Identifying the effect of persuasion

Sun Jae Jun
Sokbae Lee

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper CWP69/19

Identifying the Effect of Persuasion*

Sung Jae Jun† Sokbae Lee‡

Penn State Univ. Columbia Univ. and IFS

November 2019

Abstract. We set up an econometric model of persuasion and study identification of key parameters under various scenarios of data availability. We find that a commonly used measure of persuasion does not estimate the persuasion rate of any population in general. We provide formal identification results, recommend several new parameters to estimate, and discuss their interpretation. Further, we propose methods for carrying out inference. We revisit the empirical literature on persuasion to show that the persuasive effect is highly heterogeneous. We also show that the existence of a continuous instrument opens up the possibility of point identification for the policy-relevant population.

Key Words: Communication, Media, Persuasion, Partial Identification, Treatment Effects

JEL Classification Codes: C21, D72, L82

*We would like to thank the editor, four anonymous referees, Eric Auerbach, Stefano DellaVigna, Leonard Goff, Marc Henry, Keisuke Hirano, Joel Horowitz, Charles Manski, Joris Pinkse, Imran Rasul, and seminar participants at Northwestern University, Rutgers University, Seoul National University and Vanderbilt University for helpful comments. This work was supported in part by the European Research Council (ERC-2014-CoG-646917-ROMIA) and by the UK Economic and Social Research Council (ESRC) through research grant (ES/P008909/1) to the Centre for Microdata Methods and Practice. This paper includes links to five articles, and we would like to thank the authors of these papers for making their data sets and replication files available via journal archives or personal web pages.

†Department of Economics, Pennsylvania State University, 619 Kern Graduate Building, University Park, PA 16802, suj14@psu.edu

‡Department of Economics, Columbia University, 1022 International Affairs Building, 420 West 118th Street, New York, NY 10027, sl3841@columbia.edu

1. Introduction

How effectively one can accomplish persuasion has been of interest to ancient Greek philosophers in the Lyceum of Athens,[1] to early–modern English preachers in St Paul's Cathedral,[2] and to contemporary American news producers at Fox News in New York City.[3] Recently, economists have been endeavoring to build theoretical models of persuasion (e.g. Kamenica and Gentzkow, 2011; Che, Dessein, and Kartik, 2013; Gentzkow and Kamenica, 2017; Prat, 2018; Bergemann and Morris, 2019) and to quantify empirically to what extent persuasive effort affects the behavior of consumers, voters, donors, and investors (see DellaVigna and Gentzkow, 2010, for a survey of the recent literature).

In this paper, we set up an econometric model of persuasion, point out the key parameters of interest, and study their identification under various scenarios of data availability. Because we have observational data in mind, it is important that we allow for endogeneity (i.e. the possibility that agents' decisions on exposure to persuasive information are correlated with their potential actions). To convey this idea, we consider DellaVigna and Kaplan (2007, DK hereafter), who study the effect of exposure to Fox News on the probability of voting for a Republican presidential candidate. Here, the persuasive information of interest is the viewership of the Fox News channel, where the agent's decision about whether to watch Fox News or not may be correlated with their political orientation. In order to capture the *causal* effect of persuasion, we formulate the problem within the potential outcome framework. In the paper, we do consider nonbinary outcomes, but we use the binary outcome (e.g. voting for a Republican candidate or not) as a prototype model because it is simpler. The persuasive treatment of interest is binary throughout the paper.

Let T_i denote the binary indicator that equals 1 if individual i is exposed to persuasive information such as Fox News. Let $Y_i(t)$ be a binary indicator, which shows agent i's potential action when T_i is *exogenously* set to $t \in \{0, 1\}$. For example, $Y_i(1)$ equals 1 if individual i votes for a Republican candidate after watching Fox News. The econometrician never observes both $Y_i(0)$ and $Y_i(1)$ but can only observe either of the two: that is, $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$. Then, the fraction of the people who take the action of interest with an exposure to persuasion, among those who would not without it, is given by

$$\theta = \mathbb{P}\{Y_i(1) = 1 | Y_i(0) = 0\}, \tag{1}$$

[1] See Rapp (2010) for three technical means of persuasion in Aristotle's *Rhetoric*.

[2] See Kirby (2008) for historic details of the public persuasion at Paul's Cross, the open-air pulpit in St Paul's Cathedral in the 16th century.

[3] DellaVigna and Kaplan (2007) and Martin and Yurukoglu (2017) measure the persuasive effects of slanted news using data on Fox News.

which we call the persuasion rate. Note that θ is defined by a conditional probability so that we rule out the case of "preaching to the converted"; if $Y_i(0) = 1$, then individual i is already persuaded to take the action of interest even without the persuasive treatment.[4]

Before we further discuss θ, it is worth making a comparison with the estimand f used in DK, DellaVigna and Gentzkow (2010), and many others. For a binary outcome and by using DK's notation, f is defined by

$$f = \frac{y_T - y_C}{e_T - e_C} \cdot \frac{1}{1 - y_0}, \tag{2}$$

where T and C denote treatment and control groups (representing an instrument assignment such as having Fox News available via local cable or not), respectively. For $j \in \{T, C\}$, y_j is the share of group j adopting the target behavior (e.g. voting for a Republican candidate), and e_j is the share of group j exposed to persuasion. So, f is a rescaled version of the usual Wald statistic that estimates the local average treatment effect (LATE; see Imbens and Angrist, 1994), where the scaling factor depends on y_0, which represents the share of the converted (i.e. those who would take the action of interest even without exposure to persuasion). Here, y_0 is often unobserved, and DK propose using y_C in its place as an approximation.

Since DK first introduced the estimand f, it has been used and modified by many authors (e.g. Enikolopov, Petrova, and Zhuravskaya, 2011; Gentzkow, Shapiro, and Sinkinson, 2011; DellaVigna, Enikolopov, Mironova, Petrova, and Zhuravskaya, 2014; Bassi and Rasul, 2017; Martin and Yurukoglu, 2017). In their survey, DellaVigna and Gentzkow (2010) use f (or its approximation) as a key summary statistic to compare persuasive effects across different studies. However, we note that f might not be a proper conditional probability in a heterogeneous population and the approximation of f proposed by DK can even be larger than one. We will articulate under what assumptions f is reduced to θ and how we should interpret f in its relationship with θ in a general set-up. In this sense, we build on the work of DK but we add important clarifications to the literature.

Identification of θ is challenging for various reasons, including the fact that T_i is often difficult to observe and that it tends to be endogenous. For instance, even if subscriptions to *The Washington Post* can be randomized by design (Gerber, Karlan, and Bergan, 2009, GKB hereafter), the *actual* readership is difficult to observe and it can be highly correlated with the political orientation of agents. We will address the endogeneity issue by using an instrument Z_i under the assumption

[4] The idea of using conditional probability to define a parameter of interest can also be found in Heckman, Smith, and Clements (1997).

that $Y_i(1) \geq Y_i(0)$.[5] For the data issue on T_i, we note that some authors have used a micro-level survey to obtain auxiliary data on T_i, which motivates us to consider a few different data scenarios: the outcome and the treatment are jointly observed, they are separately observed, or the treatment is not observed at all. Under the three scenarios, we establish the sharp identified bounds of θ. Therefore, our work builds on the econometrics literature on partial identification (e.g. Manski, 2003, 2007; Tamer, 2010) as well as the literature on program evaluation (see Heckman and Vytlacil, 2007; Imbens and Wooldridge, 2009, for surveys of the literature).

The main findings of this paper are as follows. The persuasion rate θ is partially identified, and its sharp lower bound remains the same across the three data scenarios if the outcome is binary; the sharp lower bound depends only on the joint distribution of (Y_i, Z_i). We also obtain point and partial identification of the local persuasion rate (i.e. the persuasion rate for the group of compliers) under the three data scenarios. If a continuous instrument is available, then we can target a *marginal persuasion rate* that is akin to the marginal treatment effect (e.g. Heckman and Vytlacil, 2005). Therefore, having a continuous instrument opens up the possibility of point identification of the persuasion rate for the entire (or policy-relevant) population if the instrument is sufficiently rich. If the outcome is nonbinary, then we can condition on those who would not choose the outside option without the treatment. In this case, we show that the resulting lower bound is always no smaller than that of the binary outcome case.

The remainder of the paper is organized as follows. In Section 2, we discuss identification of θ, and in Section 3 we consider the case with nonbinary outcomes. In Section 4, we study the local and marginal versions of the persuasion rate. In Section 5, we discuss our recommendations on what to do in practice, including inferential issues. In Section 6, we provide an empirical illustration. We conclude in Section 7. The Appendices include additional results and examples, a detailed discussion about methods for inference, and all the proofs.

2. Identification of θ

Our objective is to quantify the causal effect of an informational treatment, such as watching a particular news channel, on convincing the agent to take a particular action of interest (e.g. Fox News promoting Republican candidates or *The Washington Post* supporting Democratic candidates). In this section, we assume that the potential outcomes, denoted by $Y_i(t)$, are binary (i.e. the agent taking the action of interest or not). Further, we assume that the treatment T_i is binary and

[5] In DK's study, they rely on the premise that Fox News availability via local cable in 2000 seems random after controlling for a set of covariates.

that a binary instrument Z_i is available. Later, we extend our results to the case where the potential outcomes are multinomial. The observed outcome is denoted by $Y_i = T_i Y_i(t) + (1 - T_i) Y_i(0)$, as mentioned earlier. We begin by making the following monotonicity assumption.

Assumption A (Monotonic Treatment Response)**.** $Y_i(0) \leq Y_i(1)$ *with probability one.*

Assumption A is a binary version of the monotonic treatment response (MTR) assumption used in Manski (1997) and Manski and Pepper (2000). Under MTR, a voter who would vote for a Republican candidate without watching Fox News would vote for a Republican with an exposure to Fox News with probability one. In Appendix A, we present a simple expected utility model to provide a microeconomic foundation of Assumption A. The following lemma shows that Assumption A has an important implication for $\theta = \mathbb{P}\{Y_i(1) = 1 | Y_i(0) = 0\}$.

Lemma 1. *Under Assumption A,*

$$\theta = \frac{\mathbb{P}\{Y_i(1) = 1\} - \mathbb{P}\{Y_i(0) = 1\}}{1 - \mathbb{P}\{Y_i(0) = 1\}}. \tag{3}$$

Therefore, identification of θ can be achieved by identifying two counterfactual probabilities $\mathbb{P}\{Y_i(1) = 1\}$ and $\mathbb{P}\{Y_i(0) = 1\}$ (i.e. we do not need to know the joint distribution of $Y_i(0)$ and $Y_i(1)$). In fact, as $Y_i(t)$ is binary, θ is the average treatment effect (ATE) divided by $\mathbb{P}\{Y_i(0) = 0\}$. The key point is that Assumption A makes it possible to obtain the *conditional probability* θ simply by rescaling the ATE.

The next assumption is concerned about the treatment assignment T_i and the instrument Z_i.

Assumption B (No Defiers and An Exogenous Instrument)**.** *T_i has a threshold structure, i.e.*

$$T_i = \mathbb{1}\{V_i \leq e(Z_i)\}, \tag{4}$$

where V_i is uniformly distributed, and $0 \leq e(0) < e(1) \leq 1$. Further, Z_i is independent of $\big(Y_i(t), V_i\big)$ for $t = 0, 1$.

Namely, the intent-to-treat (ITT) Z_i is randomly assigned; however, T_i can be endogenous via the dependence between V_i and $Y_i(t)$. The function $e(\cdot)$ is the propensity score or, more descriptively in our context, it can be referred to as the *exposure rate*. As Vytlacil (2002) has shown, the threshold structure in equation (4) is equivalent to assuming the absence of defiers.

In addition to Y_i, T_i, and Z_i, it is possible to observe covariates X_i but we suppress X_i in our identification analysis. In other words, we implicitly assume throughout the paper that all assumptions and results are conditional on the value of X_i.

As we discussed in Section 1, T_i can be difficult to observe and there are several possibilities to consider. The simplest is that everybody complies so that there is no difference between the actual treatment and the ITT; this case is referred to as the *sharp persuasion design*. When T_i and Z_i are different, which we call the *fuzzy persuasion design*, we consider three sampling scenarios: (i) (Y_i, T_i, Z_i) is jointly observed; (ii) (Y_i, Z_i) is observed and information about the persuasion rate is available from auxiliary data;[6] and (iii) (Y_i, Z_i) is observed and no other information is available.

In the following subsections, we investigate identification of θ in each of the above scenarios.[7] Let

$$\theta_L = \frac{\mathbb{P}(Y_i = 1 | Z_i = 1) - \mathbb{P}(Y_i = 1 | Z_i = 0)}{1 - \mathbb{P}(Y_i = 1 | Z_i = 0)}, \tag{5}$$

which is an identified parameter from the distribution of (Y_i, Z_i). It turns out that θ_L is the sharp lower bound of θ in each of the data scenarios described in the fuzzy persuasion design.

2.1. **The Sharp Persuasion Design.** We start with the simplest scenario in which everybody complies with the ITT.

Theorem 1. *Suppose that Assumptions A and B hold. If $e(1) - e(0) = 1$ (i.e. $T_i = Z_i$ with probability one), then we have $\mathbb{P}\{Y_i(z) = 1\} = \mathbb{P}(Y_i = 1 | Z_i = z)$, and hence $\theta = \theta_L$.*

The condition of $e(1) - e(0) = 1$ means that everybody is a complier, and hence there is essentially no difference between T_i and Z_i; so, T_i is observed and essentially randomized. However, this is rather an exceptional situation in social sciences. The key identification question should be how far we can go when the design is not sharp (i.e. not everybody is a complier). We answer this question in the following subsection.

2.2. **The Fuzzy Persuasion Design.** In the fuzzy design, we consider the three scenarios of data availability. We see that θ is only partially identified even in the most favorable case where the full joint distribution of (Y_i, T_i, Z_i) is available, and that the sharp lower bound is always given by θ_L.

2.2.1. *Identification with the Joint Distribution of (Y_i, T_i, Z_i).* Even the joint distribution of (Y_i, T_i, Z_i) does not generally tell us about the ATE: the only subpopulation we can learn from $Z = 0$ and $Z = 1$ in common is the one of compliers, so the Wald statistic estimates only the LATE, not the ATE. Therefore, θ cannot be point identified; however, we can derive its sharp bounds.

[6] For example, DK used town-level election data to estimate $\mathbb{P}(Y_i = 1 | Z_i = z)$ and micro-level media audience data to infer $e(z) = \mathbb{P}(T_i = 1 | Z_i = z)$.

[7] Throughout the discussion, we assume that T_i is correctly measured if it is observed. See Calvi, Lewbel, and Tommasi (2018), Nguimkeu, Denteh, and Tchernis (2019), and Ura (2018) for the issues of mismeasured treatment. Their subject matters are distinct from ours.

Assumption C (Full Observability)**.** *The joint distribution of* (Y_i, T_i, Z_i) *is known, where both* $e(0)$ *and* $1 - e(1)$ *are bounded away from zero.*

Theorem 2. *Suppose that Assumptions A to C are satisfied. Then, the sharp identified interval of* θ *is given by* $[\theta_L, \theta_U]$*, where* θ_L *is given in equation (5) and*

$$\theta_U = \frac{\mathbb{P}(Y_i = 1, T_i = 1 | Z_i = 1) - \mathbb{P}(Y_i = 1, T_i = 0 | Z_i = 0) + 1 - e(1)}{1 - \mathbb{P}(Y_i = 1, T_i = 0 | Z_i = 0)}.$$

To prove Theorem 2, we first derive the sharp identified bounds for $\mathbb{P}\{Y_i(1) = 1\}$ and $\mathbb{P}\{Y_i(0) = 1\}$ separately; we denote them by the intervals $[m_a, M_a]$ and $[m_b, M_b]$, respectively. These bounds are special cases of Manski and Pepper (2000) under the MTR assumption coupled with the exogeneity of the instrument. Then, we obtain the identified bounds of θ by solving

$$\max_{a,b} \text{ and } \min_{a,b} \frac{a - b}{1 - b} \quad \text{subject to} \quad a \in [m_a,\ M_a],\ b \in [m_b,\ M_b],\ a \geq b,$$

after which we appeal to continuity and the intermediate value theorem for the sharpness result.

The bounds in Theorem 2 shrink to a singleton as $\big(e(0), e(1)\big)$ approaches $(0, 1)$, which is consistent with the result in Theorem 1. Also, it is worth noting that the lower bound θ_L only depends on the distribution of (Y_i, Z_i): observing T_i along with (Y_i, Z_i) helps only for the upper bound. If $e(1)$ is too small, then the upper bound will not be very informative: θ_U converges to 1 as $e(1)$ approaches 0; that is, if nobody reads a newspaper when they receive free subscriptions, then we do not learn much about how "persuading" the newspaper is. However, even if $e(1)$ approaches 1, the upper bound does not necessarily shrink to the lower bound; i.e. we do not necessarily pin down the persuasion rate of reading the newspaper even if everybody who has free subscriptions actually reads it.

2.2.2. *Identification with the Knowledge of the Exposure Rates.* As in the case of DK, the researcher may not directly observe T_i but may have auxiliary data from which the exposure rates $e(1)$ and $e(0)$ can be estimated.[8] In this case, the sharp identified bounds of θ become generally wider than those of Theorem 2.

Assumption D (Observability of Two Marginals)**.** *Only the distribution of* (Y_i, Z_i) *and the exposure rates* $\{e(0), e(1)\}$ *are known.*

[8] The case in which the outcome and the treatment are separately observed belongs to an identification problem called the *ecological inference problem*. For instance, Cross and Manski (2002) and Manski (2018) discuss bounding a "long regression" by using information from a "short regression". Their substantive concerns are distinct from ours.

Theorem 3. *Suppose that Assumptions A, B and D are satisfied. Then, the sharp identified interval of θ is given by $[\theta_L,\ \theta_{U_e}]$, where θ_L is given in equation (5) and*

$$\theta_{U_e} = \frac{\min\{1, \mathbb{P}(Y_i = 1|Z_i = 1) + 1 - e(1)\} - \max\{0,\ \mathbb{P}(Y_i = 1|Z_i = 0) - e(0)\}}{1 - \max\{0,\ \mathbb{P}(Y_i = 1|Z_i = 0) - e(0)\}}. \tag{6}$$

Therefore, the upper bound in this case is nontrivial if and only if $e(1) > \mathbb{P}(Y_i = 1|Z_i = 1)$. Note that it is the relative size of the take-up rate $e(1)$ (i.e. the probability of reading a newspaper when a free subscription to it is offered) that determines how much we can hope to learn about the persuasion rate. The length of the identified interval becomes smaller as $e(1)$ approaches one for each value of $e(0)$.

Theorem 3 focuses on the case where $e(0)$ and $e(1)$ are known, but it is potentially useful even when the researcher's prior knowledge on them is just probabilistic (e.g. one can average out $e(0)$ and $e(1)$ with the prior).

2.2.3. *Identification with No Information Associated with T_i.* The final scenario is the least informative one. If the researcher has no information for $e(0)$ and $e(1)$, then the worst possibility from Theorem 3 is that the upper bound is equal to one. For the sake of completeness, we state this in a separate theorem.

Assumption E (Limited Observability)**.** *No information associated with T_i is available (i.e. the distribution of (Y_i, Z_i) is all that is known).*

Theorem 4. *Suppose that Assumptions A, B and E are satisfied. Then, the sharp bound of θ is given by $[\theta_L, 1]$, where θ_L is given in equation (5).*

3. The Persuasion Rate with Nonbinary Outcomes

When outcomes are not binary, one might want to treat an outside option slightly differently. In the voting example, those who would go out and vote even without any informational treatment may be the relevant subpopulation to consider in defining the rate of persuasion. Below we formalize this idea.

Suppose that $\mathcal{S} = \{0, 1, -1\}$, where 0 is an outside option, 1 is the target action of persuasion, and -1 represents taking any other action. For instance, taking action 0 can mean that the agent does not vote at all, whereas taking action 1 means that the agent votes for a candidate from party 1 and taking -1 means that the agent votes for a candidate from any other party. We then denote agent i's potential outcomes by the vector of binary variables $Y_i(t) = \big(Y_{i0}(t), Y_{i1}(t), Y_{i,-1}(t)\big)$ for $t \in$

$\{0,1\}$. Finally, we assume that the choices in $\mathcal{S}$ are exclusive and exhaustive so that $\sum_{j\in\mathcal{S}} Y_{ij}(t) = 1$ for $t \in \{0,1\}$.[9] Similarly to the binary case, we impose monotonicity on the target action of persuasion (i.e. $Y_{i1}(1) \geq Y_{i1}(0)$ with probability one). The following assumption summarizes our set-up.

Assumption F (Multinomial Outcomes). *$Y_{i1}(1) \geq Y_{i1}(0)$ and $\sum_{j\in\mathcal{S}} Y_{ij}(t) = 1$ for $t = 0,1$ with probability one.*

If we only focus on the target action $Y_{i1}(t)$, then the persuasion rate θ can be defined as in Section 2 (i.e. $\theta = \mathbb{P}\{Y_{i1}(1) = 1 | Y_{i1}(0) = 0\}$). However, if one wants to be explicit about the presence of the outside option, then conditioning on those who would not choose the outside option without the treatment seems appropriate to define the rate of persuasion; that is,

$$\theta_{\text{mult}} = \mathbb{P}\{Y_i(1) = (0,1,0) \mid Y_i(0) = (0,0,1)\} = \mathbb{P}\{Y_{i1}(1) = 1 \mid Y_{i0}(0) = 0, Y_{i1}(0) = 0\}, \quad (7)$$

where the second equality uses Assumption F. Note that θ_{mult} is a different parameter from θ, where θ_{mult} now measures the fraction of the people who would vote for the candidate of interest among those who would still vote but for somebody else without the persuasive treatment. The advantage of having a multinomial model is that we can pay attention to this extra layer of conditioning that comes from the presence of an outside option.

Similarly to the binary case, Assumption F enables us to express θ_{mult} in terms of the marginal distributions of the potential outcomes.

