Program Evaluation in the Presence of Strategic Interactions*

Daron Acemoglu[1,2], Francis J. DiTraglia[3], Camilo García-Jimeno[2,3], and Rossa O'Keeffe-O'Donovan[4]

[1]MIT [2]NBER [3]UPenn [3]Oxford

February 18, 2018

Abstract

This paper considers causal inference in randomized controlled trials with interference, settings in which an individual's outcome may depend both on her own treatment and that of her neighbors. In such settings, applied researchers commonly employ randomized saturation designs to estimate both the direct effect of an individual's own treatment and the indirect effect of her neighbors' treatments. Many such examples are subject to imperfect compliance: subjects who are offered treatment may not take it up, and subjects who are not offered treatment may obtain it nonetheless. In this case researchers use the exogenous variation in treatment offers, both at the individual and group level, as instrumental variables to estimate the causal effects of interest. But when outcomes may depend on neighbors' treatments, treatment take-up may as well, a situation we call "strategic interactions." We first show that instrumental variables estimates of direct and spillover effects from a randomized saturation experiment can only be interpreted as local average treatment effects in the absence of strategic interactions. We then show that the assumption of no strategic interactions has testable implications under this experimental design, and go on to propose a simple regression-based test. If the test suggests that strategic interactions are present, use of the potential outcomes framework becomes infeasible as the number of principal strata that must be considered is extremely large, even for simple network structures. Accordingly, when our test finds evidence of strategic interactions, we propose using a simple structural model to recover own and network effects. Regardless of its outcome, our test for strategic interactions affects subsequent inference for the causal parameters of interest. To address this challenge, we derive a procedure for valid post-selection inference conditional on the selected model. We apply our methods to data from two well-known empirical papers. We find no evidence of strategic interactions in decisions to attend a job market training program in France. We do, however, find evidence of strategic interactions in decisions to receive deworming treatment in Kenya, and use our model to estimate the structural parameters, and the corresponding direct and spillover effects of treatment.

Keywords: strategic interactions, program evaluation, spillovers, externalities.

JEL Classification: C26, C31, C52

*This is a preliminary draft - please do not cite or circulate. For the latest version of the paper, please email rossaokod@gmail.com.

1 Introduction

The evaluation of programs and policies has become a major research area in economics. Research often focuses on the direct effects of treatment on program participants, but the spillover effects of a program may also be important for two key reasons. First, ignoring spillover effects may bias estimates of the direct effect, due to 'contamination' of the control group. Second, spillover effects are often of economic interest in their own right: we might significantly over or under-estimate the overall effects of a program if we ignore its spillover effects. Recent innovations in experimental designs have improved identification of spillover effects. However, we show that in settings where direct effects and spillover effects of a treatment interact, which often lead to strategic interactions across subjects, reduced form estimates of direct effects are biased.

Direct effects and spillover effects of a treatment interact if the size of the spillover effects on an individual depend on her own treatment status.[1] For example, in public health interventions to reduce communicable diseases, spillover effects may occur through reduced infection rates, but the benefits from reduced infection might depend on whether an individual is already infected, and therefore on their own treatment status. There are therefore three possible effects of a program or treatment: the direct effect of the treatment on those who receive it; the 'pure spillover' effect of the treatment, the size of which is independent of an individual's treatment status; and an interaction effect, through which the size of the spillover depends on an individual's treatment status.

Previous research typically estimates the direct and spillover effects of a treatment using a reduced form approach, but does not account for possible interactions between the two effects. Researchers estimate a regression with some measure of individual outcomes on the left hand side, and the treatment status of the individual and her neighbors' as explanatory variables. Recognizing that treatment status may be endogenous, recent papers have used instruments for individual treatment status and for her neighbor's average treatment status, sometimes using a two-step 'randomized saturation' experimental design. Under such designs, the proportion of people assigned to treatment in a given area is randomized, as well as the treatment status of individuals within such area.

We show that when there is an interaction between direct effects and spillover effects of a treatment, 'naive' estimates of the direct effect of treatment using a reduced form approach that excludes this interaction term are biased, even when valid instruments for an individual's treatment and the proportion of her neighbors that are treated are available. In addition, we show that inclusion of the interaction term in a regression of an individual's outcomes on her own treatment, her neighbors' treatment and their interaction can actually makes things worse: estimates of both the spillover and

[1] Equivalently, direct and spillover effects interact if the direct effects an individual receives depend on the treatment status of her neighbors. This different framing is conceptual, and the key point is that the size of the direct effects and spillover effects depend on each other.

direct effects are now biased. To understand why, notice that when treatment take-up is endogenous (as in most social interventions), and an interaction effect between the direct and spillover effects of a treatment is present, the take up decision of a subject is likely to be affected by the take up decisions of her neighbors. This is because the returns to own take up will vary with the extent of neighbors' take-up, leading to strategic interactions in the take-up decisions across subjects. In this setting, individual take-up decisions have *equilibrium* effects that go beyond partial equilibrium responses, and are not captured by the reduced form estimates of the partial effect of treatment on the outcome of interest. We illustrate this point formally using a simple model with linear best responses to show that the 'naive' reduced form approach does not estimate the 'own' effect of treatment, as it ignores the best responses and corresponding equilibrium effects.

The contribution of this paper is methodological: we develop a two-step procedure that tests whether direct effects and spillover effects interact, and produces bias-corrected estimates of the effects of treatment in the presence of strategic interactions. This requires a minimal structural model of treatment take up behavior and network structure. The first step of the procedure tests for the presence of strategic interactions, by estimating an instrumental variables regression of individual treatment on neighbors' treatment. If the estimated coefficient on neighbors' treatment is significant, we have evidence of strategic interactions in subjects' treatment take-up choices, and reduced form estimates of the effects of treatment will be biased. The second step of the procedure uses the model of strategic interactions to identify the structural parameters that determine the effects of treatment. When the treatment variable is continuous, this consists of estimating a linear best response equation and an equilibrium outcomes equation, using instruments for individual and neighbors' treatment. In cases where the treatment variable is binary, we estimate the best response equation using a Probit model, and use a Heckman-style selection correction term to consistently estimate the equilibrium outcomes equation. We estimate the system of equations using a method of moments estimator.

We analyze the performance of our two-step procedure in simulations, and then apply it to data from two empirical papers, Crépon et al. [2013] and Miguel and Kremer [2004]. Crépon et al. [2013] use a randomized saturation design to test the direct and spillover effects of a job placement assistance program on labor market outcomes in France. They find evidence that individuals who were randomly assigned to the program were more likely to be employed in the short run, but that these gains were partly at the expense of eligible workers who were not assigned to the program. It is possible that these negative spillover effects depend upon whether an individual participated in the job training program or not, inducing strategic interactions in the decision to participate in the program. However, we apply the first step of our procedure and find no evidence that an individuals' decision to receive treatment depends on the treatment of her neighbors, and thus find no evidence of bias in the study's reduced form estimates.

Miguel and Kremer [2004] estimates the direct and spillover effects of deworming treatment on school participation in Kenya. In this setting, we find evidence of strategic interactions in the decision to receive deworming treatment: we estimate that individuals are more likely to comply with treatment if a higher proportion of their neighbors receive treatment.[2] This result makes intuitive sense: spillover effects result from a reduced probability of being (re)infected with worms if other people have been treated, but this spillover will depend on an individual's own infection status, and therefore on her treatment status. Reduced form estimates of the effects of treatment that ignore these strategic interactions will be biased, so we use the second step of our procedure to estimate the structural parameters of the model, and the corresponding direct and spillover effects of deworming treatment. There are three main implications of our preliminary estimates for the effects of deworming on school participation: first, there is significant heterogeneity in the direct effects of deworming and these may be negative or insignificant for a majority of students; second, there are large positive spillover effects of deworming on school participation for both treated and untreated children; finally, these spillover effects are larger for children who received treatment.[3]

Our research builds on a relatively recent program evaluation literature that seeks to estimate the spillover effects of a treatment. We classify this literature into three 'generations' of studies, with later generations employing more robust methods of identification, with weaker corresponding assumptions. The first generation of studies randomly assign units to treatment or control, but do not have plausibly exogenous variation in the intensity of neighbors' treatment with which to estimate the size of spillovers.[4] Some studies estimate spillover effects on individuals who are not eligible for treatment (see Baird et al. [forthcoming] for a good overview of these 'partial population experiments'), while others simply look at the effects on people who do not select into treatment. However, the proportion of an individual's neighbors that receive treatment may depend on eligibility rules or self-selection into treatment, so these studies rely on strong assumptions of exogeneity.

The second generation of studies use geographic variables to define a radius around a unit of observation and use the random treatment assignment of other units within this radius as exogenous variation in intensity of neighbors' treatment. Miguel and Kremer [2004] uses this approach in a study of the effects of mass school-based deworming in Kenya. Each school is randomly assigned to either the treatment or control, and the authors use schools' GPS coordinates to calculate the

[2] In this case, we do not have a randomized saturation design, which gives exogenous variation in the intensity of treatment in a given area. However, the authors create an instrument for the proportion of people treated within a given radius, by calculating the proportion of people randomly assigned to receive treatment within this radius. We use a very similar approach, though focus on smaller cutoff distances in our analysis (1.2km. 1.8km and 2.4km, rather than 0-3km and 3-6km).

[3] Note that these estimates are preliminary. We have not yet calculated standard errors for the point estimates, as this is complicated by a number of factors discussed in section 4.2.

[4] 'Units' can refer to either individuals or clusters in a cluster-randomized design. In the latter case, if spillovers occur within a cluster, and not between clusters, then overall effects of the treatment (including spillover effects) can be estimated, though this overall effect cannot be decomposed into direct effects and spillover effects.

proportion of pupils within 3km and 3-6km that are assigned to treatment. They find evidence of positive cross-school externalities, with health and education outcomes improving for students in schools neighboring treatment schools. Bobba and Gignoux [2014] applies the same idea to Progresa and finds evidence that the cash transfer program had positive spillover effects on secondary school participation.

The third generation of studies use a two-step 'randomized saturation' experimental design.[5] In the first step, clusters are defined and the proportion of individuals that will be assigned to treatment in each cluster is randomized. In the second step, individuals within each cluster are randomly assigned to treatment or control according to the proportion drawn in the first step. Crépon et al. [2013] evaluates the impacts of a job placement assistance program in France using a randomized saturation experimental design. In the first step, clusters (labor markets) are randomly assigned to one of five proportions (0%, 25%, 50%, 75% and 100%), and in the second step eligible individuals are assigned to treatment or control in these proportions. The authors find some evidence of negative spillover effects of the program on non-participants. Angelucci et al. [2015] evaluates the impact of a micro-loan program in Mexico and employs a randomized saturation design, though the authors do not use the variation in treatment intensity to estimate spillover effects in their current paper.

Our two-step procedure depends on the availability of instruments such as those used in the second and third generation of studies. In particular, our first step, the regression-based test of strategic interactions, requires exogenous variation in individual treatment and the intensity of neighbors' treatment, as does estimation of the best response and outcomes equation in the second step. Our approach is dependent on the availability of such data and therefore builds directly on the progress made in this recent literature.

The remainder of this paper is structured as follows. Section 2 sets out our methodological framework: we show why naive reduced form estimates of the direct and spillover effects are biased in the presence of strategic interactions, and propose a two-step procedure to test for and correct this bias. Section 3 demonstrates our two-step bias correction procedure using simulated data, and section 4 applies it to data from two empirical papers, evaluating a job placement assistance program in France and a deworming program in Kenya. Section 5 concludes.

[5] Baird et al. [forthcoming] uses the term 'randomized saturation' and provides a more detailed discussion of these studies.

2 Methodological Framework

2.1 A simple model of treatment effects with strategic interactions

Suppose that we are interested in the effect of some continuous treatment variable, $x_i \in X$ where $X \subset \mathbb{R}^+$, on a continuous outcome variable, $y_i \in \mathbb{R}$, and that we want to estimate the spillover effects of the treatment, as well as the direct effects. There are n individuals in our sample, and we assume that we have some measure of network connections between them, which are summarized in a weighted, undirected, symmetric $n \times n$ matrix, $\mathbf{N}$, where the ij^{th} element, n_{ij}, represents the strength of connection between individuals i and j. n_{ij} is positive if i and j are neighbors, and zero otherwise.[6] We allow i's treatment and the treatment of i's neighbors to affect outcomes in the following way:

$$y_i = c_0 + f(x_i; \kappa) + g(\mathbf{N}_i\mathbf{x}; \gamma) + h(x_i, \mathbf{N}_i\mathbf{x}; \phi) + \epsilon_i \tag{1}$$

where c_0 is a constant and $\mathbb{E}[\epsilon_i] = 0$. The function $f(\cdot)$ is the direct effect of individual i's treatment, x_i. The second term, $g(\cdot)$, is the 'pure' spillover effect: the effect of i's neighbors' treatment on i's outcomes. This is a function of $\mathbf{N}_i\mathbf{x}$, the product of the i^{th} row of the network connections matrix, $\mathbf{N}_i$, and the full vector of treatments, $\mathbf{x}$. The third term, $h(\cdot)$, allows interactions between individual i's treatment and her neighbors' treatment to affect her outcomes. Intuitively, it is possible, and in some empirical settings it is likely, that the size of the spillover effects of i's neighbors' treatment on i's outcomes depends upon i's treatment status. For example, in many public health interventions, spillover effects come from increased group immunity or decreased risk of infection, and the size of these spillover effects for individual i may depend upon i's own treatment status.

In the second and third generation studies discussed in section 1, researchers have access to instruments for individual treatment, z_i, and for the treatment of i's neighbors, z_{ic}, where the subscript c indicates i's community or 'cluster'. In the absence of the interaction term, $h(\cdot)$, one can get unbiased estimates of the direct and spillover effects of treatment if $f(\cdot)$ and $g(\cdot)$ are known, by estimating $y_i = c_0 + f(x_i; \kappa) + g(\mathbf{N}_i\mathbf{x}; \gamma) + \epsilon_i$, using valid instruments for x_i and $\mathbf{N}_i\mathbf{x}$. Similarly, if the interaction term, $h(\cdot)$, affects outcomes, but individuals do not know about these effects (or ignore them), interactions are not strategic, and equation (1) can be estimated directly by using an instrumental variables regression.

