Debiased Regression Adjustment in Completely Randomized Experiments with Moderately High-dimensional Covariates

Xin Lu[1], Fan Yang[1], Yuhao Wang[1,2]

[1]Tsinghua University; [2]Shanghai Qi Zhi Institute.

Abstract

Completely randomized experiment is the gold standard for causal inference. When the covariate information for each experimental candidate is available, one typical way is to include them in covariate adjustments for more accurate treatment effect estimation. In this paper, we investigate this problem under the randomization-based framework, i.e., that the covariates and potential outcomes of all experimental candidates are assumed as deterministic quantities and the randomness comes solely from the treatment assignment mechanism. Under this framework, to achieve asymptotically valid inference, existing estimators usually require either (i) that the dimension of covariates p grows at a rate no faster than $O(n^{2/3})$ as sample size $n \to \infty$; or (ii) certain sparsity constraints on the linear representations of potential outcomes constructed via possibly high-dimensional covariates. In this paper, we consider the moderately high-dimensional regime where p is allowed to be in the same order of magnitude as n. We develop a novel debiased estimator with a corresponding inference procedure and establish its asymptotic normality under mild assumptions. Our estimator is model-free and does not require any sparsity constraint on potential outcome's linear representations. We also discuss its asymptotic efficiency improvements over the unadjusted treatment effect estimator under different dimensionality constraints. Numerical analysis confirms that compared to other regression adjustment based treatment effect estimators, our debiased estimator performs well in moderately high dimensions.

Keywords: randomization-based inference, causal inference, regression adjustment, high-dimensional statistics

1 Introduction

Since the seminal work of Fisher [1935], completely randomized experiment has been the gold standard for causal inference. By using only the randomization in treatment assignments as the reasoned basis, completely randomized experiments can provide a valid inference of treatment effects without any model or distributional assumptions on the experimental candidates, such as being i.i.d. sampled from some superpopulation or some other model assumption that may be unverifiable in practice. Such inference is often called randomization-based or design-based inference, sometimes also called finite-population-based inference to emphasize its focus on just candidates in the experiment. Evaluating causal effects under this inferential framework has been an active area of research in the past few years [e.g. Lin, 2013, Bloniarz et al., 2016, Li et al., 2018, Lei and Ding, 2021, Wang and Li, 2022], which is also the framework we focus on in this paper.

When the covariate information of each experimental candidate is available, it is often popular to use regression adjustment in the analysis stage to utilize the additional covariate information to improve estimation precision [Lin, 2013, Bloniarz et al., 2016, Negi and Wooldridge, 2021, Lei and Ding, 2021]. Lin [2013] showed that in completely randomized experiments, regression adjustment can improve the asymptotic efficiency of average treatment effect estimation when the dimension of covariates p is fixed as the sample size n goes to infinity. However, in modern experiments, researchers can collect a large number of covariates. It is important to develop methodology and theory for high-dimensional settings where $p \to \infty$ as the sample size goes to infinity.

In the high-dimensional regime where $p \gg n$, the LASSO-adjusted estimator [Bloniarz et al., 2016] has been proposed to estimate causal effect with high accuracy. However, it is under the requirement that the potential outcomes can be well represented by a sparse linear function of the high-dimensional covariates, which can be unrealistic in practice. In the lower dimensional regime, Lei and Ding [2021] improved Lin's method by debiasing and guaranteed improvement of estimation efficiency compared to the difference in means estimator when $p = O(n^{2/3}/(\log n)^{1/3})$, without any assumption on the sparsity of potential outcome's linear representations. However, when p grows faster than that rate, the theory for Lei's estimator is not applicable anymore. In practice, we have also found that Lei's estimator may not achieve the desired performance when p is relatively large, as we will show later in the numerical analysis section of this paper.

In this paper, we consider the *moderately* high-dimensional regime where p is allowed to be in the same order of magnitude as n. We develop a novel debiased estimator with a corresponding inference procedure and establish its asymptotic normality and inference validity. Our estimator guarantees improvement of estimation efficiency over the unadjusted estimator in the regime $p = o(n)$. In the higher dimensional regime where p can be in the same order of magnitude as n, we prove that if the empirical correlation between potential outcomes and covariates is sufficiently large relative to p/n, we can guarantee efficiency improvement compared to the simple difference in means estimator that estimates causal effect without any adjustment on the collected covariates. Noteworthy, our theory for asymptotic normality and inference validity is based on some standard regularity conditions on potential outcomes and their empirical regression residuals, beyond that, no assumption is required on the observed covariate features. This allows us to provide valid inferences even with heavy-tailed covariates.

Before moving forward, it would be convenient to introduce some notations that will be used in the rest of this paper. Given n d-dimensional samples $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ of a variable $\boldsymbol{a}$, we let $\bar{\boldsymbol{a}}$ denote its empirical average, and let $\boldsymbol{S}^2_{\boldsymbol{a}} := \frac{1}{n-1}\sum_{i=1}^n (\boldsymbol{a}_i - \bar{\boldsymbol{a}})(\boldsymbol{a}_i - \bar{\boldsymbol{a}})^\top$ be the empirical covariance matrix of $\boldsymbol{a}$. Given n samples of two variables $\boldsymbol{a}$ and $\boldsymbol{b}$, we write $\boldsymbol{S}_{\boldsymbol{a},\boldsymbol{b}} := \frac{1}{n-1}\sum_{i=1}^n (\boldsymbol{a}_i - \bar{\boldsymbol{a}})(\boldsymbol{b}_i - \bar{\boldsymbol{b}})^\top$ as its empirical covariance matrix. Analogously, given a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, we define the *scaled* variance of variable $\boldsymbol{a}$, $\boldsymbol{S}^2_{\boldsymbol{A},\boldsymbol{a}}$ and *scaled* covariance of variables $\boldsymbol{a}$ and $\boldsymbol{b}$ as

$$\boldsymbol{S}^2_{\boldsymbol{A},\boldsymbol{a}} := \frac{1}{n-1}\sum_{i=1}^n\sum_{j=1}^n A_{ij}(\boldsymbol{a}_i - \bar{\boldsymbol{a}})(\boldsymbol{a}_j - \bar{\boldsymbol{a}})^\top$$

$$\boldsymbol{S}_{\boldsymbol{A},\boldsymbol{a},\boldsymbol{b}} := \frac{1}{n-1}\sum_{i=1}^n\sum_{j=1}^n A_{ij}(\boldsymbol{a}_i - \bar{\boldsymbol{a}})(\boldsymbol{b}_j - \bar{\boldsymbol{b}})^\top.$$

Apparently, $\boldsymbol{S}^2_{\boldsymbol{a}} = \boldsymbol{S}^2_{\boldsymbol{I},\boldsymbol{a}}$, where $\boldsymbol{I}$ is the identity matrix. Given a sequences of random variables U_n, we use $U_n \dot{\sim} \mathcal{N}(0,1)$ to denote that it converges in distribution to a standard normal distribution. We write $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ as the hat matrix where $H_{ij} := (n-1)^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top \boldsymbol{S}^{-2}_{\boldsymbol{X}}(\boldsymbol{X}_j - \bar{\boldsymbol{X}})$.

1.1 Framework, literature review and overview of contributions

We consider an experiment with n experimental units and two arms $z \in \{0,1\}$. We restrict the experiment to be a completely randomized experiment where the experimenter selects n_1 units uniformly at random to the treatment group, and the rest n_0 units to the control group. To describe causality, we adopt the potential outcome framework, where for each experimental candidate i, we assume there are two potential outcomes $Y_i(1)$ and $Y_i(0)$, where $Y_i(z)$ denotes the potential outcome of unit i had unit i been assigned to group z. Then the observed outcome Y_i satisfies $Y_i = Z_iY_i(1) + (1 - Z_i)Y_i(0)$, where the random variable $Z_i \in \{0,1\}$ denotes the treatment arm assigned to unit i.

In this paper, we consider the randomization-based framework where all the potential outcomes $(Y_i(1), Y_i(0))$ are considered deterministic and the randomness comes only from the randomness in the treatment assignment mechanism. This regime has a long history in the study of randomized experiments [Imbens and Rubin, 2015]. Under this regime, our target of interest then becomes estimating the sample average treatment effect:

$$\bar{\tau} := \frac{1}{n}\sum_{i=1}^{n}\tau_i \quad \text{where} \quad \tau_i = Y_i(1) - Y_i(0).$$

According to the finite-population central limit theorem [Hájek, 1960], one can prove that under some standard regularity conditions, as n goes to infinity, the simple difference in mean estimator $\hat{\tau}_{\text{unadj}} := \frac{1}{n_1}\sum_{i=1}^{n} Z_iY_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - Z_i)Y_i$ is guaranteed to provide asymptotically normal and unbiased estimation of $\bar{\tau}$. Specifically, writing $r_z := n_z/n$ as the proportion of units in treatment arm z, we have

$$\sqrt{n}(\hat{\tau}_{\text{unadj}} - \bar{\tau})/\sigma_{\text{cre}} \dot{\sim} \mathcal{N}(0,1) \quad \text{where} \quad \sigma^2_{\text{cre}} := r_1^{-1}S^2_{Y(1)} + r_0^{-1}S^2_{Y(0)} - S^2_{\tau}.$$

When each experimental unit i has a deterministic covariate information $\boldsymbol{X}_i$ of dimension p indicating its properties, such as age, education, and body weights, a typical choice is to use regression adjustment to incorporate these information for more efficient treatment effect estimation. By defining $\tilde{\boldsymbol{\beta}}_z := \boldsymbol{S}_{\boldsymbol{X}}^{-1}\boldsymbol{S}_{\boldsymbol{X},Y(z)}$, and $\hat{\boldsymbol{\beta}}_z$ as an empirical estimate of $\tilde{\boldsymbol{\beta}}_z$ using samples in the treatment arm z, Lin [2013] showed that in the regime where p is assumed as a fixed constant, the regression-adjusted estimator:

$$\hat{\tau}_{\text{adj}} := \frac{1}{n_1}\sum_{i=1}^{n} Z_i\{Y_i - \hat{\boldsymbol{\beta}}_1^\top(\boldsymbol{X}_i - \bar{\boldsymbol{X}})\} - \frac{1}{n_0}(1 - Z_i)\{Y_i - \hat{\boldsymbol{\beta}}_0^\top(\boldsymbol{X}_i - \bar{\boldsymbol{X}})\} \tag{1}$$

has the representation

$$\hat{\tau}_{\text{adj}} - \bar{\tau} = \frac{1}{n}\sum_{i=1}^{n} Z_i(r_1^{-1}e_i(1) + r_0^{-1}e_i(0)) + o_{\mathbb{P}}(1/\sqrt{n}), \tag{2}$$

where

$$e_i(z) := Y_i(z) - \bar{Y}(z) - \tilde{\boldsymbol{\beta}}_z^\top(\boldsymbol{X}_i - \bar{\boldsymbol{X}}) \tag{3}$$

corresponds to the regression residual of $Y_i(z)$. Thus, from standard results in finite population central limit theorem (see e.g. Li and Ding [2017] and the references therein), it has the asymptotic distribution

$$\sqrt{n}(\hat{\tau}_{\text{adj}} - \bar{\tau})/\sigma_{\text{adj}} \dot{\sim} \mathcal{N}(0,1) \quad \text{where} \quad \sigma^2_{\text{adj}} := r_1^{-1}S^2_{e(1)} + r_0^{-1}S^2_{e(0)} - S^2_{\tau_e},$$

where $\tau_{e,i} := e_i(1) - e_i(0)$ corresponds to the individual difference of the residuals in two treatment arms. As an extension, Lei and Ding [2021] showed that in the regime $p = o(\sqrt{n})$, the above conclusion still holds. Moreover, they proposed a debiased estimator which is asymptotically normal in the regime $p = O(n^{2/3}/(\log n)^{1/3})$.

In this paper, we develop a new regression adjustment-based ATE estimator in the higher dimensional regime where p is allowed to be in the same order of magnitude as n. Under this regime, a major challenge is that the inverse covariance matrix used in the construction of $\hat{\boldsymbol{\beta}}_z$ is based on the $\boldsymbol{X}_i$'s in treatment arm z, which is hard to analyze with large p. Because of this, in this paper, we instead consider a variant of regression adjustment estimator where the inverse covariance matrix is instead constructed using covariate information on the entire dataset, i.e. that we instead set $\hat{\boldsymbol{\beta}}_z := \boldsymbol{S}_{\boldsymbol{X}}^{-2} \boldsymbol{s}_{\boldsymbol{X},Y(z)}$[1], where $\boldsymbol{s}_{\boldsymbol{X},Y(z)}$ denotes an empirical estimate of $\boldsymbol{S}_{\boldsymbol{X},Y(z)}$ using samples from treatment arm z, and set $\hat{\tau}_{\text{adj}}$ as in (1) but with the new $\hat{\boldsymbol{\beta}}_z$. In other words, compared to the definition at Lin [2013], Lei and Ding [2021], the main difference is that we use the $\boldsymbol{X}_i$'s of the entire sample to construct the inverse covariance matrix, not only those in treatment arm z. Such a variant has been discussed before by Li and Ding [2020], Wang and Li [2022] in the lower dimensional regime. Building on this variant of regression adjustment-based estimator, we propose a new debiased estimator for average treatment effect estimation; we prove that in an asymptotic regime where $p = o(n)$, the debiased estimator is asymptotically normal with the same variance as in the fixed dimensional regime, namely σ^2_{adj}. We also derive the asymptotic distribution of the debiased estimator in the higher-dimensional regime where p can be in the same order of magnitude as n and propose sufficient conditions so that the debiased estimator is asymptotically more efficient than the unadjusted estimator. As far as we are aware, both regimes have not been well investigated by existing randomization-based framework literature.

Finally, we would like to remark that besides the randomization-based framework we consider in this manuscript, another choice is the superpopulation framework, which assumes that the experimental units must be randomly sampled from some superpopulation. This framework is also popular in literature, some examples include Tsiatis et al. [2008], Wager et al. [2016], Negi and Wooldridge [2021].

The rest of this paper is organized as follows. In Section 2, we present our new estimator, and prove its asymptotic normality in the regime $p = o(n)$. In Section 3, we prove its asymptotic convergence in the higher dimensional regime where p is allowed to be in the same order of magnitude as n and discuss conditions so that it is asymptotically more efficient than without regression adjustment. In Section 4, we present a new confidence interval construction method. In Section 5, we discuss the regularity conditions involved in Sections 2–4. We further conduct a numerical analysis in Section 6. We end with a concluding remark in Section 7.

2 A debiased regression adjustment estimator

In this section, we discuss how our debiased estimator is constructed, and present its asymptotic property in the regime $p = o(n)$. As will be demonstrated in the Supplementary Material, under some regularity conditions that will be discussed further later in this section, there exists some c_i

[1] From here and below, we assume throughout that $p < n$, so that the regression adjustment estimator is well defined.

and c_{ij} depending only on the pre-treatment variables such that the following decomposition holds:

$$\begin{aligned}\hat{\tau}_{\text{adj}} - \bar{\tau} = & -\frac{r_1 r_0}{n} \sum_{i=1}^{n} H_{ii} \left(\frac{Y_i(1) - \bar{Y}(1)}{r_1^2} - \frac{Y_i(0) - \bar{Y}(0)}{r_0^2} \right) \\ & + \frac{1}{n} \sum_{i=1}^{n} (Z_i - r_1) c_i + \frac{1}{n} \sum_{i \neq j} (Z_i - r_1)(Z_j - r_1) c_{ij} + o_{\mathbb{P}}(1/\sqrt{n}),\end{aligned} \tag{4}$$

where the form of c_i and c_{ij} will be demonstrated in the Supplementary Material. Apparently, the second term is of mean zero. For the third term, since the Z_i's are just weakly dependent, one can also show that its mean is approximately zero. Therefore, the first term constitutes the bias. Based on this, we propose a new debiased ATE estimator via stripping the original $\hat{\tau}_{\text{adj}}$ with an approximately unbiased estimator of the first term, which is constructed based on the observations in the two treatment arms:

$$\hat{\tau}_{\text{db}} := \hat{\tau}_{\text{adj}} + r_1 r_0 \left(\frac{1}{n_1} \sum_{i:Z_i=1} H_{ii} \frac{(Y_i - \bar{Y}_1)}{r_1^2} - \frac{1}{n_0} \sum_{i:Z_i=0} H_{ii} \frac{(Y_i - \bar{Y}_0)}{r_1^2} \right), \tag{5}$$

where different from $\bar{Y}(z)$, $\bar{Y}_z$ is the estimated mean using the outcome data in treatment arm z. Below we provide an asymptotic convergence guarantee of $\hat{\tau}_{\text{db}}$ in the asymptotic regime $p = o(n)$. Our new guarantee is based on the following 4 assumptions:

Assumption 1. For $z = 0, 1$, r_z tends to a limit in $(0, 1)$.

Assumption 2. For $z = 0, 1$, $\sum_{i=1}^{n} \left(Y_i(z) - \bar{Y}(z) \right)^2 = O(n)$.

Assumption 3. Consider the $e_i(z)$ in (3), as $n \to \infty$,

$$\max_z \max_i |Y_i(z) - \bar{Y}(z)|/\sqrt{n} \to 0 \quad \& \quad \max_z \max_i |e_i(z)|/\sqrt{n} \to 0.$$

Assumption 4. $\liminf_{n\to\infty} \sigma_{\text{adj}}^2 > 0$.

Assumptions 1, 2 and 4 are standard assumptions in randomization-based inference. Assumption 3 is a Lindeberg–Feller-type condition to guarantee that the representation in (2) has an approximately normal distribution in the large sample limit; similar condition has also appeared in previous regression adjustment literature; see e.g. Lei and Ding [2021] and the references therein.

With these assumptions, we are able to show that $\hat{\tau}_{\text{db}}$ has the representation

$$\begin{aligned}\hat{\tau}_{\text{db}} - \bar{\tau} = & \frac{1}{n} \sum_{i=1}^{n} Z_i (r_1^{-1} e_i(1) + r_0^{-1} e_i(0)) + \frac{1}{n} \sum_{i=1}^{n} Z_i (r_1^{-1} s_i(1) + r_0^{-1} s_i(0)) \\ & - \frac{1}{n} \sum_{i \neq j} (Z_i - r_1)(Z_j - r_1) \left(\frac{A_{ij}(1)}{r_1^2} - \frac{A_{ij}(0)}{r_0^2} \right) + o_{\mathbb{P}}(1/\sqrt{n}),\end{aligned} \tag{6}$$

where recall that $e_i(z)$ are defined as in Assumption 3,

$$s_i(z) := H_{ii}(Y_i(z) - \bar{Y}(z)) - \frac{1}{n} \sum_{i=1}^{n} H_{ii}(Y_i(z) - \bar{Y}(z)),$$

and $A_{ij}(z)$ is the (i,j)-th entry of the matrix

$$\boldsymbol{A}(z) := \boldsymbol{H} \operatorname{diag}\left\{(Y_1(z) - \bar{Y}(z), \ldots, Y_n(z) - \bar{Y}(z))^\top\right\}.$$

We now invoke the following condition on the ordered sequence of potential outcomes, which characterizes the tail of the population of potential outcomes:

Assumption 5. As $n \to \infty$, $p/n \to 0$. Moreover, let $(Y_{(1)}(z) - \bar{Y}(z))^2 \geq \ldots \geq (Y_{(n)}(z) - \bar{Y}(z))^2$ be the ordered sequence of $\{(Y_i(z) - \bar{Y}(z))^2\}_{i=1}^n$. Then, for any $z \in \{0,1\}$, we have that $\sum_{i=1}^p (Y_{(i)}(z) - \bar{Y}(z))^2 = o(n)$.

Armed with the above 5 assumptions, we are able to show that the second and third terms in the decomposition (6) are of order $o_{\mathbb{P}}(1/\sqrt{n})$, so that $\hat{\tau}_{\text{db}}$ has the same asymptotic representation (and therefore the same asymptotic distribution) as $\hat{\tau}_{\text{adj}}$ in the fixed p case. Specifically, we have the following result:

Theorem 1. *Under Assumptions 1–5, we have*

$$n^{1/2} \left(\hat{\tau}_{db} - \tau\right) / \sigma_{adj} \dot{\sim} \mathcal{N}(0,1).$$

As discussed before, Assumptions 1–4 are basic assumptions in previous literature; Assumption 5 is a novel contribution from us. In other words, our estimator requires no further assumption on the moment of covariate $\boldsymbol{X}_i$ or the leverage scores H_{ii} to obtain asymptotically normal convergence. As we will show later in Section 5, Assumption 5 holds with high probability when $Y_i(z)$ are i.i.d. generated from a superpopulation with bounded second order moment.

In the next section, we further consider the regime where p can be in the same order of magnitude as n. Under this regime, the second and third terms in the decomposition (6) are not necessarily negligible anymore, so we need new analysis to understand their asymptotic convergence. The second term is easy to deal with. The main obstacle of the analysis is that the quadratic form of centered treatment indicators (i.e., the third term in (6)) needs to be characterized by a new analytic tool, which is the central limit theorem of quadratic forms. We will discuss this further in the next section.

3 Asymptotic convergence of debiased estimator with moderately high-dimensional covariates

In this section, we consider the asymptotic convergence of our debiased estimator in the moderate high-dimensional regime where we allow p to be in the same order of magnitude as n. As mentioned before, since in this regime, the second and third terms in the decomposition (6) are not negligible, we need to derive their (joint) distribution in the large sample limit. With the help of standard results in combinatorial central limit theorem (CLT) [Hájek, 1960], the second term is easy to derive. Whilst for the third term, due to the quadratic functions in the form $(Z_i - r_1)(Z_j - r_1)$, standard combinatorial CLT is not applicable to understand its convergence anymore.

In this paper, we will use the newly developed central limit theorem of the so-called *homogeneous* sums from Koike [2022] to characterize the third term of (6). Specifically, Koike [2022] studied the convergence of random variables satisfying the following form:

Definition 1. Let $\boldsymbol{W} = (W_i)_{i=1}^n$ be a sequence of independent centered random variables with unit variance. A homogeneous sum is a random variable of the form

$$Q(f; \boldsymbol{W}) = \sum_{i_1,\ldots,i_p=1}^{n} f(i_1, \ldots, i_q) W_{i_1} \cdots W_{i_q},$$

where $n, q \in \mathbb{N}, [n] := \{1, \ldots, n\}$ and $f : [n]^q \to \mathbb{R}$ is a symmetric function vanishing on diagonals, i.e., $f(i_1, \ldots, i_q) = 0$ unless $i_1, \ldots, i_q$ are mutually different.

This is an extension of the linear statistics studied by the standard CLT. Apparently, by setting $W_i \equiv (Z_i - r_1)/(r_1 r_0)^{1/2}$, $q = 2$ and

$$f(i_1, i_2) \equiv r_1 r_0 \left(r_1^{-2} A_{i_1 i_2}(1) + r_1^{-2} A_{i_2 i_1}(1) - r_0^{-2} A_{i_1 i_2}(0) - r_0^{-2} A_{i_2 i_1}(0)\right)/(2n),$$

the third term in (6) falls into this category, with the exception that in our problem, the $(Z_i - r_1)$'s are weakly dependent.

Below we give a brief literature review of this class of CLT. Rotar' [1976] and Rotar et al. [1979] studied the invariance principles of $Q(f; \boldsymbol{W})$ regarding the law of W. De Jong [1990] established the univariate central limit theorem for $Q(f; \boldsymbol{W})$. Koike [2022] extended it to the multivariate case and obtained the bound for the error of normal approximation. The special case of $q = 1$ is the classic sum of independent random variables. The special case of $q = 2$ has been extensively studied; see, for example, De Wet and Venter [1973], de Jong [1987], Fox and Taqqu [1987]. Note that all of the results are for independent W_i's.

Nevertheless, the results of Koike [2022] are still not sufficient, since Koike [2022] assumed that all the random variables are independent, whilst in our problem, the Z_i's are weakly dependent due to simple random sampling. Mimicking the idea of Hàjek's coupling [Hájek, 1960] and its extension in Wang and Li [2022], we are able to propose a new combinatorial central limit theorem to characterize the joint distribution of the decomposition in (4), and furthermore, the asymptotic distribution of $\hat{\tau}_{\text{db}} - \bar{\tau}$.

To formally describe the new convergence result, we first define

$$\sigma_{\text{hd},l}^2 := r_1^{-1} S_{e(1)+s(1)}^2 + r_0^{-1} S_{e(0)+s(0)}^2 - S_{\tau_e+\tau_s}^2,$$

where analogous to τ_e, we write $\tau_{s,i} := s_i(1) - s_i(0)$. Moreover, we define

$$\sigma_{\text{hd},q}^2 := (r_1 r_0)^2 S^2_{\boldsymbol{Q},\, r_1^{-2} Y(1) - r_0^{-2} Y(0)},$$

where $\boldsymbol{Q}$ is an $n \times n$ dimensional matrix such that $Q_{ij} := H_{ij}^2$ whenever $i \neq j$ and $Q_{ii} := H_{ii} - H_{ii}^2$. Apparently, $\sigma_{\text{hd},l}^2$ and $\sigma_{\text{hd},q}^2$ correspond to the variances contributed by the linear statistic and quadratic statistic in (6), respectively. We also write $\sigma_{\text{hd}}^2 := \sigma_{\text{hd},l}^2 + \sigma_{\text{hd},q}^2$ as their total variance. We now invoke the following assumption regarding the asymptotics of our estimator variance.

Assumption 6. $\liminf_{n\to\infty} \sigma_{\text{hd},l}^2 > 0$ or $\liminf_{n\to\infty} \sigma_{\text{hd},q}^2 > 0$.

Armed with this assumption, we are able to show that our debiased estimator is asymptotically normal.

Theorem 2. *If Assumptions 1–3, 6 hold, we have as $n \to \infty$,*

$$n^{1/2} \left(\hat{\tau}_{db} - \tau\right) / \sigma_{hd} \dot{\sim} \mathcal{N}(0, 1).$$

Notice that in the above theorem, we do not require any assumption on the scaling of p; instead, we just require $p < n$ so that the debiased estimator is well-defined. Of course, as we will discuss later, we may still need p/n to be asymptotically bounded below by some constant in $(0, 1)$ to justify some assumptions. For more discussions, we refer the readers to Section 5.

Interestingly, without the constraint $p = o(n)$, we do not need Assumption 5 anymore. This is because we characterize more carefully the distribution of $\hat{\tau}_{\text{db}}$ by considering the second and third terms in the decomposition (6).

3.1 Efficiency improvement for debiased regression adjustment with moderately high-dimensional covariates

In Theorem 2, we present the asymptotic distribution of our new debiased estimator. In this section, we discuss conditions so that $\sigma^2_{\text{hd},l} + \sigma^2_{\text{hd},q}$ is smaller than σ^2_{cre}, i.e., that our debiased estimator is asymptotically more efficient than without doing regression adjustment at all.

To shed light on how the covariate-dimension-to-sample-size-ratio p/n influences the variance of our new estimator, we consider a class of pre-treatment covariates whose leverage scores concentrate around their mean p/n. We formalize it into the following assumption.

Assumption 7. Let $\alpha := p/n$. As $n \to \infty$, we have that:

$$\max_{1\leq i\leq n} |H_{ii} - \alpha| \to 0 \quad \& \quad \max_{z\in\{0,1\}} \max_{1\leq i\leq n} |Y_i(z) - \bar{Y}(z)|/\sqrt{n} \to 0,$$

or for some constant $\eta > 0$,

$$\frac{1}{n}\sum_{i=1}^{n}(H_{ii} - \alpha)^2 \to 0 \quad \& \quad \max_{z\in\{0,1\}} \frac{1}{n}\sum_{i=1}^{n} |Y_i(z) - \bar{Y}(z)|^{2+\eta} \to 0.$$

In Lei and Ding [2021], the authors have also assumed similar assumptions. Specifically, they require $\max_i |H_{ii}| = o(1)$, which is equivalent to our first constraint in the regime $\alpha \to 0$. To justify this assumption, they proved that this assumption holds with high probability when the covariates are randomly generated from a superpopulation with $(6+\delta)$-th moment. This proof, albeit being enough for the setting $p = o(n)$, cannot be used in the $p \asymp n$ regime. As we will discuss further in Section 5, in this paper, we provide a new proof to show that if the covariates are generated as i.i.d. realizations of a random variable with $(4+\delta)$-th moment, Assumption 7 holds with high probability.

Armed with the new assumption, we are able to propose some bounds on $\sigma_{\text{hd},l}$ and $\sigma_{\text{hd},q}$, depending on α. By defining $R^2 := 1 - \frac{\sigma^2_{\text{adj}}}{\sigma^2_{\text{cre}}}$, which is equivalent to the empirical correlation between $\boldsymbol{X}$ and $r_1^{-1}Y(1) + r_0^{-1}Y(0)$, we have the following result:

Corollary 1. Under Assumptions 1, 2 and 7, as $n \to \infty$, we have

$$\begin{aligned}
&\sigma^2_{\text{hd},l} = \left[(1+\alpha)^2 - (1+2\alpha)R^2\right]\sigma^2_{\text{cre}} + o(1),\\
&0 \leq \sigma^2_{\text{hd},q} \leq 2(r_1r_0)^2\alpha(1-\alpha)S^2_{r_1^{-2}Y(1) - r_0^{-2}Y(0)} + o(1).
\end{aligned}$$

As a consequence, there is

$$o(1) \leq \sigma^2_{\text{hd}} - \left[(1+\alpha)^2 - (1+2\alpha)R^2\right]\sigma^2_{\text{cre}} \leq 2(r_1r_0)^2\alpha(1-\alpha)S^2_{r_1^{-2}Y(1) - r_0^{-2}Y(0)} + o(1).$$

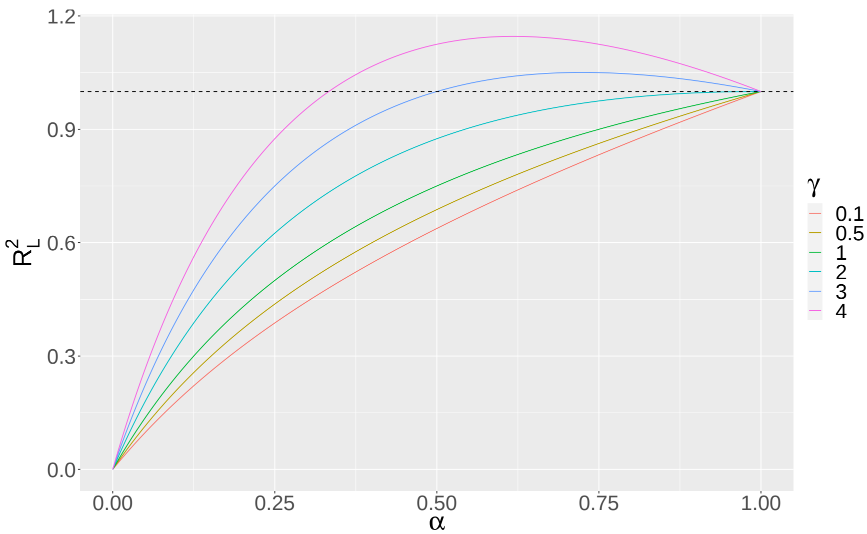


Figure 1: Curves of R_L^2 as a function of $\alpha := p/n$ with different magnitudes of $\gamma := \frac{S_\tau^2}{2S^2_{\frac{Y(1)+Y(0)}{2}}}$. The dashed line signifies 1.

Informally then, Corollary 1 shows that a necessary condition for the debiased estimator to give a variance smaller than that of $\hat{\tau}_{\text{unadj}}$ is

$$(1+\alpha)^2 - (1+2\alpha)R^2 - 1 < 0 \Leftrightarrow R^2 > \frac{\alpha^2+2\alpha}{1+2\alpha}. \tag{7}$$

At the same time, a sufficient condition for the debiased estimator to give a variance reduction is

$$R^2 > \frac{\alpha^2+2\alpha}{1+2\alpha} + 2r_1r_0\frac{\alpha(1-\alpha)}{1+2\alpha}\frac{S^2_{r_1^{-2}Y(1)-r_0^{-2}Y(0)}}{S^2_{r_1^{-1}Y(1)+r_0^{-1}Y(0)}}. \tag{8}$$

When $r_1 \equiv r_0 \equiv \frac{1}{2}$, then the above inequality can be written as

$$R^2 > \underbrace{\frac{\alpha^2+2\alpha}{1+2\alpha} + \frac{\alpha(1-\alpha)}{1+2\alpha}\frac{S_\tau^2}{2S^2_{\frac{Y(1)+Y(0)}{2}}}}_{=:R_L^2}.$$

We denote the right-hand side of the above inequality by R_L^2. Informally, the magnitude of R_L^2 depends on two quantities: the first is the covariate-dimension-to-sample size ratio α; the second is a scaled ratio between the variance of individual treatment effect $\tau_i := Y_i(1) - Y_i(0)$ and the variance of the average of two potential outcomes $\frac{Y_i(1)+Y_i(0)}{2}$, which we denote by $\gamma := \frac{S_\tau^2}{2S^2_{\frac{Y(1)+Y(0)}{2}}}$.

Figure 1 illustrates the dependency of R_L^2 on α and γ. Apparently, with a decreasing γ, R_L^2 decreases monotonically. This indicates that when the individual effects have smaller heterogeneity, less dependency between the potential outcomes and covariates is required for the debiased estimator to have an efficiency improvement compared to the unadjusted estimator. When γ is small,

R_L^2 increases monotonically as α goes from 0 to 1. When γ is increased to above 2, the trend then follows a different pattern. Regardless of the magnitude of γ, R_L^2 reduces to zero as α goes to zero. This is consistent with the theoretical findings in the low-dimensional setting. When α approaches 1, both R_L^2 and the lower bound in (7) approaches 1. This implies that we usually cannot achieve any efficiency improvement when p is close to n. This also implies that adding more covariates to the regression will usually result in a phase transition from "mostly harmless" to "harmful" to the post-experiment analysis. In practice, we recommend practitioners to choose a moderate number of covariates so that R^2 is above R_L^2.

When α is small, say 0.1, and we restrict γ to be no larger than 2, R_L^2 is at most 0.325. We believe this already includes a large number of cases in practical applications; when $\alpha = 0.5$, one may require a relatively low γ to keep R_L^2 away from 1. Of course, since $R^2 > R_L^2$ is just a sufficient condition for our estimator to be more accurate than without using regression adjustment at all, in practice one may still observe an improved accuracy even when this is violated.

4 Inference

Inference on $\hat{\tau}_{\text{db}}$ relies on a valid estimation of σ_{hd}^2, so that one can construct asymptotically valid Wald-type confidence intervals. We will derive the formula of the variance estimator in this section and show that this variance estimator is asymptotically valid with moderately high-dimensional covariates.

Our new inferential technique is constructed by a decomposition of σ_{hd}^2, which is given below in (13). In order to describe this decomposition, we define:

$$\boldsymbol{B} := \boldsymbol{M}^\top \boldsymbol{M}, \quad \text{where} \quad \boldsymbol{M} := \left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^\top\right) - \boldsymbol{H} + \left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^\top\right)\text{diag}\{\boldsymbol{H}\}, \tag{9}$$

where with a slight abuse of notation, we take $\text{diag}\{\boldsymbol{H}\}$ as a diagonal matrix with (i,i)-th entry equal to H_{ii}. As will be demonstrated in the Supplementary Material, with the above notation, we are able to rewrite σ_{hd} as

$$\sigma_{\text{hd}}^2 = (r_1 r_0) S^2_{\boldsymbol{B},\, r_1^{-1}Y(1) + r_0^{-1}Y(0)} + (r_1 r_0)^2 S^2_{\boldsymbol{Q},\, r_1^{-2}Y(1) - r_0^{-2}Y(0)}. \tag{10}$$

Now, using that for any *symmetric* matrix $\boldsymbol{A}$ and any vectors $\boldsymbol{a}, \boldsymbol{b}$,

$$S_{\boldsymbol{A},\boldsymbol{a},\boldsymbol{b}} = S_{\text{diag}\{\boldsymbol{A}\},\boldsymbol{a},\boldsymbol{b}} + S_{\text{diag}^-\{\boldsymbol{A}\},\boldsymbol{a},\boldsymbol{b}}, \tag{11}$$

where $\text{diag}^-\{\boldsymbol{A}\}$ corresponds to the matrix with all diagonal entries equal to zero and all off-diagonal entries equal to the off-diagonal entries of $\boldsymbol{A}$, and

$$S^2_{\boldsymbol{A},\boldsymbol{a}+\boldsymbol{b}} = S^2_{\boldsymbol{A},\boldsymbol{a}} + S^2_{\boldsymbol{A},\boldsymbol{b}} + 2S_{\boldsymbol{A},\boldsymbol{a},\boldsymbol{b}}, \tag{12}$$

(which we will clarify in the Supplementary Material), we further decompose σ^2_{hd} as

$$\begin{aligned}
\sigma^2_{\mathrm{hd}} := & \underbrace{(r_1 r_0) \sum_{z\in\{0,1\}} \left(S^2_{r_1 r_0 \,\mathrm{diag}\{\boldsymbol{Q}\},\, r_z^{-2} Y(z)} + S^2_{r_z^{-2}\,\mathrm{diag}\{\boldsymbol{B}\},\, Y(z)} \right)}_{=:\mathcal{I}_1} \\
& + \underbrace{(r_1 r_0) \sum_{z\in\{0,1\}} \left(S^2_{r_1 r_0 \,\mathrm{diag}^-\{\boldsymbol{Q}\},\, r_z^{-2} Y(z)} + S^2_{r_z^{-2}\,\mathrm{diag}^-\{\boldsymbol{B}\},\, Y(z)} \right)}_{=:\mathcal{I}_2} \\
& + \underbrace{2S_{\mathrm{diag}\{\boldsymbol{B}\}, Y(1), Y(0)} - 2S_{\mathrm{diag}\{\boldsymbol{Q}\}, Y(1), Y(0)}}_{=:\mathcal{I}_3} \\
& + \underbrace{2S_{\mathrm{diag}^-\{\boldsymbol{B}\}, Y(1), Y(0)} - 2S_{\mathrm{diag}^-\{\boldsymbol{Q}\}, Y(1), Y(0)}}_{=:\mathcal{I}_4}.
\end{aligned} \tag{13}$$

Informally, $\mathcal{I}_1$ and $\mathcal{I}_2$ correspond to the variances of a single world, and $\mathcal{I}_3$ and $\mathcal{I}_4$ correspond to the covariance of counterfactual worlds.

Armed with the above decomposition, we construct an estimation of σ^2_{hd} by estimating $\mathcal{I}_1, \ldots, \mathcal{I}_4$ separately. Since they represent variances from different sources, we need different estimation strategies for each term. We first consider $\mathcal{I}_1$ and $\mathcal{I}_2$. Since these quantities are quadratic functions of potential outcomes of a single arm, they can be consistently estimated using empirical observations from a single arm. Specifically, we can estimate $\mathcal{I}_1$ via

$$\hat{\mathcal{I}}_1 := (r_1 r_0) \sum_{z\in\{0,1\}} \left(s^2_{r_1 r_0 \,\mathrm{diag}\{\boldsymbol{Q}\},\, r_z^{-2} Y(z)} + s^2_{r_z^{-2}\,\mathrm{diag}\{\boldsymbol{B}\},\, Y(z)} \right),$$

where $s^2_{r_1 r_0 \,\mathrm{diag}\{\boldsymbol{Q}\},\, r_z^{-2} Y(z)}$ and $s^2_{r_z^{-2}\,\mathrm{diag}\{\boldsymbol{B}\},\, Y(z)}$ are empirical estimates of their oracle versions using samples from treatment arm z. For example, we write

$$s^2_{r_1 r_0 \,\mathrm{diag}\{\boldsymbol{Q}\},\, r_z^{-2} Y(z)} := \frac{1}{n_z} \sum_{i: Z_i = z} r_1 r_0 Q_{ii} (r_z^{-2} Y_i - r_z^{-2} \bar{Y}_z)^2.$$

We now consider $\hat{\mathcal{I}}_2$. Since it involves cross-sample products, we define $s^2_{r_1 r_0 \,\mathrm{diag}^-\{\boldsymbol{Q}\},\, r_z^{-2} Y(z)}$ instead as

$$s^2_{r_1 r_0 \,\mathrm{diag}^-\{\boldsymbol{Q}\},\, r_z^{-2} Y(z)} := \frac{1}{r_z n_z} \sum_{i\neq j: Z_i, Z_j = z} r_1 r_0 Q_{ij} (r_z^{-2} Y_i - r_z^{-2} \bar{Y}_z)(r_z^{-2} Y_j - r_z^{-2} \bar{Y}_z)$$

and similarly for $s^2_{r_z^{-2}\,\mathrm{diag}^-\{\boldsymbol{B}\},\, Y(z)}$.

