CRESTCENTER FOR RESEARCH IN ECONOMICS AND STATISTICS

WORKING PAPER SERIES

Estimating heterogeneous peer effects with partial population experiments

Felix Pasquier, Pauline Rossi, Zheng Wang

N°11/September 2026

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: communication.crest@ensae.fr
https://crest.science/

Estimating heterogeneous peer effects with partial population experiments*

Felix Pasquier
CREST-ENSAE

Pauline Rossi
CREST-École Polytechnique

Zheng Wang
NYU Abu Dhabi

September 15, 2026

Abstract

The standard linear-in-means model of peer effects assumes that the endogenous peer effect parameter is homogeneous. In the context of group interactions, we relax this assumption by allowing individuals to respond differently to the outcomes of other group members depending on the identity of these members. 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 provide a package to implement our method in R.

KEYWORDS: PEER EFFECTS, PARTIAL POPULATION EXPERIMENTS, HETEROGENEOUS EFFECTS

*We are grateful to Xavier d'Haultfoeuille, Margherita Comola, Lena Janys, Michael Visser and Nicolas Ziebarth for insightful comments. We thank Laurine Meier for excellent research assistance, in particular for creating the R package. Rossi gratefully acknowledges funding from the European Union (ERC, P3OPLE, ERC-2021-StG 101039252). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. The R software package is available at `https://github.com/LaurineMir/heterogeneouspeereffects`

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 when identity and eligibility are based on the same characteristic, we can identify within and between peer effects when the share of eligible individuals 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). Third, we show that the baseline model can be extended to study additional dimensions of heterogeneity. The first extension allows the between and within parameters to vary across sub-groups. This is important to study asymmetric social dynamics, like influencers-followers. The second extension considers designs where the identity that matters for the heterogeneous peer effect dimension is orthogonal to treatment eligibility. This is important because the eligibility dimension is not always interesting; for instance, in two-stage randomized experiments, eligible individuals are chosen randomly and therefore, have no meaningful "identity". Our methodology allows researchers to use these designs to study any dimension of heterogeneity observed in the data. We provide a package to implement our method in R.

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. Separating both effects is crucial since only the endogenous component gives rise to population multipliers.

To illustrate how our methodology can be applied, we start by simulating data corresponding to different types of group dynamics. We vary the way individuals are influenced by others in terms of magnitude and sign. For example, in the case of polarization, the "within" parameter is null and the "between" parameter is negative; in the case of role model, the between parameter is positive from influencers to followers and null from followers to influencers. We then estimate our model with heterogeneous parameters and contrast the results with the results we would have obtained under the homogeneity assumption. We show that the homogeneous model can lead to all sorts of biases whose signs and magnitudes vary with the type of group dynamics.

Next, we apply our methodology on real data 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 0.31, a "within" peer effect of 0.24, and a direct effect of $5p.p.$. Yet, we cannot reject the null that all those parameters are zero.

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 is at the intersection of two streams of literature.[2] First, it 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

[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.

[2]There is also a literature on networks interactions exploiting the structure of the network to model heterogeneity in peer effects (e.g. Lee and Liu (2010); Gupta and Robinson (2015); Wu et al. (2022); Comola, Rokhaya, and Fortin (2023)). The main limitation is that the true network is rarely observed by researchers. Instead, we focus on settings in which the researcher defines the group as a relevant unit observed in the data, such as a classroom, a school or a village, and assumes that all members of the group interact.

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. An interesting case is polarization, for which we obtain a non-monotonous effect (inverted-U) of s on the eligibles.
Second, it 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 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.
At this intersection, we are aware of two other papers. The first one is Masten (2018) that considers a linear-in-means models with random coefficients on the endogenous variable. This model implies that individuals respond differently to the average behavior of their peers. On the contrary, in our model, individuals respond differently to their peers' behavior depending on their peers' identity. The second one is Arduini, Patacchini, and Rainone (2020). They consider a similar linear-in-means model and propose a methodology to estimate "between" and "within" eligibles and non-eligibles parameters in the case of a randomized experiments. We develop a different estimation strategy, which does not require the share of eligibles to be exogenous to the treatment and is therefore compatible with natural experiments. Importantly, our methodology also allows us to study heterogeneity along characteristics that are unrelated to eligibility.
The remainder of the paper is organized as follows. Section 2 introduces notations, presents our baseline model and discusses our main assumptions. Section 3 explains our identification strategy. Section 4 presents our estimation results. In section 5, we extend the baseline model to allow for additional dimensions of heterogeneity. Finally, sections 6 and 7 show how to apply our methodology using simulated data and Progresa data.

2 Set-up, notations and main assumptions

We start by describing the set-up with randomized treatment assignment and then turn to the difference-in-difference.

2.1 Randomized experiment

We consider a setting inspired by Moffitt et al. (2001). We suppose that G reference groups are observed. Group g contains n_g units (indexed by i); and a share s_g of those units are eligible for a binary treatment, denoted D_g. Some groups receive the treatment while others don't. 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 unit i has received the treatment. We suppose perfect compliance. None of the ineligible individuals may receive the treatment, i.e.

$P(D_{ig} = 1|E_{ig} = 0) = 0$. And if group g is treated, all of the eligible units receive the treatment, i.e. $P(D_{ig} = 1|E_{ig} = 1, D_g = 1) = 1$. Let $n_g^E = \sum_{i=1}^{n_g} E_{ig} = s_g n_g$ and $n_g^N = \sum_{i=1}^{n_g}(1 - E_{ig}) = n_g - n_g^E$ 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". In our application that evaluates the effects of Progresa, a reference group corresponds to a grade level associated to a village, the treatment corresponds to Progresa's conditional cash transfer that was randomly allocated across villages, and eligibility status is determined by a threshold on the income distribution.

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

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

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 average outcome of her peers in group g. 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. Individuals may be differently influenced by their peers' outcomes depending on whether they share the same "identity" or not. In the following, for any variable w, we denote the average among eligibles and the average among non-eligibles as follows:
$\bar{w}_g^E = \frac{1}{n_g^E}\sum_{i\in\mathcal{E}_g} w_{ig}^E$ and $\bar{w}_g^N = \frac{1}{n_g^N}\sum_{i\in\mathcal{O}_g} w_{ig}^N$ where $n_g^E = \sum_{i=1}^{n_g} E_{ig}$ and $n_g^N = n_g - n_g^E$.
We make the following assumptions.

Assumption 1 (Linear-in-means model)
Within each group $g \in \{1, \dots, G\}$ of the i.i.d sample, the outcome of individual i is defined as

$$y_{ig} = E_{ig}y_{ig}^E + (1 - E_{ig})y_{ig}^N$$

with

$$y_{ig}^E = \theta_0^w s_g \bar{y}_g^E + \theta_0^b(1 - s_g)\bar{y}_g^N + \delta_0 D_{ig} + e_{ig}^E \tag{2}$$
$$y_{ig}^N = \theta_0^b s_g \bar{y}_g^E + \theta_0^w(1 - s_g)\bar{y}_g^N + e_{ig}^N \tag{3}$$

The parameters of interest are δ_0, θ_0^b and θ_0^w, with $\delta_0 \in \mathbb{R}^*$ and $\theta_0^b, \theta_0^w \in (-1, 1)$. δ_0 represents the direct effect of the treatment on the eligible individuals that receive it. It is assumed to be homogeneous across groups. θ_0^w is the coefficient for the endogenous peer effects from peers who share the same identity as the individual while θ_0^w is the one for endogenous peers from peers of a different identity. For a given group g, when $\bar{y}_g^E$ and $\bar{y}_g^N$ both increase by 1, the outcome of an eligible unit will increase by $\theta_0^w s_g + \theta_0^b(1 - s_g)$ and the outcome of a non-eligible unit will increase by $\theta_0^b s_g + \theta_0^w(1 - s_g)$. These quantities are convex combinations of the within and between endogenous peer effects, weighted by the share of the eligible units. To illustrate the interpretation of (θ_0^w, θ_0^b), 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 $\bar{y}_g^N$ by respectively Δ^E and Δ^N results in an increase in an eligible individual's outcome by $\frac{1}{2}(\theta_0^w \Delta^E + \theta_0^b \Delta^N)$. 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. For conciseness, the residual terms e_{ig}^E and e_{ig}^N capture all the individual, exogenous and correlated effects. Because our focus is on θ_0^b and θ_0^w, we do not express explicitly all these effects in our baseline model. One may think of these terms, for example, as

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

Two features of this baseline model have to be emphasized at this stage. First, eligibility to the treatment corresponds to the "identity" that matters when evaluating the heterogeneity of peer effects. In our Progresa application for instance, we evaluate how the action of a child coming from a poor (resp. relatively rich) family is impacted by the actions of her peers who also come from the same economic background and by the actions of her peers coming from a different economic background. In Section 5, we consider a setting where identity is different from treatment eligibility. In the Progresa application, for example, one may want to consider gender as a more relevant identity to model the complex social mechanisms regarding a child's decision to drop-out as boys may be more influenced by the actions of other boys than by the actions of girls. Second, by allowing for heterogeneous peer effects, we may uncover social mechanisms beyond the conventional LIM model, such as polarisation. This mechanism happens when $\theta_0^b < 0$ and $\theta_0^w > 0$. For instance, a disruptive child might have a negative influence on his friends' attitudes towards school while prompting other children to exert more effort.

From Assumption 1, we can average respectively y_{ig}^E and y_{ig}^N among eligible (resp. non-eligible) units in group g to get

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

After some development, we get the following reduced forms

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

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

Then, we specify the treatment assignment across groups.

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

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

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

The second type of designs we consider are natural experiments. It is a panel extension of the previous subsection. 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}. Let E_{igt} be the eligibility status of individual i in group g at time t. For $k \in \{E, N\}$, $\bar{y}_{gt}^k = \left(n_{gt}^k\right)^{-1}\sum_{i=1}^{n_{gt}^k} y_{igt}^k$ where n_{gt}^k is the number of individuals whose "identity" is k at time period t.

Assumption 1′ (Linear-in-means model with panel data)
Within each group g in $\{1,\dots,G\}$ present in an i.i.d panel, for each time period $t \in \{1,\dots,T\}$, the outcome of an individual i is defined as

$$y_{igt} = E_{igt}y_{igt}^E + (1-E_{igt})y_{igt}^N$$

with

$$y_{igt}^E = \theta_0^w s_{gt}\bar{y}_{gt}^E + \theta_0^b(1-s_{gt})\bar{y}_{gt}^N + \delta_0 D_{gt} + e_{igt}^E \tag{8}$$
$$y_{igt}^N = \theta_0^b s_{gt}\bar{y}_{gt}^E + \theta_0^w(1-s_{gt})\bar{y}_{gt}^N + e_{igt}^N \tag{9}$$

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

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

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

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 $(g,t) \in \{1,\dots,G\}\times\{1,2\}$,

$$s_{gt} = s_g$$

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)

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

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 $k \in \{NE, E\}$,

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

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 identify the different parts that contribute to the overall effect of the policy.

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 $M^N(s)$ are the *population multipliers* for the eligible and non-eligible individuals, 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, as illustrated in Figure 2. Indeed, s influences M in two ways:

- Through the initial shock: the strength of the shock is proportional to the share of eligible individuals
- 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. In section 6, we describe different scenarios to illustrate the non-linearities.

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

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

and

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

where $P^E(s)$ and $P^N(s)$ 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 $\theta_0^w <$

θ_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 $P^E = P^N = \frac{1}{1-\theta}$. As a consequence, the population multiplier is linear in s.
The shapes of $\tau^E(s)$ and $\tau^N(s)$ are informative about $\delta_0, \theta_0^b, \theta_0^w$. As an illustration, Figure 3 plots $\tau^E(s)$ and $\tau^N(s)$ 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:

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

First, looking at eligibles in groups with a very low fraction of eligibles is informative about the direct effect. Second, looking at eligible individuals in groups with a very high fraction of eligible individuals 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 $\tau^N(s)$ 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

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

and

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

Hence,

$$P^E(s) := \frac{\theta_0^w - (1-s)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]}{1-\theta_0^w + s(1-s)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \tag{18}$$

$$P^N(s) := \frac{1}{1-\theta_0^w + s(1-s)\left[(\theta_0^w)^2 - (\theta_0^b)^2\right]} \tag{19}$$

Proof. See Appendix □

Proposition 1 provides closed-form solutions for the total average treatment effects on the average outcome within the eligible (respectivily non-eligible) sub-population, for a given share of eligible individuals s, as well as for the strength of the propagation shock. The difference of conditional expectations between treated and untreated groups can be expressed as a single conditional moment of a weighted outcome, as shown by equations (14) and (16), whose weights are based on a transformation process à la Abadie (2005). We rewrite the treatment effects in this way because we want to replace two conditional expectations by one. This will allow us to combine the different conditional expectations into a single unconditional expectation, as explained in the estimation section. 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,*

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

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

$$u^{R,E}(\overline{y}_g^E, D_g, s_g; \lambda) = \omega^R(D_g)\overline{y}_g^E - \frac{\delta\left(1-\theta^w(1-s_g)\right)}{1-\theta^w + s_g(1-s_g)\left[(\theta^w)^2 - (\theta^b)^2\right]}$$

$$u^{R,NE}(\overline{y}_g^N, D_g, s_g; \lambda) = \omega^R(D_g)\overline{y}_g^N - \frac{\delta\theta^b s_g}{1-\theta^w + s_g(1-s_g)\left[(\theta^w)^2 - (\theta^b)^2\right]}$$

and $\omega^R(D_g) = \dfrac{D_g - \mu_D}{\mu_D(1-\mu_D)}$ *with* $\mu_D = P(D_g = 1)$

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 are obtained in the natural experiment design. In the following, for all $k \in \{E, N\}$, we use the conventional notation for first differenced variables:

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

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

$$\begin{aligned}\tau^E(s) &= \mathrm{E}\left[\Delta\bar{y}^E_g|s_g=s,D_g=1\right]-\mathrm{E}\left[\Delta\bar{y}^E_g|s_g=s,D_g=0\right]\\ &=\mathrm{E}\left[\frac{D_{g2}-\mu_D(s)}{\mu_D(s)(1-\mu_D(s))}.\Delta\bar{y}^E_g\middle|s_g=s\right] \qquad (21)\\ &=\delta_0\left(1+s.\frac{\theta^w_0-(1-s)\left[(\theta^w_0)^2-(\theta^b_0)^2\right]}{1-\theta^w_0+s(1-s)\left[(\theta^w_0)^2-(\theta^b_0)^2\right]}\right) \qquad (22)\end{aligned}$$

with $\mu_D(s)=\Pr(D_{g2}=1|s_g=s)$ *and*

$$\begin{aligned}\tau^N(s) &= \mathrm{E}\left[\Delta\bar{y}^N_g|s_g=s,T_g=1\right]-\mathrm{E}\left[\Delta\bar{y}^N_g|s_g=s,D_g=0\right]\\ &=\mathrm{E}\left[\frac{D_{g2}-\mu_D(s_g)}{\mu_D(s)(1-\mu_D(s))}.\Delta\bar{y}^N_g\middle|s_g=s\right] \qquad (23)\\ &=\frac{\delta_0\theta^b_0 s}{1-\theta^w_0+s(1-s)\left[(\theta^w_0)^2-(\theta^b_0)^2\right]} \qquad (24)\end{aligned}$$

Proposition 2 is the counterpart of Proposition 1 in the natural experiment setting. There are two main differences. First, the outcomes of interest here are the first differenced aggregate outcomes within the eligible and non-eligible subpopulations. Second, the weighting function ω^{DiD} is now a function of both the treatment assignment variable and the share of eligible individuals in the group. In this design, treatment assignment is allowed to be correlated with the share of eligible units. Once again, 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 $\lambda_0=(\delta_0,\theta^w_0,\theta^b_0)$ 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,*

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

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

$$\begin{aligned}u^{DiD,E}(\Delta\bar{y}^E_g,D_{g2},s_g;\lambda)&=\omega^{DiD}(D_{g2},s_g)\Delta\bar{y}^E_g-\frac{\delta\left(1-\theta^w(1-s_g)\right)}{1-\theta^w+s_g(1-s_g)\left[(\theta^w)^2-(\theta^b)^2\right]}\\ u^{DiD,NE}(\Delta\bar{y}^N_g,D_{g2},s_g;\lambda)&=\omega^{DiD}(D_{g2},s_g)\Delta\bar{y}^N_g-\frac{\delta\theta^b s_g}{1-\theta^w+s_g(1-s_g)\left[(\theta^w)^2-(\theta^b)^2\right]}\end{aligned}$$

and $\omega^{DiD}(D_{g2},s_g)=\dfrac{D_{g2}-\mu_D(s_g)}{\mu_D(s_g)(1-\mu_D(s_g))}$

Proof. Immediate consequence from Proposition 2 □

To improve the plausibility of the conditional common trends assumption, One may want to include covariates in the model. In Appendix B, 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.

