

Causal Spillover Effects Using Instrumental Variables*

Gonzalo Vazquez-Bare[†]

September 8, 2020

Abstract

I set up a potential-outcomes framework to analyze spillover effects using instrumental variables. I characterize the population compliance types in a setting in which spillovers can occur on both treatment take-up and outcomes, and provide conditions for identification of the marginal distribution of these compliance types. I show that intention-to-treat (ITT) parameters aggregate multiple direct and spillover effects for different compliance types, and hence do not have a clear link to causally interpretable parameters. Moreover, rescaling ITT parameters by first-stage estimands generally recovers a weighted combination of average effects where the sum of weights is larger than one. I then analyze identification of causal direct and spillover effects under one-sided noncompliance, and propose simple estimators that are consistent and asymptotically normal under mild conditions. I use the proposed methods to analyze an experiment on social interactions and voting behavior.

Keywords: causal inference, spillover effects, instrumental variables, treatment effects.

*I thank Matias Cattaneo, Clément de Chaisemartin, Xinwei Ma, Kenichi Nagasawa, Olga Namen, Dick Startz and Doug Steigerwald for valuable discussions and suggestions that greatly improved the paper. I also thank participants of the UCSB Applied Econometrics Research Group and seminar participants at Stanford University, UCSB Applied Microeconomics Lunch, Northwestern University, UCLA, UC San Diego and University of Chicago for helpful comments.

[†]Department of Economics, University of California, Santa Barbara. gvazquez@econ.ucsb.edu.

1 Introduction

An accurate assessment of spillover effects is crucial to understand the costs and benefits of policies or treatments (Athey and Imbens, 2017; Abadie and Cattaneo, 2018). Previous literature has shown that appropriately designed randomized controlled trials (RCTs) are a powerful tool to analyze spillovers (Moffit, 2001; Duflo and Saez, 2003; Hudgens and Halloran, 2008; Vazquez-Bare, 2017; Baird et al., 2018). However, RCTs are often subject to imperfect compliance, which can render actual treatment take-up endogenous even when the treatment assignment was random. In other cases, researchers may not have access to an RCT, and instead may need to rely on quasi-experimental variation from a natural experiment (see e.g. Angrist and Krueger, 2001; Titiunik, 2019).

While instrumental variables (IVs) have been a workhorse for addressing endogeneity when evaluating policy effects (Angrist, Imbens and Rubin, 1996; Imbens and Wooldridge, 2009; Abadie and Cattaneo, 2018), little is known about what an IV can identify in the presence of spillovers. This paper provides a framework to study causal spillover effects using instrumental variables in a setting in which spillovers occur between pairs such as couples, roommates, siblings, etc.

This paper offers three main contributions. First, Section 2 defines causal direct and spillover effects under two-sided noncompliance and shows that, when treatment take-up is endogenous, spillover effects can occur in both the treatment take-up and the outcome. I propose a generalization of the monotonicity assumption (Imbens and Angrist, 1994) that partitions the population into five compliance types. In addition to the usual always-takers, compliers and never-takers, units may be social-interaction compliers, who receive the treatment as soon as their peer is assigned to it, and group compliers, who only receive the treatment when both themselves and their peer are assigned to it. Section 3 provides conditions for identification of the marginal distribution of compliance types, and shows that the joint distribution is generally not identified.

Second, Section 4 analyzes intention-to-treat (ITT) parameters and shows that these estimands conflate multiple direct and indirect effects for different compliance subpopulations, and hence do not have a clear interpretation in general. Moreover, I show that rescaling the ITT by the first-stage estimand, which would recover the average effect on compliers in the absence of spillovers, will generally yield a weighted average of direct and spillover effects where the sum of the weights exceeds one.

Third, Section 5 shows that, when noncompliance is one-sided, it is possible to identify the average direct effect on compliers and the average spillover effect on units with compliant peers, and provides a way to assess the external validity of these parameters. Moreover, I show that these direct and indirect local average effects can be written as two-stage least squares (2SLS) estimands and can thus be estimated using standard regression methods.

Section 6 shows that all the parameters of interest can be consistently estimated as nonlinear combinations of sample means, and are asymptotically normal under standard conditions.

In Section 7, I implement the proposed methods to study the effect of social interactions in the household on voter turnout. I reanalyze the experiment conducted by [Foos and de Rooij \(2017\)](#) in which they randomly assign two-voter households to receive telephone calls encouraging them to vote on the West Midlands Police and Crime Commissioner election in Birmingham, UK. This experiment provides an ideal setting to employ the methods I propose, as this type of voter mobilization programs are commonly subject to severe noncompliance ([Gerber and Green, 2000](#); [John and Brannan, 2008](#)). I find evidence of large and statistically significant local average direct and spillover effects, and strong evidence of heterogeneity across compliance types. I also show that a simple 2SLS regression that ignores the presence of spillovers grossly underestimates the effects of the treatment, and I interpret these estimates in light of my framework.

Finally, Section 8 discusses generalizations of the results to IV conditional-on-observables and arbitrary group sizes, and Section 9 concludes.

This paper contributes to the literature on causal inference under interference ([Rosenbaum, 2007](#); [Hudgens and Halloran, 2008](#); [Tchetgen Tchetgen and VanderWeele, 2012](#); [Halloran and Hudgens, 2016](#); [Basse and Feller, 2018](#); [Basse et al., 2019](#)) by considering imperfect compliance. In earlier work, [Sobel \(2006\)](#) studied the performance of IVs under a completely randomized experiment with one-sided noncompliance when the SUTVA is violated. More recent studies include [Kang and Imbens \(2016\)](#), who analyze identification and estimation under personalized assignment (i.e. no spillovers on treatment take-up), [Kang and Keele \(2018\)](#) who provide bounds for average effects on compliers in cluster randomized trials and [Imai et al. \(forthcoming\)](#) who focus on estimands that average over peers’ assignments in a two-stage experimental design. My findings add to this literature by introducing a novel set of estimands and identification conditions that are independent of the experimental design and that simultaneously allow for spillovers on outcomes, spillovers on treatment take-up and multiple compliance types in a superpopulation setting.

This paper is also related to the literature on multiple instruments ([Imbens and Angrist, 1994](#); [Mogstad et al., 2019](#)). While most of the literature studies the use of multiple instruments for a single binary instrument, my setting with unrestricted spillovers introduces both multiple instruments and multiple treatments. In related work, [Blackwell \(2017\)](#) analyzes the case of two instruments and two treatments under a treatment exclusion restriction assumption that rules out “cross effects” of instruments on treatments. My findings complement these existing results by allowing for spillover effects on treatments, which generates a novel set of compliance types and causal effects.

2 Setup

Consider a random sample of independent and identically distributed households indexed by $g = 1, \dots, G$, each with 2 identically-distributed units, so that each unit i in group g has one peer. This setup has a wide range of applications in which groups consist of couples, roommates, siblings, etc (see e.g. [Babcock et al., 2015](#); [Fletcher and Marksteiner, 2017](#); [Foos and de Rooij, 2017](#); [Sacerdote, 2001](#)). Spillovers are assumed to occur between units in the same group, but not between units in different groups.

The goal is to study the effect of a binary treatment on an outcome of interest. The treatment can be endogenous in the sense that it is allowed to be arbitrarily correlated with the potential outcomes. To address this endogeneity, I will use a binary instrumental variable that can be considered “as if randomly assigned”, as formalized below.

