Analysis of Spillover Effects in Randomized Experiments

by

Gonzalo Vazquez-Bare

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2018

Doctoral Committee:

Professor Matias D. Cattaneo, Chair
Professor Lutz Kilian
Professor Melvin Stephens Jr
Professor Rocío Titiunik

Gonzalo Vazquez-Bare

gvazquez@umich.edu

ORCID iD: 0000-0002-3175-592X

© Gonzalo Vazquez-Bare 2018

ACKNOWLEDGEMENTS

I am deeply grateful to Matias Cattaneo, who, since the first years of my PhD, has been an outstanding advisor, teacher, boss, mentor and friend. This dissertation would not have been possible without his continued support and encouragement, endless patience and invaluable advice.

I am indebted to Lutz Kilian, Mel Stephens and Rocío Titiunik, who sat through countless presentations, read multiple versions of my papers and provided thoughtful feedback and discussions. I greatly benefited from conversations with Xinwei Ma and Kenichi Nagasawa, whose excellent suggestions greatly improved my dissertation. I also thank Sebastián Calonico and Max Farrell for helpful comments.

Olga Namen deserves a special mention for her unconditional encouragement, for being an example of discipline and hard work and for helping me navigate through the many ups and downs of my PhD.

Finally, I wish to thank my friends Alberto Arredondo, Chiara Ferrero, Catalina Franco, Amelia Hawkins, Nicolás Idrobo, Ana María Martínez Baquero and Nicolás Morales for making these years a truly enjoyable experience, and my family, for their constant and wholehearted support.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

This dissertation studies identification, estimation, inference and experimental design for analyzing causal spillover effects in randomized experiments. Chapter II provides a nonparametric framework based on potential outcomes to define spillover effects in a setting in which units are clustered and their potential outcomes can depend on the treatment assignments of all units within a group. Using this framework, I provide conditions for identification of average direct and spillover effects when the treatment is randomly assigned. I then study identification under three estimation strategies that are commonly employed in empirical work: a regression of an outcome on a treatment indicator, which calculates a difference in means between treated and controls, a regression that controls for the proportion of treated peers, and a regression exploiting variation in treatment probabilities in two-stage designs.

Chapter III analyzes estimation and inference for spillover effects. I start by illustrating the results from Chapter II using two empirical applications. I then study nonparametric estimation and inference for spillover effects in a setting in which both the number of groups and the group size are allowed to grow. This setting allows me to understand the effect of the number of parameters on the asymptotic properties of the proposed nonparametric estimators. Finally, I discuss the implications of these findings for the design of experiments.

Chapter IV discusses some key issues related to the empirical implementation of the results from the previous chapters: the inclusion of covariates, identification of spillover effects in experiments with imperfect compliance and optimal design of experiments.

CHAPTER I

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; Manski, 2013b), 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; Athey and Imbens, 2017). However, the literature is still evolving in this area, and most of the available methods either assume no spillovers or allow for them in restrictive ways, without a precise definition of the parameters of interest or the conditions required to recover them.

This dissertation 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. Chapter II 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. 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 2.4. 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. I then characterize the parameters that are recovered by three 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), a regression on a treatment indicator and the proportion of treated neighbors (a reduced-form linear-in-means model) and a regression exploiting variation in treatment probability across different groups.

Chapter III analyzes nonparametric estimation and inference for average spillover effects. I provide 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 later on: (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.

I then 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 3.4 presents a simulation setting that studies the performance of spillover effects estimators under simple and two-stage random assignment.

Finally, Chapter IV discusses several issues related to the empirical implementation of the findings from the previous chapters: the inclusion of covariates, identification of spillover effects under imperfect compliance and optimal design of experiments. The discussion in this chapter highlights some important challenges related to these issues and points to some directions for future work in the analysis of spillover effects.

CHAPTER II

A Framework for Analyzing Spillover Effects in Randomized Experiments

2.1 Introduction

Situations in which individuals' actions and behaviors indirectly affect other agents, a phenomenon usually known as spillovers, externalities or interference, abound in economics and other social sciences. Spillovers can materialize through different channels such as general equilibrium and displacement effects, agglomeration and economies of scale, diffusion of information and knowledge through social interactions, and contagion effects, among others. A large empirical and theoretical literature has analyzed the existence of spillover effects in different areas such as education (Epple and Romano, 2011; Sacerdote, 2011; Carrell and Hoekstra, 2010), labor (Topa, 2001; Duflo and Saez, 2003; Crépon, Duflo, Gurgand, Rathelot, and Zamora, 2013), firm agglomeration, cluster policies and R&D (Bloom, Schankerman, and Van Reenen, 2013; Greenstone, Hornbeck, and Moretti, 2010; Figal Garone, Maffioli, de Negri, Rodriguez, and Vazquez-Bare, 2015), health (Miguel and Kremer, 2004; Baird, Hicks, Kremer, and Miguel, 2016), development (Banerjee, Chandrasekhar, Duflo, and Jackson, 2013), crime and law enforcement (Di Tella and Schargrodsky, 2004; Yang, 2008; Rincke and Traxler, 2011).

Despite the longstanding and widespread interest in this phenomenon across different disciplines, identification of spillover effects of programs and policies has proven a challenging problem. In economics, most existing identification results rely on strong parametric assumptions such as the linear-in-means model (Moffit, 2001; Manski, 1993; Bramoullé, Djebbari, and Fortin, 2009). Moreover, the problem is usually analyzed from a linear regression perspective without a clear notion of causality,

so the parameters considered are best linear predictors whose causal interpretation remains unclear.

On the other hand, researchers have conducted and analyzed experiments in which treatment is assigned in such a way that some units are subject to spillovers while some units are not. A popular design in this setting is one in which groups of individuals (such as classrooms, households or villages) are randomly divided in categories, and then the treatment is assigned in each category with a different intensity. Different versions of this experimental design have been implemented, for example, by Duflo and Saez (2003); 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) and Giné and Mansuri (forthcoming), among others, and more formally analyzed, mostly in terms of estimation and inference, by Hudgens and Halloran (2008); Hirano and Hahn (2010) and Baird, Bohren, McIntosh, and Özler (forthcoming).

A common feature in the existing literature analyzing spillover effects is that the parameters of interest are defined by the estimation strategy or the experimental design. In both cases, however, the lack of a general framework leaves the causal interpretation of the parameters unclear. The goal of this chapter is to provide a causal nonparametric framework based on potential outcomes to define spillover effects and establish conditions for identification when a treatment is randomly assigned.

In Section 2.2, I describe a setup in which individuals are grouped and their potential outcomes can depend on the treatment assignments of all the units within the group. In this setting, different treatment assignment configurations give a rich set of policy parameters of interest measuring how individual outcomes vary with own and peers' treatment assignments. I analyze an assumption, *exchangeability*, according to which average potential outcomes are invariant to permutations of peers' treatment assignments, and show how this restriction can drastically reduce the dimensionality of the problem, making it more tractable in practice. Section 2.3 offers a comparison between this setup and some of the ones used in previous studies.

Section 2.4 analyzes identification of direct and average spillover effects. I show that when the treatment is randomly assigned, all the parameters of interest can be identified using variation on own and peers' treatment assignments, whenever the treatment assignment mechanism gives non-zero probability to each possible treatment configuration. Based on these results, I characterize the estimands from

some of the most common empirical approaches when analyzing spillover effects in Section 2.5. Section 2.6 discusses some extensions to the setup in this chapter, and Section 2.7 concludes.

2.2 Setup, notation and parameters of interest

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 Bogotá, Colombia. The program aimed at increasing student retention and reducing drop-out and child labor. Eligible registrants were randomly assigned between the control status and treatment. 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 included in the study. This survey data contains attendance data and was conducted in the household.

As shown in Table 2.1, 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 2.2. 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 give an incomplete assessment of 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 chapter is to analyze conditions under

Table 2.1: Distribution of household size

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

Table 2.2: Treated per household

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

which spillover effects can be precisely defined and identified.

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 dissertation, 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 Chapter IV. 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:

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

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 dissertation 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 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

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

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 II.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 II.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, 2013b;

Leung, 2017). The main difference between this assumption and similar restrictions used in the literature is that Assumption II.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 II.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 II.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 II.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 II.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 here, 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 chapter to conserve space, leaving different alternatives to this assumption for Section 2.6.1. I will define two sets of parameters of interest. First, the *average direct effect* of the treatment given s treated neighbors is defined as:

$$\tau_s = \mu(1, s) - \mu(0, s) \tag{2.1}$$

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:

$$\theta_s(d) = \mu(d, s) - \mu(d, 0) \tag{2.2}$$

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 $\theta_{s+1}(d) - \theta_s(d)$. I provide conditions to achieve identification of these treatment effects when the treatment is randomly assigned in Section 2.4.

2.3 Comparison with existing literature

For a direct comparison between the literature and my setup, I will focus on the case in which the covariate of interest is a binary treatment D_{ig}. I will follow the notation introduced above.

2.3.1 Econometric models, LIM, response functions

The linear-in-means (LIM) model is arguably the most widely used tool to analyze peer effects in most areas of economics. In its standard version, a LIM model is given by the equation:

$$Y_{ig} = \alpha + \beta D_{ig} + \gamma \bar{D}_g^{(i)} + \delta \bar{Y}_g^{(i)} + \varepsilon_{ig} \tag{2.3}$$

where $\bar{D}_g^{(i)}$ is the leave-one-out sample average of D (in this case, the proportion of ones) excluding D_{ig}, and similarly for $\bar{Y}_g^{(i)}$. In this equation, β is the direct effect of the treatment, γ is the *exogenous effect* and δ is the *endogenous effect*. A "large group" version of this equation replaces $\bar{D}_g^{(i)}$ and $\bar{Y}_g^{(i)}$ with their within-group population averages $\mathbb{E}_g[D]$ and $\mathbb{E}_g[Y]$ (see e.g. Manski, 1993). A LIM model is often interpreted as the Nash equilibrium of a game in which players maximize a quadratic utility function (Blume, Brock, Durlauf, and Jayaraman, 2015; Kline and Tamer, forthcoming):

$$\max_{Y_{ig}} U\left(Y_{ig}, \bar{Y}_g^{(i)}\right) = \xi_{ig} Y_{ig} - \frac{(1-\delta) Y_{ig}^2}{2} - \frac{\delta}{2}(Y_{ig} - \bar{Y}_g^{(i)})^2$$

In this equation, the first two terms represent a private component of utility, with marginal private benefit equal to ξ_{ig} and a convex cost, and the last term captures a "social pressure" component (Blume, Brock, Durlauf, and Jayaraman, 2015). The presence of this last term implies that an individual gets higher utility by choosing an action Y_{ig} that is close the the average action in her group. The first-order condition of this maximization problem yields Equation 2.3 after setting $\xi_{ig} = \alpha + \beta D_{ig} + \gamma \bar{D}_g^{(i)} + \varepsilon_{ig}$.

Manski (1993) pointed out two identification problems associated with model 2.3. First, the model includes endogenous variables, namely, the outcomes of other units, as regressors (the *reflection problem*). Second, the presence of a group-level fixed effect can generate a correlation between the error term and the regressors (the

problem of *correlated effects*). Several approaches have been put forward to ensure identification of γ and δ, such as exploiting the variation generated by partially-overlapping groups (Bramoullé, Djebbari, and Fortin, 2009; De Giorgi, Pellizzari, and Redaelli, 2010), using variation in group sizes (Lee, 2007; Davezies, D'Haultfoeuille, and Fougère, 2009) or combining the availability of an exogenous instrument with the panel-like structure of the data (Graham and Hahn, 2005). However, these methods only work in specific contexts and can be very demanding in terms of data requirements. A more straightforward approach taken by the literature is to give up separate identification of γ and δ, and use the fact that under appropriate restrictions on the model parameters, Equation 2.3 can be solved to obtain a reduced-form equation (corresponding to a Nash equilibrium):

$$Y_{ig} = \lambda + \mu D_{ig} + \theta \bar{D}_g^{(i)} + u_{ig} \tag{2.4}$$

where the coefficients (λ, μ, θ) are (nonlinear) functions of the structural parameters $(\alpha, \beta, \gamma, \delta)$. In this case, θ captures the composite exogenous and endogenous peer effect. Although Equation 2.4 does not allow separate identification of the exogenous and endogenous effects, Manski (2013a,b) has argued that the reduced form may actually be the object of interest from a policy perspective, since a policy intervention can affect exogenous variables but not outcomes directly.

While Equation 2.4 circumvents the endogeneity generated by the presence of $\bar{Y}_g^{(i)}$, its parameters remain unidentified when $(D_{ig}, \bar{D}_g^{(i)})$ are correlated with u_{ig}. Such correlation can arise, for example, when units in the same group are subject to common shocks. If these common shocks are correlated with the regressors, the reduced-form parameters are not identified. For instance, suppose D_{ig} indicates whether student i in classroom g has ever failed a grade, and $\bar{D}_g^{(i)}$ is the proportion of students excluding i that have failed a grade (repeaters). If classrooms with a higher proportion of repeaters are assigned better teachers, then teacher quality is a group-level shock that is correlated with the regressors, and it is impossible to disentangle the effect of having more repeaters from the effect of having better teachers.

Again, the literature has offered several alternatives to deal with this issue. A credible approach has been to rely on random assignment to eliminate the correlation between the regressors and the error term. There are two main ways in which

randomization is conducted in the peer effects literature. The first one is random assignment of group membership. For instance, Sacerdote (2001), Zimmerman (2003) and Stinebrickner and Stinebrickner (2006) exploit random (or random-like) assignment of college roommates, while Lyle (2007) and Carrell, Fullerton, and West (2009) study the case of random assignment into peer groups in West Point and the Air Force Academy, respectively. Graham (2008) takes advantage of the random assignment of students to small and large classrooms in Project STAR to identify peer effects through variance contrasts. However, random assignment of groups breaks apart when individuals refuse to interact with the peers they were assigned to (Carrell, Sacerdote, and West, 2013). The second method is direct random assignment of the treatment. Moffit (2001) argued in favor of *partial-population experiments*, in which the proportion of treated units is randomized in each group. Some examples exploiting random assignment of treatment in linear-in-means models are Lalive and Cattaneo (2009), Bobonis and Finan (2009) and Dieye, Djebbari, and Barrera-Osorio (2014).

Even when randomization is possible, identification of the parameters still relies strongly on the linearity imposed on Equations 2.3 and 2.4, and the question remains of whether (i) the linear model is an appropriate representation of the phenomenon under study and (ii) it is possible to achieve identification without imposing a linear structure. Attempts to relax the linearity assumption have been motivated by both theoretical and empirical considerations. On the one hand, the linear model is generally incorrect when outcomes are binary or discrete. This observation sparked a large literature on binary-choice models with social interactions (see e.g. Brock and Durlauf, 2001). Although this literature removes linearity, it usually does so by replacing it by alternative (and possibly equally strong) parametric or distributional assumptions. On the other hand, the linear model has been criticized on empirical grounds for the unrealistic restrictions that it imposes on the structure of peer effects (Hoxby and Weingarth, 2005; Carrell, Fullerton, and West, 2009; Sacerdote, 2011, 2014).

On the opposite end of the spectrum, Manski (2013b) and Lazzati (2015) study nonparametric partial identification of the response function, that is, the reduced-form relationship between outcomes and treatment assignments, in presence of social interactions. These papers characterize identification regions for the distribution of the potential outcomes under different restrictions on the structural model, the

response functions and the structure of social interactions.

In this dissertation, I focus on identification and estimation of (reduced-form) response functions under random assignment of the treatment. By considering the "many small groups" case with an exogenous treatment, I can achieve point identification in a setting that has wide empirical applicability. On the other hand, randomization allows me to bypass the endogeneity issues that plague observational studies and focus on the structure of the response function, with emphasis on the restrictions that justify the different models that the literature has used in practice and their causal interpretation. Specifically, after defining a potential outcome $Y_{ig}(d, \mathbf{d}_g)$, I show how a general, nonparametric potential-outcome model can become a (reduced-form) LIM model under three conditions, namely (i) exchangeability, (ii) equal spillover effects under treatment and control status, and (iii) linear spillover effects. I also analyze the parameters that can be recovered by a misspecified LIM model.

While Manski (2013a), Manski (2013b) and Angrist (2014) have questioned the relevance of such models from a causal perspective, it is interesting to ask what type of structural model can justify the response functions that I study in this dissertation. To simplify the discussion, consider a setting with groups of size 2, that is, each unit has one neighbor. Suppose the structural potential outcomes y_i are generated by the following system:

$$y_1 = f(d_1, d_2, y_2, \varepsilon_1)$$
$$y_2 = f(d_2, d_1, y_1, \varepsilon_2)$$

This implies that

$$y_1 = f(d_1, d_2, f(d_2, d_1, y_1, \varepsilon_2), \varepsilon_1)$$
$$y_2 = f(d_2, d_1, f(d_1, d_2, y_2, \varepsilon_1), \varepsilon_2)$$

Depending on the form of $f(\cdot)$, the above system may have one, zero or multiple equilibria. Suppose that $f(\cdot)$ is such that the system has a unique equilibrium. Then, the reduced form potential outcome for unit i in group g is given by:

$$Y_{ig}(d_i, d_j) = \varphi(d_i, d_j, \varepsilon_i, \varepsilon_j)$$

Integrating over the joint distribution of the error terms,

$$\mathbb{E}[Y_{ig}(d_i, d_j)] = \phi(d_i, d_j)$$

for some function $\varphi(\cdot)$, where the shape of this function will depend on the structural function $f(\cdot)$. Manski (2013b) provides an extensive discussion on the relationship between these functions in a general nonparametric setting. Now, because d_i and d_j are binary, we have that, without loss of generality,

$$\begin{aligned}\varphi(d_i, d_j) &= \varphi^{00}(1-d_i)(1-d_j) + \varphi^{10}d_i(1-d_j) + \varphi^{01}(1-d_i)d_j + \varphi^{11}d_i d_j \\ &= \varphi^{00} + (\varphi^{10} - \varphi^{00})d_i + (\varphi^{01} - \varphi^{00})(1-d_i)d_j + (\varphi^{11} - \varphi^{10})d_i d_j\end{aligned}$$

where $\varphi^{dt} = \varphi(d, t)$, so the average (reduced-form) potential outcome is a function of the treatment indicators and interactions. This shows that in this case,

$$\tau_s = \varphi^{1,s} - \varphi^{0,s}, \quad \theta_1(d) = \varphi^{d1} - \varphi^{d0}.$$

Importantly, the nonparametric nature of the reduced form does not rely on parametric assumptions on the structural equation. Since the treatment indicators are binary, the reduced form can always be written in a fully saturated form, which does not require any assumptions on the structural equations, besides the restrictions that guarantee a unique equilibrium.

As an illustration, consider the following structural function with constant coefficients (so the treatment and spillover effects are homogeneous):

$$y_i = f(d_i, y_j, \varepsilon_i) = \alpha_i + \beta d_i + \theta y_j + \delta d_i y_j$$

where $\alpha_i = \alpha(\varepsilon_i)$. Then,

$$\begin{cases} y_i = \alpha_i + \theta(\alpha_j + \theta y_i) & \text{if } d_i = 0,\ d_j = 0 \\ y_i = \alpha_i + \beta + (\theta + \delta)(\alpha_j + \theta y_i) & \text{if } d_i = 1,\ d_j = 0 \\ y_i = \alpha_i + \theta(\alpha_j + \beta + (\theta + \delta)y_i) & \text{if } d_i = 0,\ d_j = 1 \\ y_i = \alpha_i + \beta + (\theta + \delta)(\alpha_j + \beta + (\theta + \delta)y_i) & \text{if } d_i = 1,\ d_j = 1 \end{cases}$$

which implies the reduced form:

$$
\begin{aligned}
Y_{ig}(0,0) &= \frac{\alpha_i + \theta\alpha_j}{1-\theta^2} \\
Y_{ig}(1,0) &= \frac{\alpha_i + (\theta+\delta)\alpha_j + \beta}{1-\theta(\theta+\delta)} \\
Y_{ig}(0,1) &= \frac{\alpha_i + \theta\alpha_j + \beta\theta}{1-\theta(\theta+\delta)} \\
Y_{ig}(1,1) &= \frac{\alpha_i + (\theta+\delta)\alpha_j + \beta(1+\theta+\delta)}{1-(\theta+\delta)^2}
\end{aligned}
$$

as long as θ^2, $(\theta+\delta)\theta$ and $(\theta+\delta)^2$ are different from 1. Then, the reduced-form causal effects are nonlinear functions of the structural parameters. For example,

$$
\tau_0 = \frac{\alpha\theta\delta + \beta(1+\theta)}{(1+\theta)[1-\theta(\theta+\delta)]}
$$

where $\alpha = \mathbb{E}[\alpha_i]$.

