THE UNIVERSITY OF CHICAGO

ESSAYS ON CAUSAL INFERENCE AND ITS APPLICATIONS IN EMPIRICAL POLITICAL ECONOMY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF POLITICAL SCIENCE

BY
ZEYANG YU

CHICAGO, ILLINOIS
DECEMBER 2023

To my mom, who has always been supportive throughout this intellectual journey.

TABLE OF CONTENTS

LIST OF FIGURES viii

LIST OF TABLES ix

ACKNOWLEDGMENTS xi

ABSTRACT xiii

1 A BINARY IV MODEL FOR PERSUASION: PROFILING PERSUASION TYPES AMONG COMPLIERS 1

1.1 Introduction 1

1.2 Model Setup 5

1.3 Target Parameters 8

1.4 Identification of the Potential Outcome Distributions for Compliers 10

1.4.1 Identification of the Joint Distribution of Potential Outcomes for Compliers in a Binary IV Model 11

1.5 Profiling Persuasion Types 13

1.5.1 Profiling the Locally Persuadable 14

1.5.2 Identification: Compliance and Persuasion 17

1.6 Estimation and Inference 19

1.7 Discussion 20

1.7.1 Comparison with Existing Estimands 21

1.7.2 A Sharp Test of the Identification Assumptions 23

1.7.3 Sensitivity Analysis: the Monotone Treatment Response Assumption 25

1.8 Empirical Application: Revisit Green et al. [2003] 26

1.8.1 Empirical Setup 26

1.8.2 Empirical Results 26

1.8.3 Testing Identification Assumptions and Sensitivity Analysis 29

1.9 Conclusion 30

2 PROFILING PERSUASION TYPES IN A BINARY IA-IV MODEL: THREE EMPIRICAL APPLICATIONS . . . 33
2.1 Introduction . . . 33
2.2 Revisit Enikolopov et al. [2011] . . . 34
2.2.1 Empirical Setup . . . 34
2.2.2 Empirical Results . . . 36
2.3 Revisit Blattman and Annan [2016] . . . 39
2.3.1 Empirical Setup . . . 39
2.3.2 Empirical Results . . . 42
2.4 Revisit Chen and Yang [2019] . . . 45
2.4.1 Empirical Setup . . . 46
2.4.2 Empirical Results . . . 48
2.5 Conclusion . . . 51
3 WHEN IS TRIPLE DIFFERENCE SENSITIVE TO FUNCTIONAL FORM? . . . 61
3.1 Introduction . . . 61
3.2 Setup . . . 63
3.3 Invariance of Parallel Trends . . . 65
3.4 A Simulation for the Decomposition Results . . . 69
3.5 Triple Difference as Difference in Differences . . . 70
3.6 Comparison with Distributional DID . . . 72
3.7 Testable Implications . . . 74
3.8 Empirical Illustration . . . 76
3.9 Concluding Remarks . . . 77
APPENDIX A APPENDIX FOR CHAPTER 1 . . . 79
A.1 Identifiability of the Joint Distribution of Non-Binary Instruments or Outcomes 79
A.1.1 Non-Binary Instrument . . . 79
A.1.2 Non-Binary Outcome . . . 82
A.2 Profiling Compliers with a Non-Binary Instrument . . . 85

A.3 A Different Quantity of "Profiling" . . . 86
A.4 Identification: Always-Takers and Never-Takers . . . 87
A.5 More on Estimation and Inference . . . 88
A.5.1 Estimation and Inference under Strong Identification . . . 88
A.5.2 An Anderson-Rubin Test under Weak Identification . . . 90
A.6 A System of Equation for the Binary IV Model with Monotone Treatment Response . . . 92
A.7 Implementing the Test in Section 1.7.2 . . . 94
A.8 An Equivalence Result . . . 95
A.9 Proofs . . . 95
A.9.1 Proof of Lemma 1.4.1 . . . 95
A.9.2 Proof of Lemma 1.4.2 . . . 96
A.9.3 Proof of Theorem 1.5.1 . . . 97
A.9.4 Proof of Proposition 1.5.1 . . . 97
A.9.5 Proof of Theorem 1.5.2 . . . 98
A.9.6 Proof of Proposition 1.5.2 . . . 99
A.9.7 Proof of Proposition A.4.1 . . . 100
A.9.8 Proof of Claim 1.7.1 . . . 100
A.9.9 Proof of Theorem 1.7.1 . . . 101
A.9.10 Proof of Proposition A.1.1 . . . 103
A.9.11 Proof of Corollary A.1.1 . . . 104
A.9.12 Proof of Corollary A.1.2 . . . 105
A.9.13 Proof of Corollary A.2.1 . . . 107
A.9.14 Proof of Corollary A.2.2 . . . 107
A.9.15 A Glivenko-Cantelli Theorem for Conditional Cumulative Distribution Function . . . 108
A.9.16 Proof of Proposition A.8.2 . . . 111

APPENDIX B APPENDIX FOR CHAPTER 3 113
B.1 Proof of Proposition 3.3.1 113
B.2 Proof of Proposition 3.3.2 113
B.3 Proof of Proposition 3.7.1 114
B.4 Implementing the Least Favorable Method in Canay and Shaikh [2017] . . . 116
B.5 Implementing the Two-Step Method in Romano et al. [2014] and Bai et al. [2022a] 117

REFERENCES 123

LIST OF FIGURES

1.1 Test Identification Assumptions using Bai et al. [2022a] 30

2.1 Enikolopov et al. [2011]: Testing Identification Assumptions 39
2.2 Blattman and Annan [2016]: Testing Identification Assumptions 46
2.3 Chen and Yang [2019]: Testing Identification Assumptions 51

3.1 Illustration: DDD Invariance of Mixtures of Distributions 69
3.2 Implied Densities and p-Values using Least Favorable Test 78

LIST OF TABLES

1.1 Types of Individuals 8

1.2 Summary Statistics in Green et al. [2003] 27

1.3 Distribution of Potential Outcomes in Green et al. [2003] 28

1.4 Profiling Persuasion Types in Green et al. [2003] 29

1.5 Sensitivity for Distribution of Potential Outcomes in Green et al. [2003] 31

2.1 Enikolopov et al. [2011]: Summary Statistics 35

2.2 Enikolopov et al. [2011]: Joint Distribution of Potential Outcomes for Voting for OVR 36

2.3 Enikolopov et al. [2011]: Profiling Demographic Variables among Compliers for Voting for OVR 38

2.4 Enikolopov et al. [2011]: Profiling Political Variables among Compliers for Voting for OVR 40

2.5 Enikolopov et al. [2011]: Comparison Between θ_{local} and θ_{DK} 41

2.6 Blattman and Annan [2016]: Summary Statistics 41

2.7 Blattman and Annan [2016]: Joint Distribution of Potential Outcomes for Agricultural Work 52

2.8 Blattman and Annan [2016]: Joint Distribution of Potential Outcomes for Legal Activities 53

2.9 Blattman and Annan [2016]: Profiling Demographic Variables among Compliers for Agricultural Work 54

2.10 Blattman and Annan [2016]: Profiling Demographic Variables among Compliers for Legal Activities 54

2.11 Blattman and Annan [2016]: Profiling Variables that Measure the Level of Violence among Compliers for Agricultural Work 55

2.12 Blattman and Annan [2016]: Profiling Variables that Measure the Level of Violence among Compliers for Legal Activities 56

2.13 Blattman and Annan [2016]: Comparison Between θ_{local} and θ_{DK} for Agricultural Work 56

2.14 Blattman and Annan [2016]: Comparison Between θ_{local} and θ_{DK} for Legal Activities 57

2.15 Chen and Yang [2019]: Summary Statistics 57

2.16 Chen and Yang [2019]: Joint Distribution of Potential Outcomes for Trust in Foreign Media 58
2.17 Chen and Yang [2019]: Profiling Demographic Variables among Compliers for Trust in Foreign Media 59
2.18 Chen and Yang [2019]: Profiling Preference Variables among Compliers for Trust in Foreign Media 60
2.19 Chen and Yang [2019]: Comparison Between θ_{local} and θ_{DK} 60

A.1 Types of Individuals with Trinary Outcome 83

ACKNOWLEDGMENTS

I am deeply thankful for the unwavering guidance and support from my advisors, Scott Gehlbach, Bobby Gulotty, and Alex Torgovitsky. Scott has consistently uplifted my spirits and provided invaluable insights into the substantive interpretation of the findings from the methodologies developed in this dissertation. Bobby not only offered countless valuable feedback on the methodologies but also provided constant emotional support during the challenging and disheartening moments I faced over the past six years. I owe a great deal to Alex for most things I know about microeconometrics. I am especially thankful for the two invaluable lessons he imparted to me. Firstly, he taught me not to shy away from tackling difficult theories in econometrics and statistics; instead, he encouraged me to confront and overcome these difficulties with courage. Secondly, he showed me how to recognize the intellectual merits in econometrics research and craft new research based on these merits.

I would also like to thank many other professors at UChicago for their support. In particular, I would like to thank Oeindrila Dube, for the QJE paper by David Yanagizawa-Drott she asked to read in her conflict course that inspired the first two chapters. I would also like to thank the econometrics group at the economics department, Stephane Bonhomme, Azeem Shaikh, Max Tabord-Meehan, and Kirill Ponomarev, for the econometrics that they taught me, and their incessant feedback provided in our interactions in the past few years. I would also like to thank professors from the political economy group at Harris, Chris Blattman, Anthony Fowler, and James Robinson, for their comments on the substantive findings from applying the methods developed in the first chapter to interesting political economy research. Finally, I would also like to thank professors from the political science department, who have provided both intellectual and emotional support in the past six years: Mark Hansen, William Howell, Monika Nalepa, Molly Offer-Westort, Eric Oliver, Jon Rogowski, Anton Strezhnev, and Dali Yang.

During my time in graduate school, I was incredibly fortunate to have received immense support from my close friends. They served as my emotional refuge during moments of despair and hopelessness throughout this academic journey. I am deeply grateful to Josh Byun, Lautaro Cella, Yi Gu, Hongchang Guo, Lingnan He, Minju Kim, Jiarui Qian, Francesco Ruggieri, and Zhen Xie for their unwavering support.

Finally, I would like to thank my mom for her faith in me through my ups and downs and the intellectual curiosity she fostered.

ABSTRACT

This dissertation studies causal inference and its applications in empirical political economy. Chapter 1 studies a binary Imbens-Angrist instrumental variable model for persuasion. In the empirical study of persuasion, researchers often use a binary instrument to encourage individuals to consume information and take some action. We show that with the Imbens-Angrist instrumental variable model assumptions and the monotone treatment response assumption, it is possible to identify the joint distributions of potential outcomes among compliers. This is necessary to identify the percentage of persuaded individuals and their statistical characteristics. Specifically, we develop a weighting method that helps researchers identify the statistical characteristics of persuasion types: compliers and always-persuaded, compliers and persuaded, and compliers and never-persuaded. These findings extend the "κ weighting" results in Abadie [2003]. We also provide a sharp test on the two sets of identification assumptions. The test boils down to testing whether there exists a nonnegative solution to a possibly under-determined system of linear equations with known coefficients. An application based on Green et al. [2003] is provided. The result shows that among compliers, roughly 10% voters are persuaded. The results are consistent with the findings that voters' voting behaviors are highly persistent.

Chapter 2 applies the methods developed in the first chapter to three empirical examples [Enikolopov et al., 2011, Blattman and Annan, 2016, Chen and Yang, 2019]. The results illustrate the usefulness of the methods. Re-analyzing Enikolopov et al. [2011] informed us that most of the voters were persuaded, and the persuaded voters were likely to be middle-aged and male. Re-analyzing Blattman and Annan [2016] informed us that around 20% of the Liberian ex-fighters were persuaded, and the persuaded ex-fighters were more likely to be risky type. Re-analyzing Chen and Yang [2019] informed us that roughly 20% of the students were persuaded, and the persuaded students were likely to come from wealthy families, come from coastal areas, less risk-loving, and less likely to believe in the inherent

goodness of people.

Chapter 3, coauthored with Hongchang Guo, studies when the validity of triple difference depends on functional form. Here, the functional form refers to the transformations on the outcome variables (e.g., taking the logarithm of the outcome variable). Build on Roth and Sant'Anna [2023], we provide a novel characterization: the "modified" parallel trends assumption in the triple difference design holds under all measurable transformations of the outcome if and only if a stronger "modified" parallel trends-type condition holds for the cumulative distribution function of untreated potential outcomes. Another equivalent condition for "modified" parallel trends to be insensitive to functional form is that the population can be partitioned into subgroups for which the treatment is effectively not (as-if) randomly assigned and a remaining part that is stable over time, which contrasts sharply to the decomposition results in Roth and Sant'Anna [2023]. These conditions have testable implications on the distribution of the unobservable but identifiable untreated potential outcomes for the treated group in the treated period. Testing these implications boils down to testing a family of moment inequalities. We revisit Muralidharan and Prakash [2017] to illustrate the methodology we propose.

CHAPTER 1

A BINARY IV MODEL FOR PERSUASION: PROFILING PERSUASION TYPES AMONG COMPLIERS

1.1 Introduction

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

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

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

Given the ability to identify persuasion types, we can also profile them by using pre-treatment covariates. We begin by extending the κ weighting result in Abadie [2003] to the local persuasion rate developed by Jun and Lee [2023]. Specifically, we show that with the IA IV assumption, we can identify the statistical characteristics measured by pre-treatment covariates of the locally persuadable, by which we mean those who are compliers and who will not take the action of interest without being exposed to the treatment.

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

The new identification results follow from the monotone treatment response assumption, which may not be applicable in situations where researchers are uncertain about the direction of the treatment effect. To guide researchers in the applicability of these results, we provide a sharp test on the two sets of identification assumptions and a sensitivity analysis. The sharp test closely relates to the result in Balke and Pearl [1997]. The test exploits the fact that a binary IA IV model with monotone treatment response assumption implies an under-determined system of linear equations with known coefficients. Thus, testing the validity of the identification assumptions boils down to testing whether there exists a non-negative solution to the implied system of linear equations. We implement the test using the subsampling method [Bai et al., 2022a]. We also provide a sensitivity result based on the idea in Balke and Pearl [1997]. Specifically, since in the binary IV model, the observed quantity is a linear system equation of the unobserved outcome and compliance types, we can vary the size of the violation of the monotone treatment response assumption among compliers to see how our point identification results change.

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

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

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

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

This paper also closely relates to Jun and Lee [2023]. Jun and Lee [2023] provides a set of point/partial identification results for the persuasion rate and the local persuasion rate under different data scenarios. One main focus of this paper is to profile the persuasion types among compliers. Moreover, this paper provides a sharp test on the assumptions in a binary IV model for persuasion. The sharp test itself also speaks to a large literature on testing IA IV model validity [Balke and Pearl, 1997, Heckman and Vytlacil, 2005, Kitagawa, 2015, Huber and Mellace, 2015, Wang et al., 2017, Mourifié and Wan, 2017, Machado et al., 2019, Kédagni and Mourifié, 2020]. The sharp test follows the tradition of the literature by using the simple fact that the observed quantity in the data is a linear combination of the probability of the unobserved outcome and compliance types. Furthermore, we also provide a necessary and sufficient condition under which the "approximated" persuasion rate proposed by DellaVigna and Kaplan [2007] equals the local persuasion rate proposed by Jun and Lee [2023] when there is one-sided non-compliance in the experiment design. Finally, we also provide a simple sensitivity analysis approach to assess the robustness of the results for the violation of the monotone treatment response assumption.

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

1.2 Model Setup

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

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

Assumption 1.2.1. (A Binary IV Model of Persuasion)

1. Exclusion restriction: $Y_i(t, z) = Y_i(t)$, for $t, z \in \{0, 1\}$,
2. Exogenous instrument: $Z_i \perp\!\!\!\perp (Y_i(0), Y_i(1), T_i(0), T_i(1), X_i)$,
3. First stage: $\mathbb{P}[T_i = 1|Z_i = 1] \neq \mathbb{P}[T_i = 1|Z_i = 0]$,
4. IV Monotonicity: $T_i(1) \geq T_i(0)$ holds almost surely,
5. Monotone treatment response: $Y_i(1) \geq Y_i(0)$ holds almost surely, and $Y_i(0), Y_i(1) \in \{0, 1\}$.

Assumptions 1 to 4 are the assumptions in the IA IV model. In what follows, we use the IA IV assumptions and the LATE assumptions interchangeably to refer to Assumptions 1

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

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

As pointed out by Machado et al. [2019], the results in Vytlacil [2002] imply that Assumption 1.2.1 is equivalent to the following triangular system model:

1. $Y_i(t) = \mathbb{1}\{U_i \leq \gamma(t)\}$, where $\gamma : \mathcal{T} \to \mathbb{R}$ is a measurable function with $\gamma(0) < \gamma(1)$,
2. $T_i(z) = \mathbb{1}\{V_i \leq \nu(z)\}$, where $\nu : \mathcal{Z} \to \mathbb{R}$ is a measurable function with $\nu(0) < \nu(1)$,
3. $Z_i \perp\!\!\!\perp (V_i, U_i, X_i)$,

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

Assumption 1.2.1 can be applied in cases other than persuasion.[2] For instance, researchers are interested in studying the effect of participating in a job training program on the decision to join a rebellion group in a fragile state [Blattman and Annan, 2016, Blattman et al., 2017, 2020]. Blattman and Annan [2016] conducted an experiment in Liberia that randomly assigned Liberian ex-fighters to a free agricultural training program. The treatment is the actual participation in the agricultural training program. The outcome of interest is whether or not the Liberian ex-fighters are employed in the legal sector. Here, the IV

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.

monotonicity condition is likely to hold because the program should decrease the cost of the training program for all of the ex-fighters. The monotone treatment response assumption is likely to hold as the training program is expected to increase the human capital of ex-fighters, thereby increasing their wage return from getting a job in the legal sector and raising their opportunity cost of getting a job in the illegal sector.

By Assumption 1.2.1, 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 1.1.

Table 1.1: Types of Individuals

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

1.3 Target Parameters

In the empirical study of persuasion, researchers are interested in the "effect" of the information treatment on individuals' behaviors. One target parameter proposed by Jun and Lee [2023] is the local persuasion rate:

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

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 1.2.1, Jun and Lee [2023] have shown that θ_{local} is point identifiable.

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

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

Second, we are interested in the statistical characteristics measured by pre-treatment covariates for the locally persuadable. Here, the locally persuadable is the subpopulation that θ_{local} conditions on: $[Y_i(0) = 0, T_i(1) > T_i(0)]$. Learning the statistical characteristics

3. As summarized in DellaVigna and Gentzkow [2010], another popular target parameter in the empirics of persuasion is the persuasion rate: $\theta \coloneqq \mathbb{P}[Y_i(1) = 1 | Y_i(0) = 0]$. 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 [2023], θ_{DK} is not a well defined conditional probability. Hence, it does not measure the persuasion rate for any subpopulation. Moreover, Jun and Lee [2023] show that under Assumption 1.2.1, θ is not point identifiable. They instead provide sharp bounds for θ.

of the locally persuadable can help researchers assess the strength of the study's external validity. If the statistical characteristics of the locally persuadable are not similar to the general population, researchers need to be cautious about generalizing their conclusion to the general population.

The third set of target parameters refers to the statistical characteristics of the persuasion types among compliers (i.e., always-persuaded, never-persuaded, and persuaded). Understanding these characteristics can help researchers assess the experiment's success in achieving specific goals and its potential policy outcomes. For instance, in the GOTV experiment, researchers aimed to mobilize underrepresented minorities to vote, so estimating the likelihood of persuaded and compliers being part of this group is crucial. Additionally, researchers may want to determine the types of voters mobilized, such as their likelihood of being Democrats. This information can help researchers evaluate the policy impact of the mobilization effort.

1.4 Identification of the Potential Outcome Distributions for Compliers

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

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

As is well known, given the IA IV assumptions, we can point identify the marginal distribution of potential outcomes among compliers [Imbens and Rubin, 1997, Abadie, 2003, Jun and Lee, 2023]. In other words, we can know the percentage of voters who will vote if they receive the GOTV treatment and the percentage of voters who will vote if they do not receive the GOTV treatment among compliers. For the sake of completeness, we restate this classic result in Lemma 1.4.1.

Lemma 1.4.1. Assume that the 1 to 4 in Assumption 1.2.1 hold, then, with binary Y_i, the marginal distribution of potential outcomes conditional on compliers is point identified:

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

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

The intuition of the identification results in Lemma 1.4.1 is the following. To make the discussion more concrete, let us consider the untreated potential outcome in the GOTV experiment. Among the voters who are not randomly assigned to the GOTV treatment (i.e., those with $Z_i = 0$), for those who do not receive the GOTV experiment (i.e., those with $T_i = 0$), we know that: (1) we observe their untreated potential outcome, $Y_i(0)$; (2) by the IV monotonicity in Assumption 1.2.1, they are either compliers or never-takers. Among the voters who are randomly assigned to the GOTV treatment (i.e., those with $Z_i = 1$), for those who do not receive the GOTV experiment (i.e., those with $T_i = 0$), we know that: (1) we observe their untreated potential outcome; (2) by the IV monotonicity assumption, they are never-takers. Subtracting the two groups then gives us compliers. Similarly, for the treated potential outcome, subtracting a mixture of always-takers and compliers from always-takers gives us compliers.

The two estimands in Lemma 1.4.1 are similar to the Wald estimand in the IA IV model. Consider the marginal distribution of $Y_i(1)$ among compliers, the estimand is equivalent to a Wald estimand with treatment variable being T_i, instrument being Z_i, and the outcome variable being $\mathbb{1}\{Y_i = y, T_i = 1\}$ with $y \in \{0, 1\}$. For the marginal distribution of $Y_i(0)$ among compliers, it is the negative of the Wald estimand with the outcome variable being the following indicator variable: $\mathbb{1}\{Y_i = y, T_i = 0\}$ with $y \in \{0, 1\}$.

The identification results in Lemma 1.4.1 only use the IA IV assumptions. Remarkably, if we further assume the treatment response is monotone, we can point identify the joint distribution of potential outcomes among compliers. Thus, this lemma strengthens the classic results in the LATE literature that identifies the quantities of the marginal distribution of the potential outcome of compliers [Imbens and Angrist, 1994, Angrist et al., 1996]. Lemma 1.4.1, In other words, under Assumption 1.2.1, we can know the percentage of always-persuaded, never-persuaded, and persuaded among compliers.

