Identification and Estimation of Spillover Effects in Randomized Experiments *

Gonzalo Vazquez-Bare[†]

November 3, 2017

Job Market Paper

Abstract

This paper employs a nonparametric potential outcomes framework to study causal spillover effects in a setting where units are clustered and their potential outcomes can depend on the treatment assignment of all the units within a cluster. Using this framework, I define parameters of interest and provide conditions under which direct and spillover effects can be identified when a treatment is randomly assigned. In addition, I characterize and discuss the causal interpretation of the estimands that are recovered by two popular estimation approaches in empirical work: a regression of an outcome on a treatment indicator (difference in means) and a regression of an outcome on a treatment indicator and the proportion of treated peers (a reduced-form linear-in-means model). It is shown that consistency and asymptotic normality of the nonparametric spillover effects estimators require a precise relationship between the number of parameters, the total sample size and the probability distribution of the treatment assignments, which has important implications for the design of experiments. The findings are illustrated with data from a conditional cash transfer pilot study and with simulations. The wild bootstrap is shown to be consistent, and simulation evidence suggests a better performance compared to the Gaussian approximation when groups are moderately large relative to the sample size.

1 Introduction

Spillover effects, which occur when an agent's actions and 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), peer effects (Bramoullé, Djebbari, and Fortin, 2009; Epple and Romano, 2011; Sacerdote, 2014), networks (Graham, 2015; de Paula, 2016), games with multiple equilibria (de Paula, 2013; Kline and Tamer, forthcoming), design of experiments (Duflo and Saez, 2003; Hirano and Hahn, 2010; Baird, Bohren, McIntosh, and Özler, forthcoming), 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, forthcoming). 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 of a randomly assigned treatment, 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, direct and spillover effects are defined based on these potential outcomes. I discuss an interpretable restriction, *exchangeability*, according to 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, and discuss to what extent this property reduces the number of spillover effects of interest. Identification of the parameters of interest when the treatment is randomly assigned is analyzed in Section 3. This framework highlights that 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 of spillover effects. Section 4 provides general conditions that ensure uniform consistency and asymptotic normality of the 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 have on the asymptotic properties of nonparametric estimators. More precisely, consistency and asymptotic normality require two main conditions that are formalized in the paper: (i) the number of parameters 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 and treatment assignment mechanisms. 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. Section 5 presents a simulation setting that studies the performance of spillover effects estimators under simple and two-stage random assignment.

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 account for spillover effects using a difference in means or accounting for them in restrictive ways as in a linear-in-means model. Finally, Section 6 discusses some extensions and ongoing work related to the inclusion of covariates, imperfect compliance and experimental design. Section 7 concludes.

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 assump-

tion 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 3.1, 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 (forthcoming), among others. Hirano and Hahn (2010) and Baird, Bohren, McIntosh, and Özler (forthcoming) study experimental design under two-stage random assignment.

A common feature in this portion of the literature is that spillover effects are defined by the experimental design. For examples, 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 identified under mild restrictions on the assignment mechanism without the need of any specific experimental design. Furthermore, 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, forthcoming). 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.

A fourth strand of the literature on peer effects relates to statistics and epidemiology, and focuses on causal inference and two-stage randomization 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; Rigdon and Hudgens, 2015; Liu and Hudgens, 2014; Basse and Feller, forthcoming). A growing related literature studies interference in a setting that replaces a partial interference assumption with more general network structures (Eckles, Karrer, and Ugander, 2017; Leung, 2017; Athey, Eckles, and Imbens, forthcoming; 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 the program "Subsidios Condicionados a la Asistencia Escolar" in two localities in Bogotá, San Cristóbal and Suba. The program aimed at increasing student retention and reducing drop-out and child labor. The experiment consisted of a conditional cash transfer with three treatment arms:

1. Basic: participants receive 30,000 pesos per month conditional on attending at least 80 percent of the days of the month.

2. Savings: participants are paid two thirds of the 30,000 pesos on a bi-monthly basis, conditional on attendance. The remaining 10,000 pesos are held in a bank account and made available during the period in which students prepare to enroll for the next school year (not conditional on attendance).

3. Tertiary: participants are paid two thirds of the 30,000 as in the savings treatment. Upon graduating, students receive 600,000 pesos immediately if they enroll in a tertiary institution, or one year later if they fail to enroll.

Eligible registrants in San Cristóbal, ranging from grade 6-11, were randomly assigned between the control status and first two treatment arms. The tertiary treatment was evaluated separately in Suba, where students were randomly assigned between control and tertiary

Table 1: Distribution of household size				Table 2: Treated per household	

	Frequency
1	3519
2	1171
3	150
4	14
5	1
Total	4855

	Frequency
0	1459
1	2815
2	528
3	50
4	3
Total	4855

treatment. 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 data contains attendance data and was conducted in the household.

