

Inference on a New Class of Sample Average Treatment Effects*

Jasjeet S. Sekhon
(UC Berkeley)

Yotam Shem-Tov
(UC Berkeley)

October 17, 2017

Abstract

We derive new variance formulas for inference on a general class of estimands of causal average treatment effects in a Randomized Control Trial (RCT). We demonstrate the applicability of the new theoretical results using an empirical application with hundreds of online experiments with an average sample size of approximately one hundred million observations per experiment. We generalize [Robins \(1988\)](#) and show that when the estimand of interest is the Sample Average Treatment Effect of the Treated (SATT, or SATC for controls), a consistent variance estimator exists. Although these estimands are equal to the Sample Average Treatment Effect (SATE) in expectation, potentially large differences in both accuracy and coverage can occur by the change of estimand, even asymptotically. Inference on SATE, even using a conservative confidence interval, provides incorrect coverage of SATT or SATC. We derive the variance and limiting distribution of a new and general class of estimands—any mixing between SATT and SATC—of which SATE is a specific case. An R package, `estCI`, that implements all the proposed estimation procedures is available.

Keywords: Average treatment effect, causality, Neyman causal model

*Jasjeet S. Sekhon: Robson Professor of Political Science and Statistics, University of California Berkeley; sekhon@berkeley.edu; Yotam Shem-Tov: Ph.D. Candidate, Department of Economics, shemtov@berkeley.edu. A previous version of this paper was entitled “Efficient Estimation of Average Treatment Effects under Effect Heterogeneity.” We thank Peter Aronow, Max Balandat, Eytan Bakshy, Avi Feller, Johann Gagnon-Bartsch, Peng Ding, Ben Hansen, Guido Imbens, James Robins, Sören Kunzel, Winston Lin, Juliana Londoño Vélez, Fredrik Sävje, and John Myles White for helpful comments and discussions. In addition, we thank Max Balandat, Eytan Bakshy, and John Myles White for help with one of the applications. We also thank the participants of the Berkeley Statistics Annual Research Symposium 2017 and the Atlantic Causal Inference Conference 2017. Sekhon wishes to acknowledge Office of Naval Research (ONR) grants N00014-15-1-2367 and N00014-17-1-2176. The R package, `estCI`, that implements the estimation methods described in the paper is available at <https://github.com/yotamshemtov/estCI>.

1 Introduction

The Neyman variance estimator is the most commonly used variance estimator in randomized experiments (Imbens and Rubin, 2015). Under the super-population model, it is a consistent estimator for the variance of the difference-in-means, and this probably accounts for its popularity. However, under Neyman’s finite population model, a consistent variance estimator for the difference-in-means does not exist (Neyman, 1923, 1990), and Neyman’s variance estimator is conservative. Sharper, albeit still conservative, variance estimators exist (Aronow et al., 2014), but they are not often used.

To estimate the Sample Average Treatment Effect (SATE), researchers use the limiting distribution of the difference-in-means recentered around SATE. We show that changing the estimand and recentering the difference-in-means limiting distribution with respect to the Sample Average Treatment Effect for the Treated (SATT, or SATC for controls) allows one to obtain a consistent non-conservative variance estimator. Consequently, recentering with respect to SATT (or SATC) can yield a prediction interval (PI) for SATT (or SATC) that is substantially different from a confidence interval (CI) for SATE. It follows that inference on SATE has incorrect coverage and/or is inefficient for the estimation of SATT (or SATC).¹ The key result of the paper is the derivation of valid inference on a new and general class of causal estimands that results from any mixing between SATT and SATC. In addition, we present a plug-in estimator for the mixing between these two estimands that maximizes the accuracy – minimizes the variance – of the recentered difference-in-means estimator. We call this causal estimand the Sample Average Treatment Effect Optimal (SATO). SATE is a specific case of this general class in which the weighting between SATT and SATC is done by the probability of treatment assignment. SATO is also equal to SATE under a constant/homogeneous treatment effect model.²

The objective of the paper is to develop limiting distribution results for inference on sample

¹This holds even though SATE, SATT, and SATC are equal in expectation and the difference-in-means is used to estimate all three estimands. Note that under the super-population model, there is no difference between the equivalent three population estimands, as the Population Average Treatment Effect (PATE) and the Population Average Treatment Effect of the Treated (PATC) are equal when the treatment assignment is done at random.

²A related literature discusses the idea of an optimal estimand in terms of covariate balance in observational studies (Crump et al., 2009; Li et al., 2016). Crump et al. (2009) suggested a procedure for choosing the optimal estimand in observational studies where there is limited overlap in the covariates. The population overlap issue does not arise in randomized experiments.

average treatment effects (e.g., SATE, SATT). [Imbens \(2004\)](#) made explicit the difference between SATE/PATE and SATT/PATT regarding inference, and noted that “[w]ithout further assumptions, the sample contains no information about the PATE beyond the SATE.” Therefore, sample estimands are often of direct interest because one does not wish to make additional assumptions. Inference on SATO yields a PI that has correct coverage (of SATO) and is substantially shorter, more accurate, than a CI for SATE. The heterogeneity in the response of different units to the treatment is the driving force behind the gain in accuracy. The change of estimand yields a recentered difference-in-means test statistic that is sensitive to differences in the variance of the treated and control units, and not only to the mean impact of the treatment. Our approach detects treatment effect heterogeneity by changing the estimand (and corresponding variance formulas), while keeping fixed the test statistic. Detecting and testing for heterogeneity in the treatment effect is of increasing importance to applied researchers ([Ding et al., 2016](#)).

Furthermore, in some applications the estimand of interest is SATT and using a CI for SATE is inaccurate. An example of when SATT can be the estimand of interest is an attributable treatment effect model in which the treatment effect varies across units ([Rosenbaum, 2001](#); [Feng et al., 2014](#); [Keele et al., 2017](#)). In general, PIs for SATT are not guaranteed to have correct coverage of SATE. We provide analytical and simulation based evidence for when a PI for SATT has approximately correct coverage of SATE and when it does not. The key factor behind the differences is the variance of the treatment effect distribution. As the impacts of the treatment are more heterogeneous, inference on SATE differs from inference on SATO.

Our possibly surprising results do have an intuitive interpretation: unlike in the case of the super-population model in which PATE equals PATT and PATC, in the case of Neyman’s finite sample model, the three estimands differ from each other and each recentering choice yields a different variance expression for the limiting distribution of the recentered difference-in-means. Accuracy differences between inference on SATT relative to inference on SATE come from two channels. First, in the case of SATT (and SATC), one does not need to use conservative bounds for the variance estimator as it is point-identified. Second, the change of estimand from SATE to SATT (or SATC) changes the variance of the recentered difference-in-means. We discuss and

decompose the conditions for accuracy gains from each one of these channels. [Robins \(1988\)](#) studied this phenomenon for the special case of binary outcomes. He emphasized that a CI for SATE does not yield correct coverage of SATT (or STAC). We extend his result in various ways, including by providing results for non-binary outcomes and providing conditions under which the SATT and SATC prediction intervals yield gains in accuracy. Unlike Robins, we discuss inference on possible mixes between SATT and SATC and provide a criterion for a plug-in estimator for the optimal mixing between the two estimands.

Our results also extend [Rigdon and Hudgens \(2015\)](#), who showed that PIs for SATT (and SATC) can be combined to construct a CI for SATE in the context of binary outcomes. We generalize their result to the case of non-binary outcomes. In addition, we show that when using the difference-in-means test statistic, combining PIs for SATT and SATC yields the exact same conservative CI for SATE as one would have gotten by using a bound for the variance of the difference-in-means, which assumes that the correlation between potential outcomes is one. More efficient CIs can be constructed directly using sharper bounds on the unobserved correlation between potential outcomes ([Aronow et al., 2014](#)). Taken together, our results prove that as long as the test statistic is the difference-in-means, combining separate PIs for SATT and SATC, which have been derived in a non-conservative procedure, does *not* allow one to construct a CI for SATE that is more efficient than existing procedures that use conservative variance estimators and directly conduct inference on SATE.

Inference on SATT and SATC may also be of interest even when the primary parameter of interest is SATE. SATT represents the effect of the treatment on the units that have been assigned/exposed to it, and SATC represents the effect of the treatment on the units that have not been exposed to it under the observed treatment allocation. To infer whether the treatment *had* an effect on average, the key parameter of interest is SATT. However, in order to infer whether treatment *will* have an average effect if applied to all the sample units, information on both SATT and SATC is needed. The difference in PIs for each of these estimands is a diagnostic tool for understanding how much of the uncertainty about SATE arises from uncertainty about SATT and how much of it arises from our uncertainty about whether the treatment would have had

an effect on the controls (SATC). Decomposing the uncertainty over the SATE can be used as a diagnostic tool for understanding the effects of an intervention, especially in medical trials where heterogeneity in treatment effects is common.

The remainder of this paper is organized as follows. Section (2) presents an illustrative data example. Section (3) describes the theoretical framework, definitions, and notation that are used throughout the paper. Section (4) describes the key theoretical results of the paper. Section (5) provides Monte Carlo simulations from three different data generating processes. Section (6) extends some of the theoretical results to the super-population sampling model. Section (7) mentions several notes and remarks about the theoretical results and possible extensions. Section (8) discusses the main empirical data application that consists of hundreds of online experiments with millions of observations. Section (9) concludes.

2 Illustrative example

Following Rosenbaum (2001), we use data from Tunca and Egeli (1996) who studied the effect of benzene exposure on cytogenetic changes over time. Rosenbaum analyzed this data under an attributable treatment effect model in which the estimand of interest is SATT. The treated group ($m = 50$) is shoe workers in the area of Bursa, Turkey, and the control group ($N - m = 20$) is residents of Bursa area who were not exposed to benzene. Although the treatment was not assigned at random because this is an observational study, for illustrative purposes we follow previous researchers and assume that the treatment assignment was unconfounded.

Figure (1) shows a CI (SATE) and PIs (SATT and SATC) for the sample average treatment effects. The PI for SATT is 27% shorter, than a CI for the SATE, while the PI for SATC is 55% longer. The figure demonstrates that in this example SATT is more accurately estimated than SATE, which has a meaningful interpretation. It implies that there is less uncertainty about whether the treatment had an effect, and more uncertainty about the effect the treatment will have if expanded to the units that are currently under the control regime.³

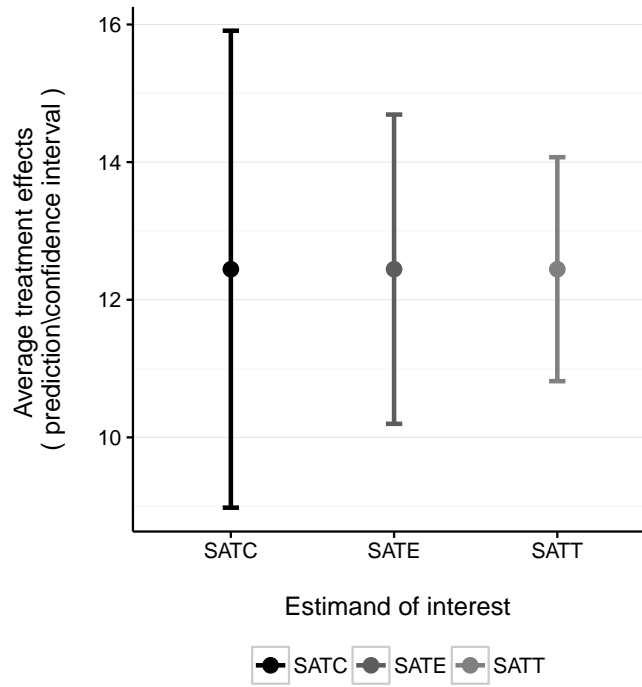
In this example if the researcher is interested in SATT, then a CI for SATE will be overly

³Note that, SATE is a weighted average of SATT and SATC by the probability of treatment assignment.

conservative and inefficient. However, if the estimand of interest is SATC, then a CI for SATE will have incorrect coverage and will be too short; i.e., it will over reject the null hypothesis that SATC is equal to zero.

In what follows we derive analytical results for inference on any weighting between SATT and SATC, which SATE is a special case of. The optimal weighting, in terms of accuracy, between the two estimands depends on the variance of the units under the control and treatment regimes and the correlation between potential outcomes. In the simulations and the main empirical application we show how to conduct inference on an estimand that is a mix between SATT and SATC and that it can detect treatment effect heterogeneity even when SATE is equal to zero.

Figure 1: **Confidence/prediction intervals for average treatment effects (Tunca and Egeli, 1996)**



3 Setting, definitions and notation

We follow Neyman’s finite population causal model. Consider a fixed finite population of $N \geq 4$ units and two dimensions, $Y(0)$ and $Y(1)$:

$$\Pi_N = \{(Y(0)_{1N}, Y(1)_{1N}), (Y(0)_{2N}, Y(1)_{2N}), \dots, (Y(0)_{NN}, Y(1)_{NN})\} \quad (1)$$

The researcher observes a random sample of m units from the finite population:

$$\Pi_N^1 = \{Y(1)_{1N}, Y(1)_{2N}, \dots, Y(1)_{NN}\} \quad (2)$$

and the vector of treatment indicators, $\mathbf{T} = (T_1, \dots, T_N)$, represents the m units that are randomly sampled to the treatment group.⁴ The remaining $N - m$ units are assigned to the control group and they form a random sample of $N - m$ units from the finite population $\Pi_N^0 = \{Y(0)_{1N}, Y(0)_{2N}, \dots, Y(0)_{NN}\}$, which is represented by the vector of indicators, $(1 - T_1, \dots, 1 - T_N)$.

The probability of each unit being assigned to the treatment regime is $p = \frac{m}{N}$ and the correlation between the treatment assignment of each two units is negative:

$$\text{Cov}(T_i, T_j) = p \cdot \left(\frac{m-1}{N-1} - p \right) < 0 \quad (3)$$

Let Y_i be the outcome of interest for unit i , which is a function of $Y(\mathbf{T})$. We assume SUTVA (Holland, 1986) is satisfied:

$$Y_i(\mathbf{T}) = Y_i(T_i) \quad (4)$$

and T_i is assigned at random:

$$Y(1), Y(0) \perp\!\!\!\perp T \quad (5)$$

⁴We study the classic Neyman causal model for a finite population. For a review of the framework and the classic CLT results under the finite population model, see Li and Ding (2016). Note that, the randomization model implies that the number of treated units, m , is a fixed number and not a random variable. The only random component is the treatment indicators.

