Causal Spillover Effects Using Instrumental Variables*

Gonzalo Vazquez-Bare†

April 27, 2020

Abstract

I set up a potential-outcomes framework to analyze spillover effects using instrumental variables. I show that intention-to-treat parameters aggregate a large number of direct and spillover effects for different compliance types, and hence they may not have a clear causal interpretation. I provide three alternative conditions to point identify average direct and spillover effects on specific subpopulations, by restricting either (i) the number of spillover effects, (ii) the degree of noncompliance, or (iii) the degree of heterogeneity in average causal parameters. I propose simple estimators that are consistent and asymptotically normal under mild conditions, and illustrate my results using data from an experiment that analyzes the effect of social interactions in the household on voting behavior.

*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 identify spillovers (Moffit, 2001; Duflo and Saez, 2003; Vazquez-Bare, 2017; Baird et al., 2018). However, RCTs are often subject to imperfect compliance, which can render actual treatment receipt 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 in this type of settings (see e.g. 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, and offers three main contributions. First, Section 2 defines direct and spillover causal effects and shows that, when treatment receipt 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: always-takers, social interaction compliers, compliers, group compliers and never-takers. 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. 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 proposes three alternative conditions to identify causal parameters consisting in partially restricting either (i) the channels through which spillovers can materialize, (ii) the degree of noncompliance, or (iii) the degree of heterogeneity in average causal parameters. More precisely, Section 5.1 shows that, when spillovers affect the outcome but not treatment take-up (personalized assignment), it is possible to identify several average effects conditional on own and peers' compliance types. On the other hand, Section 5.2 shows that, under one-sided noncompliance, 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 partially 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. Finally, Section 5.3 proposes an assumption, independence of peers' types (IPT), under which average potential outcomes do not vary with peers' compliance types, which yields identification of all average direct and spillover effects on compliers, social compliers and group compliers.

The remainder of the paper discusses estimation and inference and provides an empirical illustration. Section 6 shows that the parameters of interest can be consistently estimated as nonlinear combinations of sample means, and are asymptotically normal under standard conditions. Section 7 illustrates some of these results using data from a study on social interactions and 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. 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 program, and interpret these estimates in light of my framework. Finally, Section 8 generalizes the results to conditional-on-observables and arbitrary group sizes.

The results in this paper provide a general causal framework to interpret the findings of a growing empirical literature using IV methods to estimate spillover effects in economics and other social sciences. For instance, Duflo and Saez (2003) analyze social interaction effects in retirement plans enrollment decisions by running a two-stage experiment in which university employees in different departments are assigned to pure treatment, control in treated departments, and pure controls. Using a similar design, Crépon et al. (2013) study the labor market displacement effects of a job placement assistance program in France. Fletcher and Marksteiner (2017) study behavioral spillovers between spouses in two experiments aiming at reducing smoking and drinking habits. Ichino and Schündeln (2012) assess whether increased monitoring in the voter registration process can displace electoral irregularities to neighboring areas. Eckles et al. (2016) conduct an experiment encouraging peers to provide feedback in a large social network.

This paper is also related to the networks literature. A strand of this literature has focused on identification of peer effects and social interaction effects in linear models through random group assignment (Graham, 2008), partially overlapping groups (Bramoullé et al., 2009) or variation in group sizes (Davezies et al., 2009; De Giorgi et al., 2010). Leung (forthcoming) and Leung (2020) analyze direct and spillover effects of a randomly assigned treatment in a single-network setting. On the other hand, recent studies have analyzed treatment effects of randomized experiments allowing the treatment to affect the network structure (Kline and Tamer, 2019; Johnsson and Moon, forthcoming; Comola and Prina, forthcoming). In my paper, I will consider a setting with many independent exogenous networks, which allows me to be fully nonparametric about the within-group network structure while letting the treatment be endogenous and arbitrarily correlated with potential outcomes.

Finally, this paper complements the literature focusing on causal inference under interference. Basse and Feller (2018), Basse et al. (2019) and Puelz et al. (2019) propose randomization-based inference for causal effects in experiments with interference. Sobel (2006) studied the performance of IV strategies under a completely randomized experiment with one-sided noncompliance. Kang and Imbens (2016) and Imai et al. (2018) extend the results in Hudgens and Halloran (2008) to account for imperfect compliance by focusing on estimands that average over peers' assignments. Kang and Keele (2018) provide bounds for average effects on compliers in cluster randomized trials.

2 Setup

Consider a random sample of iid groups indexed by $g = 1, \ldots, G$, each with $n_g + 1$ identically-distributed units, so that each unit in group g has n_g peers or neighbors. Spillovers are assumed to occur between units in the same group, but not between units in different groups. A unit's potential outcomes, defined in the next paragraph, can depend on her own and her group peers' treatment status. I will start by analyzing a setup in which units are grouped in pairs, so that each unit has one peer, $n_g = 1$. This setup helps convey the main ideas and difficulties of the problem under study using minimal notation, and 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). Section 8 generalizes the results to the case with arbitrary group sizes.

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 cases in which the instruments come from a natural experiment (see e.g. Angrist and Krueger, 2001; Titiunik, 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, $((Y_{ig}(d, d'))_{(d,d')}, (D_{ig}(z, z'))_{(z,z')}) \perp\!\!\!\perp (Z_{ig}, Z_{jg})$.

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 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, $\xi_{ig} \in \{\text{AT,SC,C,GC,NT}\}$. 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) $\mathbb{P}[AT_{ig}] + \mathbb{P}[NT_{ig}] < 1$.

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)

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

First-stage analysis refers to the relationship between the instruments (Z_{ig}, Z_{jg}) and the realized treatment status (D_{ig}, D_{jg}). Assumption 3 requires the first stage to show a non-trivial relationship between D_{ig} and (Z_{ig}, Z_{jg}), which can be directly tested since both treatment status and the instruments are observed. 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 a large number of 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:

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

Under Assumption 1, we have the following decomposition:

$$\begin{aligned}\mathbb{E}[Y_{ig}|Z_{ig}=z,Z_{jg}=z'] = \mathbb{E}&[Y_{ig}(0,0)]\\ &+\mathbb{E}[(Y_{ig}(1,0)-Y_{ig}(0,0))D_{ig}(z,z')(1-D_{jg}(z',z))]\\ &+\mathbb{E}[(Y_{ig}(0,1)-Y_{ig}(0,0))(1-D_{ig}(z,z'))D_{jg}(z',z)]\\ &+\mathbb{E}[(Y_{ig}(1,1)-Y_{ig}(0,0))D_{ig}(z,z')D_{jg}(z',z)].\end{aligned}$$

Based on this fact, 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,*

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

To understand the above result, consider the effect of switching Z_{ig} from 0 to 1, while leaving Z_{jg} fixed at zero, for the different combination of compliance types. 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 $D_{jg}=0$ regardless of the value of Z_{ig}. 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. An analogous result, omitted to conserve space, can be shown for the indirect ITT parameter $\mathbb{E}[Y_{ig}|Z_{ig} = 0, Z_{jg} = 1] - \mathbb{E}[Y_{ig}|Z_{ig} = 0, Z_{jg} = 0]$.

Two further issues may complicate the interpretation of this estimand. On the one hand, the ITT parameter is not a weighted average of these effects. 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 $\mathbb{P}[C_{ig}] + \mathbb{P}[SC_{ig}] + \mathbb{P}[S_{ig}, GC_{jg}] + \mathbb{P}[SC_{ig}, NT_{jg}] + \mathbb{P}[SC_{ig}, AT_{jg}]$, 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.

