UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

An analysis of the distributive effects of public policies and their spillovers

Alan André Borges da Costa
Orientador: Prof. Dr. Ricardo Madeira
Co-orientador: Prof. Dr. Sergio Firpo

São Paulo
2020

Prof. Dr. Vahan Agopyan
Reitor da Universidade de São Paulo

Prof. Dr. Fábio Frezatti
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. José Carlos de Souza Santos
Chefe do Departamento de Economia

Prof. Dr. Ariaster Baumgratz Chimeli
Coordenador do Programa de Pós-Graduação em Economia

ALAN ANDRÉ BORGES DA COSTA

An analysis of the distributive effects of public policies and their spillovers

Tese apresentada ao Programa de Pós-Graduação em Economia do Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo como requisito parcial para a obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. Ricardo Madeira;

Co-orientador: Prof. Dr. Sergio Firpo.

Versão Corrigida

**São Paulo
2020**

Catálogo na Publicação (CIP)
Ficha Catalográfica com dados inseridos pelo autor

Costa, Alan André Borges da.
An analysis of the distributive effects of public policies and their spillovers
/ Alan André Borges da Costa. - São Paulo, 2020.
117 p.

Tese (Doutorado) - Universidade de São Paulo, 2020.
Orientador: Ricardo Madeira.
Co-orientador: Sérgio Firpo.

1. Econometria. 2. Quantil. 3. Saturação. 4. Transbordamento. I.
Universidade de São Paulo. Faculdade de Economia, Administração e
Contabilidade. II. Título.

Resumo

O objetivo deste trabalho consiste em definir, identificar e estimar os efeitos diretos, indiretos, totais e de saturação (proporção tratados) nos diversos quantis da variável de resultado na presença de spillover do tratamento. Explorando a variação advinda da aleatorização em dois estágios propomos estimadores para medir as distâncias horizontais e verticais (quantil marginal) entre os grupos de tratamento e controle. Além disso, também definimos e identificamos a derivada do quantil marginal em relação à saturação, decompondo-a em efeito direto e spillover. Para finalizar, a aplicação empírica utiliza o banco de dados de um experimento aleatório realizado no México para estimar os objetos propostos usando os índices de participação e confiança em comunidades pobres como variáveis indicadoras de impacto. Os resultados mostram que existem efeitos diretos, positivos e negativos, para as caudas da distribuição. Um exercício adicional revela possíveis efeitos spillovers heterogêneos. Utilizando este banco de dados não é possível afirmar para um dado percentil que existe relação, positiva ou negativa, entre os quantis marginais e os diversos níveis de saturação, ou seja, não fica evidente que aumentar a proporção de tratados resultará em maiores efeitos, diretos ou spillover, do tratamento.

Palavras-chave: Quantil, Saturação, Transbordamento, Efeitos distributivos.

Abstract

The aim of this work is to define, identify and estimate the direct, indirect, total and saturation effects (proportion of treated subjects) on the several quantiles of the outcome variable in the presence of treatment spillover. Exploring the variation resulting from two stage randomization we propose estimators to measure the horizontal and vertical distances (marginal quantile) between treatment and control groups. In addition we defined, identified and estimated the derivative of the marginal quantile with respect to saturation by decomposing it into direct and spillover effects. Finally, the empirical application uses the database of a random experiment performed in Mexico that allows us to estimate the proposed objects using the participation and trust indexes in poor communities as outcome variables. The results show that there are direct effects, positive and negative, for the tails of the distribution. An additional exercise reveals evidence of heterogeneous spillover effects. With this database it is not possible to state for a given percentile that there is a relationship, positive or negative, between marginal quantiles and the several levels saturation, that is, it is not evident that increasing the proportion of treated subjects will result in greater (direct or spillover) effects of treatment.

Keywords: Quantile, Saturation, Spillover, Distributive effects.

Contents

1	Introduction	5
2	Potential outcomes and assumptions	9
3	Spillover and saturation: a distributive analysis	13
4	Estimation	23
5	Empirical Application	27
5.1	Description of the experiment and data	27
5.2	Descriptive statistics	29
5.3	Results	31
5.3.1	Direct effects (horizontal and vertical) and derivatives	31
5.3.2	Exercises 1 (estimation in the saturation interval) and 2 (choosing pure control group)	34
6	Conclusion	35
A	Figures	39
B	Tables	97
C	Proofs	102
C.1	Identification	102
C.2	Unconditional quantile	106

Chapter 1

Introduction

In several areas of knowledge the main research question is to estimate the causal effect of a treatment variable (D) on an outcome variable (Y). Using the potential outcome model (Rubin 1974), define the potential outcome for the individual i to be treated or untreated (control) respectively as $Y_i(1)$ and $Y_i(0)$. The individual impact and the Average Treatment Effects (ATE) can be denoted respectively as $Y_i(1) - Y_i(0)$ and $ATE = \mathbb{E}[Y_i(1) - Y_i(0)]$. Unfortunately it is not possible to observe the same individual in both situations (treated and control), that is, we have only $Y = D_i Y_i(1) + (1 - D_i) Y_i(0)$ where $D_i \in \{0, 1\}$. Thus, under the following assumption $(Y_i(1), Y_i(0)) \perp D_i$, we can use the sample analogues to estimate the average causal impact: $ATE = \mathbb{E}[Y|D_i = 1] - \mathbb{E}[Y|D_i = 0]$.

For the identification of the object ATE the assumption $(Y_i(1), Y_i(0)) \perp D_i$ has been used. In general this assumption will be satisfied when individuals are randomized to receive the treatment. When we perform a Randomized Control Trial (RCT) we can obtain a counterfactual that represents the treatment group: the randomization mechanism provides the balancing of the characteristics of the individuals excluding the possibility of selection in observable or not observable in both groups.

The causal effect of the treatment can be heterogeneous over the distribution of the outcome variable and distributive measures may become interesting to the researcher or policy maker. In this case, the Quantile Treatment Effect (QTE) estimators can capture different effects on the tails or in the middle of the distribution of Y . When we have a RCT we can directly use the measures proposed by Doksum (1974) and Lehmann (1975) defined as, for a given percentile $\tau \in (0, 1)$, the horizontal distance between the cumulative distribution functions of treatment (F_1) and control (F_0) groups: $QTE(\tau) = F_1^{-1}(\tau) - F_0^{-1}(\tau)$. Thus we have a complete description of the treatment effect for the entire distribution of the outcome variable.

The identification assumptions of ATE and QTE assume that the potential outcome of the individual i is not affected by the treatment selection mechanism and that the treatment received by an individual does not change the potential outcome of another

individual excluding interaction between units and possible effects of spillover (Cox 1958, Imbens & Wooldridge 2009, Rubin 1980). This assumption, known as Stable Unit Treatment Value Assumption (SUTVA), may not be satisfied when, for example, treatment selection occurs at an individual level but a RCT occurs at an aggregate level (with intersection among groups) allowing the transmission of information or contagion between the units of the treatment or control groups ¹. In this case, when the the SUTVA is not valid, the estimators of interest will be biased(Sobel 2006).

Thus it is necessary to rewrite the SUTVA assumption to allow interaction among individuals at some level, such as interference within groups (households, schools, firms, etc.), and excluding contacts among them. In this case it is possible to recover the desirable properties of the objects of interest and also estimate other effects.

The seminal work of Halloran & Struchiner (1991) defines the concepts currently used to estimate spillover: direct, indirect, total and overall effects. The authors argue that depending on how the treatment and control groups are formed and, due to the interactions between individuals, the direct effects can be calculated incorrectly. In this case it is important to measure the indirect effects to account for how much public policy has spillover.

Using the random voucher experiment of the Move to Opportunity (MTO) program, Sobel (2006) shows that it is not possible to estimate the measures defined in Halloran & Struchiner (1991) without using additional assumptions, otherwise the results will be biased due to spillovers and intersections between the groups. In this case it is necessary to partition the clusters and allow an individual to affect the outcome variable of another individual only if they are within the same group naming this assumption as “partial interference”. Given this assumption, and considering a two level random experiment, the author demonstrates the non bias of several spillover estimators.

From the works of Halloran & Struchiner (1991) and Sobel (2006) the literature was divided into works that contributed to demonstrate the properties of small and large samples. In the case of properties of small samples, the authors assume that there is a randomization design in two levels: at first stage the proportion or quantity of treated individuals is randomized into several groups, then in the second stage, the subjects are randomly selected within the group.

In this context Hudgens & Halloran (2008) under some assumptions, among them partial interference, formalize from the potential outcomes model, the direct, indirect, total and overall effects proving the absence of bias for all estimators. Tchetgen Tchetgen & VanderWeele (2010) use the same previous assumptions and extend the work of Hudgens & Halloran (2008) by calculating confidence intervals and defining the inverse probability

¹One example is the Mexican income transfer program: Progresa. The RCT was carried out at the aggregate level of villages in poor rural areas of Mexico. As the villages are not isolated the intersection between the groups, in this case areas, allowed interaction between the individuals generating program spillover for several control groups.

weighting estimator. Finally, Basse & Feller (2017) do not use the assumption of groups with homogeneous sizes and propose unbiased estimators for average effects and variance.

Recently other authors have proved the asymptotic properties for the spillover effects. Unlike previous authors, Vazquez-Bare (2017) does not consider the existence of two stage randomization to identify the parameters of interest. He estimates the objects non-parametrically, performs inference, and demonstrates the asymptotic properties. Lastly, Leung (2017) remove SUTVA assumption, treating the problem as a single network, allowing interference between groups (cross-cluster links). The author establishes the conditions for the estimators to be consistent and asymptotically normal.

The above works have concentrated efforts to correctly measure spillover effects from the definitions of Halloran & Struchiner (1991). Philipson (2000) defines the concepts of external and private average effects of treatment due to variations in the proportion of treated. To identify the populational parameters, the author states that it is necessary to conduct a randomized two level experiment to generate variation between groups and between individuals. According to the author, the main difference with respect to Halloran & Struchiner (1991) is the incorporation of the change in distribution, for the average, due to variation of the proportion of treated.

Despite the advances made, there are still no papers formalizing the distributive effects of treatment spillover in two stage randomized experiments. Thus, in this context, the following general questions arise: i) how to define and identify the direct, indirect and total effects of treatment for quantiles? ii) what is the effect of the proportion of treated over the distribution of the outcome variable?

This work is interested in the saturation of the treatment by groups (proportion of treated) and uses the assumptions related to multilevel experiments to identify the effects of spillover (the details of the saturation randomization design will be discussed in the next chapter). The first central idea is to define, identify and estimate the direct, indirect and total effects of treatment on the percentiles of the outcome variable, similar to the average effects of Halloran & Struchiner (1991). The identified objects show how to measure the impact of the difference between the treatment and control groups (horizontal distance) and also on the marginal quantile (vertical distance) of the outcome variable. The effects of the interaction between the individuals may be different along the distribution, because depending on the type of program, the left tail, center or right tail may be affected in an unequal way. It may not be necessary to treat the entire target public, since the interaction between individuals may overflow the results for the other members and this effect may be different throughout the distribution. Thus, it becomes essential to have quantile estimators that allow to identify these distributive effects.

In addition we define and identify the derivative of the marginal quantile with respect to saturation by decomposing it into direct and spillover effects. This measure is important because we can estimate, for example, how the marginal quantile of the median changes

when changing the saturation within a group. When performing an experiment we can estimate the level of saturation that results in greater impacts on marginal quantiles. This measure can be used by policy makers to improve the design of public policies by allocating resources in such a way that the performance of the marginal distribution of the outcome variable, due to the treatment, is greater for the tails or center of the distribution.

Finally, the empirical application uses the database of a random experiment performed in Mexico that allows us to estimate the proposed objects using the participation and trust indexes in poor communities as outcome variables. The results show that there are direct effects, positive and negative, for the tails of the distribution. An additional exercise reveals evidence of heterogeneous spillover effects. With this database it is not possible to state for a given percentile that there is a relationship, positive or negative, between marginal quantiles and the several levels saturation, that is, it is not evident that increasing the proportion of treated subjects will result in greater (direct or spillover) effects of treatment.

The main contributions of this work are: i) to define, identify and estimate the direct, indirect, total and saturation effects (proportion of treated subjects) on the several quantiles measuring the horizontal and vertical distances (marginal quantile) between the treatment and control groups and ii) to define, identify and estimate the derivative of the marginal quantile with respect to saturation by decomposing it into direct and spillover effects.

In addition to this introduction, this work presents the potential outcomes and their assumptions in chapter 2, the objects of interest in chapter 3, estimation in chapter 4, the chapter 5 shows the empirical application and finally section 6 addresses the conclusion.

Chapter 2

Potential outcomes and assumptions

The aim of this chapter is to define the potential outcomes and the assumptions in the context of *Random Saturation Design* (RSD) (Baird, Bohren, McIntosh & Ozler 2015, Crépon, Duflo, Gurgand, Rathelot & Zamora 2013).

Let i , g and n represent respectively the individuals, $i = 1, 2, \dots, n_g$, groups of allocation, $g = 1, 2, \dots, G$, and the total number of individuals $n = n_1 + n_2 + \dots + n_G = \sum_{g=1}^G n_g$. Denote S_g as the random variable representing the saturation level and its realization as π_g . The support of saturation is written as Π and therefore S_g and $\pi_g \in \Pi \subset [0, 1]$. The random variable of treatment assignment and its realization will be denoted as D_{ig} and d_{ig} . They may assume values D_{ig} and $d_{ig} \in \mathcal{D} \subset \{0, 1\}$.

In the first stage of RSD the proportion of treated individuals, π_g , are allocated randomly across several different groups. Because in most setup G is relatively small, one typically chooses few different values of π_g . It is important to notice that those G groups were not randomly formed and existed before the intervention.

In the second stage, for each group g , one selects randomly $n_g \pi_g = \sum_{i=1}^{n_g} D_{ig}$ subjects to form the “treatment group” of g . The remainder $n_g(1 - \pi_g) = \sum_{i=1}^{n_g} (1 - D_{ig})$ are not treated, and form the “control group” of g .

The saturation vector of the random variable and the realization will be denoted respectively as $\mathbf{S} = [S_1, S_2, \dots, S_G]'$ and $\boldsymbol{\pi} = [\pi_1, \pi_2, \dots, \pi_G]'$. Following the notation similar to Vazquez-Bare (2017) denote as $\mathbf{D}_g = [D_{1g}, D_{2g}, \dots, D_{n_gg}]'$ the vector of length n_g of treatment assignments. The vector of size $n_g - 1$ containing the treatment status of the neighbors of i is given by $\mathbf{D}_{(i)g} = [D_{1g}, D_{2g}, \dots, D_{i-1g}, D_{i+1g}, \dots, D_{n_g-1g}]'$, so $\mathbf{D}_g = [D_{ig}, \mathbf{D}_{(i)g}]'$. The realization of vector $\mathbf{D}_{(i)g}$ is denoted as $\mathbf{d}_{(i)g} = [d_{1g}, d_{2g}, \dots, d_{i-1g}, d_{i+1g}, \dots, d_{n_g-1g}]' \in \{0, 1\}^{n_g-1}$ and we have therefore $\mathbf{d}_g = [d_{ig}, \mathbf{d}_{(i)g}] = [d_{ig}, d_{1g}, d_{2g}, d_{3g}, \dots, d_{i-1g}, d_{i+1g}, \dots, d_{n_g-1g}]'$. The vector of all treatment assignments in all groups and their realization is given by $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_G]$ and $\mathbf{d} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_G]$.