Let τ_i denote the treatment effect on unit i , $Y_i(1) - Y_i(0)$, and let the vector of treatment effects in the finite population be denoted by $\boldsymbol{\tau} = \mathbf{Y}(1) - \mathbf{Y}(0)$. The researcher might be interested in conducting inference on several possible average treatment effect estimands:

$$\text{SATE} = \frac{1}{N} \cdot \sum_{i=1}^N \tau_i \quad (6)$$

$$\text{SATT} = \frac{1}{m} \cdot \sum_{i=1}^N \tau_i \cdot T_i, \quad \text{where} \quad m = \sum_{i=1}^N T_i \quad (7)$$

$$\text{SATC} = \frac{1}{N - m} \cdot \sum_{i=1}^N \tau_i \cdot (1 - T_i) \quad (8)$$

We do not impose any parametric assumptions on the relationship between $Y(1)$ and $Y(0)$, and allow τ to vary across units. It is important to note that unlike SATE, both SATT and SATC are random variables, even conditional on the sample, as they are a function of \mathbf{T} . Because SATT is a random variable, a random interval that contains SATT a with probability of $1 - \alpha$ is usually referred to as a prediction interval (PI) or, in Bayesian terminology, a credible interval, rather than a CI. We are not the first ones to discuss inference on a causal estimand that is a random variable (Robins, 1988; Bowers and Hansen, 2009).

4 Theory

In this section we discuss how to conduct inference, without parametric or distributional assumptions, on a general class of average treatment effect estimands:

$$\omega \text{SATT} + (1 - \omega) \text{SATC} \quad (9)$$

We first present theoretical results on inference about SATT (or SATC) that generalize Robins (1988) to non-binary and continuous outcomes. Next we conduct a comparison between inference on SATE and SATT in terms of accuracy and coverage (i.e., type I error). We show that a PI for SATT (or SATC) can be substantially different (smaller or larger) than a CI for SATE, and we decompose the differences to those that result from a change of estimand and those that are the

result of being forced to use a conservative variance estimator when conducting inference on SATE. At the end of this section we derive the accuracy-maximizing estimand. This is the mix between SATT and SATC that can be estimated with the highest level of precision while holding fixed the test statistic. SATE is a specific case of the general class of estimands described in equation (9) when $\omega = p$.

4.1 Inference on SATT (and SATC) relative to SATE

The classic estimator for SATE (and SATT) is the difference-in-means between the outcomes of units under the two treatment regimes. For notational convenience, denote the difference-in-means estimator by t_{diff} :

$$t_{\text{diff}} \equiv \frac{1}{m} \sum_{i=1}^N Y_i \cdot T_i - \frac{1}{N-m} \sum_{i=1}^N Y_i \cdot (1 - T_i) \quad (10)$$

Lemma (1) shows that the difference-in-means can be decomposed to three terms of which only two depends on $\boldsymbol{\tau}$. In equation (11), the first expression is a function of $Y_i(0)$ and T_i and is a random variable, the second is a function of $Y_i(0)$ and is not a random quantity, and the third is SATT. Lemma (1) motivates the use of t_{diff} for estimating SATT (or SATC). When recentering t_{diff} w.r.t. SATE the variance of the recentered estimator does not change; however, recentering w.r.t. SATT (or SATC) does change the variance calculations. This raises the question of whether inference on SATT (or SATC) has correct coverage of SATE and vice versa. Next we address this question both analytically and using Monte Carlo simulations.

Lemma 1. The difference-in-means test statistic t_{diff} can be decomposed into three expressions:

$$t_{\text{diff}}(\mathbf{Y}, \mathbf{T}) = \frac{N}{m \cdot (N-m)} \cdot \sum_{i=1}^N Y_i(0) \cdot T_i - \frac{1}{N-m} \cdot \sum_{i=1}^N Y_i(0) + \text{SATT} \quad (11)$$

or

$$t_{\text{diff}}(\mathbf{Y}, \mathbf{T}) = \frac{N}{m \cdot (N-m)} \cdot \sum_{i=1}^N Y_i(1) \cdot T_i - \frac{1}{N-m} \cdot \sum_{i=1}^N Y_i(1) + \text{SATC}$$

See Appendix (B) for the proof.

Lemma (2) describes the variance of t_{diff} when it is recentered with respect to different *causal* estimands of interest. The variance of $t_{\text{diff}} - \text{SATE}$ contains the parameter ρ , which cannot be observed and must be bound when conducting inference. Lemma (1) illustrates how by changing the estimand from SATE to SATT (or SATC), still using the difference-in-means test statistic, it is possible to conduct inference without needing to either know or bound ρ .

Lemma 2. The variance of the difference-in-means when recentered w.r.t. SATE, SATT or SATC is

$$\begin{aligned}\text{Var}(t_{\text{diff}} - \text{SATE}) &= \frac{1}{N \cdot (1-p) \cdot p} \cdot [p^2 \cdot \sigma_0^2 + (1-p)^2 \cdot \sigma_1^2 + 2p(1-p) \cdot \rho \cdot \sigma_0 \cdot \sigma_1] \\ \text{Var}(t_{\text{diff}} - \text{SATT}) &= \frac{1}{p \cdot (1-p) \cdot N} \cdot \sigma_0^2 \\ \text{Var}(t_{\text{diff}} - \text{SATC}) &= \frac{1}{p \cdot (1-p) \cdot N} \cdot \sigma_1^2\end{aligned}$$

where $\rho \equiv \frac{\text{Cov}(Y(1), Y(0))}{\sigma_0 \cdot \sigma_1}$ is the correlation between the potential outcomes under the treatment and control regimes, and σ_j^2 is defined as $\sigma_j^2 = \frac{\sum_{i=1}^N (Y(j)_i - \bar{Y}(j))^2}{N-1}$. See Appendix (C) for the proof.

The $t_{\text{diff}} - \text{SATT}$ can be more accurately estimated relative to the $t_{\text{diff}} - \text{SATE}$ when $\frac{\sigma_1}{\sigma_0} > 1$, ρ is sufficiently high, and p is not too high (e.g., $p = 1/2$). Theorem (4.1) shows that regardless of the value of $\frac{\sigma_1}{\sigma_0}$ there is a threshold level of ρ that below it $\text{Var}(t_{\text{diff}} - \text{SATE}) \leq \text{Var}(t_{\text{diff}} - \text{SATT})$ and above which $\text{Var}(t_{\text{diff}} - \text{SATE}) > \text{Var}(t_{\text{diff}} - \text{SATT})$. Notice that according to Theorem (4.1), it is simple to empirically test whether $\bar{\rho}$ is negative. All that is needed is to conduct a one-sided hypothesis test of the null:

$$H_0 : \frac{\sigma_1}{\sigma_0} \leq \sqrt{\frac{1-p^2}{(1-p)^2}}$$

and if the null hypothesis is rejected we can infer that $\bar{\rho} < 0$.

Theorem 4.1. For all σ_0 and σ_1 such that $\sigma_0 < \sigma_1$:

1. There exists a threshold level of ρ , $\bar{\rho}$ such that

$$\rho \leq \bar{\rho} \Rightarrow \text{Var}(t_{\text{diff}} - \text{SATE}) \leq \text{Var}(t_{\text{diff}} - \text{SATT})$$

$$\rho > \bar{\rho} \Rightarrow \text{Var}(t_{\text{diff}} - \text{SATE}) > \text{Var}(t_{\text{diff}} - \text{SATT})$$

2. When $\frac{\sigma_1}{\sigma_0} > \sqrt{\frac{1-p^2}{(1-p)^2}}$ then, $\bar{\rho} < 0$.

See Appendix (D) for the proof.

In practice, the correlation between potential outcomes is not observed and to estimate $\text{Var}(t_{\text{diff}} - \text{SATE})$ it is required to use a bound for ρ . The most commonly used estimator was proposed by Neyman and it ignores the correlation component all together. It can be re-written as:

$$\begin{aligned} \mathbb{V}_{\text{Neyman}} &= \frac{1}{m}\sigma_1^2 + \frac{1}{N-m}\sigma_0^2 \\ &= \frac{1}{N \cdot p \cdot (1-p)} (\sigma_1^2(1-p) + \sigma_0^2 p). \end{aligned} \quad (12)$$

This variance estimator is a consistent estimator for the variance of the difference-in-means under the super-population sampling model, and it can be used to conduct inference on the Population Average Treatment Effect (PATE). It also corresponds to Neyman's second variance estimator when the estimand is SATE (Neyman, 1923, 1990). A less conservative variance estimator for the difference-in-means bounds ρ at 1:

$$\begin{aligned} \mathbb{V}_{\rho=1} &= \frac{1}{m}\sigma_1^2 + \frac{1}{N-m}\sigma_0^2 - \frac{(\sigma_1 - \sigma_0)^2}{n} \\ &= \frac{1}{N \cdot (1-p) \cdot p} (p \cdot \sigma_0 + (1-p) \cdot \sigma_1)^2 \end{aligned} \quad (13)$$

Theorem 4.2. (Asymptotic distribution of $t_{\text{diff}} - \text{SATT}$) The difference-in-means recentered w.r.t. SATT follows a standard Normal distribution under two regularity conditions. When the following two conditions are satisfied:

$$N - m \rightarrow \infty, \quad m \rightarrow \infty \quad \text{and} \quad \frac{\max_{1 \leq i \leq N} (Y(0)_{Ni} - \bar{Y}(0)_N)^2}{\sum_{i=1}^N (Y(0)_{Ni} - \bar{Y}(0)_N)^2} \cdot \max\left(\frac{N-m}{m}, \frac{m}{N-m}\right) \rightarrow 0$$

then

$$\frac{t_{\text{diff}} - \text{SATT}}{\sqrt{\text{Var}(t_{\text{diff}} - \text{SATT})}} \xrightarrow{d} N(0, 1) \quad (14)$$

see Appendix (E) for the proof.⁵

Therefore, a $1 - \alpha$ prediction interval for SATT is

$$\left[t_{\text{diff}} - z_{1-\alpha/2} \cdot \hat{\sigma}_0 \cdot \sqrt{k(N, m)}, t_{\text{diff}} + z_{1-\alpha/2} \cdot \hat{\sigma}_0 \cdot \sqrt{k(N, m)} \right] \quad (15)$$

where

$$k(N, m) = \frac{1}{p \cdot (1 - p) \cdot N}$$

Theorem (4.2) establishes that the limiting distribution of the $t_{\text{diff}} - \text{SATT}$, when standardized using the variance formulas in Lemma (2), is standard Normal. An equivalent derivation of the limiting distribution can be carried out for SATC. The theorem for $t_{\text{diff}} - \text{SATC}$ follows immediately using an analog proof.

Rigdon and Hudgens (2015) showed how to derive a CI for SATE by combining two PIs for SATT and SATC. They focused on binary outcomes for which PIs have been derived using past theoretical results (Robins, 1988; Rosenbaum, 2001). Theorem (4.3) provides two key results. First, it is a generalization of Theorem (1) in Rigdon and Hudgens (2015) and shows how a CI for SATE can be constructed in any randomized control trial regardless of whether the outcomes are binary or continuous. This part of the theorem follows directly from the derivations in Rigdon and Hudgens (2015). Second, it shows that combining PIs for SATT and SATC using a Bonferroni-type adjustment, as was done by Rigdon and Hudgens, yields *exactly* the same CI as when one constructs a CI for SATE directly and uses a conservative variance estimator that bounds ρ at 1, $\mathbb{V}_{\rho=1}$. Theorem (4.3) implies that combining PIs for SATT and SATC yields a more conservative CI than using a CI based on a sharper bound for ρ , such as that of Aronow et al. (2014).

⁵These are not the weakest possible conditions.

Theorem 4.3. Let $[L_{\text{SATT}}, U_{\text{SATT}}]$ and $[L_{\text{SATC}}, U_{\text{SATC}}]$ be PIs for SATT and SATC according to the variance formulas in Lemma (2); then a CI for SATE is

$$[p \cdot L_{\text{SATT}} + (1 - p) \cdot L_{\text{SATC}}, \quad p \cdot U_{\text{SATT}} + (1 - p) \cdot U_{\text{SATC}}]$$

and is equal to

$$\left[t_{\text{diff}} - z_{1-\alpha/2} \cdot \sqrt{\hat{V}_{\rho=1}}, \quad t_{\text{diff}} + z_{1-\alpha/2} \cdot \sqrt{\hat{V}_{\rho=1}} \right]$$

See Appendix (F) for the proof.

Theorem 4.4. When $\sigma_1 \neq \sigma_0$, a PI for either SATT or SATC is shorter than a CI for SATE that uses either $\mathbb{V}_{\text{Neyman}}$ or $\mathbb{V}_{\rho=1}$.

See Appendix (G) for the proof.

According to Theorem (4.4), whenever $\sigma_0 < \sigma_1$ ($\sigma_0 > \sigma_1$), a PI for SATT (SATC) is shorter than a CI for SATE using Neyman's variance estimator. The gain in terms of interval length (in %) is

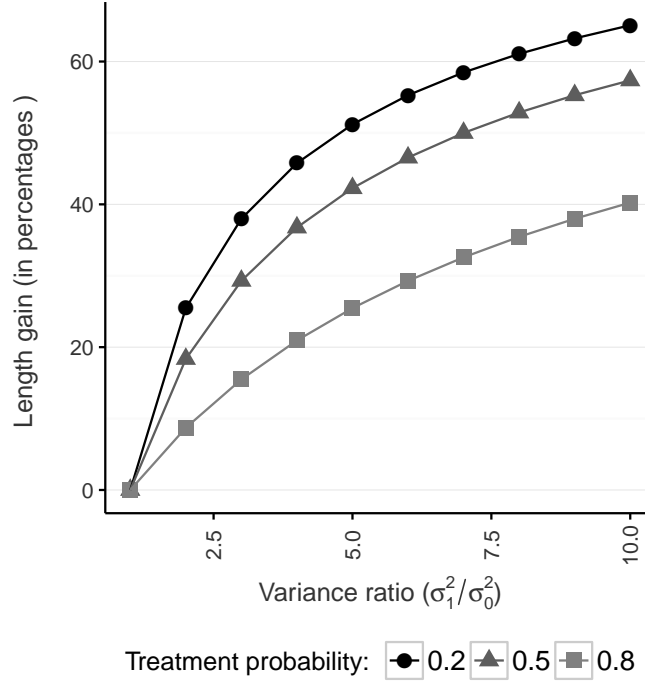
$$1 - \frac{1}{\sqrt{\left(\frac{\sigma_1^2}{\sigma_0^2}(1 - p) + p\right)}} \quad (16)$$

and the gains are decreasing with respect to p .

