The Effect of Location Based Subsidies on the Housing Market

Juan Pablo Uribe*

April 23, 2021

Abstract

This paper estimates the effect of a location-based price subsidy of utilities and property taxes on the housing market in Bogotá, Colombia. In Bogotá, neighborhood blocks are divided into 6 subsidy codes using arbitrary cutoffs on a block quality score. I use the discontinuity introduced by the cutoffs in a Regression Discontinuity Design to study the effect of the subsidy on the housing market. I find that blocks receiving a higher subsidy have newer houses - implying more construction in such areas. I also find a capitalization of the subsidy into housing prices: properties in areas receiving a small subsidy are cheaper than those receiving a bigger subsidy. These results suggest that a careful evaluation of location-based subsidies' incidence needs to consider potential capitalization into the housing market and other unintended effects such as new construction or renovations.

*I am grateful to Matt Turner, Jesse Shapiro, Andrew Foster, Ken Chay, John Friedman, Bryce Steignberg, Stefan Hut, Juan David Herreno, Diego Ramos, Nick Reynolds,Juan Miguel Gallego, Carlos Sepulveda and seminar participants at Brown University,LACEA conference in Puebla and NEUDC. I am also grateful to the District Department of Planning of for useful conversation and facilitating the data. In particular, I want to thank Ariel Carrero, Antonio Avendano and David Monroy. I also want to thank the CEDE data center from the Universidad de Los Andes Economics's' department for providing me access to different data sources. Brown University: juan_uribe@brown.edu

I. Introduction

A challenge for social assistance programs and redistributive policies is to find the poor. It is particularly challenging in settings with high informality levels, where income-based, means-tested transfers do not get to the right people.[1] An alternative targeting tool is geographic location, such as neighborhoods, states, or cities. This targeting approach is particularly appealing to developing countries, where self-employment and economic informality are prevalent. Yet, location-based subsidies could affect people's decisions about where to live, and this could affect the housing market. We need to know how the housing market is affected before evaluating the efficacy of place-based redistributive policies. Standard urban economics theory would predict the subsidy's capitalization into the housing prices (i.e., housing prices in the subsidized areas are higher).[2] If the subsidized areas become more expensive, the subsidy benefiting the landowners could affect tenants. The subsidy can also attract more people and new construction, affecting city growth. The significance of these effects is an empirical question and the main purpose of this paper.

I use a location-based subsidy scheme in Bogotá, Colombia and I quantify the effect of the subsidies on the housing market. Bogotá uses the characteristics of neighborhoods to target subsidies on public utility services (i.e. water, gas, electricity, phone, and internet), and to charge differential rates on property taxes. Each city block receives a subsidy code based on neighborhood block quality scores. There are six subsidy codes called *estratos*. Houses in estrato 1,2 and 3 receive subsidies, those in estrato 4 pay the market price and those in estratos 5 and 6 pay a tax. Neighborhood score cutoffs are arbitrarily defined, making Bogotá an ideal setting to study subsidy effects. Bogotá's scheme allows me to use quasi-experimental variation that is rare in this type of policy.

[1]See Hanna and Olken (2018) for a summary of the literature studying these proxy-mean tests to target subsidies.

[2]For example, Glaeser and Gottlieb (2008) suggest that "if high prices and low amenities offset high wages in a spatial equilibrium, there is nothing particularly equitable about taking money from rich places and giving it to poor places." In a recent paper Gaubert, Kline, and Yagan (2020) reconsider this view and suggest that using location as a targeting tool can be a successful redistributive tool under certain conditions. The empirical evidence is mixed. Chen, Glaeser, and Wessel (2019) study if the housing market in the USA responds to the Opportunity Zone program. They rule out price impacts bigger than 1.3 %. Lutz (2015) uses a school finance reform in New Hampshire and finds that lower property taxes generate new residential construction. He finds a higher response in areas with a higher elasticity of housing supply. Places with low elasticity of housing supply experienced an increase in prices. The property taxes are capitalized into housing prices.

In particular, I use the discontinuities introduced by the cutoffs defining the subsidy level, to study the effect of the subsidy scheme on the housing market. I focus on houses receiving a high subsidy (*estrato* 2) and a low subsidy (*estrato* 3). These two codes concentrate 57 percent of all housing units, which allows me get enough blocks to implement my Research Design. Additionally, the models for valuation of housing prices, are common across these two estratos,[3] and both *estratos* receive subsidies. By focusing on these two estratos, I minimize the potential effect of a behavioral response to avoid a stigma associated with receiving the subsidy and being labeled as poor.

I collect different sources of administrative data to replicate the neighborhood block quality score and to measure characteristics of the housing market. The neighborhood block quality score and *estratos* come from the Stratification Census. Using the replicated score, I correctly classify 99.9 percent of the city blocks in 1997, when the assignment method was first introduced. I use the Cadaster Census to construct measure of property size, quality and construction timing. The quality measure is a score that summarizes the interior characteristics of the property (i.e bathroom size, type of floor and materials of the kitchen). For construction timing, I observe the date when the structure was built in each lot. Having the construction date allows me to see the effect of the subsidy scheme on new construction. For land and structure price, I observe market price appraisals for each unit. These appraisals are the base for the assessed values used for property tax purpose.

I find that the probability of new construction is 43 percent higher in the heavily subsidized areas,that pay around 30 percent less in utilities and property taxes. Consistently, the average age of the units is lower. Additionally, there is some evidence of a higher quality housing and no clear differences in property size. The heavily subsidized areas are more expensive. Most of the increase in value is explained by the age of the property, the quality and the increase in land prices. I can not reject the null hypothesis that difference in prices is equal to a benchmark calculation of full capitalization.

This paper contributes to the literature on capitalization of subsidies and taxes into house prices, and more generally the literature studying the effect of policies on the housing market.Hilber (2017) highlights that empirical evidence for capitalization subsidies and taxes into housing prices is limited. On the other hand, the body of literature studying effect of different policies and regulations on the housing market is large but concentrated

[3]The officials use a different model for codes 1,2,3 and 4,5,6 and a different model for single-family units and multi-family units

in the USA.[4] As mentioned by Gyourko and Molloy (2015), while some papers investigate land regulations in England, Spain and Shanghai, the research for the developing world is scarce. This paper contributes by studying a different type of policy in a developing country using a novel Research Design.

I also contribute to the literature studying the effectiveness of location-based policies. There is little evidence of the efficacy of location-based subsidies, particularly outside of the USA. Kline and Moretti (2014) present a framework to study location-based policies and suggest that the effects of these type of policies are not well-studied and that it is important to identify who is benefiting from them. Normally the literature on location-based subsidies focuses on firms and efficiency losses. However there could be redistributive reasons to support location-based policies. Gaubert et al. (2020) propose a framework to study the equity-efficiency trade-offs when targeting policies to location and not individuals based on their income. They present some conditions under which location based subsidies could be effective redistribution tools.[5] In this paper, I provide empirical evidence on the effect of a subsidy on the housing market. I argue that the housing market effects should be included when evaluating the potential welfare gains of using location as a targeting tool.

Finally, I contribute to the literature studying the *estratos* as a targeting tool for subsidies in Colombia.The public utilities subsidies assigned using the estratos are an important social assistance program. They represent 0.27% of Colombia GDP. Despite the big proportion of national expenditure required to fund this subsidy, and the potential forgone revenue in property taxes,little is known about its cost effectiveness.[6] There are two conflicting papers that address a similar question. Medina and Morales (2007) use

[4]For example,Bayer, Ferreira, and McMillan (2007), Black (1999) show that houses in districts with better schools are more expensive. Chay and Greenstone (2005) find that less polluted counties have higher property prices. Chen et al. (2019) study how Opportunity Zone Program on housing prices. There is evidence of housing market responses to policies into particular locations. Land prices in regulated areas are higher. Turner, Haughwout, and van der Klaauw (2014) find that land regulation has an impact on land prices in the USA. Lutz (2015) uses a reform in school finances in New Hampshire and finds that lower taxes is associated with new construction. It also increases in property values in places where the elasticity of housing supply is lower. Haan and Simmler (2018) study the incidence of subsidies for wind energy on agricultural land prices in Germany.

[5]In particular, location-based subsidies could be effective if less skilled households are concentrated in Distressed areas, if few households are indifferent between targeted and not targeted locations, if productivity differences across areas are small, or if the marginal utility of consumption declines slowly with income.

[6]The other large social assistance program, the Colombian Conditional Cash Transfer, *Familias en Acción* , is 0.35 percent of the GDP. Excluding the utility subsidies, the total expenditure on social assistance programs was 0.49 of the GDP in 2014. (Harker, Lustig, Martínez, & Melendez, 2016).

a boundary discontinuity design and find evidence of capitalization. Gallego, Montoya, and Sepúlveda (2016) use the cutoff points from the score for Bogotá and find higher prices in not subsidized areas. These conflicting results suggest that the discussion remains unresolved. In contrast to the analysis by Gallego et al. (2016), I recover the assignment score for all the blocks in the city. I contribute to this discussion, by refining the existing analysis and providing evidence of the effects on new construction and other housing outcomes. As suggested in other settings by McRae (2015), Casas (2014) and Meléndez (2004), I also find that the subsidies are generating distortions in the housing market. These distortions could offset the primary redistributive goal.

The rest of the paper proceeds as follows. I first introduce the targeting tool and the subsidies in the institutional background section. Then I explain the assignment rule of the estratos and how I use it in my Identification Strategy. Then, I explain the data, show the results and present some robustness exercises. Finally, I give some concluding remarks.

II. Institutional background

1. Targeting Tool- the *estratos*, and the Subsidies

In 1994, law 142, mandated all municipalities to use a standard method to classify the houses into six different subsidy codes- the *estratos*. Each residential block in a Colombian city has an *estrato*. The *estratos* comprise a coding system designed to reflect the quality and urban characteristics of each block. In this paper, I focus on Bogotá, which implemented the new stratification method in 1997.[7] Figure 1a shows the distribution of the codes in Bogotá.

The purpose of the estratos is to create a targeting tool to materialize the criteria of solidarity and income redistribution contemplated in the tariff regime for public utility services.[8] The providers of public utilities such as electricity, water, sewage, etc., and property taxes authorities use the *estratos* to charge different tariffs. The public services have a cross-subsidy price scheme. Depending on the estrato of the house, people pay a subsidized price, pay the *market price*, or a tax. Production costs determine the regulated *market price*.[9] People living in properties located in estrato 5 and 6 pay more

[7]The city subsidizes public services since 1983,but the targeting tool to subsidize utilities was different. (Decreto Distrital 1140, 1983.)

[8]See law 142 of 1994 . This law regulates the provision of public services in the country.

[9]The market price is a regulated price that considers the production cost and a reasonable margin

than the *market price*, those in estrato 4 pay the market price, and those in estratos 1,2 or 3 pay less. Figure 1b shows the electricity price scheme. The y-axis represents the price by kwh relative to the market price. The x-axis shows the different estratos. In 2017, houses in *estrato* 4 paid a market price of 40.84$\$COP$ per kwh. In the figure, this is normalized to be 1. Households living in estrato 1 receive a 66 percent price discount, paying 44 percent of the *market price*. Those in estrato 2 receive a 45 percent discount, and those in estrato 3 receive a 15 percent discount. The higher-numbered estratos and public funds subsidize this discount. The households in estratos 5 and 6 pay 20 percent more than the *market price*.

Other policies rely on this targeting tool to charge differential rates. For example, in Bogotá, the property tax varies on two dimensions, the value of the property, and the estrato of the property. Figure 1c shows the property tax scheme for three different property values. I selected the prices to show the important breaks in house values, and to make the figure comparable to figure 1b, I show the property tax relative to that of properties with the same value but in *estrato* 4. In general, the differential tax rate is similar to the electricity price scheme. The property tax rate for houses in estratos 1 and 2 is around 62% of the rate for houses of the same value in estrato 4. If the house costs less than 43 million Colombian Pesos (COP), they do not pay taxes. Depending on their price, houses in estrato 3 pay between 81 percent and 84 percent of what they would pay if they were in estrato 4. Houses in estratos 5 and 6 pay between 10 and 30 percent more than what they would pay in estrato 4.[10]

of benefits for the provider.

[10]In 2016, there was a reform to the tax code. Starting in January 2017, the property tax only depends on the value of the house. Therefore, the estratos do not longer serve as a way to target differential rates for the property tax. Other subsidies, like differential rates for public higher education institutions use the estratos as a targeting tool. Additionally, the estratos are important in terms of signaling. It is a very salient characteristic and plays an important role with social status.

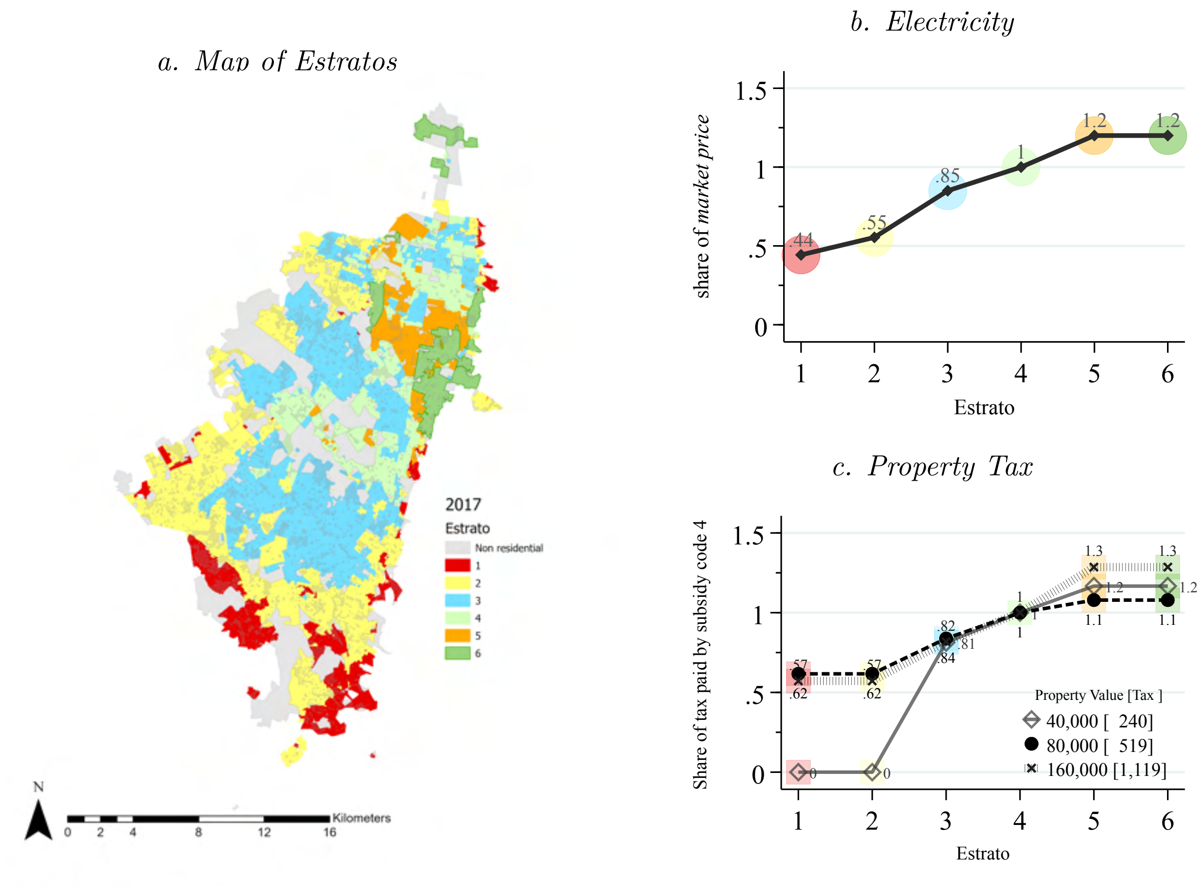


Figure 1: THE ESTRATOS AND THE SUBSIDIES IN BOGOTÁ

NOTE: Figure 1a shows the map Map of Bogota in 2017. Each color represents one of the 6 different subsidy codes-*estratos*. The figure 1b shows the electricity price scheme. The y-axis represents the price by kwh relative to the market price. The x-axis shows the different estratos. The market price is regulated and set as a function of the production costs. In 2017 the price was 440.84 COP/kwh. All of the households start paying the market price for each kwh consumed above the basic consumption of 130 kwh. The price discount only applies for consumption below 130kwh. For every extra kwh consumed above 130kwh, subsidized households pay the market price. The price scheme for the other utility services is very similar. The price scheme is similar for the other utilities (gas, water, sewage, cleaning services, phone, and internet). The figure 1c shows the property tax rate relative to the rate a house of the same value would pay if located in subsidy code 4. The x-axis represents the estratos. The y-axis represents the share of tax a house in estrato 4 would pay. Each line in the figure represents the tax scheme for a given property value. The property value and the tax rate for *estrato* 4 are at the bottom right of the figure in brackets. *Sources:* Enel-Codensa Energy Rates and Article 2 of the agreement 105 of 2003

2. How big is the subsidy?

To compare the subsidy scheme with housing prices, I convert the monthly subsidies on utilities and property taxes into a stock variable.[11] I calculate the average subsidy in each estrato using the aggregate records from the public services regulator (SUI, 2020). I use the annual subsidy on electricity, water, gas, cleaning services, and the sewage system in 2011. For the property tax, I calculate the mean property tax rate, using the Cadaster appraisals and applying the tax formula. Figure 2a shows the Net Present Value (NPV) of the subsidies and property taxes. As a reference point, I also show the average housing price. The figure shows that the subsidies are important relative to the housing prices. Figure 2b shows the difference in the property tax and subsidies between estratos. This difference is a benchmark for a complete capitalization of the subsidy.[12] This *naive* comparison shows the opposite of a capitalization effect. For example, the difference between estratos 2 and 3 is 18.8 million COP, around 35 percent of the average house price in *estrato 2*. Despite the fact that they pay more on utilities, and property taxes, the average price of a house in estrato 3 is also higher. However, average houses and blocks in each estrato are not necessarily comparable; see appendix figure 6 as an example. The purpose of this paper is precisely to find a valid counterfactual to make a valid comparison of houses and blocks that are comparable but are classified in a different estrato. When I do that, the pattern shown in this figure is reversed, and I cannot reject a full capitalization of the subsidy. Houses in comparable houses in estrato 2 are more expensive than those in estrato 3.

