Causal Spillover Effects Using Instrumental Variables*

Gonzalo Vazquez-Bare†

December 13, 2021

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 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 show that causal effects can be estimated by 2SLS in this case. I illustrate the proposed methods using data from an experiment on social interactions and voting behavior. I also introduce an alternative assumption, *independence of peers' types*, that identifies parameters of interest under two-sided noncompliance by restricting the amount of heterogeneity in average potential outcomes.

Keywords: spillover effects, instrumental variables, imperfect compliance, treatment effects.

*I thank Matias Cattaneo, Clément de Chaisemartin, Xinwei Ma, Kenichi Nagasawa, Olga Namen, Neslihan Sakarya, Dick Startz and Doug Steigerwald for valuable discussions and suggestions, and seminar participants at Stanford University, UCSB Applied Microeconomics Lunch, Northwestern University, UCLA, UC San Diego, University of Chicago, UC Davis, USC and University of Essex for helpful comments. I also thank the editor, Jane-Ling Wang, the associate editor and three anonymous referees for their detailed suggestions that greatly improved the paper.

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

1 Introduction

An accurate assessment of spillover effects is crucial to learn about the costs and benefits of treatments and policies (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; Baird et al., 2018; Vazquez-Bare, forthcoming). RCTs, however, are often subject to imperfect compliance, as individuals assigned to treatment may refuse it and individuals not assigned to treatment may find alternative sources to receive it. In other cases, researchers may not have control on the treatment assignment, and instead may need to rely on quasi-experimental variation from a natural experiment (see e.g. Angrist and Krueger, 2001; Titiunik, 2019). In such cases, actual treatment receipt becomes endogenous, even if treatment assignment is randomized (or as-if-randomized). While previous literature has shown that instrumental variables (IVs) can address this type of endogeneity and identify local average treatment effects when spillovers are ruled out (Angrist et al., 1996), 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, I define causal direct and spillover effects in a setup with two-sided noncompliance. When treatment take-up is endogenous, spillover effects can occur in two stages: (i) treatment take-up, when a unit's instrument can affect its peers' treatment status, and (ii) outcomes, when a unit's treatment can affect its peers' responses. Focusing on the case in which spillovers occur within pairs (such as spouses or roommates), I propose a generalization of the monotonicity assumption (Imbens and Angrist, 1994) that partitions the population into five compliance types: in addition to always-takers, compliers and never-takers, units may be social-interaction compliers, who receive the treatment when either themselves or their peer are assigned to it, and group compliers, who only receive the treatment when both themselves and their peer are assigned to it. Proposition 1 provides conditions for identification of the marginal distribution of compliance types, and shows that the joint distribution is generally not identified.

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

Third, I show 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. I provide a way to assess the external validity of these parameters and discuss testable impli-

cations of the identification assumptions. In addition, I show that direct and indirect local average effects can be jointly estimated by two-stage least squares (2SLS). This provides a straightforward way to estimate these effects in practice based on standard regression methods and to conduct inference that is robust to weak instruments using Fieller-Anderson-Rubin confidence intervals. The proposed methods are illustrated using data from an experiment on social interactions in the household and voter turnout.

Finally, I discuss two generalizations of my results. The first one considers the case in which the IV identification assumptions hold after conditioning on a set of observed covariates. The second one generalizes the results to arbitrary group sizes by introducing a novel assumption, *independence of peers' types*, by which potential outcomes are independent of peers' compliance types conditional on own type. I show how this alternative assumption permits identification under two-sided noncompliance by limiting the amount of heterogeneity in the distribution of potential outcomes.

This paper contributes to the literature on causal inference under interference, which generally focuses on the case of two-stage experimental designs with perfect compliance (see Halloran and Hudgens, 2016, for a recent review). Existing studies analyzing imperfect compliance (Sobel, 2006; Kang and Imbens, 2016; Kang and Keele, 2018; Imai et al., 2021) consider identification, estimation and inference for specific experimental designs or by imposing specific restrictions on the spillovers structure. 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. In related work, DiTraglia et al. (2021) analyze identification and estimation of direct and spillover effects in randomized saturation designs under one-sided noncompliance. Their approach is complementary to mine, as it assumes a specific experimental design and rules out the presence of spillovers in treatment take-up but focuses on the case of large and unequally-sized groups.

This paper can also be linked to the literature on multiple instruments (Imbens and Angrist, 1994; Mogstad et al., forthcoming) since the vectors of own and peers' instruments and treatments in spillover analysis can be rearranged as a multivalued instrument and a multivalued treatment. While most of the existing literature in this area considers the case of multiple instruments and a single binary treatment, my setting with unrestricted spillovers introduces both multiple instruments and multiple treatments (see Remark 2 for further discussion).

The rest of the article is organized as follows. Section 2 describes the setup, introduces the notation and defines the causal parameters of interest. Section 3 analyzes ITT parameters. Section 4 provides the main identification results under one-sided noncompliance. Section 5 analyzes estimation and inference, and Section 6 applies the proposed methods in an

empirical setting. Finally, Section 7 generalizes the results to conditional-on-observables IV and multiple units per group, and Section 8 concludes. The supplemental appendix contains the technical proofs and additional results and discussions.

2 Setup

Consider a random sample of independent and identically distributed groups indexed by $g = 1, \ldots, G$. Spillovers are assumed to occur between units in the same group, but not between units in different groups. I start by considering the case in which each group consists of two identically-distributed units, so that each unit i in group g has one peer. This setup has a a wide range of applications in which groups consist, for example, of roommates in college dormitories (Sacerdote, 2001; Babcock et al., 2015; Garlick, 2018), spouses (Fletcher and Marksteiner, 2017; Foos and de Rooij, 2017), siblings (Barrera-Osorio et al., 2011), etc. Section 7 generalizes this setup to the case of multiple units per group (see Sacarny et al., 2018, for an example).

The goal is to study the effect of a treatment on an outcome of interest Y_{ig}, allowing for within-group spillovers. The individual treatment status of unit i in group g is denoted by D_{ig}, taking values $d \in \{0, 1\}$.[1] For each unit i, D_{jg} with $j \neq i$ is the treatment indicator corresponding to unit i's peer. Treatment take-up can be endogenous and is allowed to be arbitrarily correlated with individual characteristics, both observable and unobservable. To address this endogeneity, I assume the researcher has access to a pair of instrumental variables, (Z_{ig}, Z_{jg}) for unit i and her peer j, taking values $(z, z') \in \{0, 1\}^2$. These instruments are as-if randomly assigned in a sense formalized below. Borrowing from the literature on imperfect compliance in RCTs, I will often refer to the instruments (Z_{ig}, Z_{jg}) as "assignments", as they indicate whether an individual is assigned (or encouraged) to get the treatment. 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; Titiunik, 2019, for general discussions and Rincke and Traxler, 2011, for an application).