The interval length gains (in percentages) of a PI for SATT relative to a CI for SATE are illustrated in Figure (2). In a balanced design, $p = \frac{1}{2}$, with a variance ratio of two, the length/accuracy gain is 18.35% relative to a CI for SATE that is based on Neyman's variance estimator. Figure (2) shows the accuracy gains from estimating SATT instead of SATE for different levels of p , which can be substantial.

Note that Theorem (4.4) does not cover other estimators for the variance of $\text{Var}(t_{\text{diff}} - \text{SATE})$, such as that of Aronow et al. (2014). CIs based on these other variance estimators may be shorter than our PIs even when $\sigma_1 \neq \sigma_0$. To address these variance estimators, we derive the accuracy gains for a more general case in which ρ can be bounded by ρ^* , $\rho \leq \rho^*$. As the variance of

Figure 2: Percentage gains from a PI for SATT instead of a CI for SATE



$\text{Var}(t_{\text{diff}} - \text{SATE})$ is increasing w.r.t. ρ , it follows that substituting ρ^* with ρ yields a conservative variance estimator that is smaller than Neyman's variance estimator. The idea of substituting a bound of ρ instead of the true parameter value was proposed before in the literature (Reichardt and Gallob, 1999; Aronow et al., 2014). The percentage gain in terms of CI length is

$$1 - \frac{1}{\sqrt{p^2 + (1-p)^2 \cdot \left(\frac{\sigma_1}{\sigma_0}\right)^2 + 2p(1-p) \cdot \rho^* \cdot \frac{\sigma_1}{\sigma_0}}}$$

To obtain intuition about how recentering the difference-in-means w.r.t. SATT provides an accuracy gain, it is useful to decompose the variance of the difference-in-means. The decomposition in equation (17) shows that when recentering w.r.t. SATT we cancel two of the elements in the variance of the difference-in-means, $\text{Var}(\text{SATT})$ and $\text{Cov}\left(\text{SATT}, \frac{N}{m(N-m)} \cdot \sum_{i=1}^N Y(0)T_i\right)$, and in doing so reduce the uncertainty.

$$\text{Var}(t_{\text{diff}} - \text{SATE}) = \text{Var}(\text{SATT}) + \text{Var}\left(\frac{N}{m(N-m)} \cdot \sum_{i=1}^N Y(0)T_i\right) + 2 \cdot \text{Cov}\left(\text{SATT}, \frac{N}{m(N-m)} \cdot \sum_{i=1}^N Y(0)T_i\right) \quad (17)$$

To understand why inference on SATE can provide incorrect coverage of SATT (or SATC), it

is useful to look at the mean square distance between the two estimands:

$$\text{MSE}(\text{SATE}, \text{SATT}) = \frac{1-p}{m} \cdot \sigma_\tau^2 \quad (18)$$

see Appendix (A) for the proof. The differences between conducting inference on SATE or SATT does increase with the heterogeneity of the treatment effect. Given a fixed and bounded value of $\sigma_\tau^2 < \infty$, those difference will converges to zero as $m \rightarrow \infty$.

To illustrate how SATT and SATE can differ substantially, consider the following extreme example. There is a finite population of four units (see Table (1)) of which two are assigned to the treatment regime and the other two to the control regime. SATE is exactly zero in this example; however, SATT can be negative, positive, or zero depending on which units are allocated to the treatment group. In this example using inference on SATE (or SATT) to approximate inference on the SATT (or SATE) is misleading. However, inference on SATT can be valuable; and an estimand of interest in its own.

Table 1: **Example of when SATT can substantially differ from SATE**

Unit	$Y(1)$	$Y(0)$
1	1	0
2	-1	0
3	-100	0
4	100	0

4.2 Inference a general class of average treatment effects

In this section we derive inference for the Sample Average Treatment Effect Optimal (SATO), which is the weighting of SATT and SATC that minimizes MSE. The inference is conditional on the sample and uses the difference-in-means test statistics. Formally, SATO is defined as

$$\text{SATO} \equiv \omega^* \cdot \text{SATT} + (1 - \omega^*) \cdot \text{SATC}$$

s.t.

$$\omega^* = \underset{\omega}{\operatorname{argmin}} \operatorname{Var} \left(\hat{Y}_1 - \hat{Y}_0 - \text{SATO} \right) = \underset{\omega}{\operatorname{argmin}} \operatorname{MSE} \left(\hat{Y}_1 - \hat{Y}_0 - \text{SATO}, 0 \right)$$

Lemma 3. The variance of the difference-in-means when recentered w.r.t. any choice of weighting between SATT and SATC, $\text{Var}(t_{\text{diff}} - (\omega\text{SATT} + (1 - \omega)\text{SATC}))$, is

$$\begin{aligned} & \frac{1}{Np(1-p)} [p^2\sigma_0^2 + (1-p)^2\sigma_1^2 + 2p(1-p) \cdot \rho\sigma_0\sigma_1] \\ & + \frac{\sigma_\tau^2}{N-1} \cdot \left[\omega^2 \cdot \frac{1-p}{p} + (1-\omega)^2 \cdot \frac{p}{1-p} - 2\omega(1-\omega) \right] \\ & - 2 \cdot \left[\frac{\omega}{(N-1)p} [\sigma_1^2 - \rho\sigma_1\sigma_0] - \frac{1}{N-1} \cdot \sigma_\tau^2 + \frac{(1-\omega)}{(N-1)(1-p)} [\sigma_0^2 - \rho\sigma_1\sigma_0] \right] \end{aligned}$$

See Appendix (I) for the proof.

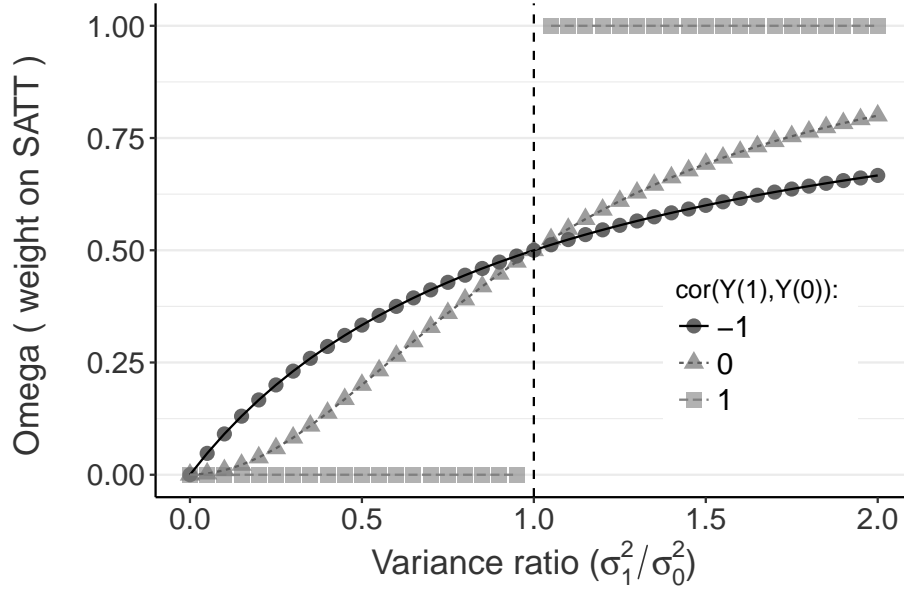
Lemma (3) presents the variance of the difference-in-means when it is recentered w.r.t. $\omega\text{SATT} + (1 - \omega)\text{SATC}$ for any choice of ω . It follows immediately from Lemma (3), that the optimal accuracy-maximizing value of ω is

$$\omega^* = \frac{\left(\frac{\sigma_1}{\sigma_0}\right)^2 - \rho \cdot \frac{\sigma_1}{\sigma_0}}{\left(\frac{\sigma_1}{\sigma_0}\right)^2 + 1 - 2\rho \left(\frac{\sigma_1}{\sigma_0}\right)}$$

The optimal choice of ω depends on two parameters, ρ and $\frac{\sigma_1}{\sigma_0}$, and is *independent* of p . This is in stark contrast to SATE, which is the weighting of SATT and SATC according *only* to the probability of treatment assignment. Figure (3) illustrates how the value of ω changes depending on ρ and $\frac{\sigma_1}{\sigma_0}$. Two clear patterns stand out. First, the weight that is assigned to SATT is monotonically increasing w.r.t. the variance ratio, $\frac{\sigma_1}{\sigma_0}$. Second, the relationship between ω^* and ρ is ambiguous and depends on $\frac{\sigma_1}{\sigma_0}$. When $\frac{\sigma_1}{\sigma_0} > 0$, as $\uparrow \rho$, the weight assigned to SATT increases, and the opposite is true when $\frac{\sigma_1}{\sigma_0} < 0$.

A key question is under what conditions SATE is equal to SATO, which is equivalent to asking whether SATE can ever be the estimand that is estimated most accurately. Two scenarios in which SATO coincides with SATE are (i) a constant treatment effect model and (ii) when $\sigma_1 = \sigma_0$ and $p = 1/2$. When the variance ratio is equal to 1, the optimal weighting between SATT and SATC is half and half ($\omega = 0.5$). Equality of variance between treatment and control units does *not* imply that SATO collapses to SATE. The right plot of Figure (4) shows the accuracy gain of SATO relative to SATE for different values of ρ and p . SATO can be estimated at a higher

Figure 3: Optimal ω weight of SATT in SATO for different $\frac{\sigma_1}{\sigma_0}$ and ρ



Notes: The above figure describes the relationship between the variance ratio of treated and control units ($\frac{\sigma_1}{\sigma_0}$) and the weight which is assigned to SATT in SATO for three different values of ρ .

degree of accuracy for most levels of ρ and for any $p \neq 1/2$, even when $\sigma_1 = \sigma_0$. The left plot of Figure (4) compares the accuracy of inference on SATE to a mixing between SATT and SATC – not SATO as ω^* is not used – for a variety of different ω weights and three values of ρ . SATE can be estimated more accurately than many mixes of SATT and SATC, but it will always – for a known ρ – be more noisily estimated than SATO. To conclude, it is hard to justify a weighting of SATT and SATC that will yield the SATE when the objective is to maximize accuracy. SATE is the only mix of SATT and SATC that yields a parameter and *not* a random variable.

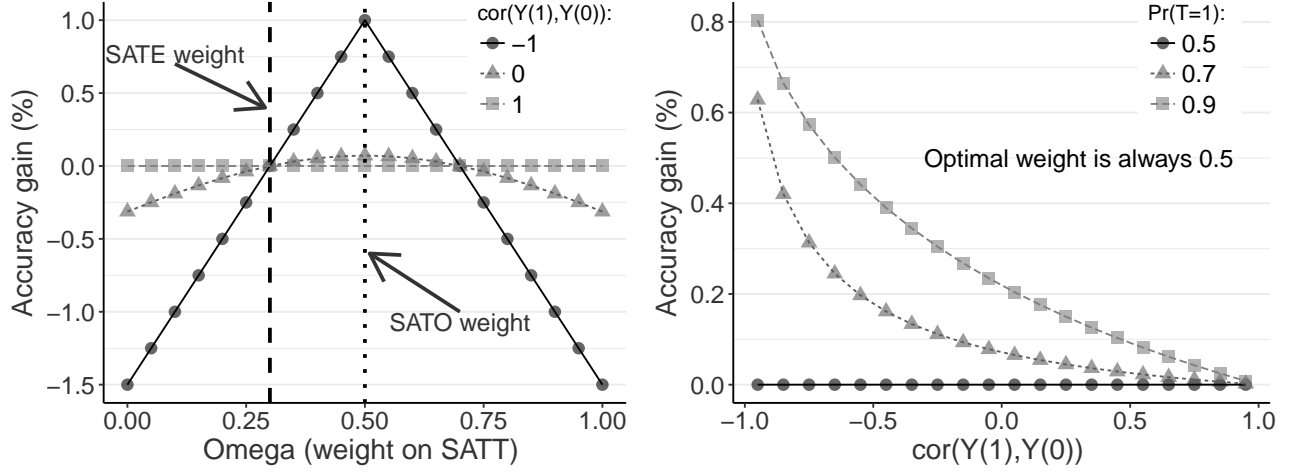
5 Analytical examples and simulations

5.1 Additive random coefficients model

We consider a simple additive treatment effect model with heterogeneous impacts across units:

$$Y_i(1) = \tau_i + Y_i(0) \tag{19}$$

Figure 4: The accuracy gain (in percentages) from estimating SATO instead of SATE when $\frac{\sigma_1}{\sigma_0} = 1$ for different values of the treatment assignment probability (p) and the correlation between potential outcomes (ρ)



Notes: The left plot describes the relationship between the accuracy gain of conducting inference on a mixing between SATT and SATC relative to SATE as a function of the weight which is assigned to SATT (ω) for three different values of ρ . The right plot illustrates the relationship between ρ and the accuracy gains of a PI for SATO relative to a CI for SATE for three different levels of the probability of treatment assignment. In both plots the relationships that are described are for the case in which $\frac{\sigma_1}{\sigma_0} = 0$, which implies that accuracy gains are not the result of a difference in the variance of the outcome among treated units relative to the control units.

where τ_i is a random variable, which for simplicity is assumed not to be correlated with $Y_i(0)$, nor to be correlated with T_i by construction due to the randomization of treatment assignment. The variance of the treated units is larger by σ_τ^2 , and this generates a potential difference between a CI for SATE and a PI for SATT. The variance ratio is:

