Universidade Federal de Pernambuco
Centro de Ciências Sociais Aplicadas
Programa de Pós-Graduação em Economia

Three Essays in Urban Economics

Klebson Humberto de Lucena Moura

Tese de Doutorado

Recife
February 2018

KLEBSON MOURA

THREE ESSAYS IN URBAN ECONOMICS

Thesis submitted to the Graduate Program in Economics of the Federal University of Pernambuco (PIMES), as a partial requirement to obtain a PhD.

Advisor: Prof. Dr. Raul da Mota Silveira Neto

FEBRUARY 2018

Catalogação na Fonte
Bibliotecária Ângela de Fátima Correia Simões, CRB4-773

M929t Moura, Klebson Humberto de Lucena
. Three essays in urban economics / Klebson Humberto de Lucena Moura.
. - 2018.
. 88 folhas : il. 30 cm.

. Orientador: Prof. Dr. Raul da Mota Silveira Neto.
. Tese (Doutorado em Economia) - Universidade Federal de Pernambuco,
. CCSA, 2018.
. Inclui bibliografia e apêndice.

. 1. Capital Humano. 2. Externalidade. 3. Moradia. I. Silveira Neto, Raul
. da Mota (Orientador) II. Título.

. 330.9 CDD (22.ed.) UFPE (CSA 2018 - 033)

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
DEPARTAMENTO DE ECONOMIA
PIMES/PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

PARECER DA COMISSÃO EXAMINADORA DA DEFESA DE TESE DO DOUTORADO EM ECONOMIA DE:

KLEBSON HUMBERTO DE LUCENA MOURA

A comissão Examinadora composta pelos professores abaixo, sob a presidência do primeiro, considera o Candidato Klebson Humberto de Lucena Moura **APROVADO**.

Recife, 28/02/2018.

Prof. Dr. Raul da Mota Silveira Neto
Orientador

Prof^a^. Dr^a^. Tatiane Almeida de Menezes
Examinadora Interna

Prof. Dr. André Matos Magalhães
Examinador Interno

Prof. Dr. Gervásio Ferreira dos Santos
Examinador Externo / UFBA

Prof^a^. Dr^a^. Roberta de Moraes Rocha
Examinadora Externa/ UFPE-CAA

ACKNOWLEDGMENTS

First of all to my parents, for having given the initial conditions that allowed me to get here and for the constant support and encouragement.

To Rosemary, for her unconditional support.

To my advisor, Prof. Dr. Raul da Mota Silveira Neto, for his extreme willingness to guide me and all considerations.

To the members of the Thesis Commitee, especially Profa. Dra. Roberta de Moraes Rocha, for the assistance of my academic trajectory.

To all Ph.D. colleagues, especially Rodrigo Carvalho Oliveira and Jorge Henrique Norões Viana.

Resumo

O objetivo desta tese é estudar algumas questões relacionadas ao ambiente urbano brasileiro. No primeiro capítulo procura-se identificar e medir os retornos externos ao capital humano no Brasil, utilizando informações do Relatório Anual de Informações Sociais (RAIS) do Ministério do Trabalho sobre todas as aglomerações urbanas do país para o período 2002-2014. Com uma estimativa de duas etapas e usando variáveis instrumentais para identificação, encontramos um efeito considerável da concentração de capital humano local nos salários locais. O conjunto de resultados indica que, no caso das áreas de mercado de trabalho brasileiro, os retornos externos ao capital humano são da ordem de 0,86% nos salários locais para um aumento de um ponto percentual nos graduados da faculdade. Em consonância com as expectativas teóricas os resultados são semelhantes aos encontrados na literatura pertinente, além disso, o impacto do capital humano local é mais forte para trabalhadores não qualificados do que para trabalhadores qualificados. Os resultados são robustos e não igualmente distribuídos entre os setores de trabalho; No segundo capítulo, analisamos a influência da exposição aos espaços públicos sobre a vitimização usando uma grande amostra representativa nacional de indivíduos brasileiros para 2009, usando técnicas de correspondência de propensão para criar contrafactuais, realizando verificações de robustez e implementando uma análise de sensibilidade baseada em simulação que suportam uma interpretação causal dos resultados. Os resultados indicam que os indivíduos com mais de uma hora de deslocamento têm um aumento global de 2.1% na probabilidade de ser vítima de roubo, sem impacto robusto no furto. Além disso, seguindo a literatura de exposição, encontramos maior efeito sobre a probabilidade de vitimização por roubo em mulheres quando comparado com homens, 2,5% e 2,2%, respectivamente; Finalmente, no último capítulo, apresentamos uma avaliação do impacto do financiamento habitacional para famílias brasileiras no tempo de deslocamento de casa ao trabalho (commuting). Com o uso de informações da PNAD 2014 e uma estratégia com base em Propensity Score Matching e análise de sensibilidade, os resultados indicam que o financiamento habitacional acarreta em aumento na probabilidade dos beneficiários em apresentar maior tempo de deslocamento dos chefes de família. Entretanto, este efeito se concentra em famílias de menor renda residentes em áreas urbanas não-metropolitanas.

Palavras-chave: *Capital Humano. Externalidades. Deslocamento ao Trabalho. Vitimização. Moradia*

Abstract

The objective of this thesis is to study some issues related to the Brazilian urban environment. In the first chapter the aim identify and to measure the external returns to human capital in Brazil using information from the Ministry of Labor's Annual Social Information Report (RAIS) on all urban agglomeration in the country for the period 2002-2014. With a two step estimation and using instrumental variables for identification we find a considerable effect of local human capital concentration on local wages. The set of results indicates that, in the case of brazilian labor market areas, the external returns to human capital are a 0.86% increase in local wages for an increase of one percentage point in college graduates. Consistent with theoretical expectations and similar to literature results, we also found that the impact of local human capital is stronger for unskilled than for skilled workers. Finally, results are robust and not equally distributed across sectors; On the second chapter we analyze the influence of exposure to public spaces on victimization using a large nationally representative cross-section sample of Brazilian individuals for 2009, using propensity score matching techniques to create counterfactuals, performing robustness checks and implementing a simulation-based sensitivity analysis that support a causal interpretation of the results. We find that individuals with more than one hour of commuting have an overall 2.1% increase in the probability of being victim of robbery, with no robust impact on theft. Also, following the exposure literature we find larger effect on the probability of robbery victimization on women when compared with men, 2.5% and 2.2% respectively; Finally, in the last chapter, presents an evaluation of the impact of the housing financing for Brazilian families in commuting time. With the use of PNAD 2014 information and a strategy based on Propensity Score Matching and sensitivity analysis, the results indicate that housing financing increases the probability of the beneficiaries having larger commuting time. However, this effect is concentrated in lower income families living in non-metropolitan urban areas.

Keywords*: Human Capital. Externalities. Commuting. Victimization. Housing*

List of Figures

List of Tables

Contents

1. EXTERNAL RETURNS TO HUMAN CAPITAL: EVIDENCE FROM BRAZILIAN CITIES

1.1 Introduction

The relation between individual human capital and earnings are quite known. Most of the empirical work confirms the intuitive notion that wage gains are in fact related to education attainment and not just to unobservable worker characteristics, leaving little doubt about the impact that education has on the private sphere. On the other hand, the relationship between concentration of local human capital in individual earnings are less known, despite human capital externalities being at the core of both theoretical models and justifications for education provision through public policies.

