A Binary IV Model for Persuasion: Profiling Persuasion Types among Compliers*

Arthur Zeyang Yu
Department of Political Science
The University of Chicago
arthurzeyangyu@uchicago.edu

May 3, 2023

Abstract

In the empirical study of persuasion, researchers often use a binary instrument to encourage individuals to consume information and take some action. We show that with the Imbens-Angrist instrumental variable model assumptions and the monotone treatment response assumption, it is possible to identify the joint distributions of potential outcomes among compliers. This is necessary to identify the percentage of persuaded individuals and their statistical characteristics. Specifically, we develop a weighting method that helps researchers identify the statistical characteristics of persuasion types: compliers and always-persuaded, compliers and persuaded, and compliers and never-persuaded. These findings extend the "κ weighting" results in Abadie (2003). We also provide a sharp test on the two sets of identification assumptions. The test boils down to testing whether there exists a nonnegative solution to a possibly under-determined system of linear equations with known coefficients. An application based on Green et al. (2003) is provided. The result shows that among compliers, roughly 10% voters are persuaded. The results are consistent with the findings that voters' voting behaviors are highly persistent.

Keywords: Instrumental variable, local persuasion rate, Abadie's κ, specification test, GOTV

*I thank my advisors, Scott Gehlbach, Robert Gulotty, Molly Offer-Westort, and Alexander Torgovitsky, who were gracious with their advice, support, and feedback. I additionally thank Eric Auerbach, Stephane Bonhomme, Federico Bugni, Joshua Byun, Matias Cattaneo, Gustavo Diaz, Wayne Yuan Gao, Justin Grimmer, Peter Hull, Kosuke Imai, Sung Jae Jun, Nadav Kunievsky, Xinran Li, Jonathan Mummolo, Maggie Penn, Kirill Ponomarev, James Robinson, Jonathan Roth, Francesco Ruggieri, Azeem Shaikh, Tymon Sloczynski, Joshua Ka Chun Shea, Liyang Sun, Max Tabord-Meehan, Christopher Walters, Linbo Wang, Yiqing Xu, and Teppei Yamamoto for helpful comments on this paper.

1 Introduction

In the empirical study of persuasion, researchers are interested in the treatment effect of information on political choices. Since the decision to consume information is endogenous, researchers often rely on instrumental variables (IVs) that capture exogenous variation in that decision making process. Previous research on instrumental variables has focused on the marginal distribution of potential outcomes: the share of people that take an action under treatment and the share of people that do so under control (Imbens and Rubin, 1997). However, persuasion involves moving a single person from one kind of action to another. This paper shows that under certain assumptions, a binary instrumental variable (IV) model can identify the proportion of individuals who are persuaded, those that are "always persuaded", and those that are "never persuaded", and describe their profiles in terms of pre-treatment covariates.

In a binary IV model of persuasion, the outcome, treatment, and instrument are all dichotomous. Therefore, we can classify individuals into four persuasion types: 1) *always-persuaded*, or those who will take the action of interest regardless of whether receive the information treatment or not; 2) *never-persuaded*, or those who will not take the action of interest regardless of the treatment; 3) *persuaded*, or those who will take the action of interest only if they are exposed to the information treatment; and 4) *dissuaded*, or those who will take the action of interest if they are not exposed to the information treatment but not take the action of interest if they are exposed to the information treatment. Similarly, we can classify individuals into four compliance types: always-takers, never-takers, compliers, and defiers.

We first show that in a binary IA IV model with the monotone treatment response assumption (Imbens and Angrist, 1994; Manski, 1997), the joint distribution of potential outcomes among compliers is point identified. Note that these two assumptions rule out the dissuaded and the defiers. Therefore, treated individuals are at least as likely to take action as an individual who is untreated. This implies that the percentage of persuaded individuals among compliers is equal to the local average treatment effect (LATE). Furthermore, under monotone treatment response, the event in which an individual is always-persuaded is equivalent to the event that an individual would take action without treatment. The latter event only involves the marginal distribution of potential outcomes, which is point identified. (Imbens and Rubin, 1997; Abadie, 2002, 2003). By applying a similar argument, we can identify the proportion of never-persuaded among compliers.

Given the ability to identify persuasion types, we can also profile them by using pre-treatment covariates. We begin by extending the κ weighting result in Abadie (2003) to the local persuasion rate developed by Jun and Lee (2018). Specifically, we show that with the IA IV assumption, we can identify the statistical characteristics measured by pre-treatment covariates of the locally persuadable, by which we mean those who are compliers and who will not take the action of interest without being exposed to the treatment.

We then extend this analysis to show that, under the monotone treatment response assumption, we can characterize the statistical characteristics across persuasion types: always-persuaded compliers, never-persuaded compliers, and persuaded compliers, by reweighting the data to "find" them. This result extends the classic κ weighting result in Abadie (2003) because we now can learn the statistical characteristics of different persuasion types among compliers.

The new identification results follow from the monotone treatment response assumption, which may not be applicable in situations where researchers are uncertain about the direction of the treatment effect.

To guide researchers in the applicability of these results, we provide a sharp test on the two sets of identification assumptions and a sensitivity analysis. The sharp test closely relates to the result in Balke and Pearl (1997). The test exploits the fact that a binary IA IV model with monotone treatment response assumption implies an under-determined system of linear equations with known coefficients. Thus, testing the validity of the identification assumptions boils down to testing whether there exists a nonnegative solution to the implied system of linear equations. We implement the test using the subsampling method (Bai et al., 2022). We also provide a sensitivity result based on the idea in Balke and Pearl (1997). Specifically, since in the binary IV model, the observed quantity is a linear system equation of the unobserved outcome and compliance types, we can vary the size of the violation of the monotone treatment response assumption among compliers to see how our point identification results change.

We also provide estimation and inference results. Our identification results show that most of the estimands share a similar flavor with the Wald estimands. Therefore, the estimation and inference results can be obtained by directly applying the classic results in the IV literature. Moreover, they can be easily implemented in standard statistical software, say, `Stata`.

Finally, we illustrate the usage of our methods by providing an application based on Green et al. (2003). Green et al. (2003) conduct a field experiment to use the Getting Out the Vote (GOTV) program to persuade voters to vote. Specifically, the instrument is the randomly assigned GOTV program. The treatment is the actual take-up of the GOTV program. The outcome is whether or not voters turn out to vote. The results show that among compliers, around 10% individuals are persuaded. Moreover, we find that among compliers, the chance for always-persuaded voters to vote in the last presidential election is the highest, and the chance for never-persuaded voters to vote in the last presidential election is the lowest. These results are consistent with the interpretation that voters' voting behaviors are habit-forming, hence are highly persistent (Gerber et al., 2003). Moreover, our results show that the voting propensity of those persuaded is close to those always-persuaded, which is consistent with the finding in Enos et al. (2014) that GOTV program mobilizes high-propensity voters. Moreover, in Bridgeport, the results show that the chance of being a Democrat among the persuaded voters and compliers is high, though the estimate is quite noisy.

This paper is closely related to Abadie (2003), who provides results on identifying the statistical characteristics measured by pre-treatment covariates for compliers. We extend Abadie's κ result by identifying statistical characteristics measured by the pre-treatment covariates of the persuasion types (i.e., always-persuaded, never-persuaded, and persuaded) among compliers under a binary IA IV model with an additional monotone treatment response assumption.

Moreover, this paper also relates to the literature on identifying the distribution of potential outcomes in an IV model. Prior work proposes three approaches: 1) focuses on identifying the marginal distribution of potential outcomes among compliers (Imbens and Rubin, 1997; Abadie, 2002; Abadie et al., 2002; Abadie, 2003); 2) makes a rank invariance assumption to point identify quantile treatment effect (Chernozhukov and Hansen, 2004, 2005; Vuong and Xu, 2017; Feng et al., 2019); 3) constructs sharp bounds on the joint distribution of potential outcomes (Torgovitsky, 2019; Russell, 2021). In this paper, the identification of the joint distribution of potential outcomes among compliers depends on the binary nature of the outcome and the assumption of the direction of the treatment effect.

This paper also closely relates to Jun and Lee (2018). Jun and Lee (2018) provides a set of point/partial identification results for the persuasion rate and the local persuasion rate under different data scenarios.

One main focus of this paper is to profile the persuasion types among compliers. Moreover, this paper provides a sharp test on the assumptions in a binary IV model for persuasion. The sharp test itself also speaks to a large literature on testing IA IV model validity (Balke and Pearl, 1997; Heckman and Vytlacil, 2005; Kitagawa, 2015; Huber and Mellace, 2015; Wang et al., 2017; Mourifié and Wan, 2017; Machado et al., 2019; Kédagni and Mourifié, 2020). The sharp test follows the tradition of the literature by using the simple fact that the observed quantity in the data is a linear combination of the probability of the unobserved outcome and compliance types. Furthermore, we also provide a necessary and sufficient condition under which the "approximated" persuasion rate proposed by DellaVigna and Kaplan (2007) equals the local persuasion rate proposed by Jun and Lee (2018) when there is one-sided non-compliance in the experiment design. Finally, we also provide a simple sensitivity analysis approach to assess the robustness of the results for the violation of the monotone treatment response assumption.

The remainder of the paper proceeds as follows. In Section 2, we set up a binary IV model of persuasion. In Section 3, we define the target parameters. Section 4 presents the point identification results of the distribution of potential outcomes among compliers. Section 5 presents the identification results that identify the statistical characteristics of persuasion types among compliers. Section 6 presents the estimation and inference results. Additional discussions can be found in Section 7. We provide an application in Section 8 and conclude in the final section.

2 Model Setup

In empirical study of persuasion, researchers often collect data on a binary information treatment T_i, and a binary behavioral outcome Y_i. In the GOTV experiment, the outcome of interest is whether or not voters vote, and the information treatment is the information on the timing and the location of the upcoming election. Since information consumption is endogenous, researchers often employ an instrument Z_i which creates exogenous variations for an individual's information consumption decision. In many experiments, the instrument Z_i is also binary. In the GOTV experiment, the instrument is the randomly assigned access to the GOTV treatment, which contains information on the timing and location of the upcoming election. Besides the aforementioned variables, researchers also collect pre-treatment covariates $X_i \in \mathbb{R}^k$.[1] Define $Y_i(1)$ and $Y_i(0)$ as the potential outcomes that an individual would attain with and without being exposed to the treatment, and $T_i(1)$ and $T_i(0)$ as the potential treatments that an individual would attain with and without being exposed to the instrument. For a particular individual, the variable $Y_i(t, z)$ represents the potential outcome that this individual would obtain if $T_i = t$ and $Z_i = z$.

Formally speaking, researchers make the following assumptions in a binary IV model of persuasion with the potential outcome and potential treatment notations.

Assumption 2.1. (A Binary IV Model of Persuasion)

1. Exclusion restriction: $Y_i(t, z) = Y_i(t)$, for $t, z \in \{0, 1\}$,
2. Exogenous instrument: $Z_i \perp\!\!\!\perp (Y_i(0), Y_i(1), T_i(0), T_i(1), X_i)$,
3. First stage: $\mathbb{P}[T_i = 1|Z_i = 1] \neq \mathbb{P}[T_i = 1|Z_i = 0]$,

[1] In what follows, we assume without loss of generality that $k = 1$.

4. IV Monotonicity: $T_i(1) \geq T_i(0)$ holds almost surely,

5. Monotone treatment response: $Y_i(1) \geq Y_i(0)$ holds almost surely, and $Y_i(0), Y_i(1) \in \{0, 1\}$.

Assumptions 1 to 4 are the assumptions in the IA IV model. In what follows, we use the IA IV assumptions and the LATE assumptions interchangeably to refer to Assumptions 1 to 4. Note that it is not new to assume the direction of the treatment effect in econometrics literature (Manski, 1997; Manski and Pepper, 2000; Okumura and Usui, 2014; Kim et al., 2018). This type of assumption is attractive when researchers have strong prior for the direction of the treatment effect. Similar to the IV monotonicity in the IA IV assumption, this assumption rules out the type of individuals who will take the action of interest if the treatment switches off but will not take the action of interest if the treatment switches on. In other words, this assumption assumes that there are no dissuaded people.

As pointed out by Machado et al. (2019), the results in Vytlacil (2002) imply that Assumption 2.1 is equivalent to the following triangular system model:

1. $Y_i(t) = \mathbb{1}\{U_i \leq \gamma(t)\}$, where $\gamma : \mathcal{T} \to \mathbb{R}$ is a measurable function with $\gamma(0) < \gamma(1)$,

2. $T_i(z) = \mathbb{1}\{V_i \leq \nu(z)\}$, where $\nu : \mathcal{Z} \to \mathbb{R}$ is a measurable function with $\nu(0) < \nu(1)$,

3. $Z_i \perp\!\!\!\perp (V_i, U_i, X_i)$,

where U_i is the latent utility in the outcome process, and V_i is the latent utility in the selection process.

Assumption 2.1 can be applied in cases other than persuasion.[2] For instance, researchers are interested in studying the effect of participating in a job training program on the decision to join a rebellion group in a fragile state (Blattman and Annan, 2016; Blattman et al., 2017, 2020). Blattman and Annan (2016) conducted an experiment in Liberia that randomly assigned Liberian ex-fighters to a free agricultural training program. The treatment is the actual participation in the agricultural training program. The outcome of interest is whether or not the Liberian ex-fighters are employed in the legal sector. Here, the IV monotonicity condition is likely to hold because the program should decrease the cost of the training program for all of the ex-fighters. The monotone treatment response assumption is likely to hold as the training program is expected to increase the human capital of ex-fighters, thereby increasing their wage return from getting a job in the legal sector and raising their opportunity cost of getting a job in the illegal sector.

By Assumption 2.1, we can classify individuals into 9 groups. Since the outcome is binary, the monotone treatment response assumption implies that we can classify individuals as always-persuaded, never-persuaded, and persuaded. By the IV monotonicity assumption, we can classify the individuals as always-takers, never-takers, and compliers. The classification is presented in Table 1.

3 Target Parameters

In the empirical study of persuasion, researchers are interested in the "effect" of the information treatment on individuals' behaviors. One target parameter proposed by Jun and Lee (2018) is the local persuasion

[2]Besides the applications mentioned in the main text, the binary IA IV model with monotone treatment response can further be applied to the study of the persuasion effect of political messages on political behavior in democracy and autocracy (DellaVigna and Kaplan, 2007; Enikolopov et al., 2011), the persuasion effect of uncensored internet on the views of censorship (Chen and Yang, 2019), persuading donors to donate (Landry et al., 2006), etc.

Table 1: Types of Individuals

$Y_i(0)$	$Y_i(1)$	$T_i(0)$	$T_i(1)$	Persuasion Types	Compliance Types
0	0	0	0	Never-Persuaded	Never-Takers
0	1	0	0	Persuaded	Never-Takers
1	1	0	0	Always-Persuaded	Never-Takers
0	0	0	1	Never-Persuaded	Compliers
0	1	0	1	Persuaded	Compliers
1	1	0	1	Always-Persuaded	Compliers
0	0	1	1	Never-Persuaded	Always-Takers
0	1	1	1	Persuaded	Always-Takers
1	1	1	1	Always-Persuaded	Always-Takers

rate:

$$\theta_{\text{local}} := \mathbb{P}[Y_i(1) = 1 | Y_i(0) = 0, T_i(1) > T_i(0)].$$

The local persuasion rate measures the percentage of compliers who take the action of interest if exposed to the treatment among those who will not take the action of interest without being exposed to the information treatment.[3] In the GOTV experiment, the local persuasion rate measures the percentage of voters who would vote if they had been exposed to the GOTV program among compliers and those who would not vote were they not exposed to the GOTV program. Given Assumption 2.1, Jun and Lee (2018) have shown that θ_{local} is point identifiable.

Compared to the LATE, the local persuasion rate focuses on a smaller subpopulation. LATE is the average treatment effect for compliers. The local persuasion rate further conditions on those who will not take the action of interest without the information treatment (i.e., $[Y_i(0) = 0]$). In the GOTV experiment, the local persuasion rate conditions on those who will not vote without being exposed to the GOTV program and those who comply with the experiment design.

We propose three sets of new target parameters in this paper. First, we are interested in the joint distribution of potential outcomes among compliers. Persuasion involves moving an individual from one kind of action to another. Therefore, to gain a deeper understanding of the effectiveness of information intervention, researchers need information about the joint distribution of potential outcomes.

Second, we are interested in the statistical characteristics measured by pre-treatment covariates for the locally persuadable. Here, the locally persuadable is the subpopulation that θ_{local} conditions on: $[Y_i(0) = 0, T_i(1) > T_i(0)]$. Learning the statistical characteristics of the locally persuadable can help researchers assess the strength of the study's external validity. If the statistical characteristics of the locally persuadable are not similar to the general population, researchers need to be cautious about generalizing their conclusion to the general population.

The third set of target parameters refers to the statistical characteristics of the persuasion types among compliers (i.e., always-persuaded, never-persuaded, and persuaded). Understanding these characteristics can help researchers assess the experiment's success in achieving specific goals and its potential policy out-

[3] As summarized in DellaVigna and Gentzkow (2010), another popular target parameter in the empirics of persuasion is the persuasion rate: $\theta := \mathbb{P}[Y_i(1) = 1 | Y_i(0) = 0]$. DellaVigna and Gentzkow (2010) suggests to use an estimand proposed in DellaVigna and Kaplan (2007) to measure θ: $\theta_{\text{DK}} = \frac{\mathbb{P}[Y_i=1|Z_i=1]-\mathbb{P}[Y_i=1|Z_i=0]}{\mathbb{P}[T_i=1|Z_i=1]-\mathbb{P}[T_i=1|Z_i=0]} \times \frac{1}{1-\mathbb{P}[Y_i(0)=1]}$, where researchers use $\mathbb{P}[Y_i = 1 | Z_i = 0]$ to approximate $\mathbb{P}[Y_i(0) = 1]$. As pointed out in Jun and Lee (2018), θ_{DK} is not a well defined conditional probability. Hence, it does not measure the persuasion rate for any subpopulation. Moreover, Jun and Lee (2018) show that under Assumption 2.1, θ is not point identifiable. They instead provide sharp bounds for θ.

comes. For instance, in the GOTV experiment, researchers aimed to mobilize underrepresented minorities to vote, so estimating the likelihood of persuaded and compliers being part of this group is crucial. Additionally, researchers may want to determine the types of voters mobilized, such as their likelihood of being Democrats. This information can help researchers evaluate the policy impact of the mobilization effort.

