University of São Paulo
School of Economics, Business, Accounting and Actuary
Department of Economics

Patrick Parente Nasser

BACHELOR THESIS

Accessibility and Social Networks as conditions for the participation of vulnerable families in the "Programa Minha Casa Minha Vida" Brazil's Government Housing Program.

São Paulo

2019

Patrick Parente Nasser

BACHELOR THESIS

Accessibility and Social Networks as conditions for the participation of vulnerable families in the "Programa Minha Casa Minha Vida" Brazil's Government Housing Program.

Monograph presented to the Department Economics from the School of Economics, Business, Accounting and Actuary of the University of São Paulo as a partial requirement to obtain the title of Bachelor of Economics.

Major fields: Economics; Regional and Urban Economics.*

Advisor: Prof. Dr. Eduardo Amaral Haddad†

São Paulo

2019

* FEA/USP. E-mail:patrick.nasser@usp.br
† FEA/USP. E-mail:ehaddad@usp.br

Acknowledgements

Firstly I would like to thank my family, especially my mother, my friends and teachers for all the teachings, patience and joy over the last few years.

Special thanks to my advisor Prof. Dr. Eduardo Amaral Haddad for giving me the opportunity to research in recent months with the fantastic team of Nereus.

Finally, I would like to thank the SAGI staff for allowing me access to the necessary databases, São José do Rio Preto's geoprocessing department for providing me with the city street map. Last but not least, I thank the taxpayer to letting me study over the years in this wonderful University.

"When people thought the earth was flat, they were wrong. When people thought the earth was spherical, they were wrong. But if you think that thinking the earth is spherical is just as wrong as thinking the earth is flat, then your view is wronger than both of them put together."

Isaac Asimov

Abstract

We investigate how accessibility, measured by derivations of Hansen indexes, and social networks are related to the final decision of accepting the benefit by the poorest participants of Brazilian housing program "Programa Minha Casa Minha Vida" (PMCMV) in the city of São José do Rio Preto (São Paulo state). From the lists of municipal draws, Census, identified Unique Register, identified RAIS and information disclosed by state bank Caixa Econômica Federal, we estimate different specifications of Probit models with one and two stages. Our results indicate that worse accessibility prospects are related to non-compliance decision and we don't reject the null hypothesis for the effect of age-based social networks on compliance besides showing that neighborhoods might affect different age-compositions households differently.

Keywords: Urban and regional economics, Programa Minha Casa Minha Vida, Housing, Housing Demand, Residential Location;
JEL: D190, R210.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

1.1 What is it and how does the program work?

The "Programa Minha Casa Minha Vida" (PMCMV)[1] is the largest housing program of the Federal Government, created in 2009, and one of its objectives is to promote dignified housing for the poorest part of the population through subsidized housing loans.

The lines of credit destined to the residential financing of individuals have as criterion the family income, being divided in four main levels: the *Band 1*, destined to families with nowadays income up to 1,800 reais [2], counting on the greater subsidy of the Federal Government, and *Bands 1.5, 2 e 3*, destined to families with incomes of up to, respectively, 2,600, 4,000 and 7,000 reais[3].

The financing program of *Band 1* - the most socially sensitive line - is carried out in four different modalities: companies, entities, FGTS, municipalities with up to 50 thousand inhabitants and rural. All the resources of these modalities are transferred to the Federal Savings Bank (Caixa) by the budget of the Ministry of Cities. Only In 2017, PMCMV received R$10.2bi from Fundo de Garantia por Tempo de Serviço (FGTS) and distributed R$8.3bi in subsidies to 350 thousand families.[4].

Despite this, according to news published by newspaper *O Estado de São Paulo*[5], only 8% of PMCMV beneficiaries in the city of São Paulo belong to the poorest family segment of the program, that is with gross income up to R$1800.

[1] Translated as "My House My Life Program".

[2] The cash transfers Benefício de Prestação Continuada (BPC) and Bolsa Família, provided by the Federal Government, do not make up family income.

[3] http://www.caixa.gov.br/voce/habitacao/minha-casa-minha-vida/urbana/Paginas/default.aspx

[4] Available in http://www.caixa.gov.br/Downloads

[5] https://sao-paulo.estadao.com.br/noticias/geral,so-8-do-minha-casa-acolhe-faixa-mais-pobre,70002296499

The beneficiary who can join the program within *Band 1* is enrolled in the Cadastro Único (CadU) - Unique Register - of the Special Secretariat for Social Development (SEDS) from the Ministério da Cidadania (MC) - Misnistry of Citzenship - and has his / her information on public domain of the acquired property and Personal Taxplayer Registration (PTR) publicly disclosed by Caixa. You also have to pay Caixa a monthly amount for managing the credit lines, that varies between 80 and 270 reais, as a part of this subsidy line. In addition, the recipient cannot be an owner, transferee or promising buyer of residential property and must not have received housing benefit from city, State, Union, FAR, SDS budget resources or housing rebates granted with resources of the FGTS (Time Work State Found), with the exception of subsidies or discounts intended for the acquisition of construction material for the purpose of completing, expanding, renovating or improving housing units.[6].

In order to enroll in *Band 1* of the program, the applicant must register his/her housing demand with the city or local responsible body, which may use its own criteria in order to carry out the selection of the candidates, while respecting the norms imposed by the MC.

1.2 Natural Experiments: An Empirical Opportunity

The relative autonomy of the selection criteria by the municipalities allowed the creation of natural experiments in some cities that would participate in the PMCMV, mainly during its initial phases. Natural experiments, as already widely discussed in the literature of potential results models, are a quasi-experimental way of empirically testing the impact of a policy from counterfactuals that can be obtained by, as the name itself tells us, the nature of the experiment.

In order to illustrate this type of problem we will use an example of Angrist (1990). The author's study analyzes the effect of US veterans' participation in the 1970 Vietnam War in their future income years after the conflict. By then, a lottery based on dates of birth was drawn up to define the order of selection of recruits, so that the young men could be summoned and recruited more fairly. Thus, the order of the sequence of dates drawn was directly associated with the probability of recruiting the young people. Since it is not all those summoned who can participate in the conflict the randomization, provided by the lottery was used as an instrument for the participation of the combatant in the regression of incomes after the conflict. The result found by the author shows that the effect of having been eligible during the 1970s (in other

[6] Same source previously mentioned.

words, being over 19 years), under normal health conditions and having been raffled by the lottery, implies 15% lower wages even in 1980.

However, treating the lottery effect in a homogeneous way may not be realistic: some more nationalist men would in any case fight by volunteering themselves even if they had not been chosen by the lottery, while others would avoid going to war at all costs. Thus, the recruitment lottery instrument does not inform us the effect of eligibility under those who were not affected by the lottery. On the other hand, part of the veterans served exclusively because they were chosen in the draw. The men who had the behavior affected by the influence of the instrument "lottery" were called *compliers* Angrist et al. (1996).

In this case, the IV instrument of potential outcomes measures LATE - *Local Average Treatment Effects* - first proposed by Angrist and Imbens (1995), corresponds to the effect of recruiting on the salary of those who have become military only because they were drawn.

1.2.1 Internal and external validity

By the knowledge that LATE has validity only for a subpopulation of the sample, it is interesting to question the validity of this method in the econometric literature. In Imbens (2010), the author attempts to respond to criticisms in the literature that the excessive use of instrumental variables and randomization in the applied economics works of the last decade, which claim that "the disguise of the hypothesis of identification through statistical procedures has have been preferred over the honest discussion of the limits of the data", in other words, transforming what may be considered valid for some economists into an *ad hoc* source of demand for causal effects.

The main argument of Imbens (2010) and other well-known economists such as Abhijit V. Banerjee e Esther Duflo (from the development literature) comes from the fact that these experiments are powerful tools of internal validation and empirically credence the scientists' economic phenomena well modeled by theory.

It is worth mentioning that the randomization of a policy is aimed at calculating the so-called ATE - Averege Treatment Effect- , that future decisions about expansion or maintenance guidelines would be, in theory, empirically based on expected effects in a way, in theory, robust if randomization was well done and the result replicable.

A robust result of this nature would be extremely desirable from the point of view of decision-making by public agents and society as a whole, but in the real world, randomization of policies is not common either for ethical reasons (where there is no population and / or political feasibility to conduct such experiments), or by the costs of implementing this method. In this context, the analysis through LATE seems to suggest an alternative that may, in some cases, mitigate these problems making it a powerful tool for economic evaluation.

