# SUPPLEMENT TO "RERANDOMIZATION WITH DIMINISHING COVARIATE IMBALANCE AND DIVERGING NUMBER OF COVARIATES"

BY YUHAO WANG[1,a] AND XINRAN LI[2,b]

[1]Institute for Interdisciplinary Information Sciences, Tsinghua University, [a]yuhaow@tsinghua.edu.cn

[2]Department of Statistics, University of Illinois at Urbana-Champaign, [b]xinranli@illinois.edu

Appendix A1 provides some finite-sample diagnosis tools, discusses the choice of covariates and threshold for rerandomization, and conducts a simulation study.

Appendix A2 studies regression adjustment after rerandomization.

Appendix A3 studies Berry–Esseen-type bound for finite population central limit theorem in simple random sampling.

Appendix A4 studies asymptotic properties for completely randomized and rerandomized experiments. It includes the proofs of Theorems 1–3, Corollary 1; and technical details about the comments of (9).

Appendix A5 studies the limiting behavior of the constrained Gaussian random variable. It includes the proof of Theorem 4.

Appendix A6 studies asymptotics for the optimal rerandomization. It includes the proofs of Theorems 5 and 6.

Appendix A7 studies large-sample inference for rerandomization. It includes the proofs of Theorems 7 and 8.

Appendix A8 studies the regularity conditions and finite-sample diagnoses for rerandomization. It also provides the technical details for the comments on $\gamma_n$ and $\sum_{i=1}^n H_{ii}^{3/2}$ in Section 7.2.

Appendix A9 studies the asymptotic properties of regression adjustment under rerandomization with a diverging number of covariates. It includes the proof of Theorem A1.

Appendix A10 studies connections to optimal design under certain hypothesized model of the potential outcomes.

## A1. Finite-sample Diagnoses and Simulation Studies.

A1.1. *Finite-sample diagnoses for rerandomization.* Our theoretical results are mostly concerned with the asymptotic properties of rerandomization designs. In this section, we further provide some additional tools for the diagnosis of rerandomization in finite samples. Our diagnosis is based on the bias and mean squared error (MSE) of the difference-in-means estimator $\hat{\tau}$ for estimating the average treatment effect $\tau$. Kapelner et al. (2020) and Nordin and Schultzberg (2020) also considered the MSE of $\hat{\tau}$ under rerandomization, but they focused mainly on the case with equal treatment group sizes (i.e., $r_1 = r_0 = 1/2$). In the following, we consider a general design $\mathcal{D}$ that randomly assigns $r_1$ proportion of units to treatment and the remaining $r_0$ proportion to control. For descriptive convenience, we introduce $\mathbb{E}_{\mathcal{D}}(\cdot)$ and $\text{Var}_{\mathcal{D}}(\cdot)$ to denote the mean and variance under the design $\mathcal{D}$. The bias and MSE of the difference-in-means estimator $\hat{\tau}$ under the design $\mathcal{D}$ can then be written as $\mathbb{E}_{\mathcal{D}}(\hat{\tau} - \tau)$ and $\mathbb{E}_{\mathcal{D}}\{(\hat{\tau} - \tau)^2\}$.

Recall that $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^\top$ is the treatment assignment vector for all $n$ units. Let $\boldsymbol{\pi} \equiv \mathbb{E}_{\mathcal{D}}(\boldsymbol{Z}) \in \mathbb{R}^n$ and $\boldsymbol{\Omega} \equiv \text{Cov}_{\mathcal{D}}(\boldsymbol{Z}) \in \mathbb{R}^{n\times n}$ be its mean and covariance matrix under the design $\mathcal{D}$. Let $y_i = r_0 Y_i(1) + r_1 Y_i(0)$ denote the weighted average of potential outcomes for unit $i$, $\bar{y} = n^{-1}\sum_{i=1}^n y_i$ be the corresponding finite population average, and $\tilde{\boldsymbol{y}} = (y_1 -$

$\bar{y}, \ldots, y_n - \bar{y})^\top \in \mathbb{R}^n$ be the vector consisting of the centered weighted averages of potential outcomes for all units. As demonstrated in Appendix A8, the bias and MSE of the difference-in-means estimator $\hat{\tau}$ have the following forms:

(A1.1)
$$\mathbb{E}_{\mathcal{D}}(\hat{\tau} - \tau) = \frac{(\boldsymbol{\pi} - r_1 \mathbf{1}_n)^\top \tilde{\boldsymbol{y}}}{n r_1 r_0}, \quad \mathbb{E}_{\mathcal{D}}\{(\hat{\tau} - \tau)^2\} = \frac{\tilde{\boldsymbol{y}}^\top \{\boldsymbol{\Omega} + (\boldsymbol{\pi} - r_1 \mathbf{1}_n)(\boldsymbol{\pi} - r_1 \mathbf{1}_n)^\top\} \tilde{\boldsymbol{y}}}{(n r_1 r_0)^2},$$

where $\mathbf{1}_n$ denotes an $n$-dimensional vector with all elements being 1. In (A1.1), $\tilde{\boldsymbol{y}}$ depends on the potential outcomes and is generally unknown in the design stage of an experiment. However, the other quantities in (A1.1) are fully determined by the design $\mathcal{D}$, and can be computed or at least approximated by Monte Carlo method before actually conducting the experiment. Below we consider the worst-case behavior of the design in terms of the estimation bias and MSE in (A1.1) over the unknown potential outcomes. Recall that $V_{\tau\tau}$ in (4) is the variance of $\hat{\tau}$ under the CRE. As verified in Appendix A8, we can equivalently write $V_{\tau\tau}$ as $V_{\tau\tau} = \tilde{\boldsymbol{y}}\tilde{\boldsymbol{y}}^\top / \{n(n-1) r_1 r_0\}$.

PROPOSITION A1. For any design $\mathcal{D}$ that randomly assign $r_1$ proportion of units to treatment and the remaining $r_0$ to control, the maximum absolute bias and the maximum root MSE of the difference-in-means estimator $\hat{\tau}$ under $\mathcal{D}$, standardized by the corresponding standard deviation of $\hat{\tau}$ under the CRE, have the following forms:

(A1.2)
$$\max_{\tilde{\boldsymbol{y}} \neq \mathbf{0}} V_{\tau\tau}^{-1/2} \left| \mathbb{E}_{\mathcal{D}}(\hat{\tau} - \tau) \right| = \sqrt{\frac{n-1}{n r_1 r_0}} \cdot \| \boldsymbol{\pi} - r_1 \mathbf{1}_n \|_2 \geq 0,$$

(A1.3)
$$\max_{\tilde{\boldsymbol{y}} \neq \mathbf{0}} V_{\tau\tau}^{-1/2} \sqrt{\mathbb{E}_{\mathcal{D}}\{(\hat{\tau} - \tau)^2\}} = \sqrt{\frac{n-1}{n r_1 r_0}} \cdot \lambda_{\max}^{1/2} \left( \boldsymbol{\Omega} + (\boldsymbol{\pi} - r_1 \mathbf{1}_n)(\boldsymbol{\pi} - r_1 \mathbf{1}_n)^\top \right) \geq 1,$$

where $\tilde{\boldsymbol{y}}$, $\boldsymbol{\pi}$ and $\boldsymbol{\Omega}$ are the same as defined before, and $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a matrix.

Proposition A1 characterizes the maximum bias and root MSE under any given design. It is not difficult to see that, when the design $\mathcal{D}$ is the CRE, the maximum mean in (A1.2) achieves its minimum value 0, and the maximum root MSE in (A1.3) achieves its minimum value 1. This implies that the CRE is minimax optimal; see also Wu (1981). However, this does not contradict with our Corollary 1, which shows that the difference-in-means estimator under ReM always has smaller or equal variance and shorter or equal symmetric quantile ranges than that under the CRE asymptotically. The reason is that Proposition A1 considers all possible configurations of potential outcomes, including the case with $R_n^2 = 0$, i.e., the potential outcomes are uncorrelated with the covariates. In this case, the asymptotic distribution of $\hat{\tau}$ under ReM reduces to that under the CRE.

More importantly, Proposition A1 can help us conduct some finite-sample diagnoses for rerandomization. Given any acceptance probability $p$ and covariates $\boldsymbol{X}$ for each unit, we can estimate $\boldsymbol{\pi}$ and $\boldsymbol{\Omega}$ by simulating treatment assignments from the corresponding ReM, based on which we can then investigate the maximum bias and root MSE in (A1.2) and (A1.3). In practice, we may consider several choices of $p$ and $\boldsymbol{X}$, and compare them taking into account both the improvement they can bring as shown in Corollary 1 and the finite-sample biases they may cause as shown in Proposition A1; see the next subsection for details.

A1.2. *Choice of covariates and acceptance probability for rerandomization.* Below we consider some practical strategy to choose the covariates and imbalance threshold for the design of rerandomization in practice. From Corollary 1, if the sample size is large and the asymptotic approximation works well, then rerandomization can reduce the MSE of the difference-in-means estimator by $100(1 - v_{K_n,a_n})R_n^2$ percent, or equivalently the standardized MSE is approximately $1 - (1 - v_{K_n,a_n})R_n^2$. On the contrary, from Appendix A1.1, with a finite sample size, the worst-case MSE of the difference-in-means estimator under rerandomization is no less than that under the CRE, and their ratio is the square of the quantity in (A1.3). Obviously, there is a trade-off for the choice of covariates and imbalance threshold. First, when the threshold (or equivalently the acceptance probability) decreases and the covariates are fixed, the asymptotic percentage reduction in MSE will increase (due to the decreasing $v_{K_n,a_n}$), while the finite-sample worst-case MSE is likely to increase. When the number of covariates increases and the acceptance probability is fixed, the asymptotic percentage reduction in MSE may increase or decrease (due to the increasing $v_{K_n,a_n}$ and $R_n^2$), while the finite-sample worst-case MSE is likely to increase.

Inspired by the above trade-off, we propose the following measure for the choice of covariates $\boldsymbol{X}$ and acceptance probability $p$ for rerandomization, which takes a geometric mean of the standardized MSEs in the worst case and the best case (in which the asymptotics works well):

$$c(\boldsymbol{X}, p) \equiv \widetilde{\text{MSE}}(\boldsymbol{X}, p) \times \{1 - (1 - v_{K,a})R^2_{\boldsymbol{X}}\}; \tag{A1.4}$$

other possible measures taking into account the best- and worst-case MSEs can also be considered for practical diagnosis. In (A1.4), $\widetilde{\text{MSE}}(\boldsymbol{X}, p)$ denotes the worst-case standardized MSE under rerandomization with covariates $\boldsymbol{X}$ and acceptance probability $p$, $K$ is the dimension of covariates, $a$ is the $p$-th quantile of the chi-squared distribution with degrees of freedom $K$, and $R^2_{\boldsymbol{X}}$ denotes the squared multiple correlation between potential outcomes and covariates $\boldsymbol{X}$. We can then use (A1.4) as a measure for comparing different rerandomization designs, and can choose the one with minimum value of (A1.4) for the actual implementation of the experiment. Note that $R^2_{\boldsymbol{X}}$ in (A1.4) depends on the potential outcomes and is thus unknown in the design stage of experiments. In practice, we can use some domain knowledge or some prior (pilot) studies to estimate $R^2_{\boldsymbol{X}}$.

Below we illustrate the use of the measure in (A1.4) using the dataset from the Student Achievement and Retention (STAR) Project (STAR, Angrist, Lang and Oreopoulos, 2009), a randomized evaluation of academic services and incentives conducted at a Canadian university. We focus on the treatment group where the students were offered some academic support (including peer-advising service) and scholarships for meeting targeted grades, and the control group receiving neither of these. Similar to Li, Ding and Rubin (2018), we dropped the students with missing covariates, resulting a treated group of size $n_1 = 118$ and $n_0 = 856$. We generate 200 covariates for each unit, where the first five are from the STAR project, i.e., high-school GPA, age, gender and indicators for whether lives at home and whether rarely puts off studying for tests, and the rest 195 are drawn independently from the $t$ distribution with degrees of freedom 2; once generated, these covariates are kept fixed, mimicking the finite population inference. We consider rerandomization with the first $K = 5, 10, 50, 100, 200$ covariates, and consider acceptance probability $p = 0.001, 0.05, 0.01, 0.1, 0.5$. The left half of Table A1 shows the worst-case MSEs, standardized by that under the CRE, under various choices of $(K, p)$, where each worst-case MSE is estimated based on at least about $10^5$ randomly generated treatment assignments from each design. It shows that the worst-case MSE generally increases as the number of covariates increases and the acceptance probability decreases. We further hypothesize that $R^2_{\boldsymbol{X}}$ takes values $0.4, 0.5, 0.6, 0.7, 0.8$, respectively, when $K$ increases from 5 to 200. Intuitively, this implies that the additional gain from including more covariates decreases with the number of included covariates. The right half of Table A1 shows the value of the measure in (A1.4), which suggests to use rerandomization with $K = 5$ covariates and acceptance probability $p = 0.01$.

TABLE A1

*The worst-case mean squared error (standardized by that under the CRE) and the measure in (A1.4) for choosing and diagnosing rerandomization designs under various choices of covariates (whose number is denoted by $K$) and acceptance probability (denoted by $p$).*

| | worst-case mean squared error | | | | | The measure in (A1.4) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $K$ \ $p$ | 0.5 | 0.1 | 0.05 | 0.01 | 0.001 | 0.5 | 0.1 | 0.05 | 0.01 | 0.001 |
| 5 | 1.012 | 1.025 | 1.033 | 1.068 | 1.216 | 0.819 | 0.704 | 0.685 | **0.674** | 0.744 |
| 10 | 1.015 | 1.095 | 1.147 | 1.264 | 1.414 | 0.839 | 0.754 | 0.752 | 0.762 | 0.792 |
| 50 | 1.023 | 1.340 | 1.523 | 1.935 | 2.477 | 0.925 | 1.083 | 1.188 | 1.411 | 1.676 |
| 100 | 1.029 | 1.448 | 1.684 | 2.225 | 2.96 | 0.948 | 1.212 | 1.367 | 1.702 | 2.114 |
| 200 | 1.038 | 1.495 | 1.752 | 2.356 | 3.189 | 0.972 | 1.294 | 1.479 | 1.892 | 2.417 |

A1.3. *A simulation study.* We now conduct a simulation study to demonstrate the potential gain from trimming, as well as investigating the inference for rerandomization in finite samples. We use again the dataset from the STAR project, and consider in total nine rerandomization designs for the $n_1 + n_0 = 974$ units, with number of covariates $K$ ranging from 0 to 200 and acceptance probability fixed at $p = 0.001$, where the covariates are generated in the same way as in Appendix A1.2. Note that when $K = 0$, rerandomization without any covariate essentially reduces to the CRE. We then simulate $10^5$ treatment assignments from each of these designs.

TABLE A2

*Properties of rerandomization with fixed acceptance probability $p_a = 0.001$ and varying number of covariates $K$. The 1st column shows the number of covariates for each design, the 2nd column shows the corresponding value of $1 - v_{K,a}$. The 3rd to 6th columns show the maximum standardized bias and mean squared error in (A1.2) and (A1.3), the summation of leverages to the power of $3/2$, and the maximum leverage over all units. The 7th–10th columns show the analogous quantities for the design using trimmed covariates. The 11th and 12th columns show the minimum possible values of $\sum_{i=1}^n H_{ii}^{3/2}$ (which equals $K^{3/2}/\sqrt{n}$) and $\max_i H_{ii}$ (which equals $K_n/n$) for each $K$.*

| $K$ | $1 - v_{K,a}$ | Covariates without trimming | | | | Trimmed covariates | | | | Minimal | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bias | RMSE | $\sum_{i=1}^n H_{ii}^{3/2}$ | $\max_i H_{ii}$ | Bias | RMSE | $\sum_{i=1}^n H_{ii}^{3/2}$ | $\max_i H_{ii}$ | $\sum_{i=1}^n H_{ii}^{3/2}$ | $\max_i H_{ii}$ |
| 0 | 0 | 0.10 | 1.10 | NA | NA | 0.10 | 1.10 | NA | NA | NA | NA |
| 5 | 0.97 | 0.13 | 1.10 | 0.39 | 0.02 | 0.11 | 1.10 | 0.39 | 0.02 | 0.36 | 0.01 |
| 9 | 0.90 | 0.76 | 1.18 | 1.78 | 0.66 | 0.16 | 1.10 | 0.94 | 0.03 | 0.87 | 0.01 |
| 15 | 0.80 | 0.91 | 1.25 | 3.48 | 0.66 | 0.20 | 1.10 | 1.97 | 0.05 | 1.86 | 0.02 |
| 24 | 0.70 | 1.28 | 1.46 | 8.00 | 0.98 | 0.23 | 1.11 | 3.92 | 0.06 | 3.77 | 0.02 |
| 37 | 0.60 | 1.33 | 1.52 | 12.71 | 0.98 | 0.25 | 1.11 | 7.41 | 0.09 | 7.21 | 0.04 |
| 60 | 0.50 | 1.47 | 1.62 | 23.45 | 0.98 | 0.27 | 1.11 | 15.13 | 0.13 | 14.89 | 0.06 |
| 100 | 0.41 | 1.56 | 1.72 | 43.87 | 0.98 | 0.29 | 1.12 | 32.33 | 0.16 | 32.04 | 0.10 |
| 200 | 0.30 | 1.62 | 1.79 | 104.57 | 0.99 | 0.31 | 1.13 | 91.01 | 0.28 | 90.63 | 0.21 |

Table A2 reports the simulation results. The first column shows the number of covariates involved in the nine rerandomization designs, and the second column shows the value of $1 - v_{K,a}$ under various values of $K$ and fixed acceptance probability $p = 0.001$. From Corollary 1, if the additional covariates do not increase the squared multiple correlation $R^2$ between potential outcomes and covariates by a relatively large amount, the improvement from rerandomization may decrease as the number of covariates increases. The 3rd–6th columns in Table A2 show the maximum absolute bias in (A1.2), the maximum root MSE in (A1.3), the summation of leverages to the power of $3/2$ as in (18) and the maximum leverage over all units, where the first two are estimated based on the $10^5$ simulated assignments from each of these designs. Note that when $K = 0$, the design is the CRE, and from the discussion after Proposition A1, the maximum mean is 0 and the maximum root MSE is 1. Thus, there is some variability for estimating the maximum mean and MSE; in practice, we may increase the number of simulated assignments to improve the precision. Nevertheless, the 3rd–6th columns in Table A2 show the trend that, as the number of covariates increases, the maximum bias and MSE under rerandomization will increase, which may render our treatment effect estimation inaccurate, and the leverages will increase as well, which may make the asymptotic approximation less accurate as discussed shortly. We further perform trimming on the covariates, as suggested in Section 7.2. Specifically, we trim each covariate at both its $2.5\%$ and $97.5\%$ quantiles. From the 7th–10th columns in Table A2, trimming significantly reduces the maximum biases, maximum MSEs and leverages. Moreover, compared to the 11th and 12th columns, the values of $\sum_{i=1}^{n} H_{ii}^{3/2}$ and $\max_i H_{ii}$ after trimming become quite close to their minimal possible values. This agrees with the suggestion we gave in Section 7.2.

TABLE A3

*Asymptotic approximation and coverage property under the nine rerandomization designs in Table A2. The top half uses the first-year GPA from the STAR dataset as the potential outcomes, and the bottom half use the average propensity score from the nine design without trimming, after a quantile transformation using $t$ distribution with degree of freedom 3, as the potential outcomes. The $K$ column shows the number of covariates in each design, Bias column shows the absolute empirical bias standardized by $V_{\tau\tau}^{1/2}$, Ratio column shows the ratio between empirical and asymptotic mean squared errors, and HC0–3 columns show the coverage probabilities in percent of the $95\%$ confidence intervals using the methods HC0–3 described in Section 6.*

| $K$ | Covariates without trimming | | | | | | Trimmed covariates | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Bias | Ratio | HC0 | HC1 | HC2 | HC3 | Bias | Ratio | HC0 | HC1 | HC2 | HC3 |
| 0 | 0.001 | 1.00 | 97.5 | 97.5 | 97.5 | 97.5 | 0.000 | 1.00 | 97.4 | 97.4 | 97.4 | 97.4 |
| 5 | 0.002 | 1.00 | 97.3 | 97.5 | 97.5 | 97.8 | 0.000 | 1.00 | 97.2 | 97.4 | 97.5 | 97.7 |
| 9 | 0.012 | 1.01 | 96.9 | 97.3 | 97.3 | 97.8 | 0.004 | 1.01 | 97.0 | 97.5 | 97.5 | 97.9 |
| 15 | 0.012 | 1.02 | 96.9 | 97.5 | 97.5 | 98.1 | 0.011 | 1.02 | 96.7 | 97.3 | 97.3 | 97.9 |
| 24 | 0.018 | 1.02 | 96.7 | 97.4 | 97.4 | 98.2 | 0.006 | 1.01 | 96.5 | 97.3 | 97.4 | 98.0 |
| 37 | 0.003 | 1.03 | 96.2 | 97.3 | 97.3 | 98.0 | 0.006 | 1.02 | 96.1 | 97.2 | 97.2 | 97.9 |
| 60 | 0.012 | 1.04 | 95.5 | 97.1 | 97.2 | 97.7 | 0.020 | 1.04 | 95.3 | 97.1 | 97.1 | 97.6 |
| 100 | 0.008 | 1.05 | 94.4 | 96.8 | 96.8 | 97.4 | 0.014 | 1.04 | 94.3 | 96.8 | 96.8 | 97.3 |
| 200 | 0.006 | 1.07 | 93.9 | 94.0 | 94.0 | 94.4 | 0.015 | 1.06 | 93.7 | 93.8 | 93.8 | 93.9 |
| 0 | 0.005 | 0.99 | 97.5 | 97.5 | 97.5 | 97.5 | 0.002 | 1.00 | 97.5 | 97.5 | 97.5 | 97.5 |
| 5 | 0.013 | 1.01 | 97.2 | 97.3 | 97.3 | 97.4 | 0.003 | 1.00 | 97.2 | 97.4 | 97.4 | 97.4 |
| 9 | 0.346 | 1.03 | 93.2 | 93.9 | 94.3 | 95.0 | 0.021 | 1.01 | 96.6 | 97.0 | 97.1 | 97.5 |
| 15 | 0.513 | 1.16 | 89.5 | 90.9 | 91.5 | 92.6 | 0.032 | 1.02 | 96.1 | 96.8 | 96.8 | 97.4 |
| 24 | 0.838 | 1.57 | 79.2 | 82.1 | 83.4 | 86.0 | 0.048 | 1.02 | 95.6 | 96.5 | 96.6 | 97.4 |
| 37 | 0.923 | 1.73 | 74.9 | 79.0 | 80.4 | 83.1 | 0.056 | 1.03 | 95.1 | 96.4 | 96.4 | 97.4 |
| 60 | 1.022 | 1.92 | 68.8 | 74.9 | 76.1 | 78.7 | 0.062 | 1.03 | 94.6 | 96.4 | 96.5 | 97.3 |
| 100 | 1.094 | 2.11 | 62.7 | 70.8 | 71.5 | 74.5 | 0.054 | 1.04 | 93.4 | 96.0 | 96.0 | 96.8 |
| 200 | 1.065 | 2.11 | 63.1 | 63.6 | 64.1 | 68.3 | 0.067 | 1.07 | 92.5 | 92.6 | 92.6 | 92.8 |

We then consider the asymptotic approximation and coverage probabilities of confidence intervals under these rerandomization designs with different numbers of covariates. We first consider the case where both potential outcomes are the same as the observed first year GPA from the STAR dataset. The top half of Table A3 shows the absolute empirical bias standardized by $V_{\tau\tau}^{1/2}$, the ratio between empirical MSE and the corresponding asymptotic variance, and the empirical coverage probabilities of $95\%$ confidence intervals using methods HC0–3 described in Section 6. Note that when $K = 200$, the number of covariates are greater than the size of treated group, under which we can only perform HC1–3 for the control group. From Table A3, under all the nine designs, the biases are close to 0, the ratios between empirical and asymptotic MSEs are close to 1, and the coverage probabilities are close to the nominal level, all of which indicate that the asymptotic approximation for rerandomization works quite well. These in some sense show the robustness of rerandomization. To illustrate the potential drawback of rerandomization with a large number of covariates, we also consider potential outcomes constructed in the following way: we first estimate the propensity scores for all units under these nine designs, then calculate the average of them for each unit, and finally take a quantile transformation using the $t$ distribution with degrees of freedom 3 to get both potential outcomes. The bottom half of Table A3 shows analogously the standardized absolute empirical bias, the ratio between empirical and asymptotic MSEs and the coverage probabilities of $95\%$ confidence intervals using HC0–3. From Table A3, as $K$ increases, the standardized biases increases, the ratio becomes further from 1, and the coverage probabilities becomes much smaller than the nominal level, all of which indicates poor asymptotic approximation under rerandomization with a large number of covariates. Comparing Tables A2 and A3, we can find that the ratio and the coverage probabilities become further off from their ideal values as the maximum standardized bias and root MSE in Table A2 get larger, which indicates that (A1.2) and (A1.3) can be used as viable tools to help assist the design of ReM in practice. Finally, we also consider rerandomization with trimmed covariates for both cases. From the right half of Table A3, trimming helps improve the finite-sample performance of rerandomization in terms of both point and interval estimates. Moreover, compared to HC0, HC1–3 help improve the coverage probabilities of the confidence intervals, especially when the number of covariates is relatively large.

**A2. Regression adjustment under rerandomization.** Regression adjustment is a popular approach to adjusting for covariate imbalance between two treatment groups after the experiments were conducted. Below we consider linearly regression adjusted estimator after ReM, which can be particularly useful when the analyzer is able to observe more covariate information after conducting the experiment. Let $\boldsymbol{W}_i \in \mathbb{R}^{J_n}$ denote the available covariate vector for unit $i$ in analysis, and $\hat{\boldsymbol{\tau}}_{\boldsymbol{W}}$ denote the corresponding difference-in-means of covariates between treatment and control groups. Following Li and Ding (2020), a general linearly regression adjusted estimator has the following form:

(A2.5)

$$\begin{aligned}\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0) &= \frac{1}{n_1}\sum_{i=1}^{n} Z_i\{Y_i - \boldsymbol{\beta}_1^\top(\boldsymbol{W}_i - \bar{\boldsymbol{W}})\} - \frac{1}{n_0}\sum_{i=1}^{n}(1-Z_i)\{Y_i - \boldsymbol{\beta}_0^\top(\boldsymbol{W}_i - \bar{\boldsymbol{W}})\} \\ &= \hat{\tau} - (r_0\boldsymbol{\beta}_1 + r_1\boldsymbol{\beta}_0)^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{W}},\end{aligned}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$ are the covariate adjustment coefficients. From (A2.5), the regression adjusted estimator $\hat{\tau}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)$ is essentially the difference-in-means estimator with adjusted treatment and control potential outcomes $Y_i(1) - \boldsymbol{\beta}_1^\top(\boldsymbol{W}_i - \bar{\boldsymbol{W}})$'s and $Y_i(0) - \boldsymbol{\beta}_0^\top(\boldsymbol{W}_i - \bar{\boldsymbol{W}})$'s. Therefore, its asymptotic property can be similarly derived as Theorems 3 and 5. Below we focus on a specific regression adjusted estimator $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$, which enjoys certain optimalities (see, e.g., Lin, 2013; Li and Ding, 2020) and uses the following least squares coefficients for covariate adjustment:

$$\tilde{\boldsymbol{\beta}}_z = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n}\{Y_i(z) - \bar{Y}(z) - \boldsymbol{\beta}_z^\top(\boldsymbol{W}_i - \bar{\boldsymbol{W}})\}^2 = (\boldsymbol{S}_{\boldsymbol{W}}^2)^{-1}\boldsymbol{S}_{\boldsymbol{W},z}, \quad (z = 0, 1)$$

where $\boldsymbol{S}_{\boldsymbol{W}}^2$ and $\boldsymbol{S}_{\boldsymbol{W},z}$ denote the finite population covariance matrices for covariates and potential outcomes.

For each unit $i$ and $z = 0, 1$, let $\tilde{e}_i(z) = Y_i(z) - \tilde{\boldsymbol{\beta}}_z^\top (\boldsymbol{W}_i - \bar{\boldsymbol{W}})$ denote the adjusted potential outcome, and $\tilde{\boldsymbol{u}}_i = (r_0 \tilde{e}_i(1) + r_1 \tilde{e}_i(0), \boldsymbol{X}_i^\top)^\top$. Define $\tilde{\gamma}_n$ and $\tilde{\Delta}_n$ analogously as (7) and (8), but with $Y_i(z)$ replaced by $\tilde{e}_i(z)$, $\boldsymbol{u}_i$ replaced by $\tilde{\boldsymbol{u}}_i$ and $\hat{\tau}$ replaced by $\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$. Analogous to (5), we define $\tilde{R}_n^2$ as the squared multiple correlation between adjusted potential outcomes and covariates $\boldsymbol{X}_i$'s, and $\rho_n^2$ as the squared multiple correlation between original (unadjusted) potential outcomes and covariates $\boldsymbol{W}_i$'s. We first invoke the following regularity condition, which essentially assumes Conditions 1 and 2 for the adjusted potential outcomes.

CONDITION A1. Conditions 1 and 2 hold with $\gamma_n$ and $\Delta_n$ replaced by $\tilde{\gamma}_n$ and $\tilde{\Delta}_n$.

Note that both adjustment coefficients $\tilde{\boldsymbol{\beta}}_1$ and $\tilde{\boldsymbol{\beta}}_0$ depend on all potential outcomes and are thus unknown. In practice, we can estimate them using the sampling analogues $\hat{\boldsymbol{\beta}}_z = (\boldsymbol{S}_{\boldsymbol{W}}^2)^{-1} \boldsymbol{s}_{z,\boldsymbol{W}}$ for $z = 0, 1$, where $\boldsymbol{s}_{z,\boldsymbol{W}}$ is the sample covariance between observed outcomes and covariates for units under treatment arm $z$, and use the regression adjusted estimator $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ with the estimated coefficients. We then invoke the following regularity condition, which ensures that the regression adjusted estimators with true and estimated coefficients have the same asymptotic distribution.

CONDITION A2. As $n \to \infty$,

$$\frac{\max_{z \in \{0,1\}} \max_{1 \le i \le n} |Y_i(z) - \bar{Y}(z)|}{\sqrt{V_{\tau\tau}(1 - \rho_n^2)\{1 - \tilde{R}_n^2\}}} \cdot J_n \cdot \frac{\max\{1, \log J_n, -\log p_n\}}{n r_1^2 r_0^2} \to 0.$$

We summarize the asymptotic distribution of $\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)$ under ReM in the theorem below.

THEOREM A1. *Under ReM and Conditions A1 and A2, as $n \to \infty$,*

$$\sup_{c \in \mathbb{R}} \left| \mathbb{P}\left\{ \frac{\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0) - \tau}{\sqrt{V_{\tau\tau}(1 - \rho_n^2)}} \le c \mid M \le a_n \right\} - \mathbb{P}\left\{ \sqrt{1 - \tilde{R}_n^2}\, \varepsilon_0 + \sqrt{\tilde{R}_n^2}\, L_{K_n, a_n} \le c \right\} \right| \to 0;$$

*If further Condition 3 holds and* $\limsup_{n \to \infty} \tilde{R}_n^2 < 1$, *then*

$$\sup_{c \in \mathbb{R}} \left| \mathbb{P}\left\{ \frac{\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0) - \tau}{\sqrt{V_{\tau\tau}(1 - \rho_n^2)}} \le c \mid M \le a_n \right\} - \mathbb{P}\left\{ \sqrt{1 - \tilde{R}_n^2}\, \varepsilon_0 \le c \right\} \right| \to 0.$$

Theorem A1 implies that we can still perform covariate adjustment under ReM with diminishing covariate imbalance threshold as well as diverging numbers of covariates in both design and analysis, which extends the discussion in Li and Ding (2020) with fixed threshold and fixed numbers of covariates. Moreover, with covariate imbalance diminishing at a proper rate, the regression adjusted estimator becomes asymptotically Gaussian distributed, and its improvement over the CRE is nondecreasing in $\tilde{R}_n^2$.

### A3. Berry–Esseen-type Bound for Finite Population Central Limit Theorem in Simple Random Sampling.

A3.1. *Main theorem.*

THEOREM A2. *Consider any finite population* $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_N\}$ *with* $\boldsymbol{u}_i \in \mathbb{R}^d$*, with* $\bar{\boldsymbol{u}} = N^{-1}\sum_{i=1}^N \boldsymbol{u}_i$ *and* $\boldsymbol{S}^2 = (N-1)^{-1}\sum_{i=1}^N (\boldsymbol{u}_i - \bar{\boldsymbol{u}})(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top$ *denoting the finite population average and covariance matrix. Let* $(Z_1, Z_2, \dots, Z_N)$ *denote the indicators for a simple random sample of size* $m$*, i.e., the probability that* $\boldsymbol{Z}$ *takes a particular value* $\boldsymbol{z} = (z_1, \dots, z_N) \in \{0,1\}^N$ *is* $m!(N-m)!/N!$ *if* $\sum_{i=1}^N z_i = m$ *and zero otherwise,* $f \equiv m/N$ *be the fraction of sampled units, and*

$$\boldsymbol{W} = \frac{1}{\sqrt{Nf(1-f)}}\boldsymbol{S}^{-1}\left\{\sum_{i=1}^N Z_i \boldsymbol{u}_i - m\bar{\boldsymbol{u}}\right\}.$$

*Let* $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$ *denote a* $d$*-dimensional standard Gaussian random vector, and define*

$$\gamma \equiv \frac{1}{\sqrt{Nf(1-f)}}\frac{d^{1/4}}{N}\sum_{i=1}^N \left\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\right\|_2^3.$$

(i) *There exists* $C_d$ *that depends only on* $d$ *such that for any* $N \ge 2$*,* $f \in (0,1)$*, any finite population* $\{\boldsymbol{u}_i : 1 \le i \le N\}$ *with nonsingular finite population covariance* $\boldsymbol{S}^2$*, and any measurable convex set* $\mathcal{Q} \subset \mathbb{R}^d$*,*

$$|\mathbb{P}(\boldsymbol{W} \in \mathcal{Q}) - \mathbb{P}(\boldsymbol{\varepsilon} \in \mathcal{Q})| \le C_d \gamma$$

(ii) *If the theorem in Raič (2015) holds, then there exists a universal constant* $C$ *such that for any* $N \ge 2$*,* $d \ge 1$*,* $f \in (0,1)$*, any finite population* $\{\boldsymbol{u}_i : 1 \le i \le N\}$ *with nonsingular finite population covariance* $\boldsymbol{S}^2$*, and any measurable convex set* $\mathcal{Q} \subset \mathbb{R}^d$*,*

$$|\mathbb{P}(\boldsymbol{W} \in \mathcal{Q}) - \mathbb{P}(\boldsymbol{\varepsilon} \in \mathcal{Q})| \le C\gamma$$

(iii) *For any* $N \ge 2$*,* $d \ge 1$*,* $f \in (0,1)$*, any finite population* $\{\boldsymbol{u}_i : 1 \le i \le N\}$ *with nonsingular finite population covariance* $\boldsymbol{S}^2$*, and any measurable convex set* $\mathcal{Q} \subset \mathbb{R}^d$*,*

$$|\mathbb{P}(\boldsymbol{W} \in \mathcal{Q}) - \mathbb{P}(\boldsymbol{\varepsilon} \in \mathcal{Q})| \le 174\gamma + 3 \cdot 2^{2/3}\frac{d^{1/2}}{\{Nf(1-f)\}^{1/6}} \le 174\gamma + 7\gamma^{1/3}.$$

(iv) *For any* $N \ge 2$*,* $d \ge 1$*,* $f \in (0,1)$*, any finite population* $\{\boldsymbol{u}_i : 1 \le i \le N\}$ *with nonsingular finite population covariance* $\boldsymbol{S}^2$*, and any measurable convex set* $\mathcal{Q} \subset \mathbb{R}^d$*,*

$$|\mathbb{P}(\boldsymbol{W} \in \mathcal{Q}) - \mathbb{P}(\boldsymbol{\varepsilon} \in \mathcal{Q})| \le 180\gamma + \frac{3(\log N)^{3/4}d^{3/4}}{N^{1/4}\sqrt{f(1-f)}} \cdot \max_{1 \le i \le n}\left\|\left(\boldsymbol{S}_{\boldsymbol{u}}^2\right)^{-1/2}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\right\|_\infty.$$

(v) *For any* $N \ge 2$*,* $d \ge 1$*,* $f \in (0,1)$*, any finite population* $\{\boldsymbol{u}_i : 1 \le i \le N\}$ *with nonsingular finite population covariance* $\boldsymbol{S}^2$*, any* $\iota \ge 2$*, and any measurable convex set* $\mathcal{Q} \subset \mathbb{R}^d$*,*

$$|\mathbb{P}(\boldsymbol{W} \in \mathcal{Q}) - \mathbb{P}(\boldsymbol{\varepsilon} \in \mathcal{Q})| \le 174\gamma + \frac{C_\iota d^{3\iota/\{4(\iota+1)\}}}{N^{\iota/\{4(\iota+1)\}}\{f(1-f)\}^{\iota/2}}\frac{1}{N}\sum_{i=1}^N \left\|\left(\boldsymbol{S}_{\boldsymbol{u}}^2\right)^{-1/2}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\right\|_\iota^\iota,$$

*where* $C_\iota$ *is a universal constant depending only on* $\iota$.

A3.2. *Proof of Theorem A2(i) and (ii) based on combinatorial central limit theorem.* To prove Theorem A2(i), we need the following lemma, which follows immediately from Bolthausen and Götze (1993, Theorem 1).

LEMMA A1. *Consider any integer $N \ge 1$ and any constant vector $\boldsymbol{a}(i,j) \in \mathbb{R}^d$ for all $1 \le i, j \le N$ satisfying that*

(A3.6)
$$\sum_{j=1}^N \boldsymbol{a}(i,j) = \boldsymbol{0}, 1 \le i \le N \ \ \text{and} \ \ \frac{1}{N-1}\sum_{i=1}^N\sum_{j=1}^N \boldsymbol{a}(i,j)\boldsymbol{a}(i,j)^\top - \frac{1}{N(N-1)}\sum_{j=1}^N \boldsymbol{b}(j)\boldsymbol{b}(j)^\top = \boldsymbol{I}_d,$$

*where $\boldsymbol{b}(j) \equiv \sum_{i=1}^N \boldsymbol{a}(i,j)$. Let $\pi$ denote a uniformly distributed random permutation of $\{1,2,\dots,N\}$, $\boldsymbol{W} = \sum_{i=1}^N \boldsymbol{a}(i,\pi(i))$, and $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$ denote a $d$-dimensional standard Gaussian random vector. Then there exists a constant $C_d$ that depends only on $d$ such that for any $N \ge 2$, any $\{\boldsymbol{a}(i,j) : 1 \le i,j \le N\}$ satisfying (A3.6), and any measurable convex set $\mathcal{Q} \subset \mathbb{R}^d$,*

$$|\mathbb{P}(\boldsymbol{W} \in \mathcal{Q}) - \mathbb{P}(\boldsymbol{\varepsilon} \in \mathcal{Q})| \le C_d \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \|\boldsymbol{a}(i,j)\|_2^3.$$

**Proof of Theorem A2(i).** Let $\boldsymbol{S}$ denote the positive definite square root of $\boldsymbol{S}^2$, and define

$$\boldsymbol{a}(i,j) = \begin{cases} \{m(1-f)\}^{-1/2}\boldsymbol{S}^{-1}(1-f)(\boldsymbol{u}_i - \bar{\boldsymbol{u}}), & \text{if } 1 \le j \le m, \\ -\{m(1-f)\}^{-1/2}\boldsymbol{S}^{-1}f(\boldsymbol{u}_i - \bar{\boldsymbol{u}}), & \text{if } m < j \le N, \end{cases} \quad (1 \le i \le N).$$

We can then verify that

$$\sum_{j=1}^N \boldsymbol{a}(i,j) = \{m(1-f)\}^{-1/2}\boldsymbol{S}^{-1}\cdot\{m(1-f) - (N-m)f\}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) = \boldsymbol{0}, \quad (1 \le i \le N)$$

$$\boldsymbol{b}(j) \equiv \sum_{i=1}^N \boldsymbol{a}(i,j) = \boldsymbol{0}, \quad (1 \le j \le N)$$

and

$$\frac{1}{N-1}\sum_{i=1}^N\sum_{j=1}^N \boldsymbol{a}(i,j)\boldsymbol{a}(i,j)^\top - \frac{1}{N(N-1)}\sum_{j=1}^N \boldsymbol{b}(j)\boldsymbol{b}(j)^\top$$
$$= \frac{m}{N-1}\frac{(1-f)^2}{m(1-f)}\boldsymbol{S}^{-1}\sum_{i=1}^N(\boldsymbol{u}_i - \bar{\boldsymbol{u}})(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top\boldsymbol{S}^{-1} + \frac{N-m}{N-1}\frac{f^2}{m(1-f)}\boldsymbol{S}^{-1}\sum_{i=1}^N(\boldsymbol{u}_i - \bar{\boldsymbol{u}})(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top\boldsymbol{S}^{-1}$$
$$= (1-f)\boldsymbol{S}^{-1}\boldsymbol{S}^2\boldsymbol{S}^{-1} + f\boldsymbol{S}^{-1}\boldsymbol{S}^2\boldsymbol{S}^{-1} = \boldsymbol{I}_d,$$

i.e., $\{\boldsymbol{a}(i,j) : 1 \le i,j \le N\}$ satisfies the conditions in (A3.6). Besides, $\sum_{i=1}^N\sum_{j=1}^N |\boldsymbol{a}(i,j)|^3$ simplifies to

$$\sum_{i=1}^N\sum_{j=1}^N \|\boldsymbol{a}(i,j)\|_2^3 = m\frac{(1-f)^3}{\{m(1-f)\}^{3/2}}\sum_{i=1}^N \left\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\right\|_2^3 + (N-m)\frac{f^3}{\{m(1-f)\}^{3/2}}\sum_{i=1}^N \left\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\right\|_2^3$$
$$= \frac{(1-f)^2 + f^2}{\sqrt{m(1-f)}}\sum_{i=1}^N \left\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\right\|_2^3 \le \frac{\{f + (1-f)\}^2}{\sqrt{Nf(1-f)}}\sum_{i=1}^N \left\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\right\|_2^3$$
$$= \frac{1}{\sqrt{Nf(1-f)}}\sum_{i=1}^N \left\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\right\|_2^3 = \frac{N}{d^{1/4}}\gamma,$$

where the last equality holds by definition.

Let $\pi$ denote a uniformly distributed random permutation of $\{1,2,\ldots,N\}$, and $\tilde{\boldsymbol{W}} = \sum_{i=1}^{N} \boldsymbol{a}(i,\pi(i))$. Then, by definition,

$$
\begin{aligned}
\tilde{\boldsymbol{W}} &= \sum_{i=1}^{N} \mathbb{1}(\pi(i) \le m)\{m(1-f)\}^{-1/2}\boldsymbol{S}^{-1}(1-f)\boldsymbol{u}_i - \sum_{i=1}^{N} \mathbb{1}(\pi(i) > m)\{m(1-f)\}^{-1/2}\boldsymbol{S}^{-1} f\boldsymbol{u}_i \\
&= \{m(1-f)\}^{-1/2}\boldsymbol{S}^{-1}\sum_{i=1}^{N}\{\mathbb{1}(\pi(i)\le m) - f\}\,\boldsymbol{u}_i = \frac{1}{\sqrt{Nf(1-f)}}\boldsymbol{S}^{-1}\left\{\sum_{i=1}^{N}\mathbb{1}(\pi(i)\le m)\boldsymbol{u}_i - m\bar{\boldsymbol{u}}\right\} \\
&\sim \frac{1}{\sqrt{Nf(1-f)}}\boldsymbol{S}^{-1}\left\{\sum_{i=1}^{N} Z_i\boldsymbol{u}_i - m\bar{\boldsymbol{u}}\right\} = \boldsymbol{W},
\end{aligned}
$$

where the last $\sim$ holds because $(\mathbb{1}(\pi(1)\le m), \mathbb{1}(\pi(2)\le m)), \ldots, \mathbb{1}(\pi(N)\le m))$ follows the same distribution as $(Z_1, Z_2, \ldots, Z_N)$. Applying Lemma A1, we can know that there exists $C_d$ that depends only on $d$ such that, for any measurable convex set $\mathcal{Q} \subset \mathbb{R}^d$,

$$
|\mathbb{P}(\boldsymbol{W}\in\mathcal{Q}) - \mathbb{P}(\boldsymbol{\varepsilon}\in\mathcal{Q})| \le C_d \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\boldsymbol{a}(i,j)\|_2^3 = C_d d^{-1/4}\cdot\gamma.
$$

This immediately implies that Theorem A2(i) holds. □

To prove Theorem A2(ii), we need the following lemma from Raič (2015). However, the author did not provide a formal proof there.

