Identification and Estimation of Spillover Effects in Randomized Experiments*

Gonzalo Vazquez-Bare†

March 20, 2019

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. I establish nonparametric identification of spillover effects, and propose nonparametric estimators that 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 then show that these results have important implications for experimental design. These findings are illustrated using data from a conditional cash transfer program in Colombia, 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 Titunik 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 the 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 (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 which are all equally important to assess the effect of a policy or treatment. I discuss an interpretable restriction, exchangeability, under which average potential outcomes do not change when swapping the identities of the treated neighbors. As shown in the paper, this restriction justifies the commonly employed assumption that outcomes depend only on the number (or proportion) of treated neighbors. 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.

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 4 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 asymp-
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. 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.

The ideas and methods put forth in this paper are illustrated by reanalyzing a randomized conditional cash transfer program studied by Barrera-Osorio, Bertrand, Linden, and Perez-Calle (2011). I discuss the empirical performance of two regression-based specifications that are widely used in empirical work: a regression of the outcome on a treatment indicator (i.e. a difference in means) and a regression on a treatment indicator and the proportion of treated neighbors (a reduced-form linear-in-means model). The results reveal the potential pitfalls of failing to flexibly account for spillover effects in policy evaluation. Finally, Section 5 discusses the inclusion of covariates. Section 6 discusses several implications for empirical work and points to upcoming and 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 3.2, 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, Rathelet, and Zamora (2013), Beuermann, Cristia, Cueto, Malamud, and Cruz-Aguayo (2015), Beshears, Choi, Laibson, Madrian, and Milkman (2015) and Gine 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 4 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 A3 of the supplemental appendix. This paper complements this important strand of the literature by offering identification, estimation and inference results for well-defined causal (reduced-form) treatment effects in the presence of spillovers.

Finally, a literature coming from 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). Given this non-random potential outcomes setting, identification issues are largely absent from this literature, and focus is placed mainly on p-values, variance and confidence interval calculations (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 replaces a partial interference assumption with more general network structures (Athey, Eckles, and Imbens, 2018; Choi, 2017). In this paper, I take a super-population approach under repeated sampling which 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 conduct 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. 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

---

1The 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.
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 underestimate the costs and benefits of this policy. Moreover, these alternative scenarios have drastically different implications on how to assign the program 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 sample consisting of independent and identically distributed groups indexed by \( g = 1, \ldots, G \), each with \( n_g + 1 \) units, so that each unit 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, an assumption sometimes known as partial interference (Sobel, 2006).

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 \( \mathbf{D}_{(i)g} = (D_{1ig}, D_{2ig}, \ldots, D_{n_{ig}}) \). This vector takes values \( \mathbf{d}_g = (d_1, d_2, \ldots, d_{n_g}) \in \mathcal{D}_g \subseteq \{0, 1\}^{n_g} \). As will be discussed in more detail later, this notation requires assigning identities to neighbors, although this requirement can be dropped under additional assumptions. For a given realization of the treatment assignment \( (d, \mathbf{d}_g) = (d, d_1, d_2, \ldots, d_{n_g}) \), the potential outcome for unit \( i \) in group \( g \) is denoted by the
random variable $Y_{ig}(d, \mathbf{d}_g)$. Throughout the paper, I will assume that all the required moments of the potential outcomes are bounded. The observed outcome for unit $i$ in group $g$ is the value of the potential outcome under the observed treatment realization, given by $Y_{ig} = Y_{ig}(D_{ig}, \mathbf{D}(i)_g)$. Note that in the presence of interference, each unit has $2^{n_g+1}$ potential outcomes, and this number reduces to the usual case with two potential outcomes when interference is ruled out. Hence, this setup relaxes the Stable Unit Treatment Value Assumption (SUTVA), according to which the potential outcomes depend only on own treatment status, $Y_{ig}(d, \mathbf{d}_g) = Y_{ig}(d)$. 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 (2019). In what follows, $0_g$ and $1_g$ will denote $n_g$-dimensional vectors of zeros and ones, respectively. The observed potential outcome can be written as:

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

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 $\mathbf{d}_g = (d_1, d_2)$ and the potential outcome is $Y_{ig}(d, d_1, d_2)$. The number of possible treatment assignments $(d, d_1, d_2)$ is $2^{n_g+1} = 8$. For example, $Y_{ig}(1, 0, 0) - Y_{ig}(0, 0, 0)$ is the effect of the treatment when both of kid $i$'s siblings are untreated, $Y_{ig}(0, 1, 0) - Y_{ig}(0, 0, 0)$ is the spillover effect on unit $i$ of treating kid $i$'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 $\mathbb{E}[Y_{ig}(d, d_1, d_2)] - \mathbb{E}[Y_{ig}(\tilde{d}, \tilde{d}_1, \tilde{d}_2)]$. For simplicity, throughout the paper I will assume that potential outcomes within a group have the same distribution, so that in particular $\mathbb{E}[Y_{ig}(d, \mathbf{d}_g)]$ does not depend on $i$ or $g$.