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^N 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 $\tau^N(s)$ 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 □

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. Both the randomized and natural experiment settings lead to conditional moment restrictions of the form

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

where $u(Z_g;\lambda,\mu)$ is a 2-dimensional vector of known functions of the i.i.d random vector of observed variables Z_g, $\lambda \in \Theta$ is a vector of parameters and μ is a propensity score function that depends on the share $s_g \in [0,1]$, μ_D being the true propensity score.
When we want to estimate these conditional moments, we may not have enough groups for a given value of s, especially when s is continuous. To overcome this issue, we use the law of iterated expectations and consider unconditional moment restrictions of the form

$$\mathrm{E}\left[m(Z_g;\lambda_0,\mu_D)\right] = \mathrm{E}\left[w(s)u(Z_g;\lambda_0,\mu_D)\right] = 0$$

where w is a $k \times 2$ matrix of instruments, with $k \geq \dim(\lambda_0)$, whose entries are functions of s_g (for e.g. polynomials). Based on results from Newey and McFadden (1994), section 5.4, and Wooldridge (2010), section 14.5.3, we consider the set of optimal instruments of the form

$$w^*(s) = M_0'\Sigma_0(s)^{-1}$$

where $M_0 := \mathrm{E}\left[\dfrac{\partial u}{\partial \lambda'}(Z;\lambda_0,p_0)\right]$ and $\Sigma_0(s) = \mathrm{E}[u(Z_g;\lambda_0,\mu_D)u(Z_g;\lambda_0,\mu_D)'|s]$ and derive Generalized Method of Moments (GMM) estimators of the form

$$\hat{\lambda} := \arg\min_{\lambda\in\Theta} \left(\frac{1}{G}\sum_{g=1}^{G}\hat{m}(Z_g;\lambda,\hat{\mu}_D)\right)'\hat{\Omega}^{-1}\left(\frac{1}{G}\sum_{g=1}^{G}\hat{m}(Z_g;\lambda,\hat{\mu})\right) \tag{26}$$

with $\hat{m}(Z_g;\lambda,\mu) = \hat{w}^*(s_g)u(Z_g;\lambda,\mu)$, $\hat{w}^*(s) = \hat{M}'\hat{\Sigma}(s)^{-1}$ where $\hat{\mu}$, $\hat{M}'$ and $\hat{\Sigma}^{-1}$ are estimators, respectively, of μ_D, M_0 and Σ_0^{-1} and $\hat{\Omega} \xrightarrow{p} \Omega_0 = \mathrm{E}\left[\left(\frac{\partial u}{\partial \lambda}(Z;\lambda_0,\mu_D)\right)'\Sigma_0(s)^{-1}\left(\frac{\partial u}{\partial \lambda}(Z;\lambda_0,\mu_D)\right)\right]$.

4.1 Known propensity score

In the randomized experiment setting, the propensity score function is supposed to be known and, without loss of generality, is such that for any $s\in\mathcal{S}$, $\mu_D(s)=\mu_D=P(D=1)$. In this particular design,

$$\begin{aligned}
\hat{m}(Z;\lambda,\mu) &= \hat{m}(Z;\lambda)\\
&= \hat{w}(s)u_B^R(Z;\lambda)\\
&= \hat{w}(s)\begin{pmatrix}\mathbf{1}\{s>0\} & 0\\ 0 & \mathbf{1}\{s<1\}\end{pmatrix}u^R(Z,\lambda)\\
&= \hat{w}(s)\left\{\omega^R(D)\begin{pmatrix}\bar{y}^E\mathbf{1}\{s>0\}\\ \bar{y}^N\mathbf{1}\{s<1\}\end{pmatrix} - \frac{\delta}{\phi(s,\lambda)}\begin{pmatrix}(1-(1-s)\theta^w)\mathbf{1}\{s>0\}\\ s\theta^b\mathbf{1}\{s<1\}\end{pmatrix}\right\}
\end{aligned}$$

with

$$\begin{aligned}
\omega^R(D) &= \frac{D}{\mu_D} - \frac{1-D}{1-\mu_D}\\
\phi(s,\lambda) &= 1-\theta^w + s(1-s)\left[(\theta^w)^2-(\theta^b)^2\right]
\end{aligned}$$

We consider the GMM estimator based on the following procedure :

1. The multivariate nonlinear least squares (MNLS) estimator $\tilde{\lambda}^R$ defined as

$$\tilde{\lambda} = \arg\min_{\lambda\in\Theta}\left(\frac{1}{G}\sum_{g=1}^{G}\tilde{w}(s_g)u_B^R(Z_g,\lambda)\right)'\left(\frac{1}{G}\sum_{g=1}^{G}\tilde{w}(s_g)u_B^R(Z_g,\lambda)\right)$$

   is computed where, for any $s\in[0,1]$, $\tilde{w}(s)$ is a $k\times 2$ matrix of arbitrary functions (e.g. polynomials) of s and $k\geq \dim(\lambda_0)$

2. $\Sigma(s) = \mathrm{E}\left[u_B^R(Z_g,\lambda_0)u_B^R(Z_g,\lambda_0)'|s\right]$ is consistently estimated, for all $s\in[0,1]$. In our application, it is obtained using series estimation, i.e. by $u_B^R(Z,\tilde{\lambda}^R)u_B^R(Z,\tilde{\lambda}^R)'$ on polynomials of s. The order of the polynomials is chosen using cross-validation. This estimator is denoted $\hat{\Sigma}(s;\tilde{\lambda})$.

3. The final estimator

$$\hat{\lambda}^* = \arg\min_{\lambda\in\Theta}\left(\frac{1}{G}\sum_{g=1}^{G}\hat{w}(s_g;\tilde{\lambda})u_B^R(Z_g;\lambda)\right)'\hat{\Omega}^{-1}\left(\frac{1}{G}\sum_{g=1}^{G}\hat{w}(s_g;\tilde{\lambda})u_B^R(Z_g;\lambda)\right)$$

   is computed with $\hat{w}(s;\tilde{\lambda}) = \left(\frac{\partial u^R}{\partial\lambda}(s;\tilde{\lambda})\right)'\hat{\Sigma}(s;\tilde{\lambda})^{-1}$ and $\hat{\Omega} = \frac{1}{G}\sum_{g=1}^{G}\left(\frac{\partial u^R}{\partial\lambda}(s_g;\tilde{\lambda})\right)'\hat{\Sigma}(s;\tilde{\lambda})^{-1}\left(\frac{\partial u^R}{\partial\lambda}(s_g;\tilde{\lambda})\right)$

Proposition 4 (Optimal GMM estimator in the randomized setting)
Let $(Z_g)_{g=1,\dots,G}$ be an i.i.d sample of G groups, where $Z_g = (\overline{y}_g^E, \overline{y}_g^N, D_g, s_g)$. 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 $K \in (0,1)$ and $\lambda_0 \in Int(\Theta)$*
3. $E\left[\left|\overline{y}_g^E\right|^2\right] < +\infty$, $E\left[\left|\overline{y}_g^N\right|^2\right] < +\infty$, $\sup_{s\in[0,1]} \|\Sigma_0(s)\| < +\infty$ *where $\|.\|$ is a sub-multiplicative 2×2 matrix norm*
4. $\Sigma_0(s) > 0$ *almost surely*
5. $\Omega_0 := \mathbb{E}\left[\left(\frac{\partial u^R}{\partial \lambda}(s;\lambda_0)\right)' \Sigma_0(s)^{-1} \left(\frac{\partial u^R}{\partial \lambda}(s;\lambda_0)\right)\right]$ *is positive-definite*
6. *For $\mathcal{N}$, a neighborhood of λ_0, then* $\sup_{\lambda\in\mathcal{N}, s\in[0,1]} \left\|\hat{\Sigma}^{-1}(s;\lambda) - \Sigma_0^{-1}(s)\right\| = o_P(1)$

Then

$$\hat{\lambda}^{R,*} = \begin{pmatrix} \delta^* \\ \theta^{w*} \\ \theta^{b*} \end{pmatrix} := \arg\min_{\lambda\in\Theta} \left(\frac{1}{G}\sum_{g=1}^{G} \hat{m}(Z_g;\lambda)\right)' \hat{\Omega}^{-1} \left(\frac{1}{G}\sum_{g=1}^{G} \hat{m}(Z_g;\lambda)\right) \tag{27}$$

is a consistent estimator of λ_0, whose asymptotic distribution is

$$\sqrt{G}(\hat{\lambda}^* - \lambda_0) \xrightarrow{d} \mathcal{N}\left(0, \Omega_0^{-1}\right) \tag{28}$$

with $\lambda_0 := (\delta_0, \theta_0^w, \theta_0^b)$

Proof. See Appendix □

4.2 Unknown propensity score

In the natural experiment setting, we have

$$\begin{aligned} \hat{m}(Z;\lambda,\mu) &= \hat{w}(s) u_B^{DiD}(Z;\lambda,\mu) \\ &= \hat{w}(s) \begin{pmatrix} \mathbf{1}\{s>0\} & 0 \\ 0 & \mathbf{1}\{s<1\} \end{pmatrix} u^{DiD}(Z;\lambda,\mu) \\ &= \hat{w}(s) \left\{ \omega^{DiD}(Z;\mu) \begin{pmatrix} \Delta\bar{y}^E \mathbf{1}\{s>0\} \\ \Delta\bar{y}^N \mathbf{1}\{s<1\} \end{pmatrix} - \frac{\delta}{\phi(s,\lambda)} \begin{pmatrix} (1-(1-s)\theta^w)\mathbf{1}\{s>0\} \\ s\theta^b \mathbf{1}\{s<1\} \end{pmatrix} \right\} \end{aligned}$$

with

$$\omega^{DiD}(Z,\mu) = \frac{D_2}{\mu_D(s)} - \frac{1-D_2}{1-\mu_D(s)}$$

In this setting, μ_D is often unknown. The following estimation procedure is thus considered:

1. The sample is split randomly into two sub-samples that are denoted $\mathcal{G}_1$ and $\mathcal{G}_2$.

2. The propensity score function μ_D is estimated non-parametrically based on the sub-sample $\mathcal{G}_2$. This estimator is denoted $\hat{\mu}_D^{(2)}$. The estimator $\hat{\mu}_D^{(1)}$ is similarly defined, using the observations in $\mathcal{G}_1$.

3. For both $k \in \{1,2\}$, the multivariate nonlinear least squares (MNLS) estimator $\tilde{\lambda}^{(k)}$, defined as

$$\tilde{\lambda}^{(k)} = \arg\min_{\lambda\in\Theta} \left(\frac{1}{G_k} \sum_{g\in\mathcal{G}_k} \tilde{w}(s_g) u_B^{DiD} \left(Z_g; \lambda, \hat{\mu}_D^{(3-k)} \right) \right)' \left(\frac{1}{G_k} \sum_{g\in\mathcal{G}_k} \tilde{w}(s_g) u_B^{DiD} \left(Z_g; \lambda, \hat{\mu}_D^{(3-k)} \right) \right)$$

   is computed with $G_k := |\mathcal{G}_k|$ and $\tilde{w}$ is defined as previously

4. The estimator

$$\tilde{\lambda} = \frac{G_1}{G_1+G_2}\tilde{\lambda}^{(1)} + \frac{G_2}{G_1+G_2}\tilde{\lambda}^{(2)}$$

   is computed

5. $\Sigma_0(s) = \mathrm{E}\left[u_B^{DiD}(Z_g;\lambda_0,\mu_D)u_B^{DiD}(Z_g;\lambda_0,\mu_D)'|s\right]$ is consistently estimated, for all $s \in [0,1]$, distinctly for each sub-sample. In our application, it is obtained using series estimation and cross-fitting. Specifically, for a given $k \in \{1,2\}$, $u_B^{DiD}(Z;\tilde{\lambda},\hat{\mu}_D^{(3-k)})u_B^{R}(Z;\tilde{\lambda},\hat{\mu}_D^{(3-k)})'$ is regressed on polynomials in s where the order of the polynomials is chosen via cross-validation. This estimator is denoted $\hat{\Sigma}^{(k)}\left(s;\tilde{\lambda},\hat{\mu}_D^{(3-k)}\right)$

6. The functions $\mu_d : s \mapsto E[\Delta\bar{y}|D_2 = d, s]$ for $d \in \{0,1\}$ are nonparametrically estimated within each sub-samples, similarly as in step 1. Their estimators are, respectfully, denoted $\hat{\mu}_0^{(k)}$ and $\hat{\mu}_1^{(k)}$, for $k \in \{1,2\}$.

7. For both $k \in \{1,2\}$, the estimator

$$\hat{\lambda}^{(k)} = \arg\min_{\lambda\in\Theta} \left(\frac{1}{G_k} \sum_{g\in\mathcal{G}_k} \hat{w}^{(k)} u_B^{DiD} \left(Z_g; \lambda, \hat{\mu}_D^{(3-k)} \right) \right)' \hat{\Omega}_{(k)}^{-1} \left(\frac{1}{G_k} \sum_{g\in\mathcal{G}_k} \hat{w}^{(k)} u_B^{DiD} \left(Z_g; \tilde{\lambda}, \hat{\mu}_D^{(3-k)} \right) \right)$$

   is computed, with

$$\hat{w}^{(k)} = g(s;\hat{\Sigma}^{(3-k)},\tilde{\lambda},\hat{\mu}_D^{(3-k)},\hat{\mu}_0^{(3-k)},\hat{\mu}_1^{(3-k)})$$

   where

$$g(s;\Sigma,\lambda,p_D,p_0,p_1) = \left(\frac{\partial u^{DiD}}{\partial\lambda}(s;\lambda)\right)' \left(\Sigma(s;\lambda,p_D) + p_D(s)(1-p_D(s))\tilde{\alpha}(s;p_D,p_0,p_1)\tilde{\alpha}(s;p_D,p_0,p_1)'\right)^{-1}$$

   and

$$\tilde{\alpha}(s;p_D,p_0,p_1) = \frac{p_1(s)}{p_D(s)} + \frac{p_0(s)}{1-p_D(s)}$$

   and

$$\hat{\Omega}_{(k)} = \frac{1}{G_k}\sum_{g\in\mathcal{G}_k} g(s_g;\hat{\Sigma}^{(3-k)},\tilde{\lambda},\hat{\mu}_D^{(3-k)},\hat{\mu}_0^{(3-k)},\hat{\mu}_1^{(3-k)}) \left(\frac{\partial u^{DiD}}{\partial\lambda}(s_g;\tilde{\lambda})\right)$$

8. The final estimator is

$$\hat{\lambda}^* = \frac{G_1}{G_1+G_2}\hat{\lambda}^{(1)} + \frac{G_1}{G_1+G_2}\hat{\lambda}^{(2)}$$

We make the following technical assumptions. For any function $f(.) = (f_1(.), \dots, f_J(.))$ with $j \geq 1$, any $\hat{f}^{(1)}(.), \hat{f}^{(2)}(.)$ that are estimators of $f(.)$ based on the subsample $\mathcal{G}_1$ and $\mathcal{G}_2$ respectively and $k \in \{1,2\}$, we define

$$\|\hat{f} - f\|_{2,\mathcal{G}_k} := \left(\frac{1}{G_k} \sum_{g \in \mathcal{G}_k} \sum_{j=1}^{J} (\hat{f}_j^{(3-k)}(s_g) - f_j(s_g))^2 \right)^{1/2}, \ \|\hat{f} - f\|_{k,\infty} := \sup_{s \in [0,1], j \in \{1,\dots,J\}} \left| \hat{f}_j^{(3-k)}(s) - f_j(s) \right|$$

Assumption 5′ (Technical Conditions)

1. *$E\left[|\Delta \bar{y}^E|\right] < +\infty$, $E\left[|\Delta \bar{y}^N|\right] < +\infty$*
2. *$s \mapsto E\left[|\Delta \bar{y}^E| \, |s\right]$ and $s \mapsto E\left[|\Delta \bar{y}^N| \, |s\right]$ are bounded on $\mathcal{S}$*
3. *For all $k \in \{1,2\}$,*
   - *(a)* $\hat{\Omega}_{(k)} \xrightarrow{p} \Omega_0$
   - *(b)* $\|\hat{\mu}_D - \mu_D\|_{k,\infty} = o_P(1)$, $\|\hat{\mu}_0 - \mu_0\|_{2,\mathcal{G}_k} = o_P(1)$, $\|\hat{\mu}_1 - \mu_1\|_{2,\mathcal{G}_k} = o_P(1)$
   - *(c)* $\sqrt{G}\|\hat{\mu}_0 - \mu_0\|_{2,\mathcal{G}_k} \times \|\hat{\mu}_0 - \mu_D\|_{2,\mathcal{G}_k} = o_P(1)$, $\sqrt{G}\|\hat{\mu}_1 - \mu_1\|_{2,\mathcal{G}_k} \times \|\hat{\mu}_D - \mu_D\|_{2,\mathcal{G}_k} = o_P(1)$

Note that as all nuisance functions are conditional expectations with respect to a single variable, the convergence-rate requirements are satisfied by classical nonparametric estimators.

Proposition 5 (Optimal GMM estimator in the natural experiment setting)
Let $(Z_g)_{g=1,\dots,G}$ be an i.i.d sample of G groups, where $Z_g = (\Delta \bar{y}_g^E, \Delta \bar{y}_g^N, D_{2g}, s_g)$. Supposing that

1. *Assumptions 1′, 2′, 3′, 4′ and 5′ hold and identification 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 $K \in (0,1)$*

Then $\hat{\lambda}^*$ is a consistent estimator of λ_0 whose asymptotic distribution is

$$\sqrt{G}(\hat{\lambda}^* - \lambda_0) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{E}\left[\left(\frac{\partial}{\partial \lambda} u^{DiD}(s;\lambda_0)\right)' \Sigma_0^{DiD}(s)^{-1} \left(\frac{\partial}{\partial \lambda} u^{DiD}(s;\lambda_0)\right)\right]^{-1}\right)$$

with $\lambda_0 := (\delta_0, \theta_0^w, \theta_0^b)$ and

$$\Sigma_0^{DiD}(s) := E\left[\left\{u_B^{DiD}(Z;\lambda_0,\mu_D) + \alpha(s;\mu_D,\mu_0,\mu_1)\right\}\left\{u_B^{DiD}(Z;\lambda_0,\mu_D) + \alpha(s;\mu_D,\mu_0,\mu_1)\right\}' \Big| s\right]$$

Proof. See Appendix □

Despite a non-parametric first stage, our estimator converges at parametric rate even in the natural experiment design. Note that this estimation procedure can be easily adapted to allow for the inclusion of covariates in the model, as explained in Appendix B.

