# Identification and Estimation of Spillover Effects in Randomized Experiments*

Gonzalo Vazquez-Bare†

July 23, 2020

Latest version here

**Abstract**

I study identification, estimation and inference for spillover effects in experiments where units' outcomes may depend on the treatment assignments of other units within a group. I show that the commonly-used linear-in-means regression identifies a weighted sum of spillover effects with some negative weights, and characterize the estimand that is recovered by a simple difference in means in the presence of spillovers. My results reveal the potential pitfalls of failing to flexibly account for spillover effects in policy evaluation. I then propose nonparametric estimators that overcome these issues and are consistent and asymptotically normal under a precise relationship between the number of parameters of interest, the total sample size and the treatment assignment mechanism. I show that these results can be used to rank assignment mechanisms for experimental design. These findings are illustrated using data from a conditional cash transfer program and with simulations. (*JEL* C10, C13, C14, C90)

*I am deeply grateful to Matias Cattaneo for continued advice and support. I am indebted to Lutz Kilian, Mel Stephens and Rocío Titiunik for thoughtful feedback and discussions that greatly improved the paper. I thank Clément de Chaisemartin, Catalina Franco, Amelia Hawkins, Nicolás Idrobo, Xinwei Ma, Nicolás Morales, Kenichi Nagasawa and Olga Namen for valuable discussions and suggestions. I also thank seminar participants at UChicago, UCLA, UC San Diego, University of Michigan, UC Santa Barbara, UChicago Harris School of Public Policy, Cornell University, UChicago Booth School of Business, UT Austin, Stanford University and UC Berkeley for helpful comments.

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

# 1 Introduction

Spillover effects, which occur when an agent's actions or behaviors indirectly affect other agents' outcomes through peer effects, social interactions, externalities or other types of interference, are pervasive in economics and social sciences. The widespread importance of this phenomenon across fields and disciplines has led to a rich literature focusing on social interactions (Manski, 1993; Brock and Durlauf, 2001; Graham, 2008; Manski, 2013), peer effects (Bramoullé, Djebbari, and Fortin, 2009; Epple and Romano, 2011; Sacerdote, 2014), networks (Graham, 2015; de Paula, 2017), games with multiple equilibria (de Paula, 2013; Kline and Tamer, 2019), design of experiments (Duflo and Saez, 2003; Hirano and Hahn, 2010; Baird, Bohren, McIntosh, and Özler, 2018), and causal inference (Hong and Raudenbush, 2006; Hudgens and Halloran, 2008; Tchetgen Tchetgen and VanderWeele, 2012; Halloran and Hudgens, 2016).

A thorough account of spillover effects is crucial to assess the causal impact of policies and programs (Abadie and Cattaneo, 2018; Athey and Imbens, 2017). However, the literature is still evolving in this area, and most of the available methods either assume no spillovers or allow for them in restrictive ways, without a precise definition of the parameters of interest or the conditions required to recover them. This paper studies identification and estimation of direct and spillover effects in randomized controlled trials, and offers three main contributions. First, I precisely define causal spillover effects and provide conditions to identify them. Section 2 sets up a causal potential-outcomes based framework that nests several models commonly used to analyze spillovers. Under the assumption that interference can occur within (but not between) the groups in which units are clustered, I define a rich set of direct and spillover treatment effects based on a function that links treatment assignments and outcomes. Nonparametric identification of all the treatment effects of interest is analyzed in Section 3. This framework highlights that the whole vector of direct and spillover effects can be identified regardless of the treatment assignment mechanism, as long as the assignments occur with non-zero probability. In Section 4, I analyze the parameters that are identified by the commonly used difference in means and linear-in-means regressions. The results reveal the potential pitfalls of failing to flexibly account for spillover effects in policy evaluation. I illustrate these findings by reanalyzing a conditional cash transfer program studied by Barrera-Osorio, Bertrand, Linden, and Perez-Calle (2011).

Second, I analyze nonparametric estimation and inference for spillover effects. In the presence of spillovers, estimation faces two main challenges: the number of treatment effects to estimate can be large, and the probability of observing units under different treatment assignments can be small. Section 5 provides general conditions that ensure uniform consistency and asymptotic normality of the direct and spillover effects estimators with special focus on the role of group size on estimation and inference. This approach formalizes the requirement of "many small groups" that is commonly invoked in the literature, and specifies the role that the number of parameters and the assignment mechanism play on the

asymptotic properties of nonparametric estimators. More precisely, consistency and asymptotic normality are shown under two main conditions that are formalized in the paper: (i) the number of treatment effects should not be "too large" with respect to the sample size, and (ii) the probability of each treatment assignment should not be "too small". These two requirements are directly linked to modeling assumptions on the potential outcomes, the choice of the set of parameters of interest and the treatment assignment mechanism. As an alternative approach to inference based on the normal approximation, the wild bootstrap is shown to be consistent, and simulation evidence suggests that it can yield better performance compared to the Gaussian approximation for moderately large groups.

The third main contribution is to show how these results can be used to guide the design of experiments to estimate spillover effects. Specifically, the rate of convergence of the spillover effects estimators and the rate of convergence of the distributional approximation are shown to depend on the treatment assignment mechanism, which gives a principled criterion to rank different procedures to assign the treatment. In Section 6 I demonstrate that a two-stage design that fixes the number of treated units in each group can improve the performance of the estimators in terms of inference, compared to simple random assignment, when groups are moderately large. I illustrate these findings with simulations in Section 7.

Finally, Section 8 discusses several implications for empirical work and points to future work in the analysis of spillover effects.

## 1.1 Related literature

Despite the longstanding and widespread interest across different disciplines, identification and estimation of spillover effects of programs and policies have proven a challenging problem. This subsection gives a brief description of some of the main approaches for analyzing spillovers; Section A1 in the supplemental appendix offers a more detailed review of the literature.

One strand of the literature builds on the linear-in-means (LIM) model, which has been the workhorse model for estimating peer effects in many areas of economics. Manski (1993) pointed out several identification problems in the LIM model. Since Manski's critique, the literature has offered several alternatives to deal with endogeneity issues in these models. The most credible ones rely on random assignment of peers (see Sacerdote, 2014, for a recent survey) or random assignment of a treatment (Lalive and Cattaneo, 2009; Bobonis and Finan, 2009; Dieye, Djebbari, and Barrera-Osorio, 2014).

Even in randomized contexts, identification in LIM models relies on the linearity assumption imposed on the structure of spillover effects. The parametric assumptions in the LIM models have been criticized for the unrealistic restrictions that they impose on the structure of peer effects (see Sacerdote, 2014). While some empirical specifications have attempted to relax parametric assumptions (Hoxby and Weingarth, 2005; Carrell, Fullerton, and West, 2009; Graham, 2008; Sacerdote, 2011, 2014), these models have only been analyzed from a

linear regression perspective; as such, the identified parameters can be interpreted as best linear predictors, but their causal interpretation remains unclear, and Angrist (2014) has criticized the usefulness of LIM models to recover causal effects. These limitations reflect the lack of a causal framework to analyze spillover effects. This paper contributes to this strand of the literature by providing a framework that does not rely on parametric assumptions for identification and estimation. In Section 4, I also characterize the estimand from the LIM model and provide conditions on potential outcomes to ensure that the LIM identifies a meaningful causal parameter.

In a second strand of the literature, researchers have conducted and analyzed experiments in which different units are assigned to treatment with varying probabilities, a design that Moffit (2001) called *partial population experiments*. A popular design in this setting is one in which groups of individuals (such as classrooms or households) are randomly divided into two categories, and then the treatment is randomized in one of the categories, leaving the other one as a pure control. This design was pioneered in an influential study by Duflo and Saez (2003), and later implemented in different versions by Miguel and Kremer (2004); Ichino and Schündeln (2012); Sinclair, McConnell, and Green (2012), Crépon, Duflo, Gurgand, Rathelot, and Zamora (2013), Beuermann, Cristia, Cueto, Malamud, and Cruz-Aguayo (2015), Beshears, Choi, Laibson, Madrian, and Milkman (2015) and Giné and Mansuri (2018), among others. Hirano and Hahn (2010) and Baird, Bohren, McIntosh, and Özler (2018) study experimental design under two-stage random assignment.

A common feature in the analysis of partial population experiments is that spillover effects are defined as comparisons between groups facing different probabilities of treatment. For example, Duflo and Saez (2003) define spillover effects as the average difference in outcomes between untreated units in treated groups and untreated units in pure control groups. This definition requires a specific experimental design. On the other hand, in the framework described in Section 2, spillover effects are defined based exclusively on potential outcomes, and have therefore a clear causal interpretation. These causal effects are shown to be identified under mild restrictions on the assignment mechanism without the need of any specific experimental design. Finally, Section 5 shows that two-stage designs can, under some conditions, significantly improve the performance of the nonparametric spillover effects estimators I recommend.

A third strand of the literature focuses on identification in games with social interactions or related strategic considerations (see e.g. Brock and Durlauf, 2001; de Paula, 2013; Blume, Brock, Durlauf, and Jayaraman, 2015; Kline and Tamer, 2019). This game-theoretic approach is sometimes used to justify the LIM model under some assumptions, and more generally highlights the key role of multiplicity of equilibria in this context. A related approach is provided by Manski (2013), who studies partial identification under different restrictions on the structural model, the response functions and the structure of social interactions. The relationship between reduced-form and structural response functions is discussed in Section A2 of the supplemental appendix. This paper complements this important strand of the

literature by offering point identification, estimation and inference results for well-defined causal (reduced-form) treatment effects in the presence of spillovers.

Finally, the literature in statistics and epidemiology focuses on causal inference and two-stage designs in a setting where potential outcomes are fixed and all randomness is generated by the assignment mechanism (see Halloran and Hudgens, 2016, for a recent review). This literature focuses mainly on constructing valid p-values, variance estimators and confidence intervals that are robust to interference (Rosenbaum, 2007; Hudgens and Halloran, 2008; Tchetgen Tchetgen and VanderWeele, 2012; Liu and Hudgens, 2014; Basse and Feller, 2018). A growing related literature studies interference in a setting that allows for more general network structures (Athey, Eckles, and Imbens, 2018; Choi, 2017). In this paper, I take a super-population approach under repeated sampling with an alternative asymptotic framework that complements the results available in this literature.

# 2 Setup

As a motivating example, consider the study by Barrera-Osorio, Bertrand, Linden, and Perez-Calle (2011). The authors conducted a pilot experiment designed to evaluate the effect of a conditional cash transfer program, *Subsidios Condicionados a la Asistencia Escolar*, in Bogotá, Colombia. The program aimed at increasing student retention and reducing drop-out and child labor. Eligible registrants ranging from grade 6-11 were randomly assigned to treatment and control.[1] The assignment was performed at the student level. In addition to administrative and enrollment data, the authors collected baseline and follow-up data from students in the largest 68 of the 251 schools. This survey contains attendance data and was conducted in the household. As shown in Table 1, 1,594 households have more than one registered children, and since the treatment was assigned at the child level, this gives variation in the number of treated children per household, as can be seen in Table 2.