Lemma 2. *Under Assumption F, we have*

$$\theta_{\text{mult}} = \frac{\mathbb{P}\{Y_{i1}(1) = 1\} - \mathbb{P}\{Y_{i1}(0) = 1\}}{1 - \mathbb{P}\{Y_{i0}(0) = 1\} - \mathbb{P}\{Y_{i1}(0) = 1\}}.$$

Lemma 2 shows that the conditional probability θ_{mult} can be obtained by simply rescaling the ATE as before. One complication (compared with the binary case) is that we have an extra term in the denominator (i.e. $\mathbb{P}\{Y_{i0}(0) = 1\}$), which substantially complicates the identification analysis.

The assumption of $Y_{i1}(1) \geq Y_{i1}(0)$ is sufficient for Lemma 2, but it is not necessary. Indeed, Assumption F rules out the possibility of having $Y_i(1) = (1,0,0)$ and $Y_i(0) = (0,1,0)$, but this is an irrelevant event for θ_{mult} because θ_{mult} focuses only on the case where $Y_{i0}(t) = 0$ for both $t = 0,1$. However, Assumption F turns out to be quite convenient to obtain informative bounds of θ_{mult}.

[9] Here, we note that there is no loss of generality in assuming that $\mathcal{S}$ has only three options; if not, we can simply define $Y_{i,-1}(t) = \sum_{j\in\mathcal{S}\setminus\{0,1\}} Y_{ij}(t)$.

Assumption G (No Defiers and an Exogenous Instrument)**.** *T_i satisfies Assumption B. Further, for $t = 0, 1$ and $j \in \{0, 1, -1\}$, $\big(Y_{ij}(t), V_i\big)$ is independent of Z_i.*

Assumption G is a trivial extension of Assumption B. In Appendix C, we provide a detailed identification analysis of θ_{mult} under Assumption G in each of the three scenarios of data availability. However, we present here only the sharp lower bound of θ_{mult} when the joint distribution of (Y_i, T_i, Z_i) is available, as it seems to be the most useful result in practice. Here, $Y_i = (Y_{i0}, Y_{i1}, Y_{i,-1}) = T_i Y_i(1) + (1 - T_i) Y_i(0)$ is now a three-dimensional vector of observed binary variables.

Theorem 5. *Suppose that Assumptions C, F and G are satisfied. If $\mathbb{P}(Y_{i1} = 1 | Z_i = 1) + \mathbb{P}(Y_{i0} = 1, T_i = 0 | Z_i = 0) < 1$, then the lower bound of the sharp identified interval of θ_{mult} is given by*

$$\theta_{L,\text{mult}} = \frac{\mathbb{P}(Y_{i1} = 1 | Z_i = 1) - \mathbb{P}(Y_{i1} = 1 | Z_i = 0)}{1 - \mathbb{P}(Y_{i0} = 1, T_i = 0 | Z_i = 0) - \mathbb{P}(Y_{i1} = 1 | Z_i = 0)}.$$

Theorem 5 shows the sharp lower bound of θ_{mult} in the most favorable data scenario; the complete characterization of the sharp identified set of θ_{mult} in the other data scenarios can be found in Appendix C. The condition of $\mathbb{P}(Y_{i1} = 1 | Z_i = 1) + \mathbb{P}(Y_{i0} = 1, T_i = 0 | Z_i = 0) \geq 1$ represents an extreme situation, in which case θ_{mult} can be shown to be equal to 1. The intuition is clear; if there are too many people who do not vote while they are untreated, then there are too few people to "persuade" as θ_{mult} focuses only on the group of people who would vote even without the treatment. Including this trivial situation, the sharp lower bound $\theta_{L,\text{mult}}$ is always no less than that of the binary persuasion rate. Therefore, ignoring the presence of an outside option can lead to underestimating the persuasive effect of the treatment outside the people who do not vote without the treatment.

Unlike the lower bound θ_L in the binary case, $\theta_{L,\text{mult}}$ depends on the joint distribution of (Y_i, T_i) given Z_i. Therefore, if the sampling scheme does not allow the econometrician to access the full joint distribution, then even the lower bound will change. By applying a version of the Fréchet–Hoeffding bounds, we can derive the sharp identified bounds of θ_{mult} under the other two sampling schemes (i.e. Assumptions D and E). Not surprisingly, in the least informative case of Assumption E, the sharp lower bound of θ_{mult} becomes identical to that of θ (i.e. θ_L).

4. The Local and Marginal Persuasion Rates

Focusing on the binary case again, we consider the local and marginal persuasion rates that are defined as follows: for $0 < v < 1$,

$$\begin{cases} \theta_{\text{local}} = \mathbb{P}\{Y_i(1) = 1 | Y_i(0) = 0, e(0) < V_i \leq e(1)\}, \\ \theta_{\text{mte}}(v) = \mathbb{P}\{Y_i(1) = 1 | Y_i(0) = 0, V_i = v\}. \end{cases} \tag{8}$$

Here, θ_{local} is the persuasion rate for the subpopulation characterized by $e(0) < V_i \leq e(1)$, that is, the *compliers* (e.g. Imbens and Angrist, 1994), whereas $\theta_{\text{mte}}(v)$ is for the subpopulation such that $V_i = v$ (e.g. Heckman and Vytlacil, 2005).

First, we obtain identification results for θ_{local} under the three sampling scenarios in the fuzzy persuasion design. The first step for this purpose is to note that the same reasoning as Lemma 1 yields

$$\theta_{\text{local}} = \frac{\mathbb{E}\{Y_i(1) - Y_i(0) \mid e(0) < V_i \leq e(1)\}}{\mathbb{P}(Y_i(0) = 0 \mid e(0) < V_i \leq e(1)\}}, \tag{9}$$

where the numerator is the LATE, which has received great attention in the econometrics literature (see Deaton, 2010; Heckman, 2010; Imbens, 2010, for a recent debate). The denominator that rescales the LATE is also conditioned on the compliers so that we achieve proper conditioning again.

Just like the LATE, it is contentious whether or not θ_{local} should be the parameter of interest, because the compliers are concerned with an unidentified subgroup of the population. We take a practical view that the identification results on θ_{local} can complement the results obtained in Section 2. The following theorem shows the identification of θ_{local} under the three scenarios of data availability.

Theorem 6. *Suppose that Assumptions A and B are satisfied.*

(i) Under Assumption C, θ_{local} is point identified by $\theta_{local} = \theta^*$, where

$$\theta^* = \frac{\mathbb{P}(Y_i = 1 | Z_i = 1) - \mathbb{P}(Y_i = 1 | Z_i = 0)}{\mathbb{P}(Y_i = 0, T_i = 0 | Z_i = 0) - \mathbb{P}(Y_i = 0, T_i = 0 | Z_i = 1)}.$$

(ii) Under Assumption D, the sharp identified interval of θ_{local} is given by $[\theta_L^*, 1]$, where

$$\theta_L^* = \frac{\mathbb{P}(Y_i = 1 | Z_i = 1) - \mathbb{P}(Y_i = 1 | Z_i = 0)}{\min\{1 - \mathbb{P}(Y_i = 1 | Z_i = 0),\ e(1) - e(0)\}}.$$ [10]

(iii) Under Assumption E, the sharp identified interval of θ_{local} coincides with that of θ, i.e. $[\theta_L,\ 1]$.

[10] Alternatively, θ_L^* can be written as the maximum between θ_L and the probability limit of the Wald statistic.

The identification of the LATE requires the distribution of (T_i, Z_i) and that of (Y_i, Z_i) separately, but not the joint distribution of (Y_i, T_i, Z_i). Unlike the LATE, the point identification in Theorem 6(i) demands the knowledge of the joint distribution of (Y_i, T_i, Z_i).[11] Theorem 6(ii) shows that this requirement is not only sufficient but also necessary to achieve the point identification of θ_{local}.

The local persuasion rate θ_{local} represents the average persuasive effect for a population that is different from the entire population. Given this caveat, it is interesting to note that, in Theorem 6(ii), the upper bound of θ_{local} is always trivial in contrast to θ, but the lower bound of θ_{local} can never be worse than that of θ. Therefore, in principle, the length of the identified interval of θ can be smaller than that of θ_{local}. If T_i is not observed at all, then there is no advantage in focusing on the compliers. Theorem 6(iii) confirms the intuition that the bounds for θ_{local} are identical to those for θ if the distribution of (Y_i, Z_i) is the only piece of information available. This corresponds to an uninteresting case for θ_{local} though, as we have no information on compliers.

Data requirement for the identification of $\theta_{\text{mte}}(v)$ is generally quite demanding; if Y_i and T_i are jointly observed along with a *continuous* instrument Z_i, then $\theta_{\text{mte}}(v)$ can be point identified as in Heckman and Vytlacil (2005); Carneiro, Heckman, and Vytlacil (2011). Examples of continuous instruments can be found in the literature on the media effects on voting. For instance, Enikolopov, Petrova, and Zhuravskaya (2011) and DellaVigna, Enikolopov, Mironova, Petrova, and Zhuravskaya (2014) use the signal strength of NTV and Serbian radio as instruments, respectively; in both of the papers, (Y_i, T_i, Z_i) are jointly observed.

The following assumption describes the situation in which we can obtain point identification of $\theta_{\text{mte}}(v)$. We use the standard results in the literature (e.g. Heckman and Vytlacil, 2005) for the subsequent theorem.

Assumption H (Marginal Treatment Effects). *(i) The joint distribution of (Y_i, T_i, Z_i) is known.*

(ii) T_i has the threshold structure in equation (4), where V_i is uniformly distributed, and Z_i is independent of $\big(Y_i(t), V_i\big)$ for $t = 0, 1$.

(iii) The distribution of $e(Z_i)$ is absolutely continuous with respect to Lebesgue measure, where v is in the interior of the support of $e(Z_i)$.

Theorem 7. *Suppose that Assumptions A and H are satisfied. Then $\theta_{mte}(v)$ is point identified by*

$$\theta_{mte}(v) = \frac{\partial \mathbb{P}\{Y_i = 1 | e(Z_i) = e\} / \partial e \big|_{e=v}}{1 - \partial \mathbb{P}\{Y_i = 1, T_i = 0 | e(Z_i) = e\} / \partial e \big|_{e=v}}, \tag{10}$$

[11] The denominator of equation (9) requires that we know the marginal distribution of $Y_i(0)$ for the compliers. Imbens and Rubin (1997) show that the marginal distributions of $Y_i(1)$ and $Y_i(0)$ for the compliers are identified if the joint distribution of (Y_i, T_i, Z_i) is known; however, they did not consider the local persuasion rate.

provided that $\mathbb{P}\{Y_i = 1|e(Z_i) = e\}$ *and* $\mathbb{P}\{Y_i = 1, T_i = 0|e(Z_i) = e\}$ *are continuously differentiable with respect to* e.

Theorem 7 does not consider the other two scenarios of data availability. This is mainly because continuous instruments are rare in the context of persuasion and we are not aware of any applications where continuous instruments are available while the outcome and treatment are not jointly observed.

If the support of the exposure rate $e(Z_i)$ is equal to the unit interval $[0,1]$, then Theorem 7 shows the identification of $\theta_{\text{mte}}(v)$ for all v in the unit interval. Then, we can use $\theta_{\text{mte}}(v)$ to construct different policy-oriented quantities as in Heckman and Vytlacil (2005) and Carneiro, Heckman, and Vytlacil (2011). For instance, the persuasion rate of the entire population θ will be given by

$$\theta = \int_0^1 \theta_{\text{mte}}(v) dv,$$

because V_i is uniformly distributed on the unit interval.

5. Discussion: Main Takeaways and Inferential Issues

In this section, we focus on the binary outcome/binary instrument case to discuss the main takeaways from our identification results as well as issues on estimation and inference; the same principles also apply to the multinomial case.

The population version of DK's proposal to approximate f is

$$\tilde{\theta}_{DK} = \frac{\mathbb{P}(Y_i = 1|Z_i = 1) - \mathbb{P}(Y_i = 1|Z_i = 0)}{e(1) - e(0)} \frac{1}{1 - \mathbb{P}(Y_i = 1|Z_i = 0)}. \tag{11}$$

Note that $\tilde{\theta}_{DK}$ would be a well-defined conditional probability if there were no exposure measures $e(1), e(0)$ and Y_i were replaced with T_i, which is equivalent to the sharp persuasion design.[12] However, it turns out that $\tilde{\theta}_{DK}$ may not be a conditional probability in general: see Section 6 for numerical demonstration. Therefore, this paper clarifies identification issues when we insert exposure as a choice variable and employ a proper causal framework that is used in policy evaluation to model two causal links (i.e. $Z_i \to T_i$ and $T_i \to Y_i$).[13]

[12] The only difference is now relabeling Y_i, T_i with T_i, Z_i, respectively.

[13] We are grateful to an anonymous referee who provided us with insightful comments.

In the most favorable data scenario, we recommend reporting $[\theta_L, \theta_U]$ along with θ^*; these can be consistently estimated by their sample analogs. If (Y_i, Z_i) is observed with some auxiliary information for $e(1)$ and $e(0)$, then $[\theta_L, \theta_{U_e}]$ and $[\theta_L^*, 1]$ should be reported. If T_i is not observed at all, then θ_L is the only parameter we can hope to learn about.

Note that θ_L should *always* be estimated as it only requires data on (Y_i, Z_i). Because the actual T_i can be difficult to observe, some authors, such as DK and GKB, have used a micro-level survey to obtain data on T_i, which seems quite costly. However, the value of an attempt to observe T_i can be limited, depending on which parameter the researcher wants to learn about. For instance, if the researcher cares about the persuasion rate of the entire population, then the value of observing T_i is only in tightening the upper bound. If the group of compliers is of interest, then whether we observe T_i or not, and how we observe it, can be relevant issues; we have $\theta_L^* \geq \theta_L$ in the second data scenario and θ^* is point identified if (Y_i, T_i, Z_i) is jointly observed. If Z_i is continuously distributed, the value of observing (Y_i, T_i, Z_i) jointly increases dramatically as well. In summary, our identification analysis shows that the value of observing T_i depends crucially on which population is of interest to the researcher.

TABLE 1. Persuasion Rates: Papers on Voter Turnout

Paper	$\hat{y}(1)$	$\hat{y}(0)$	$\hat{e}(1) - \hat{e}(0)$	$\hat{\tilde{\theta}}_{DK}$	$\hat{\theta}_L$	$\hat{\theta}_L^*$
	(1)	(2)	(3)	(4)	(5)	(6)
Green and Gerber (2000)	47.2	44.8	27.9	15.6	4.3	8.6
Green, Gerber, and Nickerson (2003)	31.0	28.6	29.3	11.5	3.4	8.2
Green and Gerber (2001)	71.1	66.0	73.7	20.4	15.0	15.0
Green and Gerber (2001)	41.6	40.5	41.4	4.5	1.8	2.7
Gentzkow (2006)	45.5	43.5	80.0	4.4	3.5	3.5
Gentzkow, Shapiro, and Sinkinson (2011)	70.0	69.0	25.0	12.9	3.2	4.0

[a] The outcome variable is voter turnout except for Gentzkow (2006), where exposure to television discouraged voters to go to the polls. Thus, to have positive persuasive effects in all rows, the outcome variable for Gentzkow (2006) is not to vote.
[b] In columns (1) and (2), $\hat{y}(t)$ denotes estimates of $\mathbb{P}(Y_i = 1 | Z_i = t)$ for $t = 1, 0$.
[c] $\tilde{\theta}_{DK}$ is the persuasion rate reported in DellaVigna and Gentzkow (2010, table 1).
[d] θ_L and θ_L^* are the sharp lower bounds of the average and local persuasion rates, respectively.
[e] The third row corresponds to the row in table 1 of DellaVigna and Gentzkow (2010) under the treatment labeled "phone calls by youth vote".
[f] The fourth row corresponds to the row in table 1 of DellaVigna and Gentzkow (2010) under the treatment labeled "phone calls 18–30-year-olds".
[g] Each entry in the table is a percentage.

To illustrate the importance of measuring the effect of persuasion properly, we compare DK's persuasion rates with our lower bounds in Table 1. Specifically, we focus on the results reported in DellaVigna and Gentzkow (2010, their table 1) when the outcome variable is voter turnout. We have chosen this type of study as the turnout is among the most studied outcome variables in the literature and it is naturally a binary measure. It can be seen that DK's persuasion rates are higher than the lower bound of θ or that of θ_{local}, thereby suggesting that the persuasive effects might not be as large as those reported in DellaVigna and Gentzkow (2010, table 1).

As we are under partial identification, inference should also account for that. The method proposed by Stoye (2009) is useful for that purpose, at least in the most favorable data scenario, in which case the sample analog principle and the Delta method show that we can construct the estimators $\hat{\theta}_L$ and $\hat{\theta}_U$ that are asymptotically jointly normal. Therefore, by Stoye (2009), a $(1-\alpha)$ confidence interval for θ can be obtained by $[\hat{\theta}_L - c_\alpha \hat{\sigma}_L, \hat{\theta}_U + c_\alpha \hat{\sigma}_U]$, where $\hat{\sigma}_L$ and $\hat{\sigma}_U$ are the estimated standard errors of $\hat{\theta}_L$ and $\hat{\theta}_U$, respectively, and c_α is chosen by solving

$$\Phi\Big(c_\alpha + \frac{\hat{\Delta}}{\max(\hat{\sigma}_L, \hat{\sigma}_U)}\Big) - \Phi(-c_\alpha) = 1 - \alpha,$$

where Φ is the distribution function of the standard normal and $\hat{\Delta}$ is the estimated length of the identified interval.

The second data scenario is slightly more complicated, because θ_{U_e} and θ_L^* contain the min or max function that is not smooth; so, the Delta method does not apply. In Appendices F and G, we propose a two-step method for inference to overcome this problem, which we have applied to the empirical example we discuss in Section 6. In the third data scheme, confidence intervals for θ and θ^* always coincide and they can be obtained by using a one-side critical value on $\hat{\theta}_L$. Specifically, they are given by $[\hat{\theta}_L - z_{1-\alpha}\hat{\sigma}_L, 1]$, where $z_{1-\alpha}$ is the $(1-\alpha)$ quantile of the standard normal distribution. Please refer to Appendices F and G for a more detailed discussion on inference.

6. An Empirical Example

In this section, we illustrate our proposed methods using data from GKB, who report findings from a field experiment to measure the effect of political news. We have chosen this example because it contains a credible binary instrument from the field experiment and we can also illustrate all of the three sampling scenarios as well as the case of nonbinary outcomes. In GKB, there are three statuses in the intention to treat: a control group, an offer of free subscription to *The Washington Post*, and one to *The Washington Times*. To illustrate the usefulness of our paper, we focus on *The*

Washington Post and drop all observations from *The Washington Times* subscription. That is, $Z_i = 1$ if the ith individual received free subscription to *The Washington Post*, and $Z_i = 0$ if not.

Focusing on the ITT analysis, GKB have reported ITT estimates for various outcomes Y_i. DellaVigna and Gentzkow (2010) compute persuasion rates for GKB, for which they simply set $T_i = 1$ if the ith individual opted into the free subscription and $T_i = 0$ if they opted out of it. In this section, for the purpose of illustrating our identification results, we consider a different treatment variable: $T_i = 1$ if the ith individual reads a newspaper at least several times per week and $T_i = 0$ otherwise, which is a variable that GKB kept track of in a follow-up survey. Therefore, the relevant treatment we consider differs from that of DellaVigna and Gentzkow (2010), but it is whether individuals have *actually* read the newspaper or not. The outcome variables we consider are as follows. For the binary case, $Y_i = 1$ if the ith individual reported voting for the Democratic candidate in the 2005 gubernatorial election, and $Y_i = 0$ if the subject did not vote for the Democratic candidate or did not vote at all. For the multinomial case, not voting at all is treated as an outside option. We use only a subsample of the GKB data with those who responded to the follow-up survey to use information on (Y_i, T_i, Z_i) jointly. After dropping observations for *The Washington Times* subscription and removing missing data, we summarize the GKB data in Table 2. Although the joint distribution of (Y_i, T_i, Z_i) is observed in this example, we also consider using the two marginals of (Y_i, Z_i) and (T_i, Z_i) separately, to make a comparison. The estimates are summarized in Table 3. Because the size of the sample extract we use is relatively modest ($n = 701$) for an interval-identified object, we report the 80% confidence intervals obtained by the inference methods described in Section 5 as well as in Appendices F and G.

First, we discuss the case where the full joint distribution of (Y_i, T_i, Z_i) is used. In this data scenario, the average effect of persuasion by reading the newspaper is bounded between 7% and 63%. In contrast, the persuasion rate for the group of compliers is point estimated by 81%. It is interesting to note that the estimate of θ_{local} is so large that it is greater than the upper bound of θ. This suggests that individuals are highly heterogeneous in this example, indicating that $\tilde{\theta}_{DK}$ might not be a well-defined conditional probability here. Indeed, the estimate of $\tilde{\theta}_{DK}$ in equation (11) is $\hat{\tilde{\theta}}_{DK} = 1.1027$, which is greater than one.

When the marginals of (Y_i, Z_i) and (T_i, Z_i) are used separately, the upper bound of θ increases from 63% to 78%. Further, θ_{local} is no longer point estimated but we only know that it is bounded between 78% and 100%. This difference illustrates the loss of identification power if we do not observe the joint distribution of (Y_i, T_i, Z_i).

TABLE 2. Summary Statistics of the GKB Data

The Washington Post ($Z_i = 1$)			
	Reads a newspaper		Total
Voted for Democrat	$T_i = 0$	$T_i = 1$	
$Y_i = 0$	94	93	187
$Y_i = 1$	31	68	99
Total	125	161	286
Control ($Z_i = 0$)			
	Reads a newspaper		Total
Voted for Democrat	$T_i = 0$	$T_i = 1$	
$Y_i = 0$	162	130	292
$Y_i = 1$	46	77	123
Total	208	207	415

The GKB data are used after dropping observations for *The Washington Times* subscription and removing missing data.

TABLE 3. Estimates of the Key Parameters

	(Y_i, T_i, Z_i)	(Y_i, Z_i) and (T_i, Z_i)	(Y_i, Z_i) only
ITT	0.0498		
	$\{0.0036,\ 0.0959\}$		
θ	$[0.0707,\ 0.6343]$	$[0.0707,\ 0.7832]$	$[0.0707,\ 1]$
	$\{0.0289,\ 0.6610\}$	$\{0.0286,\ 0.8143\}$	$\{0.0288,\ 1\}$
LATE	0.7759		-
	$\{0,\ 1\}$		
θ_{local}	0.8067	$[0.7759,\ 1]$	$[0.0707,\ 1]$
	$\{0.1243,\ 1\}$	$\{0.0069,\ 1\}$	$\{0.0288,\ 1\}$

[a] The first and third rows show the estimates of ITT and LATE, respectively. The second row corresponds to $[\hat{\theta}_L, \hat{\theta}_U]$, $[\hat{\theta}_L, \hat{\theta}_{U_e}]$, and $[\hat{\theta}_L, 1]$, respectively. The fourth row shows $\hat{\theta}^*$, $[\hat{\theta}_L^*, 1]$, and $[\hat{\theta}_L, 1]$, respectively.
[b] 80% confidence intervals are given in curly braces.