However, in cases where the interaction term, $h(\cdot)$, affects outcomes and individuals understand this, interactions are strategic (i.e. an individual's treatment decision will depend on the treatment decisions of her neighbors) and estimates of a regression using equation (1) will be biased. The

[6] Each individual is not connected to itself, i.e. $n_{ij} = 0 \; \forall i = j$. We allow for the strength of connections to vary between different pairs of neighbors (i.e. $\mathbf{N}$ is a weighted matrix), and this nests cases where the measure of network connections is binary, $n_{ij} \in \{0, 1\}$.

intuition is that if we increase an individual's treatment, x_i, then this has a direct effect through $f(\cdot)$ and $h(\cdot)$, but it also induces i's neighbors to change their treatment status, and therefore has an additional equilibrium effect through $g(\cdot)$ and $h(\cdot)$. To show this formally, we assume that utility of individual i is linear in outcomes and the cost of obtaining treatment, $C(\cdot)$:

$$U_i = y_i - C(x_i; \theta) \tag{2}$$

Individual i chooses her treatment, x_i, to maximize utility, U_i, given the actions (treatment take-up) of her neighbors. If $f(x_i; \kappa) + h(x_i, \mathbf{N}_i\mathbf{x}; \phi) - C(x_i; \theta)$ is continuous, real valued and strictly quasi-concave in x_i, and X is a non-empty, convex and compact set, then i has a unique best response to her neighbors' treatment, $x_i^*(\mathbf{N}_i\mathbf{x})$, with interior solutions characterized by:

$$f_x(x_i^*; \kappa) + h_x(x_i^*, \mathbf{N}_i\mathbf{x}; \phi) - C_x(x_i^*; \theta) = 0 \tag{3}$$

Note that if there is no interaction term $h(\cdot)$, then x_i^* is independent of $\mathbf{N}_i\mathbf{x}$, there are no strategic interactions between i and her neighbors, and estimates of the direct and spillover effects are unbiased, as noted above.[7] We therefore propose a test of whether strategic interactions are present, by estimating a best response function of the form:

$$x_i = b(\mathbf{N}_i\mathbf{x}, z_i; \beta) + e_i \tag{4}$$

where the optimal treatment of individual i depends on the treatment of her neighbors, and possibly on exogenous variables (instruments), z_i. We can estimate this equation non-parametrically, or using a flexible functional form for $b(\cdot)$. However, if $b(\cdot)$ is monotonic in $\mathbf{N}_i\mathbf{x}$, and $\mathbf{N}_i\mathbf{x}$ and z_i are separable, then we can test the null hypothesis by estimating a linear best response of the form:

$$x_i = \beta_0 + \beta_1\mathbf{N}_i\mathbf{x} + \beta_2 z_i + e_i \tag{5}$$

In many cases, we may reasonably expect $\mathbf{N}_i\mathbf{x}$ to be endogenous, as it is likely that neighbors share unobserved characteristics or shocks that may affect their treatment choices, and so we should estimate this best response equation by using z_{ic} as an instrument for $\mathbf{N}_i\mathbf{x}$. We can now test whether there are strategic interactions by testing the null hypothesis, $H_0 : \beta_1 = 0$ against the two-sided alternative, using standard two stage least squares estimation.

If there is evidence of strategic interactions (i.e. we reject null hypothesis, $H_0 : \beta_1 = 0$), then

[7] In cases where there is a boundary solution for x_i^* for all individuals, marginal changes in $\mathbf{N}_i\mathbf{x}$ will not change x_i^*. Therefore x_i^* is locally independent of $\mathbf{N}_i\mathbf{x}$ and omission of $h(\cdot)$ will not bias estimates of κ and γ. This will be the case in empirical settings where all individuals take the maximum amount of treatment, which we commonly see in programs where the benefits of treatment far outweigh the costs, for example in unconditional cash transfer programs. It is also the case in intent to treat estimates, where we are interested in the effect of *assignment* to treatment, and so there is necessarily perfect 'compliance'.

we cannot get unbiased estimates of the direct effects, spillover effects or their interaction without understanding, and modeling, the equilibrium effects of a change in x_i. This requires a model with more structure, to allow us to estimate the best response equation and how this affects the outcomes observed in equilibrium. We give an outline of such a model, and how it can be estimated, in the next section.

2.2 A two-step structural bias-correction procedure

We use a simple version of the model presented in section 2.1 to propose a two-step procedure to test for strategic interactions and, in their presence, identify the structural parameters that determine direct effects and spillover effects of a treatment.[8] The solution to the model, and how it is estimated, depends on whether the treatment variable is continuous or binary, and on the choice set of agents assigned to treatment and control. In particular, we consider three cases: settings with a continuous treatment variable; settings with a binary treatment variable and two-sided non-compliance (i.e. the control group has access to treatment, and the treatment group can opt-out of treatment); and settings with a binary treatment variable and one-sided non-compliance (individuals can opt out of treatment, but only the treatment group can access treatment).

We assume that the researcher has access to instruments for both an individual's treatment, z_i, and her neighbors' treatment, z_{ic} (henceforth z_c for simplicity). It is most intuitive to think of these treatments as generated by a two-step experimental design, similar to that employed by Crépon et al. [2013]. The experimental design in Crépon et al. [2013] is intended to measure the externalities of a job placement assistance program in France on individuals not assigned to treatment, as well as the direct benefits to those assigned to the program.[9] The paper uses a two step experimental design: in the first step, municipalities are randomly assigned to an intensity of treatment (0%, 25%, 50%, 75% or 100% of job seekers to be assigned to treatment); in the second step, individuals in these municipalities are randomly assigned to treatment in the proportions determined by the first step.

Given access to valid instruments, z_i and z_{ic}, the best response equation can be easily estimated to test for strategic interactions and, in their presence, we use our model to estimate the structural parameters for each of the three cases outlined above.[10] We give an overview of the model, and how to estimate its structural parameters, in each of the three cases in the following sections.

[8] The model is similar to that in Acemoglu et al. [2015], which estimates the direct and spillover effects of local state capacity in Colombia, using a network game in which local municipalities have strategic interactions in their choice of investments in state capacity. The outcome in municipality i is a function of its own investment, the investment of its neighbors, and an interaction between its own investment and its neighbors' investments (given by equation 1 in the paper).

[9] The authors find a positive effect of assignment to the job training program on employment outcomes, but that these gains are transitory and appear to come partly at the expense of workers who did not benefit from the program.

[10] The three cases are: continuous treatment, binary treatment with two-sided non-compliance and binary treatment with one-sided non-compliance

2.2.1 Continuous treatment variable

In the case of a continuous treatment variable, $x_i \in \mathbb{R}^+$, we can identify the structural parameters by estimating the best response equation and equilibrium outcomes equation in a simplified version of our model of strategic interactions with linear best responses. Outcomes, y_i, are given by the linear specification:

$$y_i = c_0 + \kappa_i x_i + \gamma \mathbf{N}_i\mathbf{x} + \phi x_i \mathbf{N}_i\mathbf{x} + \epsilon_i \tag{6}$$

As in section 2.1, the first term is the direct effect of individual i's treatment on her outcomes, the second term is the 'pure spillover' effect of i's neighbors' treatment on her outcomes, and the third term is the interaction between individual i's treatment and her neighbors' treatment. Note that we model the direct effect as heterogeneous, in particular using the form:

$$\kappa_i = \kappa + \varphi z_i + \xi_i \tag{7}$$

where ξ_i is a mean-zero random shock, and $cov(z_i, \epsilon_i) = cov(z_i, \xi_i) = 0$. There are two reasons for this functional form. First, we include an idiosyncratic shock to the direct effect of treatment, ξ_i, because in its absence the best response function for each individual would be deterministic (conditional on z_i) and thus unable to rationalize any variance in observed responses. Second, we must specify how the instrument, z_i, in our case the random assignment of an individual to treatment or control, affects the treatment received by an individual, x_i. There are two natural ways to include z_i in our model: as an augmentation of the direct effects of treatment (as part of κ_i), or as a decrease in the cost of treatment (as part of $C(\cdot)$ in equation (2)). Note that including z_i linearly in κ_i is conceptually equivalent to including it as a linear cost of treatment in U_i; we choose the former option because it results in a cleaner and simpler outcomes equation in equilibrium, which is more straightforward to estimate.[11]

Individual i chooses her level of treatment, x_i, to maximize quadratic utility:

$$U_i = y_i - \frac{\theta}{2} x_i^2 \tag{8}$$

Taking first order conditions gives the equilibrium best response equation:[12]

$$x_i = \frac{\phi}{\theta} \mathbf{N}_i\mathbf{x} + \frac{\kappa_i}{\theta} \tag{9}$$

[11] We are working on generalizations and departures from our simple model, and how these affect estimation of the structural parameters, and will add this to future versions of this paper.

[12] Existence of pure strategy equilibria follows immediately from concavity and Kakutani's fixed point theorem, and there is a unique interior equilibrium given by the solution to a set of linear best response equations. If $|\lambda_{min}(\mathbf{N}(\delta))| < \left(\frac{\phi}{\theta}\right)^{-1}$ then this interior equilibrium is unique, as shown in Bramoullé et al. [2014].

We can substitute the best response equation into the outcomes equation to give equilibrium outcomes:

$$y_i = c_0 + \theta x_i^2 + \gamma \mathbf{N}_i \mathbf{x} + \epsilon_i \tag{10}$$

because in equilibrium $\kappa_i x_i + \phi x_i \mathbf{N}_i \mathbf{x} = \theta x_i^2$. Note that the equilibrium outcomes equation is a function of x_i^2 and $\mathbf{N}_i \mathbf{x}$, and that neither the direct effect of treatment, or the effect of the interaction term can be identified by using a simple regression of outcomes on treatment. However, we can identify each of the parameters of the model by estimating both the best response equation, (9), and the equilibrium outcome equation, (10), provided that we have valid instruments for x_i and $\mathbf{N}_i \mathbf{x}$. The equilibrium outcomes equation identifies θ and γ. Then, given $\hat{\theta}$, estimating the best response equation, $x_i = \frac{\phi}{\theta} \mathbf{N}_i \mathbf{x} + \frac{1}{\theta}(\kappa + \varphi z_i + \xi_i)$, identifies ϕ, κ and φ. We can estimate these two equations separately, as a two-step estimator, but it will be more efficient to estimate them jointly using a method of moments estimator with the sample analogs of the following moments:

$$\mathbb{E}\left[\begin{pmatrix} z_i \\ z_c \end{pmatrix} \left(x_i - \frac{\phi}{\theta} \mathbf{N}_i \mathbf{x} - \frac{1}{\theta}(\kappa + \varphi z_i) \right) \right] = 0 \tag{11}$$

$$\mathbb{E}\left[\begin{pmatrix} z_i \\ z_c \end{pmatrix} \left(y_i - c_0 - \theta x_i^2 - \gamma \mathbf{N}_i \mathbf{x} \right) \right] = 0 \tag{12}$$

To test the performance of our proposed methodology, we simulate data from the linear model with a continuous treatment variable, and use this method to estimate the structural parameters in section 3.

2.2.2 Binary treatment variable, with two-sided non-compliance

In cases with a binary treatment variable, $x_i \in \{0, 1\}$, we do not have a first order condition to derive the best response equation, and as a result cannot obtain a single equilibrium equation for outcomes as we did in section 2.2.1. There are two separate cases to consider: binary treatment with two-sided and one-sided noncompliance.[13] There are only minor differences between the cases, and we consider each in turn.

Consider our benchmark outcomes equation:

$$y_i = c_0 + \kappa_i x_i + \gamma \mathbf{N}_i \mathbf{x} + \phi x_i \mathbf{N}_i \mathbf{x} + \epsilon_i \tag{13}$$

Here $x_i \in \{0, 1\}$, and as before we have $\kappa_i = \kappa + \varphi z_i + \xi_i$, $\mathbb{E}[\xi_i] = 0$, $\mathbb{E}[\epsilon_i] = 0$. Note that $x_i^2 = x_i$, so

[13] In settings with two-sided non-compliance, the control group has access to treatment, and the treatment group can opt-out of treatment. However, in settings with one-sided non-compliance, individuals can still opt out of treatment but only the treatment group can access treatment (i.e. there is nobody with $z_i = 0$ and $x_i = 1$).

we rewrite the utility function in equation (8) so that it is linear:

$$U_i = y_i - \theta x_i \tag{14}$$

The choice of x_i is given by the best response function:

$$x_i = \begin{cases} 1 & if\ \kappa + \varphi z_i - \theta + \phi \mathbf{N}_i \mathbf{x} \geq -\xi_i \\ 0 & if\ \kappa + \varphi z_i - \theta + \phi \mathbf{N}_i \mathbf{x} < -\xi_i \end{cases} \tag{15}$$

We assume that the ξ_i terms are normally distributed and estimate equation (15) with a Probit IV using z_c as an instrument for $\mathbf{N}_i\mathbf{x}$. Testing whether the coefficient on $\mathbf{N}_i\mathbf{x}$ is statistically significant from zero provides a test of $H_0 : \phi = 0$, and therefore provides the first step of our two step procedure. Estimating this best response equation allows us to recover estimates for:

$$\tilde{\alpha} \equiv \frac{\kappa - \theta}{\sigma_\xi} \tag{16}$$

$$\tilde{\varphi} \equiv \frac{\varphi}{\sigma_\xi} \tag{17}$$

$$\tilde{\phi} \equiv \frac{\phi}{\sigma_\xi} \tag{18}$$

Next, consider the equilibrium outcome equations for the treated and untreated groups:

$$y_{i|x_i=0} = c_0 + \gamma \mathbf{N}_i \mathbf{x} + \epsilon_{i|x_i=0} \tag{19}$$

$$y_{i|x_i=1} = c_0 + \kappa + \varphi z_i + \xi_{i|x_i=1} + (\phi + \gamma)\mathbf{N}_i \mathbf{x} + \epsilon_{i|x_i=1} \tag{20}$$

There will be selection into treatment based on the unobserved ξ_i. To see this, note that the best response of individual i follows a cutoff rule, whereby i chooses $x_i = 1$ if and only if ξ_i is above some threshold value. As shown by equation (15), this threshold depends on the value of z_i and $\mathbf{N}_i\mathbf{x}$ (and by extension its instrument, z_c). Treated individuals with a higher value of z_i and z_c are likely to have a lower value of ξ_i on average, giving $cov(z_i, \xi_{i|x_i=1}) \neq 0$, $cov(z_i, \xi_{i|x_i=0}) \neq 0$, $cov(z_c, \xi_{i|x_i=1}) \neq 0$ and $cov(z_c, \xi_{i|x_i=0}) \neq 0$. This will bias our estimates of equation (20), and also our estimates of equation (19) if $cov(\xi_i, \epsilon_i) \neq 0$. Therefore, to estimate these equations consistently, we use Heckman's selection correction. The selection correction for equation (19) is relatively standard, but estimating equation (20) is complicated slightly by the composite error term, $\xi_{i|x_i=1} + \epsilon_{i|x_i=1}$. We derive each in turn.

