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

[Link to Most Recent Version]

October 29, 2022

Abstract

In the empirical study of persuasion, researchers often use a binary instrument to encourage individuals to consume information and take some action. Under the Imbens-Angrist instrumental variable model assumptions, the binary IV model can recover the proportion of people who take the action under the treatment and control conditions. I show that with 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 persuadable individuals and their statistical characteristics. Specifically, I 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). I 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. I 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 10% voters are persuadable. 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, sensitivity analysis, 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, 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, Azeem Shaikh, Tymon Sloczynski, Joshua Ka Chun Shea, Liyang Sun, Max Tabord-Meehan, Christopher Walters, Linbo Wang, 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. That is, persuasion requires the joint distribution of potential outcomes. This paper shows that under certain assumptions, a binary instrumental variable (IV) model can identify the frequency of individuals who are persuadable, those that are "always persuaded", and those that are "never persuaded", and describe their profiles in terms of pre-treatment covariates.

The results that follow require two sets of assumptions: 1) the Imbens-Angrist IV (the IA IV, hereafter) assumptions (Imbens and Angrist, 1994), 2) the monotone treatment response assumption (Manski, 1997; Jun and Lee, 2018). Under the first set of assumptions, we can point identify the marginal distribution of potential outcomes (Imbens and Rubin, 1997; Abadie, 2002, 2003). With the monotone treatment response assumption we can classify individuals to three outcome types: 1) always-persuaded, those who will take the action of interest regardless of whether receive the information treatment or not; 2) never-persuaded, those who will not take the action of interest regardless of the treatment; 3) persuaded, those who will take the action of interest only if they are exposed to the information treatment.

We first show that in a binary IA IV model with the monotone treatment response assumption, the joint distribution of potential outcomes among compliers is point identified. The intuition is that with the monotone treatment response assumption, the percentage of persuaded individuals among compliers is equal to the local average treatment effect (LATE). 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 by the argument in Imbens and Rubin (1997). This is because, under monotone treatment response, treated individuals are at least as likely to take action as an individual who is untreated. Similarly, under monotone treatment response, the event in which an individual is never-persuaded is equivalent to the event that an individual would not take action with treatment.

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 I 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. This result greatly extends the classic κ weighting result in Abadie (2003) because we now can learn the statistical characteristics of different outcome types among compliers.

The new identification results follow from the monotone treatment response assumption, which may not

apply in all settings. To guide researchers in the applicability of these results, I 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 two sets of identification assumptions boils down to testing whether there exists a nonnegative solution to the implied system of linear equations. We implement the test by using a recent result with a subsampling method (Bai et al., 2022). I 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. Given the two sets of assumptions we make, the estimation can be done by using sample analogs. We also provide a formal justification for the bootstrap validity by using a powerful result in Fang and Santos (2019) when there is no weak identification. To incorporate the weak identification case, we also provide a Anderson-Rubin type test that is robust to weak identification (Staiger and Stock, 1997).

Finally, I provide an application based on Green et al. (2003). Green et al. (2003) conduct a field experiment to use the 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 persuadable. 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.

My analysis is closely related to Abadie (2003), who provides results on identifying the statistical characteristics measured by pre-treatment covariates for compliers. We extend the Abadie's κ result by studying a binary IA IV model with an additional monotone treatment response assumption. With both assumptions, researchers can learn the statistical characteristics measured by the pre-treatment covariates of the outcome types (i.e., always-persuaded, never-persuaded, and persuaded) among compliers.

Moreover, my 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.

My 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 my paper is to profile the persuasion types among compliers. Moreover, I provide a sharp test on the assumptions in the 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, I 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 encouragement design. Finally, I 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 an econometric model of persuasion. We then define the target parameters in Section 3. Section 4 provides the point identification results of the distribution of potential outcomes among compliers. Section 5 provides results identifying the statistical characteristics of persuasion types among compliers. Section 6 provides the estimation and inference results. Section 7 provides discussions on comparing the local persuasion rate with existing estimands, a sharp test on the identification assumptions, and a sensitivity analysis on the monotone treatment response assumption. Section 8 provides an application. The last section concludes.

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 $Z_i = z$.

An econometric model of persuasion is a binary IA IV model with the monotone treatment response assumption. Formally speaking, researchers make the following assumptions in an econometric model of persuasion with the potential outcome and potential treatment notations.

Assumption 2.1. (Potential Outcome and Potential Treatment Model)

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

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: $\mathbb{P}[T_i = 1|Z_i = 1] \neq \mathbb{P}[T_i = 1|Z_i = 0]$,
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 with $Y_i(0)$ and $Y_i(1)$ being binary.

Remark 2.1. 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 and nontrivial function of t 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 and nontrivial function of z 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. ■

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.

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) conduct an experiment in Liberia which randomly assigns a free agricultural training program to Liberian ex-fighters. The treatment is the actual participation in the agricultural training program. The outcome of interest is whether or not the Liberian ex-fighters work 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 also likely to hold: the training program should weakly increase the human capital of the ex-fighters. Hence, the training program should weakly increase ex-fighters' wage return from getting a job in the legal sector, which should increase 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.

