Model-Assisted Complier Average Treatment Effect Estimates in Randomized Experiments with Noncompliance

Jiyang Ren

Center for Statistical Science, Department of Industrial Engineering, Tsinghua University, Beijing, China

ABSTRACT

Noncompliance is a common problem in randomized experiments in various fields. Under certain assumptions, the complier average treatment effect is identifiable and equal to the ratio of the intention-to-treat effects of the potential outcomes to that of the treatment received. To improve the estimation efficiency, we propose three model-assisted estimators for the complier average treatment effect in randomized experiments with a binary outcome. We study their asymptotic properties, compare their efficiencies with that of the Wald estimator, and propose the Neyman-type conservative variance estimators to facilitate valid inferences. Moreover, we extend our methods and theory to estimate the multiplicative complier average treatment effect. Our analysis is randomization-based, allowing the working models to be misspecified. Finally, we conduct simulation studies to illustrate the advantages of the model-assisted methods and apply these analysis methods in a randomized experiment to evaluate the effect of academic services or incentives on academic performance.

ARTICLE HISTORY

Received May 2022
Accepted May 2023

KEYWORDS

Causalinference; Instrumental variable; Logistic regression; Oaxaca--Blinder estimator; Regression adjustment

1. Introduction

Randomized experiments are widely used to discover causality in social science, medical research, and many other fields. Under the Neyman–Rubin potential outcomes framework (Splawa-Neyman, Dabrowska, and Speed 1990; Rubin 1974), the average treatment effect can be identified and estimated without bias under random treatment assignment and the stable unit treatment value assumption (SUTVA) (Rubin 1980; Imbens and Rubin 2015).

However, in practice, some experimental units may not comply with their treatment assignments. For example, if someone is randomly assigned to enter a job-training program, he/she still has the right to be absent from the training. It is often difficult, or even impossible, to force experimental units to receive the assigned treatments. In this case, although the treatment assignment is randomized, the treatment received may not be. Noncompliance problems are common in randomized experiments in a variety of fields, such as clinical trials (Hirano et al. 2000), social experiments (Sherman and Berk 1984), policy evaluations (Schochet, McConnell, and Burghardt 2003), and educational studies (Susanne et al. 2005). Because the experimental units are assigned to only one of the treatment arms, we cannot observe their treatment received status under the alternative treatment arm. Thus, we do not know who the compliers are, and it is difficult to interpret the estimated treatment effect.

A highly successful application of the potential outcomes framework is to clarify the fundamental assumptions necessary to identify treatment effects in the noncompliance problems (Imbens and Angrist 1994; Angrist and Imbens 1995; Angrist, Imbens, and Rubin 1996; Imbens and Rubin 1997). Angrist, Imbens, and Rubin (1996) proposed a set of identification assumptions. Under these assumptions, the instrumental variable estimand could be interpreted as the local average treatment effect (LATE) for the subpopulation of compliers, also called the complier average treatment effect. Moreover, they showed that the LATE is equal to the ratio of intention-to-treat (ITT) effects of the potential outcomes to that of the potential treatment received. Because the ITT effects can be consistently estimated by the difference in the sample means of the potential outcomes (or potential treatment received) under the treatment and control groups, we can consistently estimate the LATE using a plug-in method, often known as the Wald estimator (Wald 1940).

Imbens and Angrist (1994) showed the consistency and the asymptotic normality of the Wald estimator. However, it does not incorporate covariate information, which is often collected for valid or more efficient causal inferences. In randomized experiments with perfect compliance, researchers have proposed regression adjustment methods with treatment assignment by covariate interactions to estimate the average treatment effect, and showed that the resulting estimator is generally more efficient than the simple difference-in-means estimator (Lin 2013; Bloniarz et al. 2016; Fogarty 2018; Yue et al. 2019; Liu and Yang 2020; Li and Ding 2020). For arbitrary potential outcomes, Guo and Basse (2023) proposed a generalized Oaxaca–Blinder...
estimator and showed its consistency and asymptotic normality under the conditions of prediction unbiasedness, stability, and simple prediction models. To guarantee the efficiency gains compared with the unadjusted estimator, Cohen and Fogarty (2022) calibrated the Oaxaca–Blinder estimator on the basis of Guo and Basse (2023).

When there is noncompliance, the most popular method in econometrics to incorporate covariate information in LATE estimation is the two-stage least squares (2SLS), using the simultaneous equations model (SEM) (Angrist and Pischke 2008; Wooldridge 2010). Moreover, under a fully parametric model specification, maximum likelihood and Bayesian inferential methods are also used to estimate the LATE (Imbens and Rubin 1997; Hirano et al. 2000). However, these methods lack validity if the working models are misspecified. Semiparametric approaches have been proposed for robust estimation (Abadie 2003; Tan 2006; Clarke and Windmeijer 2012; Ogburn, Rotnitzky, and Robins 2015; Wang et al. 2021). To further reduce the risk of model misspecification, Frölich (2007) proposed a nonparametric imputation-based LATE estimator, showed its asymptotic normality, and provided a semiparametric efficiency bound. Moreover, many researchers favor inverse probability weighting methods, some of which share the doubly robust properties (Tan 2006; Frölich 2007; Donald, Hsu, and Lieli 2014; Heiler 2022; Sun and Tan 2022). All of these studies assumed that the observed data, outcomes, covariates, treatment received, and treatment assignments, are independent and identically distributed, and sampled from a large super-population.

Super-population framework, based on the perspective of random sampling, considers the randomness of not only the treatment assignment but also the potential outcomes and covariates. It is convenient for theoretical analysis and widely used in causal inference literature, but is unnatural in some problems (Abadie et al. 2020; Guo and Basse 2023). For example, Gerber and Green (2000) conducted a field experiment in New Haven, Connecticut to evaluate the effect of personal canvassing, telephone calls, and direct mail appeals on voter turnout. All households in New Haven with one or two registered voters were randomly assigned to the treatment or control group. In this example, there is no random sampling procedure and the only source of randomness is the treatment assignment. This motivates us to consider the design-based inference, also known as the finite-population framework, under which the source of uncertainty arises from the random treatment assignment, not the random sampling. It can also be viewed as a conditional inference method under the super-population framework, that is, drawing causal inference conditional on the potential outcomes and covariates. The design-based inference is agnostic — the validity of the resulting estimates and inference procedures does not require correct specification of the underlying outcome model. It has a long history dating back to Fisher and Neyman, and has become increasingly popular in both theory and practice (Fisher 1935; Neyman, Iwaszkiewicz, and Kolodziejczyk 1935; Robins 1988; Freedman 2008; Lin 2013; Imbens and Rubin 2015; Athey and Imbens 2017; Fogarty 2018; Guo and Basse 2023).

The design-based inference of the noncompliance problem receives attention recently (Kang, Peck, and Keele 2018; Rambachan and Roth 2020). Kang, Peck, and Keele (2018) proposed the almost exact method, which is a non-asymptotic inference method for LATE in completely randomized experiments. Rambachan and Roth (2020) showed the asymptotic normality of the Wald estimator under a Poisson rejective assignment mechanism, of which the completely randomized assignment mechanism can be interpreted as a special case.

Observing that the Wald estimator is the ratio of the two simple difference-in-means estimators for the ITT effects, it equals the ratio of a reduced-form difference-in-means estimator and a first-stage difference-in-means estimator after minus the true LATE. Then, we obtain its design-based asymptotic behavior by using the classic finite-population central limit theorem (Li and Ding 2017) and Slutsky’s theorem. More importantly, it naturally motivates us to improve the LATE estimation efficiency by adjusting the difference-in-means estimators for the ITT effects. Specifically, we propose three model-assisted estimators, indirect least squares (ILS), Logistic Oaxaca–Blinder (OB), and calibrated Oaxaca–Blinder (COB) to estimate LATE more efficiently, utilizing the ideas of Lin (2013), Guo and Basse (2023), and Cohen and Fogarty (2022) in randomized experiments with perfect compliance. We study their consistency and asymptotic normality based on existing results in Lin (2013), Li and Ding (2017), Guo and Basse (2023), Cohen and Fogarty (2022), and Slutsky’s theorem. We compare their efficiencies with that of the Wald estimator, and find that the ILS estimator and COB estimator can always improve (at least never hurt) the estimation efficiency. Furthermore, we provide conservative variance estimators to construct asymptotically conservative confidence intervals for the LATE. We also extend the proposed methods to estimate and infer the multiplicative local average treatment effect (MLATE) in the supplementary materials. In our theoretical analysis, we allow the working models to be arbitrarily misspecified. Finally, we evaluate the finite-sample performance of the proposed estimators through simulation studies and a real data application.