Let $Y_{ig} \in \mathcal{Y} \subset \mathbb{R}$ be the outcome variable for individual i at group g . It can be written as:

$$Y_{ig} = Y_{ig}(\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_{g-1}, \mathbf{D}_g, \mathbf{D}_{g+1}, \dots, \mathbf{D}_{G-1}, \mathbf{D}_G) \quad (1)$$

equation (1) depends on the entire vector of treatment assignments. We then impose some assumptions in order to precisely write Y_{ig} as a function of potential outcomes. Then the first restriction in equation (1) is given by:

Assumption 1 (A1: Interference at group level). *There is interference only within the group:*

$$Y_{ig}(\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_G) = Y_{ig}^*(\mathbf{D}_g)$$

This assumption excludes spillover between groups, but allows spillover within the group. In this case, the outcome variable remains unchanged for all possible values of treatment outside the group the individual belongs to. Whether these groups have some contact, such as in neighborhoods or households, this assumption may not be valid, but if the groups are disjoint, such as schools, then there is probably no spillover between these units. Several authors have also used this approach, and each one adapted the assumption conform their question (Baird, Bohren, McIntosh & Ozler 2015, Graham, Imbens & Ridder 2010, Hudgens & Halloran 2008, Liu & Hudgens 2014, Manski 2013, Sobel 2006). For example, Manski (2013, p.4-5) and Sobel (2006, p.1405) call this assumptions respectively as “*constant treatment response*” and “*partial interference assumption*”. Similar to Vazquez-Bare (2017), under the assumption **A1** and for a given realization $\mathbf{d}_g = [d_{ig}, \mathbf{d}_{(i)g}] = [d_{ig}, d_{1g}, d_{2g}, d_{3g}, \dots, d_{i-1g}, d_{i+1g}, \dots, d_{n_g-1g}]'$ the equation (1) can be rewritten as:

$$Y_{ig} = \sum_{d_{ig} \in \{0,1\}} \sum_{\mathbf{d}_{(i)g} \in \mathcal{D}} Y_{ig}^*(d_{ig}, \mathbf{d}_{(i)g}) \mathbb{I}\{D_{ig} = d_{ig}\} \mathbb{I}\{\mathbf{D}_{(i)g} = \mathbf{d}_{(i)g}\} \quad (2)$$

This equation depend on the vector $\mathbf{d}_{(i)g}$ that contains information about the treatment assignment for all individuals within the same group. Individual i is connected to all other $n_g - 1$ individuals in that group by a given network structure. Thus, the fact that a given individual j and not individual k is treated affects i differently than another configuration in which individual k and not individual j is treated, where both j and k belong to the same group g as i . In other words the identity of the individual who is being treated matters.

In this context, if one does not have any information on the network structure, it is necessary to discuss what kind of assumptions about the potential outcomes are necessary to avoid using that information.

We can rewrite this structure for including the treatment saturation and reducing the dimensionality of equation (2). Thus define the potential outcomes for individual i at the saturation level $S_g = \pi_g$ for treated and control status respectively as:

$$\begin{aligned} Y_{ig}^{**}(1, \pi_g) \\ Y_{ig}^{**}(0, \pi_g) \end{aligned} \tag{3}$$

then consider the following assumption:

Assumption 2 (A2: interference between individuals inside the group). *For a realization $\mathbf{d}_g = [d_{ig}, \mathbf{d}_{(i)g}]$ there is a function $Y_{ig} : \{0, 1\} \times \{0, 1\}^{n_g-1} \rightarrow \mathbb{R}$ such that:*

$$Y_{ig}^*(d_{ig}, \mathbf{d}_{(i)g}) = Y_{ig}^{**}(d_{ig}, \pi_g)$$

$$\text{satisfying } \sum_{i=1}^{n_g} \frac{d_{ig}}{n_g} = \pi_g.$$

The **A2** assumption says that, conditional in saturation level ($\sum_{i=1}^{n_g} \frac{d_{ig}}{n_g} = \pi_g$), the potential outcome for individual i at group with treatment vector $\mathbf{d}_{(i)g}$ will be equal to the potential outcome with saturation level at $S_g = \pi_g$.

Note that perhaps a previous network may exist, but the random experiment turns information coming from the contacts irrelevant to explain the observed outcome. Additionally, it is implicit that the treatment does not induce the formation of new networks. Thus, the two level experiment and the above assumption allows to ignore the lack of information about network structure and, in this case, only the proportion of treated units (saturation) become relevant to determine the potential outcomes.

Consider the following example related to **A2** assumption: suppose there is a group with three individuals, with saturation level at $S_g = 2/3$ and that individual i has been treated. Then, the potential outcome can be written as $Y_{ig}^*(d_{ig}, d_1, d_2)$. Thus, assume **A2** is equivalent to stating that $Y_{ig}^*(1, 1, 0) = Y_{ig}^*(1, 0, 1) = Y_{ig}^{**}(1, \pi_g)$. Whether, for example, we reduce the saturation level to $S_g = 1/3$ and suppose that individual i was not treated, then $Y_{ig}^*(0, 1, 0) = Y_{ig}^*(0, 0, 1) = Y_{ig}^{**}(0, \pi_g)$.

The **A2** assumption reduces the dimensionality of the potential outcome and the codomain of $Y_{i,g}^*$ become $Y_{ig}^{**} : \{0, 1\} \times [0, 1] \rightarrow \mathbb{R}$. Thus, assuming **A2** we can rewrite equation (2) as:

$$Y_{ig}^{**}(d_{ig}, \pi_g) \tag{4}$$

The equation (4) shows that individual's potential outcome will depend on his own treatment status, but will also be a function of the proportion of treated individuals within his group. Using this structure we would like to compare, for example, the potential outcome of individual i when he participates in the treatment with his respective counterfactual: $Y_{ig}^{**}(1, \pi_g) - Y_{ig}^{**}(0, \pi_g)$.

Under **A1** and **A2** assumption the observed outcome of the individual i within group g if $S_g = \pi_g$ can be written as:

$$Y_{ig} = \sum_{\pi_g \in \Pi} \left\{ Y_{ig}^{**}(1, \pi_g) D_{ig} \mathbb{I}\{S_g = \pi_g\} + Y_{ig}^{**}(0, \pi_g) (1 - D_{ig}) \mathbb{I}\{S_g = \pi_g\} \right\} \quad (5)$$

where $\mathbb{I}\{\cdot\}$ is the indicator function which assumes the value 1 at saturation level $S_g = \pi_g$.

Hence, if $Y_{ig}^{**}(d_{ig}, \pi_g)$ is independent of D_{ig} given $S_g = \pi_g$, and $d_{ig} = 0, 1$, then

$$\mathbb{E}[Y_{ig}^{**}(d_{ig}, S_g = \pi_g) | D_{ig} = d_{ig}, S_g = \pi_g] = \mathbb{E}[Y_{ig}^{**}(d_{ig}, S_g = \pi_g) | S_g = \pi_g]$$

this is ensured if there is randomization of who will receive the treatment, since the proportion is fixed as $S_g = \pi_g$. Finally, because of the randomization of the proportion π_g among the groups

$$\mathbb{E}[Y_{ig}^{**}(d_{ig}, S_g = \pi_g) | S_g = \pi_g] = \mathbb{E}[Y_{ig}^{**}(d_{ig}, S_g = \pi_g)]$$

The following section address the measures that relate the saturation of treatment with the effects along the distribution of Y .

Chapter 3

Spillover and saturation: a distributive analysis

The main purpose of this chapter is to define and identify several objects for quantiles in the context of spillover.

The seminal work of Halloran & Struchiner (1991) defines the concepts currently used to estimate spillover: direct, indirect, total and overall effects. Other authors estimated these objects and proved absence of bias for the mean and variance (Basse & Feller 2017, Hudgens & Halloran 2008, Sobel 2006). Recently, Vazquez-Bare (2017) and Basse & Feller (2017) analyzed some questions about average effects such as inference and properties of the estimators in large sample. However, despite advances, effects along the distribution for two-level randomized trials have not yet been studied.

In general, let $F_Y = \mathbb{P}(Y \leq y)$ and $F_{Y|X} = \mathbb{P}(Y \leq y|X)$ be the cumulative, unconditional and conditional respectively, distributions of $Y \in \mathbb{R}$, where X is any random variable. Let τ a real number where $\tau \in (0, 1)$, then the τ^th unconditional and conditional quantile of Y is defined respectively as $q_\tau = F_Y^{-1}(\tau) = \inf_q \{\mathbb{P}[Y \leq q] \geq \tau\}$ and $q_{\tau|X} = F_{Y|X}^{-1}(\tau) = \inf_q \{\mathbb{P}[Y \leq q | X] \geq \tau\}$. Thus, we can write the quantile for potential outcome d , at group π_g for a given percentile τ as:

$$q_{d,\pi_g,\tau} = \inf_q \left\{ \mathbb{P}[Y_{ig}^{**}(d_{ig}, \pi_g) \leq q] \geq \tau \right\} \quad (6)$$

in a two-level experiment the saturation level is independent of the potential outcome and therefore: $q_{d,\pi_g,\tau|S_g=\pi_g} = q_{d,\pi_g,\tau}$.

From the equation (6) some objects of interest can be written. For $S_g = \pi_g > 0$ and $S_g = \pi'_g \geq 0$ we define the **Quantile Direct Effect (QDE)**, **Quantile Indirect Effect (QIE)**, **Quantile Total Effect (QTE)** and **Quantile Saturation Effects (QSE)** as follows:

$$\varphi_{\pi_g,\tau}^D = q_{1,\pi_g,\tau} - q_{0,\pi_g,\tau} \quad (7)$$

$$\varphi_{\pi_g, \tau}^I = q_{0, \pi_g, \tau} - q_{0, 0, \tau} \quad (8)$$

$$\varphi_{\pi_g, \tau}^T = q_{1, \pi_g, \tau} - q_{0, 0, \tau} \quad (9)$$

$$\varphi_{d, \pi_g, \pi'_g, \tau}^S = q_{d, \pi_g, \tau} - q_{d, \pi'_g, \tau} \quad (10)$$

note that the φ effects compare the quantile of the cumulative distribution between the groups for a given fixed percentile. These sets of objects will be referred to as horizontal distances. We can calculate these objects for each $\tau \in (0, 1)$ by obtaining a curve describing the effects for the entire distribution.

The last step to obtain the objects QDE, QIE, QTE and QSE is to show how we can write the quantiles as a function of the observed data coming from the random experiment. To identify these effects it will be necessary an additional assumption:

Assumption 3 (A3: independence of potential outcomes). *The potential outcomes are independent of treatment allocation and saturation level:*

$$Y_{ig}^{**}(d_{ig}, \pi_g) \perp (D_{ig}, S_g); \forall (d_{ig}, \pi_g) \in \{0, 1\} \times [0, 1]$$

the assumption **A3** will be satisfied if the random experiment is conducted on two level as described in the previous section.

Finally, the assumptions **A1-A3** allows us to rewrite the above objects as a function of the observed variables, which can be seen in lemma 1:

Lemma 1 (Identification of quantiles (horizontal distances)). *Under assumptions **A1-A3** the quantiles can be written as:*

$$\begin{aligned} \tau &= \mathbb{E} \left[\frac{D_{ig}}{\pi_g} \frac{\mathbb{I}\{S_g = \pi_g\}}{\mathbb{P}(S_g = \pi_g)} \mathbb{I}\{Y_{ig} \leq q_{1, \pi_g, \tau}\} \right] \\ \tau &= \mathbb{E} \left[\frac{(1 - D_{ig})}{(1 - \pi_g)} \frac{\mathbb{I}\{S_g = \pi_g\}}{\mathbb{P}(S_g = \pi_g)} \mathbb{I}\{Y_{ig} \leq q_{0, \pi_g, \tau}\} \right] \\ \tau &= \mathbb{E} \left[(1 - D_{ig}) \frac{\mathbb{I}\{S_g = 0\}}{\mathbb{P}(S_g = 0)} \mathbb{I}\{Y_{ig} \leq q_{0, 0, \tau}\} \right] \end{aligned}$$

Proof. See appendix □

from this lemma, the φ effects can be identified since all variables are observable for the researcher:

Corollary 1 (Identification of QDE, QIE, QTE and QSE). *Under assumptions A1-A3 the quantiles effects, $\varphi_{\pi_g, \tau}^D$, $\varphi_{\pi_g, \tau}^I$, $\varphi_{\pi_g, \tau}^T$, $\varphi_{d, \pi_g, \pi'_g, \tau}^S$, are identified.*

Proof. The proof is direct since they are functional of the data. $\square$

Using QDE, QIE, QTE, and QSE, we can capture several horizontal distances by forming a complete drawing of spillover effects. Such measures are similar to the first definitions contained in Halloran & Struchiner (1991), however, the φ effects address the entire distribution¹ of Y and apply the concepts to the random saturation design case.

Comparisons can be made using several levels of saturation providing a better understanding of spillover effects throughout the distribution. Figure 3.1 is illustrative and aims to exemplify the types of analysis that can be made from these definitions. Suppose we have two saturated group at the following levels: $(\pi_1, \pi_2) = (> 0, 0)$. Before treatment the individuals are identical in observable and unobservable characteristics and after the treatment we observe the distributions in the figure 3.1. The graphs show the cumulative distribution for the pure control group, resulting from null saturation ($\pi_2 = 0$) (green curve), the treatment (red curve) and (within) control group (black curve) with positive saturation ($\pi_1 > 0$) (note the spillover on the tail).

Figure 3.1: Cumulative distribution for two levels of saturation

Source: elaborated by the author

For a given percentile we can have different direct, indirect or total effects at each point in the distribution. In the absence of spillover it makes no difference to use the pure or within control group and, therefore, we would return to the estimators already defined in the literature (Doksum 1974, Firpo 2007, Lehmann 1975).

The interesting case arises when comparing the tails of the distributions. Note that it is not possible to identify, a priori, how spillover affects the treatment and control groups in the black and red curves, since an interaction may occur that benefits (or not)

¹Note that we can recover the average through the quantile because $F(Y) = U \Rightarrow Y = F^{-1}(U)$ and $\int_{-\infty}^{+\infty} Y f(Y) dY = \int_0^1 F^{-1}(U) dU$.

individuals in both groups. Thus, the effect of treatment (direct, indirect or total) may be heterogeneous and such measures will capture differences along the distribution, leading to a better understanding of the interaction between individuals.

The previous measures compare the horizontal distances between the groups with different saturation levels. It may also be of interest to the researcher (or policy maker) to estimate the direct, indirect and total effects on the marginal (unconditional)² quantile in a given group.

We know that when estimating a regression, due to the law of iterated expectations, the estimated parameter can also be interpreted as the (average) effect on the marginal distribution of Y . The same is not true for quantile because there is no the "law of iterated quantile". In this case it may be that the effect of the difference between treatment and control groups is different from the effect on the quantile on unconditional distribution.

Suppose, for example, a school program applied in a classroom with saturation level at $S_g = \pi_g$. The program may have no direct effect on conditional quantiles, however, due to the interaction of individuals, there may be some effect on the unconditional quantile on this classroom outcome variable. Otherwise: does the treatment make the classroom perform (notice that it is not the difference between groups), along the distribution, increase (reduce)? We can do similar reasoning for different programs, such as training in the labor market and the effects on the marginal distribution of wages; conditional cash transfers and the effects on the marginal distribution of school performance (or health indicators); primary care programs in community centers and the marginal distribution of interaction in the community (measured through the local participation index); etc.

We can go further and think of marginal measures as representing a group's well-being. In this sense, measuring the causal effect of treatment on the marginal quantile, within a group, is similar to measuring the well-being of the group for a given percentile, as we are looking at the marginal distribution of the outcome variable. So, we can ask several questions about the relationship between the effects of treatment on marginal quantile: i) is the group's well-being, measured through marginal distribution, higher (lower) due to the program?, ii) Does the saturation have a greater impact on the marginal distribution on the left tail than on the right tail? If so, can we say that the saturation generates an increase in well-being? iii) What is the total, indirect and direct effect on the marginal quantile? Can these effects be related to well being in a group? etc. In this case, we could estimate whether the spillover have positive (negative) impact on the marginal quantile, generating indirect welfare effects.