In this setting, spillovers can occur at two different stages: treatment take-up and outcomes. The first stage occurs when an individual’s decision to take up treatment depends on the values of the peers’ instrument. To fix ideas, consider an encouragement design in which smoking spouses are randomly assigned to a smoking cessation program, as in [Fletcher and Marksteiner \(2017\)](#). In this setting, it is possible that a person that is not assigned to the program decides to attend because her spouse is assigned to do so. The second stage in which spillovers can materialize is the outcome stage. For example, an individual who did not attend the smoking cessation program can learn about the health risks of smoking through their spouse and decide to quit smoking.

Individual treatment status of unit i in group g is denoted by D_{ig} , taking values $d \in \{0, 1\}$. For each unit i , D_{jg} with $j \neq i$ is the treatment indicator corresponding to unit i ’s peer. For a given realization of the treatment status $(D_{ig}, D_{jg}) = (d, d')$, the potential outcome for unit i in group g is a random variable denoted by $Y_{ig}(d, d')$. In this setting, we say there are spillover effects on unit i ’s outcome when $Y_{ig}(d, 1) - Y_{ig}(d, 0) \neq 0$ for some $d = 0, 1$. The observed outcome for unit i in group g is the value of the potential outcome under the observed treatment realization, given by $Y_{ig} = Y_{ig}(D_{ig}, D_{jg})$. The observed outcome can be written as:

$$Y_{ig} = \sum_{d \in \{0,1\}} \sum_{d' \in \{0,1\}} Y_{ig}(d, d') \mathbb{1}(D_{ig} = d) \mathbb{1}(D_{jg} = d').$$

Let (Z_{ig}, Z_{jg}) be the vector of instruments for unit i and her peer, taking values $(z, z') \in \{0, 1\}^2$. Borrowing from the literature on imperfect compliance in RCT’s, I will often refer to the instruments (Z_{ig}, Z_{jg}) as “treatment assignments”. However, all the results in the paper apply not only to cases in which the researcher has control on the assignment mechanism of (Z_{ig}, Z_{jg}) , as in an encouragement design, but also to cases in which the instruments come from a natural experiment (see e.g. [Angrist and Krueger, 2001](#); [Titunik, 2019](#)).

The potential treatment status for unit i in group g will be denoted by $D_{ig}(z, z')$, and we say there are spillover effects on unit i 's treatment status if $D_{ig}(z, 1) - D_{ig}(z, 0) \neq 0$ for some $z = 0, 1$. The observed treatment status is $D_{ig}(Z_{ig}, Z_{jg})$. The following assumption imposes some restrictions on the relationship between potential outcomes, potential treatment statuses and the instruments.

Assumption 1 (Existence of instruments)

- (a) *Exclusion restriction:* $Y_{ig}(d, d')$ is not a function of (z, z') .
- (b) *Independence:* for all $i, j \neq i$ and g , $((Y_{ig}(d, d'))_{(d, d')}, (D_{ig}(z, z'))_{(z, z')}) \perp (Z_{ig}, Z_{jg})$.

Part (a) asserts that the instrument does not have a direct effect on the potential outcome. Part (b) imposes statistical independence between the treatment assignment and potential outcomes and treatment statuses, and hence the instrument can be considered “as-if” randomly assigned. Section 8.1 offers an alternative version of this assumption in which independence holds after conditioning on a set of observable covariates.

A unit's compliance type is determined by the vector $(D_{ig}(0, 0), D_{ig}(0, 1), D_{ig}(1, 0), D_{ig}(1, 1))$, which indicates the unit's treatment status for each possible instrument assignment. For example, a unit with $D_{ig}(z, z') = 0$ for all (z, z') always refuses the treatment regardless of her own and her peer's assignment. A unit with $D_{ig}(z, z') = 1$ for all (z, z') always receives the treatment regardless of her own and peer's assignment. A unit with $D_{ig}(1, 1) = D_{ig}(1, 0) = 1$ and $D_{ig}(0, 1) = D_{ig}(0, 0) = 0$ only receives the treatment when she is assigned to it, regardless of her peer's assignment, and so on. Without further restrictions, there are a total of 16 different compliance types in the population. I will introduce the following monotonicity assumption to restrict the number of compliance types.

Assumption 2 (Monotonicity) For all i and g ,

- (a) $D_{ig}(1, z') \geq D_{ig}(0, z')$ for $z' = 0, 1$
- (b) $D_{ig}(z, 1) \geq D_{ig}(z, 0)$ for $z = 0, 1$,
- (c) $D_{ig}(1, 0) \geq D_{ig}(0, 1)$.

Part (a) states that, for a fixed peer's assignment z' , being assigned to treatment cannot push unit i away from treatment. Part (b) states that, for a fixed own assignment z , having a peer assigned to treatment cannot push unit i away from the treatment. Finally, part (c) states that if unit i is treated when only her peer is assigned to it ($D_{ig}(0, 1) = 1$), then she would also be treated when she is assigned to treatment ($D_{ig}(1, 0) = 1$), and that if unit i is not treated when she is the only one assigned to treatment ($D_{ig}(1, 0) = 0$), she would not be treated when her peer is the only one assigned to treatment ($D_{ig}(0, 1) = 0$). In other words, condition (c) means that the effect of own assignment on treatment take-up cannot

Table 1: Population types

$D_{ig}(1, 1)$	$D_{ig}(1, 0)$	$D_{ig}(0, 1)$	$D_{ig}(0, 0)$	Type
1	1	1	1	Always-taker (AT)
1	1	1	0	Social-interaction complier (SC)
1	1	0	0	Complier (C)
1	0	0	0	Group complier (GC)
0	0	0	0	Never-taker (NT)

be “weaker” than the effect of peer’s assignment. Note that this assumption is not testable, as one can only observe one out of the four possible potential treatment statuses, and hence its validity needs to be assessed on a case-by-case basis.

Assumption 2 implies the following ordering:

$$D_{ig}(1, 1) \geq D_{ig}(1, 0) \geq D_{ig}(0, 1) \geq D_{ig}(0, 0),$$

which reduces the compliance types to five. Table 1 lists the five different compliance types in the population under Assumption 2. Always-takers (AT) are units who receive treatment regardless of own and peer treatment assignment. Social-interaction compliers (SC), a term coined by Duflo and Saez (2003), are units who receive the treatment as soon as someone in their group (either themselves or their peer) is assigned to it. Compliers (C) are units that receive the treatment if and only if they are assigned to it. Group compliers (GC) are units who only receive the treatment when their whole group (i.e. both themselves and their peer) is assigned to treatment. Finally, never-takers (NT) are never treated regardless of own and peer’s assignment. The categories in Table 1 are listed in decreasing order of likelihood of being treated.

In what follows, let ξ_{ig} denote a random variable indicating unit i ’s compliance type, $\xi_{ig} \in \{\text{AT}, \text{SC}, \text{C}, \text{GC}, \text{NT}\}$. Also, let C_{ig} denote the event that unit i in group g is a complier, $C_{ig} = \{\xi_{ig} = \text{C}\}$, and similarly for $AT_{ig} = \{\xi_{ig} = \text{AT}\}$, $SC_{ig} = \{\xi_{ig} = \text{SC}\}$ and so on.

Finally, in order to exploit variation in (Z_{ig}, Z_{jg}) to identify causal effects, we need to ensure that the instruments have a non-trivial effect on treatment take-up, usually known as “instrument relevance”. In this setting, this requirement can be stated as follows.

Assumption 3 (Relevance) $\mathbb{P}[AT_{ig}] + \mathbb{P}[NT_{ig}] < 1$.