1.3 Estimation by IV and PMCMV

As already mentioned, there was, at least at the beginning of the PMCMV, enough freedom on the part of the municipalities to choose who would be the beneficiaries served by the program (since they were, of course, within the minimum criteria of the responsible ministry). Since the housing *defict* is much larger than the program's home supply, it was almost natural for the municipalities with many registered for the housing demand of *Band 1* to draw lots among the plaintiffs.

In his dissertation work, Rocha (2018) concludes that the PMCMV in the cities of Rio de Janeiro and São José do Rio Preto, both cities that fit the previous example, have a formal work reduction effect as well as an increase of beneficiaries of the Bolsa Família (Family Allowance) Program.

In order to get such results, Rocha (2018) makes use of an IV estimator, similarly to Angrist (1990)), using as instrument the ones drawn to participate in the program. The author examines the exogeneity of draws in Rio de Janeiro and São José do Rio Preto (from 2011 to 2015 in the case of Rio de Janeiro and 2013 in the case of São José do Rio Preto) and observed, when crossing the bases of Caixa Bank (containing the registered ones to receive the benefit of the program) with the RAIS (Annual Ratio of Social Information)[7] and the Unique Register[8] for the Individual Registry (CPF) that, when comparing the socioeconomic profile of those who were drawn and did not accept the benefit (*noncompliers*) with those who were drawn and accepted to participate in the program (*compliers*), on average these profiles are different, and the beneficiaries, besides all drawn families being poor, *noncompliers* appear to be more vulnerable. In addition, on average, respectively, 16% and 30% of those drawn from the program effectively turned beneficiaries.

[7] The complete and identified base one form the Ministry of Labour.

[8] The complete and identified base one form the Ministry of Citizenship.

1.4 *What about the non-compliers?*

It is not a very common rite in literature to study the *noncompliers* of a quasi-experiment and the reasons for this are quite reasonable: the great source of curiosity of the researcher usually resides in knowing if the experiment had any effect on the treaties so that it is possible to critically analyze the impacts of the implemented policy.

As the study of Rocha (2018) was assertively focused on the impact of the program on the labor market, little emphasis was placed on the comparison between *compliers* and *noncompliers*, which raises doubts about what could lead to a person[9] who, in theory, fulfills all the requirements of participation in the program and has a real housing demand (after all, one registered to be drawn in the PMCMV) to decide not to participate in the last analysis.

It is intuitive that the way in which *noncompliers* behavior is governed, in the sense that a person who initially showed interest in participating in a social program for the vulnerable eventually (for unknown reasons) did not participate in the same in face of this possibility (by being drawn), points us to a problem of preference realization that may have its roots both in poor socio-economic segmentation of those eligible for program benefit and in poor program implementation (with housing being built, for example, in worse places for the raffled person).

It is then evident the importance from the perspective of the application and elaboration of public policy to understand in depth the factors which may affect the participation decisions in order to improve its design making the process of sweepstakes and housing construction more efficient to supply the demand for housing[10].

1.5 *Objectives*

As said above,the descriptive statistics pointed by Rocha (2018) shows that the beneficiaries of the PMCMV and the ones that choose to not become part of them might have different socioeconomic profiles. The segmentation found among those could be used to better understand what is happening in the decision-making process of the selected agents. It is important from the

[9] The use of the words “person” or “individual” could freely be replaced by “family” along this section.

[10] It is also noteworthy that in the work on the PMCMV of Rocha (2018), unlike the limitation of Angrist (1990) on the inference that cannot be made on the eligibility of individuals not drawn, in the case presented about the São José City Hall this problem does not exist, as it has been checked whether the housing demand requirements of the families are met prior to the draw or not.

point of view of the application and elaboration of public policy to understand in depth how this difference is configured in order to improve the program design to make the process of selecting families and housing construction more efficient to meet the demand for housing.

Given this initial scenario, our main research question is: *Why do families choose noncompiance?*. We will take advantage on the fact that both cities (São José do Rio Petro and Rio de Janeiro) had used randomized draws in the selection processes, according to the exceptions made by the regulatory specifications imposed by the MC, with registered families in each list of municipal habitation demand to decide who would occupy the properties. We also chose only the municipality of São José do Rio Preto as our case study[11].

The first hypothesis we will test (hypothesis *i*) is that job accessibility is a determining factor from the point of view of the randomly selected families in shaping the choice of whether or not to participate in the program.

Rejecting the null hypothesis in this case could be a great explanatory evidence to corroborate the common literature argument that poor location of households in the program is a complicating factor for social policy: some families who need the benefit enough to enroll in the selection may consider at some point that the new location will put them into a worse situation and decide not to move to the program house Units (HU). If not rejected, it would contribute to the exploration of this aspect of the literature as a possible non-cause to the still barely explored phenomenon of non-participation.

There is, however, a possible endogeneity that deserves to be tested as hypothesis *ii* and may constitute a relevant factor for the decision making of families when deciding whether or not to leave their home: *social networks*.

Social networks are understood here as the ties between families and neighboring individuals identified by similar socioeconomic characteristics in nearby space units (such as proximity to people who share the same characteristics as age or migratory heritage) which are assumed to generate unobserved benefits (like friendships, labour market connectivity etc) and may constitute a determinant factor for the decision of families to leave or remain in the region.

In short, we will work in such a way as to test two hypotheses: *i*) Accessibility (in the locational sense) can be a determining factor for the choice of the individuals selected to participate or not in the program; and *ii*) *Social networks* (or social ties) may influence individuals' final decision whether or not to participate in the program.

[11] The reason is that CadU information is complete for this municipality.

This research will be presented as follows: in the next part, Chapter 2, we will present a brief review of the literature; then, in Chapter 3, the proposed methodological framework to be executed will be presented; next, Chapter 4 will detail our empirical considerations and descriptive statistics. Finally, in Charpter 5 we present our results followed by a critical reflection and summary of the work will be in the concluding Chapter 6.

Chapter 2

Literature review

2.1 *Accessibility*

2.1.1 *Housing Programs*

The effectiveness of PMCMV is commonly questioned in the literature by factors such as the location of HU.

In the article Lima Neto et al. (2015), the authors seek to answer, firstly, from a temporal and spatial analysis, if the PMCMV projects were being built far from the centers of metropolitan regions. Next, the authors check whether the PMCMV production region is spatially disconnected from the housing *deficit* region using the 2010 IBGE Demographic Census.

The authors analyzed the location of the metropolitan intra-urban housing *deficit* and the location of housing of social interest for the metropolitan regions of Belém (PA), Fortaleza (CE) and Belo Horizonte (MG).

Despite finding that during the program phases real estate projects tended to move away from the central regions, the authors cannot answer whether or not there is a link between production and the housing *deficit*, as the intentionality of the policy cannot be verified.

Accessibility is obviously not an exclusive Brazilian problem. The authors Apparicio and Séguin (2006) try to calculate the degree of accessibility to various services for residents of public housing developments in Montréal using distance measurements based *on road's networks* or *networks distance*, witch consist on the shortest possible physical distance that can be achieved to match the nearest establishment of interest that can be accessed based on the geographical structure of the city streets.

The services selected by the authors range from cultural (bookstores and movie theaters), educational (secondary and elementary schools), health services (hospitals and clinics), recreational services (parks, public pools and skating rinks), banking and others (supermarkets, police stations, subway etc.).

Thus, the authors were able to cluster the average differences in general accessibility and compare the inequality of accessibility of public housing in the city.

However, the methodology employed by the authors, as they themselves point out, is limited by some important factors, such as the assumption that a service, such as a park, is accessed by its nearest border (which is problematic because parks are very large), in addition to the assumption that people move exclusively on foot and preferentially access the nearest services offered.

We try to address this aspects of accessibility by creating metrics of distance from the original families households to the relevant central metropolitan area and the new program HU.

2.1.2 General aspects

Transportation costs also play a key role when discussing housing location issues. At work Pacheco et al. (2018) points out the city of São Paulo as a monocentric model from its center (which concentrates jobs). Thus, delimited the *Central Business District* (CBD) and, from the research of Origin and Destination (OD) of the Metro, the authors point out that about 44% of the trips which are made in the city pass or are concentrated in this region.