Estimating the impact on the marginal quantile is similar to the work of Firpo, Fortin & Lemieux (2009), however, the authors do not define measures for spillover effects. Using the previously defined structure the marginal quantile will be formalized below and then new objects will be defined and identified in the presence of spillover.

²The marginal or unconditional quantile is referring to the quantile of the marginal distribution of Y .

We know that the marginal quantile, for a given group g and a fixed percentile (τ), will be located between the distributions of the treatment and control group. We can use the law of iterated expectations³ (Wooldridge 2010, p. 31) and write the cumulative distribution function and the marginal quantile respectively as:

$$\begin{aligned}\mathbb{F}_{Y_{ig}}(q_{\pi_g, \tau}) &= \pi_g \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1] + \\ &+ (1 - \pi_g) \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0] = \tau\end{aligned}\quad (11)$$

$$\mathbb{F}_{Y_{ig}}^{-1}(\tau) = q_{\pi_g, \tau} \quad (12)$$

Note that the cumulative distribution evaluated at this point ($q_{\pi_g, \tau}$) is a combination of both curves weighted by the proportion of treated (control) subjects. The effect of treatment and saturation on the marginal quantile ($q_{\pi_g, \tau}$) should take into account the characteristics of both distributions.

Using the previous structure and the equation (12) we can write the proportion of individuals below the marginal quantile for the potential outcome d in the group π_g for a given percentile τ as:

$$p_{d, \pi_g, \tau} = \mathbb{P}[Y_{ig}^{**}(d_{ig}, \pi_g) \leq q_{\pi_g, \tau}] \quad (13)$$

equation (13), unlike the previous objects, provides the vertical distance in a distribution. We can see, similar to the literature of potential outcome, that we will have a missing variable problem applied to the proportion $p_{d, \pi_g, \tau}$.

Define the density of the distribution $\mathbb{F}_{Y_{ig}}$ evaluated at marginal quantile as $f_{Y_{ig}}(q_{\pi_g, \tau})$. For $S_g = \pi_g > 0$ and $S_g = \pi'_g \geq 0$ we define new objects of interest respectively as follows: **Marginal Quantile Direct Effect (MQDE)**, **Marginal Quantile Indirect Effect (MQIE)**, **Marginal Quantile Total Effect (MQTE)** and **Marginal Quantile Saturation Effect (MQSE)**

$$\varphi_{\pi_g, \tau}^{MD} = \frac{-1}{f_Y(q_{\pi_g, \tau})} (p_{1, \pi_g, \tau} - p_{0, \pi_g, \tau}) \quad (14)$$

$$\varphi_{\pi_g, \tau}^{MI} = \frac{-1}{f_Y(q_{\pi_g, \tau})} (p_{0, \pi_g, \tau} - p_{0, 0, \tau}) \quad (15)$$

$$\varphi_{\pi_g, \tau}^{MT} = \frac{-1}{f_Y(q_{\pi_g, \tau})} (p_{1, \pi_g, \tau} - p_{0, 0, \tau}) \quad (16)$$

³We can also use the law of total probability to get the same result (Shiryaev 1995, p. 25).

$$\varphi_{d,\pi_g,\pi'_g,\tau}^{MS} = \frac{-1}{f_Y(q_{\pi_g,\tau})} (p_{d,\pi_g,\tau} - p_{d,\pi'_g,\tau}) \quad (17)$$

note that the φ^M effects compare the proportion of the cumulative distribution between the groups for a given percentile. These sets of objects will be referred to as vertical distances. We did not observe, at the same time, the potential outcome in the condition of the individual belonging to the treatment and control group. Thus, it is also not possible to observe these proportions in both conditions and, therefore, it is necessary to show how to identify these objects.

The next lemma and corollary identify the φ^M effects under the assumptions **A1-A3**:

Lemma 2 (Identification of proportions (vertical distances)). *Under assumptions **A1-A3** the proportions can be written as:*

$$\begin{aligned} p_{1,\pi_g,\tau} &= \mathbb{E} \left[\frac{D_{ig} \mathbb{I}\{S_g = \pi_g\}}{\pi_g \mathbb{P}(S_g = \pi_g)} \mathbb{I}\{Y_{ig} \leq q_{\pi_g,\tau}\} \right] \\ p_{0,\pi_g,\tau} &= \mathbb{E} \left[\frac{(1 - D_{ig}) \mathbb{I}\{S_g = \pi_g\}}{(1 - \pi_g) \mathbb{P}(S_g = \pi_g)} \mathbb{I}\{Y_{ig} \leq q_{\pi_g,\tau}\} \right] \\ p_{0,0,\tau} &= \mathbb{E} \left[\frac{(1 - D_{ig}) \mathbb{I}\{S_g = 0\}}{(1 - \pi_g) \mathbb{P}(S_g = 0)} \mathbb{I}\{Y_{ig} \leq q_{\pi_g,\tau}\} \right] \end{aligned}$$

Proof. See appendix □

Corollary 2 (Identification of MQDE, MQIE, MQTE and MQSE). *Under assumptions **A1-A3** the marginal quantiles effects, $\varphi_{\pi_g,\tau}^{MD}$, $\varphi_{\pi_g,\tau}^{MI}$, $\varphi_{\pi_g,\tau}^{MT}$, $\varphi_{d,\pi_g,\pi'_g,\tau}^{MS}$ are identified.*

Proof. The proof is direct since they are functional of the data. □

the lemma 2 shows how to write the proportions as a function of observable variables for the researcher. The weights are the same as the horizontal distances, however, the indicator function is evaluated in the marginal quantile for all measures returning the proportion of individuals per group. Finally, it is clear that the objects identified in lemmas 1 and 2 are simple averages as a function of the sample size of the groups. Note that the weights work as locators for the two levels of randomization, showing where is the individual (individual weight) and in which saturation group it is located (group weight).

Finally, it remains to interpret these differences. The next figures illustrate the vertical distances

Figure 3.2: Vertical differences between proportions for within control group

Figure 3.3: Vertical differences between proportions for pure control group

Graph (3.2) shows the cumulative distributions for the treatment and control groups at the same saturation level $S_g = \pi_g$. Define $q'_{\pi_g, \tau} = q_{\pi_g, \tau} + \delta$ and write the slope for this graph as $\frac{p_{0, \pi_g, \tau} - p_{1, \pi_g, \tau}}{q'_{\pi_g, \tau} - q_{\pi_g, \tau}} = m$. Simplifying we have

$$-\frac{p_{1, \pi_g, \tau} - p_{0, \pi_g, \tau}}{m} = q_{\pi_g, \tau} - q'_{\pi_g, \tau} = \delta$$

where $m = [\pi_g f_Y(q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1) + (1 - \pi_g) f_Y(q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0)] = f_Y(q_{\pi_g, \tau})$ is the slope of the cumulative distribution (density) used in the vertical distances. Note that the fraction on the left side is the representation of the MQDE and shows the direct effect of the treatment on the marginal quantile.

Figure 3.3 is similar to the previous one and adds to the graph the pure control group (green curve) whose saturation is $S_g = 0$. Then, define $q''_{\pi_g, \tau} = q_{\pi_g, \tau} + \delta + \gamma$ and following the same steps we can show that

$$-\frac{p_{0, \pi_g, \tau} - p_{0, 0, \tau}}{m} = q_{\pi_g, \tau} - q''_{\pi_g, \tau} = \gamma$$

getting the definition of the MQIE. We can measure the vertical distances for each percentile by obtaining a complete description of the effects of treatment on the marginal

distribution. Finally, we say that the increase (decrease) in the marginal quantile can show the increase (decrease) in well-being within the group as we are measuring the impact of treatment on the group as a whole, not the difference between treatment and controle groups.

We can go beyond and explore the relationship between variation in saturation and variation in marginal quantile within a group g . This measure is important because we can estimate, for example, how the marginal quantile of the median changes when changing the saturation within a group, more specifically, the derivative of the marginal quantile with respect to saturation. Would it be possible to decompose this derivative by separating the direct effects from the spillover effects? The next paragraphs formalize these ideas and show how to decompose the derivative of the marginal quantile with respect to saturation for a given group g .

The work of Philipson (2000) proposes to divide the changes in the distribution of Y , more specifically $\mathbb{E}(Y_{ig}|d_{ig}, S_g = \pi_g)$, into two parts: i) a first part attributed to treatment status named as “private effects” and ii) a second part from changes in distribution due to proportions of treated denominated “external effects” (or spillover). The general idea is similar to this author applied to the marginal distribution of Y .

The proposal is to understand how the changes in saturation affect the changes in the quantile, that is, to estimate the following relationship $\frac{dq_{\pi_g, \tau}}{d\pi_g}$. For this we need to derive, implicitly, equation (11) with respect to π_g . The following lemma shows the effect of π_g on the marginal quantile ($q_{\pi_g, \tau}$):

Lemma 3 (Saturation effects on the marginal quantile). *Under assumptions A1-A2 the saturation effects on the marginal quantile can be written as:*

$$\begin{aligned} \varphi_{\pi_g, \tau}^u = & \overbrace{\frac{-1}{f_Y(q_{\pi_g, \tau})} \left(\mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1] - \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0] \right)}^{\text{Marginal Quantile Direct Effect (MQDE)}} + \\ & \underbrace{\frac{-1}{f_Y(q_{\pi_g, \tau})} \left(\pi_g \frac{\partial \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1]}{\partial \pi_g} + (1 - \pi_g) \frac{\partial \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0]}{\partial \pi_g} \right)}_{\text{Marginal Quantile Spillover Effect (MQSPE)}} \end{aligned} \quad (18)$$

where $\varphi_{\pi_g, \tau}^u = \frac{dq_{\pi_g, \tau}}{d\pi_g}$.

Proof. See appendix □

It can be seen that the relationship between the proportion of treated subjects and the marginal quantile can be decomposed on two parts called the Marginal Quantile Direct Effect (MQDE), difference between the proportions of treated and controls given the saturation level, and the Marginal Quantile Spillover Effect (MQSPE), defined by

the relation between the derivatives and the saturation level, that is, the change in the distribution due to saturation. It should be noted that the expansion of the program, in this case, does not necessarily increase or reduce $q_{\pi_g, \tau}$, since the final variation will depend on the sizes of MQDE, MQSPE and also the density of Y valued at $q_{\pi_g, \tau}$. The effects of saturation on each quantile may be different because, depending on type of program and the interaction, the left tail, center and right tail of the distribution may be unequally affected.

This measure, differently from the previous analysis, decomposes the saturation effects on the marginal quantile in two parts: the direct effect and the spillover effect. The contribution consists in identifying the vertical movement of the distributions (MQDE) and also the changes in the derivatives of the curves (MQSPE).

Analyzing equation (18) it is possible to identify the objects of interest with randomization, since the variation between groups generates MQSPE (note that if there is no variability between each g , then the derivatives with respect to π_g will be null) and the variation within each group, randomization of treatment, allows get the MQDE. If saturation is 100%, that is, $\pi_g = 1$, then, as expected, the marginal quantile depends only on the direct effects. Suppose that the $MQDE > 0$ and $MQSPE > 0$, that is, direct and spillover effects positive, then $\frac{dq_{\pi_g, \tau}}{d\pi_g} > 0 \Leftrightarrow MQDE < -MQSPE$. From the equation (18) several comparisons can be made by fixing a percentile and then conditioning the desire values of saturation and treatment.

The direct effect is exactly the same as previously presented and the idea of vertical distances is similar; it remains, therefore, to interpret the derivatives that are exemplified in the next graph.

Figure 3.4: Derivative and vertical distances

Source: elaborated by the author

Figure 3.4 shows, graphically, how to recover the derivatives of equation (18). In this case we will use the discrete approximation of the derivative to compare the saturation level of group A with another group B (suppose, without loss of generality, $\pi_A > \pi_B$). The approximation of the derivative conditional in π_A and $D_{iA} = 1$ is given by:

$$\begin{aligned}
& \frac{\partial \mathbb{P}[Y_{ig} \leq q_{\pi_A, \tau} | S_A = \pi_A, D_{iA} = 1]}{\partial \pi_A} \approx \\
& \approx \frac{\mathbb{P}[Y_{ig} \leq q_{\pi_A, \tau} | S_A = \pi_A, D_{iA} = 1] - \mathbb{P}[Y_{ig} \leq q_{\pi_A, \tau} | S_B = \pi_B, D_{iB} = 1]}{\pi_A - \pi_B} \approx \frac{p_{1, \pi_A, \tau} - p_{1, \pi_B, \tau}}{\pi_A - \pi_B}
\end{aligned}$$

Note that we are using the same idea that was presented in figure 3.2 and, therefore, we can estimate the derivative by vertical distances comparing two groups with different saturation level.

We can relate equation (18) to the marginal effects defined above. Rewriting the derivative we have

$$\varphi_{\pi_g, \tau}^u = \varphi_{\pi_g, \tau}^{MD} + \pi_g \left. \frac{\partial \varphi_{\pi_g, \tau}^{MD}}{\partial \pi_g} \right|_{S_g = \pi_g} + \left. \frac{\partial \varphi_{\pi_g, \tau}^{MI}}{\partial \pi_g} \right|_{S_g = \pi_g}$$

therefore, the derivative of the marginal quantile with respect to saturation can be decomposed into the direct effect and the rates of variation of the direct and indirect effects. It is clear that the spillover effect is the derivative from vertical distances (direct and indirect) showing how sensitive these measures are when changing the proportion of treated subjects.

After presenting and interpreting the objects of interest, it remains to discuss how to estimate the equations above.

Chapter 4

Estimation

In this chapter we will discuss the estimation of equations (7)-(10), (12), (14)-(17) and (18).

We propose to estimate the quantiles by weighting the cumulative distributions using the weights from lemmas 1 and 2. The idea of estimation is similar to the works of Firpo & Pinto (2015), Donald & Hsu (2014) and Fan & Liu (2016) among others. The main difference are the weights, while these authors estimate the propensity score in the first stage, this work weighted the above equations using proportions and unconditional binary variables.

Denote the estimated φ effects as:

$$\hat{\varphi}_{\pi_g, \tau}^D = \hat{q}_{1, \pi_g, \tau} - \hat{q}_{0, \pi_g, \tau} \quad (19)$$

$$\hat{\varphi}_{\pi_g, \tau}^I = \hat{q}_{0, \pi_g, \tau} - \hat{q}_{0, 0, \tau} \quad (20)$$

$$\hat{\varphi}_{\pi_g, \tau}^T = \hat{q}_{1, \pi_g, \tau} - \hat{q}_{0, 0, \tau} \quad (21)$$

$$\hat{\varphi}_{d, \pi_g, \pi'_g, \tau}^S = \hat{q}_{d, \pi_g, \tau} - \hat{q}_{d, \pi'_g, \tau} \quad (22)$$

By lemma 1 and 2 we can define the weights and the probabilities respectively as:

$$\hat{\omega}_i^{(1, \pi_g)} = \frac{D_{ig} \mathbb{I}\{S_g = \pi_g\}}{\pi_g \widehat{\mathbb{P}}(S_g = \pi_g)} \quad (23)$$

$$\hat{\omega}_i^{(0, \pi_g)} = \frac{(1 - D_{ig}) \mathbb{I}\{S_g = \pi_g\}}{(1 - \pi_g) \widehat{\mathbb{P}}(S_g = \pi_g)} \quad (24)$$

$$\hat{\omega}_i^{(0, 0)} = (1 - D_{ig}) \frac{\mathbb{I}\{S_g = 0\}}{\widehat{\mathbb{P}}(S_g = 0)} \quad (25)$$

$$\widehat{\mathbb{P}}(S_g = \pi_g) = \frac{1}{G} \sum_{g=1}^G \mathbb{I}\{S_g = \pi_g\} \quad (26)$$

$$\widehat{\mathbb{P}}(S_g = 0) = \frac{1}{G} \sum_{g=1}^G \mathbb{I}\{S_g = 0\} \quad (27)$$

Thus the proposed estimators can be written as weighted CDF and can be represented by sample analogue as:

$$\widehat{\mathbb{F}}_{Y_{ig}}^{(\widehat{\omega}_i^{(d, \pi_g)})}(q) = \frac{1}{n_{d,g}} \sum_{i=1}^{n_g} \left[\widehat{\omega}_i^{(d, \pi_g)} \mathbb{I}\{Y_{ig} \leq q\} \right] \quad (28)$$

and the estimated quantiles can be written as follows:

$$\widehat{q}_{d, \pi_g, \tau} = \inf_q \left\{ \widehat{\mathbb{F}}_{Y_{ig}}^{(\widehat{\omega}_i^{(d, \pi_g)})}(q) \geq \tau \right\} \quad (29)$$

we can see that the φ effects are differences between weighted minimizations. These weights, coming from lemma 1, take into account the random selection process of the individual as well as the random selection of the group. The variables are all observable and have already been defined.