First, define

$$\begin{pmatrix} \epsilon_i \\ \xi_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\epsilon^2 & \rho\sigma_\epsilon\sigma_\xi \\ \rho\sigma_\epsilon\sigma_\xi & \sigma_\xi^2 \end{pmatrix} \right] \tag{21}$$

In estimating equation (19), we must estimate $E[\epsilon_i|x_i = 0]$ which is given by:

$$E[\epsilon_i|\xi_i < -(\kappa + \varphi z_i - \theta + \phi \mathbf{N}_i\mathbf{x})] = \frac{\rho\sigma_\epsilon}{\sigma_\xi} \frac{-\phi\left(\frac{-(\kappa+\varphi-\theta+\phi\mathbf{N}_i\mathbf{x})}{\sigma_\xi}\right)}{\Phi\left(\frac{-(\kappa+\varphi-\theta+\phi\mathbf{N}_i\mathbf{x})}{\sigma_\xi}\right)} = \frac{\rho\sigma_\epsilon}{\sigma_\xi}\lambda_i^0 \tag{22}$$

where $\phi()$ and $\Phi()$ represent the density and cumulative distribution functions for the standard normal distribution, and λ_i^0 is an inverse mills ratio. We show the derivation of this equation in section A.1 in the appendix. We can use the estimated coefficients from our best response equation, given in equations (16), (17) and (18), to give us an estimate of λ_i^0:

$$\hat{\lambda}_i^0 = \frac{-\phi\left(-(\tilde{\alpha} + \tilde{\varphi} z_i + \tilde{\phi}\mathbf{N}_i\mathbf{x})\right)}{\Phi\left(-(\tilde{\alpha} + \tilde{\varphi} z_i + \tilde{\phi}\mathbf{N}_i\mathbf{x})\right)} \tag{23}$$

We can now estimate equation (19), correcting for the selection bias:

$$y_{i|x_i=0} = c_0 + \gamma\mathbf{N}_i\mathbf{x} + \frac{\rho\sigma_\epsilon}{\sigma_\xi}\hat{\lambda}_i^0 + \eta_i \tag{24}$$

We estimate this on the untreated subsample, using z_c as an instrument for $\mathbf{N}_i\mathbf{x}$ which gives us consistent estimates of the pure spillover effect, $\hat{\gamma}$, and the constant term, $\hat{c}_0$. Note that the coefficient on $\hat{\lambda}_i^0$ gives us an estimate of $\frac{\rho\sigma_\epsilon}{\sigma_\xi}$.

Estimating the outcomes equation for the treated group follows a very similar methodology, although the Heckman selection correction is non-standard because there is a composite error term, with $\xi_{i|x_i=1}$ in the outcomes equation. We calculate the expected value of the composite error term, given that an individual has selected into treatment (derivation shown in section A.1 in the appendix):

$$E[\xi_i + \epsilon_i|\xi_i \geq -(\kappa + \varphi z_i - \theta + \phi\mathbf{N}_i\mathbf{x})] = \left(1 + \frac{\rho\sigma_\epsilon}{\sigma_\xi}\right) \frac{\phi\left(\frac{-(\kappa+\varphi-\theta+\phi\mathbf{N}_i\mathbf{x})}{\sigma_\xi}\right)}{1 - \Phi\left(\frac{-(\kappa+\varphi-\theta+\phi\mathbf{N}_i\mathbf{x})}{\sigma_\xi}\right)} = \left(1 + \frac{\rho\sigma_\epsilon}{\sigma_\xi}\right)\lambda_i^1 \tag{25}$$

As before, we can use the estimated coefficients from the best response to give us an estimate of λ_i^1:

$$\hat{\lambda}_i^1 = \frac{\phi\left(-(\tilde{\alpha} + \tilde{\varphi} z_i + \tilde{\phi}\mathbf{N}_i\mathbf{x})\right)}{1 - \Phi\left(-(\tilde{\alpha} + \tilde{\varphi} z_i + \tilde{\phi}\mathbf{N}_i\mathbf{x})\right)} \tag{26}$$

Finally, we can use z_c as an instrument for $\mathbf{N}_i\mathbf{x}$ to estimate the outcomes equation for the treated group:

$$y_{i|x_i=1} = c_0 + \kappa + \varphi z_i + (\phi + \gamma)\mathbf{N}_i\mathbf{x} + \left(1 + \frac{\rho\sigma_\epsilon}{\sigma_\xi}\right)\hat{\lambda}_i^1 + \eta_i \tag{27}$$

Note that z_c is now a valid instrument for $\mathbf{N}_i\mathbf{x}$. We previously worried that $cov(z_c, \epsilon_{i|x_i=1}) \neq 0$ because of selection into treatment, but now that we are controlling for selection, with $\hat{\lambda}_i^1$, we have $cov(z_c, \eta_i) = 0$. Estimation of equation (27) gives us estimates of $\hat{\varphi}$ as well as the reduced form parameters:

$$\hat{a} = c_0 + \kappa \tag{28}$$

$$\hat{\beta} = \phi + \gamma \tag{29}$$

Combining this with our estimates of $\hat{\gamma}$ and $\hat{c}_0$ from estimation of the outcomes equation for the untreated subgroup, equation (24), we can identify $\hat{\kappa} = \hat{a} - \hat{c}_0$ and $\hat{\phi} = \hat{\beta} - \hat{\gamma}$. These estimates, combined with those from the best response equation, given by equations (16), (17) and (18), gives us two sources of identification for σ_ξ, and allows us to identify $\hat{\theta} = \hat{\kappa} - \tilde{\alpha}\hat{\sigma_\xi}$. The variance of the disturbance terms is given by $var(\eta_i) = \sigma_\epsilon^2(1-\rho^2)$, which we can estimate with $\hat{\sigma_\eta}^2 = var(\hat{\eta_i})$. This gives us two equations in two unknowns for ρ and σ_ϵ, and so they are separately identified.[14] We therefore have point identification for all of the structural parameters in our model.

Note that the estimate of the outcomes equation for the treated group, equation (27), may suffer from collinearity. This is because the key input variables to $\hat{\lambda}_i^1$ are the same as the explanatory variables in the outcomes equation (z_i and $\mathbf{N}_i\mathbf{x}$) and because the inverse mills ratio is close to a linear function over a significant range of values. However, the outcomes equation for the untreated group, equation (24), does not generally suffer from collinearity, because $\hat{\lambda}_i^0$ is a function of z_i which does not appear as an explanatory variable in the outcomes equation.[15] Therefore $\frac{\rho\sigma_\epsilon}{\sigma_\xi}$ is identified in equation (24), and can be used to estimate the coefficient on $\hat{\lambda}_i^1$ in equation (27), $(1 + \frac{\rho\sigma_\epsilon}{\sigma_\xi})$. Holding this coefficient fixed, collinearity is no longer a problem for estimation of equation (27).

The above explanation assumes that the outcomes equations, (24) and (27), are estimated sequentially (for example using 2SLS) for ease of exposition. However, they can also be estimated jointly with a cross equation restriction on the coefficients on $\hat{\lambda}_i^0$ and $\hat{\lambda}_i^1$, using the method of moments estimator, and joint estimation will be more efficient.[16] First, define the exogenous variables in each

[14] Solving this system of two equations in two unknowns gives: $\hat{\sigma_\epsilon} = \left(var(\hat{\eta_i}) + \hat{\sigma_\xi}^2(\widehat{\frac{\rho\sigma_\epsilon}{\sigma_\xi}})^2\right)^{\frac{1}{2}}$ and $\hat{\rho} = \frac{\hat{\sigma_\xi}}{\hat{\sigma_\epsilon}}\widehat{\frac{\rho\sigma_\epsilon}{\sigma_\xi}}$.

[15] Estimation of the outcomes equation for the untreated subgroup, equation (24), might suffer from collinearity if there is little or no variation in z_i.

[16] Note that we cannot estimate the best response equation, (15), jointly with the outcomes equations by adding it as a moment inequality. This is because we need to make a functional form assumption on the ξ_i terms in order to model and correct for the selection bias. In particular, we assume that the ξ_i terms are normally distributed to estimate the Heckman correction terms that we use in the outcomes equations, (24) and (27). We need to estimate these correction terms, and therefore the best response equation, separately to the method of moments estimator, and fix these correction terms in the method of moments estimator to make sure that they are consistent with how we have modeled the selection into treatment.

outcomes equation:

$$\mathbf{z_i^0} = \begin{pmatrix} z_c \\ \hat{\lambda}_i^0 \end{pmatrix} \qquad \text{and} \qquad \mathbf{z_i^1} = \begin{pmatrix} z_i \\ z_c \\ \hat{\lambda}_i^1 \end{pmatrix}$$

Note that $\mathbb{E}[\mathbf{z_i^0}\eta_i] = 0$ and $\mathbb{E}[\mathbf{z_i^1}\eta_i] = 0$, allowing us to specify moment conditions for the outcomes equations of the treated and untreated subgroups:

$$\mathbb{E}\left[\mathbf{z_i^0}\left(y_i - c_0 - \gamma\mathbf{N}_i\mathbf{x} - \frac{\rho\sigma_\epsilon}{\sigma_\xi}\hat{\lambda}_i^0\right) \;\middle|\; x_i = 0\right] = 0$$

$$\mathbb{E}\left[\mathbf{z_i^1}\left(y_i - (c_0 + \kappa) - \varphi z_i - (\phi + \gamma)\mathbf{N}_i\mathbf{x} - \left(1 + \frac{\rho\sigma_\epsilon}{\sigma_\xi}\right)\hat{\lambda}_i^1\right) \;\middle|\; x_i = 1\right] = 0$$

We can use the sample analogs of these moments in our estimation. We have seven moment conditions and six unknown parameters $(c_0, \gamma, \frac{\rho\sigma_\epsilon}{\sigma_\xi}, \kappa, \varphi, \phi)$, so estimate the parameters by GMM.[17] Calculation of standard errors must account for the fact that $\hat{\lambda}_i^0$ and $\hat{\lambda}_i^1$ are 'generated regressors'.[18]

2.2.3 Binary treatment variable, with one-sided non-compliance

When the treatment variable is binary and there is only one-sided non-compliance, nobody assigned to the control group can access the treatment (i.e. there are no always takers or defiers, with $z_i = 0$ and $x_i = 1$). As in the case of binary treatment with two-sided non-compliance, there is no first order condition for the best response equation:

$$x_i = \begin{cases} 0 & if\ z_i = 0 \\ 1 & if\ z_i = 1\ and\ \kappa + \varphi - \theta + \phi\mathbf{N}_i\mathbf{x} \geq -\xi_i \\ 0 & if\ z_i = 1\ and\ \kappa + \varphi - \theta + \phi\mathbf{N}_i\mathbf{x} < -\xi_i \end{cases} \tag{30}$$

We now have three groups to consider: the control group ($z_i = 0$), compliers ($z_i = 1, x_i = 1$) and never takers ($z_i = 1, x_i = 0$). Note that with one-sided non-compliance, we can never separately identify κ and φ. To see why, consider our outcomes equation $y_i = c_0 + (\kappa + \varphi z_i + \xi_i)x_i + \gamma\mathbf{N}_i\mathbf{x} + \phi x_i\mathbf{N}_i\mathbf{x} + \epsilon_i$. Separate identification of κ and φ requires variation in z_i among individuals who have $x_i = 1$, but with one-sided non-compliance ($x_i = 1$) implies that ($z_i = 1$), so this is clearly not possible.

[17] The seven moment conditions include the five shown above plus:

$$\mathbb{E}\left[y_i - c_0 - \gamma\mathbf{N}_i\mathbf{x} - \frac{\rho\sigma_\epsilon}{\sigma_\xi}\hat{\lambda}_i^0 \;\middle|\; x_i = 0\right] = 0$$

$$\mathbb{E}\left[y_i - (c_0 + \kappa) - \varphi z_i - (\phi + \gamma)\mathbf{N}_i\mathbf{x} - \left(1 + \frac{\rho\sigma_\epsilon}{\sigma_\xi}\right)\hat{\lambda}_i^1 \;\middle|\; x_i = 1\right] = 0$$

[18] Depending on the empirical context in question, researchers should consider also allowing standard errors to be spatially correlated.

