022).

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 and 3.4–3.6,*

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

as $G \to \infty$, where

$$\nu^2 = E[\tilde{Y}_g^2(1)] + E[\tilde{Y}_g^2(0)] - \frac{1}{2}E[(E[\tilde{Y}_g(1) + \tilde{Y}_g(0)|X_g, N_g])^2] \ , \tag{4}$$

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:

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

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:

$$\hat{v}_G^2 = \hat{\tau}_G^2 - \frac{1}{2}\hat{\lambda}_G^2 , \tag{5}$$

where

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

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. (2023b): 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

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

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

Assumption 3.8. The pairs used in determining treatment status satisfy

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

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

As noted in Bai et al. (2022), given pairs which satisfy Assumptions 3.2 or 3.5, it is frequently possible to reorder the pairs so that Assumptions 3.7 or 3.8 are satisfied. We then obtain the following two consistency results for the estimator $\hat{v}_G^2$:

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

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

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

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

Next, we derive the limits in probability 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 $\hat{\omega}^2_{\mathrm{CR,G}}$, is simply the cluster-robust variance estimator of the coefficient of D_g as defined in equation (21) in the appendix. Theorem 3.4 derives the limit in probability of $\hat{\omega}^2_{\mathrm{CR,G}}$ 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 and 3.1–3.3,*

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

with equality if and only if

$$E[\tilde{Y}_g(1) + \tilde{Y}_g(0)|X_g] = 0 \ . \tag{6}$$

The next variance estimator we consider, which we denote by $\hat{\omega}^2_{\mathrm{PCVE,G}}$, is the variance estimator of the coefficient of D_g obtained from clustering on the assignment *pairs* of clusters as defined in equation (22) in the appendix. de Chaisemartin and Ramirez-Cuellar (2019) call this the pair-cluster variance estimator (PCVE). Theorem 3.5 derives the limit in probability of $\hat{\omega}^2_{\mathrm{PCVE,G}}$ in the special case where $N_g = k$ for $g = 1, \ldots, 2G$ 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 and 3.1–3.3 hold. If in addition we impose that $N_g = k$ for $g = 1, \ldots, 2G$ 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

$$E[\tilde{Y}_g(1) - \tilde{Y}_g(0)|X_g] = 0 \ . \tag{7}$$

Although we do not derive the limit in probability 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 (7). 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 the potential outcomes in an appropriate sense. In particular, equation (7) 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. 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.

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, and 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:

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

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

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

The randomization test we consider is then given by

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

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 $\hat{\mathbf{H}}_G = \{h_1, h_2, \dots, h_B\}$, where h_1 is the identity transformation and $h_2, \dots, h_B$ 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 present 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:

$$H_0^{X,N} : (Y_{i,g}(1) : 1 \leq i \leq N_g)|(X_g, N_g) \stackrel{d}{=} (Y_{i,g}(0) : 1 \leq i \leq N_g)|(X_g, N_g) \ . \tag{8}$$

We then establish the following result on the finite sample validity of our randomization test for testing (8):

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 (8) at level $\alpha \in (0,1)$, $\phi_G^{\text{rand}}(Z^{(G)})$ satisfies*

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

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 $Z^{(G)}$. ■

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

$$H_0 : \Delta(Q_G) = 0 \ . \tag{9}$$

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 (9) at level $\alpha \in (0,1)$, $\phi_G^{\text{rand}}(Z^{(G)})$ satisfies

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

under the null hypothesis.

Theorems 3.6 and 3.7 highlight that the randomization test $\phi_G^{\text{rand}}(Z^{(G)})$ is asymptotically valid for testing (9) while additionally retaining the finite-sample validity described in Section 3.3.1 under the null hypothesis (8). In Section 4.1 we illustrate the benefit of this additional robustness on the small-sample behavior of $\phi_G^{\text{rand}}(Z^{(G)})$ relative to tests constructed using Gaussian critical values. We note that, unlike for the null hypothesis considered in Section 3.3.1, the choice of test statistic T_G is crucial for establishing Theorem 3.7. Similar observations have been made in related contexts in Janssen (1997), Chung and Romano (2013), Bugni et al. (2018) and 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. ■

3.4 Covariate Adjustment

In this section, we consider a linearly covariate-adjusted modification of $\hat{\Delta}_G$ that is designed to improve precision by exploiting additional observed baseline covariates that were not used for treatment assignment. To that end, we consider a setting in which we observe two sets of baseline covariates, X_g and C_g, where $X_g \in \mathbf{R}^k$ denotes the original set of baseline covariates used for treatment assignment, and $C_g \in \mathbf{R}^\ell$ denotes the covariates in addition to X_g that were not used for treatment assignment. Note that C_g could also include cluster-level aggregates of individual-level outcomes, including intracluster means and quantiles. Before proceeding, we note that for the remainder of Section 3.4, the assumptions in Section 2 should be modified such that X_g is replaced by (X_g, C_g) throughout. In particular, references to Assumption 2.1 below should be understood to hold with (X_g, C_g) in place of X_g.

Our primary focus will be on settings in which the cluster size N_g is used in determining the pairs. We comment on the case when N_g is not used in determining pairs in Remark 3.4, and, importantly, note that in such settings the adjustments we consider here are *not* guaranteed to improve precision). As in Section 3.1.2, let $\pi = \pi_G(X^{(G)}, N^{(G)})$ denote the permutation that determines the pairs. We then assume that treatment status is assigned as follows:

Assumption 3.9. Treatment status is assigned so that

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

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 consider a linearly covariate-adjusted estimator of $\Delta(Q)$ based on a set of regressors generated by X_g, N_g, C_g. To this end, define $\psi_g = \psi(X_g, N_g, C_g)$, where $\psi : \mathbf{R}^k \times \mathbf{R} \times \mathbf{R}^\ell \to \mathbf{R}^p$. We impose the following assumptions on ψ:

Assumption 3.10. The function ψ is such that

(a) No component of ψ is a constant and $E[\text{Var}[\psi_g | X_g, N_g]]$ is nonsingular.

(b) $\text{Var}[\psi_g] < \infty$.

(c) $E[\psi_g | W_g = w]$, $E[\psi_g \psi_g' | W_g = w]$, and $E[\psi_g \bar{Y}_g^r(d) | W_g = w]$ for $d \in \{0,1\}$ and $r \in \{1,2\}$ are Lipschitz.

(d) For some $c < \infty$, $P\{E[\|\psi_g\|^2 \bar{Y}_g^2(d) | X_g, N_g] \leq c\} = 1$ for $d \in \{0,1\}$.