4 Identification of the Potential Outcome Distributions for Compliers

In this section, we present the results of the identification of the joint distribution of potential outcomes among compliers. We first show that in a binary IA IV model with monotone treatment response assumption, the joint distribution of potential outcomes among compliers can be identified from the marginal distribution of potential outcomes among compliers. We then show that the results can be extended to the case of a non-binary instrument.

4.1 Identification of the Joint Distribution of Potential Outcomes for Compliers in a Binary IV Model

As is well known, given the IA IV assumptions, we can point identify the marginal distribution of potential outcomes among compliers (Imbens and Rubin, 1997; Abadie, 2003; Jun and Lee, 2018). In other words, we can know the percentage of voters who will vote if they receive the GOTV treatment and the percentage of voters who will vote if they do not receive the GOTV treatment among compliers. For the sake of completeness, we restate this classic result in Lemma 4.1.

Lemma 4.1. Assume that the 1 to 4 in Assumption 2.1 hold, then, with binary Y_i, the marginal distribution of potential outcomes conditional on compliers is point identified:

$$\mathbb{P}[Y_i(0) = y \mid T_i(1) > T_i(0)] = \frac{\mathbb{P}[Y_i = y, T_i = 0 \mid Z_i = 0] - \mathbb{P}[Y_i = y, T_i = 0 \mid Z_i = 1]}{\mathbb{E}[T_i \mid Z_i = 1] - \mathbb{E}[T_i \mid Z_i = 0]}$$
$$\mathbb{P}[Y_i(1) = y \mid T_i(1) > T_i(0)] = \frac{\mathbb{P}[Y_i = y, T_i = 1 \mid Z_i = 1] - \mathbb{P}[Y_i = y, T_i = 1 \mid Z_i = 0]}{\mathbb{E}[T_i \mid Z_i = 1] - \mathbb{E}[T_i \mid Z_i = 0]},$$

where $y \in \{0, 1\}$.

The intuition of the identification results in Lemma 4.1 is the following. To make the discussion more concrete, let us consider the untreated potential outcome in the GOTV experiment. Among the voters who are not randomly assigned to the GOTV treatment (i.e., those with $Z_i = 0$), for those who do not receive the GOTV experiment (i.e., those with $T_i = 0$), we know that: 1) we observe their untreated potential outcome, $Y_i(0)$; 2) by the IV monotonicity in Assumption 2.1, they are either compliers or never-takers. Among the voters who are randomly assigned to the GOTV treatment (i.e., those with $Z_i = 1$), for those who do not receive the GOTV experiment (i.e., those with $T_i = 0$), we know that: 1) we observe their untreated potential outcome; 2) by the IV monotonicity assumption, they are never-takers. Subtracting the two groups then gives us compliers. Similarly, for the treated potential outcome, subtracting a mixture of always-takers and compliers from always-takers gives us compliers.

The two estimands in Lemma 4.1 are similar to the Wald estimand in the IA IV model. Consider the marginal distribution of $Y_i(1)$ among compliers, the estimand is equivalent to a Wald estimand with treatment variable being T_i, instrument being Z_i, and the outcome variable being $\mathbb{1}\{Y_i = y, T_i = 1\}$ with $y \in \{0, 1\}$. For the marginal distribution of $Y_i(0)$ among compliers, it is the negative of the Wald estimand with the outcome variable being the following indicator variable: $\mathbb{1}\{Y_i = y, T_i = 0\}$ with $y \in \{0, 1\}$.

The identification results in Lemma 4.1 only use the IA IV assumptions. Remarkably, if we further assume the treatment response is monotone, we can point identify the joint distribution of potential outcomes among compliers. In other words, under Assumption 2.1, we can know the percentage of always-persuaded, never-persuaded, and persuaded among compliers.

Lemma 4.2. Suppose Assumption 2.1 holds, the joint distribution of potential outcomes among compliers is point identified:

$$\begin{aligned}
\mathbb{P}[Y_i(1) = 1, Y_i(0) = 1 \mid T_i(1) > T_i(0)] &= \frac{\mathbb{P}[Y_i = 1, T_i = 0 \mid Z_i = 0] - \mathbb{P}[Y_i = 1, T_i = 0 \mid Z_i = 1]}{\mathbb{E}[T_i \mid Z_i = 1] - \mathbb{E}[T_i \mid Z_i = 0]} \\
\mathbb{P}[Y_i(1) = 1, Y_i(0) = 0 \mid T_i(1) > T_i(0)] &= \frac{\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0]}{\mathbb{E}[T_i \mid Z_i = 1] - \mathbb{E}[T_i \mid Z_i = 0]} \\
\mathbb{P}[Y_i(1) = 0, Y_i(0) = 0 \mid T_i(1) > T_i(0)] &= \frac{\mathbb{P}[Y_i = 0, T_i = 1 \mid Z_i = 1] - \mathbb{P}[Y_i = 0, T_i = 1 \mid Z_i = 0]}{\mathbb{E}[T_i \mid Z_i = 1] - \mathbb{E}[T_i \mid Z_i = 0]}.
\end{aligned}$$

Here is the intuition behind the identification results in Lemma 4.2. By the monotone treatment response in Assumption 2.1, we know the following three things: 1) for those who will vote without receiving the GOTV treatment (i.e., those with $Y_i(0)$ being 1), they will also vote with receiving the GOTV treatment (i.e., their $Y_i(1)$ is also 1); 2) for those who will not vote with receiving the GOTV treatment (i.e., those with $Y_i(1)$ being 0), they will also not vote without receiving the GOTV treatment (i.e., their $Y_i(0)$ is also 0); 3) $Y_i(1) - Y_i(0) = 1$ if and only if $Y_i(1) = 1, Y_i(0) = 0$, thus, LATE becomes the proportion of mobilizable voters among compliers.[4]

Note that we only need the monotone treatment response assumption to hold among compliers for Lemma 4.2, because we are "solving" the joint distribution of potential outcomes among compliers from the marginal distribution. However, throughout the text, we maintain the assumption that the monotone treatment response holds almost surely for simplicity.

5 Profiling Persuasion Types

This section offers results that profile the persuasion types among compliers, in addition to determining the size of the persuasion effect. We present a series of results that help identify the statistical characteristics of the locally persuadable (that is, $[Y_i(0) = 0, T_i(1) > T_i(0)]$) as well as the three other persuasion types defined by the marginal potential outcomes. Next, we provide results that identify the statistical characteristics of the three persuasion types among compliers as defined in Table 1.[5]

[4]We discuss the extension of the identification results in Lemma 4.2 to non-binary outcomes and instruments in Appendix A. The results are negative for the former and positive for the latter.

[5]We also extend some of our findings to always-takers and never-takers, see Appendix D.

5.1 Profiling the Locally Persuadable

Given the IA IV assumption, we can identify the statistical characteristics of the subpopulation defined by the following event: $[Y_i(0) = 0, T_i(1) > T_i(0)]$, i.e., the locally persuadable. We do not directly observe this subpopulation because it involves potential outcomes and a pair of potential treatments. In the GOTV experiment, the locally persuadable are those who are compliers and those who will not vote if they do not receive the GOTV treatment. We formally state the results below.[6]

Theorem 5.1. Suppose that 1 to 4 in Assumption 2.1 hold. Let $g : \mathbb{R} \to \mathbb{R}$ be measurable such that $\mathbb{E}[|g(X_i)|] < \infty$, then, $\mathbb{E}[g(X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)]$ is point identified:

$$\begin{aligned}
&\mathbb{E}[g(X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)] \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0, T_i = 0\} \mid Z_i = 0] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0, T_i = 0\} \mid Z_i = 1]}{\mathbb{P}[Y_i = 0, T_i = 0 \mid Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0 \mid Z_i = 1]}.
\end{aligned}$$

We provide examples of $g(X_i)$ below. For instance, if we choose $g(X_i) = X_i^p$ where $p \in \mathbb{R}^+$, we can identify any moments of a covariate X_i that exist. In the GOTV experiment, X_i can be a binary partisanship variable, indicating whether or not i is a Democrat. Choosing $p = 1$, we can identify the probability of a locally persuadable voter being a Democrat. Another example is $g(X_i) = \mathbb{1}\{X_i \leq x\}$ with $x \in \mathbb{R}$. With this choice, we can identify the cumulative distribution function of X_i among the locally persuadable. For instance, if X_i is personal income, we can identify the cumulative distribution function of income among the locally persuadable voters.

Theorem 3.1 in Abadie (2003) shows that any statistical characteristic that can be defined in terms of moments of the joint distribution of (Y_i, T_i, X_i) is identified for compliers:

$$\mathbb{E}[g(Y_i, T_i, X_i) \mid T_i(1) > T_i(0)] = \frac{1}{\mathbb{P}[T_i(1) > T_i(0)]}\mathbb{E}[\kappa g(Y_i, T_i, X_i)],$$

where $\kappa := 1 - \frac{T_i(1-Z_i)}{\mathbb{P}[Z_i=0]} - \frac{(1-T_i)Z_i}{\mathbb{P}[Z_i=1]}$. Theorem 5.1 strengthens Abadie's κ by further conditioning on those with an untreated potential outcome of 0. Thus, a natural question is whether or not we can point identify $\mathbb{E}[g(Y_i, T_i, X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)]$ under the IA IV assumption. The answer is no. To see this:

$$\begin{aligned}
&\mathbb{E}[g(Y_i, T_i, X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)] \\
&= \mathbb{E}[g(Y_i(1)Z_i + Y_i(0)(1 - Z_i), Z_i, X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)] \\
&= \mathbb{E}[g(Y_i(1)Z_i, Z_i, X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)] \\
&= \mathbb{E}[g(Y_i(1), 1, X_i) \mid Z_i = 1, Y_i(0) = 0, T_i(1) > T_i(0)]\mathbb{P}[Z_i = 1 \mid Y_i(0) = 0, T_i(1) > T_i(0)] \\
&\quad + \mathbb{E}[g(0, 0, X_i) \mid Z_i = 1, Y_i(0) = 0, T_i(1) > T_i(0)]\mathbb{P}[Z_i = 0 \mid Y_i(0) = 0, T_i(1) > T_i(0)] \\
&= \mathbb{E}[g(Y_i(1), 1, X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)]\mathbb{P}[Z_i = 1] \\
&\quad + \mathbb{E}[g(0, 0, X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)]\mathbb{P}[Z_i = 0],
\end{aligned}$$

where the first equality uses the fact that $T_i = Z_i$ for compliers, the fourth equality uses the IV independence assumption. Due to the presence of $\mathbb{E}[g(Y_i(1), 1, X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)]\mathbb{P}[Z_i = 1]$, which is about

[6] In Appendix H, we show that we can use the weighting results in Abadie (2003) to derive the same result in Theorem 5.1.

the joint distribution of potential outcomes, $\mathbb{E}[g(Y_i, T_i, X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)]$ is not point identified with the IA IV assumptions.

Theorem 5.1 can be applied to continuous Y_i by defining a new indicator variable, $\tilde{Y}_i = \mathbb{1}\{Y_i \in B\}$, where B is a measurable set, and a new potential outcome, $\tilde{Y}_i(0) = \mathbb{1}\{Y_i(0) \in B\}$. The result in Theorem 5.1 holds for $\tilde{Y}_i$ under the IA IV assumptions in Assumption 2.1. An example of B is: $B = \mathbb{1}\{Y_i(0) \leq \tilde{y}\}$. That is, researchers can identify characteristics measured by X_i of compliers and those with untreated outcomes less than $\tilde{y}$.

Since the marginal distribution of potential outcomes among compliers is identifiable, a natural extension of Theorem 5.1 is to extend the results to the following subpopulations: $[Y_i(0) = 1, T_i(1) > T_i(0)]$, $[Y_i(1) = 0, T_i(1) > T_i(0)]$, and $[Y_i(1) = 1, T_i(1) > T_i(0)]$.

Proposition 5.1. Assume that 1 to 4 in Assumption 2.1 hold, and let $g : \mathbb{R} \to \mathbb{R}$ be measurable such that $\mathbb{E}[|g(X_i)|] < \infty$, then, the following conditional expectations of $g(X_i)$ are point identified:

$$
\begin{aligned}
&\mathbb{E}[g(X_i) \mid Y_i(0) = 1, T_i(1) > T_i(0)] \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1, T_i = 0\} \mid Z_i = 0] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1, T_i = 0\} \mid Z_i = 1]}{\mathbb{P}[Y_i = 1, T_i = 0 \mid Z_i = 0] - \mathbb{P}[Y_i = 1, T_i = 0 \mid Z_i = 1]}, \\
&\mathbb{E}[g(X_i) \mid Y_i(1) = 0, T_i(1) > T_i(0)] \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0, T_i = 1\} \mid Z_i = 1] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0, T_i = 1\} \mid Z_i = 0]}{\mathbb{P}[Y_i = 0, T_i = 1 \mid Z_i = 1] - \mathbb{P}[Y_i = 0, T_i = 1 \mid Z_i = 0]}, \\
&\mathbb{E}[g(X_i) \mid Y_i(1) = 1, T_i(1) > T_i(0)] \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1, T_i = 1\} \mid Z_i = 1] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1, T_i = 1\} \mid Z_i = 0]}{\mathbb{P}[Y_i = 1, T_i = 1 \mid Z_i = 1] - \mathbb{P}[Y_i = 1, T_i = 1 \mid Z_i = 0]}.
\end{aligned}
$$

By the identical reasoning after Theorem 5.1, we have the following three remarks on Proposition 5.1. First, the results show that any conditional moments defined by pre-treatment covariate X_i can be identified as long as the moments are finite. Second, pick $g(X_i) = \mathbb{1}\{X_i \leq x\}$ with $x \in \mathbb{R}$, the results show that the conditional cumulative functions are identified. Third, Proposition 5.1 strengthens Abadie's κ by further conditioning on the potential outcome. However, by the same token in the discussion before, the power of Abadie's κ is not fully preserved here, because we cannot identify $g(Y_i, T_i, X_i)$ conditional on the three subpopulations above.

5.2 Identification: Compliance and Persuasion

An implication of Lemma 4.2 is that we can point identify the statistical properties of always-persuaded, never-persuaded, and persuaded among compliers. The results follow because the joint distribution of potential outcomes among compliers is point identified under the monotone treatment response assumption in the binary IA IV model. The results are summarized in Theorem 5.2.

Theorem 5.2 (Compliance and Persuasion)**.** Suppose Assumption 2.1 holds, let $g : \mathbb{R} \to \mathbb{R}$ be measurable such that $\mathbb{E}[|g(X_i)|] < \infty$, then, the moments of $g(X_i)$ conditional on always-persuaded compliers, never-persuaded compliers, and persuaded compliers are point identified:

$$\mathbb{E}[g(X_i)|Y_i(1) = Y_i(0) = 1, T_i(1) > T_i(0)]$$

$$= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1, T_i = 0\}|Z_i = 0] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1, T_i = 0\}|Z_i = 1]}{\mathbb{P}[Y_i = 1, T_i = 0|Z_i = 0] - \mathbb{P}[Y_i = 1, T_i = 0|Z_i = 1]},$$

$$\mathbb{E}[g(X_i)|Y_i(1) = Y_i(0) = 0, T_i(1) > T_i(0)]$$
$$= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0, T_i = 1\}|Z_i = 1] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0, T_i = 1\}|Z_i = 0]}{\mathbb{P}[Y_i = 0, T_i = 1|Z_i = 1] - \mathbb{P}[Y_i = 0, T_i = 1|Z_i = 0]},$$

$$\mathbb{E}[g(X_i)|Y_i(1) = 1, Y_i(0) = 0, T_i(1) > T_i(0)]$$
$$= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1\}|Z_i = 1] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1\}|Z_i = 0]}{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}.$$

We now give three remarks for Theorem 5.2. By the identical argument in Theorem 5.1, the conditional distribution functions of X_i given persuasion types and compliers are also identifiable, because we can let $g(X_i)$ being $g(X_i) = \mathbb{1}\{X_i \leq x\}$ with $x \in \mathbb{R}$. Furthermore, for measurable g, the expectations of $g(X_i)$ conditional on the three subpopulations are also identifiable given the expectation is well-defined. An implication is any statistical moments of the always-persuaded, never-persuaded, and persuaded among compliers are identifiable. Thus, this theorem extends the weighting results in Abadie (2003) by further conditioning on the persuasion types defined by the pair of potential outcomes.

The aforementioned statistics provide significant aid in comprehending the intervention's impact and mechanism. To illustrate, consider the GOTV experiment. Theorem 5.2 establishes the identification of the probability of a complier-persuaded voter being a Democrat. In other words, although GOTV experiments are not typically partisan ex ante, they can produce partisan mobilization outcomes. For instance, the data may indicate that among compliers, the likelihood of a persuaded voter being a Democrat is exceedingly high. If conducted in a swing state, the mobilization experiment could potentially alter the election results. Furthermore, the results of Theorem 5.2 can facilitate our evaluation of the mechanisms by which the treatment affects the outcome. In the GOTV experiment, the aforementioned results can be employed to evaluate the hypothesis that voting is habit-forming (Gerber et al., 2003). We can utilize prior voting records as a metric for the voting propensity. If the hypothesis in Gerber et al. (2003) is accurate, we should observe that always-persuaded voters among compliers exhibit the highest voting propensity while never-persuaded voters demonstrate the lowest voting propensity.[7]

In addition to Theorem 5.2, there are several other ways to profile voters using observable covariates. For instance, researchers might be interested in the following quantity: among the compliers and those who will not vote without being exposed to the treatment (i.e., the locally persuadable), what are the characteristics of those who will vote with being exposed to the treatment. For example, in the GOTV experiment, this quantity would be the chance of locally persuadable individuals being a Democrat and will if they are exposed to the treatment. Due to the monotone treatment response and binary outcome, there are five other estimands that share a similar flavor with this example. The identifiability of these estimands follows from the fact that the monotone treatment response assumption implies the identifiability of the joint distribution of the potential outcomes among compliers. These results are formally stated in Proposition 5.2.

