Identifying Causal Effects in Experiments with Spillovers 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
[3]Development Research Institute, New York University

First Version: September 19, 2019 This Version: February 4, 2022

Abstract

This paper shows how to use a randomized saturation experimental design to identify and estimate causal effects in the presence of spillovers–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 spillovers occur within known groups, and take-up decisions are invariant to peers' realized offers. In this setting we point identify the effects of treatment-on-the-treated, both direct and indirect, in a flexible random coefficients model that allows for heterogeneous treatment effects and endogenous selection into treatment. We go on to propose a feasible estimator that is consistent and asymptotically normal as the number and size of groups increases. We apply our estimator to data from a large-scale job placement services experiment, and find negative indirect treatment effects on the likelihood of employment for those willing to take up the program. These negative spillovers are offset by positive direct treatment effects from own take-up.

Keywords: spillovers, non-compliance, randomized saturation, treatment effects

JEL Codes: C21, C26

*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 Esther Duflo, Roland Rathelot, and Philippe Zamora for their help securing our access to the experimental data set we use in this paper. We also thank Steve Bond, Christina Goldschmidt, Luojia Hu, seminar participants at The Philadelphia Fed, the 2018 IAAE Annual Conference, UPenn, Oxford, the 2018 SEA Annual Meetings, and the 2020 Econometric Society World Congress 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 spillovers—also known as 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 spillovers. 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 spillovers 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. Intent-to-treat effects are generally insufficient for policy analysis: comparing costs and benefits requires an estimate of the average effect of treatment on those who experience it. For this reason, we go beyond intent-to-treat effects. In particular, we use the randomized saturation design as a source of instrumental variables to estimate treatment-on-the-treated and treatment-on-the-untreated effects 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 spillovers occur only within groups.

This is reasonable in many experimental settings where, for example, groups correspond to villages, and spillovers 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 is invariant to the realized treatment offers made to her peers. While IOR is a strong assumption, it is *a priori* reasonable in many contexts, for example in online settings where other subjects' treatment offers are unobserved by others (Anderson et al., 2014; Bond et al., 2012; Eckles et al., 2016) confidential (Yi et al., 2015), or observed with a delay. IOR limits but does not rule out strategic behavior. For example, it holds when agents act strategically on their own beliefs about others' actions provided that they are unaware of their peers' offers when making their own take-up decisions. (Bhattacharya et al. (2021) call this an "incomplete information equilibrium.") Most importantly, IOR has testable implications and we find no evidence against it in our empirical example.

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 show how to construct valid and relevant instruments that identify the average causal effects of interest. The key to our approach is a result showing that conditioning on group size n and the share of compliers $\bar{c}$ in a group 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 $\bar{c}$. Using this insight,

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

we provide feasible estimators of direct and indirect causal effects that are consistent and asymptotically normal in the limit as group size grows at an appropriate rate relative to the number of groups. After constructing the appropriate instruments, our estimators can be implemented as simple IV regressions without the need for non-parametric estimation. In a series of simulations we demonstrate that our estimator works well at reasonable sample sizes.

We apply our methods to experimental data from Crépon et al. (2013), a large-scale randomized saturation experiment carried out across French labor markets that offered job-placement services to young adults. In particular, we estimate direct and indirect treatment effects of program take-up for compliers (the treated) and spillovers for never-takers (the untreated). We find large negative indirect effects for compliers, a more vulnerable sub-population than never-takers based on their observed characteristics at baseline. Take-up of the program by these individuals, however, shields them from the negative spillovers induced by the increased take-up of job-placement services by others in their city. The never-taker sub-population, in contrast, is unaffected by such negative spillover effects. Our results go beyond the intent-to-treat effects estimated by Crépon et al. (2013). Whereas they estimate the spillovers from *offering* job placement services, we estimate the labor market displacement effects of *providing* them.

This paper relates most closely to recent work by Kang and Imbens (2016) and Imai et al. (2020), who also study randomized saturation experiments with social interactions under non-compliance. Imai et al. (2020) 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 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. (2020) using an assumption they call "personalized encouragement," the equivalent of our IOR assumption. Both Kang and Imbens (2016) and Imai et al. (2020) 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 assumption of anonymous interactions places more restrictions on the potential outcome functions, we recover "fully structural" causal effects that are not specific to the design of the experiment.

In a recently and closely related paper, Vazquez-Bare (2021) uses instrumental variables

to identify spillovers without relying on a particular experimental design. Vazquez-Bare (2021) focuses on settings with pairs of people, for example roommates or couples, and considers spillovers both in outcomes and take-up. Under one-sided non-compliance and a novel monotonicity restriction, he identifies two causal effects without invoking the IOR assumption: a direct effect for compliers whose partner is untreated, and an indirect effect for untreated individuals whose partner is a complier. This identification result does not extend to groups of more than two people. In larger groups, Vazquez-Bare (2021) identifies average potential outcomes under anonymous interactions without IOR, instead assuming that individuals' potential outcomes are independent of their peers' compliance types. While our results rely on IOR, we do not invoke his latter assumption because in many applied settings a person's potential outcomes may be related to the characteristics of her peers.

Our paper also relates to the applied literature that estimates spillover effects, including "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 spillovers 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., 2021; 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. (2020). Our identification approach also relates to a large literature on random coefficients models, the closest being Wooldridge (2004), Masten and Torgovitsky (2016), and Graham and de Xavier Pinto (2022), as well as methods that identify structural effects using control functions (Altonji and Matzkin, 2005; Imbens and Newey, 2009).

The remainder of the paper is organized as follows. Section 2 details our notation and assumptions, section 3 presents our identification results, and section 4 provides consistent and asymptotically normal estimators of the effects identified in section 3. In section 5 we implement our estimator on data from a well-known labor market experiment, and discuss our findings. In section 6 we present a brief simulation study illustrating the behavior of our estimator. Section 7 concludes. Proofs and additional results 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 spillovers. 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:

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

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 eleven 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 a fixed number m_j of groups are assigned to saturation s_j, and $\sum_{j=1}^{J} m_j = G$. 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 (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. For simplicity we assume that treatment offers in the second stage follow a *Bernoulli design*, in which S_g determines the probability of treatment rather than the number of treatment offers.[2]

[2]With minor modifications, all of our results can be extended to a completely randomized design, in

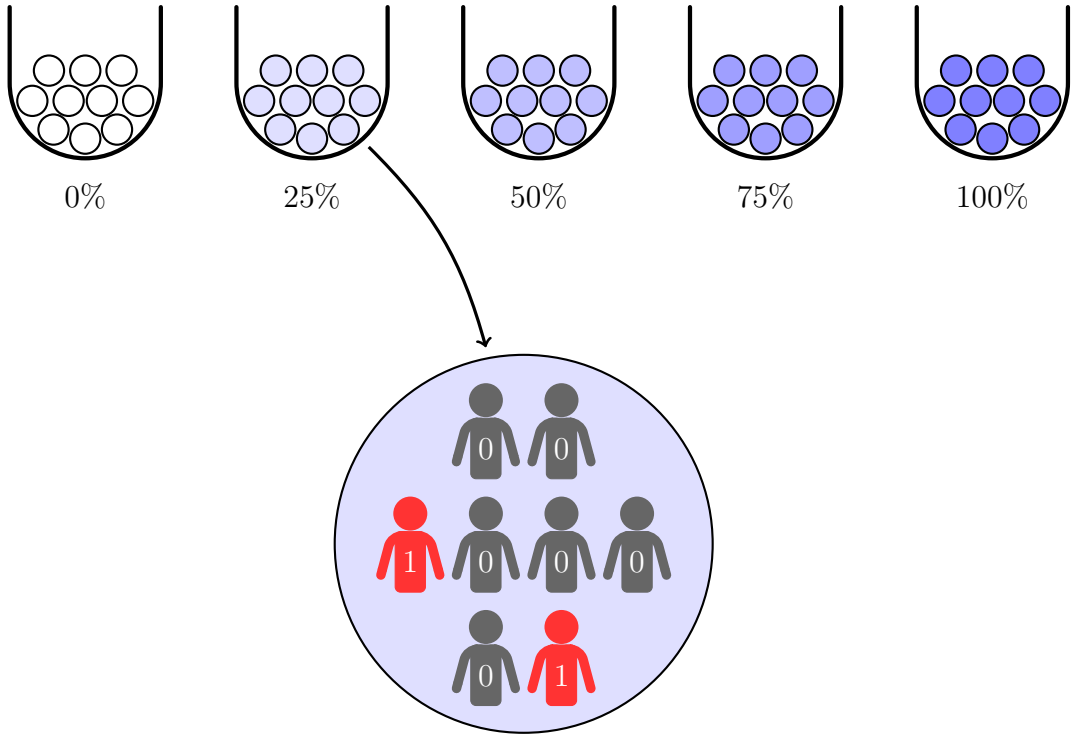

Figure 1: Randomized Saturation Design. In the first stage groups (balls) are randomly assigned to saturations (urns). In the second stage, individuals within a group are randomly assigned treatment offers at the saturation selected in the first stage. The figure zooms in on a group of size eight that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

Assumption 2 (Bernoulli Offers)**.**

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

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 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 subsection 3.3 below, the number of interior saturations in the design will determine the flexibility with which we can model potential outcome functions.

Assumptions 1–2 concern the design of the experiment. Our remaining assumptions, in

which the number of treatment offers made to a given group is fixed conditional on S_g.

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 3 restricts $Y_{ig}(\cdot)$ to depend only on D_{ig} and $\bar{D}_{ig}$ via a random coefficients model.

Assumption 3 (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 3 is the so-called *partial interference* assumption, used widely in the literature on spillover effects. This assumption states that there are no spillovers between people 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 3 states that person (i, g)'s potential outcome is only affected by the treatment take-up 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 3 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,

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

we allow arbitrary statistical dependence between $(D_{ig}, \bar{D}_{ig})$ and $\mathbf{B}_{ig}$. While our assumption of a random coefficients model is not in itself restrictive, our use of a finite basis function expansion is. As explained below in subsection 3.3, the design of a randomized saturation experiment determines the maximum number of basis functions that can be used in practice.

Ideally, our goal would be to identify the average direct and indirect causal effects of the binary treatment D_{ig}. Under Assumption 3, 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

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

where the expectations are taken over all individuals in the population from which our experimental subjects were drawn. Recall that $\bar{D}_{ig}$ excludes person (i,g), ensuring that $\mathrm{DE}(\bar{d})$ 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 holding her own treatment D_{ig} fixed at d, in other words,

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

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 3, this is our object of interest below. For example, if $\mathbf{f}(x)' = [1 \quad x]$ we obtain a linear model of the form

$$Y_{ig} = \alpha_{ig} + \beta_{ig}D_{ig} + \gamma_{ig}\bar{D}_{ig} + \delta_{ig}D_{ig}\bar{D}_{ig}. \tag{5}$$

In this case the direct effect is $\mathrm{DE}(\bar{d}) = \mathbb{E}[\beta_{ig}] + \mathbb{E}[\delta_{ig}]\bar{d}$ while the indirect effects are

$$\mathrm{IE}_0(\bar{d},\Delta) = \Delta \times \mathbb{E}[\gamma_{ig}], \quad \mathrm{IE}_1(\bar{d},\Delta) = \Delta \times \mathbb{E}[\gamma_{ig}+\delta_{ig}].$$

Figure 2 presents a hypothetical example of (5) in a setting with employment displacement effects. Suppose that Y_{ig} is Alice's probability of long-term employment. Both $Y_{ig}(1,\bar{d})$ and $Y_{ig}(0,\bar{d})$ have a negative slope. This means that Alice's probability of long-term employment *decreases* if more of her neighbors obtain job placement services. But since δ_{ig} is

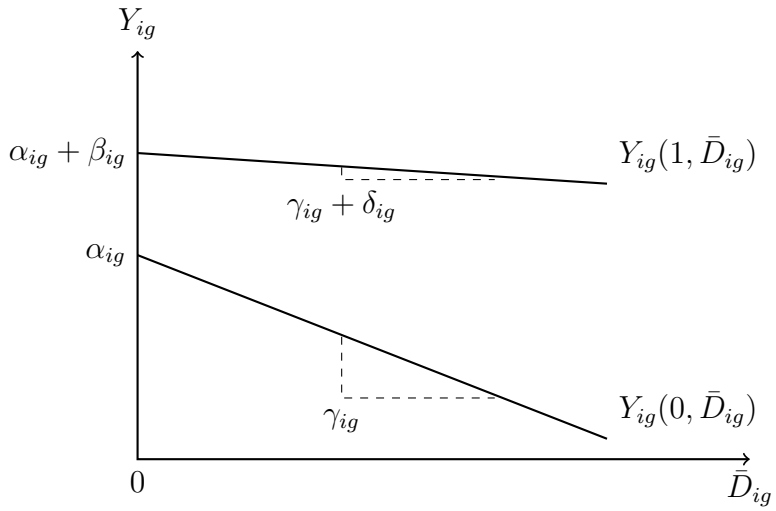

Figure 2: A hypothetical example of the linear potential outcomes model from (5). The slope of the bottom line, γ_{ig}, is the indirect marginal effect when untreated while that of the top line, $\gamma_{ig} + \delta_{ig}$, is the marginal indirect effect when treated. The distance between the two lines is the direct treatment effect.

positive, the spillover is more harmful if Alice is untreated. Alice's direct effect of treatment $Y_{ig}(1, \bar{d}) - Y_{ig}(0, \bar{d})$ is positive for all $\bar{d}$ in this example and increases as $\bar{d}$ does: job placement services are more valuable to Alice when more of her neighbors obtain them. By averaging these effects for everyone in the population, we obtain IE_0, IE_1, and DE.

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

common in practice and simplifies the analysis.[4]

Assumption 4 (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 5 (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 5 is a partial interference assumption: it requires that person (i, g)'s take-up decision is invariant to the realized treatment offers made to people in *different groups*. The second equality of Assumption 5 states that person (i, g)'s take-up decision depends on the realized treatment offers of others in her group only through the fraction $\bar{Z}_{ig}$ of treatment offers made to the others in her group. 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 is invariant to the realized offers made to her peers. In settings where participants do not observe others' offers or the saturation assigned to their group, for example, IOR holds. IOR restricts but does not rule out strategic take-up. For example, it also holds when agents act strategically on their beliefs about others' actions, provided that they are unaware of their peers' offers when making their own take-up decisions. IOR is a strong assumption, but one that has also appeared in the existing literature. Kang and Imbens (2016), for example, employ an equivalent assumption they call "personalized encouragement." And while Imai et al. (2020) derive their so-called "complier average direct effect (CADE)" under a weaker condition than IOR, the CADE is a hybrid of direct and indirect effects unless one is willing to assume that there are no social interactions in take-up. Fortunately, IOR has testable implications: 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 4 and 5), 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.[5] Defining C_{ig} to be the indicator that person (i, g) is a

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

[5] Under one-sided non-compliance, Assumption 4, there are no always-takers.

complier, Assumptions 4–5 imply that

$$D_{ig} = C_{ig} Z_{ig}, \quad \bar{D}_{ig} = \frac{1}{N_g - 1} \sum_{j \neq i} C_{jg} Z_{jg}. \tag{6}$$

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

$$\bar{C}_{ig} = \frac{1}{N_g - 1} \sum_{j \neq i} C_{jg}. \tag{7}$$

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 (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 6 (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 6 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, its counterparts are the "unconfounded type" assumption and the independence of potential outcomes and treatment offers. Assumption 6 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 6 partially embeds IOR by ruling out "selection into compliance." More prosaically, Assumption 6 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 6 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.
[6]

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

3 Identification

3.1 Conditioning on the Share of Compliers

Under Assumption 3, the functional form of the random coefficients model is known. So why not simply use (Z_{ig}, S_g) as instrumental variables for D_{ig} and $\mathbf{f}(\bar{D}_{ig})$? As shown in a number of papers from the literature on random coefficients models (Heckman and Vytlacil, 1998; Wooldridge, 1997, 2003, 2016), two-stage least squares identifies average effects when the causal effect of the instruments on the endogenous regressors is homogeneous. In our setting, however, this result does not apply because the conditional distribution of $\bar{D}_{ig}$ given S_g varies with $(\bar{C}_{ig}, N_g)$, as the following lemma shows.

Lemma 1. *Let $\bar{c}$ be a value in $[0,1]$ such that $(n-1)\bar{c}$ is a non-negative integer. Under Assumptions 1–2 and 4–6 and conditional on $(N_g = n, S_g = s, \boldsymbol{C}_g = \boldsymbol{c}, \bar{C}_{ig} = \bar{c}, Z_{ig} = z)$, $(n-1)\bar{D}_{ig}$ follows a Binomial$((n-1)\bar{c}, s)$ distribution.*

Intuitively, the problem presented by Lemma 1 is as follows. Although S_g is randomly assigned, the variation that it induces in $\bar{D}_{ig}$ is mediated by 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 not identify the effects of interest. To make this problem more concrete, consider the linear potential outcomes model from (5) and let $\boldsymbol{\vartheta}_{\text{IV}}$ be the IV estimand using instruments $(1, Z_{ig}, S_g, Z_{ig}S_g)$. Throughout, we will refer to it as the "naïve IV". In this example $\boldsymbol{\vartheta}_{\text{IV}}$ takes a particularly simple form, as shown in the following lemma.

Lemma 2. *Let $\boldsymbol{\vartheta}_{IV}$ be the IV estimand from a regression of Y_{ig} on $\mathbf{X}_{ig} \equiv (1, D_{ig}, \bar{D}_{ig}, D_{ig}\bar{D}_{ig})'$ with instruments $\boldsymbol{\mathcal{Z}}_{ig} \equiv (1, Z_{ig}, S_g, Z_{ig}S_g)'$, namely*

$$\boldsymbol{\vartheta}_{IV} \equiv \begin{bmatrix} \alpha_{IV} & \beta_{IV} & \gamma_{IV} & \delta_{IV} \end{bmatrix}' = \mathbb{E}\left[\boldsymbol{\mathcal{Z}}_{ig}\mathbf{X}_{ig}'\right]^{-1} \mathbb{E}\left[\boldsymbol{\mathcal{Z}}_{ig} Y_{ig}\right].$$

assuming that $\mathbb{E}[\boldsymbol{\mathcal{Z}}_{ig}\mathbf{X}_{ig}']$ is invertible. Then, under (5) and Assumptions 1–2 and 4–6,

$$\begin{aligned} \alpha_{IV} &= \mathbb{E}\left[\alpha_{ig}\right] & \beta_{IV} &= \mathbb{E}\left[\beta_{ig} | C_{ig} = 1\right] \\ \gamma_{IV} &= \mathbb{E}\left[\gamma_{ig}\right] + \frac{Cov(\bar{C}_{ig}, \gamma_{ig})}{\mathbb{E}(\bar{C}_{ig})} & \delta_{IV} &= \mathbb{E}\left[\delta_{ig} | C_{ig} = 1\right] + \frac{Cov(\bar{C}_{ig}, \delta_{ig} | C_{ig} = 1)}{\mathbb{E}(\bar{C}_{ig} | C_{ig} = 1)}. \end{aligned}$$

As we see from Lemma 2, IV identifies the population average of α_{ig}, along with the population average of β_{ig} for the subset of individuals who select into treatment. Neither of these, however, is itself a causal effect. In general, IV recovers neither direct nor indirect

causal effects for any well-defined group of individuals. Specializing (4) to the linear model from (5) gives $\text{IE}_0(\bar{d}, \Delta) = \mathbb{E}[\gamma_{ig}]\Delta$. In other words, $\mathbb{E}[\gamma_{ig}]$ is an average *spillover*. Lemma 2 shows that IV fails to identify this quantity unless the individual-specific spillovers γ_{ig} are uncorrelated with the share of compliers $\bar{C}_{ig}$. 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 spillovers are more intense in such cities, IV will not recover the average indirect effect. A similar problem hampers the interpretation of δ_{IV}. Under (5) the average direct effect for compliers, as a function of $\bar{d}$, is given by $\mathbb{E}[\beta_{ig}|C_{ig} = 1] + \mathbb{E}[\delta_{ig}|C_{ig} = 1]\bar{d}$. While IV identifies the intercept of this function, it only identifies the slope if δ_{ig} is uncorrelated with $\bar{C}_{ig}$ for compliers.

As this example illustrates, identifying direct and indirect causal effects requires us to correct for possible dependence between individual-specific coefficients and group-level take-up that arises from the first-stage relationship in Lemma 1. The key to our approach, as shown in the following theorem, is to condition on $\bar{C}_{ig}$ and N_g.

Theorem 1. *Under Assumptions 1–2 and 4–6,* $(S_g, Z_{ig}, \bar{D}_{ig}) \perp\!\!\!\perp (\mathbf{B}_{ig}, C_{ig}) | (\bar{C}_{ig}, N_g)$.

Theorem 1 implies that conditioning on $(\bar{C}_{ig}, N_g)$ is sufficient to break any dependence between $\mathbf{f}(\bar{D}_{ig})$ and $(\mathbf{B}_{ig}, C_{ig})$ that may be present. 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 6, 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. In effect, our identification approach is a combination of instrumental variables and control function methods. First $(\bar{C}_{ig}, N_g)$ serves as a control function for the endogenous regressor $\bar{D}_{ig}$, similar to Masten and Torgovitsky (2016). Second, Z_{ig} serves as an instrument for D_{ig}, because this regressor remains endogenous even conditional on $(\bar{C}_{ig}, N_g)$.