Given the distribution of treated siblings within households, there are several reasons to expect spillover effects in this study. On the one hand, the cash transfer may alleviate a financial constraint that was preventing the parents from sending their children to school on a regular basis. The program could also help raise awareness on the importance of school attendance, encouraging parents to worry more about sending their children to school. In both these cases, untreated children may indirectly benefit from the program when they have a treated sibling. On the other hand, the program could create incentives for the parents to reallocate resources towards their treated children and away from their untreated siblings, decreasing school attendance for the latter. In all cases, ignoring spillover effects can severely underestimate the costs and benefits of this policy, as I show in Section 4. Moreover, these alternative scenarios have drastically different implications on how to assign the program

[1] The experiment had two different treatments that varied the timing of the payments, but, following the authors, I pool the two treatment arms to increase the sample size. See Barrera-Osorio, Bertrand, Linden, and Perez-Calle (2011) for details.

Table 1: Distribution of household size

| | Frequency |
|---|---|
| 1 | 5205 |
| 2 | 1410 |
| 3 | 168 |
| 4 | 15 |
| 5 | 1 |
| Total | 6799 |

Table 2: Treated per household

| | Frequency |
|---|---|
| 0 | 2355 |
| 1 | 3782 |
| 2 | 607 |
| 3 | 52 |
| 4 | 3 |
| Total | 6799 |

when scaling it up. In the first two situations, treating one child per household can be a cost-effective way to assign the treatment, whereas in the second case, treating all the children in a household can be more beneficial.

With these ideas in mind, the goal of this paper is to analyze conditions under which spillover effects can be precisely defined, identified and estimated.

## 2.1 Notation and parameters of interest

In light of the motivating example, consider a random sample of groups indexed by $g = 1, \ldots, G$, each with $n_g + 1$ units, so that each unit $i$ in group $g$ has $n_g$ neighbors or peers. I assume group membership is observable. Units in each group are assigned a binary treatment, and a unit's potential outcomes, defined in the next paragraph, can depend on the assignment of all other units in the same group. I refer to this phenomenon as *interference*, and to the effect of a neighbor's treatment assignment on unit $i$'s potential outcome as *spillover effect*. Interference is assumed to occur between units in the same group, but not between units in different groups.

The individual treatment assignment of unit $i$ in group $g$ is denoted by $D_{ig}$, taking values $d \in \{0, 1\}$, and the vector of treatment assignments in each group is given by $\mathbf{D}_g = (D_{1g}, \ldots, D_{n_g+1g})$. For each unit $i$, $D_{jig}$ is the treatment indicator corresponding to unit $i$'s $j$-th neighbor, collected in the vector $\mathbf{D}_{(i)g} = (D_{1ig}, D_{2ig}, \ldots, D_{n_gig})$. This vector takes values $\mathbf{d}_g = (d_1, d_2, \ldots, d_{n_g}) \in \mathcal{D} \subseteq \{0, 1\}^{n_g}$.

A key element in this setup will be a function $h_0(\cdot)$ that summarizes how the vector $\mathbf{d_g}$ enters the potential outcome. More precisely, define a function

$$h_0 : \mathcal{D} \to \mathcal{H}_0$$

that maps $\mathbf{d}_g$ into some value $h_0(\mathbf{d}_g)$ of the same or smaller dimension, so that $\dim(\mathcal{H}_0) \leq \dim(\mathcal{D})$. Following Manski (2013)'s terminology, for $h_0(\mathbf{d}_g) = \mathbf{h}_0$, I will refer to the tuple $(d, \mathbf{h}_0)$ as the *effective treatment assignment*, an element in the set $\{0, 1\} \times \mathcal{H}_0$. The function $h_0(\cdot)$ is sometimes known in the literature as an *exposure mapping*. The potential outcome for unit $i$ in group $g$ depends on the treatment assignment $(d, \mathbf{d}_g)$ through this mapping $h_0(\cdot)$ and is denoted by the random variable $Y_{ig}(d, \mathbf{h})$ where $\mathbf{h} = h_0(\mathbf{d}_g) \in \mathcal{H}^0$.

**Example 1 (SUTVA)** If $h_0(\cdot)$ is a constant function, the vector of peers' assignments is ignored and the set of effective treatment assignments becomes $\{0, 1\}$, so the potential outcomes do not depend on peers' assignments. In this case the only effective treatment assignments are $D_{ig} = 1$ and $D_{ig} = 0$ (treated and control). This assumption is often known as the stable unit treatment value assumption or SUTVA (Imbens and Rubin, 2015). □

**Example 2 (Exchangeability)** When potential outcomes depend on how many peers, but not which ones, are treated, peers are said to be *exchangeable*. Exchangeability can be modeled by setting $h_0(\mathbf{d}_g) = \mathbf{1}_g'\mathbf{d}_g$ so $h_0(\cdot)$ summarizes $\mathbf{d}_g$ through the sum of its elements (i.e. the number of peers assigned to treatment). The set of effective treatment assignments in this case is given by $\{(d, s) : d = 0, 1, s = 0, 1, \ldots, n_g\}$. The exchangeability assumption is very commonly invoked in the literature, and is further discussed in Section 4. □

**Example 3 (Stratified exchangeability)** Exchangeability may also be imposed by subgroups. For instance, the vector of assignments may be summarized by the number of male and female treated peers separately. In settings where units are geographically located, peers are commonly assumed to be exchangeable within groups defined by distance such as within one block, between one and two blocks, etc, or by different distance radiuses (e.g. within 100 meters, between 100 and 200 meters and so on). See Di Tella and Schargrodsky (2004) or Blattman, Green, Ortega, and Tobón (2019) for some empirical examples. □

**Example 4 (Reference groups)** When each unit interacts only with a strict subset of her peers, we can define for example $h_0(\cdot) : \{0, 1\}^{n_g} \to \{0, 1\}^{k_g}$ where $k_g < n_g$. For instance, under the assumption that each unit interacts with her two closest neighbors, $h_0(\mathbf{d}_g) = (d_1, d_2)$ so that $k_g = 2$. The subset of peers with which each unit interacts is known as the *reference group* (Manski, 2013). □

**Example 5 (Identity mapping)** The case in which $h_0(\mathbf{d}_g) = \mathbf{d}_g$ does not imply any restrictions on the structure of the potential outcomes, and therefore does not provide any dimensionality reduction. This level of generality requires the researcher to be able to determine an ordering between peers to determine who is unit $i$'s nearest neighbor, second nearest neighbor and so on. This ordering can be based for example on geographic distance, a spatial weights matrix as used in spatial econometrics, frequency of interaction in a social network, etc. □

In what follows, $\mathbf{0}_g$ and $\mathbf{1}_g$ will denote $n_g$-dimensional vectors of zeros and ones, respectively. Throughout the paper, I will assume that all the required moments of the potential outcomes are bounded. Unit-level direct effects are defined as differences in potential outcomes switching own treatment assignment for a fixed peer assignment $\mathbf{h}$, $Y_{ig}(1, \mathbf{h}) - Y_{ig}(0, \mathbf{h})$. Unit-level spillover effects are defined as differences in potential outcomes switching peer assignments for a fixed own assignment $d$, $Y_{ig}(d, \mathbf{h}) - Y_{ig}(d, \tilde{\mathbf{h}})$.

Given a vector of observed assignments $(D_{ig}, \mathbf{D}_{(i)g})$, the observed outcome is given by $Y_{ig}(D_{ig}, h_0(\mathbf{D}_{(i)g}))$ and can be rewritten as:

$$Y_{ig} = \sum_{d\in\{0,1\}} \sum_{\mathbf{h}\in\mathcal{H}^0} Y_{ig}(d, \mathbf{h}) \mathbb{1}(D_{ig} = d) \mathbb{1}(h_0(\mathbf{D}_{(i)g}) = \mathbf{h}).$$

To fix ideas, consider a household with three children, $n_g + 1 = 3$. In this household, each kid has two siblings, with assignments $d_1$ and $d_2$, so $\mathbf{d}_g = (d_1, d_2)$. If the true mapping $h_0(\cdot)$ is the identity mapping, the potential outcome has the form $Y_{ig}(d, d_1, d_2)$. This mapping does not restrict the potential outcomes, and hence each unit can have up to $2^{(n_g+1)} = 8$ different potential outcomes. In this case, $Y_{ig}(1, 0, 0) - Y_{ig}(0, 0, 0)$ is the effect of the treatment when both of unit $i$'s siblings are untreated, $Y_{ig}(0, 1, 0) - Y_{ig}(0, 0, 0)$ is the spillover effect on unit $i$ of treating unit $i$'s first sibling, and so on. The average effect of assignment $(d, d_1, d_2)$ compared to $(\tilde{d}, \tilde{d}_1, \tilde{d}_2)$ is thus given by $\mathbb{E}[Y_{ig}(d, d_1, d_2)] - \mathbb{E}[Y_{ig}(\tilde{d}, \tilde{d}_1, \tilde{d}_2)]$. On the other hand, under an exchangeable mapping, the potential outcome can be written as $Y_{ig}(d, s)$ where $s = 0, 1, 2$ is the number of treated siblings. The total number of different potential outcomes is $2(n_g + 1) = 6$, so exchangeability reduces the dimensionality of the effective assignments set from exponential to linear in group size.

I will assume perfect compliance, which means that all units receive the treatment they are assigned to. I analyze the case of imperfect compliance in Vazquez-Bare (2020). The following assumption establishes the relationship between the treatment and the potential outcomes.

**Assumption 1 (Random assignment)** *The potential outcomes* $\{Y_{ig}(d, \mathbf{h}), d \in \{0, 1\}, \mathbf{h} \in \mathcal{H}_0\}_{i=1}^{n_g}$ *are identically distributed across* $i$ *and* $g$*, and for each* $g$*,*

$$\{Y_{ig}(d, \mathbf{h}), d \in \{0, 1\}, \mathbf{h} \in \mathcal{H}_0\}_{i=1}^{n_g} \perp\!\!\!\perp \mathbf{D}_g.$$

The first part of Assumption 1 states that potential outcomes are identically distributed, and thus moments are not indexed by either $i$ or $g$. The second part states that the treatment is randomly assigned within each group.

In practice, the true mapping $h_0(\cdot)$ is usually unknown, and the researcher needs to posit a mapping $h(\cdot)$ that may or may not coincide with $h_0(\cdot)$. Given the lack of knowledge on the true mapping, a mapping $h(\cdot)$ that imposes fewer restrictions on the potential outcomes has a lower risk of misspecification. To formalize this idea, I introduce the following definition.