Proposition 5.2. Suppose Assumption 2.1 holds, let $g : \mathbb{R} \to \mathbb{R}$ be measurable such that $\mathbb{E}[|g(X_i)|] < \infty$, then, the following conditional expectations are identifiable:

$$\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1) = 0\} \mid Y_i(0) = 0, T_i(1) > T_i(0)]$$

[7] In Appendix C, we present results that identify the proportion of persuasion types among compliers while conditioning on covariates.

$$
= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0, T_i = 1\} \mid Z_i = 1] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0, T_i = 1\} \mid Z_i = 0]}{\mathbb{P}[Y_i = 0, T_i = 0 \mid Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0 \mid Z_i = 1]}
$$

$$
\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1) = 1\} \mid Y_i(0) = 0, T_i(1) > T_i(0)]
$$

$$
= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1\} \mid Z_i = 1] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1\} \mid Z_i = 0]}{\mathbb{P}[Y_i = 0, T_i = 0 \mid Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0 \mid Z_i = 1]}
$$

$$
\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1) = 1\} \mid Y_i(0) = 1, T_i(1) > T_i(0)]
$$

$$
= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1, T_i = 0\} \mid Z_i = 0] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1, T_i = 0\} \mid Z_i = 1]}{\mathbb{P}[Y_i = 1, T_i = 0 \mid Z_i = 0] - \mathbb{P}[Y_i = 1, T_i = 0 \mid Z_i = 1]}
$$

$$
\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(0) = 0\} \mid Y_i(1) = 0, T_i(1) > T_i(0)]
$$

$$
= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0, T_i = 1\} \mid Z_i = 1] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0, T_i = 1\} \mid Z_i = 0]}{\mathbb{P}[Y_i = 0, T_i = 1 \mid Z_i = 1] - \mathbb{P}[Y_i = 0, T_i = 1 \mid Z_i = 0]}
$$

$$
\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(0) = 1\} \mid Y_i(1) = 1, T_i(1) > T_i(0)]
$$

$$
= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1, T_i = 0\} \mid Z_i = 0] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1, T_i = 0\} \mid Z_i = 1]}{\mathbb{P}[Y_i = 1, T_i = 1 \mid Z_i = 1] - \mathbb{P}[Y_i = 1, T_i = 1 \mid Z_i = 0]}
$$

$$
\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(0) = 0\} \mid Y_i(1) = 1, T_i(1) > T_i(0)]
$$

$$
= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1\} \mid Z_i = 1] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1\} \mid Z_i = 0]}{\mathbb{P}[Y_i = 1, T_i = 1 \mid Z_i = 1] - \mathbb{P}[Y_i = 1, T_i = 1 \mid Z_i = 0]}
$$

6 Estimation and Inference

This section provides estimation and inference results for the estimands we proposed in Sections 4 and 5. Note that the estimands we proposed in prior sections usually take the form of a Wald estimand:

$$
\frac{\mathbb{E}[f(X_i, Y_i, T_i) \mid Z_i = 1] - \mathbb{E}[f(X_i, Y_i, T_i) \mid Z_i = 0]}{\mathbb{E}[h(Y_i, T_i) \mid Z_i = 1] - \mathbb{E}[h(Y_i, T_i) \mid Z_i = 0]}. \tag{1}
$$

where f and h are measurable functions that map from $\mathbb{R}$ to $\mathbb{R}$.[8] It is easy to see that the numerator in Equation 1 is the coefficient of Z_i from regressing $f(X_i, Y_i, T_i)$ on Z_i and a constant, while the denominator in Equation 1 is the coefficient of Z_i from regressing $h(Y_i, T_i)$ on Z_i and a constant. Therefore, the standard estimation and inference theory for Wald estimand applies immediately to the current case with i.i.d. data of (Y_i, T_i, Z_i, X_i). We can either employ the conventional asymptotic results for hypothesis testing or use the Anderson-Rubin test which is robust to weak identification. [9] Note that both inferential methods can be easily implemented in standard statistical software, say, `ivreg2` and `weakiv` in `Stata`.[10]

[8]To provide an example of f and g for the identifiable estimand we introduced earlier, consider $\mathbb{E}[g(X_i) \mid Y_i(1) = Y_i(0) = 0, T_i(1) > T_i(0)]$ in Theorem 5.2:

$$
\begin{aligned}
f(X_i, Y_i, T_i) &= g(X_i)\mathbb{1}\{Y_i = 0, T_i = 1\} \\
h(Y_i, T_i) &= \mathbb{1}\{Y_i = 0, T_i = 1\}.
\end{aligned}
$$

[9]We provide a more detailed discussion on inference issues in Appendix E.

[10]`ivreg2` does not produce a confidence interval for the Anderson-Rubin test, while `weakiv` does. To see the implementation in `Stata`, consider again $\mathbb{E}[g(X_i) \mid Y_i(1) = Y_i(0) = 0, T_i(1) > T_i(0)]$ in Theorem 5.2. Suppose $g(X_i)\mathbb{1}\{Y_i = 0, T_i = 1\}$ is stored as `gxy0t1` and $\mathbb{1}\{Y_i = 0, T_i = 1\}$ is stored as `y0t1` in `Stata`. Then, the point estimate and confidence interval can be obtained by: `ivreg2 gxy0t1 (y0t1 = z)`. The Anderson-Rubin confidence interval can be obtained by: `weakiv ivreg2 gxy0t1 (y0t1 = z)`.

7 Discussion

In this section, we discuss three points on identification results from previous sections. Firstly, we compare θ_{local} with classic estimands. Next, we provide necessary and sufficient conditions for approximated θ_{DK} to equal θ_{local} under one-sided non-compliance. Additionally, we propose a test for Assumption 2.1, and a simple method to assess the sensitivity of results to the monotone treatment response assumption.

7.1 Comparison with Existing Estimands

7.1.1 Complier Causal Attribution Rate

The most closely related target parameter to the local persuasion rate is the causal attribution rate, which measures the proportion of observed outcome prevented by the hypothetical absence of the treatment (Pearl, 1999). With the presence of a binary instrument, Yamamoto (2012) defines the complier causal attribution rate denoted by p_C:

$$p_C = \mathbb{P}[Y_i(0) = 0 | Y_i(1) = 1, T_i = 1, T_i(1) > T_i(0)],$$

which measures the proportion of observed outcome prevented by the hypothetical absence of treatment among compliers.

One main difference between p_C and θ_{local} is that p_C conditions on $[Y_i(1) = 1, T_i = 1, T_i > T_i(0)]$ but θ_{local} conditions on $[Y_i(0) = 0, T_i > T_i(0)]$. Therefore, a natural way to extend the local persuasion rate is to define the local persuasion rate on the untreated:

$$\theta_{\text{local untreated}} := \mathbb{P}[Y_i(1) = 1 | Y_i(0) = 0, T_i = 0, T_i(1) > T_i(0)].$$

We can point identify $\theta_{\text{local untreated}}$ given Assumption 2.1. The intuition of the identification of $\theta_{\text{local untreated}}$ is that conditioning on compliers implies that $T_i = Z_i$, thus, $\theta_{\text{local untreated}} = \theta_{\text{local}}$. We formally state the result in Claim 7.1.

Claim 7.1. Assume that Assumption 2.1 holds, then, $\theta_{\text{local untreated}}$ is point identifiable:

$$\theta_{\text{local untreated}} = \frac{\mathbb{P}[Y_i = 1 | Z_i = 1] - \mathbb{P}[Y_i = 1 | Z_i = 0]}{\mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 1]}.$$

7.1.2 Equivalence between the Approximated Persuasion Rate and the Local Persuasion Rate with One-Sided Non-Compliance

As summarized in DellaVigna and Gentzkow (2010), one popular estimand in the empirics of persuasion is the "approximated" persuasion rate $\tilde{\theta}_{\text{DK}}$:

$$\tilde{\theta}_{\text{DK}} = \frac{\mathbb{P}[Y_i = 1 | Z_i = 1] - \mathbb{P}[Y_i = 1 | Z_i = 0]}{\mathbb{P}[T_i = 1 | Z_i = 1] - \mathbb{P}[T_i = 1 | Z_i = 0]} \times \frac{1}{1 - \mathbb{P}[Y_i = 1 | Z_i = 0]}.$$

As noted in Jun and Lee (2018), $\tilde{\theta}_{\text{DK}}$ is not a well-defined conditional probability. Therefore, $\tilde{\theta}_{\text{DK}}$ does not measure persuasion rate for any subpopulation.

In this subsection, we present conditions for $\tilde{\theta}_{\text{DK}}$ to equal θ_{local} in experiments with one-sided non-compliance, which is empirically relevant in some persuasion experiments. For instance, non-compliance issues arise in the treatment group of the GOTV experiment in Green et al. (2003).

The results below show that for one-sided non-compliance, $\tilde{\theta}_{\text{DK}}$ equals θ_{local} under specific conditions on the distribution of potential outcomes and treatments. If there is one-sided non-compliance in the treatment group, the two estimands are equivalent if and only if the untreated potential outcome is independent of the treated potential treatment. If there is none-sided non-compliance in the control group, the two estimands are equal if and only if the proportion of untreated potential outcome being 0 among untreated potential treatment being 0 equals the proportion of never-persuaded among the never-takers.

Theorem 7.1. Assume that Assumption 2.1 holds, if there is one-sided non-compliance in the control group, then $\theta_{\text{DK}} = \theta_{\text{local}}$ if and only if $\mathbb{P}[Y_i(0) = 0 | T_i(0) = 0] = \mathbb{P}[Y_i(1) = 0 | T_i(0) = 1]$, if there is one-sided non-compliance in the treatment group, then $\theta_{\text{DK}} = \theta_{\text{local}}$ if and only if $Y_i(0) \perp\!\!\!\perp T_i(1)$.

These results contrast sharply with the results in Jun and Lee (2018), which state that these two quantities are equivalent to each other if: 1) $T_i = Z_i$ holds almost surely, that is, we are in the sharp persuasion design; 2) $T_i \perp\!\!\!\perp (Y_i(0), Y_i(1))$; 3) $Y_i(1) = Y_i(0) = 1$ for all i, or $Y_i(1) = Y_i(0) = 0$ for all i.

7.2 A Sharp Test of the Identification Assumptions

The main identification results in Theorem 5.2 rely on two assumptions: the IA IV assumptions and the monotone treatment response assumption. These assumptions impose restrictions on individuals' choice behaviors by ruling out the dissuaded and the defiers and are thus subject to criticism for being too strong. To address this issue, we propose a sharp test for Assumption 2.1.

The idea of the test proposed here closely relates to Balke and Pearl (1997). A binary IA IV model with monotone treatment response assumption implies that the observed quantity, say $\mathbb{P}[Y_i = 0, T_i = 0, Z_i = 0, X_i \in A]$, with A measurable, is a linear combination of the probability of the unobserved outcome and compliance types:

$$A_{\text{obs}} \mathbf{p} = \mathbf{b}, \tag{2}$$

where A_{obs} is a matrix that reflects the restrictions on the data, $\mathbf{p}$ is a vector of the unobserved persuasion and compliance types defined in Table 1, $\mathbf{b}$ is a collection of observed quantities, for example $\mathbb{P}[Y_i = 0, T_i = 0, X_i \in A \mid Z_i = 0]$.[11] Thus, the observed quantity $\mathbf{b}$ is consistent with Assumption 2.1 if there exists a solution to the system of linear equations in 2. We now summarize this observation to Proposition 7.1.

Proposition 7.1. If Assumption 2.1 holds, then, there exists $\mathbf{p} \geq \mathbf{0}$ such that $A_{\text{obs}} \mathbf{p} = \mathbf{b}$ for all measurable set A.

[11] We provide an example of A_{obs}, $\mathbf{p}$, and $\mathbf{b}$ in Appendix F.

An implication of Proposition 7.1 is that to test the validity of Assumption 2.1, for observed data $\{Y_i, T_i, Z_i, X_i\}_{i=1}^n$ that is an independently and identically distributed sample drawn from $P \in \mathbf{P}$, it suffices to test the null hypothesis:

$$H_0 : P \in \mathbf{P}_0 \text{ versus } H_1 : P \in \mathbf{P} \setminus \mathbf{P}_0 \tag{3}$$

where $\mathbf{P}_0 := \{P \in \mathbf{P} : \exists \mathbf{p} \geq \mathbf{0} \text{ s.t. } A_{\text{obs}}\mathbf{p} = \mathbf{b}\}$, which is the set of distributions that is consistent with Assumption 2.1. Thus, if H_0 is rejected, we have strong evidence against the validity of Assumption 2.1. However, if H_0 is not rejected, we cannot confirm the validity of Assumption 2.1. In this precise sense, Assumption 2.1 is a refutable but nonverifiable hypothesis (Kitagawa, 2015).

In terms for the implementation of testing 3, with discrete X_i, we can set A to be the support of X_i, and proceed the test using the recent advancement on testing whether there exists a nonnegative solution to a possibly under-determined system of linear equations with known coefficients (Bai et al., 2022; Fang et al., 2023). One computationally intensive, yet feasible method for testing H_0 proposed in Bai et al. (2022) is to use subsampling method. With the subsampling method, by using the classic results in Romano and Shaikh (2012), Bai et al. (2022) shows that the test controls size uniformly over $\mathbf{P}$. The test statistic in Bai et al. (2022) is given by:

$$T_n := \inf_{\mathbf{p} \geq \mathbf{0} : B\mathbf{p} = 1} \sqrt{n} \left| A_{\text{obs}}\mathbf{p} - \hat{\mathbf{b}} \right|,$$

where $\hat{\mathbf{b}}$ is an estimator of $\mathbf{b}$.[12] For the subsampling-based test, Bai et al. (2022) defines the following quantity:

$$L_n(t) := \frac{1}{N_n} \sum_{1 \leq 1 \leq N_n} \mathbb{1}\left\{ \inf_{\mathbf{p} \geq \mathbf{0} : B\mathbf{p} = 1} \sqrt{n} \left| A_{\text{obs}}\mathbf{p} - \hat{\mathbf{b}}_j \right| \leq t \right\},$$

where $N_n = \binom{n}{b}$, j indexes the jth subsample of size b, $\hat{\mathbf{b}}_j$ is $\hat{\mathbf{b}}$ evaluated at jth subset of the data. The subsampling-based test in Bai et al. (2022) is:

$$T_n^{\text{sub}} := \mathbb{1}\{T_n > L_n^{-1}(1-\alpha)\}.$$

7.3 Sensitivity Analysis: the Monotone Treatment Response Assumption

Besides testing the identification assumptions jointly in the previous subsection, we now develop a sensitivity analysis approach to help researchers assess to what extent the point identification results are sensitive to the monotone treatment response assumption. Note that we apply the sensitivity analysis to the identification results in Lemma 4.2.

The sensitivity analysis builds on the idea in Balke and Pearl (1997). Note that the marginal distribution of potential outcomes is the marginal distribution of the potential outcomes among compliers can be

[12] We choose ℓ_2 norm when computing the test statistic. One advantage of using ℓ_2 norm is that it formulates a convex optimization problem that can be efficiently solved by standard statistical software, say, R (Boyd and Vandenberghe, 2004; Fu et al., 2017). For more discussions on computing the test statistic, see Appendix G.

represented as the following linear systems of equations:

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbb{P}[Y_i(0)=0, Y_i(1)=0 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(0)=0, Y_i(1)=1 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(0)=1, Y_i(1)=0 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(0)=1, Y_i(1)=1 | T_i(1) > T_i(0)] \end{bmatrix} = \begin{bmatrix} \mathbb{P}[Y_i(0)=0 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(0)=1 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(1)=0 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(1)=1 | T_i(1) > T_i(0)]. \end{bmatrix}$$

Therefore, we can vary the size of $\mathbb{P}[Y_i(0)=1, Y_i(1)=0|T_i(1) > T_i(0)]$ to see how the point identification results for the joint distribution of potential outcomes change. Here, with known $\mathbb{P}[Y_i(0)=1, Y_i(1)=0|T_i(1) > T_i(0)]$, we can point identify $\mathbb{P}[Y_i(0)=0, Y_i(1)=0|T_i(1) > T_i(0)]$, $\mathbb{P}[Y_i(0)=0, Y_i(1)=1|T_i(1) > T_i(0)]$, and $\mathbb{P}[Y_i(0)=1, Y_i(1)=1|T_i(1) > T_i(0)]$ from the system of equations above.

8 Empirical Application: Revisit Green et al. (2003)

This section demonstrates the application of the methods using Green et al. (2003) as an example. First, we provide information on the empirical setup. Then, we illustrate our main identification results with data from Green et al. (2003). Finally, we conduct the test for the identification assumptions and sensitivity analysis.

8.1 Empirical Setup

Green et al. (2003) conducted randomized voter mobilization experiments before the November 6, 2001 election in the following six cities: Bridgeport, Columbus, Detroit, Minneapolis, Raleigh, and St. Paul. The instrument Z_i is a randomly assigned face-to-face contact from a coalition of nonpartisan student and community organizations, encouraging voters to vote. The treatment T_i is whether or not voters indeed received face-to-face contact. The outcome variable Y_i is voter turnout in various elections in 2001. There are two pre-treatment covariates that we are interested in. For the full sample, we are interested in whether or not voters voted in the 2000 presidential election. We also restrict the analysis to Bridgeport. For Bridgeport, we are interested in whether or not voters are Democrats. A summary statistics table is provided in Table 2.

Table 2: Summary Statistics in Green et al. (2003)

	Observations	Mean	Std. Dev.	Min	Max
Panel A: Full Sample					
Y_i: Vote	18,933	0.296	0.457	0	1
T_i: Take-up of the GOTV	18,933	0.136	0.342	0	1
Z_i: Assignment the GOTV	18,933	0.461	0.498	0	1
Voted in 2000	18,933	0.608	0.488	0	1
Panel B: Bridgeport					
Y_i: Vote	1,806	0.118	0.323	0	1
T_i: Take-up of the GOTV	1,806	0.137	0.344	0	1
Z_i: Assignment the GOTV	1,806	0.496	0.5	0	1
Democrat	1,806	0.539	0.499	0	1

Note: This table provides summary statistics for Green et al. (2003). Std. Dev. stands for standard deviation.

8.2 Empirical Results

We first present the results for the marginal and joint distribution of potential outcomes of compliers in Table 3. Our results reveal two interesting patterns. First, conditional on compliers, most of them are never-persuaded in both samples. Second, only 7.9% of voters are persuaded conditional on compliers in the full sample, and 13.9% of voters are persuaded conditional on compliers in Bridgeport.

