Estimating heterogeneous peer effects with partial population experiments

Felix Pasquier
CREST-ENSAE

Pauline Rossi
CREST-École Polytechnique

Zheng Wang
CREST-École Polytechnique

DRAFT – February 27, 2024

[PRELIMINARY AND INCOMPLETE - PLEASE DO NOT QUOTE]

Abstract

The standard linear-in-means model of peer effects assumes that the endogenous peer effect parameter is homogeneous. We relax this assumption by allowing individuals to respond differently to the outcomes of other group members depending on the identity of these members. Specifically, we distinguish peer effects *within* members sharing the same identity and peer effects *between* members of different identities. We propose a simple methodology to identify and estimate the model using partial population experiments (i.e. designs in which only some individuals in a group are eligible for treatment) with variation in the share of eligible individuals across groups. We discuss two cases: randomized experiments and differences-in-differences. The estimation procedure builds on the Generalized Method of Moments. We are able to quantify the direct effect on the eligibles, the indirect effect on the non-eligibles and the population multipliers. We apply our methodology to study peer effects in school attendance using a conditional cash transfer program targeting the poor in Mexico.

KEYWORDS: PEER EFFECTS, PARTIAL POPULATION EXPERIMENTS, HETEROGENEOUS EFFECTS

1 Introduction

In the field of peer effects, applied economists often use the linear-in-means model. The model postulates that individuals are influenced by the average outcome in a given reference group. This is a way to formalize the concept of conformism: people incur a cost when their behavior is different from the average behavior. An important parameter in this model is the *endogenous peer effect*, which captures the change in the outcome of an individual in response to a change in the group average outcome. The endogenous peer effect is assumed to be homogeneous, in the sense that (i) all individuals in a reference group respond to the change in the same way, and (ii) all individuals have the same weight in the group average.

This assumption fails to account for complex group dynamics, such as polarization, social distinction, or role models. Sociologists have long documented that the relevant reference group may consist of several sub-groups who do not necessarily interact in a symmetric way. Some individuals may put more weight on their own sub-group; other individuals may put more weight on the other sub-groups, either emulating them or opposing them. These dynamics depend on what defines the identity of these sub-groups, their number and relative size, as well as the social hierarchy.

In this paper, we relax the homogeneity assumption by allowing individuals to respond differently to changes in the outcomes of other group members depending on the identity of these members. We distinguish peer effects *within* members sharing the same identity and peer effects *between* members of different identities. We propose a methodology to identify and estimate these parameters using experiments in which only a share of the population is eligible for treatment. These designs are called partial population experiments. We discuss two cases: (i) controlled experiments, where the treatment is randomly allocated to groups; and (ii) natural experiments, more specifically differences-in-differences, where the treatment is allocated to different groups at different times and the common trend condition holds.

First, we show that we can identify within and between peer effects when the share of eligibles varies across groups; in theory, two values are enough for identification. Second, we propose a simple estimation procedure building on the Generalized Method of Moments (GMM). In theory, we rely on well-established results in this literature. In practice, we use packages provided by standard statistical software. Third, we provide simulations showing that our procedure works well when we observe around 500 groups. We show how to include covariates to improve precision if the number of groups is small, or to improve the plausibility of the common trend assumption. We also discuss how many treated and control groups with a similar share of eligibles are needed to get precise estimates. The procedure is not computationally intensive, making it accessible to a wide audience.

Estimating endogenous peer effects is particularly important from a policy evaluation perspective because they generate population multipliers. The treatment is a shock affecting the behaviors of eligible individuals; then, non-eligibles respond to the change in eligibles' behaviors; and the shock propagates further (between and within sub-groups) until a new equilibrium is reached. Our estimates are useful to quantify (i) the *direct effect* of the policy on the eligibles, (ii) the *indirect effect* of the policy on the non-eligibles and (iii) the population multipliers, i.e. by how much the direct and indirect effects are amplified by social interactions between and within sub-groups. In contrast, research designs based on the random allocation of peers cannot separately identify whether individuals are influenced by the behavior of their peers (endogeneous peer effects) or by the characteristics of their peers (exogeneous peer effects). Separating both effects is crucial since only the endogenous component gives rise to population multipliers.

We apply our methodology to explore peer effects in school attendance using a conditional cash transfer program in Mexico – Progresa. This program has been extensively studied: it was randomly allocated across villages and it targeted poor households. Previous research finds a positive effect on school attendance of the eligibles and a positive spillover effects on school attendance of the non-eligibles belonging to the same group, defined as grade $\times$ village. For instance, Lalive and Cattaneo (2009) estimate a direct effect of $3p.p.$: in the absence of social interactions, school attendance among the poor should increase by 3 percentage points in treated villages. They estimate an homogeneous endogenous peer effect parameter of 0.5: the individual probability of attending school increases by 5 percentage points when the average attendance in a group increases by 10 percentage points. When we allow for heterogeneity, we estimate a "between" peer effect of 0.6, a "within" peer effect of 0.9, and a direct effect of $2p.p.$. We derive two conclusions in this context. First, conformism is extremely strong within social classes (poor and non-poor) and relatively strong between social classes. Second, failing to account for heterogeneous peer effects leads to underestimating the contribution of social interactions and overestimating the direct effect of the policy.

Other potential applications cover important topics in education (e.g. estimate peer effects in graduation rates using scholarships targeting some categories of students and variation across majors), labor (e.g. estimate peer effects in parental leave take-up using collective agreements targeting some professions and variation across neighborhoods), health (e.g. estimate peer effects in contraception take-up using rules specific to minors and variation across classes), political economy (e.g. estimate peer effects in support for local authorities using public infrastructure devoted to the elderly or to young children and variation across neighborhoods), and crime (e.g. estimate peer effects in criminal activities using interventions targeting at-risk youth and variation across classes). Our methodology is adequate in settings where the non-eligibles are affected by the treatment only through changes in the eligibles' outcomes. By assumption, we rule out any direct effect on the non-eligibles.[1]

Our paper relates to the literature using partial population experiments to identify the endogenous peer effect parameter in linear-in-means models (Moffitt et al. (2001); Bobonis and Finan (2009); Brown and Laschever (2012); Hirano and Hahn (2010)). We contribute to this literature by allowing for heterogeneous parameters. Relaxing the homogeneity assumption is important because this assumption restricts the shape of the population multipliers: they have to be the same for eligibles and non-eligibles, and they have to be linear in the share of eligibles (s). For instance, the population multiplier is equal to $2s$ in Lalive and Cattaneo (2009) This restriction implies that the total effect on the eligibles and the total effect on the non-eligibles should be linear in s, a prediction which is not always verified in the data. By allowing for different "between" and "within" parameters, we can rationalize diverse empirical patterns: convex or concave, increasing or decreasing, depending on the relative magnitudes and on the signs of the "between" and "within" parameters. In the case of Progresa, we find that the total effects are increasing and convex, and that the multiplier is larger for the non-eligibles than for the eligibles.

Our paper also relates to the literature on the estimation of spillover effects in experiments. Previous research has studied how the outcome of an individual is affected by the treatment status of her peers in a non-parametric way (Hudgens and Halloran (2008), Tchetgen and VanderWeele (2012)). In particular, Vazquez-Bare (2023) discusses the case of heterogeneous spillover effects in a flexible framework, allowing the treatment status of peers to have heterogeneous effects depending on the characteristics of the peers. The interpretation of the parameters in this

[1]For instance, in the case of Progresa, this assumption could be violated if cash transfers were shared with non-poor households, or if they caused inflation in the village: the treatment would directly affect the non-eligibles' budget constraint. Lalive and Cattaneo (2009) provide arguments to rule out these possibilities.

approach is reduced-form, in the sense that they capture different mechanisms: endogenous peer effects, exogenous peer effects, general equilibrium effects through prices, etc. In contrast, our approach focuses on a specific mechanism. We can only study a subset of the questions explored in the literature. However, for these questions, we have a structural interpretation of the parameters and we learn something fundamental about the drivers of behaviors. We learn about how individuals weigh other members of their reference groups, whether they imitate some members in particular, or try to distinguish themselves from other members. Our results are useful to deepen our understanding of social influence, in particular to study dynamics going beyond plain conformism.

2 A Partial-Population Design

2.1 Randomized Experiment

We consider a setting inspired by Moffitt et al. (2001). We suppose that G reference groups are observed. Group g has n_g units (indexed by i); and a share s_g of units are eligible for a binary treatment, denoted D_g. Some groups are treated while others are not. Let E_{ig} be the binary variable indicating whether or not unit i in group g is eligible for the treatment and let D_{ig} be a binary variable indicating whether or not the unit has received the treatment. By construction, $P(D_{ig} = 1|E_{ig} = 0) = 0$. If the group is treated, all eligible units are effectively treated: $P(D_{ig} = 1|E_{ig} = 1, D_g = 1) = 1$. Let $n_g^E = \sum_{i=1}^{n_g} E_{ig} = s_g n_g$ and $n_g^{NE} = \sum_{i=1}^{n_g}(1 - E_{ig}) = n_g - n_g^E$ be the number of units that are respectively eligible and non-eligible for the treatment in group g. We consider the case where eligibility E_{ig} is *not* randomly determined: eligibility reflects an observable characteristic, that we call "identity". For instance, in the case of Progresa, a reference group is a grade level in a village, the treatment is a cash transfer, and eligibility is based on wealth.

We start from the commonly used linear-in-means (LIM) model of social interactions (Manski (1993), Blume et al. (2011), Bramoullé, Djebbari, and Fortin (2009)), defined as

$$y_{ig} = \alpha_g + x_{ig}'\eta_0 + z_g'\gamma + \delta_0 D_{ig} + \theta_0 \times \frac{1}{n_g}\sum_{i=1}^{n_g} y_{ig} + \varepsilon_{ig} \qquad (1)$$

where y_{ig} is individual i's scalar outcome, z_g is a vector of attributes characterising individual i's reference group g, including peer characteristics within the group, x_{ig} and ε_{ig} are respectively individual i's observed (resp. unobserved) attributes that directly affect y_{ig}. If $\theta_0 \neq 0$, this model expresses an endogenous peer effect: individual i's outcome varies with the mean of her peers in group g's outcomes. It is standard to rule out explosive trajectories by assuming that $|\theta_0| < 1$. The vector (η_0, δ_0) captures the exogenous effects while α_g captures the correlated effects, as defined by Manski (1993). Moffitt et al. (2001) shows that all parameters, and thus all different types of effects, are identified as long as $\delta_0 \neq 0$.

We extend the standard model by allowing for heterogeneous endogenous peer effects. Units may be differently influenced by their peers' outcomes depending on whether they have the same "identity" or not. Let $\mathcal{E}_g = \{i \in \{1, \dots, n_g\} : E_{ig} = 1\}$ be the set of individuals in group g whose "identity" makes them eligible for the treatment. Let $\mathcal{O}_g = \{1, \dots, n_g\} \backslash \mathcal{E}_g$ be the set of individuals in group g whose "identity" makes them non-eligible for the treatment. In the following, for any variable w, $\bar{w}_g^E = \frac{1}{n_g^E}\sum_{i \in \mathcal{E}_g} w_{ig}^E$ and $\bar{w}_g^{NE} = \frac{1}{n_g^{NE}}\sum_{i \in \mathcal{O}_g} w_{ig}^{NE}$. We make the following assumptions.

Assumption 1 (Partial-Population Design with Heterogeneous Endogenous Effects)**.** *Within each group $g \in \{1, \ldots, G\}$ of the i.i.d sample,*

$$y_{ig}^{E} = \theta_0^w s_g \bar{y}_g^E + \theta_0^b (1 - s_g) \bar{y}_g^{NE} + \delta_0 D_{ig} + e_{ig}^E, \quad i \in \mathcal{E}_g \tag{2}$$

$$y_{ig}^{NE} = \theta_0^b s_g \bar{y}_g^E + \theta_0^w (1 - s_g) \bar{y}_g^{NE} + e_{ig}^{NE}, \quad i \in \mathcal{O}_g \tag{3}$$

The parameters of interest are $\theta_0^b, \theta_0^w \in (-1, 1)$. They capture the "between" and "within" endogenous peer effects. When the group average outcomes for both eligible and non-eligible are increased by 1, an eligible unit's outcome will increase by $\theta_0^w s_g + \theta_0^b(1 - s_g)$, a weighted average of the within and the between endogenous peer effects, weighted by the share of the eligible units. To clarify the impact of (θ_0^w, θ_0^b) and s_g, consider a scenario where $s_g = 0.5$, indicating an equal number of eligible and non-eligible units within a group. Under this condition, a simultaneous increase in $\bar{y}_g^E$ and $\bar{y}_g^{NE}$ by Δ^E and Δ^{NE} respectively, results in an increase in the eligible unit's outcome by $\frac{1}{2}(\theta_0^w \Delta^E + \theta_0^b \Delta^{NE})$. This demonstrates that (θ_0^w, θ_0^b) quantifies the relative impact of within-group and between-group peer effects. Conversely, when $\theta_0^w = \theta_0^b$, the increase in an eligible unit's outcome depends linearly on the group composition, reflecting a stronger response to peers' average outcomes when they constitute a larger proportion of the group.
$\delta_0 \in \mathbb{R}^*$ captures the direct effect of treatment on the eligible units, which is assumed to be a constant across groups. For conciseness, the residual terms e_{ig}^E and e_{ig}^{NE} capture all the individual, exogenous and correlated effects. One may think of these terms, for example, as

$$\begin{aligned} e_{ig}^{E} &= \alpha_g^E + x_{ig}' \eta_0^E + z_g' \gamma_0^E + \varepsilon_{ig}^E, \quad i \in \mathcal{E}_g \\ e_{ig}^{NE} &= \alpha_g^{NE} + x_{ig}' \eta_0^{NE} + z_g' \gamma_0^{NE} + \varepsilon_{ig}^{NE}, \quad i \in \mathcal{O}_g \end{aligned}$$

Because our focus is on θ_0^b and θ_0^w, we do not express explicitly all these effects in our baseline model. From Assumption 1, we can average respectively y_{ig}^E and y_{ig}^{NE} among eligible (resp. non-eligible) units in group g to get

$$\begin{aligned} \bar{y}_g^{E} &= \theta_0^w s_g \bar{y}_g^E + \theta_0^b (1 - s_g) \bar{y}_g^{NE} + \delta_0 D_g + \bar{e}_g^E \\ \bar{y}_g^{NE} &= \theta_0^b s_g \bar{y}_g^E + \theta_0^w (1 - s_g) \bar{y}_g^{NE} + \bar{e}_g^{NE} \end{aligned}$$

After some development, we get the following reduced forms

$$\begin{aligned} \bar{y}_g^E = \; & \frac{\theta_0^b (1 - s_g)}{1 - \theta_0^w + s_g(1 - s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \bar{e}_g^{NE} \\ & + \frac{1 - \theta_0^w (1 - s_g)}{1 - \theta_0^w + s_g(1 - s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \bar{e}_g^{E} \\ & + \delta_0 \left(1 + s_g . \frac{\theta_0^w - (1 - s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}{1 - \theta_0^w + s_g(1 - s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\right) D_g \end{aligned} \tag{4}$$

$$\begin{aligned} \bar{y}_g^{NE} = \; & \frac{1 - \theta_0^w s_g}{1 - \theta_0^w + s_g(1 - s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \bar{e}_g^{NE} \\ & + \frac{\theta_0^b s_g}{1 - \theta_0^w + s_g(1 - s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \bar{e}_g^{E} \\ & + \frac{\delta_0 \theta_0^b s_g}{1 - \theta_0^w + s_g(1 - s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} D_g \end{aligned} \tag{5}$$

Then, we specify the treatment assignment across groups.