As before, we can use z_c as an instrument for $\mathbf{N}_i\mathbf{x}$ to estimate the best response equation for individuals assigned to treatment (compliers and never takers) using a Probit IV, to give:

$$\tilde{\alpha} = \frac{\kappa + \varphi - \theta}{\sigma_\xi} \tag{31}$$

$$\tilde{\phi} = \frac{\phi}{\sigma_\xi} \tag{32}$$

There are now outcomes equations for each of the three subgroups:

$$y_{i|z_i=0} = c_0 + \gamma\mathbf{N}_i\mathbf{x} + \epsilon_i \tag{33}$$

$$y_{i|z_i=1,x_i=0} = c_0 + \gamma\mathbf{N}_i\mathbf{x} + \epsilon_{i|z_i=1,x_i=0} \tag{34}$$

$$y_{i|z_i=1,x_i=1} = c_0 + (\kappa + \varphi + \xi_{i|z_i=1,x_i=1}) + (\phi + \gamma)\mathbf{N}_i\mathbf{x} + \epsilon_{i|z_i=1,x_i=1} \tag{35}$$

We estimate the outcomes equation for the compliers, equation (35) and never takers, equation (34) using selection bias correction terms similar to those in section 2.2.2. We take our estimates of $\tilde{\alpha}$ and $\tilde{\phi}$ from the best response equation and calculate:

$$\hat{\lambda}_i^0 = \frac{-\phi\left(-(\tilde{\alpha} + \tilde{\phi}\mathbf{N}_i\mathbf{x})\right)}{\Phi\left(-(\tilde{\alpha} + \tilde{\phi}\mathbf{N}_i\mathbf{x})\right)} \tag{36}$$

$$\hat{\lambda}_i^1 = \frac{\phi\left(-(\tilde{\alpha} + \tilde{\phi}\mathbf{N}_i\mathbf{x})\right)}{1 - \Phi\left(-(\tilde{\alpha} + \tilde{\phi}\mathbf{N}_i\mathbf{x})\right)} \tag{37}$$

As before, we can now estimate outcomes equations (34) and (35), correcting for the selection bias:

$$y_i|_{z_i=1,x_i=0} = c_0 + \gamma\mathbf{N}_i\mathbf{x} + \frac{\rho\sigma_\epsilon}{\sigma_\xi}\hat{\lambda}_i^0 + \eta_i \tag{38}$$

$$y_i|_{z_i=1,x_i=1} = (c_0 + \kappa + \varphi) + (\phi + \gamma)\mathbf{N}_i\mathbf{x} + \left(1 + \frac{\rho\sigma_\epsilon}{\sigma_\xi}\right)\hat{\lambda}_i^1 + \eta_i \tag{39}$$

Estimation of the outcomes equation for the control group, equation (33), does not require a selection correction because individuals in the control group did not face a compliance decision, and this identifies c_0 and γ.

In this case, collinearity is a problem for estimation of both equations (38) and (39). This is because the input variable for $\hat{\lambda}_i^0$ and $\hat{\lambda}_i^1$ is the same as the single explanatory variable in the outcomes equations. However, as in section 2.2.2, we can use 'cross-equation restrictions' coming from the structure of the model to ensure that the parameter estimates are consistent. For example, if we estimate the outcomes equations sequentially, we can proceed as follows:

1. Estimate the outcomes equation for the control group, equation (33) to identify c_0 and γ.

2. Fixing our estimates, $\hat{c}_0, \hat{\gamma}$, estimate the outcomes equation for the never-takers using the selection bias correction, equation (38). This identifies $\frac{\rho\sigma_\epsilon}{\sigma_\xi}$.

3. Fixing our estimates, $\hat{c}_0, \hat{\gamma}, \widehat{\frac{\rho\sigma_\epsilon}{\sigma_\xi}}$, estimate the outcomes equation for compliers using the selection bias correction, equation (39). This identifies ϕ and $(\kappa + \varphi)$.

4. As before, we can then use the estimates from the best response to identify σ_ξ and θ. We can use the variance of the residuals, $\hat{\eta}_i$, and $\widehat{\frac{\rho\sigma_\epsilon}{\sigma_\xi}}$ to separately identify ρ and σ_ϵ, because $var(\eta_i) = \sigma_\epsilon^2(1-\rho^2)$.[19] Therefore all structural parameters are separately identified apart from κ and φ.

As in section 2.2.2, although I have described this identification argument sequentially as if each of these outcomes equations is estimated separately, it is more efficient to estimate these equations jointly, using the method of moments estimator. As before, we define the exogenous variables in the outcomes equations (38) and (39):

$$\mathbf{z_i^0} = \begin{pmatrix} z_c \\ \hat{\lambda}_i^0 \end{pmatrix} \qquad \text{and} \qquad \mathbf{z_i^1} = \begin{pmatrix} z_c \\ \hat{\lambda}_i^1 \end{pmatrix}$$

We use the sample analogs of the following moment conditions in estimation:

$$\mathbb{E}\left[z_c \left(y_i - c_0 - \gamma \mathbf{N}_i \mathbf{x} \right) \,\middle|\, z_i = 0 \right] = 0$$

$$\mathbb{E}\left[\mathbf{z_i^0} \left(y_i - c_0 - \gamma \mathbf{N}_i \mathbf{x} - \frac{\rho\sigma_\epsilon}{\sigma_\xi} \hat{\lambda}_i^0 \right) \,\middle|\, z_i = 1 \cap x_i = 0 \right] = 0$$

$$\mathbb{E}\left[\mathbf{z_i^1} \left(y_i - (c_0 + \kappa + \varphi) - (\phi + \gamma)\mathbf{N}_i \mathbf{x} - \left(1 + \frac{\rho\sigma_\epsilon}{\sigma_\xi} \right) \hat{\lambda}_i^1 \right) \,\middle|\, z_i = 1 \cap x_i = 1 \right] = 0$$

We have eight moment conditions and five unknown parameters $(c_0, \gamma, \frac{\rho\sigma_\epsilon}{\sigma_\xi}, \kappa + \varphi, \phi)$, and estimate the parameters by GMM.[20]

[19] As before, solving this system of two equations in two unknowns gives: $\hat{\sigma}_\epsilon = \left(var(\hat{\eta}_i) + \hat{\sigma_\xi}^2 (\widehat{\frac{\rho\sigma_\epsilon}{\sigma_\xi}})^2 \right)^{\frac{1}{2}}$ and $\hat{\rho} = \frac{\hat{\sigma_\xi}}{\hat{\sigma_\epsilon}} \widehat{\frac{\rho\sigma_\epsilon}{\sigma_\xi}}$.

[20] We have the five moment conditions above, plus:

$$\mathbb{E}\left[y_i - c_0 - \gamma \mathbf{N}_i \mathbf{x} \,\middle|\, z_i = 0 \right] = 0$$

$$\mathbb{E}\left[y_i - c_0 - \gamma \mathbf{N}_i \mathbf{x} - \frac{\rho\sigma_\epsilon}{\sigma_\xi} \hat{\lambda}_i^0 \,\middle|\, z_i = 1 \cap x_i = 0 \right] = 0$$

$$\mathbb{E}\left[y_i - (c_0 + \kappa + \varphi) - (\phi + \gamma)\mathbf{N}_i \mathbf{x} - \left(1 + \frac{\rho\sigma_\epsilon}{\sigma_\xi} \right) \hat{\lambda}_i^1 \,\middle|\, z_i = 1 \cap x_i = 1 \right] = 0$$

3 Simulations

To analyze the performance of our proposed two-step procedure, we simulate data using the linear model with a continuous treatment variable, as presented in section 2.2.1. We estimate the average direct effect of the treatment and the pure spillover effect using a 'naive' two stage least squares regression, both including and excluding the interaction term, $x_i\mathbf{N}_i\mathbf{x}$, and show that these estimates are biased.[21] We then implement our two step procedure, by estimating the best response equation and equilibrium outcomes equation, and show that our estimates are unbiased.

We broadly follow the experimental design used in Crépon et al. [2013] in our simulations. We define 150 clusters, and randomly assign each cluster to have 0, 25, 50, 75 or 100 per cent of individuals assigned to treatment, $z_c \in \{0, 0.25, 0.5, 0.75, 1\}$. Each cluster contains 50 individuals, whom we randomly assign to treatment in these proportions, $P(z_i = 1|i \text{ in cluster } c) = z_c$. For the simulations presented in this section, we set $\kappa = 5, \varphi = 3, \phi = 1, \gamma = 2, \theta = 2$. We give more details of our simulation approach, and present results using other values of these key parameters in Section A.2.

We estimate two regressions using the 'naive' reduced form approach, both excluding and including the interaction term, $x_i\mathbf{N}_i\mathbf{x}$. In particular, we use z_i, z_c and their interaction as instruments for $x_i, \mathbf{N}_i\mathbf{x}$ and their interaction in two-stage least squares regressions that gives the following estimated outcomes equations:

$$y_i = \hat{c_0} + \hat{\bar{\kappa}}_{NE}x_i + \hat{\gamma}_{NE}\mathbf{N}_i\mathbf{x} + \epsilon_i \tag{40}$$

$$y_i = \hat{c_0} + \hat{\bar{\kappa}}_{NI}x_i + \hat{\gamma}_{NI}\mathbf{N}_i\mathbf{x} + \hat{\phi}_{NI}x_i\mathbf{N}_i\mathbf{x} + \epsilon_i \tag{41}$$

Note that this approach gives us an estimate of the 'average direct effect of treatment', $\bar{\kappa}$, rather than our structural parameter, κ.[22] This is because of the functional form of the heterogeneous effects, defined in section 2.1, which gives us $\bar{\kappa} = \kappa + \varphi\bar{z} + \bar{\xi}$, where $\bar{z}$ and $\bar{\xi}$ represent the mean values of z_i and ξ_i respectively. Note that $\tilde{\xi_i} \to_p \xi_i$ because $\bar{\xi_i} \to_p 0$.

Figure 1 shows the distribution of the parameter estimates from the TSLS approach excluding the interaction term, equation (40). It is clear that the estimates for $\bar{\kappa}$ have a large positive bias, although the estimates for γ are approximately normally distributed around the true parameter value (given by the red vertical line). The intuition for this can be easily seen by comparing the regression specification in equation (40) to the equilibrium outcomes equation in our model (equation (10)):

[21] As we discuss in more detail shortly, we define the 'average direct effect of the treatment' as $\bar{\kappa} = \kappa + \varphi\bar{z} + \bar{\xi}$. This is not analogous to the 'average treatment effect', or ATE. In our setting, the ATE would be the average marginal effect of an increase in treatment, which would include effects through the interaction term, $\mathbf{N}_i\mathbf{x}$, and second order 'equilibrium effects' that result from strategic interactions.

[22] This difference is reflected in the 'true value' of the parameters reported in Figures 1 and 2, represented by the vertical red line.

$y_i = c_0 + \theta x_i^2 + \gamma \mathbf{N}_i \mathbf{x} + \epsilon_i$. We can see that $\mathbf{N}_i \mathbf{x}$ appears linearly in both equations, and because our instruments are valid, we get unbiased estimates of γ. In contrast, the estimates of $\bar{\kappa}$ are biased because we are mistakenly estimating outcomes as a function of $\bar{\kappa} x_i$ rather than θx_i^2, and at these parameter values the bias is large.[23]

Figure 2 shows that estimating the outcomes equation including the interaction term, equation (41), by TSLS can actually make your estimates *worse*. In this case, the estimates of $\bar{\kappa}, \gamma$ and ϕ are all biased. Again, to understand why, it helps to compare the estimated equation (41) to the equilibrium outcomes equation (10): $y_i = c_0 + \theta x_i^2 + \gamma \mathbf{N}_i \mathbf{x} + \epsilon_i$. By including the interaction term, we are mistakenly introducing an extraneous variable that is correlated with both x_i and $\mathbf{N}_i \mathbf{x}$.

Finally, we use our two-step procedure to estimate the structural parameters and the corresponding direct effects, spillovers effects and their interaction. In contrast to the naive reduced form approach, Figure 3 shows that our estimates are unbiased and normally distributed around the true structural parameters.

[23] In these simulations, the equilibrium values of x_i are between 3 and 10, so the discrepancy between estimating an equation in x_i instead of x_i^2 is large, even though the values of the structural parameters θ and κ are relatively close.

Figure 1: Distribution of parameter estimates using the naive reduced form approach, excluding the interation term. 1000 simulations of the model, true parameter values shown in red.

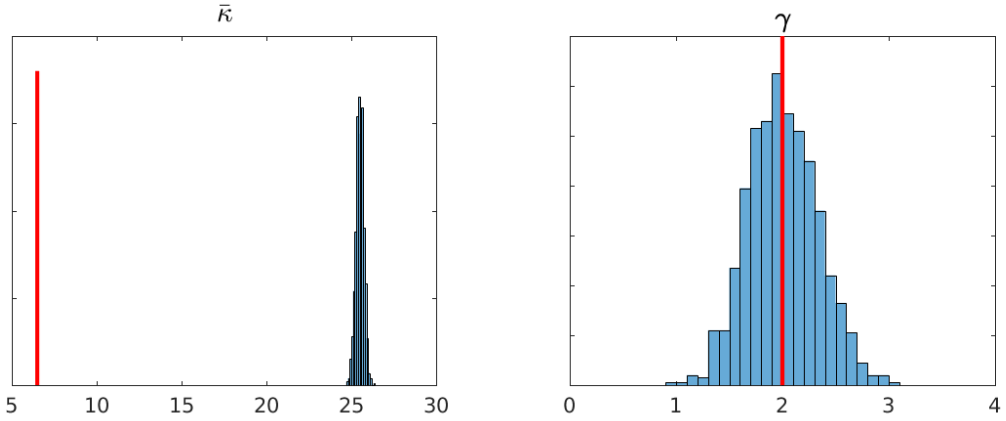


Figure 2: Distribution of parameter estimates using the naive reduced form approach, including the interation term. 1000 simulations of the model, true parameter values shown in red.