Assumption 3 rules out cases in which all units in the population are combinations of always-takers and never-takers only, which are the cases in which treatment take-up is not affected by the instruments, that is, $D_{ig}(z, z')$ takes on the same value for all combinations of z and z' . Some identification results will require strengthening this assumption, as will be made clear in the upcoming sections.

In what follows, I analyze identification of the distribution of compliance types, intention-to-treat parameters and local average treatment effects.

3 Distribution of Compliance Types

Under Assumptions 1 and 2, the marginal distribution of compliance types in the population is identified, as the following proposition shows.

Proposition 1 (Distribution of compliance types) *Under Assumptions 1 and 2,*

$$\begin{aligned}\mathbb{P}[AT_{ig}] &= \mathbb{E}[D_{ig}|Z_{ig} = 0, Z_{jg} = 0] \\ \mathbb{P}[SC_{ig}] &= \mathbb{E}[D_{ig}|Z_{ig} = 0, Z_{jg} = 1] - \mathbb{E}[D_{ig}|Z_{ig} = 0, Z_{jg} = 0] \\ \mathbb{P}[C_{ig}] &= \mathbb{E}[D_{ig}|Z_{ig} = 1, Z_{jg} = 0] - \mathbb{E}[D_{ig}|Z_{ig} = 0, Z_{jg} = 1] \\ \mathbb{P}[GC_{ig}] &= \mathbb{E}[D_{ig}|Z_{ig} = 1, Z_{jg} = 1] - \mathbb{E}[D_{ig}|Z_{ig} = 1, Z_{jg} = 0]\end{aligned}$$

and $\mathbb{P}[NT_{ig}] = 1 - \mathbb{P}[AT_{ig}] - \mathbb{P}[SC_{ig}] - \mathbb{P}[C_{ig}] - \mathbb{P}[GC_{ig}]$. Finally,

$$\begin{aligned}\mathbb{P}[AT_{ig}, AT_{jg}] &= \mathbb{E}[D_{ig}D_{jg}|Z_{ig} = 0, Z_{jg} = 0] \\ \mathbb{P}[NT_{ig}, NT_{jg}] &= \mathbb{E}[(1 - D_{ig})(1 - D_{jg})|Z_{ig} = 1, Z_{jg} = 1].\end{aligned}$$

All the proofs can be found in the supplemental appendix. Notice that the entire joint distribution of compliance types is not identified without further assumptions. The proof in the supplemental appendix shows that the possible combinations of $(D_{ig}, D_{jg}, Z_{ig}, Z_{jg})$ in this setting provide a system of 7 linearly independent equations which identify four marginal probabilities (where the fifth one is recovered from the restriction that they sum to one), the two joint probabilities $\mathbb{P}[AT_{ig}, AT_{jg}]$ and $\mathbb{P}[NT_{ig}, NT_{jg}]$, and the sum of probabilities $\mathbb{P}[AT_{ig}, NT_{jg}] + \mathbb{P}[AT_{ig}, GC_{jg}] + \mathbb{P}[SC_{ig}, GC_{jg}] + \mathbb{P}[SC_{ig}, NT_{jg}]$

Proposition 1 can be used to test for the presence of average spillover effects on treatment status. Note that under Assumption 2, $\mathbb{E}[D_{ig}(0, 1) - D_{ig}(0, 0)] = \mathbb{P}[SC_{ig}]$ and $\mathbb{E}[D_{ig}(1, 1) - D_{ig}(1, 0)] = \mathbb{P}[GC_{ig}]$, and thus testing for the presence of average spillover effects on treatment status amounts to testing for the presence of social-interaction compliers and group compliers. Because the instrument is as-if randomly assigned, these issues can be analyzed within the framework in Vazquez-Bare (2017).

4 Intention-to-Treat Analysis

Intention-to-treat (ITT) analysis focuses on the variation in Y_{ig} generated by the instrument. [Imbens and Angrist \(1994\)](#) showed that, in the absence of spillovers, the ITT estimand $\mathbb{E}[Y_{ig}|Z_{ig} = 1] - \mathbb{E}[Y_{ig}|Z_{ig} = 0]$ is an attenuated measure of the average treatment effect on compliers, or local average treatment effect (LATE). Furthermore, the LATE can be easily recovered by rescaling the ITT parameter by the proportion of compliers, which is identified under monotonicity and as-if random assignment of the instrument. Importantly, even in cases where the LATE is not identified (for example, when actual treatment status is not observed), the ITT parameter still provides valuable information as it measures the effect of offering the treatment. This section shows that, in the presence of spillovers, the link between ITT parameters and local average effects is much less clear, as the former will conflate multiple potentially different effects into a single number that may be hard to interpret in the presence of effect heterogeneity.

I will focus on the observed conditional means:

$$\mathbb{E}[Y_{ig}|Z_{ig} = z, Z_{jg} = z']$$

exploiting variation over assignments (z, z') . I will refer to differences in average outcomes changing own instrument leaving the peer's instrument fixed as *direct* ITT parameters:

$$\mathbb{E}[Y_{ig}|Z_{ig} = 1, Z_{jg} = z'] - \mathbb{E}[Y_{ig}|Z_{ig} = 0, Z_{jg} = z']$$

and differences fixing own instrument and varying the peer's instrument as *indirect* or *spillover* ITT parameters:

$$\mathbb{E}[Y_{ig}|Z_{ig} = z, Z_{jg} = 1] - \mathbb{E}[Y_{ig}|Z_{ig} = z, Z_{jg} = 0].$$

Finally, the *total* ITT is defined as $\mathbb{E}[Y_{ig}|Z_{ig} = 1, Z_{jg} = 1] - \mathbb{E}[Y_{ig}|Z_{ig} = z, Z_{jg} = 0]$.

The following result links the direct ITT estimand to potential outcomes. In what follows, the notation $\{C_{ig}, SC_{ig}\} \times \{AT_{jg}\}$ refers to the event $(C_{ig} \cap AT_{jg}) \cup (SC_{ig} \cap AT_{jg})$, that is, unit j is an always-taker and unit i can be a complier or a social complier. Similarly, $\{C_{ig}, SC_{ig}\} \times \{C_{jg}, GC_{jg}, NT_{jg}\}$ represents all the combinations in which unit i is a complier or a social complier and unit j is a complier, a group complier or a never-taker, and so on.

Lemma 1 (Direct ITT effects) *Under Assumptions 1-3,*

$$\begin{aligned}
& \mathbb{E}[Y_{ig}|Z_{ig} = 1, Z_{jg} = 0] - \mathbb{E}[Y_{ig}|Z_{ig} = 0, Z_{jg} = 0] = \\
& \quad \mathbb{E}[Y_{ig}(1, 0) - Y_{ig}(0, 0)|\{C_{ig}, SC_{ig}\} \times \{C_{jg}, GC_{jg}, NT_{jg}\}] \\
& \quad \times \mathbb{P}[\{C_{ig}, SC_{ig}\} \times \{C_{jg}, GC_{jg}, NT_{jg}\}] \\
& + \mathbb{E}[Y_{ig}(1, 1) - Y_{ig}(0, 0)|\{C_{ig}, SC_{ig}\} \times \{SC_{jg}\}] \\
& \quad \times \mathbb{P}[\{C_{ig}, SC_{ig}\} \times \{SC_{jg}\}] \\
& + \mathbb{E}[Y_{ig}(1, 1) - Y_{ig}(0, 1)|\{C_{ig}, SC_{ig}\} \times \{AT_{jg}\}] \\
& \quad \times \mathbb{P}[\{C_{ig}, SC_{ig}\} \times \{AT_{jg}\}] \\
& + \mathbb{E}[Y_{ig}(0, 1) - Y_{ig}(0, 0)|\{GC_{ig}, NT_{ig}\} \times \{SC_{jg}\}] \\
& \quad \times \mathbb{P}[\{GC_{ig}, NT_{ig}\} \times \{SC_{jg}\}] \\
& + \mathbb{E}[Y_{ig}(1, 1) - Y_{ig}(1, 0)|AT_{ig}, SC_{jg}] \\
& \quad \times \mathbb{P}[AT_{ig}, SC_{jg}].
\end{aligned}$$

The corresponding results for the indirect ITT and the total ITT are analogous, and are presented in Section A1 of the supplemental appendix to conserve space.

