Preliminary Version
Comments Welcome.

Identifying Causal Effects in Experiments with Social Interactions and Non-compliance*†

Francis J. DiTraglia‡[1], Camilo García-Jimeno[2], Rossa O'Keeffe-O'Donovan[1], and Alejandro Sánchez-Becerra[3]

[1]Department of Economics, University of Oxford
[2]Federal Reserve Bank of Chicago & NBER
[3]Department of Economics, University of Pennsylvania

This Version: April 7, 2020, First Version: April 7, 2020

Abstract

This paper shows how to use a randomized saturation experimental design to identify and estimate causal effects in the presence of social interactions—one person's treatment may affect another's outcome–and one-sided non-compliance–subjects can only be offered treatment, not compelled to take it up. Two distinct causal effects are of interest in this setting: direct effects quantify how a person's own treatment changes her outcome, while indirect effects quantify how her peers' treatments change her outcome. We consider the case in which social interactions occur only within known groups, and take-up decisions do not depend on peers' offers. In this setting we point identify local average treatment effects, both direct and indirect, in a flexible random coefficients model that allows for both heterogenous treatment effects and endogeneous selection into treatment. We go on to propose a feasible, kernel-based estimator.

Keywords: social interactions, spillovers, non-compliance, randomized saturation

JEL Codes: C14, C21, C26, C90

*The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of Chicago or the Federal Reserve System.

†We thank Steve Bond, Max Kasy, Christina Goldschmidt, seminar participants at The Philadelphia Fed, the 2018 IAAE Annual Conference, Oxford, UPenn, and the 2018 SEA Annual Meetings for helpful comments and suggestions.

‡Corresponding Author: francis.ditraglia@economics.ox.ac.uk, Manor Road, Oxford OX1 3UQ, UK.

1 Introduction

Random saturation experiments provide a powerful tool for estimating causal effects in the presence of social interactions—also known as spillovers or interference—by generating exogenous variation in both individuals' own treatment offers and the fraction of their peers who are offered treatment (Hudgens and Halloran, 2008). These two sources of variation allow researchers to study both direct causal effects—the effect of Alice's treatment on her own outcome—and indirect causal effects—the effect of Bob's treatment on Alice's outcome. A complete understanding of both direct and indirect effects is crucial for program evaluation in settings with social interactions. When considering a national job placement program, for example, policymakers may worry that the indirect effects of the program could completely offset the direct effects: in a slack labor market, job placement could merely change who is employed without affecting the overall employment rate (Crépon et al., 2013).

In this paper we provide methods that use data from a randomized saturation design to identify and estimate direct and indirect causal effects in the presence of social interactions and one-sided non-compliance. In real-world experiments non-compliance is the norm rather than the exception. In their study of the French labor market, Crépon et al. (2013) found that only 35% of workers offered job placement services took them up. Despite pervasive non-compliance in practice, most of the existing literature on randomized saturation designs either assumes perfect compliance—all subjects adhere to their experimentally-assigned treatment allocation–or identifies only intent-to-treat-effects—the effect of being *offered* treatment. In contrast, we use the experimental design as a source of instrumental variables to estimate local average treatment effects (LATE) when subjects endogenously select into treatment on the basis of their experimental offers. In a world of homogeneous treatment effects, a simple instrumental variables (IV) regression using individual treatment offers and group saturations as instruments would identify both direct and indirect effects. In most if not all real-world settings, however, treatment effects vary across individuals. In the presence of heterogeneity, this "naïve" IV approach will not in general recover interpretable causal effects. To allow for realistic patterns of heterogeneity in a tractable framework, we study a flexible random coefficients model in which causal effects may depend on an individual's treatment take-up as well as that of her peers.

Our approach relies on four key assumptions. First is *partial interference*: we assume that each subject belongs to a single, known group and that social interactions occur only within groups. This is reasonable in many experimental settings where, for example, groups correspond to villages, and social interactions across them are negligible. Second is *anonymous interactions*: we assume that individuals' potential outcome functions depend on their peers'

treatment take-up only through the *average* take-up in their group. Under this assumption only the number of treated neighbors matters, not their identities (Manski, 2013). In the absence of detailed network data, the assumption of anonymous interactions is a natural starting point and is likely to be reasonable in settings such as the labor market example described above. Third is *one-sided non-compliance*: we assume that the only individuals who can take up treatment are those to whom treatment was offered via the experimental design. One-sided non-compliance is relatively common in practice, for example when an "encouragement design" is used to introduce a new program, product or technology that is otherwise unavailable (e.g. Crépon et al., 2013; Miguel and Kremer, 2004). We refer to our fourth key assumption as *individualized offer response*, or IOR for short. IOR requires that each subject's treatment take-up decision depends only on her own treatment offer, and not on the offers made to her peers. While IOR is a strong assumption, it is testable and *a priori* reasonable in many contexts. In Crépon et al. (2013), for example, local labor markets are large and potential participants in the job placement program are unlikely to know each other in advance. As such, they are unlikely to influence each other's treatment take-up decisions, even if they may impose employment externalities on one another. IOR is also reasonable in online settings where other subjects' take-up decisions are unobserved (Anderson et al., 2014; Bond et al., 2012; Eckles et al., 2016) or confidential (Yi et al., 2015).

Because it rules out any form of strategic take-up, IOR allows us to divide the population into never-takers and compliers, two of the traditional LATE strata.[1] Under the randomized saturation design and a standard exclusion restriction, we provide linear population IV regressions that identify average causal effects for a certain type of person: someone in a group of size n with a share $\bar{c}$ of compliers among her neighbors. Averaging these "localized" effects over the distribution of n and $\bar{c}$ in the population identifies LATE-type direct and indirect causal effects.[2] The key to our approach is a result showing that conditioning on n and $\bar{c}$ breaks any dependence between peers' average take-up and an individual's random coefficients. Under the randomized saturation design, the share of Alice's neighbors who are offered treatment is exogenous. Under IOR, their average take-up depends only on how many of them are compliers and whether they are offered treatment. Thus, conditional on n and $\bar{c}$, any residual variation in the take-up of Alice's neighbors comes solely from the experimental design. Although group size is observed, the share of compliers in a given group is not. In a large group, however, the rate of take-up among those offered treatment, call it $\widehat{c}$, closely approximates $\bar{c}$. Moreover, when all groups are large, conditioning on group size becomes irrelevant. Accordingly, we propose kernel-based estimators of the "localized"

[1] One-sided non-compliance rules out always-takers and defiers.

[2] Hoderlein and Sherman (2015) call this two-step procedure the "localize-then-average" approach.

treatment effects that condition only on $\widehat{c}$.

This paper relates most closely to recent work by Kang and Imbens (2016) and Imai et al. (2018), who also study randomized saturation experiments with social interactions under non-compliance. Imai et al. (2018) identify a "complier average direct effect" (CADE), in essence a Wald estimand calculated for all groups with the same share of offers (saturation). While it is identified under a weaker condition than IOR, the CADE is in fact a hybrid of direct and indirect effects unless one is willing to impose IOR. Under IOR, the CADE quantifies the effect of an individual's own treatment take-up, given that her group has been assigned a particular saturation. In contrast, the direct effects that we recover below quantify the effect of an individual's own treatment take-up given that a certain share of her neighbors have *taken up* treatment. Kang and Imbens (2016) identify effects similar to those of Imai et al. (2018) using a variant of our IOR assumption that they call "personalized encouragement." Both Kang and Imbens (2016) and Imai et al. (2018) identify well-defined effects while placing limited structure on the potential outcome functions. The cost of this generality is that the effects they recover have a "reduced form" flavor, and are only defined relative to the specific saturations used in the experiment. While our random coefficients model places more restrictions on the potential outcome functions, it allows us to recover "fully structural" causal effects that are not specific to the design of the experiment.

Our paper also relates to the applied literature that estimates spillover effects in various settings. This includes "partial population" studies in which a subset of subjects in the treatment group are left untreated and their outcomes are compared to those of subjects in a control group (Angelucci and De Giorgi, 2009; Barrera-Osorio et al., 2011; Bobonis and Finan, 2009; Duflo and Saez, 2003; Haushofer and Shapiro, 2016). It also includes cluster-randomized trials where groups are defined by a spatial radius within which social interactions may arise (Bobba and Gignoux, 2014; Miguel and Kremer, 2004) and more recent papers that use a randomized saturation design (Banerjee et al., 2012; Bursztyn et al., 2019; Giné and Mansuri, 2018; Sinclair et al., 2012). In general, this literature estimates intent-to-treat (ITT) effects. Two notable exceptions are Crépon et al. (2013) and Akram et al. (2018) who estimate effects that are similar in spirit to the CADE of Imai et al. (2018). Our identification approach also relates to a large literature on random coefficients models, the closest being Masten and Torgovitsky (2016), as well as methods that identify structural effects using control functions (Altonji and Matzkin, 2005; Imbens and Newey, 2009). In our setting, the share of compliers in a group plays the role of a control function. Our model is also conceptually similar to the functional IV model of Cai et al. (2006).

The remainder of the paper is organized as follows. Section 2 details our notation and assumptions, while Section 3 presents our identification results. Section 3.1 suggests a fea-

sible, kernel-based estimator of the effects identified in Section 3 and Section 4 concludes. Proofs appear in the appendix.

2 Notation and Assumptions

We observe N individuals divided between G groups. We assume throughout the paper that each group has at least two members so there is scope for social interactions. Let $g = 1, \ldots, G$ index groups and $i = 1, \ldots, N_g$ index individuals within a given group g. Using this notation, $N = \sum_g N_g$. For each individual (i, g) we observe a binary treatment offer Z_{ig}, an indicator of treatment take-up D_{ig}, and an outcome Y_{ig}. For each group g we observe a saturation $S_g \in [0, 1]$ that determines the fraction of individuals offered treatment in that group. A bold letter indicates a vector and a g-subscript shows that this vector is restricted to members of a particular group. For example $\boldsymbol{Z}$ is the N-vector of all treatment offers Z_{ig} while $\boldsymbol{Z}_g$ is the N_g-vector obtained by restricting $\boldsymbol{Z}$ to group g. Define $\boldsymbol{D}$ and $\boldsymbol{D}_g$ analogously and let $\boldsymbol{S}$ denote the G-vector of all S_g. At various points in our discussion we will need to refer to the average value of a variable for everyone in a group *besides* person (i, g). As shorthand, we refer to these other individuals as person (i, g)'s *neighbors*. To indicate such an average, we use a bar along with an (i, g) subscript. For instance, $\bar{D}_{ig}$ denotes the treatment take-up rate in group g excluding (i, g), while $\bar{Z}_{ig}$ is the analogous treatment offer rate:

$$\bar{D}_{ig} \equiv \frac{1}{N_g - 1} \sum_{j \neq i} D_{jg}, \quad \bar{Z}_{ig} \equiv \frac{1}{N_g - 1} \sum_{j \neq i} Z_{jg}. \tag{1}$$

Note that, under this definition, $\bar{D}_{ig}$ and $\bar{Z}_{ig}$ vary across individuals in the same group depending on their values of D_{ig} or Z_{ig}. For example in a group of ten people, of whom five take up treatment, $\bar{D}_{ig} = 0.5$ if $D_{ig} = 0$ and 0.4 if $D_{ig} = 1$. We now introduce our basic assumptions, beginning with the experimental design.

Assumption 1 (Assignment of Saturations)**.** *Let $\mathcal{S} = \{s_1, s_2, \ldots, s_J\}$ where $s_j \in [0, 1]$ for all j. Saturations are assigned to groups completely at random from $\mathcal{S}$ such that m_j groups are assigned to saturation s_j with probability one, where $\sum_{j=1}^{J} m_j = G$. In other words,*

$$\mathbb{P}(S_g = s_j) = \begin{cases} m_j/G & \text{for } j = 1, \ldots, J \\ 0 & \text{otherwise} \end{cases}$$

Assumption 1 details the first stage of the randomized saturation design. In this stage, each group g is assigned a saturation S_g drawn completely at random from a set $\mathcal{S}$. In the example from Figure 1, fifty groups (balls) are divided equally between five saturations

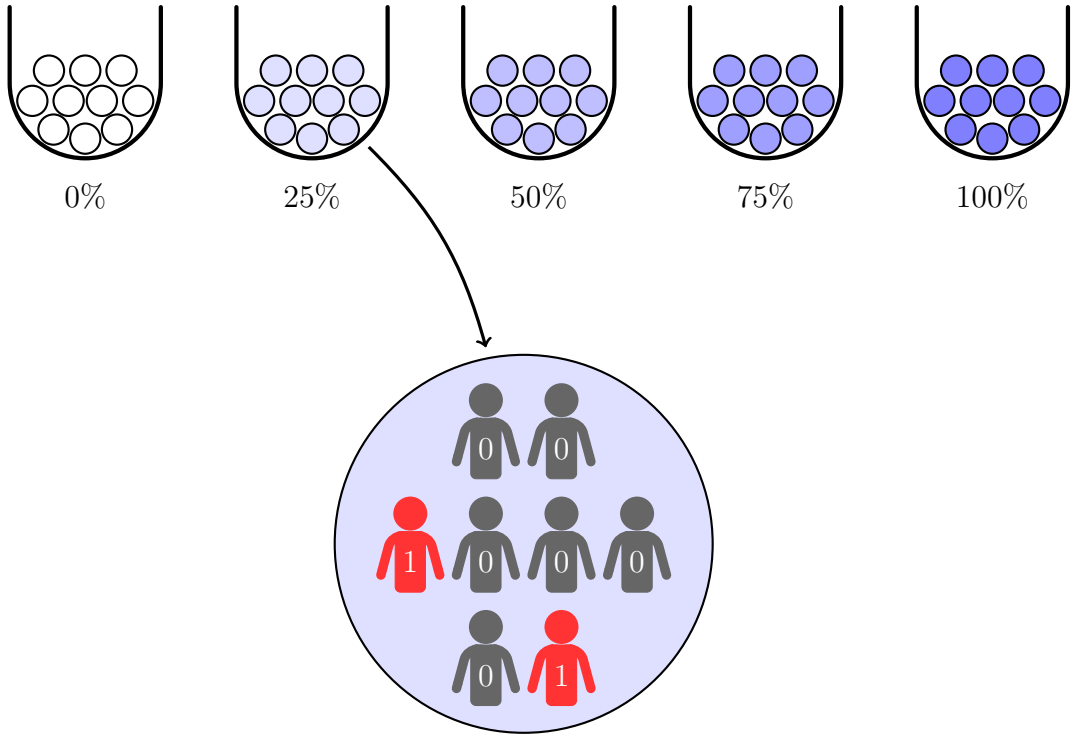

Figure 1: Randomized Saturation Design. In the first stage groups (balls) are randomly assigned to saturations (urns). In the second stage, individuals within a group are randomly assigned treatment offers at the saturation selected in the first stage. The figure zooms in on a group of size eight that has been assigned to a 25% saturation: two individuals are offered treatment.

