ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2020)
82, *Part* 1, *pp.* 241–268

Rerandomization and regression adjustment

Xinran Li

University of Illinois at Urbana—Champaign, USA

and Peng Ding

University of California, Berkeley, USA

[Received August 2018. Final revision October 2019]

Summary. Randomization is a basis for the statistical inference of treatment effects without strong assumptions on the outcome-generating process. Appropriately using covariates further yields more precise estimators in randomized experiments. R. A. Fisher suggested blocking on discrete covariates in the design stage or conducting analysis of covariance in the analysis stage. We can embed blocking in a wider class of experimental design called rerandomization, and extend the classical analysis of covariance to more general regression adjustment. Rerandomization trumps complete randomization in the design stage, and regression adjustment trumps the simple difference-in-means estimator in the analysis stage. It is then intuitive to use both rerandomization and regression adjustment. Under the randomization inference framework, we establish a unified theory allowing the designer and analyser to have access to different sets of covariates. We find that asymptotically, for any given estimator with or without regression adjustment, rerandomization never hurts either the sampling precision or the estimated precision, and, for any given design with or without rerandomization, our regression-adjusted estimator never hurts the estimated precision. Therefore, combining rerandomization and regression adjustment yields better coverage properties and thus improves statistical inference. To quantify these statements theoretically, we discuss optimal regression-adjusted estimators in terms of the sampling precision and the estimated precision, and then measure the additional gains of the designer and the analyser. We finally suggest the use of rerandomization in the design and regression adjustment in the analysis followed by the Huber–White robust standard error.

Keywords: Covariate balance; Experimental design; Potential outcome; Randomization

1. Introduction

In his seminal book *Design of Experiments*, Fisher (1935) first formally discussed the value of randomization in experiments: randomization balances observed and unobserved covariates on average and serves as a basis for statistical inference. Since then, randomized experiments have been widely used in agricultural sciences (e.g. Fisher (1935) and Kempthorne (1952)), industry (e.g. Box *et al.* (2005) and Wu and Hamada (2011)) and clinical trials (e.g. Rosenberger and Lachin (2015)). Recent years have witnessed the popularity of using randomized experiments in social sciences (e.g. Duflo *et al.* (2007), Gerber and Green (2012) and Athey and Imbens (2017)) and technology companies (e.g. Kohavi and Longbotham (2017)). Those modern applications often have richer covariates.

In completely randomized experiments, covariate imbalance often occurs by chance. Fisher (1935) proposed the use of analysis of covariance (ANCOVA) to adjust for covariate imbalance

Address for correspondence: Peng Ding, Department of Statistics, University of California, Berkeley, 425 Evans Hall, Berkeley, CA 94720, USA.
E-mail: pengdingpku@berkeley.edu

 1369–7412/20/82241

and thus to improve estimation efficiency. Fisher's (1935) ANCOVA uses the coefficient of the treatment in the ordinary least squares (OLS) fit of the outcome on the treatment and covariates. However, Freedman (2008a, b) criticized ANCOVA by showing that it can be even less efficient than the simple difference-in-means estimator under Neyman's (1923) potential outcomes framework. Freedman's (2008a, b) analyses allowed for treatment effect heterogeneity, in contrast with the existing literature on ANCOVA which often assumed additive treatment effects (Fisher, 1935; Kempthorne, 1952; Cox and Reid, 2000). Lin (2013) proposed a solution to Freedman's critique by running the OLS regression of the outcome on the treatment, covariates and their *interactions*. Li and Ding (2017) showed the 'optimality' of Lin's (2013) estimator within a class of regression-adjusted estimators.

Aware of the covariate imbalance issue, Fisher (1926) also proposed a strategy to avoid it actively in experiments. With a few discrete covariates, he proposed to use blocking, i.e. to conduct completely randomized experiments (CREs) within blocks of covariates. This remains a powerful tool in modern experiments (Miratrix *et al.*, 2013; Higgins *et al.*, 2016; Athey and Imbens, 2017). Blocking is a special case of rerandomization (Morgan and Rubin, 2012), which rejects 'bad' random allocations that violate certain covariate balance criteria. Rerandomization can also deal with more general covariates. Morgan and Rubin (2012) demonstrated that rerandomization improves covariate balance. Li *et al.* (2018) further derived the asymptotic distribution of the difference-in-means estimator and demonstrated that rerandomization improves its precision compared with complete randomization.

Rerandomization and regression adjustment are two ways to use covariates to improve efficiency. The former uses covariates in the design stage, and the latter uses covariates in the analysis stage. It is then natural to combine them in practice, i.e. to conduct rerandomization in the design and to use regression adjustment in the analysis. Several theoretical challenges remain. First, how do we conduct statistical inference? We shall derive the asymptotic distribution of the regression-adjusted estimator under rerandomization without assuming any outcome-generating model. Our theory is purely randomization based, in which potential outcomes are fixed numbers and the only randomness comes from the treatment allocation.

Second, what is the optimal regression adjustment under rerandomization? The optimality depends on the criterion. We shall introduce two notions of optimality: one based on the *sampling precision* and the other based on the *estimated precision*. Because our general theory allows the designer and analyser to have different sets of covariates, it is possible that the estimated precision differs from the sampling precision asymptotically, even under the case with additive treatment effects. We shall show that, asymptotically, rerandomization never hurts either the sampling precision or the estimated precision, and Lin's (2013) regression adjustment never hurts the estimated precision. Therefore, combining rerandomization and regression adjustment improves the coverage properties of the associated confidence intervals. On the basis of these findings, we suggest the use of Lin's (2013) estimator in general settings and show that the Huber–White variance estimator is a convenient approximation to its variance under rerandomization. Importantly, our theory does not rely on the linear model assumption.

Third, how do we quantify the gains from the designer and analyser? In particular, if the analyser uses an optimal regression adjustment, what is the additional gain of rerandomization compared with complete randomization? If the designer uses rerandomization, what is the additional gain of using an optimal regression adjustment compared with the simple difference in means? Our theory can quantitatively answer these questions.

This paper proceeds as follows. Section 2 introduces the framework and notation. Section 3 derives the sampling distribution of the regression-adjusted estimator under rerandomization. Section 4 discusses optimal regression adjustment in terms of the sampling precision. Section 5

addresses estimation and inference issues. Section 6 discusses optimal regression adjustment in terms of the estimated precision. Section 7 quantifies the gains from the analyser and the designer in both the sampling precision and the estimated precision. Section 8 unifies the discussion and gives practical suggestions. Section 9 uses examples to illustrate the theory. Section 10 concludes, and the on-line supplementary material contains all the technical details.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

`https://rss.onlinelibrary.wiley.com/hub/journal/14679868/series-b-datasets`.

2. Framework and notation

Consider an experiment on n units, with n_1 of them assigned to the treatment and n_0 of them assigned to the control. Let $r_1 = n_1/n$ and $r_0 = n_0/n$ be the proportions of units receiving the treatment and control. We use potential outcomes to define treatment effects (Neyman, 1923). For unit i, let $Y_i(1)$ and $Y_i(0)$ be the potential outcomes under the treatment and control, and $\tau_i = Y_i(1) - Y_i(0)$ be the individual treatment effect. For this finite population of n units, the average potential outcome under treatment arm z ($z = 0, 1$) is $\bar{Y}(z) = n^{-1}\Sigma_{i=1}^n Y_i(z)$, and the average treatment effect is $\tau = n^{-1}\Sigma_{i=1}^n \tau_i = \bar{Y}(1) - \bar{Y}(0)$. Let Z_i be the treatment assignment for unit i ($Z_i = 1$ for the treatment; $Z_i = 0$ for the control), and $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)'$ be the treatment assignment vector. The observed outcome for unit i is $Y_i = Z_i Y_i(1) + (1 - Z_i)Y_i(0)$.

2.1. Regression adjustment in the analysis

In a CRE, the probability that $\mathbf{Z}$ takes a particular value $\mathbf{z} = (z_1, \ldots, z_n)$ is $\binom{n}{n_1}^{-1}$, where $\Sigma_{i=1}^n z_i = n_1$ and $\Sigma_{i=1}^n (1 - z_i) = n_0$ are fixed and do not depend on the values of covariates or potential outcomes. Equivalently, $\mathbf{Z}$ is a random permutation of a vector of n_1 1s and n_0 0s. Let $\mathbf{w}_i = (w_{i1}, \ldots, w_{iJ})'$ be the J observed pretreatment covariates that are available to the analyser. For descriptive convenience, we centre these covariates at mean 0, i.e. $n^{-1}\Sigma_{i=1}^n \mathbf{w}_i = \mathbf{0}$. Let

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

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

be the difference in means of the outcome Y and covariates $\mathbf{w}$ respectively. Without covariate adjustment, $\hat{\tau}$ is unbiased for τ. After the experiment, the analyser can improve the estimation precision for the average treatment effect by adjusting for the observed covariate imbalance $\hat{\boldsymbol{\tau}}_{\mathbf{w}}$. A general linear regression-adjusted estimator has the following equivalent forms:

$$\begin{aligned} \hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) &= n_1^{-1} \sum_{i=1}^n Z_i (Y_i - \boldsymbol{\beta}_1' \mathbf{w}_i) - n_0^{-1} \sum_{i=1}^n (1 - Z_i)(Y_i - \boldsymbol{\beta}_0' \mathbf{w}_i) \\ &= \hat{\tau} - (r_0 \boldsymbol{\beta}_1 + r_1 \boldsymbol{\beta}_0)' \hat{\boldsymbol{\tau}}_{\mathbf{w}} = \hat{\tau} - \boldsymbol{\gamma}' \hat{\boldsymbol{\tau}}_{\mathbf{w}}, \end{aligned} \tag{1}$$

where $\boldsymbol{\beta}_1, \boldsymbol{\beta}_0$ and $\boldsymbol{\gamma} = r_0 \boldsymbol{\beta}_1 + r_1 \boldsymbol{\beta}_0$ are J-dimensional coefficients. From equation (1), $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ depends on $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ only through $\boldsymbol{\gamma} = r_0 \boldsymbol{\beta}_1 + r_1 \boldsymbol{\beta}_0$. Therefore, the choice of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ is not unique to achieve the same gain in efficiency. For simplicity, we shall also call equation (1) an *adjusted estimator* from now on.

Fisher's (1935) ANCOVA chose $\beta_1 = \beta_0$ to be the coefficient of $\mathbf{w}$ in the OLS fit of the observed outcome Y on the treatment Z and covariates $\mathbf{w}$. Freedman (2008b) criticized ANCOVA because

(a) the resulting estimator can be even less efficient than $\hat{\tau}$ and
(b) the standard error based on OLS can be inconsistent under the potential outcomes framework.

Lin (2013) fixed criticism (a) by choosing β_1 and β_0 to be the coefficients of $\mathbf{w}$ in the OLS fit of Y on $\mathbf{w}$ for treated and control units respectively. The resulting adjusted estimator is numerically identical to the coefficient of Z in the OLS fit of Y on Z, $\mathbf{w}$ and $Z \times \mathbf{w}$. Lin (2013) fixed criticism (b) by using the Huber–White robust standard error for linear models. Asymptotically, Lin's (2013) estimator has smaller standard error and estimated standard error than does $\hat{\tau}$.

As a side note, Lin's (2013) estimator also appeared in the semiparametric efficiency theory for the average treatment effect under independent sampling from a superpopulation (Koch *et al.*, 1998; Yang and Tsiatis, 2001; Leon *et al.*, 2003; Tsiatis *et al.*, 2008; Rubin and van der Laan, 2011).

2.2. Rerandomization in the design

The above regression adjustment uses covariates in the analysis stage. We can also use covariates in the design stage to improve the quality of randomization and the efficiency of estimates. Before conducting the experiment, the designer collects K covariates $\mathbf{x}_i = (x_{i1}, \ldots, x_{iK})'$ for unit i. Similarly, we centre the covariates at mean 0, i.e. $n^{-1}\sum_{i=1}^{n} \mathbf{x}_i = \mathbf{0}$. Note that we allow $\mathbf{x}$ to be different from $\mathbf{w}$. The CRE balances covariates on average, but an unlucky draw of the treatment vector can result in large covariate imbalance (Student, 1938; Cox, 1982, 2009; Bruhn and McKenzie, 2009; Morgan and Rubin, 2012). Therefore, it is sensible for the designer to check the covariate balance before conducting the experiment. Let $\hat{\boldsymbol{\tau}}_{\mathbf{x}} = n_1^{-1}\sum_{i=1}^{n} Z_i \mathbf{x}_i - n_0^{-1}\sum_{i=1}^{n}(1 - Z_i)\mathbf{x}_i$ be the difference in means of the covariates $\mathbf{x}$ between the treatment and control groups. It has mean 0 under the CRE. However, imbalance in covariate distributions often occurs for a realized treatment allocation. We can discard those unlucky treatment allocations with large covariate imbalance and rerandomize until the allocation satisfies a certain covariate balance criterion. This is rerandomization, which has the following steps.

Step 1: collect covariate data and specify a covariate balance criterion.
Step 2: randomize the units into treatment and control groups.
Step 3: if the allocation satisfies the balance criterion, proceed to step 4; otherwise, return to step 2.
Step 4: conduct the experiment by using the accepted allocation from step 3.

The balance criterion in step 1 can be a general function of the treatment assignment $\mathbf{Z}$ and the covariates $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$. Morgan and Rubin (2012) focused on rerandomization using the Mahalanobis distance, which accepts a randomization if and only if $M \equiv \hat{\boldsymbol{\tau}}_{\mathbf{x}}' \operatorname{cov}(\hat{\boldsymbol{\tau}}_{\mathbf{x}})^{-1} \hat{\boldsymbol{\tau}}_{\mathbf{x}} \leqslant a$, where M is the Mahalanobis distance between the covariate means in two groups and $a > 0$ is the predetermined threshold. Since we shall focus on rerandomization using the Mahalanobis distance throughout the paper, we shall simply call it 'rerandomization' from now on. Li *et al.* (2018) derived the asymptotic distribution of $\hat{\tau}$ under rerandomization, and showed that it is more precise than $\hat{\tau}$ under the CRE. They further showed that, when a is small and $\mathbf{x} = \mathbf{w}$, the asymptotic variance of $\hat{\tau}$ under rerandomization is nearly identical to Lin's (2013) adjusted estimator under the CRE. Therefore, rerandomization and regression adjustment both use covariates to improve efficiency of treatment effect estimates, but in the design and analysis stages respectively.

3. Sampling distributions of regression adjustment under rerandomization

Section 2 shows that rerandomization trumps the CRE in the design stage and regression adjustment trumps the difference in means in the analysis stage. Therefore, it is natural to combine rerandomization and regression adjustment. Then a key question is how do we conduct statistical inference? This requires us to study the sampling distribution of the adjusted estimator $\hat{\tau}(\beta_1, \beta_0)$ in equation (1) under rerandomization.

3.1. Basics of randomization-based inference

To facilitate the discussion, we introduce some basic results from finite population causal inference. The first part describes fixed finite population quantities without randomness. The second part describes the repeated sampling properties and asymptotics under the CRE.

3.1.1. Finite population quantities, projections and regularity conditions

For the treatment arm z ($z = 0, 1$), let $S^2_{Y(z)} = (n-1)^{-1}\Sigma_{i=1}^n\{Y_i(z) - \bar{Y}(z)\}^2$ be the finite population variance of the potential outcomes, and $\mathbf{S}_{Y(z),\mathbf{x}} = \mathbf{S}'_{\mathbf{x},Y(z)} = (n-1)^{-1}\Sigma_{i=1}^n\{Y_i(z) - \bar{Y}(z)\}\mathbf{x}_i'$ be the finite population covariance between the potential outcomes and covariates. Let $\mathbf{S}^2_{\mathbf{x}} = (n-1)^{-1}\Sigma_{i=1}^n\mathbf{x}_i\mathbf{x}_i'$ be the finite population covariance of the covariates. We can similarly define $\mathbf{S}_{Y(z),\mathbf{w}}$, $\mathbf{S}^2_{\mathbf{w}}$ and other covariances.

We introduce linear projections among these fixed quantities. For example, the linear projection of the potential outcome $Y(z)$ on covariates $\mathbf{w}$ is $\bar{Y}(z) + \tilde{\beta}_z'\mathbf{w}_i$ for unit i, with the coefficient

$$\tilde{\beta}_z = \arg\min_{\mathbf{b}\in\mathbb{R}^J} n^{-1}\sum_{i=1}^n\{Y_i(z) - \bar{Y}(z) - \mathbf{b}'\mathbf{w}_i\}^2 = (\mathbf{S}^2_{\mathbf{w}})^{-1}\mathbf{S}_{\mathbf{w},Y(z)} \qquad (z = 0, 1). \tag{2}$$

The residual from this projection is $Y_i(z) - \bar{Y}(z) - \tilde{\beta}_z'\mathbf{w}_i$ for unit i. Let $S^2_{Y(z)|\mathbf{w}} \equiv \mathbf{S}_{Y(z),\mathbf{w}}(\mathbf{S}^2_{\mathbf{w}})^{-1}\mathbf{S}_{\mathbf{w},Y(z)}$ denote the finite population variance of the linear projections, and $S^2_{Y(z)\backslash\mathbf{w}} \equiv S^2_{Y(z)} - S^2_{Y(z)|\mathbf{w}}$ the finite population variance of the residuals. We can similarly define $S^2_{\tau|\mathbf{w}}$, $S^2_{\tau\backslash\mathbf{w}}$, $S^2_{Y(z)|\mathbf{x}}$, $S^2_{Y(z)\backslash\mathbf{x}}$, $S^2_{\tau|\mathbf{x}}$, $\mathbf{S}^2_{\mathbf{w}|\mathbf{x}}$ and $\mathbf{S}^2_{\mathbf{w}\backslash\mathbf{x}}$.

The exact distributions of the estimators depend on unknown potential outcomes in general. We shall use asymptotic approximations. Finite population asymptotics embed the n units into a sequence of finite populations with increasing sizes. Technically, all quantities above depend on n, but we keep their dependence on n implicit for notational simplicity. Moreover, the sequence of finite populations must satisfy some regularity conditions to ensure the existence of the limiting distributions of the estimators. We use the regularity conditions that are motivated by the finite population central limit theorems (Li and Ding, 2017).

Condition 1. As $n \to \infty$, the sequence of finite populations satisfies that, for $z = 0, 1$,

(a) $r_z = n_z/n$, the proportion of units receiving treatment z, has a positive limit,
(b) the finite population variances and covariances, $S^2_{Y(z)}, S^2_\tau, \mathbf{S}^2_{\mathbf{x}}, \mathbf{S}^2_{\mathbf{w}}, \mathbf{S}_{Y(z),\mathbf{x}}, \mathbf{S}_{Y(z),\mathbf{w}}$ and $\mathbf{S}_{\mathbf{x},\mathbf{w}}$, have limiting values, and the limits of $\mathbf{S}^2_{\mathbf{x}}$ and $\mathbf{S}^2_{\mathbf{w}}$ are non-singular;
(c) $\max_{1\leqslant i\leqslant n}|Y_i(z) - \bar{Y}(z)|^2/n \to 0$, $\max_{1\leqslant i\leqslant n}\|\mathbf{x}_i\|_2^2/n \to 0$ and $\max_{1\leqslant i\leqslant n}\|\mathbf{w}_i\|_2^2/n \to 0$.

In condition 1, conditions (a) and (b) are natural, and condition (c) holds almost surely if all the variables are independent and identically distributed (IID) draws from a superpopulation with more than two moments (Li and Ding, 2017). Throughout the paper, we assume that the numbers of covariates K in the design and J in the analysis are both fixed and do not increase with the sample size n.

3.1.2. Repeated sampling inference under the completely randomized experiment

Under the CRE, over all $\binom{n}{n_1}$ randomizations, $n^{1/2}(\hat{\tau}-\tau, \hat{\boldsymbol{\tau}}_{\mathbf{x}}', \hat{\boldsymbol{\tau}}_{\mathbf{w}}')'$ has mean $\mathbf{0}$ and covariance

$$\mathbf{V} \equiv \begin{pmatrix} r_1^{-1}S_{Y(1)}^2 + r_0^{-1}S_{Y(0)}^2 - S_\tau^2 & r_1^{-1}\mathbf{S}_{Y(1),\mathbf{x}} + r_0^{-1}\mathbf{S}_{Y(0),\mathbf{x}} & r_1^{-1}\mathbf{S}_{Y(1),\mathbf{w}} + r_0^{-1}\mathbf{S}_{Y(0),\mathbf{w}} \\ r_1^{-1}\mathbf{S}_{\mathbf{x},Y(1)} + r_0^{-1}\mathbf{S}_{\mathbf{x},Y(0)} & (r_1 r_0)^{-1}\mathbf{S}_{\mathbf{x}}^2 & (r_1 r_0)^{-1}\mathbf{S}_{\mathbf{x},\mathbf{w}} \\ r_1^{-1}\mathbf{S}_{\mathbf{w},Y(1)} + r_0^{-1}\mathbf{S}_{\mathbf{w},Y(0)} & (r_1 r_0)^{-1}\mathbf{S}_{\mathbf{w},\mathbf{x}} & (r_1 r_0)^{-1}\mathbf{S}_{\mathbf{w}}^2 \end{pmatrix}$$

$$\equiv \begin{pmatrix} V_{\tau\tau} & \mathbf{V}_{\tau\mathbf{x}} & \mathbf{V}_{\tau\mathbf{w}} \\ \mathbf{V}_{\mathbf{x}\tau} & \mathbf{V}_{\mathbf{xx}} & \mathbf{V}_{\mathbf{xw}} \\ \mathbf{V}_{\mathbf{w}\tau} & \mathbf{V}_{\mathbf{wx}} & \mathbf{V}_{\mathbf{ww}} \end{pmatrix}. \tag{3}$$

The finite population central limit theorem of Li and Ding (2017) ensures that $n^{1/2}(\hat{\tau}-\tau, \hat{\boldsymbol{\tau}}_{\mathbf{x}}', \hat{\boldsymbol{\tau}}_{\mathbf{w}}')'$ is asymptotically Gaussian with mean $\mathbf{0}$ and covariance matrix $\mathbf{V}$ under the CRE and condition 1. We use a tilde with a dot for two sequences of random vectors (or distributions) converging weakly to the same distribution. Therefore, $n^{1/2}(\hat{\tau}-\tau, \hat{\boldsymbol{\tau}}_{\mathbf{x}}', \hat{\boldsymbol{\tau}}_{\mathbf{w}}')' \dot{\sim} \mathcal{N}(\mathbf{0}, \mathbf{V})$.

We define linear projections for random variables. We use $\mathbb{E}(\cdot)$, $\mathrm{var}(\cdot)$ and $\mathrm{cov}(\cdot)$ for the mean, variance and covariance, and $\mathrm{proj}(\cdot|\cdot)$ and $\mathrm{res}(\cdot|\cdot)$ for linear projections and corresponding residuals, exclusively under the CRE. For example, the linear projection of $\hat{\tau}$ on $\hat{\boldsymbol{\tau}}_{\mathbf{w}}$ is $\mathrm{proj}(\hat{\tau}|\hat{\boldsymbol{\tau}}_{\mathbf{w}}) = \tau + \tilde{\boldsymbol{\gamma}}'\hat{\boldsymbol{\tau}}_{\mathbf{w}}$, with the coefficient

$$\tilde{\boldsymbol{\gamma}} = \arg\min_{\mathbf{b}\in\mathbb{R}^J} \mathbb{E}(\hat{\tau}-\tau-\mathbf{b}'\hat{\boldsymbol{\tau}}_{\mathbf{w}})^2 = \mathrm{cov}(\hat{\boldsymbol{\tau}}_{\mathbf{w}})^{-1}\mathrm{cov}(\hat{\boldsymbol{\tau}}_{\mathbf{w}}, \hat{\tau}) = \mathbf{V}_{\mathbf{ww}}^{-1}\mathbf{V}_{\mathbf{w}\tau}. \tag{4}$$

The residual from this projection is $\mathrm{res}(\hat{\tau}|\hat{\boldsymbol{\tau}}_{\mathbf{w}}) = \hat{\tau} - \mathrm{proj}(\hat{\tau}|\hat{\boldsymbol{\tau}}_{\mathbf{w}}) = \hat{\tau} - \tau - \tilde{\boldsymbol{\gamma}}'\hat{\boldsymbol{\tau}}_{\mathbf{w}}$. We can similarly define $\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\mathbf{x}}|\hat{\boldsymbol{\tau}}_{\mathbf{w}})$ and $\mathrm{res}(\hat{\boldsymbol{\tau}}_{\mathbf{x}}|\hat{\boldsymbol{\tau}}_{\mathbf{w}})$.

Finally, the three linear projection coefficients $\tilde{\boldsymbol{\beta}}_1$, $\tilde{\boldsymbol{\beta}}_0$ and $\tilde{\boldsymbol{\gamma}}$ that are defined in equations (2) and (4) have the following relationship.

Proposition 1. $r_0\tilde{\boldsymbol{\beta}}_1 + r_1\tilde{\boldsymbol{\beta}}_0 = \tilde{\boldsymbol{\gamma}}$.

Proposition 1 is related to the non-uniqueness of the regression adjustment in equation (1). It is important for the discussion below.

3.2. Asymptotic distribution of regression adjustment under rerandomization

Equipped with the tools in Section 3.1, we now can derive the asymptotic distribution of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ under rerandomization. We first fix the coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$, and we shall devote several sections to discuss the optimal choices of them.

For unit i, let $Y_i(z;\boldsymbol{\beta}_z) \equiv Y_i(z) - \boldsymbol{\beta}_z'\mathbf{w}_i$ be the 'adjusted' potential outcome under the treatment level z ($z=0,1$) and $\tau_i(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \equiv \tau_i - (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)'\mathbf{w}_i$ be the adjusted individual treatment effect. The average adjusted treatment effect $\tau(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \equiv n^{-1}\sum_{i=1}^n \tau_i(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) = \tau$ is identical to the average unadjusted treatment effect because of the centring of $\bar{\mathbf{w}} = \mathbf{0}$. The adjusted observed outcome is $Y_i(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) = Z_i Y_i(1;\boldsymbol{\beta}_1) + (1-Z_i)Y_i(0;\boldsymbol{\beta}_0)$. The adjusted estimator (1) is essentially the difference-in-means estimator with the adjusted potential outcomes. For $z=0,1$, let $S_{Y(z;\boldsymbol{\beta}_z)}^2$ and $S_{Y(z;\boldsymbol{\beta}_z)|\mathbf{x}}^2$ be the finite population variances of $Y_i(z;\boldsymbol{\beta}_z)$ and its linear projection on $\mathbf{x}_i$. Let $S_{\tau(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)}^2$ and $S_{\tau(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)|\mathbf{x}}^2$ be the finite population variances of $\tau_i(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ and its linear projection on $\mathbf{x}_i$. From Section 3.1.2, under the CRE, the variance of $n^{1/2}\{\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) - \tau\}$ is

$$V_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) = r_1^{-1}S_{Y(1;\boldsymbol{\beta}_1)}^2 + r_0^{-1}S_{Y(0;\boldsymbol{\beta}_0)}^2 - S_{\tau(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)}^2, \tag{5}$$

and the squared multiple correlation between $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ and $\hat{\boldsymbol{\tau}}_{\mathbf{x}}$ is (Li *et al.* (2018), proposition 1)

$$R^2_{\tau,\mathbf{x}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)=\frac{\text{var}[\text{proj}\{\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)|\hat{\boldsymbol{\tau}}_{\mathbf{x}}\}]}{\text{var}\{\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\}}=\frac{r_1^{-1}S^2_{Y(1;\boldsymbol{\beta}_1)|\mathbf{x}}+r_0^{-1}S^2_{Y(0;\boldsymbol{\beta}_0)|\mathbf{x}}-S^2_{\tau(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)|\mathbf{x}}}{r_1^{-1}S^2_{Y(1;\boldsymbol{\beta}_1)}+r_0^{-1}S^2_{Y(0;\boldsymbol{\beta}_0)}-S^2_{\tau(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)}}. \quad (6)$$

The asymptotic distribution of $\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)$ under rerandomization is a linear combination of two independent random variables ε and $L_{K,a}$, where $\varepsilon\sim\mathcal{N}(0,1)$ is a standard Gaussian random variable and $L_{K,a}\sim D_1|\mathbf{D}'\mathbf{D}\leqslant a$ is a truncated Gaussian random variable with $\mathbf{D}=(D_1,\ldots,D_K)\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)$. Let $\mathcal{M}$ denote the event $M\leqslant a$.