**Definition 1 (Coarseness)** *Given two mappings* $h(\cdot) : \mathcal{D} \to \mathcal{H}$ *and* $\tilde{h}(\cdot) : \mathcal{D} \to \tilde{\mathcal{H}}_g$*, we say* $h(\cdot)$ *is coarser than* $\tilde{h}(\cdot)$ *if there exists another function* $f(\cdot) : \tilde{\mathcal{H}}_g \to \mathcal{H}$ *such that* $h(\mathbf{d}_g) = f \circ \tilde{h}(\mathbf{d}_g)$ *for all* $\mathbf{d}_g \in \mathcal{D}$*.*

Intuitively, this means that $h(\cdot)$ gives a "cruder" summary of $\mathbf{d}_g$ (i.e. discards more information) than $\tilde{h}(\cdot)$. In other words, a coarser mapping imposes more restrictions on the potential outcomes. For example, the exchangeable mapping $h(\mathbf{d}_g) = \mathbf{1}_g'\mathbf{d}_g$ is coarser than

the identity mapping $\tilde{h}(\mathbf{d}_g) = \mathbf{d}_g$, and the reference group mapping $h(\mathbf{d}_g) = (d_1, d_2)$ is coarser than the mapping $\tilde{h}(\mathbf{d}_g) = (d_1, d_2, d_3, d_4)$.

The next section addresses identification of average potential outcomes when the true mapping $h_0(\cdot)$ is unknown.

# 3 Identification

In what follows, let $\mathbf{H}_{ig} = h(\mathbf{D}_{(i)g})$ be the observed value of the chosen exposure mapping, and let $\mathbf{H}^0_{ig} = h_0(\mathbf{D}_{(i)g})$. The following result links observed outcomes, potential outcomes and exposure mappings.

**Lemma 1 (Identification)** *Suppose Assumption 1 holds and let $h_0(\cdot) : \mathcal{D} \to \mathcal{H}_0$ be the true exposure mapping. Given a mapping $h(\cdot) : \mathcal{D} \to \mathcal{H}$, for any pair $(d, \mathbf{h}) \in \{0, 1\} \times \mathcal{H}$ such that $\mathbb{P}[D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}] > 0$,*

$$\mathbb{E}[Y_{ig}|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}] = \sum_{\mathbf{h}_0 \in \mathcal{H}_0} \mathbb{E}[Y_{ig}(d, \mathbf{h}_0)]\mathbb{P}[\mathbf{H}^0_{ig} = \mathbf{h}_0|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}].$$

*In particular, if $h_0(\cdot)$ is coarser than $h(\cdot)$, then*

$$\mathbb{E}[Y_{ig}|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}] = \mathbb{E}[Y_{ig}(d, h_0(\mathbf{h}))].$$

Lemma 1 shows that the average observed outcome among units facing $D_{ig} = d$ and $\mathbf{H}_{ig} = \mathbf{h}$ averages the potential outcomes over all the assignments $\mathbf{h}_0$ that are consistent with $(D_{ig}, \mathbf{H}_{ig}) = (d, \mathbf{h})$, as long as the probability of $(d, \mathbf{h})$ is not zero. For example, consider the previous example with three units and where the true mapping is the identity mapping so the potential outcome has the form $Y_{ig}(d, d_1, d_2)$. Suppose we posit an exchangeable mapping $h(\mathbf{d}_g) = \mathbf{1}'_g\mathbf{d}_g$ and thus $\mathbf{H}_{ig} = S_{ig} = \sum_{j \neq i} D_{jg}$ which is a scalar counting how many of unit $i$'s peers are treated. By Lemma 1, if $\mathbb{P}[D_{ig} = 0, S_{ig} = 1] > 0$, $\mathbb{E}[Y_{ig}|D_{ig} = 0, S_{ig} = 1]$ equals a weighted average of $\mathbb{E}[Y_{ig}(0, 1, 0)]$ and $\mathbb{E}[Y_{ig}(0, 0, 1)]$, with weights given by the conditional probabilities of these different assignments. In general, $(D_{ig}, \mathbf{H}_{ig}) = (d, \mathbf{h})$ may be consistent with many different effective assignments $\mathbf{h}_0$. When $h_0(\cdot)$ is coarser than $h(\cdot)$, however, the value of $\mathbf{h_0}$ is uniquely determined. In such cases, the second part of Lemma 1 shows that $\mathbb{E}[Y_{ig}|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}]$ identifies the value of the average potential outcome consistent with that assignment. For example, suppose that the true mapping is exchangeable, so that outcomes have the form $Y_{ig}(d, s)$ with $s = 0, 1, 2$, and suppose we posit an identity mapping $h(\mathbf{d}_g) = \mathbf{d}_g = (d, d_1, d_2)$. Setting $\mathbf{H}_{ig} = (0, 1, 1)$ implies that the sum of treated peers is equal to 2, and therefore $\mathbb{E}[Y_{ig}|D_{ig} = 0, \mathbf{H}_{ig} = (1, 1)] = \mathbb{E}[Y_{ig}(0, 2)]$. In particular, this result implies that if the mapping $h(\cdot)$ is correctly specified, $\mathbb{E}[Y_{ig}|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}] = \mathbb{E}[Y_{ig}(d, \mathbf{h})]$.

**Remark 1 (Overidentification)** When $h_0(\cdot)$ is coarser than $h(\cdot)$, there may be multiple values of $\mathbb{E}[Y_{ig}|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}]$ that equal the same value of the average potential outcome

$\mathbb{E}[Y_{ig}(d, h_0(\mathbf{h}))]$. In the example above, if $h(\cdot)$ is the identity mapping but the true mapping is exchangeable, we have that $\mathbb{E}[Y_{ig}|D_{ig} = 0, \mathbf{H}_{ig} = (1,0)] = \mathbb{E}[Y_{ig}|D_{ig} = 0, \mathbf{H}_{ig} = (0,1)] = \mathbb{E}[Y_{ig}(0,1)]$. This fact can be used to test, to some extent, hypotheses about the true unknown mapping $h_0(\cdot)$. For instance, $\mathbb{E}[Y_{ig}|D_{ig} = 0, \mathbf{H}_{ig} = (1,0)] \neq \mathbb{E}[Y_{ig}|D_{ig} = 0, \mathbf{H}_{ig} = (0,1)]$ would suggest that the true mapping is not exchangeable. Note, however, that failing to reject this hypothesis does not imply that exchangeability holds, as exchangeability could hold for $\mathbb{E}[Y_{ig}(d, \mathbf{d}_g)]$ but not for $Y_{ig}(d, \mathbf{d}_g)$ or other moments. □

**Remark 2 (Pooled estimands)** Coarse exposure mappings can be used not only as a modeling assumption on potential outcomes but also as summary measures of average potential outcomes and treatment effects. This may be particularly useful when the true mapping $h_0(\cdot)$ is suspected to be high-dimensional. For instance, setting $h(\cdot)$ equal to a constant function, which ignores $\mathbf{d}_g$, averages over all possible peers' assignments: $\mathbb{E}[Y_{ig}|D_{ig} = d] = \sum_{\mathbf{h}_0} \mathbb{E}[Y_{ig}(d, \mathbf{h}_0)]\mathbb{P}[\mathbf{H}^0_{ig} = \mathbf{h}_0|D_{ig} = d]$. As a less extreme example, let $s = \mathbf{1}_g'\mathbf{d}_g$, and define $h(\mathbf{d}_g) = \mathbb{1}(s > 0)$ which equals one if there is at least one treated peer. Let $S_{ig} = \sum_{j \neq i} D_{jg}$ be the observed number of treated peers for unit $i$. Then, by Lemma 1, $\mathbb{E}[Y_{ig}|D_{ig} = d, S_{ig} > 0] = \sum_{\mathbf{h}_0} \mathbb{E}[Y_{ig}(d, \mathbf{h}_0)]\mathbb{P}[\mathbf{H}^0_{ig} = \mathbf{h}_0|D_{ig} = d, S_{ig} > 0]$. Consider the difference between untreated units with at least one treated peer and untreated units with no treated peers:

$$\Delta = \mathbb{E}[Y_{ig}|D_{ig} = 0, S_{ig} > 0] - \mathbb{E}[Y_{ig}|D_{ig} = 0, S_{ig} = 0].$$

Then, given that $S_{ig} = 0$ implies that $\mathbf{D}_{(i)g} = \mathbf{0}_g$, we have that:

$$\Delta = \sum_{\mathbf{h}_0} \mathbb{E}[Y_{ig}(0, \mathbf{h}_0) - Y_{ig}(0, \mathbf{0})]\mathbb{P}[\mathbf{H}^0_{ig} = \mathbf{h}_0|D_{ig} = 0, S_{ig} > 0]$$

where $\mathbf{0} = h_0(\mathbf{0}_g)$. Thus, $\Delta$ recovers a weighted average of spillover effects on untreated units weighted by the probabilities of the different assignments, a well defined causal parameter. A natural generalization of this idea is to split $S_{ig}$ into categories such as $S_{ig} = 0$, $1 \leq S_{ig} \leq m$, $m + 1 \leq S_{ig} \leq n_g$ and so on. □

**Remark 3 (Partial population experiments)** A popular design when analyzing spillover effects is the partial population design (Moffit, 2001; Baird, Bohren, McIntosh, and Özler, 2018; Duflo and Saez, 2003). In its simplest form, groups are randomly divided into treated and controls based on binary indicator $T_g$. Then, within the groups with $T_g = 1$, treatment $D_{ig}$ is randomly assigned at the individual level. In these type of experiments, spillover effects are often estimated as the average difference between control units in treated groups and control units in pure control groups,

$$\Delta_{\mathsf{PP}} = \mathbb{E}[Y_{ig}|D_{ig} = 0, T_g = 1] - \mathbb{E}[Y_{ig}|T_g = 0].$$

For recent examples of this or similar strategies, see Ichino and Schündeln (2012); Sinclair,

McConnell, and Green (2012), Beuermann, Cristia, Cueto, Malamud, and Cruz-Aguayo (2015), Beshears, Choi, Laibson, Madrian, and Milkman (2015) and Giné and Mansuri (2018), among others. Redefining the vector of treatment assignments as $(D_{ig}, \mathbf{D}_{(i)g}, T_g) = (d, \mathbf{d}_g, t)$ and setting $h(\mathbf{d}_g, t) = t$, if $(\mathbf{D}_g, T_g)$ is independent of potential outcomes, then Lemma 1 implies that:

$$\Delta_{\mathsf{PP}} = \sum_{\mathbf{h}_0} \mathbb{E}[Y_{ig}(0, \mathbf{h}_0) - Y_{ig}(0, \mathbf{0})]\mathbb{P}[\mathbf{H}_{ig}^0 = \mathbf{h}^0 | D_{ig} = 0, T_g = 1]$$

which averages over all the possible number of treated peers that an untreated unit can have in a treated group. The generalization to experiments with more than two categories (see e.g. Crépon, Duflo, Gurgand, Rathelot, and Zamora, 2013) is straightforward. □

# 4 Comparison with existing methods

Using the results from the previous section as a benchmark, in this section I analyze the performance of two of the most common strategies used in experiments: the difference in means and the linear-in-means regression. I illustrate these ideas with the data from the experiment described in Section 2. I analyze direct and spillover effects restricting the sample to households with three registered siblings, which gives a total of 168 households and 504 observations. The outcome of interest will be school attendance.

