# ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

*An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics*

https://www.econometricsociety.org/

*Econometrica*, Vol. 92, No. 5 (September, 2024), 1503–1519


## EXACT BIAS CORRECTION FOR LINEAR ADJUSTMENT OF RANDOMIZED CONTROLLED TRIALS


HAOGE CHANG
*Department of Economics, Columbia University*

JOEL A. MIDDLETON
*Institution for Social and Policy Studies, Yale University*

P. M. ARONOW
*Departments of Statistics and Data Science, Political Science, Biostatistics, and Economics, Yale University*

*Econometrica*, Vol. 92, No. 5 (September, 2024), 1503–1519

# EXACT BIAS CORRECTION FOR LINEAR ADJUSTMENT OF RANDOMIZED CONTROLLED TRIALS

HAOGE CHANG
Department of Economics, Columbia University

JOEL A. MIDDLETON
Institution for Social and Policy Studies, Yale University

P. M. ARONOW
Departments of Statistics and Data Science, Political Science, Biostatistics, and Economics, Yale University

Freedman (2008a,b) showed that the linear regression estimator is biased for the analysis of randomized controlled trials under the randomization model. Under Freedman's assumptions, we derive exact closed-form bias corrections for the linear regression estimator. We show that the limiting distribution of the bias corrected estimator is identical to the uncorrected estimator. Taken together with results from Lin (2013), our results show that Freedman's theoretical arguments against the use of regression adjustment can be resolved with minor modifications to practice.

KEYWORDS: Randomized experiments, design-based model, regression adjustment.

## 1. INTRODUCTION

RANDOMIZED CONTROLLED TRIALS (RCTS) are increasingly popular in the social sciences. When estimating average treatment effects, adjustment for pretreatment covariates with linear regression is a common practice because it can reduce the variability of estimates. However, adjusting for covariates remains somewhat controversial, in large part because of Freedman (2008a,b).

Freedman argued that randomization does not justify the use of linear regression for completely randomized experiments. Freedman's theoretical arguments relied on three results under the randomization-based (Splawa-Neyman, Dabrowska, and Speed (1923), Imbens and Rubin (2015)) inferential paradigm:

1. asymptotically, the linear regression estimator can be inefficient relative to the unadjusted (difference-in-means) estimator if the design is imbalanced;
2. the classical homoscedastic standard error for linear regression is not valid asymptotically;
3. the regression estimator has an $O_p(n^{-1})$ bias term.

Freedman's third argument garnered attention among social scientists. For example, Deaton and Cartwright (2018)'s critique of randomization in empirical economics argued that the bias introduced by regression undermines the gold standard argument for RCTs.

In general, the literature has concluded that these issues are qualitatively small, at least relative to broader concerns about power and the quantification of uncertainties in RCTs. Using Freedman's framework, Lin (2013) showed that arguments 1 and 2 were resolved by small modifications to practice. Freedman's efficiency result may be addressed simply by including treatment by covariate interactions. Then it can be shown that the adjusted

Haoge Chang: hc3516@columbia.edu
Joel A. Middleton: joel.middleton@gmail.com
P. M. Aronow: p.aronow@yale.edu

 https://doi.org/10.3982/ECTA20289

estimator is never less asymptotically efficient than the unadjusted estimator. Regarding argument 2, Lin (2013) proved that robust standard errors (White (1980)) are asymptotically conservative in Freedman's setting, guaranteeing the validity of large-sample inference. On argument 3, Lin (2013) noted that the leading term of the bias is in fact estimable and can be shown to be small in a real-world empirical example. However, the small-sample bias of the regression estimator was not yet fully resolved.

Since Lin (2013), there have been notable papers that have proposed unbiased regression-type estimators for experimental data. Miratrix, Sekhon, and Yu (2013) demonstrated that if the regression model is fully saturated (see also Athey and Imbens (2017) and Imbens (2010)), then the associated effect estimate is unbiased conditional on the event that treatment is not collinear with any covariate stratum. This approach cannot generally be used without coarsening continuous covariates. In addition, Tan (2014) studied first-order bias corrections in the survey sampling setting. Lei and Ding (2021) and Chiang, Matsushita, and Otsu (2023) studied bias corrections in a setting where the number of covariates increases with the sample size.[1]

The primary contribution of this paper is to resolve Freedman's third argument by proposing finite-sample-exact, closed-form bias corrections. Our idea builds on Lin (2013)'s proposal to estimate the leading term of the bias, but further develops a finite-sample-exact bias correction encompassing all higher-order terms. We derive these bias corrections for both the noninteracted and interacted linear regression estimators. We prove that the estimators have the same limiting distributions as the non-bias-adjusted estimators.

Finally, we remind the readers that the practice of debiasing estimators is not uncontroversial. Tibshirani and Efron (1993) have warned that the bias correction could have costs in practice due to its high variability in finite samples.[2] In real-world decision-making processes, people may express different preferences for different statistical properties (i.e., unbiasedness or low Mean Squared Error). Our results shall imply that at least in large samples, the additional variation caused by the bias correction is negligible.

The organization of the paper is as follows: Section 2 includes the model setup and assumptions. Section 3 considers the characterization of bias terms of the OLS estimators and proposes bias correction for the noninteracted ATE estimators. Section 4 considers the case of the interacted estimators. In the Appendix, one can find proofs for the theorems.

## 2. SETTING, ASSUMPTIONS, AND NOTATIONS

We follow the setting of Freedman (2008a), Lin (2013), Abadie, Athey, Imbens, and Wooldridge (2020), and Lei and Ding (2021), which assume a Neyman model with covariates (Splawa-Neyman, Dabrowska, and Speed (1923)). There are $n$ subjects indexed by $i = 1, \dots, n$. For each subject, we observe an outcome $Y_i$ and a column vector of covariates $x_i = (x_{i1}, x_{i2}, \dots, x_{id})' \in \mathbb{R}^d$.

Each subject has two potential outcomes $y_i(1)$ and $y_i(0)$. We observe $Y_i = y_i(1)$ if subject $i$ is assigned to the treatment arm $T$, and $Y_i = y_i(0)$ if subject $i$ is assigned to the control arm $C$. Let $D_i$ be a binary variable, where $D_i = 1$ indicates that subject $i$ is assigned to the treatment arm.

[1]Compared with Lei and Ding (2021) and Chiang, Matsushita, and Otsu (2023), we add to the literature by proposing exactly unbiased estimators for both interacted and noninteracted estimators.

[2]We thank Winston Lin for suggesting this reference.

The experiment is assumed to be completely randomized, where $n_T$ out of $n$ subjects are randomly assigned to the treatment arm $T$ and the remaining $n_C = n - n_T$ subjects are assigned to the control arm $C$. Random assignment is the sole source of randomness in our statistical analysis. The potential outcomes and covariates are considered fixed, and the bias of an estimator is assessed relative to the randomization distribution of the estimator. We do not assume the existence of a superpopulation: the $n$ subjects are the population of interest.

We define $[T]$ as the set of subjects chosen for the treatment arm, given by $\{i|D_i = 1\}$, and similarly $[C]$ as the set of subjects chosen for the control arm, given by $\{i|D_i = 0\}$. For a possibly matrix-valued variable $a_i$, we use the notation $\bar{a} = \frac{1}{n}\sum_{i=1}^n a_i$ to represent the population average, $\bar{a}_T = \frac{1}{n_T}\sum_{i\in[T]} a_i$ to denote the treatment group average, and $\bar{a}_C = \frac{1}{n_C}\sum_{i\in[C]} a_i$ to denote the control group average. The average treatment effect (ATE) can be expressed in this notation as $\text{ATE} = \overline{y(1)} - \overline{y(0)}$, and the difference-in-means estimator is given by $\overline{Y}_T - \overline{Y}_C$. Similarly, we can write $\frac{1}{n}\sum_{i=1}^n x_i x_i' = \overline{xx'}$ for $x_i \in \mathbb{R}^d$ and $\frac{1}{n}\sum_{i=1}^n y_i(1)x_i = \overline{y(1)x}$ for $y_i(1) \in \mathbb{R}$ and $x_i \in \mathbb{R}^d$.

We make the following assumptions throughout the paper, which are standard in the literature.