Table 3: Distribution of Potential Outcomes in Green et al. (2003)

	Estimates	95% CI	95% AR CI
Panel A: Full Sample			
$\mathbb{P}[Y_i(0)=1 \mid T_i(1)>T_i(0)]$	0.302	(0.261, 0.343)	(0.263, 0.343)
$\mathbb{P}[Y_i(1)=1 \mid T_i(1)>T_i(0)]$	0.381	(0.364, 0.398)	(0.365, 0.397)
$\mathbb{P}[Y_i(0)=1, Y_i(1)=1 \mid T_i(1)>T_i(0)]$	0.302	(0.261, 0.343)	(0.263, 0.343)
$\mathbb{P}[Y_i(0)=0, Y_i(1)=0 \mid T_i(1)>T_i(0)]$	0.619	(0.602, 0.636)	(0.603, 0.635)
$\mathbb{P}[Y_i(0)=0, Y_i(1)=1 \mid T_i(1)>T_i(0)]$	0.079	(0.035, 0.123)	(0.036, 0.122)
Panel B: Bridgeport			
$\mathbb{P}[Y_i(0)=1 \mid T_i(1)>T_i(0)]$	0.111	(0.019, 0.202)	(0.02, 0.202)
$\mathbb{P}[Y_i(1)=1 \mid T_i(1)>T_i(0)]$	0.25	(0.197, 0.303)	(0.196, 0.303)
$\mathbb{P}[Y_i(0)=1, Y_i(1)=1 \mid T_i(1)>T_i(0)]$	0.111	(0.019, 0.202)	(0.02, 0.202)
$\mathbb{P}[Y_i(0)=0, Y_i(1)=0 \mid T_i(1)>T_i(0)]$	0.75	(0.697, 0.803)	(0.697, 0.804)
$\mathbb{P}[Y_i(0)=0, Y_i(1)=1 \mid T_i(1)>T_i(0)]$	0.139	(0.033, 0.245)	(.034, 0.244)

Note: This table provides estimated marginal and joint distributions of potential outcomes among compliers for Green et al. (2003). CI stands for confidence interval. AR stands for Anderson-Rubin.

We now apply Theorem 5.1 and Theorem 5.2 to this experiment. The results are presented in Table 4. For the full sample, the probability of voting in the 2000 presidential election conditional on the locally persuadable (that is, those who do not vote without the information treatment and compliers) is 60.3%. A more interesting finding is that the subpopulation of always-persuaded compliers has the highest probability (that is, 95.4%) of voting in the 2000 presidential election. The results show that if always-persuaded and complier voters vote in the low-profile local elections regardless of the GOTV intervention, they will very likely vote in the high-profile 2000 presidential elections. This empirical pattern is consistent with the robust findings on the persistent of voting behavior (Gerber et al., 2003). One potential explanation of the persistent of the voting behavior is that voting behavior is habit-forming (Gerber et al., 2003). As expected, the subpopulation of never-persuaders and compliers has the lowest probability of voting in the 2000 presidential election.

Another interesting finding is that the voting propensity in the 2000 presidential election of the persuaded and compliers is very close to the always-persuaded and compliers. It is consistent with the findings that GOTV experiments mobilize the high-propensity voters (Enos et al., 2014). One potential explanation is that the GOTV programs only mobilize the voters who are on the margin of not voting. Hence, the persuaded voters should have a voting propensity that is close to the always-persuaded voters.

For the Bridgeport sample, the most interesting result is that among compliers and persuaded, the chance of them being a Democrat is very high. However, its confidence interval is pretty wide. Mobilizing more Democrats in the school board election in Bridgeport has practical implications for two reasons. First, Democrats are more pro-union. Second, the turnout rate in school board elections is usually low.[13] The

[13] According to Green et al. (2003), the turnout rate in Bridgeport school board election in the control arm is 9.9%

mobilized voters might vote for pro-union candidates and help select candidates who were more likely to increase teachers' salaries and benefits and improve their working conditions (Anzia, 2011).

Table 4: Profiling Persuasion Types in Green et al. (2003)

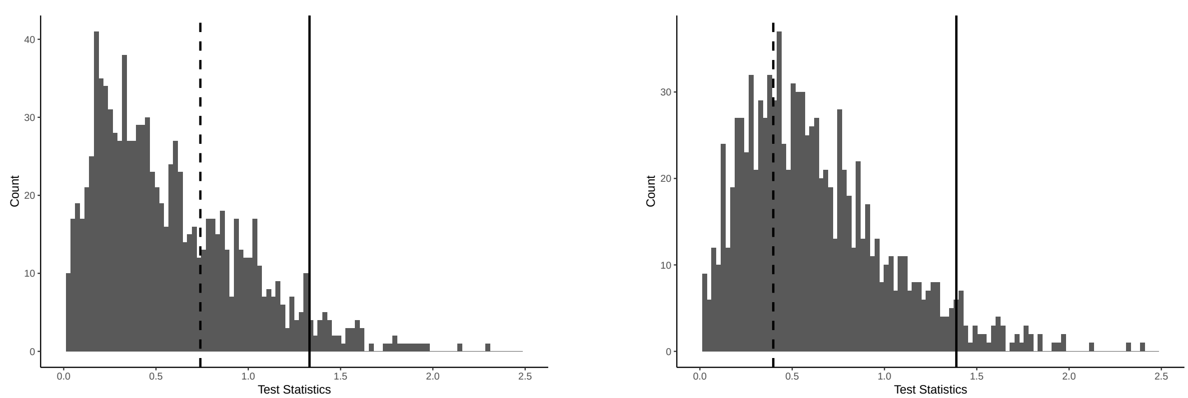
	Estimates	95% CI	95% AR CI
Panel A: Full Sample			
$\mathbb{P}[\text{Voted in 2000} = 1 \vert Y_i(0) = 0, T_i(1) > T_i(0)]$	0.603	(0.547, 0.659)	(0.549, 0.659)
$\mathbb{P}[\text{Voted in 2000} = 1 \vert Y_i(0) = 1, Y_i(1) = 1, T_i(1) > T_i(0)]$	0.954	(0.914, 0.994)	(0.914, 0.994)
$\mathbb{P}[\text{Voted in 2000} = 1 \vert Y_i(0) = 0, Y_i(1) = 0, T_i(1) > T_i(0)]$	0.511	(0.489, 0.534)	(0.489, 0.533)
$\mathbb{P}[\text{Voted in 2000} = 1 \vert Y_i(0) = 0, Y_i(1) = 1, T_i(1) > T_i(0)]$	0.885	(0.715, 1)	(0.657, 1)
Panel B: Bridgeport			
$\mathbb{P}[\text{Democrat} = 1 \vert Y_i(0) = 0, T_i(1) > T_i(0)]$	0.515	(0.35, 0.68)	(0.349, 0.681)
$\mathbb{P}[\text{Democrat} = 1 \vert Y_i(0) = 1, Y_i(1) = 1, T_i(1) > T_i(0)]$	0.507	(0.078, 0.935)	(0, 0.920)
$\mathbb{P}[\text{Democrat} = 1 \vert Y_i(0) = 0, Y_i(1) = 0, T_i(1) > T_i(0)]$	0.538	(.467, 0.609)	(0.467, 0.609)
$\mathbb{P}[\text{Democrat} = 1 \vert Y_i(0) = 0, Y_i(1) = 1, T_i(1) > T_i(0)]$	0.813	(0.437, 1)	(0.346, 1)

Note: This table provides the results of profiling different persuasion types using pre-treatment covariates. CI refers to confidence interval. AR refers to Anderson-Rubin.

8.3 Testing Identification Assumptions and Sensitivity Analysis

We implement the test for the Assumption 2.1 by using Proposition 7.1. We use the subsampling method in Bai et al. (2022) for this test.[14] The results in Figure 1 show that we cannot reject the validity of the identification assumptions at the 5% level for both the full sample and the Bridgeport sample. Furthermore, we provide the sensitivity analysis result on the joint distribution of potential outcomes in Table 5 by varying the degree to which the monotone treatment response assumption is violated among compilers. Interestingly, when the violation becomes larger, the proportion of persuaded among compliers increases.

Figure 1: Test Identification Assumptions using Bai et al. (2022)



(a) Full Sample

(b) Bridgeport

Note. These figures present the results for testing identification assumptions. Figure 1a presents the results using the full sample. Figure 1b presents the results using the sample from Bridgeport. The solid lines are the critical value for a 5% level test. The dashed lines are the test statistics.

[14]The subsampling test in Bai et al. (2022) requires us to pick a size for the subsample with $b_n \to \infty$ and $\frac{b_n}{n} \to 0$. We set b_n to $n^{\frac{2}{3}}$ here.

Table 5: Sensitivity for Distribution of Potential Outcomes in Green et al. (2003)

Panel A: Full Sample						
Sensitivity Parameter						
$\mathbb{P}[Y_i(1)=0, Y_i(0)=1 \vert T_i(1)>T_i(0)]$	0.1	0.12	0.14	0.16	0.18	0.2
Identified Parameters						
$\mathbb{P}[Y_i(1)=1, Y_i(0)=1 \vert T_i(1)>T_i(0)]$	0.202	0.182	0.162	0.142	0.122	0.102
$\mathbb{P}[Y_i(1)=0, Y_i(0)=0 \vert T_i(1)>T_i(0)]$	0.519	0.499	0.479	0.459	0.439	0.419
$\mathbb{P}[Y_i(1)=1, Y_i(0)=0 \vert T_i(1)>T_i(0)]$	0.179	0.199	0.219	0.239	0.259	0.279
Panel B: Bridgeport						
Sensitivity Parameter						
$\mathbb{P}[Y_i(1)=0, Y_i(0)=1 \vert T_i(1)>T_i(0)]$	0.05	0.06	0.07	0.08	0.09	0.1
Identified Parameters						
$\mathbb{P}[Y_i(1)=1, Y_i(0)=1 \vert T_i(1)>T_i(0)]$	0.061	0.051	0.041	0.031	0.021	0.011
$\mathbb{P}[Y_i(1)=0, Y_i(0)=0 \vert T_i(1)>T_i(0)]$	0.7	0.69	0.68	0.67	0.66	0.65
$\mathbb{P}[Y_i(1)=1, Y_i(0)=0 \vert T_i(1)>T_i(0)]$	0.189	0.199	0.209	0.219	0.229	0.239

Note: This table provides sensitivity analysis on the joint distribution of potential outcomes among compliers by varying the size of the dissuaded among compliers.

9 Conclusion

In the empirical study of persuasion, researchers often use a binary instrument to encourage individuals to consume information. The outcome of interest is also binary. Under the IA IV assumptions and the monotone treatment response assumption, we first show that it is possible to identify the joint distributions of potential outcomes among compliers. In other words, we can identify the percentage of the always-persuaded (that is, individuals who take the action of interest with and without the information treatment), the percentage of the never-persuaded (that is, individuals who do not take the action of interest with and without the information treatment), and the persuaded (that is, those who are mobilized by the treatment into taking the action of interest). These new quantities can thus provide richer information on the distribution of the treatment effects of the information treatment.

Furthermore, we develop a weighting method that helps researchers identify the statistical characteristics measured by the pre-treatment covariates of persuasion types: compliers and always-persuaded, compliers and persuaded, and compliers and never-persuaded. These findings extend the "κ weighting" results in Abadie (2003), which can profile the characteristics of compliers measured by pre-treatment covariates. This method can provide richer information on the treatment effect. For instance, some GOTV experiments aim at mobilizing underrepresented minorities. With this methodology, researchers can estimate the chance of the compliers and mobilizable voters being underrepresented minorities. Thus, researchers can assess whether or not their interventions achieve their normative goals.

To address the criticism on the monotone treatment response assumption, we provide two sets of solutions. First, we provide a sharp test on these two identification assumptions. The test boils down to testing whether there exists a nonnegative solution to a possibly under-determined system of linear equations with known coefficients. we also develop a simple sensitivity analysis to assess the sensitivity of the results with respect to the monotone treatment response assumption.

An application based on Green et al. (2003) is provided. The result shows that among compliers, roughly 11% voters are persuaded. Moreover, we find that among compliers, the chance for always-persuaded voters to vote in the 2000 presidential election is the highest, and the chance for never-persuaded voters to vote in the 2000 presidential election is the lowest. These results are consistent with the interpretation that voters' voting behaviors are habit-forming, hence are highly persistent (Gerber et al., 2003). Moreover, our results show that the voting propensity of those persuaded is close to those always-persuaded, which is consistent with the finding that GOTV programs mobilize high-propensity voters (Enos et al., 2014). Furthermore, in Bridgeport, the results show that the chance of being a Democrat among the persuaded voters and compliers in Bridgeport is high, though the estimate is quite noisy.

As pointed out in the paper, the results for the binary instrument can be easily generalized to discrete-valued instrument. However, the composition of compliers changes with any components in $\{z, z'\}$ changes. This creates an aggregation problem. Furthermore, with discrete-valued instrument, researchers can apply the partial identification approach in Mogstad et al. (2018) to partially identify the persuasion rate, which can help researchers assess the welfare impact of the information treatment. These constitute interesting topics for future research.

References

ABADIE, A. (2002): "Bootstrap tests for distributional treatment effects in instrumental variable models," *Journal of the American statistical Association*, 97, 284–292.

——— (2003): "Semiparametric instrumental variable estimation of treatment response models," *Journal of Econometrics*, 113, 231–263.

ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings," *Econometrica*, 70, 91–117.

ANZIA, S. F. (2011): "Election timing and the electoral influence of interest groups," *The Journal of Politics*, 73, 412–427.

ATHEY, S. AND S. WAGER (2021): "Policy learning with observational data," *Econometrica*, 89, 133–161.

BAI, Y., A. SANTOS, AND A. M. SHAIKH (2022): "On Testing Systems of Linear Inequalities with Known Coefficients," *Working Paper*.

BALKE, A. AND J. PEARL (1997): "Bounds on treatment effects from studies with imperfect compliance," *Journal of the American Statistical Association*, 92, 1171–1176.

BLATTMAN, C. AND J. ANNAN (2016): "Can employment reduce lawlessness and rebellion? A field experiment with high-risk men in a fragile state," *American Political Science Review*, 110, 1–17.

BLATTMAN, C., N. FIALA, AND S. MARTINEZ (2020): "The long-term impacts of grants on poverty: Nine-year evidence from Uganda's youth opportunities program," *American Economic Review: Insights*, 2, 287–304.

BLATTMAN, C., J. C. JAMISON, AND M. SHERIDAN (2017): "Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in Liberia," *American Economic Review*, 107, 1165–1206.

BOYD, S. AND L. VANDENBERGHE (2004): *Convex optimization*, Cambridge university press.

Carneiro, P. and S. Lee (2009): "Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality," *Journal of Econometrics*, 149, 191–208.

Chen, Y. and D. Y. Yang (2019): "The impact of media censorship: 1984 or brave new world?" *American Economic Review*, 109, 2294–2332.

Chernozhukov, V. and C. Hansen (2004): "The impact of 401 (k) participation on the wealth distribution: An instrumental quantile regression analysis," *Review of Economics and statistics*, 86, 735–751.

——— (2005): "An IV model of quantile treatment effects," *Econometrica*, 73, 245–261.

DellaVigna, S. and M. Gentzkow (2010): "Persuasion: empirical evidence," *Annual Review of Economics*, 2, 643–669.

DellaVigna, S. and E. Kaplan (2007): "The Fox News effect: Media bias and voting," *The Quarterly Journal of Economics*, 122, 1187–1234.

Durrett, R. (2010): *Probability: theory and examples*, Cambridge university press.

Enikolopov, R., M. Petrova, and E. Zhuravskaya (2011): "Media and political persuasion: Evidence from Russia," *American Economic Review*, 101, 3253–85.

Enos, R. D., A. Fowler, and L. Vavreck (2014): "Increasing inequality: The effect of GOTV mobilization on the composition of the electorate," *The Journal of Politics*, 76, 273–288.

Fang, Z., A. Santos, A. M. Shaikh, and A. Torgovitsky (2023): "Inference for Large-Scale Linear Systems With Known Coefficients," *Econometrica*, 91, 299–327.

Feng, Q., Q. Vuong, and H. Xu (2019): "Estimation of heterogeneous individual treatment effects with endogenous treatments," *Journal of the American Statistical Association*.

Fu, A., B. Narasimhan, and S. Boyd (2017): "CVXR: An R package for disciplined convex optimization," *arXiv preprint arXiv:1711.07582*.

Gerber, A. S., D. P. Green, and R. Shachar (2003): "Voting may be habit-forming: evidence from a randomized field experiment," *American journal of political science*, 47, 540–550.

Green, D. P., A. S. Gerber, and D. W. Nickerson (2003): "Getting out the vote in local elections: Results from six door-to-door canvassing experiments," *The Journal of Politics*, 65, 1083–1096.

Hansen, B. (2022): *Econometrics*, Princeton University Press.

Heckman, J. J. and E. Vytlacil (2005): "Structural equations, treatment effects, and econometric policy evaluation 1," *Econometrica*, 73, 669–738.

Huber, M. and G. Mellace (2015): "Testing instrument validity for LATE identification based on inequality moment constraints," *Review of Economics and Statistics*, 97, 398–411.

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 467–475.

IMBENS, G. W. AND D. B. RUBIN (1997): "Estimating outcome distributions for compliers in instrumental variables models," *The Review of Economic Studies*, 64, 555–574.

JUN, S. J. AND S. LEE (2018): "Identifying the effect of persuasion," *arXiv preprint arXiv:1812.02276*.

KÉDAGNI, D. AND I. MOURIFIÉ (2020): "Generalized instrumental inequalities: testing the instrumental variable independence assumption," *Biometrika*, 107, 661–675.

KIM, W., K. KWON, S. KWON, AND S. LEE (2018): "The identification power of smoothness assumptions in models with counterfactual outcomes," *Quantitative Economics*, 9, 617–642.

KITAGAWA, T. (2015): "A test for instrument validity," *Econometrica*, 83, 2043–2063.

KITAGAWA, T. AND A. TETENOV (2018): "Who should be treated? empirical welfare maximization methods for treatment choice," *Econometrica*, 86, 591–616.

LANDRY, C. E., A. LANGE, J. A. LIST, M. K. PRICE, AND N. G. RUPP (2006): "Toward an understanding of the economics of charity: Evidence from a field experiment," *The Quarterly journal of economics*, 121, 747–782.