A salient feature of this model is that each unit has a specific identity in the sense that, for example, with a group of size 3, $\mathbb{E}[Y_{ig}(d, 1, 0) - Y_i(d, 0, 0)] \neq \mathbb{E}[Y_{ig}(d, 0, 1) - Y_{ig}(d, 0, 0)]$ in general, that is, the effect on unit $i$ of giving treatment to neighbor 1 may differ in general from the effect of giving treatment to neighbor 2. Hence, allowing for units to have specific identities requires a natural labeling or ordering between units in each group, which can be given for example by distance according to some specified metric. A natural example would be geographical distance that orders neighbors from closest to farthest. Another example would be the case where one can rank the relationships according to its strength, e.g. closest friend, second-closest friend, etc.

Allowing for different neighbor identities leaves the structure of within-group spillovers completely unrestricted. This level of generality, however, may easily introduce a dimensionality problem. The number of potential outcomes increases exponentially with group size, and it can quickly become larger than the number of observations. More precisely, with equally-sized groups, the number of observations is $(n_g + 1)G$, whereas the number of potential outcomes is $2^{n_g+1}$, so there are at least as many potential outcomes as observations.

---

2This assumption can be relaxed by allowing the averages to depend on $i$, and switching focus to the within-group average $(n_g + 1)^{-1} \sum_{i=1}^{n_g+1} \mathbb{E}[Y_{ig}(d, \mathbf{d}_g)]$. 

6
whenever \(2^{n_g+1} \geq (n_g + 1)G\). As a simple illustration, with 200 groups, \(G = 200\), the number of potential outcomes exceeds the total sample size as soon as \(n_g + 1 \geq 12\). Even when the condition \((n_g + 1)G > 2^{n_g+1}\) holds, the number of potential outcomes may be too high for estimation results to be reliable. For example, with \(G = 200\) and \(n_g + 1 = 10\), we have 2000 observations to estimate 1024 average potential outcomes.

One way to reduce this dimensionality problem is to impose an “anonymity” assumption under which the spillover effects do not depend on the specific identity of each treated neighbor. Intuitively, this condition states that, given the number of treated neighbors for a specific unit, the potential outcome does not change when swapping the treatment assignment between neighbors, so that neighbors are exchangeable. In this case, the number of possible potential outcome values in each group drops from from \(2^{n_g+1}\) to \(2(n_g + 1)\). To formalize this idea, I assume the following condition.

**Assumption 1 (Exchangeability)** Let \(d_g, \tilde{d}_g \in D_g\) such that \(1_g' d_g = 1_g' \tilde{d}_g\). Then, for each \(d = 0, 1\),

\[
\mathbb{E}[Y_{ig}(d, d_g)] = \mathbb{E}[Y_{ig}(d, \tilde{d}_g)].
\]

Assumption 1 states that the average potential outcome is invariant to permutations of the neighbor’s assignment vector \(d_g\). Exchangeability implies the following restriction on the potential outcome.

**Lemma 1 (Potential outcomes under exchangeability)** For any \(d_g \in D_g\), let \(s = 1_g' d_g = \sum_{j=1}^{n_g} d_j\). Under Assumption 1, for \(d = 0, 1\), there is a function \(\mu(d, \cdot) : \{0, 1, \ldots, n_g\} \to \mathbb{R}\) such that:

\[
\mathbb{E}[Y_{ig}(d, d_g)] = \mu(d, s).
\]

Lemma 1 states that, for each unit \(i\) in group \(g\), the average potential outcome only depends on the neighbors’ assignment \(d_g\) through \(s = \sum_{j=1}^{n_g} d_j\). In this case, \(s = 0\) indicates that unit \(i\) in group \(g\) has no treated neighbors, whereas \(s = n_g\) corresponds to the case where all neighbors are treated, and so on. For any pair of vectors \(d_g\) and \(\tilde{d}_g\) such that \(1_g' d_g = 1_g' \tilde{d}_g\), exchangeability restricts the average spillover effect to zero, that is,

\[
\mathbb{E}[Y_{ig}(d, d_g)] - \mathbb{E}[Y_{ig}(d, \tilde{d}_g)] = 0.
\]

This restriction is what reduces the number of parameters in the model.

Previous studies have considered this or similar versions of this assumption (see e.g. Hudgens and Halloran, 2008; Manski, 2013). In other cases, result in Lemma 1 is stated as an assumption (see e.g. Baird, Bohren, McIntosh, and Özler, 2018; Ferracci, Jolivet, and van den Berg, 2014). Lemma 1 explicitly states the restrictions that this condition imposes on the potential outcomes. On the other hand, exchangeability is invoked, either explicitly or implicitly, in nearly all empirical studies analyzing spillover effects. For example, 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). Hence, the motivation for imposing Assumption 1 in this paper is understanding the extent to which this commonly invoked assumption affects identification, estimation and inference.

The plausibility of the exchangeability assumption needs to be considered on a case-by-case basis. Consider, for example, a program that assigns vaccines to students in classrooms to prevent some contagious disease. It is possible that this program prevents the unvaccinated children from getting sick through herd immunity as long as the number of treated children is large enough. In this case, it may be reasonable to assume that what matters is not which students receive the vaccine, but how many of them, since all students share a common closed space. In other cases, the exchangeability assumption may be less plausible. For example, Banerjee, Chandrasekhar, Duflo, and Jackson (2013) study the diffusion of information through social interactions in Indian villages, and show how adoption of a new technology (microfinance loans) in a village depends on the degree of centrality of the individuals who are first informed about it. In such a case, it is clear than the effect of treating a neighbor will vary depending on whether the neighbor is a “leader” or a “follower”.