(urns), namely $\mathcal{S} = \{0, 0.25, 0.5, 0.75, 1\}$. The saturation drawn in this first stage determines the fraction of individuals in the group that will be offered treatment in the second stage. Figure 1, for example, depicts a group of eight individuals that has been assigned to the 25% saturation: two are offered treatment and six are not. These second-stage treatment offers can be made in two possible ways, detailed in the following two assumptions.

Assumption 2 (Bernoulli Offers)**.**

$$\mathbb{P}(\boldsymbol{Z}_g = \boldsymbol{z} | S_g = s, N_g = n) = \prod_{i=1}^{n} s^{z_i}(1-s)^{1-z_i}.$$

Assumption 3 (Completely Randomized Offers)**.**

$$\mathbb{P}\left(\boldsymbol{Z}_g = \boldsymbol{z} | S_g = s, N_g = n\right) = \begin{cases} \binom{n}{\lfloor ns \rfloor}^{-1}, & \text{if } \sum_{i=1}^{n} z_i = \lfloor ns \rfloor \\ 0, & \text{otherwise.} \end{cases}$$

Assumption 2 and Assumption 3 are mutually exclusive ways to assign treatment offers. Under Assumption 2, the assigned saturation s determines the probability that a person in group g will be offered treatment, but the number of treatment offers actually made within the group is random: offers within a group are independent Bernoulli random variables with probability of success s, the realization of S_g. Rather than determining the probability of treatment, under Assumption 3 the assigned saturation s fixes the number of treatment offers made within a group at $\lfloor ns \rfloor$ where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. For example, in a group of 99 people and an assigned saturation of 50% we would make $\lfloor 0.5 \times 99 \rfloor = 49$ treatment offers. While Bernoulli treatment offers are statistically independent, completely randomized offers exhibit dependence within a given group. Because the total number of offers is fixed, the fact that Alice was offered makes it slightly less likely that Bob will be.

The randomized saturation design creates exogenous variation at the individual and group levels. Within a group some individuals are offered while others are not. Between groups, some have a large number of individuals offered treatment—a high saturation—while others do not. Many randomized saturation experiments, like the illustration in Figure 1, feature a 0% saturation or even a 100% saturation. We refer to 0% and 100% saturations collectively as *corner saturations* to distinguish them from all other saturations, which we call *interior*. There is no variation in treatment offers between individuals in a group assigned a corner saturation. For this reason, as we discuss in section 3 below, the number of interior saturations in the design will determine the flexibility with which we can model potential outcome functions.

Assumptions 1–3 concern the design of the experiment. Our remaining assumptions, in contrast, concern the potential outcome and treatment functions. Without imposing any restrictions, an individual's potential outcome function $Y_{ig}(\cdot)$ could in principle depend on the treatment take-up of all individuals in the sample. We denote this unrestricted potential outcome function by $Y_{ig}(\boldsymbol{D})$. Assumption 4 restricts $Y_{ig}(\cdot)$ to depend only on D_{ig} and $\bar{D}_{ig}$ via a random coefficients model.

Assumption 4 (Random Coefficients Model)**.** *Let* $\mathbf{f}(\cdot)$ *be a* K*-vector of known functions* $f_k \colon [0,1] \mapsto \mathbb{R}$*, each of which satisfies* $\sup_{x \in [0,1]} |f_k(x)| < \infty$*. We assume that*

$$Y_{ig}(\boldsymbol{D}) = Y_{ig}(\boldsymbol{D}_g) = Y_{ig}(D_{ig}, \bar{D}_{ig}) = \mathbf{f}(\bar{D}_{ig})' \left[(1 - D_{ig}) \boldsymbol{\theta}_{ig} + D_{ig} \boldsymbol{\psi}_{ig} \right]$$

where $\boldsymbol{\theta}_{ig}$ *and* $\boldsymbol{\psi}_{ig}$ *are* K*-dimensional random vectors that may be dependent on* $(D_{ig}, \bar{D}_{ig})$*.*

The first equality in Assumption 4 is the so-called *partial interference* assumption, used widely in the literature on spillover effects. This assumption states that there are no social in-

teractions between individuals in different groups: only the treatment take-up of individuals in group g affects the potential outcome of person (i, g). The second equality in Assumption 4 states that person (i, g)'s potential outcome is only affected by the treatment take-up of the others in her group through the *aggregate* $\bar{D}_{ig}$.[3] This is related to the *anonymous interactions* assumption from the network literature as it implies that only the number of (i, g)'s neighbors who take up treatment matters for her outcome; the identities of the neighbors are irrelevant (Manski, 2013). The third equality in Assumption 4 posits a finite basis function expansion for the potential outcome functions $Y_{ig}(0, \bar{D}_{ig})$ and $Y_{ig}(1, \bar{D}_{ig})$, namely

$$Y_{ig}(0, \bar{D}_{ig}) = \sum_{k=1}^{K} \theta_{ig}^{(k)} f_k(\bar{D}_{ig}), \quad Y_{ig}(1, \bar{D}_{ig}) = \sum_{k=1}^{K} \psi_{ig}^{(k)} f_k(\bar{D}_{ig})$$

or, written more compactly in matrix form,

$$Y_{ig} = \mathbf{X}_{ig}' \mathbf{B}_{ig}, \quad \mathbf{X}_{ig} \equiv \begin{bmatrix} 1 \\ D_{ig} \end{bmatrix} \otimes \mathbf{f}(\bar{D}_{ig}), \quad \mathbf{B}_{ig} \equiv \begin{bmatrix} \boldsymbol{\theta}_{ig} \\ \boldsymbol{\psi}_{ig} - \boldsymbol{\theta}_{ig} \end{bmatrix} \tag{2}$$

where the coefficient vectors $\boldsymbol{\theta}_{ig}$ and $\boldsymbol{\psi}_{ig}$, and hence $\mathbf{B}_{ig}$, are allowed to vary arbitrarily across groups and individuals. If, for example, person (i, g) has some prior knowledge of her potential outcome function $Y_{ig}(\cdot, \cdot)$, her take-up decision may depend on $\boldsymbol{\theta}_{ig}$ and $\boldsymbol{\psi}_{ig}$. More generally, the same unobserved characteristics that determine a person's decision to take up treatment could affect her potential outcomes. To account for these possibilities, we allow arbitrary statistical dependence between $(D_{ig}, \bar{D}_{ig})$ and $\mathbf{B}_{ig}$.

Ideally, our goal would be to identify the average direct and indirect causal effects of the binary treatment D_{ig}. Under Assumption 4, we define these as follows, building on the definitions of Hudgens and Halloran (2008). The direct treatment effect, DE, gives the average effect of exogenously changing an individual's own treatment D_{ig} from 0 to 1 while holding the share of her treated neighbors $\bar{D}_{ig}$ fixed at $\bar{d}$, namely

$$\text{DE}(\bar{d}) \equiv \mathbb{E}\left[Y_{ig}(1, \bar{d}) - Y_{ig}(0, \bar{d})\right] = \mathbf{f}(\bar{d})' \mathbb{E}\left[\boldsymbol{\psi}_{ig} - \boldsymbol{\theta}_{ig}\right] \tag{3}$$

where the expectations are taken over all individuals in the population from which our experimental subjects were drawn. Note that we define $\bar{D}_{ig}$ to exclude person (i, g), ensuring that $\text{DE}(\bar{d})$ is well-defined. An indirect treatment effect, in contrast, gives the average effect of exogenously increasing a person's share of treated neighbors $\bar{D}_{ig}$ from $\bar{d}$ to $\bar{d} + \Delta$ while

[3] Recall that $\bar{D}_{ig}$ is defined to exclude person (i, g).

holding her own treatment D_{ig} fixed at d, in other words

$$\begin{aligned} \text{IE}_d(\bar{d}, \Delta) &\equiv \mathbb{E}\left[Y_{ig}(d, \bar{d}+\Delta) - Y_{ig}(d, \bar{d})\right] \\ &= \left[\mathbf{f}(\bar{d}+\Delta) - \mathbf{f}(\bar{d})\right]' \left\{(1-d)\mathbb{E}\left[\boldsymbol{\theta}_{ig}\right] + d\mathbb{E}\left[\boldsymbol{\psi}_{ig}\right]\right\} \end{aligned} \tag{4}$$

where Δ is a positive increment. There are two indirect treatment effect functions, IE_0 and IE_1, corresponding to the two possible values at which we could hold D_{ig} fixed: a spillover on the untreated, and a spillover on the treated. Because the direct and indirect causal effects are fully determined by $\mathbb{E}[\mathbf{B}_{ig}]$ under Assumption 4, this is our object of interest below.

Under perfect compliance D_{ig} would simply equal Z_{ig}, making both D_{ig} and $\bar{D}_{ig}$ exogenous. In this case a sample analogue of $\mathbb{E}[Y_{ig}(d, \bar{d})]$ could be used to recover all of the treatment effects discussed above, at least at values of $\bar{d}$ that arise in the experimental design. Unfortunately non-compliance is pervasive in real-world experiments, greatly complicating the identification of causal effects. In a large-scale experiment carried out in France, for example, only 35% of unemployed workers offered job placement services took them up (Crépon et al., 2013). Those who did take up treatment likely differ in myriad ways from those who did not: they may, for example, be more conscientious. One way to to avoid this problem of self-selection is to carry out an intent-to-treat (ITT) analysis, conditioning on Z_{ig} and S_g rather than D_{ig} and $\bar{D}_{ig}$. But with take-up rates as low as 35%, ITT estimates could be very far from the causal effects of interest. In this paper we adopt a different approach. Following the tradition in the local average treatment effect (LATE) literature, we provide conditions under which direct and indirect causal effects–rather than ITT effects–can be identified for well-defined sub-populations of individuals. We focus on the case of *one-sided noncompliance*, in which only those offered treatment can take it up. One-sided non-compliance is fairly common in practice (e.g. Crépon et al., 2013) and simplifies the analysis considerably.[4]

Assumption 5 (One-sided Non-compliance)**.** *If* $Z_{ig} = 0$ *then* $D_{ig} = 0$.

To account for endogenous treatment take-up, we define potential treatment functions $D_{ig}(\cdot)$. In principle these could depend on the treatment offers of every individual, $\boldsymbol{Z}$ in the experiment. The following assumption restricts $D_{ig}(\cdot)$ to permit identification of the direct and indirect causal effects described above.

Assumption 6 (IOR)**.** $D_{ig}(\boldsymbol{Z}) = D_{ig}(\boldsymbol{Z}_g) = D_{ig}(Z_{ig}, \bar{Z}_{ig}) = D_{ig}(Z_{ig})$.

The first equality of Assumption 6 is a partial interference assumption: it requires that there are no social interactions in *take-up* between individuals in different groups. The

[4] An extension of our results to two-sided non-compliance is currently in progress.

second equality of Assumption 6 states that person (i, g)'s take-up decision depends on the treatment offers of others in her group only through the fraction $\bar{Z}_{ig}$ of treatment offers made to the others in her group.[5] Unfortunately these first two equalities are not in general sufficient to point identify direct and indirect causal effects. The third equality, which we call *individualistic offer response* or IOR for short, imposes the further restriction that each person's take-up decision depends only on her own treatment offer. IOR states that there are no social interactions in *take-up*.[6] This is a strong assumption, but one that has also appeared in the existing literature. Kang and Imbens (2016), for example, employ a variant of IOR that they called "personalized encouragement." And while Imai et al. (2018) derive their so-called "complier average direct effect (CADE)" under a weaker condition than IOR, the CADE is in fact a hybrid of direct and indirect effects unless one is willing to assume that there are no social interactions in take-up. Fortunately, IOR is testable: it implies, for example, that $\mathbb{E}[D_{ig}|Z_{ig} = 1, S_g = s]$ does not vary with s. If the observed average take-up rate among individuals who are offered treatment varies with saturation, this indicates a violation of IOR.

Under IOR and one-sided non-compliance (Assumptions 5 and 6), we can divide individuals into never-takers and compliers, two of the principal strata from the LATE literature. Never-takers are defined as those for whom $D_{ig}(0) = D_{ig}(1) = 0$, while compliers are those for whom $D_{ig}(z) = z$ for all z.[7] Defining C_{ig} to be the indicator that person (i, g) is a complier, Assumptions 5–6 imply that

$$D_{ig} = C_{ig} Z_{ig}, \quad \bar{D}_{ig} = \frac{1}{N_g - 1} \sum_{j \neq i} C_{jg} Z_{jg}. \tag{5}$$

By analogy to $\bar{Z}_{ig}$ and $\bar{D}_{ig}$, we define $\bar{C}_{ig}$ to be the share of compliers among person (i, g)'s neighbors in group g, namely

$$\bar{C}_{ig} = \frac{1}{N_g - 1} \sum_{j \neq i} C_{jg}. \tag{6}$$

Note that $\bar{C}_{ig}$ varies across individuals in the same group, depending on their values of C_{ig}. Finally, let $\boldsymbol{C}_g$ denote the vector of C_{ig} for all individuals in group g.