Assumption 2 (Randomized Experiment)**.** *The group level treatment D_g is randomly assigned, i.e.*

$$(\bar{e}_g^{NE}, \bar{e}_g^{E}, s_g) \perp\!\!\!\perp D_g \tag{6}$$

Assumption 2 states that treatment is randomly assigned. In particular, groups receive the treatment independently of their share of eligible units. We also impose a restriction on the support of D_g conditional on the share of eligible units s_g. Let $\mathcal{S} \subseteq (0,1]$ be the support of shares that are observed in the sample,

Assumption 3 (Common Support)**.** *For each $s \in \mathcal{S}$,*

$$0 < P(D_g = 1|s) < 1 \tag{7}$$

Assumption 3 states that for any share of eligible units that is observed in the population of reference groups, there exist some groups that are treated and some that are not.

2.2 Natural Experiment

In this section, we consider a panel extension of our baseline model. The G groups may be observed over T time periods indexed by t. Group g treatment status for period t is given by D_{gt}. For $k \in \{E, NE\}$, $\bar{y}_{gt}^k = \left(n_{gt}^k\right)^{-1} \sum_{i=1}^{n_{gt}^k} y_{igt}^k$ where n_{gt}^k is the number of units whose "identity" is k at time period t. Let $\mathcal{E}_{gt} := \{i \in \{1, \dots, n_{gt}\} : E_{igt} = 1\}$ and $\mathcal{O}_{gt} := \{i \in \{1, \dots, n_{gt}\} : E_{igt} = 0\}$ be respectively the set of eligible (resp. non-eligible) units in group g at time t.

Assumption 1′ (Panel Partial-Population Design with Heterogeneous Endogenous Effects)**.** *Within each group g in $\{1,\dots, G\}$ present in the i.i.d panel, for each time period $t \in \{1, \dots, T\}$,*

$$y_{igt}^E = \theta_0^w s_{gt} \bar{y}_{gt}^E + \theta_0^b (1 - s_{gt}) \bar{y}_{gt}^{NE} + \delta_0 D_{gt} + e_{igt}^E, \quad i \in \mathcal{E}_{gt} \tag{8}$$

$$y_{igt}^{NE} = \theta_0^b s_{gt} \bar{y}_{gt}^E + \theta_0^w (1 - s_{gt}) \bar{y}_{gt}^{NE} + e_{igt}^{NE}, \quad i \in \mathcal{O}_{gt} \tag{9}$$

Assumption 1′ is the counterpart of Assumption 1 in a panel data context. Note that we rule out time dependence, in the sense that former own outcomes and former peers' outcomes do not directly influence current outcomes. As before, we can average the outcomes of eligible and non-eligible units in group g for each time period t and develop to get the following reduced forms

$$\begin{aligned} \bar{y}_{gt}^E = & \frac{\theta_0^b (1 - s_{gt})}{1 - \theta_0^w + s_{gt}(1 - s_{gt}) \left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \bar{e}_{gt}^{NE} \\ & + \frac{1 - \theta_0^w (1 - s_{gt})}{1 - \theta_0^w + s_{gt}(1 - s_{gt}) \left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \bar{e}_{gt}^{E} \\ & + \delta_0 \left(1 + s_{gt} . \frac{\theta_0^w - (1 - s_{gt}) \left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}{1 - \theta_0^w + s_{gt}(1 - s_{gt}) \left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\right) D_{gt} \end{aligned} \tag{10}$$

$$\begin{aligned} \bar{y}_{gt}^{NE} = & \frac{1 - \theta_0^w s_{gt}}{1 - \theta_0^w + s_{gt}(1 - s_{gt}) \left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \bar{e}_{gt}^{NE} \\ & + \frac{\theta_0^b s_{gt}}{1 - \theta_0^w + s_{gt}(1 - s_{gt}) \left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \bar{e}_{gt}^{E} \\ & + \frac{\delta_0 \theta_0^b s_{gt}}{1 - \theta_0^w + s_{gt}(1 - s_{gt}) \left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} D_{gt} \end{aligned} \tag{11}$$

For the rest of the paper, we suppose that $T = 2$. However, the natural experiment setting we consider could potentially be extended to multiple periods. In this setting, treatment assignment may not be random but we make other restrictions on the distribution of the data.

Assumption 2′ (Stable Shares)**.** *For all* $(g, t) \in \{1, \ldots, G\} \times \{1, 2\}$,

$$s_{gt} = s_g$$

Assumption 2′ states that, for each group, the share of eligible units does not change from period 1 to period 2. It implies in particular that the composition of groups is not affected by the treatment. This assumption is thus credible in designs where the time span between the 2 periods is short.

Assumption 3′ (Treatment Distribution)**.**

$$D_{g1} = 0 \text{ a.s and for all } s \in \mathcal{S}, 0 < P(D_{g2} = 1|s) < 1 \tag{12}$$

Assumption 3′ states, first, that no group is treated at the initial period. Second, for all the values of shares of eligible units that are observed in the population, there are some treated groups and some control groups.

Assumption 4′ (Conditional Common Trends)**.** *For any* $k \in \{NE, E\}$,

$$\mathbb{E}[\bar{e}^k_{g2} - \bar{e}^k_{g1}|s_g, D_{g2} = 1] = \mathbb{E}[\bar{e}^k_{g2} - \bar{e}^k_{g1}|s_g, D_{g2} = 0] \tag{13}$$

Assumption 4′ is a common trend assumption conditional on the share of eligible units. Intuitively, we form pairs of treated and control groups with the same share of eligibles. For each pair, we assume that, in the absence of the treatment, the average change in the aggregate outcome among eligible units in treated groups would have been the same as the average change in the aggregate outcome among eligible units in control groups. We make the same assumption regarding the average change in aggregate outcome among non-eligible units. Note that the value of the conditional trend may be different for the eligible and the non-eligible sub-populations.

3 Identification

In this section, we show how one can recover the direct and endogenous peer effects in the two settings we have described.

3.1 Intuition

Figure 1 illustrates how the initial shock (the treatment received by the eligibles) propagates to the whole group through social interactions.

δ_0 is the *direct effect* of the treatment, loosely defined as the first step in the causal chain (the effect of the treatment on the eligibles before any social interaction takes place).

$\delta_0\theta_0^b$ is the *indirect effect* of the treatment, loosely defined as the second step in the causal chain (the response of the non-eligibles to changes in the eligibles' outcomes before any other social interaction takes place).

$M^E(s)$ and $M^{NE}(s)$ are the *population multipliers* for the eligibles and non-eligibles, loosely defined as the third step in the causal chain (the propagation of the initial shock through social interactions).

With these definitions, the direct effect and the indirect effect are independent of population structure, while the intensity of the population multipliers depends on the population structure. Indeed, s influences M in two ways:

- Through the initial shock: the strength of the shock is proportional to the share of eligibles
- Through the propagation: the shock is amplified by social interactions in a non-linear way that depends on the relative magnitudes of θ_0^b and θ_0^w. Intuitively, if $\theta_0^w > \theta_0^b$ (within peer effects are more intense than between peer effects), the amplification is stronger in homogeneous groups (low s and high s) than in mixed groups ($s \approx \frac{1}{2}$). Conversely, if $\theta_0^w < \theta_0^b$, the amplification is weaker is homogeneous groups than in mixed groups.

As formally shown in the next section, the treatment effects and the population multipliers for the eligibles and the non-eligibles can be written as follows:

$$\tau^E(s) = \delta_0 M^E(s) \text{ and } \tau^{NE}(s) = \delta_0 \theta_0^b M^{NE}(s)$$

and

$$M^E(s) = 1 + sP^E(s) \text{ and } M^{NE}(s) = sP^{NE}(s)$$

Where $P^E(s)$ and $P^{NE}(s)$ capture the strength of the propagation in the eligible and non-eligible sub-populations. These functions have a U-shape when $\theta_0^w > \theta_0^b$ and an inverted-U shape when $\theta_0^w < \theta_0^b$. They are different in the eligible and non-eligible sub-populations because their situations are not symmetric: e.g. a low s implies a lot of "between" interactions for eligibles and a lot of "within" interactions for non-eligibles. By contrast, in the homogeneous case where $\theta_0^w = \theta_0^b = \theta$, these functions are identical for the eligibles and the non-eligibles, and they do not depend on s: we have $P^E = P^{NE} = \frac{1}{1-\theta}$. As a consequence, the population multiplier is linear in s.
The shapes of $\tau^E(s)$ and $\tau^{NE}(s)$ are informative about $\delta_0, \theta_0^b, \theta_0^w$. As an illustration, Figure 2 plots $\tau^E(s)$ and $\tau^{NE}(s)$ as a function of s in the case when $\theta_0^w > \theta_0^b$ and in the case when $\theta_0^w < \theta_0^b$. We can get an intuition of the identification by looking at the limits:

$$\underset{s \to 0}{\tau^E(s) \to \delta_0} \qquad \underset{s \to 1}{\tau^E(s) \to \frac{\delta_0}{1-\theta_0^w}} \qquad \underset{s \to 1}{\tau^{NE}(s) \to \frac{\delta_0 \theta_0^b}{1-\theta_0^w}}$$

First, looking at eligibles in groups with a very low fraction of eligibles is informative about the direct effect. Second, looking at eligibles in groups with a very high fraction of eligibles is informative about "within" peer effects. Third, looking at non-eligibles in groups with a very high fraction of eligibles is informative about "between" peer effects.
In practice, we observe few eligibles in groups with low s and few non-eligibles in groups with high s. That is why our procedure exploits the whole distribution of s, and not only the limits. The next section derives the formulas for $\tau^E(s)$ and $\tau^{NE}(s)$ and discusses more formally the identification in the case of randomized experiments and in the case of natural experiments.

3.2 Randomized Experiment

This subsection presents how the treatment effects and population multipliers can be recovered in the randomized experiment setting described in section 2.1.

Proposition 1. *Provided Assumptions 1, 2 and 3 hold and all the mentioned conditional expectations are well-defined, we have*

$$\begin{aligned}\tau^E(s) &= \mathbb{E}\left[\bar{y}_g^E|s_g = s, D_g = 1\right] - \mathbb{E}\left[\bar{y}_g^E|s_g = s, D_g = 0\right] \\ &= \mathbb{E}\left[\frac{D_g - \Pr(D_g = 1)}{\Pr(D_g = 1)(1 - \Pr(D_g = 1))}.\bar{y}_g^E\Big|s_g = s\right] \quad (14) \\ &= \delta_0\left(1 + s.\frac{\theta_0^w - (1-s)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}{1 - \theta_0^w + s(1-s)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\right) \quad (15)\end{aligned}$$

and

$$\begin{aligned}\tau^{NE}(s) &= \mathbb{E}\left[\bar{y}_g^{NE}|s_g = s, T_g = 1\right] - \mathbb{E}\left[\bar{y}_g^{NE}|s_g = s, D_g = 0\right] \\ &= \mathbb{E}\left[\frac{D_g - \Pr(D_g = 1)}{\Pr(D_g = 1)(1 - \Pr(D_g = 1))}.\bar{y}_g^{NE}\Big|s_g = s\right] \quad (16) \\ &= \frac{\delta_0\theta_0^b s}{1 - \theta_0^w + s(1-s)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \quad (17)\end{aligned}$$

Hence,

$$\begin{aligned}P^E(s) &:= \frac{\theta_0^w - (1-s)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}{1 - \theta_0^w + s(1-s)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \quad (18) \\ P^{NE}(s) &:= \frac{1}{1 - \theta_0^w + s(1-s)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \quad (19)\end{aligned}$$

Proof. See Appendix □

Proposition 1 shows that the treatment effect and the functional form of the population multiplier for the eligible (respectively non-eligible) population can be recovered by looking at the difference of the averages of aggregated outcomes within treated groups and within non-treated groups that have exactly the same share s of eligible units. This difference of averages between treated and untreated groups can be expressed as a single average of a weighted outcome, as shown by equations (14) and (16), whose weights are based a transformation process à la Abadie (2005). This alternative formulation leads to some natural conditional moment conditions, as expressed by the following corollary.

Corollary 1 (Conditional Moment Conditions - Randomized Experiment)**.** *Let I be an interval on $\mathbb{R}\backslash\{0\}$ and $\lambda_0 = (\delta_0, \theta_0^w, \theta_0^b) \in \Theta := I \times (-1,1) \times (-1,1)$, be the true value of the parameters. For any $s \in (0,1)$, provided Assumptions 1, 2 and 3 hold and all the mentioned conditional moments are well-defined,*

$$\mathbb{E}\left[u^R(\bar{y}_g^E, \bar{y}_g^{NE}, D_g, s_g; \lambda_0)|s_g = s\right] = \begin{pmatrix} \mathbb{E}\left[u^{R,E}(\bar{y}_g^E, D_g, s_g; \lambda_0)|s_g = s\right] \\ \mathbb{E}\left[u^{R,NE}(\bar{y}_g^{NE}, D_g, s_g; \lambda_0)|s_g = s\right]\end{pmatrix} = 0 \quad (20)$$

where for all $\lambda \in \Theta$,

$$\begin{aligned}u^{R,E}(\bar{y}_g^E, D_g, s_g; \lambda) &= \rho^R(D_g)\bar{y}_g^E - \frac{\delta\left(1 - \theta^w(1 - s_g)\right)}{1 - \theta^w + s_g(1-s_g)\left[(\theta^w)^2 - (\theta^b)^2\right]} \\ u^{R,NE}(\bar{y}_g^{NE}, D_g, s_g; \lambda) &= \rho^R(D_g)\bar{y}_g^{NE} - \frac{\delta\theta^b s_g}{1 - \theta^w + s_g(1-s_g)\left[(\theta^w)^2 - (\theta^b)^2\right]}\end{aligned}$$

and $\rho^R(D_g) = \dfrac{D_g - P(D_g = 1)}{P(D_g = 1)(1 - P(D_g = 1))}$

Proof. It is an immediate consequence of Proposition 4 since the right-hand sides of equations (15) and (17) are functions of s. □

3.3 Natural Experiment

In this subsection, we show that similar results can be obtained in the natural experiment setting we have described in Section 2.2. In the following, for all $k \in \{E, NE\}$, we use the conventional notation for first differenced variables:

$$\Delta\bar{y}_g^k = \bar{y}_{g2}^k - \bar{y}_{g1}^k$$

Proposition 2. *Suppose Assumptions 1′, 2′, 3′, and 4′ hold and that all the mentioned conditional expectations are well-defined. Then, we have*

$$\begin{aligned}
\tau^E(s) &= \mathbb{E}\left[\Delta\bar{y}_g^E|s_g = s, D_g = 1\right] - \mathbb{E}\left[\Delta\bar{y}_g^E|s_g = s, D_g = 0\right] \\
&= \mathbb{E}\left[\frac{D_{g2} - \Pr(D_{g2} = 1|s_g = s)}{\Pr(D_g = 1|s_g = s)(1 - \Pr(D_g = 1|s_g = s))}.\Delta\bar{y}_g^E\middle|s_g = s\right] && (21) \\
&= \delta_0\left(1 + s.\frac{\theta_0^w - (1-s)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}{1 - \theta_0^w + s(1-s)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\right) && (22)
\end{aligned}$$

and

$$\begin{aligned}
\tau^{NE}(s) &= \mathbb{E}\left[\Delta\bar{y}_g^{NE}|s_g = s, T_g = 1\right] - \mathbb{E}\left[\Delta\bar{y}_g^{NE}|s_g = s, D_g = 0\right] \\
&= \mathbb{E}\left[\frac{D_{g2} - \Pr(D_{g2} = 1|s_g = s)}{\Pr(D_{g2} = 1|s_g = s)(1 - \Pr(D_{g2} = 1|s_g = s))}.\Delta\bar{y}_g^{NE}\middle|s_g = s\right] && (23) \\
&= \frac{\delta_0\theta_0^b s}{1 - \theta_0^w + s(1-s)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} && (24)
\end{aligned}$$

Proposition 2 is the counterpart of Proposition 1 in the natural experiment setting. Note the two main differences. First, the outcomes of interest here are the first differenced aggregated outcomes for the eligible and non-eligible subpopulations. Second, the weighting function is now a function of both the treatment assignment variable and the share of eligible units in the group. Indeed, in this setting, treatment assignment is allowed to be correlated with the share of eligible units. Similarly, immediate conditional moment conditions arise, as stated by the following corollary.