[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

$Y_i(0)$	$Y_i(1)$	$T_i(0)$	$T_i(1)$	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

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 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 two sets of new target parameters in this paper. First, 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 second set of target parameters are the statistical characteristics measured by the pre-treatment covariates of the persuasion types (i.e., always-persuaded, never-persuaded, and persuaded) among compliers. Learning the statistical characteristics of the persuasion types among compliers can help researchers

[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.

assess whether the experiment achieves specific goals or to assess the potential policy outcome of the experiment. In the GOTV experiment, the researchers aim to mobilize the underrepresented minority to vote. Hence, researchers can estimate the likelihood of the persuadable and compliers being an underrepresented minority. Furthermore, researchers may also want to assess what types of voters they mobilized. For example, they may want to know the likelihood of the mobilized voters being Democrats. If most of the mobilized voters are Democrats, researchers can judge the policy impact of the mobilization effort.

4 Identification of the Potential Outcome Distributions for Compliers

In this section, we present the results on the identification of the joint distribution of potential outcomes among compliers. We first present the well known point identification results on the marginal distribution of potential outcomes among compliers under the IA IV assumptions. We then can use these marginal distributions, as well as the monotone treatment response assumption, to point identify the joint distribution of potential outcomes among compliers. Finally, I show that the results are extendable to the case of a non-binary instrument but not the the case of non-binary outcomes.

4.1 Identification of the Marginal Distribution of Potential Outcomes for Compliers

As is well known, given the IA IV assumptions, we can point identify the marginal distribution of potential outcomes among compliers. This is a classic result in the LATE literature (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. Given that we are working with a binary IA IV model, we provide slightly more simplified results. The results are presented 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:

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

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

Remark 4.1. The two estimands 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 the following indicator variable: $\mathbb{1}\{Y_i = y, T_i = 1\}$ with $y \in \{0, 1\}$. For the marginal distribution of $Y_i(0)$ among compliers, it is the negative of the Wald estimand with outcome variable defined as the following indicator variable: $\mathbb{1}\{Y_i = y, T_i = 0\}$ with $y \in \{0, 1\}$. ■

4.2 Identification of the Joint Distribution of Potential Outcomes for Compliers

Lemma 4.1 only uses the IA IV assumptions. Remarkably, if we further assume the monotone treatment response, 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. The new results are presented in Lemma 4.2.

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

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

Remark 4.2. 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 of potential outcomes among compliers. However, throughout the text, we maintain the assumption that monotone treatment response holds almost surely for simplicity. ■

4.3 Identification of the Joint Distribution of Potential Outcome with Non-Binary Outcome and Non-Binary Instrument

Given the identification results in Lemma 4.2, it is natural to ask whether or not we can extend the results to the case when the outcome or the instrument is non-binary. The answer to the case with the trinary outcome is negative: we cannot point identify the joint distribution of potential outcomes among compliers with the monotone treatment response assumption. In contrast, the answer to the case with non-binary instrument is positive: we can point identify the joint distribution of potential outcomes among compliers with the monotone treatment response assumption with discrete and continuous instrument.

4.3.1 Case 1: Trinary 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.[4] Table 2 presents the classification.

[4] 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. There is a duality between the notation in Jun and Lee (2018) and conventional potential outcome notation used in Table 2.

Table 2: 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

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 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 2 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 4.1. 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.

Remark 4.3. We can partially identify the joint distribution of potential outcomes among compliers using the ideas 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 4.1. ■

Remark 4.4. 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. ■

4.3.2 Case 2: Discrete Instrument

Another direction of extending Lemma 4.2 is to extend the results to the case in which researchers have a discrete valued instrument. With discrete valued instrument, we can modify Assumption 2.1 to:

Assumption 4.1. (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 $t, z \in \text{supp}(T_i, Z_i)$,
3. Exogenous instrument: $Z_i \perp\!\!\!\perp (Y_i(0), Y_i(1), T_i(0), T_i(1), X_i)$,
4. First stage: $\mathbb{P}[T_i = 1|Z_i = z] \neq \mathbb{P}[T_i = 1|Z_i = z']$, for $z \neq z'$ with $z, z' \in \text{supp}(Z_i)$,
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 $z, z' \in \text{supp}(Z_i)$.

With Assumption 4.1, we can point identify the joint distribution of potential outcomes among each complier group. The intuition of the result is that with Assumption 4.1, the proof proceeds "as-if" we are using a binary IV with support being $\{z, z'\}$. We now formally state the results in Corollary 4.1.

Corollary 4.1. Suppose Assumption 4.1 holds, conditional on z, z' compliers, the joint distribution of potential outcome is point identified, that is, for $z, z' \in \text{supp}(Z_i)$ with $T_i(z) \geq T_i(z')$:

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

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

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

Remark 4.5. The results in Corollary 4.1 can also be extended to the case in which the instrument is continuous. Let $p(Z_i)$ be the propensity score:

$$p(Z_i) \equiv \mathbb{P}[T_i = 1|Z_i = z].$$

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:

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

where the first equality uses the result in Heckman and Vytlacil (2005), the last two lines use the results in Carneiro and Lee (2009). ■

5 Profiling Persuasion Types Among Compliers

So far we have results that allow researchers to determine the size of the persuasion effect among compliers. This section provides a set of results on identifying the statistical characteristics of the locally persuadable (that is, $[Y_i(0) = 0, T_i(1) > T_i(0)]$) and the three persuasion types among compliers in Table 1. Moreover, we also extend the results to always-takers and never-takers.