2.3.2 Analysis of experiments with interference

By "analysis of experiments with interference" I refer to a body of research, developed primarily in statistics and epidemiology, that studies causal inference in experiments when the potential outcome of a unit can depend on the treatment assignments of other units (i.e., a failure of the SUTVA). Rubin (1990) and later Halloran and Struchiner (1991, 1995) extended the potential-outcomes causal framework by letting each unit's potential outcome to depend on the vector of treatment assignments in a sample. In this setting, the literature has mostly focused on four estimands. Given a sample with units $i = 1, \ldots, M$, the *direct effect* is the difference in potential outcomes for unit i under treatment and control, given a vector of assignments for the remaining $M-1$ units. The *indirect effect* is the difference in the outcomes of unit i, given own treatment assignment, for two possible assignments of the remaining $M-1$ units. The *total effect* is the sum of the direct and indirect effects. Finally, the *overall effect* is the difference between the potential outcomes of unit i under two alternative vector of treatment assignments. The corresponding average effects are defined by simply averaging these estimands over the whole sample, as described below. As it is common in analysis of experiments (see e.g. Imbens and

Rubin, 2015), potential outcomes are seen as fixed, and all the randomness comes through the treatment assignment mechanism.

The main complication that arises in a setting with interference is that the number of potential outcomes for each unit can become very large, taking up to 2^M values. Sobel (2006) introduced the assumption of *partial interference*, under which units in a sample are partitioned into groups, and interference can only occur between units in the same group. This assumption greatly simplifies the problem and seems to have been adopted by a vast majority of studies in this literature.

Given the focus on finite populations with non-random outcomes, identification issues are largely absent from this literature, and interest is placed instead on finding unbiased estimators for the estimands of interest, estimating their variance and performing inference. Hudgens and Halloran (2008) discuss unbiased estimation and variance calculations under partial interference under two-stage randomization designs. They focus on finite-population versions of the estimands described above, in which individual potential outcomes are averaged over the distribution of the vector of neighbors' assignments. More precisely, given a probability distribution of treatment assignment parameterized by ψ, the individual average potential outcome under assignment d is defined as:

$$\bar{Y}_{ig}(d,\psi) = \sum_{\mathbf{d}_g \in \mathcal{D}_g} Y_{ig}(d,\mathbf{d}_g)\mathbb{P}_\psi[\mathbf{D}_{(i)g} = \mathbf{d}_g | D_{ig} = d]$$

Based on this magnitude, the group average potential outcome and the (finite) population average potential outcome are given by:

$$\bar{Y}_g(d,\psi) = \frac{1}{n_g+1}\sum_{i=1}^{n_g+1} \bar{Y}_{ig}(d,\psi), \qquad \bar{Y}(d,\psi) = \frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{n_g+1}\sum_{i=1}^{n_g+1}\bar{Y}_{ig}(d,\psi)\right)$$

Then, the population average direct effect is given by $\bar{Y}(1,\psi) - \bar{Y}(0,\psi)$; given two parameterizations of the treatment assignment distribution, ψ and ϕ, the population average indirect effect is $\bar{Y}(0,\psi) - \bar{Y}(0,\phi)$; the population average total effect is $\bar{Y}(1,\psi) - \bar{Y}(0,\phi)$ (which is the sum of the direct and indirect effects). Hudgens and Halloran (2008) propose unbiased estimators for the above estimands, and provide variance estimators under exchangeability. Tchetgen Tchetgen and VanderWeele (2012) and Rigdon and Hudgens (2015) propose finite sample confidence

intervals, while Liu and Hudgens (2014) study confidence intervals in a large-sample randomization inference context. Basse and Feller (forthcoming) adapt the variance estimators to the case of varying group sizes and link the randomization inference framework with the regression framework.

The statistics literature focuses almost exclusively on inference for finite populations with fixed potential outcomes, in which all the randomness comes through the assignment mechanism. The super-population approach, under which potential outcomes are drawn from a certain (infinite) population distribution, has two advantages over the finite population one. First, the parameters are defined with respect to a population of interest instead of a particular realization of a sample. In this sense, a super-population parameter has more external validity and more policy relevance than a magnitude which is only defined for a particular sample. Incidentally, some common population estimands of interest are not well defined when potential outcomes are fixed; for instance, the average treatment effect on the treated (ATT) is a random variable in a finite sample. Second, the population approach allows me to clearly distinguish the assumptions needed for identification from the ones needed for estimation and inference. This distinction gives a clearer conceptual picture of what conditions are required to identify the parameters of interest and what conditions are simplifications that permit estimating them and conducting hypothesis testing.

Among the few studies that analyze interference under a super-population approach, Philipson (2000) suggests conducting two-stage randomization experiments to analyze how average outcomes change in response to different shares of treated units in the population ("external effects"), while Baird, Bohren, McIntosh, and Özler (forthcoming) perform power calculations for the above estimands (in its super-population version) under what they call saturation designs. Both studies consider a setting in which units are partitioned into groups and potential outcomes are exchangeable.

Under both the finite-population and the super-population approaches, two-stage randomization has played a crucial role in the interference literature. In fact, some estimands of interest like the average indirect effect $\bar{Y}(0,\psi) - \bar{Y}(0,\phi)$ in Hudgens and Halloran (2008) are generally undefined under other designs like, for instance, simple randomization. Furthermore, the few studies discussing identification of population parameters tend to attribute identifying power to the two-stage design. For example, the abstract from Philipson (2000) states that "two-stage randomization

schemes, which randomize allocation of treatments across communities and randomizes the treatments themselves within communities, are useful for identifying private and external treatment effects.", while Baird, Bohren, McIntosh, and Özler (forthcoming) claim: "[w]e show that [randomized saturation] designs identify a set of novel estimands" (page 2). However, I show below that (i) spillover effects can be defined without reference to a specific assignment mechanism or experimental design, and (ii) these spillover effects can be identified as long as the assignment mechanism puts non-zero probability on each possible assignment. Specifically, I argue that a simple randomized (Bernoulli) experiment is enough to identify all the parameters of interest.

While most of this literature assumes partial interference, a recent body of research seeks to adapt the potential outcomes framework to more general structures of social interactions through arbitrary networks. When allowing for general interference, potential outcomes can depend on the treatment assignment of the whole population. In fact, the partial interference assumption can be seen precisely as a way to simplify this problem; in a networks setting, partial interference corresponds to the case with many independent networks (or alternatively, a large network with a block diagonal adjacency matrix). Since estimation and inference can become infeasible when the structure of interactions is completely arbitrary, the main challenge faced by this literature is therefore to provide reasonable restrictions on the type of interference that can occur between units.

Some studies replace the partial interference by similar but more general restrictions on the spillovers structure. For instance, Leung (2017) proposes restrictions on the structure of dependency graphs, which describe the correlation structure in a network, to perform asymptotic inference in a super-population framework. On the other hand, Eckles, Karrer, and Ugander (2017) study the bias of the global ATE estimator under different modeling assumptions and experimental designs. Choi (2017) considers identification under the assumption that the treatment effect is monotone. Basse, Feller, and Toulis (2018) propose conditional randomization tests under general forms of interference.

2.4 Identification of 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 II.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 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 II.2 (Identification under exchangeability) *Under Assumptions II.1 and II.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 II.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 II.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 Chapter III).

Lemma II.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 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:

$$\mathbb{E}[Y_{ig}|D_{ig}, S_{ig}] = \alpha + \tau_0 D_{ig} + \sum_{s=1}^{n_g} \theta_s(0)\mathbb{1}(S_{ig} = s)(1 - D_{ig}) + \sum_{s=1}^{n_g} \theta_s(1)\mathbb{1}(S_{ig} = s)D_{ig} \tag{2.5}$$

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 (2.5) 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 the next chapter.

2.5 Analysis of common empirical approaches

Although there is a rich set of causal effects of interests that can be identified under the conditions stated in Lemma II.2, the vast majority of empirical work either ignores the presence of spillovers or accounts for them in very specific ways. The setup described in Section 2.2 can be used to understand the performance of some common empirical approaches. In this section, I will analyze identification of spillover effects under three approaches that are very commonly employed: a regression of the outcome on a treatment indicator, which basically ignores the presence of spillovers by comparing treated and control units, a linear-in-means model that controls for the proportion of treated siblings, and a regression that exploits variation in treatment probabilities across groups in two-stage designs.

2.5.1 Difference in means

The population difference in means between treated and controls, defined as

$$\beta_{\mathsf{D}} = \mathbb{E}[Y_{ig}|D_{ig} = 1] - \mathbb{E}[Y_{ig}|D_{ig} = 0], \tag{2.6}$$

is a standard estimand when analyzing the effects of randomly assigned treatments. This parameter can be recovered through a regression of the form

$$Y_{ig} = \alpha_{\mathsf{D}} + \beta_{\mathsf{D}} D_{ig} + u_{ig} \tag{2.7}$$

where by construction (given that the treatment is binary), $\alpha_{\mathsf{D}} = \mathbb{E}[Y_{ig}|D_{ig} = 0]$ and $\mathbb{E}[u_{ig}|D_{ig}] = 0$. In the absence of spillovers, the exogeneity implied by random assignment ensures that β_{D} equals the average treatment effect. When analyzing

spillovers, a natural question is what causal parameter can be recovered through β_{D}. The following result addresses this issue:

Lemma II.3 (Difference in means) *Under the conditions for Lemma II.2, β_D 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. Under simple random assignment, 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.

In the absence of spillovers, $Y_{ig}(d, \mathbf{d}_g) = Y_{ig}(d)$, which implies $\theta_s(d) = 0$ for all s and d, and $\beta_{\mathsf{D}} = \tau_0 = \mathbb{E}[Y_{ig}(1) - Y_{ig}(0)] = ATE$. 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. In this case,

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

so a large treatment effect with an equally large spillover effect of the same sign would be observationally equivalent to a zero treatment effect.

Intuitively, the presence of spillovers generates a correlation between units' treatment assignments that is not controlled for in the specification in (2.7). To illustrate

this intuition, suppose that each household has only two siblings, $n_g = 1$, so that under simple random assignment,

$$\beta_{\mathsf{D}} = \tau_0 + [\theta_1(1) - \theta_1(0)]p$$

where $\mathbb{P}[D_{ig} = 1] = p$. It is easy to see that $\mathbb{C}\text{ov}(D_{ig}, D_{ig}D_{jg}) = p^2(1-p)$ and $\mathbb{V}[D_{ig}] = p(1-p)$, and thus:

$$\beta_{\mathsf{D}} = \tau_0 + [\theta_1(1) - \theta_1(0)] \cdot \frac{\mathbb{C}\text{ov}(D_{ig}, D_{ig}D_{jg})}{\mathbb{V}[D_{ig}]}$$

The second term in the above expression equals the well-known omitted variable bias in OLS, which is the effect of the omitted variable (in this case, the interaction between own and sibling's treatment indicators) multiplied by the coefficient from the regression of the omitted variable on the included regressor. This expression suggests that the difference in means is biased for the direct effect of the treatment when spillover effects are non-zero. Interestingly, this bias does not disappear when the treatment is randomly assigned, since $\mathbb{C}\text{ov}(D_{ig}, D_{ig}D_{jg})$ can never be zero unless $p = 0$ or $p = 1$.

2.5.2 Linear-in-means models

Equation (2.7) 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:

$$Y_{ig} = \alpha_\ell + \beta_\ell D_{ig} + \gamma_\ell \bar{D}_g^{(i)} + \eta_{ig}, \qquad \bar{D}_g^{(i)} = \frac{1}{n_g} \sum_{j \neq i} D_{jg} \tag{2.8}$$

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.

Unlike Equations (2.5) or (2.7), which exploit the binary nature of the treatment indicator, Equation (2.8) imposes parametric assumptions on the type of spillover effects that are allowed for. In particular, Equation (2.8) requires three conditions to

be correctly specified: (i) exchangeability, (ii) equal spillover effects under treatment and control, $\theta_s(0) = \theta_s(1) := \theta_s$ for all s and (iii) linearity of spillover effects in s, that is, $\theta_s = \kappa s$ for some constant κ. Under these three conditions, Equation (2.5) 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). In this case, both coefficients β_ℓ and γ_ℓ have clear causal interpretations.

On the other hand, the following result characterizes the coefficient γ_ℓ under general conditions.

Lemma II.4 (LIM regression) *Under the conditions for Lemma II.2 and simple random assignment, the coefficient γ_ℓ from Equation (2.8) can be written as:*

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

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. The reason for this counterintuitive weighting scheme is that when conditions (i)-(iii) do not hold, the LIM model is misspecified.

A straightforward way to make Equation (2.8) 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:

$$Y_{ig} = \alpha_\ell + \beta_\ell D_{ig} + \gamma_\ell^0 \bar{D}_g^{(i)}(1 - D_{ig}) + \gamma_\ell^1 \bar{D}_g^{(i)} D_{ig} + \xi_{ig} \tag{2.9}$$

Then the following holds.

Lemma II.5 (LIM with interactions) *Under the conditions for Lemma II.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.

2.5.3 Variation in assignment probabilities

Researchers in different fields in the social sciences 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.

The two-stage randomization mechanism can be executed in the following way. In the first stage, groups $g = 1, \ldots, G$ are randomly assigned to one of K mutually exclusive categories, denoted by $T_g = 0, 1, \ldots, K-1$, with $\mathbb{P}[T_g = t] = q_t$. In the second stage, within each group, unit-level treatment is randomly assigned with a probability that is determined by the category to which the group was assigned, $\mathbb{P}[D_{ig} = 1 | T_g = t] = p_t$. Random assignment of the group and treatment indica-

tors ensure that the potential outcomes are statistically independent of treatment assignments:

Assumption II.3 (Independence) *For all* $(d, \mathbf{d}_g) \in \{0,1\} \times \mathcal{D}_g$ *and for all* i *and* g*,*

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

While in principle the way in which the treatment is assigned is irrelevant for identification as long as the conditions in Lemma II.2 hold, researchers employing two-stage randomization often exploit the variability the treatment probability to analyze spillover effects. As described above, a very common approach in this case is to set $K = 2$, $p_0 = 0$ and $p_1 = p \in (0,1)$, thus obtaining two categories, one of pure controls and one in which units get the treatment with probability p. In this case, a popular strategy is to run the regression:

$$Y_{ig} = \alpha + \beta D_{ig} + \gamma T_g + u_{ig} \tag{2.10}$$

The intuition this specification is that $\gamma = \mathbb{E}[Y_{ig}|D_{ig} = 0, T_g = 1] - \mathbb{E}[Y_{ig}|T_g = 0]$ should estimate a spillover effect by comparing the average outcomes of control units in treated groups with control units in pure control groups. However, it is clear from the previous setting that there are $2(n_g + 1)$ different average spillover effects, so the question that arises is what is the relationship between the γ coefficient in Equation (2.10) and the spillover effects defined in (2.2). The following result addresses this issue.

Lemma II.6 (Identification through variation in assignment probabilities) *Under Assumptions II.1 and II.3,*

$$\mathbb{E}[Y_{ig}|T_g = 0] = \mathbb{E}[Y_{ig}(0)]$$

and for $d = 0, 1$ *and* $t = 1, 2, \ldots, K-1$*,*

$$\mathbb{E}[Y_{ig}|D_{ig} = d, T_g = t] = \mathbb{E}[Y_{ig}(d)] + \sum_{s=1}^{n_g} \mathbb{P}[S_{ig} = s|T_g = t]\theta_s(d)$$

where $\mathbb{P}[S_{ig} = 0|T_g = 0] = 1$*.*

The case where $p_{K-1} = 1$ can be handled analogously, with the only difference that $\mathbb{E}[Y_{ig}|D_{ig} = d, T_g = K-1] = \mathbb{E}[Y_{ig}(d)] + \theta_{n_g}(d)$.

Following the idea from Equation 2.10, a natural estimand in this context is the average difference between untreated units in groups assigned to $T_g = 0$ and $T_g = t > 0$,

$$\Delta_t(d) = \mathbb{E}[Y_{ig}|D_{ig} = d, T_g = t] - \mathbb{E}[Y_{ig}|T_g = 0]$$

which corresponds to the coefficients from a regression of the outcome on indicators of the form $\mathbb{1}(T_g = t)$. By simple algebra we can see that when comparing untreated units in treated and control groups, under simple random assignment within groups,

$$\begin{aligned}\Delta_t(0) &= \sum_{s=1}^{n_g} \theta_s(0)\mathbb{P}[S_{ig} = s|T_g = t] \\ &= \sum_{s=1}^{n_g} \theta_s(0)\binom{n_g}{s} p_t^s(1-p_t)^{n_g-s}\end{aligned}$$

Thus, each $\Delta_t(0)$ captures a weighted average of average spillover effects of all the orders, where the weights are given by the conditional probability of observing each possible number of treated neighbors. The lower the probability of treatment in category T_g, the higher the weight given to low order spillover effects, and vice versa.

The main limitation of this approach is that by pooling parameters that are potentially very different, it may miss important heterogeneity in spillover effects. For example, consider the case of herd immunity in a vaccination program. When only a few individuals in a group are vaccinated, it is unlikely that this will prevent an untreated individual from getting sick, since the probability of contagion will be high. However, as the number of treated neighbor increases, the probability of contagion goes down and untreated units will be less likely get sick. In this setting, we would expect the spillover effects of low orders to be close to zero and the spillover effects of high orders to be large. Depending on how the two-stage randomization is conducted, these effects can go undetected. For instance, in the simple case with two categories, $T_g = 0, 1$ with $p_0 = 0$, $p_1 = 0.5$, most of the weight will be placed on the spillover effect of treating half the neighbors, which is possibly small, and the effect of treating every neighbor, which is expected to be high, would receive a weight close to zero. A program like the one analyzed by Crépon, Duflo, Gurgand, Rathelot, and Zamora (2013) can be another example of this phenomenon. Consider a job training

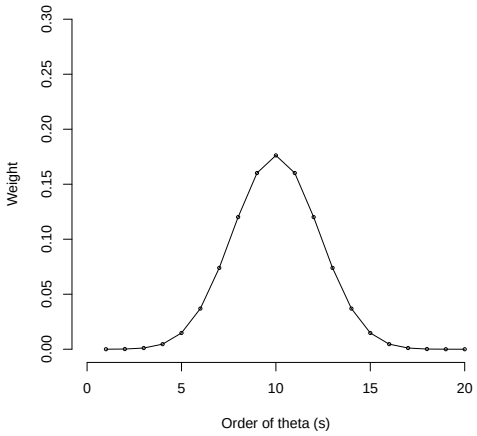
program aimed at helping unemployed workers to find jobs. If a few individuals get the treatment in a city or village, that is unlikely to affect the individuals who did not receive the program. But when a large number of people receive training, assuming the program has a positive effect, treated individuals will become better at finding jobs, increasing the competition and displacing the unemployed workers who did not receive training. Again, these effects are likely to be observed only when the number of treated neighbors is large enough.

One way to overcome this problem is to increase the number of categories K. With more variation in the assignment probabilities, the difference in means across categories $\Delta_t(d)$ will be better approximations to the average spillover effects corresponding to different proportions of treated neighbors. This idea is illustrated in Figures 2.1 to 2.3. Each figure depicts, for a given group size, the weights given by each $\Delta_t(d)$ to the different spillover effects $\theta_s(d)$ using four different two-stage randomization designs. The upper left subfigure correspond to the case of two categories, one with probability $p_0 = 0$ and one with probability $p_1 = 0.5$. In this case, $\Delta_1(d) = \mathbb{E}[Y_{ig}|D_{ig} = d, T_g = 1] - \mathbb{E}[Y_{ig}|T_g = 0]$ gives the highest weight to the spillover effect corresponding to half (or nearly half) neighbors treated, and almost zero weights to the spillover effects on the tails (no neighbors treated, all neighbors treated). The upper right subfigure uses five categories, with probabilities 0, 0.25, 0.50, 0.75 and 1. In this case, $\Delta_1(d)$ gives highest weight to the $\theta_s(d)$ corresponding to a fourth of the neighbors treated, $\Delta_2(d)$ gives highest weights to the $\theta_s(d)$ corresponding to half the neighbors treated and $\Delta_3(d)$ to three quarters of the neighbors treated (the weights given by $\Delta_4(d)$ to the spillover effects are not shown because by construction it gives weight equal to one to the effect of all neighbors treated). The subfigures in the bottom panel depict the case with six and ten categories. These figures illustrate how increasing the number of categories can give a better approximation to the different average spillover effects.

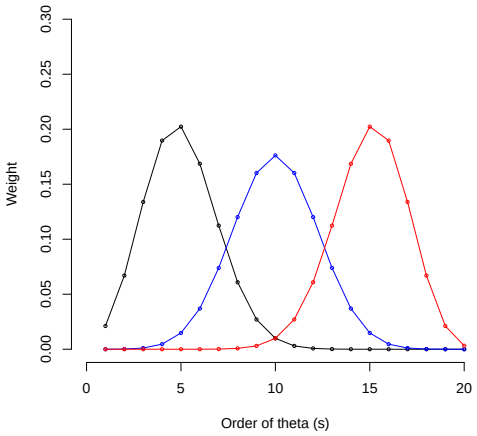
In fact, an extreme version of two-stage randomization would be one in which $K = n_g + 2$ and the number of treated units, instead of the probability, is assigned to each group. For example, in the program studied by Barrera-Osorio, Bertrand, Linden, and Perez-Calle (2011), one would take all the households with two registered siblings and randomly assign them to get zero, one or two treated children, and assign households with three registered children to get zero, one, two or three treated children. This mechanism is analyzed by Baird, Bohren, McIntosh, and Özler (forth-

coming), and has the advantage that each difference in means gives all the weight to one of the spillover effects. I analyze this assignment mechanism in the next chapter.