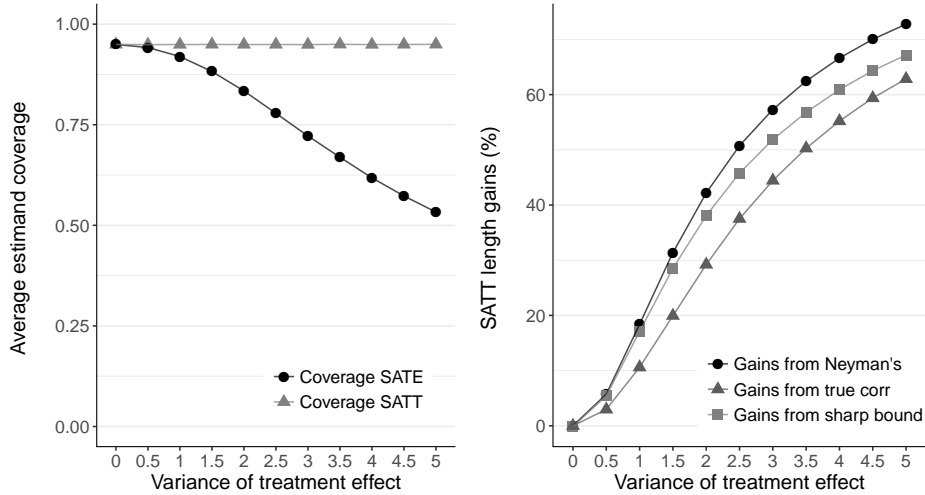
$$\frac{\sigma_\tau^2}{\sigma_0^2} + 1$$

and is increasing with the heterogeneity of the treatment effect. Consider the data-generating process in (20) with a sample of $n = 1,000$ units. We performed 1,000 draws of samples and for each one simulated 1,000 different allocations of the treatment according to $p = 1/2$. Figure (5) reports the simulation results. The PIs for SATT are substantially shorter than the CIs for the SATE; this holds both when using Neyman's variance estimator and when using a variance estimator based on a sharp bound for ρ , such as that of Aronow et al. (2014). The comparison to the CI based on the true ρ parameter shows that the accuracy differences are mainly due to the

change of estimand, rather than a conservative variance estimator. This is contrary to the binary outcome example (see below). The simulation is supported by our theoretical results that a change of estimand has larger accuracy impacts as the variance of the treatment effect is higher. The left plot in Figure (5) shows that coverage of the PI for SATT w.r.t. both SATT and SATE. As the treatment effect becomes more heterogeneous the PI for SATT provides worse coverage of SATE and, vice versa, a CI for SATE provides worse coverage of SATT. In a heterogeneous treatment effect environment it is required to use variance formulas that guarantee correct coverage of the estimand of interest, as SATE, SATT, and SATC can differ substantially from one another.

$$\begin{aligned}
Y_i(0) &\sim N(\mu = 10, \sigma_0^2 = 1) \\
\tau_i &\sim N(\mu = 0, \sigma_\tau^2) \\
Y_i(1) &= \tau_i + Y_i(0)
\end{aligned} \tag{20}$$

Figure 5: **Additive and heterogeneous treatment effect (random coefficient) simulation results: Length and coverage differences**



Notes: The left plot shows the coverage of a PI for SATT w.r.t. SATT and SATE. The coverage of SATT has correct size and there is no evidence of over rejection of the null hypothesis. On the other hand, the coverage of SATE becomes worse as the variance of the treatment effect increases. The right plot illustrates the accuracy gains of using inference on SATT relative to SATE. The PI for SATT is compared to CIs for SATE based on three different values of ρ . The circles represent the accuracy gains (in percentages) relative to Neyman's variance estimator which assumes that $\rho = 0$. The squares dots use a sharp bound for ρ that have been derived by [Aronow et al. \(2014\)](#). And the triangles compare the length of a PI for SATT to a CI for SATE when the true value of ρ is known to the researcher.

5.2 Binary outcomes

Consider a finite population of size N :

$$\Pi_N = \{(Y(0)_{1N}, Y(1)_{1N}), (Y(0)_{2N}, Y(1)_{2N}), \dots, (Y(0)_{NN}, Y(1)_{NN})\} \quad (21)$$

where $(Y(0)_{iN}, Y(1)_{iN}) \in \{0, 1\}^2$, and

$$p_0 = \frac{1}{N} \cdot \sum_{i=1}^N Y_i(0) \quad \text{and} \quad p_1 = \frac{1}{N} \cdot \sum_{i=1}^N Y_i(1),$$

The outcome variable is binary and we assume the treatment has a positive effect on average:

$$0 \leq p_0 \leq 1/2 \quad \text{and} \quad p_0 < p_1 \leq 1/2 \quad (22)$$

These conditions insure that the positive treatment effect increases the variance of the outcomes among the treated units. SATE is equal to $p_1 - p_0$, which implies that when the conditions in equation (22) are met, the variance of the treated units is strictly higher than that of the controls, $p_0(1 - p_0) < p_1(1 - p_1)$, and the variance ratio is a function of the treatment effect,

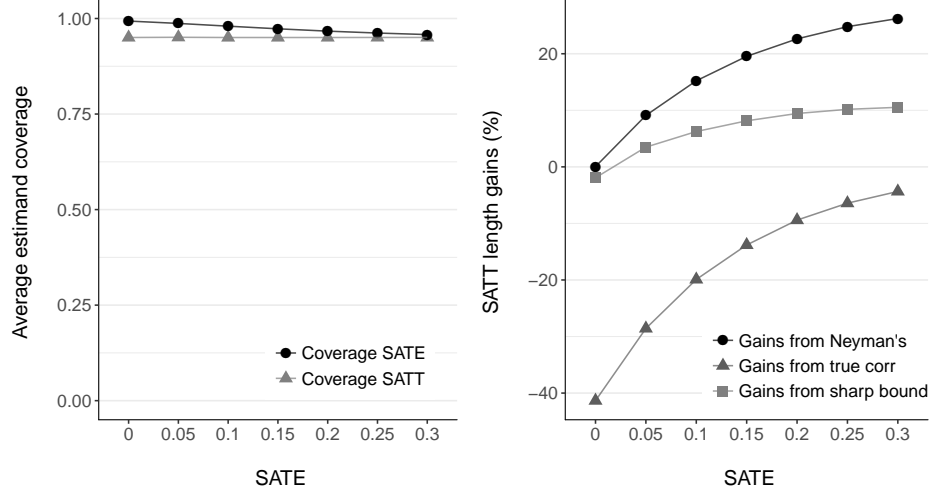
$$\frac{\sigma_1^2}{\sigma_0^2} = \frac{(p_0 + SATE)(1 - p_0 - SATE)}{p_0(1 - p_0)} \quad (23)$$

The accuracy gains are increasing with the treatment effect as long as the treatment generates a variance increase among the treated units. Figure (6) shows the interval length gains (in percentages) from using a PI for SATT in the case of a binary outcome. When $p_0 = 0.1$ and $p_1 = 0.2$, the interval for SATT is lower by 23.3% relative to the standard CI that is based on Neyman's variance estimator.

Next we decompose how much of the efficiency gain is due to a change of estimand, and how much is due to the conservative variance estimation. The lower line (i.e., triangles) in Figure (6) shows the percentage difference in length relative to a CI for SATE when ρ is known. This CI is substantially shorter than the ones for SATT, which demonstrates that in this simulation the

efficiency differences are due to the conservative variance estimation and not because of the change in estimand. Unlike the random coefficient model above, there are small differences in coverage under this data-generating model, although, there are differences in accuracy. The binary outcomes model has less heterogeneity in the treatment effect distribution (i.e., σ_τ^2 is smaller) which mitigates the differences between the different estimands.

Figure 6: **Binary outcome simulation: Length and coverage differences.**



Notes: See the notes in Figure (5).

5.3 Censored outcomes (Tobit model)

Consider a finite population of size N ,

$$\Pi_N = \{(Y(0)_{1N}, Y(1)_{1N}), (Y(0)_{2N}, Y(1)_{2N}), \dots, (Y(0)_{NN}, Y(1)_{NN})\} \quad (24)$$

where $Y(0)$ is a continuous outcome and $Y(1)$ is,

$$Y(1) = \begin{cases} Y(0) + \tau, & Y(0) \geq 0 \\ Y(0), & Y(0) < 0 \end{cases} \quad \text{and} \quad \tau > 0$$

The above data generating-process implies that the variance of the potential outcomes under treatment is higher than the variance of units under the control regime:

$$\text{Var}(Y(1)) = \sigma_0^2 + \Pr(Y(0) > 0) \cdot \tau \cdot [\tau \cdot (1 - \Pr(Y(0) > 0) + \mathbb{E}[Y(0)|Y(0) > 0] - \mathbb{E}[Y(0)])] > \sigma_0^2$$

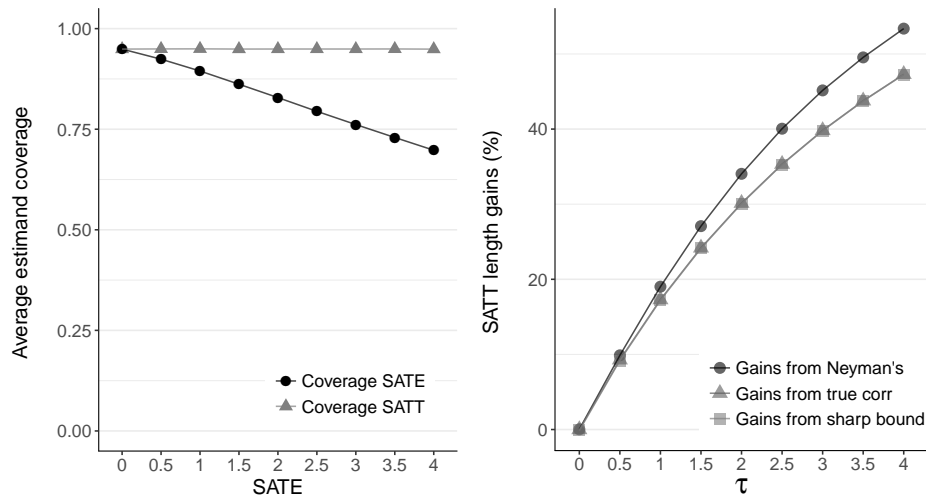
and the variance ratio is:

$$\frac{\sigma_1^2}{\sigma_0^2} = 1 + \frac{\Pr(Y(0) > 0) \cdot \tau \cdot [\tau \cdot (1 - \Pr(Y(0) > 0) + \mathbb{E}[Y(0)|Y(0) > 0] - \mathbb{E}[Y(0)])]}{\sigma_0^2}$$

The variance ratio is increasing with respect to τ and there is a potential for accuracy gains by changing the estimand from SATE to SATT. The simulation results in Figure (7) confirm the above derivation. Similarly to the random coefficient model the change of estimand is the main cause of the differences in length. Similarly to the previous simulations, as σ_τ^2 increases the differences between estimands become more stark and inference for SATT (SATE) has bad coverage w.r.t. SATE (SATT):

$$Y_i(0) \sim N(\mu = 0, \sigma_0^2 = 1) \tag{25}$$

Figure 7: **Censored outcome (Tobit) simulation results: Length and coverage differences**



Notes: See the notes in Figure (5).

6 Super-population model

In the super-population model the sampling procedure has two steps. First, a sample of N units is drawn from a super-population, $F_{Y(1),Y(0)}(\cdot)$, which can be either finite or infinite. Second, m units are randomly allocated to the treatment regime and the remaining $N - m$ units are assigned to the control regime.⁶ Both the treatment indicator and the potential outcomes are random variables, and they are independent, due to the random assignment of units to treatment regimes. SATE, SATT, and SATC are all random variables that are unbiased, and consistent, estimators of PATE. Under the super-population model the difference-in-means test statistic can be used to construct a CI for PATE. Neyman's variance estimator is a consistent estimator for the variance of the difference-in-means test statistic (Imbens and Rubin, 2015). PATE, PATT, and PATC are all equal and recentering the difference-in-means w.r.t. PATE is equivalent to recentering it w.r.t. PATT (or PATC). There are no efficiency gains or inaccuracies in making inference on PATT instead of PATE.

Next we describe the behavior of the difference-in-means when it is recentered w.r.t. different sample average treatment effect estimands (e.g., SATE, SATT). The variance of $t_{\text{diff}} - \text{SATE}$ is lower than that of $t_{\text{diff}} - \text{PATE}$:

$$\text{Var}(t_{\text{diff}} - \text{SATE}) = \text{Var}(t_{\text{diff}}) - \underbrace{\frac{1}{N} \cdot \sigma_\tau^2}_{\text{Var}(\text{SATE})} \leq \text{Var}(t_{\text{diff}}) = \text{Var}(t_{\text{diff}} - \text{PATE}) \quad (26)$$

and

$$\text{Var}(t_{\text{diff}}) = \frac{\sigma_1^2(\text{SP})}{m} + \frac{\sigma_0^2(\text{SP})}{N - m} \quad (27)$$

where $\sigma_j^2(\text{SP})$ ($j = 0, 1$) denotes the variance of units under treatment regime j in the super-population. See Appendix (H.1) for a proof of equation (26). In the fixed population case, SATE is a parameter and not a random variable, which changes when sampling uncertainty is introduced and the potential outcomes become random variables as well.

⁶The number of treated and control units is fixed: m is not a random variable.

Lemma (4) establishes that the variances calculations we derived in Section (4) hold also under the super-population framework. Although the variance calculations are over repeated samples, the correlation term between potential outcomes, ρ , does not go away and to conduct inference on the SATE all the issues we discussed in Section (4) arise here as well. More specifically, the uncertainty calculation over repeated draws of data cancels the ρ term in the variance of the difference-in-means test statistic. However, it also introduces a correlation term ρ in the variance of SATE, which is now a random variable. The two elements cancel each other out and yield a variance formula that is exactly the same as the one that was derived in Lemma (2).