MACHADO, C., A. M. SHAIKH, AND E. J. VYTLACIL (2019): "Instrumental variables and the sign of the average treatment effect," *Journal of Econometrics*, 212, 522–555.

MANSKI, C. (1997): "Monotone Treatment Response," *Econometrica*, 65, 1311–1334.

MANSKI, C. F. AND J. V. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010.

MOGSTAD, M., A. SANTOS, AND A. TORGOVITSKY (2018): "Using instrumental variables for inference about policy relevant treatment parameters," *Econometrica*, 86, 1589–1619.

MOURIFIÉ, I. AND Y. WAN (2017): "Testing local average treatment effect assumptions," *Review of Economics and Statistics*, 99, 305–313.

OKUMURA, T. AND E. USUI (2014): "Concave-monotone treatment response and monotone treatment selection: With an application to the returns to schooling," *Quantitative Economics*, 5, 175–194.

PEARL, J. (1999): "Probabilities of causation: three counterfactual interpretations and their identification," *Synthese*, 121, 93–149.

ROMANO, J. P. AND A. M. SHAIKH (2012): "On the uniform asymptotic validity of subsampling and the bootstrap," *The Annals of Statistics*, 40, 2798–2822.

RUSSELL, T. M. (2021): "Sharp bounds on functionals of the joint distribution in the analysis of treatment effects," *Journal of Business & Economic Statistics*, 39, 532–546.

STAIGER, D. AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica: Journal of the Econometric Society*, 557–586.

TORGOVITSKY, A. (2019): "Nonparametric inference on state dependence in unemployment," *Econometrica*, 87, 1475–1505.

Vuong, Q. and H. Xu (2017): "Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity," *Quantitative Economics*, 8, 589–610.

Vytlacil, E. (2002): "Independence, monotonicity, and latent index models: An equivalence result," *Econometrica*, 70, 331–341.

Wang, L., J. M. Robins, and T. S. Richardson (2017): "On falsification of the binary instrumental variable model," *Biometrika*, 104, 229–236.

Yamamoto, T. (2012): "Understanding the past: Statistical analysis of causal attribution," *American Journal of Political Science*, 56, 237–256.

Appendix A Identifiability of the Joint Distribution of Non-Binary Instruments or Outcomes

This section covers two potential directions for extending Lemma 4.2. The first direction explores the positive outcomes that arise from utilizing a non-binary instrument to extend Lemma 4.2. Following this, we delve into the negative outcomes associated with using a non-binary outcome to extend 4.2.

A.1 Non-Binary Instrument

Assumption 2.1 is adjusted to accommodate a discrete-valued instrument in two ways. Firstly, the IV monotonicity condition is crucially modified. With a discrete-valued instrument, the IV "monotonicity" condition must be satisfied for each pair of instruments. That is, changing the instrument from z to z' will either encourage or discourage every individual from taking up the treatment. Secondly, the IV relevance assumption is also revised. In this case, at least one instrument value must lead to changes in selection behavior. The formal statement of the revised assumption is now presented as Assumption A.1.

Assumption A.1. (Potential Outcome and Treatment Model with Discrete Valued Instrument)

1. Monotone treatment response: $Y_i(1) \geq Y_i(0)$ holds almost surely with $Y_i(0)$ and $Y_i(1)$ binary,
2. Exclusion restriction: $Y_i(t,z) = Y_i(t)$, for $t, z \in \operatorname{supp}(T_i, Z_i)$,
3. Exogenous instrument: $Z_i \perp\!\!\!\perp (Y_i(0), Y_i(1), T_i(0), T_i(1), X_i)$,
4. First stage: $\mathbb{P}[T_i = 1 | Z_i = z]$ is a non-trivial function of z,
5. IV Monotonicity: either $T_i(z) \geq T_i(z')$ or $T_i(z) \leq T_i(z')$ holds almost surely for $z \neq z'$ with $z, z' \in \operatorname{supp}(Z_i)$.

With Assumption A.1, we can point identify the joint distribution of potential outcomes among each complier group. The intuition of the result is that with Assumption A.1, the proof proceeds "as-if" we are using a binary IV with support being $\{z, z'\}$. We now formally state the results in Corollary A.1.

Corollary A.1. Suppose Assumption A.1 holds, conditional on z, z' compliers (that is, $z, z' \in \operatorname{supp}(Z_i)$ and $T_i(z) = T_i(z')$ does not hold almost surely), the joint distribution of potential outcome is point identified,:

$$
\begin{aligned}
\mathbb{P}[Y_i(1) = 1, Y_i(0) = 1 \mid T_i(z) \geq T_i(z')] &= \frac{\mathbb{P}[Y_i = 1, T_i = z' \mid Z_i = z'] - \mathbb{P}[Y_i = 1, T_i = z' | Z_i = z]}{\mathbb{E}[T_i | Z_i = z] - \mathbb{E}[T_i \mid Z_i = z']} \\
\mathbb{P}[Y_i(1) = 1, Y_i(0) = 0 \mid T_i(z) \geq T_i(z')] &= \frac{\mathbb{E}[Y_i \mid Z_i = z] - \mathbb{E}[Y_i \mid Z_i = z']}{\mathbb{E}[T_i \mid Z_i = z] - \mathbb{E}[T_i \mid Z_i = z']} \\
\mathbb{P}[Y_i(1) = 0, Y_i(0) = 0 \mid T_i(z) \geq T_i(z')] &= \frac{\mathbb{P}[Y_i = 0, T_i = z \mid Z_i = z] - \mathbb{P}[Y_i = 0, T_i = z | Z_i = z']}{\mathbb{E}[T_i \mid Z_i = z] - \mathbb{E}[T_i | Z_i = z']}.
\end{aligned}
$$

Just as with a discrete-valued instrument, the identification assumptions will be modified for a continuous instrument. These modifications concern the IV monotonicity and IV relevance assumptions. In this case, we use an indicator selection equation to describe the first stage selection process. With this representation, it is easy to characterize the treatment effect on different margins of self-selecting into the treatment. We also assume that at least one instrument value leads to changes in the treatment-taking behavior. Assumption A.2 formally states the identification assumptions for this scenario.

Assumption A.2. (Binary Treatment and Outcome Model with a Continuous Instrument)

1. $Y_i(0) \leq Y_i(1)$ holds almost surely, and $Y_i(0), Y_i(1) \in \{0, 1\}$,
2. $T_i(z) = \mathbb{1}\{V_i \leq \nu(z)\}$, where $\nu : \mathcal{Z} \to \mathbb{R}$ is a non-trivial measurable function with respect to z and assume without loss of generality that $V_i \sim U[0, 1]$,
3. $Z_i \perp\!\!\!\perp (Y_i(0), Y_i(1), V_i, X_i)$.

Before proceeding to present the identification results, we give two remarks related to Assumption A.2. First, the indicator selection equation is equivalent to the monotonicity condition in the IA IV model (Vytlacil, 2002). To see this, observe that a change in z induces a shift either toward or away from treatment for the support of V_i. Second, instead of assuming $V_i \sim U[0, 1]$, we can also assume V_i being continuously distributed. This implies that we can normalize the distribution of V_i to be uniformly distributed over $[0, 1]$. A consequence of this normalization is that $\nu(z) = P(z)$, where $P(z)$ is the propensity score: $P(z) \equiv \mathbb{P}[T_i = 1 | Z_i = z]$.

Corollary A.2. Assume that Assumption A.2 holds, furthermore, assume that $\text{supp}(P(Z_i)) = [0, 1]$, then, the joint distribution of potential outcomes at each margin of selecting into the treatment is identified:

$$\begin{aligned}
&\mathbb{P}[Y_i(1) = 1, Y_i(0) = 0 \mid V_i = v] = \frac{\partial}{\partial v}\mathbb{E}[Y_i \mid P(Z_i) = v], \\
&\mathbb{P}[Y_i(1) = Y_i(0) = 1 \mid V_i = v] = \mathbb{P}[Y_i = 1 \mid P(Z_i) = v, T_i = 0] - (1 - v)\frac{\partial \mathbb{P}[Y_i = 1 \mid P(Z_i) = v, T_i = 0]}{\partial v}, \\
&\mathbb{P}[Y_i(1) = Y_i(0) = 0 \mid V_i = v] = \mathbb{P}[Y_i = 0 \mid P(Z_i) = v, T_i = 1] + v\frac{\partial \mathbb{P}[Y_i = 0 \mid P(Z_i) = v, T_i = 1]}{\partial v}.
\end{aligned}$$

A.2 Non-Binary Outcome

We now discuss whether we can extend the identification of the joint distribution of potential outcomes in Lemma 4.2 to the case when the outcome is trinary. In the empirical study of persuasion, there are three possible outcomes: 0 is an outside option, 1 is the target action of persuasion, and -1 is any other action. Without the monotone treatment response assumption, we can classify individuals into nine types according to the potential outcomes.[15] Table 6 presents the classification.

With the trinary outcome, two types of monotone treatment response assumptions were made in the previous literature. Jun and Lee (2018) assumed that the information treatment has a monotone treatment effect on the target action of persuasion: we rule out the type of individuals who will take the action of

[15] Jun and Lee (2018) does not use the conventional potential outcome notation in their discussion. Jun and Lee (2018) first writes out the choice set facing agent i. They use the following notation: $S = \{0, 1, -1\}$. To write out agent i's potential outcomes, Jun and Lee (2018) uses the following notation: $Y_i(t) = (Y_{i0}(t), Y_{i1}(t), Y_{i,-1}(t))$, where $t \in \{0, 1\}$. $Y_{i0}(t)$ denotes whether the individual choose to take the action 0 if the treatment is t. $Y_{i1}(t)$ and $Y_{i,-1}(t)$ are defined similarly. Moreover, $\sum_{j \in S} Y_{ij}(t) = 1$ for $t \in \{0, 1\}$. That is, the choices in S are exclusive and exhaustive. It is easy to see that there is a duality between the notation in Jun and Lee (2018) and conventional potential outcome notation used in Table 6.

Table 6: Types of Individuals with Trinary Outcome

$Y_i(0)$	$Y_i(1)$
-1	-1
-1	0
-1	1
0^{**}	-1^{**}
0	0
0	1
1^{*}	-1^{*}
1^{*}	0^{*}
1	1

interest without being exposed to the treatment but will choose the outside action or any other action with being exposed to the treatment. In other words, with the monotone treatment response assumption made in Jun and Lee (2018), the seventh and eighth row (those with *) in Table 6 occur with probability zero.

A stronger monotone treatment response assumption was made in Manski (1997). The monotone treatment response assumption in Manski (1997) assumes that $Y_i(1) \geq Y_i(0)$ holds with probability one: the fourth row (those with **), and the seventh and the eighth rows (those with *) happen with zero probability. Manski (1997) further assumes out the type of individuals who will take the outside action without being exposed to the treatment but will take any other action with being exposed to the treatment.

Given the monotone treatment response assumption in Jun and Lee (2018), we know that there are seven unknown probabilities for the joint distribution of potential outcomes among compliers. Moreover, by the classic results of Imbens and Rubin (1997), we know that the marginal distribution of potential outcomes among compliers is point identifiable. Among compliers, the joint distribution of potential outcomes is a function of the marginal distribution of potential outcomes. In other words, we have a system of linear equations with six known probabilities of the marginal distribution of potential outcomes among compliers and seven unknown probabilities of the joint distribution of potential outcomes among compliers. Therefore, the marginal distribution of potential outcomes is not point identified given the monotonicity assumption in the trinary outcome case in Jun and Lee (2018).

A remaining question to ask is whether we can point identify the joint distribution of potential outcomes with the monotone treatment response assumption made in Manski (1997). Again, the answer is no. The reason is that even though we have six unknowns and six equations, the information in the data is repetitive. We formally state the show the impossibility results in the following Proposition.

Proposition A.1. Assume that the potential outcomes are trinary, i.e., $Y_i(t) \in \{-1, 0, 1\}$ for $t \in \{0, 1\}$. Furthermore, assume the following monotone treatment response assumption: $Y_i(1) \geq Y_i(0)$ holds with probability one. Moreover, assume assumptions 1 to 4 in Assumption 2.1 hold. Then, the joint distribution of potential outcomes among compliers is not point identified.

Even though we cannot point identify the joint distribution of potential outcomes among compliers in this case, We can still partially identify the joint distribution of potential outcomes among compliers using the approaches in Balke and Pearl (1997). For example, to construct sharp bounds for $\mathbb{P}[Y_i(0) = -1, Y_i(1) = -1 | T_i(1) > T_i(0)]$, we can form a linear program with the objective function being $\mathbb{P}[Y_i(0) = -1, Y_i(1) = -1 | T_i(1) > T_i(0)]$ and the constraints being the linear system of equations in the proof of Proposition A.1.

One way to restore the point identification of the joint distribution of potential outcomes with non-binary Y_i under the monotone treatment response and IA IV assumptions is to binarize the outcome variable. To see this, assume without loss of generality that $Y_i(1) \geq Y_i(0)$ holds almost surely. Define the following two binary random variables: $\mathbb{1}\{Y_i(1) \geq x\}$ and $\mathbb{1}\{Y_i(0) \geq x\}$ with $x \in \mathbb{R}$. Then, by the monotone treatment response, it follows immediately that $\mathbb{1}\{Y_i(1) \geq x\} \geq \mathbb{1}\{Y_i(0) \geq x\}$ holds almost surely. Thus, the results in Lemma 4.2 hold for the new binarized outcome variable.

Appendix B Profiling Compliers with a Non-Binary Instrument

In Appendix A.1, we have shown that the joint distribution of potential outcomes is identifiable with a non-binary instrument. As a result, the profiling results presented in Theorem 5.2 can be readily applied to this case. The profiling results for a discrete instrument and a continuous instrument are presented in Corollary B.1 and Corollary B.2 respectively.

Corollary B.1. Assume that Assumption A.1 holds, and let $g : \mathbb{R} \to \mathbb{R}$ be measurable with $\mathbb{E}[|g(X_i)|] < \infty$, then, conditional on z, z' compliers (that is, $z, z' \in \text{supp}(Z_i)$, $T_i(z) = T_i(z')$ does not hold almost surely, and assume without loss of generality that $T_i(z) \geq T_i(z')$ holds almost surely), the expectation of $g(X_i)$ is identified:

$$
\begin{aligned}
&\mathbb{E}[g(X_i)|Y_i(1) = Y_i(0) = 1, T_i(z) \geq T_i(z')] \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1, T_i = 0\}|Z_i = z'] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1, T_i = 0\}|Z_i = z\}]}{\mathbb{P}[Y_i = 1, T_i = 0|Z_i = z'] - \mathbb{P}[Y_i = 1, T_i = 0|Z_i = z]}, \\
&\mathbb{E}[g(X_i)|Y_i(1) = Y_i(0) = 0, T_i(z) \geq T_i(z')] \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0, T_i = 1\}|Z_i = z] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0, T_i = 1\}|Z_i = z']}{\mathbb{P}[Y_i = 0, T_i = 1|Z_i = z] - \mathbb{P}[Y_i = 0, T_i = 1|Z_i = z']}, \\
&\mathbb{E}[g(X_i)|Y_i(1) = 1, Y_i(0) = 0, T_i(z) \geq T_i(z')] \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1\}|Z_i = z] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1\}|Z_i = z']}{\mathbb{E}[Y_i|Z_i = z] - \mathbb{E}[Y_i|Z_i = z']}.
\end{aligned}
$$

Corollary B.2. Assume that Assumption A.2 holds, furthermore, assume that $\text{supp}(P(Z_i)) = [0, 1]$. Let $g : \mathbb{R} \to \mathbb{R}$ be measurable with $\mathbb{E}[|g(X_i)|] < \infty$, then, conditional at each margin of selecting into the treatment, the expectation of $g(X_i)$ is identified:

$$
\begin{aligned}
&\mathbb{E}[g(X_i) \mid Y_i(1) = 1, Y_i(0) = 0, V_i = v] \\
&= \frac{\frac{\partial}{\partial v}\mathbb{E}[g(X_i)Y_i \mid P(Z_i) = v]}{\frac{\partial}{\partial v}\mathbb{E}[Y_i \mid P(Z_i) = v]} \\
&\mathbb{E}[g(X_i) \mid Y_i(1) = Y_i(0) = 1, V_i = v] \\
&= \frac{\mathbb{E}[g(X_i)Y_i \mid P(Z_i) = v, T_i = 0] - (1 - v)\frac{\partial \mathbb{E}[g(X_i)Y_i|P(Z_i)=v,T_i=0]}{\partial v}}{\mathbb{P}[Y_i = 1 \mid P(Z_i) = v, T_i = 0] - (1 - v)\frac{\partial \mathbb{P}[Y_i=1|P(Z_i)=v,T_i=0]}{\partial v}} \\
&\mathbb{E}[g(X_i) \mid Y_i(1) = Y_i(0) = 0, V_i = v] \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0\} \mid P(Z_i) = v, T_i = 1] + v\frac{\partial \mathbb{E}[g(X_i)\mathbb{1}\{Y_i=0\}|P(Z_i)=v,T_i=1]}{\partial v}}{\mathbb{P}[Y_i = 0 \mid P(Z_i) = v, T_i = 1] + v\frac{\partial \mathbb{P}[Y_i=0|P(Z_i)=v,T_i=1]}{\partial v}}.
\end{aligned}
$$

Appendix C A Different Quantity of "Profiling"

A different quantity of interest is the following: conditional on compliers and the pretreatment covariates, the probability of being different persuasion types (i.e., always-persuaded, persuaded, never-persuaded). Given the strong IV independence assumption, such quantity is point identifiable because the strong IV independence assumption, we have:

$$\mathbb{P}[Y_i(1)=0, Y_i(0)=0 \mid T_i(1)>T_i(0), X_i] = \frac{\mathbb{P}[Y_i=1, T_i=0 \mid Z_i=0, X_i] - \mathbb{P}[Y_i=1, T_i=0 \mid Z_i=1, X_i]}{\mathbb{E}[T_i \mid Z_i=1, X_i] - \mathbb{E}[T_i \mid Z_i=0, X_i]},$$
$$\mathbb{P}[Y_i(1)=1, Y_i(0)=0 \mid T_i(1)>T_i(0), X_i] = \frac{\mathbb{E}[Y_i \mid Z_i=1, X_i] - \mathbb{E}[Y_i \mid Z_i=0, X_i]}{\mathbb{E}[T_i \mid Z_i=1, X_i] - \mathbb{E}[T_i \mid Z_i=0, X_i]},$$
$$\mathbb{P}[Y_i(1)=1, Y_i(0)=1 \mid T_i(1)>T_i(0), X_i] = \frac{\mathbb{P}[Y_i=0, T_i=1 \mid Z_i=1, X_i] - \mathbb{P}[Y_i=0, T_i=1 \mid Z_i=0, X_i]}{\mathbb{E}[T_i \mid Z_i=1, X_i] - \mathbb{E}[T_i \mid Z_i=0, X_i]}.$$

These quantities might be useful for optimal treatment allocation with non-compliance (Kitagawa and Tetenov, 2018; Athey and Wager, 2021). This is beyond the scope of this paper, and we leave it for future research.