To facilitate the comparison with the existing literature and simplify the exposition, in this section I will assume that exchangeability holds as in Example 2, and hence potential outcomes depend on the vector of peers' assignments only through the total number of treated peers, $Y_{ig}(d, s)$. This assumption is imposed here only for brevity and convenience, and Section A6 in the supplemental appendix generalizes all the results in this section to arbitrary exposure mappings.

The exchangeability assumption is ubiquitous both in the applied and theoretical literature (see e.g. Hudgens and Halloran, 2008; Manski, 2013; Ferracci, Jolivet, and van den Berg, 2014; Baird, Bohren, McIntosh, and Özler, 2018). In particular, the requirement that potential outcomes depend only on the number (or proportion) of treated neighbors is a key assumption in linear-in-means models (Manski, 1993; Moffit, 2001; Bramoullé, Djebbari, and Fortin, 2009; Lalive and Cattaneo, 2009; Bobonis and Finan, 2009; Dieye, Djebbari, and Barrera-Osorio, 2014). Section 4.2 analyzes this estimation strategy in detail.

In this application, exchangeability may be a reasonable assumption as parents can be expected to make schooling decisions based on how many of their children are treated (for example, to determine whether the cash transfer covers the direct and opportunity costs of sending their kids to school), regardless of which of their children are treated, especially so given that all eligible children are similar in age and often attend the same schools. Nevertheless, Section A5 in the supplemental appendix provides additional empirical results that do not impose this assumption, and also a test of exchangeability.

Letting $\mu(d,s) = \mathbb{E}[Y_{ig}(d,s)]$, the goal is to estimate the direct effects:

$$\tau(s) = \mathbb{E}[Y_{ig}(1,s) - Y_{ig}(0,s)] = \mu(1,s) - \mu(0,s)$$

and the spillover effects:

$$\theta_d(s) = \mathbb{E}[Y_{ig}(d,s) - Y_{ig}(d,0)] = \mu(d,s) - \mu(d,0)$$

where $s = 0, 1, 2$.

The parameters of interest are estimated through the following regression:

$$\mathbb{E}[Y_{ig}|D_{ig}, S_{ig}] = \alpha + \tau(0)D_{ig} + \sum_{s=1}^{n_g} \theta_0(s)\mathbb{1}(S_{ig} = s)(1 - D_{ig}) + \sum_{s=1}^{n_g} \theta_1(s)\mathbb{1}(S_{ig} = s)D_{ig} \quad (1)$$

where $\alpha = \mu(0,0)$. Because it is equivalent to estimating averages at each cell separately, Equation (1) does not impose any parametric assumptions. The total number of parameters in this regression is $2(n_g + 1) = 6$, so the number of coefficients equals the number of average potential outcomes to estimate.

The estimates from Equation (1) are shown in the right panel of Table 3. These estimates suggest a positive direct effect of the treatment of 16.4 percentage points, significant at the 5 percent level, with almost equally large spillover effects on the untreated units. More precisely, the estimated effect on an untreated kid of having one treated sibling is 14.6 percentage points, while the effect of having two treated siblings is 14 percentage points. The fact that we cannot reject the hypothesis that $\theta_0(1) = \theta_0(2)$ suggests some form of crowding-out: given that one sibling is treated, treating one more sibling does not affect attendance. These facts could be consistent with the idea that, for example, the conditional cash transfer alleviates some financial constraint that was preventing the parents from sending their children to school regularly, or with the program increasing awareness on the importance of school attendance, since in these cases the effect occurs as soon as at least one kid in the household is treated, and does not amplify with more treated kids.

On the other hand, spillover effects on treated children are much smaller in magnitude and negative. Notice that the fact that these estimates are negative does not mean that the program hurts treated children, but that treating more siblings reduces the benefits of the program. For example, the effect of being treated with two treated siblings, compared to nobody treated, can be written as $\mu(1,2) - \mu(0,0) = \mu(1,0) - \mu(0,0) + \mu(1,2) - \mu(1,0) = \tau(0) + \theta_1(2)$, so it can be estimated by $\hat{\tau}(0) + \hat{\theta}_1(2) \approx 0.113$. Thus, a treated kid with two treated siblings increases her attendance in 11 percentage points starting from a baseline in which nobody in the household is treated.

In all, the estimates suggest large and positive direct and spillover effects on the untreated, with some evidence of crowding-out between treated siblings. In addition to these results, Table A1 in the supplemental appendix shows the estimates when relaxing exchangeability

and defining sibling identities based on difference in ages. The results and the formal test of equality of coefficients suggest that exchangeability cannot be rejected in this case.

## 4.1 Difference in means

The above results can be used to understand how some specifications commonly used in empirical studies perform in this type of contexts. Suppose initially that the experiment was analyzed using a difference in means between treated and controls, ignoring the presence of spillovers. The left panel of Table 3 shows the difference in means, which is the estimator that is used when spillovers are ignored, usually calculated as the OLS estimator for $\beta_{\mathsf{D}}$ in the model:

$$Y_{ig} = \alpha_{\mathsf{D}} + \beta_{\mathsf{D}} D_{ig} + u_{ig}. \tag{2}$$

The results show that the difference in means is close to zero and not significant. Hence, by ignoring the presence of spillover effects, a researcher estimating the effect of the program in this way would conclude that the treatment has no effect. This finding captures the intuition that in the presence of spillovers, the "contamination" of the control group pushes the difference between treated and controls towards zero. The following result formalizes this finding.

**Lemma 2 (Difference in means)** *Under Assumption 1, the coefficient $\beta_D$ from Equation (2) can be written as:*

$$\beta_{\mathsf{D}} = \tau(0) + \sum_{s=1}^{n_g} \theta_1(s)\mathbb{P}[S_{ig} = s|D_{ig} = 1] - \sum_{s=1}^{n_g} \theta_0(s)\mathbb{P}[S_{ig} = s|D_{ig} = 0]$$

Hence, the (population) difference in means equals the direct effect without treated siblings plus the difference in weighted averages of spillover effects under treatment and under control. A common treatment assignment mechanism is simple random assignment. Under this mechanism, the treatment is assigned independently and with the same probability to each unit in the sample. In this case, the above expression reduces to:

$$\beta_{\mathsf{D}} = \tau(0) + \sum_{s=1}^{n_g} (\theta_1(s) - \theta_0(s))\mathbb{P}[S_{ig} = s]$$

The effect of the presence of spillovers in the difference in means, captured by the term $\sum_{s=1}^{n_g}(\theta_1(s) - \theta_0(s))\mathbb{P}[S_{ig} = s]$, is undetermined in general, and it could be positive, negative or zero depending on the relative magnitudes of the spillover effects under treatment and control. If all the spillover effects are equal under treatment and control, $\theta_0(s) = \theta_1(s)$ for all $s$, then the difference in means $\beta_{\mathsf{D}}$ equals the direct effect of the treatment without treated siblings, $\tau(0)$. On the other hand, if all the spillovers under treatment are zero and the spillovers under control have the same sign as the direct effects, the spillover effects will drive the difference in means towards zero, which captures the idea of "contamination" of

the control group.

From Table 3, the estimated spillover effects in this case are larger under control that under treatment, and have different signs, so $\hat{\theta}_1(s) - \hat{\theta}_0(s) < 0$. Therefore, the spillover effects push the difference in means towards zero in this case.

## 4.2 Linear-in-means models

Equation (2) may give an incomplete assessment of the effect of a program because it completely ignores the presence of spillovers. When trying to explicitly estimate spillover effects, a common strategy is to estimate a reduced-form linear-in-means model (see e.g. Lalive and Cattaneo, 2009; Bobonis and Finan, 2009; Dieye, Djebbari, and Barrera-Osorio, 2014), which is given by:

$$Y_{ig} = \alpha_\ell + \beta_\ell D_{ig} + \gamma_\ell \bar{D}_g^{(i)} + \eta_{ig}, \qquad \bar{D}_g^{(i)} = \frac{1}{n_g} \sum_{j \neq i} D_{jg} \tag{3}$$

that is, a regression of the outcome on a treatment indicator and the proportion of treated neighbors. In this specification, $\gamma_\ell$ is usually seen as a measure of spillover effects, since it captures the average change in outcomes in response to a change in the proportion of treated neighbors.

The estimates from Equation (3) are given in the first column of the middle panel in Table 3. The estimates suggest slightly negative and not significant direct and spillover effects, substantially different from results using Equation (1). To better understand this point, Equation (1) suggests the assumptions required for a LIM model to be correctly specified. In particular, we can see that if (i) the spillover effects are equal under treatment and control, $\theta_0(s) = \theta_1(s) = \theta(s)$ for all $s$ and (ii) the spillover effects are linear in $s$, that is, $\theta(s) = \kappa s$ for some constant $\kappa$, then Equation (1) reduces to:

$$\mathbb{E}[Y_{ig} | D_{ig}, S_{ig} = s] = \alpha + \tau(0) D_{ig} + \theta(n_g) \bar{D}_g^{(i)}$$

so that $\gamma_\ell = \theta(n_g)$ and thus the coefficient on the proportion of treated neighbors recovers the spillover effect of treating all neighbors (and the remaining effects can be obtained using linearity of the spillovers). However, the spillover effect estimates in Table 3 suggest that the LIM assumptions may not hold in this case. More generally, the following result holds.

**Lemma 3 (LIM regression)** *Under Assumption 1 and simple random assignment, the coefficient $\gamma_\ell$ from Equation (3) can be written as:*

$$\gamma_\ell = n_g \sum_{s=1}^{n_g} [\theta_0(s)(1-p) + \theta_1(s) p] \left( \frac{s - \mathbb{E}[S_{ig}]}{\mathbb{V}[S_{ig}]} \right) \mathbb{P}[S_{ig} = s]$$

*where $p = \mathbb{P}[D_{ig} = 1]$.*

This results shows that $\gamma_\ell$ captures a rather complicated linear combination of all the spillover effects under treatment and control. More precisely, $\gamma_\ell$ first averages the spillover

effects under treatment and control, $\theta_0(s)(1-p)+\theta_1(s)p$, and then combines all these terms. Importantly, the "weights" assigned to each of the terms $\theta_0(s)(1-p)+\theta_1(s)p$ are not bounded between zero and one, and they sum to zero. In fact, these weights are negative for all values $s$ below the mean of $S_{ig}$, and positive for all the values above. In this case, we have that $\hat{\gamma}_\ell$ will assign negative weight to the first term and positive weight to the second one.