Finally, we discuss the estimation of $\mathcal{I}_3$ and $\mathcal{I}_4$. Since $\mathcal{I}_3$ corresponds to the covariances of potential outcomes from two worlds, it cannot be estimated consistently from observed data directly. Instead, it is only identifiable up to an upper bound. As we will show in the proof of Theorem 3, $\mathcal{I}_3$ can be decomposed as

$$\begin{aligned}
\mathcal{I}_3 = & \sum_{z\in\{0,1\}} \left(S^2_{\mathrm{diag}\{\boldsymbol{B}\}, Y(z)} - S^2_{\mathrm{diag}\{\boldsymbol{Q}\}, Y(z)} - S^2_{\mathrm{diag}^-\{\boldsymbol{H}\}, Y(z)} \right) + 2S_{\mathrm{diag}^-\{\boldsymbol{H}\}, Y(1), Y(0)} \\
& - S^2_{\mathrm{diag}\{\boldsymbol{H}\}, Y(1)-Y(0)} - S^2_{e(1)-e(0)} + O(n^{-1}),
\end{aligned}$$

where the last two terms ($S^2_{\text{diag}\{\boldsymbol{H}\},Y(1)-Y(0)}$ and $S^2_{e(1)-e(0)}$) represent treatment effect variation and thus can not be estimated consistently. Fortunately, the last two terms are non-negative; this allows us to provide a consistent estimation of an upper bound of $\mathcal{I}_3$ just with the first two terms in the above decomposition, which we denote by $\mathcal{I}_{3,\text{ub}}$. Noteworthy, besides the variance of a single world, the $\mathcal{I}_{3,\text{ub}}$ involves a term representing the covariance of counterfactual worlds. We define its empirical estimate as

$$s_{\text{diag}^-\{\boldsymbol{H}\},Y(1),Y(0)} := \frac{1}{nr_1r_0} \sum_{i\neq j: Z_i=1, Z_j=0} H_{ij}(Y_i - \bar{Y}_1)(Y_j - \bar{Y}_0)$$

and similarly we can define $s_{\text{diag}^-\{\boldsymbol{B}\},Y(1),Y(0)}$ and $s_{\text{diag}^-\{\boldsymbol{Q}\},Y(1),Y(0)}$. Therefore, we have an empirical estimate of $\mathcal{I}_{3,\text{ub}}$ as

$$\hat{\mathcal{I}}_{3,\text{ub}} = \sum_{z\in\{0,1\}} \left(s^2_{\text{diag}\{\boldsymbol{B}\},Y(z)} - s^2_{\text{diag}\{\boldsymbol{Q}\},Y(z)} - s^2_{\text{diag}^-\{\boldsymbol{H}\},Y(z)} \right) + 2s_{\text{diag}^-\{\boldsymbol{H}\},Y(1),Y(0)}.$$

For $\mathcal{I}_4$, mimicking the estimates for $\mathcal{I}_3$, we propose to estimate it via

$$\hat{\mathcal{I}}_4 = 2\left(s_{\text{diag}^-\{\boldsymbol{B}\},Y(1),Y(0)} - s_{\text{diag}^-\{\boldsymbol{Q}\},Y(1),Y(0)} \right).$$

Putting together, we get the variance estimator $\hat{\sigma}^2_{\text{hd}} := \hat{\mathcal{I}}_1 + \hat{\mathcal{I}}_2 + \hat{\mathcal{I}}_{3,\text{ub}} + \hat{\mathcal{I}}_4$. The following theorem characterizes the asymptotic convergence of this estimator.

Theorem 3. *If Assumptions 1–3 and 5 hold, we have*

$$\hat{\sigma}^2_{hd} = \sigma^2_{adj} + S^2_{e(1)-e(0)} + o_{\mathbb{P}}(1).$$

Otherwise, if Assumptions 1–3 hold, we have

$$\hat{\sigma}^2_{hd} = \sigma^2_{hd} + S^2_{e(1)-e(0)} + S^2_{\text{diag}\{\boldsymbol{H}\},Y(1)-Y(0)} + o_{\mathbb{P}}(1).$$

Due to the unidentifiability of the counterfactual covariance, the estimated $\hat{\sigma}^2_{\text{hd}}$ contains a variance inflation. In the regime $p = o(n)$, our variance estimation has the same inflation as in the lower dimensional regime where $p = O(n^{2/3}/(\log n)^{1/3})$, see Lei and Ding [2021]. This variance inflation is always no greater than the usual inflation *without* any covariate adjustment, namely S^2_τ. Nevertheless, in the regime $p \asymp n$, the variance inflation $S^2_{e(1)-e(0)} + S^2_{\text{diag}\{\boldsymbol{H}\},Y(1)-Y(0)}$ is not always smaller than S^2_τ, especially when there is strong co-linearity between H_{ii} and τ_i. We will demonstrate this in numerical analysis. On the other hand, as we will show in Section 5, when the data exhibit sufficient linearity and light tail, one can still expect $S^2_{e(1)-e(0)} + S^2_{\text{diag}\{\boldsymbol{H}\},Y(1)-Y(0)} < S^2_\tau$.

We now showcase an alternative variance estimator with variance inflation equal to S^2_τ. In this estimator, instead of estimating $\mathcal{I}_{3,\text{ub}}$, we focus on the following alternative upper bound of $\mathcal{I}_3$:

$$\mathcal{I}'_{3,\text{ub}} := \sum_{z\in\{0,1\}} \left(S^2_{\text{diag}\{\boldsymbol{B}\},Y(z)} - S^2_{\text{diag}\{\boldsymbol{Q}\},Y(z)} \right),$$

which can be estimated via

$$\hat{\mathcal{I}}'_{3,\text{ub}} := \sum_{z\in\{0,1\}} \left(s^2_{\text{diag}\{\boldsymbol{B}\},Y(z)} - s^2_{\text{diag}\{\boldsymbol{Q}\},Y(z)} \right).$$

Armed with $\hat{\mathcal{I}}'_{3,\text{ub}}$, we define the alternative variance estimator as $(\hat{\sigma}'_{\text{hd}})^2 := \hat{\mathcal{I}}_1 + \hat{\mathcal{I}}_2 + \hat{\mathcal{I}}'_{3,\text{ub}} + \hat{\mathcal{I}}_4$. The following corollary characterizes the asymptotic convergence of the alternative variance estimator.

Corollary 2. If Assumptions 1–3 and 5 hold, we have

$$(\hat{\sigma}'_{\mathrm{hd}})^2 = \sigma^2_{\mathrm{adj}} + S^2_\tau + o_{\mathbb{P}}(1).$$

Otherwise, if Assumptions 1–3 hold, we have

$$(\hat{\sigma}'_{\mathrm{hd}})^2 = \sigma^2_{\mathrm{hd}} + S^2_\tau + o_{\mathbb{P}}(1).$$

In practice, we recommend the use of $\min\{\hat{\sigma}^2_{\mathrm{hd}}, (\hat{\sigma}'_{\mathrm{hd}})^2\}$ for a shorter confidence interval. Then, no matter $S^2_{e(1)-e(0)} + S^2_{\mathrm{diag}\{\boldsymbol{H}\},Y(1)-Y(0)} > S^2_\tau$ or not, the variance inflation is always no greater than S^2_τ, i.e., the inflation without using any covariate adjustment. Therefore, the confidence interval length from our inferential procedure is always asymptotically shorter than the unadjusted estimator whenever $\sigma_{\mathrm{hd}} < \sigma_{\mathrm{cre}}$. In practice, we recommend constructing confidence intervals using both our procedure and the unadjusted estimator and choosing the shorter one for downstream analysis.

Since the inferential procedure of Lin [2013] may behave poorly in practice when the covariate dimension is relatively large, existing literature recommended to use HC3-type standard error to boost finite sample performance which heavily penalizes dimension p used by the analysis. However, the HC3-type standard error is typically conservative and has no theoretical guarantee in the moderately high-dimensional regime. Based on our theory, we provide an inference procedure that is valid under this regime and the estimated variance is "tight" in that the bias is the variance of unit-level treatment effect which can not be estimated from data.

5 Justification of assumptions

In this section, we justify Assumptions 5–7. The following proposition implies that Assumption 5 holds with high probability if the potential outcomes are i.i.d. generated from a superpopulation with bounded variance.

Proposition 1. *Fix $z \in \{0,1\}$. If $p = o(n)$ and $\{Y_i(z)\}_{i=1}^n$ are i.i.d. random variables with* $\mathrm{var}(Y_1(z)) < \infty$*, then there exists a positive sequence $c_n \to 0$ such that*

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{p}\left(Y_{(i)}(z) - \bar{Y}(z)\right)^2 > c_n\right) \to 0.$$

We now focus on the justification of Assumption 7. The assumptions on $Y_i(z)$'s are common in randomization-based literature; therefore, we only need to justify the assumptions on $\boldsymbol{H}$. Since $\max_{1\le i\le n}|H_{ii} - \alpha| \to 0$ is a sufficient assumption for $\sum_{i=1}^n (H_{ii}-\alpha)^2/n \to 0$, we only need to show that the former condition holds with high probability under some superpopulation assumption on $\boldsymbol{X}_i$'s. In Proposition 2, we show that when $\boldsymbol{X}_i$'s are i.i.d. realizations from some superpopulation with entry-wise bounded $(4+\eta)$-th order moment up to some transformation, $\max_{1\le i\le n}|H_{ii}-\alpha| \to 0$ holds with high probability.

Proposition 2. *Suppose that $\{\boldsymbol{X}_i\}_{i=1}^n$ are i.i.d. random vectors generated from independent random variables as $\boldsymbol{X}_i = \mathbf{O}\boldsymbol{V}_i$, where $\mathbf{O}$ is a deterministic non-singular matrix and $\boldsymbol{V}_i$ have independent entries with mean 0, variance 1, and $\max_j \mathbb{E}|V_1(j)|^{4+\eta} < C$ for some constants $\eta, C > 0$;*

suppose also $\limsup_{n\to\infty}\alpha = \limsup_{n\to\infty}(p/n) < 1$. *Then, for any constant* δ *satisfying that* $0 < \delta < \frac{\eta}{8+2\eta}$,

$$\mathbb{P}\left(\max_i |H_{ii} - \alpha| > n^{-\delta}\right) \to 0. \tag{14}$$

Finally, we turn to Assumption 6. In fact, it can be justified by simply applying Corollary 1. Specifically, under Assumptions 1, 2 and 7, we have $\liminf_{n\to\infty}\sigma^2_{\text{hd},l} > 0$ when either (i) Assumption 4 holds or (ii) $\liminf_{n\to\infty}\sigma^2_{\text{cre}} > 0$ and $\liminf_{n\to\infty}(p/n) > 0$. These requirements are natural in randomization-based literature.

Finally, we investigate the relationship between S^2_τ and $S^2_{\text{diag}\{\boldsymbol{H}\},Y(1)-Y(0)} + S^2_{e(1)-e(0)}$ under a superpopulation framework where $(Y_i(1), Y_i(0), \boldsymbol{X}_i, \varepsilon_i(1), \varepsilon_i(0))$ are i.i.d. generated from some distribution. We assume a linear model where $\boldsymbol{X}_i, \varepsilon_i(1), \varepsilon_i(0)$ are independent and μ_z, $z \in \{0,1\}$, are deterministic scalars:

$$Y_i(1) = \mu_1 + \boldsymbol{X}_i^\top \boldsymbol{\beta}_1 + \varepsilon_i(1), \quad Y_i(0) = \mu_0 + \boldsymbol{X}_i^\top \boldsymbol{\beta}_0 + \varepsilon_i(0). \tag{15}$$

Proposition 3 shows that under this superpopulation framework, the confidence interval given by $\hat{\sigma}^2_{\text{hd}}$ is asymptotically no larger than the confidence interval from $(\hat{\sigma}'_{\text{hd}})^2$ with high probability.

Proposition 3. *Under model (15), for* $z \in \{0,1\}$, *we assume that* $\mathbb{E}|\varepsilon_1(z)|^4 < C$ *and* $\mathbb{E}|\boldsymbol{X}_1^\top\boldsymbol{\beta}_z|^4 < C$ *for some constant* $C > 0$, *and* $\boldsymbol{X}_i$ *satisfies the conditions in Proposition 2. Then, there exists a positive sequence* $c_n \to 0$ *such that*

$$\mathbb{P}\left(S^2_\tau - S^2_{e(1)-e(0)} - S^2_{\text{diag}\{\boldsymbol{H}\},Y(1)-Y(0)} > -c_n\right) \to 0.$$

6 Numerical analysis

In this section, we conduct a numerical analysis to examine the finite sample performance of our debiased estimator and its corresponding inference procedure, together with several competitors.

6.1 Experimental setup

Pre-treatment variable generation Let Scale() be a standardization function: for a finite population $\{a_i\}_{i=1}^n$ with $\boldsymbol{a} = (a_1, \ldots, a_n)$, $\text{Scale}(a_i) := (a_i - \bar{a})/\left(\sum_{i=1}^n (a_i - \bar{a})^2/n\right)^{1/2}$ and $\text{Scale}(\boldsymbol{a}) = (\text{Scale}(a_1), \ldots, \text{Scale}(a_n))$. Set $n = 1000$ and $r_1 = 0.35$. We first generate a matrix $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n\times n}$ and two vectors, $\boldsymbol{\beta} \in \mathbb{R}^n$ and $\boldsymbol{\Delta} \in \mathbb{R}^n$, with i.i.d. entries from t distribution with 3 degrees of freedom. We keep $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\beta}$ and $\boldsymbol{\Delta}$ fixed throughout the simulation. For each covariate-dimension-to-sample-size-ratio α, let $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)^\top$ be the first $p := \alpha n$ columns of $\boldsymbol{\mathcal{X}}$. We generate the potential outcomes according to the following model

$$Y_i(1) = \mu_1 + \text{Scale}(\boldsymbol{X}_i^\top\boldsymbol{\beta}_1) + \varepsilon_i(1)/\sqrt{\gamma}; \quad Y_i(0) = \mu_0 + \text{Scale}(\boldsymbol{X}_i^\top\boldsymbol{\beta}_0) + \varepsilon_i(0)/\sqrt{\gamma}.$$

Here μ_z ($z \in \{0,1\}$) are generated i.i.d. from t distribution with 3 degrees of freedom. For any vector $\boldsymbol{a}$, let $\boldsymbol{a}_{[p]}$ be the subvector of the first p elements; the coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$ are generated by

$$\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{[p]} + \delta\boldsymbol{\Delta}_{[p]}, \quad \boldsymbol{\beta}_0 = \boldsymbol{\beta}_{[p]} - \delta\boldsymbol{\Delta}_{[p]}.$$

The factor δ is introduced to control the heterogeneity of individual-level treatment effect.

For the noise terms, γ is the scaling factor for the magnitude of the noise. In addition, we consider 2 generating models of $\varepsilon_i(z)$:

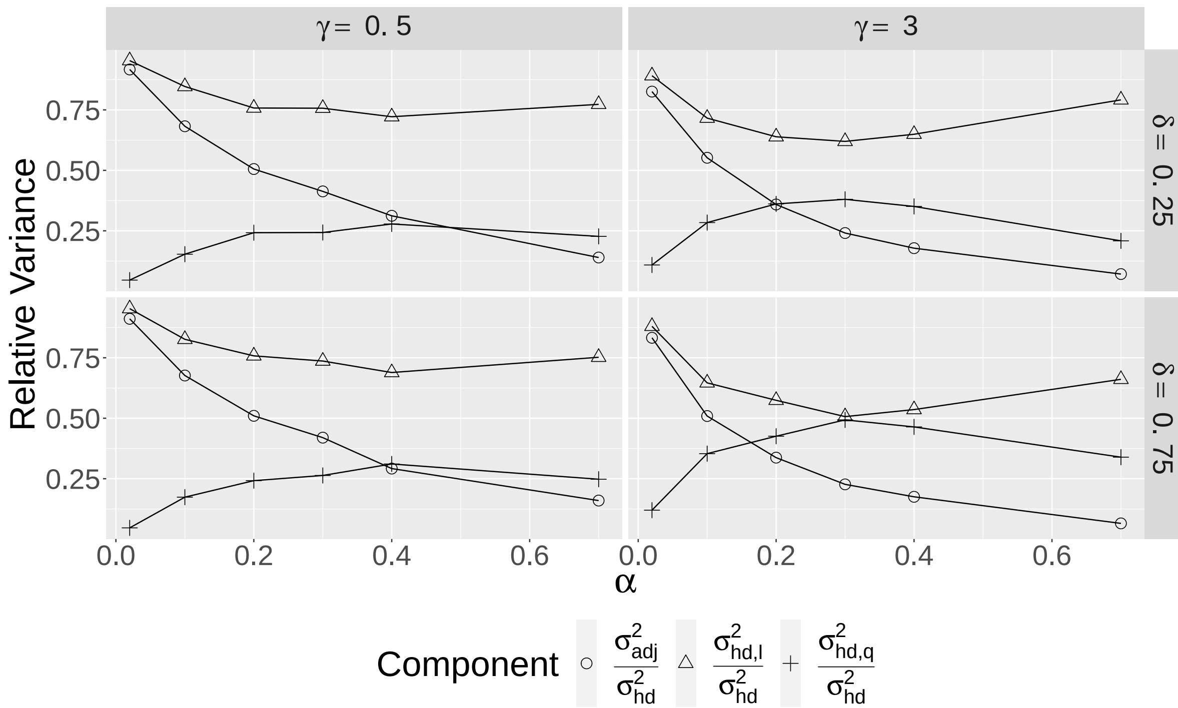


Figure 2: Relative size of variance components for different choices of γ, δ and α under the independent t residual.

- Worst-case residual: let $\boldsymbol{\varepsilon}(z) := (\varepsilon_1(z), \ldots, \varepsilon_n(z))$,

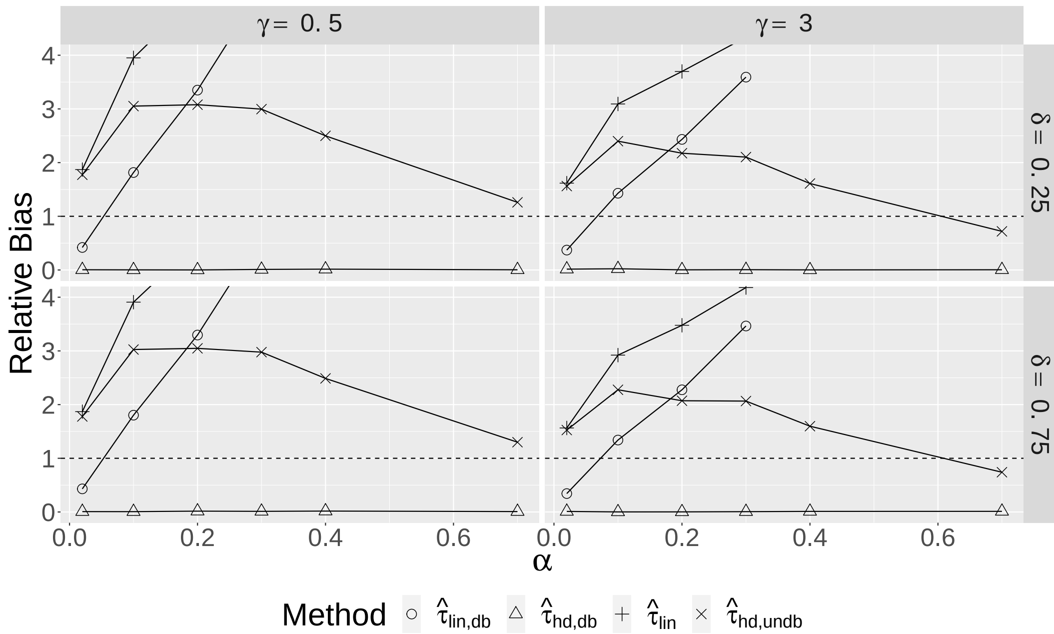
$$\boldsymbol{\varepsilon}(1) = \text{Scale}((\boldsymbol{I} - \boldsymbol{H})(H_{11}, \ldots, H_{nn})^\top); \quad \boldsymbol{\varepsilon}(0) = -2\,\text{Scale}((\boldsymbol{I} - \boldsymbol{H})(H_{11}, \ldots, H_{nn})^\top).$$

  This residual is motivated by Lei and Ding [2021] and produces a large bias for regression-adjusted estimators in theory.

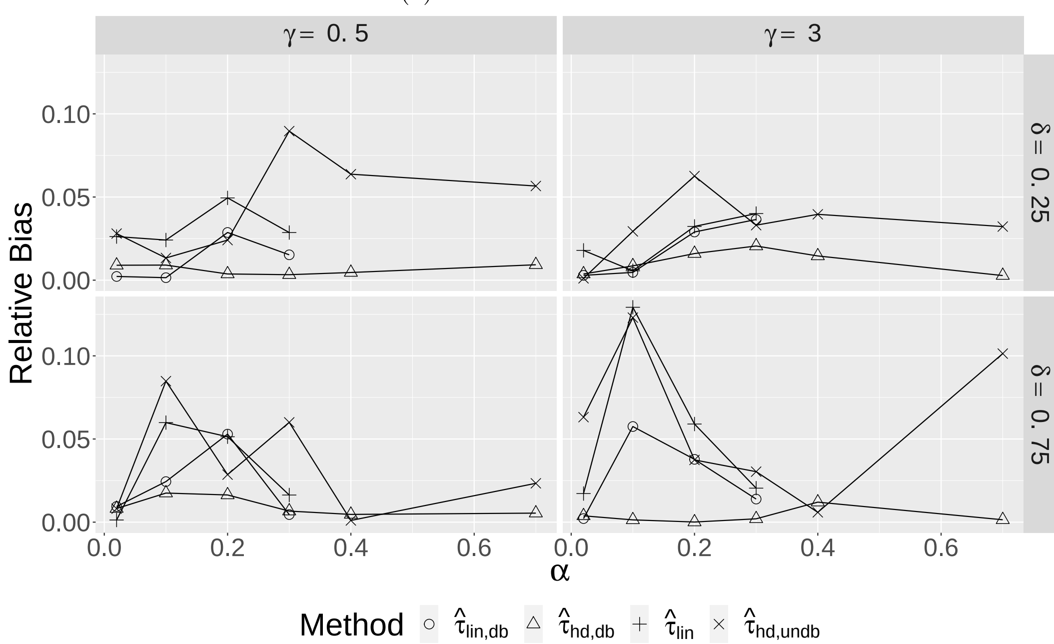
- t residual: $\varepsilon_i(z) = \text{Scale}(\breve{\varepsilon}_i(z))$. $\breve{\varepsilon}_i(z)$ is generated i.i.d. from t distribution with 3 degrees of freedom.

We view the simulation as a full factorial experiment and generate the data under all combinations of the following 4 factors: $\delta = \{0.25, 0.75\}$; $\gamma = \{0.5, 3\}$ and the covariate-dimension-to-sample-size-ratio $\alpha = \{0.02, 0.1, 0.2, 0.3, 0.4, 0.7\}$ and generating models of $\varepsilon_i(z)$. Throughout this section we fix $\boldsymbol{\mathcal{X}}$ to be generated from t distribution with 3 degrees of freedom; in the Supplementary Material we further provide simulations with $\boldsymbol{\mathcal{X}}$ generated from Cauchy distribution.

Repeated sampling evaluation Once the pre-treatment variables $\{(\boldsymbol{X}_i, Y_i(1), Y_i(0))\}_{i=1}^n$ are generated, we fix them and draw 10000 random assignments. For evaluation criterion, We consider the empirical relative root mean squared error (relative RMSE) defined by the empirical RMSE of the estimators divided by the oracle standard errors of the unadjusted estimator $\hat{\tau}_{\text{unadj}}$; and the empirical relative absolute bias defined by the absolute of the empirical bias divided by the asymptotic standard error of $\hat{\tau}_{\text{hd}}$, $\sigma_{\text{hd}}/\sqrt{n}$. For inference procedures, we then compare, under a 0.05 significance level, the empirical coverage probabilities and empirical averages of relative confidence



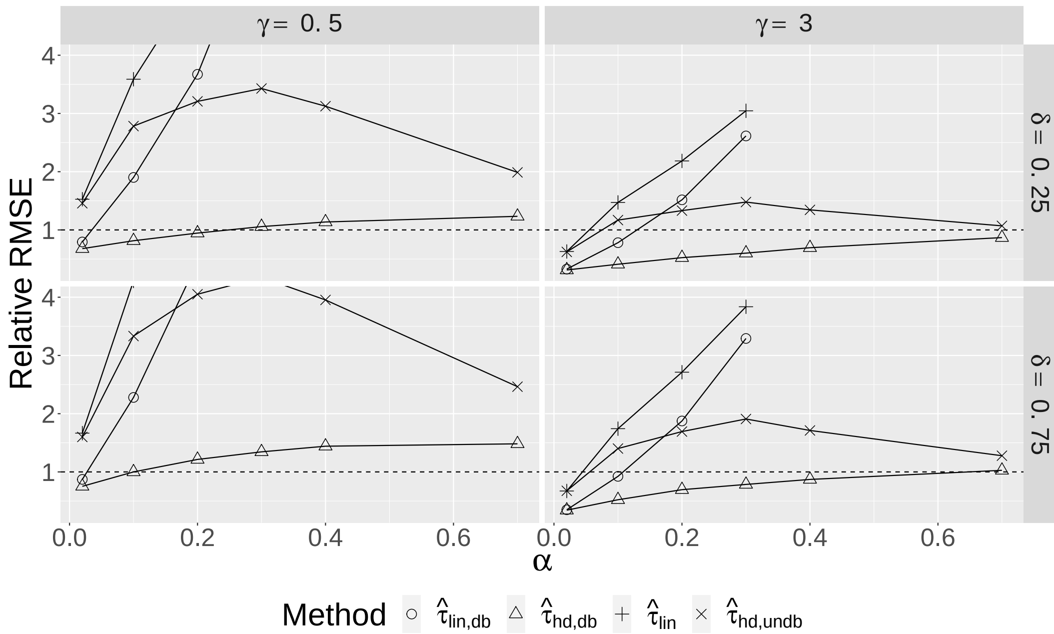
(a) Worst case residual



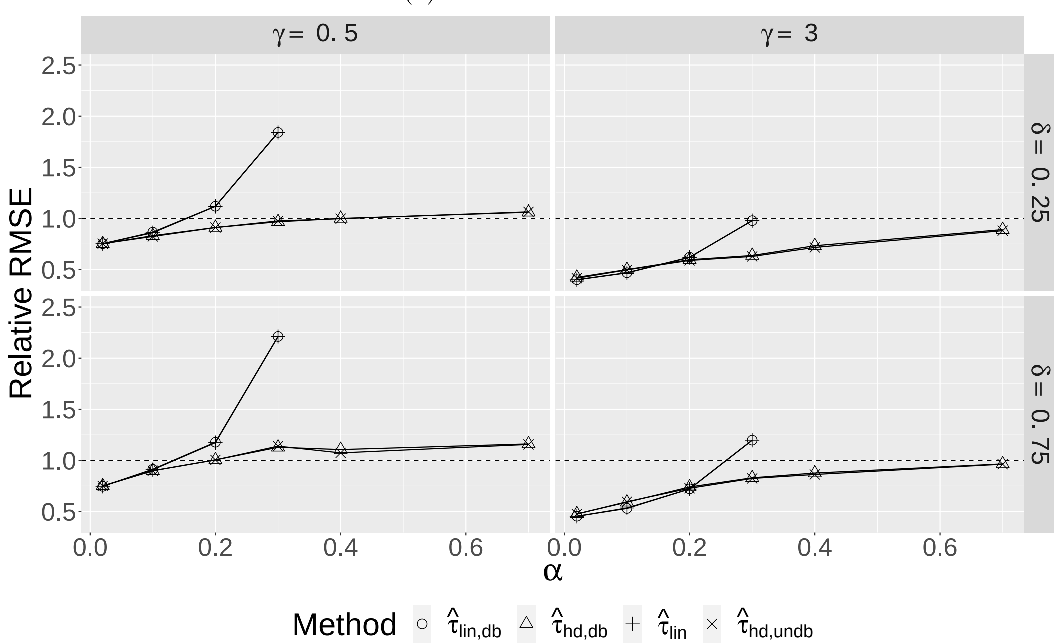
(b) Independent t residual

Figure 3: Relative bias for different choices of γ, δ and α under the worst-case residual and independent t residual. The dashed lines signify 1. Notice that for the first figure, we zoom in on part of the y-axis to better display the curve.

interval length defined by the corresponding confidence interval length divided by the length of the confidence interval constructed *without* covariate adjustment.



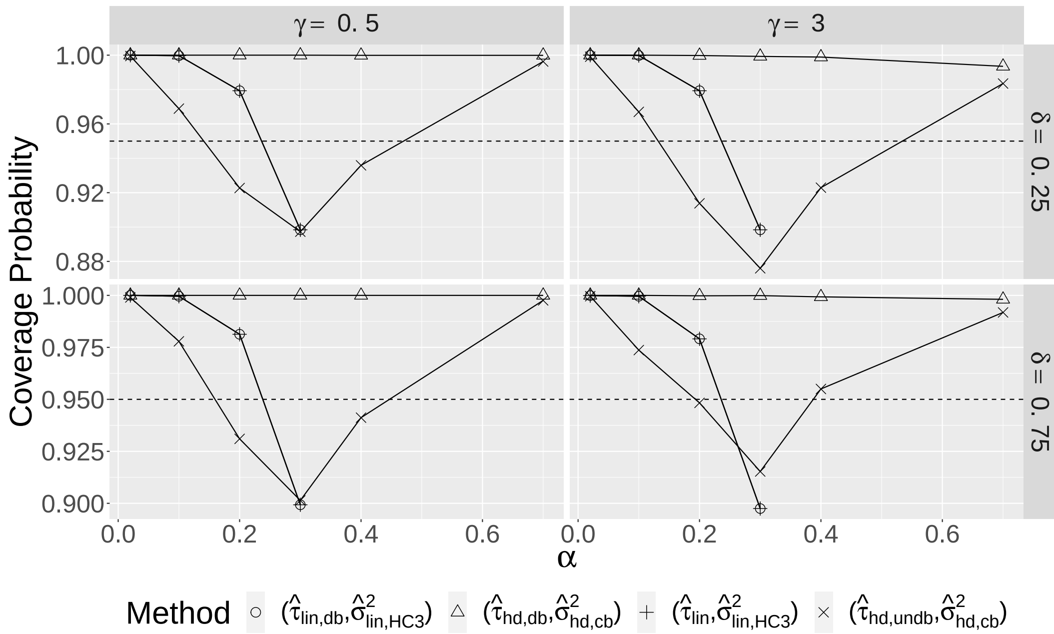
(a) worst-case residual



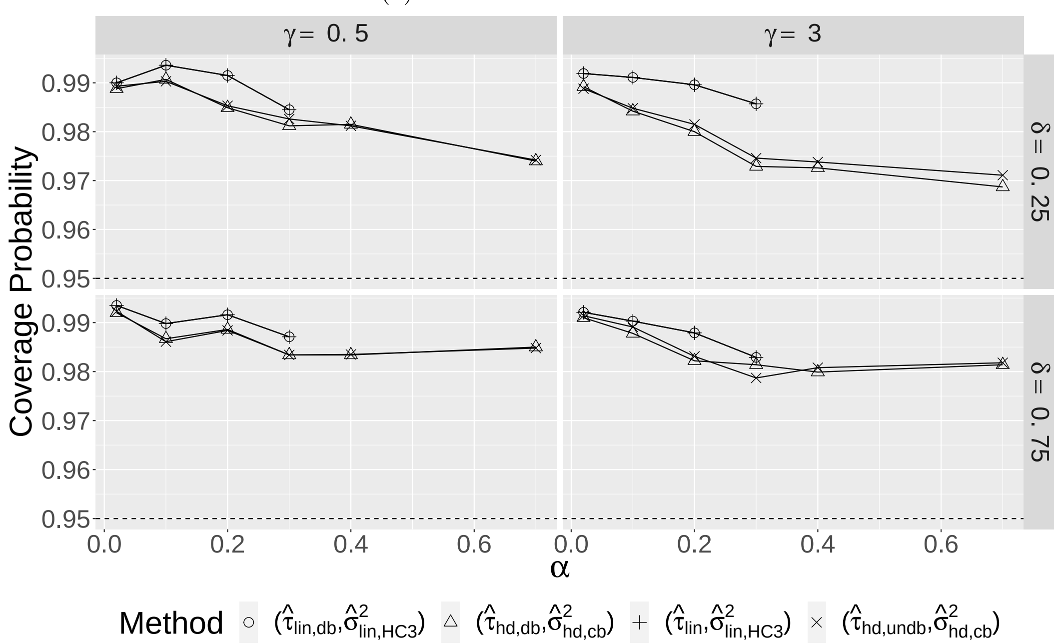
(b) independent t residual

Figure 4: Relative RMSE for different choices of γ, δ and α under the worst-case residual and independent t residual. The dashed lines signify 1. Notice that for the first figure, we zoom in on part of the y-axis to better display the curve.

Methods for comparison For estimators, we consider our proposed high-dimensional regression estimator $\hat{\tau}_{\text{hd}}$ and its "un-debiased" version $\hat{\tau}_{\text{hd,undb}}$, i.e., the one without the debiasing step in (5);



(a) the worst-case residual



(b) the independent t residual

Figure 5: Empirical coverage probabilities for different choices of γ, δ and α under the worst-case residual and independent t residual. The dashed lines signify 0.95.

Lin's regression estimator [Lin, 2013] $\hat{\tau}_{\text{lin}}$, i.e., $\hat{\tau}_{\text{adj}}$ with $\hat{\boldsymbol{\beta}}_z$ constructed by only using samples from treatment arm z, and its debiased version [Lei and Ding, 2021], $\hat{\tau}_{\text{lin,db}}$ defined by

$$\hat{\tau}_{\text{lin,db}} = \hat{\tau}_{\text{lin}} + \frac{n_0}{n_1^2} \sum_{Z_i=1} H_{ii}\hat{e}_i(1) - \frac{n_1}{n_0^2} \sum_{Z_i=0} H_{ii}\hat{e}_i(0),$$

where $\hat{e}_i(z) = Y_i - \bar{Y}_z - \hat{\boldsymbol{\beta}}_z^\top (\boldsymbol{X}_i - \bar{\boldsymbol{X}}_z)$ for $Z_i = z$ with $\bar{\boldsymbol{X}}_z$ being the sample mean of covariates under treatment arm z.

For the inference procedure, we consider 4 Wald-type confidence intervals based on the 4 point estimators and their corresponding variance estimators. In particular, for $\hat{\tau}_{\text{hd}}$ and $\hat{\tau}_{\text{hd,undb}}$, we pair them with our recommended variance estimator $\hat{\sigma}^2_{\text{hd,cb}} := \min\{\hat{\sigma}^2_{\text{hd}}, (\hat{\sigma}'_{\text{hd}})^2\}$ ("cb" for combine); for $\hat{\tau}_{\text{lin}}$ and $\hat{\tau}_{\text{lin,db}}$, we pair them with HC3 variance estimator $\hat{\sigma}^2_{\text{lin,HC3}}$ defined by

$$\hat{\sigma}^2_{\text{lin,HC3}} := \frac{n}{(n_1-1)n_1} \sum_{i:Z_i=1} \frac{\hat{e}_i(1)^2}{(1-H_{ii,1})^2} + \frac{n}{(n_0-1)n_0} \sum_{i:Z_i=0} \frac{\hat{e}_i(0)^2}{(1-H_{ii,0})^2},$$

where and $H_{ii,z} = (\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top \left\{\sum_{j:Z_j=z} (\boldsymbol{X}_j - \bar{\boldsymbol{X}})(\boldsymbol{X}_j - \bar{\boldsymbol{X}})^\top\right\}^{-1} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})$.

6.2 Results

Relative magnitude of variance components Figure 2 shows relative magnitude of $\sigma^2_{\text{hd},l}$, $\sigma^2_{\text{hd},q}$, σ^2_{adj} divided by σ^2_{hd}. With a relatively high dimension, the quadratic component $\sigma^2_{\text{hd},q}$ is non-negligible and, sometimes, becomes the main source of the asymptotic variance. Besides, as the dimension increases, σ^2_{adj} becomes an inaccurate approximation even for the linear component of variance, $\sigma^2_{\text{hd},l}$, not to mention σ^2_{hd}.

The effectiveness of debiasing Figure 3 shows the relative bias of different methods under the worst-case residual and independent t residual, respectively. Apparently, the bias of $\hat{\tau}_{\text{hd}}$ is significantly below 1 in all cases, even under large α and worst-case residual. At the same time, under the worst-case residual, the relative bias of $\hat{\tau}_{\text{hd,undb}}$ can be significantly above 1. This suggests the necessity of debiasing. Moreover, under again the worst-case residual, not only $\hat{\tau}_{\text{lin}}$ but also $\hat{\tau}_{\text{lin,db}}$ have explosive growth in bias as α grows. $\hat{\tau}_{\text{lin,db}}$ has a bias smaller than $\hat{\tau}_{\text{lin}}$; nevertheless its bias is still non-negligible under all the worst-case residual setups, except for $\alpha = 0.02$. This is consistent with the theory of Lei and Ding [2021] requiring p tending to infinity slow enough.

Relative RMSE Figure 4 shows the relative RMSE of different methods under the worst-case residual and independent t residual. Under the worst-case residual, $\hat{\tau}_{\text{hd}}$ has the best performance among the 4 methods. The other methods have large RMSE due to their non-negligible bias. When signal to noise ratio is relatively high ($\gamma = 3$), $\hat{\tau}_{\text{hd}}$ can exploit the covariate information to keep relative RMSE smaller than 1, i.e., to perform better than the unadjusted estimator, even with high dimensions. When the signal-to-noise ratio is low ($\gamma = 0.5$) and the degree of heterogeneity is high $\delta = 0.75$, the efficiency improvement from our estimator is less compelling. This is consistent with our theory that the empirical correlation needs to be relatively larger than α to secure efficiency improvement.

Under the independent t residual, when α is small (say less than 0.1), all methods tend to have similar RMSE. As α gets larger, their RMSE's begin to diverge. Interestingly, both $\hat{\tau}_{\text{hd}}$ and $\hat{\tau}_{\text{hd,undb}}$ have the smallest relative RMSE, whilst the relative bias of $\hat{\tau}_{\text{hd}}$ is obviously smaller than $\hat{\tau}_{\text{hd,undb}}$ with independent t residual.

Finally, by checking the figures of both types of residual, we can conclude that when the signal is weak ($\gamma = 0.5$) and the dimension is high, no method can guarantee improvement. But even in the least favorable case, the relative RMSE of $\hat{\tau}_{\text{hd}}$ is just slightly above 1. In other words, $\hat{\tau}_{\text{hd}}$ never does significant harm to RMSE.

Inference performance Figure 5 shows the empirical coverage probabilities of different methods under the worst-case residual and independent t residual. Only the combination of $(\hat{\tau}_{\text{hd}}, \hat{\sigma}_{\text{hd,cb}})$ gives a valid empirical coverage in all cases. With worst-case residual, the other methods cannot guarantee a correct empirical coverage as α grows.

Figure 6 shows the relative confidence interval length produced by $\hat{\sigma}_{\text{hd,cb}}$ and $\hat{\sigma}_{\text{lin,HC3}}$. The trend for the curve of $\hat{\sigma}_{\text{hd,cb}}$ is very similar to that of $\hat{\tau}_{\text{hd}}$ in Figure 4. In particular, as long as the relative RMSE of $\hat{\tau}_{\text{hd}}$ is less than 1, the relative confidence interval length of $\hat{\sigma}_{\text{hd,cb}}$ is also less than 1. This echoes our discussion of Corollary 2. In the least favorable case ($\gamma = 0.5$, $\delta = 0.75$, $\alpha = 0.7$, worst-case residual), the relative confidence interval length is about 1.1, better than the relative RMSE of $\hat{\tau}_{\text{hd}}$ (about 1.5).

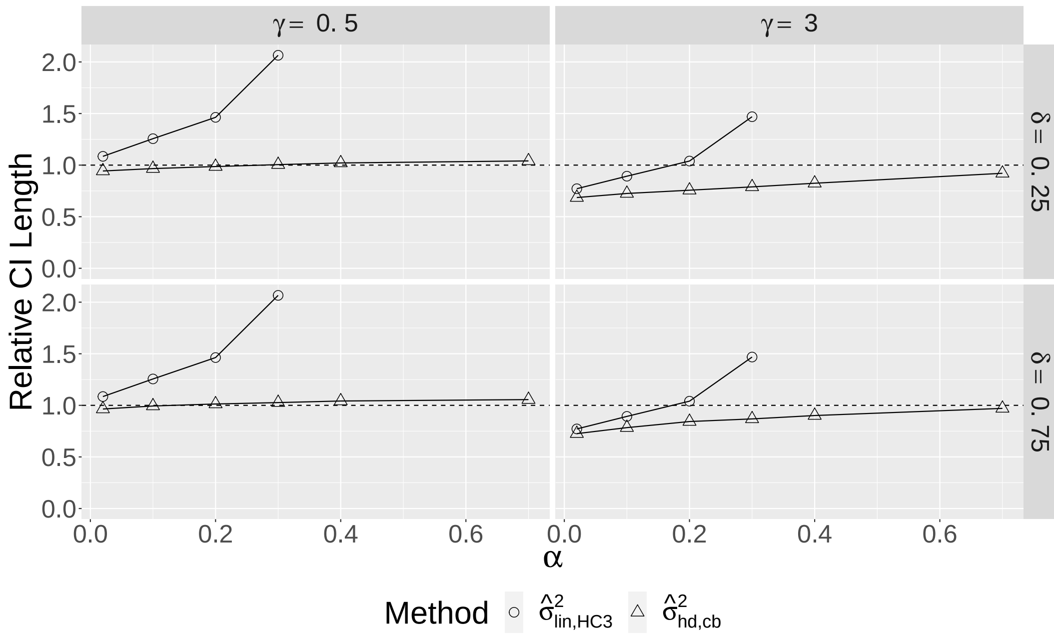
Usefulness of $\hat{\sigma}_{\text{hd,cb}}$ Figures 7b and 7a demonstrate the ratio of $\hat{\sigma}^2_{\text{hd}}$ to ${\hat{\sigma}'_{\text{hd}}}^2$. Under independent t residual, $\hat{\sigma}^2_{\text{hd}}$ is smaller which echoes Proposition 3. whilst under the worst-case residual ${\hat{\sigma}'_{\text{hd}}}^2$ is overall smaller. This supports our claim that $\hat{\sigma}^2_{\text{hd,cb}}$ improves the estimation precision by taking the advantages of both $\hat{\sigma}^2_{\text{hd}}$ and ${\hat{\sigma}'_{\text{hd}}}^2$.