Appendix D Identification: Always-Takers and Never-Takers

For always-takers, we observe their $Y_i(1)$. For never-takers, we observe their $Y_i(0)$. Therefore, the weighting method developed in Theorem 5.1 can be extended to always-takers and never-takers. The results are presented in Proposition D.1.

Proposition D.1. Assume that Assume that 1 to 4 in Assumption 2.1 hold, furthermore, assume that we observe pre-treatment covariates X_i, and let $g(\cdot)$ be any measurable real function of X_i such that $\mathbb{E}[|g(X_i)|] < \infty$, then, for $y \in \{0,1\}$, we have the following:

$$\mathbb{E}[g(X_i)|Y_i(1)=y, T_i(1)=T_i(0)=1] = \mathbb{E}[g(X_i)|Y_i=y, T_i=1, Z_i=0]$$
$$\mathbb{E}[g(X_i)|Y_i(0)=y, T_i(1)=T_i(0)=0] = \mathbb{E}[g(X_i)|Y_i=y, T_i=0, Z_i=1].$$

With the IA IV assumption, Proposition D.1 states that the conditional moments of X_i conditional on always-takers and their treated potential outcomes and the conditional moments of X_i conditional on never-takers and their untreated potential outcomes are identifiable. Furthermore, Proposition D.1 implies that the conditional cumulative distribution functions are identifiable. This follows because $g(x) = \mathbb{1}\{X_i \le x\}$ is a bounded measurable map.

For always-takers, if we further assume the monotone treatment response, we can identify the statistical characteristics measured by pre-treatment covariates of the never-persuaded and always-takers. For never-takers, if we further assume the monotone treatment response, we can identify the statistical characteristics measured by pre-treatment covariates of the always-persuaded and never-takers.

Appendix E More on Estimation and Inference

In this appendix, we offer more detailed discussions on the estimation and inference issues related to the estimands proposed in Section 4 and 5. Our first focus is on the estimation and inference results with strong identification. Afterward, we shift our discussion to the inference results when identification is weak.

E.1 Estimation and Inference under Strong Identification

Recall that our identification results give us the following β_{IV} estimand:

$$\beta_{IV} = \frac{\mathbb{E}[f(X_i, Y_i, T_i) \mid Z_i = 1] - \mathbb{E}[f(X_i, Y_i, T_i) \mid Z_i = 0]}{\mathbb{E}[h(Y_i, T_i) \mid Z_i = 1] - \mathbb{E}[h(Y_i, T_i) \mid Z_i = 0]}.$$

We can use the sample analog to estimate β_{IV}:

$$\hat{\boldsymbol{\beta}}_{IV} = \left(\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ Z_i \end{pmatrix} (1, h(Y_i, T_i)) \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ Z_i \end{pmatrix} f(X_i, Y_i, T_i) \right),$$

with $\hat{\beta}_{IV}$ being the second component of $\hat{\boldsymbol{\beta}}_{IV}$. Using a standard argument (e.g., see Chapter 12 in Hansen (2022)), we can show the consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}_{IV}$ under suitable regularity conditions. We now formally claim the results below.

Proposition E.1. Assume that the following conditions hold:

1. $\mathbb{E}[f(X_i, Y_i, T_i)^4] < \infty$,
2. $\mathbb{E}\left[\begin{pmatrix} 1 \\ Z_i \end{pmatrix} (1, Z_i) \right]$ is positive definite,
3. $\mathbb{E}\left[\begin{pmatrix} 1 \\ Z_i \end{pmatrix} (1, h(Y_i, T_i)) \right]$ is rull rank,
4. $\mathbb{E}\left[\begin{pmatrix} 1 \\ Z_i \end{pmatrix} e_i \right] = 0$, where e_i is the residual from regressing $f(X_i, Y_i, T_i)$ on $h(Y_i, T_i)$,
5. $\mathbb{E}[h(Y_i, T_i)^4] < \infty$,
6. $\mathbb{E}[Z_i^4] < \infty$,
7. $\Omega = \mathbb{E}\left[\begin{pmatrix} 1 \\ Z_i \end{pmatrix} (1, Z_i) e_i \right]$ is positive definite,

then, $\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}_{IV}\right)$ is asymptotically normal:

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}_{IV}\right) \xrightarrow{\mathcal{D}} N\left(0, \mathbb{E}\left[\begin{pmatrix} 1 \\ Z_i \end{pmatrix} (1, h(X_i, T_i)) \right]^{-1} \Omega \mathbb{E}\left[\begin{pmatrix} 1 \\ h(X_i, T_i) \end{pmatrix} (1, Z_i) \right]^{-1} \right).$$

Moreover, a consistent estimator for $\mathbb{E}\left[\begin{pmatrix}1\\ Z_i\end{pmatrix}(1, h(X_i, T_i))\right]^{-1}\Omega\mathbb{E}\left[\begin{pmatrix}1\\ h(X_i, T_i)\end{pmatrix}(1, Z_i)\right]^{-1}$ is:

$$\left(\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ Z_i\end{pmatrix}(1, h(X_i, T_i))\right)^{-1}\hat{\Omega}\left(\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ h(X_i, T_i)\end{pmatrix}(1, Z_i)\right)^{-1},$$

where $\hat{\Omega} = \left(\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ Z_i\end{pmatrix}(1, Z_i)\left(f(X_i, Y_i, T_i) - (1, h(Y_i, T_i))\hat{\boldsymbol{\beta}}_{IV}\right)\right)$.

Before we proceed, we now give a remark on the consistency of the estimator we proposed. Let $g(X_i) = \mathbb{1}\{X_i \leq x\}$, Theorem 5.1 shows that we can point identify the conditional distribution function among the locally persuadable:

$$\begin{aligned}
&\mathbb{P}[X_i \leq x | Y_i(0) = 0, T_i(1) > T_i(0)] \\
&= \frac{\mathbb{P}[X_i \leq x, Y_i = 0, T_i = 0 | Z_i = 0] - \mathbb{P}[X_i \leq x, Y_i = 0, T_i = 0 | Z_i = 1]}{\mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 1]}.
\end{aligned}$$

It is easy to see that $\hat{\mathbb{P}}[X_i \leq x | Y_i(0) = 0, T_i(1) > T_i(0)]$ is a (pointwise) consistent estimator for $\mathbb{P}[X_i \leq x | Y_i(0) = 0, T_i(1) > T_i(0)]$. By the same idea in the Glivenko-Cantelli Theorem (see, e.g., Theorem 2.4.7 in Durrett (2010)), we can strengthen the pointwise consistency to uniform consistency:

$$\sup_{x \in \mathbb{R}}\left|\hat{\mathbb{P}}[X_i \leq x | Y_i(0) = 0, T_i(1) > T_i(0)] - \mathbb{P}[X_i \leq x | Y_i(0) = 0, T_i(1) > T_i(0)]\right| \xrightarrow{\mathbb{P}} 0.$$

We prove this uniform consistent result in Appendix I.15.

E.2 An Anderson-Rubin Test under Weak Identification

Note that the estimand in Equation 1 is a function of two regression coefficients:

$$p = \frac{\beta_1}{\beta_2} \equiv \frac{\mathbb{E}[f(X_i, Y_i, T_i) \mid Z_i = 1] - \mathbb{E}[f(X_i, Y_i, T_i) \mid Z_i = 0]}{\mathbb{E}[h(Y_i, T_i) \mid Z_i = 1] - \mathbb{E}[h(Y_i, T_i) \mid Z_i = 0]}.$$

A concern regarding the asymptotic approximation discussed in the previous section is that the denominator β_2 may be close to zero. When faced with weak identification, the asymptotic approximation discussed earlier may not perform well. Fortunately, in the current exact identified scenario, we can use the Anderson-Rubin test to circumvent the issue of weak identification.

Note that under the null hypothesis $H_0 : p = p_0$, we have that $p_0\beta_2 - \beta_1 = 0$. Therefore, by using the delta method, the limiting distribution of $\sqrt{n}(p_0\hat{\beta}_1 - \hat{\beta}_2)$ under H_0 is:

$$\sqrt{n}(p_0\hat{\beta}_1 - \hat{\beta}_2) \xrightarrow{\mathcal{D}} N(0, \gamma),$$

where $\gamma = \text{Var}(\beta_1) - 2p_0\,\text{Cov}(\beta_1, \beta_2)) + p_0^2\,\text{Var}(\beta_2)$.

Therefore, a test statistic is:

$$T_n = \frac{n(p_0\hat{\beta}_1 - \hat{\beta}_2)^2}{\hat{\gamma}},$$

where $\hat{\gamma}$ is a consistent estimator for γ. By Slutsky's Lemma, we further know that:

$$T_n \xrightarrow{\mathcal{D}} \chi(1).$$

Using the AR statistic, we can form an AR test of $H_0 : p = p_0$ as:

$$\phi_{AR}(p_0) = \mathbb{1}\{T_n > \chi^2_{1,1-\alpha}\},$$

where $\chi^2_{1,1-\alpha}$ is the $1-\alpha$ quantile of χ^2_1 distribution. As noted by Staiger and Stock (1997), this yields a size-α test that is robust to weak identification. We then can form a level $1-\alpha$ weak-identification-robust confidence set by collecting the nonrejected values.

Appendix F A System of Equation for the Binary IV Model with Monotone Treatment Response

Assumption 1 to 4 in Assumption 2.1 implies the following system of linear equations:

$$A_{\text{obs}}\mathbf{p} = \mathbf{b},$$

where A_{obs}, $\mathbf{p}$, and $\mathbf{b}$ are defined as the following with A being a measurable set:

$$A_{\text{obs}} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix},$$

$$\mathbf{p} = \begin{bmatrix} \mathbb{P}[Y_i(0)=0, Y_i(1)=0, T_i(0)=0, T_i(1)=0, X_i \in A] \\ \mathbb{P}[Y_i(0)=0, Y_i(1)=0, T_i(0)=0, T_i(1)=1, X_i \in A] \\ \mathbb{P}[Y_i(0)=0, Y_i(1)=0, T_i(0)=1, T_i(1)=1, X_i \in A] \\ \mathbb{P}[Y_i(0)=0, Y_i(1)=1, T_i(0)=0, T_i(1)=0, X_i \in A] \\ \mathbb{P}[Y_i(0)=0, Y_i(1)=1, T_i(0)=0, T_i(1)=1, X_i \in A] \\ \mathbb{P}[Y_i(0)=0, Y_i(1)=1, T_i(0)=1, T_i(1)=1, X_i \in A] \\ \mathbb{P}[Y_i(0)=1, Y_i(1)=1, T_i(0)=0, T_i(1)=0, X_i \in A] \\ \mathbb{P}[Y_i(0)=1, Y_i(1)=1, T_i(0)=0, T_i(1)=1, X_i \in A] \\ \mathbb{P}[Y_i(0)=1, Y_i(1)=1, T_i(0)=1, T_i(1)=1, X_i \in A] \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} \mathbb{P}[Y_i = 0, T_i = 0, X_i \in A \mid Z_i = 0] \\ \mathbb{P}[Y_i = 0, T_i = 0, X_i \in A \mid Z_i = 1] \\ \mathbb{P}[Y_i = 0, T_i = 1, X_i \in A \mid Z_i = 0] \\ \mathbb{P}[Y_i = 0, T_i = 1, X_i \in A \mid Z_i = 1] \\ \mathbb{P}[Y_i = 1, T_i = 0, X_i \in A \mid Z_i = 0] \\ \mathbb{P}[Y_i = 1, T_i = 0, X_i \in A \mid Z_i = 1] \\ \mathbb{P}[Y_i = 1, T_i = 1, X_i \in A \mid Z_i = 0] \\ \mathbb{P}[Y_i = 1, T_i = 1, X_i \in A \mid Z_i = 1] \end{bmatrix}.$$

Appendix G Implementing the Test in Section 7.2

Recall that in Section 7.2, the test statistic is given by:

$$T_n := \inf_{\mathbf{p} \geq \mathbf{0}: B\mathbf{p} = 1} \sqrt{n} \left| A_{\text{obs}} \mathbf{p} - \hat{\mathbf{b}} \right|.$$

To compute the test statistic, we choose the ℓ_2 norm. Thus, the minimizer to the minimization problem in the test statistic can be obtained by solving:

$$\begin{aligned} &\min_{\mathbf{p}} \left|\left| A_{\text{obs}} \mathbf{p} - \hat{\mathbf{b}} \right|\right|_2 \\ &\text{subject to } \mathbf{p} \geq \mathbf{0}, \sum_{i=1}^{\dim(\mathbf{p})} p_i = 1, \end{aligned}$$

where the inequality in the constraint is interpreted to hold component-wise. Note that the minimizer of the optimization problem above is equivalent to the minimizer of the following minimization problem:

$$\begin{aligned} &\min_{\mathbf{p}} \mathbf{p}^T A_{\text{obs}}^T A_{\text{obs}} \mathbf{p} - 2\mathbf{p}^T A_{\text{obs}}^T \hat{\mathbf{b}} \\ &\text{subject to } \mathbf{p} \geq \mathbf{0}, \sum_{i=1}^{\dim(\mathbf{p})} p_i = 1, \end{aligned}$$

The minimization problem above is a convex problem (Boyd and Vandenberghe, 2004), and can be efficiently solved by using **CVXR** package in R (Fu et al., 2017).

After solving the optimal $\mathbf{p}^*$, we then can compute the test statistics by computing:

$$T_n = \sqrt{n} \left| A_{\text{obs}} \mathbf{p}^* - \hat{\mathbf{b}} \right|.$$

Appendix H An Equivalence Result

We now use the weighting methods developed in Abadie (2003) to derive the results in Theorem 5.1. The results in Abadie (2003) reweight the observations, which enables us to "find" the compliers and those who do not take the action of interest without being exposed to the treatment. We now formally state the results in Proposition H.1.

Proposition H.1. Assume that 1 to 4 in Assumption 2.1 hold, then, the distribution of X_i conditional on $[Y_i(0) = 0, T_i(1) > T_i(0)]$ is point identified. Let A be a measurable set:

$$\begin{aligned}
&\mathbb{P}[X_i \in A | Y_i(0) = 0, T_i(1) > T_i(0)] \\
&= \frac{\mathbb{P}[X_i \in A] \times (\mathbb{P}[T_i = 1 | X_i \in A, Z_i = 1] - \mathbb{P}[T_i = 1 | X_i \in A, Z_i = 0] - \mathbb{E}[\kappa_0 Y_i | X_i \in A])}{\mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 1]},
\end{aligned}$$

where $\kappa_0 = (1 - T_i) \frac{(1-Z_i) - \mathbb{P}[Z_i = 0]}{\mathbb{P}[Z_i = 0]\mathbb{P}[Z_i = 1]}$.

We can also show that the identification results in Theorem 5.1 and Proposition H.1 are equivalent. We formally state this equivalence result in Proposition H.2.

Proposition H.2. The identification results for $\mathbb{P}[X_i \in A \mid Y_i(0) = 0, T_i(1) > T_i(0)]$ in Theorem 5.1 and Proposition H.1 are equivalent.

Appendix I Proofs

I.1 Proof of Lemma 4.1

The results have been shown by Imbens and Rubin (1997) and Abadie (2003). Since the proof is brief, we include it here for completeness.

For $\mathbb{P}[Y_i(t) = y | T_i(1) > T_i(0)]$ where $y \in \{0, 1\}$ and $t \in \{0, 1\}$, we have the following:

$$\begin{aligned}
\mathbb{P}[Y_i(t) = y | T_i(1) > T_i(0)] &= \frac{\mathbb{P}[Y_i(t) = y, T_i(1) = 1, T_i(0) = 0]}{\mathbb{P}[T_i(1) = 1, T_i(0) = 0]} \\
&= \frac{\mathbb{P}[Y_i(t) = y, T_i(1) = 1, T_i(0) = 0]}{\mathbb{E}[T_i | Z_i = 1] - \mathbb{E}[T_i | Z_i = 0]},
\end{aligned}$$

where the second equality uses Lemma 2.1 in Abadie (2003).

For $\mathbb{P}[Y_i(t) = y, T_i(1) = 1, T_i(0) = 0]$ with $y \in \{0, 1\}$ and $t \in \{0, 1\}$:

$$\begin{aligned}
&\mathbb{P}[Y_i(t) = y, T_i(1) = 1, T_i(0) = 0] \\
&= \mathbb{P}[Y_i(t) = y, T_i(t) = t] - \mathbb{P}[Y_i(t) = y, T_i(t) = t, T_i(1-t) = t] \\
&= \mathbb{P}[Y_i(t) = y, T_i(t) = t] - \mathbb{P}[Y_i(t) = y, T_i(1-t) = t] \\
&= \mathbb{P}[Y_i(t) = y, T_i(t) = t | Z_i = t] - \mathbb{P}[Y_i(t) = y, T_i(1-t) = t | Z_i = 1-t] \\
&= \mathbb{P}[Y_i = y, T_i = t | Z_i = t] - \mathbb{P}[Y_i = y, T_i = t | Z_i = 1-t],
\end{aligned}$$

where the first and the second equality uses IV monotonicity in Assumption 2.1, the third equality uses IV exogeneity in Assumption 2.1. Now, the desired results follow immediately.

I.2 Proof of Lemma 4.2

By the monotone treatment response assumption in Assumption 2.1, $\mathbb{P}[Y_i(1) = 1, Y_i(0) = 1 | T_i(1) > T_i(0)] = \mathbb{P}[Y_i(0) = 1 | T_i(1) > T_i(0)]$. The desired result follows immediately from Lemma 4.1 that $\mathbb{P}[Y_i(0) = 1 | T_i(1) > T_i(0)]$ is identifiable.