The empirical distribution can be written as follows: $\widehat{\mathbb{F}}_{Y_{ig}}(q_{\pi_g, \tau}) = \tau$. Then the marginal quantile estimator can be defined as:

$$\widehat{\mathbb{F}}_{Y_{ig}}^{-1}(\tau) = \widehat{q}_{\pi_g, \tau} \quad (30)$$

denote the kernel density estimator as:

$$\widehat{f}_{Y_{ig}}(\widehat{q}_{\pi_g, \tau}) = \frac{1}{n_{\pi_g} b} \sum_{i=1}^n K\left(\frac{Y_{ig} - \widehat{q}_{\pi_g, \tau}}{b}\right) \quad (31)$$

where $K(\cdot)$ is the kernel function and b is the bandwidth. The density estimation follows the traditional methods proposed in Pagan & Ullah (1999) using the gaussian kernel and optimal bandwidth. Denote the estimated marginal effects (vertical distances) as:

$$\widehat{\varphi}_{\pi_g, \tau}^{MD} = \frac{-1}{\widehat{f}_Y(\widehat{q}_{\pi_g, \tau})} (\widehat{p}_{1, \pi_g, \tau} - \widehat{p}_{0, \pi_g, \tau}) \quad (32)$$

$$\widehat{\varphi}_{\pi_g, \tau}^{MI} = \frac{-1}{\widehat{f}_Y(\widehat{q}_{\pi_g, \tau})} (\widehat{p}_{0, \pi_g, \tau} - \widehat{p}_{0, 0, \tau}) \quad (33)$$

$$\widehat{\varphi}_{\pi_g, \tau}^{MT} = \frac{-1}{\widehat{f}_Y(\widehat{q}_{\pi_g, \tau})} (\widehat{p}_{1, \pi_g, \tau} - \widehat{p}_{0, 0, \tau}) \quad (34)$$

$$\hat{\varphi}_{d,\pi_g,\pi'_g,\tau}^{MS} = \frac{-1}{\hat{f}_Y(\hat{q}_{\pi_g,\tau})} (\hat{p}_{d,\pi_g,\tau} - \hat{p}_{d,\pi'_g,\tau}) \quad (35)$$

The estimated proportions can be written as follows:

$$\hat{p}_{1,\pi_g,\tau} = \frac{1}{n_{1,g}} \sum_{i=1}^{n_g} \left[\hat{\omega}_i^{(1,\pi_g)} \mathbb{I}\{Y_{ig} \leq \hat{q}_{\pi_g,\tau}\} \right] \quad (36)$$

$$\hat{p}_{0,\pi_g,\tau} = \frac{1}{n_{0,g}} \sum_{i=1}^{n_g} \left[\hat{\omega}_i^{(0,\pi_g)} \mathbb{I}\{Y_{ig} \leq \hat{q}_{\pi_g,\tau}\} \right] \quad (37)$$

$$\hat{p}_{0,0,\tau} = \frac{1}{n_{0,g}} \sum_{i=1}^{n_g} \left[\hat{\omega}_i^{(0,0)} \mathbb{I}\{Y_{ig} \leq \hat{q}_{\pi_g,\tau}\} \right] \quad (38)$$

The estimation of equation (18) can be divided into three parts: i) the difference denominated MQDE; ii) the spillover effect MQSPE and iii) the density evaluated in $q_{\pi_g,\tau}$ (equation (31)). Denote the estimator of the equation (18) as:

$$\begin{aligned} \hat{\varphi}_{\pi_g,\tau}^u = & \overbrace{\frac{-1}{\hat{f}_Y(\hat{q}_{\pi_g,\tau})} \left(\hat{\mathbb{P}}[Y_{ig} \leq \hat{q}_{\pi_g,\tau} | S_g = \pi_g, D_{ig} = 1] - \hat{\mathbb{P}}[Y_{ig} \leq \hat{q}_{\pi_g,\tau} | S_g = \pi_g, D_{ig} = 0] \right)}^{\text{Marginal Quantile Direct Effect (MQDE)}} + \\ & \underbrace{\frac{-1}{\hat{f}_Y(\hat{q}_{\pi_g,\tau})} \left(\pi_g \frac{\partial \hat{\mathbb{P}}[Y_{ig} \leq \hat{q}_{\pi_g,\tau} | S_g = \pi_g, D_{ig} = 1]}{\partial \pi_g} + (1 - \pi_g) \frac{\partial \hat{\mathbb{P}}[Y_{ig} \leq \hat{q}_{\pi_g,\tau} | S_g = \pi_g, D_{ig} = 0]}{\partial \pi_g} \right)}_{\text{Marginal Quantile Spillover Effect (MQSPE)}} \end{aligned} \quad (39)$$

we can write the the difference in MQDE as:

$$\begin{aligned} & \hat{\mathbb{P}}[Y_{ig} \leq \hat{q}_{\pi_g,\tau} | S_g = \pi_g, D_{ig} = 1] - \hat{\mathbb{P}}[Y_{ig} \leq \hat{q}_{\pi_g,\tau} | S_g = \pi_g, D_{ig} = 0] = \\ & = \frac{1}{n_{1,g}} \sum_{i=1}^{n_g} \left[\hat{\omega}_i^{(1,\pi_g)} \mathbb{I}\{Y_{ig} \leq \hat{q}_{\pi_g,\tau}\} \right] - \frac{1}{n_{0,g}} \sum_{i=1}^{n_g} \left[\hat{\omega}_i^{(0,\pi_g)} \mathbb{I}\{Y_{ig} \leq \hat{q}_{\pi_g,\tau}\} \right] \end{aligned} \quad (40)$$

finally, the discrete approximation for the derivative estimator for the treatment group will be given by:

$$\begin{aligned}
& \frac{\partial \widehat{\mathbb{P}}[Y_{ig} \leq \hat{q}_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1]}{\partial \pi_g} = \\
& = \frac{(n_{1,g})^{-1} \sum_{i=1}^{n_g} \left[\hat{\omega}_i^{(1, \pi_g)} \mathbb{I}\{Y_{ig} \leq \hat{q}_{\pi_g, \tau}\} \right] - (n_{1,g'})^{-1} \sum_{i=1}^{n_g} \left[\hat{\omega}_i^{(1, \pi'_{g'})} \mathbb{I}\{Y_{ig} \leq \hat{q}_{\pi_g, \tau}\} \right]}{\pi_g - \pi'_{g'}} \quad (41)
\end{aligned}$$

The next chapter addresses the empirical application using the database of a two-stage experiment (RSD) conducted in municipalities in Mexico.

Chapter 5

Empirical Application

5.1 Description of the experiment and data

This chapter will present the empirical application using McIntosh, Alegria, Ordonez & Zenteno (2018) data set experiment. The program to be analyzed, Hábitat, has the general objective of improving the local infrastructure and the quality of life of the inhabitants of poor regions of Mexico. During the years 2009-2011, around \$68 million were invested, of which approximately 20% was allocated to community social development.

The experiment was carried out by the federal government of Mexico with the support of states, municipalities and the local community. To understand the eligibility criteria and randomization it is necessary to explain the geographical division. The country is divided into 32 states and 2458 municipalities and within each municipality there are geographical subdivisions called polygons, defined for the execution of the program. These areas are collections of blocks called in Mexico as “manzanas”. The boundaries of these polygons are defined by geographic and eligibility criteria. Thus, we have four geographic levels: states, municipalities, polygons and manzanas. The experiment was carried out at the level of municipalities and polygons and will be described in the following paragraphs.

For the polygon to be eligible, it is necessary to satisfy the following criteria (McIntosh, Alegria, Ordonez & Zenteno 2018, p.266): i) to be located within a state and municipality where governments are willing to cooperate with the cost sharing rules (local governments (50%); municipalities (40%); states (8%) and the beneficiaries (2%)); ii) concentrations of asset poverty greater than 50 percent; iii) to be located in cities of 15,000 inhabitants or more; iv) deficit of infrastructure and urban services; v) with at least 80 percent of the lots having no active conflict over property rights; vi) municipalities that had only one polygon were excluded. The polygon will be defined according to criteria ii to v, that is, all regions that satisfy these conditions will be spatially delimited and will participate in the random experiment. In addition to these criteria, the policy maker required that saturation values be greater than zero and less than one and, therefore, there is no pure

control group.

The Random Saturation Design (RSD)¹ was conducted in 2008: “first the saturation of treatment was randomly assigned at the municipality level between 0.1 and 0.9, and then treatment was randomly assigned at the polygon level to match the municipality saturation as closely as possible” (McIntosh, Alegria, Ordonez & Zenteno 2018, p.267). In the experiment 65 municipalities participated of the randomization (there are 62 in the sample) totaling 370 polygons (5.7 polygons per municipality) of which 176 are treated and 194 control.

When the program starts operating in a polygon, a program agent (program officials) makes recommendations about the infrastructure deficit. Subsequently a second agent, called project executant, submits a project conditional to the operating rules. The program’s activities are divided into two parts: the first consists of urban development, which are works to improve infrastructure such as water sanitation, side-walks, curbs, wheelchair ramps, construction or improvement of roads, community sports fields, etc.; the second axis encompasses social and community development by carrying out actions related to job training, workshop on equality and domestic violence, the development and updating of community development plans (community centers, parks, sports and cultural facilities), social service provision for students, etc. In summary, we can say that the program’s actions are divided into two major groups, infrastructure and social development, and to measure the impact we will need different indicators that represent these two axes.

The individuals were interviewed at two points in time: the baseline questionnaire was conducted between March to July 2009 and the follow-up between January to March 2012. McIntosh et al. (2018) estimated the average effects of treatment on several outcome variables. Infrastructure indicators show positive average impacts in different segments, such as piped water, sidewalks, paved roads, concrete floor, flush toilet, monthly rent, etc. Due to the improvement of infrastructure, the price of square meter of land has increased. In the social area, a general social capital index was built and the program had no significant impact on this indicator. In view of this result, the authors examined the index in a disaggregated way, dividing it into indexes of security, youth behavior, social participation and trust. Positive effects were found for youth behavior and security indexes and not significant for the others.

There are several indicators that could be used in this empirical application, however, not all meet the criterion of continuity of the outcome variable. Some variables are binary and others, despite being continuous, when conditioned by saturation group the number of observations drastically reduces. In view of the aforementioned, only two outcome

¹The authors do not detail how the match described below was conducted. They state that an RSD was carried out, but it is not explained how the problem of treating, for example, 7.84% polygons was solved. Note that polygons are well-defined geographic divisions and therefore a rule is needed to select only part of the regions. Another type of solution would be to round off the saturation values that were drawn.

variables could be used: participation and trust indexes. The indexes were built using principal-component method: “The participation index is comprised of underlying questions on participation in nine different types of community organizations and questions on the existence of community groups, knowledge of community-level problems, and the degree of information exchange between neighbors. The trust index is built from two questions about the level of confidence in neighbors and local institutions and two questions about the degree of trust within the household and trust of neighbors” (McIntosh et al. 2018, p.280). The questionnaires applied to build these indexes are in Appendix B

As the authors found no significant effect for these indices, it may be the case that the behavior, measured through participation and trust in the community, is different throughout the distribution. The welfare of the community may be different throughout the distribution and the right tail may behave differently from the left and center tail. More than that, increasing the proportion of treated individuals can increase or decrease confidence in institutions and neighbors and also participation in community events. Using the theory developed in the previous chapters some questions can be raised: Is there a heterogeneous spillover effect for participation and trust indexes? Does increasing saturation change social participation and trust throughout distribution? Is there a heterogeneous spillover effect on the marginal quantile (well-being)?

The next sections will cover descriptive statistics and results using the database of this experiment.

5.2 Descriptive statistics

The database has information for the baseline (2009) and follow-up (2012) and all the results presented below use the information from 2012. The sample has 6214 households, 3233 of which are treated and 2981 control. There are 62 saturation groups, the average saturation is 46.25 with a standard deviation of 24.77. The municipality with the lowest saturation has $S_{min} = 7.84\%$ treated subjects whereas the one with the highest saturation has $S_{max} = 93.47$. Figure A1 shows the cumulative distribution for the saturation vector. Each point in this graph represents a municipality² that has a treatment and control group. For the municipality with minimum saturation we have 8 treated households and 94 in the control group, for the group with maximum saturation there are 229 treated and 16 in the control group.

In principle, the proposed objects could be estimated for the 62 groups, however, some parts of the saturation distribution have a reduced sample which could generate imprecise results. At first moment, only 3 groups, percentiles 25%, 50% and 75%, of the

²Strictly it would be a municipality, with some polygons that have treatment and control groups. The following applications should be carried out at the polygon level, however, it will not be possible to expand the level of analysis due to the reduced sample size of polygons.

saturation vector will be used to make the empirical application. It should be noted that the groups between the percentiles are not used, but only the three points mentioned and the respective treatment and control groups within each point. Table 1 in Appendix B shows the sample size and the proportion of treated by group. Note that greater saturation does not imply a larger sample size, the group with the lowest number of households is the one located in the third quartile.

Graphs A2 and A3 show, for the complete sample, the cumulative distributions for the participation and trust indexes. The distribution for the participation index has right skew with its values concentrated around -1.11 (mode), with median of -0.44, mean of -0.16, and standard deviation of 1.21. The distribution for the trust index is more symmetrical with the median (-0.28) close to the mean (-0.26) and standard deviation of 1.04. In both cases the distributions show, due to the negative values, that individuals have low participation and confidence in the regions where they live.

The general characteristics of the distributions, right skew for participation and symmetry for trust, are similar when dividing the sample by saturation groups. Graphs A4-A7 show the cumulative distributions for the participation index for the treatment and control groups for the sample as a whole (A4) and for the saturation groups defined previously (A5-A7). Graph A4 shows that the distributions of the treatment and control groups are similar and, in this case, we can say that for each percentile the horizontal difference would be close to zero. When we analyze by group we can see that there is no clear pattern regarding saturation. In the group with the smallest and largest proportion of treated subjects, there are large horizontal distances, which does not occur for the group with median saturation. Note that the tails of graphs A5 and A7 have a closer approximation to the treatment group, which may raise evidence about the heterogeneity of spillover effects. For $\tau < 0.4$, for all groups, we have differences close to zero indicating null direct effects.

Graphs A8-A11 show the cumulative distributions for the trust index for the treatment and control groups for the sample as a whole (A8) and for the saturation groups defined previously (A9-A11). These graphs also reveal distinct patterns regarding the distances between the two distributions by saturation group. Graphs A9 and A11 show that, from $\tau > 0.6$, the curves of the treatment and control group are similar showing distances close to zero. What does not happen for the group with median saturation where there are positive distances from the percentile $\tau > 0.6$. The graphs A9-A10 suggest that treatment may reduce trust in the community for the left tail.

Figures A12-A19 adds the unconditional (marginal) distribution to the previous graphs. The greatest vertical distance, visually analyzing it, is represented in graph A13 (participation index in the first saturation quartile) showing dominance in the distribution of the treatment group when compared to the control group. Graphs A17 and A18 indicate negative distances over the marginal quantile (from $\tau < 0.5$). For the other graphs, it is

not possible to say whether the distances are negative or positive.