Lemma 1.4.2. Suppose Assumption 1.2.1 holds, the joint distribution of potential outcomes among compliers is point identified:

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

Here is the intuition behind the identification results in Lemma 1.4.2. By the monotone treatment response in Assumption 1.2.1, we know the following three things: (1) for those who will vote without receiving the GOTV treatment (i.e., those with $Y_i(0)$ being 1), they will also vote with receiving the GOTV treatment (i.e., their $Y_i(1)$ is also 1); (2) for those who will not vote with receiving the GOTV treatment (i.e., those with $Y_i(1)$ being 0), they will also not vote without receiving the GOTV treatment (i.e., their $Y_i(0)$ is also 0); (3) $Y_i(1) - Y_i(0) = 1$ if and only if $Y_i(1) = 1, Y_i(0) = 0$, thus, LATE becomes the proportion of mobilizable voters among compliers.[4]

Note that we only need the monotone treatment response assumption to hold among compliers for Lemma 1.4.2, because we are "solving" the joint distribution of potential outcomes among compliers from the marginal distribution. However, throughout the text, we maintain the assumption that the monotone treatment response holds almost surely for simplicity.

1.5 Profiling Persuasion Types

This section offers results that profile the persuasion types among compliers, in addition to determining the size of the persuasion effect. We present a series of results that help identify the statistical characteristics of the locally persuadable (that is, $[Y_i(0) = 0, T_i(1) > T_i(0)]$) as well as the three other persuasion types defined by the marginal potential outcomes. Next, we provide results that identify the statistical characteristics of the three persuasion types among compliers as defined in Table 1.1.[5]

4. We discuss the extension of the identification results in Lemma 1.4.2 to non-binary outcomes and instruments in Appendix A.1. The results are negative for the former and positive for the latter.

5. We also extend some of our findings to always-takers and never-takers, see Appendix A.4.

1.5.1 *Profiling the Locally Persuadable*

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

Theorem 1.5.1. Suppose that 1 to 4 in Assumption 1.2.1 hold. Let $g : \mathbb{R} \to \mathbb{R}$ be measurable such that $\mathbb{E}[|g(X_i)|] < \infty$, then, $\mathbb{E}[g(X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)]$ is point identified:

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

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

Theorem 3.1 in Abadie [2003] shows that any statistical characteristic that can be defined

6. In Appendix A.8, we show that we can use the weighting results in Abadie [2003] to derive the same result in Theorem 1.5.1.

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 $\kappa \coloneqq 1 - \frac{T_i(1-Z_i)}{\mathbb{P}[Z_i=0]} - \frac{(1-T_i)Z_i}{\mathbb{P}[Z_i=1]}$. Theorem 1.5.1 strengthens Abadie's κ by further conditioning on those with an untreated potential outcome of 0. Thus, a natural question is whether or not we can point identify $\mathbb{E}[g(Y_i, T_i, X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)]$ under the IA IV assumption. The answer is no. To see this:

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

where the first equality uses the fact that $T_i = Z_i$ for compliers, the fourth equality uses the IV independence assumption. Due to the presence of $\mathbb{E}[g(Y_i(1), 1, X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)]\mathbb{P}[Z_i = 1]$, which is about the joint distribution of potential outcomes, $\mathbb{E}[g(Y_i, T_i, X_i) \mid Y_i(0) = 0, T_i(1) > T_i(0)]$ is not point identified with the IA IV assumptions.

Theorem 1.5.1 can be applied to continuous Y_i by defining a new indicator variable, $\tilde{Y}_i = \mathbb{1}\{Y_i \in B\}$, where B is a measurable set, and a new potential outcome, $\tilde{Y}_i(0) = \mathbb{1}\{Y_i(0) \in B\}$. The result in Theorem 1.5.1 holds for $\tilde{Y}_i$ under the IA IV assumptions in Assumption 1.2.1. An example of B is: $B = \mathbb{1}\{Y_i(0) \leq \tilde{y}\}$. That is, researchers can identify characteristics measured by X_i of compliers and those with untreated outcomes less than $\tilde{y}$.

Since the marginal distribution of potential outcomes among compliers is identifiable, a natural extension of Theorem 1.5.1 is to extend the results to the following subpopulations: $[Y_i(0) = 1, T_i(1) > T_i(0)]$, $[Y_i(1) = 0, T_i(1) > T_i(0)]$, and $[Y_i(1) = 1, T_i(1) > T_i(0)]$.

Proposition 1.5.1. Assume that 1 to 4 in Assumption 1.2.1 hold, and let $g : \mathbb{R} \to \mathbb{R}$ be measurable such that $\mathbb{E}[|g(X_i)|] < \infty$, then, the following conditional expectations of $g(X_i)$ are point identified:

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

By the identical reasoning after Theorem 1.5.1, we have the following three remarks on Proposition 1.5.1. First, the results show that any conditional moments defined by pre-treatment covariate X_i can be identified as long as the moments are finite. Second, pick $g(X_i) = \mathbb{1}\{X_i \leq x\}$ with $x \in \mathbb{R}$, the results show that the conditional cumulative functions are identified. Third, Proposition 1.5.1 strengthens Abadie's κ by further conditioning on the potential outcome. However, by the same token in the discussion before, the power of Abadie's κ is not fully preserved here, because we cannot identify $g(Y_i, T_i, X_i)$ conditional on the three subpopulations above.

1.5.2 Identification: Compliance and Persuasion

An implication of Lemma 1.4.2 is that we can point identify the statistical properties of always-persuaded, never-persuaded, and persuaded among compliers. The results follow because the joint distribution of potential outcomes among compliers is point identified under the monotone treatment response assumption in the binary IA IV model. The results are summarized in Theorem 1.5.2.

Theorem 1.5.2 (Compliance and Persuasion)**.** Suppose Assumption 1.2.1 holds, let $g : \mathbb{R} \to \mathbb{R}$ be measurable such that $\mathbb{E}[|g(X_i)|] < \infty$, then, the moments of $g(X_i)$ conditional on always-persuaded compliers, never-persuaded compliers, and persuaded compliers are point identified:

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

We now give three remarks for Theorem 1.5.2. By the identical argument in Theorem 1.5.1, the conditional distribution functions of X_i given persuasion types and compliers are also identifiable, because we can let $g(X_i)$ being $g(X_i) = \mathbb{1}\{X_i \leq x\}$ with $x \in \mathbb{R}$. Furthermore, for measurable g, the expectations of $g(X_i)$ conditional on the three subpopulations are also identifiable given the expectation is well-defined. An implication is any statistical moments of the always-persuaded, never-persuaded, and persuaded among compliers are identifiable. Thus, this theorem extends the weighting results in Abadie [2003] by further conditioning on the persuasion types defined by the pair of potential outcomes.

The aforementioned statistics provide significant aid in comprehending the intervention's impact and mechanism. To illustrate, consider the GOTV experiment. Theorem 1.5.2 establishes the identification of the probability of a complier-persuaded voter being a Democrat. In other words, although GOTV experiments are not typically partisan ex ante, they can produce partisan mobilization outcomes. For instance, the data may indicate that among compliers, the likelihood of a persuaded voter being a Democrat is exceedingly high. If conducted in a swing state, the mobilization experiment could potentially alter the election results. Furthermore, the results of Theorem 1.5.2 can facilitate our evaluation of the mechanisms by which the treatment affects the outcome. In the GOTV experiment, the aforementioned results can be employed to evaluate the hypothesis that voting is habit-forming [Gerber et al., 2003]. We can utilize prior voting records as a metric for the voting propensity. If the hypothesis in Gerber et al. [2003] is accurate, we should observe that always-persuaded voters among compliers exhibit the highest voting propensity while never-persuaded voters demonstrate the lowest voting propensity.[7]

In addition to Theorem 1.5.2, there are several other ways to profile voters using observable covariates. For instance, researchers might be interested in the following quantity: among the compliers and those who will not vote without being exposed to the treatment (i.e., the locally persuadable), what are the characteristics of those who will vote with being exposed to the treatment. For example, in the GOTV experiment, this quantity would be the chance of locally persuadable individuals being a Democrat and will if they are exposed to the treatment. Due to the monotone treatment response and binary outcome, there are five other estimands that share a similar flavor with this example. The identifiability of these

7. In Appendix A.3, we present results that identify the proportion of persuasion types among compliers while conditioning on covariates.

estimands follows from the fact that the monotone treatment response assumption implies the identifiability of the joint distribution of the potential outcomes among compliers. These results are formally stated in Proposition 1.5.2.

Proposition 1.5.2. Suppose Assumption 1.2.1 holds, let $g : \mathbb{R} \to \mathbb{R}$ be measurable such that $\mathbb{E}[|g(X_i)|] < \infty$, then, the following conditional expectations are identifiable:

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

1.6 Estimation and Inference

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

$$\frac{\mathbb{E}[f(X_i, Y_i, T_i) \mid Z_i = 1] - \mathbb{E}[f(X_i, Y_i, T_i) \mid Z_i = 0]}{\mathbb{E}[h(Y_i, T_i) \mid Z_i = 1] - \mathbb{E}[h(Y_i, T_i) \mid Z_i = 0]}. \tag{1.1}$$

where f and h are measurable functions that map from $\mathbb{R}$ to $\mathbb{R}$. It is easy to see that the numerator in Equation 1.1 is the coefficient of Z_i from regressing $f(X_i, Y_i, T_i)$ on Z_i and a constant, while the denominator in Equation 1.1 is the coefficient of Z_i from regressing $h(Y_i, T_i)$ on Z_i and a constant. Therefore, the standard estimation and inference theory for Wald estimand applies immediately to the current case with i.i.d. data of (Y_i, T_i, Z_i, X_i). We can either employ the conventional asymptotic results for hypothesis testing or use the Anderson-Rubin test which is robust to weak identification. [8] Note that both inferential methods can be easily implemented in standard statistical software, say, `ivreg2` and `weakiv` in `Stata`.[9]

1.7 Discussion

In this section, we discuss three points on identification results from previous sections. Firstly, we compare θ_{local} with classic estimands. Next, we provide necessary and sufficient conditions for approximated θ_{DK} to equal θ_{local} under one-sided non-compliance. Additionally, we propose a test for Assumption 1.2.1, and a simple method to assess the sensitivity of results to the monotone treatment response assumption.

8. We provide a more detailed discussion on inference issues in Appendix A.5.

9. `ivreg2` does not produce a confidence interval for the Anderson-Rubin test, while `weakiv` does.

1.7.1 Comparison with Existing Estimands

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:

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

We can point identify $\theta_{\text{local untreated}}$ given Assumption 1.2.1. The intuition of the identification of $\theta_{\text{local untreated}}$ is that conditioning on compliers implies that $T_i = Z_i$, thus, $\theta_{\text{local untreated}} = \theta_{\text{local}}$. We formally state the result in Claim 1.7.1.

Claim 1.7.1. Assume that Assumption 1.2.1 holds, then, $\theta_{\text{local untreated}}$ is point identifiable:

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

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 $\tilde{\theta}_{\mathrm{DK}}$:

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

As noted in Jun and Lee [2023], $\tilde{\theta}_{\mathrm{DK}}$ is not a well-defined conditional probability. Therefore, $\tilde{\theta}_{\mathrm{DK}}$ does not measure persuasion rate for any subpopulation.

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

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

Theorem 1.7.1. Assume that Assumption 1.2.1 holds, if there is one-sided non-compliance in the control group, then $\theta_{\mathrm{DK}} = \theta_{\mathrm{local}}$ 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 $\theta_{\mathrm{DK}} = \theta_{\mathrm{local}}$ if and only if $Y_i(0) \perp\!\!\!\perp T_i(1)$.

These results contrast sharply with the results in Jun and Lee [2023], 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.

1.7.2 A Sharp Test of the Identification Assumptions

The main identification results in Theorem 1.5.2 rely on two assumptions: the IA IV assumptions and the monotone treatment response assumption. These assumptions impose restrictions on individuals' choice behaviors by ruling out the dissuaded and the defiers and are thus subject to criticism for being too strong. To address this issue, we propose a sharp test for Assumption 1.2.1.

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

$$A_{\text{obs}}\mathbf{p} = \mathbf{b}, \tag{1.2}$$

where A_{obs} is a matrix that reflects the restrictions on the data, $\mathbf{p}$ is a vector of the unobserved persuasion and compliance types defined in Table 1.1, $\mathbf{b}$ is a collection of observed quantities, for example $\mathbb{P}[Y_i = 0, T_i = 0, X_i \in A \mid Z_i = 0]$. An example of A_{obs}, $\mathbf{p}$, and $\mathbf{b}$ can be found in Appendix A.6. Thus, the observed quantity $\mathbf{b}$ is consistent with Assumption 1.2.1 if there exists a solution to the system of linear equations in 1.2. We now summarize this observation to Proposition 1.7.1.

Proposition 1.7.1. If Assumption 1.2.1 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 1.7.1 is that to test the validity of Assumption 1.2.1, 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:

$$H_0 : P \in \mathbf{P}_0 \text{ versus } H_1 : P \in \mathbf{P} \setminus \mathbf{P}_0 \tag{1.3}$$

where $\mathbf{P}_0 \coloneqq \{P \in \mathbf{P} : \exists \mathbf{p} \geq \mathbf{0} \text{ s.t. } A_{\text{obs}}\mathbf{p} = \mathbf{b}\}$, which is the set of distributions that is consistent with Assumption 1.2.1. Thus, if H_0 is rejected, we have strong evidence against the validity of Assumption 1.2.1. However, if H_0 is not rejected, we cannot confirm the validity of Assumption 1.2.1. In this precise sense, Assumption 1.2.1 is a refutable but nonverifiable hypothesis [Kitagawa, 2015].

In terms for the implementation of testing 1.3, with discrete X_i, we can set A to be the support of X_i, and proceed the test using the recent advancement on testing whether there exists a nonnegative solution to a possibly under-determined system of linear equations with known coefficients [Bai et al., 2022a, Fang et al., 2023]. One computationally intensive, yet feasible method for testing H_0 proposed in Bai et al. [2022a] is to use subsampling method. With the subsampling method, by using the classic results in Romano and Shaikh [2012], Bai et al. [2022a] shows that the test controls size uniformly over $\mathbf{P}$. The test statistic in Bai et al. [2022a] is given by:

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

where $\hat{\mathbf{b}}$ is an estimator of $\mathbf{b}$.[10] For the subsampling-based test, Bai et al. [2022a] defines

10. 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 A.7.

the following quantity:

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

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

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

1.7.3 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 1.4.2.

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.

1.8 Empirical Application: Revisit Green et al. [2003]

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

1.8.1 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 1.2.

1.8.2 Empirical Results

We first present the results for the marginal and joint distribution of potential outcomes of compliers in Table 1.3. Our results reveal two interesting patterns. First, conditional on

Table 1.2: Summary Statistics in Green et al. [2003]

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

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

compliers, most of them are never-persuaded in both samples. Second, only 7.9% of voters are persuaded conditional on compliers in the full sample, and 13.9% of voters are persuaded conditional on compliers in Bridgeport.

We now apply Theorem 1.5.1 and Theorem 1.5.2 to this experiment. The results are presented in Table 1.4. For the full sample, the probability of voting in the 2000 presidential election conditional on the locally persuadable (that is, those who do not vote without the information treatment and compliers) is 60.3%. A more interesting finding is that the subpopulation of always-persuaded compliers has the highest probability (that is, 95.4%) of voting in the 2000 presidential election. The results show that if always-persuaded and complier voters vote in the low-profile local elections regardless of the GOTV intervention, they will very likely vote in the high-profile 2000 presidential elections. This empirical pattern is consistent with the robust findings on the persistent of voting behavior [Gerber et al., 2003]. One potential explanation of the persistent of the voting behavior is that voting behavior is habit-forming [Gerber et al., 2003]. As expected, the subpopulation of never-persuaders and compliers has the lowest probability of voting in the 2000 presidential election.

Table 1.3: Distribution of Potential Outcomes in Green et al. [2003]

	Estimates	95% CI	95% AR CI
Panel A: Full Sample			
$\mathbb{P}[Y_i(0) = 1 \mid T_i(1) > T_i(0)]$	0.302	(0.261, 0.343)	(0.263, 0.343)
$\mathbb{P}[Y_i(1) = 1 \mid T_i(1) > T_i(0)]$	0.381	(0.364, 0.398)	(0.365, 0.397)
$\mathbb{P}[Y_i(0) = 1, Y_i(1) = 1 \mid T_i(1) > T_i(0)]$	0.302	(0.261, 0.343)	(0.263, 0.343)
$\mathbb{P}[Y_i(0) = 0, Y_i(1) = 0 \mid T_i(1) > T_i(0)]$	0.619	(0.602, 0.636)	(0.603, 0.635)
$\mathbb{P}[Y_i(0) = 0, Y_i(1) = 1 \mid T_i(1) > T_i(0)]$	0.079	(0.035, 0.123)	(0.036, 0.122)
Panel B: Bridgeport			
$\mathbb{P}[Y_i(0) = 1 \mid T_i(1) > T_i(0)]$	0.111	(0.019, 0.202)	(0.02, 0.202)
$\mathbb{P}[Y_i(1) = 1 \mid T_i(1) > T_i(0)]$	0.25	(0.197, 0.303)	(0.196, 0.303)
$\mathbb{P}[Y_i(0) = 1, Y_i(1) = 1 \mid T_i(1) > T_i(0)]$	0.111	(0.019, 0.202)	(0.02, 0.202)
$\mathbb{P}[Y_i(0) = 0, Y_i(1) = 0 \mid T_i(1) > T_i(0)]$	0.75	(0.697, 0.803)	(0.697, 0.804)
$\mathbb{P}[Y_i(0) = 0, Y_i(1) = 1 \mid T_i(1) > T_i(0)]$	0.139	(0.033, 0.245)	(0.034, 0.244)

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

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

For the Bridgeport sample, the most interesting result is that among compliers and persuaded, the chance of them being a Democrat is very high. However, its confidence interval is pretty wide. Mobilizing more Democrats in the school board election in Bridgeport has practical implications for two reasons. First, Democrats are more pro-union. Second, the turnout rate in school board elections is usually low.[11] 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].

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

Table 1.4: Profiling Persuasion Types in Green et al. [2003]

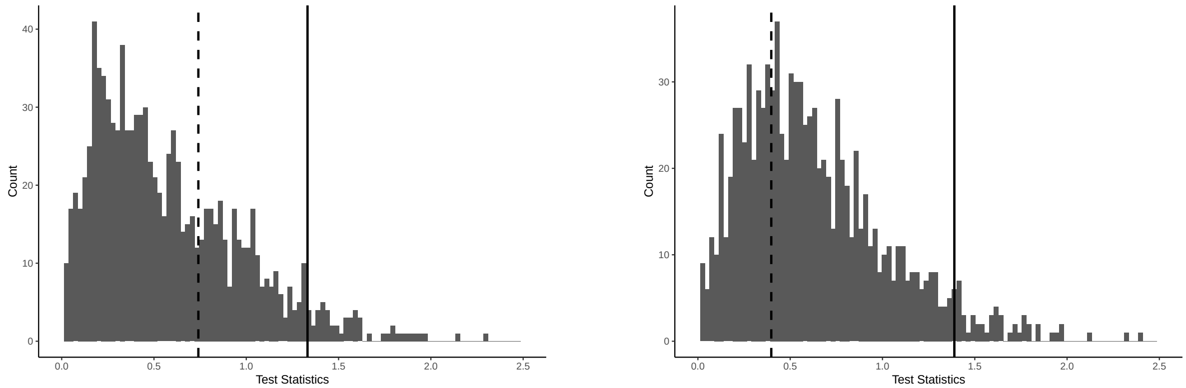
	Estimates	95% CI	95% AR CI
Panel A: Full Sample			
$\mathbb{P}$[Voted in 2000 = 1|Locally Persuadable]	0.603	(0.547, 0.659)	(0.549, 0.659)
$\mathbb{P}$[Voted in 2000 = 1|AP, C]	0.954	(0.914, 0.994)	(0.914, 0.994)
$\mathbb{P}$[Voted in 2000 = 1|NP, C]	0.511	(0.489, 0.534)	(0.489, 0.533)
$\mathbb{P}$[Voted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

Note: This table provides the results of profiling different persuasion types using pre-treatment covariates. CI refers to confidence interval. AR refers to Anderson-Rubin. Locally persuadable refers to the following event: $[Y_i(0) = 0, T_i(1) > T_i(0)]$. C refers to the following event: $[T_i(1) > T_i(0)]$. AP refers to the following event: $[Y_i(1) = Y_i(0) = 1]$. NP refers to the following event: $[Y_i(1) = Y_i(0) = 0]$. P refers to the following event: $[1 = Y_i(1) > Y_i(0) = 0]$.

1.8.3 Testing Identification Assumptions and Sensitivity Analysis

We implement the test for the Assumption 1.2.1 by using Proposition 1.7.1. We use the subsampling method in Bai et al. [2022a] for this test.[12] The results in Figure 1.1 show that we cannot reject the validity of the identification assumptions at the 5% level for both the full sample and the Bridgeport sample. Furthermore, we provide the sensitivity analysis result on the joint distribution of potential outcomes in Table 1.5 by varying the degree to which the monotone treatment response assumption is violated among compilers. Interestingly, when the violation becomes larger, the proportion of persuaded among compliers increases.

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



(a) Full Sample

(b) Bridgeport

Figure 1.1: Test Identification Assumptions using Bai et al. [2022a]

Note. These figures present the results for testing identification assumptions. Figure 1.1a presents the results using the full sample. Figure 1.1b presents the results using the sample from Bridgeport. The solid lines are the critical values for a 5% level test. The dashed lines are the test statistics.

1.9 Conclusion

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

Furthermore, we develop a weighting method that helps researchers identify the statistical characteristics measured by the pre-treatment covariates of persuasion types: compliers and always-persuaded, compliers and persuaded, and compliers and never-persuaded. These

Table 1.5: Sensitivity for Distribution of Potential Outcomes in Green et al. [2003]

Panel A: Full Sample						
Sensitivity Parameter						
$\mathbb{P}[Y_i(1)=0, Y_i(0)=1 \mid T_i(1) > T_i(0)]$	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.

findings extend the "κ weighting" results in Abadie [2003], which can profile the characteristics of compliers measured by pre-treatment covariates. This method can provide richer information on the treatment effect. For instance, some GOTV experiments aim at mobilizing underrepresented minorities. With this methodology, researchers can estimate the chance of the compliers and mobilizable voters being underrepresented minorities. Thus, researchers can assess whether or not their interventions achieve their normative goals.