The result for $\mathbb{P}[Y_i(1) = 0, Y_i(0) = 0 | T_i(1) > T_i(0)]$ can be derived analogously by observing that monotone treatment response assumption in Assumption 2.1 implies $[Y_i(1) = 0, Y_i(0) = 0] = [Y_i(1) = 0]$ and using Lemma 4.1.

For $\mathbb{P}[Y_i(1) = 1, Y_i(0) = 0 | T_i(1) > T_i(0)]$, note that the monotone treatment response assumption in Assumption 2.1 implies $\mathbb{P}[Y_i(1) = 1, Y_i(0) = 0 | T_i(1) > T_i(0)] = \mathbb{E}[Y_i(1) - Y_i(0) | T_i(1) > T_i(0)]$. By Theorem 1 in Imbens and Angrist (1994), $\mathbb{E}[Y_i(1) - Y_i(0) | T_i(1) > T_i(0)]$ is identifiable under the IA IV assumptions.

I.3 Proof of Theorem 5.1

For $\mathbb{E}[g(X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)]$:

$$\begin{aligned}
\mathbb{E}[g(X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)] &= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(0) = 0, T_i(1) > T_i(0)\}]}{\mathbb{P}[Y_i(0) = 0, T_i(1) > T_i(0)]} \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(0) = 0, T_i(1) > T_i(0)\}]}{\mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 1]},
\end{aligned}$$

where the second equality uses Lemma 4.1.

For $\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(0) = 0, T_i(1) > T_i(0)\}]$:

$$\begin{aligned}
&\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(0) = 0, T_i(1) > T_i(0)\}] \\
&= \mathbb{E}[g(X_i)\mathbb{1}\{Y_i(0) = 0, T_i(0) = 0\}] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i(0) = 0, T_i(1) = 0\}] \\
&= \mathbb{E}[g(X_i)\mathbb{1}\{Y_i(0) = 0, T_i(0) = 0\} \mid Z_i = 0] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i(0) = 0, T_i(1) = 0\} \mid Z_i = 1] \\
&= \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0, T_i = 0\} \mid Z_i = 0] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 0, T_i = 0\} \mid Z_i = 1],
\end{aligned}$$

where the first equality uses the IV monotonicity in Assumption 2.1, the second equality uses the IV independence in Assumption 2.1 and a fact that independence is preserved under measurable transform (e.g., see Theorem 2.1.6. in Durrett (2010)).

I.4 Proof of Proposition 5.1

The desired results follow immediately by using the identical arguments in Theorem 5.1.

I.5 Proof of Theorem 5.2

For $\mathbb{E}[g(X_i) \mid Y_i(1) = Y_i(0) = 1, T_i(1) > T_i(0)]$. Note that the monotone treatment response assumption in Assumption 2.1 implies $[Y_i(1) = Y_i(0) = 1] = [Y_i(0) = 1]$. Now, the desired result follows immediately from Proposition 5.1.

Similarly, by Proposition 5.1 and the fact that $[Y_i(1) = Y_i(0) = 0] = [Y_i(1) = 0]$ which is implied by the monotone treatment response assumption in Assumption 2.1, the desired result for $\mathbb{E}[g(X_i) \mid Y_i(1) = Y_i(0) = 1, T_i(1) > T_i(0)]$ follows immediately.

For $\mathbb{E}[g(X_i) \mid Y_i(1) = 1, Y_i(0) = 0, T_i(1) > T_i(0)]$, we have the following:

$$\begin{aligned}
&\mathbb{E}[g(X_i) \mid Y_i(1) = 1, Y_i(0) = 0, T_i(1) > T_i(0)] \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1) = 1, Y_i(0) = 0, T_i(1) > T_i(0)\}]}{\mathbb{P}[Y_i(1) = 1, Y_i(0) = 0, T_i(1) > T_i(0)]} \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1) = 1, Y_i(0) = 0, T_i(1) > T_i(0)\}]}{\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0]},
\end{aligned}$$

where the second equality uses Theorem 1 in Imbens and Angrist (1994).

For $\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1) = 1, Y_i(0) = 0, T_i(1) > T_i(0)\}]$:

$$\begin{aligned}
&\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1) = 1, Y_i(0) = 0, T_i(1) > T_i(0)\}] \\
&= \mathbb{E}[g(X_i)(Y_i(1) - Y_i(0))(T_i(1) - T_i(0))] \\
&= \mathbb{E}[g(X_i)(T_i(1)Y_i(1) + (1 - T_i(1))Y_i(0)] \\
&\quad - \mathbb{E}[g(X_i)(T_i(0)Y_i(1) + (1 - T_i(0))Y_i(0)] \\
&= \mathbb{E}[g(X_i)(T_i(1)Y_i(1) + (1 - T_i(1))Y_i(0) \mid Z_i = 1] \\
&\quad - \mathbb{E}[g(X_i)(T_i(0)Y_i(1) + (1 - T_i(0))Y_i(0) \mid Z_i = 0] \\
&= \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1\} \mid Z_i = 1] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1\} \mid Z_i = 0],
\end{aligned}$$

where the third equality uses the IV independence assumption in Assumption 2.1.

I.6 Proof of Proposition 5.2

First, consider $\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1) = 1\} \mid Y_i(0) = 1, T_i(1) > T_i(0)]$ and $\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(0) = 0\} \mid Y_i(1) = 0, T_i(1) > T_i(0)]$. For $t \in \{0, 1\}$:

$$\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(t) = t\} \mid Y_i(1-t) = t, T_i(1) > T_i(0)] = \mathbb{E}[g(X_i) \mid Y_i(1-t) = t, T_i(1) > T_i(0)],$$

where the equality follows from the outcome monotonicity assumption in Assumption 2.1.

Second, consider $\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1) = 1\} \mid Y_i(0) = 0, T_i(1) > T_i(0)]$ and $\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(0) = 0\} \mid Y_i(1) = 1, T_i(1) > T_i(0)]$. For $t \in \{0, 1\}$:

$$\begin{aligned}
&\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1-t) = 1-t\} \mid Y_i(t) = t, T_i(1) > T_i(0)] \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1-t) = 1-t, Y_i(t) = t, T_i(1) > T_i(0)\}]}{\mathbb{P}[Y_i(t) = t, T_i(1) > T_i(0)]} \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1\} \mid Z_i = 1] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i = 1\} \mid Z_i = 0]}{\mathbb{P}[Y_i = t, T_i = t \mid Z_i = t] - \mathbb{P}[Y_i = t, T_i = t \mid Z_i = 1-t]},
\end{aligned}$$

where the second equality uses Lemma 4.1 and Theorem 5.2.

Finally, consider $\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1)=0\} \mid Y_i(0)=0, T_i(1)>T_i(0)]$ and $\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(0)=1\} \mid Y_i(1)=1, T_i(1)>T_i(0)]$. For $t \in \{0,1\}$:

$$\begin{aligned}
&\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1-t)=t\} \mid Y_i(t)=t, T_i(1)>T_i(0)] \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1-t)=t, Y_i(t)=t, T_i(1)>T_i(0)\}]}{\mathbb{P}[Y_i(t)=t, T_i(1)>T_i(0)]} \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1-t)=t, T_i(1)>T_i(0)\}]}{\mathbb{P}[Y_i(t)=t, T_i(1)>T_i(0)]} \\
&= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i=t, T_i=1-t\} \mid Z_i=1-t] - \mathbb{E}[g(X_i)\mathbb{1}\{Y_i=t, T_i=1-t\} \mid Z_i=t]}{\mathbb{P}[Y_i=t, T_i=t \mid Z_i=t] - \mathbb{P}[Y_i=t, T_i=t \mid Z_i=1-t]},
\end{aligned}$$

where the second eqaulity uses the monotone treatment response assumption in Assumption 2.1, the third equality uses Lemma 4.1 and Theorem 5.2.

I.7 Proof of Proposition D.1

For $\mathbb{E}[g(X_i)|Y_i(t)=y, T_i(1)=T_i(0)=t]$, where $t \in \{0,1\}$ and $y \in \{0,1\}$, we have the following:

$$\begin{aligned}
\mathbb{E}[g(X_i)|Y_i(t)=y, T_i(1)=T_i(0)=t] &= \mathbb{E}[g(X_i)|Y_i(t)=y, T_i(1-t)=t] \\
&= \mathbb{E}[g(X_i)|Y_i(t)=y, T_i(1-t)=t, Z_i=1-t] \\
&= \mathbb{E}[g(X_i)|Y_i=y, T_i=t, Z_i=1-t],
\end{aligned}$$

where the first equality uses the IV monotonicity assumption in Assumption 2.1, the second equality uses the IV independence assumption in Assumption 2.1.

I.8 Proof of Claim 7.1

Note that among compliers, $T_i = Z_i$. Now the desired result follows immediately by observing that Z_i is exogenous assumed in Assumption 2.1 and using Theorem 6 in Jun and Lee (2018).

I.9 Proof of Theorem 7.1

Recall the formulas of the approximated $\tilde{\theta}_{\mathrm{DK}}$ and the identified θ_{local} from Theorem 6 in Jun and Lee (2018):

$$\begin{aligned}
\tilde{\theta}_{\mathrm{DK}} &= \frac{\mathbb{P}[Y_i=1|Z_i=1] - \mathbb{P}[Y_i=1|Z_i=0]}{(\mathbb{P}[T_i=1|Z_i=1] - \mathbb{P}[T_i=1|Z_i=0]) \times (1-\mathbb{P}[Y_i=1|Z_i=0])} \\
\theta_{\mathrm{local}} &= \frac{\mathbb{P}[Y_i=1|Z_i=1] - \mathbb{P}[Y_i=1|Z_i=0]}{\mathbb{P}[Y_i=0, T_i=0|Z_i=0] - \mathbb{P}[Y_i=0, T_i=0|Z_i=1]},
\end{aligned}$$

thus, $\tilde{\theta}_{\mathrm{DK}} = \theta_{\mathrm{local}}$ if and only if:

$$(\mathbb{P}[T_i=1|Z_i=1] - \mathbb{P}[T_i=1|Z_i=0]) \times \mathbb{P}[Y_i=0|Z_i=0]$$

$$= \mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 1]. \tag{4}$$

Consider the first case in which there is non-compliance in the control group, i.e., $\mathbb{P}[T_i = 1 | Z_i = 1] = 1$. In this case, there is no never-taker. Then, for the denominator of $\tilde{\theta}_{\text{DK}}$:

$$\begin{aligned}
&(\mathbb{P}[T_i = 1|Z_i = 1] - \mathbb{P}[T_i = 1|Z_i = 0]) \times (1 - \mathbb{P}[Y_i = 1|Z_i = 0]) \\
&= (1 - \mathbb{P}[T_i = 1|Z_i = 0]) \times (\mathbb{P}[Y_i = 0|Z_i = 0]) \\
&= \mathbb{P}[T_i = 0|Z_i = 0] \times (\mathbb{P}[Y_i = 0, T_i = 0|Z_i = 0] + \mathbb{P}[Y_i = 0, T_i = 1|Z_i = 0]) \\
&= \mathbb{P}[T_i(0) = 0] \times (\mathbb{P}[Y_i(0) = 0, T_i(0) = 0] + \mathbb{P}[Y_i(1) = 0, T_i(0) = 1]),
\end{aligned}$$

where the first equality uses the assumption that there is non-compliance in the control group. For the denominator of $\tilde{\theta}_{\text{DK}}$, by the assumption that there is non-compliance in the control group:

$$\begin{aligned}
&\mathbb{P}[Y_i = 0, T_i = 0|Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 1] \\
&= \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 0] \\
&= \mathbb{P}[Y_i(0) = 0, T_i(0) = 0].
\end{aligned}$$

Thus, by Equation 4, $\tilde{\theta}_{\text{DK}} = \theta_{\text{local}}$ if and only if:

$$\begin{aligned}
&\mathbb{P}[Y_i(0) = 0, T_i(0) = 0] = \mathbb{P}[T_i(0) = 0] \times (\mathbb{P}[Y_i(0) = 0, T_i(0) = 0] + \mathbb{P}[Y_i(1) = 0, T_i(0) = 1]) \\
&\Leftrightarrow \mathbb{P}[T_i(0) = 1] \times \mathbb{P}[Y_i(0) = 0, T_i(0) = 0] = \mathbb{P}[T_i(0) = 0] \times \mathbb{P}[Y_i(1) = 0, T_i(0) = 1] \\
&\Leftrightarrow \mathbb{P}[Y_i(0) = 0|T_i(0) = 0] = \mathbb{P}[Y_i(1) = 0|T_i(0) = 1].
\end{aligned}$$

Consider the second case in which there is non-compliance in the treatment group, i.e., $\mathbb{P}[T_i = 0|Z_i = 0] = 1$. In this case, there is no always-taker. Then, for the denominator of $\tilde{\theta}_{\text{DK}}$:

$$\begin{aligned}
&(\mathbb{P}[T_i = 1|Z_i = 1] - \mathbb{P}[T_i = 1|Z_i = 0]) \times (1 - \mathbb{P}[Y_i = 1|Z_i = 0]) \\
&= \mathbb{P}[T_i = 1|Z_i = 1] \times \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 0] \\
&= \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 0] - \mathbb{P}[T_i = 0|Z_i = 1] \times \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 0],
\end{aligned}$$

where the first equality uses the assumption that there is non-compliance in the treatment group. Thus, by Equation 4, $\tilde{\theta}_{\text{DK}} = \theta_{\text{local}}$ if and only if:

$$\begin{aligned}
&\mathbb{P}[Y_i = 0, T_i = 0|Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 1] \\
&= \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 0] - \mathbb{P}[T_i = 0|Z_i = 1] \times \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 0] \\
&\Leftrightarrow \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 1] = \mathbb{P}[T_i = 0|Z_i = 1] \times \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 0] \\
&\Leftrightarrow \mathbb{P}[Y_i(0) = 0, T_i(1) = 0] = \mathbb{P}[T_i(1) = 0] \times \mathbb{P}[Y_i(0) = 0, T_i(0) = 0] \\
&\Leftrightarrow \mathbb{P}[Y_i(0) = 0|T_i(1) = 0] = \mathbb{P}[Y_i(0) = 0] \\
&\Leftrightarrow Y_i(0) \perp\!\!\!\perp T_i(1),
\end{aligned}$$

where the third line uses the assumption that $\mathbb{P}[T_i(0) = 0] = 1$.

I.10 Proof of Proposition A.1

Note that the marginal distribution of potential outcomes among compliers is point identified (Imbens and Rubin, 1997; Abadie, 2003). Moreover, we can rewrite the marginal distribution of potential outcomes among compliers as a system of linear equations of the joint distribution of potential outcomes among compliers:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mathbb{P}[Y_i(0) = -1, Y_i(1) = -1 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(0) = -1, Y_i(1) = 0 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(0) = -1, Y_i(1) = 1 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(0) = 0, Y_i(1) = 0 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(0) = 0, Y_i(1) = 1 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(0) = 1, Y_i(1) = 1 | T_i(1) > T_i(0)] \end{bmatrix} = \begin{bmatrix} \mathbb{P}[Y_i(0) = -1 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(0) = 0 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(0) = 1 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(1) = -1 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(1) = 0 | T_i(1) > T_i(0)] \\ \mathbb{P}[Y_i(1) = 1 | T_i(1) > T_i(0)] \\ 1 \end{bmatrix},$$

where the rank of the coefficient matrix is five. Thus, there is no unique solution to the system of linear equations above.

I.11 Proof of Corollary A.1

The desired results follow immediately using the identical arguments in Lemma 4.1 and Lemma 4.2.

I.12 Proof of Corollary A.2

The desired result follows immediately by using the result in Heckman and Vytlacil (2005) and Carneiro and Lee (2009) and the monotone treatment response assumption in Assumption A.1. Since the argument is brief, we include it here for completeness.

Note that by the monotone treatment response assumption in Assumption A.2 and the fact that Y_i is binary:

$$\begin{aligned} \mathbb{P}[Y_i(1) = 1, Y_i(0) = 0 \mid V_i = v] &= \mathbb{E}[Y_i(1) - Y_i(0) \mid V_i = v] \\ \mathbb{P}[Y_i(1) = Y_i(0) = 1 \mid V_i = v] &= \mathbb{P}[Y_i(0) = 1 \mid V_i = v] \\ \mathbb{P}[Y_i(1) = Y_i(0) = 0 \mid V_i = v] &= \mathbb{P}[Y_i(1) = 0 \mid V_i = v]. \end{aligned}$$

To identify $\mathbb{E}[Y_i(1) - Y_i(0) \mid V_i = v]$, consider $\mathbb{E}[Y_i \mid V_i = v]$:

$$\begin{aligned} \mathbb{E}[Y_i \mid V_i = v] &= \mathbb{E}[Y_i(0) \mid P(Z_i) = v] + \mathbb{E}[T_i(Y_i(1) - Y_i(0)) \mid P(Z_i) = v] \\ &= \mathbb{E}[Y_i(0) \mid P(Z_i) = v] + \mathbb{E}[Y_i(1) - Y_i(0) \mid T_i = 1, P(Z_i) = v]\mathbb{P}[T_i = 1 \mid P(Z_i) = v] \\ &= \mathbb{E}[Y_i(0) \mid P(Z_i) = v] + \mathbb{E}[Y_i(1) - Y_i(0) \mid V_i \le v, P(Z_i) = v]\mathbb{P}[V_i \le v \mid P(Z_i) = v] \\ &= \mathbb{E}[Y_i(0)] + \mathbb{E}[Y_i(1) - Y_i(0) \mid V_i \le v]v \\ &= \mathbb{E}[Y_i(0)] + \mathbb{E}[(Y_i(1) - Y_i(0))\mathbb{1}\{V_i \le v\}] \end{aligned}$$

$$= \mathbb{E}[Y_i(0)] + \mathbb{E}[\mathbb{1}\{V_i \leq v\}\mathbb{E}[Y_i(1) - Y_i(0) \mid V_i = u]]$$
$$= \mathbb{E}[Y_i(0)] + \int_0^v \mathbb{E}[Y_i(1) - Y_i(0) \mid V_i = u]du,$$

where the third equality uses the selection equation in Assumption A.2, the fourth equality uses the independence of Z_i and $V_i \sim U[0,1]$ in Assumption A.2. Now the desired result follows immediately by differentiating both sides of the equation with respect to v.