Corollary 2 (Conditional Moment Conditions - Natural Experiment)**.** *Let I be an interval on $\mathbb{R}\backslash\{0\}$. Let $\lambda_0 = (\delta_0, \theta_0^w, \theta_0^b)$ be the true value of the parameters with $\lambda_0 \in \Theta := I \times (-1,1) \times (-1,1)$. For any $s \in (0,1)$, provided Assumptions 1′, 2′, 3′ and 4′ hold and all the mentioned conditional moments are well-defined,*

$$\mathbb{E}\left[u^{DiD}(\Delta\bar{y}_g^E, \Delta\bar{y}_g^{NE}, D_{g2}, s_g; \lambda_0)|s_g = s\right] = \begin{pmatrix} \mathbb{E}\left[u^{DiD,E}(\Delta\bar{y}_g^E, D_{g2}, s_g; \lambda_0)|s_g = s\right] \\ \mathbb{E}\left[u^{DiD,NE}(\Delta\bar{y}_g^{NE}, D_{g2}, s_g; \lambda_0)|s_g = s\right] \end{pmatrix} = 0 \quad (25)$$

where for all $\lambda \in \Theta$,

$$u^{DiD,E}(\Delta \overline{y}_g^E, D_{g2}, s_g; \lambda) = \rho^{DiD}(D_{g2}, s_g)\Delta \overline{y}_g^E - \frac{\delta\left(1-\theta^w(1-s_g)\right)}{1-\theta^w + s_g(1-s_g)\left[(\theta^w)^2 - (\theta^b)^2\right]}$$

$$u^{DiD,NE}(\Delta \overline{y}_g^{NE}, D_{g2}, s_g; \lambda) = \rho^{DiD}(D_{g2}, s_g)\Delta \overline{y}_g^{NE} - \frac{\delta \theta^b s_g}{1-\theta^w + s_g(1-s_g)\left[(\theta^w)^2 - (\theta^b)^2\right]}$$

and $\rho^{DiD}(D_{g2}, s_g) = \dfrac{D_{g2} - P(D_{g2}=1|s_g)}{P(D_{g2}=1|s_g)(1-P(D_{g2}=1|s_g))}$

Proof. Immediate consequence from Proposition 2 □

3.4 Sufficient Conditions for Identification

The following proposition provides some sufficient conditions to ensure the identification of λ_0, the vector of the true parameters, quantifying the direct treatment effect and the within and between endogenous effects.

Proposition 3 (Sufficient Conditions for Identification of λ_0)**.** *Provided* $\delta_0 \neq 0$ *and* $\theta_0^b \neq 0$*, if Assumptions 1, 2 and 3 hold OR if Assumptions 1′, 2′, 3′ and 4′ hold and if one of the two following conditions is satisfied*

1. *there exist at least 2 shares* $s_1, s_2 \in (0,1)$*,* $s_1 \neq s_2$ *that have positive probability mass and for which both* τ^{NE} *and* τ^E *are well-defined*
2. *there is a continuum of shares* $I_s \subseteq (0,1]$ *such that* $P(s \in I_s) > 0$ *and both* $\tau^{NE}(s)$ *and* $\tau^E(s)$ *are well-defined, for any* $s \in I_s$

then λ_0 *is the unique vector defined on* Θ *that satisfies equation* (20)

Proof. See Appendix □

3.5 Extension - Including Covariates

One may want to include covariates in the model to either improve precision or the plausibility of the common trends assumption in the case of a natural experiment setting. In Appendix B.1, we provide alternative moment conditions that are based on a new conditional common trends assumption. Intuitively, we compare treated and control groups that have the same share of eligible units and that experience the same evolution of their observed covariates from period 1 to 2. Then, we assume that, in the absence of the treatment, the average change in the aggregate outcome among eligible units in treated groups would have been the same as the average change in the aggregate outcomes among eligible units in control groups. We make the same assumption regarding the change in group average outcome among non-eligible units.

4 Estimation

In this section, we show how the vector of true parameters $\lambda_0 := (\delta_0, \theta_0^w, \theta_0^b)$ can be estimated (in a scenario without covariates). Both the randomized and natural experiment settings lead to conditional moment restrictions of the form

$$\mathbb{E}\left[u(Z_g;\lambda_0)|s_g\right] = 0$$

where $u(Z_g;\lambda)$ is a 2-dimensional vector of known functions of the i.i.d random vector of observed variables Z_g and $\lambda \in \Theta$. Let f be a 4×2 matrix such that, for any $s \in (0,1]$

$$f(s) = \begin{pmatrix} s & 0 \\ 0 & s \\ s^2 & 0 \\ 0 & s^2 \end{pmatrix}$$

We consider the class of Generalized Method of Moments (GMM) estimators of the form

$$\hat{\lambda} := \arg\min_{\Theta} \, [\overline{m}(\lambda)]' W_G \overline{m}(\lambda) \tag{26}$$

where $\overline{m}(\lambda) = \dfrac{1}{G}\sum_{g=1}^{G} f(s_g)\hat{u}(Z_g;\lambda)$ for some 2-dimensional vector $\hat{u}$ of known functions of Z_g and λ that depend on the design and W_G is a non-negative definite, symmetric matrix such that $W_G \xrightarrow{p} W_0$.

4.1 Randomized Experiment

In the randomized experiment setting, since $P(D_g = 1)$ is supposed to be known,

$$\hat{u}(Z_g;\lambda) = u^R(Z_g;\lambda) = \begin{pmatrix} \rho^R(D_g)\overline{y}_g^E - \dfrac{\delta\left(1-\theta^w(1-s_g)\right)}{1-\theta^w + s_g(1-s_g)\left[(\theta^w)^2 - (\theta^b)^2\right]} \\ \rho^R(D_g)\overline{y}_g^{NE} - \dfrac{\delta\theta^b s_g}{1-\theta^w + s_g(1-s_g)\left[(\theta^w)^2 - (\theta^b)^2\right]} \end{pmatrix} \tag{27}$$

Then, we have the following result

Proposition 4 (GMM Estimator in the Randomized Setting)**.** *Let $(Z_g)_{g=1,\ldots,G}$ be an i.i.d sample of G groups, where $Z_g = (\overline{y}_g^E, \overline{y}_g^{NE}, D_g, s_g)$. Supposing that*

1. *Assumptions 1, 2 and 3 hold and identification of λ_0 is ensured*
2. *$\Theta := I_1 \times [-K,K] \times [-K,K]$ where I is a compact subset of $\mathbb{R}$ with $0 \notin I$ and $K \in (0,1)$*
3. *$E[|\overline{y}_g^E|^2] < +\infty$ and $E[|\overline{y}_g^{NE}|^2] < +\infty$*

Then

$$\hat{\lambda}^* = \begin{pmatrix} \delta^* \\ \theta^{w*} \\ \theta^{b*} \end{pmatrix} := \arg\min_{\Theta} \, [\overline{m}(\lambda)]' \hat{V}_G^{-1} \overline{m}(\lambda) \tag{28}$$

where $\hat{V}_G \xrightarrow{p} V_0 := \mathbb{E}\left[f(s_g)u(Z_g;\lambda_0)u(Z_g;\lambda_0)'f(s_g)'\right]$ is a consistent estimator of λ_0, whose asymptotic distribution is

$$\sqrt{G}(\hat{\lambda}^* - \lambda_0) \xrightarrow{d} \mathcal{N}\left(0, \left[M_0' V_0^{-1} M_0\right]^{-1}\right) \tag{29}$$

with $\lambda_0 := (\delta_0, \theta_0^w, \theta_0^b)$ and $M_0 := \mathbb{E}\left[f(s_g)\dfrac{\partial u^R(Z_g;\lambda_0)}{\partial \lambda'}\right]$

Proof. See Appendix □

Since the estimator reaches the parametric rate of convergence, a sample of a hundred groups can already provide quite accurate estimates of the direct effect of the treatment and of the two endogenous peer effects.

4.2 Natural Experiment without Covariates

In the natural experiment setting, we consider

$$\hat{u}(Z_g;\lambda) = \begin{pmatrix} \hat{\rho}^{DiD}(D_{g2},s_g)\Delta\overline{y}_g^E - \dfrac{\delta\left(1-\theta^w(1-s_g)\right)}{1-\theta^w+s_g(1-s_g)\left[(\theta^w)^2-(\theta^b)^2\right]} \\ \hat{\rho}^{DiD}(D_{g2},s_g)\Delta\overline{y}_g^{NE} - \dfrac{\delta\theta^b s_g}{1-\theta^w+s_g(1-s_g)\left[(\theta^w)^2-(\theta^b)^2\right]} \end{pmatrix}$$

with

$$\hat{\rho}^{DiD}(D_{g2},s_g) = \frac{D_{g2}-\hat{P}(D_{g2}=1|s_g)}{\hat{P}(D_{g2}=1|s_g)(1-\hat{P}(D_{g2}=1|s_g))}$$

and

$$\hat{P}(D_{g2}=1|s_g) = \frac{\sum_{j=1}^{G} K_{h_G}(s_g-s_j)D_{j2}}{\sum_{j=1}^{G} K_{h_G}(s_g-s_j)}$$

where, for any $s\in(0,1]$, $K_{h_G}(s) = \dfrac{1}{h_G}\sum_{j=1}^{G} K\left(\dfrac{s}{h_G}\right)$ with K, a kernel of order (at least) 1 and h_G a bandwidth sequence that depends on the total number of groups in the sample. The main difference compared to the previous situation is that, now, $P(D_g=1|s_g)$ has to be nonparametrically estimated first. This first stage estimation is going to change the asymptotic variance of the GMM estimator. However, based on results from Newey and McFadden (1994) (section 8.3), we conjecture the following inference results. Simulations presented in the following sections tend to support this conjecture.

Conjecture 1 (GMM Estimator in the Natural Experiment Setting)**.** *Let $(Z_g)_{g=1,\dots,G}$ be an i.i.d sample of G groups, where $Z_g=(\Delta\overline{y}_g^E,\Delta\overline{y}_g^{NE},D_g,s_g)$. Supposing that*

- *Assumptions 1′, 2′, 3′ and 4′ hold and identification λ_0 is ensured*
- *$\Theta := I_1\times[-K,K]\times[-K,K]$ where I is a compact subset of $\mathbb{R}$ with $0\notin I$ and $K\in(0,1)$*
- *$E[|\overline{y}_g^E|^2]<+\infty$ and $E[|\overline{y}_g^{NE}|^2]<+\infty$*
- *The bandwidth for the nonparametric first stage is such that $h_G\propto G^{-\frac{1}{3}}$*

Then

$$\hat{\lambda}^* := \arg\min_{\Theta}\,[\overline{m}(\lambda)]'\hat{V}_G^{-1}\overline{m}(\lambda) \tag{30}$$

where $\hat{V}_G \xrightarrow{p} V_0 := \mathbb{E}\left[f(s_g)u^{DiD}(Z_g;\lambda_0)u^{DiD}(Z_g;\lambda_0)'f(s_g)\right]$ is a consistent estimator of λ_0 and it is asymptotically normal with convergence rate $G^{-\frac{1}{2}}$.

5 Simulations

We illustrate our estimation strategy in both the randomized and natural experiment settings with simulations. For each simulation, a sample of 500 i.i.d groups are generated. Let

$$\lambda_0 = \begin{pmatrix}\delta_0\\ \theta_0^w\\ \theta_0^b\end{pmatrix} = \begin{pmatrix}5\\ 0.5\\ 0.2\end{pmatrix}$$

5.1 Randomized Experiment

For the randomized setting, we consider the following DGP. For each group g, we observe a vector $Z_g = (\overline{y}_g^E, \overline{y}_g^{NE}, s_g, D_g)$ such that

$$\begin{cases} \overline{y}_g^E &= (\alpha_g^E + z_g^E \beta^E) + \theta_0^w s_g \overline{y}_g^E + \theta_0^b (1-s_g) \overline{y}_g^{NE} + \delta_0 D_g \\ \overline{y}_g^{NE} &= (\alpha_g^{NE} + z_g^{NE} \beta^{NE}) + \theta_0^b s_g \overline{y}_g^E + \theta_0^w (1-s_g) \overline{y}_g^{NE} \end{cases}$$

with

$$\begin{aligned} \begin{pmatrix} \beta^E \\ \beta^{NE} \end{pmatrix} &= \begin{pmatrix} -1 \\ -2 \end{pmatrix} \\ s_g &\sim \mathcal{U}_{(0,1]} \\ D_g &\sim \mathcal{B}(1/2) \\ \begin{pmatrix} \alpha_g^{NE} \\ \alpha_g^E \\ z_g^{NE} \\ z_g^E \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 8 \\ 4 \\ -3 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 & 1 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 0 & 0 & 4 & -4.2 \\ 0 & 0 & -4.2 & 9 \end{pmatrix} \right) \end{aligned}$$

This DGP is consistent with Assumption 1 with

$$\begin{aligned} \overline{e}_g^E &= \alpha_g^E + z_g^E \beta^E \\ \overline{e}_g^{NE} &= \alpha_g^{NE} + z_g^{NE} \beta^{NE} \end{aligned}$$

Moreover, $D_g \perp\!\!\!\perp (s_g, z_g^E, z_g^{NE}, \alpha_g^E, \alpha_g^{NE})$, so Assumption 2 is also satisfied. Assumption 3 is also satisfied with probability one. For each simulation, the estimator presented in Proposition 4 is computed. Figure 3 plots the distribution of $\hat{\lambda}^*$ that is obtained from 500 simulations. The dashed red lines correspond to the true value of the coefficients. As expected, the GMM estimator is consistent and is asymptotically normal.