In this context, we can ask ourselfs what externalities would arise with the educational level of a given place and what would be their net effect. Answering this question matters to support the education subsidies and local development policies associated with the provision of education attaiment, since in the case of positive externalities the gains would not be limited to individuals but also present at the social level through educational spillovers. Despite the importance of this research agenda, empirical works with this specific goal are more recent and relatively scarce.

Among theoretical explanation for the existence of human capital externalities found in the literature, we have Rauch (1993)'s argument that knowledge and skills are shared between workers in formal and informal interactions, with higher level of human capital of individuals increasing probability that meaningful knowledge is shared when individual agents meet[1]. Additionally, Glaeser (1999) argues that larger cities, in addition to the possible existence of more formal education sources and specific training programs, would also offer unique opportunities through a non formal learning environment through

[1]Moddeling approach used by Jovanovic & Rob (1989)

social and business contacts, and therefore, would also benefit ordinary workers through the imitation of more skillful neighbors, leading to widespread impacts of human capital concentration and not just focused on innovation.

Acemoglu & Angrist (2001) divide theoretical arguments into two types, pecuniary and non-pecuniary. The first type are those that manifest themselves through interactions and sharing of knowledge and ideas in the lines of *Jacobs externalities*, due to Jacobs (1970) argument that cities are the main engine of growth because they facilitate exchange of ideas. The pecuniary ones, are in lines with *Marshalian externalities*, with the argument that increasing the geographic concentration of specialized inputs increases productivity, since the matching between factor inputs and industries is improved. Moretti (2004a) calls the human capital externalities of productive externalities, which occur when the presence of skilled workers makes other more productive workers so that an increase in aggregate human capital generates an effect on aggregate productivity.

Despite the clear mechanisms postulated by economic theory to explain external effects of human capital concentration, empirical estimation allways faced difficulties. The most important ones are mainly related to bias introduced by unobservable factors correlated with wages on one hand, and human capital concentration on the other. The first source of bias is the existence of sorting, the concentration of individuals of high levels of unobserved ability into cities. On the local level, a second problem is the bias introduced by city-specific unobserved characteristics that are correlated with human capital concentration. As suggested by Moretti (2004a), some examples are, unobserved differences in industrial mix, technology or natural resources. The empirical literature tries to control for these problems with, panel data for worker fixed effect, introduction of labor shock controls, and instrumental variables respectively.

However, only throughout recent years improvements in access to worker-level information fomented a rise in studies focused on wage disparities across regions which implements these estimation strategies. This allowed the emergence of research focused on corectly identify and measure of aglomeration and huma capital concentration effects (Moretti, 2004a,b; Combes *et al.* , 2008, 2010; Heuermann, 2011; Groot *et al.* , 2014; Broersma *et al.* , 2015; Barufi *et al.* , 2016). Empirical findings for develop countries shows that externalities due to human capital concentration are large and also affect unskilled workers. For the U.S estimates range from a 1% to 5% wage increase for a one year increase in local average scholling (Rauch, 1993; Acemoglu & Angrist, 2001). For West Germany, results are 1.8% and 0.6% (Heuermann, 2011). Unfortunately, research for developing countries is scarcer. Liu (2007), which following the empirical strategy of Moretti (2004a), assess the external returns to education in China. For Kenya, Manda *et al.* (2002) uses district-level average education attaiment to capture the effect of human capital on earnings, finding positive effects for male and female workers. Thus, there is a gap to be filled in the research program on the effects of human capital concentration, namely, increase evidence of human capital agglomeration in developing countries.

Specifically for Brazil, some facts can be considered stylized in the context of human capital. The first is the existence of high private return of education, ranging from 9.8 % to 18 % depending on the methodology used (Sachsida *et al.* , 2004; Barbosa Filho & Pessoa, 2008), secondly is the large regional wage inequality related to a inequalities in regional education attaiment (Barros *et al.* , 2007). In fact, the country has historically low human capital stock and presents a stagnation of labor productivity (Galeano & Feijó, 2013). In this setup the assessement of external effects of human capital is especially important because identify social returns to human capital concentration would serve as yet another argument to promote policies that aim human capital increase within the labor force. Despite those facts, studies with the specific goal of identifing external returns of human capital concentration are not numerous, some exceptions are Araujo Junior & Silveira Neto (2004) and Falcão & Silveira Neto (2007), however these initial studies do not fully address bias issues. Therefore, the main goal of this paper is to identify and measure external returns to human capital concentration on local wages in Brazil.

Specific contributions of the paper are twofold. First, despite the high known private returns of human capital in Brazil, the external effects are still unknown. We help to fill that gap in the national and developing countries literature by assessing evidence of human capital externalities. Second, although we focus on the effect of human capital concentration on local wages, we simultaneously include local employment density in our estimations. This inclusion is important to control for what Combes *et al.* (2010) calls “endogenous labor quantity”, that is, more productive places and therefore with higher wages, can attract more skilled and unskilled workers becoming denser, which may affect the correct estimates of the effect of interest by, possibly, changing local workforce composition.

Our results indicate that, for Brazilian labor market areas, the effect of human capital concentration, defined by the fraction of individuals with at least a college degree, in local wages is about 0.839, that is, a one percentage point in the concentration o college graduates increases local wages by about 0.84%. Also the effect is larger than purely agglomeration effects arising from population density only. Furthermore, the effect is higher for unskilled workers then for skilled and are not equally distributed across sectors, with returns concentrated in the manufacturing and service sectors. Results were robust to different specifications and estimation strategies.

The remainder of the paper is organized as follows. The next section presents a theoretical framework. Section 1.4 describes the empirical strategy and the treatment of the underlying data. Section 1.5 presents and analyzes the results and robustness checks, with the conclusions being presented in the last section.

1.2 Previous Empirical Evidence

The formalization of the impacts of productivity gains from education (human capital) in the per capita production growth are laid out in Lucas (1988), with it's endogenous growth model based of heterogenous human capital distribution between workers as the main explanation from the cross-country incomes differences. However, The first paper to exploit differences in human capital across cities to identify externalities was Rauch (1993). Using the cross sectional data from the 1980 US Census he found that a one year increase in average education raises wages by 3% to 5% percent in 1980. As point out by Moretti (2004a), Rauch's paper suffers from two methodology limitation, first is that he does not directly account for the endogeneity of aggregate human capital and, second, he does not distinguish between externalities and complementarity between skilled and unskilled workers.

In order to circunvey these estimation problems, Acemoglu & Angrist (2001) uses variations in compulsory education laws and the birth quarter of individuals as an instrumental variable for secondary schooling in a panel fix effects of U.S. states, finding smaller external returns to education, between 1% and 3%. However, the authors themselves argue that variations in compulsory scholling laws would affect mainly secondary education. Ciccone & Peri (2002) argue that studies with a Mincerian approach that try to identify human capital externalites with a standard Micerian wage regression, such as the previous cited ones, may confound positive externalities with wage changes arising from a negatively inclined demand curve for human capital. The authors propose to use a new approach holding the labour force skill composition constant, not finding significant externalities for cities and states between 1970 and 1990.