Looking at the cumulative distributions it does not seem that saturating more or less results in a clear pattern with respect to the direct effects of the treatment (without doing any tests using the standard deviation). For the sample as a whole, the distances between the treatment and control groups, along the distribution, are close to zero. As we can recover the average integrating the quantiles in the interval $(0, 1)$ we have that the average impact would be null which is in agreement with the work of McIntosh, Alegria, Ordonez & Zenteno (2018), because the authors did not find statistically significant impacts for both indexes

The following sections will address the results of the empirical application using the estimators proposed in the previous chapters.

5.3 Results

This section is divided into two subsections: the first shows the results of the direct effects for the groups previously presented. Due to the database restrictions, absence of the pure control group, it will not be possible to estimate the indirect and total effects. The second subsection shows two exercises: a first aggregating the sample from the saturation groups and a second exercise, using this aggregation, and a proxy for the pure control group generating the indirect and total effects.

5.3.1 Direct effects (horizontal and vertical) and derivatives

In addition to correctly identifying distances, standard deviations were estimated using 1500 bootstrap repetitions³. All tests presented are performed at a significance level of 5%.

Graphs A20-A22 show the results for the participation index. There are direct positive and statistically significant effects for the right tail and center of distribution at first quartile saturation group. The effects are greater for the right tail showing impacts above a standard deviation. The left tail, 10th and 15th percentiles, at third quartile saturation group also had statistically significantly and positive impact.

Graphs A23-A25 show the results for the trust index. There are direct effect, negative and statistically significant, for the left tail in the first quartile of the saturation group. The difference between the distributions of the treatment and control group, for the left tail, is around -0.5 standard deviation. The 15th percentile in the second saturation group ($S = 39.81$) also had a negative impact. We can say that trust in institutions, for the left tail, reduces when a smaller proportion of individuals is treated. For the participation

³Davidson & MacKinnon (2000) show that there is a loss of power due to finite repetitions when using bootstrap. The authors recommend at least 399 repetitions to do a hypothesis test with a significance level of 0.05 and at least 1499 repetitions for a significance level of 0.01.

index the opposite occurs, participation in the community increases, for the right tail, in the group with less saturation.

In general, graphs A20-A25 show that, for a given percentile, there are no statistically significant direct effects of treatment when increasing saturation, that is, a higher proportion of treated subjects does not imply a greater or lesser impact⁴ for this sample. However, statistically significant impacts were identified on the tails and center of distribution in groups with lower saturation $S = 29.46$ suggesting that a lower proportion of treaties could result in changes in the tails of both indexes.

Graphs A26-A28a show the direct marginal effects for the participation index. Similar to the results of horizontal distances we can say that there is a direct effect, positive and statistically significant, for the right tail of the group with less saturation ($S = 29.46$) and for the left tail for the group of the third quartile saturation ($S = 69.72$). Despite not showing statistically significant results for some percentiles, we noticed that the direct marginal effect is increasing throughout the distribution for the group in the first saturation quartile. The size of the coefficient, when compared to previous results, is the main difference between QDE and MQDE. For the first quartile group the greatest impact on the horizontal distance is below 2 standard deviations and for the marginal distance it reaches 3 standard deviations. We can make similar reasoning for the statistically significant value of the third saturation quartile.

The quantile marginal effects for the trust index are shown in graphs A29-A31. The results for the first quartile saturation group are similar to those presented for horizontal distances (negative effects for the left tail around -0.5 standard deviation). For groups located in the second ($S = 39.81$) and third quartiles ($S = 69.72$) there are positive effects for the right tail, the largest of which is close to two standard deviations ($S = 69.72, \tau = 0.94$). The 30th and 35th percentiles, in the ($S = 69.72$) group, also show an increase in the confidence index when compared to the control group, however, the impact value is lower (around 1 standard deviation).

The results for the participation index do not suggest a positive relationship between saturation and the marginal effect, that is, increasing the proportion of treated subject would not imply greater effects. For the confidence index it seems that there are larger effects for the right tail when increasing the saturation. Furthermore, in both cases for the trust index (horizontal and vertical distances), a change in pattern is noticed when increasing the saturation (positive values in the graphs with greater saturation), however, most of the results are not significant.

Graphs A32-A37 show the exercise of decomposition of the derivative of the marginal quantile with respect to saturation. The derivative is divided into two parts (MQDE and MQSE) and the direct effect has been previously presented. To estimate the spillover effect

⁴It is worth mentioning that we are not testing the difference in quantiles between the saturation groups, but only measuring the impact within each group separately.

using the discrete derivative approximation, it is necessary to use a reference group. The graphs below show the derivative of the marginal quantile with respect to the saturation level ($S = 39.81$) (second quartile), with the previous quartile being used as a reference group ($S = 29.82$). Therefore, the estimates related to spillover are compared with the first quartile saturation group showing the marginal effect of increasing approximately 10% of treated individuals.

The spillover effect and the derivative for the participation index are shown in graphs A33-A34 in appendix A. The MQSE is decreasing, statistically significantly and negative from the median. To understand this result it is necessary to see that the discrete approximation of the derivative compares the number of individuals of the marginal quantile between two groups, in this specific case, the groups with saturation equal to ($S = 39.81$) and ($S = 29.46$). Thus, the final result of this comparison are the effects represented in graph A33 showing that by increasing saturation, individuals in the second group ($S = 39.81$) are relatively worse than those located in the first group ($S = 29.46$). As the MQSE values are high this will affect the derivative since this is the sum of MQDE and MQSE. Thus, the derivative presents negative values from the median showing that increasing saturation for the participation index will reduce the marginal quantile of the middle of the distribution onwards.

The reasoning in the previous paragraph (comparison between two groups) is obviously valid for the trust index (graphs A36-A37). Note that the spillover effects on the A29 and A30 graphs are almost a mirror of these results. Thus, the MQSE is positive for the left tail when comparing the groups ($S = 39.81$) e ($S = 29.46$). Again, as the MQSE values are high and the MQDE values are low, the sum of both will show that the derivative will have a positive effect on the left tail following the spillover effects.

These results make it possible to question the relationship between saturation and the derivative: what is the relationship between the derivative, and its components, with the saturation level? For a given percentile, for example the median, how do these components behave in the 62 saturation groups? For a given percentile, is the MQDE higher for any group? Does the data show any pattern when increasing or decreasing saturation? Graphs A38-A55 show the results by fixing three percentiles (5, 50 and 95) and estimating the derivative components for all 62 saturation groups. For example, charts A38-A40 show, respectively, the MQDE for the participation index for the 5th, 50th and 95th percentiles for the 62 saturation groups (the horizontal axis are the groups in ascending order of saturation, that is, the first group has the minimum saturation and the last the maximum saturation). The other graphs are similar showing the MQSE and the derivative for the same percentiles and for the two outcome variables. No standard deviation was estimated for these estimations. The results show that there is no defined standard with respect to saturation and the effects MQDE, MQSE and the derivative. The graphs present a point cloud where these components can assume different values as we change the saturation.

Thus, it is not possible to state, for these percentiles and this database, that greater proportions of treated individuals will monotonically increase (or reduce) the effects on marginal quantiles.

The results show that the actions carried out by the program, improvements in infrastructure and community development, result in positive impacts on the participation index and negative impacts on the trust index. Improvements in the community make the right tail of distribution of the participation index better for the treatment group than for the control group, however, there is no gain in trust within those regions.

There are a lot of direct and non-significant effects on the results presented. It may be necessary to analyze the indirect effects, as the control group may be contaminated. Furthermore, the sample is reduced and we know that the results will be more accurate in larger samples. Thus, the next subsection presents two exercises that will address the increase in sample size and the choice of a proxy for the pure control group.

5.3.2 Exercises 1 (estimation in the saturation interval) and 2 (choosing pure control group)

This section presents two exercises to test the robustness of the results. The first exercise consists of grouping the saturation groups and re-estimating the direct effects. When aggregating the groups, the sample size will be larger, leading to more precise results. The second exercise aims to choose a proxy for the pure control group, among the 62 saturation groups, and to estimate the indirect and total effects using the aggregate base. As there are a large number of groups, we can use several strategies for aggregating individuals. The following exercise aggregates individuals according to the proximity of the saturation groups considering that nearby groups are more homogeneous. The database was divided into four parts according to the saturation quartiles previously presented and each part became unique. Thus, previous estimates (direct effects) can be performed again within each new group. Figure A56 in Appendix A illustrates the aggregation strategy, we will have 4 intervals with sample sizes as shown in Table 2 of Appendix B. Notice that the sample size increases and we have 4 new saturation categories $(S_1; S_2; S_3; S_4) = (18.72; 35.57; 51.14; 82.88)$.

Graphs A57-A64 show the cumulative distribution for the participation and trust index in the new groups. The participation index shows a pattern, for the right tail, in which the treatment group curve is located to the right of the control group curve at the lowest saturation levels ($S_1 = 18.72$) and ($S_2 = 35.57$). The opposite is true for higher levels of saturation. For the trust index, it is not possible to verify visual differences in distributions up to the level ($S_3 = 51.14$). For the highest level of saturation, the control group dominates the treatment distribution for the percentiles between 0.2 and 0.8. Figures A65-A72 add the marginal distribution curve to the previous graphs making

it clear that there is no pattern, in addition, the distance between the curves is reduced so that the marginal distribution is confused with the distributions of the treatment and control group (except for the participation variable for the first two groups).

The direct effects for exercise 1 are shown in graphs A73-A80. The participation index has a statistically significant impact for the right tail in all saturation groups. There is a negative impact for the left tail in groups with higher saturation. The confidence index had negative impacts on the left tail and center of distribution for the groups S_3 and S_4 . QDE values are lower when compared to previous estimates.

Graphs A81-A88 show MQDE results for the same aggregate saturation intervals. For the participation variable, the effects are increasing for the first two groups of saturation (significant for the right tail) and decreasing for the last two groups (significant for the left tail). For the trust index, only the last group had a negative and statistically significant impact at the center of the distribution.

Finally, exercise 2 proposes the choice of a proxy for the pure control group among the 62 database groups. The ideal pure control group would be one that was not contaminated. A possible choice would be to use the control group with the lowest saturation level ($S = 7.84$) assuming that smaller proportions of treated people lead to less contamination. Figures A89-A92 and A97-A100 add the accumulated distribution of this group to the previous graphs. The distribution of the proxy pure control group was expected to be to the left of the other distributions, however, it crosses or even lies to the right of the graphs. A second option is to use the control group for the next saturation level ($S = 9.36$). Graphs A93-A96 and A101-A104 show this new attempt to find the proxy for the pure control group. In all cases, this group is to the left of the other distributions, looking like a reasonable group to estimate the indirect and total effects. Finally, it should be noted that this effort to use a group as a proxy is only to show the application of indirect and total effects and not really propose a method for choosing such groups since the choice was arbitrary.

The results of this exercise are shown in figures A105-A136. As expected, due to the choice of a proxy group to the left of the other distributions, all results are positive and most are statistically significant. In general, the greater impact is right tail. For a given percentile there seems to be no relationship between saturation and indirect and total effects.

Chapter 6

Conclusion

The aim of this work was to identify several distributive objects of the treatment effect on the outcome variable in the presence of spillover. The literature has developed several estimators to obtain the effect of treatment with and without the presence of spillover for the mean, but there are still no work for quantile and treatment saturation.

As a continuation of this work, we can develop the asymptotic theory regarding the limit distribution of the estimators as well as carry out Monte Carlo studies among the different saturation groups. In addition, one can also remove the assumptions related to randomization and rewrite the measures using some kind of weighting to correct the treatment selection or include the network information in the context of the treatment effect throughout the distribution.

Finally, the estimators presented in this work become inconsistent if there is selection in observable or unobservable variables, if the program induces network formation, if not all treated unit are compliers and if the subjects systematically drop out of the study after randomization (attrition).

Bibliography

- Baird, S., Bohren, A., McIntosh, C. & Ozler, B. (2015), Designing experiments to measure spillover effects, Technical Report 15-021, Penn Institute for Economic Research.
- Basse, G. & Feller, A. (2017), ‘Analyzing two-stage experiments in the presence of interference’, *Journal of the American Statistical Association* **0**, 1–34.
- Cox, D. R. (1958), *Planning of Experiments*, John Wiley, New York.
- Crépon, B., Duflo, E., Gurgand, M., Rathelot, R. & Zamora, P. (2013), ‘Do labor market policies have displacement effects? evidence from a clustered randomized experiment’, *The Quarterly Journal of Economics* **128**(2), 531–580.
- Davidson, R. & MacKinnon, J. G. (2000), ‘Bootstrap tests: how many bootstraps?’, *Econometric Reviews* **19**(1), 55–68.
- Doksum, K. (1974), ‘Empirical probability plots and statistical inference for nonlinear models in the two-sample case’, *The Annals of Statistics* **2**(2), 267–277.
- Donald, S. G. & Hsu, Y.-C. (2014), ‘Estimation and inference for distribution functions and quantile functions in treatment effect model’, *Journal of Econometrics* **178**(3), 383–397.
- Fan, Y. & Liu, R. (2016), ‘A direct approach to inference in nonparametric and semiparametric quantile models’, *Journal of Econometrics* **191**(1), 196–216.
- Firpo, S. (2007), ‘Efficiente semiparemetric estimation of quantile treatment effects’, *Econometrica* **75**(1), 259–276.
- Firpo, S., Fortin, N. M. & Lemieux, T. (2009), ‘Unconditional quantile regressions’, *Econometrica* **77**(3), 953–973.
- Firpo, S. & Pinto, C. (2015), ‘Identification and estimation of distributional impacts of interventions using changes in inequality measures’, *Journal of Applied Econometrics* **31**(3), 457–486.

- Graham, B., Imbens, G. W. & Ridder, G. (2010), Measuring the effects of segregation in the presence of social spillovers: a nonparametric approach, Technical Report 16499, National Bureau of Economic Research.
- Halloran, M. E. & Struchiner (1991), ‘Study designs for dependent happenings’, *Epidemiology Resources* **2**(5), 331–338.
- Hudgens, M. G. & Halloran, M. E. (2008), ‘Toward causal inference with interference’, *Journal of the American Statistical Association* **103**(482), 832–842.
- Imbens, G. W. & Wooldridge, J. M. (2009), ‘Recent developments in the econometrics of program evaluation’, *Journal of Economic Literature* **47**(1), 5–86.
- Lehmann, E. L. (1975), *Nonparametrics: Statistical Methods Based on Ranks*, Springer, New York.
- Leung, M. P. (2017), Treatment and spillover effects under network interference. Working paper.
- Liu, L. & Hudgens, M. G. (2014), ‘Large sample randomization inference of causal effects in the presence of interference’, *Journal of the American Statistical Association* **109**(505), 288–301.
- Manski, C. F. (2013), ‘Identification of treatment response with social interactions’, *Econometrics Journal* **16**(1), 1–23.
- McIntosh, C., Alegria, T., Ordonez, G. & Zenteno, R. (2018), ‘The neighborhood impacts of local infrastructure investment: Evidence from urban mexico’, *American Economic Journal: applied economics* **10**(3), 263–286.
- Pagan, A. & Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge University Press, New York.
- Philipson, T. J. (2000), External treatment effects and program implementation bias, Technical Report 250, National Bureau of Economic Research.
- Rubin, D. B. (1974), ‘Estimating causal effects of treatments in randomized and non randomized studies’, *Journal of Educational Psychology* **66**(5), 688–701.
- Rubin, D. B. (1980), ‘Discussion of randomization analysis of experimental data in the fisher randomization test’, *Journal of the American Statistical Association* **75**(371), 591–593.
- Shiryaev, A. N. (1995), *Probability*, Springer, New York.