5 Identification

The previous section shows that the close link between the ITT and the LATE (namely, the fact that the LATE is the ITT rescaled by the first stage) breaks down in the presence of spillovers. The large number of potential outcomes and combinations of compliance types suggests that, without further restrictions, it is not possible to isolate the different average effects that are entangled in the ITT. In this section, I propose three alternative approaches to identify several causal parameters that are policy-relevant and easy to interpret.

The first approach consists in reducing the dimensionality of the parameter set by ruling out some types of spillover effects. Section 5.1 shows that, when spillover effects occur on the outcome but not on treatment status (personalized assignment), it is possible to identify several causal effects conditional on own and peer's treatment status. In addition, the supplemental appendix discusses the case in which spillovers occur only on treatment status.

The second approach allows for spillovers on both the outcome and treatment status, but restricts the amount of noncompliance. Section 5.2 shows that, when noncompliance is one-sided, it is possible to identify the average direct on compliers and the average spillover effects on units with compliant peers.

Finally, the third approach allows for spillovers on both channels and two-sided noncompliance, but restricts the heterogeneity of treatment effects. Specifically, Section 5.3 shows identification of several direct and average effects conditional on own type, under the assumption that these effects do not vary with peer's compliance types (independence of peers' types).

In what follow, 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 $\mathbb{1}(Y_{ig} \leq y)$.

5.1 Personalized Assignment

One alternative to identify treatment effects in this context is to limit the possible number of spillovers that can occur. In this section, I rule out the possibility of spillovers in treatment take-up, so that spillovers can only occur on outcomes.

Assumption 4 (Personalized assignment - PA) *For all i and g, $D_{ig}(z, z') = D_{ig}(z)$.*

This assumption reduces the number of potential treatment statuses to two, and is also considered by Kang and Imbens (2016) and Kang and Keele (2018). Under this assumption and monotonicity, there are three compliance types: always-taker, complier or never-taker. This restriction may hold, for example, in cases in which units cannot observe their peers' instrument assignments. On the other hand, as explained in Section 3, this assumption is testable. A simple way to test it is to estimate the parameters in the following regression:

$$D_{ig} = \alpha_0 + \alpha_1 Z_{ig} + \alpha_2 Z_{jg}(1 - Z_{ig}) + \alpha_3 Z_{jg} Z_{ig} + \eta_{ig}$$

where $\eta_{ig} = D_{ig} - \mathbb{E}[D_{ig}|Z_{ig}, Z_{jg}]$ and test the null hypothesis that $\alpha_2 = \alpha_3 = 0$. See

Vazquez-Bare (2017) for further details and discussion.

The following proposition discusses identification of local average potential outcomes under personalized assignment. In what follows, the notation $\xi_{ig} = \mathrm{AT} \cdot d + \mathrm{NT} \cdot (1-d)$ means that $\xi_{ig} = \mathrm{AT}$ when $d = 1$ and $\xi_{ig} = \mathrm{NT}$ if $d = 0$, and similarly for ξ_{jg}.

Proposition 2 (Local average potential outcomes under PA) *Suppose Assumptions 1-3 and 4 hold, and consider a vector of treatment statuses (d, d') such that $\mathbb{P}[D_{ig} = d, D_{jg} = d'] > 0$.*

(1) If $\mathbb{P}[Z_{ig} = 1-d, Z_{jg} = 1-d' | D_{ig} = d, D_{jg} = d'] > 0$, then

$$\mathbb{P}[\xi_{ig} = \mathrm{AT} \cdot d + \mathrm{NT} \cdot (1-d), \xi_{jg} = \mathrm{AT} \cdot d' + \mathrm{NT} \cdot (1-d')]$$

and

$$\mathbb{E}[Y_{ig}(d, d') | \xi_{ig} = \mathrm{AT} \cdot d + \mathrm{NT} \cdot (1-d), \xi_{jg} = \mathrm{AT} \cdot d' + \mathrm{NT} \cdot (1-d')]$$

are identified.

(2) If, in addition to (1), $\mathbb{P}[Z_{ig} = d, Z_{jg} = 1-d' | D_{ig} = d, D_{jg} = d'] > 0$, then

$$\mathbb{P}[\xi_{ig} = C, \xi_{jg} = \mathrm{AT} \cdot d' + \mathrm{NT} \cdot (1-d')]$$

and

$$\mathbb{E}[Y_{ig}(d, d') | \xi_{ig} = \mathrm{C}, \xi_{jg} = \mathrm{AT} \cdot d' + \mathrm{NT} \cdot (1-d')]$$

are identified.

(3) If, in addition to (1), $\mathbb{P}[Z_{ig} = 1-d, Z_{jg} = d' | D_{ig} = d, D_{jg} = d'] > 0$, then

$$\mathbb{P}[\xi_{ig} = \mathrm{AT} \cdot d + \mathrm{NT} \cdot (1-d), \xi_{jg} = \mathrm{C}]$$

and

$$\mathbb{E}[Y_{ig}(d, d') | \xi_{ig} = \mathrm{AT} \cdot d + \mathrm{NT} \cdot (1-d), \xi_{jg} = \mathrm{C}]$$

are identified.

(4) If, in addition to (1), (2) and (3), $\mathbb{P}[Z_{ig} = d, Z_{jg} = d' | D_{ig} = d, D_{jg} = d'] > 0$, then

$$\mathbb{P}[\xi_{ig} = \mathrm{C}, \xi_{jg} = \mathrm{C}]$$

and

$$\mathbb{E}[Y_{ig}(d, d') | \xi_{ig} = \mathrm{C}, \xi_{jg} = \mathrm{C}]$$

are identified.

Proposition 2 shows which average potential outcomes can be identified, depending on which combinations of the pairs (d, d') and (z, z') can be observed. In particular, this result

Table 2: Identification under PA

	Z_{ig}	Z_{jg}	D_{ig}	D_{jg}	Unit i's type	Unit j's type	Obs. outcome
1	0	0	1	1	AT	AT	$Y_{ig}(1,1)$
2	0	1	1	1	AT	AT ∪ C	$Y_{ig}(1,1)$
3	1	0	1	1	AT ∪ C	AT	$Y_{ig}(1,1)$
4	1	1	1	1	AT ∪ C	AT ∪ C	$Y_{ig}(1,1)$
5	1	0	0	1	NT	AT	$Y_{ig}(0,1)$
6	1	1	0	1	NT	AT ∪ C	$Y_{ig}(0,1)$
7	0	0	0	1	NT ∪ C	AT	$Y_{ig}(0,1)$
8	0	1	0	1	NT ∪ C	AT ∪ C	$Y_{ig}(0,1)$
9	0	1	1	0	AT	NT	$Y_{ig}(1,0)$
10	0	0	1	0	AT	NT ∪ C	$Y_{ig}(1,0)$
11	1	1	1	0	AT ∪ C	NT	$Y_{ig}(1,0)$
12	1	0	1	0	AT ∪ C	NT ∪ C	$Y_{ig}(1,0)$
13	1	1	0	0	NT	NT	$Y_{ig}(0,0)$
14	1	0	0	0	NT	NT ∪ C	$Y_{ig}(0,0)$
15	0	1	0	0	NT ∪ C	NT	$Y_{ig}(0,0)$
16	0	0	0	0	NT ∪ C	NT ∪ C	$Y_{ig}(0,0)$

shows that if all combinations occur with non-zero probability, then all the average potential outcomes (and therefore all the average direct and spillover effects) are identified for groups where both units are compliers. Table 3 provides a visualization of this identification result.