I will maintain this assumption throughout this section to conserve space, but all the ideas extend straightforwardly to the non-exchangeable case. Section A2 of the supplemental appendix provides a general treatment without exchangeability, discusses some alternatives to relax this assumption and provides a formal test of exchangeability including some additional empirical results. Moreover, Lemma 3 below analyzes the effect of incorrectly imposing exchangeability.

I will define two sets of parameters of interest. First, the **average direct effect** of the treatment given $s$ treated neighbors is defined as:

$$\tau_s = \mu(1, s) - \mu(0, s)$$  

so each $\tau_s$ represents the average effect of giving treatment to a unit, holding the number of treated neighbors fixed at $s$. For a group of size $n_g + 1$, there are $n_g + 1$ of these parameters, one for each possible value of $s$. Second, the **average spillover effect** of $s$ treated siblings given own treatment status $d$ is:

$$\theta_s(d) = \mu(d, s) - \mu(d, 0)$$  

so $\theta_s(d)$ captures the average effect of giving treatment to $s$ neighbors, compared to having no treated neighbors, for a unit under treatment status $d$. Importantly, these effects can depend on $n_g$, since in principle direct and spillover effects can be different in large versus small groups. However, I will omit the $n_g$ from the arguments to avoid excessive notation (see Section A4 on the supplemental appendix for further discussion). These two sets of parameters do not exhaust all the possible comparisons between potential outcomes, but any other effect of interest can be reconstructed as a linear combination of $\tau_s$ and $\theta_s(d)$. For
instance, the marginal effect of an additional treated neighbor can be constructed as \( \theta_{s+1}(d) - \theta_s(d) \). In the next section I provide conditions to achieve identification of these treatment effects when the treatment is randomly assigned. Section 3.2 will link these parameters to the estimands of the difference in means and the linear-in-means regression.

### 3 Identification

The key feature of random assignment is that it ensures that potential outcomes are unrelated to treatment assignment. I formalize this condition as follows.

**Assumption 2 (Random assignment)** For all \((d_i, d_g) \in \{0, 1\} \times D_g\) and for all \(i\) and \(g\),

\[
Y_{ig}(d, d_g) \perp D_g.
\]

This condition states that potential outcomes are statistically independent of the treatment assignment vector, and rules out selection into treatment. Under SUTVA, this condition reduces to \((Y_{ig}(0), Y_{ig}(1)) \perp D_{ig}\), which means for example that the average potential outcome under no treatment is equal between treated and control units. In the presence of spillovers, independence needs to be strengthened to ensure that the potential outcomes are independent not only of own treatment assignment, but also of neighbors’ treatment assignments. This requirement is guaranteed when the treatment is randomly assigned.

Let \(S_{ig} := \sum_{j \neq i}^{n_g} D_{jg}\) be the observed number of treated neighbors for unit \(i\) in group \(g\). The following result shows identification of average direct and spillover effects under exchangeability.

**Lemma 2 (Identification)** Under Assumptions 1 and 2, for \(d = 0, 1\) and \(s = 0, 1, \ldots, n_g\), for any assignment such that \(P[D_{ig} = d, S_{ig} = s] > 0\),

\[
E[Y_{ig}|D_{ig} = d, S_{ig} = s] = \mu(d, s).
\]

Lemma 2 shows how, under random assignment of the treatment, all the average potential outcomes can be nonparametrically identified by exploiting variation in all the possible configurations of own and neighbors’ observed treatment assignments. It follows immediately from Lemma 2 that:

\[
\tau_s = E[Y_{ig}|D_{ig} = 1, S_{ig} = s] - E[Y_{ig}|D_{ig} = 0, S_{ig} = s]
\]

and

\[
\theta_s(d) = E[Y_{ig}|D_{ig} = d, S_{ig} = s] - E[Y_{ig}|D_{ig} = d, S_{ig} = 0].
\]

The main condition to achieve identification under random assignment is that the treatment assignment mechanism puts non-zero probability on each \((d, s)\), that is, \(P[D_{ig} = d, S_{ig} = s] > 0\). In the absence of spillovers, this condition is trivially satisfied, since there are only
two treatment assignments, treated and control, that occur with non-zero probability as long as $P[D_{ig} = 1] \in (0, 1)$. In the presence of spillovers, this requirement becomes non-trivial because the number of possible treatment assignments is potentially large, and some assignment mechanisms could place zero probability in some of them. For example, consider a cluster randomized trial in which groups, instead of units, are assigned to treatment with probability $1/2$, so that in each group either everybody is treated or nobody is. This assignment mechanism implies that $P[D_{ig} = 1, S_{ig} = n_g] = P[D_{ig} = 0, S_{ig} = 0] = 1/2$ and $P[D_{ig} = d, S_{ig} = s] = 0$ for all assignments $(d, s)$ different from $(1, n_g)$ and $(0, 0)$. Hence, the only treatment effect that can be identified under this assignment mechanism is $\mu(1, n_g) - \mu(0, 0)$, that is, the effect of being treated with all treated neighbors compared to being untreated with no treated neighbors. Assigning the treatment at the individual level is therefore a necessary (but not sufficient) condition to identify all the direct and spillover effects.

