

Inference in Experiments with Matched Pairs and Imperfect Compliance*

Yuehao Bai

Department of Economics
University of Southern California

yuehao.bai@usc.edu

Hongchang Guo

Department of Economics
Northwestern University

hongchangguo2028@u.northwestern.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

July 26, 2023

Abstract

This paper studies inference for the local average treatment effect in randomized controlled trials with imperfect compliance where treatment status is determined according to “matched pairs.” By “matched pairs,” we mean that units are sampled i.i.d. from the population of interest, paired according to observed, baseline covariates and finally, within each pair, one unit is selected at random for treatment. Under weak assumptions governing the quality of the pairings, we first derive the limiting behavior of the usual Wald (i.e., two-stage least squares) estimator of the local average treatment effect. We show further that the conventional heteroskedasticity-robust estimator of its limiting variance is generally conservative in that its limit in probability is (typically strictly) larger than the limiting variance. We therefore provide an alternative estimator of the limiting variance that is consistent for the desired quantity. Finally, we consider the use of additional observed, baseline covariates not used in pairing units to increase the precision with which we can estimate the local average treatment effect. To this end, we derive the limiting behavior of a two-stage least squares estimator of the local average treatment effect which includes both the additional covariates in addition to pair fixed effects, and show that the limiting variance is always less than or equal to that of the Wald estimator. To complete our analysis, we provide a consistent estimator of this limiting variance. A simulation study confirms the practical relevance of our theoretical results. We use our results to revisit a prominent experiment studying the effect of macroinsurance on microenterprise in Egypt.

KEYWORDS: Matched pairs, randomized controlled trial, experiments, noncompliance, imperfect compliance

JEL classification codes: C12, C14

*We thank Alex Torgovitsky for helpful discussions. The fourth author acknowledges support from NSF grant SES-2149408.

1 Introduction

This paper studies inference for the local average treatment effect in randomized controlled trials with imperfect compliance where treatment status is determined according to “matched pairs.” By “matched pairs,” we mean that units are sampled i.i.d. from the population of interest, paired according to observed, baseline covariates and finally, within each pair, one unit is selected at random for treatment. This method is used routinely in all parts of sciences. Indeed, commands to facilitate its implementation are included in popular software packages, such as `samps` in Stata. References to a variety of specific examples can be found, for instance, in the following surveys of various field experiments: [Glennerster and Takavarasha \(2014\)](#), [Riach and Rich \(2002\)](#), [Rosenberger and Lachin \(2015\)](#). See also [Bruhn and McKenzie \(2009\)](#), who, based on a survey of selected development economists, report that 56% of researchers have used such a design at some point. Furthermore, in many such experiments, compliance may be imperfect: some recent examples of experiments featuring both matched pairs and imperfect compliance are [Groh and McKenzie \(2016\)](#) and [Resnjanskij et al. \(2021\)](#). Under weak assumptions that ensure pairs are formed so that units within pairs are suitably “close” in terms of observed, baseline covariates, we derive a variety of results pertaining to inference about the local average treatment effect in such experiments.

We first study the behavior of the usual Wald (i.e., two-stage least squares) estimator of the local average treatment effect. When all observed, baseline covariates are used in forming pairs, we find that the estimator is efficient among all estimators for the local average treatment effect in the sense that it achieves the lower bound on the limiting variance developed in [Bai et al. \(2023b\)](#) over a broad class of treatment assignment schemes that hold the marginal probability of treatment assignment equal to one half, thereby including matched pairs, in particular, as a special case. On the other hand, we find that the conventional heteroskedasticity-robust estimator of its limiting variance is conservative in that its limit in probability is always weakly larger than the limiting variance and strictly larger unless heterogeneity is constrained in a particular fashion. We provide an alternative estimator of the limiting variance and show that it is consistent for the desired quantity. In a simulation study, we find that tests using the Wald estimator together with the conventional heteroskedasticity-robust estimator of its limiting variance may as a consequence have very poor power when compared to tests using the Wald estimator together with our estimator of its limiting variance.

Next, we consider situations in which there are additional, observed, baseline covariates that are not used when pairing units. We first derive the limiting behavior for a class of covariate-adjusted estimators indexed by different “working models” for conditional expectations. Importantly, these working models need not be correctly specified for the true conditional expectations in order for this estimator to remain consistent for the local average treatment effect, but we show that the limiting variance of the estimator is minimized, in particular, when they are correctly specified. These results are derived under a high-level assumption governing the quality of the estimators of these working models. We therefore specialize our results for the case of linear working models and characterize the optimal linear working model in the sense of minimizing the limiting variance of the resulting covariate-adjusted estimator among all such working models. We find that this limiting variance can be obtained by a familiar estimator: the two-stage least squares estimator

using treatment assignment as an instrument of the coefficient on treatment status in a linear regression of the outcome on the following quantities: treatment status; (functions of) the observed baseline covariates; and pair fixed effects. We complete our analysis by providing an estimator of the limiting variance and show that it is consistent for the desired quantity.

Our paper builds upon the analysis of [Bai et al. \(2022\)](#), who analyzed the behavior of the difference-in-means estimator of the average treatment effect in context of experiments with matched pairs and perfect compliance. Our covariate-adjusted estimator is inspired by the analysis in [Bai et al. \(2023a\)](#), who studied the use of additional, observed, baseline covariates in experiments with matched pairs and perfect compliance to improve the precision with which we can estimate the average treatment effect (similar results have also been obtained by [Cytrynbaum, 2023](#)). We emphasize, however, that none of the aforementioned papers permit imperfect compliance, which, as argued by [Athey and Imbens \(2017\)](#), is one of the most common complications in even the most well designed experiments. We note, however, that imperfect compliance has been studied in the context of randomized controlled trials with other treatment assignment schemes, such as stratified block randomization: see, for example, [Ansel et al. \(2018\)](#), [Bugni and Gao \(2021\)](#), and [Jiang et al. \(2022\)](#). We also emphasize that all of these papers, like ours, carry out their analysis in a superpopulation sampling framework. In this way, our analysis differs from the analysis of experiments with a finite population sampling framework.

The remainder of the paper is organized as follows. In [Section 2](#) we describe our setup and notation. In particular, there we describe the precise sense in which we require that units in each pair are “close” in terms of their baseline covariates. Our main results concerning the Wald estimator are contained in [Section 3](#). In [Section 4](#), we develop our results pertaining to our covariate-adjusted estimator that exploits additional observed, baseline covariates not used in pairing units to increase the precision with which we can estimate the local average treatment effect. In [Section 5](#), we examine the practical relevance of our theoretical results via a small simulation study. In [Section 6](#), we provide a brief empirical illustration of our proposed tests using data from an experiment in [Groh and McKenzie \(2016\)](#). Finally, we conclude in [Section 7](#) with some recommendations for empirical practice guided by both our theoretical results and our simulation study. As explained further in that section, we do not recommend the use of the Wald estimator with the conventional heteroskedasticity-robust estimator of its limiting variance because it is conservative in the sense described above; we instead encourage the use of the Wald estimator with our consistent estimator of its limiting variance because it is asymptotically exact, and, as a result, considerably more powerful. When there are additional, observed, baseline covariates that are not used when forming pairs, we recommend the use of our covariate-adjusted Wald estimator with our consistent estimator of its limiting variance. Proofs of all results are provided in the Appendix.

2 Setup and Notation

Let $Y_i \in \mathbf{R}$ denote the (observed) outcome of the i th unit, $A_i \in \{0, 1\}$ be an indicator for whether or not unit i is assigned to treatment, $D_i \in \{0, 1\}$ be an indicator for whether or not unit i decides to take up treatment, $X_i \in \mathbf{R}^{k_x}$ denote observed, baseline covariates for the i th unit which are used for matching,

and $W_i \in \mathbf{R}^{k_w}$ denote observed, baseline covariates for the i th unit which will be used when we consider covariate adjustment in Section 4. In contrast to the setting considered in [Bai et al. \(2022\)](#), we allow for imperfect compliance, i.e. for $D_i \neq A_i$. Further denote by $Y_i(d)$ the potential outcome of the i th unit if they make treatment decision $d \in \{0, 1\}$, and by $D_i(a)$ the potential treatment decision of the i th unit if assigned to treatment $a \in \{0, 1\}$. The observed treatment decision and potential treatment decision are related to treatment assignment via the usual relationship

$$D_i = D_i(1)A_i + D_i(0)(1 - A_i) , \quad (1)$$

and the observed outcome and potential outcome are related to treatment decision via the relationship

$$Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i) . \quad (2)$$

We will also often make use of the following alternative representation for the observed outcome, which is numerically equivalent to (2):

$$Y_i = \tilde{Y}_i(1)A_i + \tilde{Y}_i(0)(1 - A_i) , \quad (3)$$

where

$$\tilde{Y}_i(a) = Y_i(1)D_i(a) + Y_i(0)(1 - D_i(a)) \quad (4)$$

for $a \in \{0, 1\}$. In words, $\tilde{Y}_i(a)$ represents the “intent-to-treat” potential outcome for unit i when assigned to treatment $a \in \{0, 1\}$.

Following [Imbens and Angrist \(1994\)](#), each participant in the experiment can be categorized into one of four types: units for which $D_i(1) = 1$ and $D_i(0) = 0$ are referred to as compliers, units for which $D_i(1) = 1$ and $D_i(0) = 1$ are referred to as always takers, units for which $D_i(1) = 0$ and $D_i(0) = 0$ are referred to as never takers, and finally units for which $D_i(1) = 0$ and $D_i(0) = 1$ are referred to as defiers. We use the notation

$$C_i = I \{D_i(1) = 1, D_i(0) = 0\} \quad (5)$$

below to indicate whether or not unit i is a complier.

For a random variable indexed by i , say for example A_i , it will be useful to denote by $A^{(n)}$ the random vector (A_1, \dots, A_{2n}) . Denote by P_n the distribution of the observed data $(Y^{(n)}, D^{(n)}, A^{(n)}, X^{(n)}, W^{(n)})$, and by Q_n the distribution of $(Y^{(n)}(1), Y^{(n)}(0), D^{(n)}(1), D^{(n)}(0), X^{(n)}, W^{(n)})$. Note that P_n is jointly determined by (1), (2), Q_n , and the mechanism for determining treatment assignment. We assume throughout that our sample consists of $2n$ i.i.d. observations, so that $Q_n = Q^{2n}$, where Q is the marginal distribution of $(Y_i(1), Y_i(0), D_i(1), D_i(0), X_i, W_i)$. We therefore state our assumptions below in terms of assumptions on Q and the mechanism for determining treatment assignment. Indeed, we will not make reference to P_n in the sequel and all operations are understood to be under Q and the mechanism for determining treatment assignment.

Our object of interest is the local average treatment effect, which may be expressed in our notation as

$$\Delta(Q) = E[Y_i(1) - Y_i(0)|C_i = 1]. \quad (6)$$

For a pre-specified choice of Δ_0 , the testing problem of interest is

$$H_0 : \Delta(Q) = \Delta_0 \text{ versus } H_1 : \Delta(Q) \neq \Delta_0 \quad (7)$$

at level $\alpha \in (0, 1)$.

We begin by describing our primary assumptions on the data generating process.

Assumption 2.1. The distribution Q is such that

- (a) $0 < E \left[\text{Var} \left[\tilde{Y}_i(a) - \Delta(Q)D_i(a)|X_i \right] \right]$ for $a \in \{0, 1\}$.
- (b) $E[Y_i(a)^2] < \infty$ for $a \in \{0, 1\}$.
- (c) $E[Y_i(1)^r D_i(a)|X_i = x]$ and $E[Y_i(0)^r (1 - D_i(a))|X_i = x]$ are Lipschitz for $a = 0, 1$ and $r = 0, 1, 2$.
- (d) $P\{D_i(1) \geq D_i(0)\} = 1$.
- (e) $P\{C_i = 1\} > 0$.