5 Extension to additional dimensions of heterogeneity

Our baseline model allows for one dimension of heterogeneity: "within" and "between" eligibles and non-eligibles. In this section, we show that our methodology can be easily extended to more general models allowing for additional dimensions of heterogeneity.

5.1 Identity-specific "within" and "between" influences

Our baseline model imposes that θ_0^w and θ_0^b are homogeneous for eligible and non-eligible units. Yet, one may want to relax this assumption in two ways. First, we would like to allow for different "within" peer effects among eligible and among non-eligible individuals to account for the possibility that the strength of social interactions might be identity-specific. For instance, members with identity A are very conformists, while members with identity B are not. Second, and perhaps most importantly, we would like to allow for asymmetric "between" influences between eligible and non-eligible individuals. This is important to study contexts with a social hierarchy, in which individuals do not interact in a symmetric way. For instance, members with a lower status want to imitate members with a higher status, while members with a higher status are not influenced by, or potentially want to distinguish themselves from, members with a lower status.
We thus consider the more flexible model with two distinct "within" parameters, $\theta_0^{w,E}$ and $\theta_0^{w,N}$, and two distinct "between" parameters, $\theta_0^{b,E,N}$ from non-eligibles to eligibles, and $\theta_0^{b,N,E}$ from eligibles to non-eligibles, defined as follows.

Assumption 1∗ (Linear-in-means model - 5 parameters)
Within each group $g \in \{1, \dots, G\}$ of the i.i.d sample, the outcome of individual i is defined as

$$y_{ig} = E_{ig} y_{ig}^E + (1 - E_{ig}) y_{ig}^N$$

with

$$y_{ig}^E = \theta_0^{w,E} s_g \bar{y}_g^E + \theta_0^{b,E,N} (1 - s_g) \bar{y}_g^N + \delta_0 D_{ig} + e_{ig}^E \tag{29}$$

$$y_{ig}^N = \theta_0^{b,N,E} s_g \bar{y}_g^E + \theta_0^{w,N} (1 - s_g) \bar{y}_g^N + e_{ig}^N \tag{30}$$

In Appendix C, we show how to that our identification and estimation strategies can be adapted to recover five parameters instead of three in the baseline model. Identification is achieved as long as we observe at least three values of s (instead of two values in the baseline model). In theory, we demonstrate that the new GMM estimator is asymptotically normal. However, in practice, our simulations show that the asymptotic variance is very large. The main challenge for the estimator is to separately identify the "between" effect of non-eligibles on eligibles from the direct effect of the treatment.

5.2 Identity not restricted to treatment eligibility

Our baseline model restrict the identity to the dimension that matters for eligibility. In some contexts, this dimension may not be interesting. Instead, we may want to study another dimension. For instance, in the case of Progresa where eligibility is defined by household income, the researcher might be more interested in the social interactions between boys and girls.
We build on our baseline model to study any other dimension of identity which is different from eligibility. We rewrite the model as follows.

Assumption 1† (Linear-in-means model - orthogonal identity)
For a group $g \in \{1, \dots, G\}$ of the i.i.d sample, the outcome of individual i is defined as follow

$$y_{ig} = y_{ig}^M M_{ig} + y_{ig}^F (1 - M_{ig})$$

$$y_{ig}^M = \theta_0^{w,M} s_g^M \bar{y}_g^M + \theta_0^{b,M,F} (1 - s_g^M) \bar{y}_g^F + \delta_0 D_{ig} E_{ig} + e_{ig}^M \tag{31}$$

$$y_{ig}^F = \theta_0^{b,F,M} s_g^M \bar{y}_g^M + \theta_0^{w,F} (1 - s_g^M) \bar{y}_g^F + \delta_0 D_{ig} E_{ig} + e_{ig}^F \tag{32}$$

with $M_{ig} \in \{0,1\}$ equals 1 if the individual is from the "male" identity and $E_{ig} \in \{0,1\}$ equals 1 if the individual is eligible to the treatment. $s_g^M := \overline{M}_g$, i.e. it is the share of "male" individuals in the peer group.

In this design, both "male" and "female" individuals can be treated. $\theta_0^{w,M}$ and $\theta_0^{w,F}$ capture how males influence each other, and how females influence each other, respectively. $\theta_0^{b,M,F}$ captures how females influence males and $\theta_0^{b,F,M}$ captures how males influence females. We use the gender dimension as an example for clarity of exposition, but the same logic applies to any dimension observed by the researcher.
In Appendix E, we show that our identification and estimation strategies can be adapted to recover the five parameters of interest. The main adjustment resides in the need to compare treated and control groups that have the same proportion of "male", eligible "male" and eligible "female" individuals. The data requirements are therefore stronger than in the baseline model.

6 Application: simulated data

In this section, we illustrate how researchers can use our methodology to study complex group dynamics that go beyond conformism. We proceed in three steps. First, we define a Data Generating Process (DGP) corresponding to the specific group dynamic we want to study and we simulate the data. Second, we estimate the endogenous peer effect parameters using our estimation procedure and we calculate the treatment effects, $\tau^E(s)$ and $\tau^N(s)$. Third, we compare our results with the standard estimation procedure that does not allow for heterogeneous parameters.
We are interested in three types of group dynamics.
Conformism within identities. Group members are only influenced by members with the same identity and they want to imitate each other. They are not influenced by members with another identity. We suppose that $\theta^w = 0.85$ and $\theta^b = 0$.
Polarization. Group members are only influenced by members with the other identity and they want to differentiate themselves by adopting the opposite behavior. They are not influenced by members with the same identity. We suppose that $\theta^w = 0$ and $\theta^b = -0.85$.
Role model. Group members with a given identity, the *followers*, want to imitate the other identity, the *influencers*, while the latter group is not affected by anyone. Here we distinguish two situations. When the **eligibles are influencers**, we suppose that $\theta^{w,E} = \theta^{w,N} = \theta^{b,E,N} = 0$ and $\theta^{b,N,E} = 0.85$. When the **eligibles are followers**, we suppose that $\theta^{w,E} = \theta^{w,N} = \theta^{b,N,E} = 0$ and $\theta^{b,E,N} = 0.85$.
Figure 4 represents the estimated average total treatment effects within the eligible population and non-eligible population for each simulation: (i) conformism within identities; (ii) polarization; (iii) role model with eligibles as influencers; (iv) role model with eligibles as followers. The graphs on the left plot the estimations following our procedure, allowing for heterogeneous parameters. The graphs on the right plot the estimations using a 2SLS specification in which the average outcome within the group ($\bar{y}_g$) is instrumented using the share of eligible individuals within the group multiplied by the treatment indicator (i.e. $s_G D_g$). This is the standard procedure implemented when researchers want to estimate the endogenous peer effect parameter in a LIM model with homogeneous θ.
Starting with the case of conformism within identities, we find a strongly convex effect for the eligibles: the population multiplier increases exponentially as the size of the eligible sub-population increases and social interactions between treated members are more frequent. To the contrary, there is no effect at all on the non-eligibles. They only respond to changes in behaviors of non-eligible members and none of these members is affected by the policy. The average effect in the

population is even more convex that $\tau^E(s)$ because of the composition effect (the weight on $\tau^E(s)$ increases with s). If we had ignored the heterogeneity, we would have estimated a linear effect, with a similar slope for eligibles and non-eligibles. We would have overestimated the average effect in groups with a small-to-medium share of eligibles and underestimated it in groups with a high share of eligibles. Another mistake is that we would have concluded that non-eligibles responded to the policy, while they did not. Finally, we would have overestimated the direct treatment effect.
Turning to the case of polarization, we find an inverted-U shape for the eligibles: the population multiplier is maximal when $s = 0.5$ (the group is perfectly split into both identities) and relevant social interactions between members of different identities are the most frequent. The non-eligibles barely respond when $s \to 0$ because there are too few eligibles in the group; when s increases, they respond in the opposite way as the eligibles (in this example, in a negative way because $\delta > 0$). When $s \to 1$, the non-eligibles respond very strongly but they are not numerous enough to influence the eligibles so the feedback loop quickly stops. The model without heterogeneity leads to several wrong conclusions: we miss the inverted-U shape for eligibles, we strongly underestimate the magnitude of the backlash among non-eligibles, we overestimate the direct effect, and we overestimate the average effect in all groups.
Finally, in the case of role model, the results depend on who is eligible. When the policy targets the influencers, we find a concave average effect. There is no population multiplier for the eligibles (i.e. the influencers) because they are not affected by peer effects. For the non-eligibles (i.e. the followers), there is a multiplier which increases linearly with the fraction of influencers. The homogeneous model, by imposing the same parameter on both groups, underestimates the response of followers, underestimates the direct effect, and underestimates the average effect as long as $s < 0.75$. In contrast, when the policy targets the followers, the homogeneous model performs well because there is no social multiplier. The estimated homogeneous θ is equal to zero and all the relevant θs are also equal to zero. The average effect is simply the weighted average of a null effect among non-eligibles and the direct effect among eligibles. In this example, neither our model nor the homogeneous model can detect the non-null peer effect from influencers to followers because the behavior of influencers is not affected by the policy; in other words, the initial shock does not propagate.

7 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 kids from eligible households 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. Note that their estimates of θ and δ are barely significant at the 5% level. This is a first hint that we will struggle to have precise estimates when we allow for heterogeneity in this application.

We use the same data and implement our procedure allowing for heterogeneous θ. Figure 5 plots the distribution of s, by treatment status, supporting the validity of the common support condition. Table 1 provides some descriptive statistics on peer groups, that support the correct randomization of the treatment. In total, we observe 1,798 peer groups, in 496 villages. Based on our simulations, the number of observations is enough to provide us consistent estimates of the parameters and of their associated standard errors in our baseline scenario. However, it is too small to add several pre-program control variables, as in Lalive and Cattaneo (2009), or to consider a model with four distinct parameters for the endogenous peer effects.

Estimates are reported in Table 2. In column (2), we include the number of kids in the peer group as a control variable; this is our preferred specification, as it allows us to compare groups of the same size. We find a "between" peer effect of $\hat{\theta}^b = 0.31$, relatively close to, and not significantly different from, the homogeneous $\hat{\theta}$ estimated by Lalive and Cattaneo (2009). However, our estimate is noisy and we cannot reject the null hypothesis that the poor children have no influence on non-poor children. Turning to "within" peer effects, we find that $\hat{\theta}^w = 0.24$. Once again, we cannot reject the null that students are not influenced by their peers' actions and $\hat{\theta}^w$ belongs to the confidence interval of $\hat{\theta}$. To interpret the magnitude of the point estimate, suppose that one poor child drops out of school in a group of 10 children (the average number of peers in our sample). In the rest of the group, the likelihood that a child attends school decreases by $\frac{1}{10} \times 0.24 = 2.4$ percentage points among the poor. [3]. 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 $\hat{\delta} = 0.048$. Our estimate belongs to the confidence interval estimated by Lalive and Cattaneo (2009).

In terms of policy evaluation, Figure 8 plots the estimates of τ^E and τ^N as a function of s based on the point estimates of the coefficients in Table 2 . In groups with a very low share of poor children ($s \to 0$), Progresa raises school attendance by 4.7 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 slightly concave way. In groups with a very high share of poor children ($s \to 1$), Progresa raises school attendance by 6.2 percentage points for poor children and by 2 percentage points for the (very rare) non-poor children. The population multipliers are equal to $\frac{1}{1-\hat{\theta}^w} = 1.32$ when $s = 1$. The dotted line plots the total effect

[3] 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 $\theta_0^w s \frac{-1}{10s}$ and the change in y_i^N is equal to $\theta_0^b s \frac{-1}{10s}$.

of the policy: $s\tau^E + (1-s)\tau^N$. Magnitudes range from 0 when $s = 0$ to 6.2 percentage points when $s = 1$. The total effect for the average s in our sample (roughly 70%) is equal to 5.9 p.p. for eligible children and 1.4p.p. for non-eligible children. We conclude that the policy effect is mildly magnified by peer effects. In the case of Progresa, the total effect has an almost linear shape because $\hat{\theta}^w$ and $\hat{\theta}^b$ are not very different. Unlike cases discussed in section 6, the homogeneity assumption does not lead to substantial biases in this application.

8 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 subpopulations of eligibles and non-eligibles; (ii) the endogenous peer effects "within" each subpopulation; (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.

Arduini, Tiziano, Eleonora Patacchini, and Edoardo Rainone (2020). “Treatment effects with heterogeneous externalities”. In: *Journal of Business & Economic Statistics* 38(4), pp. 826–838.

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.

Comola, Marguerita, Dieye Rokhaya, and Bernard Fortin (2023). “Heterogeneous peer effects and gender-based interventions for teenage obesity”. In.

Gupta, Abhimanyu and Peter M Robinson (2015). “Inference on higher-order spatial autoregressive models with increasingly many parameters”. In: *Journal of Econometrics* 186(1), pp. 19–31.

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`.

Lee, Lung-fei and Xiaodong Liu (2010). “Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances”. In: *Econometric Theory* 26(1), pp. 187–230.

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

Masten, Matthew A (2018). “Random coefficients on endogenous variables in simultaneous equations models”. In: *The Review of Economic Studies* 85(2), pp. 1193–1250.

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.

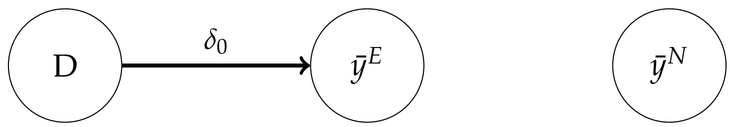
Wooldridge, Jeffrey M (2010). *Econometric analysis of cross section and panel data*. MIT press.

Wu, Yujia et al. (2022). “Inward and outward network influence analysis”. In: *Journal of Business & Economic Statistics* 40(4), pp. 1617–1628.

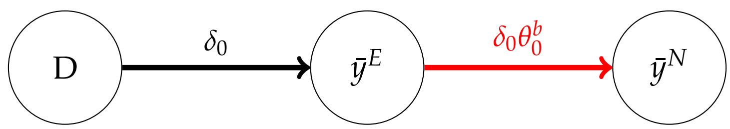
Figures and Tables

Figure 1: Decomposition of Total Treatment Effects

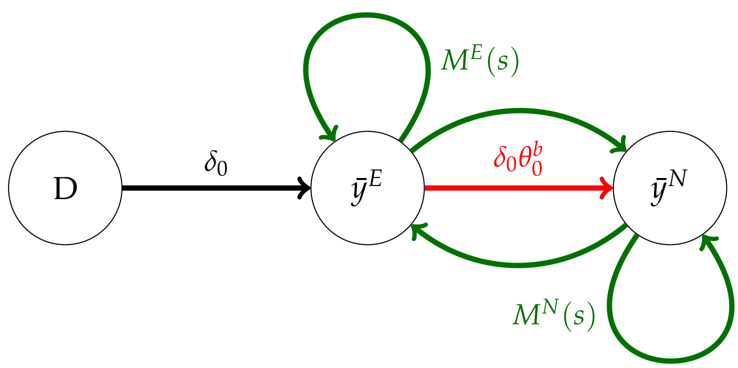
Step 1 : direct treatment effect on eligible individuals



Step 2 : indirect treatment effect of non-eligible individuals



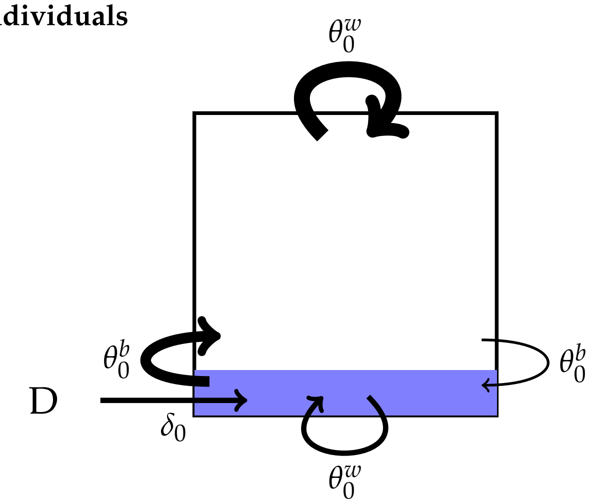
Step 3 : amplification of initial effects through social interactions



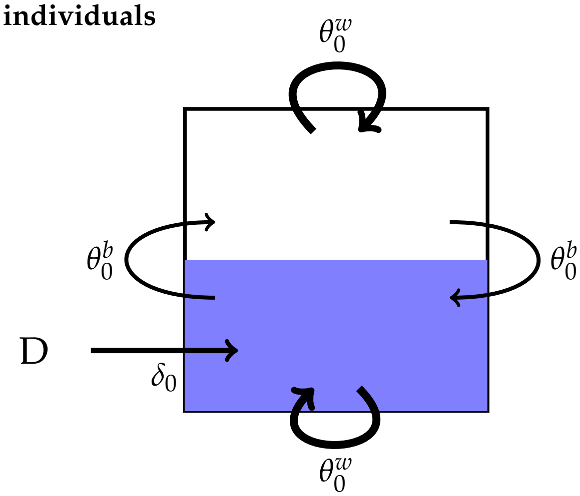
Note: The figure illustrates how the initial shock to eligible individuals propagates through peer effects. First, the eligible individuals respond to the treatment: y^E changes by δ_0 (direct effect). Second, the non-eligible individuals respond to the change in y^E: y^N changes by $\delta_0\theta_0^b$ (indirect effect). Third, these initial changes lead to a cascade process, in which eligible and non-eligible individuals keep responding to each other (in proportion to θ_0^b) and also respond to other members of their own sub-group (in proportion to θ_0^w).