Our final assumption is an exclusion restriction for the treatment offers $\boldsymbol{Z}_g$ and saturation S_g. To state it we require two additional pieces of notation. First, let $\mathbf{B}_g$ denote the vector that stacks $\mathbf{B}_{ig}$ for all individuals in group g. Second, following Dawid (1979), let "$\perp\!\!\!\perp$" denote

[5] Recall that the average $\bar{Z}_{ig}$ is defined to exclude (i, g).

[6] Work in progress explores the possibility of relaxing IOR in specific settings to obtain point, or at least partial identification.

[7] Under one-sided non-compliance, Assumption 5, there are no always-takers.

(conditional) independence so that $X \perp\!\!\!\perp Y$ indicates that X is statistically independent of Y while $X \perp\!\!\!\perp Y|Z$ indicates that X is *conditionally* independent of Y given Z. Using this notation, the exclusion restriction is as follows.

Assumption 7 (Exclusion Restriction)**.**

(i) $S_g \perp\!\!\!\perp (\boldsymbol{C}_g, \mathbf{B}_g, N_g)$

(ii) $\boldsymbol{Z}_g \perp\!\!\!\perp (\boldsymbol{C}_g, \mathbf{B}_g)|(S_g, N_g)$

Intuitively, Assumption 7 states that $(\boldsymbol{C}_g, \mathbf{B}_g, N_g)$ are "predetermined" with respect to the treatment offers and saturations. In a traditional LATE setting, the counterparts of Assumption 7 are the "unconfounded type" assumption and the independence of potential outcomes and treatment offers. Assumption 7 could be violated in a number of ways. If, for example, individuals chose their group membership based on knowledge of their group's saturation, N_g would not be independent of S_g. Similarly, if some individuals decided to comply with their treatment offers only because their group was assigned a high saturation, $\boldsymbol{C}_g$ would not be independent of S_g. This latter possibility illustrates that Assumption 7 partially embeds IOR by ruling out "selection into compliance." More prosaically, Assumption 7 would be violated if either S_g or Z_{ig} had a direct effect on the random coefficients $\mathbf{B}_g$. Notice that part (ii) of Assumption 7 conditions on (S_g, N_g). This is because the second stage of the randomized saturation experiment assigns $\boldsymbol{Z}_g$ conditional on this information: see Assumption 2 and Assumption 3.

3 Identification

Given that the functional form of the random coefficients model in Assumption 4 is known, one might be tempted to suppose that a simple IV estimation strategy using functions of (Z_{ig}, S_g) as instruments for $\left(D_{ig}, \mathbf{f}(\bar{D}_{ig})\right)$ could recover $\mathbb{E}[\boldsymbol{\theta}_{ig}]$ and $\mathbb{E}[\boldsymbol{\psi}_{ig} - \boldsymbol{\theta}_{ig}]$. Indeed, two-stage least squares identifies the average effects in a random coefficients model provided that the first stage relationship between instruments and endogenous regressors is homogeneous (Heckman and Vytlacil, 1998; Wooldridge, 1997, 2003, 2016). Unfortunately the first stage in our model is *heterogenous* so this result does not apply: the conditional distribution of $\bar{D}_{ig}$ given S_g varies with $(\bar{C}_{ig}, N_g)$, as shown in the following lemma.

Lemma 1. *Let $\bar{c}$ be a value in the interval $[0, 1]$ such that $(n-1)\bar{c}$ is a positive integer. Now define X as the sum of $(n-1)\bar{c}$ independent Bernoulli trials, each with probability of success equal to s, so that $X \sim$ Binomial$((n-1)\bar{c}, s)$. Further let Y be the number of successes*

obtained when $\lfloor ns \rfloor - z$ *draws are made without replacement from a population of size* $(n-1)$ *containing* $(n-1)\bar{c}$ *successes, so that* $Y \sim$ *Hypergeometric* $(n-1, (n-1)\bar{c}, \lfloor ns \rfloor - z)$*. Then, conditional on* $(N_g = n, S_g = s, \boldsymbol{C}_g = \boldsymbol{c}, \bar{C}_{ig} = \bar{c}, Z_{ig} = z)$ *and under Assumptions 1 and 5–7,*

(i) $\bar{D}_{ig} =^d X/(n-1)$ *under Bernoulli offers (Assumption 2), and*

(ii) $\bar{D}_{ig} =^d Y/(n-1)$ *under completely randomized offers (Assumption 3)*

where $=^d$ *denotes equality in distribution.*

Intuitively, the problem presented by the Lemma 1 is as follows. Although S_g is randomly assigned, the variation that it induces in $\bar{D}_{ig}$ is mediated through the share of compliers $\bar{C}_{ig}$. Accordingly if $\bar{C}_{ig}$—a source of first-stage heterogeneity—is correlated with the random coefficients in the second stage, the IV estimator will be inconsistent. Consider, for example, a simple linear model for the potential outcomes:

$$Y_{ig} = \alpha_{ig} + \beta_{ig} D_{ig} + \gamma_{ig} \bar{D}_{ig} + \delta_{ig} D_{ig} \bar{D}_{ig}.$$

If we restrict attention to individuals who are not offered treatment, $Z_{ig} = 0$, the model becomes $Y_{ig} = \alpha_{ig} + \gamma_{ig} \bar{D}_{ig}$. Note that there is no sample selection bias because Z_{ig} is randomly assigned. Defining $\alpha \equiv \mathbb{E}(\alpha_{ig})$ and $\gamma \equiv \mathbb{E}(\gamma_{ig})$, we can re-write the model as

$$Y_{ig} = \alpha + \gamma \bar{D}_{ig} + \varepsilon_{ig}, \quad \varepsilon_{ig} \equiv (\alpha_{ig} - \alpha) + (\gamma_{ig} - \gamma)\bar{D}_{ig}$$

so that the IV estimand for γ, using S_g to instrument for $\bar{D}_{ig}$, is given by

$$\gamma_{IV} \equiv \frac{\mathrm{Cov}(Y_{ig}, S_g)}{\mathrm{Cov}(D_{ig}, S_g)} = \gamma + \frac{\mathrm{Cov}(\varepsilon_{ig}, S_g)}{\mathrm{Cov}(S_g, \bar{D}_{ig})} = \gamma + \frac{\mathrm{Cov}\left[(\gamma_{ig} - \gamma)\bar{D}_{ig}, S_g\right]}{\mathrm{Cov}(\bar{D}_{ig}, S_g)} \tag{7}$$

since α is constant and α_{ig} is uncorrelated with S_g by Assumption 7 (i). To express the bias term from (7) in a more transparent form, we rely on the following lemma.

Lemma 2. *Let* β_{ig} *be an element of* $\mathbf{B}_g$ *and* $\beta \equiv \mathbb{E}(\beta_{ig})$*. Under Assumptions 1–2 and 5–7,*

(i) $Cov(S_g, \bar{D}_{ig}) = Var(S_g)\mathbb{E}(\bar{C}_{ig})$ *and*

(ii) $Cov\left[(\beta_{ig} - \beta)\bar{D}_{ig}, S_g\right] = Cov(\beta_{ig}, \bar{C}_{ig})\, Var(S_g)$.

Applying Lemma 2 to (7) gives $(\gamma_{IV} - \gamma) = \mathrm{Cov}(\gamma_{ig}, \bar{C}_{ig})/\mathbb{E}(\bar{C}_{ig})$ under Bernoulli offers.[8] Hence, IV fails to identify the average spillover effect unless the individual-specific spillover

[8] An analogous but slightly more complicated expression holds for completely randomized offers.

effects are uncorrelated with the share of compliers. This condition could easily fail in practice. In the labor market example from the introduction, cities with a particularly depressed labor market might be expected to contain a large share of compliers. If negative employment spillovers are more intense in such cities, IV will not recover the true indirect effect. To address this challenge, our identification results rely on a widely-used strategy from the literature on random coefficients models. Aptly dubbed "localize then average" by Hoderlein and Sherman (2015), this approach proceeds in two steps. First we localize. Conditional on $(\bar{C}_{ig} = \bar{c}, N_g = n)$ we show that certain population IV regressions identify expectations of $\mathbf{B}_{ig}$ for a particular type of individual: someone in a group of size n who has $\bar{c}(n-1)$ complier neighbors. Next we average, obtaining an overall causal effect by integrating out $(\bar{C}_{ig}, N_g)$. While N_g is observed, $\bar{C}_{ig}$ is not. If N_g is large, however, $\bar{C}_{ig}$ is well-approximated by $\bar{D}_{ig}/\bar{Z}_{ig}$. In subsection 3.1 we use this idea to provide feasible estimators that are consistent when both the size and number of groups grow. For compliers we recover the direct causal effect along with both indirect effects, IE_0 and IE_1; for never-takers, and the population as a whole, we obtain the indirect effect when untreated, IE_0.[9] Our localize-then average approach relies on the following conditional independence result.

Theorem 1. *Suppose that either Assumption 2 or Assumption 3 holds. Then, under Assumptions 1 and 4–7,* $(Z_{ig}, \bar{D}_{ig}, S_g) \perp\!\!\!\perp (\mathbf{B}_{ig}, C_{ig}) | (\bar{C}_{ig}, N_g)$.

Theorem 1 implies that conditioning on $(\bar{C}_{ig}, N_g)$ is sufficient to break the dependence between $\mathbf{f}(\bar{D}_{ig})$ and $(\mathbf{B}_{ig}, C_{ig})$. The intuition for this result is as follows. Conditional on $\bar{C}_{ig}$ and N_g, we know precisely how many of (i,g)'s neighbors are compliers. Given this information, IOR implies that all remaining variation in $\bar{D}_{ig}$ is arises solely from experimental variation in the saturation S_g assigned to different groups, and the share of compliers offered treatment across groups assigned the same saturation. So long as Z_{ig} and S_g do not affect $(\mathbf{B}_{ig}, C_g)$, Assumption 7, it follows that $(Z_{ig}, \bar{D}_{ig}, S_g)$ are exogenous given $(\bar{C}_{ig}, N_g)$, even when individuals decide whether or not to take up treatment based on knowledge of their potential outcome functions. Theorem 1 allows us to identify "localized" treatment effects via the following result.

Theorem 2. *Let* $\widetilde{\mathbf{X}}_{ig} \equiv \mathbf{f}(\bar{D}_{ig})$ *and* $\mathbf{W}_{ig} \equiv (1, Z_{ig})' \otimes \widetilde{\mathbf{X}}_{ig}$. *Suppose that either Assumption 2 or Assumption 3 holds. Then, under Assumptions 1 and 4–7, and whenever the respective inverses exist,*

$$(i) \quad \begin{bmatrix} \mathbb{E}(\boldsymbol{\theta}_{ig} | \bar{C}_{ig}, N_g) \\ \mathbb{E}(\boldsymbol{\psi}_{ig} - \boldsymbol{\theta}_{ig} | C_{ig} = 1, \bar{C}_{ig}, N_g) \end{bmatrix} = \mathbb{E}\left[\mathbf{W}_{ig}\mathbf{X}_{ig}' | \bar{C}_{ig}, N_g\right]^{-1} \mathbb{E}\left[\mathbf{W}_{ig} Y_{ig} | \bar{C}_{ig}, N_g\right],$$

[9] For definitions of the direct and indirect effects in terms of our random coefficient model see (4) and (3).

(ii) $\mathbb{E}[\boldsymbol{\psi}_{ig}|C_{ig}=1,\bar{C}_{ig},N_g]=\mathbb{E}[D_{ig}Z_{ig}\widetilde{\mathbf{X}}_{ig}\widetilde{\mathbf{X}}_{ig}'|\bar{C}_{ig},N_g]^{-1}\mathbb{E}\left[D_{ig}Z_{ig}\widetilde{\mathbf{X}}_{ig}Y_{ig}|\bar{C}_{ig},N_g\right]$, *and*

(iii) $\mathbb{E}\left[\boldsymbol{\theta}_{ig}|C_{ig}=0,\bar{C}_{ig},N_g\right]=\mathbb{E}[(1-D_{ig})Z_{ig}\widetilde{\mathbf{X}}_{ig}\widetilde{\mathbf{X}}_{ig}'|\bar{C}_{ig},N_g]^{-1}\mathbb{E}[(1-D_{ig})Z_{ig}\widetilde{\mathbf{X}}_{ig}Y_{ig}|\bar{C}_{ig},N_g]$

(iv) $\mathbb{E}\left[\boldsymbol{\theta}_{ig}|\bar{C}_{ig},N_g\right]=\mathbb{E}\left[(1-Z_{ig})\widetilde{\mathbf{X}}_{ig}'\widetilde{\mathbf{X}}_{ig}'|\bar{C}_{ig},N_g\right]^{-1}\mathbb{E}\left[(1-Z_{ig})\widetilde{\mathbf{X}}_{ig}Y_{ig}|\bar{C}_{ig},N_g\right]$

where $\mathbf{X}_{ig}$ *is as defined in* (2).

Theorem 2 is an identification result for four sets of just-identified conditional IV moment equations. Each set restricts attention to a particular kind of person: if $(\bar{C}_{ig}=\bar{c},N_g=n)$, we only average over individuals in groups of size n who have $\bar{c}(n-1)$ compliers among their neighbors. Subject to this restriction, part (ii) of Theorem 2 and the lower block of part (i) further restrict attention to individuals who are *themselves* compliers. Part (iii), in contrast, considers only *never-takers* with $(\bar{C}_{ig}=\bar{c},N_g=n)$ while part (iv) and the upper block of part (i) impose no further restrictions beyond $(\bar{C}_{ig}=\bar{c},N_g=n)$. Comparing the left-hand side of each of the four equalities with (4) and (3), we see that Theorem 2 identifies conditional direct and indirect causal effects, provided that the requisite inverses exist. The bottom block of part (i) identifies the direct effect DE for compliers with $(\bar{C}_{ig}=\bar{c},N_g=n)$, while part (ii) identifies the indirect effect IE_1 for treated compliers with $(\bar{C}_{ig}=\bar{c},N_g=n)$. Combining the bottom block of (i) with (ii) allows us to solve for $\mathbb{E}\left[\boldsymbol{\theta}_{ig}|C_{ig}=1,\bar{C}_{ig},N_g\right]$, identifying the other indirect effect, IE_0, for compliers with $(\bar{C}_{ig}=\bar{c},N_g=n)$. Thus, we recover "localized" versions of all causal effects for compliers. Because, by definition, they never take up treatment, we cannot identify DE or IE_1 for individuals with $C_{ig}=0$. Part (iii) of Theorem 2, however, allows us to identify IE_0 for never-takers with $(\bar{C}_{ig}=\bar{c},N_g=n)$ while part (iv) identifies the same indirect effect for all individuals with $(\bar{C}_{ig}=\bar{c},N_g=n)$. Notice that, while we use Z_{ig} to instrument for D_{ig}, $\mathbf{f}(\bar{D}_{ig})$ serves as its *own instrument* in the moment equations from Theorem 2. As explained above, $\bar{D}_{ig}$ is exogenous *conditional* on $(\bar{C}_{ig},N_g)$, and is hence its "own best instrument."

