A New Design-Based Variance Estimator for Finely Stratified Experiments *

Yuehao Bai
Department of Economics
University of Southern California
yuehao.bai@usc.edu

Xun Huang
Department of Economics
University of Chicago
xhuang520@uchicago.edu

Joseph P. Romano
Departments of Economics & Statistics
Stanford University
romano@stanford.edu

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

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

March 13, 2025

Abstract

This paper considers the problem of design-based inference for the average treatment effect in finely stratified experiments. Here, by "design-based" we mean that the only source of uncertainty stems from the randomness in treatment assignment itself; by "finely stratified" we mean units are first stratified into groups of a fixed size k according to baseline covariates and then, within each group, a fixed number $\ell < k$ are assigned uniformly at random to treatment and the remainder to control. In this setting, we first show under mild conditions that inference using the usual difference-in-means estimator requires an estimator of its variance that is at least asymptotically upward-biased. We then present a novel estimator of the variance and show that it is upward-biased; furthermore, the magnitude of the bias depends in a natural way on the quality of the stratification. Importantly, this estimator remains well-defined even in the challenging setting in which $\ell = 1$ or $k - \ell = 1$. We then compare our estimator with some well-known estimators that have been proposed previously for this case. We first show that, while these estimators are also upward-biased, the magnitude of their bias does not change in the natural way with the quality of stratification. To further discriminate among these different estimators, we then introduce a framework motivated by a thought experiment in which the finite population can be modeled as having been drawn once in an i.i.d. fashion from a well-behaved probability distribution. In this framework, we argue that our estimator dominates the others in terms of limiting bias, which leads to more precise inferences, and that these improvements are strict except under exceptionally strong restrictions on the treatment effects. Finally, we illustrate the practical relevance of our theoretical results through a simulation study, which reveals that our estimator can in fact lead to substantially more precise inferences, especially when the quality of stratification is high.

KEYWORDS: Experiments, Finite Population, Average Treatment Effect, Matched Pairs, Stratification

JEL classification codes: C12, C31, C35, C36

1   Introduction

This paper considers the problem of design-based inference on the average treatment effect in finely stratified experiments. Here, by "design-based" we mean that there is no sampling uncertainty and the only randomness stems from the variation in treatment assignment itself; by "finely stratified" we mean that treatment is assigned in the following fashion: units are first stratified into groups of a fixed size k according to baseline covariates and then, within each group, a fixed number $\ell < k$ are assigned uniformly at random to treatment and the remainder to control. A prominent special case of this framework is a matched pairs design, in which $k = 2$ and $\ell = 1$.

In this setting, we first establish, by way of motivation, a result that shows under mild conditions that inference using the usual difference-in-means estimator requires an estimator of its variance that is at least asymptotically upward-biased. After having motivated the importance of such estimators, we review the standard argument for why constructing such estimators is particularly challenging for the case in which $\ell = 1$ or $k - \ell = 1$; in particular, in this case the sample variance of the outcomes under treatment (if $\ell = 1$) or control (if $k - \ell = 1$) computed in each stratum are identically zero. With this in mind, we propose a novel estimator that remains well-defined even in this challenging case, and demonstrate that our estimator is upward-biased under minimal assumptions. The construction of our variance estimator involves a pairing of the strata, and its bias depends on the corresponding differences of the average treatment effects in these paired strata. As a result, if the strata are paired in such a way that ensures "similar" strata are adjacent, then the bias of our estimator will naturally be small.

We then compare our variance estimator with two well-known variance estimators that have been proposed previously for the case in which $\ell = 1$ or $k - \ell = 1$. The first estimator, proposed by Imai (2008), is based on the sample variance obtained by viewing the difference-in-means estimates for each stratum as independent observations. The second estimator, proposed by Fogarty (2018a), modifies the previous estimator by partialling out the average covariate values in each stratum. Existing results from these papers show that both estimators are also upward-biased for the variance of the difference-in-means estimator but, in contrast to our estimator, their biases do not depend directly on the quality of the stratification. To further discriminate among these estimators and ours, we introduce a framework motivated by a thought experiment in which the finite population can be modeled as being drawn once in an i.i.d. fashion from a well-behaved probability distribution, and the strata are formed in such a way that the covariate values of the units in each stratum, as well as the paired strata, get close in the limit. In this framework, we show that the limiting bias of our estimator is strictly smaller than that of Imai (2008) unless the treatment effects are homogeneous, and strictly smaller than that of Fogarty (2018a) unless the conditional average treatment effect is linear in the covariates. As a result, confidence intervals based on our variance estimator will be strictly shorter than those based on the other variance estimators unless these exceptional restrictions hold.

The literature on stratified block randomization dates back to at least Fisher (1935). For the case in which $\ell = 1$ or $k - \ell = 1$, the variance estimator in Imai (2008) has been extended to different settings by Imai et al. (2009), Pashley and Miratrix (2021), and Zhu et al. (2024). Fogarty (2018b) studies regression adjustment for matched pair designs, and Liu and Yang (2020) studies regression adjustment for stratified designs primarily for the case where $\min\{\ell, k - \ell\} > 1$. Ding (2017) considers randomization inference

for matched pairs, alongside other randomization schemes. We note that all of these papers employ a design-based framework as we do in this paper. As a consequence, their analyses differ from work which studies inference in finely stratified experiments from a superpopulation perspective (Bai, 2022; Bai et al., 2022, 2024a,b,d; Jiang et al., 2024; Cytrynbaum, 2024a,b; Bai et al., 2025); see also Abadie and Imbens (2008), who consider superpopulation estimation of the variance in matched pair designs, conditional on the covariates, in an alternative sampling framework where pairs are sampled instead of the units.

The remainder of the paper is organized as follows. In Section 2, we describe our setup and notation. Section 3 contains our main results. There we first highlight the importance of upward-biased variance estimators, and then review existing proposals before introducing our novel variance estimator. In Section 4, we compare our variance estimators with existing ones in a particular limiting thought experiment. We examine the finite-sample behavior of confidence intervals based on all these estimators through a simulation study in Section 5.

2 Setup and Notation

Consider an experiment consisting of $i \in \{1, \ldots, n\}$ units. For the ith unit, let $Y_i \in \mathbf{R}$ denote their observed outcome, $D_i \in \{0, 1\}$ denote their received treatment, $X_i \in \mathbf{R}^p$ denote their observed, baseline covariates, and $Y_i(d)$ denote their potential outcome under treatment $d \in \{0, 1\}$. As usual, the observed outcomes are related to the potential outcomes via the relationship

$$Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i) \ . \tag{1}$$

In what follows, it will be convenient to use the following shorthand notation: for a generic random vector A_i indexed by i, let $A^{(n)} := (A_i : 1 \le i \le n)$. For later use, we also define $W_i := (Y_i(1), Y_i(0), X_i)'$.

In the design-based framework that we maintain throughout the paper, the potential outcomes and covariates of the units in the experiment are modeled as nonrandom quantities, with the only source of randomness arising from the treatment assignment mechanism. Our analysis concerns "finely stratified" designs in which the covariates are used to stratify units in the experiment into groups, i.e., strata, of *fixed* size k, and then, within each group, $\ell < k$ units are chosen uniformly at random to assign to treatment and the remainder to control; of particular interest to us will be the special case when $\min\{\ell, k - \ell\} = 1$. For ease of exposition, we assume throughout that $n = mk$, so that m denotes the number of strata. Formally, we model the strata as a partition of $\{1, \ldots, n\}$:

$$\Lambda_n := \{\lambda_j \subset \{1, \ldots, n\} : 1 \le j \le m\} \ ,$$

with $|\lambda_j| = k$. It is worth emphasizing that we have suppressed in the notation the fact that each λ_j (and therefore Λ_n itself) can depend on $X^{(n)}$. Since we maintain that k is fixed throughout the paper, when we write that $n \to \infty$, it should be understood that $m \to \infty$. Using this notation, the (joint) distribution of treatment assignment is characterized by the following assumption:

Assumption 2.1. Treatment status $D^{(n)}$ is assigned independently for each $1 \leq j \leq m$ so that

$$(D_i : i \in \lambda_j) \sim \text{Unif}\left(\left\{(d_1, \ldots, d_k) \in \{0,1\}^k : \sum_{1 \leq i \leq k} d_i = \ell\right\}\right) .$$

In other words, ℓ out of k units in each stratum are treated uniformly at random, independently across strata.

Our parameter of interest is the finite population average treatment effect, given by

$$\Delta_n := \bar{Y}_n(1) - \bar{Y}_n(0) ,$$

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

$$\bar{Y}_n(d) := \frac{1}{n} \sum_{1 \leq i \leq n} Y_i(d) .$$

A natural estimator of Δ_n is given by the usual difference-in-means estimator, i.e.,

$$\hat{\Delta}_n := \frac{1}{n(1)} \sum_{1 \leq i \leq n} Y_i D_i - \frac{1}{n(0)} \sum_{1 \leq i \leq n} Y_i(1 - D_i) ,$$

where $n(1) = \sum_{1 \leq i \leq n} D_i = n\eta$, with $\eta := \ell/k$ the proportion of the n units that are treated, and $n(0) = \sum_{1 \leq i \leq n}(1 - D_i) = n(1 - \eta)$. Note that because we have assumed that $n = mk$, both $n\eta = m\ell$ and $n(1 - \eta) = m(k - \ell)$ are integers.

Remark 2.1. The analysis in this paper can be readily extended to cluster RCTs where the treatment is implemented at the cluster level (see, for instance, Imai et al., 2009; Su and Ding, 2021; De Chaisemartin and Ramirez-Cuellar, 2024; Bai et al., 2024c). Let $i \in \{1, \ldots, n\}$ denote the ith cluster and g_i denote the number of units in the ith cluster. Let D_i denote the treatment assignment for the ith cluster, which we emphasize applies to all units in this cluster. For each i and $1 \leq t \leq g_i$, let $Y_{i,t}(1), Y_{i,t}(0)$ denote the nonrandom potential outcomes for the tth unit in the ith cluster and let $Y_{i,t}$ denote its observed outcome. Suppose the parameter of interest is the average treatment effect across all units:

$$\Delta_n^\dagger = \frac{1}{\sum_{1 \leq i \leq n} g_i} \sum_{1 \leq i \leq n} \sum_{1 \leq t \leq g_i} (Y_{i,t}(1) - Y_{i,t}(0)) .$$

Note that $\Delta_n^\dagger = \frac{1}{n} \sum_{1 \leq i \leq n}(Y_i^\dagger(1) - Y_i^\dagger(0))$, where for $d \in \{0,1\}$,

$$Y_i^\dagger(d) = \frac{1}{\frac{1}{n} \sum_{1 \leq i \leq n} g_i} \sum_{1 \leq t \leq g_i} Y_{i,t}(d) .$$

Accordingly, let $Y_i^\dagger = Y_i^\dagger(1)D_i + Y_i^\dagger(0)(1 - D_i)$, and a natural estimator for $\Delta_n^\dagger$ is given by

$$\hat{\Delta}_n^\dagger = \frac{1}{n(1)} \sum_{1 \leq i \leq n} Y_i^\dagger D_i - \frac{1}{n(0)} Y_i^\dagger(1 - D_i) ,$$

where $n(1)$ and $n(0)$ are defined as before, and in this case are the numbers of treated and untreated clusters. The analysis in this paper then immediately applies by replacing Y_i by $Y_i^\dagger$ throughout. ∎

3 Main Results

In this section, we begin, by way of motivation, with a result that highlights the importance of variance estimators that are at least *asymptotically* upward-biased, defined precisely in the statement of Theorem 3.1 below. Under mild regularity conditions, we show, in particular, that valid inference for Δ_n based on a standard t-statistic requires a variance estimator that is asymptotically upward-biased for $\mathrm{Var}[\hat{\Delta}_n]$. Of course, a simple, sufficient condition for an estimator to be asymptotically upward-biased is that it is upward-biased. In Section 3.2, we review some prior proposals for constructing variance estimators in finely stratified experiments that are upward-biased under only the assumption that $D^{(n)}$ satisfies Assumption 2.1. We further discuss in Remark 3.1 some improvements upon these estimators that typically result in estimators that are only asymptotically upward-biased under assumptions stronger than Assumption 2.1 and therefore result in valid inference less generally than their upward-biased counterparts. Finally, in Section 3.3, we present our main results. We introduce a novel estimator of $\mathrm{Var}[\hat{\Delta}_n]$ for finely stratified experiments and show that, like the estimators that we review in Section 3.2, it is upward-biased for $\mathrm{Var}[\hat{\Delta}_n]$ under only the assumption that $D^{(n)}$ satisfies Assumption 2.1. In contrast to these other estimators, however, we further argue that the magnitude of the bias of our estimator depends explicitly on the quality of the groupings of the experimental units into strata, in a sense made precise by Theorem 3.2 below.

3.1 Motivating Result

The following theorem formalizes the sense in which an asymptotically upward-biased estimator for $\mathrm{Var}[\hat{\Delta}_n]$ is required for valid inference on Δ_n.

