Asymptotic Theory of Rerandomization in Treatment-Control Experiments

Abstract

Although complete randomization ensures covariate balance on average, the chance for observing significant differences between treatment and control covariate distributions increases with many covariates. Rerandomization discards randomizations that do not satisfy a predetermined covariate balance criterion, generally resulting in better covariate balance and more precise estimates of causal effects. Previous theory has derived finite sample theory for rerandomization under the assumptions of equal treatment group sizes, Gaussian covariate and outcome distributions, or additive causal effects, but not for the general sampling distribution of the difference-in-means estimator for the average causal effect. To supplement existing results, we develop asymptotic theory for rerandomization without these assumptions, which reveals a non-Gaussian asymptotic distribution for this estimator, specifically a linear combination of a Gaussian random variable and a truncated Gaussian random variable. This distribution follows because rerandomization affects only the projection of potential outcomes onto the covariate space but does not affect the corresponding orthogonal residuals. We also demonstrate that, compared to complete randomization, rerandomization reduces the asymptotic sampling variances and quantile ranges of the difference-in-means estimator. Moreover, our work allows the construction of accurate large-sample confidence intervals for the average causal effect, thereby revealing further advantages of rerandomization over complete randomization.

Keywords: Causal inference; Covariate balance; Geometry of rerandomization; Mahalanobis distance; Quantile range; Tiers of covariates.

1. Introduction

Ever since Fisher (1925, 1926, 1935)'s seminal work, randomized experiments have become the "gold standard" for drawing causal inferences. Complete randomization balances the covariate distributions between treatment groups in expectation, thereby ensuring the existence of unbiased estimators of the average causal effect. Covariate imbalance, however, often occurs in specific randomized experiments, as recognized by Fisher (1926) and later researchers (e.g., Student 1938; Hansen and Bowers 2008; Keele et al. 2009; Bruhn and McKenzie 2009). The standard approach advocated by Fisher (1935), stratification or blocking (Cochran and Cox 1992), ensures balance with a few discrete covariates.

When a randomized allocation is unbalanced, it is reasonable to discard that allocation and re-draw another one until a certain pre-determined covariate balance criterion is satisfied. This is rerandomization, an experimental design hinted by R. A. Fisher (cf. Savage 1962, page 88) and Cox (1982, 2009), and formally proposed by Rubin (2008) and Morgan and Rubin (2012). Note that rerandomization is conceptually the same as restricted or constrained randomization (e.g., Yates 1948; Grundy and Healy 1950; Youden 1972; Bailey 1983). For more historical discussion, see Fienberg and Hinkley (1980, page 45), Speed (1992), Lehmann (2011, page 57), and Morgan and Rubin (2012).

Morgan and Rubin (2012) showed that the difference-in-means estimator is generally unbiased for the average causal effect under rerandomization with equal-sized treatment groups, and obtained the sampling variance of this estimator under additional assumptions of Gaussian covariate and outcome distributions and additive causal effects. When rerandomization is applied when these assumptions do not hold, statistical inference becomes more challenging, because the Gaussian distributional theory that is justified by the central

limit theorem under complete randomization (cf. Hájek 1960; Lin 2013) no longer generally holds. Some applied researchers believe that "the only analysis that we can be completely confident in is a permutation test or rerandomization test" (Bruhn and McKenzie 2009). However, randomization-based tests require sharp null hypotheses that all individual causal effects are known from observed values.

Analogous to the repeated sampling properties for complete randomization (Neyman 1923; Imbens and Rubin 2015), we evaluate the sampling properties of the difference-in-means estimator when rerandomization is used, where all potential outcomes and covariates are regarded as fixed quantities and all randomness arises solely from the random treatment assignments. The geometry of rerandomization reveals non-Gaussian asymptotic distributions, which serve as the foundation for constructing large-sample confidence intervals for average causal effects. Furthermore, we compare the lengths of quantile ranges of the asymptotic distributions of the difference-in-means estimator under rerandomization and complete randomization, extending Morgan and Rubin (2012, 2015)'s comparison of their sampling variances.

2. Framework, Notation, and Basic Results

2.1. Covariate imbalance and rerandomization

Inferring the causal effect of some binary treatment on an outcome Y is of central interest in many studies. We consider an experiment with n units, with n_1 assigned to treatment and n_0 assigned to control, $n = n_1 + n_0$, indexed by $i = 1, \ldots, n$. Before conducting the experiment, we collect K covariates $\boldsymbol{X}_i = (X_{1i}, X_{2i}, \ldots, X_{Ki})$ for each unit, which can possibly include transformations of basic covariates and their interactions. Let Z_i be the indicator variable for unit i assigned to treatment ($Z_i = 1$ if active treatment level; $Z_i = 0$ if the control level), and $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)'$ be the treatment assignment column vector. In a completely randomized experiment (CRE), the distribution of $\boldsymbol{Z}$ is such that each value,

$\boldsymbol{z} = (z_1, \ldots, z_n)'$, of $\boldsymbol{Z}$ has probability $n_1!n_0!/n!$, where $\sum_{i=1}^n z_i = n_1$ and $\sum_{i=1}^n (1 - z_i) = n_0$, which does not depend on the values of any observed or unobserved covariates. The difference-in-means vector of the covariates between treatment and control groups is

$$\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} = \frac{1}{n_1} \sum_{i=1}^n Z_i \boldsymbol{X}_i - \frac{1}{n_0} \sum_{i=1}^n (1 - Z_i) \boldsymbol{X}_i.$$

Although on average $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ has mean zero over all $\binom{n}{n_1}$ randomizations, for any realized value of $\boldsymbol{Z}$, imbalancedness in covariate distributions between treatment groups often occurs. As pointed out by Morgan and Rubin (2012), with 10 independent covariates and significance level 5%, the probability of a significant difference for at least one covariate is 40%.

When significant covariate imbalance arises in a drawn allocation, it is reasonable to discard the unlucky allocation and draw another treatment assignment vector until some a priori covariate balance criterion is satisfied. This is rerandomization, an intuitive experimental design tool apparently personally advocated by R. A. Fisher (see the discussion by Rubin 2008) and formally discussed by Morgan and Rubin (2012).

In general, rerandomization entails the following steps:

(1) collect covariate data;

(2) specify a balance criterion to determine whether a randomization is acceptable or not;

(3) randomize the units to treatment and control groups;

(4) if the balance criterion is satisfied, proceed to Step (5); otherwise, return to Step (3);

(5) conduct the experiment using the final randomization obtained in Step (4);

(6) analyze the data taking into account the rerandomization used in Steps (2)–(4).

Although the balance criterion in Step (2) can be general, Morgan and Rubin (2012) suggested using the Mahalanobis distance between covariate means in treatment and control groups, and Morgan and Rubin (2015) suggested considering tiers of covariates according to

their presumed importance in predicting the outcomes in this experiment. We will discuss these two types of rerandomization in detail, and apposite statistical inference after these rerandomizations as implied by Step (6).