2. Framework, Notation, and Assumptions

Consider a completely randomized experiment with n units, of which n1 units are randomly assigned to the treatment group and the remaining n0 units to the control group. We denote by i the treatment assignment indicator. For unit i, Zi = 1 if it is assigned to the treatment group, and Zi = 0 otherwise. The probability distribution of the treatment assignment indicator is pr((Zi1, . . . , Zin)pT = (z1, . . . , zn)pT | n1 = n1) = n1/n!n/z1 = 0, 1. Let i(z) denote the binary potential treatment received status. If unit i is assigned to the treatment arm z, z = 0, 1, D(z) = 1 if it receives the active treatment and D(z) = 0 if it receives the control treatment. The observed treatment received is D(z) = Zi1D1(1) + (1−Zi)D0(0). According to the joint values of the potential treatment received (D0, D1), the experimental units can be divided into four strata: always takers with (1, 1), never takers with (0, 0), compliers with (0, 1), and defiers with (1, 0).

Let Y(z, d) denote the binary potential outcomes if unit i is assigned to treatment arm z and receives treatment arm d, z = 0, 1, d = 0, 1. Let Y(z) denote the potential value of outcome that would be observed when unit i is assigned to the treatment arm z, and d is set to the value that naturally occurs,
that is, \( Y_{i}(z) = Y_{i}(z, D_{i}(z)) \). The observed response value is 
\( Y_{i}^{\text{obs}} = Z_{i} Y_{i}(1) + (1 - Z_{i}) Y_{i}(0) \). The average treatment effect of \( Z \) on \( Y \), also called the ITT effect, is 
\( \tau_{Y} = \frac{1}{n} \sum_{i=1}^{n} (Y_{i}(1) - Y_{i}(0)) / n \). Similarly, the ITT effect of \( Z \) on \( D \) is 
\( \tau_{D} = \frac{1}{n} \sum_{i=1}^{n} (D_{i}(1) - D_{i}(0)) / n \).

Many studies have investigated the treatment effect of the treatment received on the response, that is, the treatment effect of \( D \) on \( Y \) instead of \( Z \) on \( Y \). For example, if we want to evaluate the effect of a drug, we need to compare the outcomes between patients taking the drug and not taking the drug. The ITT effect \( \tau_{Y} \) considers all of the experimental units, including patients who always take the drug and those who never take the drug. Only for compliers does the treatment received status equal the treatment assigned status. Thus, we prefer to study the LATE, which is \( \tau = \frac{1}{n} \sum_{i\in C} (Y_{i}(1) - Y_{i}(0)) / n \), where \( C = \{ i : D_{i}(1) = 1, D_{i}(0) = 0 \} \) is the set of compliers, and \( n_{c} \) is the total number of compliers.

As only one of the potential treatment received, \( D_{i}(1) \) or \( D_{i}(0) \), can be observed, we do not know who the compliers are. Thus, to determine the LATE, we use the following identification assumptions proposed by Angrist, Imbens, and Rubin (1996).

**Condition 1.** (i) Exclusion restriction: \( Y_{i}(1, d) = Y_{i}(0, d) \) for \( d = 0, 1 \); (ii) Monotonicity: \( D_{i}(1) \geq D_{i}(0) \); (iii) Strong instrument: \( \tau_{D} > C_{0} > 0 \), where \( C_{0} \) is a positive constant independent of \( n \).

**Remark 1.** First, the exclusion restriction implies that the treatment assignment \( Z \) affects the potential outcome \( Y \) only through the treatment received \( D \). Thus, \( Y_{i}(1, D_{i}(1)) = Y_{i}(1, D_{i}(0)) = Y_{i}(0) \) for all units with \( D_{i}(1) = D_{i}(0) \). Therefore, for always takers and never takers, the treatment effect of \( Z \) on \( Y \) is 0. Second, the monotonicity assumption rules out the defiers. Third, the strong instrument assumption implies that the ITT effect of \( Z \) on \( D \) is not equal to 0.

Under **Condition 1**, Angrist, Imbens, and Rubin (1996) showed that the LATE is equal to the ratio of the ITT effect of \( Z \) on \( Y \) to that of \( Z \) on \( D \), denoted by \( \tau = \tau_{Y} / \tau_{D} \). This identification formula has an intuitive explanation. The average treatment effect of \( Z \) on \( Y \) in the entire finite-population can be decomposed into the summation of the multiplication of the LATE and the proportion of its corresponding subpopulation. Under **Condition 1**, the LATE for always takers and never takers are 0, and the proportion of defiers is 0. Thus, the ITT effect of \( Z \) on \( Y \) equals to the LATE multiplied by the proportion of compliers. The reason is as follows (Angrist, Imbens, and Rubin 1996): The ITT effect of \( Z \) on \( Y \) can be decomposed as

\[
\frac{1}{n} \sum_{i=1}^{n} (Y_{i}(1) - Y_{i}(0)) = \frac{1}{n} \sum_{i\in C} (Y_{i}(1) - Y_{i}(0)) + \frac{1}{n} \sum_{i\notin C} (Y_{i}(1) - Y_{i}(0))
\]

\[
= \frac{1}{n} \sum_{i\in C} (Y_{i}(1) - Y_{i}(0))
\]

\[
= \frac{n_{c}}{n} \times \frac{1}{n} \sum_{i\in C} (Y_{i}(1) - Y_{i}(0))
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} (D_{i}(1) - D_{i}(0)) \times \tau.
\]

Then \( \tau = \frac{1}{n} \sum_{i=1}^{n} (Y_{i}(1) - Y_{i}(0)) / [\frac{1}{n} \sum_{i=1}^{n} (D_{i}(1) - D_{i}(0))] = \tau_{Y} / \tau_{D} \).

To proceed, we introduce the following notation. Let \( X_{i}^{*} = (X_{i1}, \ldots, X_{ip})^{T} \) be a p-dimensional vector of baseline covariates for unit \( i \). As the baseline covariates are pretreatment and not affected by the treatment assignment, we have \( X_{i}^{\text{obs}} = X_{i} \). For \( R = D \) (and \( R = Y \)), the average values of the treatment received (and potential outcomes) in the finite-population, the treatment group, and the control group are

\[
\hat{R}(z) = n^{-1} \sum_{i=1}^{n} R_{i}(z), \quad \hat{R}_{1}^{\text{obs}} = n_{1}^{-1} \sum_{i=1}^{n} Z_{i} r_{1}^{\text{obs}},
\]

\[
\hat{R}_{0}^{\text{obs}} = n_{0}^{-1} \sum_{i=1}^{n} (1 - Z_{i}) r_{0}^{\text{obs}},
\]

where * indicates the finite-population mean, and the additional subscripts 1 and 0 indicate sample means in the treatment and control groups, respectively. Similarly, define

\[
\hat{X} = n^{-1} \sum_{i=1}^{n} X_{i}, \quad \hat{X}_{1}^{\text{obs}} = n_{1}^{-1} \sum_{i=1}^{n} Z_{i} X_{i},
\]

\[
\hat{X}_{0}^{\text{obs}} = n_{0}^{-1} \sum_{i=1}^{n} (1 - Z_{i}) X_{i}.
\]

Let \( R_{i}(z) \) and \( Q_{i}(z) \) be any potential outcomes of unit \( i \), for example, \( R = Y \) and \( Q = D \). For each treatment arm \( z \), \( z = 0, 1 \), let \( S_{R(z)}^{2} \) and \( S_{R(1) - R(0)}^{2} \) be the finite-population variances of \( R_{i}(z) \) and the unit-level treatment effect \( R_{i}(1) - R_{i}(0) \), respectively. Let \( S_{R(z), Q(z)}^{2} \) and \( S_{R(1) - R(0), Q(1) - Q(0)}^{2} \) be the finite-population covariances between \( R_{i}(z) \) and \( Q_{i}(z) \) and between their unit-level treatment effects, respectively. We replace the uppercase \( S \) with a lowercase \( s \) to denote the sample analog.