By using the National Family Budget Survey (POF) between 2003-2009 of the Brazilian Institute of Geography and Statistics (IBGE), the work done by Carvalho and Pereira (2012) shows that, on average, 15% of the family budget is formed by transportation costs and transportation time is, on average, 31% longer in the metropolitan regions of São Paulo and Rio de Janeiro.

When the sample is segmented by income, in the first 5 *per capita* family income deciles presented, the family commitment of transportation income ranges from 21.83% to 16.67%, and the ratio of private transport to public transport is higher for families in higher income deciles than for the poor, indicating that public transport is an inferior good in the urban environment. The author shows that, on average, there's a 20% difference in transport time spending between the richest and the poorest deciles between metropolitan regions, and São Paulo stands out for having a clear worsening in shift to all income deciles.

In addition, residents of Brazilian urban areas spend on average 5 times more on private transport than on public one, and there is an inequality between families living in inland and outlying metropolitan areas: the latter would be more dependent on public transportation, due to lower family income.

In a different approach to measuring accessibility, Vieira et al. (2014) create Hansen Indexes[1] for the metropolitan region of São Paulo based on distances calculated by the *Google Maps Directions* tool and the 2007 OD search.

The index prepared by the authors thus corresponded to the number of job openings in a given region, weighted by a function of time displacement for two modes of transport: public and private.

After clustering SPMR eligible regions (regions were chosen because at the time, the use of the Google Directions API was not considered, what limited the researchers' data collection range) by LISA statistics, the authors were able to compare and rank travel time by public and private transport throughout the metropolitan region.

We incorporate those aspects by using the Hansen indexes related to job market similarry to the ones developed by Vieira et al. (2014), time-weighted as poor families appear to be more sensitive to costs and transportation time as pointed by Carvalho and Pereira (2012), as a way to determine the city CBD used in Pacheco et al. (2018).

2.2 Peer Effects and Social Networks

2.2.1 Peer effects

Another aspect inherent in housing programs is the so-called *neighborhood effects* also known in the literature as social effects or peer effects. These neighborhood effects, as pointed out in Manski (2000), are a way for trying to explain stylized facts in which similar groups somehow tend to behave similarly.

A figurative example of this phenomenon is found in Schelling (1971). The author attempts to explain through intuitive and simplified demonstrations based on two spatial analytical models, one linear and one matrix, how white and black racial clusters are formed in the U.S.A., assuming simple rules of preferences and mobility of individuals within each group. However, as the author points out, predicting the collective impact of individual actions is not trivial.

By discussing the model specification problems that aim to identify the impact of social behaviors on individual behavior, Manski (1993) presents the problem of "reflection", that is an allusion to the attempt to interpret these models by comparing them with the observation of an agent and a mirror: if an observer does not know the difference between the optical object and

[1] See Hansen (1959)

the human, it becomes difficult to know if the human is the one who is imitated by the mirror reflection or if the human is the one who imitates the reflected image.

Thus, the author suggests analyzing the composition of the social effect of similar groups in three separate hypotheses.

The first one is the endogenous effect: if the effect of the individual behaving in a certain way varies according to the behavior of the group. The second one is the exogenous effect, that tests whether the behavior of an individual belonging to a group varies according to the exogenous characteristics of the group (the characteristics that contextualize the group). The last one is the correlational effect, whose idea is that the group behaves similarly because of similar individual or institutional characteristics.

Each of these effects has different implications from the public policy point of view. As an example, Manski (1993) asks us to consider a situation where a new teaching program is applied to some (but not all) students in one school.

If individual student performance increases along with the average increase in school performance (the social effect is endogenous in nature) this means that the expected effect of the program is to generate externalities which indirectly contribute to improved overall performance of all students in the school. However, this “social multiplier” would not be expected to occur if the dominant effect were exogenous or correlated.

An example of a natural experiment that allowed the study of neighborhood effects in housing involves the Moving to Opportunity (MTO) American program, which was a social experiment conducted by the Department of Housing and Urban Development (HUD) in the 1990s in which vouchers were distributed for rent payments to poor families.

In this experiment, that involved nearly 7,000 low-income families with children living in public housing developments that signed up for the program, the family sets were randomized into three groups. The first group was given a voucher and the use of each one was conditional on being spent on rent from neighborhoods with low poverty rates (less than 10% according to the local 1990 census). The second group was given the voucher without any conditions restricting the neighborhood to be filled. Finally, the last group was kept under observation as control of the experiment.

Katz et al. (2001) analyzed the impacts of the program in Boston on the peer effects that neighborhood changes had on the quality of life of families involved.

Although recognizing the difficulty of correctly measuring these peer effects, the authors could use the randomization of the experiment to isolate the *intention-to-treat* (IIT) as it did Rocha (2018), by using a potential results method based on *instrumental variables* (IV) by making use of randomization itself as an instrument for a two-stage regression that would instrumentalize the participation effect.

The authors thus concluded that the voucher made conditionally available to the treatment group was more effective in removing families from less marginalized regions while the non-conditional one was more effective in removing families from more calamitous neighborhoods.

Moreover, it was observed that both treatment groups obtained significant results regarding the improvement of the children's behavior quality, and other indicators such as accidents, chronic diseases and crimes of minors by the restricted voucher group. Overall gains were also observed in terms of overall quality of physical and mental health for adults and children of the residents of the treated households.

In contrast to the results obtained, the research was not able to identify changes caused by neighborhood change in the MTO that affected variables such as employment rate, income or other welfare gains caused by the change in the number of inhabitants in the household.

In another study, with the objective of estimating the effects of living in marginalized neighborhoods on unemployment, Dujardin and Goffette-Nagot (2005) made use of public housing as a proxy for the location preference of Lyon families with data available from the French Census.

By creating the proxy of preference, the authors are able to isolate the neighborhood effects that come from the correlational effect pointed out by Manski (1993), and can control the endogenous and contextual effect due to the locational choice for public housing.

The authors recall that methods of controlling neighborhood effects by instrumental variable (such as MTO) are rare and do not apply to the French case. Knowing that misusing tools to aggregate these effects can further increase bias, the authors opt for a methodological "third way", which consists of using aggregate statistics and spatial variance to try to access the importance of these social effects.

Thus, the authors chose to run three simultaneous *probits* relating to unemployment, neighborhood type and public housing accommodation, transforming the interest equation (of unemployment) into a linear reduction of the other two.

Finally, the authors conclude that living in housing developments has no direct effect on unemployment, but living in the 35% most marginalized regions may increase this probability.

2.2.2 Social networks

In addition to “neighborhood effects”, interpersonal hauls can contribute to the allocative decisions of individuals in urban space. To get around this problem, the literature has developed solutions that are directly expressed by the expression of “social ties”, commonly called social networks. In other words, besides not being a market interaction approach either, the focus of the analysis is not only on peer externalities but also on the structure that relates each individual (or unit of analysis) to each other.

Chapter 9 of Zenou (2009) explores different ways to measure the effects of peer effects and social networks on the labor market. The structure of interpersonal interactions of social networks, according to the author, can be defined by affinity ties among individuals, whether these are strong ties or week ones. The first type refers to direct connections between individuals, while the second to indirect connections.

To illustrate strong ties and week ties, we can base these on individuals A, B and C. Assuming that individuals A and C are friends of individual B only, and conversely, individual B is friend of individuals A and C, we can say that A, being a friend of B, has a strong tie with B and a week tie with individual C, since he is a direct friend of B but not of C, but is connected to C through B.

The author then uses different approaches to model the US white and black labor market, showing the difficulty of relating models that take into account individuals’ social networks with land market models expressed primarily by bid rent functions.

As an assumption of his approach, Zenou (2009) separates the population from his theoretical model into whites, blacks who interact with whites, and blacks who do not interact with whites. The three models pointed out by the author demonstrate how important social networks are in explaining the high unemployment rates present among social minorities, especially black Americans.

It is assumed in the three models that the labor market is separated for whites and blacks, with both ethnicities having the same productivity at work. In the first one, a bid rent function is built exploiting only the strong ties for the three types of individuals. In the second one, it is adopted a premise that allows the exploitation of week ties between individuals (which is the most important information factor to help them to get a job), but resulting in the loss of their ability to represent the land market through a bid rent function. Finally, in the third model, based on Calvó-Armengol and Zenou (2005), we can explore the relationships described by week ties to the fullest based on a model described by graphs in which there is no explicit land market.