A straightforward way to make Equation (3) more flexible is to include an interaction between $D_{ig}$ and $\bar{D}_g^{(i)}$ to allow for the spillover effects to be different under treatment and control:

$$Y_{ig} = \alpha_\ell + \beta_\ell D_{ig} + \gamma_\ell^0 \bar{D}_g^{(i)}(1 - D_{ig}) + \gamma_\ell^1 \bar{D}_g^{(i)} D_{ig} + \xi_{ig} \tag{4}$$

The third column of the middle panel in Table 3 shows that the estimates for the spillover effects for treated and control are actually quite close to the estimates from the full model, which could suggest that this strategy can be a good approximation to the true spillover effects. However, in this case we have the following result.

**Lemma 4 (Interacted LIM regression)** *Under Assumption 1 and simple random assignment, for $d = 0, 1$ the coefficients $\gamma_\ell^d$ from Equation (4) can be written as:*

$$\gamma_\ell^d = n_g \sum_{s=1}^{n_g} \theta_d(s) \left( \frac{s - \mathbb{E}[S_{ig}]}{\mathbb{V}[S_{ig}]} \right) \mathbb{P}[S_{ig} = s]$$

Thus, the only difference is that each $\gamma_\ell^d$ combines the spillover effects under a fixed treatment status $d$, instead of averaging $\theta_0(s)$ and $\theta_1(s)$. As before, this expression shows that the coefficients $\gamma_\ell^d$ are not weighted averages of the spillover effects $\theta_d(s)$. More precisely, they assign negative weights to the parameters $\theta_d(s)$ with $s$ below $\mathbb{E}[S_{ig}]$ and positive weights when $s$ is above $\mathbb{E}[S_{ig}]$. Hence, these coefficients will not in general lie between the true spillover effects.

## 4.3 Pooled effects

Finally, I illustrate how to estimate pooled effects by averaging over the possible number of treated siblings (2 and 3 in this case). For this, I estimate the following regression:

$$Y_{ig} = \alpha_p + \beta_p D_{ig} + \gamma_p^0 \mathbb{1}(S_{ig} > 0)(1 - D_{ig}) + \gamma_p^1 \mathbb{1}(S_{ig} > 0) D_{ig} + \nu_{ig}$$

where

$$\gamma_p^d = \sum_{s=1}^{2} \theta_d(s) \mathbb{P}[S_{ig} = s | D_{ig} = d, S_{ig} > 0]$$

(see Remark 2). From Table 3 we can see that the estimated pooled spillover effects are 0.144 for controls and $-0.045$ for treated, which as expected lie between the effects found with the saturated regression. These results illustrate how this type of pooling can provide a useful summary of spillover effects, which may be a feasible alternative when the total number of

Table 3: Estimation results

| | **Diff. Means** | | **Linear-in-Means** | | | | **Full** | | **Pooled** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | coef | s.e. | coef | s.e. | coef | s.e. | coef | s.e. | coef | s.e. |
| $D_{ig}$ | 0.006 | 0.016 | 0.007 | 0.016 | 0.102** | 0.042 | 0.164** | 0.066 | 0.165** | 0.065 |
| $\bar{D}_g^{(i)}$ | | | 0.027 | 0.034 | | | | | | |
| $\bar{D}_g^{(i)}(1-D_{ig})$ | | | | | 0.169** | 0.068 | | | | |
| $\bar{D}_g^{(i)}D_{ig}$ | | | | | -0.064 | 0.039 | | | | |
| $\mathbb{1}(S_{ig}=1)(1-D_{ig})$ | | | | | | | 0.146** | 0.066 | | |
| $\mathbb{1}(S_{ig}=2)(1-D_{ig})$ | | | | | | | 0.14** | 0.056 | | |
| $\mathbb{1}(S_{ig}=1)D_{ig}$ | | | | | | | -0.041* | 0.023 | | |
| $\mathbb{1}(S_{ig}=2)D_{ig}$ | | | | | | | -0.051** | 0.025 | | |
| $\mathbb{1}(S_{ig}>0)(1-D_{ig})$ | | | | | | | | | 0.144** | 0.06 |
| $\mathbb{1}(S_{ig}>0)D_{ig}$ | | | | | | | | | -0.045** | 0.02 |
| Constant | 0.822*** | 0.013 | 0.811*** | 0.024 | 0.756*** | 0.037 | 0.706*** | 0.057 | 0.705*** | 0.057 |
| Observations | | 504 | | 504 | | 504 | | 504 | | 504 |

**Notes**: S.e. clustered at the household level. Regressions include school FE. $^{***}p<0.01,^{**}p<0.05,^{*}p<0.1$.

spillover effects is too large to estimate them separately. I address this issue in detail in the next section.

# 5 Estimation and inference

The previous sections show that, under random assignment of the treatment, all the parameters of interest are nonparametrically identified as long as the required probabilities are non-zero, and suggest a straightforward way to estimate them as sample means in each assignment cell. The main challenge of this strategy arises when groups are large. A large number of units per group requires estimating a large number of means in each of the cells defined by the treatment assignments. When groups have many units (as in households with many siblings or classrooms with a large number of students), the probability of observing some assignments can be close to zero and the number of observations in each cell can be too small to estimate the average potential outcomes.

For example, suppose the treatment is assigned as an independent coin flip with probability $p = 1/2$. Under this assignment we would expect most groups to have about half its units treated, so when groups have, say, 10 units, 5 of them would be treated on average. The probability of observing groups with zero or all treated units, on the other hand, will be close to zero, and thus the average potential outcomes corresponding to these "tail assignments" will be very hard to estimate.

So far, the analysis has been done taking group size as fixed. When group size is fixed, small cells are a finite sample problem that disappears as the sample grows. To account for this phenomenon asymptotically, in this section I will generalize this setting to allow group size to grow with the sample size. The goal is to answer the question of how large groups can be relative to the total sample size to allow for valid estimation and inference. The key issue to obtain consistency and asymptotic normality will be to ensure that the number of observations in all cells grows to infinity as the sample size increases. This setup is not intended to model a population in which groups are effectively infinitely large, but as a statistical device to approximate the distribution of estimators in a finite sample when the number of parameters can be "moderately" large, in a sense that will be made more precise in this section. The case with fixed group size is a particular case in this setting.

In this section I will assume that groups are equally sized, so that $n_g = n$. Recall that given an exposure mapping $h(\cdot)$ and $\mathbf{h} = h(\mathbf{d}_g)$, the effective treatment assignments are given by $(d, \mathbf{h}_g) \in \{0, 1\} \times \mathcal{H}$. As formalized in Assumption 2 below, $h(\cdot)$ is not assumed to equal the true mapping, but the true mapping $h_0(\cdot)$ has to be coarser than $h(\cdot)$ as specified in Definition 1. To make the notation more compact, I will let $\mathcal{A}_n = \{0, 1\} \times \mathcal{H}$ where the notation makes the dependence of this set on the group size explicit. The cardinality of this set is denoted by $|\mathcal{A}_n|$, which indicates the total number of parameters to be estimated. The observed effective treatment assignments will be $(D_{ig}, \mathbf{H}_{ig}) = \mathbf{A}_{ig}$, taking values $\mathbf{a} \in \mathcal{A}_n$, and $\mu(\mathbf{a}) = \mathbb{E}[Y_{ig}|\mathbf{A}_{ig} = \mathbf{a}]$.

Each treatment assignment mechanism determines a distribution $\pi(\cdot)$ over $\mathcal{A}_n$ where $\pi(\mathbf{a}) = \mathbb{P}[\mathbf{A}_{ig} = \mathbf{a}]$ for $\mathbf{a} \in \mathcal{A}_n$. For example, when $\mathcal{A}_n = \{0, 1\}$, if the treatment is assigned independently as a coin flip, $\pi(1) = \mathbb{P}[D_{ig} = 1] = p$ and $\pi(0) = 1 - p$. Under the same assignment, with an exchangeable exposure mapping, $\pi(d, s) = \mathbb{P}[D_{ig} = d, S_{ig} = s] = \binom{n}{s} p^{s+d}(1-p)^{n+1-s-d}$. Importantly, since the size of the set $\mathcal{A}_n$ can increase with group size, the probabilities $\pi(\mathbf{a})$ can shrink towards zero for some (or all) assignments $\mathbf{a} \in \mathcal{A}_n$, as illustrated in the previous example. Finally, define:

$$\underline{\pi}_n = \min_{\mathbf{a} \in \mathcal{A}_n} \pi(\mathbf{a})$$

which is the probability of the least likely treatment assignment. This probability, together with the total sample size, will determine the number of observations in the smallest assignment cell, that is, the number of observations available to estimate the "hardest" average potential outcome.

Let $\mathbf{A}_g = (\mathbf{A}_{1g}, \ldots, \mathbf{A}_{n_g+1,g})$, $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_G)$, and $\mathbf{Y}_g = (Y_{1g}, Y_{2g}, \ldots Y_{n_g+1,g})'$. I will assume the following.

**Assumption 2 (Sampling and design)**

*(i)* *For $g = 1, \ldots, G$, $(\mathbf{Y}_g', \mathbf{A}_g')$ are iid, and $n_g = n$.*

*(ii)* *The true mapping $h_0(\cdot)$ is coarser than $h(\cdot)$.*

*(iii)* *The potential outcomes are independent across $i$ within groups.*

*(iv)* *$|\mathcal{A}_n| = O(G(n+1)\underline{\pi}_n)$, as $G \to \infty$ and $n \to \infty$.*

Part (i) in Assumption 2 states that the researcher has access to a sample of $G$ independent and identically distributed equally-sized groups. When groups have different sizes (for example, households with 3, 4 or 5 siblings), the analysis can be performed separately for each group size. Section A3 of the supplemental appendix further discusses the case of unequally-sized groups. Part (ii) allows the posited mapping $h(\cdot)$ to be different from the true mapping, but requires it to be flexible enough to break the dependence between $Y_{ig}$ and $\mathbf{A}_{jg}$ conditional on $\mathbf{A}_{ig}$ for $j \neq i$. Part (iii) assumes that potential outcomes are independent within groups, and hence the only source of dependence between the observed outcomes is the assignment $\mathbf{A}_g$. In the setting with growing group sizes, this assumption can be relaxed at the expense of introducing more restrictions on the within-group network structure (for example, assuming the rate at which links are formed is slow enough). On the other hand, this condition is easily relaxed to arbitrary dependence structures when the group size is fixed. Together, conditions (ii) and (iii) imply that observed outcomes are independent conditional on the assignments. Finally, part (iv) requires that the total number of parameters not to grow faster than the effective sample size, that is, the expected sample size in the smallest cell.