On the other hand, Lemma 2 also shows that complex assignment mechanisms like two-stage designs assignments like the ones discussed by Moffit (2001), Duflo and Saez (2003), Hirano and Hahn (2010) and Baird, Bohren, McIntosh, and Özler (2018), among others, are not required for identification purposes (although they can improve estimation and inference, as discussed in Section 4).

A natural question that stems from this discussion is what parameters can be recovered when exchangeability is incorrectly assumed. The following lemma addresses this issue.

**Lemma 3 (Identification with misspecified exchangeability)** Under Assumption 2, for any assignment such that $P[D_{ig} = d, S_{ig} = s] > 0$,

$$
E[Y_{ig}|D_{ig} = d, S_{ig} = s] = \sum_{d_g, d_g'} E[Y_{ig}(d, d_g)] \frac{P[D_{ig} = d|D_{(i)g} = d_g] P[D_{(i)g} = d_g]}{P[D_{ig} = d|S_{ig} = s] P[S_{ig} = s]}
$$

This result shows that when incorrectly assuming exchangeability, exploiting variation in the number of treated neighbors recovers a weighted average of all the spillover effects for a fixed value of $s$ using the assignment probabilities. In particular, when individual treatment assignments are iid within a group, this expression reduces to:

$$
E[Y_{ig}|D_{ig} = d, S_{ig} = s] = \left( \frac{n_g}{s} \right)^{-1} \sum_{d_g} E[Y_{ig}(d, d_g)]
$$

and thus $E[Y_{ig}|D_{ig} = d, S_{ig} = s]$ is a simple average of the average potential outcomes $E[Y_{ig}(d, d_g)]$ taken over all the possible assignments with $s$ treated neighbors.

Hence, even when exchangeability does not hold, it can still be used as a device to reduce the number of parameters to be estimated, at the expense of losing the heterogeneity in spillover effects generated by the different neighbor identities.
3.1 Pooled estimands

Pooling spillover effects can provide a useful summary measure of spillovers, and can be a feasible alternative when the total number of parameters is large and hence each $\theta_s(d)$ may be imprecisely estimated, an issue that will be analyzed in detail in Section 4.

Lemma 2 implies that, for any known (or estimable) vector of weights $\omega = \{\omega_s(d)\}_{(d,s)}$, the pooled spillover effect:

$$\theta_p(\omega) = \sum_{s=1}^{n_g} \omega_s(0)\theta_s(0) + \sum_{s=1}^{n_g} \omega_s(1)\theta_s(1)$$

is identified. I will discuss two ways to pool spillover effects.

The first way consists on directly grouping units facing different number of treated peers together. For example, one possibility would be to compare control units with no treated peers to control units with one or more treated peers, effectively computing:

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

From the results in previous section, under the conditions of Lemma 2,

$$\Delta = \sum_{s=1}^{n_g} \theta_s(0)P[S_{ig} = s|D_{ig} = 0, S_{ig} > 0].$$

Thus, $\Delta$ is a weighted average of spillover effects on the controls, with weights given by the conditional probabilities of having each possible number of treated peers. Since the weights depend only on the experimental design, they are known to the researcher when conducting a randomized trial. A natural generalization of this idea is to divide the values of $S_{ig}$ into categories, such as $S_{ig} \in \{1, \ldots, k\}$, $S_{ig} = \{k+1, \ldots, n_g\}$ and so on.

The second way to pool spillover effects is based on the experimental design. More specifically, a popular design when analyzing spillover effects is the partial population design (Moffit, 2001; Baird, Bohren, McIntosh, and Özler, 2018). Consider the experiment by Duflo and Saez (2003), in which groups are randomly divided into treated and control using a 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,

$$\Delta_{PP} = E[Y_{ig}|D_{ig} = 0, T_g = 1] - E[Y_{ig}|T_g = 0].$$ (3)

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. The following result shows the causal parameter that is recovered by $\Delta_{PP}$.
Lemma 4 (Identification in partial population experiments) If $Y_{ig}(d, d_g) \perp (D_g, T_g)$, under Assumption 1,

$$\Delta_{PP} = \sum_{s=1}^{n_g} \theta_s(0) \mathbb{P}[S_{ig} = s | D_{ig} = 0, T_g = 1].$$

Hence, $\Delta_{PP}$ recovers a weighted average of all the spillover effects on the controls, where the weights are given by the probability of each possible number of treated neighbors in treated groups. The generalization of this result to experiments with more than two categories, as in Crépon, Duflo, Gurgand, Rathelot, and Zamora (2013), is straightforward.

3.2 Empirical illustration

This section employs the data from the experiment described in Section 2 to illustrate the above results. 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.
Table 3: Estimation results