As discussed in Bai et al. (2023a) and Cytrynbaum (2023), standard covariate adjustments based on a regression using treatment-covariate interactions (see, for instance, Negi and Wooldridge, 2021, for a succinct treatment) are not guaranteed to improve efficiency when treatment assignment is not completely randomized. For this reason, we consider a modified version of the adjusted estimator developed in Bai et al. (2023a) for individual-level matched pair experiments. Let $\hat{\beta}_G$ denote the OLS estimator of the slope coefficient in the linear regression of of $(\bar{Y}_{\pi(2g-1)} N_{\pi(2g-1)} - \bar{Y}_{\pi(2g)} N_{\pi(2g)})(D_{\pi(2g-1)} - D_{\pi(2g)})$ on a constant and $(\psi_{\pi(2g-1)} - \psi_{\pi(2g)})(D_{\pi(2g-1)} - D_{\pi(2g)})$. We then define our covariate-adjusted estimator as

$$\hat{\Delta}_G^{\text{adj}} = \frac{\frac{1}{G} \sum_{1 \leq g \leq 2G} (\bar{Y}_g N_g - (\psi_g - \bar{\psi}_G)' \hat{\beta}_G) D_g}{\frac{1}{G} \sum_{1 \leq g \leq 2G} N_g D_g} - \frac{\frac{1}{G} \sum_{1 \leq g \leq 2G} (\bar{Y}_g N_g - (\psi_g - \bar{\psi}_G)' \hat{\beta}_G)(1 - D_g)}{\frac{1}{G} \sum_{1 \leq g \leq 2G} N_g (1 - D_g)} , \tag{10}$$

where

$$\bar{\psi}_G = \frac{1}{2G} \sum_{1 \leq g \leq 2G} \psi_g .$$

Theorem 3.8 derives the limiting distribution of $\hat{\Delta}_G^{\text{adj}}$, and, importantly, it shows that the limiting variance of $\hat{\Delta}_G^{\text{adj}}$ is no larger than that of $\hat{\Delta}_G$ in (3) and can be strictly smaller.

Theorem 3.8. *Under Assumptions 2.1, 3.5, 3.6, 3.9, and 3.10,*

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

as $G \to \infty$, *where*

$$\varsigma^2 = E[\text{Var}[Y_g^*(1) | X_g, N_g]] + E[\text{Var}[Y_g^*(0) | X_g, N_g]] + \frac{1}{2} E[(E[Y_g^*(1) - Y_g^*(0) | X_g, N_g] - \Delta)^2] ,$$

with

$$Y_g^*(d) = \frac{\bar{Y}_g(d) N_g - (\psi_g - E[\psi_g])' \beta^*}{E[N_g]} - \frac{N_g}{E[N_g]} \frac{E[\bar{Y}_g(d) N_g - (\psi_g - E[\psi_g])' \beta^*]}{E[N_g]} = \tilde{Y}_g(d) - \frac{(\psi_g - E[\psi_g])' \beta^*}{E[N_g]} ,$$

and

$$\beta^* = (2E[\mathrm{Var}[\psi_g|X_g, N_g]])^{-1} E[\mathrm{Cov}[\psi_g, \bar{Y}_g(1)N_g + \bar{Y}_g(0)N_g|X_g, N_g]] \ . \tag{11}$$

Moreover,

$$\varsigma^2 = \nu^2 - \kappa^2 \ , \tag{12}$$

where ν^2 is as in (4) and

$$\kappa^2 = \frac{E[((\psi_g - E[\psi_g|X_g, N_g])'\beta^*)^2]}{E[N_g]^2} \ .$$

As a consequence, $\varsigma^2 \leq \nu^2$, with equality if and only if $\kappa^2 = 0$.

Note that the asymptotic variance ς^2 has the same form as the variance ν^2, but with new transformed outcomes $Y_g^*(d)$ which can be expressed as covariate-adjusted versions of the original transformed outcomes $\tilde{Y}_g(d)$. Exploiting this observation is what allows us to establish that $\varsigma^2 = \nu^2 - \kappa^2$. As a consequence, we find that the asymptotic variance of $\hat{\Delta}_G^{\mathrm{adj}}$ is lower than that of $\hat{\Delta}_G$ whenever the adjustment is appropriately "relevant," in the sense that $\kappa^2 \neq 0$.

Remark 3.4. In order to guarantee that $\varsigma^2 \leq \nu^2$ in Theorem 3.8, it was crucial to assume that N_g is contained in the set of matching variables. If instead clusters are only matched according to X_g as in Section 3.1.1, then under suitable modifications of Assumptions 3.9 and 3.10 it can be shown that the limiting variance of $\hat{\Delta}_G^{\mathrm{adj}}$ is given by

$$E[\mathrm{Var}[Y_g^*(1)|X_g]] + E[\mathrm{Var}[Y_g^*(0)|X_g]] + \frac{1}{2}E[(E[Y_g^*(1) - Y_g^*(0)|X_g] - \Delta)^2] \ ,$$

where in this case $Y_g^*(d) = \tilde{Y}_g(d) - \frac{(\psi_g - E[\psi_g])'\beta^*}{E[N_g]}$, with

$$\beta^* = (2E[\mathrm{Var}[\psi_g|X_g]])^{-1} E[\mathrm{Cov}[\psi_g, \bar{Y}_g(1)N_g + \bar{Y}_g(0)N_g|X_g]] \ .$$

Note that this expression mirrors the expression for ς^2 but removes the conditioning on N_g throughout. It can then be shown that the decomposition obtained in (12) no longer applies, and in general the covariate-adjusted estimator is no longer guaranteed to have a smaller limiting variance than the unadjusted estimator $\hat{\Delta}_G$. We illustrate this point via simulation in Section 4.2. ■

Remark 3.5. Although the estimator in (10) is closely related to the class of covariate-adjusted estimators in Bai et al. (2023a), a direct application of the results therein is prohibited because the two denominators in (10) are the average cluster sizes of treated and untreated clusters and are therefore random. As a result, unlike in Bai et al. (2023a), the demeaning of ψ in (10) is crucial for the results in Theorem 3.8 to hold. In particular, some remainder terms in the proof of Theorem 3.8 are no longer $o_P(1)$ without the demeaning. Moreover, unlike for individual-level experiments, $\hat{\Delta}_G^{\mathrm{adj}}$ cannot be interpreted as the intercept of a linear regression as in Bai et al. (2023a). ■

For variance estimation, define

$$\mathring{Y}_g = \frac{1}{\frac{1}{2G}\sum_{1 \leq j \leq 2G} N_j} \left(N_g \bar{Y}_g - N_g \frac{\frac{1}{G}\sum_{1 \leq j \leq 2G} \bar{Y}_j I\{D_j = D_g\} N_j}{\frac{1}{G}\sum_{1 \leq j \leq 2G} I\{D_j = D_g\} N_j} - \psi_g' \hat{\beta}_G \right) \ .$$

We then propose the following variance estimator:

$$\mathring{\varsigma}_G^2 = \mathring{\tau}_G^2 - \frac{1}{2}\mathring{\lambda}_G^2 \ , \tag{13}$$

where

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

The following theorem establishes the consistency of the variance estimator:

Theorem 3.9. *Under Assumptions 3.5, 3.6, 3.8, 3.9, and 3.10,*

$$\mathring{\varsigma}_G^2 \xrightarrow{P} \varsigma^2 \ .$$

4 Simulations

4.1 Unadjusted Estimation

In this section, we examine the finite-sample behavior of the estimation and inference procedures considered in Sections 3.1-3.3. 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 $d \in \{0,1\}$, $1 \le g \le 2G$, 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, $(X_g, X_g^{(N)})$, $g = 1, \dots, 2G$ are i.i.d. with $X_g, X_g^{(N)} \sim \ Beta(2,4)$, and $(\epsilon_{0,i,g}, \epsilon_{1,i,g})$, $g = 1, \dots, 2G$, $i = 1, \dots, N_g$ are i.i.d. with $\epsilon_{0,i,g}, \epsilon_{1,i,g} \sim N(0,1)$ independently. 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 (7) 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 feature 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 in 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.[1] 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 also size-distortions in small samples. In contrast, since Model 1 satisfies the null hypothesis considered in (8), our randomization test is valid in finite samples by construction, and displays comparable power to the PCVE-based test even when the latter 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. Importantly, 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 $G = 12$.

[1]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.

4.2 Covariate-Adjusted Estimation

In this section, we examine the finite-sample behavior of the covariate-adjusted estimator considered in Section 3.4. In particular, we contrast the efficiency properties of $\hat{\Delta}_G^{\text{adj}}$ when matching versus not matching on cluster size. We consider the following modification of Model 2:

Model Adj.: $\mu_1(X_g, X_g^{(N)}) = 10(X_g^2 - 1/7) + 6(X_g^{(N)} - 1/3) + 25$ and $\mu_0(X_g, X_g^{(N)}) = 0$.

In addition, we introduce a new matching variable H_g, $g = 1, \ldots, 2G$, i.i.d. with $H_g \sim U[0, 1]$ generated independently of all other variables, and modify the distribution of N_g so that $N_g \sim Binomial(R, 1 - X_g^{(N)}) + (500 - R)$.

Table 7 reports the coverage and average length of 95% confidence intervals constructed using our variance estimators when matching using both H_g and N_g, for $\hat{\Delta}_G$ versus $\hat{\Delta}_G^{\text{adj}}$ with $\psi_g = (X_g, X_g^{(N)})$. In accordance with Theorem 3.8, we find that for moderate to large samples ($G \geq 50$), covariate adjustment leads to smaller average CI lengths even as we increase the amount of cluster size heterogeneity. In contrast, Table 8 reports the coverage and average lengths of 95% confidence intervals (CIs) constructed using our variance estimators when matching using *only* H_g, for $\hat{\Delta}_G$ versus $\hat{\Delta}_G^{\text{adj}}$ with $\psi_g = (X_g, X_g^{(N)})$. In general, we find that when cluster-size heterogeneity is low, covariate adjustment leads to smaller average CI lengths. However, as the amount of heterogeneity increases, the average CI length for the adjusted estimator rapidly overtakes the length for the unadjusted estimator. We emphasize that this does not seem to be a small-sample issue: even with $G = 250$, the average CI length for the adjusted estimator is over two times larger than for the unadjusted estimator in the most extreme case.

5 Recommendations for Empirical Practice

Based on our theoretical results as well as the simulation study above, we conclude with some recommendations for practitioners when conducting inference about the size-weighted ATE in our super-population framework. Our recommendations below depend on whether the number of clusters is moderately large (e.g., at least 50 pairs) or small (e.g., less than 50 pairs).

If the number of clusters is moderately large, then our recommendation is that practitioners should employ either the covariate-adjusted tests based on the covariate-adjusted estimator $\hat{\Delta}_G^{\text{adj}}$ defined in Section 3.4 paired with its corresponding variance estimator $\mathring{\varsigma}_G^2$ and a normal critical value or the unadjusted tests based on the unadjusted estimator $\hat{\Delta}_G$ introduced in Section 2 paired with its corresponding variance estimator $\hat{v}_G^2$ and a normal critical value. Whenever cluster size is used in determining the pairs, our results show that covariate-adjusted tests are more powerful in large samples than unadjusted tests; in practice, this feature continues to hold even when cluster size was not used in determining the pairs, provided that cluster-size heterogeneity is not too great (i.e., in our simulations, a ratio of largest to smallest cluster size of less than 2). Outside of these circumstances, we recommend that practitioners employ the unadjusted tests.

If, on the other hand, the number of clusters is small, then we recommend instead that practitioners use the randomization test based on the un-adjusted estimator $\hat{\Delta}_G$ paired with its corresponding variance estimator $\hat{v}_G^2$ outlined in Section 3.3. In our simulations, this test controlled size more reliably than any of the other inference procedures we considered in the paper, while delivering comparable power. Note that by modifying the test as in Remark 3.3, the test could also be inverted to construct confidence intervals if desired.

Table 1: Model 1 - Matching on X_g *

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

* 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*

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

* 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 *

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

* 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 [*]

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

[*] 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 *

		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

* 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*

N_{max}/N_{min}		Size under H_0 $G=12$	Size under H_0 $G=26$	Size under H_0 $G=50$	Power under $H_1: \Delta_0 + 1/4$ $G=12$	Power under $H_1: \Delta_0 + 1/4$ $G=26$	Power under $H_1: \Delta_0 + 1/4$ $G=50$
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

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

Table 7: Covariate Adjustment - Matching on H_g and N_g *

N_{max}/N_{min}	ψ_g	$G = 12$	$G = 26$	$G = 50$	$G = 100$	$G = 150$	$G = 200$	$G = 250$
				Coverage				
1.11	-	0.9120	0.9275	0.9475	0.9395	0.9425	0.9510	0.9425
	$(X_g, X_g^{(N)})$	0.8625	0.8970	0.9360	0.9405	0.9440	0.9495	0.9495
1.42	-	0.9135	0.9245	0.9415	0.9445	0.9495	0.9425	0.9425
	$(X_g, X_g^{(N)})$	0.8990	0.9195	0.9375	0.9515	0.9470	0.9515	0.9455
1.99	-	0.9085	0.9250	0.9420	0.9470	0.9455	0.9545	0.9520
	$(X_g, X_g^{(N)})$	0.9175	0.9355	0.9500	0.9520	0.9505	0.9505	0.9470
3.31	-	0.9090	0.9265	0.9340	0.9515	0.9465	0.9465	0.9535
	$(X_g, X_g^{(N)})$	0.9335	0.9365	0.9480	0.9515	0.9510	0.9525	0.9550
9.80	-	0.9070	0.9245	0.9330	0.9375	0.9510	0.9455	0.9440
	$(X_g, X_g^{(N)})$	0.9325	0.9340	0.9475	0.9470	0.9575	0.9500	0.9555
				Average Length				
1.11	-	1.77556	1.21499	0.88201	0.62584	0.51123	0.44346	0.39699
	$(X_g, X_g^{(N)})$	1.30671	0.93116	0.68816	0.49242	0.40372	0.35104	0.31400
1.42	-	1.74117	1.20501	0.87067	0.62002	0.50712	0.43888	0.39274
	$(X_g, X_g^{(N)})$	1.46021	0.96656	0.69879	0.49479	0.40412	0.35025	0.31292
1.99	-	1.72916	1.19588	0.86887	0.61669	0.50509	0.43677	0.39112
	$(X_g, X_g^{(N)})$	1.81983	1.09008	0.74580	0.50919	0.41110	0.35398	0.31603
3.31	-	1.71004	1.19463	0.86708	0.61577	0.50301	0.43573	0.39127
	$(X_g, X_g^{(N)})$	2.36813	1.30774	0.83203	0.54137	0.42815	0.36460	0.32354
9.80	-	1.72505	1.19952	0.86484	0.61768	0.50429	0.43672	0.39197
	$(X_g, X_g^{(N)})$	3.06889	1.60986	0.97620	0.59917	0.46025	0.38545	0.33953