7 Conclusion

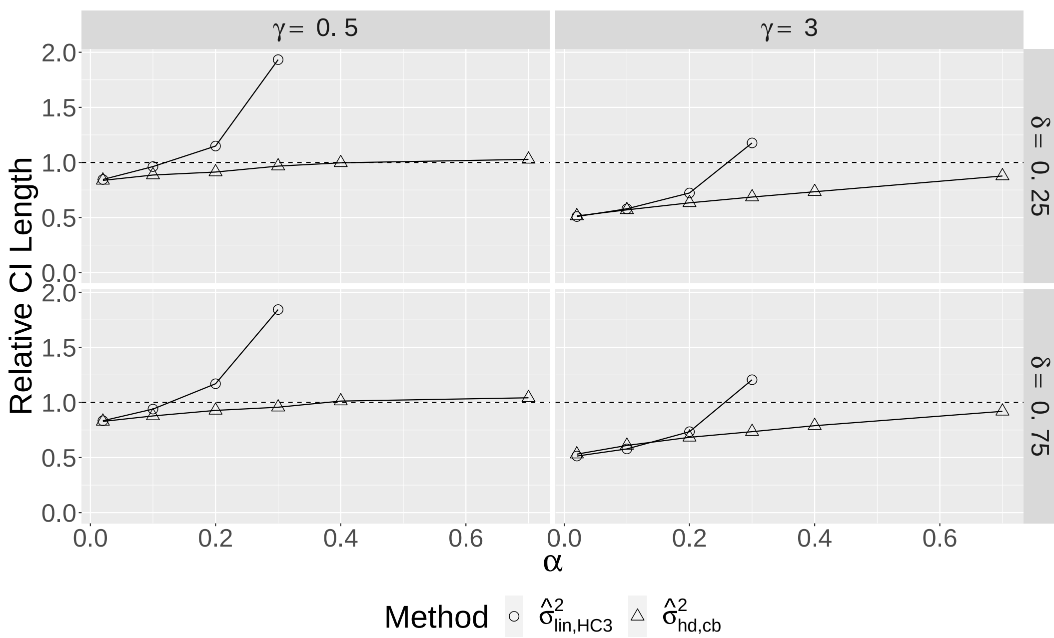
In practical applications, ignoring covariate dimension can result in catastrophic finite sample performance. Yet, at least under the context of finite-population inference, to the best of our knowledge, no theory explains this phenomenon. In this paper, we fill this gap by proposing a new debiased regression adjustment based average treatment effect estimator; and we study the conditions so that our estimator can have an advantage over the unadjusted estimator. In general, we require that the multiple correlation between covariates and potential outcomes increases with the covariate-dimension-to-sample-size ratio. Therefore, we recommend that practitioners use a moderate number of covariates that are predictive of the potential outcomes.

Our numerical analysis shows that compared to the other competitors, our estimator achieves the best performance in terms of estimation precision, bias reduction, inference reliability, and confidence interval length when the covariate-dimension-to-sample-size ratio is high; and an improved efficiency compared to the unadjusted estimator with a sufficiently large signal to noise ratio. It would be of interest to design new covariate adjustment based estimators that can bring improved accuracy even with low signal to noise ratio, which we leave for future work. Our additional numerical analysis in the Supplementary Material shows that our estimator is able to provide valid confidence interval even for heavy-tailed covariates, such as from Cauchy distribution.

Our theory builds upon a new central limit theorem of homogeneous sums [Koike, 2022]. It would also be interesting to use this new central limit theorem to study rerandomization [Morgan



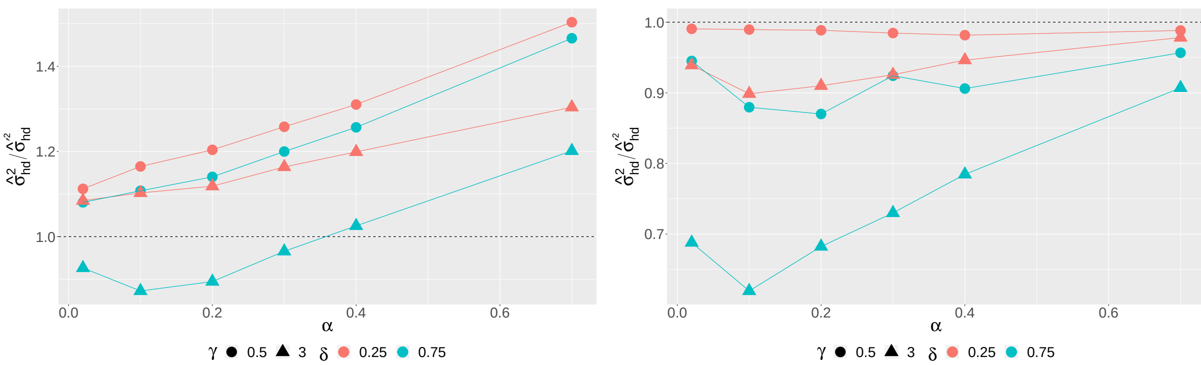
(a) the worst-case residual



(b) the independent t residual

Figure 6: Relative confidence interval length for different choices of γ, δ and α under the worst-case residual and independent t residual. The dashed lines signify 1.

and Rubin, 2012, Li and Ding, 2017] in the moderately high-dimensional regime. In this paper, we mainly focus on completely randomized experiments. It would be interesting to extend our theory to more complex experiments such as stratified experiments [Liu and Yang, 2020], and factorial



(a) the worst-case residual

(b) the independent t residual

Figure 7: Ratios of $\hat{\sigma}^2_{\text{hd}}$ to $\hat{\sigma}'^{2}_{\text{hd}}$ for different choices of γ, δ and α under the worst-case residual and independent t residual. The dashed lines signify 1.

experiments [Liu et al., 2022]. We study a high-dimensional extension of the OLS estimator and it would be interesting to consider the high-dimensional extension of the generalized linear estimator [Guo and Basse, 2023].

References

Z. D. Bai and Jack W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.*, 26(1):316–345, 1998.

Z. D. Bai, B. Q. Miao, and G. M. Pan. On asymptotics of eigenvectors of large sample covariance matrix. *The Annals of Probability*, 35(4):1532 – 1572, 2007. doi: 10.1214/009117906000001079. URL https://doi.org/10.1214/009117906000001079.

Zhigang Bao, Guangming Pan, and Wang Zhou. Universality for the largest eigenvalue of sample covariance matrices with general population. *The Annals of Statistics*, 43(1):382 – 421, 2015. doi: 10.1214/14-AOS1281. URL https://doi.org/10.1214/14-AOS1281.

Howard E Bell. Gershgorin's theorem and the zeros of polynomials. *The American Mathematical Monthly*, 72:292–295, 1965.

Andrew C Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the american mathematical society*, 49:122–136, 1941.

A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.*, 19(33):1–53, 2014.

Adam Bloniarz, Hanzhong Liu, Cun-Hui Zhang, Jasjeet S Sekhon, and Bin Yu. Lasso adjustments of treatment effect estimates in randomized experiments. *Proceedings of the National Academy of Sciences*, 113:7383–7390, 2016.

William G Cochran. *Sampling techniques*. John Wiley & Sons, 1977.

Peter de Jong. A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields*, 75:261–277, 1987.

Peter De Jong. A central limit theorem for generalized multilinear forms. *Journal of Multivariate Analysis*, 34:275–289, 1990.

T De Wet and JH Venter. Asymptotic distributions for quadratic forms with applications to tests of fit. *The Annals of Statistics*, 1:380–387, 1973.

Xiucai Ding and Fan Yang. A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *Ann. Appl. Probab.*, 28(3):1679–1738, 2018.

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

Robert Fox and Murad S Taqqu. Central limit theorems for quadratic forms in random variables having long-range dependence. *Probability Theory and Related Fields*, 74:213–240, 1987.

Kevin Guo and Guillaume Basse. The generalized oaxaca-blinder estimator. *Journal of the American Statistical Association*, 118:524–536, 2023.

Jaroslav Hájek. Limiting distributions in simple random sampling from a finite population. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:361–374, 1960.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

Antti Knowles and Jun Yin. Anisotropic local laws for random matrices. *Probability Theory and Related Fields*, pages 1–96, 2016.

Yuta Koike. High-dimensional central limit theorems for homogeneous sums. *Journal of Theoretical Probability*, pages 1–45, 2022.

Walter Kristof. A theorem on the trace of certain matrix products and some applications. *Journal of Mathematical Psychology*, 7:515–530, 1970.

Lihua Lei and Peng Ding. Regression adjustment in completely randomized experiments with a diverging number of covariates. *Biometrika*, 108:815–828, 2021.

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

Xinran Li and Peng Ding. Rerandomization and regression adjustment. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1):241–268, 2020.

Xinran Li, Peng Ding, and Donald B. Rubin. Asymptotic theory of rerandomization in treatment-control experiments. *Proceedings of the National Academy of Sciences*, 115:9157–62, 2018.

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

Hanzhong Liu and Yuehan Yang. Regression-adjusted average treatment effect estimates in stratified randomized experiments. *Biometrika*, 107:935–948, 2020.

Hanzhong Liu, Jiyang Ren, and Yuehan Yang. Randomization-based joint central limit theorem and efficient covariate adjustment in randomized block 2 k factorial experiments. *Journal of the American Statistical Association*, pages 1–15, 2022.

V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1:457, 1967.

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

Akanksha Negi and Jeffrey M Wooldridge. Revisiting regression adjustment in experiments with heterogeneous treatment effects. *Econometric Reviews*, 40:504–534, 2021.

Vladimir I Rotar et al. Limit theorems for polylinear forms. *Journal of Multivariate analysis*, 9: 511–530, 1979.

Vladimir Il'ich Rotar'. Limit theorems for multilinear forms and quasipolynomial functions. *Theory of Probability & Its Applications*, 20:512–532, 1976.

Anastasios A Tsiatis, Marie Davidian, Min Zhang, and Xiaomin Lu. Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Statistics in medicine*, 27:4658–4677, 2008.

Stefan Wager, Wenfei Du, Jonathan Taylor, and Robert J Tibshirani. High-dimensional regression adjustments in randomized experiments. *Proceedings of the National Academy of Sciences*, 113: 12673–12678, 2016.

Yuhao Wang and Xinran Li. Rerandomization with diminishing covariate imbalance and diverging number of covariates. *The Annals of Statistics*, 50(6):3439–3465, 2022.

Peter Whittle. Bounds for the moments of linear and quadratic forms in independent variables. *Theory of Probability & Its Applications*, 5:302–305, 1960.

Haokai Xi, Fan Yang, and Jun Yin. Convergence of eigenvector empirical spectral distribution of sample covariance matrices. *The Annals of Statistics*, 48(2):953 – 982, 2020. doi: 10.1214/19-AOS1832. URL https://doi.org/10.1214/19-AOS1832.

SUPPLEMENT TO "DEBIASED REGRESSION ADJUSTMENT IN COMPLETELY RANDOMIZED EXPERIMENTS WITH MODERATELY HIGH-DIMENSIONAL COVARIATES"

Appendix A provides some useful lemmas. It includes the technical details for the comments of (11) and (12).

Appendix B studies the decomposition of $\hat{\tau}_{\text{adj}}$ and $\hat{\tau}_{\text{db}}$. It includes the technical details for the comment of (4).

Appendix C studies an extension of the Hàjek's coupling.

Appendix D studies the asymptotic normality of $\hat{\tau}_{\text{db}}$ in the regime $p = o(n)$. It includes the proof of Theorem 1.

Appendix E studies the asymptotic normality of $\hat{\tau}_{\text{db}}$ in the moderately high-dimensional regime. It includes the proof of Theorem 2.

Appendix F studies the validity of the proposed inference procedure. It includes the proof of Theorem 3, Corollary 2, and the technical details for the comment of (10) and (13).

Appendix G studies the justification of assumptions. It includes the proof of Propositions 1–3 and Corollary 1.

Notations and definitions. Define $[n] := \{1, \dots, n\}$. We use $\sum_{[i_1 \dots i_k]}$ to denote summation over all $(i_1, \dots, i_k)$ with mutually distinct elements in $[n]$. So we may use $\sum_{[i,j]}$ and $\sum_{i \neq j}$ interchangeably. For any matrix $\boldsymbol{A}$, let A_{ij} be its (i,j)th element. We use $\tilde{\boldsymbol{A}}$ to denote a centered matrix with

$$\tilde{A}_{ij} := \begin{cases} A_{ij} - \frac{\sum_{k \neq l} A_{kl}}{n(n-1)}, & i \neq j; \\ A_{ii} - \frac{\sum_k A_{kk}}{n}, & i = j. \end{cases}$$

Let $\|\boldsymbol{A}\|_2$, $\text{tr}(\boldsymbol{A})$ be the l_2 norm and trace of matrix $\boldsymbol{A}$, respectively. For any random variable U, we define $\tilde{U} := U - \mathbb{E}U$ as a centered random variable. Let $\boldsymbol{Z} := (Z_1, \dots, Z_n)$. and $\tilde{\boldsymbol{Z}} := (\tilde{Z}_1, \dots, \tilde{Z}_n)$. Let $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$, otherwise. For a vector $\boldsymbol{y}$, let $\text{diag}(\boldsymbol{y})$ be the diagonal square matrix having $\boldsymbol{y}$ as its diagonal elements.

$\boldsymbol{Z} = (Z_1, \dots, Z_n) \in \{0,1\}^n$ is the indicator of a completely randomized experiment with $\sum_i Z_i = n_1$ and $n_1/n = r_1$. Let $\boldsymbol{T} = (T_1, \dots, T_n) \in \{0,1\}^n$ be the indicator of Bernoulli random sampling with each element i.i.d. generated from Bernoulli random variable with probability r_1. Moreover, we construct $\boldsymbol{T}$ so that the joint distribution of $\boldsymbol{T}$ and $\boldsymbol{Z}$ follows the so-called "Hàjek's Coupling" which will be discussed further in Appendix C.

A Some useful lemmas

We start by stating several useful lemmas and then proceed to prove the main results. Lemma A.1 shows the technical details of the comments of (11) and (12).

Lemma A.1. *For any symmetric matrix $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$, any constant c, d, and any population $\{\boldsymbol{a}_i\}_{i=1}^n$, $\{\boldsymbol{b}\}_{i=1}^n$, we have*

$$S_{\boldsymbol{A}+\boldsymbol{B},\boldsymbol{a},\boldsymbol{b}} = S_{\boldsymbol{A},\boldsymbol{a},\boldsymbol{b}} + S_{\boldsymbol{B},\boldsymbol{a},\boldsymbol{b}}, \quad S_{\boldsymbol{A},c\boldsymbol{a},d\boldsymbol{b}} = cdS_{\boldsymbol{A},\boldsymbol{a},\boldsymbol{b}}$$
$$S^2_{\boldsymbol{A},\boldsymbol{a}+\boldsymbol{b}} = S^2_{\boldsymbol{A},\boldsymbol{a}} + S^2_{\boldsymbol{A},\boldsymbol{b}} + 2S_{\boldsymbol{A},\boldsymbol{a},\boldsymbol{b}}.$$

Proof of Lemma A.1. By definition, we have

$$\begin{aligned}S_{\boldsymbol{A}+\boldsymbol{B},\boldsymbol{a},\boldsymbol{b}} =& \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}(A_{ij}+B_{ij})(\boldsymbol{a}_i-\bar{\boldsymbol{a}})(\boldsymbol{b}_j-\bar{\boldsymbol{b}})^\top \\ =& \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}(\boldsymbol{a}_i-\bar{\boldsymbol{a}})(\boldsymbol{b}_j-\bar{\boldsymbol{b}})^\top + \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}B_{ij}(\boldsymbol{a}_i-\bar{\boldsymbol{a}})(\boldsymbol{b}_j-\bar{\boldsymbol{b}})^\top \\ =& S_{\boldsymbol{A},\boldsymbol{a},\boldsymbol{b}} + S_{\boldsymbol{B},\boldsymbol{a},\boldsymbol{b}},\end{aligned}$$

and

$$S_{\boldsymbol{A},c\boldsymbol{a},d\boldsymbol{b}} = \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}cd(\boldsymbol{a}_i-\bar{\boldsymbol{a}})(\boldsymbol{b}_j-\bar{\boldsymbol{b}})^\top = cdS_{\boldsymbol{A},\boldsymbol{a},\boldsymbol{b}}.$$

Using that $A_{ij} = A_{ji}$, we have

$$\begin{aligned}S^2_{\boldsymbol{A},\boldsymbol{a}+\boldsymbol{b}} &= \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}(\boldsymbol{a}_i+\boldsymbol{b}_i-\bar{\boldsymbol{a}}-\bar{\boldsymbol{b}})(\boldsymbol{a}_j+\boldsymbol{b}_j-\bar{\boldsymbol{a}}-\bar{\boldsymbol{b}})^\top \\ &= \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}(\boldsymbol{a}_i-\bar{\boldsymbol{a}})(\boldsymbol{a}_j-\bar{\boldsymbol{a}})^\top + \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}(\boldsymbol{b}_i-\bar{\boldsymbol{b}})(\boldsymbol{b}_j-\bar{\boldsymbol{b}})^\top \\ &\quad + \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}(\boldsymbol{a}_i-\bar{\boldsymbol{a}})(\boldsymbol{b}_j-\bar{\boldsymbol{b}})^\top + \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}(\boldsymbol{b}_i-\bar{\boldsymbol{b}})(\boldsymbol{a}_j-\bar{\boldsymbol{a}})^\top \\ &= \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}(\boldsymbol{a}_i-\bar{\boldsymbol{a}})(\boldsymbol{a}_j-\bar{\boldsymbol{a}})^\top + \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}(\boldsymbol{b}_i-\bar{\boldsymbol{b}})(\boldsymbol{b}_j-\bar{\boldsymbol{b}})^\top \\ &\quad + \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}(\boldsymbol{a}_i-\bar{\boldsymbol{a}})(\boldsymbol{b}_j-\bar{\boldsymbol{b}})^\top \\ &= S^2_{\boldsymbol{A},\boldsymbol{a}} + S^2_{\boldsymbol{A},\boldsymbol{b}} + 2S_{\boldsymbol{A},\boldsymbol{a},\boldsymbol{b}}.\end{aligned}$$

□

Lemma A.2. *If* $(a_1-\bar{a},\ldots,a_n-\bar{a})^\top = \boldsymbol{A}(b_1-\bar{b},\ldots,b_n-\bar{b})^\top$, *we have*

$$S^2_a = S^2_{\boldsymbol{A}^\top\boldsymbol{A},b}.$$

Proof of Lemma A.2. Let $\boldsymbol{a} := (a_1-\bar{a},\ldots,a_n-\bar{a})^\top$ and $\boldsymbol{b} := (b_1-\bar{b},\ldots,b_n-\bar{b})^\top$. Using $\boldsymbol{a} = \boldsymbol{A}\boldsymbol{b}$, we get that

$$S^2_a = \frac{1}{n-1}\sum_{i=1}^{n}(a_i-\bar{a})^2 = \frac{1}{n-1}\boldsymbol{a}\boldsymbol{a}^\top = \frac{1}{n-1}\boldsymbol{b}\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{b}^\top = S^2_{\boldsymbol{A}^\top\boldsymbol{A},b}$$

□

Lemma A.3. *For any populations* $\{a_i\}_{i=1}^n$ *and* $\{b_i\}_{i=1}^n$, *we have that*

$$(r_1r_0)S^2_{r_1^{-1}a+r_0^{-1}b} = r_1^{-1}S^2_a + r_0^{-1}S^2_b - S^2_{a-b}.$$

Proof of Lemma A.3. Using Lemma A.1, we obtain that

$$\begin{aligned}(r_1 r_0) S^2_{r_1^{-1}a + r_0^{-1}b} &= \frac{r_0}{r_1} S^2_a + \frac{r_1}{r_0} S^2_b + 2S^2_{a,b} \\ &= (\frac{1}{r_1} - 1) S^2_a + (\frac{1}{r_0} - 1) S^2_b + 2S^2_{a,b} \\ &= \frac{1}{r_1} S^2_a + \frac{1}{r_0} S^2_b - S^2_a - S^2_b + 2S^2_{a,b} \\ &= \frac{1}{r_1} S^2_a + \frac{1}{r_0} S^2_b - S^2_{a-b}.\end{aligned}$$

□

Lemma A.4. *Consider any finite population $\{y_i\}_{i\in[n]}$, the variance of its sample total is*

$$\mathrm{var}\left(\sum_{i:Z_i=1} y_i\right) = \frac{n_1 n_0}{n(n-1)} \sum_i (y_i - \bar{y})^2.$$

Proof of Lemma A.4. See Theorem 2.2 of Cochran [1977]. □

Lemma A.5. *If $\sum_i (y_i - \bar{y})^2 = O(n)$ and r_1 tends to a limit in $(0, 1)$, then we have*

$$\sum_{i:Z_i=1} (y_i - \bar{y})/n_1 = O_{\mathbb{P}}(n^{-1/2}).$$

Proof of Lemma A.5. It follows from Lemma A.4 and Chebyshev's inequality. □

Lemma A.6. *Let $\boldsymbol{A}_l$, $l = 1, \ldots, q$, be $n \times n$ deterministic matrices with $n \geqslant 2$. Let $\alpha^2_{l1}, \ldots, \alpha^2_{ln}$ be the eigenvalues of $\boldsymbol{A}_l \boldsymbol{A}_l^\top$ in descending order with $\alpha_{li} \geq 0$ for $l \in [q]$ and $i \in [n]$. Then, we have that*

$$-\sum_{i=1}^{n} \alpha_{1i} \cdots \alpha_{qi} \leqslant \mathrm{tr}\,(\boldsymbol{A}_1 \cdots \boldsymbol{A}_q) \leqslant \sum_{i=1}^{n} \alpha_{1i} \cdots \alpha_{qi}.$$

Proof of Lemma A.6. This Lemma follows directly from Theorem (second version) of Kristof [1970] with $\Gamma_l = \boldsymbol{I}$, $l = 1, \ldots, q$. □

We will use Lemma A.6 repeatedly. For example, to bound the following quadratic form of y_i's, $\sum_{[i,j]} H^4_{ij} y^2_i y^2_j$. Let $\boldsymbol{y} = (y_1, \ldots, y_n)$. We rewrite the quadratic form as the trace of the product of several matrices

$$\sum_{[i_1,i_2]} H^4_{i_1 i_2} y^2_{i_1} y^2_{i_2} = \mathrm{tr}\left(\mathrm{diag}(\boldsymbol{y})^2\, \mathrm{diag}^-(\boldsymbol{Q})\, \mathrm{diag}(\boldsymbol{y})^2\, \mathrm{diag}^-(\boldsymbol{Q})\right).$$

We can apply Lemma A.6 with A_1, A_2, A_3, A_4 being $\mathrm{diag}(\boldsymbol{y})^2, \mathrm{diag}^-(\boldsymbol{Q}), \mathrm{diag}(\boldsymbol{y})^2, \mathrm{diag}^-(\boldsymbol{Q})$, respectively. Let $|y_{(1)}| \geq \ldots \geq |y_{(n)}|$ be the ordered sequence of $\{|y_i|\}_{i=1}^n$. Note that, for $i = 1, \ldots, n$,

$$\alpha_{1i} = \alpha_{3i} = |y_{(i)}|, \quad \alpha_{2i} = \alpha_{4i} < \|\,\mathrm{diag}^-(\boldsymbol{Q})\|_2.$$

Therefore, there is

$$\left|\sum_{[i_1,i_2]} H^4_{i_1 i_2} y^2_{i_1} y^2_{i_2}\right| < \|\,\mathrm{diag}^-(\boldsymbol{Q})\|_2^2 \sum_i y^2_{(i)} = \|\,\mathrm{diag}^-(\boldsymbol{Q})\|_2^2 \sum_i y^2_i.$$

Lemma A.7. *We have that*

$$\|\operatorname{diag}^{-}\{\boldsymbol{Q}\}\|_2 \leq 1, \quad \|\operatorname{diag}^{-}\{\boldsymbol{H}\}\|_2 \leq 2$$

Proof of lemma A.7. Using the Gershgorin circle theorem (see Theorem 0 of Bell [1965]), we get that

$$\|\operatorname{diag}^{-}\{\boldsymbol{Q}\}\|_2 < \max_i \sum_{j\in[n]\setminus i} H_{ij}^2 = \max_i (H_{ii} - H_{ii}^2) < 1.$$

On the other hand, by the triangle inequality, we have

$$\|\operatorname{diag}^{-}\{\boldsymbol{H}\}\|_2 \leq \|\boldsymbol{H}\|_2 + \|\operatorname{diag}\{\boldsymbol{H}\}\|_2 \leq 2.$$

□

Recall the definition of $\boldsymbol{A}(z)$ in the main text.

Lemma A.8. *Fix $z \in \{0, 1\}$. Let $d_i(z)$ be the i-th entry of $\boldsymbol{H}(Y_1(z) - \bar{Y}(z), \ldots, Y_n(z) - \bar{Y}(z))^\top$. We have that*

$$\sum_{i\in[n]\setminus\{j\}} \tilde{A}_{ij}(z) = -s_j(z), \quad \sum_{j\in[n]\setminus\{i\}} \tilde{A}_{ij}(z) = d_i(z) - s_i(z), \quad \tilde{A}_{ii}(z) = s_i(z).$$

Proof of Lemma A.8. By the fact $\sum_{i,j} A_{ij}(z) = 0$, we see that

$$\tilde{\boldsymbol{A}}(z) = \boldsymbol{A}(z) - \frac{\sum_i A_{ii}(z)}{n-1}\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^\top\right). \tag{16}$$

Let $\boldsymbol{Y}(z) := (Y_1(z), \ldots, Y_n(z))$ and $\boldsymbol{d}(z) := (d_1(z), \ldots, d_n(z))$. Therefore, we have

$$\tilde{\boldsymbol{A}}(z)\boldsymbol{1} = \boldsymbol{A}(z)\boldsymbol{1} = \boldsymbol{H}\operatorname{diag}(\boldsymbol{Y}(z) - \bar{Y}(z)\boldsymbol{1})\boldsymbol{1} = \boldsymbol{H}(\boldsymbol{Y}(z) - \bar{Y}(z)\boldsymbol{1}) = \boldsymbol{d}(z),$$

and

$$\tilde{\boldsymbol{A}}(z)^\top\boldsymbol{1} = \boldsymbol{A}(z)^\top\boldsymbol{1} = \operatorname{diag}(\boldsymbol{Y}(z) - \bar{Y}(z)\boldsymbol{1})\boldsymbol{H}\boldsymbol{1} = 0,$$

which implies that

$$\sum_j \tilde{A}_{ij}(z) = d_i(z), \quad \sum_i \tilde{A}_{ij}(z) = 0. \tag{17}$$

By (16), we have

$$\begin{aligned}\tilde{A}_{ii}(z) &= A_{ii}(z) - \frac{\sum_i A_{ii}(z)}{n} \\ &= H_{ii}(Y_i(z) - \bar{Y}(z)) - \frac{1}{n}\sum_{i=1}^n H_{ii}(Y_i(z) - \bar{Y}(z)) = s_i(z).\end{aligned} \tag{18}$$

Combining (17) and (18), we get that

$$\sum_{i\in[n]\setminus\{j\}} \tilde{A}_{ij}(z) = -s_j(z), \quad \sum_{j\in[n]\setminus\{i\}} \tilde{A}_{ij}(z) = d_i(z) - s_i(z).$$

This concludes the proof. □

B Decompostion of $\hat{\tau}_{\mathbf{adj}}$

In this section, we derive the decompositions of $\hat{\tau}_{\text{adj}}$ and $\hat{\tau}_{\text{db}}$, which correspond to Propositions B.1 and B.3. Before proving these results, we first state some useful lemmas.

Lemma B.1. *Fix $z \in \{0,1\}$. Under Assumption 2, we have*

$$\frac{\sum_i A_{ii}(z)}{n} = \frac{\sum_i H_{ii}\left(Y_i(z) - \bar{Y}(z)\right)}{n} = O(1).$$

Proof of Lemma B.1. By Cauchy-Schwartz inequality, we have

$$\frac{\sum_i H_{ii}\left(Y_i(z) - \bar{Y}(z)\right)}{n} \leq \left(\frac{\sum_i H_{ii}^2}{n}\right)^{1/2} \left(\frac{\sum_i \left(Y_i(z) - \bar{Y}(z)\right)^2}{n}\right)^{1/2}.$$

Recall that for $i \in [n]$, $H_{ii} \leq 1$, we have $\sum_i H_{ii}^2 \leq \sum_i H_{ii} = p$. Therefore,

$$\frac{\sum_i H_{ii}\left(Y_i(z) - \bar{Y}(z)\right)}{n} \leq \left(\frac{p}{n}\right)^{1/2} \left(\frac{\sum_i \left(Y_i(z) - \bar{Y}(z)\right)^2}{n}\right)^{1/2} = O(1),$$

where in the last step we applied Assumption 2. □

Lemma B.2. *Fix $z \in \{0,1\}$. Under Assumptions 2 and 3, we have that*

$$\boldsymbol{Z}^\top \tilde{\boldsymbol{A}}(z) \boldsymbol{Z} = o_{\mathbb{P}}(n).$$

Moreover, we have

$$\boldsymbol{Z}^\top \tilde{\boldsymbol{H}} \boldsymbol{Z} = o_{\mathbb{P}}(n).$$

Proof of Lemma B.2. Since $\sum_{i,j} A_{ij}(z) = 0$ and $\sum_{i,j} H_{ij} = 0$, we have

$$\sum_{i \neq j} A_{ij}(z) = -\sum_i A_{ii}(z), \quad \sum_{i \neq j} H_{ij} = -\sum_i H_{ii},$$

which gives that

$$\tilde{\boldsymbol{A}}(z) = \boldsymbol{A}(z) - \frac{\sum_i A_{ii}(z)}{n-1}\left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top\right), \quad \tilde{\boldsymbol{H}} = \boldsymbol{H} - \frac{\sum_i H_{ii}}{n-1}\left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top\right).$$

In light of these equations, we now analyze $\boldsymbol{Z}^\top \boldsymbol{A}(z) \boldsymbol{Z}, \boldsymbol{Z}^\top \boldsymbol{H} \boldsymbol{Z}$ and $\boldsymbol{Z}^\top \left(\boldsymbol{I} - \mathbf{1}\mathbf{1}^\top/n\right) \boldsymbol{Z}$ one by one.

We first consider $\boldsymbol{Z}^\top \boldsymbol{A}(z) \boldsymbol{Z}$. Observe that

$$\boldsymbol{Z}^\top \boldsymbol{A}(z) \boldsymbol{Z} = \sum_{[i,j]} Z_i Z_j H_{ij}(Y_j(z) - \bar{Y}(z)) + \sum_i Z_i H_{ii}(Y_i(z) - \bar{Y}(z)).$$

Applying Lemmas F.2 and F.3 with $y_i = 1$, $g_i = Y_i(z) - \bar{Y}(z)$, $D_{ij} = H_{ij}$, and $a_i = H_{ii}$, we get

$$\sum_{[i,j]} Z_i Z_j H_{ij}(Y_j(z) - \bar{Y}(z)) = r_1^2 \sum_{[i,j]} H_{ij}(Y_j(z) - \bar{Y}(z)) + o_{\mathbb{P}}(n),$$

and

$$\sum_i Z_i H_{ii}(Y_i(z) - \bar{Y}(z)) = r_1 \sum_i H_{ii}(Y_i(z) - \bar{Y}(z)) + o_{\mathbb{P}}(n).$$

Therefore, we have

$$\boldsymbol{Z}^\top \boldsymbol{A}(z)\boldsymbol{Z} = \sum_{[i,j]} r_1^2 H_{ij}(Y_j(z) - \bar{Y}(z)) + \sum_i r_1 H_{ii}(Y_i(z) - \bar{Y}(z)) + o_{\mathbb{P}}(n).$$

Applying similar analysis to $\boldsymbol{Z}^\top \boldsymbol{H}\boldsymbol{Z}$ and $\boldsymbol{Z}^\top(\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^\top/n)\boldsymbol{Z}$, we get

$$\begin{aligned}
\boldsymbol{Z}^\top \boldsymbol{H}\boldsymbol{Z} &= \sum_{[i,j]} r_1^2 H_{ij} + \sum_i r_1 H_{ii} + o_{\mathbb{P}}(n), \\
\boldsymbol{Z}^\top(\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^\top/n)\boldsymbol{Z} &= -\sum_{[i,j]} r_1^2 \frac{1}{n} + \sum_i r_1 \frac{n-1}{n} + o_{\mathbb{P}}(n).
\end{aligned}$$

Putting together, we have

$$\begin{aligned}
&\boldsymbol{Z}^\top \tilde{\boldsymbol{H}}\boldsymbol{Z} = \boldsymbol{Z}^\top \left\{ \boldsymbol{H} - \frac{\sum_i H_{ii}}{n-1}(\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^\top/n) \right\} \boldsymbol{Z} \\
&= \sum_{[i,j]} r_1^2 \left(H_{ij} + \frac{\sum_i H_{ii}}{n(n-1)} \right) + \sum_i r_1 \left(H_{ii} - \frac{\sum_i H_{ii}}{n} \right) + o_{\mathbb{P}}(n) + o_{\mathbb{P}}(n)\frac{\sum_i H_{ii}}{n-1} \\
&= 0 + 0 + o_{\mathbb{P}}(n) + o_{\mathbb{P}}(n) \cdot p/(n-1) = o_{\mathbb{P}}(n),
\end{aligned}$$

where the last equality again uses $\sum_{i,j} H_{ij} = 0$. Similarly, we have

$$\begin{aligned}
\boldsymbol{Z}^\top \tilde{\boldsymbol{A}}(z)\boldsymbol{Z} &= \boldsymbol{Z}^\top \left\{ \boldsymbol{A}(z) - \frac{\sum_i A_{ii}(z)}{n-1}(\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^\top/n) \right\} \boldsymbol{Z} \\
&= \sum_{[i,j]} r_1^2 \left(A_{ij}(z) + \frac{\sum_i A_{ii}(z)}{n(n-1)} \right) + \sum_i r_1 \left(A_{ii}(z) - \frac{\sum_i A_{ii}(z)}{n} \right) \\
&\quad + o_{\mathbb{P}}(n) + o_{\mathbb{P}}(n)\frac{\sum_i A_{ii}(z)}{n-1}.
\end{aligned}$$

Apparently, the second term on the right-hand side of the above decomposition is equal to zero. For the first term, using that

$$\sum_i \sum_j A_{ij}(z) = \sum_j \sum_i H_{ij}(Y_j(z) - \bar{Y}(z)) = 0,$$

we see that the first term is equal to zero as well. For the last term, applying Lemma B.1 yields $\frac{\sum_i A_{ii}(z)}{n-1} = O(1)$. Putting together, we have $\boldsymbol{Z}^\top \tilde{\boldsymbol{A}}(z)\boldsymbol{Z} = o_{\mathbb{P}}(n)$, which concludes the proof. □

The following proposition gives the detailed formulation and proof of (4).

Proposition B.1. *If Assumptions 1–3 hold, then we have*

$$
\begin{aligned}
&\hat{\tau}_{adj} - \bar{\tau} + \frac{r_1 r_0}{n} \sum_{i=1}^{n} H_{ii} \left(\frac{Y_i(1) - \bar{Y}(1)}{r_1^2} - \frac{Y_i(0) - \bar{Y}(0)}{r_0^2} \right) \\
&= \frac{1}{n} \sum_i (Z_i - r_1) c_i - \frac{1}{n} \sum_{[i,j]} (Z_i - r_1)(Z_j - r_1) \left(\frac{A_{ij}(1)}{r_1^2} - \frac{A_{ij}(0)}{r_0^2} \right) + o_{\mathbb{P}}(n^{-1/2}),
\end{aligned}
$$

where

$$
c_i = \alpha r_0 \frac{Y_i(1) - \bar{Y}(1)}{r_1^2} + \alpha r_1 \frac{Y_i(0) - \bar{Y}(0)}{r_0^2} - (r_0 - r_1) \left(\frac{s_i(1)}{r_1^2} - \frac{s_i(0)}{r_0^2} \right) + \frac{e_i(1)}{r_1} + \frac{e_i(0)}{r_0}.
$$

proof of Proposition B.1. In the following proof, for ease of presentation, we write $\boldsymbol{X}_i := \boldsymbol{X}_i - \bar{\boldsymbol{X}}$. We observe that

$$
\hat{\tau}_{\text{adj}} = \frac{\sum_i Z_i Y_i(1)}{n_1} - \frac{\sum_i (1 - Z_i) Y_i(0)}{n_0} - (r_1 \hat{\boldsymbol{\beta}}_0 + r_0 \hat{\boldsymbol{\beta}}_1)^\top \left(\sum_i Z_i \boldsymbol{X}_i \frac{n}{n_1 n_0} \right).
$$

Expanding the third term in the expression, we get

$$
(r_1 \hat{\boldsymbol{\beta}}_0 + r_0 \hat{\boldsymbol{\beta}}_1)^\top \left(\sum_i Z_i \boldsymbol{X}_i \frac{n}{n_1 n_0} \right) = \left(\sum_i Z_i \boldsymbol{X}_i^\top \frac{n}{n_1 n_0} \right) \boldsymbol{S}_{\boldsymbol{X}}^{-2} \left(r_1 \boldsymbol{s}_{\boldsymbol{X},\boldsymbol{Y(0)}} + r_0 \boldsymbol{s}_{\boldsymbol{X},\boldsymbol{Y(1)}} \right).
$$

We define

$$
M_1 = r_0 \left(\sum_i Z_i \boldsymbol{X}_i^\top \frac{n}{n_1 n_0} \right) \boldsymbol{S}_{\boldsymbol{X}}^{-2} \boldsymbol{s}_{\boldsymbol{X},\boldsymbol{Y(1)}}, \quad M_2 = r_1 \left(\sum_i Z_i \boldsymbol{X}_i^\top \frac{n}{n_1 n_0} \right) \boldsymbol{S}_{\boldsymbol{X}}^{-2} \boldsymbol{s}_{\boldsymbol{X},\boldsymbol{Y(0)}}.
$$

We now analyze the two terms M_1 and M_2.