In Patacchini and Zenou (2012), the authors examine the relationship between residential proximity of individuals in a single ethnic group and the likelihood of finding a job through social networks in the United Kingdom.

In order to reconcile a social network model with locational properties, the authors assume that social proximity among agents, by the way, social networks, can be approximated by geographical distance, so that the density of social network connections can be well represented by the density of the presence of a particular ethnic group in a locality. This premise, besides being based on the literature, is well justified by the authors in:

> We believe that this approximation makes sense because the two spaces (social and geographic) are highly correlated. For example, individuals in established immigrant communities typically provide information, seed capital, shelter, and legal sponsorship to other immigrants from the same origin communities (family or friends). In this way, not only employment outcomes but also geographic origin is shared by immigrants in the same national group living in proximity (e.g. Mexicans from a particular town in rural Michoacan may tend to live in the same neighborhood in East Los Angeles). This is the idea of ethnic enclaves that have positive (employment) effects on the local ethnic community

Thus, the authors use the network analysis framework elaborated by Calvó-Armengol and Zenou (2005) and come to the conclusion that the higher the percentage of the same ethnic group of an individual who lives near it, the greater the chances of one finding employment through (social) networks. Moreover, this effect rapidly declines with the distance from the social group and is not homogeneous among all ethnic groups studied.

We will incorporate Patacchini and Zenou (2012) approach to defining social ties within 500m family radius circular bands.

Chapter 3

Methodology

3.1 Accessibility

Based on Vieira et al. (2014), we will elaborate accessibility indexes based on Hansen's index:

$$A_i = \sum_j W_j/F(c_{ij}) \tag{3.1}$$

Where A_i the accessibility of region i, W_j is the number of opportunities W of destination j and $F(c_{ij})$ the impedance function (when i equals j, F is unitary).

The elaboration of the origin and destination matrix for our regions of interest will be made using the Google Maps API (which will eliminate the need of searching distances individually on the commercial platform) and in addition it would allow the impedance function of our accessibility indicators to be the travel time from each point to each destination.

To create a labor market accessibility index we will use RAIS[1], which gives the location of enterprises and the volume of labor, that is the number of W (opportunities). We can thus organize the number of jobs according to spatial granularity we want, obtaining W_j.

Next, we will use the fact that we are working with a randomized draw (the same draws worked on Rocha (2018)) to estimate an equation that gives us the effect that accessibility has on individuals' decision whether or not to participate in the program. Our ideal Probit model would be:

$$C_i = constant + \beta_1 A_i^{Diff} + \beta_3 dist_i^{CBD} + \beta_4 time_i^{CBD} + \beta_5 dist_i^{HU} + \beta_6 time_i^{HU} + controls_i + \epsilon_i \tag{3.2}$$

[1] The complete and identified base one form the Ministry of Labour.

With:

$$A_i^{Diff} = \sum_j (A_j^e - A_{ij}^m) \tag{3.3}$$

Where C_i is a *dummy* indicating whether the drawn family has become benefited. A_i^{Diff} is defined in Equation 3.3 as our accessibility index for the difference between the PMCMV Hansen index, defined by $\sum_j A_j^e$ (the sum of accessibility's from the program HU to all j locations), and the Hansen index for the origin household residence, defined by $\sum_j A_{ij}^m$ (the sum of accessibility from the origin drawn household i to all j locations). $dist_i$ and $time_i$ are, respectively, the Euclidean distance and travel time of the original household residence until the draw date to the city CBD and the PMCMV HU.

The households variables will be obtained from the base crossing that contains the CPF of the registered people. We will use the Single Registry (which will be requested from the Ministry of Citizenship), the information of those who became beneficiaries of *Band 1* with public places available by Caixa Bank and the raffles in the periods of interest to the municipalities of interest available on their related websites.

So, we can obtain socioeconomic information by individual/family who participated or not in the program, including aspects such as address, ethnicity, household information, income, participation in the *Bolsa Família* Program etc.

Unfortunately, Google's API are limited expensive and this research budget constraint is tight[2]. We can't afford all requests we need to calculate distances and travel time, so we came up with some simplifications.

The entire city, which its approximately 432km^2, was then divided into a grid containing 445 squares of dimensions approximately 1 km x 1 km (except for the borders). Although we are able to identify the exact households and firms addresses, we will make use of the square's centroids to calculate relative distances and travel time, especially in Hansen's Indexes. But despite not being the most precise approach, and considering that the city is small, we can approximate it's geometry to a plane in a Euclidian space, which make most of our estimations unbiased and consistent.

This 'budget' problem is more sensitive when dealing with the Hanssen's Indexes calculations. More specifically, any type of Hanssen's Indexes as in equation 3.1 can be summarized in a big matrix, as showed in 3.4:

[2] Actually we have only 'constraints' and no 'budget' at all.

$$
H = \begin{array}{c} \\ i_1 \\ \vdots \\ i_n \end{array}
\begin{array}{c}
\begin{array}{ccc} j_1 & \cdots & j_n \end{array} \\
\begin{pmatrix}
W_1 & \dots & W_{j_n}/F(c_{i_1 j_n}) \\
\vdots & \ddots & \vdots \\
W_{j_1}/F(c_{i_n j_1}) & \dots & W_n
\end{pmatrix}
\end{array}
\tag{3.4}
$$

As the number of impedance functions requests in Google's API must be, approximately, the number of the matrix elements (445×445), which is far beyond our constraint, we decided to pick up some randomized elements of our matrix and run an OLS regression to determine estimated impedance functions, which in our case will be given by travel time by car (accounting traffic information).

$$
TravelTime = constant + \beta_1 W_i + \beta_2 W_j + \beta_3 distance_{ij} + \beta_4 distance_{CBD} \tag{3.5}
$$

Where W_i and W_j are respectively the number of opportunities in the origin and destination centroids and, finally, $distance_{ij}$ and $distance_{CBD}$ are respectively the euclidian distance between regions i to j and i to the CBD.

3.2 Social Networks

In order to control endogenous social networking effects, which may justify the preference of certain individuals for their original location and therefore lead to refusal to participate in the program if drawn, we will typify our entire sample into distinct socioeconomic profiles and, under the same criteria, we will do the same typification with IBGE 2010 Census data for the analysis region.

We will be inspired by the methodologies employed in Zenou (2009) , Calvó-Armengol and Zenou (2005), and Patacchini and Zenou (2012), thus defining a social network based on graphs that will be related, as they did Patacchini and Zenou (2012), by the approximation between the social and geographical proximity of families/individuals with similar socioeconomic characteristics.

Our criterion of social network size will be according the density of individuals with similar characteristics among themselves in 500 meters radius. The characteristics that we will analyze will be the belonging to age groups, as it is the most granulated relevant information that we can obtain from Brazilian Census tract since 2000.

As the choice of housing for individuals is, in theory, endogenous, we will instrumentalize the age groups densities in each housing location from the 2000 Census over the 2010 Census. Once the *land market* evolves over time, our assumption is that the pattern that influenced housing choices for each age group in 2000 is independent of the 2010 pattern apart from their effect through the present-day age group density variable. Patacchini and Zenou (2012) did something similar with ethnic population densities in the labor market.

The density estimation will thus consist of the first stage of our regression, and the second stage will be the equation presented in the previous subsection, bringing our model closer to the LATE estimator used in Angrist and Imbens (1995). The densities for each age group will thus represent the social network effect associated with the neighborhood as we can approximate geographical distance into social ties.

Chapter 4

Empirical Considerations and Descriptive Statistics

4.1 Empirical Considerations

4.1.1 Matching different databases using unique CPF as key

As described in the methodology section, we used unique CPF key from the individuals enrolled in the "program lottery" in order to converge the databases from the chosen city (for subscribed and selected individuals), Caixa Federal Bank (for program compliers) and the Unique Register (with socioeconomic details about families and individuals who subscribed to the lottery).

In order to overthrow compatibilization difficulties concerning conflicting and persistent typos on CPF keys from the databases, we made use of a Levenshtein distance algorithm[1] that proved to be a great approach, as we were able to match 2,482 selected individuals from 2,513 ones. We have found 1,749 compliers: a slightly different number from Rocha (2018).

4.1.2 Matching the addresses

São José do Rio Preto Georeferecing Office kindly gave us the city road-map in order to help us to confirm the firms and families addresses precisely.