- Sobel, M. E. (2006), ‘what do randomized studies of housing mobility demonstrate?: causal inference in the face of interference’, *Journal of the American Statistical Association* **101**(476), 1398–1407.
- Tchetgen Tchetgen, E. J. & VanderWeele, T. J. (2010), ‘On causal inference in the presence of interference’, *Statistical Methods in Medical Research* **21**(1), 55–75.
- Vazquez-Bare, G. (2017), Identification and estimation of spillover effects in randomized experiments. Working paper.
- Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, The Mit Press, Cambridge, Massachusetts.

Appendix A

Figures

Descriptive Statistics

Figure A1: Cumulative distribution for saturation vector

Source: elaborated by the author

All the results presented in this work use the participation and trust indexes. The indexes were built using principal-component method (McIntosh et al. 2018, p.269). “The participation index is comprised of underlying questions on participation in nine different types of community organizations and questions on the existence of community groups, knowledge of community-level problems, and the degree of information exchange between neighbors. The trust index is built from two questions about the level of confidence in neighbors and local institutions and two questions about the degree of trust within the household and trust of neighbors” (McIntosh et al. 2018, p.280). The indexes were standardized. No control variables were included.

Figure A2: Cumulative distribution for participation index (complete sample) Figure A3: Cumulative distribution for trust index (complete sample)

Source: elaborated by the author

Source: elaborated by the author

Figure A4: Cumulative distribution for participation index (complete sample)

Source: elaborated by the author

Figure A5: Cumulative distribution for participation index at first quartile saturation group ($S = 29.46$)

Source: elaborated by the author

Figure A6: Cumulative distribution for participation index at second quartile saturation group ($S = 39.81$)

Source: elaborated by the author

Figure A7: Cumulative distribution for participation index at third quartile saturation group ($S = 69.72$)

Source: elaborated by the author

Figure A8: Cumulative distribution for trust index (complete Figure A9: Cumulative distribution for trust index at first quartile saturation group ($S = 29.46$) sample)

Source: elaborated by the author

Source: elaborated by the author

Figure A10: Cumulative distribution for trust index at second quartile saturation group ($S = 39.81$) Figure A11: Cumulative distribution for trust index at third quartile saturation group ($S = 69.72$)

Source: elaborated by the author

Source: elaborated by the author

Figure A12: Cumulative distribution for participation index (complete sample with unconditional distribution)

Source: elaborated by the author

Figure A13: Cumulative distribution for participation index at first quartile saturation group ($S = 29.46$) (with unconditional distribution)

Source: elaborated by the author

Figure A14: Cumulative distribution for participation index at second quartile saturation group ($S = 39.81$) (with unconditional distribution)

Source: elaborated by the author

Figure A15: Cumulative distribution for participation index at third quartile saturation group ($S = 69.72$) (with unconditional distribution)

Source: elaborated by the author

Figure A16: Cumulative distribution of trust index (complete sample with unconditional distribution)

Source: elaborated by the author

Figure A17: Cumulative distribution of trust index at first quartile saturation group ($S = 29.46$) (with unconditional distribution)

Source: elaborated by the author

Figure A18: Cumulative distribution of trust index at second quartile saturation group ($S = 39.81$) (with unconditional distribution)

Source: elaborated by the author

Figure A19: Cumulative distribution of trust index at third quartile saturation group ($S = 69.72$) (with unconditional distribution)

Source: elaborated by the author

Quantile Direct Effects (horizontal distances)

Figure A20: QDE for participation index at first quartile saturation group ($S = 29.46$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A21: QDE for participation index at second quartile saturation group ($S = 39.81$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A22: QDE for participation index at third quartile saturation group ($S = 69.72$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A23: QDE for trust index at first quartile saturation group ($S = 29.46$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A24: QDE for trust index at second quartile saturation group ($S = 39.81$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A25: QDE for trust index at third quartile saturation group ($S = 69.72$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Marginal Quantile Direct Effects (vertical distances)

Figure A26: MQDE for participation at first quartile saturation group ($S = 29.46$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A27: MQDE for participation at second quartile saturation group ($S = 39.81$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A28: MQDE for participation at third quartile saturation group ($S = 69.72$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A28a: MQDE for participation at third quartile saturation group ($S = 69.72$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included. This figure replicates the chart A28. The estimate for the 95th percentile was removed, as the standard deviation was large and distorted the presentation of the results (see the graph 28). For this percentile, the MQDE value is zero and not statistically significant at 5 %.

Figure A29: MQDE for trust index at first quartile saturation group ($S = 29.46$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A30: MQDE for trust index at second quartile saturation group ($S = 39.81$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A31: MQDE for trust index at third quartile saturation group ($S = 69.72$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Marginal Quantile Effects (derivative)

Figure A32: MQDE for participation index at second quartile saturation group ($S = 39.81$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included. The group with saturation at the level $S = 29.82$ (first quartile) was used as a reference to estimate the derivative in a discrete way.

Figure A33: MQSE for participation index at second quartile saturation group ($S = 39.81$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included. The group with saturation at the level $S = 29.82$ (first quartile) was used as a reference to estimate the derivative in a discrete way.

Figure A34: MQE for participation index at second quartile saturation group ($S = 39.81$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included. The group with saturation at the level $S = 29.82$ (first quartile) was used as a reference to estimate the derivative in a discrete way.

Figure A35: MQDE for trust index at second quartile saturation group ($S = 39.81$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included. The group with saturation at the level $S = 29.82$ (first quartile) was used as a reference to estimate the derivative in a discrete way.

Figure A36: MQSE for trust index at second quartile saturation group ($S = 39.81$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included. The group with saturation at the level $S = 29.82$ (first quartile) was used as a reference to estimate the derivative in a discrete way.

Figure A37: MQE for trust index at second quartile saturation group ($S = 39.81$)

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included. The group with saturation at the level $S = 29.82$ (first quartile) was used as a reference to estimate the derivative in a discrete way.

What is the relationship between the derivative and its components with the saturation level?

Figure A38: MQDE for participation index at percentile 5 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A39: MQDE for participation index at percentile 50 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A40: MQDE for participation index at percentile 95 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A41: MQSE for participation index at percentile 5 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A42: MQSE for participation index at percentile 50 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A43: MQSE for participation index at percentile 95 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A44: Derivative for participation index at percentile 5 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A45: Derivative for participation index at percentile 50 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A46: Derivative for participation index at percentile 95 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A47: MQDE for trust index at percentile 5 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A48: MQDE for trust index at percentile 50 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A49: MQDE for trust index at percentile 95 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A50: MQSE for trust index at percentile 5 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A51: MQSE for trust index at percentile 50 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A52: MQSE for trust index at percentile 95 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A53: Derivative for trust index at percentile 5 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A54: Derivative for trust index at percentile 50 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Figure A55: Derivative for trust index at percentile 95 (all groups)

Source: elaborated by the author

Note: The horizontal axis represents the 61 saturation groups. The vertical axis shows the effect of the treatment within the respective group. No control variables were included. The group used as a reference to estimate the derivative in a discrete way was the one immediately before.

Exercise 1: saturation interval

Figure A56: Saturation intervals

Figure A57: Cumulative distribution for participation index at first interval saturation group $S_1 = 18.72$

Source: elaborated by the author

Figure A58: Cumulative distribution for participation index at second interval saturation group $S_2 = 35.57$

Source: elaborated by the author

Figure A59: Cumulative distribution for participation index at third interval saturation group $S_3 = 51.14$

Source: elaborated by the author

Figure A60: Cumulative distribution for participation index at fourth interval saturation group $S_4 = 82.88$

Source: elaborated by the author

Figure A61: Cumulative distribution for trust index at first interval saturation group $S_1 = 18.72$

Source: elaborated by the author

Figure A62: Cumulative distribution for trust index at second interval saturation group $S_2 = 35.57$

Source: elaborated by the author

Figure A63: Cumulative distribution for trust index at third interval saturation group $S_3 = 51.14$

Source: elaborated by the author

Figure A64: Cumulative distribution for trust index at fourth interval saturation group $S_4 = 82.88$

Source: elaborated by the author

Figure A65: Cumulative distribution for participation index at first interval saturation group $S_1 = 18.72$ with marginal distribution

Source: elaborated by the author

Figure A66: Cumulative distribution for participation index at second interval saturation group $S_2 = 35.57$ with marginal distribution

Source: elaborated by the author

Figure A67: Cumulative distribution for participation index at third interval saturation group $S_3 = 51.14$ with marginal distribution

Source: elaborated by the author

Figure A68: Cumulative distribution for participation index at fourth interval saturation group $S_4 = 82.88$ with marginal distribution

Source: elaborated by the author

Figure A69: Cumulative distribution for trust index at first interval saturation group $S_1 = 18.72$ with marginal distribution

Source: elaborated by the author

Figure A70: Cumulative distribution for trust index at second interval saturation group $S_2 = 35.57$ with marginal distribution

Source: elaborated by the author

Figure A71: Cumulative distribution for trust index at third interval saturation group $S_3 = 51.14$ with marginal distribution

Source: elaborated by the author

Figure A72: Cumulative distribution for trust index at fourth interval saturation group $S_4 = 82.88$ with marginal distribution

Source: elaborated by the author

Quantile Direct Effects (horizontal distances) for estimates within the saturation interval (exercise 1)

Figure A73: QDE for participation index at first interval saturation group $S_1 = 18.72$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A74: QDE for participation index at second interval saturation group $S_2 = 35.57$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A75: QDE for participation index at third interval saturation group $S_3 = 51.14$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A76: QDE for participation index at fourth interval saturation group $S_4 = 82.88$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A77: QDE for trust index at first interval saturation group $S_1 = 18.72$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A78: QDE for trust index at second interval saturation group $S_2 = 35.57$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A79: QDE for trust index at third interval saturation group $S_3 = 51.14$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A80: QDE for trust index at fourth interval saturation group $S_4 = 82.88$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Marginal Quantile Direct Effects (vertical distances) for estimates within the saturation interval (exercise 1)

Figure A81: MQDE for participation index at first interval saturation group $S_1 = 18.72$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A82: MQDE for participation index at second interval saturation group $S_2 = 35.57$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A83: MQDE for participation index at third interval saturation group $S_3 = 51.14$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A84: MQDE for participation index at fourth interval saturation group $S_4 = 82.88$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A85: MQDE for trust index at first interval saturation group $S_1 = 18.72$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A86: MQDE for trust index at second interval saturation group $S_2 = 35.57$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A87: MQDE for trust index at third interval saturation group $S_3 = 51.14$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A88: MQDE for trust index at fourth interval saturation group $S_4 = 82.88$

Source: elaborated by the author

Note: Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Exercise 2: choosing within control group as pure control

Figure A89: Cumulative distribution for participation index at first interval saturation group $S_1 = 18.72$ with “pure control” (group 1; $S = 7.84$)

Source: elaborated by the author

Figure A90: Cumulative distribution for participation index at second interval saturation group $S_2 = 35.57$ with “pure control” (group 1; $S = 7.84$)

Source: elaborated by the author

Figure A91: Cumulative distribution for participation index at third interval saturation group $S_3 = 51.14$ with “pure control” (group 1; $S = 7.84$)

Source: elaborated by the author

Figure A92: Cumulative distribution for participation index at fourth interval saturation group $S_4 = 82.88$ with “pure control” (group 1; $S = 7.84$)

Source: elaborated by the author

Figure A93: Cumulative distribution for participation index at first interval saturation group $S_1 = 18.72$ with “pure control” (group 2; $S = 9.36$)

Source: elaborated by the author

Figure A94: Cumulative distribution for participation index at second interval saturation group $S_2 = 35.57$ with “pure control” (group 2; $S = 9.36$)

Source: elaborated by the author

Figure A95: Cumulative distribution for participation index at third interval saturation group $S_3 = 51.14$ with “pure control” (group 2; $S = 9.36$)

Source: elaborated by the author

Figure A96: Cumulative distribution for participation index at fourth interval saturation group $S_4 = 82.88$ with “pure control” (group 2; $S = 9.36$)

Source: elaborated by the author

Figure A97: Cumulative distribution for trust index at first interval saturation group $S_1 = 18.72$ with “pure control” (group 1; $S = 7.84$)

Source: elaborated by the author

Figure A98: Cumulative distribution for trust index at second interval saturation group $S_2 = 35.57$ with “pure control” (group 1; $S = 7.84$)

Source: elaborated by the author

Figure A99: Cumulative distribution for trust index at third interval saturation group $S_3 = 51.14$ with “pure control” (group 1; $S = 7.84$)

Source: elaborated by the author

Figure A100: Cumulative distribution for trust index at fourth interval saturation group $S_4 = 82.88$ with “pure control” (group 1; $S = 7.84$)

Source: elaborated by the author

Figure A101: Cumulative distribution for trust index at first interval saturation group $S_1 = 18.72$ with “pure control” (group 2; $S = 9.36$)

Source: elaborated by the author

Figure A102: Cumulative distribution for trust index at second interval saturation group $S_2 = 35.57$ with “pure control” (group 2; $S = 9.36$)

Source: elaborated by the author

Figure A103: Cumulative distribution for trust index at third interval saturation group $S_3 = 51.14$ with “pure control” (group 2; $S = 9.36$)

Source: elaborated by the author

Figure A104: Cumulative distribution for trust index at fourth interval saturation group $S_4 = 82.88$ with “pure control” (group 2; $S = 9.36$)

Source: elaborated by the author

Quantile Indirect Effects (horizontal distances) for estimates within the saturation interval with “pure control” (group 2; $S = 9.36$) (exercise 2)

Figure A105: QIE for participation index at first interval saturation group $S_1 = 18.72$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A106: QIE for participation index at second interval saturation group $S_2 = 35.57$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A107: QIE for participation index at third interval saturation group $S_3 = 51.14$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A108: QIE for participation index at fourth interval saturation group $S_4 = 82.88$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A109: QIE for trust index at first interval saturation group $S_1 = 18.72$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A110: QIE for trust index at second interval saturation group $S_2 = 35.57$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A111: QIE for trust index at third interval saturation group $S_3 = 51.14$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A112: QIE for trust index at fourth interval saturation group $S_4 = 82.88$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Quantile Total Effects (horizontal distances) for estimates within the saturation interval with “pure control” (group 2; $S = 9.36$) (exercise 2)

Figure A113: QTE for participation index at first interval saturation group $S_1 = 18.72$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A114: QTE for participation index at second interval saturation group $S_2 = 35.57$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A115: QTE for participation index at third interval saturation group $S_3 = 51.14$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A116: QTE for participation index at fourth interval saturation group $S_4 = 82.88$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A117: QTE for trust index at first interval saturation group $S_1 = 18.72$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A118: QTE for trust index at second interval saturation group $S_2 = 35.57$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A119: QTE for trust index at third interval saturation group $S_3 = 51.14$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A120: QTE for trust index at fourth interval saturation group $S_4 = 82.88$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Marginal Quantile Indirect Effects (vertical distances) for estimates within the saturation interval with “pure control” (group 2; $S = 9.36$) (exercise 2)

Figure A121: MQIE for participation index at first interval saturation group $S_1 = 18.72$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A122: MQIE for participation index at second interval saturation group $S_2 = 35.57$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A123: MQIE for participation index at third interval saturation group $S_3 = 51.14$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A124: MQIE for participation index at fourth interval saturation group $S_4 = 82.88$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A125: MQIE for trust index at first interval saturation group $S_1 = 18.72$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A126: MQIE for trust index at second interval saturation group $S_2 = 35.57$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A127: MQIE for trust index at third interval saturation group $S_3 = 51.14$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A128: MQIE for trust index at fourth interval saturation group $S_4 = 82.88$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Marginal Quantile Total Effects (vertical distances) for estimates within the saturation interval with “pure control” (group 2; $S = 9.36$) (exercise 2)

Figure A129: MQTE for participation index at first interval saturation group $S_1 = 18.72$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A130: MQTE for participation index at second interval saturation group $S_2 = 35.57$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A131: MQTE for participation index at third interval saturation group $S_3 = 51.14$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A132: MQTE for participation index at fourth interval saturation group $S_4 = 82.88$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A133: MQTE for trust index at first interval saturation group $S_1 = 18.72$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A134: MQTE for trust index at second interval saturation group $S_2 = 35.57$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A135: MQTE for trust index at third interval saturation group $S_3 = 51.14$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Figure A136: MQTE for trust index at fourth interval saturation group $S_4 = 82.88$

Source: elaborated by the author

Note: The control group used as a proxy for pure control was the one located in the second lowest saturation level (group 2; $S = 9.36$). Standard deviations were estimated using 1500 bootstrap replications. We report 95% confidence interval. No control variables were included.