Figure 2: Population Multipliers by Population Structure

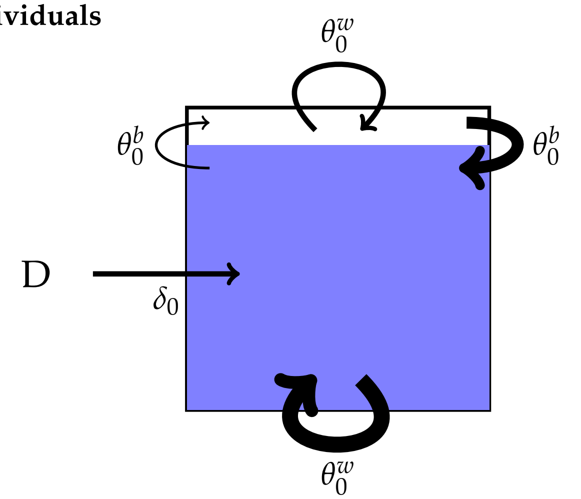
Small share of eligible individuals



Medium share of eligible individuals

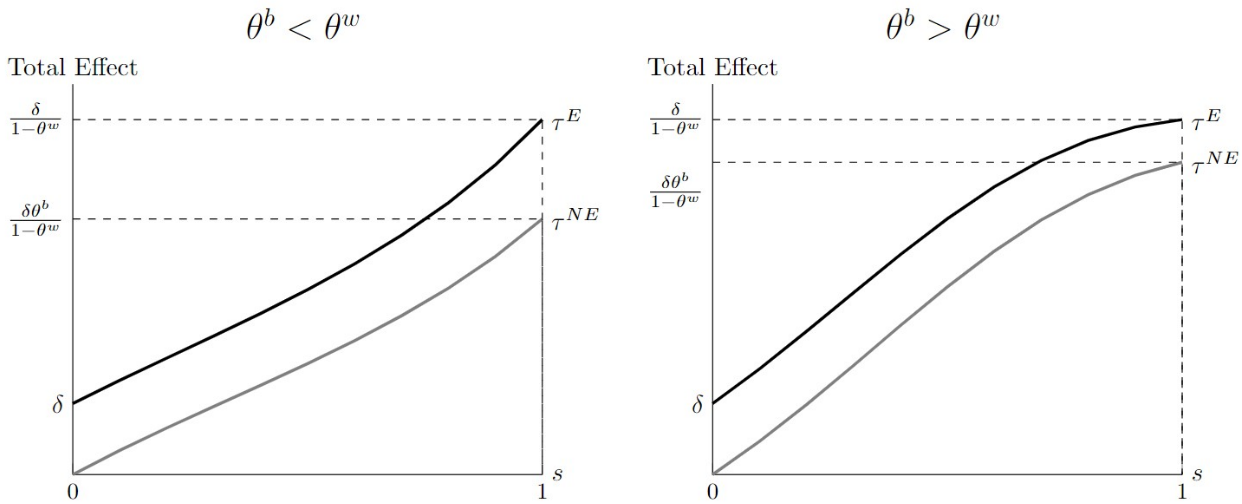


High share of eligible individuals



Note: The figure illustrates how the process intensity varies with the share of eligible individuals, s, represented in blue. First, the initial shock mechanically affects more individuals in the group when s is higher. Second, the propagation depends on s: arrows originating from the blue area are proportional to s while arrows originating from the white area are proportional to $(1-s)$. In this illustration, we consider the case when $|\theta_0^w| > |\theta_0^b|$ and we vary the size of the arrows accordingly (arrows within blue area and within white area are thicker than arrows between blue and white areas).

Figure 3: Treatment effects on eligibles and non-eligibles



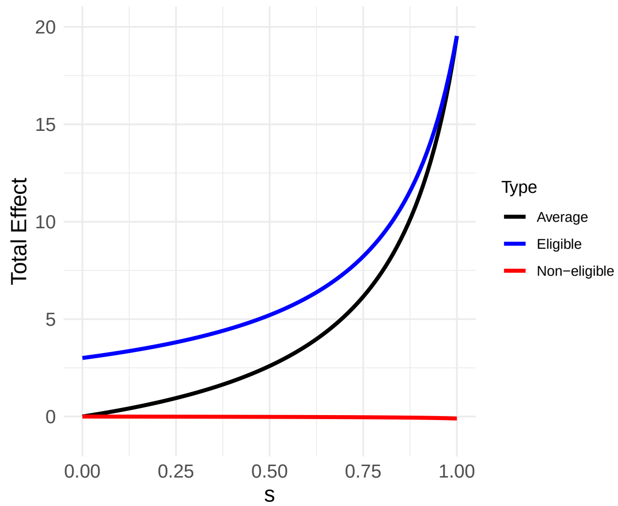
Note: The figure plots the total effect on the eligibles $\tau^E(s)$, in black, and the total effect on the non-eligibles $\tau^N(s)$, in gray, as a function of s. The graph on the left illustrates the case when $\theta^b < \theta^w$ and the graph on the right illustrates the case when $\theta^b > \theta^w$.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

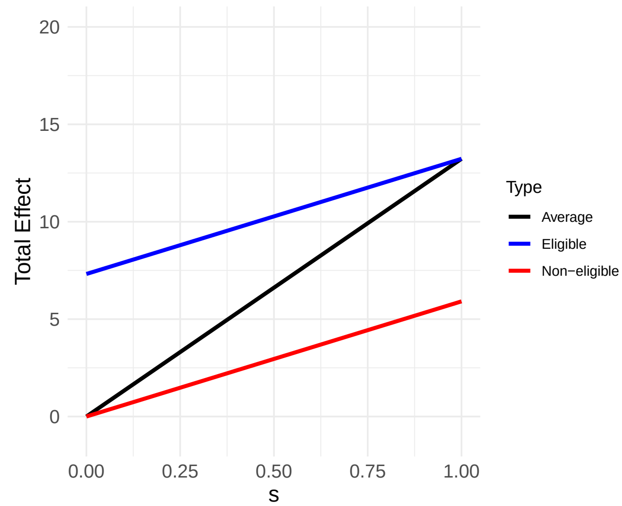
$$M^E(s) = 1 + sP^E(s), M^N(s) = sP^N(s), P^E(s) = \frac{\theta^w - (1-s)\left[(\theta^w)^2 - (\theta^b)^2\right]}{1-\theta^w + s(1-s)\left[(\theta^w)^2 - (\theta^b)^2\right]}, P^N(s) = \frac{1}{1-\theta^w + s(1-s)\left[(\theta^w)^2 - (\theta^b)^2\right]}$$

Figure 4: Results - simulated data

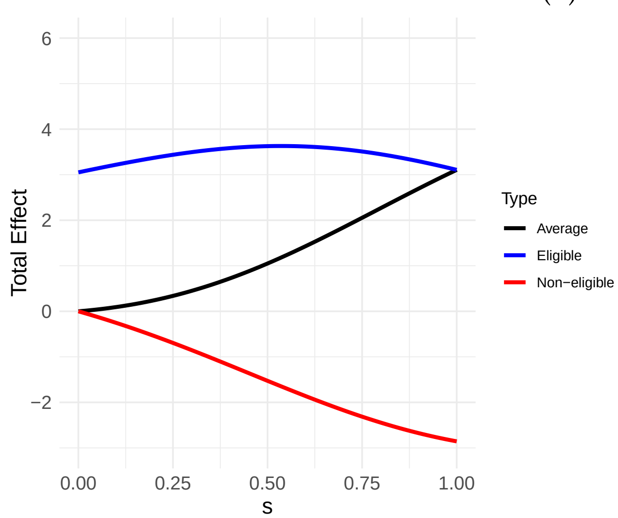
(a) Heterogeneous parameters | (b) Homogeneous parameter

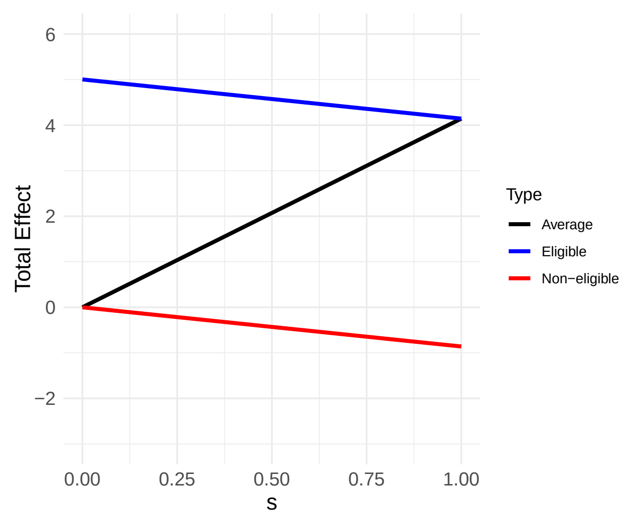
(i) Conformism within identities





(ii) Polarization



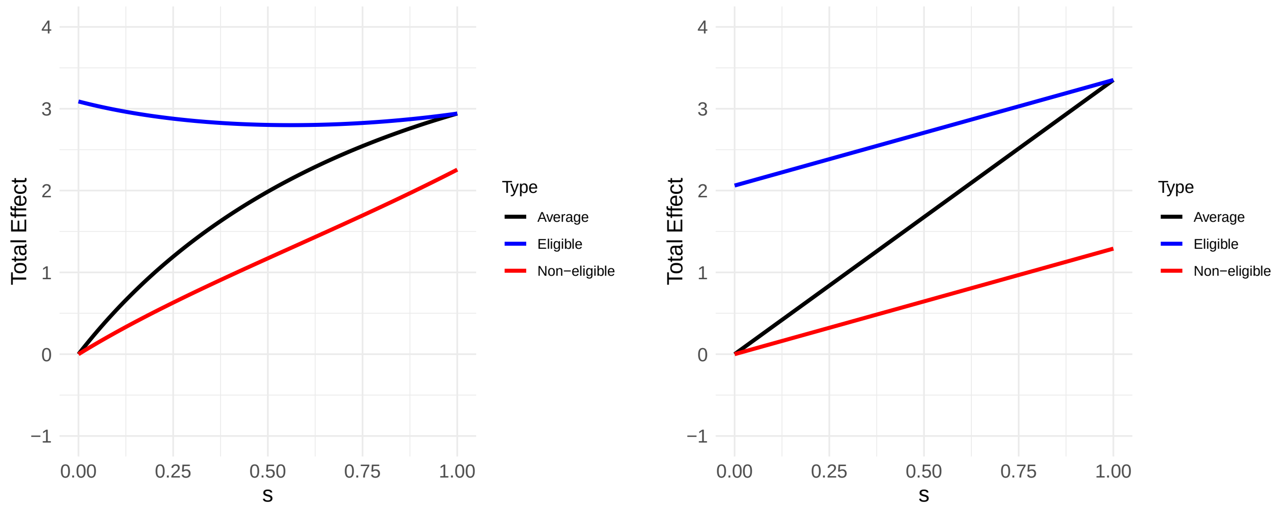


Note: The figure continues on the next page

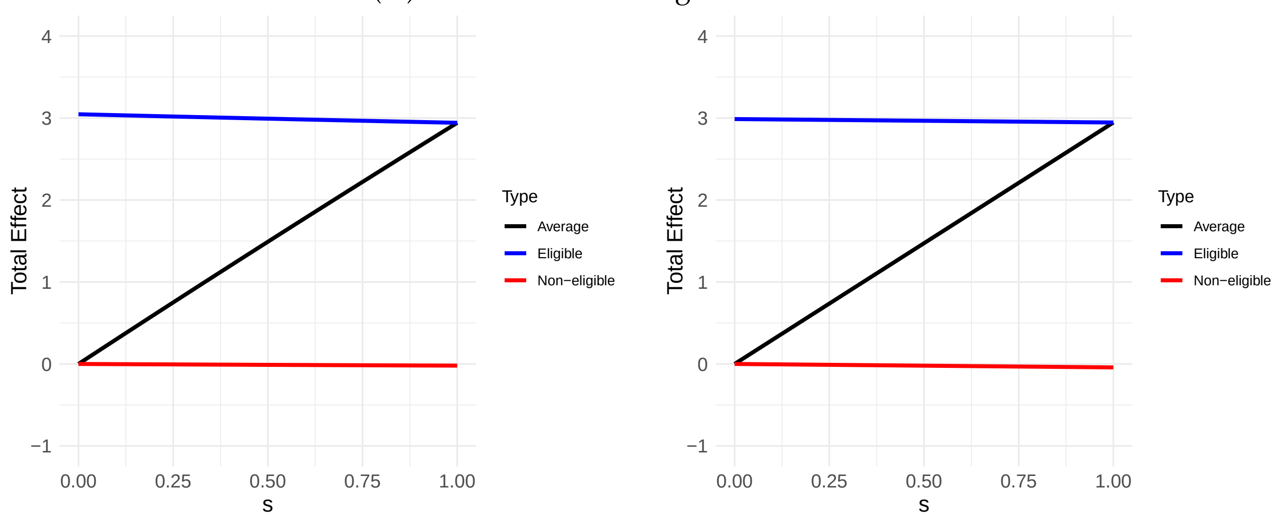
Figure 4 (continued): Results - simulated data

(a) Heterogeneous parameters (b) Homogeneous parameter

(iii) Role model with eligibles as influencers



(iv) Role model with eligibles as followers



Note: The figure plots the estimated total effect on the eligibles $\tau^E(s)$, in blue, and the estimated total effect on the non-eligibles $\tau^N(s)$, in red, as a function of s. The black line is the estimated average effect in the population given by $s\tau^E(s) + (1-s)\tau^N(s)$. The data is simulated following the DGP described in section 6 for four types of group dynamics: (i) $\theta^w = 0.85$ and $\theta^b = 0$; (ii) $\theta^w = 0$ and $\theta^b = -0.85$; (iii) $\theta^{w,E} = \theta^{w,N} = \theta^{b,E,N} = 0$ and $\theta^{b,N,E} = 0.85$; and (iv) $\theta^{w,E} = \theta^{w,N} = \theta^{b,N,E} = 0$ and $\theta^{b,E,N} = 0.85$. The graphs on the left (panel a) present the results using our estimation procedure. The graphs on the right (panel b) present the results using the standard procedure that does not allow for heterogeneous parameters.