### 3. Wald Estimator

As the treatment assignment is completely randomized, the ITT effects \( \tau_{Y} \) and \( \tau_{D} \) can be estimated without bias by the difference in the means of the observed outcomes under the treatment and control groups. The difference-in-means estimators are \( \hat{\tau}_{Y} = \hat{Y}_{1}^{\text{obs}} - \hat{Y}_{0}^{\text{obs}} \) and \( \hat{\tau}_{D} = \hat{D}_{1}^{\text{obs}} - \hat{D}_{0}^{\text{obs}} \). As \( \tau = \tau_{Y} / \tau_{D} \), the Wald estimator for \( \tau \) is to plug in the estimators for \( \tau_{Y} \) and \( \tau_{D} \). That is, \( \hat{\tau}_{\text{wald}} = \hat{\tau}_{Y} / \hat{\tau}_{D} \).

Under the SEM framework, which is widely used in econometrics, the Wald estimator is equivalent to the 2SLS estimator without covariates (Angrist and Pischke 2008). The 2SLS method by definition consists of two regression stages, that is, fitting a linear regression of \( D \) on \( Z \) and then fitting a linear regression of \( Y \) on the fitted values of \( D \):

\[
P_{i}^{\text{obs}} = \beta_{0} + \beta_{1} Z_{i} + e_{D,i}, \quad \hat{Y}_{i}^{\text{obs}} = \beta_{0} + \beta_{2}\hat{D}_{i} + e_{Y,i},
\]

where \( \hat{D}_{i} \) is the fitted value in the first regression. The 2SLS estimator is the estimator for the second-stage regression coefficient of \( \hat{D}_{i} \), \( \beta_{2}\text{SLS} \). The 2SLS has been investigated extensively in the econometric literature (Angrist and Pischke 2008; Wooldridge 2010), under a super-population framework. In what follows, we study its asymptotic properties under the design-based inference using the finite-population central limit theorem (Li and Ding 2017).
As \( \hat{\tau}_{\text{wald}} - \tau = (\hat{\tau}_Y - \tau \hat{D}_\text{wald}) / \hat{\sigma}_{\text{wald}} \), under mild conditions, \( \hat{\tau}_{\text{wald}} \) has the same asymptotic distribution as \((\hat{\tau}_Y - \tau \hat{D}) / \hat{\sigma}_{\text{wald}} = \hat{\tau}_A / \hat{\sigma}_D \), where \( \hat{\tau}_A \) is the difference-in-means estimator for the ITT effect of \( Z \) on the transformed potential outcomes \( A: A_i(z) = Y_i(z) - \tau D_i(z), z = 0, 1 \). The asymptotic variance of \( n^{1/2} \hat{\tau}_{\text{wald}} \) is the limit of \( \sigma_{\text{wald}}^2 \):

\[
\sigma_{\text{wald}}^2 = \frac{n}{D} \left( \frac{S_A(1)}{n_1} + \frac{S_A(0)}{n_0} - \frac{S_A(1) - A(0)}{n} \right).
\]

**Remark 2.** Under the super-population framework, one assumes that \( \mathcal{F} = \{Y_i(z, d), D_i(z)\} \) are independent and identically distributed samples drawn from a super-population. Then we have

\[
\text{var}(\hat{\tau}_{\text{wald}}) = E[\text{var}(\hat{\tau}_{\text{wald}} | \mathcal{F})] + \text{var}(E(\hat{\tau}_{\text{wald}} | \mathcal{F})) \\
\geq E[\text{var}(\hat{\tau}_{\text{wald}} | \mathcal{F})].
\]

That is, the variance under the super-population framework is generally greater than that under the design-based inference.

As we cannot observe or estimate \( A_i(1) \) and \( A_i(0) \) simultaneously, the last term in \( \sigma_{\text{wald}}^2 \) is not estimable. We directly drop it and obtain a Neyman-type conservative variance estimator:

\[
\hat{\sigma}_{\text{wald}}^2 = \frac{n_D}{D} \left( \frac{S_A(1)}{n_1} + \frac{S_A(0)}{n_0} \right),
\]

where \( S_A(z) = (n_z - 1)^{-1} \sum_{i \in Z = z} (\hat{A}_i(z) - n_z^{-1} \sum_{\tilde{Z} = z} \hat{A}_z)^2 \) is the sample variance for the estimated outcome \( \hat{A}_i(z) = Y_i(z) - \hat{\tau}_{\text{wald}} D_i(z), z = 0, 1 \). (A more rigorous notation for \( S_A(z) \) is \( S_A^2(z) \). For ease of notation, we still use \( S_A(z) \) when it does not confuse.)

To establish the asymptotic normality of \( \hat{\tau}_{\text{wald}} \), the following regularity conditions are required by the finite-population central limit theorem (Li and Ding 2017).

**Condition 2.** As \( n \to \infty \), (i) the proportions of the treated and control units have limits between 0 and 1, that is, \( n_1 / n \to p_1 \in (0, 1) \) and \( n_0 / n \to p_0 \in (0, 1) \); (ii) for \( z = 0, 1 \), the finite-population means, \( Y(z) \) and \( D(z) \), the finite-population variances, \( \hat{S}_Y(z), \hat{S}_D(z), \hat{S}_{Y(z) - Y(0)}, \) and the finite-population covariances, \( \hat{S}_{Y(z) D(z)} \) and \( \hat{S}_{Y(z) - Y(0) D(z) - D(0)} \), tend to finite limits, and the limit of \( \sigma_{\text{wald}}^2 \) is positive.

**Proposition 1.** Under Conditions 1 and 2, \( \hat{\tau}_{\text{wald}} - \tau \) converges in probability to 0 and \( n^{1/2} (\hat{\tau}_{\text{wald}} - \tau) / \sigma_{\text{wald}} \) converges in distribution to \( N(0, 1) \). Furthermore, \( \hat{\sigma}_{\text{wald}}^2 \) converges in probability to the limit of \( n_D \sigma_{\text{wald}}^2 \), which is no less than that of \( \sigma_{\text{wald}}^2 \).

**Proposition 1** provides a normal approximation for the distribution of \( \hat{\tau}_{\text{wald}} \) in completely randomized experiments, which can be interpreted as a special case under a Poisson reactive assignment mechanism studied by Rambachan and Roth (2020).

Unlike Rambachan and Roth (2020), we provide a proof of **Proposition 1** by directly applying the finite-population central limit theorem (Li and Ding 2017) under complete randomization to the ITT effect estimators and Slutsky’s theorem.

The variance estimator is generally conservative. It is consistent if and only if the unit-level treatment effect of \( A_i(z) \) is constant, that is, \( Y_i(1) - Y_i(0) - \tau (D_i(1) - D_i(0)) = C \) for some constant \( C \). Based on **Proposition 1**, an asymptotically conservative confidence interval for \( \tau \) is \( [\hat{\tau}_{\text{wald}} - q_{\alpha/2} n^{-1/2} \hat{\sigma}_{\text{wald}}, \hat{\tau}_{\text{wald}} + q_{\alpha/2} n^{-1/2} \hat{\sigma}_{\text{wald}}] \), where \( \alpha \in (0, 1) \) is the significance level and \( q_{\alpha/2} \) is the upper \( \alpha/2 \) quantile of a standard normal distribution. The asymptotic coverage rate of the above confidence interval is greater than or equal to \( 1 - \alpha \).

### 4. Model-assisted LATE Estimators

Although the Wald estimator exhibits good asymptotic properties, it does not use covariate information. If these baseline covariates can predict the potential outcomes, covariate adjustment tends to reduce the variance of the estimated treatment effect (Lin 2013; Bloniarz et al. 2016; Fogarty 2018; Yue et al. 2019; Liu and Yang 2020; Li and Ding 2020). In this section, we propose three model-assisted LATE estimators and study their asymptotic properties under the design-based inference. The working models used to obtain these estimators can be arbitrarily misspecified.