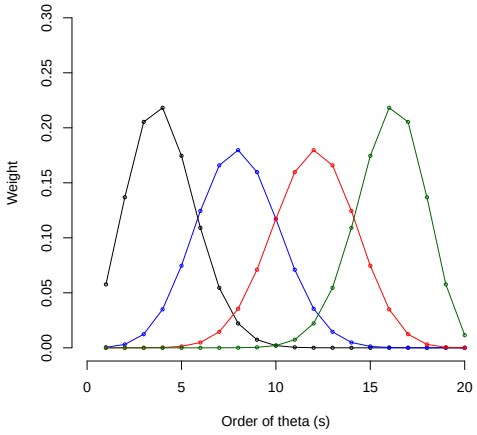
Figure 2.1: Weights assigned by $\Delta_t(d)$ to each $\theta_s(0)$ with $n = 20$.



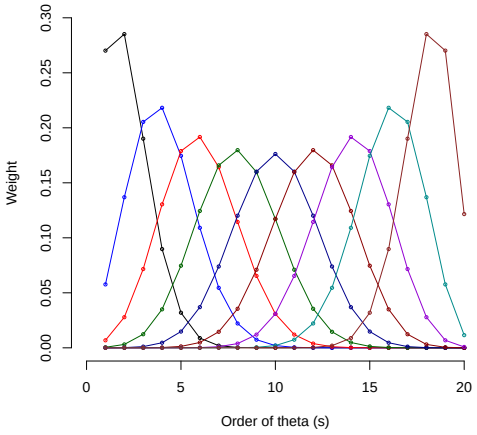
(a) $T_g \in \{0, 1\}$, $p \in \{0, 0.5\}$



(b) $T_g \in \{0, 1, 2, 3, 4\}$, $p \in \{0, 0.25, 0.50, .75, 1\}$



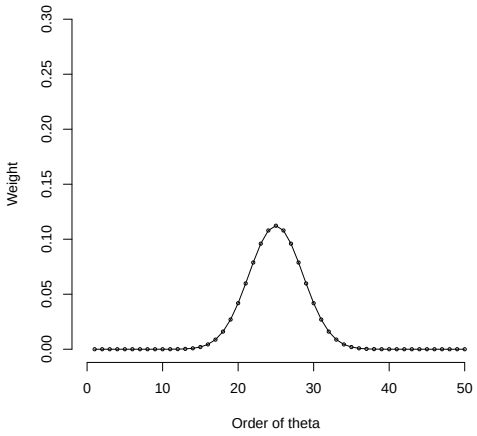
(c) $T_g \in \{0, 1, \ldots, 5\}$, $p \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$



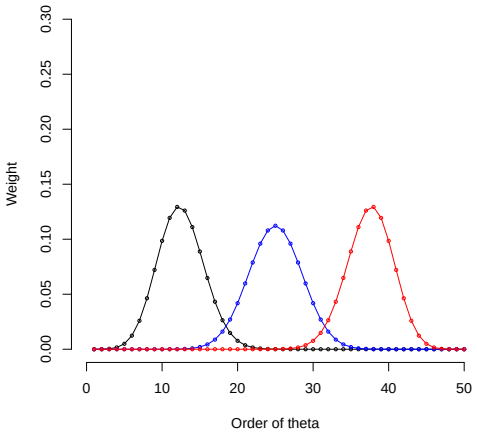
(d) $T_g \in \{0, 1, \ldots, 10\}$, $p \in \{0, 0.1, 0.2, \ldots, 1\}$

Notes: each curve plots the weights given by each difference in means $\Delta_t(d) = \mathbb{E}[Y_{ig}|D_{ig} = d, T_g = t] - \mathbb{E}[Y_{ig}|T_g = 0]$ to each $\mathbb{E}[\theta^k_{ig}(d)]$. In Panel (a) there are only two categories, therefore only one difference in means. In Panel (b) there are five categories, one in which there are no treated units and one in which everybody is treated (so $\Delta_4(d)$ gives weight equal to one to the spillover effect of higher order and zero to the remaining ones). Panels (c) and (d) are interpreted analogously.

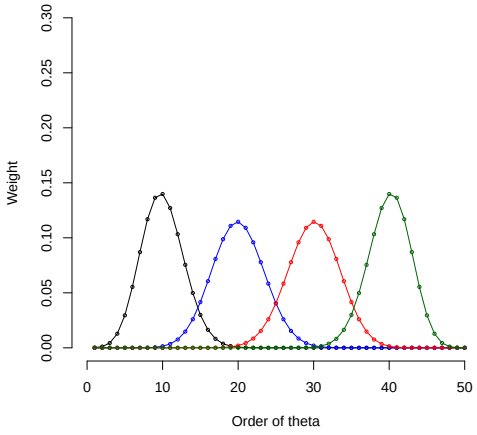
Figure 2.2: Weights assigned by $\Delta_t(d)$ to each $\theta_s(0)$ with $n = 50$.



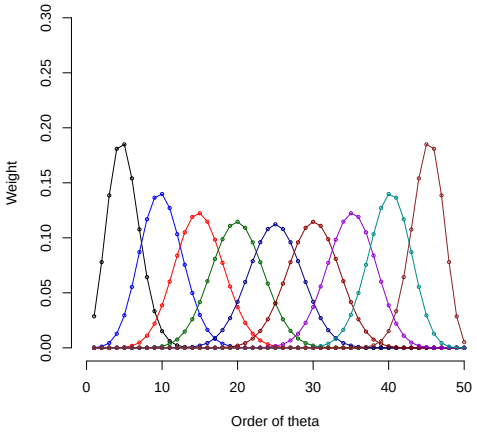
(a) $T_g \in \{0, 1\}$, $p \in \{0, 0.5\}$



(b) $T_g \in \{0, 1, 2, 3, 4\}$, $p \in \{0, 0.25, 0.50, .75, 1\}$



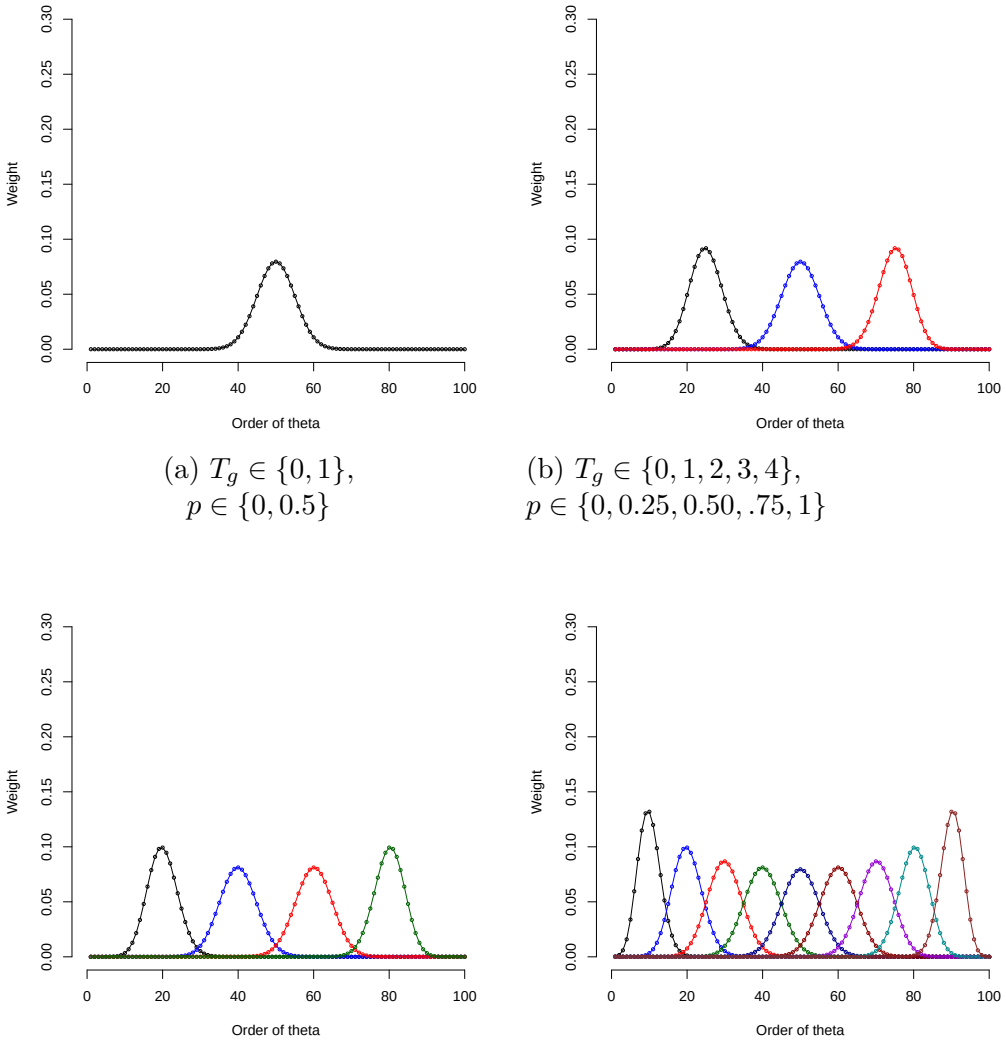
(c) $T_g \in \{0, 1, \ldots, 5\}$, $p \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$



(d) $T_g \in \{0, 1, \ldots, 10\}$, $p \in \{0, 0.1, 0.2, \ldots, 1\}$

Notes: each curve plots the weights given by each difference in means $\Delta_t(d) = \mathbb{E}[Y_{ig}|D_{ig} = d, T_g = t] - \mathbb{E}[Y_{ig}|T_g = 0]$ to each $\mathbb{E}[\theta_{ig}^k(d)]$. In Panel (a) there are only two categories, therefore only one difference in means. In Panel (b) there are five categories, one in which there are no treated units and one in which everybody is treated (so $\Delta_4(d)$ gives weight equal to one to the spillover effect of higher order and zero to the remaining ones). Panels (c) and (d) are interpreted analogously.

Figure 2.3: Weights assigned by $\Delta_t(d)$ to each $\theta_s(0)$ with $n = 100$.



(a) $T_g \in \{0, 1\}$, $p \in \{0, 0.5\}$

(b) $T_g \in \{0, 1, 2, 3, 4\}$, $p \in \{0, 0.25, 0.50, .75, 1\}$

(c) $T_g \in \{0, 1, \ldots, 5\}$, $p \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$

(d) $T_g \in \{0, 1, \ldots, 10\}$, $p \in \{0, 0.1, 0.2, \ldots, 1\}$

Notes: each curve plots the weights given by each difference in means $\Delta_t(d) = \mathbb{E}[Y_{ig}|D_{ig} = d, T_g = t] - \mathbb{E}[Y_{ig}|T_g = 0]$ to each $\mathbb{E}[\theta^k_{ig}(d)]$. In Panel (a) there are only two categories, therefore only one difference in means. In Panel (b) there are five categories, one in which there are no treated units and one in which everybody is treated (so $\Delta_4(d)$ gives weight equal to one to the spillover effect of higher order and zero to the remaining ones). Panels (c) and (d) are interpreted analogously.

2.6 Further issues and extensions

2.6.1 Alternatives to the exchangeability assumption

2.6.1.1 Lack of exchangeability

Although the exchangeability assumption can make potential outcomes more tractable by reducing the dimensionality, it is not necessary for identification and inference. In this section I show how to extend the results above to the general case without exchangeability. The potential outcome is given by $Y_{ig}(d, \mathbf{d}_g)$ where $d = 0, 1$ and $\mathbf{d}_g \in \mathcal{D} \subseteq \{0,1\}^{n_g}$, with observed outcome $Y_{ig} = \sum_d \sum_{\mathbf{d}_g} Y_{ig}(d, \mathbf{d}_g) \mathbb{1}(D_{ig} = d, \mathbf{D}_{(i)g} = \mathbf{d}_g)$. The parameters of interest are the average potential outcomes $\mathbb{E}[Y_{ig}(d, \mathbf{d}_g)]$ for all d and $\mathbf{d}_g \in \mathcal{D}_g$, which allow us to construct all possible treatment effects,

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

for any possible pair of treatment assignments $(d, \mathbf{d}_g)$ and $(\tilde{d}, \tilde{\mathbf{d}}_g)$. In particular,

$$\begin{aligned} \tau_{\mathbf{d}} &= \mathbb{E}[Y_{ig}(1, \mathbf{d}_g)] - \mathbb{E}[Y_{ig}(0, \mathbf{d}_g)] \\ \theta_{\mathbf{d}}(d) &= \mathbb{E}[Y_{ig}(d, \mathbf{d}_g)] - \mathbb{E}[Y_{ig}(d, \mathbf{0}_g)] \end{aligned}$$

which are analogous to the direct and average spillover effects defined above.

Under the assumption of independence between potential outcomes and treatment assignments (Assumption II.2), we have the following result.

Lemma II.7 (Identification without exchangeability) *Under Assumption II.2, for any assignment* $(d, \mathbf{d}_g)$ *such that* $\mathbb{P}[D_{ig} = d, \mathbf{D}_{(i)g} = \mathbf{d}_g] > 0$,

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

This result is analogous to Lemma II.2, with the difference that the left-hand side conditions on the whole vector of neighbors' assignments, instead of just the total number of treated neighbors.

2.6.1.2 Reference groups and covariates

Exchangeability fails when each unit in a group is only affected by a strict subset of neighbors, since in this case some of the spillover effects are automatically zero,

while others may not. If unit i's outcome can be affected by unit j's assignment but not by unit k's, then it is clear that units j and k cannot be exchangeable for unit i, since switching treatment from unit j to k will change unit i's outcome. However, when one has some information on the network structure, this situation is easy to handle in the above framework by manually setting some coefficients to zero and assuming exchangeability within each unit's reference group. More precisely, for each unit i, let R_i be unit i's reference group, that is, the set of indices corresponding to units that are linked to (that is, can potentially affect) unit i (for example, the set of unit i's friends). Then, the number of treated neighbors can be redefined as:

$$S_{ig}^R = \sum_{j \in R_i} D_{jg}$$

and all the previous results hold replacing S_{ig} by S_{ig}^R. This condition gives a middle ground between the case where each neighbor has a specific identity and the case where all the members in the group are exchangeable. The same reasoning can be used to make exchangeability hold conditional on a set of observed covariates. Say, for example, that exchangeability holds only after conditioning a binary covariate such as gender (so that sisters are exchangeable and brothers are exchangeable), $X_{ig} = 0, 1$. Then simply let:

$$S_{ig}(x) = \sum_{j : X_{jg} = x} D_{jg}$$

be the observed number of treated siblings with characteristic $X_{jg} = x$ (such as gender), and the results above can be adapted to this case using $\mathbb{E}[Y_{ig} | D_{ig}, S_{ig}(0), S_{ig}(1)]$.

2.6.2 Misspecification

There are two types of misspecification that are worth studying in this context, and both may stem from the need to reduce the dimensionality of the model when the number of neighbors is large relative to the number of observations. The first type of misspecification occurs when one or more neighbor assignment indicators are omitted. For example, suppose that groups are classrooms are units are students. If classrooms are relatively large, it may be hard to include all indicators and interactions, but the researcher may have information for instance on friendship relationships. If each

student only interacts with some of her peers, this additional information can be used to reduce the number of covariates to include in the regressions. However, social interactions are hard to measure in practice, so it is likely that the researcher can omit some of the student's peers. The second type of misspecification occurs when one incorrectly imposes exchangeability. I discuss the effect of these two types of misspecification below.

2.6.2.1 Omitting neighbors' treatment assignments

Take a group of size $n+1$ and partition the vector $\mathbf{D}_{(i)g}$ into $\mathbf{D}^k_{(i)g}$ and $\mathbf{D}^{n-k}_{(i)g}$. Then,

$$\mathbb{E}[Y_{ig}|D_{ig}=d,\mathbf{D}^k_{(i)g}=\mathbf{d}^k_g]=\sum_{\mathbf{d}^{n-k}_g}\mathbb{E}[Y_{ig}(d,\mathbf{d}^k_g,\mathbf{d}^{n-k}_g)] \times\mathbb{P}[\mathbf{D}^{n-k}_{(i)g}=\mathbf{d}^{n-k}_g|D_{ig}=d,\mathbf{D}^k_{(i)g}=\mathbf{d}^k_g]$$

so omitting a subset of neighbors' treatment indicators amounts to averaging the potential outcomes over all possible treatment assignments of the omitted neighbors. In the exchangeable case, letting

$$S_{ig}=\sum_{j=1}^{k}D^j_{ig}+\sum_{j=k+1}^{n}D^j_{ig}=S^k_{ig}+S^{n-k}_{ig}$$

we have that

$$\mathbb{E}[Y_{ig}|D_{ig}=d,S^k_{ig}=s_k]=\sum_{s_{n-k}=0}^{n-k}\mu(d,s_k+s_{n-k})\mathbb{P}[S^{n-k}_{ig}=s_{n-k}|D_{ig}=d,S^k_{ig}=s_k]$$

Suppose for simplicity that the treatment assignments are independent within group. Then,

$$\mathbb{E}[Y_{ig}|D_{ig}=d,S^k_{ig}=s_k]-\mathbb{E}[Y_{ig}|D_{ig}=d,S^k_{ig}=0]= \sum_{s_{n-k}=0}^{n-k}(\mu(d,s_k+s_{n-k})-\mu(d,s_{n-k}))\mathbb{P}[S^{n-k}_{ig}=s_{n-k}].$$

Hence, in both cases omitting neighbors' treatment assignments information im-

plies averaging over their assignments.

2.6.2.2 Incorrectly assuming exchangeability

Incorrectly assuming exchangeability amounts to exploiting variation in (D_{ig}, S_{ig}) to recover average effects when the average potential outcomes depend on the whole vector $(D_{ig}, \mathbf{D}_{(i)g})$. In this case we have that:

$$\begin{aligned}\mathbb{E}[Y_{ig}|D_{ig}=d, S_{ig}=s] &= \sum_{\mathbf{d}_g \in \mathcal{D}_g} \mathbb{E}[Y_{ig}|D_{ig}=d, S_{ig}=s, \mathbf{D}_{(i)g}=\mathbf{d}_g] \\ &\qquad \times \mathbb{P}[\mathbf{D}_{(i)g}=\mathbf{d}_g|D_{ig}=d, S_{ig}=s] \\ &= \sum_{\mathbf{1}_g'\mathbf{d}_g=s} \mathbb{E}[Y_{ig}(d,\mathbf{d}_g)] \frac{\mathbb{P}[D_{ig}=d|\mathbf{D}_{(i)g}=\mathbf{d}_g]\mathbb{P}[\mathbf{D}_{(i)g}=\mathbf{d}_g]}{\mathbb{P}[D_{ig}=d|S_{ig}=s]\mathbb{P}[S_{ig}=s]}\end{aligned}$$

In particular, when individual treatment assignments are mutually independent within a group, this expression reduces to:

$$\mathbb{E}[Y_{ig}|D_{ig}=d, S_{ig}=s] = \binom{n_g}{s}^{-1} \sum_{\mathbf{1}_g'\mathbf{d}_g=s} \mathbb{E}[Y_{ig}(d,\mathbf{d}_g)]$$

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

2.6.3 Unequally-sized groups

Suppose now group size is a random variable M_g with finite support $\{2, 3, \ldots, \bar{M}\}$ (if the support contains zero or one, these realizations would need to be excluded anyway since they do not provide information on spillovers). This implies that the number of neighbors for each unit is random as well. Let this variable be denoted by $N_g = M_g - 1$ with maximum value $\bar{N}$. Notice that in this case, the vector of neighbors assignments is also a function of N_g, $\mathbf{D}_{(i)g}(N_g)$. Let $\mathbf{d}(n_g)$ be a realization of this vector for a given realization n_g of N_g, and let $\mathbf{0}(N_g)$ be a vector of zeros of length N_g. Given a realization $N_g = n_g$, the potential outcome is given by $Y_{ig}(d, \mathbf{d}(n_g))$.