3.2 An Inverse-Weighting Instrumental Variables Approach

Before stating our identification results, we require some additional notation and one further assumption. Define the vector $\mathbf{W}_{ig}$ and matrix-valued functions $\mathbf{Q}, \mathbf{Q}_0, \mathbf{Q}_1$ as follows:

$$\mathbf{Q}(\bar{c}, n) \equiv \mathbb{E}\left[\mathbf{W}_{ig}\mathbf{W}_{ig}' | \bar{C}_{ig} = \bar{c}, N_g = n\right], \quad \mathbf{W}_{ig} \equiv \begin{bmatrix} 1 & Z_{ig} \end{bmatrix}' \otimes \mathbf{f}(\bar{D}_{ig}) \tag{8}$$

$$\mathbf{Q}_0(\bar{c}, n) \equiv \mathbb{E}\left[(1 - Z_{ig})\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})' | \bar{C}_{ig} = \bar{c}, N_g = n\right] \tag{9}$$

$$\mathbf{Q}_1(\bar{c}, n) \equiv \mathbb{E}\left[Z_{ig}\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})' | \bar{C}_{ig} = \bar{c}, N_g = n\right]. \tag{10}$$

These functions depend only on the distribution of $\bar{D}_{ig}|(Z_{ig}, \bar{C}_{ig}, N_g)$, which can be calculated from Lemma 1, and the distribution of $Z_{ig}|(\bar{C}_{ig}, N_g)$, which coincides with its unconditional distribution by Lemma A.2. As such, under our assumptions $\mathbf{Q}, \mathbf{Q}_0, \mathbf{Q}_1$ are *completely determined* by the design of the randomized saturation experiment. We can always calculate them by simulating the experimental design. Depending on the choice of $\mathbf{f}$, analytical expressions may also be available, as shown in subsection 3.3 for the linear potential outcomes model from (5).

We use $\mathbf{Q}, \mathbf{Q}_0, \mathbf{Q}_1$ to construct valid instrumental variables by inverse-weighting. Rather than using the randomly assigned saturation S_g as a source of instruments for $\mathbf{f}(\bar{D}_{ig})$ we *transform* the endogenous regressors $\mathbf{W}_{ig}$ into a set of exogenous instruments using $\mathbf{Q}_0(\bar{C}_{ig}, N_g)^{-1}$ and $\mathbf{Q}_1(\bar{C}_{ig}, N_g)^{-1}$, in particular

$$\boldsymbol{\mathcal{Z}}_{ig}^{W} \equiv \mathbf{Q}(\bar{C}_{ig}, N_g)^{-1}\mathbf{W}_{ig} \tag{11}$$

$$\boldsymbol{\mathcal{Z}}_{ig}^{0} \equiv \mathbf{Q}_0(\bar{C}_{ig}, N_g)^{-1}\mathbf{f}(\bar{D}_{ig}) \tag{12}$$

$$\boldsymbol{\mathcal{Z}}_{ig}^{1} \equiv \mathbf{Q}_1(\bar{C}_{ig}, N_g)^{-1}\mathbf{f}(\bar{D}_{ig}). \tag{13}$$

Constructing these instruments requires us to evaluate $\mathbf{Q}_0$ and $\mathbf{Q}_1$ at $(\bar{C}_{ig}, N_g)$.[7] The group size N_g is observed, while the share of compliers $\bar{C}_{ig}$ is not. In large groups, $\bar{C}_{ig}$ can be precisely estimated by calculating the rate of treatment take-up among the neighbors of (i, g) who are offered treatment. We formally establish the rates of convergence of IV estimators that plug-in a proxy for $\bar{C}_{ig}$ in Section 4. For the remainder of this section, however, we consider identification *conditional* on knowledge of $\bar{C}_{ig}$.

To understand the intuition behind $\boldsymbol{\mathcal{Z}}_{ig}^{W}$, $\boldsymbol{\mathcal{Z}}_{ig}^{0}$, and $\boldsymbol{\mathcal{Z}}_{ig}^{1}$, consider the linear potential

[7] The function $\mathbf{Q}$ can be constructed from $\mathbf{Q}_0$ and $\mathbf{Q}_1$, as shown in Equation 14.

outcomes example from (5) above. Here we have $\mathbf{f}(x) = (1, x)'$ and thus

$$\mathbf{Q}_z(\bar{C}_{ig}, N_g) = \mathbb{P}(Z_{ig} = z)\mathbb{E}\left[\begin{pmatrix} 1 & \bar{D}_{ig} \\ \bar{D}_{ig} & \bar{D}_{ig}^2 \end{pmatrix} \middle| \bar{C}_{ig}, N_g, Z_{ig} = z\right], \quad z \in \{0, 1\}$$

using the fact that $Z_{ig} \perp\!\!\!\perp (\bar{C}_{ig}, N_g)$ by Lemma A.2. It follows after a few steps of algebra that

$$\mathbf{Q}_z(\bar{C}_{ig}, N_g)^{-1}\mathbf{f}(\bar{D}_{ig}) = \frac{1}{\mathbb{P}(Z_{ig} = z)} \begin{bmatrix} \dfrac{\mathbb{E}(\bar{D}_{ig}^2|\bar{C}_{ig}, N_g, Z_{ig} = z) - \bar{D}_{ig}\mathbb{E}(\bar{D}_{ig}|\bar{C}_{ig}, N_g, Z_{ig} = z)}{\mathrm{Var}(\bar{D}_{ig}|\bar{C}_{ig}, N_g, Z_{ig} = z)} \\ \\ \dfrac{\bar{D}_{ig} - \mathbb{E}(\bar{D}_{ig}|\bar{C}_{ig}, N_g, Z_{ig} = z)}{\mathrm{Var}(\bar{D}_{ig}|\bar{C}_{ig}, N_g, Z_{ig} = z)} \end{bmatrix}.$$

While $\bar{D}_{ig}$ is endogenous, the scaled difference between $\bar{D}_{ig}$ and its conditional expectation is a valid instrument under the linear potential outcomes model. Intuitively, this transformation *adjusts* for the first-stage heterogeneity discussed in subsection 3.1: after controlling for differences in $(\bar{C}_{ig}, N_g)$, the remaining variation in $\bar{D}_{ig}$ arises only from the experimentally–assigned saturations. Thus, rather than using S_g as an instrument directly, we use it indirectly to generate variation in $\bar{D}_{ig}$ given $(\bar{C}_{ig}, N_g)$. The final ingredient that we require is a rank condition.

Assumption 7 (Rank Condition)**.**

(i) $0 < \mathbb{E}(C_{ig}) < 1$

(ii) $\mathbf{Q}(\bar{c}, n)$ *is invertible at every point* $(\bar{c}, n)$ *in the support of* $(\bar{C}_{ig}, N_g)$*.*

Part (i) of Assumption 7 asserts that there is at least some degree of non-compliance with the experimental treatment offers, $\mathbb{E}(C_{ig}) < 1$, and that the population contains at least some compliers, $\mathbb{E}(C_{ig}) > 0$. Part (ii) requires that the matrix-valued function $\mathbf{Q}$ defined in (8) is full rank when evaluated at any share of compliers $\bar{c}$ and group size n that occur in the population. Assumption 7 does not explicitly restrict $\mathbf{Q}_0$ or $\mathbf{Q}_1$. By the linearity of conditional expectation, however,

$$\mathbf{Q}(\bar{c}, n) = \begin{bmatrix} \mathbf{Q}_0(\bar{c}, n) + \mathbf{Q}_1(\bar{c}, n) & \mathbf{Q}_1(\bar{c}, n) \\ \mathbf{Q}_1(\bar{c}, n) & \mathbf{Q}_1(\bar{c}, n) \end{bmatrix} \tag{14}$$

so Assumption 7(ii) could equivalently be stated in terms of $\mathbf{Q}_0$ and $\mathbf{Q}_1$.

Lemma 3. *$\mathbf{Q}(\bar{c}, n)$ is invertible iff $\mathbf{Q}_0(\bar{c}, n)$ and $\mathbf{Q}_1(\bar{c}, n)$ are both invertible, in which case*

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

We discuss low-level conditions for the invertibility of $(\mathbf{Q}_0, \mathbf{Q}_1)$, and hence $\mathbf{Q}$, below in subsection 3.3. Having assumed the necessary rank condition, we can now state our main identification results. The following theorem shows how $\mathbf{Q}_0(\bar{C}_{ig}, N_g)$ and $\mathbf{Q}_1(\bar{C}_{ig}, N_g)$ can be used to construct instrumental variables that identify average values of the random coefficients for well-defined groups of individuals.

Theorem 2. *Let $\boldsymbol{\mathcal{Z}}^W_{ig}$, $\boldsymbol{\mathcal{Z}}^0_{ig}$, and $\boldsymbol{\mathcal{Z}}^1_{ig}$ be as defined in (11)–(13). Then, under Assumptions 3–5 and 7 and assuming that $(Z_{ig}, \bar{D}_{ig}) \perp\!\!\!\perp (\mathbf{B}_{ig}, C_{ig}) | (\bar{C}_{ig}, N_g)$, we have*

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

(ii) $$\mathbb{E}\left[\boldsymbol{\psi}_{ig} | C_{ig} = 1\right] = \mathbb{E}\left[\boldsymbol{\mathcal{Z}}^1_{ig} \left\{D_{ig} \mathbf{f}(\bar{D}_{ig})\right\}'\right]^{-1} \mathbb{E}\left[\boldsymbol{\mathcal{Z}}^1_{ig} \left\{D_{ig} Y_{ig}\right\}\right]$$

(iii) $$\mathbb{E}\left[\boldsymbol{\theta}_{ig} | C_{ig} = 0\right] = \mathbb{E}\left[\boldsymbol{\mathcal{Z}}^1_{ig} \left\{Z_{ig}(1 - D_{ig}) \mathbf{f}(\bar{D}_{ig})\right\}'\right]^{-1} \mathbb{E}\left[\boldsymbol{\mathcal{Z}}^1_{ig} \left\{Z_{ig}(1 - D_{ig}) Y_{ig}\right\}\right], \text{ and}$$

(iv) $$\mathbb{E}\left[\boldsymbol{\theta}_{ig}\right] = \mathbb{E}\left[\boldsymbol{\mathcal{Z}}^0_{ig} \left\{(1 - Z_{ig}) \mathbf{f}(\bar{D}_{ig})\right\}'\right]^{-1} \mathbb{E}\left[\boldsymbol{\mathcal{Z}}^0_{ig} \left\{(1 - Z_{ig}) Y_{ig}\right\}\right].$$

The first part of Theorem 2 identifies the average effects that the naïve IV approach from Lemma 2 in general fails to. Parts (ii) and (iii) use a similar approach to obtain moment equations for the average value of $\boldsymbol{\psi}_{ig}$ for compliers and $\boldsymbol{\theta}_{ig}$ for never-takers. Given part (i), part (iv) is technically redundant, but it is convenient to have an expression for $\mathbb{E}(\boldsymbol{\theta}_{ig})$ in isolation. As discussed below in Section 3.3, having sufficient variation in the saturations is crucial for part (ii) of Assumption 7.

Notice that Theorem 2 does not explicitly invoke the randomized saturation design, Assumptions 1–2, or the exclusion restriction, Assumption 6. Using this result for identification, however, requires us to satisfy $(Z_{ig}, \bar{D}_{ig}) \perp\!\!\!\perp (\mathbf{B}_{ig}, C_{ig}) | (\bar{C}_{ig}, N_g)$. As shown in Theorem 1 above, the randomized saturation design and exclusion restriction are sufficient for this condition to hold under one-sided non-compliance and IOR, Assumptions 4 and 5. The following result catalogues the full set of causal effects that are identified under our assumptions.

Theorem 3. *Given knowledge of $\bar{C}_{ig}$ the following are identified under Assumptions 1–7:*

(i) $IE_0(\bar{d}, \Delta) \equiv \mathbb{E}[Y_{ig}(0, \bar{d} + \Delta) - Y_{ig}(0, \bar{d})]$,

(ii) $DE(\bar{d}|D_{ig}=1) \equiv \mathbb{E}\left[Y_{ig}(1,\bar{d}) - Y_{ig}(0,\bar{d})|D_{ig}=1\right]$,

(iii) $IE_0(\bar{d},\Delta|D_{ig}=1) \equiv \mathbb{E}[Y_{ig}(0,\bar{d}+\Delta) - Y_{ig}(0,\bar{d})|D_{ig}=1]$,

(iv) $IE_1(\bar{d},\Delta|D_{ig}=1) \equiv \mathbb{E}[Y_{ig}(1,\bar{d}+\Delta) - Y_{ig}(1,\bar{d})|D_{ig}=1]$,

(v) $IE_0(\bar{d},\Delta|C_{ig}=0) \equiv \mathbb{E}[Y_{ig}(0,\bar{d}+\Delta) - Y_{ig}(0,\bar{d})|C_{ig}=0]$,

Part (i) of Theorem 3 is a population average indirect treatment effect, as defined in (4) above. It measures the causal impact of increasing the treatment take-up rate among Alice's neighbors from $\bar{d}$ to $(\bar{d}+\Delta)$ when Alice's own treatment is held fixed at zero. In the Crépon et al. (2013) experiment discussed in our empirical example below, this corresponds to the average labor market displacement effect. Whereas part (i) is an average treatment effect, parts (ii)–(iv) are the effects of treatment-on-the-treated.[8] Part (ii) gives the direct effect of treating Alice while holding the treatment take-up rate of her neighbors fixed at $\bar{d}$, while (iii) and (iv) give the indirect effect of increasing her neighbors' treatment take-up from $\bar{d}$ to $\bar{d}+\Delta$ while holding Alice's treatment fixed at either zero, part (iii), or one, part (iv). Part (v) is a treatment-on-the-untreated version of Equation 4: it gives the indirect effect for never-takers, holding their treatment fixed at zero. While we identify the full set of direct and indirect effects for the treated sub-population, we only identify a subset of these effects for other groups. By definition, never-takers cannot be observed with $D_{ig}=1$. As such, we cannot identify direct treatment effects for this group or indirect treatment effects when D_{ig} is held fixed at one. This in turn implies that we cannot identify the average direct effect for the population as a whole, $\text{DE}(\bar{d})$, or the average indirect effect when D_{ig} is held fixed at one, $\text{IE}_1(\bar{d},\Delta)$.

3.3 Identification in Practice: An Example

Given that $\mathbf{Q}_0$ and $\mathbf{Q}_1$ are completely determined by the experimental design, we can directly check part (ii) of Assumption 7 for any choice of basis functions $\mathbf{f}$ and probability distribution over saturations. Consider again the linear potential outcomes model from (5). In this example $\mathbf{f}(x) = (1,x)'$ and thus,

$$\mathbf{Q}_0(\bar{c},n) = \begin{bmatrix} \mathbb{E}\{1-S_g\} & \bar{c}\,\mathbb{E}\{S_g(1-S_g)\} \\ \bar{c}\,\mathbb{E}\{S_g(1-S_g)\} & \bar{c}^2\,\mathbb{E}\{S_g^2(1-S_g)\} + \frac{\bar{c}}{n-1}\mathbb{E}\{S_g(1-S_g)^2\} \end{bmatrix} \tag{15}$$

$$\mathbf{Q}_1(\bar{c},n) = \begin{bmatrix} \mathbb{E}\{S_g\} & \bar{c}\,\mathbb{E}\{S_g^2\} \\ \bar{c}\,\mathbb{E}\{S_g^2\} & \bar{c}^2\,\mathbb{E}\{S_g^3\} + \frac{\bar{c}}{n-1}\mathbb{E}\{S_g^2(1-S_g)\}. \end{bmatrix} \tag{16}$$

[8] Because this is a setting with one-sided non-compliance, any participant with $D_{ig}=1$ must be a complier.

by Bayes' Theorem, the Law of Total Probability, and Lemmas 1 and A.2. Suppose first that there is a single saturation s. Then (15) and (16) simplify to yield

$$|\mathbf{Q}_0(\bar{c}, n)| = \frac{\bar{c}s(1-s)^3}{n-1}, \quad |\mathbf{Q}_1(\bar{c}, n)| = \frac{\bar{c}s^3(1-s)}{n-1}.$$

so that $\mathbf{Q}_0(\bar{c}, n)$ and $\mathbf{Q}_1(\bar{c}, n)$ are both invertible for any n and all $\bar{c}$ greater than zero provided that $0 < s < 1$. The identifying power of this "degenerate" randomized saturation design, however, is weak: $\mathbf{Q}_0, \mathbf{Q}_1$ are arbitrarily close to being singular for any $\bar{c}$ if n is sufficiently large. Consider next a so-called "cluster randomized" experiment in which there are two saturations, 0 and 1, and $\mathbb{P}(S_g = 1) = p$. Calculating the expectations in (15) and (16),

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

In this case neither $\mathbf{Q}_0$ nor $\mathbf{Q}_1$ is invertible for *any* values of n and $\bar{c}$. Finally, consider a design with two distinct, equally likely saturations $s_L < s_H$. For this design, straightforward but tedious algebra gives

$$|\mathbf{Q}_0(\bar{c}, n)| = \frac{\bar{c}^2}{4}(1-s_L)(1-s_H)(s_H-s_L)^2 + \frac{\bar{c}\left[(1-s_L)+(1-s_H)\right]\left[s_L(1-s_L)^2 + s_H(1-s_H)^2\right]}{4(n-1)}$$
$$|\mathbf{Q}_1(\bar{c}, n)| = \frac{\bar{c}^2}{4}s_L s_H(s_H-s_L)^2 + \frac{\bar{c}\left(s_L+s_H\right)\left[s_L^2(1-s_L) + s_H^2(1-s_H)\right]}{4(n-1)}.$$

So long as neither s_L nor s_H equals zero or one, both terms in each expression are strictly positive for any $\bar{c} > 0$, so that $\mathbf{Q}_0$ and $\mathbf{Q}_1$ are invertible. Moreover, in contrast to the single saturation design discussed above, this design does not suffer from a weak identification problem. While the second term in each of the preceding equalities vanishes for large n, the first term does not. Thus, two interior saturations are sufficient to strongly identify the linear potential outcomes model.[9]

As the three preceding examples show, two distinct sources of experimental variation determine the rank of $\mathbf{Q}_0(\bar{c}, n)$ and $\mathbf{Q}_1(\bar{c}, n)$: "between" saturation variation, and "within" saturation variation. Our first example lacks "between" variation because each group is assigned the same saturation, $S_g = s$. Yet even with a single saturation, there is still "within" variation under Assumption 2, because the number of offers made to a given group is random. This "within" variation, however, is negligible when n is large. In our second

[9] In general, sufficient conditions for Assumption 7(ii) will depend on the specific choice of basis functions $\mathbf{f}$. For large n, however, a necessary condition is that the design contains at least as many distinct interior saturations as there are elements in $\mathbf{f}$. For details, see Appendix D.

example, the cluster randomized experiment, the situation is 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, with two saturations $0 < s_L < s_H < 1$, features sufficient "between" variation to identify the effects of interest even when n is so large that "within" variation becomes negligible.

4 Estimation and Inference

If $\bar{C}_{ig}$ were observed, a handful of just-identified IV regressions would suffice to estimate the causal effects from Theorem 3. While $\bar{C}_{ig}$ is unobserved in practice, fortunately we can estimate it under one-sided non-compliance by comparing treatment take-up to the share of treatment offers, i.e.

$$\widehat{C}_{ig} \equiv \begin{cases} \bar{D}_{ig}/\bar{Z}_{ig}, & \text{if } \bar{Z}_{ig} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

where we arbitrarily define $\widehat{C}_{ig} = 0$ if none of (i, g)'s neighbors are offered treatment.[10] In this section we use (17) to derive feasible, consistent, and asymptotically normal estimators of the direct and indirect causal effects identified in section 3. For simplicity, we assume throughout that the random saturation S_g is bounded below by $\underline{s} > 0$. Because we cannot estimate $\bar{C}_{ig}$ when $S_g = 0$, experiments that include a 0% saturation require a slightly different approach. We explain these differences in Appendix C.

In the interest of brevity, we introduce shorthand notation and high-level regularity conditions that apply to all four of our sample analogue estimators. These take the form

$$\widehat{\boldsymbol{\vartheta}} \equiv \left(\sum_{g=1}^{G} \sum_{i=1}^{N_g} \widehat{\boldsymbol{\mathcal{Z}}}_{ig} \boldsymbol{X}'_{ig} \right)^{-1} \left(\sum_{g=1}^{G} \sum_{i=1}^{N_g} \widehat{\boldsymbol{\mathcal{Z}}}_{ig} Y_{ig} \right), \quad \widehat{\boldsymbol{\mathcal{Z}}}_{ig} \equiv \mathbf{R}(\widehat{C}_{ig}, N_g)^{+} \boldsymbol{W}_{ig} \tag{18}$$

where Y_{ig} is the outcome variable from Assumption 3, and $\mathbf{M}^{+}$ denotes the Moore-Penrose inverse of a square matrix $\mathbf{M}$. Table 1 gives the definitions of $\boldsymbol{X}_{ig}, \mathbf{R}$, and $\boldsymbol{W}_{ig}$ corresponding to each part of Theorem 2. The "estimated" instrument $\widehat{\boldsymbol{\mathcal{Z}}}_{ig}$ is a stand-in for the unobserved "true" instrument $\boldsymbol{\mathcal{Z}}_{ig} \equiv \mathbf{R}(\bar{C}_{ig}, N_g)^{-1} \boldsymbol{W}_{ig}$. While $\mathbf{R}(\bar{C}_{ig}, N_g)$ is invertible under Assumption 7, $\mathbf{R}(\widehat{C}_{ig}, N_g)$ may not be so, since $\widehat{C}_{ig}$ could fall outside the support set of $\bar{C}_{ig}$ or even equal zero. For this reason we define $\widehat{\boldsymbol{\mathcal{Z}}}_{ig}$ using the Moore-Penrose inverse, which always

[10] Under Assumption 2 it is possible, although unlikely, that $\bar{Z}_{ig}$ could be zero even if $S_g > 0$.