#### 4.1. Indirect Least Squares Estimator with Interactions

In econometrics, the ILS and 2SLS are two widely used methods to estimate the LATE under the SEM framework (Hayashi 2000; Angrist and Pischke 2008; Wooldridge 2010). The ILS regresses the observed potential outcomes \( Y_{i1}^{\text{obs}} \) and \( Y_{i2}^{\text{obs}} \) on the treatment assigned status \( Z_i \) and covariates \( \mathbf{X}_i \) as:

\[
Y_{i1}^{\text{obs}} = \beta_0^1 + \beta_1 Z_i + \mathbf{X}_i^\top \alpha_1 + e_{i1}, \\
Y_{i2}^{\text{obs}} = \beta_0^2 + \beta_1 Z_i + \mathbf{X}_i^\top \alpha_2 + e_{i2},
\]

The ILS estimator estimates the ratio of two regression coefficients of \( Z_i \), \( \beta_{i1} = \beta_1 / \alpha_1 \) (Angrist and Pischke 2008). The 2SLS method, using information from the covariates, performs the following two regressions:

\[
Y_{i1}^{\text{obs}} = \beta_0 + \beta_1 Z_i + \mathbf{X}_i^\top \mathbf{\eta}_1 + e_{i1}, \\
Y_{i2}^{\text{obs}} = \beta_0 + \beta_2 \mathbf{D}_i + \mathbf{X}_i^\top \mathbf{\eta}_2 + e_{i2},
\]

where \( \mathbf{D}_i \) is the fitted value in the first regression. The 2SLS estimator estimates the second stage regression coefficient of \( \mathbf{D}_i \), \( \beta_{2\text{SLS}} = \text{cov}(Y, \mathbf{Z}) / \text{cov}(D, \mathbf{Z}) \), where \( \mathbf{Z} \) is the residual from a regression of \( Z \) on \( \mathbf{X} \). The 2SLS estimator with a single instrument \( \mathbf{Z} \) is identical to the ILS estimator (Angrist and Pischke 2008).

Considering a randomization-based inference without imposing the linear modeling assumptions on the potential outcomes, neither regression in the ILS estimator can ensure efficiency gains relative to the estimator without adjusting the covariates (Freedman 2008). For example, in the second regression, the OLS estimator for \( \beta_1 \) may have a larger asymptotic variance than that of the unadjusted difference-in-means estimator \( \hat{\tau}_Y \) for estimating \( \tau_Y \). Thus, the ILS estimator may not always be more efficient than the Wald estimator \( \hat{\tau}_{\text{wald}} \). To address this issue, we can add the treatment assignment by covariate interactions in both ILS regressions, utilizing the idea of Lin (2013). Intuitively, if we can estimate \( \tau_Y \) and \( \tau_D \) more accurately, we can improve the efficiency of the LATE estimator.
Adding interactions is equivalent to regressing $Y_1^{obs}$ and $D_1^{obs}$ on $X_i$, in the treatment and control groups separately. For $R = Y$ and $R = D$, the OLS estimators for the slopes are
\[
\hat{\beta}_R(z) = \arg\min_{\beta \in \mathbb{R}} \sum_{i \in z} \left[ R_i^{obs} - \hat{\mu}_{z}^{obs} - (X_i - \bar{X}_z)^T \beta \right]^2, \quad z = 0, 1.
\]
The OLS estimators for $\tau_Y$ and $\tau_D$ are, for $R = Y$ and $R = D$,
\[
\hat{\tau}_{ols,R} = \left\{ \frac{\hat{R}_1^{obs} - (\bar{X}_1 - \bar{X})^T \hat{\beta}_R(1)}{n/2} \right\} - \left\{ \frac{\hat{R}_0^{obs} - (\bar{X}_0 - \bar{X})^T \hat{\beta}_R(0)}{n/2} \right\}.
\]
Then, we can obtain the ILS estimator with interactions: $\hat{\tau}_{ils} = \hat{\tau}_{ols,Y} + \hat{\tau}_{ols,D}$.

To study the asymptotic properties of $\hat{\tau}_{ils}$, we first define the projection coefficients:
\[
\beta_R(z) = \arg\min_{\beta \in \mathbb{R}} \sum_{i = 1}^n \left[ R_i(z) - \bar{R}(z) - (X_i - \bar{X})^T \beta \right]^2
\]
for $R = Y, D$, and $z = 0, 1$. We decompose the potential outcomes $Y_i(z)$ and $D_i(z)$ into two parts: $R_i(z) = \bar{R}(z) + (X_i - \bar{X})^T \beta_R(z) + e_{R,i}(z)$. Then, the above equations are only used to define the projection errors $e_{R,Y,i}(z)$ and $e_{R,D,i}(z)$, and all quantities in these equations are fixed.

Similarly to the Wald estimator, we define the transformed potential outcomes $\hat{A}_{ ils}: \hat{A}_{ ils}(z) = Y_i(z) - \hat{\tau}_{ils}D_i(z) - (X_i - \bar{X})^T \beta_R(z)$. Then, we use the sample variance of $\hat{A}_{ ils}(z)$ to estimate $S^2_{A_{ ils}(z)}$, denoted by $S^2_{A_{ ils}(z)} = (n_z - p - 1)^{-1} \sum_{i \in z} (\hat{A}_{ ils}(z) - n_z^{-1} \sum_{i \in z} \hat{A}_{ ils}(z))^2$, where the factor $(n_z - p - 1)^{-1}$ adjusts for the degrees of freedom to achieve a better finite-sample performance. Then, a conservative estimator for $\sigma^2_{A_{ ils}}$ is $\hat{\sigma}^2_{A_{ ils}} = S^2_{A_{ ils}(1)} + S^2_{A_{ ils}(0)} - S^2_{A_{ ils}(1) - A_{ ils}(0)}$.

To establish the asymptotic normality of $\hat{\tau}_{ils}$, we need Condition 3. Let $S^2_{XY}$, $S_{XY}(z)$, and $S_{XD}(z)$ denote the finite-population variances of $X_i$ between $X_i$ and $Y_i(z)$, and between $X_i$ and $D_i(z)$, respectively.

**Condition 3.** As $n \to \infty$, (i) for each covariate $X_k (k = 1, \ldots, p)$, $\max_{i = 1, \ldots, n} |X_{ik} - \bar{X}_z|^2/n \to 0$; (ii) for $z = 0, 1$, the finite-population variances $S^2_{X}$, $S_{XY}(1)$, and $S_{XD}(z)$, tend to finite-population variances, $S^2_{\bar{X}}$, $S_{\bar{X}Y}(1)$, and $S_{\bar{X}D}(z)$, and its limit being strictly positive definite, and the limit of $\sigma^2_{A_{ ils}}$ is positive.

**Theorem 1.** Under Conditions 1–3, $\hat{\tau}_{ils} - \tau$ converges in probability to $0$ and $n^{1/2}(\hat{\tau}_{ils} - \tau)/\sigma_{A_{ ils}}$ converges in distribution to $N(0, 1)$. Furthermore, $\sigma^2_{A_{ ils}}$ in probability to the limit of $n^{1/2}(S^2_{A_{ ils}(1)} + S^2_{A_{ ils}(0)})/n_0$, which is less than that of $\sigma^2_{A_{ ils}}$.

**Theorem 1** provides a normal approximation for the distribution of $\hat{\tau}_{ils}$ under complete randomization. The variance estimator is generally conservative. It is consistent if and only if the unit-level treatment effect of $A_{ obs}(z)$ is constant, that is, $A_{ obs}(1) - A_{ obs}(0) = C$ for some constant $C$. Based on Theorem 1, an asymptotically conservative confidence interval for $\tau$ is $[\hat{\tau}_{ils} - q_{\alpha/2}n^{-1/2}\hat{\sigma}_{A_{ ils}}$, $\hat{\tau}_{ils} + q_{\alpha/2}n^{-1/2}\hat{\sigma}_{A_{ ils}}]$, whose asymptotic coverage rate is greater than or equal to $1 - \alpha$. Comparing the asymptotic variances and variance estimators of $\hat{\tau}_{ils}$ and $\hat{\tau}_{wald}$, we have the following result:

**Theorem 2.** Under the conditions of Proposition 1 and Theorem 1, the difference between the asymptotic variances of $n^{1/2}\hat{\tau}_{ils}$ and $n^{1/2}\hat{\tau}_{wald}$ is the limit of $-(\hat{\tau}_{D_0}^2\hat{\beta}_D(1) - \hat{\tau}_{D_0}\hat{\beta}_D(0))$, and the difference between the variance estimators, $\hat{\sigma}_{A_{ ils}}^2$ and $\hat{\sigma}_{wald}^2$, converges in probability to the limit of $-n\hat{\tau}_{D_0}^2\hat{\beta}_D(1)\hat{\beta}_D(0)/n_0 \leq 0$, where $\delta(1) = \beta_Y(1) - \beta_Y(0)$ and $\delta(0) = \beta_Y(0) - \beta_Y(0)$.