5.1 Identification: Who Are Locally Persuadable

We first consider the characterization of those, among compliers, who do not take the action of interest without being exposed to the treatment. Formally, we can identify the statistical characteristics of the subpopulation defined by the following event: $[Y_i(0) = 0, T_i(1) > T_i(0)]$. We do not directly observe the subpopulation because it involves potential outcomes and treatments. Nevertheless, we can profile this unobserved subpopulation using the results below.

Theorem 5.1. Suppose 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:

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

Furthermore, the conditional distribution function of X_i is point identified, for $x \in \mathbb{R}$:

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

Finally, let $g : \mathbb{R} \to \mathbb{R}$ be measurable such that $\mathbb{E}[|g(X_i)|] < \infty$, then, $\mathbb{E}[g(X_i) | Y_i(0) = 0, T_i(1) > T_i(0)]$ is point identified.

Remark 5.1. The proof of Theorem 5.1 does not use the monotone treatment response assumption in Assumption 2.1. The weighting results in Abadie (2003) show that any statistical characteristic that can be

defined in terms of moments of X_i is identified. By focusing on the case when the outcome variable is binary in the IA IV model, we extend Abadie's κ results to identify the subpopulation who are not only compliers, but also do not take the action of interest without being exposed to the treatment. ■

Remark 5.2. We can apply the idea in Theorem 5.1 to the case in which Y_i is continuous. Specifically, we can define a new indicator variable, $\tilde{Y}_i = \mathbb{1}\{Y_i \in B\}$, where B is a measurable set. A new potential outcome is defined is: $\tilde{Y}_i(0) = \mathbb{1}\{Y_i(0) \in B\}$. Then, the result in Theorem 5.1 holds for $\tilde{Y}_i$ under the IA IV assumptions we make in Assumption 2.1. A example is $B = \mathbb{1}\{Y_i(0) \leq \tilde{y}\}$. Thus, among compliers and those untreated outcome being less than $\tilde{y}$, researchers can identify the distribution of X_i. ■

Remark 5.3. If we further assume that the distributions $X_i|Y_i = 0, T_i = 0, Z_i = z$ for $z \in \{0,1\}$ have a Radon-Nikodym density with respect to a common dominating, positive $\sigma-$finite measure, the conditional probability density function is also identified:

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

■

Remark 5.4. The results can also be easily extended to the case with discrete valued instrument under Assumption 4.1. Suppose $T_i(z) \geq T_i(z')$ holds almost surely, then:

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

where A is a measurable set. ■

Remark 5.5. 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 $\kappa := 1 - \frac{T_i(1-Z_i)}{\mathbb{P}[Z_i=0]} - \frac{(1-T_i)Z_i}{\mathbb{P}[Z_i=1]}$. Thus, a natural question is whether or not we can point identify $\mathbb{E}[g(Y_i, T_i, X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)]$. The answer is no. To see this:

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

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

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

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 $[Y_i(1) = 1, T_i(1) > T_i(0)]$. The results are presented in Proposition 5.1.

Proposition 5.1. Assume that 1 to 4 in Assumption 2.1 hold, then, the following conditional distributions of X_i are point identified. Let A be a measurable set:

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

Remark 5.6. By the identical argument in Theorem 5.1, the conditional distribution functions of X_i are also identifiable, because $\{(-\infty, x] : x \in \mathbb{R}\}$ is measurable. Furthermore, for measurable g, the expectations of $g(X_i)$ conditional on the three subpopulations are also identifiable assuming that the expectation is well-defined. An implication thus is any statistical moments of the pre-treatment covariates for the three subpopulations defined by $[Y_i(0) = 1, T_i(1) > T_i(0)]$, $[Y_i(1) = 0, T_i(1) > T_i(0)]$, and $[Y_i(1) = 1, T_i(1) > T_i(0)]$ are identifiable. Finally, if we further assume the existence of the Radon-Nikodym density for $X_i|Y_i = y, T_i = t, Z_i = z$ for all $y, t, z \in \{0, 1\}$ with respect to a common dominating, positive $\sigma-$ finite measure, the conditional probability densities are also identified. ■

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 5.2.

Proposition 5.2. 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 also show that the identification results in Theorem 5.1 and Proposition 5.2 are equivalent. We formally state this equivalence result in Proposition 5.3.

Proposition 5.3. 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 5.2 are equivalent.

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 usefulness of this result is that it can assist us in understanding how persuasion works among compliers. The results are summarized in Theorem 5.2.

Theorem 5.2 (Compliance and Persuasion)**.** Suppose Assumption 2.1 holds, then, the distribution of X_i conditional on always-persuadable compliers, never-persuadable compliers, and persuadable compliers are point identified. Let A be a measurable set:

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

Remark 5.7. By the identical argument in Theorem 5.1, the conditional distribution functions of X_i given persuasion types and compliers are also identifiable, because $\{(-\infty, x] : x \in \mathbb{R}\}$ is measurable. 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 thus is any statistical moments of the always-persuaded, never-persuaded, and persuaded among compliers are identifiable. ■

Remark 5.8. This theorem extends the weighting results in Abadie (2003). The theorem says that we can learn the statistical characteristics of the persuasion types defined in terms of the joint distribution of potential outcomes given Assumption 2.1. ■

Remark 5.9. 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 implies that the joint distribution of potential outcomes and treatments is independent of the instrument conditioning on the covariates. In other words, we have:

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

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

■

Remark 5.10. The results in Theorem 5.2 hold with discrete Z_i under Assumption 4.1. Again, the results hold because of Corollary 4.1. With discrete Z_i, the results in Theorem 5.2 become:

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

where $\{z, z'\} \in \text{supp}(Z_i)$ and A is a measurable set. ■

5.3 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 5.4.