As shown in Table 1, 1,336 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 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 will be 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. Some examples could be students in classrooms, persons in villages, family members in households or firms in industrial clusters. 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 can 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 $d \in \{0, 1\}$. For each unit, the vector $\mathbf{D}_{(i)g} = (D_{1ig}, D_{2ig}, \ldots, D_{n_g ig})$ will collect the treatment assignments of that unit's neighbors, so that D_{jig} is the treatment indicator corresponding to unit i's j-th neighbor. 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 $Y_{ig}(d, \mathbf{d}_g)$. Throughout the paper, I will assume that all the required moments of the potential outcome 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 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 will discuss the possibility of imperfect compliance in Section 6. In what follows, $\mathbf{0}_g$ and $\mathbf{1}_g$ will denote n_g-dimensional vectors of zeros and ones, respectively. The observed potential outcome can be written without loss of generality 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 containing 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 $2^{n_g+1} = 8$, giving a total of $\binom{2^{n_g+1}}{2} = 28$ possible treatment effects that can be defined at the individual level. 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 outcomes of units within a group have the same distribution of potential outcomes, so that in particular $\mathbb{E}[Y_{ig}(d, \mathbf{d}_g)]$ does not depend on i or g.[1]

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

---

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

identities requires a natural labeling or ordering between units in each group, which can be given for example by (i) distance according to some specified metric or (ii) "type" of the relationship. A natural example of (i) would be geographical distance that orders neighbors from closest to farthest, for instance, if units are schools in counties and neighbor 1 is the closest school, neighbor 2 is second closest school, etc. Another example would be the case where one can rank the relationships according to its strength, e.g. closest friend, second-closest friend, etc. An example of (ii) would be kinship: in this case, neighbors would be mother, father, youngest sibling, oldest sibling, etc. An advantage of this approach is that it allows the researcher to estimate the within-group network structure, that is, to identify the subset of units affecting each individual in a group (as long as the assumption of no interference between groups holds). This issue is analyzed by Manresa (2016) in linear panel data models.

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 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 $n_g + 1 = 10$ the model has 2000 observations and 1024 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* $\mathbf{d}_g, \tilde{\mathbf{d}}_g \in \mathcal{D}_g$ *such that* $\mathbf{1}'_g\mathbf{d}_g = \mathbf{1}'_g\tilde{\mathbf{d}}_g$. *Then, for each* $d = 0, 1$,

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

Assumption 1 states that the average potential outcome is invariant to permutations of the neighbor's assignment vector $\mathbf{d}_g$. Several studies have considered stronger versions of this assumption (see e.g. Hudgens and Halloran, 2008; Manski, 2013; Leung, 2017). The main difference between this assumption and similar restrictions used in the literature is that Assumption 1 only imposes exchangeability on the first moment of the potential outcome, and not on the potential outcome function itself. On the other hand, the result in Lemma 1

below is sometimes stated as an assumption (see e.g. Baird, Bohren, McIntosh, and Özler, forthcoming; Ferracci, Jolivet, and van den Berg, 2014) without explicitly stating the restrictions on the potential outcomes that this condition requires. The condition 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), as discussed later.

Exchangeability implies the following restriction on the potential outcome.

Lemma 1 (Potential outcome under exchangeability) *For any* $\mathbf{d}_g \in \mathcal{D}_g$, *let* $s := \mathbf{1}_g' \mathbf{d}_g = \sum_{j=1}^{n_g} d_j$. *Under Assumption 1, for* $d = 0, 1$, *there is a function* $\mu(d, \cdot) : \{0, 1, \ldots, n_g\} \to \mathbb{R}$ *such that* $\mathbb{E}[Y_{ig}(d, \mathbf{d}_g)] = \mu(d, s)$.

Lemma 1 states that, for each unit i in group g, the average potential outcome only depends on the neighbors' assignment $\mathbf{d}_g$ 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 $\mathbf{d}_g$ and $\tilde{\mathbf{d}}_g$ such that $\mathbf{1}_g' \mathbf{d}_g = \mathbf{1}_g' \tilde{\mathbf{d}}_g$, exchangeability restricts the average spillover effect to zero, that is,

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

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

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".

The plausibility of the exchangeability assumption also depends on the definition of groups, since social interactions are often endogenous results of individual decisions; in other words, being members of the same group does not imply that two individuals will interact. By changing the definition of group from, say, classroom to group of friends, exchangeability may be more likely to hold. There is a growing literature studying endogenous networks; see for example Christakis, Fowler, Imbens, and Kalyanaraman (2010), Goldsmith-Pinkham and Imbens (2013), Graham (2015), Chandrasekhar (2016), de Paula (2016) and Graham (2017). I will not discuss these issues in this paper, and I assume that groups are known and their size and composition are unaffected by the treatment.

The exchangeability assumption will be maintained throughout the rest of the paper to conserve space, leaving different alternatives to this assumption for Section A2 of the Supplemental Appendix. I will define two sets of parameters of interest. First, the *average direct effect* of the treatment given s treated neighbors will be defined as:

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

so each τ_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:

$$\theta_s(d) = \mu(d, s) - \mu(d, 0) \tag{2}$$

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. 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 τ_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 3.1 will link these parameters to the estimands of the difference in means and the linear-in-means regression.

3 Identifying spillover effects

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 (Independence) *For all* $(d, \mathbf{d}_g) \in \{0, 1\} \times \mathcal{D}_g$ *and for all i and g,*

$$Y_{ig}(d, \mathbf{d}_g) \perp\!\!\!\perp (D_{ig}, \mathbf{D}_{(i)g})$$

This condition states that potential outcomes are independent of the treatment assignment vector, and rules out selection into treatment. Under SUTVA, this condition reduces to $(Y_{ig}(0), Y_{ig}(1)) \perp\!\!\!\perp D_{ig}$, which means for example that the average potential outcome under no treatment is equal between treated and control units. In 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 type of independence requires, for example, that the average potential outcomes that would be observed in absence of treated units have to coincide between groups in which nobody is treated and in groups in which at least some units are treated.