Theorem 3.1. *Suppose $D^{(n)}$ satisfies Assumption 2.1 and consider a sequence of finite populations such that*

$$\frac{1}{n} \max_{d \in \{0,1\}} \max_{1 \leq i \leq n} Y_i(d)^2 \to 0 \tag{2}$$

as $n \to \infty$, and

$$0 < \liminf_{n \to \infty} n \cdot \mathrm{Var}[\hat{\Delta}_n] \leq \limsup_{n \to \infty} n \cdot \mathrm{Var}[\hat{\Delta}_n] < \infty . \tag{3}$$

Further assume that $\tilde{V}_n$ is an estimator of $\mathrm{Var}[\hat{\Delta}_n]$ such that

$$n \cdot \left| \tilde{V}_n - E[\tilde{V}_n] \right| \xrightarrow{P} 0 \tag{4}$$

as $n \to \infty$. Then,

$$\liminf_{n \to \infty} P\left\{ \hat{\Delta}_n - \sqrt{\tilde{V}_n} \cdot z_{1-\alpha/2} \leq \Delta_n \leq \hat{\Delta}_n + \sqrt{\tilde{V}_n} \cdot z_{1-\alpha/2} \right\} \geq 1 - \alpha , \tag{5}$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution, if and only if

$$\liminf_{n \to \infty} n \left(E[\tilde{V}_n] - \mathrm{Var}[\hat{\Delta}_n] \right) \geq 0 . \tag{6}$$

Note that because $\tilde{V}_n$ is defined as an estimator of $\mathrm{Var}[\hat{\Delta}_n]$ and not the asymptotic variance of $\sqrt{n}(\hat{\Delta}_n -$

Δ_n), the conditions in the theorem are all stated with a scaling by n. When (6) holds, we say that $\tilde{V}_n$ is asymptotically upward-biased for $\mathrm{Var}[\hat{\Delta}_n]$. A simple sufficient condition for (6) is, of course, that

$$E[\tilde{V}_n] \geq \mathrm{Var}[\hat{\Delta}_n] \text{ for all } n \geq 1 . \tag{7}$$

When (7) holds, we say that $\tilde{V}_n$ is upward-biased for $\mathrm{Var}[\hat{\Delta}_n]$. Theorem 3.1 thus demonstrates that, under mild regularity conditions, confidence intervals for Δ_n of the form $[\hat{\Delta}_n \pm \sqrt{\tilde{V}_n} \cdot z_{1-\alpha/2}]$ are valid in the sense of having the right limiting coverage probability, if and only if $\tilde{V}_n$ is asymptotically upward-biased for $\mathrm{Var}[\hat{\Delta}_n]$. These assumptions include weak restrictions on the finite population (i.e., (2)), a non-degeneracy condition (i.e., (3)), and a requirement that $n \cdot \tilde{V}_n$ is consistent for its expectation (i.e., (4)). This last condition will typically hold under some further restrictions on finite population moments; see, e.g., Theorem 4.2 below.

The estimators discussed in Sections 3.2 and 3.3 below can be shown to satisfy (7), and therefore (6), whenever $D^{(n)}$ satisfies Assumption 2.1. Such estimators therefore lead to confidence intervals for Δ_n that are valid whenever the hypotheses of Theorem 3.1 are satisfied. It is worth emphasizing, however, that establishing (6) for other estimators may in general require additional assumptions beyond the hypotheses of Theorem 3.1. In Remarks 3.1 and 3.2 we discuss some examples of such estimators, and we illustrate using our simulations in Section 5 that their resulting confidence intervals may lead to under-coverage when these additional conditions are not satisfied.

3.2 Review of Some Upwards Biased Estimators of $\mathrm{Var}[\hat{\Delta}_n]$

In this section, we review some upward-biased estimators of $\mathrm{Var}[\hat{\Delta}_n]$. To this end, recall that it can be shown using standard arguments (see, for instance, Imbens and Rubin, 2015) that

$$\mathrm{Var}[\hat{\Delta}_n] = \frac{1}{nm} \sum_{1 \leq j \leq m} \left(\frac{S_j^2(1)}{\eta} + \frac{S_j^2(0)}{1-\eta} - S_{j,\Delta}^2 \right) , \tag{8}$$

where

$$S_j^2(d) := \frac{1}{k-1} \sum_{i \in \lambda_j} \left(Y_i(d) - \bar{Y}_{j,n}(d) \right)^2 , \qquad S_{j,\Delta}^2 := \frac{1}{k-1} \sum_{i \in \lambda_j} \left(Y_i(1) - Y_i(0) - \Delta_{j,n} \right)^2 ,$$

with

$$\bar{Y}_{j,n}(d) := \frac{1}{k} \sum_{i \in \lambda_j} Y_i(d) , \qquad \Delta_{j,n} := \frac{1}{k} \sum_{i \in \lambda_j} \left(Y_i(1) - Y_i(0) \right) .$$

In settings where $\min\{\ell, k - \ell\} > 1$, the construction of an upward-biased variance estimator is straightforward: an unbiased estimator of $S_j^2(d)$ for $d \in \{0,1\}$ is given by the sample variance of the outcomes for units within stratum j assigned to treatment d, which we denote by $\hat{S}_j^2(d)$ (see Imbens and Rubin, 2015; Pashley and Miratrix, 2021). A simple estimator of $\mathrm{Var}[\hat{\Delta}_n]$ is thus given by

$$\frac{1}{nm} \sum_{1 \leq j \leq m} \left(\frac{\hat{S}_j^2(1)}{\eta} + \frac{\hat{S}_j^2(0)}{1-\eta} \right) , \tag{9}$$

and its bias is $\frac{1}{nm} \sum_{1 \leq j \leq m} S_{j,\Delta}^2 \geq 0$. Note that this estimation strategy exploits the lower bound $S_{j,\Delta}^2 \geq 0$; we return to this observation in Remark 3.1 below.

This estimation strategy fails, however, when there is exactly one treated or control unit in a stratum, i.e., when $\min\{\ell, k - \ell\} = 1$, since in this case the corresponding sample variance is identically zero. In these settings, a canonical estimator considered in the literature instead computes the variance of the stratum-level average treatment effect estimates (Imai, 2008):

$$\hat{V}_n^{\text{IM}} := \frac{1}{m(m-1)} \sum_{1 \leq j \leq m} \left(\hat{\Delta}_{j,n} - \hat{\Delta}_n \right)^2 ,$$

where

$$\hat{\Delta}_{j,n} := \frac{1}{\ell} \sum_{i \in \lambda_j} Y_i D_i - \frac{1}{k - \ell} \sum_{i \in \lambda_j} Y_i (1 - D_i)$$

is the difference in means in the jth stratum. This variance estimator serves as the basic scaffolding for many of the recent estimators proposed in the literature on inference in stratified randomized experiments: it is a special case of the small-block variance estimator proposed in Pashley and Miratrix (2021) when all strata are the same size (see also Zhu et al. (2024)), a special case of the pair-cluster variance estimator of De Chaisemartin and Ramirez-Cuellar (2024) in the setting of an individual-level randomized experiment, and a special case of the regression-based estimator due to Fogarty (2018a), which we now describe. Let R be an $m \times L$ matrix for $L < m$ with rank L and $H_R = R(R'R)^{-1}R'$ denote the projection matrix onto its column space. For each R, consider the estimator given by

$$\hat{V}_n^{\text{F}}(R) = \frac{1}{m^2} \hat{\delta}_n' (\text{diag}(I - H_R))^{-1/2} (I - H_R) (\text{diag}(I - H_R))^{-1/2} \hat{\delta}_n, \tag{10}$$

where $\hat{\delta}_n = (\hat{\Delta}_{1,n}, \ldots, \hat{\Delta}_{m,n})'$. Note that $\hat{V}_n^{\text{F}}(\iota_m) = \hat{V}_n^{\text{IM}}$, where ι_m is an $m \times 1$ vector of ones, so that $\hat{V}_n^{\text{F}}(R)$ is indeed a generalization of $\hat{V}_n^{\text{IM}}$. Fogarty (2018a) focuses on the estimator given by $\hat{V}_n^{\text{F}} = \hat{V}_n^{\text{F}}(Q_2)$, where

$$Q_2 = \begin{pmatrix} & (\bar{X}_{1,n} - \mu_{X,n})' \\ \iota_m & \vdots \\ & (\bar{X}_{m,n} - \mu_{X,n})' \end{pmatrix}$$

for $\bar{X}_{j,n} = \frac{1}{k} \sum_{i \in \lambda_j} X_i$ and $\mu_{X,n} = \frac{1}{n} \sum_{1 \leq i \leq n} X_i$. In particular, Fogarty (2018a) argues that $\hat{V}_n^{\text{F}}$ can be less conservative than $\hat{V}_n^{\text{IM}}$ whenever the covariates are predictive of the treatment effects in an appropriate sense.

Note that the bias of $\hat{V}_n^{\text{IM}}$ is given by (see Imbens and Rubin, 2015; Fogarty, 2018a)

$$E[\hat{V}_n^{\text{IM}}] - \text{Var}[\hat{\Delta}_n] = \frac{1}{m(m-1)} \sum_{1 \leq j \leq m} (\Delta_{j,n} - \Delta_n)^2 \geq 0 , \tag{11}$$

so that $\hat{V}_n^{\text{IM}}$ is an upward-biased estimator of $\text{Var}[\hat{\Delta}_n]$. The bias for $\hat{V}_n^{\text{F}}$ is (see Fogarty, 2018a)

$$E[\hat{V}_n^{\text{F}}] - \text{Var}[\hat{\Delta}_n] = \frac{1}{m^2} \delta_n' (\text{diag}(I - H_{Q_2}))^{-1/2} (I - H_{Q_2}) (\text{diag}(I - H_{Q_2}))^{-1/2} \delta_n \geq 0 ,$$

where $\delta_n = (\Delta_{1,n}, \ldots, \Delta_{m,n})'$, so $\hat{V}_n^{\mathrm{F}}$ is also an upward-biased estimator for $\mathrm{Var}[\hat{\Delta}_n]$. Moreover, under mild assumptions, it is straightforward to establish using Chebyshev's inequality that $\hat{V}_n^{\mathrm{IM}}$ and $\hat{V}_n^{\mathrm{F}}$ are consistent for their expectations in the sense of (4), and thus can be used to construct valid confidence intervals in the sense of (5).

Remark 3.1. In some circumstances, it may be further possible to bound $S_{j,\Delta}^2$ from below by a positive number instead of zero, and this tighter bound could then be used to construct less conservative estimators of $\mathrm{Var}[\hat{\Delta}_n]$. For instance, it follows from the Cauchy-Schwarz inequality that

$$\sum_{i \in \lambda_j} (Y_i(1) - \bar{Y}_{j,n}(1))(Y_i(0) - \bar{Y}_{j,n}(0)) \leq \left(\sum_{i \in \lambda_j} (Y_i(1) - \bar{Y}_{j,n}(1))^2 \right)^{1/2} \left(\sum_{i \in \lambda_j} (Y_i(0) - \bar{Y}_{j,n}(0))^2 \right)^{1/2} ,$$

from which we can deduce the lower bound

$$S_{j,\Delta}^2 \geq (S_j(1) - S_j(0))^2 \geq 0 .$$

In fact, it is possible to achieve even tighter bounds by further exploiting the structure of $S_{j,\Delta}^2$; see Aronow et al. (2014) for details. We note, however, that we are not aware of estimators based on lower bounds other than $S_{j,\Delta}^2 \geq 0$ which are necessarily guaranteed to be upward-biased, and in general additional assumptions need to be imposed in order to guarantee that the resulting estimators are at least asymptotically upward-biased; see, for instance, Corollary 1 in Aronow et al. (2014). As a consequence, these improved estimators may not lead to valid confidence intervals whenever these additional assumptions fail to hold. ∎

3.3 A Novel Paired-Strata Estimator of $\mathrm{Var}[\hat{\Delta}_n]$

Building on our earlier work on super-population approaches to inference for finely stratified designs (Bai et al., 2022, 2024a,b,d, 2025), we now propose a novel paired-strata estimator of $\mathrm{Var}[\hat{\Delta}_n]$, primarily for settings where $\min\{\ell, k - \ell\} = 1$. For simplicity, we pair adjacent strata together, so the pairs are given by $\{(\lambda_{2j-1}, \lambda_{2j}) : 1 \leq j \leq \lfloor m/2 \rfloor\}$, but we emphasize that this pairing should be understood as being obtained after permuting the indices of the strata $\{\lambda_j : 1 \leq j \leq m\}$. In practice, it will be desirable to do this in a way that ensures "similar" strata are adjacent. With this is mind, our estimator is defined as follows:

$$\hat{V}_n := \frac{1}{m} \left(\hat{\tau}_n^2 - \hat{\kappa}_n \right), \tag{12}$$

where

$$\hat{\tau}_n^2 = \frac{1}{m} \sum_{1 \leq j \leq m} \hat{\Delta}_{j,n}^2$$

and

$$\hat{\kappa}_n = \frac{2}{m} \sum_{1 \leq j \leq \lfloor \frac{m}{2} \rfloor} \hat{\Delta}_{2j-1,n} \hat{\Delta}_{2j,n} .$$

Note that it follows from straightforward algebraic manipulation that $\hat{V}_n \geq 0$. The construction of $\hat{V}_n$ can be motivated using a specific limiting thought experiment that we introduce in Section 4 below. The following theorem shows that $\hat{V}_n$ is an upward-biased estimator of $\mathrm{Var}[\hat{\Delta}_n]$ when $D^{(n)}$ satisfies Assumption 2.1 and,

under mild additional restrictions, is consistent for its expectation.

Theorem 3.2. *Suppose $D^{(n)}$ satisfies Assumption 2.1. Then,*

(a) *The bias of $\hat{V}_n$ is given by*

$$E[\hat{V}_n] - \text{Var}[\hat{\Delta}_n] = \xi_n^2 := \frac{1}{m^2} \sum_{1 \le j \le \lfloor \frac{m}{2} \rfloor} (\Delta_{2j-1,n} - \Delta_{2j,n})^2 \ge 0 .$$

(b) *Suppose in addition that*

$$\frac{1}{n} \sum_{1 \le i \le n} Y_i(d)^4 = o(n) ,$$

holds for $d \in \{0, 1\}$. Then, (4) holds as $n \to \infty$.

Therefore, by combining Theorems 3.1 and 3.2, we immediately obtain the asymptotic validity of confidence intervals constructed using $\hat{V}_n$.

Theorem 3.2 justifies the use of $\hat{V}_n$ for inference about Δ_n and further demonstrates that the magnitude of its bias depends on the differences of the stratum-level average treatment effects across pairs of strata, and hence the bias decreases whenever the pairs of strata are formed in a way that increases their homogeneity in this sense. The bias will therefore be small not only in the extreme instance in which the stratum-level average treatment effects are similar across all strata, but also when stratum-level average treatment effects vary across strata and are only similar for adjacent strata. The latter scenario is what one might naturally expect when strata are formed by grouping individuals with similar values of X_i. In contrast, the biases of $\hat{V}_n^{\text{IM}}$ and $\hat{V}_n^{\text{F}}$ do not share this property. We explore this more formally in Section 4 through further restrictions on $W^{(n)}$ and Λ_n that are motivated by a particular limiting thought experiment.