Proposition 5.4. 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:

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

Remark 5.11. Proposition 5.4 implies that the conditional distributions are identifiable. This follows because $g(x) = \mathbb{1}\{x \in A\}$, with A being a measurable set, is a bounded measurable map. Furthermore, the conditional distribution function is also identifiable. This observation, again, follows from the fact that $\{(-\infty, x] : x \in \mathbb{R}\}$ is measurable. ■

Remark 5.12. 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. ■

6 Estimation and Inference

So far we have developed a set of identification results that allow researchers to determine the proportion of persuasion types among compliers and to profile them with pre-treatment covariates. This section provides estimation and inference results. We provide results on the estimation and inference for identifying the statistical characteristics measured by pre-treatment covariates for the locally persuadable in Theorem 5.1. The results in this section apply directly to other identification results, since they share similar flavor with the results in Theorem 5.1.

6.1 Estimation

Recall that in Theorem 5.1, we identify the following probability for the locally persuadable:

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

with A being a measurable set. Note this estimand can be equivalently represented as the ratios of two regression coefficient:

$$\mathbb{P}[X_i \in A | Y_i(0) = 0, T_i(1) > T_i(0)] = \frac{\beta_1}{\beta_2}, \tag{1}$$

where β_1 is the regression coefficient of Z_i when regressing $\mathbb{1}\{X_i \in A, Y_i = 0, T_i = 0\}$ on Z_i, β_2 is the coefficient of Z_i when regressing $\mathbb{1}\{Y_i = 0, T_i = 0\}$ on Z_i.

A natural estimator for $\mathbb{P}[X_i \in A | Y_i(0) = 0, T_i(1) > T_i(0)]$ is to use its sample analog:

$$\hat{\mathbb{P}}[X_i \in A | Y_i(0) = 0, T_i(1) > T_i(0)] = \frac{\hat{\beta}_1}{\hat{\beta}_2}, \tag{2}$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ are estimated regression coefficient on Z_i from regressing $\mathbb{1}\{X_i \in A, Y_i = 0, T_i = 0\}$ and $\mathbb{1}\{Y_i = 0, T_i = 0\}$ on Z_i, respectively. It is easy to see that $\hat{\mathbb{P}}[X_i \in A | Y_i(0) = 0, T_i(1) > T_i(0)]$ is a consistent estimator for $\mathbb{P}[X_i \in A | Y_i(0) = 0, T_i(1) > T_i(0)]$ under standard assumptions. We now formally state the result in Proposition 6.1.

Proposition 6.1. Assume that 1 to 4 in Assumption 2.1 hold. Moreover, assume that $\mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0 | Z_i = 1] \neq 0$ and $\mathbb{P}[Z_i = 1] > 0$. Finally, assume that $\{Y_i, T_i, Z_i, X_i\}_{i=1}^n$ is an is an independent and identically distributed sample. Then, we have:

$$\hat{\mathbb{P}}[X_i \in A | Y_i(0) = 0, T_i(1) > T_i(0)] \xrightarrow{\mathbb{P}} \mathbb{P}[X_i \in A | Y_i(0) = 0, T_i(1) > T_i(0)],$$

where $\hat{\mathbb{P}}[X_i \in A | Y_i(0) = 0, T_i(1) > T_i(0)]$ is defined in Equation 2.

Remark 6.1. Proposition 6.1 is useful when X_i has discrete support with $\{x_1, \ldots, x_k\}$. Then, Proposition 6.1 implies that for $x \in \{x_1, \ldots, x_k\}$, $\hat{\mathbb{P}}[X_i = x | Y_i(0) = 0, T_i(1) > T_i(0)]$ is a consistent estimator for $\mathbb{P}[X_i =$

$x_i|Y_i(0) = 0, T_i(1) > T_i(0)]$. Then, by the Weak Law of Large Numbers and continuous mapping theorem, $\mathbb{E}[g(X_i)|Y_i(0) = 0, T_i(1) > T_i(0)]$ can be consistently estimated by its sample analog:

$$\hat{\mathbb{E}}[g(X_i)|Y_i(0) = 0, T_i(1) > T_i(0)] = \sum_{j=1}^{k} g(x_j)\hat{\mathbb{P}}[X_i = x_j|Y_i(0) = 0, T_i(1) > T_i(0)].$$

■

Remark 6.2. Theorem 5.1 shows that we can point identify the conditional distribution function:

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

Thus, by Proposition 6.1, $\hat{\mathbb{P}}[X_i \leq x|Y_i(0) = 0, T_i(1) > T_i(0)]$ 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 the results in Appendix C.14. ■

6.2 Bootstrap Validity Under Strong Identification

After having a consistent estimator for $\mathbb{P}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)]$, we now provide a result on its limiting distribution. Specifically, we provide a justification that the standard bootstrap is valid for $\hat{\mathbb{P}}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)]$ when the instrument is not weak.