ASSUMPTION 1: *For all $n$, there exists a finite constant $K$ such that*

$$\frac{1}{n}\sum_{i=1}^n y_i(1)^2 \leq K, \qquad \frac{1}{n}\sum_{i=1}^n y_i(0)^2 \leq K, \qquad \frac{1}{n}\sum_{i=1}^n x_{ik}^2 \leq K,$$

*for all $k = 1, \dots, d$ and $d \leq n$.*

ASSUMPTION 2: *For all $n$ large enough*, (1) $\overline{x} = 0$, *and* (2) $\overline{xx'} = I_d$, *the $d \times d$ identity matrix.*

ASSUMPTION 3: *Let $p_{T,n} = \frac{n_T}{n}$ and $p_{C,n} = \frac{n-n_T}{n}$ denote the inclusion probabilities into the treatment arm $T$ and control arm $C$, respectively. There exist positive constants $p_{\min}$ and $p_{\max}$ such that $0 < p_{\min} < p_{\max} < 1$ and $p_{\min} < p_{T,n} < p_{\max}$ for all $n$.*

These assumptions are employed regularly in the literature. They are used to derive consistency and the rate of convergence for the estimators below. Assumption 2 rules out perfect collinearity. For data sets that are not perfectly collinear, Assumption 2 is without loss of generality: in practice, researchers can just demean each covariate and orthonormalize the columns.[3] Assumption 3 requires each arm to receive a nontrivial fraction of subjects throughout the asymptotic sequence of the models. We shall hereafter omit the $n$ subscript in $p_{T,n}$ and $p_{C,n}$, and write $p_T$ and $p_C$ unless otherwise noted.

We define two regression-adjusted ATE estimators for reference. The first estimator results from a OLS regression:

$$Y_i \sim \alpha + \tau D_i + x_i'\beta, \tag{1}$$

[3]For example, assuming the covariate matrix has full column rank, one can use the procedure proposed in Lei and Ding (2021). We denote the SVD decomposition of the centered covariate matrix by $X = U\Sigma V \in \mathbb{R}^{n\times d}$, where $U \in \mathbb{R}^{n\times d}$, and $\Sigma, V \in \mathbb{R}^{d\times d}$. One can replace the covariate matrix $X$ with $\sqrt{n}U$.

where one regresses observed outcome $Y_i$ on the treatment indicator $D_i$ and covariates $x_i$. We denote the OLS estimators for this case as $(\widehat{\alpha}, \widehat{\tau}, \widehat{\beta})$. The estimator $\widehat{\tau}$ is hereafter referred to as the *noninteracted* ATE estimator.

The second estimator results from an interacted OLS regression where researchers expand the covariates by including terms interacting the treatment indicator and covariates:

$$Y_i \sim \alpha_I + \tau_I D_i + x_i' \beta_{I,C} + D_i x_i' \beta_{I,T}. \tag{2}$$

We denote the OLS estimators for the interacted case as $(\widehat{\alpha}_I, \widehat{\tau}_I, \widehat{\beta}_{I,C}, \widehat{\beta}_{I,T})$. The estimator $\widehat{\tau}_I$ is hereafter referred to as the *interacted* ATE estimator.

## 3. BIAS CHARACTERIZATION AND CORRECTION FOR THE NONINTERACTED CASE

We characterize the bias of the noninteracted ATE estimator and present a bias-corrected estimator in this section. Section 4 contains results for the interacted ATE estimator.

As shown in Lin (2013), the noninteracted ATE estimator can be written as

$$\widehat{\text{ATE}} = \overline{y(1)}_T - \overline{y(0)}_C - (\overline{x}_T' \widehat{\beta} - \overline{x}_C' \widehat{\beta}), \tag{3}$$

where $\widehat{\beta}$ is the OLS coefficient estimators for the covariates. The noninteracted ATE estimator can be written as a sum of the difference-in-means estimator, adjusted by group means and OLS coefficients. The bias can be viewed as coming from the regression adjustment term, particularly from estimating the coefficients for the covariates.

The coefficient estimators $\widehat{\beta}$ can be algebraically written as $\widehat{\beta} = \widehat{L}^{-1}\widehat{N}$, where $\widehat{L} = I_d - p_T \overline{x}_T \overline{x}_T' - p_C \overline{x}_C \overline{x}_C'$ and $\widehat{N} = p_T(\overline{y(1)x}_T - \overline{y(1)}_T \overline{x}_T) + p_C(\overline{y(0)x}_C - \overline{y(0)}_C \overline{x}_C)$ by an application of the Frisch–Waugh–Lovell theorem. The matrix $\widehat{L}$ consists of the variance-covariance matrix of the covariates and two additional stochastic terms that converge to 0 as the sample size increases. The vector $\widehat{N}$ is a weighted average of sample covariances between covariates and potential outcomes.

The OLS coefficient estimators can be thought of as estimating the (finite) population coefficients $\beta^* = L^{-1}N$, where $L = I_d$ and $N = p_T \overline{y(1)x} + p_C \overline{y(0)x}$. Note that $I_d$ does not involve unknowns, so in principle one does not need to estimate it with $\widehat{L}$. This view suggests that the randomness (bias) of $\widehat{L}$ can be entirely avoided if we replace $\widehat{L}$ with $L$ when estimating $\beta^*$. With this replacement, the regression adjustments in (3) can be written as $\overline{x}_T' \hat{N} - \overline{x}_C' \hat{N}$. The remaining bias can be characterized by analyzing quantities of forms $\overline{x}_T' \overline{y(1)x}_T$ and $\overline{x}_T' \overline{x}_T \overline{y(1)}_T$, which can be seen in the theorem below.[4]

Let $y_i^*(1)$ and $y_i^*(0)$ be the centered potential outcomes, that is, $y_i^*(1) = y_i(1) - \overline{y(1)}$ and $y_i^*(0) = y_i(0) - \overline{y(0)}$. Denote the (rescaled) leverage of the $i$th subject as $h_i = \|x_i\|_2^2$. As in Lei and Ding (2021), we define the maximum leverage as

$$\kappa = \max_{i=1,\dots,n} \frac{h_i}{n} = \max_{i=1,\dots,n} \frac{\|x_i\|_2^2}{n}. \tag{4}$$

Our first theorem characterizes the bias of the noninteracted ATE estimator.

[4]Note we shall hereafter assume for simplicity that all design matrices, $\widehat{L}$, are invertible. In the case of noninvertible design matrices, our debiased procedure will still work after choosing an arbitrary generalized inverse matrix and computing the ATE estimators accordingly.

THEOREM 3.1: *Under Assumptions* 1–3, *the OLS coefficient estimators for the covariates*, $\widehat{\beta}$, *can be decomposed as* $\widehat{\beta} = \beta^* + \nu_1 + \nu_2 + \nu_3$ *with*

$$\nu_1 = p_T\left(\overline{y^*(1)x}_T - \overline{y^*(1)x}\right) + p_C\left(\overline{y^*(0)x}_C - \overline{y^*(0)x}\right),$$
$$\nu_2 = \left(\widehat{L}^{-1} - I_d^{-1}\right)\widehat{N},$$
$$\nu_3 = -\left(p_T\overline{y^*(1)}_T\overline{x}_T + p_C\overline{y^*(0)}_C\overline{x}_C\right).$$

*The bias of the* $\widehat{\text{ATE}}$ *estimator is* $\mathbf{E}[(\overline{x}_C - \overline{x}_T)'(\nu_1 + \nu_2 + \nu_3)]$, *where*

$$\mathbf{E}\left[(\overline{x}_C - \overline{x}_T)'\nu_1\right] = \frac{1}{n-1}\left(\overline{hy(0)} - \overline{h} \times \overline{y(0)}\right) - \frac{1}{n-1}\left(\overline{hy(1)} - \overline{h} \times \overline{y(1)}\right) \tag{5}$$

*and*

$$\mathbf{E}\left[(\overline{x}_C - \overline{x}_T)'\nu_3\right] = \frac{n_C - n_T}{(n-1)(n-2)n_T}\left(\overline{hy(1)} - \overline{h} \times \overline{y(1)}\right) - \frac{n_T - n_C}{(n-1)(n-2)n_C}\left(\overline{hy(0)} - \overline{h} \times \overline{y(0)}\right). \tag{6}$$

*If* $\frac{d}{n} = o(1)$, *we have the stochastic expansion*

$$\widehat{\text{ATE}} = \text{ATE} + \frac{1}{n_T}\sum_{i\in[T]}\left(y_i(1) - \overline{y(1)} - x_i'\beta^*\right) - \frac{1}{n_C}\sum_{i\in[C]}\left(y_i(0) - \overline{y(0)} - x_i'\beta^*\right) + O_p\left(\sqrt{\frac{\kappa d}{n}}\right).$$

REMARK 1: The first bias term (5) is the main component of the bias. It is the scaled covariance between the leverage $h_i$ and individual effects $y_i(1) - y_i(0)$. Note that the formula suggests that the first bias term is 0 when the treatment effect is additive, for example, if there is no treatment effect on all subjects. However, a large bias may result from the presence of highly heterogeneous effects. The third term $\mathbf{E}[(\overline{x}_C - \overline{x}_T)'\nu_3]$ is 0 when $n_T = n_C = \frac{1}{2}n$. See also Freedman (2008b) and Lin (2013).