For a given realization of the treatment statuses $(D_{ig}, D_{jg}) = (d, d')$ and the instruments $(Z_{ig}, Z_{jg}) = (z, z')$, the potential outcome for unit i in group g is a random variable denoted by $Y_{ig}(d, d', z, z')$. I assume that instruments do not directly affect potential outcomes, an assumption commonly known as the exclusion restriction.

Assumption 1 (Exclusion restriction) $Y_{ig}(d, d', z, z') = Y_{ig}(d, d', \tilde{z}, \tilde{z}')$ *for all* $(z, z', \tilde{z}, \tilde{z}')$.

Under Assumption 1, potential outcomes are a function of treatment status only, $Y_{ig}(d, d')$.

[1] Section A1.4 in the supplemental appendix generalizes the setup to multiple treatment levels.

In this setting, the direct effect of the treatment on unit i given peer's treatment status d' is defined as $Y_{ig}(1,d') - Y_{ig}(0,d')$ and the spillover effect on unit i given own treatment status d is defined as $Y_{ig}(d,1) - Y_{ig}(d,0)$. The observed outcome for unit i in group g is the value of the potential outcome under the observed treatment realization, $Y_{ig} = Y_{ig}(D_{ig}, D_{jg})$.

The possibility of endogenous treatment status introduces an additional channel through which spillovers can materialize. For example, in an encouragement design targeted at couples, unit i may not be offered the incentive to participate in the program, so that $Z_{ig} = 0$, but she may hear about the program from her spouse who is assigned the incentive, $Z_{jg} = 1$, and decide to participate so $D_{ig} = 1$. More generally, treatment status may depend on both own and peer's assignment. To formalize this phenomenon, the potential treatment status for unit i in group g given instruments values $(Z_{ig}, Z_{jg}) = (z, z')$ will be denoted by $D_{ig}(z, z')$, and the spillover effect on unit i's treatment status is $D_{ig}(z,1) - D_{ig}(z,0)$ for $z = 0, 1$. The observed treatment status is $D_{ig}(Z_{ig}, Z_{jg})$.

The following assumption formalizes the requirement that instruments are as-if randomly assigned.

Assumption 2 (Independence) *Let* $\mathbf{y}_{ig} = (Y_{ig}(d,d'))_{(d,d')}$ *and* $\bar{\mathbf{d}}_{ig} = (D_{ig}(z,z'))_{(z,z')}$. *Then,* $(\mathbf{y}_{ig}, \bar{\mathbf{d}}_{ig}, \mathbf{y}_{jg}, \bar{\mathbf{d}}_{jg}) \perp\!\!\!\perp (Z_{ig}, Z_{jg})$.

Section 7.1 offers an alternative version of this assumption in which independence holds after conditioning on a set of observable covariates.

The different values that $D_{ig}(z, z')$ can take determine each unit's *compliance type*, which indicates how each unit's treatment status responds to different configurations of instruments (Z_{ig}, Z_{jg}). A careful analysis of compliance types is crucial for defining and identifying causal parameters in this setting: the instruments affect each compliance type in a different way, which in turns determines what parameters can and cannot be identified by harnessing the exogenous variation generated by the instruments, as I discuss next.

2.1 Compliance Types

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 assignment, determines each unit's compliance type. 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 her 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 is a total of 16 different compliance types in the population. To reduce the number of compliance types, I

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)

will introduce the following restriction that generalizes the commonly invoked monotonicity assumption to the spillovers case.

Assumption 3 (Monotonicity) *For all i and g, $D_{ig}(1,1) \geq D_{ig}(1,0) \geq D_{ig}(0,1) \geq D_{ig}(0,0)$.*

This is a strict generalization of the monotonicity assumption in Angrist et al. (1996) and Kang and Imbens (2016) who require that $D_{ig}(1,1) = D_{ig}(1,0)$ and $D_{ig}(0,1) = D_{ig}(0,0)$. Assumption 3 reduces the compliance types to five, listed in Table 1 in decreasing order of likelihood of being treated. 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. Note that monotonicity 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.

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.

Remark 1 (Monotonicity and ordering) *The ordering in Assumption 3 is without loss of generality and can be rearranged depending on the context. For example, if the treatment is alcohol consumption and the instrument is a randomly assigned incentive to reduce alcohol intake, the inequalities can be reverted, $D_{ig}(1,1) \leq D_{ig}(1,0) \leq D_{ig}(0,1) \leq D_{ig}(0,0)$. More generally, any ordering is possible provided the same ordering holds for all units in the population.*

Remark 2 (Connection to multi-valued instruments and treatments) *The setup in this paper can be mapped into one where the pair (Z_{ig}, Z_{jg}) is viewed a multi-valued instrument $A_{ig} = 2Z_{ig} + Z_{jg} \in \{0,1,2,3\}$ and (D_{ig}, D_{jg}) is viewed a multi-valued treatment*

$T_{ig} = 2D_{ig} + D_{jg} \in \{0, 1, 2, 3\}$, *where* $T_{ig}(a)$ *denotes potential treatment status under assignment* a. *This setup is analyzed under general conditions by Heckman and Pinto (2018). As shown in the upcoming sections, the spillovers case introduces specific modeling restrictions and a different way to interpret heterogeneity in treatment effects. In particular, unordered monotonicity (Assumption A-3 in Heckman and Pinto, 2018) does not hold in this setting. For example, unordered monotonicity requires that either* $\mathbb{1}(T_{ig}(1) = 3) \geq \mathbb{1}(T_{ig}(2) = 3)$ *for all* i, g *or* $\mathbb{1}(T_{ig}(1) = 3) \leq \mathbb{1}(T_{ig}(2) = 3)$ *for all* i, g. *In my setting,* $\mathbb{1}(T_{ig}(1) = 3) = D_{ig}(0,1)D_{jg}(1,0)$ *and* $\mathbb{1}(T_{ig}(2) = 3) = D_{ig}(1,0)D_{jg}(0,1)$. *Now, in the event* $\{AT_{ig}, C_{jg}\}$, *we have that* $D_{ig}(0,1)D_{jg}(1,0) = 1 > D_{ig}(1,0)D_{jg}(0,1) = 0$, *whereas in the event* $\{C_{ig}, AT_{jg}\}$, $D_{ig}(0,1)D_{jg}(1,0) = 0 < D_{ig}(1,0)D_{jg}(0,1) = 1$ *and hence the weak inequality does not hold.*

Under Assumptions 1, 2 and 3, the marginal distribution of compliance types in the population is identified, as the following proposition shows.