Finally, we estimate the lower bound of θ_{mult} in equation (7) with the outside action of not voting at all. By additionally conditioning on those who would vote even without reading the newspaper, the estimated lower bound of the average persuasion increases from 0.0707 (0.0289) to 0.0975 (0.0554), where the numbers in the parentheses are the left-end point of the 80% confidence intervals. Therefore, in terms of the lower bound, the persuasive effect increases from 5% in the ITT analysis to 7% in the average persuasion rate, and to 10% if we further focus on those who would vote without reading the newspaper.

7. Conclusions

We have set up a simple econometric model of persuasion, we have introduced several parameters of interest, and we have analyzed their identification. The empirical example in Section 6 as well as the examples in Appendices D and E demonstrate that the persuasive effects are highly heterogeneous in the settings of media and fund raising. We have focused on the case of binary treatments. In applications, treatments could be multivalued (e.g. watching Fox News, CNN or MSNBC). It would be fruitful to build on recent developments in multiple treatments (e.g. Heckman and Pinto, 2018; Lee and Salanié, 2018) to investigate identification of persuasive effects. It would also be interesting to estimate deep parameters in an economic model of persuasion by using a more structural approach in the set-up of multiple treatments. These are topics of future research.

Appendices to "Identifying the Effect of Persuasion"

The appendices of the paper are grouped into four parts. Part I includes additional results that are omitted from the main text. In appendix A, we present a simple economic model to motivate our setup. Focusing on a binary treatment and a binary outcome, we formulate a model of persuasion within the framework of expected utility maximization. This formulation naturally leads to a potential outcome setup with a certain monotonicity restriction. In appendix B, we clarify the difference between f defined in equation (2) and θ as well as the relationship between f and θ_{local}. In appendix C, we give further results on identification with nonbinary outcomes.

Part II provides additional empirical examples. In appendix D, we revisit the empirical literature on the effects of news media on voting, where we apply our identification results to two published articles. In particular, when we revisit DK using their original data, we find that the identification region for the persuasion rate θ is between 1% and 99% and that the lower bound for the local persuasion rate θ_{local} is either 12% or 37%, depending on the specification of the fixed effects. These results suggest that the persuasive effect of Fox News is fairly large for compliers, i.e. those who would watch Fox News if and only if it is randomly available, but that DK's data are uninformative about the general population. In appendix E, we look at the literature on door-to-door fund raising and we illustrate the usefulness of our results by applying them to two published papers.

Part III deals with estimation and inference problems. In appendix F, we explain methods for Inference on the average persuasion rate; in appendix G, we describe methods for Inference on the local persuasion rate. In appendix H, we consider semiparametrically efficient estimation of the two key parameters, i.e. the lower bounds of θ and θ_{local}, and we provide an empirical illustration.

Part IV, which is appendix I, contains all the proofs.

PART I.

APPENDIX A. A MICROECONOMIC FOUNDATION FOR ASSUMPTION A

In this section, we consider a binary choice problem under binary states. The states are unknown to the agent at the time of the decision and the agent relies on her subjective belief about them to make a decision. Suppose that two possible states are denoted by $S \in \mathscr{S} = \{\text{High}, \text{Low}\}$. Let $T_i \in \{0,1\}$ indicate individual i's status of the informational treatment. Further, let $q_i(t)$ describe individual i's subjective belief about the state when T_i is set to $t \in \{0,1\}$: i.e. $q_i(t) = \mathbb{P}(S = \text{High} | T_i = t, I_i)$, where I_i denotes all other information available to individual i. Table A.1 describes the utility individual i receives from each choice conditional on the state. The payoffs matrix in table A.1 is from Bergemann and Morris (2019, see matrix (5)).

TABLE A.1. Utility by choice and state

	$S = \text{Low}$	$S = \text{High}$
Vote (1)	-1	$U_i \geq 0$
Not vote (0)	0	0

The utility from option 0 is normalized to be 0 for each state. Since the expected utility is all that matters for the decision, the utility from option 1 when the state is "low" is normalized to be -1: the sign restrictions are to make the choice nontrivial. The utility term U_i is not observed by the econometrician.

Suppose that individual i maximizes her expected utility. Then, individual i chooses option 1 if and only if her expected utility, $-\{1 - q_i(t)\} + q_i(t)U_i$ for $t \in \{0,1\}$, is positive with her belief $q_i(t)$ about the state. Therefore, when the informational treatment is set to be $t \in \{0,1\}$, the potential outcome $Y_i(t)$ can be written as follows:

$$Y_i(t) = \mathbb{1}\big[-\{1 - q_i(t)\} + q_i(t)U_i \geq 0\big], \tag{A.1}$$

where $\mathbb{1}[\cdot]$ is the usual indicator function. We now make the following assumptions.

Assumption A.1. *U_i has a conditional density $f\{\cdot|q_i(0), q_i(1)\}$ such that $f\{u|q_i(0), q_i(1)\} > 0$ for all $u \in [0, \infty)$ with probability one.*

Assumption A.2. *$q_i(0) \leq q_i(1)$ with probability one.*

Assumption A.1 says that U_i is continuously distributed given $q_i(0), q_i(1)$. However, it does not rule out the possibility that U_i and $q_i(t)$ are dependent on each other. Assumption A.2 simply means that the informational treatment may shift an agent's belief *only in one direction*.

Lemma A.1. *Under assumption A.1, assumption A.2 is equivalent to $Y_i(0) \leq Y_i(1)$ with probability one.*

Therefore, lemma A.1 provides a microeconomic foundation for assumption A.

Appendix B. Measuring Persuasive Effects in the Literature

We now discuss the relationship between our parameters of persuasive effects and the ones that were used in the literature. For this purpose we focus on the case with binary outcomes and a binary instrument.

The population version of DK's proposal f is

$$\theta_{DK} = \frac{\mathbb{P}(Y_i = 1|Z_i = 1) - \mathbb{P}(Y_i = 1|Z_i = 0)}{e(1) - e(0)} \frac{1}{1 - \mathbb{P}\{Y_i(0) = 1\}}, \tag{A.2}$$

which is often approximated by

$$\tilde{\theta}_{DK} = \frac{\mathbb{P}(Y_i = 1|Z_i = 1) - \mathbb{P}(Y_i = 1|Z_i = 0)}{e(1) - e(0)} \frac{1}{1 - \mathbb{P}(Y_i = 1|Z_i = 0)}. \tag{A.3}$$

Note here that $\tilde{\theta}_{DK}$ does not require any knowledge about the joint distribution of (Y_i, T_i) given Z_i.

We now discuss the relationship between $\theta, \theta_{\text{local}}$, and θ_{DK}. By equation (A.24) in the proof of theorem 6, we have

$$\mathbb{P}\{Y_i(0) = 0\}\theta_{DK} = \mathbb{P}\{Y_i(0) = 0|e(0) < V_i \leq e(1)\}\theta_{\text{local}} = \mathbb{E}\{Y_i(1) - Y_i(0)|e(0) < V_i \leq e(1)\}$$

under assumptions A and B. Further, recall from lemma 1 that

$$\mathbb{P}\{Y_i(0) = 0\}\theta = \mathbb{E}\{Y_i(1) - Y_i(0)\}.$$

Therefore, $\theta, \theta_{\text{local}}$, and θ_{DK} are all different parameters in general. For example, θ_{DK} rescales the LATE with an unconditional probability, and hence it does not render a well–defined conditional probability in general.

There are some special cases where the three parameters coincide. For example, $\theta = \theta_{DK}$ holds if and only if the ATE equals the LATE. This happens, for example, if at least one of the following three conditions holds:

(i) the entire population consists of compliers, i.e. $e(0) = 0$ and $e(1) = 1$, as in the sharp persuasion design;

(ii) $Y_i(1) - Y_i(0)$ is a constant;

(iii) V_i is independent of $\big(q_i(t), U_i\big)$ for $t = 0, 1$, in which case the potential outcome $Y_i(t)$ is independent of T_i conditional on Z_i.

Condition (ii) corresponds to the situation with no heterogeneity in the treatment effect. This is probably the least interesting condition because there are only two unrealistic possibilities for this: either $Y_i(1) - Y_i(0) = 1$ (everyone is persuaded) or $Y_i(1) - Y_i(0) = 0$ (no one has room for persuasion). Under condition (i), there is no difference between the intent–to–treat and the actual treatment, in which case randomizing the intent–to–treat is sufficient to identify θ. Condition (iii) is often referred to as the condition of *unconfoundedness or selection on observables* in econometrics. Since $\mathbb{P}\{Y_i(0) = 1\} = \mathbb{P}\{Y_i(0) = 1 \mid e(0) < V_i \leq e(1)\}$ under conditions (i) or (iii), we have $\theta = \theta_{\text{local}} = \theta_{DK}$ under either of the two conditions.

Unlike the three parameters, $\tilde{\theta}_{DK}$ generally does not measure the effect of persuasion even under condition (iii). However, as DK correctly pointed out, it is an approximation of $\theta_{DK} (= \theta = \theta_{\text{local}})$ when $e(0)$ is close to zero or $\theta = 0$.[14] $\tilde{\theta}_{DK}$ has some interesting features: observing the two marginals of (Y_i, Z_i) and (T_i, Z_i) is sufficient for its identification, and it has a simple lower bound θ_L that can be identified without observing T_i at all. Indeed, DellaVigna and Gentzkow (2010) extensively reports $\tilde{\theta}_{DK}$ or its lower bound θ_L, depending on whether T_i is observed or not. However, $\tilde{\theta}_{DK}$ should be interpreted with caution: θ_L is *always* a meaningful estimand but $\tilde{\theta}_{DK}$ is not. When information about $e(0)$ and $e(1)$ is available, it seems a better practice to report θ_L together with θ_L^* than to estimate $\tilde{\theta}_{DK}$.

It is worth pointing out that θ_L is not only a lower bound of $\tilde{\theta}_{DK}$ but also the sharp lower bound of θ in a much more general sense. Specifically, neither condition (iii) nor approximation by $\tilde{\theta}_{DK}$ is needed, and therefore the bound is robust to both the presence of endogeneity in the treatment assignment and poor approximation of θ_{DK} by $\tilde{\theta}_{DK}$.

Finally, as an aside we point out that without condition (iii), θ_{DK} does not measure the persuasion rate of any subpopulation correctly: the first factor on the right–hand side of equation (A.2) focuses

[14] Under condition (iii), we have $\mathbb{P}(Y_i = 1|Z_i = z) = \mathbb{P}(Y_i = 1, T_i = 1|Z_i = z) + \mathbb{P}(Y_i = 1, T_i = 0|Z_i = z) = \mathbb{P}\{Y_i(0) = 1\} + \big[\mathbb{P}\{Y_i(1) = 1\} - \mathbb{P}\{Y_i(0) = 1\}\big] e(z)$.

on a subpopulation of "compliers," while the second factor is not conditioned on the complier group.

Appendix C. Further Results on Identification with Nonbinary Outcomes

In this section we provide further results on identification with nonbinary outcomes. In order to present the results, we introduce the following notation. Let

$$p_j(y|z) = \mathbb{P}(Y_{ij} = y|Z_i = z) \quad \text{and} \quad p_j(y,t|z) = \mathbb{P}(Y_{ij} = y, T_i = t|Z_i = z). \tag{A.4}$$

The first theorem is a complete characterization of the sharp identified set of θ_{mult} when (Y_i, T_i, Z_i) is jointly observed.

Theorem A.1. *Suppose that assumptions C, F and G are satisfied. Then, the sharp identified set of θ_{mult} is given as follows.*

(i) If $p_1(1|1) + p_0(1,0|0) \geq 1$, then $\theta_{\text{mult}} = 1$;

(ii) If $p_1(1|1) + p_0(1,0|0) < 1 \leq p_1(1,1|1) + p_0(1,0|0) + 1 - e(1) + e(0)$, then

$$\frac{p_1(1|1) - p_1(1|0)}{1 - p_0(1,0|0) - p_1(1|0)} \leq \theta_{\text{mult}} \leq 1.$$

(iii) If $p_1(1,1|1) + p_0(1,0|0) + 1 - e(1) + e(0) < 1$, then

$$\frac{p_1(1|1) - p_1(1|0)}{1 - p_0(1,0|0) - p_1(1|0)} \leq \theta_{\text{mult}} \leq \frac{p_1(1,1|1) + 1 - e(1) - p_1(1,0|0)}{1 - p_0(1,0|0) - p_1(1,0|0)}.$$

We now present the results under assumption D.

Theorem A.2. *Suppose that assumptions F and G are satisfied. Further, suppose that the econometrician observes the distribution of (Y_i, Z_i) with the knowledge of the exposure rates $e(0)$ and $e(1)$. Then, the sharp identified set of θ_{mult} is given as follows.*

(i) If $p_1(1|1) + \max\{0,\ p_0(1|0) - e(0)\} \geq 1$, then $\theta_{\text{mult}} = 1$;

(ii) If $p_1(1|1) + \max\{0,\ p_0(1|0) - e(0)\} < 1 \leq \min\{p_1(1|1),\ e(1)\} + \min\{p_0(1|0),\ 1 - e(0)\} + 1 - e(1) + e(0)$, then

$$\max\left\{\frac{p_1(1|1) - p_1(1|0)}{1 - p_1(1|0)}, \frac{p_1(1|1) - p_1(1|0)}{1 - p_0(1|0) + e(0) - p_1(1|0)}\right\} \leq \theta_{\text{mult}} \leq 1.$$

(iii) If $\min\{p_1(1|1),\ e(1)\} + \min\{p_0(1|0),\ 1 - e(0)\} + 1 - e(1) + e(0) < 1$, then

$$\max\left\{\frac{p_1(1|1) - p_1(1|0)}{1 - p_1(1|0)}, \frac{p_1(1|1) - p_1(1|0)}{1 - p_0(1|0) + e(0) - p_1(1|0)}\right\} \leq \theta_{\text{mult}}$$

$$\leq \frac{\min\{p_1(1|1),\ e(1)\} + 1 - e(1) - \max\{0,\ p_1(1|0) - e(0)\}}{1 - \min\{p_0(1|0),\ 1 - e(0)\} + e(0) - \max\{0,\ p_1(1|0) - e(0)\}}.$$

We finally give the result under the assumption that T_i is not observed at all.

Theorem A.3. *Suppose that assumptions F and G are satisfied. Further, suppose that the econometrician observes the distribution of (Y_i, Z_i) without the knowledge of the exposure rates $e(0)$ and $e(1)$. Then, the sharp identified set of θ_{mult} is the same as the binary case, i.e.*

$$\frac{p_1(1|1) - p_1(1|0)}{1 - p_1(1|0)} \leq \theta_{\text{mult}} \leq 1.$$

Therefore, unlike the binary case, the sharp lower bound of θ_{mult} depends on the sampling scheme. If T_i is not observed at all, then observing the choices of the outside option does not help to infer the persuasion rate with the extra conditioning on those who would participate without the treatment.

PART II.

APPENDIX D. ADDITIONAL EXAMPLES: THE EFFECTS OF MEDIA ON VOTING

In this section we revisit the recent empirical literature on the effects of media on voting and apply our identification results.

D.1. **The Effect of Fox News: DellaVigna and Kaplan (2007) Revisited.** In DK, the entry of Fox News in cable markets plays a role of an instrument conditional on a set of covariates. That is, Z_i is a binary variable that equals one if Fox News was part of local cable package in the town where the i^{th} individual was living in 2000. To apply our result to DK, let Y_i be the binary dependent variable that equals one if individual i voted for the Republican candidate in the 2000 presidential election. As DK argue in their paper, Fox News availability in 2000 is likely to be idiosyncratic, only after controlling for a set of covariates. We will be explicit about conditioning on covariates X_i to apply our identification results, and we write the lower bound as a function of the values of X_i: i.e.

$$\tilde{\theta}_L(x) = \frac{\mathbb{P}(Y_i = 1|Z_i = 1, X_i = x) - \mathbb{P}(Y_i = 1|Z_i = 0, X_i = x)}{1 - \mathbb{P}(Y_i = 1|Z_i = 0, X_i = x)}, \tag{A.5}$$

which is the sharp lower bound of $\mathbb{P}\{Y_i(1) = 1|Y_i(0) = 0, X_i = x\}$, the conditional persuasion rate. Then, to obtain the lower bound for the persuasion rate in the population, we integrate (A.5) with respect to the distribution F_X of X_i, so that

$$\theta_L = \int \tilde{\theta}_L(x) dF_X(x).$$

Note that X_i is first controlled for and is averaged out.

To estimate θ_L, we use DK's data[15] and adopt similar specifications as in DK. They estimated $\mathbb{P}(Y_i = 1|Z_i, X_i)$ using a town–level linear regression model, where the dependent variable is the Republican two–party vote share for the 2000 presidential election minus the same variable for the 1996 election. To be consistent with our econometric framework, we modify the dependent variable to be the votes cast for the Republican candidate in the 2000 presidential election divided by the population of age 18 and older. Recall that in our setup, $Y_i = 0$ if individual i did not voted for the Republican candidate. This event includes the case of voting for different candidates or that of not voting for any candidate at all. As the town–level covariates, we include the Republican vote share as a share of the voting–age population in the 1996 election, census controls for both 1990 and 2000, cable system controls, and U.S. House district fixed effects (or county fixed effects). These specifications correspond to the main specifications of DK (see columns (4) and (5) of table IV in DK). In the regression, the town–level observations are weighted by the population of age 18 and older in 1996.

DK used two different data sources for (Y_i, Z_i) and (T_i, Z_i). Hence, we can look at the upper bound for θ and the lower bound for θ_{local} using these. Again, making use of the covariates explicitly, we rewrite (6) as

$$\theta_{U_e} = \int \tilde{\theta}_{U_e}(x) dF_X(x),$$

where $\tilde{\theta}_{U_e}(x)$ equals

$$\frac{\min\{1, \mathbb{P}(Y_i = 1|Z_i = 1, X_i = x) + 1 - e(1,x)\} - \max\{0,\ \mathbb{P}(Y_i = 1|Z_i = 0, X_i = x) - e(0,x)\}}{1 - \max\{0,\ \mathbb{P}(Y_i = 1|Z_i = 0, X_i = x) - e(0,x)\}},$$

and $e(z,x) = \mathbb{P}(T_i = 1|Z_i = z, X_i = x)$. We also re-write the bounds in part (ii) of Theorem 6 as

$$\int \frac{\mathbb{P}(Y_i = 1|Z_i = 1, X_i = x) - \mathbb{P}(Y_i = 1|Z_i = 0, X_i = x)}{\min\{1 - \mathbb{P}(Y_i = 1|Z_i = 0, X_i = x),\ e(1,x) - e(0,x)\}} dF_X\big(x|e(0) < V_i \le e(1)\big) \le \theta_{\text{local}} \le 1. \quad \text{(A.6)}$$

By Bayes' theorem, the lower bound in (A.6) can be re-written as

$$\begin{aligned} &\int \frac{\mathbb{P}(Y_i = 1|Z_i = 1, X_i = x) - \mathbb{P}(Y_i = 1|Z_i = 0, X_i = x)}{\min\{1 - \mathbb{P}(Y_i = 1|Z_i = 0, X_i = x),\ e(1,x) - e(0,x)\}} dF_X(x|e(0) < V_i \le e(1)) \\ &= \int \frac{\mathbb{P}(Y_i = 1|Z_i = 1, X_i = x) - \mathbb{P}(Y_i = 1|Z_i = 0, X_i = x)}{\min\{1 - \mathbb{P}(Y_i = 1|Z_i = 0, X_i = x),\ e(1,x) - e(0,x)\}} \frac{e(1,x) - e(0,x)}{\int \{e(1,x) - e(0,x)\} dF_X(x)} dF_X(x), \end{aligned}$$

[15] The data used in DK are available at http://eml.berkeley.edu/~sdellavi/index.html.

which can be estimated directly from data.

DK estimated $e(z, x)$ using the microlevel Scarborough data on television audiences. We focus on "diary audience" measure in DK[16] and take the same specifications as in columns (2) and (3) of table VIII from DK.

TABLE A.2. Persuasion Rates: Fox News Effects

	(1) U.S. House district fixed effects	(2) County fixed effects
θ	[0.005,0.991]	[0.011,0.992]
θ_{local}	[0.115,1]	[0.370,1]

Table A.2 summarizes our empirical results.[17] Column (1) shows estimation results when U.S. House district fixed effects are controlled for and column (2) displays corresponding results for county fixed effects.

The bounds for θ are wide and uninformative. However, the lower bounds for θ_{local} are sizable and also comparable to the estimates of the persuasion rates reported in DK (0.11 and 0.28, respectively). In sum, we conclude that the persuasive effect of Fox News seems fairly large for the compliers, that is, those who would watch the Fox News channel if and only if it is randomly available, although the data do not say much about the entire population.

D.2. **The NTV Effect: Enikolopov, Petrova, and Zhuravskaya (2011) Revisited.** As mentioned earlier, Enikolopov, Petrova, and Zhuravskaya (2011, EPZ hereafter) used a continuous instrument, i.e. the signal strength of NTV, to measure the persuasive effect of watching NTV (the anti-Putin TV station) on a parliamentary election in 1999. Further, in the individual–level survey data in EPZ, (Y_i, T_i, Z_i) are jointly observed. Therefore, in this subsection, we apply the identification result of the marginal persuasion rate to this example using the EPZ data.