Assumption 2.1(a)–(b) are mild restrictions imposed to rule out degenerate situations and to permit the application of suitable laws of large numbers and central limit theorems. Assumption 2.1(c) is a smoothness requirement that ensures that units that are “close” in terms of their baseline covariates are suitably comparable. Similar smoothness requirements are also considered in Bai et al. (2022), Cytrynbaum (2023), and generally play a key role in establishing the asymptotic exactness of our proposed tests. Assumptions 2.1(d)–(e) are the standard “monotonicity” and “relevance” conditions of Imbens and Angrist (1994) which ensure that the probability limit of the Wald estimator which we define in Section 3 can be interpreted as the local average treatment effect $\Delta(Q)$.

Next, we describe our assumptions on the mechanism determining treatment assignment. Following the notation in Bai et al. (2022), the n pairs can be represented by the sets

$$\{\pi(2j - 1), \pi(2j)\} \text{ for } j = 1, \dots, n,$$

where $\pi = \pi_n(X^{(n)})$ is a permutation of $2n$ elements. Given such a π , we assume that treatment status is assigned as described in the following assumption:

Assumption 2.2. Treatment status is assigned so that $(Y^{(n)}(1), Y^{(n)}(0), D^{(n)}(1), D^{(n)}(0), W^{(n)}) \perp\!\!\!\perp A^{(n)}|X^{(n)}$, and conditional on $X^{(n)}$, $(A_{\pi(2j-1)}, A_{\pi(2j)})$, $j = 1, \dots, n$ are i.i.d. and each uniformly distributed over the values in $\{(0, 1), (1, 0)\}$.

We further require that the units in each pair be “close” in terms of their baseline covariates in the following sense:

Assumption 2.3. The pairs used in determining treatment status satisfy

$$\frac{1}{n} \sum_{1 \leq j \leq n} \|X_{\pi(2j)} - X_{\pi(2j-1)}\|_2^r \xrightarrow{P} 0 ,$$

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

We will also sometimes require that the distances between units in adjacent pairs be “close” in terms of their baseline covariates:

Assumption 2.4. The pairs used in determining treatment status satisfy

$$\frac{1}{n} \sum_{1 \leq j \leq \lfloor \frac{n}{2} \rfloor} \|X_{\pi(4j-k)} - X_{\pi(4j-l)}\|_2^2 \xrightarrow{P} 0 ,$$

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

Bai et al. (2022) and Bai et al. (2023c) provide several examples of pairing algorithms which satisfy Assumptions 2.3–2.4. The simplest such example is when $X_i \in \mathbf{R}$, in which case we can order units from smallest to largest according to X_i and pair adjacent units. It then follows from Theorem 4.1 in Bai et al. (2022) that Assumptions 2.3–2.4 are satisfied as long as $E[X_i^2] < \infty$.

3 Main Results

3.1 Asymptotic Behavior of the Wald Estimator

In this section, we study the asymptotic properties of the standard Wald estimator (i.e., the two-stage least squares estimator of Y_i on D_i using A_i as an instrument) of $\Delta(Q)$ under a matched pairs design. In order to introduce this estimator, define

$$\hat{\psi}_n(a) = \frac{1}{n} \sum_{1 \leq i \leq 2n: A_i=a} Y_i , \tag{8}$$

$$\hat{\phi}_n(a) = \frac{1}{n} \sum_{1 \leq i \leq 2n: A_i=a} D_i . \tag{9}$$

Using this notation, the Wald estimator is defined as

$$\hat{\Delta}_n = \frac{\hat{\psi}_n(1) - \hat{\psi}_n(0)}{\hat{\phi}_n(1) - \hat{\phi}_n(0)} . \tag{10}$$

Note that this estimator may be obtained as the ratio of the estimator of the coefficient of A_i in an ordinary least squares regression of Y_i on a constant and A_i (the “reduced-form”) to the estimator of the coefficient of A_i in an ordinary least squares regression of D_i on a constant and A_i (the “first-stage”). Theorem 3.1 establishes the limiting distribution of $\hat{\Delta}_n$ under a matched pairs design.

Theorem 3.1. *Suppose Q satisfies Assumption 2.1 and the treatment assignment mechanism satisfies Assumptions 2.2–2.3. Then,*

$$\sqrt{n} \left(\hat{\Delta}_n - \Delta(Q) \right) \xrightarrow{d} N(0, \nu^2),$$

where

$$\begin{aligned} \nu^2 = & \frac{1}{P\{C_i = 1\}^2} \left(\text{Var}[Y_i^*(1)] + \text{Var}[Y_i^*(0)] \right. \\ & \left. - \frac{1}{2} E \left[((E[Y_i^*(1)|X_i] - E[Y_i^*(1)]) + (E[Y_i^*(0)|X_i] - E[Y_i^*(0)]))^2 \right] \right), \end{aligned}$$

with

$$Y_i^*(a) = \tilde{Y}_i(a) - \Delta(Q)D_i(a), \quad (11)$$

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

Note that the expression for ν^2 corresponds exactly to the limiting variance obtained in Bai et al. (2022) with the usual potential outcomes $Y_i(a)$ replaced with the transformed outcomes $Y_i^*(a)$. In particular, when there is perfect compliance, so that $D_i = A_i$, $P\{C_i = 1\} = 1$, and $D_i(a) = a$, the limiting variance we obtain in Theorem 3.1 corresponds exactly to the limiting variance derived in Bai et al. (2022). Moreover, it can be shown that our expression for ν^2 attains the efficiency bound derived in Bai et al. (2023b) over a broad class of treatment assignments which include matched pairs as a special case (it can also be shown that this bound coincides with the bounds derived in Frölich, 2007; Hong and Nekipelov, 2010, in settings with observational data with i.i.d assignment when $P\{A_i = a|X_i\} = 1/2$, where the local average treatment effect is estimated non-parametrically using the covariates X_i : see Lemma A.1 for details.)

Remark 3.1. It can be shown that under our matched pairs design, the two-stage least squares estimator of β_1 in the linear regression

$$Y_i = \beta_0 + \beta_1 D_i + \gamma' X_i + \epsilon_i,$$

with A_i used as an instrument for D_i , is asymptotically equivalent to the Wald estimator $\hat{\Delta}_n$. This mirrors similar observations made in Remark 3.8 in Bai et al. (2022) for the case of perfect compliance. ■

3.2 Variance Estimation

In this section, we construct a consistent variance estimator for the limiting variance ν^2 , and then contrast this to the asymptotic behavior of standard regression-based variance estimators. As noted in the discussion following Theorem 3.1, the expression for ν^2 corresponds exactly to the limiting variance obtained in Bai et al. (2022) with the usual potential outcomes $Y_i(a)$ replaced with the transformed outcomes $Y_i^*(a)$. We thus follow the variance construction from Bai et al. (2022), but with a feasible version of $Y_i^*(a)$ defined as

$$\hat{Y}_i = Y_i - \hat{\Delta}_n D_i.$$

This strategy leads to the following variance estimator:

$$\hat{\nu}_n^2 = \frac{\hat{\tau}_n^2 - \frac{1}{2}(\hat{\lambda}_n^2 + \hat{\Gamma}_n^2)}{(\hat{\phi}_n(1) - \hat{\phi}_n(0))^2}, \quad (12)$$

where

$$\begin{aligned} \hat{\tau}_n^2 &= \frac{1}{n} \sum_{1 \leq j \leq n} (\hat{Y}_{\pi(2j)} - \hat{Y}_{\pi(2j-1)})^2 \\ \hat{\lambda}_n^2 &= \frac{2}{n} \sum_{1 \leq j \leq \lfloor \frac{n}{2} \rfloor} (\hat{Y}_{\pi(4j-3)} - \hat{Y}_{\pi(4j-2)}) (\hat{Y}_{\pi(4j-1)} - \hat{Y}_{\pi(4j)}) (A_{\pi(4j-3)} - A_{\pi(4j-2)}) (A_{\pi(4j-1)} - A_{\pi(4j)}) \\ \hat{\Gamma}_n &= \frac{1}{n} \sum_{1 \leq i \leq 2n: A_i=1} \hat{Y}_i - \frac{1}{n} \sum_{1 \leq i \leq 2n: A_i=0} \hat{Y}_i. \end{aligned}$$

Note that the construction of the numerator of $\hat{\nu}_n^2$ can be motivated using a similar intuition to what has been previously discussed in [Bai et al. \(2022\)](#): to consistently estimate

$$E[(E[Y_i^*(1)|X_i] - E[Y_i^*(0)|X_i])^2],$$

ideally we would like access to four different units with similar values of X_i , of which two are treated. However, because each pair only contains two units, we need to average across “pairs of pairs” of units, where two pairs are grouped together so that they are “close” in terms of X_i . We establish the following consistency result for $\hat{\nu}_n^2$:

Theorem 3.2. *Suppose Q satisfies Assumption 2.1 and the treatment assignment mechanism satisfies Assumptions 2.2–2.4. Then,*

$$\hat{\nu}_n^2 \xrightarrow{P} \nu^2.$$

We thus immediately obtain the following corollary which establishes the asymptotic exactness of a t -test for the null hypothesis (7) constructed using the variance estimator $\hat{\nu}_n^2$:

Corollary 3.1. *Suppose Q satisfies Assumption 2.1 and the treatment assignment mechanism satisfies Assumptions 2.2–2.4. Then,*

$$\frac{\sqrt{n}(\hat{\Delta}_n - \Delta(Q))}{\hat{\nu}_n} \xrightarrow{d} N(0, 1).$$

Next, we consider the limiting behavior of the usual heteroskedasticity-robust variance estimator from a standard two-stage least squares regression, which is commonly used in practice: see [Groh and McKenzie \(2016\)](#), and [Resnjanskij et al. \(2021\)](#) for examples. Let \hat{U}_i denote the i th residual generated by the two-stage least squares estimator in a linear regression of Y_i on a constant and D_i using A_i as an instrument for D_i .

The heteroskedasticity-robust variance estimator is defined as

$$\hat{\omega}_n^2 = \left(n \left(\sum_{1 \leq i \leq 2n} \begin{pmatrix} 1 \\ A_i \end{pmatrix} \begin{pmatrix} 1 & D_i \end{pmatrix} \right)^{-1} \sum_{1 \leq i \leq 2n} \hat{U}_i^2 \begin{pmatrix} 1 \\ A_i \end{pmatrix} \begin{pmatrix} 1 & A_i \end{pmatrix} \left(\sum_{1 \leq i \leq 2n} \begin{pmatrix} 1 \\ D_i \end{pmatrix} \begin{pmatrix} 1 & A_i \end{pmatrix} \right)^{-1} \right)_{2,2}, \quad (13)$$

where the notation $(\cdot)_{2,2}$ denotes the $(2, 2)$ -element of its (matrix) argument. Theorem 3.3 derives the limit in probability of $\hat{\omega}_n^2$.

Theorem 3.3. *Suppose Q satisfies Assumption 2.1 and the treatment assignment mechanism satisfies Assumptions 2.2–2.3. Then,*

$$\hat{\omega}_n^2 \xrightarrow{P} \omega^2,$$

where

$$\omega^2 = \frac{1}{P\{C_i = 1\}^2} (\text{Var}[Y_i^*(1)] + \text{Var}[Y_i^*(0)]).$$

We thus immediately obtain the following corollary which establishes the *conservativeness* of a t -test constructed using the variance estimator $\hat{\omega}_n^2$ for the null hypothesis (7).

Corollary 3.2. *Suppose Q satisfies Assumption 2.1 and the treatment assignment mechanism satisfies Assumptions 2.2–2.3. Then,*

$$\frac{\sqrt{n} \left(\hat{\Delta}_n - \Delta(Q) \right)}{\hat{\omega}_n} \xrightarrow{d} N(0, \xi^2),$$

where $\xi^2 \leq 1$ and the inequality is strict unless

$$E[\tilde{Y}_i(1) - \Delta(Q)D_i(1) + \tilde{Y}_i(0) - \Delta(Q)D_i(0) | X_i] = E[\tilde{Y}_i(1) - \Delta(Q)D_i(1) + \tilde{Y}_i(0) - \Delta(Q)D_i(0)]$$

with probability one under Q .

4 Covariate Adjustment

In this section, we consider a generalization of the estimator $\hat{\Delta}_n$ defined in Section 3.1 that allows for covariate adjustment using the additional, observed, baseline covariates $W^{(n)}$ that are not used when forming pairs. We then show how a careful application of this covariate-adjusted estimator can ensure an improvement over the unadjusted Wald estimator $\hat{\Delta}_n$ in terms of precision.