* 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 8: Covariate Adjustment - Matching on ${H_g}^*$

N_{max}/N_{min}	ψ_g	$G=12$	$G=26$	$G=50$	$G=100$	$G=150$	$G=200$	$G=250$
				Coverage				
1.11	-	0.9015	0.9235	0.9435	0.9395	0.9365	0.9445	0.9485
	$(X_g, X_g^{(N)})$	0.8485	0.9060	0.9275	0.9425	0.9420	0.9510	0.9430
1.42	-	0.9070	0.9315	0.9365	0.9405	0.9455	0.9490	0.9525
	$(X_g, X_g^{(N)})$	0.9005	0.9230	0.9465	0.9510	0.9430	0.9475	0.9520
1.99	-	0.9050	0.9310	0.9450	0.9450	0.9480	0.9530	0.9465
	$(X_g, X_g^{(N)})$	0.9190	0.9395	0.9485	0.9470	0.9520	0.9495	0.9515
3.31	-	0.9100	0.9340	0.9410	0.9535	0.9520	0.9490	0.9485
	$(X_g, X_g^{(N)})$	0.9155	0.9325	0.9475	0.9485	0.9435	0.9535	0.9510
9.80	-	0.8975	0.9305	0.9410	0.9435	0.9420	0.9430	0.9545
	$(X_g, X_g^{(N)})$	0.9190	0.9440	0.9345	0.9455	0.9405	0.9490	0.9410
				Average Length				
1.11	-	1.86744	1.31289	0.95830	0.68388	0.55761	0.48368	0.43289
	$(X_g, X_g^{(N)})$	1.30222	0.94977	0.70427	0.50804	0.41405	0.36055	0.32280
1.42	-	1.86822	1.30105	0.95121	0.67677	0.55462	0.48111	0.43046
	$(X_g, X_g^{(N)})$	1.76667	1.22571	0.89458	0.63665	0.52213	0.45247	0.40482
1.99	-	1.85639	1.29289	0.94626	0.67421	0.55160	0.47822	0.42849
	$(X_g, X_g^{(N)})$	2.54781	1.72304	1.25092	0.87988	0.72210	0.62598	0.55911
3.31	-	1.83716	1.29155	0.94173	0.67099	0.54871	0.47588	0.42645
	$(X_g, X_g^{(N)})$	3.56010	2.39697	1.73381	1.22024	0.99619	0.86635	0.77370
9.80	-	1.83555	1.28894	0.93697	0.66756	0.54602	0.47402	0.42411
	$(X_g, X_g^{(N)})$	4.86067	3.24720	2.34399	1.64604	1.34678	1.16835	1.04106

* Number of clusters= $2G$ with $G = 12, 26, 50, 100, 150, 200, 250$. 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

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

$\frac{1}{G}\sum_{g=1}^{G} N_{\pi(2g)}^{2\ell} \leq \frac{1}{G}\sum_{g=1}^{2G} N_{g}^{2\ell} = O_P(1)$ 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

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

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

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

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,

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

and similarly for the other three terms. The desired conclusion then follows from Lemma A.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

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

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

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

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

Next, the terms involving conditional expectations are

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

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

■

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

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

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

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

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

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

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

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

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

For the upper left component, we have

$$\frac{1}{G}\sum_{1\leq g\leq 2G} \text{Var}[\bar{Y}_g(1)N_g|X_g]D_g = \frac{1}{G}\sum_{1\leq g\leq 2G} E[\bar{Y}_g^2(1)N_g^2|X_g]D_g - \frac{1}{G}\sum_{1\leq g\leq 2G} E[\bar{Y}_g(1)N_g|X_g]^2 D_g . \tag{14}$$

Note

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

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

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

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

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

Therefore,

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

Meanwhile,

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

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,

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

and hence it follows from (14) 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

$$\frac{1}{G}\sum_{1\leq g\leq 2G}\mathrm{Cov}[\bar{Y}_g(1)N_g,N_g|X_g]D_g=\frac{1}{G}\sum_{1\leq g\leq 2G}E[\bar{Y}_g(1)N_g^2|X_g]D_g-\frac{1}{G}\sum_{1\leq g\leq 2G}E[\bar{Y}_g(1)N_g|X_g]E[N_g|X_g]D_g\ . \tag{15}$$

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

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

Meanwhile,

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

Note that

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

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

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

and for each j,

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

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

Thus we have that

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

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

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

Thus we have established that

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

Similarly,

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

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

$$
\rho(\mathcal{L}((\mathbb{L}_{1,G}^{\mathrm{YN1}}, \mathbb{L}_{1,G}^{\mathrm{N1}}, \mathbb{L}_{1,G}^{\mathrm{YN0}}, \mathbb{L}_{1,G}^{\mathrm{N0}})'|X^{(G)}, D^{(G)}), N(0, \mathbb{V}_1)) \xrightarrow{P} 0 , \tag{16}
$$