REMARK 2: The bias in $E[(\overline{x}_C - \overline{x}_T)'v_2]$ has been discussed previously. $(\overline{x}_C - \overline{x}_T)'v_2$ contains no unknowns and hence can be subtracted directly for debiasing purposes.

The first bias component (5) consists of covariances of leverages and outcomes. For intuition, consider a simple problem of estimating the average centered treated outcomes with a centered and standardized one-dimensional covariate.[5] Recall the definition of the constant $\overline{xy^*(1)} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i^*(1)$. We consider a simple adjusted estimator of the form

$$\frac{1}{n_T}\sum_{i\in[T]} y_i^*(1) - \overline{x}_T \times \left(\frac{1}{n_T}\sum_{i\in[T]} x_i y_i^*(1)\right)$$
$$= \frac{1}{n_T}\sum_{i\in[T]} y_i^*(1) - \overline{x}_T \times \overline{xy^*(1)} - \overline{x}_T \times \left(\frac{1}{n_T}\sum_{i\in[T]}\left(x_i y_i^*(1) - \overline{xy^*(1)}\right)\right).$$

[5]This example is only for illustration purposes as the average of the centered treatment outcomes is 0.

The first two terms of the equation above have expectation 0, and the third term can be further written as (up to a minus sign):

$$\frac{1}{n_T^2}\frac{n_C}{n-1}\sum_{i\in[T]} x_i\big(x_i y_i^*(1) - \overline{xy^*(1)}\big)$$
$$+\left(\frac{1}{n_T^2}\frac{n_T-1}{n-1}\sum_{i\in[T]} x_i\big(x_i y_i^*(1) - \overline{xy^*(1)}\big) + \frac{1}{n_T^2}\sum_{i,j\in[T],i\neq j} x_i\big(x_j y_j^*(1) - \overline{xy^*(1)}\big)\right).$$

The term in the parentheses has mean 0. The weightings in the parenthesized term reflect different values of $E[D_i]$ and $E[D_iD_j]$. The first term has expectation proportional to $\frac{1}{n}\sum_{i=1}^{n} x_i^2 y_i^*(1)$. In this one-dimensional covariate case (and up to a scale factor), $x_i^2$ is exactly the leverage. With multiple covariates, the leverage $h_i$ plays the role of $x_i^2$.

The third bias component (6) can also be shown to be proportional to $\frac{1}{n}\sum_{i=1}^{n} x_i^2 y_i^*(1)$. This is a direct result of the third-moment calculations in simple random sampling, where it can be shown that $\mathbf{E}[\overline{x_1}_T\overline{x_1}_T\overline{y^*(1)}_T]$ is proportional to $\frac{1}{n}\sum_{i=1}^{n} x_{i1}^2 y_i^*(1)$. For this example, we note that it is important for all three variables to have mean 0. For more details, see Lemma A.2 and also Finucan, Galbraith, and Stone (1974).

Finally, some may find it redundant to center the outcomes when characterizing the bias. However, this decomposition naturally yields an estimator of the bias that remains invariant to the location of the potential outcome distributions, as can be seen below. The bias estimator essentially consists of the estimators of the covariances of leverages and outcomes.

REMARK 3: Our current analysis accommodates the case where the number of covariates $d$ increases slowly with the sample size $n$, in the spirit of Lei and Ding (2021).[6] If, in addition, we assume d is fixed and with additional moment assumptions, it can be shown that the bias is of order $O(\frac{1}{n})$. See Lin (2013).

Our bias characterization leads to a formula for exact bias correction.[7]

THEOREM 3.2: *Under Assumption 2, an unbiased estimator for the bias of the noninteracted ATE estimator is*

$$\widehat{\text{Bias}} = \frac{1}{n-2}\big(\overline{hy(0)}_C - \overline{h}_C\overline{y(0)}_C\big) - \frac{1}{n-2}\big(\overline{hy(1)}_T - \overline{h}_T\overline{y(1)}_T\big) \tag{7}$$
$$+ (\overline{x}_C - \overline{x}_T)'\big(\widehat{L}^{-1} - I_d^{-1}\big)\widehat{N}. \tag{8}$$

[6]We note that Assumption 1 on treated and control outcomes implies constraints on the magnitude of the values of the associated coefficient vector (e.g., sparsity) when $d$ increases with $n$.

[7]We thank an anonymous referee for suggesting a simplification of our initial bias correction formula.

*The estimator* $\widehat{\text{ATE}}_{\text{Debiased}} = \widehat{\text{ATE}} - \widehat{\text{Bias}}$ *is unbiased for estimating the ATE. Under Assumptions* 1–3 *and if* $\kappa = o(1)$, *the estimator* $\widehat{\text{ATE}}_{\text{Debiased}}$ *has the asymptotic linear expansion*

$$\widehat{\text{ATE}}_{\text{Debiased}} = \text{ATE} + \frac{1}{n_T}\sum_{i\in[T]}\big(y_i(1) - \overline{y(1)} - x_i'\beta^*\big) - \frac{1}{n_C}\sum_{i\in[C]}\big(y_i(0) - \overline{y(0)} - x_i'\beta^*\big) + o_p\left(\sqrt{\frac{1}{n}}\right).$$

REMARK 4: Lei and Ding (2021) gave low-level conditions for $\kappa = o(1)$. In particular, $\kappa = o(1)$ if d is fixed and $\frac{1}{n}\sum_{k=1}^{d}\sum_{i=1}^{n} x_{ik}^4 = O(1)$. See also Lin (2013).

REMARK 5: There exist prior closed-form bias estimates in the literature, including Cochran (1977), Lin (2013), Tan (2014), and Lei and Ding (2021). Most bias correction estimators are for the interacted ATE estimator and involve estimating the covariances between regression residuals and leverages, as in Tan (2014) and Lei and Ding (2021). To our knowledge, none of the aforementioned papers provide an exactly unbiased correction for RCTs.

## 4. BIAS CHARACTERIZATION AND CORRECTION FOR THE INTERACTED CASE

We write the regression coefficient estimators of the pretreatment covariates in the interacted case as $\widehat{\beta}_{I,T} = \widehat{L}_T^{-1}\widehat{N}_T$ and $\widehat{\beta}_{I,C} = \widehat{L}_C^{-1}\widehat{N}_C$, and their (finite) population counterparts as $\beta_{I,T} = I_d^{-1}N_T$ and $\beta_{I,C} = I_d^{-1}N_C$, with $\widehat{L}_T = \overline{xx'}_T - \overline{x}_T\overline{x}_T'$, $\widehat{N}_T = \overline{y(1)x}_T - \overline{y(1)}_T\overline{x}_T$, $\widehat{L}_C = \overline{xx'}_C - \overline{x}_C\overline{x}_C'$, $\widehat{N}_C = \overline{y(0)x}_C - \overline{y(0)}_C\overline{x}_C$, $N_T = \overline{y(1)x}$, and $N_C = \overline{y(0)x}$.

As shown in Lin (2013), the OLS regression-adjusted ATE estimator can be written as

$$\widehat{\text{ATE}}_I = \overline{y(1)}_T - \overline{y(0)}_C - \big(\overline{x}_T'\widehat{\beta}_{I,T} - \overline{x}_C'\widehat{\beta}_{I,C}\big)$$

for the interacted case, where $\widehat{\beta}_{I,T}$ and $\widehat{\beta}_{I,C}$ are the OLS coefficients on covariates $D_i x_i$ and $x_i$, repsectively.