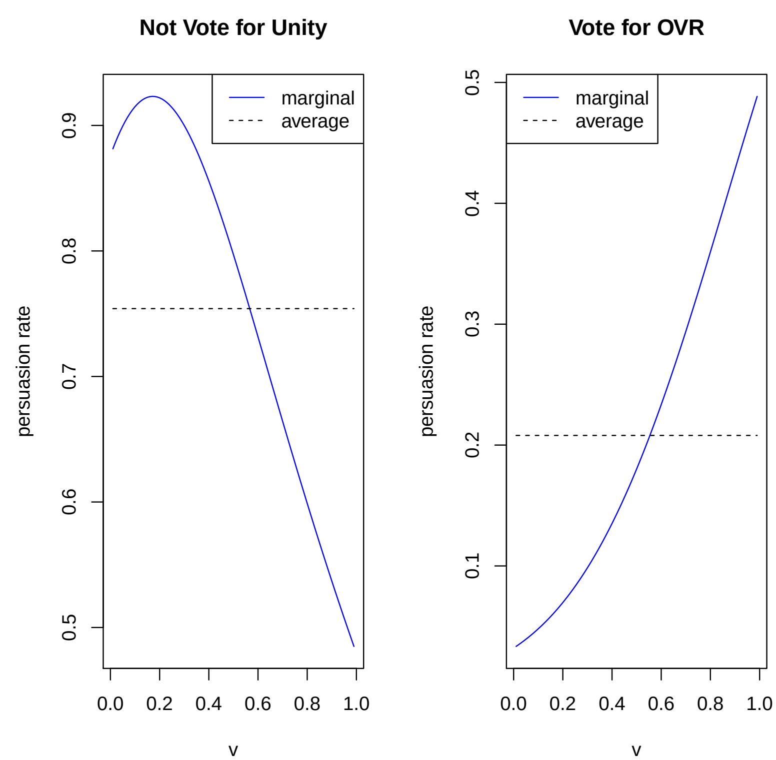
We look at two parties: the progovernment party "Unity" and the most popular opposition party OVR ("Fatherland–All Russsia"). During the 1999 election campaign, Unity was opposed by NTV, while OVR were supported by NTV. Thus, EPZ presumed a negative persuasion rate for voting for Unity but a positive persuasion rate for OVR. To be consistent with our theoretical framework

[16]The microlevel Scarborough data contain the "recall" measure regarding whether a respondent watched a given channel in the past seven days and the "diary" measure on whether a respondent watched a channel for at least one full half-an-hour block according to the seven-day diary.

[17]To estimate the unconditional bounds reported in the table, the conditional ones are weighted by the number of respondents in a town for the Scarborough data. In addition, the predicted probabilities are truncated to be between 0 and 1.

and other empirical examples, Y_i is either $Y_{\text{Unity},i}$ or $Y_{\text{OVR},i}$, depending on which party we consider. Specifically, we let $Y_{\text{Unity},i} = 1$ if an individual did not vote for Unity and $Y_{\text{Unity},i} = 0$ otherwise; $Y_{\text{OVR},i} = 1$ if an individual voted for OVR and $Y_{\text{OVR},i} = 0$ otherwise. As in the previous section, it is necessary to condition on covariates. We take the baseline covariates as in columns (1) and (2) of table 6 and table 7 in EPZ. They include individual characteristics such as gender, age, marital status, and education, and subregional variables such as population size and average wage.

FIGURE A.1. Estimates of Marginal and Average Persuasion Rates



Notes: The left and right panels of the figure show estimates of the marginal and average persuasion rates for not voting for Unity and voting for OVR, respectively.

For the sake of simplicity, we estimate θ_{mte} parametrically. The population conditional probabilities, $e(z, x)$, $\mathbb{P}(Y_i = 1 | e(Z_i) = e, X_i = x)$ and $\mathbb{P}(Y_i = 1, T_i = 0 | e(Z_i) = e, X_i = x)$, are estimated by probit,[18] and the conditional estimates of equation (10) are averaged out with respect to covariates by sample survey weight.

[18]The exposure rate $e(z, x)$ is first estimated and its predicted values are included linearly as a regressor to estimate the other two conditional probabilities.

Figure A.1 presents the estimation results. In the left panel, $\theta_{\text{mte}}(v)$ and θ are plotted as a function of v, when the outcome variable is not to vote for Unity. It can be seen that the marginal persuasive rate is about 90% at $v = 0.1$ but just 54% for $v = 0.9$. In view of equation (4), V_i can be interpreted as the unobserved cost of watching NTV. The estimation results suggest that the persuasive effect for not voting for Unity is much stronger for those whose unobserved cost of watching NTV is lower. In the right panel, corresponding results are shown for OVR. In this case, the persuasive effect is much weaker for those with lower values of v.

A striking patten we can learn from figure A.1 is that persuasive effects are highly heterogenous. This may partially answer the puzzle reported in EPZ. They found relatively modest positive persuasive effects for opposition parties but much stronger persuasive effects for Unity using aggregate voting outcomes, while the magnitudes are similar using individual survey data.[19] Our estimation results indicate that the marginal persuasive effects are highly heterogenous, thereby implying that different aggregate averages can be substantially different from each other. The average persuasive effect θ is plotted as a horizontal line in each panel of figure A.1: it is 75.4% against Unity and 20.8% for OVR. In short, this application exemplifies the identification power of continuous instruments that can uncover the patterns of heterogeneity in persuasive effects.

Appendix E. Additional Examples: Door-to-Door Fundraising

Landry, Lange, List, Price, and Rupp (2006) and DellaVigna, List, and Malmendier (2012) designed field experiments of door–to–door fund raising to examine various aspects of charity giving. In this section, we use their data to illustrate the usefulness of our identification results.

The common data structure in both papers is that for each type of experimental treatments, we observe (Y_i, T_i, Z_i):

- $Y_i = 1$ if a household made a contribution to door–to–door fund raising,
- $T_i = 1$ if a household answered the door and spoke to a solicitor,
- $Z_i = 1$ if a household was approached by a solicitor.

If $Z_i = 0$ (a household was not approached by a solicitor), then $T_i = 0$ and furthermore it is very likely that $Y_i = 0$. Hence, in this section, we assume that $\mathbb{P}(Y_i = 1, T_i = 0 | Z_i = 0) = \mathbb{P}(Y_i = 1 | Z_i = 0) = 0$. In addition, we assume that if $Y_i = 1$, it must be the case that $T_i = 1$. In other words, we assume that it is impossible to have both $Y_i = 1$ and $T_i = 0$ (making a contribution without answering the door). Thus, $\mathbb{P}(Y_i = 1, T_i = 1 | Z_i = 1) = \mathbb{P}(Y_i = 1 | Z_i = 1)$. These assumptions were

[19] EPZ estimated the persuasion rate using a continuous version of DK. See equations (3) and (4) in EPZ for their formulae.

also used in computation of the persuasion rates for donors in DellaVigna and Gentzkow (2010). Under these assumptions, we have the bound for θ as

$$\theta_L = \mathbb{P}(Y_i = 1|Z_i = 1) \quad \text{and} \quad \theta_U = \mathbb{P}(Y_i = 1|Z_i = 1) + 1 - e(1).$$

In addition,

$$\theta_{\text{local}} = \mathbb{P}(Y_i = 1|Z_i = 1)/e(1);$$

θ_{local} is the same as the usual LATE.

E.1. **Landry, Lange, List, Price, and Rupp (2006) Revisited.** In this study, there were four treatments: VCM (voluntary contributions mechanism), VCM with seed money, single-prize lottery, and multiple-prize lottery. Using Table II of Landry, Lange, List, Price, and Rupp (2006), we compute the persuasive effects by treatment and report results in table A.3.

TABLE A.3. Persuasive Effect by Treatment in Landry, Lange, List, Price, and Rupp (2006)

Treatment	$\mathbb{P}(Y_i = 1\|Z_i = 1)$	$e(1)$	θ_L	θ_U	θ_{local}
VCM	9.5%	37.6%	9.5%	71.9%	25.3%
VCM with seed money	5.2%	35.3%	5.2%	69.9%	14.8%
Single-prize lottery	17.1%	37.7%	17.1%	79.4%	45.5%
Multiple-prize lottery	12.6%	35.2%	12.6%	77.5%	35.9%
All	10.8%	36.3%	10.8%	74.5%	29.7%

Based on the lower bound and the LATE parameter, it seems that the single–prize lottery is the most effective fund raising tool, whereas the VCM with seed money is the least effective. However, the identification regions for θ of all four treatments overlap and there is no clear ranking based on those. This suggests that if one cares about the persuasive effect for the population, the evidence is inconclusive.

E.2. **DellaVigna, List, and Malmendier (2012) Revisited.** In their study of charity giving, DellaVigna, List, and Malmendier (2012, DLM hereafter) designed both fund raising and survey treatments to test for altruism and social pressure in charity giving. In this section, we focus only on three fund raising treatments: namely, the baseline treatment, the flyer treatment, and the opt-out treatment. The baseline treatment is the standard door–to–door funding raising campaign, the flyer treatment is with the flyer that provided information on fund raising the date before the solicitation, and the opt–out treatment is with the flyer that had an additional feature of a “Do Not Disturb” checkbox.

There were two charities in each of the fund raising treatments: La Rabida Children's Hospital and the East Carolina Hazard Center.

TABLE A.4. Persuasive Effect by Treatment in DLM

Treatment	$\mathbb{P}(Y_i = 1 \mid Z_i = 1)$	$e(1)$	θ_L	θ_U	θ_{local}
La Rabida Children's Hospital					
Baseline	7.1%	40.5%	7.1%	66.6%	17.5%
Flyer	6.8%	36.4%	6.8%	70.4%	18.8%
Opt–Out	5.4%	30.4%	5.4%	74.9%	17.7%
East Carolina Hazard Center					
Baseline	4.7%	43.0%	4.7%	61.7%	10.9%
Flyer	5.1%	39.6%	5.1%	65.5%	12.9%
Opt–Out	3.0%	34.4%	3.0%	68.6%	8.6%

DLM pointed out that treatments were randomized within a date–solicitor time block and estimated linear probability models with covariates: solicitor fixed effects, date–town fixed effects, hourly time block fixed effects, and area rating dummies. We use the same specification as in DLM, estimate $\mathbb{P}(Y_i = 1 \mid Z_i = 1, X_i = x)$ and $e(1, x)$, and then average out the conditional estimates as in appendix D.1.[20] The resulting estimates are reported in table A.4, where we report the persuasive effect by treatment/charity.

The local persuasion rate is point identified and is higher for the in–state charity, La Rabida Children's Hospital. The estimates of θ_{local} are the highest for the flyer treatment in both charities. This does not mean that the flyer treatment is the most effective in fund raising for the general population. Note that the compliers of the baseline treatment are different from those of the flyer treatment. For example, it could be the case that households at the margin of giving might have decided to not answer the door after they noticed the flyer. Unlike θ_{local}, θ_L and θ_U are comparable across different treatments. However, as in the previous section, it is difficult to see whether there is a significant difference across treatments if we focus on the bounds for θ.[21]

[20] The data collected in DLM are available at `http://eml.berkeley.edu/~sdellavi/index.html`. As before, the predicted probabilities are truncated to be between 0 and 1, when they are averaged out.
[21] In addition to the fund raising treatments, DLM relied on survey treatments and structural estimates to draw conclusions in their paper.

Part III.

Appendix F. Inference on the Average Persuasion Rate

In this section, we provide methods for carrying out inference on the average persuasion rate, for which we assume that the data are independent and identically distributed.

F.1. **The case of Theorem 2.** In this subsection, we consider Theorem 2, where the sharp identified interval of θ is given by $[\theta_L, \theta_U]$. Recall that

$$\theta_L = \frac{\mathbb{P}(Y_i = 1|Z_i = 1) - \mathbb{P}(Y_i = 1|Z_i = 0)}{1 - \mathbb{P}(Y_i = 1|Z_i = 0)},$$

$$\theta_U = \frac{\mathbb{P}(Y_i = 1, T_i = 1|Z_i = 1) - \mathbb{P}(Y_i = 1, T_i = 0|Z_i = 0) + 1 - e(1)}{1 - \mathbb{P}(Y_i = 1, T_i = 0|Z_i = 0)}.$$

To define the sample analog estimators of θ_L and θ_U, define

$$\widehat{\mathbb{P}}(Y_i = 1|Z_i = 1) = \frac{\sum_{i=1}^n 1(Y_i = 1, Z_i = 1)}{\sum_{i=1}^n 1(Z_i = 1)},$$

$$\widehat{\mathbb{P}}(Y_i = 1|Z_i = 0) = \frac{\sum_{i=1}^n 1(Y_i = 1, Z_i = 0)}{\sum_{i=1}^n 1(Z_i = 0)},$$

$$\widehat{\mathbb{P}}(Y_i = 1, T_i = 1|Z_i = 1) = \frac{\sum_{i=1}^n 1(Y_i = 1, T_i = 1, Z_i = 1)}{\sum_{i=1}^n 1(Z_i = 1)},$$

$$\widehat{\mathbb{P}}(Y_i = 1, T_i = 0|Z_i = 0) = \frac{\sum_{i=1}^n 1(Y_i = 1, T_i = 0, Z_i = 0)}{\sum_{i=1}^n 1(Z_i = 0)},$$

$$\widehat{e}(1) = \frac{\sum_{i=1}^n 1(T_i = 1, Z_i = 1)}{\sum_{i=1}^n 1(Z_i = 1)}.$$

Then we define

$$\widehat{\theta}_L = \frac{\widehat{\mathbb{P}}(Y_i = 1|Z_i = 1) - \widehat{\mathbb{P}}(Y_i = 1|Z_i = 0)}{1 - \widehat{\mathbb{P}}(Y_i = 1|Z_i = 0)},$$

$$\widehat{\theta}_U = \frac{\widehat{\mathbb{P}}(Y_i = 1, T_i = 1|Z_i = 1) - \widehat{\mathbb{P}}(Y_i = 1, T_i = 0|Z_i = 0) + 1 - \widehat{e}(1)}{1 - \widehat{\mathbb{P}}(Y_i = 1, T_i = 0|Z_i = 0)}.$$

Since $\widehat{\theta}_L$ and $\widehat{\theta}_U$ are asymptotically jointly normal, we follow Imbens and Manski (2004) and Stoye (2009) to construct a confidence interval for θ. Standard arguments based on the delta method yield that $\widehat{\theta}_L$ and $\widehat{\theta}_U$ have asymptotically linear approximations:

$$\widehat{\theta}_L - \theta_L = n^{-1} \sum_{i=1}^n \varphi_{L,i} + o_p(n^{-1/2}),$$

$$\widehat{\theta}_U - \theta_U = n^{-1} \sum_{i=1}^n \varphi_{U,i} + o_p(n^{-1/2}),$$

where $\varphi_{L,i}$ and $\varphi_{U,i}$ are influence functions that can be approximated by the following sample analogs:

$$
\begin{aligned}
\widehat{\varphi}_{L,i} = & \frac{1}{[1-\widehat{\mathbb{P}}(Y_i=1|Z_i=0)]}\frac{1}{\widehat{\mathbb{P}}(Z_i=1)}\left\{1(Y_i=1,Z_i=1)-\widehat{\mathbb{P}}(Y_i=1,Z_i=1)\right\} \\
& -\frac{1}{[1=\widehat{\mathbb{P}}(Y_i=1|Z_i=0)]}\frac{\widehat{\mathbb{P}}(Y_i=1|Z_i=1)}{\widehat{\mathbb{P}}(Z_i=1)}\left\{1(Z_i=1)-\widehat{\mathbb{P}}(Z_i=1)\right\} \\
& +\frac{\widehat{\theta}_L-1}{[1-\widehat{\mathbb{P}}(Y_i=1|Z_i=0)]}\frac{1}{\widehat{\mathbb{P}}(Z_i=0)}\left\{1(Y_i=1,Z_i=0)-\widehat{\mathbb{P}}(Y_i=1,Z_i=0)\right\} \\
& -\frac{\widehat{\theta}_L-1}{[1-\widehat{\mathbb{P}}(Y_i=1|Z_i=0)]}\frac{\widehat{\mathbb{P}}(Y_i=1|Z_i=0)}{\widehat{\mathbb{P}}(Z_i=0)}\left\{1(Z_i=0)-\widehat{\mathbb{P}}(Z_i=0)\right\}, \\
\widehat{\varphi}_{U,i} = & \frac{1}{[1-\widehat{\mathbb{P}}(Y_i=1,T_i=0|Z_i=0)]}\frac{1}{\widehat{\mathbb{P}}(Z_i=1)} \\
& \times\left\{1(Y_i=1,T_i=1,Z_i=1)-\widehat{\mathbb{P}}(Y_i=1,T_i=1,Z_i=1)+1(T_i=0,Z_i=1)-\widehat{\mathbb{P}}(T_i=0,Z_i=1)\right\} \\
& -\frac{1}{[1-\widehat{\mathbb{P}}(Y_i=1,T_i=0|Z_i=0)]}\frac{\widehat{\mathbb{P}}(Y_i=1,T_i=1,Z_i=1)+\mathbb{P}(T_i=0,Z_i=1)}{\widehat{\mathbb{P}}(Z_i=1)} \\
& \times\left\{1(Z_i=1)-\widehat{\mathbb{P}}(Z_i=1)\right\} \\
& +\frac{\widehat{\theta}_U-1}{[1-\widehat{\mathbb{P}}(Y_i=1,T_i=0|Z_i=0)]}\frac{1}{\widehat{\mathbb{P}}(Z_i=0)}\left\{1(Y_i=1,T_i=0,Z_i=0)-\widehat{\mathbb{P}}(Y_i=1,T_i=0,Z_i=0)\right\} \\
& -\frac{\widehat{\theta}_U-1}{[1-\widehat{\mathbb{P}}(Y_i=1,T_i=0|Z_i=0)]}\frac{\widehat{\mathbb{P}}(Y_i=1,T_i=0|Z_i=0)}{\widehat{\mathbb{P}}(Z_i=0)}\left\{1(Z_i=0)-\widehat{\mathbb{P}}(Z_i=0)\right\}.
\end{aligned}
$$

Now define

$$
\widehat{\sigma}_L^2 = n^{-2}\sum_{i=1}^n \widehat{\varphi}_{L,i}^2, \quad \widehat{\sigma}_U^2 = n^{-2}\sum_{i=1}^n \widehat{\varphi}_{U,i}^2, \quad \text{and} \quad \widehat{\Delta} = \widehat{\theta}_U - \widehat{\theta}_L.
$$

That is, $\widehat{\sigma}_L$ and $\widehat{\sigma}_U$ are standard errors of $\widehat{\theta}_U$ and $\widehat{\theta}_L$, respectively, and $\widehat{\Delta}$ is the estimated length of the identification region. Let

$$
\mathrm{CI}_\alpha^{\text{Theorem 2}} = \left[\widehat{\theta}_L - c_\alpha\widehat{\sigma}_L, \widehat{\theta}_U + c_\alpha\widehat{\sigma}_U\right], \tag{A.7}
$$

where c_α solves

$$
\Phi\left(c_\alpha + \frac{\widehat{\Delta}}{\max\{\widehat{\sigma}_L, \widehat{\sigma}_U\}}\right) - \Phi\left(-c_\alpha\right) = 1-\alpha.
$$

Since $\widehat{\theta}_L \leq \widehat{\theta}_U$ by construction, Lemma 3 and Proposition 1 of Stoye (2009) imply that $\theta \in \mathrm{CI}_\alpha^{\text{Theorem 2}}$ with probability $1-\alpha$ uniformly as $n \to \infty$, provided that the data generating process satisfies mild regularity conditions given in Assumption 1 (i) and (ii) of Stoye (2009).

F.2. **The case of Theorem 3.** Recall that in this case, the sharp identified interval of θ is given by $[\theta_L,\ \theta_{U_e}]$, where

$$\theta_{U_e} = \frac{\min\{1, \mathbb{P}(Y_i = 1|Z_i = 1) + 1 - e(1)\} - \max\{0,\ \mathbb{P}(Y_i = 1|Z_i = 0) - e(0)\}}{1 - \max\{0,\ \mathbb{P}(Y_i = 1|Z_i = 0) - e(0)\}}.$$

It is convenient to introduce additional notation. Let

$$\xi_1 = \mathbb{P}(Y_i = 1|Z_i = 1) + 1 - e(1) \quad \text{and} \quad \xi_2 = \mathbb{P}(Y_i = 1|Z_i = 0) - e(0).$$

Then ξ_1 and ξ_2 can be estimated by their sample analogs as before: i.e.

$$\widehat{\xi}_1 = \widehat{\mathbb{P}}(Y_i = 1|Z_i = 1) + 1 - \widehat{e}(1) \quad \text{and} \quad \widehat{\xi}_2 = \widehat{\mathbb{P}}(Y_i = 1|Z_i = 0) - \widehat{e}(0).$$

Furthermore, let $\widehat{\sigma}(\xi_1)$ and $\widehat{\sigma}(\xi_2)$ denote respective standard errors. For any $\alpha \in (0, 1/2)$, define z_α such that $\Phi(z_\alpha) = 1 - \alpha$. Hence, $z_{\alpha/2}$ is the two-sided standard normal critical value. We now choose $\overline{\alpha}$ that is close to zero, say $\overline{\alpha} = 0.001$, by which we define the following three possibilities.

(i) $\widehat{\xi}_1 - z_{\overline{\alpha}/4} \cdot \widehat{\sigma}(\xi_1) \geq 1$;

(ii) $\widehat{\xi}_1 + z_{\overline{\alpha}/4} \cdot \widehat{\sigma}(\xi_1) \leq 1$ and $\widehat{\xi}_2 + z_{\overline{\alpha}/4} \cdot \widehat{\sigma}(\xi_2) \leq 0$;

(iii) neither (i) nor (ii).

The critical value $z_{\overline{\alpha}/4}$ is used here to reflect the fact that the two equalities are tested jointly against both positive and negative directions, which can be viewed as pretesting. Case (i) corresponds to the case that $\widehat{\xi}_1$ is much larger than 1, implying that the upper bound is 1. Then we recommend using the confidence set such that $\left[\widehat{\theta}_L - z_{\alpha-\overline{\alpha}}\widehat{\sigma}_L, 1\right]$. To accommodate the error in the pretesting stage, we recommend using the $z_{\alpha-\overline{\alpha}}$ critical value here. Case (ii) suggests that $\widehat{\xi}_1$ is much smaller than 1 and $\widehat{\xi}_2$ is sufficiently less than zero. Therefore, in this subcase, the upper bound reduces to $\theta_{U_e} = \xi_1$. Then the estimators of both lower and upper bounds are asymptotically linear, implying that one can use the confidence interval similar to (A.7). Again, to accommodate the error in the pretesting stage, we recommend using the $z_{\alpha-\overline{\alpha}}$ critical value in applying (A.7). This ensures that the asymptotic coverage probability is at least as large as $1 - \alpha$ in applying (A.7). It turns out that case (ii) was the relevant case for the empirical example reported section 6.