5.2 Natural Experiment

For the natural experiment setting, the following DGP has been considered. For each group g, we observe a vector $Z_g = (\Delta\overline{y}_g^E, \Delta\overline{y}_g^{NE}, s_g, D_g)$ such that

$$\begin{cases} \overline{y}_{g1}^E &= (\alpha_g^E + \varepsilon_{g1}^E) + \theta_0^w s_g \overline{y}_{g1}^E + \theta_0^b (1-s_g) \overline{y}_{g1}^{NE} \\ \overline{y}_{g2}^E &= (\alpha_g^E + \lambda^E s_g + \varepsilon_{g2}^E) + \theta_0^w s_g \overline{y}_{g2}^E + \theta_0^b (1-s_g) \overline{y}_{g2}^{NE} + \delta_0 D_{g2} \\ \\ \overline{y}_{g1}^{NE} &= (\alpha_g^{NE} + \varepsilon_{g1}^{NE}) + \theta_0^b s_g \overline{y}_{g1}^E + \theta_0^w (1-s_g) \overline{y}_{g1}^{NE} \\ \overline{y}_{g2}^{NE} &= (\alpha_g^{NE} + \lambda^{NE} s_g^2 + \varepsilon_{g2}^{NE}) + \theta_0^b s_g \overline{y}_{g2}^E + \theta_0^w (1-s_g) \overline{y}_{g2}^{NE} \end{cases}$$

with

$$\begin{pmatrix} \lambda^E \\ \lambda^{NE} \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$$

$$\varepsilon_g = \begin{pmatrix} \varepsilon_{g1}^E \\ \varepsilon_{g2}^E \\ \varepsilon_{g1}^{NE} \\ \varepsilon_{g2}^{NE} \end{pmatrix} \sim \mathcal{N}\left(0, \Sigma_\varepsilon\right)$$

$$s_g \sim F^{-1}_{\mathcal{U}_{(0,1]}}(V_g^1)$$

$$D_{g2} \sim F^{-1}_{\mathcal{B}(1/2)}(V_g^2)$$

$$\begin{pmatrix} \alpha_g^E \\ \alpha_g^{NE} \\ V_g^1 \\ V_g^2 \end{pmatrix} \sim \mathcal{N}\left(0, \Sigma_{\alpha,V}\right)$$

where Σ_ε and $\Sigma_{\alpha,V}$ are 4×4 positive definite symmetric matrices whose eigenvalues are respectively 2 and 1 (resp. 1, 4 and 5). By construction, Assumptions $1'$ and $2'$ are satisfied with

$$\begin{aligned} \bar{e}_{g1}^E &= \alpha_g^E + \varepsilon_{g1}^E \\ \bar{e}_{g1}^{NE} &= \alpha_g^{NE} + \varepsilon_{g1}^{NE} \\ \bar{e}_{g2}^E &= \alpha_g^E + \lambda^E s_g + \varepsilon_{g2}^E \\ \bar{e}_{g2}^{NE} &= \alpha_g^{NE} + \lambda^{NE} s_g^2 + \varepsilon_{g2}^{NE} \end{aligned}$$

Note that

$$\mathbb{E}\left[\bar{e}_{g2}^E - \bar{e}_{g1}^E \middle| s_g, D_{g2} = 1\right] = \mathbb{E}\left[\bar{e}_{g2}^E - \bar{e}_{g1}^E \middle| s_g, D_{g2} = 0\right] = \lambda^E s_g$$

$$\mathbb{E}\left[\bar{e}_{g2}^{NE} - \bar{e}_{g1}^{NE} \middle| s_g, D_{g2} = 1\right] = \mathbb{E}\left[\bar{e}_{g2}^{NE} - \bar{e}_{g1}^{NE} \middle| s_g, D_{g2} = 0\right] = \lambda^{NE} s_g^2$$

so Assumption $4'$ is also satisfied: conditional on s_g, treated and control groups have the same time trend, and this trend is potentially different for eligibles and non-eligibles. Finally, Assumption $3'$ holds with probability 1. For each simulation, we compute the estimator presented in Conjecture 1. For the estimator's first step, the Nadaraya-Watson estimator of the conditional expectation $E(D_{g2}|s_g)$ was computed using a gaussian kernel and $h_G = 500^{-\frac{1}{3}}$. Figure 4 plots the distribution of $\hat{\lambda}^*$ that is obtained from 500 simulations. The dashed red lines correspond to the true value of the coefficients. As conjectured, the estimator seems to be consistent and asymptotically normal.

5.3 Natural Experiment with Common Trends Conditional on Covariates

In Appendix B.2, we provide some preliminary results on ways of estimating λ_0 when Assumption $4'$ fails but Assumption $3''$ is satisfied.

6 Application: Conditional cash transfers in Mexico (Progresa)

Progresa is a conditional cash transfer (CCT) program introduced in Mexico in 1997 and aimed at developing the human capital of poor households. The program conditioned cash payments on children regularly attending school and health checkups at clinics. The program was means tested, with a two-step targeting procedure. First, the poorest villages were identified using socio-economic characteristics in census data. Second, within a village, the poorest households were identified using a specific survey collecting data on assets and demographic composition. Only the poorest households were eligible for the program.

Progresa has been widely studied for two reasons. First, it was one of the earliest CCT implemented at a large scale; since then, similar programs spread around the world. Second, the implementation of the program was experimental during the first 18 months. In the spring 1998, among the 506 poorest villages, 320 were randomly chosen to participate in the program and eligible households started receiving transfers. The program was extended to the 186 control villages at the end of 1999, and then gradually to a larger set of villages. The randomization was exploited by several studies to estimate the short-term impact on education (see Parker and Todd (2017) for a review). Most studies focus on eligible households. A few others also estimate the treatment effect on the non-eligibles, taking advantage of the fact that post-program evaluation surveys interviewed all households, including non-eligible households, in treated and control villages.

Of particular interest for us is a study by Lalive and Cattaneo (2009) who use Progresa as a partial-population design to study peer effects in school attendance. They define the reference group as all children living in the same village who have reached the same grade level. Using the interaction between treatment and share of eligibles as an instrument for average group attendance in October 1998, they estimate an endogenous peer effect parameter of $\theta = 0.5$ (95% CI=[0;1]). This implies that increasing the average attendance in a child's group by 10 percentage points will raise her likelihood to attend school by 5 percentage points. The population multiplier is equal to $\frac{1}{1-\theta}s = 2s$. Furthermore, they estimate a direct effect of $\delta_0 = 0.03$ (95% CI=[0;0.06]), which represents a 4% increase compared to the control mean of 0.69.

We use the same data and implement our procedure allowing for heterogeneous θ. Figure A3 in Appendix plots the distribution of s, by treatment status. The support for the share of poor ranges from 0.1 to 1. We observe both treated and control groups in all bins, except $(0.25, 0.30]$ and $(0.90, 1]$. In total, we observe 663 groups. Estimates are reported in Table 1 (computation of standard errors in progress). We find a "between" peer effect of $\theta_0^b = 0.6$, close to the homogeneous θ estimated by Lalive and Cattaneo (2009). This makes sense because their instrumental variable strategy exploits the response of non-eligible children to the introduction of Progresa in the peer group. In other words, their identification of θ comes from "between" interactions.

In addition, our methodology gives us an estimate of "within" peer effects: we find that $\theta_0^w = 0.9$. Endogenous peer effects are therefore stronger within social classes than between social classes. Poor children do influence non-poor children, and this is picked up by the instrumental variable strategy of Lalive and Cattaneo (2009). But poor children influence even more other poor children. Conformism is therefore very strong in this context. For example, suppose that one poor child drops out of school in a group of 12 children (the average group size in our sample). In the rest of the group, the likelihood that a child attends school decreases by $\frac{1}{12} \times 0.9 = 7.5$ percentage points among the poor and by $\frac{1}{12} \times 0.6 = 5$ percentage points among the non-poor.[2]. This is the effect *ceteris paribus*, looking at the response of a given child and holding the attendance of other children constant.

These magnitudes can be compared with the direct effect of Progresa. We find that $\delta_0 = 0.02$,

[2] In this example, the total effect is independent of s because two effects cancel out: when s is large, (i) a change in the average attendance of poor children matters more, and (ii) the average attendance of poor children changes less in response to one poor dropping out. Going back to structural equations, the change in y_i^E is equal to $\theta_0^w s \frac{-1}{12s}$ and the change in y_i^{NE} is equal to $\theta_0^b s \frac{-1}{12s}$.

which implies that, in a group with only one poor child, receiving cash transfers would increase her school attendance by 2 percentage points. Combined with the estimate of θ_0^b, we find an indirect effect of $0.02 \times 0.6 = 0.012$. This implies that, in a group with only one poor child and one non-poor child, the school attendance of the non-poor child would increase by 1.2 percentage points in response to the change in the school attendance of the poor child once she receives cash transfers (*ceteris paribus*). We discuss the magnitude of population multipliers below.

Our estimate of δ_0 is lower than the estimate of Lalive and Cattaneo (2009) (although it belongs to their confidence interval). This suggests that, in this context, the homogeneity assumption leads to an attenuation bias. Our explanation is the following: θ is identified through "between" interactions; since $\theta_0^b < \theta_0^w$, the homogeneity assumption leads to underestimating the strength of peer effects, and consequently to underestimating the total effect when s is very high and "within" interactions matter a lot; therefore, the procedure overestimates δ_0 in order to fit the data (recall that $\tau^E(s=1) = \frac{\delta_0}{1-\theta_0^w}$).

In terms of policy evaluation, Figure 5 represents the total effects and the population multipliers as a function of s. Figure (a) plots the estimates of τ^E and τ^{NE}. In groups with a very low share of poor children ($s \to 0$), Progresa raises school attendance by 2 percentage points for the (very rare) poor children and this is not enough to trigger any response by non-poor children. As the fraction of poor increases, the effects on poor and non-poor increase in a convex way. In groups with a very high share of poor children ($s \to 1$), Progresa raises school attendance by 20 percentage points for poor children and by 12 percentage points for the (very rare) non-poor children. The population multipliers are equal to $\frac{1}{1-\theta_0^w} = 10$ when $s = 1$. The dotted line plots the total effect of the policy: $s\tau^E + (1-s)\tau^{NE}$. Magnitudes range from 0 when $s = 0$ to 20 percentage points when $s = 1$. The total effect for the average s in our sample (roughly 62.6%) is equal to 6p.p. for eligible children and 3.5p.p. for non-eligible children. We conclude that the policy effect is strongly magnified by peer effects.

To better understand why the total effects are convex, Figures (b) and (c) plot the population multipliers, $M(s)$, for the non-eligibles and for the eligibles, respectively. Recall that the population multipliers are the products of two components: the first one, s, captures the strength of the initial shock; the second one, $P(s)$, is not linear in s and captures the propagation of the shock. In the case of Progresa, $P(s)$ have a U-shape because $\theta_0^w > \theta_0^b$. The intuition is the following. In mixed groups, when $s \approx 0.5$, there are a lot of "between" interactions. If θ_0^b is low, the propagation rapidly loses momentum: P^E and P^{NE} are low. In homogeneous groups, there are a lot of "within" interactions. If θ_0^w is high, they generate a snowball effect within the majority group, which also propagates to the minority group. We can distinguish two cases: low s and high s. Both cases are symmetric for the non-eligibles: either they respond a lot to a small propagation within eligibles (low s) or they respond little to a large propagation within eligibles (high s). P^{NE} is equally large in both cases. In contrast, there is no symmetry for the eligibles because their own snowball effect is a first order mechanism. P^E is the largest when s is high.

7 Conclusion

We propose a simple methodology to estimate heterogeneous endogenous peer effects using partial population experiments. We discuss the case of randomized experiments and differences-in-differences. The procedure requires the following conditions: (i) there is no direct effect of the treatment of the non-eligibles; (ii) there is sufficient variation in the share of eligibles across groups and common support in the distribution by treatment status; (iii) there is a sufficient number of groups. Intuitively, we match treated and control groups with a similar share of eligibles. For each pair, we estimate the treatment effect on the eligibles and the treatment effect on the non-eligibles. The relationship between the treatment effects and the share of eligibles is informative about the propagation of the initial shock through peer effects.

Our methodology provides an estimate of (i) the endogenous peer effects "between" the sub-populations of eligibles and non-eligibles; (ii) the endogenous peer effects "within" each sub-population; (iii) the direct effect of the policy; (iv) the population multipliers. These estimates are useful in two ways. First, they help us understand how social influence works: who is influenced by whom, and how much. Second, we can compute the policy effect and decompose the total effect into a direct effect and an indirect effect generated by peers. Models with homogeneous peer effects strongly restrict the relationship between the policy effects and the share of eligibles, while our model with heterogeneous peer effects is less restrictive.

The estimation procedure relies on GMM. It is easy to implement with standard statistical software and is not computationally intensive. Therefore, we think that our methodology has the potential to be used broadly by applied economists interested in peer effects and/or policy evaluation.

References

Abadie, Alberto (2005). "Semiparametric difference-in-differences estimators". In: *The review of economic studies* 72(1), pp. 1–19.

Blume, Lawrence E et al. (2011). "Identification of social interactions". In: *Handbook of social economics*. Vol. 1. Elsevier, pp. 853–964.

Bobonis, Gustavo J. and Frederico Finan (2009). "Neighborhood Peer Effects in Secondary School Enrollment Decisions". In: *Review of Economics and Statistics* 91, pp. 695–716.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009). "Identification of peer effects through social networks". In: *Journal of econometrics* 150(1), pp. 41–55.

Brown, Kristine M. and Ron A. Laschever (2012). "When They're Sixty-For: Peer Effects and the Timing of Retirement". In: *American Economic Journal: Applied Economics* 4, pp. 90–115.

Hirano, Keisuke and Jinyong Hahn (2010). "Design of randomized experiments to measure social interaction effects". In: *Economics Letters* 106(1), pp. 51–53.

Hudgens, Michael G and M Elizabeth Halloran (2008). "Toward causal inference with interference". In: *Journal of the American Statistical Association* 103(482), pp. 832–842.

Lalive, Rafael and Matias A Cattaneo (2009). "Social interactions and schooling decisions". In: *The Review of Economics and Statistics* 91(3), pp. 457–477. URL: http://www.jstor.org/stable/25651352.

Manski, Charles F (1993). "Identification of endogenous social effects: The reflection problem". In: *The review of economic studies* 60(3), pp. 531–542.