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

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

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

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

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

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

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

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

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

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

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

Because (16) 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
$$\hat{\Delta}_G = \frac{\frac{1}{G}\sum_{1\le g\le 2G}\bar{Y}_g(1)N_gD_g}{\frac{1}{G}\sum_{1\le g\le 2G}N_gD_g} - \frac{\frac{1}{G}\sum_{1\le g\le 2G}\bar{Y}_g(0)N_g(1-D_g)}{\frac{1}{G}\sum_{1\le g\le 2G}N_g(1-D_g)} .$$
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 3.4,
$$\sqrt{G}\Big(\frac{1}{G}\sum_{1\le g\le 2G}\bar{Y}_gN_gD_g - E[\bar{Y}_g(1)N_g]\Big) = \frac{1}{\sqrt{G}}\sum_{1\le g\le 2G}(\bar{Y}_g(1)N_gD_g - E[\bar{Y}_g(1)N_g]D_g)$$
and similarly for the other three terms. The desired conclusion then follows from Lemma A.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{\mathrm{Var}[E[\bar{Y}_g(1)N_g|W_g]]}{2E[N_g]^2} - \frac{\mathrm{Var}[N_g]E[\bar{Y}_g(1)N_g]^2}{2E[N_g]^4} \\
&\quad - \frac{\mathrm{Var}[E[\bar{Y}_g(0)N_g|W_g]]}{2E[N_g]^2} - \frac{\mathrm{Var}[N_g]E[\bar{Y}_g(0)N_g]^2}{2E[N_g]^4} \\
&\quad + \frac{\mathrm{Cov}[E[\bar{Y}_g(1)N_g|W_g], N_g]E[\bar{Y}_g(1)N_g]}{E[N_g]^3} + \frac{\mathrm{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{\mathrm{Cov}[E[\bar{Y}_g(1)N_g|W_g], E[\bar{Y}_g(0)N_g|W_g]]}{E[N_g]^2} + \frac{\mathrm{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{\mathrm{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{\mathrm{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*

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

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

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

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

$$\frac{1}{G} \sum_{1 \le g \le 2G} \mathrm{Var}[\bar{Y}_g(1) N_g | W_g] D_g = \frac{1}{G} \sum_{1 \le g \le 2G} E[\bar{Y}_g^2(1) N_g^2 | W_g] D_g - \frac{1}{G} \sum_{1 \le g \le 2G} E[\bar{Y}_g(1) N_g | W_g]^2 D_g ~. \tag{17}$$

Note

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

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,

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

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 (17) that

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

Similarly,

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

We now establish

$$\rho(\mathcal{L}((\mathbb{L}^{\text{YN1}}_{1,G},0,\mathbb{L}^{\text{YN0}}_{1,G},0)|W^{(G)},D^{(G)}),N(0,\mathbb{V}_1)) \xrightarrow{P} 0\ , \tag{18}$$

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

$$\rho(\mathcal{L}((\mathbb{L}^{\text{YN1}}_{1,G_{k_l}},0,\mathbb{L}^{\text{YN0}}_{1,G_{k_l}},0)|W^{(G_{k_l})},D^{(G^{k_l})},N(0,\mathbb{V}_1)) \to 0 \text{ with probability one}\ . \tag{19}$$

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,

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

Next, We will use the inequality

$$\left|\sum_{1\leq j\leq k}a_j\right| I\left\{\left|\sum_{1\leq j\leq k}a_j\right|>\epsilon\right\} \leq \sum_{1\leq j\leq k}k|a_j|I\left\{|a_j|>\frac{\epsilon}{k}\right\}\ . \tag{20}$$

It follows from (20) that

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

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

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

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 (19) and hence (18) 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 (16) 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),

$$\hat{\omega}_{\mathrm{CR,G}}^2 = \frac{1}{2} \left(\hat{\omega}_{\mathrm{CR,G}}^2(1) + \hat{\omega}_{\mathrm{CR,G}}^2(0) \right) , \tag{21}$$

(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*:

$$\left(\frac{1}{G}\sum_{1\le j\le G}\sum_{g\in\lambda_j} X_g'X_g\right)^{-1}\left(\frac{1}{G}\sum_{1\le j\le G}\left(\sum_{g\in\lambda_j}X_g'\hat{\epsilon}_g\right)\left(\sum_{g\in\lambda_j}X_g'\hat{\epsilon}_g\right)'\right)\left(\frac{1}{G}\sum_{1\le g\le G}\sum_{g\in\lambda_j}X_g'X_g\right)^{-1}, \tag{22}$$

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. (2023b), it can be shown that

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

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

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

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

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

The conclusion then follows. ■

A.7 Proof of Theorem 3.6

Proof. Note that the null hypothesis (8) combined with Assumption 2.1(e) implies that

$$\bar{Y}_g(1)|(X_g,N_g) \stackrel{d}{=} \bar{Y}_g(0)|(X_g,N_g)\ . \tag{23}$$

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 (23) is in fact equivalent to the statement

$$(\bar{Y}_g(1),N_g)|X_g \stackrel{d}{=} (\bar{Y}_g(0),N_g)|X_g\ . \tag{24}$$

The result then follows by applying Theorem 3.4 in Bai et al. (2022) using (24) as the null hypothesis. To establish this equivalence, we first begin with (23) 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

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

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

where the first and second equalities follow from the definition of conditional expectations, the the third follows from (23), 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

Note that

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

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

$$\tilde{R}_G(t) := P\left\{\sqrt{G}\check{\Delta}(\epsilon_1, \ldots, \epsilon_G) \le t \middle| Z^{(G)}\right\} , \quad (25)$$

where

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

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

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

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

$$\check{v}_G^2(\epsilon_1, \ldots, \epsilon_G) = \hat{\tau}_G^2 - \frac{1}{2}\check{\lambda}_G^2(\epsilon_1, \ldots, \epsilon_G) \quad (26)$$

where $\hat{\tau}_G^2$ is defined in (5), and

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

The desired conclusion then follows from Lemmas B.7 and B.9, along with Theorem 5.2 in Chung and Romano (2013). ■

A.9 Proof of Theorem 3.8

We first show that $\hat{\beta}_G \xrightarrow{P} \beta^*$. The proof follows almost verbatim Theorem 4.2 in Bai et al. (2023a) with a few minor differences because we match on N_g, which could all be resolved as in the proof of Lemma B.3. To establish the limiting distribution, first define

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

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

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

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

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

and

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

The last equality follows from the arguments that establish (50) in Bai et al. (2023a). Define

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

It follows from previous arguments that

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

Recall that

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

It then follows from the proof of Theorem 3.2 that $\sqrt{G} (\hat{\Delta}_G^{\text{adj}} - \Delta) \xrightarrow{d} N(0, \varsigma^2)$, where

$$\varsigma^2 = E[\text{Var}[Y_g^*(1) | X_g, N_g]] + E[\text{Var}[Y_g^*(0) | X_g, N_g]] + \frac{1}{2} E[(E[Y_g^*(1) - Y_g^*(0) | X_g, N_g] - \Delta)^2] ,$$

where

$$Y_g^*(d) = \frac{\bar{Y}_g(d) N_g - (\psi_g - E[\psi_g])' \beta^*}{E[N_g]} - \frac{N_g}{E[N_g]} \frac{E[\bar{Y}_g(d) N_g - (\psi_g - E[\psi_g])' \beta^*]}{E[N_g]} = \tilde{Y}_g(d) - \frac{(\psi_g - E[\psi_g])' \beta^*}{E[N_g]}$$

for $d \in \{0, 1\}$. All relevant assumptions for Theorem 3.2 have their counterparts stated in Theorem 3.8.