Theorem 1. Under rerandomization and condition 1,

$$n^{1/2}\{\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)-\tau\}|\mathcal{M}\,\dot{\sim}\,V^{1/2}_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)[\{1-R^2_{\tau,\mathbf{x}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\}^{1/2}\varepsilon+R^2_{\tau,\mathbf{x}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)^{1/2}L_{K,a}]. \quad (7)$$

The ε-component in expression (7) represents the part of $\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)$ that cannot be explained by $\hat{\boldsymbol{\tau}}_{\mathbf{x}}$ and is thus unaffected by rerandomization. The $L_{K,a}$-component in expression (7) represents the part of $\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)$ that can be explained by $\hat{\boldsymbol{\tau}}_{\mathbf{x}}$ and is thus affected by rerandomization. Moreover, the asymptotic distribution (7) is symmetric around zero, and the adjusted estimator is consistent for the average treatment effect, for any fixed values of the coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$. Theorem 1 immediately implies the following two important special cases.

3.2.1. *Special case: regression adjustment under the completely randomized experiment*

The CRE is a special case of rerandomization with $a=\infty$. Therefore, theorem 1 implies that, under the CRE, $n^{1/2}\{\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)-\tau\}$ is asymptotically Gaussian with mean 0 and variance $V_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)$.

Corollary 1. Under CRE and condition 1, $n^{1/2}\{\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)-\tau\}\dot{\sim}V^{1/2}_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\varepsilon$.

Corollary 1 is a known result from Lin (2013) and Li and Ding (2017).

3.2.2. *Special case: no covariate adjustment under rerandomization*

Using theorem 1 with $\boldsymbol{\beta}_1=\boldsymbol{\beta}_0=\mathbf{0}$, we can immediately obtain the asymptotic distribution of $\hat{\tau}\equiv\hat{\tau}(\mathbf{0},\mathbf{0})$ under rerandomization. Let $R^2_{\tau,\mathbf{x}}\equiv R^2_{\tau,\mathbf{x}}(\mathbf{0},\mathbf{0})$ be the squared multiple correlation between $\hat{\tau}$ and $\hat{\boldsymbol{\tau}}_{\mathbf{x}}$ under the CRE:

$$R^2_{\tau,\mathbf{x}}=\frac{\text{var}\{\text{proj}(\hat{\tau}|\hat{\boldsymbol{\tau}}_{\mathbf{x}})\}}{\text{var}(\hat{\tau})}=\frac{\mathbf{V}_{\tau\mathbf{x}}\mathbf{V}^{-1}_{\mathbf{xx}}\mathbf{V}_{\mathbf{x}\tau}}{V_{\tau\tau}}=\frac{r_1^{-1}S^2_{Y(1)|\mathbf{x}}+r_0^{-1}S^2_{Y(0)|\mathbf{x}}-S^2_{\tau|\mathbf{x}}}{r_1^{-1}S^2_{Y(1)}+r_0^{-1}S^2_{Y(0)}-S^2_{\tau}}. \quad (8)$$

Then $\hat{\tau}$ has the following asymptotic distribution.

Corollary 2. Under rerandomization and condition 1,

$$n^{1/2}(\hat{\tau}-\tau)|\mathcal{M}\dot{\sim}V^{1/2}_{\tau\tau}\{(1-R^2_{\tau,\mathbf{x}})^{1/2}\varepsilon+(R^2_{\tau,\mathbf{x}})^{1/2}L_{K,a}\}. \quad (9)$$

Corollary 2 is a main result of Li *et al.* (2018).

4. $\mathcal{S}$-optimal regression adjustment

How do we choose the adjustment coefficients $(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)$ or $\boldsymbol{\gamma}$? It is an important practical question. From theorem 1, the adjusted estimator is consistent for any fixed coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$. Therefore, it is intuitive to choose the coefficients to achieve better precision. A measure of precision is based on the quantile ranges of an estimator.

We introduce the notion of the $\mathcal{S}$-optimal adjusted estimator, using $\mathcal{S}$ to emphasize its dependence on the *sampling* distribution.

Definition 1. Given the design, $\hat{\tau}(\beta_1, \beta_0)$ is an $\mathcal{S}$-optimal estimator if $n^{1/2}\{\hat{\tau}(\beta_1, \beta_0) - \tau\}$ has the shortest asymptotic $1-\alpha$ quantile range among all adjusted estimators in equation (1), for any $\alpha \in (0, 1)$.

In general, quantile ranges are not unique. Importantly, the asymptotic distribution in expression (7) is symmetric and unimodal around τ (Li *et al.*, 2018). We consider only symmetric quantile ranges because they have the shortest lengths (Casella and Berger (2002), theorem 9.3.2). The $\mathcal{S}$-optimal adjusted estimator also has the smallest asymptotic variance among all estimators in equation (1) (Li *et al.* (2019), proposition 4). Conversely, if all adjusted estimators in equation (1) are asymptotically Gaussian, then the estimator with the smallest asymptotic variance is an $\mathcal{S}$-optimal estimator.

Theorem 1 shows a complicated relationship between the coefficients (β_1, β_0) and the asymptotic distribution (7). Below we simplify expression (7). Let $\text{proj}(\hat{\boldsymbol{\tau}}_{\mathbf{w}}|\hat{\boldsymbol{\tau}}_{\mathbf{x}}) \equiv \mathbf{V}_{\mathbf{wx}}\mathbf{V}_{\mathbf{xx}}^{-1}\hat{\boldsymbol{\tau}}_{\mathbf{x}}$ be the linear projection of $\hat{\boldsymbol{\tau}}_{\mathbf{w}}$ on $\hat{\boldsymbol{\tau}}_{\mathbf{x}}$, and $\text{res}(\hat{\boldsymbol{\tau}}_{\mathbf{w}}|\hat{\boldsymbol{\tau}}_{\mathbf{x}}) \equiv \hat{\boldsymbol{\tau}}_{\mathbf{w}} - \mathbf{V}_{\mathbf{wx}}\mathbf{V}_{\mathbf{xx}}^{-1}\hat{\boldsymbol{\tau}}_{\mathbf{x}}$ be the residual from this linear projection. We further consider two projections. First, the linear projection of $\text{proj}(\hat{\tau}|\hat{\boldsymbol{\tau}}_{\mathbf{x}})$ on $\text{proj}(\hat{\boldsymbol{\tau}}_{\mathbf{w}}|\hat{\boldsymbol{\tau}}_{\mathbf{x}})$ has coefficient

$$\tilde{\gamma}_{\text{proj}} = (\mathbf{V}_{\mathbf{wx}}\mathbf{V}_{\mathbf{xx}}^{-1}\mathbf{V}_{\mathbf{xw}})^{-1}\mathbf{V}_{\mathbf{wx}}\mathbf{V}_{\mathbf{xx}}^{-1}\mathbf{V}_{\mathbf{x}\tau} \tag{10}$$

and squared multiple correlation R^2_{proj}. Second, the linear projection of $\text{res}(\hat{\tau}|\hat{\boldsymbol{\tau}}_{\mathbf{x}})$ on $\text{res}(\hat{\boldsymbol{\tau}}_{\mathbf{w}}|\hat{\boldsymbol{\tau}}_{\mathbf{x}})$ has coefficient

$$\tilde{\gamma}_{\text{res}} \equiv (\mathbf{V}_{\mathbf{ww}} - \mathbf{V}_{\mathbf{wx}}\mathbf{V}_{\mathbf{xx}}^{-1}\mathbf{V}_{\mathbf{xw}})^{-1}(\mathbf{V}_{\mathbf{w}\tau} - \mathbf{V}_{\mathbf{wx}}\mathbf{V}_{\mathbf{xx}}^{-1}\mathbf{V}_{\mathbf{x}\tau}) \tag{11}$$

and squared multiple correlation R^2_{res}.

Technically, the expressions for $\tilde{\gamma}_{\text{proj}}$ and $\tilde{\gamma}_{\text{res}}$ above are well defined only if the covariance matrices of $\text{proj}(\hat{\boldsymbol{\tau}}_{\mathbf{w}}|\hat{\boldsymbol{\tau}}_{\mathbf{x}})$ and $\text{res}(\hat{\boldsymbol{\tau}}_{\mathbf{w}}|\hat{\boldsymbol{\tau}}_{\mathbf{x}})$ are non-singular. Otherwise, they are not unique. However, this will not cause any issues in the later discussions because the linear projections themselves are always unique.

Recall that we have defined $\mathbf{V}$ in equation (3), and $\mathbf{S}^2_{\mathbf{w}|\mathbf{x}}$ and $\mathbf{S}^2_{\mathbf{w}\backslash\mathbf{x}}$ as the finite population covariances of the linear projections of $\mathbf{w}$ on $\mathbf{x}$ and the corresponding residuals. The following proposition shows the relationship between the three linear projection coefficients $(\tilde{\gamma}, \tilde{\gamma}_{\text{proj}}, \tilde{\gamma}_{\text{res}})$.

Proposition 2. $\mathbf{S}^2_{\mathbf{w}\backslash\mathbf{x}}(\tilde{\gamma} - \tilde{\gamma}_{\text{res}}) + \mathbf{S}^2_{\mathbf{w}|\mathbf{x}}(\tilde{\gamma} - \tilde{\gamma}_{\text{proj}}) = \mathbf{0}$.

The linear projection coefficients $(\tilde{\gamma}, \tilde{\gamma}_{\text{proj}}, \tilde{\gamma}_{\text{res}})$ are different in general. However, if any two of them are equal, all of them must be equal with non-singular $\mathbf{S}^2_{\mathbf{w}\backslash\mathbf{x}}$ and $\mathbf{S}^2_{\mathbf{w}|\mathbf{x}}$. The following theorem decomposes the asymptotic distribution (7) based on $(\tilde{\gamma}_{\text{proj}}, \tilde{\gamma}_{\text{res}})$.

Theorem 2. Under rerandomization and condition 1, recalling that $\gamma \equiv r_0\beta_1 + r_1\beta_0$, we have

$$\begin{aligned} n^{1/2}\{\hat{\tau}(\beta_1, \beta_0) - \tau\}|\mathcal{M} \dot{\sim} &\{V_{\tau\tau}(1 - R^2_{\tau,\mathbf{x}})(1 - R^2_{\text{res}}) + (r_1 r_0)^{-1}(\gamma - \tilde{\gamma}_{\text{res}})'\mathbf{S}^2_{\mathbf{w}\backslash\mathbf{x}}(\gamma - \tilde{\gamma}_{\text{res}})\}^{1/2}\varepsilon \\ &+ \{V_{\tau\tau}R^2_{\tau,\mathbf{x}}(1 - R^2_{\text{proj}}) + (r_1 r_0)^{-1}(\gamma - \tilde{\gamma}_{\text{proj}})'\mathbf{S}^2_{\mathbf{w}|\mathbf{x}}(\gamma - \tilde{\gamma}_{\text{proj}})\}^{1/2}L_{K,a}. \end{aligned} \tag{12}$$

The asymptotic distribution (12) has two independent components. The ε-component in expression (12) represents the part of $\hat{\tau}(\beta_1, \beta_0)$ that is orthogonal to $\hat{\boldsymbol{\tau}}_{\mathbf{x}}$. The coefficient of ε attains its minimal value at $\gamma = \tilde{\gamma}_{\text{res}}$ with squared minimal value $V_{\tau\tau}(1 - R^2_{\tau,\mathbf{x}})(1 - R^2_{\text{res}})$. The

first term $V_{\tau\tau}$ is the variance of $n^{1/2}(\hat{\tau}-\tau)$. The second term $1-R^2_{\tau,\mathbf{x}}$ represents the proportion of the variance of $\hat{\tau}$ unexplained by $\hat{\boldsymbol{\tau}}_{\mathbf{x}}$. The third term $1-R^2_{\text{res}}$ represents the proportion of the variance of $\hat{\tau}$ unexplained by $\hat{\boldsymbol{\tau}}_{\mathbf{w}}$, after projecting onto the space that is orthogonal to $\hat{\boldsymbol{\tau}}_{\mathbf{x}}$.

The $L_{K,a}$-component in expression (12) represents the linear projection of $\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)$ on $\hat{\boldsymbol{\tau}}_{\mathbf{x}}$ with the rerandomization constraint. The coefficient of $L_{K,a}$ attains its minimal value at $\boldsymbol{\gamma}=\tilde{\boldsymbol{\gamma}}_{\text{proj}}$ with squared minimal value $V_{\tau\tau}R^2_{\tau,\mathbf{x}}(1-R^2_{\text{proj}})$. The first term $V_{\tau\tau}$ is again the variance of $n^{1/2}(\hat{\tau}-\tau)$. The second term $R^2_{\tau,\mathbf{x}}$ represents the proportion of the variance of $\hat{\tau}$ explained by $\hat{\boldsymbol{\tau}}_{\mathbf{x}}$. The third term $1-R^2_{\text{proj}}$ represents the proportion of the variance of $\hat{\tau}$ unexplained by $\hat{\boldsymbol{\tau}}_{\mathbf{w}}$, after projecting onto the space of $\hat{\boldsymbol{\tau}}_{\mathbf{x}}$.

Because $\tilde{\boldsymbol{\gamma}}_{\text{res}}$ and $\tilde{\boldsymbol{\gamma}}_{\text{proj}}$ are different in general, the coefficients of ε and $L_{K,a}$ cannot attain their minimal values simultaneously. Consequently, the adjusted estimator $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)$ may not be an $\mathcal{S}$-optimal estimator under rerandomization, i.e. it may not have the shortest asymptotic $1-\alpha$ quantile range among equation (1) for all $\alpha\in(0,1)$.

The $\mathcal{S}$-optimal adjustment is complicated under rerandomization, especially when the designer and the analyser have different covariate information. We shall consider different scenarios based on the relative amount of covariate information that is used by the designer and the analyser.

4.1. The analyser has no less covariate information than the designer

We first consider the scenario under which the covariates $\mathbf{w}$ in the analysis can linearly represent the covariates $\mathbf{x}$ in the design.

Condition 2. There is a constant matrix $\mathbf{B}_1\in\mathbb{R}^{K\times J}$ such that $\mathbf{x}_i=\mathbf{B}_1\mathbf{w}_i$ for all unit i.

Condition 2 holds when the analyser has access to all the covariates that are used in the design, and possibly collects more covariates after the experiment. For example, condition 2 holds if $\mathbf{x}$ is a subset of $\mathbf{w}$. Under condition 2, we can simplify the asymptotic distribution of the adjusted estimator under rerandomization. Analogously to equations (6) and (8), let $R^2_{\tau,\mathbf{w}}$ be the squared multiple correlation between $\hat{\tau}$ and $\hat{\boldsymbol{\tau}}_{\mathbf{w}}$ under the CRE (Li *et al.*, 2018):

$$R^2_{\tau,\mathbf{w}}=\frac{\text{var}\{\text{proj}(\hat{\tau}|\hat{\boldsymbol{\tau}}_{\mathbf{w}})\}}{\text{var}(\hat{\tau})}=\frac{\mathbf{V}_{\tau\mathbf{w}}\mathbf{V}_{\mathbf{ww}}^{-1}\mathbf{V}_{\mathbf{w}\tau}}{V_{\tau\tau}}=\frac{r_1^{-1}S^2_{Y(1)|\mathbf{w}}+r_0^{-1}S^2_{Y(0)|\mathbf{w}}-S^2_{\tau|\mathbf{w}}}{r_1^{-1}S^2_{Y(1)}+r_0^{-1}S^2_{Y(0)}-S^2_{\tau}}. \tag{13}$$

Corollary 3. Under conditions 1 and 2,

$$\begin{gathered}\tilde{\boldsymbol{\gamma}}_{\text{proj}}=\tilde{\boldsymbol{\gamma}}_{\text{res}}=\tilde{\boldsymbol{\gamma}},\\ R^2_{\text{res}}=(R^2_{\tau,\mathbf{w}}-R^2_{\tau,\mathbf{x}})/(1-R^2_{\tau,\mathbf{x}}),\\ R^2_{\text{proj}}=1,\end{gathered}$$

and the asymptotic distribution of $\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)$ under rerandomization is

$$\begin{aligned}n^{1/2}\{\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)-\tau\}|\mathcal{M}\dot{\sim}&\{V_{\tau\tau}(1-R^2_{\tau,\mathbf{w}})+(r_1r_0)^{-1}(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}})'\mathbf{S}^2_{\mathbf{w}\backslash\mathbf{x}}(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}})\}^{1/2}\varepsilon\\&+\{(r_1r_0)^{-1}(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}})'\mathbf{S}^2_{\mathbf{w}|\mathbf{x}}(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}})\}^{1/2}L_{K,a}.\end{aligned} \tag{14}$$

From corollary 3, the coefficients of ε and $L_{K,a}$ attain minimum values at the same $\boldsymbol{\gamma}=\tilde{\boldsymbol{\gamma}}$. We can then derive the $\mathcal{S}$-optimal adjusted estimator and its asymptotic distribution.

Theorem 3. Under rerandomization and conditions 1 and 2, the $\mathcal{S}$-optimal adjusted estimator is attainable when $\boldsymbol{\gamma}=\tilde{\boldsymbol{\gamma}}$ or $r_0\boldsymbol{\beta}_1+r_1\boldsymbol{\beta}_0=r_0\tilde{\boldsymbol{\beta}}_1+r_1\tilde{\boldsymbol{\beta}}_0$, with the asymptotic distribution

$$n^{1/2}\{\hat{\tau}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)-\tau\}|\mathcal{M}\dot{\sim}\{V_{\tau\tau}(1-R^2_{\tau,\mathbf{w}})\}^{1/2}\varepsilon. \tag{15}$$

From proposition 1, an optimal choice is $(\beta_1, \beta_0) = (\tilde{\beta}_1, \tilde{\beta}_0)$ under condition 2. An important feature of theorem 3 is that the limiting distribution in expression (15) does not depend on covariates $\mathbf{x}$ and the threshold a of rerandomization. Theorem 3 has many implications, as discussed below.

4.1.1. *Special case: $\mathcal{S}$-optimal adjustment under the completely randomized experiment*

Theorem 3 holds for the CRE (rerandomization with $\mathbf{x} = \emptyset$ and $a = \infty$). It thus recovers the optimality property of Lin's (2013) estimator under the CRE previously proved by Li and Ding (2017). Therefore, when the analyser has no less covariate information than the designer, the $\mathcal{S}$-optimal adjusted estimators under rerandomization and the CRE are the same and follow the same asymptotic distribution. This implies that, with more covariates in the analysis, there is no additional gain from the designer through rerandomization as long as the analyser performs the optimal adjustment. Section 7.1.1 later contains related discussions.

4.1.2. *Special case: the designer and analyser have the same covariates*

Consider the case where the analyser has the same covariates as the designer ($\mathbf{x} = \mathbf{w}$). Compare $\hat{\tau}$ under rerandomization using the Mahalanobis distance with the $\mathcal{S}$-optimal adjusted estimator $\hat{\tau}(\tilde{\beta}_1, \tilde{\beta}_0)$ under the CRE. From corollary 2 and Section 4.1.1, the former has an additional independent component of $(V_{\tau\tau} R^2_{\tau,\mathbf{x}})^{1/2} L_{K,a}$ in the asymptotic distribution. When the threshold a is small, this additional component is approximately 0, and thus they have almost the same asymptotic distribution. Therefore, we can view rerandomization as covariate adjustment in the design stage (Li *et al.*, 2018). Moreover, using $\hat{\tau}$ in rerandomization has the following advantages. First, rerandomization in the design stage does not use the outcome data. Second, $\hat{\tau}$ is simpler and thus provides a more transparent analysis (Cox, 2007; Freedman, 2008c; Rosenbaum, 2010; Lin, 2013). Using $\hat{\tau}$ in rerandomization can thus avoid bias due to a specification search of the outcome model (i.e. data snooping). Remark 3 later contains related discussions.

4.2. *The analyser has no more covariate information than the designer*

We now consider the scenario under which the covariates $\mathbf{x}$ in the design can linearly represent the covariates $\mathbf{w}$ in the analysis.

Condition 3. There is a constant matrix $\mathbf{B}_2 \in \mathbb{R}^{J \times K}$ such that $\mathbf{w}_i = \mathbf{B}_2 \mathbf{x}_i$ for all unit i.

Condition 3 is reasonable when the analyser has access to only part of the covariates that are used in the design due to privacy or other reasons. For example, condition 3 holds if $\mathbf{w}$ is a subset of $\mathbf{x}$. It also reflects the situation where the analyser uses only the difference-in-means estimator with $\mathbf{w} = \emptyset$ even though the designer conducts rerandomization with $\mathbf{x}$. Condition 3 implies that $\mathbf{S}^2_{\mathbf{w}\backslash\mathbf{x}} = 0$, which further implies that the coefficient of ε in the asymptotic distribution (12) does not depend on (β_1, β_0). We can then simplify the asymptotic distribution of the adjusted estimator.

Corollary 4. Under conditions 1 and 3,

$$\begin{gathered}
\tilde{\gamma}_{\text{proj}} = \tilde{\gamma}, \\
R^2_{\text{res}} = 0, \\
R^2_{\text{proj}} = R^2_{\tau,\mathbf{w}} / R^2_{\tau,\mathbf{x}}, \\
\mathbf{S}^2_{\mathbf{w}\backslash\mathbf{x}} = \mathbf{0}, \\
\mathbf{S}^2_{\mathbf{w}|\mathbf{x}} = \mathbf{S}^2_{\mathbf{w}},
\end{gathered}$$

and the asymptotic distribution of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ under rerandomization is

$$n^{1/2}\{\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) - \tau\} | \mathcal{M} \dot{\sim} \{V_{\tau\tau}(1 - R^2_{\tau,\mathbf{x}})\}^{1/2}\varepsilon + \{V_{\tau\tau}(R^2_{\tau,\mathbf{x}} - R^2_{\tau,\mathbf{w}}) + (r_1 r_0)^{-1}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}})' \mathbf{S}^2_{\mathbf{w}}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}})\}^{1/2} L_{K,a}.$$

Under condition 3, $\text{res}(\hat{\boldsymbol{\tau}}_{\mathbf{w}} | \hat{\boldsymbol{\tau}}_{\mathbf{x}}) = \mathbf{0}$, and thus, as discussed earlier, the projection coefficient $\tilde{\boldsymbol{\gamma}}_{\text{res}}$ is not unique. Nevertheless, corollary 4 does not depend on $\tilde{\boldsymbol{\gamma}}_{\text{res}}$. On the basis of corollary 4, we can obtain the $\mathcal{S}$-optimal adjusted estimator and its asymptotic distribution under rerandomization. Let

$$\rho^2_{\tau,\mathbf{x}\backslash\mathbf{w}} = (R^2_{\tau,\mathbf{x}} - R^2_{\tau,\mathbf{w}})/(1 - R^2_{\tau,\mathbf{w}}) \in [0, 1] \tag{16}$$

be the additional proportion of the variance of $\hat{\tau}$ explained by the covariates $\mathbf{x}$ in the design, after being explained by the covariates $\mathbf{w}$ in the analysis.

Theorem 4. Under rerandomization and conditions 1 and 3, the $\mathcal{S}$-optimal adjusted estimator is attainable when $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$ or $r_0\boldsymbol{\beta}_1 + r_1\boldsymbol{\beta}_0 = r_0\tilde{\boldsymbol{\beta}}_1 + r_1\tilde{\boldsymbol{\beta}}_0$, with the asymptotic distribution

$$\begin{aligned} n^{1/2}\{\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) - \tau\} | \mathcal{M} &\dot{\sim} V^{1/2}_{\tau\tau}\{(1 - R^2_{\tau,\mathbf{x}})^{1/2}\varepsilon + (R^2_{\tau,\mathbf{x}} - R^2_{\tau,\mathbf{w}})^{1/2} L_{K,a}\} \\ &\sim \{V_{\tau\tau}(1 - R^2_{\tau,\mathbf{w}})\}^{1/2}\{(1 - \rho^2_{\tau,\mathbf{x}\backslash\mathbf{w}})^{1/2}\varepsilon + (\rho^2_{\tau,\mathbf{x}\backslash\mathbf{w}})^{1/2} L_{K,a}\}. \end{aligned} \tag{17}$$

From theorem 4, although the analyser has less covariate information than the designer of rerandomization, she or he can still obtain the $\mathcal{S}$-optimal adjusted estimator by using only the covariate information in the analysis.

Theorems 3 and 4 give identical optimal coefficients, but different asymptotic distributions of the optimal estimators. When the designer and the analyser have the same covariates ($\mathbf{x} = \mathbf{w}$), both theorem 3 and theorem 4 hold and give identical results. Specifically, $\rho^2_{\tau,\mathbf{x}\backslash\mathbf{w}}$ in equation (16) reduces to 0, and the asymptotic distribution in expression (17) simplifies to expression (15): a Gaussian limiting distribution.

From corollary 2 and theorem 4, under rerandomization, the asymptotic distribution of the $\mathcal{S}$-optimal adjusted estimator $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ differs from that of $\hat{\tau}$ only in the coefficient of the truncated Gaussian random variable $L_{K,a}$. With a small threshold a, $L_{K,a}$ is close to 0 and thus the gain from adjustment is small. Similarly to the discussion in Section 4.1.2, although $\hat{\tau}$ loses a little sampling precision compared with the optimal adjusted estimator, it does have the advantage of avoiding data snooping and improving transparency.

4.3. General scenarios

A practical complication is that the designer and analyser may not communicate. Then it is possible that the designer and the analyser do not use the same covariate information (e.g. Bruhn and McKenzie (2009) and Ke *et al.* (2017)). Consequently, the analyser has part of the covariate information in the design and additional covariate information. Neither condition 2 nor condition 3 holds. Under general scenarios, unfortunately, the $\mathcal{S}$-optimal adjusted estimator may not exist, in the sense that there is not an estimator among equation (1) that has the shortest asymptotic $1 - \alpha$ quantile range for all $\alpha \in (0, 1)$.

Some suboptimal strategies exist. First, we can consider the adjusted estimator with the smallest asymptotic variance or the shortest asymptotic $1 - \alpha$ quantile range for a particular $\alpha \in (0, 1)$. Section A2.3 of the on-line supplementary material gives the formulae for the estimator with the smallest variance. However, explicit formulae for the estimator with the shortest quantile range do not exist.