Lemma 4. Under the super-population model, the variance of the difference-in-means when re-centered w.r.t. the SATE, SATT, or SATC is

$$\begin{aligned}\text{Var}(t_{\text{diff}} - \text{SATE}) &= \frac{1}{N \cdot (1-p) \cdot p} \cdot [p^2 \cdot \sigma_0^2(\text{SP}) + (1-p)^2 \cdot \sigma_1^2(\text{SP}) + 2p(1-p) \cdot \rho(\text{SP}) \cdot \sigma_0(\text{SP}) \cdot \sigma_1(\text{SP})] \\ \text{Var}(t_{\text{diff}} - \text{SATT}) &= \frac{1}{p \cdot (1-p) \cdot N} \cdot \sigma_0^2(\text{SP}) \\ \text{Var}(t_{\text{diff}} - \text{SATC}) &= \frac{1}{p \cdot (1-p) \cdot N} \cdot \sigma_1^2(\text{SP})\end{aligned}$$

See Appendix (H.1) for the proof.

7 Comments

Several notes on the previous results and possible extensions are in order.

Remark 1. The previous results can be extended to include covariate adjustment of pre-treatment characteristics according to the procedure that was proposed by Rosenbaum (2002). Denote by X_i a $1 \times p$ dimensional vector of the pre-treatment characteristics of unit i . The matrix \mathbf{X} has dimensions $n \times p$ and each row i contains the pre-treatment characteristics of unit i . It is common to adjust Y_i for X_i for efficiency purposes. Define $Y_i^{\text{adjusted}} = X_i(X'X)^{-1}X'Y_i$ as the adjusted/residualized responses. All the results for inference on Y_i also apply for inference on Y_i^{adjusted} .

Remark 2. The inference results for SATT (and SATC) can also be extended to different randomization models. For example, Theorem 7.1 provides variance and limiting distribution

results for inference on SATT when the treatment assignment is done by random independent Bernoulli trials. This illustrates how inference on SATT can be derived under random treatment assignment models that differ from the classic complete randomization mechanism.

Theorem 7.1. (Limiting distribution of t_{diff} - SATT) When treatment is assigned according to random independent Bernoulli trials, the standardized and recentered, w.r.t. SATT, difference-in-means follows a standard Normal distribution under two regularity conditions. When the following is satisfied:

$$N - m \rightarrow \infty, \quad m \rightarrow \infty, \quad \text{and} \quad \sigma_1^2, \sigma_0^2 < \infty \quad (28)$$

then:

$$\frac{\frac{N-m}{N} \cdot (t_{\text{diff}} - \text{SATT})}{\sqrt{\text{Var} \left(\frac{1}{m} \cdot \sum_{i=1}^N Y_i(0) \cdot T_i \right)}} \xrightarrow{d} N(0, 1) \quad (29)$$

where

$$\text{Var} \left(\frac{1}{m} \cdot \sum_{i=1}^N Y_i(0) \cdot T_i \right) = \begin{pmatrix} \frac{1}{m} \\ -\frac{\sum_{i=1}^N Y_i(0) \cdot T_i}{m^2} \end{pmatrix}^T \cdot \Sigma \cdot \begin{pmatrix} \frac{1}{m} \\ -\frac{\sum_{i=1}^N Y_i(0) \cdot T_i}{m^2} \end{pmatrix} \quad (30)$$

and

$$\Sigma = \begin{pmatrix} N \cdot p(1-p) [\sigma_0^2 + \mathbb{E}[Y(0)]^2] & p(1-p) \cdot N \cdot \mathbb{E}[Y(0)] \\ p(1-p) \cdot N \cdot \mathbb{E}[Y(0)] & N \cdot p(1-p) \end{pmatrix} \quad (31)$$

See Appendix (J) for the proof.

Remark 3. The regularity conditions of the above theoretical results will usually be satisfied in applied research, yet it is important to understand when they will not. For example, imagine a control regime in which all the individuals die ($Y_i(0) = 0 \quad \forall_i$), while under the treatment regime units have a strictly positive survival probability, $\mathbb{E}[Y(1)] = p_1$ and $\text{Var}(Y(1)) = p_1(1 - p_1)$. As the variance of the control units is strictly lower (zero) than that of the treated units, there are

potential accuracy gains from estimating SATT instead of SATE. In this scenario as $\sigma_0^2 = 0$ the PI for SATT contains only one point, the difference-in-means estimate, which is clearly wrong. The example above does not stand in contradiction to Theorem (4.2), as in the above case the regularity condition of the theorem is not satisfied:

$$\frac{\max_{1 \leq i \leq N} (Y(0)_{Ni} - \bar{Y}(0)_N)^2}{\sum_{i=1}^N (Y(0)_{Ni} - \bar{Y}(0)_N)^2} = \frac{0}{0}$$

and $\frac{0}{0}$ is not a well-defined expression.

8 Real data application: Online experiments

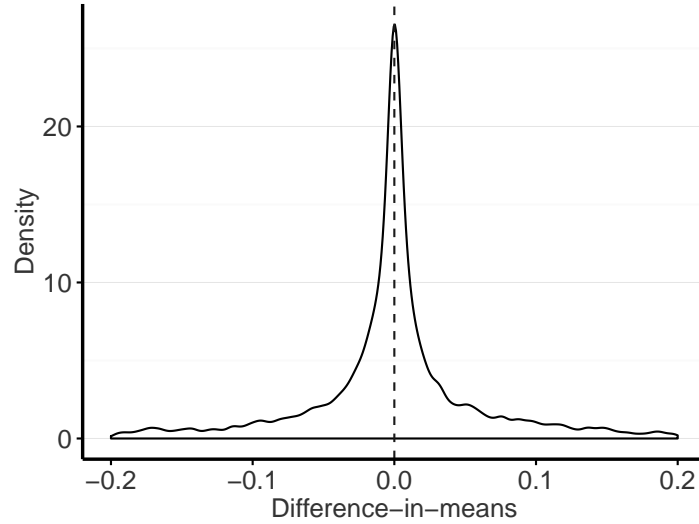
To better understand the trade-offs in conducting inference on different average treatment effect estimands, we analyze a sample of online field experiments that have been conducted by a large internet firm as product improvement tests. Our sample consists of 278 experiments with an average sample size of approximately one hundred million units per experiment. Many different outcome metrics are analyzed for various subgroups. The average subgroup consists of 1.1 million observations, and there are 826 unique outcome metrics across all of the experiments. In total, twenty-five thousand different treatment effects are estimated. The data analyzed were aggregated and de-identified.

Figure (8) shows the distribution of the difference-in-means across all of the online experiments. It is clear that on average, across experiments, the treatment had a zero effect. The distribution of difference-in-means is tightly centered around zero. However, this does not imply that the treatment had no effect. Treatment effect heterogeneity can generate positive effects for some units and negative effects for others that cancel each other on average. Elaborate tests and computations are difficult to carry out with millions of observations. Inference on SATO provides a simple solution as it is sensitive to heterogeneous treatment effects and is guaranteed to have correct coverage and valid inference on SATO. Another advantage of SATO is that for the simple cases of $\rho \in \{-1, 0, 1\}$, the change of estimand from SATE to SATO does not require any additional information for inference. It can be carry out using only aggregate summary statistics that are

already collected for inference on SATE.

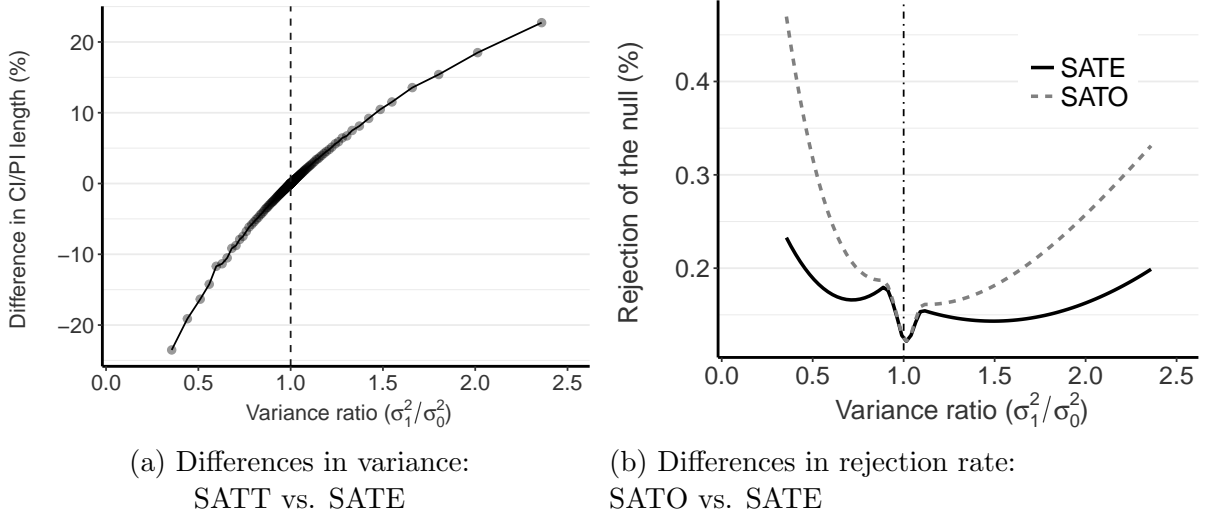
The left plot of Figure (9) replicates Figure (2) in a real empirical application and demonstrates that variance differences across experiments exist. As the variance ratio $\frac{\sigma_1^2}{\sigma_0^2}$ increases, the PI for SATT becomes shorter, and in some cases there can be large variance gains from changing the estimand to SATO. The left plot of Figure (9) shows the share of experiments in which the null hypothesis that the estimand is equal to zero was rejected, as a function of the variance ratio. As the variance ratio departs from one there are substantial efficiency gains to be had from conducting inference on SATO (with $\rho = 1$) relative to SATE. The differences in rejection rates between inference on SATE and SATO are evidence of treatment effect heterogeneity.

Figure 8: **Distribution of the difference-in-means across experiments**



Notes: The figure shows the distribution of the raw difference-in-means over all online experiments in our sample.

Figure 9: **Inference accuracy: A comparison of SATO and SATT to SATE across online experiments**



Notes: The left plot describes the empirical analogue of Figure (2). It plots the relationship between the variance ratio in the experiment and the gain in accuracy from conducting inference on SATT relative to SATE. The right plot compares the rejection rate of the null hypothesis that the estimand of interest is equal to zero for SATE and SATO when $\rho = 1$.

9 Discussion

In some applications, the estimand of interest is SATT (or SATC). Making inferences about SATT (or SATC) using a CI for SATE relies on variance estimators that are not consistent and are not guaranteed to have correct coverage or be efficient. We derive efficient variance formulas for inference on a new and general class of estimands derived from any mixing between SATT and SATC. The variance formulas are used to construct PIs that are non-parametrically guaranteed to have correct coverage and to be non-conservative – unlike inference on SATE. All inference procedures discussed in the paper use the difference-in-means as the test statistic, and therefore have the same point estimates as existing methods. Note that all three estimands, SATE, SATT, and SATC, are equal in expectation. The key difference is in the variance calculations. In addition, we present a new diagnostic tool, a decomposition of the uncertainty about the value of SATE into two components: uncertainty about SATT and about SATC, which both have an intuitive interpretation.

Taken together, the Monte Carlo simulations demonstrate that: (i) the choice of estimand has a direct implication on the accuracy of the inference that can be conducted; (ii) using variance

formulas that have correct size for SATE will *not* have correct coverage of other sample average treatment effect estimands such as SATT or SATO. When a researcher is interested in conducting inference on SATT, she is required to use variance formulas that have correct coverage for this estimand. Such formulas are derived in the paper.

The application to online experiments is a clear example of a scenario in which the methods developed in the paper can provide accuracy gains over standard practice. The gains mainly arise because inference on SATO is sensitive to treatment effect heterogeneity through the variance calculations. The large potential for accuracy gains emphasizes that researchers and other applied users should think carefully about which causal estimand they want to conduct inference on. If the answer is SATE, then they should not use any of the results in this paper; however, if SATO is also of interest, then large accuracy gains can be obtained by a change of estimand. In addition, if SATT is the primary object of interest, then variance formulas for inference on SATE will have incorrect coverage of SATT.

Our results can be used in a variety of empirical applications from medicine and social sciences to online ad experiments. In many cases one would like to have more accuracy when testing the null hypothesis of no average treatment effect. This occurs even when, as in the case of digital experiments, the sample size is large, but the power is limited ([Lewis and Rao, 2014](#)). The accuracy gains we document hold even asymptotically, and could be of special value in massive experiments because our approach requires computing only summary statistics that are simple to obtain and easy to compute, and lead to possibly large gains in accuracy.

References

- Aronow, Peter M, Donald P Green, and Donald KK Lee**, “Sharp bounds on the variance in randomized experiments,” *The Annals of Statistics*, 2014, *42* (3), 850–871.
- Bowers, Jake. and Ben B. Hansen**, “Attributing Effects to A Cluster Randomized Get-Out-The-Vote Campaign,” *Journal of the American Statistical Association*, 2009, *104* (487), 873–885.
- Crump, Richard K., V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik**, “Dealing with limited overlap in estimation of average treatment effects,” *Biometrika*, 2009, *96* (1), 187–199.
- Ding, Peng, Avi Feller, and Luke Miratrix**, “Randomization inference for treatment effect variation,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2016, *78* (3), 655–671.
- Feng, Xingdong, Yang Feng, Yuguo Chen, and Dylan S. Small**, “Randomization inference for the trimmed mean of effects attributable to treatment,” *Statistica Sinica*, 2014, *24* (2), 773–797.
- Holland, P.**, “Statistics and Causal Inference (with discussion),” *Journal of the American Statistical Association*, 1986, *86*, 945–970.
- Imbens, Guido and Donald Rubin**, “Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction,” *Cambridge University Press*, 2015.
- Imbens, Guido W.**, “Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review,” *The Review of Economics and Statistics*, 2004, *86* (1), 4–29.
- Keele, Luke, Dylan Small, and Richard Grieve**, “Randomization-based instrumental variables methods for binary outcomes with an application to the IMPROVE trial,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2017, *180* (2), 569–586.
- Lewis, Randall A and Justin M Rao**, “The unfavorable economics of measuring the returns to advertising,” *Available at SSRN 2367103*, 2014.
- Li, Fan, Kari Lock Morgan, and Alan M. Zaslavsky**, “Balancing Covariates via Propensity Score Weighting,” *Journal of the American Statistical Association*, 2016, *0*, 0–0.
- Li, Xinran and Peng Ding**, “General forms of finite population central limit theorems with applications to causal inference,” Technical Report 2016.
- Neyman, Jerzy**, “On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9,” *Statistical Science*, 1923, 1990, *5* (4), 465–472. Trans. Dorota M. Dabrowska and Terence P. Speed.
- Reichardt, Charles and Harry Gallob**, “Justifying the use and increasing the power of a t test for a randomized experiment with a convenience sample,” *Psychological Methods*, 1999, *4* (1), 117–128.
- Rigdon, Joseph and Michael G. Hudgens**, “Randomization inference for treatment effects on a binary outcome,” *Statistics in medicine*, 2015, *34* (6).