<table>
<thead>
<tr>
<th>Model</th>
<th>Diff. Means</th>
<th>Linear-in-Means</th>
<th>Full</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coef</td>
<td>s.e.</td>
<td>coef</td>
<td>s.e.</td>
</tr>
<tr>
<td>$D_{ig}$</td>
<td>0.006</td>
<td>0.016</td>
<td>0.007</td>
<td>0.016</td>
</tr>
<tr>
<td>$D_{ig}^{(i)}$</td>
<td></td>
<td>0.027</td>
<td>0.034</td>
<td></td>
</tr>
<tr>
<td>$D_{ig}^{(i)} (1 - D_{ig})$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_{ig}^{(i)} D_{ig}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mathbb{I}(S_{ig} = 1)(1 - D_{ig})$</td>
<td></td>
<td></td>
<td></td>
<td>0.146**</td>
</tr>
<tr>
<td>$\mathbb{I}(S_{ig} = 2)(1 - D_{ig})$</td>
<td></td>
<td></td>
<td></td>
<td>0.14**</td>
</tr>
<tr>
<td>$\mathbb{I}(S_{ig} = 1)D_{ig}$</td>
<td></td>
<td></td>
<td></td>
<td>-0.041*</td>
</tr>
<tr>
<td>$\mathbb{I}(S_{ig} = 2)D_{ig}$</td>
<td></td>
<td></td>
<td></td>
<td>-0.051**</td>
</tr>
<tr>
<td>$\mathbb{I}(S_{ig} &gt; 0)(1 - D_{ig})$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mathbb{I}(S_{ig} &gt; 0)D_{ig}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.822***</td>
<td>0.013</td>
<td>0.811***</td>
<td>0.024</td>
</tr>
<tr>
<td>Observations</td>
<td>504</td>
<td>504</td>
<td>504</td>
<td>504</td>
</tr>
</tbody>
</table>

Notes: S.e. clustered at the household level. Regressions include school FE. ***p < 0.01, **p < 0.05, *p < 0.1.
The results are obtained by estimating the following regression:

$$E[Y_{ig}|D_{ig},S_{ig}] = \alpha + \tau_0 D_{ig} + \sum_{s=1}^{n_g} \theta_s(0) \mathbb{1}(S_{ig} = s)(1 - D_{ig}) + \sum_{s=1}^{n_g} \theta_s(1) \mathbb{1}(S_{ig} = s) D_{ig} \quad (4)$$

where $\alpha = \mu(0,0)$. Because it is equivalent to estimating averages at each cell separately, Equation (4) 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 (4) 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 $\theta_1(0) = \theta_2(0)$ 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 $\mu(1,2) - \mu(0,0) = \mu(1,0) - \mu(0,0) + \mu(1,2) - \mu(1,0) = \tau_0 + \theta_2(1)$, so it can be estimated by $\hat{\tau}_0 + \hat{\theta}_2(1) \approx 0.113$. 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 clearly show that exchangeability cannot be rejected in this case.

### 3.3 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 $\beta_D$ in the model:

$$Y_{ig} = \alpha_D + \beta_D D_{ig} + u_{ig}. \quad (5)$$

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. More formally, we have the following result:

**Lemma 5 (Difference in means)** Under the conditions for Lemma 2, the coefficient $\beta_D$ from Equation (5) can be written as:

$$\beta_D = \tau_0 + \sum_{s=1}^{n_g} \theta_s(1)P[S_{ig} = s | D_{ig} = 1] - \sum_{s=1}^{n_g} \theta_s(0)P[S_{ig} = s | D_{ig} = 0]$$

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:

$$\beta_D = \tau_0 + \sum_{s=1}^{n_g} (\theta_s(1) - \theta_s(0))P[S_{ig} = s]$$

The effect of the presence of spillovers in the difference in means, captured by the term $\sum_{s=1}^{n_g} (\theta_s(1) - \theta_s(0))P[S_{ig} = s]$, 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, $\theta_s(0) = \theta_s(1)$ for all $s$, then the difference in means $\beta_D$ 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 much larger under control that under treatment, and have different signs, so $\hat{\theta}_s(1) - \hat{\theta}_s(0) < 0$. Therefore, the spillover effects push the difference in means towards zero in this case.

### 3.4 Linear-in-means models

Equation (5) 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:

\[ Y_{ig} = \alpha + \beta \ell D_{ig} + \gamma \ell \bar{D}_{g(i)} + \eta_{ig}, \]

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 (6) 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 (4). To better understand this point, Equation (4) 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, \( \theta_s(0) = \theta_s(1) = \theta_s \) for all \( s \) and (ii) the spillover effects are linear in \( s \), that is, \( \theta_s = \kappa s \) for some constant \( \kappa \), then Equation (4) reduces to:

\[ E[Y_{ig}|D_{ig}, S_{ig} = s] = \alpha + \tau_0 D_{ig} + \theta_{n_g} \bar{D}_g^{(i)} \]

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 do not seem to hold in this case. More in general, the following result holds.

**Lemma 6 (LIM regression)** Under the conditions for Lemma 2 and simple random assignment, the coefficient \( \gamma \ell \) from Equation (6) can be written as:

\[
\gamma \ell = n_g \sum_{s=1}^{n_g} \left[ \theta_s(0)(1 - p) + \theta_s(1)p \right] \frac{\text{Cov}(S_{ig}, 1(S_{ig} = s))}{\text{V}[S_{ig}]} \\
= n_g \sum_{s=1}^{n_g} \left[ \theta_s(0)(1 - p) + \theta_s(1)p \right] \left( s - \frac{\text{E}[S_{ig}]}{\text{V}[S_{ig}]} \right) \text{P}[S_{ig} = s]
\]

where \( p = \text{P}[D_{ig} = 1] \).

