Design-Based Ratio Estimators and Central Limit Theorems for Clustered, Blocked RCTs

May 2020

Peter Z. Schochet, Ph.D. (Corresponding Author)
Senior Fellow, Mathematica
P.O. Box 2393
Princeton, NJ 08543-2393
Phone: (609) 936-2783
pschochet@mathematica-mpr.com

Nicole E. Pashley
Department of Statistics
Harvard University
1 Oxford St
Cambridge, MA 02138
npashley@g.harvard.edu

Luke W. Miratrix, Ph.D.
Graduate School of Education
Harvard University
14 Appian Way
Cambridge, MA 02138
luke_miratrix@gse.harvard.edu

Tim Kautz, Ph.D.
Senior Researcher, Mathematica
P.O. Box 2393
Princeton, NJ 08543-2393
Phone: (609) 297-4544
tkautz@mathematica-mpr.com

Abstract

This article develops design-based ratio estimators for clustered, blocked randomized controlled trials (RCTs), with an application to a federally funded, school-based RCT testing the effects of behavioral health interventions. We consider finite population weighted least squares estimators for average treatment effects (ATEs), allowing for general weighting schemes and covariates. We consider models with block-by-treatment status interactions as well as restricted models with block indicators only. We prove new finite population central limit theorems for each block specification. We also discuss simple variance estimators that share features with commonly used cluster-robust standard error estimators. Simulations show that the design-based ATE estimator yields nominal rejection rates with standard errors near true ones, even with few clusters.

Keywords: Randomized controlled trials; clustered designs; blocked designs; design-based estimators; finite population central limit theorems

1. Introduction

There is a growing literature on design-based methods for analyzing randomized controlled trials (RCTs) (e.g., Yang and Tsiatis 2001; Freedman (2008); Schochet 2010, 2016; Lin 2013; Miratrix et al. 2013; Imbens and Rubin 2015; Middleton and Aronow 2015; Li and Ding 2017). These nonparametric methods are built on the potential outcomes framework, introduced by Neyman (1923) and later developed in seminal works by Rubin (1974, 1977) and Holland (1986). They leverage a fundamental component of experimental designs—the known treatment assignment mechanism—to achieve results that rely on minimal assumptions.

The design-based literature has largely focused on non-clustered designs in which individuals are randomly assigned to research conditions. A much smaller literature has considered design-based methods for clustered RCTs where groups (such as schools, hospitals, or communities) rather than individuals are randomized. Clustered designs are common in evaluations that test interventions targeted to a group and are sometimes preferred to non-clustered designs as they can help minimize bias due to the potential spillover of intervention effects from treatment to control subjects. Clustered designs are becoming increasingly prevalent in social policy research (Schochet 2008) and have grown exponentially in medical trials (Bland 2004).

For example, the evaluation of the Social and Character Development (SACD) Research Program was a major federal initiative, co-funded by the Institute of Education Sciences at the U.S. Department of Education and the Centers for Disease Control and Prevention, to test interventions promoting positive social and character development among elementary school children, with the goal of ultimately improving their academic performance (SACD Consortium 2010). The study was conducted in seven large school districts (blocks), where half the schools (clusters) within each district were randomly assigned to a treatment group and half to a control group, yielding a final sample of 84 schools (42 treatment and 42 control). Intervention features included materials and lessons on social skills, behavior management, social and emotional learning, self-control, anger management, and violence prevention.

Several key aspects of the SACD study motivate the theory underlying this article. First, in each of the seven volunteer SACD districts, a convenience sample of between 10 and 14 schools was recruited for the study. This suggests a finite population framework for estimating average treatment effects (ATEs), where the sample and their potential outcomes are considered fixed, with treatment assignments being the only source of randomness. Second, flexible weighting schemes are needed to help adjust for data item nonresponse and to accommodate decisions on how clusters and blocks are to be weighted to estimate pooled effects. Third, the theory should address common approaches for including block effects in the estimation models. Finally, the estimation strategy should allow for the inclusion of baseline covariates to improve precision; this is especially important for clustered designs where power is often a concern due to design effects from clustering and the typical high cost of adding clusters to the study.

We achieve these objectives in this work by developing covariate-adjusted design-based methods for obtaining point estimates and associated inference for clustered RCTs. Our results rely on new finite population central limit theorems (CLTs) for design-based ATE ratio estimators that apply to the general case where randomization of clusters is conducted within blocks (strata). Ratio estimators for clustered RCTs have intuitive appeal because they parallel differences-in-means estimators for non-clustered designs.

Our methods allow for general weighting schemes and the inclusion of baseline covariates to improve precision. We consider models with block-by-treatment status interactions as well as models with block fixed effects only (a common specification that yields biased but more precise

ATE estimates than our primary consistent estimator). For each specification, we show how the estimated ATE coefficient from a weighted least squares (WLS) regression is a ratio estimator subject to our finite population CLTs. The technical results of our CLTs and the subsequent design-based versions of the WLS approach are the primary contributions of this work.

We provide consistent variance estimators and compare them, both analytically and through simulations, to widely used ordinary least squares estimators with cluster-robust standard errors (CRSE) (Liang and Zeger 1986; Cameron and Miller 2011). Our simulations suggest that the design-based ratio estimators yield Type 1 error rates near nominal levels, even with relatively few clusters. We also conduct an empirical analysis using data from the SACD study to compare different specifications of our estimators to each other and to the standard CRSE estimator.

This article is structured as follows. Section 2 discusses the literature this work is built on and Section 3 provides our theoretical framework. Sections 4 and 5 present our finite population CLTs and variance estimators. Section 6 presents simulation results and Section 7 presents empirical results using our motivating SACD example. Section 8 concludes.

2. Related Work

Our finite population CLTs build on Li and Ding (2017), who consider CLTs for unbiased estimators for clustered RCTs using the Horvitz-Thompson estimator developed by Middleton and Aronow (2015), but not for ratio estimators or blocked designs with general weighting schemes. Our theory also builds on results in Scott and Wu (1981) who consider CLTs for ratio estimators for finite population totals, but not for clustered designs or RCTs. We extend the design-based results in Imai et al. (2009) who examine clustered RCTs with pairwise matching but not general blocked designs, models with covariates, or CLTs. We also extend the design-based results in Schochet (2013) who examines clustered designs without blocking, and Pashley and Miratrix (2017) who consider blocked designs without clustering.

Other literature in this area has a different focus. Abadie et al. (2017) discuss reasons for adjusting for clustering and investigate differences between the true asymptotic finite population variance and the CRSE variance estimator, but do not consider impact estimation. Hansen and Bowers (2009) propose model-assisted estimators combined with randomization inference for regression models in a specific context without deriving design-based estimators. Samii and Aronow (2012) compare design-based and robust estimators for non-clustered designs, but not for clustered designs or models with covariates. While there is a large statistical literature on related design-based methods for analyzing survey data with complex sample designs (e.g. Fuller 1975, 2009; Cochran 1977; Rao and Shao 1999; Wolter 2007; Lohr 2009), these works do not focus on RCT settings.

3. Framework and Definitions

We assume that a clustered RCT of m total clusters is conducted across h blocks, with block b having m_b clusters ($b = 1, \ldots, h$). Randomization of clusters is conducted separately by block, with $m_b^1 = m_b p_b$ assigned to the treatment group and $m_b^0 = m_b(1 - p_b)$ assigned to the control group ($0 < p_b < 1$). We assume a sample of n_{jb} individuals in cluster j in block b, with n_b individuals in the block and n individuals in total. For each cluster, either all individuals are treated or not. We index individuals by ijb for individual i in cluster j in block b. Let $Y_{ijb}(1)$ be a person's outcome if assigned to a treated cluster, and $Y_{ijb}(0)$ be the outcome in a control cluster. These potential outcomes can be continuous, binary, or discrete. Let T_{jb} equal 1 if cluster jb is randomly assigned to the treatment condition and 0 otherwise. Let $S_{ijb,s}$ and $S_{jb,s}$ denote indicator variables of block membership for individuals and clusters (that is, $S_{ijb,s} = 1$ or $S_{jb,s} =$

1 if the specified person or cluster belongs to block *s*). We also allow weights, with individual weights of $w_{ijb} > 0$, cluster weights of $w_{jb} = \sum_{i=1}^{n_{jb}} w_{ijb}$, and block weights of $w_b = \sum_{j=1}^{m_b} w_{jb}$.

We assume a finite population model, where the potential outcomes are assumed fixed for the study. The alternative super-population model (that underlies the CRSE estimator) assumes study samples are random draws from a broader (infinite) population (even if vaguely defined).

We assume the following two conditions that generalize those in Imbens and Rubin (2015) for the non-clustered RCT design to our context. The first is the stable unit treatment value assumption (SUTVA) (Rubin 1986):

(*C1*)*: SUTVA*: Let $Y_{ijb}(\mathbf{T}_{clus})$ denote the potential outcome for an individual given the random vector of all cluster treatment assignments, $\mathbf{T}_{clus}$. Then, if $T_{jb} = T'_{jb}$ for cluster j, then $Y_{ijb}(\mathbf{T}_{clus}) = Y_{ijb}(\mathbf{T}'_{clus})$.

SUTVA allows us to express $Y_{ijb}(\mathbf{T}_{clus})$ as $Y_{ijb}(T_{jb})$, so that an individual's potential outcomes depend only on the individual's cluster treatment assignment and not on the treatment assignments of other clusters in the sample, although the outcomes of individuals within the same cluster could be correlated. SUTVA could be more plausible for clustered designs than non-clustered designs because there are likely to be fewer meaningful interactions between sample members across clusters than within clusters. SUTVA also assumes a particular treatment unit cannot receive different forms of the treatment.

Our second condition is the randomization itself:

(*C2*)*: Randomization:* $T_{jb} \perp (Y_{ijb}(1), Y_{ijb}(0))$ for all i, j, and b. This condition is assumed to hold within each block due to the stratified RCT design.

Under SUTVA, the block ATE parameter of interest for the finite population model is

$$\beta_{1,b} = \frac{\sum_{j=1}^{m_b} w_{jb}(\bar{Y}_{jb}(1) - \bar{Y}_{jb}(0))}{\sum_{j=1}^{m_b} w_{jb}} = \bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0), \tag{1}$$

where, for $t \in \{1,0\}$, $\bar{Y}_{jb}(t) = \frac{1}{w_{jb}} \sum_{i:S_{ijb}=1}^{n_{jb}} w_{ijb} Y_{ijb}(t)$ is the weighted mean potential outcome in the treatment or control condition. Depending on the research questions of interest, the weights can be set, for example, so that intervention effects pertain to the average individual in the block ($w_{ijb} = 1$ and $w_{jb} = n_{jb}$) or the average cluster in the block ($w_{ijb} = 1/n_{jb}$ and $w_{jb} = 1$). They can also be further modified to handle various forms of data nonresponse.

The ATE parameter across all blocks is a weighted average of the block ATE parameters:

$$\beta_1 = \frac{\sum_{b=1}^{h} w_b \beta_{1,b}}{\sum_{b=1}^{h} w_b}. \tag{2}$$

4. ATE Estimators for the Finite Population Model

Under the potential outcomes framework and SUTVA, the data generating process for the observed outcome measure, y_{ijb}, is as follows:

$$y_{ijb} = T_{jb} Y_{ijb}(1) + (1 - T_{jb}) Y_{ijb}(0). \tag{3}$$

This relation states that we can observe $y_{ijb} = Y_{ijb}(1)$ for those in the treatment group and $y_{ijb} = Y_{ijb}(0)$ for those in the control group, but not both. Rearranging (3) generates the following nominal regression model for any given block:

$$y_{ijb} = \beta_{0,b} + \beta_{1,b}(T_{jb} - p_b^*) + u_{ijb}, \tag{4}$$

where $\beta_{1,b} = \bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0)$ is the block-specific ATE parameter, $p_b^* = \frac{1}{w_b} \sum_{j=1}^{m_b} T_{jb} w_{jb}$ is the weighted treatment group sampling rate, $\beta_{0,b} = p_b^* \bar{\bar{Y}}_b(1) + (1 - p_b^*) \bar{\bar{Y}}_b(0)$ is the mean potential outcome in the block, and the "error" term, u_{ijb}, can be expressed as

$$u_{ijb} = T_{jb}\left(Y_{ijb}(1) - \bar{\bar{Y}}_b(1)\right) + (1 - T_{jb})\left(Y_{ijb}(0) - \bar{\bar{Y}}_b(0)\right).$$

We center the treatment indicator in (4) to facilitate the theory without changing the estimator.

In contrast to usual formulations of the regression model, our residual, u_{ijb}, is random solely because of T_{jb} (that is, due to random assignment) (see Freedman 2008 and Lin 2013 for a more detailed discussion of this approach for non-clustered RCTs). This framework allows treatment effects to differ across individuals and clusters and is nonparametric because it makes no assumptions about the distribution of potential outcomes. Note that our model does not satisfy key assumptions of the usual regression model for correlated data: u_{ijb} does not have mean zero, u_{ijb} is heteroscedastic, $Cov(u_{ijb}, u_{ij'b})$ is not constant for individuals in the same cluster, $Cov(u_{ijb}, u_{i'j'b})$ is nonzero for individuals in different clusters, and u_{ijb} is correlated with the regressor $\left(T_{jb} - p_b^*\right)$ (see Schochet 2016).

The model in (4) can also be expressed using block indicator variables as follows:

$$y_{ijb} = \sum_{s=1}^{h} \beta_{1,s} S_{ijb,s} \tilde{T}_{js} + \sum_{s=1}^{h} \beta_{0,s} S_{ijb,s} + u_{ijb}, \tag{5}$$

where $\tilde{T}_{jb} = (T_{jb} - p_b^*)$ is the block centered treatment status indicator. Due to blocked random assignment, the errors are independent across blocks. We include terms for all h blocks in the model and exclude a grand intercept term.

For estimation, we use the following working model, a version of (5), that provides covariate-adjusted ATE estimates by including a $1xv$ vector of fixed, block-mean centered baseline covariates, $\tilde{\mathbf{x}}_{ijb}$, with associated parameter vector $\boldsymbol{\gamma}$:

$$y_{ijb} = \sum_{s=1}^{h} \beta_{1,k} S_{ijb,s} \tilde{T}_{js} + \sum_{s=1}^{h} \beta_{0,s} S_{ijb,s} + \tilde{\mathbf{x}}_{ijb} \boldsymbol{\gamma} + e_{ijb},$$

where $\tilde{\mathbf{x}}_{ijb} = (\mathbf{x}_{ijb} - \bar{\bar{\mathbf{x}}}_b)$, $\bar{\bar{\mathbf{x}}}_b = \frac{1}{w_b} \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb} \mathbf{x}_{ijb}$, and e_{ijb} is the error term. These covariates, unaffected by the treatment, can be at the individual or cluster level. We assume there are sufficient degrees of freedom for variance estimation (see Section 4.2). While the covariates do not enter the true block-specific RCT models in (4) and the ATE estimands do not change, the covariates will increase precision to the extent they are correlated with the potential outcomes. Note that we do not need to assume that the true conditional distribution of y_{ijb} given $\mathbf{x}_{ijb}$ is linear in $\mathbf{x}_{ijb}$.

We do not consider models that interact $\tilde{\mathbf{x}}_{ijb}$ and $\tilde{T}_{jb}$ due to associated degrees of freedom losses that can seriously reduce the power of clustered RCTs that, in practice, often contain relatively few clusters for cost reasons. Similarly, we pool $\boldsymbol{\gamma}$ across blocks. Our $\boldsymbol{\gamma}$ parameter is well defined: it is the finite population regression coefficient that would be obtained if we could run the weighted regression on the full schedule of potential outcomes.

Using individual-level data, we can fit this working model using weighted least squares (WLS) with weights w_{ijb}. This yields the following closed-form WLS estimator for $\beta_{1,b}$ (see Supplementary Materials A for the derivation):

$$\hat{\beta}_{1,b} = \frac{1}{w_b^1}\sum_{j:T_{jb}=1}^{m_b} w_{jb}\bar{y}_{jb} - \frac{1}{w_b^0}\sum_{j:T_{jb}=0}^{m_b} w_{jb}\bar{y}_{jb} - \left(\frac{1}{w_b^1}\sum_{j:T_{jb}=1}^{m_b} w_{jb}\bar{\mathbf{x}}_{jb} - \frac{1}{w_b^0}\sum_{j:T_{jb}=0}^{m_b} w_{jb}\bar{\mathbf{x}}_{jb}\right)\hat{\boldsymbol{\gamma}} \tag{6}$$

$$= \bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) - (\bar{\bar{\mathbf{x}}}_b^1 - \bar{\bar{\mathbf{x}}}_b^0)\,\hat{\boldsymbol{\gamma}},$$

where, for $t \in \{1,0\}$, $\bar{\bar{y}}_b(t)$ is the weighted average of the observed outcome across subjects in the treatment or control group, $w_b^t = \sum_{j:T_{jb}=t}^{m_b} w_{jb}$ is the sum of the weights, and $\hat{\boldsymbol{\gamma}}$ is the WLS parameter estimate for $\boldsymbol{\gamma}$.

4.1. Theoretical Results

To consider the asymptotic properties of $\hat{\beta}_{1,b}$, we consider a hypothetical increasing sequence of finite populations where $m_b \to \infty$ in each block, so that $m = \sum_{b=1}^{h} m_b \to \infty$. The number of blocks, h, however, remains fixed. In principle, parameters should be subscripted by m, but we omit this notation for simplicity. We further assume that the proportion of all clusters in a block converges to a constant, that is, $m_b/m \rightarrow q_b$ as $m \rightarrow \infty$, where $0 < q_b < 1$. We finally assume that p_b is (approximately) constant as $m \rightarrow \infty$, so that the number of treated and control clusters in each block increases with m (that is, $m_b^1 \rightarrow \infty$ and $m_b^0 \rightarrow \infty$).

Given this framework, we present a CLT for the WLS estimator that provides design-based standard errors and associated inference. Before presenting our theorem, we first need to define several quantities pertaining to finite population variances and covariances. First, for $t \in \{1,0\}$, we define $D_b(t) = \frac{w_{jb}}{\overline{w}_b}\left(\overline{Y}_{jb}(t) - \overline{\overline{Y}}_b(t) - \left(\overline{\mathbf{x}}_{jb} - \overline{\overline{\mathbf{x}}}_b\right)\boldsymbol{\gamma}\right)$ as the residualized potential outcomes at the cluster level in the treatment and control conditions. Second, we define $S_{D_b}^2(t)$ as the variance of these residuals,

$$S_{D_b}^2(t) = \frac{1}{m_b - 1}\sum_{j=1}^{m_b}\frac{w_{jb}^2}{\overline{w}_b^2}\left(\overline{Y}_{jb}(t) - \overline{\overline{Y}}_b(t) - \left(\overline{\mathbf{x}}_{jb} - \overline{\overline{\mathbf{x}}}_b\right)\boldsymbol{\gamma}\right)^2,$$

and $S_{D_b}^2(1,0)$ as the associated treatment-control covariance,

$$S_{D_b}^2(1,0) = \frac{1}{m_b - 1}\sum_{j=1}^{m_b}\frac{w_{jb}^2}{\overline{w}_b^2}\left(\overline{Y}_{jb}(1) - \overline{\overline{Y}}_b(1) - (\overline{\mathbf{x}}_{jb} - \overline{\overline{\mathbf{x}}}_b)\boldsymbol{\gamma}\right)\left(\overline{Y}_{jb}(0) - \overline{\overline{Y}}_b(0) - (\overline{\mathbf{x}}_{jb} - \overline{\overline{\mathbf{x}}}_b)\boldsymbol{\gamma}\right).$$

Third, we define $\mathrm{Var}(\hat{D}_b)$ as the variance of the mean difference in residuals between the observed (randomized) treatment and control group samples,

$$\mathrm{Var}\left(\hat{D}_b\right) = \frac{S_{D_b}^2(1)}{m_b^1} + \frac{S_{D_b}^2(0)}{m_b^0} - \frac{S^2(D_b)}{m_b}, \tag{8}$$

where $S^2(D_b)$ is the variance (heterogeneity) of the ATEs of the clusters in block b,

$$S^2(D_b) = \frac{1}{m_b - 1}\sum_{j=1}^{m_b}\frac{w_{jb}^2}{\overline{w}_b^2}\left(\overline{Y}_{jb}(1) - \overline{Y}_{jb}(0) - \left(\overline{\overline{Y}}_b(1) - \overline{\overline{Y}}_b(0)\right)\right)^2.$$

Fourth, we define the variance of the weights as $S^2(w_b) = \frac{1}{m_b-1}\sum_{j=1}^{m_b}(w_{jb} - \overline{w}_b)^2$, where $\overline{w}_b = \frac{1}{m_b}\sum_{j=1}^{m_b} w_{jb}$. Fifth, we need the weighted variances, $S^2_{x_b,k}$, of each covariate k,

$$S^2_{x_b,k} = \frac{1}{m_b - 1}\sum_{j=1}^{m_b} \frac{w_{jb}^2}{\overline{w}_b^2}\left([\overline{\mathbf{x}}_{jb} - \overline{\overline{\mathbf{x}}}_b]_k\right)^2,$$

and the weighted variance-covariance matrix of the covariates with themselves, $\mathbf{S}^2_{\mathbf{x},b}$, which is analogous to the classic $\mathbf{X}'\mathbf{W}\mathbf{X}$ matrix in WLS,

$$\mathbf{S}^2_{\mathbf{x},b} = \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\left(\mathbf{x}_{ijb} - \overline{\overline{\mathbf{x}}}_b\right)'\left(\mathbf{x}_{ijb} - \overline{\overline{\mathbf{x}}}_b\right).$$

Finally, we need two matrices that are analogous to the classic $\mathbf{X}'\mathbf{W}\mathbf{Y}$ matrix in WLS,

$$\mathbf{S}^2_{\mathbf{x},Y,b}(t) = \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\,\mathbf{x}'_{ijb}Y_{ijb}(t) - \overline{\overline{\mathbf{x}}}'_b\overline{wY(t)}_b$$

and

$$\mathbf{S}^2_{\mathbf{x}Y,b}(t) = \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}}\left(w_{ijb}\mathbf{x}'_{ijb}\mathbf{x}_{ijb}Y_{ijb}(t) - \overline{wxY(t)}_b\right)^2,$$

where $\overline{wY(t)}_b = \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\,Y_{ijb}(t)$ and $\overline{wxY(t)}_b = \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\,\mathbf{x}'_{ijb}Y_{ijb}(t)$.

We now present our CLT theorem, proved in Supplementary Materials A, which adapts finite population CLT results in Li and Ding (2017) and Scott and Wu (1981) to our setting.

Theorem 1. Assume (*C1*) and (*C2*) and the following conditions for $t \in \{1,0\}$:

(*C3*) Letting $g_b(t) = \max_{1\le j\le m_b}\left(w_{jb}\left(\overline{Y}_{jb}(t) - \overline{\overline{Y}}_b(t) - \left(\overline{\mathbf{x}}_{jb} - \overline{\overline{\mathbf{x}}}_b\right)\boldsymbol{\gamma}\right)/\overline{w}_b\right)^2$, as $m \to \infty$,

$$\max_{t\in\{0,1\}}\frac{1}{(m_b^t)^2}\frac{g_b(t)}{\mathrm{Var}(\hat{D}_b)} \to 0.$$

(*C4*) $f_b^t = m_b^t/m_b$ has a limiting value in $(0,1)$ and $S^2_{D_b}(t)$, and $S^2_{D_b}(1,0)$ also have finite limiting values.

($C5$) As $m \to \infty$,

$$(1 - f_b^t)\frac{S^2(w_b)}{m_b^t \overline{w}_b^2} \to 0.$$

($C6$) Letting $h_{b,k}(t) = \max_{1 \le j \le m_b}\left(w_{jb}\left([\overline{x}_{jb} - \overline{\overline{\mathrm{x}}}_b]_k\right)/\overline{w}_b\right)^2$ for all b and k, as $m \to \infty$,

$$\frac{1}{\min(m_b^1, m_b^0)}\frac{h_{b,k}(t)}{S_{x_b,k}^2} \to 0.$$

($C7$) $S_{x_b,k}^2$, $\mathbf{S}_{\mathbf{x},b}^2$, $\mathbf{S}_{\mathbf{x},Y,b}^2(t)$, and $\mathbf{S}_{\mathbf{x}Y,b}^2(t)$ have finite (positive definite) limiting values.

Then, as $m \to \infty$, $\hat{\beta}_{1,b}$ is a consistent estimator for $\beta_{1,b}$ and

$$\frac{\hat{\beta}_{1,b} - \left(\overline{\overline{Y}}_b(1) - \overline{\overline{Y}}_b(0)\right)}{\sqrt{\mathrm{Var}\left(\hat{D}_b\right)}} \xrightarrow{d} N(0,1),$$

where $\mathrm{Var}(\hat{D}_b)$ is defined as in (8).

Remark 1. Condition ($C3$) allows us to invoke the CLT in Theorem 4 of Li and Ding (2017) that underlies our finite population CLT (see Supplementary Materials A for details). ($C4$) ensures the stability of the treatment assignment probabilities in each block and also ensures limiting values of asymptotic variances and covariances of the residualized potential outcomes. ($C5$) provides a weak law of large numbers for the weights so that $\overline{w}_b^t/\overline{w}_b \xrightarrow{p} 1$, where $\overline{w}_b^t = \frac{1}{m_b^t}\sum_{j:T_j=t}^{m_b} w_{jb}$. *(C7)* specifies limiting values of the covariate variances and outcome-covariate covariances, which in turn, provide regularity conditions on $\hat{\boldsymbol{\gamma}}$.

Remark 2. The above theorem is proved as a two-step process. We first assume $\boldsymbol{\gamma}$ is known and obtain a CLT with this known parameter. We then show $\hat{\boldsymbol{\gamma}}$ converges to $\boldsymbol{\gamma}$ and use *(C6)* to ensure that the vector of covariate means, $(\overline{\overline{\mathbf{x}}}_b^1 - \overline{\overline{\mathbf{x}}}_b^0)$, is asymptotically normal with zero mean, so that the ATE estimator still converges to a standard normal.

Remark 3. The first two terms in (8) pertain to separate variances for the treatment and control groups because we allow for heterogeneous treatment effects. The third term pertains to the covariance of cluster-level average potential outcomes in the treatment and control conditions, $S^2_{D_b}(1,0)$, that we express in terms of the heterogeneity of treatment effects across clusters, $S^2(D_b)$. We hereafter label this term the "finite population heterogeneity" term. It cannot be identified from the data but can be bounded (as discussed in Section 4.2).

Remark 4. Under (*C1*)-(*C5*), Theorem 2 also applies to models without covariates by setting $\boldsymbol{\gamma} = \mathbf{0}$, yielding a simple differences-in-means ATE ratio estimator, $\hat{\beta}_{1,b} = \bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0)$. We discuss the finite sample bias of this estimator in Supplementary Materials A.