Focusing on college level education, Moretti (2004a,b), using a panel data, estimate a model of non-random selection of workers among cities, accounting for unobservable city-specifc demand shocks by using two instrumental variables for human capital concentration, namely, lagged demographic scholling structure of the city and the existence of a land-grant college. He finds that a percentage point increase in the supply of college graduates raises high school drop-outs' wages by 1.9%, high school graduates' wages by 1.6%, and college graduates wages by 0.4%. Focusing on wage inequality across US metro regions, Florida & Mellander (2014) find that regional variation in wage inequality, on the one hand, is associated with human capital, skill levels and occupational structure, in line with previous studies of skill-biased change and job polarization.

Considering the evidence for European urban centers, Heuermann (2011) using panel data and employing instrumental variables estimate human capital externalities for highly and non-highly educational groups and different sectors. Although the authors finds high a positive and significant effect, on contrary to Moretti (2004a), the effect is larger for the highly educated group, a 1.8% increase in the high skilled wage group and 0.6% for the non-highly educated. Also, the effects are larger in the manufacturing sector than in the service sector. More recently, Broersma *et al.* (2015), analyzes the effect of human capital externalities specifically for low-educated workers and for jobs with low skill requirements, using a employer-employee matched dataset for Dutch workers. Through a mincerian estimation in a multilevel model approach, including regional level data, the authors find that, workers on low-skilled jobs earn higher wages when working in cooperation with workers in high-skilled jobs, while for low-educated workers such cooperation with high-educated workers is negative. In a cross-country analysis of productivity in five OECD countries (Germany, Mexico, Spain, United Kingdom, and United States), Ahrend *et al.* (2014) uses a two-step econometric approach that enable to capture the productivity clean for individual unobserved heterogenity and sorting of more productive individuals into cities. The autors find that city productivity premiums tend to increase with city size. Specifically, a twofold increase in city size is associated with a 2-5% increase in productivity.

For developing Countries the lirature is scarcer. Liu (2007), follows the empirical strategy of Moretti (2004a) for assessing the external returns to education in China. He finds that the external returns are at least as high as the private returns to education, ranging from a low of 4.9% to a high of 6.7%. Two-stage least squares estimates indicate that a one-year increase in city average education could increase individual earnings by between 11 and 13%, and social returns can be as high as 16%. A significantly higher return than those found in developed countries. For Kenya, Manda *et al.* (2002) uses a Mincerian equation approach with district-level average education attaiment to capture the effect of human capital on monthly earnings, finding positive effects for male and female workers. More interestings, an increase in average human capital for females has a positive impact on earnings of male workers relative to female workers.

Specifically for Brazil, most papers focus on the role of density through an agglomeration economies approach such as Rocha *et al.* (2011) and more recently, Barufi *et al.* (2016) analyzes the effect of sector-specific agglomeration economies. They find that urbanisation economies positively and significantly affect the manufacturing and servive sectors, with impact of populationd density in local wages ranging from 5.1% to 9.4%. Despite the scarce literature some exceptions can be found, early works have studied the impact of human capital concentration on wages (Araujo Junior & Silveira Neto, 2004; Falcão & Silveira Neto, 2007), however these initial studies do not address bias issues.

1.3 Theoretical Framework

In order to estimate the effect of human capital concentration on productivity (trought wages), we must first present a theorical model. We use a simple framework proposed by Moretti (2004a) based on the general equilibrium model of Roback (1980, 1982). In this proposed theoretical framework the identification of externalities generated by the concentration of human capital, considers the imperfect substitution by different types of workers, skilled and unskilled. Considering, like Heuermann (2011), that the production function of the Cobb-Douglas type depends on two inputs, skilled and unskilled workers, we have to:

$$Y_j = (\theta_{1j} N_{1j})^{\alpha} (\theta_{2j} N_{2j})^{1-\alpha}, \quad \alpha \in (0,1) \tag{1.1}$$

where N_{1j} quantifies the total number of skilled workers in municipality *j* and N_{2j} the same measure for unskilled workers, with productivity mesured by θ_{1j} and θ_{2j} respectively. However, we use a more general and flexible functional that relates productivity and the share of skilled workers in region j (s_j), specifically:

$$\log(\theta_{ij}) = \phi_{ij} + f_i(s_j, \gamma), \quad i = 1, 2 \tag{1.2}$$

Where $f_i(s_j, \gamma)$ associates the productivity of each worker to the portion of skilled workers, and a positive parameter . Therefore, equation (1.2) indicates that the productivity of worker i in the municipality j depends on its own level of human capital, but also on the local stock of human capital of the locality j that must act by increasing the productivity of both workers positively. However, if the externalities generated by the concentration of human capital are not correlated with the productivity of the workers, it should be equal to zero. The generalization made by equation (1.2) is important because, unlike Moretti (2004a) and Heuermann (2011), the effects associated with human capital extrenalities does not necessarily act with equal intensity for both types of workers. Therefore, as seen bellow, with the generalization it is possible to have differentiated gains according to different levels of workers schooling.

Equating the wage of each type of worker to the marginal product of labor from the production function specified by equation (1.1), using equation (1.2) and $s_j = N_{1j}/N_{1j} +$

N_{2j} It follows:

$$\begin{aligned} \log(w_{1j}) = log(\alpha_1) + \alpha_1 \log(\theta_{1j}) + (\alpha - 1)\log(s_j) + (1 - \alpha_1)\log(\theta_{2j}) \\ + (1 - \alpha_1)\log(1 - s_j) \end{aligned} \tag{1.3}$$

$$\begin{aligned} \log(w_{2j}) = log(1 - \alpha_1) + \alpha_1 \log(\theta_{1j}) + (\alpha - 1)\log(s_j) + (1 - \alpha_1)\log(\theta_{2j}) \\ + (\alpha_1)\log(1 - s_j) \end{aligned} \tag{1.4}$$

By deriving equations (1.3) and (1.4) for region *j* as a function of the share of skilled workers, the impact of a marginal increase in the number of skilled workers on the salary of both workers is obtained as follows :

$$\frac{d\log(w_{1j})}{ds_j} = f_1^{'}(s_j, \gamma) + \frac{\alpha - 1}{s_j - s_j^2}, \tag{1.5}$$

$$\frac{d\log(w_{2j})}{ds_j} = f_2^{'}(s_j, \gamma) + \frac{\alpha}{s_j - s_j^2}. \tag{1.6}$$

Skilled workers will benefit the most by higher wages, by taking advantage of the externalities generated by the concentration of human capital if the following condition is observed:

$$f_1^{'}(s_j, \gamma) > f_2^{'}(s_j, \gamma) + \frac{1}{s_j - s_j^2} \tag{1.7}$$

Here two consideration must be made. According to expression (1.7), a marginal increase in the stock of human capital in a given region will have a greater positive effect on skilled workers wages than of unskilled workers if $f_1^{'}(s_j, \gamma)$ the "externality effect" overcome the wage gains of unskilled workers $f_2^{'}(s_j, \gamma)$ added to the "neoclassical net effect" $1/s_j - s_j^2$. Also, the neoclassical "net effect" is minimum when $s_j = 1/2$, taking te low percentage of skill workers in Brazil we can assume that $s_j < 1/2$, implying a large “neoclassical effect” on regions. Therefore, for an empirical confirmation of equation (1.7), externalities must be stronger than the “neoclassic net effect” of the region. Analogously, under these conditions, in the case when $f_1^{'}(s_j, \gamma) = f_2^{'}(s_j, \gamma) = \gamma s_j$, external returns to education for unskilled workers will be higher for low human capital stocks in "j". According to this model, it is possible that empirical evidence for countries or even regions of the same country with different levels of human capital presents different results regarding the magnitude of the relation between concentration of human capital and productivity gains, not being directly comparable.