Remark 3.2. An alternative construction for an estimator of $\text{Var}[\hat{\Delta}_n]$ based on our paired strata $\{(\lambda_{2j-1}, \lambda_{2j}) : 1 \le j \le \lfloor m/2 \rfloor\}$ would be to apply the estimation strategy described in (9) to the paired strata directly. However, such an estimator is not guaranteed to be upward-biased in general. In particular, it can be shown that in the case of a matched pairs design (i.e. when $\ell = 1, k = 2$) that this estimator coincides with the estimator $\hat{V}_n^{\text{alt}}$ proposed in Section 5. There we demonstrate via simulation that confidence intervals for Δ_n constructed using this estimator may not have correct coverage in general. ■

Remark 3.3. Because the bias of $\hat{V}_n^{\text{IM}}$ only depends $\Delta_{j,n} - \Delta_n$, it may be minimized in extreme instances when strata are formed quite unconventionally. To illustrate this counterintuitive phenomenon, suppose for example $\eta = \frac{1}{2}$, n is even, and $X_i = \frac{i}{n}$ for $1 \le i \le n$. Further suppose both $Y(1)$ and $Y(0)$ are linear functions of X. Suppose the strata are formed as $\lambda_j = \{j, n + 1 - j\}$ for $1 \le j \le \frac{n}{2}$, i.e., the unit with the smallest value of X is matched with the unit with the largest value of X, the second smallest with the second largest, and so on. Then, the linearity of potential outcomes implies $\Delta_{j,n} \equiv \Delta_n$, so that the bias in (11) is 0. On the other hand, a natural way to form the strata, and one that minimizes the sum of squared pairwise distances of X, is given by $\lambda_j = \{2j - 1, 2j\}$ for $1 \le j \le \frac{n}{2}$, under which the bias in (11) is not 0. ■

8

4 Comparison with Other Variance Estimators

In this section, we provide a framework that permits us to further discriminate among the different variance estimators of $\text{Var}[\hat{\Delta}_n]$ discussed in Section 3. This framework is formalized by restrictions on the sequence of finite populations and experimental designs $(W^{(n)}, \Lambda_n)$ specified in the assumptions below. These restrictions are motivated by a limiting thought experiment in which $W^{(n)}$ is itself a realization of an i.i.d. sample from a fixed probability distribution, and the strata Λ_n are formed in such a way that units with similar observable characteristics are grouped together, but, as explained in Remark 4.1 below, the restrictions can also be shown to hold in other contexts as well.

Assumption 4.1. For some random variable $\tilde{W} := (\tilde{Y}(1), \tilde{Y}(0), \tilde{X}) \sim Q$ with $\max_{d \in \{0,1\}} E_Q[\tilde{Y}^2(d)] < \infty$, $W^{(n)}$ and Λ_n satisfy the following requirements:

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

$$\frac{1}{n} \sum_{1 \leq i \leq n} Y_i^r(d) \to E_Q[\tilde{Y}^r(d)] \text{ for } d \in \{0,1\} \text{ and } r \in \{1,2\}$$

$$\frac{1}{n} \sum_{1 \leq i \leq n} Y_i(1)Y_i(0) \to E_Q[\tilde{Y}(1)\tilde{Y}(0)] .$$

(b) For any $(d, d') \in \{0,1\}^2$ and sequence $\{(i_j, i_j') : (i_j, i_j') \in (\lambda_{2j-1} \cup \lambda_{2j})^2, i_j \neq i_j', 1 \leq j \leq \lfloor m/2 \rfloor\}$, as $n \to \infty$,

$$\frac{2}{m} \sum_{1 \leq j \leq \lfloor \frac{m}{2} \rfloor} Y_{i_j}(d)Y_{i_j'}(d') \to E_Q[E_Q[\tilde{Y}(d)|\tilde{X}]E_Q[\tilde{Y}(d')|\tilde{X}]] .$$

For the analysis of $\hat{V}_n^{\text{F}}$ only, we need to further impose the following assumption:

Assumption 4.2. For $\tilde{W} \sim Q$ in Assumption 4.1, $W^{(n)}$ and Λ_n satisfy the following requirements:

(a) As $n \to \infty$,
$$\frac{1}{n} \max_{1 \leq i \leq n} \|X_i\|^2 \to 0 .$$

(b) As $n \to \infty$,

$$\frac{1}{n} \sum_{1 \leq i \leq n} X_i \to E_Q[\tilde{X}]$$

$$\frac{1}{n} \sum_{1 \leq i \leq n} X_i X_i' \to E_Q[\tilde{X}\tilde{X}'] > 0$$

$$\frac{1}{n} \sum_{1 \leq i \leq n} X_i Y_i(d) \to E_Q[\tilde{X}\tilde{Y}(d)] \text{ for } d \in \{0,1\} .$$

(c) For any $d \in \{0,1\}$ and sequence $\{(i_j, i_j') : (i_j, i_j') \in \lambda_j, i_j \neq i_j', 1 \leq j \leq m\}$, as $n \to \infty$,

$$\frac{1}{m} \sum_{1 \leq j \leq m} X_{i_j} X_{i_j'} \to E_Q[\tilde{X}\tilde{X}'] > 0$$

9

$$\frac{1}{m} \sum_{1 \leq j \leq m} X_{i_j} Y_{i'_j}(d) \to E_Q[\tilde{X}\tilde{Y}(d)] \ .$$

Assumption 4.1(a) states that the finite population "moments" of the outcomes should converge to well-defined limit quantities; the condition holds almost surely by the strong law of large numbers if $W^{(n)}$ is modeled as an i.i.d. sample from the probability distribution Q. Assumption 4.1(b) states that the average products of the outcomes in stratum-pairs should converge to well-defined limit quantities in such a way that their covariate values are also being matched in the limit; this condition can also be justified if $W^{(n)}$ is modeled as an i.i.d. sample from Q and the stratification Λ_n satisfies the property that

$$\frac{1}{m} \sum_{1 \leq j \leq \lfloor m/2 \rfloor} \max_{i,i' \in (\lambda_{2j-1} \cup \lambda_{2j})} \|X_i - X_{i'}\|^2 \xrightarrow{P} 0 \ , \tag{13}$$

following similar arguments to those used in Lemma C.2. in Bai et al. (2024d). In words, (13) requires that the average squared distances of the covariate values in a paired stratum converges to zero asymptotically. This property is guaranteed by specific matching algorithms under appropriate assumptions; see, for instance, Bai et al. (2022) and Cytrynbaum (2024b) for details. Assumption 4.2 introduces similar conditions to Assumption 4.1 but also requires that these conditions hold for the covariate vectors $X^{(n)}$. Note in particular that $E_Q[\tilde{X}\tilde{Y}(d)] = E_Q[\tilde{X}E_Q[\tilde{Y}(d)|\tilde{X}]]$ by the law of iterated expectations.

Remark 4.1. Another setting in which Assumptions 4.1–4.2 are naturally satisfied is as follows. Let $X_i = \frac{i}{n}$ for $1 \leq i \leq n$, so that $X^{(n)}$ are equi-distant in the unit interval. Further assume $Y_i(d) = f_d(X_i)$ for $d \in \{0, 1\}$, where f_0 and f_1 are deterministic Lipschitz functions. Define Q such that $\tilde{X} \sim \text{Unif}[0, 1]$ under Q and $\tilde{Y}(d) = f_d(\tilde{X})$ degenerately for $d \in \{0, 1\}$. Then, it can be shown using elementary properties of the Riemann integral that Assumptions 4.1–4.2 hold. ∎

Using Assumption 4.1, we have the following expression for the limit of $\text{Var}[\hat{\Delta}_n]$ after normalizing appropriately:

Theorem 4.1. *Under Assumption 4.1,*

$$n \cdot \text{Var}[\hat{\Delta}_n] \to V \ ,$$

where

$$V := E_Q\left[\frac{\text{Var}_Q[\tilde{Y}(1)|\tilde{X}]}{\eta} + \frac{\text{Var}_Q[\tilde{Y}(0)|\tilde{X}]}{1 - \eta}\right] - E_Q\left[\text{Var}_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}]\right] \ .$$

We can now motivate the construction of $\hat{V}_n$ in Section 3.3 through Theorem 4.1. For the reasons outlined in Remark 3.1, we focus on estimating

$$V^{\text{obs}} := E_Q\left[\frac{\text{Var}_Q[\tilde{Y}(1)|\tilde{X}]}{\eta} + \frac{\text{Var}_Q[\tilde{Y}(0)|\tilde{X}]}{1 - \eta}\right] \ , \tag{14}$$

which is an upper bound for V. The construction of $\hat{V}_n$ can be motivated following arguments similar to those used in our earlier work on the analysis of finely stratified experiments from a super-population perspective

([Bai et al., 2022, 2024a,d, 2025]). First, by expanding the square, it can be shown under Assumption 4.1 that

$$k \cdot \hat{\tau}_n^2 \xrightarrow{P} \frac{1}{\eta} E_Q[\text{Var}_Q[\tilde{Y}(1)|\tilde{X}]] + \frac{1}{1-\eta} E_Q[\text{Var}_Q[\tilde{Y}(0)|\tilde{X}]] + k E_Q[E_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}]^2]$$
$$= V^{\text{obs}} + k E_Q[E_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}]^2] .$$

It thus remains to construct a consistent estimator of the final term. For this purpose, it can also be shown under Assumption 4.1 that

$$k \cdot \hat{\kappa}_n \xrightarrow{P} k E_Q[E_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}]^2] . \tag{15}$$

To understand why it is natural to expect $\hat{\kappa}_n$ to satisfy (15) under Assumption 4.1(b), note that $\hat{\kappa}_n$ is the average of products of the differences in means in a pair of strata. Each pair of differences in means are independent conditional on the stratification. Furthermore, for a sufficiently good stratification, each of them has conditional expectation close to $E_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}]$ for the same value of $\tilde{X}$. For this reason, we expect $\hat{\kappa}_n$ to converge in the desired way.

In the remainder of the section we compare the limits of $\hat{V}_n$, $\hat{V}_n^{\text{IM}}$ and $\hat{V}_n^{\text{F}}$ relative to the asymptotic variance V defined above. Recall from Section 3 that it is possible to establish under weak assumptions on the sequence of population "moments" that each of these variances is consistent for their expectations in the sense of (4). Accordingly, in what follows it suffices to study the limit of the expectations of these estimators under Assumption 4.1, and compare them to V.

Theorem 4.2. *Under Assumption 4.1,*

$$n \cdot E[\hat{V}_n] \to V^{\text{obs}}$$
$$n \cdot E[\hat{V}_n^{\text{IM}}] \to V^{\text{obs}} + k \cdot \text{Var}_Q[E_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}]] =: V^{\text{IM}} ,$$

for V^{obs} defined in (14). Under Assumptions 4.1 and 4.2,

$$n \cdot E[\hat{V}_n^{\text{F}}] \to V^{\text{obs}} + k \cdot \text{Var}_Q\left[E_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}] - \text{BLP}_Q(\tilde{Y}(1) - \tilde{Y}(0)|1, \tilde{X})\right] =: V^{\text{F}} \leq V^{\text{IM}} ,$$

where BLP *denotes the best linear predictor, i.e., the unique solution under Q such that $E[(\tilde{Y}(1) - \tilde{Y}(0) - \text{BLP}_Q(\tilde{Y}(1) - \tilde{Y}(0)|1, \tilde{X}))(1, \tilde{X})'] = 0$. Furthermore, if (2)–(4) hold, then for $\alpha \in (0,1)$,*

$$P\left\{\hat{\Delta}_n - \sqrt{\hat{V}_n} \cdot z_{1-\alpha/2} \leq \Delta_n \leq \hat{\Delta}_n + \sqrt{\hat{V}_n} \cdot z_{1-\alpha/2}\right\} \to 1 - 2\Phi(z_{\alpha/2}\varsigma)$$
$$P\left\{\hat{\Delta}_n - \sqrt{\hat{V}_n^{\text{IM}}} \cdot z_{1-\alpha/2} \leq \Delta_n \leq \hat{\Delta}_n + \sqrt{\hat{V}_n^{\text{IM}}} \cdot z_{1-\alpha/2}\right\} \to 1 - 2\Phi(z_{\alpha/2}\varsigma^{\text{IM}})$$
$$P\left\{\hat{\Delta}_n - \sqrt{\hat{V}_n^{\text{F}}} \cdot z_{1-\alpha/2} \leq \Delta_n \leq \hat{\Delta}_n + \sqrt{\hat{V}_n^{\text{F}}} \cdot z_{1-\alpha/2}\right\} \to 1 - 2\Phi(z_{\alpha/2}\varsigma^{\text{F}})$$

where $\varsigma = (V/V^{\text{obs}})^{1/2}$, $\varsigma^{\text{IM}} = (V/V^{\text{IM}})^{1/2}$, $\varsigma^{\text{F}} = (V/V^{\text{F}})^{1/2}$, and $\varsigma^{\text{IM}} \leq \varsigma^{\text{F}} \leq \varsigma \leq 1$.

Theorem 4.2 demonstrates that $\hat{V}_n$ converges to our desired upper bound V^{obs}, but that in general $\hat{V}_n^{\text{IM}}$ and $\hat{V}_n^{\text{F}}$ do not. Specifically, $\hat{V}_n^{\text{IM}}$ only attains V^{obs} when treatment effects are sufficiently homogeneous, in the sense that $E_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}]$ is constant; $\hat{V}_n^{\text{F}}$ only attains V^{obs} when the conditional average treatment

effect $E_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}]$ is linear and therefore equal to the best linear predictor of $\tilde{Y}(1) - \tilde{Y}(0)$ given 1 and $\tilde{X}$. As a consequence, Theorem 4.2 demonstrates that, although confidence intervals constructed based on these variance estimators all cover Δ_n with probability at least $1 - \alpha$ in the limit, the confidence interval based on $\hat{V}_n$ is the least conservative whenever Assumptions 4.1 (and 4.2) are satisfied. We further illustrate this phenomenon via simulation in Section 5.