The equalities in Theorem 2 are only valid, and hence the relevant effects are only identified, when the appropriate matrix inverses exist. The following lemma provides high-level necessary and sufficient conditions under Bernoulli offers (Assumption 2). Analogous results for completely randomized offers appear in the Appendix.

Lemma 3. *For* $z=0,1$, *define* $\mathbf{Q}_z(\bar{c},n)\equiv\mathbb{E}[\widetilde{\mathbf{X}}_{ig}\widetilde{\mathbf{X}}_{ig}'|Z_{ig}=z,\bar{C}_{ig}=\bar{c},N_g=n]$ *where* $\widetilde{X}_{ig}\equiv\mathbf{f}(\bar{D}_{ig})$. *Then, under Assumptions 1, 2, and 4–7,*

(i) $\mathbb{E}[\mathbf{W}_{ig}\mathbf{X}_{ig}'|\bar{C}_{ig}=\bar{c},N_g=n]$ *is invertible* $\iff$ $\mathbb{E}\left(C_{ig}|\bar{C}_{ig}=\bar{c},N_g=n\right)\neq 0$, $\mathbb{E}(S_g)\notin\{0,1\}$, *and both* $\mathbf{Q}_0(\bar{c},n)$ *and* $\mathbf{Q}_1(\bar{c},n)$ *are invertible;*

(ii) $\mathbb{E}[D_{ig}Z_{ig}\widetilde{\mathbf{X}}_{ig}\widetilde{\mathbf{X}}'_{ig}|\bar{C}_{ig} = \bar{c}, N_g = n]$ *is invertible* $\iff$ $\mathbb{E}(C_{ig}|\bar{C}_{ig} = \bar{c}, N_g = n) \neq 0$, $\mathbb{E}(S_g) \neq 1$, *and* $\mathbf{Q}_1(\bar{c}, n)$ *is invertible;*

(iii) $\mathbb{E}[(1 - D_{ig})Z_{ig}\widetilde{\mathbf{X}}_{ig}\widetilde{\mathbf{X}}'_{ig}|\bar{C}_{ig} = \bar{c}, N_g = n]$ *is invertible* $\iff$ $\mathbb{E}(C_{ig}|\bar{C}_{ig} = \bar{c}, N_g = n) \neq 1$, $\mathbb{E}(S_g) \neq 0$, *and* $\mathbf{Q}_1(\bar{c}, n)$ *is invertible; and*

(iv) $\mathbb{E}[(1 - Z_{ig})\widetilde{\mathbf{X}}_{ig}\widetilde{\mathbf{X}}'_{ig}|\bar{C}_{ig} = \bar{c}, N_g = n]$ *is invertible* $\iff$ $\mathbb{E}(S_g) \neq 1$ *and* $\mathbf{Q}_0(\bar{c}, n)$ *is invertible.*

The four parts of Lemma 3 correspond to those of Theorem 2. Parts (i) and (ii) concern conditional effects for compliers. Unsurprisingly, these effects are only identified at $(\bar{C}_{ig} = \bar{c}, N_g = n)$ if there *are in fact* compliers in the population who reside in a group of size n and have $(n-1)\bar{c}$ compliers among their neighbors. This is the case precisely when $\mathbb{E}(C_{ig}|\bar{C}_{ig} = \bar{c}, N_g = n) \neq 0$. Similarly, part (iii) concerns a conditional effect for never-takers. This effect is only identified at $(\bar{c}, n)$ if there are in fact never-takers in the population who reside in a group of size n and have $(n-1)\bar{c}$ compliers among their neighbors, i.e. if $\mathbb{E}(C_{ig}|\bar{C}_{ig} = \bar{c}, N_g = n) \neq 1$. Because part (iv) identifies a conditional effect for the whole population, it does not require a restriction on $\mathbb{E}(C_{ig}|\bar{C}_{ig}, N_g)$. The restrictions placed on $\mathbb{E}(S_g)$ in Lemma 3 are weak: we need only impose that the average saturation is neither zero or one. This condition merely ensures that some individuals are offered treatment and others are not: without it, there would be no variation in Z_{ig} across the experiment.

The second step of our localize-then-average approach integrates the conditional causal effects identified by Theorem 2 over the distribution of $(\bar{C}_{ig}, N_g)$ to yield LATE-type direct and indirect causal effects.

Theorem 3. *Let* $\mathcal{Q}_z$, $z = 0, 1$, *denote the set of all* $(\bar{c}, n)$ *in the support of* $(\bar{C}_{ig}, N_g)$ *such that* $\mathbf{Q}_z(\bar{c}, n)$ *is invertible and define the shorthand* $H_{ig} \equiv (\bar{C}_{ig}, N_g)$. *Then, so long as* $\mathbb{E}(S_g) \neq 0, 1$ *following effects are point identified under Assumptions 1, 2, and 4–7:*

(i) $\mathbb{E}[\boldsymbol{\psi}_{ig} - \boldsymbol{\theta}_{ig}|C_{ig} = 1, H_{ig} \in \{\mathcal{Q}_0 \cap \mathcal{Q}_1\}]$

(ii) $\mathbb{E}[\boldsymbol{\psi}_{ig}|C_{ig} = 1, H_{ig} \in \mathcal{Q}_1]$

(iii) $\mathbb{E}[\boldsymbol{\theta}_{ig}|C_{ig} = 0, H_{ig} \in \mathcal{Q}_0]$

(iv) $\mathbb{E}[\boldsymbol{\theta}_{ig}|H_{ig} \in \mathcal{Q}_0]$

Theorem 3 uses iterated expectations to average the "localized" causal effects from Theorem 2 over the appropriate conditional distributions of $(\bar{C}_{ig}, N_g)$. This yields the average direct and indirect effects for compliers, parts (i) and (ii), the indirect effect for untreated

never-takers, part (iii), and the indirect effect for the untreated, part (iv). Notice that the conditions $\mathbb{E}(C_{ig}|\bar{C}_{ig} = \bar{c}, N_g = n) \neq 0, 1$ from parts (i)–(iii) of Lemma 3 are not among the assumptions required for Theorem 3. Any points $(\bar{c}, n)$ for which $\mathbb{E}(C_{ig}|\bar{C}_{ig} = \bar{c}, N_g = n) = 0$ by definition contains *no compliers*. Accordingly we need not include this point when computing average effects for compliers in parts (i) and (ii), so the failure of invertibility from Lemma 3 is irrelevant. Analogously, if $\mathbb{E}(C_{ig}|\bar{C}_{ig} = \bar{c}, N_g = n) = 1$, then we need not include $(\bar{c}, n)$ when computing average effects for never-takers in part (iii).

Note that the four parts of Theorem 3 average over different conditional distributions for $(\bar{C}_{ig}, N_g)$. For simplicity, suppose for the moment that $\mathbf{Q}_z(\bar{c}, n)$ is invertible over the full support of $(\bar{C}_{ig}, N_g)$. Then, part (iv) of Theorem 3 simply averages over the unconditional distribution of $(\bar{C}_{ig}, N_g)$. In contrast, parts (i)–(iii) average over the conditional distribution of $(\bar{C}_{ig}, N_g)|C_{ig}$ where C_{ig} indicates the type of person—complier or never-taker—for whom the effect is identified. By Bayes' rule, we can re-write this in terms of the unconditional distribution of $(\bar{C}_{ig}, N_g)$ as

$$\mathbb{P}(\bar{C}_{ig}, N_g|C_{ig}) = \left[\frac{\mathbb{P}(C_{ig}|\bar{C}_{ig}, N_g)}{\mathbb{P}(C_{ig})}\right] \mathbb{P}(\bar{C}_{ig}, N_g)$$

where $\mathbb{P}(C_{ig}|\bar{C}_{ig}, N_g)/\mathbb{P}(C_{ig})$ is an *importance weight*, capturing the share of compliers (or never-takers) among individuals with $(\bar{C}_{ig} = \bar{c}, N_g = n)$ relative to the population as a whole. In parts (i) and (ii) of Theorem 3, the importance weight is $\mathbb{E}(C_{ig}|\bar{C}_{ig}, N_g)/\mathbb{E}(C_{ig})$. Similarly, in part (iii) the importance weight is given by $\mathbb{E}(1 - C_{ig}|\bar{C}_{ig}, N_g)/\mathbb{E}(1 - C_{ig})$.

The key question for interpreting Theorem 3, and indeed the real substance of Lemma 3, concerns the invertibility of $\mathbf{Q}_0(\bar{c}, n)$ and $\mathbf{Q}_1(\bar{c}, n)$. Because $\widetilde{X}_{ig} \equiv \mathbf{f}(\bar{D}_{ig})$, it follows that the invertibility of $\mathbf{Q}_0$ and $\mathbf{Q}_1$ depends only on $\mathbf{f}$ and the distribution of $\bar{D}_{ig}|(\bar{C}_{ig}, N_g, Z_{ig})$. Under the conditions of Theorem 2, this distribution is *known*: it depends solely on the experimental design, as shown in Lemma 1. As such, one can always calculate the rank of $\mathbf{Q}_0$ and $\mathbf{Q}_1$ in a particular application. Indeed, one can *design* the experiment to ensure that these matrices are invertible. The following lemma establishes an important guiding principle of the design: the experiment should contain at least K saturations.

Lemma 4. *Let K be the dimension of $\mathbf{f}$ and J be the number of saturations, and define $\mathbf{Q}_z^*(\bar{c}, n) \equiv \mathbf{Q}_z\left(\lfloor (n-1)\bar{c} \rfloor/(n-1), n\right)$. Under the conditions of Theorem 2 and for any $\bar{c} \in [0, 1]$, $K > J$ implies*

$$\lim_{n\to\infty} \det[\mathbf{Q}_0^*(\bar{c}, n)] = \lim_{n\to\infty} \det[\mathbf{Q}_1^*(\bar{c}, n)] = 0.$$

Lemma 4 shows that $\mathbf{Q}_0(\bar{c}, n)$ and $\mathbf{Q}_1(\bar{c}, n)$ will be *nearly* singular when n is large unless

there are fewer basis functions in $\mathbf{f}$ than there are saturations in the experimental design. Because our estimators in subsection 3.1 rely on the assumption of large groups, it is particularly important that the experiment be designed, or $\mathbf{f}$ be chosen, to avoid this potential weak identification problem. We introduce the function $\mathbf{Q}_z^*(\bar{c}, n)$ merely as a notational device to allow us to take limits as $n \to \infty$ while holding $\bar{c}$ fixed: $\mathbf{Q}_z(\bar{c}, n)$ is only well-defined if $(n-1)\bar{c}$ is a positive integer whereas $\mathbf{Q}_z^*(\bar{c}, n)$ is defined for any $\bar{c} \in [0, 1]$ and $n \in \mathbb{N}$.[10]

To understand the relationship between Lemma 3 and Lemma 4, consider a simple example with Bernoulli offers and linear basis functions: $\mathbf{f}(x)' = (1, x)$. Conditional on $(\bar{C}_{ig} = \bar{c}, N_g = n, Z_{ig} = z, S_g = s)$, Lemma 1 establishes that $\bar{C}_{ig}$ is equal in distribution to a Binomial$((n-1)\bar{c}, s)$ random variable divided by $(n-1)$. Hence,

$$\mathbf{Q}_z(\bar{c}, n) = \sum_{j=1}^{J} \begin{bmatrix} 1 & \bar{c}s_j \\ \bar{c}s_j & (\bar{c}s_j)^2 + \frac{\bar{c}s_j(1-s_j)}{n-1} \end{bmatrix} \mathbb{P}(S_g = s_j) \left[\frac{s_j}{\mathbb{E}(S_g)}\right]^z \left[\frac{1-s_j}{\mathbb{E}(1-S_g)}\right]^{(1-z)}$$

by iterated expectations, Bayes' Theorem and Assumptions 1 and 2. Simplifying,

$$\mathbf{Q}_0(\bar{c}, n) = \frac{1}{\mathbb{E}(1-S_g)} \begin{bmatrix} \mathbb{E}\{1-S_g\} & \bar{c}\,\mathbb{E}\{S_g(1-S_g)\} \\ \bar{c}\,\mathbb{E}\{S_g(1-S_g)\} & \bar{c}^2\,\mathbb{E}\{S_g^2(1-S_g)\} + \frac{\bar{c}}{n-1}\mathbb{E}\{S_g(1-S_g)^2\} \end{bmatrix} \tag{8}$$

$$\mathbf{Q}_1(\bar{c}, n) = \frac{1}{\mathbb{E}(S_g)} \begin{bmatrix} \mathbb{E}\{S_g\} & \bar{c}\,\mathbb{E}\{S_g^2\} \\ \bar{c}\,\mathbb{E}\{S_g^2\} & \bar{c}^2\,\mathbb{E}\{S_g^3\} + \frac{\bar{c}}{n-1}\mathbb{E}\{S_g^2(1-S_g)\}. \end{bmatrix} \tag{9}$$

We consider three special cases, corresponding to different choices for the saturations S_g. Suppose first that there is a single saturation s. Although this is a degenerate random saturation experiment—there is only one saturation—it is still compatible with the conditions of Theorem 2. Since $\mathbb{P}(S_g = s) = 1$ the expectation of any function $h(S_g)$ simply equals $h(s)$. Accordingly, (8) and (9) simplify to yield

$$\det[\mathbf{Q}_0(\bar{c}, n)] = \det[\mathbf{Q}_1(\bar{c}, n)] = \begin{vmatrix} 1 & \bar{c}s \\ \bar{c}s & (\bar{c}s)^2 + \frac{\bar{c}s(1-s)}{n-1} \end{vmatrix} = \frac{\bar{c}s(1-s)}{n-1}.$$

As long as the single saturation s is interior, $0 < s < 1$, we see that both $\mathbf{Q}_0(\bar{c}, n)$ and $\mathbf{Q}_1(\bar{c}, n)$ are invertible for all n and any $\bar{c} \neq 0$. At the same time, the two matrices are arbitrarily close to being singular for *any* $\bar{c}$ provided that n is sufficiently large. Now consider a so-called "cluster randomized" experiment in which there are two saturations, 0 and 1, and

[10] Note that, by construction, $\mathbf{Q}_z^*$ equals $\mathbf{Q}_z$ wherever the latter is well-defined.