Proposition 1 (Distribution of compliance types) *Under Assumptions 1-3,*

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

and $\mathbb{P}[NT_{ig}] = 1 - \mathbb{P}[AT_{ig}] - \mathbb{P}[SC_{ig}] - \mathbb{P}[C_{ig}] - \mathbb{P}[GC_{ig}]$. *Finally,* $\mathbb{P}[AT_{ig}, AT_{jg}] = \mathbb{E}[D_{ig}D_{jg}|Z_{ig} = 0, Z_{jg} = 0]$ *and* $\mathbb{P}[NT_{ig}, NT_{jg}] = \mathbb{E}[(1 - D_{ig})(1 - D_{jg})|Z_{ig} = 1, Z_{jg} = 1]$.

All the proofs can be found in the supplemental appendix. Proposition 1 can be used to test for the presence of average spillover effects on treatment status. Note that under Assumption 3, $\mathbb{E}[D_{ig}(0,1) - D_{ig}(0,0)] = \mathbb{P}[SC_{ig}]$ 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 instruments are as-if randomly assigned, these issues can be analyzed within the framework in Vazquez-Bare (forthcoming).

2.2 Causal Parameters and Estimands of Interest

In the presence of spillovers, average direct and spillover effects can be defined as differences between average potential outcomes under different treatment configurations, $\mathbb{E}[Y_{ig}(d, d') - Y_{ig}(\tilde{d}, \tilde{d}')]$. When the treatment vector (D_{ig}, D_{jg}) is randomly assigned, average potential outcomes (and thus treatment effects) are identified by the relationship $\mathbb{E}[Y_{ig}(d, d')] = \mathbb{E}[Y_{ig}|D_{ig} = d, D_{jg} = d']$ (see Vazquez-Bare, forthcoming, and references therein).

When the treatment is endogenous, average causal effects are generally not point identified. In the absence of spillovers, that is, when $Y_{ig}(d,d') = Y_{ig}(d)$ and $D_{ig}(z,z') = D_{ig}(z)$, the instrument's exogenous variation can be leveraged to identify the local average effect on compliers, $\mathbb{E}[Y_{ig}(1) - Y_{ig}(0)|D_{ig}(1) > D_{ig}(0)]$, emphasizing that average potential outcomes and treatment effects can vary over compliance types, and that the instrument only provides identifying variation on the specific subpopulation whose behavior is affected by the instrument. In the presence of spillovers, without further restrictions, average potential outcomes can depend on both own and peer's compliance types, $\mathbb{E}[Y_{ig}(d,d')|\xi_{ig} = \xi, \xi_{jg} = \xi']$. I refer to these parameters as "local average potential outcomes".

The upcoming section shows that, in general, the simultaneous presence of spillovers on treatment status and outcomes can impede identification of causally interpretable parameters even when the instruments are randomly assigned. However, Section 4 shows that when noncompliance is one-sided, it is possible to identify the average direct effect on compliers, $\mathbb{E}[Y_{ig}(1,0) - Y_{ig}(0,0)|C_{ig}]$ and the average spillover effect on units with compliant peers, $\mathbb{E}[Y_{ig}(0,1) - Y_{ig}(0,0)|C_{jg}]$. Section 7 provides an alternative identification assumption that applies to the case of multiple units per group.

3 Intention-to-Treat Parameters

Intention-to-treat (ITT) analysis focuses on the variation in Y_{ig} generated by the instruments. 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 by the proportion of compliers, which is identified under monotonicity and as-if random assignment of the instrument. 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 treatment effect heterogeneity.

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

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,*

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

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 interpret the above result, consider the effect of switching Z_{ig} from 0 to 1, 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 assignment, the ITT parameter is unable to isolate direct and indirect effects, which blurs its link to causal effects.

In some contexts, ITT parameters are deemed policy relevant as they measure the "effect" of offering the treatment or making the treatment available, as opposed to the effect of the

treatment itself (Abadie and Cattaneo, 2018). This interpretation is based on the fact that, without spillovers, the ITT is $\mathbb{E}[Y_{ig}|Z_{ig}=1]-\mathbb{E}[Y_{ig}|Z_{ig}=0]=\mathbb{E}[Y_{ig}(1)-Y_{ig}(0)|C_{ig}]\mathbb{P}[C_{ig}]$ which is the local average treatment effect, down-weighted by the compliance rate. In particular, this well-known fact has three implications that facilitate the interpretation of the ITT as a policy-relevant parameter: (i) it has the same sign as the LATE, and it equals zero if and only if the LATE is zero (unless the instrument is completely irrelevant) (ii) it is a lower bound for the LATE (in absolute value) and (iii) it is proportional to the LATE, so it can be easily rescaled to recover the LATE.

Lemma 1 shows that the close link between the ITT and the LATE breaks down in the presence of spillovers. First, the ITT now combines multiple average direct and spillover effects which can have different signs and magnitudes. In particular, the ITT could be zero even if all treatment effects are non-zero. Second, for this same reason, the ITT is no longer a lower bound for any of the direct or spillover effects. Finally, rescaling the ITT by the first stage $\mathbb{E}[D_{ig}|Z_{ig}=1,Z_{jg}=0]-\mathbb{E}[D_{ig}|Z_{ig}=0,Z_{jg}=0]$ does not recover a treatment effect.[2] Specifically, 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.

Remark 3 (Spillovers and instrument validity) *Another way to interpret the result in Lemma 1 is to think of spillovers in treatment take-up as violating instrument validity. 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.*

In all, this section shows that ITT parameters are generally not a useful measure of average direct and spillover effects. Instead of focusing on ITT parameters, which rely exclusively on variation generated by the instruments (Z_{ig},Z_{jg}), an alternative approach for identification is to exploit the combined variation in $(Z_{ig},Z_{jg},D_{ig},D_{jg})$. Imbens and Rubin (1997) show that, in the absence of spillovers, this approach allows for separate point identification of average (or distributions of) potential outcomes for compliers. This approach, however, breaks down in the presence of spillovers. The reason is that, without further assumptions, the possible combinations of treatment and instrument values are not enough to disentangle all the different compliance types. Further details on this issue are provided in Section A1.3 of the supplemental appendix. In what follows, I show that point identification can be restored when the degree of noncompliance is restricted.

[2]Notice that rescaling the ITT by the first stage recovers the estimand from a 2SLS regression instrumenting D_{ig} with Z_{ig} conditional on $Z_{jg}=0$.

4 Identification Under One-sided Noncompliance

This section shows that identification of some causal parameters can be achieved by limiting the amount of noncompliance. I will analyze the case in which noncompliance is one-sided. One-sided noncompliance (OSN) 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 a voter turnout experiment in which individuals in two-voter households are randomly assigned to receive a telephone call encouraging them to vote (as in Foos and de Rooij, 2017). In this case, units that are assigned $Z_{ig} = 1$ may fail to receive the actual phone call (for example, because they refuse to 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 $D_{ig} = 0$. More generally, one-sided noncompliance is common when the experimenter is the only provider of a treatment (Abadie and Cattaneo, 2018). I formalize this case as follows.