Figure 5: Progresa - Common Support Assumption

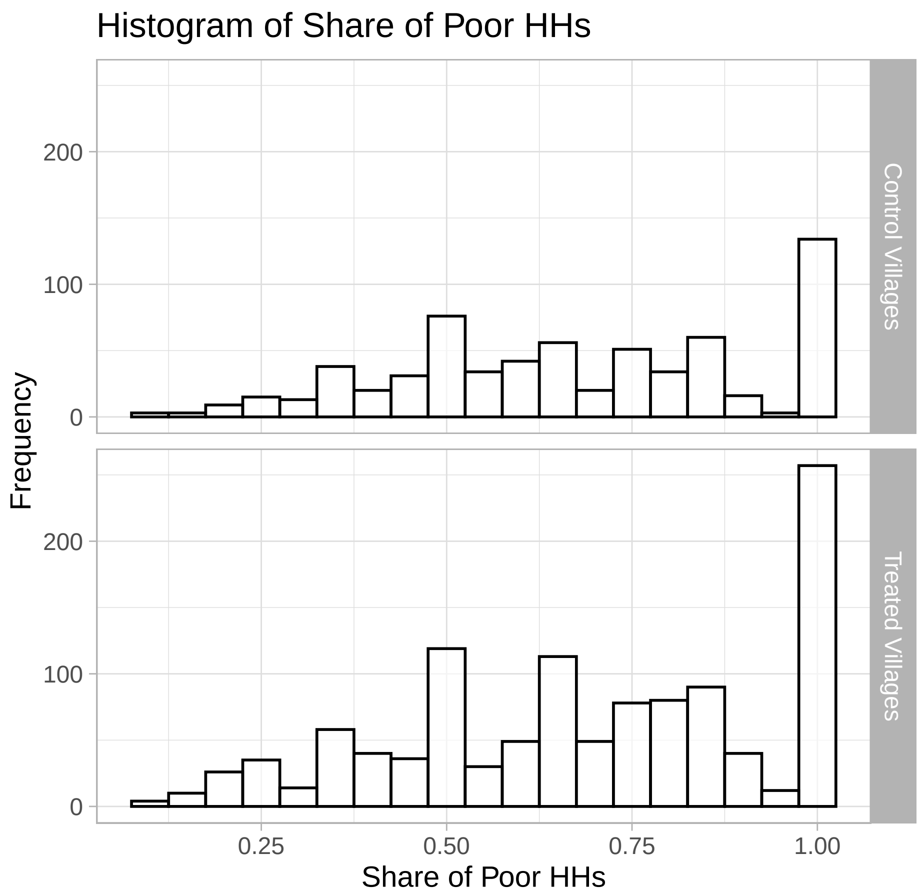
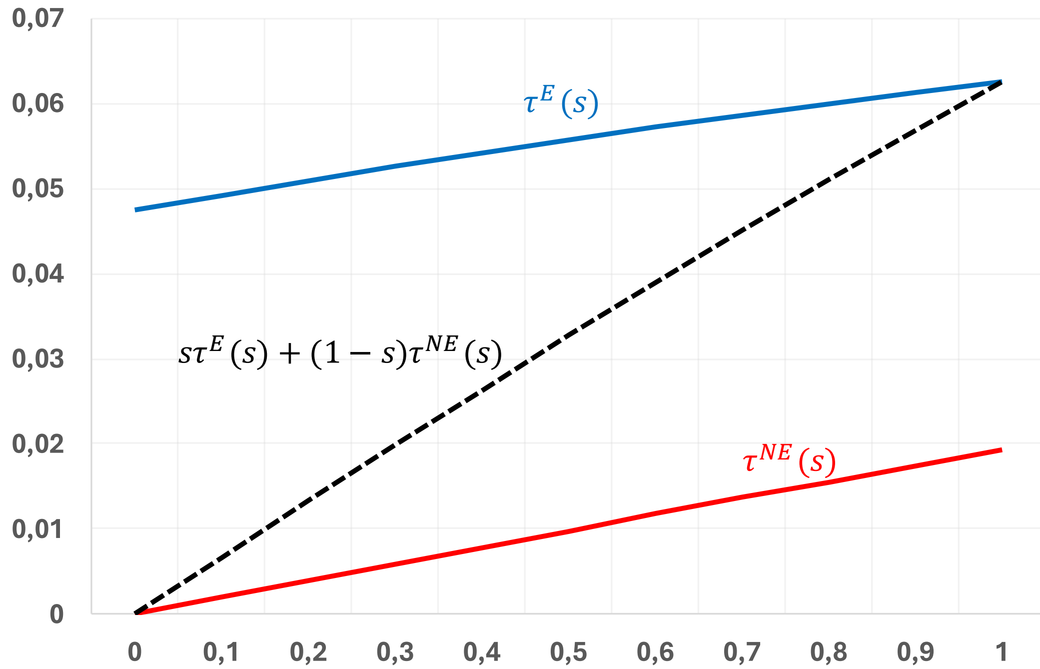


Figure 6: Results - Progresa



Note: The figure shows the total effects based on the Progresa estimates reported in Table 2, column (2).The graph plots the total effect on the eligibles $\tau^E(s)$, in blue, and the total effect on the non-eligibles $\tau^N(s)$, in red, as a function of s. The dotted line plots the average effect of the policy in the population.

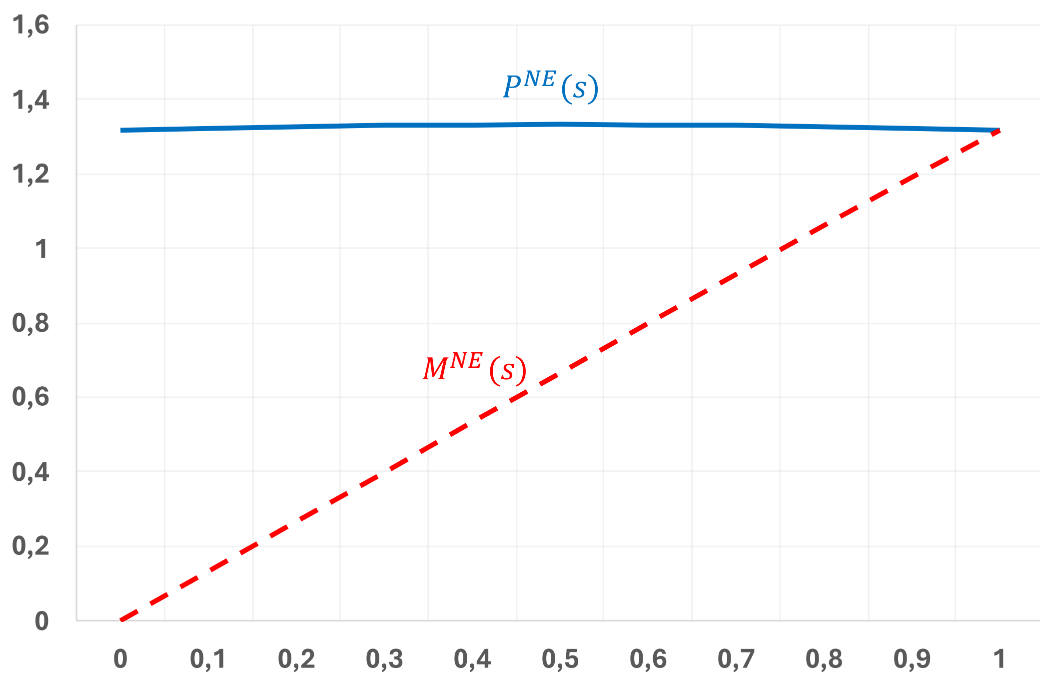
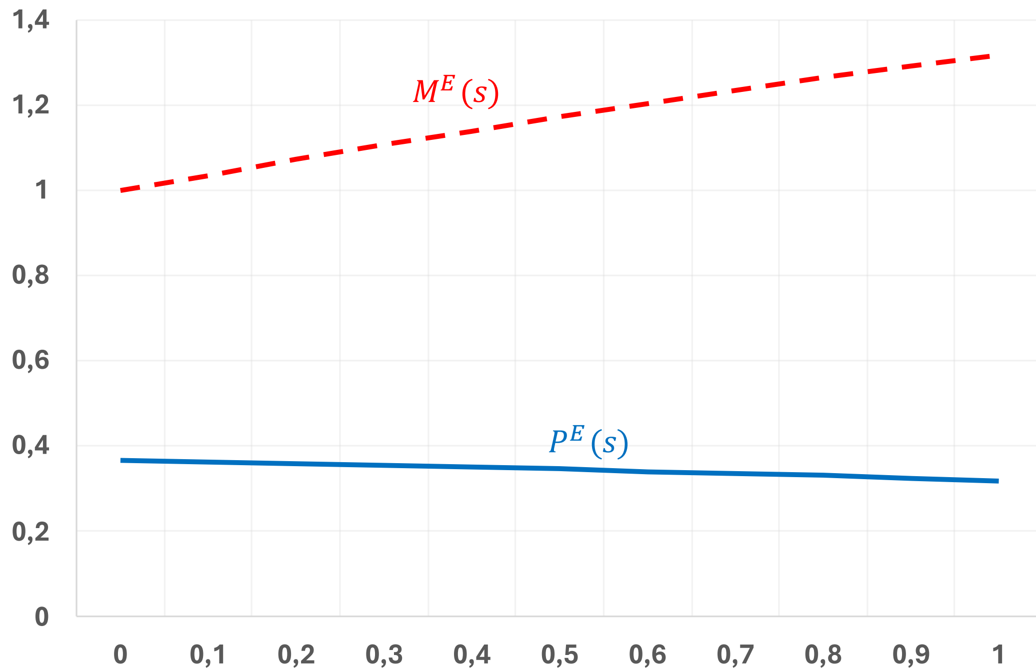


Figure 6: Results - Progresa

Figure 7: Population multipliers for eligibles



Note: The figure shows the total effects and population multipliers based on the Progresa estimates reported in Table 2. Graph (c) plots the population multiplier $M(s)$, in dashed lines, and the propagation function $P(s)$, in solid lines. See formulas in the note below Figure 3.

Table 1: Descriptive Statistics - Progresa

	Control Peer Groups	Treated Peer Groups
N	658	1140
Average Peer Group Size	8.72	8.31
	(6.08)	(5.74)
Share of Poor	0.68	0.69
	(0.24)	(0.24)
Share Grade 4	0.25	0.25
	(0.43)	(0.43)
Share Grade 5	0.24	0.24
	(0.43)	(0.43)
Share Grade 6	0.25	0.25
	(0.43)	(0.43)
Share of Males	0.51	0.51
	(0.23)	(0.23)

Table 2: Estimation results - Progresa

	(1)	(2)
$\hat{\theta}^b$	0.463	0.308
	(0.366)	(0.225)
$\hat{\theta}^w$	0.352	0.241
	(0.592)	(0.635)
$\hat{\delta}$	0.037	0.048
	(0.027)	(0.031)
Control	No	Yes
# peer groups	1,798	1,798
# villages	496	496

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. Standard errors are clustered at the village level. Specification (1) is the basic specification while in Specification (2), we control for the number of kids in the peer group.

Appendix A Proofs

Appendix A.1 Reduced forms in the randomized experiment design

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

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

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

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

Rearranging terms, we get

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

Hence,

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

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

$$\begin{aligned}
\bar{y}_g^E =& \frac{1-\theta_0^w(1-s_g)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2-(\theta_0^b)^2\right]}\bar{e}_g^E \\
&+ \frac{\theta_0^b(1-s_g)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2-(\theta_0^b)^2\right]}\bar{e}_g^N \\
&+ \frac{\delta_0(1-\theta_0^w(1-s_g))}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2-(\theta_0^b)^2\right]}D_g \\
=& \frac{\theta_0^b(1-s_g)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2-(\theta_0^b)^2\right]}\bar{e}_g^N \\
&+ \frac{1-\theta_0^w(1-s_g)}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2-(\theta_0^b)^2\right]}\bar{e}_g^E \\
&+ \delta_0\left(1 + s_g.\frac{\theta_0^w - (1-s_g)\left[(\theta_0^w)^2-(\theta_0^b)^2\right]}{1-\theta_0^w + s_g(1-s_g)\left[(\theta_0^w)^2-(\theta_0^b)^2\right]}\right)D_g
\end{aligned} \tag{33}$$

Appendix A.2 Proposition 2

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

$$\tau^k(s) = \mathrm{E}\left[\bar{y}_g^k | D_g = 1, s_g = s\right] - \mathrm{E}\left[\bar{y}_g^k | D_g = 0, s_g = s\right]$$

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

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

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

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,

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

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

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

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

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

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

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

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

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

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

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

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

Now,

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

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

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

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

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

$$\tau^E(s_1) = \frac{\delta_0\left(1-\theta_0^w(1-s_1)\right)}{1-\theta_0^w+s_1(1-s_1)\left[(\theta_0^w)^2-(\theta_0^b)^2\right]} \tag{34}$$

$$\tau^N(s_1) = \frac{\delta_0\theta_0^b s_1}{1-\theta_0^w+s_1(1-s_1)\left[(\theta_0^w)^2-(\theta_0^b)^2\right]} \tag{35}$$

$$\tau^E(s_2) = \frac{\delta_0\left(1-\theta_0^w(1-s_2)\right)}{1-\theta_0^w+s_2(1-s_2)\left[(\theta_0^w)^2-(\theta_0^b)^2\right]} \tag{36}$$

$$\tau^N(s_2) = \frac{\delta_0\theta_0^b s_2}{1-\theta_0^w+s_2(1-s_2)\left[(\theta_0^w)^2-(\theta_0^b)^2\right]} \tag{37}$$

where equations (34) and (36) are written in the reduced form, as derived in (33). 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 (35) by (34) and (37) by (36), we get

$$\frac{\tau^N(s_1)}{\tau^E(s_1)} = \frac{\theta_0^b s_1}{1 - \theta_0^w(1-s_1)} \tag{38}$$

$$\frac{\tau^N(s_2)}{\tau^E(s_2)} = \frac{\theta_0^b s_2}{1 - \theta_0^w(1-s_2)} \tag{39}$$

which can be expressed as new moment conditions

$$(1 - \theta_0^w(1-s_1))\tau^N(s_1) - \theta_0^b s_1 \tau^E(s_1) = 0 \tag{40}$$

$$(1 - \theta_0^w(1-s_2))\tau^N(s_2) - \theta_0^b s_2 \tau^E(s_2) = 0 \tag{41}$$

Solving the system composed of (40) and (41), provided $s_1 \neq s_2$, leads to

$$\begin{cases} \theta_0^b = & \dfrac{(1-\theta_0^w(1-s_1))\tau^N(s_1)}{s_1\tau^E(s_1)} \\[2ex] \theta_0^w = & \dfrac{\tau^N(s_2) - s_2\tau^E(s_2)\theta_0^b}{\tau^N(s_2)(1-s_2)} \end{cases}$$

$$\implies \begin{cases} \theta_0^b = & \dfrac{(1-\theta_0^w(1-s_1))\tau^N(s_1)}{s_1\tau^E(s_1)} \\[2ex] \theta_0^w = & \dfrac{s_1\tau^E(s_1)\tau^N(s_2) - s_2\tau^E(s_2)\tau^N(s_1)}{s_1\tau^E(s_1)(1-s_2)\tau^N(s_2) - (1-s_1)\tau^N(s_1)s_2\tau^E(s_2)} \end{cases}$$

$$\implies \begin{cases} \theta_0^w = & \dfrac{s_1\tau^E(s_1)\tau^N(s_2) - s_2\tau^E(s_2)\tau^N(s_1)}{s_1(1-s_2)\tau^E(s_1)\tau^N(s_2) - (1-s_1)s_2\tau^N(s_1)\tau^E(s_2)} \\[2ex] \theta_0^b = & \dfrac{\tau^N(s_1)\tau^N(s_2)(s_1-s_2)}{s_1(1-s_2)\tau^E(s_1)\tau^N(s_2) - (1-s_1)s_2\tau^N(s_1)\tau^E(s_2)} \end{cases}$$

Hence, θ_0^b and θ_0^w are identified. As a consequence, δ_0 is also identified based on (34), (36), (35) or (37)

Appendix A.5 Proposition 4

Let

$$\tilde{\lambda} = \arg\min_{\lambda\in\Theta} \left(\frac{1}{G}\sum_{g=1}^{G} \tilde{w}(s_g)u_B^R(Z_g,\lambda)\right)' \left(\frac{1}{G}\sum_{g=1}^{G} \tilde{w}(s_g)u_B^R(Z_g,\lambda)\right)$$

where, for any $s \in [0,1]$, $\tilde{w}(s)$ is a $k \times 2$ matrix of arbitrary functions (e.g. polynomials) of s and $k \geq \dim(\lambda_0)$. The assumptions made in the statement ensure that Theorem 2.5 of Newey and McFadden (1994) applies and $\tilde{\lambda} \xrightarrow{p} \lambda_0$.

We then must show that

$$\hat{\Omega} := \frac{1}{G}\sum_{g=1}^{G}\left(\frac{\partial u^R}{\partial \lambda}(s_g;\tilde{\lambda})\right)'\hat{\Sigma}^{-1}(s_g;\tilde{\lambda})\left(\frac{\partial u^R}{\partial \lambda}(s_g;\tilde{\lambda})\right) \xrightarrow{p} \Omega_0 := \mathrm{E}\left[\left(\frac{\partial u^R}{\partial \lambda}(s;\lambda_0)\right)'\Sigma_0^{-1}(s;\lambda_0)\left(\frac{\partial u^R}{\partial \lambda}(s;\lambda_0)\right)\right]$$

In the following, to shorten notations, we set

$$f(s;\lambda,\Sigma) = \left(\frac{\partial u^R}{\partial \lambda}(s;\lambda)\right)'\Sigma^{-1}(s;\lambda)\left(\frac{\partial u^R}{\partial \lambda}(s;\lambda)\right)$$

Let $\|.\|$ be a 2×2 matrix sub-multiplicative norm. We have that

$$\begin{aligned}\left\|\frac{1}{G}\sum_{g=1}^{G} f(s_g;\tilde{\lambda},\hat{\Sigma}) - \mathrm{E}\left[f(s;\lambda_0,\Sigma_0)\right]\right\| \leq \quad & \left\|\frac{1}{G}\sum_{g=1}^{G} f(s_g;\tilde{\lambda},\hat{\Sigma}) - \frac{1}{G}\sum_{g=1}^{G} f(s_g;\tilde{\lambda},\Sigma_0)\right\| && (42)\\ & + \left\|\frac{1}{G}\sum_{g=1}^{G} f(s_g;\tilde{\lambda},\Sigma_0) - \frac{1}{G}\sum_{g=1}^{G} f(s_g;\lambda_0,\Sigma_0)\right\| && (43)\\ & + \left\|\frac{1}{G}\sum_{g=1}^{G} f(s_g;\lambda_0,\Sigma_0) - \mathrm{E}\left[f(s;\lambda_0,\Sigma_0)\right]\right\| && (44)\end{aligned}$$