The results in this chapter can be generalized to this setting by conditioning on

group size. For example, the exchangeability assumption becomes:

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

for any pair of vectors $\mathbf{d}(n_g))$ and $\tilde{\mathbf{d}}(n_g))$ with the same number of ones. This restriction implies that:

$$\mathbb{E}[Y_{ig}(d, \mathbf{d}(n_g))|N_g = n_g] = \mu(d, s(n_g), n_g)$$

for some function $\mu(\cdot, \cdot, \cdot)$. The mean independence assumption can also be modified to hold conditional on group size:

$$\mathbb{E}[Y_{ig}(d, \mathbf{d}(n_g))|D_{ig}, \mathbf{D}_{(i)g}(n_g), N_g = n_g] = \mathbb{E}[Y_{ig}(d, \mathbf{d}(n_g))|N_g = n_g]$$

and identification can be established by:

$$\mathbb{E}[Y_{ig}|D_{ig}, \mathbf{D}_{(i)g}(n_g), N_g = n_g] = \mu(d, s(n_g), n_g)$$

In this setting, the easiest approach is to simply run separate analyses for each group size. This approach is very flexible in the sense that it allows both the baseline potential outcomes and the average direct and spillover effects to vary with group size. Equation (2.5) becomes:

$$\mathbb{E}[Y_{ig}|D_{ig}, S_{ig}(N_g), N_g] = \alpha(N_g) + \tau_0(N_g)D_{ig} + \sum_{s=1}^{\bar{N}} \theta_s(D_{ig}, N_g)\mathbb{1}(S_{ig}(N_g) = s)\mathbb{1}(s \leq N_g)$$

where $\alpha(N_g) = \mu(0, 0, N_g)$. This is a fully saturated model that includes interactions between own treatment indicator, indicators by number of treated siblings and groups sizes.

In practice, there may be cases in which group size has a rich support with only a few groups at each value n_g, so separate analyses may not be feasible. One possibility to overcome this difficulty is to strengthen the conditional mean independence assumption, and require that the effect of group size be additive:

$$\tau_0(n_g) = \tau_0, \quad \theta_s(d, n_g) = \theta_s(d)$$

Intuitively, this means that group size is allowed to affect the average baseline potential outcomes (when no neighbors are treated), but not the average direct and spillover effects. This condition rules out cases in which group size is related to potential gains from the treatment (for example, if larger groups are more likely to benefit from a program), but allows for the baseline potential outcomes under no treatment to differ across group sizes. For instance, it could be the case that large households have lower income and hence would have lower average outcomes compared to smaller households even when no unit gets the treatment, as long as these differences do not interact with treatment effects. This requirement is analogous to the identification assumption in difference-in-differences and panel data models, in which individuals unobserved characteristics are allowed to differ at the baseline, as long as these differences do not affect the potential outcomes' time trends. Under this assumption, group size only affects the baseline average outcome, but not the average direct and spillover effects. In this case,

$$\mathbb{E}[Y_{ig}|D_{ig}, S_{ig}(N_g), N_g] = \alpha(N_g) + \tau_0 D_{ig} + \sum_{s=1}^{\bar{N}} \theta_s(D_{ig}) \mathbb{1}(S_{ig}(N_g) = s) \mathbb{1}(s \leq N_g)$$

which can be interpreted as a regression that includes group size fixed effects, where

$$\alpha(N_g) = \sum_{n_g \in \mathcal{N}_g} \alpha(n_g) \mathbb{1}(N_g = n_g)$$

where $\mathcal{N}_g$ is the support of N_g. Thus, all the identification results and estimation strategies in this chapter are valid after controlling for group-size fixed effects.

2.7 Conclusion

This chapter 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 including the difference in means, linear-in-means models and two-stage designs.

An important takeaway of this setup is that, while the number of parameters of interest in the presence of spillovers can be large, identification can be established un-

der mild restrictions on the treatment assignment mechanism. But the large number of parameters of interest, which make the model very rich in terms of identification, also make estimation and inference harder. In the following chapter, I analyze estimation and inference for average direct and spillover effects with a special focus on the effect of group size and the number of parameters.

The setup in this chapter points to several directions for future research. Two generalizations worth studying are allowing for more complex interaction structures in a population, for example, when a population cannot be partitioned into groups, and allowing for the treatment to affect the group structure. In both cases, the growing literature on networks seems like a natural path to analyze these issues. Other avenues for future work are problems related to measurement error or treatment missclasification and misspecification of the network structure.

CHAPTER III

Estimation and Inference for Spillover Effects

3.1 Introduction

The results in Chapter II showed that when a treatment is randomly assigned, all the average potential outcomes can be identified under mild restrictions on the treatment assignment mechanism. The identification result in Lemma II.2 also suggests a simple yet nonparametric strategy to recover all the parameters of interest by running a regression of the outcome of interest on a set of indicator variables and interactions. This chapter studies estimation and inference for average spillover effects. I start by illustrating how to empirically implement the ideas in the previous chapter in Section 3.2 using two empirical applications: a conditional cash transfer pilot program in Colombia and an experiment in Ghana aiming at reducing electoral irregularities.

These illustrations highlight the advantages and disadvantages of estimating the whole set of average direct and spillover effects nonparametrically, which are formally analyzed in Section 3.3. To understand the role of group size and the total number of parameters in estimation and inference, I consider a setting in which both the number of groups and the group size are allowed to grow. I establish consistency and asymptotic normality of nonparametric estimators of the average potential outcomes under the conditions that the number of parameters to be estimated is small relative to the effective sample size, defined formally below. I also show validity of inference based on the wild bootstrap, which is found in simulations to perform better in terms of coverage compared to the normal approximation.

Finally, I show how my results can be used to guide the design of experiments when estimating spillover effects. Specifically, I show that the performance of the estimators in terms of inference depends crucially on the treatment assignment mech-

anism. This finding can be used as a principled criterion to rank among experimental designs. I compare two specific treatment assignment mechanisms, a Bernoulli trial and a two-stage design that fixes the number of treated units in each group, and show that the latter can yield improved inference when the goal is to estimate the whole set of average spillover effects.

3.2 Empirical applications

3.2.1 Conditional cash transfers in Colombia

Barrera-Osorio, Bertrand, Linden, and Perez-Calle (2011) 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 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.

Table 3.1: Distribution of household size

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

Table 3.2: Treated per household

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

Table 3.3: 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. $^{***}p<0.01,^{**}p<0.05,^{*}p<0.1$.

As shown in Table 3.1, 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 3.2.

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.

3.2.1.1 Nonparametric estimation

The results are obtained by estimating Equation (2.5). The estimates are shown in the right panel of Table 3.3. 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 the hypothesis that $\theta_1(0) = \theta_2(0)$ cannot be rejected 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.

3.2.1.2 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 3.3 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:

$$Y_{ig} = \alpha_{\mathsf{D}} + \beta_{\mathsf{D}} D_{ig} + u_{ig} \tag{3.1}$$

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. In Chapter II, I showed that:

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

From Table 3.3, 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.

3.2.1.3 Linear-in-means models

When trying to explicitly estimate spillover effects, a common strategy is to estimate a reduced-form linear-in-means model, which is given by:

$$Y_{ig} = \alpha_\ell + \beta_\ell D_{ig} + \gamma_\ell \bar{D}_g^{(i)} + \eta_{ig}, \qquad \bar{D}_g^{(i)} = \frac{1}{n_g} \sum_{j \neq i} D_{jg} \tag{3.2}$$

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 this specification are given in the first column of the middle panel in Table 3.3. The estimates suggest slightly negative and not significant direct and spillover effects, substantially different from results using Equation 2.5. To better understand this point, in Chapter II I showed that a LIM model requires

three conditions to be correctly specified: (i) exchangeability, (ii) equal spillover effects under treatment and control, $\theta_s(0) = \theta_s(1) := \theta_s$ for all s and (iii) linearity of spillover effects. However, the results from the nonparametric specification suggest that these conditions do not hold in this case: spillover effects have different signs for treated and control units, and the effects seem to be nonlinear. In general,

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

where $p = \mathbb{P}[D_{ig} = 1]$. Hence, γ_ℓ captures a linear combination of spillover effects using weights that are negative for all $s \leq \mathbb{E}[S_{ig}]$. 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 3.3. 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 the LIM model 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:

$$Y_{ig} = \alpha_\ell + \beta_\ell D_{ig} + \gamma_\ell^0 \bar{D}_g^{(i)}(1 - D_{ig}) + \gamma_\ell^1 \bar{D}_g^{(i)} D_{ig} + \xi_{ig} \tag{3.3}$$

The third column of the middle panel in Table 3.3 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. From Chapter II,

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

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 $\theta_s(1)$. Hence, these coefficients will not in general lie between the true spillover effects, which can be seen in Table 3.3 from the fact that -0.088 is not a weighted average of -0.026 and -0.057.

A possible concern with the above analysis is that the exchangeability assumption may not hold in this case, since the spillover effects may differ depending on

several factors such as siblings' age, gender or school they attend. While the results in Chapter II suggests that incorrectly imposing exchangeability still allows the researcher to recover weighted averages of the effects of each sibling, I explore this possibility by defining distance between siblings in terms of age. I order siblings in terms of the age difference, so that for each child, sibling one will be the sibling closest in age (older or younger) and sibling two will be the farthest in age (older or younger). The rationale for this ordering is that we could expect that siblings with similar ages go to similar grades and can affect each other more than siblings that are far away in terms of age. Other ordering criteria are possible and would identify different dimensions of effect heterogeneity.

The effects are estimated using the following fully-saturated regression:

$$\begin{aligned} Y_{ig} = \alpha &+ \tau D_{ig} + \gamma_1 D_{1g}(1 - D_{2ig})(1 - D_{ig}) + \gamma_2 (1 - D_{1g}) D_{2ig}(1 - D_{ig}) \\ &+ \gamma_{12} D_{1ig} D_{2ig}(1 - D_{ig}) \\ &+ \lambda_1 D_{1g}(1 - D_{2ig}) D_{ig} + \lambda_2 (1 - D_{1g}) D_{2ig} D_{ig} \\ &+ \lambda_{12} D_{1g} D_{2ig} D_{ig} \end{aligned} \tag{3.4}$$

where γ_k recovers the spillover effect of sibling k for an untreated kid, γ_{12} is the effect of having both siblings treated on an untreated kid, λ_k is the spillover effect of sibling k on a treated kid, and λ_{12} the effect of having both siblings treated on a treated kid.

The results are shown in Table 3.4. The estimates are qualitatively similar to the ones under exchangeability, but suggest a larger spillover effect for the closest sibling on an untreated kid. The above specification gives a straightforward way to test exchangeability, by evaluating whether $\gamma_1 = \gamma_2$ and $\lambda_1 = \lambda_2$. In this particular case, exchangeability is rejected at the 5 percent level, so the estimates in Table 3.3 should be interpreted as weighted averages of the individual effects in Table 3.4.

In sum, the empirical results in this section reveal how the saturated regression given by Equation (2.5) 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.

Table 3.4: Estimation results - non-exchangeable case

	Saturated	
	coef	s.e.
D_{ig}	0.129*	0.068
$D_{1ig}(1-D_{2ig})(1-D_{ig})$	0.151**	0.064
$(1-D_{1ig})D_{2ig}(1-D_{ig})$	0.110*	0.063
$D_{1ig}D_{2ig}(1-D_{ig})$	0.118*	0.061
$D_{1ig}(1-D_{2ig})D_{ig}$	0.008	0.026
$(1-D_{1ig})D_{2ig}D_{ig}$	-0.082*	0.042
$D_{1ig}D_{2ig}D_{ig}$	-0.060**	0.030
Constant	0.751***	0.059
Observations		363

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

3.2.2 Electoral irregularities in Ghana

To illustrate the results in Chapter II under two-stage designs, I will use the data from Ichino and Schündeln (2012). This experiment aims at assessing the effect of assigning observers to voter registration sites in Ghana. The hypothesis is that if observers reduce electoral misconduct in voter registration, the change in the number of registered voters in locations where observers are assigned should be smaller than the change in locations without observers. To test this hypothesis, the authors propose a two-stage experimental design:

1. In the first stage, 39 constituencies (administrative units) are assigned to the treatment group with probability 1/3. For this stage, constituencies were grouped in blocks of three in each of which one was assigned to treatment and two to control status.

2. In the second stage, electoral areas (ELAs) in treated constituencies are selected with probability 0.25 to be visited by registration observers.

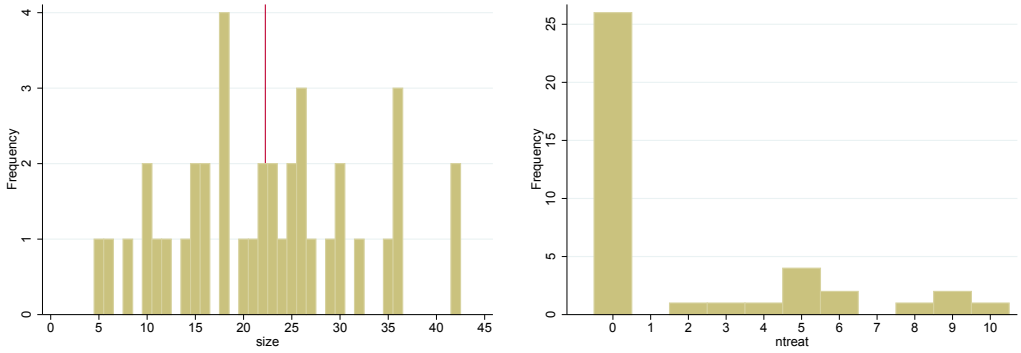
Registration observers randomly visited their assigned centers in treated ELAs during the registration period. Each ELA was visited at least twice during this period (twice for rural areas, three times or more in urban areas); each first visit lasted between one and two hours, while follow up visits lasted up to a whole day.

The sample consists of 868 ELAs in 39 constituencies, divided in 13 treated and 26 controls. In the 13 treated constituencies, 77 ELAs are treated. Average group

Table 3.5: Distribution of number of treated neighbors

	Frequency
0	592
1	2
2	7
3	19
4	34
5	59
6	33
7	8
8	35
9	47
10	32

Figure 3.1: Distribution of group size and number of treated units.



(a) Group size

(b) Number of treated units

size (number of ELAs in each constituency) is 22.3, and the distribution is plotted in Figure 3.1. Table 3.5 reports the distribution of the number of treated neighbors. We see from the table that 592 out of the 868 ELAs have no treated neighbors; these correspond to the ELAs in the 26 control constituencies. On the other hand, ELAs in treated constituencies can have up to 10 treated neighbors.

Unlike in the Barrera-Osorio, Bertrand, Linden, and Perez-Calle (2011) data, in this case group size varies too much to perform separate analyses for each group size. For this reason, I will run regressions pooling all group sizes together and controlling for group-size fixed effects. The assumption that justifies this approach is that the effect of group size in average potential outcomes is additive, that is, that changing the size of a group generates a constant shift in the potential outcomes without affecting the spillover effects, as described in Chapter II. Furthermore, because the sample sizes are small, there are several assignments that are not observed in the sample, and therefore the estimators for the corresponding effects are not defined.

I will focus on estimating average spillover effects for untreated units assuming exchangeability. Specifically, I run the regression

$$Y_{ig} = \alpha_{n_g} + \sum_{s=1}^{\bar{n}} \theta_s \mathbb{1}(S_{ig} = s)\mathbb{1}(s \leq n_g) + \eta_{ig}$$

on the sample of untreated units, where $\bar{n}$ is the maximum group size and α_{n_g} are group-size fixed effects. I will compare these estimates to the ones obtained by the pooling approach:

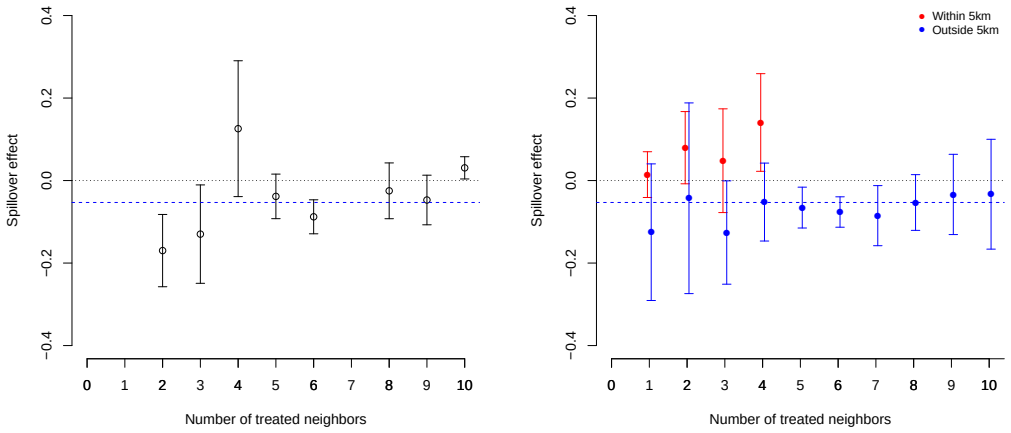
$$Y_{ig} = \alpha + \gamma T_g + \varepsilon_{it}$$

on the sample of untreated units, where $T_g = 1$ if constituency g is selected to received treatment.

The results are depicted in Figure 3.2. Figure 3.2(a) shows the estimated average spillover effect as a function of the number of treated neighbors. The dotted blue line corresponds to the estimate of the pooled parameter γ, and the dots are estimates of θ_s. Following Ichino and Schündeln (2012), a positive effect can be interpreted as a displacement effect, suggesting that electoral irregularities move from treated towards untreated ELAs. According to Figure 3.2(a), having more treated tends to increase voter registration, providing some evidence of displacement effects.

Since closer ELAs are more likely to affect neighboring areas, the exchangeability

Figure 3.2: Results under exchangeability.



(a) Full exchangeability

(b) Heterogeneity in distance

assumption may be too strong in this case. To study this possibility, in Figure 3.2(b) I estimate separate effects for treated ELAs within a 5km radius and outside the 5km radius. The figure clearly shows that the effect is positive for close treated neighbors and negative for treated neighbors far away. A possible interpretation is that there are two opposing effects, deterrence and displacement. When treating a neighbor that is close to an untreated unit, the displacement effect is stronger and the estimated effects are positive. On the other hand, for neighbors that are located outside the 5km radius, there is no displacement effect and hence the deterrence effect is stronger. However, it is important to keep in mind that small sample sizes the estimates are very noisy and only a few of them are statistically significant.

3.3 Estimation and inference

The previous chapter showed 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 $2(n_g + 1)$.[1] However, as shown by the empirical applications described above, the main challenge for es-

[1] As stated in Assumption III.1, I will assume in this section that groups are equally sized.

timation and inference arises when groups are large. A large number of units per group requires estimating a large number of means in each of the cells defined by the assignments (d, s). When groups have many units (as in households with many siblings or classrooms with a large number of students), the probability of observing some assignments can be close to zero and the number of observations in each cell can be too small to estimate the average potential outcomes. For example, suppose the treatment is assigned as an independent coin flip with probability $p = 1/2$. Under this assignment we would expect most groups to have about half its units treated, so when groups have, say, 10 units, 5 of them would be treated on average. The probability of observing groups with zero or all treated units, on the other hand, will be close to zero, and thus the average potential outcomes corresponding to these "tail assignments" will be very hard to estimate.

So far, the analysis has been done taking group size as fixed. When group size is fixed, small cells are a finite sample problem that disappears in a large enough sample. 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 III.1 (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 III.1 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 2.6.3 in Chapter II further discusses the case of unequally-sized groups. Finally, part (iv) requires that the total number of parameters does not grow faster than the effective sample size, that is, the expected sample size in the smallest cell.

Random assignment of the treatment implies that potential outcomes are statistically independent of the treatment assignments. I restate this assumption as follows to match the notation in this section.

Assumption III.2 (Independence) *For all* $\mathbf{a} \in \mathcal{A}_n$ *and for all* n*,* i*, and* g*,*

$$Y_{ig}(\mathbf{a}) \perp\!\!\!\perp \mathbf{A}_g.$$

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

$$\frac{\log |\mathcal{A}_n|}{G \underline{\pi}_n^2} \to 0 \tag{3.5}$$

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

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

Theorem III.1 says that, under condition (3.5), 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 (3.5) formalizes the meaning of "large sample" in this context, and states that the number of groups has to be large relative to the total number of parameters and the probability of the least likely assignment. 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, conditional on $N(\mathbf{a}) > 1$ (which occurs with probability approaching one under previously stated conditions), 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 III.2 (Consistency and asymptotic normality) *Under the conditions of Theorem III.1,*

$$\begin{aligned}\max_{\mathbf{a}\in\mathcal{A}_n} |\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})| &= O_{\mathbb{P}}\left(\sqrt{\frac{\log|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n}}\right), \\ \max_{\mathbf{a}\in\mathcal{A}_n} \left|\hat{\sigma}^2(\mathbf{a}) - \sigma^2(\mathbf{a})\right| &= O_{\mathbb{P}}\left(\sqrt{\frac{\log|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n}}\right),\end{aligned} \tag{3.6}$$

and

$$\max_{\mathbf{a}\in\mathcal{A}_n} \sup_{x\in\mathbb{R}} \left|\mathbb{P}\left[\frac{\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}} \le x\right] - \Phi(x)\right| = O\left(\frac{1}{\sqrt{G(n+1)\underline{\pi}_n}}\right) \tag{3.7}$$

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

Equation (3.6) shows that both the average potential outcome and standard error estimators converge in probability to their true values, uniformly over treatment assignments, at the rate $\sqrt{\log|\mathcal{A}_n|/(G(n+1)\underline{\pi}_n)}$. The denominator in this rate can be seen as the effective sample size in the smallest cell, whereas the numerator is a penalty for having an increasing number of parameters. Equation (3.7) bounds the difference between the distributions of the standardized potential outcomes estimators and the standard normal distribution, uniformly over the treatment assignments. Under condition (3.5), $G(n+1)\underline{\pi}_n \to 0$, 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 3.3.2.