LEMMA A2. *Consider the same setting as in Lemma A1. There exists a universal constant $C$ such that for any $N \ge 2$, any $d \ge 1$, any $\{\boldsymbol{a}(i,j) : 1 \le i,j \le N\}$ satisfying (A3.6), and any measurable convex set $\mathcal{Q}\subset\mathbb{R}^d$,*

$$
|\mathbb{P}(\boldsymbol{W}\in\mathcal{Q}) - \mathbb{P}(\boldsymbol{\varepsilon}\in\mathcal{Q})| \le C d^{1/4}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|\boldsymbol{a}(i,j)|^3.
$$

**Proof of Theorem A2(ii).** Theorem A2(ii) follows from Lemma A2, by almost the same logic as the proof of Theorem A2(i). Therefore, we omit its proof here. □

A3.3. *Proof of Theorem A2(iii) based on Hájek coupling.*

A3.3.1. *Technical lemmas.* To prove Theorem A2(iii), we need the following eight lemmas.

LEMMA A3. *Consider a finite population $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_N\}$ for $N$ units, where $\boldsymbol{u}_i \in \mathbb{R}^d$ for all $i$. Let $\bar{\boldsymbol{u}} = N^{-1}\sum_{i=1}^{N}\boldsymbol{u}_i$ denote the finite population average. There must exist a pair of random vectors $\boldsymbol{Z}$ and $\boldsymbol{T}$ in $\{0,1\}^N$ such that*

(i) *$\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_N)$ is an indicator vector for a simple random sample of size $m$, i.e., the probability that $\boldsymbol{Z}$ takes a particular value $\boldsymbol{z} = (z_1, \ldots, z_N) \in \{0,1\}^N$ is $m!(N-m)!/N!$ if $\sum_{i=1}^{N} z_i = m$ and zero otherwise;*
(ii) *$\boldsymbol{T} = (T_1, T_2, \ldots, T_N) \in \{0,1\}^N$ is an indicator vector for a Bernoulli random sample with equal probability $m/N$ for all units, i.e., $T_i \overset{i.i.d.}{\sim}$ Bern$(m/N)$;*

(iii) *the covariances for* $\boldsymbol{A} \equiv \sum_{i=1}^N Z_i \boldsymbol{u}_i$, $\boldsymbol{B} \equiv \sum_{i=1}^N T_i(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) + m\bar{\boldsymbol{u}}$ *and their difference satisfy* $\mathit{Cov}(\boldsymbol{B}) = (1 - N^{-1}) \cdot \mathit{Cov}(\boldsymbol{A})$ *and*

$$\mathit{Cov}^{-1/2}(\boldsymbol{B}) \cdot \mathbb{E}\big\{(\boldsymbol{B}-\boldsymbol{A})(\boldsymbol{B}-\boldsymbol{A})^\top\big\} \cdot \mathit{Cov}^{-1/2}(\boldsymbol{B}) \le \sqrt{\frac{1}{m} + \frac{1}{N-m}} \cdot \boldsymbol{I}_d.$$

LEMMA A4 (Raič (2019)). *Let* $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_N$ *be* $N$ *independent* $d$*-dimensional random vectors, satisfying* $\mathbb{E}\boldsymbol{\xi}_i = \boldsymbol{0}$ *for all* $1 \le i \le N$ *and* $\sum_{i=1}^N \mathit{Cov}(\boldsymbol{\xi}_i) = \boldsymbol{I}_d$, *and* $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$ *be a* $d$*-dimensional standard Gaussian random vector. Define* $\boldsymbol{W} = \sum_{i=1}^N \boldsymbol{\xi}_i$. *Then for any measurable convex set* $\mathcal{Q} \subset \mathbb{R}^d$,

$$|\mathbb{P}(\boldsymbol{W} \in \mathcal{Q}) - \mathbb{P}(\boldsymbol{\varepsilon} \in \mathcal{Q})| \le 58 d^{1/4} \sum_{i=1}^N \mathbb{E}\|\boldsymbol{\xi}_i\|_2^3.$$

LEMMA A5. *Let* $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_N\}$ *be a finite population of* $N$ *units, with* $\boldsymbol{u}_i \in \mathbb{R}^d$ *for all* $i$, *and* $\boldsymbol{T} = (T_1, T_2, \dots, T_N) \in \{0,1\}^N$ *be an indicator vector for a Bernoulli random sample with equal probability* $f \equiv m/N$ *for all units, i.e.,* $T_i \overset{i.i.d.}{\sim} \mathit{Bern}(f)$. *Define* $\bar{\boldsymbol{u}} = N^{-1}\sum_{i=1}^N \boldsymbol{u}_i$, $\boldsymbol{S}^2 = (N-1)^{-1}\sum_{i=1}^N (\boldsymbol{u}_i - \bar{\boldsymbol{u}})(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top$, *and* $\boldsymbol{B} \equiv \sum_{i=1}^N T_i(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) + m\bar{\boldsymbol{u}}$. *Let* $\boldsymbol{\varepsilon}$ *be a* $d$*-dimensional standard Gaussian random vector. Then for any* $N \ge 2$, $d \ge 1$, $f \in (0,1)$, *and any finite population* $\{\boldsymbol{u}_i : 1 \le i \le N\}$ *with nonsingular finite population covariance* $\boldsymbol{S}^2$, *we have, for any measurable convex set* $\mathcal{Q} \subset \mathbb{R}^d$,

$$\big|\mathbb{P}\big\{\mathit{Cov}^{-1/2}(\boldsymbol{B}) \cdot (\boldsymbol{B} - \mathbb{E}\boldsymbol{B}) \in \mathcal{Q}\big\} - \mathbb{P}(\boldsymbol{\varepsilon} \in \mathcal{Q})\big| \le \frac{165}{\sqrt{Nf(1-f)}} \frac{d^{1/4}}{N} \sum_{i=1}^N \big\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\big\|_2^3.$$

LEMMA A6. *Let* $\mathcal{Q} \subset \mathbb{R}^d$ *be a convex set in* $\mathbb{R}^d$.

(i) *For any* $c > 0$, $\overline{\mathcal{Q}}_c \equiv \{\boldsymbol{x} \in \mathbb{R}^d : \exists \boldsymbol{x}' \in \mathcal{Q} \text{ s.t. } \|\boldsymbol{x} - \boldsymbol{x}'\|_2 < c\}$ *is a convex set in* $\mathbb{R}^d$.
(ii) *For any* $c > 0$, $\underline{\mathcal{Q}}_c \equiv \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}' - \boldsymbol{x}\|_2 \ge c \ \forall \boldsymbol{x}' \notin \mathcal{Q}\}$ *is a convex set in* $\mathbb{R}^d$.
(iii) *For any matrix* $\boldsymbol{\Delta} \in \mathbb{R}^{d \times d}$, $\tilde{\mathcal{Q}} \equiv \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{\Delta}\boldsymbol{x} \in \mathcal{Q}\}$ *is a convex set in* $\mathbb{R}^d$.

LEMMA A7. *For any set* $\mathcal{Q} \subset \mathbb{R}^d$ *and any* $c > 0$, *define*

$$\overline{\mathcal{Q}}_c \equiv \{\boldsymbol{x} \in \mathbb{R}^d : \exists \boldsymbol{x}' \in \mathcal{Q} \text{ s.t. } \|\boldsymbol{x} - \boldsymbol{x}'\|_2 < c\}, \quad \underline{\mathcal{Q}}_c \equiv \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}' - \boldsymbol{x}\|_2 \ge c \ \forall \boldsymbol{x}' \notin \mathcal{Q}\}.$$

(i) *For any set* $\mathcal{Q} \subset \mathbb{R}^d$ *and any positive* $c, h$, $\overline{(\overline{\mathcal{Q}}_c)}_h = \overline{\mathcal{Q}}_{c+h}$.
(ii) *For any set* $\mathcal{Q} \subset \mathbb{R}^d$ *and any* $c > 0$, $(\underline{\mathcal{Q}}_c)^\complement = \overline{\mathcal{B}}_c$, *where* $\mathcal{B} = \mathcal{Q}^\complement$.
(iii) *For any set* $\mathcal{Q} \subset \mathbb{R}^d$ *and any positive* $c, h$, $\underline{(\underline{\mathcal{Q}}_c)}_h = \underline{\mathcal{Q}}_{c+h}$.

LEMMA A8. *Let* $\boldsymbol{\varepsilon}$ *be a* $d$*-dimensional standard Gaussian random variable,* $\phi_d(\cdot)$ *be the probability density function of* $\boldsymbol{\varepsilon}$, *and* $\mathcal{C}_d$ *be the collection of convex sets in* $\mathbb{R}^d$. *We have that*

$$\sup_{c>0, \mathcal{Q} \in \mathcal{C}_d} \frac{\int_{\overline{\mathcal{Q}}_c \setminus \mathcal{Q}} \phi_d(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon}}{c} \le 4 d^{\frac{1}{4}}$$

*and*

$$\sup_{c>0, \mathcal{Q} \in \mathcal{C}_d} \frac{\int_{\mathcal{Q} \setminus \underline{\mathcal{Q}}_c} \phi_d(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon}}{c} \le 4 d^{\frac{1}{4}}.$$

LEMMA A9. *Let $\boldsymbol{B}$ and $\boldsymbol{A}$ be two $d$-dimensional random vectors with equal means $\mathbb{E}\boldsymbol{B} = \mathbb{E}\boldsymbol{A}$ and nonsingular covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{B}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{A}}$, and $\boldsymbol{\varepsilon}$ be a $d$-dimensional standard Gaussian random vector. Let $\mathcal{C}_d$ denote the collection of convex sets in $\mathbb{R}^d$. If $\boldsymbol{\Sigma}_{\boldsymbol{B}} = (1-l)^2\boldsymbol{\Sigma}_{\boldsymbol{A}}$ for some $l \in (0,1)$,*

$$\sup_{\mathcal{Q}\in\mathcal{C}_d} \left|\mathbb{P}\big(\boldsymbol{\Sigma}_{\boldsymbol{B}}^{-1/2}(\boldsymbol{B}-\mathbb{E}\boldsymbol{B}) \in \mathcal{Q}\big) - \mathbb{P}(\boldsymbol{\varepsilon}\in\mathcal{Q})\right| \le a \quad \textit{for some finite } a>0$$

*and*

$$\boldsymbol{\Sigma}_{\boldsymbol{B}}^{-1/2}\cdot\mathbb{E}\big\{(\boldsymbol{B}-\boldsymbol{A})(\boldsymbol{B}-\boldsymbol{A})^\top\big\}\cdot\boldsymbol{\Sigma}_{\boldsymbol{B}}^{-1/2} \le b^2\boldsymbol{I}_d \quad \textit{for some } b\in(0,1),$$

*then for any positive constants $c$ and $h$,*

$$\begin{aligned}
&\sup_{\mathcal{Q}\in\mathcal{C}_d} \left|\mathbb{P}\big(\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{A}-\mathbb{E}\boldsymbol{A}) \in \mathcal{Q}\big) - \mathbb{P}(\boldsymbol{\varepsilon}\in\mathcal{Q})\right| \\
&\le 4d^{1/4}(c+h) + 2d\cdot\exp\left(-\frac{h^2}{2dl^2}\right) + a + \mathbb{P}\big(\|\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{A}-\boldsymbol{B})\|_2 \ge c\big) \\
&\le 4d^{1/4}(c+h) + 2d\cdot\exp\left(-\frac{h^2}{2dl^2}\right) + a + \frac{(1-l)^2b^2d}{c^2}.
\end{aligned}$$

LEMMA A10. *Under the same setting as in Theorem A2,*

(i) $\sum_{i=1}^N \|\boldsymbol{S}^{-1}(\boldsymbol{u}_i-\bar{\boldsymbol{u}})\|_2^2 = (N-1)d$;
(ii) $N^{-1}\sum_{i=1}^N \|\boldsymbol{S}^{-1}(\boldsymbol{u}_i-\bar{\boldsymbol{u}})\|_2^3 \ge (d/2)^{3/2}$;
(iii) $\gamma$ *defined in Theorem A2 satisfies that*

$$\gamma \equiv \frac{1}{\sqrt{Nf(1-f)}}\frac{d^{1/4}}{N}\sum_{i=1}^N \left\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i-\bar{\boldsymbol{u}})\right\|_2^3 \ge \frac{d^{7/4}}{2^{3/2}\sqrt{Nf(1-f)}}.$$

A3.3.2. *Proofs of the lemmas.*

PROOF OF LEMMA A3. Let $\boldsymbol{T} = (T_1, T_2, \dots, T_N) \in \{0,1\}^N$ be an indicator vector for a Bernoulli random sample with equal probability $n/N$ for all units. Below we construct the indicator vector $\boldsymbol{Z}$ for a simple random sample of size $n$ based on $\boldsymbol{T}$. We consider the following three different cases depending on the size $\tilde{m}$ of the set $\mathcal{T} \equiv \{i : T_i = 1, 1 \le i \le N\}$, i.e., $\tilde{m} \equiv |\mathcal{T}|$.

(i) If $\tilde{m} = m$ by accident, we define $\boldsymbol{Z} = \boldsymbol{T}$;
(ii) If $\tilde{m} > m$, we introduce a set $\mathcal{D}$ to be a simple random sample of size $\tilde{m}-m$ from $\mathcal{T}$, and define $Z_i = 1$ if $i \in \mathcal{T}\setminus\mathcal{D}$ and 0 otherwise;
(iii) If $\tilde{m} < m$, we introduce a set $\mathcal{D}$ to be a simple random sample of size $m-\tilde{m}$ from $\{1,2,\dots,N\}\setminus\mathcal{T}$, and define $Z_i = 1$ if $i \in \mathcal{T}\cup\mathcal{D}$ and 0 otherwise.

We can verify that $\boldsymbol{Z}$ must be an indicator vector for a simple random sample of size $m$. This is essentially the coupling between simple random sampling and Bernoulli random sampling used in Hájek (1960).

By definition, $\boldsymbol{B}-\boldsymbol{A} = \sum_{i=1}^N (T_i - Z_i)(\boldsymbol{u}_i - \bar{\boldsymbol{u}})$. By the construction of $\boldsymbol{T}$ and $\boldsymbol{Z}$, conditioning on $\tilde{m}$, the difference between $\boldsymbol{B}$ and $\boldsymbol{A}$ is essentially the summation of a simple random sample of size

(A3.7)
$$\begin{cases} \tilde{m}-m \text{ from the population } \{\boldsymbol{u}_1-\bar{\boldsymbol{u}}, \boldsymbol{u}_2-\bar{\boldsymbol{u}}, \dots, \boldsymbol{u}_N-\bar{\boldsymbol{u}}\}, & \text{if } \tilde{m}\ge m; \\ m-\tilde{m} \text{ from the population } \{-(\boldsymbol{u}_1-\bar{\boldsymbol{u}}), -(\boldsymbol{u}_2-\bar{\boldsymbol{u}}), \dots, -(\boldsymbol{u}_N-\bar{\boldsymbol{u}})\}, & \text{if } \tilde{m}< m. \end{cases}$$

Let $\Delta = |\tilde{m} - m|$. By the property of simple random sampling, this difference satisfies $\mathbb{E}(\boldsymbol{B} - \boldsymbol{A} \mid \tilde{m}) = \boldsymbol{0}$ and

$$\text{Cov}(\boldsymbol{B} - \boldsymbol{A} \mid \tilde{m}) = \frac{\Delta}{N} \cdot \frac{N - \Delta}{N - 1} \sum_{i=1}^{N} (\boldsymbol{u}_i - \bar{\boldsymbol{u}})(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top \leq \frac{\Delta}{N} \sum_{i=1}^{N} (\boldsymbol{u}_i - \bar{\boldsymbol{u}})(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top.$$

By the law of total expectation and total variance, and the fact that $\{\mathbb{E}(\Delta)\}^2 \leq \mathbb{E}(\Delta^2) = \text{Var}(\tilde{m})$, we have $\mathbb{E}(\boldsymbol{B} - \boldsymbol{A}) = \mathbb{E}\{\mathbb{E}(\boldsymbol{B} - \boldsymbol{A} \mid \tilde{m})\} = \boldsymbol{0}$, and

$$\begin{aligned}
\text{Cov}(\boldsymbol{B} - \boldsymbol{A}) &= \mathbb{E}\left\{\text{Cov}(\boldsymbol{B} - \boldsymbol{A} \mid \tilde{m})\right\} + \text{Cov}\left\{\mathbb{E}(\boldsymbol{B} - \boldsymbol{A} \mid \tilde{m})\right\} \leq \frac{\mathbb{E}(\Delta)}{N} \sum_{i=1}^{N} (\boldsymbol{u}_i - \bar{\boldsymbol{u}})(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top \\
&\leq \frac{\sqrt{\text{Var}(\tilde{m})}}{N} \sum_{i=1}^{N} (\boldsymbol{u}_i - \bar{\boldsymbol{u}})(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top \\
&= \sqrt{N \frac{m}{N}\left(1 - \frac{m}{N}\right)} \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{u}_i - \bar{\boldsymbol{u}})(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top. \qquad \text{(A3.8)}
\end{aligned}$$

By the property of bernoulli sampling and simple random sampling, we can derive that

$$\text{Cov}(\boldsymbol{B}) = \frac{m}{N}\left(1 - \frac{m}{N}\right) \sum_{i=1}^{N} (\boldsymbol{u}_i - \bar{\boldsymbol{u}})(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top. \qquad \text{(A3.9)}$$

and

$$\text{Cov}(\boldsymbol{A}) = \frac{m(N - m)}{N(N - 1)} \sum_{i=1}^{N} (\boldsymbol{u}_i - \bar{\boldsymbol{u}})(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top = \frac{N}{N - 1} \text{Cov}(\boldsymbol{B}).$$

(A3.8), (A3.9) and the fact that $\mathbb{E}(\boldsymbol{B} - \boldsymbol{A}) = \boldsymbol{0}$ immediately imply that

$$\begin{aligned}
\text{Cov}^{-1/2}(\boldsymbol{B}) \cdot \mathbb{E}\big\{(\boldsymbol{B} - \boldsymbol{A})(\boldsymbol{B} - \boldsymbol{A})^\top\big\} \cdot \text{Cov}^{-1/2}(\boldsymbol{B}) &= \text{Cov}^{-1/2}(\boldsymbol{B}) \cdot \text{Cov}(\boldsymbol{B} - \boldsymbol{A}) \cdot \text{Cov}^{-1/2}(\boldsymbol{B}) \\
&\leq \sqrt{\frac{1}{m} + \frac{1}{N - m}} \cdot \boldsymbol{I}_d.
\end{aligned}$$

From the above, Lemma A3 holds. □

PROOF OF LEMMA A4. Lemma A4 follows immediately from Raič (2019, Theorem 1.1). □

PROOF OF LEMMA A5. By definition, we can derive that

$$\text{Cov}^{-1/2}(\boldsymbol{B}) \cdot (\boldsymbol{B} - \mathbb{E}\boldsymbol{B}) = \text{Cov}^{-1/2}(\boldsymbol{B}) \cdot \sum_{i=1}^{N} T_i(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) = \text{Cov}^{-1/2}(\boldsymbol{B}) \sum_{i=1}^{N} \big(T_i - f\big)(\boldsymbol{u}_i - \bar{\boldsymbol{u}}),$$

where the last equality holds due to the centering of the $\boldsymbol{u}_i$'s. Define $\boldsymbol{\xi}_i = (T_i - f)\text{Cov}^{-1/2}(\boldsymbol{B}) \cdot (\boldsymbol{u}_i - \bar{\boldsymbol{u}})$. We can verify that $\boldsymbol{\xi}_i$'s satisfy the condition in Lemma A4. Thus, from Lemma A4, for any measurable convex set $\mathcal{Q} \subset \mathbb{R}^d$,

$$\Big|\mathbb{P}\big\{\text{Cov}^{-1/2}(\boldsymbol{B}) \cdot (\boldsymbol{B} - \mathbb{E}\boldsymbol{B}) \in \mathcal{Q}\big\} - \mathbb{P}(\boldsymbol{\varepsilon} \in \mathcal{Q})\Big| \leq 58 d^{1/4} \sum_{i=1}^{N} \mathbb{E}\|\boldsymbol{\xi}_i\|_2^3. \qquad \text{(A3.10)}$$

By definition, $\mathbb{E}\{|T_i - f|^3\} = f(1-f)\{f^2 + (1-f)^2\}$. From (A3.9), $\text{Cov}(\boldsymbol{B}) = f(1-f)(N-1)\boldsymbol{S}^2$. We can then simplify $\sum_{i=1}^N \mathbb{E}\|\boldsymbol{\xi}_i\|_2^3$ as

$$\sum_{i=1}^N \mathbb{E}\|\boldsymbol{\xi}_i\|_2^3 = \sum_{i=1}^N \frac{\mathbb{E}\{|T_i - f|^3\}}{\{f(1-f)(N-1)\}^{3/2}} \big\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\big\|_2^3 = \frac{f^2 + (1-f)^2}{(N-1)^{3/2}\sqrt{f(1-f)}} \sum_{i=1}^N \big\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\big\|_2^3$$
$$\le \frac{2^{3/2}}{\sqrt{Nf(1-f)}} \frac{1}{N} \sum_{i=1}^N \big\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\big\|_2^3,$$

where the last inequality holds becomes $f^2 + (1-f)^2 \le \{f + (1-f)\}^2 = 1$ and $N - 1 \ge N/2$. From (A3.10), we then have, for any measurable convex set $\mathcal{Q} \subset \mathbb{R}^d$,

$$\big|\mathbb{P}\{\text{Cov}^{-1/2}(\boldsymbol{B}) \cdot (\boldsymbol{B} - \mathbb{E}\boldsymbol{B}) \in \mathcal{Q}\} - \mathbb{P}(\boldsymbol{\varepsilon} \in \mathcal{Q})\big| \le 58 d^{1/4} \frac{2^{3/2}}{\sqrt{Nf(1-f)}} \frac{1}{N} \sum_{i=1}^N \big\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\big\|_2^3$$
$$\le \frac{165}{\sqrt{Nf(1-f)}} \frac{d^{1/4}}{N} \sum_{i=1}^N \big\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\big\|_2^3$$

Therefore, Lemma A5 holds. □

PROOF OF LEMMA A6. We first prove (i). Consider any $\boldsymbol{x}, \boldsymbol{y} \in \overline{\mathcal{Q}}_c$ and any $\lambda \in (0,1)$. By definition, there must exist $\boldsymbol{x}', \boldsymbol{y}' \in \mathcal{Q}$ such that $\|\boldsymbol{x} - \boldsymbol{x}'\|_2 < c$ and $\|\boldsymbol{y} - \boldsymbol{y}'\|_2 < c$. Because $\mathcal{Q}$ is convex, $\lambda\boldsymbol{x}' + (1-\lambda)\boldsymbol{y}' \in \mathcal{Q}$. Moreover, by the triangle inequality,

$$\|\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} - \{\lambda\boldsymbol{x}' + (1-\lambda)\boldsymbol{y}'\}\|_2 \le \lambda\|\boldsymbol{x} - \boldsymbol{x}'\|_2 + (1-\lambda)\|\boldsymbol{y} - \boldsymbol{y}'\|_2 < c.$$

Thus, we must have $\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in \overline{\mathcal{Q}}_c$. Therefore, $\overline{\mathcal{Q}}_c$ must be a convex set.

We then prove (ii). Consider any $\boldsymbol{x}, \boldsymbol{y} \in \underline{\mathcal{Q}}_c$ and any $\lambda \in (0,1)$. We prove that $\boldsymbol{z} \equiv \lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in \underline{\mathcal{Q}}_c$ by contradiction. Suppose that $\boldsymbol{z} \notin \underline{\mathcal{Q}}_c$. By definition, there must exist $\boldsymbol{z}' \notin \mathcal{Q}$ such that $\|\boldsymbol{z} - \boldsymbol{z}'\|_2 < c$. Define $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{z}' - \boldsymbol{z}$ and $\boldsymbol{y}' = \boldsymbol{y} + \boldsymbol{z}' - \boldsymbol{z}$. Because $\|\boldsymbol{x} - \boldsymbol{x}'\|_2 = \|\boldsymbol{y} - \boldsymbol{y}'\|_2 = \|\boldsymbol{z} - \boldsymbol{z}'\|_2 < c$ and $\boldsymbol{x}, \boldsymbol{y} \in \underline{\mathcal{Q}}_c$, by definition, we must have $\boldsymbol{x}', \boldsymbol{y}' \in \mathcal{Q}$. Due to the convexity of $A$, this further implies that $\lambda\boldsymbol{x}' + (1-\lambda)\boldsymbol{y}' \in \mathcal{Q}$. By some algebra, we can show that $\boldsymbol{z}' = \lambda\boldsymbol{x}' + (1-\lambda)\boldsymbol{y}' \in \mathcal{Q}$, which contradicts with $\boldsymbol{z}' \notin \mathcal{Q}$. Therefore, we must have $\boldsymbol{z} = \lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in \underline{\mathcal{Q}}_c$. Consequently, $\underline{\mathcal{Q}}_c$ is a convex set.

Finally, we prove (iii). Consider any $\boldsymbol{x}, \boldsymbol{y} \in \tilde{\mathcal{Q}}$ and any $\lambda \in (0,1)$. By definition, $\boldsymbol{\Delta}\boldsymbol{x}, \boldsymbol{\Delta}\boldsymbol{y} \in \mathcal{Q}$. By the convexity of $\mathcal{Q}$, this implies that $\boldsymbol{\Delta}\{\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}\} = \lambda\boldsymbol{\Delta}\boldsymbol{x} + (1-\lambda)\boldsymbol{\Delta}\boldsymbol{y} \in \mathcal{Q}$. Consequently, $\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in \tilde{\mathcal{Q}}$. Therefore, $\tilde{\mathcal{Q}}$ must be a convex set.

From the above, Lemma A6 holds. □

PROOF OF LEMMA A7. We first prove (i). We first prove $\overline{(\overline{\mathcal{Q}}_c)}_h \subset \overline{\mathcal{Q}}_{c+h}$. For any $\boldsymbol{x} \in \overline{(\overline{\mathcal{Q}}_c)}_h$, by definition, there exists $\boldsymbol{x}' \in \overline{\mathcal{Q}}_c$ such that $\|\boldsymbol{x} - \boldsymbol{x}'\|_2 < h$. By the same logic, there exists $\boldsymbol{x}'' \in \mathcal{Q}$ such that $\|\boldsymbol{x}' - \boldsymbol{x}''\|_2 < c$. By the triangle inequality, $\|\boldsymbol{x} - \boldsymbol{x}''\|_2 \le \|\boldsymbol{x} - \boldsymbol{x}'\|_2 + \|\boldsymbol{x}' - \boldsymbol{x}''\|_2 < c + h$, which then implies that $\boldsymbol{x} \in \overline{\mathcal{Q}}_{c+h}$. Therefore, we must have $\overline{(\overline{\mathcal{Q}}_c)}_h \subset \overline{\mathcal{Q}}_{c+h}$. We then prove $\overline{(\overline{\mathcal{Q}}_c)}_h \supset \overline{\mathcal{Q}}_{c+h}$. For any $\boldsymbol{x} \in \overline{\mathcal{Q}}_{c+h}$, by definition, there exists $\boldsymbol{x}' \in \mathcal{Q}$ such that $\|\boldsymbol{x} - \boldsymbol{x}'\|_2 < c + h$. Let $\lambda = c/(c+h)$, and $\boldsymbol{x}'' = \boldsymbol{x}' + \lambda(\boldsymbol{x} - \boldsymbol{x}')$. We then have $\|\boldsymbol{x}' - \boldsymbol{x}''\|_2 = \lambda\|\boldsymbol{x} - \boldsymbol{x}'\|_2 < c$, and $\|\boldsymbol{x}'' - \boldsymbol{x}\|_2 = (1-\lambda)\|\boldsymbol{x} - \boldsymbol{x}'\|_2 < h$. Consequently, $\boldsymbol{x}'' \in \overline{\mathcal{Q}}_c$, and $\boldsymbol{x} \in \overline{(\overline{\mathcal{Q}}_c)}_h$. Therefore, we must have $\overline{\mathcal{Q}}_{c+h} \subset \overline{(\overline{\mathcal{Q}}_c)}_h$. From the above, we have $\overline{(\overline{\mathcal{Q}}_c)}_h = \overline{\mathcal{Q}}_{c+h}$.

We then prove (ii). By definition,

$$
\begin{aligned}
\left(\underline{\mathcal{Q}}_c\right)^{\complement} &= \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}' - \boldsymbol{x}\|_2 \geq c \ \forall \boldsymbol{x}' \notin \mathcal{Q}\}^{\complement} = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}' - \boldsymbol{x}\|_2 \geq c \ \forall \boldsymbol{x}' \in \mathcal{B}\}^{\complement} \\
&= \{\boldsymbol{x} \in \mathbb{R}^d : \exists \boldsymbol{x}' \in \mathcal{B} \ s.t. \ \|\boldsymbol{x} - \boldsymbol{x}'\|_2 < c\} = \overline{\mathcal{B}}_c.
\end{aligned}
$$

Finally, we prove (iii) using (i) and (ii). From (ii), we have $\underline{\left(\underline{\mathcal{Q}}_c\right)}_h = (\overline{\mathcal{D}}_h)^{\complement}$, where $\mathcal{D} = (\underline{\mathcal{Q}}_c)^{\complement}$. By the same logic, $\mathcal{D} = (\underline{\mathcal{Q}}_c)^{\complement} = \overline{\mathcal{B}}_c$, where $\mathcal{B} = \mathcal{Q}^{\complement}$. Consequently, using (i) and (ii), we have

$$
\underline{\left(\underline{\mathcal{Q}}_c\right)}_h = (\overline{\mathcal{D}}_h)^{\complement} = \left(\overline{\left(\overline{\mathcal{B}}_c\right)}_h\right)^{\complement} = \left(\overline{\mathcal{B}}_{c+h}\right)^{\complement} = \underline{\mathcal{Q}}_{c+h}.
$$

From the above, Lemma A7 holds. □

PROOF OF LEMMA A8. This is a direct consequence of (1.3)-(1.4) of Bentkus (2005). See also (Ball, 1993; Nazarov, 2003) □

PROOF OF LEMMA A9. Let $\boldsymbol{\zeta} \equiv \boldsymbol{\Sigma}_{\boldsymbol{B}}^{-1/2}(\boldsymbol{A} - \boldsymbol{B})$ and $\boldsymbol{\Delta} \equiv \boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}\boldsymbol{\Sigma}_{\boldsymbol{B}}^{1/2} = (1-l)\boldsymbol{I}_d$. Then, by definition, $\mathbb{E}(\boldsymbol{\zeta}\boldsymbol{\zeta}^\top) \leq b^2 \boldsymbol{I}_d$, and

$$
\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{A} - \mathbb{E}\boldsymbol{A}) = \boldsymbol{\Delta}\boldsymbol{\Sigma}_{\boldsymbol{B}}^{-1/2}\{\boldsymbol{B} - \mathbb{E}\boldsymbol{B} + (\boldsymbol{A} - \boldsymbol{B})\} = \boldsymbol{\Delta}\boldsymbol{\Sigma}_{\boldsymbol{B}}^{-1/2}(\boldsymbol{B} - \mathbb{E}\boldsymbol{B}) + \boldsymbol{\Delta}\boldsymbol{\zeta}.
$$

First, for any convex set $\mathcal{Q} \subset \mathbb{R}^d$ and any $c > 0$, define

$$
\overline{\mathcal{Q}}_c \equiv \{\boldsymbol{x} \in \mathbb{R}^d : \exists \boldsymbol{x}' \in \mathcal{Q} \ s.t. \ \|\boldsymbol{x} - \boldsymbol{x}'\|_2 < c\} \quad \text{and} \quad \underline{\mathcal{Q}}_c \equiv \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}' - \boldsymbol{x}\|_2 \geq c \ \forall \boldsymbol{x}' \notin \mathcal{Q}\}.
$$

Intuitively, $\overline{\mathcal{Q}}(c)$ contains all the points whose distance from $\mathcal{Q}$ is at most $c$, and $\overline{\mathcal{Q}}(c)$ contains all the points whose distance from $\mathcal{Q}^{\complement}$ is at least $c$. Then, by definition,

$$
\begin{aligned}
\mathbb{P}\left\{\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{A} - \mathbb{E}\boldsymbol{A}) \in \mathcal{Q}\right\} &\leq \mathbb{P}\left\{\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{A} - \mathbb{E}\boldsymbol{A}) \in \mathcal{Q}, \|\boldsymbol{\Delta}\boldsymbol{\zeta}\|_2 < c\right\} + \mathbb{P}\left(\|\boldsymbol{\Delta}\boldsymbol{\zeta}\|_2 \geq c\right) \\
&\leq \mathbb{P}\left\{\boldsymbol{\Delta}\boldsymbol{\Sigma}_{\boldsymbol{B}}^{-1/2}(\boldsymbol{B} - \mathbb{E}\boldsymbol{B}) \in \overline{\mathcal{Q}}_c\right\} + \mathbb{P}\left(\|\boldsymbol{\Delta}\boldsymbol{\zeta}\|_2 \geq c\right), \qquad \text{(A3.11)}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{P}\left\{\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{A} - \mathbb{E}\boldsymbol{A}) \in \mathcal{Q}\right\} &\geq \mathbb{P}\left\{\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{A} - \mathbb{E}\boldsymbol{A}) \in \mathcal{Q}, \|\boldsymbol{\Delta}\boldsymbol{\zeta}\|_2 < c\right\} \\
&\geq \mathbb{P}\left\{\boldsymbol{\Delta}\boldsymbol{\Sigma}_{\boldsymbol{B}}^{-1/2}(\boldsymbol{B} - \mathbb{E}\boldsymbol{B}) \in \underline{\mathcal{Q}}_c, \|\boldsymbol{\Delta}\boldsymbol{\zeta}\|_2 < c\right\} \\
&\geq \mathbb{P}\left\{\boldsymbol{\Delta}\boldsymbol{\Sigma}_{\boldsymbol{B}}^{-1/2}(\boldsymbol{B} - \mathbb{E}\boldsymbol{B}) \in \underline{\mathcal{Q}}_c\right\} - \mathbb{P}\left(\|\boldsymbol{\Delta}\boldsymbol{\zeta}\|_2 \geq c\right). \qquad \text{(A3.12)}
\end{aligned}
$$

From Lemma A6 and the condition in Lemma A9, these imply that

$$
\mathbb{P}\left\{\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{A} - \mathbb{E}\boldsymbol{A}) \in \mathcal{Q}\right\} \leq \mathbb{P}\left(\boldsymbol{\Delta}\boldsymbol{\varepsilon} \in \overline{\mathcal{Q}}_c\right) + a + \mathbb{P}\left(\|\boldsymbol{\Delta}\boldsymbol{\zeta}\|_2 \geq c\right),
$$

and

$$
\mathbb{P}\left\{\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{A} - \mathbb{E}\boldsymbol{A}) \in \mathcal{Q}\right\} \geq \mathbb{P}\left(\boldsymbol{\Delta}\boldsymbol{\varepsilon} \in \underline{\mathcal{Q}}_c\right) - a - \mathbb{P}\left(\|\boldsymbol{\Delta}\boldsymbol{\zeta}\|_2 \geq c\right).
$$

Second, by definition, $\left\|(\boldsymbol{\Delta} - \boldsymbol{I}_d)\boldsymbol{\varepsilon}\right\|_2 = l\left\|\boldsymbol{\varepsilon}\right\|_2$. By the Gaussian tail bound, for any $h > 0$,

$$
\begin{aligned}
\mathbb{P}\left\{\left\|(\boldsymbol{\Delta}_N - \boldsymbol{I}_d)\boldsymbol{\varepsilon}\right\|_2 \geq h\right\} &\leq \mathbb{P}\left\{l\|\boldsymbol{\varepsilon}\|_2 \geq h\right\} = \mathbb{P}\left(\sum_{k=1}^d \varepsilon_k^2 \geq \frac{h^2}{l^2}\right) \leq \sum_{k=1}^d \mathbb{P}\left(\varepsilon_k^2 \geq \frac{h^2}{d \cdot l^2}\right) \\
&= 2d \cdot \mathbb{P}\left(\varepsilon_k \geq \frac{h}{d^{1/2} \cdot l}\right) \leq 2d \cdot \exp\left(-\frac{h^2}{2dl^2}\right).
\end{aligned}
$$

By the same logic as (A3.11) and using Lemma A7,

$$
\begin{aligned}
\mathbb{P}\big(\boldsymbol{\Delta}\boldsymbol{\varepsilon} \in \overline{\mathcal{Q}}_c\big) &\le \mathbb{P}\big(\boldsymbol{\Delta}\boldsymbol{\varepsilon} \in \overline{\mathcal{Q}}_c, \big\|(\boldsymbol{\Delta} - \boldsymbol{I}_d)\boldsymbol{\varepsilon}\big\|_2 < h\big) + \mathbb{P}\big(\big\|(\boldsymbol{\Delta} - \boldsymbol{I}_d)\boldsymbol{\varepsilon}\big\|_2 \ge h\big) \\
&\le \mathbb{P}\big(\boldsymbol{\varepsilon} \in \overline{\mathcal{Q}}_{c+h}\big) + 2d \cdot \exp\left(-\frac{h^2}{2dl^2}\right),
\end{aligned}
$$

and by the same logic as (A3.12) and using Lemma A7,

$$
\begin{aligned}
\mathbb{P}\big(\boldsymbol{\Delta}_N\boldsymbol{\varepsilon} \in \underline{\mathcal{Q}}_c\big) \ge \mathbb{P}\big(\boldsymbol{\Delta}\boldsymbol{\varepsilon} \in \underline{\mathcal{Q}}_c, \big\|(\boldsymbol{\Delta} - \boldsymbol{I}_d)\boldsymbol{\varepsilon}\big\|_2 < h\big) &\ge \mathbb{P}\big(\boldsymbol{\varepsilon} \in \underline{\mathcal{Q}}_{c+h}, \big\|(\boldsymbol{\Delta} - \boldsymbol{I}_d)\boldsymbol{\varepsilon}\big\|_2 < h\big) \\
&\ge \mathbb{P}\big(\boldsymbol{\varepsilon} \in \underline{\mathcal{Q}}_{c+h}\big) - \mathbb{P}\big(\big\|(\boldsymbol{\Delta} - \boldsymbol{I}_d)\boldsymbol{\varepsilon}\big\|_2 \ge h\big) \\
&\ge \mathbb{P}\big(\boldsymbol{\varepsilon} \in \underline{\mathcal{Q}}_{c+h}\big) - 2d \cdot \exp\left(-\frac{h^2}{2dl^2}\right).
\end{aligned}
$$

These imply that

$$
\begin{aligned}
\mathbb{P}\big\{\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{A} - \mathbb{E}\boldsymbol{A}) \in \mathcal{Q}\big\} &\le \mathbb{P}\big(\boldsymbol{\Delta}\boldsymbol{\varepsilon} \in \overline{\mathcal{Q}}_c\big) + a + \frac{b^2 d}{c^2(1-b)^2} \\
&\le \mathbb{P}\big(\boldsymbol{\varepsilon} \in \overline{\mathcal{Q}}_{c+h}\big) + 2d \cdot \exp\left(-\frac{h^2}{2dl^2}\right) + a + \mathbb{P}\big(\|\boldsymbol{\Delta}\boldsymbol{\zeta}\|_2 \ge c\big),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{P}\big\{\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{A} - \mathbb{E}\boldsymbol{A}) \in \mathcal{Q}\big\} &\ge \mathbb{P}\big(\boldsymbol{\Delta}\boldsymbol{\varepsilon} \in \underline{\mathcal{Q}}_c\big) - a - \frac{b^2 d}{c^2(1-b)^2} \\
&\ge \mathbb{P}\big(\boldsymbol{\varepsilon} \in \underline{\mathcal{Q}}_{c+h}\big) - 2d \cdot \exp\left(-\frac{h^2}{2dl^2}\right) - a - \mathbb{P}\big(\|\boldsymbol{\Delta}\boldsymbol{\zeta}\|_2 \ge c\big).
\end{aligned}
$$

Third, from Lemma A8, we have

$$
\mathbb{P}\big(\boldsymbol{\varepsilon} \in \overline{\mathcal{Q}}_{c+h}\big) \le \mathbb{P}\big(\boldsymbol{\varepsilon} \in \mathcal{Q}\big) + 4d^{1/4}(c+h) \ \text{ and } \ \mathbb{P}\big(\boldsymbol{\varepsilon} \in \underline{\mathcal{Q}}_{c+h}\big) \ge \mathbb{P}\big(\boldsymbol{\varepsilon} \in \mathcal{Q}\big) - 4d^{1/4}(c+h).
$$

From the above, we must have that, for any $\mathcal{Q} \in \mathcal{C}_d$,

$$
\begin{aligned}
&\big|\mathbb{P}\big\{\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{A} - \mathbb{E}\boldsymbol{A}) \in \mathcal{Q}\big\} - \mathbb{P}\big(\boldsymbol{\varepsilon} \in \mathcal{Q}\big)\big| \\
&\le 4d^{1/4}(c+h) + 2d \cdot \exp\left(-\frac{h^2}{2dl^2}\right) + a + \mathbb{P}\big(\|\boldsymbol{\Delta}\boldsymbol{\zeta}\|_2 \ge c\big).
\end{aligned}
$$

By Chebyshev's inequality, we can bound the tail probability of $\boldsymbol{\Delta}\boldsymbol{\zeta} = \boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{A} - \boldsymbol{B})$ by

$$
\begin{aligned}
\mathbb{P}(\|\boldsymbol{\Delta}\boldsymbol{\zeta}\|_2 \ge c) = \mathbb{P}\left\{(1-l)\|\boldsymbol{\zeta}\|_2 \ge c\right\} &\le \frac{(1-l)^2}{c^2}\mathbb{E}\big(\boldsymbol{\zeta}^\top\boldsymbol{\zeta}\big) = \frac{(1-l)^2}{c^2}\mathrm{tr}\big\{\mathbb{E}\big(\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\big)\big\} \\
&\le \frac{(1-l)^2 b^2 d}{c^2}.
\end{aligned}
$$

Therefore, Lemma A9 holds. □

Proof of Lemma A10. By definition and some algebra, we can verify that

$$
\begin{aligned}
\sum_{i=1}^N \big\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\big\|_2^2 = \sum_{i=1}^N (\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top \boldsymbol{S}^{-2}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) &= \mathrm{tr}\left(\sum_{i=1}^N \boldsymbol{S}^{-2}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top\right) \\
&= \mathrm{tr}\left(\boldsymbol{S}^{-2} \cdot (N-1)\boldsymbol{S}^2\right) = \mathrm{tr}\left((N-1)\boldsymbol{I}_d\right) = (N-1)d.
\end{aligned}
$$

By Hölder's inequality,

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i-\bar{\boldsymbol{u}})\right\|_2^3 \geq \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i-\bar{\boldsymbol{u}})\right\|_2^2\right)^{3/2} \frac{(N-1)^{3/2}d^{3/2}}{N^{3/2}} \geq (d/2)^{3/2},$$

where the last inequality holds because $(N-1)/N \geq 1/2$. Consequently, we have

$$\gamma \equiv \frac{1}{\sqrt{Nf(1-f)}}\frac{d^{1/4}}{N}\sum_{i=1}^{N}\left\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i-\bar{\boldsymbol{u}})\right\|_2^3 \geq \frac{d^{1/4}}{\sqrt{Nf(1-f)}}\frac{d^{3/2}}{2^{3/2}} = \frac{d^{7/4}}{2^{3/2}\sqrt{Nf(1-f)}}.$$

Therefore, Lemma A10 holds. □

A3.3.3. *Proof of Theorem A2(iii).*

**Proof of Theorem A2(iii).** Let $\boldsymbol{Z}$ and $\boldsymbol{T}$ be the pair of indicator vectors for simple random sampling and Bernoulli sampling satisfying Lemma A3. Recall that $\boldsymbol{A} \equiv \sum_{i=1}^{N} Z_i\boldsymbol{u}_i$ and $\boldsymbol{B} \equiv \sum_{i=1}^{N} T_i(\boldsymbol{u}_i-\bar{\boldsymbol{u}}) + m\bar{\boldsymbol{u}}$, and define further $\boldsymbol{\Sigma}_{\boldsymbol{A}} = \text{Cov}(\boldsymbol{A})$ and $\boldsymbol{\Sigma}_{\boldsymbol{B}} = \text{Cov}(\boldsymbol{B})$. By definition, we can verify that $\boldsymbol{W} = \boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{A}-\mathbb{E}\boldsymbol{A})$.