$\mathbb{P}(S_g = 1) = p$. Calculating the expectations in (8) and (9) under this distribution for S_g,

$$\mathbf{Q}_0(\bar{c}, n) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{Q}_1(\bar{c}, n) = \begin{bmatrix} 1 & \bar{c} \\ \bar{c} & \bar{c}^2 \end{bmatrix}.$$

In this case neither $\mathbf{Q}_0$ nor $\mathbf{Q}_1$ is invertible for *any* values of n and $\bar{c}$, so none of the inverses from Lemma 3 exist in a cluster randomized experiment. Finally, consider a design with two distinct, equally likely saturations: $s_L < s_H$, $\mathbb{P}(S_g = s_L) = \mathbb{P}(S_g = s_H) = 1/2$. Taking the limits of (8) and (9) as $n \to \infty$, we have

$$\begin{aligned} \lim_{n\to\infty} \det\left[\mathbf{Q}_0^*(\bar{c}, n)\right] &= \left[\frac{\bar{c}}{\mathbb{E}(1 - S_g)}\right]^2 \left[\mathbb{E}\left\{1 - S_g\right\} \mathbb{E}\left\{S_g^2(1 - S_g)\right\} - \mathbb{E}\left\{S_g(1 - S_g)\right\}^2\right] \\ \lim_{n\to\infty} \det\left[\mathbf{Q}_1^*(\bar{c}, n)\right] &= \left[\frac{\bar{c}}{\mathbb{E}(S_g)}\right]^2 \left[\mathbb{E}\left\{S_g\right\} \mathbb{E}\left\{S_g^3\right\} - \mathbb{E}\left\{S_g^3\right\}^2\right] \end{aligned}$$

and, under the specified distribution for S_g, straightforward but tedious algebra gives

$$\begin{aligned} \mathbb{E}\left\{1 - S_g\right\} \mathbb{E}\left\{S_g^2(1 - S_g)\right\} - \mathbb{E}\left\{S_g(1 - S_g)\right\}^2 &= \frac{1}{4}(1 - s_L)(1 - s_H)(s_H - s_L)^2 \\ \mathbb{E}\left\{S_g\right\} \mathbb{E}\left\{S_g^3\right\} - \mathbb{E}\left\{S_g^3\right\}^2 &= \frac{1}{4} s_L s_H (s_H - s_L)^2. \end{aligned}$$

Therefore the limit of $\mathbf{Q}_0^*(\bar{c}, n)$ is invertible at any $\bar{c} \neq 0$ if and only if $s_L \neq s_H$ and both saturations are less than one. Similarly, the limit of $\mathbf{Q}_1^*(\bar{c}, n)$ is invertible at any $\bar{c} \neq 0$ whenever $s_L \neq s_H$ and neither saturation is zero.

These three examples demonstrate that there two distinct sources of experimental variation that determine the rank of $\mathbf{Q}_0$ and $\mathbf{Q}_1$: variation in S_g *between* groups, and variation in the fraction of compliers offered treatment *within* the groups assigned to a given saturation. Our first example, that of a single saturation, had no between variation: only within variation. And, as we have seen, all within variation vanishes in the limit as group size increases. In our second example, the cluster randomized experiment, the situation was reversed. Because everyone in a given group is either offered ($S_g = 0$) or unoffered ($S_g = 1$), this design generates no within variation. While a cluster randomized design does generate some between variation, it is too coarse to identify our effects of interest: under our assumptions $\bar{D}_{ig}$ equals zero when $S_g = 0$ and $\bar{C}_{ig}$ when $S_g = 1$. Our third example eliminated all within variation by taking group size to infinity. We showed that two interior saturations provide sufficient between variation to identify all of the effects in Theorem 2 for any $\bar{c} \neq 0$ when the potential outcome functions are assumed to be linear.

3.1 A Kernel Estimator

If C_{ig} were observed, it would be straightforward to estimate the causal parameters identified in Theorem 3, using kernel averages to approximate the conditional expectations in Theorem 2 and the empirical distribution of $(\bar{C}_{ig}, N_g)$ scaled by the appropriate importance weights to average these "localized" effects over the cross-section. Standard results for kernel estimation could then be applied to establish consistency as the number of groups grows and the bandwidth approaches zero at an appropriate rate. Unfortunately, $\bar{C}_{ig}$ is unobserved and must instead be estimated.[11] Unfortunately, as G grows, so does the number of unknown values $\bar{C}_{ig}$ that we must estimate, a classic incidental parameters problem.[12] To address this challenge, we consider an asymptotic sequence in which group sizes grow sufficiently fast relative to the number of groups that $\bar{C}_{ig}$ can be replaced by an estimated value while still permitting consistent estimation of the causal effects of interest. In practical terms, our estimator is appropriate for settings with a large number of relatively large groups, e.g. the experiment of Crépon et al. (2013).

While more demanding in its requirements on sample size, the large group setting brings with it two simplifications. First, while our identification results from section 3 involve conditioning on both the share of compliers $\bar{C}_{ig}$ and group size N_g, conditioning on N_g is superfluous in large groups. As explained above, conditioning is necessary because $(\bar{C}_{ig}, N_g)$ induce heterogeneity in the "first-stage," i.e. the distribution of $\bar{D}_{ig}$ depends on them. Inspection of Lemma 1, however, reveals that $\bar{D}_{ig} \approx S_g \bar{C}_{ig}$ when N_g is large, so that only $\bar{C}_{ig}$ is a source of first-stage heterogeneity in the limit. Second, as group sizes approach infinity the distinction between the individual-specific variables $\bar{C}_{ig}$ and $\bar{D}_{ig}$ and their group-specific counterparts

$$\bar{C}_g \equiv \frac{1}{N_g} \sum_{i=1}^{N_g} C_{ig}, \quad \bar{D}_g \equiv \frac{1}{N_g} \sum_{i=1}^{N_g} D_{ig}$$

becomes irrelevant. In other words, including or excluding one additional person is negligible in a large group. These observations motivate our use of a simple kernel estimator that uses only group-level information and conditions only on an estimate of the share of compliers.

To simplify the discussion below we first introduce some additional notation. Let $\boldsymbol{\beta}(c, n)$

[11] Under one-sided non-compliance, $\bar{C}_{ig}$ is in fact observed for individuals in a group with a saturation of 100%. Because everyone in such a group has $Z_{ig} = 1$, the compliers are precisely those individuals with $D_{ig} = 1$. At any other saturation, however, $\bar{C}_{ig}$ cannot be observed. As argued in the preceding section, at least two interior saturations (i.e. strictly between zero and one) are required to avoid problems of weak identification so this problem cannot be averted by changing the experimental design.

[12] While $\bar{C}_{ig}$ can vary across individuals in the same group, it can take on at most two different values for fixed g. If a group contains T total individuals, of whom c are compliers and n never-takers, then the share of compliers among a given person's neighbors is either $(c-1)/(T-1)$ if she is a complier or $c/(T-1)$ if she is a never-taker. Thus, the number of incidental parameters is $2G$.

be a generic localized parameter vector from one of the four parts of Theorem 2. Then

$$\boldsymbol{\beta}(c,n) \equiv \mathbb{E}\left[\mathbf{A}_{ig}|\bar{C}_{ig}=c, N_g=n\right]^{-1}\mathbb{E}\left[\mathbf{P}_{ig}|\bar{C}_{ig}=c, N_g=n\right]$$

where $\mathbf{A}_{ig}$ and $\mathbf{P}_{ig}$ are placeholders for the matrices in Theorem 2. For example, part (iv) of the theorem has

$$\boldsymbol{\beta}(c,n) = \mathbb{E}[\boldsymbol{\theta}_{ig}|\bar{C}_{ig}=c, N_g=n], \quad \mathbf{A}_{ig} = (1-Z_{ig})\widetilde{X}_{ig}\widetilde{X}_{ig}', \quad \mathbf{P}_{ig} = (1-Z_{ig})\widetilde{X}_{ig}Y_{ig}.$$

Define $\boldsymbol{\beta}(c)$ to be the analogous quantity conditioning *only* on $\bar{C}_{ig}=c$ but not $N_g=n$.

We construct a kernel estimator of $\boldsymbol{\beta}(c)$ that conditions on the estimated share of compliers $\widehat{C}_g$ within a given group, defined as

$$\widehat{C}_g \equiv \begin{cases} \bar{D}_g/\bar{Z}_g & \text{if } \sum_j Z_{jg} \neq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

In a group in which no one is offered treatment, there is no information upon which to base an estimate of the share of compliers. For this reason we exclude zero-saturation groups, $S_g = 0$, from our analysis. Under Assumption 7(i), doing so does not bias our estimates of localized or average effects. To simplify the notation, we assume without loss of generality that the experimental design contains only positive saturations throughout the remainder of this section. Under Bernoulli offers (Assumption 2) it is possible, although unlikely, that $\bar{Z}_{ig}$ could be zero even if $S_g > 0$. For convenience, our definition in (10) sets $\widehat{C}_g = 0$ in this case.

Let $K(\cdot)$ be a Lipschitz-continuous kernel function with bounded support. Further define $K_h(x) = h^{-1}K(x/h)$ where h is a bandwidth. Our estimator is given by

$$\widehat{\boldsymbol{\beta}}(c) \equiv \left[\frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N_g} K_h(\widehat{C}_g - c)\mathbf{A}_{ig}\right]^{-1}\left[\frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N_g} K_h(\widehat{C}_g - c)\mathbf{P}_{ig}\right]$$

where $N = \sum_{g=1}^{G} N_g$ is the total sample size of the experiment. Because the argument of K_h is constant within group, this expression can be re-written in terms of group aggregates by defining

$$\bar{\mathbf{A}}_g \equiv \frac{1}{N_g}\sum_{i=1}^{N_g}\mathbf{A}_{ig}, \quad \bar{\mathbf{P}}_g \equiv \frac{1}{N_g}\sum_{i=1}^{N_g}\mathbf{P}_{ig}, \quad \widehat{\rho}_g \equiv \frac{N_g}{\frac{1}{G}\sum_{g=1}^{G} N_g}.$$

Note that $\{\widehat{\rho}_g\}_{g=1}^{G}$ is a set of weights that accounts for the different relative group sizes.

Using this notation,

$$\widehat{\boldsymbol{\beta}}(c) = \left[\frac{1}{G}\sum_{g=1}^{G}\widehat{\rho}_g K_h(\widehat{C}_g - c)\bar{\mathbf{A}}_g\right]^{-1}\left[\frac{1}{G}\sum_{g=1}^{G}\widehat{\rho}_g K_h(\widehat{C}_g - c)\bar{\mathbf{P}}_g\right]. \tag{11}$$

To estimate one the four "average" effects $\boldsymbol{\beta}$ from Theorem 3, we average over the observed values of $\widehat{C}_g$, scaling by the appropriate importance weight. Specifically,

$$\widehat{\boldsymbol{\beta}} \equiv \frac{1}{G}\sum_{g=1}^{G}\hat{\omega}_g\widehat{\boldsymbol{\beta}}(\widehat{C}_g). \tag{12}$$

where the specific form of $\widehat{\omega}_g$ depends on which effect from Theorem 3 we have chosen to estimate. Parts (i) and (ii), for example, concern effects for compliers. Accordingly, our estimators for these effects gives additional weight to groups with more compliers. Similarly, our estimator for part (ii) gives additional weight to groups with more never-takers. The specific form of the importance weights is as follows

$$\widehat{\omega}_g \equiv \begin{cases} \widehat{\rho}_g\widehat{C}_g \Big/ \left[\frac{1}{G}\sum_{\ell=1}^{G}\widehat{\rho}_\ell\widehat{C}_\ell\right] & \text{Complier Weights: (i) + (ii)} \\ \widehat{\rho}_g(1-\widehat{C}_g) \Big/ \left[\frac{1}{G}\sum_{\ell=1}^{G}\widehat{\rho}_\ell(1-\widehat{C}_\ell)\right] & \text{Never-taker Weights: (iii)} \\ \widehat{\rho}_g & \text{Full-sample Weights: (iv)} \end{cases} \tag{13}$$

where (i)–(iv) refer to the corresponding parts of Theorem 3. Notice that, as it is concerns an average effect for all individuals in the population, our estimator of the quantity from part (iv) of Theorem 3 uses the "full-sample" weights $\widehat{\rho}_g$, defined above.

Under regularity conditions, the difference between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$ converges in probability to zero along an asymptotic sequence in which the bandwidth approaches zero at an appropriate rate relative to the total number of groups and the minimum group size.

4 Conclusion

In this paper we have proposed methods to identify and estimate direct and indirect causal effects under one-sided non-compliance, using data from a randomized saturation experiment. A possible extension of the methods described above would be to consider settings with two-sided non-compliance. In this case the localize-then-average approach would condition on the share of always-takers in addition to the share of compliers. Another interesting extension would be to consider relaxing Assumption 6 to allow some dependence of individuals' take-up

decisions on the offers of their peers. Work currently in progress explores this possibility.