Corollary 1. Under the conditions of Theorem 1 and assuming $\overline{w}_b = \frac{w_b}{m_b}$ has a finite limit for all b, the pooled ATE estimator across blocks, $\hat{\beta}_1 = \frac{1}{h\overline{w}} \sum_{b=1}^{h} w_b \hat{\beta}_{1,b}$, is consistent for β_1 in (2) and $\frac{1}{\sqrt{\text{Var}(\widehat{D})}}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0,1)$, where $\text{Var}(\widehat{D})$= $\frac{1}{(h\overline{w})^2} \sum_{b=1}^{h} w_b^2 Var(\widehat{D}_b)$ and $\overline{w} = \frac{1}{h} \sum_{b=1}^{h} w_b$.

This result follows because the $\hat{\beta}_{1,b}$estimators are asymptotically independent.

4.2. Variance Estimation

We can estimate the block-specific variance in (8) with a consistent (upper bound) plug-in variance estimator based on the regression residuals averaged to the cluster level as follows:

$$\widehat{\text{Var}}\left(\hat{D}_b\right) = \frac{s^2_{D_b}(1)}{m_b^1} + \frac{s^2_{D_b}(0)}{m_b^0}, \tag{9}$$

where

$$s^2_{D_b}(1) = \frac{1}{(m_b^1 - v^* p_b^* q_b^* - 1)} \sum_{j:S_{jb}=1, T_{jb}=1}^{m_b^1} \frac{w_{jb}^2}{(\overline{w}_b^1)^2} \left(\bar{y}_{jb} - \hat{\beta}_{0,b} - (1 - p_b^*)\hat{\beta}_{1,b} - \bar{\bar{\mathbf{x}}}_{jb}\hat{\boldsymbol{\gamma}}\right)^2,$$

$$s^2_{D_b}(0) = \frac{1}{(m_b^0 - v^*(1 - p_b^*) q_b^* - 1)} \sum_{j:S_{jb}=1, T_{jb}=0}^{m_b^0} \frac{w_{jb}^2}{(\overline{w}_b^0)^2} \left(\bar{y}_{jb} - \hat{\beta}_{0,b} + p_b^*\hat{\beta}_{1,b} - \bar{\bar{\mathbf{x}}}_{jb}\hat{\boldsymbol{\gamma}}\right)^2,$$

$q_b^* = \frac{w_b}{\sum_{b=1}^{h} w_b}$ is the weighted share of all clusters in block b, and v^* is a degrees of freedom adjustment for the covariates. Plausible values for v^* are $v^* = 0$ or $v^* = v$ (the number of covariates) (Donald and Lang 2007), although other approaches have been proposed, such as adjusting individual sample sizes for design effects due to clustering (Hedges 2007) and minimum distance methods (Wooldridge 2006).

In our simulations (see Section 6), we also use two variants of (9). First, we multiply (9) by $(1 - R_{TXb}^2)^{-1}$, where R_{TXb}^2 is the R-squared value from a regression of $S_{ijb,s}\tilde{T}_{jb}$ on $\tilde{\mathbf{x}}_{ijb}$ and the other block-by-treatment status interactions in (5) (with no intercept). This term captures the finite sample collinearity between $\tilde{T}_{jb}$ and $\tilde{\mathbf{x}}_{ijb}$(which inflates the variances). This estimator performs well in our simulations. The second variant subtracts $\frac{1}{m_b}\left(\sqrt{s_{D_b}^2(1)} - \sqrt{s_{D_b}^2(0)}\right)^2$, a lower bound on the finite population heterogeneity term based on the Cauchy-Schwarz inequality. Aronow et al. (2014) discuss sharper bounds on this heterogeneity term by approximating the marginal distributions of potential outcomes.

We note a few features of (9). First, the same variance estimator applies when using non-centered data in the regressions instead of centered data. Second, the covariates will only affect the ATE estimates and increase precision if mean covariate values vary across clusters. Third, the estimator pertains to continuous, binary, and discrete outcomes (and covariates). Finally, the model can be estimated using data averaged to the cluster level, so that setting $v^* = v$ yields well-defined degrees of freedom. Estimators based on the individual and aggregated data will be identical for models without covariates and for models with cluster-level covariates only; for models with individual-level covariates, power losses using the aggregated data are likely to be tolerable with relatively small numbers of covariates (see Section 6 below and Schochet 2020).

We can obtain pooled ATE estimators across all blocks by inserting $\hat{\beta}_{1,b}$ into (2) and using $\frac{1}{(h\bar{w})^2}\sum_{b=1}^{h} w_b^2 V\hat{a}r(\widehat{D}_b)$ for variance estimation. Hypothesis testing can be conducted using z-tests. Alternatively, results in Bell and McCaffery (2002), Hansen (2007), and Cameron and Miller (2015) for the CRSE estimator suggest that t-tests with $(m-2h-v^*)$ degrees of freedom perform better in small samples and is what we use hereafter.

4.3 Comparing Design-Based and CRSE Estimators

The CRSE variance estimator is an extension of robust standard errors (Huber 1967; White 1980) to clustered designs (Liang and Zeger 1986). The CRSE approach, which assumes *iid* sampling of units from some (infinite) super-population, allows for errors to be correlated within clusters but not across clusters. Using individual-level data, the CRSE variance estimator for the WLS coefficients in (5) that includes baseline covariates is:

$$V\hat{a}r_{CRSE}(\widehat{\boldsymbol{\delta}}) = g(\mathbf{Z}'\mathbf{W}\mathbf{Z})^{-1}\left(\sum_{b=1}^{h}\sum_{j=1}^{m_b}\mathbf{Z}_{jb}'\,\mathbf{W}_{jb}\hat{\mathbf{e}}_{jb}\hat{\mathbf{e}}_{jb}'\mathbf{W}_{jb}\mathbf{Z}_{jb}\right)(\mathbf{Z}'\mathbf{W}\mathbf{Z})^{-1}, \tag{10}$$

where $\mathbf{Z}$ is an $nx(2h+v)$ matrix with the full set of independent variables (the block-by-treatment status interactions, block indicators, and $\tilde{\mathbf{x}}_{ijb}$ covariates); $\widehat{\boldsymbol{\delta}}$ is the corresponding vector of coefficient estimates; $\mathbf{W}$ is an nxn symmetric weight matrix with diagonal entries w_{ijb} and all other entries 0; $\mathbf{W}_{jb}$ are analogous $n_{jb}xn_{jb}$ matrices defined for each cluster; $\hat{\mathbf{e}}_{jb}$ is a vector of WLS residuals for each cluster; and g is a small sample correction term discussed below. CRSE estimators are asymptotically normal (Liang and Zeger 1986). Hypothesis testing is commonly conducted using t-tests with $(m-1)$ degrees of freedom (Cameron and Miller 2015).

As with the design-based estimators, the CRSE estimator is based on WLS using the pooled data across blocks. Therefore, both approaches yield the same ATE estimate when the covariates and weights are the same, but the variance estimators differ in several ways. To illustrate the key differences, consider the model without covariates. In that case, the CRSE variance estimator in (10) for a single block-by-treatment ATE estimate, $\hat{\beta}_{1,b}$, reduces to

$$V\hat{a}r_{CRSE}(\hat{\beta}_{1,b}) = g\,\frac{s_{D_b}^{2*}(1)}{m_b^1} + g\,\frac{s_{D_b}^{2*}(0)}{m_b^0}, \qquad (11)$$

where $s_{D_b}^{2*}(1) = \frac{(m_b^1-1)}{m_b^1} s_{D_b}^2(1)$ and $s_{D_b}^{2*}(0) = \frac{(m_b^0-1)}{m_b^0} s_{D_b}^2(0)$. Here, we use $g = \left(\frac{m}{m-1}\right)\left(\frac{n-1}{n-2h-v}\right)$, a common value in statistical software packages such as Stata (Cameron and Miller 2015), although other approaches have been proposed, such as bias-corrected CRSE estimators (Mackinnon and White 1985; Bell and McCaffery 2002; Angrist and Lavy 2002; Pustejovsky and Tipton 2018) and bootstrap methods (Cameron et al. 2008; Webb 2013) that can adjust for the known Type 1 error inflation of the CRSE estimator in small samples.

Compare (11) to the following parallel expression for the design-based estimator in (9):

$$\mathrm{V\hat{a}r}\left(\hat{D}_b\right) = \frac{m_b^1}{(m_b^1-1)}\frac{s_{D_b}^{2*}(1)}{m_b^1} + \frac{m_b^0}{(m_b^0-1)}\frac{s_{D_b}^{2*}(0)}{m_b^0}. \qquad (12)$$

There are two key differences between (11) and (12) that pertain to the degrees of freedom adjustments. First, the adjustments for the design-based variance estimator are applied separately for treatments and controls based on m_b^1 and m_b^0, whereas the standard CRSE estimator applies a single adjustment, g, based on total sample sizes (m and n). Second, the design-based estimator uses $(m-2)$ degrees of freedom for the t-tests, reflecting separation of the two research groups, whereas the CRSE estimator commonly uses $(m-1)$. These two differences will typically lead to larger design-based variances and lower rejection rates (yielding Type 1 errors closer to nominal levels as shown later in our simulations). Note also that the finite population heterogeneity term does not apply to the CRSE estimator as it assumes a super-population sampling framework. Similar issues apply to models with covariates.

5. Restricted ATE Estimators with Fixed Block Effects Only

A commonly used estimation strategy for blocked designs is to include block indicator variables in the regression model but to exclude block-by-treatment status interaction terms:

$$y_{ijb} = \beta_{1,R}\tilde{T}_{jb} + \sum_{s=1}^{h} \delta_{0,s} S_{ijb,s} + \epsilon_{ijb}, \tag{13}$$

where ϵ_{ijb} is the error term. Because this framework imposes restrictions on the assumed data structure, it typically produces asymptotically biased estimates of the true ATE parameter in (2). Nevertheless, it has practical appeal due to its parsimony and additional degrees of freedom.

Consider WLS estimation of (13) where the model includes the $\tilde{\mathbf{x}}_{ijb}$ covariates with parameter vector $\boldsymbol{\gamma}$. As shown in Supplementary Materials A, the WLS estimator for $\beta_{1,R}$ is a weighted average of block-level ATE estimates with weights, $\tilde{w}_{b,R} = \frac{1}{m w_b} w_b^0 w_b^1$:

$$\hat{\beta}_{1,R} = \sum_{b=1}^{h} \frac{\tilde{w}_{b,R}}{\sum_{a=1}^{h} \tilde{w}_{a,R}} \left(\bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) - \left(\bar{\bar{\mathbf{x}}}_b^1 - \bar{\bar{\mathbf{x}}}_b^0 \right) \hat{\boldsymbol{\gamma}} \right). \tag{14}$$

Note that the weights can also be expressed in the limit as $\tilde{w}_{b,R} = q_b p_b (1 - p_b) \overline{w}_b$, where $q_b = m_b / m$. Thus, this approach uses a form of precision weighting to weight the block-specific treatment effects and is analogous to a fixed effects regression model using non-clustered data.

5.1. Theoretical Results

We now present a CLT for $\hat{\beta}_{1,R}$ that is proved in Supplementary Materials A. Let

$$\beta_{1,R} = \sum_{b=1}^{h} \frac{q_b p_b (1 - p_b) \overline{w}_b}{\sum_{b=1}^{h} q_b p_b (1 - p_b) \overline{w}_b} \left(\bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0) \right)$$

denote the treatment effect parameter for the restricted model. Also define the vector of block-level estimators as

$$\mathbf{t} = (\overline{w}_1^1(\bar{\bar{y}}_1(1) - \bar{\bar{\mathbf{x}}}_1^1 \boldsymbol{\gamma}), \overline{w}_1^0(\bar{\bar{y}}_1(0) - \bar{\bar{\mathbf{x}}}_1^0 \boldsymbol{\gamma}), \overline{w}_1^1, \cdots, \overline{w}_1^h(\bar{\bar{y}}_h(1) - \bar{\bar{\mathbf{x}}}_h^1 \boldsymbol{\gamma}), \overline{w}_h^0(\bar{\bar{y}}_h(0) - \bar{\bar{\mathbf{x}}}_h^0 \boldsymbol{\gamma}), \overline{w}_h^1).$$

Theorem 2. Assume (*C1*), (*C2*), *(C4)* for f_b^t, (*C5*), (*C6*), (*C7*) and the following conditions for $t \in \{1,0\}$:

(*C8*) As $m \to \infty$,

$$\max_{1\le b\le h}\max_{t\in\{0,1\}}\frac{a_{Y,b}(t)}{p_b(1-p_b)m_b v_{Y,b}(t)} \to 0, \text{ where}$$

$$a_{Y,b}(t) = \max_{1\le j\le m_b}\left(w_{jb}\left(\overline{Y}_{jb}(t) - \overline{\mathbf{x}}_{jb}\boldsymbol{\gamma}\right) - \overline{w}_b\left(\overline{\overline{Y}}_b(t) - \overline{\overline{\mathbf{x}}}_b\boldsymbol{\gamma}\right)\right)^2 \text{ and}$$

$$v_{Y,b}(t) = \frac{1}{m_b-1}\sum_{j=1}^{m_b}\left(w_{jb}\left(\overline{Y}_{jb}(t) - \overline{\mathbf{x}}_{jb}\boldsymbol{\gamma}\right) - \overline{w}_b\left(\overline{\overline{Y}}_b(t) - \overline{\overline{\mathbf{x}}}_b\boldsymbol{\gamma}\right)\right)^2.$$

(*C9*) As $m \to \infty$,

$$\max_{1\le b\le h}\frac{a_{w,b}}{p_b(1-p_b)m_b v_{w,b}} \to 0,$$

where $a_{w,b} = \max_{1\le j\le m_b}\left(w_{jb} - \overline{w}_b\right)^2$ and $v_{w,b} = \frac{1}{m_b-1}\sum_{j=1}^{m_b}\left(w_{jb} - \overline{w}_b\right)^2$.

(*C10*) The correlation matrix of $\mathbf{t}$ has a finite limiting value $\boldsymbol{\Sigma}$.

(*C11*) The variance expressions, $v_{w,b}$ and $v_{Y,b}(t)$, have finite limiting values for $b \in \{1,\ldots,h\}$.

(*C12*) $\overline{w}_b\left(\overline{\overline{Y}}_b(1) - \overline{\overline{\mathbf{x}}}_b\boldsymbol{\gamma}\right) \neq 0$ or $\overline{w}_b\left(\overline{\overline{Y}}_b(0) - \overline{\overline{\mathbf{x}}}_b\boldsymbol{\gamma}\right) \neq 0$ for some b.

Then, as $m \to \infty$, $\hat{\beta}_{1,R}$ is a consistent estimator for $\beta_{1,R}$ and

$$\frac{\hat{\beta}_{1,R} - \beta_{1,R}}{\sqrt{\operatorname{Var}\left(\tilde{\beta}_{1,R}\right)}} \xrightarrow{d} N(0,1), \text{ where}$$

$$\begin{aligned}
&\operatorname{Var}\left(\tilde{\beta}_{1,R}\right) \\
&= \sum_{b=1}^{h}\frac{1}{m_b(m_b-1)}\frac{\left(q_b p_b(1-p_b)\overline{w}_b\right)^2}{\left(\sum_{a=1}^{h} q_a p_a(1-p_a)\overline{w}_a\right)^2}\sum_{j=1}^{m_b}\left(\sqrt{\frac{1-p_b}{p_b}}\left(\frac{w_{jb}\left(\overline{Y}_{jb}(1) - \overline{\overline{Y}}_b(1) - \left(\overline{\mathbf{x}}_{jb} - \overline{\overline{\mathbf{x}}}_b\right)\boldsymbol{\gamma}\right)}{\overline{w}_b}\right)\right. \\
&+ \sqrt{\frac{p_b}{1-p_b}}\left(\frac{w_{jb}\left(\overline{Y}_{jb}(0) - \overline{\overline{Y}}_b(0) - \left(\overline{\mathbf{x}}_{jb} - \overline{\overline{\mathbf{x}}}_b\right)\boldsymbol{\gamma}\right)}{\overline{w}_b}\right) \\
&\left.+ \frac{(1-2p_b)}{\sqrt{p_b(1-p_b)}\overline{w}_b}\left(\beta_{1,b} - \beta_{1,R}\right)\left(w_{jb} - \overline{w}_b\right)\right)^2.
\end{aligned} \tag{15}$$

Remark. The first two terms inside the brackets in (15) pertain to block-specific variances for the treatment and control groups that are analogous to the corresponding variance terms in the unrestricted model in (8). The third term represents the covariance between the block treatment effects and block weights that is induced by the restricted model. This term differentiates the variances for the restricted and unrestricted models, along with the block weights used for pooling and the presence of the finite population heterogeneity term. This third term is 0 if $p_b = 0.5$ or $\mathrm{Cov}(\beta_{1,b}, w_{jb}) = 0$, but otherwise can be positive or negative.

5.2. Variance Estimation

A consistent variance estimator for (15) can be obtained by multiplying out the squared term and using plug-in estimators for each of the resulting six terms. However, it is simpler to use the following consistent estimator based on cluster-level model residuals:

$$\widehat{\mathrm{Var}}\left(\hat{\beta}_{1,R}\right) = \frac{1}{m(m-h-v^*-1)} \frac{\sum_{b=1}^{h}\sum_{j=1}^{m_b} w_{jb}^2 \tilde{T}_{jb}^2 (\bar{y}_{jb} - \hat{\beta}_{1,R}\tilde{T}_{jb} - \hat{\delta}_{0,b} - \tilde{\bar{\bar{\mathbf{x}}}}_{jb}\hat{\boldsymbol{\gamma}})^2}{\left(\sum_{b=1}^{h} q_b p_b (1-p_b)\bar{w}_b\right)^2}, \tag{16}$$

recalling that $\tilde{T}_{jb} = (T_{jb} - p_b^*)$. Following Schochet (2016), the expression in (16) can be justified using the following standard asymptotic expansion for the WLS estimator:

$$\sqrt{m}\left(\hat{\beta}_{1,R} - \beta_{1,R}\right) = \frac{\sum_{b=1}^{h}\sum_{j=1}^{m_b} w_{jb}\tilde{T}_{jb}\left(\bar{y}_{jb} - \beta_{1,R}\tilde{T}_{jb} - \delta_{0,b} - \tilde{\bar{\bar{\mathbf{x}}}}_{jb}\boldsymbol{\gamma}\right)}{\sqrt{m}\sum_{b=1}^{h} q_b p_b (1-p_b)\bar{w}_b} + o_p(1), \tag{17}$$

where $o_p(1)$ signifies a term that converges in probability to zero. Suppose we insert into (17), $\bar{y}_{jb} = T_{bj}\bar{Y}_{jb}(1) + (1-T_{bj})\bar{Y}_{jb}(0)$ and $\delta_{0,b} = p_b\bar{\bar{Y}}_b(1) + (1-p_b)\bar{\bar{Y}}_b(0)$ (see Supplementary Materials A), and then add and subtract $\left(\beta_{1,b} - \beta_{1,R}\right)\tilde{T}_{jb}$. If we then calculate $\mathrm{Var}\left(\hat{\beta}_{1,R}\right)$ over the randomization distribution, we obtain (15) after some algebra. Hypothesis testing can be conducted using t-tests with $(\sum_{b=1}^{h}(m_b^1 + m_b^0) - v^* - h - 1)$ degrees of freedom. Note that with

$v^* = 0$, estimation requires at least 1 treatment and 1 control cluster per block rather than two as for the fully interacted model.

The CRSE estimator for the restricted model has the same form as (16). For the model without covariates, the only difference is that the standard CRSE estimator uses g for the degrees of freedom correction term. In studies with few clusters and many blocks, the design-based estimator will tend to exceed the CRSE estimator (that is, it will yield larger standard errors).

6. Simulation Results

To examine the statistical properties of the design-based estimator, we conducted simulations for a clustered, non-blocked design ($h = 1$) using the variance estimator in (9) with and without covariates, the R^2_{TX} adjustment, and the finite population heterogeneity term based on the Cauchy-Schwarz inequality. We also included the standard CRSE estimator to identify sources of differences between the two approaches (our goal is not to compare various proposed CRSE estimators to each other). Note that if $h > 1$, differences between the two approaches when examining block-level estimates will tend to be greater than the simulation results presented here, because the degrees of freedom adjustments will differ more (see (11) and (12)).

For the simulations, we created a single base dataset that included all potential outcomes and covariates, and then for each of 1,000 replications, we randomly assigned half the clusters to treatment and half to control, storing the associated outcomes. We then fit our models and recorded results. Finally, we calculated Type 1 errors across the replications and compared average empirical values of the standard errors produced by the estimators to their "true" sampling variability as measured by standard deviations of the 1,000 ATE estimates. To avoid unusual base datasets, we repeated this process for 100 base datasets and calculated average statistics. For all WLS estimations, clusters were weighted by their sample sizes.

To generate our initial full schedule of potential outcomes (see Supplementary Materials B for more details), we used the following model:

$$Y_{ij}(0) = x_{ij1} + x_{ij2} + u_j + e_{ij}$$
$$Y_{ij}(1) = Y_{ij}(0) + \theta_j T_j, \tag{19}$$

where u_j, θ_j (which captures treatment effect heterogeneity), and e_{ij} are each *iid* mean zero random errors and x_{ij1} and x_{ij2} are independent covariates. We ran separate simulations for $m =$ 8 to 50 clusters. We allowed cluster sample sizes to vary around a pre-set mean of 100 (or 40 for some runs) that were drawn to be correlated with both u_j and θ_j. For each replication, we calculated $y_{ij} = T_j Y_{ij}(1) + (1 - T_j) Y_{ij}(0)$ to generate the observed outcomes.

We examined a range of simulation scenarios for the covariates and model distributions. We generated data with (1) no covariates (excluding x_{ij1} and x_{ij2} from (19)); (2) two individual-level covariates (applying an intraclass correlation coefficient of $\rho_X = 0$); (3) two cluster-level covariates (applying $\rho_X = 1$); and (4) one individual-level and one cluster-level covariate. For the models with individual-level covariates, we calculated the degrees of freedom in various ways, letting v^* equal 0, the total number of covariates (v), or the number of cluster-level covariates (if applicable). Finally, we generated data assuming normal, bimodal, and chi-square distributions for the errors and covariates in (19) (see Supplementary Materials B).

Finite population results. The simulation results for the models without covariates indicate that the design-based estimator yields Type 1 errors near the 5 percent nominal level and standard errors near true values for specifications that exclude the correction for the finite population heterogeneity term, even with relatively few clusters (see Figure 1 which assumes normal errors and Supplementary Materials Table B.1 for the full results). In contrast, the standard CRSE estimator in (10) yields inflated Type 1 errors similar to those found in the literature using a super-population simulation framework (see, for example, Cameron et al. 2008; Green and Vavreck 2008; Angrist and Pischke 2009) (Figure 1 and Table B.1). The key reason is that the CRSE estimator applies a *single* degrees of freedom variance adjustment based on the total sample size, whereas the design-based estimator applies a *separate* degrees of freedom adjustment for the treatment and control groups which inflates the variances. A more minor, but related reason is that the CRSE approach uses $(m-1)$ degrees of freedom for the t-tests rather than $(m-2)$ as for the design-based approach. The design-based variance estimator that includes a correction for the finite population heterogeneity term also overrejects (so we do not focus on this estimator in what follows), but less so than the CRSE estimator (Figure 1; Table B.3). We find similar simulation results using different model distributions (Table B.1) and using an average of 40 individuals per cluster rather than 100 (Table B.3).

A similar pattern of results arises when covariates are included in the model (and standard errors decrease), but the results depend on whether the model includes individual- and/or cluster-level covariates and whether the R_{TX}^2 adjustment is applied (Figure 2 and Tables B.2 and B.3). For the model with individual-level covariates only ($\rho_X = 0$), we find that the design-based estimator with $v^* = 0$ yields Type 1 errors near the nominal level, even with relatively few clusters (Model (1) in Figure 2). In this case, the R_{TX}^2 adjustment has little effect on the results due to the relatively large number of individuals per cluster (with a mean of 100) (Table B.3). If we instead apply $v^* = 2$, the design-based approach becomes conservative (Model (2) in Figure 2). As before, the standard CRSE estimator yields inflated Type 1 errors (Model (3) in Figure 2).

For the model with cluster-level covariates only ($v^* = 2;\ \rho_X = 1$), which is identical to aggregating the data to the cluster level, the design-based estimator yields Type 1 errors at the nominal level if the R_{TX}^2 adjustment is applied, even with $m = 8$, but overrejects without the

R_{TX}^2 adjustment (Models (4) and (5) in Figure 2 and Tables B.2 and B.3). For this specification, the CRSE estimator produces inflated Type 1 errors more pronounced than with individual-level covariates (Model (6) in Figure 2). The performance of the CRSE estimator improves using $(m - v - 2) = (m - 4)$ degrees of freedom for the t-tests rather than $(m - 1)$ (Table B.3).

Note that the aggregated model could also be estimated if individual-level covariates are available for the analysis. However, the true standard errors are larger using this approach than using the individual data, leading to losses in statistical power. Precision losses using the aggregated data occur because of larger expected correlations between the covariates and the treatment indicator (Schochet 2020 quantifies these design effects).

Finally, we ran simulations for models containing one individual-level covariate ($v_1 = 1$) and one cluster-level covariate ($v_2 = 1$) (Table B.3). For the design-based estimator, the simulations suggest that using $v^* = 0$ is liberal, setting $v^* = v_2 = 1$ yields Type 1 errors close to the nominal rate; and setting $v^* = v_1 + v_2 = 2$ is conservative. As before, the CRSE tends to overreject with few clusters.

Super-population simulation results. To compare our results to those in the literature for the CRSE estimator, we also conducted select simulations using a super-population framework by generating 50,000 separate datasets and calculating Type 1 errors and standard errors across them. These results show a similar pattern of results to the above, but with somewhat larger true standard errors (Table B.4).

Discussion. Overall, the simulation results suggest that the design-based estimator has beneficial statistical properties with few clusters for models with or without covariates. For models with covariates and relatively large n (which is typical for clustered RCTs in practice), the results suggest that adjusting the degrees of freedom for the number of cluster-level covariates by setting $v^* = v_2$ (the number of cluster-level covariates) could be a good general strategy, and that setting $v^* = v$ is conservative for models with some individual-level covariates. The simulations further indicate that with individual-level covariates, the design-based estimator using data averaged to the cluster-level with the R^2_{TX} adjustment and $v^* = v$ yields nominal rejection rates. While this approach can result in losses in statistical power, it could be a good strategy if n is small, in which case precision losses could be tolerable.