To understand the above result, consider the effect of switching Z_{ig} from 0 to 1, while leaving Z_{jg} fixed at zero. First, if unit i is either a complier or a social complier, switching Z_{ig} from 0 to 1 will change her treatment status D_{ig} from 0 to 1. This case corresponds to the first three expectations on the right-hand side of Lemma 1. Now, if unit j is a complier, a group complier or a never taker, her observed treatment status would be $D_{jg} = 0$. Hence, in these cases, switching Z_{ig} from 0 to 1 while leaving Z_{jg} fixed at zero would let us observe $Y_{ig}(1, 0) - Y_{ig}(0, 0)$. This corresponds to the first expectation on the right-hand side of Lemma 1. On the other hand, if unit j was a social complier, switching Z_{ig} from 0 to 1 would push her to get the treatment, and hence in this case we would see $Y_{ig}(1, 1) - Y_{ig}(0, 0)$. This case corresponds to the second expectation on the right-hand side of Lemma 1. If instead unit j was an always-taker, she would be treated in both scenarios, so we would see $Y_{ig}(1, 1) - Y_{ig}(0, 1)$ (third expectation of the above display). Next, suppose unit i was a group complier or a never-taker. Then, switching Z_{ig} from 0 to 1 would not affect her treatment status, which would be fixed at 0, but it would affect unit j 's treatment status if she is a social complier. This case is shown in the fourth expectation on the right-hand side of Lemma 1. Finally, if unit i was an always-taker, her treatment status would be fixed at 1 but her peer's treatment status would switch from 0 to 1 if unit j was a social complier. This case is shown in the last expectation on the right-hand side of Lemma 1.

Hence the direct ITT effect is averaging five different treatment effects, $Y_{ig}(1, 0) - Y_{ig}(0, 0)$, $Y_{ig}(1, 1) - Y_{ig}(0, 0)$, $Y_{ig}(1, 1) - Y_{ig}(1, 0)$, $Y_{ig}(0, 1) - Y_{ig}(0, 0)$, and $Y_{ig}(1, 1) - Y_{ig}(0, 1)$, each one over different combinations of compliance types. Therefore, Lemma 1 shows that, even

when fixing the peer's treatment assignment, the ITT parameter is unable to isolate direct and indirect effects, which may complicate its interpretation as a causal effect.

Remark 1 (Spillovers and instrument validity) *One way to interpret the result in Lemma 1 is to think of spillovers in treatment take-up as a violation of the exclusion restriction. Since D_{jg} is a function of Z_{ig} , the instrument Z_{ig} can affect the outcome Y_{ig} not only through the variable it is instrumenting, D_{ig} , but also through another variable D_{jg} . Thus, spillovers on treatment take-up may render an instrument invalid even when the instrument would have been valid in the absence of spillovers. This fact shows that identification of causal parameters based on (Z_{ig}, Z_{jg}) will require further assumptions, as discussed in the next section. \square*

Two further issues may complicate the interpretation of the estimand in Lemma 1. On the one hand, the ITT parameter is not a proper weighted average. Specifically, the weights (given by the probabilities of the compliance types combinations described above) are non-negative, but their sum is less than one. This happens because the probabilities of the cases in which units i and j 's treatment status do not change (e.g. when they are both always-takers or both never-takers) are multiplied by a zero and hence dropped from the sum.

On the other hand, based on the identification results in the absence of spillovers, one may be inclined to rescale the ITT by the first stage $\mathbb{E}[D_{ig}|Z_{ig} = 1, Z_{jg} = 0] - \mathbb{E}[D_{ig}|Z_{ig} = 0, Z_{jg} = 0]$. However, this rescaling makes the sum of the weights larger than one. More precisely, the weights from the direct ITT sum to $\mathbb{P}[C_{ig}] + \mathbb{P}[SC_{ig}] + \mathbb{P}[SC_{ig}, GC_{jg}] + \mathbb{P}[SC_{ig}, NT_{jg}] + \mathbb{P}[SC_{ig}, AT_{jg}]$, whereas $\mathbb{E}[D_{ig}|Z_{ig} = 1, Z_{jg} = 0] - \mathbb{E}[D_{ig}|Z_{ig} = 0, Z_{jg} = 0] = \mathbb{P}[C_{ig}] + \mathbb{P}[SC_{ig}]$ from Proposition 1. As an illustration, consider the (extreme) case in which types are independent and equally likely, so that $\mathbb{P}[\xi_{ig} = \xi, \xi_{jg} = \xi'] = 1/25$. In this case the rescaled weights equal $(0.6, 0.2, 0.2, 0.2, 0.1)$ which sum to 1.3. On the other hand, if the probabilities of always-takers, social compliers, compliers, group compliers and never-takers are, respectively, $(0.1, 0.2, 0.4, 0.2, 0.1)$, and types are independent, the weights become $(0.7, 0.2, 0.2, 0.1, 0.03)$ which sum to 1.13.

Remark 2 (Naive ITT) *The naive ITT, that is, the ITT that compares units with $Z_{ig} = 1$ to units with $Z_{ig} = 0$ ignoring the peer's assignment, can be written as:*

$$\begin{aligned} \mathbb{E}[Y_{ig}|Z_{ig} = 1] - \mathbb{E}[Y_{ig}|Z_{ig} = 0] &= (\mathbb{E}[Y_{ig}|Z_{ig} = 1, Z_{jg} = 0] - \mathbb{E}[Y_{ig}|Z_{ig} = 0, Z_{jg} = 0]) \\ &\quad \times \mathbb{P}[Z_{jg} = 0|Z_{ig} = 1] \\ &\quad + (\mathbb{E}[Y_{ig}|Z_{ig} = 1, Z_{jg} = 1] - \mathbb{E}[Y_{ig}|Z_{ig} = 0, Z_{jg} = 0]) \\ &\quad \times \mathbb{P}[Z_{jg} = 1|Z_{ig} = 1] \\ &\quad - (\mathbb{E}[Y_{ig}|Z_{ig} = 0, Z_{jg} = 1] - \mathbb{E}[Y_{ig}|Z_{ig} = 0, Z_{jg} = 0]) \\ &\quad \times \mathbb{P}[Z_{jg} = 1|Z_{ig} = 0] \end{aligned}$$

which equals the direct ITT plus the difference between the total and the indirect ITTs,

weighted by their corresponding probabilities. This implies that, without further assumptions, it is not possible to predict whether the presence of spillovers drives the usually employed naive ITT upward or downward, as this depends on the relative magnitudes of the different effects that are combined. \square

5 Identification Under One-sided Noncompliance

Since failure of point identification of average effects in this setting is due to imperfect compliance, identification of some causal parameters can be achieved by restricting the amount of noncompliance. In this section I will analyze the case in which noncompliance is one-sided. One-sided noncompliance refers to the case in which individual deviations from their assigned treatment, $D_{ig} \neq Z_{ig}$, can only occur in one direction.