To identify $\mathbb{P}[Y_i(0) = 1 \mid V_i = v]$, consider $(1-v)\mathbb{E}[g(Y_i) \mid P(Z_i) = v, T_i = 0]$, where g is a measurable map:

$$\begin{aligned}(1-v)\mathbb{E}[g(Y_i) \mid P(Z_i) = v, T_i = 0] &= (1-v)\mathbb{E}[g(Y_i(0)) \mid V_i > v] \\ &= \mathbb{E}[g(Y_i(0))\mathbb{1}\{V_i > v\}] \\ &= \mathbb{E}[\mathbb{1}\{V_i > v\}\mathbb{E}[g(Y_i(0)) \mid V_i = u]] \\ &= \int_v^1 \mathbb{E}[g(Y_i(0)) \mid V_i = u]du,\end{aligned}$$

where the first equality uses the selection equation in Assumption A.2, the fourth equality uses $V_i \sim U[0,1]$ in Assumption A.2. Now the desired result follows immediately by differentiating both sides of the equation with respect to v and defining g as: $g(Y_i) = \mathbb{1}\{Y_i = 1\}$.

To identify $\mathbb{P}[Y_i(1) = 0 \mid V_i = v]$, consider $v\mathbb{E}[g(Y_i) \mid P(Z_i) = v, T_i = 1]$, where g is a measurable map:

$$\begin{aligned}v\mathbb{E}[g(Y_i) \mid P(Z_i) = v, T_i = 1] &= v\mathbb{E}[g(Y_i(1)) \mid V_i \leq v] \\ &= \mathbb{E}[g(Y_i(1))\mathbb{1}\{V_i \leq v\}] \\ &= \mathbb{E}[\mathbb{1}\{V_i \leq v\}\mathbb{E}[g(Y_i(1)) \mid V_i = u]] \\ &= \int_0^v \mathbb{E}[g(Y_i(1)) \mid V_i = u]du,\end{aligned}$$

where the first equality uses the selection equation in Assumption A.2, the fourth equality uses $V_i \sim U[0,1]$ in Assumption A.2. Now the desired result follows immediately by differentiating both sides of the equation with respect to v and defining g as: $g(Y_i) = \mathbb{1}\{Y_i = 0\}$.

I.13 Proof of Corollary B.1

The desired results follow immediately using the identical arguments in Theorem 5.2.

I.14 Proof of Corollary B.2

For $\mathbb{E}[g(X_i) \mid Y_i(1) = 1, Y_i(0) = 0, V_i = v]$:

$$\begin{aligned}\mathbb{E}[g(X_i) \mid Y_i(1) = 1, Y_i(0) = 0, V_i = v] &= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(1) = 1, Y_i(0) = 0\} \mid V_i = v]}{\mathbb{P}[Y_i(1) = 1, Y_i(0) = 0 \mid V_i = v]} \\ &= \frac{\mathbb{E}[g(X_i)(Y_i(1) - Y_i(0)) \mid V_i = v]}{\mathbb{E}[Y_i(1) - Y_i(0) \mid V_i = v]}\end{aligned}$$

$$= \frac{\frac{\partial}{\partial v}\mathbb{E}[g(X_i)Y_i \mid P(Z_i) = v]}{\frac{\partial}{\partial v}\mathbb{E}[Y_i \mid P(Z_i) = v]},$$

where the second equality uses the monotone treatment response assumption, and the third equality uses the independence assumption in Assumption A.2 and Corollary A.2.

Now. consider $\mathbb{E}[g(X_i) \mid Y_i(1) = Y_i(0) = 1, V_i = v]$ and $\mathbb{E}[g(X_i) \mid Y_i(1) = Y_i(0) = 0, V_i = v]$. For $t \in \{0,1\}$:

$$\begin{aligned}\mathbb{E}[g(X_i) \mid Y_i(1) = Y_i(0) = 1-t, V_i = v] &= \mathbb{E}[g(X_i) \mid Y_i(t) = 1-t, V_i = v] \\ &= \frac{\mathbb{E}[g(X_i)\mathbb{1}\{Y_i(t) = 1-t\} \mid V_i = v]}{\mathbb{P}[Y_i(t) = 1-t \mid V_i = v]},\end{aligned}$$

where the second equality uses the monotone treatment response assumption. Now the desired result follows immediately from the independence assumption in Assumption A.2 and Corollary A.2.

I.15 A Glivenko-Cantelli Theorem for Conditional Cumulative Distribution Function

In fact, we can strengthen the statement in Appendix E from convergence in probability to almost sure convergence:

$$\sup_{x\in\mathbb{R}} \left|\hat{\mathbb{P}}[X_i \le x|Y_i(0) = 0, T_i(1) > T_i(0)] - \mathbb{P}[X_i \le x|Y_i(0) = 0, T_i(1) > T_i(0)]\right| \xrightarrow{\text{a.s.}} 0.$$

Moreover, the uniform convergence result follows immediately from the uniform convergence of the empirical conditional cumulative distribution function. Thus, we only provide a proof for the uniform convergence of the empirical conditional cumulative distribution function in this section.

Theorem I.1. Consider a pair of random variable $(X_i, Z_i) : (\Omega, \mathcal{F}) \to (\mathbb{R}^2, \sigma(\mathcal{B}(\mathbb{R}^2)))$, where $\mathcal{F}$ is a sigma field on the outcome space Ω, and $\sigma(\mathcal{B}(\mathbb{R}^2))$ denotes the Borel sigma algebra on $\mathbb{R}^2$. Let $A \in \sigma(\mathcal{B}(\mathbb{R}^2))$ with $\mathbb{P}[Z_i \in A] \neq 0$. Then:

$$\sup_{x\in\mathbb{R}} \left|\hat{\mathbb{P}}[X_i \le x|Z_i \in A] - \mathbb{P}[X_i \le x|Z_i \in A]\right| \xrightarrow{\text{a.s.}} 0,$$

where $\hat{\mathbb{P}}[X_i \le x|Z_i \in A] = \frac{\mathbb{E}_n[\mathbb{1}\{X_i \le x, Z_i \in A\}]}{\mathbb{E}_n[\mathbb{1}\{Z_i \in A\}]}$ with $\mathbb{E}_n$ denotes sample average.

Proof. We first show that $\sup_{x\in\mathbb{R}} |\mathbb{E}_n[X_i \le x, Z_i \in A] - \mathbb{P}[X_i \le x, Z_i \in A]| \xrightarrow{\text{a.s.}} 0$. For $1 \le j \le k-1$, let $x_{j,k} = \inf\{y : \mathbb{P}[X_i \le x, Z_i \in A] \ge \frac{j}{k}\mathbb{P}[Z_i \in A]\}$. Thus, by the Strong Law of Large Numbers, there exists N_k such that if $n \ge N_k$, then:

$$\begin{aligned}&|\mathbb{E}_n[Z_i \in A] - \mathbb{P}[Z_i \in A]| < \frac{\mathbb{P}[Z_i \in A]}{k}, \\ &\left|\mathbb{E}_n[X_i \le x_{j,k}, Z_i \in A] - \mathbb{P}[Z_i \in A]\right| < \frac{\mathbb{P}[Z_i \in A]}{k}, \\ &\left|\mathbb{E}_n[X_i < x_{j,k}, Z_i \in A] - \mathbb{P}[X_i < x_{j,k} Z_i \in A]\right| < \frac{\mathbb{P}[Z_i \in A]}{k},\end{aligned}$$

for $1 \le j \le k-1$. With $x_{0,k} = -\infty$ and $x_{k,k} = \infty$, then the last two inequalities hold for $j = 0$ and $j = k$.

For $x \in (x_{j-1,k}, x_{j,k})$ with $1 \le j \le k$ and $n \ge N_k$:

$$\begin{aligned}
\mathbb{E}_n[X_i \le x, Z_i \in A] &\le \mathbb{E}_n[X_i < x_{j,k}, Z_i \in A] \le \mathbb{E}[X_i < x_{j,k}, Z_i \in A] + \frac{\mathbb{P}[Z_i \in A]}{k} \\
&\le \mathbb{E}[X_i < x_{j-1,k}, Z_i \in A] + \frac{2\mathbb{P}[Z_i \in A]}{k} \le \mathbb{E}[X_i \le x, Z_i \in A] + \frac{2\mathbb{P}[Z_i \in A]}{k}, \\
\mathbb{E}_n[X_i \le x, Z_i \in A] &\ge \mathbb{E}_n[X_i \le x_{j-1,k}, Z_i \in A] \ge \mathbb{E}[X_i \le x_{j-1,k}, Z_i \in A] - \frac{\mathbb{P}[Z_i \in A]}{k} \\
&\ge \mathbb{E}[X_i \le x_{j,k}, Z_i \in A] - \frac{2\mathbb{P}[Z_i \in A]}{k} \ge \mathbb{E}[X_i \le x, Z_i \in A] - \frac{2\mathbb{P}[Z_i \in A]}{k},
\end{aligned}$$

thus, we conclude that $\sup_{x\in\mathbb{R}} |\mathbb{E}_n[X_i \le x, Z_i \in A] - \mathbb{P}[X_i \le x, Z_i \in A]| \xrightarrow{\text{a.s.}} 0$.

For $\sup_{x\in\mathbb{R}} \left|\hat{\mathbb{P}}[X_i \le x | Z_i \in A] - \mathbb{P}[X_i \le x | Z_i \in A]\right|$:

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} \left|\hat{\mathbb{P}}[X_i \le x | Z_i \in A] - \mathbb{P}[X_i \le x | Z_i \in A]\right| \\
&= \sup_{x\in\mathbb{R}} \left| \frac{\mathbb{E}_n[\mathbb{1}\{X_i \le x, Z_i \in A\}]}{\mathbb{E}_n[\mathbb{1}\{Z_i \in A\}]} - \mathbb{P}[X_i \le x | Z_i \in A] \right| \\
&= \sup_{x\in\mathbb{R}} \left| \frac{\mathbb{E}_n[\mathbb{1}\{X_i \le x, Z_i \in A\}]}{\mathbb{E}_n[\mathbb{1}\{Z_i \in A\}]} - \frac{\mathbb{E}_n[\mathbb{1}\{X_i \le x, Z_i \in A\}]}{\mathbb{P}[\{Z_i \in A\}]} + \frac{\mathbb{E}_n[\mathbb{1}\{X_i \le x, Z_i \in A\}]}{\mathbb{P}[\{Z_i \in A\}]} - \mathbb{P}[X_i \le x | Z_i \in A] \right| \\
&\le \sup_{x\in\mathbb{R}} \left| \frac{\mathbb{E}_n[\mathbb{1}\{X_i \le x, Z_i \in A\}]}{\mathbb{E}_n[\mathbb{1}\{Z_i \in A\}]} - \frac{\mathbb{E}_n[\mathbb{1}\{X_i \le x, Z_i \in A\}]}{\mathbb{P}[\{Z_i \in A\}]} \right| \\
&\quad + \sup_{x\in\mathbb{R}} \left| \frac{\mathbb{E}_n[\mathbb{1}\{X_i \le x, Z_i \in A\}]}{\mathbb{P}[\{Z_i \in A\}]} - \mathbb{P}[X_i \le x | Z_i \in A] \right| \\
&= \left| \frac{1}{\mathbb{E}_n[\mathbb{1}\{Z_i \in A\}]} - \frac{1}{\mathbb{P}[\{Z_i \in A\}]} \right| \sup_{x\in\mathbb{R}} |\mathbb{E}_n[\mathbb{1}\{X_i \le x, Z_i \in A\}]| \\
&\quad + \frac{1}{\mathbb{P}[Z_i \in A]} \sup_{x\in\mathbb{R}} |\mathbb{E}_n[\mathbb{1}\{X_i \le x, Z_i \in A\}] - \mathbb{P}[X_i \le x, Z_i \in A]| \\
&\le \left| \frac{1}{\mathbb{E}_n[\mathbb{1}\{Z_i \in A\}]} - \frac{1}{\mathbb{P}[\{Z_i \in A\}]} \right| + \frac{1}{\mathbb{P}[Z_i \in A]} \sup_{x\in\mathbb{R}} |\mathbb{E}_n[\mathbb{1}\{X_i \le x, Z_i \in A\}] - \mathbb{P}[X_i \le x, Z_i \in A]| \\
&\xrightarrow{\text{a.s.}} 0,
\end{aligned}$$

where the first inequality uses the triangle inequality, the second inequality uses the fact that:

$$\sup_{x\in\mathbb{R}} |\mathbb{E}_n[\mathbb{1}\{X_i \le x, Z_i \in A\}]| \le 1,$$

which holds by construction, and the last line uses the Strong Law of Large Numbers and the continuous mapping theorem. ■

I.16 Proof of Proposition H.2

First note that for $\mathbb{P}[X_i \in A, Y_i = 0, T_i = 0 | Z_i = 0] - \mathbb{P}[X_i \in A, Y_i = 0, T_i = 0 | Z_i = 1]$:

$$\mathbb{P}[X_i \in A, Y_i = 0, T_i = 0 | Z_i = 0] - \mathbb{P}[X_i \in A, Y_i = 0, T_i = 0 | Z_i = 1]$$

$$= \mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 0, X_i \in A]\mathbb{P}[X_i \in A | Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 1, X_i \in A]\mathbb{P}[X_i \in A | Z_i = 1]$$
$$= (\mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 0, X_i \in A] - \mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 1, X_i \in A]) \times \mathbb{P}[X_i \in A],$$

where the second equality uses the assumption that $X_i \perp\!\!\!\perp Z_i$.

Thus, to show the numerical equivalence between the two formulas in Theorem 5.1 and Proposition H.1, it suffices to show the equivalence between the numerators in the two formulas:

$$\begin{aligned}
&\mathbb{P}[T_i = 1 | X_i \in A, Z_i = 1] - \mathbb{P}[T_i = 1 | X_i \in A, Z_i = 0] - \mathbb{E}[\kappa_0 Y_i | X_i \in A] \\
&= \mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 0, X_i \in A] - \mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 1, X_i \in A].
\end{aligned}$$

Observe that for $\mathbb{P}[T_i = 1 | X_i \in A, Z_i = 1] - \mathbb{P}[T_i = 1 | X_i \in A, Z_i = 0] - \mathbb{E}[\kappa_0 Y_i | X_i \in A]$:

$$\begin{aligned}
&\mathbb{P}[T_i = 1 | X_i \in A, Z_i = 1] - \mathbb{P}[T_i = 1 | X_i \in A, Z_i = 0] - \mathbb{E}[\kappa_0 Y_i | X_i \in A] \\
&= \mathbb{P}[T_i = 0 | X_i \in A, Z_i = 0] - \mathbb{P}[T_i = 0 | X_i \in A, Z_i = 1] - \mathbb{E}[\kappa_0 Y_i | X_i \in A] \\
&= \mathbb{P}[Y_i = 1, T_i = 0 | X_i \in A, Z_i = 0] + \mathbb{P}[Y_i = 0, T_i = 0 | X_i \in A, Z_i = 0] \\
&\quad - \mathbb{P}[Y_i = 1, T_i = 0 | X_i \in A, Z_i = 1] - \mathbb{P}[Y_i = 0, T_i = 0 | X_i \in A, Z_i = 1] - \mathbb{E}[\kappa_0 Y_i | X_i \in A]
\end{aligned}$$

We now proceed to simplify $\mathbb{E}[\kappa_0 Y_i | X_i \in A]$:

$$\begin{aligned}
&\mathbb{E}[\kappa_0 Y_i | X_i \in A] \\
&= \mathbb{E}[\kappa_0 Y_i | X_i \in A, T_i = 0, Z_i = 0] \times \mathbb{P}[T_i = 0, Z_i = 0 | X_i] \\
&\quad + \mathbb{E}[\kappa_0 Y_i | X_i \in A, T_i = 0, Z_i = 1] \times \mathbb{P}[T_i = 0, Z_i = 1 | X_i] \\
&\quad + \mathbb{E}[\kappa_0 Y_i | X_i \in A, T_i = 1, Z_i = 0] \times \mathbb{P}[T_i = 1, Z_i = 0 | X_i] \\
&\quad + \mathbb{E}[\kappa_0 Y_i | X_i \in A, T_i = 1, Z_i = 1] \times \mathbb{P}[T_i = 1, Z_i = 1 | X_i] \\
&= \mathbb{E}[\kappa_0 Y_i | X_i \in A, T_i = 0, Z_i = 0] \times \mathbb{P}[T_i = 0, Z_i = 0 | X_i] \\
&\quad + \mathbb{E}[\kappa_0 Y_i | X_i \in A, T_i = 0, Z_i = 1] \times \mathbb{P}[T_i = 0, Z_i = 1 | X_i] \\
&= \frac{1}{\mathbb{P}[Z_i = 0]} \times \mathbb{P}[Y_i = 1 | X_i \in A, T_i = 0, Z_i = 0] \times \mathbb{P}[T_i = 0, Z_i = 0 | X_i \in A] \\
&\quad - \frac{1}{\mathbb{P}[Z_i = 1]} \times \mathbb{P}[Y_i = 1 | X_i \in A, T_i = 0, Z_i = 1] \times \mathbb{P}[T_i = 0, Z_i = 1 | X_i \in A] \\
&= \frac{1}{\mathbb{P}[Z_i = 0 | X_i \in A]} \times \mathbb{P}[Y_i = 1 | X_i \in A, T_i = 0, Z_i = 0] \times \mathbb{P}[T_i = 0, Z_i = 0 | X_i \in A] \\
&\quad - \frac{1}{\mathbb{P}[Z_i = 1 | X_i \in A]} \times \mathbb{P}[Y_i = 1 | X_i \in A, T_i = 0, Z_i = 1] \times \mathbb{P}[T_i = 0, Z_i = 1 | X_i \in A] \\
&= \mathbb{P}[Y_i = 1, T_i = 0 | Z_i = 0, X_i \in A] - \mathbb{P}[Y_i = 1, T_i = 0 | Z_i = 1, X_i \in A]
\end{aligned}$$

where the second equality uses the fact that $T_i = 1$ implies $\kappa_0 = 0$, the fourth inequality uses IV independence assumption, the fifth equality uses the Bayes rule.

Now the desired equivalence result follows immediately.