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, \( \theta_s(0)(1 - p) + \theta_s(1)p \), and then combines all these terms. Importantly, the “weights” assigned to each of the terms \( \theta_s(0)(1 - p) + \theta_s(1)p \) 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 (6) more flexible is to include an interaction between $D_{ig}$ and $\bar{D}^{(i)}_g$ to allow for the spillover effects to be different under treatment and control:

$$Y_{ig} = \alpha + \beta D_{ig} + \gamma^0_0 \bar{D}^{(i)}_g (1 - D_{ig}) + \gamma^1_0 \bar{D}^{(i)}_g D_{ig} + \xi_{ig}$$  \hspace{1cm} (7)$$

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 7 (LIM with interactions)** Under the conditions for Lemma 2 and simple random assignment, for $d = 0, 1$ the coefficients $\gamma_d^{d}$ can be written as:

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

Thus, the only difference is that each $\gamma^d_d$ combines the spillover effects under a fixed treatment status $d$, instead of averaging $\theta_s(0)$ and $\theta_s(1)$. As before, this expression shows that the coefficients $\gamma^d_d$ are not weighted averages of the spillover effects $\theta_s(d)$. More precisely, they assign negative weights to the parameters $\theta_s(d)$ 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.

### 3.5 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:

$$Y_{ig} = \alpha + \beta D_{ig} + \gamma^0_p (S_{ig} > 0) (1 - D_{ig}) + \gamma^1_p (S_{ig} > 0) D_{ig} + \nu_{ig}$$

where as shown in Section 3.1,

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

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

The previous section illustrates how, under random assignment of the treatment, all the parameters of interest can be recovered using a fully-saturated regression with the number of coefficients equal to the number of average potential outcomes to estimate. 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 assignments \((d, s)\). 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 asymptotically. 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 can groups be, relative to the total sample size, to allow for valid estimation and inference. More formally, I will provide conditions for consistency and asymptotic normality in a setting in which group size is allowed to grow with the sample size. The key issue will be to ensure that the number of observations in all cells grows to infinity as the sample size increases. It is important to clarify that 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 will be shown to be a particular case in this setting.

In addition, I will relax the assumption of exchangeability to allow for a general treatment of potential outcomes. This generalization will allow me to analyze the role of different potential outcome restrictions and different choices of parameters of interest in the asymptotic properties of the proposed estimators.

I will start by defining two concepts that will play a crucial role in estimation and inference. First, let \(\mathcal{A}_n\) be the set of effective treatment assignments, with cardinality \(|\mathcal{A}_n|\). This set contains all the treatment assignments that are of interest for the researcher in a particular study. For example, if the researcher suspects the absence of spillovers and

---

3As stated in Assumption 3, I will assume in this section that groups are equally sized. The case with unequally-sized groups is briefly discussed in Section A4 of supplemental appendix.
only cares about comparisons between treated and controls, this set can be defined as $A_n = \{0, 1\}$. When the goal is to estimate spillover effects under exchangeability, as in Equation (4), $A_n = \{(d, s) : d = 0, 1, s = 0, 1, \ldots, n_g\}$ and $|A_n| = 2(n_g + 1)$. On the other hand, when aiming at estimating all the possible direct and spillover effects, $A_n = \{(d, d_g) : d = 0, 1, d_g \in \{0, 1\}\}$ and $|A_n| = 2(n_g + 1)$. Thus, the set $A_n$ indicates which (and how many) conditional means have to be estimated. The more information one wants to extract from the sample to study spillover effects, the more complex this set becomes, and the larger the number of parameters to estimate. The observed effective assignment for unit $i$ in group $g$ is denoted by $A_{ig}$, and $\mu(a) = \mathbb{E}[Y_{ig}|A_{ig} = a]$ for $a \in A_n$.

Second, each treatment assignment mechanism determines a distribution $\pi(\cdot)$ over $A_n$ where $\pi(a) = \mathbb{P}[A_{ig} = a]$ for $a \in A_n$. For example, in an experiment without spillovers in which 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, by allowing for spillovers with exchangeability, $\pi(d, s) = \mathbb{P}[D_{ig} = d, S_{ig} = s] = \binom{n_g}{s}p^{s+d}(1-p)^{n_g + 1 - s - d}$. In the latter case, as group size increases, $|A_n| \to \infty$ and $\pi(a) \to 0$ for all $a$. Finally, define:

$$\overline{\pi}_n = \min_{a \in A_n} \pi(a)$$

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 $A_g = (A_{1g}, \ldots, A_{n_g + 1, g})$, $A = (A_1, \ldots, A_G)$, and $Y_g = (Y_{1g}, Y_{2g}, \ldots Y_{n_g + 1, g})'$. I will assume the following sampling scheme.

**Assumption 3 (Sampling and design)**

(i) For $g = 1, \ldots, G$, $(Y_g', A_g')$ is a random sample.

(ii) Within each group $g$, the observed outcomes $Y_{ig}$ are independent and identically distributed across units, conditional on $A_g$.

(iii) $n_g = n$ for all $g = 1, \ldots, G$.

(iv) $|A_n| = O(G(n + 1)\overline{\pi}_n)$, as $G \to \infty$ and $n \to \infty$.