Focusing first on the term 42, we have that

$$\begin{aligned}\left\|\frac{1}{G}\sum_{g=1}^{G} f(s_g;\tilde{\lambda},\hat{\Sigma}) - \frac{1}{G}\sum_{g=1}^{G} f(s_g;\tilde{\lambda},\Sigma_0)\right\| \quad & \leq \frac{1}{G}\sum_{g=1}^{G}\|f(s_g;\tilde{\lambda},\hat{\Sigma}) - f(s_g;\tilde{\lambda},\Sigma_0)\|\\ & \leq \frac{1}{G}\sum_{g=1}^{G}\left\|\left(\frac{\partial u^R}{\partial \lambda}(s_g;\tilde{\lambda})\right)'\left(\hat{\Sigma}^{-1}(s_g;\tilde{\lambda}) - \Sigma_0^{-1}(s_g;\tilde{\lambda})\right)\left(\frac{\partial u^R}{\partial \lambda}(s_g;\tilde{\lambda})\right)\right\|\\ & \leq \frac{2}{G}\sum_{g=1}^{G}\left\|\frac{\partial u^R}{\partial \lambda}(s_g;\tilde{\lambda})\right\|^2\left\|\hat{\Sigma}^{-1}(s_g;\tilde{\lambda}) - \Sigma_0^{-1}(s_g;\tilde{\lambda})\right\|\end{aligned}$$

where the last inequality comes from the fact that $\|.\|$ is sub-multiplicative. Then, since $\tilde{\lambda} \xrightarrow{p} \lambda_0$, for G large enough,

$$\left\|\hat{\Sigma}^{-1}(s_g;\tilde{\lambda}) - \Sigma_0^{-1}(s_g;\tilde{\lambda})\right\| \leq \sup_{\lambda\in\mathcal{N}, s\in[0,1]}\left\|\hat{\Sigma}^{-1}(s;\lambda) - \Sigma_0^{-1}(s;\lambda)\right\|$$

Moreover, u^R is by construction twice-differentiable with respect to λ on Θ. By the continuous mapping theorem (CMT),

$$\frac{\partial u^R}{\partial \lambda}(s;\tilde{\lambda}) = \frac{\partial u^R}{\partial \lambda}(s;\lambda_0) + o_P(1)$$

Therefore,

$$\left\|\frac{1}{G}\sum_{g=1}^{G} f(s_g;\tilde{\lambda},\hat{\Sigma}) - \frac{1}{G}\sum_{g=1}^{G} f(s_g;\tilde{\lambda},\Sigma_0)\right\| = O_P(1)\times o_P(1) = o_P(1)$$

Now, focusing on the term (43), we have that

$$\left\|\frac{1}{G}\sum_{g=1}^{G} f(s_g;\tilde{\lambda},\Sigma_0) - \frac{1}{G}\sum_{g=1}^{G} f(s_g;\lambda_0,\Sigma_0)\right\| \leq \frac{1}{G}\sum_{g=1}^{G}\|f(s_g;\tilde{\lambda},\Sigma_0) - f(s_g;\lambda_0,\Sigma_0)\|$$

As $\tilde{\lambda} \xrightarrow{p} \lambda_0$, there exists $(\delta_G)_G \in (0;+\infty)^{\mathbb{N}}$ such that $\delta_G \underset{G\to+\infty}{\longrightarrow} 0$ and $P(\|\tilde{\lambda}-\lambda_0\| \leq \delta_G) \underset{G\to+\infty}{\longrightarrow} 1$.
Let

$$\Delta_G(s) := \sup_{\|\lambda-\lambda_0\|\leq\delta_G} \|f(s;\lambda,\Sigma_0) - f(s;\lambda_0,\Sigma_0)\|$$

Since for all $s \in [0,1]$, $f(s;.,\Sigma_0)$ is continuous, by construction, in λ_0, then $P\left(\Delta_G(s) \underset{G\to+\infty}{\longrightarrow} 0\right) = 1$ and for G large enough, $\Delta_G(s) \leq 2\sup_{\lambda\in\mathcal{N}}\|f(s;\lambda,\Sigma_0)\|$. Using the dominated convergence theorem,

$$\mathrm{E}\left[\Delta_G(s)\right] = \int \Delta_G(z)\, dP_s(z) \underset{G\to+\infty}{\longrightarrow} \int 0\, dP_s(z) = 0$$

Then, using Markov's inequality, we have, for all $\varepsilon > 0$,

$$P\left(\frac{1}{G}\sum_{g=1}^{G}\Delta_G(s_g) > \varepsilon\right) \leq \frac{\mathrm{E}\left[\Delta_G(s)\right]}{\varepsilon} \underset{G\to+\infty}{\longrightarrow} 0$$

Hence, $\frac{1}{G}\sum_{g=1}^{G}\Delta_G(s_g) \xrightarrow{p} 0$ and

$$\begin{aligned}\left\|\frac{1}{G}\sum_{g=1}^{G} f(s_g;\tilde{\lambda},\Sigma_0) - \frac{1}{G}\sum_{g=1}^{G} f(s_g;\lambda_0,\Sigma_0)\right\| &\leq \frac{1}{G}\sum_{g=1}^{G}\|f(s_g;\tilde{\lambda},\Sigma_0) - f(s_g;\lambda_0,\Sigma_0)\| \\ &\leq \frac{1}{G}\sum_{g=1}^{G}\Delta_G(s_g) \text{ with probability 1}\end{aligned}$$

Therefore, $\left\|\frac{1}{G}\sum_{g=1}^{G} f(s_g;\tilde{\lambda},\Sigma_0) - \frac{1}{G}\sum_{g=1}^{G} f(s_g;\lambda_0,\Sigma_0)\right\| = o_P(1)$.
Finally, related to term (44), we have by the weak law of large number that

$$\left\|\frac{1}{G}\sum_{g=1}^{G} f(s_g;\lambda_0,\Sigma_0) - \mathrm{E}\left[f(s;\lambda_0,\Sigma_0)\right]\right\| = o_P(1)$$

Combining the results on the three terms (42)-(44), we obtain that $\hat{\Omega} \xrightarrow{p} \Omega_0$.
Second, we suppose that

1. Θ is compact
2. The assumptions such that Proposition 3 holds are met
3. Ω_0 is definite-positive
4. $\mathrm{E}\left[\left|\bar{y}_g^E\right|^2\right] < +\infty$, $\mathrm{E}\left[\left|\bar{y}_g^N\right|^2\right] < +\infty$ and $\sup_{s\in[0,1]}\|\Sigma_0^{-1}(s)\| < +\infty$

The latter condition implies that

$$\mathrm{E}\left[\sup_{\lambda\in\Theta}\|g(Z,\lambda)\|\right] < +\infty$$

with

$$g(Z,\lambda) := \left(\frac{\partial u^R}{\partial \lambda}(s;\tilde{\lambda})\right)' \hat{\Sigma}^{-1}(s;\tilde{\lambda})u_B^R(Z;\lambda)$$

since, by construction,

$$\sup_{\lambda\in\Theta}\left\|\frac{\partial u^R}{\partial \lambda}(s;\lambda)\right\| < +\infty$$

As $g(Z;.)$ is continuous at each $\lambda \in \Theta$ with probability one, all the conditions of Theorem 2.6 of Newey and McFadden (1994) are met and $\hat{\lambda} \xrightarrow{p} \lambda_0$. Third, as we also suppose that $\lambda_0 \in \text{Int}(\Theta)$ and as, by construction :

- $g(z;\lambda)$ is continuously differentiable for a neighborhood $\mathcal{N}$ of λ_0, with probability one
- $E\left[\sup_{\lambda\in\mathcal{N}}\left\|w^R(s)\left(\frac{\partial}{\partial\lambda}u_B^R(s;\lambda)\right)\right\|\right] < +\infty$

then, Theorem 3.4 from Newey and McFadden (1994) applies and we have

$$\sqrt{G}(\hat{\lambda}-\lambda_0) \xrightarrow{d} \mathcal{N}(0,\Omega_0^{-1})$$

as the weighting matrix has been chosen to obtain the efficient GMM estimator.

Appendix A.6 Proposition 5

TO BE DONE

Appendix B Inclusion of 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 $k \in \{E, NE\}$,

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

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 $(g,t) \in \{1,\dots,G\} \times \{1,2\}$,

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

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

Proposition 6 (Conditional Moment Conditions - Natural Experiment with Covariates)
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,

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

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

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

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

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

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

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

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

Now,

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

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

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

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

□

Appendix B.2 Estimation

We consider

$$
m^{DiD}(\tilde{Z}, \lambda, \hat{\mu}_D^X) = \tilde{w}(s, \Delta X)\left\{\hat{\omega}^{DiDX}(\tilde{Z}, \hat{\mu}_D^X)\bar{Y} - \frac{\delta}{\phi(s, \lambda)}\begin{pmatrix} 1 - (1 - s)\theta^w \\ s\theta^b \end{pmatrix}\right\}
$$

where $\tilde{w}(s, \Delta X)$ is a $r \times 2$ matrix of functions of $(s, \Delta X)$ such that $r \geq 4$,

$$
\hat{\omega}^{DiDX}(\tilde{Z}, \hat{\mu}_D^X) = \frac{D}{\hat{\mu}_D^X(s, \Delta X)} - \frac{1 - D}{1 - \hat{\mu}_D^X(s, \Delta X)}
$$

and $\hat{\mu}_D^X(s, \Delta X)$ is a non-parametric estimator of $\mu_D^X(s, \Delta X) = E[D_2 | s, \Delta X]$ based on a series logistic regression of D_2 on a power series of $(s, \Delta X)$. To prove asymptotic normality, the technical condition on K_G in Assumption 5′ - point 5 has to be adapted, by requiring that K_G goes at a slower rate to infinity, so as to avoid overfitting. The greater the dimension of ΔX, the slower will be the rate. Then, one can show that

Conjecture 1 (GMM Estimator in the Natural Experiment Setting with Covariates)**.** *Let $(\tilde{Z}_g)_{g=1,\dots,G}$ be an i.i.d sample of G groups, where $\tilde{Z}_g = (\Delta\bar{y}_g^E, \Delta\bar{y}_g^N, D_{2g}, s_g, \Delta X_g)$. Let*

$$
\hat{V}_G \xrightarrow{p} \tilde{V}_0 := \mathrm{E}\left[\left\{m^{DiD}(\tilde{Z}, \lambda_0, \mu_D^X) + \alpha(\tilde{Z})\right\}\left\{m^{DiD}(\tilde{Z}, \lambda_0, \mu_D^X) + \alpha(\tilde{Z})\right\}'\right]
$$

with

$$
\alpha(\tilde{Z}) = -\tilde{w}(s, \Delta X)\left(\frac{\tilde{\mu}_1(s, \Delta X)}{\mu_D^X(s, \Delta X)^2} + \frac{\tilde{\mu}_0(s, \Delta X)}{(1 - \mu_D^X(s, \Delta X))^2}\right)(D_2 - \mu_D^X(s, \Delta X))
$$

where, for any $s \in \mathcal{S}$, $\tilde{\mu}_1(s, \Delta X) = E[\overline{Y}D_2 | s, \Delta X]$ and $\tilde{\mu}_0(s, \Delta X) = E[\overline{Y}(1 - D_2) | s, \Delta X]$. Let

$$
m^{\tilde{D}iD}(\tilde{Z}, \lambda_0) = \frac{\partial}{\partial\lambda'}m^{DiD}(\tilde{Z}, \lambda_0, \mu_D^X)
$$

which does not depend on the propensity score. Supposing that

1. *Assumptions 1′, 2′, 3′, 4′ and 5′ hold and identification 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* $K \in (0,1)$

Then

$$\hat{\lambda}^* := \arg\min_{\lambda \in \Theta} \left(\frac{1}{G} \sum_{g=1}^{G} m^{DiD}(\tilde{Z}_g, \lambda, \hat{\mu}_D^X) \right)' \hat{V}_G^{-1} \left(\frac{1}{G} \sum_{g=1}^{G} m^{DiD}(\tilde{Z}_g, \lambda, \hat{\mu}_D^X) \right) \tag{45}$$

is a consistent estimator of λ_0 *whose asymptotic distribution is*

$$\sqrt{G}(\hat{\lambda}^* - \lambda_0) \xrightarrow{d} \mathcal{N}\left(0, \left[\tilde{M}_0' \tilde{V}_0^{-1} \tilde{M}_0\right]^{-1}\right)$$

with $\lambda_0 := (\delta_0, \theta_0^w, \theta_0^b)$ *and* $\tilde{M}_0 := \mathrm{E}\left[m^{\tilde{D}iD}(\tilde{Z}, \lambda_0)\right]$

Appendix C Identity-specific "within" and "between" influences

In this section, we consider the more general LIM model that allows for 2 different "within" endogenous peer effects for eligible and non-eligible individuals. They are denoted by $\theta^{w,E}$ and $\theta^{w,N}$. Moreover, we suppose that influences may be non-symmetric between eligible and non-eligible individuals. Let $\theta^{b,N,E}$ be the "between" endogenous effects from eligible individuals on non-eligible individuals and $\theta^{b,E,N}$ be the "between" endogenous effects from non-eligible individuals on eligible individuals. We remind the

Assumption 1$*$ (Linear-in-means model - 5 parameters)
Within each group $g \in \{1, \ldots, G\}$ *of the i.i.d sample, the outcome of individual i is defined as*

$$y_{ig} = E_{ig} y_{ig}^E + (1 - E_{ig}) y_{ig}^N$$

with

$$\begin{aligned} y_{ig}^E &= \theta_0^{w,E} s_g \bar{y}_g^E + \theta_0^{b,E,N}(1 - s_g)\bar{y}_g^N + \delta_0 D_{ig} + e_{ig}^E \\ y_{ig}^N &= \theta_0^{b,N,E} s_g \bar{y}_g^E + \theta_0^{w,N}(1 - s_g)\bar{y}_g^N + e_{ig}^N \end{aligned}$$

Appendix C.1 Identification

Averaging, within group g, the outcome variables for eligible and non-eligible individuals, we get the following expressions

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

Then, plugging-in the expression of $\bar{y}_g^E$ into the one of $\bar{y}_g^N$, we get

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

Rearranging terms, we get

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

Hence,

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

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

$$\begin{aligned}
\bar{y}_g^E =& \frac{1-\theta_0^{w,N}(1-s_g)}{1-s_g\theta_0^{w,E}-(1-s_g)\theta_0^{w,N}+s_g(1-s_g)\left(\theta_0^{w,E}\theta_0^{w,N}-\theta_0^{b,N,E}\theta_0^{b,E,N}\right)}\bar{e}_g^E \\
&+\frac{\theta_0^{b,E,N}(1-s_g)}{1-s_g\theta_0^{w,E}-(1-s_g)\theta_0^{w,N}+s_g(1-s_g)\left(\theta_0^{w,E}\theta_0^{w,N}-\theta_0^{b,N,E}\theta_0^{b,E,N}\right)}\bar{e}_g^N \\
&+\frac{\delta_0(1-\theta_0^{w,N}(1-s_g))}{1-s_g\theta_0^{w,E}-(1-s_g)\theta_0^{w,N}+s_g(1-s_g)\left(\theta_0^{w,E}\theta_0^{w,N}-\theta_0^{b,N,E}\theta_0^{b,E,N}\right)}D_g \\
=& \frac{1-\theta_0^{w,N}(1-s_g)}{1-s_g\theta_0^{w,E}-(1-s_g)\theta_0^{w,N}+s_g(1-s_g)\left(\theta_0^{w,E}\theta_0^{w,N}-\theta_0^{b,N,E}\theta_0^{b,E,N}\right)}\bar{e}_g^E \\
&+\frac{\theta_0^{b,E,N}(1-s_g)}{1-s_g\theta_0^{w,E}-(1-s_g)\theta_0^{w,N}+s_g(1-s_g)\left(\theta_0^{w,E}\theta_0^{w,N}-\theta_0^{b,N,E}\theta_0^{b,E,N}\right)}\bar{e}_g^N \\
&+\delta_0\left(1+s_g.\frac{\theta_0^{w,E}-(1-s_g)\left[\theta_0^{w,E}\theta_0^{w,N}-\theta_0^{b,N,E}\theta_0^{b,E,N})\right]}{1-s_g\theta_0^{w,E}-(1-s_g)\theta_0^{w,N}+s_g(1-s_g)\left(\theta_0^{w,E}\theta_0^{w,N}-\theta_0^{b,N,E}\theta_0^{b,E,N}\right)}\right)D_g
\end{aligned} \tag{46}$$

We then get the following proposition

Proposition 7
Provided Assumptions 1∗, 2 and 3 hold and all the mentioned conditional expectations are well-defined, we have

$$\begin{aligned}
\tau^E(s) &= \mathrm{E}\left[\bar{y}_g^E|s_g=s,D_g=1\right]-\mathrm{E}\left[\bar{y}_g^E|s_g=s,D_g=0\right] \\
&= \mathrm{E}\left[\frac{D_g-\Pr(D_g=1|s_g=s)}{\Pr(D_g=1|s_g=s)(1-\Pr(D_g=1|s_g=s))}.\bar{y}_g^E\Big|s_g=s\right] \\
&= \delta_0\left(1+s_g.\frac{\theta_0^{w,E}-(1-s_g)\left[\theta_0^{w,E}\theta_0^{w,N}-\theta_0^{b,N,E}\theta_0^{b,E,N})\right]}{1-s_g\theta_0^{w,E}-(1-s_g)\theta_0^{w,N}+s_g(1-s_g)\left(\theta_0^{w,E}\theta_0^{w,N}-\theta_0^{b,N,E}\theta_0^{b,E,N}\right)}\right)
\end{aligned} \tag{47}$$

and

$$\begin{aligned}\tau^N(s) &= \mathrm{E}\left[\bar{y}_g^N | s_g = s, T_g = 1\right] - \mathrm{E}\left[\bar{y}_g^N | s_g = s, D_g = 0\right] \\ &= \mathrm{E}\left[\frac{D_g - \Pr(D_g = 1|s_g = s)}{\Pr(D_g = 1|s_g = s)(1 - \Pr(D_g = 1|s_g = s))} . \bar{y}_g^N \middle| s_g = s\right] \\ &= \frac{\delta_0 \theta_0^{b,N,E} s_g}{1 - s_g\theta_0^{w,E} - (1-s_g)\theta_0^{w,N} + s_g(1-s_g)\left(\theta_0^{w,E}\theta_0^{w,N} - \theta_0^{b,N,E}\theta_0^{b,E,N}\right)} \end{aligned} \tag{48}$$

Hence,

$$P^E(s) := \frac{\theta_0^{w,E} - (1-s_g)\left[\theta_0^{w,E}\theta_0^{w,N} - \theta_0^{b,N,E}\theta_0^{b,E,N})\right]}{1 - s_g\theta_0^{w,E} - (1-s_g)\theta_0^{w,N} + s_g(1-s_g)\left(\theta_0^{w,E}\theta_0^{w,N} - \theta_0^{b,N,E}\theta_0^{b,E,N}\right)} \tag{49}$$

$$P^N(s) := \frac{1}{1 - s_g\theta_0^{w,E} - (1-s_g)\theta_0^{w,N} + s_g(1-s_g)\left(\theta_0^{w,E}\theta_0^{w,N} - \theta_0^{b,N,E}\theta_0^{b,E,N}\right)} \tag{50}$$

Similarly, we could state the analogous proposition to proposition 2 in the 5 parameter scenario, provided Assumptions $1*$, $2'$, $3'$ and $4'$ hold.