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 (3.5) holds automatically as long as the number of groups goes to infinity. Consistency and asymptotic normality of the estimators can be achieved under the usual regularity conditions as $G \to \infty$, and the variance estimator can easily account for both heteroskedasticity and intragroup correlation using standard techniques. The particular case with homoskedasticity and a random-effects structure is analyzed by Baird, Bohren, McIntosh, and Özler (forthcoming).

3.3.1 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 III.3 (Wild bootstrap) *Under the conditions of Theorem III.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 3.4 using simulation data.

3.3.2 Implications for experimental design

Theorem III.2 shows that the quality of the standard normal as an approximation to the distribution of the t-statistics 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 III.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 III.1 (SR) *Under simple random assignment, condition (3.5) holds whenever:*

$$\frac{n+1}{\log G} \to 0 \tag{3.8}$$

Corollary III.2 (2SR-FM) *Under a 2SR-FM mechanism, condition (3.5) holds whenever:*

$$\frac{\log(n+1)}{\log G} \to 0 \tag{3.9}$$

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

Corollary III.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 III.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. I discuss these issues in Chapter IV.

3.4 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:

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

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 III.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:

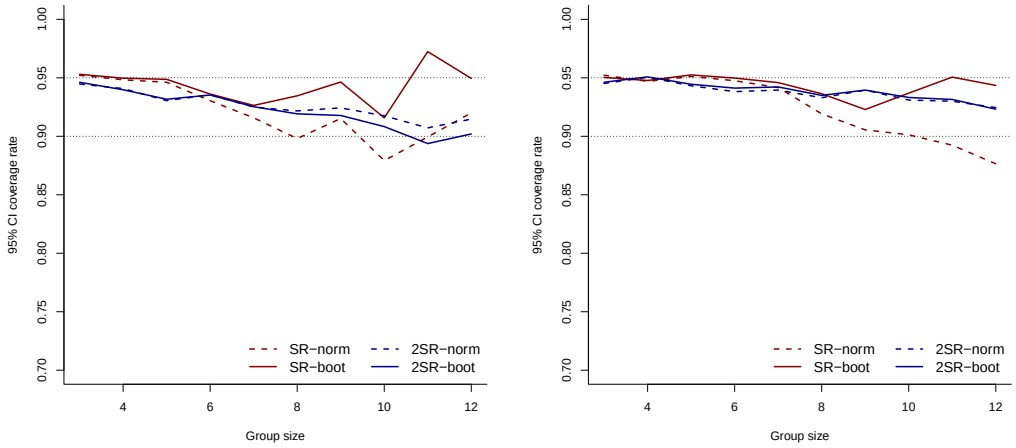
$$\hat{\theta}_n(0) = \frac{\sum_g \sum_i Y_{ig} \mathbb{1}_{ig}(0,n)}{N(0,n)} - \frac{\sum_g \sum_i Y_{ig} \mathbb{1}_{ig}(0,0)}{N(0,0)}$$

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 3.6 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 about 92% in both cases. 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

Figure 3.3: 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.

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 3.7 shows 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 (94.9 versus 92), although both the normal-based and the bootstrap-based confidence intervals perform similarly under 2SR. These results are also illustrated in Figure 3.3.

Table 3.6: 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	−0.0002	−0.0016	−0.0017	−0.0012
Variance	0.0027	0.0128	0.0416	0.0598
95% CI coverage	0.9522	0.9304	0.9152	0.9200
Prop. empty cells	0.0000	0.0083	0.5775	0.9833
Two-stage rand.				
$\log(n+1)/\log(G)$	0.1926	0.3141	0.3852	0.4357
Bias	−0.0003	0.0006	−0.0012	−0.0014
Variance	0.0024	0.0034	0.0046	0.0059
95% CI coverage	0.9447	0.9355	0.9243	0.9147
Prop. empty cells	0.0000	0.0000	0.0000	0.0000

Table 3.7: 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	−0.0002	−0.0015	−0.0016	−0.0017
Variance	0.0027	0.0121	0.0326	0.0312
95% CI coverage	0.9530	0.9361	0.9464	0.9495
Prop. empty cells	0.0000	0.0083	0.5775	0.9833
Two-stage rand.				
$\log(n+1)/\log(G)$	0.1926	0.3141	0.3852	0.4357
Bias	−0.0003	0.0006	−0.0012	−0.0014
Variance	0.0023	0.0034	0.0045	0.0056
95% CI coverage	0.9462	0.9353	0.9178	0.9020
Prop. empty cells	0.0000	0.0000	0.0000	0.0000

3.5 Conclusion

This chapter provides two empirical applications illustrating how to estimate the average direct and spillover effects of a randomly assigned treatment based on the identification results in Chapter II. Nonparametric estimation and inference are analyzed in a setting in which both the number of groups and group size are allowed to grow to approximate a setting in which groups can be "moderately" large. I establish consistency and asymptotic normality of the nonparametric estimators under the restriction that the number of parameters is small relative to the total sample size, and show consistency of the wild bootstrap.

Finally, I show how my results can be used to guide the design of experiments to estimate spillover effects. More precisely, my main result for inference reveals that the rates at which the distribution of the estimators approach the normal distribution depend on the treatment assignment mechanism. This finding provides a ranking among experimental designs based on their performance in terms of asymptotic inference. I illustrate this fact by comparing two commonly employed experimental designs, a Bernoulli trial and a two-stage design, and show that the latter can provide improved inference on the whole vector of spillover effects.

The framework in this chapter provides sufficient conditions for consistency and asymptotic normality of the estimators when the number of parameters grows slowly enough relative to the total sample size. In intuitive terms, this case corresponds to the analysis of groups that are "moderately" large, such as households or classrooms. Several empirically relevant cases may not fit well under this framework, for example, when groups are villages or large firms. Drawing from the growing literature on high dimensional models, future work should analyze estimation and inference when the number of parameters is proportional to, or larger than, the sample size.

CHAPTER IV

Further Issues in the Analysis of Spillover Effects

4.1 Introduction

This chapter addresses some issues related to the practical implementation of the results in the previous chapters. Section 4.2 discusses the inclusion of covariates. In practice, a researcher may want to include covariates in the analysis even when the treatment is randomly assigned to attenuate finite sample biases and to improve precision. Covariates can also be used to relax the random assignment assumption, replacing it by a conditional independence assumption, or to relax the exchangeability assumption.

Imperfect compliance, which occurs when units do not comply with their treatment assignments, is a pervasive problem in randomized controlled trials. Section 4.3 extends the setup in Chapter II to include this possibility. I define parameters of interest in this setting and analyze identification of these parameters by considering two cases: spillovers on treatment take-up and spillovers on outcomes.

Finally, Section 4.4 discusses experimental design for estimating spillover effects, considering different sets of parameters of interest, assignment mechanisms and optimality criteria.

4.2 Including covariates

Because estimation in this dissertation 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 (2.5) with a vector of covariates $\boldsymbol{\gamma}'\mathbf{x}_{ig}$ 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 heterogeneity across observable characteristics (for example, by separately estimating effects for males and females), although this strategy can decrease precision.

A second reason to include exogenous covariates is 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

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

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 results in previous chapters 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, at the risk of introducing misspecification bias.

4.3 Imperfect compliance

Imperfect compliance, which occurs when units do not comply with their treatment assignments, is a pervasive problem when conducting randomized controlled trials. Imbens and Angrist (1994) showed that, in the absence of spillovers, instrumenting treatment status with treatment assignment can, under some conditions, point identify the average treatment effect on the compliers, that is, the subpopulation who is pushed by the instrument to get the treatment.

Very few studies have considered identification and estimation of spillover effects under imperfect compliance. Duflo and Saez (2003) analyze social network effects in retirement plans enrollment decisions within a linear regression framework, while Eckles, Kizilcec, and Bakshy (2016) conduct an experiment encouraging peers to provide feedback in a large social network. Sobel (2006) studied the performance of the usual IV strategies in the presence of interference in a finite population setting with non-random potential outcomes. Within the same framework, Kang and Imbens (2016) extend the results in Hudgens and Halloran (2008) to account for imperfect compliance. Their paper considers estimation of spillover effects in two-stage designs when spillovers occur on outcomes but not on treatment status.

In this section, I generalize the framework in Chapter II to analyze spillovers under imperfect compliance. Under imperfect compliance, spillovers can occur at two different stages: treatment take-up and outcomes. The first stage occurs when the probability of an individual receiving the treatment depends on whether their neighbors are assigned to the treatment or not. For instance, consider the experiment in Duflo and Saez (2003), in which they encourage university employees to attend a retirement benefits fair through a letter and a cash reward. In this setting, it is possible that an employee does not receive the letter but decides to attend anyway because many of her coworkers were encouraged to attend. The second stage in which spillovers can materialize is the outcome stage. In the previous example, an individual who did not attend the benefits fair can still decide to enroll in a retirement plan by learning about the plan through social interactions with individuals who did attend the fair. I will separately discuss the cases in which one of these channels is assumed away.

4.3.1 Setup

Let $Z_{ig} \in \{0,1\}$ denote the (randomized) treatment assignment, with $(Z_{ig}, \mathbf{Z}_{(i)g})$ being the vector of treatment assignments in group g, taking values $(z, \mathbf{z}_g) \in \{0,1\} \times \mathcal{Z}_g$. Expanding the previous notation, $D_{ig}(z, \mathbf{z}_g)$ will denote potential treatment status of unit i in group g, and $D_{jig}(z, \mathbf{z}_g)$ will be the treatment status of unit i's j-th neighbor. Similarly, Z_{ig}^j is unit i's j-th neighbor's treatment assignment.

Assumption IV.1 (existence of instruments)

1. *(independence) For all i and g, $(Y_{ig}(d, \mathbf{d}_g, z, \mathbf{z}_g), D_{ig}(z, \mathbf{z}_g)) \perp\!\!\!\perp (Z_{ig}, \mathbf{Z}_{(i)g})$ for all d, $\mathbf{d}_g$, z, $\mathbf{z}_g$.*

2. *(exclusion restriction) $Y_{ig}(d, \mathbf{d}_g, z, \mathbf{z}_g) = Y_{ig}(d, \mathbf{d}_g)$.*

3. *(relevance) $\mathbb{E}[D_{ig} | Z_{ig} = z, \mathbf{Z}_{(i)g} = \mathbf{z}_g]$ is a non-trivial function of $(z, \mathbf{z}_g)$.*

This assumption extends the usual instrumental-variables assumption to the case of spillovers. Part 1 imposes statistical independence between treatment assignment and potential outcomes and treatment statuses. Parts 2 and 3 assert that the instrument does not have a direct effect on the potential outcome but is correlated with observed treatment assignment.

Both the potential outcome and the potential treatment status can in principle depend on the whole vector of treatment assignments, which, as discussed previously, has the advantage of being very flexible at the cost of a high dimensionality and strong data requirements. To simplify the setting, I will assume treatment status depends on own treatment assignment and number of assigned neighbors, and potential outcomes depend on own treatment status and number of treated neighbors. Let $W_{ig} = \sum_{j=1}^{n_g} Z_{jig}$ be the number of neighbors assigned to treatment. I will also consider Bernoulli assignment mechanism in which Z_{ig} is assigned as a coin flip.

4.3.2 Case 1: spillovers on treatment status

When spillovers only occur on the treatment take-up stage, individual outcomes are not affected by neighbors' treatment status, so that $Y_{ig}(d, \mathbf{d}_g) = Y_{ig}(d)$. Assuming exchangeability in the first stage, the observed outcome is therefore:

$$Y_{ig} = Y_{ig}(0) + \tau_{ig} D_{ig}(Z_{ig}, W_{ig}) \tag{4.1}$$

The difference in average observed outcomes for assignments (z, w) and (z', w') is:

$$\begin{aligned}\Delta(z, w, z', w') &= \mathbb{E}[Y_{ig}|Z_{ig} = z, W_{ig} = w] - \mathbb{E}[Y_{ig}|Z_{ig} = z', W_{ig} = w'] \\ &= \mathbb{E}[\tau_{ig}(D_{ig}(z, w) - D_{ig}(z', w'))] \\ &= \mathbb{E}[\tau_{ig}|D_{ig}(z, w) > D_{ig}(z', w')]\mathbb{P}[D_{ig}(z, w) > D_{ig}(z', w')] \\ &\quad - \mathbb{E}[\tau_{ig}|D_{ig}(z, w) < D_{ig}(z', w')]\mathbb{P}[D_{ig}(z, w) < D_{ig}(z', w')]\end{aligned}$$

A fact that stems from the above expression is that LATE-type estimands are identified only for the assignments such that either $\mathbb{P}[D_{ig}(z, w) > D_{ig}(z', w')] = 0$ or $\mathbb{P}[D_{ig}(z, w) < D_{ig}(z', w')] = 0$, a well-known fact (see e.g. Imbens and Angrist, 1994). The following result formalizes this idea in the presence of spillovers. While the result is analogous to Theorem 1 in Imbens and Angrist (1994), I state it as a proposition for future reference.

Proposition IV.1 (Identification of LATE) *Under Assumption IV.1, for any pair of assignments* (z, w) *and* (z', w') *such that* $\mathbb{P}[D_{ig}(z, w) < D_{ig}(z', w')] = 0$ *and* $\mathbb{P}[D_{ig}(z, w) > D_{ig}(z', w')] > 0$*,*

$$\mathbb{E}[\tau_{ig}|D_{ig}(z, w) > D_{ig}(z', w')] = \frac{\mathbb{E}[Y_{ig}|Z_{ig} = z, W_{ig} = w] - \mathbb{E}[Y_{ig}|Z_{ig} = z', W_{ig} = w']}{\mathbb{E}[D_{ig}|Z_{ig} = z, W_{ig} = w] - \mathbb{E}[D_{ig}|Z_{ig} = z', W_{ig} = w']}$$

For instance, one may be willing to assume that $\mathbb{P}[D_{ig}(1, w) < D_{ig}(0, w)] = 0$ for all w so that given a number of neighbors assigned to treatment, being assigned to treatment can never reduce the likelihood of receiving it, which is analogous to the monotonicity assumption in Imbens and Angrist (1994). In particular, the following parameters have clear interpretations. First,

$$\mathbb{E}[\tau_{ig}|D_{ig}(1, w) > D_{ig}(0, w)]$$

is the average treatment effect for the compliers with w neighbors assigned to treatment. Observe that this is a function of w, so there are in fact $n_g + 1$ of these parameters. On the other hand, for example,

$$\mathbb{E}[\tau_{ig}|D_{ig}(0, w + k) > D_{ig}(0, w)]$$

is the average treatment effect for units assigned to control but are pushed to get the treatment by k additional neighbors being assigned to treatment.

In fact, Equation 4.1 can be seen as a standard IV setting with a multivalued instrument and a binary treatment, where the values of the instrument are given by the different possible combinations of own and neighbors' assignments. For this reason, Proposition IV.1 is very similar to the result in Imbens and Angrist (1994). The main difference is that Proposition IV.1 does not require monotonicity to hold equally for all treatment assignments. More precisely, it could be the case that for some values the instrument never decreases the probability of receiving treatment while for other values the instrument never increases it, and the LATEs for those assignments are identified (although for different populations). The proposition also allows for some assignments to violate monotonicity, but the corresponding LATEs are not identified.

It is also interesting to analyze identification using the methods usually employed by applied researchers. Under imperfect compliance, the most common estimand of interest is the Wald estimand:

$$\tau_{\mathsf{W}} = \frac{\mathbb{E}[Y_{ig}|Z_{ig}=1] - \mathbb{E}[Y_{ig}|Z_{ig}=0]}{\mathbb{E}[Y_{ig}|D_{ig}=1] - \mathbb{E}[D_{ig}|Z_{ig}=0]}$$

Without spillovers, this ratio equals the average treatment effect for compliers (Imbens and Angrist, 1994). When spillovers can occur at the treatment take-up stage, and assuming $\mathbb{P}[D_{ig}(1,w) < D_{ig}(0,w)] = 0$ for all w, this estimand becomes:

$$\tau_{\mathsf{W}} = \sum_{w=0}^{n_g} \mathbb{E}[\tau_{ig}|D_{ig}(1,w) > D_{ig}(0,w)]\rho(w)$$

where

$$\rho(w) = \frac{\mathbb{P}[D_{ig}(1,w) > D_{ig}(0,w)]\mathbb{P}[W_{ig}=w]}{\sum_{w=0}^{n_g}\mathbb{P}[D_{ig}(1,w) > D_{ig}(0,w)]\mathbb{P}[W_{ig}=w]}$$

so τ_{W} recovers an average of LATEs weighted by the proportion of each type of complier in the population and the probability of observing each possible number of neighbors assigned to treatment. This result is similar to the ones in Angrist and Imbens (1995) for the case of variable treatment intensity.

The presence of spillovers on treatment status is actually straightforward to test in practice by exploring the variation in D_{ig} induced by $(Z_{ig}, \mathbf{Z}_{(i)g})$. By the previous

results above, assuming exchangeability for treatment take-up,

$$\mathbb{E}[D_{ig}|Z_{ig} = z, W_{ig}] = \mathbb{E}[D_{ig}(z)] + \sum_{w=1}^{n_g} \delta_w(z)\mathbb{1}(W_{ig} = w)$$

so the average spillover effects on treatment take-up, captured by the coefficients $\delta_w(z)$. Then, failure to reject the null that $\delta_w(z) = 0$ for $w = 1, \ldots, n_g$ would indicate the absence of average spillover effects in treatment status.

4.3.3 Case 2: spillovers on outcomes

When spillovers do not affect treatment take-up, we have that $D_{ig}(z, \mathbf{z}_g) = D_{ig}(z)$ which reduces the number of potential treatment statuses to two. I will maintain the assumption that potential outcomes satisfy exchangeability, but I will strengthen the condition to hold on the potential outcome function, instead of just the first moment, to simplify the discussion. I will refer to this condition as *strong exchangeability*.

Assumption IV.2 (Strong exchangeability) *For any assignment $(d, \mathbf{d}_g)$, the potential outcome is given by:*

$$Y_{ig}(d, \mathbf{d}_g) = Y_{ig}(d, s_g), \quad \mathbf{1}_g'\mathbf{d}_g = s_g.$$

This condition implies that the potential outcome can be written as:

$$Y_{ig}(d, s) = Y_{ig}(d, 0) + \theta_{ig}^s(d)$$

where

$$\theta_{ig}^s(d) = Y_{ig}(d, s) - Y_{ig}(d, 0).$$

As is customary in the literature, I will assume monotonicity, which means that being assigned to treatment cannot decrease the probability of receiving treatment.

Assumption IV.3 (Monotonicity) $\mathbb{P}[D_{ig}(1) < D_{ig}(0)] = 0$.

The number of treated neighbors for each unit will be given by $S_{ig}(\mathbf{z}_g)$. Exchangeability of potential outcomes is not enough to ensure that $S_{ig}(\mathbf{z}_g)$ is a function of

$\mathbf{1}_g'\mathbf{z}_g = w_g$. More precisely, without spillovers in treatment take-up we have that:

$$S_{ig}(\mathbf{z}_g) = \sum_{j=1}^{n_g} D_{jig}(z_j) = \sum_{j=1}^{n_g} D_{jig}(0) + \sum_{j=1}^{n_g} (D_{jig}(1) - D_{jig}(0))z_j \tag{4.2}$$

We see from the above equation that $S_{ig}(\mathbf{z}_g)$ can be written as a function of w_g only when $D_{jig}(1) - D_{jig}(0)$ does not depend on j, which means that the instrument has the same effect on all units. The assumption of relevance of the instrument prevents this difference to be zero for every unit, and hence if the instrument has the same effect on the treatment status of all units, either all units are compliers (so there is perfect compliance) or all units are defiers. Hence, $S_{ig}(\mathbf{z}_g)$ will depend on the whole vector of treatment assignments whenever compliance is imperfect. The intuition behind this fact is that neighbor identities cannot be irrelevant as different neighbors can have different compliance types. As a result, even when the potential outcome only depends on the treatment status vector through the number of treated neighbors, in general the number of neighbors assigned to treatment is not enough to provide identification, and the whole vector $\mathbf{Z}_{(i)g}$ has to be used instead.

Under monotonicity, $D_{ig}(0) = 1$ for always-takers and $D_{ig}(0) = 0$ for compliers and never-takers. On the other hand, $D_{ig}(1) - D_{ig}(0) = 1$ for compliers and $D_{ig}(1) - D_{ig}(0) = 0$ for always-takers and never-takers. Hence, $S_{ig}(\mathbf{z}_g)$ equals the number of unit i's neighbors that are always-takers plus the number of treated complier neighbors given a treatment assignment $\mathbf{z}_g$. Furthermore, monotonicity implies that $S_{ig}(\mathbf{z}_g)$ is non-decreasing in $\mathbf{1}_g'\mathbf{z}_g = w_g$, the number of neighbors assigned to treatment.