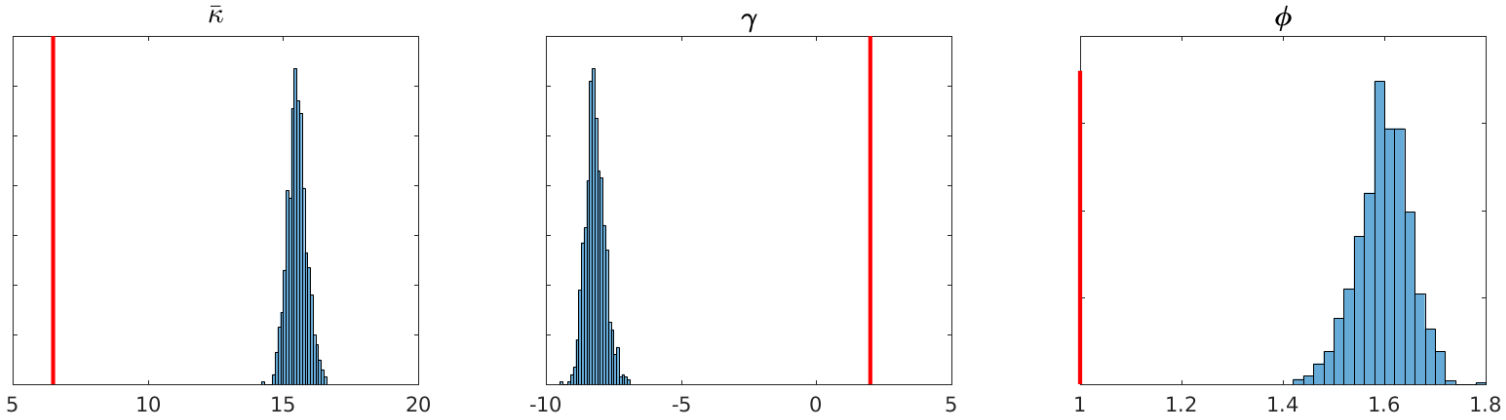
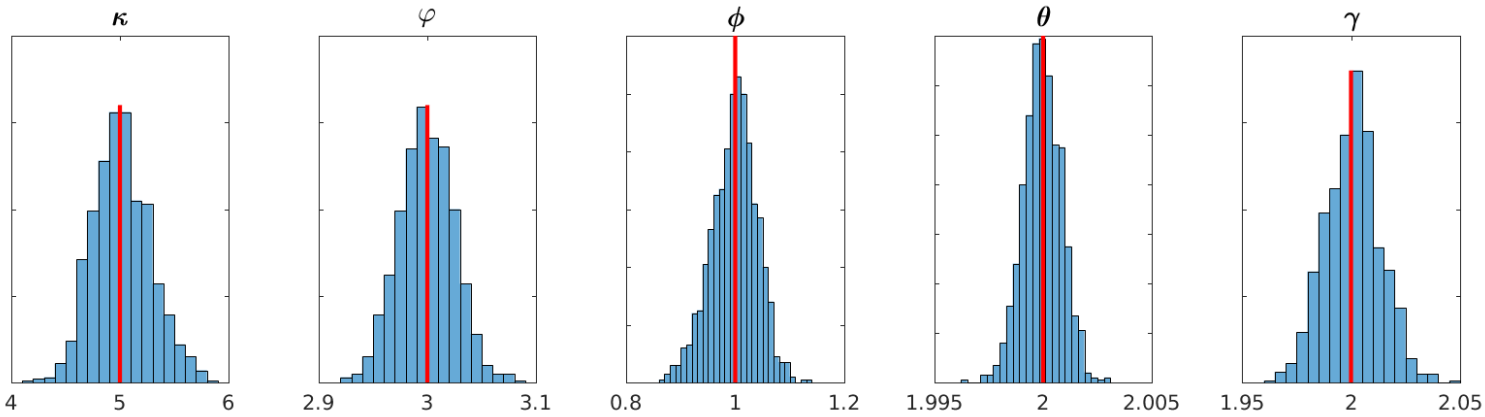


Figure 3: Distribution of estimated parameters using the two-step structural approach. 1000 simulations of the model, true parameter values shown in red.



4 Empirical Applications

We apply our two-step estimation procedure to data from two papers: Crépon et al. [2013], which evaluates the effects of a job placement assistance program in France, and Miguel and Kremer [2004], which estimates the effects of a deworming program in Kenya. The key variables for our analysis for each of the papers are summarized in Table 1. We give a short summary of the results of these tests here, and discuss these in more detail in sections 4.1 and 4.2.

We do not find evidence of strategic interactions in decisions to attend the job placement training program in France (the setting in Crépon et al. [2013]). This is not surprising, as the empirical context is unemployed individuals seeking employment in relatively large labor markets, with very limited opportunity for strategic interactions between individuals in the study. We therefore do not find evidence that the reduced form approach is biased, and do not proceed to the second step of our procedure.

However, we do find evidence of strategic interactions in the decisions of whether children would receive deworming treatment or not in Kenya (Miguel and Kremer [2004]). In particular, we estimate that children are more likely to receive deworming treatment if a higher proportion of their neighbors also receive treatment. This is intuitive: the 'payoff' for a child being dewormed is higher if more of their neighbors are also dewormed, as this will reduce the probability of reinfection after treatment. 'Naive' reduced form estimates of the outcomes equation that ignore strategic interactions will therefore produce biased results, so we estimate the structural parameters of our model and the corresponding direct and spillover effects of deworming treatment on school participation. As in Miguel and Kremer [2004], we estimate large positive spillover effects of deworming treatment, however our preliminary results imply significant heterogeneity in the direct effects of the treatment, and that these direct effects might be negative for the majority of individuals. We discuss these results in more detail in section 4.2.

Table 1: Summary of papers used for empirical applications

Paper	Program	Outcomes, y_i	Individual assignment, z_i	Treatment compliance, x_i	'Neighborhood' (cluster), N_i	Proportion assigned to treatment, z_c
Crepon et al (2013)	Job placement assistance program in France	Fixed term contract (8, 12, 16, 20 months after program)	11,806 unemployed individuals randomly assigned to job placement program	Program participation (compliance $\sim 1/3$)	235 labor markets (Local Employment Areas, LEAs)	Proportions of job seekers assigned to treatment randomized at cluster level (0%, 25%, 50%, 75%, 100%)
Miguel and Kremer (2004)	School-based deworming in Kenya	School participation. (Labor market outcomes in Baird et al. [2015])	75 schools randomly assigned to treatment and control	Children/families chose whether to receive treatment. 44% to 78% received treatment in each round.	Children within 1.2km, 1.8km and 2.4km of each school	Proportion of kids randomly assigned to treatment within a given radius

4.1 Job placement training in France

Crépon et al. [2013] uses a two-step randomized saturation design to estimate the effects of a job placement assistance program. In the first step, the proportion of job seekers to be assigned to treatment is randomized at the labor market level (Local Employment Area, LEA, which we will refer to as a 'cluster'), creating an instrument for the proportion of people receiving treatment in each cluster, $z_c \in \{0, 0.25, 0.5, 0.75, 1\}$. In the second step, eligible job seekers in these labor markets are randomly assigned to treatment or control, creating an instrument for individual receipt of treatment, $z_i \in \{0, 1\}$. We use these instruments to test for evidence of strategic interactions in compliance decisions, by estimating the regression in equation (5) and testing the null hypothesis $\beta_1 = 0$:

$$x_i = \beta_0 + \beta_1 \mathbf{N}_i \mathbf{x} + \beta_2 z_i + e_i$$

Our dependent variable is whether an individual participates in the job placement assistance program, and our explanatory variable of interest is the proportion of job seekers who participated in the program, which we instrument for using z_c. As there were no always-takers (nobody in the control group could access the program), we estimate the 'best response' equation on those assigned to treatment only (with $z_i = 1$, as in equation 30 in section 2.2.3).

Our estimates of the best response equation are given in Table 2. We present probit, OLS, 2SLS and ivprobit estimates, both with and without fixed effects.[24] The first four specifications, in which we do not instrument for the proportion of people participating in each cluster, give a large and significant estimate for β_1, suggesting that strategic interactions may be important. However, once we instrument for x_c using the random proportion of people assigned to treatment, z_c, the estimates for β_1 are close to zero and statistically insignificant (specifications 5 to 8).[25] The first stage of our 2SLS estimation confirms that our instrument is relevant, and the experimental design ensures that it is valid. We therefore do not find evidence of strategic interactions in compliance with treatment in this context, do not find evidence that reduced form estimates excluding the interaction term will be biased and do not execute the second step of our two-step procedure.

[24] Specifications (2), (4) and (6) include 'quintuplet' fixed effects and 'cohort' fixed effects, the same fixed effects used in the paper. A quintuplet is a group of 5 adjacent labor markets (clusters), and we use the authors' definition exactly. The experiment was rolled out in 14 monthly cohorts, and we also include these fixed effects, using cohorts 3-11 only, in line with the data used in the paper. We report robust standard errors, and standard errors clustered at the labour market level and the quintuplet level.

[25] The results therefore indicate that there are some unobservable characteristics that are correlated within a labor market (cluster) which are driving the observed relationship between program participation and the proportion of people participating in a cluster.

Table 2: Regressions of the best response equation, of compliance of person i on the proportion of people complying in person i's labor market (Local Employment Agency, LEA). Sample restricted to those assigned to treatment as those not assigned could not access the training program (there were no always-takers). Specifications (1) to (4) show positive correlation between x_i and $\mathbf{N}_i\mathbf{x}$, but the causal instrumental variable estimates (specifications (5), (6), (7) and (8)) are not significant.

VARIABLES	(1) Probit	(2) Probit	(3) OLS	(4) OLS	(5) 2SLS	(6) 2SLS	(7) ivprobit	(8) ivprobit
Cluster proportion received treatment	0.605	0.427	0.602	0.428	-0.0134	-0.0350	-0.0791	-0.155
	(0.0377)***	(0.0451)***	(0.0371)***	(0.0450)***	(0.0590)	(0.0606)	(0.164)	(0.169)
	[0.0384]***	[0.0396]***	[0.0369]***	[0.0387]***	[0.115]	[0.0817]		
	[0.0485]***	[0.0415]***	[0.0456]***	[0.0404]***	[0.0927]	[0.0942]		
Constant			0.188	0.188	0.355	0.327	-0.373	-0.453
			(0.0106)***	(0.0490)***	(0.0166)***	(0.0486)***	(0.0459)***	(0.133)***
			[0.0130]***	[0.0488]***	[0.0315]***	[0.0676]***		
			[0.0156]***	[0.0168]***	[0.0265]***	[0.0243]***		
Quintuplet FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	12,001	12,001	12,001	12,001	12,001	12,001	12,001	12,001
R^2			0.021	0.030	-0.001	0.021		
Psuedo R^2	0.0163	0.0237						
Uncentered R^2					0.351	0.365		
First stage								
Cluster proportion assigned to treatment					0.271	0.266		
					[0.0242]***	[0.0183]***		
Constant					0.0808	0.0996		
					[0.0149]***	[0.0587]*		
F-statistic					10357	13954		
F p-value					0	0		

Robust standard errors in round brackets. Standard errors clustered at the Local Employment Agency (LEA) level in first square brackets.
Unable to cluster standard errors for ivprobit using Newey's two step estimator. ivprobit first stage the same as 2SLS first stage.
Standard errors clustered at the quintuplet level (5 LEAs) in the second square brackets. *** p<0.01, ** p<0.05, * p<0.1
Marginal effects reported at the means of independent variables for Probit estimates; coefficients reported for ivprobit estimates.

4.2 Deworming treatment in Kenya

We apply our two-step methodology to data from Miguel and Kremer [2004], which evaluates the effects of a deworming project in Kenya in 1998. Although this paper used a cluster-randomization design rather than a 'randomized saturation' design, the authors calculate the proportion of people randomly assigned to treatment within a certain radius of each cluster, which gives exogenous variation in the intensity of treatment within that area. In particular, 75 schools were randomly assigned to either treatment or control, and the authors calculate the proportion of children assigned to treatment within 3km and 6km of each school. We take a similar approach, a detailed overview of which can be found in section A.3 in the Appendix, and consider the direct and spillover effects of deworming treatment on individuals' school participation rates, one of the key outcomes analyzed in the paper.[26]

4.2.1 Step 1: Testing for strategic interactions

We test for the presence of strategic interactions in individuals' decisions to receive deworming treatment by estimating the best response equation and testing the null hypothesis $\beta_1 = 0$:

$$x_i = \beta_0 + \beta_1 \mathbf{N}_i\mathbf{x} + \beta_2 z_i + e_i$$

Our dependent variable, x_i, is whether an individual child received deworming treatment, and our explanatory variable of interest, $\mathbf{N}_i\mathbf{x}$ is the proportion of children within a certain radius of their school who received deworming treatment. We use the proportion of children randomly assigned to treatment within a certain radius, z_c, as an instrument for $\mathbf{N}_i\mathbf{x}$. Note that because treatment was not available to students in the control groups (i.e. there are no 'always-takers' in the data), we restrict our sample to those assigned to treatment ($z_i = 1$). We estimate the best response equation using probit and IV probit, and present the results in Table 3.[27] We estimate the best response equation using cutoffs of 1.2km, 1.8km and 2.4km.[28]

[26] "The participation rate is computed among pupils enrolled in the school at the start of 1998. Pupils present in school during an unannounced NGO visit are considered participants. Pupils had 3.8 participation observations per year on average." Miguel and Kremer [2004], page 179.

[27] Because the intensity of treatment is defined at the school level (i.e. the distance radius is calculated from the school), we cannot use school-level fixed effects.

[28] We also estimated the best response using cutoffs of up to 6km (results shown in Table 8 in the Appendix), as this is the distance over which Miguel and Kremer [2004] originally tested the effects of deworming. However, the 2015 re-analysis of the paper (Aiken et al. [2015], Davey et al. [2015]) found evidence that spillovers occur at smaller distances, so we focused on cutoff distances between 1km and 3km). More detail about how we chose our cutoff distances is available in the appendix, in section A.3.

Table 3: The best response equation for the deworming data is estimated at cutoff distances of 1.2km, 1.8km and 2.4km, by Probit and IV Probit. The IV Probit estimates show that individuals are more likely to receive deworming treatment if their neighbors are also treated.