[11] To do that, I use the net present value of an annual over a period of 30 years. I use as a discount factor, β, the interest rates for a 17 year national bond, $\beta = 0.0365$. $NPV(x) = \sum_{t=1}^{30} \frac{x}{(1+\beta)^t}$

[12] An alternative way to analyse the size of the subsidy is to take the monthly savings or taxes payed because of the subsidy scheme. The households living in estrato 2 received on average 87 thousand COP per month, and households in estrato 3 around 25 thousand COP per month. Households in the best neighbourhoods pay a tax of around one million COP per year. Appendix Figure 1a shows the average subsidy or contribution for the average household in each subsidy code in Bogotá in 2011. Appendix Figure 1b shows the subsidy relative to the average income in each subsidy code. The subsidy represents around 8 percent of the annual income of households living in code 1 and less than 2 percent of the income of households in code 3.

a. NVP of subsidies and housing prices *b. Difference between estratos*

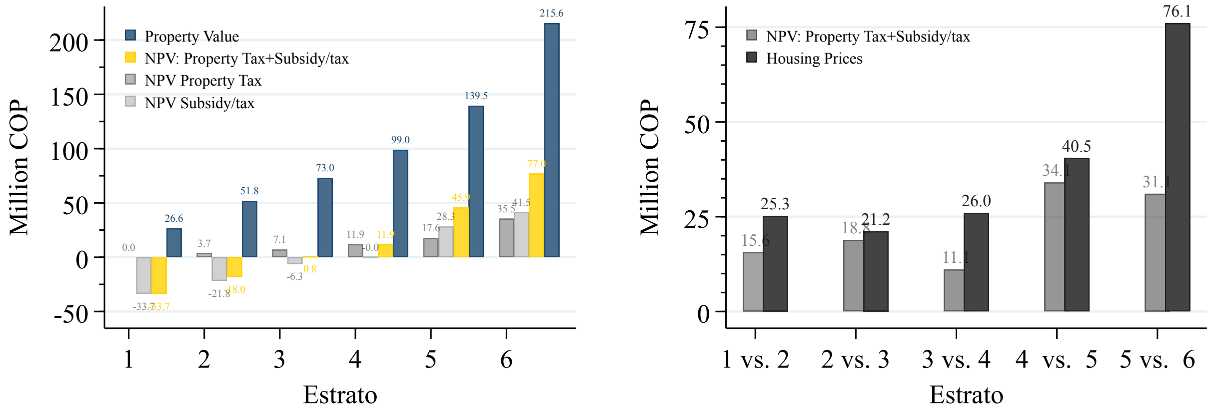


Figure 2: Subsidies and Housing Prices

Note: The value of utilities to calculate the NPV is the average yearly payments for a household in each *estrato*. I include expenditures in electricity, water, gas, cleaning services, and the sewage system in 2011. The data comes from SUI (2020). To calculate the property taxes, I take the average house appraisal in each *estrato* and I apply the tax formula. The property value is the average appraisals for each *estrato*.

III. Data

To study the subsidy scheme's effect on the housing market, I need to recover the estratos' assignment formula and observe housing characteristics, including housing prices. For this, I rely on administrative data. The two main data sources are the stratification census and the Cadaster census for Bogotá. Additionally, I use population census and some historical and recent shapefiles to create a constant geographic unit over time. In this section, I describe the different data sources.

1. Housing Characteristics

The *district cadaster* has a census of all the properties in the city. The authorities use the census to estimate the housing prices for each property.[13] With the census information, I construct variables to study the effect on new construction, house characteristics, and housing prices.

To analyze the subsidy's impact on the construction of new units, I test if the subsidized

[13]The 2017 and 2018 censuses are publicly available for a subset of variables. I only observe the appraisal and construction date for each unit for 2011. That is why my analysis uses 2011. I accessed the restricted data trough the CEDE data center, at the Universidad de Los Andes.

areas have a higher concentration of new properties. I use the construction date, and create an indicator variable for housing units built after the introduction of the stratification system. I also evaluate the effect on the average age of the unit's structure in each block.

Individuals could make quality or size upgrades in their units if the elasticity of substitution between consumption in public services and housing quality is negative. To study the effect on construction size, I use the construction's reported size in square meters. For the quality, I use the physical characteristics collected in the cadaster census. The data has measures for the kitchen's quality, the floors, and other detailed features of the houses. My quality measure is the *puntaje catastral*, a score summarizing the overall quality of the interior of the house.[14]

2. Housing Prices

The other outcome of interest is housing and land prices. Unlike the US and the UK, Bogota does not have a systematic data collection system of transactions to use as my main price data. However, I can observe the appraisals and assessments of the houses for all the properties in Bogota. This measure has some advantages and some disadvantages. The main advantage is that I have a price measure for all the properties using the same criteria. Ideally, I would like to observe transaction prices for all the existing units in the city. Transactions prices will give me the market valuation of the property. However, the decision of selling a house could be affected by the subsidy, and therefore a sample of transactions could be bias. Also, only a subset of units is sold simultaneously, which creates power issues for identification in my setting.

I have overcome this challenge by having appraisal for all the properties. The district authorities try to get accurate market value for each housing unit for tax purposes. The city is widely recognized for its quality on the appraisals (Tsivanidis, 2018). In their effort to get market prices, they appraise a subset of properties and combine it with transaction data, the cadaster census, and neighborhood amenities. They calculate a property price for all the city's properties. This creates the problem that for many of the units I observe, the property price is model-based. This could generate concerns that I may be capturing some mechanical effects of the valuation formula or the appraiser.

[14]Note that this is a different index that the block quality score use in the assignment of the estratos. First, it is constructed for a unit and not a block level. Second, it measures the interior quality in contrast with the block quality score's exterior characteristics. For a complete description of the score, see IGAC instructions document.

Additionally, it is usually the case that the valuations and assessed prices are below the market price.[15] With those caveats in mind, I use these appraisals as my housing prices measures.

The appraisals have several steps. First, the district acquires market prices for a subset of properties. These prices are transaction prices from property sales and individual appraisals by district appraisers.[16] Second, a statistical model relates these market values to a set of observable characteristics in the cadaster census and some amenities. The model includes variables like quality, age, and the type of construction. It also considers land use, location, general nearby amenities, like available parks, and access to public transportation, and estimates of land prices. The model includes 48 different attributes (see models in *Registro, Distrital Acuerdo Número 657* (2016)). Multi-family units under the horizontal property regime and the single-family units owning the lot are estimated in separate models. Estratos 1,2 and 3 and estratos 4,5 and 6 have two independent models. The combination of the two types of properties and the two groups leads to 4 different models. The main comparison in the paper uses blocks with subsidy codes 2 and 3. I separate multi-family units from single-family units to avoid mechanical effects of the statistical model. Once the district cadaster has the appraisals, they estimate the *avalúo catastral*, the assessed value for tax purposes.

Using the statistical model and the census information, the houses without an appraisal get an estimated market price for the property $\hat{p}^{catastro}_{m^2}$. This estimate can be used to calculate the structure value by square meter as the residual of the value of the property minus the lot value.[17]

The land price per square meter $\hat{p}^{land}_{m^2}$ is a block-level variable. To calculate the land price, the officials consider sales, offers, transactions, or leases of real estate, in addition to appraisals made by the district cadaster. It is an input in the appraisal models. Using

[15] Details for the calculation of the estratos are in *Registro, Distrital Acuerdo Número 657* (2016). This document includes an example of the statistical models used in the appraisals, p.3 Annex 2. For Bogota's case, the assessed value represents, on average, 80 percent of the commercial price (González, 2014). By law all the parcels are surveyed on around 60 variables.

[16] The process include officials to pose as potential buyers in order to negotiate a sales price under the premise of a cash payment and professional assessments for at least one property in the more than 16,000 *homogenous zones* (Lozano-Gracia and Anselin (2012), Ruiz and Vallejo (2015),Tsivanidis (2018))

[17] The price is defined as, $P^{structure}_{m^2} = \frac{size^{property}_{m^2} \times \hat{p}^{catastro}_{m^2} - size^{lot}_{m^2} \times \hat{p}^{land}_{m^2}}{size^{property}_{m^2}}$ if Single Family Units and $\frac{size^{unit}_{m^2} \times \hat{p}^{catastro}_{m^2} - size^{lot}_{m^2} \times \hat{p}^{land}_{m^2} \times \frac{size^{unit}_{m^2}}{size^{building}_{m^2}}}{size^{unit}_{m^2}}$ if Multi Family Units

information from the housing market, the cadaster office assigns a value per sq meter of land to each city block.

These appraisals are the base for the assessed value for property tax purposes. The appraisal has a land price component and a structure price component. For properties in subsidy code 1,2, 3, the assessed value, the *avalúo catastral*, corresponds to 70 percent of the land appraisal and 60 percent of the structure. The houses in subsidy code 4 will have an assessment equal to 85 percent of the land appraisal and 65 percent of the structure appraisal. The houses' assessment in subsidy codes 5 and 6 is 85 percent of the land value and 75 percent of the structure.[18] I use the appraisals, and not the assessed value used for property taxes, to estimate the effect of the subsidy scheme on the housing market.

3. Stratificaton Census

To implement the Research Design I need the assignment score. Because the exact score and cutoffs are only available to the department that process the raw data, I use the raw data, and the score's formula, to replicate the assignment score and recover the cutoff points. The data is collected by the District Department of Planning in the Stratification Census.[19] I use the 2009 Stratification Census. This census defines the estratos in 2011 the year I observe the main outcomes.

The composition of estratos in the city has been relatively stable. Of the 36'985 existing in 1997 only 4.2 percent had a change in *estrato* between 1997 and 2009. Only 80 blocks had a change between 2009 and 2017. Appendix table 4 shows the changes in the estratos in these years.

To ensure the accuracy of my replications, I compare the prediction using my formula and the observed estratos for all of the available years of raw data. For 1997, I correctly predict 99.9 percent of all the blocks. For 2009, I correctly classify 99.4 percent of the blocks. Appendix Table 2 shows the percentage of correctly predicted blocks by estrato for all the years with available information. The drop in accuracy for later years is consistent with internal revisions of the assignments. In fact, a technical committee

[18]The full description of the assessment formula can be found the annex 2 of the Acuerdo Número 657

[19]I have complete access to different updates of the stratification census.The district planning office collects and updates the information every 2 to 5 years. The city had seven updates in the last 20 years.The updates were in 1999, 2002 2004 2007, 2009, 2013 2017 and 2019.The form used to collect the information is in the appendix figure 5.

was created to deal with appeals in the estrato assignments. Even with this committee, however, very few blocks are revised and assigned to a different estrato.[20]

4. Score Replication

The model to assign the estratos uses two main inputs; a code that summarizes the urban context of each block – the *habitat zones*, and a composite index score that summarizes the quality of each block.

The *Habitat Zones* are urban areas with a similar set of characteristics (e.g., Residential with low density, Industrial, etc.). The type of roads, the topography, the availability of public services like schools and parks, the current use of the land and type of urban development, among others, define the zones.[21] There are 12 habitat zones, and each Colombian city is sub-divided into one of these zones. For Bogotá in particular, some zones are split in two; *zone* x $(-)$ and *zone* x $(+)$. Thus, the city has a total of 20 different zones, coded from 1 to 20. Each block receives a code based on its urban context characteristics. A lower code indicates a lower quality of urban context. The codes 18-20, are for institutional use or green space. In this paper, I focus the analysis in habitat zones with codes 7-10 corresponding to blocks in estratos 2 and 3. Those Habitat Zones correspond to Industrial Use, Consolidated progressive development (-) and (+), and Commercial (-). The coding of the different blocks did not change in the last 20 years.[22] Land-use regulations are not associated directly with the habitat zones; In Bogotá the only practical use is for the assignment of the estratos. Figure 3a shows the distribution of the zones in the city. Appendix Table 3 describes in detail the characteristics defining each zone.

The quality score summarizes 7 characteristics of the block. i) the type of access road,

[20]For example,Departamento Administrativo de Planeación Distrital (2004) document the changes from 1997 until 2004 (p35-p41). Up to 2004 680 blocks got a reassignment of the estrato. Note that the number of estratos that I cannot correctly predict in 2009,245, is below this number. The figure 8 in the appendix shows the number of requests for reassignment and the number of accepted reassignments. The figure clearly shows how only a low percentages of requests get approved.

[21]In addition to the habitat zones there are two other relevant zoning definitions in Bogotá: The Physical Homogeneous Zone and the Geo economic zones. They are the geographical spaces determined from Physical Homogeneous Zones with similar unit values in terms of their price, according to the conditions of the real estate market.

[22]For example in an study by Econometria (1999) they conclude: "[...] the zoning used to carry out the stratification procedure responded to urban and socio-economic concepts of the environment (habitat zones), evaluated with a high degree of subjectivity and not to a zoning based on cadastral information, such as it was the alternative methodology" note 7 chapter 3 in (Sepulveda, Lopez Camacho, & Gallego, 2014)

ii) the type of sidewalk, iii) the existence of a front yard, iv) the parking type, v) the front of the houses, vi) the type of roof, and vii) predominant materials of the houses in the block. The unit of observation is a residential block-side. Each block-side receives a categorical value for each of the seven variables. For example, the variable predominant materials of the houses in the block has five categories; 1-Precarious materials, 2-basic materials, 3-unpainted low-quality bricks, 4-painted low-quality bricks, and 5-polished bricks or veneer.

To create the quality score, a log-rank transformation known as a *savage score* converts the categorical variables into a continuous index. The savage score has four main steps. First, construct a single value for each variable and each block. The value is the average of the categorical values of the different block's sides (normally a block has 4 side, but there are some exceptions). For example, if a block has precarious materials (code 1) in two sides,and basic materials (code 2) on the other two sides, the block will have a value of 1.5 for the variable predominant materials of a block. Second, rank the blocks using their value on each variable. r_i^k is the rank of block i and $k = \{1, ..., 7\}$ is an index for each of the 7 variables. Third, use the ranks r_i^k each variable k receives a savage score h_i^k.

$$h_i^k = \left(\sum_{j=N-r_i+1}^{N} \frac{1}{j} \right) - 1.$$

Finally, the 7 h_i^k savage scores are added to create a quality score, $score_i$ for each block. The score is a sum of the savage score for each block, $score_i = \sum_{k=1}^{7} h_i^k$.[23] Figure 3b represents the distribution of the score in Bogotá.

[23]I use the formula described in SDP (2004) p.65

a. Habitat Zones *b. Neighborhood Quality Score*

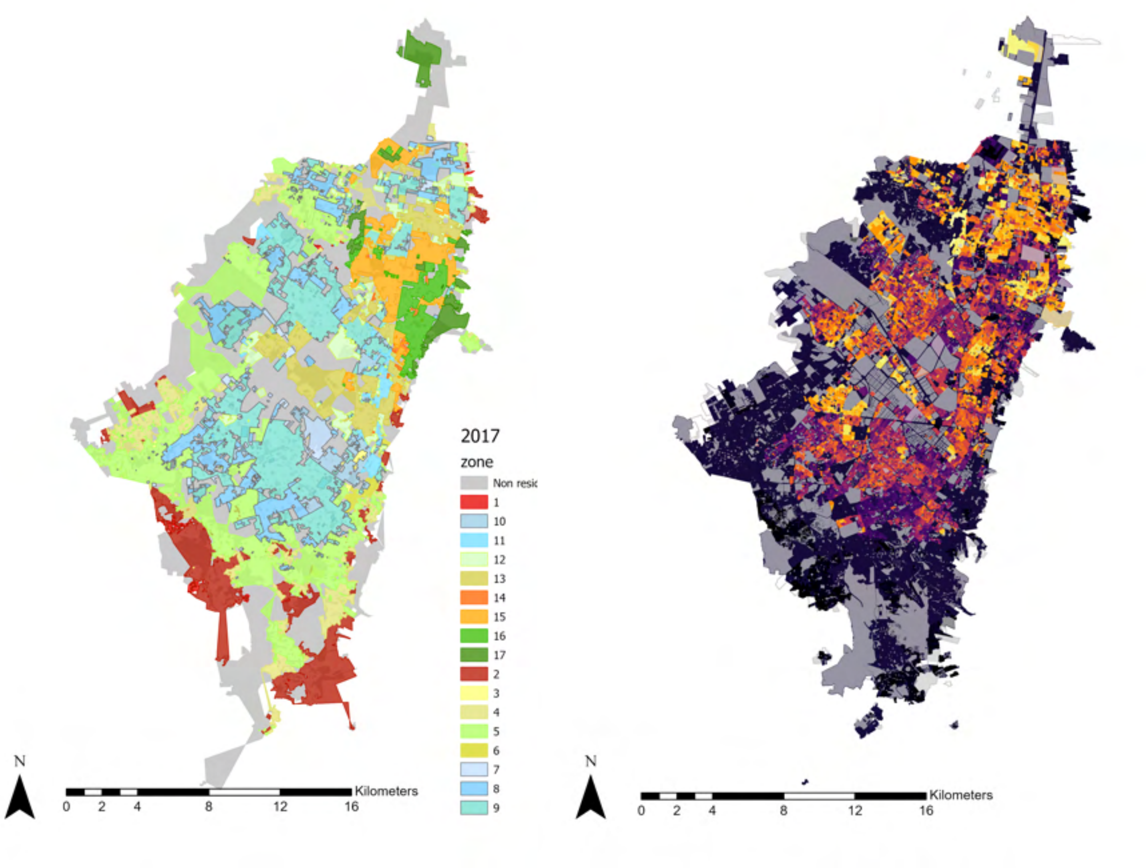


Figure 3: Distribution of Main Assignment Variables in Bogotá

Note: Panel A shows the distribution of the habitat zones in the city. Panel B shows the distribution of the quality score in the city. The light colors represent better quality areas and the dark colors lower quality areas.

The habitat zones and quality score are the two main inputs to define the estratos. Once the score is calculated, it is combined with the habitat zones. Neighborhood blocks are divided among the 6 estratos. The cutoffs are defined by the methodology Dalenius-Hodges, which tries to maximize the variance among groups and minimize the variance within groups.[24] The National Planning Department receives the raw data and processes the information. They construct the quality score, and assign an estrato to each block. They provide the estratos without making the composite index or the cutoffs public.

[24]Bogotá uses an adaptation called Bi-variate Dalenius-Hodges. This adaptation was implemented to treat the habitat zone independently and not as an additional value in the quality score.

The public service providers receive the information and use the codes to charge the differential price rate.