Assumption 4 (One-sided Noncompliance) $\mathbb{P}[D_{ig} = 1|Z_{ig} = 0] = 0$.

Under Assumption 3 (monotonicity), one-sided noncompliance implies the absence of always-takers and social-interaction compliers. Notice that Assumption 4 is testable. 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 instruments $(D_{ig}, D_{jg}, Z_{ig}, Z_{jg})$. Proposition 2 has the following implication.

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 and $\mathbb{E}[Y_{ig}(0,1) - Y_{ig}(0,0)|C_{jg}]$ is the average effect on untreated units with compliant peers. See Section 6 for a detailed discussion on these estimands in the context of an empirical application. Section A1.4 generalizes these results to the case of multiple treatment levels.

In addition to identifying these treatment effects, Proposition 2 can be used to assess whether average baseline 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 a non-complier, that is, $C_{ig}^c = NT_{ig} \cup GC_{ig}$.

Corollary 2 (Heterogeneity over compliance types) *Under Assumptions 1-4, if* $0 < \mathbb{P}[C_{ig}] < 1$,

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

and

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

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, 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.

4.1 Testing Instrument Validity

While identification assumptions are generally not testable, the existing literature on instrumental variables has provided ways to indirectly assess their validity by deriving testable implications (Balke and Pearl, 1997; Imbens and Rubin, 1997; Kitagawa, 2015). Based on these results, the following proposition provides a way to test for instrument validity in the presence of spillovers.

Proposition 3 (Instrument Validity) *Under Assumptions 1-4, for any Borel set $\mathcal{Y}$,*

$$\mathbb{P}[Y_{ig} \in \mathcal{Y}, D_{ig} = 0 | Z_{ig} = 1, Z_{jg} = 0] - \mathbb{P}[Y_{ig} \in \mathcal{Y} | Z_{ig} = 0, Z_{jg} = 0] \leq 0$$

and

$$\mathbb{P}[Y_{ig} \in \mathcal{Y}, D_{jg} = 0 | Z_{ig} = 0, Z_{jg} = 1] - \mathbb{P}[Y_{ig} \in \mathcal{Y} | Z_{ig} = 0, Z_{jg} = 0] \leq 0.$$

Because all the terms in Proposition 3 only involve observable variables, estimating them can provide a way to assess the validity of the identification assumptions. Estimation and inference for these objects can be conducted using the local regression distribution methods in Cattaneo et al. (forthcoming a) and Cattaneo et al. (forthcoming b).

5 Estimation and Inference

This section discusses estimation and inference for the causal effects in Corollary 1. In what follows, let $\mathbf{y}_g = (Y_{1g}, Y_{2g})'$, $\mathbf{Z}_g = (Z_{1g}, Z_{2g})'$ and $\mathbf{D}_g = (D_{1g}, D_{2g})'$.

Assumption 5 (Sampling and moments)

(a) $(\mathbf{y}_g', \mathbf{Z}_g', \mathbf{D}_g')_{g=1}^G$ *are independent and identically distributed across* g.

(b) For each g, (Y_{1g}, Z_{1g}, D_{1g}) *and* (Y_{2g}, Z_{2g}, D_{2g}) *are identically distributed, not necessarily independent.*

(c) $\mathbb{E}[Y_{ig}^4] < \infty$.

Part (a) of Assumption 5 states that the researcher has access to a random sample of iid groups. Part (b) states that observations within each group are identically distributed, but allows for an unrestricted correlation structure within groups. Part (c) is a standard regularity condition to ensure the appropriate moments are bounded.

The average direct and spillover effects in Corollary 1 can be estimated straightforwardly by replacing the right-hand side of each expression by their sample analog. Because these estimators are IV estimators using a binary instrument and a binary treatment on a specific subsample, I will refer to these estimators as *conditional Wald estimators*. More precisely, I

define conditional Wald estimators and their corresponding cluster-robust variance estimator as follows.

Definition 1 (Direct conditional Wald estimator) *Let* $\tilde{\mathbf{z}}_{ig} = (1, Z_{ig})'$, $\tilde{\mathbf{d}}_{ig} = (1, D_{ig})'$, $\tilde{\mathbf{z}}_g = (\tilde{\mathbf{z}}_{1g}', \tilde{\mathbf{z}}_{2g}')'$. *The direct conditional Wald estimator* $\hat{\boldsymbol{\delta}} = (\hat{\delta}_0, \hat{\delta}_1)'$ *is defined as the estimator from the 2SLS regression of* Y_{ig} *on an intercept and* D_{ig} *using* Z_{ig} *as an instrument, on the subsample of observations with* $Z_{jg} = 0$*, that is,*

$$\hat{\boldsymbol{\delta}} = \begin{bmatrix} \hat{\delta}_0 \\ \hat{\delta}_1 \end{bmatrix} = \left(\sum_g \tilde{\mathbf{z}}_g' \tilde{\mathbf{w}}_g \tilde{\mathbf{d}}_g \right)^{-1} \sum_g \tilde{\mathbf{z}}_g' \tilde{\mathbf{w}}_g \mathbf{y}_g$$

whenever $\sum_g \sum_i (1 - Z_{ig})(1 - Z_{jg}) > 0$, $\sum_g \sum_i Z_{ig}(1 - Z_{jg}) > 0$ *and* $\sum_g \sum_i D_{ig} Z_{ig}(1 - Z_{jg}) > 0$*, where*

$$\tilde{\mathbf{w}}_g = \begin{bmatrix} 1 - Z_{2g} & 0 \\ 0 & 1 - Z_{1g} \end{bmatrix}.$$

The cluster-robust variance estimator for $\hat{\boldsymbol{\delta}}$ *is:*

$$\hat{\mathbf{V}}_{\mathsf{cr}}(\hat{\boldsymbol{\delta}}) = \left(\sum_g \tilde{\mathbf{z}}_g' \tilde{\mathbf{w}}_g \tilde{\mathbf{d}}_g \right)^{-1} \sum_g \tilde{\mathbf{z}}_g' \tilde{\mathbf{w}}_g \hat{\mathbf{u}}_g \hat{\mathbf{u}}_g' \tilde{\mathbf{w}}_g \tilde{\mathbf{z}}_g \left(\sum_g \tilde{\mathbf{d}}_g' \tilde{\mathbf{w}}_g \tilde{\mathbf{z}}_g \right)^{-1}$$

where $\hat{u}_{ig} = Y_{ig} - \tilde{\mathbf{d}}_{ig}' \hat{\boldsymbol{\delta}}$ *and* $\hat{\mathbf{u}}_g = (\hat{u}_{1g}, \hat{u}_{2g})'$.