2.2. Potential outcomes and definitions of finite population quantities

We use the potential outcomes framework (sometimes called the Rubin Causal Model; see Holland 1986; Imbens and Rubin 2015) to define causal effects, and let $Y_i(1)$ and $Y_i(0)$ denote the potential outcomes of unit i under active treatment and control, respectively. On the difference scale, the individual causal effect for unit i is $\tau_i = Y_i(1) - Y_i(0)$, and the average causal effect in the finite population of n units is $\tau_Y = \sum_{i=1}^n \tau_i/n$. Let $\bar{Y}(z) = \sum_{i=1}^n Y_i(z)/n$ be the finite population average of potential outcomes under treatment arm z, $\bar{\boldsymbol{X}}$ the finite population average of covariates, $S^2_{Y(z)}$ the finite population variance (with divisor $n-1$) of the potential outcomes under treatment arm z, $S_{Y(z),\boldsymbol{X}} = S'_{\boldsymbol{X},Y(z)}$ the finite population covariance between potential outcomes and covariates, and $\boldsymbol{S}^2_{\boldsymbol{X}}$ the finite population covariance matrix of covariates. For simplicity, we avoid notation for these quantities' dependence on n. Notice that these quantities are fixed, and are not dependent on the randomization or rerandomization scheme.

2.3. Repeated sampling inference in a CRE

The observed outcome for unit i is $Y_i = Z_i Y_i(1) + (1 - Z_i)Y_i(0)$, a function of treatment assignment and potential outcomes. In a CRE, Neyman (1923) showed that, for estimating τ_Y, the difference-in-means estimator

$$\hat{\tau}_Y = \frac{1}{n_1}\sum_{i=1}^n Z_i Y_i - \frac{1}{n_0}\sum_{i=1}^n (1 - Z_i)Y_i$$

is unbiased (the expectation of $\hat{\tau}_Y$ over all randomizations is τ_Y), and obtained its sampling variance over all randomizations for constructing a large-sample confidence interval for τ_Y. However, Neyman (1923)'s interval is not accurate if rerandomization is used, except in an asymptotic conservative sense.

Let $r_1 = n_1/n$ and $r_0 = n_0/n$ be the proportions of units receiving treatment and control. According to the finite population central limit theorem (Hájek 1960), under some regularity conditions, the large n sampling distribution, over all randomizations, of $\sqrt{n}(\hat{\tau}_Y - \tau_Y, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}')$ is Gaussian with mean zero and covariance matrix $\boldsymbol{V}$, where

$$\boldsymbol{V} = \begin{pmatrix} V_{\tau\tau} & \boldsymbol{V}_{\tau\boldsymbol{x}} \\ \boldsymbol{V}_{\boldsymbol{x}\tau} & \boldsymbol{V}_{\boldsymbol{xx}} \end{pmatrix} = \begin{pmatrix} r_1^{-1} S_{Y(1)}^2 + r_0^{-1} S_{Y(0)}^2 - S_\tau^2 & r_1^{-1} \boldsymbol{S}_{Y(1),\boldsymbol{X}} + r_0^{-1} \boldsymbol{S}_{Y(0),\boldsymbol{X}} \\ r_1^{-1} \boldsymbol{S}_{\boldsymbol{X},Y(1)} + r_0^{-1} \boldsymbol{S}_{\boldsymbol{X},Y(0)} & (r_1 r_0)^{-1} \boldsymbol{S}_{\boldsymbol{X}}^2 \end{pmatrix}.$$

Note again that we are conducting randomization-based inference, where all the covariates and potential outcomes are fixed numbers, and randomness comes solely from the treatment assignment. We embed n units into an infinite sequence of finite populations with increasing sizes, and a sufficient condition for the asymptotic Gaussianity of $\sqrt{n}(\hat{\tau}_Y - \tau_Y, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}')$ is as follows (Li and Ding 2016).

Condition 1. As $n \to \infty$, for $z = 0, 1$,

(i) r_z, the proportion of units under treatment arm z, has positive limits,

(ii) the finite population variances and covariances $S_{Y(z)}^2, S_\tau^2, \boldsymbol{S}_{\boldsymbol{X}}^2$ and $S_{\boldsymbol{X},Y(z)}$ have limiting values,

(iii) $\max_{1 \le i \le n} |Y_i(z) - \bar{Y}(z)|^2/n \to 0$ and $\max_{1 \le i \le n} \|\boldsymbol{X}_i - \bar{\boldsymbol{X}}\|_2^2/n \to 0$.

We introduce the notation $\dot{\sim}$ for two sequences of random vectors converging weakly to the same distribution. Therefore, under CRE and Condition 1, $\sqrt{n}(\hat{\tau}_Y - \tau_Y, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}') \dot{\sim} (A, \boldsymbol{B}')$, where $(A, \boldsymbol{B}')$ is a random vector from $\mathcal{N}(\boldsymbol{0}, \boldsymbol{V})$.

3. Rerandomization using the Mahalanobis distance

3.1. Mahalanobis distance

The Mahalanobis distance between the covariate means in treatment and control groups is

$$\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}'\{\mathrm{Var}(\hat{\boldsymbol{\tau}}_{\boldsymbol{X}})\}^{-1}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} = \left(\sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}\right)' \boldsymbol{V}_{\boldsymbol{xx}}^{-1} \left(\sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}\right),$$

recalling that $\boldsymbol{V}_{\boldsymbol{xx}} = (r_1 r_0)^{-1}\boldsymbol{S}_{\boldsymbol{X}}^2$ is a fixed and known $K \times K$ matrix in our finite population setting. A rerandomization scheme proposed by Morgan and Rubin (2012) accepts only those randomizations with the Mahalanobis distance less than or equal to a, a pre-specified threshold. Let

$$\mathcal{M} = \{\boldsymbol{\mu} : \boldsymbol{\mu}'\boldsymbol{V}_{\boldsymbol{xx}}^{-1}\boldsymbol{\mu} \leq a\}$$

denote the acceptance region for $\sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$; that is, a treatment assignment vector $\boldsymbol{Z}$ is accepted if and only if the corresponding $\sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \in \mathcal{M}$. Below we use ReM to denote rerandomization using this criterion.

Several practical issues with ReM are worth mentioning. First, if we include transformations and interactions of $\boldsymbol{X}$, then ReM can incorporate a wide class of rerandomization schemes. Second, for small sample sizes, it can be that there does not exist any randomization satisfying the balance criterion. However, according to the finite population central limit theorem, the acceptance probability of a randomization is asymptotically $p_a = P(\chi_K^2 \leq a)$. Therefore, for relatively large sample size, there usually exist many randomizations satisfying the balance criterion with $a > 0$. In practice, we would like to choose the asymptotic acceptance probability to be small, e.g., $p_a = 0.001$. However, we do not want p_a to be too small, such as accepting only those assignments with the smallest Mahalanobis distance. Too small p_a will result in few randomizations, making the repeated sampling inference intractable, even asymptotically, as well as the randomization tests powerless (Morgan and Rubin 2012).

Furthermore, as illustrated by later examples, the gain from reducing p_a usually decreases as p_a becomes smaller.