First, from Lemma A3, $\boldsymbol{\Sigma}_{\boldsymbol{B}} = (1-l)^2\cdot\boldsymbol{\Sigma}_{\boldsymbol{A}}$ with $l = 1-\sqrt{1-N^{-1}}$, and

$$\boldsymbol{\Sigma}_{\boldsymbol{B}}^{-1/2}\cdot\mathbb{E}\{(\boldsymbol{B}-\boldsymbol{A})(\boldsymbol{B}-\boldsymbol{A})^\top\}\cdot\boldsymbol{\Sigma}_{\boldsymbol{B}}^{-1/2} \leq b^2\boldsymbol{I}_d$$

with $b^2 = \sqrt{1/m+1/(N-m)} = 1/\sqrt{Nf(1-f)}$.

Second, from Lemma A5,

$$\sup_{\mathcal{Q}\in\mathcal{C}_d}\left|\mathbb{P}\{\boldsymbol{\Sigma}_{\boldsymbol{B}}^{-1/2}\cdot(\boldsymbol{B}-\mathbb{E}\boldsymbol{B})\in\mathcal{Q}\}-\mathbb{P}(\boldsymbol{\varepsilon}\in\mathcal{Q})\right| \leq a \equiv \frac{165}{\sqrt{Nf(1-f)}}\frac{d^{1/4}}{N}\sum_{i=1}^{N}\left\|\boldsymbol{S}^{-1}(\boldsymbol{u}_i-\bar{\boldsymbol{u}})\right\|_2^3 = 165\gamma,$$

where the last equality follows from the definition of $\gamma$ in Theorem A2.

Third, let

$$c = \left\{\frac{1}{2}(1-l)^2b^2d^{3/4}\right\}^{1/3}, \quad \text{and} \quad h = \left\{dl^2\cdot\log N\right\}^{1/2}.$$

From Lemma A9, we have

$$\begin{aligned}
&\sup_{\mathcal{Q}\in\mathcal{C}_d}\left|\mathbb{P}(\boldsymbol{W}\in\mathcal{Q})-\mathbb{P}(\boldsymbol{\varepsilon}\in\mathcal{Q})\right| \\
&\leq 4d^{1/4}(c+h)+2d\cdot\exp\left(-\frac{h^2}{2dl^2}\right)+a+\frac{(1-l)^2b^2d}{c^2} \\
&= a+\left\{4d^{1/4}c+\frac{(1-l)^2b^2d}{c^2}\right\}+2d\cdot\exp\left(-\frac{h^2}{2dl^2}\right)+4d^{1/4}h \\
&= a+3\cdot 2^{2/3}(1-l)^{2/3}d^{1/2}b^{2/3}+2dN^{-1/2}+4d^{3/4}l\cdot\sqrt{\log N}.
\end{aligned}$$

Fourth, from Lemma A10 and by definition, we have

$$\left\{3\cdot 2^{2/3}(1-l)^{2/3}d^{1/2}b^{2/3}\right\}^3 = 108(1-l)^2d^{3/2}b^2 \leq \frac{108d^{3/2}}{\sqrt{Nf(1-f)}} \leq 108\cdot 2^{3/2}\gamma \leq 7^3\gamma,$$

$$2dN^{-1/2} \leq \frac{d}{\sqrt{N/4}} \leq \frac{d^{7/4}}{\sqrt{Nf(1-f)}} = 2^{3/2}\gamma,$$

$$4d^{3/4}l\cdot\sqrt{\log N} = 4d^{3/4}\frac{N^{-1}}{1+\sqrt{1-N^{-1}}}\cdot\sqrt{\log N} \leq \frac{2d^{7/4}}{\sqrt{Nf(1-f)}}\sqrt{\frac{\log N}{N}} \leq 2^{5/2}\gamma$$

These then imply that

$$\sup_{\mathcal{Q}\in\mathcal{C}_d}\left|\mathbb{P}\big(\boldsymbol{W}\in\mathcal{Q}\big)-\mathbb{P}(\boldsymbol{\varepsilon}\in\mathcal{Q})\right| \le 165\gamma+2^{3/2}\gamma+2^{5/2}\gamma+3\cdot 2^{2/3}(1-l)^{2/3}d^{1/2}b^{2/3}$$
$$\le 174\gamma+3\cdot 2^{2/3}\frac{d^{1/2}}{\{Nf(1-f)\}^{1/6}}$$
$$\le 174\gamma+7\gamma^{1/3}.$$

From the above, Theorem A2(iii) holds. □

A3.4. *Proof of Theorem A2(iv).* To prove Theorem A2(iv), we need the following two lemmas.

LEMMA A11. *Let $\mathcal{X}\equiv\{x_i\}_{i=1}^N$ be $N$ zero-centered real valued quantities; and let $X_1,\cdots,X_m$ be $m$ random sample drawn without replacement from $\mathcal{X}$, then for all $\varepsilon>0$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^m X_i\right|\ge\varepsilon\right)\le 2\exp\left(-\frac{2\varepsilon^2}{m(\max_{1\le i\le N}x_i-\min_{1\le i\le N}x_i)^2}\right)$$

PROOF OF LEMMA A11. Lemma A11 follows immediately from Bardenet and Maillard (2015, Proposition 1.2). □

LEMMA A12. *Let $\boldsymbol{Z}$ and $\boldsymbol{T}$ be the pair of random vectors constructed as in Lemma A3. For any $c,t>0$,*

$$\mathbb{P}\left\{\left\|Cov^{-1/2}(\boldsymbol{A})\cdot(\boldsymbol{B}-\boldsymbol{A})\right\|_2\ge c\right\}\le 2d\exp\left(-\frac{c^2f(1-f)\sqrt{N}}{2td\xi^2}\right)+2\exp\left(-2t^2\right),$$

*where $f=m/N$ and $\xi=\max_{1\le i\le n}\left\|\left(\boldsymbol{S}^2_{\boldsymbol{u}}\right)^{-1/2}(\boldsymbol{u}_i-\bar{\boldsymbol{u}})\right\|_\infty$.*

PROOF OF LEMMA A12. For $1\le i\le N$ and $1\le k\le d$, let $\boldsymbol{v}_i\equiv\left(\boldsymbol{S}^2_{\boldsymbol{u}}\right)^{-1/2}(\boldsymbol{u}_i-\bar{\boldsymbol{u}})$, and $v_{ik}$ denote the $k$-th coordinate of $\boldsymbol{v}_i$. From the proof of Lemma A3, we then have

$$\text{Cov}^{-1/2}(\boldsymbol{A})\cdot(\boldsymbol{B}-\boldsymbol{A})=\sqrt{\frac{1}{f(1-f)N}}\cdot\sum_{i=1}^N(T_i-Z_i)\boldsymbol{v}_i. \tag{A3.13}$$

We first consider bounding the tail probability of $\sum_{i=1}^N(T_i-Z_i)\boldsymbol{v}_i$. By definition, $2\xi=2\max_{1\le i\le n}\|\boldsymbol{v}_i\|_\infty\ge 2\max_{1\le i\le n}|v_{ik}|\ge\max_i v_{ik}-\min_i v_{ik}$ for $1\le k\le d$. From (A3.7) and Lemma A11, we then have, for $1\le k\le d$ and any $c>0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n(T_i-Z_i)v_{ik}\right|\ge c\mid\tilde{m}\right)\le 2\exp\left(-\frac{2c^2}{|\tilde{m}-m|(\max_{1\le i\le N}v_{ik}-\min_{1\le i\le N}v_{ik})^2}\right)$$
$$\le 2\exp\left(-\frac{c^2}{2|\tilde{m}-m|\xi^2}\right),$$

where $c^2/(2|\tilde{m}-m|\xi^2)$ is defined to be infinity when $\tilde{m}=m$. This further implies that

$$\mathbb{P}\left(\left\|\sum_{i=1}^n(T_i-Z_i)\boldsymbol{v}_i\right\|_2\ge c\mid\tilde{m}\right)\le\sum_{k=1}^d\mathbb{P}\left(\left|\sum_{i=1}^n(T_i-Z_i)v_{ik}\right|\ge\frac{c}{\sqrt{d}}\mid\tilde{m}\right)$$
$$\le 2d\exp\left(-\frac{c^2}{2|\tilde{m}-m|d\xi^2}\right).$$

Note that, by Hoeffding's inequality, for any $t>0$, $\mathbb{P}(|\tilde{m}-m|\geq t)\leq 2\exp(-2t^2/N)$. From the above, we can know that, for any $c,t>0$,

$$\begin{aligned}
&\mathbb{P}\left(\left\|\sum_{i=1}^{n}(T_i-Z_i)\boldsymbol{v}_i\right\|_2\geq c\right)\\
&\leq \mathbb{P}\left(\left\|\sum_{i=1}^{n}(T_i-Z_i)\boldsymbol{v}_i\right\|_2\geq c \mid |\tilde{m}-m|<t\sqrt{N}\right)+\mathbb{P}(|\tilde{m}-m|\geq t\sqrt{N})\\
&\leq 2d\exp\left(-\frac{c^2}{2td\xi^2\sqrt{N}}\right)+2\exp\left(-2t^2\right).
\end{aligned}$$

Consequently, for any $c,t>0$,

$$\mathbb{P}\left\{\left\|\mathrm{Cov}^{-1/2}(\boldsymbol{A})\cdot(\boldsymbol{B}-\boldsymbol{A})\right\|_2\geq c\right\}=\mathbb{P}\left\{\left\|\sum_{i=1}^{N}(T_i-Z_i)\boldsymbol{v}_i\right\|_2\geq c\sqrt{f(1-f)N}\right\}$$

$$\leq 2d\exp\left(-\frac{c^2f(1-f)\sqrt{N}}{2td\xi^2}\right)+2\exp\left(-2t^2\right).$$

From the above, Lemma A12 holds. □

**Proof of Theorem A2(iv).** Let $\boldsymbol{Z}$ and $\boldsymbol{T}$ be the pair of indicator vectors for simple random sampling and Bernoulli sampling satisfying Lemma A3, and adopt the same notation from the proof of Theorem A2(iii). From the proof of Lemma A9 and Theorem A2(iii), for any $c,h>0$,

$$\begin{aligned}
&\sup_{\mathcal{Q}\in\mathcal{C}_d}\left|\mathbb{P}\big(\boldsymbol{W}\in\mathcal{Q}\big)-\mathbb{P}(\boldsymbol{\varepsilon}\in\mathcal{Q})\right|\\
&\leq 4d^{1/4}(c+h)+2d\cdot\exp\left(-\frac{h^2}{2dl^2}\right)+a+\frac{(1-l)^2b^2d}{c^2}\\
&=a+\left\{2d\cdot\exp\left(-\frac{h^2}{2dl^2}\right)+4d^{1/4}h\right\}+\left[4d^{1/4}c+\mathbb{P}\left\{\left\|\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{B}-\boldsymbol{A})\right\|_2\geq c\right\}\right],
\end{aligned}$$

where $a=165\gamma$ and $l=1-\sqrt{1-N^{-1}}$. Letting $h=\{dl^2\cdot\log N\}^{1/2}$ and from the proof of Theorem A2(iii), we can know that, for any $c>0$,

$$\sup_{\mathcal{Q}\in\mathcal{C}_d}\left|\mathbb{P}\big(\boldsymbol{W}\in\mathcal{Q}\big)-\mathbb{P}(\boldsymbol{\varepsilon}\in\mathcal{Q})\right|\leq 174\gamma+\left[4d^{1/4}c+\mathbb{P}\left\{\left\|\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{B}-\boldsymbol{A})\right\|_2\geq c\right\}\right].$$

Applying Lemma A12 and letting

$$c=\left\{\left(\frac{\log N}{N}\right)^{1/2}\frac{1}{f(1-f)}\frac{\log N}{2}d\xi^2\right\}^{1/2},\quad t=\left(\frac{\log N}{4}\right)^{1/2},$$

we have

$$\begin{aligned}
4d^{1/4}c+\mathbb{P}\left\{\left\|\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{B}-\boldsymbol{A})\right\|_2\geq c\right\}&\leq 4d^{1/4}c+2d\exp\left(-\frac{c^2f(1-f)\sqrt{N}}{2td\xi^2}\right)+2\exp\left(-2t^2\right)\\
&=2\sqrt{2}\frac{(\log N)^{3/4}d^{3/4}}{N^{1/4}\sqrt{f(1-f)}}\cdot\xi+2dN^{-1/2}+2N^{-1/2}\\
&\leq 3\frac{(\log N)^{3/4}d^{3/4}}{N^{1/4}\sqrt{f(1-f)}}\cdot\xi+4dN^{-1/2},
\end{aligned}$$

where $\xi = \max_{1 \le i \le n} \left\| \left(\boldsymbol{S}_{\boldsymbol{u}}^2\right)^{-1/2} (\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \right\|_\infty$. From Lemma A10 and the proof of Theorem A2(iii), we can know that $4dN^{-1/2} \le 2^{5/2}\gamma \le 6\gamma$. From the above,

$$
\begin{aligned}
\sup_{\mathcal{Q} \in \mathcal{C}_d} \left| \mathbb{P}\left(\boldsymbol{W} \in \mathcal{Q}\right) - \mathbb{P}(\boldsymbol{\varepsilon} \in \mathcal{Q}) \right| &\le 174\gamma + 3 \frac{(\log N)^{3/4} d^{3/4}}{N^{1/4}\sqrt{f(1-f)}} \cdot \xi + 6\gamma \\
&\le 180\gamma + 3 \frac{(\log N)^{3/4} d^{3/4}}{N^{1/4}\sqrt{f(1-f)}} \cdot \max_{1 \le i \le n} \left\| \left(\boldsymbol{S}_{\boldsymbol{u}}^2\right)^{-1/2} (\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \right\|_\infty .
\end{aligned}
$$

Therefore, Theorem A2(iv) holds. □

A3.5. *Proof of Theorem A2(v).* To prove Theorem A2(v), we need the following two lemmas.

LEMMA A13. *Let $\mathcal{X} \equiv \{x_i\}_{i=1}^N$ be $N$ zero-centered real valued quantities, and let $X_1, \cdots, X_m$ be $m$ random sample drawn without replacement from $\mathcal{X}$. Then for any $t > 0$ and $\iota \ge 2$, we have*

$$
\mathbb{P}\left( \left| \sum_{i=1}^m X_i \right| \ge t \right) \le R_\iota \frac{\left(f \sum_{i=1}^N x_i^2\right)^{\iota/2} + f \sum_{i=1}^N |x_i|^\iota}{t^\iota},
$$

*where $f = m/N$ and $R_\iota$ is a universal constant depending only on $\iota$.*

PROOF OF LEMMA A13. From Markov's inequality, for any $\iota \ge 2$,

$$
\mathbb{P}\left( \left| \sum_{i=1}^m X_i \right| \ge t \right) = \mathbb{P}\left( \left| \sum_{i=1}^m X_i \right|^\iota \ge t^\iota \right) \le \frac{\mathbb{E} \left|\sum_{i=1}^m X_i\right|^\iota}{t^\iota}.
$$

From Hoeffding (1963, Theorem 4), we further have $\mathbb{E}|\sum_{i=1}^m X_i|^\iota \le \mathbb{E}|\sum_{i=1}^m \tilde{X}_i|^\iota$, where $\tilde{X}_1, \cdots, \tilde{X}_m$ are i.i.d. random samples drawn with replacement from $\mathcal{X}$. From Rosenthal's inequality, we then have

$$
\mathbb{P}\left( \left| \sum_{i=1}^m X_i \right| \ge t \right) \le \frac{\mathbb{E}\left|\sum_{i=1}^m \tilde{X}_i\right|^\iota}{t^\iota} \le R_\iota \frac{\left(\sum_{i=1}^m \mathbb{E}\tilde{X}_i^2\right)^{\iota/2} + \sum_{i=1}^m \mathbb{E}|\tilde{X}_i|^\iota}{t^\iota},
$$

where $R_\iota$ is a universal constant depending only on $\iota$. Note that $\mathbb{E}\tilde{X}_i^2 = N^{-1}\sum_{i=1}^N x_i^2$ and $\mathbb{E}|\tilde{X}_i|^\iota = N^{-1}\sum_{i=1}^N |x_i|^\iota$. We can then derive Lemma A13. □

LEMMA A14. *Let $\boldsymbol{Z}$ and $\boldsymbol{T}$ be the pair of random vectors constructed as in Lemma A3. For any $\iota \ge 2$ and $t > 0$,*

$$
\mathbb{P}\left\{ \left\| Cov^{-1/2}(\boldsymbol{A}) \cdot (\boldsymbol{B} - \boldsymbol{A}) \right\|_2 \ge c \right\} = \frac{C_\iota d^{\iota/2+1}}{c^\iota N^{\iota/4}\{f(1-f)\}^{\iota/2}} + \frac{C_\iota d^{\iota/2}\xi_\iota}{c^\iota N^{(\iota-1)/2}\{f(1-f)\}^{(\iota-1)/2}}
$$

*where $C_\iota$ is a constant depending only on $\iota$, and $\xi_\iota = N^{-1}\sum_{i=1}^N \| \left(\boldsymbol{S}_{\boldsymbol{u}}^2\right)^{-1/2} (\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \|_\iota^\iota$.*

PROOF OF LEMMA A14. We construct $\boldsymbol{Z}$ and $\boldsymbol{T}$ in the same way as in the proof of Lemmas A3 and A12, and we adopt the same notation as in Lemma A12. We further define $\xi_{k,\iota} = N^{-1}\sum_{i=1}^N |v_{ik}|^\iota$.

We first consider bounding the tail probability of $\|\sum_{i=1}^n (T_i - Z_i)\boldsymbol{v}_i\|_2$. From Lemma A13, for any $c > 0$ and $1 \le k \le d$,

$$\mathbb{P}\left( \left| \sum_{i=1}^n (T_i - Z_i) v_{ik} \right| \ge c \mid \tilde{m} \right) \le R_\iota \frac{\left( |\tilde{m} - m| \cdot N^{-1} \sum_{i=1}^N v_{ik}^2 \right)^{\iota/2} + |\tilde{m} - m| \cdot N^{-1} \sum_{i=1}^N |v_{ik}|^\iota}{c^\iota}$$
$$\le R_\iota \frac{|\tilde{m} - m|^{\iota/2} + |\tilde{m} - m| \cdot \xi_{k,\iota}}{c^\iota},$$

where the last inequality holds because $N^{-1} \sum_{i=1}^N v_{ik}^2 = (N-1)/N \le 1$. This then implies that

$$\mathbb{P}\left( \left\| \sum_{i=1}^n (T_i - Z_i) \boldsymbol{v}_i \right\|_2 \ge c \mid \tilde{m} \right) \le \sum_{k=1}^d \mathbb{P}\left( \left| \sum_{i=1}^n (T_i - Z_i) v_{ik} \right| \ge c/\sqrt{d} \mid \tilde{m} \right)$$
$$\le R_\iota \sum_{k=1}^d \frac{|\tilde{m} - m|^{\iota/2} + |\tilde{m} - m| \cdot \xi_{k,\iota}}{c^\iota d^{-\iota/2}} = R_\iota d^{\iota/2} \cdot \frac{d|\tilde{m} - m|^{\iota/2} + |\tilde{m} - m| \cdot \sum_{k=1}^d \xi_{k,\iota}}{c^\iota}.$$

By the law of iterated expectation,

$$\mathbb{P}\left( \left\| \sum_{i=1}^n (T_i - Z_i) \boldsymbol{v}_i \right\|_2 \ge c \right) = \mathbb{E}\left\{ \mathbb{P}\left( \left\| \sum_{i=1}^n (T_i - Z_i) \boldsymbol{v}_i \right\|_2 \ge c \mid \tilde{m} \right) \right\}$$
$$\le R_\iota d^{\iota/2} \cdot \frac{d \cdot \mathbb{E}\{|\tilde{m} - m|^{\iota/2}\} + \mathbb{E}\{|\tilde{m} - m|\} \cdot \sum_{k=1}^d \xi_{k,\iota}}{c^\iota}.$$

We then consider bounding the moments of $|\tilde{m} - m|$. By Hoeffding's inequality, for any $t > 0$, $\mathbb{P}(|\tilde{m} - m| \ge t) \le 2\exp(-2t^2/N)$. Using (Rigollet and Hütter, 2015, Lemma 1.4), this implies that

$$\mathbb{E}\{|\tilde{m} - m|^{\iota/2}\} \le \left( \frac{N}{2} \right)^{\iota/4} \cdot (\iota/2) \cdot \Gamma(\iota/4).$$

Besides, $\mathbb{E}\{|\tilde{m} - m|\} \le \sqrt{\text{Var}(\tilde{m} - m)} = \sqrt{N f(1-f)}$.

Finally, we consider bounding the tail probability of $\text{Cov}^{-1/2}(\boldsymbol{A}) \cdot (\boldsymbol{B} - \boldsymbol{A})$. From (A3.13),

$$\mathbb{P}\left\{ \left\| \text{Cov}^{-1/2}(\boldsymbol{A}) \cdot (\boldsymbol{B} - \boldsymbol{A}) \right\|_2 \ge c \right\} = \mathbb{P}\left\{ \left\| \sum_{i=1}^N (T_i - Z_i) \boldsymbol{v}_i \right\|_2 \ge c\sqrt{f(1-f)N} \right\}$$
$$\le R_\iota d^{\iota/2} \cdot \frac{d \cdot \mathbb{E}\{|\tilde{m} - m|^{\iota/2}\} + \mathbb{E}\{|\tilde{m} - m|\} \cdot \sum_{k=1}^d \xi_{k,\iota}}{c^\iota N^{\iota/2} \{f(1-f)\}^{\iota/2}}$$
$$\le \frac{R_\iota d^{\iota/2}}{c^\iota N^{\iota/2} \{f(1-f)\}^{\iota/2}} \left\{ d \cdot \left( \frac{N}{2} \right)^{\iota/4} \cdot (\iota/2) \cdot \Gamma(\iota/4) + \sqrt{N f(1-f)} \cdot \sum_{k=1}^d \xi_{k,\iota} \right\}$$
$$\le \frac{C_\iota d^{\iota/2}}{c^\iota N^{\iota/2} \{f(1-f)\}^{\iota/2}} \left\{ d \cdot N^{\iota/4} + \sqrt{N f(1-f)} \cdot N^{-1} \sum_{i=1}^N \|\boldsymbol{v}_i\|_\iota^\iota \right\}$$
$$= \frac{C_\iota d^{\iota/2+1}}{c^\iota N^{\iota/4} \{f(1-f)\}^{\iota/2}} + \frac{C_\iota d^{\iota/2}}{c^\iota N^{(\iota-1)/2} \{f(1-f)\}^{(\iota-1)/2}} \cdot \frac{1}{N} \sum_{i=1}^N \|\boldsymbol{v}_i\|_\iota^\iota.$$

From the above, we can immediately derive Lemma A14. □

**Proof of Theorem A2(v).** Letting $\boldsymbol{Z}$ and $\boldsymbol{T}$ be the pair of indicator vectors for simple random sampling and Bernoulli sampling satisfying Lemma A3, and by the same logic as the proof of Theorem A2(iv), for any $c > 0$,

$$\sup_{\mathcal{Q}\in\mathcal{C}_d} \left|\mathbb{P}\big(\boldsymbol{W}\in\mathcal{Q}\big) - \mathbb{P}(\boldsymbol{\varepsilon}\in\mathcal{Q})\right| \le 174\gamma + \left[4d^{1/4}c + \mathbb{P}\left\{\left\|\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{B}-\boldsymbol{A})\right\|_2 \ge c\right\}\right].$$

Applying Lemma A12 and letting $c = N^{-\iota/\{4(\iota+1)\}}\cdot d^{(2\iota-1)/\{4(\iota+1)\}}$, we have

$$\begin{aligned}
&4d^{1/4}c + \mathbb{P}\left\{\left\|\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}(\boldsymbol{B}-\boldsymbol{A})\right\|_2 \ge c\right\}\\
&\le 4d^{1/4}c + \frac{C_\iota d^{\iota/2+1}}{c^\iota N^{\iota/4}\{f(1-f)\}^{\iota/2}} + \frac{C_\iota d^{\iota/2}\xi_\iota}{c^\iota N^{(\iota-1)/2}\{f(1-f)\}^{(\iota-1)/2}}\\
&= 4\frac{d^{3\iota/\{4(\iota+1)\}}}{N^{\iota/\{4(\iota+1)\}}} + \frac{C_\iota d^{3\iota/\{4(\iota+1)\}}\cdot d}{N^{\iota/\{4(\iota+1)\}}\{f(1-f)\}^{\iota/2}} + \frac{C_\iota\cdot d^{3\iota/\{4(\iota+1)\}}\xi_\iota}{N^{(\iota^2-2)/\{4(\iota+1)\}}\{f(1-f)\}^{(\iota-1)/2}}\\
&\le 4\frac{d^{3\iota/\{4(\iota+1)\}}}{N^{\iota/\{4(\iota+1)\}}} + \frac{C_\iota d^{3\iota/\{4(\iota+1)\}}\cdot (d+\xi_\iota)}{N^{\iota/\{4(\iota+1)\}}\{f(1-f)\}^{\iota/2}} \le \max\{4, C_\iota\}\cdot \frac{d^{3\iota/\{4(\iota+1)\}}\cdot(2d+\xi_\iota)}{N^{\iota/\{4(\iota+1)\}}\{f(1-f)\}^{\iota/2}},
\end{aligned}$$

where $\xi_\iota = N^{-1}\sum_{i=1}^N \|\left(\boldsymbol{S}_{\boldsymbol{u}}^2\right)^{-1/2}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\|_\iota^\iota$. Adopting the notation from the proof of Lemma A14,

$$\begin{aligned}
\xi_\iota &= N^{-1}\sum_{i=1}^N\sum_{k=1}^d |v_{ik}|^\iota = \sum_{k=1}^d\left(N^{-1}\sum_{i=1}^N |v_{ik}|^\iota\right) \ge \sum_{k=1}^d\left(N^{-1}\sum_{i=1}^N v_{ik}^2\right)^{\iota/2} = \sum_{k=1}^d\left(\frac{N-1}{N}\right)^{\iota/2}\\
&\ge 2^{-\iota/2}\cdot d.
\end{aligned}$$

From the above, we then have

$$\begin{aligned}
&\sup_{\mathcal{Q}\in\mathcal{C}_d}\left|\mathbb{P}\big(\boldsymbol{W}\in\mathcal{Q}\big) - \mathbb{P}(\boldsymbol{\varepsilon}\in\mathcal{Q})\right|\\
&\le 174\gamma + \max\{4, C_\iota\}\cdot\frac{d^{3\iota/\{4(\iota+1)\}}\cdot(1+d+\xi_\iota)}{N^{\iota/\{4(\iota+1)\}}\{f(1-f)\}^{\iota/2}}\\
&\le 174\gamma + \max\{4, C_\iota\}\cdot\frac{d^{3\iota/\{4(\iota+1)\}}\cdot(2^{\iota/2+1}+1)\xi_\iota}{N^{\iota/\{4(\iota+1)\}}\{f(1-f)\}^{\iota/2}}\\
&\le 174\gamma + C_\iota'\cdot\frac{d^{3\iota/\{4(\iota+1)\}}}{N^{\iota/\{4(\iota+1)\}}\{f(1-f)\}^{\iota/2}}\cdot\frac{1}{N}\sum_{i=1}^N\|\left(\boldsymbol{S}_{\boldsymbol{u}}^2\right)^{-1/2}(\boldsymbol{u}_i-\bar{\boldsymbol{u}})\|_\iota^\iota,
\end{aligned}$$

where $C_\iota' = \max\{4, C_\iota\}\cdot(2^{\iota/2+1}+1)$ is a universal constant depending only on $\iota$. Therefore, Theorem A2(v) holds. □

## A4. Asymptotic Distributions in Completely Randomized and Rerandomized Experiments.

**Proof of Theorems 1 and 2.** Following the notation in Section 3.2 and from (6), the difference-in-means vector $(\hat\tau, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top)^\top$ is essentially the sample total of a simple random sample of size $n_1$ from the finite population of $\{\boldsymbol{u}_i = (r_0 Y_i(1) + r_1 Y_i(0), \boldsymbol{X}_i^\top)^\top : i = 1, 2, \ldots, n\}$, up to some constant scaling and shifting. This then implies that

$$\Delta_n \equiv \sup_{\mathcal{Q}\in\mathcal{C}_{K+1}}\left|\mathbb{P}\left\{\boldsymbol{V}^{-1/2}\begin{pmatrix}\hat\tau-\tau\\ \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}\end{pmatrix}\in\mathcal{Q}\right\} - \mathbb{P}(\boldsymbol{\varepsilon}\in\mathcal{Q})\right| = \sup_{\mathcal{Q}\in\mathcal{C}_{K+1}}\left|\mathbb{P}(\boldsymbol{W}\in\mathcal{Q}) - \mathbb{P}(\boldsymbol{\varepsilon}\in\mathcal{Q})\right|,$$

where $\boldsymbol{W} = \text{Cov}^{-1/2}(\sum_{i=1}^n Z_i \boldsymbol{u}_i) \sum_{i=1}^n Z_i \boldsymbol{u}_i$ is the standardization of $\sum_{i=1}^n Z_i \boldsymbol{u}_i$. By the definition of $\gamma_n$ in (7) and the definitions of $r_1, r_0, K$, Theorem 1 then follows immediately from Theorem A2 (i - iii); Theorem 2 follows from Theorem A2 (iv - v). □

**Proof of Theorem 3.** Under Condition 2, there must exist $\underline{n} \geq 2$ such that $p_n > \Delta_n$ for all $n \geq \underline{n}$. Let $(\tilde{\tau}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top)^\top$ denote a Gaussian random vector with mean $(\tau, \boldsymbol{0}_K^\top)$ and covariance matrix $\boldsymbol{V}$ in (4). By the definition of $\Delta_n$ in (8), we can know that, for any measurable convex set $\mathcal{Q}$ in $\mathbb{R}^{K+1}$,

$$\left|\mathbb{P}\left\{\begin{pmatrix}\hat{\tau}-\tau\\ \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}\end{pmatrix} \in \mathcal{Q}\right\} - \mathbb{P}\left\{\begin{pmatrix}\tilde{\tau}-\tau\\ \tilde{\boldsymbol{\tau}}_{\boldsymbol{X}}\end{pmatrix} \in \mathcal{Q}\right\}\right| = \left|\mathbb{P}\left\{\boldsymbol{V}^{-1/2}\begin{pmatrix}\hat{\tau}-\tau\\ \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}\end{pmatrix} \in \boldsymbol{V}^{-1/2}\mathcal{Q}\right\} - \mathbb{P}\left(\boldsymbol{\varepsilon} \in \boldsymbol{V}^{-1/2}\mathcal{Q}\right)\right| \leq \Delta_n.$$

This implies that,

$$\left|\mathbb{P}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top \boldsymbol{V}_{\boldsymbol{xx}}^{-1} \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \leq a_n\right) - \mathbb{P}\left(\tilde{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top \boldsymbol{V}_{\boldsymbol{xx}}^{-1} \tilde{\boldsymbol{\tau}}_{\boldsymbol{X}} \leq a_n\right)\right| = \left|\mathbb{P}(M \leq a_n) - \mathbb{P}(\tilde{M} \leq a_n)\right| \leq \Delta_n,$$

and for any $c \in \mathbb{R}$,

$$\begin{aligned}
&\left|\mathbb{P}\left(\hat{\tau} - \tau \leq c, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top \boldsymbol{V}_{\boldsymbol{xx}}^{-1} \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \leq a_n\right) - \mathbb{P}\left(\tilde{\tau} - \tau \leq c, \tilde{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top \boldsymbol{V}_{\boldsymbol{xx}}^{-1} \tilde{\boldsymbol{\tau}}_{\boldsymbol{X}} \leq a_n\right)\right| \\
&= \left|\mathbb{P}\left(\hat{\tau} - \tau \leq c, M \leq a_n\right) - \mathbb{P}\left(\tilde{\tau} - \tau \leq c, \tilde{M} \leq a_n\right)\right| \leq \Delta_n,
\end{aligned} \tag{A4.14}$$

where $\tilde{M} \equiv \tilde{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top \boldsymbol{V}_{\boldsymbol{xx}}^{-1} \tilde{\boldsymbol{\tau}}_{\boldsymbol{X}} \sim \chi^2_{K_n}$. By definition, $p_n = \mathbb{P}(\tilde{M} \leq a_n)$. Thus, for $n \geq \underline{n}$, we must have $\mathbb{P}(M \leq a_n) \geq p_n - \Delta_n > 0$, and consequently

$$\frac{1}{p_n + \Delta_n} \leq \frac{1}{\mathbb{P}(M \leq a_n)} \leq \frac{1}{p_n - \Delta_n}. \tag{A4.15}$$

From (A4.14) and (A4.15), we then have, for all $n \geq \underline{n}$ and $c \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{P}(\hat{\tau} - \tau \leq c \mid M \leq a_n) &= \frac{\mathbb{P}(\hat{\tau} - \tau \leq c, M \leq a_n)}{\mathbb{P}(M \leq a_n)} \leq \frac{\mathbb{P}(\tilde{\tau} - \tau \leq c, \tilde{M} \leq a_n) + \Delta_n}{\mathbb{P}(\tilde{M} \leq a_n) - \Delta_n} \\
&= \frac{\mathbb{P}(\tilde{\tau} - \tau \leq c \mid \tilde{M} \leq a_n) + \Delta_n/p_n}{1 - \Delta_n/p_n} \\
&\leq \frac{\mathbb{P}(\tilde{\tau} - \tau \leq c \mid \tilde{M} \leq a_n)(1 - \Delta_n/p_n) + 2\Delta_n/p_n}{1 - \Delta_n/p_n} \\
&= \mathbb{P}(\tilde{\tau} - \tau \leq c \mid \tilde{M} \leq a_n) + \frac{2\Delta_n/p_n}{1 - \Delta_n/p_n},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}(\hat{\tau} - \tau \leq c \mid M \leq a_n) &= \frac{\mathbb{P}(\hat{\tau} - \tau \leq c, M \leq a_n)}{\mathbb{P}(M \leq a_n)} \geq \frac{\mathbb{P}(\tilde{\tau} - \tau \leq c, \tilde{M} \leq a_n) - \Delta_n}{\mathbb{P}(\tilde{M} \leq a_n) + \Delta_n} \\
&= \frac{\mathbb{P}(\tilde{\tau} - \tau \leq c \mid \tilde{M} \leq a_n) - \Delta_n/p_n}{1 + \Delta_n/p_n} \\
&\geq \frac{\mathbb{P}(\tilde{\tau} - \tau \leq c \mid \tilde{M} \leq a_n)(1 + \Delta_n/p_n) - 2\Delta_n/p_n}{1 + \Delta_n/p_n} \\
&= \mathbb{P}(\tilde{\tau} - \tau \leq c \mid \tilde{M} \leq a_n) - \frac{2\Delta_n/p_n}{1 + \Delta_n/p_n}.
\end{aligned}$$

These imply that, for all $n \ge \underline{n}$,

$$\sup_{c\in\mathbb{R}} \left| \mathbb{P}\left\{ V_{\tau\tau}^{-1/2}(\hat{\tau}-\tau) \le c \mid M \le a_n \right\} - \mathbb{P}\left\{ V_{\tau\tau}^{-1/2}(\tilde{\tau}-\tau) \le c \mid \tilde{M} \le a_n \right\} \right|$$

$$\le \max\left\{ \frac{2\Delta_n/p_n}{1-\Delta_n/p_n}, \frac{2\Delta_n/p_n}{1+\Delta_n/p_n} \right\} \le \frac{2\Delta_n/p_n}{1-\Delta_n/p_n}.$$

Under Condition 2, we then have, as $n \to \infty$,

$$\sup_{c\in\mathbb{R}} \left| \mathbb{P}\left\{ V_{\tau\tau}^{-1/2}(\hat{\tau}-\tau) \le c \mid M \le a_n \right\} - \mathbb{P}\left\{ V_{\tau\tau}^{-1/2}(\tilde{\tau}-\tau) \le c \mid \tilde{M} \le a_n \right\} \right| \to 0.$$

Finally, from the proof of Li, Ding and Rubin (2018, Theorem 1), for any $c \in \mathbb{R}$,

$$\mathbb{P}\left\{ V_{\tau\tau}^{-1/2}(\tilde{\tau}-\tau) \le c \mid \tilde{M} \le a_n \right\} = \mathbb{P}\left\{ \left( \sqrt{1-R^2}\varepsilon_0 + \sqrt{R^2} L_{K_n,a_n} \right) \le c \right\},$$

with $\varepsilon_0$ and $L_{K_n,a_n}$ defined as in Section 4. Therefore, we derive Theorem 3. □

**Comment on Condition 1 and regularity conditions in Li, Ding and Rubin (2018).** By the definition in (7),

$$\begin{aligned}
\gamma_n &\le \frac{(K+1)^{1/4}}{\sqrt{n r_1 r_0}} \cdot \max_{1\le i\le n} \left\| \boldsymbol{S}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \right\|_2 \cdot \frac{1}{n}\sum_{i=1}^n \left\| \boldsymbol{S}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \right\|_2^2 \\
&= \frac{(K+1)^{1/4}}{\sqrt{n r_1 r_0}} \cdot \max_{1\le i\le n} \left\| \boldsymbol{S}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \right\|_2 \cdot \frac{(n-1)(K+1)}{n} \\
&\le \frac{(K+1)^{5/4}}{\sqrt{r_1 r_0}} \cdot \left\| \boldsymbol{S}_{\boldsymbol{u}}^{-1} \right\|_2 \cdot \frac{1}{\sqrt{n}} \max_{1\le i\le n} \left\| \boldsymbol{u}_i - \bar{\boldsymbol{u}} \right\|_2 .
\end{aligned}$$

Note that

$$\frac{1}{n}\max_{1\le i\le n} \|\boldsymbol{u}_i - \bar{\boldsymbol{u}}\|_2^2 = \frac{1}{n}\max_{1\le i\le n} \left[ r_0\{Y_i(1)-\bar{Y}(1)\} + r_1\{Y_i(0)-\bar{Y}(0)\} \right]^2 + \frac{1}{n}\max_{1\le i\le n} \left\| \boldsymbol{X}_i - \bar{\boldsymbol{X}} \right\|_2^2$$

$$\le \frac{2}{n}\max_{1\le i\le n} \{Y_i(1)-\bar{Y}(1)\}^2 + \frac{2}{n}\max_{1\le i\le n} \{Y_i(0)-\bar{Y}(0)\}^2 + \frac{1}{n}\left\| \boldsymbol{X}_i - \bar{\boldsymbol{X}} \right\|_2^2. \tag{A4.16}$$

Under Li, Ding and Rubin (2018, Condition 1), as $n \to \infty$, both $r_1$ and $r_0$ have positive limits, $\boldsymbol{S}_{\boldsymbol{u}}^2$ has a limiting value (in particular, the limit of $\boldsymbol{S}_{\boldsymbol{X}}^2$ is nonsingular), and the quantities on the right hand side of (A4.16) converge to zero. If additionally the limit of $R^2$ is less than 1, then the limit of $\boldsymbol{S}_{\boldsymbol{u}}^2$ will be invertible, and thus $\gamma_n$ must converge to zero as $n \to \infty$, i.e., Condition 1 holds. □

**Proof of Corollary 1.** Corollary 1 follows by the same logic as Li, Ding and Rubin (2018, Corollaries 1–3). □

**Comments on the lower bound of $\gamma_n$ in (9).** The lower bound of $\gamma_n$ follows by the same logic as Lemma A10. □

**Comments on the upper bound of $\gamma_n$.** By the definition in (7),

$$\gamma_n \equiv \frac{(K_n+1)^{1/4}}{\sqrt{n r_1 r_0}} \frac{1}{n}\sum_{i=1}^n \left\| \boldsymbol{S}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \right\|_2^3 \le \frac{(K_n+1)^{1/4}}{\sqrt{n r_1 r_0}} \frac{1}{n}\sum_{i=1}^n (K_n+1)^{3/2} \left\| \boldsymbol{S}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \right\|_\infty^{3/2}$$

$$\le \frac{(K_n+1)^{7/4}}{\sqrt{n r_1 r_0}} \max_{1\le i\le n} \left\| \boldsymbol{S}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \right\|_\infty^{3/2}.$$

If the standardized finite population $\{\boldsymbol{S}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) : 1 \le i \le n\}$ is coordinate-wise bounded, and the proportions of treated and control units are bounded away from zero, then there exist finite positive constants $c$ and $C$ such that for all $n$ and $1 \le i \le n$, $\|\boldsymbol{S}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\|_\infty \le C$ and $\min\{r_1, r_0\} > c$. Consequently,

$$\gamma_n \le \frac{(K_n+1)^{7/4}}{\sqrt{nr_1r_0}} \max_{1\le i\le n} \left\|\boldsymbol{S}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\right\|_\infty^{3/2} \le \frac{(K_n+1)^{7/4}}{\sqrt{nc^2}} C^{3/2} = \left(\frac{K_n+1}{n^{2/7}}\right)^{7/4} \frac{C^{3/2}}{c},$$

under which $K_n = o(n^{2/7})$ implies that $\gamma_n = o(1)$. □

**A5. Limiting Behavior of the Constrained Gaussian Random Variable.** In this section, we prove Theorem 4 regarding the limiting behavior of the constrained Gaussian random variable $L_{K_n,a_n}$. We first give some technical lemmas in Section A5.1, and then study the limiting behavior of $L_{K_n,a_n}$ in Sections A5.2–A5.5 under various relationship between $\log(p_n^{-1})$ and $K_n$. Sections A5.2–A5.5 essentially prove Theorem 4(i)–(iv) respectively, as briefly commented in Section A5.6. For descriptive convenience, we introduce $\chi^2_K$ to denote a random variable following the chi-square distribution with degrees of freedom $K$.

A5.1. *Technical lemmas and their proofs.*

A5.1.1. *Lemmas for the acceptance probability* $p = \mathbb{P}(\chi^2_K \le a)$.

LEMMA A15. *For any integer* $K \ge 1$, $\sqrt{\pi K}\{K/(2e)\}^{K/2} \le \Gamma(K/2+1) \le 2\sqrt{\pi K}\{K/(2e)\}^{K/2}$.