Appendix B

Tables

Table 1: Sample size by saturation group

Saturation level	Treated	Control	Total	Proportion of treated
First quartile saturation group	236	565	801	29.46
Second quartile saturation group	164	248	412	39.81
Third quartile saturation group	76	33	109	69.72

Note: The cells in columns 2 and 3 show the number of individuals in the treatment and control groups according to saturation level. The column 4 is the sum of columns 2 and 3. The last column shows the proportion of treated individuals (saturation) for each group.

Table 2: Sample size by saturation interval group

Interval	Treated	Control	Total	Proportion of treated
First interval	605	2626	3231	18.72
Second interval	889	1610	2499	35.57
Third interval	1503	1436	2939	51.14
Fourth interval	2314	478	2792	82.88

Note: The cells in columns 2 and 3 show the number of individuals in the treatment and control groups according to saturation level for interval saturation. The column 4 is the sum of columns 2 and 3. The last column shows the proportion of treated individuals (saturation) for each group.

7. Capital social

Ahora le voy a hacer algunas preguntas relacionadas con la participación de usted o algún miembro de su hogar en organizaciones, proyectos o actividades que benefician a su familia y a sus vecinos, así como con las relaciones de convivencia que existen en su comunidad.

Grupos y redes

7.1 ¿Dentro de esta colonia existen grupos u organizaciones de vecinos que se reúnan para convivir o para realizar actividades en favor de la comunidad?

CIRCULE UN SOLO CÓDIGO

Si.....1
No.....2 PASE A 7.2

7.1.1 ¿Participa usted o algún miembro de su hogar en alguno o algunos de estos grupos, redes o asociaciones?

CIRCULE UN SOLO CÓDIGO

Si.....1
No.....2 PASE A 7.2

7.1.2 Dentro de su colonia, ¿en cuáles de los siguientes grupos u organizaciones participan?

LEA CADA UNA DE LAS OPCIONES Y CIRCULE LA RESPUESTA EN CADA CASO

	Sí	No
Una agrupación religiosa o iglesia	1	2
Una asociación de padres (madres) de familia	1	2
Una asociación o grupo de la tercera edad	1	2
Un sindicato	1	2
Una organización o grupo de vecinos(as)	1	2
Un grupo deportivo o recreativo	1	2
Una asociación de autoayuda (AA, Neuróticos Anónimos)	1	2
Una agrupación o partido político	1	2
Otra (especifique):	1	2

7.1.3 De todos los grupos, redes o asociaciones a los que pertenecen usted o algún miembro de su hogar dentro de su colonia, ¿cuál de ellos es el que considera más importante para su familia?

Nombre del grupo, asociación, red: _____

7.2 ¿Hay en su colonia algún representante de la colonia o jefe de manzana?

CIRCULE UN SOLO CÓDIGO

Si.....1
No.....2 PASE A 7.3
No sé.....3 PASE A 7.3

7.2.1 ¿Este representante o jefe de manzana ayuda a resolver los problemas o demandas de la comunidad?

CIRCULE UN SOLO CÓDIGO

Si.....1
No.....2
No sé.....3

7.3 De los siguientes grupos, asociaciones u organizaciones que realizan actividades fuera de esta colonia, ¿en cuáles de ellos participan usted o algún miembro de su hogar?

LEA CADA UNA DE LAS OPCIONES Y CIRCULE LA RESPUESTA EN CADA CASO

	Sí	No
Una agrupación religiosa o iglesia	1	2
Una asociación de padres (madres) de familia	1	2
Una asociación o grupo de la tercera edad	1	2
Un sindicato	1	2
Una organización o grupo de vecinos(as)	1	2
Un grupo deportivo o recreativo	1	2
Una asociación de autoayuda (AA, Neuróticos Anónimos)	1	2
Una agrupación o partido político	1	2
Otra (especifique):	1	2

Participación y cooperación

7.4 En los 2 últimos años, ¿usted o alguien de su hogar participó con sus vecinos en alguna obra para beneficio de quienes viven en su calle, barrio o colonia?

CIRCULE UN SOLO CÓDIGO

Sí1 PASE A 7.4.2
No2

7.4.1 ¿Por qué no participaron?

CIRCULE UN SOLO CÓDIGO

No hubo ninguna obra o proyecto1
Acabamos de cambiarnos a esta colonia2
No nos interesó participar3 PASE A 7.9
No pudimos participar por falta de tiempo o dinero4
Otra razón (especifique):5

7.4.2 ¿Cuál fue la principal obra que realizaron en los 2 últimos años?

ESCRIBA LA RESPUESTA TEXTUAL

7.5 En esta obra, ¿en cuáles de las siguientes actividades tuvieron participación usted o algún miembro de su hogar?*LEA CADA UNA DE LAS OPCIONES Y CIRCULE LA RESPUESTA EN CADA CASO*

	Sí	No
Asistieron a alguna reunión de vecinos en la que se discutió la realización de la obra	1	2
Aportaron dinero	1	2
Aportaron materiales o comida	1	2
Realizaron algún tipo de trabajo	1	2
Participaron en un Comité o grupo de vecinos que supervisara que los trabajos se realizaran en el tiempo y con la calidad acordados	1	2
Participaron en un Comité o grupo de vecinos que vigilara el correcto uso de los recursos	1	2
Actualmente, colaboran en el mantenimiento y cuidado de las obras	1	2

7.6 ¿Qué lugar utilizaron principalmente para llevar a cabo reuniones entre los vecinos para conversar o tomar decisiones sobre esta obra?*CIRCULE UN SOLO CÓDIGO*

En la calle	1
En nuestras propias casas	2
En un centro comunitario	3
En un parque	4
En una cancha	5
En una iglesia	6
En otro lugar (especifique):	7
Nunca nos reunimos	8

7.7 ¿Esta obra recibió apoyo de alguna dependencia o programa de gobierno?*CIRCULE UN SOLO CÓDIGO*

Sí 1
 No 2 **PASE A 7.8**

7.7.1 ¿De cuál dependencia o programa de gobierno recibieron apoyo?*ESCRIBA LA RESPUESTA TEXTUAL***7.8.2 ¿Esta organización...***LEA CADA UNA DE LAS OPCIONES Y CIRCULE LA RESPUESTA*

Ya existía antes de iniciar la obra 1
 Se formó durante la realización de la obra 2 **PASE A 7.8.4**

7.8.3 ¿Usted o algún miembro de su familia ya formaban parte de esta organización antes de iniciar la obra, o se integraron durante su realización?*CIRCULE UN SOLO CÓDIGO*

Ya era/eran miembros 1
 Se integró o nos integramos durante la realización de la obra 2 **PASE A 7.9**

7.8.4 Actualmente ¿existe esta organización?*CIRCULE UN SOLO CÓDIGO*

Sí 1
 No 2

7.8 Para llevar a cabo la obra, ¿contaron con el apoyo de alguna organización social, sindical, deportiva, religiosa o política?*CIRCULE UN SOLO CÓDIGO*

Sí 1
 No 2 **PASE A 7.9**

7.8.1 ¿De qué tipo es esta organización?*CIRCULE UN SOLO CÓDIGO*

Una organización de colonos	1
Una asociación religiosa	2
Una asociación deportiva	3
Un sindicato	4
Una asociación o partido político	5
De otro tipo (especifique)	6

7.9 En los 2 últimos años, ¿usted o alguien de su hogar participó con sus vecinos en alguna actividad para resolver otras necesidades o problemas comunes?*CIRCULE UN SOLO CÓDIGO*

Sí 1 **PASE A 7.9.2**
 No 2

7.9.1 ¿Por qué no participaron?*CIRCULE UN SOLO CÓDIGO*

No hubo necesidades o problemas comunes	1
Acabamos de cambiarnos a esta colonia	2
No nos interesa participar	3
No podemos participar por falta de tiempo o dinero	4
Otra razón (especifique)	5

PASE A 7.10**7.9.2 ¿Cuál fue el problema o la necesidad más importante que buscaban resolver?***ESCRIBA LA RESPUESTA TEXTUAL*

7.9.3 ¿Qué lugar utilizaron principalmente para llevar a cabo reuniones entre los vecinos para conversar o tomar decisiones sobre este problema o necesidad?

CIRCULE UN SOLO CÓDIGO

En la calle 1
 En nuestras propias casas 2
 En un centro comunitario 3
 En un parque 4
 En una cancha 5
 En una iglesia 6
 En otro lugar (especifique): 7
 Nunca nos reunimos 8

7.9.4 Para resolver este problema o necesidad, ¿usted o alguien de su familia aportó...

LEA CADA UNA DE LAS OPCIONES Y CIRCULE LA RESPUESTA EN CADA CASO

	Si	No
Dinero	1	2
Trabajo	1	2
Materiales	1	2
Comida	1	2
Tiempo	1	2
Otro (especifique):	1	2

7.9.5 ¿Acudieron a alguna oficina o dependencia de gobierno en busca de apoyo para resolver este problema o necesidad?

CIRCULE UN SOLO CÓDIGO

Sí 1
 No 2 **PASE A 7.9.7**

7.9.6 ¿En cuál oficina o dependencia de gobierno solicitaron apoyo?

ESCRIBA LA RESPUESTA TEXTUAL

7.9.7 ¿Lograron resolver este problema o necesidad?

CIRCULE UN SOLO CÓDIGO

Sí 1
 No 2

Confianza y solidaridad

7.10 De los siguientes grupos de personas u organizaciones, ¿qué tanta confianza le tiene a...

LEA CADA UNA DE LAS OPCIONES Y CIRCULE LA RESPUESTA EN CADA CASO

Grupo u organización

	Ninguna	Algo/ Regular	Mucha
La policía	1	2	3
La familia	1	2	3
Los vecinos	1	2	3
Los líderes o representantes de la colonia	1	2	3
Los partidos políticos	1	2	3
Las amistades	1	2	3
La iglesia	1	2	3
El ejército	1	2	3
Los maestros	1	2	3
Los funcionarios del gobierno	1	2	3

7.11 Ahora se le van a leer un conjunto de frases. De cada una de ellas por favor mencione si está de acuerdo con las mismas.

LEA CADA UNA DE LAS OPCIONES Y CIRCULE LA RESPUESTA EN CADA CASO

	Sí	No
Se puede confiar en la mayoría de las personas de esta colonia	1	2
Puedo dejar a mis hijos encargados con mis vecinos porque sé que estarán bien cuidados	1	2
La mayoría de los vecinos de la colonia está dispuesta a ayudar cuando es necesario	1	2
Aquí en la colonia es necesario estar alerta porque siempre hay quien se quiere aprovechar de los demás	1	2
Es necesario tener armas para defenderse en la colonia	1	2
Se puede caminar de noche por la colonia porque si se ofreciera, los vecinos ayudarían	1	2
Cuando estoy fuera de mi casa, los vecinos están al pendiente de la misma por si algo se ofreciera	1	2
En esta colonia, los vecinos generalmente tienen confianza para pedir o prestar dinero	1	2

98 NO SABE, 99 NO RESPUESTA O NO CONTESTA, 97 INSUFICIENTEMENTE ESPECIFICADO

7.12 Si un proyecto de la comunidad no lo beneficia a usted o a su familia directamente, pero tiene beneficios para muchas otras personas de su colonia, contribuiría con:

LEA CADA UNA DE LAS OPCIONES Y CIRCULE LA RESPUESTA EN CADA

CASO

	Sí	No
Trabajo	1	2
Dinero	1	2
Materiales	1	2
Comida	1	2
Tiempo	1	2
Otro (especifique)	1	2

7.13 Si usted necesitara dinero para pagar algún gasto imprevisto como una enfermedad, ¿existen vecinos a quienes pediría ayuda?

CIRCULE UN SOLO CÓDIGO

Definitivamente sí	1
Probablemente	2
No está seguro	3
Probablemente no	4
Definitivamente no	5

Cohesión social

7.14 ¿Qué tan frecuentemente se reúne usted o su familia con sus vecinos para...

LEA CADA UNA DE LAS OPCIONES Y CIRCULE LA RESPUESTA EN CADA

CASO

Frecuentemente Algunas Nunca
ocasiones

Conversar	1	2	3
Jugar o hacer deporte	1	2	3
Celebrar fiestas	1	2	3
Para tratar problemas personales	1	2	3

7.15 ¿Cómo es la relación de convivencia entre sus vecinos?

LEA CADA UNA DE LAS OPCIONES Y CIRCULE UN SOLO CÓDIGO

Hay buena relación entre todos los vecinos	1
Algunos vecinos tienen conflictos entre sí	2
La relación entre los vecinos es problemática	3
No tengo convivencia con mis vecinos	4

7.16 ¿En su opinión, en su colonia existen conflictos frecuentes ocasionados por diferencias en:

LEA CADA UNA DE LAS OPCIONES Y CIRCULE LA RESPUESTA EN CADA

CASO

	Sí	No
En educación	1	2
Por la posesión de tierras	1	2
En nivel social	1	2
Entre hombres y mujeres	1	2
Entre las generaciones más jóvenes y las más viejas	1	2
Entre residentes antiguos y residentes nuevos	1	2
El lugar de origen	1	2
Por preferencias políticas	1	2
En las creencias religiosas	1	2
En el origen étnico o la lengua que hablan	1	2
Entre pandillas	1	2
Otra (especifique):	1	2

Si todos son NO, PASE A 7.17

7.16.1 ¿Al menos alguno de los problemas causados por alguna de estas diferencias han llevado a situaciones de violencia en su colonia?

CIRCULE UN SOLO CÓDIGO

Sí	1
No	2

PASE A 7.17

7.16.2 ¿Cuál fue esa situación de violencia

ESCRIBA LA RESPUESTA TEXTUAL

Información

7.17 ¿Qué tan enterados están dentro de su familia sobre los problemas de su comunidad?