Under Assumptions IV.1 and IV.2, we have that

$$\mathbb{E}[Y_{ig}|Z_{ig} = z, \mathbf{Z}_{(i)g} = \mathbf{z}_g] = \mathbb{E}[Y_{ig}(0)] + \mathbb{E}[\tau_{ig} D_{ig}(z)] + \sum_{s=1}^{n_g} \mathbb{E}[\theta_{ig}^s (D_{ig}(z)) \mathbb{1}(S_{ig}(\mathbf{z}_g) = s)]$$

The difference between units assigned and not assigned to treatment given $\mathbf{z}_g$ is:

$$\begin{aligned}
\Delta_{10}(\mathbf{z}_g) :&= \mathbb{E}[Y_{ig}|Z_{ig}=1,\mathbf{Z}_{(i)g}=\mathbf{z}_g]-\mathbb{E}[Y_{ig}|Z_{ig}=0,\mathbf{Z}_{(i)g}=\mathbf{z}_g]\\
&=\mathbb{E}[\tau_{ig}(D_{ig}(1)-D_{ig}(0))]\\
&\quad+\sum_{s=1}^{n_g}\mathbb{E}[(\theta_{ig}^s(D_{ig}(1))-\theta_{ig}^s(D_{ig}(0)))\mathbb{1}(S_{ig}(\mathbf{z}_g)=s)]\\
&=\mathbb{E}[\tau_{ig}(D_{ig}(1)-D_{ig}(0))]\\
&\quad+\sum_{s=1}^{n_g}\mathbb{E}[(\theta_{ig}^s(1)-\theta_{ig}^s(0))(D_{ig}(1)-D_{ig}(0))\mathbb{1}(S_{ig}(\mathbf{z}_g)=s)]
\end{aligned}$$

and by monotonicity,

$$\begin{aligned}
\Delta_{10}(\mathbf{z}_g)=&\mathbb{E}[\tau_{ig}|D_{ig}(1)>D_{ig}(0)]\mathbb{P}[D_{ig}(1)>D_{ig}(0)]\\
&+\sum_{s=1}^{n_g}\mathbb{E}[\theta_{ig}^s(1)-\theta_{ig}^s(0)|D_{ig}(1)>D_{ig}(0),S_{ig}(\mathbf{z}_g)=s]\\
&\times\mathbb{P}[D_{ig}(1)>D_{ig}(0)]\mathbb{P}[S_{ig}(\mathbf{z}_g)=s]
\end{aligned}$$

The first term is the direct intention-to-treat (ITT) effect, which is the LATE multiplied by the proportion of compliers in the population. The second term is a weighted average of the average difference between spillover effects under treatment and control for different number of treated neighbors, conditional on being a complier and on having s treated neighbors given assignment $\mathbf{z}_g$. This expression shows that, even after controlling for neighbors' treatment assignment, the difference between treated and control units does not in general capture a treatment effect, but a combination of average direct and indirect effects. As a result, the usual instrumental-variables strategy that divides by compliance rates will recover:

$$\begin{aligned}
&\frac{\mathbb{E}[Y_{ig}|Z_{ig}=1,\mathbf{Z}_{(i)g}=\mathbf{z}_g]-\mathbb{E}[Y_{ig}|Z_{ig}=0,\mathbf{Z}_{(i)g}=\mathbf{z}_g]}{\mathbb{E}[D_{ig}|Z_{ig}=1]-\mathbb{E}[D_{ig}|Z_{ig}=0]}=\\
&\mathbb{E}[\tau_{ig}|D_{ig}(1)>D_{ig}(0)]+\sum_{s=1}^{n_g}\mathbb{E}[\theta_{ig}^s(1)-\theta_{ig}^s(0)|D_{ig}(1)>D_{ig}(0),S_{ig}(\mathbf{z}_g)=s]\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\times\mathbb{P}[S_{ig}(\mathbf{z}_g)=s]
\end{aligned}$$

Observe that this expression is similar to the formula for the difference in means (Equation (2.7)) in that it captures a direct average treatment effect plus a weighted average of average spillover effects. In both cases, the estimands fail to correctly control for neighbors' treatment status: in the difference in means, because neighbors' treatment status is ignored; in the Wald estimand, because under imperfect compliance treatment assignment does not completely determine treatment status.

There are some particular cases in which $\Delta_{10}(\mathbf{z}_g)$ can recover an ITT effect. For example, if for any number of treated neighbors, the spillover effect on the outcome is the same under treatment and control, $\theta^s_{ig}(1) = \theta^s_{ig}(0)$, then the second term disappears. The same would happen if all the terms inside the sum, $\mathbb{E}[\theta^s_{ig}(1) - \theta^s_{ig}(0)|D_{ig}(1) > D_{ig}(0), S_{ig}(\mathbf{z}_g) = s]$, are equal to zero. Another case in which the second term vanishes is when for some assignment $\mathbf{z}_g$, $\mathbb{P}[S_{ig}(\mathbf{z}_g) = 0] = 1$, that is, for some given neighbors' assignment $\mathbf{z}_g$, no neighbor receives the treatment. When any of these conditions hold, dividing $\Delta_{10}(\mathbf{z}_g)$ by $\mathbb{E}[D_{ig}|Z_{ig} = 1] - \mathbb{E}[D_{ig}|Z_{ig} = 0]$ yields the direct LATE, $\mathbb{E}[\tau_{ig}|D_{ig}(1) > D_{ig}(0)]$.

On the other hand, fixing own assignment and using variation in neighbors' assignment,

$$\begin{aligned}
\Delta_z(\mathbf{z}_g, \tilde{\mathbf{z}}_g) :&= \mathbb{E}[Y_{ig}|Z_{ig} = z, \mathbf{Z}_{(i)g} = \mathbf{z}_g] - \mathbb{E}[Y_{ig}|Z_{ig} = z, \mathbf{Z}_{(i)g} = \tilde{\mathbf{z}}_g] \\
&= \sum_{s=1}^{n_g} \mathbb{E}[\theta^s_{ig}(D_{ig}(z))(\mathbb{1}(S_{ig}(\mathbf{z}_g) = s) - \mathbb{1}(S_{ig}(\tilde{\mathbf{z}}_g) = s))] \\
&= \sum_{s=1}^{n_g} \mathbb{E}[\theta^s_{ig}(D_{ig}(z))|S_{ig}(\mathbf{z}_g) = s, S_{ig}(\tilde{\mathbf{z}}_g) \neq s]\mathbb{P}[S_{ig}(\mathbf{z}_g) = s, S_{ig}(\tilde{\mathbf{z}}_g) \neq s] \\
&\quad - \sum_{s=1}^{n_g} \mathbb{E}[\theta^s_{ig}(D_{ig}(z))|S_{ig}(\mathbf{z}_g) \neq s, S_{ig}(\tilde{\mathbf{z}}_g) = s]\mathbb{P}[S_{ig}(\mathbf{z}_g) \neq s, S_{ig}(\tilde{\mathbf{z}}_g) = s]
\end{aligned}$$

Unlike the no-spillovers case, monotonicity is not enough to ensure that either $\mathbb{P}[S_{ig}(\mathbf{z}_g) = s, S_{ig}(\tilde{\mathbf{z}}_g) \neq s] = 0$ or $\mathbb{P}[S_{ig}(\mathbf{z}_g) \neq s, S_{ig}(\tilde{\mathbf{z}}_g) = s] = 0$. In fact, these two probabilities will in general be non-zero for at least some values of s. Suppose for example that $\mathbf{1}'_g\mathbf{z}_g = w_g > \mathbf{1}'_g\tilde{\mathbf{z}}_g = \tilde{w}_g$ so that $\mathbf{z}_g$ assigns more units to treatment than $\tilde{\mathbf{z}}_g$. By monotonicity, $S_{ig}(\mathbf{z}_g) > S_{ig}(\tilde{\mathbf{z}}_g)$, and thus $\mathbb{P}[S_{ig}(\mathbf{z}_g) = s, S_{ig}(\tilde{\mathbf{z}}_g) \neq s]$ will be non-zero for larger values of s, whereas $\mathbb{P}[S_{ig}(\mathbf{z}_g) \neq s, S_{ig}(\tilde{\mathbf{z}}_g) = s]$ will be non-zero for low values of s. This fact highlights why average spillover effects are not identified without further restrictions.

Adding more structure to the problem can help give this estimand a clearer interpretation. In particular, suppose that noncompliance is one sided:

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

This assumption asserts that individuals assigned to control will not get the treatment, and thus the only type of noncompliance occurs because units can refuse to get treatment. In other words, Assumption IV.4 rules out the presence of always-takers, and therefore under this condition $S_{ig}(\mathbf{z}_g)$ equals the number of treated complier neighbors. In particular, $S_{ig}(\mathbf{0}_g) = 0$. Therefore,

$$\begin{aligned}\Delta_0(\mathbf{z}_g, \mathbf{0}_g) :&= \mathbb{E}[Y_{ig}|Z_{ig} = 0, \mathbf{Z}_{(i)g} = \mathbf{z}_g] - \mathbb{E}[Y_{ig}|Z_{ig} = 0, \mathbf{Z}_{(i)g} = \mathbf{0}_g] \\ &= \sum_{s=1}^{w_g} \mathbb{E}[\theta_{ig}^s(0)|S_{ig}(\mathbf{z}_g) = s]\mathbb{P}[S_{ig}(\mathbf{z}_g) = s]\end{aligned}$$

Each term $\mathbb{E}[\theta_{ig}^s(0)|S_{ig}(\mathbf{z}_g) = s]$ is an average spillover effect for the subgroup of units that have s complier neighbors assigned to treatment given assignment $\mathbf{z}_g$. These terms are combined into an average, weighted by the probability of having s complier neighbors assigned to treatment given $\mathbf{z}_g$. Note that the sum goes from 1 to w_g, since when a vector $\mathbf{z}_g$ assigns w_g units to treatment, the number of complier neighbors assigned to treatment cannot exceed w_g. To get a more precise characterization of the weighs, recall that in this case $S_{ig}(\mathbf{z}_g) = \sum_{j=1}^{n_g} D_{jig}(1)z_j = \sum_{z_j=1} D_{jig}(1) \sim \mathsf{Binomial}(w_g, p_c)$ where $p_c = \mathbb{P}[D_{ig}(1) = 1] = \mathbb{P}[\text{complier}]$. Therefore,

$$\mathbb{P}[S_{ig}(\mathbf{z}_g) = s] = \binom{w_g}{s} p_c^s (1 - p_c)^{w_g - s}$$

While the parameters $\mathbb{E}[\theta_{ig}^s(0)|S_{ig}(\mathbf{z}_g) = s]$ are not separately identified, the weights are, using the fact that under one-sided noncompliance, $p_c = \mathbb{E}[D_{ig}|Z_{ig} = 1]$.

4.4 Experimental design

Given the difficulties that can arise in practice when analyzing spillover effects, experimental design plays a crucial role for estimation and inference. Among the few studies on experimental design, Hirano and Hahn (2010) consider variance minimization in a two-stage design under a linear model like the one used in Duflo and

Saez (2003), whereas Baird, Bohren, McIntosh, and Özler (forthcoming) analyze statistical power for spillover effect estimation in a setting with small groups (that is, when n is fixed).

In a related but different literature, Toulis and Kao (2013), Kao (2017) and Eckles, Karrer, and Ugander (2017) study design of experiments in networks. The main challenge in this literature is to find treatment assignment mechanisms that ensure that the different potential outcomes of interest can be observed or estimated in a setting in which a unit's assignment can affect the assignments of all the other units in the population.

The results in Chapter III suggests several guidelines and criteria for the design of experiments to estimate spillover effects. In this section, I start by defining the parameters of interest and discussing the implications of this choice on the design of experiments. I then consider two criteria for experimental design. First, MSE optimality trades off bias and variance of the spillover effects estimators. When focusing on unbiased estimators, this problem reduces to minimizing an asymptotic variance, which is the case considered by Hirano and Hahn (2010) and Baird, Bohren, McIntosh, and Özler (forthcoming). When trying to estimate a possible large number of spillover effects, however, estimation may be complicated by the possibility of empty cells as discussed in Chapter III. I propose an approximation to the MSE of the nonparametric spillover estimators that incorporates this possibility.

Finally, I discuss optimality in terms of the rates of convergence of the estimators. As suggested by Theorem III.2, the treatment assignment mechanism plays a crucial role in the asymptotic properties of the estimators, and this fact can be exploited to improve inference through the design of the experiment.

4.4.1 Defining the parameters of interest

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

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

The case analyzed in this dissertation 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, $\beta(\mathcal{E}) = (\mathbb{E}[Y_{ig}(0,0)], \ldots, E[Y_{ig}(0,n)])$. 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 $\beta(\mathcal{E}) = \mathcal{E}$.

One way to accommodate different choices for $\beta(\cdot)$ within the framework described above is to let the set of effective treatment assignments $\mathcal{A}_n$ defined in Section 3.3 depend on $\beta(\cdot)$, that is, $\mathcal{A}_n = \mathcal{A}_n(\beta)$. In this way, $\mathcal{A}_n(\beta)$ will contain the set of assignments that are possible under a certain potential outcome model and that are chosen by the researcher as parameters of interest. For instance, when the goal is to estimate the effect of everybody being treated versus nobody being treated, the set of effective treatment assignments can be defined as $\mathcal{A}_n = \{(0, \mathbf{0}_g), (1, \mathbf{1}_g)\}$. The distribution over the possible treatment assignments contains only two values, π and $1-\pi$, where π is the probability of a whole group being treated, $\pi = \mathbb{P}[D_{ig} = 0, \mathbf{D}_{(i)g} = \mathbf{0}_g]$. In this case, $\underline{\pi}_n = \min\{\pi, 1-\pi\}$.

Therefore, any possible choice for $\beta(\cdot)$ can be analyzed by appropriately redefining the set $\mathcal{A}_n$ to contain only the relevant assignments.

4.4.2 Assignment mechanisms

In this section I will consider two families of assignment mechanisms that are the most commonly employed in practice. The first one, the Bernoulli trial, assigns the

treatment independently and with equal probability to each unit in the sample.

Definition IV.1 (Bernoulli trial) *a Bernoulli trial is an assignment mechanism in which:*

1. $\mathbb{P}[D_{ig} = 1] = p \in (0, 1)$ *for all* i, g.
2. *For all* $(d, \mathbf{d}_g) \in \{0, 1\} \times \mathcal{D}_g$, $\mathbb{P}[D_{ig} = d, \mathbf{D}_{(i)g} = \mathbf{d}_g] = \mathbb{P}[D_{ig} = d] \prod_{j=1}^{n_g} \mathbb{P}[D_{jig} = d_j]$.

The first part of Definition IV.1 says that each unit in group g has the same probability of treatment, and this probability is neither zero nor one so that both treatment and control status are possible. The second part imposes independence between the treatment assignments of units within a group. This assignment mechanism corresponds to the case in which the treatment is assigned as a simple coin toss for each unit. Other mechanisms like fixed-margins randomization, in which the number of treated units in each group is fixed, satisfy part 1 but not part 2, since, for example, the probability of an individual getting the treatment conditional on another unit having the treatment is lower than the unconditional probability of receiving the treatment.

The second family of assignment mechanisms that I consider is the family of two-stage designs. These designs are nowadays common practice when estimating spillover effects (see e.g. Duflo and Saez, 2003; Hudgens and Halloran, 2008; Crépon, Duflo, Gurgand, Rathelot, and Zamora, 2013; Baird, Bohren, McIntosh, and Özler, forthcoming). Generally, a two-stage design consists on dividing the groups $g = 1, \ldots, G$ into K mutually exclusive categories, denoted by a random variable $T_g = 0, 1, \ldots, K - 1$, and then assigning treatment at the individual level according to some distribution that depends on the realized value of T_g.

Definition IV.2 (2SR) *A two-stage design (2SR) is a treatment assignment mechanism in which:*

1. *Groups are assigned a value of* $T_g = 0, 1, \ldots, K-1$ *with probability* $\mathbb{P}[T_g = t] = q_t \in (0, 1)$ *where* $\sum_{t=0}^{K-1} q_t = 1$ *and* $\mathbb{P}[T_1 = t_1, \ldots, T_G = t_G] = \prod_{g=1}^{G} q_{t_g}$.
2. *In a group with* $T_g = t$, *individual-level treatment is assigned according to some distribution* $\mathbb{P}[D_{ig} = d, \mathbf{D}_{(i)g} = \mathbf{d}_g | T_g = t]$.

Part 1 in Definition IV.2 asserts that group categories T_g are assigned independently and with the same probability to each group. Part 2 indicates that the realized value of T_g determines the probability of individual-level treatment assignments within each group. The literature has considered two variations on the 2SR mechanism. I will formalize these two versions as follows.

Definition IV.3 (2SR-BE) *A two-stage design with Bernoulli trials (2SR-BE) is a 2SR mechanism in which:*

1. $\mathbb{P}[D_{ig} = d, \mathbf{D}_{(i)g} = \mathbf{d}_g | T_g = t] = \mathbb{P}[D_{ig} = d | T_g = t] \prod_{j=1}^{n_g+1} \mathbb{P}[D_{jig} = d_j | T_g = t]$,
2. $\mathbb{P}[D_{ig} = 1 | T_g = t] = p_t$ *for all* i, g,
3. $p_0 < p_1 < \ldots < p_{K-1}$.

Part 1 in Definition IV.3 states that within each category $T_g = t$ individual treatment is assigned following a simple randomization mechanism. Parts 2 and 3 indicate that all units with the same assignment have the same treatment probability, and that the categories are ordered in such a way that the probability of treatment increases (this ordering is without loss of generality). A common way to implement this randomization scheme is to set $K = 2$, $\mathbb{P}[D_{ig} = 1 | T_g = 0] = p_0 = 0$ and $\mathbb{P}[D_{ig} = 1 | T_g = 1] = p_1 > 0$. Hence, in this case groups are split into two categories; group assigned to category $T_g = 0$ are pure control groups and no unit is treated, whereas in groups assigned to category $T_g = 1$ units are assigned to treatment with probability p_1. The extreme case where $p_1 = 1$ corresponds to a cluster-randomized trial, as in Miguel and Kremer (2004). Duflo and Saez (2003) use this design with $p_1 = 0.5$, while Ichino and Schündeln (2012) set $p_1 = 0.25$. On the other hand, Crépon, Duflo, Gurgand, Rathelot, and Zamora (2013) set $K = 5$, $p_0 = 0$, $p_1 = 0.25$, $p_2 = 0.5$, $p_3 = 0.75$ and $p_4 = 1$.

The second type of 2SR design fixed the number of treated units in each group, instead of assignment treatment with a specified probability.

Definition IV.4 (2S-FM) *A two-stage design with fixed margins (2SR-FM) is a 2SR assignment mechanism in which:*

1. *For each* t *there is an integer* m_t *such that a group with* $T_g = t$ *will have exactly* m_t *treated and* $n + 1 - m_t$ *controls,*

2. $\mathbb{P}[D_{ig} = d, \mathbf{D}_{(i)g} = \mathbf{d} | T_g = t] = \binom{n+1}{m_t}^{-1} \mathbb{1}(m_t = d + s)$ *with* $\mathbf{1}_g' \mathbf{d}_g = s$,

3. $m_0 \leq m_1 \leq \ldots \leq m_{K-1}$.

According to Definition IV.4, in a 2SR-FM each realization $T_g = t$ for group g determines a fixed number of treated units m_t in that group. Part 2 gives the probability of each assignment implied by this mechanism, and part 3 simply sorts the values of m_t in increasing order. This mechanism implies that there is no variation in the number of treated units conditional on $T_g = t$. Baird, Bohren, McIntosh, and Özler (forthcoming) study a 2SR-FM mechanism where $K = n + 1$.

4.4.3 Optimality criteria for experimental design

4.4.3.1 Point estimation and MSE

The MSE is a natural criterion to decide how to optimally design an experiment balancing the bias and the variance of an estimator. Start by considering the simple case in which each group has only two units so that $G \to \infty$ and $n = 1$. The parameters can be estimated through the regression model:

$$\begin{aligned} Y_{ig} &= \alpha + \tau D_{ig} + \theta_0 D_{1ig}(1 - D_{ig}) + \theta_1 D_{1ig} D_{ig} + u_{ig} \\ &= \alpha + \mathbf{x}_{ig}' \boldsymbol{\beta} + u_{ig} \end{aligned}$$

The estimator is conditionally unbiased when the treatment is randomly assigned. On the other hand, assuming homoskedasticity, under standard conditions,

$$\sqrt{G}(\hat{\boldsymbol{\beta}} - \beta) \to_d \mathcal{N}(0, \mathbf{V})$$

where

$$\mathbf{V} = 2\sigma^2 \begin{bmatrix} p(1-p) & p(\pi_{11} - p) & \pi_{11}(1-p) \\ \pi_{11}(1-p) & \pi_{01}(1-\pi_{01}) & -\pi_{01}\pi_{11} \\ \pi_{11}(1-p) & -\pi_{01}\pi_{11} & \pi_{11}(1-\pi_{11}) \end{bmatrix}^{-1}$$

with $p = \mathbb{P}[D_{ig} = 1]$ and $\pi_{d1} = \mathbb{P}[D_{ig} = d, D_{1ig} = 1]$. The limiting variance is therefore a function of the assignment mechanism through (p, π_{01}, π_{11}), and this assignment can be chosen to minimize some scalar function of the variance such as its trace. For example, if the treatment follows a Bernoulli trial, the distribution is determined by p, which can be easily chosen to minimize the trace of the matrix.