Second, when a is small, $L_{K,a} \approx 0$, the asymptotic distribution (12) under rerandomization

depends mainly on the ε-component. The coefficient of ε attains its minimal value at $\tilde{\gamma}_{\text{res}}$. Ignoring the $L_{K,a}$-component, $\tilde{\gamma}_{\text{res}}$ gives the $\mathcal{S}$-optimal adjusted estimator. However, this result is not useful because it is infeasible for the analyser to estimate $\tilde{\gamma}_{\text{res}}$ consistently because of the incomplete information of the covariates in the design.

Third, we can still use $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ as a convenient adjusted estimator because we can easily obtain it via OLS. Not surprisingly, this estimator is not an $\mathcal{S}$-optimal estimator in general and it can be even worse than $\hat{\tau}$ under rerandomization. When a is small, the ε-components are the dominating terms in their asymptotic distributions under rerandomization. Therefore, we compare the coefficients of ε.

Theorem 5. Under rerandomization using the Mahalanobis distance and condition 1, the squared coefficient of ε is $V_{\tau\tau}(1 - R^2_{\tau,\mathbf{w}})\{1 - R^2_{\tau,\mathbf{x}}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)\}$ in the asymptotic distribution (7) of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ and is $V_{\tau\tau}(1 - R^2_{\tau,\mathbf{x}})$ in the asymptotic distribution (9) of $\hat{\tau}$. The former is smaller than or equal to the latter if and only if

$$R^2_{\tau,\mathbf{w}} + (1 - R^2_{\tau,\mathbf{w}})\, R^2_{\tau,\mathbf{x}}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) \geqslant R^2_{\tau,\mathbf{x}}. \tag{18}$$

Remark 1. A sufficient condition for inequality (18) is $R^2_{\tau,\mathbf{w}} \geqslant R^2_{\tau,\mathbf{x}}$, which holds under condition 2. Recall that $R^2_{\tau,\mathbf{x}}$ in expression (8) measures the covariate information of the designer, and $R^2_{\tau,\mathbf{w}}$ in expression (13) measures the covariate information of the analyser. From theorem 5, when the analyser has more covariate information, $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is more precise than $\hat{\tau}$ if the threshold a for rerandomization is small.

Remark 2. A counterexample for condition (18) is that the finite population partial covariance between $Y(z)$ and $\mathbf{w}$ given $\mathbf{x}$ is 0 for $z = 0, 1$. In this case, the squared coefficient of ε for $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is larger than or equal to that for $\hat{\tau}$. Intuitively, this is because the covariates in the analysis are unrelated to the potential outcomes after adjusting for the covariates in the design and using them only introduces additional variability. In the extreme case where $\mathbf{x}$ can linearly represent $Y(1)$ and $Y(0)$, the squared coefficient of ε for $\hat{\tau}$ is 0, whereas that for $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is generally positive. See section A2.4 of the on-line supplementary material for more details.

Below we use a numerical example to illustrate the results above. It shows that $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ can be superior or inferior to $\hat{\tau}$.

4.3.1. *Example 1*

We choose $n = 1000$ and $r_1 = r_0 = 0.5$, and generate the covariates and the potential outcomes by using IID samples from the following model:

$$\left.\begin{aligned} x, \eta, \delta &\overset{\text{IID}}{\sim} \mathcal{N}(0, 1), \\ w &= x + \eta, \\ Y(0) &= 2x + \rho\eta + (1 - \rho^2)^{1/2}\delta, \\ Y(1) &= Y(0) + 1. \end{aligned}\right\} \tag{19}$$

Once generated, the covariates and potential outcomes are all fixed. We use rerandomization based on the covariate x, and we choose the threshold a to be the 0.001th quantile of the χ^2_1 random variable. We then use regression adjustment based on the covariate w. Fig. 1(a) shows the histograms of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ and $\hat{\tau}$ under rerandomization when $\rho = 0.9$. In this case, regression adjustment increases the sampling precision. Fig. 1(b) shows the histograms when $\rho = 0$. In

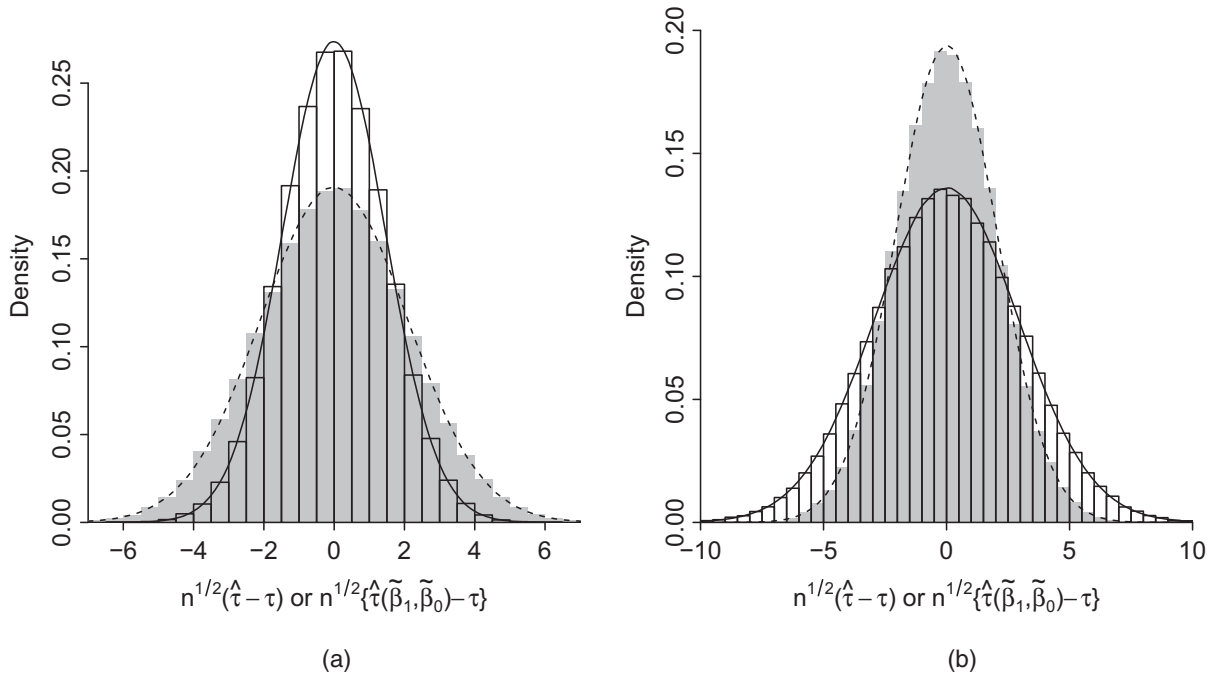

Fig. 1. Histograms of $n^{1/2}\{\hat{\tau}(\tilde{\beta}_1, \tilde{\beta}_0) - \tau\}$ and $n^{1/2}(\hat{\tau} - \tau)$ under rerandomization based on 10^5 simulated treatment assignments (□, adjusted; ■, unadjusted; ———, adjusted; - - - - -, unadjusted): (a) $\rho = 0.9$; (b) $\rho = 0$

this case, regression adjustment decreases the sampling precision. The case with $\rho = 0$ reflects the scenario that the designer gives only a covariate with measurement error to the analyser, possibly because of some privacy consideration.

5. Estimating the sampling distributions of the estimators

Sections 3 and 4 focused on the asymptotic distribution of the adjusted estimator and discussed optimal choices of the coefficients. In practice, we usually report the uncertainty of estimators in terms of confidence intervals when conducting frequentist inference. Confidence intervals are related to the quantile ranges of the estimated distributions of the corresponding estimators. Therefore, compared with $\mathcal{S}$-optimality, a more practical definition of the optimal adjusted estimator should be based on the quantile ranges of the estimated distributions. This subtle issue does not exist in many other statistical inference problems, because usually consistent estimators exist for the true asymptotic distributions of the estimators. For example, in standard statistical problems, we can consistently estimate the variance of a Gaussian limiting distribution. Because of the possible miscommunication between the designer and analyser, the analyser may not be able to estimate all quantities on the basis of the observed data in general. This is a feature of our framework.

In this section, we discuss the estimation of the sampling distributions for fixed (β_1, β_0). In the next section, we shall discuss the optimal choice of these coefficients. In what follows, we avoid the degenerate case that the estimated and the true distributions of the adjusted estimator is a point mass at zero. This requires that at least one of $S^2_{Y(1)\backslash \mathbf{w}}$ and $S^2_{Y(0)\backslash \mathbf{w}}$ has strictly positive limit.

5.1. The analyser knows all the information in the design

We first consider the scenario under which the analyser knows all the information of the designer.

Condition 4. The analyser knows all the information in the design, including the covariates $\mathbf{x}$ and the threshold a for rerandomization.

Condition 4 implies condition 2. However, condition 2 does not imply condition 4, because the analyser may not know which covariates are used in the design or which threshold a is chosen for rerandomization. From theorem 1, the asymptotic distribution (7) of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ under rerandomization depends on $V_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ and $R^2_{\tau,\mathbf{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$. Under treatment arm z ($z = 0, 1$), let $s^2_{Y(z;\boldsymbol{\beta}_z)}$, $\mathbf{s}_{Y(z;\boldsymbol{\beta}_z),\mathbf{w}}$ and $\mathbf{s}_{Y(z;\boldsymbol{\beta}_z),\mathbf{x}}$ be the sample variance and covariances for the adjusted observed outcome $Y_i - \boldsymbol{\beta}_z'\mathbf{w}_i$, covariates in the analysis $\mathbf{w}_i$ and covariates in the design $\mathbf{x}_i$; let $s^2_{Y(z;\boldsymbol{\beta}_z)|\mathbf{x}}$ be the sample variance of the linear projection of $Y_i - \boldsymbol{\beta}_z'\mathbf{w}_i$ on $\mathbf{x}_i$. We estimate $V_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ by

$$\hat{V}_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) = r_1^{-1} s^2_{Y(1;\boldsymbol{\beta}_1)} + r_0^{-1} s^2_{Y(0;\boldsymbol{\beta}_0)} - (\mathbf{s}_{Y(1;\boldsymbol{\beta}_1),\mathbf{w}} - \mathbf{s}_{Y(0;\boldsymbol{\beta}_0),\mathbf{w}})(\mathbf{S}^2_{\mathbf{w}})^{-1}(\mathbf{s}_{\mathbf{w},Y(1;\boldsymbol{\beta}_1)} - \mathbf{s}_{\mathbf{w},Y(0;\boldsymbol{\beta}_0)}), \tag{20}$$

$R^2_{\tau,\mathbf{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ by

$$\begin{aligned}\hat{R}^2_{\tau,\mathbf{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) = \hat{V}^{-1}_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)\{&r_1^{-1} s^2_{Y(1;\boldsymbol{\beta}_1)|\mathbf{x}} + r_0^{-1} s^2_{Y(0;\boldsymbol{\beta}_0)|\mathbf{x}} \\ &- (\mathbf{s}_{Y(1;\boldsymbol{\beta}_1),\mathbf{x}} - \mathbf{s}_{Y(0;\boldsymbol{\beta}_0),\mathbf{x}})(\mathbf{S}^2_{\mathbf{x}})^{-1}(\mathbf{s}_{\mathbf{x},Y(1;\boldsymbol{\beta}_1)} - \mathbf{s}_{\mathbf{x},Y(0;\boldsymbol{\beta}_0)})\},\end{aligned} \tag{21}$$

and the asymptotic distribution of $n^{1/2}\{\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) - \tau\}$ by

$$\hat{V}^{1/2}_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)[\{1 - \hat{R}^2_{\tau,\mathbf{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)\}^{1/2}\varepsilon + \hat{R}^2_{\tau,\mathbf{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)^{1/2} L_{K,a}]. \tag{22}$$

The estimated distribution (22) provides a basis for constructing confidence intervals for τ. However, it is not convenient for theoretical analyses because it is random. Below we find the probability limit of expression (22).

Theorem 6. Under rerandomization and conditions 1 and 4, the estimated distribution of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ in expression (22) has the same limit as

$$\{V_{\tau\tau}(1 - R^2_{\tau,\mathbf{w}}) + S^2_{\tau\backslash\mathbf{w}} + (r_1 r_0)^{-1}(\gamma - \tilde{\gamma})'\mathbf{S}^2_{\mathbf{w}\backslash\mathbf{x}}(\gamma - \tilde{\gamma})\}^{1/2}\varepsilon + \{(r_1 r_0)^{-1}(\gamma - \tilde{\gamma})'\mathbf{S}^2_{\mathbf{w}|\mathbf{x}}(\gamma - \tilde{\gamma})\}^{1/2} L_{K,a}. \tag{23}$$

From corollary 3 and theorem 6, expression (23) differs from the true asymptotic distribution (14) in $S^2_{\tau\backslash\mathbf{w}}$. We cannot estimate $S^2_{\tau\backslash\mathbf{w}}$ consistently by using the observed data. Consequently, the probability limit has wider quantile ranges than the true asymptotic distribution, which results in conservative confidence intervals. This kind of conservativeness is a feature of finite population causal inference known ever since Neyman's (1923) seminal work. See section A3.3 of the on-line supplementary material for a rigorous proof of the conservativeness. We discuss two special cases of theorem 6 below.

5.1.1. Special case: regression adjustment under the completely randomized experiment

Again, the CRE is rerandomization with $a = \infty$ and $\mathbf{x} = \emptyset$. Condition 4 holds automatically under the CRE. Theorem 6 immediately implies the following result.

Corollary 5. Under the CRE and condition 1, the estimated distribution of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ in expression (22) has the same limit as

$$\{V_{\tau\tau}(1 - R^2_{\tau,\mathbf{w}}) + S^2_{\tau\backslash\mathbf{w}} + (r_1 r_0)^{-1}(\gamma - \tilde{\gamma})'\mathbf{S}^2_{\mathbf{w}}(\gamma - \tilde{\gamma})\}^{1/2}\varepsilon. \tag{24}$$

5.1.2. Special case: no covariate adjustment under rerandomization

Using theorem 6 with $\beta_1 = \beta_0 = \mathbf{0}$, we can immediately obtain the probability limit of the estimated distribution of $\hat{\tau}$ under rerandomization.

Corollary 6. Under rerandomization and conditions 1 and 4, the estimated distribution of $\hat{\tau}$ in expression (22) has the same limit as

$$\{V_{\tau\tau}(1 - R^2_{\tau,\mathbf{x}}) + S^2_{\tau\backslash\mathbf{w}}\}^{1/2}\varepsilon + (V_{\tau\tau}R^2_{\tau,\mathbf{x}})^{1/2}L_{K,a}. \tag{25}$$

5.2. General scenarios with partial knowledge of the design

We now consider scenarios without condition 4. The analyser either does not have all the covariate information that is used in the design or does not know the balance criterion for rerandomization. We can still estimate $V_{\tau\tau}(\beta_1, \beta_0)$ by $\hat{V}_{\tau\tau}(\beta_1, \beta_0)$ in equation (20). However, we cannot consistently estimate $R^2_{\tau,\mathbf{x}}(\beta_1, \beta_0)$ because of incomplete information of the covariates that are used in the design stage. We can underestimate $R^2_{\tau,\mathbf{x}}(\beta_1, \beta_0)$ by 0, and then estimate the sampling distribution of $\hat{\tau}(\beta_1, \beta_0)$ by

$$\hat{V}^{1/2}_{\tau\tau}(\beta_1, \beta_0)\varepsilon. \tag{26}$$

An important fact is that the lengths of quantile ranges of the asymptotic distribution (7) are non-increasing in $R^2_{\tau,\mathbf{x}}(\beta_1, \beta_0)$. This fact guarantees that the estimated distribution (26) provides a conservative variance estimator of $\hat{\tau}(\beta_1, \beta_0)$ and conservative confidence intervals for τ. See section A3.3 of the on-line supplementary material for a rigorous proof of the conservativeness. Moreover, expression (26) equals expression (22) with $a = \infty$, the estimated distribution pretending that rerandomization does not happen in the design stage. Consequently, the probability limit of expression (26) equals expression (24), as the following theorem states.

Theorem 7. Under rerandomization and condition 1, the estimated distribution of $\hat{\tau}(\beta_1, \beta_0)$ in expression (26) has the same limit as expression (24).

From theorems 1 and 7, under rerandomization, the limit of the estimated distribution of $\hat{\tau}(\beta_1, \beta_0)$ in expression (26) differs from the corresponding true asymptotic distribution in $S^2_{\tau\backslash\mathbf{w}}$ and $R^2_{\tau,\mathbf{x}}(\beta_1, \beta_0)$. Neither $S^2_{\tau\backslash\mathbf{w}}$ nor $R^2_{\tau,\mathbf{x}}(\beta_1, \beta_0)$ has consistent estimators based on the observed data. The difficulty in estimating $S^2_{\tau\backslash\mathbf{w}}$ comes from the fact that for each unit we can observe at most one potential outcome, but the difficulty in estimating $R^2_{\tau,\mathbf{x}}(\beta_1, \beta_0)$ comes from the incomplete information of the design.

Compared with expression (22), the estimated distribution (26) has unnecessarily wider quantile ranges due to the lack of information for rerandomization, including $R^2_{\tau,\mathbf{x}}(\beta_1, \beta_0)$, K and a. In expression (26), we conduct conservative inference and consider the worse-case scenario, which is the CRE with $\mathbf{x} = \emptyset$ and $a = \infty$. In practice, we can also conduct a sensitivity analysis and check how the conclusions change as $R^2_{\tau,\mathbf{x}}(\beta_1, \beta_0)$, K and a vary. Without additional information, we still use expression (26) to construct confidence intervals in the next section.

6. Optimal adjustment based on the estimated distribution

$\mathcal{S}$-optimality is based on the uncertainty of the sampling distribution. As discussed in Section 5, we may not have consistent estimators for the sampling distributions, and we often report conservative confidence intervals based on the estimated distributions. Because the lengths of confidence intervals provide important measures of uncertainty in frequentist inference, we focus on the optimal choice of (β_1, β_0) based on the estimated distributions that were proposed in Section 5.

Under rerandomization, among the estimators in equation (1), a choice of (β_1, β_0) is optimal in terms of the estimated precision if the estimated distribution based on expression (22) or (26) has the shortest asymptotic $1-\alpha$ quantile range for any $\alpha \in (0,1)$. The corresponding estimated distribution then must have the smallest variance. Conversely, if the estimated distributions are Gaussian, then the distribution with the smallest estimated variance is optimal in terms of the estimated precision.

Apparently, the optimal adjusted estimator in terms of the estimated precision depends on the approach to constructing confidence intervals. In the ideal case, we want the probability limits of the estimated distributions to be identical to the true sampling distributions. Section 5, however, shows that this is generally impossible because of treatment effect heterogeneity or the information only known to the designer. In some cases, the confidence intervals based on expressions (22) and (26) are asymptotically exact given the analyser's observed information. For example, if the residual from the linear projection of individual treatment effect on covariates $\mathbf{w}$ is constant across all units, or equivalently $S^2_{\tau\backslash\mathbf{w}} = 0$, expression (22) provides asymptotically exact confidence intervals; if further the designer conducts CRE (i.e. rerandomization with $\mathbf{x} = \emptyset$ or $a = \infty$), expression (26) provides asymptotically exact confidence intervals. In general, these confidence intervals can be conservative and can be improved (Aronow *et al.*, 2014; Fogarty, 2018; Ding *et al.*, 2019). We focus on expressions (22) and (26) for their simplicity, and divide this section into two subsections in parallel with Section 5.

6.1. When the analyser knows all the information in the design stage

From theorem 6, we can obtain the optimal adjusted estimator in terms of the estimated precision.

Corollary 7. Under rerandomization and conditions 1 and 4, based on theorem 6, the optimal adjusted estimator in terms of the estimated precision is attainable when $\gamma = \tilde{\gamma}$ or $r_0\beta_1 + r_1\beta_0 = r_0\tilde{\beta}_1 + r_1\tilde{\beta}_0$, with the estimated distribution having the same limit as

$$\{V_{\tau\tau}(1 - R^2_{\tau,\mathbf{w}}) + S^2_{\tau\backslash\mathbf{w}}\}^{1/2}\varepsilon. \tag{27}$$

Again, from proposition 1, an optimal choice is $(\beta_1, \beta_0) = (\tilde{\beta}_1, \tilde{\beta}_0)$ under condition 4. Importantly, corollary 7 does not depend on covariates $\mathbf{x}$ and the threshold a that is used in rerandomization, and thus also holds under the CRE.

6.2. General scenarios without condition 4

From theorem 7, we can obtain the optimal adjusted estimator in terms of the estimated precision.

Corollary 8. Under rerandomization and condition 1, based on theorem 7, the optimal adjusted estimator in terms of the estimated precision is attainable when $\gamma = \tilde{\gamma}$ or $r_0\beta_1 + r_1\beta_0 = r_0\tilde{\beta}_1 + r_1\tilde{\beta}_0$, with the estimated distribution having the same limit as expression (27).

The optimal estimators in terms of the estimated precision are identical in corollaries 7 and 8, no matter whether the analyser knows all the information in the design or not. The optimal adjustment $\hat{\tau}(\tilde{\beta}_1, \tilde{\beta}_0)$ can never hurt the estimated precision. In contrast, Section 4.3 shows that $\hat{\tau}(\tilde{\beta}_1, \tilde{\beta}_0)$ may hurt the sampling precision in general. This is an important difference between the two notions of optimality in terms of the sampling precision and estimated precision.

Below we give some intuition for corollary 8. Under general scenarios without condition 4, the analyser does not know the information of the design. She or he pretends that the design

Table 1. Sampling standard errors se and average estimated standard errors multiplied by $n^{1/2}$ under rerandomization based on 10^5 simulated treatment assignments

Estimator	*Results for* $\rho = 0.9$		*Results for* $\rho = 0$	
	Sampling se	*Average estimated se*	*Sampling se*	*Average estimated se*
$\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$	1.47	1.86	2.95	3.61
$\hat{\tau}$	2.10	4.69	2.07	4.71

was a CRE and estimates the sampling distributions of the estimators under the CRE. Luckily, the resulting confidence intervals are still conservative. Dropping the term $S^2_{\tau\backslash\mathbf{w}}$, the estimated distribution converges to the sampling distribution under the CRE. Based on the discussion of $\mathcal{S}$-optimality under the CRE in Section 4.1.1, the adjusted estimator $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is also optimal in terms of the estimated precision.

6.2.1. Example 1 (continued)

We revisit example 1 and study the estimated distributions of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ and $\hat{\tau}$ under rerandomization. Because the estimated distributions are Gaussian from theorem 7, it suffices to report the estimated standard errors. Table 1 shows the sampling standard errors and the average estimated standard errors. On average, $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ results in shorter confidence intervals than does $\hat{\tau}$ when $\rho = 0$ or $\rho = 0.9$. Interestingly, when $\rho = 0$, $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ has less sampling precision as Fig. 1(b) and Table 1 show.

7. Gains from the analyser and the designer

In the design stage, we can use the CRE or rerandomization. In the analysis stage, we can use the unadjusted estimator $\hat{\tau}$ or the adjusted estimator $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$. On the basis of the results in previous sections, we now study the additional gains of the designer and analyser in the sampling precision, the estimated precision and the coverage probability.

7.1. Sampling precision

We first study the additional gains in the sampling precision from the analyser and the designer. We measure the additional gain of the analyser by comparing the asymptotic distributions of $\hat{\tau}$ and $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ under rerandomization. We measure the additional gain of the designer by comparing the asymptotic distributions of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ under the CRE and rerandomization. Similarly to Section 4, we consider different scenarios based on the relative amount of covariate information for the analyser and the designer. Let $v_{K,a} = P(\chi^2_{K+2} \leqslant a)/P(\chi^2_K \leqslant a) \leqslant 1$ be the variance of $L_{k,a}$ (Morgan and Rubin, 2012) and $q_{1-\alpha/2}(\rho^2)$ be the $(1-\alpha/2)$th quantile of $(1-\rho^2)^{1/2}\varepsilon + |\rho| L_{K,a}$.

7.1.1. The analyser has no less covariate information than the designer

First, we measure the additional gain of the analyser.

Corollary 9. Under rerandomization and conditions 1 and 2, compare $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ with $\hat{\tau}$. The

percentage reduction in the asymptotic variance is $\{R^2_{\tau,\mathbf{w}} - (1-v_{K,a})R^2_{\tau,\mathbf{x}}\}/\{1-(1-v_{K,a})R^2_{\tau,\mathbf{x}}\}$. For any $\alpha \in (0,1)$, the percentage reduction in the length of the asymptotic $1-\alpha$ quantile range is $1-(1-R^2_{\tau,\mathbf{w}})^{1/2} q_{1-\alpha/2}(0)/q_{1-\alpha/2}(R^2_{\tau,\mathbf{x}})$. Both percentage reductions are non-negative and non-decreasing in $R^2_{\tau,\mathbf{w}}$.

From corollary 9, the gain from the analyser through regression adjustment is non-decreasing in the analyser's covariate information. Both percentage reductions in corollary 9 converge to 1 as $R^2_{\tau,\mathbf{w}}$ converges to 1.

Second, we measure the additional gain of the designer. Section 4.1.1 demonstrates that $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ has the same asymptotic distribution under the CRE and rerandomization. Therefore, under condition 2, the gain from the designer is zero. Nevertheless, this also implies that using rerandomization in the design will not hurt the sampling precision of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$. Moreover, under rerandomization, we can use the additional covariate information, in the same way as in the CRE, to improve the sampling precision.

7.1.2. The analyser has no more covariate information than the designer

First, we measure the additional gain of the analyser. Both the asymptotic distributions of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ and $\hat{\tau}$ are linear combinations of ε and $L_{K,a}$. The coefficients of ε are identical, but the coefficient of $L_{K,a}$ for $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is smaller than that for $\hat{\tau}$.

Corollary 10. Under rerandomization and conditions 1 and 3, compare $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ with $\hat{\tau}$. The percentage reduction in the asymptotic variance is $v_{K,a}R^2_{\tau,\mathbf{w}}/\{1-(1-v_{K,a})R^2_{\tau,\mathbf{x}}\}$. For any $\alpha \in (0,1)$, the percentage reduction in the length of the asymptotic $1-\alpha$ quantile range is $1-(1-R^2_{\tau,\mathbf{w}})^{1/2} q_{1-\alpha/2}(\rho^2_{\tau,\,\mathbf{x}\setminus\mathbf{w}})/q_{1-\alpha/2}(R^2_{\tau,\mathbf{x}})$. Both percentage reductions are non-negative and non-decreasing in $R^2_{\tau,\mathbf{w}}$.

From corollary 10, the improvement from regression adjustment is non-decreasing in the analyser's covariate information. However, this improvment is small when the designer uses a small threshold a for rerandomization. Both percentage reductions in corollary 10 converge to 0 as a converges to 0. Intuitively, when the designer uses rerandomization with a small threshold, she or he has used more covariate information thoroughly in the design, and thus the analyser has only a small additional gain through regression adjustment.

Second, we measure the additional gain of the designer.

Corollary 11. Under conditions 1 and 3, compare $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ under rerandomization with that under the CRE. The percentage reduction in the asymptotic variance is $(1-v_{K,a})\rho^2_{\tau,\,\mathbf{x}\setminus\mathbf{w}}$. For any $\alpha \in (0,1)$, the percentage reduction in the length of the asymptotic $1-\alpha$ quantile range is $1-q_{1-\alpha/2}(\rho^2_{\tau,\mathbf{x}\setminus\mathbf{w}})/q_{1-\alpha/2}(0)$. Both percentage reductions are non-negative and non-decreasing in $R^2_{\tau,\mathbf{x}}$.