As described before, all addresses firms and households were geocoded using Google Geocoding API. The firms database was geocoded following a hybrid approach between Campos (2018) and the Levenshtein algorithm with Google's API results.

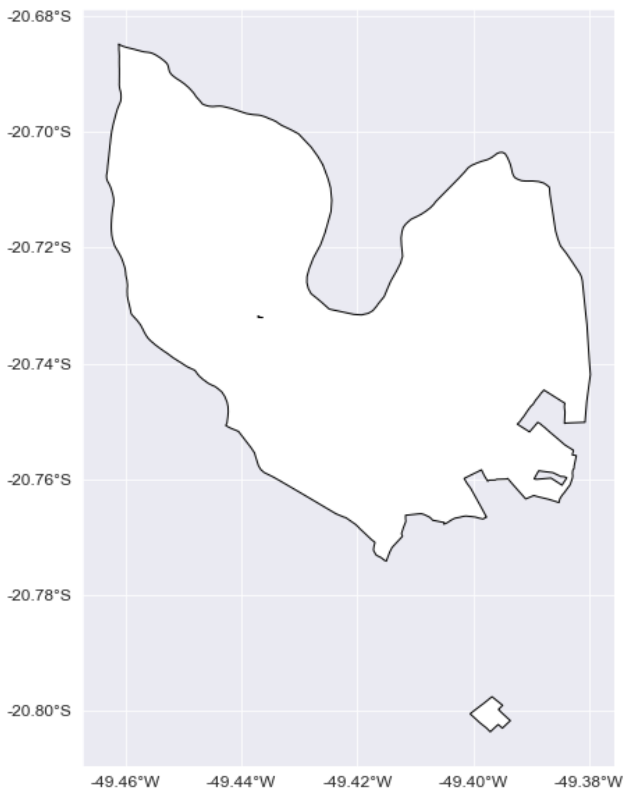
[1] Levenshtein distance is a matrix based method to measure the difference between two string data-types (commonly called "text") while looking, most importantly, to the order in which each individual character appears.

4.1.3 Minimum Comparable Areas

In order to instrumentalize variables from 2000 Census tract over Census 2010 tract we need to build the Minimum Comparable Areas (MCA) between each tract.

Unfortunately it is impossible to reconstruct each 2010 Census tract back to 2000 without creating discontinuities between tracts, in other words, it is impossible recreating the MCA without creating methodological mistakes as "islands" for the same 2000's Census tract due to IBGE tract creation approach, as illustrated for one Census tract in Figure 4.1.

Figure 4.1: MCA Census tract island

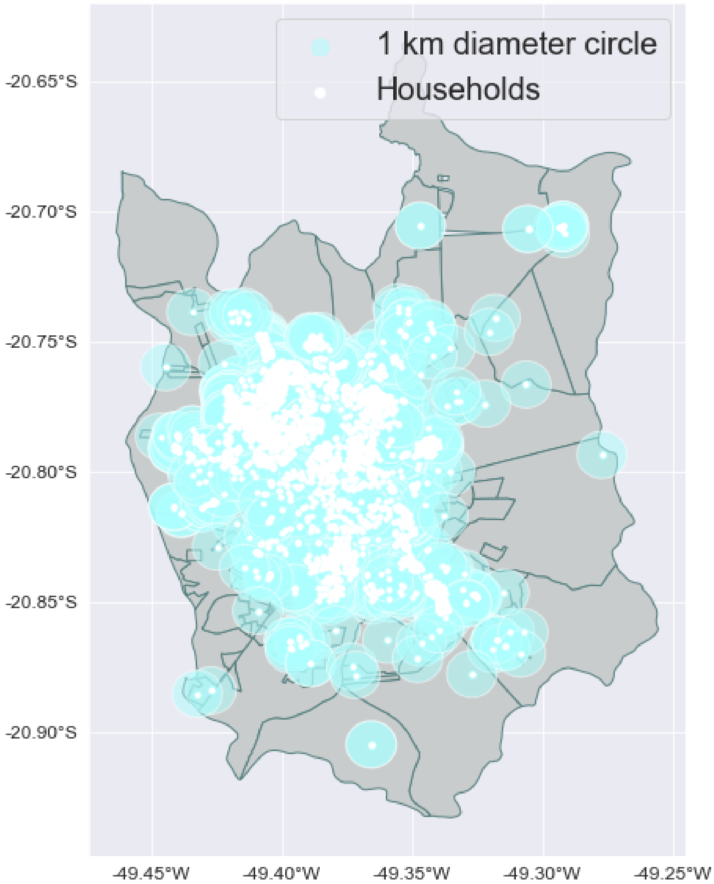


Source: author's elaboration; data from IBGE.
Note: MCA obtained following IBGE's tract movement log.

Because of this we have given up calculating the MCA based on the original Census tracts. Instead we use the areas defined by the 500-meter radius around each household to define the variables to be instrumentalized.

We assume that each variable is homogeneously distributed within each sector[2], its values corresponding to the proportion of the area of each polygon within the area of each circle, as illustrated in Figure 4.2. In other words, if a sector has 1000 inhabitants and 10% of the polygon area of that sector intersects a given circle, this circle will have at least 100 inhabitants (since the other polygon intersecting inhabitants are added).

Figure 4.2: Households circles over 2010's Census tract polygons



Source: author's elaboration; data from IBGE and CadU.

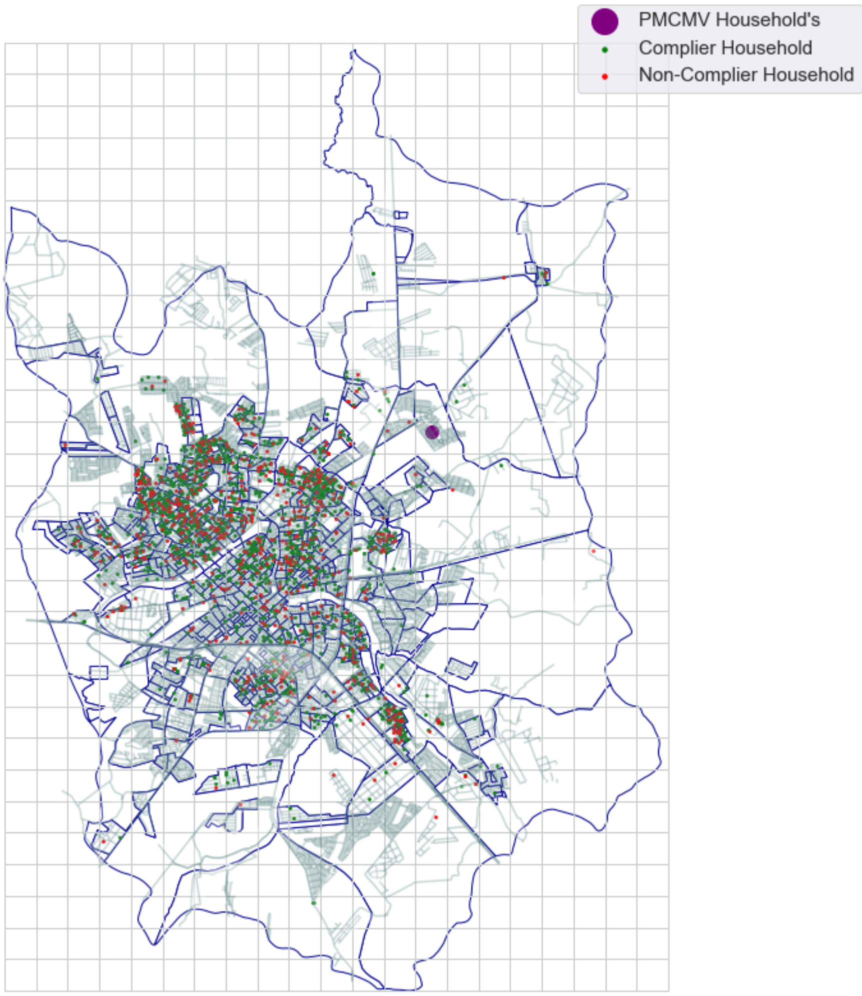
[2] Census tract

4.2 Descriptive Statistics

4.2.1 First analyzes

Households are shown in Figure 4.3, where we can see Compliers and non-Compliers along the city. The purple big dot is an approximation centroid from the PMCMV's Households complex. The light grey lines form the city roadmap. Blue lines represent the city 2010's Censu's tract sections. The grey checkered grid overlapping the map will be used to cut the city into 1 km^2 areas for the following analyses.