	$\boldsymbol{X}_{ig}$	$\mathbf{R}$	$\boldsymbol{W}_{ig}$
(i)	$\begin{bmatrix} 1 \\ D_{ig} \end{bmatrix} \otimes \mathbf{f}(\bar{D}_{ig})$	$\mathbf{Q}$	$\begin{bmatrix} 1 \\ Z_{ig} \end{bmatrix} \otimes \mathbf{f}(\bar{D}_{ig})$
(ii)	$\mathbf{f}(\bar{D}_{ig})$	$\mathbf{Q}_1$	$\mathbf{f}(\bar{D}_{ig})D_{ig}$
(iii)	$\mathbf{f}(\bar{D}_{ig})$	$\mathbf{Q}_1$	$\mathbf{f}(\bar{D}_{ig})Z_{ig}(1-D_{ig})$
(iv)	$\mathbf{f}(\bar{D}_{ig})$	$\mathbf{Q}_0$	$\mathbf{f}(\bar{D}_{ig})(1-Z_{ig})$

Table 1: This table defines the shorthand from (18) for the four sample analogue estimators corresponding to the parts of Theorem 2. In each part, the vector of regressors is $\boldsymbol{X}_{ig}$, the true instrument vector is $\boldsymbol{\mathcal{Z}}_{ig} \equiv \mathbf{R}(\bar{C}_{ig}, N_g)^{-1}\boldsymbol{W}_{ig}$, and the estimated instrument vector is $\widehat{\boldsymbol{\mathcal{Z}}}_{ig} \equiv \mathbf{R}(\widehat{C}_{ig}, N_g)^{+}\boldsymbol{W}_{ig}$, where $\mathbf{M}^{+}$ denotes the Moore-Penrose inverse of a square matrix $\mathbf{M}$, and $\widehat{C}_{ig}$ is as defined in (17). The functions $\mathbf{Q}, \mathbf{Q}_0, \mathbf{Q}_1$ are as defined in (8)–(10).

exists and coincides with the ordinary matrix inverse when $\mathbf{R}(\widehat{C}_{ig}, N_g)$ is indeed invertible.

As G grows, so does the number of unknown values $\bar{C}_{ig}$ that we must estimate to construct the instrument vectors $\widehat{\boldsymbol{\mathcal{Z}}}_{ig}$.[11] For this reason, we consider an asymptotic sequence in which the minimum group size $\underline{n}$ grows along with the number of groups G. Under appropriate assumptions, this implies that the limit behavior of $\widehat{\boldsymbol{\vartheta}}$, which we refer to as the "random saturation IV" (RS-IV), coincides with that of the infeasible estimator that uses the true instrument vector $\boldsymbol{\mathcal{Z}}_{ig}$ instead of its estimate $\widehat{\boldsymbol{\mathcal{Z}}}_{ig}$.

Like Baird et al. (2018), we take an infinite population approach to inference, assuming that the researcher observes a random sample of size G from a population of groups. Unlike Baird et al. (2018), we allow these groups to differ in size. Upon drawing a group g from the population, we observe the group-level random variables (S_g, N_g) along with the individual-level random variables (Y_{ig}, D_{ig}, Z_{ig}) for each member of the group: $1 \leq i \leq N_g$. We further assume that observations are identically distributed, but not independent, within groups.[12]

Groups are only observed *as a unit*: either everyone from the group appears in the sample or no one does. For this reason, some care is needed in defining random variables to represent our sampling procedure and expectations to represent the population averages that define our causal effects of interest. The expectations in Theorems 2–3 are averages that give equal weight to each individual in the population, or sub-population if we condition

[11] While $\bar{C}_{ig}$ can vary across individuals in the same group, it takes on at most two distinct 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 $2G$.

[12] The assumption that observations are identically distributed within group amounts to stipulating that the indices $1 \leq i \leq N_g$ are assigned at random.

on C_{ig}. Analogously, the estimator in (18) is an average that gives equal weight to each individual in the sample. Both of these are precisely what we want, as our goal is to identify and estimate average causal effects for individuals. Under iid sampling of groups, however, $(Y_{ig}, D_{ig}, Z_{ig}, \bar{D}_{ig})$ represent a single person chosen at random from a randomly-selected group. If all groups were the same size, this would be equivalent to choosing a person uniformly at random from the population of *individuals*. When groups vary in size, however, the equivalence no longer holds.[13] This creates the possibility for ambiguity when taking the expectation of an individual-level random variable, such as Y_{ig}, without conditioning on group size.

Fortunately this is only a question of defining appropriate notation. Our group sampling procedure unambiguously gives equal weight to each individual in the population because we observe not isolated individuals but whole groups. While small groups are just as likely to be drawn as large groups, large groups make a greater contribution to the sample averages from (18) because they contain more people.[14] The question is merely how to represent this mathematically. Let $\rho_g \equiv N_g/\mathbb{E}(N_g)$ denote the *relative size* of group g. We write $\mathbb{E}[Y_{ig}]$ to denote the average that gives equal weight to groups—choosing one person at random from a randomly-chosen group—and $\mathbb{E}[\rho_g Y_{ig}]$ to denote the average that gives equal weight to individuals—observing an entire group chosen at random. It is the latter expectation that appears in our results below, as it denotes the population equivalent of the double sums from (18). While this is a slight abuse of notation, expectations from section 3 that involve individual-level random variables but do not condition on group size should be interpreted as (implicitly) weighting by relative group size. Using the notation and sampling scheme defined above, we now state high-level sufficient conditions for the consistency of $\widehat{\boldsymbol{\vartheta}}$ in (18).

Theorem 4. *Let $\rho_g \equiv N_g/\mathbb{E}(N_g)$ and suppose that*

(i) we observe a random sample of G groups, where observations within a given group are identically distributed although not necessarily independent,

(ii) $Y_{ig} = \boldsymbol{X}_{ig}'\boldsymbol{\vartheta} + U_{ig}$ *for* $1 \leq g \leq G$, $1 \leq i \leq N_g$,

(iii) $\mathbb{E}\left(\rho_g \boldsymbol{\mathcal{Z}}_{ig} U_{ig}\right) = \mathbf{0}$ *and* $\mathbb{E}\left(\rho_g \boldsymbol{\mathcal{Z}}_{ig} \boldsymbol{X}_{ig}'\right) = \mathbb{I}$,

[13] Consider a population of 100 groups, half of which have 5 members and the rest of which have 15 members so that 250 of the 1000 people in the population belong to a small group and the remaining 750 belong to a large group. Suppose first that we choose a single group at random and then a single person within the selected group. Then someone from a small group has probability 1/500 of being selected while someone from a large group has probability 1/1500 of being selected.

[14] Continuing from the example in the preceding footnote: suppose we randomly sample 10 groups and observe *everyone* in them. Then, on average, our sample will contain 5 small groups and 5 large groups. While the total sample size is random, on average we will observe 100 people, of whom 25 come from small groups and the rest from large groups, matching the shares of each kind of person in the population.

(iv) $\mathbb{E}\left[\rho_g^2 ||\boldsymbol{\mathcal{Z}}_{ig}\boldsymbol{X}'_{ig}||^2\right] = o(G)$,

(v) $\mathbb{E}\left[\rho_g^2 ||\boldsymbol{\mathcal{Z}}_{ig}U_{ig}||^2\right] = o(G)$,

(vi) $||\sum_{g=1}^{G}\frac{1}{N_g}\sum_{i=1}^{N_g}\rho_g(\widehat{\boldsymbol{\mathcal{Z}}}_{ig} - \boldsymbol{\mathcal{Z}}_{ig})\boldsymbol{X}'_{ig}|| = o_{\mathbb{P}}(G)$, *and*

(vii) $||\sum_{g=1}^{G}\frac{1}{N_g}\sum_{i=1}^{N_g}\rho_g(\widehat{\boldsymbol{\mathcal{Z}}}_{ig} - \boldsymbol{\mathcal{Z}}_{ig})U_{ig}|| = o_{\mathbb{P}}(G)$.

Then $\widehat{\boldsymbol{\vartheta}}$, *defined in* (18), *is consistent for* $\boldsymbol{\vartheta}$ *as* $G \to \infty$.

Condition (i) of Theorem 4 simply restates our group sampling assumption. Conditions (ii) and (iii) hold under the assumptions of Theorem 2, as shown in the proof of that result: for each average effect $\boldsymbol{\vartheta}$ from the theorem, we can define an appropriate error term U_{ig}, vector of regressors $\boldsymbol{X}_{ig}$, and vector of instruments $\boldsymbol{\mathcal{Z}}_{ig}$ such that $Y_{ig} = \boldsymbol{X}'_{ig}\boldsymbol{\vartheta} + U_{ig}$ where $\boldsymbol{\mathcal{Z}}_{ig}$ is an exogenous and relevant instrument. Moreover, for each part of Theorem 2, $\mathbb{E}(\rho_g\boldsymbol{\mathcal{Z}}_{ig}\boldsymbol{X}'_{ig})$ equals the identity matrix.[15,16] Conditions (iv) and (v) of Theorem 4 would be implied by requiring that the second moments of $\rho_g\boldsymbol{\mathcal{Z}}_{ig}\mathbf{X}'_{ig}$ and $\rho_g\boldsymbol{\mathcal{Z}}_{ig}U_{ig}$ exist and are bounded. We choose to state these conditions in a slightly weaker form because the distribution of ρ_g necessarily changes with G if we consider an asymptotic sequence in which the minimum group size $\underline{n}$ increases with the number of groups, as we will assume below. Requiring the relevant expectations to be $o(G)$ in principle allows the variance of relative group size ρ_g to grow along with the number of groups, provided that it does not grow too quickly. Conditions (i)–(v) together are sufficient for the consistency of

$$\widetilde{\boldsymbol{\vartheta}} \equiv \left(\sum_{g=1}^{G}\sum_{i=1}^{N_g}\boldsymbol{\mathcal{Z}}_{ig}\boldsymbol{X}'_{ig}\right)^{-1}\left(\sum_{g=1}^{G}\sum_{i=1}^{N_g}\boldsymbol{\mathcal{Z}}_{ig}Y_{ig}\right), \tag{19}$$

an infeasible estimator that uses the true instrument vector $\boldsymbol{\mathcal{Z}}_{ig}$ instead of its estimate $\widehat{\boldsymbol{\mathcal{Z}}}_{ig}$. The final two conditions of Theorem 4 assume that $\widehat{\boldsymbol{\mathcal{Z}}}_{ig}$ is a sufficiently accurate estimator of $\boldsymbol{\mathcal{Z}}_{ig}$ to ensure that $\widehat{\boldsymbol{\vartheta}} = \widetilde{\boldsymbol{\vartheta}} + o_{\mathbb{P}}(1)$. In the setting we consider here, this will require a condition on how quickly the minimum group size $\underline{n}$ grows relative to G, as we discuss in detail below. Strengthening conditions (v) and (vii) and adding one further assumption implies that $\widehat{\boldsymbol{\vartheta}}$ is asymptotically normal.

Theorem 5. *Suppose that*

[15] For effects that condition on $C_{ig} = c$, e.g. those from parts (ii) and (iii) of Theorem 2, the appropriate definition of ρ_g becomes $N_g\mathbb{E}[\mathbb{1}(C_{ig} = c)]/\mathbb{E}[N_g\mathbb{1}(C_{ig} = c)]$.

[16] Given that $\mathbb{E}(\rho_g\boldsymbol{\mathcal{Z}}_{ig}\boldsymbol{X}'_{ig}) = \mathbb{I}$, we could have defined our estimator to be $\frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N_g}\widehat{\boldsymbol{\mathcal{Z}}}_{ig}Y_{ig}$ rather than $\widehat{\boldsymbol{\vartheta}}$. It is more convenient both for our asymptotic derivations and practical implementation, however, to work with an IV estimator.

(i) $Var\left(\frac{1}{N_g}\sum_{i=1}^{N_g}\rho_g \boldsymbol{\mathcal{Z}}_{ig}U_{ig}\right) \to \boldsymbol{\Sigma}$ *as* $G \to \infty$,

(ii) $\mathbb{E}\left[\rho_g^{2+\delta}||\boldsymbol{\mathcal{Z}}_{ig}U_{ig}||^{2+\delta}\right] = o(G^{\delta/2})$ *for some* $\delta > 0$, *and*

(iii) $||\sum_{g=1}^{G}\frac{1}{N_g}\sum_{i=1}^{N_g}\rho_g(\widehat{\boldsymbol{\mathcal{Z}}}_{ig} - \boldsymbol{\mathcal{Z}}_{ig})U_{ig}|| = o_{\mathbb{P}}(G^{1/2})$.

Then, under the conditions of Theorem 4, $\sqrt{G}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \to_d N(\mathbf{0}, \boldsymbol{\Sigma})$.

Combined with the first four conditions of Theorem 4, (i) and (ii) from Theorem 5 are sufficient for the asymptotic normality of $\widetilde{\boldsymbol{\vartheta}}$, the infeasible estimator defined in (19). Condition (i) implies that the rate of convergence of $\widetilde{\boldsymbol{\vartheta}}$ is $G^{-1/2}$. Obtaining a rate of convergence that depends on the total number of *individuals* rather than groups in the sample would require assumptions that are implausible in typical applications of the randomized saturation design.[17] Conditions (ii) and (iii) strengthen (v) and (vii), respectively, from Theorem 4: (ii) is sufficient for the Lindeberg condition, which we use to establish a central limit theorem, while (iii) ensures that the limit distribution of the feasible estimator $\widehat{\boldsymbol{\vartheta}}$ coincides with that of the infeasible estimator $\widetilde{\boldsymbol{\vartheta}}$.

Conditions (vi)–(vii) of Theorem 4, along with condition (iii) of Theorem 5, require the difference $(\widehat{\boldsymbol{\mathcal{Z}}}_{ig} - \boldsymbol{\mathcal{Z}}_{ig})$ to be sufficiently small on average that the limiting behavior of $\widehat{\boldsymbol{\vartheta}}$ coincides with that of the infeasible estimator. We now provide low-level sufficient conditions for this to obtain. By definition,

$$\widehat{\boldsymbol{\mathcal{Z}}}_{ig} - \boldsymbol{\mathcal{Z}}_{ig} = \left[\mathbf{R}(\widehat{C}_{ig}, N_g)^{+} - \mathbf{R}(\bar{C}_{ig}, N_g)^{-1}\right]\boldsymbol{W}_{ig}. \tag{20}$$

Accordingly, so long as $\mathbf{R}$ is a sufficiently well-behaved function, $(\widehat{\boldsymbol{\mathcal{Z}}}_{ig} - \boldsymbol{\mathcal{Z}}_{ig})$ will be small if $|\widehat{C}_{ig} - \bar{C}_{ig}|$ is. As shown in the following lemma, a sufficient condition for this difference to vanish *uniformly* over (i, g) is for the minimum group size $\underline{n}$ to be large relative to $\log G$.

Lemma 4. *Suppose that* $0 < \underline{s} \geq S_g$ *and* $\underline{n} \leq N_g$. *Under Assumptions 1–2 and 4–6*

$$\max_{1\leq g\leq G}\left(\max_{1\leq i\leq N_g}\left|\widehat{C}_{ig} - \bar{C}_{ig}\right|\right) = O_{\mathbb{P}}\left(\sqrt{\frac{\log G}{\underline{n}}}\right) \text{ as } (\underline{n}, G) \to \infty.$$

The following regularity conditions are sufficient for $\mathbf{R}(\widehat{C}_{ig}, N_g)^{+} - \mathbf{R}(\bar{C}_{ig}, N_g)^{-1}$ to inherit the asymptotic behavior of $(\widehat{C}_{ig} - \bar{C}_{ig})$.

[17] Obtaining the faster rate of convergence would require $\text{Var}\left(\frac{1}{N_g}\sum_{i=1}^{N_g}\rho_g \boldsymbol{\mathcal{Z}}_{ig}U_{ig}\right) \to \mathbf{0}$ as $G \to \infty$. Because we consider an asymptotic sequence in which the minimum group size grows with G, this is technically possible. It would, however, require us to assume that both heterogeneity between groups and dependence within groups to vanish in the limit.

Assumption 8 (Regularity Conditions for $\mathbf{R}$).

(i) $\mathbf{R}(\bar{c}, n)$ *is well-defined and symmetric for all* $\bar{c} \in [\bar{c}_L/2, 1)$, $n \geq \underline{n}$ *where* $0 < \bar{c}_L \leq \bar{C}_{ig}$;

(ii) $\inf_{\bar{c} \geq \bar{c}_L/2, n \geq \underline{n}} \sigma\left(\mathbf{R}(\bar{c}, n)\right) > \underline{\sigma} > 0$, *where* $\sigma(\mathbf{M})$ *denotes the minimum eigenvalue of* $\mathbf{M}$;

(iii) $||\mathbf{R}(\bar{c}_1, n) - \mathbf{R}(\bar{c}_2, n)|| \leq L\left\{|\bar{c}_1 - \bar{c}_2| + O(n^{-1/2})\right\}$ *as* $n \to \infty$ *for some* $0 < L < \infty$.

Parts (i) and (ii) of Assumption 8 require that $\mathbf{R}$ is well-defined and uniformly invertible over a range of values for $\bar{c}$ that includes the support of $\bar{C}_{ig}$ and excludes zero. Part (iii) is a variant of Lipschitz continuity that holds in the limit as n grows. These conditions are mild: they amount to a slight strengthening of the rank condition from Assumption 7. In the linear basis function example from (15) and (16), for instance, Assumption 8 holds whenever $\bar{C}_{ig}$ is bounded away from zero and S_g takes on at least two distinct values between zero and one.[18] More generally, provided that Assumption 7 holds, whenever $\bar{C}_{ig}$ is bounded away from zero and the basis functions $\mathbf{f}$ are well-behaved, we can always *extend* the definitions of $\mathbf{Q}_0, \mathbf{Q}_1$ from (9)–(10) to ensure that Assumption 8 holds. See Appendix D for full details. Under this assumption, we can derive sufficient conditions on the rates at which G and $\underline{n}$ approach infinity to ensure that the difference between $\widehat{\boldsymbol{\mathcal{Z}}}_{ig}$ and $\boldsymbol{\mathcal{Z}}_{ig}$ is negligible.

Theorem 6. *Suppose that* $\mathbb{E}\left[\rho_g^2 ||\boldsymbol{W}_{ig}\boldsymbol{X}_{ig}'||^2\right]$ *and* $\mathbb{E}\left[\rho_g^2 ||\boldsymbol{W}_{ig}U_{ig}||^2\right]$ *are both* $o(G)$. *Then, under condition (i) of Theorem 4 and the conditions of Lemma 4,*

(i) $\log G/\underline{n} \to 0$ *is sufficient for conditions (vi)–(vii) of Theorem 4.*

(ii) $G \log G/\underline{n} \to 0$ *is sufficient for condition (iii) of Theorem 5.*

Taken together, Theorems 4–6 establish that $\widehat{\boldsymbol{\vartheta}}$ from (18) is consistent, and asymptotically normal in the limit as G and $\underline{n}$ grow at an appropriate rate. In practical terms, our estimators are appropriate for settings with many large groups such as the experiment of Crépon et al. (2013). To implement them in practice, all that is required is to calculate the estimated instrument $\widehat{\boldsymbol{\mathcal{Z}}}_{ig}$ and then run the appropriate just-identified IV regression from Table 1 with standard errors clustered by group.

5 Application: Job Placement Program in the French Labor Market

In this section we illustrate our methods using data from Crépon et al. (2013), who implemented a large-scale randomized saturation experiment across French cities, offering job

[18] See the discussion in section 3 immediately following (15) for details.

placement program services to young workers seeking employment. In doing so, we uncover patterns of spillovers that could prove relevant for the design of similar labor market programs. The intervention included 235 cities (labor markets), covering a sample of 21,431 workers of whom 11,806 were unemployed at the time of randomization.[19] Two questions of interest arise in this setting. First, the presence of direct effects: whether receiving job placement services impacts subsequent labor market outcomes of participants, in particular the likelihood of being employed. Second, the presence of indirect (spillover) effects: whether the receipt of job placement services by others in the same labor market impacts subsequent labor market outcomes of participants. For example, in such a large-scale experiment one may worry that increasing some workers' likelihood of obtaining a job may hurt the labor market prospects of other workers.

Cities were initially randomly assigned to five saturation bins $\mathcal{S} = \{0, 0.25, 0.5, 0.75, 1\}$. For reasons outside of the experiment, 43 of the 47 cities initially assigned to the 25% saturation bin in fact received a 50% saturation, and 12 of the 47 cities initially assigned to the 75% saturation bin received a 100% saturation.[20] For this reason all of the results we present below restrict attention to the subset of cities that received their initially assigned saturation.[21] Thus, our estimation sample consists of 47 cities in the 0% saturation bin, 4 cities in the 25% saturation bin, 47 cities in the 50% saturation bin, 35 cities in the 75% saturation bin, and 47 cities in the 100% saturation bin.

Eligible workers in each city then received offers with a probability equal to the saturation assigned to their city. As mentioned in the introduction, the overall take-up rate of job placement services was 35%. Only workers who were assigned to treatment could receive it, so Assumption 4 (one-sided non-compliance) holds. In addition, Assumption 5 (IOR) is reasonable in this setting: using a simple regression-based test, Appendix E shows that an individual's probability of treatment take-up is statistically unrelated to her group's randomly assigned saturation.[22] Researchers collected data on labor market outcomes in a follow-up 8 months after treatment receipt. Here we present results for two outcome variables: long-term employment (indefinite contract or fixed-term contract longer than 6 months) and any employment. Throughout we fix a linear potential outcomes function, so

[19] The formal criteria for eligibility included "aged below 30, with at least a two-year college degree, and having spent either 12 out of the last 18 months or 6 months continuously unemployed or underemployed" (Crépon et al., 2013, p. 545).

[20] The reassignment of cities across bins is not a problem for the analysis in Crépon et al. (2013), because the main results in that study make only a binary comparison between the cities assigned to the 0% saturation bin and the pooled group of cities assigned to positive saturation bins.

[21] Naturally, the validity of this restriction relies on the assumption that the reassignment of cities across saturation bins was unrelated to their underlying characteristics.

[22] As far as we are aware, subjects in the experiment of Crépon et al. (2013) were not informed of their groups' saturations, making IOR *a priori* plausible as well.