In many applications, units who are not assigned to treatment are unable to get the treatment through other channels. For example, consider the experiment analyzed by Foos and de Rooij (2017) in which individuals in two-voter households are randomly assigned to receive a telephone call encouraging them to vote. In this case, units that are assigned $Z_{ig} = 1$ may fail to receive the actual phone call (e.g. because they don't pick up the phone), in which case $Z_{ig} = 1$ and $D_{ig} = 0$, but whenever a unit is assigned $Z_{ig} = 0$, this automatically implies $D_{ig} = 0$. I formalize this case as follows.

Assumption 4 (One-sided Noncompliance - OSN) *For all i and g , $D_{ig}(0, z') = 0$ for $z' = 0, 1$.*

Hence, one-sided noncompliance implies the absence of always-takers and social-interaction compliers, reducing the total number of compliance types from five to three: compliers, group compliers and never-takers. In other words, units cannot be treated unless they are themselves offered the treatment. Moreover, based on Proposition 1, this assumption can be tested by assessing whether $\mathbb{P}[AT_{ig}] = \mathbb{P}[SC_{ig}] = 0$.

In what follows, all the results focus on identifying the expectation of potential outcomes, but these results immediately generalize to identification of marginal distributions of potential outcomes by replacing Y_{ig} by $\mathbb{1}(Y_{ig} \leq y)$.

Proposition 2 (Local average potential outcomes under OSN) *Under Assumptions*

1-4, the following equalities hold:

$$\begin{aligned}
\mathbb{E}[Y_{ig}(0, 0)] &= \mathbb{E}[Y_{ig}|Z_{ig} = 0, Z_{jg} = 0] \\
\mathbb{E}[Y_{ig}(1, 0)|C_{ig}]\mathbb{P}[C_{ig}] &= \mathbb{E}[Y_{ig}D_{ig}|Z_{ig} = 1, Z_{jg} = 0] \\
\mathbb{E}[Y_{ig}(0, 1)|C_{jg}]\mathbb{P}[C_{jg}] &= \mathbb{E}[Y_{ig}D_{jg}|Z_{ig} = 0, Z_{jg} = 1] \\
\mathbb{E}[Y_{ig}(0, 0)|C_{ig}]\mathbb{P}[C_{ig}] &= \mathbb{E}[Y_{ig}|Z_{ig} = 0, Z_{jg} = 0] - \mathbb{E}[Y_{ig}(1 - D_{ig})|Z_{ig} = 1, Z_{jg} = 0] \\
\mathbb{E}[Y_{ig}(0, 0)|C_{jg}]\mathbb{P}[C_{jg}] &= \mathbb{E}[Y_{ig}|Z_{ig} = 0, Z_{jg} = 0] - \mathbb{E}[Y_{ig}(1 - D_{jg})|Z_{ig} = 0, Z_{jg} = 1] \\
\mathbb{E}[Y_{ig}(0, 0)|NT_{ig}, NT_{jg}]\mathbb{P}[NT_{ig}, NT_{jg}] &= \mathbb{E}[Y_{ig}(1 - D_{ig})(1 - D_{jg})|Z_{ig} = 1, Z_{jg} = 1]
\end{aligned}$$

where

$$\mathbb{P}[NT_{ig}, NT_{jg}] = \mathbb{E}[(1 - D_{ig})(1 - D_{jg})|Z_{ig} = 1, Z_{jg} = 1].$$

Combined with Proposition 1, the above result shows which local average potential outcomes can be identified by exploiting variation in the observed treatment status and assignments $(D_{ig}, D_{jg}, Z_{ig}, Z_{jg})$. This idea was proposed by Imbens and Rubin (1997) in a setting without spillovers.

Proposition 2 implies that the following treatment effects are also identified.

Corollary 1 (Local average direct and spillover effects under OSN) *Under Assumptions 1-4, if $\mathbb{P}[C_{ig}] > 0$,*

$$\mathbb{E}[Y_{ig}(1, 0) - Y_{ig}(0, 0)|C_{ig}] = \frac{\mathbb{E}[Y_{ig}|Z_{ig} = 1, Z_{jg} = 0] - \mathbb{E}[Y_{ig}|Z_{ig} = 0, Z_{jg} = 0]}{\mathbb{E}[D_{ig}|Z_{ig} = 1, Z_{jg} = 0]}$$

and

$$\mathbb{E}[Y_{ig}(0, 1) - Y_{ig}(0, 0)|C_{jg}] = \frac{\mathbb{E}[Y_{ig}|Z_{ig} = 0, Z_{jg} = 1] - \mathbb{E}[Y_{ig}|Z_{ig} = 0, Z_{jg} = 0]}{\mathbb{E}[D_{jg}|Z_{ig} = 0, Z_{jg} = 1]}.$$

In the above result, $\mathbb{E}[Y_{ig}(1, 0) - Y_{ig}(0, 0)|C_{ig}]$ represents the average direct effect on compliers with untreated peers, whereas $\mathbb{E}[Y_{ig}(0, 1) - Y_{ig}(0, 0)|C_{jg}]$ is the average effect on untreated units with compliant peers. See Section 7 for a detailed discussion on these estimands in the context of an empirical application.

In addition to identifying these treatment effects, Proposition 2 can be used to assess, at least partially, whether average potential outcomes vary across own and peer compliance types, as the following corollary shows. In what follows, C_{ig}^c represents the event in which unit i is not a complier, that is, $C_{ig}^c = NT_{ig} \cup GC_{ig}$.

Corollary 2 (Assessing heterogeneity over compliance types) *Under 1-4, if $0 <$*

$$\mathbb{P}[C_{ig}] < 1,$$

$$\begin{aligned} & \mathbb{E}[Y_{ig}(0,0)|C_{ig}] - \mathbb{E}[Y_{ig}(0,0)|C_{ig}^c] = \\ & \left\{ \frac{\mathbb{E}[Y_{ig}D_{ig}|Z_{ig}=1, Z_{jg}=0]}{\mathbb{E}[D_{ig}|Z_{ig}=1, Z_{jg}=0]} - \mathbb{E}[Y_{ig}|Z_{ig}=0, Z_{jg}=0] \right\} \frac{1}{1 - \mathbb{E}[D_{ig}|Z_{ig}=1, Z_{ig}=0]}. \end{aligned}$$

and

$$\begin{aligned} & \mathbb{E}[Y_{ig}(0,0)|C_{jg}] - \mathbb{E}[Y_{ig}(0,0)|C_{jg}^c] = \\ & \left\{ \frac{\mathbb{E}[Y_{ig}D_{jg}|Z_{ig}=0, Z_{jg}=1]}{\mathbb{E}[D_{jg}|Z_{ig}=0, Z_{jg}=1]} - \mathbb{E}[Y_{ig}|Z_{ig}=0, Z_{jg}=0] \right\} \frac{1}{1 - \mathbb{E}[D_{jg}|Z_{ig}=0, Z_{ig}=1]}. \end{aligned}$$

The first term in the above corollary, $\mathbb{E}[Y_{ig}(0,0)|C_{ig}] - \mathbb{E}[Y_{ig}(0,0)|C_{ig}^c]$, is the difference in the average baseline outcome $Y_{ig}(0,0)$ between compliers and non-compliers (i.e. group compliers or never-takers), whereas $\mathbb{E}[Y_{ig}(0,0)|C_{jg}] - \mathbb{E}[Y_{ig}(0,0)|C_{jg}^c]$ is the difference in average baseline potential outcomes among units with compliant and non-compliant peers. These differences can be used to determine whether average baseline potential outcomes vary with own and peers' compliance types, which can help assess the external validity of the local average effects. More precisely, if these differences are small, the local effects may be considered informative, at least to some extent, about average effects for the whole population, whereas finding marked heterogeneity across types would emphasize the local nature of the parameters in Corollary 1.