3.2. Multiple correlation between potential outcomes and covariates

We define the finite population squared multiple correlation between the potential outcome $Y(z)$ and the covariates $\boldsymbol{X}$ as $R^2(z)$ for $z = 1, 0$, and the finite population squared multiple correlation between the individual causal effect and the covariates as $R^2(\tau)$. Note that $R^2(1), R^2(0)$ and $R^2(\tau)$ are quantities of the finite population, which do not depend on the randomization or rerandomization scheme. Similar measures also appeared in Cochran (1965) and Rubin (1976).

We further define an R^2-type measure that is a function of the finite population quantities as well as the proportions of the group sizes:

$$R^2 = \frac{S^2_{Y(1)}}{r_1 V_{\tau\tau}} R^2(1) + \frac{S^2_{Y(0)}}{r_0 V_{\tau\tau}} R^2(0) - \frac{S^2_\tau}{V_{\tau\tau}} R^2(\tau).$$

When the causal effect is additive, $S^2_\tau = 0$ and $S^2_{Y(1)} = S^2_{Y(0)}$, and then $R^2 = R^2(1) = R^2(0)$ reduces to the squared multiple correlation between $\boldsymbol{X}$ and $Y(1)$ or $Y(0)$.

The following proposition states that under CRE R^2 is the proportion of the sampling variance of $\hat{\tau}_Y$ explained by $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ in linear projection.

Proposition 1. The sampling squared multiple correlation between $\hat{\tau}_Y$ and $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ under CRE is R^2, which can be equivalently written as

$$R^2 = \text{Corr}(\hat{\tau}_Y, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}) = \frac{r_1^{-1} S^2_{Y(1)|\boldsymbol{X}} + r_0^{-1} S^2_{Y(0)|\boldsymbol{X}} - S^2_{\tau|\boldsymbol{X}}}{r_1^{-1} S^2_{Y(1)} + r_0^{-1} S^2_{Y(0)} - S^2_\tau},$$

where $S^2_{Y(z)|\boldsymbol{X}}$ and $S^2_{\tau|\boldsymbol{X}}$ are the finite population variances of the linear projections of the potential outcomes and individual causal effects on covariates.

3.3. Asymptotic sampling distribution of $\hat{\tau}_Y$ under ReM

With rerandomization, we accept the randomizations satisfying the covariate balance criterion, and therefore the sampling distribution of $\sqrt{n}(\hat{\tau}_Y - \tau_Y)$ over rerandomizations is the same as its sampling distribution over a CRE conditional on $\sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ satisfying the covariate balance criterion. Although the following proposition holds for rerandomization with more general balance criteria, we first state it for ReM.

Proposition 2. Under ReM and Condition 1,

$$\begin{pmatrix} \sqrt{n}(\hat{\tau}_Y - \tau_Y) \\ \sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \end{pmatrix} \;\middle|\; \sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \in \mathcal{M} \;\; \dot{\sim} \;\; \begin{pmatrix} A \\ \boldsymbol{B} \end{pmatrix} \;\middle|\; \boldsymbol{B} \in \mathcal{M}, \tag{1}$$

recalling from earlier that $(A, \boldsymbol{B}')$ is a random vector following $\mathcal{N}(\boldsymbol{0}, \boldsymbol{V})$.

Simply stated, $\sqrt{n}(\hat{\tau}_Y - \tau_Y)$ has two parts: the part unrelated to the covariates, which we call ε_0, and thus unaffected by rerandomization, and the other part related to the covariates, which we call $L_{K,a}$, and thus affected by rerandomization. Therefore, the asymptotic distribution of $\hat{\tau}_Y$ is a linear combination of two independent random variables: $\varepsilon_0 \sim \mathcal{N}(0,1)$ is a standard Gaussian random variable, and $L_{K,a}$ is a random variable following the distribution of $D_1 \mid \boldsymbol{D}'\boldsymbol{D} \leq a$, where $\boldsymbol{D} = (D_1, \ldots, D_K)' \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_K)$.

Theorem 1. Under ReM and Condition 1,

$$\sqrt{n}(\hat{\tau}_Y - \tau_Y) \mid \sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \in \mathcal{M} \,\dot{\sim}\, \sqrt{V_{\tau\tau}} \left(\sqrt{1 - R^2} \cdot \varepsilon_0 + \sqrt{R^2} \cdot L_{K,a} \right), \tag{2}$$

where ε_0 is independent of $L_{K,a}$.

The coefficients of the linear combination are functions of R^2, which measures the association between the potential outcomes and the covariates. When $R^2 = 0$, the right hand side of (2) becomes a Gaussian random variable, the same as the asymptotic distribution of

$\sqrt{n}(\hat{\tau}_Y - \tau_Y)$ under CRE in Section 2.3; when $R^2 = 1$, (2) reduces to $\sqrt{V_{\tau\tau}} \cdot L_{K,a}$, a random variable with bounded support $[-\sqrt{aV_{\tau\tau}}, \sqrt{aV_{\tau\tau}}]$. Importantly, the definition of R^2 is based on linear projections but not linear models of the potential outcomes. Our asymptotic theory is based on the distribution of the randomization without imposing any modeling assumptions on the potential outcomes.

3.4. Representation and simulation of the asymptotic distribution under ReM

The asymptotic distribution in (2) involves a random variable $L_{K,a}$ that does not appear in standard statistical problems. Algebraically, $L_{K,a} \sim D_1 \mid \boldsymbol{D}'\boldsymbol{D} \leq a$ is the first coordinate of a K dimensional standard Gaussian vector, subject to the constraint that the squared length of the vector does not exceed a. This type of truncation of Gaussian distributions is apparently unstudied except for Tallis (1963) and Morgan and Rubin (2012). Because the standard Gaussian vector is spherically symmetric (Dempster 1969; Rubin 1976; Fang et al. 1989), it can be written as a product of two independent random components, a χ_K random variable and a random vector uniformly distributed on the $(K-1)$ dimensional unit sphere. The truncation condition, $\boldsymbol{D}'\boldsymbol{D} \leq a$, affects only the first component χ_K, leaving the second component unchanged. Basic properties of spherically symmetrical distributions allow us to represent $L_{K,a}$ using some known distributions, which allows for easy simulation of $L_{K,a}$.

Let $\chi^2_{K,a} \sim \chi^2_K \mid \chi^2_K \leq a$ be a truncated χ^2 random variable, U_K the first coordinate of the uniform random vector over the $(K-1)$ dimensional unit sphere, S a random sign taking ± 1 with probability $1/2$, and $\beta_K \sim \text{Beta}\,(1/2, (K-1)/2)$ a Beta random variable degenerating to a point mass at 1 when $K = 1$.

Proposition 3. $L_{K,a}$ can be represented as

$$L_{K,a} \sim D_1 \mid \boldsymbol{D}'\boldsymbol{D} \leq a \sim \chi_{K,a} U_K \sim \chi_{K,a} S \sqrt{\beta_K}, \tag{3}$$

where $(\chi_{K,a}, U_K)$ are mutually independent, and $(\chi_{K,a}, S, \beta_K)$ are mutually independent. $L_{K,a}$ is symmetric and unimodal around zero, with variance $\text{Var}(L_{K,a}) = v_{K,a} = P(\chi^2_{K+2} \leq a)/P(\chi^2_K \leq a) < 1$.