	α	γ	α^n	γ^n	α^c	γ^c	β^c	δ^c
Outcome: long-term employment								
Estimate	0.47	-0.09	0.47	0.14	0.48	-0.51	-0.09	0.62
Std. error	0.01	0.07	0.02	0.09	0.04	0.24	0.05	0.25
Outcome: any employment								
Estimate	0.60	-0.11	0.57	0.14	0.66	-0.56	-0.10	0.62
Std. error	0.01	0.06	0.02	0.09	0.04	0.24	0.05	0.25
Observations	7,440		5,814		3,104			

Table 2: Estimated coefficients for long-term employment and any employment. Standard errors are clustered at the city level. See Equation 21 for the coefficient definitions.

that $\mathbf{f}(\bar{d}) = (1, \bar{d})$:

$$Y_{ig} = \alpha_{ig} + \beta_{ig} D_{ig} + \gamma_{ig} \bar{D}_{ig} + \delta_{ig} D_{ig} \bar{D}_{ig}. \tag{21}$$

Recall that our RS-IV estimator recovers average coefficients for compliers $(\alpha^c, \beta^c, \gamma^c, \delta^c)$, for never-takers (α^n, γ^n), and for the whole population (α, γ).[23] Using these, we can reconstruct the average potential outcome functions for treated and untreated compliers, for untreated never-takers, and for the whole population.[24]

Using the estimator proposed in section 4,[25] Table 2 presents estimates and standard errors (clustered at the city level) of the average effects for the whole population, for never-takers, and for compliers using long-term employment and any employment as outcome variables. We estimate large negative spillovers ($\gamma^c = -0.51$) for untreated compliers, and effectively no spillovers ($\gamma^c + \delta^c = 0.62 - 0.51 = 0.11$) for treated compliers. For the average untreated complier, increasing the treated share among his neighbors from 10 percent to 50 percent would decrease his likelihood of employment by 20 percentage points. This is a considerable negative indirect effect of the policy intervention. However, this negative spillover effect is nullified –and possibly reversed– when compliers are assigned to, and therefore receive, the treatment.

For completeness, Figure 3 depicts the implied average potential outcome functions for untreated and treated compliers, using long-term employment as the outcome variable. Figure 4 depicts the corresponding functions using any employment as the outcome variable instead. We report average functions as bold lines, and corresponding (pointwise) 95% confidence intervals as dashed curves. The downward sloping functions on the left of both figures

[23] In the more general notation in section section 2, $\boldsymbol{\theta}_{ig} = (\alpha_{ig}, \gamma_{ig})$ and $(\boldsymbol{\psi}_{ig} - \boldsymbol{\theta}_{ig}) = (\beta_{ig}, \delta_{ig})$.

[24] Because non-compliance is one-sided, compliers are synonymous with "the treated" and never-takers with "the untreated."

[25] We include the observations from 0% saturation cities, using the over-identified 2SLS estimator described in Appendix C.

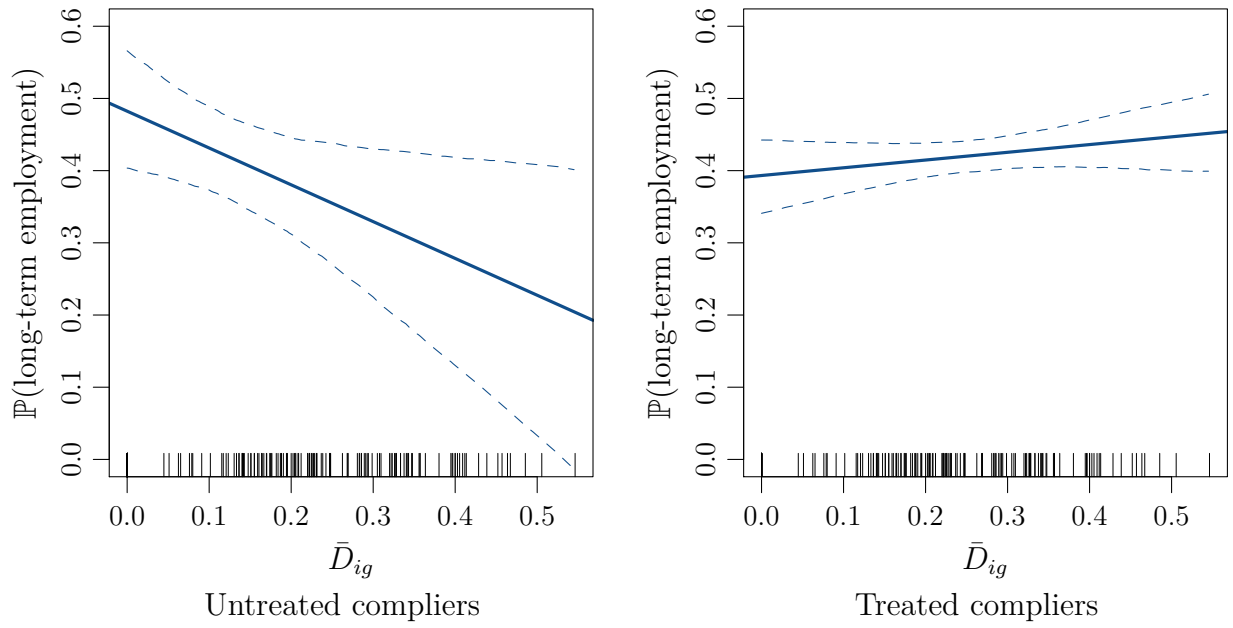

Figure 3: Potential outcomes as a function of $\bar{d}_{ig}$ using the probability of long-term employment as outcome. The left-hand side panel illustrates the average potential outcome function for untreated compliers: $\alpha^c + \gamma^c \bar{d}_{ig}$. The right-hand side panel illustrates the average potential outcome function for treated compliers: $(\alpha^c+\beta^c)+(\gamma^c+\delta^c)\bar{d}_{ig}$. The dashed curves represent 95% confidence intervals. Each tick in the rug plot on the horizontal axis represents a realized value of $\bar{D}_g$ in a city in the experiment.

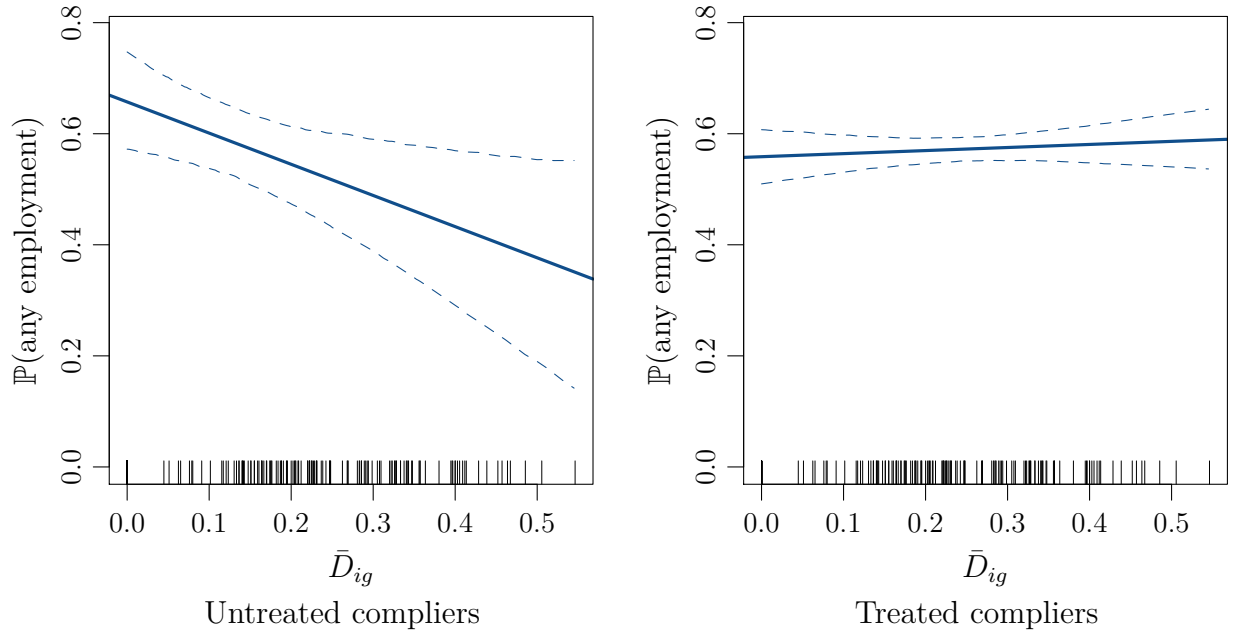

Figure 4: Potential outcomes as a function of $\bar{d}_{ig}$ using the probability of any employment as outcome. The left-hand side panel illustrates the average potential outcome function for untreated compliers: $\alpha^c + \gamma^c \bar{d}_{ig}$. The right-hand side panel illustrates the average potential outcome function for treated compliers: $(\alpha^c+\beta^c)+(\gamma^c+\delta^c)\bar{d}_{ig}$. The dashed curves represent 95% confidence intervals. Each tick in the rug plot on the horizontal axis represents a realized value of $\bar{D}_g$ in a city in the experiment.

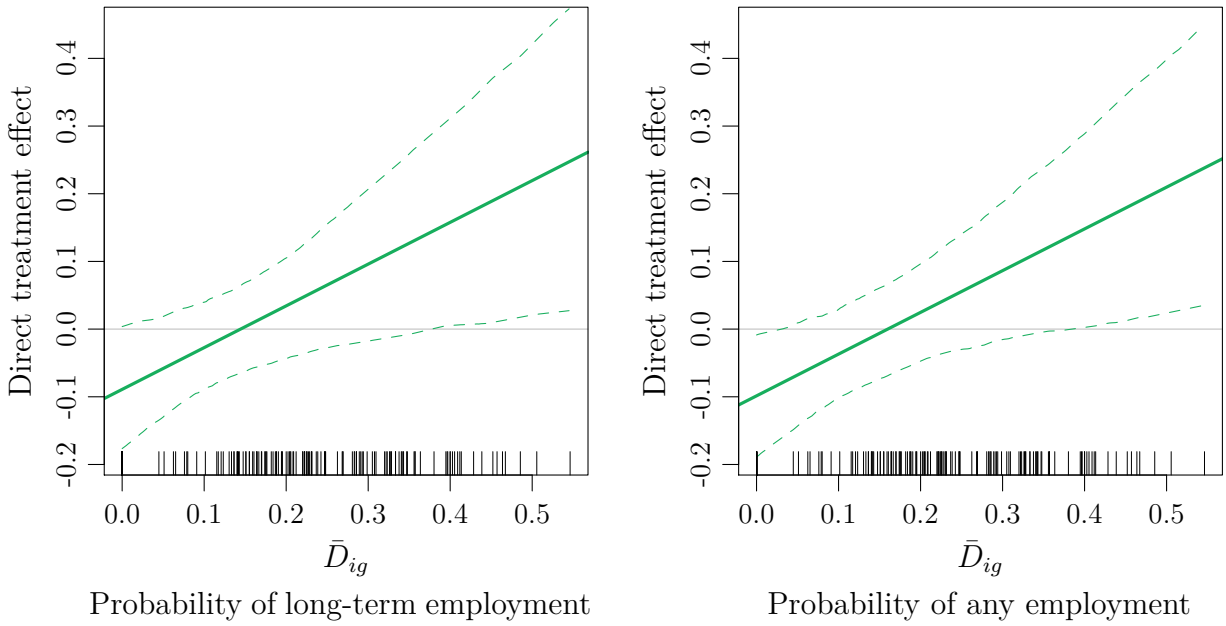

Figure 5: Direct treatment effect as a function of $\bar{d}_{ig}$ for compliers: $\beta^c + \delta^c \bar{d}_{ig}$. The left-hand side figure uses long-term employment as outcome variable. The right-hand side figure uses any employment as outcome. The dashed curves represent 95% confidence intervals. Each tick in the rug plot on the horizontal axis represents a realized value of $\bar{D}_g$ in a city in the experiment.

illustrate the negative estimated spillover for untreated compliers: employment prospects for those who would have taken up treatment if offered worsen rapidly as more job seekers in their city take up the job placement program. The flat curves on the right, in contrast, reveal that employment prospects for those who take up treatment are unaffected by the average city-level treatment take up. These patterns are consistent with the idea that compliers who did not receive job placement assistance are hurt by competition in the labor market, while job placement assistance shields those who take it up from these negative spillovers.

Thus, among those willing to receive job placement services, more widespread take-up of the program, possibly via increased labor market competition, has a differential impact across those who do receive and those who do not receive treatment. This difference is driven by the direct treatment effects on compliers, which we plot in Figure 5. The estimated direct effect increases with $\bar{D}_{ig}$ and is positive for most values of $\bar{D}_{ig}$ observed in the data, although the 95% confidence interval contains an effect size of zero for most observations. Finally, although we cannot recover full treatment effects for never-takers or for the population as a whole, Table 2 also illustrates that the average spillover γ^n for never takers is positive albeit statistically insignificant. The resulting average spillover for the population as a whole, γ, although much smaller in magnitude compared to the one for compliers, is negative and marginally significant for any employment ($\gamma = -0.11$).

In settings with potential non-compliance such as this one, participants' take-up decisions may be driven by the expected gains from participation. Our findings are consistent with such behavior: those who decline participation may do so precisely if they expect they will not suffer negative spillovers from others receiving the program. In turn, compliance may in part be driven by the knowledge that, in the absence of treatment, program receipt by others hurts own labor market prospects.[26] Indeed, in Table B.1 we report results from a regression of compliance indicators on pre-treatment characteristics for the sub-sample of offered individuals. Compared to never-takers, compliers appear to be a more vulnerable sub-population: at baseline they are less likely to cohabit, less educated, less likely to be employed or to have a stable labor contract, and are more likely to receive unemployment insurance.[27] Knowledge of this pattern of effects may prove valuable for the design of other similar large-scale labor market programs.

6 Simulation study

We now present the results of a simulation study to demonstrate the performance of our estimator. As in section 5, we assume $\mathbf{f}(x)' = [1 \quad x]$ to obtain a linear model of the form given in (5):

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

We compare the results of our estimator to those of a 'naïve' IV regression of Y_{ig} on $\mathbf{X}_{ig} \equiv (1, D_{ig}, \bar{D}_{ig}, D_{ig}\bar{D}_{ig})'$ with instruments $\boldsymbol{\mathcal{Z}}_{ig} \equiv (1, Z_{ig}, S_g, Z_{ig}S_g)'$. Lemma 2 shows that this regression yields unbiased estimates of α and β^c, and biased estimates of γ and δ^c.

Our simulation design broadly follows the sampling and experimental design of Crépon et al. (2013) and follows a simple data generating process that allows for correlation between the random coefficients and the share of compliers in a city, $\bar{C}_g$. We present results from three simulation studies with different numbers of groups, G. Our main simulations set $G = 235$ to match the experimental design in Crépon et al. (2013), and comparison exercises use 150 and 500 groups. In all cases, we consider groups of equal size, with 116 individuals in each, which is the average group size in Crépon et al. (2013). We randomly assign exactly 1/5 of groups to one of five treatment assignment saturations, $s_g \in \mathcal{S} = \{0, 0.25, 0.5, 0.75, 1\}$, and make individual Bernoulli offers.

We randomly assign to each group g a share of compliers $\bar{C}_g \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ with equal probability, and individuals are assigned a compliance status in the corresponding

[26] Notice that these forms of 'selection on gains' are compatible with the IOR assumption holding.

[27] Compliers are also less likely to have young children, which may indicate that never takers are less able to participate in the program and possibly in the labor market.

	α	γ	α^n	γ^n	α^c	γ^c	β^c	δ^c
True values	0.50	-0.70	0.50	-0.73	0.50	-0.63	0.20	0.94
150 groups								
RS-IV								
Average coefficient	0.50	-0.69	0.50	-0.73	0.50	-0.59	0.21	0.89
Std. dev.	0.00	0.08	0.01	0.10	0.04	0.36	0.07	0.44
Coverage	0.97	0.95	0.91	0.91	0.98	0.97	0.96	0.96
Naïve IV								
Average coefficient	0.50	-0.63					0.21	1.02
Std. dev.	0.00	0.05					0.06	0.29
Coverage	0.97	0.65					0.95	0.91
235 groups								
RS-IV								
Average coefficient	0.50	-0.69	0.50	-0.73	0.50	-0.60	0.20	0.91
Std. dev.	0.00	0.06	0.01	0.08	0.03	0.28	0.05	0.34
Coverage	0.97	0.94	0.91	0.92	0.98	0.97	0.96	0.96
Naïve IV								
Average coefficient	0.50	-0.63					0.20	1.03
Std. dev.	0.00	0.04					0.05	0.22
Coverage	0.97	0.50					0.95	0.90
500 groups								
RS-IV								
Average coefficient	0.50	-0.69	0.50	-0.73	0.50	-0.60	0.20	0.91
Std. dev.	0.00	0.04	0.01	0.05	0.02	0.19	0.04	0.23
Coverage	0.97	0.95	0.91	0.91	0.98	0.97	0.96	0.95
Naïve IV								
Average coefficient	0.50	-0.63					0.20	1.04
Std. dev.	0.00	0.02					0.03	0.15
Coverage	0.97	0.20					0.95	0.87

Table 3: Comparison of our RS-IV and the 'naïve' IV in simulations with 150, 235 or 500 groups. We show the mean, standard deviation, and coverage for estimates using the RS-IV and 'naïve' IV over 5000 simulations.

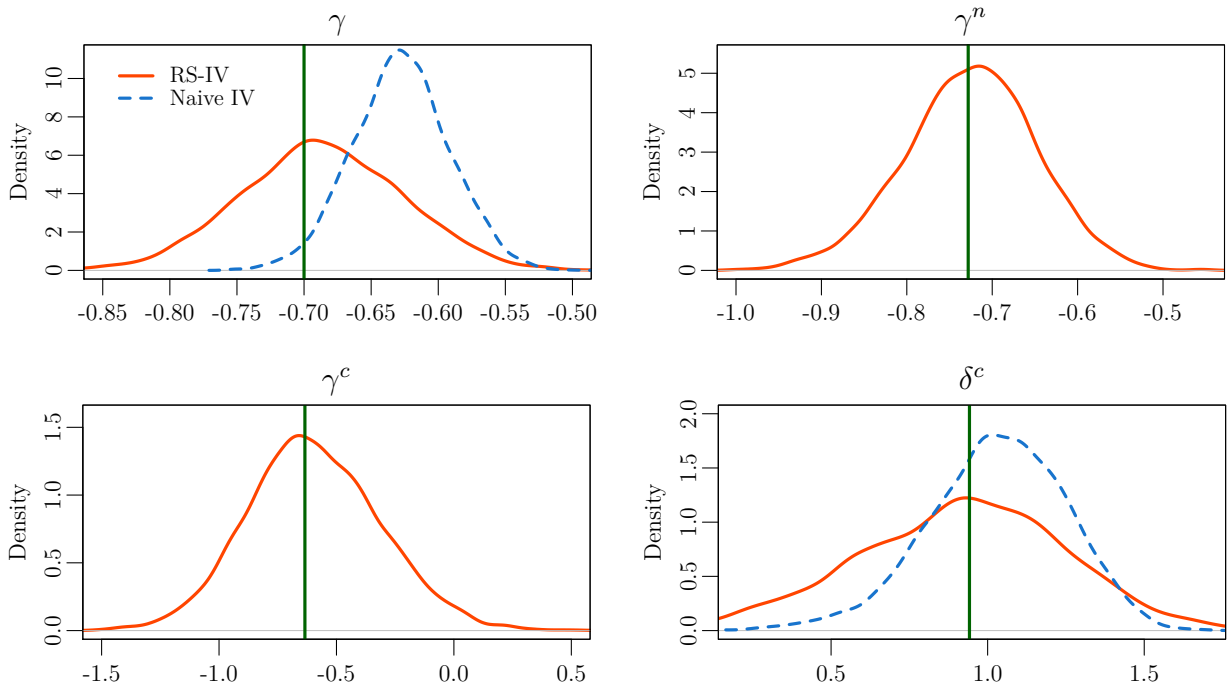

Figure 6: Distribution of the estimates of the spillover terms, $(\gamma, \gamma^n, \gamma^c, \delta^c)$, for our IV and the 'naïve' IV (where available) for simulations with 235 groups, over 5000 simulations. The true parameter values are given by green vertical lines. Analogous figures for simulations with 150 and 500 groups appear in Appendix B.

proportion. To generate the random coefficients, we first choose and hold fixed arbitrary values for the four unconditional average parameters, $(\alpha, \beta, \gamma, \delta) = (0.5, 0.2, -0.7, 0.8)$. We then draw individual level random coefficients according to (with analogous expressions for $\beta_{ig}, \gamma_{ig}, \delta_{ig}$):

$$\alpha_{ig} = \alpha + \left[\frac{\bar{C}_{ig} - \mathbb{E}[\bar{C}_{ig}]}{sd(\bar{C}_{ig})} \frac{\kappa_\alpha}{\sqrt{\kappa_\alpha^2 + 1}} + \frac{u_{ig}}{\sqrt{\kappa_\alpha^2 + 1}} \right] \sigma_\alpha, \qquad u_{ig} \overset{iid}{\sim} \mathcal{N}(0,1) \tag{22}$$

where $\boldsymbol{\kappa} = (\kappa_\alpha, \kappa_\beta, \kappa_\gamma, \kappa_\delta)$ controls the strength of correlation between $\bar{C}_{ig}$ and each of the random coefficients $(\alpha_{ig}, \beta_{ig}, \gamma_{ig}, \delta_{ig})$, such that $corr(\alpha_{ig}, \bar{C}_{ig}) = \kappa_\alpha / \sqrt{\kappa_\alpha^2 + 1}$ (with analogous expressions for $\beta_{ig}, \gamma_{ig}, \delta_{ig}$). Thus, we normalize the random coefficients so that their means are given by the unconditional parameters, $(\alpha, \beta, \gamma, \delta)$, and their standard deviations are given by $\boldsymbol{\sigma} = (\sigma_\alpha, \sigma_\beta, \sigma_\gamma, \sigma_\delta)$. In the simulations below, we set $\boldsymbol{\kappa} = (0, 0, 1.2, 1.5)$ and $\boldsymbol{\sigma} = (0.3, 0.3, 0.2, 0.4)$, which gives $corr(\gamma_{ig}, \bar{C}_{ig}) \approx 0.77$ and $corr(\delta_{ig}, \bar{C}_{ig}) \approx 0.83$.