Given a sample of $G$ groups with $n+1$ units each, let $\mathbb{1}_{ig}(\mathbf{a}) = \mathbb{1}(\mathbf{A}_{ig} = \mathbf{a})$, $N_g(\mathbf{a}) = \sum_{i=1}^{n+1} \mathbb{1}_{ig}(\mathbf{a})$ and $N(\mathbf{a}) = \sum_{g=1}^{G} N_g(\mathbf{a})$, so that $N_g(\mathbf{a})$ is the total number of observations

receiving effective assignment $\mathbf{a}$ in group $g$ and $N(\mathbf{a})$ is the total number of observations receiving effective assignment $\mathbf{a}$ in the sample. The estimator for $\mu(\mathbf{a})$ is defined as:

$$\hat{\mu}(\mathbf{a}) = \begin{cases} \frac{\sum_{g=1}^{G}\sum_{i=1}^{n+1} Y_{ig}\mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})} & \text{if } N(\mathbf{a}) > 0 \\ \nexists & \text{if } N(\mathbf{a}) = 0 \end{cases}$$

Thus, the estimator for $\mu(\mathbf{a})$ is simply the sample average of the outcome for observations receiving assignment $\mathbf{a}$, whenever there is at least one observation receiving this assignment.

The following assumption imposes some regularity conditions that are required for upcoming theorems. Let $\sigma^2(\mathbf{a}) = \mathbb{V}[Y_{ig}|\mathbf{A}_{ig} = \mathbf{a}]$.

**Assumption 3 (Moments)** *There are constants $\underline{\sigma}$ and $b$ such that:*

$$(i) \quad \inf_n \min_{\mathbf{a}\in\mathcal{A}_n} \sigma^2(\mathbf{a}) \geq \underline{\sigma}^2 > 0, \qquad (ii) \quad \sup_n \max_{\mathbf{a}\in\mathcal{A}_n} \mathbb{E}[Y_{ig}^6|\mathbf{A}_{ig} = \mathbf{a}] \leq b < \infty$$

Then we have the following result.

**Lemma 5 (Effective sample size)** *Suppose Assumptions 2 and 3 hold, and consider an assignment mechanism $\pi(\cdot)$ such that $\pi(\mathbf{a}) > 0$ for all $\mathbf{a} \in \mathcal{A}_n$. If*

$$\frac{\log|\mathcal{A}_n|}{G\underline{\pi}_n} \to 0 \tag{5}$$

*then for any $c \in \mathbb{R}$*

$$\mathbb{P}\left[\min_{\mathbf{a}\in\mathcal{A}_n} N(\mathbf{a}) > c\right] \to 1.$$

Lemma 5 says that, under condition (5), the number of observations in the smallest cell will go to infinity, which implies that all the estimators are well defined asymptotically. Hence, condition (5) formalizes the meaning of "large sample" in this context, and states that the number of groups has to be large relative to the total number of parameters and the probability of the least likely assignment. This expression can be interpreted as an invertibility condition for the design matrix of a linear regression model, in the specific case in which the regressors are mutually exclusive indicator variables. This requirement can be seen as a low-level condition that justifies the assumption of invertibility of the design matrix (see e.g. Assumption 2 in Cattaneo, Jansson, and Newey, 2018).

Next, let

$$\hat{\sigma}^2(\mathbf{a}) = \frac{\sum_{g=1}^{G}\sum_{i=1}^{n+1}(Y_{ig} - \hat{\mu}(\mathbf{a}))^2\mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})}\mathbb{1}(N(\mathbf{a}) > 0)$$

be the standard error estimators. Then we have the following result.

**Theorem 1 (Consistency and asymptotic normality)** *Suppose Assumptions 1, 2 and*

*3 hold. Under the conditions for Lemma 5,*

$$\begin{aligned}\max_{\mathbf{a}\in\mathcal{A}_n}|\hat{\mu}(\mathbf{a})-\mu(\mathbf{a})| &= O_{\mathbb{P}}\left(\sqrt{\frac{\log|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n}}\right),\\ \max_{\mathbf{a}\in\mathcal{A}_n}\left|\hat{\sigma}^2(\mathbf{a})-\sigma^2(\mathbf{a})\right| &= O_{\mathbb{P}}\left(\sqrt{\frac{\log|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n}}\right),\end{aligned} \tag{6}$$

*and*

$$\max_{\mathbf{a}\in\mathcal{A}_n}\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left[\frac{\hat{\mu}(\mathbf{a})-\mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}}\leq x\right]-\Phi(x)\right| = O\left(\frac{1}{\sqrt{G(n+1)\underline{\pi}_n}}\right) \tag{7}$$

*where $\Phi(x)$ is the cdf of a standard Gaussian random variable.*

Equation (6) shows that both the average potential outcome and standard error estimators converge in probability to their true values, uniformly over treatment assignments, at the rate $\sqrt{\log|\mathcal{A}_n|/(G(n+1)\underline{\pi}_n)}$. The denominator in this rate can be seen as the minimum expected cell size, whereas the numerator is a penalty for having an increasing number of parameters. Equation (7) bounds the difference between the distributions of the standardized potential outcomes estimators and the standard normal distribution, uniformly over the treatment assignments. Under condition (5), $G(n+1)\underline{\pi}_n \to \infty$, which gives asymptotic normality. Furthermore, this bound also reveals the rate at which the distribution of the standardized estimator approaches the standard normal, namely, $\sqrt{G(n+1)\underline{\pi}_n}$.

Importantly, both the rate of convergence and the rate of the distributional approximation depend on the assignment mechanism through $\underline{\pi}_n$, and this finding has important implications for the design of experiments to estimate spillovers, as discussed in section 6.

**Remark 4 (Inference with many small groups)** When the number of units per group is small compared to the total sample size, the effect of group size disappears asymptotically and inference can be based on a large $G$ small $n$ approximation as commonly done in panel data models. In this context, $n$, $|\mathcal{A}_n|$ and $\underline{\pi}_n$ are fixed so condition (5) holds automatically as long as the number of groups goes to infinity. Consistency and asymptotic normality of the estimators can be achieved under the usual regularity conditions as $G \to \infty$, and the variance estimator can easily account for both heteroskedasticity and intragroup correlation using standard techniques. The particular case with homoskedasticity and a random-effects structure is analyzed by Baird, Bohren, McIntosh, and Özler (2018). □

**Remark 5 (Inference for pooled parameters)** When focusing on pooled estimands as described in Section 3 (see Remarks 2 and 3), the set $\mathcal{A}_n$ does not change with $n$. For example, for the parameter $\Delta_{\mathsf{pp}}$ in Remark 3, the set of effective treatment assignments can be defined as $\mathcal{A}_n = \{(t,d) = (0,0),(1,0),(1,1)\}$ corresponding to the average outcomes for units in pure control groups ($T_g = 0$), control units in treated groups ($T_g = 1, D_{ig} = 0$) and treated units in treated groups ($T_g = 1, D_{ig} = 1$), respectively. In this case, $|\mathcal{A}_n| = 3$

and hence the number of parameters and the probabilities of each assignment do not change with group size. Hence, inference can be conducted using standard tools when when $G \to \infty$. Inference with a small number of large groups, as in the case where groups are villages or large geographical units (see e.g. Ichino and Schündeln, 2012; Giné and Mansuri, 2018; Crépon, Duflo, Gurgand, Rathelot, and Zamora, 2013) may require other strategies such as the wild bootstrap; see MacKinnon and Webb (2020) for a recent review. □

## 5.1 Bootstrap approximation

An alternative approach to perform inference in this setting is the bootstrap. Since the challenge for inference is that cells can have too few observations for the Gaussian distribution to provide a good approximation, the wild bootstrap (Shao, 1995) can offer a more accurate approximation when groups are relatively large. This type of bootstrap can be performed by defining weights $w_{ig} \in \{-1, 1\}$ with probability $1/2$ independently of the sample. The bootstrap estimator for $\mu(\mathbf{a})$ is given by:

$$\hat{\mu}^*(\mathbf{a}) = \frac{\sum_g \sum_i Y_{ig}^* \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})}$$

whenever the denominator is non-zero, where

$$Y_{ig}^* \mathbb{1}_{ig}(\mathbf{a}) = (\bar{Y}(\mathbf{a}) + (Y_{ig} - \bar{Y}(\mathbf{a}))w_{ig})\mathbb{1}_{ig}(\mathbf{a}) = (\bar{Y}(\mathbf{a}) + \hat{\varepsilon}_{ig} w_{ig})\mathbb{1}_{ig}(\mathbf{a})$$

In what follows, $\mathbb{P}^*[\cdot]$ denotes a probability calculated over the distribution of $w_{ig}$, conditional on the sample, and $\mathbb{E}^*[\cdot]$ and $\mathbb{V}^*[\cdot]$ the expectation and variance calculated over $\mathbb{P}^*[\cdot]$. The validity of the wild bootstrap is established in the following theorem.

**Theorem 2 (Wild bootstrap)** *Under Assumptions 1, 2 and 3,*

$$\max_{\mathbf{a} \in \mathcal{A}_n} \sup_{x \in \mathbb{R}} \left| \mathbb{P}^* \left[ \frac{\hat{\mu}^*(\mathbf{a}) - \hat{\mu}(\mathbf{a})}{\sqrt{\mathbb{V}^*[\hat{\mu}^*(\mathbf{a})]}} \leq x \right] - \mathbb{P} \left[ \frac{\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}} \leq x \right] \right| \to_{\mathbb{P}} 0.$$

This theorem shows that the wild bootstrap can be used to approximate the distribution of the estimator as an alternative to the standard normal, which may not be accurate when cells have few observations. The performance of the wild bootstrap will be illustrated in Section 7 using simulation data.

# 6 Implications for experimental design

Theorem 1 shows that the accuracy of the standard normal to approximate the distribution of the standardized statistic depends on the treatment assignment mechanism through $\underline{\pi}_n$. The intuition behind this result is that the amount of information to estimate each $\mu(\mathbf{a})$ depends on the number of observations facing assignment $\mathbf{a}$, and this number depends on $\pi(\mathbf{a})$. When

the goal is to estimate all the $\mu(\mathbf{a})$ simultaneously, the binding factor will be the number of observations in the smallest cell, controlled by $\underline{\pi}_n$. When an assignment sets a value of $\underline{\pi}_n$ that is very close to zero, the Gaussian distribution may provide a poor approximation to the distribution of the estimators.

When designing an experiment to estimate spillover effects, the researcher can choose distribution of treatment assignments $\pi(\cdot)$. Theorem 1 provides a way to rank different assignment mechanisms based on their rate of the approximation, which gives a principled way to choose between different assignment mechanisms.

To illustrate these issues, consider the case of an exchangeable exposure mapping $\mathcal{A}_n = \{(d, s) : d = 0, 1, s = 0, 1, \ldots, n\}$. The results below compare two treatment assignment mechanisms. The first one, *simple random assignment (SR)*, consists in assigning the treatment independently at the individual level with probability $\mathbb{P}[D_{ig} = 1] = p$. The second mechanism will be two-stage randomization. Although there are several ways to implement a two-stage design, I will focus on the case in which each group is assigned a fixed number of treated units between 0 and $n+1$ with equal probability. For example, if groups have size 3, then this mechanism assigns each group to receive 0, 1, 2 or 3 treated units with probability 1/4. This mechanism will be referred to as *two-stage randomization with fixed margins (2SR-FM)*. This mechanism is analyzed in Baird, Bohren, McIntosh, and Özler (2018), although its benefits in terms of asymptotic inference have not been previously studied.