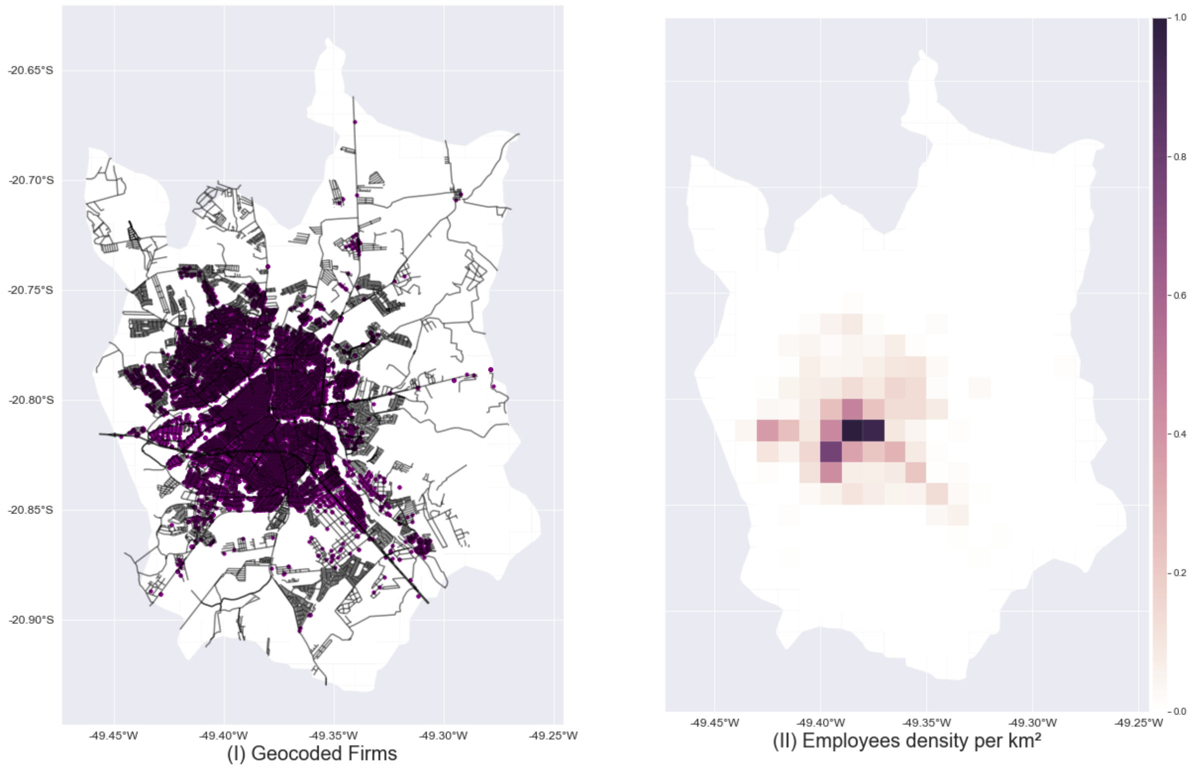
Figure 4.3: Compliers and non-Compliers



Source: author's elaboration; data from IBGE and CadU.

In Figure 4.4 first panel we can see all the geocoded firms found in RAIS in 2013. We were able to find about 32 thousand unique firms addresses in RAIS. Figure 4.4 shows us the distribution of the absolute number of employees reported by those firms in each km^2.

Figure 4.4: Firms and Employees per square kilometer



Source: author's elaboration; data from IBGE and RAIS.

Dealing with such a small town has its empirical advantages and disadvantages. Although we have already listed several disadvantages, clearly one of the biggest facilities can be shown by trying to calculate the city's CBD.

There are lots of approaches to calculate one citys CBD using kernel density estimation methods. All those methods can be highly precise depending on the quality of the data used, but are unnecessary in the case illustrated in this research panel *II* in Figure 4.4. It's not difficult at all to recognize that São José's CDB centroid is settled in the left darker square.

4.2.2 Data

We runned the OLS represented by the Equation 3.5 using the absolute number of reported employees as a variable for the number of opportunities and the Euclidian distances between the centroids (distance matrix) calculated by QGis. As we expected, we obtained a satisfactory (yet not the most precise) estimation based only on correlational aspects, illustrated by an adjusted

Table 4.1 - Travel Time - OLS model

	I
const	541.9127***
	(30.5393)
wi	-0.0248**
	(0.0096)
wj	-0.0493***
	(0.0090)
d	0.0950***
	(0.0017)
d_cbd	0.0213***
	(0.0027)
Observations	1993
R^2	0.68

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: units are measured in seconds and meters.

R^2 of 0.68 in Table 4.1.

It is worth pointing out (but not reporting on our results) that we strive to increase the obtained R^2. In order to do so we made use of a shortest path routing algorithm to try to calculate a distance matrix with real route distance approximation. We tried using both our given São José's Road Map and a road map obtained using the open source platform *OpenStreetMap*. Despite our best efforts and our obtained variable being significant to 1%, the found variable wasn't able to increase the predictability of the OLS model[3]. Due to results similar to those found in the specification of the Equation 3.5, we chose to report only the latter, as it is simpler and easier to reproduce.

Table 4.2, made with data provided from both 2010 Census trac and CadU, shows us the slightly differences between *compliers* and *noncompliers* households. Most notably, on average, *noncompliers* are richer (more per capta income among the poor), pay more rent, have more house-members among the adulthood and middle age groups and more dorm-rooms in their houses than *compliers*, while these last ones have less people home, proportionally more women,

[3] We think that it happened because the generated networks have no directions embedded.

and receive more government assistance (denote by “Programa Bolsa Família”) and, finally, live in more populated areas.

Table 4.2 - Difference in means

	noncompliers(1)	Mean(1)	compliers(2)	Mean(2)	MeanDiff(1 - 2)
A^{Diff}	723	-8200	1740	-8000	-253.4
$dist^{CBD}$	723	5815	1740	5780	35.40
$time^{CBD}$	723	873.3	1740	876.4	-3.087
$dist^{HU}$	723	10000	1740	10000	95.30
$time^{HU}$	723	1119	1740	1110	9.246
ln per capta income	723	6.016	1740	5.874	0.142***
dummy Bolsa Familia	723	0.123	1740	0.160	-0.037**
nº people house	722	2.896	1738	2.736	0.160**
piped water	723	0.985	1740	0.982	0.00300
house rooms	718	4.543	1733	4.475	0.0680
house dorm-rooms	718	1.724	1733	1.668	0.057*
ln expenditure rentals	722	3.673	1738	3.265	0.408***
dummy bathroom	723	0.993	1740	0.995	-0.00200
sewerage system	723	0.942	1740	0.944	-0.00200
garbage collection	723	0.992	1740	0.991	0
nº people CadU	723	2.490	1740	2.125	0.364***
nº of women	723	1.308	1740	1.264	0.0440
women %	723	0.516	1740	0.599	-0.083***
average age	723	37.65	1740	37.36	0.290
ages under 17	723	0.512	1740	0.570	-0.0580
ages up 18 under 35	723	0.909	1740	0.699	0.210***
ages up 36 under 60	723	0.882	1740	0.657	0.226***
ages up 61	723	0.187	1740	0.199	-0.0130
private household	723	0.993	1740	0.996	-0.00300
Circle population 2010	723	14000	1740	15000	-712.496**
density under 17	723	0.234	1740	0.235	-0.00200
density up 18 under 35	723	0.311	1740	0.311	0
density up 36 under 60	723	0.333	1740	0.333	0
density up 61	723	0.123	1740	0.121	0.00100

[1]Note: distances are measured in meters and time in minutes.

[2]Note: variables starting with "density" correspond to the ratio of the corresponding age group among circle area population.

Chapter 5

Results

5.1 Accessibility

Based on Equation 3.2, trying to isolate accessibility aspects, we made three model specifications, as showed in Table 5.1. Firstly, in model I, we employ a simple Probit regression to see if we are able to find any correlation between our interest variables and the probability of household compliance. We were unable to found any statically significant result for both model I and, secondly, model II, which contained control variables to household quality.