Next we show that $\varsigma^2 \leq \nu^2$. First note that by definition it follows immediately that

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

It thus remains to show that

$$E[\text{Var}[Y_g^*(1)|X_g, N_g]] + E[\text{Var}[Y_g^*(0)|X_g, N_g]] \leq E[\text{Var}[\tilde{Y}_g(1)|X_g, N_g]] + E[\text{Var}[\tilde{Y}_g(0)|X_g, N_g]] \ .$$

To that end,

$$\begin{aligned}
&E[\text{Var}[Y_g^*(1)|X_g, N_g]] + E[\text{Var}[Y_g^*(0)|X_g, N_g]] \\
&= E\left[\text{Var}\left[\tilde{Y}_g(1) - \frac{(\psi_g - E[\psi_g])'\beta^*}{E[N_g]}\Big|X_g, N_g\right]\right] + E\left[\text{Var}\left[\tilde{Y}_g(0) - \frac{(\psi_g - E[\psi_g])'\beta^*}{E[N_g]}\Big|X_g, N_g\right]\right] \\
&= E[\text{Var}[\tilde{Y}_g(1)|X_g, N_g]] + E[\text{Var}[\tilde{Y}_g(0)|X_g, N_g]] - \frac{2E[((\psi_g - E[\psi_g|X_g, N_g])'\beta^*)^2]}{E[N_g]^2} - 2E[\text{Cov}[N_g, \psi_g'\beta^*|X_g, N_g]]\frac{E[\bar{Y}_g(1)N_g] + E[\bar{Y}_g(0)N_g]}{E[N_g]^3} \ ,
\end{aligned}$$

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

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

or equivalently,

$$E[\text{Cov}[\bar{Y}_g(1)N_g + \bar{Y}_g(0)N_g, \psi_g'\beta^*|X_g, N_g]] = 2E[\text{Var}[\psi_g'\beta^*|X_g, N_g]] \ . \tag{27}$$

We thus obtain

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

once we notice that $\text{Cov}[N_g, \psi_g'\beta^*|X_g, N_g] = 0$, as desired. Finally, note that if we do not match on N_g, then we have that

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

but the last term no longer necessarily evaluates to zero.

A.10 Proof of Theorem 3.9

The theorem follows from combining the arguments used to establish Theorem 3.3 and those used to establish Theorem 3.2 in Bai et al. (2023a). ■

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.

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

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

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

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,

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

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,

$$\begin{aligned}
&\frac{1}{G}\sum_{j=1}^{G}\left|\left(\mu_1(W_{\pi(2j)})-\mu_1(W_{\pi(2j-1)})\right)\left(\mu_0(W_{\pi(2j)})-\mu_0(W_{\pi(2j-1)})\right)\right| \\
&\quad \le \frac{1}{G}\sum_{j=1}^{G}\left|\mu_1(W_{\pi(2j)})-\mu_1(W_{\pi(2j-1)})\right|\left|\mu_0(W_{\pi(2j)})-\mu_0(W_{\pi(2j-1)})\right| \\
&\quad \lesssim \frac{1}{G}\sum_{j=1}^{G}N^2_{\pi(2j)}\left|W_{\pi(2j)}-W_{\pi(2j-1)}\right|^2 \xrightarrow{P} 0 \ ,
\end{aligned}$$

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:*

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

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

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

PROOF. It suffices to show that

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

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

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

Note that

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

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

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

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

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

Thus, the result follows. ■

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

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

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

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

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

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

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

for which the first term is given as follows:

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

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

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

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

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

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 et al. (2022) and Lemma B.3. Hence, in the case where we match on cluster size,

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

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

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

Note that

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

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 (25)). Then under the null hypothesis (9), 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,

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

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

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

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

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

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

Combining this with Lemma B.8 and a straightforward modification of Lemma A.3. in Chung and Romano (2013) to two dimensional distributions, we obtain that

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

where when we match on cluster size

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

and when we do *not* match on cluster size

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

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

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

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

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

with

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

and when we do not match on cluster size,

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

with

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

PROOF. Using the fact that ϵ_j, $j = 1, \dots, G$ and $\epsilon_j(D_{\pi(2j)} - D_{\pi(2j-1)})$, $j = 1, \dots, G$ 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

$$\rho\left(\mathcal{L}\left((\tilde{\mathbb{K}}_G^{YN}, \tilde{\mathbb{K}}_G^N)' | Z^{(G)} \right), N(0, \mathbb{V}_R) \right) \xrightarrow{P} 0 , \tag{28}$$

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

$$\rho\left(\mathcal{L}\left((\tilde{\mathbb{K}}_{G_{k_l}}^{YN}, \tilde{\mathbb{K}}_{G_{k_l}}^N)' | Z^{(G_{k_l})} \right), N(0, \mathbb{V}_R) \right) \to 0 \text{ with probability one} . \tag{29}$$

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,

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

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

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

where the first inequality follows from exploiting the fact that $|a-b|^2 \le 2(a^2+b^2)$ and the definition of $\bar{Y}_g$, and the final equality follows from Lemma 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.

$$\mathbb{V}_{R,n} \xrightarrow{P} \mathbb{V}_R . \tag{30}$$

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 (30) holds. Next, we verify the Lindeberg condition in Proposition 2.27 of van der Vaart (1998). Note that

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

$$+\frac{1}{G}\sum_{1\leq j\leq G}(N_{\pi(2j)}-N_{\pi(2j-1)})^2 I\{(N_{\pi(2j)}-N_{\pi(2j-1)})^2>\epsilon^2 G/2\}\ .$$

where the inequality follows from (20) and the fact that $(N_g,\bar{Y}_g), 1\leq g\leq 2G$ are all constants conditional on $Z^{(G)}$. The last line converges in probability to zero as long as we can show

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

Note

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

Where the first inequality follows from the fact that $|a-b|^2\leq 2(a^2+b^2)$, 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 (29) holds, and the conclusion of the lemma follows. ■

Lemma B.9. *Let $\check{v}^2_G(\epsilon_1,\ldots,\epsilon_G)$ be defined as in equation (26). If Assumption 2.1 holds, and Assumptions 3.6-3.5 (or Assumptions 3.3-3.2) hold,*

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

where τ^2 is defined in (B.7).