Case (iii) is more complicated. We rely on a simple projection method to construct a valid confidence set. It follows from the proofs of lemma A.5 and theorem 3 that the two end points of the sharp identified interval of θ have the form:

$$\max_{a,b\in[0,1]^2} \text{ and } \min_{a,b\in[0,1]^2} \frac{a-b}{1-b} \quad \text{s.t.} \quad \begin{cases} a \geq b, \\ a \in \left[\mathbb{P}(Y_i = 1|Z_i = 1),\ \mathbb{P}(Y_i = 1|Z_i = 1) + 1 - e(1)\right], \\ b \in \left[\mathbb{P}(Y_i = 1|Z_i = 0) - e(0),\ \mathbb{P}(Y_i = 1|Z_i = 0)\right]. \end{cases} \tag{A.8}$$

Rewrite the constraints above as

$$a \in \left[\mathbb{E}(\mathbb{1}\{Y_i = 1\}|Z_i = 1),\ \mathbb{E}(\mathbb{1}\{Y_i = 1\}|Z_i = 1) + \mathbb{E}(\mathbb{1}\{T_i = 0\}|Z_i = 1)\right] \equiv [m_a, M_a],$$
$$b \in \left[\mathbb{E}(\mathbb{1}\{Y_i = 1\}|Z_i = 0) - \mathbb{E}(\mathbb{1}\{T_i = 1\}|Z_i = 0),\ \mathbb{E}(\mathbb{1}\{Y_i = 1\}|Z_i = 0)\right] \equiv [m_b, M_b].$$

Since a and b are bounded by conditional expectations, it is simple to construct a joint confidence set for a and b in the form of the Cartesian product (e.g. using a Bonferroni correction). That is, we can find the confidence set such that the following holds asymptotically with probability at least $1-\alpha$:

$$a \in \left[\hat{m}_a,\ \hat{M}_a\right] \quad \text{and} \quad b \in \left[\hat{m}_b,\ \hat{M}_b\right].$$

Consequently, the following optima include the true identified set of θ defined in equation (A.8), asymptotically with probability at least $1-\alpha$:

$$\max_{a,b\in[0,1]^2} \text{ and } \min_{a,b\in[0,1]^2} \frac{a-b}{1-b} \quad \text{s.t.} \quad \begin{cases} a \geq b, \\ a \in \left[\hat{m}_a,\ \hat{M}_a\right], \\ b \in \left[\hat{m}_b,\ \hat{M}_b\right]. \end{cases}$$

F.3. **The case of Theorem 4.** In this case, the sharp bound of θ is given by $[\theta_L, 1]$. Thus, the one-sided confidence interval for θ_L provides the valid confidence set for θ. That is, $\theta \in \left[\widehat{\theta}_L - z_\alpha \widehat{\sigma}_L, 1\right]$ holds asymptotically with probability at least $1-\alpha$, where z_α satisifies $\Phi(z_\alpha) = 1-\alpha$.

Appendix G. Inference on the Local Persuasion Rate

G.1. **Case under assumption C.** Recall that in this case, θ_{local} is point identified by $\theta_{\text{local}} = \theta^*$, where

$$\theta^* = \frac{\mathbb{P}(Y_i = 1|Z_i = 1) - \mathbb{P}(Y_i = 1|Z_i = 0)}{\mathbb{P}(Y_i = 0, T_i = 0|Z_i = 0) - \mathbb{P}(Y_i = 0, T_i = 0|Z_i = 1)}.$$

Thus, in this case, one can use the standard delta method to construct the two-sided confidence interval based on the asymptotic normality of the sample analog estimator of θ^*. Let $\widehat{\theta}^*$ and $\widehat{\sigma}^*$

denote the sample analog estimator of θ^* and its standard error, respectively. Then the random interval $[\widehat{\theta}^* - z_{\alpha/2}\widehat{\sigma}^*, \widehat{\theta}^* + z_{\alpha/2}\widehat{\sigma}^*]$ includes θ_{local} asymptotically with probability $1-\alpha$.

G.2. **Case under assumption D.** The sharp identified interval of θ_{local} is given by $[\theta_L^*, 1]$, where $\theta_L^* = \max(\text{Wald}, \theta_L)$ with

$$\text{Wald} = \frac{\mathbb{P}(Y_i = 1 | Z_i = 1) - \mathbb{P}(Y_i = 1 | Z_i = 0)}{e(1) - e(0)}.$$

Thus, the lower bound of θ_{local} has the form of the intersection bound (Chernozhukov, Lee, and Rosen, 2013). To describe how to conduct inference, define the sample analog estimator of $\widehat{\text{Wald}}$:

$$\widehat{\text{Wald}} = \frac{\widehat{\mathbb{P}}(Y_i = 1 | Z_i = 1) - \widehat{\mathbb{P}}(Y_i = 1 | Z_i = 0)}{\widehat{e}(1) - \widehat{e}(0)}.$$

Let $\widehat{\sigma}_{\text{Wald}}$ be its standard error. Then, the random interval $[\widehat{\theta}_L^*(\alpha), 1]$ includes θ_{local} with probability at least $1-\alpha$, where

$$\widehat{\theta}_L^*(\alpha) = \max\{0, \widehat{\text{Wald}} - z_{\alpha/2}\widehat{\sigma}_{\text{Wald}}, \widehat{\theta}_L - z_{\alpha/2}\widehat{\sigma}_L\}.$$

Here, the critical value $z_{\alpha/2}$ is based on simple Bonferroni correction. An adaptive inequality selection proposed in Chernozhukov, Lee, and Rosen (2013) can be adopted to construct a sharper critical value than $z_{\alpha/2}$.

G.3. **Case under assumption E.** In this case, the sharp identified interval of θ_{local} coincides with that of θ. Hence, one can use the inference method presented in appendix F.3.

Appendix H. Semiparametric Estimation

H.1. **Efficient Estimation.** In this section, we are explicit about the vector X_i of exogenous covariates and consider semiparametrically efficient estimation of the two key parameters, i.e. θ_L and θ^*.[22] We focus on independent and identically distributed (i.i.d.) data again. For θ_L we work with the dataset $\{(Y_i, Z_i, X_i^\intercal)^\intercal : i = 1, 2, \cdots n\}$, whereas we use $\{(Y_i, T_i, Z_i, X_i^\intercal)^\intercal : i = 1, 2, \cdots n\}$ for θ^*. Since we are now explicit about X_i, the objects of interest will be defined by integrating X_i out: when we analyze the local persuasion rate, we will use the conditional distribution of X_i given the group of compliers. Recall that $\tilde{\theta}_L(x)$ is defined in (A.5). Let

$$\tilde{\theta}^*(x) = \frac{\mathbb{P}(Y_i = 1 | Z_i = 1, X_i = x) - \mathbb{P}(Y_i = 1 | Z_i = 0, X_i = x)}{\mathbb{P}(Y_i = 0, T_i = 0 | Z_i = 0, X_i = x) - \mathbb{P}(Y_i = 0, T_i = 0 | Z_i = 1, X_i = x)},$$

[22] Efficient estimation of θ_L^* or θ_{U_e} is significantly more challenging because of the min function. It also requires a careful construction of an estimator for the marginal persuasion rate. These are beyond the scope of the current paper but interesting topics for future research.

which correspond to the local persuasion rate conditional on $X_i = x$. Now, we average out X_i with respect to an appropriate distribution to obtain the parameters we consider estimating in this section: i.e.

$$\theta_L = \int \tilde{\theta}_L(x) dF(x), \tag{A.9}$$

$$\theta^* = \int \tilde{\theta}^*(x) dF\{x | e(0, X_i) < V_i \leq e(1, X_i)\}, \tag{A.10}$$

where $F(\cdot)$ and $F(\cdot|\cdot)$ denote the marginal and conditional distribution function of X_i, respectively, and $e(j,x) = \mathbb{P}(T_i = 1 | Z_i = j, X_i = x)$ for $j = 0, 1$ is the exposure rate conditional on $X_i = x$: we will let $f(\cdot)$ denote the density of X_i. By Bayes' theorem θ^* can be alternatively written as

$$\theta^* = \int \frac{\tilde{\theta}^*(x)\{e(1,x) - e(0,x)\} f(x)}{\mathbb{P}\{e(0,X_i) < V_i \leq e(1,X_i)\}} dx = \frac{\int \tilde{\theta}^*(x)\{e(1,x) - e(0,x)\} f(x) dx}{\int \{e(1,x) - e(0,x)\} f(x) dx}. \tag{A.11}$$

Here, we emphasize that θ^* is defined by averaging out X_i conditional on the event that individual i is a complier—since the local persuasion rate is meaningful only for the group of compliers, it would not make sense to average over the entire population.

Semiparametric estimation of θ_L and θ^* is straightforward. Let

$$\hat{\theta}_L = \frac{1}{n} \sum_{i=1}^{n} \hat{\tilde{\theta}}(X_i) \quad \text{and} \quad \hat{\theta}^* = \frac{\sum_{i=1}^{n} \hat{\tilde{\theta}}(X_i)\{\hat{e}(1,X_i) - \hat{e}(0,X_i)\}}{\sum_{i=1}^{n} \{\hat{e}(1,X_i) - \hat{e}(0,X_i)\}}, \tag{A.12}$$

where $\hat{e}(j,x)$ is a consistent estimator of $e(j,x)$ for $j = 0, 1$, $\hat{\tilde{\theta}}_L(x)$ and $\hat{\tilde{\theta}}^*(x)$ are defined by replacing all the probabilities in the definition of $\tilde{\theta}_L(x)$ and $\tilde{\theta}^*(x)$ with their consistent estimators, respectively.

Semiparametric estimators like the ones in equation (A.12) converge at the usual $\sqrt{n}$ rate. Instead of listing all regularity conditions, which are well understood in the literature (see, e.g. Newey, 1994; Ai and Chen, 2003; Chen, Linton, and Van Keilegom, 2003; Ai and Chen, 2012; Ackerberg, Chen, Hahn, and Liao, 2014), we will derive the pathwise derivatives of θ_L and θ^*. The theorem stated below is the first main result of this section.

Theorem A.4. *Suppose that $e(0,x) < e(1,x)$ with $\inf_x e(0,x) > 0$ and $\sup_x e(1,x) < 1$. Further, suppose that $\inf_x \mathbb{P}\{Y_i(0) = 0 \mid e(0,x) < V_i \leq e(1,x), X_i = x\} > 0$. Then, the parameters θ_L and θ^* in equations (A.9) and (A.10) are well defined and they are pathwise differentiable in the sense of Newey (1994). Also, their (mean–zero) pathwise derivatives, denoted by $F_L(Y, Z, X)$ and $F^*(Y, T, Z, X)$, respectively, depend only on the objects that can be directly identified from data.*

Theorem A.4 does not display the specific forms of F_L and F^* simply because their expressions are too long and distracting; they are provided in lemmas A.10 and A.12, respectively. Below we discuss the relevance of theorem A.4.

The pathwise differentiability can tell us a couple of things about the semiparametric estimators $\hat{\theta}_L$ and $\hat{\theta}^*$. First, $F_L(Y_i, Z_i, X_i)$ and $F^*(Y_i, T_i, Z_i, X_i)$ will coincide with the influence functions of the semiparametric estimators $\hat{\theta}_L$ and $\hat{\theta}^*$ as long as they are asymptotically linear. Therefore, the asymptotic variance of $\hat{\theta}_L$ and $\hat{\theta}^*$ will be $\mathbb{V}\{F_L(Y_i, Z_i, X_i)\}$ and $\mathbb{V}\{F^*(Y_i, T_i, Z_i, X_i)\}$, respectively. Further, we show in the appendix that F_L and F^* are contained in the appropriate tangent space, which means that $\mathbb{V}\{F_L(Y_i, Z_i, X_i)\}$ and $\mathbb{V}\{F^*(Y_i, T_i, Z_i, X_i)\}$ are in fact the semiparametric efficiency bounds of θ_L and θ^*, respectively. We summarize these implications in the following theorem.

Theorem A.5. *Suppose that $\hat{\theta}_L$ and $\hat{\theta}^*$ are $\sqrt{n}$–consistent and asymptotically linear. Then, their asymptotic variances are given by $V_L = \mathbb{E}\{F_L^2(Y_i, Z_i, X_i)\}$ and $V^* = \mathbb{E}\{F^{*2}(Y_i, T_i, Z_i, X_i)\}$. Further, V_L and V^* are the semiparametric efficiency bounds for estimating θ_L and θ^*, respectively.*

By using the formulas of F_L and F^* provided in lemmas A.10 and A.12, we can consistently estimate V_L and V^*. Alternatively, one can simply rely on some resampling techniques such as the bootstrap. Once we obtain the estimates $\hat{V}_L$ and $\hat{V}^*$ of the asymptotic variances, we can conduct inference on θ_L and θ^*. In fact, inference on θ_L can be naturally extended to that of θ. Consider the simplest case where there is no uncertainty in the upper bound: i.e. we only have data on $(Y_i, Z_i, X_i^\intercal)^\intercal$. In this case, the sharp identified bounds of θ is simply $[\theta_L,\ 1]$, the asymptotically valid confidence interval for θ with the shortest length will be $\left[\hat{\theta}_L - z_\alpha \times \sqrt{\hat{V}_L/n},\ 1\right]$, where z_α is the one–sided critical value from the standard normal distribution (e.g. when $\alpha = 0.05$, we have $z_\alpha = 1.645$). When we observe all of $(Y_i, T_i, Z_i, X_i^\intercal)^\intercal$, θ^* is point identified and therefore the shortest confidence interval of θ^* can be obtained by using the usual two–sided critical value as in $\hat{\theta}^* \pm z_{\alpha/2} \times \sqrt{\hat{V}^*/n}$.

H.2. **Revisiting the NTV Example.** We revisit the NTV example in appendix D.2 to illustrate the results of appendix H.1. For simplicity, we focus on estimating the lower bound θ_L using a two–step parametric approach. For this exercise, we first create a binary instrument

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

by using the original continuous instrument. The conditional lower bound $\tilde{\theta}_L(x)$ is estimated using probit models that are linear in covariates used in appendix D.2 and is averaged out with respect to covariates by sample survey weight. The standard error is obtained by replacing population quantities of $F_L(Y_i, Z_i, X_i)$ in lemma A.10 with parametric estimates.

TABLE A.5. Persuasion Rates: NTV Effects Using a Binary Instrument

	(1) Not voting for Unity	(2) Voting for OVR
Point estimate of the lower bound	0.191	0.056
Standard error of the lower bound	0.083	0.024
One-sided 95% confidence interval for θ	[0.055,1]	[0.016,1]

The estimation results are summarized in table A.5. Both lower bounds are significantly different from zero; however, they are far smaller than the estimates of θ based on the original continuous instrument. This again illustrates the limitation of the identifying power of a binary instrument.

PART IV.

APPENDIX I. PROOFS

I.1. **Proofs of lemma 1 and lemma A.1.** If $q_i(0) \leq q_i(1)$, then $Y_i(0) = 1$ and $Y_i(1) = 0$ cannot happen. Now, conversely, suppose that $Y_i(0) \leq Y_i(1)$. If $\mathbb{P}\{q_i(1) < q_i(0)\} > 0$, then assumption A.1 implies that $\mathbb{P}\{q_i(1) < 1/(1 + U_i) < q_i(0)\} > 0$. This contradicts $\mathbb{P}\{Y_i(0) \leq Y_i(1)\} = 1$. The denominator on the right-hand side of equation (3) is nonzero since $\mathbb{P}\{Y_i(0) = 1\} < 1$ by assumption A.1. Finally, equation (3) follows from the fact that $Y_i(1) - Y_i(0) = \mathbb{1}\{Y_i(1) = 1, Y_i(0) = 0\}$ with probability one. □

I.2. **Proof of theorem 1.** By assumption B,

$$\mathbb{P}\{Y_i(z) = 1\} = \mathbb{P}(Y_i(z) = 1|Z_i = z) = \mathbb{P}(Y_i = 1|T_i = z) = \mathbb{P}(Y_i = 1|Z_i = z). \tag{A.13}$$

So, the assertion follows from lemma 1 and the definition of θ_L. □

I.3. **Under assumption C.**

Lemma A.1. $\mathbb{P}\{Y_i(1) = 1 \mid e(0) < V_i \leq e(1)\}$ *is identified by*

$$\frac{\mathbb{P}(Y_i = 1, T_i = 1|Z_i = 1) - \mathbb{P}(Y_i = 1, T_i = 1|Z_i = 0)}{e(1) - e(0)}.$$

Similarly, $\mathbb{P}\{Y_i(0) = 1 \mid e(0) < V_i \le e(1)\}$ *is identified by*

$$\frac{\mathbb{P}(Y_i = 1, T_i = 0|Z_i = 0) - \mathbb{P}(Y_i = 1, T_i = 0|Z_i = 1)}{e(1) - e(0)}.$$

Proof. The first assertion follows from

$$\mathbb{P}(Y_i = 1, T_i = 1|Z_i = z) = \mathbb{P}\{Y_i(1) = 1, V_i \le e(z)\}. \tag{A.14}$$

The second statement is similar. □

Lemma A.2. *For* $z = 0, 1$, $\mathbb{P}\{Y_i(1) = 1 \mid V_i \le e(z)\}$ *is identified by*

$$\mathbb{P}(Y_i = 1, T_i = 1|Z_i = z)/e(z).$$

Similarly, $\mathbb{P}\{Y_i(0) = 1 \mid V_i > e(z)\}$ *is identified by*

$$\mathbb{P}(Y_i = 1, T_i = 0|Z_i = z)/\{1 - e(z)\}.$$

Proof. The first assertion follows from

$$\mathbb{P}(Y_i = 1, T_i = 1|Z_i = z) = \mathbb{P}\{Y_i(1) = 1, V_i \le e(z)\}. \tag{A.15}$$

The second assertion is similar. □

Lemma A.3. *The sharp identified interval of* $\mathbb{P}\{Y_i(1) = 1\}$ *is given by*

$$\big[\mathbb{P}(Y_i = 1|Z_i = 1),\ \mathbb{P}(Y_i = 1, T_i = 1|Z_i = 1) + 1 - e(1)\big].$$

Similarly, the sharp identified interval of $\mathbb{P}\{Y_i(0) = 1\}$ *is given by*

$$\big[\mathbb{P}(Y_i = 1, T_i = 0|Z_i = 0),\ \mathbb{P}(Y_i = 1|Z_i = 0)\big].$$

Proof. For the first assertion, note that

$$\begin{aligned}\mathbb{P}\{Y_i(1) = 1\} &= \mathbb{P}\{Y_i(1) = 1 \mid e(0) < V_i \le e(1)\}\{e(1) - e(0)\} \\ &\quad + \mathbb{P}\{Y_i(1) = 1 \mid V_i \le e(0)\}e(0) + \mathbb{P}\{Y_i(1) = 1 \mid V_i > e(1)\}\{1 - e(1)\}.\end{aligned} \tag{A.16}$$

By lemmas A.1 and A.2, the first two terms on the right–hand side of equation (A.16) are identified and their sum is equal to $\mathbb{P}(Y_i = 1, T_i = 1|Z_i = 1)$. For the third term on the right–hand side of equation (A.16), note that

$$\mathbb{P}\{Y_i(1) = 1 \mid V_i > e(1)\}\{1 - e(1)\}$$
$$\geq \mathbb{P}\{Y_i(0) = 1 \mid V_i > e(1)\}\{1 - e(1)\} = \mathbb{P}(Y_i = 1, T_i = 0 | Z_i = 1), \quad \text{(A.17)}$$

where $\mathbb{P}\{Y_i(1) = 1 \mid V_i > e(1)\} \leq 1$. Therefore, the sharp bounds of the third term on the right–hand side of equation (A.16) is the interval between $\mathbb{P}(Y_i = 1, T_i = 0|Z_i = 1)$ and $1 - e(1)$. Combining all these proves the first assertion. The second assertion is similar. □

Proof of Theorem 2: Let $a = \mathbb{P}\{Y_i(1) = 1\}$ and $b = \mathbb{P}\{Y_i(0) = 1\}$: so, $\theta = (a - b)/(1 - b)$. Let m_a, M_a be the lower and upper bounds of a provided in lemma A.3. Similarly, let m_b, M_b be the bounds of b given in lemma A.3. By lemma A.3 and the fact that the dependence between $Y_i(0)$ and $Y_i(1)$ is unrestricted except that $Y_i(0) \leq Y_i(1)$, the identified bounds of θ can be obtained by

$$\max_{a,b} \text{ and } \min_{a,b} \frac{a-b}{1-b} \quad \text{subject to} \quad a \in [m_a,\ M_a],\ b \in [m_b,\ M_b],\ a \geq b. \quad \text{(A.18)}$$

Here, note that the restriction $a \geq b$ is redundant, because

$$m_a - M_b = \mathbb{P}(Y_i = 1|Z_i = 1) - \mathbb{P}(Y_i = 1|Z_i = 0)$$
$$= \mathbb{P}\{Y_i(1) = 1,\ e(0) < V_i \leq e(1)\} - \mathbb{P}\{Y_i(0) = 1,\ e(0) < V_i \leq e(1)\} \geq 0, \quad \text{(A.19)}$$

where the last inequality is from $Y_i(1) \geq Y_i(0)$. So, the minimum is $\theta_L = (m_a - M_b)/(1 - M_b) \geq 0$ and the maximum is $\theta_U = (M_a - m_b)/(1 - m_b)$: the monotonicity of the probability measure trivially shows that $\theta_U \leq 1$. Finally, sharpness follows from the intermediate value theorem because $(a - b)/(1 - b)$ varies continuously between θ_L and θ_U. □

I.4. Under assumption D.

Lemma A.4. *For any events A, B and for any probability measure $\mathbb{P}^*$, we have*

$$\max\{0, \mathbb{P}^*(A) - \mathbb{P}^*(B^c)\} \leq \mathbb{P}^*(A \cap B) \leq \min\{\mathbb{P}^*(A),\ \mathbb{P}^*(B)\},$$

where the bounds are sharp in that $\mathbb{P}^*(A \cap B)$ can be anything between the bounds without changing $\mathbb{P}^*(A)$ and $\mathbb{P}^*(B)$.