From Proposition 2, we see that for example:

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

which is the Wald estimand obtained from instrumenting $D_{jg}(1-D_{ig})$ with Z_{jg} using $Y_{ig}(1-D_{ig})$ as an outcome on the subsample of units with $Z_{ig}=1$. On the other hand,

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

which is a ratio of double differences.

Notice that not all the average potential outcomes can be identified. For example, it is not possible to identify $\mathbb{E}[Y_{ig}(1,1)|C_{ig}, NT_{jg}]$, because if j is a never-taker, the treatment status $(1,1)$ will never be observed. This fact implies that average potential outcomes for compliers are not identified in general, since, for example,

$$\begin{aligned}\mathbb{E}[Y_{ig}(1,1)|C_{ig}] = \mathbb{E}[Y_{ig}(1,1)|C_{ig}, C_{jg}]\mathbb{P}[C_{jg}|C_{ig}] \\ + \mathbb{E}[Y_{ig}(1,1)|C_{ig}, AT_{jg}]\mathbb{P}[AT_{jg}|C_{ig}] \\ + \mathbb{E}[Y_{ig}(1,1)|C_{ig}, NT_{jg}]\mathbb{P}[NT_{jg}|C_{ig}]\end{aligned}$$

and the third term is not identified without further assumptions. More precisely, out of the 36 possible average potential outcomes $\mathbb{E}[Y_{ig}(d,d')|\xi_{ig}, \xi_{jg}]$ for each unit, we can identify 16 (which is the number of rows in Table 3), since for each d we can only recover the effect conditional on two of the three possible types.

5.2 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 degree 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. The most common case of one-sided noncompliance is one in which units can refuse to receive treatment when they are assigned to it, but cannot receive the treatment when they are assigned to the untreated condition. I formalize this assumption as follows.

Assumption 5 (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. Notice that this assumption can be verified in practice, by checking that there are no units with $Z_{ig} = 0$ and $D_{ig} = 1$. By reducing the set of compliance types in the population, Assumption 5 allows for point identification of several local average potential outcomes, as shown in the following proposition.

Proposition 3 (Local average potential outcomes under OSN) *Under Assumptions*

1-3 and 5, 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 3 allows identifying the following treatment effects.

Corollary 1 (Local average direct and spillover effects under OSN) *Under Assumptions 1-3 and 5, 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 3 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^c_{ig} represents the event in which unit i is not a complier, that is, $C^c_{ig} = GC_{ig} \cup NT_{ig}$.

Corollary 2 (Assessing heterogeneity over compliance types) *Under 1-3 and 5, if*

$0 < \mathbb{P}[C_{ig}] < 1$,

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

and

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

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 4 (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 5, if $0 < \mathbb{P}[C_{ig}] < 1$,

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

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.

5.3 Independence of Peers' Types

Another way to achieve point identification of causal parameters is to restrict the amount of heterogeneity over types in the treatment effects. Specifically, in this section I impose the assumption that local average potential outcomes do not vary over peer's type, so that, for example, the average $Y_{ig}(d,d')$ for unit $i's$ is the same when unit j is an always-taker, a social complier, a group complier and so on, conditional on unit i's type. This requirement is stated as follows.

Assumption 6 (Independence of Peers' Types - IPT) *For all $z_l = 0, 1$, $l = 1, 2, 3, 4$,*

$$((Y_{ig}(d,d'))_{(d,d')} \perp\!\!\!\perp (D_{jg}(z_1,z_2))_{(z_1,z_2)} \big| (D_{ig}(z_3,z_4))_{(z_3,z_4)}, \quad i \neq j.$$

This assumption states that the potential outcomes are independent of peers' type, conditional on own type.

It is important to note that Assumption 6 is not a statement on spillover effects, but on treatment effect heterogeneity, as it only restricts the way in which average effects can vary with the different compliance types in the population. The following remarks discuss two cases in which this assumption holds.

Remark 1 (Independent sampling) *If units within a group are sampled independently from the same population, it follows that*

$$((Y_{ig}(d_1,d_2))_{(d_1,d_2)}, (D_{ig}(z_3,z_4))_{(z_3,z_4)}) \perp\!\!\!\perp ((Y_{jg}(d_3,d_4))_{(d_3,d_4)}, D_{jg}(z_1,z_2))_{(z_1,z_2)}),$$

which implies that IPT holds.

Remark 2 (Perfectly correlated types) *Suppose that types are perfectly correlated, so that $\mathbb{P}[\xi_{ig} = \xi_{jg}] = 1$. Then, conditional on own type, peer's type is non-random, and thus IPT holds.*

The following proposition shows identification of local average potential outcomes under IPT. Recall that ξ_{ig} indicates unit i's compliance type, $\xi_{ig} \in \{\text{AT,SC,C,GC,CT}\}$. In what follows, $\xi_{ig} = \text{AT} \cdot d + \text{NT} \cdot (1-d)$ means that $\xi_{ig} = \text{AT}$ when $d = 1$ and $\xi_{ig} = \text{NT}$ when $d = 0$.

Proposition 5 (Local average potential outcomes under IPT) *Suppose Assumptions 1-3 and 6 hold, and consider a vector of treatment statuses $(D_{ig}, D_{jg}) = (d, d')$ such that $\mathbb{P}[D_{ig} = d, D_{jg} = d'] > 0$.*

(1) If $\mathbb{P}[Z_{ig} = 1-d, Z_{jg} = 1-d | D_{ig} = d, D_{jg} = d'] > 0$, then

$$\mathbb{E}[Y_{ig}(d,d') | \xi_{ig} = \text{AT} \cdot d + \text{NT} \cdot (1-d)]$$

is identified.

(2) If, in addition to (1), $\mathbb{P}[Z_{ig} = 1 - d, Z_{jg} = d | D_{ig} = d, D_{jg} = d'] > 0$*, then*

$$\mathbb{E}[Y_{ig}(d, d') | \xi_{ig} = \mathrm{SC} \cdot d + \mathrm{GC} \cdot (1 - d)]$$

is identified.

(3) If, in addition to (1) and (2), $\mathbb{P}[Z_{ig} = d, Z_{jg} = 1 - d | D_{ig} = d, D_{jg} = d'] > 0$*, then*

$$\mathbb{E}[Y_{ig}(d, d') | \xi_{ig} = \mathrm{C}]$$

is identified.

(4) If, in addition to (1), (2) and (3), $\mathbb{P}[Z_{ig} = d, Z_{jg} = d | D_{ig} = d, D_{jg} = d'] > 0$*, then*

$$\mathbb{E}[Y_{ig}(d, d') | \xi_{ig} = \mathrm{GC} \cdot d + \mathrm{SC} \cdot (1 - d)]$$

is identified.