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

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

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

CHAPTER 2

PROFILING PERSUASION TYPES IN A BINARY IA-IV MODEL: THREE EMPIRICAL APPLICATIONS

2.1 Introduction

In this chapter, I apply the methods I developed for profiling persuasion types to three papers [Enikolopov et al., 2011, Blattman and Annan, 2016, Chen and Yang, 2019]. Applying the methods to Enikolopov et al. [2011] yields the following findings. First, among the compliers, the majority of them were persuaded, and very few of them were always-persuaded. The empirical results here contrast sharply to the other applications in which the percentage of persuaded among compliers is usually very small. Second, the persuaded voters were less likely to be male and the average persuaded voters were middle-aged, while the never-persuaded voters were more likely to be male and the average never-persuaded voters were more likely to be young adults.

Applying the methods to Blattman and Annan [2016] yields the following findings. Around 25% of the Liberian ex-fighters were persuaded, while more than half of them were always-persuaded. The persuaded Liberian ex-fights were likely to have fewer kids, fewer years of schooling and younger. The always-persuaded Liberian ex-fighters, on the other hand, were likely to have more kids, higher years of schooling and be in their early 30s. The demographic variables of the never-persuaded Liberian ex-fighters were between the always-persuaded and persuaded ex-fighters. Moreover, the persuaded ex-fighters were likely to be more aggressive, less easy to reintegrate, and riskier. Meanwhile, the always-persuaded ex-fighters were less aggressive, easier to reintegrate, and less risky.

Applying the methods to Chen and Yang [2019] yields the following findings. Among the compliers, the majority of the students were never persuaded, and roughly 20% of the students who comply were persuaded. Furthermore, always-persuaded students were more likely to come from wealthier families and less likely to be members of the CCP. The persuaded students were also from financially well-off families and were more likely from the coastal areas. The always-persuaded students exhibited higher levels of risk index, patience index, and were more likely to believe that people are good. In contrast, the persuaded students displayed the opposite traits compared to the always-persuaded students.

2.2 Revisit Enikolopov et al. [2011]

This section demonstrates the application of the methods using Enikolopov et al. [2011] as an example. First, we provide information on the empirical setup. Then, we present the empirical results on the joint distribution of potential outcomes. Finally, we present the results for profiling the persuasion types among compliers using voters' demographic and political variables.

2.2.1 Empirical Setup

What is the effect of watching independent media on whether or not voters voted for Putin? To address this question, researchers need to resolve the endogeneity problem. For example, voters who actively seek and access uncensored media might have systematically different political preferences. To answer this question, Enikolopov et al. [2011] finds exogenous variations to the access of NTV, an independent media in Russia around 2000, to evaluate the impact of watching NTV on voters' voting choices.

Enikolopov et al. [2011] used the data on the location of NTV transmitters inherited from a Soviet educational channel and geographical variation in signal propagation, they then calculated the strength of the signal in each locality in Russia, and on the basis of the signal strength predict the availability of NTV. Following Jun and Lee [2023], we create a binary instrument

$$Z_i = \mathbb{1}\{\text{Signal Power}_i > \text{median}(\text{Signal Power}_i)\},$$

by using the original continuous instrument. As documented in Enikolopov et al. [2011], there is noncompliance with this instrument.

Enikolopov et al. [2011] examined multiple outcomes, including voting for the most popular opposition party OVR ("Fatherland–All Russsia"), voting for the pro-government party "Unity", etc. In our analysis below, we will focus on the outcome variable that measures whether the voters voted for OVR. The summary statistics are presented in Table 2.1.

Table 2.1: Enikolopov et al. [2011]: Summary Statistics

	N	Mean	Standard Deviation	Min	Max
Vote OVR	1,624	0.094	0.292	0	1
NTV	1,624	0.629	0.483	0	1
IV	1,624	0.477	0.500	0	1
Male	1,624	0.373	0.484	0	1
Age	1,624	30.326	16.863	0	71
High School	1,619	0.786	0.410	0	1
Marriage	1,620	0.582	0.493	0	1
Vote Yabloko	646	0.099	0.299	0	1
Vote KPRF	646	0.296	0.457	0	1
Vote LDPR	646	0.080	0.272	0	1
Vote DVR	646	0.031	0.173	0	1
Vote 1995	789	0.828	0.378	0	1

This table presents summary statistics for the sample used in our empirical analysis.

2.2.2 Empirical Results

We first present the results for the joint distribution of potential outcomes of compliers in Table 2.2. There are three interesting findings. First, we cannot reject the null hypothesis that the proportion of the always-persuaded voters is 0. Second, 31.7% of the voters were never-persuaded. Third, the majority (that is, around 66.9%) of voters were persuaded. This effect is very large compared with the other applications that we used.

Table 2.2: Enikolopov et al. [2011]: Joint Distribution of Potential Outcomes for Voting for OVR

	Share
AP	0.013
	(0.027)
NP	0.317
	(0.064)
P	0.669
	(0.057)

This table presents the estimated joint distribution of potential outcomes among compliers. AP refers to the following event: $[Y_i(1) = Y_i(0) = 1]$. NP refers to the following event: $[Y_i(1) = Y_i(0) = 0]$. P refers to the following event: $[1 = Y_i(1) > Y_i(0) = 0]$.

The results for profiling the persuasion types among compliers using the demographic variables are presented in Table 2.3. Note that the results for the always-persuaded are noisy, which is consistent with the finding in Table 2.2 that there are very few always-persuaded voters. However, the results for the never-persuaded and persuaded voters are informative.

Regarding the male variable, the results show that the never-persuaded voters are more likely to be male, compared with average voter. For the persuaded voters, we fail to reject that the chance of them being a man is 0.

Regarding the age variable, there are two interesting findings. First, the average age of the never-persuaded voters was around 26 years old. In other words, the never-persuaded voters were likely to be young adults. Second, the average age of the always-persuaded voters was around 37. In other words, the persuaded voters were likely to be middle-aged, and were older than average voter.

Regarding the education variable which measures whether or not the person at least attended high school, the results show that both the never-persuaded and persuaded voters were likely to attend at least high school.

Regarding the marriage status, roughly half of the never-persuaded and persuaded voters married.

The results for profiling the persuasion types among compliers using the political variables are presented in Table 2.4. Considering that approximately 50% of the samples have missing values on the political variables, the results should be interpreted with caution.

Three findings of profiling persuasion types with political variables merit discussion. First, among the never-persuaded voters, 36.5% of them voted for Yabloko ("Russian United Democratic Party") in 1995. Second, among the never-persuaded and persuaded voters, the likelihood of them voting in 1995 was approximately 70%. Third, the other voting variables are imprecisely estimated.

We implement a joint test for testing the identification assumptions in the binary IV model of persuasion. We use the subsampling method in Bai et al. [2022a] for this test.

Table 2.3: Enikolopov et al. [2011]: Profiling Demographic Variables among Compliers for Voting for OVR

	AP	NP	P	Summary
Male	1.433	0.695	0.085	0.373
	(2.708)	(0.121)	(0.103)	(0.484)
Age	16.137	25.999	37.037	30.326
	(49.945)	(3.744)	(3.355)	(16.863)
High School	1.543	0.953	0.823	0.786
	(1.996)	(0.089)	(0.070)	(0.410)
Marriage	0.071	0.477	0.502	0.582
	(1.493)	(0.115)	(0.099)	(0.493)

This table presents the results for profiling the persuasion types among compliers. For the first three columns, standard errors are presented in the parentheses. The last column presents the sample averages of the pre-treatment covariates, and standard deviations are presented in parentheses. AP refers to the following event: $[Y_i(1) = Y_i(0) = 1]$. NP refers to the following event: $[Y_i(1) = Y_i(0) = 0]$. P refers to the following event: $[1 = Y_i(1) > Y_i(0) = 0]$.

The subsampling test in Bai et al. [2022a] requires us to pick a size for the subsample with $b_n \to \infty$ and $\frac{b_n}{n} \to 0$ as $n \to \infty$. We set b_n to $n^{\frac{2}{3}}$ here. The results in Figure 2.1 show that we cannot reject the validity of the identification assumptions at the 5% level.

Finally, we provide a comparison between the local persuasion rate proposed by Jun and Lee [2023] and the persuasion measure proposed by DellaVigna and Kaplan [2007]. The results are presented in Table 2.5. The results show a huge discrepancy between the two persuasion measures. Moreover, θ_{DK} is far above one, which illustrates the fact that it is not a well-defined conditional probability.

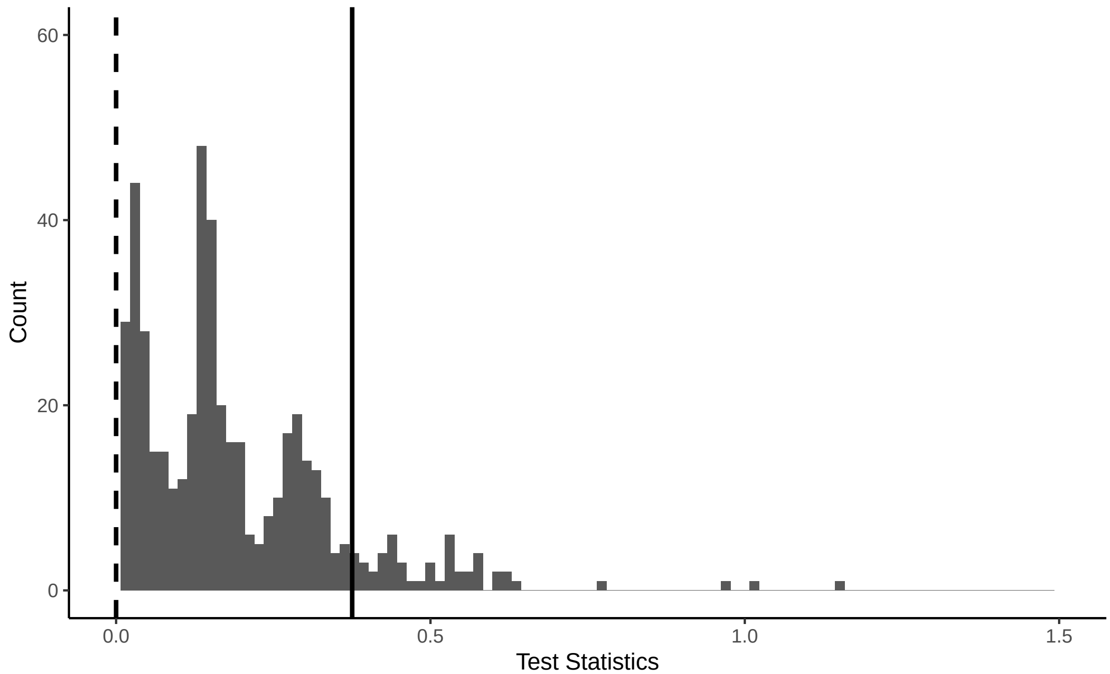


Figure 2.1: Enikolopov et al. [2011]: Testing Identification Assumptions

The figure presents the results for testing identification assumptions in Enikolopov et al. [2011]. The solid line is the critical value for a 5% level test. The dashed line is the test statistic.

2.3 Revisit Blattman and Annan [2016]

This section demonstrates the application of the methods using Blattman and Annan [2016] as an example. First, we provide information on the empirical setup. Then, we present the empirical results on the joint distribution of potential outcomes. Finally, we present the results for profiling the persuasion types among compliers using Liberian ex-fighters' demographic and violence variables.

2.3.1 Empirical Setup

What is the effect of job training programs for ex-fighters on rehabilitating high-risk men in a post-war fragile society? To address this question, researchers need to resolve the endogeneity problem, specifically, the fact that ex-fighters possess a comparative advantage

Table 2.4: Enikolopov et al. [2011]: Profiling Political Variables among Compliers for Voting for OVR

	AP	NP	P	Summary
Vote Yabloko	0.324	0.365	0.036	0.099
	(0.506)	(0.119)	(0.100)	(0.299)
Vote KPRF	1.295	-0.146	-0.046	0.296
	(1.429)	(0.177)	(0.281)	(0.457)
Vote LDPR	0.647	0.137	-0.098	0.080
	(0.826)	(0.084)	(0.212)	(0.272)
Vote DVR	-0.530	0.115	0.186	0.031
	(0.933)	(0.067)	(0.119)	(0.173)
Vote 1995	3.253	0.751	0.712	0.828
	(11.985)	(0.121)	(0.174)	(0.378)

This table presents the results for profiling the persuasion types among compliers. For the first three columns, standard errors are presented in the parentheses. The last column presents the sample averages of the pre-treatment covariates, and standard deviations are presented in parentheses. AP refers to the following event: $[Y_i(1) = Y_i(0) = 1]$. NP refers to the following event: $[Y_i(1) = Y_i(0) = 0]$. P refers to the following event: $[1 = Y_i(1) > Y_i(0) = 0]$.

in violence and frequently lack the human, social, and physical capital needed to succeed in peacetime labor markets [Blattman and Annan, 2016]. To answer this question, Blattman and Annan [2016] conducted a randomized controlled trial to evaluate the impact of job training programs on the career choices of Liberian ex-fighters who were illegally mining or occupying rubber plantations.

In Blattman and Annan [2016], the instrument Z_i is the randomly assigned job training program by the nonprofit Action on Armed Violence (AoAV). Specifically, there are four components in the job training program: (1) residential coursework and practical training in agricultural work; (2) counseling and a "life skills" class; (3) transportation to a community of the trainees' choice; and (4) a two-stage package of tools/supplies tailored to the trainee's

Table 2.5: Enikolopov et al. [2011]: Comparison Between θ_{local} and θ_{DK}

	Persuasion Rate Measures
DK Persuasion Rate	2.969
	(0.747)
Local Persuasion Rate	0.322
	(0.062)

This table presents the results of comparing θ_{local} and θ_{DK}. Standard errors are presented in the parentheses.

interests, such as vegetable farming or animal husbandry.

There were noncompliance issues in this experiment. 74% of those assigned to the job training program complied, in that they attended the program at least a day. 94% of those who attended a day graduated at the end [Blattman and Annan, 2016]. Blattman and Annan [2016] coded treatment T_i as 1 if person i attended the program for at least a day. The summary statistics are presented in Table 2.6.

Table 2.6: Blattman and Annan [2016]: Summary Statistics

	N	Mean	Standard Deviation	Min	Max
Agricultural Work	1,164	0.731	0.444	0	1
Legal Work	1,167	0.653	0.476	0	1
Treatment	1,274	0.398	0.490	0	1
IV	1,274	0.562	0.496	0	1
Number of Children	1,274	2.263	2.072	0	13
Education	1,274	5.632	3.831	0	16
Age	1,274	29.790	7.704	18	57
Closeness to Former Commanders	1,274	-0.021	0.970	-0.365	5.299
Was a Combatant	1,274	0.670	0.471	0	1
Aggression Index	1,274	1.338	1.940	0	12
Ease of Reintegration	1,274	-0.042	0.976	-0.891	2.919
Patience Index	1,274	2.985	0.872	0	4
Risk Index	1,274	0.329	0.587	0	3

This table presents summary statistics for the sample used in our empirical analysis.

Blattman and Annan [2016] examined the effect of the training program on many outcomes in their paper. In this empirical study, we will investigate the two main outcomes: (1) whether the individual engaged in agricultural work after the training, and (2) whether the individual fully participated in legal activities. We focus on these two outcomes for two reasons. First, Blattman and Annan [2016] found strong effects of the training program on these two outcome variables. Second, the monotone treatment response assumption is more likely to hold for those two outcome variables. The job training program enhances each trainee's human capital in agricultural work and other soft skills necessary for success in the labor market. Consequently, it increases the trainee's likelihood of working in the legal sector.

2.3.2 Empirical Results

We first present the results for the joint distribution of potential outcomes of compliers for agricultural work and legal activities in Table 2.7 and Table 2.8, respectively. There are two interesting patterns. First, conditional on compliers, more than half of them were always-persuaded for both outcomes. Second, in terms of the agricultural work, 16% of the ex-fighters who were compliers were persuaded, while 34.8% of the ex-fighters who were compliers were persuaded. The observation that the effect was more significant for legal activities aligns with findings suggesting that job training programs teaching ex-fighters "soft skills" enhance their abilities not only to succeed in agricultural jobs but also to excel in the labor market during peacetime [Blattman et al., 2017].

We now present the results for profiling compliers using the demographic variables. We look at three specific demographic variables, namely, the number of kids, education level, and age. The results for agricultural work and legal activities are presented in Table 2.9 and Table 2.10, respectively.

There are three interesting empirical findings. First, for the persuaded trainees, we fail to reject the null that the number of children they have is 0. This is consistent with the finding that they are the youngest among the three persuasion types. On the other hand, trainees who are always persuaded have roughly three children. This is consistent with the interpretation that they bear the responsibility of childbearing and provide financial support through the labor market. Moreover, the never-persuaded trainees have fewer kids than the always-persuaded trainees.

Second, as for the education variable, the level of education measured by years of schooling is the lowest among the persuaded trainees, although the point estimate for the legal activities outcome is not precisely estimated. This finding can be interpreted as evidence of the higher marginal return of the training program for individuals with lower levels of education. In contrast, the always-persuaded trainees have the highest level of education. This finding is consistent with the opportunity cost hypothesis of crime [Becker, 1968, Grossman, 1991]. Individuals with higher levels of education possess higher human capital, allowing them to earn higher wages in the labor market. This raises their opportunity cost of joining a rebellious group, making it more likely for them to work in the legal sector.

Thirdly, regarding the age variable, persuaded trainees are the youngest, whereas always-persuaded individuals are the oldest. It's worth noting that the average age is not precisely estimated for the persuaded group when legal activities are used as the outcome variable. These findings are consistent with the findings for the number of children.

We now present the results for profiling compliers using the variables that measure the tendency for violence. Specifically, we look at six variables, namely, closeness to former commanders, whether the person was a combatant or not, aggression index, ease of reintegration, patience index, and risk index. The results for agricultural work and legal activities are presented in Table 2.11 and Table 2.12, respectively.

There are six interesting findings. Regarding the closeness to former commanders, persuaded trainees have the lowest score on this, though the estimates when the outcome is engaging in legal activities are not precisely estimated. Interestingly, always-persuaded trainees have the highest level of being close to former commanders. Moreover, the persuaded ex-fighters are much less close to former commanders than average ex-fighters.

For the variable that measures whether or not the trainees were a combatant before, the results show that the chance of being a combatant is around 70% across three persuasion types when the outcome variable is agricultural work. We observe similar patterns when engaging in legal activities is used as the outcome variable, although the estimates for the persuaded group become much noisier.

For the variable that measures individuals' aggression level, the always-persuaded trainees have the lowest measure for the aggression index. This finding is again consistent with the opportunity cost hypothesis for crime [Becker, 1968, Grossman, 1991]. Individuals with lower aggression levels have a comparative advantage in the labor market, enabling them to earn higher wages. This increased earning potential raises their opportunity cost of joining a rebellious group, making it more likely for them to choose legal employment. However, the results for the never-persuaded and persuaded are less conclusive. Moreover, the persuaded ex-fighters' aggression level is higher than average ex-fighters.

For the variable that measures individuals' ease of integration, an interesting finding is that this index is low for the never-persuaded trainees. This finding again is consistent with the opportunity cost hypothesis for crime [Becker, 1968, Grossman, 1991]. Individuals who are more difficult to integrate might have a comparative advantage in rebellious activities, which raises their opportunity cost of working in the legal sector, making it less likely for them to choose legal employment. The results for always-persuaded and persuaded are less conclusive. However, the point estimates show that, compared with never-persuaded trainees, always-persuaded trainees are easier to integrate, whereas persuaded trainees are more difficult to integrate. Moreover, the persuaded ex-fighters' ease of reintegration is far lower than average ex-fighters.

For the variable that measures individuals' level of patience, we find that the persuaded trainees have the highest level of patience, while the never-persuaded trainees have the lowest level of patience. The patience index of the always-persuaded trainees is in the middle.

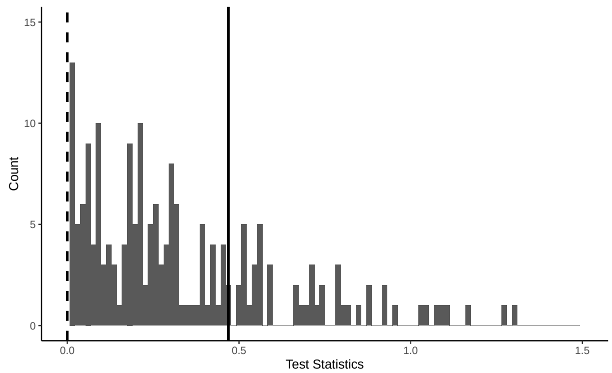
For the variable that measures individuals' level of risk, we find that the always-persuaded trainees have the lowest level of risk, and the persuaded trainees have the highest level of risk. However, a caveat is that the estimates for the persuaded trainees are noisy. The risk index of the never-persuaded trainees is in the middle.

We implement a joint test for testing the identification assumptions in the binary IV model of persuasion. We use the subsampling method in Bai et al. [2022a] for this test. The subsampling test in Bai et al. [2022a] requires us to pick a size for the subsample with $b_n \to \infty$ and $\frac{b_n}{n} \to 0$ as $n \to \infty$. We set b_n to $n^{\frac{2}{3}}$ here. The results in Figure 2.2 show that we cannot reject the validity of the identification assumptions at the 5% level for both outcome variables.