Moffitt, Robert A et al. (2001). "Policy interventions, low-level equilibria, and social interactions". In: *Social dynamics* 4(45-82), pp. 6–17.

Newey, Whitney K and Daniel McFadden (1994). "Large sample estimation and hypothesis testing". In: *Handbook of econometrics* 4, pp. 2111–2245.

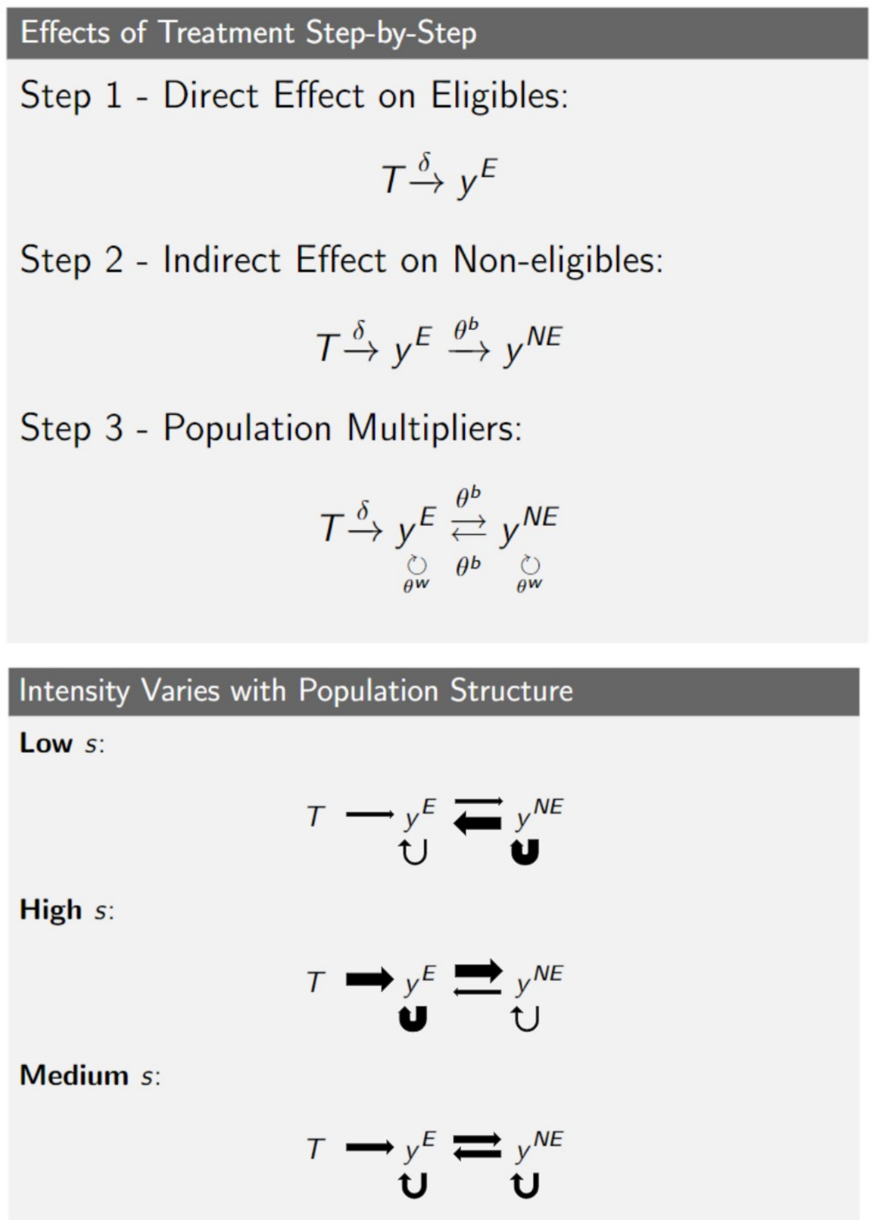
Parker, Susan and Petra Todd (2017). "Conditional Cash Transfers: The Case of Progresa/Oportunidades". In: *Journal of Economic Literature* 55(3), pp. 866–915.

Tchetgen, Eric J Tchetgen and Tyler J VanderWeele (2012). "On causal inference in the presence of interference". In: *Statistical methods in medical research* 21(1), pp. 55–75.

Vazquez-Bare, Gonzalo (2023). "Identification and estimation of spillover effects in randomized experiments". In: *Journal of Econometrics* 237(1), p. 105237.

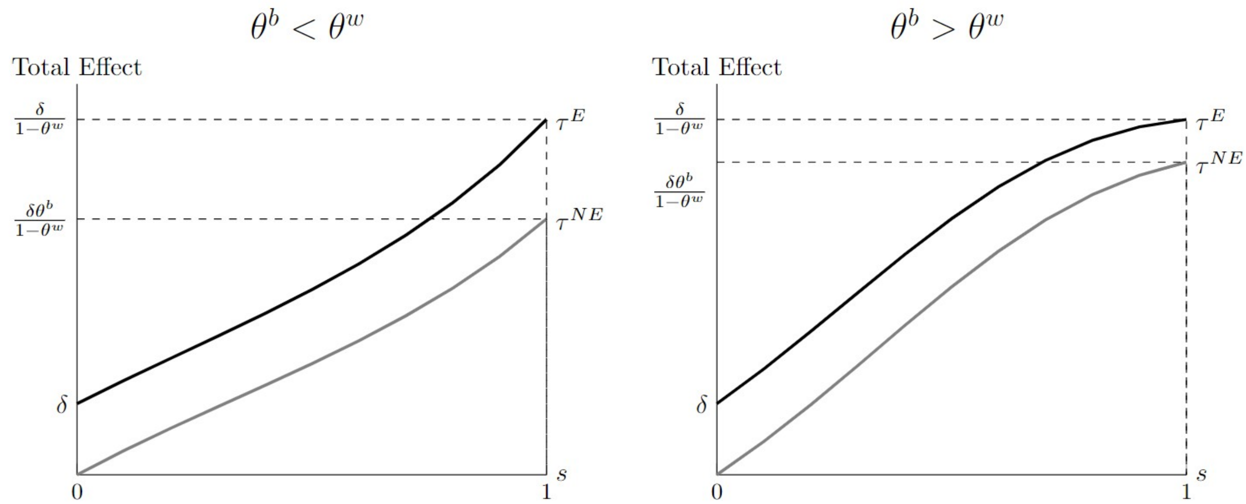
Figures and Tables

Figure 1: Intuition: initial shock and propagation



Note: The top figure illustrates how the initial shock to eligibles propagates through peer effects. First, the eligibles respond to the treatment: y^E changes by δ (direct effect). Second, the non-eligibles respond to the change in y^E: y^{NE} changes by $\theta^b \times \delta$ (indirect effect). Third, these initial changes lead to a cascade process, in which eligibles and non-eligibles keep responding to each other (in proportion to θ^b) and also respond to other members of their own sub-group (in proportion to θ^w).
The bottom figure illustrates how the process intensity varies with the share of eligibles, s. First, the initial shock is proportional to s: the total direct effect on the group average is equal to $\delta \times s$. Second, the propagation depends on s: arrows originating from y^E are proportional to s while arrows originating from y^{NE} are proportional to $(1 - s)$. "Between" and "within" interactions therefore matter more or less depending on s, and for a given s, differently for eligibles and non-eligibles.

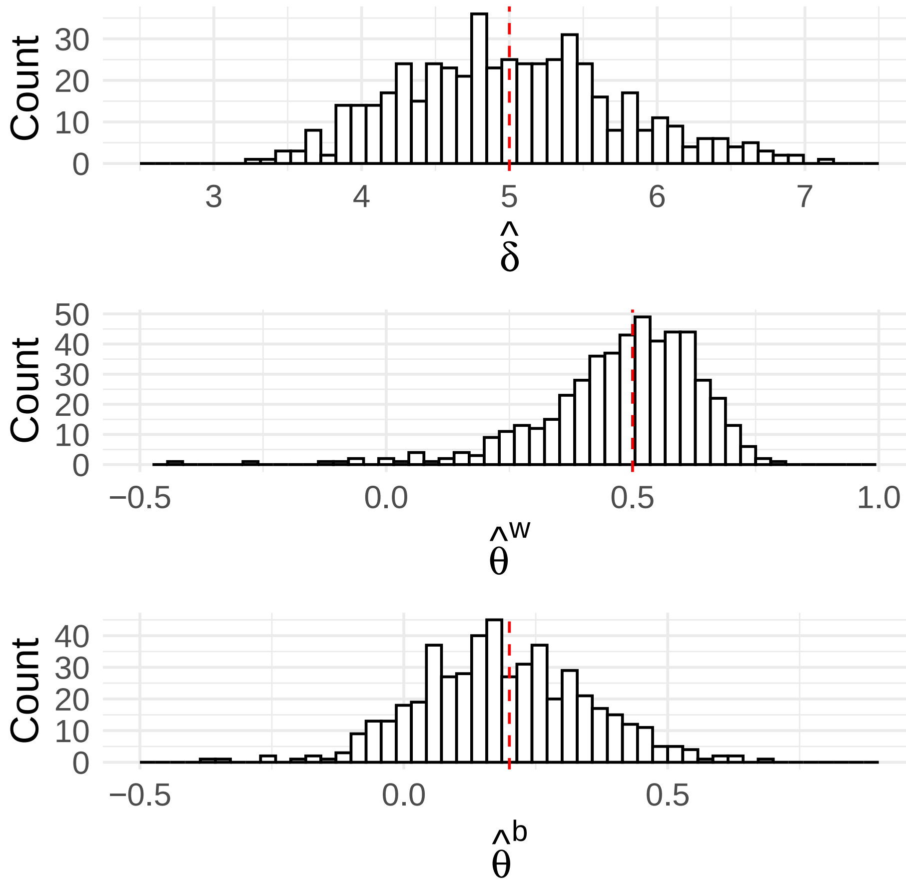
Figure 2: Treatment effects on eligibles and non-eligibles



Note: The figure plots the total effect on the eligibles $\tau^E(s)$, in black, and the total effect on the non-eligibles $\tau^{NE}(s)$, in gray, as a function of s. The graph on the left illustrates the case when $\theta^b < \theta^w$ and the graph on the right illustrates the case when $\theta^b > \theta^w$. As shown in section 3, $\tau^E(s) = \delta M^E(s)$ and $\tau^{NE}(s) = \delta\theta^b M^{NE}(s)$, where:

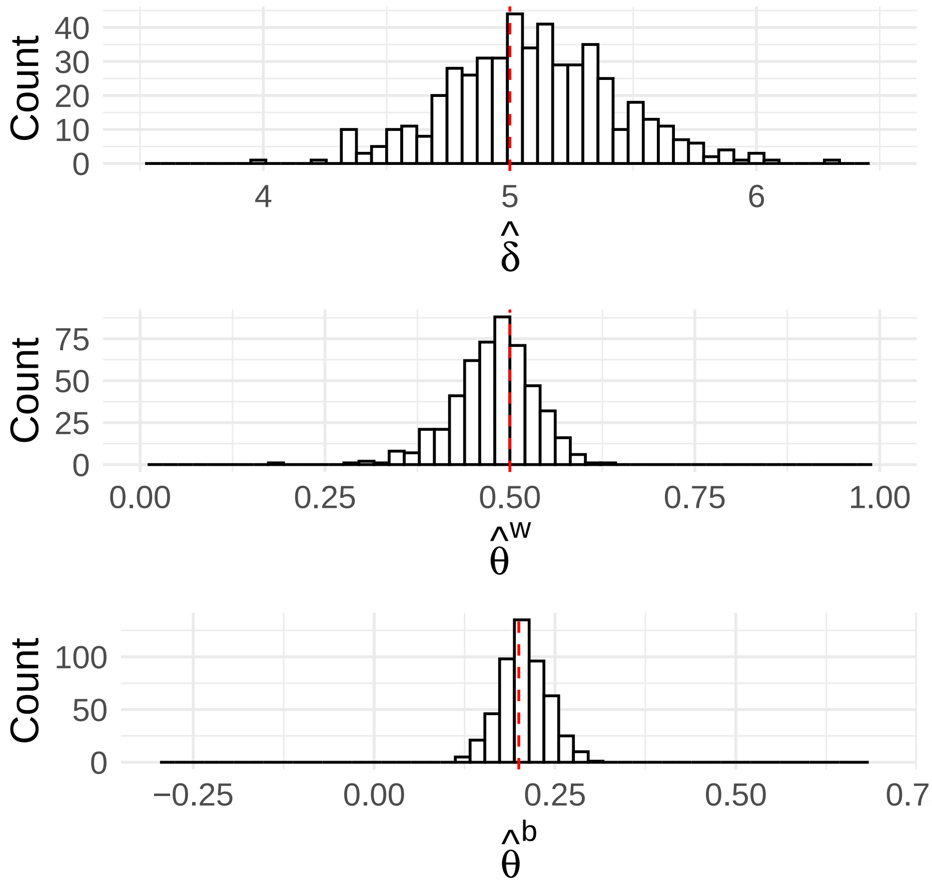
$$M^E(s) = 1 + sP^E(s), M^{NE}(s) = sP^{NE}(s), P^E(s) = \frac{\theta^w - (1-s)\left[(\theta^w)^2 - (\theta^b)^2\right]}{1 - \theta^w + s(1-s)\left[(\theta^w)^2 - (\theta^b)^2\right]}, P^{NE}(s) = \frac{1}{1 - \theta^w + s(1-s)\left[(\theta^w)^2 - (\theta^b)^2\right]}$$

Figure 3: Distribution of $\hat{\lambda}^*$ in the Randomized Setting



Note: The Figure plots the distribution of the GMM estimator λ^* defined in the Estimation section (Randomized Experiment subsection), based on 500 simulations of the DGP described in the Simulations Section (Randomized Experiment subsection). The dashed red lines indicate the true value of the parameter.