To illustrate Proposition 5, let $(d, d)' = (1, 1)$. If $\mathbb{P}[Z_{ig} = 0, Z_{jg} = 0 | D_{ig} = 1, D_{jg} = 1] > 0$, so that, among the pairs in which both units are treated, we can observe pairs in which both were assigned to control, we would know that unit i is an always-taker since $D_{ig}(0, 0) = 1$. From those groups we can therefore identify $\mathbb{E}[Y_{ig}(1, 1) | AT_{ig}]$. Next, suppose that $\mathbb{P}[Z_{ig} = 0, Z_{jg} = 1 | D_{ig} = 1, D_{jg} = 1] > 0$. In this case we know that $D_{ig}(0, 1) = 1$ from which we deduce that unit i can be an always-taker or a social complier. From the observed outcomes for these units we can recover:

$$\mathbb{E}[Y_{ig}(1,1)|AT_{ig}]\frac{\mathbb{P}[AT_{ig}]}{\mathbb{P}[AT_{ig}] + \mathbb{P}[SC_{ig}]} + \mathbb{E}[Y_{ig}(1,1)|SC_{ig}]\frac{\mathbb{P}[SC_{ig}]}{\mathbb{P}[AT_{ig}] + \mathbb{P}[SC_{ig}]},$$

but since the probabilities are identified under Proposition 1 and $\mathbb{E}[Y_{ig}(1, 1) | AT_{ig}]$ was identified in the previous step, we can recover $\mathbb{E}[Y_{ig}(1, 1) | SC_{ig}]$. For the third step, if $\mathbb{P}[Z_{ig} = 1, Z_{jg} = 0 | D_{ig} = 1, D_{jg} = 1] > 0$, we have that $D_{ig}(1, 0) = 1$ so unit i can be an always-taker, a social complier or a complier. Again using Proposition 1 and the two previous steps we can isolate $\mathbb{E}[Y_{ig}(1, 1) | C_{ig}]$. Finally, if $\mathbb{P}[Z_{ig} = 1, Z_{jg} = 1 | D_{ig} = 1, D_{jg} = 1] > 0$ we can use the same reasoning the identify $\mathbb{E}[Y_{ig}(1, 1) | GC_{ig}]$. Notice that $\mathbb{E}[Y_{ig}(1, 1) | NT_{ig}]$ is not identified because if unit i is a never-taker she can never be observed to receive treatment $d = 1$.

Table 3 illustrates that IPT provides point identification for a wide array of local average effects. In particular, the four average potential outcomes are identified for compliers, social compliers and group compliers, which allows us to identify all the direct and spillover effects for these subpopulations. On the other hand, Proposition 5 also identifies the spillover effect on treated always-takers, $\mathbb{E}[Y_{ig}(1, 1) - Y_{ig}(1, 0) | AT_{ig}]$, and the spillover effect on untreated never-takers, $\mathbb{E}[Y_{ig}(0, 1) - Y_{ig}(0, 0) | NT_{ig}]$. Hence, in this setting, it is possible to assess how the treatment affects subpopulations whose behavior is unaffected by the instrument.

Table 3: Identification under IPT

ξ_{ig}	$\mathbb{E}[Y_{ig}(1,1)\vert\xi_{ig}]$	$\mathbb{E}[Y_{ig}(1,0)\vert\xi_{ig}]$	$\mathbb{E}[Y_{ig}(0,1)\vert\xi_{ig}]$	$\mathbb{E}[Y_{ig}(0,0)\vert\xi_{ig}]$
AT	✓	✓	-	-
SC	✓	✓	✓	✓
C	✓	✓	✓	✓
GC	✓	✓	✓	✓
NT	-	-	✓	✓

Notes: a check mark (✓) indicates that the corresponding average potential outcome (column) given own type (row) is identified by Proposition 5. A minus sign (-) indicates that the corresponding average potential outcome given own type is not identified by Proposition 5.

6 Estimation and Inference

Based on the results in the previous sections, the parameters of interest can be recovered by estimating expectations using sample means. More precisely, let

$$\mathbb{1}^{\mathbf{z}}_{ig} = \begin{bmatrix} \mathbb{1}(Z_{ig}=0, Z_{jg}=0) \\ \mathbb{1}(Z_{ig}=1, Z_{jg}=0) \\ \mathbb{1}(Z_{ig}=0, Z_{jg}=1) \\ \mathbb{1}(Z_{ig}=1, Z_{jg}=1) \end{bmatrix}$$

and let $\mathbf{H}(\cdot)$ be a vector-valued function whose exact shape depends on the parameters to be estimated, as illustrated below. Then the goal is to estimate:

$$\boldsymbol{\mu} = \mathbb{E}\begin{bmatrix} \mathbb{1}^{\mathbf{z}}_{ig} \\ \mathbf{H}(Y_{ig}, D_{ig}, D_{jg}) \otimes \mathbb{1}^{\mathbf{z}}_{ig} \end{bmatrix}$$

where the first four elements correspond to the assignment probabilities $\mathbb{P}[Z_{ig}=z, Z_{jg}=z']$ and the remaining elements corresponds to estimands of the form $\mathbb{E}[Y_{ig}\mathbb{1}(D_{ig}=d, D_{jg}=d')\mathbb{1}(Z_{ig}=z, Z_{jg}=z')]$. The most general choice of $\mathbf{H}$ in this setup is the following:

$$\mathbf{H}(Y_{ig}, D_{ig}, D_{jg}) = \begin{bmatrix} \mathbb{1}(D_{ig}=0, D_{jg}=0) \\ \vdots \\ \mathbb{1}(D_{ig}=1, D_{jg}=1) \\ Y_{ig}\mathbb{1}(D_{ig}=0, D_{jg}=0) \\ \vdots \\ Y_{ig}\mathbb{1}(D_{ig}=1, D_{jg}=1) \end{bmatrix}$$

which is a vector of dimension equal to eight that can be used to estimate all the first-stage estimands $\mathbb{E}[\mathbb{1}(D_{ig}=d, D_{jg}=d')|Z_{ig}=z, Z_{jg}=z']$ and average outcomes $\mathbb{E}[Y_{ig}\mathbb{1}(D_{ig}=d, D_{jg}=d')|Z_{ig}=z, Z_{jg}=z']$. In this general case, the total number of equations to be estimated is 36: four probabilities $\mathbb{P}[Z_{ig} = z, Z_{jg} = z']$ plus the four indicators $\mathbb{1}(Z_{ig} = z, Z_{jg} = z')$ times each of the eight elements in $\mathbf{H}(\cdot)$. The dimension of $\mathbf{H}(\cdot)$ can be reduced, for example, by focusing on ITT parameters $\mathbb{E}[Y_{ig}|Z_{ig} = z, Z_{jg} = z']$, which corresponds to:

$$\mathbf{H}(Y_{ig}, D_{ig}, D_{jg}) = Y_{ig},$$

or by imposing the assumptions described in previous sections. For instance, under one-sided noncompliance, the parameters in Corollaries 1 and 2 can be estimated by defining:

$$\mathbb{1}^{\mathbf{z}}_{ig} = \begin{bmatrix} \mathbb{1}(Z_{ig} = 0, Z_{jg} = 0) \\ \mathbb{1}(Z_{ig} = 1, Z_{jg} = 0) \\ \mathbb{1}(Z_{ig} = 0, Z_{jg} = 1) \end{bmatrix}$$

and

$$\mathbf{H}(Y_{ig}, D_{ig}, D_{jg}) = \begin{bmatrix} D_{ig} \\ Y_{ig} \\ Y_{ig}(1 - D_{ig}) \\ Y_{ig}(1 - D_{jg}) \end{bmatrix}.$$

Regardless of the choice of $\mathbb{1}^{\mathbf{z}}_{ig}$ and $\mathbf{H}(\cdot)$, the dimension of the vector of parameters to be estimated is fixed (and at most 36). Consider the following sample mean estimator:

$$\hat{\boldsymbol{\mu}} = \frac{1}{G}\sum_{g=1}^{G} W_g$$

where

$$W_g = \begin{bmatrix} (\mathbb{1}^{\mathbf{z}}_{1g} + \mathbb{1}^{\mathbf{z}}_{2g})/2 \\ (\mathbf{H}(Y_{1g}, D_{1g}, D_{2g}) \otimes \mathbb{1}^{\mathbf{z}}_{1g} + \mathbf{H}(Y_{2g}, D_{2g}, D_{1g}) \otimes \mathbb{1}^{\mathbf{z}}_{2g})/2 \end{bmatrix}.$$

I will assume the following.