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

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

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.

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, $\mathbb{P}[D_{ig} = d, S_{ig} = s] > 0$. In 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 $\mathbb{P}[D_{ig} = 1] \in (0, 1)$. In 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 $\mathbb{P}[D_{ig} = 1, S_{ig} = n_g] = \mathbb{P}[D_{ig} = 0, S_{ig} = 0] = 1/2$ and $\mathbb{P}[D_{ig} = d, S_{ig} = s] = 0$ 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 (forthcoming), among others, are not required for identification purposes (although they can improve estimation and inference, as discussed in Section 4).

Lemma 2 provides a straightforward way to identify both direct and spillover effects. More precisely, we have that:

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

and

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

In terms of the empirical implementation, there are two ways to estimate these average treatment effects. The first one is to construct outcome sample means for each cell defined by

Table 3: Estimation results

	Diff. Means		Linear-in-Means				Full	
	coef	s.e.	coef	s.e.	coef	s.e.	coef	s.e.
D_{ig}	-0.018	0.019	-0.018	0.019	0.072	0.048	0.127*	0.068
$\bar{D}_g^{(i)}$			-0.020	0.040				
$\bar{D}_g^{(i)}(1-D_{ig})$					0.120*	0.072		
$\bar{D}_g^{(i)}D_{ig}$					-0.088*	0.050		
$\mathbb{1}(S_{ig}=1)(1-D_{ig})$							0.128**	0.063
$\mathbb{1}(S_{ig}=2)(1-D_{ig})$							0.119*	0.061
$\mathbb{1}(S_{ig}=1)D_{ig}$							-0.026	0.025
$\mathbb{1}(S_{ig}=2)D_{ig}$							-0.057*	0.030
Constant	0.857***	0.013	0.866***	0.026	0.808***	0.039	0.751***	0.060
Observations		363		363		363		363

Notes: Cluster-robust s.e. Regressions include school FE. ***$p < 0.01$,**$p < 0.05$,*$p < 0.1$.

the assignments (d, s), and then construct the differences between estimated average potential outcomes. Equivalently, we can consider a saturated linear-in-parameters nonparametric regression of the form:

$$\mathbb{E}[Y_{ig}|D_{ig}, S_{ig}] = \alpha + \tau_0 D_{ig} + \sum_{s=1}^{n_g} \theta_s(0)\mathbb{1}(S_{ig} = s)(1 - D_{ig}) + \sum_{s=1}^{n_g} \theta_s(1)\mathbb{1}(S_{ig} = s)D_{ig} \quad (3)$$

where $\alpha = \mu(0,0)$, which provides identical point estimates and standard errors as the ones estimating cell means after accounting for heteroskedasticity. Because it is equivalent to estimating averages at each cell separately, Equation (3) does not impose any parametric assumptions. The total number of parameters in this regression is $2(n_g + 1)$, so the number of coefficients equals the number of average potential outcomes that we need to estimate. I will discuss this issue in detail in Section 4.

3.1 Empirical results

This section employs the data from the experiment described in Section 2 to illustrate the above results. Barrera-Osorio, Bertrand, Linden, and Perez-Calle (2011) analyzed spillover effects on enrollment and attendance in households with two registered siblings pooling the three treatment arms (Table 9 in their paper). The authors find negative and statistically significant spillover effects of about 3 percentage points for attendance and 7 percentage points on enrollment, which they interpret as suggestive evidence that the conditional cash transfer drives family resources toward treated children and away from their untreated siblings. In this section, I will analyze direct and spillover effects in households with three registered siblings, which gives a larger number of possible treatment effects. The outcome of interest will be attendance.

The results are obtained by estimating Equation (3). The estimates are shown in the right panel of Table 3. These estimates suggest a positive direct effect of the treatment of about

12.7 percentage points, significant at the 10 percent level, with equally large spillover effects on the untreated units. More precisely, the estimated effect on an untreated kid of having one treated sibling is 12.8 percentage points, while the effect of having two treated siblings is 11.9 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 smaller in magnitude and negative, with the effect on a treated kid of having two treated siblings being significant at the 10 percent level. 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 $\hat{\tau}_0 + \hat{\theta}_2(1) \approx 0.07$. Thus, a treated kid with two treated siblings increases her attendance in 7 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.

3.2 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 β_{D} in the model:

$$Y_{ig} = \alpha_{\mathsf{D}} + \beta_{\mathsf{D}} D_{ig} + u_{ig} \tag{4}$$

where $\beta_{\mathsf{D}} = \mathbb{E}[Y_{ig}|D_{ig} = 1] - \mathbb{E}[Y_{ig}|D_{ig} = 0]$. 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 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 3 (Difference in means) *Under the conditions for Lemma 2, the coefficient β_D*

from Equation (4) can be written as:

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

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, which is the one used in this application, 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:

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

The effect of the presence of spillovers in the difference in means, captured by the term $\sum_{s=1}^{n_g}(\theta_s(1) - \theta_s(0))\mathbb{P}[S_{ig} = s]$, 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, $\theta_s(0) = \theta_s(1)$ for all s, then the difference in means β_{D} equals the direct effect of the treatment without treated siblings, τ_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 $\hat\theta_1(1) - \hat\theta_1(0) = -0.155$ and $\hat\theta_2(1) - \hat\theta_2(0) = -0.176$. Therefore, the spillover effects push the difference in means towards zero in this case.