LEA CADA UNA DE LAS OPCIONES Y CIRCULE UN SOLO CÓDIGO

Siempre estamos enterados	1
Nos enteramos ocasionalmente	2
Nunca nos enteramos	3

7.18 Entre sus vecinos comparten información útil para...

LEA CADA UNA DE LAS OPCIONES Y CIRCULE LA RESPUESTA EN CADA
CASO

	Sí	No
Identificar problemáticas comunes	1	2
Obtener un empleo	1	2
Acceder a algún servicio	1	2
Defender sus derechos	1	2
Enterarse de las acciones o programas de gobierno	1	2
Llevar a cabo proyectos comunitarios	1	2
Cuidar el medio ambiente	1	2
Prevenir riesgos por desastres naturales	1	2

Appendix C

Proofs

C.1 Identification

Proof. Proof of lemma 1

Identification of the quantile for: $q_{1,\pi_g,\tau}$

The definition of τ is given by:

$$\tau = \mathbb{P}(Y_{ig}^{**}(1, \pi_g) \leq q_{1,\pi_g,\tau})$$

Under the assumption **A3** (independence):

$$\tau = \mathbb{P}(Y_{ig}^{**}(1, \pi_g) \leq q_{1,\pi_g,\tau} | S_g = \pi_g, D_{ig} = 1)$$

Replacing $Y_{ig} = \sum_{\pi_g \in \Pi} \left\{ Y_{ig}^{**}(1, \pi_g) D_{ig} \mathbb{I}\{S_g = \pi_g\} + Y_{ig}^{**}(0, \pi_g) (1 - D_{ig}) \mathbb{I}\{S_g = \pi_g\} \right\}$:

$$\tau = \mathbb{P}(Y_{ig} \leq q_{1,\pi_g,\tau} | S_g = \pi_g, D_{ig} = 1)$$

Knowing that $\mathbb{P}(A) = \mathbb{E}[\mathbb{I}\{A\}]$. When replacing it multiplies by D_{ig} (because it is conditional on $D_{ig} = 1$) and $\mathbb{I}\{S_g = \pi_g\}$ (because it is conditional in the group $S_g = \pi_g$):

$$\tau = \mathbb{E}[D_{ig} \mathbb{I}\{S_g = \pi_g\} \mathbb{I}\{Y_{ig} \leq q_{1,\pi_g,\tau}\} | S_g = \pi_g, D_{ig} = 1]$$

Under the law of total probability:

$$\begin{aligned} \mathbb{E}[Y | S_g = \pi_g] &= \pi_g \mathbb{E}[Y | S_g = \pi_g, D_{ig} = 1] + (1 - \pi_g) \mathbb{E}[Y | S_g = \pi_g, D_{ig} = 0] \\ &\xrightarrow{D_{ig}=1} \mathbb{E}[Y | S_g = \pi_g, D_{ig} = 1] = \frac{\mathbb{E}[Y | S_g = \pi_g]}{\pi_g} \end{aligned}$$

Rewriting

$$\tau = \mathbb{E} \left[\frac{D_{ig}}{\pi_g} \mathbb{I}\{S_g = \pi_g\} \mathbb{I}\{Y_{ig} \leq q_{1,\pi_g,\tau}\} \middle| S_g = \pi_g \right]$$

Applying the law of iterated expectations: $\mathbb{E}[\mathbb{E}[Y|X]] = p_1\mathbb{E}[Y|x = c_1] + p_2\mathbb{E}[Y|x = c_2] + \dots + p_M\mathbb{E}[Y|x = c_M] = \mathbb{E}[Y] \xrightarrow{x=c_1} \mathbb{E}[Y|x = c_1] = \frac{\mathbb{E}[Y]}{p_1}$

$$\tau = \mathbb{E} \left[\frac{D_{ig}}{\pi_g} \frac{\mathbb{I}\{S_g = \pi_g\}}{\mathbb{P}(S_g = \pi_g)} \mathbb{I}\{Y_{ig} \leq q_{1,\pi_g,\tau}\} \right]$$

Identification of the quantile for: $q_{0,\pi_g,\tau}$

The definition of τ is given by:

$$\tau = \mathbb{P}(Y_{ig}^{**}(0, \pi_g) \leq q_{0,\pi_g,\tau})$$

Under the assumption **A3** (independence):

$$\tau = \mathbb{P}(Y_{ig}^{**}(0, \pi_g) \leq q_{0,\pi_g,\tau} | S_g = \pi_g, D_{ig} = 0)$$

Replacing $Y_{ig} = \sum_{\pi_g \in \Pi} \left\{ Y_{ig}^{**}(1, \pi_g) D_{ig} \mathbb{I}\{S_g = \pi_g\} + Y_{ig}^{**}(0, \pi_g) (1 - D_{ig}) \mathbb{I}\{S_g = \pi_g\} \right\}$:

$$\tau = \mathbb{P}(Y_{ig} \leq q_{0,\pi_g,\tau} | S_g = \pi_g, D_{ig} = 0)$$

Knowing that $\mathbb{P}(A) = \mathbb{E}[\mathbb{I}\{A\}]$. When replacing it multiplies by $(1 - D_{ig})$ (because it is conditional on $D_{ig} = 0$) and $\mathbb{I}\{S_g = \pi_g\}$ (because it is conditional in the group $S_g = \pi_g$):

$$\tau = \mathbb{E}[(1 - D_{ig}) \mathbb{I}\{S_g = \pi_g\} \mathbb{I}\{Y_{ig} \leq q_{0,\pi_g,\tau}\} | S_g = \pi_g, D_{ig} = 0]$$

Under the law of total probability:

$$\begin{aligned} \mathbb{E}[Y|S_g = \pi_g] &= \pi_g \mathbb{E}[Y|S_g = \pi_g, D_{ig} = 1] + (1 - \pi_g) \mathbb{E}[Y|S_g = \pi_g, D_{ig} = 0] \\ &\xrightarrow{D_{ig}=0} \mathbb{E}[Y|S_g = \pi_g, D_{ig} = 0] = \frac{\mathbb{E}[Y|S_g = \pi_g]}{(1 - \pi_g)} \end{aligned}$$

Rewriting

$$\tau = \mathbb{E} \left[\frac{(1 - D_{ig})}{(1 - \pi_g)} \mathbb{I}\{S_g = \pi_g\} \mathbb{I}\{Y_{ig} \leq q_{0,\pi_g,\tau}\} \middle| S_g = \pi_g \right]$$

Applying the law of iterated expectations: $\mathbb{E}[\mathbb{E}[Y|X]] = p_1\mathbb{E}[Y|x = c_1] + p_2\mathbb{E}[Y|x = c_2] + \dots + p_M\mathbb{E}[Y|x = c_M] = \mathbb{E}[Y] \xrightarrow{x=c_1} \mathbb{E}[Y|x = c_1] = \frac{\mathbb{E}[Y]}{p_1}$

$$\tau = \mathbb{E} \left[\frac{(1 - D_{ig}) \mathbb{I}\{S_g = \pi_g\}}{(1 - \pi_g) \mathbb{P}(S_g = \pi_g)} \mathbb{I}\{Y_{ig} \leq q_{0,\pi_g,\tau}\} \right]$$

Identification of the quantile for: $q_{0,0,\tau}$

The definition of τ is given by:

$$\tau = \mathbb{P}(Y_{ig}^{**}(0, S_g = 0) \leq q_{0,0,\tau})$$

Under the assumption **A3** (independence):

$$\tau = \mathbb{P}(Y_{ig}^{**}(0, S_g = 0) \leq q_{0,0,\tau} | S_g = 0, D_{ig} = 0)$$

Replacing $Y_{ig} = \sum_{\pi_g \in \Pi} \left\{ Y_{ig}^{**}(1, \pi_g) D_{ig} \mathbb{I}\{S_g = \pi_g\} + Y_{ig}^{**}(0, \pi_g) (1 - D_{ig}) \mathbb{I}\{S_g = \pi_g\} \right\}$:

$$\tau = \mathbb{P}(Y_{ig} \leq q_{0,0,\tau} | S_g = 0, D_{ig} = 0)$$

Knowing that $\mathbb{P}(A) = \mathbb{E}[\mathbb{I}\{A\}]$. When replacing it multiplies by $(1 - D_{ig})$ (because it is conditional on $D_{ig} = 0$) and $\mathbb{I}\{S_g = 0\}$ (because it is conditional in the group $S_g = 0$):

$$\tau = \mathbb{E}[(1 - D_{ig}) \mathbb{I}\{S_g = 0\} \mathbb{I}\{Y_{ig} \leq q_{0,0,\tau}\} | S_g = 0, D_{ig} = 0]$$

Under the law of total probability:

$$\begin{aligned} \mathbb{E}[Y | S_g = \pi_g] &= \pi_g \mathbb{E}[Y | S_g = \pi_g, D_{ig} = 1] + (1 - \pi_g) \mathbb{E}[Y | S_g = \pi_g, D_{ig} = 0] \\ &\xrightarrow{D_{ig}=0} \mathbb{E}[Y | S_g = \pi_g, D_{ig} = 0] = \frac{\mathbb{E}[Y | S_g = \pi_g]}{(1 - \pi_g)} \end{aligned}$$

Rewriting

$$\tau = \mathbb{E} \left[(1 - D_{ig}) \mathbb{I}\{S_g = 0\} \mathbb{I}\{Y_{ig} \leq q_{0,0,\tau}\} \middle| S_g = 0 \right]$$

Applying the law of iterated expectations: $\mathbb{E}[\mathbb{E}[Y | X]] = p_1 \mathbb{E}[Y | x = c_1] + p_2 \mathbb{E}[Y | x = c_2] + \dots + p_M \mathbb{E}[Y | x = c_M] = \mathbb{E}[Y] \xrightarrow{x=c_1} \mathbb{E}[Y | x = c_1] = \frac{\mathbb{E}[Y]}{p_1}$

$$\tau = \mathbb{E} \left[(1 - D_{ig}) \frac{\mathbb{I}\{S_g = 0\}}{\mathbb{P}(S_g = 0)} \mathbb{I}\{Y_{ig} \leq q_{0,0,\tau}\} \right]$$

□

Proof. Proof of lemma 2

Identification of the proportion for: $p_{1,\pi_g,\tau}$

The definition of $p_{1,\pi_g,\tau}$ is given by:

$$p_{1,\pi_g,\tau} = \mathbb{P}(Y_{ig}^{**}(1, \pi_g) \leq q_{\pi_g,\tau})$$

Under the assumption **A3** (independence):

$$p_{1,\pi_g,\tau} = \mathbb{P}(Y_{ig}^{**}(1, \pi_g) \leq q_{\pi_g,\tau} | S_g = \pi_g, D_{ig} = 1)$$

Replacing $Y_{ig} = \sum_{\pi_g \in \Pi} \left\{ Y_{ig}^{**}(1, \pi_g) D_{ig} \mathbb{I}\{S_g = \pi_g\} + Y_{ig}^{**}(0, \pi_g) (1 - D_{ig}) \mathbb{I}\{S_g = \pi_g\} \right\}$:

$$p_{1,\pi_g,\tau} = \mathbb{P}(Y_{ig} \leq q_{\pi_g,\tau} | S_g = \pi_g, D_{ig} = 1)$$

Knowing that $\mathbb{P}(A) = \mathbb{E}[\mathbb{I}\{A\}]$. When replacing it multiplies by D_{ig} (because it is conditional on $D_{ig} = 1$) and $\mathbb{I}\{S_g = \pi_g\}$ (because it is conditional in the group $S_g = \pi_g$):

$$p_{1,\pi_g,\tau} = \mathbb{E}[D_{ig} \mathbb{I}\{S_g = \pi_g\} \mathbb{I}\{Y_{ig} \leq q_{\pi_g,\tau}\} | S_g = \pi_g, D_{ig} = 1]$$

Under the law of total probability:

$$\begin{aligned} \mathbb{E}[Y | S_g = \pi_g] &= \pi_g \mathbb{E}[Y | S_g = \pi_g, D_{ig} = 1] + (1 - \pi_g) \mathbb{E}[Y | S_g = \pi_g, D_{ig} = 0] \\ &\xrightarrow{D_{ig}=1} \mathbb{E}[Y | S_g = \pi_g, D_{ig} = 1] = \frac{\mathbb{E}[Y | S_g = \pi_g]}{\pi_g} \end{aligned}$$

Rewriting

$$p_{1,\pi_g,\tau} = \mathbb{E} \left[\frac{D_{ig}}{\pi_g} \mathbb{I}\{S_g = \pi_g\} \mathbb{I}\{Y_{ig} \leq q_{\pi_g,\tau}\} \middle| S_g = \pi_g \right]$$

Applying the law of iterated expectations: $\mathbb{E}[\mathbb{E}[Y|X]] = p_1 \mathbb{E}[Y|x = c_1] + p_2 \mathbb{E}[Y|x = c_2] + \dots + p_M \mathbb{E}[Y|x = c_M] = \mathbb{E}[Y] \xrightarrow{x=c_1} \mathbb{E}[Y|x = c_1] = \frac{\mathbb{E}[Y]}{p_1}$

$$p_{1,\pi_g,\tau} = \mathbb{E} \left[\frac{D_{ig}}{\pi_g} \frac{\mathbb{I}\{S_g = \pi_g\}}{\mathbb{P}(S_g = \pi_g)} \mathbb{I}\{Y_{ig} \leq q_{\pi_g,\tau}\} \right]$$

The proof for objects $p_{0,\pi_g,\tau}$ and $p_{0,0,\tau}$ follow the same steps. □

C.2 Unconditional quantile

In general, define $f(x, y(x)) = c \Rightarrow f(x, y(x)) - c = f^*(x, y(x)) = 0$. The following derivative $\frac{dy}{dx}$ can be calculated using the implicit function theorem:

$$\begin{aligned} \frac{\partial f(x, y(x))}{\partial x} + \frac{\partial f(x, y(x))}{\partial y} \frac{dy}{dx} - \frac{dc}{dx} &= 0 \\ \Rightarrow \frac{dy}{dx} &= -\frac{\frac{\partial f(x, y(x))}{\partial x}}{\frac{\partial f(x, y(x))}{\partial y}} = -\frac{\frac{\partial f^*(x, y(x))}{\partial x}}{\frac{\partial f^*(x, y(x))}{\partial y}} \end{aligned}$$

Proof. Saturation effects on the unconditional quantile

Using the law of iterated expectations (Wooldridge 2010, p.31):

$$\tau = F_{Y|S_g=\pi_g}(q_{\pi_g, \tau}) = \pi_g \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1] + (1 - \pi_g) \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0]$$

rewriting:

$$H = \tau - \pi_g \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1] - (1 - \pi_g) \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0]$$

Deriving equation H with respect to π_g

$$\begin{aligned} \frac{\partial H}{\partial \pi_g} &= -\mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1] - \pi_g \frac{\partial \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1]}{\partial \pi_g} - \\ &\quad - \frac{\partial \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0]}{\partial \pi_g} + \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0] + \\ &\quad + \pi_g \frac{\partial \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0]}{\partial \pi_g} = \\ &= -\mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1] + \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0] - \\ &\quad - \pi_g \frac{\partial \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1]}{\partial \pi_g} - (1 - \pi_g) \frac{\partial \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0]}{\partial \pi_g} \end{aligned}$$

Knowing that $\frac{\partial F_Y(a|x)}{\partial a} = f_Y(a|x)$ and deriving equation H with respect to $q_{\pi_g, \tau}$:

$$\begin{aligned} \frac{\partial H}{\partial q_{\pi_g, \tau}} &= -\pi_g f_Y(q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1) - (1 - \pi_g) f_Y(q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0) = \\ &= -[\pi_g f_Y(q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1) + (1 - \pi_g) f_Y(q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0)] = -f_Y(q_{\pi_g, \tau}) \end{aligned}$$

Applying the implicit function theorem:

$$\begin{aligned} \frac{dq_{\pi_g, \tau}}{d\pi_g} &= -\frac{\frac{\partial H}{\partial \pi_g}}{\frac{\partial H}{\partial q_{\pi_g, \tau}}} = \\ &= -\left(\left(\frac{-1}{f_Y(q_{\pi_g, \tau})} \right) \left[-\mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1] + \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0] - \right. \right. \\ &\quad \left. \left. - \pi_g \frac{\partial \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1]}{\partial \pi_g} - (1 - \pi_g) \frac{\partial \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0]}{\partial \pi_g} \right] \right) \end{aligned}$$

Simplifying :

$$\begin{aligned} \frac{dq_{\pi_g, \tau}}{d\pi_g} &= \overbrace{\frac{-1}{f_Y(q_{\pi_g, \tau})} \left(\mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1] - \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0] \right)}^{\text{Marginal Quantile Direct Effect (MQDE)}} + \\ &\quad \underbrace{\frac{-1}{f_Y(q_{\pi_g, \tau})} \left(\pi_g \frac{\partial \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 1]}{\partial \pi_g} + (1 - \pi_g) \frac{\partial \mathbb{P}[Y_{ig} \leq q_{\pi_g, \tau} | S_g = \pi_g, D_{ig} = 0]}{\partial \pi_g} \right)}_{\text{Marginal Quantile Spillover Effect (MQSPE)}} \end{aligned}$$

□

The **A1-A2** assumptions have no explicit role in the proof. Such assumptions allow the potential outcome $Y_{ig}(d_{ig}, \pi_g)$ to depend only on the treatment status and the saturation level.