PROOF OF LEMMA A15. We can numerically verify that Lemma A15 holds when $K = 1$. Below we consider only the case with $K \ge 2$. From Karatsuba (2001), the Gamma function can be bounded by

$$\sqrt{\pi}\left(\frac{K}{2e}\right)^{K/2}\left(K^3 + K^2 + \frac{K}{2} + \frac{1}{100}\right)^{1/6} \le \Gamma(K/2+1) \le \sqrt{\pi}\left(\frac{K}{2e}\right)^{K/2}\left(K^3 + K^2 + \frac{K}{2} + \frac{1}{30}\right)^{1/6},$$

$$1 \le \left(1 + \frac{1}{K} + \frac{1}{2K^2} + \frac{1}{100K^3}\right)^{1/6} \le \frac{\Gamma(K/2+1)}{\sqrt{\pi K}\{K/(2e)\}^{K/2}} \le \left(1 + \frac{1}{K} + \frac{1}{2K^2} + \frac{1}{30K^3}\right)^{1/6} \le 2.$$

From the above, Lemma A15 holds. □

LEMMA A16. *For any integer* $K \ge 1$ *and* $a > 0$, *define* $p = \mathbb{P}(\chi^2_K \le a)$. *Then*

$$\frac{\log(p^{-1})}{K} \le \frac{\log(4\pi K)}{2K} + \frac{1}{2}\left\{\frac{a}{K} - 1 - \log\left(\frac{a}{K}\right)\right\}.$$

*Moreover, if* $a/K < 1$, *then*

$$\frac{\log(p^{-1})}{K} \ge \frac{1}{2}\left\{\frac{a}{K} - 1 - \log\left(\frac{a}{K}\right)\right\} + \frac{\log(\pi K)}{2K} + \frac{1}{K}\log\left(1 - \frac{a}{K}\right).$$

PROOF OF LEMMA A16. By definition and using integration by parts, we have

$$p = \frac{1}{2^{K/2}\Gamma(K/2)}\int_0^a t^{K/2-1}e^{-t/2}\mathrm{d}t = \frac{1}{2^{K/2}\Gamma(K/2)} \cdot \frac{t^{K/2}}{K/2}e^{-t/2}\Big|_0^a + \frac{1}{2^{K/2}\Gamma(K/2)} \cdot \int_0^a \frac{t^{K/2}}{K/2}e^{-t/2}\frac{1}{2}\mathrm{d}t$$

$$= \frac{a^{K/2}e^{-a/2}}{2^{K/2}\Gamma(K/2+1)} + \frac{1}{K}\frac{1}{2^{K/2}\Gamma(K/2)}\int_0^a t \cdot t^{K/2-1}e^{-t/2}\mathrm{d}t \qquad \text{(A5.17)}$$

$$\leq \frac{a^{K/2}e^{-a/2}}{2^{K/2}\Gamma(K/2+1)} + \frac{a}{K}\frac{1}{2^{K/2}\Gamma(K/2)}\int_0^a t^{K/2-1}e^{-t/2}\mathrm{d}t$$
$$= \frac{a^{K/2}e^{-a/2}}{2^{K/2}\Gamma(K/2+1)} + \frac{a}{K}p,$$

and

$$p = \frac{a^{K/2}e^{-a/2}}{2^{K/2}\Gamma(K/2+1)} + \frac{1}{K}\frac{1}{2^{K/2}\Gamma(K/2)}\int_0^a t\cdot t^{K/2-1}e^{-t/2}\mathrm{d}t \geq \frac{a^{K/2}e^{-a/2}}{2^{K/2}\Gamma(K/2+1)}.$$

These implies that

$$\left(1-\frac{a}{K}\right)p \leq \frac{a^{K/2}e^{-a/2}}{2^{K/2}\Gamma(K/2+1)} = \frac{\left(a/K\cdot e^{1-a/K}\right)^{K/2}}{\sqrt{\pi K}}\frac{\sqrt{\pi K}\{K/(2e)\}^{K/2}}{\Gamma(K/2+1)} \leq p.$$

From Lemma A15, we then have

$$p \geq \frac{1}{2}\frac{\left(a/K\cdot e^{1-a/K}\right)^{K/2}}{\sqrt{\pi K}} \quad \text{and} \quad \left(1-\frac{a}{K}\right)p \leq \frac{\left(a/K\cdot e^{1-a/K}\right)^{K/2}}{\sqrt{\pi K}}.$$

Consequently,

$$\frac{\log(p^{-1})}{K} \leq \frac{\log(4\pi K)}{2K} + \frac{1}{2}\left\{\frac{a}{K} - 1 - \log\left(\frac{a}{K}\right)\right\}.$$

and, when $a/K < 1$,

$$\frac{\log(p^{-1})}{K} \geq \frac{1}{2}\left\{\frac{a}{K} - 1 - \log\left(\frac{a}{K}\right)\right\} + \frac{\log(\pi K)}{2K} + \frac{1}{K}\log\left(1-\frac{a}{K}\right).$$

Therefore, Lemma A16 holds. □

A5.1.2. *Lemmas for the variance of $L_{K,a}$ and its bounds.*

LEMMA A17. *For any integer $K > 0$ and $a > 0$,*

(i) *$Var(L_{K,a}) = K^{-1}\mathbb{E}(\chi^2_K \mid \chi^2_K \leq a) = \mathbb{P}(\chi^2_{K+2} \leq a)/\mathbb{P}(\chi^2_K \leq a)$.*
(ii) *$Var(L_{K,a})$ is nondecreasing in $a$ for any given fixed $K \geq 1$.*

PROOF OF LEMMA A17. Lemma A17 follows immediately from Morgan and Rubin (2012, Theorem 3.1) and Li, Ding and Rubin (2018, Lemma A5). □

LEMMA A18. *For any integer $K \geq 1$ and $a \geq 0$, it holds that*

$$\min\left\{\frac{a}{4K}, \frac{K-2}{4K}\right\} \leq Var(L_{K,a}) \leq \frac{a}{K}.$$

PROOF OF LEMMA A18. The upper bound of $\mathrm{Var}(L_{K,a})$ is a direct consequence of Lemma A17(i). The lower bound of $\mathrm{Var}(L_{K,a})$ holds obviously when $a = 0$ or $K \leq 2$. Below we consider only the lower bound of $\mathrm{Var}(L_{K,a})$ when $a > 0$ and $K \geq 3$. Define $\tilde{a} = \min\{a, K-2\}$. By the property of chi-square distribution, the density function of $\chi^2_K$ is monotonically increasing on the interval $[0, K-2] \supset [0, \tilde{a}]$. This implies that $\mathbb{P}(\chi^2_K \leq \tilde{a}/2) \leq \mathbb{P}(\tilde{a}/2 \leq \chi^2_K \leq \tilde{a})$ and

$$\mathbb{P}(\tilde{a}/2 \leq \chi^2_K \leq \tilde{a} \mid \chi^2_K \leq \tilde{a}) = \frac{\mathbb{P}(\tilde{a}/2 \leq \chi^2_K \leq \tilde{a})}{\mathbb{P}(\chi^2_K \leq \tilde{a}/2) + \mathbb{P}(\tilde{a}/2 \leq \chi^2_K \leq \tilde{a})} \geq 1/2. \tag{A5.18}$$

Consequently, from Lemma A17, the variance of $L_{K,a}$ multiplied by $K$ can be bounded by

$$K\text{Var}(L_{K,a}) \geq K\text{Var}(L_{K,\tilde{a}}) = \mathbb{E}[\chi^2_K \mid \chi^2_K \leq \tilde{a}] \geq \mathbb{P}(\tilde{a}/2 \leq \chi^2_K \leq \tilde{a} \mid \chi^2_K \leq \tilde{a}) \cdot \mathbb{E}[\chi^2_K \mid \tilde{a}/2 \leq \chi^2_K \leq \tilde{a}]$$
$$\geq \frac{1}{2} \cdot \frac{\tilde{a}}{2} = \frac{\min\{a, K-2\}}{4}$$

where the first inequality holds because $a \geq \tilde{a}$ and the last inequality holds due to (A5.18). From the above, Lemma A18 holds. □

LEMMA A19. *For any integer $K > 0$ and $a > 0$, with $p = \mathbb{P}(\chi^2_K \leq a)$,*

$$-\log\{1 - Var(L_{K,a})\} \geq \frac{K}{2}\left\{\frac{a}{K} - 1 - \log\left(\frac{a}{K}\right) - \frac{2\log(p^{-1})}{K} + \frac{\log(\pi K)}{K}\right\}.$$

PROOF OF LEMMA A19. From (A5.17),

$$p = \mathbb{P}(\chi^2_K \leq a) = \frac{a^{K/2}e^{-a/2}}{2^{K/2}\Gamma(K/2+1)} + \frac{1}{K}\frac{1}{2^{K/2}\Gamma(K/2)}\int_0^a t^{K/2}e^{-t/2}\text{d}t$$
$$= \frac{a^{K/2}e^{-a/2}}{2^{K/2}\Gamma(K/2+1)} + \frac{1}{2^{(K+2)/2}\Gamma((K+2)/2)}\int_0^a t^{(K+2)/2-1}e^{-t/2}\text{d}t$$
$$= \frac{a^{K/2}e^{-a/2}}{2^{K/2}\Gamma(K/2+1)} + \mathbb{P}(\chi^2_{K+2} \leq a).$$

From Lemmas A15 and A17, this implies that

(A5.19)
$$1 - \text{Var}(L_{K,a}) = 1 - \frac{\mathbb{P}(\chi^2_{K+2} \leq a)}{\mathbb{P}(\chi^2_K \leq a)} = \frac{a^{K/2}e^{-a/2}}{2^{K/2}\Gamma(K/2+1)} \cdot \frac{1}{p} \leq \frac{a^{K/2}e^{-a/2}}{2^{K/2}\sqrt{\pi K}\{K/(2e)\}^{K/2}} \cdot \frac{1}{p}$$
$$= \frac{(a/K)^{K/2}e^{(K-a)/2}}{p\sqrt{\pi K}}.$$

Consequently,

$$-\log\{1 - \text{Var}(L_{K,a})\} \geq -\frac{K}{2}\log\left(\frac{a}{K}\right) - \frac{K-a}{2} + \log(p) + \frac{1}{2}\log(\pi K)$$
$$= \frac{K}{2}\left\{\frac{a}{K} - 1 - \log\left(\frac{a}{K}\right) - \frac{2\log(p^{-1})}{K} + \frac{\log(\pi K)}{K}\right\}.$$

Therefore, Lemma A19 holds. □

LEMMA A20. *For any $K > 2$, $a \in (0, K-2]$ and $\zeta \in (0,1)$,*

$$-\log\{1 - Var(L_{K,a})\} \geq -\log(2) + \log(K\zeta) - \frac{\zeta}{2(1-\zeta)}\{a\zeta + (K-2-a)\}.$$

PROOF OF LEMMA A20. From (A5.19),

$$1 - \text{Var}(L_{K,a}) = \frac{a^{K/2}e^{-a/2}}{2^{K/2}\Gamma(K/2+1)} \cdot \frac{1}{\mathbb{P}(\chi^2_K \leq a)} = \frac{a^{K/2}e^{-a/2}}{2^{K/2}\Gamma(K/2+1)} \cdot \frac{2^{K/2}\Gamma(K/2)}{\int_0^a t^{K/2-1}e^{-t/2}\text{d}t}$$
$$= \frac{2}{K}\frac{a^{K/2}e^{-a/2}}{\int_0^a t^{K/2-1}e^{-t/2}\text{d}t}.$$

By the property of chi-square distribution, $t^{K/2-1}e^{-t/2}$ is nondecreasing in $t \in [0, K-2] \supset [0, a]$, which implies that

$$\int_0^a t^{K/2-1}e^{-t/2}\mathrm{d}t \geq \int_{(1-\zeta)a}^a t^{K/2-1}e^{-t/2}\mathrm{d}t \geq \zeta a \cdot \{(1-\zeta)a\}^{K/2-1}e^{-(1-\zeta)a/2}$$
$$= \zeta(1-\zeta)^{K/2-1}a^{K/2}e^{-(1-\zeta)a/2}.$$

Thus,

$$1 - \text{Var}(L_{K,a}) = \frac{2}{K}\frac{a^{K/2}e^{-a/2}}{\int_0^a t^{K/2-1}e^{-t/2}\mathrm{d}t} \leq \frac{2}{K}\frac{a^{K/2}e^{-a/2}}{\zeta(1-\zeta)^{K/2-1}a^{K/2}e^{-(1-\zeta)a/2}} = \frac{2}{K}\frac{e^{-\zeta a/2}}{\zeta(1-\zeta)^{K/2-1}},$$

and consequently

$$-\log\{1 - \text{Var}(L_{K,a})\} \geq -\log(2) + \log(K\zeta) + \frac{\zeta a}{2} + \left(\frac{K}{2} - 1\right)\log(1-\zeta).$$

Using the inequality that $\log(1+x) \geq x/(1+x)$ for all $x > -1$, we have $\log(1-\zeta) \geq -\zeta/(1-\zeta)$, and thus

$$\begin{aligned}
&-\log\{1 - \text{Var}(L_{K,a})\} \\
&\geq -\log(2) + \log(K\zeta) + \frac{\zeta a}{2} - \left(\frac{K}{2} - 1\right)\frac{\zeta}{1-\zeta} = -\log(2) + \log(K\zeta) + \frac{\zeta}{2(1-\zeta)}\{a - a\zeta - (K-2)\} \\
&= -\log(2) + \log(K\zeta) - \frac{\zeta}{2(1-\zeta)}\{a\zeta + (K-2-a)\}.
\end{aligned}$$

Therefore, Lemma A20 holds. □

A5.2. *Limiting behavior when* $\lim_{n\to\infty}\log(p_n^{-1})/K_n = \infty$.

LEMMA A21. *As $n \to \infty$, if $\log(p_n^{-1})/K_n \to \infty$, then $a_n/K_n \to 0$ and Var$(L_{K_n,a_n}) \to 0$.*

PROOF OF LEMMA A21. From Lemma A18, it suffices to show that $a_n/K_n \to 0$ as $n \to \infty$. We prove this by contradiction. If $a_n/K_n$ does not converge to zero, then there must exist a positive constant $c > 0$ and a subsequence $\{n_j : j = 1, 2, \ldots\}$ such that $a_{n_j}/K_{n_j} \geq c$ for all $j \geq 1$. Thus, for any $j \geq 1$, $p_{n_j} = \mathbb{P}(\chi^2_{K_{n_j}} \leq a_{n_j}) \geq \mathbb{P}(\chi^2_{K_{n_j}} \leq cK_{n_j})$. From Lemma A16, this then implies that

$$\frac{\log(p_{n_j}^{-1})}{K_{n_j}} \leq \frac{\log\{\mathbb{P}(\chi^2_{K_{n_j}} \leq cK_{n_j})^{-1}\}}{K_{n_j}} \leq \frac{\log(4\pi K_{n_j})}{2K_{n_j}} + \frac{c - 1 - \log(c)}{2} \leq \frac{\log(4\pi)}{2} + \frac{c-1-\log(c)}{2},$$

where the last inequality holds because $\log(4\pi K)/(2K)$ is decreasing in $K$ for $K \geq 1$. However, this contradicts the fact that $\log(p_n^{-1})/K_n \to \infty$. Therefore, we must have $a_n/K_n \to 0$ as $n \to \infty$. From the above, Lemma A21 holds. □

A5.3. *Limiting behavior when* $\limsup_{n\to\infty}\log(p_n^{-1})/K_n < \infty$.

LEMMA A22. *If $\limsup_{n\to\infty}\log(p_n^{-1})/K_n < \infty$, then $\liminf_{n\to\infty} a_n/K_n > 0$ and $\liminf_{n\to\infty}$ Var$(L_{K_n,a_n}) > 0$.*

PROOF OF LEMMA A22. We first prove $\liminf_{n\to\infty} a_n/K_n > 0$ by contradiction. If $\liminf_{n\to\infty} a_n/K_n = 0$, then there exists a subsequence $\{n_j : j = 1, 2, \dots\}$ such that $a_{n_j}/K_{n_j} \to 0$ as $j \to \infty$ and $a_{n_j}/K_{n_j} < 1$ for all $j$. From Lemma A16, we then have, for any $j \geq 1$,

$$\begin{aligned}\frac{\log(p_{n_j}^{-1})}{K_{n_j}} &\geq \frac{1}{2}\left\{\frac{a_{n_j}}{K_{n_j}} - 1 - \log\left(\frac{a_{n_j}}{K_{n_j}}\right)\right\} + \frac{\log(\pi K_{n_j})}{2K_{n_j}} + \frac{1}{K_{n_j}}\log\left(1 - \frac{a_{n_j}}{K_{n_j}}\right) \\ &\geq -\frac{1}{2} - \frac{1}{2}\log\left(\frac{a_{n_j}}{K_{n_j}}\right) + \log\left(1 - \frac{a_{n_j}}{K_{n_j}}\right),\end{aligned}$$

which converges to infinity as $j \to \infty$. However, this contradicts with the fact that $\limsup_{n\to\infty} \log(p_n^{-1})/K_n < \infty$. Thus, we must have $\liminf_{n\to\infty} a_n/K_n > 0$.

Second, we prove that $\liminf_{n\to\infty} \text{Var}(L_{K_n,a_n}) > 0$ by contradiction. If $\liminf_{n\to\infty} \text{Var}(L_{K_n,a_n}) = 0$, then there exists a subsequence $\{n_j : j = 1, 2, \dots\}$ such that $\text{Var}(L_{K_{n_j},a_{n_j}}) \to 0$ as $n \to \infty$ and $\text{Var}(L_{K_{n_j},a_{n_j}}) < 1/12$ for all $j$. Below we consider two cases, depending on whether $\limsup_{j\to\infty} K_{n_j}$ is greater than or equal to 3. If $\limsup_{j\to\infty} K_{n_j} \geq 3$, then there exists a further subsequence $\{m_1, m_2, \dots\} \subset \{n_1, n_2, \dots\}$ such that $K_{m_j} \geq 3$ for all $j$. Because $(K_{m_j} - 2)/(4K_{m_j}) \geq 1/12$, from Lemma A18, we must have $\text{Var}(L_{K_{m_j},a_{m_j}}) \geq a_{m_j}/(4K_{m_j})$. This then implies that

$$0 = \lim_{j\to\infty} \text{Var}(L_{K_{m_j},a_{m_j}}) \geq \liminf_{j\to\infty} a_{m_j}/(4K_{m_j}) \geq \liminf_{n\to\infty} a_n/(4K_n) > 0,$$

a contradiction. If $\limsup_{j\to\infty} K_{n_j} < 3$, then there exists a further subsequence $\{m_1, m_2, \dots\} \subset \{n_1, n_2, \dots\}$ such that $K_{m_j} \leq 2$ for all $j$. Because

$$0 < \liminf_{n\to\infty} a_n/(4K_n) \leq \liminf_{j\to\infty} a_{m_j}/(4K_{m_j}) \leq \liminf_{j\to\infty} a_{m_j}/4,$$

there must exist a positive constant $\underline{a} > 0$ such that $a_{m_j} > \underline{a}$ for all $j$. From Lemma A17, this then implies that

$$0 = \lim_{j\to\infty} \text{Var}(L_{K_{m_j},a_{m_j}}) \geq \liminf_{j\to\infty} \text{Var}(L_{K_{m_j},\underline{a}}) \geq \min\left\{\text{Var}(L_{1,\underline{a}}), \text{Var}(L_{2,\underline{a}})\right\} > 0,$$

a contradiction. Therefore, we must have $\liminf_{n\to\infty} \text{Var}(L_{K_n,a_n}) > 0$, i.e., Lemma A22 holds. □

A5.4. *Limiting behavior when* $\liminf_{n\to\infty} \log(p_n^{-1})/K_n > 0$.

LEMMA A23. *If* $\liminf_{n\to\infty} \log(p_n^{-1})/K_n > 0$, *then*

(i) *for any subsequence* $\{n_j : j = 1, 2, \dots\}$ *with* $\lim_{j\to\infty} K_{n_j} = \infty$, $\limsup_{j\to\infty} a_{n_j}/K_{n_j} < 1$.

(ii) $\limsup_{n\to\infty}$ *Var*$(L_{K_n,a_n}) < 1$.

PROOF OF LEMMA A23. First, we prove (i) by contradiction. Suppose that there exists a subsequence $\{n_j : j = 1, 2, \dots\}$ such that $K_{n_j} \to \infty$ as $j \to \infty$ and $\limsup_{j\to\infty} a_{n_j}/K_{n_j} \geq 1$. Then there exists a further subsequence $\{m_j : j = 1, 2, \dots\} \subset \{n_j : j = 1, 2, \dots\}$ such that $\lim_{j\to\infty} a_{m_j}/K_{m_j} \geq 1$. Define $\tilde{a}_{m_j} = \min\{1, a_{m_j}/K_{m_j}\} \cdot K_{m_j}$. We can then verify that $\tilde{a}_{m_j} \leq a_{m_j}$ and $\lim_{j\to\infty} \tilde{a}_{m_j}/K_{m_j} = 1$. From Lemma A16, for any $j \geq 1$,

$$\frac{\log(p_{m_j}^{-1})}{K_{m_j}} \leq \frac{\log\{\mathbb{P}(\chi^2_{K_{m_j}} \leq \tilde{a}_{m_j})^{-1}\}}{K} \leq \frac{\log(4\pi K_{m_j})}{2K_{m_j}} + \frac{1}{2}\left\{\frac{\tilde{a}_{m_j}}{K_{m_j}} - 1 - \log\left(\frac{\tilde{a}_{m_j}}{K_{m_j}}\right)\right\},$$

where the right hand side converges to 0 as $j \to \infty$. Consequently, $\log(p_{m_j}^{-1})/K_{m_j} \to 0$ as $j \to \infty$. However, this contradicts with the fact that $\liminf_{n\to\infty} \log(p_n^{-1})/K_n > 0$. Therefore, for any subsequence $\{n_j : j = 1, 2, \ldots\}$ with $\lim_{j\to\infty} K_{n_j} = \infty$, we must have $\limsup_{j\to\infty} a_{n_j}/K_{n_j} < 1$.

Second, we prove (ii) by contradiction. If $\limsup_{n\to\infty} \text{Var}(L_{K_n,a_n}) = 1$, then there exists a subsequence $\{n_j : j = 1, 2, \ldots\}$ such that $\text{Var}(L_{K_{n_j},a_{n_j}}) \to 1$ as $j \to \infty$. Below we consider two cases, depending on whether $\limsup_{j\to\infty} K_{n_j}$ is finite. If $\limsup_{j\to\infty} K_{n_j} = \infty$, then there exists a further subsequence $\{m_j : j = 1, 2, \ldots\} \subset \{n_j : j = 1, 2, \ldots\}$ such that $K_{m_j} \to \infty$ as $j \to \infty$. From Lemma A18 and the discussion before, we have

$$1 > \limsup_{j\to\infty} a_{m_j}/K_{m_j} \geq \limsup_{j\to\infty} \text{Var}(L_{K_{m_j},a_{m_j}}) = 1,$$

a contradiction. If $\limsup_{j\to\infty} K_{n_j} < \infty$, then there exists a finite integer $\overline{K}$ such that $K_{n_j} \leq \overline{K}$ for all $j$. Because $\liminf_{n\to\infty} \log(p_n^{-1})/K_n > 0$, there must exists a positive constant $c > 0$ such that $\log(p_n^{-1})/K_n \geq c$ for all $n$. This immediately implies that $\log(p_n^{-1}) \geq c$ and $p_n \leq e^{-c}$ for all $n$. Consequently, for all $j$, we have $a_{n_j} = F_{K_{n_j}}^{-1}(p_{n_j}) \leq F_{K_{n_j}}^{-1}(e^{-c}) \leq \max_{1\leq K\leq\overline{K}} F_K^{-1}(e^{-c}) \equiv \overline{a}$, where $F_K^{-1}(\cdot)$ denotes the quantile function for the chi-square distribution with degrees of freedom $K$. From Lemma A17, we then have

$$1 = \lim_{j\to\infty} \text{Var}(L_{K_{n_j},a_{n_j}}) \leq \limsup_{j\to\infty} \text{Var}(L_{K_{n_j},\overline{a}}) \leq \max_{1\leq K\leq\overline{K}} \text{Var}(L_{K,\overline{a}}) < 1,$$

a contradiction. Therefore, we must have $\limsup_{n\to\infty} \text{Var}(L_{K_n,a_n}) < 1$.

From the above, Lemma A23 holds. □

A5.5. *Limiting behavior when* $\lim_{n\to\infty} \log(p_n^{-1})/K_n = 0$.

LEMMA A24. *If* $\log(p_n^{-1})/K_n \to 0$ *as* $n \to \infty$, *then* $\liminf_{n\to\infty} a_n/K_n \geq 1$.

PROOF OF LEMMA A24. We prove the lemma by contradiction. Assume that $\liminf_{n\to\infty} a_n/K_n < 1$. Then there must exist a subsequence $\{n_j : j = 1, 2, \ldots\}$ such that $a_{n_j}/K_{n_j} < 1$ and $a_{n_j}/K_{n_j}$ converges to some $c \in [0, 1)$ as $n \to \infty$. From Lemma A16, for any $j \geq 1$,

$$\frac{\log(p_{n_j}^{-1})}{K_{n_j}} \geq \frac{1}{2}\left\{\frac{a_{n_j}}{K_{n_j}} - 1 - \log\left(\frac{a_{n_j}}{K_{n_j}}\right)\right\} + \frac{\log(\pi K_{n_j})}{2K_{n_j}} + \frac{1}{K_{n_j}} \log\left(1 - \frac{a_{n_j}}{K_{n_j}}\right).$$

If $\limsup_{j\to\infty} K_{n_j} = \infty$, then there exists a further subsequence $\{m_j : j = 1, 2, \ldots\}$ such that $K_{m_j} \to \infty$ as $j \to \infty$, under which we have

$$\begin{aligned}\frac{\log(p_{m_j}^{-1})}{K_{m_j}} &\geq \frac{1}{2}\left\{\frac{a_{m_j}}{K_{m_j}} - 1 - \log\left(\frac{a_{m_j}}{K_{m_j}}\right)\right\} + \frac{\log(\pi K_{m_j})}{2K_{m_j}} + \frac{1}{K_{m_j}} \log\left(1 - \frac{a_{m_j}}{K_{m_j}}\right) \\ &\to \frac{1}{2}(c - 1 - \log(c)) > 0.\end{aligned}$$

However, this contradicts with $\lim_{n\to\infty} \log(p_n^{-1})/K_n = 0$. If $\limsup_{j\to\infty} K_{n_j} < \infty$, then there exists a finite $\overline{K} < \infty$ such that $K_{n_j} \leq \overline{K}$ for all $j$. Thus, $a_{n_j} = a_{n_j}/K_{n_j} \cdot K_{n_j} \leq \overline{K}$ for all $j$, under which we have

$$p_{n_j} = \mathbb{P}(\chi^2_{K_{n_j}} \leq a_{n_j}) \leq \mathbb{P}(\chi^2_{K_{n_j}} \leq \overline{K}) \leq \max_{1\leq K\leq\overline{K}} \mathbb{P}(\chi^2_K \leq \overline{K}),$$

and

$$\frac{\log(p_{n_j}^{-1})}{K_{n_j}} \geq \frac{\log(p_{n_j}^{-1})}{\overline{K}} \geq \frac{1}{\overline{K}} \log\left\{\frac{1}{\max_{1\leq K\leq\overline{K}} \mathbb{P}(\chi^2_K \leq \overline{K})}\right\} > 0.$$

However, this contradicts with $\lim_{n\to\infty} \log(p_n^{-1})/K_n = 0$. From the above, Lemma A24 holds. $\square$

LEMMA A25. *If* $\log(p_n^{-1})/K_n \to 0$, *then* $Var(L_{K_n,a_n}) \to 1$.

PROOF OF LEMMA A25. We prove the lemma by contradiction. Assume $\liminf_{n\to\infty} \text{Var}(L_{K_n,a_n}) < 1$. Then there exist a constant $c \in [0,1)$ and a subsequence $\{n_j : j = 1,2,\ldots\}$ such that $\text{Var}(L_{K_{n_j},a_{n_j}}) \le c$ for all $j$. Below we consider several cases, depending on the values of $\liminf_{j\to\infty} K_{n_j}$, $\limsup_{j\to\infty} a_{n_j}/K_{n_j}$ and $\liminf_{j\to\infty}(K_{n_j} - a_{n_j})/\sqrt{K_{n_j}}$.

First, we consider the case in which $\liminf_{j\to\infty} K_{n_j} < \infty$. Then there exist a finite constant $\overline{K} < \infty$ and a subsequence $\{m_j : j = 1,2,\ldots\} \subset \{n_j : j = 1,2,\ldots\}$ such that $K_{m_j} \le \overline{K}$. Thus, for any $j \ge 1$, we have

$$\min_{1\le K\le \overline{K}} \log\left\{\frac{1}{\mathbb{P}(\chi^2_K \le a_{m_j})}\right\} \le \log\left\{\frac{1}{\mathbb{P}(\chi^2_{K_{m_j}} \le a_{m_j})}\right\} = \log(p_{m_j}^{-1}) \le \overline{K}\frac{\log(p_{m_j}^{-1})}{K_{m_j}} \to 0,$$

which must implies that $a_{m_j} \to \infty$ as $j \to \infty$. Consequently, from Lemma A17,

$$1 > c \ge \text{Var}(L_{K_{m_j},a_{m_j}}) \ge \min_{1\le K\le\overline{K}} \text{Var}(L_{K,a_{m_j}}) = \min_{1\le K\le\overline{K}} \frac{\mathbb{P}(\chi^2_{K+2} \le a_{m_j})}{\mathbb{P}(\chi^2_K \le a_{m_j})} \to 1,$$

a contradiction.

Second, we consider the case in which $\liminf_{j\to\infty} K_{n_j} = \infty$ and $\limsup_{j\to\infty} a_{n_j}/K_{n_j} > 1$. Then there exist a constant $\delta > 1$ and a further subsequence $\{m_j : j = 1,2,\ldots\} \subset \{n_j : j = 1,2,\ldots\}$ such that $a_{m_j}/K_{m_j} > \delta$ and $K_{m_j} \ge 3$ for all $j$. Note that the function $x - 1 - \log(x)$ is increasing in $x \in [1,\infty)$ and takes positive value when $x > 1$. From Lemma A19, we then have

$$-\frac{2\log(1-c)}{K_{m_j}} \ge -\frac{2\log\{1 - \text{Var}(L_{K_{m_j},a_{m_j}})\}}{K_{m_j}} \ge \frac{a_{m_j}}{K_{m_j}} - 1 - \log\left(\frac{a_{m_j}}{K_{m_j}}\right) - \frac{2\log(p_{m_j}^{-1})}{K_{m_j}} + \frac{\log(\pi K_{m_j})}{K_{m_j}}$$
$$\ge \delta - 1 - \log\delta - \frac{2\log(p_{m_j}^{-1})}{K_{m_j}} + \frac{\log(\pi K_{m_j})}{K_{m_j}} \to \delta - 1 - \log\delta > 0.$$

However, as $j \to \infty$, $K_{m_j} \to \infty$ and thus $-2\log(1-c)/K_{m_j} \to 0$, a contradiction.

Third, we consider the case in which $\liminf_{j\to\infty} K_{n_j} = \infty$ and $\liminf_{j\to\infty}(K_{n_j} - a_{n_j})/\sqrt{K_{n_j}} < \infty$. Then there exists a finite constant $\beta$ and a subsequence $\{m_j : j = 1,2,\ldots\} \subset \{n_j : j = 1,2,\ldots\}$ such that $(K_{m_j} - a_{m_j})/\sqrt{K_{m_j}} \le \beta$. By the central limit theorem,

$$p_{m_j} = \mathbb{P}(\chi^2_{K_{m_j}} \le a_{m_j}) = \mathbb{P}\left(\frac{\chi^2_{K_{m_j}} - K_{m_j}}{\sqrt{2K_{m_j}}} \le -\frac{K_{m_j} - a_{m_j}}{\sqrt{2K_{m_j}}}\right) \ge \mathbb{P}\left(\frac{\chi^2_{K_{m_j}} - K_{m_j}}{\sqrt{2K_{m_j}}} \le -\frac{\beta}{\sqrt{2}}\right) \to \Phi\left(-\frac{\beta}{\sqrt{2}}\right),$$

which implies that $\limsup_{j\to\infty} \log(p_{m_j}^{-1}) < \infty$. From Lemma A19 and the inequality $x - 1 - \log(x) \ge 0$ for all $x > 0$, we then have

$$-\log(1-c) \ge -\log\{1 - \text{Var}(L_{K_{m_j},a_{m_j}})\} \ge \frac{K_{m_j}}{2}\left\{\frac{a_{m_j}}{K_{m_j}} - 1 - \log\left(\frac{a_{m_j}}{K_{m_j}}\right)\right\} - \log(p_{m_j}^{-1}) + \frac{\log(\pi K_{m_j})}{2}$$
$$\ge -\log(p_{m_j}^{-1}) + \frac{\log(\pi K_{m_j})}{2} \to \infty,$$

a contradiction.

Finally, we consider the case in which $\liminf_{j\to\infty} K_{n_j} = \infty$, $\limsup_{j\to\infty} a_{n_j}/K_{n_j} \le 1$ and $\liminf_{j\to\infty}(K_{n_j} - a_{n_j})/\sqrt{K_{n_j}} = \infty$. From Lemma A24, $\liminf_{j\to\infty} a_{n_j}/K_{n_j} \ge \liminf_{n\to\infty} a_n/K_n \ge 1$, which implies that $a_{n_j}/K_{n_j} \to 1$ as $j \to \infty$. Moreover, there exists a finite $\underline{j}$ such that $a_{n_j} < K_{n_j} - 3$ for all $j \ge \underline{j}$. For any $j \ge \underline{j}$, define $\Delta_{n_j} = (K_{n_j} - 2 - a_{n_j})/K_{n_j}$ and $\zeta_{n_j} = \min\{\Delta_{n_j}^{-1/2}/K_{n_j}, K_{n_j}^{-3/4}\} \in (0,1)$. We can then verify that, as $j \to \infty$, $\Delta_{n_j} \to 0$, $\zeta_{n_j} \to 0$, $K_{n_j}\zeta_{n_j} = \min\{\Delta_{n_j}^{-1/2}, K_{n_j}^{1/4}\} \to \infty$, and

$$a_{n_j}\zeta_{n_j}^2 \le a_{n_j}K_{n_j}^{-3/2} = K_{n_j}^{-1/2}\frac{a_{n_j}}{K_{n_j}} \to 0, \quad \zeta_j(K_{n_j} - 2 - a_{n_j}) \le \Delta_{n_j}^{-1/2}\frac{K_{n_j} - 2 - a_{n_j}}{K_{n_j}} = \Delta_{n_j}^{1/2} \to 0.$$

From Lemma A20, this further implies that, for any $j \ge \underline{j}$,

$$-\log(1-c) \ge -\log\left\{1 - \text{Var}(L_{K_{n_j},a_{n_j}})\right\} \ge -\log(2) + \log(K_{n_j}\zeta_{n_j}) - \frac{a_{n_j}\zeta_{n_j}^2 + \zeta_{n_j}(K_{n_j} - 2 - a_{n_j})}{2(1-\zeta_{n_j})}$$
$$\to \infty,$$

a contradiction.

From the above, Lemma A25 holds. □

A5.6. *Proof of Theorem 4 and an additional proposition .*

**Proof of Theorem 4.** (i) is a direct consequence of Lemma A21. (ii) is a direct consequence of Lemma A22. (iii) is a direct consequence of Lemma A23. (iv) is a direct consequence of Lemma A25. □

The following proposition establishes the equivalence between convergence in probability and convergence of variance for the constrained Gaussian random variable discussed in Section 5.1.

PROPOSITION A2. As $n \to \infty$, $L_{K_n,a_n} \xrightarrow{\mathbb{P}} 0$ if and only if $\text{Var}(L_{K_n,a_n}) \to 0$.

PROOF OF PROPOSITION A2. The "if" is a direct consequence of Chebyshev's inequality. Below we focus on the "only if" direction. Suppose that $L_{K_n,a_n} \xrightarrow{\mathbb{P}} 0$. Note that when $a_n = \infty$, $L_{K_n,a_n} \sim \varepsilon$, a standard Gaussian random variable. From Lemma A17, for all $n \ge 1$, $\mathbb{E}(L_{K_n,a_n}^2) = \text{Var}(L_{K_n,a_n}) \le \text{Var}(L_{K_n,\infty}) = \text{Var}(\varepsilon) = 1$. From Durrett (2019, Theorem 4.6.3), to prove that $\text{Var}(L_{K_n,a_n}) = \mathbb{E}(L_{K_n,a_n}^2) \to 0$, it suffices to show that $\{L_{K_n,a_n}^2 : n \ge 1\}$ is uniformly integrable. From Li, Ding and Rubin (2018, Lemma A5), $|L_{K_n,a_n}|$ is stochastically smaller than or equal to $|\varepsilon|$. Because $x\mathbb{1}(x > c)$ is a nondecreasing function of $x \in [0,\infty)$ for any given $c > 0$, by the property of stochastic ordering, we have $\mathbb{E}\{L_{K_n,a_n}^2\mathbb{1}(L_{K_n,a_n}^2 > c)\} \le \mathbb{E}\{\varepsilon^2\mathbb{1}(\varepsilon^2 > c)\}$. By the dominated convergence theorem, this further implies that

$$\lim_{c\to\infty}\left(\sup_{n\ge1}\mathbb{E}\{L_{K_n,a_n}^2\mathbb{1}(L_{K_n,a_n}^2 > c)\}\right) \le \lim_{c\to\infty}\mathbb{E}\{\varepsilon^2\mathbb{1}(\varepsilon^2 > c)\} = 0,$$

i.e., $\{L_{K_n,a_n}^2 : n \ge 1\}$ is uniformly integrable. From the above, Proposition A2 holds. □

**A6. Asymptotics for Optimal Rerandomization.**

LEMMA A26. *For any two random variables $\psi$ and $\tilde{\psi}$, and any constant $\delta > 0$,*

$$\sup_{c\in\mathbb{R}}\left|\mathbb{P}(\tilde{\psi} \le c) - \mathbb{P}(\psi \le c)\right| \le \mathbb{P}(|\tilde{\psi} - \psi| > \delta) + \sup_{b\in\mathbb{R}}\mathbb{P}(b < \psi \le b + \delta)$$

PROOF OF LEMMA A26. For any $c \in \mathbb{R}$ and $\delta > 0$, we have

$$\begin{aligned}\mathbb{P}(\psi \le c) &= \mathbb{P}(\psi \le c-\delta) + \mathbb{P}(c-\delta < \psi \le c)\\ &\le \mathbb{P}(\psi \le c-\delta, |\tilde{\psi}-\psi| \le \delta) + \mathbb{P}(|\tilde{\psi}-\psi| > \delta) + \mathbb{P}(c-\delta < \psi \le c)\\ &\le \mathbb{P}(\tilde{\psi} \le c) + \mathbb{P}(|\tilde{\psi}-\psi| > \delta) + \mathbb{P}(c-\delta < \psi \le c)\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}(\psi \le c) &= \mathbb{P}(\psi \le c+\delta) - \mathbb{P}(c < \psi \le c+\delta) \ge \mathbb{P}(\psi \le c+\delta, |\tilde{\psi}-\psi| \le \delta) - \mathbb{P}(c < \psi \le c+\delta)\\ &\ge \mathbb{P}(\tilde{\psi} \le c, |\tilde{\psi}-\psi| \le \delta) - \mathbb{P}(c < \psi \le c+\delta)\\ &\ge \mathbb{P}(\tilde{\psi} \le c) - \mathbb{P}(|\tilde{\psi}-\psi| > \delta) - \mathbb{P}(c < \psi \le c+\delta).\end{aligned}$$

These imply that

$$\left|\mathbb{P}(\tilde{\psi} \le c) - \mathbb{P}(\psi \le c)\right| \le \mathbb{P}(|\tilde{\psi}-\psi| > \delta) + \sup_{b\in\mathbb{R}} \mathbb{P}(b < \psi \le b+\delta).$$

Taking supremum over $c$, we then derive Lemma A26. □

LEMMA A27. *Let $\{\psi_n\}$ and $\{\tilde{\psi}_n\}$ be two sequence of random variables satisfying that $\psi_n = \beta_n\varepsilon_0 + \zeta_n$ and $\psi_n - \tilde{\psi}_n = o_{\mathbb{P}}(\beta_n)$, where $\{\beta_n\}$ is a sequence of positive constants, $\{\zeta_n\}$ is a sequence of random variables independent of $\varepsilon_0$, and $\varepsilon_0$ is a random variable with bounded density. Then we have, as $n \to \infty$, $\sup_{c\in\mathbb{R}} |\mathbb{P}(\tilde{\psi}_n \le c) - \mathbb{P}(\psi_n \le c)| \to 0$.*

PROOF OF LEMMA A27. For any constant $\eta > 0$, using Lemma A26 with $\delta = \beta_n\eta$, we have

$$\begin{aligned}\sup_{c\in\mathbb{R}}\left|\mathbb{P}(\tilde{\psi}_n \le c) - \mathbb{P}(\psi_n \le c)\right| &\le \mathbb{P}(|\tilde{\psi}_n - \psi_n| > \beta_n\eta) + \sup_{b\in\mathbb{R}} \mathbb{P}(b < \psi_n \le b+\beta_n\eta)\\ &= \mathbb{P}(|\tilde{\psi}_n - \psi_n|/\beta_n > \eta) + \sup_{b\in\mathbb{R}} \mathbb{P}(b < \psi_n/\beta_n \le b+\eta).\end{aligned} \tag{A6.20}$$

Below we consider the two terms in (A6.20), separately. First, by the fact that $\psi_n - \tilde{\psi}_n = o_{\mathbb{P}}(\beta_n)$, the first term in (A6.20) satisfies that $\mathbb{P}(|\tilde{\psi}_n - \psi_n|/\beta_n > \eta) \to 0$ as $n \to \infty$. Second, let $C$ be the upper bound of the density of $\varepsilon_0$. For any $b \in \mathbb{R}$, we then have,

$$\mathbb{P}(b < \psi_n/\beta_n \le b+\eta \mid \zeta_n) = \mathbb{P}(b - \zeta_n/\beta_n < \varepsilon_0 \le b - \zeta_n/\beta_n + \eta \mid \zeta_n) \le C\eta,$$

and thus, by the law of iterated expectation,

$$\mathbb{P}(b < \psi_n/\beta_n \le b+\eta) = \mathbb{E}\{\mathbb{P}(b < \psi_n/\beta_n \le b+\eta \mid \zeta_n)\} \le C\eta.$$

Consequently, we have $\sup_{b\in\mathbb{R}} \mathbb{P}(b < \psi_n/\beta_n \le b+\eta) \le C\eta$.