- Robins, James M.**, “Confidence Intervals for Causal Parameters,” *Statistics in Medicine*, 1988, 7, 773–785.
- Rosenbaum, P.**, “Effects Attributable to Treatment: Inference in Experiments and Observational Studies with a Discrete Pivot,” *Biometrika*, 2001, 88 (1), 219–231.
- Rosenbaum, Paul R.**, “Covariance Adjustment in Randomized Experiments and Observational Studies,” *Statistical Science*, 2002, 17 (3), 286–304.
- Tunca, T. and U. Egeli**, “Cytogenetic Findings on Shoe Workers Exposed Long-Term to Benzene,” *Environmental Health Perspectives*, 1996, 104, 13131317.

Online appendix: Proofs

A Derivation of MSE(SATE,SATT)

$$\begin{aligned}
\text{MSE}(\text{SATE}, \text{SATT}) &= \text{MSE}(t_{\text{diff}} - \text{SATE}, t_{\text{diff}} - \text{SATT}) \\
&= [\mathbb{E}[\text{SATE} - \text{SATT}]]^2 + \text{Var}(\text{SATT}) \\
&= \text{Var}(\text{SATT}) \\
&= \left(\frac{1}{m} - \frac{1}{N} \right) \cdot \sigma_\tau^2 \\
&= \frac{N-m}{Nm} \cdot \text{Var}(\tau_i) = \frac{1-p}{m} \cdot \sigma_\tau^2
\end{aligned}$$

B Proof of Lemma 1

The sum of the responses of the treated units can be decomposed into two components, the sum of the potential outcomes under control and the sum of the treatment effects.

$$\sum_{i=1}^N Y_i \cdot T_i = \sum_{i=1}^N T_i \cdot Y_i(1) = \sum_{i=1}^N T_i \cdot Y_i(0) + \sum_{i=1}^N T_i \cdot (Y_i(1) - Y_i(0)) \quad (32)$$

$$\Rightarrow \frac{1}{m} \cdot \sum_{i=1}^N Y_i \cdot T_i = \frac{1}{m} \cdot \sum_{i=1}^N T_i \cdot Y_i(0) + \text{SATT} \quad (33)$$

We can re-write the difference in means as,

$$\begin{aligned}
t_{\text{diff}} &= \frac{1}{m} \cdot \sum_{i=1}^N Y_i \cdot T_i - \frac{1}{N-m} \cdot \sum_{i=1}^N Y_i \cdot (1 - T_i) \\
&= \frac{1}{m} \cdot \sum_{i=1}^N T_i \cdot Y_i(0) + \text{SATT} - \frac{1}{N-m} \cdot \sum_{i=1}^N Y_i(0) \cdot (1 - T_i) \\
&= \frac{1}{m} \cdot \sum_{i=1}^N T_i \cdot Y_i(0) - \frac{1}{N-m} \cdot \sum_{i=1}^N Y_i(0) \cdot (1 - T_i) + \text{SATT} \\
&= \frac{N}{m \cdot (N-m)} \cdot \sum_{i=1}^N Y_i(0) T_i - \underbrace{\frac{1}{N-m} \cdot \sum_{i=1}^N Y_i(0)}_{\text{constant}} + \text{SATT}
\end{aligned} \quad (34)$$

The decomposition of t_{diff} as a function of $Y(1)$ and the SATC can be proved using a similar proof to the one above.

C Proof of Lemma 2

The variance of $t_{\text{diff}} - \text{SATE}$ can be written as,

$$\begin{aligned}
\text{Var}(t_{\text{diff}} - \text{SATE}) &= \frac{m}{N \cdot (N - m)} \sigma_0^2 + \frac{N - m}{N \cdot m} \sigma_1^2 + \frac{2}{N} \cdot \rho \cdot \sigma_0 \cdot \sigma_1 \\
&= \frac{p}{N \cdot (1 - p)} \cdot \sigma_0^2 + \frac{N \cdot (1 - p)}{N^2 \cdot p} \cdot \sigma_1^2 + \frac{2}{N} \cdot \rho \cdot \sigma_0 \cdot \sigma_1 \\
&= \frac{p}{N \cdot (1 - p)} \cdot \sigma_0^2 + \frac{(1 - p)}{N \cdot p} \cdot \sigma_1^2 + \frac{2}{N} \cdot \rho \cdot \sigma_0 \cdot \sigma_1 \\
&= \frac{1}{N \cdot (1 - p) \cdot p} \cdot [p^2 \cdot \sigma_0^2 + (1 - p)^2 \cdot \sigma_1^2 + 2p(1 - p) \cdot \rho \cdot \sigma_0 \cdot \sigma_1]
\end{aligned}$$

The variance derivations of $t_{\text{diff}} - \text{SATT}$ are described in [E](#), and the variance calculation of $t_{\text{diff}} - \text{SATC}$ follows exactly the same steps as the variance calculation of $t_{\text{diff}} - \text{SATT}$.

D Proof of Theorem 4.1

Proof of part (1) of Theorem 4.1

The proof is a simple implementation of the intermediate value theorem. We show that the difference in variance is negative when $\rho = -1$, and is positive when $\rho = 1$, and that the difference in variance is a continuous and increasing function with respect to ρ . Then the desired result follows immediately from the intermediate value theorem.

When $\rho = -1$, the variance of $t_{\text{diff}} - \text{SATE}$ can be written as:

$$\begin{aligned}
\text{Var}(t_{\text{diff}} - \text{SATE}) &= \frac{m}{N \cdot (N - m)} \sigma_0^2 + \frac{N - m}{N \cdot m} \sigma_1^2 + \frac{2}{N} \cdot \rho \cdot \sigma_0 \cdot \sigma_1 \\
&= \frac{p}{N \cdot (1 - p)} \cdot \sigma_0^2 + \frac{N \cdot (1 - p)}{N^2 \cdot p} \cdot \sigma_1^2 - \frac{2}{N} \cdot \sigma_0 \cdot \sigma_1 \\
&= \frac{p}{N \cdot (1 - p)} \cdot \sigma_0^2 + \frac{(1 - p)}{N \cdot p} \cdot \sigma_1^2 - \frac{2}{N} \cdot \sigma_0 \cdot \sigma_1 \\
&= \frac{1}{N \cdot (1 - p) \cdot p} \cdot [p^2 \cdot \sigma_0^2 + (1 - p)^2 \cdot \sigma_1^2 - 2p(1 - p) \cdot \sigma_0 \cdot \sigma_1] \\
&= \frac{1}{N \cdot (1 - p) \cdot p} \cdot (p \cdot \sigma_0 - (1 - p) \cdot \sigma_1)^2
\end{aligned}$$

and as,

$$p \cdot \sigma_0 < \sigma_0 - (1 - p) \cdot \sigma_1 < \sigma_0$$

it follows that when $\rho = -1$, then $\text{Var}(t_{\text{diff}} - \text{SATE}) < \text{Var}(t_{\text{diff}} - \text{SATT})$. When $\rho = 1$, the variance of $t_{\text{diff}} - \text{SATE}$ can be written as:

$$\begin{aligned}
\text{Var}(t_{\text{diff}} - \text{SATE}) &= \frac{p}{N \cdot (1 - p)} \cdot \sigma_0^2 + \frac{N \cdot (1 - p)}{N^2 \cdot p} \cdot \sigma_1^2 + \frac{2}{N} \cdot \sigma_0 \cdot \sigma_1 \\
&= \frac{p}{N \cdot (1 - p)} \cdot \sigma_0^2 + \frac{(1 - p)}{N \cdot p} \cdot \sigma_1^2 + \frac{2}{N} \cdot \sigma_0 \cdot \sigma_1 \\
&= \frac{1}{N \cdot (1 - p) \cdot p} \cdot [p^2 \cdot \sigma_0^2 + (1 - p)^2 \cdot \sigma_1^2 + 2p(1 - p) \cdot \sigma_0 \cdot \sigma_1] \\
&= \frac{1}{N \cdot (1 - p) \cdot p} \cdot (p \cdot \sigma_0 + (1 - p) \cdot \sigma_1)^2
\end{aligned}$$

and as $\sigma_1 > \sigma_0$, it follows immediately that when $\rho = 1$, $\text{Var}(t_{\text{diff}} - \text{SATE}) > \text{Var}(t_{\text{diff}} - \text{SATT})$.

As $\frac{\partial \text{Var}(t_{\text{diff}} - \text{SATE}) - \text{Var}(t_{\text{diff}} - \text{SATT})}{\partial \rho} > 0$, it follows directly from the intermediate value theorem

that there exists a value of ρ , $\bar{\rho}$, such that:

$$\rho \leq \bar{\rho} \Rightarrow \text{Var}(t_{\text{diff}} - \text{SATE}) \leq \text{Var}(t_{\text{diff}} - \text{SATT})$$

$$\rho > \bar{\rho} \Rightarrow \text{Var}(t_{\text{diff}} - \text{SATE}) > \text{Var}(t_{\text{diff}} - \text{SATT})$$

Proof of part (2) of Theorem 4.1

Consider the case in which $\rho = 0$:

$$\text{Var}(t_{\text{diff}} - \text{SATE}) = \frac{1}{N \cdot (1-p) \cdot p} \cdot [p^2 \cdot \sigma_0^2 + (1-p)^2 \cdot \sigma_1^2]$$

$\text{Var}(t_{\text{diff}} - \text{SATT})$ will be lower than $\text{Var}(t_{\text{diff}} - \text{SATE})$, when:

$$p^2 \cdot \sigma_0^2 + (1-p)^2 \cdot \sigma_1^2 \geq \sigma_0^2 \Rightarrow \frac{(1-p)^2}{1-p^2} \cdot \sigma_1^2 \geq \sigma_0^2$$

Note, $\frac{(1-p)^2}{1-p^2} = \frac{1+p^2-2p}{1-p^2} \leq 1 \Rightarrow p^2 - 2p \leq -p^2 \Rightarrow 2p^2 \leq 2p \Rightarrow p^2 \leq p$, which is always satisfied. Therefore, $0 \leq \frac{(1-p)^2}{1-p^2} \leq 1$. To conclude, when $\frac{\sigma_1}{\sigma_0} > \sqrt{\frac{1-p^2}{(1-p)^2}}$ the variance difference will be positive $\text{Var}(t_{\text{diff}} - \text{SATE}) > \text{Var}(t_{\text{diff}} - \text{SATT})$. As the variance difference is negative when $\rho = -1$ it follows directly from the intermediate value theorem that the desired $\bar{\rho}$ exists and is strictly larger than -1 and lower than 0 — $\bar{\rho}$ is negative. This concludes the proof.

E Proof of Theorem 4.4

According to Lemma 1, the adjusted difference in means, $\frac{N-m}{N} \cdot (t_{\text{diff}} - \text{SATT})$, has the same distribution of $\frac{1}{m} \cdot \sum_{i=1}^N Y_i(0) \cdot T_i$ with an additive shift of $\frac{1}{N} \cdot \sum_{i=1}^N Y_i(0)$. [?, ?, ?, and Li and Ding \(2016\)](#) all provide proofs showing that in a finite population model with complete randomization the standardized mean of the treated (or sampled) units will follow a standard Normal distribution under two regularity conditions,

$$N - m \rightarrow \infty, \quad \text{and} \quad m \rightarrow \infty \tag{35}$$

and

$$\frac{\max_{1 \leq i \leq N} (Y(0)_{Ni} - \bar{Y}(0)_N)^2}{\sum_{i=1}^N (Y(0)_{Ni} - \bar{Y}(0)_N)^2} \cdot \max\left(\frac{N-m}{m}, \frac{m}{N-m}\right) \rightarrow \infty \tag{36}$$

The expectation and variance of $\frac{N-m}{N} \cdot (t_{\text{diff}} - \text{SAT})$ are derived next and complete the proof of (4.2). The expectation is,

$$\begin{aligned}\mathbb{E} \left[\frac{N-m}{N} \cdot (t_{\text{diff}} - \text{SAT}) \right] &= \frac{N-m}{N} \cdot \left[\frac{N}{m(N-m)} \cdot \sum_{i=1}^N \mathbb{E}[Y(0)] \cdot \mathbb{E}[T] - \frac{1}{N} \cdot \sum_{i=1}^N \mathbb{E}[Y(0)] \right] \\ &= \frac{N}{m} \cdot \mathbb{E}[Y(0)] \cdot \frac{m}{N} - \frac{1}{N} \cdot N \cdot \mathbb{E}[Y(0)] \\ &= 0\end{aligned}$$

where the first equality follows from Lemma 1. The variance of the adjusted difference in means is,

$$\begin{aligned}\text{Var} \left(\frac{N-m}{N} \cdot (t_{\text{diff}} - \text{SAT}) \right) &= \text{Var} \left(\frac{1}{m} \cdot \sum_{i=1}^N Y_i(0) \cdot T_i - \frac{1}{N} \cdot \sum_{i=1}^N Y_i(0) \right) \\ &= \frac{1}{m^2} \cdot \text{Var} \left(\sum_{i=1}^N Y_i(0) \cdot T_i \right)\end{aligned}$$

where the first equality follows from Lemma 1, and the second equality follows as $\frac{1}{N} \cdot \sum_{i=1}^N Y_i(0)$ is a constant and not a random variable. The variance of the sum $\sum_{i=1}^N Y_i(0) \cdot T_i$, is the variance of m units who are randomly sampled from the population $Y_{N1}(0), \dots, Y_{NN}(0)$.⁷ Denote the m chosen variables by V_1, \dots, V_m , and the variance of the sum $V_1 + \dots + V_m$ is,

$$\text{Var}(V_1 + \dots + V_m) = \sum_{i=1}^m \text{Var}(V_i) + \sum_{i=1}^m \sum_{j \neq i}^m \text{Cov}(V_i, V_j) = m \cdot \sigma_0^2 + m(m-1) \cdot \rho \quad (37)$$

where the second equality follows as the covariance between each two units is the same and is denoted by ρ . When $m = N$, the variance of the sum is zero and therefore,

$$\rho = -\frac{\sigma_0^2}{N-1} \quad (38)$$

⁷Our calculation of the variance of $\frac{1}{N} \cdot \sum_{i=1}^N Y_i(0)$ follows the variance calculation of the Wilcoxon Rank Sum Test statistic in ?.