THEOREM 4.1: *Under Assumptions* 1–3, *the OLS coefficient vectors for the covariates of the interacted ATE estimator can be written as* $\widehat{\beta}_{I,T} = \beta_{I,T} + \nu_{1T} + \nu_{2T} + \nu_{3T}$, *and* $\widehat{\beta}_{I,C} = \beta_{I,C} + \nu_{1C} + \nu_{2C} + \nu_{3C}$, *with*

$$\nu_{1T} = \overline{y^*(1)x}_T - \overline{y^*(1)x}, \qquad \nu_{2T} = \big(\widehat{L}_T^{-1} - I_d^{-1}\big)\hat{N}_T, \qquad \nu_{3T} = -\big(\overline{x}_T\overline{y^*(1)}_T\big),$$
$$\nu_{1C} = \overline{y^*(0)x}_C - \overline{y^*(0)x}, \qquad \nu_{2C} = \big(\widehat{L}_C^{-1} - I_d^{-1}\big)\hat{N}_C, \qquad \nu_{3C} = -\big(\overline{x}_C\overline{y^*(0)}_C\big).$$

*The bias of the* $\widehat{\text{ATE}}_I$ *estimator is* $E[\widehat{\text{ATE}}_I - \text{ATE}] = E[\overline{x}_C'(\nu_{1C} + \nu_{2C} + \nu_{3C})] - E[\overline{x}_T'(\nu_{1T} + \nu_{2T} + \nu_{3T})]$. *We have*

$$E\big[\overline{x}_C'\nu_{1C}\big] - E\big[\overline{x}_T'\nu_{1T}\big] = \frac{n_T}{(n-1)n_C}\big(\overline{hy(0)} - \overline{h}\times\overline{y(0)}\big) - \frac{n_C}{(n-1)n_T}\big(\overline{hy(1)} - \overline{h}\times\overline{y(1)}\big),$$

$$E\big[\overline{x}_C'\nu_{3C}\big] - E\big[\overline{x}_T'\nu_{3T}\big] = \frac{n_T(n_T - n_C)}{(n-1)(n-2)n_C^2}\big(\overline{hy(0)} - \overline{h}\times\overline{y(0)}\big) - \frac{n_C(n_C - n_T)}{(n-1)(n-2)n_T^2}\big(\overline{hy(1)} - \overline{h}\times\overline{y(1)}\big).$$

*If* $\kappa \log d = o(1)$,[8]

$$\widehat{\mathrm{ATE}}_I = \mathrm{ATE} + \frac{1}{n_T}\sum_{i\in[T]}\big(y_i(1) - \overline{y(1)} - x_i'\beta_{I,T}\big) - \frac{1}{n_C}\sum_{i\in[C]}\big(y_i(0) - \overline{y(0)} - x_i'\beta_{I,C}\big) + O_p\left(\sqrt{\frac{\kappa d}{n}}\right).$$

THEOREM 4.2: *Under Assumption 2, an unbiased estimator for the bias of the interacted ATE estimator is*

$$\widehat{\mathrm{Bias}}_I = \frac{n_T}{(n-2)n_C}\big(\overline{hy(0)}_C - \overline{h}_C\overline{y(0)}_C\big) + \overline{x}_C'\big(\widehat{L}_C^{-1} - I_d^{-1}\big)\hat{N}_C - \frac{n_C}{(n-2)n_T}\big(\overline{hy(1)}_T - \overline{h}_T\overline{y(1)}_T\big) - \overline{x}_T'\big(\widehat{L}_T^{-1} - I_d^{-1}\big)\hat{N}_T.$$

*The estimator* $\widehat{\mathrm{ATE}}_{I,\mathrm{Debiased}} = \widehat{\mathrm{ATE}}_I - \widehat{\mathrm{Bias}}_I$ *is unbiased for estimating the ATE. Under Assumptions* 1–3 *and* $\kappa = o(1)$, *the estimator* $\widehat{\mathrm{ATE}}_{I,\mathrm{Debiased}}$ *has the asymptotic linear expansion*

$$\widehat{\mathrm{ATE}}_{I,\mathrm{Debiased}} = \overline{y(1)} - \overline{y(0)} + \frac{1}{n_T}\sum_{i\in[T]}\big(y_i(1) - \overline{y(1)} - x_i'\beta_{I,T}\big) - \frac{1}{n_C}\sum_{i\in[C]}\big(y_i(0) - \overline{y(0)} - x_i'\beta_{I,C}\big) + o_p\left(\sqrt{\frac{1}{n}}\right).$$

# APPENDIX A: PROOFS

## A.1. *Constants*

We define the following three constants:

$$N_{\mathrm{TTT}} = \frac{(n-n_T)(n-2n_T)}{(n-1)(n-2)n_T^2} = \frac{n_C(n_C-n_T)}{(n-1)(n-2)n_T^2},$$

$$N_{\mathrm{CCC}} = \frac{(n-n_C)(n-2n_C)}{(n-1)(n-2)n_C^2} = \frac{n_T(n_T-n_C)}{(n-1)(n-2)n_C^2},$$

$$N_{\mathrm{TTC}} = -\frac{(n-2n_T)}{(n-1)(n-2)n_T} = \frac{n_T-n_C}{(n-1)(n-2)n_T}.$$

## A.2. *Auxiliary Lemmas*

Let $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ and $b = (b_1, \dots, b_n) \in \mathbb{R}^n$ be arbitrary $n$-vectors with fixed elements. We define the covariance estimator for the treated group as $\widehat{\mathrm{Cov}}_T(a,b) = \overline{ab}_T - \overline{a}_T\overline{b}_T$, and similarly for the control group as $\widehat{\mathrm{Cov}}_C(a,b) = \overline{ab}_C - \overline{a}_C\overline{b}_C$. The following lemma characterizes the mean and variance of these covariance estimators.

[8]The upper bound on the stochastic order of the bias of the interacted case is characterized in equation (15) in Lei and Ding (2021).

LEMMA A.1: *Let* $t \in \{T, C\}$ *and* $n \geq 4$. *Then*,

$$\mathbf{E}\big[\widehat{\mathrm{Cov}}_t(a,b)\big] = \frac{(n_t-1)n}{n_t(n-1)}(\overline{ab} - \bar{a}\bar{b}),$$

$$\mathrm{Var}\big(\widehat{\mathrm{Cov}}_t(a,b)\big) \leq \frac{2(n-n_t)}{n_t(n-1)}\frac{1}{n}\sum_{i=1}^{n}(a_ib_i - \overline{ab})^2 + \frac{2(n-n_t)n}{n_t^3(n-1)}\frac{1}{n}\sum_{i=1}^{n}a_i^2 \times \frac{1}{n}\sum_{i=1}^{n}b_i^2.$$

PROOF: We will prove the case for the treated group. The proof for the control group case is analogous. To begin, we have

$$\mathbf{E}[\overline{ab}_T - \bar{a}_T\bar{b}_T] = \mathbf{E}\left[\frac{1}{2n_T^2}\sum_{i,j\in[T]}(a_i-a_j)(b_i-b_j)\right] = \frac{1}{2n_T^2}\frac{n_T(n_T-1)}{n(n-1)}\sum_{i\neq j}(a_i-a_j)(b_i-b_j)$$
$$= \frac{(n_T-1)n}{n_T(n-1)}\frac{1}{2n^2}\sum_{i,j}(a_i-a_j)(b_i-b_j) = \frac{(n_T-1)n}{n_T(n-1)}(\overline{ab} - \bar{a}\bar{b}).$$

To upper-bound the variance, we first note that the covariance estimator is location invariant and we can assume, without loss of generality, $\bar{a} = 0$ and $\bar{b} = 0$.[9] We then have

$$\mathrm{Var}(\overline{ab}_T - \bar{a}_T\bar{b}_T) \leq 2\,\mathrm{Var}(\overline{ab}_T) + 2\,\mathrm{Var}(\bar{a}_T\bar{b}_T) = 2\,\mathrm{Var}(\overline{ab}_T) + \frac{2}{n_T^4}\mathrm{Var}\Big(\sum_{i,j\in[T]}a_ib_j\Big)$$
$$\leq \frac{2(n-n_T)}{n_T(n-1)}\frac{1}{n}\sum_{i=1}^{n}(a_ib_i - \overline{ab})^2 + \frac{2(n-n_T)n}{n_T^3(n-1)}\frac{1}{n}\sum_{i=1}^{n}a_i^2\frac{1}{n}\sum_{i=1}^{n}b_i^2,$$

where the last inequality follows from the variance calculation of simple random samplings (first term) and Lemma A.5 in Lei and Ding (2021) (second term). *Q.E.D.*

Let $x_i$, $y_i$, and $z_i$ be three possibly identical vectors such that $\bar{x} = \bar{y} = \bar{z} = 0$.