We report the main results of our simulations in Table 3, which shows the mean and standard deviation of our estimated coefficients and the coverage of our estimated 95%

confidence intervals for both our estimator and the 'naïve' IV across 5000 simulations.[28] The second panel presents the results for a sample size similar to the experimental design in Crépon et al. (2013), with 235 groups. Our estimator performs well at this sample size—the average coefficients are very close to the true values and the coverage is close to the nominal level for all eight parameter values—and its performance improves in larger samples, as expected. In contrast, the naïve IV estimates of γ and δ^c are biased, as shown in Lemma 2. Naturally, the performance of the naïve IV does not improve as we increase the sample size—the average coefficients do not change and the coverage worsens as the standard errors shrink. Figure 6 shows the empirical distribution of the point estimates for our estimator in the simulations with 235 groups and compares this to the naïve IV for γ and δ^c. Again, our estimator performs well and the bias of the naïve IV is clearly visible, as is the mean-variance tradeoff between the two estimators. Appendix B presents similar figures for simulations with 150 and 500 groups.

7 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. Under appropriate assumptions, we show that the key source of unobserved heterogeneity is the share of compliers within a given group. In a setting with many large groups, this quantity can be estimated and yields a simple IV estimator that is consistent and asymptotically normal in the limit as group size and the number of groups grow. We have also illustrated the applicability of our methods using data from a large-scale job-placement program randomized saturation experiment. In this setting, we find negative spillover effects on the sub-population willing to take up the program. The direct effects, however, shield those who take up treatment from those negative indirect effects. A possible extension of the methods described above would be to consider settings with two-sided non-compliance. In this case our identification approach would condition on the share of always-takers in addition to the share of compliers.

[28] In principle, one could estimate (α^n, γ^n) by estimating a naïve IV regression of Y_{ig} on a constant and $\bar{D}_{ig}$ on a subset of the data with $(Z_{ig} = 1, D_{ig} = 0)$, using S_g as an instrument for $\bar{D}_{ig}$. Similarly, one could estimate $(\alpha^c + \beta^c, \gamma^c + \delta^c)$ by estimating the same regression on a subset of the data with $(Z_{ig} = 1, D_{ig} = 1)$. However, both sets of estimated parameters would be biased if γ is correlated with $\bar{C}_{ig}$, and the second set of estimates would be biased if δ is correlated with $\bar{C}_{ig}$.

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, 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, 486–508.

Baird, S., Bohren, J.A., McIntosh, C., Özler, B., 2018. Optimal design of experiments in the presence of interference. Review of Economics and Statistics 100, 844–860.

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, 167–95.

Bhattacharya, D., Dupas, P., Kanaya, S., 2021. Demand and welfare analysis in discrete choice models with social interactions. Technical Report.

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, 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, 295.

Bursztyn, L., Cantoni, D., Yang, D.Y., Yuchtman, N., Zhang, Y.J., 2021. Persistent political engagement: Social interactions and the dynamics of protest movements. American Economic Review: Insights 3, 233–50.

Constantinou, P., Dawid, A.P., 2017. Extended conditional independence and applications in causal inference. The Annals of Statistics 45, 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, 531–580.

Dawid, A.P., 1979. Conditional independence in statistical theory. Journal of the Royal Statistical Society: Series B (Methodological) 41, 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, 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, 7316–7322.

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

Graham, B.S., de Xavier Pinto, C.C., 2022. Semiparametrically efficient estimation of the average linear regression function. Journal of Econometrics 226, 115–138.

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, 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.

Hoeffding, W., 1963. Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association 58, 13–30.

Hudgens, M.G., Halloran, M.E., 2008. Toward causal inference with interference. Journal of the American Statistical Association 103, 832–842. doi:10.1198/016214508000000292.

Imai, K., Jiang, Z., Malani, A., 2020. Causal inference with interference and noncompliance in two-stage randomized experiments. Journal of the American Statistical Association , 1–13.

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

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

Manski, C.F., 2013. Identification of treatment response with social interactions. Econometrics Journal 16, 1–23. doi:10.1111/j.1368-423X.2012.00368.x.

Masten, M.A., Torgovitsky, A., 2016. Identification of instrumental variable correlated random coefficients models. Review of Economics and Statistics 98, 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, 1055–1069.

Vazquez-Bare, G., 2021. Causal spillover effects using instrumental variables. Technical Report just-accepted.

Wooldridge, J.M., 1997. On two stage least squares estimation of the average treatment effect in a random coefficient model. Economics Letters 56, 129–133. doi:10.1016/s0165-1765(97)81890-3.

Wooldridge, J.M., 2003. Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. Economics Letters 79, 185–191. doi:10.1016/S0165-1765(02)00318-X.

Wooldridge, J.M., 2004. Estimating average partial effects under conditional moment independence assumptions. Technical Report. cemmap working paper.

Wooldridge, J.M., 2016. Instrumental variables estimation of the average treatment effect in the correlated random coefficient model. Advances in Econometrics 21, 93–116. doi:10.1016/S0731-9053(07)00004-7.

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.

A Proofs

The following lemma, taken from Constantinou and Dawid (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 6(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. Under Assumption 2, (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 $\sum_{j\in\mathcal{C}(i)} Z_{jg} \sim \text{Binomial}\big(\bar{c}(n-1), s\big)$. □

Proof of Lemma 2. Under (5), $Y_{ig} = \mathbf{X}_{ig}'\mathbf{B}_{ig}$ where $\mathbf{B}_{ig} = (\alpha_{ig}, \beta_{ig}, \gamma_{ig}, \delta_{ig})'$. Now, let $\mathcal{R}_{ig} \equiv \{S_g, Z_{ig}, N_g, \bar{C}_{ig}, C_{ig}, \mathbf{B}_{ig}\}$ and $\boldsymbol{\Lambda}_{ig} \equiv \text{diag}\{1, C_{ig}, \bar{C}_{ig}, C_{ig}\bar{C}_{ig}\}$. From Lemma 1 we see that $\mathbb{E}[\bar{D}_{ig}|\mathcal{R}] = \bar{C}_{ig}S_g$. Since $D_{ig} = C_{ig}Z_{ig}$ under one-sided non-compliance and IOR, it follows that $\mathbb{E}[\mathbf{X}_{ig}'|\mathcal{R}_{ig}] = \boldsymbol{\mathcal{Z}}_{ig}'\boldsymbol{\Lambda}_{ig}$. Hence,

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\mathcal{Z}}_{ig}Y_{ig}\right] &= \mathbb{E}\left[\boldsymbol{\mathcal{Z}}_{ig}\mathbb{E}(\mathbf{X}_{ig}'|\mathcal{R}_{ig})\mathbf{B}_{ig}\right] = \mathbb{E}\left[\left(\boldsymbol{\mathcal{Z}}_{ig}\boldsymbol{\mathcal{Z}}_{ig}'\right)\left(\boldsymbol{\Lambda}_{ig}\mathbf{B}_{ig}\right)\right] \\ \mathbb{E}\left[\boldsymbol{\mathcal{Z}}_{ig}\mathbf{X}_{ig}'\right] &= \mathbb{E}\left[\boldsymbol{\mathcal{Z}}_{ig}\mathbb{E}\left(\mathbf{X}_{ig}'|\mathcal{R}_{ig}\right)\right] = \mathbb{E}\left[\left(\boldsymbol{\mathcal{Z}}_{ig}\boldsymbol{\mathcal{Z}}_{ig}'\right)\boldsymbol{\Lambda}_{ig}\right]\end{aligned}$$

since $\boldsymbol{\mathcal{Z}}_{ig}$ and $\mathbf{B}_{ig}$ are $\mathcal{R}_{ig}$–measurable. Now, applying Decomposition and Corollary A.1 to part (ii) of Assumption 6 gives $Z_{ig} \perp\!\!\!\perp (C_{ig}, \bar{C}_{ig}, \mathbf{B}_{ig})|(S_g, N_g)$. Under Bernoulli offers, however, this conditional distribution does not involve N_g, so we obtain

$$(C_{ig}, \bar{C}_{ig}, \mathbf{B}_{ig}) \perp\!\!\!\perp Z_{ig}|S_g. \tag{A.2}$$

Similarly, applying Decomposition to part (ii) of Corollary A.1, we see that $(C_{ig}, \bar{C}_{ig}, \mathbf{B}_{ig}) \perp\!\!\!\perp S_g$. Combining this with (A.2), the Contraction axiom yields $(C_{ig}, \bar{C}_{ig}, \mathbf{B}_{ig}) \perp\!\!\!\perp (Z_{ig}, S_g)$, implying that $(\boldsymbol{\mathcal{Z}}_{ig}\boldsymbol{\mathcal{Z}}_{ig}')$ is independent of both $\boldsymbol{\Lambda}_{ig}$ and $(\boldsymbol{\Lambda}_{ig}\mathbf{B}_{ig})$. Accordingly,

$$\boldsymbol{\vartheta}_{\text{IV}} = \left\{\mathbb{E}\left[\left(\boldsymbol{\mathcal{Z}}_{ig}\boldsymbol{\mathcal{Z}}_{ig}'\right)\boldsymbol{\Lambda}_{ig}\right]\right\}^{-1}\mathbb{E}\left[\left(\boldsymbol{\mathcal{Z}}_{ig}\boldsymbol{\mathcal{Z}}_{ig}'\right)\left(\boldsymbol{\Lambda}_{ig}\mathbf{B}_{ig}\right)\right] = \mathbb{E}\left[\boldsymbol{\Lambda}_{ig}\right]^{-1}\mathbb{E}\left[\boldsymbol{\Lambda}_{ig}\mathbf{B}_{ig}\right].$$

By the definitions of $\boldsymbol{\vartheta}_{\text{IV}}$, $\boldsymbol{\Lambda}_{ig}$ and $\mathbf{B}_{ig}$ it follows that

$$\alpha_{\text{IV}} = \mathbb{E}\left[\alpha_{ig}\right], \quad \beta_{\text{IV}} = \frac{\mathbb{E}\left[C_{ig}\beta_{ig}\right]}{\mathbb{E}\left[C_{ig}\right]}, \quad \gamma_{\text{IV}} = \frac{\mathbb{E}\left[\bar{C}_{ig}\gamma_{ig}\right]}{\mathbb{E}\left[\bar{C}_{ig}\right]}, \quad \delta_{\text{IV}} = \frac{\mathbb{E}\left[C_{ig}\bar{C}_{ig}\delta_{ig}\right]}{\mathbb{E}\left[C_{ig}\bar{C}_{ig}\right]}.$$

By iterated expectations over C_{ig}, we obtain $\beta_{\text{IV}} = \mathbb{E}\left[\beta_{ig}|C_{ig} = 1\right]$ while

$$\gamma_{\text{IV}} = \frac{\mathbb{E}\left[\bar{C}_{ig}\gamma_{ig}\right]}{\mathbb{E}\left[\bar{C}_{ig}\right]} = \frac{\text{Cov}(\bar{C}_{ig}, \gamma_{ig}) + \mathbb{E}(\bar{C}_{ig})\mathbb{E}(\gamma_{ig})}{\mathbb{E}(\bar{C}_{ig})} = \mathbb{E}[\gamma_{ig}] + \frac{\text{Cov}(\bar{C}_{ig}, \gamma_{ig})}{\mathbb{E}(\bar{C}_{ig})}.$$

Similarly, again taking iterated expectations over C_{ig},

$$\delta_{\text{IV}} = \frac{\mathbb{E}\left[\bar{C}_{ig}\delta_{ig}|\bar{C}_{ig} = 1\right]}{\mathbb{E}\left[\bar{C}_{ig}|C_{ig} = 1\right]} = \mathbb{E}\left[\gamma_{ig}\right] + \frac{\text{Cov}(\bar{C}_{ig}, \delta_{ig}|C_{ig} = 1)}{\mathbb{E}(\bar{C}_{ig}|C_{ig} = 1)}.$$

□

Proof of Theorem 1. Assumption 6(i) implies $(\boldsymbol{C}_g, \mathbf{B}_g) \perp\!\!\!\perp S_g | N_g$ by Weak Union and Decomposition. Combining this with Assumption 6(ii) gives

$$(\mathbf{Z}_g, S_g) \perp\!\!\!\perp (\mathbf{B}_g, \boldsymbol{C}_g) | N_g \tag{A.3}$$

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.3),

$$(S_g, \boldsymbol{Z}_g) \perp\!\!\!\perp (B_{ig}, C_{ig}, \boldsymbol{C}_{-ig}, N_g) | (N_g, \bar{C}_{ig}). \tag{A.4}$$

because $\bar{C}_{ig}$ is a function of $(\boldsymbol{C}_g, N_g)$. By Lemma 1,

$$\bar{D}_{ig} \perp\!\!\!\perp \boldsymbol{C}_{-ig} | (N_g, \bar{C}_{ig}, S_g, Z_{ig}). \tag{A.5}$$

Applying Decomposition to (A.4) gives $\boldsymbol{C}_{-ig} \perp\!\!\!\perp (S_g, Z_{ig}) | (N_g, \bar{C}_{ig})$. Combining this with (A.5),

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

by Contraction. Now, applying Weak Union, Decomposition, and Corollary A.1 to (A.4),

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

since $\bar{D}_{ig}$ is a function of $(\boldsymbol{Z}_g, \boldsymbol{C}_{-ig}, N_g)$. Finally, applying Contraction to (A.6) and (A.7),

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

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

Proof of Lemma 3. Define the shorthand $U \equiv \mathbf{Q}(\bar{c}, n)$, $A \equiv \mathbf{Q}_0(\bar{c}, n)$, and $B = \mathbf{Q}_1(\bar{c}, n)$ so that

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

Using this notation, we are asked to show that U is invertible if and only if A and B are both invertible, in which case $U^{-1} = V$ where

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

The "if" direction follows by direct calculation: $VU = UV = \mathbb{I}$. 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

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

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 $AC + B(C + E) = \mathbb{I}$; 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

$CA+(C+D)B = \mathbb{I}$; multiplying the first row of U^{-1} by the second column of U gives $(C+D)B = 0$. Combining these, $CA = \mathbb{I}$. Since $AC = CA = \mathbb{I}$, 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 $B(D + F) = \mathbb{I}$. Combining these, $AD = -\mathbb{I}$ 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 $(E + F)B = \mathbb{I}$. Combining these, $EA = -\mathbb{I}$ 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 $B(D + F) = \mathbb{I}$, but since $D = -C$ this becomes $B(F - C) = \mathbb{I}$ 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 $(F - C)B = \mathbb{I}$. Thus, $B(F - C) = (F - C)B = \mathbb{I}$ so we have shown that B is invertible with $B^{-1} = F - C$. □

Proof of Theorem 2. For each part, it suffices to find an appropriate outcome variable $\widetilde{Y}_{ig}$, regressor vector $\widetilde{\mathbf{X}}_{ig}$, and instrument set $\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}$ such that we can write $\widetilde{Y}_{ig} = \widetilde{\mathbf{X}}_{ig}'\boldsymbol{\vartheta} + U_{ig}$ where $\boldsymbol{\vartheta}$ is the parameter of interest, $\mathbb{E}[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}U_{ig}] = \mathbf{0}$, and $\mathbb{E}[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}\widetilde{\mathbf{X}}_{ig}']$ is invertible. Note that $(\widetilde{\mathbf{X}}_{ig}, \widetilde{Y}_{ig}, \widetilde{\boldsymbol{\mathcal{Z}}}_{ig})$ are placeholders for quantities that differ in each part of the proof: for part (i) they represent $(\mathbf{X}_{ig}, Y_{ig}, \boldsymbol{\mathcal{Z}}_{ig}^W)$ while for part (ii) they stand for $\left(D_{ig}\mathbf{f}(\bar{D}_{ig}), D_{ig}Y_{ig}, \boldsymbol{\mathcal{Z}}_{ig}^1\right)$, for example. The definitions of U_{ig} and $\boldsymbol{\vartheta}$ are also specific to each part of the proof.

Part (i) By (2) we can write $\widetilde{Y}_{ig} = \widetilde{\mathbf{X}}_{ig}'\boldsymbol{\vartheta} + U_{ig}$ where $\boldsymbol{\vartheta}' \equiv \left[\mathbb{E}(\boldsymbol{\theta}_{ig}') \quad \mathbb{E}(\boldsymbol{\psi}_{ig}' - \boldsymbol{\vartheta}_{ig}'|C_{ig} = 1)\right]$, $\widetilde{Y}_{ig} \equiv Y_{ig}$, $\widetilde{\mathbf{X}}_{ig} \equiv \mathbf{X}_{ig}$, and $U_{ig} \equiv \mathbf{X}_{ig}'(\mathbf{B}_{ig} - \boldsymbol{\vartheta})$. Under IOR $D_{ig} = C_{ig}Z_{ig}$. Hence, defining $\mathbf{M}_{ig} \equiv \operatorname{diag}\{1, C_{ig}\} \otimes \mathbb{I}_K$,

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

Since $\mathbf{M}_{ig}$ is symmetric, $U_{ig} = \mathbf{W}_{ig}'\left[\mathbf{M}_{ig}\left(\mathbf{B}_{ig} - \boldsymbol{\vartheta}\right)\right]$. Thus, taking $\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}^W \equiv \boldsymbol{\mathcal{Z}}_{ig}$, we have

$$\mathbb{E}[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}U_{ig}] = \mathbb{E}\left\{\mathbb{E}\left[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}U_{ig}\middle|\,\bar{C}_{ig}, N_g\right]\right\} = \mathbb{E}\left\{\mathbf{Q}(\bar{C}_{ig}, N_g)^{-1}\mathbb{E}\left[\mathbf{W}_{ig}\mathbf{W}_{ig}'\mathbf{M}_{ig}\left(\mathbf{B}_{ig} - \boldsymbol{\vartheta}\right)\middle|\,\bar{C}_{ig}, N_g\right]\right\}$$

by iterated expectations. By assumption $(Z_{ig}, \bar{D}_{ig}) \perp\!\!\!\perp (C_{ig}, \mathbf{B}_{ig})|(\bar{C}_{ig}, N_g)$. Hence,

$$\mathbb{E}\left[\mathbf{W}_{ig}\mathbf{W}_{ig}'\mathbf{M}_{ig}(\mathbf{B}_{ig} - \boldsymbol{\vartheta})\middle|\,\bar{C}_{ig}, N_g\right] = \mathbb{E}\left[\mathbf{W}_{ig}\mathbf{W}_{ig}'\middle|\,\bar{C}_{ig}, N_g\right]\mathbb{E}\left[\mathbf{M}_{ig}(\mathbf{B}_{ig} - \boldsymbol{\vartheta})\middle|\,\bar{C}_{ig}, N_g\right]$$

by Decomposition, since $\mathbf{W}_{ig}\mathbf{W}_{ig}'$ is a measurable function of $(Z_{ig}, \bar{D}_{ig})$ and $\mathbf{M}_{ig}(\mathbf{B}_{ig} - \boldsymbol{\vartheta})$ is a measurable function of $(C_{ig}, \mathbf{B}_{ig})$. Substituting into the expression for $\mathbb{E}[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}U_{ig}]$,

$$\mathbb{E}\left[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}U_{ig}\right] = \mathbb{E}\left\{\mathbb{E}\left[\mathbf{M}_{ig}(\mathbf{B}_{ig} - \boldsymbol{\vartheta})\middle|\,\bar{C}_{ig}, N_g\right]\right\} = \mathbb{E}\left[\mathbf{M}_{ig}\left(\mathbf{B}_{ig} - \boldsymbol{\vartheta}\right)\right]$$

by iterated expectations, since $\mathbf{Q}(\bar{C}_{ig}, N_g)^{-1} = \mathbb{E}[\mathbf{W}_{ig}\mathbf{W}_{ig}'|\bar{C}_{ig}, N_g]^{-1}$. Now, substituting the definitions of $\mathbf{M}_{ig}$, $\mathbf{B}_{ig}$, and $\boldsymbol{\vartheta}$,

$$\mathbb{E}\left[\mathbf{M}_{ig}\left(\mathbf{B}_{ig} - \boldsymbol{\vartheta}\right)\right] = \mathbb{E}\begin{bmatrix}\left(\boldsymbol{\theta}_{ig} - \mathbb{E}\left\{\boldsymbol{\theta}_{ig}\right\}\right)\\ C_{ig}\left(\left\{\boldsymbol{\psi}_{ig} - \boldsymbol{\theta}_{ig}\right\} - \mathbb{E}\left\{\boldsymbol{\psi}_{ig} - \boldsymbol{\theta}_{ig}\middle|\,C_{ig} = 1\right\}\right)\end{bmatrix} = \mathbf{0}$$

since $\mathbb{E}\left[C_{ig}\left(\boldsymbol{\psi}_{ig}-\boldsymbol{\theta}_{ig}\right)\right]=\mathbb{E}(C_{ig})\mathbb{E}\left(\boldsymbol{\psi}_{ig}-\boldsymbol{\theta}_{ig}\middle|\, C_{ig}=1\right)$. Therefore $\mathbb{E}[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}U_{ig}]=\mathbf{0}$. Similarly,

$$\begin{aligned}\mathbb{E}\left[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}\widetilde{\mathbf{X}}_{ig}'\right] &= \mathbb{E}\left\{\mathbf{Q}(\bar{C}_{ig},N_g)^{-1}\mathbb{E}\left[\mathbf{W}_{ig}\mathbf{W}_{ig}'\mathbf{M}_{ig}|\bar{C}_{ig},N_g\right]\right\}\\ &= \mathbb{E}\left\{\mathbf{Q}(\bar{C}_{ig},N_g)^{-1}\mathbb{E}\left[\mathbf{W}_{ig}\mathbf{W}_{ig}'|\bar{C}_{ig},N_g\right]\mathbb{E}\left[\mathbf{M}_{ig}|\bar{C}_{ig},N_g\right]\right\}=\mathbb{E}\left[\mathbf{M}_{ig}\right].\end{aligned}$$

Since $[\mathbf{M}_{ig}]$ is invertible if and only if $\mathbb{E}(C_{ig})\neq 0$, it follows that $\mathbb{E}[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}\widetilde{\mathbf{X}}_{ig}']$ is invertible by Assumption 7.