Finally, the following result shows that, under one-sided noncompliance, the direct and indirect local average effects in Corollary 1 are equal to the estimands from a 2SLS regression using $(Z_{ig}, Z_{jg}, Z_{ig}Z_{jg})$ as instruments for $(D_{ig}, D_{jg}, D_{ig}D_{jg})$.

Proposition 3 (2SLS) *Consider the regression:*

$$Y_{ig} = \beta_0 + \beta_1 D_{ig} + \beta_2 D_{jg} + \beta_3 D_{ig}D_{jg} + u_{ig}, \quad \mathbb{E}[u_{ig}|Z_{ig}, Z_{jg}] = 0$$

to be estimated by 2SLS using $(Z_{ig}, Z_{jg}, Z_{ig}Z_{jg})$ as instruments. Under Assumptions 1-3 and 4, if $0 < \mathbb{P}[C_{ig}] < 1$,

$$\begin{aligned} \beta_0 &= \mathbb{E}[Y_{ig}(0,0)] \\ \beta_1 &= \mathbb{E}[Y_{ig}(1,0) - Y_{ig}(0,0)|C_{ig}] \\ \beta_2 &= \mathbb{E}[Y_{ig}(0,1) - Y_{ig}(0,0)|C_{jg}]. \end{aligned}$$

The coefficient β_3 does not have a straightforward interpretation, as it combines several different effects. Its exact shape is shown in the proof of the proposition in the supplemental appendix.

6 Estimation and Inference

Because all the estimands proposed in the previous section are (nonlinear) combinations of cell averages, the parameters of interest can be estimated using sample counterparts. Inference on these magnitudes is straightforward based on the normal approximation and the delta method in a setting in which the number of groups G grows to infinity, and can allow for an unrestricted correlation structure for both outcomes and treatment assignments between groups. In particular, the estimands in Proposition 3 can be estimated using standard two-stage least squares methods.

Because estimation and inference can be conducted using standard procedures in this context and do not provide any specific challenge, I provide further details in the supplemental appendix to conserve space.

7 Application: Spillovers in Voting Behavior

In this section I illustrate the results in this paper using data from a randomized experiment on voter mobilization conducted by [Foos and de Rooij \(2017\)](#). Their study contributes to a literature analyzing the effect of social interactions on voting behavior (see e.g. [Sinclair, 2012](#)). Broadly, the goal is to assess if political discussions within close social networks such as the household have an effect on voter turnout, and, if so, in what direction. [Foos and de Rooij \(2017\)](#) study to what extent intrahousehold mobilization during an election campaign is conditioned by both the degree of heterogeneity of party preferences within the household and the partisan intensity of a campaign message.

To this end, the authors conducted a randomized experiment in which two-voter households in Birmingham, UK were randomly assigned to receive a telephone message encouraging people to vote on the West Midlands Police and Crime Commissioner (PCC) election, held on November 15, 2012. The telephone message was delivered by the Labour party volunteers, and provided information such as the election date and their local polling station, and encouraged people to vote. The experiment was designed as follows. A sample of 5,190 two-voter households with landline numbers were stratified into three blocks based on their last recorded party preference (Labor party supporter, rival party supported, unattached) and randomly assigned to one of three treatment arms:

- High-intensity treatment: the telephone message had a strong partisan tone, explicitly mentioning the Labour party and policies, taking an antagonistic stance toward the main rival party.
- Low-intensity treatment: the telephone message avoided statements about party competition and did not mention the candidate’s affiliation nor the rival party.

- Control: did not receive any form of contact from the campaign.

Finally, within the households assigned to the low- or high-intensity treatment arms, only one household member was randomly selected to receive the telephone message.

Because the telephone message is delivered by landline, this type of experiments is usually subject to rather severe rates of nonresponse, since individuals assigned to treatment are likely to be unavailable, refuse to participate, may have moved or their phone numbers can be outdated or wrong. For these and other reasons, it is common to find compliance rates below 50 percent (see e.g. [Gerber and Green, 2000](#); [John and Brannan, 2008](#)). In the experiment described here, the response rate among individuals assigned to receive the message is about 45 percent. To account for the potential endogeneity of this type of noncompliance, the randomized treatment assignment can be used as an instrument for actual treatment receipt.

For illustration purposes, I will pool the low- and high-intensity treatments into a single combined treatment. To analyze this experiment in the framework set up in previous sections, for each household g , let (Z_{ig}, Z_{jg}) be the randomized treatment assignment for each unit, where $Z_{ig} = 1$ if individual i is randomly assigned to receive the phone call. Let (D_{ig}, D_{jg}) be the treatment indicators, where $D_{ig} = 1$ if individual i actually receives the phone message. Finally, the outcome of interest Y_{ig} , voter turnout, equals 1 if individual i voted in the election.

In this experiment, noncompliance is one-sided, as units assigned to treatment can fail to receive the phone call, but units assigned to control will never receive it. Hence, we can analyze this experiment using the results from Section 5, Proposition 2, Corollaries 1 and 2 and Proposition 3. Since only one member of each treated household was selected to receive the call, we also have that $\mathbb{P}[Z_{ig} = 1, Z_{jg} = 1] = 0$.

The estimation results are shown in Table 2. Given the experimental design, the first stage reduces to estimating $\mathbb{E}[D_{ig}|Z_{ig} = 1, Z_{jg} = 0] = \mathbb{E}[D_{ig}|Z_{ig} = 1]$. The estimated coefficient is 0.451, significantly different from zero at the one percent level and with an F -statistic of 1759.03, which suggests a strong instrument.

The right column shows the estimated direct and indirect ITT and LATE parameters. The results reveal strong evidence of both local average direct and indirect effects. More precisely, the phone message increases voter turnout on compliers with untreated peers by about 7 percentage points, and turnout for untreated individuals with treated compliant peers by about 10 percentage points, both effects significant at the 1 percent level.

The finding that the estimated spillover effect is larger than the direct effect may seem surprising, as one may intuitively expect indirect effects to be weaker than direct ones. This comparison, however, must be done with care, as the estimated effects correspond to different subpopulations. More precisely, the direct effect is estimated for compliers, whereas

the spillover effect is estimated for units with compliant peers, averaging over own compliance types. This is different than comparing the direct and spillover effects on, say, the population of compliers. Note that the indirect LATE is:

$$\begin{aligned}\mathbb{E}[Y_{ig}(0, 1) - Y_{ig}(0, 0)|C_{jg}] &= \mathbb{E}[Y_{ig}(0, 1) - Y_{ig}(0, 0)|C_{ig}, C_{jg}]\mathbb{P}[C_{ig}|C_{jg}] \\ &\quad + \mathbb{E}[Y_{ig}(0, 1) - Y_{ig}(0, 0)|C_{ig}^c, C_{jg}]\mathbb{P}[C_{ig}^c|C_{jg}]\end{aligned}$$

so it combines the effects on compliers and non-compliers, conditional on them having compliant peers.