In summary, the main result of the model for equation (1.2), suggests that human capital concentration influences skilled and unskilled workers wages differently through two forces, "spillovers effect" and the "neoclassical effect". The former operates through productive externalities generated by the concentration of skilled workers - which makes access to information and learning easier and faster, among other reasons - increasing the productivity of both workers. The later derives from imperfect substitution between skilled and unskilled workers - an increase in the supply of skilled workers reduces the productivity of these individuals and thus their wages, but raises the productivity of the unskilled.

1.4 Empirical Strategy and Descriptive Statistics

To achieve the goal of correctly estimate human capital externalities in local productivity, measure trought real wages, we must consider the possible sources of omitted variable bias and the subsequent difficulties that its existence imposes the estimation

The first source of bias, on the individual level, is the workers unobserved heterogeneity may affect estimation. Individuals observed in cities with a higher concentration of human capital may have more ability than individuals with the same formal educational level residing in a location with a lower concentration of human capital. As stated by Moretti (2004a), this type of sorting can be problematic if cities with a higher share of college are associated with a high return for unobserved abilities, attracting the most skillful and affecting the labor force composition.

Second, on the local level, another potential source of bias is related to shocks in the local labor market, which may be correlated with the concentration of skilled workers in the locality. Even with cities fixed effects, time-varying factors may affect the stock of human capital or wages between cities. For example, shocks can attract skilled workers to a particular region and to raise wages (Moretti, 2004a; Falcão & Silveira Neto, 2007). This problem is adressed by constructing a demand shock index for each educational group and area, which represents the expected change in employment for an educational group in the area.

Finally, as explained in Combes *et al.* (2010), density and measures of productivity (wages) may be simultaneously determined, because more productive places tend to attract more workers and as a result become denser. This issue is refered as the "endogenous quantity of labor" problem, a higher employment density can exacerbate the knowledge spillover, despite of human capital concentration. Including employment density as a control variable, seeks to obtain a effect of human capital externalities "clean" of density effect.

1.4.1 Model Specification

To simultaneously control all of these bias sources, we use a two step estimation approach[2] adapting from Heuermann (2011) and Groot *et al.* (2014). The first step consists in estimate a Mincerian type equation for individual-level data given by equation (1.8):

$$\begin{aligned} \ln(w_{i,j,t}) &= \beta_0 + \theta_i + \beta_1 age_{i,j,t} + \beta_2 age^2_{i,j,t} + \beta_3 tenure_{i,j,t} + \textstyle\sum_{edu} \beta_{edu} D^{edu}_{i,j} \\ &+ \textstyle\sum_{sector} \beta_{sector} D^{sector}_{i,j} + \sum_{size} \beta_{size} D^{size}_{i,j} + \sum_j \sum_t \lambda_{j,t} D_{j,t} + \epsilon_{i,t}, \end{aligned} \tag{1.8}$$

where $w_{i,j,t}$ is the wage from individual i in local j on year t, θ_i is the individual fix effect, $D^{edu}_{i,j}$ are a series of educational dummies, $D^{sector}_{i,j}$ are sector dummies, include to control for sector-specific effects, $D^{size}_{i,j}$ are firm size dummies, measure by number of employees. Most importantly, $D_{j,t}$ are dummies for each local and year, providing a local-year specific wage index, given by $\lambda_{j,t}$. Ideally, estimating equation (1.8) using longitudinal data, allows to control for an important source of bias, the unobserved heterogeneity in workers abilities.

The second step regression uses first step estimates of local wage indexes, $\hat{\lambda}_{j,t}$, as dependent variable in a regression on local human capital concentration, employment density. Also, as we know from economic theory, market structure can affect prices and wages, therefore, we include the share of jobs in industry, a sector with historically higher wages and three measures for, specialization, diversity and competition[3] (that will be define below) as shown in equation (1.9):

$$\begin{aligned} \hat{\lambda}_{j,t} &= \alpha_0 + \alpha_1 humancapital_{j,t} + \alpha_2 \ln(employ.density_{j,t}) + \alpha_3 \ln(area_j) \\ &+\alpha_4 industryshare_{j,t} + \alpha_5 specialization_{j,t} + \alpha_6 diversity_{j,t} \\ &+\alpha_7 competition_{j,t} + \alpha_8 demandshocks_j + \varepsilon_{j,t} \end{aligned} \tag{1.9}$$

In this set up we are interested in correctly estimate α_1, the effect of human capital

[2] A similiar approach was used by Ahrend *et al.* (2014) for estimate agglomeration benefits based on city productivity differentials across five OECD countries (Germany, Mexico, Spain, United Kingdom, and United States).

[3] see Barufi *et al.* (2016).

concentration on the local wage index. However, as discussed earlier, in the second step we also should use strategies to control for possible sources of bias.

First, in order to control for the "endogenous quantity of labor" we first use lagged population density as instrument for current employment density. As argued by Combes *et al.* (2010) for the french case, considering a relatively constant urban hierarchy, highly dense regions today were highly dense in the past, however, in those days the most important sector was agriculture and density was affect only by the capability of sustain population. In a relative young country as Brazil, the urban hierarchy may not be highly stable and the drivers of urban center development in Brazil can be potentialy related to main historical economic cycles from the colonial period, such as the sugar-cane, gold mining and coffee production as discussed by Naritomi *et al.* (2012). The autors construct a variable which indicates the influence of each of those economic cycles in the brazilian municipalites with value 1 for those directly affected and defined by

$$I_i = \begin{cases} \left(\dfrac{200 - d_i}{200}\right) & \text{se } d_i \leq 200\text{km}, \\ 0 & \text{otherwise}, \end{cases} \tag{1.10}$$

where d_i is the distance from municipality i to the closest municipality directly involved in the respective resource boom. The assumption is that these economic episodes help define the brazilian urban structure by partialy deteming urban centers location and development and no longer are related to the current productive structure, making then a plausable instrument for employment density.