Definition 2 (Indirect conditional Wald estimator) *Let* $\check{\mathbf{z}}_{ig} = (1, Z_{jg})'$, $\check{\mathbf{d}}_{ig} = (1, D_{jg})'$, $\check{\mathbf{z}}_g = (\check{\mathbf{z}}_{1g}', \check{\mathbf{z}}_{2g}')'$. *The indirect conditional Wald estimator* $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_0, \hat{\lambda}_1)'$ *is defined as the estimator from the 2SLS regression of* Y_{ig} *on an intercept and* D_{jg} *using* Z_{jg} *as an instrument, on the subsample of observations with* $Z_{ig} = 0$*, that is,*

$$\hat{\boldsymbol{\lambda}} = \begin{bmatrix} \hat{\lambda}_0 \\ \hat{\lambda}_1 \end{bmatrix} = \left(\sum_g \check{\mathbf{z}}_g' \check{\mathbf{w}}_g \check{\mathbf{d}}_g \right)^{-1} \sum_g \check{\mathbf{z}}_g' \check{\mathbf{w}}_g \mathbf{y}_g$$

whenever $\sum_g \sum_i (1 - Z_{ig})(1 - Z_{jg}) > 0$, $\sum_g \sum_i Z_{jg}(1 - Z_{ig}) > 0$ *and* $\sum_g \sum_i D_{jg} Z_{jg}(1 - Z_{ig}) > 0$*, where*

$$\check{\mathbf{w}}_g = \begin{bmatrix} 1 - Z_{1g} & 0 \\ 0 & 1 - Z_{2g} \end{bmatrix}.$$

The cluster-robust variance estimator for $\hat{\boldsymbol{\lambda}}$ *is:*

$$\hat{\mathbf{V}}_{\mathsf{cr}}(\hat{\boldsymbol{\lambda}}) = \left(\sum_g \check{\mathbf{z}}_g' \check{\mathbf{w}}_g \check{\mathbf{d}}_g \right)^{-1} \sum_g \check{\mathbf{z}}_g' \check{\mathbf{w}}_g \hat{\boldsymbol{\nu}}_g \hat{\boldsymbol{\nu}}_g' \check{\mathbf{w}}_g \check{\mathbf{z}}_g \left(\sum_g \check{\mathbf{d}}_g' \check{\mathbf{w}}_g \check{\mathbf{z}}_g \right)^{-1}$$

where $\hat{\nu}_{ig} = Y_{ig} - \check{\mathbf{d}}_{ig}' \hat{\boldsymbol{\lambda}}$ *and* $\hat{\boldsymbol{\nu}}_g = (\hat{\nu}_{1g}, \hat{\nu}_{2g})'$.

An alternative estimation strategy in a setting with multiple instruments and multiple endogenous variables is to combine all regressors and instruments into a single 2SLS regression.

In a constant coefficients setting, this estimation strategy yields consistent and asymptotically normal estimators. However, these features do not generally extend to a setting with heterogeneous effects. In what follows I show that, under one-sided noncompliance, if the 2SLS regression combining both instruments and endogenous variables is fully saturated, the resulting estimators and cluster-robust standard errors are in fact equivalent to the conditional Wald estimators. I start by defining the fully-saturated 2SLS regression estimator as follows.

Definition 3 (Saturated 2SLS regression) *Let* $\mathbf{z}_{ig} = (1, Z_{ig}, Z_{jg}, Z_{ig}Z_{jg})'$, $\mathbf{d}_{ig} = (1, D_{ig}, D_{jg}, D_{ig}D_{jg})'$, $\mathbf{z}_g = (\mathbf{z}_{1g}', \mathbf{z}_{2g}')'$. *The saturated 2SLS regression estimator* $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)'$ *is defined as the estimator from the 2SLS regression of* Y_{ig} *on an intercept,* D_{ig}, D_{jg} *and* $D_{ig}D_{jg}$ *using* Z_{ig}, Z_{jg} *and* $Z_{ig}Z_{jg}$ *as instruments on the full sample, that is,*

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \left(\sum_g \mathbf{z}_g' \mathbf{d}_g \right)^{-1} \sum_g \mathbf{z}_g' \mathbf{y}_g.$$

whenever $\sum_g \sum_i (1 - Z_{ig})(1 - Z_{jg}) > 0$, $\sum_g \sum_i Z_{ig}(1 - Z_{jg}) > 0$, $\sum_g \sum_i Z_{ig}Z_{jg} > 0$, $\sum_g \sum_i D_{ig}Z_{ig}(1 - Z_{jg}) > 0$ *and* $\sum_g \sum_i D_{ig}D_{jg}Z_{ig}Z_{jg} > 0$. *The cluster-robust variance estimator for* $\hat{\boldsymbol{\beta}}$ *is:*

$$\hat{\mathbf{V}}_{\mathsf{cr}}(\hat{\boldsymbol{\beta}}) = \left(\sum_g \mathbf{z}_g' \mathbf{d}_g \right)^{-1} \sum_g \mathbf{z}_g' \hat{\boldsymbol{\varepsilon}}_g \hat{\boldsymbol{\varepsilon}}_g' \mathbf{z}_g \left(\sum_g \mathbf{d}_g' \mathbf{z}_g \right)^{-1}$$

where $\hat{\varepsilon}_{ig} = Y_{ig} - \mathbf{d}_{ig}'\hat{\boldsymbol{\beta}}$ *and* $\hat{\boldsymbol{\varepsilon}}_g = (\hat{\varepsilon}_{1g}, \hat{\varepsilon}_{2g})'$.

The following theorem establishes the equivalence between conditional Wald estimators and the saturated 2SLS estimators.

Theorem 1 (Equivalence between conditional Wald and 2SLS) *Consider the estimators* $\hat{\boldsymbol{\delta}}$, $\hat{\boldsymbol{\lambda}}$ *and* $\hat{\boldsymbol{\beta}}$ *from Definitions 1, 2 and 3. Suppose that Assumptions 1-5 hold, and that* $\sum_g \sum_i (1 - Z_{ig})(1 - Z_{jg}) > 0$, $\sum_g \sum_i Z_{ig}(1 - Z_{jg}) > 0$, $\sum_g \sum_i Z_{ig}Z_{jg} > 0$, $\sum_g \sum_i D_{ig}Z_{ig}(1 - Z_{jg}) > 0$ *and* $\sum_g \sum_i D_{ig}D_{jg}Z_{ig}Z_{jg} > 0$. *Then,* $\hat{\delta}_0 = \hat{\lambda}_0 = \hat{\beta}_0$, $\hat{\delta}_1 = \hat{\beta}_1$, $\hat{\lambda}_1 = \hat{\beta}_2$ *and* $\hat{\mathbf{V}}_{\mathsf{cr},11}(\hat{\boldsymbol{\delta}}) = \hat{\mathbf{V}}_{\mathsf{cr},11}(\hat{\boldsymbol{\lambda}}) = \hat{\mathbf{V}}_{\mathsf{cr},11}(\hat{\boldsymbol{\beta}})$, $\hat{\mathbf{V}}_{\mathsf{cr},22}(\hat{\boldsymbol{\delta}}) = \hat{\mathbf{V}}_{\mathsf{cr},22}(\hat{\boldsymbol{\beta}})$ *and* $\hat{\mathbf{V}}_{\mathsf{cr},22}(\hat{\boldsymbol{\lambda}}) = \hat{\mathbf{V}}_{\mathsf{cr},33}(\hat{\boldsymbol{\beta}})$.