**Theorem 2** shows that both the asymptotic variance and variance estimator of $\hat{\tau}_{ils}$ are no greater than those of $\hat{\tau}_{wald}$. Thus, the ILS estimator with interactions improves or at least does not degrade the estimation and inference efficiencies. The improvement in asymptotic efficiency mainly depends on the $R_2^2$ of the projection of $Y_i(z)$ on $X_i$, which measures the variance of the potential outcomes that is explained by the covariates. In particular, if $\delta(1) = \delta(0) = 0$, that is, $X_i$ does not affect $Y_i(z) - \tau D_i(z)$, $\hat{\tau}_{ils}$ has no asymptotic variance gain compared with $\hat{\tau}_{wald}$. For another special case with $p_0\delta(1) + p_1\delta(0) = 0$, although $\hat{\tau}_{ils}$ has no asymptotic efficiency improvement, it can produce a shorter confidence interval because of the variance estimator, if at least one of $\delta(1)$ and $\delta(0)$ is not equal to 0.

### 4.2. Logistic Oaxaca–Blinder Estimator

The ILS estimator uses a linear regression model to fit the data. For binary or count outcomes, it is natural to consider nonlinear working models, such as logistic regression or Poisson regression, to improve the efficiency further. To motivate the method, we discuss an imputation-based interpretation of the OLS estimator $\hat{\tau}_{ils,Y}$ (Imbens and Rubin 2015; Guo and Basse 2023). We can derive $\hat{\tau}_{ils,Y}$ by the following imputation procedure: we fit a linear regression model of $Y_i^{obs}$ on $(1, X_i')^T$ using the data in the treatment group. Thereafter, for any unit $i$ in the control group, we impute the unobserved potential outcome $Y_i(1)$, denoted by $\hat{Y}_{i,1}$, by the fitted model and obtain $\hat{\tau}_{Y,i} = \hat{Y}_{i,1} - Y_i(0)$. Similarly, we can obtain $\hat{\tau}_{Y,i}$ for unit $i$ in the treatment group. Imbens and Rubin (2015) and Guo and Basse (2023) showed that $\hat{\tau}_{ils,Y} = n^{-1} \sum_{i=1}^n \hat{\tau}_{Y,i}$ in econometrics, this double-imputation procedure is known as the Oaxaca–Blinder method (Blinder 1973; Oaxaca 1973; Kline 2011).

We can use nonlinear models to impute the unobserved potential outcomes if they fit the data better than the linear model and obtain a generalized Oaxaca–Blinder estimator (Guo and Basse 2023). We extend this idea to the noncompliance case.

For $z = 0, 1$, to fill in the unobserved potential outcome $Y_i(z)$ for units assigned to the treatment arm 1 $\rightarrow$ $z$, we use data from the treatment arm $z$ and fit a model to predict $Y_i(z)$ using $X_i$. 
The predicted value is denoted as \( \hat{\mu}_{Y_i}(z) \). Then, we fill in the unobserved values of \( Y_i(1) \) and \( Y_i(0) \) using
\[
\hat{Y}_i(1) = \begin{cases} 
Y_i(1) & Z_i = 1 \\
\hat{\mu}_{Y_i}(1) & Z_i = 0 
\end{cases} \\
\hat{Y}_i(0) = \begin{cases} 
\hat{\mu}_{Y_i}(0) & Z_i = 1 \\
Y_i(0) & Z_i = 0 
\end{cases}
\]
and obtain the estimator for \( \tau \): \( \hat{\tau}_{\text{OR,Y}} = n^{-1} \sum_{i=1}^{n} (\hat{Y}_i(1) - \hat{Y}_i(0)) \). Similarly, we can obtain \( \hat{\tau}_{\text{OR,D}} = n^{-1} \sum_{i=1}^{n} (\hat{D}_i(1) - \hat{D}_i(0)) \). Then, the generalized Oaxaca–Blinder estimator for \( \tau \) is \( \hat{\tau}_{\text{OB}} = \hat{\tau}_{\text{OR,Y}} - \hat{\tau}_{\text{OR,D}} \).

For \( R = Y, D, z = 0, 1 \), we define \( \mu_{R_i}(z) \) as the population prediction of \( Y_i(z) \) using \( X_i \), corresponding to \( \mu_{R_i}(z) \), and \( \eta_{R_i}(z) = R_i(z) - \mu_{R_i}(z) \) as the residuals. The transformed potential outcomes \( A_{OB} \) are \( A_{OB}(z) = Y_i(z) - \eta(D_i - \{\mu_{Y_i}(z) - \tau\mu_{D_i}(z)\}, z = 0, 1 \). As shown in Theorem 3, the asymptotic variance of \( n^{1/2} \hat{\tau}_{\text{OB}} \) is the limit of the form
\[
\sigma_{\hat{\tau}_{\text{OB}}}^2 = \frac{n}{\tau_D^2} \left\{ S_{AOR(1)}^2(\eta_1) + S_{AOR(0)}^2(\eta_0) - S_{AOR(1)}^2 - S_{AOR(0)}^2 \right\}/n.
\]

Similar to the \( \sigma_{\hat{\tau}}^2 \) estimation, we can obtain a conservative variance estimator: \( \sigma_{\hat{\tau}_{\text{OB}}}^2 = n^{1/2} \hat{\tau}_{\text{OR,D}}^2 S_{AOR(1)}^2(n^{1/2} / (n + S_{AOR(0)}^2(n)) \hat{\tau}_{\text{OB}} \).

### Condition 4
For \( R = Y, D, z = 0, 1 \), (i) \( \bar{\sigma}_{\hat{\tau}_{\text{OB}}} \) holds with probability one; (ii) \( \left\{ \hat{\mu}(z) - \mu(z) \right\} \rightangle _n \stackrel{p}{\rightarrow} 0; \) (iii) for some sequence of function class \( \left\{ F_{\text{R}}(z) \right\} \), \( pr(\hat{\mu}(z) \in \left\{ F_{\text{R}}(z) \right\} \to 1 \) and \( N(\hat{\tau}(z), \| \cdot \|, s) \leq C/s^k \), where \( \| \cdot \| \) is the s-covering number of the metric space \( (F_{\text{R}}(z), \| \cdot \|, s) \); (v) as \( n \to \infty \), \( \max_{i=1,\ldots,n} A_{OB}(z) / n \to 0 \); (v) the finite-population variances, \( S_{AOR(1)}^2 \), \( S_{AOR(0)}^2 \), and \( S_{AOR(1)}^2 - S_{AOR(0)}^2 \), tend to finite limits, and the limit of \( \sigma_{\hat{\tau}_{\text{OB}}}^2 \) is positive.

### Theorem 3
Under Conditions 1, 2, and 4, \( \hat{\tau}_{\text{OR}} - \tau \) converges in probability to 0 and \( n^{1/2} (\hat{\tau}_{\text{OR}} - \tau) / \sigma_{\text{OR}} \) converges in distribution to \( N(0, 1) \). Furthermore, \( \sigma_{\hat{\tau}_{\text{OB}}}^2 \) converges in probability to the limit of \( n^{1/2} S_{AOR(1)}^2(n^{1/2} / n + S_{AOR(0)}^2(n)) \), which is no less than that of \( \sigma_{\hat{\tau}_{\text{OB}}}^2 \).

Theorem 3 provides a normal approximation for the distribution of \( \sigma_{\text{OB}} \). Again, the variance estimator is generally conservative, and is consistent if and only if the unit-level treatment effect \( A_{OB}(1) - A_{OB}(0) \) is constant.

For binary responses, we can fill in the unobserved potential outcome and treatment received \( Y_i(z) \) and \( D_i(z) \) for units assigned to the treatment arm \( 1 - z \) using logistic regression models: for \( R = Y, D \),
\[
\theta_R(z) = \arg \min_{\theta \in \mathbb{R}^{p+1}} \sum_{i:Z_i=z} \left\{ -R_i^\text{obs} \mathbf{X}_i^\top \theta + \log \left[ 1 + \exp \left( \mathbf{X}_i^\top \theta \right) \right] \right\}.
\]

We define \( \theta_D(z) \) and \( \theta_Y(z) \) as the solutions of the following finite-population logistic regression problems: for \( R = Y, D, z = 0, 1 \),
\[
\theta_R(z) = \arg \min_{\theta \in \mathbb{R}^{p+1}} L_R^{(n)}(\theta),
\]
\[
L_R^{(n)}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left[ -R_i \mathbf{X}_i^\top \theta + \log \left[ 1 + \exp \left( \mathbf{X}_i^\top \theta \right) \right] \right].
\]

Throughout the article, we assume that \( \theta_D(z) \) and \( \theta_Y(z) \) exist and are unique. Then \( \mu_{R_i}(z) = \exp(\mathbf{X}_i^\top \theta_R(z)) / [1 + \exp(\mathbf{X}_i^\top \theta_R(z))] \) are the predicted probabilities. As shown by Guo and Basse (2023), when \( Y \) is binary, Condition 4 is implied by Condition 5.