7. Empirical Application Using the Motivating SACD Example

To demonstrate the considered estimators, we use outcome and baseline data from the SACD evaluation on 4,018 4th graders (2,147 treatments and 1,871 controls) in 84 schools in 7 large school districts. The data were obtained from student reports administered in the classroom, primary caregiver telephone interviews, and teacher reports on students (SACD Consortium, 2010). We analyze six primary study outcome scales (Table 1) and adjust for baseline covariates selected in an initial step from the 46 available, along with their two-way interactions, using Least Absolute Shrinkage and Selection Operator (LASSO) methods with 5-fold cross-validation (Tibshirani 1996). Our goal is not to replicate study results but to illustrate the ATE estimators.

Table 2 presents the estimation results for various model specifications: (1) with and without baseline covariates, (2) with and without block-by-treatment status interaction terms, and (3) with equal weighting of individuals (to estimate ATEs for the average student in the sample) versus equal weighting of sites and clusters (to estimate ATEs for the average school in the average district in the sample). Our methodology can easily accommodate other weighting schemes, such as those that adjust for data item nonresponse. For some specifications, we compare the design-based results to those using the standard CRSE estimator (for illustration).

The results indicate that for all specifications, the behavioral health interventions had no statistically significant effect on any outcome scale, although the negative estimate on the scale measuring fear in school is marginally statistically significant at the 10 percent level for most models with covariates. Across the six outcomes, standard errors are about 16 to 35 percent smaller when covariates are included in the models. Further, we find very similar results for the fully-interacted and restricted models for two reasons: (1) the estimated treatment effects vary little across sites (0.07 standard deviations on average across the outcomes) and (2) the two sets of site weights are highly correlated (greater than 0.95) because sample sizes do not vary substantially across sites (they range from 425 to 650 students and 10 to 14 schools and $p_b = 0.5$ in all sites). For similar reasons, findings do not materially differ when individuals versus blocks and clusters are weighted equally, although in the latter case, standard errors increase due to design effects from weighting and the marginally significant impact on the scale measuring fear at school disappears. Finally, consistent with the theory and simulations, standard errors are somewhat larger using the design-based estimators than the parallel CRSE estimators.

8. Conclusions

This article considered design-based ratio estimators for clustered, blocked RCTs using the Neyman-Rubin-Holland model and weighted least squares methods. We developed finite population CLTs for the ATE estimators, allowing for baseline covariates to improve precision, general weighting schemes, and several common approaches for handling blocks in the models. We showed that the design-based ratio estimators are attractive in that they yield consistent and asymptotically normal ATE estimators with simple variance estimators based on cluster-level model residuals; apply to continuous, binary, and discrete outcomes; and yield Type 1 errors at nominal levels for models with and without covariates, even in small samples. An unexpected finding is that the "conservative" variance estimator that excludes a correction for the finite population heterogeneity term based on the Cauchy-Schwarz inequality improves statistical performance. Further, for models with covariates, an R^2_{TX} adjustment for the collinearity between the covariates and treatment indicator improves results in designs with few clusters and subjects.

Our findings justify the CRSE estimator from a finite population perspective (even though it estimates a super-population ATE parameter); this contribution follows similar literature for the individual randomized case (see, for example, Freedman 2008 and Lin 2013). However, while the structure of the design-based and standard CRSE variance estimators are similar, differences in their degrees of freedom adjustments do affect their statistical performance in small samples (the standard CRSE estimator overrejects in this case). The key difference is simple: the randomization mechanism leads to separate degrees of freedom adjustments for the treatment and control groups based on their respective numbers of clusters (in each block), whereas the CRSE approach often used in practice applies a single adjustment based on the total number of clusters. These differences tend to increase with more blocks.

As discussed in the article, other corrections for the CRSE estimator have been proposed that can improve the Type 1 error inflation rate in general settings. In the RCT setting, however, the advantage of the design-based variance estimator is that it is tailored to experiments, as it is derived directly from principles underlying them. Further, it is simple to apply and parallels design-based estimators for non-clustered RCTs. The free *RCT-YES* software (www.rct-yes.com), funded by the U.S. Department of Education, estimates ATEs for full sample and baseline subgroup analyses using the design-based methods discussed in this article using either R or Stata, and also allows for multi-armed trials with multiple treatment conditions.

Acknowledgements

The authors thank Charles Tilley, Joel Middleton, and session participants at the Society for Research on Education Effectiveness conference for useful feedback on the manuscript. Nicole Pashley is supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE1745303. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

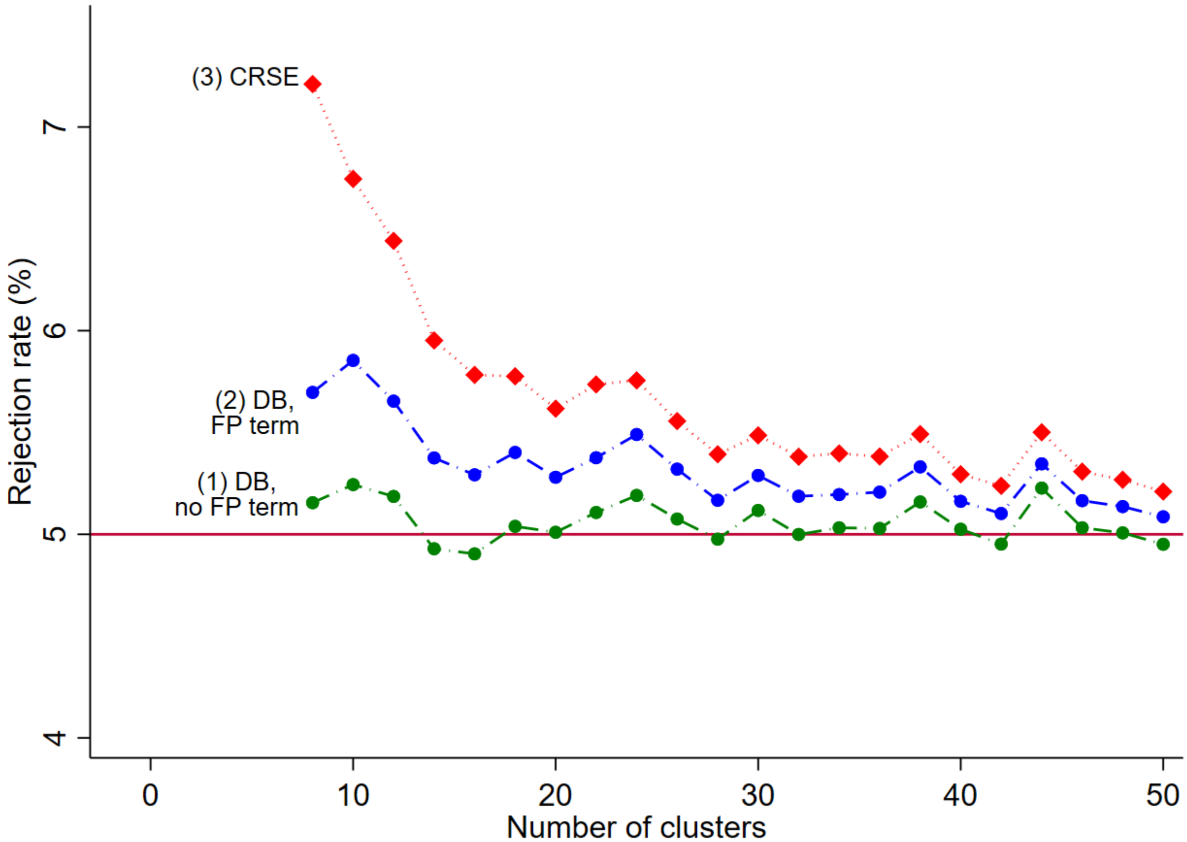

Figure 1. Type I Error Rates for Models without Covariates

Abbreviations. DB = Design-based variance estimator; FP term = Finite population heterogeneity term included based on the Cauchy-Schwarz inequality; CRSE = Standard cluster-robust standard error estimator.

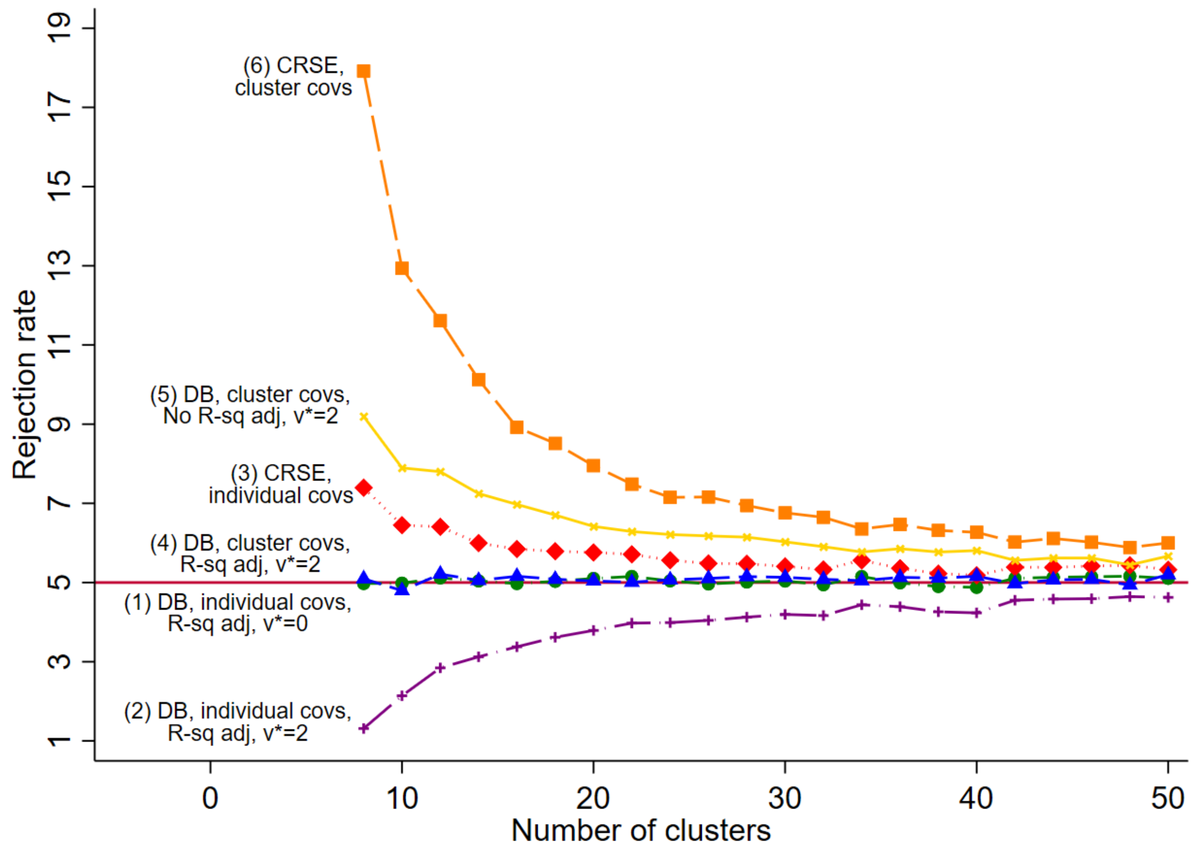

Figure 2. Type I Error Rates for Models with Covariates

Abbreviations. DB = Design-based variance estimator in (9) (without the finite population heterogeneity term); "cluster covs" = Two cluster-level covariates included in the model; "Individual covs" = Two individual-level covariates included in the model; "R-sq-adj" = R^2_{TX} adjustment applied to the DB estimator; v^*= Degrees of freedom adjustment for the covariates for the DB estimator; CRSE = Standard cluster-robust standard error estimator.

Table 1. Outcome variables for the empirical analysis using SACD RCT data.

Outcome	Data source (Spring 2007)	Description of variable
Problem behavior	Child report	Scale ranges from 0 to 3 and contains 6 items from the Frequency of Delinquent Behavior Scale and 6 items from the Aggression Scale; Reliability = 0.86.
Normative beliefs about aggression	Child report	Scale ranges from 1 to 4 and contains 12 items from the Normative Beliefs About Aggression Scale; Reliability = 0.83.
Student afraid at school	Child report	Scale ranges from 1 to 4 and contains 4 items from the Feelings of Safety at School scale; Reliability = 0.79.
Altruistic behavior	Primary caregiver report	Scale ranges from 1 to 4 and contains 8 items from the Altruism Scale, Primary Caregiver Version; Reliability = 0.88
Positive social behavior	Teacher report	Scale ranges from 1 to 4 and contains 6 items from the Responsibility Scale and 19 items from the Social Competence Scale and 8 items from the Altruism Scale, Teacher Version; Reliability = 0.97.
Problem behavior	Teacher report	Scale ranges from 1 to 4 and contains 14 items from the BASC Aggression Subscale, Teacher Version, 7 items from the BASC Conduct Problems Subscale, Teacher Version and 2 items from the Responsibility Scale; Reliability = 0.95.

Note: See SACD Research Consortium (2010) for a complete description of the construction of these scales.

Table 2. Estimated ATEs and standard errors for the SACD study, by model specification.

Outcome variable and covariate specification	Model with site-by-treatment interaction terms: Individuals weighted equally, Design-based	Model with site-by-treatment interaction terms: Individuals weighted equally, Standard CRSE	Model with site-by-treatment interaction terms: Schools and sites weighted equally, Design-based	Model with site fixed effects only: Design-based	Model with site fixed effects only: Standard CRSE
Model without covariates					
Problem behavior (CR)	0.006 (0.037)	0.006 (0.034)	0.011 (0.041)	0.006 (0.036)	0.006 (0.035)
Normative beliefs about aggression (CR)	0.003 (0.031)	0.003 (0.029)	0.000 (0.038)	0.003 (0.031)	0.003 (0.030)
Student afraid at school (CR)	-0.064 (0.052)	-0.064 (0.048)	-0.041 (0.061)	-0.064 (0.050)	-0.064 (0.048)
Altruistic behavior (PCR)	-0.006 (0.035)	-0.006 (0.032)	-0.011 (0.041)	-0.006 (0.034)	-0.006 (0.033)
Positive social behavior (TR)	-0.046 (0.061)	-0.046 (0.056)	-0.036 (0.065)	-0.045 (0.060)	-0.045 (0.056)
Problem behavior (TR)	0.019 (0.040)	0.019 (0.036)	0.006 (0.044)	0.019 (0.039)	0.019 (0.038)
Model with covariates					
Problem behavior (CR)	-0.006 (0.027)	-0.006 (0.025)	-0.002 (0.031)	-0.006 (0.027)	-0.006 (0.025)
Normative beliefs about aggression (CR)	-0.005 (0.026)	-0.005 (0.024)	-0.009 (0.033)	-0.005 (0.026)	-0.005 (0.025)
Student afraid at school (CR)	-0.067* (0.040)	-0.067* (0.035)	-0.047 (0.045)	-0.067* (0.037)	-0.067* (0.036)
Altruistic behavior (PCR)	-0.016 (0.028)	-0.016 (0.024)	-0.013 (0.031)	-0.017 (0.027)	-0.017 (0.026)
Positive social behavior (TR)	-0.011 (0.045)	-0.011 (0.039)	-0.015 (0.045)	-0.010 (0.043)	-0.010 (0.040)
Problem behavior (TR)	-0.009 (0.025)	-0.009 (0.023)	-0.011 (0.026)	-0.009 (0.025)	-0.009 (0.023)

Abbreviations. CR = child report, PCR = primary caregiver report, TR = teacher report, CRSE = Cluster-robust standard error estimator.

* Statistically significant at the 10 percent level, two-tailed test.

References

Abadie A., Athey, S., Imbens, G., and Wooldridge, J. (2017), "When Should You Adjust Standard Errors for Clustering?" arxiv: 1710.02926[Math.ST]

Angrist, J. and Pischke, S. (2009), *Mostly Harmless Econometrics*, Princeton NJ: Princeton University Press.

Angrist, J. D. and Lavy, V. (2002), "The Effect of High School Matriculation Awards: Evidence from Randomized Trials," *American Economic Review,* 99, 1384–1414.

Aronow, P. M. and Middleton, J. A. (2013), "A Class of Unbiased Estimators of the Average Treatment Effect in Randomized Experiments," *Journal of Causal Inference,* 1, 135–154.

Aronow, P. M., Green, D. P. and Lee, D. K. K. (2014), "Sharp Bounds on the Variance in Randomized Experiments," *Annals of Statistics*," 42, 850-871.

Bell, R. and Mccaffrey, D. (2002), "Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples," *Survey Methodology,* 28, 169–181.

Bland, J. M. (2004), "Cluster Randomised Trials in the Medical Literature: Two Bibliometric Surveys," *BMC Medical Research Methodology*, 4, 21.

Cameron, A. C., and Miller, D. L. (2015), "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 50, 317-372.

Cameron, A. C., Gelbach, J. G., and Miller, D. L. (2008), "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics,* 90, 414–27.

Cochran, W. (1977), *Sampling Techniques*, New York: John Wiley and Sons.

Donald, S. G., and Lang, K. L. (2007), "Inference with Difference-in-Differences and Other Panel Data," *Review of Economics and Statistics,* 89, 221–33.

Freedman, D. (2008), "On Regression Adjustments to Experimental Data," *Advances in Applied Mathematics* 40, 180-193.

Fuller, W. A. (1975), "Regression Analysis for Sample Survey," *Sankya*, 37, Series C, 117–132.

Fuller, W. A. (2009), *Sampling Statistics*, Hoboken, NJ: Wiley.

Green, D.P. and Vavrek, L. (2008), "Analysis of Cluster-Randomized Experiments: A Comparison of Alternative Estimation Approaches," *Political Analysis*, 16, 138–152.

Hansen, C. B. (2007), "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data When T Is Large," *Journal of Econometrics,* 141*,* 597-620.

Hansen, B. B., and Bowers, J. (2009), "Attributing Effects to a Cluster-Randomized Get-Out-the-Vote Campaign," *Journal of the American Statistical Association*, 104, 873-885.

Hedges, L. (2007), "Correcting a Significance Test for Clustering," *Journal of Educational and Behavioral Statistics,* 32, 151-179.

Holland, P. W. (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association,* 81, 945–960.

Huber, P. J. (1967), "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," *Procedures of the Fifth Berkeley Symposium on Math and Statistical Probability*, 1, 221-233.

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

Imbens G. (2004), "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *Review of Economics and Statistics*, 86, 4-29.

Imbens, G. and Rubin, D. (2015), *Causal inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge, UK: Cambridge University Press.

Kish, L. (1995), *Survey Sampling*, New York, NY: John Wiley and Sons.

Li, X. and Ding, P. (2017), "General Forms of Finite Population Central Limit Theorems with Applications to Causal inference," *Journal of the American Statistical Association,* 112, 1759-1769.

Liang, K. and Zeger, S. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika,* 73, 13-22.

Lin, W. (2013), "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique," *Annals of Applied Statistics*, 7, 295-318.

Lohr, S. L. (2009), *Sampling: Design and Analysis*, Second Edition. Pacific Grove, CA: Duxbury Press.

Mackinnon, J. G. and White, H. (1985), "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics* 29, 305–25.

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.

Miratrix, L.W., Sekhon, J. B. and Yu, B. (2013), "Adjusting Treatment Effect Estimates in Randomized Experiments," *Journal of the Royal Statistical Society B*, 75, 369-396.

Neyman, J. (1923), “On the Application of Probability Theory to Agricultural Experiments: Essay on Principles,” Section 9, Translated in *Statistical Science*, 1990, 5, 465-472.

Pashley. N. and Miratrix, L. (2017), “Insights on Variance Estimation for Blocked and Matched Pair Designs,” Arxiv:1710.10342v1 [Stat.ME] 27-Oct 2017.

Pustejovsky, J. E., and Tipton, E. (2018), “Small-Sample Methods for Cluster-Robust Variance Estimation and Hypothesis Testing in Fixed Effects Models,” *Journal of Business and Economic Statistics*, 36, 672–683.

Rao, J. N. K., and Shao, J. (1999), “Modified Balanced Repeated Replication for Complex Survey Data,” *Biometrika,* 86*,* 403-415.

Rubin, D. B. (1974), “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies,” *Journal of Education Psychology,* 66, 688–701.

Rubin, D. B. (1977), “Assignment to Treatment Group on the Basis of a Covariate,” *Journal of Education Statistics,* 2, 1–26.

Rubin, D. B. (1986), “Which Ifs Have Causal Answers? Discussion of Holland’s “Statistics and Causal inference.”,” *Journal of the American Statistical Association,* 81, 961-962.

SACD Research Consortium (2010). “Efficacy of Schoolwide Programs to Promote Social and Character Development and Reduce Problem Behavior in Elementary School Children.” Washington DC: Final Report: Institute for Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncer/pubs/20112001/

Samii, C. and Aronow, P. M. (2012), “On Equivalencies Between Design-Based and Regression-Based Variance Estimators for Randomized Experiments,” *Statistics & Probability Letters*, 82, 365-370.

Schochet, P. Z. (2008), “Statistical Power for Random Assignment Evaluations of Education Programs,” *Journal of Educational and Behavioral Statistics,* 33, 62-87.

Schochet, P. Z. (2010), “Is Regression Adjustment Supported by the Neyman Model for Causal inference?” *Journal of Statistical Planning and inference,* 140, 246–259.

Schochet, P. Z. (2013), “Estimators for Clustered Education RCTs Using the Neyman Model for Causal inference,” *Journal of Educational and Behavioral Statistics,* 38, 219–238.

Schochet, P. Z. (2016) Second Edition. *Statistical Theory for the RCT-YES Software: Design-Based Causal inference for RCTs* (NCEE 2015–4011), Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. **https://ies.id.gov/ncee/pubs/20154011/pdf/20154011.pdf**.

Schochet P. Z. (2020), "Analyzing Grouped Administrative Data for RCTs Using Design-Based Methods," *Journal of Educational and Behavioral Statistics*, 45: 32-57.

Scott, A. and Wu, C. F. (1981), "On the Asymptotic Distribution of Ratio and Regression Estimators," *Journal of the American Statistical Association*, 112, 1759-1769.

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B,* 58, 267-288.

Webb, M. D. (2013), "Reworking Wild Bootstrap Based Inference for Clustered Errors," Queens Economics Department Working Paper 1315.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica,* 48, 817-838.

Wolter, K.M. (2007), *Introduction to Variance Estimation*, Second Edition. Springer Verlag.

Wooldridge, J. M. (2006), "Cluster Sample Methods in Applied Econometrics," Michigan State University, Working Paper.

Yang, L. and Tsiatis, A. (2001), "Efficiency Study of Estimators for a Treatment Effect in a Pretest-Posttest Trial," *American Statistician*, 55, 314-321.

Supplementary Materials for "Design-Based Estimators for Clustered RCTs and How They Compare to Robust Estimators"

A Proofs for asymptotic results

A.1 Asymptotic properties of ratio estimators

In this section, we develop two results that will help us prove asymptotic normality of our weighted least squares regression estimators. We derive these results generally, assuming arbitrary cluster level[1] potential outcomes $C_j(t)$ for cluster j with $t \in \{0, 1\}$. These results could alternatively be obtained using a finite population delta method (see Pashley, 2019).

A.1.1 Asymptotic normality of one ratio

We first derive a result on the asymptotic distribution of a single ratio estimator, very similar to Theorem 1 in Scott and Wu (1981), but using conditions from Li and Ding (2017). To do this, we redefine the potential outcomes in a way that allows us to use known normality results for an estimator that is close to our ratio estimator. Then we use an application of Slutsky's theorem to get the normality result for our estimator of interest. Before the result, we must define some notation. Let us have arbitrary cluster level potential outcomes $C_j(t)$ for $t \in \{0, 1\}$ and cluster weights w_j, with T_j being the indicator of treatment assignment for cluster j. Following Li and Ding (2017), our finite population is within a sequence of finite populations where m and m^t go to ∞, which naturally holds if we grow the population with a fixed proportion of treated clusters, p. Further define

$$\overline{w} = \frac{1}{m}\sum_{j=1}^{m} w_j,$$

$$\overline{w}^t = \frac{1}{m^t}\sum_{j:T_j=t} w_j,$$

$$\overline{\overline{C}}(t) = \frac{\frac{1}{m}\sum_{j=1}^{m} w_j C_j(t)}{\overline{w}} \quad \text{(the overall weighted mean of } C_j(t)\text{)},$$

$$\overline{\overline{c}}(t) = \frac{\frac{1}{m^t}\sum_{j:T_j=t} w_j C_j(t)}{\overline{w}^t},$$

[1] The same mathematical arguments are immediately extendable to the case where we have individuals rather than clusters.

and

$$\hat{\overline{\overline{C}}}(t) = \overline{\overline{c}}(t)\overline{w}.$$

In particular, we are interested in the asymptotic distribution of the ratio estimator $\overline{\overline{c}}(t)$, unlike Scott and Wu (1981) who found the distribution for an estimator closer to the form $\hat{\overline{\overline{C}}}(t)$. In our case, $\overline{\overline{c}}(t)$ is an estimator of $\overline{\overline{C}}(t)$. We will see this is a straightforward change.

Define the following finite population variance of the weights:

$$S^2(w) = \frac{1}{m-1}\sum_{j=1}^{m}(w_j - \overline{w})^2.$$

Let $D_j(t)$ by a scaled deviation of $C_j(t)$ from the overall weighted mean. That is,

$$D_j(t) = \left(w_j C_j(t) - w_j \overline{\overline{C}}(t)\right)/\overline{w}.$$

Then the mean of the $D_j(t)$'s, $\bar{D}(t)$, equals 0 as

$$\sum_{j=1}^{m} w_j \overline{\overline{C}}(t) = m\overline{w}\overline{\overline{C}}(t) = \sum_{j=1}^{m} w_j C_j(t),$$

and the variance expression for the $D_j(t)$, which is a weighted variance of the $C_j(t)$, is

$$S_D^2(t) = \frac{1}{m-1}\sum_{j=1}^{m} D_j^2(t) = \frac{1}{m-1}\sum_{j=1}^{m}\frac{w_j^2}{\overline{w}^2}\left(C_j(t) - \overline{\overline{C}}(t)\right)^2.$$

Lemma A.1.1. *Assume we have the following conditions:*

(a) Defining $g(t) = max_{1\leq j\leq m}(D_j(t))^2$

$$\frac{1}{min(m^1, m-m^1)}\frac{g(t)}{S_D^2(t)} \to 0, \quad \textit{as } m \to \infty. \tag{S1}$$

(b) For $t \in \{1, 0\}$,

$$\left(1 - f^t\right)\frac{S^2(w)}{m^t\overline{w}^2} \to 0 \textit{ as } m \to \infty. \tag{S2}$$

Under these conditions,

$$\frac{\overline{\overline{c}}(t) - \overline{\overline{C}}(t)}{\sqrt{\frac{1-f^t}{m^t}}S_D(t)/\overline{w}} \xrightarrow{d} N(0,1). \tag{S3}$$

Proof. The proof of this result closely follows the proof in Scott and Wu (1981). We are interested in the asymptotic distribution of the ratio estimator $\bar{\bar{c}}(t)$, but to find this we will first find the distribution of $\hat{\bar{C}}(t)$.