For M_1, we write it as

$$
\begin{aligned}
M_1 =& r_0 \left(\sum_i Z_i \boldsymbol{X}_i^\top \frac{n}{n_1 n_0} \right) \boldsymbol{S}_{\boldsymbol{X}}^{-2} \boldsymbol{s}_{\boldsymbol{X},\boldsymbol{Y(1)}} \\
=& r_0 \left(\sum_i Z_i \boldsymbol{X}_i^\top \frac{n}{n_1 n_0} \right) \boldsymbol{S}_{\boldsymbol{X}}^{-2} \left(\frac{1}{n_1 - 1} \sum_i Z_i \boldsymbol{X}_i (Y_i(1) - \bar{Y}_1) \right) \\
=& \frac{n-1}{(n_1 - 1) n_1} \boldsymbol{Z}^\top \boldsymbol{A}(1) \boldsymbol{Z} + \frac{n-1}{(n_1 - 1) n_1} \boldsymbol{Z}^\top \boldsymbol{H} \boldsymbol{Z} (\bar{Y}(1) - \bar{Y}_1) =: M_{11} + M_{12}.
\end{aligned}
$$

For M_{11}, by the definition of $\tilde{\boldsymbol{A}}(1)$ and the fact that

$$
\boldsymbol{Z}^\top (\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^\top / n) \boldsymbol{Z} = \frac{n_1 n_0}{n},
$$

we decompose it as

$$
\begin{aligned}
M_{11} =& \frac{n-1}{(n_1 - 1) n_1} \boldsymbol{Z}^\top \boldsymbol{A}(1) \boldsymbol{Z} = \frac{n-1}{(n_1 - 1) n_1} \left(\boldsymbol{Z}^\top \tilde{\boldsymbol{A}}(1) \boldsymbol{Z} + \frac{\sum_i H_{ii} \left(Y_i(1) - \bar{Y}(1) \right)}{n-1} \frac{n_1 n_0}{n} \right) \\
=:& M_{111} + M_{112}.
\end{aligned}
$$

For M_{111}, we further expand $\boldsymbol{Z}^\top \tilde{\boldsymbol{A}}(1)\boldsymbol{Z}$. We introduce $d_i(z)$ as the i-th entry of $\boldsymbol{H}(Y_1(z) - \bar{Y}(z), \dots, Y_n(z) - \bar{Y}(z))^\top$. Applying Lemma A.8 and using the fact $\sum_{[i,j]} \tilde{A}_{ij}(z) = 0$ repeatedly, we obtain that

$$\begin{aligned}
\boldsymbol{Z}^\top \tilde{\boldsymbol{A}}(1)\boldsymbol{Z} &= \sum_i Z_i s_i(1) + \sum_{[i,j]} Z_i Z_j \tilde{A}_{ij}(1) \\
&= \sum_i Z_i s_i(1) + \sum_{[i,j]} (Z_i - r_1)(Z_j - r_1)\tilde{A}_{ij}(1) + \sum_{[i,j]} (Z_i - r_1) r_1 \tilde{A}_{ij}(1) \\
&\quad + \sum_{[i,j]} (Z_j - r_1) r_1 \tilde{A}_{ij}(1) + \sum_{[i,j]} r_1^2 \tilde{A}_{ij}(1) \\
&= \sum_i (Z_i - r_1) s_i(1) + \sum_{[i,j]} (Z_i - r_1)(Z_j - r_1)\tilde{A}_{ij}(1) + \sum_i r_1 (Z_i - r_1)(d_i(1) - s_i(1)) \\
&\quad + \sum_j r_1 (Z_j - r_1)(-s_j(1)) \\
&= \sum_i (Z_i - r_1)(s_i(1) + r_1 d_i(1) - 2 r_1 s_i(1)) + \sum_{[i,j]} (Z_i - r_1)(Z_j - r_1)\tilde{A}_{ij}(1) \\
&= \sum_i (Z_i - r_1)(s_i(1) + r_1 d_i(1) - 2 r_1 s_i(1)) + \sum_{[i,j]} (Z_i - r_1)(Z_j - r_1) A_{ij}(1) + O_{\mathbb{P}}(1),
\end{aligned}$$

where in the last step, we used that

$$\sum_{[i,j]} (Z_i - r_1)(Z_j - r_1) = -\sum_i (Z_i - r_1)^2 = O(n),$$

and that by Lemma B.1,

$$A_{ij}(1) - \tilde{A}_{ij}(1) = -\frac{\sum_i H_{ii} \left(Y_i(1) - \bar{Y}(1)\right)}{(n-1)n} = O(n^{-1}), \quad \forall i \neq j.$$

Moreover, by Lemma B.2, $\boldsymbol{Z}^\top \tilde{\boldsymbol{A}}(1)\boldsymbol{Z} = o_{\mathbb{P}}(n)$. Thus, we obtain that

$$\begin{aligned}
M_{111} &= \left(\frac{n}{n_1^2} + O(n^{-2})\right) \boldsymbol{Z}^\top \tilde{\boldsymbol{A}}(1)\boldsymbol{Z} = \frac{n}{n_1^2} \boldsymbol{Z}^\top \tilde{\boldsymbol{A}}(1)\boldsymbol{Z} + o_{\mathbb{P}}(n^{-1}) \\
&= \frac{1}{n r_1^2} \sum_i (Z_i - r_1)(s_i(1) + r_1 d_i(1) - 2 r_1 s_i(1)) \\
&\quad + \frac{1}{n r_1^2} \sum_{[i,j]} (Z_i - r_1)(Z_j - r_1) A_{ij}(1) + O_{\mathbb{P}}(n^{-1}).
\end{aligned}$$

On the other hand, for M_{112}, we use Lemma B.1 to get that

$$\begin{aligned}
M_{112} &= \left(\frac{n_0}{n_1} + O(n^{-1})\right) \frac{\sum_i H_{ii} \left(Y_i(1) - \bar{Y}(1)\right)}{n} \\
&= \frac{r_0}{r_1} \frac{\sum_i H_{ii} \left(Y_i(1) - \bar{Y}(1)\right)}{n} + O(n^{-1}).
\end{aligned}$$

For M_{12}, we see that

$$\boldsymbol{Z}^\top \boldsymbol{H} \boldsymbol{Z} = \boldsymbol{Z}^\top \tilde{\boldsymbol{H}} \boldsymbol{Z} + \frac{\sum_i H_{ii}}{n-1} \frac{n_1 n_0}{n} = \boldsymbol{Z}^\top \tilde{\boldsymbol{H}} \boldsymbol{Z} + \frac{\alpha n_1 n_0}{n-1},$$

with which we can decompose M_{12} as

$$M_{12} = \frac{n-1}{(n_1-1)n_1} \boldsymbol{Z}^\top \tilde{\boldsymbol{H}} \boldsymbol{Z} (\bar{Y}(1) - \bar{Y}_1) + \frac{\alpha n_0}{n_1 - 1} (\bar{Y}(1) - \bar{Y}_1) =: M_{121} + M_{122}. \tag{19}$$

By Lemmas A.5 and B.2, we have that

$$M_{121} = O(n^{-1}) o_{\mathbb{P}}(n) O_{\mathbb{P}}(n^{-1/2}) = o_{\mathbb{P}}(n^{-1/2}).$$

For M_{122}, we can derive that

$$M_{122} = \left(\frac{\alpha r_0}{r_1} + O(n^{-1}) \right) (\bar{Y}(1) - \bar{Y}_1) = \frac{\alpha r_0}{r_1} (\bar{Y}(1) - \bar{Y}_1) + O_{\mathbb{P}}(n^{-3/2}).$$

Combining the above results, we obtain that

$$\begin{aligned} M_1 = {} & \frac{r_0}{r_1} \frac{\sum_i H_{ii} \left(Y_i(1) - \bar{Y}(1) \right)}{n} \\ & + \frac{1}{n r_1^2} \sum_i (Z_i - r_1) \left[s_i(1) + r_1 d_i(1) - 2 r_1 s_i(1) - \alpha r_0 (Y_i(1) - \bar{Y}(1)) \right] \\ & + \frac{1}{n r_1^2} \sum_{[i,j]} (Z_i - r_1)(Z_j - r_1) A_{ij}(1) + o_{\mathbb{P}}(n^{-1/2}). \end{aligned} \tag{20}$$

Now, notice that

$$\begin{aligned} M_1 &= r_0 \left(\sum_{i:Z_i=1} \boldsymbol{X}_i^\top / n_1 - \sum_{i:Z_i=0} \boldsymbol{X}_i^\top / n_0 \right) \boldsymbol{S}_{\boldsymbol{X}}^{-2} \boldsymbol{s}_{\boldsymbol{X}, \boldsymbol{Y}(\mathbf{1})}, \\ M_2 &= -r_1 \left(\sum_{i:Z_i=0} \boldsymbol{X}_i^\top / n_0 - \sum_{i:Z_i=1} \boldsymbol{X}_i^\top / n_1 \right) \boldsymbol{S}_{\boldsymbol{X}}^{-2} \boldsymbol{s}_{\boldsymbol{X}, \boldsymbol{Y}(\mathbf{0})}. \end{aligned}$$

So, similar arguments also apply to M_2. By symmetry, replacing Z_i with $1 - Z_i$, replacing the treatment-group-specific quantities with their control-group analogues in the formula of (20), and multiplying with a negative sign, we obtain that

$$\begin{aligned} M_2 = {} & -\frac{r_1}{r_0} \frac{\sum_i H_{ii} \left(Y_i(0) - \bar{Y}(0) \right)}{n} \\ & + \frac{1}{n r_0^2} \sum_i (Z_i - r_1) \left[s_i(0) + r_0 d_i(0) - 2 r_0 s_i(0) - \alpha r_1 (Y_i(0) - \bar{Y}(0)) \right] \\ & - \frac{1}{n r_0^2} \sum_{[i,j]} (Z_i - r_1)(Z_j - r_1) A_{ij}(0) + o_{\mathbb{P}}(n^{-1/2}). \end{aligned}$$

Finally, the conclusion follows immediately from the equation

$$\begin{aligned}\hat{\tau}_{\text{adj}} - \bar{\tau} &= \frac{\sum_i Z_i(Y_i(1) - \bar{Y}(1))}{nr_1} - \frac{\sum_i (1 - Z_i)(Y_i(0) - \bar{Y}(0))}{nr_0} - M_1 - M_2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(Z_i - r_1)\left(\frac{Y_i(1) - \bar{Y}(1)}{r_1} + \frac{Y_i(1) - \bar{Y}(0)}{r_0}\right) - M_1 - M_2,\end{aligned}$$

and that $e_i(z) = Y_i(z) - \bar{Y}(z) - d_i(z)$ for $z \in \{0, 1\}$. □

As a direct consequence of Proposition B.1, the bias term is

$$b = -\frac{r_1 r_0}{n}\sum_{i=1}^{n} H_{ii}\left(\frac{Y_i(1) - \bar{Y}(1)}{r_1^2} - \frac{Y_i(0) - \bar{Y}(0)}{r_0^2}\right).$$

Recall that we estimate the bias via (see also (5))

$$\hat{b} =: -r_1 r_0\left(\frac{1}{n_1}\sum_{i:Z_i=1} H_{ii}\frac{(Y_i - \bar{Y}_1)}{r_1^2} - \frac{1}{n_0}\sum_{i:Z_i=0} H_{ii}\frac{(Y_i - \bar{Y}_0)}{r_0^2}\right).$$

We apply the following proposition to characterize $\hat{b}$:

Proposition B.2. *If Assumptions 1–3 hold, then we have that*

$$\hat{b} = b - \frac{1}{n}\sum_i (Z_i - r_1)\left\{r_0\frac{s_i(1)}{r_1^2} - r_0\alpha\frac{Y_i(1) - \bar{Y}(1)}{r_1^2} + r_1\frac{s_i(0)}{r_0^2} - r_1\alpha\frac{Y_i(0) - \bar{Y}(0)}{r_0^2}\right\} + o_{\mathbb{P}}(n^{-1/2}).$$

Proof of Proposition B.2. We see that

$$\hat{b} = -r_1 r_0\left(\frac{1}{n_1}\sum_{i:Z_i=1}(H_{ii} - \alpha)\frac{Y_i - \bar{Y}_1}{r_1^2} - \frac{1}{n_0}\sum_{i:Z_i=0}(H_{ii} - \alpha)\frac{Y_i - \bar{Y}_0}{r_0^2}\right) =: M_1 + M_2,$$

where

$$\begin{aligned}M_1 &= -r_1 r_0\left(\frac{1}{n_1}\sum_{i:Z_i=1}(H_{ii} - \alpha)\frac{\bar{Y}(1) - \bar{Y}_1}{r_1^2} - \frac{1}{n_0}\sum_{i:Z_i=0}(H_{ii} - \alpha)\frac{\bar{Y}(0) - \bar{Y}_0}{r_0^2}\right), \\ M_2 &= -r_1 r_0\left(\frac{1}{n_1}\sum_{i:Z_i=1}(H_{ii} - \alpha)\frac{Y_i - \bar{Y}(1)}{r_1^2} - \frac{1}{n_0}\sum_{i:Z_i=0}(H_{ii} - \alpha)\frac{Y_i - \bar{Y}(0)}{r_0^2}\right).\end{aligned}$$

For any $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$ with empirical averages $\bar{a}$ and $\bar{b}$, there is

$$\begin{aligned}\sum_{i:Z_i=1} a_i/n_1 - \sum_{i:Z_i=0} b_i/n_0 =& \bar{a} - \bar{b} + \frac{1}{n}\sum_i Z_i\left(\frac{a_i - \bar{a}}{r_1} + \frac{b_i - \bar{b}}{r_0}\right) \\ =& \bar{a} - \bar{b} + \frac{1}{n}\sum_i (Z_i - r_1)\left(\frac{a_i - \bar{a}}{r_1} + \frac{b_i - \bar{b}}{r_0}\right).\end{aligned}$$

Applying the above equation with a_i and b_i replaced by $(H_{ii}-\alpha)\frac{Y_i(1)-\bar{Y}(1)}{r_1^2}$ and $(H_{ii}-\alpha)\frac{Y_i(0)-\bar{Y}(0)}{r_0^2}$, respectively, we obtain that

$$M_2 = b - \frac{1}{n}\sum_i (Z_i - r_1)\left\{r_0\frac{s_i(1)}{r_1^2} - r_0\alpha\frac{Y_i(1)-\bar{Y}(1)}{r_1^2} + r_1\frac{s_i(0)}{r_0^2} - r_1\alpha\frac{Y_i(0)-\bar{Y}(0)}{r_0^2}\right\}.$$

It suffice to show that $M_1 = o_{\mathbb{P}}(n^{-1/2})$. Applying Lemma A.5 with $y_i = H_{ii}$ or $Y_i(z)$, we get $\sum_{i:Z_i=z}(H_{ii}-\alpha)/n_z = O_{\mathbb{P}}(n^{-1/2})$ and $\bar{Y}_z - \bar{Y}(z) = O_{\mathbb{P}}(n^{-1/2})$, which implies that

$$M_1 = O_{\mathbb{P}}(n^{-1/2})O_{\mathbb{P}}(n^{-1/2}) = o_{\mathbb{P}}(n^{-1/2}).$$

This concludes the proof. □

Combining Propositions B.1 and B.2, it is straightforward to derive the following result.

Proposition B.3. *If Assumptions 1–3 hold, then we have that*

$$\hat{\tau}_{db} - \bar{\tau} = n^{-1}\sum_i (Z_i - r_1)\left\{\frac{e_i(1)}{r_1} + \frac{e_i(0)}{r_0} + \frac{s_i(1)}{r_1} + \frac{s_i(0)}{r_0}\right\} - n^{-1}\sum_{[i,j]}(Z_i - r_1)(Z_j - r_1)\left(\frac{A_{ij}(1)}{r_1^2} - \frac{A_{ij}(0)}{r_0^2}\right) + o_{\mathbb{P}}(n^{-1/2}).$$

C Hàjek‘s coupling

In this section, we study Hàjek's coupling for sampling without replacement. We prove the second-order Hàjek's coupling which is Proposition C.2. Then we use it to prove that $\hat{\tau}_{\mathrm{db}}$ is asymptotically equal to the summation of several homogeneous sums which is Proposition C.3.

For ease of presentation, we consider a finite population $\{y_i\}_{i\in[n]}$ with $\sum_{i=1}^n y_i = 0$. Let A_{ij} be the (i,j)-th element of $\boldsymbol{A} := \boldsymbol{H}\,\mathrm{diag}\,(y_1,\dots,y_n)$. Let d_i be the ith element of $\boldsymbol{H}(y_1,\dots,y_n)^\top$. We can see that

$$\sum_j \tilde{A}_{ij} = d_i; \quad \sum_i \tilde{A}_{ij} = 0.$$

Recall that $\boldsymbol{Z} = (Z_1,\dots,Z_n) \in \{0,1\}^n$ is the indicator of a completely randomized experiment with $\sum_i Z_i = n_1$ and $n_1/n = r_1$. Let $\boldsymbol{T} = (T_1,\dots,T_n) \in \{0,1\}^n$ be the indicator of Bernoulli random sampling with each element i.i.d. generated from Bernoulli random variable with probability r_1. Let $n_1' = \sum_i T_i$ and $\mathcal{T} = \{i : T_i = 1\}$. We assume the following coupling between $\boldsymbol{T}$ and $\boldsymbol{Z}$:

- If $n_1' = n_1$, $\boldsymbol{Z} = \boldsymbol{T}$,
- If $n_1' > n_1$, we select a random sample $\mathcal{D}$ of size $n_1' - n_1$ in $\mathcal{T}$ and define $Z_i = 0$ for $i \in \mathcal{D}$ and $Z_i = T_i$ for $i \in [n]\backslash\mathcal{D}$,
- If $n_1' < n_1$, we select a random sample $\mathcal{D}$ of size $n_1 - n_1'$ in $[n]\backslash\mathcal{T}$ and define $Z_i = 1$ for $i \in \mathcal{D}$ and $Z_i = T_i$ for $i \in [n]\backslash\mathcal{D}$.

Proposition C.1 (First-order Hàjek's coupling)**.** *If Assumption 1 holds and $\sum_i y_i^2 = O(n)$, then we have that*

$$n^{-1/2}\sum_i (Z_i - T_i)y_i = o_{\mathbb{P}}(1).$$

Proof of Proposition C.1. The proposition follows from Lemma A3 (iii) of Wang and Li [2022] with $u_i = y_i$. □

Proposition C.2 (Second-order Hàjek's coupling)**.** *Under Assumption 1 and* $\sum_i y_i^2 = O(n)$*, we have*

$$n^{-1/2} \sum_{[i,j]} (Z_i Z_j - T_i T_j) \tilde{A}_{ij} = o_{\mathbb{P}}(1).$$

Proof of Proposition C.2. Let $v = (v_1, \ldots, v_n)^\top$ be a uniform at random permutation of $\{1, \ldots, n\}$ and is independent from n_1'. Write $D := \sum_{[i_1,i_2]} \tilde{A}_{i_1 i_2}(T_{i_1} T_{i_2} - Z_{i_1} Z_{i_2})$; apparently $\mathbb{E}D = 0$. We now bound $\mathbb{E}[D^2]$. First, from the coupling between $\boldsymbol{T}$ and $\boldsymbol{Z}$, by conditioning on n_1', the random variable D is equal in distribution to

$$\sum_{i=n_1+1}^{n_1'} \sum_{j=1}^{n_1} \tilde{A}_{v_i v_j} + \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n_1'} \tilde{A}_{v_i v_j} + \sum_{i=n_1+1}^{n_1'} \sum_{j=n_1+1}^{n_1'} \tilde{A}_{v_i v_j}(1 - \delta_{ij})$$

if $n_1' > n_1$,

$$-\sum_{i=n_1'+1}^{n_1} \sum_{j=1}^{n_1'} \tilde{A}_{v_i v_j} - \sum_{i=1}^{n_1'} \sum_{j=n_1'+1}^{n_1} \tilde{A}_{v_i v_j} - \sum_{i=n_1'+1}^{n_1} \sum_{j=n_1'+1}^{n_1} \tilde{A}_{v_i v_j}(1 - \delta_{ij})$$

if $n_1' < n_1$, and 0 if $n_1' = n_1$.

We first consider D^2 conditioning on some $n_1' > n_1$. Under this event, we can write $D = \sum_{(i,j)\in\mathcal{S}} \tilde{A}_{v_i v_j}$, where

$$\mathcal{S} := \{(i,j) : \ n_1 + 1 \le i \le n_1', \ 1 \le j \le n_1\} \cup \{(i,j) : \ 1 \le i \le n_1, \ n_1 + 1 \le j \le n_1'\} \cup \\ \{(i,j) : n_1 + 1 \le i, j \le n_1', i \ne j\}.$$

Then, we have that

$$\begin{aligned} D^2 = &\sum_{i\ne j,(i,j)\in\mathcal{S}} \tilde{A}_{v_i v_j}^2 + \sum_{i\ne j,(i,j)\in\mathcal{S}} \tilde{A}_{v_i v_j} \tilde{A}_{v_j v_i} + \sum_{i\ne j\ne k,(i,j),(k,j)\in\mathcal{S}} \tilde{A}_{v_i v_j} \tilde{A}_{v_k v_j} + \\ &+ \sum_{i\ne j\ne k,(i,j),(j,k)\in\mathcal{S}} \tilde{A}_{v_i v_j} \tilde{A}_{v_j v_k} + \sum_{i\ne j\ne k,(i,j),(i,k)\in\mathcal{S}} \tilde{A}_{v_i v_j} \tilde{A}_{v_i v_k} + \sum_{i\ne j\ne k\ne l,(i,j),(k,l)\in\mathcal{S}} \tilde{A}_{v_i v_j} \tilde{A}_{v_k v_l}. \end{aligned}$$

For the first term, we have that for each index,

$$\mathbb{E}\tilde{A}_{v_i v_j}^2 = \frac{\sum_{[i_1,i_2]} \tilde{A}_{i_1 i_2}^2}{n(n-1)}.$$

Similarly, we have

$$\begin{aligned} &\mathbb{E}\tilde{A}_{v_i v_j}\tilde{A}_{v_j v_i} = \frac{\sum_{[i_1,i_2]} \tilde{A}_{i_1 i_2}\tilde{A}_{i_2 i_1}}{n(n-1)}, \quad \mathbb{E}\tilde{A}_{v_i v_j}\tilde{A}_{v_k v_j} = \frac{\sum_{[i_1\ldots i_3]} \tilde{A}_{i_1 i_2}\tilde{A}_{i_3 i_2}}{n(n-1)(n-2)}, \\ &\mathbb{E}\tilde{A}_{v_i v_j}\tilde{A}_{v_j v_k} = \frac{\sum_{[i_1\ldots i_3]} \tilde{A}_{i_3 i_2}\tilde{A}_{i_2 i_1}}{n(n-1)(n-2)}, \quad \mathbb{E}\tilde{A}_{v_i v_j}\tilde{A}_{v_i v_k} = \frac{\sum_{[i_1\ldots i_3]} \tilde{A}_{i_2 i_3}\tilde{A}_{i_2 i_1}}{n(n-1)(n-2)}, \\ &\mathbb{E}\tilde{A}_{v_i v_j}\tilde{A}_{v_k v_l} = \frac{\sum_{[i_1\ldots i_4]} \tilde{A}_{i_1 i_2}\tilde{A}_{i_3 i_4}}{n(n-1)(n-2)(n-3)}. \end{aligned}$$

To understand the order of the above terms, we introduce $M_1, \ldots, M_5$ as

$$M_1 := \sum_{i_1} \tilde{A}^2_{i_1 i_1}, \quad M_2 := \sum_{[i_1,i_2]} \tilde{A}^2_{i_1 i_2}, \quad M_3 := \Big| \sum_{[i_1,i_2]} \tilde{A}_{i_1 i_2} \tilde{A}_{i_2 i_1} \Big|,$$
$$M_4 := \sum_{i_1} d^2_{i_1}, \quad M_5 := \Big| \sum_{i_1} d_{i_1} \tilde{A}_{i_1 i_1} \Big|,$$

Now, by repeatedly applying $\sum_j \tilde{A}_{ij} = d_i$ and $\sum_i \tilde{A}_{ij} = 0$, we obtain that

$$\begin{aligned}
\sum_{[i_1 \ldots i_3]} \tilde{A}_{i_1 i_2} \tilde{A}_{i_1 i_3} &= \sum_{[i_1,i_2]} d_{i_1} \tilde{A}_{i_1 i_2} - \sum_{[i_1,i_2]} (\tilde{A}_{i_1 i_2} \tilde{A}_{i_1 i_1} + \tilde{A}_{i_1 i_2} \tilde{A}_{i_1 i_2}) \\
&= \sum_{[i_1,i_2]} d_{i_1} \tilde{A}_{i_1 i_2} - \sum_{[i_1,i_2]} \tilde{A}_{i_1 i_2} \tilde{A}_{i_1 i_1} - M_2 \\
&= \sum_{[i_1,i_2]} d_{i_1} \tilde{A}_{i_1 i_2} - \sum_{i_1} (d_{i_1} \tilde{A}_{i_1 i_1} - \tilde{A}_{i_1 i_1} \tilde{A}_{i_1 i_1}) - M_2 \\
&= \sum_{[i_1,i_2]} d_{i_1} \tilde{A}_{i_1 i_2} + O(M_1 + M_2 + M_5) \\
&= \sum_{i_1} (d^2_{i_1} - d_{i_1} \tilde{A}_{i_1 i_1}) + O(M_1 + M_2 + M_5) = O(M_1 + M_2 + M_4 + M_5); \\
\sum_{[i_1 \ldots i_3]} \tilde{A}_{i_1 i_2} \tilde{A}_{i_2 i_3} &= \sum_{[i_1,i_2]} \tilde{A}_{i_1 i_2} d_{i_2} - \sum_{[i_1,i_2]} (\tilde{A}_{i_1 i_2} \tilde{A}_{i_2 i_1} + \tilde{A}_{i_1 i_2} \tilde{A}_{i_2 i_2}) \\
&= \sum_{[i_1,i_2]} \tilde{A}_{i_1 i_2} d_{i_2} - M_3 + \sum_{i_2} \tilde{A}^2_{i_2 i_2} \\
&= \sum_{[i_1,i_2]} \tilde{A}_{i_1 i_2} d_{i_2} + O(M_1 + M_3) \\
&= -\sum_{i_2} \tilde{A}_{i_2 i_2} d_{i_2} + O(M_1 + M_3) = O(M_1 + M_3 + M_5); \\
\sum_{[i_1 \ldots i_3]} \tilde{A}_{i_1 i_2} \tilde{A}_{i_3 i_2} &= -\sum_{[i_1,i_2]} (\tilde{A}_{i_1 i_2} \tilde{A}_{i_1 i_2} + \tilde{A}_{i_1 i_2} \tilde{A}_{i_2 i_2}) = O(M_1 + M_2).
\end{aligned}$$

Applying $\sum_i d_i = 0$ and $\sum_i \tilde{A}_{ii} = 0$,, we obtain that

$$\begin{aligned}
&\sum_{[i_1 \ldots i_4]} \tilde{A}_{i_1 i_2} \tilde{A}_{i_3 i_4} = \sum_{[i_1 \ldots i_3]} \tilde{A}_{i_1 i_2} d_{i_3} - \sum_{[i_1 \ldots i_3]} (\tilde{A}_{i_1 i_2} \tilde{A}_{i_3 i_1} + \tilde{A}_{i_1 i_2} \tilde{A}_{i_3 i_2} + \tilde{A}_{i_1 i_2} \tilde{A}_{i_3 i_3}) \\
&= -\sum_{[i_1,i_2]} (\tilde{A}_{i_1 i_2} d_{i_1} + \tilde{A}_{i_1 i_2} d_{i_2}) - \sum_{[i_1 \ldots i_3]} (\tilde{A}_{i_1 i_2} \tilde{A}_{i_2 i_3} + \tilde{A}_{i_1 i_2} \tilde{A}_{i_3 i_2} + \tilde{A}_{i_1 i_2} \tilde{A}_{i_3 i_3}) \\
&= -\sum_{[i_1,i_2]} (\tilde{A}_{i_1 i_2} d_{i_1} + \tilde{A}_{i_1 i_2} d_{i_2}) - \sum_{[i_1 \ldots i_3]} \tilde{A}_{i_1 i_2} \tilde{A}_{i_3 i_3} - \sum_{[i_1 \ldots i_3]} (\tilde{A}_{i_1 i_2} \tilde{A}_{i_2 i_3} + \tilde{A}_{i_1 i_2} \tilde{A}_{i_3 i_2}) \\
&= -\sum_{[i_1,i_2]} (\tilde{A}_{i_1 i_2} d_{i_1} + \tilde{A}_{i_1 i_2} d_{i_2}) + \sum_{[i_1,i_2]} (\tilde{A}_{i_1 i_2} \tilde{A}_{i_1 i_1} + \tilde{A}_{i_1 i_2} \tilde{A}_{i_2 i_2}) \\
&\quad - \sum_{[i_1 \ldots i_3]} (\tilde{A}_{i_1 i_2} \tilde{A}_{i_2 i_3} + \tilde{A}_{i_1 i_2} \tilde{A}_{i_3 i_2}).
\end{aligned}$$

Then, using the derivations in the analysis of $\sum_{[i_1\ldots i_3]}\tilde{A}_{i_1i_2}\tilde{A}_{i_1i_3}$, $\sum_{[i_1\ldots i_3]}\tilde{A}_{i_1i_2}\tilde{A}_{i_2i_3}$ and $\sum_{[i_1\ldots i_3]}\tilde{A}_{i_1i_2}\tilde{A}_{i_3i_2}$, we get that

$$\sum_{[i_1\ldots i_4]}\tilde{A}_{i_1i_2}\tilde{A}_{i_3i_4}=O(M_1+\cdots+M_5).$$

On the other hand, writing $\Delta:=|n_1'-n_1|$, we have

$$|\mathcal{S}|\le 2n\Delta,\quad |\{(i,j,k):i\neq j\neq k,(i,j),(k,j)\in\mathcal{S}\}|\le 2n^2\Delta,$$
$$|\{(i,j,k,l):i\neq j\neq k\neq l,(i,j),(k,l)\in\mathcal{S}\}|\le 4n^2\Delta^2\le 4n^3\Delta.$$

Putting together, we obtain that when $n_1'>n_1$, there exists a universal constant $C>0$ which does not depend on Δ such that

$$\mathbb{E}[D^2\mid n_1']\le C\Delta n^{-1}\sum_{t=1}^{5}M_t.$$

With similar arguments, we can obtain the same bound for $\mathbb{E}[D^2\mid n_1']$ when $n_1'<n_1$. Finally, with the law of total expectation, we obtain that

$$\mathbb{E}[D^2]\le C\mathbb{E}[\Delta]n^{-1}(M_1+M_2+M_3+M_4+M_5).$$

Now, by Assumption 1, we can bound $\mathbb{E}\Delta$ as

$$\mathbb{E}\Delta\le(\mathbb{E}\Delta^2)^{1/2}=(nr_1r_0)^{1/2}=O(n^{1/2}).$$

Combining these results, we get

$$\mathbb{E}D^2=O\left(n^{-1/2}(M_1+M_2+M_3+M_4+M_5)\right).$$

It remains to bound M_i, $i=1,\ldots,5$. Since $H_{i_1i_1}\le 1$ and $H_{i_2i_2}=\sum_{i_1}H_{i_1i_2}^2$ (due to the fact $\boldsymbol{H}=\boldsymbol{H}^2$), we have that

$$M_1=\sum_{i_1}\tilde{A}_{i_1i_1}^2\le\sum_{i_1}H_{i_1i_1}^2y_{i_1}^2\le\sum_{i_1}y_{i_1}^2=O(n),$$
$$M_2=\sum_{[i_1,i_2]}\tilde{A}_{i_1i_2}^2\le\sum_{[i_1,i_2]}H_{i_1i_2}^2y_{i_2}^2\le\sum_{i_1}H_{i_1i_1}y_{i_1}^2\le\sum_{i_1}y_{i_1}^2=O(n).$$

By Cauchy-Schwarz inequality, we have that $M_3\le M_2=O(n)$. Finally, we have

$$M_4=\sum_i d_i^2\le\sum_i y_i^2=O(n),\quad M_5\le(M_4M_1)^{1/2}=O(n).$$

The above bounds give that $\mathbb{E}D^2=O(n^{1/2})$. Therefore, by Chebyshev's inequality, we have

$$\sum_{[i_1,i_2]}\tilde{A}_{i_1i_2}(T_{i_1}T_{i_2}-Z_{i_1}Z_{i_2})=O_p(n^{1/4})=o_{\mathbb{P}}(n^{1/2}).$$

The conclusion then follows. □

Equipped with Propositions C.1 and C.2, we can now approximate $\hat{\tau}_{\mathrm{db}}$ with a polynomial of T_i's.

Proposition C.3. *If Assumptions 1–3 hold, then we have that*

$$\hat{\tau}_{db} - \bar{\tau} = n^{-1}\sum_i (T_i - r_1)\left(\frac{e_i(1)}{r_1} + \frac{e_i(0)}{r_0} + \frac{s_i(1)}{r_1} + \frac{s_i(0)}{r_0}\right)$$
$$- n^{-1}\sum_{[i,j]} (T_i - r_1)(T_j - r_1)\left(\frac{A_{ij}(1)}{r_1^2} - \frac{A_{ij}(0)}{r_0^2}\right) + o_{\mathbb{P}}(n^{-1/2}).$$

Proof. For ease of presentation, we write $\tilde{Z}_i = Z_i - r_1$ and $\tilde{T}_i = T_i - r_1$. By proposition B.3, it remains to show that

$$M_1 - M_2 := n^{-1}\sum_i (\tilde{T}_i - \tilde{Z}_i)\left(\frac{e_i(1)}{r_1} + \frac{e_i(0)}{r_0} + \frac{s_i(1)}{r_1} + \frac{s_i(0)}{r_0}\right)$$
$$- n^{-1}\sum_{[i,j]} (\tilde{T}_i\tilde{T}_j - \tilde{Z}_i\tilde{Z}_j)\left(\frac{A_{ij}(1)}{r_1^2} - \frac{A_{ij}(0)}{r_0^2}\right) = o_{\mathbb{P}}(n^{-1/2}).$$

For the term M_2, using $\sum_{i,j} A_{ij}(z) = 0$, we obtain the decomposition

$$\frac{1}{n}\sum_{[i,j]} (\tilde{T}_i\tilde{T}_j - \tilde{Z}_i\tilde{Z}_j)\left(\frac{A_{ij}(1)}{r_1^2} - \frac{A_{ij}(0)}{r_0^2}\right)$$
$$= \frac{1}{n}\sum_{[i,j]} (\tilde{T}_i\tilde{T}_j - \tilde{Z}_i\tilde{Z}_j)\left(-\frac{\sum_i A_{ii}(1)}{r_1^2 n(n-1)} + \frac{\sum_i A_{ii}(0)}{r_0^2 n(n-1)}\right)$$
$$+ \frac{1}{n}\sum_{[i,j]} (\tilde{T}_i\tilde{T}_j - \tilde{Z}_i\tilde{Z}_j)\left(\frac{\tilde{A}_{ij}(1)}{r_1^2} - \frac{\tilde{A}_{ij}(0)}{r_0^2}\right) =: M_{21} + M_{22}.$$

For M_{21}, as shown in the proof of Proposition B.1, we have

$$\sum_{[i,j]} \tilde{Z}_i\tilde{Z}_j = O(n).$$

This, together with Lemma B.1, yields that

$$\sum_{[i,j]} \tilde{Z}_i\tilde{Z}_j \frac{\sum_i A_{ii}(z)}{n(n-1)} = O(1).$$

Moreover, we see that for $i \neq j \neq k \neq l$,

$$\mathbb{E}\tilde{T}_i^2\tilde{T}_j\tilde{T}_k = 0, \quad \mathbb{E}\tilde{T}_i\tilde{T}_j\tilde{T}_k\tilde{T}_l = 0,$$

which implies that

$$\mathbb{E}\left(\sum_{[i,j]} \tilde{T}_i\tilde{T}_j \frac{\sum_i A_{ii}(z)}{n(n-1)}\right)^2 = 2\sum_{[i,j]} \mathbb{E}(\tilde{T}_i\tilde{T}_j)^2\left(\frac{\sum_i A_{ii}(z)}{n(n-1)}\right)^2$$
$$= 2(r_1 r_0)^2 n(n-1)\left(\frac{\sum_i A_{ii}(z)}{n(n-1)}\right)^2 = O(1),$$

where the last inequality follows from Lemma B.1. Then, by Chebyshev's inequality, we have

$$\sum_{[i,j]} \tilde{T}_i \tilde{T}_j \frac{\sum_i A_{ii}(z)}{n(n-1)} = O_{\mathbb{P}}(1).$$

Combining these results, we obtain $M_{21} = O_{\mathbb{P}}(n^{-1})$.

We now focus on M_{22}. We first expand it as

$$\begin{aligned}
M_{22} &= \frac{1}{n} \sum_{[i,j]} (T_i T_j - Z_i Z_j - r_1 T_i + r_1 Z_i - r_1 T_j + r_1 Z_j) \left(\frac{\tilde{A}_{ij}(1)}{r_1^2} - \frac{\tilde{A}_{ij}(0)}{r_0^2} \right) \\
&= \frac{1}{n} \sum_{[i,j]} (T_i T_j - Z_i Z_j) \left(\frac{\tilde{A}_{ij}(1)}{r_1^2} - \frac{\tilde{A}_{ij}(0)}{r_0^2} \right) + \frac{1}{n} \sum_{[i,j]} r_1 (Z_i - T_i) \left(\frac{\tilde{A}_{ij}(1)}{r_1^2} - \frac{\tilde{A}_{ij}(0)}{r_0^2} \right) \\
&\quad + \frac{1}{n} \sum_{[i,j]} r_1 (Z_j - T_j) \left(\frac{\tilde{A}_{ij}(1)}{r_1^2} - \frac{\tilde{A}_{ij}(0)}{r_0^2} \right).
\end{aligned}$$

Then, by Lemma A.8, we see that

$$\begin{aligned}
M_{22} &= \frac{1}{n} \sum_{[i,j]} (T_i T_j - Z_i Z_j) \left(\frac{\tilde{A}_{ij}(1)}{r_1^2} - \frac{\tilde{A}_{ij}(0)}{r_0^2} \right) \\
&\quad + \frac{1}{n} \sum_{i} r_1 (Z_i - T_i) \left(\frac{d_i(1)}{r_1^2} - \frac{s_i(1)}{r_1^2} - \frac{d_i(0)}{r_0^2} + \frac{s_i(0)}{r_0^2} \right) \\
&\quad + \frac{1}{n} \sum_{i} r_1 (Z_i - T_i) \left(-\frac{s_i(1)}{r_1^2} + \frac{s_i(0)}{r_0^2} \right).
\end{aligned}$$

Applying Proposition C.2 with $y_i = Y_i(z) - \bar{Y}(z)$, we get

$$\sum_{[i,j]} (T_i T_j - Z_i Z_j) \left(\frac{\tilde{A}_{ij}(1)}{r_1^2} - \frac{\tilde{A}_{ij}(0)}{r_0^2} \right) = o_{\mathbb{P}}(n^{1/2}).$$

Then, applying Proposition C.1 with $y_i = s_i(z)$ and $d_i(z)$, we get

$$\sum_i (Z_i - T_i) s_i(z) = o_{\mathbb{P}}(n^{1/2}), \quad \sum_i (Z_i - T_i) d_i(z) = o_{\mathbb{P}}(n^{1/2}), \tag{21}$$

which implies that $M_{22} = o_{\mathbb{P}}(n^{-1/2})$. Together with the bound on M_{21}, it implies that

$$M_2 = o_{\mathbb{P}}(n^{-1/2}).$$

It remains to bound M_1. By Proposition C.1 with $y_i = e_i(z)$, we have

$$\sum_i (Z_i - T_i) e_i(z) = o_{\mathbb{P}}(n^{1/2}),$$

which, combined with (21), implies that $M_1 = o_{\mathbb{P}}(n^{-1/2})$. This concludes the proof. □

D Asymptotic normality of $\hat{\tau}_{\text{db}}$ when $p = o(n)$

In this section, we study the asymptotic normality of $\hat{\tau}_{\text{db}}$ when $p = o(n)$ and give the proof of Theorem 1. We first state some lemmas that will be used in the proof.

The following Lemma demonstrates that for a sequence of population $\{y_i\}_{i\in[n]}$ satisfying a bounded second moment and a Lindeberg–Feller–type condition, the linear type statistics is asymptotically normal.

Lemma D.1. *Let $\sigma_y^2 = r_1 r_0 \sum_i (y_i - \bar{y})^2/n$. If $\liminf_{n\to\infty} \sigma_y > 0$ and $\max_i (y_i - \bar{y})^2 = o(n)$, then we have*

$$\frac{\sum_i T_i(y_i - \bar{y})}{\sqrt{n}\sigma_y} \dot{\sim} \mathcal{N}(0,1).$$

Proof of Lemma D.1. By the Theorem 1 of Berry [1941], we have

$$d_K\left(\frac{\sum_i T_i(y_i - \bar{y})}{\sqrt{n}\sigma_y}, \mathcal{N}(0,1)\right) < C \max_i \left|\frac{(y_i - \bar{y})}{\sqrt{n}\sigma_y}\right|,$$

where d_K denotes the Kolmogorov distance between two distributions. If $\liminf_{n\to\infty} \sigma_y > 0$ and $\max_i (y_i - \bar{y})^2 = o(n)$, then

$$\max_i \left|\frac{y_i - \bar{y}}{\sqrt{n}\sigma_y}\right| = o(1).$$

The conclusion follows. □

Proof of Theorem 1. First, we see that

$$\text{var}\left(\frac{1}{\sqrt{n}} \sum_i T_i \left(\frac{e_i(1)}{r_1} + \frac{e_i(0)}{r_0}\right)\right) = \frac{1}{n} \sum_i \left(\frac{e_i(1)}{r_1} + \frac{e_i(0)}{r_0}\right)^2 = \frac{n-1}{n} S^2_{r_1^{-1}e(1)+r_0^{-1}e(0)}.$$

By Lemma A.3, we have

$$S^2_{r_1^{-1}e(1)+r_0^{-1}e(0)} = r_1^{-1} S^2_{e(1)} + r_0^{-1} S^2_{e(0)} - S^2_{\tau_e} = \sigma^2_{\text{adj}}.$$

Putting together, we get

$$\text{var}\left(\frac{1}{\sqrt{n}} \sum_i T_i \left(\frac{e_i(1)}{r_1} + \frac{e_i(0)}{r_0}\right)\right) = \sigma^2_{\text{adj}} + o(1).$$

Then, by Lemma D.1, we see that under Assumptions 1-4,

$$\frac{1}{\sqrt{n}\sigma_{\text{adj}}} \sum_i T_i \left(\frac{e_i(1)}{r_1} + \frac{e_i(0)}{r_0}\right) \dot{\sim} \mathcal{N}(0,1).$$

Recall that $\tilde{T}_i = T_i - r_1$. By Proposition C.3, it remains to show that for $z \in \{0,1\}$

$$n^{-1/2} \sum_i T_i s_i(z) = o_{\mathbb{P}}(1), \quad n^{-1/2} \sum_{[i,j]} \tilde{T}_i \tilde{T}_j A_{ij}(z) = o_{\mathbb{P}}(1).$$

Noticing that $\sum_i H_{ii}^2 \le \sum_i H_{ii} = p$, we have

$$\mathbb{E}\left(n^{-1/2}\sum_i T_i s_i(z)\right)^2 = \frac{r_1 r_0}{n}\sum_i s_i(z)^2 \le \frac{r_1 r_0}{n}\sum_i H_{ii}^2 (Y_i(z) - \bar{Y}(z))^2$$
$$\le \frac{r_1 r_0}{n}\sum_{i=1}^{p}(Y_{(i)}(z) - \bar{Y}(z))^2 = O(n^{-1})o(n) = o(1),$$

which implies that

$$n^{-1/2}\sum_i T_i s_i(z) = o_{\mathbb{P}}(1).$$

Next, using that for $i \neq j \neq k \neq l$,

$$\mathbb{E}\tilde{T}_i^2 \tilde{T}_j \tilde{T}_k = 0, \quad \mathbb{E}\tilde{T}_i \tilde{T}_j \tilde{T}_k \tilde{T}_l = 0,$$

we can derive that

$$\mathbb{E}\left(n^{-1/2}\sum_{[i,j]} \tilde{T}_i \tilde{T}_j A_{ij}(z)\right)^2 = n^{-1}\sum_{[i,j]} \mathbb{E}(\tilde{T}_i \tilde{T}_j)^2 \left(A_{ij}(z)^2 + A_{ij}(z)A_{ji}(z)\right)$$
$$= \frac{(r_1 r_0)^2}{n}\sum_{[i,j]}\left(A_{ij}(z)^2 + A_{ij}(z)A_{ji}(z)\right) \le \frac{2(r_1 r_0)^2}{n}\sum_{[i,j]} A_{ij}(z)^2$$
$$= \frac{2(r_1 r_0)^2}{n}\sum_{[i,j]} H_{ij}^2 \left(Y_j(z) - \bar{Y}(z)\right)^2 \le \frac{2(r_1 r_0)^2}{n}\sum_i H_{ii}\left(Y_i(z) - \bar{Y}(z)\right)^2$$
$$\le \frac{2(r_1 r_0)^2}{n}\sum_{i=1}^{p}(Y_{(i)}(z) - \bar{Y}(z))^2 = o(1).$$

Thus, by Chebyshev's inequality, we have

$$n^{-1/2}\sum_i \tilde{T}_i \tilde{T}_j A_{ij}(z) = o_{\mathbb{P}}(1).$$

Then, the conclusion follows. □

E The CLT of quadratic forms and the asymptotic normality of $\hat{\tau}_{\mathrm{db}}$ when $p \asymp n$

In this section, we study the asymptotic normality of $\hat{\tau}_{\mathrm{db}}$ when $p \asymp n$ and give the proof of Theorem 2. The main intermediate step is to show that the Kolmogorov distance between the normal distribution and the joint distribution of the linear and quadratic terms of $\hat{\tau}_{\mathrm{db}}$ is negligible (see Proposition E.3).