Also, we must include other market area variables related to local wages. For market measures we first use the specialization index given by equation (1.11), in which a value close to 0 indicates industrial composition in the region is similar to the national one, for values closed to 1 the labor market area is completely specialized. As Henderson (2010) affirms, standardized manufacturing activity benefits from agglomeration generating productivity improvents that can affect local wages.

$$Specialization_{region} = \sum_i \left(\frac{E_{i,region}}{E_{i,region}} - \frac{E_{industry}}{E_{country}}\right)^2 \tag{1.11}$$

The second market measures included as control is for diversity, since Jacobs (1970) arguments that a city with larger sector diversity, because in diverse cities there is more interchange of diferent ideas (Glaeser *et al.* , 1992). Following Combes *et al.* (2011), we use a Inverse Herfindahl Index given by equation (1.12), in which higher values represents higher diversity within labor market area.

$$IHI = \left(\frac{E^2_{region}}{\sum_{ind} E^2_{industry,region}} \right) \tag{1.12}$$

For the last market measure, we include a degree of competition measure given by equation (1.13), where values larger then one indicates larger competition in the industry within a particular labor market area. Competition can have a dual effect on productivity improvements, which can affect local wages. A greater competition accelerate imitation and improvement of inovative ideias on one hand and reduces returns to inovators (Glaeser *et al.* , 1992).

$$C_{industry,region} = \frac{F_{industry,region}/E_{industry,region}}{F_{industry,country}/E_{industry,country}} \tag{1.13}$$

Finally, we use an instrumental variable approach proposed by Moretti (2004a) to control for possible demand shocks that attract skilled workers to a particular region and raise wages. The proposed instrument seeks to capture the generational change in the schooling of the regions considered in the analysis, such instrument is supposedly correlated with the human capital stock, but is not influenced by current shocks in the local labor market. Moretti (2004a) argues the existence of a long-run trend of incresing education, as younger and more educated cohorts enters the labor force, as relative population shares of diferent cohorts vary across cities, this will lead to diferential trends in college share across cities. The instrument is defined as:

$$IV = \sum_m \omega_{mj} \Delta P_m, \tag{1.14}$$

where m identifies age groups, j the geographical unit, ω_{mj} the proportion of individuals living in the region j in 1980 and who, in 2000's, belonged to the m age group, finally ΔP_m captures national change in the share of skilled workers by group m between the various years of the study (2002 to 2014). The authors note that, if the age distribution of cities may reflects expected changes in the local economy the estimates would still be biased, usage of lagged values circunvent this issue.

1.4.2 Data

As previously discussed, in order to achieve proposed objective, it's necessary to initially estimate equation (1.8), generating a local wage index . For this, we must have individual-level wage information, socioeconomic characteristics, labor relations type and individual location. For the brazilian case, at the national level, the main database that meets these requirements is given by the Ministry of Labor's Annual Social Information Report (RAIS).

The RAIS dataset consists of information on all formal workers in the country, whose advantages are its scope, longitudinal structure of the information and geographical breakdown. These advantages allow include worker fixed effects in order to correct the unobserved heteroneneity bias. However, because it is restricted to formal workers, it leaves aside a large part of the individuals employed in the country. Fortunately, as emphasized by Baruffi et al. (2016), during the first decades of economic stabilization (first and second decades of the current century), the process of formalization guarantees a greater representativeness of the RAIS dataset. To put into perspective, for 2002 the RAIS dataset contains about 46 million workers, increasing in 2014 to more then 76 million workers.

In order to avoid computational issues we decide to keep only seven time periods, all the even year from 2002 to 2014. Also, we keep only the workes who had a permanent job contract with at least 20 weekly hours that was active at the end of the year. In order to create a more homogeneus sample we also removed the public administration workers, since wages in this sector do not necessarily reflect workers productivity and presents a specific contract relation. Finally, we considered only men between 18 and 56 years of age. With the selections discussed above, we built a panel at individual level for the workers present in every considered year. With this in mind our final sample consists of 1,327,411 workers for each year, for a total of 9,291,877 observations.

For the Brazilian case the most disaggregated administative unit is the municipality, however, labor market interactions often cross administratively defined boundaries. For this reason, Combes *et al.* (2008, 2010) for the french case, use employment zones built by the French Ministry of Labor in their analisys of agglomeration externalities and Ahrend *et al.* (2014) uses *functional urban areas*, built by for OECD countries, in their analysis cross-country analysis of productivity. A direct equivalent for Brazil are the so-called *Regiões de Influência* (Regions of Influence) which are areas of influence of local urban centers that takes into account the brazilian urban network and all daily commuting and transportation connections among the municipalities, created by the Brazilian Statistics Bureau IBGE (2013). In it's most desagregates form, the country is divided into 482

Regions of Influence (REGIC) that can be thought as labor market areas and will be reffered to as such from now on. Before presenting results, we perform a brief descriptive analysis of the data used, in order to identify the main characteristics of the dataset and to substantiate the later discussion of results.

Table 1.1: Descriptive Statistics for Individual Workers Sample (First Step)

	2002		2014		Whole Period	
Variables	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Wage (R$ per month)	741.07	26,677.3	2,304.4	4,762.7	1,370.94	11,592.8
Age (years)	33.89	9.10	45.9	9.11	39.63	10.03
Tenure (months)	61.6	69.5	128.6	109.6	97.53	91.1
Educational Levels						
Illiterate (%)	0.02		0.01		0.01	
Incomplete Middle School (%)	0.36		0.24		0.29	
Complete Middle School (%)	0.28		0.22		0.25	
Complete High School (%)	0.26		0.38		0.32	
Incomplet College (%)	0.03		0.02		0.03	
College (%)	0.06		0.13		0.10	
Firm Size (Employee Number)						
(0;4)	0.095		0.076		0.059	
(5;9)	0.099		0.087		0.087	
(10;19)	0.109		0.095		0.097	
(20;49)	0.138		0.131		0.122	
(50;99)	0.101		0.097		0.106	
(100;249)	0.128		0.119		0.115	
(250;499)	0.103		0.099		0.108	
(500;999)	0.091		0.097		0.096	
1000+	0.134		0.198		0.147	

Source: author's own calculation based on the RAIS dataset.

In Table 1.1, we present the descriptive statistics for the first step variables, that is, in equation (1.8) for 2002, 2014 and the role period. As can be seen, the average monthly wage for the period is R$ 1,370 with a high standard deviation indicating great wage inequality among sample workers. Also, the average age is almost 40 years and a 98 months tenure, this can be reflected in the restriction of the sample to individuals that are present in the sample in all seven periods considered between 2002 and 2014. The low percentage educational level obtaneid by the workforce is striking, Table 1.1 shows that only 10% of them have at least a college degree, the majority finishes high school. By this metric it is easy to observe the low overall stock of human capital in brazilian cities.

The descriptive statistics of second step (labor market area level) information are show in Table 1.2. As previously discussed, local wages indexes where estimated by including dummies for each labor market area and year combination, thus it becomes a measure of wage for each labor area. On average, 6% of a regic workers have college

Table 1.2: Descriptive Statistics for Local Variables (Second Step)

Variable	Mean	Std. Dev	Min.	Max
Local Wage Index (estimated coefficients)	4.34	0.17	3.17	4.90
Local Human Capital (share of college graduates)	0.06	0.12	0.01	0.22
ln *Employment Density* $(jobs/km^2)$	0.32	1.64	-5.44	4.31
ln *populational density in 1940* (hab/km^2)	2.15	1.54	-3.63	5.24
ln *area* (km^2)	8.85	1.28	5.70	13.6
Jobs in Industry (%)	0.22	0.09	0.01	0.49
Skilled Demand Shock	1.07	0.21	0.00	1.41
Unskilled Demand Shock	0.39	0.15	-1.15	0.89
Market Measures[b]				
Specialization Index	0.02	0.03	0.00	0.37
Diversity Index	0.31	0.12	0.08	0.76
Competition Index	0.84	0.26	0.36	4.88
Historical Variables[c]				
Sugar Cane Boom	0.08	0.21	0.00	1.00
Gold Boom	0.11	0.26	0.00	1.00
Coffee Boom (Colonial Era)	0.09	0.26	0.00	1.00
Coffee Boom (Post-Colonial Era)	0.15	0.33	0.00	1.00

Source: author's own calculation based on the RAIS dataset.