Part (i) in Assumption 3 states that the researcher has access to a sample of $G$ independent groups. Part (ii) requires that the outcomes have the same distribution within a group, and are independent conditional on the vector of effective treatment assignments. This assumption rules out the presence of within-group correlations or group-level random effects, but can be relaxed to arbitrary covariance structures when the group size is fixed using standard cluster variance estimators, as discussed later. Part (iii) imposes that all groups have equal size. When groups may have different sizes (for example, households with 3, 4 or 5 siblings), the analysis can be performed separately for each group size. Section A4 of
the supplemental appendix further discusses the case of unequally-sized groups. Finally, part (iv) requires that the total number of parameters do not 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 $1_{ig}(a) = 1(A_{ig} = a)$, $N_g(a) = \sum_{i=1}^{n+1} 1_{ig}(a)$ and $N(a) = \sum_{g=1}^{G} N_g(a)$, so that $N_g(a)$ is the total number of observations receiving effective assignment $a$ in group $g$ and $N(a)$ is the total number of observations receiving effective assignment $a$ in the sample. The estimator for $\mu(a)$ is defined as:

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

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

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

**Assumption 4 (Moments)**

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

Then we have the following result.

**Theorem 1 (Effective sample size)** Suppose Assumptions 3 and 4 hold, and consider an assignment mechanism $\pi(\cdot)$ such that $\pi(a) > 0$ for all $a \in \mathcal{A}_n$. If

$$\frac{\log |\mathcal{A}_n|}{G_{\Sigma_n}^2} \to 0 \quad (8)$$

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

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

Theorem 1 says that, under condition (8), the number of observations in the smallest cell will go to infinity, which implies that all the estimators are well defined asymptotically. Hence, condition (8) 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

$$\hat{\sigma}^2(a) = \frac{\sum_{g=1}^{G} \sum_{i=1}^{n+1} (Y_{ig} - \hat{\mu}(a))^2 1_{ig}(a)}{N(a)} 1(N(a) > 0)$$
be the standard error estimators. Then we have the following result.

**Theorem 2 (Consistency and asymptotic normality)** Under the conditions of Theorem 1,

\[
\max_{a \in A_n} |\hat{\mu}(a) - \mu(a)| = O_P \left( \sqrt{\frac{\log |A_n|}{G(n+1) \pi_n}} \right),
\]

\[
\max_{a \in A_n} |\hat{\sigma}^2(a) - \sigma^2(a)| = O_P \left( \sqrt{\frac{\log |A_n|}{G(n+1) \pi_n}} \right),
\]

and

\[
\max_{a \in A_n} \sup_{x \in \mathbb{R}} \left| \mathbb{P} \left[ \frac{\hat{\mu}(a) - \mu(a)}{\sqrt{V[\hat{\mu}(a) | A]}} \leq x \right] - \Phi(x) \right| = O \left( \frac{1}{\sqrt{G(n+1) \pi_n}} \right)
\]

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

Equation (9) 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 \( \sqrt{\log |A_n|/(G(n+1) \pi_n)} \). The denominator in this rate can be seen as the effective sample size in the smallest cell, whereas the numerator is a penalty for having an increasing number of parameters. Equation (10) bounds the difference between the distributions of the standardized potential outcomes estimators and the standard normal distribution, uniformly over the treatment assignments. Under condition (8), \( G(n+1) \pi_n \to 0 \), which gives asymptotic normality. Furthermore, this bound also reveals the rate at which the distribution of the standardized estimator approaches the standard normal, namely, \( \sqrt{G(n+1) \pi_n} \), where \( G(n+1) \pi_n \) is the minimum expected number of observations across cells, \( \min_{a \in A_n} \mathbb{E}[N(a)] \).

Importantly, both the rate of convergence and the rate of the distributional approximation depend on the assignment mechanism through \( \pi_n \), and this finding has key implications for the design of experiments to estimate spillovers, as discussed in section 4.2.

**Remark 1 (inference with many small groups).** The case of fixed group size corresponds to a setting in which the number of units per group is small compared to the total sample size, so that the effect of group size disappears asymptotically. In this context, condition (8) 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 was analyzed by Baird, Bohren, McIntosh, and Özler (2018). \( \square \)

**Remark 2 (inference for pooled parameters).** Estimation of pooled estimands as described in Section 3.1 is easily seen to satisfy the conditions for Theorems 1 and 2.
Consider for example the parameter $\Delta_{pp}$ in Equation 3. In this case, the set of effective treatment assignments can be defined as $A_n = \{(t,d) = (0,0), (1,0), (1,1)\}$ 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, $|A_n| = 3$ and hence the number of parameters and the probabilities of each assignment do not change with group size. Hence, condition (8) holds when $G \to \infty$, and $\Delta$ from Equation 3 can be estimated at rate $\sqrt{G(n+1)}$. In fact, because the number of parameters is fixed, these pooled estimands can be consistently estimated even in cases in which $G$ is fixed and $n \to \infty$, at least when observations are assumed to be iid. This asymptotic approximation may be more appropriate for studies with a small number of large groups, such as cases in which the groups are villages or large geographical units (see e.g. Ichino and Schindeln, 2012; Giné and Mansuri, 2018; Crépon, Dufo, Gurgand, Rاطelot, and Zamora, 2013). Hence, pooled estimands provide a feasible alternative that is easier to estimate when the number of parameters is large, at the expense of ignoring spillover effect heterogeneity. □

4.1 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 (Wu, 1986; Mammen, 1993; Kline and Santos, 2012) 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 $\mu(a)$ is given by:

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

whenever the denominator is non-zero, where

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

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

**Theorem 3 (Wild bootstrap)** Under the conditions of Theorem 2,

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

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 4.3 using simulation data.