Remark 4.2. Our focus in this paper has been primarily on settings where $\min\{\ell, k - \ell\} = 1$, since, as explained in Section 3.2, this case poses particular analytical challenges. We note however that $\hat{V}_n$ is still well-defined even in the case where $\min\{\ell, k - \ell\} > 1$, and Theorems 3.2 and 4.2 apply equally well to this setting. However, in this case, we note that the estimation strategy proposed in (9) also provides valid inference, and it can be shown under appropriate assumptions that its corresponding probability limit is also V^{obs}. ∎

5 Simulations

In this section we examine the finite-n behavior of several confidence intervals of the form described in (5), constructed using different variance estimators. The variance estimators we consider are our proposed estimator $\hat{V}_n$, the Imai (2008) estimator $\hat{V}_n^{\mathrm{IM}}$, the Fogarty (2018a) estimator $\hat{V}_n^{\mathrm{F}}$, and an alternative variance estimator $\hat{V}_n^{\mathrm{alt}}$ which can also be motivated using Theorem 4.1 in our limiting thought experiment. Importantly, however, $\hat{V}_n^{\mathrm{alt}}$ is *not* necessarily upward-biased under Assumption 2.1 alone. As a consequence, we will demonstrate that the resulting confidence interval will not appropriately cover Δ_n whenever the limiting thought experiment described in Assumption 4.1 fails to hold. To construct $\hat{V}_n^{\mathrm{alt}}$, consider the following decomposition:

$$E_Q[\mathrm{Var}_Q[\tilde{Y}(d)|\tilde{X}]] = E_Q[\tilde{Y}(d)^2] - E_Q[E_Q[\tilde{Y}(d)|\tilde{X}]^2] = \mathrm{Var}_Q[\tilde{Y}(d)] + E_Q[\tilde{Y}(d)]^2 - E_Q[E_Q[\tilde{Y}(d)|\tilde{X}]^2] \ .$$

Under Assumption 4.1(a), straightforward consistent estimators of $\mathrm{Var}_Q[\tilde{Y}(d)]$ and $E_Q[\tilde{Y}(d)]^2$ are given by $\hat{\sigma}_n^2(d)$ and $\hat{\mu}_n(d)^2$, where

$$\hat{\mu}_n(d) = \frac{1}{n(d)} \sum_{1 \le i \le n} Y_i I\{D_i = d\}$$

$$\hat{\sigma}_n^2(d) = \frac{1}{n(d)} \sum_{1 \le i \le n} (Y_i - \hat{\mu}_n(d))^2 I\{D_i = d\} \ .$$

Under Assumption 4.1(b), an estimator of $E_Q[E_Q[\tilde{Y}(d)|\tilde{X}]^2]$ is

$$\hat{\varsigma}_n(d) = \frac{2}{m} \sum_{1 \le j \le \lfloor \frac{m}{2} \rfloor} \sum_{i \in \lambda_{2j}, i' \in \lambda_{2j-1}} Y_i Y_{i'} I\{D_i = D_{i'} = d\} \ .$$

An alternative estimator of V^{obs} under Assumption 4.1 is thus given by

$$\hat{V}_n^{\mathrm{alt}} := \frac{1}{n} \left(\frac{\hat{\sigma}_n^2(1) + \hat{\mu}_n(1)^2 - \hat{\varsigma}_n(1)}{\eta} + \frac{\hat{\sigma}_n^2(0) + \hat{\mu}_n(0)^2 - \hat{\varsigma}_n(0)}{1 - \eta} \right) \ .$$

We remark that, in the case of matched pairs, $\hat{V}_n^{\text{alt}}$ also coincides with the estimator that would be obtained if we applied the estimation strategy in (9) to the paired strata directly; see Remark 3.2 for further discussion.

We generate our population $((Y_i(1), Y_i(0), X_i) : 1 \le i \le n)$ using i.i.d. draws from a probability distribution. The potential outcomes are generated according to the equation:

$$Y_i(d) = \mu_d + \mu_d(X_i) + \epsilon_{di} ,$$

where $((X_i, \epsilon_i) : 1 \le i \le n)$ are i.i.d., X_i and ϵ_i are independent with $X_i \sim U[0, 1]$, $\epsilon_{id} \sim N(0, 1)$, the parameters μ_d are given by $\mu_1 = 0.25$, $\mu_0 = 0$, and $\mu_d(\cdot)$ are specified as

Model 1: $\mu_0(X_i) = 20(X_i - 1/2)$, $\mu_1(X_i) = 10(X_i - 1/2)$

Model 2: $\mu_0(X_i) = 40(X_i^2 - 4/3)$, $\mu_1(X_i) = 10(X_i^2 - 4/3)$.

For a population of size $n = 1000$, the potential outcomes and covariates are generated as i.i.d. draws *once* and then fixed in repeated samples. When considering populations of size $n \in \{100, 250, 500, 750\}$, we draw a subsample from our population of size 1000 *once* and then fix this in repeated samples.

In each Monte Carlo iteration, we assign units to treatment using a matched pairs design (i.e., a finely stratified design with $\ell = 1$ and $k = 2$), under two alternative matching methods, which we call "good match" and "bad match." For "good match," we sort units in increasing values of X_i and then pair adjacent units. For "bad match," we sort units in increasing values of X_i and then pair them such that the unit with the *smallest* value of X is matched with the unit with the *largest* value of X, the second smallest with the second largest, and so on. In both cases, when re-arranging the strata to construct our paired strata $\{(\lambda_{2j-1}, \lambda_{2j}) : 1 \le j \le \lfloor m/2 \rfloor\}$, we sort strata in increasing values of their covariate *averages* and then pair adjacent pairs. As a result, the "good match" design minimizes the average squared distances of the covariate values in paired strata and can be shown to satisfy the conditions of our limiting thought experiment in Assumptions 4.1 and 4.2. In contrast, the "bad match" design is constructed specifically to fail the conditions of the limiting thought experiment.

Tables 1 and 2 report the coverage and average length of 95% confidence intervals computed across 5000 Monte Carlo replications. When matches are good, we see that all estimators have appropriate (albeit conservative) coverage. In Table 1 with good matches, the average length produced by $\hat{V}^{\text{IM}}$ is largest while the average lengths produced by the other three estimators are comparable; this is in line with the theoretical results presented in Theorem 4.2 given that the conditional average treatment effect is linear under Model 1. In Table 2 with good matches, the average lengths produced by $\hat{V}_n$ and $\hat{V}_n^{\text{alt}}$ are shortest, again in line with the theoretical results presented in Theorem 4.2.

When matches are bad, the conditions in Theorem 4.2 no longer apply, and instead we must rely on the asymptotic validity provided by Theorem 3.1 alone. With this in mind we see that in both models, the estimators $\hat{V}_n^{\text{IM}}$, $\hat{V}_n^{\text{F}}$ and $\hat{V}_n$ all provide appropriate coverage; this is in line with our theoretical results given that all three estimators are upward-biased under Assumption 2.1. The average lengths of the confidence intervals based on these estimators are comparable. On the other hand, $\hat{V}_n^{\text{alt}}$, which is not guaranteed to be upward-biased in general, undercovers even in large populations.

Population size	Estimator	Good matches		Bad matches	
		Coverage	Length	Coverage	Length
100	$\hat{V}_n^{\mathrm{IM}}$	1.000	1.963	0.944	4.681
	$\hat{V}_n^{\mathrm{F}}$	0.985	0.842	0.938	4.671
	$\hat{V}_n$	0.976	0.793	0.929	4.634
	$\hat{V}_n^{\mathrm{alt}}$	0.989	0.859	0.844	3.547
250	$\hat{V}_n^{\mathrm{IM}}$	1.000	1.189	0.946	3.018
	$\hat{V}_n^{\mathrm{F}}$	0.989	0.511	0.944	3.016
	$\hat{V}_n$	0.992	0.534	0.942	3.005
	$\hat{V}_n^{\mathrm{alt}}$	0.998	0.618	0.863	2.281
500	$\hat{V}_n^{\mathrm{IM}}$	1.000	0.825	0.952	2.155
	$\hat{V}_n^{\mathrm{F}}$	0.993	0.368	0.952	2.155
	$\hat{V}_n$	0.992	0.369	0.951	2.153
	$\hat{V}_n^{\mathrm{alt}}$	0.992	0.361	0.862	1.628
750	$\hat{V}_n^{\mathrm{IM}}$	1.000	0.651	0.950	1.745
	$\hat{V}_n^{\mathrm{F}}$	0.994	0.293	0.950	1.745
	$\hat{V}_n$	0.994	0.294	0.948	1.743
	$\hat{V}_n^{\mathrm{alt}}$	0.997	0.312	0.858	1.316
1000	$\hat{V}_n^{\mathrm{IM}}$	1.000	0.565	0.949	1.518
	$\hat{V}_n^{\mathrm{F}}$	0.995	0.247	0.949	1.518
	$\hat{V}_n$	0.995	0.245	0.950	1.516
	$\hat{V}_n^{\mathrm{alt}}$	0.996	0.246	0.857	1.143

Table 1: Coverage probabilities and average length (Model 1)

Population size	Estimator	Good matches		Bad matches	
		Coverage	Length	Coverage	Length
100	$\hat{V}_n^{\mathrm{IM}}$	1.000	5.288	0.944	8.185
	$\hat{V}_n^{\mathrm{F}}$	1.000	1.389	0.939	8.173
	$\hat{V}_n$	0.980	0.848	0.930	8.102
	$\hat{V}_n^{\mathrm{alt}}$	0.990	0.910	0.880	6.813
250	$\hat{V}_n^{\mathrm{IM}}$	1.000	3.274	0.947	5.174
	$\hat{V}_n^{\mathrm{F}}$	1.000	0.909	0.944	5.153
	$\hat{V}_n$	1.000	0.812	0.940	5.133
	$\hat{V}_n^{\mathrm{alt}}$	1.000	0.992	0.892	4.292
500	$\hat{V}_n^{\mathrm{IM}}$	1.000	2.256	0.952	3.587
	$\hat{V}_n^{\mathrm{F}}$	1.000	0.675	0.952	3.580
	$\hat{V}_n$	0.992	0.370	0.951	3.567
	$\hat{V}_n^{\mathrm{alt}}$	0.992	0.363	0.905	3.003
750	$\hat{V}_n^{\mathrm{IM}}$	1.000	1.816	0.951	2.919
	$\hat{V}_n^{\mathrm{F}}$	1.000	0.548	0.951	2.918
	$\hat{V}_n$	1.000	0.359	0.950	2.913
	$\hat{V}_n^{\mathrm{alt}}$	1.000	0.406	0.908	2.459
1000	$\hat{V}_n^{\mathrm{IM}}$	1.000	1.588	0.952	2.552
	$\hat{V}_n^{\mathrm{F}}$	1.000	0.462	0.952	2.545
	$\hat{V}_n$	0.995	0.245	0.950	2.537
	$\hat{V}_n^{\mathrm{alt}}$	0.996	0.246	0.898	2.130

Table 2: Coverage probabilities and average length (Model 2)

A Proofs of Main Results

A.1 Proof of Theorem 3.1

First, we show that Assumptions (2) and (3) imply the conditions in Theorem A.1, so that

$$\frac{\hat{\Delta}_n - \Delta_n}{\text{Var}[\hat{\Delta}_n]^{1/2}} \xrightarrow{d} N(0,1) . \tag{16}$$

To that end, we first show

$$\frac{1}{n} \max_{1 \le j \le m} \max_{i \in \lambda_j} (R_i - \bar{R}_j)^2 \to 0 , \tag{17}$$

where R_i, $\bar{R}_j$ are defined in the statement of Theorem A.1, and then show that $\frac{1}{n} \sum_{1 \le j \le m} \sum_{i \in \lambda_j} (R_i - \bar{R}_j)^2$ is bounded away from zero. Note that from the inequality $(a + b)^2 \le 2(a^2 + b^2)$,

$$\max_{i \in \lambda_j} (R_i - \bar{R}_j)^2 \lesssim \max_{i \in \lambda_j} \frac{1}{\eta^2} (Y_i(1) - \bar{Y}_{j,n}(1))^2 + \frac{1}{(1-\eta)^2} \max_{i \in \lambda_j} (Y_i(0) - \bar{Y}_{j,n}(0))^2 .$$

Since by the triangle inequality $\max_{i \in \lambda_j} |Y_i(d) - \bar{Y}_{j,n}(d)| \le 2 \max_{i \in \lambda_j} |Y_i(d)|$,

$$\frac{1}{n} \max_{1 \le j \le m} \max_{i \in \lambda_j} (Y_i(d) - \bar{Y}_{j,n}(d))^2 \lesssim \frac{1}{n} \max_{1 \le i \le n} Y_i(d)^2 \to 0 ,$$

where the final convergence follows from (2). (17) therefore follows. Next, since it is shown in the proof of Theorem A.1 that $\frac{1}{n} \sum_{1 \le j \le m} \sum_{i \in \lambda_j} (R_i - \bar{R}_j)^2$ is proportional to $n \cdot \text{Var}[\hat{\Delta}_n]$ we immediately obtain from Assumption (3) that $\frac{1}{n} \sum_{1 \le j \le m} \sum_{i \in \lambda_j} (R_i - \bar{R}_j)^2$ is bounded away from zero. We thus obtain (33) and so (16) follows. To complete the proof, we apply Lemma A.1 with $X_n = \sqrt{n}(\hat{\Delta}_n - \Delta_n)$, $\sigma_n^2 = n \text{Var}[\hat{\Delta}_n]$, and $\hat{\sigma}_n^2 = n\hat{V}_n$. ∎

A.2 Proof of Theorem 3.2

First we establish (a). By direct calculation using Assumption 2.1 and Theorem 2.1 in Cochran (1977), we have

$$
\begin{aligned}
&E[\hat{\tau}_n^2 - \hat{\kappa}_n] \\
&= \frac{1}{n\ell} \sum_{1 \le i \le n} Y_i(1)^2 + \frac{2(\ell - 1)}{n(k-1)\ell} \sum_{1 \le j \le m} \sum_{i < i' \in \lambda_j} Y_i(1)Y_{i'}(1) \\
&\quad + \frac{1}{n(k-\ell)} \sum_{1 \le i \le n} Y_i(0)^2 + \frac{2(k-\ell-1)}{n(k-1)(k-\ell)} \sum_{1 \le j \le m} \sum_{i < i' \in \lambda_j} Y_i(0)Y_{i'}(0) \\
&\quad - \frac{2}{n(k-1)} \sum_{1 \le j \le m} \sum_{i \ne i' \in \lambda_j} Y_i(1)Y_{i'}(0) \\
&\quad - \frac{2}{m} \sum_{1 \le j \le \frac{m}{2}} (\bar{Y}_{2j-1,n}(1) - \bar{Y}_{2j-1,n}(0))(\bar{Y}_{2j,n}(1) - \bar{Y}_{2j,n}(0)) .
\end{aligned}
$$