[a] All calculations based on the Region of Influence (REGIC) geographical unity.

[b] Following Barufi *et al.* (2016) the respective measures are: degree of specialization, inverse herfindahl and degree of competition.

[c] Refers to the value of the index created by Naritomi *et al.* (2012) for mesuring the proximity to each of the refered colonial economic booms.

degrees, our chosen metric for human capital concentration. The high standard deviation (0.12) imply a high human capital inequality between labor market areas. Indeed, many recent articles in Brazil discuss the importance of human capital and wage differentials in income inequality and its recent decline (Barros *et al.* , 2007; Tavares & Menezes Filho, 2011; Silva *et al.* , 2016).

As can be seen, on average labor areas have similar composition to national one. The index shows a larger variation with an average of 0.31 and 0.36 as standard deviation. As we can see, on average, labor market areas presents less competition compared to the country. However, there is cases of high competition as the maximum copetition index approachs 5.

Finally, in Table 1.2 we have the measure of proximity of each labor market area to major colonial economic cicles. As previously explained, values close to 1 shows greater proximity and therefore greater influence. As shown by Naritomi *et al.* (2012), these economic episodes have a long standing impact on local institutional framework, measure by access to justice, land inequality and governance. For our purpuses, these institutional outcomes can affect human capital concentration and density today, and must be considered.

1.5 Results

In Table 1.3, we show the results for the first step estimation, following equation (1.8). As we can see, the overall fit is satisfactory with a R^2 of 0.45 with all control variables presenting highly significant coefficients. Wages are higher for older and longer tenured workers, although there is a nonlinearity in age presente by the negative coefficiente of age^2. Higher educational levels and working on larger firms positively correlate with higher wages. The inclusion of dummies for REGIC and year interactions, allows the estimation of $\lambda_{j,t}$ to be use in the second step.

As previously discussed, the unobserved worker heterogeneities can turn OLS estimations inconsistent, therefore, we perform a panel with individual fixed-effect estimation, the last two columns of Table 1.3 presentes the results. As we can see, effects of educational levels an firm size on individual wages, although still highly significant, are strongly reduced, indicating that a considerable share of those returns were from unobserved ability. This highlights the importance of controls for unobserve ability. Given that fixed-effect estimation appears to be the most adequate[4], in the second step the wage indexes from this estimation will be used.

In the second step, with the estimates of local wage indexes and control variables aggregated in labor market area level, we estimate equation (1.9). Again, to provide stronger conclusions, the existence of sorting and consequent endogeneity of human capital concentration and local wage indexes requires an estimation strategy with instrumental variables, therefore the second step was estimated using a 2SLS approach. In order to assess their relevance, in Table 1.4 we present the first stage estimation. We can see that as expected the proposed instruments are relevant only for it's endogenous counterparts, that is, the demographic schooling structure instrument is relevant for Human Capital and population density, sugar, gold and coffee booms are relevant for employment density.

Second step results are presented in Table 1.5. For comparison, we present, in the first column, OLS estimation of the equation (1.9), we can see that the concentration of human capital has a positive and significant effect on the hourly wages of workers, a one point increase in the share of workers with college increases labor market areas wages by 0.70%. However, employment density presents a negative impact of wages, contradicting the aglomeration literature when we don't control for the "endogenous labor quantitiy".

[4]As in Combes *et al.* (2010) indetification of parameters come from worker movement between labor

Table 1.3: Human Capital Concentration and Local Wages (Mincerian Equation)

	Depent Variable.: Individual Hourly Wage			
	OLS		Panel	
Explanatory Variables	Coef.	Stand. Dev.	Coef.	Stand. Dev.
age (years)	0.044***	(0.000)	0.097***	(0.001)
age^2	-0.001***	(0.000)	-0.002***	(0.000)
tenure (months)	0.002***	(0.000)	0.002***	(0.000)
tenure^2	0.000***	(0.000)	0.000***	(0.000)
Educational Levels				
< Middle School	0.177***	(0.002)	0.007***	(0.002)
Middle School	0.296***	(0.002)	0.001***	(0.002)
High School	0.581***	(0.002)	0.017***	(0.002)
College	1.397***	(0.002)	0.131***	(0.002)
Firm Size (Employee Number)				
(0;4)	-0.192***	(0.012)	-0.124***	(0.012)
(5;9)	-0.123***	(0.012)	-0.090***	(0.012)
(10;19)	-0.041***	(0.012)	-0.045***	(0.012)
(20;49)	0.051***	(0.012)	-0.002***	(0.012)
(50;99)	0.136***	(0.012)	0.046***	(0.012)
(100;249)	0.229***	(0.012)	0.094***	(0.012)
(250;499)	0.275***	(0.012)	0.130***	(0.012)
(500;999)	0.298***	(0.012)	0.151***	(0.012)
1000+	0.326***	(0.012)	0.167***	(0.012)
Ignored	0.305***	(0.012)	0.180***	(0.012)
Intercept	0.250***	(0.072)	0.356***	(0.075)
Sector Dummies	Yes		Yes	
$\lambda_{j,t}$ *Dummies*	Yes		Yes	
R^2	0.452		0.455	
N	8,928,383		8,928,383	

Significance levels: *** 1% , ** 5% and * 10% .

The 2SLS estimation results are presented in the third column of Table 1.5. As we can see, the effect of interest grows to 0.859, a 22% increase, indicating that a one point increase in the share of workers with college increases labor market areas wages by 0.86%. This effect in similar to Moretti (2004a) findings of positive externalities for human capital concentration (around 1.14%). With inclusion of instruments for employment density it's estimate becomes positive and significant, implying that denser labor market areas presents higher wages as in Barufi *et al.* (2016). Market measures introduced as controls are not statistically significant but their inclusion in important for an obtaining

market areas and from time variation for worker ho didn't move

Table 1.4: Human Capital Concentration and Local Wage Index (2SLS First Stage)

	Human Capital[a]		ln *Employ. Density*	
Explanatory Variable	Coef.	S.D	Coef	S.D
Age Structure	2.36***	(0.104)	-5.240	(3.604)
Pop. Density 1940	0.000	(0.000)	0.325***	(0.062)
Sugar Boom	-0.003	(0.004)	2.63***	(0.255)
Gold Boom	-0.003	(0.003)	0.544**	(0.242)
Coffee Boom (Colonial)	0.003	(0.005)	1.453***	(0.193)
Coffee Boom (Post-Colonial)	0.016**	(0.006)	0.603***	(0.235)
Constant	0.035***	(0.010)	-3.902***	(0.716)
Exogenous Regressors	Yes		Yes	
F-test	453.5***		35.40***	
Sanderson-Widjmeier[a]	680.09***		243.95***	
N	390		390	

Significance levels: *** 1% , ** 5% and * 10% .
[a] Human Capital Concentration is measure as the share of workers with at least a college degree;
[b] Consists of a weak identification test of individual endogenous regressors. They are constructed by "partialling-out" linear projections of the remaining endogenous regressors.

an unbiased estimate of human capital concentration on local wages. Finally, although skilled demand shock no longer plays a significant role, the impact of unskilled demand shocks remains important for the wages in the period.

These results offer evidence on the existence of human capital externalities, however, in order to empiricaly test the assumption of a distinct effect between educational groups as proposed by the theoretical model of Section 1.3, we must perform separate estimates for different educational groups.