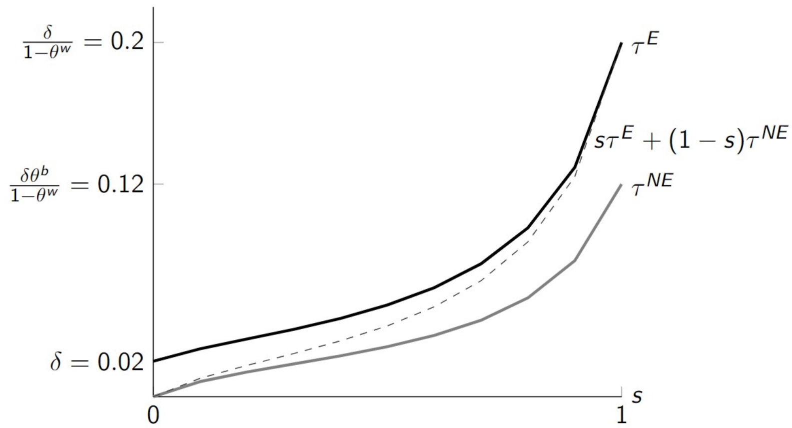
Figure 4: Distribution of $\hat{\lambda}^*$ in the Natural Experiment Setting



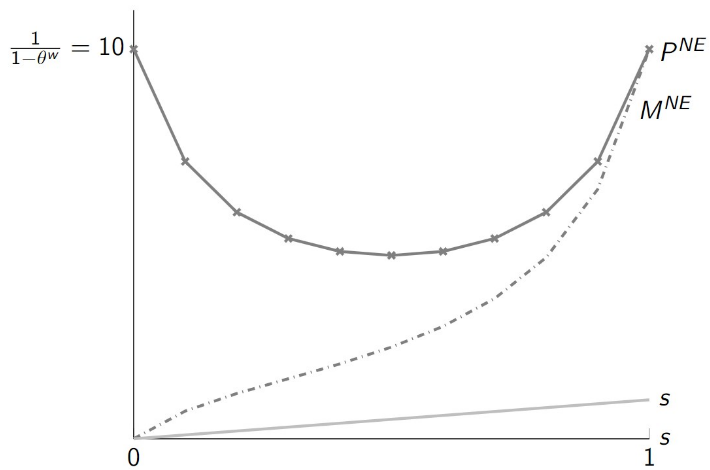
Note: The Figure plots the distribution of the GMM estimator λ^* defined in the Estimation section (Natural Experiment subsection), based on 500 simulations of the DGP described in the Simulations Section (Natural Experiment subsection). The dashed red lines indicate the true value of the parameter.

Figure 5: Results: Progresa

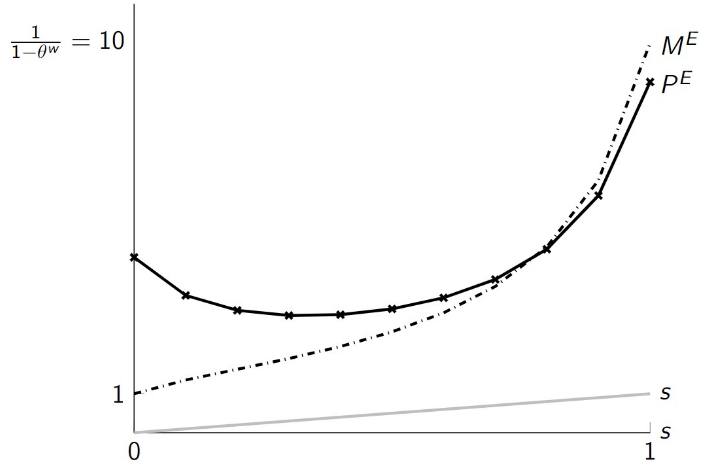
(a) Total effect on eligibles and non-eligibles



(b) Population multipliers for non-eligibles



(c) Population multipliers for eligibles



Note: The figure shows the total effects and population multipliers based on the Progresa estimates reported in Table 1. Graph (a) plots the total effect on the eligibles $\tau^E(s)$, in black, and the total effect on the non-eligibles $\tau^{NE}(s)$, in gray, as a function of s. The dotted line plots the total effect of the policy in the peer group.
Graph (b) (resp. graph (c)) plots the population multiplier $M(s)$, in dashed lines, the propagation function $P(s)$, in solid lines, and the 45-degree line for non-eligibles (resp. eligibles) as a function of s. See formulas in the note below Figure 2.

Table 1: Estimation results: Progresa

θ^b	0.60
θ^w	0.90
δ	0.02
N of groups	663
N of obs.	15,292

Note: The table reports the estimates of our three parameters of interest: the "between" peer effect parameter θ^b, the "within" peer effect parameter θ^w, and the direct effect δ. We implement the procedure described in section 4 using Progresa evaluation data. The computation of standard errors is in progress.

Appendix A Proofs

Appendix A.1 Reduced Forms in the Randomized Experiment Setting

Let's start from the aggregated model, based on Assumption 1

$$\begin{aligned} \bar{y}_g^E &= \theta_0^w s_g \bar{y}_g^E + \theta_0^b (1-s_g)\bar{y}_g^{NE} + \delta_0 D_g + \bar{e}_g^E \\ \bar{y}_g^{NE} &= \theta_0^b s_g \bar{y}_g^E + \theta_0^w (1-s_g)\bar{y}_g^{NE} + \bar{e}_g^{NE} \end{aligned}$$

Plugging-in the expression of $\bar{y}_g^E$ into the one of $\bar{y}_g^{NE}$, we get

$$\bar{y}_g^{NE} = \frac{\theta_0^b s_g}{1-\theta_0^w s_g}\left(\theta_0^b(1-s_g)\bar{y}_g^{NE} + \delta_0 D_g + \bar{e}_g^E\right) + \theta_0^w(1-s_g)\bar{y}_g^{NE} + \bar{e}_g^{NE}$$

Rearranging terms, we get

$$\left[1-\theta_0^w + s_g(1-s_g)((\theta^w)^2 - (\theta^b)^2)\right]\bar{y}_g^{NE} = \theta_0^b s_g(\delta_0 D_g + \bar{e}_g^E) + (1-\theta_0^w s_g)\bar{e}_g^{NE}$$

Hence,

$$\begin{aligned} \bar{y}_g^{NE} = & \frac{1-\theta_0^w s_g}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\bar{e}_g^{NE} \\ & + \frac{\theta_0^b s_g}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\bar{e}_g^{E} \\ & + \frac{\delta_0\theta_0^b s_g}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}D_g \end{aligned}$$

Now, plugging-in the reduced form expression of $\bar{y}_g^{NE}$ into $\bar{y}_g^E$ and developping, we get

$$\begin{aligned} \bar{y}_g^E = & \frac{1-\theta_0^w(1-s_g)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\bar{e}_g^{E} \\ & + \frac{\theta_0^b(1-s_g)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\bar{e}_g^{NE} \\ & + \frac{\delta_0(1-\theta_0^w(1-s_g))}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}D_g \\ = & \frac{\theta_0^b(1-s_g)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\bar{e}_g^{NE} \\ & + \frac{1-\theta_0^w(1-s_g)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\bar{e}_g^{E} \\ & + \delta_0\left(1 + s_g.\frac{\theta_0^w - (1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\right)D_g \end{aligned} \tag{31}$$

Appendix A.2 Proposition 2

Suppose Assumptions 1, 2 and 3 are satisfied. For any $k \in \{E, NE\}$ and any $s \in \mathcal{S}$,

$$\tau^k(s) = \mathbb{E}\left[\bar{y}_g^k | D_g = 1, s_g = s\right] - \mathbb{E}\left[\bar{y}_g^k | D_g = 1, s_g = s\right]$$

This quantity exists as long as Assumption 3 holds and the relevant conditional expectations are well-defined. Let's show first that

$$\tau^k(s) = \mathbb{E}\left[\rho^R(D_g)\bar{y}_g^k | s_g = s\right] \text{ with } \rho^R(D_g) = \frac{D_g - P(D_g = 1)}{P(D_g = 1)(1 - P(D_g = 1))}$$

$$\begin{aligned}
\mathbb{E}\left[\rho^R(D_g)\bar{y}_g^k | s_g = s\right] &= \mathbb{E}\left[\frac{D_g - P(D_g = 1|s_g = s)}{P(D_g = 1|s_g = s)(1 - P(D_g = 1|s_g = s))}\bar{y}_g^k | s_g = s\right] \\
&= P(D_g = 1|s_g = s)\,\mathbb{E}\left[\frac{D_g - P(D_g = 1|s_g = s)}{P(D_g = 1|s_g = s)(1 - P(D_g = 1|s_g = s))}\bar{y}_g^k | s_g = s, D_g = 1\right] \\
&\quad + P(D_g = 0|s_g = s)\,\mathbb{E}\left[\frac{D_g - P(D_g = 1|s_g = s)}{P(D_g = 1|s_g = s)(1 - P(D_g = 1|s_g = s))}\bar{y}_g^k | s_g = s, D_g = 0\right] \\
&= \frac{P(D_g = 1)(1 - P(D_g = 1))}{P(D_g = 1)(1 - P(D_g = 1))}\,\mathbb{E}\left[\bar{y}_g^k | s_g = s, D_g = 1\right] \\
&\quad - \frac{(1 - P(D_g = 1))P(D_g = 1)}{P(D_g = 1)(1 - P(D_g = 1))}\,\mathbb{E}\left[\bar{y}_g^k | s_g = s, D_g = 0\right] \\
&= \mathbb{E}\left[\bar{y}_g^k | s_g = s, D_g = 1\right] - \mathbb{E}\left[\bar{y}_g^k | s_g = s, D_g = 0\right] \\
&= \tau^k(s)
\end{aligned}$$

The second equality is obtained using the law of iterated expectations and Assumption 2, in particular the fact that $D_g \perp\!\!\!\perp s_g$. The third and forth equalities are algebra. Now,

$$\begin{aligned}
\mathbb{E}\left[\bar{y}_g^E | s_g = s, D_g\right] = &\ \frac{\theta_0^b(1-s)}{1 - \theta_0^w + s(1-s)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\,\mathbb{E}\left[\bar{e}_g^{NE} | s_g = s\right] \\
&+ \frac{1 - \theta_0^w(1 - s_{gt})}{1 - \theta_0^w + s_{gt}(1 - s_{gt})\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\,\mathbb{E}\left[\bar{e}_g^E | s_g = s\right] \\
&+ \delta_0\left(1 + s_g.\frac{\theta_0^w - (1 - s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}{1 - \theta_0^w + s_g(1 - s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\right) D_g
\end{aligned}$$

Since, by Assumption 2, $D_g \perp\!\!\!\perp (\bar{e}_g^E, \bar{e}_g^{NE})$. It follows that

$$\tau^E(s) = \delta_0\left(1 + s_g.\frac{\theta_0^w - (1 - s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}{1 - \theta_0^w + s_g(1 - s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}\right)$$

Similarly, one can show that

$$\tau^{NE}(s) = \frac{\delta_0\theta_0^b s_g}{1 - \theta_0^w + s_g(1 - s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}$$

Finally, equations (18) and (19) are immediately identified, based on the definition of the population multipliers.

Appendix A.3 Proposition 2

From Assumption 1′, using the same steps as in Appendix A.1, we get

$$\begin{aligned}\bar{y}^{E}_{gt} = & \frac{\theta^b_0(1-s_{gt})}{1-\theta^w_0+s_{gt}(1-s_{gt})\left[(\theta^w_0)^2-(\theta^b_0)^2\right]}\bar{e}^{NE}_{gt} \\ & + \frac{1-\theta^w_0(1-s_{gt})}{1-\theta^w_0+s_{gt}(1-s_{gt})\left[(\theta^w_0)^2-(\theta^b_0)^2\right]}\bar{e}^{E}_{gt} \\ & + \frac{\delta_0\left(1-\theta^w_0(1-s_{gt})\right)}{1-\theta^w_0+s_{gt}(1-s_{gt})\left[(\theta^w_0)^2-(\theta^b_0)^2\right]}D_{gt}\end{aligned}$$

$$\begin{aligned}\bar{y}^{NE}_{gt} = & \frac{1-\theta^w_0 s_{gt}}{1-\theta^w_0+s_{gt}(1-s_{gt})\left[(\theta^w_0)^2-(\theta^b_0)^2\right]}\bar{e}^{NE}_{gt} \\ & + \frac{\theta^b_0 s_{gt}}{1-\theta^w_0+s_{gt}(1-s_{gt})\left[(\theta^w_0)^2-(\theta^b_0)^2\right]}\bar{e}^{E}_{gt} \\ & + \frac{\delta_0\theta^b_0 s_{gt}}{1-\theta^w_0+s_{gt}(1-s_{gt})\left[(\theta^w_0)^2-(\theta^b_0)^2\right]}D_{gt}\end{aligned}$$

Using Assumptions 2′ and the first part of 3′, we get

$$\begin{aligned}\Delta\bar{y}^{E}_{g} = & \frac{\theta^b_0(1-s_{g})}{1-\theta^w_0+s_{g}(1-s_{g})\left[(\theta^w_0)^2-(\theta^b_0)^2\right]}(\bar{e}^{NE}_{g2}-\bar{e}^{NE}_{g1}) \\ & + \frac{1-\theta^w_0(1-s_{g})}{1-\theta^w_0+s_{g}(1-s_{g})\left[(\theta^w_0)^2-(\theta^b_0)^2\right]}(\bar{e}^{E}_{g2}-\bar{e}^{E}_{g1}) \\ & + \frac{\delta_0\left(1-\theta^w_0(1-s_{g})\right)}{1-\theta^w_0+s_{g}(1-s_{g})\left[(\theta^w_0)^2-(\theta^b_0)^2\right]}D_{g2}\end{aligned}$$

$$\begin{aligned}\Delta\bar{y}^{NE}_{g} = & \frac{1-\theta^w_0 s_{g}}{1-\theta^w_0+s_{g}(1-s_{g})\left[(\theta^w_0)^2-(\theta^b_0)^2\right]}(\bar{e}^{NE}_{g2}-\bar{e}^{NE}_{g1}) \\ & + \frac{\theta^b_0 s_{g}}{1-\theta^w_0+s_{g}(1-s_{g})\left[(\theta^w_0)^2-(\theta^b_0)^2\right]}(\bar{e}^{E}_{g2}-\bar{e}^{E}_{g1}) \\ & + \frac{\delta_0\theta^b_0 s_{g}}{1-\theta^w_0+s_{g}(1-s_{g})\left[(\theta^w_0)^2-(\theta^b_0)^2\right]}D_{g2}\end{aligned}$$

Then, since there are supposed to be well-defined, we can take the conditional expectations of these quantities with respect to (D_{g2}, s_g), we get

$$\begin{aligned}
\mathbb{E}\left[\Delta\bar{y}_g^E \middle| D_{g2}, s_g\right] = \; & \frac{\theta_0^b(1-s_g)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \mathbb{E}\left[\bar{e}_{g2}^{NE} - \bar{e}_{g1}^{NE} \middle| D_{g2}, s_g\right] \\
& + \frac{1-\theta_0^w(1-s_g)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \mathbb{E}\left[\bar{e}_{g2}^{E} - \bar{e}_{g1}^{E} \middle| D_{g2}, s_g\right] \\
& + \frac{\delta_0\left(1-\theta_0^w(1-s_g)\right)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} D_{g2}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\Delta\bar{y}_g^{NE} \middle| D_{g2}, s_g\right] = \; & \frac{1-\theta_0^w s_g}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \mathbb{E}\left[\bar{e}_{g2}^{NE} - \bar{e}_{g1}^{NE} \middle| D_{g2}, s_g\right] \\
& + \frac{\theta_0^b s_g}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \mathbb{E}\left[\bar{e}_{g2}^{E} - \bar{e}_{g1}^{E} \middle| D_{g2}, s_g\right] \\
& + \frac{\delta_0 \theta_0^b s_g}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} D_{g2}
\end{aligned}$$