Assumption 7 (Sampling and moments) *Let $V_{ig} = (Y_{ig}, D_{ig}, Z_{ig})'$ for $i = 1, 2$, and $V_g = (V_{1g}', V_{2g}')'$.*

1. *$(V_g)_{g=1}^{G}$ are independent and identically distributed.*
2. *For each g, V_{1g} and V_{2g} are identically distributed but not necessarily independent.*
3. *$\mathbb{E}[Y_{ig}^4] < \infty$.*

Assumption 7 requires the groups to be independent and identically distributed, but allows units within groups to be arbitrarily correlated. Hence, this assumption permits cases in which outcomes are correlated between units, and also treatment assignment mechanisms in which assignments are dependent such as fixed margins.

It is straightforward to see that under assumption 7, $\hat{\boldsymbol{\mu}}$ is unbiased and consistent for $\boldsymbol{\mu}$, and it converges in distribution to a normal random variable after centering and rescaling as $G \to \infty$:

$$\sqrt{G}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma} = \mathbb{E}[(W_g - \boldsymbol{\mu})(W_g - \boldsymbol{\mu})']$, and where the limiting variance can be consistently estimated by:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{G} \sum_g (W_g - \hat{\boldsymbol{\mu}})(W_g - \hat{\boldsymbol{\mu}})'.$$

Finally, once $\hat{\boldsymbol{\mu}}$ has been estimated, the treatment effects of interest can be estimated as (possibly nonlinear) transformations of $\hat{\boldsymbol{\mu}}$, and their variance estimated using the delta method.

7 Empirical Illustration: Social Interactions and Voter Turnout

In this section I illustrate some of 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. 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 5.2, Proposition 3, Corollaries 1 and 2 and Proposition 4. 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 4. 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 5 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 4.

Regarding the external validity of these local effects, the estimates in Table 5 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 4 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 3. Under this treatment assignment mechanism, it can be seen that:

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

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 4: 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 5: Assessing heterogeneity over types

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

8.1 Conditional-on-observables

I this section I extend 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. For brevity, I will only consider the case of one-sided noncompliance, as in Section 5.2.

Assumption 8 (Conditional-on-observables)

1. *Exclusion restriction:* $Y_{ig}(d, d')$ *is not a function of* (z, z')*,*
2. *Independence: For all* i*,* $j \neq i$ *and* g*,* $((Y_{ig}(d, d'))_{(d,d')}, (D_{ig}(z, z'))_{(z,z')}) \perp\!\!\!\perp (Z_{ig}, Z_{jg})|X_g$*,*
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 6 (Identification conditional on observables) *Under Assumption 8,*

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

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

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

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 3, 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 Multiple units per group

In this section I generalize the results to the case where each group g has $n_g + 1$ identically-distributed units, so that each unit in group g has n_g neighbors or peers. The vector of treatment statuses in each group is given by $\mathbf{D}_g = (D_{1g}, \ldots, D_{n_g+1,g})$. For each unit i, $D_{j,ig}$ is the treatment indicator corresponding to unit i's j-th neighbor, collected in the vector $\mathbf{D}_{(i)g} = (D_{1,ig}, D_{2,ig}, \ldots, D_{n_g,ig})$. This vector takes values $\mathbf{d}_g = (d_1, d_2, \ldots, d_{n_g}) \in \mathcal{D}_g \subseteq \{0,1\}^{n_g}$. For a given realization of the treatment status $(d, \mathbf{d}_g)$, the potential outcome for unit i in group g is $Y_{ig}(d, \mathbf{d}_g)$ with observed outcome $Y_{ig} = Y_{ig}(D_{ig}, \mathbf{D}_{(i)g})$. In what follows, $\mathbf{0}_g$ and $\mathbf{1}_g$ will denote n_g-dimensional vectors of zeros and ones, respectively. The observed outcome can be written as:

$$Y_{ig} = \sum_{d\in\{0,1\}} \sum_{\mathbf{d}_g \in \mathcal{D}_g} Y_{ig}(d, \mathbf{d}_g) \mathbb{1}(D_{ig} = d) \mathbb{1}(\mathbf{D}_{(i)g} = \mathbf{d}_g).$$

Let $\mathbf{Z}_{(i)g}$ be the vector of unit i's peers' instruments, taking values $\mathbf{z}_g \in \{0,1\}^{n_g}$. For simplicity, I will assume that potential statuses and outcomes satisfy an exchangeability condition under which the identities of the treated peers do not matter, and thus the variables depend on the vectors $\mathbf{z}_g$ and $\mathbf{d}_g$, respectively, only through the sum of its elements. See Vazquez-Bare (2017) for a detailed discussion on this assumption. Under this condition, we have that $D_{ig}(z, \mathbf{z}_g) = D_{ig}(z, w_g)$ where $w_g = \mathbf{1}_g' \mathbf{z}_g$ and $Y_{ig}(d, \mathbf{d}_g) = Y_{ig}(d, s_g)$ where $s_g = \mathbf{1}_g' \mathbf{d}_g$.

The monotonicity assumption can be adapted to the general case as:

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

for all $w_g = 0, 1, \ldots, n_g$. Under this assumption, we can define five compliance classes. First, always-takers, AT, are units with $D_{ig}(0,0) = 1$ which implies $D_{ig}(z, w_g) = 1$ for all (z, w_g). Next, w^*-social compliers, SC(w^*), are units for whom $D_{ig}(1, w_g) = 1$ for all w_g, and for which there exists a $0 < w^* < n_g$ such that $D_{ig}(0, w_g) = 1$ for all $w_g \geq w^*$. Thus, w^*-social compliers start receiving treatment as soon as w^* of their peers are assigned to treatment. Compliers, C, are units with $D_{ig}(1, w_g) = 1$ and $D_{ig}(0, w_g) = 0$ for all w_g. Next, w^*-group compliers, GC(w^*) have $D_{ig}(0, w_g) = 0$ for all w_g and there exists a $0 < w^* < n_g$ such that

$D_{ig}(1, w_g) = 1$ for all $w_g \geq w^*$. That is, w^*-group compliers need to be assigned to treatment and have at least w^* peers assigned to treatment to actually receive the treatment. Finally, never-takers, NT, are units with $D_{ig}(z, w_g) = 0$ for all (z, w_g).

Let ξ_{ig} be a random variable indicating unit i's compliance type, with

$$\xi_{ig} \in \Xi = \{\mathrm{NT}, \mathrm{GC}(w^*), \mathrm{C}, \mathrm{SC}(w^*), \mathrm{AT} | w^* = 1, \ldots, n_g - 1\},$$

and $\boldsymbol{\xi}_{(i)g}$ the vector collecting ξ_{jg} for $j \neq i$. As before, let the event $AT_{ig} = \{\xi_{ig} = \mathrm{AT}\}$, $C_{ig} = \{\xi_{ig} = \mathrm{C}\}$ and similarly for the other compliance types. I will also assume that peers' types are exchangeable, which means that average potential outcomes depend only on how many of their peers belong to each compliance class, regardless of their identities.

The following assumption collects the required conditions for the upcoming results.