### Condition 5
(i) For all large \( n, R = Y, D, z = 0, 1 \), \( \nabla^2 L_R^{(n)}(\theta_2) - CI_{(p+1) \times (p+1)} \) is positive semidefinite for some constant \( C > 0 \) independent of \( n \) where \( I_{(p+1) \times (p+1)} \) is a \( (p + 1) \times (p + 1) \) identity matrix; (ii) the fourth moments of the covariates are uniformly bounded, that is, \( n^{-1} \sum_{i=1}^{n} X_{ik}^4 \leq M \) for \( k = 1, \ldots, p \) and some constant \( M < \infty \); (iii) as \( n \to \infty \), for \( z = 0, 1 \), the finite-population variances, \( S_{AOR(1)}, S_{AOR(0)} \), and \( S_{AOR(1)} - S_{AOR(0)} \), tend to finite limits, and the limit of \( \sigma_{\hat{\tau}_{\text{OB}}}^2 \) is positive.

### Remark 3
As pointed out by Guo and Basse (2023), when \( \mathbf{X} \) is rank-deficient, or the 1’s and 0’s among the observed outcomes in the treatment or control group can be perfectly separated by a hyperplane, which violates Condition 5, the logistic regression coefficient estimators \( \hat{\theta}_D(z) \) or \( \hat{\theta}_Y(z) \) may not be unique or even exist. In practice, if some of the potential outcomes are sparse (the proportion of elements 1 is close to zero or one, known as unbalanced data), the number of predictors used in the logistic regression model should be limited (Agresti 2013). Otherwise, the logistic regression model might overfit the data such that the fitted probabilities are numerically 0 or 1, which also violates Condition 5. In these cases, variable selection or regularization is necessary to obtain the estimators.

Based on Theorem 3, we can obtain an asymptotically conservative confidence interval for \( \tau \): \( \hat{\tau}_{\text{OB}} - q_{a/2} n^{-1/2} \hat{\sigma}_{\text{OB}} \leq \hat{\tau}_{\text{OB}} + q_{a/2} n^{-1/2} \hat{\sigma}_{\text{OB}} \). However, \( \hat{\tau}_{\text{OR,Y}} \) and \( \hat{\tau}_{\text{OR,D}} \) cannot always improve the asymptotic efficiency compared with the difference-means estimators \( \hat{\tau}_Y \) and \( \hat{\tau}_D \). We refer to Cohen and Fogarty (2022) for a counterexample. Thus, \( \hat{\tau}_{\text{OB}} \) may degrade the efficiency in some extreme cases compared with \( \hat{\tau}_{\text{wald}} \).

### 4.3. Calibrated Oaxaca–Blinder Estimator
Although the logistic regression model might be more appropriate to predict binary potential outcomes than the linear regression model, \( \hat{\tau}_{\text{OB}} \) cannot ensure efficiency gains. To solve
this non-superiority problem, we propose a calibrated Oaxaca–Blinder estimator, borrowing techniques from Cohen and Fogarty (2022).

The basic idea is to take the logistic regression fitted values \( \hat{\mu}_Y(z) \) and \( \hat{\mu}_{D,z}(z) \) as new covariates \( W_i^{\text{obs}} = (\hat{\mu}_Y(z), \hat{\mu}_{Y}(0), \hat{\mu}_{D,z}(1), \hat{\mu}_{D,z}(0))^T \), and regress \( y_i^{\text{obs}} \) and \( D_i^{\text{obs}} \) on \( W_i^{\text{obs}} \) in the treatment and control groups separately. The coefficient estimators are

\[
\hat{\gamma}_z(z) = \arg\min_{\gamma \in \mathbb{R}^4} \sum_{i=2}^n \left[ (\hat{y}_i^{\text{obs}} - \hat{W}_i^{\text{obs}})^T \gamma \right]^2,
\]

\[ R = Y, D_z \quad z = 0, 1. \]

We obtain the estimators for \( \gamma_Y \) and \( \gamma_D \):

\[
\hat{\gamma}_{\text{COB},R} = \left\{ \begin{array}{ll}
\hat{R}_0^{\text{obs}} - (\hat{W}_i^{\text{obs}} - \hat{W}_0^{\text{obs}})^T \hat{\gamma}_R(1) \\
\hat{R}_0^{\text{obs}} - (\hat{W}_i^{\text{obs}} - \hat{W}_0^{\text{obs}})^T \hat{\gamma}_R(0),
\end{array} \right.
\]

where \( \hat{W}_0^{\text{obs}} = n^{-1} \sum_{i=1}^{n} W_i^{\text{obs}} \). Then, the calibrated Oaxaca–Blinder estimator for \( \tau \) is \( \hat{\tau}_{\text{COB}} = \hat{\gamma}_{\text{COB},R} / \hat{\gamma}_{\text{COB},D} \).

To study the asymptotic properties of \( \hat{\tau}_{\text{COB}} \), we decompose the potential outcomes \( Y_i(z) \) and \( D_i(z) \) into projections on the space spanned by the linear combination of the covariates \( W_i = (\mu_Y(z), \mu_Y(0), \mu_{D,z}(1), \mu_{D,z}(0))^T \) and projection errors:

\[
\hat{R}_i(z) = \hat{R}_i(z) + (\hat{W}_i - \hat{W})^T \gamma_Y(z) + \xi_{R,i}(z), \quad R = Y, D_z, \quad z = 0, 1,
\]

and \( \xi_{R,i}(z) \) is the projection error.

We define the transformed potential outcomes \( A_{\text{COB}} \) as \( A_{\text{COB},i}(z) = Y_i(z) - \tau D_i(z) - (W_i - \hat{W})^T \gamma_Y(z) - \gamma_D(z), \) \( z = 0, 1 \). The asymptotic variance of \( n^{1/2} \hat{\tau}_{\text{COB}} \) is the limit of \( \sigma^2_{\text{COB}} \),

\[
\sigma^2_{\text{COB}} = \frac{n}{f_D} \left\{ S^2_{A_{\text{COB},1}}(1)/n_1 + S^2_{A_{\text{COB},0}}(0)/n_0 - S^2_{A_{\text{COB},1} - A_{\text{COB},0}(1)}/n \right\}.
\]

Similar to the \( \sigma^2_{\text{ls}} \) estimation, let \( \hat{A}_{\text{COB},i}(z) \) be the estimated value of \( A_{\text{COB},i}(z) \) and let

\[
S^2_{A_{\text{COB},1}}(1) = (n_2 - 1)^{-1} \sum_{i \in Z_1} \left[ \hat{A}_{\text{COB},i}(z) - n_2^{-1} \sum_{i \in Z_1} \hat{A}_{\text{COB},i}(z) \right]^2,
\]

be the sample variance of \( \hat{A}_{\text{COB},i}(z) \) under treatment arm \( z \). Then, \( \sigma^2_{\text{COB}} \) can be conservatively estimated by \( \sigma^2_{\text{COB}} = n^{1/2} \hat{\tau}_{\text{COB}} [S^2_{A_{\text{COB},1}}(1)/n_1 + S^2_{A_{\text{COB},0}}(0)/n_0] \).

**Condition 6.** As \( n \to \infty \), for \( z = 0, 1 \), the finite-population covariances, \( S^2_W, S^2_{WY(2)} \), and \( S^2_{D(z)} \) tend to finite limits with \( S^2_W \) and its limit being strictly positive definite, and the limit of \( \sigma^2_{\text{COB}} \) is positive.