We suspect that this result is due to the lack of controls for neighborhood quality, so we assign dummy variables representing the belonging to each 1 km^2 square that we displayed over the city, illustrated in Figure 4.3, clustering the errors in these same assigned dummy variables.

As a result we find that willingness to complie is associated with positive expectations on accessibility to job market gains and inversely related to the number of people living in the house, income per capta and household rent expenditure.

Also, we tried to see this effect among different households per capta incomes percentiles, but choose to not report the results due to lack of observations. You can see then in Appendix A Table A.

Table 5.1 - Compliers - accessibility model

	I	II	III
A^{Diff}	2.69e-06	2.13e-06	0.00364***
	(4.40e-06)	(4.57e-06)	(0.000710)
$dist^{CBD}$	-5.34e-05	-5.89e-05	4.97e-07
	(3.60e-05)	(3.60e-05)	(0.000108)
$time^{CBD}$	0.000422	0.000415	0.000504
	(0.000373)	(0.000376)	(0.000879)
$dist^{HU}$	1.29e-05	1.45e-05	3.45e-05
	(3.54e-05)	(3.65e-05)	(0.000107)
$time^{HU}$	-0.000124	-0.000138	-0.00116
	(0.000342)	(0.000350)	(0.000981)
ln per capta income		-0.0909**	-0.0918**
		(0.0359)	(0.0360)
dummy Bolsa Familia		0.0454	0.0269
		(0.102)	(0.0884)
n^{o} people house		-0.0575**	-0.0569**
		(0.0232)	(0.0260)
piped water		-0.302	-0.307
		(0.274)	(0.292)
house rooms		0.00353	0.0101
		(0.0213)	(0.0199)
house dorm-rooms		-0.0534	-0.0764*
		(0.0485)	(0.0435)
ln expenditure rentals		-0.0389***	-0.0356***
		(0.00994)	(0.0106)
dummy bathroom		-	-
sewerage system		0.0841	0.103
		(0.138)	(0.179)
garbage collection		-0.693	-0.716
		(0.562)	(0.520)
private household		-	-
Constant	0.512**	2.360***	-5.651**
	(0.253)	(0.697)	(2.474)
Observations	2,463	2,450	2,388
R^2	0.0011	0.0142	0.0387
Dummy for each region			YES
Cluster by region			YES

Robust standard errors in parentheses

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

5.2 Social Networks

To discriminate our Social Network we keep the main Equation 3.2, specification adding control groups and neighborhood age-group density, that correspond to the ratio between the number of people in a certain age group around the 1km^2 diameter circle within the household at it's center (as illustrated in Figure 4.2), these densities were instrumentalized by corresponding 2000's Census densities and "circle population".

The densities corresponding to each age group tie would thus consist in social network effect. We thus divide our sample in four not mutually exclusive groups, corresponding to the presence of the same age-bands for each household. Then we test two slightly different model specifications: in Table 5.2 we retain households that have at least one person in the assigned age-group; in Table 5.3 we reattain only the households that have at least the same density ratio with the neighborhood for the age group[1].

Our results in both Tables 5.2 and 5.3 were not able to reject the null hypothesis for the effects of social networks in all interest groups. This result, whatever, might be strong related to the lack of available observations, as even accessibility effect disappeared due to much higher standard errors.

Besides that data limitation, we were able to show that age-groups might behave differently according to neighborhood characteristics, as the elderly seams to avoid more populated areas, on the opposite side of young adults. Also, the number of people in the house, expenditure on rentals and per capta income appears to have greater anchoring relations only among younger age tiers composition. Finally, we find that households with relative more children living in far Euclidian distances from the CBD are less willingness to complie.

[1] Note that last column for each table is the same, because the elderly are less representative among the population.

Table 5.2 - IVProbits for households with at least one representative age group member

	under 17	up 18 under 35	up 36 under 60	up 61
A^{Diff}	0.000537	-0.000531	0.000466	-0.000236
	(0.000534)	(0.000394)	(0.000590)	(0.00154)
$dist^{CBD}$	-0.000649**	-9.01e-05	-0.000225	-0.000741
	(0.000321)	(0.000171)	(0.000212)	(0.000662)
$time^{CBD}$	0.00178	0.000902	-1.91e-05	0.00431
	(0.00173)	(0.00137)	(0.00138)	(0.00374)
$dist^{HU}$	-0.000164	-3.88e-05	3.01e-05	7.04e-05
	(0.000208)	(0.000154)	(0.000162)	(0.000398)
$time^{HU}$	-0.000260	-0.00137	-0.00170	0.000289
	(0.00155)	(0.00128)	(0.00136)	(0.00252)
ln per capta income	-0.174**	-0.151***	-0.00969	-0.151
	(0.0799)	(0.0520)	(0.0418)	(0.132)
dummie Bolsa Familia	-0.0258	-0.105	0.0360	-0.302
	(0.134)	(0.115)	(0.0934)	(0.716)
nº people house	-0.163***	-0.102***	-0.0618*	-0.130
	(0.0463)	(0.0308)	(0.0367)	(0.108)
ln expenditure rentals	0.00343	-0.0269**	-0.0633***	-0.0720*
	(0.0163)	(0.0135)	(0.0140)	(0.0404)
women	(0.249)	(0.133)	(0.107)	(0.239)
average age	-0.0474***	-0.0264***	-0.00467	0.00760
	(0.00871)	(0.00533)	(0.00485)	(0.00980)
density under 17	26.11	6.064	17.59	
	(16.07)	(8.840)	(11.45)	
density up 18 under 35	-3.536	8.303	3.694	-100.2
	(15.40)	(9.908)	(12.16)	(87.30)
density up 36 under 60	-0.356	-11.67	-8.877	-47.08
	(15.37)	(10.39)	(12.78)	(52.33)
Circle population 2010	-4.94e-06	4.83e-05***	-2.79e-05	0.000172***
	(2.80e-05)	(1.66e-05)	(2.56e-05)	(6.27e-05)
density up 61				-35.79
				(42.68)
Constant	4.738	4.143	1.866	49.48
	(7.250)	(4.732)	(5.559)	(40.72)
Observations	873	1,354	1,394	319
Dummy for each region	YES	YES	YES	YES
Cluster by region	YES	YES	YES	YES
Accessibility controls	YES	YES	YES	YES

Robust standard errors in parentheses

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

Table 5.3 - IVProbits for households age groups denser than in the neighborhood

	under 17	up 18 under 35	up 36 under 60	up 61
A^{Diff}	0.000487	0.00110	0.000622	-0.000236
	(0.000526)	(0.000711)	(0.000677)	(0.00154)
$dist^{CBD}$	-0.000598*	-9.74e-05	-0.000217	-0.000741
	(0.000320)	(0.000186)	(0.000224)	(0.000662)
$time^{CBD}$	0.00122	0.000602	-0.000342	0.00431
	(0.00173)	(0.00146)	(0.00150)	(0.00374)
$dist^{HU}$	-0.000202	-9.90e-05	9.31e-05	7.04e-05
	(0.000192)	(0.000162)	(0.000178)	(0.000398)
$time^{HU}$	-4.87e-05	-0.000561	-0.00190	0.000289
	(0.00150)	(0.00131)	(0.00154)	(0.00252)
ln per capta income	-0.159**	-0.137**	0.0184	-0.151
	(0.0787)	(0.0560)	(0.0470)	(0.132)
dummie Bolsa Familia	-0.00422	-0.181	0.0243	-0.302
	(0.139)	(0.134)	(0.169)	(0.716)
n° people house	-0.158***	-0.108***	-0.0701*	-0.130
	(0.0494)	(0.0351)	(0.0383)	(0.108)
ln expenditure rentals	-0.000379	-0.0214	-0.0668***	-0.0720*
	(0.0178)	(0.0147)	(0.0177)	(0.0404)
women	(0.252)	(0.141)	(0.116)	(0.239)
average age	-0.0436***	-0.0288***	-0.00602	0.00760
	(0.00940)	(0.00550)	(0.00495)	(0.00980)
density under 17	26.99*	7.760	22.96*	
	(15.53)	(10.39)	(12.92)	
density up 18 under 35	-6.270	11.08	5.253	-100.2
	(14.88)	(13.54)	(12.95)	(87.30)
density up 36 under 60	-4.775	-12.82	-1.458	-47.08
	(14.85)	(11.62)	(15.44)	(52.33)
Circle population 2010	-1.58e-06	4.48e-05***	-2.80e-05	0.000172***
	(3.18e-05)	(1.69e-05)	(2.74e-05)	(6.27e-05)
density up 61				-35.79
				(42.68)
Constant	6.976	6.001	-2.648	49.48
	(7.319)	(4.986)	(6.854)	(40.72)
Observations	813	1,168	1,118	319
Dummy for each region	YES	YES	YES	YES
Cluster by region	YES	YES	YES	YES
Accessibility controls	YES	YES	YES	YES

Robust standard errors in parentheses

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

Chapter 6

Discussion and Conclusion

It's important to mention that our empirical approach was being limited by factors such as the number of observations, street map quality, Census compatibility and absence of HU spatial variation, however this research can also be applied in cities with a higher offer of public data, just like Rocha (2018) did with Rio de Janeiro.