Let $\bar{d}(t) = \sum_{j:T_j=t} D_j(t)/m^t$. Then

$$\bar{d}(t) = \left(\bar{\bar{c}}(t) - \bar{\bar{C}}(t)\right)\frac{\overline{w}^t}{\overline{w}}.$$

Assuming our first condition given by Equation S1, by Theorem 1 of Li and Ding (2017),

$$\frac{\bar{d}(t)}{\sqrt{\frac{1-f^t}{m^t}S_D(t)}} \xrightarrow{d} N(0,1).$$

Further, from Theorem B of Scott and Wu (1981), given our second condition given by Equation S2,

$$\overline{w}^t/\overline{w} - 1 \xrightarrow{p} 0 \text{ as } m \to \infty.$$

Thus, $\overline{w}^t/\overline{w} \xrightarrow{p} 1$. Note the condition on the weights would cause issues if the average weight were to go to 0 as the sample size increased, but, as discussed in Section A.2.3, this is easy to avoid by using unnormalized weights. This in turn means by Slutsky's Theorem, recalling that $\bar{\bar{c}}(t) - \bar{\bar{C}}(t) = \bar{d}(t)\overline{w}/\overline{w}^t$, we have our result in Equation S3.

□

A.1.2 Asymptotic normality of difference of ratios (Main result)

We will now show the asymptotic normality of a difference of ratios. In particular, we are interested in the asymptotic distribution of $\bar{\bar{c}}(1) - \bar{\bar{c}}(0)$, using notation from Section A.1.1. We will again define new potential outcomes that are easier to work with than the direct ratio. First we define some notation. Construct new potential outcomes $D_j(t) = (w_jC_j(t) - w_j\bar{\bar{C}}(t))/\overline{w}$ with notation defined in Section A.1.1, so $\bar{D}(1) = 0$, $\bar{D}(0) = 0$ and

$$\bar{d}(t) = \frac{1}{m^t}\sum_{j:T_j=t} D_j(t).$$

Further define an intermediate estimator $\hat{D} = \bar{d}(1) - \bar{d}(0)$ and

$$S_D^2(1,0) = \frac{1}{m-1}\sum_{j=1}^{m} D_j(1)D_j(0).$$

Recall for $t \in \{0,1\}$

$$S_D^2(t) = \frac{1}{m-1}\sum_{j=1}^{m} D_j^2(t).$$

and further define

$$S^2(D) = \frac{1}{m-1}\sum_{j=1}^{m}(D_j(1) - D_j(0))^2.$$

We will use these notations to get the final result.

Lemma A.1.2. *Let us have the following conditions*[2]*:*

(a) Let $g(t) = max_{1 \leq j \leq m}(D_j(t))^2$ and as $m \to \infty$

$$max_{t \in \{0,1\}} \frac{1}{(m^t)^2} \frac{g(t)}{Var(\hat{D})} \to 0. \tag{S4}$$

(b) m^t/m has a limiting value in (0,1) and $S_D^2(t)$ and $S_D^2(1,0)$ have limiting values.

(c) For $t \in \{1, 0\}$,

$$\left(1 - f^t\right) \frac{S^2(w)}{m^t \overline{w}^2} \to 0 \text{ as } m \to \infty. \tag{S5}$$

Then we have that

$$\frac{\left(\bar{\bar{c}}(1) - \bar{\bar{c}}(0)\right) - \left(\overline{\overline{C}}(1) - \overline{\overline{C}}(0)\right)}{\sqrt{Var(\hat{D})}} \xrightarrow{d} N(0,1) \tag{S6}$$

where

$$Var(\hat{D}) = \frac{S_D^2(1)}{m^1} + \frac{S_D^2(0)}{m^0} - \frac{S^2(D)}{m}.$$

Proof. First we derive a normality result for $\hat{D}$. If we randomly select pm clusters to assign to treatment, then from Theorem 3 in Li and Ding (2017), $\hat{D}$ has mean 0 and variance

$$\mathrm{Var}(\hat{D}) = \frac{S_D^2(1)}{m^1} + \frac{S_D^2(0)}{m^0} - \frac{S^2(D)}{m}.$$

If we have the condition of Equation S4, then under Theorem 4 of Li and Ding (2017),

$$\frac{\hat{D}}{\sqrt{\mathrm{Var}(\hat{D})}} \xrightarrow{d} \mathrm{N}(0, 1).$$

Now note that

$$\begin{aligned}
\frac{\hat{D}}{\sqrt{\mathrm{Var}(\hat{D})}} &= \frac{\bar{d}(1) - \bar{d}(0)}{\sqrt{\mathrm{Var}(\hat{D})}} \\
&= \frac{\frac{1}{m^1} \sum_{j=1}^{m} T_j \left(w_j C_j(1) - w_j \overline{\overline{C}}(1)\right) / \overline{w} - \frac{1}{m^0} \sum_{j=1}^{m} (1 - T_j) \left(w_j C_j(0) - w_j \overline{\overline{C}}(0)\right) / \overline{w}}{\sqrt{\mathrm{Var}(\hat{D})}} \\
&= \frac{\frac{\overline{w}^1}{\overline{w}} \left(\bar{\bar{c}}(1) - \overline{\overline{C}}(1)\right) - \frac{\overline{w}^0}{\overline{w}} \left(\bar{\bar{c}}(0) - \overline{\overline{C}}(0)\right)}{\sqrt{\mathrm{Var}(\hat{D})}}.
\end{aligned}$$

[2] It is interesting to note that a delta method approach would rely on slightly different conditions. In particular, the delta method does not explicitly require limiting values on asymptotic variances, as given in the second condition, but it does require a convergence in probability result that is satisfied by those limiting values for the asymptotic variances. Comparing this derivation to proofs via the delta method is left for future work.

We can now get back to our estimator of interest, $\bar{\bar{c}}(1) - \bar{\bar{c}}(0)$:

$$\begin{aligned}
&\frac{\left(\bar{\bar{c}}(1) - \overline{\overline{C}}(1)\right) - \left(\bar{\bar{c}}(0) - \overline{\overline{C}}(0)\right)}{\sqrt{\mathrm{Var}(\hat{D})}} \\
&= \frac{\left(1 - \frac{\overline{w}^1}{\overline{w}}\right)\left(\bar{\bar{c}}(1) - \overline{\overline{C}}(1)\right) - \left(1 - \frac{\overline{w}^0}{\overline{w}}\right)\left(\bar{\bar{c}}(0) - \overline{\overline{C}}(0)\right)}{\sqrt{\mathrm{Var}(\hat{D})}} \\
&\quad + \frac{\frac{\overline{w}^1}{\overline{w}}\left(\bar{\bar{c}}(1) - \overline{\overline{C}}(1)\right) - \frac{\overline{w}^0}{\overline{w}}\left(\bar{\bar{c}}(0) - \overline{\overline{C}}(0)\right)}{\sqrt{\mathrm{Var}(\hat{D})}} \\
&= \left(1 - \frac{\overline{w}^1}{\overline{w}}\right)\frac{\left(\bar{\bar{c}}(1) - \overline{\overline{C}}(1)\right)}{\sqrt{\mathrm{Var}(\hat{D})}} - \left(1 - \frac{\overline{w}^0}{\overline{w}}\right)\frac{\left(\bar{\bar{c}}(0) - \overline{\overline{C}}(0)\right)}{\sqrt{\mathrm{Var}(\hat{D})}} \\
&\quad + \frac{\hat{D}}{\sqrt{\mathrm{Var}(\hat{D})}}.
\end{aligned}$$

We know the asymptotic distribution of the third term, so we need to show that the first two terms vanish. As stated in Lemma A.1.1, if we have the conditions given by Equation S4 and Equation S5, then

$$\frac{\bar{\bar{c}}(t) - \overline{\overline{C}}(t)}{\sqrt{\frac{1-f^t}{m^t}S_D(t)}} \xrightarrow{d} N(0,1).$$

Recall from Section A.1.1 that the fourth condition gives us

$$\frac{\overline{w}^t}{\overline{w}} \xrightarrow{p} 1.$$

To use these notes, we need to rewrite out previous result:

$$\begin{aligned}
\frac{\left(\bar{\bar{c}}(1) - \overline{\overline{C}}(1)\right) - \left(\bar{\bar{c}}(0) - \overline{\overline{C}}(0)\right)}{\sqrt{\mathrm{Var}(\hat{D})}} &= \left(1 - \frac{\overline{w}^1}{\overline{w}}\right)\frac{\sqrt{\frac{1}{m^1} - \frac{1}{m}S_D(1)}}{\sqrt{\mathrm{Var}(\hat{D})}}\frac{\left(\bar{\bar{c}}(1) - \overline{\overline{C}}(1)\right)}{\sqrt{\frac{1}{m^1} - \frac{1}{m}S_D(1)}} \\
&\quad - \left(1 - \frac{\overline{w}^0}{\overline{w}}\right)\frac{\sqrt{\frac{1}{m^0} - \frac{1}{m}S_D(0)}}{\sqrt{\mathrm{Var}(\hat{D})}}\frac{\left(\bar{\bar{c}}(0) - \overline{\overline{C}}(0)\right)}{\sqrt{\frac{1}{m^0} - \frac{1}{m}S_D(0)}} \\
&\quad + \frac{\hat{D}}{\sqrt{\mathrm{Var}(\hat{D})}}.
\end{aligned}$$

The second condition puts limiting values on our asymptotic variances. That is, the second condition (see Theorems 3 and 5 of Li and Ding (2017)), gives us that for $t \in \{0,1\}$,

$m\left(\frac{1-f^t}{m^t}\right) S_D^2(t)$ has some limiting value $\sigma_D^2(t)$ and $m\text{Var}(\hat{D})$ has some limiting value V. This allows us to use Slutsky's theorem to get

$$\left(1 - \frac{\overline{w}^1}{\overline{w}}\right) \frac{\sqrt{\frac{1}{m^1} - \frac{1}{m}} S_D(t)/\overline{w}}{\sqrt{\text{Var}(\hat{D})}} \frac{\left(\bar{\bar{c}}(t) - \overline{\overline{C}}(t)\right)}{\sqrt{\frac{1-f^t}{m^t}} S_D(t)} \xrightarrow{p} 0.$$

To break this step down more, note that the first term goes to 0 in probability, the second goes to a constant by the third condition, and the final term is asymptotically normal.

Finally, Slutsky's theorem gives the result in Equation S6. $\square$

A.2 Theorem 1 without covariates

In this section, we explore the impact estimator of a weighted least squares regression without covariate adjustment. We prove Theorem 1 under this special case and also find the bias of our estimator. We start by deriving the form of the estimator, then finding bias and consistency results. In particular, the estimator has finite sample bias but is consistent. We then show the asymptotic normality result given in Theorem 1. Without covariates, the interactions in our regression imply that block means and effects are estimated independently, and so we can focus on the estimator for a single block, b. In other words, for now we can act as if there is only one block.

A.2.1 The estimator

We assume that we have one block, arbitrarily block b, with m_b clusters. We assign $m_b^1 = p_b m_b$ clusters to treatment and the rest of the m_b^0 clusters to control. Define shorthand $f_b^t = m_b^t/m_b$. When we have multiple blocks, we have block indicators S_{ijb} and S_{jb} which equals 1 if unit i or cluster j belong to block b, and 0 otherwise. Each unit i, of n_{jb} units, in cluster j of block b has associated weight w_{ijb} and we have $w_{jb} = \sum w_{ijb}$. Denote $w_b^0 = \sum_{j=1}^{m_b}(1 - T_{jb})w_{jb}$, $w_b^1 = \sum_{j=1}^{m_b} T_{jb}w_{jb}$, and $w_b = \sum_{j=1}^{m_b} w_{jb}$. Each unit i in cluster j has potential outcomes $Y_{ijb}(1)$ under treatment and $Y_{ijb}(0)$ under control, with treatment assigned at the cluster level. y_{ijb} is the observed outcome for unit i in cluster j of block b and $\bar{y}_{jb}$ is the observed weighted average outcome for all units in cluster j of block b. We have by definition that $y_{ijb} = T_{jb}Y_{ijb}(1) + (1 - T_{jb})Y_{ijb}(0)$ and for $t \in \{0, 1\}$,

$$\bar{Y}_{jb}(t) = \frac{1}{w_{jb}} \sum_{i:T_{jb}=t} w_{ijb}Y_{ijb}(t)$$

and $\bar{y}_{jb} = T_{jb}\bar{Y}_{jb}(1) + (1 - T_{jb})\bar{Y}_{jb}(0)$. We also define for $t \in \{0, 1\}$,

$$\bar{\bar{y}}_b(t) = \frac{1}{\sum_{j:T_{jb}=t} w_{jb}} \sum_{j:T_{jb}=t} w_{jb}\bar{y}_{jb}.$$

We have finite population parameters

$$\bar{\bar{Y}}_b(t) = \frac{1}{\sum_{j=1}^{m_b} w_{jb}} \sum_{j=1}^{m_b} w_{jb}\bar{Y}_{jb}(t).$$

Throughout, we will treat $\beta_{1,b}$ as the *finite population* parameter of the weighted average of cluster impacts,

$$\beta_{1,b} = \frac{\sum_{j=1}^{m_b} w_{jb}(\bar{Y}_{jb}(1) - \bar{Y}_{jb}(0))}{\sum_{j=1}^{m_b} w_{jb}} = \bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0).$$

We start by finding the exact form of the regression estimator $\hat{\beta}_1$. For our regression, we have $\boldsymbol{z}_{ijb} = (S_{ijb}\tilde{T}_{jb}, S_{ijb}) = (\tilde{T}_{jb}, 1)$ because we only have one block. Here we define

$$p_b^* = \frac{1}{w_b}\sum_{j=1}^{m_b} T_{jb}w_{jb} = \frac{w_b^1}{w_b}$$

and

$$\tilde{T}_{jb} = T_{jb} - p_b^*,$$

so that

$$\sum_{j=1}^{m_b} w_{j_b}\tilde{T}_{jb} = \sum_{j=1}^{m_b} w_{j_b}\left(T_{jb} - p_b^*\right) = 0.$$

The estimated parameter vector from weighted least squares regressing the individual y_{ijb} on an intercept and centered treatment indicator $\tilde{T}_{jb}$ with weights w_{ijb} is

$$\begin{pmatrix}\hat{\beta}_{1,b}\\ \hat{\beta}_{0,b}\end{pmatrix} = \left[\left(\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\boldsymbol{z}'_{ijb}\boldsymbol{z}_{ijb}\right)^{-1}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\boldsymbol{z}'_{ijb}y_{ijb}\right].$$

Result: $\hat{\beta}_{1,b} = \bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0)$.

We have by simple algebra simplifications and matrix inversion,

$$\begin{aligned}
\begin{pmatrix}\hat{\beta}_{1,b}\\ \hat{\beta}_{0,b}\end{pmatrix} &= \begin{pmatrix}\sum_{j=1}^{m_b} w_{jb}\tilde{T}_{jb}^2 & \sum_{j=1}^{m_b} w_{jb}\tilde{T}_{jb}\\ \sum_{j=1}^{m_b} w_{jb}\tilde{T}_{jb} & \sum_{j=1}^{m_b} w_{jb}\end{pmatrix}^{-1}\begin{pmatrix}\sum_{j=1}^{m_b} w_{jb}\tilde{T}_{jb}\bar{y}_{jb}\\ \sum_{j=1}^{m_b} w_{jb}\bar{y}_{jb}\end{pmatrix}\\
&= \begin{pmatrix}\sum_{j=1}^{m_b} w_{jb}\tilde{T}_{jb}^2 & 0\\ 0 & w_b\end{pmatrix}^{-1}\begin{pmatrix}\sum_{j=1}^{m_b} w_{jb}\tilde{T}_{jb}\bar{y}_{jb}\\ w_b^1\bar{\bar{y}}_b(1) + w_b^0\bar{\bar{y}}_b(0)\end{pmatrix}\\
&= \begin{pmatrix}\frac{1}{\sum_{j=1}^{m_b} w_{jb}\tilde{T}_{jb}^2} & 0\\ 0 & \frac{1}{w_b}\end{pmatrix}\begin{pmatrix}\sum_{j=1}^{m_b} w_{jb}\tilde{T}_{jb}\bar{y}_{jb}\\ w_b^1\bar{\bar{y}}_b(1) + w_b^0\bar{\bar{y}}_b(0)\end{pmatrix}
\end{aligned}$$

To simplify this final form, note a few useful algebra results:

$$\begin{aligned}
\sum_{j=1}^{m_b} w_{jb}\tilde{T}_{jb}^2 &= \sum_{j=1}^{m_b} w_{jb}\left(T_{jb}^2 - 2T_{jb}p_b^* + (p_b^*)^2\right)\\
&= w_b^1 - \frac{(w_b^1)^2}{w_b}\\
&= \frac{w_b^1 w_b^0}{w_b}
\end{aligned}$$

and

$$\begin{aligned}
\sum_{j=1}^{m_b} w_{jb}\tilde{T}_{jb}\bar{y}_{jb} &= \sum_{j=1}^{m_b} w_{jb}(T_{jb} - p_b^*)\bar{y}_{jb} \\
&= w_b^1\bar{\bar{y}}_b(1) - \frac{w_b^1}{w_b}(w_b^1\bar{\bar{y}}_b(1) + w_b^0\bar{\bar{y}}_b(0)) \\
&= \frac{w_b^1 w_b^0}{w_b}\left(\bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0)\right).
\end{aligned}$$

We then have the following for our regression estimator for the treatment impact (and control mean):

$$\begin{aligned}
\begin{pmatrix} \hat{\beta}_{1,b} \\ \hat{\beta}_{0,b} \end{pmatrix} &= \begin{pmatrix} \frac{w_b}{w_b^1 w_b^0} & 0 \\ 0 & \frac{1}{w_b} \end{pmatrix} \begin{pmatrix} \frac{w_b^1 w_b^0}{w_b}\left(\bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0)\right) \\ w_b^1\bar{\bar{y}}_b(1) + w_b^0\bar{\bar{y}}_b(0) \end{pmatrix} \\
&= \begin{pmatrix} \bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) \\ \frac{w_b^1}{w_b}\bar{\bar{y}}_b(1) + \frac{w_b^0}{w_b}\bar{\bar{y}}_b(0) \end{pmatrix}.
\end{aligned}$$

A.2.2 Finite sample bias

Here we explore the finite sample bias of the regression estimator. Using Equation 7 in Aronow and Middleton (2013), originally from Hartley and Ross (1954), the bias for a ratio estimator is

$$E\left[\frac{u}{v}\right] = \frac{1}{E[v]}\left[E[u] - \mathrm{Cov}\left(\frac{u}{v}, v\right)\right],$$

assuming that $v > 0$.

We then have, assuming that $w^1 > 0$ and $w^0 > 0$,

$$\begin{aligned}
E[\bar{\bar{y}}_b(1)] &= E\left[\frac{\sum_{j=1}^{m_b} T_{jb}w_{jb}\bar{y}_{bj}}{\sum_{j=1}^{m_b} w_{jb}T_{jb}}\right] \\
&= \frac{1}{\frac{p_b}{m_b^1}\sum_{j=1}^{m_b} w_{jb}}\left[\frac{p_b}{m_b^1}\sum_{j=1}^{m_b} w_{jb}\bar{Y}_{jb}(1) - \mathrm{Cov}\left(\frac{\sum_{j=1}^{m_b} T_{jb}w_{jb}\bar{y}_{jb}}{\sum_{j=1}^{m_b} w_{jb}T_{jb}}, \frac{1}{m_b^1}\sum_{j=1}^{m_b} T_{jb}w_{jb}\right)\right] \\
&= \bar{\bar{Y}}_b(1) - \frac{m_b}{w_b}\mathrm{Cov}\left(\bar{\bar{y}}_b(1), \frac{w_b^1}{m_b^1}\right).
\end{aligned}$$

It is then straightforward to find the bias of our estimator as follows:

$$E[\hat{\beta}_{1,b}] - \beta_{1,b} = -\frac{m_b}{w_b}\mathrm{Cov}\left(\bar{\bar{y}}_b(1), \frac{w_b^1}{m_b^1}\right) + \frac{m_b}{w_b}\mathrm{Cov}\left(\bar{\bar{y}}_b(0), \frac{w_b^0}{m_b^0}\right).$$

This means, as pointed out by Aronow and Middleton (2013), that the bias will depend on (1) the covariance between cluster size (as captured by the w_{jb}) and outcomes and (2) the variability of cluster sizes.

A.2.3 Consistency

Here we give conditions for consistency of the regression estimator. It is useful to assume limiting values on the finite population parameters. This can be handled in a few ways. Following Aronow and Middleton (2013), we may envision a sequence of h finite populations such that as $h \to \infty$, the finite population increases by copying the original m_b clusters in block b h times and then randomization occurs independently within each copy with probability p_b. This keeps our estimands and parameters constant as the population grows.

A bit more generally, we may have limiting values for the finite population parameters. For a given finite population with m clusters, define the following quantities:

$$\overline{wY(1)}_b = \frac{1}{m_b} \sum_{j=1}^{m_b} w_{jb} \bar{Y}_{jb}(1),$$

$$\overline{wY(0)}_b = \frac{1}{m_b} \sum_{j=1}^{m_b} w_{jb} \bar{Y}_{jb}(0),$$

and

$$\overline{w}_b = \frac{1}{m_b} \sum_{j=1}^{m_b} w_{jb}.$$

Given these definitions, our $\beta_{1,b}$ is then

$$\beta_{1,b} = \frac{\overline{wY(1)}_b}{\overline{w}_b} - \frac{\overline{wY(0)}_b}{\overline{w}_b}.$$

Assume limiting values $\mu_b^*(1)$, $\mu_b^*(0)$, and ω_b such that as $m_b \to \infty$,

$$\overline{wY(1)}_b \xrightarrow{p} \mu_b^*(1),$$
$$\overline{wY(0)}_b \xrightarrow{p} \mu_b^*(0),$$

and

$$\overline{w}_b \xrightarrow{p} \omega_b > 0.$$

If we were normalizing the weights across the whole population, it could be the case that $\omega_b \to 0$; we eliminate this issue by using unnormalized weights (which don't change the estimand or estimator because we use weights in the numerator and denominator). This gives a limiting value for $\beta_{1,b}$ of $\beta_{1,b}^* = (\mu_b^*(1) - \mu_b^*(0))/\omega_b$.

We next show that $\hat{\beta}_{1,b} \xrightarrow{p} \beta_{1,b}^*$. Denote

$$S_{w,b}^2(1) = \frac{1}{m_b - 1} \sum_{j=1}^{m_b} \left(w_{jb} \bar{Y}_{jb}(1) - \overline{wY(1)}_b \right)^2,$$

$$S_{w,b}^2(0) = \frac{1}{m_b - 1} \sum_{j=1}^{m_b} \left(w_j \bar{Y}_{jb}(0) - \overline{wY(0)}_b \right)^2,$$

and

$$S^2(w_b) = \frac{1}{m_b - 1} \sum_{j=1}^{m_b} (w_{jb} - \overline{w}_b)^2 .$$

Assume that $m_b^1/m_b \to p_b \in (0,1)$. Using Theorem B from Scott and Wu (1981), under simple random sampling of clusters into treatment, if our weighted variances do not go to infinity as we increase sample size, specifically if as $m_b \to \infty$,

$$S^2_{w,b}(1)/m_b \to 0,$$
$$S^2_{w,b}(0)/m_b \to 0,$$

and

$$S^2(w_b)/m_b \to 0$$

then

$$\frac{1}{m_b^1} \sum_{j=1}^{m_b} w_{jb} T_{jb} \bar{Y}_{jb}(1) - \overline{wY(1)}_b \xrightarrow{p} 0,$$

$$\frac{1}{m_b^0} \sum_{j=1}^{m_b} w_{jb}(1 - T_{jb}) \bar{Y}_{jb}(0) - \overline{wY(0)}_b \xrightarrow{p} 0,$$

$$\frac{1}{m_b^1} \sum_{j=1}^{m_b} w_{jb} T_{jb} - \overline{w}_b \xrightarrow{p} 0,$$

and

$$\frac{1}{m_b^0} \sum_{j=1}^{m_b} w_{jb}(1 - T_{jb}) - \overline{w}_b \xrightarrow{p} 0$$

as $m_b \to \infty$.

We next need a small convergence lemma:

Lemma A.2.1. *If $A_m - B_m \xrightarrow{p} 0$ and $B_m \xrightarrow{p} k$, with k a constant, as $m \to \infty$ then $A_m \xrightarrow{p} k$ as $m \to \infty$.*

Proof.

$$A_m - k = A_m - B_m + B_m - k \xrightarrow{p} 0.$$

□

Using Lemma A.2.1, we have

$$\frac{1}{m_b^t} \sum_{j:T_{jb}=t} w_{jb} \bar{Y}_{jb}(t) = \frac{w_b^t}{m_b^t} \bar{\bar{y}}_b(t) \xrightarrow{p} \mu_b^*(t)$$

and

$$\frac{1}{m_b^t} \sum_{j:T_{jb}=t} w_{jb} = \frac{w_b^t}{m_b^t} \xrightarrow{p} \omega_b.$$

Then by Slutsky's theorem,

$$\hat{\beta}_{1,b} = \bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) \xrightarrow{p} \frac{\mu_b^*(1)}{\omega_b} - \frac{\mu_b^*(0)}{\omega_b} = \beta_{1,b}^*.$$

Hence in this setting our estimator is consistent.

A.2.4 Asymptotic normality of one ratio

We are interested in the asymptotic behavior of, for $t \in \{0, 1\}$,

$$\bar{\bar{y}}_b(t) = \frac{1}{\sum_{j=1}^{m_b} w_{jb} T_{jb}} \sum_{j=1}^{m_b} T_{jb} w_{jb} \bar{Y}_{jb}(t).$$

We use the notation from Section A.1.1 but now add subscripts b to indicate that we are referring to block b. For $t \in \{1, 0\}$, let $C_{jb}(t) = \bar{Y}_{jb}(t)$, so that $\bar{\bar{c}}_b(t) = \bar{\bar{y}}_b(t)$ and $\overline{\overline{C}}_b(t) = \bar{\bar{Y}}_b(t)$.