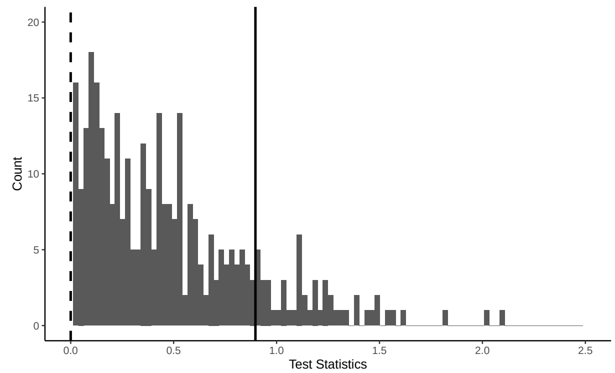
Finally, we provide a comparison between the local persuasion rate proposed by Jun and Lee [2023] and the persuasion measure proposed by DellaVigna and Kaplan [2007]. The results for agricultural work and legal activities are presented in Table 2.5 and Table 2.14 respectively.. The results show a huge discrepancy between the two persuasion measures. Moreover, θ_{DK} is far above one, which illustrates the fact that it is not a well-defined conditional probability.

2.4 Revisit Chen and Yang [2019]

This section demonstrates the application of the methods using Chen and Yang [2019] as an example. First, we provide information on the empirical setup. Then, we present the



(a) Agricultural Work



(b) Legal Activities

Figure 2.2: Blattman and Annan [2016]: Testing Identification Assumptions

The figures present the results for testing identification assumptions in Blattman and Annan [2016]. The left panel presents the results when the outcome variable is agricultural work. The right panel presents the results when the outcome variable is legal work. The solid lines are the critical values for a 5% level test. The dashed lines are the test statistics.

empirical results on the joint distribution of potential outcomes. Finally, we present the results for profiling the persuasion types among compliers using students' demographic and preference variables.

2.4.1 Empirical Setup

What is the effect of uncensored media on individuals' political outcomes in authoritarian regimes? To address this question, researchers need to resolve the endogeneity problem. For example, individuals living under autocracies who actively seek and access uncensored media might have systematically different political preferences. To answer this question, Chen and Yang [2019] conducted a randomized controlled trial among the students at Peking University to evaluate the impact of uncensored internet on various political and economic outcomes.

In Chen and Yang [2019], the instrument Z_i is the encouragement to view foreign news websites and/or the randomly assigned access to the virtual proxy network (VPN) that can aid internet users in China to gain access to websites blocked by the Great Firewall.[1] In Chen and Yang [2019], the instrument is discrete. Specifically, there were four assigned groups in the original study: (1) the control group; (2) an encourage treatment that encourages students to visit foreign news websites blocked by the Great Firewall; (3) an access treatment that gives students free access to uncensored internet; (4) an access and encouragement treatment that gives students both free access to uncensored internet and the encouragement treatment.

In the analysis below, we only keep the control group and the treatment group for both received the encouragement and the access. We focus on these two groups because the treatment effects in Chen and Yang [2019] are the strongest for the access and the encouragement group.

There were noncompliance issues in this experiment. In the treatment group that both receives the access and the encouragement, roughly 50% of students become active users of the VPN. Note that active users are defined as students whose accounts show at least one browsing activity per day for more than 40 days after the encouragement treatment ends [Chen and Yang, 2019]. The summary statistics are presented in Table 2.15.

Chen and Yang [2019] examined the effect of uncensored internet on many outcomes, including media-related behaviors and beliefs, knowledge of recent events, economic beliefs, political beliefs, and planned behaviors. In this empirical study, we will focus on one outcome variable, which measures whether or not the student trusts foreign media.

1. The Great Firewall, a major part of the umbrella Golden Shield Project directed by China's Ministry of Public Security, has operated since 2003 and serves as the main infrastructure blocking access to potentially unfavorable incoming data from foreign media outlets [Chen and Yang, 2019].

2.4.2 Empirical Results

We first present the results for the joint distribution of potential outcomes of compliers in Table 2.16. There are three interesting patterns. First, conditional on compliers, more than half of them were never-persuaded. Second, roughly one-quarter of the students were always-persuaded. Third, around 20% of the students were persuaded.

We now present the results for profiling compliers using the demographic variables. We look at nine specific demographic variables, namely, whether or not they are a member of the Chinese Communist Party (CCP), household income, Han ethnicity, Hukou status, gender, number of siblings, born in costal areas or not, live in costal areas or not, and English ability. The results are presented in Table 2.17.

For the CCP member, there are three interesting findings. First, for the always-persuaded students, we cannot reject the null hypothesis that the probability of them being a CCP member is 0. Second, among the never-persuaded and persuaded students, the chance of them being a CCP member is around 10%. Note that in this sample, 6.77% of the students were CCP members. Moreover, the always-persuaded students were much less likely to be CCP members, whereas the never-persuaded students were more likely to be CCP members.

For the household income, there are three findings. First, the always-persuaded students' household income is the highest among the three persuasion groups. Second, the never-persuaded students' household income is the lowest among the three persuasion groups. The findings are consistent with the findings in Roberts [2018] that the people who circumvented the censorship tool in China were those who were richer. Moreover, the always-persuaded students were richer than average students, and the never-persuaded students were less rich than average students.

Regarding Hukou status, the three groups have approximately the same probability of having a city Hukou.

For the gender variable, we find that the always-persuaded students are more likely to be men than the never-persuaded and persuaded. Note that there were 53.41% of the respondents in the sample were men.

For the number of siblings, there are two interesting findings. First, for the always-persuaded students, we cannot reject the null that the number of siblings was 0. In other words, we cannot reject the null hypothesis that the always-persuaded student was a single child. However, the never-persuaded and persuaded students were more likely to have siblings.

For whether or not the students were born and lived in coastal areas, there are two findings. First, the point estimates for always-persuaded are noisy, that is, we cannot reject the null hypothesis that the chance that always-persuaded students were born and lived in coastal areas is 0. Second, we find that the persuaded students are more likely to be born and live in coastal areas than the never-persuaded students. The second finding is also consistent with the findings in Roberts [2018].

For English ability measures, the profiling results are noisy for the three outcome groups. Nevertheless, the point estimates for the persuaded are larger: the persuaded students have higher English ability than the other two groups.

We now present the results for profiling compliers using the preference variables. We look at seven preference variables collected by Chen and Yang [2019], namely, risk attitude, patience index, whether you will punish others or not when you are treated unfairly, whether you will punish others or not when others are treated unfairly, altruism, reciprocity, and belief that people are good. The results are presented in Table 2.18.

In terms of the risk measure, the always-persuaded students had the highest risk preferences, whereas the risk preferences of never-persuaded and persuaded students were quite similar.

Regarding the patience measure, the always-persuaded students exhibited the highest level of patience, while the patience measures for the never-persuaded and persuaded students were similar.

When it came to the question of whether one would punish those who treated them unfairly, our findings indicated that both always-persuaded and never-persuaded students were more inclined to punish those who treated them unfairly compared to persuaded students.

Regarding the question of whether individuals would punish others when other people were treated unfairly, all three groups demonstrated an equal willingness to punish in such scenarios.

Additionally, our study revealed that the three persuasion types have similar measures for altruism and reciprocity measures.

Finally, our results show that the always-persuaded students were more inclined to believe that people are good, compared with the never-persuaded and persuaded. Furthermore, the never-persuaded and persuaded students shared a similar inclination to believe in the inherent goodness of people.

We implement a joint test for testing the identification assumptions in the binary IV model of persuasion. We use the subsampling method in Bai et al. [2022a] for this test. The subsampling test in Bai et al. [2022a] requires us to pick a size for the subsample with $b_n \to \infty$ and $\frac{b_n}{n} \to 0$ as $n \to \infty$. We set b_n to $n^{\frac{2}{3}}$ here. The results in Figure 2.3 show that we cannot reject the validity of the identification assumptions at the 5% level.

Finally, we provide a comparison between the local persuasion rate proposed by Jun and Lee [2023] and the persuasion measure proposed by DellaVigna and Kaplan [2007]. The results are presented in Table 2.19. The results show a discrepancy between the two persuasion measures, although the difference is not as large as the differences in the other two applications.

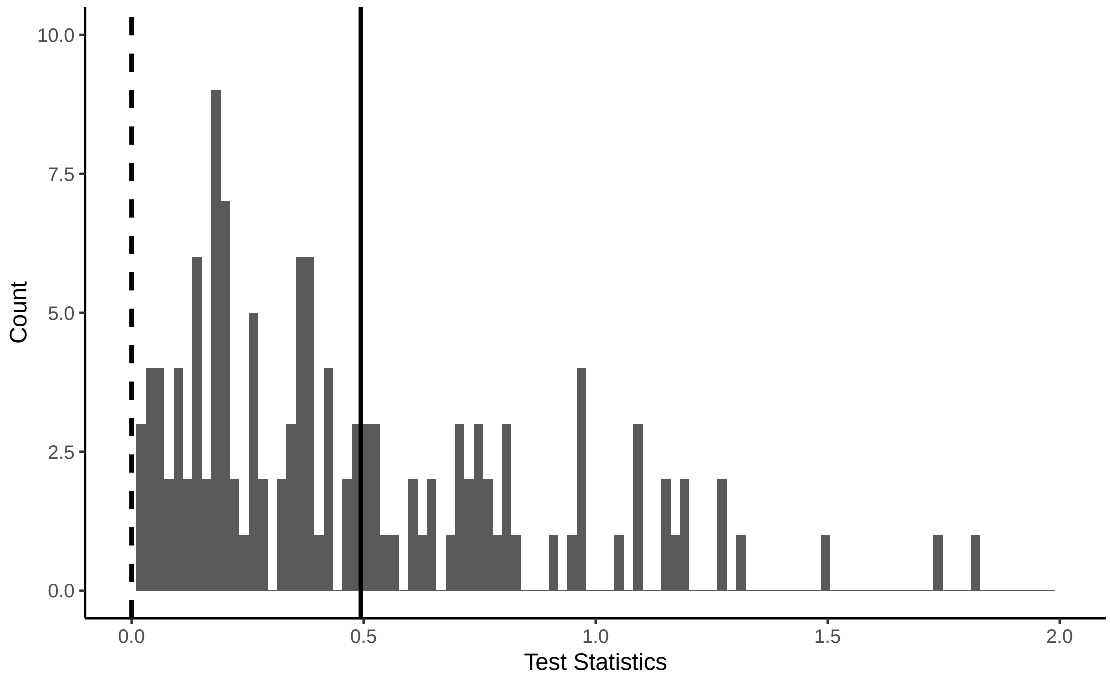


Figure 2.3: Chen and Yang [2019]: Testing Identification Assumptions

The figure presents the results for testing identification assumptions in Chen and Yang [2019]. The solid line is the critical value for a 5% level test. The dashed line is the test statistic.

2.5 Conclusion

We applied the methods developed in the first chapter to three empirical examples [Enikolopov et al., 2011, Blattman and Annan, 2016, Chen and Yang, 2019]. The results illustrate the usefulness of the methods. Re-analyzing Enikolopov et al. [2011] informed us that most of the voters were persuaded, and the persuaded voters were likely to be middle-aged and male. Re-analyzing Blattman and Annan [2016] informed us that around 20% of the Liberian ex-fighters were persuaded, and the persuaded ex-fighters were more likely to be risky type. Re-analyzing Chen and Yang [2019] informed us that roughly 20% of the students were persuaded, and the persuaded students were likely to come from wealthy families, come from coastal areas, less risk-loving, and less likely to believe in the inherent goodness of people.

Table 2.7: Blattman and Annan [2016]: Joint Distribution of Potential Outcomes for Agricultural Work

	Share
AP	0.703
	(0.029)
NP	0.137
	(0.035)
P	0.160
	(0.020)

This table presents the estimated joint distribution of potential outcomes among compliers. All regressions include block dummies for block randomization. Standard errors are presented in the parentheses. AP refers to the following event: $[Y_i(1) = Y_i(0) = 1]$. NP refers to the following event: $[Y_i(1) = Y_i(0) = 0]$. P refers to the following event: $[1 = Y_i(1) > Y_i(0) = 0]$.

Table 2.8: Blattman and Annan [2016]: Joint Distribution of Potential Outcomes for Legal Activities

	Share
AP	0.614
	(0.030)
NP	0.038
	(0.038)
P	0.348
	(0.025)

This table presents the estimated joint distribution of potential outcomes among compliers. All regressions include block dummies for block randomization. Standard errors are presented in the parentheses. AP refers to the following event: $[Y_i(1) = Y_i(0) = 1]$. NP refers to the following event: $[Y_i(1) = Y_i(0) = 0]$. P refers to the following event: $[1 = Y_i(1) > Y_i(0) = 0]$.

Table 2.9: Blattman and Annan [2016]: Profiling Demographic Variables among Compliers for Agricultural Work

	AP	NP	P	Summary
Number of Children	2.975	1.640	-0.940	2.263
	(0.169)	(0.250)	(1.453)	(2.072)
Education	5.890	5.215	4.624	5.632
	(0.268)	(0.539)	(1.913)	(3.831)
Age	31.486	27.313	23.913	29.790
	(0.591)	(0.970)	(4.480)	(7.704)

This table presents the results profiling the compliers with demographic variables. All regressions include block dummies for block randomization. For the first three columns, standard errors are presented in the parentheses. The last column presents the sample averages of the pre-treatment covariates, and standard deviations are presented in parentheses. AP refers to the following event: $[Y_i(1) = Y_i(0) = 1]$. NP refers to the following event: $[Y_i(1) = Y_i(0) = 0]$. P refers to the following event: $[1 = Y_i(1) > Y_i(0) = 0]$.

Table 2.10: Blattman and Annan [2016]: Profiling Demographic Variables among Compliers for Legal Activities

	AP	NP	P	Summary
Number of Children	2.836	1.991	-5.452	2.263
	(0.171)	(0.157)	(8.771)	(2.072)
Education	5.956	5.288	3.418	5.632
	(0.299)	(0.325)	(6.940)	(3.831)
Age	30.885	28.948	20.623	29.790
	(0.603)	(0.648)	(16.551)	(7.704)

This table presents the results profiling the compliers with demographic variables. All regressions include block dummies for block randomization. For the first three columns, standard errors are presented in the parentheses. The last column presents the sample averages of the pre-treatment covariates, and standard deviations are presented in parentheses. AP refers to the following event: $[Y_i(1) = Y_i(0) = 1]$. NP refers to the following event: $[Y_i(1) = Y_i(0) = 0]$. P refers to the following event: $[1 = Y_i(1) > Y_i(0) = 0]$.

Table 2.11: Blattman and Annan [2016]: Profiling Variables that Measure the Level of Violence among Compliers for Agricultural Work

	AP	NP	P	Summary
Closeness to Former Commanders	0.140	-0.010	-1.064	-0.021
	(0.080)	(0.147)	(0.574)	(0.970)
Was a Combatant	0.767	0.692	0.780	0.670
	(0.033)	(0.061)	(0.222)	(0.471)
Aggression Index	1.178	1.606	1.591	1.338
	(0.123)	(0.278)	(0.918)	(1.940)
Ease of Reintegration	0.090	-0.322	-0.434	-0.042
	(0.074)	(0.098)	(0.517)	(0.976)
Patience Index	2.985	2.754	3.106	2.985
	(0.060)	(0.136)	(0.441)	(0.872)
Risk Index	0.314	0.387	0.421	0.329
	(0.043)	(0.091)	(0.300)	(0.587)

This table presents the results profiling the compliers with variables that measure individuals' levels of violence. All regressions include block dummies for block randomization. For the first three columns, standard errors are presented in the parentheses. The last column presents the sample averages of the pre-treatment covariates, and standard deviations are presented in parentheses. AP refers to the following event: $[Y_i(1) = Y_i(0) = 1]$. NP refers to the following event: $[Y_i(1) = Y_i(0) = 0]$. P refers to the following event: $[1 = Y_i(1) > Y_i(0) = 0]$.

Table 2.12: Blattman and Annan [2016]: Profiling Variables that Measure the Level of Violence among Compliers for Legal Activities

	AP	NP	P	Summary
Closeness to Former Commanders	0.137	-0.069	-2.910	-0.021
	(0.085)	(0.081)	(3.403)	(0.970)
Was a Combatant	0.737	0.739	1.120	0.670
	(0.037)	(0.038)	(0.872)	(0.471)
Aggression Index	1.176	1.419	2.253	1.338
	(0.134)	(0.182)	(3.169)	(1.940)
Ease of Reintegration	0.094	-0.164	-1.267	-0.042
	(0.080)	(0.081)	(2.075)	(0.976)
Patience Index	2.947	2.842	4.381	2.985
	(0.064)	(0.078)	(2.008)	(0.872)
Risk Index	0.308	0.344	0.777	0.329
	(0.048)	(0.056)	(1.091)	(0.587)

This table presents the results profiling the compliers with variables that measure individuals' levels of violence. All regressions include block dummies for block randomization. For the first three columns, standard errors are presented in the parentheses. The last column presents the sample averages of the pre-treatment covariates, and standard deviations are presented in parentheses. AP refers to the following event: $[Y_i(1) = Y_i(0) = 1]$. NP refers to the following event: $[Y_i(1) = Y_i(0) = 0]$. P refers to the following event: $[1 = Y_i(1) > Y_i(0) = 0]$.

Table 2.13: Blattman and Annan [2016]: Comparison Between θ_{local} and θ_{DK} for Agricultural Work

	Persuasion Rate Measures
DK Persuasion Rate	2.349
	(0.854)
Local Persuasion Rate	0.459
	(0.084)

This table presents the results of comparing θ_{local} and θ_{DK} when the outcome variable is agricultural work. Standard errors are presented in the parentheses.

Table 2.14: Blattman and Annan [2016]: Comparison Between θ_{local} and θ_{DK} for Legal Activities

	Persuasion Rate Measures
DK Persuasion Rate	6.918
	(67.334)
Local Persuasion Rate	0.147
	(0.098)

This table presents the results of comparing θ_{local} and θ_{DK} when the outcome variable is legal activities. Standard errors are presented in the parentheses.

Table 2.15: Chen and Yang [2019]: Summary Statistics

	N	Mean	Standard Deviation	Min	Max
Trust Foreign Media	886	0.644	0.479	0	1
Treatment	886	0.275	0.447	0	1
IV	886	0.553	0.497	0	1
CCP	886	0.068	0.251	0	1
House Income	886	136,168	175,178	5,000	1,050,000
Han	886	0.910	0.287	0	1
Hukou	886	0.228	0.420	0	1
Male	865	0.534	0.499	0	1
Siblings	886	0.572	1.121	0	9
Born Costal	886	0.394	0.489	0	1
Live Costal	886	0.409	0.492	0	1
English Ability	886	-0.051	0.960	-0.891	2.235
Risk	886	5.589	1.918	0	10
Patience	886	5.992	2.184	0	10
Punish when Unfair (Self)	886	5.472	2.421	0	10
Punish when Unfair (Others)	886	4.519	2.298	0	10
Altruism	886	6.924	2.219	0	10
Reciprocity	886	8.909	1.290	0	10
People Are Good	886	5.959	2.694	0	10

This table presents summary statistics for the sample used in our empirical analysis.

Table 2.16: Chen and Yang [2019]: Joint Distribution of Potential Outcomes for Trust in Foreign Media

	Share
AP	0.239
	(0.061)
NP	0.552
	(0.067)
P	0.209
	(0.029)

This table presents the estimated joint distribution of potential outcomes among compliers. Standard errors are presented in the parentheses. AP refers to the following event: $[Y_i(1) = Y_i(0) = 1]$. NP refers to the following event: $[Y_i(1) = Y_i(0) = 0]$. P refers to the following event: $[1 = Y_i(1) > Y_i(0) = 0]$.

Table 2.17: Chen and Yang [2019]: Profiling Demographic Variables among Compliers for Trust in Foreign Media

	AP	NP	P	Summary
CCP	-0.011	0.137	0.101	0.068
	(0.091)	(0.054)	(0.048)	(0.251)
House Income	150,560.393	98,039.216	136,803.978	136,168.172
	(63,244.897)	(17,534.713)	(33,530.237)	(175,178.169)
Han	0.837	0.902	0.959	0.910
	(0.107)	(0.046)	(0.055)	(0.287)
Hukou	0.209	0.216	0.199	0.228
	(0.154)	(0.064)	(0.081)	(0.420)
Male	0.646	0.490	0.487	0.534
	(0.193)	(0.079)	(0.098)	(0.499)
Siblings	0.426	0.549	0.402	0.572
	(0.464)	(0.136)	(0.214)	(1.121)
Born Costal	0.143	0.373	0.465	0.394
	(0.197)	(0.075)	(0.097)	(0.489)
Live Costal	0.095	0.392	0.515	0.409
	(0.205)	(0.076)	(0.098)	(0.492)
English Ability	-0.293	-0.016	0.113	-0.051
	(0.377)	(0.119)	(0.196)	(0.960)

This table presents the results for profiling the persuasion types among compliers. Standard errors are presented in the parentheses. AP refers to the following event: $[Y_i(1) = Y_i(0) = 1]$. NP refers to the following event: $[Y_i(1) = Y_i(0) = 0]$. P refers to the following event: $[1 = Y_i(1) > Y_i(0) = 0]$.

Table 2.18: Chen and Yang [2019]: Profiling Preference Variables among Compliers for Trust in Foreign Media

	AP	NP	P	Summary
Risk	6.201	5.667	5.319	5.589
	(0.747)	(0.292)	(0.379)	(1.918)
Patience	6.454	5.725	5.856	5.992
	(0.797)	(0.324)	(0.408)	(2.184)
Punish when Unfair (Self)	6.010	6.275	4.904	5.472
	(0.936)	(0.337)	(0.482)	(2.421)
Punish when Unfair (Others)	4.348	4.392	4.577	4.519
	(0.850)	(0.298)	(0.447)	(2.298)
Altruism	6.941	6.412	6.654	6.924
	(0.798)	(0.342)	(0.437)	(2.219)
Reciprocity	8.994	8.843	8.713	8.909
	(0.507)	(0.176)	(0.263)	(1.290)
People Are Good	7.126	5.784	5.661	5.959
	(1.023)	(0.408)	(0.520)	(2.694)

This table presents the results for profiling the persuasion types among compliers. Standard errors are presented in the parentheses. AP refers to the following event: $[Y_i(1) = Y_i(0) = 1]$. NP refers to the following event: $[Y_i(1) = Y_i(0) = 0]$. P refers to the following event: $[1 = Y_i(1) > Y_i(0) = 0]$.

Table 2.19: Chen and Yang [2019]: Comparison Between θ_{local} and θ_{DK}

	Persuasion Rate Measures
DK Persuasion Rate	0.919
	(0.085)
Local Persuasion Rate	0.725
	(0.044)

This table presents the results of comparing θ_{local} and θ_{DK}. Standard errors are presented in the parentheses.