PROOF. From Lemma B.5, we see that $\hat{\tau}^2_G\xrightarrow{P}\tau^2$. It therefore suffices to show that $\check{\lambda}^2_G(\epsilon_1,\ldots,\epsilon_G)\xrightarrow{P}0$. In order to do so, note that $\check{\lambda}^2_G(\epsilon_1,\ldots,\epsilon_G)$ may be decomposed into sums of the form

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

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

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

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

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

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.

$$\frac{2}{G}\sum_{1\leq j\leq\lfloor\frac{G}{2}\rfloor}\epsilon_{2j-1}\epsilon_{2j}\left(\hat{Y}_{\pi(4j-3)}\hat{Y}_{\pi(4j-1)}-\tilde{Y}_{\pi(4j-3)}\tilde{Y}_{\pi(4j-1)}\right)D_{\pi(4j-3)}D_{\pi(4j-1')}$$
$$=\frac{2}{G}\sum_{1\leq j\leq\lfloor\frac{G}{2}\rfloor}\epsilon_{2j-1}\epsilon_{2j}\left(\hat{Y}_{\pi(4j-3)}(1)\hat{Y}_{\pi(4j-1)}(1)-\tilde{Y}_{\pi(4j-3)}(1)\tilde{Y}_{\pi(4j-1)}(1)\right)D_{\pi(4j-3)}D_{\pi(4j-1')}$$

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

for which the first term is given as follows:

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

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

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

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

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

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

The result then follows immediately. ■

C Analysis of Matched Tuples designs

In this section we state generalizations of the results presented in Sections 3.1 and 3.2 to settings with more than two treatments. We focus on the case when not matching on cluster size; similar results should follow for the case of matching on cluster size analogously.

C.1 Setup and Main Results

We follow the general setup of Bai et al. (2023b) generalized to a setting with clustered assignment. Let $D_g \in \mathcal{D}$ denote treatment status for the gth cluster, where $\mathcal{D} = \{1, \dots, |\mathcal{D}|\}$ denotes a finite set of values of the treatment. For $d \in \mathcal{D}$, let $Y_{i,g}(d)$ denote the potential outcome for the ith unit in the gth cluster if its treatment status were d. The observed outcome and potential outcomes are related to treatment status by the expression

$$Y_{i,g} = \sum_{d \in \mathcal{D}} Y_{i,g}(d) I\{D_g = d\} .$$

We suppose our sample consists of $J_G := (|\mathcal{D}|)G$ i.i.d. clusters. Now we have

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

and

$$((((Y_{i,g}(d) : d \in \mathcal{D}) : 1 \le i \le N_g), \mathcal{M}_g, X_g, N_g) : 1 \le g \le J_G) .$$

Our object of interest will generically be defined as a vector of linear contrasts over the collection of size-weighted cluster-level expected potential outcomes across treatments. Formally, let

$$\Gamma(Q_G) := (\Gamma_1(Q_G), \dots, \Gamma_{|\mathcal{D}|}(Q_G))',$$

where

$$\Gamma_d(Q_G) := \frac{1}{E[N_g]} E\left[\sum_{i=1}^{N_g} Y_{ig}(d)\right]$$

for $d \in \mathcal{D}$. Let ν be a real-valued $m \times |\mathcal{D}|$ matrix. Define

$$\Delta_\nu(Q_G) := \nu \Gamma(Q_G) \in \mathbf{R}^m ,$$

as our generic parameter of interest. We maintain the following generalization of Assumptions 2.1 and 3.3.

Assumption C.1. The distribution Q_G is such that

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

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

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

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

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

(e) $\mathcal{M}_g \perp\!\!\!\perp ((Y_{i,g}(d) : d \in \mathcal{D}) : 1 \le i \le N_g) \bigm| X_g, N_g$ for all $1 \le g \le J_G$.

(f) For $d \in \mathcal{D}$ and $1 \le g \le J_G$,

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

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

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

Following Bai et al. (2023b), the G blocks in a matched tuples design may then be represented by the sets

$$\lambda_j = \lambda_j(X^{(G)}) \subseteq \{1, 2, \ldots, J_G\} ,$$

for $1 \le j \le G$. We then maintain the following two assumptions on the treatment assignment mechanism which generalize Assumptions 3.1, 3.2, and 3.7:

Assumption C.2. Treatment status is assigned so that $\left\{\left((Y_{ig}(d) : d \in \mathcal{D})_{1 \le i \le N_g}, N_g\right)\right\}_{g=1}^G \perp\!\!\!\perp D^{(G)} | X^{(G)}$, and, conditional on $X^{(G)}$,

$$\{(D_g : g \in \lambda_j) : 1 \le j \le G\} ,$$

are i.i.d. and each uniformly distributed over all permutations of $(1, 2, \ldots, |\mathcal{D}|)$.

Assumption C.3. The blocks satisfy

$$\frac{1}{G} \sum_{1 \le j \le G} \max_{i,k \in \lambda_j} |X_i - X_k|^2 \xrightarrow{P} 0 .$$

Assumption C.4. The blocks satisfy

$$\frac{1}{G} \sum_{1 \le j \le \lfloor \frac{G}{2} \rfloor} \max_{i \in \lambda_{2j-1}, k \in \lambda_{2j}} |X_i - X_k|^2 \xrightarrow{P} 0 .$$

The estimator for $\Delta_\nu(Q_G)$ is given by

$$\hat{\Delta}_{\nu,G} := \nu \hat{\Gamma}_G ,$$

where for $d \in \mathcal{D}$ we define

$$\hat{\Gamma}_G(d) := \frac{1}{N(d)} \sum_{1 \le g \le J_G} I\{D_g = d\} \frac{N_g}{|\mathcal{M}_g|} \sum_{i \in \mathcal{M}_g} Y_{ig} ,$$

with

$$N(d) = \sum_{1 \le g \le J_G} N_g I\{D_g = d\} .$$

and let $\hat{\Gamma}_G = (\hat{\Gamma}_G(1), \ldots, \hat{\Gamma}_G(|\mathcal{D}|))'$.

Our first result derives the limiting distribution of $\hat{\Delta}_{\nu,G}$ under our maintained assumptions.

Theorem C.1. *Suppose Assumptions C.1-C.3 holds. Then,*

$$\sqrt{G}(\hat{\Delta}_{\nu,G} - \Delta_\nu(Q)) \xrightarrow{d} N(0, \mathbf{V}_\nu) ,$$

where $\mathbf{V}_\nu := \nu \mathbf{V} \nu'$, *with*

$$\mathbf{V} := \mathbf{V}_1 + \mathbf{V}_2 , \tag{31}$$

$$\mathbf{V}_1 := \operatorname{diag}(E[\operatorname{Var}[\tilde{Y}_g(d)|X_i]] : d \in \mathcal{D}) ,$$

$$\mathbf{V}_2 := \left[\frac{1}{|\mathcal{D}|} \operatorname{Cov}[E[\tilde{Y}_g(d)|X_i], E[\tilde{Y}_g(d'))|X_i]]\right]_{d,d' \in \mathcal{D}} .$$