Corollary A.2.2. *Under the conditions of Lemma A.1.1, we have*

$$\begin{aligned} \frac{\bar{\bar{c}}_b(t) - \overline{\overline{C}}_b(t)}{\sqrt{\frac{1-f_b^t}{m_b^t}} S_{R,b}(t)/\overline{w}} &= \frac{(\bar{\bar{y}}_b(t) - \bar{\bar{Y}}_b(t))}{\sqrt{\frac{1-f_b^t}{m_b^t}} \sqrt{\frac{1}{m_b-1} \sum_{j=1}^{m_b} (R_{jb}(t)/\overline{w}_b)^2}} \\ &\xrightarrow{d} N(0,1). \end{aligned}$$

The term in the denominator expands as follows:

$$\begin{aligned} \frac{1}{m_b - 1} \sum_{j=1}^{m_b} (R_{jb}(t)/\overline{w}_b)^2 &= \frac{1}{m_b - 1} \sum_{j=1}^{m_b} \left(\left(w_{jb} C_{jb}(t) - w_{jb} \overline{\overline{C}}_b(t) \right) / \overline{w}_b \right)^2 \\ &= \frac{1}{(m_b - 1)} \sum_{j=1}^{m_b} \frac{w_{jb}^2}{\overline{w}_b^2} \left(\bar{Y}_{jb}(t) - \bar{\bar{Y}}_b(t) \right)^2. \end{aligned}$$

A.2.5 Asymptotic normality

We are interested in the ratio difference estimator, $\bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0)$.

Let $D_{jb}(t) = (w_{jb} C_{jb}(t) - w_{jb} \overline{\overline{C}}_b(t))/\overline{w}_b = w_{jb}(\bar{Y}_{jb}(t) - \bar{\bar{Y}}_b(t))/\overline{w}_b$, with $C_{jb}(t)$ defined as in Section A.2.4. Note that this means that we still have $\bar{\bar{c}}_b(t) = \bar{\bar{y}}_b(t)$ and $\overline{\overline{C}}_b(t) = \bar{\bar{Y}}_b(t)$. Now we have reformulated our complicated difference in ratio estimators into a form that's easier to work with.

Corollary A.2.3. *Under the conditions of Lemma A.1.2, we have*

$$\begin{aligned} \frac{\left(\bar{\bar{c}}_b(1) - \overline{\overline{C}}_b(1)\right) - \left(\bar{\bar{c}}_b(0) - \overline{\overline{C}}_b(0)\right)}{\sqrt{Var(\hat{D}_b)}} &= \frac{\left(\bar{\bar{y}}_b(1) - \bar{\bar{Y}}_b(1)\right) - \left(\bar{\bar{y}}_b(0) - \bar{\bar{Y}}_b(0)\right)}{\sqrt{Var(\hat{D}_b)}} \\ &\xrightarrow{d} N(0,1). \end{aligned}$$

The denominator simplifies as follows:

$$\begin{aligned}
\mathrm{Var}(\hat{D}_b) =& \frac{S^2_{D,b}(1)}{m^1_b} + \frac{S^2_{D,b}(0)}{m^0_b} - \frac{S^2(D_b)}{m_b} \\
=& \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} D^2_{jb}(1)}{m^1_b} + \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} D^2_{jb}(0)}{m^0_b} - \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b}(D_{jb}(1) - D_{jb}(0))^2}{m_b} \\
=& \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} \left(w_{jb}(\bar{Y}_{jb}(1) - \bar{\bar{Y}}_b(1))/\overline{w}_b\right)^2}{m^1_b} + \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} \left(w_{jb}(\bar{Y}_{jb}(0) - \bar{\bar{Y}}_b(0))/\overline{w}_b\right)^2}{m^0_b} \\
&- \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} \left(w_{jb}(\bar{Y}_{jb}(1) - \bar{\bar{Y}}_b(1))/\overline{w}_b - w_{jb}(\bar{Y}_{jb}(0) - \bar{\bar{Y}}_b(0))/\overline{w}_b\right)^2}{m_b} \\
=& \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} \frac{w^2_{jb}}{\overline{w}^2_b}\left(\bar{Y}_{jb}(1) - \bar{\bar{Y}}_b(1)\right)^2}{m^1_b} + \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} \frac{w^2_{jb}}{\overline{w}^2_b}\left(\bar{Y}_{jb}(0) - \bar{\bar{Y}}_b(0))\right)^2}{m^0_b} \\
&- \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} \frac{w^2_{jb}}{\overline{w}^2_b}\left(\bar{Y}_{jb}(1) - \bar{Y}_{jb}(0) - (\bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0))\right)^2}{m_b}.
\end{aligned}$$

A.3 Theorem 1 with covariates, multiple blocks

In this section we explore the impact estimator from a weighted least squares regression with additional covariate adjustment. We start by deriving the closed form of this estimator, which is $\hat{\beta}_{1,b} = \bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) - \left(\overline{\overline{\boldsymbol{x}}}^1_b - \overline{\overline{\boldsymbol{x}}}^0_b\right)\hat{\boldsymbol{\gamma}}$ for block b. We then find consistency results and show the asymptotic normality result given in Theorem 1. We do not specifically include interactions between the additional covariates and the block indicators as this immediately follows by generating the fully interacted set of covariates and then using our results on that extended set.

A.3.1 The estimator

We now have multiple blocks with covariate adjustment but not interactions between blocks and additional covariates in our regression. However, we continue to include an interaction between block and treatment. We assume that $m_b/m \xrightarrow{p} q_b$ where $0 < q_b < 1$. Let there be h blocks. Accordingly, $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_0$ are vectors with an entry for each block. Thus we have $\hat{\beta}_{1,b}$, the bth entry of $\hat{\boldsymbol{\beta}}_1$, is still the treatment effect estimator for block b.

Now let $\tilde{\boldsymbol{x}}_{ijb} = \boldsymbol{x}_{ijb} - \overline{\overline{\boldsymbol{x}}}_b$ with

$$\overline{\overline{\boldsymbol{x}}}_b = \frac{1}{w_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\boldsymbol{x}_{ijb}$$

and $\tilde{\overline{\boldsymbol{x}}}_{jb} = \overline{\boldsymbol{x}}_{jb} - \overline{\overline{\boldsymbol{x}}}_b$ with

$$\overline{\boldsymbol{x}}_{jb} = \frac{1}{w_{jb}}\sum_{i=1}^{n_{jb}} w_{ijb}\boldsymbol{x}_{ijb}.$$

We also define

$$\bar{\bar{\boldsymbol{x}}}_b^1 = \frac{\sum_{j=1}^{m_b} w_{jb} T_{jb} \overline{\boldsymbol{x}}_{jb}}{\sum_{j=1}^{m_b} w_{jb} T_{jb}}$$

and

$$\bar{\bar{\boldsymbol{x}}}_b^0 = \frac{\sum_{j=1}^{m_b} w_{jb}(1 - T_{jb}) \overline{\boldsymbol{x}}_{jb}}{\sum_{j=1}^{m_b} w_{jb}(1 - T_{jb})}.$$

We have $\boldsymbol{z}_{ijb} = (S_{ij1}\tilde{T}_{j1}, \ldots, S_{ijh}\tilde{T}_{jh}, S_{ij1}, \ldots, S_{ijh}, \tilde{\boldsymbol{x}}_{ijb})$. The estimated parameter vector is

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_0 \end{pmatrix} = \left[\left(\sum_{b=1}^{h} \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb} \boldsymbol{z}'_{ijb} \boldsymbol{z}_{ijb} \right)^{-1} \sum_{b=1}^{h} \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb} \boldsymbol{z}'_{ijb} y_{ijb} \right].$$

Result:

$$\hat{\beta}_{1,b} = \bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) - \left(\bar{\bar{\boldsymbol{x}}}_b^1 - \bar{\bar{\boldsymbol{x}}}_b^0 \right) \hat{\boldsymbol{\gamma}}$$

Remark. When we have interactions between blocks and additional covariates, we will have a different $\hat{\boldsymbol{\gamma}}$ for each block and the estimators for each block will be independent. In that case, $\hat{\beta}_{1,b}$ is the same as the estimator as if we had only run the regression with block b, i.e. as if we only had one block. Therefore, these results directly extend to that case.

We have

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_0 \\ \hat{\boldsymbol{\gamma}} \end{pmatrix}$$

$$= \begin{pmatrix}
\sum_{j=1}^{m_1} w_{j1} \tilde{T}_{j1}^2 & \cdots & 0 & 0 & \cdots & 0 & \sum_{j=1}^{m_1} w_{j1} \tilde{T}_{j1} \bar{\tilde{\boldsymbol{x}}}_{j1} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & \sum_{j=1}^{m_h} w_{jh} \tilde{T}_{jh}^2 & 0 & \cdots & 0 & \sum_{j=1}^{m_h} w_{jh} \tilde{T}_{jh} \bar{\tilde{\boldsymbol{x}}}_{jh} \\
0 & \cdots & 0 & w_1 & \cdots & 0 & \mathbf{0}_v \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \mathbf{0}_v \\
0 & \cdots & 0 & 0 & \cdots & w_h & \mathbf{0}_v \\
\sum_{j=1}^{m_1} w_{j1} \tilde{T}_{j1} \bar{\tilde{\boldsymbol{x}}}'_{j1} & \cdots & \sum_{j=1}^{m_h} w_{jh} \tilde{T}_{jh} \bar{\tilde{\boldsymbol{x}}}'_{jh} & \mathbf{0}'_v & \cdots & \mathbf{0}'_v & \sum_{b=1}^{h} \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb} \tilde{\boldsymbol{x}}'_{ijb} \tilde{\boldsymbol{x}}_{ijb}
\end{pmatrix}^{-1}$$

$$\times \begin{pmatrix}
\sum_{j=1}^{m_1} w_{j1} \tilde{T}_{j1} \bar{y}_{j1} \\
\vdots \\
\sum_{j=1}^{m_h} w_{jh} \tilde{T}_{jh} \bar{y}_{jh} \\
\sum_{j=1}^{m_1} w_{j1} \bar{y}_{j1} \\
\vdots \\
\sum_{j=1}^{m_h} w_{jh} \bar{y}_{jh} \\
\sum_{b=1}^{h} \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb} \tilde{\boldsymbol{x}}'_{ijb} y_{ijb}
\end{pmatrix}.$$

We start by performing the matrix inversion. We can break this matrix into the following blocks:

$$
\begin{aligned}
\boldsymbol{A} &= \begin{pmatrix}
\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}^2 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}^2 & 0 & \cdots & 0 \\
0 & \cdots & 0 & w_1 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & w_h
\end{pmatrix} \\
&= \begin{pmatrix}
\frac{w_1^1 w_1^0}{w_1} & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \frac{w_h^1 w_h^0}{w_h} & 0 & \cdots & 0 \\
0 & \cdots & 0 & w_1 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & w_h
\end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{B} &= \begin{pmatrix}
\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}\tilde{\overline{\boldsymbol{x}}}_{j1} \\
\vdots \\
\sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}\tilde{\overline{\boldsymbol{x}}}_{jh} \\
\boldsymbol{0}_v \\
\vdots \\
\boldsymbol{0}_v
\end{pmatrix}
= \begin{pmatrix}
\frac{w_1^1 w_1^0}{w_1}\left(\overline{\overline{\boldsymbol{x}}}_1^1 - \overline{\overline{\boldsymbol{x}}}_1^0\right) \\
\vdots \\
\frac{w_h^1 w_h^0}{w_h}\left(\overline{\overline{\boldsymbol{x}}}_h^1 - \overline{\overline{\boldsymbol{x}}}_h^0\right) \\
\boldsymbol{0}_v \\
\vdots \\
\boldsymbol{0}_v
\end{pmatrix} \\
\boldsymbol{C} &= \begin{pmatrix} \sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}\tilde{\overline{\boldsymbol{x}}}_{j1}' & \cdots & \sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}\tilde{\overline{\boldsymbol{x}}}_{jh}' & \boldsymbol{0}_v' & \cdots & \boldsymbol{0}_v' \end{pmatrix} \\
&= \begin{pmatrix} \frac{w_1^1 w_1^0}{w_1}\left(\overline{\overline{\boldsymbol{x}}}_1^1 - \overline{\overline{\boldsymbol{x}}}_1^0\right)' & \cdots & \frac{w_h^1 w_h^0}{w_h}\left(\overline{\overline{\boldsymbol{x}}}_h^1 - \overline{\overline{\boldsymbol{x}}}_h^0\right)' & \boldsymbol{0}_v' & \cdots & \boldsymbol{0}_v' \end{pmatrix} \\
\boldsymbol{D} &= \left(\sum_{b=1}^{h} \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb}\tilde{\boldsymbol{x}}_{ijb}'\tilde{\boldsymbol{x}}_{ijb} \right).
\end{aligned}
$$

We can then use the following matrix inversion formula:

$$
\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}^{-1} = \begin{pmatrix} (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} & (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \end{pmatrix}. \tag{S7}
$$

To make things easier, we will first derive a result for

$$\hat{\boldsymbol{\beta}}_1 + \begin{pmatrix} \overline{\overline{\boldsymbol{x}}}_1^1 - \overline{\overline{\boldsymbol{x}}}_1^0 \\ \vdots \\ \overline{\overline{\boldsymbol{x}}}_h^1 - \overline{\overline{\boldsymbol{x}}}_h^0 \end{pmatrix} \hat{\boldsymbol{\gamma}}.$$

We have, based on our matrix inversion, that

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_1 &= \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix} \left((\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} \quad -\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \right) \\
&\quad \times \begin{pmatrix} \sum_{j=1}^{m_1} w_{j1} \tilde{T}_{j1} \bar{y}_{j1} \\ \vdots \\ \sum_{j=1}^{m_h} w_{jh} \tilde{T}_{jh} \bar{y}_{jh} \\ \sum_{j=1}^{m_1} w_{j1} \bar{y}_{j1} \\ \vdots \\ \sum_{j=1}^{m_h} w_{jh} \bar{y}_{jh} \\ \sum_{b=1}^{h} \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb} \tilde{\boldsymbol{x}}'_{ijb} y_{ijb} \end{pmatrix} \\
&= \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix} \left((\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} \quad -\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \right) \\
&\quad \times \begin{pmatrix} \frac{w_1^1 w_1^0}{w_1} \left(\bar{\bar{y}}_1(1) - \bar{\bar{y}}_1(0) \right) \\ \vdots \\ \frac{w_h^1 w_h^0}{w_h} \left(\bar{\bar{y}}_h(1) - \bar{\bar{y}}_h(0) \right) \\ \sum_{j=1}^{m_1} w_{j1} \bar{y}_{j1} \\ \vdots \\ \sum_{j=1}^{m_h} w_{jh} \bar{y}_{jh} \\ \sum_{b=1}^{h} \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb} \tilde{\boldsymbol{x}}'_{ijb} y_{ijb} \end{pmatrix}
\end{aligned}$$

and

$$\begin{aligned}
&\begin{pmatrix} \overline{\overline{\boldsymbol{x}}}_1^1 - \overline{\overline{\boldsymbol{x}}}_1^0 \\ \vdots \\ \overline{\overline{\boldsymbol{x}}}_h^1 - \overline{\overline{\boldsymbol{x}}}_h^0 \end{pmatrix} \hat{\boldsymbol{\gamma}} \\
&= \begin{pmatrix} \overline{\overline{\boldsymbol{x}}}_1^1 - \overline{\overline{\boldsymbol{x}}}_1^0 \\ \vdots \\ \overline{\overline{\boldsymbol{x}}}_h^1 - \overline{\overline{\boldsymbol{x}}}_h^0 \end{pmatrix} \left(-\boldsymbol{D}^{-1}\boldsymbol{C}(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} \quad (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \right) \begin{pmatrix} \frac{w_1^1 w_1^0}{w_1} \left(\bar{\bar{y}}_1(1) - \bar{\bar{y}}_1(0) \right) \\ \vdots \\ \frac{w_h^1 w_h^0}{w_h} \left(\bar{\bar{y}}_h(1) - \bar{\bar{y}}_h(0) \right) \\ \sum_{j=1}^{m_1} w_{j1} \bar{y}_{j1} \\ \vdots \\ \sum_{j=1}^{m_h} w_{jh} \bar{y}_{jh} \\ \sum_{b=1}^{h} \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb} \tilde{\boldsymbol{x}}'_{ijb} y_{ijb} \end{pmatrix}.
\end{aligned}$$

We see there are a lot of common terms when we add these expressions. Let's first simplify

$$-\begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix} \boldsymbol{A}^{-1}\boldsymbol{B} + \begin{pmatrix} \overline{\overline{\boldsymbol{x}}}_1^1 - \overline{\overline{\boldsymbol{x}}}_1^0 \\ \vdots \\ \overline{\overline{\boldsymbol{x}}}_h^1 - \overline{\overline{\boldsymbol{x}}}_h^0 \end{pmatrix}.$$

$$\begin{aligned}
\begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix} \boldsymbol{A}^{-1}\boldsymbol{B} &= \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \frac{w_1}{w_1^1 w_1^0} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{w_h}{w_h^1 w_h^0} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \frac{1}{w_1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \frac{1}{w_h} \end{pmatrix} \\
&\quad \times \begin{pmatrix} \frac{w_1^1 w_1^0}{w_1}\left(\overline{\overline{\boldsymbol{x}}}_1^1 - \overline{\overline{\boldsymbol{x}}}_1^0\right) \\ \vdots \\ \frac{w_h^1 w_h^0}{w_h}\left(\overline{\overline{\boldsymbol{x}}}_h^1 - \overline{\overline{\boldsymbol{x}}}_h^0\right) \\ \boldsymbol{0}_v \\ \vdots \\ \boldsymbol{0}_v \end{pmatrix} \\
&= \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \overline{\overline{\boldsymbol{x}}}_1^1 - \overline{\overline{\boldsymbol{x}}}_1^0 \\ \vdots \\ \overline{\overline{\boldsymbol{x}}}_h^1 - \overline{\overline{\boldsymbol{x}}}_h^0 \\ \boldsymbol{0}_v \\ \vdots \\ \boldsymbol{0}_v \end{pmatrix} \\
&= \begin{pmatrix} \overline{\overline{\boldsymbol{x}}}_1^1 - \overline{\overline{\boldsymbol{x}}}_1^0 \\ \vdots \\ \overline{\overline{\boldsymbol{x}}}_h^1 - \overline{\overline{\boldsymbol{x}}}_h^0 \end{pmatrix}
\end{aligned}$$

We see that

$$-\begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix} \boldsymbol{A}^{-1}\boldsymbol{B} + \begin{pmatrix} \overline{\overline{\boldsymbol{x}}}_1^1 - \overline{\overline{\boldsymbol{x}}}_1^0 \\ \vdots \\ \overline{\overline{\boldsymbol{x}}}_h^1 - \overline{\overline{\boldsymbol{x}}}_h^0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{0}_v \\ \vdots \\ \boldsymbol{0}_v \end{pmatrix}.$$

Now we need to simplify

$$-\begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix} - \begin{pmatrix} \overline{\overline{\boldsymbol{x}}}_1^1 - \overline{\overline{\boldsymbol{x}}}_1^0 \\ \vdots \\ \overline{\overline{\boldsymbol{x}}}_h^1 - \overline{\overline{\boldsymbol{x}}}_h^0 \end{pmatrix} \boldsymbol{D}^{-1}\boldsymbol{C}.$$

But first let's look at $\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}$.

$$\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}=\begin{pmatrix}\frac{w_1^1w_1^0}{w_1} & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & \frac{w_h^1w_h^0}{w_h} & 0 & \cdots & 0\\ 0 & \cdots & 0 & w_1 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & 0 & \cdots & w_h\end{pmatrix}-\begin{pmatrix}\frac{w_1^1w_1^0}{w_1}\left(\overline{\overline{\boldsymbol{x}}}_1^1-\overline{\overline{\boldsymbol{x}}}_1^0\right)\\ \vdots\\ \frac{w_h^1w_h^0}{w_h}\left(\overline{\overline{\boldsymbol{x}}}_h^1-\overline{\overline{\boldsymbol{x}}}_h^0\right)\\ \boldsymbol{0}_v\\ \vdots\\ \boldsymbol{0}_v\end{pmatrix}\boldsymbol{D}^{-1}\boldsymbol{C}$$

$$=\begin{pmatrix}\frac{w_1^1w_1^0}{w_1} & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & \frac{w_h^1w_h^0}{w_h} & 0 & \cdots & 0\\ 0 & \cdots & 0 & w_1 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & 0 & \cdots & w_h\end{pmatrix}\left(\begin{pmatrix}1 & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 1 & 0 & \cdots & 0\\ 0 & \cdots & 0 & 1 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & 0 & \cdots & 1\end{pmatrix}-\begin{pmatrix}\left(\overline{\overline{\boldsymbol{x}}}_1^1-\overline{\overline{\boldsymbol{x}}}_1^0\right)\\ \vdots\\ \left(\overline{\overline{\boldsymbol{x}}}_h^1-\overline{\overline{\boldsymbol{x}}}_h^0\right)\\ \boldsymbol{0}_v\\ \vdots\\ \boldsymbol{0}_v\end{pmatrix}\boldsymbol{D}^{-1}\boldsymbol{C}\right)$$

Now let's return to

$$\begin{pmatrix}1 & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 1 & 0 & \cdots & 0\\ 0 & \cdots & 0 & 1 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & 0 & \cdots & 1\end{pmatrix}-\begin{pmatrix}\overline{\overline{\boldsymbol{x}}}_1^1-\overline{\overline{\boldsymbol{x}}}_1^0\\ \vdots\\ \overline{\overline{\boldsymbol{x}}}_h^1-\overline{\overline{\boldsymbol{x}}}_h^0\end{pmatrix}\boldsymbol{D}^{-1}\boldsymbol{C}.$$

$$\begin{pmatrix}1 & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 1 & 0 & \cdots & 0\\ 0 & \cdots & 0 & 1 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & 0 & \cdots & 1\end{pmatrix}-\begin{pmatrix}\overline{\overline{\boldsymbol{x}}}_1^1-\overline{\overline{\boldsymbol{x}}}_1^0\\ \vdots\\ \overline{\overline{\boldsymbol{x}}}_h^1-\overline{\overline{\boldsymbol{x}}}_h^0\end{pmatrix}\boldsymbol{D}^{-1}\boldsymbol{C}=\begin{pmatrix}\frac{w_1}{w_1^1w_1^0} & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & \frac{w_2}{w_2^1w_2^0} & 0 & \cdots & 0\end{pmatrix}\left(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)$$

Hence,

$$\left(\begin{pmatrix}1 & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 1 & 0 & \cdots & 0\\ 0 & \cdots & 0 & 1 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & 0 & \cdots & 1\end{pmatrix}-\begin{pmatrix}\overline{\overline{\boldsymbol{x}}}_1^1-\overline{\overline{\boldsymbol{x}}}_1^0\\ \vdots\\ \overline{\overline{\boldsymbol{x}}}_h^1-\overline{\overline{\boldsymbol{x}}}_h^0\end{pmatrix}\boldsymbol{D}^{-1}\boldsymbol{C}\right)\left(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1}=\begin{pmatrix}\frac{w_1}{w_1^1w_1^0} & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & \frac{w_2}{w_2^1w_2^0} & 0 & \cdots & 0\end{pmatrix}.$$

Putting it all together:

$$\hat{\boldsymbol{\beta}}_1 + \begin{pmatrix} \overline{\overline{\boldsymbol{x}}}_1^1 - \overline{\overline{\boldsymbol{x}}}_1^0 \\ \vdots \\ \overline{\overline{\boldsymbol{x}}}_h^1 - \overline{\overline{\boldsymbol{x}}}_h^0 \end{pmatrix} \hat{\boldsymbol{\gamma}} = \begin{pmatrix} \frac{w_1}{w_1^1 w_1^0} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{w_2}{w_2^1 w_2^0} & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \frac{w_1^1 w_1^0}{w_1}\left(\bar{\bar{y}}_1(1) - \bar{\bar{y}}_1(0)\right) \\ \vdots \\ \frac{w_h^1 w_h^0}{w_h}\left(\bar{\bar{y}}_h(1) - \bar{\bar{y}}_h(0)\right) \\ \sum_{j=1}^{m_1} w_{j1}\bar{y}_{j1} \\ \vdots \\ \sum_{j=1}^{m_h} w_{jh}\bar{y}_{jh} \\ \sum_{b=1}^{h}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\tilde{\boldsymbol{x}}'_{ijb} y_{ijb} \end{pmatrix}$$

$$= \begin{pmatrix} \bar{\bar{y}}_1(1) - \bar{\bar{y}}_1(0) \\ \vdots \\ \bar{\bar{y}}_h(1) - \bar{\bar{y}}_h(0) \end{pmatrix}.$$

Hence, we have the desired result,

$$\hat{\boldsymbol{\beta}}_1 = \begin{pmatrix} \bar{\bar{y}}_1(1) - \bar{\bar{y}}_1(0) \\ \vdots \\ \bar{\bar{y}}_h(1) - \bar{\bar{y}}_h(0) \end{pmatrix} - \begin{pmatrix} \overline{\overline{\boldsymbol{x}}}_1^1 - \overline{\overline{\boldsymbol{x}}}_1^0 \\ \vdots \\ \overline{\overline{\boldsymbol{x}}}_h^1 - \overline{\overline{\boldsymbol{x}}}_h^0 \end{pmatrix} \hat{\boldsymbol{\gamma}}.$$

A.3.2 Consistency

Here we will show that $\hat{\beta}_{1,b} \xrightarrow{p} \frac{1}{\omega_b}\left(\mu_b^*(1) - \mu_b^*(0)\right) = \beta_{1,b}^*$. We assume the same limiting values for our average potential outcomes and weights as in Section A.2.3. Assume that we have finite limiting values on the following weighted variance/covariance expressions, denoted as follows:

$$\boldsymbol{S}^2_{\boldsymbol{x},b} = \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\left(\boldsymbol{x}_{ijb} - \overline{\overline{\boldsymbol{x}}}_b\right)'\left(\boldsymbol{x}_{ijb} - \overline{\overline{\boldsymbol{x}}}_b\right) \xrightarrow{p} \boldsymbol{\Sigma}^2_{\boldsymbol{x},b}$$

and

$$\boldsymbol{S}^2_{\boldsymbol{x},Y,b}(t) = \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\boldsymbol{x}'_{ijb}Y_{ijb}(t) - \overline{\overline{\boldsymbol{x}}}'_b\overline{wY(t)}_b \xrightarrow{p} \boldsymbol{\Sigma}^2_{\boldsymbol{x},Y(t),b}.$$

Assume we also have a (unnamed) limiting value on the following variance expression:

$$\boldsymbol{S}^2_{\boldsymbol{x}Y,b}(t) = \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}}\left(w_{ijb}\boldsymbol{x}'_{ijb}Y_{ijb}(t) - \overline{wxY(t)}_b\right)^2$$

with

$$\overline{wxY(t)}_b = \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\boldsymbol{x}'_{ijb}Y_{ijb}(t).$$