**Corollary 1 (SR)** *Under simple random assignment, condition (5) holds whenever:*

$$\frac{n+1}{\log G} \to 0. \tag{8}$$

**Corollary 2 (2SR-FM)** *Under a 2SR-FM mechanism, condition (5) holds whenever:*

$$\frac{\log(n+1)}{\log G} \to 0. \tag{9}$$

In qualitative terms, both results imply that estimation and inference for spillover effects require group size to be small relative to the total number of groups. Thus, these results formalize the requirement of "many small groups" that is commonly invoked, for example, when estimating LIM models (see e.g. Davezies, D'Haultfoeuille, and Fougère, 2009; Kline and Tamer, 2019).

Corollary 1 shows that when the treatment is assigned using a simple random assignment, group size has to be small relative to $\log G$. Given the concavity of the log function, this is a strong requirement; for instance, with a sample of $G = 300$ groups, having $n = 5$ neighbors already gives $n + 1 > \log G$. Hence, groups have to be very small relative to the sample size for inference to be asymptotically valid. The intuition behind this result is that under a SR, the probability of the tail assignments $(0, 0)$ and $(1, n)$ decreases exponentially fast with group size.

On the other hand, Corollary 2 shows that a 2SR-FM mechanism reduces the requirement

to $\log(n+1)/\log G \approx 0$, so now the log of group size has to be small compared to the log of the number of groups. This condition is much more easily satisfied, which in practical terms implies that a 2SR-FM mechanism can handle larger groups compared to SR. The intuition behind this result is that, by fixing the number of treated units in each group, a 2SR-FM design has better control on how small the probabilities of each assignment can be, hence facilitating the estimation of the tail assignments.

# 7 Simulations

This section illustrates the above findings in a simulation setting. More precisely, I will study the performance of the spillover effects estimators under simple random assignment and 2SR-FM, as described in the previous section. The outcome will be binary and generated by the following DGP:

$$\mathbb{P}[Y_{ig}(d,s)=1] = \mu(d,s) = 0.75 + 0.13 \times d + 0.12 \times (1-d)\mathbb{1}(s>0)$$

which corresponds to the case with $\mu(0,0) = 0.75$, $\tau(0) = 0.13$, $\theta_0(s) = 0.12$ for all $s$ and $\theta_1(s) = 0$ for all $s$. That is, the spillover effects on an untreated unit is equal to 0.12 whenever at least one neighbor is treated, and zero for treated units.

The simulations consider two assignment mechanisms: SR with $\mathbb{P}[D_{ig} = 1] = 0.5$ and 2SR-FM in which groups are equally likely to be assigned to have any number from 0 to $n+1$ treated units. From Corollary 2, this assignment mechanism weakens the conditions for consistency and asymptotic normality from $(n+1)/\log G \to 0$ to $\log(n+1)/\log G \to 0$.

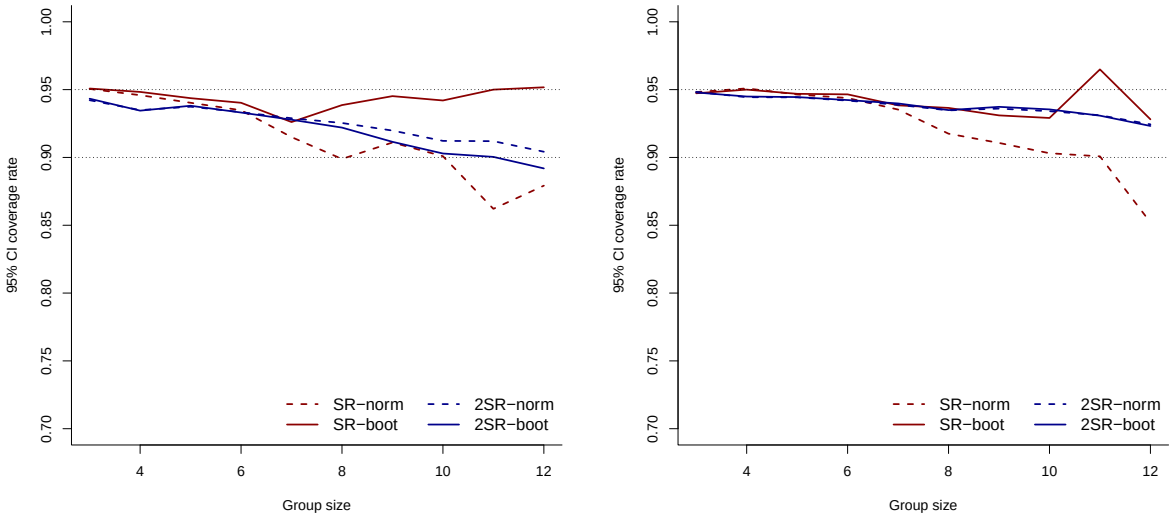
The parameter of interest will be $\theta_0(n) = \mathbb{E}[Y_{ig}(0,n)] - \mathbb{E}[Y_{ig}(0,0)]$, which is the average spillover effect for an untreated units with all neighbors treated. In this simulation, $\theta_0(n) = 0.12$. This parameters can be seen as a "worst-case scenario" given that the probability of the assignment $(D_{ig}, S_{ig}) = (0,n)$ is one of the smallest (in fact, the smallest under 2SR-FM). The estimator will be the difference in cell means:

$$\hat{\theta}_n(0) = \frac{\sum_g \sum_i Y_{ig}\mathbb{1}_{ig}(0,n)}{N(0,n)} - \frac{\sum_g \sum_i Y_{ig}\mathbb{1}_{ig}(0,0)}{N(0,0)}$$

whenever $N(0,n) > 1$ and $N(0,0) > 1$, so that both the estimator and its standard error can be calculated. When at least one of the cells has one or zero observations, the estimator is undefined.

Table 4 presents the results for a sample with 300 groups, for four group sizes, $n+1 = 3, 6, 9, 12$. The upper panel shows the results under SR while the lower panel corresponds to the 2SR-FM assignment. In each panel, the first row gives the value of the condition to achieve consistency and asymptotic normality; intuitively, the closer this value is to zero, the better the approximation based on the Gaussian distribution should be. The second and third rows show the bias and the variance of $\hat{\theta}_0(n)$, calculated over the values of the simulated estimates

Figure 1: Coverage rate of the 95% confidence interval.




(a) $\hat{\theta}_n(0)$, $G = 300$ (b) $\hat{\theta}_n(0)$, $G = 600$

**Notes**: the dashed lines show the coverage rate of the 95% confidence interval for $\theta_n(0)$ based on the normal approximation under simple random assignment (red line) and two-stage randomization (blue line) for a sample with 300 (left) and 600 (right) groups. The solid lines show the coverage rates for the confidence interval constructed using wild bootstrap.

conditional on the estimate being well defined (i.e. when the cells have enough observations to calculate the estimator). The third and fourth rows show the coverage rate of a 95% confidence interval based on the Gaussian approximation and a wild bootstrap confidence interval. Finally, the sixth row, labeled "proportion of empty cells", gives the proportion of the simulations in which the estimator or its standard error could not be calculated due to insufficient number of observations.

The simulations reveal that under both assignment mechanisms, the estimators perform well for $n = 2$ and $n = 5$, with biases close to zero and coverage rate close to 95%. In both cases the coverage rate decreases as group size increases reaching 88% under SR and 90% for 2SR-FM. For $n = 11$, the variance under SR is much larger than the one under 2SR-FM. These sharp differences in precision are due to the fact that, under simple randomization, when $n = 11$ the probability of observing observations in the cells $(0, 0)$ and $(1, n)$ is very close to zero; as shown in the fourth row of the upper panel, the estimator is undefined in almost 99% of the simulations, and, when it is defined, it relies on a very small number of observations. In fact, the expected number of observations in these cells is about 1.6, not enough to calculate a standard error. On the other hand, the variance under 2SR-FM is much more stable across group sizes, and the estimator can be defined in 100% of the cases. The difference in coverage rates under the two assignment mechanisms becomes more evident when $G = 600$, as shown in Figure 1. On the other hand, the wild bootstrap-

Table 4: Simulation results, $G = 300$

| | $n = 2$ | $n = 5$ | $n = 8$ | $n = 11$ |
|---|---|---|---|---|
| **Simple randomization** | | | | |
| $(n+1)/\log(G)$ | 0.5260 | 1.0519 | 1.5779 | 2.1039 |
| Bias | −0.0006 | 0.0007 | 0.0041 | −0.0118 |
| Variance | 0.0027 | 0.0128 | 0.0433 | 0.0654 |
| 95% CI coverage - normal | 0.9505 | 0.9348 | 0.9110 | 0.8792 |
| 95% CI coverage - bootstrap | 0.9508 | 0.9403 | 0.9452 | 0.9517 |
| Prop. empty cells | 0.0000 | 0.0087 | 0.5730 | 0.9851 |
| **Two-stage randomization** | | | | |
| $\log(n+1)/\log(G)$ | 0.1926 | 0.3141 | 0.3852 | 0.4357 |
| Bias | −0.0001 | 0.0000 | 0.0003 | −0.0003 |
| Variance | 0.0024 | 0.0034 | 0.0046 | 0.0058 |
| 95% CI coverage - normal | 0.9422 | 0.9334 | 0.9198 | 0.9042 |
| 95% CI coverage - bootstrap | 0.9433 | 0.9331 | 0.9115 | 0.8919 |
| Prop. empty cells | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

**Notes**: simulation results for $G = 300$ groups. The second and third rows in each panel show the bias and variance of $\hat{\theta}_n(0)$. The fourth and fifth rows show the coverage rate of a normal-based and wild-bootstrap-based 95% confidence intervals, respectively. The sixth row shows the proportion of simulations in which $\hat{\theta}_n(0)$ is undefined due to the small number of observations in the corresponding cell. Results from 10,000 simulations with 2,000 bootstrap repetitions.

based confidence interval maintains coverage close to 95% for all group sizes under simple randomization, whereas both the normal-based and the bootstrap-based confidence intervals perform similarly under 2SR. These results are also illustrated in Figure 1.

Table 5 shows the same results for a sample with 600 groups. As expected, the estimator and confidence intervals show better performance compare to the case with $G = 300$.

# 8 Discussion

This paper develops a potential-outcome-based nonparametric framework to analyze spillover effects that nests several models used in existing theoretical and empirical work. Within this framework, I define parameters of interest, provide identification conditions for these parameters and evaluate the performance of commonly applied methods such as the difference in means, linear-in-means models and partial population designs. Finally, I study estimation and inference with a special focus on the effect of the number of parameters on the asymptotic properties of the estimators, and discuss the implications of my results for experimental design.