From the above, for any constant $\eta > 0$,

$$\limsup_{n\to\infty}\sup_{c\in\mathbb{R}}\left|\mathbb{P}(\tilde{\psi}_n \le c) - \mathbb{P}(\psi_n \le c)\right| \le \limsup_{n\to\infty} \mathbb{P}(|\tilde{\psi}_n - \psi_n|/\beta_n > \eta) + C\eta \le C\eta.$$

Because the above inequality holds for any $\eta > 0$, we must have $\limsup_{n\to\infty}\sup_{c\in\mathbb{R}} |\mathbb{P}(\tilde{\psi}_n \le c) - \mathbb{P}(\psi_n \le c)| = 0$. Therefore, Lemma A27 holds. □

**Proof of Theorem 5.** From Theorem 4(i) and Proposition A2, under Condition 3, we must have $L_{K_n,a_n} = o_{\mathbb{P}}(1)$. From Condition 4, we can know that for sufficiently large $n$,

$1 - R_n^2$ is greater than certain positive constant, and $\sqrt{R_n^2}\, L_{K_n,a_n} = o_{\mathbb{P}}(\sqrt{1-R_n^2})$. Using Lemma A27 with $\psi_n = \sqrt{1-R_n^2}\,\varepsilon_0$ and $\tilde{\psi}_n = \sqrt{1-R_n^2}\,\varepsilon_0 + \sqrt{R_n^2}\, L_{K_n,a_n}$, we then have

$$\sup_{c\in\mathbb{R}} \left| \mathbb{P}\left(\sqrt{1-R_n^2}\,\varepsilon_0 \le c\right) - \mathbb{P}\left\{\left(\sqrt{1-R_n^2}\,\varepsilon_0 + \sqrt{R_n^2}\, L_{K_n,a_n}\right) \le c\right\} \right| \to 0.$$

From Theorem 3, as $n \to \infty$,

$$\begin{aligned}
&\sup_{c\in\mathbb{R}} \left| \mathbb{P}\left\{V_{\tau\tau}^{-1/2}(\hat{\tau}-\tau) \le c \mid M \le a_n\right\} - \mathbb{P}\left\{\sqrt{1-R_n^2}\,\varepsilon_0 \le c\right\} \right| \\
&\le \sup_{c\in\mathbb{R}} \left| \mathbb{P}\left\{V_{\tau\tau}^{-1/2}(\hat{\tau}-\tau) \le c \mid M \le a_n\right\} - \mathbb{P}\left\{\left(\sqrt{1-R_n^2}\,\varepsilon_0 + \sqrt{R_n^2}\, L_{K_n,a_n}\right) \le c\right\} \right| \\
&+ \sup_{c\in\mathbb{R}} \left| \mathbb{P}\left\{\sqrt{1-R_n^2}\,\varepsilon_0 \le c\right\} - \mathbb{P}\left\{\left(\sqrt{1-R_n^2}\,\varepsilon_0 + \sqrt{R_n^2}\, L_{K_n,a_n}\right) \le c\right\} \right| \\
&\to 0.
\end{aligned}$$

Therefore, Theorem 5 holds. □

**Proof of Theorem 6.** Note that Condition 2 is that $p_n/\Delta_n \to \infty$ as $n \to \infty$, and Condition 3 is that $\log(p_n^{-1})/K_n \to \infty$ as $n \to \infty$. Below we prove Theorem 6(i)–(iv) respectively.

First, we prove (i). Consider first the "only if" part. If both Conditions 2 and 3 hold for some sequence $\{p_n\}$, then we must have, for sufficiently large $n$, $\log(\Delta_n^{-1})/K_n = \log(p_n^{-1})/K_n + \log(p_n/\Delta_n)/K_n \ge \log(p_n^{-1})/K_n$, which must imply that $\log(\Delta_n^{-1})/K_n \to \infty$ as $n \to \infty$. Consider then the "if" part. Because $\log(\Delta_n^{-1})/K_n \to \infty$ as $n \to \infty$, we can construct a sequence $\{p_n\}$ such that, as $n \to \infty$, $\Delta_n/p_n \to 0$, and $\log(\Delta_n/p_n)/K_n + \log(\Delta_n^{-1})/K_n \to \infty$. For such a choice of $\{p_n\}$, Condition 2 holds obviously, and $\log(p_n^{-1})/K_n = \log(\Delta_n/p_n)/K_n + \log(\Delta_n^{-1})/K_n \to \infty$, i.e., Condition 3 holds.

Second, we prove (ii). For any sequence $\{p_n\}$ such that Condition 2 holds, we have

$$\limsup_{n\to\infty} \frac{\log(p_n^{-1})}{K_n} \le \limsup_{n\to\infty} \frac{\log(\Delta_n/p_n)}{K_n} + \limsup_{n\to\infty} \frac{\log(\Delta_n^{-1})}{K_n} \le \limsup_{n\to\infty} \frac{\log(\Delta_n^{-1})}{K_n} < \infty.$$

Fro Theorem 4(ii), this further implies that $\liminf_{n\to\infty} v_{K_n,a_n} > 0$.

Third, we prove (iii). Because Condition 1 holds, from Theorem 1, we can construct a sequence $\{p_n\}$ such that $\lim_{n\to\infty} p_n/\Delta_n = \infty$ and $\limsup_{n\to\infty} \log(p_n/\Delta_n)/\log(\Delta_n^{-1}) < 1 - c$ for some $c > 0$. This then implies that

$$\liminf_{n\to\infty} \frac{\log(p_n^{-1})}{K_n} = \liminf_{n\to\infty} \left[\frac{\log(\Delta_n^{-1})}{K_n}\left\{1 - \log(p_n/\Delta_n)/\log(\Delta_n^{-1})\right\}\right] \ge c \liminf_{n\to\infty} \frac{\log(\Delta_n^{-1})}{K_n} > 0.$$

From Theorem 4(iii), we then have $\limsup_{n\to\infty} v_{K_n,a_n} < 1$.

Fourth, we prove (iv). For any sequence $\{p_n\}$ such that Condition 2 holds, we have

$$\limsup_{n\to\infty} \frac{\log(p_n^{-1})}{K_n} \le \limsup_{n\to\infty} \frac{\log(\Delta_n/p_n)}{K_n} + \limsup_{n\to\infty} \frac{\log(\Delta_n^{-1})}{K_n} \le \limsup_{n\to\infty} \frac{\log(\Delta_n^{-1})}{K_n} = 0.$$

From Theorem 4(iv), this further implies that $v_{K_n,a_n} \to 0$ as $n \to \infty$. □

**A7. Asymptotic Validity of Confidence Intervals.**

A7.1. *Technical lemmas.* For descriptive convenience, throughout this section, we define $a/b$ as $+\infty$ when $a > 0$ and $b = 0$.

LEMMA A28. *Let $\{u_i \in \mathbb{R} : i = 1, 2, \ldots, n\}$ be a finite population of $N > 0$ units, with $\bar{u} = N^{-1}\sum_{i=1}^{N} u_i$ and $\sigma_u^2 = N^{-1}\sum_{i=1}^{N}(u_i - \bar{u})^2$. Let $(Z_1, \cdots, Z_N)$ denote a sampling indicator vector for a simple random sample of size $m > 0$, and $\hat{u} = m^{-1}\sum_{i=1}^{N} Z_i u_i$ denote the corresponding sample average. Define $f = m/N$. Then for any $t > 0$,*

$$\mathbb{P}\left(|\hat{u} - \bar{u}| \geq t\right) \leq 2\exp\left(-\frac{70^2}{71^2}\frac{Nf^2t^2}{\sigma_u^2}\right).$$

LEMMA A29. *Let $\{(u_i, w_i) \in \mathbb{R}^2 : i = 1, 2, \ldots, N\}$ be a finite population of $N \geq 2$ units, with finite population averages and covariance $\bar{u} \equiv N^{-1}\sum_{i=1}^{N} u_i$, $\bar{w} = N^{-1}\sum_{i=1}^{N} w_i$ and $S_{uw} = (N-1)^{-1}\sum_{i=1}^{N}(u_i - \bar{u})(w_i - \bar{w})$. Let $(Z_1, \cdots, Z_N)$ denote a sampling indicator vector for a simple random sample of size $m \geq 2$, with corresponding sample averages and covariance $\hat{u} = m^{-1}\sum_{i=1}^{N} Z_i u_i$, $\hat{w} = m^{-1}\sum_{i=1}^{N} Z_i w_i$ and $s_{uw} = (m-1)^{-1}\sum_{i=1}^{N} Z_i(u_i - \hat{u})(w_i - \hat{w})$. Define $f = m/N$,*

$$\Delta_u = \hat{u} - \bar{u}, \quad \Delta_w = \hat{w} - \bar{w}, \quad \Delta_{uw} = \frac{1}{m}\sum_{i=1}^{N} Z_i(u_i - \bar{u})(w_i - \bar{w}) - \frac{N-1}{N}S_{uw},$$

*and*

$$\sigma_u^2 = \frac{1}{N}\sum_{i=1}^{N}(u_i - \bar{u})^2, \;\; \sigma_w^2 = \frac{1}{N}\sum_{i=1}^{N}(w_i - \bar{w})^2, \;\; \sigma_{u\times w}^2 = \frac{1}{N}\sum_{i=1}^{N}\left\{(u_i - \bar{u})(w_i - \bar{w}) - \frac{N-1}{N}S_{uw}\right\}^2.$$

*Then $|s_{uw} - S_{uw}| \leq 2|\Delta_{u\times w}| + 2|\Delta_u||\Delta_w| + 2(1-f)|S_{uw}|/m$, and for any $t > 0$,*

$$\mathbb{P}\left(|\Delta_u| \geq t\right) \leq 2\exp\left(-\frac{70^2}{71^2}\frac{Nf^2t^2}{\sigma_u^2}\right), \quad \mathbb{P}\left(|\Delta_w| \geq t\right) \leq 2\exp\left(-\frac{70^2}{71^2}\frac{Nf^2t^2}{\sigma_w^2}\right),$$

$$\mathbb{P}\left(|\Delta_{u\times w}| \geq t\right) \leq 2\exp\left(-\frac{70^2}{71^2}\frac{Nf^2t^2}{\sigma_{u\times w}^2}\right).$$

LEMMA A30. *Let $\{(u_i, \boldsymbol{w}_i^\top) \in \mathbb{R}^{1+K} : i = 1, 2, \ldots, N\}$ be a finite population of $N \geq 2$ units, with $\boldsymbol{w}_i = (w_{1i}, w_{2i}, \ldots w_{Ki})^\top$ and finite population averages and covariance $\bar{u} \equiv N^{-1}\sum_{i=1}^{N} u_i$, $\bar{\boldsymbol{w}} = (\bar{w}_1, \ldots, \bar{w}_K)^\top = N^{-1}\sum_{i=1}^{N} \boldsymbol{w}_i$ and $\boldsymbol{S}_{u\boldsymbol{w}} = (S_{uw_1}, \ldots, S_{uw_K}) = (N-1)^{-1}\sum_{i=1}^{N}(u_i - \bar{u})(\boldsymbol{w}_i - \bar{\boldsymbol{w}})^\top$. Let $(Z_1, \cdots, Z_N)$ denote a sampling indicator vector for a simple random sample of size $m \geq 2$, with corresponding sample averages and covariance $\hat{u} = m^{-1}\sum_{i=1}^{N} Z_i u_i$, $\hat{\boldsymbol{w}} = m^{-1}\sum_{i=1}^{N} Z_i \boldsymbol{w}_i$ and $\boldsymbol{s}_{u\boldsymbol{w}} = (s_{uw_1}, \ldots, s_{uw_K}) = (m-1)^{-1}\sum_{i=1}^{N} Z_i(u_i - \hat{u})(\boldsymbol{w}_i - \hat{\boldsymbol{w}})^\top$. Let $f = m/N$, and for $1 \leq k \leq K$, define*

$$\Delta_u = \hat{u} - \bar{u}, \quad \Delta_{w_k} = \hat{w}_k - \bar{w}_k, \quad \Delta_{uw_k} = \frac{1}{m}\sum_{i=1}^{N} Z_i(u_i - \bar{u})(w_{ki} - \bar{w}_k) - \frac{N-1}{N}S_{uw_k},$$

*and*

$$\sigma_u^2 = \frac{1}{N}\sum_{i=1}^{N}(u_i - \bar{u})^2, \;\; \sigma_{w_k}^2 = \frac{1}{N}\sum_{i=1}^{N}(w_{ki} - \bar{w}_k)^2, \;\; \sigma_{u\times w_k}^2 = \frac{1}{N}\sum_{i=1}^{N}\left\{(u_i - \bar{u})(w_{ki} - \bar{w}_k) - \frac{N-1}{N}S_{uw_k}\right\}^2.$$

*Then*

$$\|\boldsymbol{s}_{u\boldsymbol{w}} - \boldsymbol{S}_{u\boldsymbol{w}}\|_2^2 \leq 12\sum_{k=1}^{K}\Delta_{u\times w_k}^2 + 12\Delta_u^2\sum_{k=1}^{K}\Delta_{w_k}^2 + \frac{12(1-f)^2}{m^2}\sum_{k=1}^{K}S_{uw_k}^2,$$

*and for any $t > 0$,*

$$\mathbb{P}\left(\Delta_u^2 \geq t\right) \leq 2\exp\left(-\frac{70^2}{71^2}\frac{Nf^2 t}{\sigma_u^2}\right), \qquad \mathbb{P}\left(\sum_{k=1}^K \Delta_{w_k}^2 \geq t\right) \leq 2K\exp\left(-\frac{70^2}{71^2}\frac{Nf^2 t}{\sum_{k=1}^K \sigma_{w_k}^2}\right),$$

$$\mathbb{P}\left(\sum_{k=1}^K \Delta_{u\times w_k}^2 \geq t\right) \leq 2K\exp\left(-\frac{70^2}{71^2}\frac{Nf^2 t}{\sum_{k=1}^K \sigma_{u\times w_k}^2}\right).$$

LEMMA A31. *Consider the same setting as in Lemma A30 and any event $\boldsymbol{Z} \in \mathcal{E} \subset \{0,1\}^N$ with positive probability $p = \mathbb{P}(\boldsymbol{Z} \in \mathcal{E})$. Define*

$$\xi = \frac{\max\{1, \log K, -\log p\}}{Nf^2}\sum_{k=1}^K \sigma_{u\times w_k}^2 + \frac{\max\{1, -\log p\}\cdot\max\{1,\log K, -\log p\}}{N^2 f^4}\sigma_u^2 \sum_{k=1}^K \sigma_{w_k}^2$$
$$+ \frac{(1-f)^2}{N^2 f^2}\sum_{k=1}^K S_{uw_k}^2.$$

*Then for any $t \geq 3\cdot 71^2/70^2$,*

$$\mathbb{P}\left(\|\boldsymbol{s}_{u\boldsymbol{w}} - \boldsymbol{S}_{u\boldsymbol{w}}\|_2^2 > 36 t^2 \xi \mid \boldsymbol{Z} \in \mathcal{E}\right) \leq 6\exp\left(-\frac{1}{3}\frac{70^2}{71^2}t\right).$$

LEMMA A32. *Under ReM with threshold $a_n$, along the sequence of finite populations with increasing sample size $n$, if $\min\{n_1, n_0\} \geq 2$ when $n$ is sufficiently large, then the estimators $\hat{V}_{\tau\tau}$ and $\hat{R}^2$ satisfy that*

$$\hat{V}_{\tau\tau} - V_{\tau\tau} - n^{-1}S^2_{\tau\setminus \boldsymbol{X}} = O_{\mathbb{P}}\left(\frac{\xi_{11}^{1/2}}{n_1} + \frac{\xi_{00}^{1/2}}{n_0} + \frac{\xi_{1\boldsymbol{w}} + \xi_{0\boldsymbol{w}}}{n} + \|S_{1\boldsymbol{w}} - S_{0\boldsymbol{w}}\|_2 \frac{\xi_{1\boldsymbol{w}}^{1/2} + \xi_{0\boldsymbol{w}}^{1/2}}{n}\right),$$

*and*

$$\hat{V}_{\tau\tau}\hat{R}_n^2 - V_{\tau\tau}R_n^2 = O_{\mathbb{P}}\left(\frac{\xi_{1\boldsymbol{w}}}{n_1} + \frac{\xi_{0\boldsymbol{w}}}{n_0} + \|S_{1\boldsymbol{w}}\|_2\frac{\xi_{1\boldsymbol{w}}^{1/2}}{n_1} + \|S_{0\boldsymbol{w}}\|_2\frac{\xi_{0\boldsymbol{w}}^{1/2}}{n_1} + \|S_{1\boldsymbol{w}} - S_{0\boldsymbol{w}}\|_2\frac{\xi_{1\boldsymbol{w}}^{1/2} + \xi_{0\boldsymbol{w}}^{1/2}}{n}\right),$$

*where $\boldsymbol{w}_i = (w_{1i}, \ldots, w_{K_n i})^\top = \boldsymbol{S}_{\boldsymbol{X}}^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})$ is the standardized covariates, $S_{z\boldsymbol{w}} = (S_{zw_1}, \ldots, S_{zw_{K_n}})$ is the finite population covariance between $Y(z)$ and $\boldsymbol{w}$,*

$$\xi_{zz} = \frac{\max\{1, -\log\tilde{p}_n\}}{nr_z^2}\sigma_{z\times z}^2 + \frac{\max\{1, (-\log\tilde{p}_n)^2\}}{n^2 r_z^4}\sigma_z^4 + \frac{(1-r_z)^2}{n^2 r_z^2}S_z^4,$$

$$\xi_{z\boldsymbol{w}} = \frac{\max\{1, \log K_n, -\log\tilde{p}_n\}}{nr_z^2}\sum_{k=1}^K \sigma_{z\times w_k}^2 + \frac{\max\{1, -\log\tilde{p}_n\}\cdot\max\{1, \log K_n, -\log\tilde{p}_n\}}{n^2 r_z^4}\sigma_u^2\sum_{k=1}^{K_n}\sigma_{w_k}^2$$
$$+ \frac{(1-r_z)^2}{n^2 r_z^2}\sum_{k=1}^{K_n} S_{zw_k}^2,$$

*$\tilde{p}_n = \mathbb{P}(M \leq a_n)$ is the actual acceptance probability under ReM, and*

$$\sigma_z^2 = \frac{1}{n}\sum_{i=1}^n\{Y_i(z) - \bar{Y}(z)\}^2 = \frac{n-1}{n}S_z^2, \qquad \sigma_{w_k}^2 = \frac{1}{n}\sum_{i=1}^n (w_{ki} - \bar{w}_k)^2 = \frac{n-1}{n},$$

$$\sigma_{z\times z}^2 = \frac{1}{n}\sum_{i=1}^n\left[\{Y_i(z) - \bar{Y}(z)\}^2 - \sigma_z^2\right]^2, \;\; \sigma_{z\times w_k}^2 = \frac{1}{n}\sum_{i=1}^n\left[\{Y_i(z) - \bar{Y}(z)\}(w_{ki} - \bar{w}_k) - \frac{n-1}{n}S_{zw_k}\right]^2.$$

LEMMA A33. *Under the same setting as Lemma A32, if* $\max\{1, \log K_n, -\log \tilde{p}_n\} = O(nr_1^2 r_0^2)$*, then*

$$\max\left\{\left|\hat{V}_{\tau\tau} - V_{\tau\tau} - n^{-1}S^2_{\tau\setminus \boldsymbol{X}}\right|, \ \left|\hat{V}_{\tau\tau}\hat{R}_n^2 - V_{\tau\tau}R_n^2\right|\right\}$$
$$= \max_{z\in\{0,1\}} \max_{1\le i\le n} \{Y_i(z) - \bar{Y}(z)\}^2 \cdot O_{\mathbb{P}}\left(\max\{K_n, 1\} \cdot \frac{\sqrt{\max\{1, \log K_n, -\log \tilde{p}_n\}}}{n^{3/2} r_1^2 r_0^2}\right).$$

LEMMA A34. *Under the same setting as Lemmas A32 and A33,*

(i) *if Condition 2 holds, then* $\max\{1, -\log \tilde{p}_n\} = O(\max\{1, -\log p_n\})$*, recalling that* $\tilde{p}_n = \mathbb{P}(M \le a_n)$ *is the actually acceptance probability under ReM, while* $p_n = \mathbb{P}(\chi^2_{K_n} \le a_n)$ *is the approximate acceptance probability;*
(ii) $\max_{z\in\{0,1\}} \max_{1\le i\le n}\{Y_i(z) - \bar{Y}(z)\}^2/(r_0 S^2_{1\setminus \boldsymbol{X}} + r_1 S^2_{0\setminus \boldsymbol{X}}) \ge 1/2$*;*
(iii) *if Conditions 2 and 5 hold, then,* $\max\{1, \log K_n, -\log \tilde{p}_n\} = o(nr_1^2 r_0^2)$.

A7.2. *Proofs of the lemmas.*

PROOF OF LEMMA A28. When $\sigma_u^2 = 0$, $u_1 = \ldots = u_N = \bar{u}$, and thus $\hat{u} - \bar{u}$ must be a constant zero, under which Lemma A28 holds obviously. Below we consider only the case where $\sigma_u^2 > 0$. From Bloniarz et al. (2016, Lemma S1), for any $t > 0$,

$$\mathbb{P}\left(|\hat{u} - \bar{u}| \ge t\right) = \mathbb{P}\left(\hat{u} - \bar{u} \ge t\right) + \mathbb{P}\left\{(-\hat{u}) - (-\bar{u}) \ge t\right\} \le 2\exp\left(-\frac{fmt^2}{(1+c)^2\sigma_u^2}\right),$$

where $c \equiv \min\{1/70, (3f)^2/70, (3-3f)^2/70\} \le 1/70$. This then implies that for any $t > 0$,

$$\mathbb{P}\left(|\hat{u} - \bar{u}| \ge t\right) \le 2\exp\left(-\frac{fmt^2}{(1+c)^2\sigma_u^2}\right) \le 2\exp\left(-\frac{70^2}{71^2}\frac{Nf^2t^2}{\sigma_u^2}\right),$$

i.e., Lemma A28 holds. □

PROOF OF LEMMA A29. First, by definition, the sample covariance between $u$ and $w$ has the following equivalent forms:

$$s_{uw} = \frac{1}{m-1}\sum_{i=1}^N Z_i(u_i - \hat{u})(w_i - \hat{w}) = \frac{m}{m-1}\frac{1}{m}\sum_{i=1}^N Z_i(u_i - \bar{u})(w_i - \bar{w}) - \frac{m}{m-1}(\hat{u} - \bar{u})(\hat{w} - \bar{w})$$
$$= \frac{m}{m-1}\left\{\frac{1}{m}\sum_{i=1}^N Z_i(u_i - \bar{u})(w_i - \bar{w}) - \frac{N-1}{N}S_{uw}\right\} - \frac{m}{m-1}(\hat{u} - \bar{u})(\hat{w} - \bar{w}) + \frac{m(N-1)}{(m-1)N}S_{uw}.$$

Consequently, we can bound the difference between $s_{uw}$ and $S_{uw}$ by

$$|s_{uw} - S_{uw}|$$
$$= \left|\frac{m}{m-1}\left\{\frac{1}{m}\sum_{i=1}^N Z_i(u_i - \bar{u})(w_i - \bar{w}) - \frac{N-1}{N}S_{uw}\right\} - \frac{m}{m-1}(\hat{u} - \bar{u})(\hat{w} - \bar{w}) + \frac{1-f}{m-1}S_{uw}\right|$$
$$\le 2\left|\Delta_{u\times w}\right| + 2\left|\Delta_u\right|\left|\Delta_w\right| + \frac{2(1-f)}{m}\left|S_{uw}\right|,$$

where the last inequality holds because $m/(m-1) \le 2$.

Second, applying Lemma A28 to the finite populations of $\{u_i\}_{i=1}^n$, $\{w_i\}_{i=1}^n$ and $\{(u_i - \bar{u})(w_i - \bar{w})\}_{i=1}^n$, we can immediately derive the probability bounds for $\Delta_u$, $\Delta_w$ and $\Delta_{u\times w}$.

From the above, Lemma A29 holds. □

PROOF OF LEMMA A30. First, we consider the bound for $\|\boldsymbol{s}_{\boldsymbol{uw}}-\boldsymbol{S}_{\boldsymbol{uw}}\|_2^2$. From Lemma A29 and the Cauchy-Schwartz inequality,

$$\begin{aligned}\|\boldsymbol{s}_{\boldsymbol{uw}}-\boldsymbol{S}_{\boldsymbol{uw}}\|_2^2 &= \sum_{k=1}^K (s_{uw_k}-S_{uw_k})^2 \le 4\sum_{k=1}^K (|\Delta_{u\times w_k}|+|\Delta_u||\Delta_{w_k}|+(1-f)|S_{uw_k}|/m)^2\\ &\le 12\sum_{k=1}^K \left(\Delta^2_{u\times w_k}+\Delta_u^2\Delta_{w_k}^2+(1-f)^2S^2_{uw_k}/m^2\right)\\ &= 12\sum_{k=1}^K \Delta^2_{u\times w_k}+12\Delta_u^2\sum_{k=1}^K\Delta^2_{w_k}+\frac{12(1-f)^2}{m^2}\sum_{k=1}^K S^2_{uw_k}.\end{aligned}$$

Second, the probability bound for $\Delta_u^2$ follows immediately from Lemma A29.

Third, we consider the probability bound for $\sum_{k=1}^K\Delta^2_{w_k}$. We consider two cases separately, depending on whether $\sum_{j=1}^K\sigma^2_{w_j}$ is positive. When $\sum_{j=1}^K\sigma^2_{w_j}>0$, we introduce $a_k=\sigma^2_{w_k}/\sum_{j=1}^K\sigma^2_{w_j}$ for $1\le k\le K$. Obviously, $a_k\ge 0$ for all $k$ and $\sum_{k=1}^K a_k=1$. Note that if $a_k=0$ for some $1\le k\le K$, then it follows from Lemma A28 that the corresponding $\Delta_{w_k}$ is a constant zero. With this in mind, from Lemma A29, we have that for any $t>0$,

$$\begin{aligned}\mathbb{P}\left(\sum_{k=1}^K\Delta^2_{w_k}\ge t\right) &= \mathbb{P}\left(\sum_{k:a_k>0}\Delta^2_{w_k}\ge\sum_{k:a_k>0}a_kt\right)\le\sum_{k:a_k>0}\mathbb{P}\left(\Delta^2_{w_k}\ge a_kt\right)\le 2\sum_{k:a_k>0}\exp\left(-\frac{70^2}{71^2}\frac{Nf^2a_kt}{\sigma^2_{w_k}}\right)\\ &= 2\sum_{k:a_k>0}\exp\left(-\frac{70^2}{71^2}\frac{Nf^2t}{\sum_{j=1}^K\sigma^2_{w_j}}\right)\le 2K\exp\left(-\frac{70^2}{71^2}\frac{Nf^2t}{\sum_{k=1}^K\sigma^2_{w_k}}\right).\end{aligned}$$

When $\sum_{j=1}^K\sigma^2_{w_j}=0$, $\sum_{k=1}^K\Delta^2_{w_k}$ is a constant zero, under which the above probability bound holds obviously.

Fourth, we consider the probability bound for $\sum_{k=1}^K\Delta^2_{u\times w_k}$. By the same logic as the proof above for the probability bound of $\sum_{k=1}^K\Delta^2_{w_k}$, we can derive that, for any $t>0$,

$$\mathbb{P}\left(\sum_{k=1}^K\Delta^2_{u\times w_k}\ge t\right)\le 2K\exp\left(-\frac{70^2}{71^2}\frac{Nf^2t}{\sum_{k=1}^K\sigma^2_{u\times w_k}}\right).$$

From the above, Lemma A30 holds. □

PROOF OF LEMMA A31. From Lemma A30, we have that for any $t>0$,

$$\begin{aligned}&\mathbb{P}\left(\|\boldsymbol{s}_{\boldsymbol{uw}}-\boldsymbol{S}_{\boldsymbol{uw}}\|_2^2>36t^2\xi\mid\boldsymbol{Z}\in\mathcal{E}\right)\\ &\le\frac{\mathbb{P}\left(\|\boldsymbol{s}_{\boldsymbol{uw}}-\boldsymbol{S}_{\boldsymbol{uw}}\|_2^2>36t^2\xi\right)}{\mathbb{P}\left(\boldsymbol{Z}\in\mathcal{E}\right)}\le\frac{1}{p}\mathbb{P}\left(12\sum_{k=1}^K\Delta^2_{u\times w_k}+12\Delta_u^2\sum_{k=1}^K\Delta^2_{w_k}+\frac{12(1-f)^2}{N^2f^2}\sum_{k=1}^K S^2_{uw_k}>36t^2\xi\right)\end{aligned}$$

(A7.21)

$$\le\frac{1}{p}\mathbb{P}\left(\sum_{k=1}^K\Delta^2_{u\times w_k}>t^2\xi\right)+\frac{1}{p}\mathbb{P}\left(\Delta_u^2\sum_{k=1}^K\Delta^2_{w_k}>t^2\xi\right)+\frac{1}{p}\mathbb{P}\left(\frac{(1-f)^2}{N^2f^2}\sum_{k=1}^K S^2_{uw_k}>t^2\xi\right).$$

Below we consider the three terms in (A7.21) separately.

First, we prove that, for any $t^2 \geq 3 \cdot 71^2/70^2$,

$$\text{(A7.22)} \qquad \frac{1}{p}\mathbb{P}\left(\sum_{k=1}^{K} \Delta^2_{u\times w_k} > t^2\xi\right) \leq 2\exp\left(-\frac{1}{3}\frac{70^2}{71^2}t^2\right).$$

Note that if $\sum_{k=1}^{K} \sigma^2_{u\times w_k} = 0$, then $\sum_{k=1}^{K} \Delta^2_{u\times w_k}$ is a constant zero and the above inequality holds obviously. Below we consider only the case where $\sum_{k=1}^{K} \sigma^2_{u\times w_k} > 0$. By definition, for any $t^2 \geq 3 \cdot 71^2/70^2$,

$$\begin{aligned}
\frac{70^2}{71^2}\frac{Nf^2t^2\xi}{\sum_{k=1}^{K}\sigma^2_{u\times w_k}} - \log K + \log p &\geq \frac{70^2}{71^2}t^2 \max\{1, \log K, -\log p\} - \log K + \log p \\
&\geq \frac{70^2}{71^2}t^2 \frac{1 + \log K - \log p}{3} - \log K + \log p \\
&\geq \frac{1}{3}\frac{70^2}{71^2}t^2.
\end{aligned}$$

Thus, from Lemma A30, for any $t^2 \geq 3 \cdot 71^2/70^2$,

$$\begin{aligned}
\frac{1}{p}\mathbb{P}\left(\sum_{k=1}^{K}\Delta^2_{u\times w_k} > t^2\xi\right) &\leq 2\frac{K}{p}\exp\left(-\frac{70^2}{71^2}\frac{Nf^2t^2\xi}{\sum_{k=1}^{K}\sigma^2_{u\times w_k}}\right) = 2\exp\left(-\frac{70^2}{71^2}\frac{Nf^2t^2\xi}{\sum_{k=1}^{K}\sigma^2_{u\times w_k}} + \log K - \log p\right) \\
&\leq 2\exp\left(-\frac{1}{3}\frac{70^2}{71^2}t^2\right).
\end{aligned}$$

Second, we prove that, for any $t \geq 3 \cdot 71^2/70^2$,

$$\text{(A7.23)} \qquad \frac{1}{p}\mathbb{P}\left(\Delta^2_u \sum_{k=1}^{K}\Delta^2_{w_k} > t^2\xi\right) \leq 4\exp\left(-\frac{1}{3}\frac{70^2}{71^2}t\right).$$

Note that if $\sigma^2_u = 0$ or $\sum_{k=1}^{K}\sigma^2_{w_k} = 0$, then $\Delta^2_u\sum_{k=1}^{K}\Delta^2_{w_k}$ is a constant zero and the above inequality holds obviously. Below we consider only the case where both $\sigma^2_u$ and $\sum_{k=1}^{K}\sigma^2_{w_k}$ are positive. By definition, for any $t > 0$,

$$t^2\xi \geq t\frac{\max\{1, -\log p\}}{Nf^2}\sigma^2_u \cdot t\frac{\max\{1, \log K, -\log p\}}{Nf^2}\sum_{k=1}^{K}\sigma^2_{w_k}.$$

From Lemma A30, this implies that, for any $t > 0$,

(A7.24)

$$\begin{aligned}
&\frac{1}{p}\mathbb{P}\left(\Delta^2_u\sum_{k=1}^{K}\Delta^2_{w_k} > t^2\xi\right) \\
&\leq \frac{1}{p}\mathbb{P}\left(\Delta^2_u > t\frac{\max\{1, -\log p\}}{Nf^2}\sigma^2_u\right) + \frac{1}{p}\mathbb{P}\left(\sum_{k=1}^{K}\Delta^2_{w_k} > t\frac{\max\{1, \log K, -\log p\}}{Nf^2}\sum_{k=1}^{K}\sigma^2_{w_k}\right) \\
&\leq \frac{2}{p}\exp\left(-\frac{70^2}{71^2}t\max\{1, -\log p\}\right) + \frac{2K}{p}\exp\left(-\frac{70^2}{71^2}t\max\{1, \log K, -\log p\}\right) \\
&= 2\exp\left(-\frac{70^2}{71^2}t\max\{1, -\log p\} - \log p\right) + 2\exp\left(-\frac{70^2}{71^2}t\max\{1, \log K, -\log p\} + \log K - \log p\right).
\end{aligned}$$

Note that when $t \geq 2 \cdot 71^2/70^2$,

$$\frac{70^2}{71^2} t \max\{1, -\log p\} + \log p \geq \frac{70^2}{71^2} t \frac{1 - \log p}{2} + \log p \geq \frac{1}{2} \frac{70^2}{71^2} t,$$

and when $t \geq 3 \cdot 71^2/70^2$,

(A7.25)

$$\begin{aligned}\frac{70^2}{71^2} t \max\{1, \log K, -\log p\} - \log K + \log p &\geq \frac{70^2}{71^2} t \frac{1 + \log K - \log p}{3} - \log K + \log p \\ &\geq \frac{1}{3} \frac{70^2}{71^2} t.\end{aligned}$$

Thus, when $t \geq 3 \cdot 71^2/70^2$, we have

$$\frac{1}{p} \mathbb{P}\left(\Delta_u^2 \sum_{k=1}^K \Delta_{w_k}^2 > t^2 \xi\right) \leq 2 \exp\left(-\frac{1}{2} \frac{70^2}{71^2} t\right) + 2 \exp\left(-\frac{1}{3} \frac{70^2}{71^2} t\right) \leq 4 \exp\left(-\frac{1}{3} \frac{70^2}{71^2} t\right).$$

Third, by definition, when $t \geq 1$, $t^2 \xi \geq \xi \geq (1-f)^2/(N^2 f^2) \cdot \sum_{k=1}^K S_{uw_k}^2$. This immediately implies that, when $t \geq 1$,

$$\frac{1}{p} \mathbb{P}\left(\frac{(1-f)^2}{N^2 f^2} \sum_{k=1}^K S_{uw_k}^2 > t^2 \xi\right) = 0. \tag{A7.26}$$

From (A7.21)–(A7.26), we can know that, when $t \geq 3 \cdot 71^2/70^2$,

$$\begin{aligned}&\mathbb{P}\left(\|\boldsymbol{s}_{u\boldsymbol{w}} - \boldsymbol{S}_{u\boldsymbol{w}}\|_2^2 > 36 t^2 \xi \mid \boldsymbol{Z} \in \mathcal{E}\right) \\ &\leq \frac{1}{p} \mathbb{P}\left(\sum_{k=1}^K \Delta_{u \times w_k}^2 > t^2 \xi\right) + \frac{1}{p} \mathbb{P}\left(\Delta_u^2 \sum_{k=1}^K \Delta_{w_k}^2 > t^2 \xi\right) + \frac{1}{p} \mathbb{P}\left(\frac{(1-f)^2}{N^2 f^2} \sum_{k=1}^K S_{uw_k}^2 > t^2 \xi\right) \\ &\leq 2 \exp\left(-\frac{1}{3} \frac{70^2}{71^2} t^2\right) + 4 \exp\left(-\frac{1}{3} \frac{70^2}{71^2} t\right) \leq 6 \exp\left(-\frac{1}{3} \frac{70^2}{71^2} t\right).\end{aligned}$$

Therefore, Lemma A31 holds. □

PROOF OF LEMMA A32. By definition, Lemma A31 immediately implies that, under ReM,

$$\left|s_z^2 - S_z^2\right| = O_{\mathbb{P}}\left(\xi_{zz}^{1/2}\right), \quad \|s_{z\boldsymbol{w}} - S_{z\boldsymbol{w}}\|_2 = O_{\mathbb{P}}\left(\xi_{z\boldsymbol{w}}^{1/2}\right).$$

This implies that, for $z = 0, 1$,

$$\begin{aligned}\left|\|s_{z\boldsymbol{w}}\|_2^2 - \|S_{z\boldsymbol{w}}\|_2^2\right| &= \left|(s_{z\boldsymbol{w}} - S_{z\boldsymbol{w}})(s_{z\boldsymbol{w}} - S_{z\boldsymbol{w}} + 2S_{z\boldsymbol{w}})^\top\right| = \left|\|s_{z\boldsymbol{w}} - S_{z\boldsymbol{w}}\|_2^2 + 2(s_{z\boldsymbol{w}} - S_{z\boldsymbol{w}}) S_{z\boldsymbol{w}}^\top\right| \\ &\leq \|s_{z\boldsymbol{w}} - S_{z\boldsymbol{w}}\|_2^2 + 2\|s_{z\boldsymbol{w}} - S_{z\boldsymbol{w}}\|_2 \|S_{z\boldsymbol{w}}\|_2 = O_{\mathbb{P}}\left(\xi_{z\boldsymbol{w}} + \|S_{z\boldsymbol{w}}\|_2 \xi_{z\boldsymbol{w}}^{1/2}\right).\end{aligned}$$

By the same logic,

$$\begin{aligned}\left|s_{\tau|\boldsymbol{X}}^2 - S_{\tau|\boldsymbol{X}}^2\right| &= \left|\|(s_{1\boldsymbol{w}} - s_{0\boldsymbol{w}}) - (S_{1\boldsymbol{w}} - S_{0\boldsymbol{w}})\|_2^2 + 2\{(s_{1\boldsymbol{w}} - s_{0\boldsymbol{w}}) - (S_{1\boldsymbol{w}} - S_{0\boldsymbol{w}})\}(S_{1\boldsymbol{w}} - S_{0\boldsymbol{w}})^\top\right| \\ &\leq 2\left(\|s_{1\boldsymbol{w}} - S_{1\boldsymbol{w}}\|_2^2 + \|s_{0\boldsymbol{w}} - S_{0\boldsymbol{w}}\|_2^2\right) + 2\|S_{1\boldsymbol{w}} - S_{0\boldsymbol{w}}\|_2 \{\|s_{1\boldsymbol{w}} - S_{1\boldsymbol{w}}\|_2 + \|s_{0\boldsymbol{w}} - S_{0\boldsymbol{w}}\|_2\} \\ &= O_{\mathbb{P}}\left(\xi_{1\boldsymbol{w}} + \xi_{0\boldsymbol{w}} + \|S_{1\boldsymbol{w}} - S_{0\boldsymbol{w}}\|_2 \xi_{1\boldsymbol{w}}^{1/2} + \|S_{1\boldsymbol{w}} - S_{0\boldsymbol{w}}\|_2 \xi_{0\boldsymbol{w}}^{1/2}\right).\end{aligned}$$

From the above and by definition, we then have

$$\begin{aligned}\left|\hat{V}_{\tau\tau} - V_{\tau\tau} - n^{-1}S^2_{\tau\setminus \boldsymbol{X}}\right| &\le n_1^{-1}\left|s_1^2 - S_1^2\right| + n_0^{-1}\left|s_0^2 - S_0^2\right| + n^{-1}\left|s^2_{\tau\mid\boldsymbol{X}} - S^2_{\tau\mid\boldsymbol{X}}\right| \\ &= O_{\mathbb{P}}\left(\frac{\xi_{11}^{1/2}}{n_1} + \frac{\xi_{00}^{1/2}}{n_0} + \frac{\xi_{1\boldsymbol{w}} + \xi_{0\boldsymbol{w}}}{n} + \|S_{1\boldsymbol{w}} - S_{0\boldsymbol{w}}\|_2\frac{\xi_{1\boldsymbol{w}}^{1/2} + \xi_{0\boldsymbol{w}}^{1/2}}{n}\right),\end{aligned}$$

and

$$\begin{aligned}&\left|\hat{V}_{\tau\tau}\hat{R}_n^2 - V_{\tau\tau}R_n^2\right| \\ &= \left|n_1^{-1}\|s_{1\boldsymbol{w}}\|_2^2 + n_0^{-1}\|s_{0\boldsymbol{w}}\|_2^2 - n^{-1}s^2_{\tau\mid\boldsymbol{X}} - \left(n_1^{-1}\|S_{1\boldsymbol{w}}\|_2^2 + n_0^{-1}\|S_{0\boldsymbol{w}}\|_2^2 - n^{-1}S^2_{\tau\mid\boldsymbol{X}}\right)\right| \\ &\le n_1^{-1}\left|\|s_{1\boldsymbol{w}}\|_2^2 - \|S_{1\boldsymbol{w}}\|_2^2\right| + n_0^{-1}\left|\|s_{0\boldsymbol{w}}\|_2^2 - \|S_{0\boldsymbol{w}}\|_2^2\right| + n^{-1}\left|s^2_{\tau\mid\boldsymbol{X}} - S^2_{\tau\mid\boldsymbol{X}}\right| \\ &= O_{\mathbb{P}}\left(\frac{\xi_{1\boldsymbol{w}}}{n_1} + \|S_{1\boldsymbol{w}}\|_2\frac{\xi_{1\boldsymbol{w}}^{1/2}}{n_1} + \frac{\xi_{0\boldsymbol{w}}}{n_1} + \|S_{0\boldsymbol{w}}\|_2\frac{\xi_{0\boldsymbol{w}}^{1/2}}{n_0} + \frac{\xi_{1\boldsymbol{w}} + \xi_{0\boldsymbol{w}}}{n} + \|S_{1\boldsymbol{w}} - S_{0\boldsymbol{w}}\|_2\frac{\xi_{1\boldsymbol{w}}^{1/2} + \xi_{0\boldsymbol{w}}^{1/2}}{n}\right) \\ &= O_{\mathbb{P}}\left(\frac{\xi_{1\boldsymbol{w}}}{n_1} + \frac{\xi_{0\boldsymbol{w}}}{n_0} + \|S_{1\boldsymbol{w}}\|_2\frac{\xi_{1\boldsymbol{w}}^{1/2}}{n_1} + \|S_{0\boldsymbol{w}}\|_2\frac{\xi_{0\boldsymbol{w}}^{1/2}}{n_0} + \|S_{1\boldsymbol{w}} - S_{0\boldsymbol{w}}\|_2\frac{\xi_{1\boldsymbol{w}}^{1/2} + \xi_{0\boldsymbol{w}}^{1/2}}{n}\right).\end{aligned}$$

Therefore, Lemma A32 holds. □

PROOF OF LEMMA A33. First, we consider bounding some finite population quantities. For descriptive convenience, we introduce $\psi = \max_{z\in\{0,1\}}\max_{1\le i\le n}\{Y_i(z) - \bar{Y}(z)\}^2$. By definition, for $z = 0, 1$ and $1 \le k \le K$,

(A7.27)
$$\sigma_z^2 = \frac{1}{n}\sum_{i=1}^n\{Y_i(z) - \bar{Y}(z)\}^2 \le \psi, \quad S_z^2 = \frac{n}{n-1}\sigma_z^2 \le 2\psi, \quad \sigma^2_{w_k} = \frac{1}{n}\sum_{i=1}^n(w_{ki} - \bar{w}_k)^2 = \frac{n-1}{n} \le 1.$$

and

$$\begin{aligned}\sigma^2_{z\times z} &= \frac{1}{n}\sum_{i=1}^n\left[\{Y_i(z) - \bar{Y}(z)\}^2 - \sigma_z^2\right]^2 \le \frac{1}{n}\sum_{i=1}^n\{Y_i(z) - \bar{Y}(z)\}^4 \le \psi^2, \\ \sigma^2_{z\times w_k} &= \frac{1}{n}\sum_{i=1}^n\left[\{Y_i(z) - \bar{Y}(z)\}(w_{ki} - \bar{w}_k) - \frac{n-1}{n}S_{zw_k}\right]^2 \le \frac{1}{n}\sum_{i=1}^n\{Y_i(z) - \bar{Y}(z)\}^2(w_{ki} - \bar{w}_k)^2 \\ &\le \psi\cdot\frac{1}{n}\sum_{i=1}^n(w_{ki} - \bar{w}_k)^2 \le \psi.\end{aligned}$$