Now,

$$\mathbb{E}\left[\Delta\bar{y}_g^E \middle| D_{g2}=1, s_g\right] - \mathbb{E}\left[\Delta\bar{y}_g^E \middle| D_{g2}=0, s_g\right] = \frac{\delta_0\left(1-\theta_0^w(1-s_g)\right)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}$$

$$\mathbb{E}\left[\Delta\bar{y}_g^{NE} \middle| D_{g2}=1, s_g\right] - \mathbb{E}\left[\Delta\bar{y}_g^{NE} \middle| D_{g2}=0, s_g\right] = \frac{\delta_0 \theta_0^b s_g}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}$$

using Assumption 4′. Then, using the same steps as in the proof for Proposition 4, one can show that, for any $k \in \{E, NE\}$ and any $s \in \mathcal{S}$,

$$\tau^k(s) = \mathbb{E}\left[\rho^{DiD}(D_{g2}, s_g)\Delta\bar{y}_g^k | s_g = s\right] \text{ with } \rho^{DiD}(D_{g2}, s_g) = \frac{D_{g2} - P(D_{g2}=1|s_g)}{P(D_{g2}=1|s_g)(1-P(D_{g2}=1|s_g))}$$

Appendix A.4 Proposition 3

Let s_1, s_2 be two elements of $(0,1)$, such that $s_1 \neq s_2$. Whether Assumptions 1, 2 and 3 hold together or Assumptions 1′, 2′, 3′ and 4′ hold together, we have the following conditions

$$\tau^E(s_1) = \frac{\delta_0\left(1-\theta_0^w(1-s_1)\right)}{1-\theta_0^w + s_1(1-s_1)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \tag{32}$$

$$\tau^{NE}(s_1) = \frac{\delta_0 \theta_0^b s_1}{1-\theta_0^w + s_1(1-s_1)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \tag{33}$$

$$\tau^E(s_2) = \frac{\delta_0\left(1-\theta_0^w(1-s_2)\right)}{1-\theta_0^w + s_2(1-s_2)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \tag{34}$$

$$\tau^{NE}(s_2) = \frac{\delta_0 \theta_0^b s_2}{1-\theta_0^w + s_2(1-s_2)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \tag{35}$$

where equations (32) and (34) are written in the reduced form, as derived in (31). Note that for any $s \in (0,1)$ and for any $(x_1, x_2) \in (-1,1) \times (-1,1)$,

$$1 - x_1 + s(1-s)\left[(x_1)^2 - (x_2)^2\right] > 0$$

Dividing (33) by (32) and (35) by (34), we get

$$\frac{\tau^{NE}(s_1)}{\tau^E(s_1)} = \frac{\theta_0^b s_1}{1 - \theta_0^w(1-s_1)} \tag{36}$$

$$\frac{\tau^{NE}(s_2)}{\tau^E(s_2)} = \frac{\theta_0^b s_2}{1 - \theta_0^w(1-s_2)} \tag{37}$$

Since it is supposed that $\theta_0^b \neq 0$, (36) can be divided by (37), and we get

$$\kappa = \frac{\frac{\tau^{NE}(s_1)}{\tau^E(s_1)}}{\frac{\tau^{NE}(s_2)}{\tau^E(s_2)}} = \frac{s_1(1 - \theta_0^w(1-s_2))}{s_2(1 - \theta_0^w(1-s_1))} \tag{38}$$

From there, it is clear that if there is a unique θ_0^w that solves (38), then there is a unique λ_0 that solves the system defined by (32), (33), (34) and (35). Let's suppose there exists $\tilde{\theta}^w \in (-1,1)$, $\tilde{\theta}^w \neq \theta_0^w$ such that

$$\frac{s_1(1 - \theta_0^w(1-s_2))}{s_2(1 - \theta_0^w(1-s_1))} = \frac{s_1(1 - \tilde{\theta}^w(1-s_2))}{s_2(1 - \tilde{\theta}^w(1-s_1))}$$

which can be re-expressed as

$$\begin{aligned}
&(1 - \theta_0^w(1-s_2))(1 - \tilde{\theta}^w(1-s_1)) = (1 - \theta_0^w(1-s_1))(1 - \tilde{\theta}^w(1-s_2)) \\
\implies & -\tilde{\theta}^w(1-s_1) - \theta_0^w(1-s_2) = -\tilde{\theta}^w(1-s_2) - \theta_0^w(1-s_1) \\
\implies & (\tilde{\theta}^w - \theta_0^w)(s_2 - s_1) = 0
\end{aligned}$$

Since $s_1 \neq s_2$ and $\tilde{\theta}^w \neq \theta_0^w$, this leads to a contradiction so necessarily, $\tilde{\theta}^w = \theta_0^w$. Consequently, the whole vector λ_0 is identified

Appendix A.5 Proposition 4

Let

$$\hat{\lambda}^* = \begin{pmatrix} \delta^* \\ \theta^{w*} \\ \theta^{b*} \end{pmatrix} := \arg\min_{\Theta} \, [\overline{m}(\lambda)]' \hat{V}_G^{-1} \overline{m}(\lambda)$$

with

$$\begin{aligned}
Z_g &:= (\overline{y}_g^E, \overline{y}_g^{NE}, D_g, s_g) \\
\overline{m}(\lambda) &= \frac{1}{G} \sum_{g=1}^{G} f(s_g) \hat{u}(Z_g; \lambda) \\
\hat{u}(Z_g; \lambda) &= \begin{pmatrix} \rho^R(D_g)\overline{y}_g^E - \dfrac{\delta\left(1 - \theta^w(1-s_g)\right)}{1 - \theta^w + s_g(1-s_g)\left[(\theta^w)^2 - (\theta^b)^2\right]} \\ \rho^R(D_g)\overline{y}_g^{NE} - \dfrac{\delta\theta^b s_g}{1 - \theta^w + s_g(1-s_g)\left[(\theta^w)^2 - (\theta^b)^2\right]} \end{pmatrix}
\end{aligned}$$

and $\hat{V}_G \xrightarrow{p} V_0 := \mathbb{E}\left[f(s_g)u(Z_g;\lambda_0)u(Z_g;\lambda_0)'f(s_g)'\right]$. According to Theorem 2.6 of Newey and McFadden (1994), if

1. V_0 is positive semi-definite and $V_0\,\mathbb{E}\left[f(s_g)\hat{u}(Z_g;\lambda)\right] = 0$ only if $\lambda = \lambda_0$
2. Θ is compact
3. $f(s_g)\hat{u}(Z_g;\lambda)$ is continuous at each $\lambda \in \Theta$ almost surely
4. $\mathbb{E}\left[\sup_{\lambda\in\Theta}||f(s_g)\hat{u}(Z_g;\lambda)||\right] < +\infty$

Then, $\hat{\lambda}^*$ is a consistent estimator of λ_0.

By assumptions, Conditions 1 and 2 are satisfied (1 holds for instance when the conditions of Proposition 3 are met). Since $\hat{u}$ is differentiable with respect to λ on Θ, it is also continuous with respect to λ. So Condition 3 is also satisfied. Finally, since for $k \in \{E, NE\}$, $\mathbb{E}\left[|\overline{y}^k|\right] < +\infty$ and $s_g \in (0,1]$, there exists $\kappa \in \mathbb{R}^+$ such that for any $\lambda \in \Theta$,

$$\frac{\delta \times \max(\theta^b s_g, 1-\theta^w(1-s_g))}{1-\theta^w + s_g(1-s_g)\left[(\theta^w)^2 - (\theta^b)^2\right]} < \kappa$$

which implies that $\mathbb{E}\left[\sup_{\lambda\in\Theta}||f(s_g)\hat{u}(Z_g;\lambda)||\right] < +\infty$ so Condition 4 is also satisfied. As a conclusion, $\hat{\lambda}^*$ is a consistent estimator of λ_0. Then, according to Theorem 3.4 of Newey and McFadden (1994), if

1. $\lambda_0 \in \text{Int}(\Theta)$
2. $f(s_g)\hat{u}(Z_g;\lambda)$ is continuous differentiable in a neighborhood of λ_0, almost surely
3. $\mathbb{E}\left[f(s_g)\hat{u}(Z_g;\lambda_0)\right] = 0$ and $\mathbb{E}\left[||f(s_g)\hat{u}(Z_g;\lambda_0)||^2\right] < +\infty$
4. $\mathbb{E}\left[\sup_{\lambda\in\Theta}||\nabla_\lambda \hat{u}(Z_g;\lambda)||^2\right] < +\infty$
5. $M_0'V_0^{-1}M_0$ is non-singular with $M_0 := \mathbb{E}\left[f(s_g)\frac{\partial u^R(Z_g;\lambda_0)}{\partial\lambda'}\right]$

Then,

$$\sqrt{G}(\hat{\lambda}^* - \lambda_0) \xrightarrow{d} \mathcal{N}\left(0, \left[M_0'V_0^{-1}M_0\right]^{-1}\right)$$

By assumption, Condition 1 holds. Since $\hat{u}(Z_g;\lambda)$ is twice differentiable with respect to λ on Θ, it is continuously differentiable, so condition 2 holds as well. Since, for $k \in \{E, NE\}$, $\mathbb{E}\left[|\overline{y}^k|^2\right] < +\infty$ and $s_g \in (0,1]$, then Assumption 3 holds as well. Points 4 and 5 hold as well, due to the restriction on the distribution of $\overline{y}^k$, on the support of s_g and on Θ.

Appendix B Extension - Including Covariates

In this section, we show how the model can be adapted, in the natural experiment setting, so as to make the common trends assumption more plausible.

Appendix B.1 Identification

Let ΔX_g denote a vector of observed changes of some of group g's exogenous characteristics and let $\mathcal{X}$ denote the support of ΔX_g. Let's suppose Assumption $4'$ does not hold but, instead, the following condition is satisfied

Assumption 3$''$ (Extended Conditional Common Trends)**.** *For any $k \in \{E, NE\}$,*

$$E\left[\bar{e}^k_{g2} - \bar{e}^k_{g1}|s_g, \Delta X_g, D_{g2} = 1\right] = E\left[\bar{e}^k_{g2} - \bar{e}^k_{g1}|s_g, \Delta X_g, D_{g2} = 0\right]$$

Intuitively, we are going to compare pairs of treated and control groups with the same share of eligible units and the same evolution of the covariates X from period 1 to 2. Then, we assume that, in the absence of the treatment, the change in the average aggregate outcome among eligible units in treated groups would have been the same as the change in the average aggregate outcome among eligible units in control groups. We make the same assumption regarding the change in group average outcome among non-eligible units. Finally,

Assumption 4$''$ (Extended Conditional Common Support)**.** *For every $(g, t) \in \{1, \ldots, G\} \times \{1, 2\}$,*

$$D_{g1} = 0 \text{ a.s and for all } s \in \mathcal{S} \text{ and } x \in \mathcal{X}, 0 < P(D_{g2} = 1|s, x) < 1$$

Then, the previous results can be adapted to this new context

Proposition 5 (Conditional Moment Conditions - Natural Experiment with Covariates)**.** *Let I be an interval on $\mathbb{R}\backslash\{0\}$. Let $\lambda_0 = (\delta_0, \theta^w_0, \theta^b_0)$ be the true value of the parameters with $\lambda_0 \in \Theta := I \times (-1, 1) \times (-1, 1)$. For any $s \in (0, 1)$, provided Assumptions $1'$, $2'$, $3''$ and $4''$ hold and all the mentioned conditional moments are well-defined,*

$$\mathbb{E}\left[u^{DiDX}(Z_g; \lambda_0)|s_g = s, \Delta X_g = x\right] = \begin{pmatrix} \mathbb{E}\left[u^{DiDX,E}(Z_g; \lambda_0)|s_g = s, \Delta X_g = x\right] \\ \mathbb{E}\left[u^{DiDX,NE}(Z_g; \lambda_0)|s_g = s, \Delta X_g = x\right] \end{pmatrix} = 0$$

where $Z_g := (\Delta\bar{y}^E_g, \Delta\bar{y}^{NE}_g, D_{g2}, s_g, \Delta X_g)$ and for all $\lambda \in \Theta$,

$$u^{DiDX,E}(Z_g; \lambda) = \rho^{DiDX}(D_{g2}, s_g, \Delta X_g)\Delta\bar{y}^E_g - \frac{\delta\left(1 - \theta^w(1 - s_g)\right)}{1 - \theta^w + s_g(1 - s_g)\left[(\theta^w)^2 - (\theta^b)^2\right]}$$

$$u^{DiDX,NE}(Z_g; \lambda) = \rho^{DiDX}(D_{g2}, s_g, \Delta X_g)\Delta\bar{y}^{NE}_g - \frac{\delta\theta^b s_g}{1 - \theta^w + s_g(1 - s_g)\left[(\theta^w)^2 - (\theta^b)^2\right]}$$

and $\rho^{DiDX}(D_{g2}, s_g, \Delta X_g) = \dfrac{D_{g2} - P(D_{g2} = 1|s_g, \Delta X_g)}{P(D_{g2} = 1|s_g, \Delta X_g)(1 - P(D_{g2} = 1|s_g, \Delta X_g))}$

Proof. Provided they are well-defined, the conditional expectations of $\Delta\bar{y}_g^E$ and $\Delta\bar{y}_g^{NE}$ with respect to $(D_{g2}, s_g, \Delta X_g)$ are

$$\begin{aligned}
\mathbb{E}\left[\Delta\bar{y}_g^E \middle| D_{g2}, s_g, \Delta X_g\right] = & \frac{\theta_0^b(1-s_g)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \mathbb{E}\left[\bar{e}_{g2}^{NE} - \bar{e}_{g1}^{NE} \middle| D_{g2}, s_g, \Delta X_g\right] \\
& + \frac{1-\theta_0^w(1-s_g)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \mathbb{E}\left[\bar{e}_{g2}^{E} - \bar{e}_{g1}^{E} \middle| D_{g2}, s_g, \Delta X_g\right] \\
& + \frac{\delta_0\left(1-\theta_0^w(1-s_g)\right)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} D_{g2}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\Delta\bar{y}_g^{NE} \middle| D_{g2}, s_g, \Delta X_g\right] = & \frac{1-\theta_0^w s_g}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \mathbb{E}\left[\bar{e}_{g2}^{NE} - \bar{e}_{g1}^{NE} \middle| D_{g2}, s_g, \Delta X_g\right] \\
& + \frac{\theta_0^b s_g}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \mathbb{E}\left[\bar{e}_{g2}^{E} - \bar{e}_{g1}^{E} \middle| D_{g2}, s_g, \Delta X_g\right] \\
& + \frac{\delta_0 \theta_0^b s_g}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} D_{g2}
\end{aligned}$$