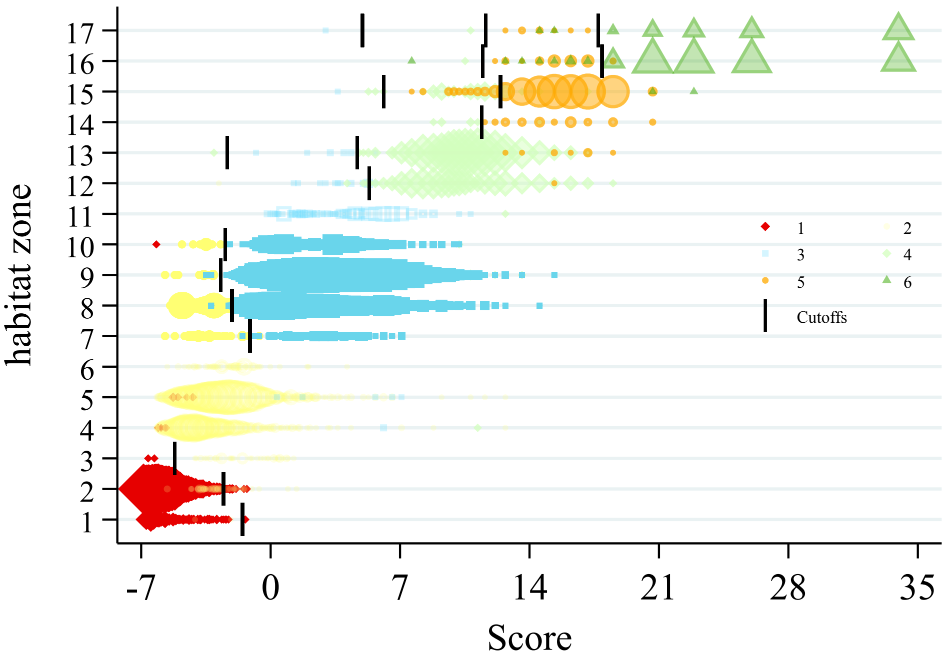


Figure 4: Subsidy Codes Based on Habitat Zones and the Score

Note: The figure presents the distribution of estratos by habitat zone and block quality score. A different color represents each estrato. The darker blue and yellow are the focus of the analysis.

I combine the score with the *habitat zones* to infer the location of the cutoffs. To define the cutoffs, I try to emulate the matrix employed in the bi-variate Dalenius-Hodge methodology. Within each Habitat Zones and the quality score, I define the cutoff as the value that maximizes the probability that I correctly predict the observed distribution of estratos. The figure 4 shows the scores (x-axis) and the *Habitat Zones* (y-axis) for 2009. The six different colors represent the six *estratos*. The black solid lines are the cutoff that I assign. The symbols are weighted to represent the number of blocks with a particular score and zone. I highlight the areas that are the main focus of the analysis (i.e. habitat zones 7-10). The blue squares represent blocks in *estrato* 3, and the yellow circles represent blocks in *estrato* 2. The figure shows that only a few habitat zones have more than one estrato. I use those zones in my identification strategy. The distribution of scores and estratos are similar in the other years with available data (see appendix figures 7).

5. Baseline Characteristics and Constant Geographic Unit

Some of the blocks change the geographic administrative codes over time. This makes it difficult to track a particular location at different points in time. To do that, I identify which codes correspond to the 2017 administrative codes based on the location and assign that code to all of the years based in a particular location over time. Additionally, there are two different administrative codes, one used by the district authorities and the other one used by the National Department of Statistics (DANE). To check for balance in preexisting characteristics I use data from the 1993 population Census. I am able to identify the exact block for each observation. However, the code for the block is not the same used in the stratification census. To create a crosswalk between the two blocks I use the location of based on a shapefile and assign the 2017 administrative codes based on the location. Like I did to create the constant geographic unit over time. For more details see appendix. 5..1

6. Main Sample of analysis

The main analysis of this paper, focuses on estratos 2 and 3. There are several reasons. First, these two groups represent an important share of the residential properties in the city. In 2011 Bogotá had 559,328 residential properties in estrato 2 and 654,136 in estrato 3, corresponding to 61 percent of all the residential properties in the city. Second, the subsidy for houses in estrato 2 and 3 represents an important share of household income and housing prices (see figure 2a and Appendix Figure 1b and 1a). Third, the appraisals for those two codes are calculated with the same statistical model. I also avoid getting effects mainly explain by the difference in appraisal methods. This will be the case if I compare *estratos* 3 and 4.[25] Fourth, individuals may want to avoid the stigma associated with receiving the subsidy, the fact that the two groups receive a subsidy reduces the role of this mechanism.

The quasi-experimental variation I exploit in my research design is at the block level.Therefore, I create block-level variables and perform the analysis at the block level.[26] I restrict the analysis to residential properties because the subsidy scheme does not apply to non-residential properties. I restrict the analysis to residential properties because the subsidy

[25]The officials use different model for estratos 1,2,3 and 4,5,6 and for single and multifamily homes. In section 3 I explain more in detail.

[26]The results are very similar when I use individual level variables but to be more transparent about the source of variation I am using, and to have more conservative SE, I keep the block level specification as my main specification.

scheme does not apply to non-residential properties.

IV. Research Design

Studying the incidence of a location-based subsidy is challenging for at least two reasons. The targeted areas are particular, and the policy can induce sorting. Usually, the targeted locations are under-performing areas; they are poorer, have more crime, or are more polluted. Thus, a counterfactual that allows studying the incidence and effectiveness of the policy is hard to find. In contrast with other settings, randomly assigning the treatment to some areas is not easy, and is rarely implemented. Second, the policies can encourage people to move to take advantage of the subsidy. For example, people in the US decide where to live based on the school quality of each town (Bayer et al. (2007) and Black (1999)). A careful evaluation of location-based policies has to consider this sorting and the potential effect on the housing market. A comprehensive assessment of these subsidies has to account for any impact of the policies on the housing market.[27]

Bogotá's particular targeting tool, the estratos, allows me to address these challenges using a quasi-experimental research design. As explained in section 3.4, Bogotá uses a uses a score summarizing the quality of the block and the urban context characteristics to assign the estratos. These two characteristics and arbitrary cutoffs separate the city into the six estratos. I use these cutoffs to apply a Regression Discontinuity Design and evaluate the effect of the subsidy scheme on the housing market. In this section, I explain how I use the features of the assignment in my Research Design and my estimation approach.

Figure 4 shows the cutoffs that define the estratos of different blocks within a habitat zone. The figure shows that there is not a single discontinuity, and a standard uni-dimensional RDD does not directly apply. There are different estratos and different cutoffs. In this paper, I am focusing on comparing the houses in estratos 2 and 3. I can reduce the discontinuities in the habitat zones 7-10 to a single discontinuity. Choi and Lee (2018b) among others show that the standard RDD will work under the assumption that the effect is the same in all the discontinuities.[28] Figure 5 shows the discontinuities at the cutoff for the particular areas where I focus my analysis (i.e habitat zones 7-10

[27] for a detailed discussion of the difficulties of evaluating location base policies see Kline and Moretti (2014)

[28] Choi and Lee (2018a) Reardon and Joseph P. (2012) Wong, Steiner, and Cook (2013) Choi and Lee (2018b) Papay, Willett, and Murnane (2011)

which are highlighted in figure 4). I normalize the score in order to have the cutoff at 0 in each habitat zone. This figure shows a clear discontinuity in the probability of being treated at the cutoff. I use this discrete jump in the subsidy level to identify the effect of the subsidy on the housing market.

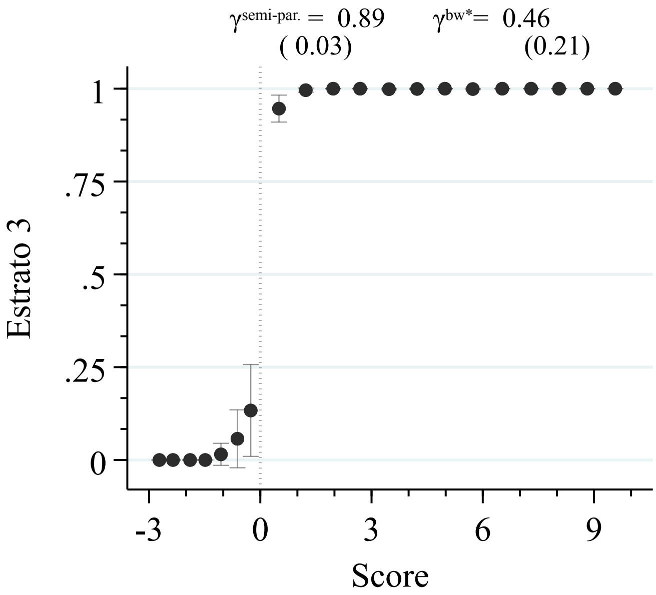


Figure 5: First Stage (habitat zones 7-10)

NOTE: This Figure represents the discontinuity in probability of being in estrato 3 introduced by the arbitrary cutoffs. The dots are local average for equally spaced bins. The number of bins minimizes the integrated mean square error (IMSE) on each side. The point estimates from 3 different estimation methods are at the top. γ^{parm} and $\gamma^{semipar}$ use different approaches to estimate $h\left(S_{izt}\right)$ in the model: $\mathbb{1}_{[S_i=3]} = \alpha + \gamma_{2,3}\mathbb{1}_{\left[S_{izt}\geq\delta_2^3\right]} + h\left(S_{izt}\right) + \varepsilon_{izt}$.$\gamma^{semipar}$ estimates $h(S_{izt})$ non parametrically using a partially linear model (Robinson, 1988), and γ^{parm} uses a parametric approximation using a polynomial of degree 1. $\gamma^{bw^*} = \lim_{s_{izt}\to\delta^{3+}_2} \mathrm{E}\left(y_{izt}|S_{izt}=s_{izt}\right) - \lim_{s_{izt}\to\delta^{3,2-}} \mathrm{E}\left(y_{izt}|S_{izt}=s_{izt}\right)$. In this approach I use the method proposed by Cattaneo, Idrobo, and Titiunik (2018) to select the optimal bandwidth. The numbers in parenthesis are robust standard errors. The figure contains all the blocks in estratos 2 and 3 in habitat zones 7,8,9 and 10. I drop the observations with a score lower than -3 or higher than 10. I normalize the score to have the cutoff at 0 for all the habitat zones. .

The arbitrary cutoffs and the associated discontinuities in the subsidy levels allow me to identify the causal effect of the subsidy. I analyze how different outcomes Y_i, related to the housing market are affected by the subsidy scheme. The unit of study is a block i. Each block has two potential outcomes $Y_i\left(\mathbb{1}_{[S_i=2]}\right)$ and $Y_i\left(\mathbb{1}_{[S_i=3]}\right)$. I am interested in studying the difference between the potential outcomes under treatment, belonging to estrato 3, $\mathbb{1}_{[S_i=3]}$, and control belonging to estrato 2, $\mathbb{1}_{[S_i=2]}$. The parameter of interest,

$\theta_i^{2,3}$ is therefore the difference between these two potential outcomes.

$$\theta_i^{2,3} = Y_i\left(\mathbb{1}_{[S=3]}\right) - Y_i\left(\mathbb{1}_{[S=2]}\right) \tag{1}$$

To estimate the causal effect of the subsidy scheme, I use the discontinuous shift in the subsidy level at an arbitrary cutoff in the block quality score. The idea behind this approach is that the only discrete change at the cutoff is the estrato level. Comparing blocks to the left and the right of the cutoff will allow me to identify $\theta_{(2,3)}$ under the assumption that the blocks on the left of the cutoff are a good counterfactual to what would had happen to the blocks on the right if they received a higher subsidy. The effect of the subsidy on the housing market will be identified by the discontinuous mean shift at the cutoff, if two assumptions are satisfied. First, when the estratos were assigned 1997, all the observable characteristics not involved in the assignement of the estratos and the unobservable characteristics should be continuous at cutoff. Second, the distribution of blocks around the cutoff should be the same at both sides of the cutoff, i.e., there is not manipulation around the cutoff. I provide evidence that these two assumptions are credible. First, in Figure 6, I present observable characteristics before the estratos were assigned. Second, In Figure 7, I show the distribution of the score for 2009 and a McCrary Manipulation test.

Figure 6 A and figure 6 B shows characteristics of the households (number of kids, number of people in the house, type of employment and education level) and housing characteristics (House with brick walls,Building,House with cement floor, number of rooms) by block quality score in 1993. I use this information to compare the blocks before the assignment of the policy. The baseline characteristics of the blocks seem to be balanced. With the exception of the number of rooms, there is not evidence of a discontinuous jump around the cutoff. To summarize the information, figure 6 C) shows the prediction on the main outcomes variables (i.e. age, and probability of building built after 1997, and prices) using the individual and housing characteristics in 1993. There is a small apparent discontinuity that is minor compared to the mean shift I will show in the next section. The small discontinuity in prices goes is the opposite of the discontinuity I observe in the prices in 2011. In addition to this exercise, In my empirical analysis I include this baseline characteristics and the results are unchanged. The fact that the variables are balanced in the period before the assignment of the estratos, gives me confidence on my research design.

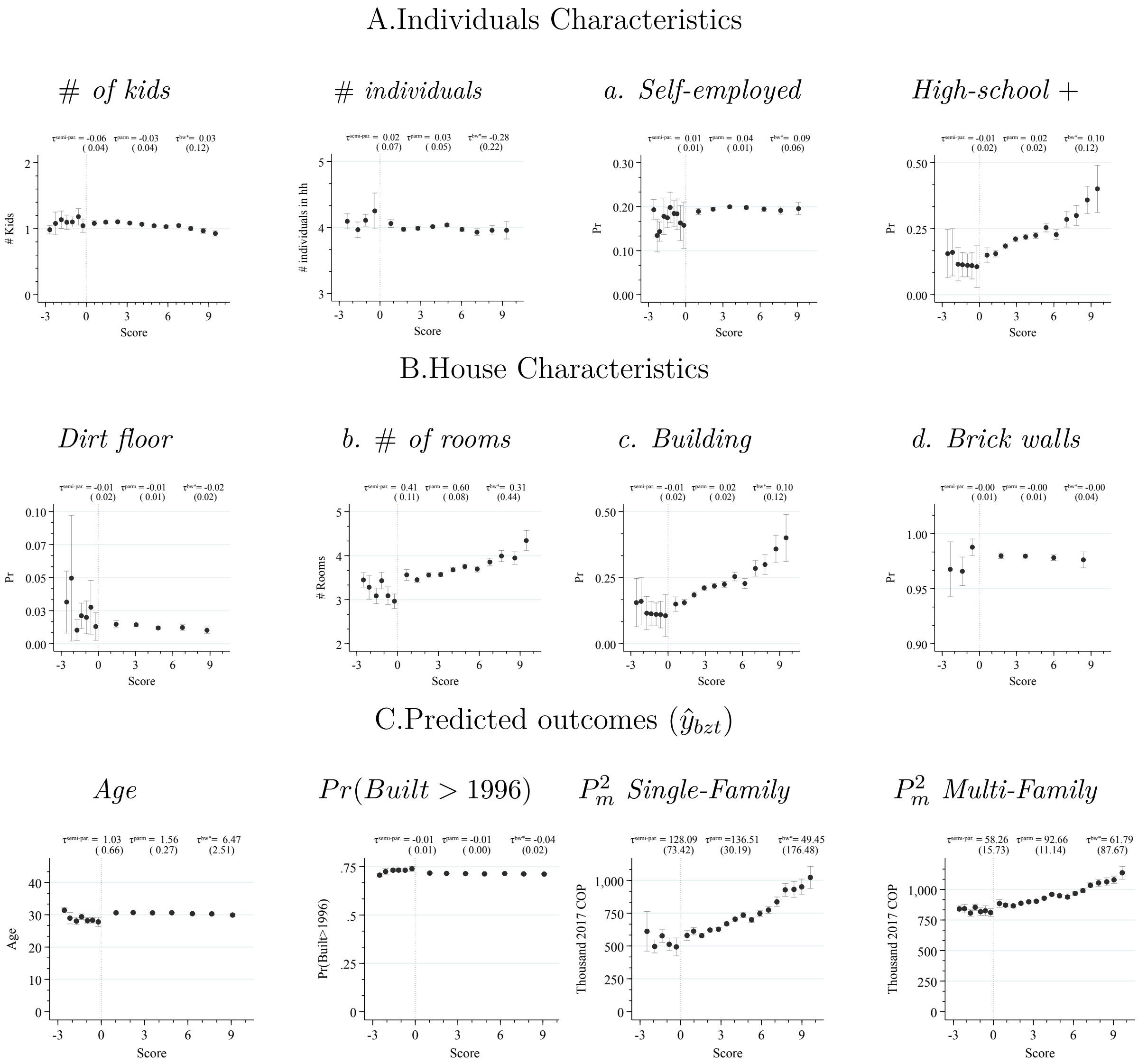


Figure 6: 1993 Balance

Note: The independent variables, $\hat{y}_{izt}$ is the prediction of an outcome y_{izt} using 1993 demographic and housing characteristics (i.e. $y_{izt} = x_{izt}^{93}\beta + \varepsilon_{izt}$ and $\hat{y}_{izt} = x_{izt}^{93}\prime\hat{\beta}$. The 1993 demographic characteristics are the number of individuals, the number of kids in a house, the educational level, and the type of education. The house characteristics are the type of floor, walls, the number of rooms, and the type of house (i.e., buildings, houses). The running variables uses the 1997 score.

Figure 7 shows the distribution of the block quality score (panel A) and the McCrary Manipulation test (panel B). The distribution blocks by the blocks quality does not have any apparent bunching at the cutoff. Given that the unit of observation is the block and the exact cutoff follows an objective methodology to classify the blocks, this is unsurprising. However, the cutoff is located at a point right before the slope's increase in the bell shape. This is consistent with the fact that the cutoff minimize the variance within groups and maximize the variance among groups (Delano-Hodges methodology). Because this results from an algorithm following and objective method, this should not be a major concern. The McCrary Manipulation test confirms that there is not manipulation around the cutoff. I cannot reject the null hypothesis that the distributions on both sides of the cutoff are the same (p-value is 0.55).

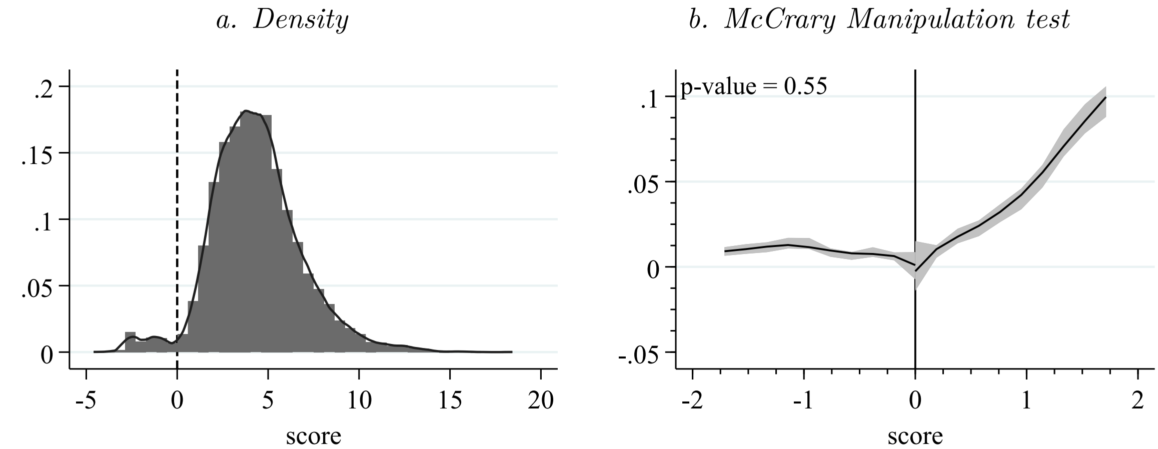


Figure 7: Score Distribution

Note: This Figure shows the distribution of the score for 2009. (left panel) and the McCrary Manipulation test (right panel). H_0: The density is the same on the right and on the left. The p-value of the test is on the top of the right panel. I cannot reject the null hypothesis. To implement the manipulation test I follow Cattaneo, Jansson, and Ma (2018).