We also have the following proposition

Proposition 8 (Sufficient Conditions for Identification of $\lambda_0 := (\delta_0, \theta^{w,E}, \theta^{w,N}, \theta^{b,N,E}, \theta^{b,E,N})$)
Provided $\delta_0 \neq 0$ and $\theta_0^b \neq 0$, if Assumptions $1*$, 2 and 3 hold and if one of the two following conditions is satisfied

1. *there exist at least 3 shares s_1, s_2, $s_3 \in (0,1)$, $s_1 \neq s_2 \neq s_3$, that have positive probability mass and for which both τ^N 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 $\tau^N(s)$ 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 the moment condition

$$\mathrm{E}\left[\tilde{u}(\overline{y}_g^E, \overline{y}_g^N, D_g, s_g; \lambda_0)|s_g = s\right] = \begin{pmatrix} \mathrm{E}\left[\tilde{u}^E(\overline{y}_g^E, D_g, s_g; \lambda_0)|s_g = s\right] \\ \mathrm{E}\left[\tilde{u}^N(\overline{y}_g^N, D_g, s_g; \lambda_0)|s_g = s\right]\end{pmatrix} = 0 \tag{51}$$

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

$$\tilde{u}^E(\overline{y}_g^E, D_g, s_g; \lambda) = \omega^R(D_g, s_g)\overline{y}_g^E - \frac{\delta\left(1 - \theta^{w,N}(1-s_g)\right)}{1 - s_g\theta^{w,E} - (1-s_g)\theta^{w,N} + s_g(1-s_g)\left(\theta^{w,E}\theta^{w,N} - \theta^{b,N,E}\theta^{b,E,N}\right)}$$

$$\tilde{u}^N(\overline{y}_g^N, D_g, s_g; \lambda) = \omega^R(D_g, s_g)\overline{y}_g^N - \frac{\delta\theta^{b,N,E}s_g}{1 - s_g\theta^{w,E} - (1-s_g)\theta^{w,N} + s_g(1-s_g)\left(\theta^{w,E}\theta^{w,N} - \theta^{b,N,E}\theta^{b,E,N}\right)}$$

and $\omega^R(D_g, s_g) = \dfrac{D_g - P(D_g = 1|s_g)}{P(D_g = 1|s_g)(1 - P(D_g = 1|s_g))}$

Proof. Let s_1, s_2, s_3 be three distinct elements of $(0,1)$, such that $s_1 \neq s_2 \neq s_3$. Supposing that Assumptions $1*$, 2 and 3 hold, we have the following conditions

$$\tau^E(s_1) = \frac{\delta_0 \left(1 - \theta_0^{w,N}(1-s_1)\right)}{1 - s_1\theta_0^{w,E} - (1-s_1)\theta_0^{w,N} + s_1(1-s_1)\left(\theta_0^{w,E}\theta_0^{w,N} - \theta_0^{b,N,E}\theta_0^{b,E,N}\right)} \tag{52}$$

$$\tau^N(s_1) = \frac{\delta_0\theta_0^{b,N,E}s_1}{1 - s_1\theta_0^{w,E} - (1-s_1)\theta_0^{w,N} + s_1(1-s_1)\left(\theta_0^{w,E}\theta_0^{w,N} - \theta_0^{b,N,E}\theta_0^{b,E,N}\right)} \tag{53}$$

$$\tau^E(s_2) = \frac{\delta_0 \left(1 - \theta_0^{w,N}(1-s_2)\right)}{1 - s_2\theta_0^{w,E} - (1-s_2)\theta_0^{w,N} + s_2(1-s_2)\left(\theta_0^{w,E}\theta_0^{w,N} - \theta_0^{b,N,E}\theta_0^{b,E,N}\right)} \tag{54}$$

$$\tau^N(s_2) = \frac{\delta_0\theta_0^{b,N,E}s_2}{1 - s_2\theta^{w,E} - (1-s_2)\theta_0^{w,N} + s_2(1-s_2)\left(\theta_0^{w,E}\theta_0^{w,N} - \theta_0^{b,N,E}\theta_0^{b,E,N}\right)} \tag{55}$$

$$\tau^E(s_3) = \frac{\delta_0 \left(1 - \theta_0^{w,N}(1-s_3)\right)}{1 - s_3\theta_0^{w,E} - (1-s_3)\theta_0^{w,N} + s_3(1-s_3)\left(\theta_0^{w,E}\theta_0^{w,N} - \theta_0^{b,N,E}\theta_0^{b,E,N}\right)} \tag{56}$$

$$\tau^N(s_3) = \frac{\delta_0\theta_0^{b,N,E}s_3}{1 - s_3\theta_0^{w,E} - (s_3)\theta_0^{w,N} + s_3(1-s_3)\left(\theta_0^{w,E}\theta_0^{w,N} - \theta_0^{b,N,E}\theta_0^{b,E,N}\right)} \tag{57}$$

Note that for any $s \in (0,1)$ and for any $(x_1, x_2, x_3, x_4) \in (-1,1)^4$,

$$1 - x_1 s - x_2(1-s) + s(1-s)\left[x_1x_2 - x_3x_4\right] > 0$$

Dividing (53) by (52) and (55) by (54), we get

$$\frac{\tau^N(s_1)}{\tau^E(s_1)} = \frac{\theta_0^{b,N,E}s_1}{1 - \theta_0^{w,N}(1-s_1)} \tag{58}$$

$$\frac{\tau^N(s_2)}{\tau^E(s_2)} = \frac{\theta_0^{b,N,E}s_2}{1 - \theta_0^{w,N}(1-s_2)} \tag{59}$$

which can be expressed as new moment conditions

$$(1 - \theta_0^{w,N}(1-s_1))\tau^N(s_1) - \theta_0^{b,N,E}s_1\tau^E(s_1) = 0 \tag{60}$$

$$(1 - \theta_0^{w,N}(1-s_2))\tau^N(s_2) - \theta_0^{b,N,E}s_2\tau^E(s_2) = 0 \tag{61}$$

Solving the system composed of (60) and (61), provided $s_1 \neq s_2$, leads to

$$\begin{cases} \theta_0^{w,N} = \dfrac{s_1\tau^E(s_1)\tau^N(s_2) - s_2\tau^E(s_2)\tau^N(s_1)}{s_1(1-s_2)\tau^E(s_1)\tau^N(s_2) - (1-s_1)s_2\tau^N(s_1)\tau^E(s_2)} \\ \theta_0^{b,N,E} = \dfrac{\tau^N(s_1)\tau^N(s_2)(s_1-s_2)}{s_1(1-s_2)\tau^E(s_1)\tau^N(s_2) - (1-s_1)s_2\tau^N(s_1)\tau^E(s_2)} \end{cases}$$

Hence, $\theta_0^{b,N,E}{}_0$ and $\theta_0^{w,N}$ are identified. As a consequence, δ_0 is also identified based on (56), (57) and one equation from (52) to (55) can be used to identify the 3 other parameters. $\square$

Appendix C.2 Estimation

The true vector of parameters $\lambda_0 := (\delta_0, \theta_0^{w,E}, \theta_0^{w,N}, \theta_0^{b,N,E}, \theta_0^{b,E,N})$ can be estimated via an adapted version of our Generalized Method of Moments (GMM) estimators. Let $Z_g := (\overline{y}^E, \overline{y}^N, s_g, D_g)$,

$$\hat{\lambda}^* := \arg\min_{\lambda\in\Theta} \left(\frac{1}{G}\sum_{g=1}^{G} m(Z_g, \lambda, \hat{p})\right)' \hat{V}_G^{-1} \left(\frac{1}{G}\sum_{g=1}^{G} m(Z_g, \lambda, \hat{p})\right) \tag{62}$$

with

$$m(Z_g, \lambda, \hat{p}) = w(s_g)\left[\left(\frac{D_g}{\hat{p}(s_g)} - \frac{1-D_g}{1-\hat{p}(s_g)}\right)\begin{pmatrix}\overline{y}^E \\ \overline{y}^N\end{pmatrix} - \frac{\delta}{\phi(s_g,\lambda)}\begin{pmatrix}1-(1-s_g)\theta^{w,N} \\ s\theta^{b,N,E}\end{pmatrix}\right]$$

and

$$\phi(s_g, \lambda) = 1 - \theta^{w,E}s_g - \theta^{w,N}(1-s_g) + s_g(1-s_g)\left(\theta^{w,E}\theta^{w,N} - \theta^{b,N,E}\theta^{b,E,N}\right)$$

$w(s_g)$ is a $k\times 2$ matrix of functions of s_g with $k \geq 5$. $\hat{p}(s_g)$ is a series logit estimate of $\mu_D(s_g) := P(D_g = 1|s_g)$. Finally,

$$\hat{V}_G \xrightarrow{p} V_0 := \mathrm{E}\left[m^{DiD}(Z, \lambda_0, \mu_D)m^{DiD}(Z, \lambda_0, \mu_D)'\right]$$

with

$$m^{DiD}(Z, \lambda_0, \mu_D) = m(Z, \lambda_0, \mu_D) - w(s)\left(\frac{E\left[\overline{Y}D|s\right]}{\mu_D(s)^2} + \frac{E\left[\overline{Y}(1-D)|s\right]}{(1-\mu_D(s))^2}\right)(D - \mu_D(s))$$

and $\overline{Y} = \begin{pmatrix}\overline{y}^E & \overline{y}^N\end{pmatrix}'$. We get the following result

Proposition 9 (GMM Asymptotic Properties)
Provided Assumptions 1∗, 2, 3 and technical conditions hold,

$$\sqrt{G}(\hat{\lambda}^* - \lambda_0) \xrightarrow{d} \mathcal{N}\left(0, \left[M_0'V_0^{-1}M_0\right]^{-1}\right) \text{ with } M_0 := \mathrm{E}\left[\frac{\partial m}{\partial\lambda}(Z, \lambda_0)\right]$$

In practice, standard errors can be estimated via bootstrap.

Appendix D Simulations

We evaluate our estimation procedure for the 3 and 5 parameter scenarios via simulations. The following Data Generating Process. An i.i.d sample of G groups is simulated. For each group g, a vector $Z_g = (\Delta\overline{y}_g^E, \Delta\overline{y}_g^N, s_g, D_g, \Delta X_g^E, \Delta X_g^N)$ is drawn such that

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

with $(\beta^E, \beta^N) = (-1, -2)$ and

$$s_g \sim \mathcal{U}_{(0,1]}$$
$$D_g \sim \mathcal{B}(1/2)$$
$$\begin{pmatrix}\alpha_g^E \\ \alpha_g^N \\ X_g^E \\ X_g^N\end{pmatrix} \sim \mathcal{N}\left(\mu_{\alpha,X}, \Sigma_{\alpha,X}\right)$$

300 simulations were performed. Figures 8 and 9 show that having 1,000 or 10,000 groups is not enough to obtain consistent estimators in the 5 parameter scenario. It appears that $\hat{\theta}^{b,N,E}$ and $\hat{\theta}^{w,E}$ converge relatively fast while $\hat{\theta}^{b,E,N}$, $\hat{\theta}^{w,N}$ and $\hat{\delta}$ are quite imprecise. Global convergence is assured for sample of a million groups, as shown by Figure 10. We explain this slow convergence, coming from a large asymptotic variance, by the difficulty to disentangle the endogenous peer effects from non-eligible to eligible individuals, $\theta_0^{b,E,N}$, from the direct effect of the treatment δ_0. In our baseline scenario with 3 parameter, convergence is faster. Asymptotic normality is almost reached with 1,000 groups, as shown by Figure 11.

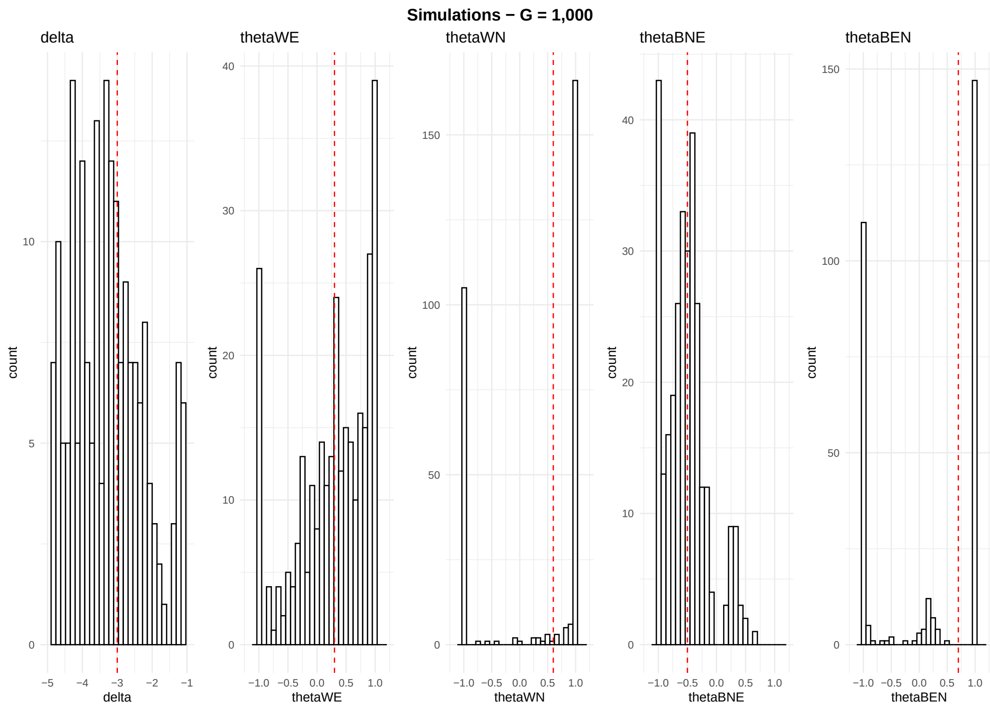


Figure 8: Simulations - 4 endogenous peer effects parameters - 1,000 groups

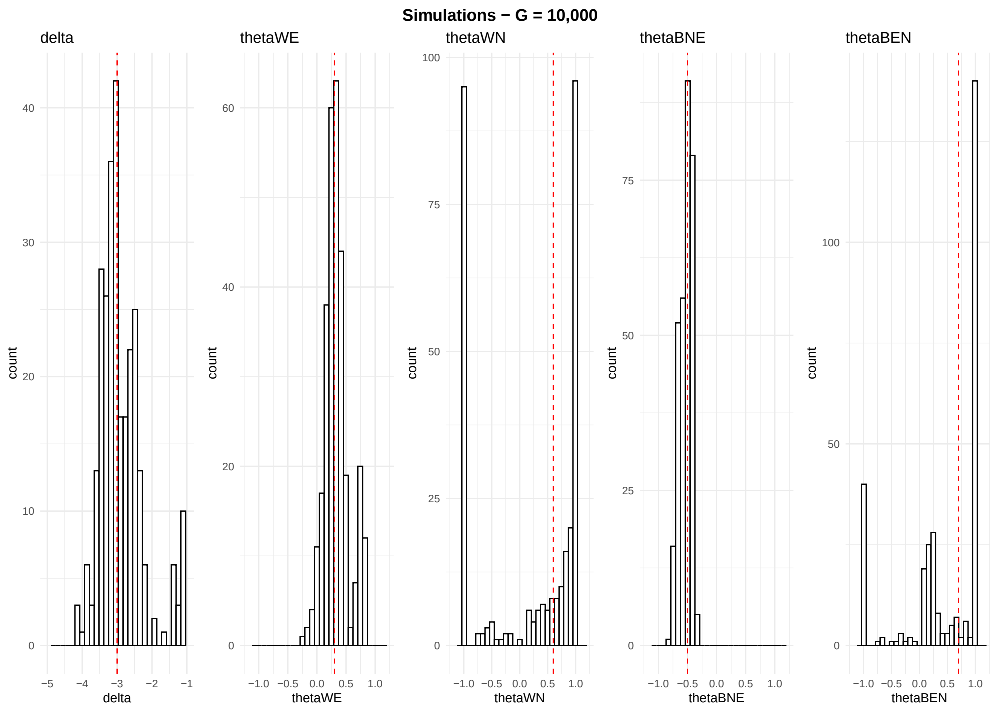


Figure 9: Simulations - 4 endogenous peer effects parameters - 10,000 groups

Note: the dotted red line represents the true values of the coefficients, in order δ_0, $\theta_0^{w,E}$, $\theta_0^{w,N}$, $\theta_0^{b,N,E}$ and $\theta_0^{b,E,N}$