Substituting (38) in (37) and re-arranging yields that:

$$\begin{aligned}
\text{Var} \left(\frac{N-m}{N} \cdot (t_{\text{diff}} - \text{SATT}) \right) &= \text{Var} \left(\frac{1}{m} \cdot \sum_{i=1}^N Y_i(0) \cdot T_i - \frac{1}{N} \cdot \sum_{i=1}^N Y_i(0) \right) \\
&= \frac{1}{m^2} \cdot \text{Var} \left(\sum_{i=1}^N Y_i(0) \cdot T_i \right) = \frac{1}{m^2} \cdot \frac{m(N-m)}{N-1} \cdot \sigma_0^2 \\
&= \frac{N-m}{m(N-1)} \cdot \sigma_0^2 \\
&\approx \frac{N-m}{mN} \cdot \sigma_0^2 = \left(\frac{1}{m} - \frac{1}{N} \right) \cdot \sigma_0^2
\end{aligned}$$

Hence, it follows that:

$$\begin{aligned}
\text{Var}((t_{\text{diff}} - \text{SATT})) &= \left(\frac{N}{N-m} \right)^2 \cdot \text{Var} \left(\frac{N-m}{N} \cdot (t_{\text{diff}} - \text{SATT}) \right) \\
&= \left(\frac{N}{N-m} \right)^2 \cdot \left(\frac{N-m}{mN} \right) \cdot \sigma_0^2 \\
&= \frac{N}{(N-m) \cdot m} \cdot \sigma_0^2 \\
&= \frac{1}{Np(1-p)} \cdot \sigma_0^2
\end{aligned}$$

F Proof of Theorem (4.3)

The first part of the theorem, that $[p \cdot L_{\text{SATT}} + (1-p) \cdot L_{\text{SATC}}, \quad p \cdot U_{\text{SATT}} + (1-p) \cdot U_{\text{SATC}}]$, follows immediately from the proof of Theorem (1) in [Rigdon and Hudgens \(2015\)](#). Next we show the derivations of the second part of the theorem.

$$\begin{aligned}
p \cdot L_{\text{SATT}} + (1-p) \cdot L_{\text{SATC}} &= t_{\text{diff}} - z_{1-\alpha/2} \cdot \sqrt{k(N, m)} \cdot (p\sigma_0 + (1-p)\sigma_1) \\
&= t_{\text{diff}} - z_{1-\alpha/2} \cdot \sqrt{k(N, m)} \cdot \sqrt{\hat{\mathbb{V}}_{\rho=1}}
\end{aligned}$$

and

$$\begin{aligned} p \cdot U_{\text{SATT}} + (1 - p) \cdot U_{\text{SATC}} &= t_{\text{diff}} + z_{1-\alpha/2} \cdot \sqrt{k(N, m)} \cdot (p\sigma_0 + (1 - p)\sigma_1) \\ &= t_{\text{diff}} + z_{1-\alpha/2} \cdot \sqrt{k(N, m)} \cdot \sqrt{\hat{\mathbb{V}}_{\rho=1}} \end{aligned}$$

The second equality in both of the above equations follows from substituting $\rho = 1$ in the variance formula for $\text{Var}(t_{\text{diff}} - \text{SATE})$ in Lemma (2).

G Proof of Theorem 4.4

The desired result follows immediately from a comparison of the different variance expressions in Lemma 2.

H Super-population model

H.1 Proof of Lemma 4

$$\text{Var}(t_{\text{diff}} - \text{PATE}) = \text{Var}(t_{\text{diff}}) = \frac{\sigma_0^2(\text{SP})}{N - m} + \frac{\sigma_1^2(\text{SP})}{m}$$

see [Imbens and Rubin \(2015\)](#) for proof.

Next we derive the variance of $t_{\text{diff}} - \text{SATE}$:

$$\begin{aligned} \text{Var}(t_{\text{diff}} - \text{SATE}) &= \mathbb{E}[(t_{\text{diff}} - \text{SATE})^2] - (t_{\text{diff}} - \text{SATE})^2 \\ &= \mathbb{E}[(t_{\text{diff}} - \text{SATE})^2] \\ &= \mathbb{E}[t_{\text{diff}}^2] + \mathbb{E}[\text{SATE}^2] - 2 \cdot \mathbb{E}[t_{\text{diff}} \cdot \text{SATE}] \\ &= \mathbb{E}[t_{\text{diff}}^2] + \mathbb{E}[\text{SATE}^2] - 2 \cdot (\text{Var}(\text{SATE}) + \text{PATE}^2) \\ &= \text{Var}(t_{\text{diff}}) + \text{Var}(\text{SATE}) - 2 \cdot \text{Var}(\text{SATE}) \\ &= \text{Var}(t_{\text{diff}}) - \text{Var}(\text{SATE}) \end{aligned}$$

Note that,

$$\begin{aligned}
\mathbb{E}[t_{\text{diff}} \cdot \text{SATE}] &= \mathbb{E}_{\text{SP}} [\mathbb{E}_{\text{T}} [t_{\text{diff}} \cdot \text{SATE} | (\mathbf{Y}(\mathbf{1}), \mathbf{Y}(\mathbf{0}))]] \\
&= \mathbb{E}_{\text{SP}} [\text{SATE}^2] \\
&= \text{Var}(\text{SATE}) + [\mathbb{E}_{\text{SP}} [\text{SATE}]]^2 \\
&= \text{Var}(\text{SATE}) + \text{PATE}^2
\end{aligned}$$

The variance of SATE is:

$$\text{Var}(\text{SATE}) = \frac{1}{N} \cdot \text{Var}(Y_i(1) - Y_i(0)) = \frac{1}{N} \cdot (\sigma_1^2(\text{SP}) + \sigma_0^2(\text{SP}) - 2\rho(\text{SP})\sigma_1(\text{SP})\sigma_0(\text{SP}))$$

Therefore the $\text{Var}(t_{\text{diff}} - \text{SATE})$ is:

$$\begin{aligned}
\text{Var}(t_{\text{diff}} - \text{SATE}) &= \frac{1}{m} \cdot \sigma_1^2(\text{SP}) + \frac{1}{N-m} \cdot \sigma_0^2(\text{SP}) - \frac{1}{N} \cdot (\sigma_1^2(\text{SP}) + \sigma_0^2(\text{SP}) - 2\rho(\text{SP})\sigma_1(\text{SP})\sigma_0(\text{SP})) \\
&= \frac{N-m}{Nm} \cdot \sigma_1^2(\text{SP}) + \frac{m}{(N-m)N} \cdot \sigma_0^2(\text{SP}) + \frac{2\rho(\text{SP})\sigma_1(\text{SP})\sigma_0(\text{SP})}{N} \\
&= \frac{1-p}{Np} \cdot \sigma_1^2(\text{SP}) + \frac{p}{N(1-p)} \cdot \sigma_0^2(\text{SP}) + \frac{2\rho(\text{SP})\sigma_1(\text{SP})\sigma_0(\text{SP})}{N} \\
&= \frac{1}{Np(1-p)} \cdot [(1-p)^2 \cdot \sigma_1^2(\text{SP}) + p^2 \cdot \sigma_0^2(\text{SP})] + \frac{2\rho(\text{SP})\sigma_1(\text{SP})\sigma_0(\text{SP})}{N} \\
&= \frac{1}{N \cdot (1-p) \cdot p} \cdot [p^2 \cdot \sigma_0^2(\text{SP}) + (1-p)^2 \cdot \sigma_1^2(\text{SP}) + 2p(1-p) \cdot \rho(\text{SP}) \cdot \sigma_0(\text{SP}) \cdot \sigma_1(\text{SP})]
\end{aligned}$$

Next we derive the variance of $t_{\text{diff}} - \text{SATT}$ under the super-population model.

$$\begin{aligned}
\text{Var} \left(\frac{N-m}{N} \cdot (t_{\text{diff}} - \text{SATT}) \right) &= \mathbb{E}_{Y(1), Y(0)} \left[\text{Var}_T \left(\frac{N-m}{N} \cdot (t_{\text{diff}} - \text{SATT}) \mid Y(1), Y(0) \right) \right] \\
&\quad + \text{Var}_{Y(1), Y(0)} \left(\mathbb{E}_T \left[\frac{N-m}{N} \cdot (t_{\text{diff}} - \text{SATT}) \mid Y(1), Y(0) \right] \right) \\
&= \mathbb{E}_{Y(1), Y(0)} \left[\text{Var}_T \left(\frac{N-m}{N} \cdot (t_{\text{diff}} - \text{SATT}) \mid Y(1), Y(0) \right) \right] \\
&= \mathbb{E}_{Y(1), Y(0)} \left[\frac{N-m}{m(N-1)} \cdot \sigma_0^2(\text{FS}) \right] \\
&= \frac{N-m}{m(N-1)} \cdot \mathbb{E}_{Y(1), Y(0)} [\sigma_0^2(\text{FS})] \\
&= \frac{N-m}{m(N-1)} \cdot \sigma_0^2(\text{SP})
\end{aligned}$$

where the first equality follows from the law of total variance (i.e., law of conditional variance). There are a few more technical steps to complete the derivation. We neglect them as they already appeared in detail in Appendix C. The calculation of the variance of $t_{\text{diff}} - \text{SATC}$ is similar to the one above.

I Proof: Variance of $(t_{\text{diff}} - \text{SATO})$

Consider the following class of average treatment effect estimands:

$$\text{SATO} = \omega \cdot \text{SATT} + (1 - \omega) \cdot \text{SATC} \tag{39}$$

We discuss inferences on this class of estimands using the difference-in-means test statistic.

$$\text{Var}(t_{\text{diff}} - \text{SATO}) = \text{Var}(t_{\text{diff}}) + \text{Var}(\text{SATO}) - 2 \cdot \text{Cov}(\text{SATO}, t_{\text{diff}})$$

According to the derivations that are detailed later on in this section we can re-write the variance of t_{diff} re-centered w.r.t SATO as:

$$\begin{aligned}\text{Var}(t_{\text{diff}} - \text{SATO}) &= \frac{1}{Np(1-p)} [p^2\sigma_0^2 + (1-p)^2\sigma_1^2 + 2p(1-p) \cdot \rho\sigma_0\sigma_1] \\ &\quad + \frac{\sigma_\tau^2}{N-1} \cdot \left[\omega^2 \cdot \frac{1-p}{p} + (1-\omega)^2 \cdot \frac{p}{1-p} - 2\omega(1-\omega) \right] \\ &\quad - 2 \cdot \left[\frac{\omega}{(N-1)p} [\sigma_1^2 - \rho\sigma_1\sigma_0] - \frac{1}{N-1} \cdot \sigma_\tau^2 + \frac{(1-\omega)}{(N-1)(1-p)} [\sigma_0^2 - \rho\sigma_1\sigma_0] \right]\end{aligned}$$

Calculation of Var (SATO)

$$\begin{aligned}\text{Var}(\text{SATO}) &= \omega^2 \cdot \text{Var}(\text{SATT}) + (1-\omega)^2 \cdot \text{Var}(\text{SATC}) - 2 \cdot \omega(1-\omega) \cdot \text{Cov}(\text{SATT}, \text{SATC}) \\ &= \omega^2 \cdot \frac{N-m}{m(N-1)} \cdot \sigma_\tau^2 + (1-\omega)^2 \cdot \frac{m}{(N-m)(N-1)} \cdot \sigma_\tau^2 - \omega(1-\omega) \cdot \frac{2}{N-1} \cdot \sigma_\tau^2 \\ &= \frac{\sigma_\tau^2}{N-1} \cdot \left[\omega^2 \cdot \frac{1-p}{p} + (1-\omega)^2 \cdot \frac{p}{1-p} - 2\omega(1-\omega) \right]\end{aligned}$$

Calculation of Cov (SATO, t_{diff})

The covariance term is:

$$\begin{aligned}\text{Cov} \left(\omega \cdot \text{SATT} + (1-\omega) \cdot \text{SATC}, \frac{N}{m \cdot (N-m)} \cdot \sum_{i=1}^N Y_i(0) \cdot T_i - \frac{1}{N-m} \cdot \sum_{i=1}^N Y_i(0) + \text{SATT} \right) \\ = \omega \cdot \frac{N}{m \cdot (N-m)} \cdot \text{Cov} \left(\text{SATT}, \sum_{i=1}^N Y_i(0) \cdot T_i \right) + \omega \cdot \text{Var}(\text{SATT}) \\ + (1-\omega) \cdot \frac{N}{m \cdot (N-m)} \cdot \text{Cov} \left(\text{SATC}, \sum_{i=1}^N Y_i(0) \cdot T_i \right) + (1-\omega) \cdot \text{Cov}(\text{SATC}, \text{SATT})\end{aligned}$$