Part (ii) Since $D_{ig}^2=D_{ig}$ and $D_{ig}(1-D_{ig})=0$, multiplying both sides of (2) by D_{ig} and simplifying gives $D_{ig}Y_{ig}=D_{ig}\mathbf{f}(\bar{D}_{ig})\boldsymbol{\psi}_{ig}$. Thus $\widetilde{Y}_{ig}=\widetilde{\mathbf{X}}_{ig}'\boldsymbol{\vartheta}+U_{ig}$ where $\boldsymbol{\vartheta}\equiv\mathbb{E}(\boldsymbol{\psi}_{ig}|C_{ig}=1)$, $\widetilde{Y}\equiv D_{ig}Y_{ig}$, $\widetilde{\mathbf{X}}_{ig}\equiv D_{ig}\mathbf{f}(\bar{D}_{ig})$, and $U_{ig}\equiv\left[D_{ig}\mathbf{f}(\bar{D}_{ig})\right]'(\boldsymbol{\psi}_{ig}-\boldsymbol{\vartheta})$. The remainder of the argument is similar to that of part (i). Taking $\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}\equiv\boldsymbol{\mathcal{Z}}_{ig}^1$ and substituting $D_{ig}=Z_{ig}C_{ig}$ gives

$$\begin{aligned}\mathbb{E}[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}U_{ig}] &= \mathbb{E}\left\{\mathbf{Q}_1(\bar{C}_{ig},N_g)^{-1}\mathbb{E}\left[\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})'Z_{ig}|\bar{C}_{ig},N_g\right]\mathbb{E}\left[C_{ig}(\boldsymbol{\psi}_{ig}-\boldsymbol{\vartheta})\middle|\,\bar{C}_{ig},N_g\right]\right\}\\ &= \mathbb{E}\left\{\mathbb{E}\left[C_{ig}(\boldsymbol{\psi}_{ig}-\boldsymbol{\vartheta})|\bar{C}_{ig},N_g\right]\right\}=\mathbb{E}\left[C_{ig}(\boldsymbol{\psi}_{ig}-\boldsymbol{\vartheta})\right].\end{aligned}$$

Since $\mathbb{E}[C_{ig}\boldsymbol{\psi}_{ig}]=\mathbb{E}(C_{ig})\mathbb{E}(\boldsymbol{\psi}_{ig}|C_{ig}=1)=\mathbb{E}(C_{ig}\boldsymbol{\vartheta})$, we obtain $\mathbb{E}(\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}U_{ig})=\mathbf{0}$. Similarly,

$$\begin{aligned}\mathbb{E}\left[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}\widetilde{\mathbf{X}}_{ig}'\right] &= \mathbb{E}\left\{\mathbf{Q}_1(\bar{C}_{ig},N_g)^{-1}\mathbb{E}\left[\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})'Z_{ig}C_{ig}|\bar{C}_{ig},N_g\right]\right\}\\ &= \mathbb{E}\left\{\mathbf{Q}_1(\bar{C}_{ig},N_g)^{-1}\mathbb{E}\left[\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})'Z_{ig}|\bar{C}_{ig},N_g\right]\mathbb{E}\left[C_{ig}|\bar{C}_{ig},N_g\right]\right\}=\mathbb{E}(C_{ig})\mathbb{I}_K.\end{aligned}$$

Hence, $\mathbb{E}[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}\widetilde{\mathbf{X}}_{ig}]'$ is invertible by Assumption 7.

Part (iii) Since $(1-D_{ig})^2=(1-D_{ig})$ and $D_{ig}(1-D_{ig})=0$, multiplying both sides of (2) by $Z_{ig}(1-D_{ig})$ and simplifying gives $Z_{ig}(1-D_{ig})Y_{ig}=Z_{ig}(1-D_{ig})\mathbf{f}(\bar{D}_{ig})\boldsymbol{\theta}_{ig}$. Thus we have $\widetilde{Y}_{ig}=\widetilde{\mathbf{X}}_{ig}'\boldsymbol{\vartheta}+U_{ig}$ where $\boldsymbol{\vartheta}\equiv\mathbb{E}(\boldsymbol{\theta}_{ig}|C_{ig}=0)$, $\widetilde{Y}_{ig}\equiv Z_{ig}(1-D_{ig})Y_{ig}$, $\widetilde{\mathbf{X}}_{ig}\equiv Z_{ig}(1-D_{ig})\mathbf{f}(\bar{D}_{ig})$, and $U_{ig}\equiv\left[Z_{ig}(1-D_{ig})\mathbf{f}(\bar{D}_{ig})\right]'(\boldsymbol{\theta}_{ig}-\boldsymbol{\vartheta})$. The remainder of the argument is similar to that of part (i). Taking $\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}\equiv\boldsymbol{\mathcal{Z}}_{ig}^1$ and substituting $Z_{ig}(1-D_{ig})=Z_{ig}(1-C_{ig})$ gives

$$\begin{aligned}\mathbb{E}[\widetilde{\boldsymbol{\mathcal{Z}}}_{\boldsymbol{ig}}U_{ig}] &= \mathbb{E}\left\{\mathbf{Q}_1(\bar{C}_{ig},N_g)^{-1}\mathbb{E}\left[\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})'Z_{ig}|\bar{C}_{ig},N_g\right]\mathbb{E}\left[(1-C_{ig})(\boldsymbol{\theta}_{ig}-\boldsymbol{\vartheta})|\,\bar{C}_{ig},N_g\right]\right\}\\ &= \mathbb{E}\left\{\mathbb{E}\left[(1-C_{ig})(\boldsymbol{\theta}_{ig}-\boldsymbol{\vartheta})|\bar{C}_{ig},N_g\right]\right\}=\mathbb{E}\left[(1-C_{ig})(\boldsymbol{\theta}_{ig}-\boldsymbol{\vartheta})\right].\end{aligned}$$

Since $\mathbb{E}[(1-C_{ig})\boldsymbol{\theta}_{ig}]=\mathbb{E}(1-C_{ig})\mathbb{E}(\boldsymbol{\theta}_{ig}|C_{ig}=1)=\mathbb{E}[(1-C_{ig})\boldsymbol{\vartheta}]$, we obtain $\mathbb{E}(\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}U_{ig})=\mathbf{0}$. Similarly,

$$\begin{aligned}\mathbb{E}\left[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}\widetilde{\mathbf{X}}_{ig}'\right] &= \mathbb{E}\left\{\mathbf{Q}_1(\bar{C}_{ig},N_g)^{-1}\mathbb{E}\left[\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})'Z_{ig}(1-C_{ig})|\bar{C}_{ig},N_g\right]\right\}\\ &= \mathbb{E}\left\{\mathbf{Q}_1(\bar{C}_{ig},N_g)^{-1}\mathbb{E}\left[\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})'Z_{ig}|\bar{C}_{ig},N_g\right]\mathbb{E}\left[(1-C_{ig})|\bar{C}_{ig},N_g\right]\right\}=\mathbb{E}(1-C_{ig})\mathbb{I}_K.\end{aligned}$$

It follows that $\mathbb{E}[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}\widetilde{\mathbf{X}}_{ig}']$ is invertible by Assumption 7.

Part (iv) Under one-sided non-compliance and IOR, $(1-Z_{ig})(1-D_{ig})=(1-Z_{ig})$. Hence, multiplying both sides of (2) by $(1-Z_{ig})$, we obtain $(1-Z_{ig})Y_{ig}=(1-Z_{ig})\mathbf{f}(\bar{D}_{ig})'\boldsymbol{\theta}_{ig}$, using the fact that $Z_{ig}(1-Z_{ig})=0$. Thus we can write $\widetilde{Y}_{ig}=\widetilde{\mathbf{X}}_{ig}'\boldsymbol{\vartheta}+U_{ig}$ where $\boldsymbol{\vartheta}\equiv\mathbb{E}(\boldsymbol{\theta}_{ig})$, $\widetilde{Y}_{ig}\equiv(1-Z_{ig})Y_{ig}$, $\widetilde{\mathbf{X}}_{ig}\equiv(1-Z_{ig})\mathbf{f}(\bar{D}_{ig})$, and $U_{ig}\equiv(1-Z_{ig})\mathbf{f}(\bar{D}_{ig})'(\boldsymbol{\theta}_{ig}-\boldsymbol{\vartheta})$. The remainder of the argument is

similar to that of part (i). Taking $\widetilde{\boldsymbol{\mathcal{Z}}}_{ig} \equiv \boldsymbol{\mathcal{Z}}^0_{ig}$, we obtain

$$\begin{aligned}\mathbb{E}[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}U_{ig}] &= \mathbb{E}\left\{\mathbf{Q}_0(\bar{C}_{ig},N_g)^{-1}\mathbb{E}\left[\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})'(1-Z_{ig})|\bar{C}_{ig},N_g\right]\mathbb{E}\left[\boldsymbol{\theta}_{ig}-\boldsymbol{\vartheta}|\,\bar{C}_{ig},N_g\right]\right\}\\ &= \mathbb{E}\left\{\mathbb{E}\left[\boldsymbol{\theta}_{ig}-\mathbb{E}(\boldsymbol{\theta}_{ig})|\bar{C}_{ig},N_g\right]\right\}=\mathbf{0}\end{aligned}$$

and $\mathbb{E}\left[\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}\widetilde{\mathbf{X}}'_{ig}\right] = \mathbb{E}\left\{\mathbf{Q}_0(\bar{C}_{ig},N_g)^{-1}\mathbb{E}\left[\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})'(1-Z_{ig})|\bar{C}_{ig},N_g\right]\right\} = \mathbb{I}_K.$ □

Lemma A.2. *Under Assumptions 2 and 6,* $(S_g, Z_{ig}) \perp\!\!\!\perp (C_{ig}, \bar{C}_{ig}, N_g, \mathbf{B}_{ig})$.

Proof of Lemma A.2. By Assumption 2 $Z_{ig} \perp\!\!\!\perp N_g | S_g$ and by Assumption 6 (ii) and Decomposition $Z_{ig} \perp\!\!\!\perp (C_{ig}, \mathbf{B}_{ig})|(S_g, N_g)$. Combining these by Contraction yields

$$Z_{ig} \perp\!\!\!\perp (\boldsymbol{C}_g, \mathbf{B}_{ig}, N_g)|S_g. \tag{A.8}$$

Now, by Assumption 6 (i) we have $S_g \perp\!\!\!\perp (\boldsymbol{C}_g, \mathbf{B}_{ig}, N_g)$. Combining this with (A.8) by a second application of Contraction gives $(Z_{ig}, S_g) \perp\!\!\!\perp (\boldsymbol{C}_g, \mathbf{B}_{ig}, N_g)$. The result follows by a final application of Decomposition. □

Proof of Theorem 3. Assumptions 1–6 imply that $(Z_{ig}, \bar{D}_{ig}) \perp\!\!\!\perp (\mathbf{B}_{ig}, C_{ig})|(\bar{C}_{ig}, N_g)$ by Theorem 1. Hence Assumptions 1–7 are sufficient for the conclusions of Theorem 2 to hold. Now, by Lemma 1, Assumptions 1–2 and 4–6 imply that the conditional distribution of $\bar{D}_{ig}|(\bar{C}_{ig}, N_g, Z_{ig})$ is known. Moreover, by Lemma A.2, $Z_{ig} \perp\!\!\!\perp (\bar{C}_{ig}, N_g)$ so the distribution of $\boldsymbol{\mathcal{Z}}_{ig}|(\bar{C}_{ig}, N_g)$ is likewise known. It follows that $\mathbf{Q}, \mathbf{Q}_0$ and $\mathbf{Q}_1$ are *known* functions of $(\bar{C}_{ig}, N_g)$. Since N_g is observed, knowledge of $\bar{C}_{ig}$ is thus sufficient to identify the quantities

$$\mathbb{E}(\boldsymbol{\theta}_{ig}), \quad \mathbb{E}(\boldsymbol{\psi}_{ig}-\boldsymbol{\theta}_{ig}|C_{ig}=1), \quad \mathbb{E}(\boldsymbol{\psi}_{ig}|C_{ig}=1), \quad \mathbb{E}(\boldsymbol{\theta}_{ig}|C_{ig}=0)$$

by the relevant parts of Theorem 2. Now, by iterated expectations,

$$\mathbb{E}(\boldsymbol{\theta}_{ig}|C_{ig}=1) = \mathbb{E}(\boldsymbol{\theta}_{ig}|C_{ig}=0) + \frac{1}{\mathbb{E}(C_{ig})}\left[\mathbb{E}(\boldsymbol{\theta}_{ig}) - \mathbb{E}(\boldsymbol{\theta}_{ig}|C_{ig}=0)\right].$$

Since $\mathbb{E}(C_{ig}) = \mathbb{E}(D_{ig}|Z_{ig}=1)$, it follows that $\mathbb{E}(\boldsymbol{\theta}_{ig}|C_{ig}=1)$ is identified. Under IOR and one-sided non-compliance $\{D_{ig}=1\} = \{C_{ig}=1, Z_{ig}=1\}$, and applying Weak Union and Decomposition to Lemma A.2, we see that $Z_{ig} \perp\!\!\!\perp \mathbf{B}_{ig}|C_{ig}$. Thus,

$$\mathbb{E}(\mathbf{B}_{ig}|D_{ig}=1) = \mathbb{E}(\mathbf{B}_{ig}|C_{ig}=1, Z_{ig}=1) = \mathbb{E}(\mathbf{B}_{ig}|C_{ig}=1).$$

The result follows since $Y_{ig}(d,\bar{d}) = \mathbf{f}(\bar{d})'\boldsymbol{\theta}_{ig} + d\mathbf{f}(\bar{d})'(\boldsymbol{\psi}_{ig}-\boldsymbol{\theta}_{ig})$ under Assumption 3. □

Proof of Theorem 4. Substituting the model into the definition of $\widehat{\boldsymbol{\vartheta}}$ and $\rho_g \equiv N_g/\mathbb{E}(N_g)$,

$$\begin{aligned}\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta} &= \left(\sum_{g=1}^{G}\sum_{i=1}^{N_g}\widehat{\boldsymbol{\mathcal{Z}}}_{ig}\boldsymbol{X}'_{ig}\right)^{-1}\left(\sum_{g=1}^{G}\sum_{i=1}^{N_g}\widehat{\boldsymbol{\mathcal{Z}}}_{ig}U_{ig}\right)\\ &= \left(\frac{1}{G}\sum_{g=1}^{G}\mathbf{A}_g + \frac{1}{G}\sum_{g=1}^{G}\mathbf{R}_g^{(1)}\right)^{-1}\left(\frac{1}{G}\sum_{g=1}^{G}\mathbf{P}_g + \frac{1}{G}\sum_{g=1}^{G}\mathbf{R}_g^{(2)}\right)\end{aligned}$$

where we define

$$\mathbf{A}_g \equiv \frac{1}{N_g}\sum_{i=1}^{N_g} \rho_g \widehat{\boldsymbol{\mathcal{Z}}}_{ig} \boldsymbol{X}'_{ig} \qquad \mathbf{R}_g^{(1)} \equiv \frac{1}{N_g}\sum_{i=1}^{N_g} \rho_g (\widehat{\boldsymbol{\mathcal{Z}}}_{ig} - \boldsymbol{\mathcal{Z}}_{ig}) \boldsymbol{X}'_{ig}$$

$$\mathbf{P}_g \equiv \frac{1}{N_g}\sum_{i=1}^{N_g} \rho_g \widehat{\boldsymbol{\mathcal{Z}}}_{ig} U'_{ig} \qquad \mathbf{R}_g^{(2)} \equiv \frac{1}{N_g}\sum_{i=1}^{N_g} \rho_g (\widehat{\boldsymbol{\mathcal{Z}}}_{ig} - \boldsymbol{\mathcal{Z}}_{ig}) U_{ig}.$$

By assumption, both $||\sum_{g=1}^G \mathbf{R}_g^{(1)}||$ and $||\sum_{g=1}^G \mathbf{R}_g^{(2)}||$ are $o_{\mathbb{P}}(G)$ and thus

$$\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta} = \left(\frac{1}{G}\sum_{g=1}^G \mathbf{A}_g + o_{\mathbb{P}}(1)\right)^{-1} \left(\frac{1}{G}\sum_{g=1}^G \mathbf{P}_g + o_{\mathbb{P}}(1)\right)$$

Now, since we observe a random sample of groups and $\mathbf{A}_g$ is a group-level random variable

$$\mathbb{E}\left[\frac{1}{G}\sum_{g=1}^G \mathbf{A}_g\right] = \mathbb{E}(\mathbf{A}_g) = \mathbb{E}\left[\frac{1}{N_g}\sum_{i=1}^{N_g} \mathbb{E}\left(\rho_g \boldsymbol{\mathcal{Z}}_{ig} \mathbf{X}'_{ig} | N_g\right)\right] = \mathbb{E}\left[\mathbb{E}\left(\rho_g \boldsymbol{\mathcal{Z}}_{ig} \mathbf{X}'_{ig} | N_g\right)\right] = \mathbb{E}(\rho_g \boldsymbol{\mathcal{Z}}_{ig} \mathbf{X}'_{ig})$$

where the second equality uses iterated expectations and linearity, the third uses the assumption of identical distribution within groups, and the the fourth uses iterated expectations a second time. Now consider an arbitrary entry $A_g^{(j,k)}$ of the matrix $\mathbf{A}_g$ and let $\|\cdot\|_F$ denote the Frobenius norm. By the triangle and Cauchy-Schwarz inequalities, and using the assumption of identical distribution with group, we have

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{G}\sum_{g=1}^G A_g^{(j,k)}\right) &= \frac{1}{G}\mathrm{Var}\left(A_g^{(j,k)}\right) \leq \frac{1}{G}\mathbb{E}\left[||\mathbf{A}_g||_F^2\right] = \frac{1}{G}\mathbb{E}\left(\frac{1}{N_g^2}\left\|\sum_{i=1}^{N_g} \rho_g \boldsymbol{\mathcal{Z}}_{ig} \mathbf{X}'_{ig}\right\|_F^2\right) \\
&\leq \frac{1}{G}\mathbb{E}\left[\frac{1}{N_g^2}\left(\sum_{i=1}^{N_g} \left\|\rho_g \boldsymbol{\mathcal{Z}}_{ig} \mathbf{X}'_{ig}\right\|_F\right)^2\right] \\
&= \frac{1}{G}\mathbb{E}\left[\frac{1}{N_g^2}\mathbb{E}\left(\sum_{i,j\leq N_g} \left\|\rho_g \boldsymbol{\mathcal{Z}}_{ig} \mathbf{X}'_{ig}\right\|_F \left\|\rho_g \boldsymbol{\mathcal{Z}}_{jg} \mathbf{X}'_{jg}\right\|_F \middle| N_g\right)\right] \\
&\leq \frac{1}{G}\mathbb{E}\left[\frac{1}{N_g^2}\mathbb{E}\left(\sum_{i,j\leq N_g} \left\|\rho_g \boldsymbol{\mathcal{Z}}_{ig} \mathbf{X}'_{ig}\right\|_F^2 \middle| N_g\right)\right] \\
&= \frac{1}{G}\mathbb{E}\left[\mathbb{E}\left(\left\|\rho_g \boldsymbol{\mathcal{Z}}_{ig} \mathbf{X}'_{ig}\right\|_F^2 \middle| N_g\right)\right] = \frac{1}{G}\mathbb{E}\left[\rho_g^2 \left\|\boldsymbol{\mathcal{Z}}_{ig} \mathbf{X}'_{ig}\right\|_F^2\right] \to 0
\end{aligned}$$

since all finite-dimensional norms are equivalent and $\mathbb{E}\left[\rho_g^2 \left\|\boldsymbol{\mathcal{Z}}_{ig} \mathbf{X}'_{ig}\right\|_F^2\right] = o(G)$. Hence, by the L^2 weak law of large numbers $G^{-1}\sum_{g=1}^G \mathbf{A}_g \to_p \mathbb{E}(\rho_g \boldsymbol{\mathcal{Z}}_{ig} \mathbf{X}'_{ig}) = \mathbb{I}$. An analogous argument shows that $G^{-1}\sum_{g=1}^G \mathbf{P}_g \to_p \mathbb{E}(\rho_g \boldsymbol{\mathcal{Z}}_{ig} U_{ig}) = \mathbf{0}$. The result follows by the continuous mapping theorem. □

Proof of Theorem 5. Continuing the argument from the proof of Theorem 4, we have

$$\sqrt{G}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) = [\mathbb{I} + o_{\mathbb{P}}(1)]^{-1} \left(\frac{1}{\sqrt{G}} \sum_{g=1}^{G} \mathbf{P}_g + \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \mathbf{R}_g^{(2)} \right).$$