Following Bai et al. (2023a), note that it can be shown under Assumption 2.2 that for any $a \in \{0, 1\}$, $m_{a, \tilde{Y}} : \mathbb{R}^{k_x} \times \mathbb{R}^{k_w} \rightarrow \mathbb{R}$, and $m_{a, D} : \mathbb{R}^{k_x} \times \mathbb{R}^{k_w} \rightarrow \mathbb{R}$ such that $E[|m_{a, \tilde{Y}}(X_i, W_i)|] < \infty$, $E[|m_{a, D}(X_i, W_i)|] < \infty$,

$$E[2I\{A_i = a\}(Y_i - m_{a, \tilde{Y}}(X_i, W_i)) + m_{a, \tilde{Y}}(X_i, W_i)] = E[\tilde{Y}_i(a)], \quad (14)$$

$$E[2I\{A_i = a\}(D_i - m_{a,D}(X_i, W_i)) + m_{a,D}(X_i, W_i)] = E[D_i(a)]. \quad (15)$$

Here, we view $m_{a,D}$ and $m_{a,\tilde{Y}}$ as “working models” of $E[\tilde{Y}_i(a)|X_i, W_i]$ and $E[D_i(a)|X_i, W_i]$, respectively, but we emphasize that (14)–(15) hold even if they are incorrectly specified. This observation suggests a covariate-adjusted estimator defined as

$$\hat{\Delta}_n^{\text{adj}} = \frac{\hat{\psi}_n^{\text{adj}}(1) - \hat{\psi}_n^{\text{adj}}(0)}{\hat{\phi}_n^{\text{adj}}(1) - \hat{\phi}_n^{\text{adj}}(0)},$$

where

$$\begin{aligned} \hat{\psi}_n^{\text{adj}}(a) &= \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2I\{A_i = a\}(Y_i - \hat{m}_{a,\tilde{Y}}(X_i, W_i)) + \hat{m}_{a,\tilde{Y}}(X_i, W_i)), \\ \hat{\phi}_n^{\text{adj}}(a) &= \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2I\{A_i = a\}(D_i - \hat{m}_{a,D}(X_i, W_i)) + \hat{m}_{a,D}(X_i, W_i)), \end{aligned}$$

and $\hat{m}_{a,\tilde{Y}}$ and $\hat{m}_{a,D}$ are suitable estimators of $m_{a,\tilde{Y}}$ and $m_{a,D}$. Note that if we set $\hat{m}_{a,\tilde{Y}} = \hat{m}_{a,D} = 0$, then $\hat{\Delta}_n^{\text{adj}}$ simplifies to $\hat{\Delta}_n$. Theorem 4.1 below establishes the limiting distribution of $\hat{\Delta}_n^{\text{adj}}$ for a matched-pairs design under the following high-level assumption on the working models:

Assumption 4.1. Let $m_{a,\tilde{Y}D} = m_{a,\tilde{Y}} - \Delta(Q)m_{a,D}$. The functions $m_{a,\tilde{Y}D}$ for $a \in \{0, 1\}$ satisfy

(a) For $a \in \{0, 1\}$,

$$\lim_{\lambda \rightarrow \infty} E \left[m_{a,\tilde{Y}D}(X_i, W_i)^2 I \left\{ |m_{a,\tilde{Y}D}(X_i, W_i)| > \lambda \right\} \right] = 0.$$

(b) $E[m_{a,\tilde{Y}D}(X_i, W_i)|X_i = x]$, $E[(m_{a,\tilde{Y}D}(X_i, W_i))^2|X_i = x]$, and $E[m_{a,\tilde{Y}D}(X_i, W_i)Y_i^*(a)|X_i = x]$ for $d \in \{0, 1\}$, and $E[m_{1,\tilde{Y}D}(X_i, W_i)m_{0,\tilde{Y}D}(X_i, W_i)|X_i = x]$ are Lipschitz.

Before proceeding, we note that we later provide low-level sufficient conditions for this high-level assumption for the special case in which $m_{a,\tilde{Y}}$ and $m_{a,D}$ correspond to linear working models that are optimal in the sense of minimizing the limiting variance of $\hat{\Delta}_n^{\text{adj}}$ among all linear working models.

Theorem 4.1. Suppose Q satisfies Assumption 2.1, the treatment assignment mechanism satisfies Assumptions 2.2–2.3, and $m_{a,\tilde{Y}}$ and $m_{a,D}$ for $a \in \{0, 1\}$ satisfy Assumption 4.1. Further suppose $\hat{m}_{a,\tilde{Y}}$ and $\hat{m}_{a,D}$ satisfy that

$$\frac{1}{\sqrt{2n}} \sum_{1 \leq i \leq 2n} (2A_i - 1)(\hat{m}_{a,\tilde{Y}}(X_i, W_i) - m_{a,\tilde{Y}}(X_i, W_i)) \xrightarrow{P} 0, \quad (16)$$

$$\frac{1}{\sqrt{2n}} \sum_{1 \leq i \leq 2n} (2A_i - 1)(\hat{m}_{a,D}(X_i, W_i) - m_{a,D}(X_i, W_i)) \xrightarrow{P} 0. \quad (17)$$

Then, $\hat{\Delta}_n^{\text{adj}}$ satisfies

$$\sqrt{n} \left(\hat{\Delta}_n^{\text{adj}} - \Delta(Q) \right) \xrightarrow{d} N(0, \nu_{\text{adj}}^2),$$

where $\nu_{\text{adj}}^2 = \frac{1}{P\{C_i=1\}^2}(\nu_{1,\text{adj}}^2 + \nu_{2,\text{adj}}^2 + \nu_{3,\text{adj}}^2)$ with

$$\begin{aligned}\nu_{1,\text{adj}}^2 &= \frac{1}{2}E[\text{Var}[E[Y_i^*(1) + Y_i^*(0)|X_i, W_i] - (m_{1,\tilde{Y}D}(X_i, W_i) + m_{0,\tilde{Y}D}(X_i, W_i))|X_i]] , \\ \nu_{2,\text{adj}}^2 &= \frac{1}{2}\text{Var}[E[Y_i^*(1) - Y_i^*(0)|X_i, W_i]] , \\ \nu_{3,\text{adj}}^2 &= E[\text{Var}[Y_i^*(1)|X_i, W_i] + \text{Var}[Y_i^*(0)|X_i, W_i]] .\end{aligned}$$

Note that the expression for ν_{adj}^2 corresponds exactly to the limiting variance obtained in [Bai et al. \(2023a\)](#) with the usual potential outcomes $Y_i(a)$ replaced with the transformed outcomes $Y_i^*(a)$ and the working models in [Bai et al. \(2023a\)](#) replaced with $m_{a,\tilde{Y}}(X_i, W_i) - \Delta(Q)m_{a,D}(X_i, W_i)$. In general, ν_{adj}^2 is not guaranteed to be weakly smaller than ν^2 for all choices of working models, but we note that ν_{adj}^2 is minimized when $\nu_{1,\text{adj}}^2 = 0$, i.e., when the working models satisfy

$$\begin{aligned}E[Y_i^*(1) + Y_i^*(0)|X_i, W_i] - E[Y_i^*(1) + Y_i^*(0)|X_i] \\ = m_{1,\tilde{Y}D}(X_i, W_i) + m_{0,\tilde{Y}D}(X_i, W_i) - E[m_{1,\tilde{Y}D}(X_i, W_i) + m_{0,\tilde{Y}D}(X_i, W_i)|X_i]\end{aligned}$$

with probability one. This property holds, in particular, when $m_{a,\tilde{Y}}$ and $m_{a,D}$ are correctly specified.

We now specialize [Theorem 4.1](#) to the case in which $m_{a,\tilde{Y}}$ and $m_{a,D}$ are the optimal linear working models in a sense to be described below. To that end, let $\zeta_i = \zeta(X_i, W_i)$ be a user-specified transformation of the baseline characteristics (X_i, W_i) given by some function $\zeta : \mathbb{R}^{k_x} \times \mathbb{R}^{k_w} \rightarrow \mathbb{R}^p$. Denote by $\hat{\beta}_n^Y$ and $\hat{\beta}_n^D$ the ordinary least squares estimators of β^Y and β^D in the following two linear regressions with pair fixed effects:

$$Y_i = \alpha^Y A_i + \zeta_i' \beta^Y + \sum_{1 \leq j \leq n} \theta_j^Y I\{i \in \{\pi(2j-1), \pi(2j)\}\} + \epsilon_i^Y \quad (18)$$

$$D_i = \alpha^D A_i + \zeta_i' \beta^D + \sum_{1 \leq j \leq n} \theta_j^D I\{i \in \{\pi(2j-1), \pi(2j)\}\} + \epsilon_i^D \quad (19)$$

Using this notation, consider $\hat{\Delta}_n^{\text{adj}}$ with $\hat{m}_{a,\tilde{Y}}(X_i, W_i) = \zeta_i' \hat{\beta}_n^Y$ and $\hat{m}_{a,D}(X_i, W_i) = \zeta_i' \hat{\beta}_n^D$ for $a \in \{0, 1\}$. In order to analyze the large-sample behavior of this estimator, we introduce the following assumption:

Assumption 4.2. The function $\zeta(\cdot)$ is such that

- (a) No component of ζ_i is a constant and $E[\text{Var}[\zeta_i|X_i]]$ is nonsingular.
- (b) $\text{Var}[\zeta_i] < \infty$.
- (c) $E[\zeta_i|X_i = x]$, $E[\zeta_i \zeta_i'|X_i = x]$, $E[\zeta_i \tilde{Y}_i(a)|X_i = x]$, $E[\zeta_i D_i(a)|X_i = x]$ are Lipschitz.

[Theorem 4.2](#) shows that [Assumption 4.2](#) provides low-level sufficient conditions for [Assumption 4.1](#). In this way, it establishes the large-sample behavior of $\hat{\Delta}_n^{\text{adj}}$ for the special case of these working models and further verifies their optimality among all linear working models:

Theorem 4.2. *Suppose Q satisfies Assumption 2.1, the treatment assignment mechanism satisfies Assumptions 2.2–2.3, and in addition Assumption 4.2 is satisfied. Then,*

$$\begin{aligned}\hat{\beta}_n^Y &\xrightarrow{P} \beta^{\tilde{Y}} = (2E[\text{Var}[\zeta_i|X_i]])^{-1}E[\text{Cov}[\zeta_i, \tilde{Y}_i(1) + \tilde{Y}_i(0)|X_i]] \\ \hat{\beta}_n^D &\xrightarrow{P} \beta^D = (2E[\text{Var}[\zeta_i|X_i]])^{-1}E[\text{Cov}[\zeta_i, D_i(1) + D_i(0)|X_i]] .\end{aligned}$$

In addition, (16)–(17) and Assumption 4.1 are satisfied for $\hat{m}_{a,\tilde{Y}}(X_i, W_i) = \zeta_i' \hat{\beta}_n^Y$, $\hat{m}_{a,D}(X_i, W_i) = \zeta_i' \hat{\beta}_n^D$, $m_{1,\tilde{Y}}(X_i, W_i) = m_{0,\tilde{Y}}(X_i, W_i) = \zeta_i' \beta^{\tilde{Y}}$, and $m_{1,\bar{D}}(X_i, W_i) = m_{0,\bar{D}}(X_i, W_i) = \zeta_i' \beta^D$. Moreover, $\hat{\Delta}_n^{\text{adj}}$ is optimal in the sense of minimizing ν_{adj}^2 among all choices of $m_{a,D}$ and $m_{a,\tilde{Y}}$ that are linear in ζ_i .