Furthermore, by the Cauchy–Schwarz inequality,

$$\begin{aligned}S^2_{zw_k} &= \left[\frac{1}{n-1}\sum_{i=1}^n\{Y_i(z) - \bar{Y}(z)\}(w_{ki} - \bar{w}_k)\right]^2 \le \frac{1}{(n-1)^2}\sum_{i=1}^n\{Y_i(z) - \bar{Y}(z)\}^2\cdot\sum_{i=1}^n(w_{ki} - \bar{w}_k)^2 \\ &= \frac{1}{n-1}\sum_{i=1}^n\{Y_i(z) - \bar{Y}(z)\}^2 \le 2\psi.\end{aligned}$$

Second, we consider the bounds on $\xi_{zz}$ and $\xi_{z\boldsymbol{w}}$ for $z = 0, 1$. For descriptive convenience, we introduce $b_n = \max\{1, -\log\tilde{p}_n\}$ and $c_n = \max\{1, \log K_n, -\log\tilde{p}_n\}$. By definition and from the bounds we derived above, for $z=0,1$,

$$\xi_{zz}=\frac{b_n}{nr_z^2}\sigma^2_{z\times z}+\frac{b_n^2}{n^2r_z^4}\sigma_z^4+\frac{(1-r_z)^2}{n^2r_z^2}S_z^4\le\psi^2\left(\frac{b_n}{nr_z^2}+\frac{b_n^2}{n^2r_z^4}+\frac{4}{n^2r_z^2}\right)\le\psi^2\left(\frac{2b_n}{nr_z^2}+\frac{b_n^2}{n^2r_z^4}\right)$$

where the last inequality holds because $b_n\ge 1$ and $n\ge 4$. From the condition in Lemma A33, $b_n\le c_n=O(nr_1^2r_0^2)$, and thus

$$\xi_{zz}=\psi^2\frac{b_n}{nr_z^2}\left(2+\frac{b_n}{nr_z^2}\right)=O\left(\psi^2\frac{b_n}{nr_z^2}\right).$$

Similarly, we can derive that, for $z=0,1$,

(A7.28)

$$\begin{aligned}\xi_{z\boldsymbol{w}}&=\frac{c_n}{nr_z^2}\sum_{k=1}^{K_n}\sigma^2_{z\times w_k}+\frac{b_nc_n}{n^2r_z^4}\sigma_z^2\sum_{k=1}^{K_n}\sigma^2_{w_k}+\frac{(1-r_z)^2}{n^2r_z^2}\sum_{k=1}^{K_n}S^2_{zw_k}\le\psi K_n\left(\frac{c_n}{nr_z^2}+\frac{b_nc_n}{n^2r_z^4}+\frac{2}{n^2r_z^2}\right)\\&\le\psi K_n\left(2\frac{c_n}{nr_z^2}+\frac{b_nc_n}{n^2r_z^4}\right)=\psi K_n\frac{c_n}{nr_z^2}\left(2+\frac{b_n}{nr_z^2}\right)=O\left(\psi K_n\frac{c_n}{nr_z^2}\right).\end{aligned}$$

Third, we consider the probability bounds for $\hat{V}_{\tau\tau}-V_{\tau\tau}-n^{-1}S^2_{\tau\setminus\boldsymbol{X}}$ and $\hat{V}_{\tau\tau}\hat{R}_n^2-V_{\tau\tau}R_n^2$. From the bounds we derived before, for $z=0,1$,

$$\|S_{z\boldsymbol{w}}\|_2\le\left(\sum_{k=1}^{K_n}S^2_{z\boldsymbol{w}_k}\right)^{1/2}\le\sqrt{2K_n\psi},\quad \|S_{1\boldsymbol{w}}-S_{0\boldsymbol{w}}\|_2\le\|S_{1\boldsymbol{w}}\|_2+\|S_{0\boldsymbol{w}}\|_2\le 2\sqrt{2K_n\psi}.$$

Consequently, we have

$$\begin{aligned}&\frac{\xi_{11}^{1/2}}{n_1}+\frac{\xi_{00}^{1/2}}{n_0}+\frac{\xi_{1\boldsymbol{w}}+\xi_{0\boldsymbol{w}}}{n}+\|S_{1\boldsymbol{w}}-S_{0\boldsymbol{w}}\|_2\frac{\xi_{1\boldsymbol{w}}^{1/2}+\xi_{0\boldsymbol{w}}^{1/2}}{n}\\&=\psi\cdot O\left(\frac{\sqrt{b_n}}{n^{3/2}r_1^2}+\frac{\sqrt{b_n}}{n^{3/2}r_0^2}+K_n\frac{c_n}{n^2r_1^2}+K_n\frac{c_n}{n^2r_0^2}+K_n\frac{\sqrt{c_n}}{n^{3/2}r_1}+K_n\frac{\sqrt{c_n}}{n^{3/2}r_0}\right)\\&=\psi\cdot O\left(\frac{\sqrt{b_n}}{n^{3/2}r_1^2r_0^2}+K_n\frac{c_n}{n^2r_1^2r_0^2}+K_n\frac{\sqrt{c_n}}{n^{3/2}r_1r_0}\right)=\psi\cdot O\left\{\frac{\sqrt{b_n}}{n^{3/2}r_1^2r_0^2}+K_n\frac{\sqrt{c_n}}{n^{3/2}r_1r_0}\left(1+\sqrt{\frac{c_n}{nr_1^2r_0^2}}\right)\right\}\\&=\psi\cdot O\left(\frac{\sqrt{b_n}}{n^{3/2}r_1^2r_0^2}+K_n\frac{\sqrt{c_n}}{n^{3/2}r_1r_0}\right)=\psi\cdot O\left(\frac{\sqrt{b_n}+K_n\sqrt{c_n}}{n^{3/2}r_1^2r_0^2}\right).\end{aligned}$$

where the second last equality holds because $c_n=O(nr_1^2r_0^2)$. Similarly, we can derive that

$$\begin{aligned}&\frac{\xi_{1\boldsymbol{w}}}{n_1}+\frac{\xi_{0\boldsymbol{w}}}{n_0}+\|S_{1\boldsymbol{w}}\|_2\frac{\xi_{1\boldsymbol{w}}^{1/2}}{n_1}+\|S_{0\boldsymbol{w}}\|_2\frac{\xi_{0\boldsymbol{w}}^{1/2}}{n_0}+\|S_{1\boldsymbol{w}}-S_{0\boldsymbol{w}}\|_2\frac{\xi_{1\boldsymbol{w}}^{1/2}+\xi_{0\boldsymbol{w}}^{1/2}}{n}\\&=\psi\cdot O\left(K_n\frac{c_n}{n^2r_1^3}+K_n\frac{c_n}{n^2r_0^3}+K_n\frac{\sqrt{c_n}}{n^{3/2}r_1^2}+K_n\frac{\sqrt{c_n}}{n^{3/2}r_0^2}+K_n\frac{\sqrt{c_n}}{n^{3/2}r_1}+K_n\frac{\sqrt{c_n}}{n^{3/2}r_0}\right)\\&=\psi\cdot O\left(K_n\frac{c_n}{n^2r_1^3r_0^3}+K_n\frac{\sqrt{c_n}}{n^{3/2}r_1^2r_0^2}\right)=\psi\cdot O\left\{K_n\frac{\sqrt{c_n}}{n^{3/2}r_1^2r_0^2}\left(1+\sqrt{\frac{c_n}{nr_1^2r_0^2}}\right)\right\}\\&=\psi\cdot O\left(K_n\frac{\sqrt{c_n}}{n^{3/2}r_1^2r_0^2}\right).\end{aligned}$$

Note that, by definition, $b_n\le c_n$. Thus, we must have

$$\frac{\sqrt{b_n}+K_n\sqrt{c_n}}{n^{3/2}r_1^2r_0^2}\le 2\max\{K_n,1\}\cdot\frac{\sqrt{c_n}}{n^{3/2}r_1^2r_0^2},\qquad K_n\frac{\sqrt{c_n}}{n^{3/2}r_1^2r_0^2}\le\max\{K_n,1\}\cdot\frac{\sqrt{c_n}}{n^{3/2}r_1^2r_0^2}.$$

From the above and Lemma A32, these then imply that

$$\max\left\{\left|\hat{V}_{\tau\tau} - V_{\tau\tau} - n^{-1}S^2_{\tau\setminus \boldsymbol{X}}\right|,\ \ \left|\hat{V}_{\tau\tau}\hat{R}^2_n - V_{\tau\tau}R^2_n\right|\right\} = \psi \cdot O_{\mathbb{P}}\left(\max\{K_n, 1\}\cdot \frac{\sqrt{c_n}}{n^{3/2}r_1^2r_0^2}\right).$$

Therefore, Lemma A33 holds. □

PROOF OF LEMMA A34. We first prove (i). By definition, $|\tilde{p}_n - p_n| \le \Delta_n$. Because $\Delta_n/p_n = o(1)$, this implies that

$$-\log \tilde{p}_n \le -\log(p_n - \Delta_n) = -\log p_n - \log(1 - \Delta_n/p_n) = -\log p_n + O(1) = O(\max\{1, -\log p_n\}).$$

Thus, $\max\{1, -\log \tilde{p}_n\} = O(\max\{1, -\log p_n\})$.

We then prove (ii). By definition,

$$\begin{aligned} r_0S^2_{1\setminus \boldsymbol{X}} + r_1S^2_{0\setminus \boldsymbol{X}} &\le r_0S^2_1 + r_1S^2_0 \le 2(r_0+r_1)\max_{z\in\{0,1\}}\max_{1\le i\le n}\{Y_i(z) - \bar{Y}(z)\}^2 \\ &= 2\max_{z\in\{0,1\}}\max_{1\le i\le n}\{Y_i(z) - \bar{Y}(z)\}^2. \end{aligned}$$

Thus, (ii) holds.

Last, we prove (iii). From (i) and Condition 5, we can verify that

$$\begin{aligned} &\frac{\max_{z\in\{0,1\}}\max_{1\le i\le n}\{Y_i(z) - \bar{Y}(z)\}^2}{r_0S^2_{1\setminus \boldsymbol{X}} + r_1S^2_{0\setminus \boldsymbol{X}}}\cdot\frac{\max\{K_n,1\}}{r_1r_0}\cdot\sqrt{\frac{\max\{1, \log K_n, -\log\tilde{p}_n\}}{n}} \\ &= O\left(\frac{\max_{z\in\{0,1\}}\max_{1\le i\le n}\{Y_i(z) - \bar{Y}(z)\}^2}{r_0S^2_{1\setminus \boldsymbol{X}} + r_1S^2_{0\setminus \boldsymbol{X}}}\cdot\frac{\max\{K_n,1\}}{r_1r_0}\cdot\sqrt{\frac{\max\{1, \log K_n, -\log p_n\}}{n}}\right) \\ &= o(1). \end{aligned}$$

From (ii), this then implies that

$$\begin{aligned} o(1) &= \frac{\max_{z\in\{0,1\}}\max_{1\le i\le n}\{Y_i(z) - \bar{Y}(z)\}^2}{nr_1r_0V_{\tau\tau}(1-R^2)}\cdot\frac{\max\{K_n,1\}}{r_1r_0}\cdot\sqrt{\frac{\max\{1, \log K_n, -\log\tilde{p}_n\}}{n}} \\ &\ge \frac{1}{2}\cdot\frac{\max\{K_n,1\}}{r_1r_0}\cdot\sqrt{\frac{\max\{1, \log K_n, -\log\tilde{p}_n\}}{n}} \ge \frac{1}{2}\sqrt{\frac{\max\{1, \log K_n, -\log\tilde{p}_n\}}{nr_1^2r_0^2}}. \end{aligned}$$

Consequently, we must have $\max\{1, \log K_n, -\log\tilde{p}_n\} = o(nr_1^2r_0^2)$, i.e., (iii) holds. □

A7.3. *Proofs of Theorems 7 and 8.*

**Proof of Theorem 7(i).** From Lemmas A33 and A34, under ReM and Conditions 2 and 5, we must have

$$\begin{aligned} &\max\left\{\left|\hat{V}_{\tau\tau} - V_{\tau\tau} - n^{-1}S^2_{\tau\setminus \boldsymbol{X}}\right|,\ \ \left|\hat{V}_{\tau\tau}\hat{R}^2_n - V_{\tau\tau}R^2_n\right|\right\} \\ &= \max_{z\in\{0,1\}}\max_{1\le i\le n}\{Y_i(z) - \bar{Y}(z)\}^2\cdot O_{\mathbb{P}}\left(\max\{K_n,1\}\cdot\frac{\sqrt{\max\{1, \log K_n, -\log p_n\}}}{n^{3/2}r_1^2r_0^2}\right) \\ &= \frac{\max_{z\in\{0,1\}}\max_{1\le i\le n}\{Y_i(z) - \bar{Y}(z)\}^2}{nr_1r_0}\cdot O_{\mathbb{P}}\left(\frac{\max\{K_n,1\}}{r_1r_0}\cdot\sqrt{\frac{\max\{1, \log K_n, -\log p_n\}}{n}}\right) \\ &= \frac{r_0S^2_{1\setminus \boldsymbol{X}} + r_1S^2_{0\setminus \boldsymbol{X}}}{nr_1r_0}\cdot o_{\mathbb{P}}(1) = \left(n_1^{-1}S^2_{1\setminus \boldsymbol{X}} + n_0^{-1}S^2_{0\setminus \boldsymbol{X}}\right)\cdot o_{\mathbb{P}}(1). \end{aligned}$$

Note that, by definition,

$$V_{\tau\tau}(1-R_n^2)+n^{-1}S^2_{\tau\setminus \boldsymbol{X}} = \left(n_1^{-1}S_1^2+n_0^{-1}S_0^2-n^{-1}S_\tau^2\right)-\left(n_1^{-1}S^2_{1|\boldsymbol{X}}+n_0^{-1}S^2_{0|\boldsymbol{X}}-n^{-1}S^2_{\tau|\boldsymbol{X}}\right)+n^{-1}S^2_{\tau\setminus \boldsymbol{X}}$$

(A7.29)

$$= n_1^{-1}S^2_{1\setminus \boldsymbol{X}}+n_0^{-1}S^2_{0\setminus \boldsymbol{X}}.$$

From the above, Theorem 7(i) holds. □

To prove Theorem 7(ii), we need the following two lemmas.

LEMMA A35. *Let $\{\psi_n\}$ and $\{\tilde\psi_n\}$ be two sequences of continuous random variables such that, as $n\to\infty$, $\sup_{c\in\mathbb{R}}\left|\mathbb{P}(\psi_n\le c)-\mathbb{P}(\tilde\psi_n\le c)\right|\to 0$. For any $n$ and $\alpha\in(0,1)$, let $q_n(\alpha)$ and $\tilde q_n(\alpha)$ be the $\alpha$th quantile of $\psi_n$ and $\tilde\psi_n$, respectively. Then for any $0<\alpha<\beta<1$, $\mathbb{1}\{\tilde q_n(\beta)\le q_n(\alpha)\}\to 0$ as $n\to\infty$.*

PROOF OF LEMMA A35. From the condition and definition in Lemma A35, $\mathbb{P}\{\psi_n\le q_n(\alpha)\}=\alpha$, $\mathbb{P}\{\tilde\psi_n\le\tilde q_n(\beta)\}=\beta$, and $\left|\mathbb{P}\{\psi_n\le\tilde q_n(\beta)\}-\mathbb{P}\{\tilde\psi_n\le\tilde q_n(\beta)\}\right|\le\sup_{c\in\mathbb{R}}\left|\mathbb{P}(\psi_n\le c)-\mathbb{P}(\tilde\psi_n\le c)\right|\to 0$ as $n\to\infty$. These imply that $\mathbb{P}\{\psi_n\le\tilde q_n(\beta)\}\to\beta>\alpha$ as $n\to\infty$. Below we prove Lemma A35 by contradiction.

Suppose that $\mathbb{1}\{\tilde q_n(\beta)\le q_n(\alpha)\}$ does not converge to zero as $n\to\infty$. Then there exists a subsequence $\{n_j\}$ such that $\tilde q_{n_j}(\beta)\le q_{n_j}(\alpha)$ for all $j$. This implies that, for all $j$, $\mathbb{P}\{\psi_{n_j}\le\tilde q_{n_j}(\beta)\}\le\mathbb{P}\{\psi_{n_j}\le q_{n_j}(\alpha)\}=\alpha$. Consequently, we must have $\limsup_{j\to\infty}\mathbb{P}\{\psi_{n_j}\le\tilde q_{n_j}(\beta)\}\le\alpha$. However, this contradicts with the fact that $\lim_{n\to\infty}\mathbb{P}\{\psi_n\le\tilde q_n(\beta)\}=\beta$. From the above, Lemma A35 holds. □

LEMMA A36. *Let $\varepsilon_0\sim\mathcal{N}(0,1)$, and $L_{K_n,a_n}$ be the truncated Gaussian random variables defined as in Section 2.3, where $\{K_n\}$ and $\{a_n\}$ are sequences of positive integers and thresholds, and $\varepsilon_0$ is independent of $L_{K_n,a_n}$ for all $n$. Let $\{A_n\}$, $\{B_n\}$, $\{\tilde A_n\}$ and $\{\tilde B_n\}$ be sequences of nonnegative constants, and for each $n$, define $\psi_n=A_n^{1/2}\cdot\varepsilon_0+B_n^{1/2}\cdot L_{K_n,a_n}$ and $\tilde\psi_n=\tilde A_n^{1/2}\cdot\varepsilon_0+\tilde B_n^{1/2}\cdot L_{K_n,a_n}$. For each $n$ and $\alpha\in(0,1)$, let $q_n(\alpha)$ and $\tilde q_n(\alpha)$ be the $\alpha$th quantile of $\psi_n$ and $\tilde\psi_n$, respectively. If $\max\{|\tilde A_n-A_n|,|\tilde B_n-B_n|\}=o(A_n)$, then for any $0<\alpha<\beta<1$, as $n\to\infty$, $\mathbb{1}\{\tilde q_n(\beta)\le q_n(\alpha)\}\to 0$ and $\mathbb{1}\{q_n(\beta)\le\tilde q_n(\alpha)\}\to 0$.*

PROOF OF LEMMA A36. Because $\operatorname{Var}(L_{K_n,a_n})\le 1$, $L_{K_n,a_n}=O_{\mathbb{P}}(1)$. Using the inequality that $|\sqrt{b}-\sqrt{c}|\le\sqrt{|b-c|}$ for any $b,c\ge 0$, we have

$$\begin{aligned}\tilde\psi_n-\psi_n &= \left(\tilde A_n^{1/2}-A_n^{1/2}\right)\varepsilon_0+\left(\tilde B_n^{1/2}-B_n^{1/2}\right)L_{K_n,a_n} = |\tilde A_n-A_n|^{1/2}\cdot O_{\mathbb{P}}(1)+|\tilde B_n-B_n|^{1/2}\cdot O_{\mathbb{P}}(1)\\ &= A_n^{1/2}\cdot o_{\mathbb{P}}(1).\end{aligned}$$

From Lemma A27, this then implies that $\sup_{c\in\mathbb{R}}|\mathbb{P}(\tilde\psi_n\le c)-\mathbb{P}(\psi_n\le c)|\to 0$ as $n\to\infty$. From Lemma A35, this further implies that, for any $0<\alpha<\beta<1$, $\mathbb{1}\{\tilde q_n(\beta)\le q_n(\alpha)\}\to 0$ as $n\to\infty$. By summery, we also have $\mathbb{1}\{q_n(\beta)\le\tilde q_n(\alpha)\}\to 0$ as $n\to\infty$. Therefore, Lemma A36 holds. □

**Proof of Theorem 7(ii).** For descriptive convenience, let $\varepsilon_0$ and $L_{K_n,a_n}$ be two independent standard and constrained Gaussian random variables defined as in Section 2.3, which are further constructed to be independent of the treatment assignment $\boldsymbol{Z}$. We then define

$$\theta_n=\sqrt{V_{\tau\tau}(1-R_n^2)}\cdot\varepsilon_0+\sqrt{V_{\tau\tau}R_n^2}\cdot L_{K_n,a_n}\equiv A_n^{1/2}\cdot\varepsilon_0+B_n^{1/2}\cdot L_{K_n,a_n},$$

$$\tilde{\theta}_n = \sqrt{V_{\tau\tau}(1-R_n^2) + n^{-1}S^2_{\tau\setminus \boldsymbol{X}}} \cdot \varepsilon_0 + \sqrt{V_{\tau\tau}R_n^2} \cdot L_{K_n,a_n} \equiv \tilde{A}_n^{1/2} \cdot \varepsilon_0 + \tilde{B}_n^{1/2} \cdot L_{K_n,a_n},$$

$$\hat{\theta}_n = \sqrt{\hat{V}_{\tau\tau}(1-\hat{R}_n^2)} \cdot \varepsilon_0 + \sqrt{\hat{V}_{\tau\tau}\hat{R}_n^2} \cdot L_{K_n,a_n} \equiv \hat{A}_n^{1/2} \cdot \varepsilon_0 + \hat{B}_n^{1/2} \cdot L_{K_n,a_n},$$

where $A_n, \tilde{A}_n, \hat{A}_n$ and $B_n, \tilde{B}_n, \hat{B}_n$ denote the squared coefficients of the standard and constrained Gaussian random variables, respectively. We introduce $q_\alpha(A, B, K, a)$ to denote the $\alpha$th quantile of $A^{1/2}\varepsilon_0 + B^{1/2}L_{K,a}$, and further define $\hat{q}_{n,\alpha} = q_\alpha(\hat{A}_n, \hat{B}_n, K_n, a_n)$, $\tilde{q}_{n,\alpha} = q_\alpha(\tilde{A}_n, \tilde{B}_n, K_n, a_n)$ and $q_{n,\alpha} = q_\alpha(A_n, B_n, K_n, a_n)$.

First, we prove that $\lim_{n\to\infty} \mathbb{P}(\hat{q}_{n,\beta} \le \tilde{q}_{n,\alpha} \mid M \le a_n) = 0$ for any $0 < \alpha < \beta < 1$. From Theorem 7(i), under ReM, $\max\{|\hat{A}_n - \tilde{A}_n|, |\hat{B}_n - \tilde{B}_n|\} = o_{\mathbb{P}}(\tilde{A}_n)$. By Durrett (2019, Theorem 2.3.2), under ReM, for any subsequence $\{n_j : j = 1, 2, \ldots\}$, there exists a further subsequence $\{m_j : j = 1, 2, \ldots\} \subset \{n_j : j = 1, 2, \ldots\}$ such that $|\hat{A}_{m_j} - \tilde{A}_{m_j}|/\tilde{A}_{m_j} \xrightarrow{a.s.} 0$ and $|\hat{B}_{m_j} - \tilde{B}_{m_j}|/\tilde{A}_{m_j} \xrightarrow{a.s.} 0$ as $j \to \infty$. From Lemma A36, this immediately implies that, for any $0 < \alpha < \beta < 1$, $\mathbb{1}\{\hat{q}_{m_j,\beta} \le \tilde{q}_{m_j,\alpha}\} \xrightarrow{a.s.} 0$ as $n \to 0$. From Durrett (2019, Theorem 2.3.2), we can know that, under ReM, for any $0 < \alpha < \beta < 1$, $\mathbb{1}(\hat{q}_{n,\beta} \le \tilde{q}_{n,\alpha}) \xrightarrow{\mathbb{P}} 0$ as $n \to 0$. Consequently, under ReM, for any $0 < \alpha < \beta < 1$, as $n \to \infty$,

$$\mathbb{P}(\hat{q}_{n,\beta} \le \tilde{q}_{n,\alpha} \mid M \le a_n) = \mathbb{E}\{\mathbb{1}(\hat{q}_{n,\beta} \le \tilde{q}_{n,\alpha}) \mid M \le a_n\} \to 0. \tag{A7.30}$$

Second, we prove the asymptotic validity of the confidence interval $\hat{\mathcal{C}}_\alpha$ for $\alpha \in (0,1)$. For any $\alpha \in (0, 1)$ and $\eta \in (0, (1-\alpha)/2)$, the coverage probability of the confidence interval $\hat{\mathcal{C}}_\alpha$ can be bounded by

$$\begin{aligned}
\mathbb{P}(\tau \in \hat{\mathcal{C}}_\alpha \mid M \le a_n) &= \mathbb{P}\{|\hat{\tau} - \tau| \le \hat{q}_{n,1-\alpha/2} \mid M \le a_n\} \\
&\ge \mathbb{P}\{|\hat{\tau} - \tau| \le \hat{q}_{n,1-\alpha/2},\ \hat{q}_{n,1-\alpha/2} \ge \tilde{q}_{n,1-\alpha/2-\eta} \mid M \le a_n\} \\
&\ge \mathbb{P}\{|\hat{\tau} - \tau| \le \tilde{q}_{n,1-\alpha/2-\eta},\ \hat{q}_{n,1-\alpha/2} \ge \tilde{q}_{n,1-\alpha/2-\eta} \mid M \le a_n\} \\
&\ge \mathbb{P}\{|\hat{\tau} - \tau| \le \tilde{q}_{n,1-\alpha/2-\eta} \mid M \le a_n\} - \mathbb{P}\{\hat{q}_{n,1-\alpha/2} < \tilde{q}_{n,1-\alpha/2-\eta} \mid M \le a_n\}.
\end{aligned}$$

From (A7.30) and Theorem 3, $\mathbb{P}\{\hat{q}_{n,1-\alpha/2} < \tilde{q}_{n,1-\alpha/2-\eta} \mid M \le a_n\} = o(1)$, and

$$\begin{aligned}
\mathbb{P}\{|\hat{\tau} - \tau| \le \tilde{q}_{n,1-\alpha/2-\eta} \mid M \le a_n\} &= \mathbb{P}(|\theta_n| \le \tilde{q}_{n,1-\alpha/2-\eta}) + o(1) \ge \mathbb{P}(|\tilde{\theta}_n| \le \tilde{q}_{n,1-\alpha/2-\eta}) + o(1) \\
&= 1 - \alpha - 2\eta + o(1),
\end{aligned}$$

where the last inequality follows from Li and Ding (2020, Lemma A3). These then imply that

$$\liminf_{n\to\infty} \mathbb{P}(\tau \in \hat{\mathcal{C}}_\alpha \mid M \le a_n) \ge 1 - \alpha - 2\eta.$$

Because the above inequality holds for any $\eta \in (0, (1-\alpha)/2)$, we must have $\liminf_{n\to\infty} \mathbb{P}(\tau \in \hat{\mathcal{C}}_\alpha \mid M \le a_n) \ge 1 - \alpha$.

From the above, Theorem 7(ii) holds. □

**Proof of Theorem 7(iii).** For any $\alpha \in (0, 1)$ and $\eta \in (0, \alpha/2)$, the coverage probability of the confidence interval $\hat{\mathcal{C}}_\alpha$ can be bounded by

$$\begin{aligned}
&\mathbb{P}(\tau \in \hat{\mathcal{C}}_\alpha \mid M \le a_n) \\
&= \mathbb{P}\{|\hat{\tau} - \tau| \le \hat{q}_{n,1-\alpha/2} \mid M \le a_n\} \\
&\le \mathbb{P}\{|\hat{\tau} - \tau| \le \hat{q}_{n,1-\alpha/2},\ \hat{q}_{n,1-\alpha/2} \le \tilde{q}_{n,1-\alpha/2+\eta} \mid M \le a_n\} + \mathbb{P}(\hat{q}_{n,1-\alpha/2} > \tilde{q}_{n,1-\alpha/2+\eta} \mid M \le a_n) \\
&\le \mathbb{P}\{|\hat{\tau} - \tau| \le \tilde{q}_{n,1-\alpha/2+\eta} \mid M \le a_n\} + \mathbb{P}(\hat{q}_{n,1-\alpha/2} > \tilde{q}_{n,1-\alpha/2+\eta} \mid M \le a_n).
\end{aligned} \tag{A7.31}$$

Below we consider the two terms in (A7.31), separately.

First, from Theorem 3, $\mathbb{P}\{|\hat{\tau}-\tau| \le \tilde{q}_{n,1-\alpha/2+\eta} \mid M \le a_n\} = \mathbb{P}(|\theta_n| \le \tilde{q}_{n,1-\alpha/2+\eta}) + o(1)$. Because

$$\tilde{\theta}_n - \theta_n = \Big\{\sqrt{V_{\tau\tau}(1-R_n^2) + n^{-1}S^2_{\tau\setminus \boldsymbol{X}}} - \sqrt{V_{\tau\tau}(1-R_n^2)}\Big\} \cdot \varepsilon_0 = \sqrt{n^{-1}S^2_{\tau\setminus \boldsymbol{X}}} \cdot O_{\mathbb{P}}(1)$$
$$= \sqrt{V_{\tau\tau}(1-R_n^2)} \cdot o_{\mathbb{P}}(1),$$

from Lemma A27, we must have $\sup_{c\in\mathbb{R}} |\mathbb{P}(\theta_n \le c) - \mathbb{P}(\tilde{\theta}_n \le c)| \to 0$. This then implies that

$$\mathbb{P}\{|\hat{\tau}-\tau| \le \tilde{q}_{n,1-\alpha/2+\eta} \mid M \le a_n\} = \mathbb{P}(|\theta_n| \le \tilde{q}_{n,1-\alpha/2+\eta}) + o(1) = \mathbb{P}(|\tilde{\theta}_n| \le \tilde{q}_{n,1-\alpha/2+\eta}) + o(1)$$
$$= 1-\alpha+2\eta+o(1).$$

Second, by the same logic as the proof of (A7.30) in Theorem 7(ii), we can derive that, for any $0<\alpha<\beta<1$, $\mathbb{P}(\tilde{q}_{n,\beta} \le \hat{q}_{n,\alpha} \mid M \le a_n) = \mathbb{E}\{\mathbb{1}(\tilde{q}_{n,\beta} \le \hat{q}_{n,\alpha}) \mid M \le a_n\} \to 0$. This immediately implies that

$$\mathbb{P}(\hat{q}_{n,1-\alpha/2} > \tilde{q}_{n,1-\alpha/2+\eta} \mid M \le a_n) = o(1).$$

From the above, we can know that $\limsup_{n\to\infty} \mathbb{P}(\tau \in \hat{\mathcal{C}}_\alpha \mid M \le a_n) \le 1-\alpha+2\eta$. Because this inequality holds for any $\eta \in (0, \alpha/2)$, we must have $\limsup_{n\to\infty} \mathbb{P}(\tau \in \hat{\mathcal{C}}_\alpha \mid M \le a_n) \le 1-\alpha$. From Theorem 7(ii), we then have $\lim_{n\to\infty} \mathbb{P}(\tau \in \hat{\mathcal{C}}_\alpha \mid M \le a_n) = 1-\alpha$. Therefore, Theorem 7(iii) holds. □

To prove Theorem 8, we need the following lemma.

LEMMA A37. *Let $\varepsilon_0 \sim \mathcal{N}(0,1)$, and $L_{K_n,a_n}$ be the truncated Gaussian random variables defined as in Section 2.3, where $\{K_n\}$ and $\{a_n\}$ are sequences of positive integers and thresholds, and $\varepsilon_0$ is independent of $L_{K_n,a_n}$ for all $n$. Let $\{A_n\}$, $\{B_n\}$, $\{\tilde{A}_n\}$ and $\{\tilde{B}_n\}$ be sequences of nonnegative constants, and for each $n$, define $\psi_n = A_n^{1/2} \cdot \varepsilon_0 + B_n^{1/2} \cdot L_{K_n,a_n}$ and $\tilde{\psi}_n = \tilde{A}_n^{1/2} \cdot \varepsilon_0 + \tilde{B}_n^{1/2} \cdot L_{K_n,a_n}$. For each $n$ and $\alpha \in (0,1)$, let $q_n(\alpha)$ and $\tilde{q}_n(\alpha)$ be the $\alpha$th quantile of $\psi_n$ and $\tilde{\psi}_n$, respectively. If $L_{K_n,a_n} = o_{\mathbb{P}}(1)$, $\tilde{A}_n - A_n = o(A_n)$ and $\tilde{B}_n - B_n = O(A_n)$, then for any $0<\alpha<\beta<1$, as $n\to\infty$, $\mathbb{1}\{\tilde{q}_n(\beta) \le q_n(\alpha)\} \to 0$ and $\mathbb{1}\{q_n(\beta) \le \tilde{q}_n(\alpha)\} \to 0$.*

PROOF OF LEMMA A37. Note that $L_{K_n,a_n} = o_{\mathbb{P}}(1)$. Using the inequality that $|\sqrt{b} - \sqrt{c}| \le \sqrt{|b-c|}$ for any $b, c \ge 0$, we then have

$$\tilde{\psi}_n - \psi_n = \big(\tilde{A}_n^{1/2} - A_n^{1/2}\big)\varepsilon_0 + \big(\tilde{B}_n^{1/2} - B_n^{1/2}\big)L_{K_n,a_n} = |\tilde{A}_n - A_n|^{1/2} \cdot O_{\mathbb{P}}(1) + |\tilde{B}_n - B_n|^{1/2} \cdot o_{\mathbb{P}}(1)$$
$$= A_n^{1/2} \cdot o_{\mathbb{P}}(1).$$

From Lemma A27, this then implies that $\sup_{c\in\mathbb{R}} |\mathbb{P}(\tilde{\psi}_n \le c) - \mathbb{P}(\psi_n \le c)| \to 0$ as $n\to\infty$. Lemma A37 then follows immediately from Lemma A35. □

**Proof of Theorem 8.** Following the notation in the proof of Theorem 7(ii), we define additionally

$$\breve{\theta}_n = \sqrt{\hat{V}_{\tau\tau}(1-\hat{R}_n^2)} \cdot \varepsilon_0 + 0 \cdot L_{K_n,a_n} \equiv \breve{A}_n^{1/2} \cdot \varepsilon_0 + \breve{B}_n^{1/2} \cdot L_{K_n,a_n},$$

and $\breve{q}_{n,\alpha} = q_\alpha(\breve{A}_n, \breve{B}_n, K_n, a_n)$. where $\breve{A}_n$ nd $\breve{B}_n$ denote the squared coefficients of the standard and constrained Gaussian random variables, respectively. From Theorem 7(i) and

Condition 4, $|\check{A}_n - \tilde{A}_n| = |\hat{A}_n - \tilde{A}_n| = o_{\mathbb{P}}(\tilde{A}_n)$, and, for sufficiently large $n$, $|\check{B} - \tilde{B}_n| = \sqrt{V_{\tau\tau} R_n^2} = \sqrt{V_{\tau\tau}} \cdot O(1) = \sqrt{V_{\tau\tau}(1-R_n^2)} \cdot O(1) = O(\tilde{A}_n)$. We can then prove Theorem 8 using almost the same steps as the proof of Theorem 7, where we will replace $\hat{q}_{n,\alpha}$ by $\check{q}_{n,\alpha}$ and use Lemma A37 instead of Lemma A36. For conciseness, we omit the detailed proof here. □

**A8. Regularity Conditions and Diagnoses for Rerandomization.** To prove Proposition 1, we need the following two lemmas.

LEMMA A38. *Let $\boldsymbol{W}_1, \boldsymbol{W}_2, \dots$ be i.i.d. random vectors in $\mathbb{R}^{K_n}$ with $\mathbb{E}[\boldsymbol{W}_i] = \boldsymbol{0}$ and $Cov(\boldsymbol{W}_i) = \boldsymbol{I}_{K_n}$. Assume that*

$$\sup_{\boldsymbol{\nu}\in\mathbb{R}^{K_n}:\boldsymbol{\nu}^\top\boldsymbol{\nu}=1} \mathbb{E}|\nu^\top \boldsymbol{W}_i|^\delta = O(1) \quad and \quad \max_{1\le i\le n} \left| \|\boldsymbol{W}_i\|_2^2 - \mathbb{E}\|\boldsymbol{W}_i\|_2^2 \right| = O_{\mathbb{P}}(\omega(n, K_n)),$$

*for some $\delta > 2$ and some function $\omega(n,K_n)$ increasing in $n$ and $K_n$. If $K_n = O(n^\beta)$ for some $0<\beta<1$, then when $n$ is sufficiently large,*

$$\|\boldsymbol{S}^2_{\boldsymbol{W}} - \boldsymbol{I}_{K_n}\|_{op} = O_{\mathbb{P}}\left( \frac{\omega(n,K_n)}{n} + \left(\frac{K_n}{n}\right)^{\frac{\delta-2}{\delta}} \log^4\left(\frac{n}{K_n}\right) + \left(\frac{K_n}{n}\right)^{\frac{\min\{\delta-2,2\}}{\min\{\delta,4\}}} \right)$$

*and*

$$\max_{1\le i\le n} \|(\boldsymbol{S}^2_{\boldsymbol{W}})^{-1/2}(\boldsymbol{W}_i - \bar{\boldsymbol{W}})\|_2^2 = O_{\mathbb{P}}\left( \omega(n,K_n) + \frac{K_n^{\frac{2\delta-2}{\delta}}}{n^{\frac{\delta-2}{\delta}}} \log^4\left(\frac{n}{K_n}\right) + n\cdot\left(\frac{K_n}{n}\right)^{\frac{\min\{2\delta-2,6\}}{\min\{\delta,4\}}} + K_n \right),$$

*where $\bar{\boldsymbol{W}} = n^{-1}\sum_{i=1}^n \boldsymbol{W}_i$ and $\boldsymbol{S}^2_{\boldsymbol{W}} = (n-1)^{-1}\sum_{i=1}^n (\boldsymbol{W}_i - \bar{\boldsymbol{W}})(\boldsymbol{W}_i - \bar{\boldsymbol{W}})^\top$.*

PROOF OF LEMMA A38. Lemma A38 follows immediately from Lei and Ding (2020, Lemma H.1). Let $\tilde{\boldsymbol{W}} = (\boldsymbol{W}_1 - \bar{\boldsymbol{W}}, \dots, \boldsymbol{W}_n - \bar{\boldsymbol{W}})^\top$, and $\tilde{\boldsymbol{H}} = \tilde{\boldsymbol{W}}(\tilde{\boldsymbol{W}}^\top\tilde{\boldsymbol{W}})^{-1}\tilde{\boldsymbol{W}}^\top$. We can verify that $\tilde{\boldsymbol{W}}^\top\tilde{\boldsymbol{W}} = \sum_{i=1}^n (\boldsymbol{W}_i - \bar{\boldsymbol{W}})(\boldsymbol{W}_i - \bar{\boldsymbol{W}})^\top = (n-1)\boldsymbol{S}^2_{\boldsymbol{W}}$, and the $i$th diagonal element of $\tilde{\boldsymbol{H}}$ has the following equivalent forms:

$$\begin{aligned} H_{ii} &= (\boldsymbol{W}_i - \bar{\boldsymbol{W}})^\top(\tilde{\boldsymbol{W}}^\top\tilde{\boldsymbol{W}})^{-1}(\boldsymbol{W}_i - \bar{\boldsymbol{W}}) = (n-1)^{-1}(\boldsymbol{W}_i - \bar{\boldsymbol{W}})^\top(\boldsymbol{S}^2_{\boldsymbol{W}})^{-1}(\boldsymbol{W}_i - \bar{\boldsymbol{W}}) \\ &= (n-1)^{-1}\|(\boldsymbol{S}^2_{\boldsymbol{W}})^{-1/2}(\boldsymbol{W}_i - \bar{\boldsymbol{W}})\|_2^2. \end{aligned}$$

From Lei and Ding (2020, Lemma H.1), we can know that

$$\left\| n^{-1}\tilde{\boldsymbol{W}}^\top\tilde{\boldsymbol{W}} - \boldsymbol{I}_{K_n} \right\|_{\text{op}} = O_{\mathbb{P}}\left( \frac{\omega(n,K_n)}{n} + \left(\frac{K_n}{n}\right)^{\frac{\delta-2}{\delta}} \log^4\left(\frac{n}{K_n}\right) + \left(\frac{K_n}{n}\right)^{\frac{\min\{\delta-2,2\}}{\min\{\delta,4\}}} \right),$$

and

$$\max_{1\le i\le n} |H_{ii} - K_n/n| = O_{\mathbb{P}}\left( \frac{\omega(n,K_n)}{n} + \left(\frac{K_n}{n}\right)^{\frac{2\delta-2}{\delta}} \log^4\left(\frac{n}{K_n}\right) + \left(\frac{K_n}{n}\right)^{\frac{\min\{2\delta-2,6\}}{\min\{\delta,4\}\}}} \right).$$

These immediately imply that

$$\begin{aligned} \|\boldsymbol{S}^2_{\boldsymbol{W}} - \boldsymbol{I}_{K_n}\|_{\text{op}} &= \left\| \frac{n}{n-1}\left(n^{-1}\tilde{\boldsymbol{W}}^\top\tilde{\boldsymbol{W}} - \boldsymbol{I}_{K_n}\right) + \frac{1}{n-1}\boldsymbol{I}_{K_n} \right\|_{\text{op}} \le \frac{n}{n-1}\left\| n^{-1}\tilde{\boldsymbol{W}}^\top\tilde{\boldsymbol{W}} - \boldsymbol{I}_{K_n} \right\|_{\text{op}} + \frac{1}{n-1} \\ &\le 2\left\| n^{-1}\tilde{\boldsymbol{W}}^\top\tilde{\boldsymbol{W}} - \boldsymbol{I}_{K_n} \right\|_{\text{op}} + \frac{2}{n} \\ &= O_{\mathbb{P}}\left( \frac{\omega(n,K_n)}{n} + \left(\frac{K_n}{n}\right)^{\frac{\delta-2}{\delta}} \log^4\left(\frac{n}{K_n}\right) + \left(\frac{K_n}{n}\right)^{\frac{\min\{\delta-2,2\}}{\min\{\delta,4\}}} \right), \end{aligned}$$

where the last equality holds because $(K_n/n)^{\min\{\delta-2,2\}/\min\{\delta,4\}} \geq (K_n/n)^{1/2} \geq 1/n$ when $n$ is sufficiently large, and

$$\max_{1\leq i\leq n} \|(\boldsymbol{S}_{\boldsymbol{W}}^2)^{-1/2}(\boldsymbol{W}_i - \bar{\boldsymbol{W}})\|_2^2 = (n-1)\max_{1\leq i\leq n} H_{ii} \leq n \max_{1\leq i\leq n} |H_{ii} - K_n/n| + K_n$$
$$= O_{\mathbb{P}}\left(\omega(n,K_n) + \frac{K_n^{\frac{2\delta-2}{\delta}}}{n^{\frac{\delta-2}{\delta}}}\log^4\left(\frac{n}{K_n}\right) + n\left(\frac{K_n}{n}\right)^{\frac{\min\{2\delta-2,6\}}{\min\{\delta,4\}}} + K_n\right).$$

Therefore, Lemma A38 holds. □

LEMMA A39. *Assume that $\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots$ are i.i.d. random vectors in $\mathbb{R}^{K_n}$ with $\max_{1\leq j\leq K_n}\mathbb{E}|W_{ij}|^\delta \leq M$ for some absolute constants $M<\infty$ and $\delta>2$, where $W_{ij}$ is the $j$th coordinate of $\boldsymbol{W}_i$. Then*

$$\max_{1\leq i\leq n} \left|\|\boldsymbol{W}_i\|_2^2 - \mathbb{E}\|\boldsymbol{W}_i\|_2^2\right| = O_{\mathbb{P}}(n^{2/\delta}K_n).$$