LEMMA A.2: *For* $n \geq 3$,[10]

$$\mathbf{E}[\bar{x}_T\bar{y}_T\bar{z}_T] = N_{\mathrm{TTT}}\frac{1}{n}\sum_{i=1}^{n}x_iy_iz_i \quad \textit{and} \quad \mathbf{E}[\bar{x}_T\bar{y}_T\bar{z}_C] = N_{\mathrm{TTC}}\frac{1}{n}\sum_{i=1}^{n}x_iy_iz_i.$$

PROOF: We only prove the first equality. The second one can be proved analogously. First, notice two useful equalities:

$$\mathbf{E}\left[\sum_{i=1}^{n}D_ix_iy_i\sum_{j\neq i}D_jz_j\right] = \sum_{i=1}^{n}\sum_{j\neq i}[D_iD_j]x_iy_iz_j = \frac{n_T(n_T-1)}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}x_iy_iz_j$$
$$= \frac{n_T(n_T-1)}{n(n-1)}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}x_iy_iz_j - \sum_{i=1}^{n}x_iy_iz_i\right)$$

[9] This is required by the conditions of Lemma A.5 in Lei and Ding (2021).
[10] We note the following equalities can also be derived using results in Finucan, Galbraith, and Stone (1974).

$$= -\frac{n_T(n_T-1)}{n(n-1)}\sum_{i=1}^n x_i y_i z_i,$$

where the fourth equality uses the fact that $\sum_{i=1}^n z_i = 0$. Also,

$$\begin{aligned}
&\mathbf{E}\Bigg[\sum_{i=1}^n D_i x_i \sum_{j\neq i} D_j y_j \sum_{s\notin\{i,j\}} D_s z_s\Bigg] \\
&= \sum_{i=1}^n \sum_{j\neq i} \sum_{s\notin\{i,j\}} \mathbf{E}[D_i D_j D_s] x_i y_j z_s \\
&= \frac{n_T(n_T-1)(n_T-2)}{n(n-1)(n-2)} \sum_{i=1}^n \sum_{j\neq i} \sum_{s\notin\{i,j\}} x_i y_j z_s \\
&= \frac{n_T(n_T-1)(n_T-2)}{n(n-1)(n-2)} \Bigg(\sum_{i=1}^n \sum_{j\neq i} \sum_{s=1}^n x_i y_j z_s - \sum_{i=1}^n \sum_{j\neq i} x_i y_j (z_i + z_j)\Bigg) \\
&= \frac{n_T(n_T-1)(n_T-2)}{n(n-1)(n-2)} \Bigg(-\sum_{i=1}^n \sum_{j=1}^n x_i y_j (z_i + z_j) + 2\sum_{i=1}^n x_i y_i z_i\Bigg) \\
&= \frac{2n_T(n_T-1)(n_T-2)}{n(n-1)(n-2)} \sum_{i=1}^n x_i y_i z_i,
\end{aligned}$$

where the fourth and fifth equality use $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i = \sum_{i=1}^n z_i = 0$. Finally,

$$\begin{aligned}
\mathbf{E}[\bar{x}_T \bar{y}_T \bar{z}_T] &= \frac{1}{n_T^3}\Bigg(\mathbf{E}\Bigg[\sum_i D_i x_i y_i z_i\Bigg] + \mathbf{E}\Bigg[\sum_{i=1}^n \sum_{j\neq i} D_i D_j (x_i y_i z_j + x_i y_j z_i + x_j y_i z_i)\Bigg] \\
&\quad + \mathbf{E}\Bigg[\sum_{i=1}^N D_i x_i \sum_{j\neq i} D_j y_j \sum_{s\notin\{i,j\}} D_s z_s\Bigg]\Bigg) \\
&= \frac{1}{n_T^3}\Bigg(\frac{n_T}{n} - \frac{3n_T(n_T-1)}{n(n-1)} + \frac{2n_T(n_T-1)(n_T-2)}{n(n-1)(n-2)}\Bigg)\sum_{i=1}^n x_i y_i z_i,
\end{aligned}$$

where, for the last equality, we apply the previous two equalities. Simplifying the coefficients gives $\frac{1}{n}N_{\text{TTT}}$. *Q.E.D.*

## APPENDIX B: PROOF OF THE MAIN THEOREMS

### B.1. *Proof of Theorem 3.1*

By the Frisch–Waugh–Lovell theorem, the OLS estimate of the coefficient on the covariates can be written as $\widehat{\beta} = \widehat{L}^{-1}\widehat{N}$, where

$$\begin{aligned}
\widehat{N} &= p_T(\overline{y(1)x}_T - \overline{y(1)}_T \overline{x}_T) + p_C(\overline{y(0)x}_C - \overline{y(0)}_C \overline{x}_C) \\
&= p_T(\overline{y^*(1)x}_T - \overline{y^*(1)}_T \overline{x}_T) + p_C(\overline{y^*(0)x}_C - \overline{y^*(0)}_C \overline{x}_C)
\end{aligned}$$

$$= N + p_T(\overline{y^*(1)x}_T - \overline{y^*(1)x}) + p_C(\overline{y^*(0)x}_C - \overline{y^*(0)x}) - p_T\overline{y^*(1)}_T\overline{x}_T - p_C\overline{y^*(0)}_C\overline{x}_C$$

and $\widehat{L} = I_d - p_T\overline{x}_T\overline{x}_T' - p_C\overline{x}_C\overline{x}_C'$.

We have the following decomposition:

$$\widehat{\beta} = N + p_T(\overline{y^*(1)x}_T - \overline{y^*(1)x}) + p_C(\overline{y^*(0)x}_C - \overline{y^*(0)x}) - p_T\overline{y^*(1)}_T\overline{x}_T - p_C\overline{y^*(0)}_C\overline{x}_C + (\widehat{L}^{-1} - I_d^{-1})\widehat{N}.$$

As a result, $\mathbf{E}[\widehat{\text{ATE}} - \text{ATE}] = \mathbf{E}[(\overline{x}_C - \overline{x}_T)'\widehat{\beta}] = \mathbf{E}[(\overline{x}_C - \overline{x}_T)'\widehat{\beta} - (\overline{x}_C - \overline{x}_T)'\beta^*]$, where we used the fact $\mathbf{E}[\overline{x}_T] = \mathbf{E}[\overline{x}_C] = 0$. See the proof of Theorem 3.2 for the characterization of the bias.

B.1.1. *Stochastic Orders of the Bias of the Noninteracted ATE Estimator*

Let $(\alpha^*, \tau^*, \beta^*)$ be the minimizer of the criterion $p_T\sum_{i=1}^n(y_i(1) - \alpha - \tau - x_i'\beta)^2 + p_C\sum_{i=1}^n(y_i(0) - \alpha - x_i'\beta)^2$. Some calculation shows that

$$\begin{bmatrix}\alpha^*\\ \tau^*\\ \beta^*\end{bmatrix} = \Bigg(\underbrace{\begin{bmatrix}1 & p_T & 0_{1\times d}\\ p_T & p_T & 0_{1\times d}\\ 0_{d\times 1} & 0_{d\times 1} & I_d\end{bmatrix}}_{O}\Bigg)^{-1}\begin{bmatrix}p_T\overline{y(1)} + p_C\overline{y(0)}\\ p_T\overline{y(1)}\\ p_T\overline{y(1)x} + p_C\overline{y(0)x}\end{bmatrix}. \tag{9}$$

Define $e_i(1) = y_i(1) - \alpha^* - \tau^* - x_i'\beta^*$ and $e_i(0) = y_i(0) - \alpha^* - x_i'\beta^*$. Note $p_T\frac{1}{n}\sum_{i=1}^n(y_i(1) - \alpha^* - \tau^* - x_i'\beta^*)^2 + p_C\frac{1}{n}\sum_{i=1}^n(y_i(0) - \alpha^* - x_i'\beta^*)^2 \le p_T\frac{1}{n}\sum_{i=1}^n y_i^2(1) + p_C\frac{1}{n}\sum_{i=1}^n y_i(0)^2$ by definition. We thus have $\frac{1}{n}\sum_{i=1}^n e_i^2(1) = O(1)$ and $\frac{1}{n}\sum_{i=1}^n e_i^2(0) = O(1)$ by Assumption 1 and Assumption 3. We can represent the observed outcome as $Y_i = \alpha^* + \tau^* D_i + x_i'\beta^* + e_i(D_i)$. Some manipulation shows that the OLS estimator has the following representation:

$$\begin{bmatrix}\widehat{\alpha}\\ \widehat{\tau}\\ \widehat{\beta}\end{bmatrix} - \begin{bmatrix}\alpha^*\\ \tau^*\\ \beta^*\end{bmatrix} = \Bigg(\underbrace{\begin{bmatrix}1 & p_T & 0_{1\times d}\\ p_T & p_T & p_T\overline{x}_T'\\ 0_{d\times 1} & p_T\overline{x}_T & I_d\end{bmatrix}}_{\widehat{O}}\Bigg)^{-1}\frac{1}{n}\sum_{i=1}^n\begin{bmatrix}1\\ D_i\\ x_i\end{bmatrix}e_i(D_i). \tag{10}$$

Define $\tilde{x}_i = (1, D_i, x_i)'$ and $x_i^o = (0, 0, x_i)'$. We have the decomposition

$$(\overline{x}_T - \overline{x}_C)'(\widehat{\beta} - \beta^*) = \underbrace{(\overline{x}_T - \overline{x}_C)'\frac{1}{n}\sum_{i=1}^n x_i e_i(D_i)}_{(*)} + \underbrace{(\overline{x}_T^o - \overline{x}_C^o)'(\widehat{O}^{-1} - O^{-1})\frac{1}{n}\sum_{i=1}^n\tilde{x}_i e_i(D_i)}_{(**)}.$$