Assumption 9 (Identification conditions for general n_g)

1. *Existence of instruments:*
   (a) $Y_{ig}(d, \mathbf{d}_g)$ *is not a function of* $(z, \mathbf{z}_g)$.
   (b) *For all* i, $j \neq i$ *and* g, $((Y_{ig}(d, \mathbf{d}_g))_{(d, \mathbf{d}_g)}, (D_{ig}(z, \mathbf{z}_g))_{z, \mathbf{z}_g}) \perp\!\!\!\perp (Z_{ig}, \mathbf{Z}_{(i)g})$.
2. *Exchangeability:*
   (a) $D_{ig}(z, \mathbf{z}_g) = D_{ig}(z, w_g)$ *where* $w_g = \mathbf{1}_g' \mathbf{z}_g$
   (b) $Y_{ig}(d, \mathbf{d}_g) = Y_{ig}(d, s_g)$ *where* $s_g = \mathbf{1}_g' \mathbf{d}_g$.
   (c) $\mathbb{E}[Y_{ig}(d, s_g) | \xi_{ig}, \boldsymbol{\xi}_{(i)g}] = \mathbb{E}[Y_{ig}(d, s_g) | \xi_{ig}, \{N_{ig}^{\xi}\}_{\xi \in \Xi}]$ *where* $N_{ig}^{\xi} = \sum_{j \neq i} \mathbb{1}(\xi_{jg} = \xi)$.
3. *Monotonicity: for all* $w_g = 0, 1, \ldots, n_g$, $D_{ig}(1, w_g) \geq D_{ig}(1, 0) \geq D_{ig}(0, n_g) \geq D_{ig}(0, w_g)$.
4. *Relevance:* $\mathbb{P}[AT_{ig}] + \mathbb{P}[NT_{ig}] < 1$.

Let $W_{ig} = \sum_{j \neq i} Z_{jg}$ be the observed number of unit i' peers assigned to treatment. The following result discusses identification of the distribution of compliance types.

Proposition 7 *Under Assumption 9,*

$$\begin{aligned}
\mathbb{P}[AT_{ig}] &= \mathbb{E}[D_{ig} | Z_{ig} = 0, W_{ig} = 0] \\
\mathbb{P}[SC_{ig}(w^*)] &= \mathbb{E}[D_{ig} | Z_{ig} = 0, W_{ig} = w^*] - \mathbb{E}[D_{ig} | Z_{ig} = 0, W_{ig} = w^* - 1], \quad 1 < w^* < n_g \\
\mathbb{P}[C_{ig}] &= \mathbb{E}[D_{ig} | Z_{ig} = 1, W_{ig} = 0] - \mathbb{E}[D_{ig} | Z_{ig} = 0, W_{ig} = n_g] \\
\mathbb{P}[GC_{ig}(w^*)] &= \mathbb{E}[D_{ig} | Z_{ig} = 1, W_{ig} = w^*] - \mathbb{E}[D_{ig} | Z_{ig} = 1, W_{ig} = w^* - 1], \quad 1 < w^* < n_g \\
\mathbb{P}[NT_{ig}] &= \mathbb{E}[1 - D_{ig} | Z_{ig} = 1, W_{ig} = n_g].
\end{aligned}$$

Next, I generalize the results for ITT parameters as follows. Given an assignment vector

$(z, \mathbf{z}_g)$, let $S_{ig}(z, \mathbf{z}_g)$ be the sum of unit i's peers potential treatment statuses:

$$S_{ig}(z, \mathbf{z}_g) = \sum_{j \neq i} D_{jg}(z_j, w_g + z - z_j)$$

where z_j is the j-th element in $\mathbf{z}_g$. Let $\mathbf{C}_{(i)g} = (\mathbb{1}(\xi_{jg}) = \mathrm{C}))_{j \neq i}$ be a vector with entries equal to 1 for each one of i's peers who is a complier and 0 otherwise, with analogous definitions for $\mathbf{GC}^{(w)}_{(i)g}$ and $\mathbf{SC}^{(w)}_{(i)g}$. Then we have the following result.

Lemma 2 *Under Assumption 9, letting $w_g = \mathbf{1}_g'\mathbf{z}_g$,*

$$\begin{aligned}
S_{ig}(z, \mathbf{z}_g) = N_{ig}^{AT} + zN_{ig}^{SC(1)} + \mathbb{1}(w_g > 0) \sum_{w=1}^{w_g} N_{ig}^{SC(w)} + z\mathbb{1}(w_g > 0)\left[N_{ig}^{SC(w_g)} - N_{ig}^{SC(1)}\right] \\
+ \mathbb{1}(w_g > 0)\mathbf{z}_g' \left[\mathbf{C}_{(i)g} + \mathbb{1}(w_g > 1)\left(\sum_{w=w_g+1}^{n_g} \mathbf{SC}_{ig}^{(w)} + \sum_{w=1}^{w_g-1} \mathbf{GC}_{ig}^{(w)}\right)\right] \\
+ z\mathbb{1}(w_g > 0)\mathbf{z}_g'\left[\mathbf{GC}_{(i)g}^{(w_g)} - \mathbf{SC}_{(i)g}^{(w_g)}\right].
\end{aligned}$$

The illustrate the above lemma, suppose $\mathbf{z}_g = \mathbf{0}_g$. Then $w_g = 0$ and

$$S_{ig}(z, \mathbf{0}_g) = N_{ig}^{AT} + zN_{ig}^{SC(1)}$$

which equals the number of peer always-takers when $z = 0$, and the number of peer always-takers plus 1-social compliers when $z = 1$. On the other hand, suppose $z = 0$ and $w_g = 1$ so that only one of i's peers is assigned to treatment. Then,

$$S_{ig}(0, \mathbf{z}_g) = N_{ig}^{AT} + N_{ig}^{SC(1)} + \mathbf{z}_g'\mathbf{C}_{(i)g}.$$

The first two terms are the total number of unit i's peers who are always-takers and 1-social compliers, who will be treated under this assignment. The remaining term is the total number of unit i's compliant peers who are assigned to treatment under assignment $\mathbf{z}_g$.

Lemma 3 (Direct ITT effects) *Under Assumption 9,*

$$\begin{aligned}
&\mathbb{E}[Y_{ig}|Z_{ig} = 1, \mathbf{Z}_{(i)g} = \mathbf{0}_g] - \mathbb{E}[Y_{ig}|Z_{ig} = 0, \mathbf{Z}_{(i)g} = \mathbf{0}_g] = \\
&\quad \sum_{s=0}^{n_g} \mathbb{E}[Y_{ig}(1, s) - Y_{ig}(0, s)|D_{ig}(1, 0) = 1, N_{ig}^{AT} + N_{ig}^{SC(1)} = s]\mathbb{P}[D_{ig}(1, 0) = 1, N_{ig}^{AT} + N_{ig}^{SC(1)} = s] \\
&\quad + \sum_{s=0}^{n_g} \mathbb{E}[Y_{ig}(1, s) - Y_{ig}(0, s)|AT_{ig}, N_{ig}^{AT} = s]\mathbb{P}[AT_{ig}, N_{ig}^{AT} = s] \\
&\quad + \sum_{s=0}^{n_g} \mathbb{E}[Y_{ig}(0, s)|N_{ig}^{AT} \neq s, N_{ig}^{AT} + N_{ig}^{SC(1)} = s]\mathbb{P}[N_{ig}^{AT} \neq s, N_{ig}^{AT} + N_{ig}^{SC(1)} = s] \\
&\quad - \sum_{s=0}^{n_g} \mathbb{E}[Y_{ig}(0, s)|N_{ig}^{AT} = s, N_{ig}^{SC(1)} > 0]\mathbb{P}[N_{ig}^{AT} = s, N_{ig}^{SC(1)} > 0].
\end{aligned}$$

The following propositions generalize the identification results under personalized assignment, one-sided noncompliance and IPT, respectively.