	Probit			IV Probit		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	1.2km	1.8km	2.4km	1.2km	1.8km	2.4km
Proportion received treatment, 0-1.2km	0.843			0.496		
	[0.0879]***			[0.0966]***		
	(0.207)***			(0.221)**		
Proportion received treatment, 0-1.8km		0.786			0.524	
		[0.0722]***			[0.0883]***	
		(0.229)***			(0.242)**	
Proportion received treatment, 0-2.4km			0.655			0.338
			[0.0771]***			[0.0841]***
			(0.227)***			(0.248)
Constant	0.109	0.121	0.203	0.264	0.236	0.326
	[0.0423]***	[0.0344]***	[0.0327]***	[0.0462]***	[0.0411]***	[0.0352]***
	(0.126)	(0.134)	(0.122)*	(0.135)*	(0.128)*	(0.115)***
Observations	5,305	7,846	8,437	5,305	7,846	8,437
First stage						
Proportion assigned to treatment				0.432	0.424	0.480
				[0.00215]***	[0.00291]***	[0.00208]***
				(0.0469)***	(0.0599)***	(0.0498)***
Constant				0.295	0.291	0.220
				[0.00187]***	[0.00224]***	[0.00125]***
				(0.0392)***	(0.0445)***	(0.0271)***
R^2				0.813	0.673	0.839
F-statistic				84.9	50.2	93.1
F p-value				0	0	0

Robust standard errors in square brackets, standard errors clustered at the school level in parentheses.
Probit coefficients reported. *** p<0.01, ** p<0.05, * p<0.1

Our estimates provide evidence that there are strategic interactions in the decision to receive deworming treatment. In particular, the IV probit estimates show that individuals are more likely to receive deworming treatment if their neighbors also receive treatment.[29] This makes intuitive sense, as spillovers occur through a reduced probability of worm infections, but the benefit of this reduced probability of infection will depend on an individual's current infection status, and therefore their treatment status. The evidence for strategic interactions is weaker when we use a 2.4km cutoff rather than 1.2km or 1.8km: the coefficient of interest is significant and positive when we use robust standard errors, but insignificant when we cluster the standard errors at the school level. Overall, we think this is reasonable evidence that ignoring strategic interactions might induce bias in reduced form estimates of the direct and spillover effects of treatment, and so estimate the structural parameters of our model using the second step of our procedure.[30]

4.2.2 Step 2: Estimating the structural parameters

In this empirical setting, children either receive treatment or do not, and only children assigned to treatment are able to access treatment. Therefore, to estimate the structural parameters of the model, we follow the method for binary treatment variables with one-sided non-compliance set out in section 2.2.3, and jointly estimate the outcomes equations by GMM. As in the previous section, x_i is the treatment status of an individual child, z_i is their assignment to treatment, and we use the proportion of children randomly assigned to treatment within a certain radius, z_c, as an instrument for the proportion that actually receive treatment, $\mathbf{N}_i\mathbf{x}$. We use school participation as the outcome variable for individual children, y_i, as this was a key focus of the initial paper, which found evidence of positive and significant direct and spillover effects of deworming treatment on school attendance.

In order to aid interpretation of the estimates, we display them in three ways. First, Table 4 compares the naive reduced form estimates of $\hat{\tilde{\kappa}}$ and $\hat{\gamma}$ (which exclude the interaction term) with the GMM estimates of the key structural parameters in our model from the second stage of our procedure. Second, Table 5 presents the GMM estimates of all of the structural parameters in the model. Finally, Table 6 presents the outcomes equation estimated by the naive reduced form approach and compares it to the key equations estimated using the structural model, and their associated marginal effects.

These estimates are preliminary and we have not yet calculated the standard errors for these esti-

[29] Although this finding may seem at odds with Kremer and Miguel [2007], which estimates that individuals are less likely to take up deworming treatment if they have social ties to people who were part of an early treatment group, we are estimating a fundamentally different relationship between own treatment status and neighbors' treatment status. Kremer and Miguel [2007] uses follow-up data on social networks in 2001 to estimate the effect of having social ties to people who were treated in 1998 and 1999, and therefore is estimating dynamic peer effects, such as learning and imitation. In contrast, we are estimating a static best response equation, in which individuals react to their neighbors' treatment contemporaneously. Our model and overall approach is static, and we do not consider dynamic considerations in estimating the effects of strategic interactions.

[30] *Note: the estimates presented here are very preliminary and do not include standard errors.*

Table 4: Parameter estimates using the 'naive' reduced form estimates (excluding the interaction term), and using the two-step procedure to estimate the structural parameters of the model, at each of our three cutoff distances.

	Reduced form		Two-step procedure			
Cutoff	$\hat{\tilde{\kappa}}$	$\hat{\gamma}$	$\widehat{\kappa+\varphi}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\theta}$
1.2km	0.063	0.26	-0.42	0.18	0.27	-0.81
1.8km	0.020	0.29	-0.47	0.16	0.32	-0.72
2.4km	0.052	0.27	-0.42	0.15	0.30	-0.81

Table 5: GMM estimates of all structural parameters in the model, at each of our three cutoff distances.

Cutoff	c_0	$\widehat{\kappa+\varphi}$	$\hat{\gamma}$	$\hat{\phi}$	$\hat{\theta}$	$\hat{\rho}$	$\hat{\sigma_\epsilon}$	$\hat{\sigma_\xi}$
1.2km	0.69	-0.42	0.27	0.18	-0.81	0.04	0.27	0.68
1.8km	0.69	-0.47	0.32	0.16	-0.72	0.10	0.27	0.48
2.4km	0.69	-0.42	0.30	0.15	-0.81	0.09	0.27	0.68

Table 6: Comparison of the outcomes equation estimated by the 'naive' reduced form approach with instrumental variables, and the key equations of the structural model, as estimated in step two of our procedure, for a 1.2km cutoff. Note - we display the same table for 1.8km and 2.4km cutoffs in section A.4.

'Naive' IV outcomes equation	$y_i = 0.69 + 0.063x_i + 0.26\mathbf{N}_i\mathbf{x} + \epsilon_i$
Structural outcomes equation	$y_i = 0.69 + (-0.42 + \xi_i)x_i + 0.27\mathbf{N}_i\mathbf{x} + 0.18x_i\mathbf{N}_i\mathbf{x} + \epsilon_i$
Average marginal effects, x_i	$\frac{\partial E[y_i\|x_i]}{\partial x_i} = -0.42 + 0.18E[\mathbf{N}_i\mathbf{x}] = -0.37$
Average marginal effects, $\mathbf{N}_i\mathbf{x}$	$\frac{\partial E[y_i\|\mathbf{N}_i\mathbf{x}]}{\partial \mathbf{N}_i\mathbf{x}} = 0.27 + 0.18E[x_i] = 0.32$
Structural individual utility	$U_i = y_i + 0.81x_i$
Structural best response function	$x_i = \begin{cases} 0 & if\ z_i = 0 \\ 1 & if\ z_i = 1\ and\ -0.42 + 0.81 + 0.18\mathbf{N}_i\mathbf{x} \geq -\xi_i \\ 0 & if\ z_i = 1\ and\ -0.42 + 0.81 + 0.18\mathbf{N}_i\mathbf{x} < -\xi_i \end{cases}$

mated parameters: they should therefore be interpreted with caution.[31] However, there are a few interesting results in these preliminary estimates that are worthy of note. First, both approaches estimate large, positive spillover effects of deworming treatment on school participation, a result that is consistent with the findings of Miguel and Kremer [2004] and Kremer and Miguel [2007]. Second, although the reduced form approach estimates a positive and significant direct effect of deworming treatment on school participation, the structural approach estimates $\widehat{\bar{\kappa}+\varphi} < 0$, which implies that the direct effect of treatment on school participation is negative on average.[32] However, the model allows for heterogeneous direct effects of treatment, and estimates that this heterogeneity is large (the standard deviation of the random element of the direct effect of treatment is given in Table 5). The model estimates that there will be substantial selection into treatment, and individuals with $\xi_i > -0.39$ will set $x_i = 1$, where we estimate that ξ_i is $N(0, 0.68^2)$ (for a cutoff of 1.2km).[33] While the model can account for this selection in its estimates, the reduced form estimates of the average treatment effect on the treated do not.[34] Finally, the model estimates $\theta < 0$, which implies that it is costly to *avoid* treatment (rather than being costly to receive treatment). This seems reasonable because children were treated at school and, to avoid their child being dewormed, parents either had to inform the head teacher or keep their child out of school on the day of treatment, both of which are potentially costly.

Although these results are preliminary and should be interpreted with caution, they imply that there may be significant differences between estimates using the naive reduced form approach, and estimates using the structural model following the procedure set out in section 2.2. In future work, we will prioritize conducting inference for the deworming estimates, and applying our methodology to other papers with exogenous variation in the intensity of treatment, such as Dupas [2014], Muralidharan et al. [2016] and Sinclair et al. [2012].[35]

[31] Calculation of the standard errors is complicated by a few factors. First, we have a GMM estimator with generated explanatory variables, the Heckman selection correction terms. Second, we estimate the structural parameters in two steps: an Probit IV gives estimates of $\tilde{\alpha}$ and $\tilde{\phi}$ from equations (31) and (32), and we must combine these with the GMM estimates from the outcomes equations to identify the structural parameters. Our estimated structural parameters are therefore combinations of random variables, and calculating their asymptotic distribution correctly is tricky. Finally, the second step of our procedure is conditional on the outcome of an (imperfect) model selection test in the first step, which will also affect the distribution of our estimated structural parameters. To correct for this, we will use 'selective inference': we will calculate the joint distribution of our two estimators (for the first and second steps) and then calculate the distribution of the second step estimator conditional on the decision rule of the first step. We may also want to correct standard errors for potential spatial correlation.

[32] This finding is similar to estimates in Kremer and Miguel [2007], which does not find evidence of significant private benefits of deworming treatment, and concludes that the majority of the benefit comes from positive spillover effects.

[33] The exact cutoff may be lower, depending on the proportion of people receiving treatment, $\mathbf{N}_i\mathbf{x}$. In areas with $\mathbf{N}_i\mathbf{x} = 0$, the cutoff is -0.39, and approximately 71.5% of individuals will comply with treatment. In areas with $\mathbf{N}_i\mathbf{x} = E[\mathbf{N}_i\mathbf{x}] = 0.276$, the cutoff is -0.44, and approximately 73.9% of individuals will comply with treatment. Finally in areas with $\mathbf{N}_i\mathbf{x} = 1$, the cutoff is -0.57, and 79.5% of individuals will comply with treatment.

[34] Note that selection into treatment will not affect estimates of the intent-to-treat effects of treatment, which are presented in Miguel and Kremer [2004].

[35] We have already tried applying our two-step procedure to a number of papers with exogenous variation in the intensity of treatment, but these have generally not had sufficient power to estimate the parameters of our model precisely. The papers we have explored applying our two-step procedure to include Angelucci et al. [2015], Baird et al.

5 Conclusion

Spillover effects may be important in the evaluation of the overall effects of many programs and policies, from health interventions to cash transfers. However, causal identification of these spillover effects is generally challenging. Recent research has made great improvements in the identification and estimation of spillover effects, in particular with the use of two step 'randomized saturation' experimental designs. However, we show that reduced form estimates of the direct effects and spillover effects of treatment are biased in the presence of strategic interactions, even when using high quality experimental data. We develop a two-step procedure, in which the first step tests for strategic interactions, and the second step gives bias-corrected estimates of the effects of treatment in their presence.

Our approach requires very high quality data, similar to that used in the most sophisticated papers that estimate the spillover effects of programs and policies. In particular, we need exogenous variation in both individual treatment and the intensity of neighbors' treatment to implement our two-step procedure. Such data is not always available, and challenges remain in the estimation of spillover effects of programs and policies, particularly in settings where effect sizes are likely to be small and diffuse. However, where appropriate data is available, researchers should test for the presence of strategic interactions and, if necessary, can use a simple model to get unbiased estimates of the structural parameters and treatment effects.

[2011] and Bobba and Gignoux [2014].

References

Daron Acemoglu, Camilo García-Jimeno, and James A Robinson. State capacity and economic development: a network approach. *The American Economic Review*, 105(8):2364–2409, 2015.

Alexander M Aiken, Calum Davey, James R Hargreaves, and Richard J Hayes. Re-analysis of health and educational impacts of a school-based deworming programme in western kenya: a pure replication. *International Journal of Epidemiology*, 44(5):1572–1580, 2015.

Manuela Angelucci, Dean Karlan, and Jonathan Zinman. Microcredit impacts: Evidence from a randomized microcredit program placement experiment by compartamos banco. *American Economic Journal: Applied Economics*, 7(1):151–82, 2015. doi: 10.1257/app.20130537. URL http://www.aeaweb.org/articles.php?doi=10.1257/app.20130537.

Sarah Baird, Craig McIntosh, and Berk Özler. Cash or condition? Evidence from a cash transfer experiment. *The Quarterly Journal of Economics*, page qjr032, 2011.

Sarah Baird, Joan Hamory Hicks, Michael Kremer, and Edward Miguel. Worms at work: Long-run impacts of a child health investment. Technical report, National Bureau of Economic Research, 2015.

Sarah Baird, J Aislinn Bohren, Craig McIntosh, and Berk Ozler. Optimal design of experiments in the presence of interference. *Review of Economics and Statistics*, forthcoming.

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

Yann Bramoullé, Rachel Kranton, and Martin D'Amours. Strategic interaction and networks. *The American Economic Review*, 104(3):898–930, 2014.

Bruno Crépon, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora. Do labor market policies have displacement effects? evidence from a clustered randomized experiment. *The Quarterly Journal of Economics*, 128(2):531–580, 2013.

Calum Davey, Alexander M Aiken, Richard J Hayes, and James R Hargreaves. Re-analysis of health and educational impacts of a school-based deworming programme in western kenya: a statistical replication of a cluster quasi-randomized stepped-wedge trial. *International Journal of Epidemiology*, page dyv128, 2015.

Pascaline Dupas. Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment. *Econometrica*, 82(1):197–228, 2014.

Michael Kremer and Edward Miguel. The illusion of sustainability. *The Quarterly Journal of Economics*, 122(3):1007–1065, 2007.

Edward Miguel and Michael Kremer. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, pages 159–217, 2004.

Karthik Muralidharan, Paul Niehaus, and Sandip Sukhtankar. General equilibrium effects of (improving) public employment programs: Experimental evidence from India. 2016.