1. Estimation

To estimate the parameter of interest I estimate the following model.

$$Y_{izt} = \alpha + \theta_{2,3}\mathbb{1}_{[S_{i}zt=3]} + k\left(S_{izt}\right) + \Gamma' X_{izt} + \varepsilon_{izt} \tag{2}$$

Where Y_{izt} is an outcome of interest for block i in habitat zone z in year t. S_{izt} is the score assigned to each block, $\mathbb{1}_{[S_{it}=3]}$ is an indicator variable equal to one if the block belongs to estrato 3 and 0 otherwise, $k\left(S_{izt}\right)$ is a control function that captures the relationship between the block quality score used in the assignment and the outcome of interest Y_{izt}. X_{izt} is a vector of observable characteristics, and ε_{bzt} are the unobserved characteristics affecting Y_{izt}. The parameter of interest is $\theta_{2,3}$. The main challenge to estimate the effect of the subsidy scheme in the housing market is to recover the control function

$k(S_{izt})$. I estimate the control function $k(S_{izt})$ non para-metrically using a partially linear model as in Robinson (1988). In this approach I use all the available information and I assuming that $k(S_{izt})$ is a smooth functions that do not change at the cutoff. I prefer this approach to an approach like Cattaneo, Titiunik, and Vazquez-Bare (2019) where I just use data only around the cutoff. Cattaneo et al. (2019) choose an optimal bandwith to estimate the effect using only observations around the cutoff. This approach, may be more accurate in other settings with more observations around the cutoff. In my case I do not have many observations and this lead to the estimation using the optimal bandwidth to include in some cases only 12 blocks. However, for completeness, I present estimates using both estimation approaches. As a robustness check I also estimate the control functions parametrically using a polynomial on S_{izt}.

The discontinuity, I am exploiting is not a sharp discontinuity, as Figure 5 shows. Thus, I apply the standard fuzzy RD framework. I use the assignment rule as an instrument for the treatment. γ_2^3 is the cutoff for blocks 2 and 3 and $\mathbb{1}_{\left[S_{izt}\geq\gamma_2^3\right]}$ is an indicator variable equal to one if the score is higher than γ_2^3 and zero otherwise.

$$\textit{First Stage}\text{: } \mathbb{1}_{[S_i=3]} = \alpha + \delta_{2,3}\mathbb{1}_{\left[S_{izt}\geq\gamma_2^3\right]} + h\left(S_{izt}\right) + \epsilon_{izt}$$

$$\textit{Reduced Form}\text{:} Y_{izt} = \beta + \tau_{2,3}\mathbb{1}_{\left[S_{izt}\geq\gamma_2^3\right]} + g\left(S_{izt}\right) + \varepsilon_{izt}$$

$$\textit{The parameter of interest}\text{: } \theta_{2,3} = \frac{\tau_{2,3}}{\delta_{2,3}}$$

I use a partially linear model Robinson (1988) to estimate $g\left(s_{bzt}\right)$ non para-metrically and $\mathbb{1}_{\left[s_{bzt}\geq\gamma_2^3\right]}$.[29] I complement the analysis using non parametric approach following Cattaneo, Idrobo, and Titiunik (2018). I use local polynomial and I calculate the optimal bandwidth for both sides of the cutoff. I adjust the weights using a triangular kernel, giving more weight to the observations close to the cutoff. To get an estimate of $\theta_{2,3}$ I get not parametric approximation to the expected value of the outcome conditional on the score S_{izt} on the right and on the left of the cutoff, $g\left(S_{izt}\right) = g^1\left(S_{izt}|S_{izt} < \gamma_2^3\right) + g^2\left(S_{izt}|S_{izt} > \gamma_2^3\right)$

[29]The parametric approach use in the robustness analysis I use a polynomial of degree three to approximate $g\left(S_{izt}\right)$. $g\left(S_{izt}\right) = S_{izt} + S_{izt}^2 + S_{izt}^3$.I fix the shape of the polynomial to be the same on both sides. The estimate for the reduced form γ_2^3, is the mean shift around the cutoff.

$$\theta_{2,3} = \frac{\lim_{s_{izt} \to \gamma_2^{3+}} \mathrm{E}\left(Y_{izt}|S_{izt} = s_{izt}\right) - \lim_{s_{izt} \to \gamma_2^{3-}} \mathrm{E}\left(Y_{izt}|S_{izt} = s_{izt}\right)}{\lim_{s_{izt} \to \gamma_2^{3+}} \mathrm{E}\left(\mathbb{1}_{\left[S_{izt} \geq \tau_2^3\right]}|S_{izt} = s_{izt}\right) - \lim_{s_{izt} \to \gamma_2^{3-}} \mathrm{E}\left(\mathbb{1}_{\left[S_{izt} \geq \tau_2^3\right]}|S_{izt} = s_{izt}\right)} \tag{3}$$

V. Results

Figure 8 shows the share of buildings in estratos 2 and 3 built before and after the new stratification system was introduced in 1997. The x axis shows the score in 2011 and the y axis shows the percentage of blocks. The data reflects the housing stock in 2011. This figure suggests that after 1997, more construction was concentrated below the cutoff dividing estratos 2 and 3. The figure shows the cumulative share of housing stock by construction year. The two dashed lines represent all the blocks in estratos 2 and 3. The solid lines represent the blocks near the cutoff. When I compare all the blocks in estrato 2 and estrato 3, there is no apparent increase in units built after 1997. When I consider only the blocks around the cutoff, the subsidized blocks seem to have a higher concentration of new construction. More than 50 percent of the properties in estrato 2 with neighborhood quality scores close to the cutoff were built after 1997. This is suggestive evidence that developers and landowners built disproportionately more units in the blocks in estrato 2 comparable in terms of quality to the blocks in estrato 3. Therefore the subsidies seemed to induce new construction.

To test this hypothesis further, I use the research design to test whether the subsidy scheme affects the structure's age and the share of units in each block built after 1997.

a. PDF of units by score and build timing *b. CDF by construction date*

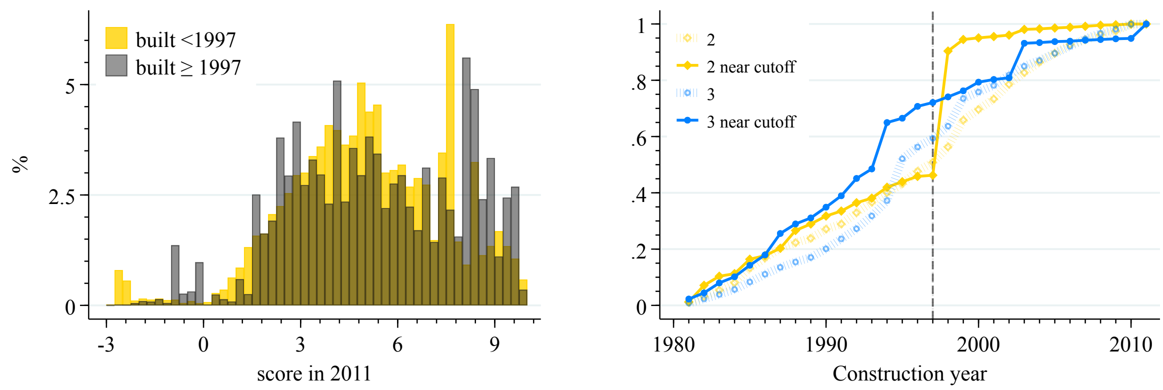


Figure 8: Stock of buildings in 2011 by construction date

1. Property Characteristics

The top graphs of Figure 9 show bin scatter plots for the variables related to the housing characteristics. Figure 9a shows the result for the age of the housing units, Figure 9b, the probability of having new construction, Figure 9c, the quality of units, and Figure 9d the average size of each unit. Following Cattaneo, Idrobo, and Titiunik (2018), I select equally spaced bins that minimize the integrated mean square error (IMSE) to the left and right of the cutoff. The bin scatters allow to visually explore any discontinuous shift at the cutoff without imposing structure on the data. The dashed line illustrate the identification approach where I impose some structure on the data. The dashed line in each figure shows the control functions $k(s_{izt})$. The idea of this identification approach is that in the absence of the subsidy scheme, the outcomes will follow $k(s_{izt})$. Therefore, the counterfactual to a world without subsidies is described by the control function. Thus the subsidy scheme's effect is the difference between the bins, and the control functions estimated with the model in equation 2.

The alternative is to estimate the effect using only the observations close to the cutoff. The idea is that the effect is the difference between the projection of the bin on the left and the right of the cutoff. The estimation uses the data around each dot to get a projection that intersects the cutoff. The point estimates for the two different approaches are at the top of the figure. These estimates are equivalent to the reduced form in a fuzzy RD design.

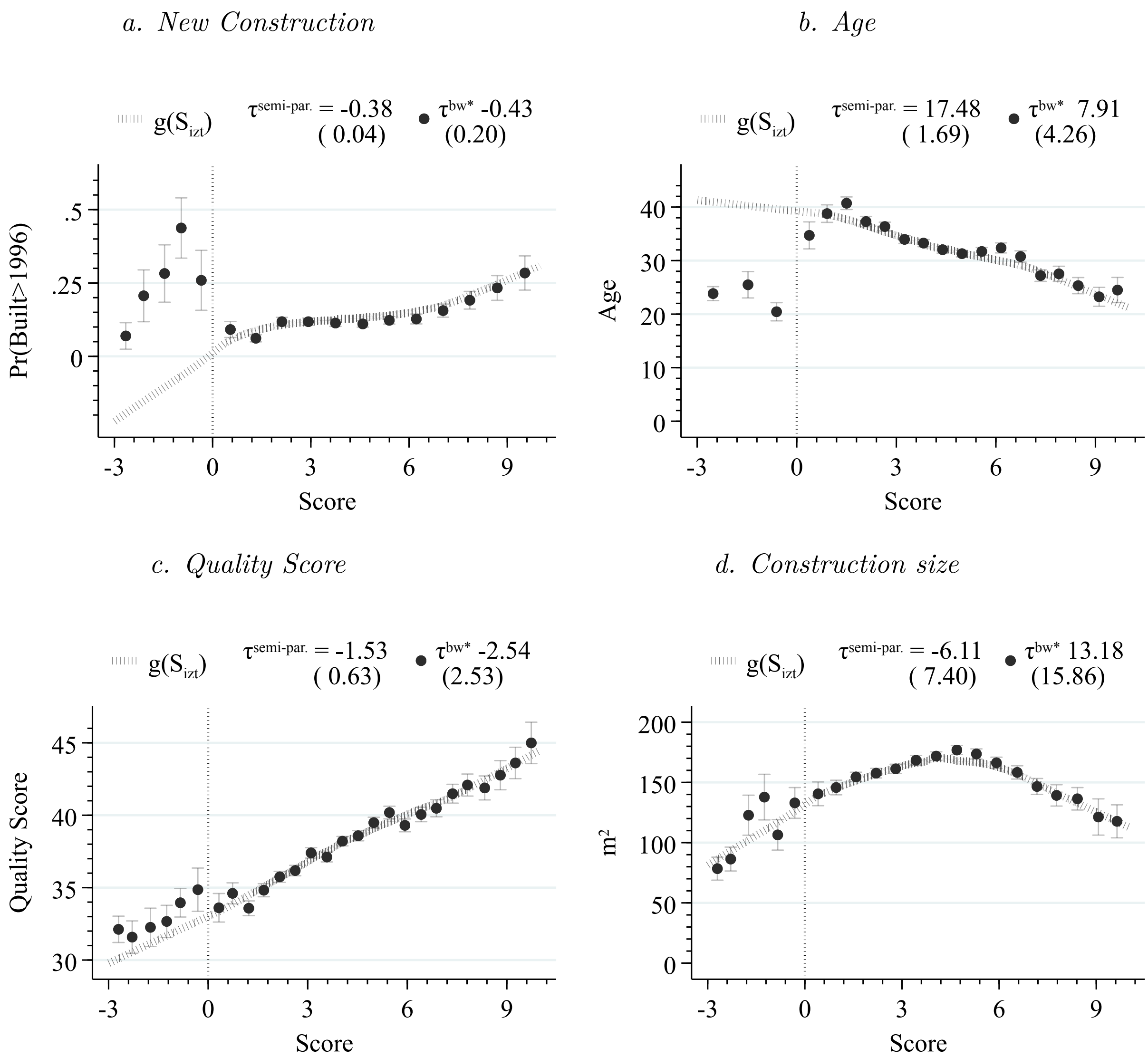


Figure 9: PROPERTY CHARACTERISTICS

NOTE: This figure represents the reduced form. The dots are local average for equally spaced bins. The number of bins minimizes the integrated mean square error (IMSE) on each side. The point estimates from 3 different estimation methods are at the top. τ^{parm} and $\tau^{semipar}$ use different approaches to estimate $g\left(S_{izt}\right)$ in the model: $Y_{izt} = \alpha + \tau_{2,3}\mathbb{1}_{\left[S_{izt} \geq \delta_2^3\right]} + g\left(S_{izt}\right) + \varepsilon_{izt}$. $\tau^{semipar}$ estimates $g(S_{izt})$ non parametrically using a partially linear model (Robinson, 1988), and τ^{parm} uses a parametric approximation using a polynomial of degree 3. $\tau^{bw^*} = \lim_{s_{izt} \to \delta_2^{3+}} \mathrm{E}\left(y_{izt} | S_{izt} = s_{izt}\right) - \lim_{s_{izt} \to \delta_2^{3-}} \mathrm{E}\left(y_{izt} | S_{izt} = s_{izt}\right)$. In this approach I use the method proposed by Cattaneo, Idrobo, and Titiunik (2018) to select the optimal bandwidth. The numbers in parenthesis are robust standard errors. The figure contains all the blocks in estratos 2 and 3 in habitat zones 7,8,9 and 10. I drop the observations with a score lower than -3 or higher than 10.

Figures 9a and 9b show a clear mean shift at the cutoff, suggesting that the subsidy scheme affects how the housing market evolves in Bogotá. The blocks in the heavily subsidized areas (i.e., estrato 2) are newer and have a higher probability of being built after the assignment methodology for the estratos started in 1997. There is a smaller and less obvious jump at the cutoff for housing quality. Depending on the estimation method, the point estimate for the reduced form mean shift on average age of the units is between 7.9 and 20.8 years, and the mean shift on new construction is between 0.21 and 0.43 percent. The point estimate for the mean shift in the quality score is between -1.5 and -2.5, depending on the estimation method. For house size, there is no clear visual discontinuity.

The evidence presented to this point shows the effect of the assignment rule and not of being part of a different estrato and therefore paying different prices for utilities. To estimate the effect of the estratos in the housing market, I use the assignment rule as an instrument for being in estrato 3. The effect of the estrato is given by the reduced form estimates presented in figure 9 divided by the first stage presented in figure 5. The estimates are in table 3. I estimate the effects with the two different estimation methods, and I include some additional controls in columns 2, 3, and 5,6, respectively. In columns 2 and 5, I control by the habitat zone, the share of single-family units in a block, and an indicator variable equal to 1 if the block had a change in the estrato since the new stratification method started in 1997. In Columns 3 and 6, I control for the characteristics of each block in 1993. Each row presents the results for a different variable. For the partially linear model, I estimate the standard errors using bootstrap (100 repetitions). For the other non parametric approach, I use the 2sls robust standard errors.

The results are consistent with the visual description presented in the figures. Heavily subsidized blocks observe higher construction levels, which is reflected in a lower average age of its units. This conclusion does not depend on the inclusion of different controls or the estimation approach. The coefficients change very little when I include controls, and the point estimates are similar across all the different estimation approaches. Because the non-parametric estimates only consider information close to the cutoff, the estimates are less precise, and the standard errors are bigger. If the control function describing the relationship between the block quality score and the outcomes of interest is a valid counterfactual, considering information farther away from the cutoff is reasonable, it allows for more precise estimates. In the figures, the difference between the non-parametric coefficients and the other models was higher than in the table, where the point estimates are very similar across specifications.

Being in estrato 2 and getting a higher subsidy causes a decrease in the block's units' average age. The units in estrato 2 are between 15.80 and 22.08 years younger because they receive a higher subsidy. The blocks in estrato 3 have a 43 percent lower probability of having properties built after the introduction of the new stratification regime. The non-parametric estimation, the quality effect is not distinguishable from 0. For the partially linear model, the effect fluctuates between -1.72 and -2.04. The results for the size of the property suggest that there is not a difference in the size of properties caused by the subsidy scheme. When I include preexisting characteristics as additional controls, the coefficients are statistically different from 0 but economically small.

These results suggest that individuals and developers are responding to the subsidy by building new units in the areas where it is cheaper to live. There is some evidence that individuals may increase the quality of the units or build better quality units. This is consistent with individuals having a negative elasticity of substitution between consumption of utilities and housing quality; individuals that save in utilities consume better quality housing. Also, the new construction induced by the subsidy may be of better quality. There is an apparent effect of an increase in property size,which suggests that the new units are a bit smaller. These results suggest that the way the city, particularly the housing market, evolves is affected by the subsidy scheme. If the locations with higher subsidies are worse, this could be a sub-optimal equilibrium; the city growth is skewed towards its lower quality areas and worst neighborhoods.

Table 1: THE EFFECT ON CONSTRUCTION AND HOUSING CHARACTERISTICS.