**Theorem 4.** Under Conditions 1, 2, 5, and 6, \( \hat{\tau}_{\text{COB}} - \tau \) converges in probability to 0 and \( n^{1/2} (\hat{\tau}_{\text{COB}} - \tau) / \sigma_{\text{COB}} \) converges in distribution to \( N(0,1) \). Furthermore, \( \sigma^2_{\text{COB}} \) converges in probability to the limit of \( n^{1/2} [S^2_{A_{\text{COB},1}}(1)/n_1 + S^2_{A_{\text{COB},0}}(0)/n_0] \), which is no less than that of \( \sigma^2_{\text{COB}} \).
where $X_i$ follows a two-dimensional normal distribution with mean zero and covariance $\Sigma$, which entries $\Sigma_{11} = \Sigma_{22} = 2$ and $\Sigma_{12} = \Sigma_{21} = \rho$.

The potential outcomes and covariates are generated once and then kept fixed. We set $n = 200, 500$, $e = n_1/n = 0.3, 0.4, 0.5$, and $\rho = 0, 1$. The proportions of compliers, always takers, and never takers are approximately 0.5, 0.36, and 0.14, respectively. The true values of LATE are $0.084$ for $n = 200$, $\rho = 0$, $0.159$ for $n = 200$, $\rho = 1$, $0.081$ for $n = 500$, $\rho = 0$, and $0.107$ for $n = 500$, $\rho = 1$. We perform completely randomized experiments 10,000 times to compare the performance of $\hat{\tau}_{\text{wald}}$, $\hat{\tau}_{\text{ils}}$, $\hat{\tau}_{\text{OB}}$, $\hat{\tau}_{\text{COB}}$, $\hat{\tau}_{\text{tls}}$, $\hat{\tau}_{\text{tls}_x}$, and $\hat{\tau}_{\text{abadie}}$, in terms of bias, standard deviation (SD), root mean square error (RMSE), empirical coverage probability (CP), mean SD estimator, and mean confidence interval length (CI length) of the 95% confidence intervals (CI).

The results for $n = 500$ are shown in Figure 1 and Tables 1 and 2, with the results for $n = 200$ presented in the Supplementary Material. From these results, we observe that, first, the biases of all of the methods are negligible in accordance with their asymptotic unbiasedness. Second, compared with $\hat{\tau}_{\text{wald}}$, $\hat{\tau}_{\text{ils}}$ reduces the RMSE and CI length by 33.8%–41.3% and 30.3%–35.4%, respectively. The logistic regression-based estimator $\hat{\tau}_{\text{OB}}$ further reduces the RMSE and CI length by 8.2%–15.3% and 7.6%–12%, respectively. The calibrated estimator $\hat{\tau}_{\text{COB}}$ performs similarly to $\hat{\tau}_{\text{OB}}$. Furthermore, the empirical coverage probabilities of the 95% confidence intervals produced by all these estimators are higher than 95% in all of the cases, because of the conservative

Figure 1. Density plot for the distributions of LATE estimators, $\sqrt{n}(\hat{\tau} - \tau)$, $n = 500$. 
variance estimators. In addition, the point estimators \( \hat{\tau}_{\text{dls}} \) and \( \hat{\tau}_{\text{wald}} \) are equivalent, but the standard deviation estimator of \( \hat{\tau}_{\text{dls}} \) is more conservative than that of \( \hat{\tau}_{\text{wald}} \). The point estimators \( \hat{\tau}_{\text{dls},x} \) and \( \hat{\tau}_{\text{abadie}} \) are equivalent, which can be seen as the Indirect Least Squares estimators without interactions. Their performances are similar to \( \hat{\tau}_{\text{dls}} \), but can not always improve the efficiency as critiqued by Freedman (2008).

### 5.2. Empirical Application

In this section, we apply the proposed methods to analyze a real dataset from the Student Achievement and Retention Project (STAR), a randomized trial to evaluate the effect of academic services or incentives on academic performance among first-year college students (Angrist, Lang, and Oreopoulos 2009a). This trial was conducted at one of the satellite campuses of a...
large Canadian university. All of the first-year students entering in September 2005 were randomly assigned to the treatment or control group. The students in the treatment group were offered academic support services (SSP), financial incentives (SFP), or a combination of services and incentives (SFSP), and they were required to sign up for consent; otherwise, they were ineligible for services and incentives. Let $Z_i$ denote whether the student $i$ was assigned to the treatment group, and $D_i$ denote whether the student $i$ signed up for consent. The monotonicity assumption seems reasonable because the students assigned to the control group were unlikely to sign up for consent. There were $n = 1461$ students in total, and the cross-tabulation of the treatment assigned and received is shown in Table 3.

The outcome of interest is whether the students were in good standing after one year. To evaluate the effect of services or incentives on college achievement, we estimate both the LATE and MLATE using the following 11 covariates to perform regression adjustment (linear or logistic): gender, age, high school GPA, whether mother tongue is English, whether lives at home, whether at first-choice school, whether plans to work while in school, whether rarely or never puts off studying for tests, whether wants more than a BA, whether intends to finish in 4 years, and parents’ education.

The point estimates and 95% CIs for LATE are presented in Table 4. The results show that the support of services or incentives has no significant effect on students’ good standing after one year. This conclusion is the same as that drawn by Angrist, Lang, and Oreopoulos (2009a). Moreover, compared with $\hat{\tau}_{wald}$, all of the model-assisted methods, $\hat{\tau}_{ls}$, $\hat{\tau}_{OB}$, and $\hat{\tau}_{COB}$, reduce the SD estimator by 3.7%, 3.7%, 5.6% for the LATE of SSP, 7.0%, 7.5%, 9.5% for the LATE of SFP, respectively. When estimating the LATE of SFSP, only $\hat{\tau}_{COB}$ reduces the SD estimator by 2.6%.

Since not all of the potential outcomes are observed, we do not know the true gains of model-assisted methods. Thus, we conduct an empirical Monte Carlo simulation by imputing the unknown potential outcomes; see the supplementary material.

### 6. Discussion

In this article, we study how to efficiently estimate the LATE and MLATE in a completely randomized experiment with noncompliance and a binary outcome.

Under the design-based inference and mild conditions, we proved that the Wald estimator is consistent and asymptotically normal, and the variance estimator is generally conservative.

To improve the estimation efficiency, we proposed three model-assisted methods, the ILS estimator with interactions, logistic Oaxaca–Blinder estimator, and calibrated Oaxaca–Blinder estimator, and established their asymptotic theories. Our analysis is purely design-based, allowing the working model to be misspecified. We showed that the ILS estimator with interactions is no worse than the Wald estimator; the logistic Oaxaca–Blinder estimator cannot ensure efficiency gains relative to the Wald estimator; the calibrated Oaxaca–Blinder estimator is generally more efficient than the Wald estimator and the logistic Oaxaca–Blinder estimator. The efficiencies of the calibrated Oaxaca–Blinder estimator and the ILS estimator with interactions are not ordered unambiguously, depending on which models, linear or logistic, fit the data better. In addition, we proposed conservative variance estimators to facilitate inferences.

This paper focuses on estimating the LATE and MLATE in completely randomized experiments with noncompliance and binary outcomes. Other causal estimands, such as the complier odds ratio, are also interesting and worthy of further investigation. The identification of the complier odds ratio is more complex than the LATE and MLATE, and is left to future research.

Moreover, regression adjustment methods have been widely used to improve the estimation efficiency in more complicated randomized experiments, such as stratified randomized experiments, paired randomized experiments, and cluster randomized experiments (Fogarty 2018; Liu and Yang 2020; Su and Ding 2021). It would also be interesting to extend the proposed methods to estimate the treatment effect in these experiments when noncompliance problems occur.

### Supplementary Materials

The supplementary material provides the extension to MLATE, proofs of main results and additional results of simulation and empirical application.