Table 3 provides some further insights to interpret these findings. The results estimate the difference in average baseline potential outcomes $Y_{ig}(0, 0)$ between compliers and non-compliers (first row) and between units with compliant and non-compliant peers (second row). The differences are about 17 and 19 percentage points, respectively, significant at the 1 percent level. Because the outcome of interest is binary, the fact that compliers start from a higher baseline leaves a smaller margin for the treatment to increase turnout. For this reason, we can expect the noncompliers to have larger average effects than compliers, which could explain the difference between the direct and indirect effects in Table 2.

Regarding the external validity of these local effects, the estimates in Table 3 suggest the presence of marked heterogeneity in average potential outcomes, both across own and peer's type. This casts doubts on the possibility of extrapolating the estimated effects on compliers for never-takers or group compliers. For these reason, we can expect the identified LATEs to be different from the average treatment effects, which are not point identified under imperfect compliance.

Finally, the left panel in Table 2 shows the estimates one would obtain by ignoring the presence of spillovers, that is, running a 2SLS using Z_{ig} as an instrument for D_{ig} without accounting for peer's assignment or treatment status. While also statistically significant, the magnitude of the coefficient is 4 percentage points, almost half of the estimated direct effect and about 40 percent of the indirect effect. These results can be interpreted using Proposition 2. Under this treatment assignment mechanism, it can be seen that:

$$\begin{aligned}\frac{\mathbb{E}[Y_{ig}|Z_{ig} = 1] - \mathbb{E}[Y_{ig}|Z_{ig} = 0]}{\mathbb{E}[D_{ig}|Z_{ig} = 1]} &= \\ &\quad \mathbb{E}[Y_{ig}(1, 0) - Y_{ig}(0, 0)|C_{ig}] - \mathbb{E}[Y_{ig}(0, 1) - Y_{ig}(0, 0)|C_{jg}]\mathbb{P}[Z_{jg} = 1|Z_{ig} = 0],\end{aligned}$$

which shows that the 2SLS estimand ignoring spillovers is a difference between the direct and indirect LATEs, where the indirect LATE is rescaled by the conditional probability of treatment assignment.

Table 2: Empirical Results

	(1)		(2)	
	coef.	p-value	coef.	p-value
ITT				
Z_{ig}	0.018	0.045	0.030	0.0076
Z_{jg}			0.046	0.0000
LATE				
D_{ig}	0.039	0.044	0.068	0.0073
D_{jg}			0.102	0.0000

Notes: rows 1 and 2 show the estimated coefficients from the reduced-form regressions (ITT parameters) of Y_{ig} on Z_{ig} (left panel) and the reduced-form regression of Y_{ig} on Z_{ig} and Z_{jg} (right panel). Rows 3 and 4 show the estimated coefficients from a 2SLS regression (LATEs) of Y_{ig} on D_{ig} using Z_{ig} as an instrument (left panel) and a 2SLS regression of Y_{ig} on D_{ig} and D_{jg} using Z_{ig} and Z_{jg} as instruments (right panel). The first-stage coefficient is 0.451, with an F -statistic of 1759.03. Results allow for clustering at the household level. Number of clusters: $G = 4,930$, total sample size $N = 9,860$.

Table 3: Assessing heterogeneity over types

	coef.	p-value
$\mathbb{E}[Y_{ig}(0,0) C_{ig}] - \mathbb{E}[Y_{ig}(0,0) C_{ig}^c]$	0.170	0.0000
$\mathbb{E}[Y_{ig}(0,0) C_{jg}] - \mathbb{E}[Y_{ig}(0,0) C_{jg}^c]$	0.191	0.0000

Notes: estimated heterogeneity measures from Corollary 2. Results allow for clustering at the household level. Number of clusters: $G = 4,930$, total sample size $N = 9,860$.

8 Further Results and Discussions

8.1 Conditional-on-observables IV

In this section I generalize my results to the case in which quasi-random assignment of (Z_{ig}, Z_{jg}) holds after conditioning on observable characteristics, following [Abadie \(2003\)](#). Let $X_g = (X'_{ig}, X'_{jg})'$ be a vector of observable characteristics for units i and j in group g .

Assumption 5 (Conditional-on-observables IV)

1. *Exclusion restriction:* $Y_{ig}(d, d')$ is not a function of (z, z') ,
2. *Independence:* For all $i, j \neq i$ and g , $((Y_{ig}(d, d'))_{(d, d')}, (D_{ig}(z, z'))_{(z, z')}) \perp (Z_{ig}, Z_{jg}) | X_g$,
3. *Monotonicity:* $\mathbb{P}[D_{ig}(1, 1) \geq D_{ig}(1, 0) \geq D_{ig}(0, 1) \geq D_{ig}(0, 0) | X_g] = 1$,
4. $\mathbb{P}[AT_{ig} | X_g] + \mathbb{P}[NT_{ig} | X_g] < 1$,
5. $\mathbb{P}[D_{ig}(0, z') = 0 | X_g] = 1$ for $z' = 0, 1$.

Let $p_{zz'}(X_g) = \mathbb{P}[Z_{ig} = z, Z_{jg} = z' | X_g]$. Then we have the following result.

Proposition 4 (Identification conditional on observables) Under Assumption 5,

$$\begin{aligned}\mathbb{P}[C_{ig} | X_g] &= \mathbb{E}[D_{ig} | Z_{ig} = 1, Z_{jg} = 0, X_g] \\ \mathbb{P}[GC_{ig} | X_g] &= \mathbb{E}[D_{ig} | Z_{ig} = 1, Z_{jg} = 1, X_g] - \mathbb{E}[D_{ig} | Z_{ig} = 1, Z_{jg} = 0, X_g] \\ \mathbb{P}[NT_{ig} | X_g] &= 1 - \mathbb{P}[GC_{ig} | X_g] - \mathbb{P}[C_{ig} | X_g]\end{aligned}$$

and for any (integrable) function $g(\cdot, \cdot)$,

$$\begin{aligned}\mathbb{E}[g(Y_{ig}(0, 0), X_g)] &= \mathbb{E}\left[g(Y_{ig}, X_g) \frac{(1 - Z_{ig})(1 - Z_{jg})}{p_{00}(X_g)}\right] \\ \mathbb{E}[g(Y_{ig}(1, 0), X_g) | C_{ig}] \mathbb{P}[C_{ig}] &= \mathbb{E}\left[g(Y_{ig}, X_g) D_{ig} \frac{Z_{ig}(1 - Z_{jg})}{p_{10}(X_g)}\right] \\ \mathbb{E}[g(Y_{ig}(0, 1), X_g) | C_{jg}] \mathbb{P}[C_{jg}] &= \mathbb{E}\left[g(Y_{ig}, X_g) D_{jg} \frac{(1 - Z_{ig})Z_{jg}}{p_{01}(X_g)}\right] \\ \mathbb{E}[g(Y_{ig}(0, 0), X_g) | C_{ig}] \mathbb{P}[C_{ig}] &= \mathbb{E}\left[g(Y_{ig}, X_g) \frac{(1 - Z_{ig})(1 - Z_{jg})}{p_{00}(X_g)}\right] - \mathbb{E}\left[g(Y_{ig}, X_g) (1 - D_{ig}) \frac{Z_{ig}(1 - Z_{jg})}{p_{10}(X_g)}\right] \\ \mathbb{E}[g(Y_{ig}(0, 0), X_g) | C_{jg}] \mathbb{P}[C_{jg}] &= \mathbb{E}\left[g(Y_{ig}, X_g) \frac{(1 - Z_{ig})(1 - Z_{jg})}{p_{00}(X_g)}\right] - \mathbb{E}\left[g(Y_{ig}, X_g) (1 - D_{jg}) \frac{(1 - Z_{ig})Z_{jg}}{p_{01}(X_g)}\right],\end{aligned}$$

whenever the required conditional probabilities $p_{zz'}(X_g)$ are positive. Furthermore, these equalities also hold conditional on X_g .