In what follows let "$\to_{\mathbb{P}}$" denote convergence in probability and "$\to_{\mathcal{D}}$" denote convergence in distribution. The following result shows that the 2SLS are consistent and asymptotically normal. This result is standard in 2SLS and included only for completion.

Lemma 2 (Consistency and asymptotic normality) *Under Assumptions 1-5, if* $\mathbb{P}[Z_{ig} = 0, Z_{jg} = 0] > 0$, $\mathbb{P}[Z_{ig} = 1, Z_{jg} = 0] > 0$, $\mathbb{P}[Z_{ig} = 1, Z_{jg} = 1] > 0$, $\mathbb{E}[D_{ig}|Z_{ig} = 1, Z_{jg} = 0] > 0$

and $\mathbb{E}[D_{ig}D_{jg}|Z_{ig}=1,Z_{jg}=1]>0$*, then as* $G\to\infty$*,*

$$\hat{\boldsymbol{\beta}} \to_{\mathbb{P}} \boldsymbol{\beta} = \begin{bmatrix} \mathbb{E}[Y_{ig}(0,0)] \\ \mathbb{E}[Y_{ig}(1,0)-Y_{ig}(0,0)|C_{ig}] \\ \mathbb{E}[Y_{ig}(0,1)-Y_{ig}(0,0)|C_{jg}] \\ \beta_3 \end{bmatrix}$$

and

$$\sqrt{2G}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \to_{\mathcal{D}} \mathcal{N}(\mathbf{0},\mathbf{V}), \quad \mathbf{V}=\mathbb{E}[\mathbf{z}_g'\mathbf{d}_g]^{-1}\mathbb{E}[\mathbf{z}_g'\boldsymbol{\varepsilon}_g\boldsymbol{\varepsilon}_g'\mathbf{z}_g]\mathbb{E}[\mathbf{d}_g'\mathbf{z}_g]^{-1}$$

where $\varepsilon_{ig}=Y_{ig}-\mathbf{d}_g'\boldsymbol{\beta}$ *and* $\boldsymbol{\varepsilon}_g=(\varepsilon_{1g},\varepsilon_{2g})'$*. In addition,* $2G\hat{\mathbf{V}}_{\mathsf{cr}}(\hat{\boldsymbol{\beta}}) \to_{\mathbb{P}} \mathbf{V}$*.*

The interaction population coefficient β_3 does not have a direct causal interpretation, and its exact formula is given in the proof in the supplemental appendix.

Remark 4 (Interaction terms) *Although the coefficient on* $\hat{\beta}_3$ *corresponding to the interaction term* $D_{ig}D_{jg}$ *does not have a direct causal interpretation, the terms* $D_{ig}D_{jg}$ *and* $Z_{ig}Z_{jg}$ *need to be included to ensure the equivalence of the estimators by saturating the model. If* $\sum_g\sum_i Z_{ig}Z_{jg}=0$*, under imperfect compliance* $D_{ig}D_{jg}=0$ *and the results from Theorem 1 hold after excluding the interaction terms* $D_{ig}D_{jg}$ *and* $Z_{ig}Z_{jg}$ *from the estimation procedure.*

Notice that the magnitudes of interest defined in Corollary 2 and Proposition 3 cannot be written as 2SLS estimators, and hence Theorem 1 and Lemma 2 do not apply directly. However, all these parameters are nonlinear functions of sample means, and hence consistency and asymptotic normality follows under standard conditions. I provide further details on estimation and inference for these parameters in the supplemental appendix.

5.1 Weak-Instrument-Robust Inference

When non-compliance is severe, the proportion of compliers can be close to zero and, as a result, instruments may be weak. In such cases, the estimators analyzed above may have poor finite-sample properties, and inference based on the normal approximation may be unreliable. The 2SLS literature has provided several alternatives to conduct inference that is robust to weak instruments (see Andrews et al., 2019, for a recent review). In particular, Fieller or Anderson-Rubin (AR) confidence intervals have been shown to provide correct coverage regardless of the strength of the instruments.

By Theorem 1, the average direct and spillover effects can be estimated by two separate regressions that involve a single binary instrument and a single binary endogenous variable (the conditional Wald estimators), which give the same estimates and standard errors as the saturated 2SLS regression. The advantage of the conditional Wald estimators is that their corresponding AR confidence intervals can be obtained by solving a quadratic inequality as

shown by Dufour and Taamouti (2005) and Mikusheva (2010), without the need of computationally intensive grid searches or projection methods. Section A3 provides further details on how to construct AR confidence intervals in this context.

6 Application: Spillovers in Voting Behavior

In this section I illustrate the above results using data from a randomized experiment on voter mobilization conducted by Foos and de Rooij (2017). The goal of their study 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.

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 providing information and encouraging people to vote on the West Midlands Police and Crime Commissioner (PCC) 2012 election. A sample of 5,190 two-voter households with landline numbers were divided into treatment and control households, and within the households assigned to the treatment, only one household member was randomly selected to receive the telephone message.[3]

Because the telephone message is delivered by landline, this type of experiments is usually subject to 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 (Gerber and Green, 2000). 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. More precisely, 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 do not receive it. Since only one member of each treated household was selected to receive the call, we also have that $\mathbb{P}[Z_{ig} = 1, Z_{jg} = 1] = 0$. Given this 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

[3]The experiment had two different treatment intensities, which I pool here for illustration purposes. Section A4 in the supplemental appendix provides a detailed analysis of the multi-level treatment.

instrument.

The estimation results are shown in Table 2. Column (1) shows the naive estimates obtained 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. Taken at face value, these estimates suggest an ITT effect of about 2 percentage points and a local average effect of about 4 percentage points on voter turnout. However, in the presence of spillovers, these magnitudes do not generally have a clear causal interpretation. In fact, by Proposition 2, given this assignment mechanism,

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

so the naive 2SLS estimand that ignores spillovers equals the difference between the direct and indirect LATEs, where the indirect LATE is rescaled by the conditional probability of treatment assignment. Therefore, the naive 2SLS estimand can be close to zero whenever direct and spillover effects have the same sign.