For a symmetric function vanishing on diagonals, define the influence of the i-th variable of f by

$$\mathrm{Inf}_i(f) := \sum_{i_2,\ldots,i_q} f(i, i_2, \ldots, i_q)^2 .$$

if $q \geq 2$ and $\mathrm{Inf}_i(f) := f(i)^2$ if $q = 1$. Denote

$$\|f\|_{\ell_2} := \left\{ \sum_{i_1,\ldots,i_q} f(i_1,\ldots,i_q)^2 \right\}^{1/2}, \quad \mathcal{M}(f) := \max_{1\leq i\leq n} \mathrm{Inf}_i(f).$$

For a random variable U, define its fourth-order cumulant $\kappa_4(U) := \mathbb{E}U^4 - 3(\mathbb{E}U^2)^2$. Set W_i in Definition 1 in the main context as $W_i = (T_i - r_1)/(r_1 r_0)^{1/2}$, $i \in [n]$. With the decomposition in Proposition C.3, we define $\boldsymbol{Q} := (Q_1, Q_2)^\top$ as

$$Q_1 := \frac{\sqrt{r_1 r_0}}{\sqrt{n}} \sum_i W_i \left(\frac{e_i(1)}{r_1} + \frac{e_i(0)}{r_0} + \frac{s_i(1)}{r_1} + \frac{s_i(0)}{r_0} \right),$$
$$Q_2 := \frac{r_1 r_0}{\sqrt{n}} \sum_{[i,j]} W_i W_j \left(-\frac{A_{ij}(1)}{r_1^2} + \frac{A_{ij}(0)}{r_0^2} \right),$$

and it is easy to see that $\sqrt{n}(\hat{\tau}_{\mathrm{db}} - \bar{\tau}) = Q_1 + Q_2 + o_{\mathbb{P}}(1)$. Moreover, we rewrite Q_2 into the form of a homogeneous sum

$$Q_2 = \frac{r_1 r_0}{2\sqrt{n}} \sum_{[i,j]} W_i W_j \left(-\frac{A_{ij}(1)}{r_1^2} - \frac{A_{ji}(1)}{r_1^2} + \frac{A_{ij}(0)}{r_0^2} + \frac{A_{ji}(0)}{r_0^2} \right).$$

Q_1 and Q_2 define two homogeneous sums

$$Q_1 = \sum_i f_1(i) W_i, \quad Q_2 = \sum_{[i,j]} f_2(i,j) W_i W_j,$$

with

$$f_1(i) = \frac{\sqrt{r_1 r_0}}{\sqrt{n}} \left(\frac{e_i(1)}{r_1} + \frac{e_i(0)}{r_0} + \frac{s_i(1)}{r_1} + \frac{s_i(0)}{r_0} \right),$$
$$f_2(i,j) = \frac{r_1 r_0}{2\sqrt{n}} \left(-\frac{A_{ij}(1)}{r_1^2} - \frac{A_{ji}(1)}{r_1^2} + \frac{A_{ij}(0)}{r_0^2} + \frac{A_{ji}(0)}{r_0^2} \right).$$

The following Proposition gives the variances of the linear and quadratic terms.

Proposition E.1. *We have*

$$\mathrm{var}(Q_1) = \frac{n-1}{n} \sigma^2_{hd,l}, \quad \mathrm{var}(Q_2) = \frac{n-1}{n} \sigma^2_{hd,q}.$$

Proof of Proposition E.1. Using $\mathbb{E}W_i = 0$, $\mathbb{E}W_i^2 = 1$, and the independence between W_i's, we obtain that

$$\mathrm{var}(Q_1) = \sum_i f_1(i)^2 \,\mathrm{var}(W_1) = \sum_i f_1(i)^2$$
$$= \frac{r_1 r_0}{n} \sum_i \left(\frac{e_i(1)}{r_1} + \frac{e_i(0)}{r_0} + \frac{s_i(1)}{r_1} + \frac{s_i(0)}{r_0} \right)^2$$
$$= \frac{n-1}{n} (r_1 r_0) S^2_{r_1^{-1} e(1) + r_1^{-1} s(1) + r_0^{-1} e(0) + r_0^{-1} s(0)} = \frac{n-1}{n} \sigma^2_{\mathrm{hd},l},$$

where the last equation follows from Lemma A.3 with $a_i = e_i(1) + s_i(1)$, $b_i = e_i(0) + s_i(0)$.

Using $\sum_{j:j\neq i} H_{ij}^2 = H_{ii} - H_{ii}^2$ and recalling the definition of $\boldsymbol{Q}$, we obtain that

$$\begin{aligned}
\operatorname{var}(Q_2) &= \sum_{[i,j]} (r_1 r_0)^2 \left(\frac{A_{ij}(1)}{r_1^2} - \frac{A_{ij}(0)}{r_0^2} \right)^2 \mathbb{E}(W_i^2 W_j^2) \\
&\quad + \sum_{[i,j]} (r_1 r_0)^2 \left(\frac{A_{ij}(1)}{r_1^2} - \frac{A_{ij}(0)}{r_0^2} \right) \left(\frac{A_{ji}(1)}{r_1^2} - \frac{A_{ji}(0)}{r_0^2} \right) \mathbb{E}(W_i^2 W_j^2) \\
&= \frac{(r_1 r_0)^2}{n} \left\{ \sum_{[i,j]} H_{ij}^2 \left(\frac{Y_i(1)}{r_1^2} - \frac{Y_i(0)}{r_0^2} \right) \left(\frac{Y_j(1)}{r_1^2} - \frac{Y_j(0)}{r_0^2} \right) + \sum_{[i,j]} H_{ij}^2 \left(\frac{Y_i(1)}{r_1^2} - \frac{Y_i(0)}{r_0^2} \right) \right\} \\
&= \frac{(r_1 r_0)^2}{n} \left\{ \sum_{[i,j]} H_{ij}^2 \left(\frac{Y_i(1)}{r_1^2} - \frac{Y_i(0)}{r_0^2} \right) \left(\frac{Y_j(1)}{r_1^2} - \frac{Y_j(0)}{r_0^2} \right) + \sum_{i} (H_{ii} - H_{ii}^2) \left(\frac{Y_i(1)}{r_1^2} - \frac{Y_i(0)}{r_0^2} \right)^2 \right\} \\
&= \frac{(r_1 r_0)^2}{n} \sum_i \sum_j Q_{ij} \left(\frac{Y_i(1)}{r_1^2} - \frac{Y_i(0)}{r_0^2} \right) \left(\frac{Y_j(1)}{r_1^2} - \frac{Y_j(0)}{r_0^2} \right) \\
&= \frac{(r_1 r_0)^2}{n} \sum_i \sum_j Q_{ij} \left(\frac{Y_i(1) - \bar{Y}(1)}{r_1^2} - \frac{Y_i(0) - \bar{Y}(0)}{r_0^2} \right) \left(\frac{Y_j(1) - \bar{Y}(1)}{r_1^2} - \frac{Y_j(0) - \bar{Y}(0)}{r_0^2} \right) \\
&= \frac{n-1}{n} \sigma_{\mathrm{hd},q}^2,
\end{aligned}$$

where we used the fact that $\boldsymbol{Q}^\top \mathbf{1} = \boldsymbol{Q}\mathbf{1} = 0$. □

Let $\boldsymbol{G} := (G_1, G_2)^\top$ be a normal approximation of (Q_1, Q_2), i.e.

$$\begin{pmatrix} G_1 \\ G_2 \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \sigma_{\mathrm{hd},l}^2 & 0 \\ 0 & \sigma_{\mathrm{hd},q}^2 \end{pmatrix} \right).$$

The following proposition shows the order of $\sigma_{\mathrm{hd},l}^2$ and $\sigma_{\mathrm{hd},q}^2$.

Proposition E.2. *If Assumptions 1 and 2 hold, then we have that*

$$\sigma_{hd,l}^2 = O(1), \quad \sigma_{hd,q}^2 = O(1).$$

Proof of Proposition E.2. Recall that $\boldsymbol{B}$ is defined as

$$\boldsymbol{B} = \left\{ \left(\boldsymbol{I} - \frac{1}{n} \mathbf{1}\mathbf{1}^\top \right) - \boldsymbol{H} + \left(\boldsymbol{I} - \frac{1}{n} \mathbf{1}\mathbf{1}^\top \right) \operatorname{diag}\{\boldsymbol{H}\} \right\}^2,$$

and $\sigma_{\mathrm{hd},l}^2$ and $\sigma_{\mathrm{hd},q}^2$ are defined as

$$\sigma_{\mathrm{hd},l}^2 = (r_1 r_0) S^2_{\boldsymbol{B},\, r_1^{-1} Y(1) + r_0^{-1} Y(0)}, \quad \sigma_{\mathrm{hd},q}^2 = (r_1 r_0)^2 S^2_{\boldsymbol{Q},\, r_1^{-2} Y(1) - r_0^{-2} Y(0)}.$$

By sub-additivity and sub-multiplicativity of the l_2-norm and the trivial bounds

$$\left\| \boldsymbol{I} - \frac{1}{n} \mathbf{1}\mathbf{1}^\top \right\|_2 \leq 1, \quad \|\boldsymbol{H}\|_2 \leq 1,$$

we see that $\|\boldsymbol{B}\|_2 = O(1)$, which, combined with Assumptions 1 and 2, yields that $\sigma^2_{\mathrm{hd},l} = O(1)$. For $\sigma^2_{\mathrm{hd},q}$, we notice that

$$\|\operatorname{diag}(\boldsymbol{Q})\|_2 = \max_i (H_{ii} - H_{ii}^2) \le \frac{1}{4} = O(1).$$

Therefore, by Lemma A.7, we have

$$\|\boldsymbol{Q}\|_2 \le \|\operatorname{diag}(\boldsymbol{Q})\|_2 + \|\operatorname{diag}^-(\boldsymbol{Q})\|_2 = O(1),$$

which, combined with Assumption 1 and 2, yields that $\sigma^2_{\mathrm{hd},q} = O(1)$. This concludes the proof. □

The following proposition shows that the Kolmogorov distance between (G_1, G_2) and (Q_1, Q_2) is negligible.

Proposition E.3. *Under Assumptions 1–3 and 6, we have that*

$$\sup_{(x_1,x_2)^\top \in \mathbb{R}^2} |\mathbb{P}(Q_1 \le x_1; Q_2 \le x_2) - \mathbb{P}(G_1 \le x_1; G_2 \le x_2)| = \delta_n \left(1 + \frac{1}{\min\{\sigma_{hd,q}, \sigma_{hd,l}\}}\right),$$

for a deterministic parameter δ_n of order $o(1)$.

Proof of Proposition E.3. For ease of presentation, we denote

$$g_i := (r_1 r_0)^{1/2} \left(\frac{e_i(1)}{r_1} + \frac{e_i(0)}{r_0} + \frac{s_i(1)}{r_1} + \frac{s_i(0)}{r_0}\right),$$
$$y_i := (r_1 r_0) \left(-\frac{Y_i(1) - \bar{Y}(1)}{r_1^2} + \frac{Y_i(0) - \bar{Y}(0)}{r_0^2}\right).$$

Therefore, we can rewrite that

$$f_1(i) = n^{-1/2} g_i, \quad f_2(i,j) = (2\sqrt{n})^{-1} H_{ij}(y_i + y_j).$$

By Theorem 2.1 of Koike [2022], we have

$$\sup_{(x_1,x_2)^\top \in \mathbb{R}^2} |\mathbb{P}(Q_1 \le x_1; Q_2 \le x_2) - \mathbb{P}(G_1 \le x_1; G_2 \le x_2)|$$
$$= C \left(\delta_0(\boldsymbol{Q})^{\frac{1}{3}} + \delta_1(\boldsymbol{Q})^{\frac{1}{3}} + \max_{1 \le k \le 2} (\mathcal{M}(f_k))^{1/2}\right) \left(1 + \frac{1}{\min\{\sigma_{\mathrm{hd},q}, \sigma_{\mathrm{hd},l}\}}\right),$$

where C is a universal constant that does not depend on n and

$$\delta_0(\boldsymbol{Q}) := \|\operatorname{cov}(\boldsymbol{Q}) - \operatorname{cov}(\boldsymbol{G})\|_\infty,$$
$$\delta_1(\boldsymbol{Q}) := \left(|\kappa_4(Q_1)| + \sum_i \operatorname{Inf}_i(f_1)^2\right)^{1/2} + \left(|\kappa_4(Q_2)| + \sum_i \operatorname{Inf}_i(f_2)^2\right)^{1/2}$$
$$+ \|f_1\|_{\ell_2} \left(|\kappa_4(Q_2)| + \sum_i \operatorname{Inf}_i(f_2)^2\right)^{1/4}.$$

Now, we set

$$\delta_n = C\left(\delta_0(\boldsymbol{Q})^{\frac{1}{3}} + \delta_1(\boldsymbol{Q})^{\frac{1}{3}} + \max_{1\le k\le 2}\left(\mathcal{M}\left(f_k\right)\right)^{1/2}\right).$$

To conclude the proof, it suffices to show that $\delta_0(\boldsymbol{Q})$, $\delta_1(\boldsymbol{Q})$, $\max_{1\le k\le 2}\mathcal{M}\left(f_k\right)$ are all of order $o(1)$.

By Proposition E.2, we have that

$$\begin{aligned}
&\operatorname{cov}(Q_1, Q_2) = \operatorname{cov}(G_1, G_2) = 0,\\
&\operatorname{var}(Q_1) - \sigma^2_{\mathrm{hd},l} = \frac{n-1}{n}\sigma^2_{\mathrm{hd},l} - \sigma^2_{\mathrm{hd},l} = -\frac{1}{n}\sigma^2_{\mathrm{hd},l} = o(1),\\
&\operatorname{var}(Q_2) - \sigma^2_{\mathrm{hd},q} = \frac{n-1}{n}\sigma^2_{\mathrm{hd},q} - \sigma^2_{\mathrm{hd},q} = -\frac{1}{n}\sigma^2_{\mathrm{hd},q} = o(1).
\end{aligned}$$

These estimates give that $\delta_0(\boldsymbol{Q}) = o(1)$. We then consider $\max_{k\in\{1,2\}}\mathcal{M}\left(f_k\right)$. We have that

$$\max_i |s_i(z)| < 2\max_i |H_{ii}(Y_i(z) - \bar{Y}(z))| < 2\max_i |Y_i(z) - \bar{Y}(z)| = o(n^{1/2}),$$

which, combined with Assumption 1 and 3, yields that $\max_i g_i^2 = o(n)$ and $\max_i y_i^2 = o(n)$. As a consequence, we obtain that

$$\begin{aligned}
\mathcal{M}(f_1) &= \max_i \operatorname{Inf}_i(f_1) = \max_i f_1(i)^2 = \max_i g_i^2/n = o(1),\\
\mathcal{M}(f_2) &= \max_i \operatorname{Inf}_i(f_2) = \max_i \sum_j f_2(i,j)^2(1-\delta_{ij}) = \max_i \sum_j H_{ij}^2(y_i + y_j)^2(1-\delta_{ij})/(4n)\\
&\le \max_i y_i^2 \sum_j H_{ij}^2(1-\delta_{ij})/n = O\Big(\max_i y_i^2/n\Big) = o(1).
\end{aligned}$$

Finally, we estimate $\delta_1(\boldsymbol{Q})$. We first focus on terms relating to Q_1 and f_1. We see that

$$\sum_i \operatorname{Inf}_i(f_1)^2 = \sum_i f_1(i)^4 = \sum_i g_i^4/n^2 = \max_i g_i^2/n \cdot \sum_i g_i^2/n = o\Big(\sum_i g_i^2/n\Big).$$

Then, using

$$\frac{1}{n}\sum_i g_i^2 = O\bigg(\max_z \frac{1}{n}\sum_i (e_i(z)^2 + s_i^2(z))\bigg) = O\bigg(\max_z \frac{1}{n}\sum_i (Y_i - \bar{Y}(z))^2\bigg) = O(1),$$

we get that $\sum_i \operatorname{Inf}_i(f_1)^2 = o(1)$ and

$$\|f_1\|_{\ell_2} = \left\{\sum_i f_1\left(i\right)^2\right\}^{1/2} = \left(\sum_i g_i^2/n\right)^{1/2} = O(1). \tag{22}$$

On the other hand, for $\kappa_4\left(Q_1\right)$, we have that

$$\begin{aligned}
&\mathbb{E}Q_1^4 = \sum_i f_1(i)^4\mathbb{E}W_i^4 + 3\sum_{i\ne j} f_1(i)^2 f_1(j)^2\mathbb{E}\left(W_i^2W_j^2\right),\\
&3\left(\mathbb{E}Q_1^2\right)^2 = 3\bigg(\sum_i f_1(i)^2\mathbb{E}W_i^2\bigg)^2 = 3\sum_i f_1(i)^4\left(\mathbb{E}W_i^2\right)^2 + 3\sum_{i\ne j} f_1(i)^2 f_1(j)^2\mathbb{E}W_i^2\cdot\mathbb{E}W_j^2,
\end{aligned}$$

which yield that

$$\kappa_4(Q_1) = \sum_i f_1(i)^4 \left[\mathbb{E}W_i^4 - 3\left(\mathbb{E}W_i^2\right)^2\right] = \kappa_4(W_1) \sum_i f_1(i)^4 = \kappa_4(W_1) \sum_i g_i^4/n^2$$
$$\leq \kappa_4(W_1)\left(\max_i g_i^2\right) \sum_i g_i^2/n^2 = o(1).$$

Next, we focus on terms relating to Q_2 and f_2. For $\mathrm{Inf}_i(f_2)^2$, using that $(y_i + y_j)^2 \leq 2(y_i^2 + y_j^2)$, we get

$$\sum_i \mathrm{Inf}_i(f_2)^2 = O\Bigg(\sum_i \Big\{\sum_j H_{ij}^2 y_i^2 (1-\delta_{ij})/n\Big\}^2 + \sum_i \Big\{\sum_j H_{ij}^2 y_j^2 (1-\delta_{ij})/n\Big\}^2\Bigg).$$

Expanding the above two terms, we get

$$\sum_i \Big\{\sum_j H_{ij}^2 y_j^2 (1-\delta_{ij})/n\Big\}^2 = \sum_{[i_1,i_2]} H_{i_1 i_2}^4 y_{i_1}^4/n^2 + \sum_{[i_1 \ldots i_3]} H_{i_1 i_2}^2 H_{i_2 i_3}^2 y_{i_1}^2 y_{i_3}^2/n^2 =: M_{11} + M_{12},$$
$$\sum_i \Big\{\sum_j H_{ij}^2 y_i^2 (1-\delta_{ij})/n\Big\}^2 = \sum_{[i_1,i_2]} H_{i_1 i_2}^4 y_{i_2}^4/n^2 + \sum_{[i_1 \ldots i_3]} H_{i_1 i_2}^2 H_{i_2 i_3}^2 y_{i_2}^4/n^2 =: M_{13} + M_{14}.$$

First, we use Lemma A.6 and Lemma A.7 to get that

$$M_{11} = M_{13} = \sum_{[i_1,i_2]} H_{i_1 i_2}^4 y_{i_1}^4/n^2 = \mathrm{tr}\left(\mathrm{diag}(\boldsymbol{y})^2 \,\mathrm{diag}^-(\boldsymbol{Q})\,\mathrm{diag}^-(\boldsymbol{Q})\,\mathrm{diag}(\boldsymbol{y})^2\right)/n^2$$
$$\leq \sum_i y_i^4/n^2 < \left(\max_i y_i^2\right) \sum_i y_i^2/n^2 = o(1).$$

For M_{12}, by repeatedly applying $\sum_{j:j\neq i} H_{ij}^2 \leq H_{ii} \leq 1$, we get

$$M_{12} = \sum_{[i_1 \ldots i_3]} H_{i_1 i_2}^2 H_{i_2 i_3}^2 y_{i_1}^2 y_{i_3}^2/n^2 \leq \left(\max_i y_i^2\right) \sum_{[i_1 \ldots i_3]} H_{i_1 i_2}^2 H_{i_2 i_3}^2 y_{i_1}^2/n^2$$
$$\leq \left(\max_i y_i^2\right) \sum_{[i_1,i_2]} H_{i_1 i_2}^2 y_{i_1}^2/n^2 \leq \left(\max_i y_i^2\right) \sum_i y_i^2/n^2 = o(1),$$

With a similar argument, we can bound M_{14} as

$$M_{14} = \sum_{[i_1 \ldots i_3]} H_{i_1 i_2}^2 H_{i_2 i_3}^2 y_{i_2}^4/n^2 \leq \sum_i y_i^4/n^2 = o(1).$$

Putting together, we see that

$$\sum_i \mathrm{Inf}_i(f_2)^2 = o(1).$$

We next show that $\kappa_4(Q_2) = o(1)$. For ease of presentation, we abbreviate $f_2(i,j)$ as f_{ij}. Since $f_{ij} = f_{ji}$, we see from some basic combinatorics that

$$
\begin{aligned}
\mathbb{E}Q_2^4 &= \mathbb{E}\Big(\sum_{[i_1,i_2]} f_{i_1i_2} W_{i_1} W_{i_2}\Big)^4 \\
&= C_1 \sum_{[i_1,i_2]} f_{i_1i_2}^4 \mathbb{E}W_{i_1}^4 W_{i_2}^4 + C_2 \sum_{[i_1\ldots i_3]} f_{i_1i_2}^2 f_{i_2i_3} f_{i_3i_1} \mathbb{E}W_{i_1}^3 W_{i_2}^3 W_{i_3}^2 \\
&\quad + C_3 \sum_{[i_1\ldots i_3]} f_{i_1i_2}^2 f_{i_2i_3}^2 \mathbb{E}W_{i_1}^2 W_{i_2}^4 W_{i_3}^2 + C_4 \sum_{[i_1\ldots i_4]} f_{i_1i_2} f_{i_2i_3} f_{i_3i_4} f_{i_4i_1} \mathbb{E}W_{i_1}^2 W_{i_2}^2 W_{i_3}^2 W_{i_4}^2 \\
&\quad + 12 \sum_{[i_1\ldots i_4]} f_{i_1i_2}^2 f_{i_3i_4}^2 \mathbb{E}W_{i_1}^2 W_{i_2}^2 W_{i_3}^2 W_{i_4}^2,
\end{aligned}
$$

where $C_1, \ldots, C_4$ are universal constant that do not depend on n. On the other hand, we can calculate that

$$
\begin{aligned}
3\left(\mathbb{E}Q_2^2\right)^2 &= 3\Big\{\mathbb{E}\Big(\sum_{[i_1,i_2]} f_{i_1i_2} W_{i_1} W_{i_2}\Big)^2\Big\}^2 = 3\Big(2\sum_{[i_1,i_2]} f_{i_1i_2}^2 \mathbb{E}W_{i_1}^2 W_{i_2}^2\Big)^2 \\
=& C_6 \sum_{[i_1,i_2]} f_{i_1i_2}^4 \left(\mathbb{E}W_{i_1}^2 W_{i_2}^2\right)^2 + C_7 \sum_{[i_1\ldots i_3]} f_{i_1i_2}^2 f_{i_2i_3}^2 \left(\mathbb{E}W_{i_1}^2 W_{i_2}^2\right)^2 + 12 \sum_{[i_1\ldots i_4]} f_{i_1i_2}^2 f_{i_3i_4}^2 \left(\mathbb{E}W_{i_1}^2 W_{i_2}^2\right)^2,
\end{aligned}
$$

where C_6 and C_7 are universal constants that do not depend on n. Using the cancellation of the term $\sum_{[i_1\ldots i_4]} f_{i_1i_2}^2 f_{i_3i_4}^2$, we obtain that

$$
\kappa_4(Q_2) = O\left(|M_{21}| + |M_{22}| + |M_{23}| + |M_{24}|\right),
$$

where

$$
\begin{aligned}
M_{21} &= \sum_{[i_1,i_2]} f_{i_1i_2}^4, \quad M_{22} = \sum_{[i_1\ldots i_3]} f_{i_1i_2}^2 f_{i_2i_3} f_{i_3i_1}, \\
M_{23} &= \sum_{[i_1\ldots i_3]} f_{i_1i_2}^2 f_{i_2i_3}^2, \quad M_{24} = \sum_{[i_1\ldots i_4]} f_{i_1i_2} f_{i_2i_3} f_{i_3i_4} f_{i_4i_1}.
\end{aligned}
$$

We handle these terms one by one.

The term M_{21} can be written as a summation of terms of the form

$$
\sum_{[i_1,i_2]} H_{i_1i_2}^4 y_{i_1}^{m_1} y_{i_2}^{m_2}/n^2, \quad (m_1, m_2) \in \mathbb{N}^2,\ m_1 + m_2 = 4.
$$

(We adopt the convention that $0 \in \mathbb{N}$.) By Lemmas A.6 and A.7, for any $(m_1, m_2) \in \mathbb{N}^2$ with $m_1 + m_2 = 4$,

$$
\begin{aligned}
\sum_{[i_1,i_2]} H_{i_1i_2}^4 y_{i_1}^{m_1} y_{i_2}^{m_2}/n^2 &= \operatorname{tr}\left(\operatorname{diag}(\boldsymbol{y})^{m_1} \operatorname{diag}^-(\boldsymbol{Q}) \operatorname{diag}(\boldsymbol{y})^{m_2} \operatorname{diag}^-(\boldsymbol{Q})\right)/n^2 \qquad (23) \\
&\le \sum_i y_i^4/n^2 = o(1),
\end{aligned}
$$

which implies that $|M_{21}| = o(1)$.

By Cauchy-Schwarz inequality, we have that $|M_{22}| \leq M_{23}$. The term M_{23} can be written as a summation of terms of the form

$$\sum_{[i_1 \dots i_3]} H_{i_1 i_2}^2 H_{i_2 i_3}^2 y_{i_1}^{m_1} y_{i_2}^{m_2} y_{i_3}^{m_3} / n^2, \quad (m_1, m_2, m_3) \in \mathbb{N}^3,\ m_1 + m_2 + m_3 = 4.$$

We can find $(m_1^{(1)}, m_2^{(1)}, m_3^{(1)}) \in \mathbb{N}^3$ and $(m_1^{(2)}, m_2^{(2)}, m_3^{(2)}) \in \mathbb{N}^3$ such that $m_i = m_i^{(1)} + m_i^{(2)}$, $i = 1, 2, 3$, and

$$m_1^{(1)} + m_2^{(1)} + m_3^{(1)} = 2, \quad m_1^{(2)} + m_2^{(2)} + m_3^{(2)} = 2.$$

Then, applying the Cauchy-Schwartz inequality, we can obtain that

$$\left| \sum_{[i_1 \dots i_3]} H_{i_1 i_2}^2 H_{i_2 i_3}^2 y_{i_1}^{m_1} y_{i_2}^{m_2} y_{i_3}^{m_3} \right| \leq$$
$$\left(\sum_{[i_1 \dots i_3]} H_{i_1 i_2}^2 H_{i_2 i_3}^2 y_{i_1}^{2m_1^{(1)}} y_{i_2}^{2m_2^{(1)}} y_{i_3}^{2m_3^{(1)}} \right)^{1/2} \left(\sum_{[i_1 \dots i_3]} H_{i_1 i_2}^2 H_{i_2 i_3}^2 y_{i_1}^{2m_1^{(2)}} y_{i_2}^{2m_2^{(2)}} y_{i_3}^{2m_3^{(2)}} \right)^{1/2}.$$

Mimicking the above proof for M_{12}, by repeatedly applying $\sum_{j:j \neq i} H_{ij}^2 \leq H_{ii} \leq 1$ and the condition $\max_i y_i^2 = o(n)$, we get that

$$\frac{1}{n^2} \sum_{[i_1 \dots i_3]} H_{i_1 i_2}^2 H_{i_2 i_3}^2 y_{i_1}^{2m_1} y_{i_2}^{2m_2} y_{i_3}^{2m_3} = o(1), \quad \forall (m_1, m_2, m_3) \in \mathbb{N}^3, m_1 + m_2 + m_3 = 2.$$

As a consequence, it implies that

$$\frac{1}{n^2} \sum_{[i_1 \dots i_3]} H_{i_1 i_2}^2 H_{i_2 i_3}^2 y_{i_1}^{m_1} y_{i_2}^{m_2} y_{i_3}^{m_3} = o(1),\ \forall (m_1, m_2, m_3) \in \mathbb{N}^3, m_1 + m_2 + m_3 = 4, \tag{24}$$

so we have $|M_{23}| = o(1)$.

Finally, M_{24} can be written as a summation of terms of the form

$$\sum_{[i_1 \dots i_4]} H_{i_1 i_2} H_{i_2 i_3} H_{i_3 i_4} H_{i_4 i_1} y_{i_1}^{m_1} y_{i_2}^{m_2} y_{i_3}^{m_3} y_{i_4}^{m_4} / n^2, \tag{25}$$

for $(m_1, m_2, m_3, m_4) \in \mathbb{N}^4$ with $m_1 + m_2 + m_3 + m_4 = 4$. To bound this term, we estimate an intermediate quantity

$$\sum_{i_1 \neq i_2, i_2 \neq i_3, i_3 \neq i_4, i_4 \neq i_1} H_{i_1 i_2} H_{i_2 i_3} H_{i_3 i_4} H_{i_4 i_1} y_{i_1}^{m_1} y_{i_2}^{m_2} y_{i_3}^{m_3} y_{i_4}^{m_4} / n^2,$$

and define

$$\Delta := \left(\sum_{i_1 \neq i_2, i_2 \neq i_3, i_3 \neq i_4, i_4 \neq i_1} - \sum_{[i_1 \dots i_4]} \right) H_{i_1 i_2} H_{i_2 i_3} H_{i_3 i_4} H_{i_4 i_1} y_{i_1}^{m_1} y_{i_2}^{m_2} y_{i_3}^{m_3} y_{i_4}^{m_4} / n^2.$$

We observe that

$$\operatorname{tr}\left[\prod_{v=1}^{4}\left(\operatorname{diag}(\boldsymbol{y})^{m_v}\operatorname{diag}^{-}(\boldsymbol{H})\right)\right]/n^2$$
$$=\sum_{i_1\neq i_2,i_2\neq i_3,i_3\neq i_4,i_4\neq i_1} H_{i_1i_2}H_{i_2i_3}H_{i_3i_4}H_{i_4i_1}y_{i_1}^{m_1}y_{i_2}^{m_2}y_{i_3}^{m_3}y_{i_4}^{m_4}/n^2,$$

and Δ can be written as a summation of terms of the forms

$$\sum_{[i_1,i_2]} H_{i_1i_2}^4 y_{i_1}^{m_1'}y_{i_2}^{m_2'}/n^2, \quad (m_1',m_2')\in\mathbb{N}^2,\ m_1'+m_2'=4,$$

and

$$\sum_{[i_1\dots i_3]} H_{i_1i_2}^2 H_{i_2i_3}^2 y_{i_1}^{m_1'}y_{i_2}^{m_2'}y_{i_3}^{m_3'}/n^2, \quad (m_1',m_2',m_3')\in\mathbb{N}^3,\ m_1'+m_2'+m_3'=4.$$

By Lemmas A.6 and A.7, we have

$$\operatorname{tr}\left[\prod_{v=1}^{4}\left(\operatorname{diag}(\boldsymbol{y})^{m_v}\operatorname{diag}^{-}(\boldsymbol{H})\right)\right]/n^2 = O\left(\sum_i y_i^4/n^2\right) = o(1).$$

On the other hand, by (23) and (24), we have $\Delta = o(1)$. Putting together, we have that for all $(m_1,m_2,m_3,m_4)\in\mathbb{N}^4$ with $m_1+m_2+m_3+m_4=4$,

$$\sum_{[i_1\dots i_4]} H_{i_1i_2}H_{i_2i_3}H_{i_3i_4}H_{i_4i_1}y_{i_1}^{m_1}y_{i_2}^{m_2}y_{i_3}^{m_3}y_{i_4}^{m_4}/n^2 = o(1), \tag{26}$$

which yields that $|M_{24}| = o(1)$. Combining all the above estimates, we conclude that $\kappa_4(Q_2) = o(1)$.

In light of (22) and our bounds on $\kappa_4(Q_k)$ and $\sum_i \operatorname{Inf}_i(f_k)^2$, $k\in\{1,2\}$, there is $\delta_1(\boldsymbol{Q}) = o(1)$, which concludes the proof. ☐

The following proposition shows the marginal convergence of Q_1 and Q_2.

Proposition E.4. *Assume Assumptions 1–3 holds. We have that*

(i) if $\liminf \sigma_{hd,l}^2 > 0$, *then* $Q_1/\sigma_{hd,l} \xrightarrow{\text{d}} \mathcal{N}(0,1)$;

(ii) if $\liminf \sigma_{hd,q}^2 > 0$, *then* $Q_2/\sigma_{hd,q} \xrightarrow{\text{d}} \mathcal{N}(0,1)$.

Proof of Proposition E.4. By Theorem 1 of De Jong [1990], we have

$$Q_k/\operatorname{var}(Q_k)^{1/2} \xrightarrow{\text{d}} \mathcal{N}(0,1), \quad k\in\{1,2\},$$

provided that the following two conditions hold: (i) $\mathbb{E}Q_k^4/(\mathbb{E}Q_k^2)^2 \to 3$; (ii) $\mathcal{M}(f_k)/\operatorname{var}(Q_k)\to 0$. Moreover, by Proposition E.1, we have

$$\operatorname{var}(Q_1) = \sigma_{\text{hd},l}^2 + o(1), \quad \operatorname{var}(Q_2) = \sigma_{\text{hd},q}^2 + o(1).$$

From the proof of Proposition E.3, we have seen that under Assumptions 1–3, $\kappa_4(Q_k)\to 0$ and $\mathcal{M}(f_k)\to 0$ for $k\in\{1,2\}$. If $\liminf \sigma_{\text{hd},l}^2 > 0$ and $k=1$, $\kappa_4(Q_1)\to 0$ and $\mathcal{M}(f_1)\to 0$ imply conditions (i) and (ii); if $\liminf \sigma_{\text{hd},q}^2 > 0$ and $k=2$, then $\kappa_4(Q_2)\to 0$ and $\mathcal{M}(f_2)\to 0$ imply conditions (i) and (ii). Thus, we conclude the proof. ☐

Proof of Theorem 2. Without loss of generality, we assume $\liminf \sigma^2_{\mathrm{hd},l} > 0$. We split the entire sequence into two subsequences. The first subsequence is such that all $\sigma_{\mathrm{hd},q}$'s are larger than $\delta_n^{1/6}$, the second is such that all $\sigma_{\mathrm{hd},q}$'s are smaller than $\delta_n^{1/6}$.

For the first subsequence, we have

$$1 + \frac{1}{\min\{\sigma_{\mathrm{hd},q}, \sigma_{\mathrm{hd},l}\}} = O\left(1 + \sigma^{-1}_{\mathrm{hd},q} + \sigma^{-1}_{\mathrm{hd},l}\right) = O(1 + 1 + \delta_n^{-1/6}) = O(\delta_n^{-1/6}).$$

which yields that

$$\sup_{(x_1,x_2)^\top \in \mathbb{R}^2} |\mathbb{P}(Q_1 \le x_1; Q_2 \le x_2) - \mathbb{P}(G_1 \le x_1; G_2 \le x_2)| \le O(\delta_n \delta_n^{-1/6}) = O(\delta_n^{5/6}).$$

We first show that given any $x \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{Q_1 + Q_2}{\sigma_{\mathrm{hd}}} \le x\right) \le \mathbb{P}\left(\frac{G_1 + G_2}{\sigma_{\mathrm{hd}}} \le x\right) + o(1).$$

Let $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ be the ceiling and floor functions, respectively. We decompose the left-hand side as

$$\begin{aligned}
&\mathbb{P}\left(\frac{Q_1 + Q_2}{\sigma_{\mathrm{hd}}} \le x\right) \\
&\le \sum_{t=\lfloor -\delta_n^{-2/3} \rfloor}^{\lceil \delta_n^{-2/3} \rceil} \mathbb{P}\left(\frac{Q_1 + Q_2}{\sigma_{\mathrm{hd}}} \le x, (t-1)\cdot \delta_n^{1/2} \le Q_1 \le t \cdot \delta_n^{1/2}\right) + \mathbb{P}\left(|Q_1| \ge \delta_n^{-1/6}\right) \\
&\le \sum_{t=\lfloor -\delta_n^{-2/3} \rfloor}^{\lceil \delta_n^{-2/3} \rceil} \mathbb{P}\left(Q_2 \le \sigma_{\mathrm{hd}} x - (t-1)\cdot \delta_n^{1/2}, (t-1)\cdot \delta_n^{1/2} \le Q_1 \le t \cdot \delta_n^{1/2}\right) + \mathbb{P}\left(|Q_1| \ge \delta_n^{-1/6}\right) \\
&\le \sum_{t=\lfloor -\delta_n^{-2/3} \rfloor}^{\lceil \delta_n^{-2/3} \rceil} \mathbb{P}\left(G_2 \le \sigma_{\mathrm{hd}} x - (t-1)\cdot \delta_n^{1/2}, (t-1)\cdot \delta_n^{1/2} \le G_1 \le t \cdot \delta_n^{1/2}\right) \\
&\quad + \mathbb{P}\left(|Q_1| \ge \delta_n^{-1/6}\right) + O\left(\delta_n^{-2/3}\delta_n^{5/6}\right) \\
&\le \sum_{t=\lfloor -\delta_n^{-2/3} \rfloor}^{\lceil \delta_n^{-2/3} \rceil} \mathbb{P}\left(\frac{G_1 + G_2}{\sigma_{\mathrm{hd}}} \le x, (t-1)\cdot \delta_n^{1/2} \le G_1 \le t \cdot \delta_n^{1/2}\right) \\
&\quad + \sum_{t=\lfloor -\delta_n^{-2/3} \rfloor}^{\lceil \delta_n^{-2/3} \rceil} \mathbb{P}\left(\sigma_{\mathrm{hd}} x - t \cdot \delta_n^{1/2} \le G_2 \le \sigma_{\mathrm{hd}} x - (t-1)\cdot \delta_n^{1/2}, (t-1)\cdot \delta_n^{1/2} \le G_1 \le t \cdot \delta_n^{1/2}\right) \\
&\quad + \mathbb{P}\left(|Q_1| \ge \delta_n^{-1/6}\right) + O\left(\delta_n^{1/6}\right).
\end{aligned}$$

The second term on the right-hand side is of order

$$\delta_n^{5/6} \cdot \delta_n^{-2/3} = \delta_n^{1/6} = o(1),$$

since, by $\liminf \sigma_{\mathrm{hd},l} > 0$ and $\sigma_{\mathrm{hd},q} > \delta_n^{1/6}$, we have that

$$\mathbb{P}\left(\sigma_{\mathrm{hd}} x - t \cdot \delta_n^{1/2} \le G_2 \le \sigma_{\mathrm{hd}} x - (t-1) \cdot \delta_n^{1/2}, (t-1) \cdot \delta_n^{1/2} \le G_1 \le t \cdot \delta_n^{1/2}\right)$$
$$= O\left(\frac{\delta_n^{1/2}}{\sigma_{\mathrm{hd},q}} \frac{\delta_n^{1/2}}{\sigma_{\mathrm{hd},l}}\right) = O\left(\delta_n^{5/6}\right).$$

By Lemma E.2, we have

$$\mathrm{var}(Q_1)/\sigma_n^{-1/6} = \frac{n-1}{n} \sigma_{\mathrm{hd},l}^2 \sigma_n^{1/6} = O(\sigma_n^{1/6}),$$

which, by Chebyshev's inequality, implies that the third term on the right-hand side and $\mathbb{P}(|G_1| \ge \delta_n^{-1/6})$ are both negligible. As a consequence, we have

$$\begin{aligned}
\mathbb{P}\left(\frac{Q_1 + Q_2}{\sigma_{\mathrm{hd}}} \le x\right) &\le \sum_{t=\lfloor -\delta_n^{-2/3} \rfloor}^{\lceil \delta_n^{-2/3} \rceil} \mathbb{P}\left(\frac{G_1 + G_2}{\sigma_{\mathrm{hd}}} \le x, (t-1) \cdot \delta_n^{1/2} \le G_1 \le t \cdot \delta_n^{1/2}\right) + o(1) \\
&\le \mathbb{P}\left(\frac{G_1 + G_2}{\sigma_{\mathrm{hd}}} \le x\right) + \mathbb{P}(|G_1| \ge \delta_n^{-1/6}) + o(1) \\
&\le \mathbb{P}\left(\frac{G_1 + G_2}{\sigma_{\mathrm{hd}}} \le x\right) + o(1).
\end{aligned}$$

For the lower bound, we apply similar arguments as above to get that

$$\mathbb{P}\left(\frac{Q_1+Q_2}{\sigma_{\mathrm{hd}}}\leq x\right)\geq\sum_{t=\lfloor-\delta_n^{-2/3}\rfloor}^{\lceil\delta_n^{-2/3}\rceil}\mathbb{P}\left(\frac{Q_1+Q_2}{\sigma_{\mathrm{hd}}}\leq x,(t-1)\cdot\delta_n^{1/2}\leq Q_1\leq t\cdot\delta_n^{1/2}\right)$$

$$\geq\sum_{t=\lfloor-\delta_n^{-2/3}\rfloor}^{\lceil\delta_n^{-2/3}\rceil}\mathbb{P}\left(Q_2\leq\sigma_{\mathrm{hd}}x-t\cdot\delta_n^{1/2},(t-1)\cdot\delta_n^{1/2}\leq Q_1\leq t\cdot\delta_n^{1/2}\right)$$

$$\geq\sum_{t=\lfloor-\delta_n^{-2/3}\rfloor}^{\lceil\delta_n^{-2/3}\rceil}\mathbb{P}\left(G_2\leq\sigma_{\mathrm{hd}}x-t\cdot\delta_n^{1/2},(t-1)\cdot\delta_n^{1/2}\leq G_1\leq t\cdot\delta_n^{1/2}\right)-O(\delta_n^{1/6})$$

$$\geq\sum_{t=\lfloor-\delta_n^{-2/3}\rfloor}^{\lceil\delta_n^{-2/3}\rceil}\mathbb{P}\left(\frac{G_1+G_2}{\sigma_{\mathrm{hd}}}\leq x,(t-1)\cdot\delta_n^{1/2}\leq G_1\leq t\cdot\delta_n^{1/2}\right)$$

$$-\sum_{t=\lfloor-\delta_n^{-2/3}\rfloor}^{\lceil\delta_n^{-2/3}\rceil}\mathbb{P}\left(\sigma_{\mathrm{hd}}x-t\cdot\delta_n^{1/2}\leq G_2\leq\sigma_{\mathrm{hd}}x-(t-1)\cdot\delta_n^{1/2},(t-1)\cdot\delta_n^{1/2}\leq G_1\leq t\cdot\delta_n^{1/2}\right)$$

$$-O(\delta_n^{1/6})$$

$$\geq\sum_{t=\lfloor-\delta_n^{-2/3}\rfloor}^{\lceil\delta_n^{-2/3}\rceil}\mathbb{P}\left(\frac{G_1+G_2}{\sigma_{\mathrm{hd}}}\leq x,(t-1)\cdot\delta_n^{1/2}\leq G_1\leq t\cdot\delta_n^{1/2}\right)+o(1)$$

$$\geq\mathbb{P}\left(\frac{G_1+G_2}{\sigma_{\mathrm{hd}}}\leq x\right)-\mathbb{P}(|G_1|\geq\delta_n^{-1/6})+o(1)\geq\mathbb{P}\left(\frac{G_1+G_2}{\sigma_{\mathrm{hd}}}\leq x\right)+o(1).$$

Putting together the upper and lower bounds, we get that the first subsequence satisfies

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{Q_1+Q_2}{\sigma_{\mathrm{hd}}}\leq x\right)-\mathbb{P}\left(\frac{G_1+G_2}{\sigma_{\mathrm{hd}}}\leq x\right)\right|=o(1).$$

We now consider the second subsequence where $\sigma_{\mathrm{hd},q}$'s are all smaller than $\delta_n^{1/6}$ which implies that $\sigma_{\mathrm{hd}}/\sigma_{\mathrm{hd},l}=1+o(1)$ in this sequence. As a consequence, we have $Q_1/\sigma_{\mathrm{hd}}=o_{\mathbb{P}}(1)$, which means that

$$\frac{Q_1+Q_2}{\sigma_{\mathrm{hd}}}=\frac{Q_1}{\sigma_{\mathrm{hd}}}+o_{\mathbb{P}}(1).$$

By Proposition E.4, we have $Q_1/\sigma_{\mathrm{hd},l}\xrightarrow{\mathrm{d}}\mathcal{N}(0,1)$. Together with Slutsky's theorem, it implies that $(Q_1+Q_2)/\sigma_{\mathrm{hd}}\xrightarrow{\mathrm{d}}\mathcal{N}(0,1)$.