On the other hand,

$$
m \cdot \operatorname{Var}[\hat{\Delta}_n] = \frac{1}{n} \sum_{1 \le j \le m} \left(\frac{k}{(k-1)\ell} \sum_{i \in \lambda_j} (Y_i(1) - \bar{Y}_{j,n}(1))^2 + \frac{k}{(k-1)(k-\ell)} \sum_{i \in \lambda_j} (Y_i(0) - \bar{Y}_{j,n}(0))^2 \right.
$$
$$
\left. - \frac{1}{k-1} \sum_{i \in \lambda_j} (Y_i(1) - Y_i(0) - (\bar{Y}_{j,n}(1) - \bar{Y}_{j,n}(0)))^2 \right)
$$
$$
= \frac{1}{n} \sum_{1 \le j \le m} \left(\frac{1}{\ell} \sum_{i \in \lambda_j} Y_i(1)^2 - \frac{2}{(k-1)\ell} \sum_{i < i' \in \lambda_j} Y_i(1)Y_{i'}(1) \right.
$$
$$
+ \frac{1}{k-\ell} \sum_{i \in \lambda_j} Y_i(0)^2 - \frac{2}{(k-1)(k-\ell)} \sum_{i < i' \in \lambda_j} Y_i(0)Y_{i'}(0)
$$
$$
\left. - \frac{1}{k-1} \left(\sum_{i \in \lambda_j} (Y_i(1) - Y_i(0))^2 - k(\bar{Y}_{j,n}(1) - \bar{Y}_{j,n}(0))^2 \right) \right) .
$$

We therefore have

$$
E[\hat{\tau}_n^2 - \hat{\kappa}_n] - m \cdot \operatorname{Var}[\hat{\Delta}_n]
$$
$$
= \frac{2}{n(k-1)} \sum_{1 \le j \le m} \sum_{i < i' \in \lambda_j} Y_i(1)Y_{i'}(1) + \frac{2}{n(k-1)} \sum_{1 \le j \le m} \sum_{i < i' \in \lambda_j} Y_i(0)Y_{i'}(0)
$$
$$
- \sum_{1 \le j \le m} \frac{2}{n(k-1)} \sum_{i \ne i' \in \lambda_j} Y_i(1)Y_{i'}(0)
$$
$$
+ \frac{1}{n(k-1)} \sum_{1 \le j \le m} \left(\sum_{i \in \lambda_j} (Y_i(1) - Y_i(0))^2 - k(\bar{Y}_{j,n}(1) - \bar{Y}_{j,n}(0))^2 \right)
$$
$$
- \frac{2}{m} \sum_{1 \le j \le \frac{m}{2}} (\bar{Y}_{2j-1,n}(1) - \bar{Y}_{2j-1,n}(0))(\bar{Y}_{2j,n}(1) - \bar{Y}_{2j,n}(0))
$$
$$
= \frac{1}{n(k-1)} \sum_{1 \le j \le m} \left(k^2 \bar{Y}_{j,n}(1)^2 - \sum_{i \in \lambda_j} Y_i(1)^2 \right) + \frac{1}{n(k-1)} \sum_{1 \le j \le m} \left(k^2 \bar{Y}_{j,n}(0)^2 - \sum_{i \in \lambda_j} Y_i(0)^2 \right)
$$
$$
- \frac{2}{n(k-1)} \sum_{1 \le j \le m} \left(k^2 \bar{Y}_{j,n}(1)\bar{Y}_{j,n}(0) - \sum_{i \in \lambda_j} Y_i(1)Y_i(0) \right)
$$
$$
+ \frac{1}{n(k-1)} \sum_{1 \le j \le m} \left(\sum_{i \in \lambda_j} (Y_i(1) - Y_i(0))^2 - k(\bar{Y}_{j,n}(1) - \bar{Y}_{j,n}(0))^2 \right)
$$
$$
- \frac{2}{m} \sum_{1 \le j \le \frac{m}{2}} (\bar{Y}_{2j-1,n}(1) - \bar{Y}_{2j-1,n}(0))(\bar{Y}_{2j,n}(1) - \bar{Y}_{2j,n}(0))
$$
$$
= \frac{1}{n(k-1)} \sum_{1 \le i \le n} (-Y_i(1)^2 - Y_i(0)^2 + 2Y_i(1)Y_i(0) + Y_i(1)^2 + Y_i(0)^2 - 2Y_i(1)Y_i(0))
$$
$$
+ \frac{1}{n(k-1)} \sum_{1 \le j \le m} (k^2 \bar{Y}_{j,n}(1)^2 + k^2 \bar{Y}_{j,n}(0)^2 - 2k^2 \bar{Y}_{j,n}(1)\bar{Y}_{j,n}(0)
$$
$$
- k\bar{Y}_{j,n}(1)^2 - k\bar{Y}_{j,n}(0)^2 + 2k\bar{Y}_{j,n}(1)\bar{Y}_{j,n}(0))
$$
$$
- \frac{2}{m} \sum_{1 \le j \le \frac{m}{2}} (\bar{Y}_{2j-1,n}(1) - \bar{Y}_{2j-1,n}(0))(\bar{Y}_{2j,n}(1) - \bar{Y}_{2j,n}(0))
$$
$$
= \frac{1}{m} \sum_{1 \le j \le m} (\bar{Y}_{j,n}(1) - \bar{Y}_{j,n}(0))^2 - \frac{2}{m} \sum_{1 \le j \le \frac{m}{2}} (\bar{Y}_{2j-1,n}(1) - \bar{Y}_{2j-1,n}(0))(\bar{Y}_{2j,n}(1) - \bar{Y}_{2j,n}(0))
$$

$$= \frac{1}{m} \sum_{1 \leq j \leq \frac{m}{2}} (\bar{Y}_{2j-1,n}(1) - \bar{Y}_{2j-1,n}(0) - (\bar{Y}_{2j,n}(1) - \bar{Y}_{2j,n}(0)))^2 \ .$$

Next, we establish (b). We'll show that

$$|\hat{\tau}_n^2 - E[\hat{\tau}_n^2]| \xrightarrow{P} 0$$

$$|\hat{\kappa}_n - E[\hat{\kappa}_n]| \xrightarrow{P} 0 \ ,$$

from which the result will follow from the triangle inequality. By Chebyshev's inequality, it suffices to show that $\text{Var}[\hat{\tau}_n^2] \to 0$ and $\text{Var}[\hat{\kappa}_n] \to 0$. To that end, define

$$\hat{\Delta}_j = \frac{1}{\ell} \sum_{i \in \lambda_j} Y_i D_i - \frac{1}{k - \ell} \sum_{i \in \lambda_j} Y_i (1 - D_i) \ .$$

Then,

$$\text{Var}[\hat{\tau}_n^2] = \frac{1}{m^2} \sum_{1 \leq j \leq m} \text{Var}[\hat{\Delta}_{j,n}^2] \tag{18}$$

$$\text{Var}[\hat{\kappa}_n] = \frac{4}{m^2} \sum_{1 \leq j \leq \frac{m}{2}} \text{Var}[\hat{\Delta}_{2j-1,n} \hat{\Delta}_{2j,n}] \ . \tag{19}$$

First, we bound (18):

$$\frac{1}{m^2} \sum_{1 \leq j \leq m} \text{Var}[\hat{\Delta}_j^2] \leq \frac{1}{m^2} \sum_{1 \leq j \leq m} E[\hat{\Delta}_j^4]$$

$$\lesssim \frac{1}{m^2} \sum_{1 \leq j \leq m} \left(E\left[\left(\frac{1}{\ell} \sum_{i \in \lambda_j} Y_i D_i \right)^4 \right] + E\left[\left(\frac{1}{k - \ell} \sum_{i \in \lambda_j} Y_i (1 - D_i) \right)^4 \right] \right)$$

$$\leq \frac{1}{m^2} \sum_{1 \leq j \leq m} \left(E\left[\frac{1}{\ell} \sum_{i \in \lambda_j} Y_i^4 D_i \right] + E\left[\frac{1}{k - \ell} \sum_{i \in \lambda_j} Y_i^4 (1 - D_i) \right] \right)$$

$$= \frac{1}{m^2} \sum_{1 \leq j \leq m} \left(\frac{1}{k} \sum_{i \in \lambda_j} Y_i(1)^4 + \frac{1}{k} \sum_{i \in \lambda_j} Y_i(0)^4 \right)$$

$$= \frac{1}{m} \left(\frac{1}{n} \sum_{1 \leq i \leq n} Y_i(1)^4 + \frac{1}{n} \sum_{1 \leq i \leq n} Y_i(0)^4 \right) \to 0 \ ,$$

where the first inequality follows from the definition of the variance, the second from repeated applications of the inequality $(a \pm b)^2 \leq 2(a^2 + b^2)$, the third from Jensen's inequality, the first equality from Assumption 2.1 and Theorem 2.1 in Cochran (1977), which together imply $E[D_i] = \ell/k$, and the final line by (2).

Next, we bound (19):

$$\frac{4}{m^2} \sum_{1 \leq j \leq \frac{m}{2}} \text{Var}[\hat{\Delta}_{2j-1,n} \hat{\Delta}_{2j,n}] \leq \frac{2}{m^2} \sum_{1 \leq j \leq \frac{m}{2}} E[2\hat{\Delta}_{2j-1,n}^2 \hat{\Delta}_{2j,n}^2]$$

$$\leq \frac{2}{m^2} \sum_{1 \leq j \leq \frac{m}{2}} E[\hat{\Delta}_{2j-1,n}^4 + \hat{\Delta}_{2j,n}^4]$$

$$\lesssim \frac{2}{m^2} \sum_{1 \leq j \leq \frac{m}{2}} \left(\frac{1}{k} \sum_{i \in \lambda_{2j-1}} \sum_{d \in \{0,1\}} Y_i(d)^4 + \frac{1}{k} \sum_{i \in \lambda_{2j}} \sum_{d \in \{0,1\}} Y_i(d)^4 \right)$$

$$= \frac{2}{m} \left(\frac{1}{n} \sum_{1 \leq i \leq n} Y_i(1)^4 + \frac{1}{n} \sum_{1 \leq i \leq n} Y_i(0)^4 \right) \to 0 \;.$$

where the first inequality follows from the definition of the variance, the second from the inequality $2|ab| \leq (a^2+b^2)$, the third from following the same argument as when bounding (18), and the final line by Assumption (2). ∎

A.3   Proof of Theorem 4.1

From the proof of Theorem A.1, we obtain that

$$n \cdot \mathrm{Var}[\hat{\Delta}_n] = k \sum_{1 \leq j \leq m} \mathrm{Var}[A_j] = k\eta^2 \left(\frac{1}{\ell} - \frac{1}{k} \right) \frac{1}{k-1} \frac{1}{m} \sum_{1 \leq j \leq m} \sum_{i \in \lambda_j} (R_i - \bar{R}_j)^2$$

Consider the following expansion of $\frac{1}{m} \sum_{1 \leq j \leq m} \sum_{i \in \lambda_j} \bar{R}_j^2$:

$$\frac{1}{m} \sum_{1 \leq j \leq n} \sum_{i \in \lambda_j} \bar{R}_j^2 = \frac{k}{m} \sum_{1 \leq j \leq m} \bar{R}_j^2 = \frac{k}{mk^2} \sum_{1 \leq j \leq m} \sum_{i,i' \in \lambda_j} R_i R_{i'}$$

$$= \frac{1}{km} \sum_{1 \leq i \leq km} R_i^2 + \frac{2}{km} \sum_{1 \leq j \leq m} \sum_{i < i' \in \lambda_j} R_i R_{i'}$$

$$= \frac{1}{n} \sum_{1 \leq i \leq n} R_i^2 + \frac{k-1}{m} \sum_{1 \leq j \leq m} \frac{1}{\binom{k}{2}} \sum_{i < i' \in \lambda_j} R_i R_{i'} \;.$$

We now show that

$$\frac{1}{n} \sum_{1 \leq i \leq n} R_i^2 \to E_Q[\tilde{R}^2] \tag{20}$$

$$\frac{1}{m} \sum_{1 \leq j \leq m} \frac{1}{\binom{k}{2}} \sum_{i < i' \in \lambda_j} R_i R_{i'} \to E_Q[E_Q[\tilde{R}|\tilde{X}]^2] \;, \tag{21}$$

where $\tilde{R} := \frac{\tilde{Y}(1)}{\eta} + \frac{\tilde{Y}(0)}{1-\eta}$. Note (20) follows from Assumption 4.1(a) immediately. To see how (21) is implied by Assumption 4.1(b), for each j, let $i_{j,1} < \cdots < i_{j,k}$ be such that $\lambda_j = \{i_{j,1}, \ldots, i_{j,k}\}$, and note Assumption 4.1(b) implies that for each $1 \leq \gamma_1 < \gamma_2 \leq k$,

$$\frac{1}{m} \sum_{1 \leq j \leq m} R_{i_{j,\gamma_1}} R_{i_{j,\gamma_2}} \to E_Q[E_Q[\tilde{R}|\tilde{X}]^2] \;.$$

The result in (21) then follows upon noting

$$\frac{1}{m} \sum_{1 \leq j \leq m} \frac{1}{\binom{k}{2}} \sum_{i < i' \in \lambda_j} R_i R_{i'} = \frac{1}{\binom{k}{2}} \sum_{1 \leq \gamma_1 < \gamma_2 < k} \frac{1}{m} \sum_{1 \leq j \leq m} R_{i_{j,\gamma_1}} R_{i_{j,\gamma_2}} .$$