Proposition 8 (Identification under PA) *Suppose that $D_{ig}(z,w_g)=D_{ig}(z)$ for all i and g. Under Assumption 9, for any $(d,\mathbf{d}_g)$ such that $\mathbb{P}[D_{ig}=d,\mathbf{D}_{(i)g}=\mathbf{d}_g]>0$ and*

$$\mathbb{P}[Z_{ig}=z,\mathbf{Z}_{(i)g}=\mathbf{z}_g|D_{ig}=d,\mathbf{D}_{(i)g}=\mathbf{d}_g]>0 \tag{1}$$

for all $(z,\mathbf{z}_g)$, $\mathbb{P}[\xi_{ig}=\xi(d_i),(\xi_{jg}=\xi(d_j))_{j\neq i}]$ and $\mathbb{E}[Y_{ig}(d,\mathbf{d}_g)|\xi_{ig}=\xi(d_i),(\xi_{jg}=\xi(d_j))_{j\neq i}]$ are identified, where $\xi(1)=\mathrm{C},\mathrm{AT}$ and $\xi(0)=\mathrm{C},\mathrm{NT}$. In particular, if

$$\mathbb{P}[Z_{ig}=z,\mathbf{Z}_{(i)g}=\mathbf{z}_g,D_{ig}=d,\mathbf{D}_{(i)g}=\mathbf{d}_g]>0,$$

for all $(z,\mathbf{z}_g,d,\mathbf{d}_g)$, then $\mathbb{E}[Y_{ig}(d,s_g)|\cap_{i=1}^{n_g}C_{ig}]$ is identified for all (d,s_g).

In addition, when peers' types are exchangeable, $\mathbb{P}[\xi_{ig}=\xi(d_i),N_{ig}^{AT}=n^{AT},N_{ig}^{C}=n^{C},N_{ig}^{NT}=n^{NT}]$ and $\mathbb{E}[Y_{ig}(d,s_g)|\xi_{ig}=\xi(d_i),N_{ig}^{AT}=n^{AT},N_{ig}^{C}=n^{C},N_{ig}^{NT}=n^{NT}]$ are identified, where $0\leq n^C\leq n_g$, $\max\{0,s_g-n^C\}\leq n^{AT}\leq\min\{s,n_g-n^C\}$ and $n^{NT}=n_g-n^{AT}-n^C$.

The idea in this result is analogous to the one in Proposition 2, in which we exploit the different combinations of assignments and treatment statuses to infer each unit's type. Note that this proposition ensures identification of all average direct and spillover effects for groups in which all units are compliers, as long as all the required probabilities are non-zero.

Proposition 9 (Identification under OSN) *Suppose that $D_{ig}(0,w_g)=0$ for all i,g and w_g. Under Assumption 9, for any $j\neq i$,*

$$\begin{aligned}
\mathbb{E}[Y_{ig}(0,0)]&=\mathbb{E}[Y_{ig}|Z_{ig}=0,W_{ig}=0]\\
\mathbb{E}[Y_{ig}(1,0)|C_{ig}]\mathbb{P}[C_{ig}]&=\mathbb{E}[Y_{ig}D_{ig}|Z_{ig}=1,W_{ig}=0]\\
\mathbb{E}[Y_{ig}(0,1)|C_{jg}]\mathbb{P}[C_{jg}]&=\mathbb{E}[Y_{ig}D_{jg}|Z_{ig}=0,Z_{jg}=1,W_{ig}=1]\\
\mathbb{E}[Y_{ig}(0,0)|C_{ig}]\mathbb{P}[C_{ig}]&=\mathbb{E}[Y_{ig}|Z_{ig}=0,W_{ig}=0]-\mathbb{E}[Y_{ig}(1-D_{ig})|Z_{ig}=1,W_{ig}=0]\\
\mathbb{E}[Y_{ig}(0,0)|C_{jg}]\mathbb{P}[C_{jg}]&=\mathbb{E}[Y_{ig}|Z_{ig}=0,W_{ig}=0]-\mathbb{E}[Y_{ig}(1-D_{jg})|Z_{ig}=0,Z_{jg}=1,W_{ig}=1]\\
\mathbb{E}[Y_{ig}(0,0)|NT_{ig},N_{ig}^{NT}=n_g]\mathbb{P}[NT_{ig},N_{ig}^{NT}=n_g]&=\mathbb{E}\left[Y_{ig}\prod_{i=1}^{n_g+1}(1-D_{ig})\middle|Z_{ig}+W_{ig}=n_g+1\right]
\end{aligned}$$

where

$$\mathbb{P}[NT_{ig},N_{ig}^{NT}=n_g]=\mathbb{E}\left[\prod_{i=1}^{n_g+1}(1-D_{ig})\middle|Z_{ig}+W_{ig}=n_g+1\right].$$

Notice that this result only identifies the average direct and spillover effects when only one unit is treated. The reason is that, as soon as there is more than one unit assigned to treatment, it is not possible to distinguish between compliers and group compliers or between group compliers and never-takers. I provide this result to show that the conclusions from Proposition 3 still hold in the general case. I leave the question of what other parameters can be identified in this more general setting for future research.

Proposition 10 (Identification under IPT) *Suppose that for all $i \neq j$, $((Y_{ig}(d,s))_{(d,s)} \perp\!\!\!\perp (D_{jg}(z',w'))_{(z',w')} | (D_{ig}(z,w))_{(z,w)}$. Under Assumption 9, if $\mathbb{P}[D_{ig} = d, Z_{ig} = z, S_{ig} = s, W_{ig} = w] > 0$ for all (d,z,s,w), then $\mathbb{E}[Y_{ig}(1,s)|\xi_{ig}]$ is identified for all s and $\xi_{ig} \neq \mathrm{NT}$ and $\mathbb{E}[Y_{ig}(0,s)|\xi_{ig}]$ is identified for all s and $\xi_{ig} \neq \mathrm{AT}$.*

In particular, this result implies that all the average potential outcomes for compliers $\mathbb{E}[Y_{ig}(d,s_g)|C_{ig}]$ are identified.

Since the conditions for identification of average potential outcomes and causal effects may be harder to satisfy for arbitrary group sizes, the researcher may be willing to combine different sets of assumptions to make the problem more tractable. The next result shows that combining personalized assignment with IPT provides a simple way to identify and estimate all the average causal effects on compliers.

Proposition 11 (Identification under PA and IPT) *Suppose that $D_{ig}(z,w_g) = D_{ig}(z)$ for all i and g and $((Y_{ig}(d,s))_{(d,s)} \perp\!\!\!\perp (D_{jg}(z'))_{z'} | (D_{ig}(z))_z$, $\forall\, i \neq j$. Under Assumption 9, for any s such that $\mathbb{P}[Z_{ig} = 1, D_{ig} = 1, S_{ig} = s] > 0$ and $\mathbb{P}[Z_{ig} = 0, D_{ig} = 1, S_{ig} = s] > 0$,*

$$\mathbb{E}[Y_{ig}(1,s)|AT_{ig}] = \frac{\mathbb{E}[Y_{ig}D_{ig}|Z_{ig} = 0, S_{ig} = s]}{\mathbb{E}[D_{ig}|Z_{ig} = 0]}$$

and

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

For any s such that $\mathbb{P}[Z_{ig} = 0, D_{ig} = 0, S_{ig} = s] > 0$ and $\mathbb{P}[Z_{ig} = 1, D_{ig} = 0, S_{ig} = s] > 0$,

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

and

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

This result shows how combining personalized assignment with IPT can significantly simplify the problem. In this case, we can obtain simple closed forms for the local average potential outcomes that can be easily combined to construct treatment effects. Specifically, note that by Proposition 11,

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

which is a Wald ratio (2SLS estimand) using Z_{ig} as an instrument for D_{ig}, conditional on $S_{ig}=s$.