	Semipar			Non-parametric		
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(Built>1996)	-0.43***	-0.43***	-0.43***	-0.39+	-0.61*	-0.54*
	(0.03)	(0.09)	(0.01)	(0.24)	(0.29)	(0.21)
Age	19.64***	22.08***	22.08***	15.80***	18.16**	10.76***
	(0.27)	(0.71)	(1.44)	(4.73)	(6.83)	(3.11)
Quality Score	-1.72***	-2.04***	-2.04***	-3.94+	-4.97	-3.94*
	(0.37)	(0.22)	(0.33)	(2.18)	(3.78)	(1.98)
Size m^2	-6.87	10.27	10.27***	16.56	45.44	29.55*
	(5.97)	(8.29)	(0.96)	(18.85)	(40.35)	(14.30)
Controls	No	Yes	Yes	No	Yes	Yes
Char. 1993	No	No	Yes	No	No	Yes

NOTE: This table presents the estimates θ using three different estimation methods with and without controls.The estimates are the IV estimates using the assignment rule as an instrument for a block belonging to estrato 3. Each row is the estimate for a different variable. Columns 1 to 4 use different approaches to estimate $k(S_{izt})$ in the model $y_{izt} = \alpha + \theta_{2,3}\mathbb{1}_{[S_{izt}\geq\delta_2^3]} + k(S_{izt}) + \beta X + \varepsilon_{izt}$. Columns 1 and 2 use a partially linear model Robinson (1988) to estimate $k(S_{izt})$. Columns 5 and 6 estimate θ using non parametric approach $\theta = \lim_{s_{izt}\to\gamma_2^{3+}} \lim \mathrm{E}(y_{izt}|S_{izt} = s_{izt}) - \lim_{s_{izt}\to\gamma_2^{3-}} \mathrm{E}(y_{izt}|S_{izt} = s_{izt})$. In this approach I use the method proposed by Cattaneo, Idrobo, and Titiunik (2018) to select the optimal bandwidth.The first stage and reduced form are represented in figures 5 and 9. The controls included are a dummy equal to 1 if the block changed the estrato since 1997, habitat zones, and the share of blocks that are multi family units. The numbers in parenthesis are robust standard errors. The figure contains all the blocks in estratos 2 and 3 in habitat zones 7,8,9 and 10. I drop the observations with a score lower than -3 or higher than 10.

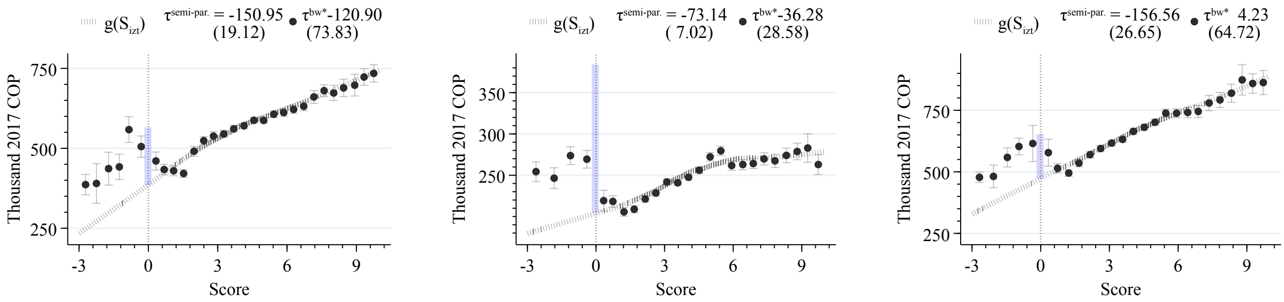
Significance level $^{+}10\%,^{*}5\%,^{**}1\%^{***}0.1\%$

2. Property Values

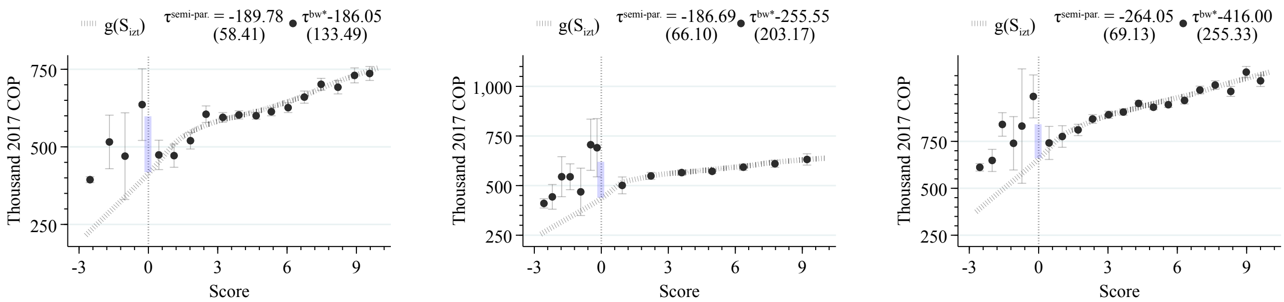
Does the subsidy scheme affect housing prices? Or does it only affect the housing market through new construction? This subsection investigates the impact on prices. As explained in section III., the appraisals use different models for multi-family and single-family units. The appraisals also separately estimate land and structure prices.

A. Single Family Units

(1) Land *(2) Structure* *(3) Property*



*B.*Multi Family Units



*C.*All Units

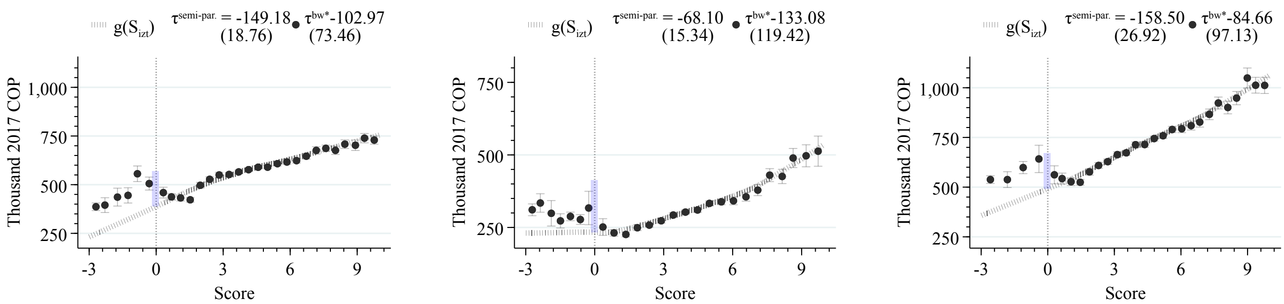


Figure 10: Housing Prices

Based on standard urban theory, the savings in utilities and property taxes should be offset by higher housing prices. All individuals have constant utility in a spatial equilibrium, and housing prices adjust to generate that equilibrium. Everything else equal, people would prefer to live in locations with higher subsidies. This higher demand to live in subsidized areas will cause housing prices to increase until everybody has the same utility, and nobody has incentives to move. This section tests this key assumption of spatial equilibrium models. My empirical setting provides a unique opportunity to perform such a test.

To set a capitalization benchmark, I take the net present value of the monthly expenditure on property taxes and utilities in each estrato. The difference between estratos 2 and 3 is 180 Thousand COP/m^2 per square meter and is the benchmark for a full capitalization in my empirical exercise. This is the number presented in section 2, divided by the average size of all the houses in estrato 3 and 2 (100 m^2).

In contrast with the case of housing characteristics, for housing prices, I only focus on blocks in habitat zones 8 and 9. To pool the four habitat zones into a single discontinuity, the assumption of constant treatment across habitat zone has to hold. For housing prices in habitat zones 7 and 10, this is not a reasonable assumption. Those zones have only few blocks with multi-family units, and housing prices are not comparable to those in habitat zones 8 and 9.[30]

Figure 10 shows the figures for the different measures of housing prices and the different types of units. Panels A and B show the results for multi and single-family units. Column 1 shows the land prices, column 2 structure prices, and column 3 overall property prices. All the prices are in thousands of 2017 COP per square meter. For completeness, the panel 3 presents the results for all the units. The point estimates for the different estimation methods are at the top of each figure. As with figure 9, I include the control function as dashed line. In this case, the control functions represent a price gradient in the absence of the subsidy scheme.

The figures show a clear discontinuous mean shift for the multi-family units. The land prices show a mean shift between 158.6 Thousand COP/m2 and 189.7 Thousand COP/m2 depending on the estimation method. The single-family units have a less clear mean shift at the cutoff when we look at the bin scatters. Yet, the graphs imposing more

[30]The results on housing characteristics are similar when I use only habitat zone 8 and 9 (See Appendix table 1). I decided to present the results with all the habitat zones in the body of the paper to have more observations and get more precise estimates.

structure in figure 10 suggest a mean shift at the cutoff for single families. In each of the graphs in figure 10, I add a line with the size of the benchmark capitalization. The difference between the bins and the control functions represents the effect of the subsidy under my preferred estimation approach. The figure shows an apparent effect that is close to the full capitalization benchmark. The non-parametric approach estimates are noisy, but except for total property value for single-family units (Column C panel 2), the estimates have the same sign. Their confidence intervals include the estimates from the other approaches. Figure 10 shows that the slope of the land gradient is steeper than the slope of the structure for both the multi and single-family units. The property structure gradient is relatively flat, suggesting that the structure's price does not vary much from the neighborhood block quality score.

Table 3 shows the estimates of the subsidy scheme's effect on the housing market when I instrument being in estrato 3 with the assignment rule. Columns 1 and 5 present the results for the different estimation methods without including controls. Columns 2 and 5 include basic controls, habitat zone fixed effects, the share of multifamily units, and an indicator variable equal to one if the block changed the subsidy code since implementing the stratification method 1997. Each row presents the point estimate and standard error for land, structure, and total property value per square meter. I include the p-value of a test for a full capitalization of the subsidy scheme. The null hypothesis is that the point estimate is equal to the full capitalization benchmark. i.e $\hat{\theta} = 180$. The housing characteristics analysis showed that the subsidy scheme affected the properties' age and possibly quality. These two characteristics could explain the differences in the value of the structure. Newer houses and better quality houses are usually more expensive. To test this hypothesis, columns 4 and 8 include the housing characteristics as additional controls. In particular, I control by the mean interior quality score, size, and age of the units in each block and the share of units built after 1997.

Table 2: Fuzzy RD estimate for appraisals

	Semipar				Non-parametric			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A.Mult-Family Units							
Land	-213.24*	-215.85	-171.53+	-107.44	-453.05	-388.23	419.70	19.12
	(107.77)	(138.12)	(100.12)	(78.62)	(408.10)	(324.65)	(765.80)	(163.96)
p-value	0.76	0.80	0.93	0.36	0.50	0.52	0.43	0.22
Structure	-209.76+	-230.84+	-197.26	6.20	-658.88	-518.64	478.16	28.10
	(117.96)	(127.19)	(126.97)	(64.85)	(662.48)	(446.57)	(1,252.39)	(224.30)
p-value	0.80	0.69	0.89	0.00	0.47	0.45	0.60	0.35
Property	-296.69	-318.96	-257.83+	-91.25	-847.15	-966.25	2,106.98	-224.02
	(197.18)	(210.86)	(140.31)	(115.71)	(719.56)	(725.20)	(3,132.02)	(268.62)
p-value	0.55	0.51	0.58	0.44	0.35	0.28	0.47	0.87
	B.Single-Family Units							
Land	-169.61***	-161.52***	-176.34***	-114.76***	-169.71	-119.45	-240.70	106.37
	(15.89)	(16.18)	(21.18)	(17.03)	(160.30)	(116.03)	(343.32)	(150.20)
p-value	0.51	0.25	0.86	0.00	0.95	0.60	0.86	0.06
Structure	-82.18***	-78.66***	-65.35***	-16.42*	-96.49	-67.19	-6.44	-12.65
	(9.35)	(5.73)	(8.55)	(6.54)	(66.85)	(44.88)	(34.37)	(35.77)
p-value	0.00	0.00	0.00	0.00	0.21	0.01	0.00	0.00
Property	-175.91***	-166.33***	-171.92***	-119.33***	-44.42	81.00	45.48	391.54
	(30.36)	(27.79)	(37.08)	(31.59)	(139.36)	(123.52)	(303.01)	(268.22)
p-value	0.89	0.62	0.83	0.05	0.33	0.03	0.46	0.03
	C.All Units							
Land	-167.62***	-166.48***	-176.78***	-99.31***	-122.79	-162.99	-38.06	59.56
	(22.65)	(20.34)	(22.16)	(18.03)	(118.60)	(139.07)	(105.06)	(122.98)
p-value	0.58	0.51	0.88	0.00	0.63	0.90	0.18	0.05
Structure	-76.52***	-122.02***	-102.75***	-4.93	-112.11	-206.70	-223.89	-27.03
	(16.88)	(8.86)	(12.35)	(7.85)	(78.85)	(166.68)	(422.83)	(36.29)
p-value	0.00	0.00	0.00	0.00	0.39	0.87	0.92	0.00
Property	-178.09***	-212.99***	-214.58***	-122.17**	-193.20	-235.26	-84.43	-94.97
	(27.80)	(23.31)	(29.03)	(38.66)	(177.38)	(189.88)	(119.01)	(97.28)
p-value	0.95	0.16	0.23	0.13	0.94	0.77	0.42	0.38
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
1993 Char.	No	No	Yes	Yes	No	No	Yes	Yes
House Char.	No	No	No	Yes	No	No	No	Yes

NOTE: This table presents the estimates for the RD design θ using three different estimation methods with and without controls.The estimates are the IV estimates using the assignment rule as an instrument for a block belonging to estrato 3. Each row is the estimate for a different variable. Columns 1 to 4 use different approaches to estimate $k\left(S_{izt}\right)$ in the model $y_{izt} = \alpha + \theta_{2,3}\mathbb{1}_{\left[S_{izt}\geq\delta_2^3\right]} + k\left(S_{izt}\right) + \beta X + \varepsilon_{izt}$. Columns 1 and 2 use a partially linear model (Robinson, 1988) to estimate $k(S_{izt})$. Columns 5 and 6 estimate $\theta usingnonparametricapproach$ $\lim_{s_{izt}\to\delta_2^3+}\mathrm{E}\left(y_{izt}|S_{izt}=s_{izt}\right) - \lim_{s_{izt}\to\delta_2^3-}\mathrm{E}\left(y_{izt}|S_{izt}=s_{izt}\right)$. In this approach I use the method proposed by Cattaneo, Idrobo, and Titiunik (2018) to select the optimal bandwidth. The first stage and reduced form are represented in figures 5 and ??. The controls included are a dummy equal to 1 if the block changed the estrato since 1997, habitat zones, and the share of blocks that are multi family units. House Char. controls includes age, quality score size and the share of houses built after 1997. The numbers in parenthesis are robust standard errors. The figure contains all the blocks in estratos 2 and 3 in habitat zones 8,9. I do not include the habitat zones 7 and 10 because they do not have enough block with multi family homes in estrato 2. I drop the observations with a score lower than 3 or higher than 10.

Significance level $^{+}10\%,^{*}5\%,^{**}1\%^{***}0.1\%$

This table shows that the housing market seems to capitalize the savings caused by the subsidy scheme. Independent of the estimation method, the land price is indistinguishable from the multifamily units' full capitalization benchmark. The point estimates for single-family units when I do not include controls are the same for the partially linear and non-parametric approaches; -169 Thousand COP/m^2. It is a bit higher in the parametric approach. Once I include controls for housing characteristics, the partially linear model's point estimate is 98.40, about half of the full capitalization benchmark. I cannot reject the null hypothesis that they are the same for any of the estimation approaches. If I pool all the units (in panel C), I get a similar result. Consistent with a spatial equilibrium prediction, these estimates suggest that the housing market capitalizes the subsidy into land prices.

In terms of structure, the mean shift in age and quality score seems to explain the discontinuity in structure prices apparent in figures 10. The point estimate is around -20 Thousand COP/m^2 for all the models using the partially linear and parametric approaches once I include the controls. The non-parametric approach estimates are very noisy, particularly for multi family units. I cannot reject the null hypothesis that they are equal to the full capitalization in all the cases. The only exception is structure when I introduce controls for the quality characteristics in the single-family units. In this case, the point estimate is -6.44 Thousand COP/m^2. This result suggests that the age and quality effect's market price is close to the full capitalization effect.

The total property effect for all the units has a point estimate of -127, -151, and -84 in each estimation method after I control for the housing characteristics. The point estimates are similar when I estimate separately for multi and single-family units. I cannot reject the null hypothesis that the housing market capitalizes on the subsidy in all the cases and estimation methods. This is the case after I control for housing characteristics and the effect of new housing described previously.

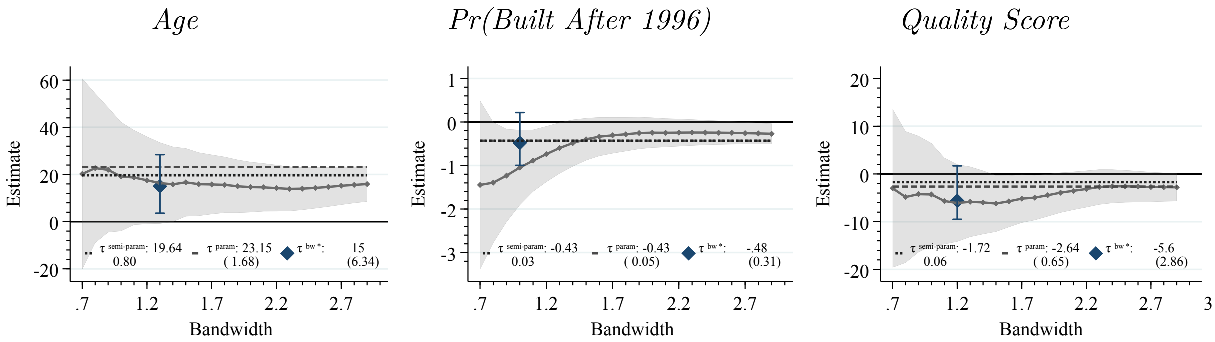
Two main results emerge. First,the city growth is being affected by the subsidy scheme. The units in the heavily subsidized areas are newer, and the share of buildings built after the stratification methodology started is larger. Second the subsidy is offset by an increase in housing prices. House characteristics and construction timing explain most of the observed mean shifts in the price of the structures. But the land values and property values in estrato 2 are higher, even after controlling by housing characteristics. I cannot reject the hypothesis that the effect is equal to the full capitalization.

VI. Robustness

The control function $g(s)$ in my setting is an object of interest, It shows the relationship between neighborhood quality and the outcome of interest. Under the assumption that this function is the same in both sides of the cutoffs, the partially linear model and a parametric approach are more efficient at estimating the causal effect of the subsidy scheme because I use all the available data, instead of using only the data around the cutoff. Even without relying on the full sample and the meaning of the control function, I can use the fact that the blocks to the right and left of the cutoff have similar probability of being treated to identify the effect of the subsidy. The optimal bandwidth estimation relies more heavily on this assumption and the control function is not relevant. This comes at the cost of using only the information around the cutoff. To check the sensitive on the non parametric approach to the selection of the band with. I compare my estimates with the estimates using different bandwidths.

Figure 11, shows the estimates for age, probability of being built after 1997, interior quality score, and the different measures of property values. The x-axis represents the bandwidth; the y-axis is the point estimate. The dashed line is the parametric approach, the dotted line is the partially linear model, and the dot is the optimal bandwidth estimate. The solid line with the shaded area is the estimates for the different bandwidth. The figure shows that the estimates for housing characteristics do not depend on the estimation method. The point estimates of the partially linear model and the parametric approach are always in the confidence interval of the estimates using a bandwidth. The figure makes the trade-off between bias and precision of the estimates. Using a bandwidth, I am more agnostic about the functional form of the control function. However, this comes at an efficiency cost, because I only use the information around the cutoff. The figures for the housing values illustrate this point clearly. With a narrow bandwidth, the point estimates have large confidence intervals and include the point estimates of the other less agnostic approaches. When I increase the bandwidth, I gain precision on my estimates, and the bandwidth estimates converge towards the partially linear model and the parametric estimates. This gives me some confidence in those approaches. While this may not be the case in some RDD applications, in this particular case, using the full sample and estimating the control function seems reasonable and allows for point estimates more precisely estimated.