According to the derivations that are described in detail later on in this section we can re-write the above covariance expression as:

$$\begin{aligned}
\text{Cov}(\text{SATO}, t_{\text{diff}}) &= \omega \cdot \frac{N}{m \cdot (N-m)} \cdot \frac{(1-p)N}{N-1} \cdot [\rho\sigma_1\sigma_0 - \sigma_0^2] + \omega \cdot \frac{N-m}{m(N-1)} \cdot \sigma_\tau^2 \\
&\quad - (1-\omega) \cdot \frac{N}{m(N-m)} \cdot \frac{m}{N-1} \cdot [\rho\sigma_1\sigma_0 - \sigma_0^2] - (1-\omega) \cdot \frac{1}{N-1} \cdot \sigma_\tau^2 \\
&= \omega \cdot \frac{1}{p(N-1)} \cdot [\rho\sigma_1\sigma_0 - \sigma_0^2] + \omega \cdot \frac{(1-p)}{p(N-1)} \cdot \sigma_\tau^2 - \frac{1}{N-1} (1-\omega) \sigma_\tau^2 \\
&\quad - \frac{(1-\omega)}{(N-1)(1-p)} [\rho\sigma_1\sigma_0 - \sigma_0^2] \\
&= \frac{\omega}{(N-1)p} [\rho\sigma_1\sigma_0 - \sigma_0^2 + (1-p)\sigma_\tau^2 + p\sigma_\tau^2] - \frac{1}{N-1} \cdot \sigma_\tau^2 \\
&\quad - \frac{(1-\omega)}{(N-1)(1-p)} [\rho\sigma_1\sigma_0 - \sigma_0^2] \\
&= \frac{\omega}{(N-1)p} [\sigma_1^2 - \rho\sigma_1\sigma_0] - \frac{1}{N-1} \cdot \sigma_\tau^2 - \frac{(1-\omega)}{(N-1)(1-p)} [\rho\sigma_1\sigma_0 - \sigma_0^2]
\end{aligned}$$

Calculation of Var (SATT) and Var (SATC)

$$\text{Var}(\text{SATT}) = \frac{N-m}{m(N-1)} \cdot \sigma_\tau^2 = \frac{N-m}{m(N-1)} \cdot [\sigma_1^2 + \sigma_0^2 - 2\rho \cdot \sigma_1 \cdot \sigma_0]$$

Similarly,

$$\text{Var}(\text{SATC}) = \frac{m}{(N-m)(N-1)} \cdot \sigma_\tau^2$$

Calculation of Cov (SATC, $\sum_{i=1}^N Y_i(0) \cdot T_i$)

$$\begin{aligned}
\text{Cov}\left(\text{SATC}, \sum_{i=1}^N Y_i(0) \cdot T_i\right) &= -\frac{m}{N-m} \cdot \text{Cov}\left(\text{SATT}, \sum_{i=1}^N Y_i(0) \cdot T_i\right) \\
&= -\frac{m}{N-m} \cdot \frac{(1-p)N}{N-1} \cdot [\rho\sigma_1\sigma_0 - \sigma_0^2] \\
&= -\frac{m}{N-1} \cdot [\rho\sigma_1\sigma_0 - \sigma_0^2] \\
&\approx -p \cdot [\rho\sigma_1\sigma_0 - \sigma_0^2]
\end{aligned}$$

where the second equality follows from the derivations in the next subsection below.

Calculation of $\text{Cov}\left(\text{SATT}, \sum_{i=1}^N Y_i(0) \cdot T_i\right)$

$$\begin{aligned}
\text{Cov}\left(\text{SATT}, \sum_{i=1}^N Y_i(0) \cdot T_i\right) &= \frac{1}{m} \cdot \text{Cov}\left(\sum_{i=1}^N Y_i(1) \cdot T_i - \sum_{i=1}^N Y_i(0) \cdot T_i, \sum_{i=1}^N Y_i(0) \cdot T_i\right) \\
&= \frac{1}{m} \cdot \left[\text{Cov}\left(\sum_{i=1}^N Y_i(1) \cdot T_i, \sum_{i=1}^N Y_i(0) \cdot T_i\right) - \text{Var}\left(\sum_{i=1}^N Y_i(0) \cdot T_i\right) \right] \\
&= \frac{1}{m} \cdot \left[\frac{p(1-p)N^2}{N-1} \cdot \rho \cdot \sigma_1 \sigma_0 - \frac{m(N-m)}{N-1} \cdot \sigma_0^2 \right] \\
&= \frac{(1-p)N}{N-1} \cdot [\rho \sigma_1 \sigma_0 - \sigma_0^2]
\end{aligned}$$

Derivations of $\text{Cov}\left(\sum_{i=1}^N Y_i(1) \cdot T_i, \sum_{i=1}^N Y_i(0) \cdot T_i\right)$

$$\begin{aligned}
\mathbb{E}\left[\sum_{l=1}^N Y_l(1) \cdot T_l \cdot \sum_{i=1}^N Y_i(0) \cdot T_i\right] &= \mathbb{E}\left[\sum_{i=1}^N \sum_{l=1}^N Y_l(1) \cdot Y_i(0) \cdot T_i \cdot T_l\right] \\
&= \mathbb{E}\left[T_i \cdot \sum_{i=1}^N Y_i(0) \cdot Y_i(1) + \sum_{i=1}^N \sum_{l \neq i}^N Y_i(0) \cdot Y_l(1) \cdot T_i \cdot T_l\right] \\
&= \Pr(T_i = 1) \cdot \sum_{i=1}^N Y_i(0) \cdot Y_i(1) + \mathbb{E}[T_i \cdot T_l | i \neq l] \cdot \sum_{i=1}^N \sum_{l \neq i}^N Y_i(0) \cdot Y_l(1) \\
&= \frac{m}{N} \cdot \sum_{i=1}^N Y_i(0) \cdot Y_i(1) + \frac{m}{N} \cdot \frac{m-1}{N-1} \cdot \sum_{i=1}^N \sum_{l \neq i}^N Y_i(0) \cdot Y_l(1) \\
&= \frac{m}{N} \cdot \frac{N-m}{N-1} \cdot \sum_{i=1}^N Y_i(0) \cdot Y_i(1) + \frac{m}{N} \cdot \frac{m-1}{N-1} \cdot \sum_{i=1}^N \sum_{l=1}^N Y_i(0) \cdot Y_l(1)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\sum_{l=1}^N Y_l(1) \cdot T_l\right] &= \frac{m}{N} \sum_{l=1}^N Y_l(1) \\
\mathbb{E}\left[\sum_{l=1}^N Y_l(0) \cdot T_l\right] &= \frac{m}{N} \sum_{l=1}^N Y_l(0) \\
\Rightarrow \mathbb{E}\left[\sum_{l=1}^N Y_l(1) \cdot T_l\right] \cdot \mathbb{E}\left[\sum_{i=1}^N Y_i(0) \cdot T_i\right] &= \left(\frac{m}{N}\right)^2 \cdot \sum_{i=1}^N \sum_{l=1}^N Y_i(0) \cdot Y_l(1)
\end{aligned}$$

Hence,

$$\begin{aligned}
\text{Cov} \left(\sum_{i=1}^N Y_i(1) \cdot T_i, \sum_{i=1}^N Y_i(0) \cdot T_i \right) &= \frac{m}{N} \cdot \left[\frac{N-m}{N-1} \cdot \sum_{i=1}^N Y_i(0) \cdot Y_i(1) - \frac{N-m}{N(N-1)} \cdot \sum_{i=1}^N \sum_{l=1}^N Y_i(0) \cdot Y_l(1) \right] \\
&= \frac{m}{N} \cdot \frac{N-m}{N-1} \cdot \left[\sum_{i=1}^N Y_i(0) \cdot Y_i(1) - \frac{1}{N} \cdot \sum_{i=1}^N \sum_{l=1}^N Y_i(0) \cdot Y_l(1) \right] \\
&= \frac{p(1-p)N}{N-1} \cdot \left[\sum_{i=1}^N Y_i(0) \cdot Y_i(1) - \frac{1}{N} \cdot \sum_{i=1}^N \sum_{l=1}^N Y_i(0) \cdot Y_l(1) \right]
\end{aligned}$$

Re-arranging the above expression yields:

$$\text{Cov} \left(\sum_{i=1}^N Y_i(1) \cdot T_i, \sum_{i=1}^N Y_i(0) \cdot T_i \right) = \frac{p(1-p)N^2}{N-1} \cdot \rho \cdot \sigma_1 \sigma_0$$

Calculation of Cov (SATT, SATC)

Note that the covariance between SATT and SATC is,

$$\begin{aligned}
\text{Cov (SATT, SATC)} &= \text{Cov} \left(\frac{1}{m} \cdot \sum_{i=1}^N (Y_i(1) - Y_i(0)) \cdot T_i, \frac{1}{N-m} \cdot \sum_{i=1}^N (Y_i(1) - Y_i(0)) \cdot (1 - T_i) \right) \\
&= -\frac{1}{m} \cdot \frac{1}{N-m} \cdot \text{Cov} \left(\sum_{i=1}^N (Y_i(1) - Y_i(0)) \cdot T_i, \sum_{i=1}^N (Y_i(1) - Y_i(0)) T_i \right) \\
&= -\frac{1}{N \cdot p(1-p)} \cdot \frac{\text{Var} \left(\sum_{i=1}^N (Y_i(1) - Y_i(0)) T_i \right)}{N} \\
&= -\frac{1}{N \cdot p(1-p)} \cdot \frac{m^2}{N} \cdot \text{Var (SATT)} \\
&= -\frac{1}{N \cdot p(1-p)} \cdot \frac{m^2}{N} \cdot \frac{N-m}{m(N-1)} \cdot \sigma_\tau^2 \\
&= -\frac{1}{N-1} \cdot \sigma_\tau^2
\end{aligned}$$

Another way of deriving the covariance between the SATT and SATC is to use the fact that

$\text{Var}(\text{SATE}) = 0$: First note that,

$$\begin{aligned}
\text{Var}(\text{SATE}) &= \text{Var}\left(\frac{m}{N} \cdot \text{SATT}\right) + \text{Var}\left(\frac{N-m}{N} \cdot \text{SATC}\right) + 2 \cdot p(1-p) \cdot \text{Cov}(\text{SATT}, \text{SATC}) \\
&= p^2 \cdot \frac{1}{m^2} \cdot \text{Var}\left(\sum_{i=1}^N (Y_i(1) - Y_i(0))T_i\right) + (1-p)^2 \cdot \frac{1}{(N-m)^2} \cdot \text{Var}\left(\sum_{i=1}^N (Y_i(1) - Y_i(0))T_i\right) \\
&\quad + 2 \cdot p(1-p) \cdot \text{Cov}(\text{SATT}, \text{SATC}) \\
&= 2 \cdot \frac{1}{N^2} \cdot \text{Var}\left(\sum_{i=1}^N (Y_i(1) - Y_i(0))T_i\right) + 2 \cdot p(1-p) \cdot \text{Cov}(\text{SATT}, \text{SATC})
\end{aligned}$$

Hence,

$$\begin{aligned}
\text{Var}(\text{SATE}) &= 0 \\
\Rightarrow \text{Cov}(\text{SATT}, \text{SATC}) &= -\frac{1}{N \cdot p(1-p)} \cdot \frac{\text{Var}\left(\sum_{i=1}^N (Y_i(1) - Y_i(0))T_i\right)}{N}
\end{aligned}$$

J Proof of Theorem (7.1)

As both m and $\sum_{i=1}^N Y_i(0) \cdot T_i$ are random variables, the distribution of the ratio can be derived using the Delta method. The result that the limiting distribution is Normal follows directly from standard results of the Delta method (?). The derivations of the variance-covariance matrix of the limiting distribution are detailed below. First notice that:

$$\begin{aligned}
\frac{1}{N} \cdot \sum_{i=1}^N Y_i(0) \cdot T_i &\xrightarrow{p} \mathbb{E}[Y(0)] \\
\frac{1}{N} \cdot \sum_{i=1}^N Y_i(0) \cdot T_i &\xrightarrow{d} N\left(\mathbb{E}[Y(0)], \frac{1}{N} \cdot p(1-p) [\sigma_0^2 + \mathbb{E}[Y(0)]^2]\right)
\end{aligned}$$

where

$$\begin{aligned}
\text{Var}\left(\frac{1}{N} \cdot \sum_{i=1}^N Y_i(0) \cdot T_i\right) &= \frac{1}{N^2} \cdot p(1-p) \sum_{i=1}^N Y_i(0)^2 \\
&= \frac{1}{N} \cdot p(1-p) [\sigma_0^2 + \mathbb{E}[Y(0)]^2]
\end{aligned}$$

and that m follows a Binomial distribution of $m \sim \text{Binom}(n, p)$ and the variance of m is $\text{Var}(m) = N \cdot p(1 - p)$. The covariance between m and $\sum_{i=1}^N Y_i(0) \cdot T_i$ is

$$\begin{aligned} \text{Cov} \left(m, \sum_{i=1}^N Y_i(0) \cdot T_i \right) &= \text{Cov} \left(\sum_{i=1}^N T_i, \sum_{i=1}^N Y_i(0) \cdot T_i \right) = \sum_{i=1}^N Y_i(0) \cdot \text{Var}(T_i) \\ &= p(1 - p) \cdot \sum_{i=1}^N Y_i(0) \\ &= p(1 - p) \cdot N \cdot \mathbb{E}[Y(0)] \end{aligned}$$

and therefore the covariance-variance matrix is,

$$\Sigma = \begin{pmatrix} N \cdot p(1 - p) [\sigma_0^2 + \mathbb{E}[Y(0)]^2] & p(1 - p) \cdot N \cdot \mathbb{E}[Y(0)] \\ p(1 - p) \cdot N \cdot \mathbb{E}[Y(0)] & N \cdot p(1 - p) \end{pmatrix} \quad (40)$$

The variance of $\frac{1}{m} \cdot \sum_{i=1}^N Y_i(0) \cdot T_i$ is,

$$\begin{pmatrix} \frac{1}{m} \\ -\frac{\sum_{i=1}^N Y_i(0) \cdot T_i}{m^2} \end{pmatrix}^T \cdot \Sigma \cdot \begin{pmatrix} \frac{1}{m} \\ -\frac{\sum_{i=1}^N Y_i(0) \cdot T_i}{m^2} \end{pmatrix}$$

where a consistent estimator for p is $\frac{1}{N} \cdot \sum_{i=1}^N T_i$ and a consistent estimator for $\sum_{i=1}^N Y_i(0) \cdot T_i$ is $\sum_{i=1}^N Y_i(0) \cdot (1 - T_i)$.