The standard bootstrap estimates the limiting distribution of $\sqrt{n}(\hat{\mathbb{P}}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)] - \mathbb{P}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)])$ by the conditional law of

$$\sqrt{n}(\hat{\mathbb{P}}^*[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)] - \hat{\mathbb{P}}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)]) \tag{3}$$

given the data. Here, $\hat{\mathbb{P}}^*[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)]$ refers to the estimates of $\mathbb{P}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)]$ using the bootstrapped sample (Bickel and Freedman, 1981; Vaart and Wellner, 1996).

When we use standard bootstrap, we want the conditional law of 3 provides a "good" approximation to the limiting distribution of $\sqrt{n}(\hat{\mathbb{P}}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)] - \mathbb{P}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)])$, so that we can construct an asymptotically valid confidence interval. Formally speaking, a "good" approximation requires that the conditional law of 3 consistently estimates the limiting distribution of $\sqrt{n}(\hat{\mathbb{P}}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)] - \mathbb{P}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)])$.

The estimator in Equation 2 is a ratio of two regression coefficient. Thus, the problem of characterizing the limiting distribution of the estimator in Equation 2 shares similar flavor with the problem of characterizing the limiting distribution of a Two Stage Least Square (TSLS) estimator. Thus, to make the asymptotic approximation work, we need the denominator β_2 in Equation 1 to be bounded a way from zero. The intu-

ition of the bootstrap validity is that when the instrument is not weak, $\mathbb{P}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)]$ is differentiable at (β_1, β_2). (Fang and Santos, 2019). We now formally state the result in Proposition 6.2

Proposition 6.2. Assume that 1 to 4 in Assumption 2.1 hold. Moreover, assume that $\mathbb{P}[Y_i = 0, T_i = 0|Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 1] \neq 0$ and $\mathbb{P}[Z_i = 1] > 0$. Finally, assume that $\{Y_i, T_i, Z_i, X_i\}_{i=1}^n$ is an is an independent and identically distributed sample. Then, we have:

$$\begin{aligned}&\sqrt{n}(\hat{\mathbb{P}}^*[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)] - \hat{\mathbb{P}}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)])\\ &= \sqrt{n}(\hat{\mathbb{P}}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)] - \mathbb{P}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)]) + o_p(1).\end{aligned}$$

Remark 6.3. Proposition 6.2 is useful when X_i has a discrete support with $\{x_1, \ldots, x_k\}$. In this case, the bootstrap is also valid. Specifically, by applying Theorem 3.1 in Fang and Santos (2019), the limiting distribution of $\sqrt{n}(\hat{\mathbb{E}}[g(X_i)|Y_i(0) = 0, T_i(1) > T_i(0)] - \mathbb{E}[g(X_i) \in A|Y_i(0) = 0, T_i(1) > T_i(0)])$ can be consistently estimated by the conditional law of $\sqrt{n}(\hat{\mathbb{E}}^*[g(X_i)|Y_i(0) = 0, T_i(1) > T_i(0)] - \hat{\mathbb{E}}[g(X_i)|Y_i(0) = 0, T_i(1) > T_i(0)])$. ■

6.3 An Anderson-Rubin Test Under Weak Identification

The result in Proposition 6.2 is valid when $\mathbb{P}[Y_i = 0, T_i = 0|Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 1]$ is bounded away from 0. However, in practice, the identification of $\mathbb{P}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)]$ might be weak because $\mathbb{P}[Y_i = 0, T_i = 0|Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 1]$ might be arbitrarily close to zero. The weak identification problem causes a poor asymptotic approximation in the finite-sample settings. This section proposes an inferential procedure that is robust to weak identification problem.

Let us denote $\mathbb{P}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)]$ by p. Recall from previous section that p is a function of (β_1, β_2):

$$p = \frac{\beta_1}{\beta_2} \Leftrightarrow p\beta_2 = \beta_1.$$

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

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

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.

7 Discussion

In this section, I offer three points of discussion on the identification results in previous sections. First, I compare θ_{local} with classic estimands. Specifically, I first compare the local persuasion rate and the complier causal attribution rate. I then provide the necessary and sufficient conditions under which approximated θ_{DK} equals θ_{local} under one-sided non-compliance, which complements the analysis in Jun and Lee (2018). I also propose a test of the identification assumptions, namely the IA IV and monotone treatment response assumptions. I then provide a simple method that can help researchers assess the sensitivity of the 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:

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

In other words, compared with the local persuasion rate, we further condition on those whose treatment switches off. 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 $T_i = Z_i$, and Z_i is exogenous, 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 provide conditions under which $\tilde{\theta}_{\text{DK}}$ equals to θ_{local} when there is one side non-compliance in the experiment. These conditions have empirical relevance. Because in some experiments on persuasion with encouragement design, there is one-sided non-compliance. For example, there is a non-compliance problem in the treatment group in the GOTV experiment in Green et al. (2003).

The results below state that $\tilde{\theta}_{\text{DK}}$ equals to θ_{local} if and only if the distribution of potential outcomes and potential treatments satisfy certain conditions. Specifically, when there is one-sided non-compliance in the treatment group, $\tilde{\theta}_{\text{DK}}$ equals to θ_{local} if and only if the potential outcome under untreated is independent of the potential treatment when the instrument switches off. Suppose there is one-sided non-compliance in the control group. In that case, the two estimands are equal if and only if the proportion of those who do not vote without being exposed to the treatment among those who comply equals the proportion of those who do not vote were they exposed to the treatment among the never-takers.