Remark 4.1. Here, we present two equivalent expressions for $\hat{\Delta}_n^{\text{adj}}$ when $\hat{m}_{a,\tilde{Y}}(X_i, W_i) = \zeta_i' \hat{\beta}_n^Y$ and $\hat{m}_{a,D}(X_i, W_i) = \zeta_i' \hat{\beta}_n^D$. Denote by $\hat{\alpha}_n^Y$ and $\hat{\alpha}_n^D$ the ordinary least squares estimators of α^Y and α^D in (18)–(19). The Frisch-Waugh-Lovell theorem implies that

$$\begin{aligned}\hat{\alpha}_n^Y &= \hat{\psi}_n^{\text{adj}}(1) - \hat{\psi}_n^{\text{adj}}(0) \\ \hat{\alpha}_n^D &= \hat{\phi}_n^{\text{adj}}(1) - \hat{\phi}_n^{\text{adj}}(0) .\end{aligned}$$

It thus follows immediately that

$$\hat{\Delta}_n^{\text{adj}} = \frac{\hat{\alpha}_n^Y}{\hat{\alpha}_n^D} .$$

Next, denote by $\hat{\alpha}_n^{\text{IV}}$ estimator of α in the following linear regression

$$Y_i = \alpha D_i + \zeta_i' \beta + \sum_{1 \leq j \leq n} \theta_j I\{i \in \{\pi(2j-1), \pi(2j)\}\} + \epsilon_i . \quad (20)$$

obtained by applying two-stage least squares with A_i as an instrument for D_i . It can be shown that

$$\hat{\Delta}_n^{\text{adj}} = \hat{\alpha}_n^{\text{IV}} .$$

See Appendix B for details. We emphasize, however, that the usual heteroskedasticity-robust estimator of the variance of $\hat{\alpha}_n^{\text{IV}}$ is not consistent for the limiting variance derived in Theorem 4.1; Bai et al. (2023c) provide a related discussion of when such an estimator may even be invalid. We therefore construct a consistent estimator of the required variance in Theorem 4.3 below. ■

Finally, we construct a consistent variance estimator for the limiting variance ν_{adj}^2 . As noted in the discussion following Theorem 4.1, the expression for ν_{adj}^2 corresponds exactly to the limiting variance obtained in Bai et al. (2023a) with the usual potential outcomes $Y_i(a)$ replaced with the transformed outcomes $Y_i^*(a)$ and working models in Bai et al. (2023a) replaced with $m_{a,\tilde{Y}}(X_i, W_i) - \Delta(Q)m_{a,D}(X_i, W_i)$. We thus follow the variance construction from Bai et al. (2023a), but with a feasible version of $Y_i^*(a)$ defined as

$$\hat{Y}_i = Y_i - \hat{\Delta}_n D_i ,$$

and a feasible version of $m_{a,\bar{Y}}(X_i, W_i) - \Delta(Q)m_{a,D}(X_i, W_i)$ defined as

$$\hat{m}_{a,\bar{Y}D}(X_i, W_i) = \hat{m}_{a,\bar{Y}}(X_i, W_i) - \hat{\Delta}_n \hat{m}_{a,D}(X_i, W_i) .$$

This strategy leads to the following variance estimator:

$$\hat{\nu}_{n,\text{adj}}^2 = \frac{\hat{\tau}_{n,\text{adj}}^2 - \frac{1}{2}(\hat{\lambda}_{n,\text{adj}} + \hat{\Gamma}_{n,\text{adj}}^2)}{(\hat{\phi}_n^{\text{adj}}(1) - \hat{\phi}_n^{\text{adj}}(0))^2} , \quad (21)$$

where

$$\begin{aligned} \hat{\tau}_{n,\text{adj}}^2 &= \frac{1}{n} \sum_{1 \leq j \leq n} (\hat{Y}_{\pi(2j-1),\text{adj}} - \hat{Y}_{\pi(2j),\text{adj}})^2 , \\ \hat{\lambda}_{n,\text{adj}} &= \frac{2}{n} \sum_{1 \leq j \leq \lfloor \frac{n}{2} \rfloor} (\hat{Y}_{\pi(4j-3),\text{adj}} - \hat{Y}_{\pi(4j-2),\text{adj}}) \\ &\quad \times (\hat{Y}_{\pi(4j-1),\text{adj}} - \hat{Y}_{\pi(4j),\text{adj}})(A_{\pi(4j-3)} - A_{\pi(4j-2)})(A_{\pi(4j-1)} - A_{\pi(4j)}) , \\ \hat{\Gamma}_{n,\text{adj}} &= \frac{1}{n} \sum_{1 \leq i \leq n: A_i=1} \hat{Y}_{i,\text{adj}} - \frac{1}{n} \sum_{1 \leq i \leq n: A_i=0} \hat{Y}_{i,\text{adj}} , \\ \hat{Y}_{i,\text{adj}} &= \hat{Y}_i - \frac{1}{2}(\hat{m}_{1,\bar{Y}D}(X_i, W_i) + \hat{m}_{0,\bar{Y}D}(X_i, W_i)) . \end{aligned}$$

The following theorem establishes the consistency of $\hat{\nu}_{n,\text{adj}}^2$ for ν_{adj}^2 :

Theorem 4.3. *Suppose Q satisfies Assumption 2.1, the treatment assignment mechanism satisfies Assumptions 2.2-2.4, and $m_{a,\bar{Y}}$ and $m_{a,D}$ for $a \in \{0, 1\}$ satisfy Assumption 4.1. Further suppose $\hat{m}_{a,\bar{Y}}$ and $\hat{m}_{a,D}$ satisfy (16)-(17) and*

$$\frac{1}{2n} \sum_{1 \leq i \leq 2n} (\hat{m}_{a,\bar{Y}}(X_i, W_i) - m_{a,\bar{Y}}(X_i, W_i))^2 \xrightarrow{P} 0 , \quad (22)$$

$$\frac{1}{2n} \sum_{1 \leq i \leq 2n} (\hat{m}_{a,D}(X_i, W_i) - m_{a,D}(X_i, W_i))^2 \xrightarrow{P} 0 . \quad (23)$$

Then,

$$\hat{\nu}_{n,\text{adj}}^2 \xrightarrow{P} \nu_{\text{adj}}^2 .$$

5 Simulations

In this section, we examine the finite-sample behavior of the estimation and inference procedures introduced in Section 3. Following the simulation design in Bai et al. (2022), the potential outcomes are generated according to the equation:

$$Y_i(d) = \mu_d + m_d(X_i) + \sigma_d(X_i)\epsilon_{d,i},$$

where μ_d , $m_d(X_i)$, and $\epsilon_{d,i}$ are specified in each model below. In Models 1–3, we model compliance status as being independent of potential outcomes and covariates, with each of the following three compliance types being equally likely: compliers, always takers, and never takers. In Models 4–6, we allow the compliance status of an individual to depend on X_i , and define $C_i = 1$ if individual i is a complier, $AT_i = 1$ if individual i is an always taker, and $NT_i = 1$ if individual i is a never taker. We consider the following model specifications:

Model 1: $X_i \sim \text{Unif}[0, 1]$; $m_1(X_i) = m_0(X_i) = \gamma(X_i - \frac{1}{2})$; $\epsilon_{d,i} \sim N(0, 1)$ for $d \in \{0, 1\}$; $\sigma_0(X_i) = \sigma_0 = 1$, and $\sigma_1(X_i) = \sigma_1$. $\gamma = 1$, $\sigma_1 = 1$, $\mu_0 = 0$ and $\mu_1 = \Delta$, where $\Delta = 0$ is to study the behavior of the tests under the null hypothesis and $\Delta = 1$ is to study the behavior of the tests under the alternative hypothesis.

Model 2: As in Model 1, but $m_0(X_i) = 0$, and $m_1(X_i) = 10(X_i^2 - \frac{1}{3})$.

Model 3: As in Model 2, but $\sigma_0(X_i) = X_i^2$ and $\sigma_1(X_i) = \sigma_1 X_i^2$.

Model 4: As in Model 1, but $P\{C_i = 1|X_i\} = X_i$, and $P\{AT_i = 1|X_i\} = P\{NT_i = 1|X_i\} = \frac{1-X_i}{2}$.

Model 5: As in Model 4, but $m_0(X_i) = 0$, and $m_1(X_i) = 10(X_i^2 - \frac{1}{3} - \frac{1}{6})$.

Model 6: As in Model 5, but $\sigma_0(X_i) = X_i^2$ and $\sigma_1(X_i) = \sigma_1 X_i^2$.

Since $\dim(X_i) = 1$, we construct pairs by sorting units according to X_i and matching adjacent units. By Theorem 4.1 in [Bai et al. \(2022\)](#), this pairing algorithm ensures that both Assumptions 2.3 and 2.4 are satisfied. Table 1 reports the rejection probabilities from 5000 Monte Carlo replications, for testing the null hypothesis (7) using t -tests constructed using either the regression-based variance $\hat{\omega}_n^2$ (Robust and Robust-S, where Robust-S employs a standard small sample correction) or the consistent estimator $\hat{\nu}_n^2$ (MPIV). As expected given our theoretical results, the regression-based estimator $\hat{\omega}_n^2$ is conservative, which leads to a loss of power relative to our consistent estimator $\hat{\nu}_n^2$. Figure 1 displays the power curves for Models 1–6 with $2n = 1000$.

6 Empirical Application

In this section, we illustrate our findings by revisiting the empirical application in [Groh and McKenzie \(2016\)](#). [Groh and McKenzie \(2016\)](#) designed a matched-pair experiment in Egypt to study the effect on microenterprises of acquiring insurance against macroeconomic shocks.¹ The eligibility to purchase macroinsurance was offered to companies in the treatment group. The take-up rate of purchasing in the treatment group was 37%: among 1481 companies in the treatment group, 548 of them purchased insurance. We also note among 1480 companies in the control group, 5 of them purchased insurance as well.

¹To most closely align the dataset with our theoretical results, we made the following modifications to the dataset: (1) for each outcome variable, we drop pairs if at least one of the individuals in that pair has a missing outcome variable, (2) we drop pairs if at least one of the individuals in that pair is missing treatment assignment (the eligibility of purchasing the insurance), treatment status (whether a company actually purchased the insurance), or any baseline covariates, (3) we keep only pairs with exactly two individuals (there were 39 pairs with only one individual and one “block” with 16 individuals), (4) if necessary, we drop one pair from the end of the resulting dataset to ensure that the sample size is divisible by 4. (5) to construct the pairs of pairs when computing $\hat{\nu}_n$ and $\hat{\nu}_{n,adj}$, we use the R package `nbpMatching` to match pairs of pairs such that the conditions in Theorem 4.3 of [Bai et al. \(2022\)](#) are satisfied. Modifications (1)-(5) result in an average sample size reduction of 96 observations (3.29% of total sample size) across outcomes.

Model	Sample Size	$H_0 : \Delta = 0$			$H_1 : \Delta = 1$		
		Robust	Robust-S	MPIV	Robust	Robust-S	MPIV
1	200	2.72	2.64	3.36	60.76	60.46	63.36
	800	4.10	4.06	5.10	99.14	99.14	99.38
	1600	4.22	4.20	4.98	100.00	100.00	100.00
	3200	4.42	4.38	5.10	100.00	100.00	100.00
2	200	0.48	0.46	3.50	10.60	10.34	25.80
	800	0.78	0.78	4.50	53.28	53.20	73.70
	1600	1.30	1.30	5.00	87.20	87.16	95.10
	3200	1.14	1.14	5.14	99.52	99.52	99.96
3	200	0.36	0.34	3.14	11.46	11.14	33.50
	800	0.34	0.34	4.44	61.84	61.72	85.04
	1600	0.76	0.74	5.04	93.24	93.20	98.62
	3200	0.72	0.72	5.16	99.94	99.94	100.00
4	200	3.58	3.46	4.56	91.64	91.52	92.58
	800	4.10	4.08	4.96	100.00	100.00	100.00
	1600	4.06	4.06	5.04	100.00	100.00	100.00
	3200	4.00	4.00	4.84	100.00	100.00	100.00
5	200	1.36	1.30	4.80	20.80	20.42	41.58
	800	1.36	1.36	4.60	90.24	90.16	96.90
	1600	1.58	1.56	4.72	99.80	99.80	99.98
	3200	1.32	1.32	4.82	100.00	100.00	100.00
6	200	1.02	0.98	4.86	23.64	23.20	52.04
	800	0.94	0.92	4.94	95.70	95.70	99.48
	1600	1.02	1.02	5.20	100.00	100.00	100.00
	3200	0.78	0.78	4.38	100.00	100.00	100.00

Table 1: Rejection rates for Models 1, 2, 3, 4, 5, and 6

Table 2 reports the estimated LATEs for a collection of outcomes, using both the unadjusted Wald estimator $\hat{\Delta}_n$ as well as the linearly adjusted estimator $\hat{\Delta}_n^{\text{adj}}$ which uses the same covariates as those considered in the analysis in Groh and McKenzie (2016).² We also report the standard errors obtained from the regression-based variance estimator $\hat{\omega}_n^2$ as well as the standard errors obtained from our consistent variance estimators $\hat{\nu}_n^2$ and $\hat{\nu}_{n,\text{adj}}^2$ (note that we do not report the standard error obtained from the standard regression-based variance estimator when considering covariate adjustment, since as mentioned in Remark 4.1 this is not guaranteed to be valid). Our findings are consistent with the theoretical results presented in Sections 3 and 4: for the unadjusted estimates, standard errors constructed from $\hat{\nu}_n^2$ are smaller than those constructed from $\hat{\omega}_n^2$ across all outcomes, and in fact result in significance tests for the effect on profits and the aggregate index which reject at the 10% level. For the adjusted estimates, we find that the standard errors constructed from $\hat{\nu}_{n,\text{adj}}^2$ are smaller than those constructed from $\hat{\nu}_n^2$ across all outcomes. However, point estimates for profits and the aggregate index change in such a way that these are no longer significant at the 10% level.