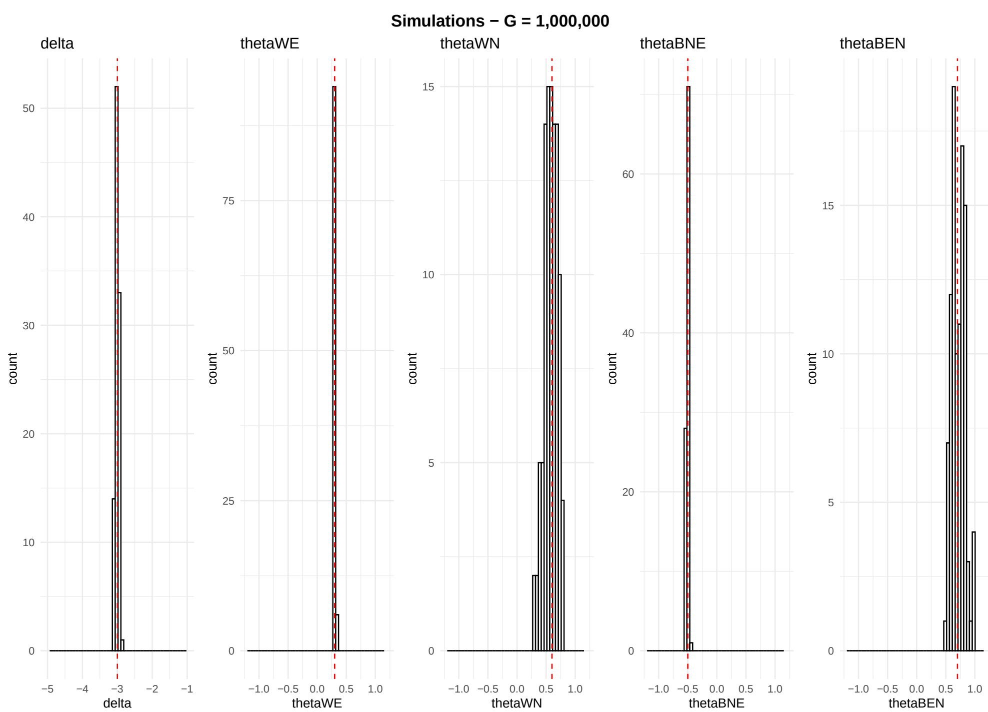


Figure 10: Simulations - 4 endogenous peer effects parameters - 1,000,000 groups

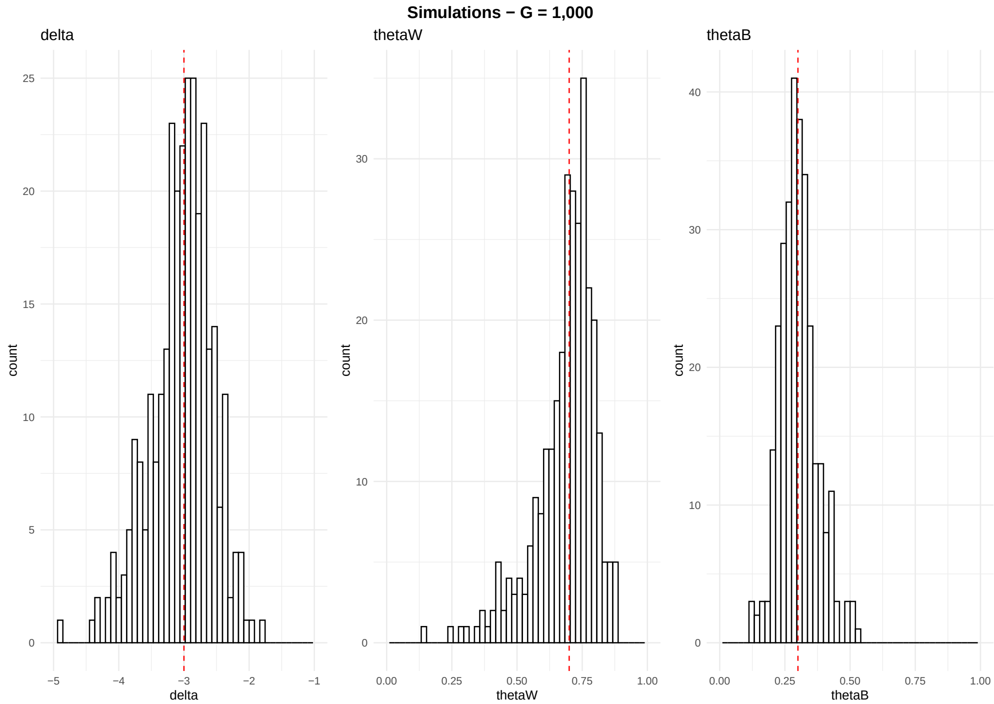


Figure 11: Simulations - 2 endogenous peer effects parameters - 1,000 groups

Note: the dotted red line represents the true values of the coefficients

Appendix E Identity not restricted to eligibility

In this extension, we suppose that the identity that matters in social interactions is different from eligibility. For instance, in the context of Progresa, we may suppose that boys (resp. girls) are more influenced by the actions of their male (resp. female) peers than by their female (resp. male) peers. Gender, which the relevant identity for social interactions, is different from being eligible to Progresa, since eligibility is only based on households' income. Formally, here is the model that we consider in this extension

Assumption 1† (Linear-in-means model - orthogonal identity)
For a group $g \in \{1, \ldots, G\}$ of the i.i.d sample, the outcome of individual i is defined as follow

$$
\begin{aligned}
y_{ig} &= y_{ig}^M M_{ig} + y_{ig}^F (1 - M_{ig}) \\
y_{ig}^M &= \theta_0^{w,M} s_g^M \bar{y}_g^M + \theta_0^{b,M,F} (1 - s_g^M) \bar{y}_g^F + \delta_0 D_{ig} E_{ig} + e_{ig}^M \qquad (63) \\
y_{ig}^F &= \theta_0^{b,F,M} s_g^M \bar{y}_g^M + \theta_0^{w,F} (1 - s_g^M) \bar{y}_g^F + \delta_0 D_{ig} E_{ig} + e_{ig}^F \qquad (64)
\end{aligned}
$$

with $M_{ig} \in \{0,1\}$ equals 1 if the individual is from the "male" identity and $E_{ig} \in \{0,1\}$ equals 1 if the individual is eligible to the treatment. $s_g^M := \overline{M}_g$, i.e. it is the share of "males" in the peer group.

Averaging, within group g, the outcome variables for "male" and "female" individuals, we get the following expressions

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

where $s_g^{E,M} = \dfrac{\sum_{i=1}^{n_g} E_{ig} M_{ig}}{\sum_{i=1}^{n_g} M_{ig}}$ and $s_g^{E,F} = \dfrac{\sum_{i=1}^{n_g} E_{ig} (1 - M_{ig})}{\sum_{i=1}^{n_g} (1 - M_{ig})}$ are, respectively, the share of "males" (resp. "females") that are eligible to the treatment. Then, plugging-in the expression of $\bar{y}_g^F$ into the one of $\bar{y}_g^M$, we get

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

Rearranging terms, we get

$$
\begin{aligned}
&\left[1 - s_g^M \theta_0^{w,M} - (1 - s_g^M) \theta_0^{w,F} + s_g^M (1 - s_g^M) \left(\theta_0^{w,M} \theta_0^{w,F} - \theta_0^{b,M,F} \theta_0^{b,F,M} \right) \right] \bar{y}_g^M \\
= \; &\delta_0 \left(\theta_0^{b,M,F} s_g^{E,F} (1 - s_g^M) + \left(1 - \theta^{w,F} (1 - s_g^M) \right) s_g^{E,M} \right) D_g + \theta^{b,M,F} (1 - s_g^M) \bar{e}_g^F + \left(1 - \theta_0^{w,F} (1 - s_g^M) \right) \bar{e}_g^M
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\bar{y}_g^M = & \frac{1-\theta_0^{w,F}(1-s_g^M)}{1-s_g^M\theta_0^{w,M}-(1-s_g^M)\theta_0^{w,F}+s_g^M(1-s_g^M)\left(\theta_0^{w,M}\theta_0^{w,F}-\theta_0^{b,M,F}\theta_0^{b,F,M}\right)}\bar{e}_g^M \\
& + \frac{\theta^{b,M,F}(1-s_g^M)}{1-s_g^M\theta_0^{w,M}-(1-s_g^M)\theta_0^{w,F}+s_g^M(1-s_g^M)\left(\theta_0^{w,M}\theta_0^{w,F}-\theta_0^{b,M,F}\theta_0^{b,F,M}\right)}\bar{e}_g^F \\
& + \frac{\delta_0\left(s_g^{E,F}\theta_0^{b,M,F}(1-s_g^M)+s_g^{E,M}\left(1-\theta^{w,F}(1-s_g^M)\right)\right)}{1-s_g^M\theta_0^{w,M}-(1-s_g^M)\theta_0^{w,F}+s_g^M(1-s_g^M)\left(\theta_0^{w,M}\theta_0^{w,F}-\theta_0^{b,M,F}\theta_0^{b,F,M}\right)}D_g
\end{aligned}
$$

Now, plugging-in the reduced form expression of $\bar{y}_g^M$ into $\bar{y}_g^F$ and developing, we get

$$
\begin{aligned}
\bar{y}_g^F = & \frac{\theta^{b,F,M}s_g^M}{1-s_g^M\theta_0^{w,M}-(1-s_g^M)\theta_0^{w,F}+s_g^M(1-s_g^M)\left(\theta_0^{w,M}\theta_0^{w,F}-\theta_0^{b,M,F}\theta_0^{b,F,M}\right)}\bar{e}_g^M \\
& + \frac{(1-\theta^{w,M}s_g^M)}{1-s_g^M\theta_0^{w,M}-(1-s_g^M)\theta_0^{w,F}+s_g^M(1-s_g^M)\left(\theta_0^{w,M}\theta_0^{w,F}-\theta_0^{b,M,F}\theta_0^{b,F,M}\right)}\bar{e}_g^F \\
& + \frac{\delta_0\left(s_g^{E,F}(1-\theta^{w,M}s_g^M)+s_g^{E,M}\theta^{b,F,M}s_g^M\right)}{1-s_g^M\theta_0^{w,M}-(1-s_g^M)\theta_0^{w,F}+s_g^M(1-s_g^M)\left(\theta_0^{w,M}\theta_0^{w,F}-\theta_0^{b,M,F}\theta_0^{b,F,M}\right)}D_g
\end{aligned} \tag{65}
$$

We then make the following "adapted" assumptions

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

$$(\bar{e}_g^F, \bar{e}_g^M, s_g^M, s_g^{E,F}, s_g^{E,M}) \perp\!\!\!\perp D_g \tag{66}$$

Assumption 2† states that treatment is randomly assigned. In particular, groups receive the treatment independently of their share of eligible units for both identities and of their share of "males" in the group. 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 $(s^M, s^{E,F}, s^{E,M}) \in \mathcal{S}^M \times \mathcal{S}^{E,F} \times \mathcal{S}^{E,M}$,

$$0 < P(D_g = 1 | s^M, s^{E,F}, s^{E,M}) < 1 \tag{67}$$

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

Proposition 10
Provided Assumptions 1†, 2† and 3† hold and all the mentioned conditional expectations are well-defined,

we have

$$\begin{aligned}\tau^M(s^M,s^{E,F},s^{E,M}) &= \mathrm{E}\left[\bar{y}^M|s^M,s^{E,F},s^{E,M},D=1\right]-\mathrm{E}\left[\bar{y}^M|s^M,s^{E,F},s^{E,M},D=0\right]\\ &= \mathrm{E}\left[\frac{D-\Pr(D=1|s^M,s^{E,F},s^{E,M})}{\Pr(D=1|s^M,s^{E,F},s^{E,M})(1-\Pr(D=1|s^M,s^{E,F},s^{E,M}))}.\bar{y}^M\Big|s^M,s^{E,F},s^{E,M}\right]\\ &= \frac{\delta_0\left(s^{E,F}\theta_0^{b,M,F}(1-s^M)+s^{E,M}\left(1-\theta^{w,F}(1-s^M)\right)\right)}{1-s^M\theta_0^{w,M}-(1-s^M)\theta_0^{w,F}+s^M(1-s^M)\left(\theta_0^{w,M}\theta_0^{w,F}-\theta_0^{b,M,F}\theta_0^{b,F,M}\right)} \qquad (68)\end{aligned}$$

and

$$\begin{aligned}\tau^F(s^M,s^{E,F},s^{E,M}) &= \mathrm{E}\left[\bar{y}^F|s^M,s^{E,F},s^{E,M},D=1\right]-\mathrm{E}\left[\bar{y}^F|s^M,s^{E,F},s^{E,M},D=0\right]\\ &= \mathrm{E}\left[\frac{D-\Pr(D=1|s^M,s^{E,F},s^{E,M})}{\Pr(D=1|s^M,s^{E,F},s^{E,M})(1-\Pr(D=1|s^M,s^{E,F},s^{E,M}))}.\bar{y}^F\Big|s^M,s^{E,F},s^{E,M}\right]\\ &= \frac{\delta_0\left(s^{E,F}(1-\theta^{w,M}s^M)+s^{E,M}\theta^{b,F,M}s^M\right)}{1-s^M\theta_0^{w,M}-(1-s^M)\theta_0^{w,F}+s^M(1-s^M)\left(\theta_0^{w,M}\theta_0^{w,F}-\theta_0^{b,M,F}\theta_0^{b,F,M}\right)} \qquad (69)\end{aligned}$$

In the following, for conciseness, we define, for any $g\in\{1,\dots,G\}$, $s_g := (s_g^M, s_g^{E,F}, s_g^{E,M})$ and we let $\lambda_0^O := (\delta_0, \theta^{w,F}, \theta^{w,M}, \theta^{b,F,M}, \theta^{b,M,F})$

Proposition 11 (Sufficient Conditions for Identification of $\lambda_0^O := (\delta_0, \theta^{w,F}, \theta^{w,M}, \theta^{b,F,M}, \theta^{b,M,F})$)
Provided $\delta_0 \neq 0$, if Assumptions 1†, 2† and 3† hold and if one of the two following conditions is satisfied

1. *there exist at least 4 different vectors of shares $(s_1^M, s_1^{E,F}, s_1^{E,M})$, $(s_2^M, s_2^{E,F}, s_2^{E,M})$, $(s_3^M, s_3^{E,F}, s_3^{E,M})$ and $(s_4^M, s_4^{E,F}, s_4^{E,M}) \in (0,1)^3$ that have positive probability mass and for which both τ^M and τ^F are well-defined*
2. *there is a continuum of shares $I_s \subseteq [0,1]\times[0,1]\times[0,1]$ such that $P(s\in I_s)>0$ and both $\tau^F(s)$ and $\tau^M(s)$ are well-defined, for any $s\in I_s$*

then λ_0^O is the unique vector defined on Θ that satisfies the moment condition

$$\mathrm{E}\left[u^O(\bar{y}_g^M,\bar{y}_g^F,D_g,s_g;\lambda_0)|s_g=s\right] = \begin{pmatrix}\mathrm{E}\left[u^{O,M}(\bar{y}_g^M,D_g,s_g;\lambda_0^O)|s_g=s\right]\\ \mathrm{E}\left[u^{O,F}(\bar{y}_g^F,D_g,s_g;\lambda_0^O)|s_g=s\right]\end{pmatrix} = 0 \qquad (70)$$

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

$$u^{O,M}(\bar{y}_g^M,D_g,s_g;\lambda) = \omega^O(D_g,s_g)\bar{y}_g^M - \frac{\delta_0\left(s_g^{E,F}\theta_0^{b,M,F}(1-s_g^M)+s_g^{E,M}\left(1-\theta^{w,F}(1-s_g^M)\right)\right)}{1-s_g^M\theta_0^{w,M}-(1-s_g^M)\theta_0^{w,F}+s_g^M(1-s_g^M)\left(\theta_0^{w,M}\theta_0^{w,F}-\theta_0^{b,M,F}\theta_0^{b,F,M}\right)}$$

$$u^{O,F}(\bar{y}_g^F,D_g,s_g;\lambda) = \omega^O(D_g,s_g)\bar{y}_g^F - \frac{\delta_0\left(s_g^{E,F}(1-\theta^{w,M}s_g^M)+s_g^{E,M}\theta^{b,F,M}s_g^M\right)}{1-s_g^M\theta_0^{w,M}-(1-s_g^M)\theta_0^{w,F}+s_g^M(1-s_g^M)\left(\theta_0^{w,M}\theta_0^{w,F}-\theta_0^{b,M,F}\theta_0^{b,F,M}\right)}$$

and $\omega^O(D_g,s_g) = \dfrac{D_g-P(D_g=1|s_g)}{P(D_g=1|s_g)(1-P(D_g=1|s_g))}$

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: communication.crest@ensae.fr
https://crest.science/