From corollary 11, the gain from the designer through rerandomization is non-decreasing in the designer's covariate information. The gain from the designer is substantial when $R^2_{\tau,\,\mathbf{x}}$ is large and the threshold for rerandomization is small. Both percentage reductions in corollary 11 converge to 1 as $R^2_{\tau,\,\mathbf{x}} \to 1$ and $a \to 0$.

Remark 3. Consider the special case where the designer and the analyser have the same covariates ($\mathbf{x} = \mathbf{w}$). The additional gain from the analyser is small given that the designer uses rerandomization with a small threshold a, and so is the additional gain from the designer given that the analyser uses the $\mathcal{S}$-optimal $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$.

7.1.3. General scenarios

The complexity that was discussed in Section 4.3 makes it difficult to evaluate the additional gains from the analyser and the designer. Given any adjusted estimator, the designer can always use rerandomization to reduce the asymptotic variance and the lengths of asymptotic quantile ranges (Li *et al.*, 2018). For the analyser, in general, the performance of regression adjustment under rerandomization depends on the covariates that are used in the design, and thus the analyser does not know the optimal adjusted estimator among equation (1). For instance, without the covariate information that is used in the design, the analyser is not sure whether $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ has smaller asymptotic variance than $\hat{\tau}$. Example 1 shows two cases where $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ increases and decreases the sampling precision.

7.2. Estimated precision

We now study the additional gains in the asymptotic estimated precision from the analyser and the designer. We measure the additional gain of the analyser by comparing the estimated distributions of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ and $\hat{\tau}$ under rerandomization. We measure the additional gain of the designer by comparing the estimated distributions of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ under the CRE and rerandomization. Similarly to Section 6, we consider two scenarios based on whether the analyser has full knowledge of the design or not. Let $\kappa = 1 + V_{\tau\mathcal{T}}^{-1} S^2_{\tau\backslash\mathbf{w}} \geqslant 1$, which reduces to 1 when $S^2_{\tau\backslash\mathbf{w}} = 0$, i.e. the adjusted individual treatment effect $\tau_i(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) = Y_i(1; \tilde{\boldsymbol{\beta}}_1) - Y_i(0; \tilde{\boldsymbol{\beta}}_0)$ is constant for all units.

7.2.1. When the analyser knows all the information in the design stage

First, we measure the additional gain of the analyser.

Corollary 12. Under rerandomization and conditions 1 and 4, compare the probability limit of the estimated distribution of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ with that of $\hat{\tau}$ based on expression (22). The percentage reduction in the variance is $\{R^2_{\tau,\mathbf{w}} - (1 - v_{K,a}) R^2_{\tau,\mathbf{x}}\} / \{\kappa - (1 - v_{K,a}) R^2_{\tau,\mathbf{x}}\}$. For $\alpha \in (0, 1)$, the percentage reduction in the length of the $1 - \alpha$ quantile range is $1 - (1 - R^2_{\tau,\mathbf{w}}/\kappa)^{1/2} \times q_{1-\alpha/2}(0) / q_{1-\alpha/2}(R^2_{\tau,\mathbf{x}}/\kappa)$. Both percentage reductions are non-negative and non-decreasing in $R^2_{\tau,\mathbf{w}}$.

From corollary 12, the gain from regression adjustment is non-decreasing in the analyser's covariate information. Both percentage reductions in corollary 12 converge to 1 as $R^2_{\tau,\mathbf{w}} \to 1$ and $\kappa \to 1$.

Second, we measure the additional gain of the designer. From corollary 7 and the comment after it, the probability limits of the estimated distributions of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ are identical under both designs. Therefore, the gain from the designer is zero.

Remark 4. Consider the special case where the analyser has the same covariate information as the designer and knows the balance criterion in the design. As discussed above, the designer has no gain. Based on corollaries 6 and 7, with a small threshold a, the estimated distributions of $\hat{\tau}$ and $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ have approximately the same probability limit. Therefore, the analyser has small additional gain.

7.2.2. General scenarios without condition 4

First, we measure the additional gain of the analyser.

Corollary 13. Under rerandomization and condition 1, compare the probability limit of the estimated distribution of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ with that of $\hat{\tau}$ based on expression (26). The percentage

reduction in the variance is $R^2_{\tau,\mathbf{w}}/\kappa$. For any $\alpha \in (0,1)$, the percentage reduction in the length of the $1-\alpha$ quantile range is $1-(1-R^2_{\tau,\mathbf{w}}/\kappa)^{1/2}$. Both percentage reductions are non-negative and non-decreasing in $R^2_{\tau,\mathbf{w}}$.

Both percentage reductions in corollary 13 converge to 1 as $R^2_{\tau,\mathbf{w}} \to 1$ and $\kappa \to 1$, and they are larger than or equal to those in corollary 12.

Second, we measure the additional gain of the designer. From corollary 5 and theorem 7, the estimated distributions of any adjusted estimator in equation (1) have the same probability limit under both designs. Therefore, the gain from the designer is zero. Nevertheless, using rerandomization will not hurt the estimated precision of the treatment effect estimators, and we can use covariates in the analysis in the same way as in the CRE to improve the estimated precision.

7.3. Coverage probabilities

From Sections 6 and 7.2,

(a) $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is optimal in terms of the estimated precision no matter whether the analyser knows all the information of the design or not, and
(b) the designer provides no gain in the estimated precision of the $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$.

From result (a), under rerandomization, the analyser can never increase the asymptotic lengths of the confidence intervals by using $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ instead of $\hat{\tau}$. Therefore, we do not measure the additional gain of the analyser in coverage probabilities. From result (b), the asymptotic lengths of the confidence intervals based on $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ are the same under the CRE and rerandomization. However, we shall show shortly that the designer can help to improve the coverage probabilities of the confidence intervals based on $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$.

7.3.1. When the analyser knows all the information in the design stage

From Sections 7.1.1 and 7.2.1, $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ has the same sampling precision and estimated precision under rerandomization and the CRE. Therefore, the coverage probabilities of the associated confidence intervals are asymptotically the same under the CRE and rerandomization. This implies that the designer provides no gain for the coverage probabilities of the confidence intervals based on $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$. We formally state the results as follows.

Corollary 14. Under conditions 1 and 4, compare the confidence intervals based on $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ and expression (22) under the CRE and rerandomization. Their lengths are asymptotically the same after being scaled by $n^{1/2}$, and they have the same asymptotic coverage probability.

7.3.2. General scenarios without condition 4

From Sections 7.1.3 and 7.2.2, $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ in equation (1) with any $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$ has better sampling precision under rerandomization than under the CRE, but it has the same estimated precision under rerandomization and the CRE. Therefore, the confidence intervals based on $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ under rerandomization have higher coverage probabilities than that under the CRE. We give a formal statement below.

Corollary 15. Under condition 1, compare the confidence intervals based on $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ and expression (26) under the CRE and rerandomization. Their lengths are asymptotically the same after being scaled by $n^{1/2}$. However, the asymptotic coverage probability under rerandomization is larger than or equal to that under the CRE.

In corollary 15, the confidence intervals under both rerandomization and the CRE are asymptotically valid and of the same length, but the interval under rerandomization has higher coverage probabilities and is more conservative. In particular, for any $\alpha \in (0,1)$ and any $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$, as $R^2_{\tau,\mathbf{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \to 1$, $S^2_{\tau\backslash\mathbf{w}} \to 0$ and $a \to 0$, the asymptotic coverage probabilities of the $1-\alpha$ confidence intervals are 1 and $1-\alpha$ under rerandomization and the CRE respectively. Corollary 15 holds for any adjusted estimator and thus holds for $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$. Therefore, under general scenarios without condition 4, the designer can provide a substantial gain in coverage probabilities of confidence intervals. This gives another justification for using rerandomization.

8. Unification and practical suggestions

8.1. Unification

In total, there are four combinations in the design and analysis of experiments. Fig. 2 summarizes the sampling distributions and the probability limits of the estimated distributions for all combinations.

Neyman (1923) started the literature by discussing the property of $\hat{\tau}$ under the CRE. Lin (2013) showed that $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ improves $\hat{\tau}$ in terms of the sampling precision and estimated precision under the CRE. Arrow (i) in Fig. 2 illustrates this improvement. Li *et al.* (2018) showed that rerandomization improves the CRE in terms of the sampling precision and the estimated precision of $\hat{\tau}$. Arrow (ii) in Fig. 2 illustrates this improvement. Interestingly, $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ under the CRE and $\hat{\tau}$ under rerandomization have almost identical asymptotic sampling distributions and estimated distributions, if we use the same sets of covariates and $a \approx 0$ in rerandomization.

However, both Lin (2013) and Li *et al.* (2018) compared suboptimal strategies. We evaluated the additional gain from the analyser given that the designer uses rerandomization. Arrow (iii) in Fig. 2 illustrates this improvement. We also evaluated the additional gain from the designer given that the analyser uses $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$. Arrow (iv) in Fig. 2 illustrates this improvement. Table 2 summarizes the results under all scenarios. We make the following conclusions.

(a) Compare the analyser and the designer on the basis of the sampling precision. From the first two rows of Table 2, when one has more covariate information than the other, the one with more covariate information provides a substantial additional gain, whereas the other provides negligible additional gain.
(b) Compare the analyser and the designer on the basis of the estimated precision. From

		Analysis		
		$\hat{\tau}$		$\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$
Design	CRE	$V_{\tau\tau}^{1/2} \cdot \varepsilon$ [$\tilde{V}_{\tau\tau}^{1/2} \cdot \varepsilon$]	Lin (2013) (i)	(15) [(27)]
		Li et al. (2018) (ii)		(iv)
	Rerandomization	(9) [(25) or $\tilde{V}_{\tau\tau}^{1/2} \cdot \varepsilon$]	(iii)	(7) & (12) with $(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ [(27)]

Fig. 2. Design and analysis strategies: the formulae without square brackets correspond to asymptotic distributions, and those with square brackets correspond to probability limits of the estimated distributions with $\tilde{V}_{\tau\tau} = V_{\tau\tau} + S^2_{\tau\backslash\mathbf{w}}$; the probability limits of the estimated distributions of $\hat{\tau}$ under rerandomization have two forms, depending on whether condition 4 holds or not

Table 2. Additional gains†

Covariates information	*Balance criterion*	*Optimal adjustment*		*Additional gain from analyser*		*Additional gain from designer*	
		Sampling	*Estimated*	*Sampling*	*Estimated*	*Sampling*	*Estimated*
$A \geqslant D$	Unknown	$(\tilde{\beta}_1, \tilde{\beta}_0)$	$(\tilde{\beta}_1, \tilde{\beta}_0)$	✓○	✓○	○	○
$A \leqslant D$	Unknown	$(\tilde{\beta}_1, \tilde{\beta}_0)$	$(\tilde{\beta}_1, \tilde{\beta}_0)$	○	✓○	✓○	○
$A = D$	Unknown	$(\tilde{\beta}_1, \tilde{\beta}_0)$	$(\tilde{\beta}_1, \tilde{\beta}_0)$	○	✓○	○	○
$A = D$	Known	$(\tilde{\beta}_1, \tilde{\beta}_0)$	$(\tilde{\beta}_1, \tilde{\beta}_0)$	○	○	○	○
General	Unknown	$\gamma \approx \tilde{\gamma}_{\text{res}}$	$(\tilde{\beta}_1, \tilde{\beta}_0)$	✓○×	✓○	✓○	○

†In the first column, $A \geqslant D$, $A \leqslant D$ and $A = D$ denote that the analyser has no less (i.e. condition 2), no more (i.e. condition 3) and the same (i.e. both conditions 2 and 3) covariate information compared with the designer. The second column shows whether the analyser knows the balance criterion in the design (i.e. condition 4). The third and fourth columns show the optimal coefficients. '✓' denotes a substantial gain, '○' denotes no gain or negligible gain and '×' denotes a negative gain.

the sixth and eighth columns of Table 2, the additional gain from the analyser can be substantial, whereas the additional gain from the designer is negligible in general.

(c) Consider the special case where the analyser has the same covariate information as the designer and knows the balance criterion in the design. From the fourth row of Table 2, the additional gain from either the analyser or the designer is negligible.

(d) From the last row of Table 2, the analyser may hurt the sampling precision through regression adjustment but can provide a substantial gain in the estimated precision. The designer can improve sampling precision of any adjusted estimator and does not hurt the estimated precision. Therefore, although the designer cannot shorten the confidence intervals, she or he can increase the coverage probabilities.

8.2. Rerandomization, Lin's estimator and the Huber–White variance estimator

On the basis of the summary in Section 8.1, we recommend the use of rerandomization in the design and $\hat{\tau}(\tilde{\beta}_1, \tilde{\beta}_0)$ in the analysis, which has better estimated precision and coverage property. However, some practical issues remain.

First, we need to estimate the population OLS coefficients $\tilde{\beta}_1$ and $\tilde{\beta}_0$. Under both the CRE and rerandomization, we can use their sample analogues as consistent estimators, with $\hat{\beta}_1$ and $\hat{\beta}_0$ being the coefficients of $\mathbf{w}$ in the OLS fit of Y on $\mathbf{w}$ under the treatment and control respectively. The corresponding adjusted estimator $\hat{\tau}(\hat{\beta}_1, \hat{\beta}_0)$ numerically equals the coefficient of Z in the OLS fit of Y on Z, $\mathbf{w}$ and $Z \times \mathbf{w}$, i.e. it is Lin's (2013) estimator. Replacing $\tilde{\beta}_1$ and $\tilde{\beta}_0$ by their sample analogues does not change the asymptotic distribution. Informally speaking, $\hat{\tau}(\tilde{\beta}_1, \tilde{\beta}_0)$ and $\hat{\tau}(\hat{\beta}_1, \hat{\beta}_0)$ have the same asymptotic behaviour and optimality. The following corollary is a formal statement.

Proposition 3. Under rerandomization and condition 1, $\hat{\tau}(\tilde{\beta}_1, \tilde{\beta}_0)$ and $\hat{\tau}(\hat{\beta}_1, \hat{\beta}_0)$ have the same asymptotic distributions and the same probability limits of the estimated distributions. Thus, among equation (1), $\hat{\tau}(\hat{\beta}_1, \hat{\beta}_0)$ is an $\mathcal{S}$-optimal estimator under condition 2 or 3 and is always optimal in terms of the estimated precision based on expressions (22) and (26).

Second, the estimated distributions (22) and (26) are identical with or without condition 4 because $\hat{R}^2_{\tau,\mathbf{x}}(\hat{\beta}_1, \hat{\beta}_0)$ based on expression (21) equals 0 under condition 4. Moreover, the variance

estimator $\hat{V}_{\tau\tau}(\hat{\beta}_1, \hat{\beta}_0)$ based on equation (20) is asymptotically equivalent to the Huber–White variance estimator $\hat{V}_{\mathrm{HW}}$ of the coefficient of Z from the OLS fit of Y on Z, $\mathbf{w}$ and $Z \times \mathbf{w}$.

Theorem 8. Under rerandomization and condition 1, $\hat{V}_{\tau\tau}(\hat{\beta}_1, \hat{\beta}_0) - \hat{V}_{\mathrm{HW}} \to 0$ in probability.

Theorem 8 extends Lin's (2013) result for the CRE to rerandomization. It requires only condition 1, but Lin (2013) required higher order moment conditions. We finally construct the Wald-type confidence intervals based on a Gaussian approximation. The statistical inference based on $\hat{\tau}(\hat{\beta}_1, \hat{\beta}_0)$, including variance estimation and confidence interval construction, is always the same no matter whether the design is a CRE or rerandomization and no matter whether the analyser knows all the information of the design or not.

From the above, using rerandomization and $\hat{\tau}(\hat{\beta}_1, \hat{\beta}_0)$ enjoys the optimal estimated precision and improves the coverage property, and the associated statistical inference can be conveniently implemented through the OLS fit and Huber–White variance estimate.

9. Illustration

9.1. A simulation study

We conduct a simulation study to investigate the performance of the asymptotic approximation and the coverage properties of the confidence intervals in finite samples. We generate the data in the same way as in example 1 with $\rho = 0$ and vary the sample size n from 100 to 1000. For each simulated data set, we generate rerandomization based on covariate x, where the threshold a is the 0.001th quantile of the χ^2_1 random variable. Figs 3(a) and 3(b) show the histograms of $\hat{\tau}$ and $\hat{\tau}(\hat{\beta}_1, \hat{\beta}_0)$ based on covariate w. From Figs 3(a) and 3(b), the asymptotic approximation works fairly well. We then construct 95% confidence intervals for the average treatment effect, using the estimated distribution (26) with either $\hat{\tau}$ or $\hat{\tau}(\hat{\beta}_1, \hat{\beta}_0)$. From Figs 3(c) and 3(d), the confidence intervals based on the estimator adjusted for w are shorter than that based on $\hat{\tau}$, and both confidence intervals are conservative with coverage probabilities that are larger than the nominal level, because of the analyser's incomplete information of the design. We further consider $\hat{\tau}(\hat{\beta}_1, \hat{\beta}_0)$ based on (x, w), assuming that the analyser has access to the covariate x in the design. The corresponding confidence interval is even shorter and becomes asymptotically exact, because of the additive treatment effects in the data-generating process. From Fig. 3(d), the coverage probabilities are close to the nominal level as the sample size increases. Even when the sample size is small, the confidence interval works fairly well with coverage probability at least 94%.

9.2. The 'opportunity knocks' experiment

The 'opportunity knocks' experiment (Angrist *et al.*, 2014) aims at evaluating the effect of a financial incentive demonstration programme on college students' academic performance. The experiment includes first- and second-year students who apply for the financial aid at a large Canadian commuter university. These students were randomly assigned to treatment and control groups. Students in the treated group have peer advisors and receive a cash reward for attaining certain grades.

We use this data set to illustrate rerandomization and regression adjustment. We consider the second-year students and choose the outcome to be the average grade for the semester right after the experiment. We exclude students with missing outcomes or covariates, resulting in a treatment group of size 199 and a control group of size 369. We evaluate the repeated sampling properties of the adjusted estimators under rerandomization, which depend on all the potential outcomes. However, half of the potential outcomes are missing from the observed data. To make

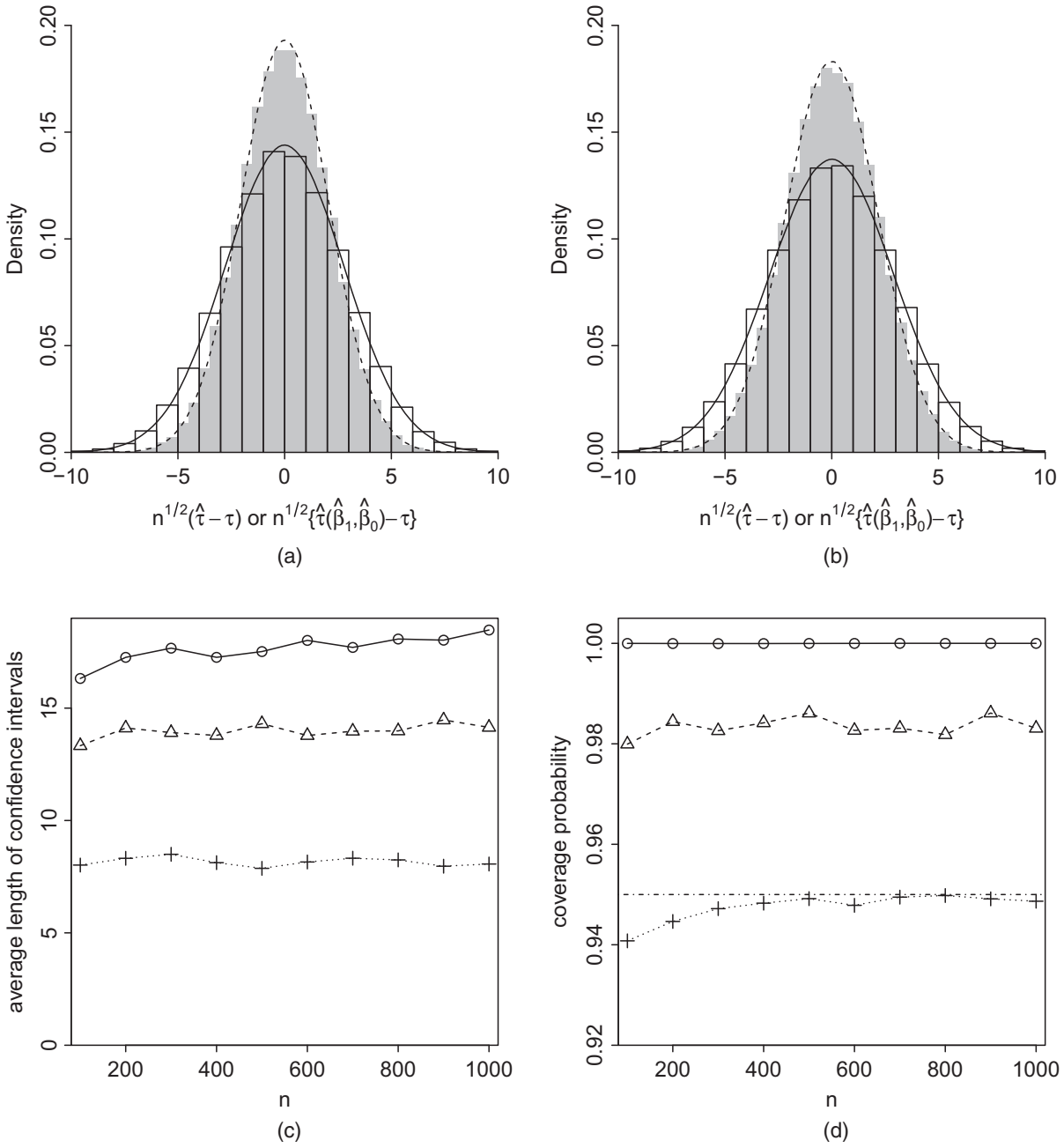

Fig. 3. Histograms of (a) ($n = 100$) and (b) ($n = 300$) of $n^{1/2}\{\hat{\tau}(\hat{\beta}_1, \hat{\beta}_0) - \tau\}$ and $n^{1/2}(\hat{\tau} - \tau)$ under rerandomization based on 10^5 simulated treatment assignments (□, adjusted; ■, unadjusted; ———, adjusted; - - - - -, unadjusted); (c) average lengths and (d) coverage probabilities of 95% confidence intervals constructed from expression (26) for three estimators (△, adjusted estimator based on w; +, adjusted estimator based on x and w; ○, unadjusted)

the simulation more realistic, we impute all the missing potential outcomes on the basis of simple model fitting. We first fit a linear model of the observed outcome on the treatment indicator and covariates within each stratum classified by sex and high school grade point average. We then impute the missing potential outcomes by using the fitted linear model.

We conduct rerandomization with two covariates and choose threshold a to be the 0.005th quantile of χ^2_2. For the covariates in the design and the analysis, we consider the following two cases:

(a) the covariates in the design are sex and high school grade, and the covariates in the analysis are whether the mother or father is a college graduate, whether they correctly answer the first or second question in a survey, whether their mother tongue is English and their grade point average in the previous year;
(b) the covariates in the design and analysis are the same as in case (a), except that we switch the high school grade to the analysis stage and switch the grade point average in the previous year to the design stage.

We first consider the sampling precision. Table 3 shows the coefficients of ε in the asymptotic distributions. We omit the coefficients of $L_{K,a}$ because the ε-components are the dominating terms in the asymptotic distributions. Fig. 4 shows the histograms of $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ and $\hat{\tau}$ under

Table 3. Sampling precision, estimated precision and coverage probabilities for $\hat{\tau}$ and $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ based on 10^5 simulated treatment assignments†

	Results for estimator $\hat{\tau}$ and the following designs:		*Results for estimator $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ and the following designs:*	
	Rerandomization	*CRE*	*Rerandomization*	*CRE*
Sampling				
Case (a)	13.88	16.73	9.83	9.86
Case (b)	9.90	16.73	11.62	13.80
Estimated				
Case (a)	18.56 (98.5%)	18.56 (96.0%)	12.75 (97.1%)	12.75 (97.0%)
Case (b)	18.57 (99.9%)	18.56 (96.0%)	15.95 (98.3%)	15.94 (96.1%)

†The first two rows show the coefficients of ε in the asymptotic distributions of $\hat{\tau}$ and $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ in expression (7). The last two rows show the average estimated standard errors multiplied by $n^{1/2}$ with the coverage probabilities of 95% confidence intervals in parentheses.

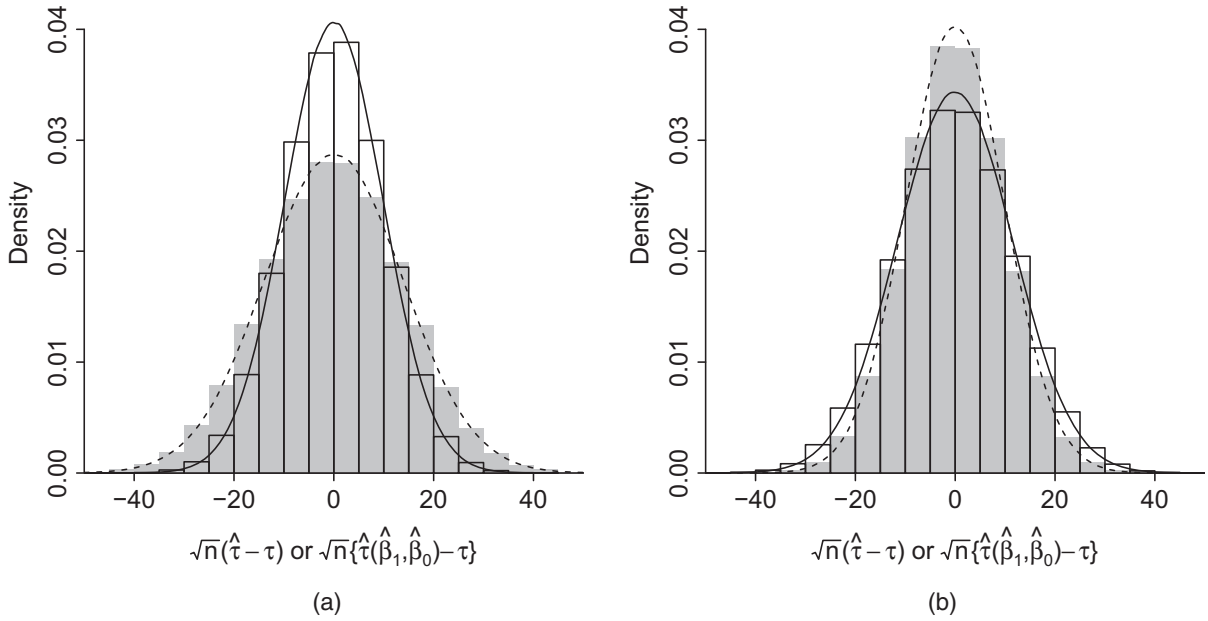

Fig. 4. Histograms of $\hat{\tau}$ and $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ under rerandomization based on 10^5 simulated treatment assignments for (a) case (a) and (b) case (b) (□, adjusted; ■, unadjusted; ———, adjusted; - - - - -, unadjusted)

rerandomization. In Table 3, compared with the second column, the reduction in coefficients in the first column shows the gain from the designer alone, and the reduction in the last column shows the gain from the analyser alone. The magnitude of the reduction suggests the relative amount of the covariate information of the designer and analyser. In case (a), the first row of Table 3 shows that $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ under the CRE is more precise than $\hat{\tau}$ under rerandomization. This holds because $R^2_{\tau,\mathbf{w}} = 0.65 \geqslant R^2_{\tau,\mathbf{x}} = 0.31$. Fig. 4(a) shows that $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ outperforms $\hat{\tau}$, coherent with theorem 5. In case (b), the second row of Table 3 shows that $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ under the CRE is less precise than $\hat{\tau}$ under rerandomization. This is because $R^2_{\tau,\mathbf{w}} = 0.32 < R^2_{\tau,\mathbf{x}} = 0.65$. Fig. 4(b) shows that $\hat{\tau}$ outperforms $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ under rerandomization.