CHAPTER 3

WHEN IS TRIPLE DIFFERENCE SENSITIVE TO FUNCTIONAL FORM?

3.1 Introduction

This paper studies when the "modified" parallel trends assumption necessary for the identification of the average treatment effect on the treated (ATT) in a triple difference design is insensitive to functional form. Studying this property is worthwhile for at least two reasons. The triple difference design has become increasingly popular in economics in recent years [Olden and Møen, 2022]. Moreover, it is often not clear from theory that the "modified" parallel trends should hold for a particular choice of functional form. For example, an empirical researcher may be interested in the ATT in levels for a particular treatment that is of economic relevance. However, it is not obvious that the treatment at the state-level generates parallel trends specifically in levels rather than in logs or some other transformation. The triple difference design thus will be more credible if its validity does not depend on a particular functional form. In this paper, we provide precise conditions under which the triple difference is robust to functional form, we also suggest that researchers should be careful when they give a justification specific for a functional form when the conditions are not plausible.

We present two characterizations of when the "modified" parallel trends assumption is insensitive to functional form, meaning that it holds for all measurable transformations of the outcome variable. Firstly, we show that the "modified" parallel trends assumption is insensitive to functional form if and only if a corresponding condition holds for the entire cumulative distribution function of the untreated potential outcomes. Secondly, we show that if the distribution of the untreated potential outcome can be decomposed into two group-specific trends and a group-specific stationary part, then, the "modified" parallel trends assumption is robust to functional form.

These conditions for triple difference being invariant to transformations have testable implications. The testable implications explore the fact that the distribution of the untreated potential outcomes in the treated group and the treated period can be identified if the triple difference design is insensitive to functional form. Therefore, if the triple difference design is insensitive to functional form, the identified (or, implied) density, should be non-negative almost everywhere. The statistical tests proposed in this paper can be useful in terms of warning researchers when they should be particularly careful about justifying the "modified" parallel trends assumption for some specific functional forms that they chose for their analysis. We illustrate the tests using Muralidharan and Prakash [2017], who use a triple difference design to study the effect of a bicycle program for girls in India's Bihar state on their academic performance.

Prior research has noted and shown that the parallel trends assumption in DID may hold in levels but not logs or vice versa [Meyer, 1995, Kahn-Lang and Lang, 2020, Athey and Imbens, 2006, Roth and Sant'Anna, 2023]. However, to the best of our knowledge, we extend their insights in DID to triple difference and provide the first full characterization of when the triple difference is insensitive to functional form. The conditions we prove are distinct from the assumptions needed for identifying distributional treatment effects in DID setting [Athey and Imbens, 2006, Bonhomme and Sauder, 2011, Callaway and Li, 2019].

The remainder of the paper is organized as follows. In Section 2, we set up a canonical triple difference model. We provide the two characterizations of when triple difference is insensitive to functional form in Section 3. Section 4 provides a small simulation to illustrate the characterizations in Section 3. Section 5 discusses why we cannot directly apply Roth and Sant'Anna [2023] to an equivalent form of the triple difference estimand, that is, a DID with the transformed outcomes. Section 6 provides comparisons with distributional DID.

Section 7 provides a sharp testable implication from the triple difference being invariant to functional form. Section 8 provides an empirical application using data from Muralidharan and Prakash [2017]. The last section provides a concluding discussion on common empirical settings in which our approach is particularly useful.

3.2 Setup

Consider the canonical triple difference design in Olden and Møen [2022]. Units are indexed by i. There are two periods $T_i \in \{0,1\}$. There are two group indicators: $G_i, H_i \in \{0,1\}$. An example of G_i can be rural and urban areas in the United States. An example of H_i can be two binary demographic groups, say, black and white [Aaronson and Mazumder, 2011]. The treated population, denoted by D_i, consists of those units for which both G_i and H_i are equal to 1:

$$D_i = G_i H_i \ . \tag{3.1}$$

Moreover, for the treated population, the treatment is assigned in the first period. All the remaining three groups are the control population.

The potential outcomes for unit i in period t is denoted by $Y_{it}(0)$, $Y_{it}(1)$. Given the stable unit treatment value (SUTVA) assumption, the observed outcome is:

$$Y_{it} = Y_{it}(1)D_i + Y_{it}(0)(1 - D_i) \ . \tag{3.2}$$

The target parameter with the triple difference design is the average treatment effect on the treated (ATT):

$$\begin{aligned}\tau_{ATT} &:= \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) \mid D_i = 1] \\ &= \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) \mid G_i = 1, H_i = 1] \ .\end{aligned}$$

We first assume that there are no anticipatory effects, that is, there is no treatment effect prior to the implementation of the treatment:

Assumption 3.2.1. (No Anticipatory Effects) $Y_{i0}(0) = Y_{i0}(1)$ for all i with $D_i = 1$.

Assumption 3.2.1 is a crucial but often hidden assumption for identifying τ_{ATT}. Without Assumption 3.2.1, the changes in the outcome for the treated population between period 0 and 1 may reflect not just the causal effect in period $T_i = 1$, but also an anticipatory effect in period $T_i = 0$ [Abbring and Van den Berg, 2003, Malani and Reif, 2015].

Olden and Møen [2022] points out that the other main identification assumption is the "modified" parallel trends assumption. The "modified" parallel trends assumption requires the relative outcome of group 1 (i.e., $H_i = 1$) and group 0 (i.e., $H_i = 0$) in the treatment state (i.e., $G_i = 1$) to trend in the same way as the relative outcome of group 1 and group 0 in the control state in the absence of treatment:

Assumption 3.2.2. ("Modified" Parallel Trends)

$$\begin{aligned}&(\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 1, H_i = 1]) - (\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 0, H_i = 1]) \\ &= (\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 1, H_i = 0]) - (\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 0, H_i = 0]) \ ,\end{aligned}$$

where the expectations above are finite.

Under the "modified" parallel trends assumption, we allow the parallel trends to be violated for the group $H_i = 1$. Since parallel trends violation is the same between $H_i = 0$

and $H_i = 1$, we can use the placebo group (i.e., $H_i = 0$) to de-bias the bias caused by the violation of parallel trends in the group $H_i = 1$. Therefore, under the no anticipatory effect and the "modified" parallel trends assumption, the ATT is identified by:

$$\begin{aligned}\tau_{\text{ATT}} = &(\mathbb{E}[Y_{i1} - Y_{i0} \mid G_i = 1, H_i = 1] - \mathbb{E}[Y_{i1} - Y_{i0} \mid G_i = 0, H_i = 1]) \\ &- (\mathbb{E}[Y_{i1} - Y_{i0} \mid G_i = 1, H_i = 0] - \mathbb{E}[Y_{i1} - Y_{i0} \mid G_i = 0, H_i = 0]) \ .\end{aligned}$$

3.3 Invariance of Parallel Trends

Following Athey and Imbens [2006] and Roth and Sant'Anna [2023], we say that the "modified" parallel trends assumption is invariant to transformations if the "modified" parallel trends assumption holds for all measurable transformations of the outcome.

Definition 3.3.1. We say that the "modified" parallel trends assumption is invariant to transformations (i.e., insensitive to functional form) if

$$\begin{aligned}&(\mathbb{E}[g(Y_{i1}(0)) - g(Y_{i0}(0)) \mid G_i = 1, H_i = 1]) - (\mathbb{E}[g(Y_{i1}(0)) - g(Y_{i0}(0)) \mid G_i = 0, H_i = 1]) \\ &= (\mathbb{E}[g(Y_{i1}(0)) - g(Y_{i0}(0)) \mid G_i = 1, H_i = 0]) - (\mathbb{E}[g(Y_{i1}(0)) - g(Y_{i0}(0)) \mid G_i = 0, H_i = 0]),\end{aligned}$$

for all measurable functions g such that the expectations above are finite.

Remark 3.3.1. Compared with the invariance criteria in Roth and Sant'Anna [2023], the class of functions in Definition 3.3.1 is larger, because Roth and Sant'Anna [2023] focus on the strictly monotone transformations. However, similar to the results in Roth and Sant'Anna [2023], the "modified" parallel trends assumption holds for all measurable transformations is in fact equivalent to the "modified" parallel trends assumption holds for all strictly monotonic transformations. ■

Proposition 3.3.1 provides a characterization when the "modified" parallel trends assumption is invariant to all measurable transformations on the outcome variables. Note that this is a direct extension of Proposition 3.1 in Roth and Sant'Anna [2023] to the triple difference setting.

Proposition 3.3.1. The "modified" parallel trends is invariant to transformations if and only if for all $y \in \mathbb{R}$,

$$
\begin{aligned}
&(F_{Y_{i1}(0)|G_i=1,H_i=1}(y) - F_{Y_{i0}(0)|G_i=1,H_i=1}(y)) - (F_{Y_{i1}(0)|G_i=0,H_i=1}(y) - F_{Y_{i0}(0)|G_i=0,H_i=1}(y)) \\
&= (F_{Y_{i1}(0)|G_i=1,H_i=0(y)} - F_{Y_{i0}(0)|G_i=1,H_i=0}(y)) - (F_{Y_{i1}(0)|G_i=0,H_i=0}(y) - F_{Y_{i0}(0)|G_i=0,H_i=0}(y)) \, .
\end{aligned}
\tag{3.3}
$$

Remark 3.3.2. The result shows that the "modified" parallel trends assumption is invariant to transformations if and only if a stronger "modified parallel trends"-type condition holds for the cumulative distribution functions of the untreated potential outcomes. ■

Remark 3.3.3. If the outcome variable is binary, $Y_i \in \{0, 1\}$. Then, Proposition 3.3.1 implies that whenever the "modified" parallel trends assumption holds, applying bijective functions on the outcome variables (i.e., replacing $\{0, 1\}$ with $\{a, b\}$, where $a \neq b$) will also satisfy the "modified" parallel trends assumption. This holds because the expectation of a binary outcome fully characterizes its distribution. ■

The following result provides a characterization of how distributions satisfying the invariance assumption (i.e., Equation 3.3) can be generated. The result below is not a necessary and sufficient characterization, as the necessary direction requires an additional assumption. Note that this is not a direct extension of Proposition 3.2 in Roth and Sant'Anna [2023] to the triple difference setting.

Proposition 3.3.2. Suppose that the distribution of $Y_{it}(0) \mid G_i = g, H_i = h$ for all $t, g, h \in \{0, 1\}$ have a Radon–Nikodym density with respect to a common dominating, positive σ-finite measure.

(i) The "modified" parallel trends is invariant to transformations if

$$F_{Y_{it}(0)|G_i=g,H_i=h} = \alpha F_t^g + \beta F_t^h + \eta F^{gh} \ ,$$

where $\alpha, \beta, \eta \in [0,1]$, $\alpha + \beta + \eta = 1$, F_t^g is any valid distribution that depends on g and t, F_t^h is any valid distribution that depends on h and t, and F^{gh} is any valid distribution that depends on g and h.

(ii) Suppose the "modified" parallel trends is invariant to transformations, and $F_{Y_{it}(0)|G_i=g,H_i=h}$ can be decomposed as

$$F_{Y_{it}(0)|G_i=g,H_i=h} = \sum_{k=1}^{K} \theta_k J_k \ ,$$

where $\theta_k \in [0,1]$ for $k = 1, ..., K$, $\sum_{k=1}^K \theta_k = 1$, $\{J_k\}_{k=1}^K$ are valid CDFs, and none of $\{J_k\}_{k=1}^K$ depends on t, g, and h simultaneously. Then, $\{J_k\}_{k=1}^K$ must include the following three components: (1) F_t^g, a valid distribution that depends on t and g; (2) F_t^h, a valid distribution that depends on t and h; (3) F^{gh}, a valid distribution that depends on g and h.

Remark 3.3.4. If F_t^g is included in the decomposition, then F_t or F^g is redundant. This is because if $F_{Y_{it}(0)|G_i=g,H_i=h}$ can be decomposed as

$$F_{Y_{it}(0)|G_i=g,H_i=h} = \theta_1 F_t^g + \theta_2 F_t + \theta_3 F^g + \sum_{k=4}^{K} \theta_k J_k \ ,$$

where $\theta_k \in [0,1]$ for $k = 1, ..., K$, and $\theta_1 > 0$, then it can be rewritten as:

$$\begin{aligned}
F_{Y_{it}(0)|G_i=g,H_i=h} &= \theta_1 F_t^g + \theta_2 F_t + \theta_3 F^g + \sum_{k=4}^{K} \theta_k J_k \\
&= (\theta_1 + \theta_2 + \theta_3)\left(\frac{\theta_1}{\theta_1 + \theta_2 + \theta_3} F_t^g + \frac{\theta_2}{\theta_1 + \theta_2 + \theta_3} F_t + \frac{\theta_3}{\theta_1 + \theta_2 + \theta_3} F^g\right) \\
&\quad + \sum_{k=3}^{K} \theta_k J_k \\
&= (\theta_1 + \theta_2 + \theta_3) H_t^g + \sum_{k=3}^{K} \theta_k J_k \ ,
\end{aligned}$$

where $H_t^g \equiv \frac{\theta_1}{\theta_1+\theta_2+\theta_3} F_t^g + \frac{\theta_2}{\theta_1+\theta_2+\theta_3} F_t + \frac{\theta_3}{\theta_1+\theta_2+\theta_3} F^g$, which is another valid CDF depends on t and g. By a similar argument, we can show that F_t or F^h is redundant if we include F_t^h in the decomposition. ■

Remark 3.3.5. Consider the following CDFs: F_t^k and F_t^l, where $k = 1$ if either $g = 1, h = 1$ or $g = 0, h = 0$, and $k = 0$ otherwise, and $l = 1$ if $g = g'$ and $h = h'$ for some $g', h' \in \{0, 1\}$ and $l = 0$ otherwise. Both CDFs cannot be included in the decomposition. The reason is that doing so will make the "modified" parallel trends assumption not invariant to functional form. ■

Remark 3.3.6. If $F_{Y_{it}(0)|G_i=g,H_i=h}$ can be decomposed as

$$F_{Y_{it}(0)|G_i=g,H_i=h} = \sum_{k=1}^{K} \theta_k J_k \ ,$$

where $\theta_k \in [0,1]$ for $k = 1, ..., K$, $\sum_{k=1}^{K} \theta_k = 1$, $\{J_k\}_{k=1}^{K}$ are valid CDFs, and none of $\{J_k\}_{k=1}^{K}$ depends on t, g, and h simultaneously, then θ_k cannot depend on any of t, g, or h. Otherwise, "modified" parallel trends is not invariant to functional form. ■

3.4 A Simulation for the Decomposition Results

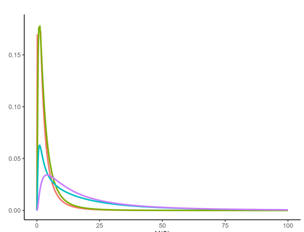
Based on the decomposition results in Proposition 3.3.2, we conduct a simulation in which the untreated potential outcome is a mixture of log normal distributions:

$$F_{Y_{it}(0)|G_i=g,H_i=h} = \alpha F_t^g + \beta F_t^h + \eta F^{gh} \ ,$$

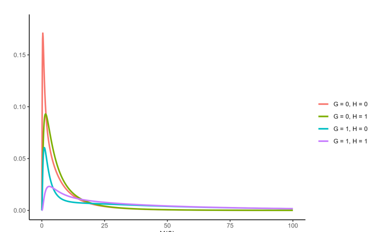
where $\alpha = \frac{1}{2}$ and $\beta = \eta = \frac{1}{4}$, and the distributions of F_t^g, F_t^h, and F^{gh} are:

$$\begin{aligned} F_t^g &\sim \text{lognormal}(1 + t + 2 \times g, 1) \ , \\ F_t^h &\sim \text{lognormal}(10 + 2 \times t + 3 \times h, 1) \ , \\ F^{gh} &\sim \text{lognormal}(g + h, 1) \ . \end{aligned}$$

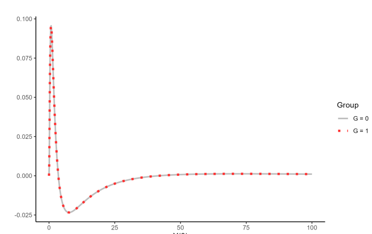
The simulation results are displayed in Figure 3.1. As we can see in the figure, the distributions of the untreated potential outcomes for the four groups differ from each other in both pre-treatment and post-treatment periods. Moreover, the distributions of the untreated potential outcomes change over time. However, the triple difference of the PDFs is the same. Therefore, our results imply that the "modified" parallel trends hold for all measurable transformations of the outcome.



(a) Pre-Treatment Period



(b) Post-Treatment Period



(c) DDD of CDF of Untreated Potential Outcomes

Figure 3.1: Illustration: DDD Invariance of Mixtures of Distributions

3.5 Triple Difference as Difference in Differences

As pointed out by Olden and Møen [2022], the triple difference estimator can be written as a difference in differences (DID) estimator with a differential as the new outcome variable. We thus can apply the same idea to the triple difference identification assumption. The "modified" parallel trends (i.e., Assumption 3.2.2) can be rewritten as a parallel trends assumption with a differential as the new outcome variable. To see this, first observe that Assumption 3.2.2 can be re-written as:

$$\begin{aligned}
&(\mathbb{E}[Y_{i1}(0) \mid G_i = 1, H_i = 1] - \mathbb{E}[Y_{i1}(0) \mid G_i = 1, H_i = 0]) \\
&- (\mathbb{E}[Y_{i0}(0) \mid G_i = 1, H_i = 1] - \mathbb{E}[Y_{i0}(0) \mid G_i = 1, H_i = 0]) \\
&= (\mathbb{E}[Y_{i1}(0) \mid G_i = 0, H_i = 1] - \mathbb{E}[Y_{i1}(0) \mid G_i = 0, H_i = 0]) \\
&- (\mathbb{E}[Y_{i0}(0) \mid G_i = 0, H_i = 1] - \mathbb{E}[Y_{i0}(0) \mid G_i = 0, H_i = 0]) \ .
\end{aligned} \tag{3.4}$$

Furthermore, some algebraic manipulations yield:

$$\mathbb{E}[Y_{it}(0) \mid G_i = g, H_i = h] = \mathbb{E}\left[\frac{Y_{it}(0)\mathbb{1}\{H_i = h\}}{\mathbb{P}[H_i = h \mid G_i = g]} \mid G_i = g\right] \ .$$

Therefore, we can re-write Equation 3.4 as:

$$\begin{aligned}
&\mathbb{E}\left[\frac{Y_{i1}(0)\mathbb{1}\{H_i = 1\}}{\mathbb{P}[H_i = 1 \mid G_i = 1]} - \frac{Y_{i1}(0)\mathbb{1}\{H_i = 0\}}{\mathbb{P}[H_i = 0 \mid G_i = 1]} \mid G_i = 1\right] \\
&- \mathbb{E}\left[\frac{Y_{i0}(0)\mathbb{1}\{H_i = 1\}}{\mathbb{P}[H_i = 1 \mid G_i = 1]} - \frac{Y_{i0}(0)\mathbb{1}\{H_i = 0\}}{\mathbb{P}[H_i = 0 \mid G_i = 1]} \mid G_i = 1\right] \\
&= \mathbb{E}\left[\frac{Y_{i1}(0)\mathbb{1}\{H_i = 1\}}{\mathbb{P}[H_i = 1 \mid G_i = 0]} - \frac{Y_{i1}(0)\mathbb{1}\{H_i = 0\}}{\mathbb{P}[H_i = 0 \mid G_i = 0]} \mid G_i = 0\right] \\
&- \mathbb{E}\left[\frac{Y_{i0}(0)\mathbb{1}\{H_i = 1\}}{\mathbb{P}[H_i = 1 \mid G_i = 0]} - \frac{Y_{i0}(0)\mathbb{1}\{H_i = 0\}}{\mathbb{P}[H_i = 0 \mid G_i = 0]} \mid G_i = 0\right] \ .
\end{aligned} \tag{3.5}$$

Equation 3.5 is the "modified" parallel trends assumption written in the parallel trends assumption for a new outcome variable, the difference in weighted outcomes between two groups (i.e., between $H_i = 0$ and $H_i = 1$).

One significance of Equation 3.5 is that our results do not directly follow from applying the results in Roth and Sant'Anna [2023] by using a DID with differential outcomes. The reason is that our invariance definition only requires transforms on the untreated potential outcomes, namely, $g(Y_{it}(0))$, $t \in \{0, 1\}$, while the invariance with respect to the transformed outcome requires transforms on the new outcome variables, namely, $g\left(\frac{Y_{it}(0)\mathbb{1}\{H_i=1\}}{\mathbb{P}[H_i=1|G_i=g]} - \frac{Y_{it}(0)\mathbb{1}\{H_i=0\}}{\mathbb{P}[H_i=0|G_i=g]}\right)$, $t \in \{0, 1\}$. Hence, there is no clear economic interpretation of this transformed outcome.

There is a special case in which Equation 3.5 can be collapsed to a simple DID at G_i with unweighted differential outcomes. Let us consider the case when $\mathbb{P}[H_i = h \mid G_i = g] = \frac{1}{2}$, with $g, h \in \{0, 1\}$. An example of such a case is when G_i is a state variable and H_i is a sex variable and there is a strong gender balance within the two states. In this case, Equation 3.5 becomes:

$$\begin{aligned}
&\mathbb{E}[Y_{i1}(0)(\mathbb{1}\{H_i = 1\} - \mathbb{1}\{H_i = 0\}) \mid G_i = 1] - \mathbb{E}[Y_{i0}(0)(\mathbb{1}\{H_i = 1\} - \mathbb{1}\{H_i = 0\}) \mid G_i = 1] \\
&= \mathbb{E}[Y_{i1}(0)(\mathbb{1}\{H_i = 1\} - \mathbb{1}\{H_i = 0\}) \mid G_i = 1] - \mathbb{E}[Y_{i0}(0)(\mathbb{1}\{H_i = 1\} - \mathbb{1}\{H_i = 0\}) \mid G_i = 0].
\end{aligned} \tag{3.6}$$

In other words, we can view Equation 3.6 as parallel trends at the G_i level with the untreated outcome being a simple difference of the untreated potential outcomes between group $H_i = 1$ and $H_i = 0$. In this case, our results still do not directly follow from Roth and Sant'Anna [2023]. To see this, applying the results in Roth and Sant'Anna [2023] requires us to define the invariance for the transformed outcome, namely, $g(Y_{it}(0)(\mathbb{1}\{H_i = 1\} - \mathbb{1}\{H_i = 0\}))$, $t \in \{0, 1\}$. Hence, there is still no clear economic interpretation of this new outcome variable.