The results below 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.

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 $Y_i(0) \perp\!\!\!\perp T_i(1)$.

7.2 A Sharp Test of the Identification Assumptions

Our main identification result in Theorem 5.2 depends on two sets of assumptions, namely the IA IV assumptions and the monotone treatment response assumption. Both assumptions impose restrictions on individuals' choice behaviors. The IV monotonicity assumption assumes that the instrument uniformly shifts individuals' treatment-taking decisions in one direction. The monotone treatment response assumption assumes that the treatment uniformly shifts individuals' outcome decisions in one direction. Both assumptions are subject to the criticism that they impose strong restrictions on choice behaviors. To address such criticism, we develop a sharp test for Assumption 2.1.

The idea of the test proposed here closely relates to Balke and Pearl (1997) and Machado et al. (2019). The 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, where the types are defined in Table 1. In other words, the identification assumptions imply the following linear system of equations:

$$A_{\text{obs}}\mathbf{p} = \mathbf{b}, \tag{4}$$

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 $\mathbb{P}[Y_i = 0, T_i = 0, Z_i = 0, X_i \in A]$.[5] Thus, the observed quantity $\mathbf{b}$ is consistent with Assumption 2.1 if there exists a solution to the system of linear equations in 4. 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.

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:

$$H_0 : P \in \mathbf{P}_0 \text{ versus } H_1 : P \in \mathbf{P} \setminus \mathbf{P}_0 \tag{5}$$

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 5, 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 (Fang et al., 2020; Bai et al., 2022). 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 $\mathbf{b}$.[6] For the subsampling-based test, Bai et al. (2022) defines the following quantity:

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

[5] We provide the definition of A_{obs}, $\mathbf{p}$, and $\mathbf{b}$ in Appendix A.

[6] 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 B.

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 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)

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 3.

Table 3: Summary Statistics in Green et al. (2003)

	Observations	Mean	Std. Dev.	Min	Max
Panel A: Full Sample					
Y_i	18,933	0.296	0.457	0	1
T_i	18,933	0.136	0.342	0	1
Z_i	18,933	0.461	0.498	0	1
Voted in 2000	18,933	0.608	0.488	0	1
Panel B: Bridgeport					
Y_i	1,806	0.118	0.323	0	1
T_i	1,806	0.137	0.344	0	1
Z_i	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 4. Our results reveal two interesting patterns. First, conditional on compliers, most of them are never-persuaded (that is, never-voters in this specific application) in both full and Bridgeport samples. Second, only 7.9% of voters are persuaded (that is, mobilizable in this specific application) conditional on compliers in the full sample, and 13.9% of voters are persuaded (that is, mobilizable in this specific application) conditional on compliers in Bridgeport.

Table 4: Distribution of Potential Outcomes in Green et al. (2003)

	Estimates	95% Bootstrap CI
Panel A: Full Sample		
$\mathbb{P}[Y_i(0) = 1 \mid T_i(1) > T_i(0)]$	0.302	(0.264, 0.348)
$\mathbb{P}[Y_i(1) = 1 \mid T_i(1) > T_i(0)]$	0.381	(0.363, 0.399)
$\mathbb{P}[Y_i(0) = 1, Y_i(1) = 1 \mid T_i(1) > T_i(0)]$	0.302	(0.264, 0.348)
$\mathbb{P}[Y_i(0) = 0, Y_i(1) = 0 \mid T_i(1) > T_i(0)]$	0.619	(0.6, 0.636)
$\mathbb{P}[Y_i(0) = 0, Y_i(1) = 1 \mid T_i(1) > T_i(0)]$	0.079	(0.039, 0.119)
Panel B: Bridgeport		
$\mathbb{P}[Y_i(0) = 1 \mid T_i(1) > T_i(0)]$	0.111	(0.029, 0.199)
$\mathbb{P}[Y_i(1) = 1 \mid T_i(1) > T_i(0)]$	0.25	(0.203, 0.305)
$\mathbb{P}[Y_i(0) = 1, Y_i(1) = 1 \mid T_i(1) > T_i(0)]$	0.111	(0.029, 0.199)
$\mathbb{P}[Y_i(0) = 0, Y_i(1) = 0 \mid T_i(1) > T_i(0)]$	0.75	(0.7, 0.8)
$\mathbb{P}[Y_i(0) = 0, Y_i(1) = 1 \mid T_i(1) > T_i(0)]$	0.139	(0.048, 0.21)

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

We now apply Theorem 5.1 and Theorem 5.2 to this experiment. We construct the 95% confidence interval by using both the bootstrap and Anderson-Rubin test. The results are presented in Table 5.