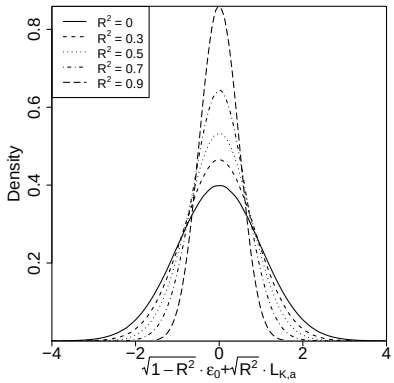
Because both ε_0 and $L_{K,a}$ are symmetric and both are unimodal at zero, their linear combination is also symmetric and unimodal at zero according to Wintner (1936)'s Theorem. The same is true for the asymptotic distribution of $\sqrt{n}(\hat{\tau}_Y - \tau_Y)$ in (2). The representation in (3) allows for easy simulation of $L_{K,a}$, as well as the asymptotic distribution of $\sqrt{n}(\hat{\tau}_Y - \tau_Y)$ in (2), which is relevant for statistical inference discussed later.

Without loss of generality, we fix $V_{\tau\tau}$ at 1, and consider the distribution of

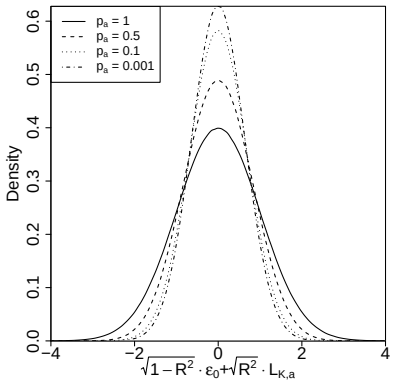
$$Q = \sqrt{1 - R^2} \cdot \varepsilon_0 + \sqrt{R^2} \cdot L_{K,a}, \tag{4}$$

which depends on R^2, the dimension of the covariates K, and the asymptotic acceptance probability of rerandomization $p_a = P(\chi^2_K \leq a)$. We simulate values of Q using independent and identically distributed (i.i.d) draws from (4). First, we fix $K = 10$ and $p_a = 0.001$. Figure 1a shows the probability densities of Q with different values of R^2, which approaches to that of $L_{K,a}$ as R^2 increases. Because ε_0 is more diffusely distributed than the truncated variable $L_{K,a}$, the probability density of Q will concentrate more around 0 with increasing R^2, as shown in Figure 1a.

Second, we fix $K = 3$ and $R^2 = 0.6$. Figure 1b shows the probability densities of Q with different values of asymptotic acceptance probability p_a; the CRE corresponds to $p_a = 1$. With smaller p_a, the distribution of Q becomes more concentrated around 0. Asymptotically, using smaller acceptance probabilities in ReM gives us more precise estimators for the average causal effect. However, when $R^2 < 1$, which is usually the case in practice, the gain of ReM by decreasing the threshold a becomes less as a becomes smaller. For example, the density of Q with $p_a = 0.0001$ is almost the same as the one with $p_a = 0.001$ in Figure 1b, and the percentage reduction in variance of Q achieved by decreasing p_a from 0.001 to 0.0001 is only



(a) $K = 10$ and $p_a = 0.001$



(b) $R^2 = 0.6$ and $K = 3$

Figure 1: Asymptotic distribution under ReM with $V_{\tau\tau}$ fixed at 1

5.7%.

3.5. Asymptotic unbiasedness, sampling variance and quantile ranges

First, the asymptotic distribution in (2) is symmetric around 0, implying that $\hat{\tau}_Y$ is asymptotically unbiased for τ_Y. Let $\mathbb{E}_{\mathrm{a}}(\cdot)$ and $\mathrm{Var}_{\mathrm{a}}(\cdot)$ denote the expectation and covariance matrix (or variance for scalar cases) of the asymptotic sampling distribution of a sequence of random vectors.

Corollary 1. Under ReM and Condition 1, $\mathbb{E}_{\mathrm{a}}\left\{\sqrt{n}(\hat{\tau}_Y - \tau_Y) \mid \sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \in \mathcal{M}\right\} = 0$.

Morgan and Rubin (2012) gave a counter-example showing that, in an experiment with unequal treatment group sizes, $\hat{\tau}_Y$ can be biased for τ_Y over ReM. As conjectured by Morgan and Rubin (2015), our result suggests that the bias is often small with large samples. Corollary 1 extends Morgan and Rubin (2012, Theorem 2.1) and ensures the asymptotic unbiasedness of $\hat{\tau}_Y$ for experiments with any ratio of group sizes. Corollary 1 also implies that any covariate asymptotically has the same means under treatment and control.

Furthermore, from Proposition 3 and Theorem 1, we can calculate the asymptotic sampling variances of $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ and $\hat{\tau}_Y$, and the percentage reductions in asymptotic sampling variances (PRIASV) under ReM compared to CRE. Recalling that $v_{K,a} = P(\chi^2_{K+2} \leq a)/P(\chi^2_K \leq a)$, we summarize the results below.

Corollary 2. Under ReM and Condition 1, the asymptotic sampling covariance of $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ is

$$\mathrm{Var}_{\mathrm{a}}\left(\sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \mid \sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \in \mathcal{M}\right) = v_{K,a}\boldsymbol{V}_{\boldsymbol{xx}},$$

and the PRIASV of any component of $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ is $1 - v_{K,a}$. The asymptotic sampling variance of $\hat{\tau}_Y$ is

$$\mathrm{Var}_{\mathrm{a}}\left\{\sqrt{n}(\hat{\tau}_Y - \tau_Y) \mid \sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \in \mathcal{M}\right\} = V_{\tau\tau}\left\{1 - (1 - v_{K,a})R^2\right\}, \tag{5}$$

and the PRIASV of $\hat{\tau}_Y$ is $(1 - v_{K,a})R^2$.

Note that the asymptotic sampling covariance and sampling variance of $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ and $\hat{\tau}_Y$ are actually the limits of $v_{K,a}\boldsymbol{V}_{\boldsymbol{xx}}$ and $V_{\tau\tau}\{1 - (1 - v_{K,a})R^2\}$ in the sequence of finite populations. However, for descriptive convenience, we omit these limit signs when discussing the expectation and covariance of asymptotic sampling distributions. When a is close to 0, or equivalently the asymptotic acceptance probability is small, the asymptotic sampling variance on the right hand side of (5) reduces to $V_{\tau\tau}(1-R^2)$, which is identical to the asymptotic sampling variance of the regression adjusted estimator under CRE discussed in Lin (2013) as an extension of Fisher (1925, 1935). Therefore, rerandomization does covariate adjustment in the design stage, and regression does covariate adjustment in the analysis stage. Cox (2009) and Morgan and Rubin (2012) discussed related issues.