We thus obtain that

$$\frac{1}{m} \sum_{1 \leq j \leq m} \sum_{i \in \lambda_j} (R_i - \bar{R}_j)^2 = \frac{1}{m} \sum_{1 \leq j \leq m} \left(\sum_{i \in \lambda_j} R_i^2 - k \bar{R}_j^2 \right) \rightarrow (k-1) E_Q[\text{Var}_Q[\tilde{R}|\tilde{X}]] .$$

Finally,

$$n \cdot \text{Var}[\hat{\Delta}_n] = k \sum_{1 \leq j \leq m} \text{Var}[A_j] \rightarrow k \eta^2 \left(\frac{1}{\ell} - \frac{1}{k} \right) E_Q[\text{Var}_Q[\tilde{R}|\tilde{X}]]$$

$$= \eta(1-\eta) E_Q[\text{Var}_Q[\tilde{R}|\tilde{X}]] = \eta(1-\eta) E_Q\left[\text{Var}_Q\left[\frac{\tilde{Y}(1)}{\eta} + \frac{\tilde{Y}(0)}{1-\eta} \Bigg| \tilde{X} \right] \right]$$

$$= E_Q\left[\frac{\text{Var}_Q[\tilde{Y}(1)|\tilde{X}]}{\eta} + \frac{\text{Var}_Q[\tilde{Y}(0)|\tilde{X}]}{1-\eta} \right] - E_Q[\text{Var}_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}]] ,$$

as desired. ∎

A.4 Proof of Theorem 4.2

(a) We start by deriving the limit of $nE[\hat{V}_n]$. Recall from the proof of Theorem 3.2 that

$$E[\hat{\tau}_n^2] = \frac{1}{n\ell} \sum_{1 \leq i \leq n} Y_i(1)^2 + \frac{2(\ell-1)}{n(k-1)\ell} \sum_{1 \leq j \leq m} \sum_{i < i' \in \lambda_j} Y_i(1) Y_{i'}(1)$$

$$+ \frac{1}{n(k-\ell)} \sum_{1 \leq i \leq n} Y_i(0)^2 + \frac{2(k-\ell-1)}{n(k-1)(k-\ell)} \sum_{1 \leq j \leq m} \sum_{i < i' \in \lambda_j} Y_i(0) Y_{i'}(0)$$

$$- \frac{2}{n(k-1)} \sum_{1 \leq j \leq m} \sum_{i \neq i' \in \lambda_j} Y_i(1) Y_{i'}(0)$$

$$= \frac{1}{m} \sum_{1 \leq j \leq m} \left(\frac{1}{\ell} \frac{1}{k} \sum_{i \in \lambda_j} Y_i^2(1) + \frac{\ell-1}{\ell} \frac{1}{\binom{k}{2}} \sum_{i < i' \in \lambda_j} Y_i(1) Y_{i'}(1) \right)$$

$$+ \frac{1}{m} \sum_{1 \leq j \leq m} \left(\frac{1}{k-\ell} \frac{1}{k} \sum_{i \in \lambda_j} Y_i^2(0) + \frac{k-\ell-1}{k-\ell} \frac{1}{\binom{k}{2}} \sum_{i < i' \in \lambda_j} Y_i(0) Y_{i'}(0) \right)$$

$$- \frac{1}{m} \sum_{1 \leq j \leq m} \frac{1}{\binom{k}{2}} \sum_{i \neq i' \in \lambda_j} Y_i(1) Y_{i'}(0) .$$

Assumption 4.1 implies that

$$\frac{1}{n} \sum_{1 \leq i \leq n} Y_i(d)^2 \rightarrow E_Q[\tilde{Y}^2(d)] \text{ for } d \in \{0,1\} \tag{22}$$

$$\frac{1}{m} \sum_{1 \leq j \leq m} \frac{1}{\binom{k}{2}} \sum_{i < i' \in \lambda_j} Y_i(d) Y_{i'}(d) \rightarrow E_Q[\tilde{E}_Q[\tilde{Y}(d)|\tilde{X}]^2] \text{ for } d \in \{0,1\} \tag{23}$$

19

$$\frac{1}{m} \sum_{1 \le j \le m} \frac{1}{\binom{k}{2}} \sum_{i \ne i' \in \lambda_j} Y_i(1) Y_{i'}(0) \to 2 E_Q[E_Q[\tilde{Y}(1)|\tilde{X}] E_Q[\tilde{Y}(0)|\tilde{X}]] \ , \tag{24}$$

where (22) follows from Assumption 4.1(a) and (23)–(24) follow from Assumption 4.1(b). To see how (23) is implied by Assumption 4.1(b), for each j, let $i_{j,1} < \cdots < i_{j,k}$ be such that $\lambda_j = \{i_{j,1}, \ldots, i_{j,k}\}$, and note Assumption 4.1(b) implies that for each $1 \le \gamma_1 < \gamma_2 \le k$,

$$\frac{1}{m} \sum_{1 \le j \le m} Y_{i_{j,\gamma_1}} Y_{i_{j,\gamma_2}} \to E_Q[E_Q[\tilde{Y}(d)|\tilde{X}]^2] \ .$$

(23) then follows by noting

$$\frac{1}{m} \sum_{1 \le j \le m} \frac{1}{\binom{k}{2}} \sum_{i < i' \in \lambda_j} Y_i(d) Y_{i'}(d) = \frac{1}{\binom{k}{2}} \sum_{1 \le \gamma_1 < \gamma_2 \le k} \frac{1}{m} \sum_{1 \le j \le m} Y_{i_{j,\gamma_1}} Y_{i_{j,\gamma_2}} \ .$$

The result in (24) can be established similarly. Therefore, noting $n = mk$, we get

$$E[\hat{\tau}_n^2] \to \frac{1}{\ell} E_Q[\mathrm{Var}_Q[\tilde{Y}(1)|\tilde{X}]] + \frac{1}{k-\ell} E_Q[\mathrm{Var}_Q[\tilde{Y}(0)|\tilde{X}]] + E_Q[E_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}]^2] \ . \tag{25}$$

Next, recall from the proof of Theorem 3.2 that

$$E[\hat{\kappa}_n] = \frac{2}{m} \sum_{1 \le j \le \frac{m}{2}} (\bar{Y}_{2j-1,n}(1) - \bar{Y}_{2j-1,n}(0))(\bar{Y}_{2j,n}(1) - \bar{Y}_{2j,n}(0)) \ .$$

=It follows from Assumption 4.1(b) applied across strata and similar arguments to those used to prove (23)–(24) that

$$E[\hat{\kappa}_n] \to E_Q[E_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}]^2] \ .$$

As a result,

$$nE[\hat{V}_n] = k(E[\hat{\tau}_n^2] - E[\hat{\kappa}_n]) \to E_Q \left[\frac{\mathrm{Var}_Q[\tilde{Y}(1)|\tilde{X}]}{\eta} + \frac{\mathrm{Var}_Q[\tilde{Y}(0)|\tilde{X}]}{1 - \eta} \right] \ .$$

(b) For $nE[\hat{V}_n^{\mathrm{IM}}]$, first observe that because $\hat{\Delta}_n = \frac{1}{m} \sum_{1 \le j \le m} \hat{\Delta}_{j,n}$,

$$\begin{aligned} n\hat{V}_n^{\mathrm{IM}} &= k \frac{m}{m-1}(m-1)\hat{V}_n^{\mathrm{IM}} \\ &= k \frac{m}{m-1} \left(\frac{1}{m} \sum_{j=1}^m \hat{\Delta}_{j,n}^2 - \hat{\Delta}_n^2 \right) \\ &= k\hat{\tau}_n^2 - k \frac{m}{m-1} \frac{2}{m^2} \sum_{1 \le j < j' \le m} \hat{\Delta}_{j,n} \hat{\Delta}_{j',n} \ . \end{aligned} \tag{26}$$

It therefore suffices to derive the limit of

$$E \left[\frac{2}{m^2} \sum_{1 \le j < j' \le m} \hat{\Delta}_{j,n} \hat{\Delta}_{j',n} \right]$$

20

and combine with the limit of $E[\hat{\tau}_n^2]$ in (25). To that end, note

$$E\left[\frac{2}{m^2}\sum_{1\leq j<j'\leq m}\hat{\Delta}_{j,n}\hat{\Delta}_{j',n}\right]$$

$$=\frac{2}{m^2}\sum_{1\leq j<j'\leq m}\Delta_{j,n}\Delta_{j',n}$$

$$=\left(\frac{1}{m}\sum_{1\leq j\leq m}\Delta_{j,n}\right)^2-\frac{1}{m^2}\sum_{1\leq j\leq n}\Delta_{j,n}^2$$

$$=\Delta_n^2-\frac{1}{m^2}\sum_{1\leq j\leq n}\frac{1}{k^2}\left(\sum_{i\in\lambda_j}(Y_i(1)-Y_i(0))^2+2\sum_{i<i'\in\lambda_j}(Y_i(1)-Y_i(0))(Y_{i'}(1)-Y_{i'}(0))\right)$$

$$\rightarrow E_Q[\tilde{Y}(1)-\tilde{Y}(0)]^2\ , \tag{27}$$

where the first equality follows because $\hat{\Delta}_{j,n}$ and $\hat{\Delta}_{j',n}$ are independent and $E[\hat{\Delta}_{j,n}]=\Delta_{j,n}$, the second follows by completing the square, the third follows by expanding the square, and the convergence follows because $\Delta_n\rightarrow E_Q[\tilde{Y}(1)-\tilde{Y}(0)]$, $\frac{1}{m}\rightarrow 0$, and

$$\frac{1}{n}\sum_{1\leq i\leq n}(Y_i(1)-Y_i(0))^2\rightarrow E_Q[(\tilde{Y}(1)-\tilde{Y}(0))^2]$$

$$\frac{1}{m}\sum_{1\leq j\leq n}\frac{1}{\binom{k}{2}}\sum_{i<i'\in\lambda_j}(Y_i(1)-Y_i(0))(Y_{i'}(1)-Y_{i'}(0))\rightarrow E_Q[E_Q[\tilde{Y}(1)-\tilde{Y}(0)|\tilde{X}]^2]$$

by similar arguments to those used to prove (22)–(23). Combining (25), (26), and (27), noting $\frac{m}{m-1}\rightarrow 1$, we have

$$n\hat{E}[V_n^{\mathrm{IM}}]\rightarrow\frac{1}{\eta}E_Q[\mathrm{Var}_Q[\tilde{Y}(1)|\tilde{X}]]+\frac{1}{1-\eta}E_Q[\mathrm{Var}_Q[\tilde{Y}(0)|\tilde{X}]]$$

$$+kE_Q[E_Q[\tilde{Y}(1)-\tilde{Y}(0)|\tilde{X}]^2]-kE_Q[\tilde{Y}(1)-\tilde{Y}(0)]^2$$

$$=V^{\mathrm{obs}}+k\,\mathrm{Var}_Q[E_Q[\tilde{Y}(1)-\tilde{Y}(0)|X]]\ ,$$

as desired.

(c) Finally, we study the limit of $nE[V_n^{\mathrm{F}}]$. We first show that the diagonal entries of the projection matrix H_{Q_2} converge to 0. Let $\Sigma_{X,n}=\frac{1}{m}\sum_{1\leq j\leq m}(\bar{X}_{j,n}-\mu_{X,n})(\bar{X}_{j,n}-\mu_{X,n})'$, and we have

$$H_{Q_2}=Q_2(Q_2'Q_2)^{-1}Q_2'$$

$$=\begin{pmatrix}\iota_m & \begin{array}{c}(\bar{X}_{1,n}-\mu_{X,n})'\\ \vdots\\ (\bar{X}_{m,n}-\mu_{X,n})'\end{array}\end{pmatrix}\begin{pmatrix}m & 0_p'\\ 0_p & m\Sigma_{X,n}\end{pmatrix}^{-1}\begin{pmatrix}\iota_m'\\ \bar{X}_{1,n}-\mu_{X,n} & \cdots & \bar{X}_{m,n}-\mu_{X,n}\end{pmatrix}$$

$$=\frac{1}{m}\iota_m\iota_m'+\frac{1}{m}\begin{pmatrix}(\bar{X}_{1,n}-\mu_{X,n})'\\ \vdots\\ (\bar{X}_{m,n}-\mu_{X,n})'\end{pmatrix}\Sigma_{X,n}^{-1}\begin{pmatrix}\bar{X}_{1,n}-\mu_{X,n} & \cdots & \bar{X}_{m,n}-\mu_{X,n}\end{pmatrix}\ .$$

Therefore, denoting the operator norm by $\|\cdot\|_{\mathrm{op}}$, we have that for $1\leq j\leq m$, the j-th diagonal entry of

21

H_{Q_2} may be uniformly bounded as

$$
\begin{aligned}
\max_{1 \le j \le m}[H_{Q_2}]_{jj} &= \max_{1 \le j \le m}\left(\frac{1}{m} + \frac{1}{m}(\bar{X}_{j,n} - \mu_{X,n})'\Sigma_{X,n}^{-1}(\bar{X}_{j,n} - \mu_{X,n})\right) \\
&\le \frac{1}{m} + \frac{1}{m}\left\|\Sigma_{X,n}^{-1}\right\|_{\mathrm{op}} \max_{1 \le j \le m}\|\bar{X}_{j,n} - \mu_{X,n}\|^2 \\
&\lesssim \frac{1}{m} + k\left\|\Sigma_{X,n}^{-1}\right\|_{\mathrm{op}} \frac{1}{n}\max_{1 \le i \le n}\|X_i\|^2 \to 0,
\end{aligned}
$$

where the first inequality follows from the definition of the operator norm, the second inequality follows from repeated application of the inequality that $(a - b)^2 \le 2(a^2 + b^2)$, and the convergence follows from Assumption 4.2(a) and that $\Sigma_{X,n} \to \mathrm{Var}_Q[\tilde{X}]$, which further follows from Assumption 4.2(b)–(c) by similar arguments to those used to prove (22)–(23).