We then consider the estimated precision. Because the estimated distributions in expression (26) are Gaussian, it suffices to compare the estimated standard errors. Table 3 shows the average estimated standard errors and the coverage probabilities of the 95% confidence intervals. In both cases, $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ has almost the same estimated precision under rerandomization and the CRE. So does $\hat{\tau}$. However, the 95% confidence intervals under rerandomization have higher coverage probability, which is coherent with corollary 15. Moreover, $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ always has higher estimated precision than $\hat{\tau}$, which is coherent with the optimality results in corollary 8.

10. Discussion

In sum, regression adjustment can improve the estimated precision but may hurt the sampling precision, and rerandomization can improve the sampling precision and never hurts the estimated precision. The resulting adjusted estimator is optimal in terms of the estimated precision among all linearly adjusted estimators in equation (1), has lower sampling variability under rerandomization than it would have had under the CRE, and the corresponding confidence intervals have higher coverage probabilities than that under the CRE. Therefore, in practice, we recommend the use of rerandomization in the design and the use of Lin's (2013) estimator in the analysis followed by the Huber–White robust standard error. Importantly, the analyser should communicate with the designer, asking for detailed covariate information and assignment mechanism in the design stage.

For the analysis, we focused on inferring the average treatment effect by using regression adjustment. It is interesting to extend the discussion to covariate adjustment in more complicated settings, such as high dimensional covariates (Bloniarz *et al.*, 2016; Wager *et al.*, 2016; Lei and Ding, 2018), logistic regression for binary outcomes (Zhang *et al.*, 2008; Freedman, 2008d; Moore and van der Laan, 2009; Moore *et al.*, 2011) and adjustment using machine learning methods (Bloniarz *et al.*, 2016; Wager *et al.*, 2016; Wu and Gagnon-Bartsch, 2018). It is also important to consider covariate adjustment for general non-linear estimands (Zhang *et al.*, 2008; Jiang *et al.*, 2019; Tian *et al.*, 2019) and general designs (Middleton, 2018), such as blocking (Miratrix *et al.*, 2013; Bugni *et al.*, 2018), matched pairs (Fogarty, 2018), and factorial designs (Lu, 2016).

For the design, we focused on rerandomization using the Mahalanobis distance. It is conceptually straightforward to extend the results to rerandomization with tiers of covariates (Morgan and Rubin, 2015; Li *et al.*, 2018). Recently, Zhou *et al.* (2018) discussed sequential rerandomization, and Li *et al.* (2019) discussed rerandomization in 2^K factorial experiments with tiers of both covariates and factorial effects. It is important to discuss regression adjustment after these rerandomizations.

The relationship between blocking and post-stratification for discrete covariates is analogous to the relationship between rerandomization and regression adjustment for general covariates.

When the number of blocks is small compared with the sample size, our results suggest conducting post-stratification, or equivalently an OLS fit of the outcome on treatment, block indicator and their interactions, followed by the Huber–White robust standard error. When the number of blocks is large, Miratrix *et al.* (2013) showed that post-stratification can be worse than blocking, which sheds light on the possible advantage of rerandomization over regression adjustment with a large number of covariates. In this case, although deriving the asymptotic properties of rerandomization is challenging, it is still straightforward to conduct Fisher randomization tests.

Acknowledgements

We thank the Associate Editor and two reviewers for constructive comments. Peng Ding gratefully acknowledges financial support from the National Science Foundation (grant 1713152).

References

Angrist, J., Oreopoulos, P. and Williams, T. (2014) When opportunity knocks, who answers?: New evidence on college achievement awards. *J. Hum. Resour.*, **49**, 572–610.

Aronow, P. M., Green, D. P. and Lee, D. K. K. (2014) Sharp bounds on the variance in randomized experiments. *Ann. Statist.*, **42**, 850–871.

Athey, S. and Imbens, G. W. (2017) The econometrics of randomized experiments. In *Handbook of Economic Field Experiments*, vol. 1 (eds A. Banerjee and E. Duflo), ch. 3, pp. 73–140. Amsterdam: North-Holland.

Bloniarz, A., Liu, H., Zhang, C. H., Sekhon, J. and Yu, B. (2016) Lasso adjustments of treatment effect estimates in randomized experiments. *Proc. Natn. Acad. Sci. USA*, **113**, 7383–7390.

Box, G. E. P., Hunter, J. S. and Hunter, W. G. (2005) *Statistics for Experimenters: Design, Innovation, and Discovery*. New York: Wiley-Interscience.

Bruhn, M. and McKenzie, D. (2009) In pursuit of balance: randomization in practice in development field experiments. *Am. Econ. J.*, **1**, 200–232.

Bugni, F. A., Canay, I. A. and Shaikh, A. M. (2018) Inference under covariate-adaptive randomization. *J. Am. Statist. Ass.*, **113**, 1784–1796.

Casella, G. and Berger, R. L. (2002) *Statistical Inference*, 2nd edn. Pacific Grove: Duxbury.

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

Cox, D. R. (2007) Applied statistics: a review. *Ann. Appl. Statist.*, **1**, 1–16.

Cox, D. R. (2009) Randomization in the design of experiments. *Int. Statist. Rev.*, **77**, 415–429.

Cox, D. R. and Reid, N. (2000) *The Theory of the Design of Experiments*. Boca Raton: CRC Press.

Ding, P., Feller, A. and Miratrix, L. (2019) Decomposing treatment effect variation. *J. Am. Statist. Ass.*, **114**, 304–317.

Duflo, E., Glennerster, R. and Kremer, M. (2007) Using randomization in development economics research: a toolkit. In *Handbook of Development Economics*, vol. 4 (eds J. A. Strauss and T. P. Schultz), ch. 61, pp. 3895–3962. New York: Elsevier.

Fisher, R. A. (1926) The arrangement of field experiments. *J. Min. Agric. Gt Br.*, **33**, 503–513.

Fisher, R. A. (1935) *The Design of Experiments*, 1st edn. Edinburgh: Oliver and Boyd.

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

Freedman, D. A. (2008a) On regression adjustments in experiments with several treatments. *Ann. Appl. Statist.*, **2**, 176–196.

Freedman, D. A. (2008b) On regression adjustments to experimental data. *Adv. Appl. Math.*, **40**, 180–193.

Freedman, D. A. (2008c) Oasis or mirage? *Chance,* **21,** 59–61.

Freedman, D. A. (2008d) Randomization does not justify logistic regression. *Statist. Sci.*, **23**, 237–249.

Gerber, A. S. and Green, D. P. (2012) *Field Experiments: Design, Analysis, and Interpretation*. New York: Norton.

Higgins, M. J., Sävje, F. and Sekhon, J. S. (2016) Improving massive experiments with threshold blocking. *Proc. Natn. Acad. Sci. USA*, **113**, 7369–7376.

Jiang, F., Tian, L., Fu, H., Hasegawa, T. and Wei, L. J. (2019) Robust alternatives to ANCOVA for estimating the treatment effect via a randomized comparative study. *J. Am. Statist. Ass.*, to be published.

Ke, C., Wang, J., Zhang, C., Jiang, Q. and Snapinn, S. (2017) On errors in stratified randomization. *Statist. Biopharm. Res.*, **9**, 225–233.

Kempthorne, O. (1952) *The Design and Analysis of Experiments*. New York: Wiley.

Koch, G. G., Tangen, C. M., Jung, J.-W. and Amara, I. A. (1998) Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. *Statist. Med.*, **17**, 1863–1892.
Kohavi, R. and Longbotham, R. (2017) Online controlled experiments and a/b testing. In *Encyclopedia of Machine Learning and Data Mining* (eds C. Sammut and G. I. Webb), pp. 922–929. New York: Springer.
Lei, L. and Ding, P. (2018) Regression adjustment in completely randomized experiments with a diverging number of covariates. *Preprint arXiv:1806.07585*.
Leon, S., Tsiatis, A. A. and Davidian, M. (2003) Semiparametric estimation of treatment effect in a pretest-posttest study. *Biometrics*, **59**, 1046–1055.
Li, X. and Ding, P. (2017) General forms of finite population central limit theorems with applications to causal inference. *J. Am. Statist. Ass.*, **112**, 1759–1769.
Li, X., Ding, P. and Rubin, D. B. (2018) Asymptotic theory of rerandomization in treatment-control experiments. *Proc. Natn. Acad. Sci. USA*, **115**, 9157–9162.
Li, X., Ding, P. and Rubin, D. B. (2019) Rerandomization in 2^K factorial experiments. *Ann. Statist.*, to be published.
Lin, W. (2013) Agnostic notes on regression adjustments to experimental data: reexamining Freedman's critique. *Ann. Appl. Statist.*, **7**, 295–318.
Lu, J. (2016) Covariate adjustment in randomization-based causal inference for 2^K factorial designs. *Statist. Probab. Lett.*, **119**, 11–20.
Middleton, J. A. (2018) A unified theory of regression adjustment for design-based inference. *Preprint arXiv:1803.06011*.
Miratrix, L. W., Sekhon, J. S. and Yu, B. (2013) Adjusting treatment effect estimates by post-stratification in randomized experiments. *J. R. Statist. Soc.* B, **75**, 369–396.
Moore, K. L. and van der Laan, M. J. (2009) Covariate adjustment in randomized trials with binary outcomes: targeted maximum likelihood estimation. *Statist. Med.*, **28**, 39–64.
Moore, K. L., Neugebauer, R., Valappil, T. and van der Laan, M. J. (2011) Robust extraction of covariate information to improve estimation efficiency in randomized trials. *Statist. Med.*, **30**, 2389–2408.
Morgan, K. L. and Rubin, D. B. (2012) Rerandomization to improve covariate balance in experiments. *Ann. Statist.*, **40**, 1263–1282.
Morgan, K. L. and Rubin, D. B. (2015) Rerandomization to balance tiers of covariates. *J. Am. Statist. Ass.*, **110**, 1412–1421.
Neyman, J. (1923) On the application of probability theory to agricultural experiments: essay on principles (with discussion); section 9 (in Polish) (Engl. transl. D. M. Dabrowska and T. P. Speed, *Statist. Sci.*, **5**, 465–472).
Rosenbaum, P. R. (2010) *Design of Observational Studies*. New York: Springer.
Rosenberger, W. F. and Lachin, J. M. (2015) *Randomization in Clinical Trials: Theory and Practice*. New York: Wiley.
Rubin, D. B. and van der Laan, M. J. (2011) Targeted ANCOVA estimator in RCTs. In *Targeted Learning* (eds M. J. van der Laan and S. Rose), pp. 201–215. New York: Springer.
Student (1938) Comparison between balanced and random arrangements of field plots. *Biometrika*, **29**, 363–378.
Tian, L., Jiang, F., Hasegawa, T., Uno, H., Pfeffer, M. and Wei, L. J. (2019) Moving beyond the conventional stratified analysis to estimate an overall treatment efficacy with the data from a comparative randomized clinical study. *Statist. Med.*, **38**, 917–932.
Tsiatis, A. A., Davidian, M., Zhang, M. and Lu, X. (2008) Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Statist. Med.*, **27**, 4658–4677.
Wager, S., Du, W., Taylor, J. and Tibshirani, R. J. (2016) High-dimensional regression adjustments in randomized experiments. *Proc. Natn. Acad. Sci. USA*, **113**, 12673–12678.
Wu, E. and Gagnon-Bartsch, J. A. (2018) The LOOP estimator: adjusting for covariates in randomized experiments. *Evaln Rev.*, **42**, 458–488.
Wu, C. F. J. and Hamada, M. S. (2011) *Experiments: Planning, Analysis, and Optimization*. New York: Wiley.
Yang, L. and Tsiatis, A. A. (2001) Efficiency study of estimators for a treatment effect in a pretest–posttest trial. *Am. Statistn*, **55**, 314–321.
Zhang, M., Tsiatis, A. A. and Davidian, M. (2008) Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. *Biometrics*, **64**, 707–715.
Zhou, Q., Ernst, P., Morgan, K. L., Rubin, D. B. and Zhang, A. (2018) Sequential rerandomization. *Biometrika*, **105**, 745–752.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supplementary material for "Rerandomization and regression adjustment"'.

Supplementary Material for "Rerandomization and regression adjustment"

Xinran Li and Peng Ding

Appendix A1 proves the results related to the sampling distributions.

Appendix A2 proves the results related to $\mathcal{S}$-optimality.

Appendix A3 proves the results related to the confidence intervals and the optimal adjusted estimators in terms of the estimated precision.

Appendix A4 proves the results related to the gains from the analyzer and the designer.

Appendix A5 proves the asymptotic equivalence of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ and $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ as well as the asymptotic equivalence of $\hat{V}_{\tau\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ and the Huber–White variance estimator.

Similar to the main paper, we shall focus on rerandomization using the Mahalanobis distance throughout the supplementary material, so we shall simply call it 'rerandomization' from now on.

A1. Sampling distributions of adjusted estimators

Proof of Proposition 1. By definition,

$$\begin{aligned} r_0\tilde{\boldsymbol{\beta}}_1 + r_1\tilde{\boldsymbol{\beta}}_0 &= r_0\left(\boldsymbol{S}_{\boldsymbol{w}}^2\right)^{-1}\boldsymbol{S}_{\boldsymbol{w},Y(1)} + r_1\left(\boldsymbol{S}_{\boldsymbol{w}}^2\right)^{-1}\boldsymbol{S}_{\boldsymbol{w},Y(0)} \\ &= \left\{(r_1r_0)^{-1}\boldsymbol{S}_{\boldsymbol{w}}^2\right\}^{-1}\left\{r_1^{-1}\boldsymbol{S}_{\boldsymbol{w},Y(1)} + r_0^{-1}\boldsymbol{S}_{\boldsymbol{w},Y(0)}\right\} = \boldsymbol{V}_{\boldsymbol{ww}}^{-1}\boldsymbol{V}_{\boldsymbol{w}\tau} = \tilde{\boldsymbol{\gamma}}. \end{aligned}$$

□

Proof of Theorem 1. The regression adjustment coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$ can depend on sample size n implicitly and have finite limits as $n \to \infty$. Recall that $Y_i(z;\boldsymbol{\beta}_z) = Y_i(z) - \boldsymbol{\beta}_z'\boldsymbol{w}_i$ is the "adjusted" potential outcome under treatment z, and $\tau_i(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0) = \tau_i - (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)'\boldsymbol{w}_i$ is the "adjusted" individual treatment effect. Under Condition 1(ii), the finite population variances and covariances $S^2_{Y(z;\boldsymbol{\beta}_z)} = S^2_{Y(z)} + \boldsymbol{\beta}_z'\boldsymbol{S}^2_{\boldsymbol{w}}\boldsymbol{\beta}_z - 2\boldsymbol{\beta}_z'\boldsymbol{S}_{\boldsymbol{w},Y(z)}$, $S^2_{\tau(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)} = S^2_\tau + (\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)'\boldsymbol{S}^2_{\boldsymbol{w}}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0) - 2(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)'\boldsymbol{S}_{\boldsymbol{w},\tau}$, $\boldsymbol{S}_{Y(z;\boldsymbol{\beta}_z),\boldsymbol{x}} = \boldsymbol{S}_{Y(z),\boldsymbol{x}} - \boldsymbol{\beta}_z'\boldsymbol{S}_{\boldsymbol{w},\boldsymbol{x}}$, and $\boldsymbol{S}^2_{\boldsymbol{x}}$ have finite limiting values. Under Condition 1(iii), the maximum squared distances satisfy that as $n \to \infty$, $\max_{1\le i\le n}\|\boldsymbol{x}_i\|_2^2/n \to 0$, and

$$\begin{aligned} n^{-1}\max_{1\le i\le n}\left|Y_i(z;\boldsymbol{\beta}_z) - \bar{Y}(z;\boldsymbol{\beta}_z)\right|^2 &= n^{-1}\max_{1\le i\le n}\left|Y_i(z) - \bar{Y}(z) - \boldsymbol{\beta}_z'\boldsymbol{w}_i\right|^2 \\ &\le n^{-1}(1+\boldsymbol{\beta}_z'\boldsymbol{\beta}_z)\max_{1\le i\le n}\left(\left|Y_i(z) - \bar{Y}(z)\right|^2 + \|\boldsymbol{w}_i\|_2^2\right) \to 0, \end{aligned}$$

where the inequality follows from the Cauchy–Schwarz inequality. Using Li et al. (2018, Theorem 1), we can show that, under rerandomization, $n^{1/2}\{\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0) - \tau\}$ has the asymptotic distribution (7). □

Proof of Corollary 1. Corollary 1 follows from Theorem 1 with $a = \infty$. □

Proof of Corollary 2. Corollary 2 follows from Theorem 1 with $(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0) = (\boldsymbol{0},\boldsymbol{0})$. □

Proof of Proposition 2. By definition, $\mathrm{Cov}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}, \hat{\tau} - \tilde{\boldsymbol{\gamma}}'\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}) = 0$ under the CRE. We have

$$\begin{aligned}\hat{\tau} - \tilde{\boldsymbol{\gamma}}'\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} &= \mathrm{proj}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) + \mathrm{res}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) - \tilde{\boldsymbol{\gamma}}' \left\{\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) + \mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})\right\} \\ &= \left\{\mathrm{proj}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) - \tilde{\boldsymbol{\gamma}}'_{\mathrm{proj}}\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})\right\} + \left\{\mathrm{res}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) - \tilde{\boldsymbol{\gamma}}'_{\mathrm{res}}\mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})\right\} \\ &\quad - (\tilde{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}_{\mathrm{proj}})'\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) - (\tilde{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}_{\mathrm{res}})'\mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}). \qquad \text{(A1)}\end{aligned}$$

Thus, the covariances between $\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}$ and the four terms in (A1) sum to 0. Below we consider the four covariances separately.

First, by definition, $\mathrm{proj}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) - \tilde{\boldsymbol{\gamma}}'_{\mathrm{proj}}\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ is uncorrelated with $\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$. Moreover, because $\mathrm{proj}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) - \tilde{\boldsymbol{\gamma}}'_{\mathrm{proj}}\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ is a linear function of $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$, it must also be uncorrelated with $\mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$. Thus, $\mathrm{proj}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) - \tilde{\boldsymbol{\gamma}}'_{\mathrm{proj}}\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ is uncorrelated with $\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) + \mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) = \hat{\boldsymbol{\tau}}_{\boldsymbol{w}}$.

Second, by definition, $\mathrm{res}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) - \tilde{\boldsymbol{\gamma}}'_{\mathrm{res}}\mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ is uncorrelated with $\mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$. Moreover, because $\mathrm{res}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) - \tilde{\boldsymbol{\gamma}}'_{\mathrm{res}}\mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ is uncorrelated with $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$, it must also be uncorrelated with $\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$. Thus, $\mathrm{res}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) - \tilde{\boldsymbol{\gamma}}'_{\mathrm{res}}\mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ is uncorrelated with $\mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) + \mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) = \hat{\boldsymbol{\tau}}_{\boldsymbol{w}}$.

Third, because $\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ is uncorrelated with $\mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$, we can simplify the covariance between $(\tilde{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}_{\mathrm{proj}})'\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ and $\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} = \mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) + \mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ as $(\tilde{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}_{\mathrm{proj}})'\mathrm{Cov}\{\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})\}$. The covariance of $\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ has the following equivalent forms:

$$\mathrm{Cov}\left\{\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})\right\} = n^{-1}\boldsymbol{V}_{\boldsymbol{wx}}\boldsymbol{V}_{\boldsymbol{xx}}^{-1}\boldsymbol{V}_{\boldsymbol{xw}} = (nr_1r_0)^{-1}\boldsymbol{S}^2_{\boldsymbol{w}|\boldsymbol{x}}. \qquad \text{(A2)}$$

Fourth, because $\mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ is uncorrelated with $\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$, we can simplify the covariance between $(\tilde{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}_{\mathrm{res}})'\mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ and $\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} = \mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) + \mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ as $(\tilde{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}_{\mathrm{res}})'\mathrm{Cov}\{\mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})\}$. The covariance of $\mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ has the following equivalent forms:

$$\mathrm{Cov}\left\{\mathrm{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})\right\} = \mathrm{Cov}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}) - \mathrm{Cov}\left\{\mathrm{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})\right\} = (nr_1r_0)^{-1}\boldsymbol{S}^2_{\boldsymbol{w}\setminus\boldsymbol{x}}. \qquad \text{(A3)}$$

From the above, the zero covariance between $\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}$ and (A1) implies that

$$\boldsymbol{0} = -(nr_1r_0)^{-1}\left\{(\tilde{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}_{\mathrm{proj}})'\boldsymbol{S}^2_{\boldsymbol{w}|\boldsymbol{x}} + (\tilde{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}_{\mathrm{res}})'\boldsymbol{S}^2_{\boldsymbol{w}\setminus\boldsymbol{x}}\right\}.$$

Therefore, Proposition 2 holds. □

Proof of Theorem 2. First, by the definitions of $V_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ in (5) and $R^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ in (6), the squared coefficients of $L_{K,a}$ and ε in (7) have the following equivalent forms:

$$V_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)R^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) = n\mathrm{Var}\left\{\mathrm{proj}\left(\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}, \qquad \text{(A4)}$$

$$V_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)\left\{1 - R^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)\right\} = n\mathrm{Var}\left\{\mathrm{res}\left(\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}. \qquad \text{(A5)}$$

Second, because $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) = \hat{\tau} - \boldsymbol{\gamma}'\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}$ by (1), the linear projection of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ on $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ under the

CRE and the corresponding residual have the following equivalent forms:

$$\begin{aligned}\text{proj}\left(\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)&=\text{proj}\left(\hat{\tau}-\boldsymbol{\gamma}'\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)=\text{proj}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)-\boldsymbol{\gamma}'\text{proj}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right),\\ \text{res}\left(\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)&=\text{res}\left(\hat{\tau}-\boldsymbol{\gamma}'\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)=\text{res}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)-\boldsymbol{\gamma}'\text{res}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right).\end{aligned}$$

Using the definitions of $\tilde{\boldsymbol{\gamma}}_{\text{proj}}$ in (10) and $\tilde{\boldsymbol{\gamma}}_{\text{res}}$ in (11), we can express the above quantities as

$$\begin{aligned}\text{proj}\left(\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)&=\text{proj}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)-\tilde{\boldsymbol{\gamma}}_{\text{proj}}'\text{proj}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)-\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{proj}}\right)'\text{proj}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\\ &=\tau+\text{res}\left\{\text{proj}(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}})\mid\text{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}})\right\}-\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{proj}}\right)'\text{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}),\qquad\text{(A6)}\\ \text{res}\left(\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)&=\text{res}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)-\tilde{\boldsymbol{\gamma}}_{\text{res}}'\text{res}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)-\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{res}}\right)'\text{res}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\\ &=\text{res}\left\{\text{res}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\mid\text{res}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}-\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{res}}\right)'\text{res}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right).\qquad\text{(A7)}\end{aligned}$$

Third, because the two terms in (A6) excluding the constant term τ are uncorrelated, the variance of the linear projection of $\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)$ on $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ is the summation of the variances of these two terms in (A6). Using (A2) and the definitions of $R^2_{\tau,\boldsymbol{x}}$ and R^2_{proj} in (8) and (10), we have

$$\begin{aligned}&\text{Var}\left\{\text{proj}\left(\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}\\ =\;&\text{Var}\left\{\text{proj}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}\cdot\left(1-R^2_{\text{proj}}\right)+\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{proj}}\right)'\text{Cov}\left\{\text{proj}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{proj}}\right)\\ =\;&n^{-1}V_{\tau\tau}\cdot R^2_{\tau,\boldsymbol{x}}\left(1-R^2_{\text{proj}}\right)+(nr_1r_0)^{-1}\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{proj}}\right)'\boldsymbol{S}^2_{\boldsymbol{w}|\boldsymbol{x}}\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{proj}}\right).\qquad\text{(A8)}\end{aligned}$$

Similarly, because the two terms in (A7) are uncorrelated, the variance of the residual of the linear projection of $\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)$ on $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ is the summation of the variances of these two terms in (A7). Using (A3) and the definitions of $R^2_{\tau,\boldsymbol{w}}$ in (13) and R^2_{res} in (11), we have

$$\begin{aligned}&\text{Var}\left\{\text{res}\left(\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}\\ =\;&\text{Var}\left\{\text{res}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}\cdot\left(1-R^2_{\text{res}}\right)+\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{res}}\right)'\text{Cov}\left\{\text{res}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{res}}\right)\\ =\;&n^{-1}V_{\tau\tau}\cdot\left(1-R^2_{\tau,\boldsymbol{x}}\right)\left(1-R^2_{\text{res}}\right)+(nr_1r_0)^{-1}\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{res}}\right)'\boldsymbol{S}^2_{\boldsymbol{w}\setminus\boldsymbol{x}}\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{res}}\right).\qquad\text{(A9)}\end{aligned}$$

Fourth, using (A4), (A5), (A8) and (A9), we have

$$\begin{aligned}V_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)R^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)&=V_{\tau\tau}R^2_{\tau,\boldsymbol{x}}\left(1-R^2_{\text{proj}}\right)+(r_1r_0)^{-1}\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{proj}}\right)'\boldsymbol{S}^2_{\boldsymbol{w}|\boldsymbol{x}}\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{proj}}\right),\\ V_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\left\{1-R^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\right\}&=V_{\tau\tau}\left(1-R^2_{\tau,\boldsymbol{x}}\right)\left(1-R^2_{\text{res}}\right)+(r_1r_0)^{-1}\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{res}}\right)'\boldsymbol{S}^2_{\boldsymbol{w}\setminus\boldsymbol{x}}\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{\text{res}}\right).\end{aligned}$$

These coupled with Theorem 1 imply Theorem 2. □

Proof of Corollary 3. First, we prove that $\tilde{\boldsymbol{\gamma}}_{\text{proj}}=\tilde{\boldsymbol{\gamma}}_{\text{res}}=\tilde{\boldsymbol{\gamma}}$ and $R^2_{\text{proj}}=1$. Under Condition 2, $\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}$ can linearly represent $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ as $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}=\boldsymbol{B}_1\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}$. Thus using the linearity of the projection operator, we have

$$\text{proj}(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}})-\tau-\tilde{\boldsymbol{\gamma}}'\text{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}})=\text{proj}(\hat{\tau}-\tau-\tilde{\boldsymbol{\gamma}}'\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}})=\text{proj}\{\text{res}(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{w}})\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\}.\qquad\text{(A10)}$$

By definition, $\text{res}(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{w}})$ is uncorrelated with $\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}$, and then it is also uncorrelated with $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}=\boldsymbol{B}_1\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}$.

Therefore, (A10) equals zero, implying that (i) $\text{proj}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) = \tau + \tilde{\boldsymbol{\gamma}}'\text{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ and $\tilde{\boldsymbol{\gamma}}$ equals the linear projection coefficient of $\text{proj}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ on $\text{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$, i.e., $\tilde{\boldsymbol{\gamma}} = \tilde{\boldsymbol{\gamma}}_{\text{proj}}$; (ii) the squared multiple correlation between $\text{proj}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ and $\text{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ equals 1, i.e., $R^2_{\text{proj}} = 1$. Moreover, (i) and Proposition 2 imply that $\tilde{\boldsymbol{\gamma}} = \tilde{\boldsymbol{\gamma}}_{\text{proj}} = \tilde{\boldsymbol{\gamma}}_{\text{res}}$.