1.5.1 Effect on Different Educational Groups

As previously discussed, human capital concentration influences skilled and unskilled workers wages differently through two forces, “spillovers effect” and the “neoclassical effect”. The former operates through productive externalities generated by the concentration of skilled workers - which makes access to information and learning easier and faster, among other reasons - increasing the productivity of both workers. Therefore we must reestimate the equation (1.8) considering this differentiation.

Considering that both groups are present in the regions labor markets and wage determination is simultaneos, in this set up separately estimate the first step for each

Table 1.5: Effects of Human Capital Externalites in Local Wage - Reduced Form

	OLS		2SLS	
	Coef.	Std. Dev	Coef.	Std. Dev
Human Capital[a]	0.699***	(0.130)	0.839***	(0.129)
ln *Employment Density*	-0.004	(0.007)	0.024**	(0.010)
ln *REGIC Area*	0.002	(0.007)	0.002	(0.006)
Jobs in Industry	-0.127	(0.097)	-0.134	(0.095)
Diversity	0.049	(0.048)	-0.005	(0.048)
Specialization	0.115	(0.307)	-0.580	(0.428)
Competition	0.043	(0.027)	0.001	(0.018)
Centrality	-0.098**	(0.042)	-0.085*	(0.045)
Foundation (year)	0.000	(0.000)	0.000	(0.000)
Major Port Dummy	0.024	(0.047)	-0.052	(0.040)
Skilled Demand Shock	-0.053**	(0.024)	-0.024	(0.021)
Unskilled Demand Shock	-0.075**	(0.024)	-0.106***	(0.035)
Constant	4.46***	(0.016)	4.39***	(0.145)
Regional Dummies	Yes		Yes	
Kleibergen-Paap[b]			78.30	
Cragg-Donald Wald F[c]			33.2	
Hansen-J[d]			8.20	
R^2	0.311			
N	482		390	

Note: Dependent variable in the local wage index generated at the first step mincerian wage equations given by equation (1.8). Excluded instruments are: Lagged age structure, 1940's population density, Sugar, Gold and Coffee boom proximity.
[a] Human Capital Concentration is measure as the share of workers with at least a college degree;
[b] test of whether the equation is identified, i.e., that the excluded instruments are "relevant", meaning correlated with the endogenous regressors.
[c] Should be compared with Stock & Yogo (2002) critical values.
[d] Hansen-J test is for overidentified restrictions. The null hipothesis is that all instruments are valid.
Significance levels: *** 1% , ** 5% and * 10% .

group is not a good strategy. Therefore, we included a interaction dummy between $D_{j,t}$ the dummies for each local and year, with a skilled worker dummy. Implicitly considering the hypothesis that there is an effect of the concentration of human capital for all the workers, with a specific return added for skilled workers trough education.

The second step regression uses first step estimates of local wage indexes $\hat{\lambda}_{j,t}$ only, for the unskilled group and $\hat{\lambda}_{j,t}$ added with a dummy for skilled group, estimation results are presented in Table 1.6. Theorical prediction, according to equation (1.7) is that a marginal increase in the stock of human capital in a given region will have a larger posi-

Table 1.6: Human Capital Concentration and Local Wage Index (2SLS Reduced Form Educational Groups)

	Depedent Variable: Local Wage Index			
	Skilled[a]		*Unskilled*[b]	
	Coef.	Stand. Dev.	Coef.	Stand. Dev.
Human Capital	0.254**	(0.134)	0.737***	(0.136)
ln *Employment Density*	0.006***	(0.005)	0.014***	(0.008)
ln *REGIC Area*	0.001	(0.004)	-0.012	(0.009)
Jobs in Industry	0.010	(0.080)	-0.117	(0.095)
Diversity	0.001	(0.043)	0.001	(0.047)
Specialization	-0.628	(0.347)	-0.279	(0.439)
Competition	-0.005	(0.014)	0.004	(0.018)
Skilled Demand Shock	-0.020	(0.024)	-0.005	(0.025)
Unskilled Demand Shock	-0.126	(0.036)	-0.107***	(0.040)
Constant	-0.804***	(0.058)	4.70***	(0.000)
Macro Dummies	Yes		Yes	
Kleibergen-Paap[c]	72.5		72.5	
Cragg-Donald Wald F[d]	37.5		37.5	
Hansen-J	2.55		9.14	
N	390		390	

Note: Excluded instruments are: Lagged age structure, 1940's population density, Sugar, Gold and Coffee boom proximity.

[a] Dependent variable is $\hat{\lambda}_{j,t} + \hat{\gamma}_{j,t}$, the local wage index generated at the first step mincerian added with a local wage index of only the skilled workers.

[b] Dependent variable is $\hat{\lambda}_{j,t}$ the local wage index generated at the first step mincerian equations given by equation (9)

[c] test of whether the equation is identified, i.e., that the excluded instruments are "relevant", meaning correlated with the endogenous regressors.

[d] Should be compared with Stock & Yogo (2002) critical values.

[e] Hansen-J test is for overidentified restrictions. The null hypothesis is that all instruments are valid.

Significance levels: *** 1% , ** 5% and * 10% .

tive effect on skilled workers wages than of unskilled if the "externality effect" overcome the neoclassical "net effect". As we can see, our results show that wages of both groups are affected by the externalities resulting from the concentration of human capital. Moreover, as the effect found is greater for non-skilled individuals (0.73%) then for skilled (0.25%), we can imply that these externalities for the skilled group are not strong enough to outweigh the gains in the wages of the unskilled when we also consider the neoclassical effect. These results are similar to Moretti (2004a), which finds a 1.6% for highschool graduates and 0.4% for college graduates.

1.5.2 Effect on Different Sectors

As stated by Heuermann (2011), labour market institutions and the relative importance of knowledge and physical capital as factors of production can be quite diverse across industries. This fact together with findings of wage differences across sectors makes interesting to know the specific impacts of human capital concentration across industries in order to promote strategies and improve their efficiency. Therefore, in order to evaluate these differences, we repeat our baseline estimation separately considering a agregation of two-digit CNAE divisions into four large groups, namelly, manufacturing industry, extractive industry, construction and services. Results are presented in Table (1.7)

Table 1.7: Human Capital Externalities Effects by Economic Sectors

	Depedent Variable: Local Wage Index			
	Manufacturing	Extractive	Construction	Services
Human Capital[b]	0.524**	-1.687	-1.141	0.365***
	(0.244)	(2.749)	(1.142)	(0.166)
ln *Employment Density*	0.031	0.230**	0.119	0.025
	(0.029)	(0.106)	(0.082)	(0.017)
Constant	5.73***	5.57***	5.11***	5.75***
	(0.270)	(1.062)	(0.764)	(0.186)
Controls	Yes	Yes	Yes	Yes
Macrorregion Dummies	Yes	Yes	Yes	Yes
Fist Stage F-*test*	75.2***	71.8***	71.8***	72.5***
Kleibergen-Paap[c]	33.2	36.1	34.8	35.6
Hansen-J	2.96	3.06	3.28	4.36
N	390	349	387	390

Note: Excluded instruments are: Lagged age structure, 1940's population density, Sugar, Gold and Coffee boom proximity.
[a] Dependent variable is $\hat{\lambda}_{j,t} + \hat{\gamma}_{j,t}$, the local wage index generated at the first step mincerian added with a local wage index of workers in that sector. This is similar with the strategy used for skilled and unskilled estimations.
[b] Human Capital Concentration is measure as the share of workers with at least a college degree;
[c] test of whether the equation is identified, i.e., that the excluded instruments are "relevant", meaning correlated with the endogenous regressors.
[d] Should be compared with Stock & Yogo (2002) critical values.
Significance levels: *** 1% , ** 5% and * 10% .