Note $\overline{x}_T - \overline{x}_C = \frac{1}{p_C}\overline{x}_T$. We bound the stochastic order of $(*)$:

$$\begin{aligned}
&\frac{1}{p_C}\overline{x}_T'\left(\frac{1}{n}\sum_{i\in[T]} x_i\big(e_i(1)-e_i(0)\big)+\frac{1}{n}\sum_{i=1}^{n} x_i e_i(0)\right)\\
&\quad=\frac{1}{p_C}\overline{x}_T'\left(\frac{1}{n}\sum_{i\in[T]}\left(x_i\big(e_i(1)-e_i(0)\big)+\frac{1}{p_T}\frac{1}{n}\sum_{i=1}^{n} x_i e_i(0)\right)\right))\\
&\quad=\frac{1}{p_C}\overline{x}_T'\left(\frac{1}{n}\sum_{i\in[T]}\left(x_i\big(e_i(1)-e_i(0)\big)-\frac{1}{n}\sum_{i=1}^{n} x_i\big(e_i(1)-e_i(0)\big)\right)\right),
\end{aligned}$$

where, for the second equality, we use the fact that $p_T\sum_{i=1}^{n} x_i e_i(1)+p_C\sum_{i=1}^{n} x_i e_i(0)=0$ implies $\sum_{i=1}^{n} x_i(e_i(1)-e_i(0)) = -\frac{1}{p_T}\sum_{i=1}^{n} x_i e_i(0)$.

By Lemma A.5 in Lei and Ding (2021), the first moment can be upper-bounded as

$$\begin{aligned}
\mathbf{E}\big[(*)\big] &= \mathbf{E}\left[\frac{1}{p_C}\overline{x}_T'\frac{1}{n}\sum_{i\in[T]} x_i\big(e_i(1)-e_i(0)\big)\right] = \frac{1}{p_C n n_T}\mathbf{E}\Big[\sum_{i\in[T]}\sum_{j\in[T]} x_i' x_j\big(e_j(1)-e_j(0)\big)\Big]\\
&= \frac{1}{p_C n n_T}\frac{n_T n_C}{n(n-1)}\sum_i \|x_i\|_2^2\big(e_i(1)-e_i(0)\big) = O\left(\frac{1}{n}\sum_{i=1}^{n} n^{-1}\|x_i\|_2^2\big(e_i(1)-e_i(0)\big)\right)\\
&\le O\left(\sqrt{\frac{\kappa d}{n}}\right),
\end{aligned}$$

by Assumption 1 and an argument similar to (15) in Lei and Ding (2021). By Lemma A.5 in Lei and Ding (2021), the variance can be upper-bounded as

$$\begin{aligned}
\mathrm{Var}\big[(*)\big] &\le \left(\frac{1}{p_C n_T n}\right)^2 \frac{n_T n_C}{n(n-1)}\sum_{i,j}\left(x_i'\left(x_j\big(e_j(1)-e_j(0)\big)-\frac{1}{n}\sum_{i=1}^{n} x_i\big(e_i(1)-e_i(0)\big)\right)\right)^2\\
&\le \left(\frac{1}{p_C n_T n}\right)^2 \frac{n_T n_C}{n(n-1)}\sum_{i,j}\big(x_i' x_j\big(e_j(1)-e_j(0)\big)\big)^2\\
&= \left(\frac{1}{p_C n_T n}\right)^2 n^2\frac{n_T n_C}{n(n-1)}\sum_i n^{-1}\|x_i\|_2^2\big(e_j(1)-e_j(0)\big)^2 = O\left(\frac{\kappa}{n}\right),
\end{aligned}$$

where, for the second inequality, we use the fact that, for each $x_i$, $\frac{1}{n}\sum_j(x_i'x_j(e_j(1) - e_j(0)))^2 \ge \frac{1}{n}\sum_j(x_i'(x_j(e_j(1)-e_j(0)) - \frac{1}{n}\sum_{j=1}^{n} x_j(e_j(1)-e_j(0))))^2$. For the second to last equality, we use the fact $\sum_{i,j}(x_i'x_j a_j)^2 = \sum_j a_j^2 x_j'(\sum_i x_i x_i')x_j = n\sum_j a_j^2\|x_j\|_2^2$ by Assumption 2. Thus, the term $(*)$ is of order $O_p(\sqrt{\frac{\kappa d}{n}})$.

Regarding the term $(**)$, we note the eigenvalues of the matrix $O$ are uniformly bounded above and below by Assumption 3. Let $|||\cdot|||_2$ be the matrix operator norm and

$\|\cdot\|_2$ be the Frobenius norm. We have $|||O-\hat{O}|||_2 \le \|O-\hat{O}\|_2 = \sqrt{2p_T^2\sum_{k=1}^{d}(\frac{1}{n_T}\sum_{i\in[T]} x_{ik})^2} = O_p(\sqrt{\frac{d}{n}})$ by Assumption 1. Thus, $|||O^{-1}-\hat{O}^{-1}|||_2 = O_p(\sqrt{\frac{d}{n}})$, if $\frac{d}{n}=o(1)$. A simple calculation shows that the term $(**)$ is of the order $o_p(\sqrt{\frac{\kappa d}{n}})$.

### B.2. *Proof of Theorem 3.2*

PROOF: We first propose estimators for the bias $\mathbf{E}[(\overline{x}_C-\overline{x}_T)(\nu_1+\nu_2+\nu_3)]$. Note, by Assumption 2, $(\overline{x}_C-\overline{x}_T)=\frac{1}{p_T}\overline{x}_C=-\frac{1}{p_C}\overline{x}_T$.

1. We first construct an estimator for $\mathbf{E}[(\overline{x}_C-\overline{x}_T)'\nu_1]$. First notice:

$$\begin{aligned}\mathbf{E}\big[(\overline{x}_C-\overline{x}_T)'\nu_1\big] &= \frac{1}{p_T}\mathbf{E}\big[\overline{x}_C'\big(p_T(\overline{y^*(1)x}_T-\overline{y^*(1)x})+p_C(\overline{y^*(0)x}_C-\overline{y^*(0)x})\big)\big]\\ &= \frac{1}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}x_i'x_iy_i^*(0)-\frac{1}{n}\sum_{i=1}^{n}x_i'x_iy_i^*(1)\right)\\ &= \frac{1}{n-1}(\overline{hy(0)}-\overline{h}\times\overline{y(0)})-\frac{1}{n-1}(\overline{hy(1)}-\overline{h}\times\overline{y(1)}),\end{aligned}$$

where the second equality follows from Proposition 1 from Freedman (2008b) and Assumption 2. An unbiased estimator of this expression is

$$\frac{1}{n-1}\left(\frac{n_C(n-1)}{(n_C-1)n}(\overline{hy(0)}_C-\overline{h}_C\overline{y(0)}_C)-\frac{n_T(n-1)}{(n_T-1)n}(\overline{hy(1)}_T-\overline{h}_T\overline{y(1)}_T)\right). \tag{11}$$

2. $(\overline{x}_C-\overline{x}_T)'\nu_2$ does not contain any unknown quantities, so it can be subtracted directly.
3. We now propose an estimator for $\mathbf{E}[(\overline{x}_C-\overline{x}_T)'\nu_3]$. First notice, by a term-wise application of Lemma A.2,

$$\begin{aligned}\mathbf{E}\big[(\overline{x}_C-\overline{x}_T)'\nu_3\big] &= \frac{1}{p_C}\mathbf{E}\big[\overline{x}_T'p_T\overline{y^*(1)}_T\overline{x}_T\big]+\frac{1}{p_C}\mathbf{E}\big[\overline{x}_T'p_C\overline{y^*(0)}_C\overline{x}_C\big]\\ &= \frac{1}{p_C}\mathbf{E}\big[\overline{x}_T'p_T\overline{y^*(1)}_T\overline{x}_T\big]-\frac{1}{p_C}\mathbf{E}\big[\overline{x}_T'p_T\overline{y^*(0)}_C\overline{x}_T\big]\\ &= \frac{p_T}{p_C}N_{\mathrm{TTT}}\frac{1}{n}\sum_{i=1}^{n}x_i'x_iy_i^*(1)-\frac{p_T}{p_C}N_{\mathrm{TTC}}\frac{1}{n}\sum_{i=1}^{n}x_i'x_iy_i^*(0)\\ &= \frac{p_T}{p_C}N_{\mathrm{TTT}}(\overline{hy(1)}-\overline{h}\times\overline{y(1)})-\frac{p_T}{p_C}N_{\mathrm{TTC}}(\overline{hy(0)}-\overline{h}\times\overline{y(0)}).\end{aligned}$$