Also assume that we have finite limiting values on the variances for the potential outcomes. Further assume we have limiting values $\overline{\boldsymbol{X}}^*_b$, $\overline{\boldsymbol{X}'\boldsymbol{X}}^*_b$, and $\overline{\boldsymbol{X}\mu}_{wb}(t)$ such that $\frac{1}{m_b}\sum_{j=1}^{m_b} w_{jb}\overline{\boldsymbol{x}}_{jb} \xrightarrow{p} \overline{\boldsymbol{X}}^*_b$, $\frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\boldsymbol{x}'_{ijb}\boldsymbol{x}_{ijb} \xrightarrow{p} \overline{\boldsymbol{X}'\boldsymbol{X}}^*_b$, and $\frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\tilde{\boldsymbol{x}}'_{ijb}Y_{ijb}(t) \xrightarrow{p} \overline{\boldsymbol{X}\mu}^*_b(t)$. Again, we assume that $m_b/m \xrightarrow{p} q_b$ where $0 < q_b < 1$, such that each block is growing to infinity with $m = \sum_{b=1}^h m_b$. In Section A.2.3, we already showed that $\bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) \xrightarrow{p} \mu_b^*(1) - \mu_b^*(0)$. Thus, we now need to examine the extra term in this new $\hat{\beta}_{1,b}$, for which we need to simplify the asymptotic form of $\hat{\boldsymbol{\gamma}}$.

$$
\begin{aligned}
&\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_0 \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} \\
&= \begin{pmatrix}
\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}^2 & \cdots & 0 & 0 & \cdots & 0 & \sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}\tilde{\bar{\boldsymbol{x}}}_{j1} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & \sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}^2 & 0 & \cdots & 0 & \sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}\tilde{\bar{\boldsymbol{x}}}_{jh} \\
0 & \cdots & 0 & w_1 & \cdots & 0 & \mathbf{0}_v \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \mathbf{0}_v \\
0 & \cdots & 0 & 0 & \cdots & w_h & \mathbf{0}_v \\
\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}\tilde{\bar{\boldsymbol{x}}}_{j1}' & \cdots & \sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}\tilde{\bar{\boldsymbol{x}}}_{jh}' & \mathbf{0}_v' & \cdots & \mathbf{0}_v' & \sum_{b=1}^h \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb}\tilde{\boldsymbol{x}}_{ijb}'\tilde{\boldsymbol{x}}_{ijb}
\end{pmatrix}^{-1} \\
&\quad \times \begin{pmatrix}
\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}\bar{y}_{j1} \\
\vdots \\
\sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}\bar{y}_{jh} \\
\sum_{j=1}^{m_1} w_{j1}\bar{y}_{j1} \\
\vdots \\
\sum_{j=1}^{m_h} w_{jh}\bar{y}_{jh} \\
\sum_{b=1}^h \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb}\tilde{\boldsymbol{x}}_{ijb}' y_{ijb}
\end{pmatrix} \\
&= \begin{pmatrix}
\frac{1}{m_1}\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}^2 & \cdots & 0 & 0 & \cdots & 0 & \frac{1}{m_1}\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}\tilde{\bar{\boldsymbol{x}}}_{j1} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & \frac{1}{m_h}\sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}^2 & 0 & \cdots & 0 & \frac{1}{m_h}\sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}\tilde{\bar{\boldsymbol{x}}}_{jh} \\
0 & \cdots & 0 & \frac{1}{m_1}w_1 & \cdots & 0 & \mathbf{0}_v \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \mathbf{0}_v \\
0 & \cdots & 0 & 0 & \cdots & \frac{1}{m_h}w_h & \mathbf{0}_v \\
\frac{1}{m}\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}\tilde{\bar{\boldsymbol{x}}}_{j1}' & \cdots & \frac{1}{m}\sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}\tilde{\bar{\boldsymbol{x}}}_{jh}' & \mathbf{0}_v' & \cdots & \mathbf{0}_v' & \frac{1}{m}\sum_{b=1}^h \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb}\tilde{\boldsymbol{x}}_{ijb}'\tilde{\boldsymbol{x}}_{ijb}
\end{pmatrix}^{-1} \\
&\quad \times \begin{pmatrix}
\frac{1}{m_1}\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}\bar{y}_{j1} \\
\vdots \\
\frac{1}{m_h}\sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}\bar{y}_{jh} \\
\frac{1}{m_1}\sum_{j=1}^{m_1} w_{j1}\bar{y}_{j1} \\
\vdots \\
\frac{1}{m_h}\sum_{j=1}^{m_h} w_{jh}\bar{y}_{jh} \\
\frac{1}{m}\sum_{b=1}^h \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb}\tilde{\boldsymbol{x}}_{ijb}' y_{ijb}
\end{pmatrix}.
\end{aligned}
$$

We have

$$\frac{1}{m_b}\sum_{j=1}^{m_b} w_{jb}\tilde{T}_{jb}^2 = \frac{1}{m_b}\frac{w_b^1 w_b^0}{w_b} \xrightarrow{p} p_b(1-p_b)\omega_b,$$

$$\frac{1}{m_b}\sum_{j=1}^{m_b} w_{jb} \xrightarrow{p} \omega_b,$$

and

$$\frac{1}{m_b}\sum_{j=1}^{m_b} w_{jb}\tilde{T}_{jb}\tilde{\overline{\boldsymbol{x}}}_{jb} = \frac{1}{m_b}\frac{w_b^0 w_b^1}{w_b}\left(\overline{\overline{\boldsymbol{x}}}_b^1 - \overline{\overline{\boldsymbol{x}}}_b^0\right) \xrightarrow{p} 0$$

because

$$\overline{\overline{\boldsymbol{x}}}_b^1 - \overline{\overline{\boldsymbol{x}}}_b^0 \xrightarrow{p} \frac{\overline{\boldsymbol{X}}_b^*}{\omega_b} - \frac{\overline{\boldsymbol{X}}_b^*}{\omega_b} = 0.$$

We also have

$$\begin{aligned}
&\frac{1}{m}\sum_{b=1}^{h}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\tilde{\boldsymbol{x}}_{ijb}'\tilde{\boldsymbol{x}}_{ijb} \\
&= \frac{1}{m}\sum_{b=1}^{h}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\left(\boldsymbol{x}_{ijb} - \overline{\overline{\boldsymbol{x}}}_b\right)'\left(\boldsymbol{x}_{ijb} - \overline{\overline{\boldsymbol{x}}}_b\right) \\
&= \sum_{b=1}^{h}\frac{m_b}{m}\frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\left(\boldsymbol{x}_{ijb} - \overline{\overline{\boldsymbol{x}}}_b\right)'\left(\boldsymbol{x}_{ijb} - \overline{\overline{\boldsymbol{x}}}_b\right) \\
&\xrightarrow{p} \sum_{b=1}^{h} q_b \boldsymbol{\Sigma}_{\boldsymbol{x},b}^2.
\end{aligned}$$

Then we have

$$\begin{pmatrix}
\frac{1}{m_1}\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}^2 & \cdots & 0 & 0 & \cdots & 0 & \frac{1}{m_1}\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}\tilde{\overline{\boldsymbol{x}}}_{j1} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & \frac{1}{m_h}\sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}^2 & 0 & \cdots & 0 & \frac{1}{m_h}\sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}\tilde{\overline{\boldsymbol{x}}}_{jh} \\
0 & \cdots & 0 & \frac{1}{m_1}w_1 & \cdots & 0 & \mathbf{0}_v \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \mathbf{0}_v \\
0 & \cdots & 0 & 0 & \cdots & \frac{1}{m_h}w_h & \mathbf{0}_v \\
\frac{1}{m}\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}\tilde{\overline{\boldsymbol{x}}}_{j1}' & \cdots & \frac{1}{m}\sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}\tilde{\overline{\boldsymbol{x}}}_{jh}' & \mathbf{0}_v' & \cdots & \mathbf{0}_v' & \frac{1}{m}\sum_{b=1}^{h}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\tilde{\boldsymbol{x}}_{ijb}'\tilde{\boldsymbol{x}}_{ijb}
\end{pmatrix}$$

$$\xrightarrow{p} \begin{pmatrix}
p_1(1-p_1)\omega_1 & \cdots & 0 & 0 & \cdots & 0 & \mathbf{0}_v \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & p_h(1-p_h)\omega_h & 0 & \cdots & 0 & \mathbf{0}_v \\
0 & \cdots & 0 & \omega_1 & \cdots & 0 & \mathbf{0}_v \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & \omega_h & \mathbf{0}_v \\
\mathbf{0}_v' & \cdots & \mathbf{0}_v' & \mathbf{0}_v' & \cdots & \mathbf{0}_v' & \sum_{b=1}^{h} q_b \boldsymbol{\Sigma}_{\boldsymbol{x},b}^2
\end{pmatrix}.$$

Hence, by continuity of the inverse and Slutsky's theorem,

$$\begin{pmatrix} \frac{1}{m_1}\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}^2 & \cdots & 0 & 0 & \cdots & 0 & \frac{1}{m_1}\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}\tilde{\bar{\boldsymbol{x}}}_{j1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \frac{1}{m_h}\sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}^2 & 0 & \cdots & 0 & \frac{1}{m_h}\sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}\tilde{\bar{\boldsymbol{x}}}_{jh} \\ 0 & \cdots & 0 & \frac{1}{m_1}w_1 & \cdots & 0 & \mathbf{0}_v \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \mathbf{0}_v \\ 0 & \cdots & 0 & 0 & \cdots & \frac{1}{m_h}w_h & \mathbf{0}_v \\ \frac{1}{m}\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}\tilde{\bar{\boldsymbol{x}}}'_{j1} & \cdots & \frac{1}{m}\sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}\tilde{\bar{\boldsymbol{x}}}'_{jh} & \mathbf{0}'_v & \cdots & \mathbf{0}'_v & \frac{1}{m}\sum_{b=1}^{h}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\tilde{\boldsymbol{x}}'_{ijb}\tilde{\boldsymbol{x}}_{ijb} \end{pmatrix}^{-1}$$

$$\xrightarrow{p} \begin{pmatrix} \frac{1}{p_1(1-p_1)\omega_1} & \cdots & 0 & 0 & \cdots & 0 & \mathbf{0}_v \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \frac{1}{p_h(1-p_h)\omega_h} & 0 & \cdots & 0 & \mathbf{0}_v \\ 0 & \cdots & 0 & \frac{1}{\omega_1} & \cdots & 0 & \mathbf{0}_v \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \frac{1}{\omega_h} & \mathbf{0}_v \\ \mathbf{0}'_v & \cdots & \mathbf{0}'_v & \mathbf{0}'_v & \cdots & \mathbf{0}'_v & \left(\sum_{b=1}^{h} q_b\boldsymbol{\Sigma}^2_{\boldsymbol{x},b}\right)^{-1} \end{pmatrix}.$$

We also have

$$\begin{aligned}\frac{1}{m_b}\sum_{j=1}^{m_b} w_{jb}\tilde{T}_{jb}\bar{y}_{jb} = \frac{1}{m_b}\frac{w_b^1 w_b^0}{w_b}\left(\bar{\bar{y}}_b(1)-\bar{\bar{y}}_b(0)\right) &\xrightarrow{p} p_b(1-p_b)\omega_b\left(\frac{\mu_b^*(1)}{\omega_b}-\frac{\mu_b^*(0)}{\omega_b}\right) \\ &= p_b(1-p_b)\left(\mu_b^*(1)-\mu_b^*(0)\right)\end{aligned}$$

and

$$\frac{1}{m_b}\sum_{j=1}^{m_b} w_{jb}\bar{y}_{jb} = p_b\frac{1}{m_b^1}\sum_{j=1}^{m_b} w_{jb}T_{jb}\bar{Y}_{jb}(1)+(1-p_b)\frac{1}{m_b^0}\sum_{j=1}^{m_b} w_{jb}(1-T_{jb})\bar{Y}_{jb}(0) \xrightarrow{p} p_b\mu_b^*(1)+(1-p_b)\mu_b^*(0).$$

Additionally,

$$\begin{aligned}&\frac{1}{m}\sum_{b=1}^{h}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\tilde{\boldsymbol{x}}'_{ijb}y_{ijb} \\ &= \frac{1}{m}\sum_{b=1}^{h}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} T_{jb}w_{ijb}(\boldsymbol{x}_{ijb}-\bar{\bar{\boldsymbol{x}}}_b)Y_{ijb}(1) + \frac{1}{m}\sum_{b=1}^{h}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}}(1-T_{jb})w_{ijb}(\boldsymbol{x}_{ijb}-\bar{\bar{\boldsymbol{x}}}_b)Y_{ijb}(0) \\ &= \sum_{b=1}^{h}\frac{m_b^1}{m}\left[\frac{1}{m_b^1}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} T_{jb}w_{ijb}\boldsymbol{x}_{ijb}Y_{ijb}(1) - \bar{\bar{\boldsymbol{x}}}_b\left(\frac{1}{m_b^1}\sum_{j=1}^{m_b} T_{jb}w_{jb}\bar{Y}_{jb}(1)\right)\right] \\ &+ \sum_{b=1}^{h}\frac{m_b^0}{m}\left[\frac{1}{m_b^0}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}}(1-T_{jb})w_{ijb}\boldsymbol{x}_{ijb}Y_{ijb}(0) - \bar{\bar{\boldsymbol{x}}}_b\left(\frac{1}{m_b^0}\sum_{j=1}^{m_b}(1-T_{jb})w_{jb}\bar{Y}_{jb}(0)\right)\right] \\ &\xrightarrow{p} \sum_{b=1}^{h} p_b q_b\boldsymbol{\Sigma}^2_{\boldsymbol{x},Y(1),b} + \sum_{b=1}^{h}(1-p_b)q_b\boldsymbol{\Sigma}^2_{\boldsymbol{x},Y(0),b}.\end{aligned}$$

This last line comes from the following two intermediate steps:

(a) Because we have limiting values on the variances of our potential outcomes, we have a law of large numbers type result (see Theorem B of Scott and Wu (1981))

$$\frac{1}{m_b^1}\sum_{j:T_{jb}=t}\sum_{i=1}^{n_{jb}} w_{ijb}\boldsymbol{x}_{ijb}Y_{ijb}(t) - \bar{\bar{\boldsymbol{x}}}_b\left(\frac{1}{m_b^1}\sum_{j:T_{jb}=t} w_{jb}\bar{Y}_{jb}(t)\right) - \boldsymbol{S}^2_{\boldsymbol{x},Y,b}(t) \xrightarrow{p} 0.$$

(b) This implies (see Lemma A.2.1) that

$$\frac{1}{m_b^1}\sum_{j:T_{jb}=t}\sum_{i=1}^{n_{jb}} w_{ijb}\boldsymbol{x}_{ijb}Y_{ijb}(t) - \bar{\bar{\boldsymbol{x}}}_b\left(\frac{1}{m_b^1}\sum_{j:T_{jb}=t} w_{jb}\bar{Y}_{jb}(t)\right) \xrightarrow{p} \boldsymbol{\Sigma}^2_{\boldsymbol{x},Y(t),b}.$$

Then we have

$$\begin{pmatrix} \frac{1}{m_1}\sum_{j=1}^{m_1} w_{j1}\tilde{T}_{j1}\bar{y}_{j1} \\ \vdots \\ \frac{1}{m_h}\sum_{j=1}^{m_h} w_{jh}\tilde{T}_{jh}\bar{y}_{jh} \\ \frac{1}{m_1}\sum_{j=1}^{m_1} w_{j1}\bar{y}_{j1} \\ \vdots \\ \frac{1}{m_h}\sum_{j=1}^{m_h} w_{jh}\bar{y}_{jh} \\ \frac{1}{m}\sum_{b=1}^{h}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\tilde{\boldsymbol{x}}'_{ijb}y_{ijb} \end{pmatrix}$$

$$\xrightarrow{p} \begin{pmatrix} p_1(1-p_1)\left(\mu_1^*(1)-\mu_1^*(0)\right) \\ \vdots \\ p_h(1-p_h)\left(\mu_h^*(1)-\mu_h^*(0)\right) \\ p_1\mu_1^*(1)+(1-p_1)\mu_1^*(0) \\ \vdots \\ p_h\mu_h^*(1)+(1-p_h)\mu_h^*(0) \\ \sum_{b=1}^{h} p_b q_b \boldsymbol{\Sigma}^2_{\boldsymbol{x},Y(1),b} + \sum_{b=1}^{h}(1-p_b)q_b\boldsymbol{\Sigma}^2_{\boldsymbol{x},Y(0),b} \end{pmatrix}.$$

Putting this all together, we have the following result:

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_0 \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \frac{1}{p_1(1-p_1)\omega_1} & \cdots & 0 & 0 & \cdots & 0 & \mathbf{0}_v \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \frac{1}{p_h(1-p_h)\omega_h} & 0 & \cdots & 0 & \mathbf{0}_v \\ 0 & \cdots & 0 & \frac{1}{\omega_1} & \cdots & 0 & \mathbf{0}_v \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \frac{1}{\omega_h} & \mathbf{0}_v \\ \mathbf{0}'_v & \cdots & \mathbf{0}'_v & \mathbf{0}'_v & \cdots & \mathbf{0}'_v & \left(\sum_{b=1}^h q_b \boldsymbol{\Sigma}^2_{\boldsymbol{x},b}\right)^{-1} \end{pmatrix}$$

$$\times \begin{pmatrix} p_1(1-p_1)\left(\mu_1^*(1)-\mu_1^*(0)\right) \\ \vdots \\ p_h(1-p_h)\left(\mu_h^*(1)-\mu_h^*(0)\right) \\ p_1\mu_1^*(1)+(1-p_1)\mu_1^*(0) \\ \vdots \\ p_h\mu_h^*(1)+(1-p_h)\mu_h^*(0) \\ \sum_{b=1}^h p_b q_b \boldsymbol{\Sigma}^2_{\boldsymbol{x},Y(1),b} + \sum_{b=1}^h (1-p_b) q_b \boldsymbol{\Sigma}^2_{\boldsymbol{x},Y(0),b} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{\omega_1}\left(\mu_1^*(1)-\mu_1^*(0)\right) \\ \vdots \\ \frac{1}{\omega_h}\left(\mu_h^*(1)-\mu_h^*(0)\right) \\ \frac{1}{\omega_1}\left(p_1\mu_1^*(1)+(1-p_1)\mu_1^*(0)\right) \\ \vdots \\ \frac{1}{\omega_h}\left(p_2\mu_h^*(1)+(1-p_h)\mu_h^*(0)\right) \\ \boldsymbol{\Gamma} \end{pmatrix}$$

$$= \begin{pmatrix} \beta_{11}^* \\ \vdots \\ \beta_{1h}^* \\ \frac{1}{\omega_1}\left(p_1\mu_1^*(1)+(1-p_1)\mu_1^*(0)\right) \\ \vdots \\ \frac{1}{\omega_h}\left(p_h\mu_h^*(1)+(1-p_h)\mu_h^*(0)\right) \\ \boldsymbol{\Gamma} \end{pmatrix},$$

where

$$\boldsymbol{\Gamma} = \left(\sum_{b=1}^h q_b \boldsymbol{\Sigma}^2_{\boldsymbol{x},b}\right)^{-1} \left(\sum_{b=1}^h p_b q_b \boldsymbol{\Sigma}^2_{\boldsymbol{x},Y(1),b} + \sum_{b=1}^h (1-p_b) q_b \boldsymbol{\Sigma}^2_{\boldsymbol{x},Y(0),b}\right).$$

Hence, we have that $\hat{\beta}_{1,b} \xrightarrow{p} \beta_{1b}^*$.

A.3.3 Asymptotic normality with known $\boldsymbol{\gamma}$

We ultimately want to find asymptotic normality results for the estimator for a single block b with covariate adjustment across blocks, $\hat{\beta}_{1,b} = \bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) - \left(\overline{\overline{\boldsymbol{x}}}_b^1 - \overline{\overline{\boldsymbol{x}}}_b^0\right)\hat{\boldsymbol{\gamma}}$. But first, let us show, similar to Li and Ding (2017), that asymptotic normality holds for $\tilde{\beta}_{1b} = \bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) - \left(\overline{\overline{\boldsymbol{x}}}_b^1 - \overline{\overline{\boldsymbol{x}}}_b^0\right)\boldsymbol{\gamma}$, where $\boldsymbol{\gamma}$ is the finite population regression estimator we would have obtained if we had run the regression on the full schedule of potential outcomes, and is thus constant for each treatment arm. Stated differently, we first find the result when we know $\boldsymbol{\gamma}$ and do not have to estimate it. In particular,

$$\boldsymbol{\gamma} = \left(\sum_{b=1}^{h} q_b \boldsymbol{S}^2_{\boldsymbol{x},b}\right)^{-1}\left(\sum_{b=1}^{h} p_b q_b \boldsymbol{S}^2_{\boldsymbol{x},Y,b}(1) + \sum_{b=1}^{h}(1-p_b) q_b \boldsymbol{S}^2_{\boldsymbol{x},Y,b}(0)\right).$$

In this setting, we have that $C_{jb}(t) = (\bar{Y}_{jb}(t) - \overline{\boldsymbol{x}}_{jb}\boldsymbol{\gamma})$ and so $D_{jb}(t) = w_{jb}(\bar{Y}_{jb}(t) - \overline{\boldsymbol{x}}_{jb}\boldsymbol{\gamma} - \bar{\bar{Y}}_b(t) + \overline{\overline{\boldsymbol{x}}}_b\boldsymbol{\gamma})/\overline{w}_b$. Hence $\bar{\bar{c}}_b(t) = \bar{\bar{y}}_b(t) - \overline{\overline{\boldsymbol{x}}}_b^1\boldsymbol{\gamma}$ and $\overline{\overline{C}}_b(t) = \bar{\bar{Y}}_b(t) - \overline{\overline{\boldsymbol{x}}}_b\boldsymbol{\gamma}$.

Corollary A.3.1. *Under the conditions of Lemma A.1.2 for block b,*

$$\begin{aligned}\frac{\left(\bar{\bar{c}}_b(1) - \bar{\bar{c}}_b(0)\right) - \left(\overline{\overline{C}}_b(1) - \overline{\overline{C}}_b(0)\right)}{\sqrt{Var(\hat{D}_b)}} &= \frac{\tilde{\beta}_{1b} - \left(\bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0)\right)}{\sqrt{Var(\hat{D}_b)}}\\ &\xrightarrow{d} N(0,1).\end{aligned}$$

Proof. The result is a direct consequence of Lemma A.1.2. The equality holds by noting that

$$\frac{\sum_{j=1}^{m_b} w_{jb}\left(\bar{Y}_{jb}(1) - \overline{\boldsymbol{x}}_{jb}\boldsymbol{\gamma}\right)}{\sum_{j=1}^{m_b} w_{jb}} - \frac{\sum_{j=1}^{m_b} w_{jb}\left(\bar{Y}_{jb}(0) - \overline{\boldsymbol{x}}_{jb}\boldsymbol{\gamma}\right)}{\sum_{j=1}^{m_b} w_{jb}} = \bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0).$$

□

The variance in the denominator of our asymptotic expression simplifies as follows:

$$
\begin{aligned}
\mathrm{Var}(\hat{D}_b) =& \frac{S^2_{D,b}(1)}{m^1} + \frac{S^2_{D,b}(0)}{m^0} - \frac{S^2(D_b)}{m} \\
=& \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} D^2_{jb}(1)}{m^1_b} + \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} D^2_{jb}(0)}{m^0_b} - \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b}(D_{jb}(1)-D_{jb}(0))^2}{m_b} \\
=& \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} \frac{w^2_{jb}}{\overline{w}^2_b}\left(\bar{Y}_{jb}(1) - \bar{\bar{Y}}_b(1) - \left(\overline{\boldsymbol{x}}_{jb} - \overline{\overline{\boldsymbol{x}}}_b\right)\boldsymbol{\gamma}\right)^2}{m^1_b} \\
&+ \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} \frac{w^2_{jb}}{\overline{w}^2_b}\left(\bar{Y}_{jb}(0) - \bar{\bar{Y}}_b(0) - \left(\overline{\boldsymbol{x}}_{jb} - \overline{\overline{\boldsymbol{x}}}_b\right)\boldsymbol{\gamma}\right)^2}{m^0_b} \\
&- \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} \frac{w^2_{jb}}{\overline{w}^2_b}\left(\bar{Y}_{jb}(1) - \bar{Y}_{jb}(0) - (\bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0))\right)^2}{m_b} \\
=& \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} \frac{w^2_{jb}}{\overline{w}^2_b}\left(\bar{Y}_{jb}(1) - \bar{\bar{Y}}_b(1) - \tilde{\boldsymbol{x}}_{jb}\boldsymbol{\gamma}\right)^2}{m^1_b} + \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} \frac{w^2_{jb}}{\overline{w}^2_b}\left(\bar{Y}_{jb}(0) - \bar{\bar{Y}}_b(0) - \tilde{\boldsymbol{x}}_{jb}\boldsymbol{\gamma}\right)^2}{m^0_b} \\
&- \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b} \frac{w^2_{jb}}{\overline{w}^2_b}\left(\bar{Y}_{jb}(1) - \bar{Y}_{jb}(0) - (\bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0))\right)^2}{m_b}.
\end{aligned}
$$

A.3.4 Joint asymptotic normality with known $\boldsymbol{\gamma}$

We now examine the joint convergence of $\tilde{\boldsymbol{\beta}}_1$. We see that each element of $\tilde{\boldsymbol{\beta}}_1$, $\tilde{\beta}_{1b}$, is independent. Thus, we have the characteristic function for $A_b = \frac{\tilde{\beta}_{1b} - \left(\bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0)\right)}{\sqrt{\mathrm{Var}(\hat{D}_b)}}$, which we denote $\phi_b(t)$, converges as follows:

$$\phi_b(t) \to e^{-t^2/2}$$

by the equivalency of convergence in distribution and point-wise convergence of characteristic functions and using Corollary A.3.1.