A. Housing Characteristics



B. Property Values (All units)

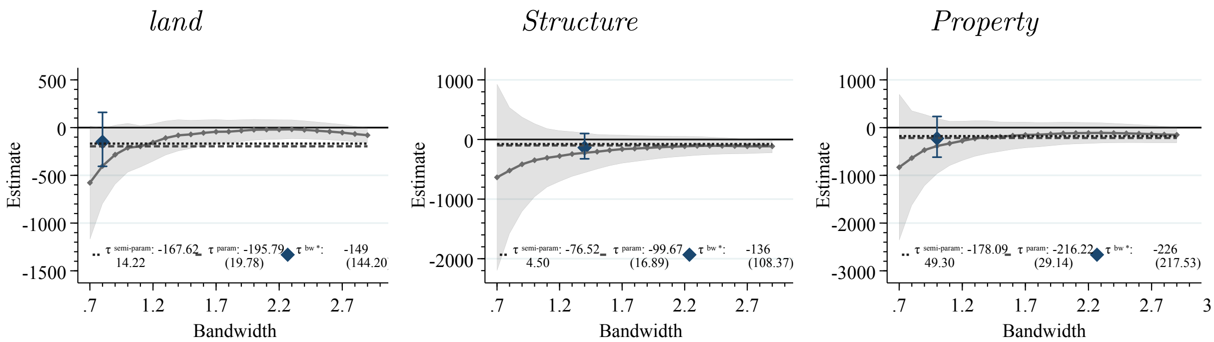


Figure 11: Estimates Using Different Bandwidths

Table 3: The Effect on Construction and Housing Characteristics.

	(1)	(2)	(3)
Pr(Built>1996)	-0.43***	-0.42***	-0.40***
	(0.05)	(0.05)	(0.05)
Age	23.15***	25.03***	19.71***
	(1.68)	(1.57)	(1.49)
Quality Score	-2.64***	-2.31***	-3.85***
	(0.65)	(0.64)	(0.63)
Size m^2	-23.25**	-8.10	0.51
	(7.83)	(7.62)	(7.56)
Controls	No	Yes	

Note: This table presents the estimates θ using three different estimation methods with and without controls.The estimates are the IV estimates using the assignment rule as an instrument for a block belonging to estrato 3. Each row is the estimate for a different variable. Columns 1 to 4 use different approaches to estimate $k\left(S_{izt}\right)$ in the model $y_{izt} = \alpha + \theta_{2,3} \mathbb{1}_{\left[S_{izt} \geq \delta_2^3\right]} + k\left(S_{izt}\right) + \beta X + \varepsilon_{izt}$. Columns 1 and 2 use a partially linear model Robinson (1988) to estimate $k(S_{izt})$. Columns 5 and 6 estimate θ using non parametric approach $\theta = \lim_{s_{izt} \to \gamma_2^{3+}} \lim \mathrm{E}\left(y_{izt} | S_{izt} = s_{izt}\right) - \lim_{s_{izt} \to \gamma_2^{3-}} \mathrm{E}\left(y_{izt} | S_{izt} = s_{izt}\right)$. In this approach I use the method proposed by Cattaneo, Idrobo, and Titiunik (2018) to select the optimal bandwidth.The first stage and reduced form are represented in figures 5 and 9. The controls included are a dummy equal to 1 if the block changed the estrato since 1997, habitat zones, and the share of blocks that are multi family units. The numbers in parenthesis are robust standard errors. The figure contains all the blocks in estratos 2 and 3 in habitat zones 7,8,9 and 10. I drop the observations with a score lower than -3 or higher than 10.

Significance level $^{+}$10%,* 5%,** 1%***0.1%

The second robustness exercise is to check that the mean shift I observed happens only at the cutoff. To check this, I plot the R^2 and t-statistic using different hypothetical cutoff points. If the biggest t-statistic and R^2 is at the cutoff, this is a good indicator that the mean shift is not an artificial jump unrelated to the subsidy scheme. I use the parametric approach, including the basic set of controls. Figure 12 shows the results for the select set of outcomes. There is an important spike on the t-stat and R^2 in all the outcomes but on the interior quality score. This gives less confidence to that result. Also, it is noteworthy that for the case of land, there seems to be another mean shift around 2.5. This may reduce the confidence in the land price estimates.

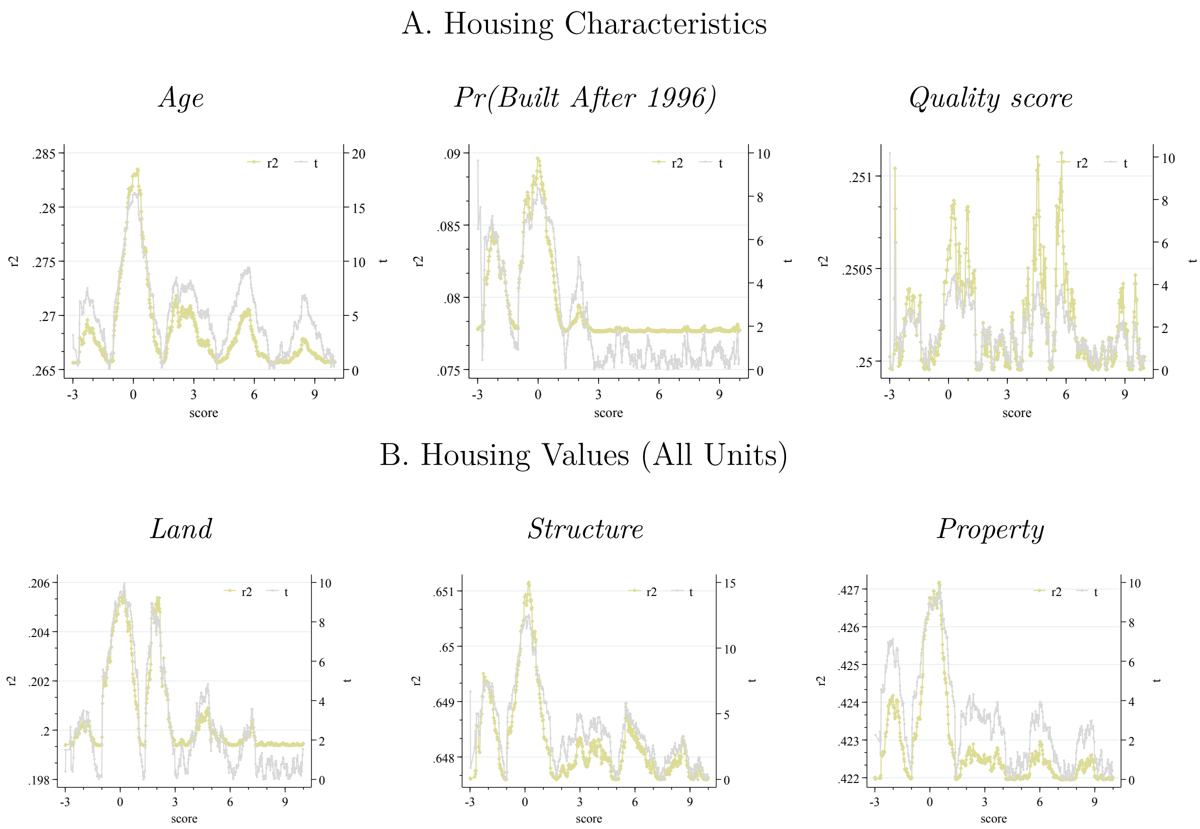


Figure 12: R^2 AND t AT DIFFERENT CUTOFFS

VII. Concluding remarks

My findings suggest that using a location to target redistribution policies has an impact on the housing market. There is any apparent effect on the new construction and capitalization of the subsidy into the housing market. Blocks in areas receiving a higher subsidy, have housing units 30 years more modern on average and have 43 percent more chances of having units built after the new targeting tool was introduced. The houses in blocks of estrato 2 capitalize on the differences in utility prices and property taxes. The NPV of the differences in costs is 180 thousand COP/m^2, and my preferred estimate for land prices is 89 and 127 thousand COP/m^2 for total property value.Under all the specifications, I cannot reject the null hypothesis of full capitalization.

These results have several implications that need to be a considered when using location as a targeting tool. First, the intended redistributive purposes can fail if the subsidy ends up being transferred to property owners in targeted areas. In the case of Bogotá, around 50 percent of households rent their houses in the neighborhoods in estrato 2 and

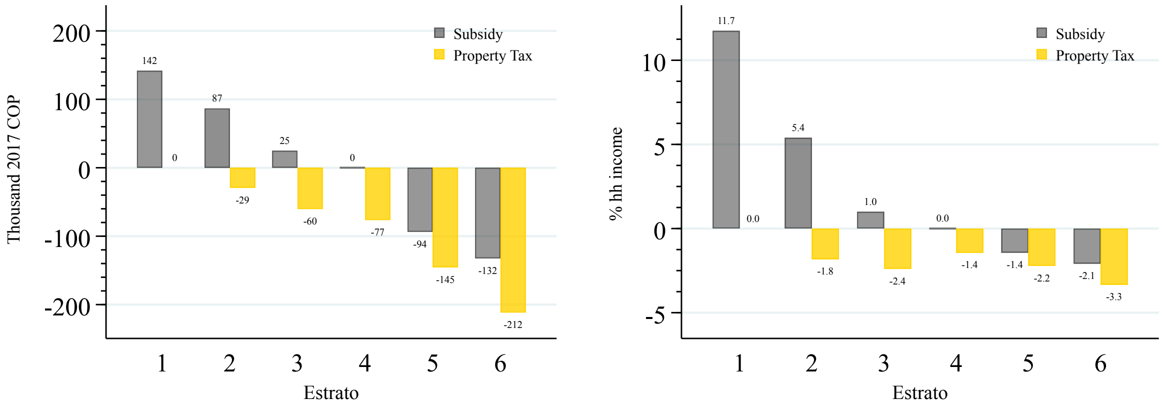
3. If the increase in housing prices get charged into monthly rents, the renters do not benefit from the transfers. In Colombia,renters usually pay the utilities bill. Moreover, if developers and not individual families drive the new construction, the subsidy scheme could be developers and not poor people.

Additionally, these types of schemes may generate inefficient city growth. If individuals and developers build new units to take advantage of the subsidy and not to get closer to job opportunities or market access, the city may grow inefficiently. People may decide to live far away from jobs and spend more time commuting, to take advantage of the subsidy. Developments driven by the subsidy will likely persist even if the subsidy scheme disappears. Finally, the transfer system may prevent some areas from experiencing dynamic improvements if the resident fears to lose the subsidy status. A careful analysis of those potential implications and the costs of this paper's findings needs to be considered when evaluating the efficacy of using a location as a targeting tool. This is an area of future research.

Recent research suggests that neighborhoods are critical for determining future income (Chetty, Hendren, and Katz (2016), Chetty, Friedman, Hendren, Jones, and Porter (2018)). A possible policy implication of these studies is to use location characteristics to target redistributive policies. A location-based transfer affects people's decisions about where to live, and therefore it may affect the housing market. These unintended effects can offset the intended impact of the location-based policy. Gaubert et al. (2020) suggests that location-based subsidy could be a useful tool for redistribution. I provide some evidence of the unintended consequences of using location. The costs of these consequences should be included in a framework analyzing the efficacy of location-based subsidies.

VIII. Appendix:

a. Monthly Subsidy and Property Tax *b. ... as a % of HH expenditure*

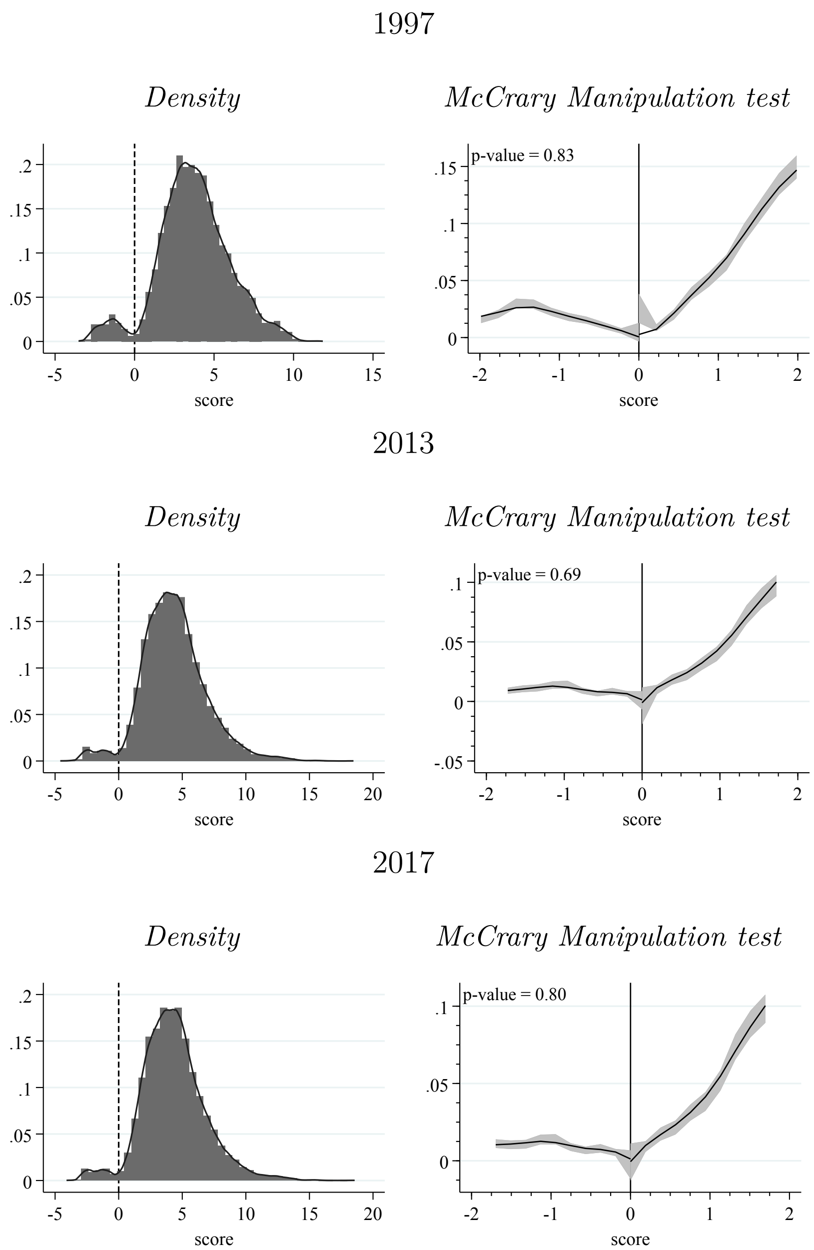


Appendix Figure 1: Subsidy Size

Note: The subsidy is the monthly average for each subsidy code. The property tax is the estimated payment of a house of average size assessed at an average price for m^2 divided by 12 within each code. As reference point, the GDP per capita in Colombia in 2011 was 13.4 million COP and the the average exchange rate for 2011 was 1848.17 COP pesos for a Dollar.

1. Baseline Characteristics

i. No manipulation assumption (individual years)



Appendix Figure 2: Score Distribution

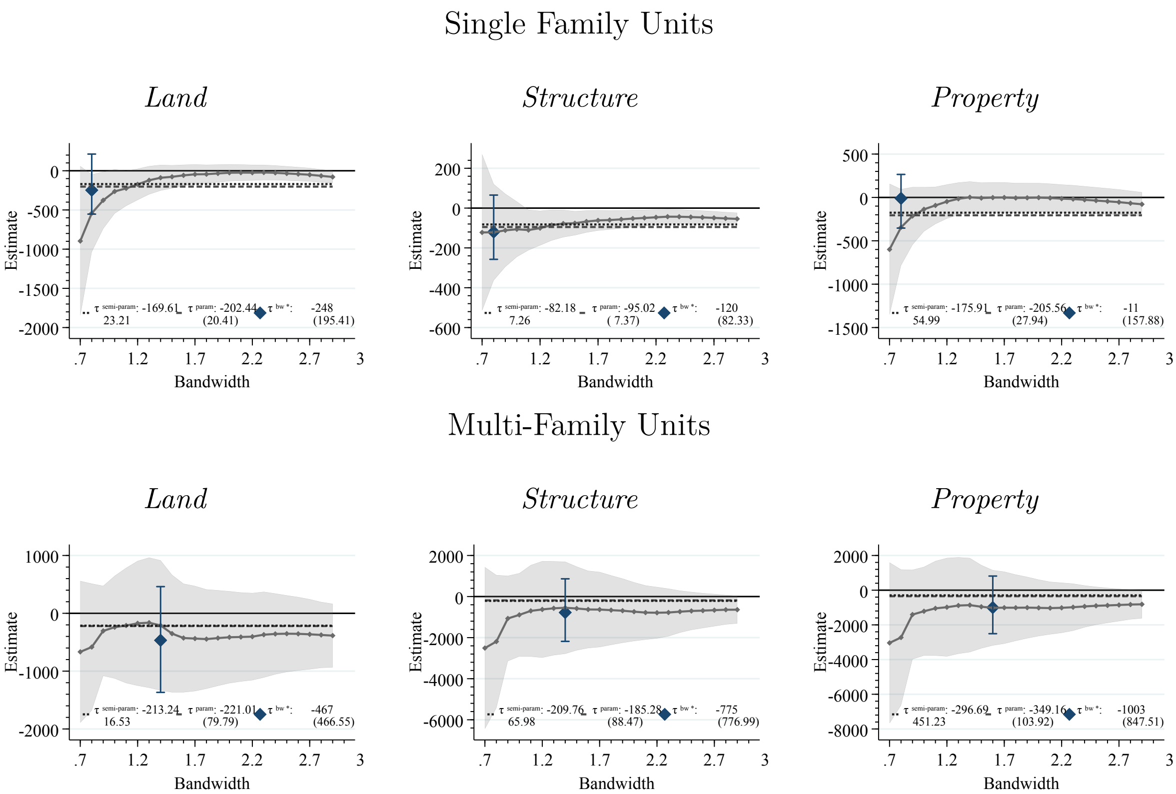
Note: H_0: The density is the same on the right and on the left. The p-value of the test is on the top of panel b). I cannot reject the null hypothesis. To implement the manipulation test I follow (Cattaneo, Jansson, & Ma, 2018)

Appendix Table 1: The Effect on Construction and Housing Characteristics. (Habitat Zones 8 and 9)