A Proofs

The following lemma, taken from Constantinou et al. (2017) summarizes several useful properties of conditional independence that we use in our proofs below. The names attached to properties (i) and (iii)–(v) are taken from Pearl (1988). For the purposes of this document, we call the second property "redundancy."

Lemma A.1 (Axioms of Conditional Independence)**.** *Let X, Y, Z, W be random vectors defined on a common probability space, and let h be a measurable function. Then:*

(i) (Symmetry): $X \perp\!\!\!\perp Y|Z \implies Y \perp\!\!\!\perp X|Z$.

(ii) (Redundancy): $X \perp\!\!\!\perp Y|Y$.

(iii) (Decomposition): $X \perp\!\!\!\perp Y|Z$ *and* $W = h(Y) \implies X \perp\!\!\!\perp W|Z$.

(iv) (Weak Union): $X \perp\!\!\!\perp Y|Z$ *and* $W = h(Y) \implies X \perp\!\!\!\perp Y|(W, Z)$.

(v) (Contraction): $X \perp\!\!\!\perp Y|Z$ *and* $X \perp\!\!\!\perp W|(Y, Z) \implies X \perp\!\!\!\perp (Y, W)|Z$.

For simplicity, our proofs below freely use the "Symmetry" property without comment, although we reference the other properties when used. We also rely on the following corollary of Lemma A.1.

Corollary A.1. $X \perp\!\!\!\perp Y|Z$ *implies* $(X, Z) \perp\!\!\!\perp Y|Z$.

Proof of Lemma 1. Applying Corollary A.1 and the Decomposition property to Assumption 7(ii) yields $\boldsymbol{Z}_g \perp\!\!\!\perp (\boldsymbol{C}_g, \bar{C}_{ig})|(N_g, S_g)$. By the definition of conditional independence, it follows that the distribution of $\boldsymbol{Z}_g|(N_g, S_g, \boldsymbol{C}_g, \bar{C}_{ig})$ is the same as that of $\boldsymbol{Z}_g|(N_g, S_g)$:

$$\mathbb{P}(\boldsymbol{Z}_g = \boldsymbol{z}|N_g = n, S_g = s, \boldsymbol{C}_g, \bar{C}_{ig}) = \mathbb{P}(\boldsymbol{Z}_g = \boldsymbol{z}|N_g = n, S_g = s). \tag{A.1}$$

Now, define the shorthand $A \equiv \{N_g = n, S_g = s, \boldsymbol{C}_g = \boldsymbol{c}, \bar{C}_{ig} = \bar{c}\}$ and let $\mathcal{C}(i)$ be the indices of all non-zero components of $\boldsymbol{c}$, *excluding* the ith component, i.e. $\mathcal{C}(i) \equiv \{j \neq i \colon c_j = 1\}$. By the definition of $\bar{D}_{ig}$, the event $\{\bar{D}_{ig} = d\}$ is equivalent to $\left\{\sum_{j \neq i} C_{jg} Z_{jg} = d(N_g - 1)\right\}$. Consequently,

$$\mathbb{P}(\bar{D}_{ig} = d|A, Z_{ig}) = \mathbb{P}\left(\left[\sum_{j \neq i} C_{jg} Z_{jg}\right] = d(n-1) \middle| A, Z_{ig}\right) = \mathbb{P}\left(\left[\sum_{j \in \mathcal{C}(i)} Z_{jg}\right] = d(n-1) \middle| A, Z_{ig}\right)$$

where the first equality uses the fact that A implies $N_g = n$, and the second uses the fact that A implies $\boldsymbol{C}_g = \boldsymbol{c}$, so we know precisely which of the indicators C_{jg} equal zero and which equal one.

For part (i), suppose that Assumption 2 holds. Then (A.1) implies that $\boldsymbol{Z}_g|A \sim$ iid Bernoulli(s). By our definition of $\mathcal{C}(i)$ it follows that, conditional on A, the subvector of $\boldsymbol{Z}_g$ that corresponds to $\mathcal{C}(i)$ constitutes an iid sequence of $\bar{c}(n-1)$ Bernoulli(s) random variables, each of which is *independent of* Z_{ig}. Hence, conditional on (A, Z_{ig}), we see that $\sum_{j \in \mathcal{C}(i)} Z_{jg} \sim$ Binomial$\big(\bar{c}(n-1), s\big)$.

For part (ii), suppose that Assumption 3 holds. Our task is to calculate the probability that $\sum_{j \in \mathcal{C}(i)} Z_{jg} = d(n-1)$ given A and $Z_{ig} = z$. By the definition of $\mathcal{C}(i)$ this is simply the probability

that exactly $d(n-1)$ of the $(n-1)\bar{c}$ compliers (excluding person i) are offered treatment, conditional on A and the treatment offer z made to person i. Under Assumption 3, (A.1) implies

$$\mathbb{P}(\boldsymbol{Z}_g = \boldsymbol{z}|A) = \begin{cases} \binom{n}{\lfloor ns \rfloor}^{-1}, & \text{if } \sum_i z_i = \lfloor sn \rfloor \\ 0, & \text{otherwise.} \end{cases}$$

Hence, conditional on $(A, Z_{ig} = z)$, the allocation of treatment offers is equivalent to drawing $\lfloor ns \rfloor - z$ balls without replacement from an urn containing $n-1$ balls in total. Of the balls $(n-1)\bar{c}$ are red, corresponding to the compliers, and $(n-1)(1-\bar{c})$ are white, corresponding to the never-takers. This follows from our definition of $\bar{C}_{ig}$, which *excludes* person (i, g). Conditional on A and $Z_{ig} = z$, the sum $\sum_{j \in \mathcal{C}(i)} Z_{jg}$ is simply the number of red balls that we draw from the urn. If $z = 0$, then person (i, g) was not offered treatment so we make $\lfloor ns \rfloor$ draws from the urn; if $z = 1$, then person (i, g) was offered treatment, so we make only $\lfloor ns \rfloor - 1$ draws from the urn. Hence, given $(A, Z_{ig} = z)$, the sum $\sum_{j \in \mathcal{C}(i)} Z_{jg}$ is a Hypergeometric$(n-1, (n-1)\bar{c}, \lfloor ns \rfloor - z)$ random variable. □

Proof of Lemma 2. Define the shorthand $\mathcal{R} \equiv \{\boldsymbol{C}_g, \mathbf{B}_g, N_g\}$. By the law of total covariance

$$\begin{aligned} \text{Cov}\left[S_g, (\gamma_{ig} - \gamma)\bar{D}_{ig}\right] &= \mathbb{E}\left\{\text{Cov}\left[S_g, (\gamma_{ig} - \gamma)\bar{D}_{ig}|\mathcal{R}\right]\right\} + \text{Cov}\left\{\mathbb{E}\left[S_g|\mathcal{R}\right], \mathbb{E}\left[(\gamma_{ig} - \gamma)\bar{D}_{ig}|\mathcal{R}\right]\right\} \\ \text{Cov}(S_g, \bar{D}_{ig}) &= \mathbb{E}\left[\text{Cov}\left(S_g, \bar{D}_{ig}|\mathcal{R}\right)\right] + \text{Cov}\left[\mathbb{E}\left(S_g|\mathcal{R}\right), \mathbb{E}\left(\bar{D}_{ig}|\mathcal{R}\right)\right]. \end{aligned}$$

And by Assumption 7 (i), $\mathbb{E}(S_g|\mathcal{R}) = \mathbb{E}(S_g)$, a constant. Hence, the the covariance of any random variable with $\mathbb{E}(S_g|\mathcal{R})$ is zero. This leaves,

$$\text{Cov}\left[S_g, (\gamma_{ig} - \gamma)\bar{D}_{ig}\right] = \mathbb{E}\left[(\gamma_{ig} - \gamma)\text{Cov}\left(S_g, \bar{D}_{ig}|\mathcal{R}\right)\right] \tag{A.2}$$

$$\text{Cov}(S_g, \bar{D}_{ig}) = \mathbb{E}\left[\text{Cov}\left(S_g, \bar{D}_{ig}|\mathcal{R}\right)\right] \tag{A.3}$$

since γ is constant and γ_{ig} is $\mathcal{R}$-measurable. The law of total covariance also holds conditionally, and as a special case of this result, $\text{Cov}(X, Y|Z) = \text{Cov}\left[X, \mathbb{E}(Y|Z, X)|Z\right]$. [13] Accordingly,

$$\text{Cov}(S_g, \bar{D}_{ig}|\mathcal{R}) = \text{Cov}\left[S_g, \mathbb{E}(\bar{D}_{ig}|\mathcal{R}, S_g)\right] = \text{Cov}\left(S_g, S_g\bar{C}_{ig}|\mathcal{R}\right) = \bar{C}_{ig}\text{Var}(S_g|\mathcal{R}) = \bar{C}_{ig}\text{Var}(S_g)$$

where the second equality uses $\mathbb{E}[\bar{D}_{ig}|\mathcal{R}, S_g] = S_g\bar{C}_{ig}$, as implied by Lemma 1 under Assumptions 1–2 and 6–7, the third equality uses the fact that $\bar{C}_{ig}$ is $\mathcal{R}$-measurable, and the fourth follows from $S_g \perp\!\!\!\perp \mathcal{R}$ (Assumption 7). The result follows by substituting $\text{Cov}(S_g, \bar{D}_{ig}|\mathcal{R}) = \bar{C}_{ig}\text{Var}(S_g)$ into (A.2) and (A.3), since $\text{Var}(S_g)$ is a constant and $\text{Cov}\left[(\gamma_{ig} - \gamma), \bar{C}_{ig}\right] = \mathbb{E}\left[(\gamma_{ig} - \gamma)\bar{C}_{ig}\right]$ since $\gamma = \mathbb{E}(\gamma_{ig})$. □

Proof of Theorem 1. Assumption 7(i) implies $(\boldsymbol{C}_g, \mathbf{B}_g) \perp\!\!\!\perp S_g|N_g$ by Weak Union and Decomposition. Combining this with Assumption 7(ii) gives

$$(\mathbf{Z}_g, S_g) \perp\!\!\!\perp (\mathbf{B}_g, \boldsymbol{C}_g)|N_g \tag{A.4}$$

by Contraction. Now let $\boldsymbol{C}_{-ig}$ denote the subvector of $\boldsymbol{C}_g$ that excludes element i. Applying Decomposition, Corollary A.1, and Weak Union to (A.4),

$$(S_g, \boldsymbol{Z}_g) \perp\!\!\!\perp (B_{ig}, C_{ig}, \boldsymbol{C}_{-ig}, N_g)|(N_g, \bar{C}_{ig}). \tag{A.5}$$

[13] The general form is $\text{Cov}(X, Y|Z) = \mathbb{E}\left[\text{Cov}(X, Y|Z, W)|Z\right] + \text{Cov}\left[\mathbb{E}\left(X|Z, W\right), \mathbb{E}(Y|Z, W)|Z\right]$.

because $\bar{C}_{ig}$ is a function of $(\boldsymbol{C}_g, N_g)$. Since neither the distribution of X nor Y from Lemma 1 depends on $\mathbf{c}$,

$$\bar{D}_{ig} \perp\!\!\!\perp \boldsymbol{C}_{-ig} | (N_g, \bar{C}_{ig}, S_g, Z_{ig}). \tag{A.6}$$

Applying Decomposition to (A.5) gives $\boldsymbol{C}_{-ig} \perp\!\!\!\perp (S_g, Z_{ig}) | (N_g, \bar{C}_{ig})$. Combining this with (A.6),

$$(S_g, Z_{ig}, \bar{D}_{ig}) \perp\!\!\!\perp \boldsymbol{C}_{-ig} | (N_g, \bar{C}_{ig}) \tag{A.7}$$

by Contraction. Now, applying Weak Union, Decomposition, and Corollary A.1 to (A.5),

$$(S_g, Z_{ig}, \bar{D}_{ig}) \perp\!\!\!\perp (B_{ig}, C_{ig}) | (\boldsymbol{C}_{-ig}, \bar{C}_{ig}, N_g). \tag{A.8}$$

since $\bar{D}_{ig}$ is a function of $(\boldsymbol{Z}_g, \boldsymbol{C}_{-ig}, N_g)$. Finally, applying Contraction to (A.7) and (A.8),

$$(S_g, Z_{ig}, \bar{D}_{ig}) \perp\!\!\!\perp (\boldsymbol{C}_{-ig}, B_{ig}, C_{ig}) | (\bar{C}_{ig}, N_g)$$

and the result follows by a final application of Decomposition. □

Proof of Theorem 2(i). Under IOR $D_{ig} = C_{ig} Z_{ig}$. Defining $\mathbf{M}_{ig} \equiv \text{diag}\{1, C_{ig}\} \otimes \mathbb{I}_K$,

$$\mathbf{X}_{ig} = \left(\begin{bmatrix} 1 & 0 \\ 0 & C_{ig} \end{bmatrix} \begin{bmatrix} 1 \\ Z_{ig} \end{bmatrix} \right) \otimes \left[\mathbb{I}_K \mathbf{f}(\bar{D}_{ig}) \right] = \left(\begin{bmatrix} 1 & 0 \\ 0 & C_{ig} \end{bmatrix} \otimes \mathbb{I}_K \right) \left(\begin{bmatrix} 1 \\ Z_{ig} \end{bmatrix} \otimes \mathbf{f}(\bar{D}_{ig}) \right) = \mathbf{M}_{ig} \mathbf{W}_{ig}$$

from which we obtain

$$\mathbf{W}_{ig} \mathbf{X}_{ig}' = \mathbf{W}_{ig} \mathbf{W}_{ig}' \mathbf{M}_{ig} \tag{A.9}$$

along with

$$\mathbf{W}_{ig} Y_{ig} = \mathbf{W}_{ig} \left(\mathbf{X}_{ig}' \mathbf{B}_{ig} \right) = \mathbf{W}_{ig} \mathbf{W}_{ig}' \mathbf{M}_{ig} \mathbf{B}_{ig} \tag{A.10}$$

using the fact that $\mathbf{M}_{ig}$ is a symmetric matrix. Substituting (A.9) gives