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

compliers 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. 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 persuadable, 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.[7] The 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 5: Profiling Persuasion Types in Green et al. (2003)

	Estimates	95% Bootstrap CI	95% AR CI
Panel A: Full Sample			
$\mathbb{P}[\text{Voted in 2000} = 1 \mid Y_i(0) = 0, T_i(1) > T_i(0)]$	0.603	(0.552, 0.643)	(0.561, 0.644)
$\mathbb{P}[\text{Voted in 2000} = 1 \mid Y_i(0) = 1, Y_i(1) = 1, T_i(1) > T_i(0)]$	0.954	(0.92, 0.99)	(0.927, 0.979)
$\mathbb{P}[\text{Voted in 2000} = 1 \mid Y_i(0) = 0, Y_i(1) = 0, T_i(1) > T_i(0)]$	0.511	(0.481, 0.532)	(0.489, 0.532)
$\mathbb{P}[\text{Voted in 2000} = 1 \mid Y_i(0) = 0, Y_i(1) = 1, T_i(1) > T_i(0)]$	0.885	(0.688, 1)	(0.692, 0.971)
Panel B: Bridgeport			
$\mathbb{P}[\text{Democrat} = 1 \mid Y_i(0) = 0, T_i(1) > T_i(0)]$	0.515	(0.34, 0.675)	(0.364, 0.659)
$\mathbb{P}[\text{Democrat} = 1 \mid Y_i(0) = 1, Y_i(1) = 1, T_i(1) > T_i(0)]$	0.507	(0, 0.913)	(0, 0.752)
$\mathbb{P}[\text{Democrat} = 1 \mid Y_i(0) = 0, Y_i(1) = 0, T_i(1) > T_i(0)]$	0.538	(0.461, 0.61)	(0.474, 0.593)
$\mathbb{P}[\text{Democrat} = 1 \mid Y_i(0) = 0, Y_i(1) = 1, T_i(1) > T_i(0)]$	0.813	(0.219, 1)	(0.364, 1)

Note: This table provides the results on profiling different persuasion types by 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 of the Assumption 2.1 by using Proposition 7.1. We use the subsampling method in Bai et al. (2022) to conduct the test. For the full sample and the Bridgeport sample, the identification assumptions are not rejected at the 5% level. Furthermore, I provide the sensitivity analysis result on the joint distribution of potential outcomes in Table 6 by varying the degree to which the monotone treatment response assumption is violated among compilers. Interestingly, when the violation becomes larger, the proportion of persuadable among compliers increases.

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

Table 6: Sensitivity for Distribution of Potential Outcomes in Green et al. (2003)

Panel A: Full Sample						
Sensitivity Parameter						
$\mathbb{P}[Y_i(1)=0, Y_i(0)=1 \mid T_i(1) > T_i(0)]$	0.1	0.12	0.14	0.16	0.18	0.2
Identified Parameters						
$\mathbb{P}[Y_i(1)=1, Y_i(0)=1 \mid T_i(1) > T_i(0)]$	0.202	0.182	0.162	0.142	0.122	0.102
$\mathbb{P}[Y_i(1)=0, Y_i(0)=0 \mid T_i(1) > T_i(0)]$	0.519	0.499	0.479	0.459	0.439	0.419
$\mathbb{P}[Y_i(1)=1, Y_i(0)=0 \mid T_i(1) > T_i(0)]$	0.179	0.199	0.219	0.239	0.259	0.279
Panel B: Bridgeport						
Sensitivity Parameter						
$\mathbb{P}[Y_i(1)=0, Y_i(0)=1 \mid T_i(1) > T_i(0)]$	0.05	0.06	0.07	0.08	0.09	0.1
Identified Parameters						
$\mathbb{P}[Y_i(1)=1, Y_i(0)=1 \mid T_i(1) > T_i(0)]$	0.061	0.051	0.041	0.031	0.021	0.011
$\mathbb{P}[Y_i(1)=0, Y_i(0)=0 \mid T_i(1) > T_i(0)]$	0.7	0.69	0.68	0.67	0.66	0.65
$\mathbb{P}[Y_i(1)=1, Y_i(0)=0 \mid T_i(1) > T_i(0)]$	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, I 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 the persuadable (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, I 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 my 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, I provide two sets of solutions. First, I provide a sharp test on the two sets of identification assumptions (that is, the IA IV assumptions and the monotone treatment response assumption). The test boils down to testing whether there exists a nonnegative solution to a possibly under-determined system of linear equations with known coefficients. I 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 persuadable. 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 by considering two instrument levels: $\{z, z'\}$. 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.

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.

BICKEL, P. J. AND D. A. FREEDMAN (1981): "Some asymptotic theory for the bootstrap," *The annals of statistics*, 9, 1196–1217.

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. AND A. SANTOS (2019): "Inference on directionally differentiable functions," *The Review of Economic Studies*, 86, 377–412.

FANG, Z., A. SANTOS, A. M. SHAIKH, AND A. TORGOVITSKY (2020): "Inference for large-scale linear systems with known coefficients," *arXiv preprint arXiv:2009.08568*.

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.

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.

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.

VAART, A. W. AND J. A. WELLNER (1996): "Weak convergence," in *Weak convergence and empirical processes*, Springer, 16–28.

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

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:

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

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

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

Appendix B 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:

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

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:

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

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).

Appendix C Proof of the Main Results

C.1 Proof of Lemma 4.1

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 the first equality uses IV monotonicity in Assumption 2.1, 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.

C.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.

C.3 Proof of Proposition 4.1

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.

C.4 Proof of Theorem 5.1

For $\mathbb{P}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)]$, where A being a measurable set:

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

where the first equality uses the Bayes' Theorem, the second equality uses Corollary 4.1, the third equality follows by using the identical argument in Corollary 4.1 under Assumption 2.1.

The identification of the conditional distribution of X_i follows immediately by observing that $\{(-\infty, x] : x \in \mathbb{R}\}$ is measurable.