By assumption, $||\sum_{g=1}^{G} \mathbf{R}_g^{(2)}|| = o_{\mathbb{P}}(G^{1/2})$, and hence $\sqrt{G}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) = \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \mathbf{P}_g + o_{\mathbb{P}}(1)$. Thus, it suffices to apply the Lindeberg-Feller central limit theorem to $\mathbf{P}_g/\sqrt{G}$. Because we observe a random sample of groups, $\text{Var}(\sum_{g=1}^{G} \mathbf{P}_g/\sqrt{G}) = \text{Var}(\mathbb{P}_g)$ which by assumption converges to Σ. All that remains is to verify the Lindeberg condition, namely

$$\mathbb{E}\left[||\mathbf{P}_g||^2 \mathbb{1}\left\{ ||\mathbf{P}_g|| > \varepsilon\sqrt{G} \right\} \right] \to 0$$

for any $\varepsilon > 0$. A sufficient condition for this to hold is $G^{-\delta/2}\mathbb{E}\left[||\mathbf{P}_g||^{2+\delta} \right] \to 0$ for some $\delta > 0$. By an argument similar to that used to establish $\mathbb{E}\left[||\mathbf{A}_g||_F^2 \right] \leq \mathbb{E}\left[\rho_g^2 ||\boldsymbol{\mathcal{Z}}_{ig} \mathbf{X}_{ig}'||_F^2 \right]$ in the proof of Theorem 4, we likewise have

$$G^{-\delta/2}\mathbb{E}\left[||\mathbf{P}_g||^{2+\delta} \right] \leq G^{-\delta/2}\mathbb{E}\left[\rho_g^{2+\delta} ||\boldsymbol{\mathcal{Z}}_{ig} \mathbf{X}_{ig}'||^{2+\delta} \right] = o(1)$$

so the result follows. □

Lemma A.3. *Let $\bar{Z}_g \equiv \sum_{j=1}^{N_g} Z_{jg}/N_g$. Under the conditions of Lemma 4,*

$$\mathbb{P}(\bar{Z}_g < \underline{s}/2) \leq \exp\left\{ -\underline{n}\underline{s}^2/2 \right\}.$$

Proof of Lemma A.3. Conditional on $(N_g = n, S_g = s)$, the treatment offers $(Z_1, \ldots, Z_{N_g})$ are a collection of n iid Bernoulli(s) random variables by Assumption 2. Hence, by Hoeffding's inequality

$$\mathbb{P}\left(\bar{Z}_g < \underline{s}/2 | N_g = n, S_g = s \right) \leq \exp\left\{ -2n(s - \underline{s}/2)^2 \right\} \leq \exp\left\{ -\underline{n}\underline{s}^2/2 \right\}$$

where the second inequality follows since $\underline{s} \leq s$. Thus,

$$\mathbb{P}(\bar{Z}_g < \underline{s}/2) = \sum_{n,s} \mathbb{P}(\bar{Z}_g \leq \underline{s}/2 | N_g = n, S_g = s) \mathbb{P}(N_g = n, S_g = s) \leq \exp\left\{ -2\underline{n}\underline{s}^2/4 \right\}$$

by the law of total probability. The result follows since $\mathbb{P}(\bar{Z}_g < \underline{s}/2) \leq \mathbb{P}(Z_g \leq \underline{s}/2)$. □

Lemma A.4. *Let $\bar{C}_g = \sum_{j=1}^{N_g} C_{jg}/N_g$ and $\widehat{C}_g \equiv \sum_{j=1}^{N_g} D_{jg}/N_g \bar{Z}_g$, where $\bar{Z}_g$ is as defined in Lemma A.3. Under the conditions of Lemma 4 and for any $t > 0$,*

$$\mathbb{P}\left(\left| \widehat{C}_g - \bar{C}_g \right| \geq t \middle| \bar{Z}_g \geq \underline{s}/2 \right) \leq 2\exp\left\{ -\underline{n}\underline{s}^2 t^2/2 \right\}.$$

Proof of Lemma A.4. Let $\mathcal{A} \equiv \left\{ \mathbf{C}_g = \mathbf{c}, N_g = n, \bar{C}_g = \bar{c}, N_g \bar{Z}_g = m, S_g = s \right\}$ where $m > 0$. Suppose first that $\bar{c} \neq 0$. In this case

$$\mathbb{P}\left(\left| \widehat{C}_g - \bar{C}_g \right| > t \middle| \mathcal{A} \right) = \mathbb{P}\left(\left| \sum_{j=1}^{n} \frac{c_j Z_{jg}}{m} - \bar{c} \right| > t \middle| \mathcal{A} \right) = \mathbb{P}\left(\left| \frac{1}{n\bar{c}} \sum_{j \in \mathcal{C}} Z_{jg}^* - \bar{c} \right| > t \middle| \mathcal{A} \right)$$

where $\mathcal{C} \equiv \{j : c_j = 1\}$ and $Z_{jg}^* \equiv n\bar{c} Z_{jg}/m$. Given $\mathcal{A}$, the $\{Z_{jg}\}_{j \in \mathcal{C}}$ are a sequence of $n\bar{c}$ draws

made without replacement from a population of m ones and $(n-m)$ zeros. Thus

$$\mathbb{E}(Z_{jg}^*) = \frac{n\bar{c}}{m}\mathbb{P}(Z_{jg}=1|\mathcal{A}) = \frac{n\bar{c}}{m}\cdot\frac{m}{n} = \bar{c}.$$

Moreover, since $Z_{jg} \in \{0,1\}$, each of the Z_{jg}^* is bounded between 0 and $n\bar{c}/m$. While these random variables are identically distributed, they are not independent—like the Z_{jg} from which they are constructed, $\left\{Z_{jg}^*\right\}_{j\in\mathcal{C}}$ are draws made without replacement from a finite population. Under this form of dependence, however, Hoeffding's Inequality continues to apply (Hoeffding, 1963, p. 28) and hence

$$\mathbb{P}\left(\left|\widehat{C}_g - \bar{C}_g\right| > t \middle| \mathcal{A}\right) \leq 2\exp\left\{\frac{-2t^2m^2}{n\bar{c}}\right\} \leq 2\exp\left\{-2n\left(\frac{m}{n}\right)^2 t^2\right\}$$

where the second inequality follows because $0 < \bar{c} \leq 1$. If $\bar{c} = 0$, we have

$$\mathbb{P}\left(\left|\widehat{C}_g - \bar{C}_g\right| > t \middle| \mathcal{A}\right) = \mathbb{P}(|0-0| > t|\mathcal{A}) = 0 \leq 2\exp\left\{-2n\left(\frac{m}{n}\right)^2 t^2\right\}$$

so this inequality holds for any $\bar{c}$. Applying the law of total probability as in the proof of Lemma A.3, we see that

$$\mathbb{P}\left(\left|\widehat{C}_g - \bar{C}_g\right| > t \middle| N_g = n, N_g\bar{Z}_g = m\right) \leq 2\exp\left\{-2n\left(\frac{m}{n}\right)^2 t^2\right\}$$

and thus

$$\begin{aligned}
\mathbb{P}\left(\left|\widehat{C}_g - \bar{C}_g\right| \geq t \middle| \bar{Z}_g \geq \underline{s}/2\right) &= \sum_{\{(m,n):\ \frac{m}{n}\geq \underline{s}/2\}} \mathbb{P}\left(\left|\widehat{C}_g - \bar{C}_g\right| > t \middle| N_g = n, N_g\bar{Z}_g = m\right) \\
&\quad \times \mathbb{P}(N_g\bar{Z}_g = m, N_g = n|\bar{Z}_g \geq \underline{s}/2) \\
&\leq \sum_{\{(m,n):\ \frac{m}{n}\geq \underline{s}/2\}} 2\exp\left\{-2n\left(\frac{m}{n}\right)^2 t^2\right\} \mathbb{P}(N_g\bar{Z}_g = m, N_g = n|\bar{Z}_g \geq \underline{s}/2) \\
&\leq \sum_{\{(m,n):\ \frac{m}{n}\geq \underline{s}/2\}} 2\exp\left\{-\underline{n}\underline{s}^2t^2/2\right\} \mathbb{P}(N_g\bar{Z}_g = m, N_g = n|\bar{Z}_g \geq \underline{s}/2) \\
&= \exp\left\{-\underline{n}\underline{s}^2t^2/2\right\}
\end{aligned}$$

by a second application of the law of total probability, since $\underline{n} \leq N_g$. □

Lemma A.5. *Suppose that $\underline{sn} > 2$. Then, under the conditions of Lemma 4,*

$$\mathbb{P}\left(\max_{1\leq i\leq N_g}\left|\widehat{C}_{ig} - \bar{C}_{ig}\right| > t \middle| \bar{Z}_g \geq \underline{s}/2\right) \leq 2\exp\left\{-\underline{n}\underline{s}^2 h(\underline{sn}, t)^2/2\right\}$$

where we define

$$h(x,t) \equiv \left(\frac{x-2}{x}\right)^2 t - \left[1 - \left(\frac{x-2}{x}\right)^2\right]\frac{4}{x-2}.$$

Proof of Lemma A.5. If $\bar{Z}_g > \underline{s}/2 > 1/\underline{n}$, then $N_g\bar{Z}_g - Z_{ig} > 0$ and $N_g\bar{Z}_g > 0$. Hence,

$$\widehat{C}_{ig} \equiv \frac{\bar{D}_{ig}}{\bar{Z}_{ig}} = \frac{N_g\bar{D}_g - D_{ig}}{N_g\bar{Z}_g - Z_{ig}} = \frac{N_g\bar{Z}_g\widehat{C}_g - D_{ig}}{N_g\bar{Z}_g - Z_{ig}} = \left(\frac{N_g\bar{Z}_g}{N_g\bar{Z}_g - Z_{ig}}\right)\widehat{C}_g - \frac{D_{ig}}{N_g\bar{Z}_g - Z_{ig}}.$$

Similar manipulations give

$$\bar{C}_{ig} = \left(\frac{N_g}{N_g - 1}\right)\bar{C}_g - \frac{C_{ig}}{N_g - 1}$$

from which it follows that

$$\left|\widehat{C}_{ig} - \bar{C}_{ig}\right| \leq \left|\left(\frac{N_g \bar{Z}_g}{N_g \bar{Z}_g - Z_{ig}}\right)\widehat{C}_g - \left(\frac{N_g}{N_g - 1}\right)\bar{C}_g\right| + \left|\frac{C_{ig}}{N_g - 1} - \frac{D_{ig}}{N_g \bar{Z}_g - Z_{ig}}\right|$$

by the triangle inequality. Using the fact that Z_{ig}, D_{ig}, and C_{ig} are binary along with $\underline{n} \leq N_g$ and $\bar{Z}_g > \underline{s}/2 > 1/\underline{n}$, tedious but straightforward algebra allows us to bound the right-hand side of the preceding inequality from above, yielding

$$\left|\widehat{C}_{ig} - \bar{C}_{ig}\right| \leq \left(\frac{\underline{sn}}{\underline{sn} - 2}\right)^2 \left|\widehat{C}_g - \bar{C}_g\right| + \left[\left(\frac{\underline{sn}}{\underline{sn} - 2}\right)^2 + 1\right]\frac{4}{\underline{sn} - 2}.$$

Since this upper bound for $|\widehat{C}_{ig} - \bar{C}_{ig}|$ does not depend on i, it follows that

$$\max_{1 \leq i \leq N_g}\left|\widehat{C}_{ig} - \bar{C}_{ig}\right| \leq \left(\frac{\underline{sn}}{\underline{sn} - 2}\right)^2 \left|\widehat{C}_g - \bar{C}_g\right| + \left[\left(\frac{\underline{sn}}{\underline{sn} - 2}\right)^2 + 1\right]\frac{4}{\underline{sn} - 2}$$

provided that $\bar{Z}_g > \underline{s}/2 > 1/\underline{n}$. In other words, so long as $\underline{sn} > 2$ we have

$$\left\{\bar{Z}_g \geq \underline{s}/2\right\} \cap \left\{\max_{1 \leq i \leq N_g}\left|\widehat{C}_{ig} - \bar{C}_{ig}\right| > t\right\} \subseteq \left\{\bar{Z}_g > \underline{s}/2\right\} \cap \left\{\left|\widehat{C}_g - \bar{C}_g\right| > h(\underline{sn}, t)\right\}.$$

Therefore, by the monotonicity of probability

$$\mathbb{P}\left(\max_{1 \leq i \leq N_g}\left|\widehat{C}_{ig} - \bar{C}_{ig}\right| > t \,\middle|\, \bar{Z}_g \geq \underline{s}/2\right) \leq \mathbb{P}\left(\left|\widehat{C}_g - \bar{C}_g\right| > h(\underline{sn}, t) \,\middle|\, \bar{Z}_g \geq \underline{s}/2\right)$$

and the result follows by Lemma A.4.

□

Proof of Lemma 4. By the law of total probability, Lemma A.4, and Lemma A.5

$$\begin{aligned}\mathbb{P}\left(\max_{1 \leq i \leq N_g}\left|\widehat{C}_{ig} - \bar{C}_{ig}\right| > t\right) &\leq \mathbb{P}\left(\max_{1 \leq i \leq N_g}\left|\widehat{C}_{ig} - \bar{C}_{ig}\right| > t \,\middle|\, \bar{Z}_g \geq \underline{s}/2\right) + \mathbb{P}(\bar{Z}_g < \underline{s}/2) \\ &\leq 2\exp\left\{-\underline{n}\underline{s}^2 h(\underline{sn}, t)^2/2\right\} + \exp\left\{-\underline{n}\underline{s}^2/2\right\}\end{aligned}$$

where $h(\cdot, \cdot)$ is as defined in Lemma A.5. Expanding and simplifying, we see that

$$h(\underline{sn}, t)^2 \geq \left(\frac{\underline{sn} - 2}{\underline{sn}}\right)^4 t^2 - \frac{16t}{\underline{sn} - 2} \equiv h^*(\underline{sn}, t).$$

Now, for any $t \geq 1$ we have $\mathbb{P}\left(\max_{1 \leq i \leq N_g}\left|\widehat{C}_{ig} - \bar{C}_{ig}\right| > t\right)$ since both $\widehat{C}_{ig}$ and $\bar{C}_{ig}$ are between

zero and one. Since $h^*(\underline{sn}, t) < 1$ for any $t < 1$, it follows that

$$\begin{aligned}\mathbb{P}\left(\max_{1\le i\le N_g}\left|\widehat{C}_{ig}-\bar{C}_{ig}\right|>t\right) &\le 2\exp\left\{-\underline{ns}^2 h(\underline{sn},t)^2/2\right\}+\exp\left\{-\underline{ns}^2/2\right\}\\ &\le 2\exp\left\{-\underline{ns}^2 h^*(\underline{sn},t)/2\right\}+\exp\left\{-\underline{ns}^2/2\right\}\\ &\le 3\exp\left\{-\underline{ns}^2 h^*(\underline{sn},t)/2\right\}\end{aligned}$$

Applying the union bound we obtain

$$\begin{aligned}\mathbb{P}\left(\max_{1\le g\le G}\max_{1\le i\le N_g}\left|\widehat{C}_{ig}-\bar{C}_{ig}\right|>t\right) &= \mathbb{P}\left(\bigcup_{g=1}^{G}\left\{\max_{1\le i\le N_g}\left|\widehat{C}_{ig}-\bar{C}_{ig}\right|>t\right\}\right)\\ &\le \sum_{g=1}^{G}\mathbb{P}\left(\max_{1\le i\le N_g}\left|\widehat{C}_{ig}-\bar{C}_{ig}\right|>t\right)\\ &\le \sum_{g=1}^{G}3\exp\left\{-\underline{ns}^2 h^*(\underline{sn},t)/2\right\}\\ &= 3G\exp\left\{-\underline{ns}^2 h^*(\underline{sn},t)/2\right\}\end{aligned}$$

and accordingly we have

$$\begin{aligned}\mathbb{P}\left(\frac{\max_{1\le g\le G}\max_{1\le i\le N_g}\left|\widehat{C}_{ig}-\bar{C}_{ig}\right|}{\sqrt{\frac{\log G}{\underline{n}}}}>M\right) &\le 3G\exp\left\{\frac{-\underline{ns}^2}{2}\left[\left(\frac{\underline{sn}-2}{\underline{sn}}\right)^4\frac{\log G}{\underline{n}}M^2-\frac{16}{\underline{sn}-2}\sqrt{\frac{\log G}{\underline{n}}}M\right]\right\}\\ &= 3G\exp\left\{\log G\left[1-\frac{\underline{s}^2}{2}\left(\frac{\underline{sn}-2}{\underline{sn}}\right)^4M^2-\frac{16}{\underline{sn}-2}\sqrt{\frac{1}{\underline{n}\log G}}\right]\right\}.\end{aligned}$$

The expression on the right-hand side converges to $3\exp\left\{\log G\left[1-\underline{s}^2M^2/2\right]\right\}$ as $(\underline{n}, G)\to\infty$ and hence can be made arbitrarily small by choosing a sufficiently large value of M. □

Proof of Theorem 6. We provide the argument for condition (vii) of Theorem 4 and (iii) of Theorem 5 only. For (vi) from Theorem 4, simply replace U_{ig} with $\boldsymbol{X}_{ig}$ in the following derivations. By (20) and the triangle inequality

$$\left\|\sum_{g=1}^{G}\frac{1}{N_g}\sum_{i=1}^{N_g}\rho_g(\widehat{\boldsymbol{\mathcal{Z}}}_{ig}-\boldsymbol{\mathcal{Z}}_{ig})U_{ig}\right\|\le\Delta_G\left(\sum_{g=1}^{G}\frac{1}{N_g}\sum_{i=1}^{N_g}||\rho_g\boldsymbol{W}_{ig}U_{ig}||\right)\tag{A.9}$$

where we define the shorthand

$$\Delta_G\equiv\max_{1\le g\le G}\left(\max_{1\le i\le N_g}\left\|\mathbf{R}(\widehat{C}_{ig},N_g)^{+}-\mathbf{R}(\bar{C}_{ig},N_g)^{-1}\right\|\right).$$

Consider the second factor on the RHS of (A.9). By an argument similar to that used in the proof

of Theorem 4,

$$\frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{N_g}\sum_{i=1}^{N_g}||\rho_g \boldsymbol{W}_{ig}U_{ig}||\right) \to_p \mathbb{E}\left[||\rho \boldsymbol{W}_{ig}U_{ig}||\right] < \infty$$

so that $\sum_{g=1}^{G}\frac{1}{N_g}\sum_{i=1}^{N_g}||\rho_g \boldsymbol{W}_{ig}U_{ig}|| = O_{\mathbb{P}}(G)$. Now, define the event $\widehat{1}_G$ as

$$\widehat{1}_G \equiv \mathbb{1}\left\{\min_{1\le g\le G}\left(\min_{1\le i\le N_g}\widehat{C}_{ig}\right) \ge \frac{\bar{c}_L}{2}\right\}.$$

By assumption $\mathbf{R}(\bar{C}_{ig}, N_{ig})$ is invertible, and conditional on $\widehat{C}_{ig} \ge \bar{c}_L/2$ it follows that $\mathbf{R}(\widehat{C}_{ig}, N_g)$ is likewise invertible. Hence, if $\widehat{1}_G = 1$ we can write

$$\begin{aligned}\left\|\mathbf{R}(\widehat{C}_{ig}, N_g)^{-1} - \mathbf{R}(\bar{C}_{ig}, N_g)^{-1}\right\| &= \left\|\mathbf{R}(\widehat{C}_{ig}, N_g)^{-1}\left[\mathbf{R}(\widehat{C}_{ig}, N_g) - \mathbf{R}(\bar{C}_{ig}, N_g)\right]\mathbf{R}(\bar{C}_{ig}, N_g)^{-1}\right\| \\ &\le \left\|\mathbf{R}(\widehat{C}_{ig}, N_g)^{-1}\right\|\left\|\mathbf{R}(\widehat{C}_{ig}, N_g) - \mathbf{R}(\bar{C}_{ig}, N_g)\right\|\left\|\mathbf{R}(\bar{C}_{ig}, N_g)^{-1}\right\|.\end{aligned}$$

Let $||\mathbf{M}||_2$ denote the spectral norm of a matrix $\mathbf{M}$, i.e. its largest singular value. Since $\mathbf{R}(\bar{C}_{ig}, N_g)$ is square, symmetric, and positive definite we have $||\mathbf{R}(\bar{C}_{ig}, N_g)^{-1}||_2 \le 1/\underline{\sigma} < \infty$. Similarly, if $\widehat{1}_G = 1$, then $||\mathbf{R}(\widehat{C}_{ig}, N_g)^{-1}||_2 \le 1/\underline{\sigma} < \infty$. Because all finite-dimensional norms are equivalent, it follows that

$$\widehat{1}_G\Delta_G \le K\max_{1\le g\le G}\left(\max_{1\le i\le N_g}\left\|\mathbf{R}(\widehat{C}_{ig}, N_g) - \mathbf{R}(\bar{C}_{ig}, N_g)\right\|\right) \le K\left\{\max_{1\le g\le G}\left(\max_{1\le i\le N_g}\left|\widehat{C}_{ig} - \bar{C}_{ig}\right|\right) + O(\underline{n}^{-1/2})\right\}$$

where $0 < K < \infty$ denotes a generic, unspecified constant. Applying Lemma 4 we see that $\widehat{1}_G\Delta_G = O_{\mathbb{P}}\left(\sqrt{\log G/\underline{n}}\right)$ as $(\underline{n}, G) \to \infty$. Thus, by (A.9),

$$\widehat{1}_G\left\|\sum_{g=1}^{G}\frac{1}{N_g}\sum_{i=1}^{N_g}\rho_g(\widehat{\boldsymbol{\mathcal{Z}}}_{ig} - \boldsymbol{\mathcal{Z}}_{ig})U_{ig}\right\| = O_{\mathbb{P}}\left(\sqrt{\frac{\log G}{\underline{n}}}\right)O_{\mathbb{P}}(G). \tag{A.10}$$

If $\log G/\underline{n} \to 0$ as $(\underline{n}, G) \to \infty$, then the rate on the RHS of (A.10) becomes $o_{\mathbb{P}}(G)$. If $G\log G/\underline{n} \to 0$, it becomes $o_{\mathbb{P}}(G^{1/2})$. Finally, since $\bar{c}_L \le \bar{C}_{ig}$, it follows that

$$\mathbb{P}\left(\widehat{1}_G \ne 1\right) \le \mathbb{P}\left[\max_{1\le g\le G}\left(\max_{1\le i\le N_g}\left|\widehat{C}_{ig} - \bar{C}_{ig}\right| \ge \frac{\bar{c}_L}{2}\right)\right]$$

Hence, applying Lemma 4, $\log G/\underline{n} \to 0$ implies $\widehat{1}_G \to_p 1$. The result follows. □

B Additional Tables and Figures

Age	0.004
	(0.002)
Cohabits	-0.02
	(0.010)
Has at least one child	-0.13
	(0.032)
Youngest child: 12+ months	0.12
	(0.027)
Education: less than Bac+2 years	-0.03
	(0.012)
Employed at baseline	-0.09
	(0.019)
Not employed at baseline	0.03
	(0.015)
Permanent contract at baseline	-0.14
	(0.017)
Fixed term contract at baseline	-0.06
	(0.015)
Duration of contract at baseline: 7-12 months	-0.04
	(0.018)
Duration of contract at baseline: 13+ months	-0.12
	(0.028)
Receives unemployment insurance at baseline	0.04
	(0.009)
Mean compliance	0.35
Observations	11,976
R^2	0.055

Table B.1: Predictors of compliance: linear probability model. OLS estimates of compliance indicators on baseline covariates, estimated on the subsample of participants assigned to treatment. Standard errors clustered at the city level. The following variables are included in the regression but are not reported and are not statistically significant: sex; number of children; youngest child 0-4 months, 4-8 months, 8-12 months; unemployment duration at baseline; did not provide employment status at baseline; unemployment duration in the last 18 months; temporary contract at baseline; 1-3 month contract at baseline; 3-6 month contract at baseline; average city unemployment rate.