Proof. This is a version of the Fréchet–Hoeffding bounds. The upper bound is trivially true. For the lower bound, simply note that

$$\mathbb{P}^*(A) \leq \mathbb{P}^*(A \cap B) + \mathbb{P}^*(B^c) \quad \text{and} \quad \mathbb{P}^*(B) \leq \mathbb{P}^*(A \cap B) + \mathbb{P}^*(A^c),$$

where $\mathbb{P}^*(A) - \mathbb{P}^*(B^c) = \mathbb{P}^*(B) - \mathbb{P}^*(A^c)$. For sharpness, note that the upper bound is achieved when $A \subset B$ or $B \subset A$. Also, the lower bound is achieved when $A \cap B = \emptyset$ or $A \cup B = \Omega$, where Ω is the entire sample space. To show that anything between the bounds can be achieved, consider the canonical probability space without loss of generality: i.e. $\Omega = [0,1]$ and $\mathbb{P}^*$ be the Lebesgue measure on the Borel σ–algebra on Ω. Choose $p_A, p_B \in [0,1]$, where we assume that $p_A \geq p_B$ without loss of generality; the other case is symmetric. Let $A = [0,\ p_A]$ and $B = x + [p_A - p_B,\ p_A]$, where $0 \leq x \leq 1 - p_A$. So, A, B are events in Ω with $\mathbb{P}^*(A) = p_A, \mathbb{P}^*(B) = p_B$ for all $x \in [0, 1 - p_A]$. Now, note that $\mathbb{P}^*(A \cap B) = \max(p_B - x,\ 0)$ is a continuous function in $x \in [0, 1 - p_A]$, where its maximum and the minimum are given by p_B and $\max(p_B - 1 + p_A, 0)$. □

Lemma A.5. *The sharp identified interval of* $\mathbb{P}\{Y_i(1) = 1\}$ *is given by*

$$\big[\mathbb{P}(Y_i = 1|Z_i = 1),\ \min\{1, \mathbb{P}(Y_i = 1|Z_i = 1) + 1 - e(1)\}\big] \tag{A.20}$$

Similarly, the sharp identified interval of $\mathbb{P}\{Y_i(0) = 1\}$ *is given by*

$$\big[\max\{0,\ \mathbb{P}(Y_i = 1|Z_i = 0) - e(0)\},\ \mathbb{P}(Y_i = 1|Z_i = 0)\big]. \tag{A.21}$$

Proof. It follows from lemmas A.3 and A.4. □

Proof of theorem 3: Similarly to the proof of theorem 2, we need to consider

$$\max_{a,b} \text{ and } \min_{a,b} \frac{a-b}{1-b} \quad \text{subject to} \quad a \in [m_a, \tilde{M}_a],\ b \in [\tilde{m}_b, M_b],\ a \geq b, \tag{A.22}$$

where $m_a, \tilde{M}_a, \tilde{m}_b, M_b$ are given in lemma A.5. Follow the same reasoning as theorem 2. □

I.5. **Under assumption E.**

Proof of theorem 4: Since theorem 3 uses more information but its lower bound only depends on the distribution of (Y_i, Z_i), it suffices to focus on the upper bound. From theorem 3, we can find the sharp upper bound in this case by

$$\max_{0<e(0)\leq e(1)<1} \theta_{U_e} = \frac{\min\{1, \mathbb{P}(Y_i = 1|Z_i = 1) + 1 - e(1)\} - \max\{0,\ \mathbb{P}(Y_i = 1|Z_i = 0) - e(0)\}}{1 - \max\{0,\ \mathbb{P}(Y_i = 1|Z_i = 0) - e(0)\}}. \tag{A.23}$$

Note that setting $e(0) = \mathbb{P}(Y_i = 1|Z_i = 0) \leq \mathbb{P}(Y_i = 1|Z_i = 1) = e(1)$ yields the maximum value 1. Sharpness follows from the fact that θ_{U_e} is continuous in $\big(e(0), e(1)\big)$. □

I.6. **For the Compliers.**

Proof of theorem 6: For part *(i)*, note that

$$\mathbb{P}(Y_i = 1|Z_i = z) = \mathbb{P}\{Y_i(1) = 1, V_i \le e(z)\} + \mathbb{P}\{Y_i(0) = 1, V_i > e(z)\},$$

from which it follows that

$$\theta_{\text{local}} = \frac{\mathbb{P}(Y_i = 1|Z_i = 1) - \mathbb{P}(Y_i = 1|Z_i = 0)}{\{e(1) - e(0)\}\mathbb{P}\{Y_i(0) = 0|e(0) < V_i \le e(1)\}}. \tag{A.24}$$

Finally, note that the denominator on the right–hand side of equation (A.24) is equal to

$$\mathbb{P}\{Y_i(0) = 0, e(0) < V_i \le e(1)\} = \mathbb{P}\{Y_i(0) = 0, e(0) < V_i\} - \mathbb{P}\{Y_i(0) = 0, e(1) < V_i\},$$

where $\mathbb{P}\{Y_i(0) = 0, e(z) < V_i\} = \mathbb{P}\{Y_i = 0, T_i = 0|Z_i = z\}$.

For part *(ii)*, we look for sharp bounds for $\mathbb{P}\{Y_i(0) = 0, e(0) < V_i \le e(1)\}$ under assumption D. Using the fact that the sharp bounds of $\mathbb{P}(A \cap B \cap C)$ when $\mathbb{P}(A \cap B), \mathbb{P}(B \cap C)$, and $\mathbb{P}(C \cap A)$ are given are equal to the interval between 0 and $\min\{\mathbb{P}(A \cap B), \mathbb{P}(B \cap C), \mathbb{P}(C \cap A)\}$, we know that

$$\begin{aligned} 0 &\le \mathbb{P}\{Y_i(0) = 0, e(0) < V_i \le e(1)\} \\ &\le \min\big[\mathbb{P}\{Y_i(0) = 0, V_i > e(0)\},\ e(1) - e(0),\ \mathbb{P}\{Y_i(0) = 0, V_i \le e(1)\}\big], \end{aligned} \tag{A.25}$$

where it suffices to look for the sharp upper bound of the expression on the utmost left–hand side. First,

$$\mathbb{P}\{Y_i(0) = 0, V_i > e(0)\} = \mathbb{P}(Y_i = 0, T_i = 0|Z_i = 0) \le \mathbb{P}(Y_i = 0|Z_i = 0),$$

where the inequality holds with equality when $\mathbb{P}(Y_i = 0, T_i = 1|Z_i = 0) = 0$. Second, note that

$$\mathbb{P}\{Y_i(0) = 0, V_i \le e(1)\} = \mathbb{P}\{Y_i(0) = 0, T_i = 1|Z_i = 1\}$$

is totally unidentified. So, we conclude that the sharp upper bound of the term on the right–hand side of equation (A.25) is

$$\min\{\mathbb{P}(Y_i = 0|Z_i = 0), e(1) - e(0)\}. \tag{A.26}$$

The bound in part *(iii)* corresponds to the case where $e(1) - e(0) = 1$. □

Proof of theorem 7: By the same reasoning as lemma 1, we have

$$\theta_{\text{mte}}(v) = \frac{\mathbb{E}\{Y_i(1) - Y_i(0)\ |V_i = v\}}{\mathbb{P}(Y_i(0) = 0\ |\ V_i = v\}}. \tag{A.27}$$

Then as shown in Heckman and Vytlacil (2005),

$$\mathbb{E}\{Y_i(1) - Y_i(0) \,|V_i = v\} = \left.\frac{\partial \mathbb{P}\{Y_i = 1 | e(Z_i) = e\}}{\partial e}\right|_{e=v}.$$

Also, by the same argument as in Heckman and Vytlacil (2005),

$$\mathbb{E}\{Y_i(0) \,|V_i = v\} = -\left.\frac{\partial \mathbb{P}\{Y_i = 1, T_i = 0 | e(Z_i) = e\}}{\partial e}\right|_{e=v}.$$

The desired result follows immediately. □

I.7. **Proofs with Nonbinary Outcomes.**

Proof of Lemma 2: Fixing $Y_{i0}(t) = 0$ for $t = 0, 1$, there are only four possibilities, where one of them can be ruled out by $Y_{i1}(1) \geq Y_{i1}(0)$: this is illustrated in the following table, where the outcomes in the third row (those with *) have probability zero.

$Y(1)$	$Y(0)$
$(0,1,0)$	$(0,1,0)$
$(0,1,0)$	$(0,0,1)$
$(0,0,1)^*$	$(0,1,0)^*$
$(0,0,1)$	$(0,0,1)$

Therefore,

$$\mathbb{P}\{Y_i(1) = (0,1,0), Y_i(0) = (0,0,1)\} = \mathbb{P}\{Y_i(1) = (0,1,0)\} - \mathbb{P}\{Y_i(0) = (0,1,0)\},$$

which takes care of the numerator of the conditional probability of θ_{mult}. For the denominator, note that

$$\mathbb{P}\{Y_i(0) = (0,0,1)\} = \mathbb{P}\{Y_{i,-1}(0) = 1\} = 1 - \mathbb{P}\{Y_{i0}(0) = 1\} - \mathbb{P}\{Y_{i1}(0) = 1\}.$$ □

Lemma A.6. *For $j \in \{0, 1, -1\}$, we have*

$$\mathbb{P}\{Y_{ij}(1) = 1 \mid e(0) < V_i \leq e(1)\} = \frac{p_j(1,1|1) - p_j(1,1|0)}{e(1) - e(0)}.$$

Similarly,

$$\mathbb{P}\{Y_{ij}(0) = 1 \mid e(0) < V_i \leq e(1)\} = \frac{p_j(1,0|0) - p_j(1,0|1)}{e(1) - e(0)}.$$

Proof. It follows from the same proof as lemma A.1. □

Lemma A.7. *For $z = 0, 1$ and $j \in \{0, 1, -1\}$, we have*

$$\mathbb{P}\{Y_{ij}(1) = 1 \mid V_i \le e(z)\} = \frac{p_j(1, 1|z)}{e(z)}.$$

Similarly,

$$\mathbb{P}\{Y_{ij}(0) = 1 \mid V_i > e(z)\} = \frac{p_j(1, 0|z)}{1 - e(z)}.$$

Proof. It follows from the same proof as lemma A.2. □

Lemma A.8. *The sharp identified region of $\big(\mathbb{P}\{Y_{i1}(1) = 1\}, \mathbb{P}\{Y_{i0}(0) = 1\}, \mathbb{P}\{Y_{i1}(0) = 1\}\big)$ is given by $\mathcal{S}_1(1) \times \mathcal{S}_{01}(0)$, where $\mathcal{S}_1(1) = \big[p_1(1|1),\ p_1(1, 1|1) + 1 - e(1)\big]$ is the sharp identified interval of $\mathbb{P}\{Y_{i1}(1) = 1\}$ and*

$$\mathcal{S}_{01}(0) = \Big\{(x, y) \in [0, 1]^2 :\ p_0(1, 0|0) \le x \le p_0(1, 0|0) + e(0),$$
$$p_1(1, 0|0) \le y \le p_1(1|0), \quad x + y \le \sum_{j=0}^{1} p_j(1, 0|0) + e(0)\Big\} \ne \emptyset$$

is the sharp identified region of $\big(\mathbb{P}\{Y_{i0}(0) = 1\}, \mathbb{P}\{Y_{i1}(0) = 1\}\big)$.

Proof. Note that

$$\mathbb{P}\{Y_{i1}(1) = 1\} = \mathbb{P}\{Y_{i1}(1) = 1 \mid V_i \le e(0)\}e(0)$$
$$+ \mathbb{P}\{Y_{i1}(1) = 1 \mid e(0) < V_i \le e(1)\}\{e(1) - e(0)\} + \mathbb{P}\{Y_{i1}(1) = 1 \mid V_i > e(1)\}\{1 - e(1)\}, \quad \text{(A.28)}$$

where the first two terms on the right-hand side are identified by lemmas A.6 and A.7. Specifically,

$$\mathbb{P}\{Y_{i1}(1) = 1\} = p_1(1, 1|1) + \mathbb{P}\{Y_{i1}(1) = 1 \mid V_i > e(1)\}\{1 - e(1)\}. \quad \text{(A.29)}$$

Similarly, for $j \in \{0, 1, -1\}$, we obtain

$$\mathbb{P}\{Y_{ij}(0) = 1\} = p_j(1, 0|0) + \mathbb{P}\{Y_{ij}(0) = 1 \mid V_i \le e(0)\}e(0). \quad \text{(A.30)}$$

Here, note that there is no restriction on the relationship between $\tilde{a} := \mathbb{P}\{Y_{i1}(1) = 1 \mid V_i > e(1)\}$ and $(\tilde{b}, \tilde{c}) := \big(\mathbb{P}\{Y_{i0}(0) = 1 \mid V_i \le e(0)\},\ \mathbb{P}\{Y_{i1}(0) = 1 \mid V_i \le e(0)\}\big)$. Therefore, it suffices to consider $\tilde{a}$ and $(\tilde{b}, \tilde{c})$ separately, after which we take the Cartesian product of the two sharp identified sets. For the sharp interval of $\tilde{a}$, we simply combine equation (A.29) with

$$1 \ge \mathbb{P}\{Y_{i1}(1) = 1 \mid V_i > e(1)\} \ge \mathbb{P}\{Y_{i1}(0) = 1 \mid V_i > e(1)\} = \mathbb{P}(Y_{i1} = 1 | T_i = 0, Z_i = 1), \quad \text{(A.31)}$$

which yields $\mathbb{P}\{Y_{i1}(1)=1\} \in \mathcal{S}_1(1)$. For the sharp region of $(\tilde{b}, \tilde{c})$, we use equation (A.30) with the fact that

$$0 \leq \mathbb{P}\{Y_{i1}(0)=1 \mid V_i \leq e(0)\} \leq \mathbb{P}\{Y_{i1}(1)=1 \mid V_i \leq e(0)\} = \mathbb{P}(Y_{i1}=1|T_i=1, Z_i=0). \quad \text{(A.32)}$$

Therefore, it follows that

$$\mathbb{P}\{Y_{i0}(0)=1\} \in \left[p_0(1,0|0),\ p_0(1,0|0)+e(0)\right], \quad \text{(A.33)}$$

$$\mathbb{P}\{Y_{i1}(0)=1\} \in \left[p_1(1,0|0),\ p_1(1|0)\right], \quad \text{(A.34)}$$

$$\mathbb{P}\{Y_{i,-1}(0)=1\} \in \left[p_{-1}(1,0|0),\ p_{-1}(1,0|0)+e(0)\right], \quad \text{(A.35)}$$

where we must have

$$\sum_{j\in\{0,1,-1\}} \mathbb{P}\{Y_{ij}(0)=1\} = 1 \quad \text{and} \quad \sum_{j\in\{0,1,-1\}} p_j(1,0|0) = 1-e(0). \quad \text{(A.36)}$$

by assumption F.[23] Therefore, we can rewrite equation (A.35) by using equation (A.36). Specifically, equation (A.35) can be written as

$$\begin{aligned}\mathbb{P}\{Y_{i,-1}(0)=1\} &= 1-\mathbb{P}\{Y_{i0}(0)=1\}-\mathbb{P}\{Y_{i1}(0)=1\} \\ &\in \left[1-e(0)-p_0(1,0|0)-p_1(1,0|0),\ 1-e(0)-p_0(1,0|0)-p_1(1,0|0)+e(0)]\right], \quad \text{(A.37)}\end{aligned}$$

which implies

$$\mathbb{P}\{Y_{i0}(0)=1\}+\mathbb{P}\{Y_{i1}(0)=1\} \in \left[\sum_{j=0}^{1} p_j(1,0|0),\ \sum_{j=0}^{1} p_j(1,0|0)+e(0)\right]. \quad \text{(A.38)}$$

But the lower bound in equation (A.38) is implied by equations (A.33) and (A.34), whereas the upper bound is not redundant. Note that $\mathcal{S}_{01}(0)$ is not empty, unless

$$-p_0(1,0|0)+\sum_{j=0}^{1} p_j(1,0|0)+e(0) < p_1(1,0|0),$$

which is not possible. □

[23]The upper end point of the bounds in equation (A.34) is no larger than the lower end point of the interval $\mathcal{S}_1(1)$ because

$$p_1(1|1)-p_1(1|0) = \mathbb{P}\{Y_{i1}(1)=1, e(0)<V_i\leq e(1)\}-\mathbb{P}\{Y_{i1}(0)=1, e(0)<V_i\leq e(1)\} \geq 0,$$

where the inequality is by $Y_{i1}(1) \geq Y_{i1}(0)$.

Proof of Theorem A.1: Let $a := \mathbb{P}\{Y_{i1}(1) = 1\}, b := \mathbb{P}\{Y_{i0}(0) = 1\}$, and $c := \mathbb{P}\{Y_{i1}(0) = 1\}$, and we have

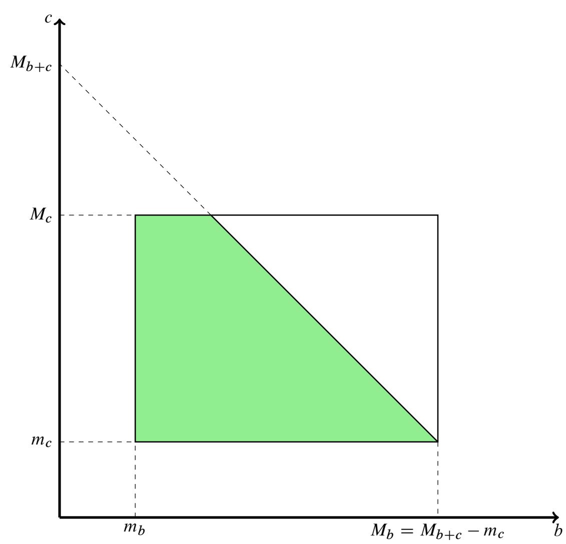
$$\theta_{\text{mult}} = \frac{a - c}{1 - b - c},$$

where the sharp identified region of (a, b, c) is $\mathcal{S}_1(1) \times \mathcal{S}_{01}(0)$ by lemma A.8. Therefore, we can obtain the sharp identified bounds of θ_{mult} by solving constrained maximization/minimization problems:

$$\max / \min_{a,b,c} \quad \frac{a - c}{1 - b - c} \qquad \text{s.t.} \quad \begin{cases} m_a \le a \le M_a, \\ m_b \le b \le M_b, \\ m_c \le c \le M_c, \\ b + c \le M_{b+c}, \end{cases} \tag{A.39}$$

where $m_a, M_a, \cdots, M_{b+c}$ are given in the definitions of $\mathcal{S}_1(1)$ and $\mathcal{S}_{10}(0)$. Here, the restriction of $a \ge c$ is automatic, because $m_a = p_1(1|1) \ge p_1(1|0) = M_c$ as we explained in footnote 23. Further, note that $m_b + m_c < M_{b+c}$ because $e(0) > 0$; therefore, the set of (b, c) that satisfies the constraints is not empty. In the following arguments, the set of the feasible values of (b, c) is important, which is illustrated in figure A.2.

FIGURE A.2. The set of the feasible values of (b, c)



Note: $M_{b+c} - m_b = p_0(1,0|0) + p_1(1,0|0) + e(0) - p_0(1,0|0) \ge p_1(1,0|0) + p_1(1,1|0) = M_c$.
Note: $M_{b+c} - m_c = p_0(1,0|0) + e(0) = M_b$.
Note: $M_{b+c} - M_b = p_0(1,0|0) + p_1(1,0|0) + e(0) - p_0(1,0|0) - e(0) \le p_1(1,0|0) + p_1(1,1|0) = M_c$.

Consider the minimization first, for which it suffices to solve

$$\min_{b,c} \quad \frac{m_a - c}{1 - b - c} \qquad \text{s.t.} \qquad \begin{cases} m_b \leq b \leq M_b, \\ m_c \leq c \leq M_c, \\ \quad b + c \leq M_{b+c}. \end{cases} \tag{A.40}$$

Here, the objective function is monotonically increasing in b for every c. Therefore,

$$\min\Big(\frac{m_a - c}{1 - m_b - c},\ 1\Big) \leq \min\Big(\frac{m_a - c}{1 - b - c},\ 1\Big) \leq \theta_{\text{mult}} \leq 1 \tag{A.41}$$

for all $m_c \leq c \leq \min(M_{b+c} - m_b,\ M_c) = M_c$, where the utmost left–hand side inequality in equation (A.41) holds with equality when $b = m_b$.

Now, we consider two possibilities: i.e. $m_a + m_b \geq 1$ and $m_a + m_b < 1$. If $m_a + m_b \geq 1$, then the utmost left–hand side of equation (A.41) becomes 1 for any $c \leq 1 - m_b$. But, taking $c = m_c$ leads to $m_c \leq M_{b+c} - m_b \leq 1 - m_b$ by

$$M_{b+c} = p_0(1,0|0) + p_1(1,0|0) + e(0) = 1 - e(0) - p_{-1}(1,0|0) + e(0) \leq 1. \tag{A.42}$$

Therefore, we conclude $\theta_{\text{mult}} = 1$ in this case, which is achieved when $a = m_a, b = m_b$, and $c = m_c$.

If $m_a + m_b < 1$, then $(m_a - c)/(1 - m_b - c)$ is decreasing in c, and therefore, we simply take the largest value of c to achieve the minimum: i.e.

$$\frac{m_a - M_c}{1 - m_b - M_c} \leq \theta_{\text{mult}}. \tag{A.43}$$

This is indeed the sharp lower bound, because $(a - c)/(1 - b - c)$ is a continuous function in (a, b, c) and the lower bound is achieved when $a = m_a, b = m_b$, and $c = M_c$.

The sharp upper bound can be similarly obtained by solving the maximization problem in equation (A.39), for which it suffices to consider

$$\max_{b,c} \quad \frac{M_a - c}{1 - b - c} \qquad \text{s.t.} \qquad \begin{cases} m_b \leq b \leq M_b, \\ m_c \leq c \leq M_c, \\ \quad b + c \leq M_{b+c}. \end{cases} \tag{A.44}$$

Since the objective function is increasing in b, it suffices to have b take its largest value within the feasible set, i.e. $b = \min(M_{b+c} - m_c,\ M_b) = M_b$.