This result shows identification of functions of potential outcomes and covariates for compliers and for units with compliant peers. In particular, note that setting $g(y, x) = y$

recovers the result from Proposition 2, which gives identification of local direct and spillover effects, both unconditionally or conditional on X_g . On the other hand, setting $g(y, x) = x$ shows that it is possible to identify the average characteristics of compliers and units with compliant peers. Hence, even if compliance type is unobservable, it is possible to characterize the distribution of observable characteristics for these subgroups (also a point made in the no-spillovers case by Abadie, 2003; Angrist and Pischke, 2009).

8.2 Arbitrary Group Sizes

Without further assumptions, identification becomes increasingly harder as group size grows. In particular, the larger the group, the larger the set of compliance types, as units may respond in different ways to the different possible combinations of own and peers' treatment assignments, and it is generally not possible to pin down each unit's type.

Section A1.3 in the supplemental appendix shows that the results from Propositions 1 and 2 still hold with arbitrary group sizes, and it is possible to identify the direct effects on compliers and indirect effects on units for which a specific neighbor is a complier. However, a more thorough exploration of what can be identified in general settings is left for future research.

9 Conclusion

This paper proposes a potential outcomes framework to analyze identification and estimation of causal spillover effects using instrumental variables. I provide conditions for identification of the marginal distribution of compliance types and show that intention-to-treat parameters identify linear combinations of direct and spillover effects over different subpopulations, but that are not proper weighted average even after rescaling by the first stage. I then show how to identify different causally interpretable estimands under one-sided noncompliance, and apply the proposed methods to study the effect of social interactions on voting behavior.

References

- Abadie, A. (2003), “Semiparametric instrumental variable estimation of treatment response models,” *Journal of Econometrics*, 113, 231 – 263.
- Abadie, A., and Cattaneo, M. D. (2018), “Econometric Methods for Program Evaluation,” *Annual Review of Economics*, 10, 465–503.
- Angrist, J., and Pischke, J.-S. (2009), *Mostly Harmless Econometrics: An Empiricist’s Companion*, Princeton University Press.
- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996), “Identification of Causal Effects Using Instrumental Variables,” *Journal of the American Statistical Association*, 91, 444–455.
- Angrist, J. D., and Krueger, A. B. (2001), “Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments,” *Journal of Economic Perspectives*, 15, 69–85.
- Athey, S., and Imbens, G. (2017), “The Econometrics of Randomized Experiments,” in *Handbook of Field Experiments*, eds. A. V. Banerjee and E. Duflo, Vol. 1 of *Handbook of Economic Field Experiments*, North-Holland, pp. 73 – 140.
- Babcock, P., Bedard, K., Charness, G., Hartman, J., and Royer, H. (2015), “Letting Down the Team? Social Effects of Team Incentives,” *Journal of the European Economic Association*, 13, 841–870.
- Baird, S., Bohren, A., McIntosh, C., and Özler, B. (2018), “Optimal Design of Experiments in the Presence of Interference,” *The Review of Economics and Statistics*, 100, 844–860.
- Basse, G., and Feller, A. (2018), “Analyzing two-stage experiments in the presence of interference,” *Journal of the American Statistical Association*, 113, 41–55.
- Basse, G. W., Feller, A., and Toulis, P. (2019), “Randomization tests of causal effects under interference,” *Biometrika*, 106, 487–494.
- Blackwell, M. (2017), “Instrumental Variable Methods for Conditional Effects and Causal Interaction in Voter Mobilization Experiments,” *Journal of the American Statistical Association*, 112, 590–599.
- Duflo, E., and Saez, E. (2003), “The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment,” *The Quarterly Journal of Economics*, 118, 815–842.

- Fletcher, J., and Marksteiner, R. (2017), “Causal Spousal Health Spillover Effects and Implications for Program Evaluation,” *American Economic Journal: Economic Policy*, 9, 144–66.
- Foos, F., and de Rooij, E. A. (2017), “All in the Family: Partisan Disagreement and Electoral Mobilization in Intimate Networks—A Spillover Experiment,” *American Journal of Political Science*, 61, 289–304.
- Gerber, A. S., and Green, D. P. (2000), “The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment,” *American Political Science Review*, 94, 653–663.
- Halloran, M. E., and Hudgens, M. G. (2016), “Dependent Happenings: a Recent Methodological Review,” *Current Epidemiology Reports*, 3, 297–305.
- Hudgens, M. G., and Halloran, M. E. (2008), “Toward Causal Inference with Interference,” *Journal of the American Statistical Association*, 103, 832–842.
- Imai, K., Jiang, Z., and Malani, A. (forthcoming), “Causal Inference with Interference and Noncompliance in Two-Stage Randomized Experiments,” *Journal of the American Statistical Association*.
- Imbens, G. W., and Angrist, J. D. (1994), “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 62, 467–475.
- Imbens, G. W., and Rubin, D. B. (1997), “Estimating Outcome Distributions for Compliers in Instrumental Variables Models,” *The Review of Economic Studies*, 64, 555–574.
- Imbens, G. W., and Wooldridge, J. M. (2009), “Recent Developments in the Econometrics of Program Evaluation,” *Journal of Economic Literature*, 47, 5–86.
- John, P., and Brannan, T. (2008), “How Different Are Telephoning and Canvassing? Results from a ‘Get Out the Vote’ Field Experiment in the British 2005 General Election,” *British Journal of Political Science*, 38, 565–574.
- Kang, H., and Imbens, G. (2016), “Peer Encouragement Designs in Causal Inference with Partial Interference and Identification of Local Average Network Effects,” *arXiv:1609.04464*.
- Kang, H., and Keele, L. (2018), “Spillover Effects in Cluster Randomized Trials with Non-compliance,” *arXiv:1808.06418*.
- Moffit, R. (2001), “Policy Interventions, Low-level Equilibria and Social Interactions,” in *Social Dynamics*, eds. S. N. Durlauf and P. Young, MIT Press, pp. 45–82.
- Mogstad, M., Torgovitsky, A., and Walters, C. R. (2019), “The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables,” *NBER Working Paper*.

- Rosenbaum, P. R. (2007), “Interference Between Units in Randomized Experiments,” *Journal of the American Statistical Association*, 102, 191–200.
- Sacerdote, B. (2001), “Peer Effects with Random Assignment: Results for Dartmouth Roommates,” *The Quarterly Journal of Economics*, 116, 681–704.
- Sinclair, B. (2012), *The Social Citizen: Peer Networks and Political Behavior*, University of Chicago Press.
- Sobel, M. E. (2006), “What Do Randomized Studies of Housing Mobility Demonstrate?: Causal Inference in the Face of Interference,” *Journal of the American Statistical Association*, 101, 1398–1407.
- Tchetgen Tchetgen, E. J., and VanderWeele, T. J. (2012), “On causal inference in the presence of interference,” *Statistical Methods in Medical Research*, 21, 55–75.
- Titunik, R. (2019), “Natural Experiments,” in *Advances in Experimental Political Science*, eds. J. Druckman and D. Green, In preparation for Cambridge University Press.
- Vazquez-Bare, G. (2017), “Identification and Estimation of Spillover Effects in Randomized Experiments,” *working paper*.