When the causal effect is additive, R^2 is equal to the finite population squared multiple correlation between $\boldsymbol{X}$ and $Y(0)$. Therefore, Corollary 2 is an asymptotic version of Theorem 3.2 in Morgan and Rubin (2012).

Under ReM, in addition to the sampling variance reduction result concerning $\hat{\tau}_Y$ in Corollary 2, we consider the reduction in the length of the $(1-\alpha)$ quantile range of $\hat{\tau}_Y$ compared to that under CRE. We choose the length of the $(1-\alpha)$ quantile range, because of its connection to constructing confidence intervals as discussed shortly.

Let z_ξ be the ξth quantile of a standard Gaussian distribution. Let $\nu_\xi(R^2, p_a, K)$ be the

ξth quantile of the distribution of Q in (4). Note that $\nu_\xi(0, p_a, K) = z_\xi$. Because p_a and K are usually known by design, we write $\nu_\xi(R^2, p_a, K)$ as $\nu_\xi(R^2)$ for notational simplicity. Under ReM, the $(1-\alpha)$ quantile range of the asymptotic distribution of $\sqrt{n}(\hat{\tau}_Y - \tau_Y)$ is

$$\mathrm{QR}_\alpha(V_{\tau\tau}, R^2) = \left[\nu_{\alpha/2}(R^2)\sqrt{V_{\tau\tau}},\quad \nu_{1-\alpha/2}(R^2)\sqrt{V_{\tau\tau}}\right], \tag{6}$$

and the corresponding quantile range under CRE is

$$\mathrm{QR}_\alpha(V_{\tau\tau}, 0) = \left[z_{\alpha/2}\sqrt{V_{\tau\tau}},\quad z_{1-\alpha/2}\sqrt{V_{\tau\tau}}\right]. \tag{7}$$

Theorem 2. Under Condition 1, the length of the $(1-\alpha)$ quantile range of the asymptotic sampling distribution of $\sqrt{n}(\hat{\tau}_Y - \tau_Y)$ under ReM is less than or equal to that under CRE, with the difference nondecreasing in R^2.

3.6. Sampling variance estimation and confidence intervals

Asymptotic sampling variance and quantile range for $\hat{\tau}_Y$ depend on $V_{\tau\tau}$ and R^2, which are determined by the finite population covariances among potential outcomes and covariates. To obtain a sampling variance estimator and to construct an asymptotic confidence interval for τ_Y, we need to estimate these finite population variances and covariances. Let $s^2_{Y(z)}$, $s^2_{Y(z)|\boldsymbol{X}}$ and $\boldsymbol{s}_{Y(z),\boldsymbol{X}}$ be the sample variance of Y, sample variance of linear projection of Y on $\boldsymbol{X}$, and sample covariance of Y and $\boldsymbol{X}$ in treatment arm z. We show in the Supplementary Material that under ReM they are asymptotically unbiased for their population analogues $S^2_{Y(z)}, S^2_{Y(z)|\boldsymbol{X}}$ and $S_{Y(z),\boldsymbol{X}}$. Therefore, we can estimate $V_{\tau\tau}$ by (Ding et al. 2016)

$$\hat{V}_{\tau\tau} = r_1^{-1}s^2_{Y(1)} + r_0^{-1}s^2_{Y(0)} - (\boldsymbol{s}_{Y(1),\boldsymbol{X}} - \boldsymbol{s}_{Y(0),\boldsymbol{X}})(\boldsymbol{S}^2_{\boldsymbol{X}})^{-1}(\boldsymbol{s}_{\boldsymbol{X},Y(1)} - \boldsymbol{s}_{\boldsymbol{X},Y(0)}).$$

The estimator $\hat{V}_{\tau\tau}$ can be further improved using the Frechét–Hoeffding inequality (Aronow et al. 2014; Ding et al. 2016), but we omit the discussion here for simplicity. We then estimate

R^2 by

$$\hat{R}^2 = \hat{V}_{\tau\tau}^{-1} \left\{ r_1^{-1} s^2_{Y(1)|\boldsymbol{X}} + r_0^{-1} s^2_{Y(0)|\boldsymbol{X}} - \left(\boldsymbol{s}_{Y(1),\boldsymbol{X}} - \boldsymbol{s}_{Y(0),\boldsymbol{X}} \right) \left(\boldsymbol{S}^2_{\boldsymbol{X}} \right)^{-1} \left(\boldsymbol{s}_{\boldsymbol{X},Y(1)} - \boldsymbol{s}_{\boldsymbol{X},Y(0)} \right) \right\}. \quad (8)$$

We set $\hat{R}^2$ to be 0 if the estimator in (8) is negative.

According to (5), we can estimate the sampling variance of $\hat{\tau}_Y$ by $\hat{V}_{\tau\tau}\{1-(1-v_{K,a})\hat{R}^2\}/n$, and according to (6), we can construct a large sample $(1-\alpha)$ confidence interval for τ_Y using $\hat{\tau}_Y - \mathrm{QR}_\alpha(\hat{V}_{\tau\tau}, \hat{R}^2)/\sqrt{n}$. The sampling variance estimator is smaller than Neyman (1923)'s sampling variance estimator for CRE, and the confidence interval is shorter than Neyman (1923)'s confidence interval for CRE. Not surprisingly, unless the residual from the linear projection of individual causal effect on the covariates is constant, the above sampling variance estimator and confidence interval are both asymptotically conservative, in the sense that the probability limit of variance estimator is larger than or equal to the actual sampling variance, and the limit of coverage probability of confidence interval is larger than or equal to $(1-\alpha)$. Therefore, if we conduct ReM in the design stage but analyze data as in CRE, the consequential sampling variance estimator and confidence intervals will be overly conservative.

4. Rerandomization with tiers of covariates

4.1. Mahalanobis distance with tiers of covariates criterion

When covariates are thought to have different levels of importance for the outcomes, Morgan and Rubin (2015) proposed rerandomization using the Mahalanobis distance with differing criteria for different tiers of covariates. We partition the covariates into T tiers indexed by $t = 1, \ldots, T$ with decreasing importance, with k_t covariates in tier t. Let $\boldsymbol{X}_i = (\boldsymbol{X}_i[1], \ldots, \boldsymbol{X}_i[T])$, where $\boldsymbol{X}_i[t]$ denotes the covariates in tier t. Define $\boldsymbol{X}_i[\bar{t}] = (\boldsymbol{X}_i[1], \ldots, \boldsymbol{X}_i[t])$, the covariates in the first t tiers. Following the notation in Morgan and