Second, we prove that $R^2_{\text{res}} = (R^2_{\tau,\boldsymbol{w}} - R^2_{\tau,\boldsymbol{x}})/(1 - R^2_{\tau,\boldsymbol{x}})$. Because $\tilde{\boldsymbol{\gamma}}_{\text{res}} = \tilde{\boldsymbol{\gamma}}$, the residual from the linear projection of $\text{res}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ on $\text{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ reduces to

$$\text{res}\left(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right) - \tilde{\boldsymbol{\gamma}}'_{\text{res}}\text{res}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right) = \text{res}\left(\hat{\tau} - \tilde{\boldsymbol{\gamma}}'\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right) = \text{res}\left\{\text{res}\left(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\right) \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right\}. \tag{A11}$$

Because, under Condition 2, $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}} = \boldsymbol{B}_1\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}$ is uncorrelated with $\text{res}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{w}})$, (A11) reduces to $\text{res}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{w}})$. Thus, the squared multiple correlation between $\text{res}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ and $\text{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ reduces to

$$R^2_{\text{res}} = 1 - \frac{\text{Var}\left\{\text{res}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{w}})\right\}}{\text{Var}\left\{\text{res}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})\right\}} = 1 - \frac{1 - R^2_{\tau,\boldsymbol{w}}}{1 - R^2_{\tau,\boldsymbol{x}}} = \frac{R^2_{\tau,\boldsymbol{w}} - R^2_{\tau,\boldsymbol{x}}}{1 - R^2_{\tau,\boldsymbol{x}}}.$$

Corollary 3 then follows immediately from Theorem 2. □

Proof of Corollary 4. First, we prove that $\tilde{\boldsymbol{\beta}}_{\text{proj}} = \tilde{\boldsymbol{\beta}}$, $R^2_{\text{proj}} = R^2_{\tau,\boldsymbol{w}}/R^2_{\tau,\boldsymbol{x}}$, $\boldsymbol{S}^2_{\boldsymbol{w}|\boldsymbol{x}} = \boldsymbol{S}^2_{\boldsymbol{w}}$, and $\boldsymbol{S}^2_{\boldsymbol{w}\backslash\boldsymbol{x}} = \boldsymbol{0}$. Under Condition 3, $\boldsymbol{S}^2_{\boldsymbol{w}|\boldsymbol{x}} = \boldsymbol{S}^2_{\boldsymbol{w}}$, $\boldsymbol{S}^2_{\boldsymbol{w}\backslash\boldsymbol{x}} = \boldsymbol{0}$, and $\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} = \boldsymbol{B}_2\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$. Then using Proposition 2, we have $\tilde{\boldsymbol{\gamma}}_{\text{proj}} = \tilde{\boldsymbol{\gamma}}$. Moreover, $\text{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) = \hat{\boldsymbol{\tau}}_{\boldsymbol{w}}$, and thus the linear projection of $\text{proj}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ on $\text{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ under the CRE reduces to

$$\text{proj}\left\{\text{proj}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) \mid \text{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})\right\} = \tau + \tilde{\boldsymbol{\gamma}}'_{\text{proj}}\text{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) = \tau + \tilde{\boldsymbol{\gamma}}'\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} = \text{proj}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{w}}).$$

Consequently, the squared multiple correlation between $\text{proj}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ and $\text{proj}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ equals

$$R^2_{\text{proj}} = \frac{\text{Var}\left\{\text{proj}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{w}})\right\}}{\text{Var}\left\{\text{proj}\left(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}} = \frac{\text{Var}(\hat{\tau})R^2_{\tau,\boldsymbol{w}}}{\text{Var}(\hat{\tau})R^2_{\tau,\boldsymbol{x}}} = \frac{R^2_{\tau,\boldsymbol{w}}}{R^2_{\tau,\boldsymbol{x}}}.$$

Second, we prove that $R^2_{\text{res}} = 0$. Under Condition 3, $\text{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) = \boldsymbol{0}$, and thus the squared multiple correlation between $\text{res}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ and $\text{res}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}})$ reduces to zero.

Corollary 4 then follows immediately from Theorem 2. □

A2. $\mathcal{S}$-optimality

A2.1. Lemmas

Lemma A1. Let $\varepsilon \sim \mathcal{N}(0,1)$, and $L_{K,a} \sim D_1 \mid \boldsymbol{D}'\boldsymbol{D} \leq a$, where $\boldsymbol{D} = (D_1, \ldots, D_K) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_K)$. Both ε and $L_{K,a}$ are symmetric and unimodal around zero.

Proof of Lemma A1. It follows from Li et al. (2018, Proposition 2). □

Lemma A2. Let ζ_0, ζ_1 and ζ_2 be three mutually independent random variables. If

(1) ζ_0 is symmetric and unimodal around zero;

(2) ζ_1 and ζ_2 are symmetric around 0;

(3) $P(|\zeta_1| \le c) \ge P(|\zeta_2| \le c)$ for any $c \ge 0$;

then $P(|\zeta_0 + \zeta_1| \le c) \ge P(|\zeta_0 + \zeta_2| \le c)$ for any $c \ge 0$.

Proof of Lemma A2. It follows from Dharmadhikari and Joag-Dev (1988, Theorem 7.5). □

Lemma A3. Let $\varepsilon \sim \mathcal{N}(0,1)$, $L_{K,a} \sim D_1 \mid \boldsymbol{D}'\boldsymbol{D} \le a$, where $\boldsymbol{D} = (D_1, \ldots, D_K) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_K)$, and ε and $L_{K,a}$ be mutually independent. For any nonnegative constants $b_1 \le c_1$, $b_2 \le c_2$, and any $\alpha \in (0,1)$, the $1-\alpha$ quantile range of $b_1\varepsilon + b_2 L_{K,a}$ is narrower than or equal to that of $c_1\varepsilon + c_2 L_{K,a}$.

Proof of Lemma A3. From Lemma A1, $b_1\varepsilon$ is symmetric and unimodal. Because $b_2 \le c_2$, $P(|b_2 L_{K,a}| \le c) \ge P(|c_2 L_{K,a}| \le c)$ for any $c \ge 0$. Then from Lemma A2, $P(|b_1\varepsilon + b_2 L_{K,a}| \le c) \ge P(|b_1\varepsilon + c_2 L_{K,a}| \le c)$ for any $c \ge 0$.

From Lemma A1, $c_2 L_{K,a}$ is symmetric and unimodal. Because $b_1 \le c_1$, $P(|b_1\varepsilon| \le c) \ge P(|c_1\varepsilon| \le c)$ for any $c \ge 0$. Then from Lemma A2, $P(|b_1\varepsilon + c_2 L_{K,a}| \le c) \ge P(|c_1\varepsilon + c_2 L_{K,a}| \le c)$ for any $c \ge 0$.

From the above two results, for any $c \ge 0$,

$$P(|b_1\varepsilon + b_2 L_{K,a}| \le c) \ge P(|b_1\varepsilon + c_2 L_{K,a}| \le c) \ge P(|c_1\varepsilon + c_2 L_{K,a}| \le c),$$

which implies Lemma A3. □

Lemma A4. For any $\alpha \ge 1/2$, the αth quantile of $(1-\rho^2)^{1/2} \cdot \varepsilon + |\rho| \cdot L_{K,a}$ is nonincreasing in ρ^2.

Proof of Lemma A4. It follows from Li et al. (2018, Lemma A3). □

A2.2. Proofs

Proof of Theorem 3. In the asymptotic distribution of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ in Corollary 3, both coefficients of ε and $L_{K,a}$ attain their minimum values at $r_1\boldsymbol{\beta}_1 + r_1\boldsymbol{\beta}_0 \equiv \boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$. From Lemma A3 and Corollary 3, the $\mathcal{S}$-optimal adjusted estimator is attainable when $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$ or $r_0\boldsymbol{\beta}_1 + r_1\boldsymbol{\beta}_0 = r_0\tilde{\boldsymbol{\beta}}_1 + r_1\tilde{\boldsymbol{\beta}}_0$, with the asymptotic distribution (15). From Proposition 1, $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is $\mathcal{S}$-optimal. □

Proof of Theorem 4. From Corollary 4, in the asymptotic distribution of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ under rerandomization, the coefficient of ε does not depend on $\boldsymbol{\gamma}$, and the coefficient of $L_{K,a}$ attains its minimum when $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$. From Lemma A3 and Corollary 4, the $\mathcal{S}$-optimal adjusted estimator is attainable when $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$ or $r_0\boldsymbol{\beta}_1 + r_1\boldsymbol{\beta}_0 = r_0\tilde{\boldsymbol{\beta}}_1 + r_1\tilde{\boldsymbol{\beta}}_0$, with asymptotic distribution (17). Moreover, from Proposition 1, $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is $\mathcal{S}$-optimal. □

Proof of Theorem 5. From Corollary 2, the squared coefficient of ε in the asymptotic distribution of $\hat{\tau}$ is $V_{\tau\tau}(1 - R^2_{\tau,\boldsymbol{x}})$. From Theorem 1, the squared coefficient of ε in the asymptotic distribution of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is $V_{\tau\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)\{1 - R^2_{\tau,\boldsymbol{x}}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)\}$. Because

$$V_{\tau\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) = n\mathrm{Var}\{\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)\} = n\mathrm{Var}(\hat{\tau} - \tilde{\boldsymbol{\gamma}}\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}) = n\mathrm{Var}\left\{\mathrm{res}(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{w}})\right\} = V_{\tau\tau}\left(1 - R^2_{\tau,\boldsymbol{w}}\right),$$

the squared coefficient of ε in the asymptotic distribution of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ under rerandomization reduces to $V_{\tau\tau}\left(1 - R^2_{\tau,\boldsymbol{w}}\right)\{1 - R^2_{\tau,\boldsymbol{x}}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)\}$. Therefore, under rerandomization, the squared coefficient of ε in the asymptotic distribution of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is smaller than or equal to that of $\hat{\tau}$ if and only if

$$\begin{aligned} & V_{\tau\tau}\left(1 - R^2_{\tau,\boldsymbol{w}}\right)\left\{1 - R^2_{\tau,\boldsymbol{x}}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)\right\} \leq V_{\tau\tau}\left(1 - R^2_{\tau,\boldsymbol{x}}\right) \\ \iff & \left(1 - R^2_{\tau,\boldsymbol{w}}\right) - \left(1 - R^2_{\tau,\boldsymbol{w}}\right) R^2_{\tau,\boldsymbol{x}}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) \leq \left(1 - R^2_{\tau,\boldsymbol{x}}\right) \\ \iff & R^2_{\tau,\boldsymbol{w}} + \left(1 - R^2_{\tau,\boldsymbol{w}}\right) R^2_{\tau,\boldsymbol{x}}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) \geq R^2_{\tau,\boldsymbol{x}}. \end{aligned}$$

□

A2.3. The adjusted estimator with the smallest asymptotic variance

From Theorem 2, the asymptotic variance of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ under rerandomization is

$$\begin{aligned} & \left\{V_{\tau\tau}\left(1 - R^2_{\tau,\boldsymbol{x}}\right)\left(1 - R^2_{\text{res}}\right) + (r_1 r_0)^{-1}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}_{\text{res}})' \boldsymbol{S}^2_{\boldsymbol{w}\backslash\boldsymbol{x}}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}_{\text{res}})\right\} \\ & \quad + \left\{V_{\tau\tau} R^2_{\tau,\boldsymbol{x}}\left(1 - R^2_{\text{proj}}\right) + (r_1 r_0)^{-1}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}_{\text{proj}})' \boldsymbol{S}^2_{\boldsymbol{w}|\boldsymbol{x}}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}_{\text{proj}})\right\} v_{K,a} \\ = & V_{\tau\tau}\left(1 - R^2_{\tau,\boldsymbol{x}}\right)\left(1 - R^2_{\text{res}}\right) + V_{\tau\tau} R^2_{\tau,\boldsymbol{x}}\left(1 - R^2_{\text{proj}}\right) v_{K,a} \\ & \quad + (r_1 r_0)^{-1}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}_{\text{res}})' \boldsymbol{S}^2_{\boldsymbol{w}\backslash\boldsymbol{x}}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}_{\text{res}}) + (r_1 r_0)^{-1}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}_{\text{proj}})' \boldsymbol{S}^2_{\boldsymbol{w}|\boldsymbol{x}}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}_{\text{proj}})\, v_{K,a}. \end{aligned} \tag{A12}$$

It is a quadratic form of $\boldsymbol{\gamma}$. The derivative of (A12) with respect to $\boldsymbol{\gamma}$ is

$$\begin{aligned} & 2(r_1 r_0)^{-1} \boldsymbol{S}^2_{\boldsymbol{w}\backslash\boldsymbol{x}}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}_{\text{res}}) + 2 v_{K,a}(r_1 r_0)^{-1} \boldsymbol{S}^2_{\boldsymbol{w}|\boldsymbol{x}}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}_{\text{proj}}) \\ = & \; 2(r_1 r_0)^{-1}\left\{\left(\boldsymbol{S}^2_{\boldsymbol{w}\backslash\boldsymbol{x}} + v_{K,a}\boldsymbol{S}^2_{\boldsymbol{w}|\boldsymbol{x}}\right)\boldsymbol{\gamma} - \left(\boldsymbol{S}^2_{\boldsymbol{w}\backslash\boldsymbol{x}}\tilde{\boldsymbol{\gamma}}_{\text{res}} + v_{K,a}\boldsymbol{S}^2_{\boldsymbol{w}|\boldsymbol{x}}\tilde{\boldsymbol{\gamma}}_{\text{proj}}\right)\right\}. \end{aligned}$$

Therefore, under rerandomization, $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ with the smallest asymptotic variance is attainable when

$$r_0\boldsymbol{\beta}_1 + r_1\boldsymbol{\beta}_0 \equiv \boldsymbol{\gamma} = \left(\boldsymbol{S}^2_{\boldsymbol{w}\backslash\boldsymbol{x}} + v_{K,a}\boldsymbol{S}^2_{\boldsymbol{w}|\boldsymbol{x}}\right)^{-1}\left(\boldsymbol{S}^2_{\boldsymbol{w}\backslash\boldsymbol{x}}\tilde{\boldsymbol{\gamma}}_{\text{res}} + v_{K,a}\boldsymbol{S}^2_{\boldsymbol{w}|\boldsymbol{x}}\tilde{\boldsymbol{\gamma}}_{\text{proj}}\right).$$

When $a \approx 0$, the above coefficient is close to $\tilde{\boldsymbol{\gamma}}_{\text{res}}$.

A2.4. Technical details for Remark 2

We first give equivalent forms for the squared coefficients of ε in the asymptotic distributions of $\hat{\tau}$ and $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ under rerandomization. From Corollary 2 and the definition of $R^2_{\tau,\boldsymbol{x}}$ in (8), the squared coefficient of ε in the asymptotic distribution of $\hat{\tau}$ under rerandomization has the following equivalent forms:

$$V_{\tau\tau}(1 - R^2_{\tau,\boldsymbol{x}}) = n \cdot \text{Var}(\hat{\tau}) \cdot (1 - R^2_{\tau,\boldsymbol{x}}) = n \cdot \text{Var}\left\{\text{res}\left(\hat{\tau} \mid \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}. \tag{A13}$$

From Theorem 1 and the definition of $R^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)$ in (6), the squared coefficient of ε in the asymptotic distribution of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)$ under rerandomization has the following equivalent forms:

$$\begin{aligned}
&V_{\tau\tau}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)\left\{1-R^2_{\tau,\boldsymbol{x}}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)\right\}\\
&= n\cdot\mathrm{Var}\left\{\hat{\tau}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)\right\}\cdot\left\{1-R^2_{\tau,\boldsymbol{x}}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)\right\} = n\cdot\mathrm{Var}\left\{\mathrm{res}\left(\hat{\tau}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}\\
&= n\cdot\mathrm{Var}\left\{\mathrm{res}\left(\hat{\tau}-\tilde{\boldsymbol{\gamma}}'\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}\\
&= n\cdot\mathrm{Var}\left\{\mathrm{res}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)-\tilde{\boldsymbol{\gamma}}'\mathrm{res}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}. \qquad \text{(A14)}
\end{aligned}$$

We then study the covariance between $\mathrm{res}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)$ and $\mathrm{res}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)$ under the CRE. For $z=0,1$, let $Y_i^{\perp}(z)$ be the residual from the linear projection of $Y_i(z)$ on $\boldsymbol{x}_i$, and $\boldsymbol{w}_i^{\perp}$ be the residual from the linear projection of $\boldsymbol{w}_i$ on $\boldsymbol{x}_i$. We have

$$\begin{aligned}
\mathrm{res}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right) &= n_1^{-1}\sum_{i=1}^{n}Z_iY_i^{\perp}(1)-n_0^{-1}\sum_{i=1}^{n}(1-Z_i)Y_i^{\perp}(0),\\
\mathrm{res}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right) &= n_1^{-1}\sum_{i=1}^{n}Z_i\boldsymbol{w}_i^{\perp}-n_0^{-1}\sum_{i=1}^{n}(1-Z_i)\boldsymbol{w}_i^{\perp},
\end{aligned}$$

and thus

$$\mathrm{Cov}\left\{\mathrm{res}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right),\mathrm{res}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\} = n_1^{-1}\boldsymbol{S}_{Y(1),\boldsymbol{w}|\boldsymbol{x}}+n_0^{-1}\boldsymbol{S}_{Y(0),\boldsymbol{w}|\boldsymbol{x}}, \qquad \text{(A15)}$$

where $\boldsymbol{S}_{Y(z),\boldsymbol{w}|\boldsymbol{x}}$ is the finite population covariance between $Y^{\perp}(z)$ and $\boldsymbol{w}^{\perp}$, or, equivalently, the finite population partial covariance between $Y(z)$ and $\boldsymbol{w}$ given $\boldsymbol{x}$, for $z=0,1$.

We finally prove Remark 2. When $\boldsymbol{S}_{Y(z),\boldsymbol{w}|\boldsymbol{x}}=\boldsymbol{0}$ for $z=0,1$, from (A15), $\mathrm{res}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)$ is uncorrelated with $\mathrm{res}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)$. This further implies that

$$\mathrm{Var}\left\{\mathrm{res}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)-\tilde{\boldsymbol{\gamma}}'\mathrm{res}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\} = \mathrm{Var}\left\{\mathrm{res}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}+\mathrm{Var}\left\{\tilde{\boldsymbol{\gamma}}'\mathrm{res}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}\geq\mathrm{Var}\left\{\mathrm{res}\left(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}\right)\right\}.$$

From (A13) and (A14), the coefficient of ε for $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)$ is larger than or equal to that for $\hat{\tau}$.

A3. Estimated distributions and the associated optimality

A3.1. Lemmas

For treatment arm z ($z=0,1$), let $s^2_{Y(z)}$ be the sample variance of observed outcome, $\boldsymbol{s}^2_{\boldsymbol{w},z}$ be the sample covariance of covariates $\boldsymbol{w}$, and $\boldsymbol{s}_{Y(z),\boldsymbol{w}}$ be the sample covariance between observed outcome and covariates $\boldsymbol{w}$.

Lemma A5. Under rerandomization and Condition 1, for $z=0,1$,

$$\begin{aligned}
&s^2_{Y(z)}-S^2_{Y(z)}=o_P(1),\quad \boldsymbol{s}^2_{\boldsymbol{w},z}-\boldsymbol{S}^2_{\boldsymbol{w}}=o_P(1),\quad \boldsymbol{s}_{Y(z),\boldsymbol{w}}-\boldsymbol{S}_{Y(z),\boldsymbol{w}}=o_P(1),\\
&\boldsymbol{s}^2_{\boldsymbol{x},z}-\boldsymbol{S}^2_{\boldsymbol{x}}=o_P(1),\quad \boldsymbol{s}_{Y(z),\boldsymbol{x}}-\boldsymbol{S}_{Y(z),\boldsymbol{x}}=o_P(1),\quad \boldsymbol{s}_{\boldsymbol{w},\boldsymbol{x}}-\boldsymbol{S}_{\boldsymbol{w},\boldsymbol{x}}=o_P(1); \qquad \text{(A16)}
\end{aligned}$$

and for any $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$ that can depend implicitly on sample size n but have finite limits,

$$s^2_{Y(z;\boldsymbol{\beta}_z)} - S^2_{Y(z;\boldsymbol{\beta}_z)} = o_P(1), \quad s^2_{Y(z;\boldsymbol{\beta}_z)|\boldsymbol{x}} - S^2_{Y(z;\boldsymbol{\beta}_z)|\boldsymbol{x}} = o_P(1),$$
$$\boldsymbol{s}_{Y(z;\boldsymbol{\beta}_z),\boldsymbol{x}} - \boldsymbol{S}_{Y(z;\boldsymbol{\beta}_z),\boldsymbol{x}} = o_P(1), \quad \boldsymbol{s}_{Y(z;\boldsymbol{\beta}_z),\boldsymbol{w}} - \boldsymbol{S}_{Y(z;\boldsymbol{\beta}_z),\boldsymbol{w}} = o_P(1). \tag{A17}$$

Proof of Lemma A5. First, we can view covariates $\boldsymbol{w}$ as "outcomes" unaffected by the treatment. Thus, (A16) follows immediately from Li et al. (2018, Lemma A16).

Second, the observed sample variances have the following equivalent forms:

$$s^2_{Y(z;\boldsymbol{\beta}_z)} = s^2_{Y(z)} + \boldsymbol{\beta}_z' \boldsymbol{s}^2_{\boldsymbol{w},z} \boldsymbol{\beta}_z - 2\boldsymbol{\beta}_z' \boldsymbol{s}_{\boldsymbol{w},Y(z)}, \quad \boldsymbol{s}_{Y(z;\boldsymbol{\beta}_z),\boldsymbol{x}} = \boldsymbol{s}_{Y(z),\boldsymbol{x}} - \boldsymbol{\beta}_z' \boldsymbol{s}_{\boldsymbol{w},\boldsymbol{x}},$$
$$\boldsymbol{s}_{Y(z;\boldsymbol{\beta}_z),\boldsymbol{w}} = \boldsymbol{s}_{Y(z),\boldsymbol{w}} - \boldsymbol{\beta}_z' \boldsymbol{s}^2_{\boldsymbol{w},z}, \quad s^2_{Y(z;\boldsymbol{\beta}_z)|\boldsymbol{x}} = s_{Y(z;\boldsymbol{\beta}_z),\boldsymbol{x}} \left(\boldsymbol{s}^2_{\boldsymbol{x},z}\right)^{-1} s_{\boldsymbol{x},Y(z;\boldsymbol{\beta}_z)}.$$

From (A16),

$$s^2_{Y(z;\boldsymbol{\beta}_z)} - S^2_{Y(z;\boldsymbol{\beta}_z)} = o_P(1), \quad \boldsymbol{s}_{Y(z;\boldsymbol{\beta}_z),\boldsymbol{x}} - \boldsymbol{S}_{Y(z;\boldsymbol{\beta}_z),\boldsymbol{x}} = o_P(1), \quad \boldsymbol{s}_{Y(z;\boldsymbol{\beta}_z),\boldsymbol{w}} - \boldsymbol{S}_{Y(z;\boldsymbol{\beta}_z),\boldsymbol{w}} = o_P(1).$$

These further imply that $s^2_{Y(z;\boldsymbol{\beta}_z)|\boldsymbol{x}} - S^2_{Y(z;\boldsymbol{\beta}_z)|\boldsymbol{x}} = o_P(1)$. Thus, (A17) holds.

□

Lemma A6. Under rerandomization and Condition 1,

$$\hat{V}_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) - \tilde{V}_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) = o_P(1), \quad \hat{R}^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) - \tilde{R}^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) = o_P(1),$$

where

$$\tilde{V}_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \equiv V_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) + S^2_{\tau\setminus\boldsymbol{w}}, \quad \tilde{R}^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \equiv \tilde{V}^{-1}_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) V_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) R^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0).$$

Proof of Lemma A6. From (20), (21) and Lemma A5,

$$\begin{aligned}
\hat{V}_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) &= r_1^{-1} S^2_{Y(1;\boldsymbol{\beta}_1)} + r_0^{-1} S^2_{Y(0;\boldsymbol{\beta}_0)} - \boldsymbol{S}^2_{\tau(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)|\boldsymbol{w}} + o_P(1) \\
&= V_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) + \boldsymbol{S}^2_{\tau(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\setminus\boldsymbol{w}} + o_P(1) \\
&= V_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) + \boldsymbol{S}^2_{\tau\setminus\boldsymbol{w}} + o_P(1) = \tilde{V}_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) + o_P(1), \\
\hat{R}^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) &= \tilde{V}^{-1}_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)^{-1} \left(r_1^{-1} S^2_{Y(1;\boldsymbol{\beta}_1)|\boldsymbol{x}} + r_0^{-1} S^2_{Y(0;\boldsymbol{\beta}_0)|\boldsymbol{x}} - \boldsymbol{S}^2_{\tau(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)|\boldsymbol{x}} \right) + o_P(1) \\
&= \tilde{V}^{-1}_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) V_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \cdot V^{-1}_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \left(r_1^{-1} S^2_{Y(1;\boldsymbol{\beta}_1)|\boldsymbol{x}} + r_0^{-1} S^2_{Y(0;\boldsymbol{\beta}_0)|\boldsymbol{x}} - \boldsymbol{S}^2_{\tau(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)|\boldsymbol{x}} \right) + o_P(1) \\
&= \tilde{V}^{-1}_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) V_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) R^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) + o_P(1) = \tilde{R}^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) + o_P(1).
\end{aligned}$$

□

Lemma A7. $V_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) = V_{\tau\tau}(1 - R^2_{\tau,\boldsymbol{w}}) + (r_1 r_0)^{-1} (\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}})' \boldsymbol{S}^2_{\boldsymbol{w}} (\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}})$.