²We note however that we exclude the female dummy and branchid dummies, which were used in the original regression specifications in Groh and McKenzie (2016), since these are perfectly collinear with our pair fixed effects.

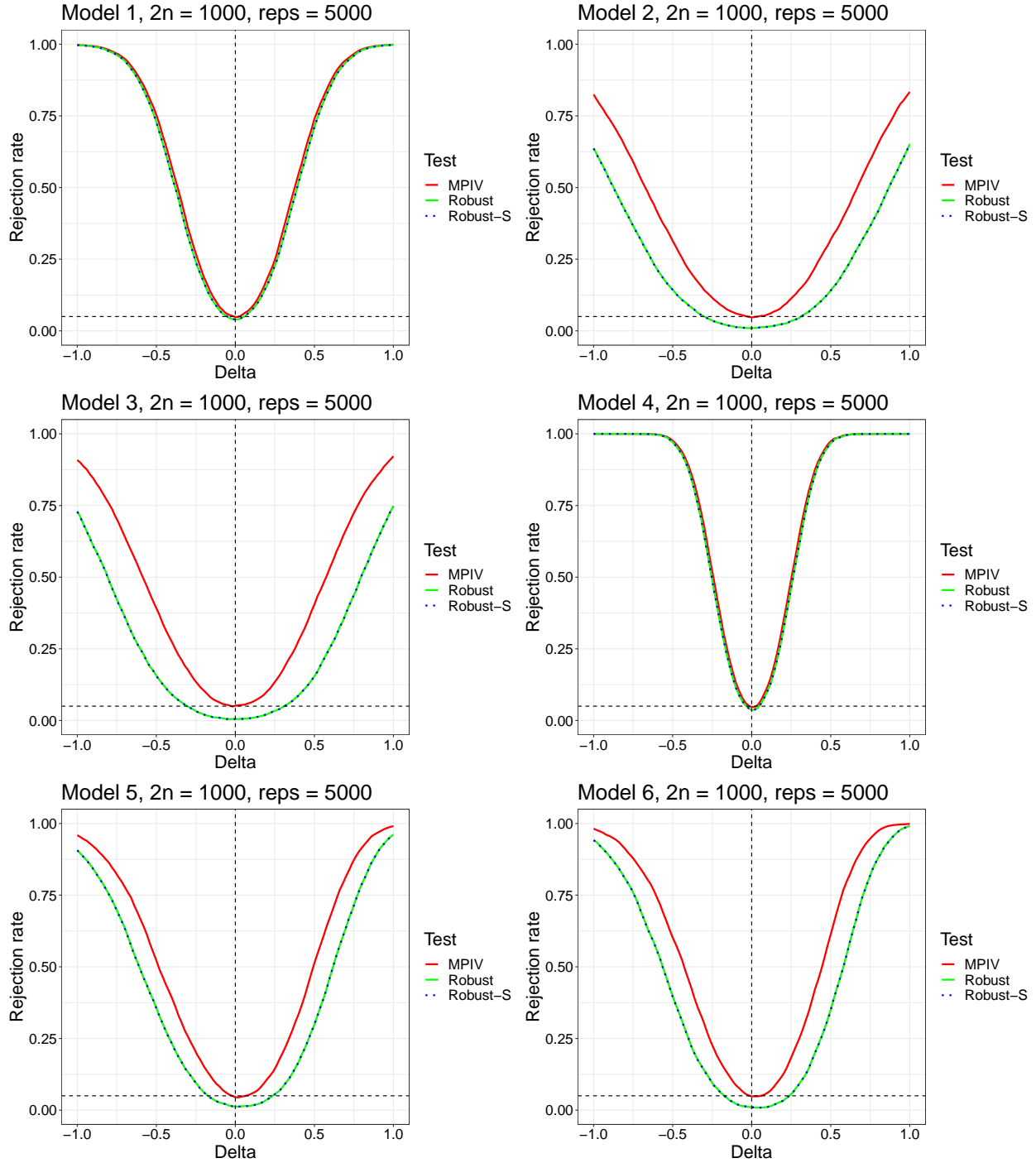


Figure 1: Power Curves

7 Recommendations for Empirical Practice

Based on our theoretical results as well as the simulation study above, we conclude with some recommendations for practitioners when conducting inference about the local average treatment effect in matched-pairs experiments. Our findings are that the standard Wald estimator is generally consistent and asymptotically

Table 2: Summary of Estimates Obtained from Empirical Application [Groh and McKenzie \(2016\)](#)

	Profits	High Profit	Revenue	High Revenue	Number Employees	Any Worker	Owner's Hours	Monthly Consumption	Aggregate Index
$\hat{\Delta}_n$	-241.602	-0.029	-2561.536	-0.063	-0.085	0.009	-1.324	-14.197	-0.104
SE from $\hat{\omega}_n^2$	(149.696)	(0.022)	(937.455)***	(0.021)***	(0.144)	(0.050)	(2.506)	(91.237)	(0.069)
SE from $\hat{\nu}_n^2$	(130.946)*	(0.021)	(841.562)***	(0.020)***	(0.138)	(0.049)	(2.179)	(79.273)	(0.061)*
$\hat{\Delta}_n^{\text{adj}}$	-151.345	-0.020	-1802.829	-0.052	-0.023	0.024	-0.401	6.541	-0.069
SE from $\hat{\nu}_{n,\text{adj}}^2$	(120.969)	(0.020)	(751.651)**	(0.018)***	(0.133)	(0.047)	(2.116)	(74.388)	(0.057)
Sample Size	2804	2804	2800	2800	2824	2824	2796	2880	2880

Notes: *: significant at 10% level. **: significant at 5% level. ***: significant at 1% level. For each outcome listed in Table 7 of [Groh and McKenzie \(2016\)](#), we report (a) the Wald estimates $\hat{\Delta}_n$ in (10), (b) the robust standard error obtained from $\hat{\omega}_n^2$ in (13), (c) the MPIV standard error obtained from $\hat{\nu}_n^2$ in (12), (d) the covariate-adjusted estimates $\hat{\Delta}_n^{\text{adj}}$ with pair fixed effects based on (18) and (19), (e) the standard error obtained from $\hat{\nu}_{n,\text{adj}}^2$ in (21), and (f) the sample sizes for the regression of each outcome variable.

normal under matched-pair designs, but its limiting variance is typically smaller than what would be obtained under i.i.d. assignment. It follows that inferences using the usual heteroskedasticity-robust estimator of the variance will typically be conservative. We therefore recommend that practitioners use our consistent variance estimator $\hat{\nu}_n^2$ instead. When considering covariate adjustment, our findings are that the two-stage least squares estimator with pair fixed effects leads to an estimator that is optimal in the sense of having smallest limiting variance among all linearly-adjusted estimators. An important caveat, however, is that the usual heteroskedasticity-robust estimator of the variance is not consistent for its variance. As a result, we recommend that practitioners use our consistent variance estimator $\hat{\nu}_{n,\text{adj}}^2$ instead.

A Proofs of Main Results

A.1 Proof of Theorem 3.1

PROOF.

$$\begin{aligned}\sqrt{n} \left(\hat{\Delta}_n - \Delta(Q) \right) &= \sqrt{n} \left(\frac{\hat{\psi}_n(1) - \hat{\psi}_n(0)}{\hat{\phi}_n(1) - \hat{\phi}_n(0)} - \Delta(Q) \right) \\ &= \frac{\sqrt{n} \left(\hat{\psi}_n(1) - \hat{\psi}_n(0) - \Delta(Q) \left(\hat{\phi}_n(1) - \hat{\phi}_n(0) \right) \right)}{\hat{\phi}_n(1) - \hat{\phi}_n(0)}.\end{aligned}$$

Following similar arguments to the proof of Lemma S.1.5 in [Bai et al. \(2022\)](#), we have

$$\hat{\phi}_n(1) - \hat{\phi}_n(0) \xrightarrow{P} P\{C_i = 1\}. \quad (24)$$

Therefore, to understand the limiting distribution of $\hat{\Delta}_n$, it suffices to show that

$$\sqrt{n}(\tilde{\psi}_n(1) - \tilde{\psi}_n(0)) \xrightarrow{d} N(0, \nu^2 P\{C_i = 1\}^2), \quad (25)$$

where

$$\tilde{\psi}_n(a) = \frac{1}{n} \sum_{1 \leq i \leq 2n: A_i = a} Y_i^*(a).$$

We may write

$$\sqrt{n} \left(\tilde{\psi}_n(1) - \tilde{\psi}_n(0) \right) = A_n - B_n + C_n - D_n,$$

where

$$\begin{aligned}A_n &= \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq 2n} \left(Y_i^*(1)A_i - E[Y_i^*(1)A_i | X^{(n)}, A^{(n)}] \right), \\ B_n &= \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq 2n} \left(Y_i^*(0)(1 - A_i) - E[Y_i^*(0)(1 - A_i) | X^{(n)}, A^{(n)}] \right), \\ C_n &= \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq 2n} \left(E[Y_i^*(1)A_i | X^{(n)}, A^{(n)}] - A_i E[Y_i^*(1)] \right), \\ D_n &= \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq 2n} \left(E[Y_i^*(0)(1 - A_i) | X^{(n)}, A^{(n)}] - (1 - A_i) E[Y_i^*(0)] \right).\end{aligned}$$

The decomposition holds because

$$\begin{aligned}E[Y_i^*(1)] - E[Y_i^*(0)] &= E[\tilde{Y}_i(1) - \Delta(Q)D_i(1)] - E[\tilde{Y}_i(0) - \Delta(Q)D_i(0)] \\ &= E[\tilde{Y}_i(1) - \tilde{Y}_i(0)] - \Delta(Q)E[D_i(1) - D_i(0)] \\ &= E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))] - \Delta(Q)E[D_i(1) - D_i(0)] \\ &= 0,\end{aligned} \quad (26)$$

where the first equation follows by (27), the second equation follows by inspection, the third equation follows by (4) two equations follow by inspection, and the last equation follows by Assumption 2.1-(d). Then the proof of (25) follows similarly to the proof of Lemma S.1.4 in Bai et al. (2022). Therefore, by Slutsky's theorem, the desired conclusion follows by (24) and (25) under Assumption 2.1-(e). ■

A.2 Proof of Theorem 3.2

PROOF. Recall that for $a \in \{0, 1\}$, we denote the adjusted potential outcome as

$$Y_i^*(a) = \tilde{Y}_i(a) - \Delta(Q)D_i(a), \quad (27)$$

and denote the (infeasible) adjusted observed outcome as

$$Y_i^* = Y_i - \Delta(Q)D_i. \quad (28)$$

With this notation, the observed adjusted outcome can be written as

$$Y^* = Y_i^*(1)A_i + Y_i^*(0)(1 - A_i). \quad (29)$$

First note that as a consequence of Lemma S.1.5 in Bai et al. (2022) and the continuous mapping theorem,

$$\left(\hat{\phi}_n(1) - \hat{\phi}_n(0)\right)^2 \xrightarrow{P} P\{C_i = 1\}^2.$$

It thus suffices to show that the numerator converges to the desired quantity.