In sum, we have that for each $x\in\mathbb{R}$,

$$\left|\mathbb{P}\left(\frac{Q_1+Q_2}{\sigma_{\mathrm{hd}}}\leq x\right)-\mathbb{P}\left(\frac{G_1+G_2}{\sigma_{\mathrm{hd}}}\leq x\right)\right|=o(1)$$

for both subsequences, showing that this estimate indeed holds for the whole sequence. This gives that

$$\frac{Q_1+Q_2}{\sigma_{\mathrm{hd}}}\xrightarrow{\mathrm{d}}\mathcal{N}(0,1),$$

and the conclusion then follows from Proposition C.3. □

F Inference

In this section, we study the validity of the proposed inference procedure. It includes the proofs for Theorem 3 and Corollary 2. The comment of (10) follows from the following proposition.

Proposition F.1. *We have*

$$\sigma^2_{hd,l} = (r_1 r_0) S^2_{\boldsymbol{B},\, r_1^{-1}Y(1)+r_0^{-1}Y(0)}.$$

Proof of Proposition F.1. Using Lemma A.3 with $a_i = e_i(1) + s_i(1)$ and $b_i = e_i(0) + s_i(0)$, we get

$$\sigma^2_{\mathrm{hd},l} = (r_1 r_0) S^2_{r_1^{-1}e(1)+r_1^{-1}s(1)+r_0^{-1}e(0)+r_0^{-1}s(0)}.$$

Denote $\boldsymbol{P} = \boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^\top$. Observe that

$$\begin{aligned}(e_1(z), \ldots, e_n(z))^\top &= (\boldsymbol{I} - \boldsymbol{H})(Y_1(z) - \bar{Y}(z), \ldots, Y_n(z) - \bar{Y}(z))^\top \\ &= (\boldsymbol{P} - \boldsymbol{H})(Y_1(z) - \bar{Y}(z), \ldots, Y_n(z) - \bar{Y}(z))^\top,\end{aligned}$$

and

$$(s_1(z), \ldots, s_n(z))^\top = \boldsymbol{P}\,\mathrm{diag}(\boldsymbol{H})(Y_1(z) - \bar{Y}(z), \ldots, Y_n(z) - \bar{Y}(z))^\top.$$

Then, applying Lemma A.2 with $a_i = r_1^{-1}e_i(1) + r_1^{-1}s_i(1) + r_0^{-1}e_i(0) + r_0^{-1}s_i(0)$, $b_i = r_1^{-1}Y_i(1) + r_0^{-1}Y_i(0)$ and $\boldsymbol{M} = (\boldsymbol{P} - \boldsymbol{H}) + \boldsymbol{P}\,\mathrm{diag}(\boldsymbol{H})$, and noticing that $\sum_i a_i = 0$, we obtain that

$$S^2_{r_1^{-1}e(1)+r_1^{-1}s(1)+r_0^{-1}e(0)+r_0^{-1}s(0)} = S^2_{\boldsymbol{M}^\top\boldsymbol{M}, r_1^{-1}Y(1)+r_0^{-1}Y(0)}.$$

The conclusion then follows by the definition of $\boldsymbol{B}$ in (9). □

Theorem 3 and Corollary 2 follow from the following Lemmas F.1–F.4.

Lemma F.1. *Under Assumption 1, we have*

$$\mathrm{cov}(Z_i Z_j, Z_k Z_l) = O(n^{-1}).$$

Proof of Lemma F.1. Observe that

$$\mathrm{cov}(Z_i Z_j, Z_k Z_l) = \mathbb{E}Z_i Z_j Z_k Z_l - \mathbb{E}(Z_i Z_j)\mathbb{E}(Z_k Z_l).$$

First, we have

$$\begin{aligned}&\mathbb{E}Z_i Z_j Z_k Z_l = \mathbb{P}(Z_i = 1, Z_j = 1, Z_k = 1, Z_l = 1) \\ &= \mathbb{P}(Z_l = 1)\mathbb{P}(Z_k = 1|Z_l = 1)\mathbb{P}(Z_j = 1|Z_k = 1, Z_l = 1)\mathbb{P}(Z_i = 1|Z_j = 1, Z_k = 1, Z_l = 1) \\ &= \frac{n_1}{n}\frac{n_1 - 1}{n - 1}\frac{n_1 - 2}{n - 2}\frac{n_1 - 3}{n - 3}.\end{aligned}$$

On the other hand, we have

$$\mathbb{E}(Z_i Z_j)\mathbb{E}(Z_k Z_l) = \left(\frac{n_1}{n}\frac{n_1 - 1}{n - 1}\right)^2.$$

In sum, under Assumption 1, we have that

$$\begin{aligned}
\mathrm{cov}(Z_iZ_j, Z_kZ_l) &= \frac{n_1}{n}\frac{n_1-1}{n-1}\frac{n_1-2}{n-2}\frac{n_1-3}{n-3} - \left(\frac{n_1}{n}\frac{n_1-1}{n-1}\right)^2 \\
&= \frac{n_1^4}{n(n-1)(n-2)(n-3)} - \frac{n_1^4}{n^2(n-1)^2} + O(n^{-1}) \\
&= n_1^4\left(\frac{1}{n(n-1)(n-2)(n-3)} - \frac{1}{n^2(n-1)^2}\right) + O(n^{-1}) \\
&= n_1^4 O(n^{-5}) + O(n^{-1}) = O(n^{-1}).
\end{aligned}$$

□

Let $g_{(1)}^2 \geq \ldots \geq g_{(n)}^2$ and $y_{(1)}^2 \geq \ldots \geq y_{(n)}^2$ be the ordered sequence of $\{g_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$, respectively.

Lemma F.2. *Assume Assumption 1 holds and $\sum_i y_i^2 = O(n)$, $\sum_i g_i^2 = O(n)$, $\max_i g_i^2 = o(n)$, $\max_i y_i^2 = o(n)$. For any symmetric matrix $\boldsymbol{D}$ with diagonal entries being 0 and $\|\boldsymbol{D}\|_2 < C$, we have*

$$\frac{1}{n}\sum_{[i,j]} D_{ij}y_ig_jZ_iZ_j = \frac{1}{n}\sum_{[i,j]} D_{ij}y_ig_jr_1^2 + o_{\mathbb{P}}(1), \tag{27}$$

and

$$\frac{1}{n}\sum_{[i,j]} D_{ij}y_ig_j(1-Z_i)(1-Z_j) = \frac{1}{n}\sum_{[i,j]} D_{ij}y_ig_jr_0^2 + o_{\mathbb{P}}(1). \tag{28}$$

Proof of Lemma F.2. We only prove (27), and (28) follows immediately by replacing Z_i with $1-Z_i$.

For $i \neq j$, using $\mathbb{E}Z_iZ_j = r_1\frac{n_1-1}{n-1}$, we get

$$\mathbb{E}\frac{1}{n}\sum_{[i,j]} D_{ij}y_ig_jZ_iZ_j = \frac{1}{n}\sum_{[i,j]} D_{ij}y_ig_jr_1\frac{n_1-1}{n-1} = \left(1+O(n^{-1})\right)\frac{1}{n}\sum_{[i,j]} D_{ij}y_ig_jr_1^2.$$

On the other hand, we have

$$\left|\frac{1}{n}\sum_{[i,j]} D_{ij}y_ig_j\right| \leq \left(\sum_i y_i^2/n\right)^{1/2}\left(\sum_i g_i^2/n\right)^{1/2}\|\boldsymbol{D}\|_2 = O(1), \tag{29}$$

which implies that

$$\mathbb{E}\frac{1}{n}\sum_{[i,j]} D_{ij}y_ig_jZ_iZ_j = \frac{1}{n}\sum_{[i,j]} D_{ij}y_ig_jr_1^2 + o(1).$$

Thus, to conclude the proof using Chebyshev's inequality, it suffices to show that

$$\mathrm{var}\left(\sum_{[i,j]} y_ig_jD_{ij}Z_iZ_j\right) = o(n^2). \tag{30}$$

Through direct calculation, we get that

$$\begin{aligned}
&\operatorname{var}\left(\sum_{[i,j]} y_i g_j D_{ij} Z_i Z_j\right) = \operatorname{var}(Z_1 Z_2) \sum_{[i_1,i_2]} \left(C_1 y_{i_1}^2 g_{i_2}^2 D_{i_1 i_2}^2 + C_2 y_{i_1} y_{i_2} g_{i_1} g_{i_2} D_{i_1 i_2}^2\right) \\
&+ \operatorname{cov}(Z_1 Z_2, Z_1 Z_3) \sum_{[i_1 \ldots i_3]} \left(C_3 D_{i_1 i_2} D_{i_2 i_3} g_{i_2}^2 y_{i_1} y_{i_3} + C_4 D_{i_1 i_2} D_{i_2 i_3} g_{i_2} y_{i_2} y_{i_1} g_{i_3} + C_5 D_{i_1 i_2} D_{i_2 i_3} y_{i_2}^2 g_{i_1} g_{i_3}\right) \\
&+ \operatorname{cov}(Z_1 Z_2, Z_3 Z_4) \sum_{[i_1 \ldots i_4]} C_6 D_{i_1 i_2} D_{i_3 i_4} y_{i_1} g_{i_2} y_{i_3} g_{i_4} \\
&=: \operatorname{var}(Z_1 Z_2) \left(C_1 M_1 + C_2 M_2\right) + \operatorname{cov}(Z_1 Z_2, Z_1 Z_3) \left(C_3 M_3 + C_4 M_4 + C_5 M_5\right) \\
&\qquad + \operatorname{cov}(Z_1 Z_2, Z_3 Z_4) C_6 M_6,
\end{aligned}$$

where, C_i, $i = 1, \ldots, 6$ are universal constants that do not depend on n. By Assumption 1 and Lemma F.1, we have

$$\operatorname{var}(Z_1 Z_2) = O(1), \quad \operatorname{cov}(Z_1 Z_2, Z_1 Z_3) = O(1), \quad \operatorname{cov}(Z_1 Z_2, Z_3 Z_4) = O(n^{-1}).$$

It remains to estimate the order of M_i, $i = 1, \ldots, 6$.

By Lemma A.6, we have

$$M_1 = \operatorname{tr}\left(\operatorname{diag}(\boldsymbol{y})^2 \boldsymbol{D} \operatorname{diag}(\boldsymbol{g})^2 \boldsymbol{D}\right) \le C^2 \sum_i y_{(i)}^2 g_{(i)}^2 \le C^2 \left(\max_i g_i^2\right) \sum_i y_i^2 = o(n^2).$$

Applying the Cauchy-Schwarz inequality, we also get $|M_2| \le M_1 = o(n^2)$.

For M_3, let $y_{D,i}$ and $g_{D,i}$ be the i-th element of $\boldsymbol{D}(y_1, \ldots, y_n)^\top$ and $\boldsymbol{D}(g_1, \ldots, g_n)^\top$. We see that $\sum_i y_{D,i}^2 = O(n)$ and $\sum_i g_{D,i}^2 = O(n)$ since $\|D\|_2 < C$. By repeatedly applying $\sum_{j \in [n] \setminus i} D_{ij} y_j = y_{D,i}$ and $\sum_{j \in [n] \setminus i} D_{ij} g_j = g_{D,i}$, we obtain that

$$\begin{aligned}
M_3 &= \sum_{[i_1 \ldots i_3]} D_{i_1 i_2} D_{i_2 i_3} g_{i_2}^2 y_{i_1} y_{i_3} = \sum_{[i_1,i_2]} D_{i_1 i_2} g_{i_2}^2 y_{i_1} y_{D,i_2} - \sum_{[i_1,i_2]} D_{i_1 i_2} D_{i_2 i_1} g_{i_2}^2 y_{i_1} y_{i_1} \\
&= \sum_{i_1} g_{i_1}^2 y_{D,i_1}^2 - \sum_{[i_1,i_2]} D_{i_1 i_2}^2 g_{i_2}^2 y_{i_1}^2 =: M_{31} - M_{32}.
\end{aligned}$$

For M_{31}, it holds that

$$M_{31} \le \left(\max_i g_i^2\right) \sum_i y_{D,i}^2 = o(n^2).$$

For M_{32}, we see that $M_{32} = M_1 = o(n^2)$. Hence, there is $M_3 = o(n^2)$.

For M_4 and M_5, we decompose them as

$$\begin{aligned}
M_4 &= \sum_{[i_1,i_2]} D_{i_1 i_2} g_{i_2} g_{D,i_2} y_{i_1} y_{i_2} - \sum_{[i_1,i_2]} D_{i_1 i_2} D_{i_2 i_1} g_{i_2} g_{i_1} y_{i_1} y_{i_2} \\
&= \sum_{i_1} g_{i_1} g_{D,i_1} y_{D,i_1} y_{i_1} - \sum_{[i_1,i_2]} D_{i_1 i_2}^2 g_{i_2} g_{i_1} y_{i_1} y_{i_2}, \\
M_5 &= \sum_{[i_1,i_2]} D_{i_1 i_2} y_{i_2}^2 g_{i_1} g_{D,i_2} - \sum_{[i_1,i_2]} D_{i_1 i_2} D_{i_2 i_1} y_{i_2}^2 g_{i_1} g_{i_1} \\
&= \sum_{i_1} y_{i_1}^2 g_{D,i_1}^2 - \sum_{[i_1,i_2]} D_{i_1 i_2}^2 y_{i_2}^2 g_{i_1}^2.
\end{aligned}$$

Using similar arguments as in the analysis of M_3, we can show that $M_5 = o(n^2)$. For M_4, by Cauchy-Schwarz inequality, we have

$$\left(\sum_{i_1} g_{i_1} g_{D,i_1} y_{D,i_1} y_{i_1}\right)^2 \leq \left(\sum_{i_1} g_{i_1}^2 y_{D,i_1}^2\right)\left(\sum_{i_1} g_{D,i_1}^2 y_{i_1}^2\right),$$
$$\left(\sum_{[i_1,i_2]} D_{i_1 i_2}^2 g_{i_2} g_{i_1} y_{i_1} y_{i_2}\right)^2 \leq \left(\sum_{[i_1,i_2]} D_{i_1 i_2}^2 y_{i_2}^2 g_{i_1}^2\right)^2,$$

which, combined with the arguments in the analysis of M_3, yields that $M_4 = o(n^2)$.

Finally, for M_6, we see that

$$M_6 - \left(\sum_{[i_1,i_2]} D_{i_1 i_2} y_{i_1} g_{i_2}\right)^2 = O\left(|M_1| + |M_2| + |M_3| + |M_4| + |M_5|\right) = o(n^2),$$

and we have shown in (29) that

$$\frac{1}{n}\sum_{[i,j]} D_{ij} y_i g_j = O(1).$$

Thus, we have $M_6 = O(n^2)$.

Putting together the above estimates, we obtain that

$$\mathrm{var}\left(\sum_{[i,j]} y_i g_j D_{ij} Z_i Z_j\right) = O(1)o(n^2) + O(1)o(n^2) + O(n^{-1})O(n^2) = o(n^2),$$

which concludes the proof. □

Lemma F.3. *Assume Assumption 1 holds and* $\sum_i y_i^2 = O(n)$, $\sum_i g_i^2 = O(n)$, $\max_i g_i^2 = o(n)$. *For any sequence* $\{a_i\}_{i=1}^n$ *with* $\max_i |a_i| < C$, *we have that*

$$\frac{1}{n}\sum_i a_i y_i g_i Z_i = \frac{1}{n}\sum_i a_i y_i g_i r_1 + o_{\mathbb{P}}(1).$$

Proof of Lemma F.3. It suffices to show that

$$\mathrm{var}\left(\sum_i a_i y_i g_i Z_i\right) = o(n^2). \tag{31}$$

Direct calculations give that

$$\begin{aligned}\mathrm{var}\left(\sum_i a_i y_i g_i Z_i\right) &= \mathrm{var}(Z_1)\sum_i a_i^2 y_i^2 g_i^2 + \mathrm{cov}(Z_1, Z_2)\sum_{[i,j]} a_i a_j y_i y_j g_i g_j \\ &=: \mathrm{var}(Z_1) M_1 + \mathrm{cov}(Z_1, Z_2) M_2.\end{aligned}$$

Under Assumption 1, we have that

$$\mathrm{var}(Z_1) = O(1), \quad \mathrm{cov}(Z_1, Z_2) = O(n^{-1}).$$

It remains to estimate the order of M_1 and M_2.

For M_1, by $\max_i |a_i| < C$, $\max_i g_i^2 = o(n)$ and $\sum_i y_i^2 = O(n)$, we have

$$\sum_i a_i^2 y_i^2 g_i^2 \le C^2 \left(\max_i g_i^2 \right) \sum_i y_i^2 = o(n^2).$$

For M_2, we have that

$$M_2 = \sum_{[i,j]} a_i a_j y_i y_j g_i g_j \le \left(\sum_i a_i y_i g_i \right)^2.$$

By Cauchy-Schwarz inequality, there is

$$\left(\sum_i a_i y_i g_i \right)^2 \le \max_i a_i^2 \left(\sum_i y_i^2 \right) \left(\sum_i g_i^2 \right) = O(n^2),$$

which implies that $M_2 = O(n^2)$. Thus, we have

$$\operatorname{var}\left(\sum_i a_i y_i g_i Z_i \right) = O(1)o(n^2) + o(n^{-1})O(n^2) = o(n^2),$$

which concludes the proof. □

Lemma F.4. *Assume Assumptions 1-3 hold. For any symmetric matrix $\boldsymbol{D}$ with $\| \operatorname{diag}^-(\boldsymbol{D}) \|_2 < C$ and $\| \operatorname{diag}(\boldsymbol{D}) \|_2 < C$, we have that for $z \in \{0, 1\}$,*

$$s^2_{\operatorname{diag}^-(\boldsymbol{D}),Y(z)} = S^2_{\operatorname{diag}^-(\boldsymbol{D}),Y(z)} + o_{\mathbb{P}}(1), \quad s^2_{\operatorname{diag}(\boldsymbol{D}),Y(z)} = S^2_{\operatorname{diag}(\boldsymbol{D}),Y(z)} + o_{\mathbb{P}}(1),$$

$$s_{\operatorname{diag}^-(\boldsymbol{D}),Y(1),Y(0)} = S_{\operatorname{diag}^-(\boldsymbol{D}),Y(1),Y(0)} + o_{\mathbb{P}}(1).$$

Proof of Lemma F.4. We can write that

$$\begin{aligned}
s^2_{\operatorname{diag}^-(\boldsymbol{D}),Y(z)} &= \frac{1}{n r_z^2} \sum_{i \ne j : Z_i = z, Z_j = z} D_{ij} (Y_i(z) - \bar{Y}_z)(Y_j(z) - \bar{Y}_z) \\
&= M_1 + M_2 + M_3,
\end{aligned}$$

where

$$\begin{aligned}
M_1 &= \frac{1}{n r_z^2} \sum_{i \ne j : Z_i = z, Z_j = z} D_{ij} (Y_i(z) - \bar{Y}(z))(Y_j(z) - \bar{Y}(z)), \\
M_2 &= 2(\bar{Y}(z) - \bar{Y}_z) \frac{1}{n r_z^2} \sum_{i \ne j : Z_i = z, Z_j = z} D_{ij} (Y_i(z) - \bar{Y}(z)), \\
M_3 &= 2(\bar{Y}(z) - \bar{Y}_z)^2 \frac{1}{n r_z^2} \sum_{i \ne j : Z_i = z, Z_j = z} D_{ij}.
\end{aligned}$$

Applying Lemma F.2 with $f_i = g_i = Y_i(z) - \bar{Y}(z)$, we get

$$M_1 = \frac{1}{n} \sum_{i \ne j} D_{ij} (Y_i(z) - \bar{Y}(z))(Y_j(z) - \bar{Y}(z)) + o_{\mathbb{P}}(1).$$

Applying Lemma A.5 and Lemma F.2 with $f_i = Y_i(z) - \bar{Y}(z)$ and $g_i = 1$, we get

$$M_2 = 2(\bar{Y}(z) - \bar{Y}_z)O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

Applying Lemma A.5 and Lemma F.2 with $f_i = g_i = 1$, we get

$$M_2 = (\bar{Y}(z) - \bar{Y}_z)^2 O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

These results together imply that

$$\begin{aligned}
s^2_{\mathrm{diag}^-(\boldsymbol{D}),Y(z)} &= \frac{1}{n}\sum_{i\neq j} D_{ij}(Y_i(z) - \bar{Y}(z))(Y_j(z) - \bar{Y}(z)) + o_{\mathbb{P}}(1) \\
&= \left(1 + O(n^{-1})\right) S^2_{\mathrm{diag}^-(\boldsymbol{D}),Y(z)} + o_{\mathbb{P}}(1) = S^2_{\mathrm{diag}^-(\boldsymbol{D}),Y(z)} + o_{\mathbb{P}}(1).
\end{aligned}$$

For $s_{\mathrm{diag}^-(\boldsymbol{D}),Y(1),Y(0)}$, we have

$$s_{\mathrm{diag}^-(\boldsymbol{D}),Y(1),Y(0)} = \frac{1}{nr_1r_0}\sum_{i\neq j} D_{ij}(Y_i(1) - \bar{Y}_1)(Y_j(0) - \bar{Y}_0)Z_i(1 - Z_j) = M_4 + M_5,$$

where

$$\begin{aligned}
M_4 &= -\frac{1}{nr_1r_0}\sum_{i\neq j} D_{ij}(Y_i(1) - \bar{Y}_1)(Y_j(0) - \bar{Y}_0)Z_iZ_j, \\
M_5 &= \frac{1}{nr_1r_0}\sum_{i\neq j} D_{ij}(Y_i(1) - \bar{Y}_1)(Y_j(0) - \bar{Y}_0)Z_i.
\end{aligned}$$

Similarly, applying Lemma F.2, we get that

$$M_4 = -\frac{r_1}{nr_0}\sum_{i\neq j} D_{ij}(Y_i(1) - \bar{Y}_1)(Y_j(0) - \bar{Y}_0) + o_{\mathbb{P}}(1).$$

The term M_5 is decomposed as $M_5 = M_{51} + M_{52} + M_{53} + M_{54}$, where

$$\begin{aligned}
M_{51} &= \frac{1}{nr_1r_0}\sum_{i\neq j} D_{ij}(Y_i(1) - \bar{Y}(1))(Y_j(0) - \bar{Y}(0))Z_i, \\
M_{52} &= (\bar{Y}(1) - \bar{Y}_1)\frac{1}{nr_1r_0}\sum_{i\neq j} D_{ij}(Y_j(0) - \bar{Y}(0))Z_i, \\
M_{53} &= (\bar{Y}(0) - \bar{Y}_0)\frac{1}{nr_1r_0}\sum_{i\neq j} D_{ij}(Y_i(1) - \bar{Y}(1))Z_i, \\
M_{54} &= (\bar{Y}(0) - \bar{Y}_0)(\bar{Y}(1) - \bar{Y}_1)\frac{1}{nr_1r_0}\sum_{i\neq j} D_{ij}Z_i.
\end{aligned}$$

Applying Lemma F.3 with $y_i = \sum_{j\in[n]\setminus i} D_{ij}(Y_j(0) - \bar{Y}(0))$, $g_i = Y_i(1) - \bar{Y}(1)$, and $a_i = 1$, we get

$$\begin{aligned}
M_{51} &= \frac{1}{nr_1r_0}\sum_{i\neq j} D_{ij}(Y_i(1) - \bar{Y}(1))(Y_j(0) - \bar{Y}(0))Z_i \\
&= \frac{1}{nr_0}\sum_{i\neq j} D_{ij}(Y_i(1) - \bar{Y}(1))(Y_j(0) - \bar{Y}(0)) + o_{\mathbb{P}}(1).
\end{aligned}$$

Applying Lemma F.3 with $y_i = \sum_{j\in[n]\setminus i} D_{ij}(Y_j(0) - \bar{Y}(0))$, $g_i = 1$, and $a_i = 1$, we get

$$M_{52} = (\bar{Y}(1) - \bar{Y}_1)O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

Applying Lemma F.3 with $y_i = \sum_{j\in[n]\setminus i} D_{ij}$, $g_i = (Y_i(1) - \bar{Y}(1))$, and $a_i = 1$, we get

$$M_{53} = (\bar{Y}(0) - \bar{Y}_0)O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

Applying Lemma F.3 with $y_i = \sum_{j\in[n]\setminus i} D_{ij}$, $g_i = 1$, and $a_i = 1$, we get

$$M_{53} = (\bar{Y}(0) - \bar{Y}_0)(\bar{Y}(1) - \bar{Y}_1)O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

These results together imply that

$$\begin{aligned} s_{\mathrm{diag}^-(\boldsymbol{D}),Y(1),Y(0)} &= \frac{1}{n}\sum_{i\neq j} D_{ij}(Y_i(1) - \bar{Y}_1)(Y_j(0) - \bar{Y}_0) + o_{\mathbb{P}}(1) \\ &= S_{\mathrm{diag}^-(\boldsymbol{D}),Y(1),Y(0)} + o_{\mathbb{P}}(1). \end{aligned}$$

Finally, for $s^2_{\mathrm{diag}(\boldsymbol{D}),Y(z)}$, we have

$$s^2_{\mathrm{diag}(\boldsymbol{D}),Y(z)} = \frac{1}{n_z}\sum_{i:Z_i=z} D_{ii}(Y_i(z) - \bar{Y}_z)^2 = M_6 + M_7 + M_8,$$

where

$$\begin{aligned} M_6 &= \frac{1}{n_z}\sum_{i:Z_i=z} D_{ii}(Y_i(z) - \bar{Y}(z))^2, \\ M_7 &= 2(\bar{Y}(z) - \bar{Y}_z)\frac{1}{n_z}\sum_{i:Z_i=z} D_{ii}(Y_i(z) - \bar{Y}(z)), \\ M_8 &= (\bar{Y}(z) - \bar{Y}_z)^2\frac{1}{n_z}\sum_{i:Z_i=z} D_{ii}. \end{aligned}$$

Applying Lemma F.3 with $y_i = g_i = Y_i(z) - \bar{Y}(z)$ and $a_i = D_{ii}$, we get

$$M_6 = \frac{1}{n}\sum_{i:Z_i=z} D_{ii}(Y_i(z) - \bar{Y}(z))^2 + o_{\mathbb{P}}(1).$$

Applying Lemma A.5 and Lemma F.3 with $y_i = 1$, $g_i = Y_i(z) - \bar{Y}(z)$, and $a_i = D_{ii}$, we get

$$M_7 = 2(\bar{Y}(z) - \bar{Y}_z)O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

Applying Lemma A.5 and Lemma F.3 with $y_i = 1$, $g_i = 1$, and $a_i = D_{ii}$, we get

$$M_8 = (\bar{Y}(z) - \bar{Y}_z)^2 O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

These results together imply that

$$s^2_{\mathrm{diag}(\boldsymbol{D}),Y(z)} = \frac{1}{n}\sum_{i} D_{ii}(Y_i(z) - \bar{Y}_z)^2 + o_{\mathbb{P}}(1) = S^2_{\mathrm{diag}(\boldsymbol{D}),Y(z)} + o_{\mathbb{P}}(1).$$

To sum up, we have concluded the proof. $\square$

Now, we are ready to prove Theorem 3. The also proof includes the technical details of the comment of (13).

Proof of Theorem 3. Recall that we denote $\boldsymbol{P} = \boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^\top$. By Lemma A.7 and the fact $0 \le H_{ii} \le 1$, there is

$$\|\operatorname{diag}(\star)\|_2 = O(1), \quad \|\operatorname{diag}^-(\star)\|_2 = O(1), \quad \star \in \{\boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{P}\}.$$

We then expand $\boldsymbol{B}$ as

$$\begin{aligned}\boldsymbol{B} &= (\boldsymbol{P} - \boldsymbol{H} + \boldsymbol{P}\operatorname{diag}(\boldsymbol{H}))^\top (\boldsymbol{P} - \boldsymbol{H} + \boldsymbol{P}\operatorname{diag}(\boldsymbol{H}))\\ &= \boldsymbol{P} - \boldsymbol{H} + \operatorname{diag}(\boldsymbol{H})\boldsymbol{P}\operatorname{diag}(\boldsymbol{H}) + (\boldsymbol{P} - \boldsymbol{H})\operatorname{diag}(\boldsymbol{H}) + \operatorname{diag}(\boldsymbol{H})(\boldsymbol{P} - \boldsymbol{H}).\end{aligned}$$

Therefore, we have

$$\begin{aligned}\operatorname{diag}(\boldsymbol{B}) &= \operatorname{diag}(\boldsymbol{P}) - \operatorname{diag}(\boldsymbol{H}) + \operatorname{diag}(\boldsymbol{H})\operatorname{diag}(\boldsymbol{P})\operatorname{diag}(\boldsymbol{H})\\ &\quad + (\operatorname{diag}(\boldsymbol{P}) - \operatorname{diag}(\boldsymbol{H}))\operatorname{diag}(\boldsymbol{H}) + \operatorname{diag}(\boldsymbol{H})(\operatorname{diag}(\boldsymbol{P}) - \operatorname{diag}(\boldsymbol{H})),\\ \operatorname{diag}^-(\boldsymbol{B}) &= \operatorname{diag}^-(\boldsymbol{P}) - \operatorname{diag}^-(\boldsymbol{H}) + \operatorname{diag}(\boldsymbol{H})\operatorname{diag}^-(\boldsymbol{P})\operatorname{diag}(\boldsymbol{H})\\ &\quad + (\operatorname{diag}^-(\boldsymbol{P}) - \operatorname{diag}^-(\boldsymbol{H}))\operatorname{diag}(\boldsymbol{H}) + \operatorname{diag}(\boldsymbol{H})(\operatorname{diag}^-(\boldsymbol{P}) - \operatorname{diag}^-(\boldsymbol{H})).\end{aligned}$$

By sub-additivity and sub-multiplicativity of the l_2-norm, we have

$$\|\operatorname{diag}(\boldsymbol{B})\|_2 = O(1), \quad \|\operatorname{diag}^-(\boldsymbol{B})\|_2 = O(1).$$

Therefore, for $\star \in \{\boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{B}, \boldsymbol{P}\}$, under Assumptions 1-3, we can derive that

$$\begin{aligned}&s^2_{\operatorname{diag}(\star),Y(z)} = S^2_{\operatorname{diag}(\star),Y(z)} + o_{\mathbb{P}}(1) \quad s^2_{\operatorname{diag}^-(\star),Y(z)} = S^2_{\operatorname{diag}^-(\star),Y(z)} + o_{\mathbb{P}}(1),\\ &s_{\operatorname{diag}^-(\star),Y(1),Y(0)} = S_{\operatorname{diag}^-(\star),Y(1),Y(0)} + o_{\mathbb{P}}(1).\end{aligned} \tag{32}$$

by using Lemma F.4. Thus, we have proved the consistency of those empirical estimators of covariances.

Next, we prove that

$$\begin{aligned}\mathcal{I}_3 = &\sum_{z\in\{0,1\}} \left(S^2_{\operatorname{diag}(\boldsymbol{B}),Y(z)} - S^2_{\operatorname{diag}(\boldsymbol{Q}),Y(z)} - S^2_{\operatorname{diag}^-(\boldsymbol{H}),Y(z)}\right)\\ &+ 2S_{\operatorname{diag}^-(\boldsymbol{H}),Y(1),Y(0)} - S^2_{\operatorname{diag}(\boldsymbol{H}),Y(1)-Y(0)} - S^2_{e(1)-e(0)} + O\left(n^{-1}\right).\end{aligned} \tag{33}$$

Some direct calculations give that

$$B_{ii} = 1 - \frac{1}{n} + \left(1 - \frac{2}{n}\right) H_{ii} - \left(1 + \frac{1}{n}\right) H_{ii}^2,$$

which implies that $B_{ii} - Q_{ii} = 1 + O(n^{-1})$.

Applying the equation

$$2S_{\operatorname{diag}(\boldsymbol{D}),Y(1),Y(0)} = \sum_{z\in\{0,1\}} S^2_{\operatorname{diag}(\boldsymbol{D}),Y(z)} - S^2_{\operatorname{diag}(\boldsymbol{D}),Y(1)-Y(0)},$$

with $\boldsymbol{D} \in \{\boldsymbol{B}, \boldsymbol{Q}, \boldsymbol{H}\}$ and the equation

$$S^2_{Y(1)-Y(0)} = S^2_{\boldsymbol{H},Y(1)-Y(0)} + S^2_{e(1)-e(0)},$$

we obtain that

$$\begin{aligned}
\mathcal{I}_3 &= 2S_{\mathrm{diag}(\boldsymbol{B}),Y(1),Y(0)} - 2S_{\mathrm{diag}(\boldsymbol{Q}),Y(1),Y(0)} \\
&= \sum_{z\in\{0,1\}} \left(S^2_{\mathrm{diag}(\boldsymbol{B}),Y(z)} - S^2_{\mathrm{diag}(\boldsymbol{Q}),Y(z)}\right) - S^2_{\mathrm{diag}(\boldsymbol{B}),Y(1)-Y(0)} + S^2_{\mathrm{diag}(\boldsymbol{Q}),Y(1)-Y(0)} \\
&= \sum_{z\in\{0,1\}} \left(S^2_{\mathrm{diag}(\boldsymbol{B}),Y(z)} - S^2_{\mathrm{diag}(\boldsymbol{Q}),Y(z)}\right) - S^2_{Y(1)-Y(0)} + O\left(n^{-1}\right) \\
&= \sum_{z\in\{0,1\}} \left(S^2_{\mathrm{diag}(\boldsymbol{B}),Y(z)} - S^2_{\mathrm{diag}(\boldsymbol{Q}),Y(z)}\right) - S^2_{\boldsymbol{H},Y(1)-Y(0)} - S^2_{e(1)-e(0)} + O\left(n^{-1}\right) \\
&= \sum_{z\in\{0,1\}} \left(S^2_{\mathrm{diag}\{\boldsymbol{B}\},Y(z)} - S^2_{\mathrm{diag}\{\boldsymbol{Q}\},Y(z)} - S^2_{\mathrm{diag}^-\{\boldsymbol{H}\},Y(z)}\right) \\
&\quad + 2S_{\mathrm{diag}^-\{\boldsymbol{H}\},Y(1),Y(0)} - S^2_{\mathrm{diag}\{\boldsymbol{H}\},Y(1)-Y(0)} - S^2_{e(1)-e(0)} + O\left(n^{-1}\right).
\end{aligned}$$

We replace all terms in the formula of σ^2_{hd} with their empirical estimators, except for the term

$$-S^2_{\mathrm{diag}\{\boldsymbol{H}\},Y(1)-Y(0)} - S^2_{e(1)-e(0)},$$

which constitutes the bias of $\hat{\sigma}^2_{\mathrm{hd}}$.

Using (32), we get that under Assumptions 1-3,

$$\hat{\sigma}^2_{\mathrm{hd}} = \sigma^2_{\mathrm{hd}} + S^2_{\mathrm{diag}\{\boldsymbol{H}\},Y(1)-Y(0)} + S^2_{e(1)-e(0)} + o_{\mathbb{P}}(1).$$

It remains to show that if Assumption 5 holds, we have

$$\sigma^2_{\mathrm{hd}} + S^2_{\mathrm{diag}\{\boldsymbol{H}\},Y(1)-Y(0)} + S^2_{e(1)-e(0)} = \sigma^2_{\mathrm{adj}} + S^2_{e(1)-e(0)} + o(1).$$

Comparing the left-hand and right-hand sides of the above formula with the formulas of σ^2_{hd} and σ^2_{adj}, we find that it suffices to prove that

$$S^2_{\mathrm{diag}\{\boldsymbol{H}\},Y(1)-Y(0)} = o(1), \quad S^2_{s(z)} = o(1).$$

Under Assumption 5, using $\sum_i H_{ii} = p$, we obtain that

$$\begin{aligned}
S^2_{\mathrm{diag}\{\boldsymbol{H}\},Y(1)-Y(0)} &\le \frac{2}{n-1}\sum_i H_{ii}(Y_i(1)-\bar{Y}(1))^2 + \frac{2}{n-1}\sum_i H_{ii}(Y_i(0)-\bar{Y}(0))^2 \\
&\le \frac{2}{n-1}\sum_{i=1}^{p}(Y_{(i)}(1)-\bar{Y}(1))^2 + \frac{2}{n-1}\sum_{i=1}^{p}(Y_{(i)}(0)-\bar{Y}(0))^2 = o(1).
\end{aligned}$$

On the other hand, using $\sum_i H^2_{ii} \le \sum_i H_{ii} = p$, we obtain that

$$S^2_{s(z)} \le \frac{1}{n-1}\sum_i H^2_{ii}(Y_i(z)-\bar{Y}(z))^2 \le \frac{1}{n-1}\sum_{i=1}^{p}(Y_{(i)}(z)-\bar{Y}(z))^2 = o(1).$$

The conclusion then follows. □

Proof of Corollary 2. In the proof of Theorem 3, we have derived that (recall (33))

$$\mathcal{I}_3 = \sum_{z\in\{0,1\}} \left(S^2_{\mathrm{diag}(\boldsymbol{B}),Y(z)} - S^2_{\mathrm{diag}(\boldsymbol{Q}),Y(z)}\right) - S^2_{Y(1)-Y(0)} + O\left(n^{-1}\right).$$

Then, the conclusion follows by replacing all terms with their empirical estimators except for $S^2_{Y(1)-Y(0)}$ and by using a proof similar to that of Theorem 3. □

G Justification of assumptions

In this section, we prove Propositions 1–3, which provide some justifications of our assumptions. For the proof of Proposition 1, we will use the classical Bernstein inequality.