3.3 Linear-in-means models

Equation (4) 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, which is given by:

$$Y_{ig} = \alpha_\ell + \beta_\ell D_{ig} + \gamma_\ell \bar{D}_g^{(i)} + \eta_{ig}, \qquad \bar{D}_g^{(i)} = \frac{1}{n_g}\sum_{j \neq i} D_{jg} \tag{5}$$

that is, a regression of the outcome on a treatment indicator and the proportion of treated neighbors. In this specification, γ_ℓ 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 5 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 restuls using Equation 3. To better understand this point, Equation (3) 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 κ, then Equation (3) 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 none of the LIM assumptions hold in this case: the spillover effects are significantly different for treated and controls, and they are decreasing in the number of treated siblings. More in general, the following result holds.

Lemma 4 (LIM regression) *Under the conditions for Lemma 2 and simple random assignment, the coefficient γ_ℓ from Equation (5) can be written as:*

$$\gamma_\ell = n_g \sum_{s=1}^{n_g} [\theta_s(0)(1-p) + \theta_s(1)p] \frac{\mathbb{C}ov(S_{ig}, \mathbb{1}(S_{ig} = s))}{\mathbb{V}[S_{ig}]}$$
$$= n_g \sum_{s=1}^{n_g} [\theta_s(0)(1-p) + \theta_s(1)p] \left(\frac{s - \mathbb{E}[S_{ig}]}{\mathbb{V}[S_{ig}]} \right) \mathbb{P}[S_{ig} = s]$$

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

This results shows that γ_ℓ captures a rather complicated linear combination of all the spillover effects under treatment and control. More precisely, γ_ℓ 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 $\hat{\theta}_1(0)(1-\hat{p}) + \hat{\theta}_1(1)\hat{p} \approx 0.027$ and $\hat{\theta}_2(0)(1-\hat{p}) + \hat{\theta}_2(1)\hat{p} \approx 0.003$. On the other hand, $\hat{\mathbb{E}}[S_{ig}] = 1.3$, so $\hat{\gamma}_\ell$ will assign negative weight to the first term and positive weight to the second one, resulting in the negative -0.02 shown in Table 3. These results suggest that the LIM model is in general sensitive to misspecification and may give a poor summary of spillover effects when the assumptions that justify it are violated.

A straightforward way to make Equation (5) 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:

$$Y_{ig} = \alpha_\ell + \beta_\ell D_{ig} + \gamma_\ell^0 \bar{D}_g^{(i)}(1 - D_{ig}) + \gamma_\ell^1 \bar{D}_g^{(i)} D_{ig} + \xi_{ig} \tag{6}$$

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 can see that, for $d = 0, 1$,

Lemma 5 (LIM with interactions) *Under the conditions for Lemma 2 and simple random assignment, for $d = 0, 1$ the coefficients γ_ℓ^d can be written as:*

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

Thus, the only difference is that each γ_ℓ^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 γ_ℓ^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, which can be seen in Table 3 from the fact that -0.088 is not a weighted average of -0.026 and -0.057. The similarity between the estimates in this case seems to be coincidental.

In sum, the empirical results in this section reveal how the saturated regression given by Equation (3) is a fully nonparametric yet easily implemented regression-based strategy that recovers all the treatment effects of interest. On the other hand, both the difference-in-means regression and the linear-in-means regression impose strong restrictions on the spillover effects that may be violated in many contexts, and can be sensitive to misspecification.

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, which assuming all groups are equally-sized, is $2(n_g + 1)$.[2] 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

---

[2]As 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 A5 of Supplemental Appendix.

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.

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* for unit i in group g, that is, the set of assignments that are possible under a certain potential outcome model. Denote by $|\mathcal{A}_n|$ the cardinality of $\mathcal{A}_n$. For example, under SUTVA (no spillovers), $\mathcal{A}_n = \{0,1\}$. When exchangeability holds, $\mathcal{A}_n = \{(d,s) : d \in \{0,1\}, s \in \{0,1,\ldots,n_g\}\}$ and $|\mathcal{A}_n| = 2(n_g + 1)$. In the general case without assuming exchangeability, $\mathcal{A}_n = \{(d,\mathbf{d}_g) : d \in \{0,1\}, \mathbf{d}_g \in \{0,1\}^{n_g}\}$ and $|\mathcal{A}_n| = 2^{n_g+1}$. Hence, the less restrictive the potential outcome modeling assumptions, the larger the number of average potential outcomes that need to be estimated. The observed effective assignment for unit i in group g is denoted by $\mathbf{A}_{ig}$, taking values in the set $\mathcal{A}_n$. The potential outcome under effective assignment $\mathbf{a}$ is given by $Y_{ig}(\mathbf{a})$ with expected value $\mathbb{E}[Y_{ig}(\mathbf{a})] = \mu(\mathbf{a})$, and the observed outcome is $Y_{ig} = Y_{ig}(\mathbf{A}_{ig})$.