Column (2) in Table 2 shows the estimated direct and indirect ITT and LATE parameters based on Corollary 1. The results reveal strong evidence of direct and indirect effects. On the one hand, ITT effects are positive and significant. ITT estimates are around 3 and 7 percentage points, respectively. While these magnitudes do not directly measure causal effects, as shown in Lemma 1, under one-sided noncompliance ITT parameters are attenuated (rescaled) measures of local average effects, that is, $\mathbb{E}[Y_{ig}|Z_{ig}=1,Z_{jg}=0]-\mathbb{E}[Y_{ig}|Z_{ig}=0,Z_{jg}=0]=\mathbb{E}[Y_{ig}(1,0)|C_{ig}]\mathbb{P}[C_{ig}]$ and $\mathbb{E}[Y_{ig}|Z_{ig}=0,Z_{jg}=1]-\mathbb{E}[Y_{ig}|Z_{ig}=0,Z_{jg}=0]=\mathbb{E}[Y_{ig}(0,1)|C_{jg}]\mathbb{P}[C_{jg}]$. These magnitudes can be interpreted as the "effect" of offering the treatment (see the discussion on ITT and LATE in Section 3).

On the other hand, 2SLS estimates reveal that 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. Although the first-stage estimate suggests a strong instrument, I also calculate weak-instrument-robust AR 95%-confidence intervals as a robustness check. These intervals are shown in parentheses in Table 2. Reassuringly, the AR confidence intervals are slightly wider but very close to the large-sample-based confidence intervals, and lead to the same conclusions.

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

Table 2: Main Estimation Results

	(1)	(2)
ITT		
Z_{ig}	0.0178	0.0305
	(0.0089)	(0.0114)
	[0.0004 , 0.0352]	[0.0081 , 0.0529]
Z_{jg}		0.0459
		(0.0116)
		[0.0232 , 0.0686]
2SLS		
D_{ig}	0.0394	0.0675
	(0.0196)	(0.0252)
	[0.0010 , 0.0778]	[0.0182 , 0.1168]
	(0.0008 , 0.0782)	(0.0175 , 0.1208)
D_{jg}		0.1017
		(0.0255)
		[0.0517 , 0.1518]
		(0.0502 , 0.1570)
N	9,860	9,860
Clusters	4,930	4,930

Notes: estimated results from reduced-form regressions ("ITT") and 2SLS regressions ("2SLS"). Column (1) shows the naive reduced-form and 2SLS estimates that ignore spillovers. Column (2) shows the ITT effects and local average direct and spillover effects obtained by 2SLS. Standard errors in parentheses. 95%-confidence intervals in brackets are based on the large-sample normal approximation. 95%-confidence intervals in parentheses are weak-instrument-robust AR confidence intervals. Estimation accounts for clustering at the household level.

type. This is different than comparing the direct and spillover effects on the population of compliers. Note that the indirect LATE is:

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

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

Finally, instrument validity can be assessed based on Proposition 3. In this application, because the outcome variable is binary, it is sufficient to run the following regression:

$$Y_{ig}(1 - D_{ig})(1 - D_{jg}) = \gamma_0 + \gamma_1 Z_{ig} + \gamma_2 Z_{jg} + u_{ig} \tag{1}$$

and test that $\gamma_1 \leq 0$ and $\gamma_2 \leq 0$. The results from this regression are shown in Table 3. As expected, both coefficients are negative and significant.

Table 3: Testing instrument validity

	(1)
Z_i	−0.0996
	(0.0094)
	[−0.1181 , −0.0811]
Z_j	−0.0893
	(0.0096)
	[−0.1082 , −0.0704]
N	9,860
Clusters	4,930

Notes: estimated results from Equation (1). Standard errors in parentheses. 95%-confidence intervals based on the large-sample normal approximation. Estimation accounts for clustering at the household level.

7 Generalizations and Extensions

7.1 Conditional-on-observables IV

In many cases, the assumption that the instruments are as good as randomly assigned is more credible after conditioning on a set of covariates. This is particularly relevant when the variation in the instruments is not controlled by the researcher, but instead comes from a natural experiment. Furthermore, Abadie (2003) shows that covariates can be exploited to identify the characteristics of the compliant subpopulation.

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. I introduce the following assumption.

Assumption 6 (Conditional-on-observables IV)

1. *Exclusion restriction:* $Y_{ig}(d, d', z, z') = Y_{ig}(d, d')$ *for all* (d, d', 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. *One-sided noncompliance:* $\mathbb{P}[D_{ig}(0, z') = 0|X_g] = 1$ *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 6,*

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

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 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. Estimation in this case can be based on reweighting methods after estimating the propensity score, as in Abadie (2003).

7.2 Multiple Units Per Group

Without further assumptions, identification becomes increasingly harder as group size grows. 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. One-sided noncompliance is not enough to identify causal parameters when more than one unit is assigned to treatment since, as soon as more than one unit is assigned to treatment, it is not possible to distinguish between compliers and group compliers or between group compliers and never-takers.

Some studies addressed this problem by restricting the way in which potential treatment

status depends on peers' assignments (see e.g. Kang and Imbens, 2016; Imai et al., 2021; DiTraglia et al., 2021). In this section I provide an alternative assumption, *independence of peers's types* (IPT), under which average potential outcomes do not depend on peers' compliance types. I then show that, under a generalization of the monotonicity assumption, average potential outcomes can be identified in the presence of two-sided noncompliance and spillovers in both outcomes and treatment statuses.

Suppose that 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, and the analysis is conducted for a given group size n_g.

Let $\mathbf{Z}_{(i)g}$ be the vector of unit i's peers' instruments, taking values $\mathbf{z}_g \in \{0,1\}^{n_g}$. To reduce the complexity of the model, 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. 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 following assumption collects the required conditions for the upcoming results.

Assumption 7 (Identification conditions for general n_g)

(a) Exclusion restriction: $Y_{ig}(d, \mathbf{d}_g, \mathbf{z}_g) = Y_{ig}(d, \mathbf{d}_g, \tilde{\mathbf{z}}_g)$ *for all* $\mathbf{z}_g, \tilde{\mathbf{z}}_g$.

(b) Independence: let $\mathbf{y}_{ig} = (Y_{ig}(d, \mathbf{d}_g))_{(d,\mathbf{d}_g)}$, $\mathbf{y}_{(i)g} = (\mathbf{y}_{jg})_{j\neq i}$, $\bar{\mathbf{d}}_{ig} = (D_{ig}(z, \mathbf{z}_g))_{(z,\mathbf{z}_g)}$ *and* $\bar{\mathbf{d}}_{(i)g} = (\bar{\mathbf{d}}_{jg})_{j\neq i}$. *Then,* $(\mathbf{y}_{ig}, \bar{\mathbf{d}}_{ig}, \mathbf{y}_{(i)g}, \bar{\mathbf{d}}_{(i)g}) \perp\!\!\!\perp (Z_{ig}, \mathbf{Z}_{(i)g})$.