In general, let $\pi(\cdot)$ be the distribution of the treatment assignments in a family of distributions Π. The variance matrix of interest $\mathbf{V}$ is a function of $\pi(\cdot)$ and possibly a vector of nuisance parameters ξ. For instance, under homoskedasticity $\xi = \sigma^2$, the variance of the error term, whereas in a random effects scenario $\xi = (\sigma_u^2, \sigma_\eta^2)$. More generally, ξ contains the variances and covariances of the potential outcomes. An optimal design problem is given by:

$$\min_{\pi(\cdot)\in\Pi} \quad \psi\left([\mathbf{V}(\pi(\cdot), \xi)]\right)$$

where $\psi(\cdot)$ is a criterion function. A natural choice for a criterion function would be $\psi(\mathbf{A}) = \text{trace}(\mathbf{V})$, which corresponds to minimizing the sum of the variances of the estimators. Another possible choice, more in line with the discussion in the previous section, is to minimize the largest variance. Other criteria used in the literature of optimal design are $\psi(\mathbf{V}) = -\det(\mathbf{V})$ or $\psi(\mathbf{V}) = -\lambda_{\min}(\mathbf{V})$, the minimum eigenvalue of $\mathbf{V}$ (see e.g. Silvey, 1980; Melas, 2006). Under a similar approach, Hirano and Hahn (2010) study variance minimization in two-categories two-stage designs like the one used by Duflo and Saez (2003), while Baird, Bohren, McIntosh, and Özler (forthcoming) analyze power in two-stage designs with fixed-margins randomization.

Because under random assignment the average potential outcome estimators are conditionally unbiased, the MSE reduces to the variance of the estimators. The approach outlined above is appropriate for the case in which groups are small. When groups are moderately large, however, the MSE needs to account for the possibility that the estimator is not defined due to insufficient observations, which can introduce a bias in finite samples. Following the notation in the previous chapter, define the estimator:

$$\hat{\mu}(\mathbf{a}) = \frac{\sum_g \sum_i Y_{ig}(\mathbf{a}) \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})} \mathbb{1}(N(\mathbf{a}) > 0)$$

The estimator is set to zero when the cell corresponding to assignment $\mathbf{a}$ is empty (the choice of the number zero is immaterial, and can be changed by any other constant). In this setting, the conditional bias of the estimator is given by

$$\mathbb{E}[\hat{\mu}(\mathbf{a})|\mathbf{A}] - \mu(\mathbf{a}) = -\mu(\mathbf{a})\mathbb{1}(N(\mathbf{a}) = 0)$$

and the conditional variance is

$$\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}] = \frac{\sigma^2(\mathbf{a})}{N(\mathbf{a})}\mathbb{1}(N(\mathbf{a}) > 0)$$

so the MSE can be approximated by

$$\mathsf{MSE}(\mathbf{a}) \approx \mu(\mathbf{a})^2\mathbb{P}[N(\mathbf{a}) = 0]^2 + \frac{\sigma^2(\mathbf{a})}{G(n+1)\pi(\mathbf{a})}$$

Using that $\mathbb{P}[N(\mathbf{a}) = 0] \leq (1 - \pi(\mathbf{a}))^G$, it follows that:

$$\mathsf{MSE}(\mathbf{a}) \leq \max_{\mathbf{a}} \mu(\mathbf{a})^2(1 - \underline{\pi}_n)^{2G} + \frac{\max_{\mathbf{a}} \sigma^2(\mathbf{a})}{G(n+1)\underline{\pi}_n}$$

where as before $\underline{\pi}_n = \min_{\mathbf{a}} \pi(\mathbf{a})$. The optimal design problem in this case is given by

$$\min_{\pi(\cdot)\in\Pi} \left\{ \max_{\mathbf{a}} \mu(\mathbf{a})^2(1 - \underline{\pi}_n)^{2G} + \frac{\max_{\mathbf{a}} \sigma^2(\mathbf{a})}{G(n+1)\underline{\pi}_n} \right\}$$

Although the criterion is different from the one in Section 4.4.3.2, both problems amount to choosing the distribution $\pi(\cdot) \in \Pi$ that makes $\underline{\pi}_n$ as large as possible.

4.4.3.2 Inference and rates of convergence

The results in Section 3.3 reveal that (i) estimators for each $\hat{\mu}(\mathbf{a})$ have different rates and (ii) both the rates of convergence and the rate at which the normal distribution approaches the distribution of the estimators depend on the assignment mechanism. In particular, when the goal is to estimate and conduct inference on the whole vector of average potential outcomes, the assignment mechanism enters the rates through the term $\underline{\pi}_n = \min_{\mathbf{a}} \pi(\mathbf{a})$. These facts suggest a principled criterion for experimental design: choose the assignment mechanism $\pi(\cdot)$ that makes these rates as fast as possible. In practical terms, this criterion corresponds to choosing the assignment mechanism such that (i) the estimators are "as close as possible" to their true values (in terms of consistency) and (ii) the normal is as good as possible an approximation to the distribution of the estimators. Given that normal-based asymptotic inference is the most common approach to conduct inference in many fields in the social sciences, these feature has a large practical appeal.

Given a family of distributions Π, an optimal design problem is given by:

$$\max_{\pi(\cdot)\in\Pi} \underline{\pi}_n$$

that is, choose the distribution $\pi(\cdot)$ within the family Π to make the minimum probability as large as possible. Top simplify the discussion, consider the case under exchangeability, so that $\mathcal{A}_n = \{(d,s); d \in \{0,1\}, s \in \{0,1,1,\ldots,n\}\}$ and a distribution $\pi(d,s)$ over $\mathcal{A}_n$. Consider a Bernoulli trial:

$$\Pi_{\mathsf{BE}} = \left\{ \pi(d,s) = \binom{n}{s} p^{d+s}(1-p)^{n+1-d-s}, d \in \{0,1\}, s \in \{0,1,\ldots,n\} \right\}$$

Under this assignment, $\underline{\pi}_n = \min\{p, 1-p\}^{n+1}$ so the optimal assignment sets $p^* = 1/2$.

Next, consider a 2S-FM assignment mechanism like the one described in Definition IV.4

$$\Pi_{\mathsf{2S-FM}} = \left\{ \pi(d,s) = q_{d+s} \left(\frac{s+1}{n+1}\right)^d \left(1 - \frac{s}{n+1}\right)^{1-d} \mathbb{1}(d+s=t \text{ for some t}), \right.$$
$$\left. d \in \{0,1\}, s \in \{0,1,\ldots,n\}, t \in \{0,1,\ldots,K-1\} \right\}$$

In this case, the design problem requires choosing values $q_0, q_2, \ldots, q_{K-1}$.

Finally, in a 2SR-BE,

$$\Pi_{\mathsf{2S-BE}} = \left\{ \pi(d,s) = \sum_{t=1}^{K-1} q_t \binom{n}{s} p_t^{d+s}(1-p_t)^{n+1-d-s}, \right.$$
$$\left. d \in \{0,1\}, s \in \{0,1,\ldots,n\}, t \in \{0,1,\ldots,K-1\} \right\}$$

so the problem requires choosing both $q_0, q_2 \ldots q_{K-1}$ and $p_0, p_1, \ldots p_{K-1}$.

4.5 Conclusion

In this chapter I discuss three key issues related to the empirical implementation of randomized control trials to estimate spillovers. The first one is the inclusion of

covariates, which is a standard practice to improve finite sample performance of the estimators or to relax the assumption of independence between potential outcomes and treatment assignments. The second issue is imperfect compliance, which occurs almost inevitably in practice since units in an experiment can be offered treatment but cannot be forced to accept it. The third issue is experimental design, which plays a crucial role in solving or at least alleviating the difficulties that arise when estimating spillover effects.

This chapter does not attempt to provide a definitive analysis of the issues mentioned above, but rather setting up a framework and pointing out the main challenges that need to be addressed in future work. Section 4.3 reveals that point identification of spillover effects in the presence of imperfect compliance is difficult and may require strong assumptions. Partial identification can be a valid alternative that should be considered in the future. On the other hand, Section 4.4 provides some optimality criteria for the design of experiments that can be appealing in practice, but further thought should be devoted to issues such as existence and uniqueness of optimal designs.

APPENDICES

APPENDIX A

Proofs for Chapter II

Proof of Lemma II.1 Fix $d \in \{0,1\}$. Let $\xi_d(\cdot) : \mathcal{D}_g \to \mathbb{R}$ be a function such that $\xi_d(\mathbf{d}_g) = \mathbb{E}[Y_{ig}(d, \mathbf{d}_g)]$ for any $\mathbf{d}_g \in \mathcal{D}_g$. For $s \in \{0, 1, \ldots, n_g\}$, let $\mathcal{A}_s = \{\mathbf{d}_g \in \mathcal{D}_g : \mathbf{1}_g'\mathbf{d}_g = s\}$, and note that $\cup_{s=0}^{n_g} \mathcal{A}_s = \mathcal{D}_g$ and $\mathcal{A}_s \cap \mathcal{A}_{\tilde{s}} = \emptyset$ for $s \neq \tilde{s}$. By Assumption II.1, the image set of each $\mathcal{A}_s$, $\xi_d(\mathcal{A}_s)$, is a singleton. Call the only element of this image set $\mu_d(s)$, and collect all these elements in $\mathcal{F}_g = \{\mu_d(s)\}_{s=0}^{n_g}$, and note that $\xi_d(\mathcal{D}_g) = \mathcal{F}_g$. Next, let $\mathcal{S}_g = \{0, 1, \ldots, n_g\}$ and define a function $\mu(d, \cdot) : \mathcal{S}_g \to \mathbb{R}$ by setting $\mu(d, s) = \mu_d(s)$ for each $s \in \mathcal{S}_g$. Then, $\mu(d, \mathcal{S}_g) = \mathcal{F}_g$ and for any $\mathbf{d}_g \in \mathcal{D}_g$ such that $\mathbf{1}_g'\mathbf{d}_g = s$, we have that $\mu(d, s) = \mathbb{E}[Y_{ig}(d, \mathbf{d}_g)]$. □

Proof of Lemma II.2 First, for all assignments with non-zero probability,

$$\begin{aligned}
\mathbb{E}[Y_{ig}|D_{ig} = d, S_{ig}] &= \mathbb{E}\left\{\mathbb{E}[Y_{ig}|D_{ig} = d, \mathbf{D}_g^{(i)}, S_{ig}]\middle| D_{ig} = d, S_{ig}\right\} \\
&= \mathbb{E}\left\{\sum_{\mathbf{d}_g} \mathbb{E}[Y_{ig}(d, \mathbf{d}_g)|D_{ig} = d, \mathbf{D}_g^{(i)}, S_{ig}]\mathbb{1}(\mathbf{D}_g^{(i)} = \mathbf{d}_g)\middle| D_{ig} = d, S_{ig}\right\} \\
&= \mathbb{E}\left\{\sum_{\mathbf{d}_g} \mathbb{E}[Y_{ig}(d, \mathbf{d}_g)]\mathbb{1}(\mathbf{D}_g^{(i)} = \mathbf{d}_g)\middle| D_{ig} = d, S_{ig}\right\} \\
&= \sum_{\mathbf{d}_g} \mathbb{E}[Y_{ig}(d, \mathbf{d}_g)]\mathbb{P}[\mathbf{D}_g^{(i)} = \mathbf{d}_g|D_{ig} = d, S_{ig}]
\end{aligned}$$

Then, since $\mathbb{P}[\mathbf{D}_g^{(i)} = \mathbf{d}_g | D_{ig} = d, S_{ig} = s_g] = 0$ if $\mathbf{1}_g'\mathbf{d}_g \neq s$,

$$\begin{aligned}
\mathbb{E}[Y_{ig}|D_{ig} = d, S_{ig} = s] &= \sum_{\mathbf{d}_g : \mathbf{1}_g'\mathbf{d}_g = s_g} \mathbb{E}[Y_{ig}(d, \mathbf{d}_g)]\mathbb{P}[\mathbf{D}_g^{(i)} = \mathbf{d}_g | D_{ig} = d, S_{ig} = s] \\
&= \mu(d, s) \sum_{\mathbf{d}_g : \mathbf{1}_g'\mathbf{d}_g = s_g} \mathbb{P}[\mathbf{D}_g^{(i)} = \mathbf{d}_g | D_{ig} = d, S_{ig} = s] \\
&= \mu(d, s).
\end{aligned}$$

The second equality follows from the definition of $\theta_s(d)$. □

Proof of Lemma II.3 Use the fact that

$$\begin{aligned}
\mathbb{E}[Y_{ig}|D_{ig} = d] &= \sum_s \mathbb{E}[Y_{ig}|D_{ig} = d, S_{ig} = s]\mathbb{P}[S_{ig} = s|D_{ig} = d] \\
&= \sum_s \mathbb{E}[Y_{ig}(d, s)]\mathbb{P}[S_{ig} = s|D_{ig} = d] \\
&= \tau_0 + \sum_s \theta_s(d)\mathbb{P}[S_{ig} = s|D_{ig} = d]
\end{aligned}$$

where the second equality uses the independence assumption. □

Proof of Lemma II.4 By independence between D_{ig} and S_{ig} under simple random assignment,

$$\gamma_\ell = \frac{\mathbb{C}\text{ov}(Y_{ig}, \bar{D}_g^{(i)})}{\mathbb{V}[\bar{D}_g^{(i)}]} = n_g \frac{\mathbb{C}\text{ov}(Y_{ig}, S_{ig})}{\mathbb{V}[S_{ig}]} = n_g \frac{\mathbb{C}\text{ov}(\mathbb{E}[Y_{ig}|S_{ig}], S_{ig})}{\mathbb{V}[S_{ig}]}$$

but $\mathbb{E}[Y_{ig}|S_{ig}] = p\mathbb{E}[Y_{ig}|D_{ig} = 1, S_{ig}] + (1-p)\mathbb{E}[Y_{ig}|D_{ig} = 0, S_{ig}]$, and

$$\mathbb{E}[Y_{ig}|D_{ig} = d, S_{ig}] = \mu(0, 0) + \tau_0 d + \sum_s \theta_s(d)\mathbb{1}(S_{ig} = s)$$

and calculating the covariance gives the result. □

Proof of Lemma II.5 Follows by the same argument as the previous lemma but conditioning on $D_{ig} = d$. □

APPENDIX B

Proofs for Chapter III

Lemma B.1 *Let* $\hat{\pi}(\mathbf{a}) := \sum_g \sum_i \mathbb{1}_{ig}(\mathbf{a})/G(n+1)$. *Then under the assumptions of Theorem III.1,*

$$\max_{\mathbf{a}\in\mathcal{A}_n} \left| \frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})} - 1 \right| \to_{\mathbb{P}} 0.$$

Proof of Lemma B.1 Take $\varepsilon > 0$, then

$$\begin{aligned}
\mathbb{P}\left[\max_{\mathbf{a}\in\mathcal{A}_n} \left| \frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})} - 1 \right| > \varepsilon \right] &\leq \sum_{\mathbf{a}\in\mathcal{A}_n} \mathbb{P}\left[\left| \frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})} - 1 \right| > \varepsilon \right] \\
&\leq |\mathcal{A}_n| \max_{\mathbf{a}\in\mathcal{A}_n} \mathbb{P}\left[|\hat{\pi}(\mathbf{a}) - \pi(\mathbf{a})| > \varepsilon \pi(\mathbf{a}) \right] \\
&\leq 2|\mathcal{A}_n| \max_{\mathbf{a}\in\mathcal{A}_n} \exp\left\{ -\frac{2G^2\varepsilon^2\pi(\mathbf{a})^2}{G} \right\} = 2|\mathcal{A}_n| \exp\left\{ -2\varepsilon^2 G \underline{\pi}_n^2 \right\} \\
&= 2\exp\left\{ -G\underline{\pi}_n^2 \left(2\varepsilon^2 - \frac{\log|\mathcal{A}_n|}{G\underline{\pi}_n^2} \right) \right\} \to 0
\end{aligned}$$

where the second line uses Hoeffding's inequality. □

Proof of Theorem III.1 Take a constant $c \in \mathbb{R}$. Then

$$\mathbb{P}\left[\min_{\mathbf{a}\in\mathcal{A}_n} N(\mathbf{a}) \leq c \right] \leq |\mathcal{A}_n| \max_{\mathbf{a}\in\mathcal{A}_n} \mathbb{P}[N(\mathbf{a}) \leq c].$$

Now, for any $\delta > 0$,

$$\begin{aligned}
\mathbb{P}[N(\mathbf{a}) \leq c] &= \mathbb{P}\left[N(\mathbf{a}) \leq c, \left|\frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})} - 1\right| > \delta\right] + \mathbb{P}\left[N(\mathbf{a}) \leq c, \left|\frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})} - 1\right| \leq \delta\right] \\
&\leq \mathbb{P}\left[\left|\frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})} - 1\right| > \delta\right] \\
&\quad + \mathbb{P}[N(\mathbf{a}) \leq c, G(n+1)\pi(\mathbf{a})(1-\delta) \leq N(\mathbf{a}) \leq \pi(\mathbf{a})G(n+1)(1+\delta)] \\
&\leq \mathbb{P}\left[\left|\frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})} - 1\right| > \delta\right] + \mathbb{1}(G(n+1)\pi(\mathbf{a}) \leq c/(1-\delta)) \\
&\leq \mathbb{P}\left[\left|\frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})} - 1\right| > \delta\right] + \mathbb{1}(G(n+1)\underline{\pi}_n \leq c/(1-\delta))
\end{aligned}$$

which implies

$$\begin{aligned}
|\mathcal{A}_n| \max_{\mathbf{a}\in\mathcal{A}_n} \mathbb{P}[N(\mathbf{a}) \leq c] \leq |\mathcal{A}_n| \max_{\mathbf{a}\in\mathcal{A}_n} \mathbb{P}\left[\left|\frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})} - 1\right| > \delta\right] \\
+ |\mathcal{A}_n| \mathbb{1}(G(n+1)\underline{\pi}_n \leq c/(1-\delta))
\end{aligned}$$

which converges to zero under condition 3.5 and using Lemma B.1. □

Lemma B.2 *Under the assumptions of Theorem III.1,*

$$\max_{\mathbf{a}\in\mathcal{A}_n} \left\{\left|\frac{\pi(\mathbf{a})}{\hat{\pi}(\mathbf{a})} - 1\right| \cdot \mathbb{1}(N(\mathbf{a}) > 0)\right\} \to_{\mathbb{P}} 0$$

Proof of Lemma B.2 follows from Lemma B.1 and Theorem III.1 using that $\mathbb{P}[\min_{\mathbf{a}\in\mathcal{A}_n} N(\mathbf{a}) = 0] \to 0$. □

Proof of Theorem III.2 All the estimators below are only defined when $\mathbb{1}(N(\mathbf{a}) > 0)$. Because under the conditions for Theorem III.1 this event occurs with probability approaching one, the indicator will be omitted to simplify the notation. For the consistency part, we have that

$$\begin{aligned}
\frac{\sum_g \sum_i \varepsilon_{ig}(\mathbf{a}) \mathbb{1}_{ig(\mathbf{a})}}{N(\mathbf{a})} &= \frac{\sum_g \sum_i (\varepsilon_{ig}(\mathbf{a}) \mathbb{1}(|\varepsilon_{ig}| > \xi_n) - \mathbb{E}[\varepsilon_{ig}(\mathbf{a}) \mathbb{1}(|\varepsilon_{ig}| > \xi_n)]) \mathbb{1}_{ig(\mathbf{a})}}{N(\mathbf{a})} \\
&\quad + \frac{\sum_g \sum_i (\varepsilon_{ig}(\mathbf{a}) \mathbb{1}(|\varepsilon_{ig}| \leq \xi_n) - \mathbb{E}[\varepsilon_{ig}(\mathbf{a}) \mathbb{1}(|\varepsilon_{ig}| \leq \xi_n)]) \mathbb{1}_{ig(\mathbf{a})}}{N(\mathbf{a})}
\end{aligned}$$

for some increasing sequence of constants ξ_n whose rate will be determined along the proof. Let

$$\underline{\varepsilon}_{ig}(\mathbf{a}) = \varepsilon_{ig}(\mathbf{a})\mathbb{1}(|\varepsilon_{ig}(\mathbf{a})| \leq \xi_n) - \mathbb{E}[\varepsilon_{ig}(\mathbf{a})\mathbb{1}(|\varepsilon_{ig}(\mathbf{a})| \leq \xi_n)]$$

and

$$\bar{\varepsilon}_{ig}(\mathbf{a}) = \varepsilon_{ig}(\mathbf{a})\mathbb{1}(|\varepsilon_{ig}(\mathbf{a})| > \xi_n) - \mathbb{E}[\varepsilon_{ig}(\mathbf{a})\mathbb{1}(|\varepsilon_{ig}(\mathbf{a})| > \xi_n)]$$