PROOF OF LEMMA A39. By Hölder's inequality, for $1\leq i\leq n$ and $1\leq j\leq K_n$, $\{\mathbb{E}(W_{ij}^2)\}^{\delta/2} \leq \mathbb{E}|W_{ij}|^\delta$,

$$\frac{1}{K_n}\left|\|\boldsymbol{W}_i\|_2^2 - \mathbb{E}\|\boldsymbol{W}_i\|_2^2\right| \leq \frac{1}{K_n}\sum_{j=1}^{K_n}\left|W_{ij}^2 - \mathbb{E}(W_{ij}^2)\right| \leq \left(\frac{1}{K_n}\sum_{j=1}^{K_n}\left|W_{ij}^2 - \mathbb{E}(W_{ij}^2)\right|^{\delta/2}\right)^{2/\delta},$$

and

$$\frac{1}{2}\left|W_{ij}^2 - \mathbb{E}(W_{ij}^2)\right| \leq \frac{1}{2}\left(W_{ij}^2 + \mathbb{E}(W_{ij}^2)\right) \leq \left[\frac{1}{2}\left\{|W_{ij}|^\delta + \left(\mathbb{E}(W_{ij}^2)\right)^{\delta/2}\right\}\right]^{2/\delta}.$$

These imply that

$$\mathbb{E}\left\{\left|\|\boldsymbol{W}_i\|_2^2 - \mathbb{E}\|\boldsymbol{W}_i\|_2^2\right|^{\delta/2}\right\}$$
$$= K_n^{\delta/2}\mathbb{E}\left\{\frac{1}{K_n^{\delta/2}}\left|\|\boldsymbol{W}_i\|_2^2 - \mathbb{E}\|\boldsymbol{W}_i\|_2^2\right|^{\delta/2}\right\} \leq K_n^{\delta/2}\mathbb{E}\left(\frac{1}{K_n}\sum_{j=1}^{K_n}\left|W_{ij}^2 - \mathbb{E}(W_{ij}^2)\right|^{\delta/2}\right)$$
$$= K_n^{\delta/2-1}2^{\delta/2}\sum_{j=1}^{K_n}\mathbb{E}\left(\frac{1}{2^{\delta/2}}\left|W_{ij}^2 - \mathbb{E}(W_{ij}^2)\right|^{\delta/2}\right) \leq K_n^{\delta/2-1}2^{\delta/2}\sum_{j=1}^{K_n}\mathbb{E}\left[\frac{1}{2}\left\{|W_{ij}|^\delta + \left(\mathbb{E}(W_{ij}^2)\right)^{\delta/2}\right\}\right]$$
$$= K_n^{\delta/2-1}2^{\delta/2-1}\sum_{j=1}^{K_n}\left\{\mathbb{E}|W_{ij}|^\delta + \left(\mathbb{E}(W_{ij}^2)\right)^{\delta/2}\right\} \leq K_n^{\delta/2-1}2^{\delta/2}\sum_{j=1}^{K_n}\mathbb{E}|W_{ij}|^\delta$$
$$\leq 2^{\delta/2}K_n^{\delta/2}M.$$

Consequently,

$$\mathbb{E}\left\{\max_{1\leq i\leq n}\left|\|\boldsymbol{W}_i\|_2^2 - \mathbb{E}\|\boldsymbol{W}_i\|_2^2\right|^{\delta/2}\right\} \leq \sum_{i=1}^n \mathbb{E}\left\{\left|\|\boldsymbol{W}_i\|_2^2 - \mathbb{E}\|\boldsymbol{W}_i\|_2^2\right|^{\delta/2}\right\} \leq 2^{\delta/2}MK_n^{\delta/2}n.$$

By the Markov's inequality, $\max_{1\leq i\leq n}\left|\|\boldsymbol{W}_i\|_2^2 - \mathbb{E}\|\boldsymbol{W}_i\|_2^2\right| = O_{\mathbb{P}}(K_n n^{2/\delta})$, i.e., Lemma A39 holds. □

**Proof of Proposition 1.** By the same logic as Lemma A10, we can bound $\gamma_n$ by

$$\begin{aligned}
\gamma_n &\equiv \frac{(K_n+1)^{1/4}}{\sqrt{nr_1r_0}} \frac{1}{n}\sum_{i=1}^n \left\| \left(\boldsymbol{S}_{\boldsymbol{u}}^2\right)^{-1/2}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \right\|_2^3 \\
&\le \frac{(K_n+1)^{1/4}}{\sqrt{nr_1r_0}} \frac{1}{n}\sum_{i=1}^n \left\| \left(\boldsymbol{S}_{\boldsymbol{u}}^2\right)^{-1/2}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \right\|_2^2 \cdot \max_{1\le i\le n}\left\| \left(\boldsymbol{S}_{\boldsymbol{u}}^2\right)^{-1/2}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \right\|_2 \\
&= \frac{(K_n+1)^{1/4}}{\sqrt{nr_1r_0}} \frac{(n-1)(K_n+1)}{n} \cdot \max_{1\le i\le n}\left\| \left(\boldsymbol{S}_{\boldsymbol{u}}^2\right)^{-1/2}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \right\|_2 \\
&= \frac{(K_n+1)^{5/4}}{\sqrt{nr_1r_0}} \frac{n-1}{n} \cdot \max_{1\le i\le n}\left\| \left(\boldsymbol{S}_{\boldsymbol{\xi}}^2\right)^{-1/2}(\boldsymbol{\xi}_i - \bar{\boldsymbol{\xi}}) \right\|_2, \qquad \text{(A8.32)}
\end{aligned}$$

where the last equality holds because the quantity $\| \left(\boldsymbol{S}_{\boldsymbol{u}}^2\right)^{-1/2}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\|_2$ is invariant under a non-singular linear transformation of $\boldsymbol{u}_i$'s. Under Condition 6 and the fact that $K_n+1 = O(n^\beta)$ for some $\beta \in (0,1)$, from Lemmas A38 and A39, we can know that

$$\begin{aligned}
&\max_{1\le i\le n} \|(\boldsymbol{S}_{\boldsymbol{\xi}}^2)^{-1/2}(\boldsymbol{\xi}_i - \bar{\boldsymbol{\xi}})\|_2^2 \\
&= O_{\mathbb{P}}\left( n^{2/\delta}(K_n+1) + \frac{(K_n+1)^{\frac{2\delta-2}{\delta}}}{n^{\frac{\delta-2}{\delta}}} \log^4\left(\frac{n}{K_n+1}\right) + n\cdot\left(\frac{K_n+1}{n}\right)^{\frac{\min\{2\delta-2,6\}}{\min\{\delta,4\}}} + K_n + 1 \right).
\end{aligned}$$

Note that as $n\to\infty$, $(K_n+1)/n = o(1)$,

$$\frac{1}{n^{2/\delta}(K_n+1)} \frac{(K_n+1)^{\frac{2\delta-2}{\delta}}}{n^{\frac{\delta-2}{\delta}}} \log^4\left(\frac{n}{K_n+1}\right) = \frac{1}{(K_n+1)^{2/\delta}} \frac{K_n+1}{n} \log^4\left(\frac{n}{K_n+1}\right) = o(1),$$

and

$$\begin{aligned}
&\frac{1}{n^{2/\delta}(K_n+1)} n\cdot\left(\frac{K_n+1}{n}\right)^{\frac{\min\{2\delta-2,6\}}{\min\{\delta,4\}}} \\
&= \mathbb{1}(\delta\le 4)\frac{1}{n^{2/\delta}(K_n+1)} n\cdot\left(\frac{K_n+1}{n}\right)^{2-2/\delta} + \mathbb{1}(\delta>4)\frac{1}{n^{2/\delta}(K_n+1)} n\cdot\left(\frac{K_n+1}{n}\right)^{3/2} \\
&= \mathbb{1}(\delta\le 4)\frac{(K_n+1)^{1-2/\delta}}{n} + \mathbb{1}(\delta>4)\frac{(K_n+1)^{1/2}}{n^{2/\delta+1/2}} = o(1). \qquad \text{(A8.33)}
\end{aligned}$$

Thus, we must have

$$\max_{1\le i\le n} \|(\boldsymbol{S}_{\boldsymbol{\xi}}^2)^{-1/2}(\boldsymbol{\xi}_i - \bar{\boldsymbol{\xi}})\|_2^2 = O_{\mathbb{P}}\left(n^{2/\delta}(K_n+1)\right).$$

From (A8.32), this then implies that

$$\gamma_n = \frac{(K_n+1)^{5/4}}{\sqrt{nr_1r_0}} \frac{n-1}{n} \cdot O_{\mathbb{P}}\left(\sqrt{n^{2/\delta}(K_n+1)}\right) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{r_1r_0}} \frac{(K_n+1)^{7/4}}{n^{1/2-1/\delta}}\right).$$

Therefore, Proposition 1 holds. □

**Proof of Corollary 2(i).** When Condition 6 holds, $r_z^{-1} = O(1)$ and $K_n = o(n^{2/7-4/(7\delta)})$, from Proposition 1, we have

$$\gamma_n = O_{\mathbb{P}}\left(\frac{1}{\sqrt{r_1r_0}} \frac{(K_n+1)^{7/4}}{n^{1/2-1/\delta}}\right) = \frac{\{n^{2/7-4/(7\delta)}\}^{7/4}}{n^{1/2-1/\delta}} \cdot o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

Therefore, Corollary 2(i) holds. □

To prove Corollary 2(ii), we need the following two lemmas.

LEMMA A40. *The squared multiple correlation $R_n^2$ defined as in (5) can be equivalently written as*

$$R_n^2 = \frac{\boldsymbol{S}_{r_0Y(1)+r_1Y(0),\boldsymbol{X}}(\boldsymbol{S}_{\boldsymbol{X}}^2)^{-1}\boldsymbol{S}_{\boldsymbol{X},r_0Y(1)+r_1Y(0)}}{S^2_{r_0Y(1)+r_1Y(0)}},$$

*where $S^2_{r_0Y(1)+r_1Y(0)}$ denotes the finite population variance of $r_0Y(1)+r_1Y(0)$ and $\boldsymbol{S}_{r_0Y(1)+r_1Y(0),\boldsymbol{X}}$ denotes the finite population covariance between $r_0Y(1)+r_1Y(0)$ and $\boldsymbol{X}$.*

PROOF OF LEMMA A40. Let $S_{10}$ be the finite population covariance between $Y(1)$ and $Y(0)$. By some algebra, $S^2_{r_0Y(1)+r_1Y(0)}$ has the following equivalent forms:

$$S^2_{r_0Y(1)+r_1Y(0)} = r_0^2S_1^2 + r_1^2S_0^2 + 2r_1r_0S_{10} = (r_0^2+r_1r_0)S_1^2 + (r_1^2+r_1r_0)S_0^2 - r_1r_0(S_1^2+S_0^2-2S_{10})$$

$$= r_0S_1^2 + r_1S_0^2 - r_1r_0S_\tau^2 = nr_1r_0(n_1^{-1}S_1^2 + n_0^{-1}S_0^2 - n^{-1}S_\tau^2) = nr_1r_0V_{\tau\tau}. \tag{A8.34}$$

By the same logic, we have

$$\boldsymbol{S}_{r_0Y(1)+r_1Y(0),\boldsymbol{X}}(\boldsymbol{S}_{\boldsymbol{X}}^2)^{-1}\boldsymbol{S}_{\boldsymbol{X},r_0Y(1)+r_1Y(0)} = nr_1r_0(n_1^{-1}S^2_{1|\boldsymbol{X}} + n_0^{-1}S^2_{0|\boldsymbol{X}} - n^{-1}S^2_{\tau|\boldsymbol{X}}).$$

Lemma A40 then follows from the definition in (5). □

LEMMA A41. *For any sequence of positive integers $\{K_n\}$ and any sequence of matrices $\boldsymbol{A}_n \in \mathbb{R}^{K_n}$, if $\|\boldsymbol{A}_n - \boldsymbol{I}_{K_n}\|_{op} = o_{\mathbb{P}}(1)$, then $\|\boldsymbol{A}_n^{-1} - \boldsymbol{I}_{K_n}\|_{op} = o_{\mathbb{P}}(1)$.*

PROOF OF LEMMA A41. Note that

$$\begin{aligned}\|\boldsymbol{A}_n^{-1} - \boldsymbol{I}_{K_n}\|_{\text{op}} &= \|\boldsymbol{A}_n^{-1}(\boldsymbol{A}_n - \boldsymbol{I}_{K_n})\|_{\text{op}} \le \|\boldsymbol{A}_n^{-1}\|_{\text{op}} \cdot \|\boldsymbol{A}_n - \boldsymbol{I}_{K_n}\|_{\text{op}} \\ &= \|\boldsymbol{I}_{K_n} + (\boldsymbol{A}_n^{-1} - \boldsymbol{I}_{K_n})\|_{\text{op}} \cdot \|\boldsymbol{A}_n - \boldsymbol{I}_{K_n}\|_{\text{op}} \\ &\le \left(\|\boldsymbol{I}_{K_n}\|_{\text{op}} + \|\boldsymbol{A}_n^{-1} - \boldsymbol{I}_{K_n}\|_{\text{op}}\right) \cdot \|\boldsymbol{A}_n - \boldsymbol{I}_{K_n}\|_{\text{op}} \\ &\le \|\boldsymbol{A}_n - \boldsymbol{I}_{K_n}\|_{\text{op}} + \|\boldsymbol{A}_n - \boldsymbol{I}_{K_n}\|_{\text{op}} \cdot \|\boldsymbol{A}_n^{-1} - \boldsymbol{I}_{K_n}\|_{\text{op}}.\end{aligned}$$

Thus, when $\|\boldsymbol{A}_n - \boldsymbol{I}_{K_n}\|_{\text{op}} < 1$, we have $\|\boldsymbol{A}_n^{-1} - \boldsymbol{I}_{K_n}\|_{\text{op}} \le \|\boldsymbol{A}_n - \boldsymbol{I}_{K_n}\|_{\text{op}}/(1 - \|\boldsymbol{A}_n - \boldsymbol{I}_{K_n}\|_{\text{op}})$. By the property of convergence in probability (e.g., Durrett, 2019, Theorem 2.3.2), we can immediately derive Lemma A41. □

**Proof of Corollary 2(ii).** We first prove that $\|\boldsymbol{S}_{\boldsymbol{\xi}}^2 - \boldsymbol{I}_{K_n+1}\|_{\text{op}} = o_{\mathbb{P}}(1)$. Under Condition 6, from Lemmas A38 and A39,

$$\|\boldsymbol{S}_{\boldsymbol{\xi}}^2 - \boldsymbol{I}_{K_n+1}\|_{\text{op}} = O_{\mathbb{P}}\left(\frac{n^{2/\delta}(K_n+1)}{n} + \left(\frac{K_n+1}{n}\right)^{\frac{\delta-2}{\delta}} \log^4\left(\frac{n}{K_n+1}\right) + \left(\frac{K_n+1}{n}\right)^{\frac{\min\{\delta-2,2\}}{\min\{\delta,4\}}}\right).$$

We can verify that $2/7 - 4/(7\delta) < 1 - 2/\delta$ for $\delta > 2$. Thus, we must have $(K_n+1)/n^{1-2/\delta} = o(1)$, which further implies that $\|\boldsymbol{S}_{\boldsymbol{\xi}}^2 - \boldsymbol{I}_{K_n+1}\|_{\text{op}} = o_{\mathbb{P}}(1)$.

We then prove that $R_n^2 - R^2_{\text{sup},n} = o_{\mathbb{P}}(1)$. By definition,

$$\begin{pmatrix} r_0Y_i(1) + r_1Y_i(0) - \mathbb{E}\{r_0Y_i(1) + r_1Y_i(0)\} \\ \boldsymbol{X}_i - \mathbb{E}(\boldsymbol{X}_i)\end{pmatrix} = \boldsymbol{u}_i = \text{Cov}(\boldsymbol{u})^{1/2}\boldsymbol{\xi}_i = \begin{pmatrix}(1, \boldsymbol{0}_{K_n}^\top)\text{Cov}(\boldsymbol{u})^{1/2}\boldsymbol{\xi}_i \\ (\boldsymbol{0}_{K_n}, \boldsymbol{I}_{K_n})\text{Cov}(\boldsymbol{u})^{1/2}\boldsymbol{\xi}_i\end{pmatrix}$$

$$\equiv \begin{pmatrix}\boldsymbol{a}^\top\boldsymbol{\xi}_i \\ \boldsymbol{B}^\top\boldsymbol{\xi}_i\end{pmatrix},$$

where $\boldsymbol{a}^\top = (1, \boldsymbol{0}_{K_n}^\top)\text{Cov}(\boldsymbol{u})^{1/2} \in \mathbb{R}^{1\times(K_n+1)}$ and $\boldsymbol{B}^\top = (\boldsymbol{0}_{K_n}, \boldsymbol{I}_{K_n})\text{Cov}(\boldsymbol{u})^{1/2} \in \mathbb{R}^{K_n\times(K_n+1)}$. We can then verify that

$$\text{Var}\{r_0Y(1)+r_1Y(0)\} = \boldsymbol{a}^\top\boldsymbol{a}, \quad \text{Var}(\boldsymbol{X}_i) = \boldsymbol{B}^\top\boldsymbol{B}, \quad \text{Cov}\{r_0Y(1)+r_1Y(0), \boldsymbol{X}\} = \boldsymbol{a}^\top\boldsymbol{B}.$$

Consequently, the super population squared multiple correlation between $r_0Y_i(1)+r_1Y_i(0)$ and $\boldsymbol{X}_i$ has the following equivalent forms:

$$R^2_{\text{sup},n} = \frac{\text{Cov}\{r_0Y(1)+r_1Y(0), \boldsymbol{X}_i\}\{\text{Var}(\boldsymbol{X})\}^{-1}\text{Cov}\{\boldsymbol{X}, r_0Y(1)+r_1Y(0)\}}{\text{Var}\{r_0Y(1)+r_1Y(0)\}} = \frac{\boldsymbol{a}^\top\boldsymbol{B}(\boldsymbol{B}^\top\boldsymbol{B})^{-1}\boldsymbol{B}^\top\boldsymbol{a}}{\boldsymbol{a}^\top\boldsymbol{a}}.$$

From Lemma A40, the finite population squared multiple correlation $R_n^2$ satisfies that

$$S^2_{r_0Y(1)+r_1Y(0)}R_n^2 = \boldsymbol{S}_{r_0Y(1)+r_1Y(0),\boldsymbol{X}}(\boldsymbol{S}^2_{\boldsymbol{X}})^{-1}\boldsymbol{S}_{\boldsymbol{X},r_0Y(1)+r_1Y(0)} = \boldsymbol{a}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\boldsymbol{B}\left(\boldsymbol{B}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\boldsymbol{B}\right)^{-1}\boldsymbol{B}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\boldsymbol{a}.$$

Let $\tilde{\boldsymbol{a}} = \boldsymbol{a}/\sqrt{\boldsymbol{a}^\top\boldsymbol{a}}$ and $\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{C}\boldsymbol{\Gamma}^\top$ be the singular value decomposition of $\boldsymbol{B}$, where $\boldsymbol{Q} \in \mathbb{R}^{(K_n+1)\times K_n}$, $\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}_{K_n}$, $\boldsymbol{C} \in \mathbb{R}^{K_n\times K_n}$ is a diagonal matrix, and $\boldsymbol{\Gamma} \in \mathbb{R}^{K_n\times K_n}$ is an orthogonal matrix. We then have

$$\begin{aligned}
&\frac{S^2_{r_0Y(1)+r_1Y(0)}}{\text{Var}\{r_0Y(1)+r_1Y(0)\}}R_n^2 - R^2_{\text{sup}}\\
&= \tilde{\boldsymbol{a}}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\boldsymbol{B}\left(\boldsymbol{B}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\boldsymbol{B}\right)^{-1}\boldsymbol{B}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\tilde{\boldsymbol{a}} - \tilde{\boldsymbol{a}}^\top\boldsymbol{B}(\boldsymbol{B}^\top\boldsymbol{B})^{-1}\boldsymbol{B}^\top\tilde{\boldsymbol{a}} = \tilde{\boldsymbol{a}}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\boldsymbol{Q}\left(\boldsymbol{Q}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\tilde{\boldsymbol{a}} - \tilde{\boldsymbol{a}}^\top\boldsymbol{Q}\boldsymbol{Q}^\top\tilde{\boldsymbol{a}}\\
&= \tilde{\boldsymbol{a}}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\boldsymbol{Q}\left\{\left(\boldsymbol{Q}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\boldsymbol{Q}\right)^{-1} - \boldsymbol{I}_{K_n}\right\}\boldsymbol{Q}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\tilde{\boldsymbol{a}} + \tilde{\boldsymbol{a}}^\top\left(\boldsymbol{S}^2_{\boldsymbol{\xi}} - \boldsymbol{I}_{K_n}\right)\boldsymbol{Q}\boldsymbol{Q}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\tilde{\boldsymbol{a}} + \tilde{\boldsymbol{a}}^\top\boldsymbol{Q}\boldsymbol{Q}^\top\left(\boldsymbol{S}^2_{\boldsymbol{\xi}} - \boldsymbol{I}_{K_n}\right)\tilde{\boldsymbol{a}}.
\end{aligned}$$

By the property of operator norm and the fact that $\tilde{\boldsymbol{a}}^\top\tilde{\boldsymbol{a}} = 1$ and $\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}_{K_n}$, we then have

$$\left|\frac{S^2_{r_0Y(1)+r_1Y(0)}R_n^2}{\text{Var}\{r_0Y(1)+r_1Y(0)\}} - R^2_{\text{sup}}\right| \le \left\|\left(\boldsymbol{Q}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\boldsymbol{Q}\right)^{-1} - \boldsymbol{I}_{K_n}\right\|_{\text{op}}\|\boldsymbol{S}^2_{\boldsymbol{\xi}}\|^2_{\text{op}} + \left\|\boldsymbol{S}^2_{\boldsymbol{\xi}} - \boldsymbol{I}_{K_n}\right\|_{\text{op}}\left(\|\boldsymbol{S}^2_{\boldsymbol{\xi}}\|_{\text{op}} + 1\right).$$

Note that $\|\boldsymbol{S}^2_{\boldsymbol{\xi}} - \boldsymbol{I}_{K_n}\|_{\text{op}} = o_{\mathbb{P}}(1)$, $\|\boldsymbol{S}^2_{\boldsymbol{\xi}}\|_{\text{op}} \le \|\boldsymbol{S}^2_{\boldsymbol{\xi}} - \boldsymbol{I}_{K_n}\|_{\text{op}} + \|\boldsymbol{I}_{K_n}\|_{\text{op}} = 1 + o_{\mathbb{P}}(1)$, and $\|\boldsymbol{Q}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\boldsymbol{Q} - \boldsymbol{I}_{K_n}\|_{\text{op}} = \|\boldsymbol{Q}^\top(\boldsymbol{S}^2_{\boldsymbol{\xi}} - \boldsymbol{I}_{K_n})\boldsymbol{Q}\|_{\text{op}} \le \|\boldsymbol{S}^2_{\boldsymbol{\xi}} - \boldsymbol{I}_{K_n}\|_{\text{op}} = o_{\mathbb{P}}(1)$. From Lemma A41, we can then derive that $\|(\boldsymbol{Q}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\boldsymbol{Q})^{-1} - \boldsymbol{I}_{K_n}\|_{\text{op}} = o_{\mathbb{P}}(1)$, $S^2_{r_0Y(1)+r_1Y(0)}/\text{Var}\{r_0Y(1) + r_1Y(0)\} \cdot R_n^2 - R^2_{\text{sup}} = o_{\mathbb{P}}(1)$,

$$\left|\frac{S^2_{r_0Y(1)+r_1Y(0)}}{\text{Var}\{r_0Y(1)+r_1Y(0)\}} - 1\right| = \left|\tilde{\boldsymbol{a}}^\top\boldsymbol{S}^2_{\boldsymbol{\xi}}\tilde{\boldsymbol{a}} - 1\right| = \left|\tilde{\boldsymbol{a}}^\top\left(\boldsymbol{S}^2_{\boldsymbol{\xi}} - \boldsymbol{I}_{K_n}\right)\tilde{\boldsymbol{a}}\right| \le \|\boldsymbol{S}^2_{\boldsymbol{\xi}} - \boldsymbol{I}_{K_n}\|_{\text{op}} = o_{\mathbb{P}}(1).$$

Consequently,

$$R_n^2 - R^2_{\text{sup},n} = \frac{S^2_{r_0Y(1)+r_1Y(0)}}{\text{Var}\{r_0Y(1)+r_1Y(0)\}}R_n^2 - R^2_{\text{sup}} - \left(\frac{S^2_{r_0Y(1)+r_1Y(0)}}{\text{Var}\{r_0Y(1)+r_1Y(0)\}} - 1\right)R_n^2 = o_{\mathbb{P}}(1).$$

From the above, Corollary 2(ii) holds. □

**Proof of Corollary 2(iii).** First, from (A7.29) and (A8.34), we can know that

$$\begin{aligned}
r_0S^2_{1\setminus\boldsymbol{X}} + r_1S^2_{0\setminus\boldsymbol{X}} &= nr_1r_0\left(n_1^{-1}S^2_{1\setminus\boldsymbol{X}} + n_0^{-1}S^2_{0\setminus\boldsymbol{X}}\right) = nr_1r_0\left\{V_{\tau\tau}(1-R_n^2) + n^{-1}S^2_{\tau\setminus\boldsymbol{X}}\right\}\\
&\ge nr_1r_0V_{\tau\tau}(1-R_n^2) = S^2_{r_0Y(1)+r_1Y(0)}(1-R_n^2).
\end{aligned}$$

From Corollary 2(ii) and its proof, and by the conditions in Corollary 2(iii), we can know that $S^2_{r_0Y(1)+r_1Y(0)} = \text{Var}(r_0Y_i(1)+r_1Y_i(0))\cdot(1+o_{\mathbb{P}}(1))$ and $1-R_n^2 = 1-R^2_{\text{sup},n}+o_{\mathbb{P}}(1) = (1-R^2_{\text{sup},n})\cdot(1+o_{\mathbb{P}}(1))$. These imply that the quantity on the left hand side of (16) satisfies

$$
\begin{aligned}
&\frac{\max_{z\in\{0,1\}}\max_{1\le i\le n}\{Y_i(z)-\bar{Y}(z)\}^2}{r_0S^2_{1\setminus \boldsymbol{X}}+r_1S^2_{0\setminus \boldsymbol{X}}}\cdot\frac{\max\{K_n,1\}}{r_1r_0}\cdot\sqrt{\frac{\max\{1,\log K_n,-\log p_n\}}{n}}\\
&=\frac{\max_{z\in\{0,1\}}\max_{1\le i\le n}\{Y_i(z)-\bar{Y}(z)\}^2}{\text{Var}(r_0Y_i(1)+r_1Y_i(0))\cdot(1-R^2_{\text{sup},n})}\cdot\frac{\max\{K_n,1\}\cdot\sqrt{\max\{1,\log K_n,-\log p_n\}}}{n^{1/2}}\cdot O_{\mathbb{P}}(1)\\
&=\frac{\max_{z\in\{0,1\}}\max_{1\le i\le n}\{Y_i(z)-\bar{Y}(z)\}^2}{\text{Var}(r_0Y_i(1)+r_1Y_i(0))}\cdot\frac{\max\{K_n,1\}\cdot\sqrt{\max\{1,\log K_n,-\log p_n\}}}{n^{1/2}}\cdot O_{\mathbb{P}}(1),
\end{aligned}
$$

where the last equality follows from the condition on $R^2_{\text{sup},n}$.

Second, by some algebra, for $1\le i\le n$,

$$
|Y_i(z)-\bar{Y}(z)|\le|Y_i(z)-\mathbb{E}(Y(z))|+|\bar{Y}(z)-\mathbb{E}(Y(z))|\le 2\max_{1\le i\le n}|Y_i(z)-\mathbb{E}(Y(z))|
$$

and

$$
\begin{aligned}
\max_{1\le i\le n}|Y_i(z)-\mathbb{E}(Y(z))|^b &\le \sum_{i=1}^n|Y_i(z)-\mathbb{E}(Y(z))|^b = n\cdot\mathbb{E}\big\{|Y(z)-\mathbb{E}(Y(z))|^b\big\}\cdot O_{\mathbb{P}}(1)\\
&= n\cdot\mathbb{E}\left\{\left|\frac{Y(z)-\mathbb{E}(Y(z))}{\sqrt{\text{Var}(Y(z))}}\right|^b\right\}\cdot\{\text{Var}(Y(z))\}^{b/2}\cdot O_{\mathbb{P}}(1)\\
&= n\{\text{Var}(Y(z))\}^{b/2}\cdot O_{\mathbb{P}}(1).
\end{aligned}
$$

These imply that $\max_{1\le i\le n}|Y_i(z)-\mathbb{E}(Y(z))|^2 = n^{2/b}\text{Var}(Y(z))\cdot O_{\mathbb{P}}(1)$. Consequently, we can further bound the quantity on the left hand side of (16) by

$$
\begin{aligned}
&\frac{\max_{z\in\{0,1\}}\max_{1\le i\le n}\{Y_i(z)-\bar{Y}(z)\}^2}{r_0S^2_{1\setminus \boldsymbol{X}}+r_1S^2_{0\setminus \boldsymbol{X}}}\cdot\frac{\max\{K_n,1\}}{r_1r_0}\cdot\sqrt{\frac{\max\{1,\log K_n,-\log p_n\}}{n}}\\
&=\frac{\max_{z\in\{0,1\}}\max_{1\le i\le n}\{Y_i(z)-\bar{Y}(z)\}^2}{\text{Var}(r_0Y_i(1)+r_1Y_i(0))}\cdot\frac{\max\{K_n,1\}\cdot\sqrt{\max\{1,\log K_n,-\log p_n\}}}{n^{1/2}}\cdot O_{\mathbb{P}}(1)\\
&=\frac{n^{2/b}\{\text{Var}(Y(1))+\text{Var}(Y(0))\}}{\text{Var}(r_0Y_i(1)+r_1Y_i(0))}\cdot\frac{\max\{K_n,1\}\cdot\sqrt{\max\{1,\log K_n,-\log p_n\}}}{n^{1/2}}\cdot O_{\mathbb{P}}(1)\\
&=\frac{\max\{K_n,1\}\cdot\sqrt{\max\{1,\log K_n,-\log p_n\}}}{n^{1/2-2/b}}\cdot O_{\mathbb{P}}(1)\\
&=\frac{\max\{K_n,1\}\cdot\sqrt{\max\{1,\log K_n\}}}{n^{1/2-2/b}}\cdot O_{\mathbb{P}}(1)+\frac{\max\{K_n,1\}\cdot\sqrt{-\log p_n}}{n^{1/2-2/b}}\cdot O_{\mathbb{P}}(1).
\end{aligned}
$$

Third, because $K_n = O(n^c)$ for some $c<1/2-2/b$, we have

$$
\frac{\max\{K_n,1\}\cdot\sqrt{\max\{1,\log K_n\}}}{n^{1/2-2/b}} = \frac{\log n}{n^{1/2-2/b-c}}\cdot O(1) = o(1),
$$

and

$$
\frac{\max\{K_n,1\}\cdot\sqrt{-\log p_n}}{n^{1/2-2/b}} = \frac{\sqrt{-\log p_n}}{n^{1/2-2/b-c}}\cdot O(1) = \sqrt{\frac{-\log p_n}{n^{1-4/b-2c}}} = o(1),
$$

where the last condition holds by the condition on $p_n$.

From the above, we can know that Corollary 2 holds. $\square$

**Proof of Corollary 3.** We choose $p_n \propto n^{-h}$ for some $0 < h < (1/2 - 1/\delta)/3$. Below we verify that Conditions 1–4 holds with high probability.

First, from Proposition 1, $\gamma_n = o_{\mathbb{P}}(1)$. Second, from Theorem 1 and Proposition 1, and by the construction of $p_n$,

$$\frac{\Delta_n}{p_n} = \frac{\gamma_n + \gamma_n^{1/3}}{p_n} \cdot O(1) = \frac{(\log n)^{(7/12)}}{n^{(1/2-1/\delta)/3}} \cdot n^h \cdot o_{\mathbb{P}}(1) = \frac{(\log n)^{(7/12)}}{n^{(1/2-1/\delta)/3-h}} \cdot o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

Third, $K_n/\log(p_n^{-1}) = K_n/\log(n^h) \cdot O(1) = h^{-1}K_n/\log(n) = o(1)$. Fourth, from Corollary 2(ii), $R_n^2 = R_{\text{sup},n}^2 + o_{\mathbb{P}}(1) \le 1 - c + o_{\mathbb{P}}(1)$.

Therefore, by the property of convergence in probability (e.g., Durrett, 2019, Theorem 2.3.2), Corollary 3 follows from Theorem 5. $\square$

**Comments on the equivalent form of $\gamma_n$ and its bounds in Section 7.2.** We first prove the equivalent form of $\gamma_n$. Because $(e_i, \boldsymbol{X}_i)$ is a non-singular linear transformation of $\boldsymbol{u}_i$, and the finite population covariance between $e_i$ and $\boldsymbol{X}_i$ is zero, we can equivalently write $(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top \boldsymbol{S}_{\boldsymbol{u}}^{-2}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})$ as

$$\begin{aligned}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top \boldsymbol{S}_{\boldsymbol{u}}^{-2}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) &= (e_i, (\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top) \begin{pmatrix} S_e^{-2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S}_{\boldsymbol{X}}^{-2} \end{pmatrix} \begin{pmatrix} e_i \\ \boldsymbol{X}_i - \bar{\boldsymbol{X}} \end{pmatrix} \\ &= e_i^2 + (\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top \boldsymbol{S}_{\boldsymbol{X}}^{-2}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}) = e_i^2 + (n-1)H_{ii},\end{aligned}$$

where the second last equality holds since the finite population variance of $e_i$, $S_e^2$ equals 1, and the last equality follows from the definition of $H_{ii}$'s. This then implies that

$$\gamma_n = \frac{(K_n+1)^{1/4}}{\sqrt{nr_1r_0}} \frac{1}{n} \sum_{i=1}^n \left\| \left(\boldsymbol{S}_{\boldsymbol{u}}^2\right)^{-1/2} (\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \right\|_2^3 = \frac{(K_n+1)^{1/4}}{\sqrt{nr_1r_0}} \frac{1}{n} \sum_{i=1}^n \left(e_i^2 + (n-1)H_{ii}\right)^{3/2}.$$

We then prove the bounds for $\gamma_n$. By Hölder's inequality,

$$\begin{aligned}\left(e_i^2 + (n-1)H_{ii}\right)^{3/2} &= 2^{3/2} \left(\frac{e_i^2 + (n-1)H_{ii}}{2}\right)^{3/2} \le 2^{3/2} \frac{|e_i|^3 + (n-1)^{3/2}H_{ii}^{3/2}}{2} \\ &\le \sqrt{2}(|e_i|^3 + n^{3/2}H_{ii}^{3/2}).\end{aligned}$$

This immediately implies that $\gamma_n \le \sqrt{2}\tilde{\gamma}_n$. Note that

$$\begin{aligned}2\left(e_i^2 + (n-1)H_{ii}\right)^{3/2} &\ge (e_i^2)^{3/2} + \{(n-1)H_{ii}\}^{3/2} \ge |e_i|^3 + \frac{n^{3/2}}{2^{3/2}} H_{ii}^{3/2} \\ &\ge \frac{1}{2^{3/2}} \left(|e_i|^3 + n^{3/2}H_{ii}^{3/2}\right).\end{aligned}$$

This immediately implies that $\gamma_n \ge \tilde{\gamma}_n/2^{5/2} = \tilde{\gamma}_n/(4\sqrt{2})$. $\square$

**Comment on $\sum_{i=1}^n H_{ii}^{3/2}$.** Because $H_{ii}$'s are the diagonal elements of a projection matrix of rank $K_n$, we have $\sum_{i=1}^n H_{ii} = K_n$. By Hölder's inequality, we then have

$$\sum_{i=1}^n H_{ii}^{3/2} = n \cdot \frac{1}{n} \sum_{i=1}^n H_{ii}^{3/2} \ge n \cdot \left(\frac{1}{n}\sum_{i=1}^n H_{ii}\right)^{3/2} = \frac{K_n^{3/2}}{\sqrt{n}}.$$

$\square$

**Comments on the first two moments of $\hat{\tau}-\tau$ under any design.** From (6) and by definition,

$$\hat{\tau}-\tau = \frac{n}{n_1 n_0}\sum_{i=1}^{n} Z_i y_i - \frac{n}{n_0}\bar{Y}(0) - \bar{Y}(1) + \bar{Y}(0) = \frac{n}{n_1 n_0}\sum_{i=1}^{n} Z_i y_i - \frac{n}{n_0}\{r_1\bar{Y}(0) + r_0\bar{Y}(1)\}$$
$$= \frac{n}{n_1 n_0}\sum_{i=1}^{n} Z_i y_i - \frac{n}{n_0}\bar{y} = \frac{n}{n_1 n_0}\sum_{i=1}^{n} Z_i y_i - \frac{n}{n_1 n_0}\sum_{i=1}^{n} Z_i\bar{y} = \frac{n}{n_1 n_0}\sum_{i=1}^{n} Z_i(y_i-\bar{y})$$
$$= \frac{n}{n_1 n_0}\sum_{i=1}^{n}\left(Z_i - \frac{n_1}{n}\right)(y_i-\bar{y}) = \frac{n}{n_1 n_0}\left(\boldsymbol{Z} - r_1\mathbf{1}_n\right)^\top \tilde{\boldsymbol{y}}.$$

Consequently,

$$\mathbb{E}_{\mathcal{D}}(\hat{\tau}-\tau) = \frac{n}{n_1 n_0}\left(\mathbb{E}\boldsymbol{Z} - r_1\mathbf{1}_n\right)^\top \tilde{\boldsymbol{y}} = \frac{1}{n r_1 r_0}\left(\boldsymbol{\pi} - r_1\mathbf{1}_n\right)^\top \tilde{\boldsymbol{y}},$$

and

$$\mathbb{E}_{\mathcal{D}}\{(\hat{\tau}-\tau)^2\} = \left(\frac{n}{n_1 n_0}\right)^2 \tilde{\boldsymbol{y}}^\top \mathbb{E}\left\{\left(\boldsymbol{Z} - r_1\mathbf{1}_n\right)\left(\boldsymbol{Z} - r_1\mathbf{1}_n\right)^\top\right\}\tilde{\boldsymbol{y}}$$
$$= \frac{1}{(n r_1 r_0)^2}\tilde{\boldsymbol{y}}^\top\left\{\mathrm{Var}_{\mathcal{D}}(\boldsymbol{Z}) + (\mathbb{E}\boldsymbol{Z} - r_1\mathbf{1}_n)(\mathbb{E}\boldsymbol{Z} - r_1\mathbf{1}_n)^\top\right\}\tilde{\boldsymbol{y}}$$
$$= \frac{1}{(n r_1 r_0)^2}\tilde{\boldsymbol{y}}^\top\left\{\boldsymbol{\Omega} + (\boldsymbol{\pi} - r_1\mathbf{1}_n)(\boldsymbol{\pi} - r_1\mathbf{1}_n)^\top\right\}\tilde{\boldsymbol{y}}. \tag{A8.35}$$

Therefore, (A1.1) holds. □

**Proof of Proposition A1.** From (A8.34), we can know that

$$V_{\tau\tau} = \frac{1}{n r_1 r_0}S^2_{r_0 Y(1) + r_1 Y(0)} = \frac{1}{n r_1 r_0}\frac{1}{n-1}\tilde{\boldsymbol{y}}^\top\tilde{\boldsymbol{y}}.$$

This implies that

$$V_{\tau\tau}^{-1/2}\mathbb{E}_{\mathcal{D}}(\hat{\tau}-\tau) = \sqrt{\frac{n-1}{n r_1 r_0}}\cdot\frac{\left(\boldsymbol{\pi} - r_1\mathbf{1}_n\right)^\top\tilde{\boldsymbol{y}}}{\|\tilde{\boldsymbol{y}}\|_2},$$

and

$$V_{\tau\tau}^{-1/2}\sqrt{\mathbb{E}_{\mathcal{D}}\{(\hat{\tau}-\tau)^2\}} = \sqrt{\frac{n-1}{n r_1 r_0}}\sqrt{\frac{\tilde{\boldsymbol{y}}^\top\left\{\boldsymbol{\Omega} + (\boldsymbol{\pi} - r_1\mathbf{1}_n)(\boldsymbol{\pi} - r_1\mathbf{1}_n)^\top\right\}\tilde{\boldsymbol{y}}}{\tilde{\boldsymbol{y}}^\top\tilde{\boldsymbol{y}}}}.$$

By some matrix properties, we can know that

$$\max_{\tilde{\boldsymbol{y}}\neq\mathbf{0}} V_{\tau\tau}^{-1/2}\mathbb{E}_{\mathcal{D}}(\hat{\tau}-\tau) = \sqrt{\frac{n-1}{n r_1 r_0}}\cdot\|\boldsymbol{\pi} - r_1\mathbf{1}_n\|_2 \geq 0,$$

and

$$\max_{\tilde{\boldsymbol{y}}\neq\mathbf{0}} V_{\tau\tau}^{-1/2}\sqrt{\mathbb{E}_{\mathcal{D}}\{(\hat{\tau}-\tau)^2\}} = \sqrt{\frac{n-1}{n r_1 r_0}}\cdot\lambda_{\max}^{1/2}\left(\boldsymbol{\Omega} + (\boldsymbol{\pi} - r_1\mathbf{1}_n)(\boldsymbol{\pi} - r_1\mathbf{1}_n)^\top\right).$$

Below we prove the inequality on the right hand side of (A1.3). Let $\boldsymbol{\Psi} = \boldsymbol{\Omega} + (\boldsymbol{\pi} - r_1\mathbf{1}_n)(\boldsymbol{\pi} - r_1\mathbf{1}_n)^\top$ From (A8.35), $\boldsymbol{\Psi} = \mathbb{E}\{(\boldsymbol{Z} - r_1\mathbf{1}_n)(\boldsymbol{Z} - r_1\mathbf{1}_n)^\top\}$. Because $\mathbf{1}_n^\top(\boldsymbol{Z} -$

$r_1\mathbf{1}_n) = \sum_{i=1}^n Z_i - n_1 = 0$, we must have $\mathbf{1}_n^\top \mathbf{\Psi} \mathbf{1}_n = 0$. This implies that $\mathbf{\Psi}$ has at most $n-1$ positive eigenvalues. Consequently,

$$
\begin{aligned}
&(n-1)\lambda_{\max}(\mathbf{\Psi}) \\
&\ge \text{tr}(\mathbf{\Psi}) = \mathbb{E}\left[\text{tr}\left((\boldsymbol{Z} - r_1\mathbf{1}_n)(\boldsymbol{Z} - r_1\mathbf{1}_n)^\top\right)\right] = \mathbb{E}\left[\text{tr}\left((\boldsymbol{Z} - r_1\mathbf{1}_n)^\top(\boldsymbol{Z} - r_1\mathbf{1}_n)\right)\right] \\
&= \mathbb{E}\left\{\text{tr}\left(\boldsymbol{Z}^\top\boldsymbol{Z} - 2r_1\mathbf{1}_n^\top\boldsymbol{Z} + r_1^2\mathbf{1}_n^\top\mathbf{1}_n\right)\right\} = \mathbb{E}(n_1 - 2r_1n_1 + r_1^2 n) = nr_1r_0,
\end{aligned}
$$

i.e., $\lambda_{\max}(\mathbf{\Psi}) \ge nr_1r_0/(n-1)$. This immediately implies the inequality on the right hand side of (A1.3).

From the above, Proposition A1 holds. □

**A9. Asymptotic analysis of regression adjustment under rerandomization.** To prove Theorem A1, we need the following four lemmas.