$$\mathbb{E}\left[\mathbf{W}_{ig} \mathbf{X}_{ig}' | \bar{C}_{ig}, N_g \right] = \mathbb{E}\left[\mathbf{W}_{ig} \mathbf{W}_{ig}' | \bar{C}_{ig}, N_g \right] \mathbb{E}\left[\mathbf{M}_{ig} | \bar{C}_{ig}, N_g \right] \tag{A.11}$$

by Theorem 1, since $\mathbf{W}_{ig} \mathbf{W}_{ig}'$ is a measurable function of $(Z_{ig}, \bar{D}_{ig})$ and $\mathbf{M}_{ig}$ is a measurable function of C_{ig}. Similarly, substituting (A.10) and again applying Theorem 1,

$$\mathbb{E}\left[\mathbf{W}_{ig} Y_{ig} | \bar{C}_{ig}, N_g \right] = \mathbb{E}\left[\mathbf{W}_{ig} \mathbf{W}_{ig}' | \bar{C}_{ig}, N_g \right] \mathbb{E}\left[\mathbf{M}_{ig} \mathbf{B}_{ig} | \bar{C}_{ig}, N_g \right] \tag{A.12}$$

since $\mathbf{M}_{ig} \mathbf{B}_{ig}$ is a measurable function of $(\mathbf{B}_{ig}, C_{ig})$. Now, By iterated expectations,

$$\begin{aligned}
\mathbb{E}\left[\mathbf{M}_{ig} \mathbf{B}_{ig} | \bar{C}_{ig}, N_g \right] &= \begin{bmatrix} \mathbb{E}\left(\boldsymbol{\theta}_{ig} | \bar{C}_{ig}, N_g \right) \\ \mathbb{E}\left(C_{ig} \left\{ \boldsymbol{\psi}_{ig} - \boldsymbol{\theta}_{ig} \right\} | \bar{C}_{ig}, N_g \right) \end{bmatrix} \\
&= \begin{bmatrix} \mathbb{E}\left(\boldsymbol{\theta}_{ig} | \bar{C}_{ig}, N_g \right) \\ \mathbb{E}\left(C_{ig} | \bar{C}_{ig}, N_g \right) \mathbb{E}\left(\left\{ \boldsymbol{\psi}_{ig} - \boldsymbol{\theta}_{ig} \right\} | C_{ig} = 1, \bar{C}_{ig}, N_g \right) \end{bmatrix} \\
&= \mathbb{E}\left[\mathbf{M}_{ig} | \bar{C}_{ig}, N_g \right] \begin{bmatrix} \mathbb{E}\left(\boldsymbol{\theta}_{ig} | \bar{C}_{ig}, N_g \right) \\ \mathbb{E}\left(\boldsymbol{\psi}_{ig} - \boldsymbol{\theta}_{ig} | C_{ig} = 1, \bar{C}_{ig}, N_g \right) \end{bmatrix}.
\end{aligned}$$

The result follows by substituting this expression for $\mathbb{E}\left[\mathbf{M}_{ig} \mathbf{B}_{ig} | \bar{C}_{ig}, N_g \right]$ into (A.12) and combining with (A.11). □

Proof of Theorem 2(ii). By iterated expectations,

$$\mathbb{E}[D_{ig}Z_{ig}\widetilde{\mathbf{X}}_{ig}'Y_{ig}|\bar{C}_{ig},N_g]=\mathbb{E}[D_{ig}Z_{ig}|\bar{C}_{ig},N_g]\mathbb{E}[\widetilde{\mathbf{X}}_{ig}'Y_{ig}|D_{ig}Z_{ig}=1,\bar{C}_{ig},N_g] \tag{A.13}$$

and similarly,

$$\mathbb{E}[D_{ig}Z_{ig}\widetilde{\mathbf{X}}_{ig}\widetilde{\mathbf{X}}_{ig}'|\bar{C}_{ig},N_g]=\mathbb{E}[D_{ig}Z_{ig}|\bar{C}_{ig},N_g]\mathbb{E}[\widetilde{\mathbf{X}}_{ig}\widetilde{\mathbf{X}}_{ig}'|D_{ig}Z_{ig}=1,\bar{C}_{ig},N_g]. \tag{A.14}$$

Under IOR and one-sided non-compliance,

$$\{D_{ig}Z_{ig}=1\}=\{D_{ig}=1,Z_{ig}=1\}=\{C_{ig}=1,Z_{ig}=1\}. \tag{A.15}$$

Hence,

$$\mathbb{E}[\widetilde{\mathbf{X}}_{ig}Y_{ig}|D_{ig}Z_{ig}=1,\bar{C}_{ig},N_g]=\mathbb{E}[\widetilde{\mathbf{X}}_{ig}\widetilde{\mathbf{X}}_{ig}'\boldsymbol{\psi}_{ig}|C_{ig}=1,Z_{ig}=1,\bar{C}_{ig},N_g] \tag{A.16}$$

by (A.15), using the fact that $Y_{ig}=\widetilde{\mathbf{X}}_{ig}'\left\{(1-D_{ig})\boldsymbol{\theta}_{ig}+D_{ig}\boldsymbol{\psi}_{ig}\right\}$. Applying Weak Union and Decomposition to Theorem 1, we obtain $\bar{D}_{ig}\perp\!\!\!\perp\mathbf{B}_{ig}|(C_{ig},\bar{C}_{ig},N_g)$, which in turn implies

$$\widetilde{X}_{ig}\perp\!\!\!\perp\boldsymbol{\psi}_{ig}|(C_{ig},Z_{ig},\bar{C}_{ig},N_g) \tag{A.17}$$

since $\widetilde{X}_{ig}$ is a measurable function of $\bar{D}_{ig}$ and $\boldsymbol{\psi}_{ig}$ is a measurable function of $\mathbf{B}_{ig}$. Proceeding similarly, $\mathbf{B}_{ig}\perp\!\!\!\perp Z_{ig}|(C_{ig},\bar{C}_{ig},N_g)$ which implies

$$\boldsymbol{\psi}_{ig}\perp\!\!\!\perp Z_{ig}|(C_{ig},\bar{C}_{ig},N_g). \tag{A.18}$$

Thus, applying (A.15), (A.17), and (A.18) to (A.16),

$$\mathbb{E}[\widetilde{\mathbf{X}}_{ig}Y_{ig}|D_{ig}Z_{ig}=1,\bar{C}_{ig},N_g]=\mathbb{E}[\widetilde{\mathbf{X}}_{ig}\widetilde{\mathbf{X}}_{ig}'|D_{ig}Z_{ig}=1,\bar{C}_{ig},N_g]\mathbb{E}[\boldsymbol{\psi}_{ig}|C_{ig}=1,\bar{C}_{ig},N_g] \tag{A.19}$$

The result follows by combining (A.13), (A.14), and (A.19). □

Proof of Theorem 2(iii). Because this proof is nearly identical to that of Theorem 2(ii), we merely outline the differences here. First, in the iterated expectations step, we condition on the event $(1-D_{ig})Z_{ig}=1$. Second, we use $\{(1-D_{ig})Z_{ig}=1\}=\{D_{ig}=0,Z_{ig}=1\}=\{C_{ig}=0,Z_{ig}=1\}$ to change the conditioning set when manipulating $\mathbb{E}[\widetilde{\mathbf{X}}_{ig}Y_{ig}|(1-D_{ig})Z_{ig},\bar{C}_{ig},N_g]$. Third, we note that Y_{ig} is equal to $\widetilde{\mathbf{X}}_{ig}'\boldsymbol{\theta}_{ig}$ given $D_{ig}=0$. Fourth, we apply Lemma A.1 to Theorem 1 to obtain $\widetilde{X}_{ig}\perp\!\!\!\perp\boldsymbol{\theta}_{ig}|(C_{ig},Z_{ig},\bar{C}_{ig},N_g)$ and $\boldsymbol{\theta}_{ig}\perp\!\!\!\perp Z_{ig}|(C_{ig},\bar{C}_{ig},N_g)$. □

Proof of Theorem 2(iv). By iterated expectations,

$$\begin{aligned}\mathbb{E}\left[(1-Z_{ig})\,\widetilde{\mathbf{X}}_{ig}Y_{ig}|\bar{C}_{ig},N_g\right]&=\mathbb{E}\left[1-Z_{ig}|\bar{C}_{ig},N_g\right]\mathbb{E}\left[\widetilde{\mathbf{X}}_{ig}Y_{ig}|Z_{ig}=0,\bar{C}_{ig},N_g\right]\\ \mathbb{E}\left[(1-Z_{ig})\,\widetilde{\mathbf{X}}_{ig}\widetilde{\mathbf{X}}_{ig}\,|\bar{C}_{ig},N_g\right]&=\mathbb{E}\left[1-Z_{ig}\,|\bar{C}_{ig},N_g\right]\mathbb{E}\left[\widetilde{\mathbf{X}}_{ig}\widetilde{\mathbf{X}}_{ig}\,|Z_{ig}=0,\bar{C}_{ig},N_g\right].\end{aligned}$$

Now, by Theorem 1 $(Z_{ig},\bar{D}_{ig},S_g)\perp\!\!\!\perp(\mathbf{B}_{ig},C_{ig})|(\bar{C}_{ig},N_g)$ which implies $\boldsymbol{\theta}_{ig}\perp\!\!\!\perp\bar{D}_{ig}|(Z_{ig},\bar{C}_{ig},N_g)$ by weak union and decomposition. Since $\widetilde{\mathbf{X}}_{ig}$ is a measurable function of $\bar{D}_{ig}$,

$$\mathbb{E}\left[\widetilde{\mathbf{X}}_{ig}Y_{ig}\,|Z_{ig}=0,\bar{C}_{ig},N_g\right]=\mathbb{E}\left[\widetilde{\mathbf{X}}_{ig}\widetilde{\mathbf{X}}_{ig}'\,|Z_{ig}=0,\bar{C}_{ig},N_g\right]\mathbb{E}\left[\boldsymbol{\theta}_{ig}\,|Z_{ig}=0,\bar{C}_{ig},N_g\right]$$

because $Z_{ig}=0$ implies $Y_{ig}=\widetilde{\mathbf{X}}_{ig}'\boldsymbol{\theta}_{ig}$ under one-sided non-compliance. The result follows by

combining the three displayed equations, using $\boldsymbol{\theta}_{ig} \perp\!\!\!\perp Z_{ig} | (\bar{C}_{ig}, N_g)$ as implied by Theorem 1 and decomposition. □

Lemma A.2. *Let A and B be $(m \times m)$ matrices and define*

$$U \equiv \begin{bmatrix} A+B & B \\ B & B \end{bmatrix}.$$

Then U is invertible if and only if A and B are both invertible, in which case $U^{-1} = V$ where

$$V \equiv \begin{bmatrix} A^{-1} & -A^{-1} \\ -A^{-1} & A^{-1} + B^{-1} \end{bmatrix}.$$

Proof of Lemma A.2. The "if" direction follows by direct calculation: $VU = UV = \mathbb{I}_{2m}$. For the "only if" direction, suppose that U is invertible. Partitioning U^{-1} into blocks (C, D, E, F) conformably with the partition of U, we have

$$UU^{-1} = \begin{bmatrix} A+B & B \\ B & B \end{bmatrix} \begin{bmatrix} C & D \\ E & F \end{bmatrix} = \begin{bmatrix} \mathbb{I}_m & 0 \\ 0 & \mathbb{I}_m \end{bmatrix} = \begin{bmatrix} C & D \\ E & F \end{bmatrix} \begin{bmatrix} A+B & B \\ B & B \end{bmatrix} = U^{-1}U.$$

We begin by showing that A is invertible. Consider the product UU^{-1}. Multiplying the first row of U by the first column of U^{-1} gives the equation $AC + B(C+E) = \mathbb{I}_m$; multiplying the second row of U by the first column of U^{-1} gives $B(C+E) = 0$. Combining these, $AC = \mathbb{I}_m$. Now consider the product $U^{-1}U$. Multiplying the first row of U^{-1} by the first column of U gives $CA + (C+D)B = \mathbb{I}_m$; multiplying the first row of U^{-1} by the second column of U gives $(C+D)B = 0$. Combining these, $CA = \mathbb{I}_m$. Since $AC = CA = \mathbb{I}_m$, we have shown that A is invertible with $A^{-1} = C$.

We next show that $D = E = -C$. Consider again the product UU^{-1}. Multiplying the first row of U by the second column of U^{-1} gives $AD + B(D+F) = 0$; multiplying the second row of U by the second column of U^{-1} gives $B(D+F) = \mathbb{I}_m$. Combining these, $AD = -\mathbb{I}_m$ and because $A^{-1} = C$ we can solve this equation to yield $D = -C$. Now consider $U^{-1}U$. Multiplying the second row of U^{-1} by the first column of U gives $EA + (E+F)B = 0$; multiplying the second row of U^{-1} by the second column of U gives $(E+F)B = \mathbb{I}_m$. Combining these, $EA = -\mathbb{I}_m$ and solving for E, we have $E = -C$ since $A^{-1} = C$.

Finally we show that B is invertible. Multiplying the second row of U by the second column of U^{-1} gives $B(D+F) = \mathbb{I}_m$, but since $D = -C$ this becomes $B(F-C) = \mathbb{I}_m$ Multiplying the second row of U^{-1} by the first column of U gives $(E+F)B + EA = 0$ but because $E = -C = A^{-1}$ this becomes $(F-C)B = \mathbb{I}_m$. Thus, $B(F-C) = (F-C)B = \mathbb{I}_m$ so we have shown that B is invertible with $B^{-1} = F - C$. □

Proof of Lemma 3. By the law of total probability,

$$\mathbb{P}(Z_{ig} = 1 | \bar{C}_{ig}, N_g) = \mathbb{E}\left[\mathbb{E}\left(Z_{ig} | S_g\right)\right] = \sum_{j=1}^{J} s_j \mathbb{P}(S_g = s_j) = \mathbb{E}(S_g) \tag{A.20}$$

since $S_g \perp\!\!\!\perp (\bar{C}_g, \mathbf{B}_g, N_g)$ by Assumption 7(i) and $\boldsymbol{Z}_g \perp\!\!\!\perp (\boldsymbol{C}_g, \mathbf{B}_{ig}) | (S_g, N_g)$ by Assumption 7(ii).