3.6 Comparison with Distributional DID

An immediate consequence of Proposition 3.3.1 is that the cumulative distribution function of the untreated outcomes conditional on the treated group in the treated period is identifiable:

$$\begin{aligned}
&F_{Y_{i1}(0)|G_i=1,H_i=1}(y) \\
&= F_{Y_{i0}(0)|G_i=1,H_i=1}(y) + (F_{Y_{i1}(0)|G_i=1,H_i=0}(y) - F_{Y_{i0}(0)|G_i=1,H_i=0}(y)) \\
&\quad + (F_{Y_{i1}(0)|G_i=0,H_i=1}(y) - F_{Y_{i0}(0)|G_i=0,H_i=1}(y)) \\
&\quad - (F_{Y_{i1}(0)|G_i=0,H_i=0}(y) - F_{Y_{i0}(0)|G_i=0,H_i=0}(y)) \ .
\end{aligned} \tag{3.7}$$

Therefore, the quantile treatment effect on the treated is identified:

$$\text{QTT}(\tau) = F^{-1}_{Y_{i1}(1)|G_i=1,H_i=1}(\tau) - F^{-1}_{Y_{i1}(0)|G_i=1,H_i=1}(\tau) \ .$$

We can also adapt distributional DID models [Athey and Imbens, 2006, Bonhomme and Sauder, 2011, Callaway and Li, 2019] to identify $\text{QTT}(\tau)$ in the triple difference setting. However, this is beyond the scope of the paper, hence, we leave a formal analysis of distributional triple difference to future research. Nevertheless, we provide some comparisons between Equation 3.7 and other distributional DID models.

Remark 3.6.1 (Relationship to Athey and Imbens [2006])**.** The first approach in distributional DID is to invoke the rank invariance assumption [Athey and Imbens, 2006]. The rank invariance assumption implies that the untreated potential outcome for the treated group in the first period can be identified by mapping between the quantiles of the untreated potential outcomes for the treated and control groups. However, Equation 3.7 does not restrict the dependence of the untreated potential outcomes between the treatment and control groups. Hence, these two approaches are non-nested. ■

Remark 3.6.2 (Relationship to Bonhomme and Sauder [2011])**.** The distributional DID model in Bonhomme and Sauder [2011] implies a “parallel trend” condition for the log of the characteristic function:

$$\ln\left(\Psi^{Y_{i1}(0)}_{D_i=1}(s)\right) - \ln\left(\Psi^{Y_{i0}(0)}_{D_i=1}(s)\right) = \ln\left(\Psi^{Y_{i1}(0)}_{D_i=0}(s)\right) - \ln\left(\Psi^{Y_{i0}(0)}_{D_i=0}(s)\right) ,$$

where $\Psi^{Y_{it}(0)}_{D_i=d}(s)$ is the characteristic function of $Y_{it}(0) \mid D_i = d$, with $d, t \in \{0, 1\}$. Instead, Equation 3.3 implies a “modified” parallel trend condition for the characteristic function:

$$\begin{aligned}&\left(\Psi^{Y_{i1}(0)}_{G_i=1,H_i=1}(s) - \Psi^{Y_{i0}(0)}_{G_i=1,H_i=1}(s)\right) - \left(\Psi^{Y_{i1}(0)}_{G_i=0,H_i=1}(s) - \Psi^{Y_{i0}(0)}_{G_i=0,H_i=1}(s)\right)\\ &= \left(\Psi^{Y_{i1}(0)}_{G_i=1,H_i=0}(s) - \Psi^{Y_{i0}(0)}_{G_i=1,H_i=0}(s)\right) - \left(\Psi^{Y_{i1}(0)}_{G_i=0,H_i=0}(s) - \Psi^{Y_{i0}(0)}_{G_i=0,H_i=0}(s)\right) .\end{aligned}$$

Thus, the two conditions above are generally non-nested. ■

Remark 3.6.3 (Relationship to Callaway and Li [2019])**.** The distributional DID model in Callaway and Li [2019] makes two identification assumptions. First, $Y_{i1}(1) - Y_{i0}(0)$ is fully independent of the treatment assignment D_i. Second, within the treated population, the dependence between $Y_{i1}(1) - Y_{i0}(0)$ and $Y_{i0}(0)$ is identical to the dependence between $Y_{i0}(1) - Y_{i-1}(0)$ and $Y_{i-1}(0)$. However, Equation 3.7 does not restrict the dependence of the untreated potential outcome across periods conditional on the treated population. Hence, these two approaches are non-nested. ■

3.7 Testable Implications

Proposition 3.3.1 also implies the following condition for the probability distribution of the untreated potential outcome in the treated group and treated period:

$$
\begin{aligned}
&\mathbb{P}_{Y_{i1}(0)|G_i=1,H_i=1}(Y_{i1}(0) \in A) \\
&= \mathbb{P}_{Y_{i0}(0)|G_i=1,H_i=1}(Y_{i0}(0) \in A) \\
&\quad + (\mathbb{P}_{Y_{i1}(0)|G_i=1,H_i=0}(Y_{i1}(0) \in A) - \mathbb{P}_{Y_{i0}(0)|G_i=1,H_i=0}(Y_{i0}(0) \in A)) \\
&\quad + (\mathbb{P}_{Y_{i1}(0)|G_i=0,H_i=1}(Y_{i1}(0) \in A) - P\mathbb{P}Y_{i0}(0) \mid G_i = 0, H_i = 1(Y_{i0}(0) \in A)) \\
&\quad - (\mathbb{P}_{Y_{i1}(0)|G_i=0,H_i=0}(Y_{i1}(0) \in A) - \mathbb{P}_{Y_{i0}(0)|G_i=0,H_i=0})(Y_{i0}(0) \in A)) \ ,
\end{aligned}
\tag{3.8}
$$

where $A \in \mathcal{B}(\mathbb{R})$. Then, by the SUTVA assumption, the implied probability distribution for $Y_{i1}(0) \mid G_i = 1, H_i = 1$ is:

$$
\begin{aligned}
&\mathbb{P}^{\text{implied}}_{Y_{i1}(0)|G_i=1,H_i=1}(Y_{i1}(0) \in A) \\
&= \mathbb{P}(Y_{i0} \in A \mid G_i = 1, H_i = 1) \\
&\quad + (\mathbb{P}(Y_{i1} \in A \mid G_i = 1, H_i = 0) - \mathbb{P}(Y_{i0} \in A \mid G_i = 1, H_i = 0)) \\
&\quad + (\mathbb{P}(Y_{i1} \in A \mid G_i = 0, H_i = 1) - \mathbb{P}(Y_{i0} \in A \mid G_i = 0, H_i = 1)) \\
&\quad - (\mathbb{P}(Y_{i1} \in A \mid G_i = 0, H_i = 0) - \mathbb{P}(Y_{i0} \in A \mid G_i = 0, H_i = 0)) \ ,
\end{aligned}
\tag{3.9}
$$

where $A \in \mathcal{B}(\mathbb{R})$.

Note that for $\mathbb{P}^{\text{implied}}_{Y_{i1}(0)|G_i=1,H_i=1}(A)$, Equation 3.9 is guaranteed to be:

$$
\begin{aligned}
\mathbb{P}^{\text{implied}}_{Y_{i1}(0)|G_i=1,H_i=1}(\Omega) &= 1, \\
\mathbb{P}^{\text{implied}}_{Y_{i1}(0)|G_i=1,H_i=1}(\emptyset) &= 0,
\end{aligned}
$$

and countably additive. Therefore, if the "modified" parallel trends assumption is invariant to transformations, the implied probability density should be non-negative almost everywhere, i.e., $\mathbb{P}^{\text{implied}}_{Y_{i1}(0)|G_i=1,H_i=1}(Y_{i1}(0) \in A) \geq 0$, for all $A \in \mathcal{B}(\mathbb{R})$. This is a sharp testable implication. We formalize this in Proposition 3.7.1.

Proposition 3.7.1. (Sharp Characterization)

(i) If the "modified" parallel trend is invariant (i.e., Definition 3.3.1 holds), then:

$$\mathbb{P}^{\text{implied}}_{Y_{i1}(0)|G_i=1,H_i=1}(Y_{i1}(0) \in A) \geq 0,$$

for any $A \in \mathcal{B}(\mathbb{R})$.

(ii) For any joint distribution of observed data $(Y_{i0}, Y_{i1}, G_i, H_i)$ such that the implied density is non-negative almost everywhere:

$$\mathbb{P}^{\text{implied}}_{Y_{i1}(0)|G_i=1,H_i=1}(Y_{i1}(0) \in A) \geq 0, \quad \forall A \in \mathcal{B}(\mathbb{R}),$$

there exists a joint distribution of: $(Y_{i1}(1), Y_{i0}(1), Y_{i1}(0), Y_{i0}(0), G_i, H_i)$, such that it induces the observed data $(Y_{i0}, Y_{i1}, G_i, H_i)$, and the "modified" parallel trends is invariant to transformations (in Definition 3.3.1).

With a discrete outcome, the testable implication implies the following null hypothesis:

$$H_0 : -\mathbb{E}[\mathbb{1}\{Y_{i1}(0) = y\} \mid G_i = 1, H_i = 1] \leq 0,$$

for all $y \in \text{supp}(Y_i)$, where $\mathbb{E}[\mathbb{1}\{Y_{it}(0) = y\} \mid G_i = 1, H_i = 1]$ is defined by the right hand side of Equation 3.9. Note that this is a standard problem of testing moment inequalities. There are a variety of methods for testing such null hypothesis [Canay and Shaikh, 2017].

Furthermore, the test can be extended to the case with non-discrete outcome variables easily by possibly converting the testing problem to a high dimensional moment inequality testing problem [Bai et al., 2022b]. Specifically, when the outcome variable is not discrete, the null hypothesis becomes:

$$H_0 : -\mathbb{E}[\mathbb{1}\{Y_{i1}(0) \in A\} \mid G_i = 1, H_i = 1] \leq 0, \forall A \in \mathcal{B}(\mathbb{R}) \ ,$$

where $\mathbb{E}[\mathbb{1}\{Y_{i1}(0) \in A\} \mid G_i = 1, H_i = 1]$ is defined by Equation 3.9. To apply Bai et al. [2022b], let $\{A_j\}_{1 \leq j \leq p}$ be a partition of real line $\mathbb{R}$ with $A_j \in \mathcal{B}(\mathbb{R})$ for $1 \leq j \leq p$. We can pick a sequence of such partitions, with the number of Borel sets equal to p_n, where p_n grows to infinity with sample size under a suitable rate. Note that the regularity conditions in Bai et al. [2022b] work here because the support of $\hat{\mathbb{P}}_{Y_{i1}(0)|G_i=1,H_i=1}(Y_{i1}(0) \in A)$ is bounded uniformly in $P \in \mathbf{P}_n$.

3.8 Empirical Illustration

We conclude with an empirical illustration of our theoretical results using data from Muralidharan and Prakash [2017], who study the effect of the Cycle program on girls' education outcome using a triple difference design.

We now provide a brief description of the empirical setting and refer the reader to Muralidharan and Prakash [2017] for a more detailed description. The Government of Bihar of India launched the Chief Minister's Bicycle program (hereafter referred to as the Cycle program) in Bihar in 2006 to boost girls' education outcomes. The program provided girls who enrolled in secondary school with a free bicycle to reduce the transportation cost of going to school. The control state is Jharkhand, and the placebo group is boys. We focus on two continuous educational outcome variables on appearance and performance in the secondary school (SSC) certificate exams: log(number of candidates who appeared for the 10th grade exam) and log(number of candidates who passed the 10th grade exam).[1] Thus, the unit of analysis here is at the school level. Due to the two-year lag between enrollment in secondary school and the exam, the control years are from 2004 to 2007, while the treatment period covers 2009-2010.[2]

We use both the least favorable method in Canay and Shaikh [2017] and the two-step method in Romano et al. [2014] and Bai et al. [2022b] to test whether or not the "modified" parallel trends assumption is invariant to transforms. To implement both methods, we discretize the continuous outcome variables with equidistant bins ($p = 24$ in this example) and treat them as discrete outcomes. See Appendix B for more details about the implementation of the tests.

Figure 3.2 presents the implied density plots and the corresponding p-values using the least favorable method in Canay and Shaikh [2017]. The results show that we fail to reject the null hypothesis that the "modified" parallel trends assumption is invariant to transforms at 5% level.. Moreover, the two-step method in Romano et al. [2014] and Bai et al. [2022b] produces the same result, namely, we fail to reject the null hypothesis that the "modified" parallel trends assumption is invariant to transforms at 5% level.

3.9 Concluding Remarks

Our paper suggests the following different paths that empirical researchers can take for justifying the identification assumption when using a triple difference design. First, researchers can use contextual knowledge and economic theory to argue for the validity of the "modified" parallel trends assumption for a particular functional form. Second, researchers can

1. These are the outcome variables in Table 4 in Muralidharan and Prakash [2017].

2. Muralidharan and Prakash [2017] dropped the year 2008 in their analysis because they argue it's the transition year.

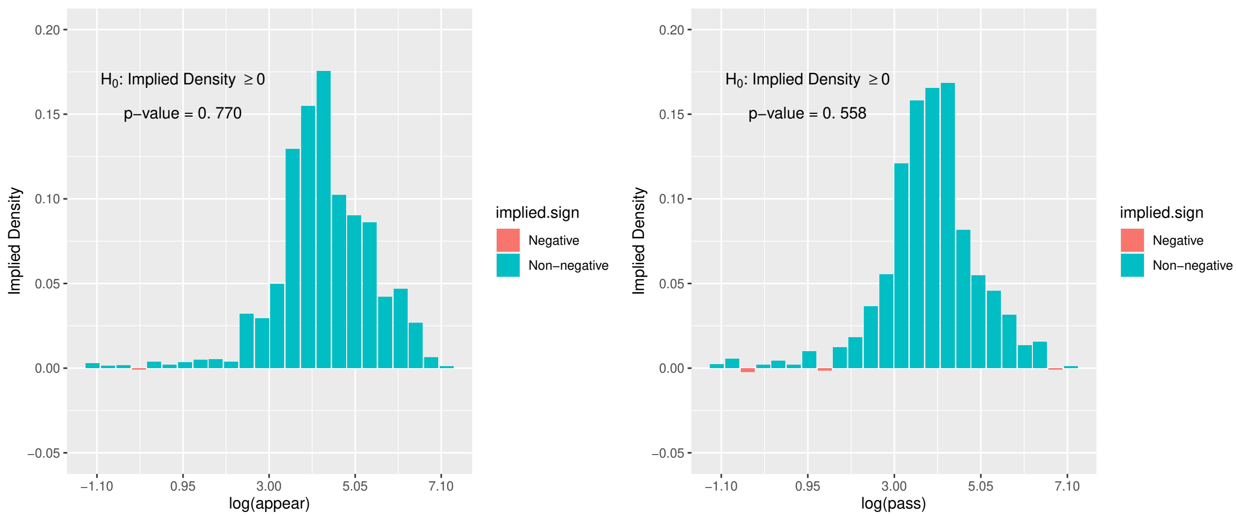


Figure 3.2: Implied Densities and p-Values using Least Favorable Test

impose additional distributional assumptions to point identify the distribution of $Y_{i1}(0)$ for the treated group. For example, imposing the "modified" parallel trends of CDFs ensures the validity of the triple difference estimator for all (measurable) transforms of the outcome variable. To conclude, we hope that the results in this paper will help applied researchers to have a more clear justification when they use triple difference design.

APPENDIX A

APPENDIX FOR CHAPTER 1

A.1 Identifiability of the Joint Distribution of Non-Binary Instruments or Outcomes

This section covers two potential directions for extending Lemma 1.4.2. The first direction explores the positive outcomes that arise from utilizing a non-binary instrument to extend Lemma 1.4.2. Following this, we delve into the negative outcomes associated with using a non-binary outcome to extend 1.4.2.

A.1.1 Non-Binary Instrument

Assumption 1.2.1 is adjusted to accommodate a discrete-valued instrument in two ways. Firstly, the IV monotonicity condition is crucially modified. With a discrete-valued instrument, the IV “monotonicity" condition must be satisfied for each pair of instruments. That is, changing the instrument from z to z' will either encourage or discourage every individual from taking up the treatment. Secondly, the IV relevance assumption is also revised. In this case, at least one instrument value must lead to changes in selection behavior. The formal statement of the revised assumption is now presented as Assumption A.1.1.

Assumption A.1.1. (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: $\mathbb{P}[T_i = 1|Z_i = z]$ is a non-trivial function of z,

5. IV Monotonicity: either $T_i(z) \geq T_i(z')$ or $T_i(z) \leq T_i(z')$ holds almost surely for $z \neq z'$ with $z, z' \in \text{supp}(Z_i)$.

With Assumption A.1.1, we can point identify the joint distribution of potential outcomes among each complier group. The intuition of the result is that with Assumption A.1.1, the proof proceeds "as-if" we are using a binary IV with support being $\{z, z'\}$. We now formally state the results in Corollary A.1.1.

Corollary A.1.1. Suppose Assumption A.1.1 holds, conditional on z, z' compliers (that is, $z, z' \in \text{supp}(Z_i)$ and $T_i(z) = T_i(z')$ does not hold almost surely), the joint distribution of potential outcome is point identified,:

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

Just as with a discrete-valued instrument, the identification assumptions will be modified for a continuous instrument. These modifications concern the IV monotonicity and IV relevance assumptions. In this case, we use an indicator selection equation to describe the first stage selection process. With this representation, it is easy to characterize the treatment effect on different margins of self-selecting into the treatment. We also assume that at least one instrument value leads to changes in the treatment-taking behavior. Assumption A.1.2 formally states the identification assumptions for this scenario.

Assumption A.1.2. (Binary Treatment and Outcome Model with a Continuous Instrument)

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

Before proceeding to present the identification results, we give two remarks related to Assumption A.1.2. First, the indicator selection equation is equivalent to the monotonicity condition in the IA IV model [Vytlacil, 2002]. To see this, observe that a change in z induces a shift either toward or away from treatment for the support of V_i. Second, instead of assuming $V_i \sim U[0, 1]$, we can also assume V_i being continuously distributed. This implies that we can normalize the distribution of V_i to be uniformly distributed over $[0, 1]$. A consequence of this normalization is that $\nu(z) = P(z)$, where $P(z)$ is the propensity score: $P(z) \equiv \mathbb{P}[T_i = 1 | Z_i = z]$.

Corollary A.1.2. Assume that Assumption A.1.2 holds, and assume $\text{supp}(P(Z_i)) = [0, 1]$, then, the joint distribution of potential outcomes at each margin of selecting into the treatment is identified:

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

A.1.2 Non-Binary Outcome

We now discuss whether we can extend the identification of the joint distribution of potential outcomes in Lemma 1.4.2 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.[1] Table A.1 presents the classification.

With the trinary outcome, two types of monotone treatment response assumptions were made in the previous literature. Jun and Lee [2023] 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

1. Jun and Lee [2023] does not use the conventional potential outcome notation in their discussion. Jun and Lee [2023] 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 [2023] uses the following notation: $Y_i(t) = (Y_{i0}(t), Y_{i1}(t), Y_{i,-1}(t))$, where $t \in \{0, 1\}$. $Y_{i0}(t)$ denotes whether the individual choose to take the action 0 if the treatment is t. $Y_{i1}(t)$ and $Y_{i,-1}(t)$ are defined similarly. Moreover, $\sum_{j \in S} Y_{ij}(t) = 1$ for $t \in \{0, 1\}$. That is, the choices in S are exclusive and exhaustive. It is easy to see that there is a duality between the notation in Jun and Lee [2023] and conventional potential outcome notation used in Table A.1.

Table A.1: 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

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 [2023], the seventh and eighth row (those with *) in Table A.1 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 [2023], 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 [2023].

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

Proposition A.1.1. 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 1.2.1 hold. Then, the joint distribution of potential outcomes among compliers is not point identified.

Even though we cannot point identify the joint distribution of potential outcomes among compliers in this case, We can still partially identify the joint distribution of potential outcomes among compliers using the approaches in Balke and Pearl [1997]. For example, to construct sharp bounds for $\mathbb{P}[Y_i(0) = -1, Y_i(1) = -1 | T_i(1) > T_i(0)]$, we can form a linear program with the objective function being $\mathbb{P}[Y_i(0) = -1, Y_i(1) = -1 | T_i(1) > T_i(0)]$ and the constraints being the linear system of equations in the proof of Proposition A.1.1.

One way to restore the point identification of the joint distribution of potential outcomes with non-binary Y_i under the monotone treatment response and IA IV assumptions is to binarize the outcome variable. To see this, assume without loss of generality that $Y_i(1) \geq Y_i(0)$ holds almost surely. Define the following two binary random variables: $\mathbb{1}\{Y_i(1) \geq x\}$ and $\mathbb{1}\{Y_i(0) \geq x\}$ with $x \in \mathbb{R}$. Then, by the monotone treatment response, it follows immediately that $\mathbb{1}\{Y_i(1) \geq x\} \geq \mathbb{1}\{Y_i(0) \geq x\}$ holds almost surely. Thus, the results in Lemma 1.4.2 hold for the new binarized outcome variable.

A.2 Profiling Compliers with a Non-Binary Instrument

In Appendix A.1.1, we have shown that the joint distribution of potential outcomes is identifiable with a non-binary instrument. As a result, the profiling results presented in Theorem 1.5.2 can be readily applied to this case. The profiling results for a discrete instrument and a continuous instrument are presented in Corollary A.2.1 and Corollary A.2.2 respectively.