Second, 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, 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, $|\mathcal{A}_n| \to \infty$ and $\pi(\mathbf{a}) \to 0$ for all $\mathbf{a}$. 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 let $\mathbf{y}_g(\mathbf{a}_g) = (Y_{1g}(\mathbf{a}_{1g}), Y_{2g}(\mathbf{a}_{2g}), \ldots, Y_{n_g+1,g}(\mathbf{a}_{n_g+1,g}))'$ be the vector of potential outcomes in group g. I will assume the following sampling scheme.

Assumption 3 (Sampling and design)

(i) For $g = 1, \ldots, G$, $(\mathbf{y}_g(\mathbf{a}_g)', \mathbf{A}_g')$ is a random sample.

(ii) *Within each group g, the potential outcomes $Y_{ig}(\mathbf{a})$ are independent and identically distributed across units for all $\mathbf{a} \in \mathcal{A}_n$, conditional on $\mathbf{A}_g$.*

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

(iv) *$|\mathcal{A}_n| = O(G(n+1)\underline{\pi}_n)$, 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. As usual, potential outcomes are only observed for the realized treatment assignments, so the vector of observed variables is $(\mathbf{Y}_g', \mathbf{A}_g')$ where $\mathbf{Y}_g = \mathbf{y}_g(\mathbf{A}_g)$. Part (ii) requires that the potential outcomes have the same distribution within a group, and are independent conditional on the vector of 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 A5 of the Supplemental Appendix further discusses the case of unequally-sized groups. Finally, part (iv) requires that the total number parameters does 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 $\mathbb{1}_{ig}(\mathbf{a}) = \mathbb{1}(\mathbf{A}_{ig} = \mathbf{a})$, $N_g(\mathbf{a}) = \sum_{i=1}^{n+1} \mathbb{1}_{ig}(\mathbf{a})$ 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.

Assumption 4 (Moments)

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

Then we have the following result.

Theorem 1 (Effective sample size) *Suppose Assumptions 2, 3 and 4 hold, and consider an assignment mechanism $\pi(\cdot)$ such that $\pi(\mathbf{a}) > 0$ for all $\mathbf{a} \in \mathcal{A}_n$. If*

$$
\frac{\log |\mathcal{A}_n|}{G \underline{\pi}_n^2} \to 0 \tag{7}
$$

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

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

Theorem 1 says that, under condition (7), the number of observations in the smallest cell will go to infinity, which implies that all the potential outcome estimators are well defined asymptotically. 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. Condition (7) 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. Because this condition implies part 1 of the above theorem, it is a low-level condition that justifies the assumption of invertibility of the design matrix (see e.g. Assumption 2 in Cattaneo, Jansson, and Newey, forthcoming).

Next, let

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

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

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

$$\max_{\mathbf{a}\in\mathcal{A}_n} |\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})| = O\left(\sqrt{\frac{\log|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n}}\right),$$

$$\max_{\mathbf{a}\in\mathcal{A}_n} \left|\hat{\sigma}^2(\mathbf{a}) - \sigma^2(\mathbf{a})\right| = O\left(\sqrt{\frac{\log|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n}}\right),$$

(8)

and

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

(9)

where $\Phi(x)$ is the cdf of a standard Gaussian random variable and where the exact form of C is given in the Supplemental Appendix.

Equation (8) 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 effective sample size in the smallest cell, whereas the numerator is a penalty for having an increasing number of parameters. Equation (9) bounds the difference between the distributions of the standardized potential outcomes estimators and the standard normal distribution, uniformly over the treatment assignments. Under condition (7), $G(n+1)\underline{\pi}_n \to 0$, which gives asymptotic normality. However, this bound also reveals the rate at which the distribution of

the standardized estimator approaches the standard normal, namely, $\sqrt{G(n+1)\underline{\pi}_n}$, where $G(n+1)\underline{\pi}_n$ is the minimum expected number of observations across cells, $\min_{\mathbf{a} \in \mathcal{A}_n} \mathbb{E}[N(\mathbf{a})]$.

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 key implications for the design of experiments to estimate spillovers, as discussed in section 4.2.

Remark. 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 (7) 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 (forthcoming).

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 $1/2$ independent 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 3 (Wild bootstrap) *Under the conditions of Theorem 2,*

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

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

using simulation data.

4.2 Implications for experimental design

Theorems 2 shows that the appropriateness of the standard normal to approximate the distribution of the standardized statistic approximation depends on the treatment assignment mechanism through $\underline{\pi}_n$. The intuition behind this result is that our ability to estimate each $\mu(\mathbf{a})$ depends on the number of observations facing assignment $\mathbf{a}$, and this number depends on $\pi(\mathbf{a})$. Since in principle all average potential outcomes are equally important, 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 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 (forthcoming), 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 $k = 1, \ldots, 6$, $\mathbb{E}[Y_{ig}^p(d, \mathbf{d}_g)] = \mathbb{E}[Y_{ig}^p(d, \tilde{\mathbf{d}}_g)]$ for any pair of vectors such that $\mathbf{1}_g' \mathbf{d}_g = \mathbf{1}_g' \tilde{\mathbf{d}}_g$. In particular, $\mathbb{V}[Y_{ig}(d, \mathbf{d}_g)] = \sigma^2(d, s)$ where $s = \mathbf{1}_g' \mathbf{d}_g$.

Corollary 1 (SR) *Under simple random assignment, condition (7) holds whenever:*

$$\frac{n+1}{\log G} \to 0 \tag{10}$$

Corollary 2 (2SR-FM) *Under a 2SR-FM mechanism, condition (7) holds whenever:*

$$\frac{\log(n+1)}{\log G} \to 0 \tag{11}$$

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, forthcoming).