We showed on Table 5.1 that the difference in accessibility to the workplace marked (which is great proxy for general accessibility to services) from the original household to the HU has a positive influence on the decision of participating in the housing program. This result dialogues with the common argue that accessibility is an important topic to families willing to move out from poor housing conditions.

This raises a special concern from policy perspective when it brings back, as pointed out by Lima Neto et al. (2015), that PMCMV HU have a tendency to be build far from the metropolitan centers over time. New forms of incentives for housing, as the conditional vouchers described by Katz et al. (2001), may draw the path through increasing welfare.

Although we were not able to determine social networks effects, we were able to observe the correlational effects described by Manski (1993), as different households compositions seem to respond differently to differently neighborhood variables.

It's reasonable to assume that different housing compositions from all types of perspective (race, age, positioning of family members, labour positioning of family members, family members birthplace etc) might demand different types of accessibility, for example, kids would demand school proximity, and the elderly would demand proximity to public health units.

We were not able to test that due to the lack of research time and data in Census tract data[1], but future research could use Vieira et al. (2014) approach on Hansen's indexes to explore this

[1] That could be overcome by using Cesus microdata, but it would only make sense in a larger city.

kind of social effect.

Finally, the PMCMV is an expensive program that aims to help vulnerable families, and in doing so, it must account it's inability to reach the poor by offering worse accessibility prospects.

Bibliography

Angrist J., Imbens G., , 1995 Identification and estimation of local average treatment effects

Angrist J. D., Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records, The American Economic Review, 1990, pp 313–336

Angrist J. D., Imbens G. W., Rubin D. B., Identification of causal effects using instrumental variables, Journal of the American statistical Association, 1996, vol. 91, p. 444

Apparicio P., Séguin A.-M., Measuring the accessibility of services and facilities for residents of public housing in Montreal, Urban studies, 2006, vol. 43, p. 187

Calvó-Armengol A., Zenou Y., Job matching, social network and word-of-mouth communication, Journal of urban economics, 2005, vol. 57, p. 500

Campos R. B. A., Subcentralidades e e prêmio salarial intra-urbano na região metropolitana de São Paulo, Universidade de São Paulo - Faculdade de Economia, Administração e Contabilidade [Departamento de Economia: Programa de Pós graduação em Economia], 2018, Ph.D. Thesis, 230 p.

Carvalho C. H. R., Pereira R. H. M., Gastos das famílias brasileiras com transporte urbano público e privado no Brasil: uma análise da POF 2003 e 2009, Instituto de Pesquisa Econômica Aplicada (Ipea), 2012

Dujardin C., Goffette-Nagot F., Neighborhood effects, public housing and unemployment in France, 2005

Hansen W. G., How accessibility shapes land use, Journal of the American Institute of planners, 1959, vol. 25, p. 73

Imbens G. W., Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009), Journal of Economic literature, 2010, vol. 48, p. 399

Katz L. F., Kling J. R., Liebman J. B., Moving to opportunity in Boston: Early results of a randomized mobility experiment, The Quarterly Journal of Economics, 2001, vol. 116, p. 607

Lima Neto V. C., Krause C., Furtado B. A., O déficit habitacional intrametropolitano e a localização de empreendimentos do Programa Minha Casa Minha Vida: mensurando possibilidades de atendimento, Instituto de Pesquisa Econômica Aplicada (Ipea), 2015

Manski C. F., Identification of endogenous social effects: The reflection problem, The review of economic studies, 1993, vol. 60, p. 531

Manski C. F., Economic analysis of social interactions, Journal of economic perspectives, 2000, vol. 14, p. 115

Pacheco T. S., Chagas A. L. S., et al., Demanda por transporte na Região Metropolitana de São Paulo e política de pedágio urbano para redução de congestionamento. In Anais do XLIV Encontro Nacional de Economia [Proceedings of the 44th Brazilian Economics Meeting] , No. 154, 2018

Patacchini E., Zenou Y., Ethnic networks and employment outcomes, Regional Science and Urban Economics, 2012, vol. 42, p. 938

Rocha G. M. S., Política habitacional e a oferta de trabalho: Evidências de sorteios do Minha Casa Minha Vida, Universidade de São Paulo - Faculdade de Economia, Administração e Contabilidade [Departamento de Economia: Programa de Pós graduação em Economia], 2018

Schelling T. C., Dynamic models of segregation, Journal of mathematical sociology, 1971, vol. 1, p. 143

Vieira R. S., Haddad E. A., et al., Índice de acessibilidade para São Paulo, XL Encontro Nacional de Economia. ANPEC-Associação Nacional dos Centros de Pós-graduação em Economia. Anais, 2014, vol. 167

Zenou Y., Urban labor economics. Cambridge University Press, 2009, 376

Appendix

Appendix A

Appendix A

Table A.1 - Accessibility among different household per capta income

	20% poorest	50% poorest	50% richer	20% richer	all
A^{Diff}	-0.000427	-0.000239	-1.75e-05	-0.000748	0.00364***
	(0.000382)	(0.000275)	(0.000319)	(0.000690)	(0.000710)
$dist^{CBD}$	-8.39e-05	-0.000180	-2.33e-06	-0.000275	4.97e-07
	(0.000369)	(0.000194)	(0.000133)	(0.000309)	(0.000108)
$time^{CBD}$	0.00251	0.00169	0.000342	0.00217	0.000504
	(0.00316)	(0.00141)	(0.00116)	(0.00277)	(0.000879)
$dist^{HU}$	-3.64e-05	0.000125	-1.34e-05	0.000696**	3.45e-05
	(0.000249)	(0.000160)	(0.000107)	(0.000293)	(0.000107)
$time^{HU}$	-0.000683	-0.00197	-0.000255	-0.00825***	-0.00116
	(0.00256)	(0.00130)	(0.00116)	(0.00271)	(0.000981)
ln per capta income		-0.0997**	0.00380	-1.575***	-0.0918**
		(0.0496)	(0.116)	(0.489)	(0.0360)
dummie Bolsa Familia	0.0175	0.0650			0.0269
	(0.165)	(0.0937)			(0.0884)
nº people house	-0.0528	-0.0923**	-0.0410	0.0252	-0.0569**
	(0.0569)	(0.0366)	(0.0308)	(0.0684)	(0.0260)
piped water	-0.368	-0.189	-0.374		-0.307
	(0.562)	(0.382)	(0.439)		(0.292)
house rooms	0.130	0.0151	0.0135	-0.0536	0.0101
	(0.104)	(0.0399)	(0.0256)	(0.0962)	(0.0199)
house dorm-rooms	-0.439**	-0.0679	-0.0895	0.0415	-0.0764*
	(0.202)	(0.0730)	(0.0589)	(0.163)	(0.0435)
ln expenditure rentals	-0.0716*	-0.0323**	-0.0395***	-0.0117	-0.0356***
	(0.0385)	(0.0151)	(0.0144)	(0.0311)	(0.0106)
dummy bathroom	-	-	-	-	-
sewerage system	-0.0921	0.00123	0.148		0.103
	(0.720)	(0.253)	(0.256)		(0.179)
garbage collection		-0.331	-0.823		-0.716
		(0.675)	(0.645)		(0.520)
private household	-	-	-	-	-
Constant	0.761	2.884*	1.495	12.83***	-5.651**
	(3.207)	(1.598)	(1.510)	(3.696)	(2.474)
Observations	384	1,163	1,399	376	2,388
Dummy for each region	YES	YES	YES	YES	YES
Cluster	YES	YES	YES	YES	YES