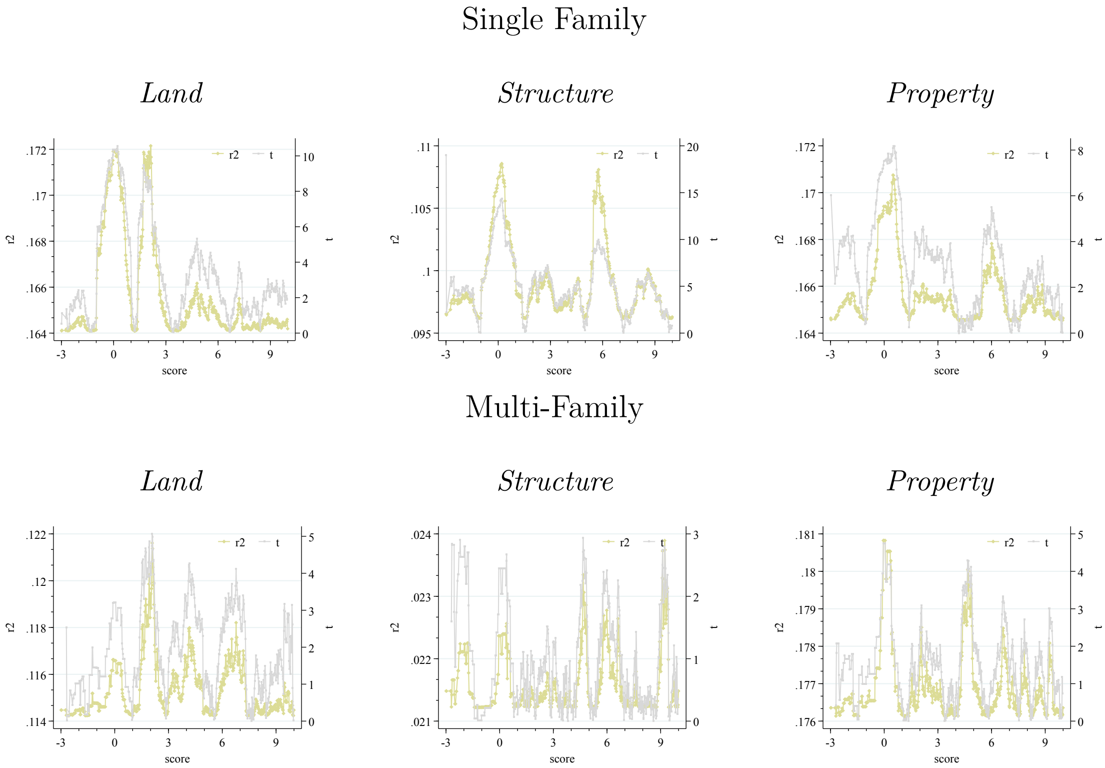
	Semipar		Parametric		Non-parametric	
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(Built>1996)	-0.49***	-0.49***	-0.48***	-0.48***	-0.68+	-1.05+
	(0.05)	(0.08)	(0.05)	(0.05)	(0.40)	(0.61)
Age	22.36***	24.47***	25.41***	27.17***	17.45***	27.09*
	(1.58)	(1.33)	(1.59)	(1.49)	(5.30)	(11.86)
Quality Score	-1.60*	-1.82***	-2.77***	-2.36***	-3.55+	-4.62
	(0.65)	(0.52)	(0.66)	(0.65)	(2.01)	(4.52)
Size m^2	2.65	21.22***	-12.86+	3.04	15.60	68.15
	(6.69)	(6.24)	(7.01)	(6.97)	(19.50)	(68.94)
Controls	No	Yes	No	Yes	No	Yes

Note: This table presents the estimates θ using three different estimation methods with and without controls.The estimates are the IV estimates using the assignment rule as an instrument for a block belonging to estrato 3. Each row is the estimate for a different variable. Columns 1 to 4 use different approaches to estimate $k\left(S_{izt}\right)$ in the model $y_{izt} = \alpha + \theta_{2,3}\mathbb{1}_{\left[S_{izt} \geq \delta_2^3\right]} + k\left(S_{izt}\right) + \beta X + \varepsilon_{izt}$. Columns 1 and 2 use a partially linear model Robinson (1988) to estimate $k(S_{izt})$. Columns 5 and 6 estimate θ using non parametric approach $\theta = \lim_{s_{izt} \to \gamma_2^{3+}} \lim \mathrm{E}\left(y_{izt} | S_{izt} = s_{izt}\right) - \lim_{s_{izt} \to \gamma_2^{3-}} \mathrm{E}\left(y_{izt} | S_{izt} = s_{izt}\right)$. In this approach I use the method proposed by Cattaneo, Idrobo, and Titiunik (2018) to select the optimal bandwidth.The first stage and reduced form are represented in figures 5 and 9. The controls included are a dummy equal to 1 if the block changed the estrato since 1997, habitat zones, and the share of blocks that are multi family units. The numbers in parenthesis are robust standard errors. The figure contains all the blocks in estratos 2 and 3 in habitat zones 7,8,9 and 10. I drop the observations with a score lower than -3 or higher than 10.

Significance level $^{+}10\%,^{*}5\%,^{**}1\%^{***}0.1\%$



Appendix Figure 3: Using Different Bandwidths



Appendix Figure 4: R^2 AND t AT DIFFERENT CUTOFFS

2. Estratos Assignment Details

Appendix Table 2: Percentage of Blocks Correctly Predicted

	1997	1999	2009	2013	2017	All
Estrato 1	99.87	99.91	99.61	99.56	99.50	99.69
Estrato 2	99.89	99.91	99.54	99.47	99.41	99.65
Estrato 3	99.98	99.99	99.86	99.78	99.74	99.87
Estrato 4	99.74	99.91	98.91	98.18	98.21	98.99
Estrato 5	98.03	98.09	94.06	93.67	93.27	95.45
Estrato 6	99.52	98.75	98.11	97.97	97.82	98.45
All	99.85	99.86	99.44	99.32	99.25	99.55
Observations	36985	37686	37089	37096	36258	185114

Note:

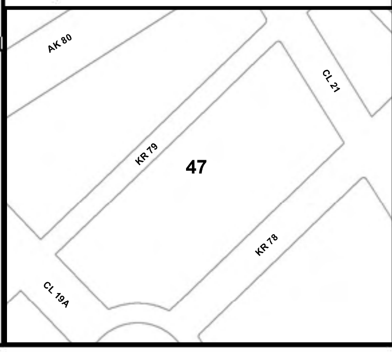
3. How Information is Collected

FORMULARIO
ESTRATIFICACION SOCIOECONÓMICA URBANA DE BOGOTÁ, D.C. LOCALIDAD 0

I. IDENTIFICACIÓN

1 [0] Urbano/Rural 4 [2 0] Barrio/Subvereda
2 [0 6] Circuito/Término 5 [4 7] Manzana
3 [3] Sector/Vereda
6 Dirección en Terreno
Calles ______________ Carreras ______________
7 Nombre del Barrio: CIUDAD HAYUELOS
8 Nombre del Conjunto: ______________
9 Código Sectorización: 006320

10. Croquis de la manzana



III. CONTEXTO URBANISTICO

La manzana pertenece a la zona [ ]

V. DATOS DE RECOLECCIÓN

FECHA:
CÓDIGO RECOLECTOR:
CÓDIGO SUPERVISOR:
CÓDIGO DIGITADOR:
Observaciones: ______________

II. CARACTERISTICAS DE LA VIVIENDA Y SU ENTORNO

PREGUNTAS	RESPUESTAS	Cód	A	B	C	D	E	F	G	H	I	J	K
1. En el lado de manzana hay Vivienda con Entrada Principal	Sí	1											
	No	2											
2. Vías de Acceso La Calle o la Vía del lado de la Manzana es :	Sendero o camino	1											
	Peatonal	2											
	Vehícular en tierra	3											
	Vehícular en recebo - balasto o gravilla -	4											
	Vehícular en cemento, asfalto o adoquín	5											
3. Tamaño del frente El tamaño predominante del frente de las Viviendas del Lado de la Manzana es :	Hasta 7 metros	1											
	Entre más de 7 y 9 metros	2											
	Entre más de 9 y 12 metros	3											
	Más de 12 metros	4											
4. Andén Predominan en el lado de la Manzana viviendas :	Sin Andén	1											
	Con Andén Sin Zona Verde	2											
	Con Andén con Zona Verde	3											
5. Antejardín Predomina en el lado de la Manzana Viviendas :	Sin Antejardín	1											
	Con Antejardín Pequeño	2											
	Con Antejardín Mediano	3											
	Con Antejardín Grande	4											
6. Garajes Predominan en el lado de la Manzanas Viviendas	Sin Garaje ni Parqueadero	1											
	Con Garaje Cubierto Usado para otros fines	2											
	Con parqueadero o Zona de Parqueo	3											
	Con Garaje Adicionado a la Vivienda	4											
	Con Garaje Sencillo que hace parte del diseño Original de la Vivienda	5											
	Con Garajes Dobles o en Sótano	6											
7. Material de las Fachadas Predominan en el lado de la Manzana Viviendas con Fachadas :	En Guadua, Caña, Esterilla, Tablas y Desechos	1											
	Sin cubrir-adobe, bahareque, tapia pisada, placa prefabricada, bloque o ladrillo común	2											
	En Revoque - Pañete o Repello - Sin Pintura	3											
	En Revoque - Pañete o Repello - Con Pintura	4											
	Con Enchapes, en Ladrillo Pulido o en Madera Fina	5											
8. Material de los Techos Predominan en el lado de la Manzana Viviendas con Techos en :	Desechos, Telas Asfálticas ó Pedazos de Tejas	1											
	Placa de Entrepiso	2											
	Terraza, Azotea o Cubierta Sencilla	3											
	Lujosa u Ornamental	4											

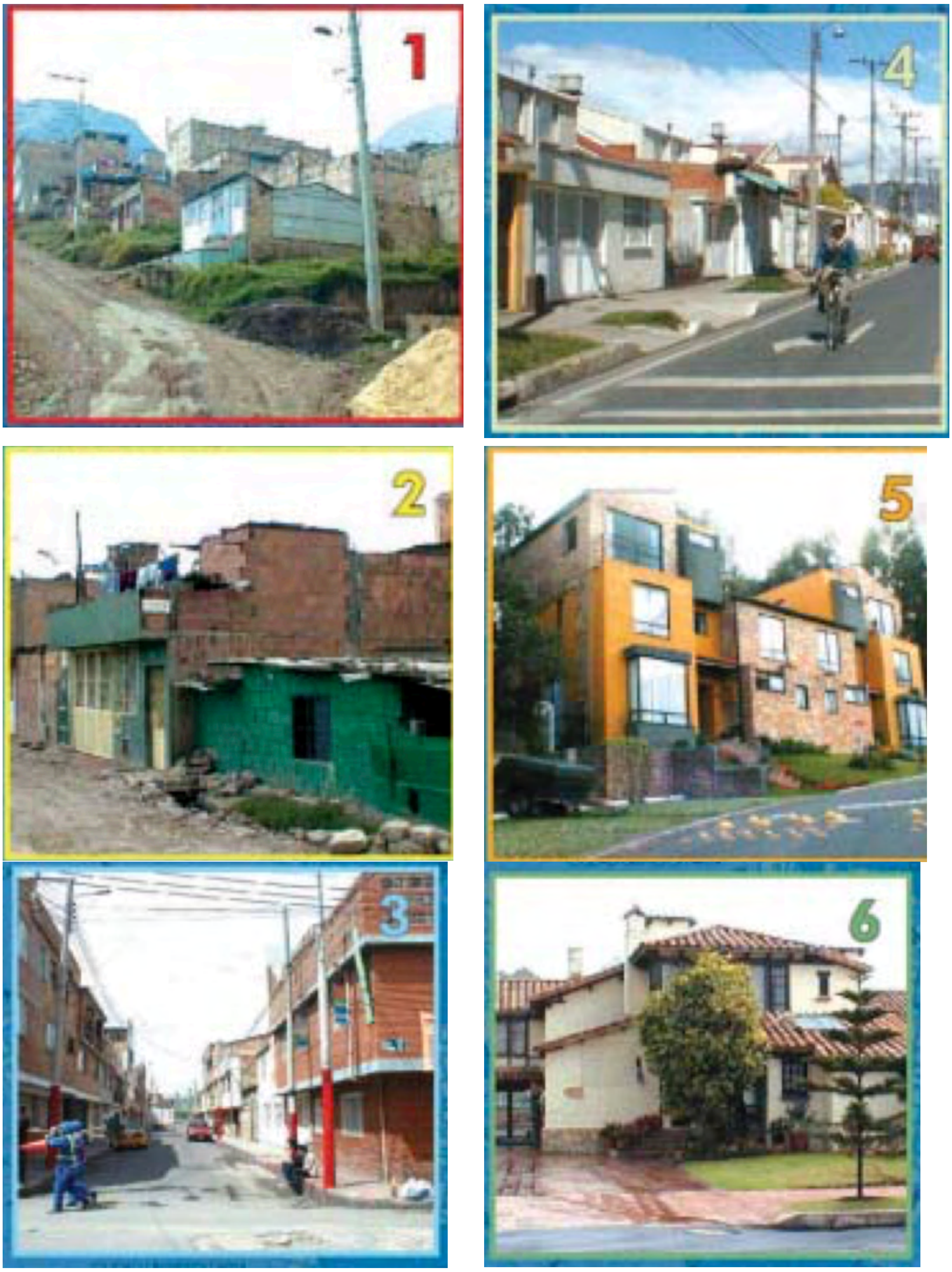
IV. LISTADO DE VIVIENDAS ATÍPICAS

Tenga en cuenta que una vivienda es ATÍPICA cuando difiere del resto en la manzana por presentar evidente contraste en el tamaño, los materiales, el terminado, el estado de deterioro o conservación.

Lado de Manzana	Dirección	Justificación	Atipicidad (+/-)

Appendix Figure 5: Stratification Census Form

NOTE: This Figure shows the form used to collect the information required to assign the *estratos*



Appendix Figure 6: EXAMPLE UNITS BY ESTRATO

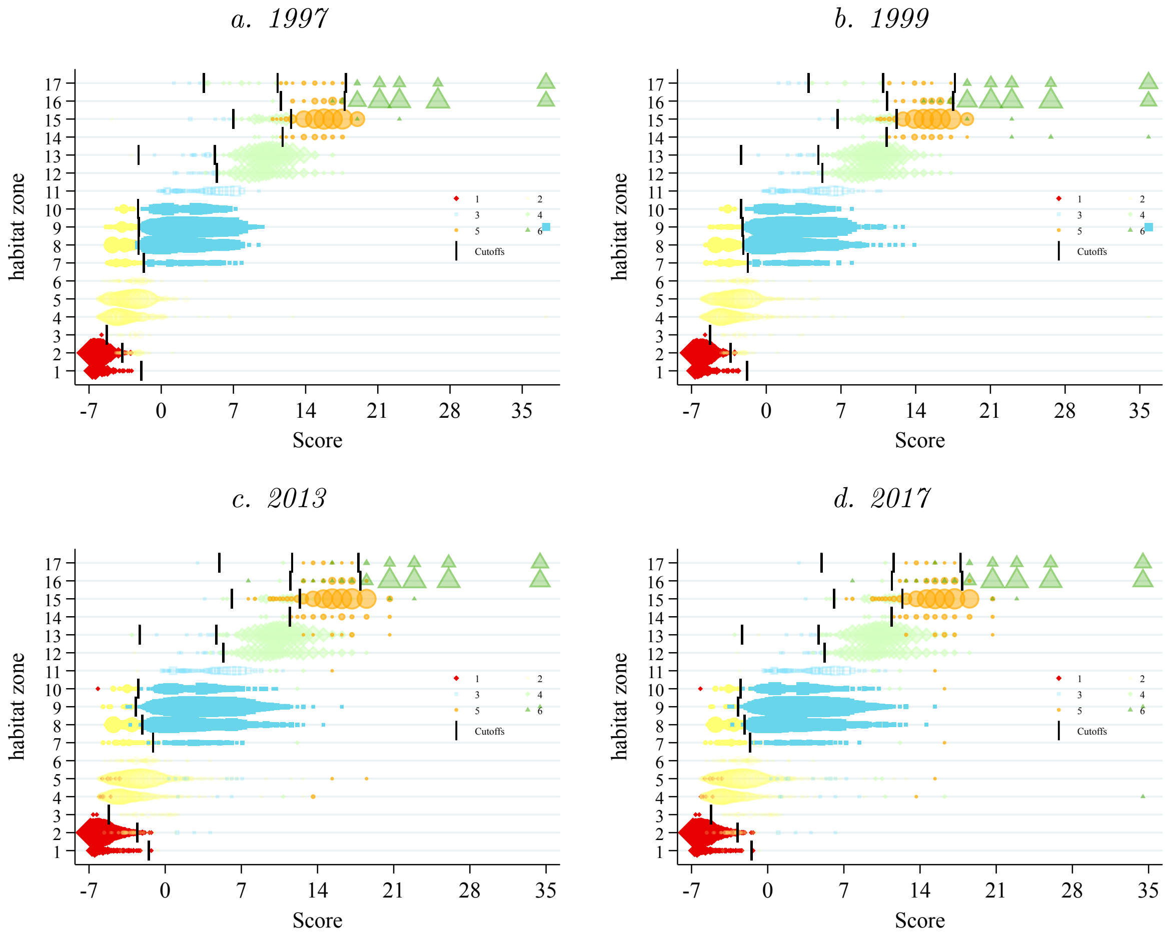
NOTE:

4. Defining the Habitat Zones

Appendix Table 3: Habitat zones.