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

A Appendices

A.1 Derivation of Heckman selection terms

We give the full derivation of the Heckman selection correction term for the untreated group in cases with a binary treatment variable and always takers (equation (22) in section 2.2.2). The first step is from an orthogonal decomposition, $\epsilon_i = \frac{\rho\sigma_\epsilon\xi_i}{\sigma_\xi} + \eta_i$ such that η_i is orthogonal to ξ_i.

$$
\begin{aligned}
&E[\epsilon_i|\xi_i < -(\kappa + \varphi z_i - \theta + \phi \mathbf{N}_i\mathbf{x})] \\
=&E\left[\frac{\rho\sigma_\epsilon\xi_i}{\sigma_\xi} + \eta_i|\xi_i < -(\kappa + \varphi z_i - \theta + \phi \mathbf{N}_i\mathbf{x})\right] \\
=&E\left[\frac{\rho\sigma_\epsilon\xi_i}{\sigma_\xi}|\xi_i < -(\kappa + \varphi z_i - \theta + \phi \mathbf{N}_i\mathbf{x})\right] + E\left[\eta_i|\xi_i < -(\kappa + \varphi z_i - \theta + \phi \mathbf{N}_i\mathbf{x})\right] \\
=&\frac{\rho\sigma_\epsilon}{\sigma_\xi}E\left[\xi_i|\xi_i < -(\kappa + \varphi z_i - \theta + \phi \mathbf{N}_i\mathbf{x})\right] \\
=&\frac{\rho\sigma_\epsilon}{\sigma_\xi}\frac{-\phi\left(\frac{-(\kappa+\varphi-\theta+\phi\mathbf{N}_i\mathbf{x})}{\sigma_\xi}\right)}{\Phi\left(\frac{-(\kappa+\varphi-\theta+\phi\mathbf{N}_i\mathbf{x})}{\sigma_\xi}\right)} = \frac{\rho\sigma_\epsilon}{\sigma_\xi}\lambda_i^0
\end{aligned}
$$

Similarly, we give the full derivation of the Heckman selection correction term for the treated group in cases with a binary treatment variable and always takers (equation (25) in section 2.2.2). The second step comes from the same orthogonal decomposition used for the untreated group above.

$$
\begin{aligned}
&E[\xi_i + \epsilon_i|\xi_i \geq -(\kappa + \varphi z_i - \theta + \phi \mathbf{N}_i\mathbf{x})] \\
=&E[\xi_i|\xi_i \geq -(\kappa + \varphi z_i - \theta + \phi \mathbf{N}_i\mathbf{x})] + E[\epsilon_i|\xi_i \geq -(\kappa + \varphi z_i - \theta + \phi \mathbf{N}_i\mathbf{x})] \\
=&E[\xi_i|\xi_i \geq -(\kappa + \varphi z_i - \theta + \phi \mathbf{N}_i\mathbf{x})] + \frac{\rho\sigma_\epsilon}{\sigma_\xi}E[\xi_i|\xi_i \geq -(\kappa + \varphi z_i - \theta + \phi \mathbf{N}_i\mathbf{x})] \\
=&\left(1 + \frac{\rho\sigma_\epsilon}{\sigma_\xi}\right)\frac{\phi\left(\frac{-(\kappa+\varphi-\theta+\phi\mathbf{N}_i\mathbf{x})}{\sigma_\xi}\right)}{1 - \Phi\left(\frac{-(\kappa+\varphi-\theta+\phi\mathbf{N}_i\mathbf{x})}{\sigma_\xi}\right)} = \left(1 + \frac{\rho\sigma_\epsilon}{\sigma_\xi}\right)\lambda_i^1
\end{aligned}
$$

A.2 Simulation of linear model with a continuous treatment variable

We use the linear model with a continuous treatment variable set out in section 2.2.1 to simulate data, and follow the experimental design used in Crépon et al. [2013] where possible. We use a relatively simple network structure, in which we define non-overlapping clusters of individuals. Individuals within a cluster are connected, but there are no network connections between clusters.[36] We define 150 clusters, each with 50 individuals, and randomly assign each cluster to have 0, 25, 50, 75 or 100 per cent of individuals assigned to treatment, $z_c \in \{0, 0.25, 0.5, 0.75, 1\}$. Given the z_c draws, we then assign individuals in each cluster to treatment using a Bernoulli draw: $P(z_i = 1|i \text{ in cluster } c) = z_c$. We draw correlated random shocks for each individual, (ϵ_i, ξ_i), and allow these to be correlated within clusters. We then calculate the unique interior solution using the vector equation analog of the best response given in equation (9):

$$\mathbf{x} = (I - \frac{\phi}{\theta}\mathbf{N})^{-1}\frac{1}{\theta}\mathbf{K} \tag{42}$$

where $\mathbf{K}$ is a vector of the κ_i terms. Finally, we generate a vector of outcomes equations using equation (10), and check whether the uniqueness condition holds for each simulation.[37]

In Section 3, we present results of our simulations using a 'baseline' set of parameters, which are also shown in the first row of Table 7. The key parameters of the model are set at: $\kappa = 5, \varphi = 3, \phi = 1, \gamma = 2, \theta = 2$. In simulations 2 to 11 in Table 7, we change exactly one of the key parameters, and leave other parameters of the model unchanged.[38] In particular, we use the following values for the other parameters in all of the simulations presented in this version of the paper: the constant term, c_0 is equal to zero; the output shock, ϵ_i, is normally distributed with a standard deviation, σ_ϵ, of 0.5; and the random shock to the direct effects of the treatment, ξ_i, is normally distributed with a standard deviation, σ_ξ, of $\sqrt{3}/2$, and it is correlated within clusters, and with ϵ_i.

[36] The network matrix, $\mathbf{N}$ is therefore block diagonal, with elements on the leading diagonal equal to zero and positive network connections within a cluster are row-normalized: $n_{ij} = 1/n_c$ if i, j are in the same cluster, where n_c is the number of individuals in a cluster, and $n_{ij} = 0$ otherwise.

[37] If $|\lambda_{min}(\mathbf{N}(\delta))| < \left(\frac{\phi}{\theta}\right)^{-1}$ then this interior equilibrium is unique, as shown in Bramoullé et al. [2014].

[38] In particular, simulations 2-4 vary ϕ, the effect of the interaction between x_i and $\mathbf{N}_i\mathbf{x}$ on outcomes, to see how the size of this parameter affects the bias. Simulations 5 and 6 vary φ, simulation 7 varies κ, simulations 8 and 9 vary γ, and simulations 10 and 11 vary θ.

Table 7: Parameter estimates using 'naive' reduced form (with and without the interaction term) and two-step structural bias correction procedure. Estimated on 200 simulated datasets for each of 11 sets of parameters. We report the mean and standard deviation of the estimates. 'Naive RF 1' is the naive reduced form estimation excluding the interaction term ('NE'), 'Naive RF 2' is the naive reduced form estimation including the interaction term ('NI').

	True parameters						Naive RF 1		Naive RF 2			Structural estimates: Best response			Structural estimates: Outcomes	
Sim	κ	φ	ϕ	γ	θ		$\hat{\bar{\kappa}}_{NE}$	$\hat{\gamma}_{NE}$	$\hat{\bar{\kappa}}_{NI}$	$\hat{\gamma}_{NI}$	$\hat{\phi}_{NI}$	$\hat{\bar{\kappa}}_{S}$	$\hat{\varphi}_{S}$	$\hat{\phi}_{S}$	$\hat{\theta}_{S}$	$\hat{\gamma}_{S}$
1	5	3	1	2	2	mean:	25.48	2.01	15.52	-8.16	1.6	5.02	3	1	2	2
						sd:	0.24	0.32	0.32	0.35	0.05	0.25	0.03	0.04	0	0.01
2	5	3	0	2	2	mean:	13	2.01	13.09	2.1	-0.03	5.02	3	-0.01	2	2
						sd:	0.11	0.14	0.53	0.55	0.16	0.25	0.03	0.08	0	0.02
3	5	3	-1	2	2	mean:	8.72	2.01	13.7	7.09	-2.33	5.02	3	-1.01	2	2
						sd:	0.09	0.13	0.65	0.68	0.3	0.25	0.03	0.11	0	0.02
4	5	3	1.5	2	2	mean:	49.05	2.01	24.84	-22.69	2.01	5.02	3	1.5	2	2
						sd:	0.53	0.7	0.35	0.4	0.02	0.25	0.02	0.02	0	0.01
5	5	1.5	1	2	2	mean:	22.54	2.01	13.87	-6.84	1.57	5.06	1.5	0.99	2	2
						sd:	0.22	0.3	0.67	0.7	0.12	0.45	0.03	0.08	0	0.03
6	5	-2	1	2	2	mean:	15.67	2.02	9.54	-4.25	1.6	5	-2	1	2	2
						sd:	0.22	0.29	0.34	0.33	0.08	0.24	0.03	0.06	0	0.02
7	3	3	1	2	2	mean:	17.64	2.01	10.74	-5.03	1.6	3.01	3	1	2	2
						sd:	0.24	0.32	0.24	0.26	0.05	0.18	0.03	0.04	0	0.01
8	5	3	1	0.2	2	mean:	25.48	0.21	15.52	-9.96	1.6	5.02	3	1	2	0.2
						sd:	0.24	0.32	0.32	0.35	0.05	0.25	0.03	0.04	0	0.01
9	5	3	1	5	2	mean:	25.48	5.01	15.52	-5.16	1.6	5.02	3	1	2	5
						sd:	0.24	0.32	0.32	0.35	0.05	0.25	0.03	0.04	0	0.01
10	5	3	1	2	0.2	mean:	-3.34	2	-2186	-2138	-265	5.02	3	1	0.2	2
						sd:	0.14	0.23	26183	25860	3186	0.27	0.03	0.03	0	0
11	5	3	1	2	5	mean:	16.17	2.01	13.31	-0.91	1.8	5.02	3	0.99	5	2
						sd:	0.14	0.18	0.47	0.49	0.29	0.25	0.03	0.16	0.01	0.04

A.3 Details of application of two-step procedure to deworming data

This section provides more detail about how we apply our two-step procedure to the data from Miguel and Kremer [2004]. The following bullet points describe how we constructed x_c and z_c.

- 75 schools were randomly assigned to one of three groups, each of which began treatment at a different time period. The 25 schools in Group 1 received free deworming treatment for their enrolled pupils in both 1998 and 1999, Group 2 schools received it in 1999, while Group 3 (the control group) did not receive free deworming until 2001. We focus only on the treatment decisions in the first year of the experiment, 1998, in our analysis because our model is static and does not explain decision-making in multiple periods.[39]

- Deworming treatment was carried out on a pre-announced day at children's schools. In order to 'opt out' of treatment in 1998, parents had to inform the head teacher. Alternatively, children who did not attend school on the day of deworming were not treated.[40]

- To estimate the spillover effects of the treatment, the paper calculates the number of school children within 3km (and within 3-6km) of each school, and the number of those children who are assigned to treatment. Following their approach, we calculated the following variables:

  $z_c^j = \frac{\text{\# of students assigned to treatment within } j\text{km}}{\text{Total \# of students within } j\text{km}}$ where j are cutoff distances that we choose. Note that the numbers of students used here *exclude* the students in school c to ensure that z_c is exogenous for each individual student, i.

  $x_c^j = \frac{\text{\# of students receiving treatment within } j\text{km}}{\text{Total \# of students within } j\text{km for which there is compliance data}}$. Note that the numbers of students used here *include* the students in school c as this is the relevant population for the best response of each individual student, i; x_c^j is therefore endogenous.[41]

Note that the 2015 'worm wars' controversy surrounding the replication and reanalysis of the original deworming studies (Aiken et al. [2015], Davey et al. [2015]) does not have an impact on our empirical analysis. The replication and reanalysis showed that spillovers are not present over 4-6km, but are still significant between schools less than 4km from each other; we focus on cutoff distances less than 4km.

A.3.1 Choosing the cutoff distances

There is some measurement error in the GPS coordinates in the data, as they are only given to two decimal places.[42] The left panel of Figure 4 plots the location of the GPS coordinates given for

[39] It is possible that agents were forwards-looking, and used their beliefs about treatment in future periods into account in their compliance decision. We can test for this in the data by testing whether treatment compliance in 1998 depends on the proportion of people assigned to treatment in 1999.

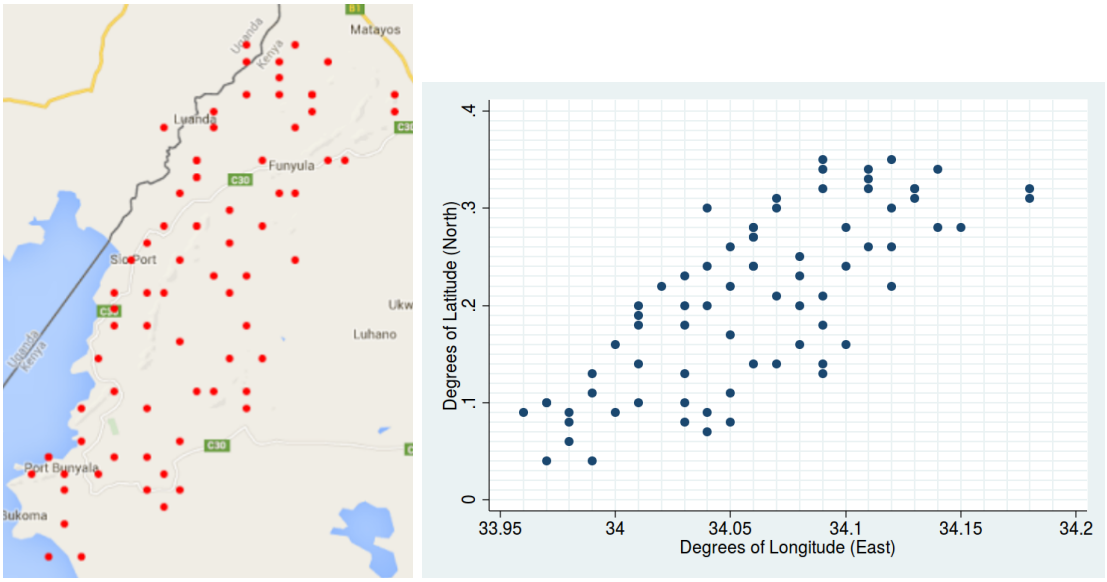
[40] For more detail, see page 169 of Miguel and Kremer [2004].