Define

$$
Q_{2,X} = \begin{pmatrix} (\bar{X}_{1,n} - \mu_{X,n})' \\ \vdots \\ (\bar{X}_{m,n} - \mu_{X,n})' \end{pmatrix}.
$$

Let H_{ι_m} denote the projection matrix for ι_m and $H_{Q_{2,X}}$ denote the projection matrix for $Q_{2,X}$. Note that by construction $\iota_m' Q_{2,X} = 0$. It follows from Lemma A.2 that

$$
\begin{aligned}
nE[\hat{V}_n^{\mathrm{F}}] &= \frac{k}{m}E[\hat{\delta}_n'(\mathrm{diag}(I - H_{Q_2}))^{-1/2}(I - H_{Q_2})(\mathrm{diag}(I - H_{Q_2}))^{-1/2}\hat{\delta}_n] \\
&= \frac{k}{m}\delta_n'(\mathrm{diag}(I - H_{Q_2}))^{-1/2}(I - H_{Q_2})(\mathrm{diag}(I - H_{Q_2}))^{-1/2}\delta_n \quad (28) \\
&\quad + \frac{k}{m}\mathrm{tr}((I - H_{Q_2})\mathrm{Var}[\mathrm{diag}(I - H_{Q_2}))^{-1/2}\hat{\delta}_n]) \quad (29)
\end{aligned}
$$

Define $\epsilon_n = (\mathrm{diag}(I - H_{Q_2}))^{-1/2}\delta_n - \delta_n$. Note (28) equals

$$
\frac{k}{m}\delta_n'(I - H_{Q_2})\delta_n + \frac{k}{m}\epsilon_n'(I - H_{Q_2})\delta_n + \frac{k}{m}\delta_n'(I - H_{Q_2})\epsilon_n + \frac{k}{m}\epsilon_n'(I - H_{Q_2})\epsilon_n \;. \quad (30)
$$

Note $\frac{1}{m}\|\delta_n\|^2$ converges to a constant by the derivations in the proof for the limit of $nE[\hat{V}_n^{\mathrm{IM}}]$. Because $\max_{1 \le j \le m}[H_{Q_2}]_{jj} \to 0$ as $n \to \infty$,

$$
\frac{1}{m}\|\epsilon_n\|^2 \le \frac{1}{m}\|\delta_n\|^2 \max_{1 \le j \le m}\left(\frac{1}{\sqrt{1 - [H_{Q_2}]_{jj}}} - 1\right)^2 \to 0 \;.
$$

Because $I - H_{Q_2}$ is a projection matrix, all of its eigenvalues are either 0 or 1, and hence its operator norm is 1. Therefore,

$$
\frac{k}{m}\epsilon_n'(I - H_{Q_2})\delta_n \le k\left(\frac{1}{m}\|\epsilon_n\|^2\right)^{1/2}\|I - H_{Q_2}\|_{\mathrm{op}}\left(\frac{1}{m}\|\delta_n\|\right)^{1/2} \to 0 \;.
$$

Similarly, the last two terms in (30) also converge to 0 as $n \to \infty$. Recalling $\iota_m' Q_{2,X} = 0$, we get that (28)

22

equals

$$\frac{k}{m}\delta_n'(I - H_{Q_2})\delta_n = \frac{k}{m}\delta_n'(I - H_{\iota_m} - H_{Q_{2,X}})\delta_n = \frac{k}{m}\delta_n'(I - H_{\iota_m})\delta_n - \frac{k}{m}\delta_n' Q_{2,X} \Sigma_{X,n}^{-1} \frac{1}{m} Q_{2,X}' \delta_n . \quad (31)$$

Next, because $\hat{\Delta}_{j,n}$ are independent across $1 \leq j \leq m$, note (29) equals

$$\frac{k}{m} \operatorname{tr}\left((I - H_{Q_2}) \operatorname{diag}\left(\frac{\operatorname{Var}[\hat{\Delta}_{j,n}]}{1 - [H_{Q_2}]_{jj}} : 1 \leq j \leq m\right)\right)$$

$$= \frac{k}{m} \sum_{1 \leq j \leq m} (1 - [H_{Q_2}]_{jj}) \frac{\operatorname{Var}[\hat{\Delta}_{j,n}]}{1 - [H_{Q_2}]_{jj}}$$

$$= \frac{k}{m} \sum_{1 \leq j \leq m} \operatorname{Var}[\hat{\Delta}_{j,n}] . \quad (32)$$

Moreover, this relationship holds with Q_2 replaced by ι_m. Therefore, the (31)–(32) imply

$$nE[\hat{V}_n^{\mathrm{F}}]$$
$$= \frac{k}{m}\delta_n'(I - H_{\iota_m})\delta_n - \frac{k}{m}\delta_n' Q_{2,X} \Sigma_{X,n}^{-1} \frac{1}{m} Q_{2,X}' \delta_n + \frac{k}{m} \operatorname{tr}((I - H_{\iota_m}) \operatorname{Var}[(\operatorname{diag}(I - H_{\iota_m}))^{-1/2} \hat{\delta}_n])$$
$$= nE[\hat{V}_n^{\mathrm{IM}}] - \frac{k}{m}\delta_n' Q_{2,X} \Sigma_{X,n}^{-1} \frac{1}{m} Q_{2,X}' \delta_n$$

Next, note $\frac{1}{m} Q_{2,X}' \delta_n \to \operatorname{Cov}_Q[\tilde{X}, \tilde{Y}(1) - \tilde{Y}(0)]$ under Assumption 4.2(a)–(c) by similar arguments to those used to prove (22)–(23). Further note $\frac{m-1}{m} \to 1$ and recall $\Sigma_{X,n} \to \operatorname{Var}_Q[\tilde{X}]$ and the limit of $nE[\hat{V}_n^{\mathrm{IM}}]$ derived above, and we have

$$nE[\hat{V}_n^{\mathrm{F}}] \to V^{\mathrm{obs}} + k \operatorname{Var}_Q[E_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}]] - k \operatorname{Cov}_Q[\tilde{X}, \tilde{Y}(1) - \tilde{Y}(0)] \operatorname{Var}_Q[\tilde{X}]^{-1} \operatorname{Cov}_Q[\tilde{X}, \tilde{Y}(1) - \tilde{Y}(0)]$$
$$= V^{\mathrm{obs}} + k \operatorname{Var}_Q[E_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}]] - k \operatorname{Var}_Q[\mathrm{BLP}_Q(\tilde{Y}(1) - \tilde{Y}(0)|1, \tilde{X})]$$
$$= V^{\mathrm{obs}} + k \operatorname{Var}_Q[E_Q[\tilde{Y}(1) - \tilde{Y}(0)|\tilde{X}] - \mathrm{BLP}_Q(\tilde{Y}(1) - \tilde{Y}(0)|1, \tilde{X})] ,$$

where the first equality follows from Lemma A.3(b) and the second follows from Lemma A.3(d).

Finally, the statement about the coverages of confidence intervals follow from Theorem A.1 (which applies because of (2)–(3)), (4), and Slutsky's theorem. ∎

A.5 Auxiliary Results

Here we state and prove a standard central limit theorem for design-based inference (see for instance Hájek, 1960; Li and Ding, 2017).

Theorem A.1. *Suppose that Assumption 2.1 holds and that*

$$\frac{\max_{1 \leq j \leq m} \max_{i \in \lambda_j} (R_i - \bar{R}_j)^2}{\sum_{1 \leq j \leq m} \sum_{i \in \lambda_j} (R_i - \bar{R}_j)^2} \to 0 \quad (33)$$

as $n \to \infty$, *where* $R_i := \frac{Y_i(1)}{\eta} + \frac{Y_i(0)}{1-\eta}$ *and* $\bar{R}_j := \frac{1}{k} \sum_{i \in \lambda_j} R_i$. *Then,*

$$\frac{\hat{\Delta}_n - \Delta_n}{\text{Var}[\hat{\Delta}_n]^{1/2}} \xrightarrow{d} N(0,1) \ .$$

Proof. Re-writing $\hat{\Delta}_n$ and Δ_n over the blocks $\{\lambda_j : 1 \le j \le n\}$, we obtain

$$\hat{\Delta}_n = \frac{1}{n} \left(\sum_{1 \le j \le m} \sum_{i \in \lambda_j} \left(\frac{Y_i(1)}{\eta} + \frac{Y_i(0)}{1-\eta} \right) D_i - \sum_{1 \le j \le m} \sum_{i \in \lambda_j} \frac{Y_i(0)}{1-\eta} \right) \ ,$$

and

$$\Delta_n = \frac{1}{n} \left(\sum_{1 \le j \le m} \sum_{i \in \lambda_j} \left(\bar{R}_j(1) + \bar{R}_j(0) \right) D_i - \sum_{1 \le j \le m} \sum_{i \in \lambda_j} \frac{Y_i(0)}{1-\eta} \right) \ ,$$

where $\bar{R}_j(1) = \frac{1}{k} \sum_{i \in \lambda_j} \frac{Y_i(1)}{\eta}$ and similarly for $\bar{R}_j(0)$. Putting both together,

$$\hat{\Delta}_n - \Delta_n = \frac{1}{n} \sum_{1 \le j \le m} \sum_{i \in \lambda_j} (R_i - \bar{R}_j) D_i \ ,$$

where $R_i = \frac{Y_i(1)}{\eta} + \frac{Y_i(0)}{1-\eta}$ and $\bar{R}_j = \bar{R}_j(1) + \bar{R}_j(0)$. Hence we obtain that

$$\sqrt{m}(\hat{\Delta}_n - \Delta_n) = \sum_{1 \le j \le m} A_j \ ,$$

where $A_j = \frac{1}{\sqrt{m}} \frac{\eta}{\ell} \sum_{i \in \lambda_j} (R_i - \bar{R}_j) D_i$. Note that $\frac{1}{\ell} \sum_{i \in \lambda_j} (R_i - \bar{R}_j) D_i$ is the sample mean when sampling a subset of size ℓ without replacement from the finite population $\{(R_i - \bar{R}_j) : i \in \lambda_j\}$. Accordingly, from the properties of the sample mean when sampling from a finite population (Theorems 2.1 and 2.2 in Cochran (1977)), we obtain immediately that

$$E[A_j] = 0 \ ,$$

$$\text{Var}[A_j] = \frac{1}{m} \eta^2 \left(\frac{1}{\ell} - \frac{1}{k} \right) \left(\frac{1}{k-1} \right) \sum_{i \in \lambda_j} (R_i - \bar{R}_j)^2 \ .$$

Moreover, by the definition of A_j and Jensen's inequality,

$$A_j^2 \lesssim \frac{1}{m} \max_{i \in \lambda_j} (R_i - \bar{R}_j)^2 \ .$$

To establish asymptotic normality, we verify the Lindeberg condition:

$$\sum_{1 \le j \le m} E \left[\frac{A_j^2}{s_n^2} I \left\{ \frac{A_j^2}{s_n^2} > \epsilon^2 \right\} \right] \to 0 \ ,$$

where $s_n^2 = \sum_{1 \le j \le m} \text{Var}[A_j]$. Note that by our previous calculations

$$s_n^2 \propto \frac{1}{m} \sum_{1 \le j \le m} \sum_{i \in \lambda_j} (R_i - \bar{R}_j)^2 \ ,$$

hence

$$\frac{A_j^2}{s_n^2} \lesssim \frac{\max_{i \in \lambda_j}(R_i - \bar{R}_j)^2}{\sum_{1 \le j \le m} \sum_{i \in \lambda_j}(R_i - \bar{R}_j)^2} \; .$$

Therefore, if (33) holds, then for each fixed $\epsilon > 0$, for m large enough,

$$\max_{1 \le j \le m} \frac{A_j^2}{s_n^2} \le \epsilon^2 \; ,$$

so

$$\sum_{1 \le j \le m} \frac{A_j^2}{s_n^2} I\left\{\frac{A_j^2}{s_n^2} > \epsilon^2\right\} \le \sum_{1 \le j \le m} \frac{A_j^2}{s_n^2} I\left\{\max_{1 \le j \le m} \frac{A_j^2}{s_n^2} > \epsilon^2\right\} = 0 \; .$$

As a result, the Lindeberg condition holds, and the conclusion follows. ∎

Lemma A.1. *Let $X_n, n \ge 1$ be a sequence of random variables such that $E[X_n] = 0$, $\mathrm{Var}[X_n] = \sigma_n^2$ and $X_n/\sigma_n \overset{d}{\to} N(0,1)$. Further suppose $\liminf_{n \to \infty} \sigma_n^2 > 0$, $\limsup_{n \to \infty} \sigma_n^2 < \infty$, and there exists a sequence of random variables $\hat{\sigma}_n^2, n \ge 1$ such that $\hat{\sigma}_n^2 - E[\hat{\sigma}_n^2] \overset{P}{\to} 0$. Then,*

$$\liminf_{n \to \infty} P\left\{\frac{X_n}{\hat{\sigma}_n} \le x\right\} \ge \Phi(x) \text{ for } x > 0$$

$$\limsup_{n \to \infty} P\left\{\frac{X_n}{\hat{\sigma}_n} \le x\right\} \le \Phi(x) \text{ for } x < 0 \; ,$$

where $\Phi(x) = P\{N(0,1) \le x\}$, if and only if $\liminf_{n \to \infty}(E[\hat{\sigma}_n^2] - \sigma_n^2) \ge 0$.