General zone	Habitat zone	Description
1. Poverty	1	Blocks with lacking planning. Fragile and short-lived materials characterize the houses. The houses usually do not have spaces specifically designed for cooking, bathing, or washing clothes. These houses have high housing density. Public areas, such as platforms and roads, are limited. In their immediate surroundings typically have black waters, garbage, and animals. They present problems of natural risk (i.g. landslides, floods, marshy lands, etc.)
	2	
2.Tolerance zone	3	Blocks characterized by high delinquency, prostitution, distribution, or consumption of drugs.
3. Unconsolidated development	4	Blocks with unfinished houses. They can be in blackwork or with uncovered or unpainted facades and temporary roofs, whose planks will eventually. Homes built in stages over the years. Blocks with scattered buildings or sides of the block with an abundant presence of lots without closing or without a known destination; These characteristics make these spaces appear in the process of consolidation.
	5	
4.Urban deterioration	6	Blocks in the historic center of the city. The facade looks old (manifested by aged ceilings, broken eaves, deteriorated facades, raised paintings, doors, corroded windows, etc.). The immediate surroundings show narrow streets and sidewalks, without front yards or green areas. We observe the presence of economic establishments of a different use (shops, mechanic workshops, canteens, or restaurants) that degrade the quality of homes. In some cities, the tenants are in these areas; it refers to large houses where several families separated in independent pieces with sanitary, kitchen, and patio clothing services.
5.Industrial	7	Blocks with "factories" - destined for the massive production of goods. Surrounding areas have warehouses, premises for the sale of food, and the permanent traffic of trucks. The waste and noise, characteristic of the factories, pollute the environment.
6.Consolidated progressive development	8	Blocks with self-construction homes. They express the culmination of progressive development. For this reason, the landscape is heterogeneous or architectural diverse. The buildings occupy the space on each side of the block in a continuous manner, in such a way that its urban structure can be considered consolidated and definitive. It can include social housing, finished, and built-in series.
	9	
7. Commercial	10	Blocks with mainly commercial buildings; initially, many of them were housing units but were refurbished for shops. Usually, homes are on the upper floors or in the interior part of the buildings. This zone is the CBD; in small cities, it is located around the plaza or along the main street. In some cities, these areas exist in the traditional historical center but could be in other places, such as neighborhoods and high vehicular traffic lanes. The public space is minimal, and the congestion of customers degrades the quality of the homes.
	11	

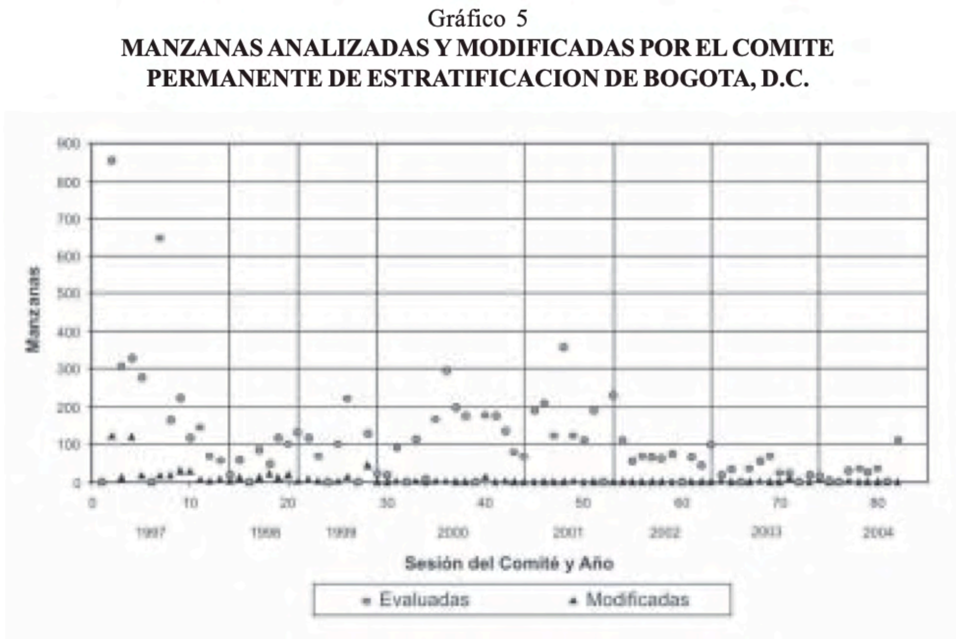
8.Intermediate residential	12	Block with finished homes in residential neighborhoods. The immediate surroundings have wide public spaces, streets in good condition, green areas, and low density of commercial establishments.
	13	
9.Commercial compatible	14	Blocks with residential homes and access to amenities such as gyms and saloons, boutiques or luxury goods stores, cigar stores, warehouses with decorated showcases, florists, etc. The immediate environment have main roads with high vehicular traffic and alternate roads with less traffic.
10.Exclusive residential	15	Blocks with residential buildings with modern designs, green areas, special systems of private surveillance, and almost no presence of economic establishments.
	16	
11.Low density residential	17	Blocks with residential homes with the architectural design of their homes and salient decoration (i.g. fountains, gardens, lighting systems, etc.). It is characterized by big houses, mansions, or majestic buildings. In some cases, the roads are exclusive for residents and visitors and have private services.
12.Institutional	18	Non-residential blocks. This zone includes blocks with institutional use, those with lots and others without homes, and those dedicated to green areas..
	19	
	20	



Appendix Figure 7: Subsidy Codes Based on Habitat Zones and the Score

Note:

The proposed cutoffs are the dark solid lines. Each dot in the figure represents one or more blocks with a given score and zone. The symbols are weighted to represent the number of blocks with a particular score and zone. The six different colors represent the six subsidy codes. For example, the green cross are blocks belonging to subsidy code 3. There are three main takeaways from this figure. First, the habitat zones play a more important role than the score at determining the different codes. Second, the worst zones (lower numbers) have lower scores than the best areas (higher numbers). However, there is some overlap in the index between some areas. Third, only a few zones have more than one code. For my empirical strategy I use the discontinuities introduced by the cutoffs.



Appendix Figure 8: Reassignment of *Estratos*
Note: *source:* (Departamento Administrativo de Planeación Distrital, 2004)

5. Data Details

i. Constant Geographic Unit

There are two geographic codes for the blocks in Bogota. The Census Block Code from the National Geo-static Framework (*MGN*-from the Spanish spelling) (DANE, 2018) and the the cadaster block code (*ManCodigo* (Igac, 2013)). The National Adminisrative Department of Statistics (*DANE*-from the Spanish spelling) uses the The *MGN* code, and the district cadaster and the secretary of district planning use *ManCodigo* code to locate each property and to produce the stratification census.

The *MGN* code starts in 1993 with that year's population census. The codes have no significant changes over time, therefore we can track units over time. This is not the case for the cadaster block code, *ManCodigo*. Some blocks in the same geographic location change the *ManCodigo* over time. Therefore, to follow the same location over time, we need to define a time-invariant code for each location and unit of observation. I use the shapefiles of the Stratification Census to create such a code.[31] The shapefiles with

[31] The secretary of district planning provided me those shape files.

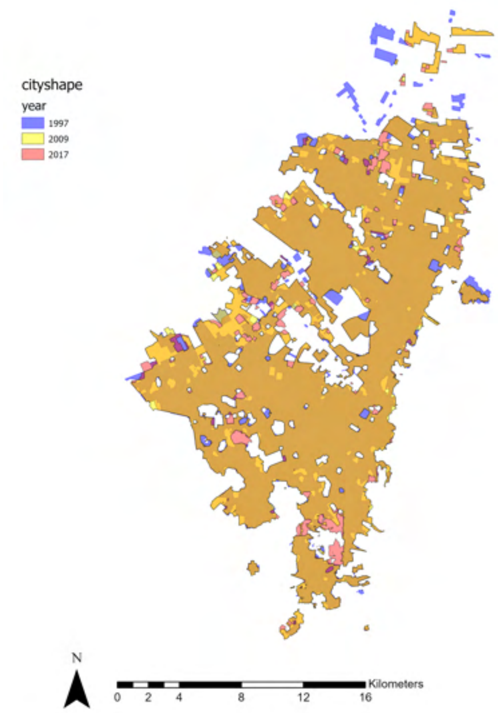
the codes corresponding to the stratification census for the all the updates, allow me to know the exact geographic location of each unit independently of the code assigned to the unit in each year. The appendix figure 9 shows the changes in the residential city boundary. The appendix figure 9a shows the changes in residential areas in the city since 1997. The appendix figure 9b shows the places with codes that existed in 1997 but not in 2017. This figure makes it clear that the changes in codes are not only related to a change in residential city boundaries. Therefore, a time invariant geographic code is essential to analyze the data.

I create a stable geographic unit over time and create a crosswalk between the the two types of codes (MGN and ManCodigo). To do that, I fix the 2017 geographic codes, *ManCodigo*, and assign that code to each unit of observation based on the location obtained with the corresponding shapefile. The crosswalk between *MGN* and *ManCodigo* allows me to use the 1993 Population Census, to check for the balance in the characteristics of each block before the policy implementation. I create the crosswalk using the shapefiles for each type of code and doing a geographic merge, similar to what I did for the time-invariant code.[32]

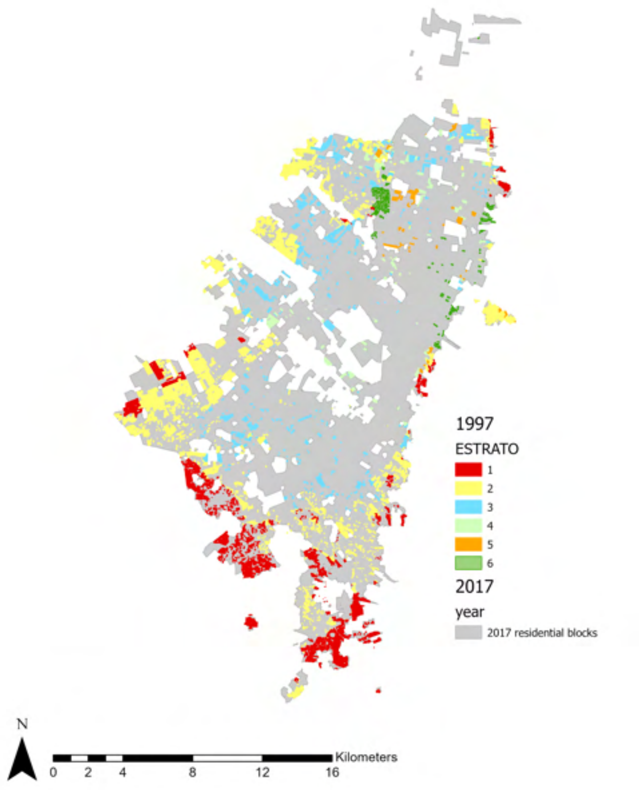
[32]For the MGN I only have the 2005 shape file. In total I geolocate an important fraction of the blocks. Aliaga-Linares and Alvarez-Rivadulla (2010) does a similar exercises.

Appendix Figure 9: Changes in codes and city boundaries

a. Changes in city boundaries



b. Code changes between 1997 and 2017



Note: The figure 9b shows the blocks that "exit" from 1997 to 2017 if we do not adjust the administrative code changes. Figure **??** shows the residential blocks for the different years where and update of the *stratification census* took place

ii. Changes over time

Because I have a time-invariant geographic code, I can check the changes in locations for each unit over time. Appendix Table 4 shows the changes in the estratos overtime. The top panel shows the changes between 2009 and 1997. Out of the 36,985 blocks in 1997, 32,425 have the same estrato in the two periods, 1,006 have a lower estrato in 2009 and 533 have a higher estrato, 3,125 blocks starting being defined as residential blocks and 3,021 were removed.

Appendix Table 4: Changes in the Subsidy Codes Over Time

	2009-1997						
	Same	Down	Up	New	Exit	Total 1997	Total 2009
1	5,249	.	114	1,063	1,023	6,386	6,426
2	12,881	289	316	1,360	1,361	14,847	14,846
3	10,674	477	69	436	455	11,675	11,656
4	2,029	181	28	151	45	2,283	2,389
5	877	25	6	46	55	963	954
6	715	34	.	69	82	831	818
All	32,425	1,006	533	3,125	3,021	36,985	37,089

	2017-2009						
	Same	Down	Up	New	Exit	Total 2009	Total 2017
1	6,375	.	35	50	193	6,603	6,460
2	14,268	4	18	43	548	14,838	14,333
3	11,379	4	12	34	188	11,583	11,429
4	2,266	2	3	11	22	2,293	2,282
5	969	2	.	5	6	977	976
6	769	.	.	9	26	795	778
All	36,026	12	68	152	983	37,089	36,258

.

VIII. References

Aliaga-Linares, L., & Alvarez-Rivadulla, M. J. (2010). Segregación residencial en bogotá a través del tiempo y diferentes escalas. *Documento de Trabajo de Lincoln Institute of Land Policy. Recuperado de http://bit. ly/19h8zP7*.

Bayer, P., Ferreira, F., & McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, *115*(4), 588-638. Retrieved from link

Black, S. E. (1999). Do better schools matter? parental valuation of elementary education*. *The Quarterly Journal of Economics*, *114*(2), 577-599. Retrieved from link

Casas, C. (2014, October). *Subsidies to Electricity Consumption and Housing Demand in Bogotá* (Borradores de Economia No. 847). Banco de la Republica de Colombia. Retrieved from link

Cattaneo, M., Idrobo, N., & Titiunik, R. (2018). *A practical introduction to regression discontinuity designs: Volume i.* Cambrigde University Press, forthcoming.

Cattaneo, M., Jansson, M., & Ma, X. (2018, November). *Simple Local Polynomial Density Estimators* (Paper). arXiv.org. Retrieved 2019-05-10, from link

Cattaneo, M., Titiunik, R., & Vazquez-Bare, G. (2019). The Regression Discontinuity Design. , 36.

Chay, K. Y., & Greenstone, M. (2005). Does air quality matter? evidence from the housing market. *Journal of Political Economy*, *113*(2), 376–424. Retrieved from link

Chen, J., Glaeser, E. L., & Wessel, D. (2019, December). *The (non-) effect of opportunity zones on housing prices* (Working Paper No. 26587). National Bureau of Economic Research. Retrieved from link

Chetty, R., Friedman, J. N., Hendren, N., Jones, M. R., & Porter, S. R. (2018, October). *The opportunity atlas: Mapping the childhood roots of social mobility* (Working Paper No. 25147). National Bureau of Economic Research. Retrieved from link

Chetty, R., Hendren, N., & Katz, L. F. (2016, April). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, *106*(4), 855-902. Retrieved from link

Choi, J.-y., & Lee, M.-j. (2018a, January). Minimum distance estimator for sharp regression discontinuity with multiple running variables. *Economics Letters*, *162*, 10–14. Retrieved 2019-04-01, from link

Choi, J.-y., & Lee, M.-j. (2018b, July). Regression Discontinuity with Multiple Running Variables Allowing Partial Effects. *Political Analysis*, *26*(3), 258–274. Retrieved 2019-03-28, from link

DANE. (2018). *Manual De Uso Del Marco Geoestadístico Nacional En El Proceso Estadístico.* Retrieved 2020-06-12, from link

Departamento Administrativo de Planeación Distrital. (2004). *La estratificación en Bogotá D.C. y estudios relacionados, 1983-2004*. Bogotá: Alcaldía Mayor de Bogotá D.C., Departamento Administrativo Planeación Distrital. (OCLC: 221308728)

Econometria. (1999). *Revisión de la metodología de estratificación socioeconómica de Santa Fe de Bogotá. Bogotá.*

Gallego, J., Montoya, S., & Sepúlveda, C. (2016). Effect of socio-economic stratification on house value in bogotá.

Gaubert, C., Kline, P., & Yagan, D. (2020). *Place-based redistribution* (Paper). Berkeley. Retrieved 2020-07-21, from link

Glaeser, E. L., & Gottlieb, J. D. (2008, October). *The economics of place-making policies* (Working Paper No. 14373). National Bureau of Economic Research. Retrieved from link

González, J. I. (2014). *Aglomeraciones y condiciones de vida en Bogotá.* Retrieved 2019-05-14, from link

Gyourko, J., & Molloy, R. (2015). Chapter 19 - regulation and housing supply. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of regional and urban economics* (Vol. 5, p. 1289 - 1337). Elsevier. Retrieved from link

Haan, P., & Simmler, M. (2018). Wind electricity subsidies — a windfall for landowners? evidence from a feed-in tariff in germany. *Journal of Public Economics*, *159*, 16-32. Retrieved from link

Hanna, R., & Olken, B. A. (2018, November). Universal basic incomes versus targeted transfers: Anti-poverty programs in developing countries. *Journal of Economic Perspectives*, *32*(4), 201-26. Retrieved from link

Harker, A., Lustig, N., Martínez, V., & Melendez, M. (2016). *The impact of taxes and transfers on inequality and poverty in colombia* (Tech. Rep.). CEQ Working Paper 24. CEQ Institute, Center for Inter-American Policy and . . . .

Hilber, C. A. L. (2017). The economic implications of house price capitalization: A synthesis. *Real Estate Economics*, *45*(2), 301-339. Retrieved from link

Igac, I. G. A. C. (2013, September). *Resolución 829 De 26 De Septiembre De 2013.* Retrieved 2020-06-12, from link (Library Catalog: legal.legis.com.co)

Kline, P., & Moretti, E. (2014). People, places, and public policy: Some simple welfare economics of local economic development programs. *Annual Review of Economics*, *6*(1), 629-662. Retrieved from link

Lozano-Gracia, N., & Anselin, L. (2012). Is the price right?: Assessing estimates of cadastral values for bogotá, colombia. *Regional Science Policy Practice*, *4*(4), 495-508. Retrieved from link

Lutz, B. (2015, February). Quasi-experimental evidence on the connection between property taxes and residential capital investment. *American Economic Journal: Economic Policy*, *7*(1), 300-330. Retrieved from link

McRae, S. (2015, January). Infrastructure quality and the subsidy trap. *American Economic Review*, *105*(1), 35-66. Retrieved from link

Medina, C., & Morales, L. (2007). Stratification and public utility services in colombia: Subsidies to households or distortion of housing prices? [with comments]. *Economía*, *7*(2), 41–99. Retrieved from link

Meléndez, M. (2004, August). Subsidios al consumo de los servicios públicos en Colombia. ¿Hacia dónde nos movemos? Retrieved 2019-03-08, from link

Papay, J. P., Willett, J. B., & Murnane, R. J. (2011, April). Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics*, *161*(2), 203–207. Retrieved 2019-05-05, from link

Reardon, S. F., & Joseph P., R. (2012). Regression Discontinuity Designs With Multiple Rating-Score Variables. *Journal of Research on Educational Effectiveness*. Retrieved 2019-03-27, from link

Registro, distrital acuerdo número 657 (No. Annex 2). (2016). Retrieved 2020-09-26, from link

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, *56*(4), 931–954. Retrieved from link

Ruiz, F., & Vallejo, G. (2015). *Using land registration as a tool to generate municipal revenue: Lessons from bogota.* Retrieved from link

SDP. (2004). *Bogotá, D.C., y su Estratificación Socioeconómica.* Retrieved 2019-03-08, from link

Sepulveda, C. E., Lopez Camacho, D., & Gallego, J. M. (2014). *Los límites de la estratificación: en busca de alternativas.* Retrieved 2019-05-02, from link

SUI, s. (2020). *Datos abiertos | superservicios - superintendencia de servicios públicos domiciliarios.* Retrieved 2020-09-23, from link

Tsivanidis, N. (2018). *The aggregate and distributional effects of urban transit infrastructure: Evidence from bogotá's transmilenio* (Tech. Rep.). Working Paper, University of Chicago Booth School of Business.

Turner, M. A., Haughwout, A., & van der Klaauw, W. (2014). Land use regulation and welfare. *Econometrica*, *82*(4), 1341-1403. Retrieved from link

Wong, V. C., Steiner, P. M., & Cook, T. D. (2013). *Analyzing Regression-Discontinuity Designs With Multiple Assignment Variables: A Comparative Study of Four Estimation Methods.* Retrieved 2019-03-28, from link