[41] Approximately 13% of observations do not have compliance data (see footnote 42 on page 189 of Miguel and Kremer [2004]) - we treat this data as missing at random.

[42] We have checked with the authors, and they confirmed that these were the most accurate GPS readings available at the time. It may be possible for us to update these coordinates using satellite images and/or school census data.

the schools in the data on a map, and shows that some of the points appear in Lake Victoria, or over the border in Uganda. The right panel shows that the points can be plotted on a grid. This coarse measurement limits the amount of variation in distance to another school seen in the data, and therefore limits the cutoff distances we can choose.[43]

Figure 4: There is some measurement error in the GPS coordinates in the data, as they are only given to two decimal places. When plotted on a map, some of the points appear in Lake Victoria, or over the border in Uganda. When plotted in Stata, these coordinates fit on a grid with lines 0.01 degrees apart.



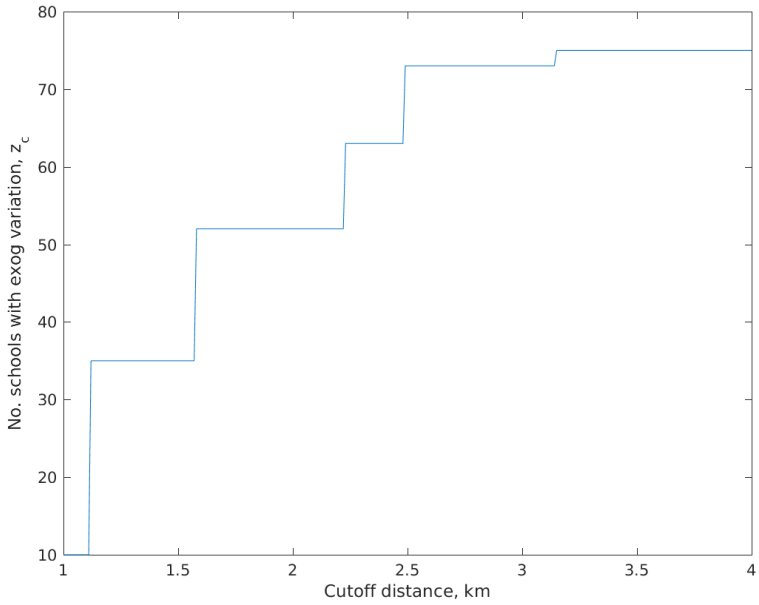
To get exogenous variation in the proportion of schools assigned to treatment within a given cutoff distance of school c, z_c, we must have at least one other school within that distance. Therefore, at small cutoff distances, there are not many schools with exogenous variation in z_c and this limits our sample size. Further, because the GPS coordinates of the schools are essentially approximated on a grid, we do not get much variance in the number of schools with exogenous variation in z_c as we increase the cutoff distance, as shown by Figure 5.

Figure 6 shows the distribution of our z_c^j and x_c^j variables, for cutoff distances of 1.2km, 1.8km and 2.4km. If we use a cutoff distance of 1.2km, we have exogenous variation in z_c in 35 schools, the majority of which only have one other school within the cutoff distance, making our instrument close to binary. There is more variation in z_c and x_c at cutoffs of 1.8km and 2.4km.Figure 7 shows a

However, the original data was collected in 1998, and it seems likely that some schools may have moved location, or changed name, since then, which might introduce measurement error.

[43]For example, we only see 5 minimum distances to another school in the data, which correspond to the horizontal, vertical or 'diagonal' distance between two points on a grid. The minimum distance to another school in the data are: 1.11km (the horizontal or vertical distance in between two gridpoints); 1.57km (the 'diagonal' distance between two gridpoints); 2.22km (twice the horizontal or vertical distance between two grid points); 2.48km (note $2.49 = \sqrt{1.11^2 + 2.22^2}$); and 3.15km (twice the 'diagonal' distance between two gridpoints).

Figure 5: The number of schools with exogenous variation in the proportion of students assigned to treatment within $j$km, z_c^j, increases in steps as we increase the distance cutoff. This is because the GPS coordinates of schools are only given to two decimal places.



scatter plot of x_c on z_c for each cutoff distance, and shows a strong positive correlation, suggesting that our instrument is relevant.

Figure 6: The distribution of z_c and x_c for each of our three cutoff distances. Note that not all schools have exogenous variation in z_c at smaller cutoff distances (number of schools given in parentheses).

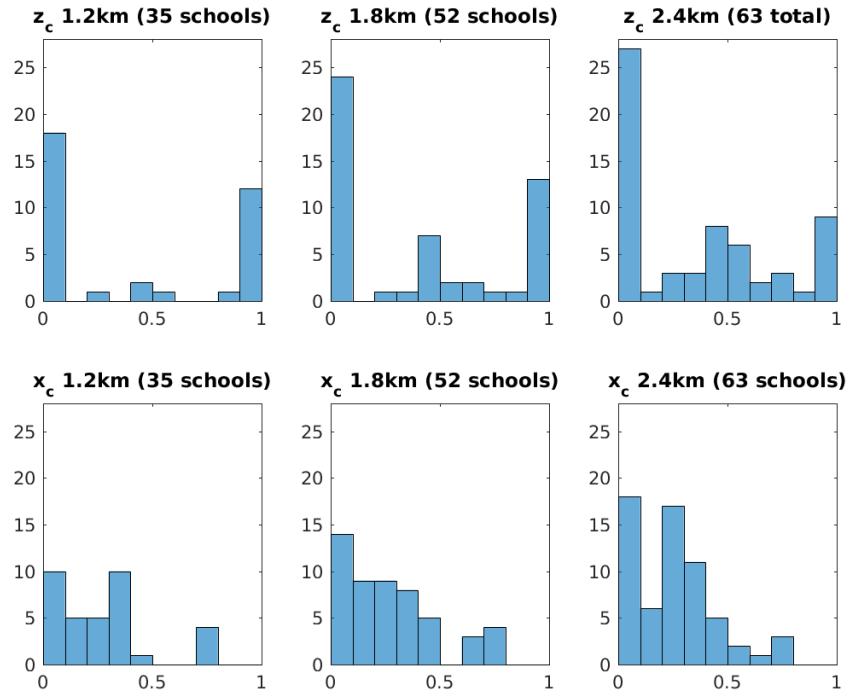


Figure 7: There is a positive correlation between x_c and z_c at each of our three cutoff distances, suggesting that our instruments are relevant at the 1.2km, 1.8km and 2.4km cutoff distances. Note that because z_c only varies at the school level, each school is an observation in these scatter plots.

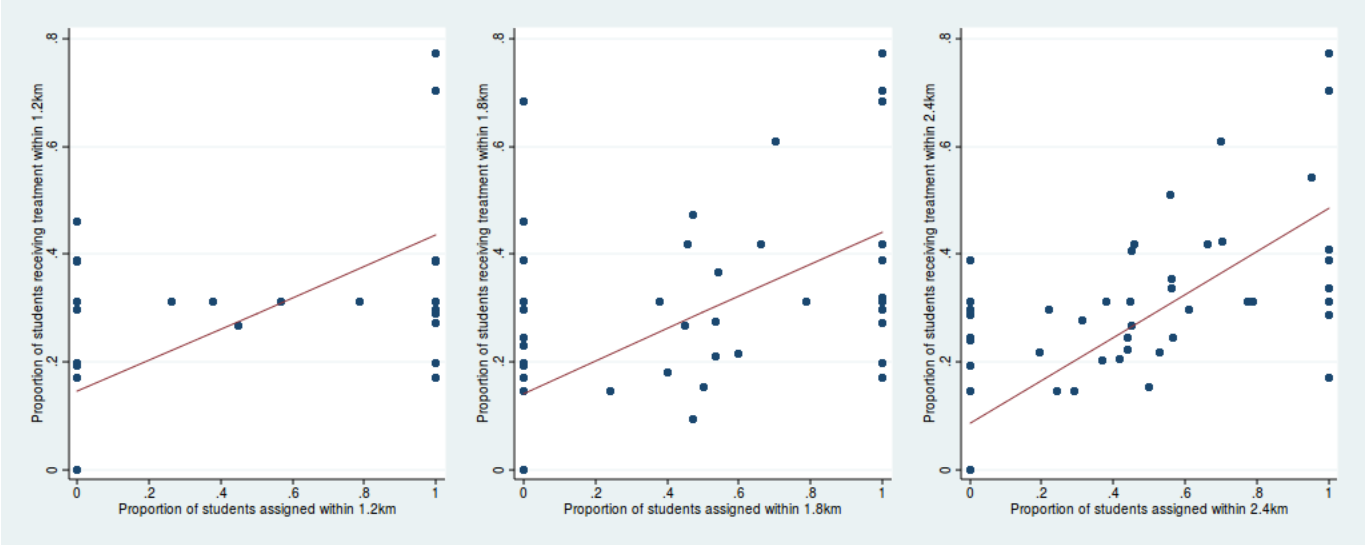


Table 8: Two stage least squares estimates of the best response equation for deworming, using cutoff distances from 1km to 6km. Note that there are no observations with exogenous variation in the proportion of individuals assigned to treatment using a cutoff of 1km, which explains the zero estimates in specification 1.

VARIABLES	(1) 1km	(2) 2km	(3) 3km	(4) 4km	(5) 5km	(6) 6km
Assigned to treatment, 1998	0.647***	0.651***	0.689***	0.677***	0.683***	0.684***
	(0.0386)	(0.0255)	(0.0149)	(0.0164)	(0.0179)	(0.0163)
Proportion received treatment, 0-1km	-0*					
	(0)					
Proportion received treatment, 0-2km		0.0930*				
		(0.0507)				
Proportion received treatment, 0-3km			-0.0368			
			(0.0363)			
Proportion received treatment, 0-4km				0.0248		
				(0.0318)		
Proportion received treatment, 0-5km					-0.0160	
					(0.0539)	
Proportion received treatment, 0-6km						-0.0489
						(0.0648)
Constant	0***	-0.0145*	0.00554	-0.00470	0.00304	0.0105
	(0)	(0.00819)	(0.00540)	(0.00599)	(0.0103)	(0.0139)
Observations	4,043	19,882	29,386	30,131	30,131	30,131
R^2	0.528	0.564	0.581	0.585	0.585	0.585
First stage						
Assigned to treatment, 1998	0.350***	0.320***	0.225***	0.147***	0.0796***	0.0594***
	(0.0714)	(0.0293)	(0.0288)	(0.0151)	(0.00914)	(0.00764)
Proportion assigned to treatment	0.263***	0.347***	0.407***	0.492***	0.593***	0.618***
	(0.0586)	(0.0354)	(0.0425)	(0.0282)	(0.0307)	(0.0215)
Constant	0	-0.00382	0.0239**	0.0109	0.00170	0.00307
		(0.0150)	(0.0113)	(0.00802)	(0.00804)	(0.00702)
Fstat	20.11	96.14	91.99	304	374.3	824.3
Fpval	0.00152	0	0	0	0	0

*** p<0.01, ** p<0.05, * p<0.1

Standard errors clustered at the school level

A.4 Deworming estimates summary tables: 1.8km and 2.4km cutoffs

Table 9: Comparison of the outcomes equation estimated by the 'naive' reduced form approach with instrumental variables, and the key equations of the structural model, as estimated in step two of our procedure, for a 1.8km cutoff.

'Naive' IV outcomes equation	$y_i = 0.70 + 0.020x_i + 0.29\mathbf{N}_i\mathbf{x} + \epsilon_i$
Structural outcomes equation	$y_i = 0.69 + (-0.47 + \xi_i)x_i + 0.32\mathbf{N}_i\mathbf{x} + 0.16x_i\mathbf{N}_i\mathbf{x} + \epsilon_i$
Average marginal effects, x_i	$\frac{\partial E[y_i\|x_i]}{\partial x_i} = -0.47 + 0.16E[\mathbf{N}_i\mathbf{x}] = -0.43$
Average marginal effects, $\mathbf{N}_i\mathbf{x}$	$\frac{\partial E[y_i\|\mathbf{N}_i\mathbf{x}]}{\partial \mathbf{N}_i\mathbf{x}} = 0.32 + 0.16E[x_i] = 0.36$
Structural individual utility	$U_i = y_i + 0.72x_i$
Structural best response function	$x_i = \begin{cases} 0 & if\ z_i = 0 \\ 1 & if\ z_i = 1\ and\ -0.47 + 0.72 + 0.16\mathbf{N}_i\mathbf{x} \geq -\xi_i \\ 0 & if\ z_i = 1\ and\ -0.47 + 0.72 + 0.16\mathbf{N}_i\mathbf{x} < -\xi_i \end{cases}$

Table 10: Comparison of the outcomes equation estimated by the 'naive' reduced form approach with instrumental variables, and the key equations of the structural model, as estimated in step two of our procedure, for a 2.4km cutoff.

'Naive' IV outcomes equation	$y_i = 0.70 + 0.052x_i + 0.27\mathbf{N}_i\mathbf{x} + \epsilon_i$
Structural outcomes equation	$y_i = 0.69 + (-0.42 + \xi_i)x_i + 0.30\mathbf{N}_i\mathbf{x} + 0.15x_i\mathbf{N}_i\mathbf{x} + \epsilon_i$
Average marginal effects, x_i	$\frac{\partial E[y_i\|x_i]}{\partial x_i} = -0.42 + 0.15E[\mathbf{N}_i\mathbf{x}] = -0.39$
Average marginal effects, $\mathbf{N}_i\mathbf{x}$	$\frac{\partial E[y_i\|\mathbf{N}_i\mathbf{x}]}{\partial \mathbf{N}_i\mathbf{x}} = 0.30 + 0.15E[x_i] = 0.34$
Structural individual utility	$U_i = y_i + 0.81x_i$
Structural best response function	$x_i = \begin{cases} 0 & if\ z_i = 0 \\ 1 & if\ z_i = 1\ and\ -0.42 + 0.81 + 0.15\mathbf{N}_i\mathbf{x} \geq -\xi_i \\ 0 & if\ z_i = 1\ and\ -0.42 + 0.81 + 0.15\mathbf{N}_i\mathbf{x} < -\xi_i \end{cases}$