For the first term,

$$\mathbb{P}\left[\max_{\mathbf{a}\in\mathcal{A}_n}\left|\frac{\sum_g\sum_i \underline{\varepsilon}_{ig}(\mathbf{a})\mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})}\right| > Mr_n \middle| \mathbf{A}\right] \leq$$
$$|\mathcal{A}_n| \max_{\mathbf{a}\in\mathcal{A}_n}\mathbb{P}\left[\left|\frac{\sum_g\sum_i \underline{\varepsilon}_{ig}(\mathbf{a})\mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})}\right| > Mr_n \middle| \mathbf{A}\right]$$

For the right-hand side, by Bernstein's inequality

$$\begin{aligned}\mathbb{P}\left[\left|\sum_g\sum_i \underline{\varepsilon}_{ig}(\mathbf{a})\mathbb{1}_{ig}(\mathbf{a})\right| > N(\mathbf{a})Mr_n \middle| \mathbf{A}\right] &\leq 2\exp\left\{-\frac{1}{2}\frac{M^2r_n^2N(\mathbf{a})^2}{\sigma^2(\mathbf{a})N(\mathbf{a}) + 2\xi_nMr_nN(\mathbf{a})/3}\right\} \\ &= 2\exp\left\{-\frac{1}{2}\frac{M^2r_n^2N(\mathbf{a})}{\sigma^2(\mathbf{a}) + 2M\xi_nr_n/3}\right\} \\ &\leq 2\exp\left\{-\frac{1}{2}\frac{M^2r_n^2\min_{\mathbf{a}\in\mathcal{A}_n}N(\mathbf{a})}{\bar{\sigma}^2 + 2M\xi_nr_n/3}\right\}\end{aligned}$$

Set

$$r_n = \sqrt{\frac{\log|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n}}$$

Next, use the fact that

$$\frac{\min_{\mathbf{a}\in\mathcal{A}_n}N(\mathbf{a})}{G(n+1)\underline{\pi}_n} \to_{\mathbb{P}} 1$$

which follows from Lemma B.1, since

$$\mathbb{P}[\pi(\mathbf{a})(1-\varepsilon) \leq \hat{\pi}(\mathbf{a}) \leq \pi(\mathbf{a})(1+\varepsilon),\ \forall\,\mathbf{a}] \to 1$$

for any $\varepsilon > 0$ and

$$\mathbb{P}[\pi(\mathbf{a})(1-\varepsilon) \leq \hat{\pi}(\mathbf{a}) \leq \pi(\mathbf{a})(1+\varepsilon),\, \forall\, \mathbf{a}] \leq \mathbb{P}[\underline{\pi}_n(1-\varepsilon) \leq \min_{\mathbf{a}} \hat{\pi}(\mathbf{a}) \leq \underline{\pi}_n(1+\varepsilon)]$$
$$= \mathbb{P}\left[\left|\frac{\min_{\mathbf{a}} \hat{\pi}(\mathbf{a})}{\underline{\pi}_n} - 1\right| \leq \varepsilon\right].$$

Then,

$$\mathbb{P}\left[\left|\sum_g \sum_i \underline{\varepsilon}_{ig}(\mathbf{a}) \mathbb{1}_{ig}(\mathbf{a})\right| > N(\mathbf{a}) M r_n \middle| \mathbf{A}\right] \leq 2\exp\left\{-\frac{1}{2}\frac{M^2 \log|\mathcal{A}_n|(1+o_{\mathbb{P}}(1))}{\bar{\sigma}^2 + 2M\xi_n r_n/3}\right\}$$

and therefore

$$\mathbb{P}\left[\max_{\mathbf{a}\in\mathcal{A}_n}\left|\frac{\sum_g \sum_i \underline{\varepsilon}_{ig}(\mathbf{a}) \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})}\right| > M r_n \middle| \mathbf{A}\right] \leq$$
$$2\exp\left\{\log|\mathcal{A}_n|\left(1 - \frac{1}{2}\frac{M^2(1+o_{\mathbb{P}}(1))}{\bar{\sigma}^2 + 2M r_n \xi_n/3}\right)\right\}$$

which can be made arbitrarily small for sufficiently large M as long as $r_n \xi_n = O(1)$.

For the second term, by Markov's inequality

$$\mathbb{P}\left[\left|\sum_g \sum_i \bar{\varepsilon}_{ig}(\mathbf{a}) \mathbb{1}_{ig}(\mathbf{a})\right| > N(\mathbf{a}) M r_n \middle| \mathbf{A}\right] \leq \frac{\mathbb{E}[\varepsilon_{ig}^2(\mathbf{a}) \mathbb{1}(|\varepsilon_{ig}(\mathbf{a})| > \xi_n)] N(\mathbf{a})}{M^2 r_n^2 N(\mathbf{a})^2}$$
$$\leq \frac{\bar{\sigma}^2}{M^2 \xi_n^{\delta}} \frac{1}{r_n^2 \min_{\mathbf{a}\in\mathcal{A}_n} N(\mathbf{a})}$$
$$= \frac{\bar{\sigma}^2}{M^2} \frac{1}{\xi_n^{\delta} \log|\mathcal{A}_n|(1+o_{\mathbb{P}}(1))}$$

so that

$$\mathbb{P}\left[\max_{\mathbf{a}\in\mathcal{A}_n}\left|\sum_g \sum_i \bar{\varepsilon}_{ig}(\mathbf{a}) \mathbb{1}_{ig}(\mathbf{a})\right| > N(\mathbf{a}) M r_n \middle| \mathbf{A}\right] \leq \frac{\bar{\sigma}^2}{M^2} \frac{|\mathcal{A}_n|}{\xi_n^{\delta} \log|\mathcal{A}_n|(1+o_{\mathbb{P}}(1))}$$

Finally, set $\xi_n = r_n^{-1}$. Then, the above term can be made arbitrarily small for M sufficiently large, as long as

$$\frac{|\mathcal{A}_n|}{\log|\mathcal{A}_n|}\left(\frac{\log|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n}\right)^{\delta/2} = O(1)$$

Setting $\delta = 2$, this condition reduces to:

$$\frac{|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n} = O(1)$$

Therefore,

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

The proof for the standard error estimator uses the same reasoning after replacing $\varepsilon_{ig}(\mathbf{a})$ by $\hat{\varepsilon}^2_{ig}(\mathbf{a})$ and using consistency of $\hat{\mu}(\mathbf{a})$.

For the second part, we want to bound

$$\Delta = \max_{\mathbf{a}\in\mathcal{A}_\mathbf{n}} \sup_{x\in\mathbb{R}} \left| \mathbb{P}\left[\frac{\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}} \leq x \right] - \Phi(x) \right|$$

$$\begin{aligned}
\Delta &= \max_{\mathbf{a}\in\mathcal{A}_\mathbf{n}} \sup_{x\in\mathbb{R}} \left| \mathbb{P}\left[\frac{\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}} \leq x \right] - \Phi(x) \right| \\
&= \max_{\mathbf{a}\in\mathcal{A}_\mathbf{n}} \sup_{x\in\mathbb{R}} \left| \mathbb{E}\left\{ \mathbb{P}\left[\frac{\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}} \leq x \middle| \mathbf{A} \right] - \Phi(x) \right\} \right| \\
&\leq \mathbb{E}\left\{ \max_{\mathbf{a}\in\mathcal{A}_\mathbf{n}} \sup_{x\in\mathbb{R}} \left| \mathbb{P}\left[\frac{\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}} \leq x \middle| \mathbf{A} \right] - \Phi(x) \right| \right\}
\end{aligned}$$

Then,

$$\left| \mathbb{P}\left[\frac{\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}} \leq x \middle| \mathbf{A} \right] - \Phi(x) \right| = \left| \mathbb{P}\left[\frac{\sum_g \sum_i \varepsilon_{ig} \mathbb{1}_{ig}(\mathbf{a})}{\sigma(\mathbf{a})\sqrt{N(\mathbf{a})}} \leq x \middle| \mathbf{A} \right] - \Phi(x) \right|$$

By the Berry-Esseen bound,

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left[\frac{\sum_g \sum_i \varepsilon_{ig} \mathbb{1}_{ig}(\mathbf{a})}{\sigma(\mathbf{a})\sqrt{N(\mathbf{a})}} \leq x \middle| \mathbf{A} \right] - \Phi(x) \right| \leq \frac{C\bar{\tau}^3}{\underline{\sigma}^3} \cdot \frac{1}{\sqrt{N(\mathbf{a})}}$$

But

$$\frac{1}{N(\mathbf{a})} = O_{\mathbb{P}}\left(\frac{1}{G(n+1)\pi(\mathbf{a})} \right)$$

Therefore,

$$\max_{\mathbf{a}\in\mathcal{A}_n}\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left[\frac{\sum_g\sum_i\varepsilon_{ig}\mathbb{1}_{ig}(\mathbf{a})}{\sigma(\mathbf{a})\sqrt{N(\mathbf{a})}}\leq x\middle|\mathbf{A}\right]-\Phi(x)\right|\leq\frac{C\bar{\tau}^3}{\underline{\sigma}^3}\cdot O_{\mathbb{P}}\left(\frac{1}{\sqrt{G(n+1)\underline{\pi}_n}}\right)$$

and the result follows. □

Proof of Theorem III.3 We want to bound:

$$\Delta^*(\mathbf{a})=\sup_x\left|\mathbb{P}^*\left[\frac{\hat{\mu}^*(\mathbf{a})-\hat{\mu}(\mathbf{a})}{\sqrt{\mathbb{V}^*[\hat{\mu}(\mathbf{a})]}}\leq x\right]-\Phi(x)\right|$$

uniformly over $\mathbf{a}$, where

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

when the denominator is non-zero, which occurs with probability approaching one, and where

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

Then,

$$\begin{aligned}\mathbb{E}^*[\hat{\mu}^*(\mathbf{a})]&=\hat{\mu}(\mathbf{a})\\ \mathbb{V}^*[\hat{\mu}^*(\mathbf{a})]&=\frac{\sum_g\sum_i\hat{\varepsilon}_{ig}^2\mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})^2}\end{aligned}$$

The centered and scaled statistic is given by:

$$\frac{\sum_g\sum_i\hat{\varepsilon}_{ig}\mathbb{1}_{ig}(\mathbf{a})w_{ig}}{\sqrt{\sum_g\sum_i\hat{\varepsilon}_{ig}^2\mathbb{1}_{ig}(\mathbf{a})}}$$

By Berry-Esseen,

$$\sup_x \left| \mathbb{P}^* \left[\frac{\sum_g \sum_i \hat{\varepsilon}_{ig} \mathbb{1}_{ig}(\mathbf{a}) w_{ig}}{\sqrt{\sum_g \sum_i \hat{\varepsilon}_{ig}^2 \mathbb{1}_{ig}(\mathbf{a})}} \leq x \right] - \Phi(x) \right| \leq C \frac{\sum_g \sum_i |\hat{\varepsilon}_{ig}|^3 \mathbb{1}_{ig}(\mathbf{a})/N(\mathbf{a})}{\left(\sum_g \sum_i \hat{\varepsilon}_{ig}^2 \mathbb{1}_{ig}(\mathbf{a})/N(\mathbf{a}) \right)^{3/2}} \cdot \frac{1}{\sqrt{N(\mathbf{a})}}$$

We also have that

$$\begin{aligned} \frac{\sum_g \sum_i |\hat{\varepsilon}_{ig}|^3 \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})} &\leq \frac{\sum_g \sum_i |Y_{ig} - \mu(\mathbf{a})|^3 \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})} + |\bar{Y}(\mathbf{a}) - \mu(\mathbf{a})|^3 + O_{\mathbb{P}}(N(\mathbf{a})^{-2}) \\ &= \mathbb{E}[|Y_{ig} - \mu(\mathbf{a})|^3] + O_{\mathbb{P}}(N(\mathbf{a})^{-1}) \end{aligned}$$

and

$$\begin{aligned} \frac{\sum_g \sum_i \hat{\varepsilon}_{ig}^2 \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})} &= \frac{\sum_g \sum_i (Y_{ig} - \mu(\mathbf{a}))^2 \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})} + (\bar{Y}(\mathbf{a}) - \mu(\mathbf{a}))^2 \\ &= \sigma^2(\mathbf{a}) + O_{\mathbb{P}}(N(\mathbf{a})^{-1}) \end{aligned}$$

Then,

$$\begin{aligned} \Delta^*(\mathbf{a}) &\leq \sup_x \left| \mathbb{P}^* \left[\frac{\sum_g \sum_i \hat{\varepsilon}_{ig} \mathbb{1}_{ig}(\mathbf{a}) w_{ig}}{\sqrt{\sum_g \sum_i \hat{\varepsilon}_{ig}^2 \mathbb{1}_{ig}(\mathbf{a})}} \leq x \right] - \Phi(x) \right| \\ &= C \frac{\mathbb{E}[|Y_{ig} - \mu(\mathbf{a})|^3] + O_{\mathbb{P}}(N(\mathbf{a})^{-1})}{[\sigma^2(\mathbf{a}) + O_{\mathbb{P}}(N(\mathbf{a})^{-1}))]^{3/2}} \cdot \frac{1}{\sqrt{N(\mathbf{a})}} \end{aligned}$$

and the result follows Lemma B.1. □

Proof of Corollary III.1 Without exchangeability, $\pi(\mathbf{a}) = \pi(d, \mathbf{d}_g) = p^d(1-p)^{1-d} \prod_{j=1}^n p^{d_j}(1-p)^{d_j} = p^{d+s}(1-p)^{n+1-s-d}$ where $s = \mathbf{1}_g' \mathbf{d}_\mathbf{g}$. On the other hand, under exchangeability $\pi(\mathbf{a}) = \pi(d, s) = p^d(1-p)^{1-d}\binom{n}{s} p^s(1-p)^{n-s} = \binom{n}{s} p^{s+d}(1-p)^{n+1-s-d}$. Observe that both distributions are minimized at $\underline{\pi}_n = \underline{p}^{n+1} \propto \underline{p}^n$ where

$\underline{p} = \min\{p, 1-p\}$. Thus,

$$\frac{\log|\mathcal{A}_n|}{G\underline{p}^{2n}} = \exp\left\{-\log G\left(1 - \frac{n+1}{\log G}2\log\underline{p} - \frac{\log\log|\mathcal{A}_n|}{\log G}\right)\right\}$$

and since $|\mathcal{A}_n|$ is at most 2^{n+1}, if $(n+1)/\log G \to 0$ the term converge to zero. □

Proof of Corollary III.2 Without exchangeability, $\mathbb{P}[D_{ig} = d, \mathbf{D}_g^{(i)} = \mathbf{d}_g] = \binom{n+1}{m_w}^{-1} q_w$ where $w = d + \mathbf{1}_g'\mathbf{d}_g$, which in this case reduces to $\binom{n+1}{m_w}^{-1}/(n+1)$. This function has a unique minimum at $(n+1)/2$ when n is odd, and two minima at $\lfloor(n+1)/2\rfloor$ and $\lfloor(n+1)/2\rfloor + 1$ when n is even. For simplicity, assume n is odd (otherwise, take $m^* = \lfloor(n+1)/2\rfloor$ to be the minimizer of the function, and use the fact that $(n+1)/2 \leq m^* \leq (n+1)/2+1$). The smallest probability is given by $\underline{\pi}_n = \frac{[(n+1)/2)!]^2}{(n+1)!(n+1)}$. Using Stirling's formula, we have that

$$\underline{\pi} = \frac{\pi(n+1)((n+1)/2)^{n+1}e^{-(n+1)}}{\sqrt{(2\pi(n+1))}(n+1)^{n+1}e^{-(n+1)}(n+1)}(1+o(1)) = \sqrt{\frac{\pi}{2}}\cdot\frac{1}{2^{n+1}\sqrt{n+1}}(1+o(1))$$

Then,

$$\frac{\log|\mathcal{A}_n|}{G\underline{\pi}_n^2} = \exp\left\{-\log G\left(1 - \frac{n+1}{\log G}\log 2 - \frac{3}{2}\cdot\frac{\log(n+1)}{\log G} + o(1)\right)\right\} \to 0$$

when $(n+1)/\log G \to 0$. With exchangeability, $\mathbb{P}[D_{ig} = d, S_{ig} = s] = q_{d+s}\left(\frac{s+1}{n+1}\right)^d \times \left(1 - \frac{s}{n+1}\right)^{1-d}$ which in this case equals $\frac{1}{n+1}\left(\frac{s+1}{n+1}\right)^d\left(1-\frac{s}{n+1}\right)^{1-d}$. This function has two minima, one at $(d,s) = (0,n)$ and one at $(d,s) = (1,0)$, giving the same minimized value of $\underline{\pi}_n = (n+1)^{-2}$. Then,

$$\frac{\log|\mathcal{A}_n|}{G\underline{\pi}_n^2} = \exp\left\{-\log G\left(1 - \frac{\log(n+1)}{\log G}4 - \frac{\log\log 2(n+1)}{\log G} + o(1)\right)\right\} \to 0$$

if $\log(n+1)/\log G \to 0$. □

BIBLIOGRAPHY

BIBLIOGRAPHY

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.

Angrist, J. D., and G. W. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90(430), 431–442.

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

Baird, S., A. Bohren, C. McIntosh, and B. Özler (forthcoming): "Optimal Design of Experiments in the Presence of Interference," *Review of Economics and Statistics*.

Baird, S., J. H. Hicks, M. Kremer, and E. Miguel (2016): "Worms at Work: Long-run Impacts of a Child Health Investment," *The Quarterly Journal of Economics*, 131(4), 1637.

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*.

Basse, G., A. Feller, and P. Toulis (2018): "Exact conditional randomization tests for causal effects under interference," *arXiv:1709.08036*.

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.

Bloom, N., M. Schankerman, and J. Van Reenen (2013): "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 81(4), 1347–1393.

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.

Carrell, S. E., and M. L. Hoekstra (2010): "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids," *American Economic Journal: Applied Economics*, 2(1), 211–28.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013): "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 81(3), 855–882.

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 Giorgi, G., M. Pellizzari, and S. Redaelli (2010): "Identification of Social Interactions through Partially Overlapping Peer Groups," *American Economic Journal: Applied Economics*, 2(2), 241–75.

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.

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

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

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

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).

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

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.

Figal Garone, L., A. Maffioli, J. A. de Negri, C. M. Rodriguez, and G. Vazquez-Bare (2015): "Cluster development policy, SME's performance, and spillovers: evidence from Brazil," *Small Business Economics*, 44(4), 925–948.

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.

Graham, B. S., and J. Hahn (2005): "Identification and estimation of the linear-in-means model of social interactions," *Economics Letters*, 88(1), 1–6.

Greenstone, M., R. Hornbeck, and E. Moretti (2010): "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 118(3), 536–598.

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

Halloran, M. E., and C. J. Struchiner (1991): "Study Designs for Dependent Happenings," *Epidemiology*, 2(5), 331–338.

——— (1995): "Causal Inference in Infectious Diseases," *Epidemiology*, 6(2), 142–151.

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.

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

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

Kao, E. K. (2017): "Causal Inference Under Network Interference: A Framework for Experiments on Social Networks," *arXiv:1708.08522*.

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.

Lazzati, N. (2015): "Treatment response with social interactions: Partial identification via monotone comparative statics," *Quantitative Economics*, 6(1), 49–83.

Lee, L.-F. (2007): "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects," *Journal of Econometrics*, 140(2), 333–374.

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.

Lyle, D. S. (2007): "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point," *The Review of Economics and Statistics*, 89(2), 289–299.

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.

——— (2013a): "Comment," *Journal of Business & Economic Statistics*, 31(3), 273–275.

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

Melas, V. B. (2006): *Functional Approach to Optimal Experimental Design*. Springer New York.

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.

Philipson, T. J. (2000): "External Treatment Effects and Program Implementation Bias," *NBER Technical Working Papers*, (250).

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

Rincke, J., and C. Traxler (2011): "Enforcement Spillovers," *The Review of Economics and Statistics*, 93(4), 1224–1234.

Rubin, D. B. (1990): "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies," *Statistical Science*, 5(4), 472–480.

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

——— (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.

Silvey, S. D. (1980): *Optimal Design: An Introduction to the Theory for Parameter Estimation*. Springer Netherlands.

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.

Stinebrickner, R., and T. R. Stinebrickner (2006): "What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds," *Journal of Public Economics*, 90(8), 1435–1454.

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.

Topa, G. (2001): "Social Interactions, Local Spillovers and Unemployment," *The Review of Economic Studies*, 68(2), 261–295.

Toulis, P., and E. Kao (2013): "Estimation of Causal Peer Influence Effects," in *Proceedings of the 30th International Conference on Machine Learning*, ed. by S. Dasgupta, and D. McAllester, vol. 28 of *Proceedings of Machine Learning Research*, pp. 1489–1497, Atlanta, Georgia, USA.

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

Yang, D. (2008): "Can Enforcement Backfire? Crime Displacement in the Context of Customs Reform in the Philippines," *The Review of Economics and Statistics*, 90(1), 1–14.

Zimmerman, D. J. (2003): "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *The Review of Economics and Statistics*, 85(1), 9–23.