Notice this bias is of order $O(\frac{\kappa}{n})$. An unbiased estimator for this quantity is

$$\begin{aligned}&\frac{N_{\mathrm{TTT}}n_T^2(n-1)}{n_C(n_T-1)n}(\overline{hy(1)}_T-\overline{h}_T\overline{y(1)}_T)\\ &\quad-\frac{N_{\mathrm{TTC}}n_T(n-1)}{(n_C-1)n}(\overline{hy(0)}_C-\overline{h}_C\overline{y(0)}_C).\end{aligned} \tag{12}$$

Collecting the constants in front of $\overline{hy(1)}_T - \overline{h}_T\overline{y(1)}_T$ gives

$$-\frac{1}{n-1}\frac{n_T(n-1)}{(n_T-1)n} + \frac{n_C(n_C-n_T)}{(n-1)(n-2)n_T^2}\frac{n_T^2(n-1)}{n_C(n_T-1)n} = \frac{n-n_Tn}{(n-2)(n_T-1)n} = -\frac{1}{n-2}.$$

Collecting the constants in front of $\overline{hy(0)}_C - \overline{hy(0)y(0)}_C$ gives

$$\frac{1}{n-1}\frac{n_C(n-1)}{(n_C-1)n} - \frac{(n_T-n_C)}{(n-1)(n-2)n_T}\frac{n_T(n-1)}{(n_C-1)n} = \frac{1}{n-2}.$$

Collecting terms gives the expression in the main text.

Now we derive the stochastic expansion for the debiased estimator:

$$\begin{aligned}
&\widehat{\mathrm{ATE}}_{\mathrm{Debiased}} \\
&= \big(\overline{y(1)}_T - \overline{y(0)}_C - (\overline{x}_T - \overline{x}_C)\widehat{L}^{-1}\hat{N}\big) - (\overline{x}_C - \overline{x}_T)'\big(\widehat{L}^{-1} - I_d^{-1}\big)\hat{N} \\
&\quad + \frac{1}{n-2}(\overline{hy(1)}_T - \overline{h}_T\overline{y(1)}_T) - \frac{1}{n-2}(\overline{hy(0)}_C - \overline{h}_C\overline{y(0)}_C) \\
&= \frac{1}{n_T}\sum_{i\in[T]}\big(y_i(1) - x_i'N\big) - \frac{1}{n_C}\sum_{i\in[T]}\big(y_i(0) - x_i'N\big) \\
&\quad \underbrace{- (\overline{x}_T - \overline{x}_C)'(\hat{N} - N) + \frac{1}{n-2}(\overline{hy(1)}_T - \overline{h}_T\overline{y(1)}_T) - \frac{1}{n-2}(\overline{hy(0)}_C - \overline{h}_C\overline{y(0)}_C)}_{(*)}.
\end{aligned}$$

We now bound the stochastic orders of the terms. The term $(*)$ can be decomposed as

$$\begin{aligned}
&(\overline{x}_C - \overline{x}_T)'\big(p_T(\overline{y(1)x}_T - \overline{y(1)}_T\overline{x}_T - \overline{y(1)x}) + p_C(\overline{y(0)x}_C - \overline{y(0)}_C\overline{x}_C - \overline{y(0)x})\big) \\
&\quad + \frac{1}{n-2}(\overline{hy(1)}_T - \overline{h}_T\overline{y(1)}_T) - \frac{1}{n-2}(\overline{hy(0)}_C - \overline{h}_C\overline{y(0)}_C) \\
&= -\frac{p_T}{p_C}\overline{x}_T'(\overline{y^*(1)x}_T - \overline{y^*(1)}_T\overline{x}_T - \overline{y^*(1)x}) + \frac{p_C}{p_T}\overline{x}_C'(\overline{y^*(0)x}_C - \overline{y^*(0)}_C\overline{x}_C - \overline{y^*(0)x}) \\
&\quad + \frac{1}{n-2}(\overline{hy(1)}_T - \overline{h}_T\overline{y(1)}_T) - \frac{1}{n-2}(\overline{hy(0)}_C - \overline{h}_C\overline{y(0)}_C) \\
&= -\frac{p_T}{p_C}\overline{x}_T'(\overline{y^*(1)x}_T - \overline{y^*(1)x}) + \frac{p_C}{p_T}\overline{x}_C'(\overline{y^*(0)x}_C - \overline{y^*(0)x}) - (11) \\
&\quad + \frac{p_T}{p_C}\overline{x}_T'\overline{y^*(1)}_T\overline{x}_T - \frac{p_C}{p_T}\overline{x}_C'\overline{y^*(0)}_C\overline{x}_C - (12) = O_p\left(\sqrt{\frac{\kappa}{n}}\right),
\end{aligned}$$

by the stochastic order estimates below and the fact that $\kappa \in [\frac{d}{n}, 1]$.

We now derive the stochastic order estimates for terms of the treated group. The stochastic order estimates for terms of the control group can be calculated analogously.

1. Since the first-order term in $\overline{x}_T'(\overline{y^*(1)x}_T - \overline{y^*(1)x})$ is canceled, we only need to characterize the variance of the term $\overline{x}_T'(\overline{y^*(1)x}_T - \overline{y^*(1)x})$. By Assumption 1, Assump-

tion 3, Lemma A.5, and (B.6) in Lei and Ding (2021), we have

$$
\begin{aligned}
&\operatorname{Var}\big(\overline{x}_T'\big(\overline{y^*(1)x}_T - \overline{y^*(1)x}\big)\big)\\
&= \frac{1}{n_T^4}\operatorname{Var}\Big(\sum_{i,j\in[T]} x_i'\big(x_j y_j^*(1) - \overline{y^*(1)x}\big)\Big)\\
&\le \frac{1}{n_T^4}\frac{n_T n_C}{n(n-1)}\sum_{i,j}\big(x_i'\big(x_j y_j^*(1) - \overline{y^*(1)x}\big)\big)^2 \le \frac{1}{n_T^4}\frac{n_T n_C}{n(n-1)}\sum_{i,j}\big(x_i' x_j y_j^*(1)\big)^2\\
&= \frac{n^2}{n_T^4}\frac{n_T n_C}{n(n-1)}\sum_{i,j}\big(n^{-1}x_i' x_j\big)^2 y_j^*(1)^2 = \frac{n^2}{n_T^4}\frac{n_T n_C}{n(n-1)}\sum_{j} n^{-1}\|x_j\|_2^2 y_j^*(1)^2\\
&= O\Big(\frac{\kappa}{n}\Big),
\end{aligned}
$$

where, for the second inequality, we use the fact that, for each $x_i$, $\frac{1}{n}\sum_j (x_i' x_j y_j^*(1) - x_i'\overline{y(1)x})^2 \le \frac{1}{n}\sum_j (x_i' x_j y_j^*(1))^2$.

2. Since the first-order term in (11) is canceled, we characterize the variance of the form $\frac{1}{n}(\overline{hy(1)}_T - \overline{h}_T\overline{y(1)}_T)$. First notice, by Lemma A.1 and Assumption 1,

$$
\begin{aligned}
&\operatorname{Var}\big[\big(\overline{hy(1)}_T - \overline{h}_T\overline{y(1)}_T\big)\big]\\
&\le \frac{2(n-n_T)}{n_T(n-1)}\frac{1}{n}\sum_{i=1}^n\big(h_i y_i(1) - \overline{hy(1)}\big)^2 + \frac{2(n-n_T)n}{n_T^3(n-1)}\frac{1}{n}\sum_{i=1}^n y_i^2(1)\frac{1}{n}\sum_{i=1}^n h_i^2\\
&\le \frac{2(n-n_T)}{n_T(n-1)}\frac{1}{n}\sum_{i=1}^n h_i^2 y_i^2(1) + \frac{2(n-n_T)n}{n_T^3(n-1)}\frac{1}{n}\sum_{i=1}^n y_i^2(1)\frac{1}{n}\sum_{i=1}^n h_i^2\\
&\le \frac{2(n-n_T)}{n_T(n-1)}n^2\kappa^2\frac{1}{n}\sum_{i=1}^n y_i^2(1) + \frac{2(n-n_T)n}{n_T^3(n-1)}\frac{1}{n}\sum_{i=1}^n y_i^2(1)\frac{1}{n}\sum_{i=1}^n h_i \times n\kappa\\
&= O\Big(\kappa^2 n + \frac{\kappa d}{n}\Big).
\end{aligned}
$$

Thus, $\operatorname{Var}(\frac{1}{n}(\overline{hy(1)}_T - \overline{h}_T\overline{y(1)}_T)) = O(\frac{\kappa^2}{n} + \frac{\kappa d}{n^3})$.