Now take any linear combination L, defined as

$$L = \sum_{b=1}^{h} s_b A_b.$$

Then the characteristic function of L is

$$\begin{aligned}
\phi_L(t) &= E\left[e^{itL}\right] \\
&= E\left[e^{it\sum_{b=1}^{h} s_b A_b}\right] \\
&= E\left[\prod_{b=1}^{h} e^{its_b A_b}\right] \\
&= \prod_{b=1}^{h} E\left[e^{its_b A_b}\right] \\
&= \prod_{b=1}^{h} \phi_b(s_b t) \\
&\to \prod_{b=1}^{h} e^{-s_b^2 t^2/2} \\
&= e^{-(\sum_{b=1}^{h} s_b^2)t^2/2}.
\end{aligned}$$

Hence, $L \xrightarrow{d} N(0, \sum_{b=1}^{h} s_b^2)$. Therefore, by the Cramer-Wold device,

$$\begin{pmatrix} \frac{\tilde{\beta}_{11} - \left(\bar{\bar{Y}}_1(1) - \bar{\bar{Y}}_1(0)\right)}{\sqrt{\mathrm{Var}(\hat{D}_1)}} \\ \vdots \\ \frac{\tilde{\beta}_{1h} - \left(\bar{\bar{Y}}_h(1) - \bar{\bar{Y}}_h(0)\right)}{\sqrt{\mathrm{Var}(\hat{D}_h)}} \end{pmatrix} \xrightarrow{d} N(\mathbf{0}_h, \boldsymbol{I}_h).$$

A.3.5 Theorem 1: Asymptotic normality with estimated $\hat{\boldsymbol{\gamma}}$

We now move to our primary result, Theorem 1. To obtain this result, we first find asymptotic normality results for $\hat{\beta}_{1,b} = \bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) - \left(\overline{\overline{\boldsymbol{x}}}_b^1 - \overline{\overline{\boldsymbol{x}}}_b^0\right)\hat{\boldsymbol{\gamma}}$ for a single block b. To do this, we need to show $\hat{\beta}_{1,b}$ has the same asymptotic distribution as $\tilde{\beta}_{1,b} = \bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) - \left(\overline{\overline{\boldsymbol{x}}}_b^1 - \overline{\overline{\boldsymbol{x}}}_b^0\right)\boldsymbol{\gamma}$, as done in Li and Ding (2017) for the unweighted case. Following that paper, we aim to show that the difference is order $o_p(m_b^{-1/2})$.

First note that we can use results from Section A.2.4 to put a convergence rate on $\overline{\overline{\boldsymbol{x}}}_b^1 - \overline{\overline{\boldsymbol{x}}}_b^0$. Convergence in probability of each element of $\overline{\overline{\boldsymbol{x}}}_b^1$ and $\overline{\overline{\boldsymbol{x}}}_b^0$ implies convergence of the entire vector. Hence we can look at one entry of $\overline{\overline{\boldsymbol{x}}}_b^1$ and $\overline{\overline{\boldsymbol{x}}}_b^0$ at a time (since we can use component-wise convergence in probability). For the kth component, let $C_{jb,k}(t) = [\overline{\boldsymbol{x}}_{jb}]_k$, noting that this does not change under treatment or control. Then under the conditions of Lemma A.1.1 we have asymptotic normality results for each of the components of $\overline{\overline{\boldsymbol{x}}}_b^1$ and $\overline{\overline{\boldsymbol{x}}}_b^0$. Assuming a limiting value on the variance, this in turn means that we have, for the kth component, $\left[\overline{\overline{\boldsymbol{x}}}_b^1 - \overline{\overline{\boldsymbol{x}}}_b\right]_k = O_p(m_b^{-1/2})$ and $\left[\overline{\overline{\boldsymbol{x}}}_b^0 - \overline{\overline{\boldsymbol{x}}}_b\right]_k = O_p(m_b^{-1/2})$.

Then we have

$$\left[\overline{\overline{\boldsymbol{x}}}_b^1 - \overline{\overline{\boldsymbol{x}}}_b^0\right]_k = \left[\overline{\overline{\boldsymbol{x}}}_b^1 - \overline{\overline{\boldsymbol{x}}}_b - \left(\overline{\overline{\boldsymbol{x}}}_b^0 - \overline{\overline{\boldsymbol{x}}}_b\right)\right]_k$$
$$= O_p(m_b^{-1/2}).$$

Now note that

$$\begin{aligned}
\hat{\beta}_{1,b} &= \bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) - \left(\overline{\overline{\boldsymbol{x}}}_b^1 - \overline{\overline{\boldsymbol{x}}}_b^0\right)\hat{\boldsymbol{\gamma}} \\
&= \frac{\sum_{j=1}^{m_b} T_{jb}w_{jb}\left(\bar{y}_{jb} - \overline{\boldsymbol{x}}_{jb}\boldsymbol{\gamma}\right)}{\sum_{j=1}^{m_b} T_{jb}w_{jb}} - \frac{\sum_{j=1}^{m_b}(1-T_{jb})w_{jb}\left(\bar{y}_{jb} - \overline{\boldsymbol{x}}_{jb}\boldsymbol{\gamma}\right)}{\sum_{j=1}^{m_b}(1-T_{jb})w_{jb}} - \left(\overline{\overline{\boldsymbol{x}}}_b^1 - \overline{\overline{\boldsymbol{x}}}_b^0\right)(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}).
\end{aligned}$$

From the limiting value assumptions in Section A.3.2, $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \xrightarrow{p} 0$ and so $\left(\overline{\overline{\boldsymbol{x}}}_b^1 - \overline{\overline{\boldsymbol{x}}}_b^0\right)(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = o_p(m_b^{-1/2})$. This means that $\hat{\beta}_{1,b}$ has the same asymptotic distribution as $\tilde{\beta}_{1b}$ and so we can use Corollary A.3.1.

In Theorem 1, we assume the conditions of Corollary A.3.1, the conditions of Lemma A.1.1 applied to each of the components of $\overline{\overline{\boldsymbol{x}}}_b^1$ and $\overline{\overline{\boldsymbol{x}}}_b^0$ as well as limiting values on the variance expression of each component. Also assume limiting values on the following variance expressions:

$$\begin{aligned}
\boldsymbol{S}^2_{\boldsymbol{x},b} &= \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\left(\boldsymbol{x}_{ijb} - \overline{\overline{\boldsymbol{x}}}_b\right)'\left(\boldsymbol{x}_{ijb} - \overline{\overline{\boldsymbol{x}}}_b\right) \\
\boldsymbol{S}^2_{\boldsymbol{x},Y,b}(t) &= \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\boldsymbol{x}'_{ijb}Y_{ijb}(t) - \overline{\overline{\boldsymbol{x}}}_b'\overline{wY(t)}_b \\
\boldsymbol{S}^2_{\boldsymbol{x}Y,b}(t) &= \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} \left(w_{ijb}\boldsymbol{x}'_{ijb}Y_{ijb}(t) - \overline{wxY(t)}_b\right)^2 \text{ for } t \in \{0,1\} \text{ with} \\
\overline{wxY(t)}_b &= \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\boldsymbol{x}'_{ijb}Y_{ijb}(t).
\end{aligned}$$

Then we have the result of Theorem 1,

$$\frac{\hat{\beta}_{1,b} - \left(\bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0)\right)}{\sqrt{\mathrm{Var}(\hat{D}_b)}} \xrightarrow{d} N(0,1).$$

with

$$\begin{aligned}
\mathrm{Var}(\hat{D}_b) =& \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b}\frac{w_{jb}^2}{\overline{w}_b^2}\left(\bar{Y}_{jb}(1) - \bar{\bar{Y}}_b(1) - \tilde{\overline{\boldsymbol{x}}}_{jb}\boldsymbol{\gamma}\right)^2}{m_b^1} + \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b}\frac{w_{jb}^2}{\overline{w}_b^2}\left(\bar{Y}_{jb}(0) - \bar{\bar{Y}}_b(0) - \tilde{\overline{\boldsymbol{x}}}_{jb}\boldsymbol{\gamma}\right)^2}{m_b^0} \\
&- \frac{\frac{1}{m_b-1}\sum_{j=1}^{m_b}\frac{w_{jb}^2}{\overline{w}_b^2}\left(\bar{Y}_{jb}(1) - \bar{Y}_{jb}(0) - (\bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0))\right)^2}{m_b}.
\end{aligned}$$

A.3.6 Joint asymptotic normality with estimated $\hat{\boldsymbol{\gamma}}$

We now investigate joint asymptotic normality results for the vector $\hat{\boldsymbol{\beta}}_1$. Each element of $\hat{\boldsymbol{\beta}}_1$, $\hat{\beta}_{1,b}$, is dependent because of the shared $\hat{\gamma}$ term.

With block-covariate interactions, each $\hat{\beta}_{1,b}$ is independent and thus the joint asymptotic result is immediate.

Without these interactions, we first define $C_b = \frac{\hat{\beta}_{1,b} - \left(\bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0)\right)}{\sqrt{\text{Var}(\hat{D}_b)}}$. Based on results from Section A.3.5, we have that C_b converges to the same distribution that

$$A_b = \frac{\tilde{\beta}_{1b} - \left(\bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0)\right)}{\sqrt{\text{Var}(\hat{D}_b)}}$$

converges to. This implies that the linear combination

$$S = \sum_{b=1}^{h} s_b C_b$$

has the same asymptotic distribution as

$$L = \sum_{b=1}^{h} s_b A_b.$$

In particular,

$$S = \sum_{b=1}^{h} s_b \frac{\hat{\beta}_{1,b} - \left(\bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0)\right)}{\sqrt{\text{Var}(\hat{D}_b)}} = \sum_{b=1}^{h} s_b A_b - \sum_{b=1}^{h} s_b \left(\bar{\bar{\boldsymbol{x}}}_b^1 - \bar{\bar{\boldsymbol{x}}}_b^0\right)(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$$

and the last sum is a finite sum of terms that are $o_p(m_b^{-1/2})$.

Hence, $S \xrightarrow{d} N(0, \sum_{b=1}^{h} s_b^2)$. Therefore, by the Cramer-Wold device,

$$\begin{pmatrix} \frac{\hat{\beta}_{1,1} - \left(\bar{\bar{Y}}_1(1) - \bar{\bar{Y}}_1(0)\right)}{\sqrt{\text{Var}(\hat{D}_1)}} \\ \vdots \\ \frac{\hat{\beta}_{1,h} - \left(\bar{\bar{Y}}_h(1) - \bar{\bar{Y}}_h(0)\right)}{\sqrt{\text{Var}(\hat{D}_h)}} \end{pmatrix} \xrightarrow{d} N(\mathbf{0}_h, \boldsymbol{I}_h).$$

A.4 Restricted model

In this section, we explore the impact estimator from a weighted least squares regression with additional covariate adjustment and without interactions between block indicators and treatment indicators. We now have a single treatment effect estimator that aggregates across blocks. We first find the closed form for this estimator as

$$\hat{\beta}_1 = \frac{\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b} \left(\bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0)\right)}{\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b}} - \frac{\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b} \left(\bar{\bar{\boldsymbol{x}}}_b^1 - \bar{\bar{\boldsymbol{x}}}_b^0\right)}{\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b}} \hat{\boldsymbol{\gamma}}.$$

We then find consistency results and show an asymptotic normality result. As before, we first assume a known $\boldsymbol{\gamma}$ to focus on the pooling of individual treatment impacts across blocks and then extend to an estimated $\hat{\boldsymbol{\gamma}}$.

A.4.1 The estimator

We now examine the model with no interactions between treatment and blocks . In this case, there is a single treatment effect estimator, $\hat{\beta}_1$ for all blocks. We have $\boldsymbol{z}_{ijb} = (\tilde{T}_{jb}, S_{ij1}, \ldots, S_{ijh}, \tilde{\boldsymbol{x}}_{ijb})$. The estimated parameter vector is

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\boldsymbol{\beta}}_0 \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} = \left[\left(\sum_{b=1}^{h} \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb} \boldsymbol{z}'_{ijb} \boldsymbol{z}_{ijb} \right)^{-1} \sum_{b=1}^{h} \sum_{j=1}^{m_b} \sum_{i=1}^{n_{jb}} w_{ijb} \boldsymbol{z}'_{ijb} y_{ijb} \right].$$

Using the same techniques as in Section A.3.1 we find the following:

Result:

$$\hat{\beta}_1 = \frac{\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b} \left(\bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) \right)}{\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b}} - \frac{\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b} \left(\bar{\bar{\boldsymbol{x}}}_b^1 - \bar{\bar{\boldsymbol{x}}}_b^0 \right)}{\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b}} \hat{\boldsymbol{\gamma}}.$$

Derivations available upon request.

A.4.2 Consistency

Result:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b} \left(\bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) \right)}{\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b}} - \frac{\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b} \left(\bar{\bar{\boldsymbol{x}}}_b^1 - \bar{\bar{\boldsymbol{x}}}_b^0 \right)}{\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b}} \hat{\boldsymbol{\gamma}} \\ &\xrightarrow{p} \frac{\sum_{b=1}^{h} q_b p_b (1 - p_b) \left(\mu_b^*(1) - \mu_b^*(0) \right)}{\sum_{b=1}^{h} q_b p_b (1 - p_b) \omega_b} \end{aligned}$$

Using results and limiting values given in Section A.2.3 and Section A.3.2 and also

$m_b/m \xrightarrow{p} q_b$ $(0 < q_b < 1)$, we have

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\boldsymbol{\beta}}_0 \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} \frac{1}{m}\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b} & 0 & \cdots & 0 & \frac{1}{m}\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b} \left(\overline{\overline{\boldsymbol{x}}}_b^1 - \overline{\overline{\boldsymbol{x}}}_b^0 \right) \\ 0 & \frac{1}{m_1} w_1 & \cdots & 0 & \mathbf{0}_v \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{m_h} w_h & \mathbf{0}_v \\ \frac{1}{m}\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b} \left(\overline{\overline{\boldsymbol{x}}}_b^{1\prime} - \overline{\overline{\boldsymbol{x}}}_b^{0\prime} \right) & \mathbf{0}_v' & \cdots & \mathbf{0}_v' & \frac{1}{m}\sum_{b=1}^{h}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb} \tilde{\boldsymbol{x}}_{ijb}' \tilde{\boldsymbol{x}}_{ijb} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \frac{1}{m}\sum_{b=1}^{h} \frac{w_b^1 w_b^0}{w_b} \left(\bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) \right) \\ \frac{1}{m_1} \left(w_1^1 \bar{\bar{y}}_1(1) + w_1^0 \bar{\bar{y}}_1(0) \right) \\ \vdots \\ \frac{1}{m_h} \left(w_h^1 \bar{\bar{y}}_h(1) + w_h^0 \bar{\bar{y}}_h(0) \right) \\ \frac{1}{m}\sum_{b=1}^{h}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb} \tilde{\boldsymbol{x}}_{ijb}' \bar{y}_{ijb} \end{pmatrix}$$

$$\xrightarrow{p} \begin{pmatrix} \sum_{b=1}^{h} q_b p_b (1-p_b)\omega_b & 0 & \cdots & 0 & 0 \\ 0 & \omega_1 & \cdots & 0 & \mathbf{0}_v \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \omega_h & \mathbf{0}_v \\ 0 & \mathbf{0}_v' & \cdots & \mathbf{0}_v' & \sum_{b=1}^{h} q_b \boldsymbol{\Sigma}_{\boldsymbol{x},b}^2 \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \sum_{b=1}^{h} q_b p_b (1-p_b) \left(\mu_b^*(1) - \mu_b^*(0) \right) \\ p_1 \mu_1(1) + (1-p_1)\mu_1(0) \\ \vdots \\ p_h \mu_h(1) + (1-p_h)\mu_h(0) \\ \sum_{b=1}^{h} p_b q_b \boldsymbol{\Sigma}_{\boldsymbol{x},Y(1),b}^2 + \sum_{b=1}^{h} (1-p_b) q_b \boldsymbol{\Sigma}_{\boldsymbol{x},Y(0),b}^2 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\sum_{b=1}^{h} q_b p_b (1-p_b) \left(\mu_b^*(1) - \mu_b^*(0) \right)}{\sum_{b=1}^{h} q_b p_b (1-p_b)\omega_b} \\ \frac{1}{\omega_1} \left(p_1 \mu_1(1) + (1-p_1)\mu_1(0) \right) \\ \vdots \\ \frac{1}{\omega_h} \left(p_h \mu_h(1) + (1-p_h)\mu_h(0) \right) \\ \boldsymbol{\Gamma} \end{pmatrix}$$

where

$$\boldsymbol{\Gamma} = \left(\sum_{b=1}^{h} q_b \boldsymbol{\Sigma}_{\boldsymbol{x},b}^2 \right)^{-1} \left(\sum_{b=1}^{h} p_b q_b \boldsymbol{\Sigma}_{\boldsymbol{x},Y(1),b}^2 + \sum_{b=1}^{h} (1-p_b) q_b \boldsymbol{\Sigma}_{\boldsymbol{x},Y(0),b}^2 \right).$$

A.4.3 Asymptotic normality with known $\boldsymbol{\gamma}$

Let $\tilde{q}_b = m_b/m$. Following Section A.3.4, we want to start by finding the asymptotic distribution of

$$\tilde{\beta}_1 = \frac{\sum_{b=1}^h \frac{w_b^1 w_b^0}{w_b}\left(\bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0)\right)}{\sum_{a=1}^h \frac{w_a^1 w_a^0}{w_a}} - \frac{\sum_{b=1}^h \frac{w_b^1 w_b^0}{w_b}\left(\bar{\bar{\boldsymbol{x}}}_b^1 - \bar{\bar{\boldsymbol{x}}}_b^0\right)}{\sum_{a=1}^h \frac{w_a^1 w_a^0}{w_a}}\boldsymbol{\gamma}$$
$$= \sum_{b=1}^h \frac{\tilde{q}_b p_b \overline{w}_b^1(\overline{w}_b - p_b\overline{w}_b^1)/\overline{w}_b}{\sum_{a=1}^h \tilde{q}_a p_a \overline{w}_a^1(\overline{w}_a - p_a\overline{w}_a^1)/\overline{w}_a}\left(\bar{\bar{y}}_b(1) - \bar{\bar{y}}_b(0) - \left(\bar{\bar{\boldsymbol{x}}}_b^1 - \bar{\bar{\boldsymbol{x}}}_b^0\right)\boldsymbol{\gamma}\right).$$

Further denote

$$\beta_1 = \sum_{b=1}^h \frac{\tilde{q}_b p_b(1-p_b)\overline{w}_b}{\sum_{a=1}^h \tilde{q}_a p_a(1-p_a)\overline{w}_a}\left(\bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0)\right).$$

Let

$$U_{jb}(t) = w_{jb}\left(\bar{Y}_{jb}(t) - \bar{\boldsymbol{x}}_{jb}\boldsymbol{\gamma}\right), \quad \bar{U}_b(t) = \frac{1}{m_b}\sum_{j=1}^{m_b} U_{jb}(t), \quad \text{and} \quad \bar{u}_b(t) = \frac{1}{m_b^t}\sum_{j:T_{jb}=t} U_{jb}(t).$$

Then we can write

$$\tilde{\beta}_1 = \sum_{b=1}^h \frac{\tilde{q}_b p_b \overline{w}_b^1(\overline{w}_b - p_b\overline{w}_b^1)/\overline{w}_b}{\sum_{a=1}^h \tilde{q}_a p_a \overline{w}_a^1(\overline{w}_a - p_a\overline{w}_a^1)/\overline{w}_a}\left(\frac{\bar{u}_b(1)}{\overline{w}^1} - \frac{\bar{u}_b(0)}{\overline{w}^0}\right).$$

It is useful to rewrite this in terms of the fewest possible random variables, so we rewrite $\overline{w}_b^1 = (\overline{w}_b - p_b\overline{w}_b^1)/(1-p_b)$ everywhere as follows:

$$\tilde{\beta}_1 = \sum_{b=1}^h \frac{\tilde{q}_b p_b \overline{w}_b^1(\overline{w}_b - p_b\overline{w}_b^1)/\overline{w}_b}{\sum_{a=1}^h \tilde{q}_a p_a \overline{w}_a^1(\overline{w}_a - p_a\overline{w}_a^1)/\overline{w}_a}\left(\frac{\bar{u}_b(1)}{\overline{w}^1} - \frac{(1-p_b)\bar{u}_b(0)}{\overline{w}_b - p_b\overline{w}_b^1}\right).$$

Let, for $t \in \{0,1\}$ and $z \in \{w, U\}$,

$$a_{z,b}(t) = \max_{1\le j\le m_b}\left(z_{j,b}(t) - \bar{z}_b(t)\right)^2$$

and

$$v_{z,b}(t) = \frac{1}{m_b - 1}\sum_{j=1}^{m_b}\left(z_{j,b}(t) - \bar{z}_b(t)\right)^2,$$

noting that $w_{jb}(t) = w_{jb}$. Let $\boldsymbol{t} = (\bar{u}_1(1), \bar{u}_1(0), \overline{w}_1^1, \cdots, \bar{u}_h(1), \bar{u}_h(0), \overline{w}_h^1)$ and $\boldsymbol{T} = (\bar{U}_1(1), \bar{U}_1(0), \overline{w}_1, \cdots, \bar{U}_h(1), \bar{U}_h(0), \overline{w}_h)$.

Theorem A.1. *Let us assume the following conditions:*

(a) As $m \to \infty$,

$$max_{1\le b\le h} max_{z\in\{w,U\}} max_{t\in\{0,1\}} \frac{a_{z,b}(t)}{p_b(1-p_b)m_b v_{z,b}(t)} \to 0.$$

(b) The correlation matrix of $\boldsymbol{t}$ has a limiting value $\boldsymbol{\Sigma}$.

(c) We have limiting values on the following variance expressions: $m\,Var(\overline{w}_b^1)$ and $m\,Var(\bar{u}_b(z))$ for all $b \in \{1, \dots, h\}$ and $z \in \{0, 1\}$.

(d) $\bar{U}_b(1) \neq 0$ or $\bar{U}_b(0) \neq 0$ for some b.

Then we have

$$\frac{\tilde{\beta}_1 - \beta_1}{\sqrt{Var\left(\tilde{\beta}_1\right)}} \xrightarrow{d} N(0,1)$$

where

$$\begin{aligned}
Var\left(\tilde{\beta}_1\right) = \sum_{b=1}^{h} \frac{1}{m_b(m_b-1)} \sum_{j=1}^{m_b} \Bigg(& \frac{\tilde{q}_b p_b(1-p_b)(1-2p_b)}{\sqrt{p_b(1-p_b)}\left(\sum_{a=1}^{h} \tilde{q}_a p_a (1-p_a)\overline{w}_a\right)} \left(\beta_{1,b} - \beta_1\right)\left(w_{jb} - \overline{w}_b\right) \\
& + \sqrt{\frac{1-p_b}{p_b}} \frac{\tilde{q}_b p_b(1-p_b)\overline{w}_b}{\sum_{a=1}^{h} \tilde{q}_a p_a(1-p_a)\overline{w}_a} \left(\frac{U_{jb}(1)}{\overline{w}_b} - \frac{w_{jb}}{\overline{w}_b}\frac{\bar{U}_b(1)}{\overline{w}_b}\right) \\
& - \sqrt{\frac{p_b}{1-p_b}} \frac{\tilde{q}_b p_b(1-p_b)\overline{w}_b}{\sum_{a=1}^{h} \tilde{q}_a p_a(1-p_a)\overline{w}_a} \left(\frac{U_{jb}(0)}{\overline{w}_b} - \frac{w_{jb}}{\overline{w}_b}\frac{\bar{U}_b(0)}{\overline{w}_b}\right) \Bigg)^2 \\
= \sum_{b=1}^{h} \frac{1}{m_b(m_b-1)} \sum_{j=1}^{m_b} \Bigg(& \frac{\tilde{q}_b p_b(1-p_b)(1-2p_b)}{\sqrt{p_b(1-p_b)}\left(\sum_{a=1}^{h} \tilde{q}_a p_a (1-p_a)\overline{w}_a\right)} \left(\beta_{1,b} - \beta_1\right)\left(w_{jb} - \overline{w}_b\right) \\
& + \sqrt{\frac{1-p_b}{p_b}} \frac{\tilde{q}_b p_b(1-p_b)\overline{w}_b}{\sum_{a=1}^{h} \tilde{q}_a p_a(1-p_a)\overline{w}_a} \left(\frac{w_{jb}\left(\bar{Y}_{jb}(1) - \overline{\boldsymbol{x}}_{jb}\boldsymbol{\gamma}\right)}{\overline{w}_b} - \frac{w_{jb}\left(\bar{\bar{Y}}_b(1) - \overline{\overline{\boldsymbol{x}}}_b\boldsymbol{\gamma}\right)}{\overline{w}_b}\right) \\
& - \sqrt{\frac{p_b}{1-p_b}} \frac{\tilde{q}_b p_b(1-p_b)\overline{w}_b}{\sum_{a=1}^{h} \tilde{q}_a p_a(1-p_a)\overline{w}_a} \left(\frac{w_{jb}\left(\bar{Y}_{jb}(0) - \overline{\boldsymbol{x}}_{jb}\boldsymbol{\gamma}\right)}{\overline{w}_b} - \frac{w_{jb}\left(\bar{\bar{Y}}_b(0) - \overline{\overline{\boldsymbol{x}}}_b\boldsymbol{\gamma}\right)}{\overline{w}_b}\right) \Bigg)^2.
\end{aligned}$$

Proof. By Theorem 4 of Li and Ding (2017), if as $m \to \infty$

$$\max_{1 \leq b \leq h} \max_{z \in \{w, U\}} \max_{t \in \{0,1\}} \frac{a_{z,b}(t)}{p_b(1-p_b)m_b v_{z,b}(t)} \to 0$$

and the correlation matrix of $\boldsymbol{t}$ has a limiting value $\boldsymbol{\Sigma}$, then

$$\left(\frac{\bar{u}_1(1) - \bar{U}_1(1)}{\sqrt{\mathrm{Var}(\bar{u}_1(1))}}, \frac{\bar{u}_1(0) - \bar{U}_1(0)}{\sqrt{\mathrm{Var}(\bar{u}_1(0))}}, \frac{\overline{w}_1^1 - \overline{w}_1}{\sqrt{\mathrm{Var}(\overline{w}_1^1)}}, \dots, \frac{\bar{u}_h(0) - \bar{U}_h(0)}{\sqrt{\mathrm{Var}(\bar{u}_h(0))}}, \frac{\overline{w}_h^1 - \overline{w}_h}{\sqrt{\mathrm{Var}(\overline{w}_h^1)}}\right) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

To use the delta method given in Pashley (2019), we also require that $\boldsymbol{t} - \boldsymbol{T} \xrightarrow{p} 0$ (i.e., $\overline{w}_b^1 - \overline{w}_b \xrightarrow{p} 0$ and $\bar{u}_b(z) - \bar{U}_b(z) \xrightarrow{p} 0$ for all $b \in \{1, \dots, h\}$ and $z \in \{0, 1\}$). This is satisfied by our assumption on limiting values of the variances (this can be seen directly from our prior results or Markov's inequality).