Consider the following infeasible version of the numerator, given by

$$\tilde{\tau}_n^2 - \frac{1}{2}(\tilde{\lambda}_n^2 + \tilde{\Gamma}_n^2),$$

where

$$\begin{aligned} \tilde{\tau}_n^2 &= \frac{1}{n} \sum_{1 \leq j \leq n} (Y_{\pi(2j)}^* - Y_{\pi(2j-1)}^*)^2, \\ \tilde{\lambda}_n^2 &= \frac{2}{n} \sum_{1 \leq j \leq \lfloor \frac{n}{2} \rfloor} \left(Y_{\pi(4j-3)}^* - Y_{\pi(4j-2)}^* \right) \left(Y_{\pi(4j-1)}^* - Y_{\pi(4j)}^* \right) \left(A_{\pi(4j-3)} - A_{\pi(4j-2)} \right) \left(A_{\pi(4j-1)} - A_{\pi(4j)} \right), \\ \tilde{\Gamma}_n &= \frac{1}{n} \sum_{1 \leq i \leq 2n: A_i=1} Y_i^* - \frac{1}{n} \sum_{1 \leq i \leq 2n: A_i=0} Y_i^*. \end{aligned}$$

It follows immediately from Assumption 2.1 that (b)–(c) from Bai et al. (2022) are satisfied for the transformed outcomes $Y_i^*(a)$, and thus it follows from Lemmas S.1.5, S.1.6, and S.1.7 in Bai et al. (2022) that this infeasible numerator converges to the desired quantity. It thus remains to show that

$$\hat{\tau}_n^2 = \tilde{\tau}_n^2 + o_P(1) \quad (30)$$

$$\hat{\lambda}_n^2 = \tilde{\lambda}_n^2 + o_P(1) \quad (31)$$

$$\hat{\Gamma}_n = \tilde{\Gamma}_n + o_P(1) . \quad (32)$$

We begin with (30). To see this, note that

$$\begin{aligned} & \frac{1}{n} \sum_{1 \leq j \leq n} \left\{ (Y_{\pi(2j)}^* - Y_{\pi(2j-1)}^*)^2 - (\hat{Y}_{\pi(2j)} - \hat{Y}_{\pi(2j-1)})^2 \right\} \\ &= \frac{1}{n} \sum_{1 \leq j \leq n} \left\{ (Y_{\pi(2j)}^* - Y_{\pi(2j-1)}^*)^2 - (\hat{Y}_{\pi(2j)} - \hat{Y}_{\pi(2j-1)})^2 \right\} A_{\pi(2j)} \\ & \quad + \frac{1}{n} \sum_{1 \leq j \leq n} \left\{ (Y_{\pi(2j)}^* - Y_{\pi(2j-1)}^*)^2 - (\hat{Y}_{\pi(2j)} - \hat{Y}_{\pi(2j-1)})^2 \right\} A_{\pi(2j-1)} . \end{aligned}$$

It thus suffices to show that each component on the RHS converges in probability to zero. We only show the first since the second follows symmetrically. From the definitions of Y_i^* and \hat{Y}_i we obtain that

$$\begin{aligned} & \frac{1}{n} \sum_{1 \leq j \leq n} \left\{ (Y_{\pi(2j)}^* - Y_{\pi(2j-1)}^*)^2 - (\hat{Y}_{\pi(2j)} - \hat{Y}_{\pi(2j-1)})^2 \right\} A_{\pi(2j)} \\ &= \frac{1}{n} \sum_{1 \leq j \leq n} \left\{ (\tilde{Y}_{\pi(2j)}(1) - \tilde{Y}_{\pi(2j-1)}(0) - \Delta(Q)(D_{\pi(2j)}(1) - D_{\pi(2j-1)}(0)))^2 \right. \\ & \quad \left. - (\tilde{Y}_{\pi(2j)}(1) - \tilde{Y}_{\pi(2j-1)}(0) - \hat{\Delta}_n(D_{\pi(2j)}(1) - D_{\pi(2j-1)}(0)))^2 \right\} A_{\pi(2j)} \\ &= -2(\Delta(Q) - \hat{\Delta}_n) \frac{1}{n} \sum_{1 \leq j \leq n} \left\{ (\tilde{Y}_{\pi(2j)}(1) - \tilde{Y}_{\pi(2j-1)}(0))(D_{\pi(2j)}(1) - D_{\pi(2j-1)}(0)) \right\} A_{\pi(2j)} \\ & \quad + (\Delta(Q)^2 - \hat{\Delta}_n^2) \frac{1}{n} \sum_{1 \leq j \leq n} \left\{ (D_{\pi(2j)}(1) - D_{\pi(2j-1)}(0))^2 \right\} A_{\pi(2j)} \end{aligned}$$

Next, note that by the triangle inequality, Assumption 2.1(b) and the weak law of large numbers,

$$\begin{aligned} & \left| \frac{1}{n} \sum_{1 \leq j \leq n} \left\{ (\tilde{Y}_{\pi(2j)}(1) - \tilde{Y}_{\pi(2j-1)}(0))(D_{\pi(2j)}(1) - D_{\pi(2j-1)}(0)) \right\} A_{\pi(2j)} \right| \\ & \leq \frac{1}{n} \sum_{1 \leq i \leq 2n} |\tilde{Y}_i(1)| + \frac{1}{n} \sum_{1 \leq i \leq 2n} |\tilde{Y}_i(0)| = O_P(1) , \end{aligned}$$

and since D and A are binary,

$$\left| \frac{1}{n} \sum_{1 \leq j \leq n} \left\{ (D_{\pi(2j)}(1) - D_{\pi(2j-1)}(0))^2 \right\} A_{\pi(2j)} \right| \leq 1 ,$$

and hence the result follows from the fact that $\hat{\Delta}_n \xrightarrow{P} \Delta(Q)$ by Theorem 3.1. To show (31), note that

$$\begin{aligned} & \hat{\lambda}_n^2 - \tilde{\lambda}_n^2 \\ &= (\hat{\Delta}_n - \Delta(Q))^2 \frac{2}{n} \sum_{1 \leq j \leq \lfloor \frac{n}{2} \rfloor} (D_{\pi(4j-3)} - D_{\pi(4j-2)})(D_{\pi(4j-1)} - D_{\pi(4j)}) \\ & \quad \times (A_{\pi(4j-3)} - A_{\pi(4j-2)})(A_{\pi(4j-1)} - A_{\pi(4j)}) \end{aligned}$$

$$\begin{aligned}
& - (\hat{\Delta}_n - \Delta(Q)) \frac{2}{n} \sum_{1 \leq j \leq \lfloor \frac{n}{2} \rfloor} (D_{\pi(4j-3)} - D_{\pi(4j-2)})(Y_{\pi(4j-1)}^* - Y_{\pi(4j)}^*) \\
& \quad \times (A_{\pi(4j-3)} - A_{\pi(4j-2)})(A_{\pi(4j-1)} - A_{\pi(4j)}) \\
& - (\hat{\Delta}_n - \Delta(Q)) \frac{2}{n} \sum_{1 \leq j \leq \lfloor \frac{n}{2} \rfloor} (D_{\pi(4j-1)} - D_{\pi(4j)})(Y_{\pi(4j-3)}^* - Y_{\pi(4j-2)}^*) \\
& \quad \times (A_{\pi(4j-3)} - A_{\pi(4j-2)})(A_{\pi(4j-1)} - A_{\pi(4j)}) .
\end{aligned}$$

Note $\hat{\Delta}_n \xrightarrow{P} \Delta(Q)$ because of Theorem 3.1. On the other hand,

$$\begin{aligned}
& \left| \frac{2}{n} \sum_{1 \leq j \leq \lfloor \frac{n}{2} \rfloor} (D_{\pi(4j-3)} - D_{\pi(4j-2)})(Y_{\pi(4j-1)}^* - Y_{\pi(4j)}^*)(A_{\pi(4j-3)} - A_{\pi(4j-2)})(A_{\pi(4j-1)} - A_{\pi(4j)}) \right| \\
& \leq \frac{2}{n} \sum_{1 \leq j \leq \lfloor \frac{n}{2} \rfloor} (|Y_{\pi(4j-1)}^*| + |Y_{\pi(4j)}^*|) \\
& \leq \frac{2}{n} \sum_{1 \leq i \leq 2n} |Y_i^*(1)| + \frac{2}{n} \sum_{1 \leq i \leq 2n} |Y_i^*(0)| ,
\end{aligned}$$

where the first inequality follows from the triangle inequality and the fact that D and A are binary, and the last inequality follows trivially. And since D and A are binary,

$$\left| \frac{2}{n} \sum_{1 \leq j \leq \lfloor \frac{n}{2} \rfloor} (D_{\pi(4j-3)} - D_{\pi(4j-2)})(D_{\pi(4j-1)} - D_{\pi(4j)})(A_{\pi(4j-3)} - A_{\pi(4j-2)})(A_{\pi(4j-1)} - A_{\pi(4j)}) \right| \leq 1 .$$

(31) then follows because

$$\frac{2}{n} \sum_{1 \leq i \leq 2n} |Y_i^*(a)| = O_P(1)$$

for $a \in \{0, 1\}$ because of Assumption 2.1(b) and the weak law of large numbers. Finally, (32) follows immediately from $\hat{\Delta}_n \xrightarrow{P} \Delta(Q)$. ■

A.3 Proof of Theorem 3.3

PROOF. Following arguments similar to those used in the proof of Lemma S.1.5 in Bai et al. (2022), it can be shown that

$$\begin{aligned}
& \frac{1}{n} \sum_{1 \leq i \leq 2n} I\{A_i = a\} D_i \xrightarrow{P} E[D_i(a)] \\
& \frac{1}{n} \sum_{1 \leq i \leq 2n} I\{A_i = a\} Y_i^r \xrightarrow{P} E[\tilde{Y}_i^r(a)]
\end{aligned}$$

for $r = 1, 2$. Note in addition that

$$\hat{U}_i = Y_i - \frac{1}{2n} \sum_{1 \leq i \leq 2n} Y_i - \left(D_i - \frac{1}{2n} \sum_{1 \leq i \leq 2n} D_i \right) \hat{\Delta}_n .$$

It follows from direct calculation that

$$\hat{\omega}_n^2 = \frac{\frac{1}{n} \sum_{1 \leq i \leq 2n} \hat{U}_i^2}{\left(\frac{2}{n} \sum_{1 \leq i \leq 2n} A_i D_i - \frac{1}{n} \sum_{1 \leq i \leq 2n} D_i \right)^2}.$$

The conclusion then follows from the above derivations, the continuous mapping theorem, and additional direct calculations. ■

A.4 Proof of Theorem 4.1

PROOF. Note

$$\sqrt{n}(\hat{\Delta}_n^{\text{adj}} - \Delta(Q)) = \frac{\sqrt{n}(\hat{\psi}_n^{\text{adj}}(1) - \hat{\psi}_n^{\text{adj}}(0) - \Delta(Q)(\hat{\phi}_n^{\text{adj}}(1) - \hat{\phi}_n^{\text{adj}}(0)))}{\hat{\phi}_n^{\text{adj}}(1) - \hat{\phi}_n^{\text{adj}}(0)}.$$

Note by similar arguments to the proof of Theorem 3.1 in Bai et al. (2023a) and (17),

$$\hat{\phi}_n^{\text{adj}}(1) - \hat{\phi}_n^{\text{adj}}(0) \xrightarrow{P} P\{C_i = 1\}. \quad (33)$$