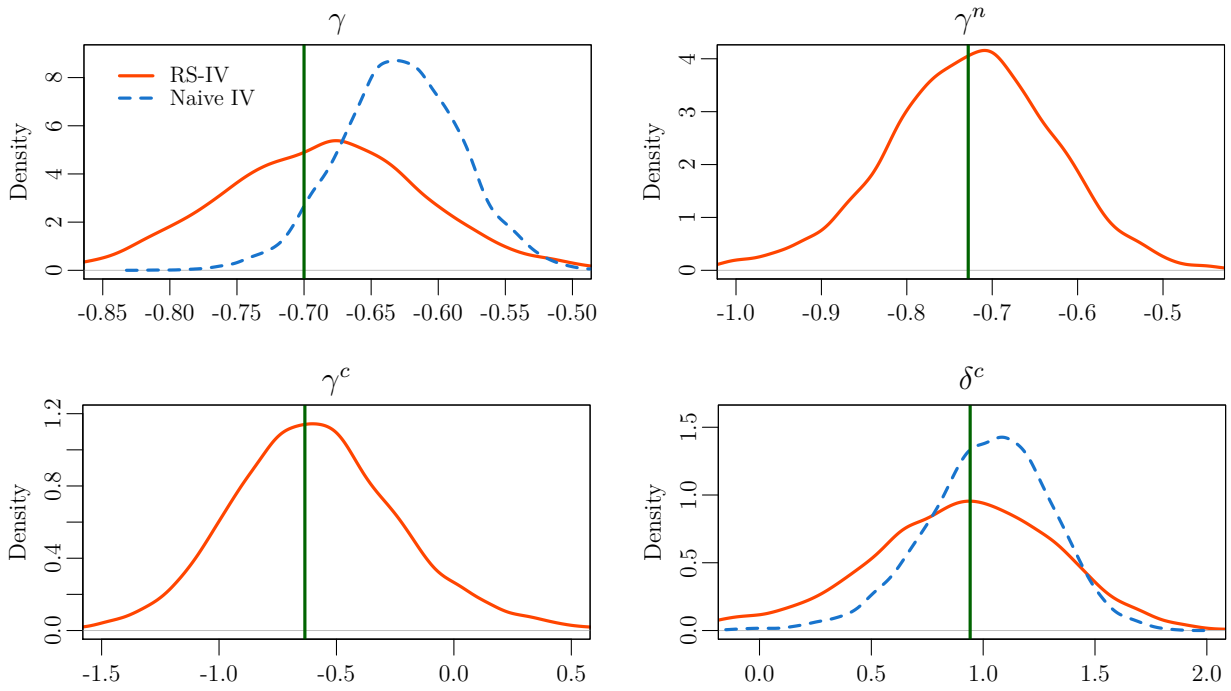

Figure B.1: Distribution of the estimates of the spillover terms, $(\gamma, \gamma^n, \gamma^c, \delta^c)$, over 5000 simulations for our IV and the 'naive' IV (where available) for simulations with 150 groups.

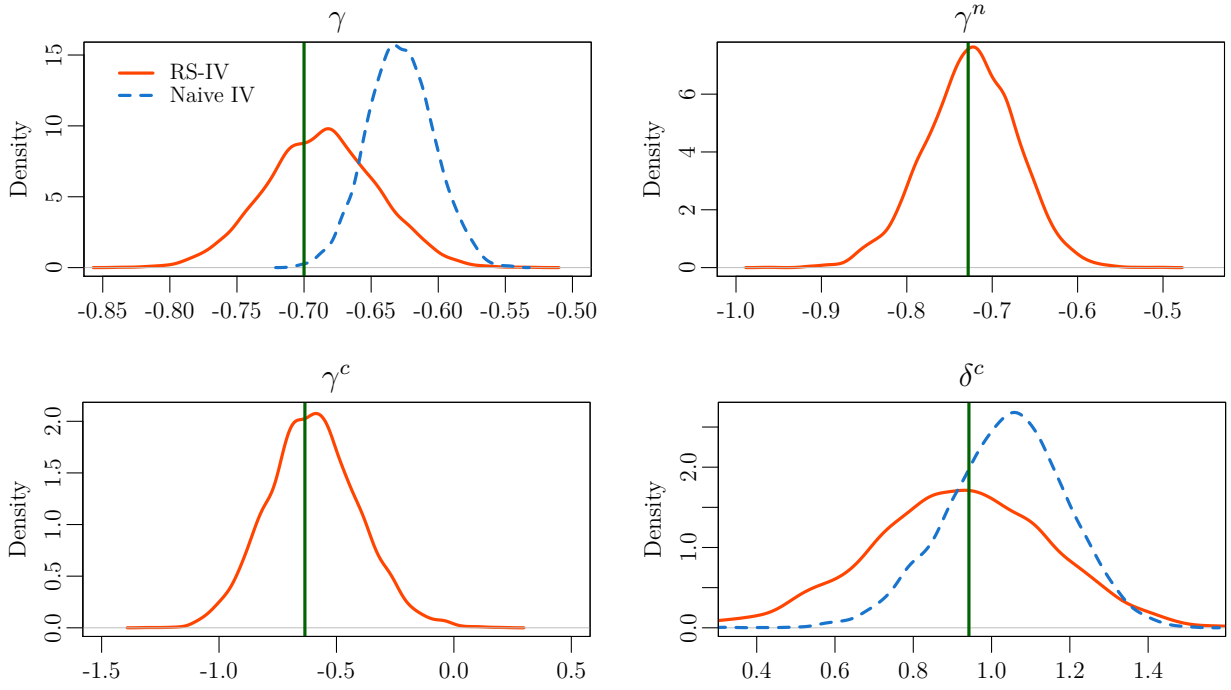

Figure B.2: Distribution of the estimates of the spillover terms, $(\gamma, \gamma^n, \gamma^c, \delta^c)$, over 5000 simulations for our IV and the 'naive' IV (where available) for simulations with 500 groups.

C Experiments with a 0% Saturation

Some randomized saturation designs, including the experiment of Crépon et al. (2013), include a zero percent saturation, also known as a "pure control" condition. Under one-sided non-compliance $S_g = 0$ implies $Z_{ig} = D_{ig} = \bar{D}_{ig} = 0$ for all $1 \leq i \leq N_g$. Accordingly, we cannot estimate the share of compliers $\widehat{C}_{ig}$ from (17) for groups assigned a saturation of zero. The easiest solution to this problem is simply to drop observations for any zero saturation groups. Under Assumptions 1–2 and 6 this has no effect on our identification or large-sample results provided that we replace $\mathbf{Q}, \mathbf{Q}_0$ and $\mathbf{Q}_1$ with expectations that condition on $S_g > 0$, namely

$$\begin{aligned}
\widetilde{\mathbf{Q}}(\bar{c}, n) &\equiv \mathbb{E}\left[\mathbf{W}_{ig}\mathbf{W}_{ig}' | \bar{C}_{ig} = \bar{c}, N_g = n, S_g > 0\right] \\
\widetilde{\mathbf{Q}}_0(\bar{c}, n) &\equiv \mathbb{E}\left[(1 - Z_{ig})\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})' | \bar{C}_{ig} = \bar{c}, N_g = n, S_g > 0\right] \\
\widetilde{\mathbf{Q}}_1(\bar{c}, n) &\equiv \mathbb{E}\left[Z_{ig}\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})' | \bar{C}_{ig} = \bar{c}, N_g = n, S_g > 0\right]
\end{aligned}$$

Zero percent saturation groups, however, *are* informative: they pin down the value of $\mathbb{E}[Y_{ig}(0,0)]$ and hence can be used to improve estimates of $\mathbb{E}[\boldsymbol{\theta}_{ig}]$. To exploit this information, we replace the instrument vectors from parts (i) and (iv) of Theorem 2 with

$$\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}^{W} \equiv \begin{bmatrix} \mathbb{1}\{S_g > 0\}\, \widetilde{\mathbf{Q}}(\bar{C}_{ig}, N_g)^{-1}\mathbf{W}_{ig} \\ \mathbb{1}\{S_g = 0\} \end{bmatrix}, \qquad \widetilde{\boldsymbol{\mathcal{Z}}}_{ig}^{0} \equiv \begin{bmatrix} \mathbb{1}\{S_g > 0\}\, \widetilde{\mathbf{Q}}_0(\bar{C}_{ig}, N_g)^{-1}\mathbf{f}(\bar{D}_{ig}) \\ \mathbb{1}\{S_g = 0\} \end{bmatrix}$$

Calculations similar to those in the proof of Theorem 2 establish that these are valid and relevant instruments. Because the dimensions of $\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}^{W}$ and $\widetilde{\boldsymbol{\mathcal{Z}}}_{ig}^{0}$ exceed those of the parameters for which they instrument by one, they provide over-identifying information. As such, the just-identified IV moment condition from parts (i) and (iv) of Theorem 2 must be replaced with a linear GMM moment equation. Subject to this small change, estimation and inference can proceed almost exactly as in section 4: we merely substitute $\widehat{C}_{ig}$ for $\bar{C}_{ig}$ in $\widetilde{\mathbf{Q}}$ and $\widetilde{\mathbf{Q}}_0$ to yield a feasible GMM estimator, e.g. two-stage least squares. With minor notational modifications, our large-sample results continue to apply.

D Extending the Definition of Q

Technically, the conditional expectations in (8)–(10) are only well-defined when $n\bar{c}$ is a positive integer, whereas Assumption 8 requires the functions $\mathbf{Q}, \mathbf{Q}_0$, and $\mathbf{Q}_1$ to be defined over a continuous range of values for $\bar{c}$. This problem is easily solved by *extending* the definitions of $\mathbf{Q}_0$ and $\mathbf{Q}_1$. In many cases, the natural extension will be obvious. In the linear potential outcomes model, for example, (15) and (16) agree with (9) and (10) when these conditional expectations are well-defined and satisfy all the conditions of Assumption 8

More generally, we can always *construct* extended definitions of $\mathbf{Q}_0$ and $\mathbf{Q}_1$ to satisfy these regularity conditions. Here we provide one such construction based on *linear interpolation*. Let

$$\bar{c}_\ell(\bar{c}, n) \equiv \frac{\lfloor (n-1)\bar{c} \rfloor}{n-1}, \quad \bar{c}_u(\bar{c}, n) \equiv \frac{\lceil (n-1)\bar{c} \rceil}{n-1}.$$

By construction, $(n-1)\bar{c}_u(\bar{c},n)$ and $(n-1)\bar{c}_\ell(\bar{c},n)$ are always non-negative integers. Now let

$$\mathbf{Q}_z^\ell(\bar{c},n) \equiv \mathbb{E}\left[\mathbb{1}(Z_{ig}=z)\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})' \middle| \bar{C}_{ig}=\bar{c}_\ell(\bar{c},n), N_g=n\right]$$
$$\mathbf{Q}_z^u(\bar{c},n) \equiv \mathbb{E}\left[\mathbb{1}(Z_{ig}=z)\mathbf{f}(\bar{D}_{ig})\mathbf{f}(\bar{D}_{ig})' \middle| \bar{C}_{ig}=\bar{c}_u(\bar{c},n), N_g=n\right]$$

for $z=0,1$. Notice that $\mathbf{Q}_0^\ell, \mathbf{Q}_1^\ell$ and $\mathbf{Q}_0^u, \mathbf{Q}_1^u$ are well-defined regardless of whether $(n-1)\bar{c}$ is an integer. From these ingredients, we construct extended definitions $\mathbf{Q}_0^*$ and $\mathbf{Q}_1^*$ of $\mathbf{Q}_0, \mathbf{Q}_1$ as

$$\mathbf{Q}_z^*(\bar{c},n) = [1-\omega(\bar{c},n)]\,\mathbf{Q}_z^\ell(\bar{c},n) + \omega(\bar{c},n)\mathbf{Q}_z^u(\bar{c},n); \quad \omega(\bar{c},n) \equiv \frac{\bar{c}-\bar{c}_\ell(\bar{c},n)}{\bar{c}_u(\bar{c},n)-\bar{c}_\ell(\bar{c},n)} \in [0,1]$$

for $z=0,1$. Since both $\mathbf{Q}_z^\ell$ and $\mathbf{Q}_z^u$ are symmetric and positive definite, their convex combination $\mathbf{Q}_z^*$ is as well. To show that this construction satisfies Assumption 8 (iii), define

$$\mathbf{Q}_0^\infty(\bar{c}) \equiv \mathbb{E}\left[(1-S_g)\mathbf{f}(\bar{c}S_g)\mathbf{f}(\bar{c}S_g)'\right], \quad \mathbf{Q}_1^\infty(\bar{c}) \equiv \mathbb{E}\left[S_g\mathbf{f}(\bar{c}S_g)\mathbf{f}(\bar{c}S_g)'\right].$$

Recall that $0 \le S_g \le 1$ a discrete random variable with finite support, $\bar{c}$ is a real number between zero and one, and $\mathbf{f}$ is a K-vector of Lipschitz-continuous functions, all of which are bounded on $[0,1]$. It follows that both $\mathbf{Q}_0^\infty$ and $\mathbf{Q}_1^\infty$ are bounded and Lipschitz-continuous on $[0,1]$. Accordingly, by Lemma 1, Jensen's inequality, and the triangle inequality we can show that

$$\left\|\mathbf{Q}_z^\ell(\bar{c},n) - \mathbf{Q}_z^\infty\left(\bar{c}_\ell(\bar{c},n)\right)\right\| \le \frac{L}{\sqrt{n-1}}, \quad \left\|\mathbf{Q}_z^u(\bar{c},n) - \mathbf{Q}_z^\infty\left(\bar{c}_u(\bar{c},n)\right)\right\| \le \frac{L}{\sqrt{n-1}}$$

where L denotes an arbitrary, finite, positive constant. Similarly,

$$\left\|\mathbf{Q}_z^\infty(\bar{c}) - \mathbf{Q}_z^\infty\left(\bar{c}_\ell(\bar{c},n)\right)\right\| \le \frac{L}{n-1}, \quad \left\|\mathbf{Q}_z^\infty(\bar{c}) - \mathbf{Q}_z^\infty\left(\bar{c}_u(\bar{c},n)\right)\right\| \le \frac{L}{n-1}.$$

Combining these inequalities an applying the triangle inequality, it follows that

$$\left\|\mathbf{Q}_z^u(\bar{c},n) - \mathbf{Q}_z^\ell(\bar{c},n)\right\| \le \frac{L}{\sqrt{n-1}}, \quad \left\|\mathbf{Q}_z^u(\bar{c},n) - \mathbf{Q}_z^\infty(\bar{c})\right\| \le \frac{L}{\sqrt{n-1}}$$

and as a consequence

$$\left\|\mathbf{Q}_z^\ell(\bar{c},n) - \mathbf{Q}_z^\infty(\bar{c})\right\| \le \frac{L}{\sqrt{n-1}}$$

where, again, L is an arbitrary, finite, positive constant. Thus,

$$\begin{aligned}
\left\|\mathbf{Q}_z^*(\bar{c},n) - \mathbf{Q}_z^\infty(\bar{c})\right\| &\le \left\|\mathbf{Q}_z^*(\bar{c},n) - \mathbf{Q}_z^\ell(\bar{c},n)\right\| + \left\|\mathbf{Q}_z^\ell(\bar{c},n) - \mathbf{Q}_z^\infty(\bar{c})\right\| \\
&\le \left\|\mathbf{Q}_z^*(\bar{c},n) - \mathbf{Q}_z^\ell(\bar{c},n)\right\| + \frac{L}{\sqrt{n-1}} \\
&= \omega(\bar{c},n)\left\|\mathbf{Q}_z^u(\bar{c},n) - \mathbf{Q}_z^\ell(\bar{c},n)\right\| + \frac{L}{\sqrt{n-1}} \\
&\le \frac{L}{\sqrt{n-1}}
\end{aligned}$$

using the definitions of $\mathbf{Q}_z^*$ and $\omega(\bar{c}, n)$ from above. Combining all of the preceding inequalities,

$$\left\| \mathbf{Q}_z^*(\widehat{C}_{ig}, N_g) - \mathbf{Q}_z^*(\bar{C}_{ig}, N_g) \right\| \leq L \left\{ \frac{1}{\sqrt{\underline{n} - 1}} + \left| \widehat{C}_{ig} - \bar{C}_{ig} \right| \right\}$$

since $\underline{n} \leq N_g$ and $\mathbf{Q}_z^\infty$ is Lipschitz-continuous.

E Regression-based test of IOR assumption

We implement a simple regression-based test of Assumption 5 (individualistic offer response, IOR) for our empirical application. In settings with two-sided non-compliance, the following regression tests whether the individuals are more likely to take up the treatment if they are in a high saturation group relative to a low saturation group as follows:

$$D_{ig} = \gamma_0 + \gamma_1 Z_{ig} + \sum_{s \in \mathcal{S}^-} \beta_s \mathbb{1}(S_g = s) + \sum_{s \in \mathcal{S}^-} \beta_{J-1+s} \mathbb{1}(S_g = s) Z_{ig} + \epsilon_i$$
$$H_0 : \beta_1 = \beta_2 = ... = \beta_{J+J-2} = 0$$

where $\mathcal{S}^-$ is the full set of J saturation bins with one excluded. Therefore $\hat{\gamma}_0$ gives the average take-up for individuals with $Z_{ig} = 0$ in the excluded saturation bin, $\hat{\gamma}_0 + \hat{\gamma}_1$ gives the average take-up for individuals with $Z_{ig} = 1$ in the excluded bin, and the $\hat{\beta}$ terms give the difference in take-up in the other saturation bins for individuals with $Z_{ig} = 0$ and $Z_{ig} = 1$. If there is evidence against the null, this suggests that there are social interactions in take-up, and IOR may fail. In our application, Assumption 4 also holds – we have one-sided non-compliance where $Z_{ig} = 0$ implies $D_{ig} = 0$ – so the test simplifies to:

$$D_{ig} = \gamma_1 + \sum_{s \in \mathcal{S}^-} \beta_{J-1+s} \mathbb{1}(S_g = s) + \epsilon_i$$
$$H_0 : \beta_{J-1+s} = \beta_2 = ... = \beta_{J+J-2} = 0$$

where this regression is estimated on the sub-sample that is offered treatment, with $Z_{ig} = 1$. We estimate this regression using our sample from Crépon et al. (2013), clustering standard errors at the city (group) level. This allows us to calculate the estimated share of compliers as $\hat{\gamma}_1$ for the excluded bin and $\hat{\gamma}_1 + \hat{\beta}_s$ for each of the other saturation bins, which we plot in Figure E.1. We do not find any evidence against IOR.

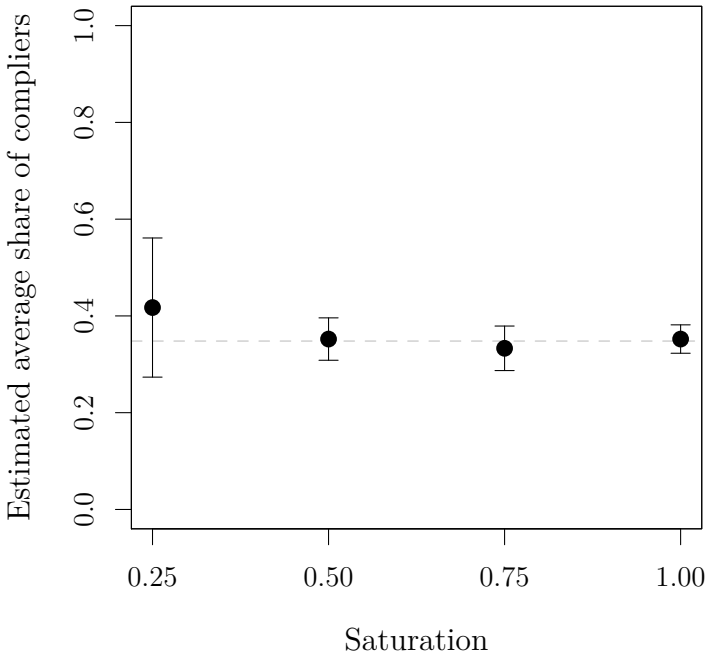

Figure E.1: Regression-based test of IOR. The estimated share of compliers is given by the dot and its 95% confidence interval is given by the bars for each of our four saturation bins. The horizontal dotted line gives the estimated share of compliers across the whole sample.