Lemma G.1 (Bernstein inequality)**.** *Let $X_1, \dots, X_n$ be independent centered random variables. Suppose that $|X_i| \le M$ almost surely for all i. Then, for all $t > 0$, we have that*

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \ge t\right) \le \exp\left(-\frac{\frac{1}{2}t^2}{\sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] + \frac{1}{3}Mt}\right).$$

Proof of Proposition 1. Fix $z \in \{0, 1\}$. For ease of presentation, we denote $Y_i(z) - \mathbb{E}Y_i(z)$ by U_i and $Y_{(i)}(z) - \mathbb{E}Y_i(z)$ by $U_{(i)}$. By definition, $(U_{(1)} - \bar{U})^2 \ge (U_{(2)} - \bar{U})^2 \ge \dots \ge (U_{(n)} - \bar{U})^2$. We further define $U^2_{<1>} \ge \dots \ge U^2_{<n>}$ as the ordered sequence of $\{U_i^2\}_{i=1}^n$. Then, we have that

$$\sum_{i=1}^{p} (Y_{(i)}(z) - \bar{Y}(z))^2 = \sum_{i=1}^{p} (U_{(i)} - \bar{U})^2 \le 2p\bar{U}^2 + 2\sum_{i=1}^{p} U^2_{(i)} \le 2p\bar{U}^2 + 2\sum_{i=1}^{p} U^2_{<i>}.$$

Since $\mathbb{E}\bar{U} = 0$ and $\mathrm{var}(\bar{U}) = \mathrm{var}(U_1)/n = O(n^{-1})$, by Chebyshev's inequality, we have that

$$\mathbb{P}(p\bar{U}^2 \ge c_{n1}) \le \frac{\mathrm{var}(\bar{U})p}{c_{n1}} = \mathrm{var}(U_1)\frac{p}{nc_{n1}}.$$

Thus, choosing $c_{n1} = (p/n)^{1/2} = o(1)$, we get that with probability $1 - (p/n)^{1/2} = 1 - o(1)$,

$$p\bar{U}^2 < c_{n1}.$$

It remains to show that there exists $c_{n2} \to 0$ such that

$$\mathbb{P}\left(\sum_{i=1}^{p} U^2_{<i>} \ge c_{n2}\right) = o(1).$$

Note $\sum_{i=1}^{p} U^2_{<i>}$ is increasing in p, so in the following proof, we assume that $p \to \infty$ without loss of generality.

Now, we consider the following two cases for the distribution of U_1^2:

(1) U_1^2 is bounded almost surely, i.e., there exists an $M > 0$ such that

$$\mathbb{P}(U_1^2 \ge M) = 0.$$

(2) U_1^2 is unbounded, i.e., for any $M > 0$, we have

$$\mathbb{P}(U_1^2 \ge M) > 0.$$

In case (1), we have that almost surely,

$$\frac{1}{n}\sum_{i=1}^{p} U^2_{<i>} < pM/n = o(1),$$

in which case we can choose $c_{n2} = pM/n$.

On the other hand, suppose case (2) holds. Then, we define the upper quantiles of U_1 as

$$Q_a := \sup\{M \in \mathbb{R} \mid \mathbb{P}(U_1^2 \geq M) \geq a\}, \quad a > 0.$$

By definition, $\mathbb{P}(U_1^2 \geq Q_a) \geq a$ and $Q_a \to \infty$ as $a \to 0$. For any $c_{n2} > 0$ and $\alpha = p/n$, we have

$$\begin{aligned}
&\mathbb{P}\left(\sum_{i=1}^{p} U_{<i>}^2/n \geq c_{n2}\right) \\
&\leq \mathbb{P}\left(\sum_i I(U_i^2 \geq Q_{2\alpha}) < p\right) + \mathbb{P}\left(\sum_{i=1}^{p} U_{<i>}^2/n \geq c_{n2}, \sum_i I(U_i^2 \geq Q_{2\alpha}) \geq p\right) \\
&\leq \mathbb{P}\left(\sum_i I(U_i^2 \geq Q_{2\alpha}) < p\right) + \mathbb{P}\left(\sum_{i=1}^{n} U_i^2 I(U_i^2 \geq Q_{2\alpha})/n \geq c_{n2}\right) \\
&=: \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2).
\end{aligned}$$

We next deal with the events $\mathcal{E}_1$ and $\mathcal{E}_2$, respectively.

For $\mathcal{E}_1$, let $e := \mathbb{P}(U_i^2 \geq Q_{2\alpha}) \geq 2\alpha$. Then, we apply Bernstein's inequality with $X_i = e - I(U_i^2 \geq Q_{2\alpha})$, $t = ne/2$, $\mathbb{E}X_i^2 < e$, and $|X_i| < 2$ to get that

$$\sum_i I(U_i^2 \geq Q_{2\alpha}) \geq ne/2 \geq n\alpha = p$$

holds with probability at least

$$1 - \exp\left(-\frac{3}{32}ne\right) \geq 1 - \exp\left(-\frac{3p}{16}\right) = 1 - o(1).$$

This implies that $\mathbb{P}(\mathcal{E}_1) = o(1)$.

For $\mathcal{E}_2$, using Markov's inequality, we get that

$$\mathbb{P}\left(\sum_{i=1}^{n} U_i^2 I(U_i^2 \geq Q_{2\alpha})/n \geq c_{n2}\right) \leq \frac{1}{c_{n2}} \mathbb{E}U_i^2 I(U_i^2 \geq Q_{2\alpha}).$$

Since $Q_{2\alpha} \to \infty$ as $\alpha \to 0$ and $\mathbb{E}U_i^2 < \infty$, we have $\mathbb{E}U_i^2 I(U_i^2 \geq Q_{2\alpha}) \to 0$. Thus, we can choose $c_{n2} = \left[\mathbb{E}U_i^2 I(U_i^2 \geq Q_{2\alpha})\right]^{1/2} = o(1)$ such that $\mathbb{P}(\mathcal{E}_2) = c_{n2} \to 0$.

In sum, we have proved that with probability $1 - o(1)$,

$$\sum_{i=1}^{p} (Y_{(i)}(z) - \bar{Y}(z))^2 < 2c_{n1} + 2c_{n2} = o(1).$$

Hence, the conclusion follows. □

Proof for Corollary 1. For simplicity of notations, we denote by $\tilde{Y}_i(z) := Y_i(z) - \bar{Y}(z)$. By definition and Proposition F.1, we have

$$\sigma^2_{\mathrm{hd},l} = \frac{1}{(n-1)(r_1 r_0)} \sum_i \left(r_0 s_i(1) + r_0 e_i(1) + r_1 s_i(0) + r_1 e_i(0)\right)^2, \tag{34}$$

$$\begin{aligned} \sigma^2_{\mathrm{hd},q} = & \frac{(r_1 r_0)^2}{n-1} \sum_{[i,j]} H_{ij}^2 \left(\frac{\tilde{Y}_i(1)}{r_1^2} - \frac{\tilde{Y}_i(0)}{r_0^2} \right) \left(\frac{\tilde{Y}_j(1)}{r_1^2} - \frac{\tilde{Y}_j(0)}{r_0^2} \right) \\ & + \frac{(r_1 r_0)^2}{n-1} \sum_i (H_{ii} - H_{ii}^2) \left(\frac{\tilde{Y}_i(1)}{r_1^2} - \frac{\tilde{Y}_i(0)}{r_0^2} \right)^2. \end{aligned} \tag{35}$$

Note $\sigma^2_{\mathrm{hd},l} \geq 0$ and $\sigma^2_{\mathrm{hd},q} \geq 0$, since both of them are variances of certain random variables. We first prove that under Assumption 7,

$$\sum_i \left(s_i(z) - \alpha \tilde{Y}_i(z) \right)^2 = o(n). \tag{36}$$

Applying the inequality $\sum_i (a_i - \bar{a})^2 \leq \sum_i a_i^2$ with $a_i = (H_{ii} - \alpha)\tilde{Y}_i(z)$, we get that

$$\sum_i \left(s_i(z) - \alpha \tilde{Y}_i(z) \right)^2 = \sum_i (a_i - \bar{a})^2 \leq \sum_i a_i^2 = \sum_i (H_{ii} - \alpha)^2 \tilde{Y}_i(z)^2,$$

where the right-hand side is bounded by

$$\sum_i (H_{ii} - \alpha)^2 \tilde{Y}_i(z)^2 \leq \max_i |H_{ii} - \alpha|^2 \cdot \sum_i \tilde{Y}_i(z)^2 = o(n)$$

when $\max_i |H_{ii} - \alpha| = o(1)$ and $\sum_i \tilde{Y}_i(z)^2 = O(n)$, or bounded by

$$\begin{aligned} \sum_i (H_{ii} - \alpha)^2 \tilde{Y}_i(z)^2 & \leq \left(\sum_i |\tilde{Y}_i(z)|^{2+\eta} \right)^{\frac{2}{2+\eta}} \left(\sum_i |H_{ii} - \alpha|^{\frac{2(2+\eta)}{\eta}} \right)^{\frac{\eta}{2+\eta}} \\ & < \left(\sum_i |\tilde{Y}_i(z)|^{2+\eta} \right)^{\frac{2}{2+\eta}} \left(\sum_i |H_{ii} - \alpha|^2 \right)^{\frac{\eta}{2+\eta}} = o(n) \end{aligned} \tag{37}$$

when $\sum_i |H_{ii} - \alpha|^2 = o(n)$ and $\sum_i |\tilde{Y}_i(z)|^{2+\eta} = O(n)$. In the derivation of (37), the first inequality uses Hölder's inequality, while the second inequality is due to $\max_i |H_{ii} - \alpha| < 1$. In either case, we have proved (36), which implies that replacing $s_i(z)$ with $\alpha \tilde{Y}_i(z)$ in the formula of $\sigma^2_{\mathrm{hd},l}$ leads to a negligible difference, i.e.,

$$\sigma^2_{\mathrm{hd},l} = \frac{1}{(n-1)(r_1 r_0)} \sum_i \left\{ r_0 \alpha \tilde{Y}_i(1) + r_0 e_i(1) + r_1 \alpha \tilde{Y}_i(0) + r_1 e_i(0) \right\}^2 + o(1).$$

We denote $d_i(z) := \tilde{Y}_i(z) - e_i(z)$. Notice that $\sum_i e_i(z) d_i(z') = 0$ for $z, z' \in \{0, 1\}$ and

$$\begin{aligned} R^2 \sum_i \left(r_0 \tilde{Y}_i(1) + r_1 \tilde{Y}_i(0) \right)^2 &= \sum_i \left(r_0 d_i(1) + r_1 d_i(0) \right)^2, \\ (1 - R^2) \sum_i \left(r_0 \tilde{Y}_i(1) + r_1 \tilde{Y}_i(0) \right)^2 &= \sum_i \left(r_0 e_i(1) + r_1 e_i(0) \right)^2, \\ \sigma^2_{\mathrm{cre}} &= \frac{1}{(n-1)(r_1 r_0)} \sum_i \left(r_0 \tilde{Y}_i(1) + r_1 \tilde{Y}_i(0) \right)^2. \end{aligned}$$

With these identities, we derive that

$$\begin{aligned}
\sigma^2_{\text{hd},l} &= \frac{1}{(n-1)(r_1 r_0)} \sum_i \left(r_0(1+\alpha)e_i(1) + r_1(1+\alpha)e_i(0) + r_0\alpha d_i(1) + r_1\alpha d_i(0)\right)^2 + o(1) \\
&= \frac{1}{(n-1)(r_1 r_0)} \left[(1+\alpha)^2 \sum_i \left(r_0 e_i(1) + r_1 e_i(0)\right)^2 + \alpha^2 \sum_i \left(r_0 d_i(1) + r_1 d_i(0)\right)^2\right] + o(1) \\
&= \left[(1+\alpha)^2 - (1+2\alpha)R^2\right] \sigma^2_{\text{cre}} + o(1).
\end{aligned}$$

Therefore, we have

$$\sigma^2_{\text{hd}} \geq \sigma^2_{\text{hd},l} = \left[(1+\alpha)^2 - (1+2\alpha)R^2\right] \sigma^2_{\text{cre}} + o(1), \tag{38}$$

which gives the lower bound on σ^2_{hd}.

For the upper bounds on σ^2_{hd} and $\sigma^2_{\text{hd},q}$, applying the Cauchy-Schwarz inequality and the identity $\sum_j H^2_{ij} = (\mathbf{H}^2)_{ii} = H_{ii}$, we get that

$$\begin{aligned}
\sum_{[i,j]} H^2_{ij} \left(\frac{\tilde{Y}_i(1)}{r_1^2} - \frac{\tilde{Y}_i(0)}{r_0^2}\right)\left(\frac{\tilde{Y}_j(1)}{r_1^2} - \frac{\tilde{Y}_j(0)}{r_0^2}\right) &\leq \sum_{[i,j]} H^2_{ij} \left(\frac{\tilde{Y}_i(1)}{r_1^2} - \frac{\tilde{Y}_i(0)}{r_0^2}\right)^2 \\
&= \sum_i (H_{ii} - H^2_{ii}) \left(\frac{\tilde{Y}_i(1)}{r_1^2} - \frac{\tilde{Y}_i(0)}{r_0^2}\right)^2.
\end{aligned}$$

Plugging it into (35) yields that

$$\sigma^2_{\text{hd},q} \leq \frac{2(r_1 r_0)^2}{n-1} \sum_i (H_{ii} - H^2_{ii}) \left(\frac{\tilde{Y}_i(1)}{r_1^2} - \frac{\tilde{Y}_i(0)}{r_0^2}\right)^2. \tag{39}$$

Now, the upper bounds on σ^2_{hd} and $\sigma^2_{\text{hd},q}$ follows immediately from the estimates

$$\sum_i (H_{ii} - \alpha) \left(\frac{\tilde{Y}_i(1)}{r_1^2} - \frac{\tilde{Y}_i(0)}{r_0^2}\right)^2 = o(n), \quad \sum_i (H^2_{ii} - \alpha^2) \left(\frac{\tilde{Y}_i(1)}{r_1^2} - \frac{\tilde{Y}_i(0)}{r_0^2}\right)^2 = o(n).$$

It suffices to prove that under Assumption 7,

$$M_1 := \sum_i |H_{ii} - \alpha| \tilde{Y}_i(z)^2 = o(n), \quad M_2 := \sum_i |H^2_{ii} - \alpha^2| \tilde{Y}_i(z)^2 = o(n). \tag{40}$$

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

When $\max_i |H_{ii} - \alpha| = o(1)$ and $\sum_i \tilde{Y}_i(z)^2 = O(n)$, M_1 is bounded as

$$\max_i |H_{ii} - \alpha| \cdot \sum_i \tilde{Y}_i(z)^2 = o(n).$$

When $\sum_i |H_{ii} - \alpha|^2 = o(n)$ and $\sum_i |\tilde{Y}_i(z)|^{2+\eta} = O(n)$, M_1 is bounded as

$$
\begin{aligned}
&\left(\sum_i |\tilde{Y}_i(z)|^{2+\eta'}\right)^{\frac{2}{2+\eta'}} \left(\sum_i |H_{ii} - \alpha|^{\frac{2+\eta'}{\eta'}}\right)^{\frac{\eta'}{2+\eta'}} \\
\leq &\left(\sum_i |\tilde{Y}_i(z)|^{2+\eta'}\right)^{\frac{2}{2+\eta'}} \left(\sum_i |H_{ii} - \alpha|^2\right)^{\frac{\eta'}{2+\eta'}} \\
\leq &\, n^{\frac{2}{2+\eta'}} \left(\sum_i |\tilde{Y}_i(z)|^{2+\eta}/n\right)^{\frac{2}{2+\eta}} \left(\sum_i |H_{ii} - \alpha|^2\right)^{\frac{\eta'}{2+\eta'}} = o(n),
\end{aligned}
$$

by using Hölder's inequality in the first two steps, where $\eta' \in (0, \eta)$ is chosen to be a small constant such that $(2+\eta')/\eta' > 2$. To sum up, under Assumption 7, we have $M_1 = o(n)$. The bound on M_2 then follows easily:

$$
M_2 \leq \max_i |H_{ii} + \alpha| \cdot \sum_i |H_{ii} - \alpha| \tilde{Y}_i(z)^2 \leq 2M_1 = o(n).
$$

Finally, using (39) and (40), we obtain

$$
\begin{aligned}
\sigma^2_{\mathrm{hd},q} &\leq \frac{2(r_1 r_0)^2}{n-1} \sum_i (\alpha - \alpha^2) \left(\frac{\tilde{Y}_i(1)}{r_1^2} - \frac{\tilde{Y}_i(0)}{r_0^2}\right)^2 + o(1) \\
&= 2(r_1 r_0)^2 \alpha(1-\alpha) S^2_{r_1^{-2} Y(1) - r_0^{-2} Y(0)} + o(1).
\end{aligned}
$$

Together with (38), it concludes the proof. □

For the proof of Proposition 3, we need to use the following lemma, which is an i.i.d. version of Theorem 2 in Whittle [1960].

Lemma G.2. *Let $\xi = (\xi_1, \ldots, \xi_n)$ be a random vector with centered i.i.d. entries. Let $\boldsymbol{A}$ be an arbitrary deterministic matrix. For any $s \geq 2$, there exists a constant $C(s)$ depending on s such that*

$$
\mathbb{E}\left|\boldsymbol{\xi}^\top \boldsymbol{A} \boldsymbol{\xi} - \mathbb{E}(\boldsymbol{\xi}^\top \boldsymbol{A} \boldsymbol{\xi})\right|^s \leq C(s) \left(\mathbb{E}|\boldsymbol{\xi}_1|^{2s}\right)^{1/2} \left(\sum_{i,j} |A_{ij}|^2\right)^{s/2}.
$$

The next lemma follows from a simple calculation.

Lemma G.3. *Let $\xi = (\xi_1, \ldots, \xi_n)$ be a random vector with centered i.i.d. entries. Let $\boldsymbol{A}$ be an arbitrary deterministic matrix. Then, we have*

$$
\mathbb{E}(\boldsymbol{\xi}^\top \boldsymbol{A} \boldsymbol{\xi}) = \mathrm{tr}(A) \mathbb{E}\xi_1^2.
$$

Proof of Lemma G.3. By the mean zero and i.i.d. conditions for the entries of ξ, we have

$$
\mathbb{E}(\boldsymbol{\xi}^\top \boldsymbol{A} \boldsymbol{\xi}) = \sum_{[i,j]} \mathbb{E}(\xi_i \xi_j A_{ij}) + \sum_i \mathbb{E}(\xi_i^2 A_{ii}) = 0 + (\mathbb{E}\xi_1^2) \sum_i \mathbb{E} A_{ii} = \mathrm{tr}(A) \mathbb{E}\xi_1^2.
$$

This concludes the proof. □

Proof of Proposition 3. We observe that

$$S_{\tau}^2 - S_{e(1)-e(0)}^2 - S_{\mathrm{diag}\{\boldsymbol{H}\},Y(1)-Y(0)}^2 = S_{\boldsymbol{H},\tau}^2 - S_{\mathrm{diag}\{\boldsymbol{H}\},Y(1)-Y(0)}^2. \tag{41}$$

Through a direct calculation, we can write $S_{\boldsymbol{H},\tau}^2$ as

$$\begin{aligned} S_{\boldsymbol{H},\tau}^2 = \frac{1}{n-1}\Big\{ &(\boldsymbol{\varepsilon}(1)-\boldsymbol{\varepsilon}(0))^\top \boldsymbol{H} (\boldsymbol{\varepsilon}(1)-\boldsymbol{\varepsilon}(0)) + (\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)^\top \boldsymbol{X}^\top \boldsymbol{P}\boldsymbol{X} (\boldsymbol{\beta}_1-\boldsymbol{\beta}_0) \\ &+ 2(\boldsymbol{\varepsilon}(1)-\boldsymbol{\varepsilon}(0))^\top \boldsymbol{P}\boldsymbol{X}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0) \\ =:\ & M_1 + M_2 + 2M_3, \end{aligned}$$

where we denote $\boldsymbol{\varepsilon}(z) = (\varepsilon_1(z), \ldots, \varepsilon_n(z))^\top$, $\boldsymbol{P} = \boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$, and $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)^\top$. We next estimate the terms M_i, $i = 1,2,3$, one by one.

For M_1, applying Lemmas G.2 and G.3 with $s = 2$, $\boldsymbol{A} = \boldsymbol{H}/(n-1)$, and $\xi_i = \varepsilon_i(1) - \varepsilon_i(0)$, and using the independence between $\boldsymbol{H}$ and $\boldsymbol{\varepsilon}(z)$, we obtain that

$$\begin{aligned} \mathbb{E}(M_1|\boldsymbol{H}) = \mathbb{E}M_1 = \frac{\mathrm{tr}(\boldsymbol{H})}{n-1}\mathrm{var}(\varepsilon_1(1)-\varepsilon_1(0)) &= \frac{p}{n-1}\left(\sigma_{\varepsilon(1)}^2 + \sigma_{\varepsilon(0)}^2\right) \\ &= \alpha\left(\sigma_{\varepsilon(1)}^2 + \sigma_{\varepsilon(0)}^2\right) + O(n^{-1}), \end{aligned}$$

with $\sigma_{\varepsilon(z)}^2$, $z \in \{0,1\}$, denoting the variance of $\varepsilon_1(z)$, and that

$$\begin{aligned} \mathbb{E}\big(|M_1 - \mathbb{E}(M_1|\boldsymbol{H})|^2 \,\big|\boldsymbol{H}\big) &\le C(2)\left(\mathbb{E}\xi_1^4\right)^{1/2}\frac{1}{(n-1)^2}\sum_{i,j} H_{ij}^2 \\ &= C(2)\left(\mathbb{E}\xi_1^4\right)^{1/2}\frac{p}{(n-1)^2} = O(n^{-1}). \end{aligned}$$

Thus, by choosing $c_{n1} = \left[C(2)(\mathbb{E}\xi_1^4)^{1/2}p(n-1)^{-2}\right]^{1/3} = o(1)$, we have

$$\begin{aligned} \mathbb{P}\big(|M_1 - \mathbb{E}M_1| \ge c_{n1}\big|\boldsymbol{H}\big) &= \mathbb{P}\big(|M_1 - \mathbb{E}(M_1|\boldsymbol{H})| \ge c_{n1}\big|\boldsymbol{H}\big) \\ &\le c_{n1}^{-2}\mathbb{E}\big[|M_1 - \mathbb{E}(M_1|\boldsymbol{H})|^2\big|\boldsymbol{H}\big] \le c_{n1}. \end{aligned}$$

Then, using the law of total expectation, we obtain that

$$\mathbb{P}\big(|M_1 - \mathbb{E}M_1| \ge c_{n1}\big) \le c_{n1} = o(1).$$

For M_2, applying Lemma G.2 with $s = 2$, $\boldsymbol{A} = \boldsymbol{P}/(n-1)$, and $\xi_i = \boldsymbol{X}_i^\top(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)$, we obtain that

$$\mathbb{E}M_2 = \frac{\mathrm{tr}(\boldsymbol{P})}{n-1}\mathbb{E}\big|\boldsymbol{X}_1^\top(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\big|^2 = \|\mathbf{O}^\top(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\|_2^2.$$

Notice that due to the condition $\mathbb{E}|\boldsymbol{X}_1^\top\boldsymbol{\beta}_z|^4 < C$, $z \in \{0,1\}$, we have

$$\|\mathbf{O}^\top\boldsymbol{\beta}_z\|_2^2 = \mathbb{E}|\boldsymbol{X}_1^\top\boldsymbol{\beta}_z|^2 \le \left(\mathbb{E}|\boldsymbol{X}_1^\top\boldsymbol{\beta}_z|^4\right)^{1/2} \le C^{1/2}. \tag{42}$$

Next, applying Lemma G.3, we obtain that

$$\begin{aligned} \mathbb{E}|M_2 - \mathbb{E}M_2|^2 &\le C(2)\left(\mathbb{E}\xi_1^4\right)^{1/2}\frac{1}{(n-1)^2}\sum_{i,j}P_{ij}^2 \\ &= C(2)\left(\mathbb{E}\xi_1^4\right)^{1/2}(n-1)^{-1} = O(n^{-1}). \end{aligned}$$

Hence, by choosing $c_{n2} = \left[C(2)(\mathbb{E}\xi_1^4)^{1/2}(n-1)^{-1}\right]^{1/3} = o(1)$, we have

$$\mathbb{P}\big(|M_2 - \mathbb{E}M_2| \geq c_{n2}\big) \leq c_{n2} = o(1).$$

For M_3, we observe that $\mathbb{E}M_3 = 0$ due to the independence between $\boldsymbol{X}$ and $\varepsilon(z)$. Denoting $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ with $\xi_i = \boldsymbol{X}_i^\top(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)$, we obtain that

$$\begin{aligned}\mathbb{E}M_3^2 &= \mathbb{E}\left[\frac{1}{n-1}(\varepsilon(1) - \varepsilon(0))\boldsymbol{P}\boldsymbol{\xi}\right]^2 = (n-1)^{-2}\left(\sigma_{\varepsilon(1)}^2 + \sigma_{\varepsilon(0)}^2\right)\mathbb{E}(\boldsymbol{\xi}^\top\boldsymbol{P}\boldsymbol{\xi}) \\ &= (n-1)^{-1}\left(\sigma_{\varepsilon(1)}^2 + \sigma_{\varepsilon(0)}^2\right)\mathbb{E}M_2 = O(n^{-1}).\end{aligned}$$

where we used $\boldsymbol{P}^2 = \boldsymbol{P}$ and (42). Then, we choose $c_{n3} = (\mathbb{E}M_3^2)^{1/3} = o(1)$ such that

$$\mathbb{P}(|M_3| \geq c_{n3}) < c_{n3}^{-2}\mathbb{E}M_3^2 = c_{n3} = o(1).$$

To sum up, we have shown that with probability $1 - o(1)$,

$$\left|S_{\boldsymbol{H},\tau}^2 - \mathbb{E}S_{\boldsymbol{H},\tau}^2\right| \leq c_{n1} + c_{n2} + 2c_{n3} = o(1), \tag{43}$$

where

$$\mathbb{E}S_{\boldsymbol{H},\tau}^2 = \alpha\left(\sigma_{\varepsilon(1)}^2 + \sigma_{\varepsilon(0)}^2\right) + \|\mathbf{O}^\top(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\|_2^2 + O(n^{-1}).$$

Next, we handle $S^2_{\mathrm{diag}\{\boldsymbol{H}\},Y(1)-Y(0)}$. It is easy to see that

$$|S^2_{\mathrm{diag}\{\boldsymbol{H}\},Y(1)-Y(0)} - \alpha S^2_{Y(1)-Y(0)}| < \max_i |H_{ii} - \alpha| \cdot S^2_{Y(1)-Y(0)}. \tag{44}$$

By Proposition 2, we have that with probability $1 - o(1)$,

$$\max_i |H_{ii} - \alpha| < c_{n4} := n^{-\delta}, \tag{45}$$

for some constant $\delta \in (0, \frac{\eta}{8+2\eta})$. For $S^2_{Y(1)-Y(0)}$, applying Lemmas G.2 and G.3 with $s = 2$, $\boldsymbol{A} = \boldsymbol{P}/(n-1)$ and $\xi_i = (Y_i(1) - Y_i(0)) - (\mu_1 - \mu_0)$, we obtain that

$$\mathbb{E}S^2_{Y(1)-Y(0)} = \frac{\mathrm{tr}(\boldsymbol{P})}{n-1}\mathbb{E}\xi_1^2 = \sigma_{\varepsilon(1)}^2 + \sigma_{\varepsilon(0)}^2 + \|\mathbf{O}^\top(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\|_2^2.$$

and

$$\begin{aligned}\mathbb{E}|S^2_{Y(1)-Y(0)} - \mathbb{E}S^2_{Y(1)-Y(0)}|^2 &\leq C(2)\left(\mathbb{E}\xi_1^4\right)^{1/2}\frac{1}{(n-1)^2}\sum_i\sum_j P_{ij}^2 \\ &= C(2)\left(\mathbb{E}\xi_1^4\right)^{1/2}(n-1)^{-2} = O(n^{-1}).\end{aligned}$$

Thus, by choosing $c_{n5} = \left[\mathbb{E}|S^2_{Y(1)-Y(0)} - \mathbb{E}S^2_{Y(1)-Y(0)}|^2\right]^{1/3} = o(1)$, we have that

$$\mathbb{P}\left(|S^2_{Y(1)-Y(0)} - \mathbb{E}S^2_{Y(1)-Y(0)}| \geq c_{n5}\right) \leq c_{n5} = o(1). \tag{46}$$

Plugging (45) and (46) into (44), we obtain that with probability $1 - o(1)$,

$$\begin{aligned} & \left| S^2_{\text{diag}\{\boldsymbol{H}\},Y(1)-Y(0)} - \alpha \left(\sigma^2_{\varepsilon(1)} + \sigma^2_{\varepsilon(0)} + \|\mathbf{O}^\top(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\|_2^2 \right) \right| \\ & < c_{n4} \left(\sigma^2_{\varepsilon(1)} + \sigma^2_{\varepsilon(0)} + \|\mathbf{O}^\top(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\|_2^2 + c_{n5} \right) + \alpha c_{n5} = o(1). \end{aligned} \tag{47}$$

Finally, combining (41), (43) and (47), we obtain that with probability $1 - o(1)$,

$$S^2_\tau - S^2_{e(1)-e(0)} - S^2_{\text{diag}\{\boldsymbol{H}\},Y(1)-Y(0)} > (1-\alpha)\|\mathbf{O}^\top(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\|_2^2 + o(1) \geq o(1).$$

The conclusion then follows. □

Finally, we give the proof of Proposition 2.

Proof of Proposition 2. For simplicity of notations, we denote

$$\mathbf{W} := n^{-1/2} \left(\boldsymbol{V}_1, \cdots, \boldsymbol{V}_n \right)^\top, \quad \mathbf{P} := \boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top.$$

Then, we can write the matrix $\boldsymbol{H}$ as

$$\boldsymbol{H} = \mathbf{P}\mathbf{W}\left(\mathbf{W}^\top\mathbf{P}\mathbf{W}\right)^{-1}\mathbf{W}^\top\mathbf{P}.$$

Now, we introduce a truncated matrix $\tilde{\boldsymbol{V}} = (\tilde{\boldsymbol{V}}_1, \ldots, \tilde{\boldsymbol{V}}_n)^\top$ as

$$\tilde{V}_{ij} := \mathbf{1}\left(|V_{ij}| \leq \varphi_n \log n\right) \cdot V_{ij}, \quad \text{with} \quad \varphi_n := n^{\frac{2}{4+\eta}}, \tag{48}$$

and denote $\tilde{\mathbf{W}} := n^{-1/2} \left(\tilde{\boldsymbol{V}}_1, \cdots, \tilde{\boldsymbol{V}}_n \right)^\top$. Combining the moment bound $\max_j \mathbb{E}|V_{ij}|^{4+\eta} < C$ with Markov's inequality, we obtain from a simple union bound that

$$P(\tilde{\boldsymbol{V}} = \boldsymbol{V}) = 1 - \mathbb{P}\left(\max_{i,j} |V_{ij}| > \varphi_n \log n \right) = 1 - O\left((\log n)^{-(4+\eta)} \right). \tag{49}$$

By definition, we have

$$\begin{aligned} \mathbb{E}\tilde{V}_{ij} &= -\mathbb{E}\left[\mathbf{1}\left(|V_{ij}| > \varphi_n \log n\right) V_{ij}\right], \\ \mathbb{E}|\tilde{V}_{ij}|^2 &= 1 - \mathbb{E}\left[\mathbf{1}\left(|V_{ij}| > \varphi_n \log n\right) |V_{ij}|^2\right]. \end{aligned} \tag{50}$$

Using the tail probability expectation formula, we can check that

$$\begin{aligned} & \mathbb{E}\left|\mathbf{1}\left(|V_{ij}| > \varphi_n \log n\right) V_{ij}\right| = \int_0^\infty \mathbb{P}\left(\left|\mathbf{1}\left(|V_{ij}| > \varphi_n \log n\right) V_{ij}\right| > s\right) \mathrm{d}s \\ & = \int_0^{\varphi_n \log n} \mathbb{P}\left(|V_{ij}| > \varphi_n \log n\right) \mathrm{d}s + \int_{\varphi_n \log n}^\infty \mathbb{P}\left(|V_{ij}| > s\right) \mathrm{d}s \\ & \lesssim \int_0^{\varphi_n \log n} \left(\varphi_n \log n\right)^{-(4+\eta)} \mathrm{d}s + \int_{\varphi_n \log n}^\infty s^{-(4+\eta)} \mathrm{d}s \lesssim \left(\varphi_n \log n\right)^{-(3+\eta)}. \end{aligned}$$

Here, for simplicity of notations, given two quantities a_n, b_n depending on n, we have used $a_n \lesssim b_n$ to mean that $|a_n| = O(|b_n|)$. Similarly, we have

$$\begin{aligned}
&\mathbb{E}\left|\mathbf{1}\left(|V_{ij}| > \varphi_n \log n\right) V_{ij}\right|^2 = 2\int_0^\infty s\mathbb{P}\left(\left|\mathbf{1}\left(|V_{ij}| > \varphi_n \log n\right) V_{ij}\right| > s\right) \mathrm{d}s \\
&= 2\int_0^{\varphi_n \log n} s\mathbb{P}\left(|V_{ij}| > \varphi_n \log n\right) \mathrm{d}s + 2\int_{\varphi_n \log n}^\infty s\mathbb{P}\left(|V_{ij}| > s\right) \mathrm{d}s \\
&\lesssim \int_0^{\varphi_n \log n} s\left(\varphi_n \log n\right)^{-(4+\eta)} \mathrm{d}s + \int_{\varphi_n \log n}^\infty s^{-(3+\eta)} \mathrm{d}s \lesssim \left(\varphi_n \log n\right)^{-(2+\eta)}.
\end{aligned}$$

From the above two estimates, we can derive that

$$|\mathbb{E}\tilde{V}_{ij}| \le n^{-3/2}, \quad \mathbb{E}|\tilde{V}_{ij}|^2 = 1 + O(n^{-1}), \tag{51}$$

$$\mathbb{E}\|\tilde{\boldsymbol{V}} - \boldsymbol{V}\|_F^2 = \sum_{i,j} \mathbb{E}\left|\mathbf{1}\left(|V_{ij}| > \varphi_n \log n\right) V_{ij}\right|^2 \lesssim n^{\frac{4}{4+\eta}} (\log n)^{-(2+\eta)}. \tag{52}$$

As a consequence, we get that

$$\|\mathbb{E}\tilde{\mathbf{W}}\|_F \le n^{-1/2}\Big(\sum_{i,j} |\mathbb{E}\tilde{V}_{ij}|^2\Big)^{1/2} \le n^{-1}, \quad \mathbb{P}\left(\|\tilde{\mathbf{W}} - \mathbf{W}\|_F \ge n^{-\frac{\eta}{8+2\eta}}\right) = o(1). \tag{53}$$

Let $\mathbf{D}$ be a $p \times p$ diagonal matrix with entries $D_{ii} = \mathrm{var}(\tilde{V}_{1i})$, $i \in [p]$. By (51), we have

$$|D_{ii} - 1| = O(n^{-1}). \tag{54}$$

Now, we introduce the matrices $\boldsymbol{\mathcal{W}} := (\tilde{\mathbf{W}} - \mathbb{E}\tilde{\mathbf{W}})\mathbf{D}^{-1/2}$ and

$$\boldsymbol{H}' = \mathbf{P}\boldsymbol{\mathcal{W}}\left(\boldsymbol{\mathcal{W}}^\top \mathbf{P}\boldsymbol{\mathcal{W}}\right)^{-1}\boldsymbol{\mathcal{W}}^\top \mathbf{P}.$$

By definition and (54), the entries of $\boldsymbol{\mathcal{W}}$ are independent random variables satisfying

$$\mathbb{E}\mathcal{W}_{ij} = 0, \quad \mathbb{E}|\mathcal{W}_{ij}|^2 = n^{-1}, \quad |\mathcal{W}_{ij}| \le \frac{2\varphi_n \log n}{n^{1/2}}, \quad i \in [n],\ j \in [p]. \tag{55}$$

Moreover, from (53), we see that

$$\mathbb{P}\left(\|\boldsymbol{\mathcal{W}}\mathbf{D}^{1/2} - \mathbf{W}\|_F \ge 2n^{-\frac{\eta}{8+2\eta}}\right) = o(1). \tag{56}$$

On the other hand, it is well-known that the empirical spectral distribution of $\mathbf{W}\mathbf{P}\mathbf{W}^\top$ satisfies the famous Marchenko-Pastur (MP) law [Marčenko and Pastur, 1967], and their eigenvalues are all inside the support of the MP law, $[(1-\sqrt{\alpha})^2, (1+\sqrt{\alpha})^2]$, with high probability [Bai and Silverstein, 1998]. In particular, the following estimate is a direct consequence of the results in Bai and Silverstein [1998]: for any small constant $0 < c < (1-\sqrt{\alpha})^2$,

$$\begin{aligned}
&\mathbb{P}\left((1-\sqrt{\alpha})^2 - c \le \lambda_{\min}(\mathbf{W}^\top\mathbf{P}\mathbf{W}) \le \lambda_{\max}(\mathbf{W}^\top\mathbf{P}\mathbf{W}) \le (1+\sqrt{\alpha})^2 + c\right) \\
&= 1 - o(1),
\end{aligned} \tag{57}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimum and maximum eigenvalues, respectively. With (54), (56) and (57), we obtain the following two estimates: there exists a constant $C_1 > 0$ (depending on $\limsup \alpha$) such that

$$\mathbb{P}\left(C_1^{-1} \leq \lambda_{\min}(\boldsymbol{\mathcal{W}}^\top \mathbf{P} \boldsymbol{\mathcal{W}}) \leq \lambda_{\max}(\boldsymbol{\mathcal{W}}^\top \mathbf{P} \boldsymbol{\mathcal{W}}) \leq C_1\right) = 1 - o(1), \tag{58}$$

and

$$\mathbb{P}\left(\|\boldsymbol{H}' - \boldsymbol{H}\|_2 \geq C_1 n^{-\frac{\eta}{8+2\eta}}\right) \to 0. \tag{59}$$

Since $|H'_{ii} - H_{ii}| \leq \|\boldsymbol{H}' - \boldsymbol{H}\|_2$, to conclude the proof, it suffices to show that

$$\mathbb{P}\left(\max_{i \in [n]} |H'_{ii} - \alpha| > n^{-\delta}\right) \to 0 \tag{60}$$

for any constant $0 < \delta < \frac{\eta}{8+2\eta}$.

Let $\varepsilon_n = n^{-1/2}$. By (58), there exists a constant $C_2 > 0$ such that

$$\mathbb{P}\left(\|\boldsymbol{H}' - \boldsymbol{H}_\varepsilon\|_2 \geq C_2 \varepsilon_n\right) = 1 - o(1),$$

where $\boldsymbol{H}_\varepsilon$ is defined as

$$\boldsymbol{H}_\varepsilon := \mathbf{P}\boldsymbol{\mathcal{W}} \frac{1}{(\boldsymbol{\mathcal{W}}^\top \mathbf{P})(\mathbf{P}\boldsymbol{\mathcal{W}}) - \mathrm{i}\varepsilon_n \boldsymbol{I}} \boldsymbol{\mathcal{W}}^\top \mathbf{P}.$$

Observe the following matrix identity

$$\boldsymbol{H}_\varepsilon = 1 + \frac{\mathrm{i}\varepsilon_n}{\mathbf{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top \mathbf{P} - \mathrm{i}\varepsilon_n \boldsymbol{I}}.$$

Now, to conclude (60), it suffices to prove that

$$\mathbb{P}\left[\left|\left(\frac{\mathrm{i}\varepsilon_n}{\mathbf{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top \mathbf{P} - \mathrm{i}\varepsilon_n \boldsymbol{I}}\right)_{ii} + 1 - \alpha\right| \geq n^{-\delta}\right] \leq n^{-C} \tag{61}$$

for any constant $0 < \delta < \frac{\eta}{8+2\eta}$ and large constant $C > 1$. Then, taking a simple union bound, we obtain that

$$\mathbb{P}\left[\max_{i \in [n]}\left|\left(\frac{\mathrm{i}\varepsilon_n}{\mathbf{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top \mathbf{P} - \mathrm{i}\varepsilon_n \boldsymbol{I}}\right)_{ii} + 1 - \alpha\right| \geq n^{-\delta}\right] \leq n^{-(C-1)}, \tag{62}$$

which concludes (60).