We can determine the variance and covariances by noting that blocks are independent of each other, but random variables within blocks are dependent. We have

$$\mathrm{Var}\left(\bar{u}_b(t)\right) = \left(\frac{1}{m_b^t} - \frac{1}{m_b}\right) v_{z,U}(t),$$

$$\mathrm{Var}\left(\overline{w}_b^t\right) = \left(\frac{1}{m_b^t} - \frac{1}{m_b}\right) v_{z,w}(t),$$

$$\mathrm{Cov}\left(\bar{u}_b(1), \bar{u}_b(0)\right) = -\frac{1}{m_b}\frac{1}{m_b - 1}\sum_{j=1}^{m_b}\left(U_{jb}(1) - \bar{U}_b(1)\right)\left(U_{jb}(0) - \bar{U}_b(0)\right),$$

$$\mathrm{Cov}\left(\bar{u}_b(1), \overline{w}_b^1\right) = \left(\frac{1}{m_b^1} - \frac{1}{m_b}\right)\frac{1}{m_b - 1}\sum_{j=1}^{m_b}\left(U_{jb}(1) - \bar{U}_b(1)\right)\left(w_{jb} - \overline{w}_b\right),$$

and

$$\mathrm{Cov}\left(\bar{u}_b(0), \overline{w}_b^1\right) = -\frac{1}{m_b}\frac{1}{m_b - 1}\sum_{j=1}^{m_b}\left(U_{jb}(0) - \bar{U}_b(0)\right)\left(w_{jb} - \overline{w}_b\right).$$

Now $g(\cdot) : \mathbb{R}^{3h} \to \mathbb{R}$ takes our vector $\boldsymbol{t}$ and returns the estimator $\tilde{\beta}_1$. As our weights are all positive and non-zero, this function is continuous and differential on the domain of $\boldsymbol{T}$. We have

$$\nabla g(\boldsymbol{T}) = \begin{pmatrix} \frac{\partial g(\boldsymbol{T})}{\partial \bar{u}_1(1)} & \frac{\partial g(\boldsymbol{T})}{\partial \bar{u}_1(0)} & \frac{\partial g(\boldsymbol{T})}{\partial \overline{w}_1^1} & \cdots & \frac{\partial g(\boldsymbol{T})}{\partial \bar{u}_h(1)} & \frac{\partial g(\boldsymbol{T})}{\partial \bar{u}_h(0)} & \frac{\partial g(\boldsymbol{T})}{\partial \overline{w}_h^1} \end{pmatrix}^T.$$

The partial derivatives are

$$\begin{aligned}
\left.\frac{\partial \hat{\beta}}{\partial \bar{u}_1(1)}\right|_{\boldsymbol{T}} =& \frac{\tilde{q}_b p_b(1-p_b)\overline{w}_b}{\sum_{a=1}^h \tilde{q}_a p_a(1-p_a)\overline{w}_a}\frac{1}{\overline{w}_b}\\
\left.\frac{\partial \hat{\beta}}{\partial \bar{u}_1(0)}\right|_{\boldsymbol{T}} =& -\frac{\tilde{q}_b p_b(1-p_b)\overline{w}_b}{\sum_{a=1}^h \tilde{q}_a p_a(1-p_a)\overline{w}_a}\frac{1}{\overline{w}_b}\\
\left.\frac{\partial \hat{\beta}}{\partial \overline{w}_b}\right|_{\boldsymbol{T}} =& -\tilde{q}_b p_b(1-2p_b)\sum_{c=1}^h \frac{\tilde{q}_c p_c(1-p_c)\overline{w}_c}{\left(\sum_{a=1}^h \tilde{q}_a p_a(1-p_a)\overline{w}_a\right)^2}\left(\frac{\bar{U}_c(1)}{\overline{w}_c} - \frac{\bar{U}_c(0)}{\overline{w}_c}\right)\\
&- \frac{\tilde{q}_b p_b(1-p_b)}{\sum_{a=1}^h \tilde{q}_a p_a(1-p_a)\overline{w}_a}\left(\frac{p_b}{1-p_b}\frac{\bar{U}_b(1)}{\overline{w}_b} + \frac{\bar{U}_b(0)}{\overline{w}_b}\right)\\
=& -\frac{\tilde{q}_b p_b(1-2p_b)}{\sum_{a=1}^h \tilde{q}_a p_a(1-p_a)\overline{w}_a}\beta_1 - \frac{\tilde{q}_b p_b(1-p_b)\overline{w}_b}{\sum_{a=1}^h \tilde{q}_a p_a(1-p_a)\overline{w}_a}\left(\frac{p_b}{1-p_b}\frac{\bar{U}_b(1)}{(\overline{w}_b)^2} + \frac{\bar{U}_b(0)}{(\overline{w}_b)^2}\right).
\end{aligned}$$

We see that all of these derivatives are continuous and, assuming that we do not have $\bar{U}_b(0) = \bar{U}_b(0) = 0$ for all b, we can therefore use the delta method result of Theorem 2 of Pashley (2019), which gives

$$\frac{g(\boldsymbol{t}) - g(\boldsymbol{T})}{\sqrt{\left(\nabla g(\boldsymbol{T})\right)^T \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}\nabla g(\boldsymbol{T})}} \xrightarrow{d} \mathrm{N}(0,1)$$

where

$$\boldsymbol{V} = \begin{pmatrix} \sqrt{\mathrm{Var}(\bar{u}_1(1))} & 0 & \cdots & 0 \\ 0 & \sqrt{\mathrm{Var}(\bar{u}_1(0))} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\mathrm{Var}(\overline{w}_h^1)} \end{pmatrix}.$$

The denominator corresponds to the variance of $\tilde{\beta}_1$. Let $\beta_{1,b} = (\bar{U}_b(1) - \bar{U}_b(0))/\overline{w}_b = \bar{\bar{Y}}_b(1) - \bar{\bar{Y}}_b(0)$. We can now do the multiplication to get the variance, noting that symmetry can greatly simplify the calculations. Replacing the limiting values of $\boldsymbol{\Sigma}$ with the sample values, we get

$$\begin{aligned} \mathrm{Var}(\tilde{\beta}_1) = \sum_{b=1}^{h} \frac{1}{m_b(m_b-1)} \sum_{j=1}^{m_b} \Bigg(& - \frac{\sqrt{1-p_b}\tilde{q}_b p_b (1-2p_b)}{\sqrt{p_b}\left(\sum_{a=1}^{h} \tilde{q}_a p_a (1-p_a)\overline{w}_a\right)} \beta_1 (w_{jb} - \overline{w}_b) \\ & - \sqrt{\frac{1-p_b}{p_b}} \frac{\tilde{q}_b p_b (1-p_b)\overline{w}_b}{\sum_{a=1}^{h} \tilde{q}_a p_a (1-p_a)\overline{w}_a} \left(\frac{p_b}{1-p_b} \frac{\bar{U}_b(1)}{(\overline{w}_b)^2} + \frac{\bar{U}_b(0))}{(\overline{w}_b)^2} \right) (w_{jb} - \overline{w}_b) \\ & + \sqrt{\frac{1-p_b}{p_b}} \frac{\tilde{q}_b p_b (1-p_b)\overline{w}_b}{\sum_{a=1}^{h} \tilde{q}_a p_a (1-p_a)\overline{w}_a} \left(\frac{U_{jb}(1)}{\overline{w}_b} - \frac{\bar{U}_b(1)}{\overline{w}_b} \right) \\ & + \sqrt{\frac{p_b}{1-p_b}} \frac{\tilde{q}_b p_b (1-p_b)\overline{w}_b}{\sum_{a=1}^{h} \tilde{q}_a p_a (1-p_a)\overline{w}_a} \left(\frac{U_{jb}(0)}{\overline{w}_b} - \frac{\bar{U}_b(0)}{\overline{w}_b} \right) \Bigg)^2. \end{aligned}$$

We can rewrite this more digestibly in terms of block estimators as

$$\begin{aligned} \mathrm{Var}(\tilde{\beta}_1) = \sum_{b=1}^{h} \frac{1}{m_b(m_b-1)} \sum_{j=1}^{m_b} \Bigg(& \frac{\tilde{q}_b p_b (1-p_b)(1-2p_b)}{\sqrt{p_b(1-p_b)}\left(\sum_{a=1}^{h} \tilde{q}_a p_a (1-p_a)\overline{w}_a\right)} \left(\beta_{1,b} - \beta_1\right)(w_{jb} - \overline{w}_b) \\ & + \sqrt{\frac{1-p_b}{p_b}} \frac{\tilde{q}_b p_b (1-p_b)\overline{w}_b}{\sum_{a=1}^{h} \tilde{q}_a p_a (1-p_a)\overline{w}_a} \left(\frac{w_{jb}\left(\bar{Y}_{jb}(1) - \overline{\boldsymbol{x}}_{jb}\boldsymbol{\gamma}\right)}{\overline{w}_b} - \frac{w_{jb}\left(\bar{\bar{Y}}_b(1) - \overline{\overline{\boldsymbol{x}}}_b\boldsymbol{\gamma}\right)}{\overline{w}_b} \right) \\ & + \sqrt{\frac{p_b}{1-p_b}} \frac{\tilde{q}_b p_b (1-p_b)\overline{w}_b}{\sum_{a=1}^{h} \tilde{q}_a p_a (1-p_a)\overline{w}_a} \left(\frac{w_{jb}\left(\bar{Y}_{jb}(0) - \overline{\boldsymbol{x}}_{jb}\boldsymbol{\gamma}\right)}{\overline{w}_b} - \frac{w_{jb}\left(\bar{\bar{Y}}_b(0) - \overline{\overline{\boldsymbol{x}}}_b\boldsymbol{\gamma}\right)}{\overline{w}_b} \right) \Bigg)^2. \end{aligned}$$

□

A.4.4 Theorem 2: Asymptotic normality with estimated $\hat{\boldsymbol{\gamma}}$

Finally, to prove Theorem 2 we can follow the same proof as in Section A.3.5.

Theorem 2 assume the conditions of Theorem A.1, the conditions of Lemma A.1.1 applied to each of the components of $\overline{\overline{\boldsymbol{x}}}_b^1$ and $\overline{\overline{\boldsymbol{x}}}_b^0$, and a limiting value on the asymptotic variance for those components. It also assumes limiting values on the following variance expressions:

$$\begin{aligned}
\boldsymbol{S}^2_{\boldsymbol{x},b} &= \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\left(\boldsymbol{x}_{ijb}-\overline{\overline{\boldsymbol{x}}}_b\right)'\left(\boldsymbol{x}_{ijb}-\overline{\overline{\boldsymbol{x}}}_b\right)\\
\boldsymbol{S}^2_{\boldsymbol{x},Y,b}(t) &= \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\boldsymbol{x}'_{ijb}Y_{ijb}(t)-\overline{\overline{\boldsymbol{x}}}'_b\overline{wY(t)}_b\\
\boldsymbol{S}^2_{\boldsymbol{x}Y,b}(t) &= \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}}\left(w_{ijb}\boldsymbol{x}'_{ijb}Y_{ijb}(t)-\overline{wxY(t)}_b\right)^2 \text{ for } t\in\{0,1\} \text{ with}\\
\overline{wxY(t)}_b &= \frac{1}{m_b}\sum_{j=1}^{m_b}\sum_{i=1}^{n_{jb}} w_{ijb}\boldsymbol{x}'_{ijb}Y_{ijb}(t).
\end{aligned}$$

Then, we have the result of Theorem 2,

$$\frac{\hat{\beta}_1-\beta_1}{\sqrt{\mathrm{Var}\left(\tilde{\beta}_1\right)}}\xrightarrow{d} N(0,1).$$

with

$$\begin{aligned}
\mathrm{Var}\left(\hat{\beta}_1\right) &= \sum_{b=1}^{h}\frac{1}{m_b(m_b-1)}\sum_{j=1}^{m_b}\Bigg(\frac{\tilde{q}_b p_b(1-p_b)(1-2p_b)}{\sqrt{p_b(1-p_b)}\left(\sum_{a=1}^{h}\tilde{q}_a p_a(1-p_a)\overline{w}_a\right)}\left(\beta_{1,b}-\beta_1\right)\left(w_{jb}-\overline{w}_b\right)\\
&\quad+\sqrt{\frac{1-p_b}{p_b}}\frac{\tilde{q}_b p_b(1-p_b)\overline{w}_b}{\sum_{a=1}^{h}\tilde{q}_a p_a(1-p_a)\overline{w}_a}\left(\frac{U_{jb}(1)}{\overline{w}_b}-\frac{w_{jb}}{\overline{w}_b}\frac{\bar{U}_b(1)}{\overline{w}_b}\right)\\
&\quad+\sqrt{\frac{p_b}{1-p_b}}\frac{\tilde{q}_b p_b(1-p_b)\overline{w}_b}{\sum_{a=1}^{h}\tilde{q}_a p_a(1-p_a)\overline{w}_a}\left(\frac{U_{jb}(0)}{\overline{w}_b}-\frac{w_{jb}}{\overline{w}_b}\frac{\bar{U}_b(0)}{\overline{w}_b}\right)\Bigg)^2\\
&= \sum_{b=1}^{h}\frac{1}{m_b(m_b-1)}\sum_{j=1}^{m_b}\Bigg(\frac{\tilde{q}_b p_b(1-p_b)(1-2p_b)}{\sqrt{p_b(1-p_b)}\left(\sum_{a=1}^{h}\tilde{q}_a p_a(1-p_a)\overline{w}_a\right)}\left(\beta_{1,b}-\beta_1\right)\left(w_{jb}-\overline{w}_b\right)\\
&\quad+\sqrt{\frac{1-p_b}{p_b}}\frac{\tilde{q}_b p_b(1-p_b)\overline{w}_b}{\sum_{a=1}^{h}\tilde{q}_a p_a(1-p_a)\overline{w}_a}\left(\frac{w_{jb}\left(\bar{Y}_{jb}(1)-\overline{\boldsymbol{x}}_{jb}\boldsymbol{\gamma}\right)}{\overline{w}_b}-\frac{w_{jb}\left(\bar{\bar{Y}}_b(1)-\overline{\overline{\boldsymbol{x}}}_b\boldsymbol{\gamma}\right)}{\overline{w}_b}\right)\\
&\quad+\sqrt{\frac{p_b}{1-p_b}}\frac{\tilde{q}_b p_b(1-p_b)\overline{w}_b}{\sum_{a=1}^{h}\tilde{q}_a p_a(1-p_a)\overline{w}_a}\left(\frac{w_{jb}\left(\bar{Y}_{jb}(0)-\overline{\boldsymbol{x}}_{jb}\boldsymbol{\gamma}\right)}{\overline{w}_b}-\frac{w_{jb}\left(\bar{\bar{Y}}_b(0)-\overline{\overline{\boldsymbol{x}}}_b\boldsymbol{\gamma}\right)}{\overline{w}_b}\right)\Bigg)^2.
\end{aligned}$$

B Detailed Simulation Methods and Results

The data generating process in (19) of the main text was used for the simulations. We parameterized all error variances in terms of the intraclass correlation coefficient for the control group, $ICC_0 = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$, the explained variance of the covariates in the control group, R_0^2, and the variance of the outcome in the control group, $\sigma_{y_0}^2$. In addition, we assumed that the two model covariates were iid normal with the same variance, $\sigma_{x_k}^2 = \sigma_x^2$ and the same parameter values, $\gamma_k = 1$. Thus, because $R_0^2 = 2\gamma_k^2\sigma_x^2/\sigma_{y_0}^2$, we have that $\sigma_x^2 = R_0^2\sigma_{y_0}^2/2$. Let $\sigma_{y_0^*}^2 = \sigma_{y_0}^2 - 2\sigma_x^2$ and assume that $\sigma_\theta^2 = f\sigma_u^2$ for $f \geq 0$. We then have that $\sigma_u^2 = b_{y_0^*}^2 ICC_0$, $\sigma_e^2 = \sigma_{y_0^*}^2(1 - ICC_0)$, and $\sigma_\theta^2 = f\sigma_{y_0^*}^2 ICC_0$. Further, we set $\sigma_{y_0}^2 = 1$ so that the outcomes are in effect size units based on the standard deviation of the control group, which is common practice.

To apply the formulas and match real-world scenarios, we assumed that $ICC_0 = .10$, $f = .10$, and $R_0^2 = .30$. These assumptions yield simulation values of $\sigma_x^2 = .15$, $\sigma_u^2 = .07$, $\sigma_\theta^2 = .007$, and $\sigma_e^2 = .63$.

Separate simulations were conducted using $m = 8$ to 50 clusters, split evenly between the treatment and control groups. We assumed n_j individuals per cluster, where n_j differed across clusters and was generated to be correlated with both u_j and θ_j. Specifically, we generated cluster sample sizes using $n_j = \mu + n_j^* + \delta_1 u_j + \delta_2\theta_j$, where n_j^* is a random variable with mean 0 and variance $\sigma_{n^*}^2$. We assumed correlations between n_j and u_j of $\rho_{nu} = .25$ and between n_j and θ_j of $\rho_{n\theta} = .05$. We assumed $\mu = 100$ (or 40 in sensitivity analyses) and that, the standard deviation of n_j, $\sigma_n = 10$ (or 5 in sensitivity analyses). In terms of these parameters, we have that $\delta_1 = \rho_{nu}\sigma_n/\sigma_u$, $\delta_2 = \rho_{n\theta}\sigma_n/\sigma_\theta$, and $\sigma_{n^*}^2 = \sigma_n^2 - \delta_1^2\sigma_u^2 - \delta_2^2\sigma_\theta^2$.

We conducted the simulations separately assuming that the random variables $(x_{ij1}, x_{j2}, u_j, \theta_j, e_{ij}, n_j^*)$ were drawn from three types of distributions: (1) normal, (2) chi-square (to allow for some skewness), and (3) bi-modal. We used a chi-squared distribution with 3 degrees of freedom that we centered at 0 and normalized to match the variances for each random variable. To simulate a bi-modal distribution for each random variable k, we drew from a normal distribution, $N(-\sqrt{.5\sigma_k^2}, .5\sigma_k^2)$, with probability 1/2 and from a normal distribution, $N(\sqrt{.5\sigma_k^2}, .5\sigma_k^2)$, with probability 1/2. The full simulation results for the clustered, non-blocked RCT are presented in Appendix Tables B.1 to B.4 below.

Table B.1. Full finite population simulation results for models without covariates

	Type 1 error		Standard error		
Number of clusters (half treatment, half control)	Design-based, t-test with df=m-2	CRSE, t-test with df=m-1	True	Mean design-based	Mean CRSE
Random variables with normal distributions					
8	5.15%	7.26%	0.220	0.221	0.205
10	5.02%	6.49%	0.193	0.194	0.183
12	5.10%	6.36%	0.192	0.193	0.184
16	4.93%	5.81%	0.163	0.164	0.159
20	4.92%	5.60%	0.147	0.149	0.145
50	5.07%	5.29%	0.092	0.093	0.093
Random variables with bimodal distributions					
8	5.01%	7.16%	0.225	0.225	0.209
10	5.44%	7.09%	0.213	0.213	0.201
12	5.04%	6.32%	0.184	0.185	0.177
16	5.22%	6.06%	0.167	0.168	0.162
20	5.17%	5.80%	0.150	0.151	0.147
50	5.03%	5.27%	0.095	0.096	0.095
Random variables with chi-square distributions					
8	4.37%	6.52%	0.199	0.200	0.185
10	4.68%	6.15%	0.204	0.203	0.192
12	4.63%	5.80%	0.189	0.190	0.181
16	4.74%	5.61%	0.171	0.171	0.166
20	4.98%	5.63%	0.145	0.146	0.142
50	4.97%	5.20%	0.095	0.096	0.095

Notes: Results obtained from 1,000 simulation draws from the randomization distribution for each of 100 datasets assuming the distributions indicated above. See the main text and Appendix B for simulation specifications. The design-based variance estimator uses (9) in the main text. CRSE = Standard cluster-robust standard error estimator; df = Degrees of freedom.

Table B.2. Full finite population simulation results for models with covariates

Number of clusters (half treatment, half control)	Type 1 error: Design-based, v*=0	Type 1 error: Design-based, v*=2	Type 1 error: CRSE, t-test with df=m-1	Standard error: True	Standard error: Mean design-based, v*=0	Standard error: Mean design-based, v*=2	Standard error: Mean CRSE
Two individual-level covariates with R^2_{TX} adjustment							
Normal distributions							
8	4.98%	1.31%	7.40%	0.187	0.188	0.230	0.174
10	4.98%	2.14%	6.45%	0.176	0.177	0.204	0.167
12	5.12%	2.84%	6.41%	0.154	0.155	0.174	0.148
16	4.98%	3.38%	5.85%	0.141	0.142	0.153	0.137
20	5.10%	3.79%	5.76%	0.124	0.125	0.133	0.122
50	5.11%	4.63%	5.32%	0.079	0.080	0.081	0.079
Chi-square distributions							
8	4.28%	1.37%	6.52%	0.189	0.188	0.230	0.174
10	4.49%	1.90%	5.93%	0.172	0.172	0.199	0.162
12	4.72%	2.50%	5.92%	0.152	0.153	0.171	0.146
16	4.84%	3.15%	5.65%	0.140	0.140	0.152	0.136
20	4.87%	3.60%	5.53%	0.125	0.126	0.133	0.122
50	5.06%	4.54%	5.31%	0.081	0.082	0.084	0.081
Two cluster-level covariates with R^2_{TX} adjustment							
Normal distributions							
8	NA	5.10%	17.92%	0.236	NA	0.229	0.162
10	NA	4.80%	12.94%	0.197	NA	0.195	0.153
12	NA	5.21%	11.61%	0.185	NA	0.184	0.152
16	NA	5.16%	8.92%	0.151	NA	0.151	0.134
20	NA	5.04%	7.95%	0.134	NA	0.134	0.122
50	NA	5.20%	6.00%	0.081	NA	0.082	0.079
Chi-square distributions							
8	NA	5.26%	18.04%	0.232	NA	0.227	0.161
10	NA	4.79%	12.20%	0.201	NA	0.200	0.158
12	NA	4.81%	10.39%	0.172	NA	0.171	0.143
16	NA	4.96%	8.52%	0.147	NA	0.147	0.130
20	NA	5.02%	7.64%	0.131	NA	0.132	0.120
50	NA	4.96%	5.76%	0.081	NA	0.082	0.079

Notes: Results obtained from 1,000 simulation draws from the randomization distribution for each of 100 datasets assuming the distributions indicated above. See the main text and Appendix B for simulation specifications. The design-based variance estimator uses (9) in the main text. NA = Not applicable, CRSE = Standard cluster-robust standard error estimator; df = Degrees of freedom; v^*= Degrees of freedom adjustment for the covariates for the DB estimator.

Table B.3. Type 1 error rates for various model specifications

	Number of clusters (half treatment, half control)					
Model specification and estimator	8	10	12	16	20	50
Design-based estimator (finite population (FP) heterogeneity term excluded except where otherwise noted)						
No covariates, average of 40 individuals per cluster	5.15%	5.45%	5.25%	5.01%	5.03%	5.18%
Individual-level covariates, no R^2_{TX} adjustment, v*=0	4.99%	5.01%	5.12%	4.99%	5.11%	5.11%
FP heterogeneity term included	6.01%	5.63%	5.67%	5.45%	5.44%	5.21%
Individual-level covariates, no R^2_{TX} adjustment, v*=2	1.31%	2.17%	2.86%	3.39%	3.80%	4.63%
FP heterogeneity term included	1.69%	2.50%	3.19%	3.76%	4.07%	4.75%
Cluster-level covariates, no R^2_{TX} adjustment, v*=2	9.20%	7.90%	7.80%	6.97%	6.42%	5.67%
FP heterogeneity term included	10.15%	8.62%	8.43%	7.46%	6.82%	5.80%
1 Individual-level, 1 cluster-level covariate, R^2_{TX} adjustment						
v*=0	7.71%	6.76%	6.48%	5.99%	5.80%	5.26%
v*=1	5.16%	5.04%	5.11%	5.06%	5.04%	5.03%
v*=2	2.61%	3.37%	3.74%	4.12%	4.34%	4.78%
Standard CRSE estimator						
No covariates, average of 40 individuals per cluster	7.14%	6.97%	6.44%	5.89%	5.67%	5.41%
Individual-level covariates: alignment with design-based estimator						
Align degrees of freedom (df) in variance formulas, v*=0	5.83%	5.27%	5.39%	5.10%	5.18%	5.12%
Align df in variance formulas and t-tests, v*=0	5.06%	4.99%	5.12%	4.99%	5.11%	5.11%
Cluster-level covariates: alignment with design-based estimator						
Align df in variance formulas, v*=2	10.21%	7.60%	7.49%	6.22%	5.84%	5.29%
Align df in variance formulas and t-tests, v*=2	6.95%	6.16%	6.41%	5.74%	5.54%	5.25%
1 Individual-level and 1 cluster-level covariate						
No alignment	12.44%	9.60%	8.62%	7.41%	6.70%	5.54%
Align df in variance formulas, v*=1	8.19%	6.72%	6.28%	5.73%	5.46%	5.10%
Align df in variance formulas and t-tests, v*=1	6.56%	5.89%	5.70%	5.45%	5.28%	5.07%

Notes: Results obtained from 1,000 simulation draws from the randomization distribution for each of 100 datasets assuming normal random variables. See the main text and Appendix B for simulation specifications.

Table B.4. Super-population simulation results for models without covariates

Number of clusters (half treatment, half control)	Type 1 error: Design-based, t-test with df=m-2	Type 1 error: CRSE, t-test with df=m-1	Standard error: True	Standard error: Mean design-based	Standard error: Mean CRSE
Random variables with normal distributions					
8	4.91%	7.00%	0.238	0.229	0.212
10	5.02%	6.58%	0.214	0.207	0.195
12	5.27%	6.40%	0.197	0.190	0.182
16	5.01%	5.89%	0.169	0.166	0.160
20	5.01%	5.66%	0.152	0.149	0.145
50	5.06%	5.33%	0.096	0.095	0.094
Random variables with bimodal distributions					
8	5.27%	7.25%	0.239	0.230	0.213
10	5.20%	6.80%	0.214	0.208	0.196
12	5.28%	6.50%	0.195	0.191	0.182
16	5.08%	5.87%	0.169	0.167	0.161
20	5.13%	5.80%	0.152	0.150	0.146
50	5.22%	5.46%	0.097	0.095	0.094
Random variables with chi-square distributions					
8	4.67%	6.72%	0.244	0.224	0.208
10	4.69%	6.21%	0.221	0.205	0.193
12	4.68%	5.92%	0.202	0.190	0.181
16	4.69%	5.61%	0.174	0.166	0.161
20	5.01%	5.67%	0.157	0.150	0.146
50	4.97%	5.21%	0.099	0.098	0.097

Notes: Results obtained from 50,000 simulation draws assuming the distributions indicated above. See the main text and Appendix B for simulation specifications. The design-based variance estimator uses (9) in the main text.

References

Aronow, P. M. and Middleton, J. A. (2013). A class of unbiased estimators of the average treatment effect in randomized experiments. *Journal of Causal Inference*, 1(1):135–154.

Hartley, H. and Ross, A. (1954). Unbiased ratio estimators. *Nature*, 174(4423):270.

Li, X. and Ding, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association*, 112(520):1759–1769.

Pashley, N. E. (2019). Note on the delta method for finite population inference with applications to causal inference. Manuscript in progress.

Scott, A. and Wu, C.-F. (1981). On the asymptotic distribution of ratio and regression estimators. *Journal of the American Statistical Association*, 76(373):98–102.