Therefore, to understand the limiting distribution of $\hat{\Delta}_n^{\text{adj}}$, it suffices to show that

$$\sqrt{n}(\tilde{\psi}_n^{\text{adj}}(1) - \tilde{\psi}_n^{\text{adj}}(0)) \xrightarrow{d} N(0, \nu_{1,\text{adj}}^2 + \nu_{2,\text{adj}}^2 + \nu_{3,\text{adj}}^2), \quad (34)$$

where

$$\begin{aligned} \tilde{\psi}_n^{\text{adj}}(a) &= \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2I\{A_i = a\}(Y_i^* - \hat{m}_{a,\tilde{Y}D}^*(X_i, W_i)) + \hat{m}_{a,\tilde{Y}D}^*(X_i, W_i)), \\ \hat{m}_{a,\tilde{Y}D}^*(X_i, W_i) &= \hat{m}_{a,\tilde{Y}}(X_i, W_i) - \Delta(Q)\hat{m}_{a,D}(X_i, W_i). \end{aligned}$$

First note that (16) and (17) imply that

$$\frac{1}{\sqrt{2n}} \sum_{1 \leq i \leq 2n} (2A_i - 1)(\hat{m}_{a,\tilde{Y}D}^*(X_i, W_i) - m_{a,\tilde{Y}D}(X_i, W_i)) \xrightarrow{P} 0, \quad (35)$$

then we have

$$\begin{aligned} \tilde{\psi}_n^{\text{adj}}(1) &= \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2A_i(Y_i^*(1) - \hat{m}_{1,\tilde{Y}D}^*(X_i, W_i)) + \hat{m}_{1,\tilde{Y}D}^*(X_i, W_i)) \\ &= \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2A_i Y_i^*(1) - (2A_i - 1)\hat{m}_{1,\tilde{Y}D}^*(X_i, W_i)) \\ &= \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2A_i Y_i^*(1) - (2A_i - 1)m_{1,\tilde{Y}D}(X_i, W_i)) + o_P(n^{-1/2}) \end{aligned}$$

$$= \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2A_i Y_i^*(1) - A_i m_{1, \tilde{Y}_D}(X_i, W_i) - (1 - A_i) m_{1, \tilde{Y}_D}(X_i, W_i)) + o_P(n^{-1/2}),$$

where the third equality follows from (35). Similarly,

$$\tilde{\psi}_n^{\text{adj}}(0) = \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2(1 - A_i) Y_i^*(0) - A_i m_{0, \tilde{Y}_D}(X_i, W_i) - (1 - A_i) m_{0, \tilde{Y}_D}(X_i, W_i)) + o_P(n^{-1/2}).$$

Then

$$\tilde{\psi}_n^{\text{adj}}(1) - \tilde{\psi}_n^{\text{adj}}(0) = \frac{1}{n} \sum_{1 \leq i \leq 2n} A_i \phi_{1,i}^* - \frac{1}{n} \sum_{1 \leq i \leq 2n} (1 - A_i) \phi_{0,i}^* + o_P(n^{-1/2}),$$

where

$$\begin{aligned} \phi_{1,i}^* &= Y_i^*(1) - \frac{1}{2}(m_{1, \tilde{Y}_D}(X_i, W_i) + m_{0, \tilde{Y}_D}(X_i, W_i)), \\ \phi_{0,i}^* &= Y_i^*(0) - \frac{1}{2}(m_{1, \tilde{Y}_D}(X_i, W_i) + m_{0, \tilde{Y}_D}(X_i, W_i)). \end{aligned}$$

Then we have the decomposition

$$\sqrt{n}(\tilde{\psi}_n^{\text{adj}}(1) - \tilde{\psi}_n^{\text{adj}}(0)) = A_n^{\text{adj}} - B_n^{\text{adj}} + C_n^{\text{adj}} - D_n^{\text{adj}} + o_P(n^{-1/2}),$$

where

$$\begin{aligned} A_n^{\text{adj}} &= \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq 2n} (A_i \phi_{1,i}^* - E[A_i \phi_{1,i}^* | X^{(n)}, A^{(n)}]), \\ B_n^{\text{adj}} &= \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq 2n} ((1 - A_i) \phi_{1,i}^* - E[(1 - A_i) \phi_{1,i}^* | X^{(n)}, A^{(n)}]), \\ C_n^{\text{adj}} &= \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq 2n} A_i (E[Y_i^*(1) | X_i] - E[Y_i^*(1)]), \\ D_n^{\text{adj}} &= \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq 2n} (1 - A_i) (E[Y_i^*(0) | X_i] - E[Y_i^*(0)]). \end{aligned}$$

This decomposition holds because of (26). Then the the proof of (34) follows similarly to the proof of Theorem 3.1 in Bai et al. (2023a). Therefore, by Slutsky's theorem, the desired conclusion follows by (33) and (34) under Assumption 2.1-(e).

Next, in order to see that $\hat{\Delta}_n^{\text{adj}}$ in Theorem 4.2 is the optimal linear adjustment, note that ν_{adj}^2 only depends on $m_{a, \tilde{Y}}(X_i, W_i)$ and $m_{a, D}(X_i, W_i)$ through $\nu_{1, \text{adj}}^2$. Then for arbitrary linear adjustments $m_{a, \tilde{Y}}(X_i, W_i) = \zeta_i' \tilde{\beta}^{\tilde{Y}}(a)$ and $m_{a, D}(X_i, W_i) = \zeta_i' \tilde{\beta}^D(a)$ for $a \in \{0, 1\}$, $\nu_{1, \text{adj}}^2$ can be re-written as

$$\begin{aligned} \nu_{1, \text{adj}}^2 &= \frac{1}{2} E[\text{Var}[E[Y_i^*(1) + Y_i^*(0) | X_i, W_i] - (m_{1, \tilde{Y}_D}(X_i, W_i) + m_{0, \tilde{Y}_D}(X_i, W_i)) | X_i]] \\ &= \frac{1}{2} E[\text{Var}[E[Y_i^*(1) + Y_i^*(0) | X_i, W_i] - \zeta_i' ((\tilde{\beta}^{\tilde{Y}}(1) + \tilde{\beta}^{\tilde{Y}}(0)) - \Delta(Q)(\tilde{\beta}^D(1) + \tilde{\beta}^D(0))) | X_i]] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2}E[E[(E[Y_i^*(1) + Y_i^*(0)|X_i, W_i] - E[Y_i^*(1) + Y_i^*(0)|X_i] - \\
&\quad (\zeta_i - E[\zeta_i|X_i])'((\tilde{\beta}^{\tilde{Y}}(1) + \tilde{\beta}^{\tilde{Y}}(0)) - \Delta(Q)(\tilde{\beta}^D(1) + \tilde{\beta}^D(0)))^2|X_i]] \\
&= \frac{1}{2}E[E[(E[Y_i^*(1) + Y_i^*(0)|X_i, W_i] - E[Y_i^*(1) + Y_i^*(0)|X_i] - \\
&\quad (\zeta_i - E[\zeta_i|X_i])'((\tilde{\beta}^{\tilde{Y}}(1) + \tilde{\beta}^{\tilde{Y}}(0)) - \Delta(Q)(\tilde{\beta}^D(1) + \tilde{\beta}^D(0)))^2],
\end{aligned}$$

which minimized when

$$(\beta^{\tilde{Y}}(1) + \beta^{\tilde{Y}}(0)) - \Delta(Q)(\beta^D(1) + \beta^D(0)) = (E[\text{Var}[\zeta_i|X_i]])^{-1}E[\text{Cov}[E[Y_i^*(1) + Y_i^*(0)|X_i, W_i], \zeta_i|X_i]] \quad (36)$$

$$= (E[\text{Var}[\zeta_i|X_i]])^{-1}E[\text{Cov}[Y_i^*(1) + Y_i^*(0), \zeta_i|X_i]], \quad (37)$$

where the first equality follows from the first order condition of minimizing $\nu_{1,\text{adj}}^2$ and the second equality follows by the fact that

$$\begin{aligned}
&E[\text{Cov}[E[Y_i^*(1) + Y_i^*(0)|X_i, W_i], \zeta_i|X_i] - \text{Cov}[Y_i^*(1) + Y_i^*(0), \zeta_i|X_i]] \\
&= E[\text{Cov}[E[Y_i^*(1) + Y_i^*(0)|X_i, W_i] - (Y_i^*(1) + Y_i^*(0)), \zeta_i|X_i]] \\
&= E[(E[Y_i^*(1) + Y_i^*(0)|X_i, W_i] - (Y_i^*(1) + Y_i^*(0)))\zeta_i] \\
&= E[E[(E[Y_i^*(1) + Y_i^*(0)|X_i, W_i] - (Y_i^*(1) + Y_i^*(0)))\zeta_i|X_i, W_i]] \\
&= E[(E[Y_i^*(1) + Y_i^*(0)|X_i, W_i] - E[Y_i^*(1) + Y_i^*(0)|X_i, W_i])\zeta_i] \\
&= 0,
\end{aligned}$$

where the fourth equality follows by the fact that ζ_i depends only on X_i and W_i , the rest of the equalities follows by inspection.

Finally, note that (36) holds when $\tilde{\beta}^{\tilde{Y}}(a) = \beta^{\tilde{Y}}$ and $\tilde{\beta}^D(a) = \beta^D$ for $a \in \{0, 1\}$, as desired. ■

A.5 Proof of Theorem 4.2

PROOF. The result follows from applying the arguments in the proof of Theorem 4.2 in Bai et al. (2023a) to $\hat{\beta}_n^Y$ and $\hat{\beta}_n^D$. ■

A.6 Proof of Theorem 4.3

PROOF. First note that (33) and continuous mapping theorem implies

$$\left(\hat{\phi}_n^{\text{adj}}(1) - \hat{\phi}_n^{\text{adj}}(0)\right)^2 \xrightarrow{P} P\{C_i = 1\}^2.$$

It thus suffices to show that the numerator converges to the desired quantity.

Consider the following infeasible version of the numerator, given by

$$\tilde{\tau}_{n,\text{adj}}^2 - \frac{1}{2}(\tilde{\lambda}_{n,\text{adj}}^2 + \tilde{\Gamma}_{n,\text{adj}}^2),$$

where

$$\tilde{\tau}_{n,\text{adj}}^2 = \frac{1}{n} \sum_{1 \leq j \leq n} (\hat{Y}_{\pi(2j),\text{adj}}^* - \hat{Y}_{\pi(2j-1),\text{adj}}^*)^2,$$

$$\tilde{\lambda}_{n,\text{adj}}^2 = \frac{2}{n} \sum_{1 \leq j \leq \lfloor \frac{n}{2} \rfloor} \left(\hat{Y}_{\pi(4j-3),\text{adj}}^* - \hat{Y}_{\pi(4j-2),\text{adj}}^* \right) \left(\hat{Y}_{\pi(4j-1),\text{adj}}^* - \hat{Y}_{\pi(4j),\text{adj}}^* \right) (A_{\pi(4j-3)} - A_{\pi(4j-2)}) (A_{\pi(4j-1)} - A_{\pi(4j)}),$$

$$\tilde{\Gamma}_{n,\text{adj}} = \frac{1}{n} \sum_{1 \leq i \leq 2n: A_i=1} \hat{Y}_{i,\text{adj}}^* - \frac{1}{n} \sum_{1 \leq i \leq 2n: A_i=0} \hat{Y}_{i,\text{adj}}^*,$$

$$\hat{Y}_{i,\text{adj}}^* = Y_i^* - \frac{1}{2}(\hat{m}_{1,\hat{Y}_D}^*(X_i, W_i) + \hat{m}_{0,\hat{Y}_D}^*(X_i, W_i)).$$

It then follows from the arguments in the proof of Theorem 3.2 in [Bai et al. \(2023a\)](#) that this infeasible numerator converges to the desired quantity. It thus remains to show that

$$\hat{\tau}_{n,\text{adj}}^2 = \tilde{\tau}_{n,\text{adj}}^2 + o_P(1) \tag{38}$$

$$\hat{\lambda}_{n,\text{adj}}^2 = \tilde{\lambda}_{n,\text{adj}}^2 + o_P(1) \tag{39}$$

$$\hat{\Gamma}_{n,\text{adj}} = \tilde{\Gamma}_{n,\text{adj}} + o_P(1). \tag{40}$$

The rest of the proof is similar to that of Theorem 3.2 and is omitted. ■

A.7 Auxiliary Results

Lemma A.1. ν^2 in Theorem 3.1 matches the expression of the efficiency bound in Theorem 2 of [Frölich \(2007\)](#).