Rubin (2015), we let $\boldsymbol{S}^2_{\boldsymbol{X}[\overline{t-1}]}$ be the finite population covariance matrix of the covariates in first $t-1$ tiers, and $\boldsymbol{S}_{\boldsymbol{X}[t],\boldsymbol{X}[\overline{t-1}]}$ be the finite population covariance matrix between $\boldsymbol{X}[t]$ and $\boldsymbol{X}[\overline{t-1}]$. We first apply a block-wise Gram–Schmidt orthogonalization to the covariates to create the orthogonalized covariates:

$$\begin{aligned}
\boldsymbol{E}_i[1] &= \boldsymbol{X}_i[1], \\
\boldsymbol{E}_i[t] &= \boldsymbol{X}_i[t] - \boldsymbol{S}_{\boldsymbol{X}[t],\boldsymbol{X}[\overline{t-1}]} \left(\boldsymbol{S}^2_{\boldsymbol{X}[\overline{t-1}]}\right)^{-1} \boldsymbol{X}_i[\overline{t-1}], \quad (2 \le t \le T)
\end{aligned}$$

where $\boldsymbol{E}_i[t]$ is the residual of the projection of the covariates $\boldsymbol{X}_i[t]$ in tier t onto the space spanned by the covariates in previous tiers; $\boldsymbol{E}_i = (\boldsymbol{E}_i[1], \ldots, \boldsymbol{E}_i[T])$. Let $\hat{\boldsymbol{\tau}}_{\boldsymbol{E}[t]}$ be the difference-in-means vector of $\boldsymbol{E}_i[t]$ between treatment and control groups, and $\boldsymbol{S}^2_{\boldsymbol{E}[t]}$ the finite population covariance matrix of $\boldsymbol{E}_i[t]$. The Mahalanobis distance in tier t is

$$M_t = \frac{n_1 n_0}{n} \hat{\boldsymbol{\tau}}'_{\boldsymbol{E}[t]} \left(\boldsymbol{S}^2_{\boldsymbol{E}[t]}\right)^{-1} \hat{\boldsymbol{\tau}}_{\boldsymbol{E}[t]},$$

and rerandomization using the Mahalanobis distance with tiers of covariates (ReMT) accepts those treatment assignments with $M_t \le a_t$, where a_t's are predetermined constants ($1 \le t \le T$). We can show that the criterion depends only on $\sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ and $\boldsymbol{V}_{\boldsymbol{xx}}$. If $T = 1$, then ReMT is simply ReM. We use $\mathcal{T}$ to denote the acceptance region for $\sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ under ReMT. The theory below extends Morgan and Rubin (2015) using the concepts from our Section 3.

4.2. Multiple correlation between potential outcomes and covariates with tiers

Similar to Section 3.2, we define the finite population squared multiple correlation between the potential outcome $Y(z)$ and the orthogonalized covariates in tier t as $\rho^2_t(z)$, and the finite population squared multiple correlation between the individual causal effect and the orthogonalized covariates in tier t as $\rho^2_t(\tau)$. We further define an R^2-type measure as the

function of these finite population quantities and the proportions of group sizes:

$$\rho_t^2 = \frac{S_{Y(1)}^2}{r_1 V_{\tau\tau}} \rho_t^2(1) + \frac{S_{Y(0)}^2}{r_0 V_{\tau\tau}} \rho_t^2(0) - \frac{S_\tau^2}{V_{\tau\tau}} \rho_t^2(\tau), \quad (1 \le t \le T)$$

which under the additive causal effect assumption reduces to $\rho_t^2 = \rho_t^2(1) = \rho_t^2(0)$, the squared multiple correlation between $\boldsymbol{E}[t]$ and $Y(1)$ or $Y(0)$.

Under CRE, ρ_t^2 is the sampling squared multiple correlation between $\hat{\tau}_Y$ and $\hat{\boldsymbol{\tau}}_{\boldsymbol{E}[t]}$, and can be equivalently written as

$$\rho_t^2 = \mathrm{Corr}(\hat{\tau}_Y, \hat{\boldsymbol{\tau}}_{\boldsymbol{E}[t]}) = \frac{r_1^{-1} S_{Y(1)|\boldsymbol{E}[t]}^2 + r_0^{-1} S_{Y(0)|\boldsymbol{E}[t]}^2 - S_{\tau|\boldsymbol{E}[t]}^2}{r_1^{-1} S_{Y(1)}^2 + r_0^{-1} S_{Y(0)}^2 - S_\tau^2}, \quad (1 \le t \le T)$$

where $S_{Y(z)|\boldsymbol{E}[t]}^2$ and $S_{\tau|\boldsymbol{E}[t]}^2$ are the finite population variances of the projections of the potential outcomes and individual causal effects on the orthogonalized covariates in tier t. For descriptive simplicity, we introduce $\rho_{T+1}^2 = 1 - \sum_{t=1}^T \rho_t^2 = 1 - R^2$ for later discussion.

4.3. Asymptotic distribution of $\hat{\tau}_Y$

The weak convergence of $\sqrt{n}(\hat{\tau}_Y - \tau_Y, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}')$ in (1) still holds for ReMT, with region $\mathcal{M}$ replaced by region $\mathcal{T}$. Intuitively, $\sqrt{n}(\hat{\tau}_Y - \tau_Y)$ can be decomposed into $(T+1)$ parts: the part unrelated to covariates and the T projections onto the space spanned by the orthogonalized covariates in T tiers. Due to the construction of the orthogonalized covariates, these $(T+1)$ parts are orthogonal to each other and the constraint for balance on the Mahalanobis distance in tier t affects only the t-th projection.

As earlier, let $\varepsilon_0 \sim \mathcal{N}(0,1)$, and extending earlier notation, let $L_{k_t,a_t} \sim D_{t1} \mid \boldsymbol{D}_t' \boldsymbol{D}_t \le a_t$, where $\boldsymbol{D}_t = (D_{t1}, \ldots, D_{tk_t}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{k_t})$ for $1 \le t \le T$.

Theorem 3. Under ReMT and Condition 1,

$$\sqrt{n}(\hat{\tau}_Y - \tau_Y) \mid \sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \in \mathcal{T} \quad \dot{\sim} \quad \sqrt{V_{\tau\tau}} \left(\rho_{T+1} \cdot \varepsilon_0 + \sum_{t=1}^T \rho_t \cdot L_{k_t,a_t} \right), \tag{9}$$

where $(\varepsilon_0, L_{k_1,a_1}, \ldots, L_{k_T,a_T})$ are mutually independent.

Obviously, in (9), ε_0 is the part of $\sqrt{n}(\hat{\tau}_Y - \tau_Y)$ that is unrelated to the covariates, and L_{k_t,a_t} is the part related to the orthogonalized covariates $\boldsymbol{E}_i[t]$ in tier t. According to Proposition 3, the distribution in Theorem 3 involves distributions that are easy to simulate.

4.4. Asymptotic unbiasedness, sampling variance and quantile ranges

First, the asymptotic distribution in (9) is symmetric around 0, implying that $\hat{\tau}_Y$ is asymptotically unbiased for τ_Y. Therefore, all observed or unobserved covariates have asymptotically balanced means.

Corollary 3. Under ReMT and Condition 1, $\mathbb{E}_{\mathrm{a}}\left\{\sqrt{n}(\hat{\tau}_Y - \tau_Y) \mid \sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \in \mathcal{T}\right\} = 0$.