Corollary A.2.1. Assume that Assumption A.1.1 holds, and let $g : \mathbb{R} \to \mathbb{R}$ be measurable with $\mathbb{E}[|g(X_i)|] < \infty$, then, conditional on z, z' compliers (that is, $z, z' \in \text{supp}(Z_i)$, $T_i(z) = T_i(z')$ does not hold almost surely, and assume without loss of generality that $T_i(z) \geq T_i(z')$ holds almost surely), the expectation of $g(X_i)$ is identified:

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

Corollary A.2.2. Assume that Assumption A.1.2 holds, and assume that $\text{supp}(P(Z_i)) = [0, 1]$. Let $g : \mathbb{R} \to \mathbb{R}$ be measurable with $\mathbb{E}[|g(X_i)|] < \infty$, then, conditional at each margin of selecting into the treatment, the expectation of $g(X_i)$ is identified:

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

A.3 A Different Quantity of "Profiling"

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

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

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

A.4 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 1.5.1 can be extended to always-takers and never-takers. The results are presented in Proposition A.4.1.

Proposition A.4.1. Assume that Assume that 1 to 4 in Assumption 1.2.1 hold, furthermore, assume that we observe pre-treatment covariates X_i, and let $g(\cdot)$ be any measurable real function of X_i such that $\mathbb{E}[|g(X_i)|] < \infty$, then, for $y \in \{0, 1\}$, we have the following:

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

With the IA IV assumption, Proposition A.4.1 states that the conditional moments of X_i conditional on always-takers and their treated potential outcomes and the conditional moments of X_i conditional on never-takers and their untreated potential outcomes are identifiable. Furthermore, Proposition A.4.1 implies that the conditional cumulative distribution functions are identifiable. This follows because $g(x) = \mathbb{1}\{X_i \leq x\}$ is a bounded measurable map.

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

A.5 More on Estimation and Inference

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

A.5.1 Estimation and Inference under Strong Identification

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

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

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

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

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

Proposition A.5.1. Assume that the following conditions hold:

1. $\mathbb{E}[f(X_i, Y_i, T_i)^4] < \infty$,

2. $\mathbb{E}\left[\begin{pmatrix}1\\Z_i\end{pmatrix}(1,Z_i)\right]$ is positive definite,

3. $\mathbb{E}\left[\begin{pmatrix}1\\Z_i\end{pmatrix}(1,h(Y_i,T_i))\right]$ is rull rank,

4. $\mathbb{E}\left[\begin{pmatrix}1\\Z_i\end{pmatrix}e_i\right]=0$, where e_i is the residual from regressing $f(X_i,Y_i,T_i)$ on $h(Y_i,T_i)$,

5. $\mathbb{E}[h(Y_i,T_i)^4]<\infty$,

6. $\mathbb{E}[Z_i^4]<\infty$,

7. $\Omega=\mathbb{E}\left[\begin{pmatrix}1\\Z_i\end{pmatrix}(1,Z_i)e_i\right]$ is positive definite,

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

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

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

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

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

Before we proceed, we now give a remark on the consistency of the estimator we proposed. Let $g(X_i) = \mathbb{1}\{X_i \leq x\}$, Theorem 1.5.1 shows that we can point identify the conditional distribution function among the locally persuadable:

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

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

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

We prove this uniform consistent result in Appendix A.9.15.

A.5.2 An Anderson-Rubin Test under Weak Identification

Note that the estimand in Equation 1.1 is a function of two regression coefficients:

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

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

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

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

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

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.

A.6 A System of Equation for the Binary IV Model with Monotone Treatment Response

Assumption 1 to 4 in Assumption 1.2.1 implies the following system of linear equations:

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

where A_{obs} is defined as:

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

and $\mathbf{p}$ is defined as the following with A being a measurable set:

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

and $\mathbf{b}$ is defined as the following with A being a measurable set:

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

A.7 Implementing the Test in Section 1.7.2

Recall that in Section 1.7.2, the test statistic is given by:

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

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

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

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

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

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

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

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

A.8 An Equivalence Result

We now use the weighting methods developed in Abadie [2003] to derive the results in Theorem 1.5.1. 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 A.8.1.

Proposition A.8.1. Assume that 1 to 4 in Assumption 1.2.1 hold, then, the distribution of X_i conditional on $[Y_i(0) = 0, T_i(1) > T_i(0)]$ is point identified. Let A be a measurable set:

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

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

We can also show that the identification results in Theorem 1.5.1 and Proposition A.8.1 are equivalent. We formally state this equivalence result in Proposition A.8.2.

Proposition A.8.2. The identification results for $\mathbb{P}[X_i \in A \mid Y_i(0) = 0, T_i(1) > T_i(0)]$ in Theorem 1.5.1 and Proposition A.8.1 are equivalent.

A.9 Proofs

A.9.1 Proof of Lemma 1.4.1

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

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

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

where the second equality uses Lemma 2.1 in Abadie [2003].

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

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

where the first and the second equality uses IV monotonicity in Assumption 1.2.1, the third equality uses IV exogeneity in Assumption 1.2.1. Now, the desired results follow immediately.

A.9.2 Proof of Lemma 1.4.2

By the monotone treatment response assumption in Assumption 1.2.1, $\mathbb{P}[Y_i(1) = 1, Y_i(0) = 1|T_i(1) > T_i(0)] = \mathbb{P}[Y_i(0) = 1|T_i(1) > T_i(0)]$. The desired result follows immediately from Lemma 1.4.1 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 1.2.1 implies $[Y_i(1) = 0, Y_i(0) = 0] = [Y_i(1) = 0]$ and using Lemma 1.4.1.

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 1.2.1 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.

A.9.3 Proof of Theorem 1.5.1

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

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

where the second equality uses Lemma 1.4.1.

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

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

where the first equality uses the IV monotonicity in Assumption 1.2.1, the second equality uses the IV independence in Assumption 1.2.1 and a fact that independence is preserved under measurable transform (e.g., see Theorem 2.1.6. in Durrett [2010]).

A.9.4 Proof of Proposition 1.5.1

The desired results follow immediately by using the identical arguments in Theorem 1.5.1.

A.9.5 Proof of Theorem 1.5.2

For $\mathbb{E}[g(X_i) \mid Y_i(1) = Y_i(0) = 1, T_i(1) > T_i(0)]$. Note that the monotone treatment response assumption in Assumption 1.2.1 implies $[Y_i(1) = Y_i(0) = 1] = [Y_i(0) = 1]$. Now, the desired result follows immediately from Proposition 1.5.1.

Similarly, by Proposition 1.5.1 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 1.2.1, the desired result for $\mathbb{E}[g(X_i) \mid Y_i(1) = Y_i(0) = 1, T_i(1) > T_i(0)]$ follows immediately.

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

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

where the second equality uses Theorem 1 in Imbens and Angrist [1994].

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

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

where the third equality uses the IV independence assumption in Assumption 1.2.1.

A.9.6 Proof of Proposition 1.5.2

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

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

where the equality follows from the outcome monotonicity assumption in Assumption 1.2.1.

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

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

where the second equality uses Lemma 1.4.1 and Theorem 1.5.2.

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

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

where the second eqaulity uses the monotone treatment response assumption in Assumption 1.2.1, the third equality uses Lemma 1.4.1 and Theorem 1.5.2.

A.9.7 Proof of Proposition A.4.1

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

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

where the first equality uses the IV monotonicity assumption in Assumption 1.2.1, the second equality uses the IV independence assumption in Assumption 1.2.1.

A.9.8 Proof of Claim 1.7.1

Note that among compliers, $T_i = Z_i$. Now the desired result follows immediately by observing that Z_i is exogenous assumed in Assumption 1.2.1 and using Theorem 6 in Jun and Lee [2023].

A.9.9 Proof of Theorem 1.7.1

Recall the formulas of the approximated $\tilde{\theta}_{\mathrm{DK}}$ and the identified θ_{local} from Theorem 6 in Jun and Lee [2023]:

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

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

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

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 $\tilde{\theta}_{\mathrm{DK}}$:

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

where the first equality uses the assumption that there is non-compliance in the control group. For the denominator of $\tilde{\theta}_{\mathrm{DK}}$, 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 A.1, $\tilde{\theta}_{\text{DK}} = \theta_{\text{local}}$ 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 A.1, $\tilde{\theta}_{\text{DK}} = \theta_{\text{local}}$ 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 $\mathbb{P}[T_i(0) = 0] = 1$.

A.9.10 Proof of Proposition A.1.1

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.

A.9.11 Proof of Corollary A.1.1

The desired results follow immediately using the identical arguments in Lemma 1.4.1 and Lemma 1.4.2.

A.9.12 Proof of Corollary A.1.2

The desired result follows immediately by using the result in Heckman and Vytlacil [2005] and Carneiro and Lee [2009] and the monotone treatment response assumption in Assumption A.1.1. Since the argument is brief, we include it here for completeness.

Note that by the monotone treatment response assumption in Assumption A.1.2 and the fact that Y_i is binary:

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

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

$$\begin{aligned}
\mathbb{E}[Y_i \mid V_i = v] &= \mathbb{E}[Y_i(0) \mid P(Z_i) = v] + \mathbb{E}[T_i(Y_i(1) - Y_i(0)) \mid P(Z_i) = v] \\
&= \mathbb{E}[Y_i(0) \mid P(Z_i) = v] \\
&\quad + \mathbb{E}[Y_i(1) - Y_i(0) \mid T_i = 1, P(Z_i) = v]\mathbb{P}[T_i = 1 \mid P(Z_i) = v] \\
&= \mathbb{E}[Y_i(0) \mid P(Z_i) = v] \\
&\quad + \mathbb{E}[Y_i(1) - Y_i(0) \mid V_i \leq v, P(Z_i) = v]\mathbb{P}[V_i \leq v \mid P(Z_i) = v] \\
&= \mathbb{E}[Y_i(0)] + \mathbb{E}[Y_i(1) - Y_i(0) \mid V_i \leq v]v \\
&= \mathbb{E}[Y_i(0)] + \mathbb{E}[(Y_i(1) - Y_i(0))\mathbb{1}\{V_i \leq v\}] \\
&= \mathbb{E}[Y_i(0)] + \mathbb{E}[\mathbb{1}\{V_i \leq v\}\mathbb{E}[Y_i(1) - Y_i(0) \mid V_i = u]] \\
&= \mathbb{E}[Y_i(0)] + \int_0^v \mathbb{E}[Y_i(1) - Y_i(0) \mid V_i = u]du,
\end{aligned}$$

where the third equality uses the selection equation in Assumption A.1.2, the fourth equality uses the independence of Z_i and $V_i \sim U[0, 1]$ in Assumption A.1.2. Now the desired result follows immediately by differentiating both sides of the equation with respect to v.

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

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

where the first equality uses the selection equation in Assumption A.1.2, the fourth equality uses $V_i \sim U[0,1]$ in Assumption A.1.2. Now the desired result follows immediately by differentiating both sides of the equation with respect to v and defining g as: $g(Y_i) = \mathbb{1}\{Y_i = 1\}$.

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

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

where the first equality uses the selection equation in Assumption A.1.2, the fourth equality uses $V_i \sim U[0,1]$ in Assumption A.1.2. Now the desired result follows immediately by differentiating both sides of the equation with respect to v and defining g as: $g(Y_i) = \mathbb{1}\{Y_i = 0\}$.

A.9.13 Proof of Corollary A.2.1

The desired results follow immediately using the identical arguments in Theorem 1.5.2.

A.9.14 Proof of Corollary A.2.2

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

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

where the second equality uses the monotone treatment response assumption, and the third equality uses the independence assumption in Assumption A.1.2 and Corollary A.1.2.

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

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

where the second equality uses the monotone treatment response assumption. Now the desired result follows immediately from the independence assumption in Assumption A.1.2 and Corollary A.1.2.

A.9.15 A Glivenko-Cantelli Theorem for Conditional Cumulative Distribution Function

In fact, we can strengthen the statement in Appendix A.5 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 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 A.9.1. 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 $1 \le j \le k-1$. With $x_{0,k} = -\infty$ and $x_{k,k} = \infty$, then the last two inequalities hold for $j = 0$ and $j = k$.

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

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

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

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

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

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. ■

A.9.16 Proof of Proposition A.8.2

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

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

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 1.5.1 and Proposition A.8.1, it suffices to show the equivalence between the numerators in the two formulas:

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

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

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

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

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

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

Now the desired equivalence result follows immediately.

APPENDIX B

APPENDIX FOR CHAPTER 3

B.1 Proof of Proposition 3.3.1

($\Leftarrow$) Integrating on both sides of the equation yields:

$$\begin{aligned}
&\left(\int g(y)dF_{Y_{i1}(0)|G_i=1,H_i=1}(y) - \int g(y)dF_{Y_{i0}(0)|G_i=1,H_i=1}(y)\right) \\
&- \left(\int g(y)dF_{Y_{i1}(0)|G_i=0,H_i=1}(y) - \int g(y)dF_{Y_{i0}(0)|G_i=0,H_i=1}(y)\right) \\
&= \left(\int g(y)dF_{Y_{i1}(0)|G_i=1,H_i=0(y)} - \int g(y)dF_{Y_{i0}(0)|G_i=1,H_i=0}(y)\right) \\
&- \left(\int g(y)dF_{Y_{i1}(0)|G_i=0,H_i=0}(y) - \int g(y)dF_{Y_{i0}(0)|G_i=0,H_i=0})(y)\right),
\end{aligned}$$

provided g is measurable and the integral exists.

($\Rightarrow$) The desired result follows immediately by defining: $g(y) = \mathbb{1}\{y \leq \tilde{y}\}$, where $\tilde{y} \in \mathbb{R}$.

B.2 Proof of Proposition 3.3.2

The first statement follows immediately from some simple calculations.

For the second statement, let's show that the decomposition $\sum_{k=1}^{K} \theta_k J_k$ must include F_t^g, and similar arguments follow for F_t^h and F^{gh}. We proceed with the proof by contradiction. Suppose the decomposition does not include F_t^g, then, we can get a new decomposition:

$$\gamma \sum_{k=1}^{K} \theta_k J_k + (1-\gamma) F_t^g \ ,$$

where $\gamma \in (0,1)$, such that the "modified" parallel trends assumption is invariant to transformations, which is a contradiction to the claim that the decomposition $\{J_k\}_{k=1}^{K}$ does not include F_t^g.

B.3 Proof of Proposition 3.7.1

The first part of the statement follows directly from Equations 3.1, 3.2, and 3.9.

For the second part of the statement, note that since $\mathbb{P}^{\text{implied}}_{Y_{i1}(0)|G_i=1,H_i=1}(Y_{i1}(0) \in A) \geq 0$ for any $A \in \mathcal{B}(\mathbb{R})$, and the right-hand side of Equation 3.9 are all probability measures, $\mathbb{P}^{\text{implied}}_{Y_{i1}(0)|G_i=1,H_i=1}$ is a valid probability measure.

Given that $P^{\text{implied}}_{Y_{i1}(0)|G_i=1,H_i=1}(Y_{i1}(0) \in A)$ is a valid marginal probability distribution, we first construct a valid joint distribution of potential outcomes and group variables (i.e., $(Y_{i1}(1), Y_{i0}(1), Y_{i1}(0), Y_{i0}(0), G_i, H_i)$), which is denoted by $\mathbb{P}^*$.

By the Product Measure Theorem (e.g., Theorem 1.7.1. in Durrett [2010]), the fact that $\mathcal{B}(\mathbb{R}^k \times \mathbb{R}^m) = \sigma(\mathcal{B}(\mathbb{R}^k) \times \mathcal{B}(\mathbb{R}^m))$ for $k, m \in \mathbb{N}^+$, and the fact that $\mathbb{P}^{\text{implied}}_{Y_{i1}(0)|G_i=1,H_i=1}$ is a valid probability measure, there exists a unique probability measure μ_{11} on $\mathcal{B}(\mathbb{R}^4)$ such that

$$
\begin{aligned}
&\mu_{11}((Y_{i0}(0), Y_{i1}(1)) \in B_1, Y_{i1}(0) \in A_1, Y_{i0}(1) \in A_2) \\
&= \mathbb{P}((Y_{i0}, Y_{i1}) \in B_1 \mid G_i = 1, H_i = 1) \times \mathbb{P}^{\text{implied}}_{Y_{i1}(0)|G_i=1,H_i=1}(Y_{i1}(0) \in A_1) \\
&\quad \times \mathbb{P}(Y_{i0} \in A_2 \mid G_i = 1, H_i = 1) \ ,
\end{aligned}
$$

where $B_1 \in \mathcal{B}(\mathbb{R}^2)$, and $A_1, A_2 \in \mathcal{B}(\mathbb{R})$. Similarly, for $g, h \in \{0,1\}$ such that $g \times h = 0$, there exists a unique probability measure μ_{gh} on $\mathcal{B}(\mathbb{R}^4)$ such that

$$\begin{aligned}&\mu_{gh}((Y_{i0}(0), Y_{i1}(0)) \in B_1, (Y_{i0}(1), Y_{i1}(1)) \in B_2)\\&= \mathbb{P}((Y_{i0}, Y_{i1}) \in B_1 \mid G_i = g, H_i = h) \times ((Y_{i0}, Y_{i1}) \in B_2 \mid G_i = g, H_i = h)\ ,\end{aligned}$$

where $B_1, B_2 \in \mathcal{B}(\mathbb{R}^2)$. Therefore, there exists a probability measure $\mathbb{P}^*$ on the Borel σ-algebra on $\mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \{0,1\} \times \{0,1\}$ such that

$$\begin{aligned}&\mathbb{P}^*((Y_{i0}(0), Y_{i1}(1)) \in B_1, Y_{i1}(0) \in A_1, Y_{i0}(1) \in A_2 \mid G_i = 1, H_i = 1)\\&= \mu_{11}((Y_{i0}(0), Y_{i1}(1)) \in B_1, Y_{i1}(0) \in A_1, Y_{i0}(1) \in A_2)\ ,\end{aligned}$$

where $B_1 \in \mathcal{B}(\mathbb{R}^2)$, and $A_1, A_2 \in \mathcal{B}(\mathbb{R})$, and for $g, h \in \{0,1\}$ such that $g \times h = 0$,

$$\begin{aligned}&\mathbb{P}^*((Y_{i0}(0), Y_{i1}(0)) \in B_1, (Y_{i0}(1), Y_{i1}(1)) \in B_2 \mid G_i = g, H_i = h)\\&= \mu_{gh}((Y_{i0}(0), Y_{i1}(0)) \in B_1, (Y_{i0}(1), Y_{i1}(1)) \in B_2)\ ,\end{aligned}$$

where $B_1, B_2 \in \mathcal{B}(\mathbb{R}^2)$.

Our next step is to check that the constructed joint distribution induces the observed distribution in the data. To see this,

$$\begin{aligned}&\mathbb{P}_{\text{induced}}((Y_{i0}, Y_{i1}) \in B | G_i = 1, H_i = 1)\\&= \mathbb{P}^*((Y_{i0}(0), Y_{i1}(1)) \in B | G_i = 1, H_i = 1)\\&= \mathbb{P}((Y_{i0}, Y_{i1}) \in B \mid G_i = 1, H_i = 1) \times \mathbb{P}^{\text{implied}}_{Y_{i1}(0)|G_i=1,H_i=1}(Y_{i1}(0) \in \mathbb{R})\\&\quad \times P(Y_{i0} \in \mathbb{R} \mid G_i = 1, H_i = 1)\\&= \mathbb{P}((Y_{i0}, Y_{i1}) \in B | G_i = 1, H_i = 1)\ ,\end{aligned}$$

and

$$
\begin{aligned}
&\mathbb{P}_{\text{induced}}((Y_{i0}, Y_{i1}) \in B | G_i = g, H_i = h) \\
&= \mathbb{P}^*((Y_{i0}(0), Y_{i1}(0)) \in B | G_i = g, H_i = h) \\
&= \mathbb{P}(Y_{i0}, Y_{i1}) \in B_1 \mid G_i = g, H_i = h) \times \mathbb{P}(Y_{i0}, Y_{i1}) \in \mathbb{R}_2 \mid G_i = g, H_i = h) \\
&= \mathbb{P}(Y_{i0}, Y_{i1}) \in B_1 \mid G_i = g, H_i = h) \ ,
\end{aligned}
$$

for $g, h \in \{0, 1\}$ such that $g \times h = 0$.

Finally, we verify that the constructed joint distribution satisfies the invariance condition. This follows immediately from the SUTVA condition.

B.4 Implementing the Least Favorable Method in Canay and Shaikh [2017]

Below are the steps of the least favorable test in Canay and Shaikh [2017]:

1. Resample the data with replacement with sample size n;
2. Compute the implied densities for the bootstrapped sample;
3. Repeat the previous two steps $B = 1000$ times and get a matrix of results, where there are B rows, and columns are implied densities for different Borel sets. Denote each row b as $(\hat{f}^*_{1,b}, ..., \hat{f}^*_{p,b})$, where p is the number of Borel sets
4. Use the results matrix in the previous step to calculate the correlation matrix. Denote the correlation matrix as $\{\hat{\sigma}^*_{m,n}\}_{1 \le m,n \le p}$

5. Calculate the test statistic

$$T_n = \max_{1 \le j \le p} \frac{-\sqrt{n}\hat{f}_j}{\hat{\sigma}^*_{j,j}} \ ,$$

where $\{\hat{f}_j\}_{1 \le j \le p}$ is the implied density of the real data

6. Denote $T^*_{n,b} = \max_{1 \le j \le p} \frac{-\sqrt{n}\hat{f}^*_{j,b}}{\hat{\sigma}^*_{j,j}}$. Compute the $(1-\alpha)$-th quantile of $\{T^*_{n,b}\}_{1 \le b \le B}$. Denote this quantile by $\hat{c}(1-\alpha)$. Moreover, $\hat{c}(1-\alpha)$ is the critical value for a level α test in the least favorable method.

7. Reject the null hypothesis if $T_n > \hat{c}(1-\alpha)$. Moreover, the p-value is the proportion of $\{T^*_{n,b}\}_{1 \le b \le B}$ that are larger than T_n.