LEMMA A42. *Under ReM and Condition A1, as $n \to \infty$,*

$$
\begin{aligned}
\sup_{c\in\mathbb{R}}\Big|&\mathbb{P}\{V_{\tau\tau}^{-1/2}(1-\rho_n^2)^{-1/2}\{\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) - \tau\} \le c \mid M \le a_n\} \\
&- \mathbb{P}\left(\sqrt{1 - R_n^2(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)}\, \varepsilon_0 + \sqrt{R_n^2(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)}\, L_{K_n, a_n} \le c\right)\Big| \to 0.
\end{aligned}
$$

PROOF OF LEMMA A42. Define $V_{\tau\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)$ analogously as $V_{\tau\tau}$ in (4), but using the adjusted potential outcomes with adjustment coefficients $\tilde{\boldsymbol{\beta}}_1$ and $\tilde{\boldsymbol{\beta}}_0$. From Li and Ding (2020, Proof of Theorem 5), $V_{\tau\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) = V_{\tau\tau}(1-\rho_n^2)$. Lemma A42 then follows immediately from Theorem 3. □

LEMMA A43. *Consider the same setting as in Lemma A30 and any event $\boldsymbol{Z} \in \mathcal{E} \subset \{0,1\}^N$ with positive probability $p \equiv \mathbb{P}(\boldsymbol{Z} \in \mathcal{E})$. Then for any $t \ge 3 \cdot 71^2/70^2$,*

$$
\mathbb{P}\left(\sum_{k=1}^K \Delta_{w_k}^2 > t\frac{\max\{1, \log K, -\log p\}}{Nf^2}\sum_{k=1}^K \sigma_{w_k}^2 \mid \boldsymbol{Z} \in \mathcal{E}\right) \le 2\exp\left(-\frac{1}{3}\frac{70^2}{71^2}t\right).
$$

PROOF OF LEMMA A43. If $\sum_{k=1}^K \sigma_{w_k}^2 = 0$, then $\sum_{k=1}^K \Delta_{w_k}^2$ is constant zero, and Lemma A43 holds obviously. Below we consider only the case in which $\sum_{k=1}^K \sigma_{w_k}^2 > 0$. From Lemma A30, for any $t > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(\sum_{k=1}^K \Delta_{w_k}^2 > t\frac{\max\{1, \log K, -\log p\}}{Nf^2}\sum_{k=1}^K \sigma_{w_k}^2 \mid \boldsymbol{Z} \in \mathcal{E}\right) \\
&\le 2\frac{K}{p}\exp\left(-\frac{70^2}{71^2}t\max\{1, \log K, -\log p\}\right) \\
&= 2\exp\left(-\frac{70^2}{71^2}t\max\{1, \log K, -\log p\} + \log K - \log p\right).
\end{aligned}
$$

From (A7.25), when $t \ge 3 \cdot 71^2/70^2$, we then have

$$
\mathbb{P}\left(\sum_{k=1}^K \Delta_{w_k}^2 > t\frac{\max\{1, \log K, -\log p\}}{Nf^2}\sum_{k=1}^K \sigma_{w_k}^2 \mid \boldsymbol{Z} \in \mathcal{E}\right) \le 2\exp\left(-\frac{1}{3}\frac{70^2}{71^2}t\right).
$$

Therefore, Lemma A43 holds. □

LEMMA A44. *Under ReM with actual acceptance probability $\tilde{p}_n = \mathbb{P}(M \le a_n)$, if $\min\{n_1, n_0\} \ge 2$ when $n$ is sufficiently large, and $\max\{1, \log J_n, -\log \tilde{p}_n\} = O(nr_1^2 r_0^2)$, then*

$$\begin{aligned}&\left\{r_0(\hat{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_1) + r_1(\hat{\boldsymbol{\beta}}_0 - \tilde{\boldsymbol{\beta}}_0)\right\}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{W}} \\ &= O_{\mathbb{P}}\left( \max_{z\in\{0,1\}} \max_{1\le i\le n} |Y_i(z) - \bar{Y}(z)| \cdot J_n \frac{\max\{1, \log J_n, -\log \tilde{p}_n\}}{nr_1^2 r_0^2} \right).\end{aligned}$$

PROOF OF LEMMA A44. First, we bound the Euclidean norms of $r_0(\hat{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_1)^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{W}}$ and $r_1(\hat{\boldsymbol{\beta}}_0 - \tilde{\boldsymbol{\beta}}_0)^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{W}}$. By definition,

$$\hat{\boldsymbol{\tau}}_{\boldsymbol{W}} = \bar{\boldsymbol{W}}_1 - \bar{\boldsymbol{W}}_0 = \frac{n}{n_1 n_0} \sum_{i=1}^n Z_i(\boldsymbol{W}_i - \bar{\boldsymbol{W}}).$$

We then have

$$\begin{aligned}\left\| r_0(\hat{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_1)^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{W}} \right\|_2 &= \left\| (\boldsymbol{s}_{1,\boldsymbol{W}} - \boldsymbol{S}_{1,\boldsymbol{W}})^\top (\boldsymbol{S}_{\boldsymbol{W}}^2)^{-1} \frac{1}{n_1} \sum_{i=1}^n Z_i(\boldsymbol{W}_i - \bar{\boldsymbol{W}}) \right\|_2 \\ &= \left\| (\boldsymbol{s}_{1,\boldsymbol{w}} - \boldsymbol{S}_{1,\boldsymbol{w}})^\top (\bar{\boldsymbol{w}}_1 - \bar{\boldsymbol{w}}) \right\|_2 \le \|\boldsymbol{s}_{1,\boldsymbol{w}} - \boldsymbol{S}_{1,\boldsymbol{w}}\|_2 \|\bar{\boldsymbol{w}}_1 - \bar{\boldsymbol{w}}\|_2,\end{aligned}$$

and

$$\begin{aligned}\left\| r_1(\hat{\boldsymbol{\beta}}_0 - \tilde{\boldsymbol{\beta}}_0)^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{W}} \right\|_2 &= \left\| (\boldsymbol{s}_{0,\boldsymbol{W}} - \boldsymbol{S}_{0,\boldsymbol{W}})^\top (\boldsymbol{S}_{\boldsymbol{W}}^2)^{-1} \frac{1}{n_0} \sum_{i=1}^n Z_i(\boldsymbol{W}_i - \bar{\boldsymbol{W}}) \right\|_2 \\ &= \left\| (\boldsymbol{s}_{0,\boldsymbol{W}} - \boldsymbol{S}_{0,\boldsymbol{W}})^\top (\boldsymbol{S}_{\boldsymbol{W}}^2)^{-1} \frac{1}{n_0} \sum_{i=1}^n (1 - Z_i)(\boldsymbol{W}_i - \bar{\boldsymbol{W}}) \right\|_2 \\ &= \left\| (\boldsymbol{s}_{0,\boldsymbol{w}} - \boldsymbol{S}_{0,\boldsymbol{w}})^\top (\bar{\boldsymbol{w}}_0 - \bar{\boldsymbol{w}}) \right\|_2 \le \|\boldsymbol{s}_{0,\boldsymbol{w}} - \boldsymbol{S}_{0,\boldsymbol{w}}\|_2 \|\bar{\boldsymbol{w}}_0 - \bar{\boldsymbol{w}}\|_2,\end{aligned}$$

where $\boldsymbol{w}_i = (\boldsymbol{S}_{\boldsymbol{W}}^2)^{-1/2}(\boldsymbol{W}_i - \bar{\boldsymbol{W}})$ denotes the standardized covariate vector for $1 \le i \le n$.

Second, from Lemma A31 and by the same logic as the proof of Lemma A33 (in particular, (A7.28)), we can know that, for $z = 0, 1$,

$$\|\boldsymbol{s}_{z,\boldsymbol{w}} - \boldsymbol{S}_{z,\boldsymbol{w}}\|_2^2 = \max_{z\in\{0,1\}} \max_{1\le i\le n} \{Y_i(z) - \bar{Y}(z)\}^2 \cdot J_n \frac{\max\{1, \log J_n, -\log \tilde{p}_n\}}{nr_z^2} \cdot O_{\mathbb{P}}(1).$$

Third, from Lemma A43, for $z = 0, 1$,

$$\|\bar{\boldsymbol{w}}_z - \bar{\boldsymbol{w}}\|_2^2 = J_n \frac{\max\{1, \log J_n, -\log \tilde{p}_n\}}{nr_z^2} \cdot O_{\mathbb{P}}(1).$$

From the above, for $z = 0, 1$,

$$\begin{aligned}\left\| r_{1-z}(\hat{\boldsymbol{\beta}}_z - \tilde{\boldsymbol{\beta}}_z)^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{W}} \right\|_2 &\le \|\boldsymbol{s}_{z,\boldsymbol{w}} - \boldsymbol{S}_{z,\boldsymbol{w}}\|_2 \|\bar{\boldsymbol{w}}_z - \bar{\boldsymbol{w}}\|_2 \\ &= \max_{z\in\{0,1\}} \max_{1\le i\le n} |Y_i(z) - \bar{Y}(z)| \cdot J_n \frac{\max\{1, \log J_n, -\log \tilde{p}_n\}}{nr_z^2} \cdot O_{\mathbb{P}}(1) \\ &= \max_{z\in\{0,1\}} \max_{1\le i\le n} |Y_i(z) - \bar{Y}(z)| \cdot J_n \frac{\max\{1, \log J_n, -\log \tilde{p}_n\}}{nr_1^2 r_0^2} \cdot O_{\mathbb{P}}(1).\end{aligned}$$

Therefore, Lemma A44 holds. □

LEMMA A45. *Under ReM with actual acceptance probability $\tilde{p}_n = \mathbb{P}(M \le a_n)$,*

(i) *if Condition A1 holds, then* $\max\{1, -\log \tilde{p}_n\} = O(\max\{1, -\log p_n\})$, *recalling that* $p_n = \mathbb{P}(\chi^2_{K_n} \le a_n)$ *is the approximate acceptance probability;*
(ii) *if Conditions A1 and A2 hold, then,* $\max\{1, \log K_n, -\log \tilde{p}_n\} = o(nr_1^2 r_0^2)$.

PROOF OF LEMMA A45. In Lemma A45, (i) follows by the same logic as Lemma A34, and below we focus only on the proof of (ii). From the proof of Lemma A34, $2\max_{z\in\{0,1\}}\max_{1\le i\le n}|Y_i(z) - \bar{Y}(z)|^2 \ge r_0 S_1^2 + r_1 S_0^2 \ge nr_1 r_0 V_{\tau\tau}$. Consequently, from Condition A2,

$$
\begin{aligned}
o(1) &= \frac{\max_{z\in\{0,1\}}\max_{1\le i\le n}|Y_i(z) - \bar{Y}(z)|}{\sqrt{V_{\tau\tau}(1-\rho_n^2)\{1 - R_n^2(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0)\}}} \cdot J_n \cdot \frac{\max\{1, \log J_n, -\log p_n\}}{nr_1^2 r_0^2} \\
&\ge \frac{2^{-1/2}\sqrt{nr_1 r_0 V_{\tau\tau}}}{\sqrt{V_{\tau\tau}}} \cdot J_n \cdot \frac{\max\{1, \log J_n, -\log p_n\}}{nr_1^2 r_0^2} \\
&\ge 2^{-1/2} \cdot \sqrt{nr_1 r_0} \cdot \frac{\max\{1, \log J_n, -\log p_n\}}{nr_1^2 r_0^2} \ge \frac{2^{-1/2}}{\sqrt{nr_1 r_0}}.
\end{aligned}
$$

This implies that $1 = o(\sqrt{nr_1 r_0})$, and thus $\max\{1, \log K_n, -\log \tilde{p}_n\} = o(nr_1^2 r_0^2)$. From the above, Lemma A45 holds. □

**Proof of Theorem A1(i).** Below we prove the first part of Theorem A1. Define

$$
\tilde{\psi}_n = V_{\tau\tau}^{-1/2}(1-\rho_n^2)^{-1/2}\{\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) - \tau\}, \qquad \hat{\psi}_n = V_{\tau\tau}^{-1/2}(1-\rho_n^2)^{-1/2}\{\hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0) - \tau\},
$$

$$
\psi_n = \sqrt{1 - \tilde{R}_n^2}\,\varepsilon_0 + \sqrt{\tilde{R}_n^2}\, L_{K_n, a_n}.
$$

Note that Conditions A1 and A2 hold. From Lemma A42, $n \to \infty$,

$$
\sup_{c\in\mathbb{R}} |\mathbb{P}(\tilde{\psi}_n \le c \mid M \le a_n) - \mathbb{P}(\psi_n \le c)| \to 0.
$$

From Lemmas A44 and A45, under ReM,

$$
\begin{aligned}
\tilde{\psi}_n - \hat{\psi}_n &= \frac{\hat{\tau}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_0) - \hat{\tau}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_0)}{V_{\tau\tau}^{1/2}(1-\rho_n^2)^{1/2}} = \frac{\{r_0(\hat{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_1) + r_1(\hat{\boldsymbol{\beta}}_0 - \tilde{\boldsymbol{\beta}}_0)\}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{W}}}{V_{\tau\tau}^{1/2}(1-\rho_n^2)^{1/2}} \\
&= \sqrt{1 - \tilde{R}_n^2} \cdot o_{\mathbb{P}}(1).
\end{aligned}
$$

For any $\eta > 0$, define $\delta_n = \sqrt{1 - \tilde{R}_n^2} \cdot \eta$. From Lemma A26,

$$
\begin{aligned}
&\sup_{c\in\mathbb{R}} \left|\mathbb{P}(\hat{\psi}_n \le c \mid M \le a_n) - \mathbb{P}(\tilde{\psi}_n \le c \mid M \le a_n)\right| \\
&\le \mathbb{P}(|\hat{\psi}_n - \tilde{\psi}_n| > \delta_n \mid M \le a_n) + \sup_{b\in\mathbb{R}} \mathbb{P}(b < \tilde{\psi}_n \le b + \delta_n \mid M \le a_n) \\
&\le \mathbb{P}(|\hat{\psi}_n - \tilde{\psi}_n| > \delta_n \mid M \le a_n) + \sup_{b\in\mathbb{R}} \mathbb{P}(b < \psi_n \le b + \delta_n) \\
&\quad + 2\sup_{c\in\mathbb{R}} \left|\mathbb{P}(\tilde{\psi}_n \le c \mid M \le a_n) - \mathbb{P}(\psi_n \le c)\right|
\end{aligned}
$$

Letting $n \to \infty$, from the discussion before, we have

$$
\limsup_{n\to\infty} \sup_{c\in\mathbb{R}} \left|\mathbb{P}(\hat{\psi}_n \le c \mid M \le a_n) - \mathbb{P}(\tilde{\psi}_n \le c \mid M \le a_n)\right| \le \limsup_{n\to\infty} \sup_{b\in\mathbb{R}} \mathbb{P}(b < \psi_n \le b + \delta_n).
$$

By definition, for any $b \in \mathbb{R}$, we have, with $b' = b/\sqrt{1-\tilde{R}_n^2}$,

$$\begin{aligned}
&\mathbb{P}(b < \psi_n \le b + \delta_n) \\
&= \mathbb{P}\left(b' < \varepsilon_0 + \sqrt{\frac{\tilde{R}_n^2}{1-\tilde{R}_n^2}}\, L_{K_n,a_n} \le b' + \eta\right) \\
&= \mathbb{E}\left\{\mathbb{P}\left(b' < \varepsilon_0 + \sqrt{\frac{\tilde{R}_n^2}{1-\tilde{R}_n^2}}\, L_{K_n,a_n} \le b' + \eta \mid L_{K,a}\right)\right\} \\
&\le \mathbb{E}\left(\eta/\sqrt{2\pi}\right) = \eta/\sqrt{2\pi},
\end{aligned}$$

where the last inequality holds because the density of $\varepsilon_0$ is bounded by $1/\sqrt{2\pi}$. This then implies that

$$\limsup_{n\to\infty} \sup_{c\in\mathbb{R}} \left|\mathbb{P}(\hat{\psi}_n \le c \mid M \le a_n) - \mathbb{P}(\tilde{\psi}_n \le c \mid M \le a_n)\right| \le \eta/\sqrt{2\pi}.$$

Because the above inequality holds for any $\eta > 0$, we must have, as $n \to \infty$,

$$\lim_{n\to\infty} \sup_{c\in\mathbb{R}} \left|\mathbb{P}(\hat{\psi}_n \le c \mid M \le a_n) - \mathbb{P}(\tilde{\psi}_n \le c \mid M \le a_n)\right| = 0.$$

From the discussion before, as $n \to \infty$,

$$\begin{aligned}
&\sup_{c\in\mathbb{R}} |\mathbb{P}(\hat{\psi}_n \le c \mid M \le a_n) - \mathbb{P}(\psi_n \le c)| \\
&\le \sup_{c\in\mathbb{R}} |\mathbb{P}(\tilde{\psi}_n \le c \mid M \le a_n) - \mathbb{P}(\psi_n \le c)| + \sup_{c\in\mathbb{R}} \left|\mathbb{P}(\hat{\psi}_n \le c \mid M \le a_n) - \mathbb{P}(\tilde{\psi}_n \le c \mid M \le a_n)\right|
\end{aligned}$$

Therefore, the first part of Theorem A1 holds. □

**Proof of Theorem A1(ii).** Because Condition 3 holds and $\limsup_{n\to\infty} \tilde{R}_n^2 < 1$, by the same logic as the proof of Theorem 5, we can know that, as $n \to \infty$,

$$\sup_{c\in\mathbb{R}} \left|\mathbb{P}\left\{\sqrt{1-\tilde{R}_n^2}\,\varepsilon_0 + \sqrt{\tilde{R}_n^2}\, L_{K_n,a_n} \le c\right\} - \mathbb{P}\left\{\sqrt{1-\tilde{R}_n^2}\,\varepsilon_0 \le c\right\}\right| \to 0.$$

From the first part of Theorem A1, we can immediately derive the second part of the theorem. □

## A10. Connection with optimal designs.

A10.1. *Optimal design via minimizing Mahalanobis distance.* In this section, we show that, under certain model assumptions, the optimal design that tries to minimize the mean squared error (MSE) of the difference-in-means estimator will seek the assignment minimizing the Mahalanobis distance for covariate imbalance between the two treatment groups. For more detailed discussion of optimally balanced designs, we refer the readers to Kasy (2016) and Kallus (2018).

Suppose that the potential outcomes satisfy the following model:

$$Y_i(z) = \alpha_z + \boldsymbol{\beta}_z^\top \boldsymbol{X}_i + e_i(z), \quad (z = 0, 1; i = 1, 2, \ldots, n), \tag{A10.36}$$

where $(e_i(1), e_i(0))$'s are mutually independent across all units, and $e_i(z)$'s have mean zero and the same variance $\sigma_z^2$ across all units for $z = 0, 1$. Throughout the discussion in this section, the covariates $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are fixed constants or equivalently being conditioned on. Under model (A10.36), the expected treatment effect for each unit $i$ is then

$$\tau_i^\star = \mathbb{E}\{\tau_i\} = \mathbb{E}\{Y_i(1) - Y_i(0)\} = \alpha_1 - \alpha_0 + (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^\top \boldsymbol{X}_i, \quad (i = 1, 2, \ldots, n)$$

and its average over all units is

$$\tau^\star = \mathbb{E}\{\tau\} = \frac{1}{n}\sum_{i=1}^n \tau_i^\star = \alpha_1 - \alpha_0 + (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^\top \bar{\boldsymbol{X}}.$$

We are interested in estimating the average treatment effect $\tau^\star$ using the difference-in-means estimator $\hat{\tau}$. Moreover, we will write the difference-in-means estimator as $\hat{\tau}(\boldsymbol{Z})$, to emphasize its dependence on the treatment assignment. For any fixed treatment assignment $\boldsymbol{z}$, the MSE of the corresponding difference-in-means estimator under model (A10.36) has the following decomposition:

$$\begin{aligned} \mathbb{E}[\{\hat{\tau}(\boldsymbol{z}) - \tau^\star\}^2] &= [\mathbb{E}\{\hat{\tau}(\boldsymbol{z}) - \tau^\star\}]^2 + \mathrm{Var}\{\hat{\tau}(\boldsymbol{z}) - \tau^\star\} \\ &= \{\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}(\boldsymbol{z})\}^2 + \frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}, \end{aligned} \tag{A10.37}$$

where $\tilde{\boldsymbol{\beta}} = r_0\boldsymbol{\beta}_1 + r_1\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}(\boldsymbol{z})$ is the difference-in-means of covariates under the treatment assignment $\boldsymbol{z}$. From (A10.37), the optimal assignment minimizing the MSE is equivalently the one minimizing the squared bias of $\hat{\tau}$. Since $\tilde{\boldsymbol{\beta}}$ is unknown, similar to Kallus (2018, Section 2.3.3), we consider the worst-case squared bias after some standardization. Specifically, let $\mu_i \equiv r_0\mathbb{E}\{Y_i(1)\} + r_1\mathbb{E}\{Y_i(0)\}$ be a certain weighted average of expected potential outcomes for each unit $i$. The finite population variance of $\mu_i$'s across all units can be equivalently written as $S_\mu^2 = \tilde{\boldsymbol{\beta}}^\top \boldsymbol{S}_{\boldsymbol{X}}^2 \tilde{\boldsymbol{\beta}}$. We then consider the worst-case squared bias of $\hat{\tau}$ standardized by $S_\mu^2$, which has the following equivalent forms:

$$\sup_{\tilde{\boldsymbol{\beta}} \neq \boldsymbol{0}} \frac{\{\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}(\boldsymbol{z})\}^2}{S_\mu^2} = \sup_{\tilde{\boldsymbol{\beta}} \neq \boldsymbol{0}} \frac{\{\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}(\boldsymbol{z})\}^2}{\tilde{\boldsymbol{\beta}}^\top \boldsymbol{S}_{\boldsymbol{X}}^2 \tilde{\boldsymbol{\beta}}} = \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}(\boldsymbol{z})^\top \boldsymbol{S}_{\boldsymbol{X}}^{-2} \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}(\boldsymbol{z}) = \frac{n}{n_1 n_0} M(\boldsymbol{z}), \tag{A10.38}$$

where $M(\boldsymbol{z})$ is the Mahalanobis distance of covariate means in two treatment groups under the treatment assignment $\boldsymbol{z}$, as defined in Section 2.2. Consequently, the assignment minimizing the Mahalanobis distance is equivalently the one that minimizes worst-case standardized squared bias. Therefore, under the proposed model and criterion, minimizing the Mahalanobis distance leads to the optimal design.

A10.2. *Model-based efficiency of rerandomization.* We now briefly discuss the efficiency of rerandomization under the proposed model in (A10.36). In short, we will show that, under ReM with properly diminishing threshold for covariate imbalance, the design can asymptotically achieve the optimal efficiency.

To gain some intuition, beyond the equal variance assumption, we further assume that $e_i(z)$'s are i.i.d. across all units, for $z = 0, 1$; the i.i.d. assumption will be relaxed later. We can verify that the difference-in-means estimator has the following decomposition:

$$\hat{\tau} - \tau^\star = \tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} + \left\{\frac{1}{n_1}\sum_{i=1}^n Z_i e_i(1) - \frac{1}{n_0}\sum_{i=1}^n (1 - Z_i) e_i(0)\right\} \equiv \tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} + \hat{\tau}_e, \tag{A10.39}$$

where $\tilde{\boldsymbol{\beta}} = r_0\boldsymbol{\beta}_1 + r_1\boldsymbol{\beta}_0$ is defined the same as before and $\hat{\tau}_e$ is the difference-in-means for the residual potential outcomes. For treatment assignment mechanisms depending only on the covariates, such as rerandomization based on $\boldsymbol{X}_i$'s, $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ and $\hat{\tau}_e$ must be mutually independent, with $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ following its randomization distribution and

$$\hat{\tau}_e \sim \frac{1}{n_1}\sum_{i=1}^{n_1} e_i(1) - \frac{1}{n_0}\sum_{i=n_1+1}^{n} e_i(0). \tag{A10.40}$$

This is because the conditional distribution of $\hat{\tau}_e$ given $\boldsymbol{Z}$ must follow the distribution on the right hand side of (A10.40). By the standard central limit theorem, when $e_i(1)$'s and $e_i(0)$'s have finite second moments, and the proportions of treated and control units $r_1$ and $r_0$ have positive limits as $n \to \infty$, $\sqrt{n}\hat{\tau}_e$ will asymptotically converge to a Gaussian distribution with mean zero and variance $r_1^{-1}\sigma_1^2 + r_0^{-1}\sigma_0^2$. By Slutsky's theorem, as long as $\sqrt{n}\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} = o_{\mathbb{P}}(1)$, $\sqrt{n}(\hat{\tau} - \tau^\star)$ will converge to the same asymptotic distribution as $\sqrt{n}\hat{\tau}_e$, which is actually the optimal efficiency that we can expect, as implied by (A10.37).

Below we rigorously study the asymptotic efficiency of rerandomization under model (A10.36). First, we allow the residuals $e_i(z)$'s to be non-identically distributed for $z = 0, 1$, but require them to have bounded third absolute moments.

CONDITION A3. There exists some finite constant $C_e$ such that, for all $n$,

$$\max_{1\le i\le n} \mathbb{E}[|e_i(1)|^3] \le C_e, \quad \& \quad \max_{1\le i\le n} \mathbb{E}[|e_i(0)|^3] \le C_e.$$

Second, to conduct the optimal rerandomization with diminishing covariate imbalance, we invoke similar regularity conditions as Conditions 1–3. Because here we care only the difference-in-means of covariates, we redefine the quantifies in the main paper by excluding the potential outcomes there. Specifically, analogous to $\gamma_n$ and $\Delta_n$ in (7) and (8), define

$$\gamma_n^\star \equiv \frac{K_n^{1/4}}{\sqrt{nr_1r_0}}\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{S}_{\boldsymbol{X}}^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})\right\|_2^3, \quad \Delta_n^\star \equiv \sup_{\mathcal{Q}\in\mathcal{C}_{K_n}}\left|\mathbb{P}\left(\boldsymbol{V}_{\boldsymbol{xx}}^{-1/2}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \in \mathcal{Q}\right) - \mathbb{P}\left(\boldsymbol{\varepsilon}^\star \in \mathcal{Q}\right)\right|,$$

where $\boldsymbol{\varepsilon}^\star \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{K_n})$. We then invoke the following regularity condition.

CONDITION A4. Conditions 1–3, with $\gamma_n$ and $\Delta_n$ replaced by $\gamma_n^\star$ and $\Delta_n^\star$, hold.

Third, we assume the following condition on model (A10.36) and proportions of treated and control units.

CONDITION A5. As the sample size $n$ increases,

(i) the residual variances $\sigma_1^2$ and $\sigma_0^2$ do not vary, and at least one of them is positive;
(ii) the proportions of treated and control units satisfy $\sqrt{n}\min\{r_1, r_0\} \to \infty$ as $n \to \infty$;
(iii) the weighted average of expected potential outcomes has bounded finite population variance, i.e., $S_\mu^2 = \tilde{\boldsymbol{\beta}}^\top \boldsymbol{S}_{\boldsymbol{X}}^2 \tilde{\boldsymbol{\beta}} \le C_\mu$ for all $n$ and some finite constant $C_\mu$.

Under the above conditions, $\sqrt{n}\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ will converge in probability to zero. This implies that, asymptotically, rerandomization will achieve the optimal efficiency (or equivalently be the optimal design) under model (A10.36). We summarize the results in the following theorem.

THEOREM A3. *Under ReM and Conditions A3–A5, as $n \to \infty$,*

$$\frac{\hat{\tau} - \tau^\star}{\sqrt{\sigma_1^2/n_1 + \sigma_0^2/n_0}} \mid M \le a_n \xrightarrow{d} \mathcal{N}(0,1).$$

Theorem A3 shows that, under model (A10.36), ReM with properly diminishing threshold can asymptotically achieve the optimal efficiency as implied by (A10.37). Both Theorem 5 and Theorem A3 show the optimality of ReM with properly diminishing covariate imbalance threshold. However, their justification is quite different. First, the two theorems rely on different sources of randomness. Theorem 5 views all the potential outcomes as fixed constant (or equivalently conditioning on all the potential outcomes), and the randomness comes solely from the treatment assignment; while Theorem A3 assumes additionally that the potential outcomes are random following model (A10.37). Second, due to the aforementioned difference, the estimands for average treatment effects have different forms in the two theorems. Theorem 5 focuses on $\tau = n^{-1}\sum_{i=1}^{n}\{Y_i(1) - Y_i(0)\}$, while Theorem A3 focuses on $\tau^\star = \mathbb{E}(\tau) = n^{-1}\sum_{i=1}^{n}\mathbb{E}\{Y_i(1) - Y_i(0)\}$ under model (A10.37).

A10.3. *Technical details.*

ADDITIONAL DETAILS FOR (A10.37)–(A10.39). First, we prove the decomposition of $\hat{\tau} - \tau^\star$ in (A10.39). By some algebra,

$$
\begin{aligned}
\hat{\tau} - \tau^\star &= \frac{1}{n_1} Z_i Y_i(1) - \frac{1}{n_0} Z_i Y_i(0) - \{\alpha_1 - \alpha_0 + (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^\top \bar{\boldsymbol{X}}\} \\
&= \alpha_1 + \boldsymbol{\beta}_1^\top \bar{\boldsymbol{X}}_1 + \bar{e}_1 - \alpha_0 - \boldsymbol{\beta}_0^\top \bar{\boldsymbol{X}}_0 - \bar{e}_0 - \{\alpha_1 - \alpha_0 + (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^\top (r_1 \bar{\boldsymbol{X}}_1 + r_0 \bar{\boldsymbol{X}}_0)\} \\
&= \boldsymbol{\beta}_1^\top r_0 (\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_0) + \boldsymbol{\beta}_0^\top r_1 (\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_0) + \bar{e}_1 - \bar{e}_0 \\
&= (r_0 \boldsymbol{\beta}_1 + r_1 \boldsymbol{\beta}_0)^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} + \hat{\tau}_e = \tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} + \hat{\tau}_e,
\end{aligned}
$$

where $\bar{e}_z$ denotes the average of residual potential outcomes $e_i(z)$'s for units under treatment arm $z$.

Second, we prove the decomposition of the model-based MSE in (A10.37). From the decomposition in (A10.39) and the property of model (A10.36), $\mathbb{E}\{\hat{\tau}(\boldsymbol{z}) - \tau^\star\} = \tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$, and $\text{Var}\{\hat{\tau}(\boldsymbol{z}) - \tau^\star\} = \text{Var}\{\hat{\tau}_e(\boldsymbol{z})\} = \sigma_1^2/n_1 + \sigma_0^2/n_0$, where we use $\hat{\tau}_e(\boldsymbol{z})$ to emphasize that it is the difference-in-means of residual potential outcomes under the treatment assignment $\boldsymbol{z}$. These then immediately imply the decomposition in (A10.37).

Third, we prove (A10.38). By the definition of matrix norm, letting $\check{\boldsymbol{\beta}} = \boldsymbol{S}_{\boldsymbol{X}} \tilde{\boldsymbol{\beta}}$, we have

$$
\sup_{\tilde{\boldsymbol{\beta}} \neq \boldsymbol{0}} \frac{\{\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}(\boldsymbol{z})\}^2}{\tilde{\boldsymbol{\beta}}^\top \boldsymbol{S}_{\boldsymbol{X}}^2 \tilde{\boldsymbol{\beta}}} = \sup_{\check{\boldsymbol{\beta}} \neq \boldsymbol{0}} \frac{\|\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}(\boldsymbol{z})^\top \boldsymbol{S}_{\boldsymbol{X}}^{-1} \check{\boldsymbol{\beta}}\|_2^2}{\check{\boldsymbol{\beta}}^\top \check{\boldsymbol{\beta}}} = \|\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}(\boldsymbol{z})^\top \boldsymbol{S}_{\boldsymbol{X}}^{-1}\|_2^2 = \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}(\boldsymbol{z})^\top \boldsymbol{S}_{\boldsymbol{X}}^{-2} \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}(\boldsymbol{z}).
$$

This immediately implies (A10.38). $\square$

PROOF OF THEOREM A3. Below we consider the two terms in the decomposition (A10.39) separately.

First, we consider the limiting distribution of $\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$. By the same logic as the proof of Theorem 1, under Condition A4, as $n \to \infty$,

$$
\sup_{c \in \mathbb{R}} \left| \mathbb{P}\left(\sqrt{n}\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \le c \mid M \le a_n\right) - \mathbb{P}\left(\sqrt{n}\tilde{\boldsymbol{\beta}}^\top \tilde{\boldsymbol{\tau}}_{\boldsymbol{X}} \le c \mid \tilde{M} \le a_n\right) \right| \to 0,
$$

By the same logic as the proof of Li, Ding and Rubin (2018, Theorem 1),

$$
\sqrt{n}\tilde{\boldsymbol{\beta}}^\top \tilde{\boldsymbol{\tau}}_{\boldsymbol{X}} \mid \tilde{M} \le a_n \;\sim\; \sqrt{n}\tilde{\boldsymbol{\beta}}^\top \boldsymbol{V}_{\boldsymbol{xx}}^{1/2} \boldsymbol{\varepsilon}^\star \mid (\boldsymbol{\varepsilon}^\star)^\top \boldsymbol{\varepsilon}^\star \le a_n \;\sim\; \sqrt{n}\|\boldsymbol{V}_{\boldsymbol{xx}}^{1/2}\tilde{\boldsymbol{\beta}}\|_2 L_{K_n, a_n},
$$

recalling that $\boldsymbol{\varepsilon}^\star \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{K_n})$. From Conditions A4 and A5 and using Theorem 4(i) and Proposition A2,

$$
\sqrt{n}\|\boldsymbol{V}_{\boldsymbol{xx}}^{1/2}\tilde{\boldsymbol{\beta}}\|_2 L_{K_n, a_n} = \sqrt{(r_1 r_0)^{-1}\tilde{\boldsymbol{\beta}}^\top \boldsymbol{S}_{\boldsymbol{X}}^2 \tilde{\boldsymbol{\beta}}}\, L_{K_n, a_n} = O(1) \cdot o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).
$$

From the above, we can derive that $\sqrt{n}\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} = o_{\mathbb{P}}(1)$.

Second, we consider the limiting of $\hat{\tau}_e$. For any fixed acceptable assignment $\boldsymbol{z}$ under ReM, from the standard univariate Berry–Esseen theorem (Esseen, 1942), there exists a universal constant $C$ such that

$$\begin{aligned}
&\sup_{c\in\mathbb{R}} |\mathbb{P}\{\text{Var}(\hat{\tau}_e)^{-1/2}\hat{\tau}_e \le c \mid \boldsymbol{Z} \equiv \boldsymbol{z}\} - \mathbb{P}(\varepsilon_0 \le c)| \\
&\le C \frac{n_1^{-2}\mathbb{E}|e_i^3(1)| + n_0^{-2}\mathbb{E}|e_i^3(0)|}{(n_1^{-1}\sigma_1^2 + n_0^{-1}\sigma_0^2)^{3/2}} \le \frac{C}{\sqrt{n}} \frac{(1/r_1 + 1/r_0)C_e}{(\sigma_1^2/r_1 + \sigma_0^2/r_0)^{3/2}} \le \frac{C}{\sqrt{n}} \frac{2C_e/\min\{r_1, r_0\}}{(\sigma_1^2 + \sigma_0^2)^{3/2}} \\
&\le \frac{2CC_e(\sigma_1^2 + \sigma_0^2)^{-3/2}}{\sqrt{n}\min\{r_1, r_0\}}.
\end{aligned}$$

where $\varepsilon_0 \sim \mathcal{N}(0,1)$. This then implies that, for any $c \in \mathbb{R}$,

$$\begin{aligned}
&|\mathbb{P}\{\text{Var}(\hat{\tau}_e)^{-1/2}\hat{\tau}_e \le c \mid M \le a_n\} - \mathbb{P}(\varepsilon_0 \le c)| \\
&= |\mathbb{E}[\mathbb{P}\{\text{Var}(\hat{\tau}_e)^{-1/2}\hat{\tau}_e \le c \mid \boldsymbol{Z}\} \mid M \le a_n] - \mathbb{P}(\varepsilon_0 \le c)| \\
&\le \mathbb{E}[|\mathbb{P}\{\text{Var}(\hat{\tau}_e)^{-1/2}\hat{\tau}_e \le c \mid \boldsymbol{Z}\} - \mathbb{P}(\varepsilon_0 \le c)| \mid M \le a_n] \\
&\le \frac{2CC_e(\sigma_1^2 + \sigma_0^2)^{-3/2}}{\sqrt{n}\min\{r_1, r_0\}},
\end{aligned}$$

i.e.,

$$\sup_{c\in\mathbb{R}} |\mathbb{P}\{\text{Var}(\hat{\tau}_e)^{-1/2}\hat{\tau}_e \le c \mid M \le a_n\} - \mathbb{P}(\varepsilon_0 \le c)| \le \frac{2CC_e(\sigma_1^2 + \sigma_0^2)^{-3/2}}{\sqrt{n}\min\{r_1, r_0\}}. \tag{A10.41}$$

Finally, we study the limiting distribution of $\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} + \hat{\tau}_e$. From Lemma A26, for any constant $\delta > 0$,

$$\begin{aligned}
&\sup_{c\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} + \hat{\tau}_e}{\text{Var}(\hat{\tau}_e)^{1/2}} \le c \mid M \le a_n \right\} - \mathbb{P}\left\{ \frac{\hat{\tau}_e}{\text{Var}(\hat{\tau}_e)^{1/2}} \le c \mid M \le a_n \right\} \right| \\
&\le \mathbb{P}\left\{ \left| \frac{\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}}{\text{Var}(\hat{\tau}_e)^{1/2}} \right| > \delta \mid M \le a_n \right\} + \sup_{b\in\mathbb{R}} \mathbb{P}\left\{ b < \frac{\hat{\tau}_e}{\text{Var}(\hat{\tau}_e)^{1/2}} \le b + \delta \mid M \le a_n \right\}.
\end{aligned} \tag{A10.42}$$

Because $\sqrt{n}\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} = o_{\mathbb{P}}(1)$ and $n\text{Var}(\hat{\tau}_e) = \sigma_1^2/r_1 + \sigma_0^2/r_0 \ge \sigma_1^2 + \sigma_0^2 > 0$, $\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}/\text{Var}(\hat{\tau}_e)^{1/2} = o_{\mathbb{P}}(1)$, and thus the first term in (A10.42) converges to zero as $n \to \infty$. From (A10.41),

$$\begin{aligned}
\sup_{b\in\mathbb{R}} \mathbb{P}\left\{ b < \frac{\hat{\tau}_e}{\text{Var}(\hat{\tau}_e)^{1/2}} \le b + \delta \mid M \le a_n \right\} &\le \sup_{b\in\mathbb{R}} \mathbb{P}(b < \varepsilon_0 \le b + \delta) + \frac{4CC_e(\sigma_1^2 + \sigma_0^2)^{-3/2}}{\sqrt{n}\min\{r_1, r_0\}} \\
&\le \frac{\delta}{2\pi} + \frac{4CC_e(\sigma_1^2 + \sigma_0^2)^{-3/2}}{\sqrt{n}\min\{r_1, r_0\}},
\end{aligned}$$

where, from Condition A5, the upper bound converges to $\delta/(2\pi)$ as $n \to \infty$. From the above, for any constant $\delta > 0$,

$$\limsup_{n\to\infty} \sup_{c\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} + \hat{\tau}_e}{\text{Var}(\hat{\tau}_e)^{1/2}} \le c \mid M \le a_n \right\} - \mathbb{P}\left\{ \frac{\hat{\tau}_e}{\text{Var}(\hat{\tau}_e)^{1/2}} \le c \mid M \le a_n \right\} \right| \le \frac{\delta}{2\pi}.$$

This immediately implies that, as $n \to \infty$,

$$\sup_{c \in \mathbb{R}} \left| \mathbb{P}\left\{ \frac{\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} + \hat{\tau}_e}{\mathrm{Var}(\hat{\tau}_e)^{1/2}} \le c \mid M \le a_n \right\} - \mathbb{P}\left\{ \frac{\hat{\tau}_e}{\mathrm{Var}(\hat{\tau}_e)^{1/2}} \le c \mid M \le a_n \right\} \right| \to 0.$$

From (A10.41) and Condition A5, we further have

$$\sup_{c \in \mathbb{R}} \left| \mathbb{P}\left\{ \frac{\tilde{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} + \hat{\tau}_e}{\mathrm{Var}(\hat{\tau}_e)^{1/2}} \le c \mid M \le a_n \right\} - \mathbb{P}\left\{ \varepsilon_0 \le c \right\} \right| \to 0,$$

i.e.,

$$\frac{\hat{\tau} - \tau^\star}{\sqrt{\sigma_1^2/n_1 + \sigma_0^2/n_0}} \mid M \le a_n \xrightarrow{d} \mathcal{N}(0, 1).$$

Therefore, Theorem A3 holds. □

## REFERENCES


ANGRIST, J., LANG, D. and OREOPOULOS, P. (2009). Incentives and Services for College Achievement: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics* **1** 136–63.

BALL, K. (1993). The reverse isoperimetric problem for Gaussian measure. *Discrete & Computational Geometry* **10** 411–420.

BARDENET, R. and MAILLARD, O.-A. (2015). Concentration inequalities for sampling without replacement. *Bernoulli* **21** 1361–1385.

BENTKUS, V. (2005). A Lyapunov-type bound in $\mathbb{R}^d$. *Theory of Probability & Its Applications* **49** 311–323.

BLONIARZ, A., LIU, H., ZHANG, C. H., SEKHON, J. S. and YU, B. (2016). Lasso adjustments of treatment effect estimates in randomized experiments. *Proceedings of the National Academy of Sciences* **113** 7383–7390.

BOLTHAUSEN, E. and GÖTZE, F. (1993). The Rate of Convergence for Multivariate Sampling Statistics. *Annals of Statistics* **21** 1692–1710.

DURRETT, R. (2019). *Probability: theory and examples* **49**. Cambridge university press.

ESSEEN, C. G. (1942). On the Liapunov limit error in the theory of probability. *Ark. Mat. Astr. Fys.* **28** 1–19.

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

HOEFFDING, W. (1963). Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association* **58** 13–30.

KALLUS, N. (2018). Optimal a priori balance in the design of controlled experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80** 85–112.

KAPELNER, A., KRIEGER, A. M., SKLAR, M., SHALIT, U. and AZRIEL, D. (2020). Harmonizing Optimized Designs With Classic Randomization in Experiments. *The American Statistician* **in press**.

KARATSUBA, E. A. (2001). On the asymptotic representation of the Euler gamma function by Ramanujan. *Journal of Computational and Applied Mathematics* **135** 225–240.

KASY, M. (2016). Why Experimenters Might Not Always Want to Randomize, and What They Could Do Instead. *Political Analysis* **24** 324–338.

LEI, L. and DING, P. (2020). Regression adjustment in completely randomized experiments with a diverging number of covariates. *Biometrika* **in press**.

LI, X., DING, P. and RUBIN, D. B. (2018). Asymptotic theory of rerandomization in treatment–control experiments. *Proceedings of the National Academy of Sciences* **115** 9157–9162.

LI, X. and DING, P. (2020). Rerandomization and regression adjustment. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82** 241–268.

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

MORGAN, K. L. and RUBIN, D. B. (2012). Rerandomization to improve covariate balance in experiments. *Annals of Statistics* **40** 1263–1282.

NAZAROV, F. (2003). On the Maximal Perimeter of a Convex Set in $\mathbb{R}^n$ with Respect to a Gaussian Measure. In *Geometric aspects of functional analysis* 169–187. Springer.

NORDIN, M. and SCHULTZBERG, M. (2020). Properties of restricted randomization with implications for experimental design. *arXiv preprint arXiv:2006.14888*.

RAIČ, M. (2015). Multivariate normal approximation: permutation statistics, local dependence and beyond.

RAIČ, M. (2019). A multivariate Berry–Esseen theorem with explicit constants. *Bernoulli* **25** 2824–2853.

RIGOLLET, P. and HÜTTER, J.-C. (2015). High dimensional statistics. *Lecture notes for course 18S997* **813** 814.

WU, C. F. (1981). On the Robustness and Efficiency of Some Randomized Designs. *The Annals of Statistics* **9** 1168 – 1177.