The asymptotic sampling variance of $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ under ReMT has a complicated but conceptually obvious form, and we give it in the Supplementary Material. Below we present only the PRIASV of $\hat{\tau}_Y$; the PRIASVs for covariates are special cases of the same corollary because covariates are formally "outcomes" unaffected by the treatment. Recall the definition of $v_{k_t,a_t} = P(\chi^2_{k_t+2} \leq a_t)/P(\chi^2_{k_t} \leq a_t)$.

Corollary 4. Under ReMT and Condition 1, the asymptotic sampling variance of $\hat{\tau}_Y$ is

$$\mathrm{Var}_{\mathrm{a}}\left\{\sqrt{n}(\hat{\tau}_Y - \tau_Y) \mid \sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \in \mathcal{T}\right\} = V_{\tau\tau}\left\{1 - \sum_{t=1}^{T}(1 - v_{k_t,a_t})\rho_t^2\right\}, \tag{10}$$

and the PRIASV of $\hat{\tau}_Y$ is $\sum_{t=1}^{T}(1 - v_{k_t,a_t})\rho_t^2$.

When the causal effect is additive, ρ_t^2 becomes the finite population squared multiple correlation between $\boldsymbol{E}[t]$ and $Y(0)$. Therefore, Corollary 4 is an asymptotic extension of Morgan and Rubin (2015, Theorem 4.2). When the thresholds a_t's are close to zero, the asymptotic sampling variance on the right hand side of (10) reduces to $V_{\tau\tau}(1 - \sum_{t=1}^{T}\rho_t^2) = V_{\tau\tau}(1 - R^2)$, which is identical to that of the regression adjusted estimator under CRE (Lin 2013).

We now compare the quantile range under ReMT to that under CRE. Let $\nu_\xi(\rho_1^2, \rho_2^2, \ldots, \rho_T^2)$ be the ξth quantile of $\rho_{T+1}\varepsilon_0 + \sum_{t=1}^{T} \rho_t L_{k_t,a_t}$. Although $\nu_\xi(\rho_1^2, \rho_2^2, \ldots, \rho_T^2)$ depends also on p_{a_t} and k_t $(1 \le t \le K)$, we omit them to avoid notational clatter. The $(1-\alpha)$ quantile range of the asymptotic distribution of $\sqrt{n}(\hat{\tau}_Y - \tau_Y)$ under ReMT is

$$\mathrm{QR}_\alpha(V_{\tau\tau}, \rho_1^2, \ldots, \rho_T^2) = \left[\nu_{\alpha/2}(\rho_1^2, \ldots, \rho_T^2)\sqrt{V_{\tau\tau}}, \quad \nu_{1-\alpha/2}(\rho_1^2, \ldots, \rho_T^2)\sqrt{V_{\tau\tau}}\right]. \tag{11}$$

The stronger the correlation between the outcome and the orthogonalized covariates in tier t, the more reduction in quantile range we have when using ReMT rather than CRE. The following theorem is immediate.

Theorem 4. Under Condition 1, the $(1-\alpha)$ quantile range of the asymptotic distribution of $\sqrt{n}(\hat{\tau}_Y - \tau_Y)$ under ReMT is narrower than, or equal to the one under CRE, and the reduction in length of the quantile range is nondecreasing in ρ_t^2 for all $1 \le t \le T$.

4.5. Sampling variance estimation and confidence interval

We can estimate $V_{\tau\tau}$ and ρ_t^2 $(1 \le t \le T)$ in the same way as in ReM, and we estimate ρ_{T+1}^2 by $1 - \hat{R}^2$. In practice, we set $\hat{\rho}_t^2$ $(1 \le t \le T)$ to 0 when it is negative due to sampling variability, and standardize their sum to $\hat{R}^2$. According to (10) and (11), we can estimate the sampling variance of $\hat{\tau}_Y$ and $(1-\alpha)$ confidence intervals for τ_Y by replacing the unknown quantities with their point estimates. The sampling variance estimator is smaller than Neyman (1923)'s sampling variance estimator for CRE, and the confidence interval is shorter than Neyman (1923)'s confidence interval for CRE; both are asymptotically conservative in general, and only when the residual from the linear projection of individual causal effect on the covariates is constant, are they asymptotically exact. Therefore, analyzing data from ReMT as from CRE, the resulting sampling variance estimator and confidence intervals are overly conservative.

5. An education example with tiers of covariates

We illustrate our theory using the data from the Student Achievement and Retention Project (Angrist et al. 2009), a randomized evaluation of academic services and incentives at one of the satellite campuses of a large Canadian university, involving college freshmen. A treatment group of 150 students was offered an array of support services and substantial cash awards for meeting a target first year grade point average (GPA), and a control group of many more (1006) students received only standard university support services.

To illustrate the benefit of rerandomization, we use the 15 covariates as listed in Table 1, and exclude students with missing values, resulting in 118 students in the treatment group and 856 in the control. To make the simulation relevant to the real data, we fix unknown parameters based on some simple model fitting: We fit a linear regression of the observed first year GPA on the treatment indicator, all covariates and their interactions, and use the fitted model to generate all potential outcomes. To make the data generating process realistic, we simulate eight pseudo sets of potential outcomes using the fitted model with different choices for the variance of the residuals. The error terms for $Y(1)$ and $Y(0)$ are independent, and therefore conditional on the covariates, the potential outcomes are simulated as uncorrelated, but they have a positive correlation marginally. The final potential outcomes are truncated to lie on $[0, 4]$, mimicking the value of the GPA. We choose different variances of residuals such that the values of R^2 for the eight simulated data sets are located approximately evenly within interval $[0, 0.5]$. One choice for the variance of residuals is the one estimated from the fitted linear model, and the corresponding R^2 is about 0.23.

Table 1 partitions the covariates into three tiers with decreasing a priori importance to the outcome. As suggested by Morgan and Rubin (2015), for tiers with increasing numbers of covariates, we choose a_t such that $P(\chi^2_{k_t} \leq a_t) = (0.001)^{1/3} = 0.1$ for $t = 1, 2, 3$. We simulate data under ReMT, and obtain the confidence intervals based on our asymptotic theory for ReMT and Neyman (1923)'s results for CRE. Figure 2a shows the empirical coverage

Table 1: Covariates in the Student Achievement and Retention Project.

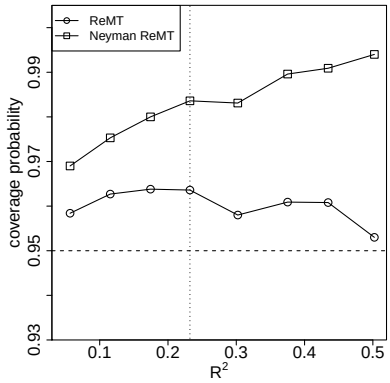
Tier	Covariates
1	high school GPA
2	whether lives at home, gender, age, whether rarely puts off studying for tests
3	whether mother is a college graduate, whether mother is a high school graduate, mother tongue (English or other), whether plans to work while in school, whether father is a college graduate, whether father is a high school graduate, whether never puts off studying for tests, whether wants more than a bachelor degree, whether intends to finish in 4 years, whether at the first choice school

probabilities of our and Neyman (1923)'s confidence intervals, showing that Neyman (1923)'s CRE confidence intervals are highly conservative. Note that there are 15 covariates and only 118 units in the treatment group, and the sample size is not extremely large. Despite this, our asymptotic confidence interval works well in this example.