PROOF: The efficiency bound in Theorem 2 of [Frölich \(2007\)](#) is

$$\begin{aligned} V &= \frac{1}{P\{C_i = 1\}^2} E \left[\frac{\text{Var}[Y_i|X_i, A_i = 1] - 2\Delta(Q) \text{Cov}[Y_i, D_i|X_i, A_i = 1] + \Delta(Q)^2 \text{Var}[D_i|X_i, A_i = 1]}{P\{A_i = 1|X_i\}} \right. \\ &\quad \left. + \frac{\text{Var}[Y_i|X_i, A_i = 0] - 2\Delta(Q) \text{Cov}[Y_i, D_i|X_i, A_i = 0] + \Delta(Q)^2 \text{Var}[D_i|X_i, A_i = 0]}{P\{A_i = 0|X_i\}} \right] \\ &\quad + \frac{1}{P\{C_i = 1\}^2} E[(E[Y_i|X_i, A_i = 1] - E[Y_i|X_i, A_i = 0] \\ &\quad \quad - \Delta(Q)E[D_i|X_i, A_i = 1] + \Delta(Q)E[D_i|X_i, A_i = 0])^2] \\ &= \frac{2}{P\{C_i = 1\}^2} E \left[\text{Var}[\tilde{Y}_i(1)|X_i] - 2\Delta(Q) \text{Cov}[\tilde{Y}_i(1), D_i(1)|X_i] + \Delta(Q)^2 \text{Var}[D_i(1)|X_i] \right. \\ &\quad \left. + \text{Var}[\tilde{Y}_i(0)|X_i] - 2\Delta(Q) \text{Cov}[\tilde{Y}_i(0), D_i(0)|X_i] + \Delta(Q)^2 \text{Var}[D_i(0)|X_i] \right] \\ &\quad + \frac{1}{P\{C_i = 1\}^2} E \left[\left(E[\tilde{Y}_i(1)|X_i] - E[\tilde{Y}_i(0)|X_i] - \Delta(Q)E[D_i(1)|X_i] + \Delta(Q)E[D_i(0)|X_i] \right)^2 \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{2}{P\{C_i = 1\}^2} E \left[\text{Var}[\tilde{Y}_i(1)|X_i] + \text{Var}[\tilde{Y}_i(0)|X_i] + \frac{1}{2} E^2[\tilde{Y}_i(1) - \tilde{Y}_i(0)|X_i] \right. \\
&\quad \left. - 2\Delta(Q) \left(\text{Cov}[\tilde{Y}_i(1), D_i(1)|X_i] + \text{Cov}[\tilde{Y}_i(0), D_i(0)|X_i] + \frac{1}{2} E[\tilde{Y}_i(1) - \tilde{Y}_i(0)|X_i] E[D_i(1) - D_i(0)|X_i] \right) \right. \\
&\quad \left. + \Delta(Q)^2 \left(\text{Var}[D_i(1)|X_i] + \text{Var}[D_i(0)|X_i] + \frac{1}{2} E^2[D_i(1) - D_i(0)|X_i] \right) \right] \\
&= \frac{2}{P\{C_i = 1\}^2} E \left[\text{Var}[Y_i^*(1)|X_i] + \text{Var}[Y_i^*(0)|X_i] + \frac{1}{2} E^2[Y_i^*(1) - Y_i^*(0)|X_i] \right] ,
\end{aligned}$$

where the first equality follows by Theorem 2 in Frölich (2007), the second equality follows by (4) and Assumption 2.2, the third equality follows by direct calculation, and the fourth equality follows by (11). Then we have

$$\begin{aligned}
&\frac{V}{2} - \nu^2 \\
&= \frac{1}{P\{C_i = 1\}^2} \left(E \left[\text{Var}[Y_i^*(1)|X_i] + \text{Var}[Y_i^*(0)|X_i] + \frac{1}{2} E^2[Y_i^*(1) - Y_i^*(0)|X_i] \right] \right. \\
&\quad \left. - \text{Var}[Y_i^*(1)] - \text{Var}[Y_i^*(0)] + \frac{1}{2} \text{Var}[E[Y_i^*(1) + Y_i^*(0)|X_i]] \right) \\
&= \frac{1}{2P\{C_i = 1\}^2} \left(-2 \text{Var}[E[Y_i^*(1)|X_i]] - 2 \text{Var}[E[Y_i^*(0)|X_i]] + E[E^2[Y_i^*(1) - Y_i^*(0)|X_i]] + \text{Var}[E[Y_i^*(1) + Y_i^*(0)|X_i]] \right) \\
&= \frac{1}{2P\{C_i = 1\}^2} \left(E[E^2[Y_i^*(1) - Y_i^*(0)|X_i]] - \text{Var}[E[Y_i^*(1) - Y_i^*(0)|X_i]] \right) \\
&= \frac{1}{2P\{C_i = 1\}^2} E^2[Y_i^*(1) - Y_i^*(0)] \\
&= 0 ,
\end{aligned}$$

where the first three equalities follow by inspection, and the fourth equation follows Assumption 2.1(e), and the fact that Assumption 2.1(d) implies $E[Y_i^*(1) - Y_i^*(0)] = 0$. ■

B Details for Remark 4.1

The subvector formula for IV says that

$$\hat{\alpha}_n^{\text{IV}} = \left(\frac{1}{n} \sum_{1 \leq i \leq 2n} \tilde{A}_i D_i \right)^{-1} \left(\frac{1}{n} \sum_{1 \leq i \leq 2n} \tilde{A}_i Y_i \right) ,$$

where \tilde{A}_i is the residual in the projection of A on ζ and fixed effects. Formally, consider the projection

$$A_i = \zeta_i' \gamma + \sum_{1 \leq j \leq n} \tau_j I\{i \in \{\pi(2j-1), \pi(2j)\}\} + U_i .$$

Let $\hat{\gamma}_n$ and $\hat{\tau}_{j,n}$ denote the OLS estimators of γ and τ_j . We first calculate $\hat{\gamma}_n$. To do so, use the subvector formula and note the residual of projection of $\zeta_{\pi(2j-1)}$ on the fixed effects is $(\zeta_{\pi(2j-1)} - \zeta_{\pi(2j)})/2$ and similarly

for $2j$. It then follows that

$$\hat{\gamma}_n = \left(\sum_{1 \leq j \leq n} (\zeta_{\pi(2j-1)} - \zeta_{\pi(2j)}) (\zeta_{\pi(2j-1)} - \zeta_{\pi(2j)})' \right)^{-1} \sum_{1 \leq j \leq n} (\zeta_{\pi(2j-1)} - \zeta_{\pi(2j)}) (A_{\pi(2j-1)} - A_{\pi(2j)}) .$$

Given this, it follows from the orthogonality condition

$$\sum_{1 \leq i \leq 2n} \left(A_i - \sum_{1 \leq i \leq 2n} \zeta_i' \hat{\gamma}_n - \sum_{1 \leq j \leq n} \hat{\tau}_{j,n} I\{i \in \{\pi(2j-1), \pi(2j)\}\} \right) I\{i \in \{\pi(2j-1), \pi(2j)\}\} = 0$$

that

$$\hat{\tau}_{j,n} = \frac{1}{2} - \frac{1}{2} (\zeta_{\pi(2j-1)} + \zeta_{\pi(2j)})' \hat{\gamma}_n .$$

Therefore,

$$\tilde{A}_{\pi(2j-1)} = \left(A_{\pi(2j-1)} - \frac{1}{2} \right) - \frac{1}{2} (\zeta_{\pi(2j-1)} - \zeta_{\pi(2j)})' \hat{\gamma}_n$$

and similarly for $\tilde{A}_{\pi(2j)}$. We then have

$$\begin{aligned} & \sum_{1 \leq i \leq 2n} \tilde{A}_i D_i \\ &= \frac{1}{2} \sum_{1 \leq i \leq 2n} (D_{\pi(2j-1)} - D_{\pi(2j)}) (A_{\pi(2j-1)} - A_{\pi(2j)}) - \frac{1}{2} \sum_{1 \leq j \leq n} (D_{\pi(2j-1)} - D_{\pi(2j)}) (\zeta_{\pi(2j-1)} - \zeta_{\pi(2j)})' \hat{\gamma}_n \\ &= \frac{1}{2} \sum_{1 \leq i \leq 2n} (D_i - \zeta_i' \hat{\beta}_n^D) (2A_i - 1) \end{aligned}$$

where the last step follows because

$$\sum_{1 \leq j \leq n} (D_{\pi(2j-1)} - D_{\pi(2j)}) (\zeta_{\pi(2j-1)} - \zeta_{\pi(2j)})' \hat{\gamma}_n = \hat{\beta}_n^{D'} \sum_{1 \leq j \leq n} (\zeta_{\pi(2j-1)} - \zeta_{\pi(2j)}) (A_{\pi(2j-1)} - A_{\pi(2j)}) ,$$

where $\hat{\beta}_n^D$ is the OLS estimator for β^D in (19) and equivalently that in the regression of pairwise difference of D on pairwise difference in ζ (see Section 4.2 in Bai et al. (2023a)). We therefore have $\hat{\alpha}_n^D = \sum_{1 \leq i \leq 2n} \tilde{A}_i D_i$ by direct calculation and similarly for $\hat{\alpha}_n^Y$.

References

- ANSEL, J., HONG, H. and JESSIE LI, A. (2018). Ols and 2sls in randomized and conditionally randomized experiments. *Jahrbücher für Nationalökonomie und Statistik*, **238** 243–293.
- ATHEY, S. and IMBENS, G. W. (2017). The econometrics of randomized experiments. In *Handbook of economic field experiments*, vol. 1. Elsevier, 73–140.
- BAI, Y., JIANG, L., ROMANO, J. P., SHAIKH, A. M. and ZHANG, Y. (2023a). Covariate adjustment in experiments with matched pairs. *arXiv preprint arXiv:2302.04380*.
- BAI, Y., LIU, J., SHAIKH, A. M. and TABORD-MEEHAN, M. (2023b). On the efficiency of finely stratified experiments.
- BAI, Y., LIU, J. and TABORD-MEEHAN, M. (2023c). Inference for Matched Tuples and Fully Blocked Factorial Designs. ArXiv:2206.04157 [econ, math, stat], URL <http://arxiv.org/abs/2206.04157>.
- BAI, Y., ROMANO, J. P. and SHAIKH, A. M. (2022). Inference in experiments with matched pairs. *Journal of the American Statistical Association*, **117** 1726–1737.
- BRUHN, M. and MCKENZIE, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics*, **1** 200–232.
- BUGNI, F. A. and GAO, M. (2021). Inference under covariate-adaptive randomization with imperfect compliance. *arXiv preprint arXiv:2102.03937*.
- CYTRYNBAUM, M. (2023). Covariate adjustment in stratified experiments.
- FRÖLICH, M. (2007). Nonparametric iv estimation of local average treatment effects with covariates. *Journal of Econometrics*, **139** 35–75.
- GLENNERSTER, R. and TAKAVARASHA, K. (2014). *Running randomized evaluations: A practical guide*. Princeton University Press.
- GROH, M. and MCKENZIE, D. (2016). Macroinsurance for microenterprises: A randomized experiment in post-revolution egypt. *Journal of Development Economics*, **118** 13–25.
- HONG, H. and NEKIPELOV, D. (2010). Semiparametric efficiency in nonlinear late models. *Quantitative Economics*, **1** 279–304.
- IMBENS, G. W. and ANGRIST, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62** 467–475.
- JIANG, L., LINTON, O. B., TANG, H. and ZHANG, Y. (2022). Improving estimation efficiency via regression-adjustment in covariate-adaptive randomizations with imperfect compliance. *arXiv preprint arXiv:2201.13004*.
- RESNJANSKIJ, S., RUHOSE, J., WIEDERHOLD, S. and WOESSMANN, L. (2021). Can mentoring alleviate family disadvantage in adolescence? a field experiment to improve labor-market prospects. *Working paper*.

RIACH, P. A. and RICH, J. (2002). Field experiments of discrimination in the market place. *The economic journal*, **112** F480–F518.

ROSENBERGER, W. F. and LACHIN, J. M. (2015). *Randomization in clinical trials: theory and practice*. John Wiley & Sons.