The identification of the conditional expectation of $g(X_i)$ follows immediately by observing that:

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

C.5 Proof of Proposition 5.1

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

C.6 Proof of Proposition 5.2

First, note that the IV independence assumption in Assumption 2.1 implies $(Y_i(1), Y_i(0), T_i(1), T_i(0)) \perp\!\!\!\perp Z_i|X_i$. For $\mathbb{P}[X_i \in A, Y_i(0) = 0, T_i(1) > T_i(0)]$:

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

For $\mathbb{P}[T_i(1) > T_i(0)|X_i \in A]$:

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

which follows from Lemma 3.1 in Abadie (2003).

For $\mathbb{P}[Y_i(0) = 0|X_i \in A, T_i(1) > T_i(0)]$:

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

where the third equality follows from Theorem 3.1 part (b) in Abadie (2003) by defining $g(Y_i(0)) = Y_i(0)$.

For $\mathbb{P}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)]$:

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

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

where the last equality uses Lemma 4.1.

C.7 Proof of Proposition 5.3

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

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

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 5.2, 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 $\mathbb{E}[\kappa_0 Y_i | X_i \in A]$:

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

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

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.

C.8 Proof of Theorem 5.2

For $\mathbb{P}[X_i \in A | Y_i(1) = Y_i(0) = 1, T_i(1) > T_i(0)]$. 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 $\mathbb{P}[X_i \in A | Y_i(1) = Y_i(0) = 1, T_i(1) > T_i(0)]$ follows immediately.

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

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

where the second equality uses Theorem 1 in Imbens and Angrist (1994). For $\mathbb{P}[X_i \in A, Y_i(1) = 1, Y_i(0) = 0, T_i(1) > T_i(0)]$:

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

where the third equality uses the assumption 2 in Assumption 2.1.

C.9 Proof of Proposition 5.4

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 exogeneity assumption in Assumption 2.1.

C.10 Proof of Proposition 6.1

Let $\mathbb{E}_n$ denote sample average. Then, for $\hat{\beta}_1$, we have::

$$\begin{aligned}\hat{\beta}_1 &= \frac{\mathbb{E}_n[Z_i Y_i] - \mathbb{E}_n[Z_i]\mathbb{E}_n[Y_i]}{\mathbb{E}_n[Z_i](1 - \mathbb{E}_n[Z_i])} \\ &\xrightarrow{\mathbb{P}} \frac{\mathbb{E}[Z_i Y_i] - \mathbb{E}[Z_i]\mathbb{E}[Y_i]}{\mathbb{E}[Z_i](1 - \mathbb{E}[Z_i])} \\ &= \beta_1,\end{aligned}$$

where the second line uses the Weak Law of Large Numbers and the continuous mapping theorem. provided that $\mathbb{P}[Z_i = 1] > 0$. Similarly, we can show that $\hat{\beta}_2 \xrightarrow{\mathbb{P}} \beta_2$. Now the desired result follows immediately from the continuous mapping theorem provided β_2 is bounded away from 0.

C.11 Proof of Proposition 6.2

Note that our estimator for $\mathbb{P}[X_i \in A|Y_i(0) = 0, T_i(1) > T_i(0)]$ is a function of (θ_1, θ_2), where θ_1 and θ_2 are the regression coefficients of Z_i from regressing $\mathbb{1}\{X_i \in A, Y_i = 0, T_i = 0\}$ and $\mathbb{1}\{Y_i = 0, T_i = 0\}$ on Z_i, respectively. By the standard application of the delta method, we know that:

$$\sqrt{n}\left(\begin{pmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{pmatrix} - \begin{pmatrix}\theta_1 \\ \theta_2\end{pmatrix}\right) \xrightarrow{\mathcal{D}} \mathbb{G}_0,$$

where $\hat{\theta}_1$ and $\hat{\theta}_2$ are sample analogs of θ_1 and θ_2, and $\mathbb{G}_0$ is a normal distribution. Furthermore, we assume that β_2 is bounded away from 0, thus, $\phi(\theta_1, \theta_2) = \frac{\theta_1}{\theta_2}$ is differentiable at (β_1, β_2). Now the desired result follows immediately from applying Theorem 3.1 in Fang and Santos (2019).

C.12 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).

C.13 Proof of Theorem 7.1

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

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

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

$$\begin{aligned}&(\mathbb{P}[T_i = 1|Z_i = 1] - \mathbb{P}[T_i = 1|Z_i = 0]) \times \mathbb{P}[Y_i = 0|Z_i = 0] \\ &= \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 0] - \mathbb{P}[Y_i = 0, T_i = 0|Z_i = 1]. \end{aligned} \tag{6}$$

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 6, $\tilde{\theta}_{\text{DK}} = \theta_{\text{local}}$ 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 6, $\tilde{\theta}_{\text{DK}} = \theta_{\text{local}}$ 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 $\mathbb{P}[T_i(0) = 0] = 1$.

C.14 A Glivenko-Cantelli Theorem for Conditional Cumulative Distribution Function

In fact, we can strengthen the statement in Remark 6.2 from convergence in probability to almost sure convergence:

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

Moreover, the uniform convergence result in Remark 6.2 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 C.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:

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

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

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

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

for $1 \leq j \leq k-1$. 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 $1 \leq j \leq k$ and $n \geq N_k$:

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

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

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

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

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

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

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