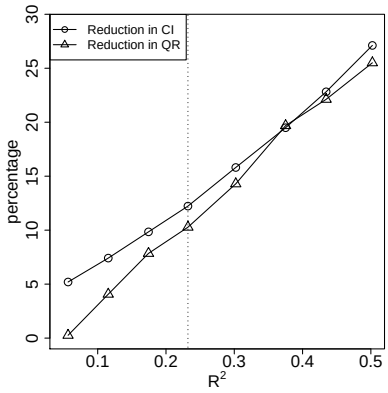
To evaluate the performance of ReMT compared to CRE, we compare the average length of Neyman (1923)'s confidence interval under CRE with the confidence interval under ReMT. From Figure 2b, the percentage reduction in average lengths of confidence intervals under ReMT compared to Neyman (1923)'s under CRE is nondecreasing in R^2. We also compare the empirical 95% quantile ranges of $\hat{\tau}_Y$ under ReMT and CRE, and the percentage reduction in the lengths of quantile ranges are close to the percentage reduction for average lengths of confidence intervals. When R^2 is close to that of the real data set (i.e. 0.23), the percentage increase in the effective sample size, that is, the sample size needed in CRE in order for $\hat{\tau}_Y$ to achieve the same 95% quantile range under ReMT, is about 24%. When R^2 is about twice as large as with the real data (i.e. 0.5), the percentage increase in the effective sample size increases to 80%.

6. Conclusions

Extending Morgan and Rubin (2012, 2015), we show that rerandomization balances covariates better than complete randomization, and provides a more precise difference-in-means



(a) Empirical coverage probabilities of our and Neyman (1923)'s 95% confidence intervals under ReMT



(b) Percentage reductions of average lengths of confidence intervals (◯) and quantile ranges (△) comparing ReMT with CRE

Figure 2: Eight data sets simulated based on the Student Achievement and Retention Project

estimator for the average causal effect. The asymptotic distributions of the difference-in-means estimator under rerandomization with strigent constraints are close to that of the regression adjusted estimator under CRE (Lin 2013), implying that rerandomization does the covariate adjustment in the design stage and avoids outcome modeling. The new asymptotic distributions allow us to construct confidence intervals for the average causal effect, when the classical Neyman (1923)'s inference for CRE is overly conservative.

REFERENCES

J. Angrist, D. Lang, and P. Oreopoulos. Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1: 136–163, 2009.

P. M. Aronow, D. P. Green, and D. K. K. Lee. Sharp bounds on the variance in randomized experiments. *The Annals of Statistics*, 42:850–871, 2014.

R. A. Bailey. Restricted randomization. *Biometrika*, 70:183–198, 1983.

M. Bruhn and D. McKenzie. In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, 1:200–232, 2009.

W. G. Cochran. The planning of observational studies of human populations. *Journal of the Royal Statistical Society. Series A (General)*, 128:234–266, 1965.

William G. Cochran and Gertrude M. Cox. *Experimental Designs*. John Wiley & Sons, Inc., 2 edition, 1992.

D. R. Cox. Randomization and concomitant variables in the design of experiments. In P. R. Krishnaiah G. Kallianpur and J. K. Ghosh, editors, *Statistics and Probability: Essays in Honor of C. R. Rao*, pages 197–202. North-Holland, Amsterdam, 1982.

D. R. Cox. Randomization in the design of experiments. *International Statistical Review*, 77:415–429, 2009.

A. P. Dempster. *Elements of Continuous Multivariate Analysis*. Addison-Wesley, Reading, Massachusetts, 1969.

P. Ding, A. Feller, and L. Miratrix. Decomposing treatment effect variation. *arXiv preprint arXiv:1605.06566*, 2016.

K. T. Fang, S. Kotz, and K. W. Ng. *Symmetric Multivariate and Related Distributions*. Chapman and Hall/CRC, 1989.

S. E. Fienberg and D. V. Hinkley. *R. A. Fisher : An Appreciation*. New York : Springer-Verlag, 1980.

R. A. Fisher. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd, 1st edition, 1925.

R. A. Fisher. The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain*, 33:503–513, 1926.

R. A. Fisher. *The Design of Experiments, 1st Edition*. Edinburgh, London: Oliver and Boyd, 1935.

P. M. Grundy and M. J. R. Healy. Restricted randomization and quasi-latin squares. *Journal of the Royal Statistical Society, Series B (Methodological)*, 12:286–291, 1950.

J. Hájek. Limiting distributions in simple random sampling from a finite population. *Publications of the Mathematics Institute of the Hungarian Academy of Science*, 5:361–74, 1960.

B. B. Hansen and J. Bowers. Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, 23:219–236, 2008.

P. W. Holland. Statistics and causal inference. *Journal of the American statistical Association*, 81:945–960, 1986.

G. W. Imbens and D. B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press, 2015.

L. Keele, C. McConnaughy, I. White, P. M. E. M. List, and D. Bailey. Adjusting experimental data. In *Experiments in Political Science Conference*. 2009.

E. L. Lehmann. *Fisher, Neyman, and the Creation of Classical Statistics*. New York: Springer, 2011.

X. Li and P. Ding. General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association*, in press, 2016.

W. Lin. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *The Annals of Applied Statistics*, 7:295–318, 2013.

K. L. Morgan and D. B. Rubin. Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 40:1263–1282, 2012.

K. L. Morgan and D. B. Rubin. Rerandomization to balance tiers of covariates. *Journal of the American Statistical Association*, 110:1412–1421, 2015.

J. Neyman. On the application of probability theory to agricultural experiments. essay on principles (with discussion). section 9 (translated). reprinted ed. *Statistical Science*, 5: 465–472, 1923.

D. B. Rubin. Multivariate matching methods that are equal percent bias reducing, I: some examples. *Biometrics*, 32:109–120, 1976.

D. B. Rubin. Comment to W. R. Shadish, M. H. Clark and P. M. Steiner. *Journal of the American Statistical Association*, 103:1350–1353, 2008.

L. J. Savage. *The Foundations of Statistical Inference*. Methuen and Co. Led., London, 1962.

T. P. Speed. Introduction to Fisher (1926). In S. Kotz and N. L. Johnson, editors, *Breakthroughs in Statistics*, pages 71–81. Springer, 1992.

Student. Comparison between balanced and random arrangements of field plots. *Biometrika*, 29:363–378, 1938.

G. M. Tallis. Elliptical and radial truncation in normal populations. *The Annals of Mathematical Statistics*, 34:940–944, 1963.

A. Wintner. On a class of Fourier transforms. *American Journal of Mathematics*, 58:45–90, 1936.

F. Yates. Comment to F. J. Anscombe. *Journal of the Royal Statistical Society, Series A (General)*, 111:204–205, 1948.

W. J. Youden. Randomization and experimentation. *Technometrics*, 14:13–22, 1972.