Proof. We will only prove the inequality when $x \ge 0$ and the other follows similarly. Because $\liminf_{n \to \infty} \sigma_n^2 > 0$, $\sigma_n^2 \ge c > 0$ for some $c > 0$, at least for n large enough. Since further $\liminf_{n \to \infty}(E[\hat{\sigma}_n^2] - \sigma_n^2) \ge 0$, $E[\hat{\sigma}_n^2] \ge \frac{3}{4}c$ for n large enough. Because further $\hat{\sigma}_n^2 - E[\hat{\sigma}_n^2] \overset{P}{\to} 0$, we have

$$P\{\hat{\sigma}_n^2 > c/2\} \to 1 \; .$$

Therefore, in what follows, all statements should be understood as conditioning on this event. In that case,

$$\frac{X_n}{\hat{\sigma}_n} = \frac{X_n}{\sigma_n} \frac{\sigma_n}{\hat{\sigma}_n} \; . \tag{34}$$

Next, we claim that for each $\epsilon > 0$,

$$P\left\{\frac{\sigma_n}{\hat{\sigma}_n} > 1 + \epsilon\right\} \to 0 \; . \tag{35}$$

Suppose for a moment that (35) holds. Then, because $X_n/\sigma_n \overset{d}{\to} N(0,1)$, $\sigma_n > 0$, and $\hat{\sigma}_n > 0$, (34) implies that for each $x \ge 0$ and $\epsilon > 0$,

$$P\left\{\frac{X_n}{\hat{\sigma}_n} > x(1 + \epsilon)\right\} \le P\left\{\frac{X_n}{\sigma_n} > x\right\} + P\left\{\frac{\sigma_n}{\hat{\sigma}_n} > 1 + \epsilon\right\} \to 1 - \Phi(x)$$

as $n \to \infty$. In other words, for $x > 0$ and $\epsilon > 0$,

$$\limsup_{n \to \infty} P\left\{\frac{X_n}{\hat{\sigma}_n} > x\right\} \le 1 - \Phi(x/(1 + \epsilon)) \; .$$

25

The conclusion of the theorem follows by letting $\epsilon \to 0$ and noting Φ is continuous everywhere. It therefore suffices to prove (35). Because (35) holds if and only if

$$I\left\{\frac{\sigma_n}{\hat{\sigma}_n} > 1 + \epsilon\right\} \xrightarrow{P} 0 , \tag{36}$$

it suffices to prove that for each subsequence $\{n(k)\}$, there exists a further subsequence $\{n(k(\ell))\}$ along which the convergence in (36) holds with probability one. Fix an arbitrary subsequence $\{n(k)\}$. Because $\liminf_{n\to\infty} \sigma_n^2 > 0$, $\liminf_{n\to\infty}(E[\hat{\sigma}_n^2] - \sigma_n^2) \geq 0$, and $\hat{\sigma}_n^2 - E[\hat{\sigma}_n^2] \xrightarrow{P} 0$, there exists a constant $c > 0$ and a further subsequence $\{n(k(\ell))\}$ along which $\sigma_{n(k(\ell))}^2 \geq c > 0$, $E[\hat{\sigma}_{n(k(\ell))}^2] \geq c/2 > 0$, and $\hat{\sigma}_{n(k(\ell))}^2 - E[\hat{\sigma}_{n(k(\ell))}^2] \to 0$ with probability one. Furthermore,

$$\limsup_{\ell\to\infty} \frac{\sigma_{n(k(\ell))}^2}{E[\hat{\sigma}_{n(k(\ell))}^2]} \leq 1 .$$

Along this subsequence,

$$\limsup_{\ell\to\infty} \frac{\sigma_{n(k(\ell))}^2}{\hat{\sigma}_{n(k(\ell))}^2} = \limsup_{\ell\to\infty} \frac{\sigma_{n(k(\ell))}^2}{E[\hat{\sigma}_{n(k(\ell))}^2]} \frac{E[\hat{\sigma}_{n(k(\ell))}^2]}{\hat{\sigma}_{n(k(\ell))}^2} \leq 1 ,$$

so (36) holds with probability one, and hence (35) follows.

Next, we show that $\liminf_{n\to\infty}(E[\hat{\sigma}_n^2] - \sigma_n^2) \geq 0$ is necessary for the result to hold. We present the proof for $x > 0$ while symmetric arguments work for $x < 0$. To that end, first suppose $\liminf_{n\to\infty}(E[\hat{\sigma}_n^2] - \sigma_n^2) = -c < 0$ and $\liminf_{n\to\infty} E[\hat{\sigma}_n^2] > 0$. Then, there exists a subsequence $\{n(k)\}$ along which $E[\hat{\sigma}_{n(k)}^2] < \sigma_{n(k)}^2 - c/2$ and $E[\hat{\sigma}_{n(k)}^2] \geq c/2$. Because by assumption $\limsup_{n\to\infty} \sigma_n^2 < \infty$, there exists $M > 0$ such that $\sigma_{n(k)}^2 \leq M$ for all k. Define $\epsilon = \frac{M}{M-c/2} - 1 > 0$. By arguing similarly as in (36), we can show

$$P\left\{\frac{\sigma_{n(k)}}{\hat{\sigma}_{n(k)}} \geq 1 + \epsilon\right\} \to 1 . \tag{37}$$

If instead $\liminf_{n\to\infty} E[\hat{\sigma}_n^2] = 0$, then by passing to a subsequence along which $E[\hat{\sigma}_{n(k)}^2] \to 0$, we can construct a further subsequence along which

$$\frac{\sigma_{n(k(\ell))}}{\hat{\sigma}_{n(k(\ell))}} \to \infty ,$$

with probability one, and thus (37) still holds. Then,

$$P\left\{\frac{X_{n(k)}}{\hat{\sigma}_{n(k)}} > x\right\} = P\left\{\frac{X_{n(k)}}{\sigma_{n(k)}} \frac{\sigma_{n(k)}}{\hat{\sigma}_{n(k)}} > x\right\} \geq P\left\{\left\{\frac{X_{n(k)}}{\sigma_{n(k)}} > \frac{x}{1+\epsilon}\right\} \cap \left\{\frac{\sigma_{n(k)}}{\hat{\sigma}_{n(k)}} \geq 1 + \epsilon\right\}\right\}$$

$$= P\left\{\frac{X_{n(k)}}{\sigma_{n(k)}} > \frac{x}{1+\epsilon}\right\} - P\left\{\left\{\frac{X_{n(k)}}{\sigma_{n(k)}} > \frac{x}{1+\epsilon}\right\} \cap \left\{\frac{\sigma_{n(k)}}{\hat{\sigma}_{n(k)}} < 1 + \epsilon\right\}\right\}$$

$$\to 1 - \Phi(x/(1+\epsilon)) > 1 - \Phi(x) .$$

Therefore,

$$\liminf_{n\to\infty} P\left\{\frac{X_n}{\hat{\sigma}_n} \leq x\right\} = 1 - \limsup_{n\to\infty} P\left\{\frac{X_n}{\hat{\sigma}_n} > x\right\} \leq \Phi(x/(1+\epsilon)) < \Phi(x) ,$$

and the desired result follows. ∎

26

Lemma A.2. *Suppose* $E[\hat{\delta}_n] = \delta_n$ *and A is deterministic. Then,* $E[\hat{\delta}_n' A \hat{\delta}_n] = \delta_n' A \delta_n + \mathrm{tr}(A \mathrm{Var}[\hat{\delta}_n])$.

Proof. Note

$$
\begin{aligned}
E[\hat{\delta}_n' A \hat{\delta}_n] &= E[(\delta_n + \hat{\delta}_n - \delta_n)' A (\delta_n + \hat{\delta}_n - \delta_n)] \\
&= \delta_n' A \delta_n + E[(\hat{\delta}_n - \delta_n)' A (\hat{\delta}_n - \delta_n)] \\
&= \delta_n' A \delta_n + \mathrm{tr}(E[(\hat{\delta}_n - \delta_n)' A (\hat{\delta}_n - \delta_n)]) \\
&= \delta_n' A \delta_n + E[\mathrm{tr}((\hat{\delta}_n - \delta_n)' A (\hat{\delta}_n - \delta_n))] \\
&= \delta_n' A \delta_n + E[\mathrm{tr}(A(\hat{\delta}_n - \delta_n)(\hat{\delta}_n - \delta_n)')] \\
&= \delta_n' A \delta_n + \mathrm{tr}(E[A(\hat{\delta}_n - \delta_n)(\hat{\delta}_n - \delta_n)']) \\
&= \delta_n' A \delta_n + \mathrm{tr}(A \mathrm{Var}[\hat{\delta}_n]) \ ,
\end{aligned}
$$

where the second equality follows because $E[\hat{\delta}_n - \delta_n] = 0$, the third follows because $E[(\hat{\delta}_n - \delta_n)' A (\hat{\delta}_n - \delta_n)]$ is a scalar, the fourth and sixth inequalities follow because the expectation operator is linear, and the fifth follows from $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ as long as the matrices are commutative. $\blacksquare$

Lemma A.3. *Let (Y, X) be random vectors where Y lives in $\mathbf{R}$ and X lives in $\mathbf{R}^k$. Then,*

(a) $E[\mathrm{BLP}(Y|1, X)] = E[Y]$.

(b) $(X - E[X])' \mathrm{Var}[X]^{-1} \mathrm{Cov}[X, Y] = \mathrm{BLP}(Y|1, X) - E[\mathrm{BLP}(Y|1, X)]$.

(c) $E[(Y - E[Y] - (\mathrm{BLP}(Y|1, X) - E[\mathrm{BLP}(Y|1, X)]))(\mathrm{BLP}(Y|1, X) - E[\mathrm{BLP}(Y|1, X)])] = 0$ *and therefore* $\mathrm{Cov}[Y, \mathrm{BLP}(Y|1, X)] = \mathrm{Var}[\mathrm{BLP}(Y|1, X)]$.

(d) $E([E[Y|X] - E[Y] - (\mathrm{BLP}(Y|1, X) - E[\mathrm{BLP}(Y|1, X)]))(\mathrm{BLP}(Y|1, X) - E[\mathrm{BLP}(Y|1, X)])] = 0$ *and therefore* $\mathrm{Cov}[E[Y|X], \mathrm{BLP}(Y|1, X)] = \mathrm{Var}[\mathrm{BLP}(Y|1, X)]$ *and* $\mathrm{Var}[E[Y|X]] = \mathrm{Var}[E[Y|X] - \mathrm{BLP}(Y|1, X)] + \mathrm{Var}[\mathrm{BLP}(Y|1, X)]$.

Proof. (a) follows because $E[(Y - \mathrm{BLP}(Y|1, X)) \cdot 1] = 0$. (b) follows because if $\mathrm{BLP}(Y|1, X) = \beta_0 + X' \beta_1$ then $\beta_1 = \mathrm{Var}[X]^{-1} \mathrm{Cov}[X, Y]$. (c) follows because recalling $E[\mathrm{BLP}(Y|1, X)] = E[Y]$, we have

$$
\begin{aligned}
&E[(Y - E[Y] - (\mathrm{BLP}(Y|1, X) - E[\mathrm{BLP}(Y|1, X)]))(\mathrm{BLP}(Y|1, X) - E[\mathrm{BLP}(Y|1, X)])] \\
&= E[(Y - \mathrm{BLP}(Y|1, X))(\mathrm{BLP}(Y|1, X) - E[Y])] \\
&= E[(Y - \mathrm{BLP}(Y|1, X)) \mathrm{BLP}(Y|1, X)] + E[Y - \mathrm{BLP}(Y|1, X)] E[Y] \\
&= 0 \ .
\end{aligned}
$$

(d) follows similarly. $\blacksquare$

References

ABADIE, A. and IMBENS, G. W. (2008). Estimation of the Conditional Variance in Paired Experiments. *Annales d'Économie et de Statistique* 175–187.

ARONOW, P. M., GREEN, D. P. and LEE, D. K. K. (2014). Sharp Bounds on the Variance in Randomized Experiments. *The Annals of Statistics*, **42** 850–871.

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

BAI, Y., GUO, H., SHAIKH, A. M. and TABORD-MEEHAN, M. (2024a). Inference in Experiments with Matched Pairs and Imperfect Compliance. *Journal of Business & Economic Statistics*, **0** 1–22.

BAI, Y., JIANG, L., ROMANO, J. P., SHAIKH, A. M. and ZHANG, Y. (2024b). Covariate adjustment in experiments with matched pairs. *Journal of Econometrics*, **241** 105740.

BAI, Y., LIU, J., SHAIKH, A. M. and TABORD-MEEHAN, M. (2024c). Inference in cluster randomized trials with matched pairs. *Journal of Econometrics*, **245** 105873.

BAI, Y., LIU, J., SHAIKH, A. M. and TABORD-MEEHAN, M. (2025). On the Efficiency of Finely Stratified Experiments. ArXiv:2307.15181 [econ], URL http://arxiv.org/abs/2307.15181.

BAI, Y., LIU, J. and TABORD-MEEHAN, M. (2024d). Inference for Matched Tuples and Fully Blocked Factorial Designs. *Quantitative Economics*, **15** 279–330.

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

COCHRAN, W. G. (1977). Sampling techniques. *Johan Wiley & Sons Inc.*

CYTRYNBAUM, M. (2024a). Covariate adjustment in stratified experiments. *Quantitative Economics*, **15** 971–998.

CYTRYNBAUM, M. (2024b). Optimal stratification of survey experiments. *arXiv preprint arXiv:2111.08157.*

DE CHAISEMARTIN, C. and RAMIREZ-CUELLAR, J. (2024). At what level should one cluster standard errors in paired and small-strata experiments? *American Economic Journal: Applied Economics*, **16** 193–212.

DING, P. (2017). A Paradox from Randomization-Based Causal Inference. *Statistical Science*, **32** 331–345.

FISHER, R. A. (1935). The design of experiments. *The design of experiments.*

FOGARTY, C. B. (2018a). On mitigating the analytical limitations of finely stratified experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **80** 1035–1056.

FOGARTY, C. B. (2018b). Regression-assisted inference for the average treatment effect in paired experiments. *Biometrika*, **105** 994–1000.

HÁJEK, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, **5** 361–374.

IMAI, K. (2008). Variance identification and efficiency analysis in randomized experiments under the matched-pair design. *Statistics in medicine*, **27** 4857–4873.

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

IMBENS, G. W. and RUBIN, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences.* Cambridge university press.

JIANG, L., LIU, X., PHILLIPS, P. C. and ZHANG, Y. (2024). Bootstrap inference for quantile treatment effects in randomized experiments with matched pairs. *Review of Economics and Statistics*, **106** 542–556.

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.

LIU, H. and YANG, Y. (2020). Regression-adjusted average treatment effect estimates in stratified randomized experiments. *Biometrika*, **107** 935–948.

PASHLEY, N. E. and MIRATRIX, L. W. (2021). Insights on variance estimation for blocked and matched pairs designs. *Journal of Educational and Behavioral Statistics*, **46** 271–296.

SU, F. and DING, P. (2021). Model-assisted analyses of cluster-randomized experiments. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **83** 994–1015.

ZHU, K., LIU, H. and YANG, Y. (2024). Design-Based Theory for Lasso Adjustment in Randomized Block Experiments and Rerandomized Experiments. *Journal of Business & Economic Statistics*, **0** 1–12.