### Acknowledgments

The author is grateful to the editor, associate editor, three anonymous referees and Dr. Hanzhong Liu for their valuable comments, which have greatly

---

**Table 4. Results for the STAR data.**

| Effect     | Method  | Point estimator | SD estimator | 95% CI       | SD reduction |
|------------|---------|-----------------|--------------|--------------|--------------|
| SSP, LATE  | $\hat{\tau}_{wald}$ | 0.077 | 0.069 | [−0.059,0.214] | 0 |
|            | $\hat{\tau}_{ls}$    | 0.054 | 0.067 | [−0.078,0.185] | 0.037 |
|            | $\hat{\tau}_{OB}$    | 0.054 | 0.067 | [−0.078,0.185] | 0.037 |
|            | $\hat{\tau}_{COB}$   | 0.054 | 0.067 | [−0.075,0.182] | 0.056 |
| SFP, LATE  | $\hat{\tau}_{wald}$ | 0.040 | 0.043 | [−0.044,0.125] | 0 |
|            | $\hat{\tau}_{ls}$    | 0.033 | 0.040 | [−0.046,0.111] | 0.070 |
|            | $\hat{\tau}_{OB}$    | 0.034 | 0.040 | [−0.044,0.112] | 0.075 |
|            | $\hat{\tau}_{COB}$   | 0.034 | 0.039 | [−0.042,0.111] | 0.095 |
| SFSP, LATE | $\hat{\tau}_{wald}$ | 0.036 | 0.061 | [−0.083,0.156] | 0 |
|            | $\hat{\tau}_{ls}$    | 0.052 | 0.062 | [−0.069,0.173] | −0.008 |
|            | $\hat{\tau}_{OB}$    | 0.060 | 0.062 | [−0.061,0.181] | −0.013 |
|            | $\hat{\tau}_{COB}$   | 0.059 | 0.059 | [−0.057,0.176] | 0.026 |

**Note:** CI, confidence interval; SD reduction, relative to the Wald estimator.
improved the quality of this article. The data of STAR is available at https://doi.org/10.3886/E116327V1, (Angrist, Lang, and Oreopoulos 2009b) which is distributed through openICPSR, a public access repository supported by the Inter-university Consortium for Political and Social Research (ICPSR). The original collector of the data, ICPSR, and the relevant funding agency bear no responsibility for use of the data or for interpretations or inferences based upon such uses.

Disclosure Statement

The author reports there are no competing interests to declare.

References

Abadie, A. (2003), “Semiparametric Instrumental Variable Estimation of Treatment Response Models,” Journal of Econometrics, 113, 231–263. [708,713]

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2020), “Sampling-based versus Design-based Uncertainty in Regression Analysis,” Econometrica, 88, 265–296. [708]

Agresti, A. (2013), Categorical Data Analysis (3rd ed.), Hoboken, NJ: Wiley. [712]

Angrist, J., Lang, D., and Oreopoulos, P. (2009a), “Incentives and Services for College Achievement: Evidence from a Randomized Trial,” American Economic Journal: Applied Economics, 1, 136–163. [715,716]

Angrist, J., Lang, D., and Oreopoulos, P. (2009b), “Replication Data for: Incentives and Services for College Achievement: Evidence from a Randomized Trial,” Dataset. Nashville, TN: American Economic Association [publisher]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. 2019-12-07. DOI:10.3886/E116327V1 [717]

Angrist, J. D., and Imbens, G. W. (1995), “Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity,” Journal of the American Statistical Association, 90, 431–442. [707]

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996), “Identification of Causal Effects Using Instrumental Variables,” Journal of the American Statistical Association, 91, 444–455. [707,709]

Angrist, J. D., and Pischke, J.-S. (2008), Mostly Harmless Econometrics: An Empiricist’s Companion, Princeton: Princeton University Press. [708,709,710]

Athey, S., and Imbens, G. W. (2017), “The State of Applied Econometrics: Blending Causal Inference with Machine Learning,” Journal of the American Statistical Association, 112, 1759–1769. [707,708,710]

Athey, S., and Imbens, G. W., and Rubin, D. B. (1997), “Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance,” The Annals of Statistics, 25, 305–327. [707,708]

——— (2015), Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction, New York: Cambridge University Press. [707,708,711]

Kang, H., Peck, L., and Keele, L. (2018), “Inference for Instrumental Variables: A Randomization Inference Approach,” Journal of the Royal Statistical Society, Series A, 181, 1231–1254. [708]

Kline, P. M. (2011), “Oaxaca-Blinder as a Reweighting Estimator,” American Economic Review, 101, 532–537. [711]

Li, X., and Ding, P. (2017), “General Formulas of Finite Population Central Limit Theorems with Applications to Causal Inference,” Journal of the American Statistical Association, 112, 1759–1769. [707,709,710]

——— (2020), “Randomization and Regression Adjustment,” Journal of the Royal Statistical Society, Series B, 82, 241–268. [707,710]

Lin, W. (2013), “Agnostic Notes on Regression Adjustments to Experimental Data: Reexaming Freedman’s Critique,” The Annals of Applied Statistics, 7, 295–318. [707,708,710]

Liu, H., and Yang, Y. (2020), “Regression-Adjusted Average Treatment Effect Estimators in Stratified Randomized Experiments,” Biometrika, 107, 935–948. [707,710,716]

Neyman, J., and Kolodziejczyk, S. (1935), “Statistical Problems in Agricultural Experimentation,” Supplement to the Journal of the Royal Statistical Society, 2, 107–180. [708]

Oaxaca, R. L. (1973), “Male-Female Wage Differentials in Urban Labor Markets,” International Economic Review, 14, 693–709. [711]

Ogburn, E. L., Rotnitzky, A., and Robins, J. M. (2015), “Doubly Robust Estimation of the Local Average Treatment Effect Curve,” Journal of the Royal Statistical Society, Series B, 77, 373–396. [708]

Rambachan, A., and Roth, J. (2020), “Design-based Uncertainty for Quasi-Experiments,” arXiv:2008.06062. [708,710]

Robins, J. M. (1988), “Confidence Intervals for Causal Parameters,” Statistics in Medicine, 7, 773–785. [708]

Rubin, D. B. (1974), “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies,” Journal of Educational Psychology, 66, 688–701. [707]

——— (1980), “Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment,” Journal of the American Statistical Association, 55, 591–593. [707]

Schochet, P. Z., McConnell, S. M., and Burghardt, J. A. (2003), “National Job Corps Study: Findings Using Administrative Earnings Records Data,” Final Report. Mathematica Policy Research, Incorporated. [707]

Sherman, L. W., and Berk, R. A. (1984), “The Specific Deterrent Effects of Incentives and Services for College Achievement: Evidence from a Randomized Trial.” Dataset. Nashville, TN: American Economic Association [publisher]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-12-07. DOI:10.3886/E116327V1 [717]

Su, F., and Ding, P. (2021), “Model-Assisted Analyses of Cluster-Randomized Experiments,” Journal of the Royal Statistical Society, Series B, 83, 994–1015. [716]

Sun, B., and Tan, Z. (2022), “High-Dimensional Model-Assisted Inference for Cluster-Randomized Experiments,” Journal of the Royal Statistical Society, Series B, 83, 994–1015. [716]

Schochet, P. Z., McConnell, S. M., and Burghardt, J. A. (2003), “National Job Corps Study: Findings Using Administrative Earnings Records Data,” Final Report. Mathematica Policy Research, Incorporated. [707]

Sherman, L. W., and Berk, R. A. (1984), “The Specific Deterrent Effects of Arrest for Domestic Assault,” American Sociological Review, 49, 261–272. [707]

Splawa-Neyman, J., Dabrowska, D. M., and Speed, T. P. (1990), “On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9,” Statistical Science, 5, 465–472. [707]

Su, F., and Ding, P. (2021), “Model-Assisted Analyses of Cluster-Randomized Experiments,” Journal of the Royal Statistical Society, Series B, 83, 994–1015. [716]

Sun, B., and Tan, Z. (2022), “High-Dimensional Model-Assisted Inference for Local Average Treatment Effects with Instrumental Variables,” Journal of Business & Economic Statistics, 40, 1732–1744. [708]

Susanne, J.-B., Dynarski, M., Moore, M., Deke, J., Mansfield, W., and Pistorino, C. (2005), “When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program:
Final Report,” U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. [707]

Tan, Z. (2006), “Regression and Weighting Methods for Causal Inference Using Instrumental Variables,” Journal of the American Statistical Association, 101, 1607–1618. [708]

Wald, A. (1940), “The Fitting of Straight Lines If both Variables are Subject to Error,” The Annals of Mathematical Statistics, 11, 284–300. [707]

Wang, L., Zhang, Y., Richardson, T. S., and Robins, J. M. (2021), “Estimation of Local Treatment Effects under the Binary Instrumental Variable Model,” Biometrika, 108, 881–894. [708]

Wooldridge, J. M. (2010), Econometric Analysis of Cross Section and Panel Data (2nd ed.), Cambridge, MA, London, England: The MIT Press. [708,709,710]

Yue, L., Li, G., Lian, H., and Wan, X. (2019), “Regression Adjustment for Treatment Effect with Multicollinearity in High Dimensions,” Computational Statistics & Data Analysis, 134, 17–35. [707,710]