9 Conclusion

This paper proposes a potential outcomes framework to analyze identification and estimation of causal spillover effects using instrumental variables. I 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 propose three alternative strategies to identify parameters of interest, based on restricting (i) the channels through which spillovers can materialize, (ii) the degree of noncompliance, or (ii) treatment effect heterogeneity.

To some extent, the choice between these alternative approaches for identifying causal parameters in a specific empirical application can be based on observable data. One-sided noncompliance can be ruled out if there are units with $Z_{ig}=0$ and $D_{ig}=1$. On the other hand, the absence of spillovers on treatment status (personalized assignment) can be tested by applying the methods in Vazquez-Bare (2017), using treatment status as an outcome. When the researcher can control instrument assignment, as in an encouragement design, this restriction can be enforced by preventing units from reacting to their peers' instrument values, for example by withholding information on the peers' assignments if possible. Finally, when these conditions do not hold in an empirical application, the researcher may be willing to assume IPT. Although IPT cannot be tested without further assumptions, it can allow the researcher to identify and estimate an array of causal effects whenever its validity can be argued in a specific setting.

References

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

_ **and Matias D. Cattaneo**, "Econometric Methods for Program Evaluation," *Annual Review of Economics*, August 2018, *10*, 465–503.

Angrist, Joshua and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2009.

Angrist, Joshua D. and Alan B. Krueger, "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, December 2001, *15* (4), 69–85.

Athey, Susan and Guido Imbens, "Chapter 3 - The Econometrics of Randomized Experiments," in Abhijit Vinayak Banerjee and Esther Duflo, eds., *Handbook of Field Experiments*, Vol. 1 of *Handbook of Economic Field Experiments*, North-Holland, 2017, pp. 73 – 140.

Babcock, Philip, Kelly Bedard, Gary Charness, John Hartman, and Heather Royer, "Letting Down the Team? Social Effects of Team Incentives," *Journal of the European Economic Association*, 2015, *13* (5), 841–870.

Baird, Sarah, Aislinn Bohren, Craig McIntosh, and Berk Özler, "Optimal Design of Experiments in the Presence of Interference," *The Review of Economics and Statistics*, 2018, *100* (5), 844–860.

Basse, G W, A Feller, and P Toulis, "Randomization tests of causal effects under interference," *Biometrika*, 02 2019, *106* (2), 487–494.

Basse, Guillaume and Avi Feller, "Analyzing two-stage experiments in the presence of interference," *Journal of the American Statistical Association*, 2018, *113* (521), 41–55.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, "Identification of peer effects through social networks," *Journal of Econometrics*, 2009, *150* (1), 41 – 55.

Comola, Margherita and Silvia Prina, "Treatment Effect Accounting for Network Changes," *The Review of Economics and Statistics*, forthcoming.

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora, "Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment," *The Quarterly Journal of Economics*, 2013, *128* (2), 531–580.

Davezies, Laurent, Xavier D'Haultfoeuille, and Denis Fougère, "Identification of peer effects using group size variation," *Econometrics Journal*, 2009, *12* (3), 397–413.

Duflo, Esther and Emmanuel Saez, "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," *The Quarterly Journal of Economics*, 2003, *118* (3), 815–842.

Eckles, Dean, René F. Kizilcec, and Eytan Bakshy, "Estimating peer effects in networks with peer encouragement designs," *Proceedings of the National Academy of Sciences*, 2016, *113* (27), 7316–7322.

Fletcher, Jason and Ryne Marksteiner, "Causal Spousal Health Spillover Effects and Implications for Program Evaluation," *American Economic Journal: Economic Policy*, November 2017, *9* (4), 144–66.

Foos, Florian and Eline A. de Rooij, "All in the Family: Partisan Disagreement and Electoral Mobilization in Intimate Networks—A Spillover Experiment," *American Journal of Political Science*, 2017, *61* (2), 289–304.

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

Giorgi, Giacomo De, Michele Pellizzari, and Silvia Redaelli, "Identification of Social Interactions through Partially Overlapping Peer Groups," *American Economic Journal: Applied Economics*, April 2010, *2* (2), 241–75.

Graham, Bryan S., "Identifying Social Interactions Through Conditional Variance Restrictions," *Econometrica*, 2008, *76* (3), 643–660.

Hudgens, Michael G. and M. Elizabeth Halloran, "Toward Causal Inference with Interference," *Journal of the American Statistical Association*, 2008, *103* (482), 832–842.

Ichino, Nahomi and Matthias Schündeln, "Deterring or Displacing Electoral Irregularities? Spillover Effects of Observers in a Randomized Field Experiment in Ghana," *The Journal of Politics*, 2012, *74* (1), 292–307.

Imai, Kosuke, Zhichao Jiang, and Anup Malani, "Causal Inference with Interference and Noncompliance in Two-Stage Randomized Experiments," *working paper*, 2018.

Imbens, Guido W. and Donald B. Rubin, "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *The Review of Economic Studies*, 1997, *64* (4), 555–574.

_ **and Jeffrey M. Wooldridge**, "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, March 2009, *47* (1), 5–86.

— **and Joshua D. Angrist**, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475.

John, Peter and Tessa Brannan, "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*, 2008, *38* (3), 565–574.

Johnsson, Ida and Hyunhgsik Roger Moon, "Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach," *The Review of Economics and Statistics*, forthcoming.

Kang, Hyunseung and Guido Imbens, "Peer Encouragement Designs in Causal Inference with Partial Interference and Identification of Local Average Network Effects," *arXiv:1609.04464*, 2016.

— **and Luke Keele**, "Spillover Effects in Cluster Randomized Trials with Noncompliance," *arXiv:1808.06418*, 2018.

Kline, Brendan and Elie Tamer, "Econometric Analysis of Models with Social Interactions," in Bryan S. Graham and Aureo De Paula, eds., *The Econometric Analysis of Network Data*, Elsevier, 2019.

Leung, Michael, "Causal Inference under Approximate Neighborhood Interference," *Working Paper*, 2020.

— , "Treatment and Spillover Effects under Network Interference," *The Review of Economics and Statistics*, forthcoming.

Moffit, Robert, "Policy Interventions, Low-level Equilibria and Social Interactions," in Stephen N. Durlauf and Peyton Young, eds., *Social Dynamics*, MIT Press, 2001, pp. 45–82.

Puelz, David, Guillaume Basse, Avi Feller, and Panos Toulis, "A Graph-Theoretic Approach to Randomization Tests of Causal Effects Under General Interference," *arXiv:1910.10862*, 2019.

Sacerdote, Bruce, "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly Journal of Economics*, 2001, *116* (2), 681–704.

Sobel, Michael E., "What Do Randomized Studies of Housing Mobility Demonstrate?: Causal Inference in the Face of Interference," *Journal of the American Statistical Association*, 2006, *101* (476), 1398–1407.

Titiunik, Rocío, "Natural Experiments," in James Druckman and Donald Green, eds., *Advances in Experimental Political Science*, In preparation for Cambridge University Press, 2019.

Vazquez-Bare, Gonzalo, "Identification and Estimation of Spillover Effects in Randomized Experiments," *working paper*, 2017.