Proof. We show that $\sqrt{G}(\hat{\Gamma}_G(d) - \Gamma_G(Q) : d \in \mathcal{D}) \xrightarrow{d} N(0, \mathbf{V})$, from which the conclusion of the theorem follows by an application of the continuous mapping theorem. To show this we repeat the arguments from the proof of Theorem 3.1 while using the Delta method for vector-valued functions with $h(x_1, y_1, \ldots, x_{|\mathcal{D}|}, y_{|\mathcal{D}|}) = (x_d/y_d : d \in \mathcal{D})$ and using the fact that

$$(\hat{\Gamma}_G(d) : d \in \mathcal{D}) = \left(\frac{\frac{1}{\sqrt{G}} \sum_{1 \le g \le J_G} \bar{Y}_g(d) N_g I\{D_g = d\}}{\frac{1}{\sqrt{G}} \sum_{1 \le g \le J_G} N_g I\{D_g = d\}} : d \in \mathcal{D}\right) .$$

The Jacobian is given by

$$D_h(x_1, y_1, \ldots, x_{|\mathcal{D}|}, y_{|\mathcal{D}|}) = \begin{bmatrix} \frac{1}{y_1} & 0 & \ldots & 0 \\ -\frac{x_1}{y_1^2} & 0 & \ldots & 0 \\ 0 & \frac{1}{y_2} & \ldots & 0 \\ 0 & -\frac{x_2}{y_2^2} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \frac{1}{y_{|\mathcal{D}|}} \\ 0 & 0 & \ldots & -\frac{x_{|\mathcal{D}|}}{y_{|\mathcal{D}|}^2} \end{bmatrix} .$$

Repeating the algebra in proof of binary case, we obtain

$$D_h((E[\bar{Y}_g(d)N_g], E[N_g]) : d \in \mathcal{D})' \mathbb{V} D_h((E[\bar{Y}_g(d)N_g], E[N_g]) : d \in \mathcal{D}) = \mathbf{V} ,$$

where $\mathbb{V}$ is defined in the statement of Lemma C.1. ∎

Following Bai et al. (2023b), our estimator for $\mathbf{V}_\nu$ is then given by $\hat{\mathbf{V}}_{\nu,G} := \nu \hat{\mathbf{V}}_G \nu'$, where

$$\begin{aligned} \hat{\mathbf{V}}_G &:= \hat{\mathbf{V}}_{1,G} + \hat{\mathbf{V}}_{2,G} \\ \hat{\mathbf{V}}_{1,G} &:= \text{diag}\left(\hat{\mathbf{V}}_{1,G}(d) : d \in \mathcal{D}\right) \\ \hat{\mathbf{V}}_{2,G} &:= \left[\hat{\mathbf{V}}_{2,G}(d, d')\right]_{d,d' \in \mathcal{D}} , \end{aligned}$$

with

$$\begin{aligned} \hat{\mathbf{V}}_{1,G}(d) &:= \hat{\sigma}_G^2(d) - \hat{\rho}_G(d,d) \\ \hat{\mathbf{V}}_{2,G}(d,d') &:= \frac{1}{|\mathcal{D}|} \hat{\rho}_G(d,d') , \end{aligned}$$

where

$$\begin{aligned} \hat{\rho}_G(d,d) &:= \frac{2}{G} \sum_{1 \le j \le \lfloor G/2 \rfloor} \Big(\sum_{g \in \lambda_{2j-1}} \hat{Y}_g I\{D_g = d\} \Big) \Big(\sum_{g \in \lambda_{2j}} \hat{Y}_g I\{D_g = d\} \Big) \\ \hat{\rho}_G(d,d') &:= \frac{1}{G} \sum_{1 \le j \le G} \Big(\sum_{g \in \lambda_j} \hat{Y}_g I\{D_g = d\} \Big) \Big(\sum_{g \in \lambda_j} \hat{Y}_g I\{D_g = d'\} \Big) \text{ if } d \ne d' \\ \hat{\sigma}_G^2(d) &:= \frac{1}{G} \sum_{1 \le g \le J_G} \hat{Y}_g^2 I\{D_g = d\} . \end{aligned}$$

Suppose Assumptions C.1–C.4 hold, then consistency of our variance estimator follows by adapting the arguments from Bai et al. (2023b) to the proof of Theorem 3.3.

Lemma C.1. *Suppose Assumptions C.1–C.3 holds. Define*

$$\begin{aligned} \mathbb{L}_G^{\text{YN}}(d) &= \frac{1}{\sqrt{G}} \sum_{1 \le g \le J_G} (\bar{Y}_g(d) N_g I\{D_g = d\} - E[\bar{Y}_g(d) N_g] I\{D_g = d\}) \\ \mathbb{L}_G^{\text{N}}(d) &= \frac{1}{\sqrt{G}} \sum_{1 \le g \le J_G} (N_g I\{D_g = d\} - E[N_g] I\{D_g = d\}) . \end{aligned}$$

Then, as $G \to \infty$,

$$((\mathbb{L}_G^{\text{YN}}(d), \mathbb{L}_G^{\text{N}}(d)) : d \in \mathcal{D})' \xrightarrow{d} N(0, \mathbb{V}) ,$$

where

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

for

$$\mathbb{V}_1 = \text{diag}(\mathbb{V}_1^d : d \in \mathcal{D})$$

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

$$\mathbb{V}_2 = \frac{1}{|\mathcal{D}|} \operatorname{Var}[((E[\bar{Y}_g(d) N_g | X_g], E[N_g | X_g]) : d \in \mathcal{D})'] \ .$$

PROOF. The proof is omitted, but follows similarly to previous results using arguments from the proofs of Theorem 3.1 in Bai et al. (2023b) and Lemma A.2. ■

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., Jiang, L., Romano, J. P., Shaikh, A. M. and Zhang, Y. (2023a). Covariate Adjustment in Experiments with Matched Pairs. ArXiv:2302.04380 [econ], URL http://arxiv.org/abs/2302.04380.

Bai, Y., Liu, J. and Tabord-Meehan, M. (2023b). 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*.

Cytrynbaum, M. (2023). Covariate adjustment in stratified experiments.

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*.

Donner, A. and Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*. London: Arnold.

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

Hansen, B. E. and Lee, S. (2019). Asymptotic theory for clustered samples. *Journal of econometrics*, **210** 268–290.

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.

Middleton, J. A. and Aronow, P. M. (2015). Unbiased estimation of the average treatment effect in cluster-randomized experiments. *Statistics, Politics and Policy*, **6** 39–75.

Negi, A. and Wooldridge, J. M. (2021). Revisiting regression adjustment in experiments with heterogeneous treatment effects. *Econometric Reviews*, **40** 504–534.

Schochet, P. Z., Pashley, N. E., Miratrix, L. W. and Kautz, T. (2021). Design-based ratio estimators and central limit theorems for clustered, blocked rcts. *Journal of the American Statistical Association* 1–12.

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.

Wang, B., Park, C., Small, D. S. and Li, F. (2022). Model-robust and efficient inference for cluster-randomized experiments. *arXiv preprint arXiv:2210.07324*.