For the proof of (61), we will adopt Theorem 11.2 of Knowles and Yin [2016]. More precisely, under the conditions on $\boldsymbol{\mathcal{W}}$ in (55), the following estimate was proved in Theorem 11.2 of Knowles and Yin [2016]: for any small constant $c > 0$ and large constant $C > 1$,

$$\mathbb{P}\left(\left|\left(\frac{\mathrm{i}\varepsilon_n}{\mathbf{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top \mathbf{P} - \mathrm{i}\varepsilon_n \boldsymbol{I}}\right)_{ii} - \left(m(\mathrm{i}\varepsilon_n)\mathbf{P} - \frac{1}{n}\mathbf{1}_n \mathbf{1}_n^\top\right)_{ii}\right| \geq \frac{\varphi_n}{n^{1/2-c}}\right) \leq n^{-C}, \tag{63}$$

where $m(z)$ is the unique analytic function in a neighborhood around the origin that satisfies the equation

$$-m(z) + \frac{p}{n}\frac{m(z)}{m(z) + z} = 1.$$

In particular, from this equation, we can solve that that

$$|m(\mathrm{i}\varepsilon_n) + (1-\alpha)| \le C_3 \varepsilon_n^{-1} \tag{64}$$

for a constant $C_3 > 0$. Plugging (64) into (63) and using $\left(n^{-1}\mathbf{1}_n\mathbf{1}_n^\top\right)_{ii} = n^{-1}$, we obtain the estimate (61), which concludes the proof.

Remark 1. An estimate of the form (63) is often called a *local law* of $(\mathbf{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top\mathbf{P} - z\boldsymbol{I})^{-1}$, the Green's function of $\mathbf{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top\mathbf{P}$. Such local laws of sample covariance matrices were also established in many other papers under different settings, see e.g., Bai et al. [2007], Bloemendal et al. [2014], Bao et al. [2015], Ding and Yang [2018], Xi et al. [2020] (we remark that this list is far from being comprehensive). The setting in Theorem 11.2 of Knowles and Yin [2016] is closest to our current one, but there is a minor difference that $|\mathcal{W}_{ij}|$ is of order $O(n^{-1/2+\varepsilon})$ in Knowles and Yin [2016]. However, using the argument in Ding and Yang [2018], it is rather straightforward to extend Theorem 11.2 of Knowles and Yin [2016] to our setting with $|\mathcal{W}_{ij}| \le \frac{\varphi_n \log n}{n^{1/2}}$ in (55) and conclude (63). We omit the details here.

□

H Additional numerical experiments

In this section, we conduct additional simulation analysis to examine the finite sample performance of the proposed estimator and inference procedure. In the main text, we consider the setup that $\boldsymbol{\mathcal{X}}$ and $\check{\boldsymbol{\varepsilon}}(z)$ (which is used to generate the independent t residual) have i.i.d. entries from t distribution with 3 degrees of freedom. Here, we consider 2 more setups:

- $\boldsymbol{\mathcal{X}}$ and $\check{\boldsymbol{\varepsilon}}(z)$ have i.i.d. entries from Cauchy distribution
- $\boldsymbol{\mathcal{X}}$ have i.i.d. entries from Cauchy distribution and $\check{\boldsymbol{\varepsilon}}(z)$ have i.i.d. entries from t distributions with degrees of freedom 3. We also modify the model to

  $$Y_i(1) = \mu_1 + \mathrm{Scale}(\mathrm{Trans}(\boldsymbol{X}_i^\top \boldsymbol{\beta}_1)) + \varepsilon_i(1)/\sqrt{\gamma},$$
  $$Y_i(0) = \mu_0 + \mathrm{Scale}(\mathrm{Trans}(\boldsymbol{X}_i^\top \boldsymbol{\beta}_0)) + \varepsilon_i(0)/\sqrt{\gamma},$$

  where for a finite population $\{a_i\}_{i=1}^n$:

  $$\mathrm{Trans}(a_i) = b_{(\pi(i))},$$

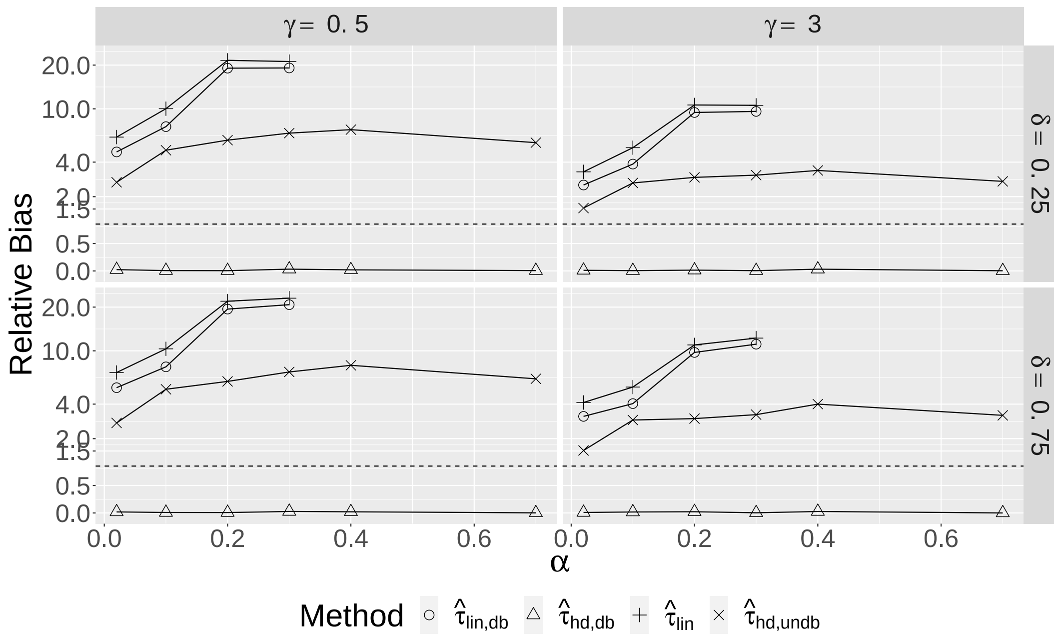
  $b_{(1)} \le b_{(2)} \le \ldots \le b_{(n)}$ is the ordered sequence of $\{b_i\}_{i=1}^n$ with b_i generated from t distribution with degrees of freedom 3, and $\pi(i)$ is the rank of a_i.

For both setups, we consider the same factorial experiments regarding γ, δ, α, and the generating models of $\varepsilon_i(z)$. Note that the first setup represents the most challenging case in which Assumption 2–3 fail. In the second setup, albeit with extremely heavy-tail covariates, the Assumption 2–3 hold. Figure 8–11 show the results for the first setup. Figure 12–15 show the results for the second setup.

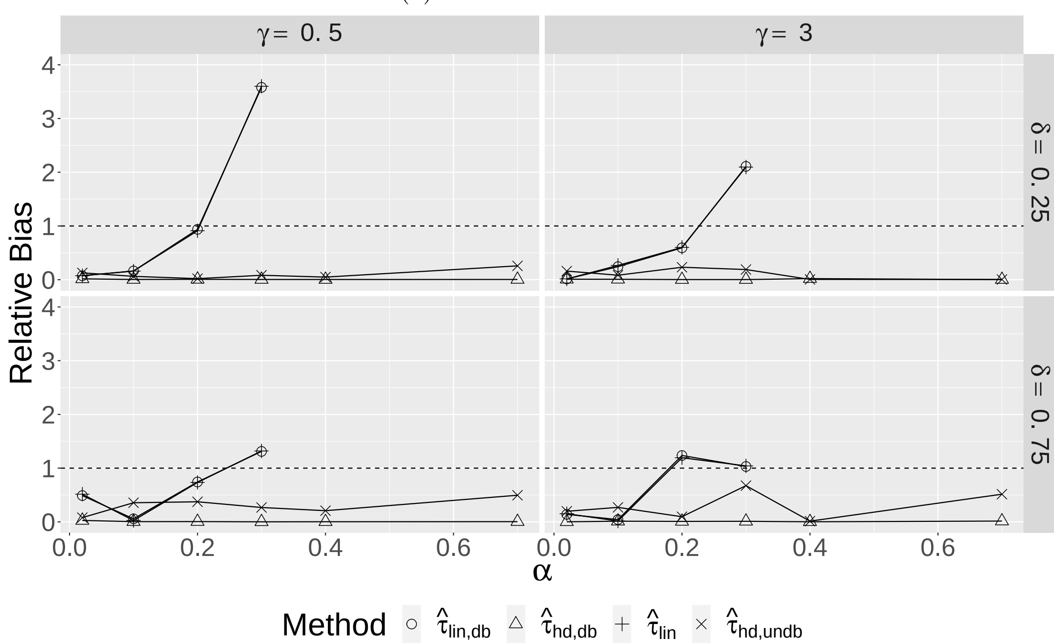
For the first setup, $(\hat{\tau}_{\mathrm{hd}}, \hat{\sigma}^2_{\mathrm{hd,cb}})$ outperforms its competitors in terms of relative RMSE, relative bias, more reliable inference and shorter confidence intervals in all cases, except under the independent t residual with $\gamma = 3$ and $\alpha \le 0.1$. The performance of $\hat{\tau}_{\mathrm{lin}}$ can be more catastrophic than in the main text when α is large. For example, the relative RMSE can be as large as 40. Interestingly, although our asymptotic theory does not apply to these extreme regimes, in most of the cases the relative RMSE and relative confidence interval length produced by our debiased estimator is not too far away from 1. In other words, $(\hat{\tau}_{\rm hd}, \hat{\sigma}^2_{\rm hd,cb})$ does not give significant harm compared to without covariate adjustment in these extreme setups. This demonstrates the robustness of our method when faced with extreme cases.

For the second setup, $\hat{\tau}_{\rm hd}$ outperforms other competitors for smaller relative bias and relative RMSE under the worst-case residual. Although our theory only guarantees that our method has a better estimation efficiency and a shorter confidence interval length than the unadjusted method under a high signal-to-noise ratio and light-tailed covariates, it is interesting that we can observe improved efficiency even with heavy-tailed covariates.

We notice that for both setups, when $\gamma = 3$, sometimes, $\hat{\tau}_{\rm hd,undb}$ slightly outperforms $\hat{\tau}_{\rm hd}$ in terms of relative RMSE but with larger bias. Since under the worst-case residual, $\hat{\tau}_{\rm hd,undb}$ has very large relative RMSE, we still recommend using $\hat{\tau}_{\rm hd}$ for heavy-tail covariates.

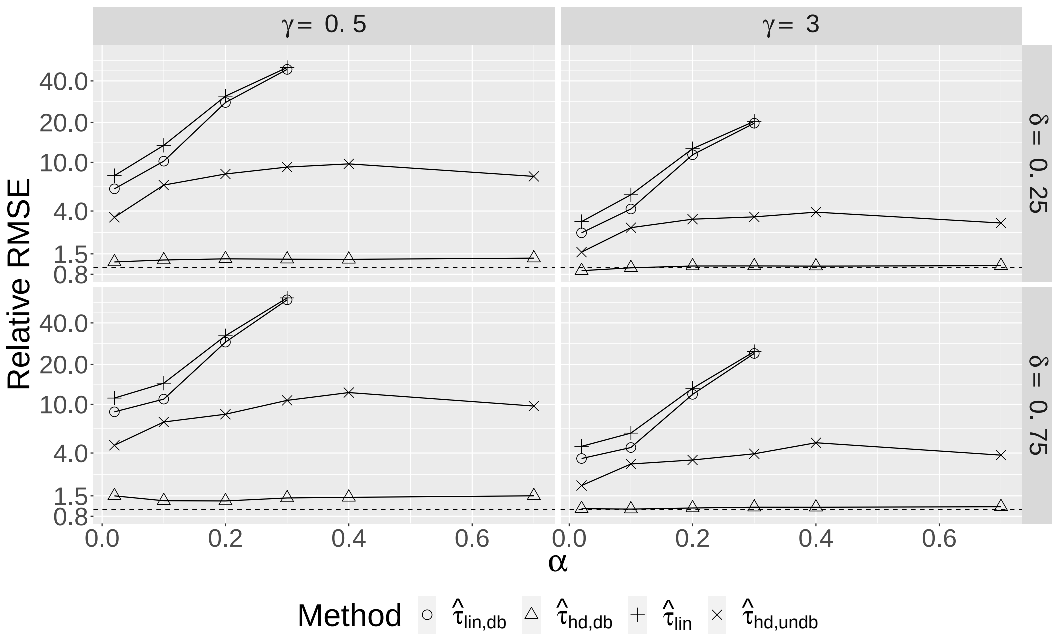


(a) Worst case residual

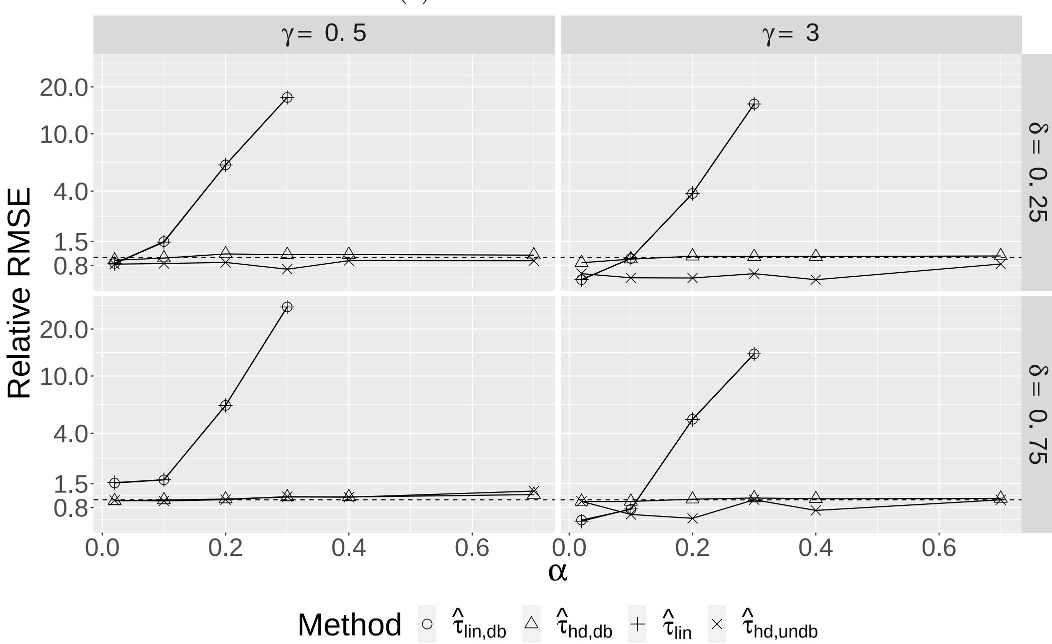


(b) Independent t residual

Figure 8: *The first set up.* Relative bias for different choices of γ, δ and α under the worst-case residual and independent t residual. The dashed lines signify 1. Notice that for the first figure, we use a transformation of $\log_{10}(1+x)$ for the y-axis to adapt the curve display.

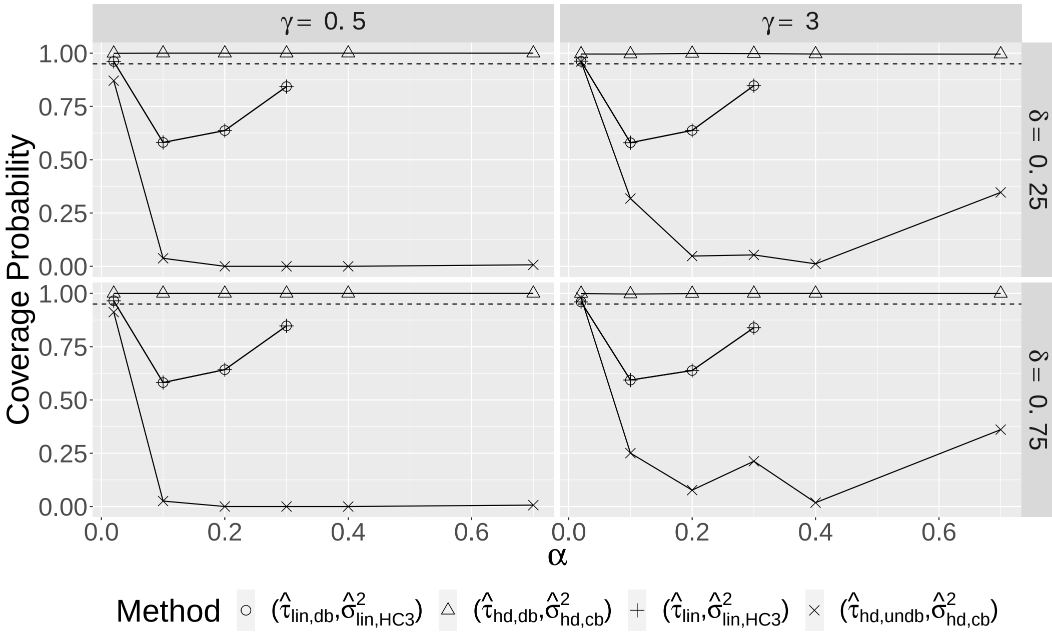


(a) Worst case residual

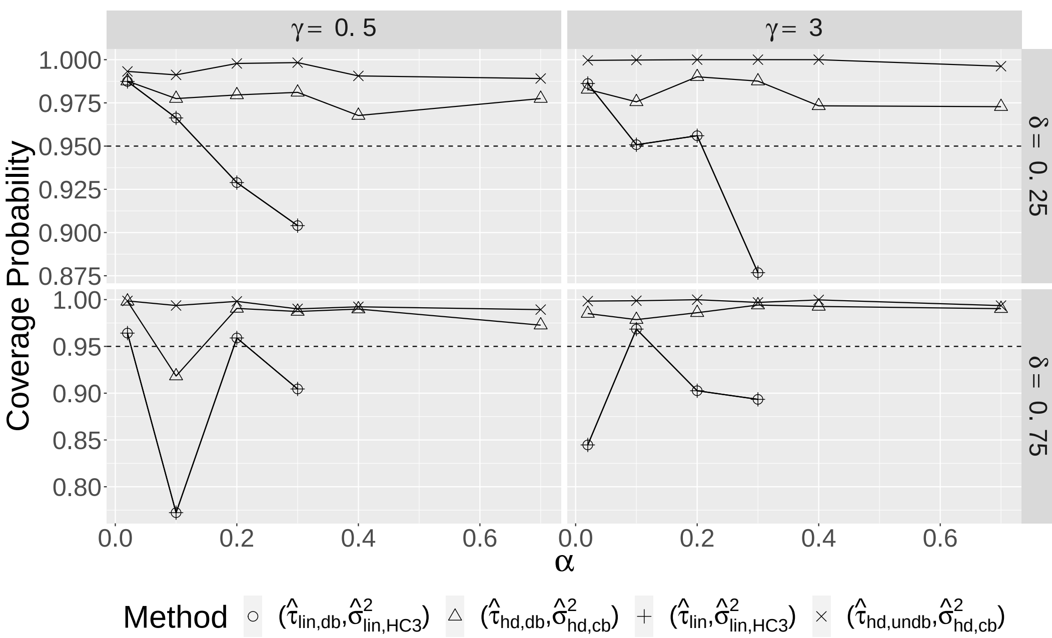


(b) Independent t residual

Figure 9: *The first set up.* Relative RMSE for different choices of γ, δ and α under the worst-case residual and independent t residual. The dashed lines signify 1. Notice that for both figures, we use a transformation of $\log_{10}(1+x)$ for the y-axis to adapt the curve display.

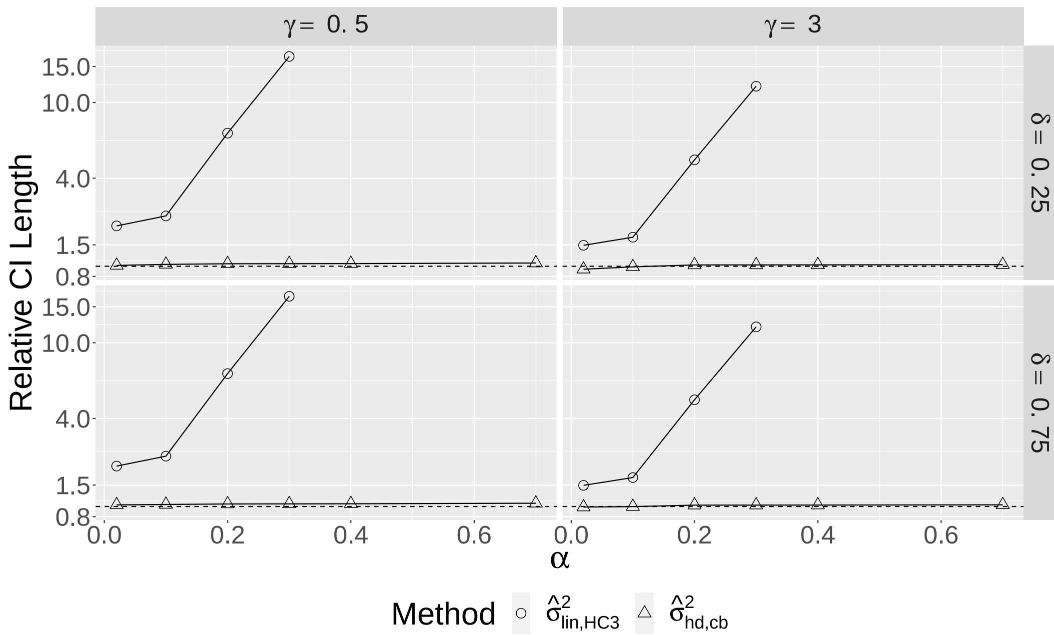


(a) Worst case residual

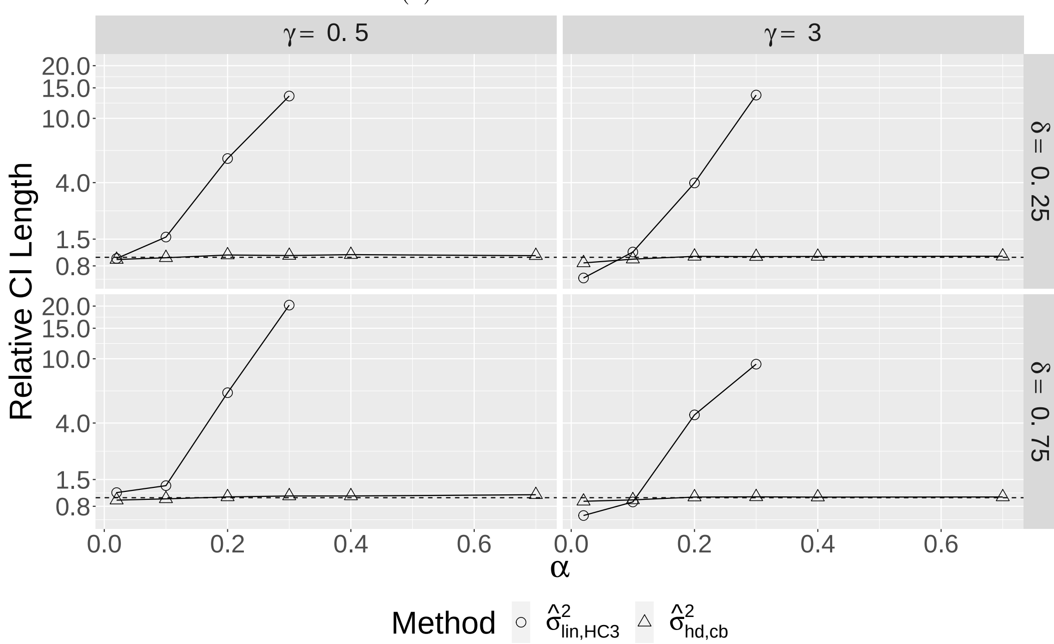


(b) Independent t residual

Figure 10: *The first set up.* Coverage probabilities for different choices of γ, δ and α under the worst-case residual and independent t residual. The dashed lines signify 0.95.

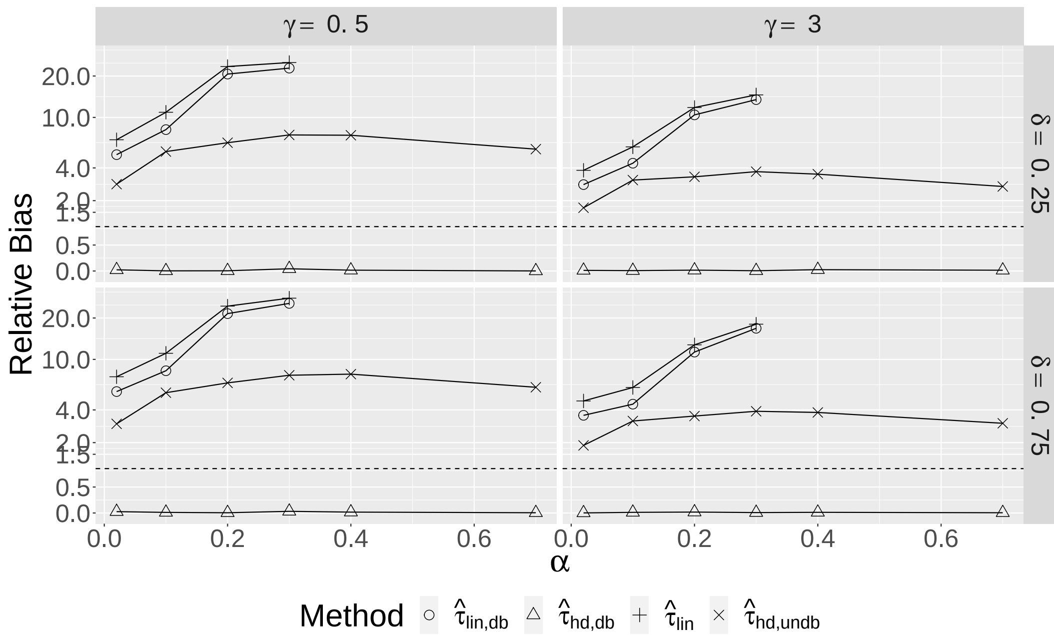


(a) Worst case residual

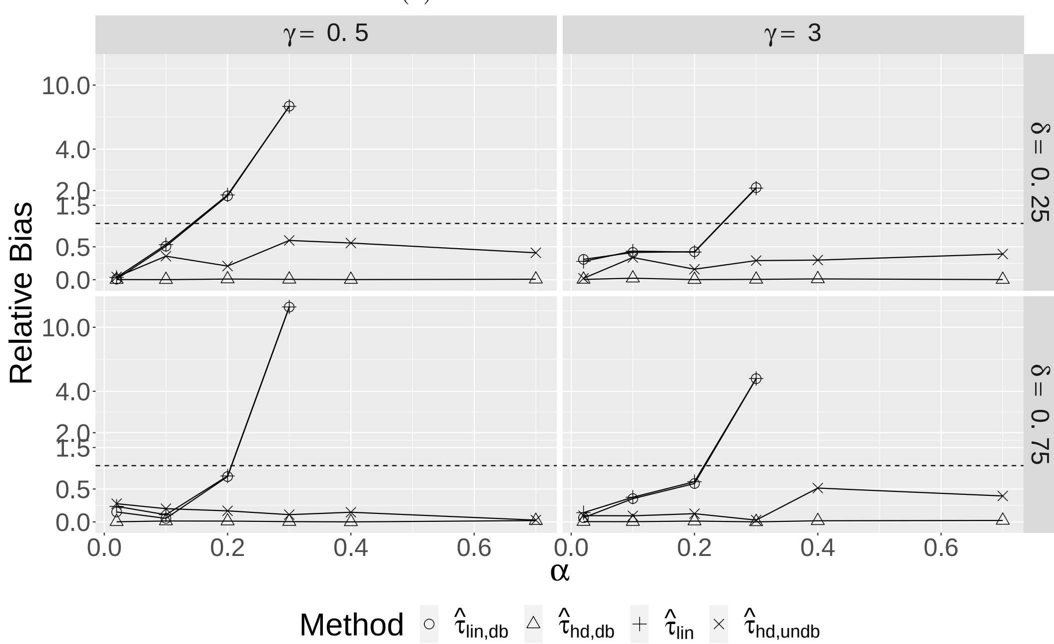


(b) Independent t residual

Figure 11: *The first set up.* Relative confidence interval length for different choices of γ, δ and α under the worst-case residual and independent t residual. The dashed lines signify 1. Notice that for both figures, we use a transformation of $\log_{10}(1 + x)$ for the y-axis to adapt the curve display.

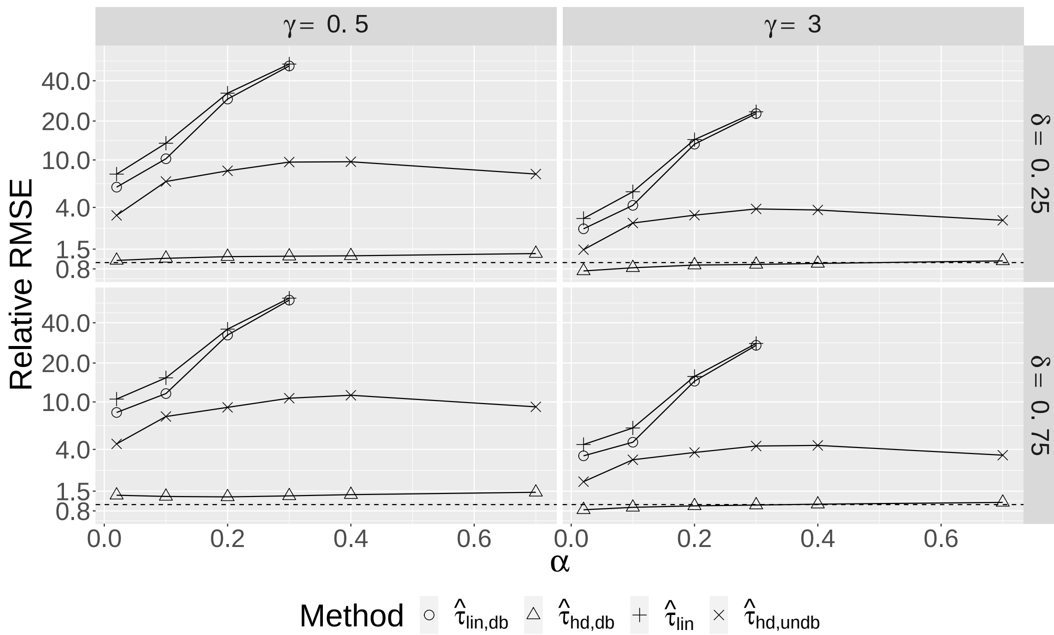


(a) Worst case residual

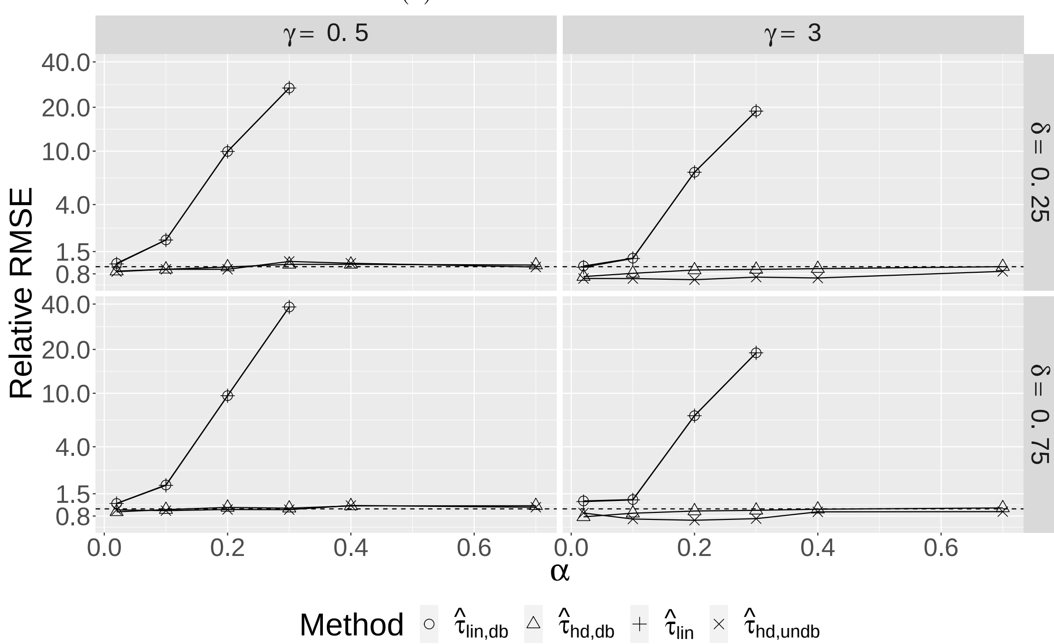


(b) Independent t residual

Figure 12: *The second set up.* Relative bias for different choices of γ, δ and α under the worst-case residual and independent t residual. The dashed lines signify 1. Notice that for both figures, we use a transformation of $\log_{10}(1 + x)$ for the y-axis to adapt the curve display.

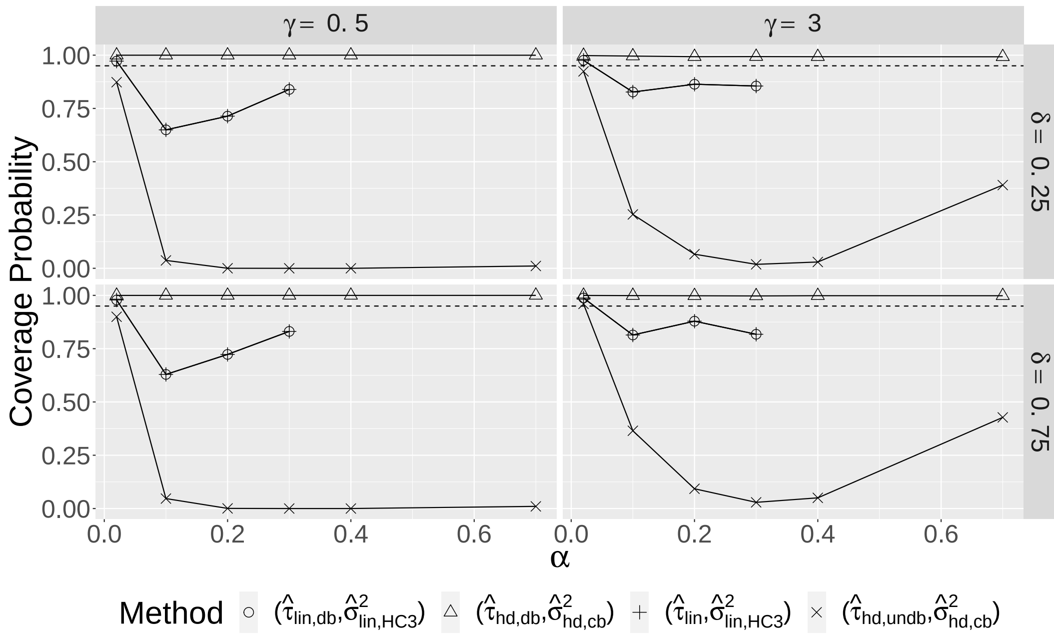


(a) Worst case residual

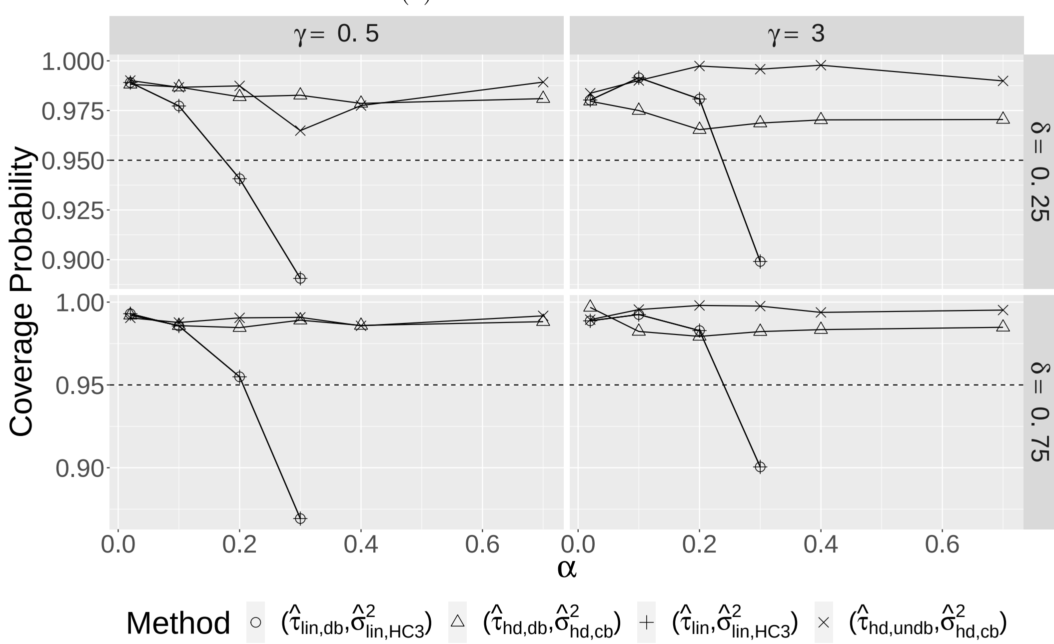


(b) Independent t residual

Figure 13: *The second set up.* Relative RMSE for different choices of γ, δ and α under the worst-case residual and independent t residual. The dashed lines signify 1. Notice that for both figures, we use a transformation of $\log_{10}(1 + x)$ for the y-axis to adapt the curve display.

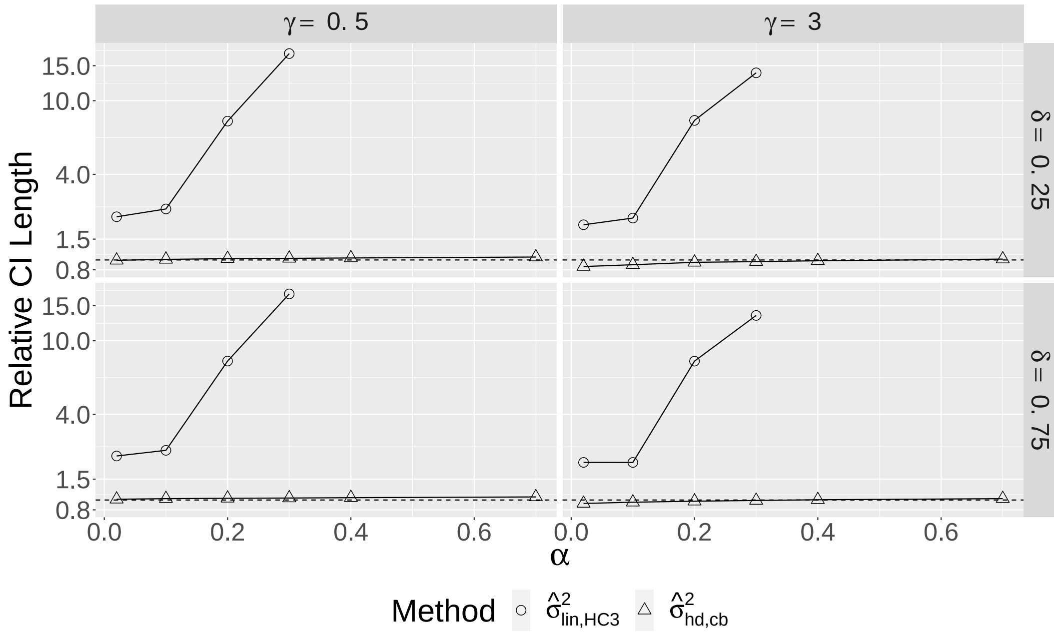


(a) Worst case residual

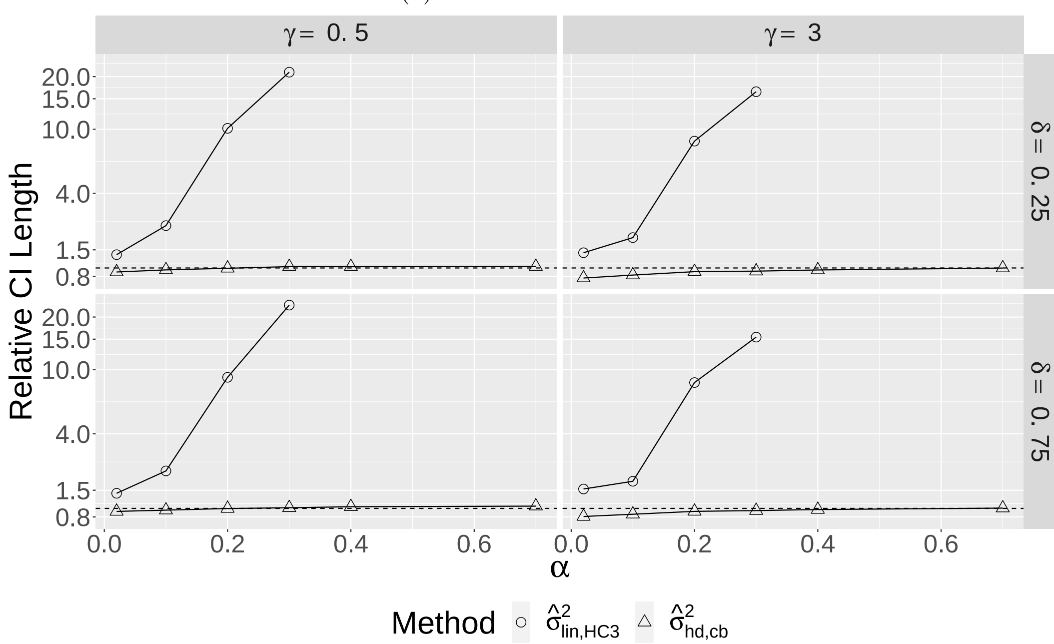


(b) Independent t residual

Figure 14: *The second set up.* Coverage probabilities for different choices of γ, δ and α under the worst-case residual and independent t residual. The dashed lines signify 0.95.



(a) Worst case residual



(b) Independent t residual

Figure 15: *The second set up.* Relative interval length for different choices of γ, δ and α under the worst-case residual and independent t residual. The dashed lines signify 1. Notice that for both figures, we use a transformation of $\log_{10}(1 + x)$ for the y-axis to adapt the curve display.