(c) 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* $s_g = \mathbf{1}_g'\mathbf{d}_g$.

(d) Monotonicity:

  (i) $D_{ig}(z, w_g) \geq D_{ig}(z, w_g')$ *for* $w_g \geq w_g'$ *and* $z = 0, 1$,

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

Parts (a) and (b) in Assumption 7 generalize Assumptions 1 and 2 to the case of general group sizes. Part (c) imposes exchangeability as discussed above. Part (d) generalizes the monotonicity assumption requiring that treatment take-up becomes more likely as the

number of peers assigned to treatment increases, and that the effect of own assignment is stronger than the effect of peers' assignments.

Under monotonicity, 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, $\text{SC}(w^*)$, are units for whom $D_{ig}(1, w_g) = 1$ for all w_g, and for which there exists $0 < w^* < n_g$ 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, $\text{GC}(w^*)$ have $D_{ig}(0, w_g) = 0$ for all w_g and there exists $0 < w^* < n_g$ 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 $\xi_{ig} \in \Xi = \{\text{NT}, \text{GC}(w^*), \text{C}, \text{SC}(w^*), \text{AT} | w^* = 1, \ldots, n_g\}$ and $\boldsymbol{\xi}_{(i)g}$ the vector collecting ξ_{jg} for $j \neq i$. As before, let the event $AT_{ig} = \{\xi_{ig} = \text{AT}\}$, $C_{ig} = \{\xi_{ig} = \text{C}\}$, $GC(\omega^*) = \{\xi_{ig} = \text{GC}(\omega^*)\}$ and similarly for the other compliance types. Finally, 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 5 *Under Assumption 7,*

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

The following assumption restricts the amount of heterogeneity in potential outcomes by ensuring that potential outcomes are independent from peers' compliance types, conditional on own type.

Assumption 8 (Independence of peers' types) *Let* $\mathbf{y}_{ig} = (Y_{ig}(d, \mathbf{d}_g))_{(d, \mathbf{d}_g)}$. *Potential outcomes are independent of peers' types:* $\mathbf{y}_{ig} \perp\!\!\!\perp \boldsymbol{\xi}_{(i)g} \,|\, \xi_{ig}$.

Intuitively, IPT states that own type summarizes all the heterogeneity in potential outcome distributions. For example, this assumption may hold when groups are randomly formed, so that types are independent, or when types are perfectly correlated (i.e. units match with other units of the same type), so that conditional on own type, peers' types are constant.

The following result shows which average potential outcomes are identified under these assumptions.

Proposition 6 (Identification under IPT) *Under Assumptions 7 and 8, if* $\mathbb{P}[D_{ig} = d, Z_{ig} = z, S_{ig} = s, W_{ig} = w] > 0$*, then* $\mathbb{E}[Y_{ig}|D_{ig} = d, Z_{ig} = z, S_{ig} = s, W_{ig} = w] = \mathbb{E}[Y_{ig}(d, s)|D_{ig}(z, w) = d]$*. In particular, 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* $\xi_{ig} \neq \text{NT}$*, and* $\mathbb{E}[Y_{ig}(0, s)|\xi_{ig}]$ *is identified for all* s *and* $\xi_{ig} \neq \text{AT}$*.*

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

8 Conclusion

This paper proposed a potential outcomes framework to analyze identification and estimation of causal spillover effects using instrumental variables. The findings in this paper highlight the challenges of analyzing spillover effects with imperfect compliance and provide practical guidance on how to address them. First, when groups consist of pairs (such as couples, roommates, siblings), local average effects can be identified under one-sided noncompliance using 2SLS methods. One advantage of this approach is that one-sided noncompliance can be straightforwardly verified in practice. While 2SLS methods do not work in general under two-sided noncompliance or when units have multiple peers, the independence of peers' types assumption introduced in Section 7.2 provides an alternative restriction on effect heterogeneity that permits identification of causally interpretable parameters. Finally, Section 7.1 generalizes the results to cases in which the assumption of as-if random assignment of the instruments holds after conditioning on a set of covariates, which allows the researcher to apply the results in this paper in more general settings such as natural experiments.

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.

Andrews, I., Stock, J. H., and Sun, L. (2019), "Weak Instruments in Instrumental Variables Regression: Theory and Practice," *Annual Review of Economics*, 11, 727–753.

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.

Balke, A., and Pearl, J. (1997), "Bounds on Treatment Effects from Studies with Imperfect Compliance," *Journal of the American Statistical Association*, 92, 1171–1176.

Barrera-Osorio, F., Bertrand, M., Linden, L. L., and Perez-Calle, F. (2011), "Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia," *American Economic Journal: Applied Economics*, 3, 167–195.

Cattaneo, M. D., Jansson, M., and Ma, X. (forthcoming a), "Local Regression Distribution Estimators," *Journal of Econometrics*.

——— (forthcoming b), "lpdensity: Local Polynomial Density Estimation and Inference," *Journal of Statistical Software*.

DiTraglia, F. J., García-Jimeno, C., O'Keeffe-O'Donovan, R., and Sánchez-Becerra, A. (2021), "Identifying Causal Effects in Experiments with Spillovers and Non-Compliance," *working paper*.

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.

Dufour, J.-M., and Taamouti, M. (2005), "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73, 1351–1365.

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.

Garlick, R. (2018), "Academic Peer Effects with Different Group Assignment Policies: Residential Tracking versus Random Assignment," *American Economic Journal: Applied Economics*, 10, 345–69.

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.

Heckman, J. J., and Pinto, R. (2018), "Unordered Monotonicity," *Econometrica*, 86, 1–35.

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. (2021), "Causal Inference with Interference and Noncompliance in Two-Stage Randomized Experiments," *Journal of the American Statistical Association*, 116, 632–644.

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.

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 Noncompliance," *arXiv:1808.06418*.

Kitagawa, T. (2015), "A Test for Instrument Validity," *Econometrica*, 83, 2043–2063.

Mikusheva, A. (2010), "Robust confidence sets in the presence of weak instruments," *Journal of Econometrics*, 157, 236–247.

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. (forthcoming), "The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables," *American Economic Review*.

Rincke, J., and Traxler, C. (2011), "Enforcement Spillovers," *The Review of Economics and Statistics*, 93, 1224–1234.

Sacarny, A., Barnett, M. L., Le, J., Tetkoski, F., Yokum, D., and Agarwal, S. (2018), "Effect of Peer Comparison Letters for High-Volume Primary Care Prescribers of Quetiapine in Older and Disabled Adults: A Randomized Clinical Trial," *JAMA Psychiatry*, 10, 1003–1011.

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

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.

Titiunik, 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. (forthcoming), "Identification and Estimation of Spillover Effects in Randomized Experiments," *Journal of Econometrics*.