B.5 Implementing the Two-Step Method in Romano et al. [2014] and Bai et al. [2022a]

Let $W_i = (Y_{i0}, Y_{i1}, G_i, H_i)$, and let $\{A_j\}_{j=1}^p$ be a finite partition of real line $\mathbb{R}$ such that $A_j \in \mathcal{B}(\mathbb{R})$ for $1 \le j \le p$. We want to test the following null hypothesis:

$$\begin{aligned}
H_0 :& \mathbb{E}[-\mathbb{1}\{Y_{i0}(0) \in A_j\} \mid G_i = 1, H_i = 1] \\
&+ (\mathbb{E}[-\mathbb{1}\{Y_{i1}(0) \in A_j\} \mid G_i = 1, H_i = 0] - \mathbb{E}[-\mathbb{1}\{Y_{i0}(0) \in A_j\} \mid G_i = 1, H_i = 0]) \\
&+ (\mathbb{E}[-\mathbb{1}\{Y_{i1}(0) \in A_j\} \mid G_i = 0, H_i = 1] - \mathbb{E}[-\mathbb{1}\{Y_{i0}(0) \in A_j \mid G_i = 0, H_i = 1\}]) \\
&- (\mathbb{E}[-\mathbb{1}\{Y_{i1}(0) \in A_j\} \mid G_i = 0, H_i = 0] - \mathbb{E}[-\mathbb{1}\{Y_{i0}(0) \in A_j\} \mid G_i = 0, H_i = 0]) \le 0,
\end{aligned}$$

$\forall 1 \le j \le p$.

Furthermore, we can rewrite the null hypothesis above as:

$$
\begin{aligned}
&\mathbb{E}[-\mathbb{1}\{Y_{i0}(0) \in A_j\} \mid G_i = 1, H_i = 1] \\
&\quad + (\mathbb{E}[-\mathbb{1}\{Y_{i1}(0) \in A_j\} \mid G_i = 1, H_i = 0] - \mathbb{E}[-\mathbb{1}\{Y_{i0}(0) \in A_j\} \mid G_i = 1, H_i = 0]) \\
&\quad + (\mathbb{E}[-\mathbb{1}\{Y_{i1}(0) \in A_j\} \mid G_i = 0, H_i = 1] - \mathbb{E}[-\mathbb{1}\{Y_{i0}(0) \in A_j\} \mid G_i = 0, H_i = 1]) \\
&\quad - (\mathbb{E}[-\mathbb{1}\{Y_{i1}(0) \in A_j\} \mid G_i = 0, H_i = 0] - \mathbb{E}[-\mathbb{1}\{Y_{i0}(0) \in A_j\} \mid G_i = 0, H_i = 0]) \\
&= \mathbb{E}[-\mathbb{1}\{Y_{i0} \in A_j\} \mid G_i = 1, H_i = 1] \\
&\quad + (\mathbb{E}[-\mathbb{1}\{Y_{i1} \in A_j\} \mid G_i = 1, H_i = 0] - \mathbb{E}[-\mathbb{1}\{Y_{i0} \in A_j\} \mid G_i = 1, H_i = 0]) \\
&\quad + (\mathbb{E}[-\mathbb{1}\{Y_{i1} \in A_j\} \mid G_i = 0, H_i = 1] - \mathbb{E}[-\mathbb{1}\{Y_{i0} \in A_j\} \mid G_i = 0, H_i = 1]) \\
&\quad - (\mathbb{E}[-\mathbb{1}\{Y_{i1} \in A_j\} \mid G_i = 0, H_i = 0] - \mathbb{E}[-\mathbb{1}\{Y_{i0} \in A_j\} \mid G_i = 0, H_i = 0]) \\
&= \mathbb{E}\Bigg[\frac{-\mathbb{1}\{Y_{i0} \in A_j\}\mathbb{1}\{G_i = 1, H_i = 1\}}{\mathbb{E}[\mathbb{1}\{G_i = 1, H_i = 1\}]} \\
&\qquad + \frac{(-\mathbb{1}\{Y_{i1} \in A_j\} + \mathbb{1}\{Y_{i0} \in A_j\})\mathbb{1}\{G_i = 1, H_i = 0\}}{\mathbb{E}[\mathbb{1}\{G_i = 1, H_i = 0\}]} \\
&\qquad + \frac{(-\mathbb{1}\{Y_{i1} \in A_j\} + \mathbb{1}\{Y_{i0} \in A_j\})\mathbb{1}\{G_i = 0, H_i = 1\}}{\mathbb{E}[\mathbb{1}\{G_i = 0, H_i = 1\}]} \\
&\qquad - \frac{(-\mathbb{1}\{Y_{i1} \in A_j\} + \mathbb{1}\{Y_{i0} \in A_j\})\mathbb{1}\{G_i = 0, H_i = 0\}}{\mathbb{E}[\mathbb{1}\{G_i = 0, H_i = 0\}]}\Bigg],
\end{aligned}
$$

where the first equality rewrites the observed potential outcomes as observed outcomes, and the second equality uses the definition of conditional expectation.

Moreover, denote X_{ij} as the following:

$$
\begin{aligned}
X_{ij} = & \frac{-\mathbb{1}\{Y_{i0} \in A_j\}\mathbb{1}\{G_i = 1, H_i = 1\}}{\mathbb{E}[\mathbb{1}\{G_i = 1, H_i = 1\}]} \\
& + \frac{(-\mathbb{1}\{Y_{i1} \in A_j\} + \mathbb{1}\{Y_{i0} \in A_j\})\mathbb{1}\{G_i = 1, H_i = 0\}}{\mathbb{E}[\mathbb{1}\{G_i = 1, H_i = 0\}]} \\
& + \frac{(-\mathbb{1}\{Y_{i1} \in A_j\} + \mathbb{1}\{Y_{i0} \in A_j\})\mathbb{1}\{G_i = 0, H_i = 1\}}{\mathbb{E}[\mathbb{1}\{G_i = 0, H_i = 1\}]} \\
& - \frac{(-\mathbb{1}\{Y_{i1} \in A_j\} + \mathbb{1}\{Y_{i0} \in A_j\})\mathbb{1}\{G_i = 0, H_i = 0\}}{\mathbb{E}[\mathbb{1}\{G_i = 0, H_i = 0\}]} , \\
X_i = & (X_{i1}, ..., X_{ip})' .
\end{aligned}
$$

Thus, our original null hypothesis can be written as

$$
H_0 : \mathbb{E}[X_i] \leq 0 .
$$

Note that X_i is infeasible in our setting since it involves unknown population parameters, $\mathbb{E}[\mathbb{1}\{G_i = g, H_i = h\}]$, $g, h \in \{0, 1\}$. We use a feasible X_i, which uses the sample analog $\hat{\mathbb{E}}[\mathbb{1}\{G_i = g, H_i = h\}]$ to estimate $\mathbb{E}[\mathbb{1}\{G_i = g, H_i = h\}]$. We use $\hat{X}_i$ to denote the feasible X_i that we propose. We then apply the two-step procedure in Romano et al. [2014] and Bai et al. [2022b] using $\hat{X}_i$. After replacing X_i with $\hat{X}_i$, we conjecture that the procedure in Romano et al. [2014] and Bai et al. [2022b] is still valid.

Moreover, denote $\hat{X}_{ij}$ by:

$$
\begin{aligned}
\hat{X}_{ij} = & \frac{-\mathbb{1}\{Y_{i0} \in A_j\}\mathbb{1}\{G_i = 1, H_i = 1\}}{\hat{\mathbb{E}}[\mathbb{1}\{G_i = 1, H_i = 1\}]} \\
& + \frac{(-\mathbb{1}\{Y_{i1} \in A_j\} + \mathbb{1}\{Y_{i0} \in A_j\})\mathbb{1}\{G_i = 1, H_i = 0\}}{\hat{\mathbb{E}}[\mathbb{1}\{G_i = 1, H_i = 0\}]} \\
& + \frac{(-\mathbb{1}\{Y_{i1} \in A_j\} + \mathbb{1}\{Y_{i0} \in A_j\})\mathbb{1}\{G_i = 0, H_i = 1\}}{\hat{\mathbb{E}}[\mathbb{1}\{G_i = 0, H_i = 1\}]} \\
& - \frac{(-\mathbb{1}\{Y_{i1} \in A_j\} + \mathbb{1}\{Y_{i0} \in A_j\})\mathbb{1}\{G_i = 0, H_i = 0\}}{\hat{\mathbb{E}}[\mathbb{1}\{G_i = 0, H_i = 0\}]} \ , \\
\hat{X}_i = & (\hat{X}_{i1}, ..., \hat{X}_{ip})' \ , \\
\overline{\hat{X}}_{j,n} = & \frac{1}{n}\sum_{i=1}^{n} \hat{X}_{ij} \ , \\
\overline{\hat{X}}_n = & \frac{1}{n}\sum_{i=1}^{n} \hat{X}_i = (\overline{\hat{X}}_1, ..., \overline{\hat{X}}_p)' \ , \\
\hat{S}^2_{j,n} = & \frac{1}{n}\sum_{i=1}^{n} (\hat{X}_{ij} - \overline{\hat{X}}_j)^2 \ ,
\end{aligned}
$$

where

$$
\hat{\mathbb{E}}[\mathbb{1}\{G_i = g, H_i = h\}] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{G_i = g, H_i = h\} \ ,
$$

for $g, h \in \{0, 1\}$.

Note that we have the following numerical relationship:

$$
\overline{\hat{X}}_{j,n} = \hat{\mathbb{P}}^{\text{implied}}_{Y_{i1}(0)|G_i=1,H_i=1}(Y_{i1}(0) \in A_j),
$$

for $1 \leq j \leq p$. Then, we can compute the test statistic by using the two-step procedure proposed by Romano et al. [2014] and Bai et al. [2022b]:

$$T_n = \max \left\{ \max_{1 \le j \le p} \frac{\sqrt{n}\overline{\hat{X}}_{j,n}}{\hat{S}_{j,n}}, 0 \right\} \ .$$

Before computing the critical value, we first introduce some notations. Let $\hat{X}_i^*$ denote an i.i.d. sequence of random vectors with distribution $\hat{\mathbb{P}}_n$ conditional on $\{\hat{X}_i\}_{i=1}^n$. We further denote $\overline{\hat{X}}_{j,n}^*$ and $\hat{S}_{j,n}^*$ as:

$$\begin{aligned}
\overline{\hat{X}}_{j,n}^* &= \frac{1}{n} \sum_{i=1}^n \hat{X}_{ij}^* \\
(\hat{S}_{j,n}^*)^2 &= \frac{1}{n} \sum_{i=1}^n (\hat{X}_{ij}^* - \overline{\hat{X}}_{j,n}^*)^2 \ .
\end{aligned}$$

Then, we can compute the critical value in the following way:

$$\begin{aligned}
&\hat{c}_n^{(2)}(1-\alpha+\beta) \\
&= \inf \left\{ c \in \mathbb{R} : \mathbb{P}\left\{ \max\left\{ T_n^{\text{two-step}}, 0 \right\} \le c \mid \{\hat{X}_i\}_{i=1}^n \right\} \ge 1-\alpha+\beta \right\} \ ,
\end{aligned}$$

where

$$\begin{aligned}
T_n^{\text{two-step}} &= \max_{1 \le j \le p} \frac{\sqrt{n}(\overline{\hat{X}}_{j,n}^* - \overline{\hat{X}}_{j,n} + \hat{u}_{j,n})}{\hat{S}_{j,n}^*} \ , \\
\hat{u}_{j,n} &= \min \left\{ \overline{\hat{X}}_{j,n} + \frac{\hat{S}_{j,n}}{\sqrt{n}} \hat{c}_n^{(1)}(1-\beta), 0 \right\} \ , \\
\hat{c}_n^{(1)}(1-\beta) &= \inf \left\{ c \in \mathbb{R} : \mathbb{P}\left\{ \max_{1 \le j \le p} \frac{\sqrt{n}(\overline{\hat{X}}_{j,n} - \overline{\hat{X}}_{j,n}^*)}{\hat{S}_{j,n}^*} \le c \mid \{\hat{X}_i\}_{i=1}^n \right\} \ge 1-\beta \right\} \ .
\end{aligned}$$

Then, the rejection rule we use is:

$$\phi_n^{\text{RSW}} = \mathbb{1}\{T_n > \hat{c}_n^{(2)}(1-\alpha+\beta)\} \ .$$

Note that β is the level of the test in the pre-testing step in the two-step method in Romano et al. [2014]. Following Romano et al. [2014], we set $\beta = \frac{\alpha}{10}$.

REFERENCES

Daniel Aaronson and Bhashkar Mazumder. The impact of rosenwald schools on black achievement. *Journal of Political Economy*, 119(5):821–888, 2011.

Alberto Abadie. Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American statistical Association*, 97(457):284–292, 2002.

Alberto Abadie. Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, 113(2):231–263, 2003.

Alberto Abadie, Joshua Angrist, and Guido Imbens. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70(1): 91–117, 2002.

Jaap H Abbring and Gerard J Van den Berg. The nonparametric identification of treatment effects in duration models. *Econometrica*, 71(5):1491–1517, 2003.

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434): 444–455, 1996.

Sarah F Anzia. Election timing and the electoral influence of interest groups. *The Journal of Politics*, 73(2):412–427, 2011.

Susan Athey and Guido W Imbens. Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497, 2006.

Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89 (1):133–161, 2021.

Yuehao Bai, Andres Santos, and Azeem M Shaikh. On testing systems of linear inequalities with known coefficients. *Working Paper*, 2022a.

Yuehao Bai, Andres Santos, and Azeem M Shaikh. A two-step method for testing many moment inequalities. *Journal of Business & Economic Statistics*, 40(3):1070–1080, 2022b.

Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176, 1997.

Gary S Becker. Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2):169–217, 1968.

Dirk Bergemann and Stephen Morris. Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95, 2019.

Peter J Bickel and David A Freedman. Some asymptotic theory for the bootstrap. *The annals of statistics*, 9(6):1196–1217, 1981.

Christopher Blattman and Jeannie Annan. Can employment reduce lawlessness and rebellion? a field experiment with high-risk men in a fragile state. *American Political Science Review*, 110(1):1–17, 2016.

Christopher Blattman, Julian C Jamison, and Margaret Sheridan. Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in liberia. *American Economic Review*, 107(4):1165–1206, 2017.

Christopher Blattman, Nathan Fiala, and Sebastian Martinez. The long-term impacts of grants on poverty: Nine-year evidence from uganda's youth opportunities program. *American Economic Review: Insights*, 2(3):287–304, 2020.

Stéphane Bonhomme and Ulrich Sauder. Recovering distributions in difference-in-differences models: A comparison of selective and comprehensive schooling. *Review of Economics and Statistics*, 93(2):479–494, 2011.

Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Brantly Callaway and Tong Li. Quantile treatment effects in difference in differences models with panel data. *Quantitative Economics*, 10(4):1579–1618, 2019.

Ivan A Canay and Azeem M Shaikh. Practical and theoretical advances in inference for partially identified models. *Advances in Economics and Econometrics*, 2:271–306, 2017.

Pedro Carneiro and Sokbae Lee. Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics*, 149(2):191–208, 2009.

Yuyu Chen and David Y Yang. The impact of media censorship: 1984 or brave new world? *American Economic Review*, 109(6):2294–2332, 2019.

Victor Chernozhukov and Christian Hansen. The impact of 401 (k) participation on the wealth distribution: An instrumental quantile regression analysis. *Review of Economics and statistics*, 86(3):735–751, 2004.

Victor Chernozhukov and Christian Hansen. An iv model of quantile treatment effects. *Econometrica*, 73(1):245–261, 2005.

Stefano DellaVigna and Matthew Gentzkow. Persuasion: empirical evidence. *Annual Review of Economics*, 2(1):643–669, 2010.

Stefano DellaVigna and Ethan Kaplan. The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234, 2007.

Rick Durrett. *Probability: theory and examples*. Cambridge university press, 2010.

Ruben Enikolopov, Maria Petrova, and Ekaterina Zhuravskaya. Media and political persuasion: Evidence from russia. *American Economic Review*, 101(7):3253–85, 2011.

Ryan D Enos, Anthony Fowler, and Lynn Vavreck. Increasing inequality: The effect of gotv mobilization on the composition of the electorate. *The Journal of Politics*, 76(1):273–288, 2014.

Zheng Fang and Andres Santos. Inference on directionally differentiable functions. *The Review of Economic Studies*, 86(1):377–412, 2019.

Zheng Fang, Andres Santos, Azeem M Shaikh, and Alexander Torgovitsky. Inference for large-scale linear systems with known coefficients. *Econometrica*, 91(1):299–327, 2023.

Qian Feng, Quang Vuong, and Haiqing Xu. Estimation of heterogeneous individual treatment effects with endogenous treatments. *Journal of the American Statistical Association*, 2019.

Anqi Fu, Balasubramanian Narasimhan, and Stephen Boyd. Cvxr: An r package for disciplined convex optimization. *arXiv preprint arXiv:1711.07582*, 2017.

Alan S Gerber, Donald P Green, and Ron Shachar. Voting may be habit-forming: evidence from a randomized field experiment. *American journal of political science*, 47(3):540–550, 2003.

Donald P Green, Alan S Gerber, and David W Nickerson. Getting out the vote in local elections: Results from six door-to-door canvassing experiments. *The Journal of Politics*, 65(4):1083–1096, 2003.

Herschell I Grossman. A general equilibrium model of insurrections. *The American Economic Review*, pages 912–921, 1991.

Bruce Hansen. *Econometrics*. Princeton University Press, 2022.

James J Heckman and Edward Vytlacil. Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica*, 73(3):669–738, 2005.

Richard Hornbeck. Barbed wire: Property rights and agricultural development. *The Quarterly Journal of Economics*, 125(2):767–810, 2010.

Martin Huber and Giovanni Mellace. Testing instrument validity for late identification based on inequality moment constraints. *Review of Economics and Statistics*, 97(2):398–411, 2015.

Guido W Imbens and Joshua D Angrist. Identification and estimation of local average treatment effects. *Econometrica*, pages 467–475, 1994.

Guido W Imbens and Donald B Rubin. Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies*, 64(4):555–574, 1997.

Sung Jae Jun and Sokbae Lee. Identifying the effect of persuasion. *Journal of Political Economy*, 131(8):2032–2058, 2023.

Ariella Kahn-Lang and Kevin Lang. The promise and pitfalls of differences-in-differences: Reflections on 16 and pregnant and other applications. *Journal of Business & Economic Statistics*, 38(3):613–620, 2020.

Désiré Kédagni and Ismael Mourifié. Generalized instrumental inequalities: testing the instrumental variable independence assumption. *Biometrika*, 107(3):661–675, 2020.

Wooyoung Kim, Koohyun Kwon, Soonwoo Kwon, and Sokbae Lee. The identification power of smoothness assumptions in models with counterfactual outcomes. *Quantitative Economics*, 9(2):617–642, 2018.

Toru Kitagawa. A test for instrument validity. *Econometrica*, 83(5):2043–2063, 2015.

Toru Kitagawa and Aleksey Tetenov. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616, 2018.

Craig E Landry, Andreas Lange, John A List, Michael K Price, and Nicholas G Rupp. Toward an understanding of the economics of charity: Evidence from a field experiment. *The Quarterly journal of economics*, 121(2):747–782, 2006.

Cecilia Machado, Azeem M Shaikh, and Edward J Vytlacil. Instrumental variables and the sign of the average treatment effect. *Journal of Econometrics*, 212(2):522–555, 2019.

Anup Malani and Julian Reif. Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform. *Journal of Public Economics*, 124:1–17, 2015.

Charles Manski. Monotone treatment response. *Econometrica*, 65(6):1311–1334, 1997.

Charles F Manski and John V Pepper. Monotone instrumental variables: With an application to the returns to schooling. *Econometrica*, 68(4):997–1010, 2000.

Eric Mbakop and Max Tabord-Meehan. Model selection for treatment choice: Penalized welfare maximization. *Econometrica*, 89(2):825–848, 2021.

Breed D Meyer. Natural and quasi-experiments in economics. *Journal of business & economic statistics*, 13(2):151–161, 1995.

Magne Mogstad, Andres Santos, and Alexander Torgovitsky. Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*, 86(5):1589–1619, 2018.

Ismael Mourifié and Yuanyuan Wan. Testing local average treatment effect assumptions. *Review of Economics and Statistics*, 99(2):305–313, 2017.

Karthik Muralidharan and Nishith Prakash. Cycling to school: Increasing secondary school enrollment for girls in india. *American Economic Journal: Applied Economics*, 9(3):321–350, 2017.

Tsunao Okumura and Emiko Usui. Concave-monotone treatment response and monotone treatment selection: With an application to the returns to schooling. *Quantitative Economics*, 5(1):175–194, 2014.

Andreas Olden and Jarle Møen. The triple difference estimator. *The Econometrics Journal*, 25(3):531–553, 2022.

Judea Pearl. Probabilities of causation: three counterfactual interpretations and their identification. *Synthese*, 121(1):93–149, 1999.

Margaret Roberts. *Censored: distraction and diversion inside China's Great Firewall*. Princeton University Press, 2018.

Joseph P Romano and Azeem M Shaikh. On the uniform asymptotic validity of subsampling and the bootstrap. *The Annals of Statistics*, 40(6):2798–2822, 2012.

Joseph P Romano, Azeem M Shaikh, and Michael Wolf. A practical two-step method for testing moment inequalities. *Econometrica*, 82(5):1979–2002, 2014.

Jonathan Roth and Pedro HC Sant'Anna. When is parallel trends sensitive to functional form? *Econometrica*, 91(2):737–747, 2023.

Jonathan Roth, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 2023.

Thomas M Russell. Sharp bounds on functionals of the joint distribution in the analysis of treatment effects. *Journal of Business & Economic Statistics*, 39(2):532–546, 2021.

Yuya Sasaki and Takuya Ura. Welfare analysis via marginal treatment effects. *arXiv preprint arXiv:2012.07624*, 2020.

Douglas Staiger and James H Stock. Instrumental variables regression with weak instruments. *Econometrica: Journal of the Econometric Society*, pages 557–586, 1997.

Alexander Torgovitsky. Nonparametric inference on state dependence in unemployment. *Econometrica*, 87(5):1475–1505, 2019.

Aad W Vaart and Jon A Wellner. Weak convergence. In *Weak convergence and empirical processes*, pages 16–28. Springer, 1996.

Quang Vuong and Haiqing Xu. Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity. *Quantitative Economics*, 8(2):589–610, 2017.

Edward Vytlacil. Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, 70(1):331–341, 2002.

Linbo Wang, James M Robins, and Thomas S Richardson. On falsification of the binary instrumental variable model. *Biometrika*, 104(1):229–236, 2017.

David Williams. *Probability with martingales*. Cambridge university press, 1991.

Teppei Yamamoto. Understanding the past: Statistical analysis of causal attribution. *American Journal of Political Science*, 56(1):237–256, 2012.