Proof of Lemma A7. It follows from Theorem 1 and Corollary 3 with $a = \infty$ and $\boldsymbol{x} = \emptyset$. We

give a more direct proof below:

$$\begin{aligned}V_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0) &= n\mathrm{Var}\left(\hat{\tau}-\tau-\boldsymbol{\gamma}'\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\right) = n\mathrm{Var}\left\{\hat{\tau}-\tau-\tilde{\boldsymbol{\gamma}}'\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}-(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}})'\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\right\}\\ &= n\mathrm{Var}\left\{\mathrm{res}(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{w}})-(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}})'\hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\right\} = n\mathrm{Var}\left\{\mathrm{res}(\hat{\tau}\mid\hat{\boldsymbol{\tau}}_{\boldsymbol{w}})\right\}+n(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}})'\mathrm{Cov}(\hat{\boldsymbol{\tau}}_{\boldsymbol{w}})(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}})\\ &= V_{\tau\tau}(1-R^2_{\tau,\boldsymbol{w}})+(r_1r_0)^{-1}(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}})'\boldsymbol{S}^2_{\boldsymbol{w}}(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}).\end{aligned}$$

□

A3.2. Proofs

Proof of Theorem 6. From Lemma A6, under rerandomization, the probability limit of the estimated distribution of $\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)$ is

$$\tilde{V}^{1/2}_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\left[\left\{1-\tilde{R}^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\right\}^{1/2}\cdot\varepsilon+\left\{\tilde{R}^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\right\}^{1/2}\cdot L_{K,a}\right].$$

By definition, it has the following equivalent forms:

$$\begin{aligned}&\left\{\tilde{V}_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)-\tilde{V}_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\tilde{R}^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\right\}^{1/2}\cdot\varepsilon+\left\{\tilde{V}_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\tilde{R}^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\right\}^{1/2}\cdot L_{K,a}\\ \sim&\left\{V_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)+S^2_{\tau\backslash\boldsymbol{w}}-V_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)R^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\right\}^{1/2}\cdot\varepsilon+\left\{V_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)R^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\right\}^{1/2}\cdot L_{K,a}\\ \sim&\left\{V_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\left\{1-R^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\right\}+S^2_{\tau\backslash\boldsymbol{w}}\right\}^{1/2}\cdot\varepsilon+\left\{V_{\tau\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)R^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)\right\}^{1/2}\cdot L_{K,a}.\end{aligned}$$

Because Condition 4 implies Condition 2, from Corollary 3 and its proof, we can further write the probability limit of the estimated distribution of $\hat{\tau}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)$ as

$$\left\{V_{\tau\tau}(1-R^2_{\tau,\boldsymbol{w}})+S^2_{\tau\backslash\boldsymbol{w}}+(r_1r_0)^{-1}(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}})'\boldsymbol{S}^2_{\boldsymbol{w}\backslash\boldsymbol{x}}(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}})\right\}^{1/2}\cdot\varepsilon+\left\{(r_1r_0)^{-1}(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}})'\boldsymbol{S}^2_{\boldsymbol{w}|\boldsymbol{x}}(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}})\right\}^{1/2}\cdot L_{K,a}.$$

□

Proof of Corollary 5. It follows from Theorem 6 with $a=\infty$ and $\boldsymbol{x}=\emptyset$. □

Proof of Corollary 6. With $\boldsymbol{\beta}_1=\boldsymbol{\beta}_0=\mathbf{0}$, Theorem 6 implies that the the probability limit of the estimated distribution of $\hat{\tau}$ is

$$\begin{aligned}&\tilde{V}^{1/2}_{\tau\tau}\left(1-\tilde{R}^2_{\tau,\boldsymbol{x}}\right)^{1/2}\cdot\varepsilon+\tilde{V}^{1/2}_{\tau\tau}\left(\tilde{R}^2_{\tau,\boldsymbol{x}}\right)^{1/2}\cdot L_{K,a}\\ \sim&\left(\tilde{V}_{\tau\tau}-\tilde{V}_{\tau\tau}\tilde{R}^2_{\tau,\boldsymbol{x}}\right)^{1/2}\cdot\varepsilon+\left(\tilde{V}_{\tau\tau}\tilde{R}^2_{\tau,\boldsymbol{x}}\right)^{1/2}\cdot L_{K,a}\\ \sim&\left(V_{\tau\tau}+S^2_{\tau\backslash\boldsymbol{w}}-V_{\tau\tau}R^2_{\tau,\boldsymbol{x}}\right)^{1/2}\cdot\varepsilon+\left(V_{\tau\tau}R^2_{\tau,\boldsymbol{x}}\right)^{1/2}\cdot L_{K,a}\\ \sim&\left\{V_{\tau\tau}(1-R^2_{\tau,\boldsymbol{x}})+S^2_{\tau\backslash\boldsymbol{w}}\right\}^{1/2}\cdot\varepsilon+\left(V_{\tau\tau}R^2_{\tau,\boldsymbol{x}}\right)^{1/2}\cdot L_{K,a}.\end{aligned}$$

□

Proof of Theorem 7. From Lemma A6, the probability limit of the estimated distribution of

$\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ in (26) is $\tilde{V}_{\tau\tau}^{1/2}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \cdot \varepsilon$. From Lemma A7, this probability limit has the following equivalent forms:

$$\begin{aligned}\tilde{V}_{\tau\tau}^{1/2}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \cdot \varepsilon &\sim \left\{V_{\tau\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) + S^2_{\tau\backslash \boldsymbol{w}}\right\}^{1/2} \cdot \varepsilon \\ &\sim \left\{V_{\tau\tau}(1 - R^2_{\tau,\boldsymbol{w}}) + S^2_{\tau\backslash \boldsymbol{w}} + (r_1 r_0)^{-1}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}})' \boldsymbol{S}^2_{\boldsymbol{w}}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}})\right\}^{1/2} \cdot \varepsilon.\end{aligned}$$

□

Proof of Corollary 7. From Theorem 6, in the probability limit of the estimated distribution of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ in (22), both coefficients of ε and $L_{K,a}$ attain their minimum values at $r_0\boldsymbol{\beta}_1 + r_1\boldsymbol{\beta}_0 \equiv \boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$. Lemma A3 then implies that the optimal adjusted estimator among (1) in terms of the estimated precision is attainable when $r_0\boldsymbol{\beta}_1 + r_1\boldsymbol{\beta}_0 \equiv \boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$. The corresponding probability limit of the estimated distribution is $\{V_{\tau\tau}(1 - R^2_{\tau,\boldsymbol{w}}) + S^2_{\tau\backslash \boldsymbol{w}}\}^{1/2} \cdot \varepsilon$. □

Proof of Corollary 8. It follows immediately from Theorem 7. □

A3.3. Additional comments on the asymptotic conservativeness under rerandomization

We first comment on a technical issue of degeneracy. From Section 6, the optimal adjusted estimator in terms of estimated precision is the same with or without Condition 4. Its estimated distribution is Gaussian with mean zero and variance equaling the limit of

$$\begin{aligned}V_{\tau\tau}(1 - R^2_{\tau,\boldsymbol{w}}) + S^2_{\tau\backslash \boldsymbol{w}} &= r_1^{-1}S^2_{Y(1)} + r_0^{-1}S^2_{Y(0)} - S^2_\tau - \left(r_1^{-1}S^2_{Y(1)|\boldsymbol{w}} + r_0^{-1}S^2_{Y(0)|\boldsymbol{w}} - S^2_{\tau|\boldsymbol{w}}\right) + S^2_{\tau\backslash \boldsymbol{w}} \\ &= r_1^{-1}S^2_{Y(1)\backslash \boldsymbol{w}} + r_0^{-1}S^2_{Y(0)\backslash \boldsymbol{w}}.\end{aligned}$$

We assume at least one of $S^2_{Y(1)\backslash \boldsymbol{w}}$ and $S^2_{Y(0)\backslash \boldsymbol{w}}$ has strictly positive limits, so the asymptotic variance $V_{\tau\tau}(1 - R^2_{\tau,\boldsymbol{w}}) + S^2_{\tau\backslash \boldsymbol{w}}$ must have a positive limit. Thus, its estimated distribution has a non-degenerate probability limit. Therefore, the estimated distribution of any adjusted estimator with or without Condition 4 has a non-degenerate probability limit. This helps to avoid the case where both sampling and estimated distributions converge weakly to zero. For any adjusted estimator, to prove the conservativeness of the confidence intervals, it suffices to show that the probability limit of the estimated distribution has wider or the same quantile ranges compared with the asymptotic sampling distribution.

We then discuss the asymptotic conservativeness of the estimated distributions.

With Condition 4 From Lemma A3, Corollary 3 and Theorem 6, the probability limit of the estimated distribution of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ has larger variance and wider quantile ranges than the asymptotic distribution of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$. Therefore, both the variance estimator and confidence intervals are asymptotically conservative.

Without Condition 4 Theorem 7 implies that the probability limit of the estimated distribution of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ has larger variance and wider quantile ranges than $V_{\tau\tau}^{1/2}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \cdot \varepsilon$. Lemma A4 implies that $V_{\tau\tau}^{1/2}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \cdot \varepsilon$ has larger variance and wider quantile ranges than the asymptotic distribution of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ in (7). Therefore, both the variance estimator and the Wald-type confidence intervals are asymptotically conservative.

A4. Gains from the analyzer and the designer

Proof of Corollary 9. First, we compare the asymptotic variances. From Corollary 2, the asymptotic variance of $\hat{\tau}$ is $V_{\tau\tau}\{1-(1-v_{K,a})R_{\tau,\boldsymbol{x}}^2\}$. From Theorem 3, the asymptotic variance of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is $V_{\tau\tau}(1 - R_{\tau,\boldsymbol{w}}^2)$. Compared to $\hat{\tau}$, the percentage reduction in the asymptotic variance of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is

$$1 - \frac{1 - R_{\tau,\boldsymbol{w}}^2}{1 - (1 - v_{K,a})R_{\tau,\boldsymbol{x}}^2} = \frac{R_{\tau,\boldsymbol{w}}^2 - (1 - v_{K,a})R_{\tau,\boldsymbol{x}}^2}{1 - (1 - v_{K,a})R_{\tau,\boldsymbol{x}}^2}.$$

Second, we compare the asymptotic quantile ranges. From Corollary 2, the length of the asymptotic $1-\alpha$ quantile range of $\hat{\tau}$ is $2V_{\tau\tau}^{1/2} \cdot q_{1-\alpha/2}(R_{\tau,\boldsymbol{x}}^2)$. From Theorem 3, the length of the asymptotic $1-\alpha$ quantile range of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is $2V_{\tau\tau}^{1/2}(1 - R_{\tau,\boldsymbol{w}}^2)^{1/2} \cdot q_{1-\alpha/2}(0)$. Compared to $\hat{\tau}$, the percentage reduction in the length of the asymptotic $1-\alpha$ quantile range of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is

$$1 - \frac{2V_{\tau\tau}^{1/2}(1 - R_{\tau,\boldsymbol{w}}^2)^{1/2} \cdot q_{1-\alpha/2}(0)}{2V_{\tau\tau}^{1/2} \cdot q_{1-\alpha/2}(R_{\tau,\boldsymbol{x}}^2)} = 1 - \left(1 - R_{\tau,\boldsymbol{w}}^2\right)^{1/2} \cdot \frac{q_{1-\alpha/2}(0)}{q_{1-\alpha/2}(R_{\tau,\boldsymbol{x}}^2)}.$$

Third, because $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is $\mathcal{S}$-optimal, both percentage reductions in the variance and the $1-\alpha$ quantile range are nonnegative. It is easy to verify that they are both nondecreasing in $R_{\tau,\boldsymbol{w}}^2$. □

Proof of Corollary 10. First, we compare the asymptotic variances. From Corollary 2, the asymptotic variance of $\hat{\tau}$ is $V_{\tau\tau}\{1 - (1 - v_{K,a})\, R_{\tau,\boldsymbol{x}}^2\}$. From Theorem 4, the asymptotic variance of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is $V_{\tau\tau}\{1-(1-v_{K,a})R_{\tau,\boldsymbol{x}}^2 - v_{K,a}R_{\tau,\boldsymbol{w}}^2\}$. Compared to $\hat{\tau}$, the percentage reduction in the asymptotic variance of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is

$$1 - \frac{V_{\tau\tau}\{1 - (1 - v_{K,a})R_{\tau,\boldsymbol{x}}^2 - v_{K,a}R_{\tau,\boldsymbol{w}}^2\}}{V_{\tau\tau}\{1 - (1 - v_{K,a})\, R_{\tau,\boldsymbol{x}}^2\}} = \frac{v_{K,a}R_{\tau,\boldsymbol{w}}^2}{1 - (1 - v_{K,a})\, R_{\tau,\boldsymbol{x}}^2}.$$

Second, we compare the asymptotic quantile ranges. From Corollary 2, the length of the asymptotic $1-\alpha$ quantile range of $\hat{\tau}$ is $2V_{\tau\tau}^{1/2} \cdot q_{1-\alpha/2}(R_{\tau,\boldsymbol{x}}^2)$. From Theorem 4, the length of the asymptotic $1-\alpha$ quantile range of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is $2V_{\tau\tau}^{1/2}(1 - R_{\tau,\boldsymbol{w}}^2)^{1/2} \cdot q_{1-\alpha/2}(\rho_{\tau,\boldsymbol{x}\setminus\boldsymbol{w}}^2)$. Compared to $\hat{\tau}$, the percentage reduction in the asymptotic $1-\alpha$ quantile range of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is

$$1 - \left(1 - R_{\tau,\boldsymbol{w}}^2\right)^{1/2} \cdot q_{1-\alpha/2}(\rho_{\tau,\boldsymbol{x}\setminus\boldsymbol{w}}^2)/q_{1-\alpha/2}(R_{\tau,\boldsymbol{x}}^2).$$

Third, because $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is $\mathcal{S}$-optimal, both percentage reductions in variance and $1-\alpha$ quantile range are nonnegative. It is easy to verify that the percentage reduction in the variance is

nondecreasing in $R^2_{\tau,\boldsymbol{w}}$. For the quantile range, from Lemma A3, $(1-R^2_{\tau,\boldsymbol{w}})^{1/2}\cdot q_{1-\alpha/2}(\rho^2_{\tau,\boldsymbol{x}\setminus\boldsymbol{w}})$, the $(1-\alpha/2)$th quantile of $(1-R^2_{\tau,\boldsymbol{x}})^{1/2}\cdot\varepsilon+(R^2_{\tau,\boldsymbol{x}}-R^2_{\tau,\boldsymbol{w}})^{1/2}\cdot L_{K,a}$, is nonincreasing in $R^2_{\tau,\boldsymbol{w}}$. Hence the percentage reduction in the $1-\alpha$ quantile range is nondecreasing in $R^2_{\tau,\boldsymbol{w}}$. □

Proof of Corollary 11. First, we compare the asymptotic variances. From Section 4.1.1, the asymptotic variance of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)$ under the CRE is $V_{\tau\tau}(1-R^2_{\tau,\boldsymbol{w}})$. From Theorem 4, the asymptotic variance of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)$ under rerandomization is $V_{\tau\tau}(1-R^2_{\tau,\boldsymbol{w}})\{1-(1-v_{K,a})\rho^2_{\tau,\boldsymbol{x}\setminus\boldsymbol{w}}\}$. Compared to the CRE, the percentage reduction in the asymptotic variance under rerandomization is

$$1-\frac{V_{\tau\tau}(1-R^2_{\tau,\boldsymbol{w}})\{1-(1-v_{K,a})\rho^2_{\tau,\boldsymbol{x}\setminus\boldsymbol{w}}\}}{V_{\tau\tau}(1-R^2_{\tau,\boldsymbol{w}})}=(1-v_{K,a})\rho^2_{\tau,\boldsymbol{x}\setminus\boldsymbol{w}}.$$

Second, we compare the asymptotic quantile ranges. From Section 4.1.1, the length of the asymptotic $1-\alpha$ quantile range of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)$ under the CRE is $2V^{1/2}_{\tau\tau}(1-R^2_{\tau,\boldsymbol{w}})^{1/2}\cdot q_{1-\alpha/2}(0)$. From Theorem 4, the length of the asymptotic $1-\alpha$ quantile range of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)$ under rerandomization is $2V^{1/2}_{\tau\tau}(1-R^2_{\tau,\boldsymbol{w}})^{1/2}\cdot q_{1-\alpha/2}(\rho^2_{\tau,\boldsymbol{x}\setminus\boldsymbol{w}})$. Compared to the CRE, the percentage reduction in the length of the asymptotic $1-\alpha$ quantile range under rerandomization is $1-q_{1-\alpha/2}(\rho^2_{\tau,\boldsymbol{x}\setminus\boldsymbol{w}})/q_{1-\alpha/2}(0)$.

Third, from Lemma A4, both percentage reductions in variance and $1-\alpha$ quantile range are nonnegative and nondecreasing in $\rho^2_{\tau,\boldsymbol{x}\setminus\boldsymbol{w}}=(R^2_{\tau,\boldsymbol{x}}-R^2_{\tau,\boldsymbol{w}})/(1-R^2_{\tau,\boldsymbol{w}})$. Consequently, both percentage reductions are nondecreasing in $R^2_{\tau,\boldsymbol{x}}$. □

Proof of Corollary 12. Recall that $\kappa=1+V^{-1}_{\tau\tau}S^2_{\tau\setminus\boldsymbol{w}}\geq 1$. From Corollary 6, under rerandomization and Conditions 1 and 4, the probability limit of the estimated distribution of $\hat{\tau}$ is

$$\begin{aligned}
&\left\{V_{\tau\tau}(1-R^2_{\tau,\boldsymbol{x}})+S^2_{\tau\setminus\boldsymbol{w}}\right\}^{1/2}\cdot\varepsilon+\left\{V_{\tau\tau}R^2_{\tau,\boldsymbol{x}}\right\}^{1/2}\cdot L_{K,a}\\
\sim\;& V^{1/2}_{\tau\tau}\left\{\left(\kappa-R^2_{\tau,\boldsymbol{x}}\right)^{1/2}\cdot\varepsilon+\left(R^2_{\tau,\boldsymbol{x}}\right)^{1/2}\cdot L_{K,a}\right\}\\
\sim\;& \kappa^{1/2}V^{1/2}_{\tau\tau}\left\{\left(1-R^2_{\tau,\boldsymbol{x}}/\kappa\right)^{1/2}\cdot\varepsilon+\left(R^2_{\tau,\boldsymbol{x}}/\kappa\right)^{1/2}\cdot L_{K,a}\right\}.
\end{aligned}$$

From Corollary 7, under rerandomization and Conditions 1 and 4, the probability limit of the estimated distribution of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)$ is

$$\left\{V_{\tau\tau}(1-R^2_{\tau,\boldsymbol{w}})+S^2_{\tau\setminus\boldsymbol{w}}\right\}^{1/2}\cdot\varepsilon\sim V^{1/2}_{\tau\tau}\left(\kappa-R^2_{\tau,\boldsymbol{w}}\right)^{1/2}\cdot\varepsilon\sim\kappa^{1/2}V^{1/2}_{\tau\tau}\left(1-R^2_{\tau,\boldsymbol{w}}/\kappa\right)^{1/2}\cdot\varepsilon.$$

First, we compare the variances. The variance of the probability limit of the estimated distributions of $\hat{\tau}$ is $\kappa V_{\tau\tau}\{1-(1-v_{K,a})R^2_{\tau,\boldsymbol{x}}/\kappa\}$. The variance of the probability limit of the estimated distributions of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)$ is $\kappa V_{\tau\tau}(1-R^2_{\tau,\boldsymbol{w}}/\kappa)$. Compared to $\hat{\tau}$, the percentage reduction in variance of the probability limit of the estimated distribution of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_0)$ is

$$1-\frac{1-R^2_{\tau,\boldsymbol{w}}/\kappa}{1-(1-v_{K,a})R^2_{\tau,\boldsymbol{x}}/\kappa}=\frac{R^2_{\tau,\boldsymbol{w}}-(1-v_{K,a})R^2_{\tau,\boldsymbol{x}}}{\kappa-(1-v_{K,a})R^2_{\tau,\boldsymbol{x}}}.$$

Second, we compare the quantile ranges. The length of $1-\alpha$ quantile range of the probability limit of the estimated distribution of $\hat{\tau}$ is $2\kappa^{1/2}V^{1/2}_{\tau\tau}\cdot q_{1-\alpha/2}(R^2_{\tau,\boldsymbol{x}}/\kappa)$. The length of $1-\alpha$ quantile

range of the probability limit of the estimated distribution of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is $2\kappa^{1/2} V_{\tau\tau}^{1/2}(1 - R^2_{\tau,\boldsymbol{w}}/\kappa)^{1/2} \cdot q_{1-\alpha/2}(0)$. Compared to $\hat{\tau}$, the percentage reduction in $1-\alpha$ quantile range of the probability limit of the estimated distribution of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is

$$1 - \frac{q_{1-\alpha/2}(0)\left(1 - R^2_{\tau,\boldsymbol{w}}/\kappa\right)^{1/2}}{q_{1-\alpha/2}(R^2_{\tau,\boldsymbol{x}}/\kappa)} = 1 - \left(1 - R^2_{\tau,\boldsymbol{w}}/\kappa\right)^{1/2} \cdot \frac{q_{1-\alpha/2}(0)}{q_{1-\alpha/2}(R^2_{\tau,\boldsymbol{x}}/\kappa)}.$$

Third, the optimality of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ in terms of the estimated precision implies that both percentage reductions are nonnegative. It is easy to verify that they are both nondecreasing in $R^2_{\tau,\boldsymbol{w}}$. □

Proof of Corollary 13. From Theorem 7, the probability limit of the estimated distribution of $\hat{\tau}$ is $\tilde{V}_{\tau\tau}^{1/2} \cdot \varepsilon \sim (V_{\tau\tau} + S^2_{\tau\backslash\boldsymbol{w}})^{1/2} \cdot \varepsilon$. From Corollary 8, the probability limit of the estimated distribution of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is $\{V_{\tau\tau}(1 - R^2_{\tau,\boldsymbol{w}}) + S^2_{\tau\backslash\boldsymbol{w}}\}^{1/2} \cdot \varepsilon$. Compared to $\hat{\tau}$, the percentage reduction in variance of the probability limit of the estimated distribution of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is

$$1 - \frac{V_{\tau\tau}(1 - R^2_{\tau,\boldsymbol{w}}) + S^2_{\tau\backslash\boldsymbol{w}}}{V_{\tau\tau} + S^2_{\tau\backslash\boldsymbol{w}}} = 1 - \frac{\kappa - R^2_{\tau,\boldsymbol{w}}}{\kappa} = \frac{R^2_{\tau,\boldsymbol{w}}}{\kappa},$$

and the percentage reduction in length of the $1-\alpha$ quantile range of the probability limit of the estimated distribution of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ is

$$1 - \frac{2q_{1-\alpha/2}(0)\left\{V_{\tau\tau}(1 - R^2_{\tau,\boldsymbol{w}}) + S^2_{\tau\backslash\boldsymbol{w}}\right\}^{1/2}}{2q_{1-\alpha/2}(0)\left(V_{\tau\tau} + S^2_{\tau\backslash\boldsymbol{w}}\right)^{1/2}} = 1 - \frac{\left(\kappa - R^2_{\tau,\boldsymbol{w}}\right)^{1/2}}{\kappa^{1/2}} = 1 - \left(1 - R^2_{\tau,\boldsymbol{w}}/\kappa\right)^{1/2}.$$

It is easy to show that both percentage reductions are nonnegative and nondecreasing in $R^2_{\tau,\boldsymbol{w}}$. □

In the following two proofs, we recall that $q_{1-\alpha/2}(0)$ is the $(1-\alpha/2)$th quantile of a standard Gaussian distribution, and use the fact that under either the CRE or rerandomization, the $1-\alpha$ confidence interval covers the average treatment effect if and only if

$$q_{1-\alpha/2}(0) \leq \hat{V}_{\tau\tau}^{-1/2}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) \times n^{1/2}\left\{\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) - \tau\right\} \leq q_{1-\alpha/2}(0). \tag{A18}$$

Therefore, the limit of the probability that (A18) holds is the asymptotic coverage probability of the confidence interval.

Proof of Corollary 14. From Corollary 7 and the comment after it, the probability limits of the estimated distributions of $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ are the same under the CRE and rerandomization, and so are the lengths of confidence intervals after being scaled by $n^{1/2}$.

From Lemma A6, $\hat{V}_{\tau\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ in (A18) has the same probability under the CRE and rerandomization. From Theorem 3 and Section 4.1.1, $n^{1/2}\{\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) - \tau\}$ in (A18) converges weakly to the same distribution under the CRE and rerandomization. From Slutsky's theorem, the quantity in the middle of (A18) converges weakly to the same distribution under the CRE and rerandomization. Therefore, for any $\alpha \in (0, 1)$, the limit of the probability that (A18) holds is the same

under the CRE and rerandomization, and so is the asymptotic coverage probability of the $1-\alpha$ confidence interval. □

Proof of Corollary 15. From Theorem 7 and Corollary 5, the probability limits of the estimated distributions of $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ are the same under the CRE and rerandomization, and so are the lengths of confidence intervals after being scaled by $n^{1/2}$.