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 $(1,n)$ decrease exponentially with group size, and thus they become very small very rapidly.

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.

The superior performance of the 2SR-FM compared to SR, formalized by the difference in convergence rates, relies crucially on the fact that we aim at estimating all the average potential outcomes simultaneously. Focusing on all the average potential outcomes is an agnostic approach that does not place any restrictions or priors on the different direct and spillover effects. This approach extracts all the information related to spillover effects, and can be used to test a wide array of hypotheses like the absence of spillovers for treated units, linearity of spillovers, tipping points, etc.

In practice, however, it is possible that the researcher wants to focus on a subset of parameters or a function thereof, such as the spillover effect of having half the neighbors treated. These alternative choices have different implications in terms of designs. In ongoing work (Vazquez-Bare, 2017a), I study how different objective functions can be translated into different experimental designs. I briefly discuss these issues in Section 6.

5 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:

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

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

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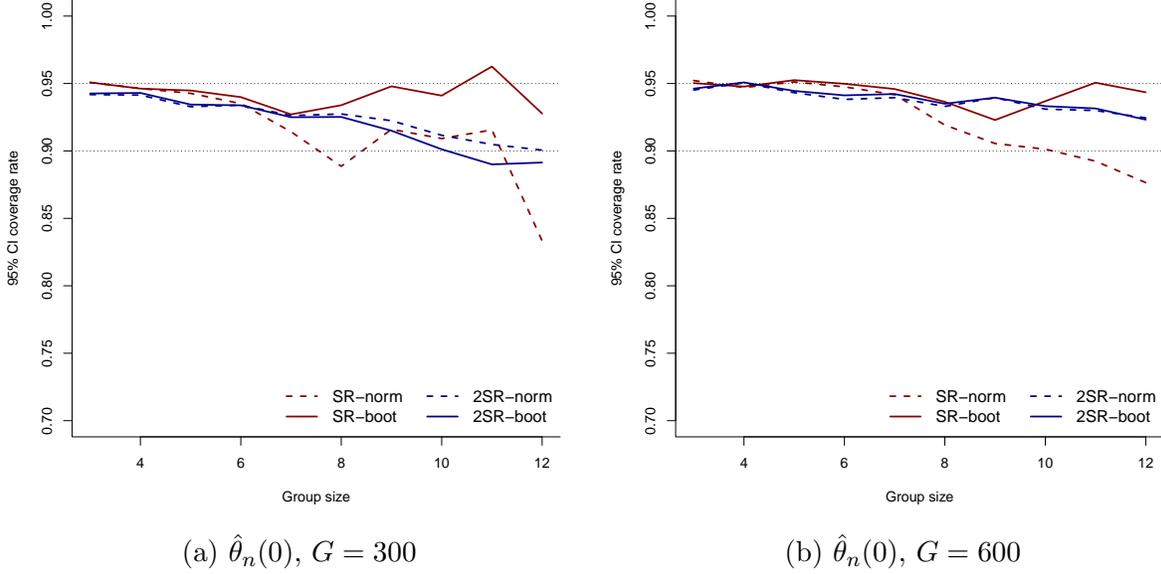
Notes: the dashed lines show the coverage rate of the 95% confidence interval for $\theta_n(0)$ based on the normal approximation under simple random assignment (red line) and two-stage randomization (blue line) for a sample with 300 (left) and 600 (right) groups. The solid lines show the coverage rates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

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:

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

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 row shows the coverage rate of a 95% confidence interval based on the Gaussian approximation. Finally, the last 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 $n = 3$ 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 90% under 2SR-FM and 83% under SR. 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 98% 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.

Table ?? show the same results but for the wild bootstrap approach. Under simple random assignment, the wild bootstrap confidence interval achieves better coverage compared to the one based on the Gaussian approximation (92.7 versus 83.3), although both the normal-based and the bootstrap-based confidence intervals perform similarly under 2SR. These results are also illustrated in Figure 1.

6 Extensions

6.1 Including covariates

Because estimation in this paper can be performed using linear regressions, the inclusion of exogenous covariates is straightforward. 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, forthcoming). This can be done by simply augmenting Equation (3) with a vector of covariates $\gamma' \mathbf{x}_{ig}$ 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

Table 4: Simulation results, $G = 300$ - normal approximation

	$n = 2$	$n = 5$	$n = 8$	$n = 11$
Simple rand.				
$(n+1)/\log(G)$	0.5260	1.0519	1.5779	2.1039
Bias	0.0000	0.0027	-0.0027	-0.0277
Variance	0.0027	0.0128	0.0427	0.0673
95% CI coverage	0.9508	0.9351	0.9160	0.8333
Prop. empty cells	0.0000	0.0117	0.5675	0.9860
Two-stage rand.				
$\log(n+1)/\log(G)$	0.1926	0.3141	0.3852	0.4357
Bias	0.0009	-0.0017	-0.0021	0.0018
Variance	0.0024	0.0035	0.0047	0.0058
95% CI coverage	0.9418	0.9340	0.9224	0.9006
Prop. empty cells	0.0000	0.0000	0.0000	0.0000