### 4.2 Implications for experimental design

Theorem 2 shows that the quality of the standard normal to approximate the distribution of the standardized statistic depends on the treatment assignment mechanism through $\pi_n$. The intuition behind this result is that our ability to estimate each $\mu(a)$ depends on the number of observations facing assignment $a$, and this number depends on $\pi(a)$. When the goal is to estimate all the $\mu(a)$ simultaneously, the binding factor will be the number of observations in the smallest cell, controlled by $\pi_n$. When an assignment sets a value of $\pi_n$ 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 2 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.

The results below consider two treatment assignment mechanisms. The first one, simple random assignment (SR), consists on assigning the treatment independently at the individual level with probability $\mathbb{P}[D_{ig} = 1] = p$. This mechanism is used in the experiment analyzed in the empirical illustration. 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.

When required, it will be assumed that exchangeability holds on the first 6 moments of the potential outcomes, that is, for $p = 1, \ldots, 6$, $\mathbb{E}[Y_{ig}^p(d, d_g)] = \mathbb{E}[Y_{ig}^p(d, \tilde{d}_g)]$ for any pair of vectors such that $1'_g d_g = 1'_g \tilde{d}_g$. In particular, $\mathbb{V}[Y_{ig}(d, d_g)] = \sigma^2(d, s)$ where $s = 1'_g d_g$.

#### Corollary 1 (SR)

Under simple random assignment, condition (8) holds whenever:

$$\frac{n + 1}{\log G} \to 0.$$  \hspace{1cm} (11)

#### Corollary 2 (2SR-FM)

Under a 2SR-FM mechanism, condition (8) holds whenever:

$$\frac{\log(n + 1)}{\log G} \to 0.$$  \hspace{1cm} (12)

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.

4.3 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:

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

which corresponds to the case with $\mu(0, 0) = 0.75$, $\tau = 0.13$, $\theta_s(0) = 0.12$ for all $s$ and $\theta_s(1) = 0$ 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 $\theta_n(0) = \mathbb{E}[Y_{ig}(0, n)] - \mathbb{E}[Y_{ig}(0, 0)]$, which is the average spillover effect for an untreated units with all neighbors treated. In this simulation, $\theta_n(0) = 0.12$ 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:

$$\hat{\theta}_n(0) = \frac{\sum_g \sum_i Y_{ig} \mathbb{1}_{ig}(0, n)}{N(0, n)} - \frac{\sum_g \sum_i Y_{ig} \mathbb{1}_{ig}(0, 0)}{N(0, 0)}$$
Figure 1: Coverage rate of the 95% confidence interval.

Notes: the dashed lines show the coverage rate of the 95% confidence interval for $\hat{\theta}_n(0)$ 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.

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 $\hat{\theta}_n(0)$, calculated over the values of the simulated estimates 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.
Table 4: Simulation results, $G = 300$

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

**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.

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-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 $G = 300$. 

26
Table 5: Simulation results, $G = 600$

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

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.
5 Including covariates

There are several reasons why one may want to include covariates when estimating direct and spillover effects. First, pre-treatment characteristics may help reduce the variability of the estimators and decrease small-sample bias, which is standard practice when analyzing randomly assigned programs. Covariates can also help get valid inference when the assignment mechanisms stratifies on baseline covariates (Bugni, Canay, and Shaikh, 2018). This can be done by simply augmenting Equation (4) with a vector of covariates $\gamma'x_{ig}$ which can vary at the unit or at the group level. The covariates can also be interacted with the treatment assignment indicators to explore effect heterogeneity across observable characteristics (for example, by separately estimating effects for males and females.

Second, exogenous covariates can be used to relax the mean-independence assumption in observational studies. More precisely, if $X_g$ is a matrix of covariates, a conditional mean-independence assumption would be

$$E[Y_{ig}(d, d_g)|X_g, D_g] = E[Y_{ig}(d, d_g)|X_g]$$

which is a version of the standard unconfoundeness condition. The vector of covariates can include both individual-level and group-level characteristics.

Third, the exchangeability assumption can be relaxed by assuming it holds after conditioning on covariates, so that for any pair of treatment assignments $d_g$ and $\tilde{d}_g$ with the same number of ones,

$$E[Y_{ig}(d, d_g)|X_g] = E[Y_{ig}(d, \tilde{d}_g)|X_g]$$

For example, exchangeability can be assumed to hold for all siblings with the same age, gender or going to the same school.

All the identification results in the paper can be adapted to hold after conditioning on covariates. In terms of implementation, when the covariates are discrete the parameters of interest can be estimated at each possible value of the matrix $X_g$, although this strategy can worsen the dimensionality problem. Alternatively, covariates can be included in a regression framework after imposing parametric assumptions, for example, assuming the covariates enter linearly.

6 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 the difference-in-means 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 simple linear 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 2 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 3.1 provide weighted averages of spillover effects with known (or estimable) weights, and inference on these parameters can be conducted under mild conditions, as shown in Theorem 2. 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 my setup. Future work should also formally address issues that arise frequently in empirical studies measuring spillovers, such as imperfectly measured groups or treatment missclassification. Another important issue is imperfect compliance, which is a pervasive problem in randomized evaluations. In ongoing work (Vazquez-Bare, 2019), I propose a setup with spillover effects under imperfect compliance, defining parameters of interest, providing conditions for nonparametric identification and analyzing estimation and inference.
References