Turning our attention to part (i), (A.11) implies that $\mathbb{E}[\mathbf{W}_{ig}\mathbf{X}'_{ig} | \bar{C}_{ig}, N_g]$ is invertible if and only if $\mathbb{E}[\mathbf{W}_{ig}\mathbf{W}'_{ig} | \bar{C}_{ig}, N_g]$ and $\mathbb{E}[\mathbf{M}_{ig} | \bar{C}_{ig}, N_g]$ are both invertible, where we define $\mathbf{M}_{ig} \equiv \text{diag}\{1, C_{ig}\} \otimes \mathbb{I}_K$ as in the proof of Theorem 2(i). The matrix $\mathbb{E}[\mathbf{M}_{ig} | \bar{C}_{ig}, N_g]$, in turn, is invertible if and only if

$\mathbb{E}[C_{ig}|\bar{C}_{ig}, N_g] \neq 0$. It remains to examine $\mathbb{E}[\mathbf{W}_{ig}\mathbf{W}_{ig}'|\bar{C}_{ig}, N_g]$. By (A.20), iterated expectations, and the definition of $\mathbf{W}_{ig}$ in Theorem 1, we have

$$\begin{aligned}\mathbb{E}[\mathbf{W}_{ig}\mathbf{W}_{ig}'|\bar{C}_{ig}, N_g] &= \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} \otimes \mathbf{Q}_0(1-\mathbb{E}[S_g]) + \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} \otimes \mathbf{Q}_1\mathbb{E}[S_g] \\ &= \begin{bmatrix}\{1-\mathbb{E}(S_g)\}\,\mathbf{Q}_0 + \mathbb{E}(S_g)\mathbf{Q}_1 & \mathbb{E}(S_g)\mathbf{Q}_1 \\ \mathbb{E}(S_g)\mathbf{Q}_1 & \mathbb{E}(S_g)\mathbf{Q}_1\end{bmatrix}\end{aligned}$$

where it is understood that the left-hand side conditional expectation, and both $\mathbf{Q}_0$ and $\mathbf{Q}_1$ are evaluated at $(\bar{c}, n)$. It follows by Lemma A.2 that $\mathbb{E}[\mathbf{W}_{ig}\mathbf{W}_{ig}|\bar{C}_{ig}, N_g]$ is invertible if and only if both $[1-\mathbb{E}(S_g)]\mathbf{Q}_0$ and $\mathbb{E}(S_g)\mathbf{Q}_1$ are invertible, which holds precisely when $\mathbb{E}(S_g) \notin \{0, 1\}$ and both $\mathbf{Q}_0$ and $\mathbf{Q}_1$ are invertible. This establishes part (i) of the result.

For part (ii), recall that $D_{ig} = C_{ig}Z_{ig}$ under IOR. Since Z_{ig} is binary, it follows that

$$\mathbb{E}[D_{ig}Z_{ig}\widetilde{X}_{ig}\widetilde{X}_{ig}'|\bar{C}_{ig}, N_g] = \mathbb{E}[C_{ig}Z_{ig}\widetilde{X}_{ig}\widetilde{X}_{ig}'|\bar{C}_{ig}, N_g] = \mathbb{E}[C_{ig}|\bar{C}_{ig}, N_g]\mathbb{E}[Z_{ig}\widetilde{X}_{ig}\widetilde{X}_{ig}'|\bar{C}_{ig}, N_g]$$

using $C_{ig} \perp\!\!\!\perp Z_{ig}\widetilde{X}_{ig}\widetilde{X}_{ig}'|(\bar{C}_{ig}, N_g)$, as implied by Theorem 1. Part (ii) follows since

$$\mathbb{E}[Z_{ig}\widetilde{X}_{ig}\widetilde{X}_{ig}'|\bar{C}_{ig}, N_g] = \mathbf{Q}_1\mathbb{E}(S_g)$$

by iterated expectations. Part (iii) follows similarly, since

$$\mathbb{E}[(1-D_{ig})Z_{ig}\widetilde{X}_{ig}\widetilde{X}_{ig}'|\bar{C}_{ig}, N_g] = \left\{1-\mathbb{E}[C_{ig}|\bar{C}_{ig}, N_g]\right\}\mathbf{Q}_1\mathbb{E}(S_g).$$

Finally, by (A.20) and by iterated expectations, conditioning on $Z_{ig} = 0$,

$$\mathbb{E}\left[(1-Z_{ig})\widetilde{X}_{ig}\widetilde{X}_{ig}\,\middle|\bar{C}_{ig}, N_g\right] = \mathbf{Q}_0\left[1-\mathbb{E}(S_g)\right]$$

from which we obtain part (iv) of the result. □

Proof of Lemma 4. Let $X_n^{(j)}$ be equal in distribution to a Bernoulli$\big(\lfloor(n-1)\bar{c}\rfloor, s_j\big)$ RV multiplied by $1/(n-1)$. Inspection of the mean and variance of $X_n^{(j)}$ establishes that $X_n^{(j)}$ converges in probability to $s_j\bar{c}$, because mean-square convergence implies convergence in probability. Hence, if we let $F_n^{(j)}$ denote the CDF of $X_n^{(j)}$, then $\lim_{n\to\infty} F_n^{(j)}(x) = \mathbb{1}\{x \geq s_j\bar{c}\}$ for all $x \neq s_j\bar{c}$, because convergence in probability to a constant implies convergence in distribution. Now let Y_n be a RV with CDF $F_n(y) \equiv \sum_{j=1}^J F_n^{(j)}(y)\frac{m_j}{G}$. By construction the random variable Y_n has the same distribution as $\bar{D}_{ig}|\big(\bar{C}_{ig} = \lfloor(n-1)\bar{c}\rfloor/(n-1), N_g = n, Z_{ig} = z\big)$ so that

$$\mathbf{Q}_z^*(\bar{c}, n) \equiv \mathbb{E}[\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})'|\bar{C}_{ig} = \lfloor(n-1)\bar{c}\rfloor/(n-1), N_g = n, Z_{ig} = z] = \mathbb{E}[\mathbf{f}(Y_n)\mathbf{f}(Y_n)'].$$

Using the convergence of $F_n^{(j)}$, it follows that Y_n converges in distribution to a random variable Y with cdf $F(y) \equiv \sum_{j=1}^J \mathbb{1}\{y \geq y_j\}\frac{m_j}{G}$ where $y_j \equiv s_j\bar{c}$. Accordingly, since $Y_n \in [0, 1]$ for all n and $\mathbf{f}$ is bounded over the same interval, the bounded convergence theorem gives

$$\lim_{n\to\infty} \mathbb{E}[\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})'|\bar{C}_{ig} = \lfloor(n-1)\bar{c}\rfloor/(n-1), N_g = n, Z_{ig} = z] = \lim_{n\to\infty} \mathbb{E}[\mathbf{f}(Y_n)\mathbf{f}(Y_n)'] = \mathbb{E}[\mathbf{f}(Y)\mathbf{f}(Y)']$$

and because the determinant of a matrix is a continuous function of its elements,

$$\lim_{n\to\infty} \det\left\{\mathbb{E}[\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})'|\bar{C}_{ig} = \lfloor (n-1)\bar{c}\rfloor/(n-1), N_g = n, Z_{ig} = z]\right\} = \det\left\{\mathbb{E}[\mathbf{f}(Y)\mathbf{f}(Y)']\right\}.$$

From its CDF, we see that Y is discrete with support set $\{y_1, y_2, \ldots, y_J\}$, and probability mass function $\mathbb{P}(Y = y_j) = m_j/G$. Hence, $\mathbb{E}\left[\mathbf{f}(Y)\mathbf{f}(Y)'\right] = \sum_{j=1}^{J} \frac{m_j}{G}\mathbf{f}(y_j)\mathbf{f}(y_j)'$, which implies

$$\operatorname{rank}\left(\mathbb{E}\left[\mathbf{f}(Y)\mathbf{f}(Y)'\right]\right) \leq \sum_{j=1}^{J} \operatorname{rank}\left[\frac{m_j}{G}\mathbf{f}(y_j)\mathbf{f}(y_j)'\right] = \sum_{j=1}^{J} \operatorname{rank}\left[\mathbf{f}(y_j)\mathbf{f}(y_j)'\right] = J$$

since $m_j/G > 0$ for all j. Because $\mathbb{E}[\mathbf{f}(Y)\mathbf{f}(Y)']$ is a $(K \times K)$ matrix, its determinant is nonzero precisely when its rank equals K. But we have shown that the rank of $\mathbb{E}[\mathbf{f}(Y)\mathbf{f}(Y)']$ cannot exceed J. Therefore $K > J$ implies $\det\{\mathbb{E}[\mathbf{f}(Y)\mathbf{f}(Y)']\} = 0$ and the result follows. □

References

Akram, A. A., Chowdhury, S., Mobarak, A. M., 2018. Effects of emigration on rural labor markets. URL http://faculty.som.yale.edu/mushfiqmobarak/papers/migrationge.pdf

Altonji, J. G., Matzkin, R. L., 2005. Cross section and panel data estimators for nonseparable models with endogenous regressors. Econometrica 73 (4), 1053–1102.

Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J., 2014. Engaging with massive online courses. In: Proceedings of the 23rd international conference on World wide web. ACM, pp. 687–698.

Angelucci, M., De Giorgi, G., 2009. Indirect effects of an aid program: how do cash transfers affect ineligibles' consumption? American Economic Review 99 (1), 486–508.

Banerjee, A. V., Chattopadhyay, R., Duflo, E., Keniston, D., Singh, N., 2012. Can institutions be reformed from within? evidence from a randomized experiment with the rajasthan police.

Barrera-Osorio, F., Bertrand, M., Linden, L. L., Perez-Calle, F., 2011. Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia. American Economic Journal: Applied Economics 3 (2), 167–95.

Bobba, M., Gignoux, J., 2014. Neighborhood effects and take-up of transfers in integrated social policies: Evidence from Progresa.

Bobonis, G. J., Finan, F., 2009. Neighborhood peer effects in secondary school enrollment decisions. The Review of Economics and Statistics 91 (4), 695–716.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., Fowler, J. H., 2012. A 61-million-person experiment in social influence and political mobilization. Nature 489 (7415), 295.

Bursztyn, L., Cantoni, D., Yang, D., Yuchtman, N., Zhang, J., 2019. Persistent political engagement: Social interactions and the dynamics of protest movements. Working Paper.

Cai, Z., Das, M., Xiong, H., Wu, X., 2006. Functional coefficient instrumental variables models. Journal of Econometrics 133 (1), 207–241.

Constantinou, P., Dawid, A. P., et al., 2017. Extended conditional independence and applications in causal inference. The Annals of Statistics 45 (6), 2618–2653.

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., Zamora, P., 2013. Do labor market policies have displacement effects? evidence from a clustered randomized experiment. The Quarterly Journal of Economics 128 (2), 531–580.

Dawid, A. P., 1979. Conditional independence in statistical theory. Journal of the Royal Statistical Society: Series B (Methodological) 41 (1), 1–15.

Duflo, E., Saez, E., 2003. The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. The Quarterly Journal of Economics 118 (3), 815–842.

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

Giné, X., Mansuri, G., 2018. Together we will: experimental evidence on female voting behavior in pakistan. American Economic Journal: Applied Economics 10 (1), 207–35.

Haushofer, J., Shapiro, J., 2016. The short-term impact of unconditional cash transfers to the poor: experimental evidence from kenya. The Quarterly Journal of Economics 131 (4), 1973–2042.

Heckman, J., Vytlacil, E., 1998. Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. Journal of Human Resources, 974–987.

Hoderlein, S., Sherman, R., 2015. Identification and estimation in a correlated random coefficients binary response model. Journal of Econometrics 188 (1), 135–149.

Hudgens, M. G., Halloran, M. E., 2008. Toward causal inference with interference. Journal of the American Statistical Association 103 (482), 832–842.

Imai, K., Jiang, Z., Anup Malani, 2018. Causal Inference with Interference and Noncompliance in Two-Stage Randomized Experiments.
URL http://imai.princeton.edu/research/files/spillover.pdf

Imbens, G. W., Newey, W. K., 2009. Identification and estimation of triangular simultaneous equations models without additivity. Econometrica 77 (5), 1481–1512.

Kang, H., Imbens, G., 2016. Peer Encouragement Designs in Causal Inference with Partial Interference and Identification of Local Average Network Effects, 1–39.
URL http://arxiv.org/abs/1609.04464

Manski, C. F., 2013. Identification of treatment response with social interactions. Econometrics Journal 16 (1), 1–23.

Masten, M. A., Torgovitsky, A., 2016. Identification of instrumental variable correlated random coefficients models. Review of Economics and Statistics 98 (5), 1001–1005.

Miguel, E., Kremer, M., 2004. Worms: identifying impacts on education and health in the presence of treatment externalities. Econometrica, 159–217.

Pearl, J., 1988. Probabilistic reasoning in intelligent systems: Networks of plausible inference.

Sinclair, B., McConnell, M., Green, D. P., 2012. Detecting spillover effects: Design and analysis of multilevel experiments. American Journal of Political Science 56 (4), 1055–1069.

Wooldridge, J. M., 1997. On two stage least squares estimation of the average treatment effect in a random coefficient model. Economics Letters 56 (2), 129–133.

Wooldridge, J. M., 2003. Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. Economics Letters 79 (2), 185–191.

Wooldridge, J. M., 2016. Instrumental variables estimation of the average treatment effect in the correlated random coefficient model. Advances in Econometrics 21, 93–116.

Yi, H., Song, Y., Liu, C., Huang, X., Zhang, L., Bai, Y., Ren, B., Shi, Y., Loyalka, P., Chu, J., et al., 2015. Giving kids a head start: The impact and mechanisms of early commitment of financial aid on poor students in rural China. Journal of Development Economics 113, 1–15.