3. The stochastic order of the term $\overline{y^*(1)}_T\overline{x}_T'\overline{x}_T$ is $O_p(\frac{d}{n^{1.5}})$. The stochastic order of the term in (12) is $O(\frac{\kappa}{n} + \sqrt{\frac{\kappa^2}{n^3}} + \sqrt{\frac{\kappa d}{n^5}})$.

We find that the dominant term is of order $O_p(\sqrt{\frac{\kappa}{n}})$.

On Remark 4, note that $\kappa = \max_i \frac{\|x_i\|_2^2}{n} \le \frac{1}{n}\sqrt{\sum_{i=1}^n \|x_i\|_2^4} = O(\sqrt{\frac{d}{n}})$, as in Proposition 1 of Wu and Ding (2021). *Q.E.D.*

### B.3. *Proof of Theorem 4.1*

The decomposition of the bias is similar to the one in Theorem 3.1. We omit the details. The stochastic order of the bias is derived in Lei and Ding (2021).

### B.4. *Proof of Theorem 4.2*

We derive the unbiased estimator and the stochastic expansion for the treated group. We first propose estimators for the bias $E[-\overline{x}'_T(\nu_{1T}+\nu_{2T}+\nu_{3T})]$. Derivation and characterization for the control group are analogous:

$$\begin{aligned}&\mathbf{E}\big[-\overline{x}'_T(\nu_{1T}+\nu_{2T}+\nu_{3T})\big]\\&\quad=-\big(\mathbf{E}\big[\overline{x}'_T\big(\widehat{L}_T^{-1}-I_d^{-1}\big)\hat{N}_T\big]+\mathbf{E}\big[\overline{x}'_T\big(\overline{y^*(1)x}_T-\overline{y^*(1)x}\big)\big]-E\big[\overline{x}'_T\overline{x}_T\overline{y^*(1)}_T\big]\big).\end{aligned}$$

1. The term $\overline{x}'_T(\widehat{L}_T^{-1}-I_d^{-1})\hat{N}_T$ does not contain any unknown quantities, so it can be subtracted directly for debiasing.
2. An unbiased estimator for the second term $\mathbf{E}[\overline{x}'_T(\overline{y^*(1)x}_T-\overline{y^*(1)x})]$ can be derived in the same way as the first term in the noninteracted case, which gives
$$\frac{n_C}{n(n_T-1)}(\overline{hy(1)}_T-\overline{h}_T\overline{y(1)}_T).$$
3. An unbiased estimator for the third term $E[\overline{x}'_T\overline{x}_T\overline{y^*(1)}_T]$ can be derived in the same way as the third term in the noninteracted case, which gives
$$N_{\text{TTT}}\frac{n_T(n-1)}{(n_T-1)n}(\overline{hy(1)}_T-\overline{h}_T\overline{y(1)}_T).$$

Collecting the constants in front of $(\overline{hy(1)}_T-\overline{h}_T\overline{y(1)}_T)$ gives

$$\frac{n_C}{n(n_T-1)}-\frac{(n-n_T)(n-2n_T)}{(n-1)(n-2)n_T^2}\times\frac{n_T(n-1)}{(n_T-1)n}=\frac{n_C}{(n-2)n_T}.$$

Thus, an unbiased estimator for $\mathbf{E}[-\overline{x}'_T(\nu_{1T}+\nu_{2T}+\nu_{3T})]$ is

$$-\frac{n_C}{(n-2)n_T}(\overline{hy(1)}_T-\overline{h}_T\overline{y(1)}_T)-\overline{x}'_T\big(\widehat{L}_T^{-1}-I_d^{-1}\big)\hat{N}_T.$$

Similarly, an unbiased estimator for $\mathbf{E}[\overline{x}'_C(\nu_{1C}+\nu_{2C}+\nu_{3C})]$ is

$$\frac{n_T}{(n-2)n_C}(\overline{hy(0)}_C-\overline{h}_C\overline{y(0)}_C)+\overline{x}'_C\big(\widehat{L}_T^{-1}-I_d^{-1}\big)\hat{N}_C.$$

We can characterize the stochastic expansion for the debiased estimator of the treated group as

$$\begin{aligned}&\widehat{\text{ATE}}_{T,I,\text{Debiased}}\\&\quad=\overline{y(1)}_T-\overline{x}_T\widehat{L}_T^{-1}\hat{N}_T-\left(-\overline{x}'_T\big(\widehat{L}_T^{-1}-I_d^{-1}\big)\hat{N}_T-\frac{n_T}{(n-2)n_C}(\overline{hy(1)}_T-\overline{h}_T\overline{y(1)}_T)\right)\\&\quad=\overline{y(1)}_T-\overline{x}_TN_T-\overline{x}_T(\hat{N}_T-N_T)+\frac{n_T}{(n-2)n_C}(\overline{hy(1)}_T-\overline{h}_T\overline{y(1)}_T)\\&\quad=\overline{y(1)}_T-\overline{x}_TN_T+O_p\left(\sqrt{\frac{\kappa}{n}}\right),\end{aligned}$$

using the stochastic order estimates we derived in the proof of Theorem 4.1 and the fact that $\kappa \in [\frac{d}{n}, 1]$.

## REFERENCES


ABADIE, ALBERTO, SUSAN ATHEY, GUIDO W. IMBENS, AND JEFFREY M. WOOLDRIDGE (2020): "Sampling-Based versus Design-Based Uncertainty in Regression Analysis," *Econometrica*, 88 (1), 265–296. [1504]

ATHEY, SUSAN, AND GUIDO W. IMBENS (2017): "The Econometrics of Randomized Experiments," in *Handbook of Economic Field Experiments*, Vol. 1. Elsevier, 73–140. [1504]

CHIANG, HAROLD D., YUKITOSHI MATSUSHITA, AND TAISUKE OTSU (2023): "Regression Adjustment in Randomized Controlled Trials With Many Covariates," arXiv preprint. arXiv:2302.00469. [1504]

COCHRAN, WILLIAM G. (1977): *Sampling Techniques*. John Wiley & Sons. [1509]

DEATON, ANGUS, AND NANCY CARTWRIGHT (2018): "Understanding and Misunderstanding Randomized Controlled Trials," *Social Science & Medicine*, 210, 2–21. [1503]

FINUCAN, HENRY M., REX F. GALBRAITH, AND MERVYN STONE (1974): "Moments Without Tears in Simple Random Sampling From a Finite Population," *Biometrika*, 151–154. [1508,1511]

FREEDMAN, DAVID A. (2008)a: "On Regression Adjustments to Experimental Data," *Advances in Applied Mathematics*, 40 (2), 180–193. [1503,1504]

——— (2008)b: "On Regression Adjustments in Experiments With Several Treatments," *The Annals of Applied Statistics*, 2 (1), 176–196. [1503,1507,1515]

IMBENS, GUIDO W. (2010): "Better Late Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic literature*, 48 (2), 399–423. [1504]

IMBENS, GUIDO W., AND DONALD B. RUBIN (2015): *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press. [1503]

LEI, LIHUA, AND PENG DING (2021): "Regression Adjustment in Completely Randomized Experiments With a Diverging Number of Covariates," *Biometrika*, 108 (4), 815–828. [1504-1506,1508-1511,1514,1517]

LIN, WINSTON (2013): "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique," *The Annals of Applied Statistics*, 7 (1), 295–318. [1503,1504,1506-1509]

MIRATRIX, LUKE W., JASJEET S. SEKHON, AND BIN YU (2013): "Adjusting Treatment Effect Estimates by Post-Stratification in Randomized Experiments," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75 (2), 369–396. [1504]

SPLAWA-NEYMAN, JERZY, DOROTA M. DABROWSKA, AND TERRY. P. SPEED (1923): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," *Statistical Science*, 465–472. [1503,1504]

TAN, ZHIQIANG (2014): "Second-Order Asymptotic Theory for Calibration Estimators in Sampling and Missing-Data Problems," *Journal of Multivariate Analysis*, 131, 240–253. [1504,1509]

TIBSHIRANI, ROBERT J., AND BRADLEY EFRON (1993): "An Introduction to the Bootstrap," *Monographs on statistics and applied probability*, 57, 1–436. [1504]

WHITE, HALBERT (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 817–838. [1504]

WU, JASON, AND PENG DING (2021): "Randomization Tests for Weak Null Hypotheses in Randomized Experiments," *Journal of the American Statistical Association*, 116 (536), 1898–1913. [1517]


---

*Co-editor Guido W. Imbens handled this manuscript.*

*Manuscript received 2 November, 2021; final version accepted 11 May, 2024; available online 15 May, 2024.*