Table 5: Simulation results, $G = 300$ - wild bootstrap

	$n = 2$	$n = 5$	$n = 8$	$n = 11$
Simple rand.				
$(n+1)/\log(G)$	0.5260	1.0519	1.5779	2.1039
Bias	0.0000	0.0027	-0.0026	-0.0273
Variance	0.0027	0.0121	0.0329	0.0337
95% CI coverage	0.9508	0.9399	0.9479	0.9277
Prop. empty cells	0.0000	0.0117	0.5675	0.9860
Two-stage rand.				
$\log(n+1)/\log(G)$	0.1926	0.3141	0.3852	0.4357
Bias	0.0009	-0.0017	-0.0021	0.0018
Variance	0.0023	0.0034	0.0045	0.0055
95% CI coverage	0.9426	0.9338	0.9150	0.8914
Prop. empty cells	0.0000	0.0000	0.0000	0.0000

heterogeneity across observable characteristics (for example, by separately estimating effects for males and females), although this strategy can decrease precision.

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

$$\mathbb{E}[Y_{ig}(d, \mathbf{d}_g)|\mathbf{X}_g, D_{ig}, \mathbf{D}_g^{(i)}] = \mathbb{E}[Y_{ig}(d, \mathbf{d}_g)|\mathbf{X}_g]$$

which is a version of the standard unconfoundeness condition (see e.g. Imbens, 2004). 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 $\mathbf{d}_g$ and $\tilde{\mathbf{d}}_g$ with the same number of ones,

$$\mathbb{E}[Y_{ig}(d, \mathbf{d}_g)|\mathbf{X}_g] = \mathbb{E}[Y_{ig}(d, \tilde{\mathbf{d}}_g)|\mathbf{X}_g]$$

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 $\mathbf{X}_g$, 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.

6.2 Imperfect compliance

Imperfect compliance, which occurs when units may not comply with their treatment assignments, is a pervasive problem in treatment and policy evaluation. The main problem generated by imperfect compliance is that, even when treatment assignment is random, actual treatment status can be endogenous. In absence of spillovers, Imbens and Angrist (1994) showed that an instrumental variables approach that instruments endogenous treatment status with exogenous treatment assignment can recover a local average treatment effect, which is the average effect on the subpopulation of units that are pushed to receive the treatment by the instrument, known as compliers.

The presence of imperfect compliance introduces several complexities in the analysis of spillover effects. When individuals may not comply with their treatment assignment, spillovers can occur at two different stages. To fix ideas, consider the experiment conducted by Duflo and Saez (2003). In this study, the authors randomly selected employees in different departments from a large research university. The selected employees received a letter and a monetary incentive to attend a benefits fair describing retirement plans. The first stage at which spillovers can occur is treatment take-up. For example, employees that did not

receive the incentive to attend the fair can still decide to attend if their selected coworkers share information with them. The second stage is the outcome stage, which is the one that was analyzed in this paper: the decision to enroll in a retirement plan may depend not only on whether a person attends the fair, but also on whether her coworkers do. These two different stages of spillovers can rapidly increase the dimensionality of the problem. For example, if each person has only one coworker, in addition to the four possible potential outcomes depending on the observed D_{ig} and D_{jg}, we also have four possible treatment statuses, $D_{ig}(0,0), D_{ig}(1,0), D_{ig}(0,1)$ and $D_{ig}(1,1)$, depending on whether individual i and her coworker receive the letter or not.

In this setting, the usual classification of always-takers, never-takers, compliers and defiers, depending on how own treatment status varies with treatment assignment, no longer partitions the population. For example, an individual may decide to comply with her treatment assignment when her coworker in assigned to control, $D_{ig}(1,0) = 1$ and $D_{ig}(0,0) = 0$, but may always take the treatment when her coworker is treated, $D_{ig}(1,1) = D_{ig}(0,1) = 1$. Thus, individual i is a complier when her coworker is assigned to control, but an always-taker when her coworker is assigned to treatment. For this reason, there are multiple local average treatment effects that can be defined.

Finally, the exchangeability assumption used to decrease the dimensionality of the problem may be harder to satisfy in presence of imperfect compliance. More precisely, even when the outcome only depends on the total number of treated neighbors, this number depends on the assignments of all the neighbors in the group. But the condition that the identities of the neighbors assigned to treatment are irrelevant may be very strong, since different neighbors can have different compliance statuses. In other words, assigning a neighbor that always manages to get the treatment cannot be equivalent to assigning a neighbor who always refuses the treatment.

In ongoing work (Vazquez-Bare, 2017b), I am studying a setup with spillover effects and imperfect compliance, defining parameters of interest, providing conditions to identify them and analyzing the performance of usual techniques like the one proposed by Imbens and Angrist (1994) in presence of spillovers.

6.3   Experimental design

Section 4.2 showed how a two-stage randomization design that fixes the number of treated units in each group can provide improved performance compared to simple random assignment. The intuition behind this result is that the former mechanism spreads observations more evenly across the cells defined by the treatment assignments. The improvement in terms of inference given by the two-stage design relies on the fact that we are trying to estimate all average potential outcomes simultaneously, and hence we need to worry about the probability of observing units under the least likely assignments.

More generally, the parameter of interest is a function of the vector of potential outcomes.