Now,

$$\mathbb{E}\left[\Delta\bar{y}_g^E \middle| D_{g2}=1, s_g, \Delta X_g\right] - \mathbb{E}\left[\Delta\bar{y}_g^E \middle| D_{g2}=0, s_g, \Delta X_g\right] = \frac{\delta_0\left(1-\theta_0^w(1-s_g)\right)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}$$

$$\mathbb{E}\left[\Delta\bar{y}_g^{NE} \middle| D_{g2}=1, s_g, \Delta X_g\right] - \mathbb{E}\left[\Delta\bar{y}_g^{NE} \middle| D_{g2}=0, s_g, \Delta X_g\right] = \frac{\delta_0 \theta_0^b s_g}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}$$

using Assumption 3″. Finally, using the same steps as in the proof for Proposition 4, one can show that, for any $k \in \{E, NE\}$,

$$\begin{aligned}
\mathbb{E}\left[\rho^{DiDX}(D_{g2}, s_g, \Delta X_g)\Delta\bar{y}_g^E | s_g, \Delta X_g\right] &= \mathbb{E}\left[\frac{D_{g2} - P(D_{g2}=1|s_g,\Delta X_g)}{P(D_g=1|s_g,\Delta X_g)(1-P(D_g=1|s_g,\Delta X_g))}\Delta\bar{y}_g^k | s_g, \Delta X_g\right] \\
&= \frac{P(D_{g2}=1|s_g,\Delta X_g)(1-P(D_{g2}=1|s_g,\Delta X_g))}{P(D_{g2}=1|s_g,\Delta X_g)(1-P(D_{g2}=1|s_g,\Delta X_g))} \mathbb{E}\left[\Delta\bar{y}_g^k | s_g, \Delta X_g, D_{g2}=1\right] \\
&\quad - \frac{(1-P(D_{g2}=1|s_g,\Delta X_g))P(D_{g2}=1|s_g,\Delta X_g)}{P(D_{g2}=1|s_g,\Delta X_g)(1-P(D_{g2}=1|s_g,\Delta X_g))} \mathbb{E}\left[\Delta\bar{y}_g^k | s_g, \Delta X_g, D_{g2}=0\right] \\
&= \mathbb{E}\left[\Delta\bar{y}_g^k | s_g, \Delta X_g, D_{g2}=1\right] - \mathbb{E}\left[\bar{y}_g^k | s_g, \Delta X_g, D_{g2}=0\right]
\end{aligned}$$

□

Appendix B.2 Estimation

In this section, we present preliminary results on a practical way of estimating $\lambda_0 := (\delta_0, \theta_0^w, \theta_0^b)$ in a natural experiment setting where Assumption 4′ fails but Assumption 3″ holds. We consider the following GMM estimator

$$\hat{\lambda}^X := \arg\min_{\Theta} \, [\overline{m}(\lambda)]'\overline{m}(\lambda) \tag{39}$$

with $Z_g := (\Delta\overline{y}_g^E, \Delta\overline{y}_g^{NE}, D_{g2}, s_g, \Delta X_g)$, $\overline{m}(\lambda) = \dfrac{1}{G}\sum_{g=1}^{G} f(s_g)\hat{u}(Z_g;\lambda)$,

$$\hat{u}(Z_g;\lambda) = \begin{pmatrix} \hat{\rho}^{DiDX}(D_{g2}, s_g, \Delta X_g)\Delta\overline{y}_g^E - \dfrac{\delta\left(1-\theta^w(1-s_g)\right)}{1-\theta^w + s_g(1-s_g)\left[(\theta^w)^2-(\theta^b)^2\right]} \\ \hat{\rho}^{DiDX}(D_{g2}, s_g, \Delta X_g)\Delta\overline{y}_g^{NE} - \dfrac{\delta\theta^b s_g}{1-\theta^w + s_g(1-s_g)\left[(\theta^w)^2-(\theta^b)^2\right]} \end{pmatrix}$$

and

$$\hat{\rho}^{DiDX}(D_{g2}, s_g, \Delta X_g) = \frac{\hat{P}(D_{g2}=1|s_g,\Delta X_g)}{\hat{P}(D_{g2}=1|s_g,\Delta X_g)(1-\hat{P}(D_{g2}=1|s_g,\Delta X_g))}$$

where $\hat{P}(D_{g2}=1|s_g,\Delta X_g)$ is estimated, in a first step, using a probit regression of D_{g2} on a constant, s_g and the vector ΔX_g.

We evaluate this estimation procedure via simulations, using the following DGP. For each group g, we observe a vector $Z_g = (\Delta\overline{y}_g^E, \Delta\overline{y}_g^{NE}, s_g, D_g, \Delta X_g^E, \Delta X_g^{NE})$ such that

$$\begin{cases} \overline{y}_{g1}^E &= (\alpha_g^E + \varepsilon_{g1}^E) + \theta_0^w s_g \overline{y}_{g1}^E + \theta_0^b(1-s_g)\overline{y}_{g1}^{NE} \\ \overline{y}_{g2}^E &= (\alpha_g^E + \lambda^E s_g + \beta^E \Delta X_g^E + \varepsilon_{g2}^E) + \theta_0^w s_g \overline{y}_{g2}^E + \theta_0^b(1-s_g)\overline{y}_{g2}^{NE} + \delta_0 D_{g2} \\ & \\ \overline{y}_{g1}^{NE} &= (\alpha_g^{NE} + \varepsilon_{g1}^{NE}) + \theta_0^b s_g \overline{y}_{g1}^E + \theta_0^w(1-s_g)\overline{y}_{g1}^{NE} \\ \overline{y}_{g2}^{NE} &= (\alpha_g^{NE} + \lambda^{NE} s_g^2 + \beta^{NE}\Delta X_g^{NE} + \varepsilon_{g2}^{NE}) + \theta_0^b s_g \overline{y}_{g2}^E + \theta_0^w(1-s_g)\overline{y}_{g2}^{NE} \end{cases}$$

with

$$\begin{pmatrix} \beta^E \\ \beta^{NE} \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}$$

$$\begin{pmatrix} \lambda^E \\ \lambda^{NE} \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$$

$$\varepsilon_g = \begin{pmatrix} \varepsilon_{g1}^E \\ \varepsilon_{g2}^E \\ \varepsilon_{g1}^{NE} \\ \varepsilon_{g2}^{NE} \end{pmatrix} \sim \mathcal{N}(0, \Sigma_\varepsilon)$$

$$s_g \sim F^{-1}_{\mathcal{U}_{(0,1]}}(V_g^1)$$

$$D_{g2} \sim F^{-1}_{\mathcal{B}(1/2)}(V_g^2)$$

$$\begin{pmatrix} \alpha_g^E \\ \alpha_g^{NE} \\ V_g^1 \\ V_g^2 \\ \Delta X_g^E \\ \Delta X_g^{NE} \end{pmatrix} \sim \mathcal{N}(0, \Sigma_{\alpha,V,X})$$

where Σ_ε and $\Sigma_{\alpha,V,X}$ are positive definite symmetric matrices whose eigenvalues are respectively 2 and 1 (resp. 1, 2, 3, 4 and 5). By construction, Assumptions $1'$ and $2'$ are satisfied with

$$\begin{aligned}
\bar{e}^E_{g1} &= \alpha^E_g + \varepsilon^E_{g1} \\
\bar{e}^{NE}_{g1} &= \alpha^{NE}_g + \varepsilon^{NE}_{g1} \\
\bar{e}^E_{g2} &= \alpha^E_g + \lambda^E s_g + \beta^E \Delta X^E_g + \varepsilon^E_{g2} \\
\bar{e}^{NE}_{g2} &= \alpha^{NE}_g + \lambda^{NE} s_g^2 + \beta^{NE} \Delta X^{NE}_g + \varepsilon^{NE}_{g2}
\end{aligned}$$

Note that

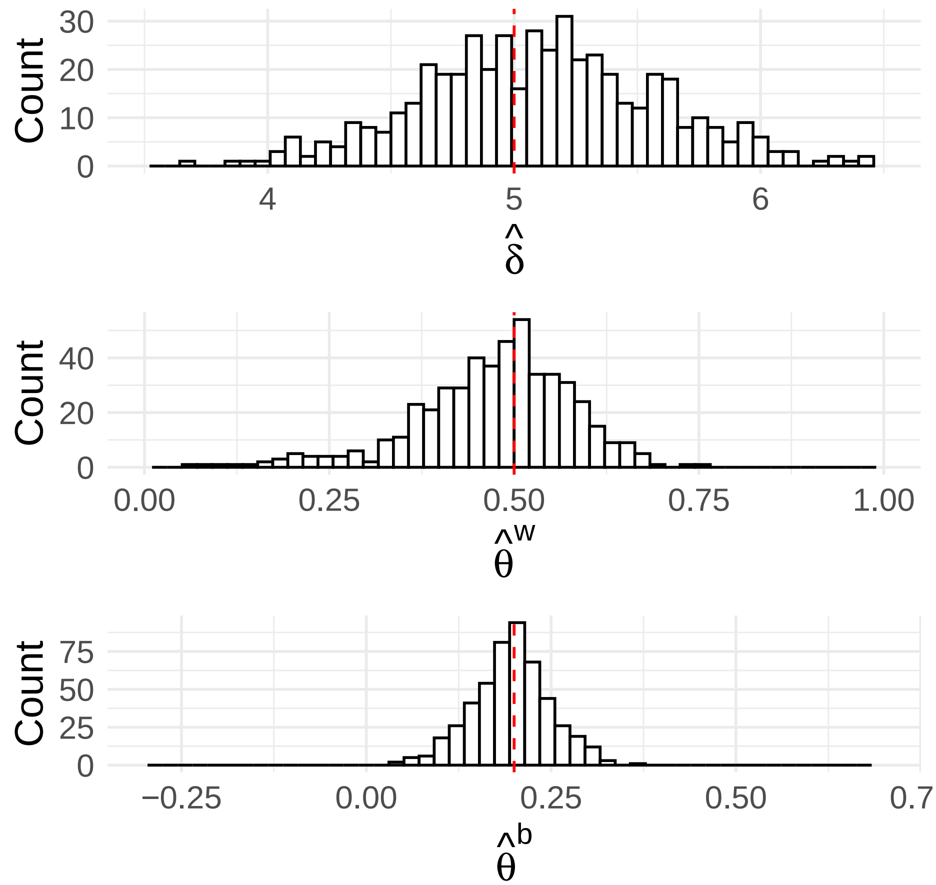
$$\begin{aligned}
\mathbb{E}\left[\bar{e}^E_{g2} - \bar{e}^E_{g1}\middle| s_g, D_{g2} = 1, \Delta X_g\right] &= \mathbb{E}\left[\bar{e}^E_{g2} - \bar{e}^E_{g1}\middle| s_g, D_{g2} = 0, \Delta X_g\right] = \lambda^E s_g + \beta^E \Delta X^E_g \\
\mathbb{E}\left[\bar{e}^{NE}_{g2} - \bar{e}^{NE}_{g1}\middle| s_g, D_{g2} = 1, \Delta X_g\right] &= \mathbb{E}\left[\bar{e}^{NE}_{g2} - \bar{e}^{NE}_{g1}\middle| s_g, D_{g2} = 0, \Delta X_g\right] = \lambda^{NE} s_g^2 + \beta^{NE} \Delta X^{NE}_g
\end{aligned}$$

so Assumption $3''$ is also satisfied: conditional on s_g and $\Delta X_g = (\Delta X^E_g, \Delta X^{NE}_g)$, treated and control groups have the same time trend, and this trend is potentially different for eligibles and non-eligibles. Finally, Assumption $4''$ holds with probability 1.

Figure A1 plots the distribution of $\hat{\lambda}^X$ from 500 simulations. Meanwhile, Figure A2 plots the distribution of $\hat{\lambda}^*$, the baseline estimator in a natural experiment setting. As expected, $\hat{\lambda}^*$ appears to be biased. Interestingly, even though we don't use a nonparametric first step to estimate $P(D_{g2} = 1|\Delta X_g, s_g)$, we significantly manage to reduce the bias when controlling for covariates using a logitistic regression.

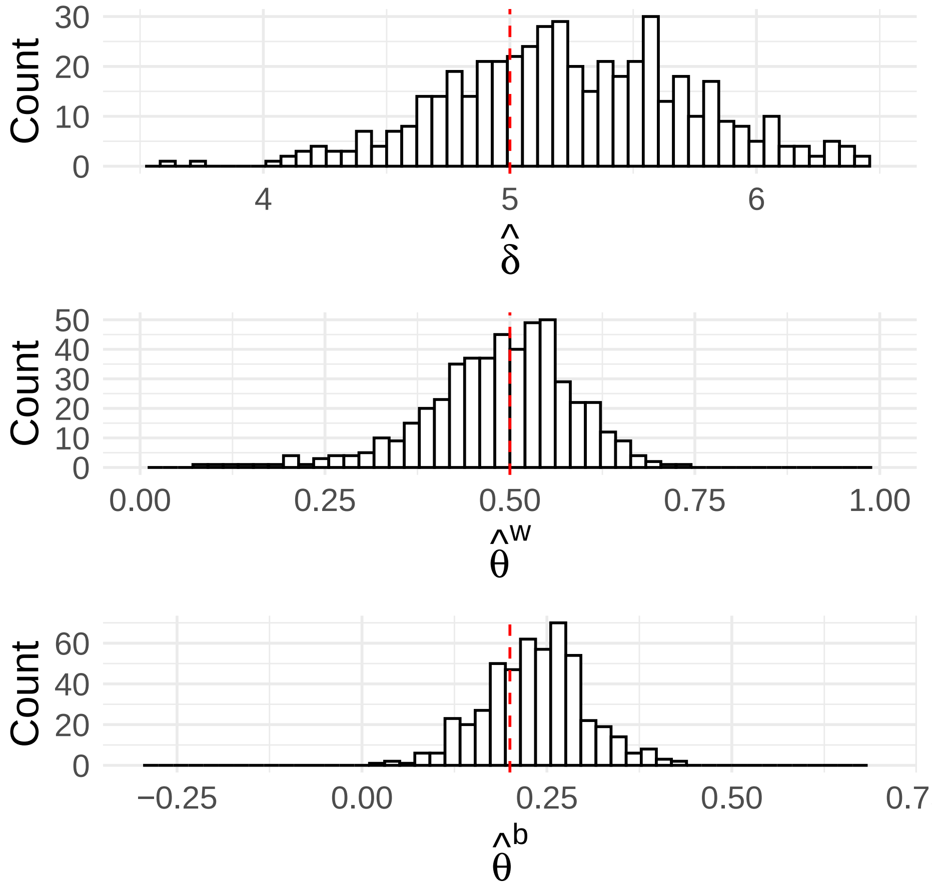
Appendix C Additional Figures and Tables

Figure A1: Distribution of $\hat{\lambda}^X$ in the Natural Experiment Setting with Conditional Common Trends - Controlling for Covariates



Note: The Figure plots the distribution of the GMM estimator λ^* defined in the Simulations section (Implementation Challenges), based on 500 simulations. The dashed red lines indicate the true value of the parameter.

Figure A2: Distribution of $\hat{\lambda}^*$ in the Natural Experiment Setting with Conditional Common Trends - not Controlling for Covariates



Note: The Figure plots the distribution of the GMM estimator λ^* defined in the Estimation section (Natural Experiment subsection), based on 500 simulations, while Assumption 4′ does not hold but Assumption 3′′ is satisfied. The dashed red lines indicate the true value of the parameter.

Figure A3: Progresa: distribution of s, by treatment status