Then, we have

$$\theta_{\text{mult}} \leq \frac{M_a - c}{1 - b - c} \leq \frac{M_a - c}{1 - M_b - c}, \tag{A.45}$$

for all $m_c \leq c \leq \min(M_{b+c} - m_b,\ M_c) = M_c$, where the utmost right–hand side inequality in equation (A.45) holds with equality when $b = M_b$.

Now, we consider two possibilities: i.e. $M_a + M_b \geq 1$ and $M_a + M_b < 1$. First, suppose that $M_a + M_b \geq 1$. Then, the utmost right–hand side of equation (A.45) is no smaller than 1 for any $c \leq 1 - M_b$. Therefore, the sharp upper bound of θ_{mult} is trivial and equal to 1 in this case. Suppose that $M_a + M_b < 1$. Then, the utmost right–hand side expression in equation (A.45) is decreasing in c, and therefore

$$\theta_{\text{mult}} \leq \frac{M_a - c}{1 - b - c} \leq \frac{M_a - m_c}{1 - M_b - m_c}. \tag{A.46}$$

Sharpness follows from the fact that $(a - c)/(1 - b - c)$ is continuous in (a, b, c) and the upper bound is achieved when $a = M_a, b = M_b$, and $c = m_c$.

Finally, since $m_a + m_b \leq M_a + M_b$, what we have shown so far yields the sharp upper and lower bounds in all of the three possible cases, i.e. (i) $1 \leq m_a + m_b$, (ii) $m_a + m_b < 1 \leq M_a + M_b$, and (iii) $m_a + m_b \leq M_a + M_b < 1$. □

Proof of Theorem A.2: We use lemmas A.4 and A.8 to obtain the sharp identified region of $\big(\mathbb{P}\{Y_{i1}(1) = 1\}, \mathbb{P}\{Y_{i0}(0) = 1\}, \mathbb{P}\{Y_{i1}(0) = 1\}\big)$, after which we follow the same reasoning as in the proof of theorem A.1. □

Proof of Theorem A.3: The case of $p_1(1|1) = 1$ is trivial, because we can simply take $e(0) = p_0(1|0)$; this leads us to case (i) in theorem A.2 but the claimed bounds yields the same conclusion. Suppose that $p_1(1|1) < 1$. Then, we can always choose $e(1) = \max\{p_0(1|0),\ p_1(1|0)\}$ and take $e(0)$ smaller than but arbitrarily close to $e(1)$, which leads us to case (ii) in theorem A.2 with the claimed bounds. □

I.8. Proofs for the Results for Efficient Estimation of θ_L.

In this part of the appendix we consider the efficiency issues for the θ_L parameter defined in equation (A.9). So, we assume that the data available to us are the i.i.d. observations of $(Y_i, Z_i, X_i^\intercal)^\intercal$. Below we write f for F', i.e. the density of X_i. Further, we use the following notation:

$$\mathcal{P}_{y_1|z_1}(x) = \mathbb{P}(Y_i = 1|Z_i = 1, X_i = x) \quad \text{and} \quad \mathcal{P}_{y_0|z_1}(x) = 1 - \mathcal{P}_{y_1|z_1}(x),$$
$$\mathcal{P}_{y_1|z_0}(x) = \mathbb{P}(Y_i = 1|Z_i = 0, X_i = x) \quad \text{and} \quad \mathcal{P}_{y_0|z_0}(x) = 1 - \mathcal{P}_{y_1|z_0}(x),$$
$$\mathcal{P}_{z_1}(x) = \mathbb{P}(Z_i = z|X_i = x) \quad \text{and} \quad \mathcal{P}_{z_0}(x) = 1 - \mathcal{P}_{z_1}(x).$$

Then, we can write the likelihood function as follows.

$$q_L(y,z,x) = f(x) \prod_{j\in\{1,0\}} \Big[\mathcal{P}_{z_j}(x)\big\{\mathcal{P}_{y_1|z_j}(x)^y \mathcal{P}_{y_0|z_j}(x)^{1-y}\big\}\Big]^{\tilde{z}_j},$$

where $\tilde{z}_1 = z$ and $\tilde{z}_0 = 1 - z$. Therefore, the loglikelihood function is given by

$$\log q_L(y,z,x) = \log f(x) + \sum_{j\in\{1,0\}} \big\{\tilde{z}_j \log \mathcal{P}_{z_j}(x) + \tilde{z}_j y \log \mathcal{P}_{y_1|z_j}(x) + \tilde{z}_j (1-y)\mathcal{P}_{y_0|z_j}(x)\big\}. \quad \text{(A.47)}$$

Lemma A.9. *The tangent space for θ_L is given by*

$$\mathcal{T}_L = \Big\{\alpha_f(x) + \{z - \mathcal{P}_{z_1}(x)\}\alpha_z(x) + \sum_{j\in\{0,1\}} \tilde{z}_j\{y - \mathcal{P}_{y_1|z_j}(x)\}\alpha_{y|z_j}(x)\Big\},$$

where α_f is any square–integrable function with $\mathbb{E}\{\alpha_f(X_i)\} = 0$ and $\alpha_z, \alpha_{y|z_j}$ are any square–integrable functions of x.

Proof. Let $\mathcal{P}_{z_j}(x|\gamma), \mathcal{P}_{y_k|z_j}(x|\gamma)$ denote regular parametric submodels indexed by γ:[24] we will denote the true value by γ_0. Then, it follows from equation (A.47) that the score is given by

$$s(y,z,x|\gamma_0) = s_X(x|\gamma_0) + s_{Z|X}(z,x|\gamma_0) + s_{Y|Z,X}(y,z,x|\gamma_0), \quad \text{(A.48)}$$

where

$$s_X(x|\gamma_0) = \frac{1}{f(x)}\frac{\partial f(x|\gamma)}{\partial\gamma},$$

$$s_{Z|X}(z,x|\gamma_0) = \left\{\frac{z}{\mathcal{P}_{z_1}(x)} - \frac{1-z}{\mathcal{P}_{z_0}(x)}\right\}\frac{\partial \mathcal{P}_{z_1}(x|\gamma)}{\partial\gamma},$$

$$s_{Y|Z,X}(y,z,x|\gamma_0) = \sum_{j\in\{1,0\}} \tilde{z}_j \left\{\frac{y}{\mathcal{P}_{y_1|z_j}(x)} - \frac{1-y}{\mathcal{P}_{y_0|z_j}(x)}\right\}\frac{\partial \mathcal{P}_{y_1|z_j}(x|\gamma)}{\partial\gamma},$$

where all the derivatives are evaluated at γ_0. The conclusion follows from the fact that all the derivatives are unrestricted here. □

Lemma A.10. *The pathwise derivative of θ_L is given by*

$$F_L(Y,Z,X) = \frac{Z}{\mathcal{P}_{z_1}(X)}\frac{Y - \mathcal{P}_{y_1|z_1}(X)}{1 - \mathcal{P}_{y_1|z_0}(X)} + \frac{1-Z}{\mathcal{P}_{z_0}(X)}\frac{\{Y - \mathcal{P}_{y_1|z_0}(X)\}\{\tilde{\theta}_L(X) - 1\}}{1 - \mathcal{P}_{y_1|z_0}(X)} + \tilde{\theta}_L(X) - \theta_L.$$

[24] There is no loss of generality in assuming that γ is a scalar.

Proof. Let $\bar{\theta}_L(\gamma)$ be the parameter corresponding to θ_L along regular parametric submodels indexed by γ: i.e.

$$\bar{\theta}_L(\gamma) = \int \frac{\mathcal{P}_{y_1|z_1}(x|\gamma) - \mathcal{P}_{y_1|z_0}(x|\gamma)}{1 - \mathcal{P}_{y_1|z_0}(x|\gamma)} f(x|\gamma)dx.$$

Then,

$$\frac{\partial \tilde{\theta}_L(\gamma_0)}{\partial \gamma} = \int \frac{1}{1 - \mathcal{P}_{y_1|z_0}(x)} \frac{\partial \mathcal{P}_{y_1|z_1}(x|\gamma)}{\partial \gamma} f(x)dx + \int \frac{\theta_L(x) - 1}{1 - \mathcal{P}_{y_1|z_0}(x)} \frac{\partial \mathcal{P}_{y_1|z_0}(x|\gamma)}{\partial \gamma} f(x)dx + \int \theta_L(x) \frac{\partial f(x|\gamma)}{\partial \gamma} dx. \quad \text{(A.49)}$$

Now, note that the score $s(Y, Z, X|\gamma_0)$ given in equation (A.48) is the sum of Bernoulli scores, and therefore it follows that

$$\frac{\partial \bar{\theta}_L(\gamma_0)}{\partial \gamma} = \mathbb{E}\{F_L(Y, Z, X)s(Y, Z, X|\gamma_0)\}.$$

□

I.9. **Proofs for the Results for Efficient Estimation of θ^*.**

We now derive the efficient influence function of the integrated local persuasion parameter defined in equation (A.10) when an i.i.d. sample of (Y_i, T_i, Z_i, X_i) is available. Similarly to the previous subsection we use the following notation:

$$\mathcal{P}_{z_1}(x) = \mathbb{P}(Z_i = 1|X_i = x) \quad \text{and} \quad \mathcal{P}_{z_0} = 1 - \mathcal{P}_{z_1}(x),$$

$$\mathcal{P}_{t_1|z_1}(x) = \mathbb{P}(T_i = 1|Z_i = 1, X_i = x) \quad \text{and} \quad \mathcal{P}_{t_0|z_1}(x) = 1 - \mathcal{P}_{t_1|z_1}(x),$$
$$\mathcal{P}_{t_1|z_0}(x) = \mathbb{P}(T_i = 1|Z_i = 0, X_i = x) \quad \text{and} \quad \mathcal{P}_{t_0|z_0}(x) = 1 - \mathcal{P}_{t_1|z_0}(x),$$

$$\mathcal{P}_{y_1|t_1,z_1}(x) = \mathbb{P}(Y_i = 1|T_i = 1, Z_i = 1, X_i = x) \quad \text{and} \quad \mathcal{P}_{y_0|t_1,z_1}(x) = 1 - \mathcal{P}_{y_1|t_1,z_1}(x),$$
$$\mathcal{P}_{y_1|t_0,z_1}(x) = \mathbb{P}(Y_i = 1|T_i = 0, Z_i = 1, X_i = x) \quad \text{and} \quad \mathcal{P}_{y_0|t_0,z_1}(x) = 1 - \mathcal{P}_{y_1|t_0,z_1}(x),$$
$$\mathcal{P}_{y_1|t_1,z_0}(x) = \mathbb{P}(Y_i = 1|T_i = 1, Z_i = 0, X_i = x) \quad \text{and} \quad \mathcal{P}_{y_0|t_1,z_0}(x) = 1 - \mathcal{P}_{y_1|t_1,z_0}(x),$$
$$\mathcal{P}_{y_1|t_0,z_0}(x) = \mathbb{P}(Y_i = 1|T_i = 0, Z_i = 0, X_i = x) \quad \text{and} \quad \mathcal{P}_{y_0|t_0,z_0}(x) = 1 - \mathcal{P}_{y_1|t_0,z_0}(x).$$

Using this notation, the likelihood function can be written as

$$q^*(y, t, z, x) = f(x) \prod_{j=\{1,0\}} \Big[\mathcal{P}_{z_j}(x) \prod_{k=\{1,0\}} \{\mathcal{P}_{t_k|z_j}(x) \prod_{\ell=\{1,0\}} \mathcal{P}_{y_\ell|t_k,z_j}(x)^{\tilde{y}_\ell}\}^{\tilde{t}_k}\Big]^{\tilde{z}_j}$$
$$= f(x) \prod_{j=\{1,0\}} \Big[\mathcal{P}_{z_j}(x)^{\tilde{z}_j} \prod_{k=\{1,0\}} \{\mathcal{P}_{t_k|z_j}(x)^{\tilde{t}_k\tilde{z}_j} \prod_{\ell=\{1,0\}} \mathcal{P}_{y_\ell|t_k,z_j}(x)^{\tilde{y}_\ell\tilde{t}_k\tilde{z}_j}\}\Big], \quad \text{(A.50)}$$

where $\tilde{z}_1 = z, \tilde{z}_0 = 1 - z$, $\tilde{t}_1 = t, \tilde{t}_0 = 1 - t$, and $\tilde{y}_1 = y, \tilde{y}_0 = 1 - y$. Therefore, the loglikelihood function is given by

$$\log q^*(y,t,z,x) = \log f(x) + \sum_{j\in\{1,0\}} \tilde{z}_j \log \mathcal{P}_{z_j}(x) + \sum_{j\in\{1,0\}} \sum_{k\in\{0,1\}} \tilde{t}_k \tilde{z}_j \log \mathcal{P}_{t_k|z_j}(x) + \sum_{\ell\in\{1,0\}} \sum_{j\in\{1,0\}} \sum_{k\in\{0,1\}} \tilde{y}_\ell \tilde{t}_k \tilde{z}_j \log \mathcal{P}_{y_\ell|t_k,z_j}(x). \tag{A.51}$$

Lemma A.11. *The tangent space for θ^* is given by*

$$\mathcal{T}^* = \Big\{\alpha_f(x) + \{z - \mathcal{P}_{z_1}(x)\}\alpha_z(x) + \sum_{j\in\{0,1\}} \tilde{z}_j\{t - \mathcal{P}_{t_1|z_j}(x)\}\alpha_{t|z_j}(x) + \sum_{j\in\{0,1\}} \sum_{k\in\{0,1\}} \tilde{t}_k \tilde{z}_j \{y - \mathcal{P}_{y_1|t_k,z_j}(x)\}\alpha_{y|t_k z_j}(x)\Big\},$$

where α_f is any square–integrable function with $\mathbb{E}\{\alpha_f(X_1)\} = 0$, and $\alpha_z, \alpha_{t|z_j}, \alpha_{y|t_k,z_j}$ are any square–integrable functions of x.

Proof. Let $\mathcal{P}_{z_j}(x|\gamma), \mathcal{P}_{t_k|z_j}(x|\gamma), \mathcal{P}_{y_\ell|t_k,z_j}(x|\gamma)$ denote regular parametric submodels indexed by γ: as in the proof of lemma A.9, γ is a scalar–valued parameter and its true value is denoted by γ_0. From the loglikelihood function described in equation (A.51), we know that the score of the regular parametric submodel can be written as follows:

$$s^*(y,t,z,x|\gamma_0) = \frac{1}{f(x)}\frac{\partial f(x|\gamma)}{\partial \gamma} + \left\{\frac{\tilde{z}_1}{\mathcal{P}_{z_1}(x)} - \frac{\tilde{z}_0}{\mathcal{P}_{z_0}(x)}\right\}\frac{\partial \mathcal{P}_{z_1}(x|\gamma)}{\partial \gamma} + \sum_{j\in\{1,0\}} \tilde{z}_j \left\{\frac{\tilde{t}_1}{\mathcal{P}_{t_1|z_j}(x)} - \frac{\tilde{t}_0}{\mathcal{P}_{t_0|z_j}(x)}\right\}\frac{\partial \mathcal{P}_{t_1|z_j}(x|\gamma)}{\partial \gamma} + \sum_{j\in\{1,0\}} \sum_{k\in\{1,0\}} \tilde{t}_k \tilde{z}_j \left\{\frac{\tilde{y}_1}{\mathcal{P}_{y_1|t_k,z_j}(x)} - \frac{\tilde{y}_0}{\mathcal{P}_{y_0|t_k,z_j}(x)}\right\}\frac{\partial \mathcal{P}_{y_1|t_k,z_j}(x|\gamma)}{\partial \gamma}, \tag{A.52}$$

where all the derivatives are evaluated at γ_0. Here, we can do further albegra by using

$$\frac{\partial \mathcal{P}_{z_1}(x|\gamma)}{\partial \gamma} = -\frac{\partial \mathcal{P}_{z_0}(x|\gamma)}{\partial \gamma}, \tag{A.53}$$

$$\frac{\partial \mathcal{P}_{t_1|z_j}(x|\gamma)}{\partial \gamma} = -\frac{\partial \mathcal{P}_{t_0|z_j}(x|\gamma)}{\partial \gamma} \quad \text{for } j = 0,1 \tag{A.54}$$

$$\frac{\partial \mathcal{P}_{y_1|t_k,z_j}(x|\gamma)}{\partial \gamma} = -\frac{\partial \mathcal{P}_{y_0|t_k,z_j}(x|\gamma)}{\partial \gamma} \quad \text{for } j,k = 0,1. \tag{A.55}$$

Therefore, the score can be rewritten as follows:

$$s^*(y,t,z,x|\gamma_0) = s_X(x|\gamma_0) + s_{Z|X}(z,x|\gamma_0) + s_{T|Z,X}(t,z,x|\gamma_0) + s_{Y|T,Z,X}(y,t,z,x|\gamma_0), \tag{A.56}$$

where

$$s_X(x|\gamma_0) = \frac{1}{f(x)}\frac{\partial f(x|\gamma)}{\partial \gamma}, \tag{A.57}$$

$$s_{Z|X}(z,x|\gamma_0) = \frac{z-\mathcal{P}_{z_1}(x)}{\mathcal{P}_{z_1}(x)\{1-\mathcal{P}_{z_1}(x)\}}\frac{\partial \mathcal{P}_{z_1}(x|\gamma)}{\partial \gamma}, \tag{A.58}$$

$$s_{T|Z,X}(t,z,x|\gamma_0) = \sum_{j\in\{1,0\}} \frac{\tilde{z}_j\{t-\mathcal{P}_{t_1|z_j}(x)\}}{\mathcal{P}_{t_1|z_j}(x)\{1-\mathcal{P}_{t_1|z_j}(x)\}}\frac{\partial \mathcal{P}_{t_1|z_j}(x|\gamma)}{\partial \gamma}, \tag{A.59}$$

$$s_{Y|T,Z,X}(y,t,z,x|\gamma_0) = \sum_{j\in\{1,0\}}\sum_{k\in\{1,0\}} \frac{\tilde{t}_k\tilde{z}_j\{y-\mathcal{P}_{y_1|t_k,z_j}(x)\}}{\mathcal{P}_{y_1|t_k,z_j}(x)\{1-\mathcal{P}_{y_1|t_k,z_j}(x)\}}\frac{\partial \mathcal{P}_{y_1|t_k,z_j}(x|\gamma)}{\partial \gamma}. \tag{A.60}$$

Interpretation of each term should be straightforward. For example, $s_{T|Z,X}(y,z,x|\gamma_0)$ is the score of T at t conditional on $Z=z, X=x$. Finally, the conclusion follows from equation (A.56) and the fact that all the derivatives here are unrestricted. □

Now, we derive the pathwise derivative of θ^*. For this purpose it is convenient to write

$$\theta^* = \int Q_1(x)Q_2(x)f(x)dx, \tag{A.61}$$

where $Q_1(x) = Q_{1n}(x)/Q_{1d}(x) = \theta^*(x)$ and $Q_2(x) = Q_{2n}(x)/Q_{2d}$ with

$$Q_{1n}(x) = \sum_{k\in\{1,0\}} \mathcal{P}_{y_1|t_k,z_1}(x)\mathcal{P}_{t_k|z_1}(x) - \sum_{k\in\{1,0\}} \mathcal{P}_{y_1|t_k,z_0}(x)\mathcal{P}_{t_k|z_0}(x),$$

$$Q_{1d}(x) = \mathcal{P}_{y_0|t_0,z_0}(x)\mathcal{P}_{t_0|z_0}(x) - \mathcal{P}_{y_0|t_0,z_1}(x)\mathcal{P}_{t_0|z_1}(x),$$

$$Q_{2n}(x) = \mathcal{P}_{t_1|z_1}(x) - \mathcal{P}_{t_1|z_0}(x),$$

$$Q_{2d} = \int \{\mathcal{P}_{t_1|z_1}(x) - \mathcal{P}_{t_1|z_0}(x)\}f(x)dx.$$

Further, we define the following functions:

$$F_1^*(Y,T,Z,X) = \frac{TZ}{\mathcal{P}_{t_1|z_1}(X)\mathcal{P}_{z_1}(X)}\{Y-\mathcal{P}_{y_1|t_1,z_1}(X)\}\frac{Q_2(X)}{Q_{1d}(X)}\mathcal{P}_{t_1|z_1}(X),$$

$$F_2^*(Y,T,Z,X) = \frac{(1-T)Z}{\mathcal{P}_{t_0|z_1}(X)\mathcal{P}_{z_1}(X)}\{Y-\mathcal{P}_{y_1|t_0,z_1}(X)\}\frac{Q_2(X)}{Q_{1d}(X)}\{1-Q_1(X)\}\mathcal{P}_{t_0|z_1}(X),$$

$$F_3^*(Y,T,Z,X) = -\frac{T(1-Z)}{\mathcal{P}_{t_1|z_0}(X)\mathcal{P}_{z_0}(X)}\{Y-\mathcal{P}_{y_1|t_1,z_0}(X)\}\frac{Q_2(X)}{Q_{1d}(X)}\mathcal{P}_{t_1|z_0}(X),$$

$$
\begin{aligned}
F_4^*(Y,T,Z,X) &= -\frac{(1-T)(1-Z)}{\mathcal{P}_{t_0|z_0}(X)\mathcal{P}_{z_0}(X)}\{Y-\mathcal{P}_{y_1|t_0,z_0}(X)\}\frac{Q_2(X)}{Q_{1d}(X)}\{1-Q_1(X)\}\mathcal{P}_{t_0|z_0}(X),\\
F_5^*(Y,T,Z,X) &= \frac{Z\{T-\mathcal{P}_{t_1|z_1}(X)\}}{\mathcal{P}_{z_1}(X)}\left(\frac{Q_2(X)}{Q_{1d}(X)}\Big[\mathcal{P}_{y_1|t_1,z_1}(X)-\{1-Q_1(X)\}\mathcal{P}_{y_1|t_0,z_1}(X)-Q_1(X)\Big]+\frac{Q_1(X)}{Q_{2d}}-S_2\right),\\
F_6^*(Y,T,Z,X) &= -\frac{(1-Z)\{T-\mathcal{P}_{t_1|z_0}(X)\}}{\mathcal{P}_{z_0}(X)}\\
&\quad\times\left(\frac{Q_2(X)}{Q_{1d}(X)}\Big[\mathcal{P}_{y_1|t_1,z_0}(X)-\{1-Q_1(X)\}\mathcal{P}_{y_1|t_0,z_0}(X)-Q_1(X)\Big]-\frac{Q_1(X)}{Q_{2d}}+S_2\right),\\
F_7^*(Y,T,Z,X) &= Q_1(X)Q_2(X)-S_2\Big\{\mathcal{P}_{t_1|z_1}(X)-\mathcal{P}_{t_1|z_0}(X)-Q_{2d}\Big\}-\theta^*,
\end{aligned}
$$

where $S_2=\int Q_1(x)Q_2(x)f(x)dx/Q_{2d}$.