Letting $\mathcal{E} = (\mathbb{E}[Y_{ig}(0,0)], \dots, \mathbb{E}[Y_{ig}(1,n)])$, the parameter of interest is:

$$\beta = \beta(\mathcal{E})$$

The case analyzed in this paper corresponds to $\beta(\cdot)$ being the identity function, $\beta(\mathcal{E}) = \mathcal{E}$. More generally, $\beta(\cdot)$ can select, for example, a subset of potential outcomes, like the potential outcomes under no treatment, $\beta(\mathcal{E}) = (\mathbb{E}[Y_{ig}(0,0)], \dots, E[Y_{ig}(0,n)])$. This would be the parameter of interest if previous literature or a theoretical model suggested no spillover effects on treated units. Another possibility would be to focus on the total effect of the program, $\mathbb{E}[Y_{ig}(1,n)] - \mathbb{E}[Y_{ig}(0,0)]$, which is the effect of treating everyone in the group compared to treating nobody. In this case, $\beta(\mathcal{E}) = \mathbb{E}[Y_{ig}(1,n)] - \mathbb{E}[Y_{ig}(0,0)]$.

Different choices of $\beta(\cdot)$ have different implications for experimental design, since each choice determines the importance given to the sample size in each assignment cell. For instance, a cluster randomized trial in which the treatment is assigned at the group level may be appropriate when $\beta(\mathcal{E}) = \mathbb{E}[Y_{ig}(1,n)] - \mathbb{E}[Y_{ig}(0,0)]$, whereas a two-stage randomization with fixed margins can be more appropriate when $\beta(\mathcal{E}) = \mathcal{E}$. Combining Theorem 2 with each the information contained in $\beta(\cdot)$ gives a principled way to rank different assignment mechanisms based on the rate of convergence they imply for the estimators, as I am currently studying in Vazquez-Bare (2017a).

7   Conclusion

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 two-stage randomization designs. Finally, I study estimation and inference when the number of parameters can be large relative to the sample size, and discuss the implications of my results for experimental design.

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.

References

ABADIE, A., AND M. D. CATTANEO (forthcoming): "Econometric Methods for Program Evaluation," *Annual Review of Economics*.

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

ATHEY, S., D. ECKLES, AND G. W. IMBENS (forthcoming): "Exact P-values for Network Interference," *Journal of the American Statistical Association*.

BAIRD, S., A. BOHREN, C. McINTOSH, AND B. ÖZLER (forthcoming): "Optimal Design of Experiments in the Presence of Interference," *Review of Economics and Statistics*.

BANERJEE, A., A. G. CHANDRASEKHAR, E. DUFLO, AND M. O. JACKSON (2013): "The Diffusion of Microfinance," *Science*, 341(6144).

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 (forthcoming): "Analyzing two-stage experiments in the presence of interference," *Journal of the American Statistical Association*.

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.

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.

BUGNI, F. A., I. A. CANAY, AND A. M. SHAIKH (forthcoming): "Inference under Covariate-Adaptive Randomization," *Journal of the American Statistical Association*.

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 (forthcoming): "Inference in Linear Regression Models with Many Covariates and Heteroskedasticity," *Journal of the American Statistical Association.*

CHANDRASEKHAR, A. (2016): "Econometrics of Network Formation," in *The Oxford Handbook of the Economics of Networks*, ed. by Y. Bramoullé, A. Galeotti, and B. Rogers.

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

CHRISTAKIS, N. A., J. H. FOWLER, G. W. IMBENS, AND K. KALYANARAMAN (2010): "An Empirical Model for Strategic Network Formation," *NBER Working Paper*, (16039).

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.

——— (2016): "Econometrics of Network Models," *Cemmap working paper*, CWP06/16.

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.

ECKLES, D., B. KARRER, AND J. UGANDER (2017): "Design and Analysis of Experiments in Networks: Reducing Bias from Interference," *Journal of Causal Inference*, 5(1).

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 (forthcoming): "Together We Will: Experimental Evidence on Female Voting Behavior in Pakistan," *American Economic Journal: Applied Economics.*

GOLDSMITH-PINKHAM, P., AND G. W. IMBENS (2013): "Social Networks and the Identification of Peer Effects," *Journal of Business & Economic Statistics*, 31(3), 253–264.

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.

——— (2017): "An Econometric Model of Network Formation with Degree Heterogeneity," *Econometrica*, 85(4), 1033–1063.

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.

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. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *The Review of Economics and Statistics*, 86(1), 4–29.

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

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

KLINE, P., AND A. SANTOS (2012): "Higher order properties of the wild bootstrap under misspecification," *Journal of Econometrics*, 171(1), 54–70.

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

LEUNG, M. (2017): "Treatment and Spillover Effects under Network Interference," *Working paper*.

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.

MAMMEN, E. (1993): "Bootstrap and Wild Bootstrap for High Dimensional Linear Models," *The Annals of Statistics*, 21(1), 255–285.

MANRESA, E. (2016): "Estimating the Structure of Social Interactions Using Panel Data," *Working Paper*.

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.

RIGDON, J., AND M. G. HUDGENS (2015): "Exact confidence intervals in the presence of interference," *Statistics & Probability Letters*, 105(6), 130–135.

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.

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.

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

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. (2017a): "Design of Experiments for Estimating Spillover Effects," *In progress*.

———— (2017b): "Identification of Spillover Effects with Imperfect Compliance," *In progress*.

WU, C. F. J. (1986): "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," *The Annals of Statistics*, 14(4), 1261–1295.