The findings in this paper offer several takeaways for analyzing spillover effects in empirical work. A first conclusion that stems from Section 3 is that that the difference-in-means

Table 5: Simulation results, $G = 600$

| | $n = 2$ | $n = 5$ | $n = 8$ | $n = 11$ |
|---|---|---|---|---|
| **Simple randomization** | | | | |
| $(n+1)/\log(G)$ | 0.4690 | 0.9379 | 1.4069 | 1.8759 |
| Bias | 0.0001 | −0.0004 | 0.0028 | 0.0022 |
| Variance | 0.0013 | 0.0059 | 0.0270 | 0.0613 |
| 95% CI coverage - normal | 0.9482 | 0.9438 | 0.9107 | 0.8521 |
| 95% CI coverage - bootstrap | 0.9473 | 0.9465 | 0.9310 | 0.9281 |
| Prop. empty cells | 0.0000 | 0.0000 | 0.3112 | 0.9263 |
| **Two-stage randomization** | | | | |
| $\log(n+1)/\log(G)$ | 0.1717 | 0.2801 | 0.3435 | 0.3885 |
| Bias | 0.0005 | 0.0002 | −0.0004 | 0.0005 |
| Variance | 0.0012 | 0.0017 | 0.0023 | 0.0028 |
| 95% CI coverage - normal | 0.9483 | 0.9419 | 0.9360 | 0.9244 |
| 95% CI coverage - bootstrap | 0.9479 | 0.9423 | 0.9373 | 0.9232 |
| Prop. empty cells | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

**Notes**: simulation results for $G = 600$ groups. The second and third rows in each panel show the bias and variance of $\hat{\theta}_n(0)$. The fourth and fifth rows show the coverage rate of a normal-based and wild-bootstrap-based 95% confidence intervals, respectively. The sixth row shows the proportion of simulations in which $\hat{\theta}_n(0)$ is undefined due to the small number of observations in the corresponding cell. Results from 10,000 simulations with 2,000 bootstrap repetitions.

estimator can give a poor measure of the effects of a treatment when spillovers are present. On the other hand, the full vector of spillover effects is easily identifiable whenever the design generates enough variation in the number of treated units in each group, and can be easily estimated using fully interacted regressions.

Second, while nonparametric estimation of all direct and spillover effects can give a complete picture of the effects of the treatment, it can be difficult to implement in practice when groups are large. As a guideline to determine in which cases spillover effects can be estimated nonparametrically, Theorem 1 formalizes the notion of a "large enough sample" in this context, and provides a way to assess the performance of the different types of treatment effect estimators depending on the number of groups, number of parameters of interest and treatment assignment mechanism.

Third, when fully-nonparametric estimation of spillover effects is not feasible, this paper compares different ways to aggregate information to facilitate the practical implementation. In particular, I show that the commonly employed linear-in-means estimator can give an inaccurate measure of spillover effects, except under strong parametric assumptions. On the other hand, the pooling estimands in Section 3 provide weighted averages of spillover effects with known weights that depend only on the design, and inference on these parameters can be conducted under mild conditions. These findings also highlight the importance of clearly defining the set of parameters of interest, and how these choices determine at least partly the performance of alternative estimators and the way in which experiments should be designed.

The analysis in this paper leaves several questions open for future research. In terms of the setup, while the partial interference assumption has wide empirical applicability, in many contexts spillovers can occur through more complex interaction structures. The currently developing literature on networks seems like a natural path to generalize the setup in this paper. Future work should also formally address issues that arise frequently in empirical studies measuring spillovers, such as imperfectly measured groups or treatment missclasification.

# References


Abadie, A., and M. D. Cattaneo (2018): "Econometric Methods for Program Evaluation," *Annual Review of Economics*, 10, 465–503.

Angrist, J. D. (2014): "The perils of peer effects," *Labour Economics*, 30, 98–108.

Athey, S., D. Eckles, and G. W. Imbens (2018): "Exact P-values for Network Interference," *Journal of the American Statistical Association*, 113(521), 230–240.

Athey, S., and G. Imbens (2017): "Chapter 3 - The Econometrics of Randomized Experiments," in *Handbook of Field Experiments*, ed. by A. V. Banerjee, and E. Duflo, vol. 1 of *Handbook of Economic Field Experiments*, pp. 73 – 140. North-Holland.

Baird, S., A. Bohren, C. McIntosh, and B. Özler (2018): "Optimal Design of Experiments in the Presence of Interference," *The Review of Economics and Statistics*, 100(5), 844–860.

Barrera-Osorio, F., M. Bertrand, L. L. Linden, and F. Perez-Calle (2011): "Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia," *American Economic Journal: Applied Economics*, 3(2), 167–195.

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

Beshears, J., J. J. Choi, D. Laibson, B. C. Madrian, and K. L. Milkman (2015): "The Effect of Providing Peer Information on Retirement Savings Decisions," *The Journal of Finance*, 70(3), 1161–1201.

Beuermann, D. W., J. Cristia, S. Cueto, O. Malamud, and Y. Cruz-Aguayo (2015): "One Laptop per Child at Home: Short-Term Impacts from a Randomized Experiment in Peru," *American Economic Journal: Applied Economics*, 7(2), 53–80.

Blattman, C., D. Green, D. Ortega, and S. Tobón (2019): "Place Based Interventions at Scale: The Direct and Spillover Effects of Policing and City Services on Crime," *Available at SSRN: https://ssrn.com/abstract=3050823*.

Blume, L. E., W. A. Brock, S. N. Durlauf, and R. Jayaraman (2015): "Linear Social Interactions Models," *Journal of Political Economy*, 123(2), 444–496.

Bobonis, G. J., and F. Finan (2009): "Neighborhood Peer Effects in Secondary School Enrollment Decisions," *The Review of Economics and Statistics*, 91(4), 695–716.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of Econometrics*, 150(1), 41 – 55.

Brock, W. A., and S. N. Durlauf (2001): "Discrete Choice with Social Interactions," *The Review of Economic Studies*, 68(2), 235–260.

Carrell, S. E., R. L. Fullerton, and J. E. West (2009): "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics*, 27(3), 439–464.

Cattaneo, M. D., M. Jansson, and W. K. Newey (2018): "Inference in Linear Regression Models with Many Covariates and Heteroskedasticity," *Journal of the American Statistical Association*, 113(523), 1350–1361.

Choi, D. (2017): "Estimation of Monotone Treatment Effects in Network Experiments," *Journal of the American Statistical Association*, 112(519), 1147–1155.

Crépon, B., E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora (2013): "Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment," *The Quarterly Journal of Economics*, 128(2), 531–580.

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

de Paula, Á. (2013): "Econometric Analysis of Games with Multiple Equilibria," *Annual Review of Economics*, 5(1), 107–131.

——— (2017): "Econometrics of Network Models," in *Advances in Economics and Econometrics: Eleventh World Congress*, ed. by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, vol. 1 of *Econometric Society Monographs*, p. 268–323. Cambridge University Press.

Di Tella, R., and E. Schargrodsky (2004): "Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack," *American Economic Review*, 94(1), 115–133.

Dieye, R., H. Djebbari, and F. Barrera-Osorio (2014): "Accounting for Peer Effects in Treatment Response," *IZA Discussion Paper*, (8340).

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

Epple, D., and R. E. Romano (2011): "Peer Effects in Education: A Survey of the Theory and Evidence," *Handbook of Social Economics*, 1, 1053 – 1163.

Ferracci, M., G. Jolivet, and G. J. van den Berg (2014): "Evidence of Treatment Spillovers Within Markets," *The Review of Economics and Statistics*, 96(5), 812–823.

Giné, X., and G. Mansuri (2018): "Together We Will: Experimental Evidence on Female Voting Behavior in Pakistan," *American Economic Journal: Applied Economics*, 10(1), 207–235.

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

——— (2015): "Methods of Identification in Social Networks," *Annual Review of Economics*, 7(1), 465–485.

Halloran, M. E., and M. G. Hudgens (2016): "Dependent Happenings: a Recent Methodological Review," *Current Epidemiology Reports*, 3(4), 297–305.

Hirano, K., and J. Hahn (2010): "Design of Randomized Experiments to Measure Social Interaction Effects," *Economics Letters*, 106(1), 51–53.

Hong, G., and S. W. Raudenbush (2006): "Evaluating Kindergarten Retention Policy," *Journal of the American Statistical Association*, 101(475), 901–910.

Hoxby, C. M., and G. Weingarth (2005): "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects," *Working paper*.

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

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

Imbens, G. W., and D. B. Rubin (2015): *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Kline, B., and E. Tamer (2019): "Econometric Analysis of Models with Social Interactions," in *The Econometric Analysis of Network Data*, ed. by B. S. Graham, and A. De Paula. Elsevier.

Lalive, R., and M. A. Cattaneo (2009): "Social Interactions and Schooling Decisions," *The Review of Economics and Statistics*, 91(3), 457–477.

Liu, L., and M. G. Hudgens (2014): "Large Sample Randomization Inference of Causal Effects in the Presence of Interference," *Journal of the American Statistical Association*, 109(505), 288–301.

MacKinnon, J. G., and M. D. Webb (2020): "When and How to Deal with Clustered Errors in Regression Models," Working Paper 1421, Economics Department, Queen's University.

Manski, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60(3), 531–542.

——— (2013): "Identification of treatment response with social interactions," *The Econometrics Journal*, 16(1), S1–S23.

Miguel, E., and M. Kremer (2004): "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 72(1), 159–217.

Moffit, R. (2001): "Policy Interventions, Low-level Equilibria and Social Interactions," in *Social Dynamics*, ed. by S. N. Durlauf, and P. Young, pp. 45–82. MIT Press.

Rosenbaum, P. R. (2007): "Interference Between Units in Randomized Experiments," *Journal of the American Statistical Association*, 102(477), 191–200.

Sacerdote, B. (2011): "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?," in *Handbook of the Economics of Education*, ed. by S. M. Eric A. Hanushek, and L. Woessmann, vol. 3, pp. 249 – 277.

——— (2014): "Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward?," *Annual Review of Economics*, 6(1), 253–272.

Shao, J. (1995): *The Jackknife and Bootstrap*. Springer New York, New York, NY.

Sinclair, B., M. McConnell, and D. P. Green (2012): "Detecting Spillover Effects: Design and Analysis of Multilevel Experiments," *American Journal of Political Science*, 56(4), 1055–1069.

Tchetgen Tchetgen, E. J., and T. J. VanderWeele (2012): "On causal inference in the presence of interference," *Statistical Methods in Medical Research*, 21(1), 55–75.

Vazquez-Bare, G. (2020): "Causal Spillover Effects Using Instrumental Variables," *arXiv:2003.06023*.