Using Lemma A6, Theorem 1 and Slutsky's theorem, we have that under rerandomization, the quantity in the middle of (A18) is asymptotically equal to

$$\tilde{V}_{\tau\tau}^{-1/2}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) V_{\tau\tau}^{1/2}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \left[\left\{ 1 - R^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \right\}^{1/2} \cdot \varepsilon + \left\{ R^2_{\tau,\boldsymbol{x}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \right\}^{1/2} \cdot L_{K,a} \right]. \tag{A19}$$

Using Lemma A6, Corollary 1 and Slutsky's theorem, we have that under the CRE, the quantity in the middle of (A18) is asymptotically equal to

$$\tilde{V}_{\tau\tau}^{-1/2}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) V_{\tau\tau}^{1/2}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) \cdot \varepsilon. \tag{A20}$$

From Lemma A4, the distribution (A19) has shorter quantile ranges than (A20). Therefore, for any $\alpha \in (0,1)$, the limit of the probability that (A18) holds under rerandomization is larger than or equal to that under the CRE. □

A5. $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ and variance estimators under rerandomization

We need additional notation. Let $\boldsymbol{U}_i = (1, Z_i, \boldsymbol{w}_i', Z_i \boldsymbol{w}_i')' \in \mathbb{R}^{2J+2}$. Let $\bar{\boldsymbol{w}}_1$ and $\bar{\boldsymbol{w}}_0$ be the averages of covariates, and $\bar{Y}_1$ and $\bar{Y}_0$ be the averages of observed outcomes in treatment and control groups. We can verify that in the OLS fit of Y on $\boldsymbol{U}$, the coefficient of Z is $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$, and the residual for unit i is $\hat{e}_i = Y_i - \hat{\boldsymbol{\beta}}_1' \boldsymbol{w}_i - (\bar{Y}_1 - \hat{\boldsymbol{\beta}}_1' \bar{\boldsymbol{w}}_1)$ for treated units with $Z_i = 1$ and $\hat{e}_i = Y_i - \hat{\boldsymbol{\beta}}_0' \boldsymbol{w}_i - (\bar{Y}_0 - \hat{\boldsymbol{\beta}}_0' \bar{\boldsymbol{w}}_0)$ for control units with $Z_i = 0$. For $z = 0, 1$, let $\hat{\sigma}^2_{e,z} = n_z^{-1} \sum_{i: Z_i = z} \hat{e}_i^2$ be the average of squared residuals, and $\boldsymbol{m}^2_{\boldsymbol{w},z} = n_z^{-1} \sum_{i: Z_i = z} \boldsymbol{w}_i \boldsymbol{w}_i'$ be the second sample moment of $\boldsymbol{w}$. Define

$$\begin{aligned} \boldsymbol{G} &= n^{-1} \sum_{i=1}^n \boldsymbol{U}_i \boldsymbol{U}_i' = \begin{pmatrix} \boldsymbol{G}_{11} & \boldsymbol{G}_{12} \\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22} \end{pmatrix} \qquad \text{(A21)} \\ &= n^{-1} \sum_{i=1}^n \begin{pmatrix} 1 & Z_i & \boldsymbol{w}_i' & Z_i \boldsymbol{w}_i' \\ Z_i & Z_i & Z_i \boldsymbol{w}_i' & Z_i \boldsymbol{w}_i' \\ \boldsymbol{w}_i & Z_i \boldsymbol{w}_i & \boldsymbol{w}_i \boldsymbol{w}_i' & Z_i \boldsymbol{w}_i \boldsymbol{w}_i' \\ Z_i \boldsymbol{w}_i & Z_i \boldsymbol{w}_i & Z_i \boldsymbol{w}_i \boldsymbol{w}_i' & Z_i \boldsymbol{w}_i \boldsymbol{w}_i' \end{pmatrix} = \left(\begin{array}{cc:cc} 1 & r_1 & \bar{\boldsymbol{w}}' & r_1 \bar{\boldsymbol{w}}_1' \\ r_1 & r_1 & r_1 \bar{\boldsymbol{w}}_1' & r_1 \bar{\boldsymbol{w}}_1' \\ \hdashline \bar{\boldsymbol{w}} & r_1 \bar{\boldsymbol{w}}_1 & \boldsymbol{S}^2_{\boldsymbol{w}} & r_1 \boldsymbol{m}^2_{\boldsymbol{w},1} \\ r_1 \bar{\boldsymbol{w}}_1 & r_1 \bar{\boldsymbol{w}}_1 & r_1 \boldsymbol{m}^2_{\boldsymbol{w},1} & r_1 \boldsymbol{m}^2_{\boldsymbol{w},1} \end{array} \right) \\ \boldsymbol{H} &= n^{-1} \sum_{i=1}^n \hat{e}_i^2 \boldsymbol{U}_i \boldsymbol{U}_i' \equiv \begin{pmatrix} \boldsymbol{H}_{11} & \boldsymbol{H}_{12} \\ \boldsymbol{H}_{21} & \boldsymbol{H}_{22} \end{pmatrix} \qquad \text{(A22)} \\ &= n^{-1} \sum_{i=1}^n \begin{pmatrix} \hat{e}_i^2 & Z_i \hat{e}_i^2 & \hat{e}_i^2 \boldsymbol{w}_i' & Z_i \hat{e}_i^2 \boldsymbol{w}_i' \\ Z_i \hat{e}_i^2 & Z_i \hat{e}_i^2 & Z_i \hat{e}_i^2 \boldsymbol{w}_i' & Z_i \hat{e}_i^2 \boldsymbol{w}_i' \\ \hat{e}_i^2 \boldsymbol{w}_i & Z_i \hat{e}_i^2 \boldsymbol{w}_i & \hat{e}_i^2 \boldsymbol{w}_i \boldsymbol{w}_i' & Z_i \hat{e}_i^2 \boldsymbol{w}_i \boldsymbol{w}_i' \\ Z_i \hat{e}_i^2 \boldsymbol{w}_i & Z_i \hat{e}_i^2 \boldsymbol{w}_i & Z_i \hat{e}_i^2 \boldsymbol{w}_i \boldsymbol{w}_i' & Z_i \hat{e}_i^2 \boldsymbol{w}_i \boldsymbol{w}_i' \end{pmatrix} \end{aligned}$$

$$= \left(\begin{array}{cc:cc} \hat{\sigma}_e^2 & r_1\hat{\sigma}_{e,1}^2 & n^{-1}\sum_{i=1}^n \hat{e}_i^2 \boldsymbol{w}_i' & n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2 \boldsymbol{w}_i' \\ r_1\hat{\sigma}_{e,1}^2 & r_1\hat{\sigma}_{e,1}^2 & n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2 \boldsymbol{w}_i' & n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2 \boldsymbol{w}_i' \\ \hdashline n^{-1}\sum_{i=1}^n \hat{e}_i^2 \boldsymbol{w}_i & n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2 \boldsymbol{w}_i & n^{-1}\sum_{i=1}^n \hat{e}_i^2 \boldsymbol{w}_i\boldsymbol{w}_i' & n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2 \boldsymbol{w}_i\boldsymbol{w}_i' \\ n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2 \boldsymbol{w}_i & n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2 \boldsymbol{w}_i & n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2 \boldsymbol{w}_i\boldsymbol{w}_i' & n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2 \boldsymbol{w}_i\boldsymbol{w}_i' \end{array} \right).$$

The Huber–White variance estimator for $n^{1/2}\{\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0) - \tau\}$ is $\hat{V}_{\text{HW}} = \left[\boldsymbol{G}^{-1}\boldsymbol{H}\boldsymbol{G}^{-1}\right]_{(2,2)}$, the $(2,2)$th element of $\boldsymbol{G}^{-1}\boldsymbol{H}\boldsymbol{G}^{-1}$.

A5.1. Lemmas

Lemma A8. Under rerandomization and Condition 1, $\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} = r_0^{-1}\bar{\boldsymbol{w}}_1 = -r_1^{-1}\bar{\boldsymbol{w}}_0 = O_P(n^{-1/2})$.

Proof of Lemma A8. For $1 \le j \le J$, define pseudo potential outcomes $(\tilde{Y}(1), \tilde{Y}(0)) = (W_j, W_j)$. We can verify that Condition 1 also holds if we replace the original potential outcomes by the pseudo ones. Corollary 2 implies that $\hat{\tau}_{W_j} = O_P(n^{-1/2})$ and thus $\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} = O_P(n^{-1/2})$. □

Lemma A9. Under rerandomization and Condition 1, for $z = 0, 1$, we have $\hat{\boldsymbol{\beta}}_z - \tilde{\boldsymbol{\beta}}_z = o_P(1)$, and

$$s^2_{Y(z;\hat{\boldsymbol{\beta}}_z)} - S^2_{Y(z;\tilde{\boldsymbol{\beta}}_z)} = o_P(1), \quad s^2_{Y(z;\hat{\boldsymbol{\beta}}_z)|\boldsymbol{x}} - S^2_{Y(z;\tilde{\boldsymbol{\beta}}_z)|\boldsymbol{x}} = o_P(1)$$
$$\boldsymbol{s}_{Y(z;\hat{\boldsymbol{\beta}}_z),\boldsymbol{x}} - \boldsymbol{S}_{Y(z;\tilde{\boldsymbol{\beta}}_z),\boldsymbol{x}} = o_P(1), \quad \boldsymbol{s}_{Y(z;\hat{\boldsymbol{\beta}}_z),\boldsymbol{w}} - \boldsymbol{S}_{Y(z;\tilde{\boldsymbol{\beta}}_z),\boldsymbol{w}} = o_P(1).$$

Proof of Lemma A9. The results follow directly from Lemma A5. □

Lemma A10. For any two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, if both $\boldsymbol{A}$ and $\boldsymbol{A} + \boldsymbol{B}$ are nonsingular, then

$$(\boldsymbol{A} + \boldsymbol{B})^{-1} - \boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{B}\,(\boldsymbol{A} + \boldsymbol{B})^{-1}\,\boldsymbol{B}\boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}.$$

Proof of Lemma A10. Lemma A10 is known, but we give a direct proof for completeness. From

$$\boldsymbol{0} = \boldsymbol{A} - \boldsymbol{A} = \boldsymbol{A}\,(\boldsymbol{A} + \boldsymbol{B})^{-1}\,(\boldsymbol{A} + \boldsymbol{B}) - \boldsymbol{A} = \boldsymbol{A}\,(\boldsymbol{A} + \boldsymbol{B})^{-1}\,\boldsymbol{A} + \boldsymbol{A}\,(\boldsymbol{A} + \boldsymbol{B})^{-1}\,\boldsymbol{B} - \boldsymbol{A},$$
$$\boldsymbol{0} = \boldsymbol{B} - \boldsymbol{B} = (\boldsymbol{A} + \boldsymbol{B})\,(\boldsymbol{A} + \boldsymbol{B})^{-1}\,\boldsymbol{B} - \boldsymbol{B} = \boldsymbol{A}\,(\boldsymbol{A} + \boldsymbol{B})^{-1}\,\boldsymbol{B} + \boldsymbol{B}\,(\boldsymbol{A} + \boldsymbol{B})^{-1}\,\boldsymbol{B} - \boldsymbol{B},$$

we have $\boldsymbol{A}\,(\boldsymbol{A} + \boldsymbol{B})^{-1}\,\boldsymbol{A} - \boldsymbol{A} = \boldsymbol{B}\,(\boldsymbol{A} + \boldsymbol{B})^{-1}\,\boldsymbol{B} - \boldsymbol{B}$, which further implies

$$\begin{aligned}(\boldsymbol{A} + \boldsymbol{B})^{-1} - \boldsymbol{A}^{-1} &= \boldsymbol{A}^{-1}\left\{\boldsymbol{A}\,(\boldsymbol{A} + \boldsymbol{B})^{-1}\,\boldsymbol{A} - \boldsymbol{A}\right\}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\left\{\boldsymbol{B}\,(\boldsymbol{A} + \boldsymbol{B})^{-1}\,\boldsymbol{B} - \boldsymbol{B}\right\}\boldsymbol{A}^{-1} \\ &= \boldsymbol{A}^{-1}\boldsymbol{B}\,(\boldsymbol{A} + \boldsymbol{B})^{-1}\,\boldsymbol{B}\boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}.\end{aligned}$$

□

Lemma A11. Under rerandomization and Condition 1, $\boldsymbol{m}^2_{\boldsymbol{w},z} = \boldsymbol{S}^2_{\boldsymbol{w}} + o_P(1)$ for $z = 0, 1$. Both $\boldsymbol{G}_{11}$ and $\boldsymbol{G}_{22}$ in (A21) converge in probability to nonsingular matrices.

Proof of Lemma A11. First, we consider $\boldsymbol{m}^2_{\boldsymbol{w},z}$. By definition,

$$\boldsymbol{m}^2_{\boldsymbol{w},1} = n_1^{-1}\sum_{i:Z_i=1} \boldsymbol{w}_i\boldsymbol{w}_i' = n_1^{-1}\sum_{i:Z_i=1} (\boldsymbol{w}_i - \bar{\boldsymbol{w}}_1)\,(\boldsymbol{w}_i - \bar{\boldsymbol{w}}_1)' + \bar{\boldsymbol{w}}_1\bar{\boldsymbol{w}}_1' = n_1^{-1}(n_1 - 1)\boldsymbol{s}^2_{\boldsymbol{w},1} + \bar{\boldsymbol{w}}_1\bar{\boldsymbol{w}}_1'.$$

From Lemmas A5 and A8, $\boldsymbol{s}^2_{\boldsymbol{w},1} = \boldsymbol{S}^2_{\boldsymbol{w}} + o_P(1)$ and $\bar{\boldsymbol{w}}_1\bar{\boldsymbol{w}}_1' = o_P(1)$. Thus, $\boldsymbol{m}^2_{\boldsymbol{w},1} = \boldsymbol{S}^2_{\boldsymbol{w}} + o_P(1)$. Similarly, $\boldsymbol{m}^2_{\boldsymbol{w},0} = \boldsymbol{S}^2_{\boldsymbol{w}} + o_P(1)$.

Second, we consider $\boldsymbol{G}_{11}$. By definition, $\boldsymbol{G}_{11}$ has a limit as $n \to \infty$. Because

$$\boldsymbol{G}_{11}^{-1} = \begin{pmatrix} 1 & r_1 \\ r_1 & r_1 \end{pmatrix}^{-1} = \frac{1}{r_1 r_0}\begin{pmatrix} r_1 & -r_1 \\ -r_1 & 1 \end{pmatrix},$$

the limit of $\boldsymbol{G}_{11}$ is nonsingular.

Third, we consider $\boldsymbol{G}_{22}$. From the above and by definition,

$$\boldsymbol{G}_{22} = \begin{pmatrix} \boldsymbol{S}^2_{\boldsymbol{w}} & r_1\boldsymbol{m}^2_{\boldsymbol{w},1} \\ r_1\boldsymbol{m}^2_{\boldsymbol{w},1} & r_1\boldsymbol{m}^2_{\boldsymbol{w},1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}^2_{\boldsymbol{w}} & r_1\boldsymbol{S}^2_{\boldsymbol{w}} \\ r_1\boldsymbol{S}^2_{\boldsymbol{w}} & r_1\boldsymbol{S}^2_{\boldsymbol{w}} \end{pmatrix} + o_P(1).$$

Thus, $\boldsymbol{G}_{22}$ has a probability limit as $n \to \infty$. Because the limit of $\boldsymbol{S}^2_{\boldsymbol{w}}$ is nonsingular, and

$$\begin{pmatrix} \boldsymbol{S}^2_{\boldsymbol{w}} & r_1\boldsymbol{S}^2_{\boldsymbol{w}} \\ r_1\boldsymbol{S}^2_{\boldsymbol{w}} & r_1\boldsymbol{S}^2_{\boldsymbol{w}} \end{pmatrix}^{-1} = (\boldsymbol{G}_{11} \otimes \boldsymbol{S}^2_{\boldsymbol{w}})^{-1} = \boldsymbol{G}_{11}^{-1} \otimes (\boldsymbol{S}^2_{\boldsymbol{w}})^{-1},$$

the probability limit of $\boldsymbol{G}_{22}$ is nonsingular.

□

Lemma A12. Under rerandomization and Condition 1,

(i) for $z = 0, 1$, $\hat{\sigma}^2_{e,z} = s^2_{Y(z;\hat{\boldsymbol{\beta}}_z)} + o_P(1) = S^2_{Y(z;\tilde{\boldsymbol{\beta}}_z)} + o_P(1) = O_P(1)$, and $\boldsymbol{H}_{11} = O_P(1)$;

(ii) $n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2\boldsymbol{w}_i' = o_P(n^{1/2})$, $n^{-1}\sum_{i=1}^n (1-Z_i)\hat{e}_i^2\boldsymbol{w}_i' = o_P(n^{1/2})$, and $\boldsymbol{H}_{12} = \boldsymbol{H}_{21}' = o_P(n^{1/2})$;

(iii) $n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2\boldsymbol{w}_i\boldsymbol{w}_i' = o_P(n)$, $n^{-1}\sum_{i=1}^n (1-Z_i)\hat{e}_i^2\boldsymbol{w}_i\boldsymbol{w}_i' = o_P(n)$, and $\boldsymbol{H}_{22} = o_P(n)$.

Proof of Lemma A12. First, we prove (i). By definition and from Lemma A9,

$$\hat{\sigma}^2_{e,z} = n_z^{-1}(n_z - 1)s^2_{Y(z;\hat{\boldsymbol{\beta}}_z)} = s^2_{Y(z;\hat{\boldsymbol{\beta}}_z)} + o_P(1) = S^2_{Y(z;\tilde{\boldsymbol{\beta}}_z)} + o_P(1) = O_P(1), \quad (z = 0, 1).$$

This further implies $\hat{\sigma}^2_e = r_1\hat{\sigma}^2_{e,1} + r_0\hat{\sigma}^2_{e,0} = O_P(1)$. Thus, $\boldsymbol{H}_{11} = O_P(1)$.

Second, we prove (ii). For any $1 \le j \le J$,

$$\left| n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2 w_{ij} \right| \le \max_{1\le i\le n} |w_{ij}| \cdot n_1^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2 = \max_{1\le i\le n} |w_{ij}| \cdot \hat{\sigma}^2_{e,1}.$$

Condition 1 implies that $\max_{1\le i\le n} |w_{ij}|/n^{1/2} \to 0$ and thus $\max_{1\le i\le n} |w_{ij}| = o(n^{1/2})$. Lemma A12(i) implies $\hat{\sigma}^2_{e,1} = O_P(1)$. Thus, $n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2 w_{ij} = o(n^{1/2})O_P(1) = o_P(n^{1/2})$. This further implies $n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2\boldsymbol{w}_i' = o_P(n^{1/2})$. Similarly, $n^{-1}\sum_{i=1}^n (1-Z_i)\hat{e}_i^2\boldsymbol{w}_i' = o_P(n^{1/2})$. By definition,

$$n^{-1}\sum_{i=1}^n \hat{e}_i^2\boldsymbol{w}_i' = n^{-1}\sum_{i=1}^n Z_i\hat{e}_i^2\boldsymbol{w}_i' + n^{-1}\sum_{i=1}^n (1-Z_i)\hat{e}_i^2\boldsymbol{w}_i' = o_P(n^{1/2}).$$

Thus, $\boldsymbol{H}_{12} = o_P(n^{1/2})$.

Third, we prove (iii). For any $1 \leq l, j \leq J$,

$$\left| n^{-1} \sum_{i=1}^{n} Z_i \hat{e}_i^2 w_{il} w_{ij} \right| \leq n^{-1} \sum_{i=1}^{n} Z_i \hat{e}_i^2 |w_{il}||w_{ij}| \leq \max_{1 \leq i \leq n} |w_{il}| \cdot \max_{1 \leq i \leq n} |w_{ij}| \cdot n_1^{-1} \sum_{i=1}^{n} Z_i \hat{e}_i^2$$
$$= \max_{1 \leq i \leq n} |w_{il}| \cdot \max_{1 \leq i \leq n} |w_{ij}| \cdot \hat{\sigma}_{e,1}^2.$$

Condition 1 implies $\max_{1 \leq i \leq n} |w_{il}| = o(n^{1/2})$ and $\max_{1 \leq i \leq n} |w_{ij}| = o(n^{1/2})$. Lemma A12(i) implies $\hat{\sigma}_{e,1}^2 = O_P(1)$. Thus, $n^{-1} \sum_{i=1}^{n} Z_i \hat{e}_i^2 w_{il} w_{ij} = o(n^{1/2}) o(n^{1/2}) O_P(1) = o_P(n)$. This then implies $n^{-1} \sum_{i=1}^{n} Z_i \hat{e}_i^2 \boldsymbol{w}_i \boldsymbol{w}_i' = o_P(n)$. Similarly, $n^{-1} \sum_{i=1}^{n} (1 - Z_i) \hat{e}_i^2 \boldsymbol{w}_i \boldsymbol{w}_i' = o_P(n)$. Thus

$$n^{-1} \sum_{i=1}^{n} \hat{e}_i^2 \boldsymbol{w}_i \boldsymbol{w}_i' = n^{-1} \sum_{i=1}^{n} Z_i \hat{e}_i^2 \boldsymbol{w}_i \boldsymbol{w}_i' + n^{-1} \sum_{i=1}^{n} (1 - Z_i) \hat{e}_i^2 \boldsymbol{w}_i \boldsymbol{w}_i' = o_P(n).$$

By definition, $\boldsymbol{H}_{22} = o_P(n)$.

□

A5.2. Proofs

Proof of Proposition 3. We first consider the asymptotic distributions. Let $\hat{\boldsymbol{\gamma}} = r_0 \hat{\boldsymbol{\beta}}_1 + r_1 \hat{\boldsymbol{\beta}}_0$, which satisfies $\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}} = o_P(1)$ by Lemma A9. From Lemma A8, $\hat{\boldsymbol{\tau}}_{\boldsymbol{w}} = O_P(n^{-1/2})$. Then

$$\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0) - \hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) = \left(\hat{\tau} - \hat{\boldsymbol{\gamma}}' \hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\right) - \left(\hat{\tau} - \tilde{\boldsymbol{\gamma}}' \hat{\boldsymbol{\tau}}_{\boldsymbol{w}}\right) = (\tilde{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}})' \hat{\boldsymbol{\tau}}_{\boldsymbol{w}} = o_P(1) O_P(n^{-1/2}) = o_P(n^{-1/2}),$$

which implies that $n^{1/2}\{\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0) - \tau\}$ has the same asymptotic distribution as $n^{1/2}\{\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) - \tau\}$.

We then consider the probability limit of the estimated distribution. From Lemma A9, we can show that $\hat{V}_{\tau\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0) - \tilde{V}_{\tau\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) = o_P(1)$ and $\hat{R}_{\tau,\boldsymbol{x}}^2(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0) - \tilde{R}_{\tau,\boldsymbol{x}}^2(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) = o_P(1)$ under rerandomization. Therefore, under rerandomization and Condition 1 the estimated distributions of $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ and $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ have the same probability limit. □

Proof of Theorem 8. Let

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{G}_{11} & \boldsymbol{0}_{2 \times 2J} \\ \boldsymbol{0}_{2J \times 2} & \boldsymbol{G}_{22} \end{pmatrix}, \quad \boldsymbol{\Delta} = \boldsymbol{G} - \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{0}_{2 \times 2} & \boldsymbol{G}_{12} \\ \boldsymbol{G}_{21} & \boldsymbol{0}_{2J \times 2J} \end{pmatrix}, \quad \boldsymbol{\Psi} = \boldsymbol{G}^{-1} - \boldsymbol{\Lambda}^{-1}.$$

First, we first find the stochastic orders of $\boldsymbol{\Delta}$ and $\boldsymbol{\Psi}$. From Lemma A8, $\boldsymbol{G}_{12} = \boldsymbol{G}_{21}' = O_P(n^{-1/2})$, and thus $\boldsymbol{\Delta} = O_P(n^{-1/2})$. From Lemma A10,

$$\boldsymbol{\Psi} = (\boldsymbol{\Lambda} + \boldsymbol{\Delta})^{-1} - \boldsymbol{\Lambda}^{-1} = \boldsymbol{\Lambda}^{-1} \boldsymbol{\Delta} (\boldsymbol{\Lambda} + \boldsymbol{\Delta})^{-1} \boldsymbol{\Delta} \boldsymbol{\Lambda}^{-1} - \boldsymbol{\Lambda}^{-1} \boldsymbol{\Delta} \boldsymbol{\Lambda}^{-1}.$$

From Lemma A11, the probability limit of $\boldsymbol{\Lambda}$ exists and is nonsingular. Thus, $\boldsymbol{\Psi} = O_P(n^{-1/2})$.

Second, we consider the difference between $\boldsymbol{G}^{-1} \boldsymbol{H} \boldsymbol{G}^{-1}$ and $\boldsymbol{\Lambda}^{-1} \boldsymbol{H} \boldsymbol{\Lambda}^{-1}$:

$$\boldsymbol{G}^{-1} \boldsymbol{H} \boldsymbol{G}^{-1} - \boldsymbol{\Lambda}^{-1} \boldsymbol{H} \boldsymbol{\Lambda}^{-1} = \left(\boldsymbol{\Lambda}^{-1} + \boldsymbol{\Psi}\right) \boldsymbol{H} \left(\boldsymbol{\Lambda}^{-1} + \boldsymbol{\Psi}\right) - \boldsymbol{\Lambda}^{-1} \boldsymbol{H} \boldsymbol{\Lambda}^{-1}$$
$$= \boldsymbol{\Psi} \boldsymbol{H} \boldsymbol{\Lambda}^{-1} + \boldsymbol{\Lambda}^{-1} \boldsymbol{H} \boldsymbol{\Psi} + \boldsymbol{\Psi} \boldsymbol{H} \boldsymbol{\Psi}. \tag{A23}$$

In (A23), we focus on the sub-matrix of the first two rows and the first two columns. We consider the corresponding submatrices of the three terms in (A23). Below let $[\cdot]_{(1\text{-}2,1\text{-}2)}$ denote the sub-matrix of the first two rows and the first two columns, and $[\cdot]_{(2,2)}$ denote the $(2,2)$th element of a matrix. The first term in (A23) is

$$\boldsymbol{\Psi}\boldsymbol{H}\boldsymbol{\Lambda}^{-1} = \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{H}_{11} & \boldsymbol{H}_{12} \\ \boldsymbol{H}_{21} & \boldsymbol{H}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{G}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{G}_{22}^{-1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Psi}_{11}\boldsymbol{H}_{11}\boldsymbol{G}_{11}^{-1} + \boldsymbol{\Psi}_{12}\boldsymbol{H}_{21}\boldsymbol{G}_{11}^{-1} & * \\ * & * \end{pmatrix}.$$

From Lemmas A11 and A12,

$$[\boldsymbol{\Psi}\boldsymbol{H}\boldsymbol{\Lambda}^{-1}]_{(1\text{-}2,1\text{-}2)} = \boldsymbol{\Psi}_{11}\boldsymbol{H}_{11}\boldsymbol{G}_{11}^{-1} + \boldsymbol{\Psi}_{12}\boldsymbol{H}_{21}\boldsymbol{G}_{11}^{-1} = O_P(n^{-1/2})O_P(1)O_P(1) + O_P(n^{-1/2})o_P(n^{1/2})O_P(1).$$

For the second term in (A23), $[\boldsymbol{\Lambda}^{-1}\boldsymbol{H}\boldsymbol{\Psi}]_{(1\text{-}2,1\text{-}2)} = [\boldsymbol{\Psi}\boldsymbol{H}\boldsymbol{\Lambda}^{-1}]'_{(1\text{-}2,1\text{-}2)} = o_P(1)$. For the third term in (A23), Lemma A12 implies $\boldsymbol{H} = o_P(n)$ and thus $\boldsymbol{\Psi}\boldsymbol{H}\boldsymbol{\Psi} = O_P(n^{-1/2})o_P(n)O_P(n^{-1/2}) = o_P(1)$. Therefore, from (A23), $[\boldsymbol{G}^{-1}\boldsymbol{H}\boldsymbol{G}^{-1}]_{(1\text{-}2,1\text{-}2)} - [\boldsymbol{\Lambda}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}^{-1}]_{(1\text{-}2,1\text{-}2)} = o_P(1)$, which further implies

$$\hat{V}_{\text{HW}} - [\boldsymbol{\Lambda}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}^{-1}]_{(2,2)} = [\boldsymbol{G}^{-1}\boldsymbol{H}\boldsymbol{G}^{-1}]_{(2,2)} - [\boldsymbol{\Lambda}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}^{-1}]_{(2,2)} = o_P(1).$$

Third, we consider the difference between $[\boldsymbol{\Lambda}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}^{-1}]_{(2,2)}$ and $\hat{V}_{\tau\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$. Because

$$\begin{aligned}[\boldsymbol{\Lambda}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}^{-1}]_{(1\text{-}2,1\text{-}2)} = \boldsymbol{G}_{11}^{-1}\boldsymbol{H}_{11}\boldsymbol{G}_{11}^{-1} &= (r_1 r_0)^{-2} \begin{pmatrix} r_1 & -r_1 \\ -r_1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\sigma}_e^2 & r_1\hat{\sigma}_{e,1}^2 \\ r_1\hat{\sigma}_{e,1}^2 & r_1\hat{\sigma}_{e,1}^2 \end{pmatrix} \begin{pmatrix} r_1 & -r_1 \\ -r_1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} r_0^{-1}\hat{\sigma}_{e,0}^2 & -r_0^{-1}\hat{\sigma}_{e,0}^2 \\ -r_0^{-1}\hat{\sigma}_{e,0}^2 & r_1^{-1}\hat{\sigma}_{e,1}^2 + r_0^{-1}\hat{\sigma}_{e,0}^2 \end{pmatrix}.\end{aligned}$$

using Lemma A12, we have

$$[\boldsymbol{\Lambda}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}^{-1}]_{(2,2)} = r_1^{-1}\hat{\sigma}_{e,1}^2 + r_0^{-1}\hat{\sigma}_{e,0}^2 = r_1^{-1}\boldsymbol{s}^2_{Y(1;\hat{\boldsymbol{\beta}}_1)} + r_0^{-1}\boldsymbol{s}^2_{Y(0;\hat{\boldsymbol{\beta}}_0)} + o_P(1).$$

The property of OLS implies $\boldsymbol{s}_{Y(z;\hat{\boldsymbol{\beta}}_z),\boldsymbol{w}} = \mathbf{0}$ for $z = 0, 1$. Using the definition in (20), we can then simplify $\hat{V}_{\tau\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ as $\hat{V}_{\tau\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0) = r_1^{-1}s^2_{Y(1;\hat{\boldsymbol{\beta}}_1)} + r_0^{-1}s^2_{Y(0;\hat{\boldsymbol{\beta}}_0)}$. Therefore, $[\boldsymbol{\Lambda}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}^{-1}]_{(2,2)} = \hat{V}_{\tau\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0) + o_P(1)$. From the above, we have $\hat{V}_{\text{HW}} = [\boldsymbol{\Lambda}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}^{-1}]_{(2,2)} + o_P(1) = \hat{V}_{\tau\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0) + o_P(1)$. □