Lemma A.12. *The pathwise derivative of θ^* is given by the function F^* defined by $F^*(Y,T,Z,X)=\sum_{j=1}^{7}F_j^*(Y,T,Z,X)$.*

Proof. Let $\bar{\theta}^*(\gamma)$ be the parameter corresponding to θ^* along regular parametric submodels indexed by γ: i.e.

$$
\bar{\theta}^*(\gamma)=\int Q_1(x|\gamma)Q_2(x|\gamma)f(x|\gamma)dx,
$$

where $Q_1(x|\gamma)$, $Q_2(x|\gamma)$, and $f(x|\gamma)$ are naturally defined from Q_1, Q_2, and f, respectively.

We now calculate the derivatives of the Q functions with respect to γ (evaluated at γ_0). For this calculation there are only seven relevant derivatives, i.e.

$$
\begin{aligned}
&\frac{\partial\mathcal{P}_{y_1|t_1,z_1}(x|\gamma)}{\partial\gamma},\frac{\partial\mathcal{P}_{y_1|t_0,z_1}(x|\gamma)}{\partial\gamma},\frac{\partial\mathcal{P}_{y_1|t_1,z_0}(x|\gamma)}{\partial\gamma},\frac{\partial\mathcal{P}_{y_1|t_0,z_0}(x|\gamma)}{\partial\gamma},\\
&\frac{\partial\mathcal{P}_{t_1|z_1}(x|\gamma)}{\partial\gamma},\frac{\partial\mathcal{P}_{t_1|z_0}(x|\gamma)}{\partial\gamma},\\
&\frac{\partial f(x|\gamma)}{\partial\gamma}.
\end{aligned}
$$

Consider the easy ones first:

$$
\begin{aligned}
\frac{\partial Q_{2n}(x|\gamma)}{\partial\gamma} &= \frac{\partial\mathcal{P}_{t_1|z_1}(x|\gamma)}{\partial\gamma}-\frac{\partial\mathcal{P}_{t_1|z_0}(x|\gamma)}{\partial\gamma},\\
\frac{\partial Q_{2d}(\gamma)}{\partial\gamma} &= \int\Big\{\frac{\partial\mathcal{P}_{t_1|z_1}(x|\gamma)}{\partial\gamma}-\frac{\partial\mathcal{P}_{t_1|z_0}(x|\gamma)}{\partial\gamma}\Big\}f(x)dx+\int\{\mathcal{P}_{t_1|z_1}(x)-\mathcal{P}_{t_1|z_0}(x)\}\frac{\partial f(x|\gamma)}{\partial\gamma}dx.
\end{aligned}
$$

Now, here are the messier ones:

$$
\begin{aligned}
\frac{\partial Q_{1n}(x|\gamma)}{\partial\gamma} &= \sum_{k\in\{1,0\}} \frac{\partial \mathcal{P}_{y_1|t_k,z_1}(x|\gamma)}{\partial\gamma}\mathcal{P}_{t_k|z_1}(x) + \left\{\mathcal{P}_{y_1|t_1,z_1}(x) - \mathcal{P}_{y_1|t_0,z_1}(x)\right\}\frac{\partial \mathcal{P}_{t_1|z_1}(x|\gamma)}{\partial\gamma} \\
&\quad - \sum_{k\in\{1,0\}} \frac{\partial \mathcal{P}_{y_1|t_k,z_0}(x|\gamma)}{\partial\gamma}\mathcal{P}_{t_k|z_0}(x) - \left\{\mathcal{P}_{y_1|t_1,z_0}(x) - \mathcal{P}_{y_1|t_0,z_0}(x)\right\}\frac{\partial \mathcal{P}_{t_1|z_0}(x|\gamma)}{\partial\gamma}, \\
\frac{\partial Q_{1d}(x|\gamma)}{\partial\gamma} &= \frac{\partial \mathcal{P}_{y_1|t_0,z_1}(x|\gamma)}{\partial\gamma}\mathcal{P}_{t_0|z_1}(x) - \frac{\partial \mathcal{P}_{y_1|t_0,z_0}(x|\gamma)}{\partial\gamma}\mathcal{P}_{t_0|z_0}(x) \\
&\quad + \mathcal{P}_{y_0|t_0,z_1}\frac{\partial \mathcal{P}_{t_1|z_1}(x|\gamma)}{\partial\gamma} - \mathcal{P}_{y_0|t_0,z_0}\frac{\partial \mathcal{P}_{t_1|z_0}(x|\gamma)}{\partial\gamma}.
\end{aligned}
$$

Now, note that $\partial\theta^*(\gamma)/\partial\gamma$ is the sum of the following terms:

$$\int \frac{\partial Q_1(x|\gamma)}{\partial\gamma}Q_2(x)f(x)dx = \int \frac{Q_2(x)}{Q_{1d}(x)}A_1(x|\gamma)f(x)dx \tag{A.62}$$

$$\int Q_1(x)\frac{\partial Q_2(x|\gamma)}{\partial\gamma}f(x)dx = \int \frac{Q_1(x)}{Q_{2d}}A_2(x|\gamma)f(x)dx \tag{A.63}$$

$$\int Q_1(x)Q_2(x)\frac{\partial f(x|\gamma)}{\partial\gamma}dx, \tag{A.64}$$

where

$$A_1(x|\gamma) = \frac{\partial Q_{1n}(x|\gamma)}{\partial\gamma} - Q_1(x)\frac{\partial Q_{1d}(x|\gamma)}{\partial\gamma} \quad \text{and} \quad A_2(x|\gamma) = \frac{\partial Q_{2n}(x|\gamma)}{\partial\gamma} - Q_2(x)\frac{\partial Q_{2d}(\gamma)}{\partial\gamma}.$$

Again, consider the easy one first:

$$A_2(x|\gamma) = A_{21}(x|\gamma) + A_{22}(x|\gamma) + A_{23}(x|\gamma), \tag{A.65}$$

where

$$
\begin{aligned}
A_{21}(x|\gamma) &= \frac{\partial \mathcal{P}_{t_1|z_1}(x|\gamma)}{\partial\gamma} - Q_2(x)\int \frac{\partial \mathcal{P}_{t_1|z_1}(r|\gamma)}{\partial\gamma}f(r)dr \\
A_{22}(x|\gamma) &= -\frac{\partial \mathcal{P}_{t_1|z_0}(x|\gamma)}{\partial\gamma} + Q_2(x)\int \frac{\partial \mathcal{P}_{t_1|z_0}(r|\gamma)}{\partial\gamma}f(r)dr, \\
A_{23}(x|\gamma) &= -Q_2(x)\int \left\{\mathcal{P}_{t_1|z_1}(r) - \mathcal{P}_{t_1|z_0}(r)\right\}\frac{\partial f(r|\gamma)}{\partial\gamma}dr.
\end{aligned}
$$

Here, note that

$$\int \frac{Q_1(x)}{Q_{2d}}A_{21}(x|\gamma)f(x)dx = \int \left\{\frac{Q_1(x)}{Q_{2d}} - S_2\right\}f(x)\frac{\partial \mathcal{P}_{t_1|z_1}(x|\gamma)}{\partial\gamma}dx, \tag{A.66}$$

$$\int \frac{Q_1(x)}{Q_{2d}}A_{22}(x|\gamma)f(x)dx = -\int \left\{\frac{Q_1(x)}{Q_{2d}} - S_2\right\}f(x)\frac{\partial \mathcal{P}_{t_1|z_0}(x|\gamma)}{\partial\gamma}dx, \tag{A.67}$$

$$\int \frac{Q_1(x)}{Q_{2d}} A_{23}(x|\gamma) f(x) dx = -S_2 \int \left\{ \mathcal{P}_{t_1|z_1}(x) - \mathcal{P}_{t_1|z_0}(x) \right\} \frac{\partial f(x|\gamma)}{\partial \gamma} dx. \tag{A.68}$$

Next, here are the messier one:

$$A_1(x|\gamma) = A_{11}(x|\gamma) + A_{12}(x|\gamma) + A_{13}(x|\gamma) + A_{14}(x|\gamma) + A_{15}(x|\gamma) + A_{16}(x|\gamma), \tag{A.69}$$

where

$$A_{11}(x|\gamma) = \mathcal{P}_{t_1|z_1}(x) \frac{\partial \mathcal{P}_{y_1|t_1,z_1}(x|\gamma)}{\partial \gamma}, \tag{A.70}$$

$$A_{12}(x|\gamma) = \{1 - Q_1(x)\} \mathcal{P}_{t_0|z_1}(x) \frac{\partial \mathcal{P}_{y_1|t_0,z_1}(x|\gamma)}{\partial \gamma}, \tag{A.71}$$

$$A_{13}(x|\gamma) = -\mathcal{P}_{t_1|z_0}(x) \frac{\partial \mathcal{P}_{y_1|t_1,z_0}(x|\gamma)}{\partial \gamma}, \tag{A.72}$$

$$A_{14}(x|\gamma) = -\{1 - Q_1(x)\} \mathcal{P}_{t_0|z_0}(x) \frac{\partial \mathcal{P}_{y_1|t_0,z_0}(x|\gamma)}{\partial \gamma}, \tag{A.73}$$

$$A_{15}(x|\gamma) = \left[\mathcal{P}_{y_1|t_1,z_1}(x) - \{1 - Q_1(x)\} \mathcal{P}_{y_1|t_0,z_1}(x) - Q_1(x) \right] \frac{\partial \mathcal{P}_{t_1|z_1}(x|\gamma)}{\partial \gamma}, \tag{A.74}$$

$$A_{16}(x|\gamma) = -\left[\mathcal{P}_{y_1|t_1,z_0}(x) - \{1 - Q_1(x)\} \mathcal{P}_{y_1|t_0,z_0}(x) - Q_1(x) \right] \frac{\partial \mathcal{P}_{t_1|z_0}(x|\gamma)}{\partial \gamma}. \tag{A.75}$$

Therefore, combining equations (A.62) to (A.64), (A.66) to (A.68) and (A.70) to (A.75), we conclude that $\partial \theta^*(\gamma)/\partial \gamma$ is the sum of the following terms:

$$\int \frac{Q_2(x)}{Q_{1d}(x)} \mathcal{P}_{t_1|z_1}(x) f(x) \frac{\partial \mathcal{P}_{y_1|t_1,z_1}(x|\gamma)}{\partial \gamma} dx,$$

$$\int \frac{Q_2(x)}{Q_{1d}(x)} \{1 - Q_1(x)\} \mathcal{P}_{t_0|z_1}(x) f(x) \frac{\partial \mathcal{P}_{y_1|t_0,z_1}(x|\gamma)}{\partial \gamma} dx,$$

$$-\int \frac{Q_2(x)}{Q_{1d}(x)} \mathcal{P}_{t_1|z_0}(x) f(x) \frac{\partial \mathcal{P}_{y_1|t_1,z_0}(x|\gamma)}{\partial \gamma} dx,$$

$$-\int \frac{Q_2(x)}{Q_{1d}(x)} \{1 - Q_1(x)\} \mathcal{P}_{t_0|z_0}(x) f(x) \frac{\partial \mathcal{P}_{y_1|t_0,z_0}(x|\gamma)}{\partial \gamma} dx,$$

$$\int \left(\frac{Q_2(x)}{Q_{1d}(x)} \left[\mathcal{P}_{y_1|t_1,z_1}(x) - \{1 - Q_1(x)\} \mathcal{P}_{y_1|t_0,z_1}(x) - Q_1(x) \right] + \frac{Q_1(x)}{Q_{2d}} - S_2 \right) f(x) \frac{\partial \mathcal{P}_{t_1|z_1}(x|\gamma)}{\partial \gamma} dx,$$

$$-\int \left(\frac{Q_2(x)}{Q_{1d}(x)} \left[\mathcal{P}_{y_1|t_1,z_0}(x) - \{1 - Q_1(x)\} \mathcal{P}_{y_1|t_0,z_0}(x) - Q_1(x) \right] - \frac{Q_1(x)}{Q_{2d}} + S_2 \right) f(x) \frac{\partial \mathcal{P}_{t_1|z_0}(x|\gamma)}{\partial \gamma} dx,$$

$$\int \left[Q_1(x) Q_x(x) - S_2 \{ \mathcal{P}_{t_1|z_1}(x) - \mathcal{P}_{t_1|z_0}(x) \} \right] \frac{\partial f(x|\gamma)}{\partial \gamma} dx.$$

Finally, we note that the function F^* satisfies that

$$\frac{\partial\theta^*(\gamma)}{\partial\gamma} = \mathbb{E}\{F(Y,T,Z,X)s^*(Y,T,Z,X|\gamma_0)\},$$

where s^* is defined in equation (A.56): in fact, this equation can be seen immediately from the fact that s^* is the sum of Bernoulli scores.[25] □

Proof of Theorem A.4 It directly follows from lemmas A.10 and A.12. □

Proof of Theorem A.5 It follows from Theorem 2.1 in Newey (1994) and the fact that the scores given in equations (A.48) and (A.52) can approximate any mean zero random variable with an arbitrarily small mean squared error. □

References

ACKERBERG, D., X. CHEN, J. HAHN, AND Z. LIAO (2014): "Asymptotic efficiency of semiparametric two-step GMM," *Review of Economic Studies*, 81(3), 919–943.

AI, C., AND X. CHEN (2003): "Efficient estimation of models with conditional moment restrictions containing unknown functions," *Econometrica*, 71(6), 1795–1843.

——— (2012): "The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions," *Journal of Econometrics*, 170(2), 442–457.

BASSI, V., AND I. RASUL (2017): "Persuasion: A Case Study of Papal Influence on Fertility Preferences and Behavior," *American Economic Journal: Applied Economics*, 9(4), 250–302.

BERGEMANN, D., AND S. MORRIS (2019): "Information Design: A Unified Perspective," *Journal of Economic Literature*, 57(1), 44–95.

CALVI, R., A. LEWBEL, AND D. TOMMASI (2018): "Women's Empowerment and Family Heath: Estimating LATE with Mismeausred Treatment," Available at SSRN: https://ssrn.com/abstract=2980250.

[25] To see this point, it is helpful to consider a simple binary example. For instance, consider a generic binary variable B such that $\mathbb{P}(B=1|X,\gamma) = p_B(X|\gamma)$. Suppose that the parameter β satisfies $\partial\beta(\gamma)/\partial\gamma = \int A(x)\partial p_B(X|\gamma)/\partial\gamma f(x)dx = \mathbb{E}\{A(X)\partial p_B(X|\gamma)/\partial\gamma\}$. Here, the score of B given X is

$$s_B(B|X) = \frac{B - p_B(X)}{p_B(X)\{1-p_B(X)\}}\frac{\partial p_B(X|\gamma)}{\partial\gamma}.$$

Now, we are looking for the function $G^*(B,X)$ such that

$$\mathbb{E}\{G^*(B,X)s_B(B|X)\} = \mathbb{E}\Big\{A(X)\frac{\partial p_B(X|\gamma)}{\partial\gamma}\Big\},$$

where

$$\mathbb{E}\{G^*(B,X)s_B(B|X)\} = \mathbb{E}\Big[\{G^*(1,X) - G^*(0,X)\}\frac{\partial p_B(X|\gamma)}{\partial\gamma}\Big].$$

Therefore, we immediately see that $G^*(B,X) = \{B - p_B(X)\}A(X)$ does the job.

CARNEIRO, P., J. J. HECKMAN, AND E. VYTLACIL (2011): "Estimating marginal returns to education," *American Economic Review*, 101(6), 2754–2781.

CHE, Y.-K., W. DESSEIN, AND N. KARTIK (2013): "Pandering to Persuade," *American Economic Review*, 103(1), 47–79.

CHEN, X., O. LINTON, AND I. VAN KEILEGOM (2003): "Estimation of semiparametric models when the criterion function is not smooth," *Econometrica*, 71(5), 1591–1608.

CHERNOZHUKOV, V., S. LEE, AND A. ROSEN (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

CROSS, P. J., AND C. F. MANSKI (2002): "Regressions, Short and Long," *Econometrica*, 70(1), 357–368.

DEATON, A. (2010): "Instruments, Randomization, and Learning about Development," *Journal of Economic Literature*, 48(2), 424–55.

DELLAVIGNA, S., R. ENIKOLOPOV, V. MIRONOVA, M. PETROVA, AND E. ZHURAVSKAYA (2014): "Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia," *American Economic Journal: Applied Economics*, 6(3), 103–32.

DELLAVIGNA, S., AND M. GENTZKOW (2010): "Persuasion: empirical evidence," *Annual Review of Economics*, 2(1), 643–669.

DELLAVIGNA, S., AND E. KAPLAN (2007): "The Fox News effect: Media bias and voting," *Quarterly Journal of Economics*, 122(3), 1187–1234.

DELLAVIGNA, S., J. A. LIST, AND U. MALMENDIER (2012): "Testing for Altruism and Social Pressure in Charitable Giving," *Quarterly Journal of Economics*, 127(1), 1–56.

ENIKOLOPOV, R., M. PETROVA, AND E. ZHURAVSKAYA (2011): "Media and Political Persuasion: Evidence from Russia," *American Economic Review*, 101(7), 3253–85.

GENTZKOW, M. (2006): "Television and Voter Turnout," *Quarterly Journal of Economics*, 121(3), 931–972.

GENTZKOW, M., AND E. KAMENICA (2017): "Competition in Persuasion," *Review of Economic Studies*, 84(1), 300–322.

GENTZKOW, M., J. M. SHAPIRO, AND M. SINKINSON (2011): "The effect of newspaper entry and exit on electoral politics," *American Economic Review*, 101(7), 2980–3018.

GERBER, A. S., D. KARLAN, AND D. BERGAN (2009): "Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions," *American Economic Journal: Applied Economics*, 1(2), 35–52.

GREEN, D. P., AND A. S. GERBER (2000): "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment," *American Political Science Review*, 94(3), 653–663.

——— (2001): "Getting out the youth vote: Results from randomized field experiments," Natural Field Experiments 00260, The Field Experiments Website.

GREEN, D. P., A. S. GERBER, AND D. W. NICKERSON (2003): "Getting Out the Vote in Local Elections: Results from Six Door-to-Door Canvassing Experiments," *The Journal of Politics*, 65(4), 1083–1096.

HECKMAN, J. J. (2010): "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy," *Journal of Economic Literature*, 48(2), 356–98.

HECKMAN, J. J., AND R. PINTO (2018): "Unordered Monotonicity," *Econometrica*, 86(1), 1–35.

HECKMAN, J. J., J. SMITH, AND N. CLEMENTS (1997): "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts," *Review of Economic Studies*, 64(4), 487–535.

HECKMAN, J. J., AND E. VYTLACIL (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73(3), 669–738.

——— (2007): "Chapter 70 Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6 Part B, pp. 4779–4874. Elsevier.

IMBENS, G., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.

IMBENS, G. W. (2010): "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, 48(2), 399–423.

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and estimation of local average treatment effects," *Econometrica*, 62(2), 467–475.

IMBENS, G. W., AND D. B. RUBIN (1997): "Estimating outcome distributions for compliers in instrumental variables models," *Review of Economic Studies*, 64(4), 555–574.

IMBENS, G. W., AND J. M. WOOLDRIDGE (2009): "Recent developments in the econometrics of program evaluation," *Journal of Economic Literature*, 47(1), 5–86.

KAMENICA, E., AND M. GENTZKOW (2011): "Bayesian Persuasion," *American Economic Review*, 101(6), 2590–2615.

KIRBY, W. T. (2008): "The Public Sermon: Paul's Cross and the culture of persuasion in England, 1534-1570," *Renaissance and Reformation/Renaissance et Réforme*, pp. 3–29.

LANDRY, C. E., A. LANGE, J. A. LIST, M. K. PRICE, AND N. G. RUPP (2006): "Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment," *Quarterly Journal of Economics*, 121(2), 747–782.

LEE, S., AND SALANIÉ (2018): "Identifying Effects of Multivalued Treatments," *Econometrica*, 86(6), 1939–1963.

MANSKI, C. F. (1997): "Monotone treatment response," *Econometrica*, pp. 1311–1334.

——— (2003): *Partial Identification of Probability Distributions*. Springer-Verlag, New York.

——— (2007): *Identification for Prediction and Decision*. Harvard University Press, New York.

——— (2018): "Credible Ecological Inference for Medical Decisions with Personalized Risk Assessment," *Quantitative Economics*, 9(2), 541–569.

MANSKI, C. F., AND J. V. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68(4), 997–1010.

MARTIN, G. J., AND A. YURUKOGLU (2017): "Bias in Cable News: Persuasion and Polarization," *American Economic Review*, 107(9), 2565–99.

NEWEY, W. K. (1994): "The asymptotic variance of semiparametric estimators," *Econometricay*, pp. 1349–1382.

NGUIMKEU, P., A. DENTEH, AND R. TCHERNIS (2019): "On the estimation of treatment effects with endogenous misreporting," *Journal of Econometrics*, 208(2), 487–506.

PRAT, A. (2018): "Media Power," *Journal of Political Economy*, 126(4), 1747–1783.

RAPP, C. (2010): "Aristotle's Rhetoric," in *The Stanford Encyclopedia of Philosophy*, ed. by E. N. Zalta. Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, spring 2010 edn., https://plato.stanford.edu/archives/spr2010/entries/aristotle-rhetoric/.

STOYE, J. (2009): "More on Confidence Regions for Partially Identified Parameters," *Econometrica*, 77(4), 1299–1315.

TAMER, E. (2010): "Partial Identification in Econometrics," *Annual Review of Economics*, 2, 167–195.

URA, T. (2018): "Heterogeneous Treatment Effects With Mismeasured Endogenous Treatment," *Quantitative Economics*, 9(3), 1335–1370.

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