As we can see, the results confirm a variety in human capital concentration effects the effects between different sectors, implying that human capital concentration benefits are unevenly distributed. For example, we find larger effects for manufacturing with (0.52%) and services (0.36%), with no significant effect for the remaining two groups, extractive industry and construction. However, these results are intuitive given their underlying caracteristics of low gains of exchange of information and social enviorment. On the other hand, concentration externalities play a important role for extractive, as the coeficient of employment density was positive and highly significant for these sectors.

Our results are in line with (Heuermann, 2011), which finds for West Germany positive effects for the manufacturing industry and services, the former with highest effects than the later. As an explanation, the author uses the notion of pecuniary externalities versus tecnological externalities as microeconomic sources for human capital externalities. The first arises when the firms relate their investments in physical capital to local human capital endowments, which could be true for physical capital intensive industries like the manufacturing. while the second arises when externalities come from the exchange of knowledge between firms and workers, more probable in services. Specifically for Brazil, we have no knowledge of directly comparable results, since Barufi *et al.* (2016) focus on the employment density effect.

1.5.3 Robustness Checks

In order to assess the validity of human capital agglomeration effects, in this section we perform a series of robustness checks. First we include as controls variables related to amenities and geographical characteristics possibly correlated to wages. With the notion of compensatory wages[5], part of the existing wage variations compensate the accessibility to amenities. Without including controls for this compensation could lead to a overestimation of the true effect, therefore, in order to correctly evaluate the effect of human capital concentration on local wage indexes, inclusion of amenities and geographic characteristics controls is necessary. Therefore, we include altitude, number of sunny days, number of rainy days, distance to equator and distance to sea as controls.

In the remainder robustness checks, we repeat the baseline estimation using subsets of labor market areas. First we exclude Rio de Janeiro e São Paulo labor makert areas from the estimation based on concerns that, given Brazil's urban hierarchy with two disproportionately big metropolitan areas, São Paulo and Rio de Janeiro influence on the overall results may be overwhelming. We then restrict the estimation for only non-Metropolitan labor makert areas is order to evaluate if results hold outside metropolitan areas. For this we considered only the areas were none of its composing municipalities lay in metropolitan regions, ending up with 357 non-metropolitan labor market areas.[6].

[5] See Roback (1982).

[6] Unfortunately, it is not possible to perform an estimation considering only the metropolitan regions by their small number, creating problems on the estimation matrix rank.

Table 1.8: Human Capital Externalities Effects - Robustness Checks

	Depedent Variable: Local Wage Index				
	(1)	(2)	(3)	(4)	(5)
Human Capital	0.770***	0.856***	0.887***	0.853***	0.825***
	(0.138)	(0.128)	(0.130)	(0.213)	(0.169)
ln *Employment Density*	0.031**	0.016***	0.017**	0.014	0.017
	(0.015)	(0.007)	(0.007)	(0.012)	(0.013)
Constant	4.59***	4.55***	4.54***	4.57***	4.54***
	(0.121)	(0.080)	(0.086)	(0.111)	(0.139)
Controls	Yes	Yes	Yes	Yes	Yes
Regional Dummies	Yes	Yes	Yes	Yes	Yes
Amenities[a]	Yes	No	No	No	No
Excluding São Paulo and Rio	No	Yes	No	No	No
Only Non Metropolitan	No	No	Yes	No	No
Only 50% Smaller Regics	No	No	No	Yes	No
Only 50% Bigger Regics	No	No	No	No	Yes
Kleibergen-Paap[b]	53.3***	75.6***	71.6***	36.9***	39.5***
Cragg-Donald Wald F[c]	16.5	38.6	35.7	17.7	17.9
Hansen-J[d]	6.47	8.57	9.29	8.43	5.95
N	390	388	357	195	195

Note: Dependent variable in the local wage index generated at the first step mincerian wage equations given by equation (1.8) for each educational group. Excluded instruments are: Lagged age structure, 1940's population density, Sugar, Gold and Coffee boom proximity.

[a] The amenity variables are altitude, number or sunny days, number of rainy days, distance to the equador and distance to sea.;

[b] test of whether the equation is identified, i.e., that the excluded instruments are "relevant", meaning correlated with the endogenous regressors.

[c] Should be compared with Stock & Yogo (2002) critical values..

Significance levels: *** 1% , ** 5% and * 10% .

Finally, as a way to evaluate if the results are generalized even in the smaller labor market areas of the sample, we separatly estimate the model for the 50% smaller and 50% larger areas, which produces a population cutpoint of about 250,000.

Results for all robustness checks are presented in Table (1.8). Column (1) shows the effect of interest on a regression including the aforementioned variables. This result should be compared with the coeficient of the 2SLS estimation of table (1.5). The inclusion of amenities and geografical control variables reduces the human capital concentration effect on local wage indexes from 0.86% to 0.77%, implying that 11% of the effect comes from compensation for amenities accessibility and suggesting that a high parcel of college degree ocurrs in less ameable places. More importantly, the effect of interest remains highly significant. On column (2) we have results from estimation without Rio de Janeiro and São Paulo, as we can see, they remain unchanged relative to the baseline estimate, indicating that the impact of the concentration of human capital on wage rates is not by the two largest brazilian labor market areas. Results are show in the third column of Table (1.8) and in this case, we found that human capital externalities are slightly larger (0.89%) and still highly significant. Results from the last two column of Table (1.8) shown a slightly higher point estimate of human capital concentration effect on local wages in the smaller regics, which could be associated implying with the relative scarcity of human capital in those areas.

1.6 Concluding Remarks

As recent research about the external returns to human capital shows, these gains can be bigger than individual returns and, thus, can play an important role in explaining spatial urban inequalities. Recent research about urban agglomeration gains in Brazil indicates that they are significant , but these researches have not explored the importance of local concentration of human capital. The present paper, fills that gap, with the objective of identify and to measure the external returns to human capital concentration, defined by the share of local workes with at least college degree, in Brazil. In agreement with the avaible national and international literature, we find strong evidence of the existence of large external effects of concentration of human capital in Brazil, with a 0.86% baseline effect of human capital concentration on local wages.

Moreover, theoretical prediction is that marginal increase in the stock of human capital in a given region will have a greater positive effect on skilled workers wages than of unskilled workers only if the "externality effect" overcome the wage gains of unskilled workers added to the neoclassical "net effect". We find larger effect for the unskilled workers so we may infer that human capital externalities for the skilled group, although existing, are not strong enough to outweigh the gains in the wages of the unskilled when we also consider the neoclassical effect. We also investigated the differentiated effects among sectors, finding large and significant effects for manufacturing and commercer, with no effect for extractive industry and construction, in line with available international results.

Results were robust to a series of checks, such as inclusion of amenities related and geographical characteristics controls variables possibly correlated to wages, we still find a highly signicant human capital concentration effect 0.77%, a 11% reduction from baseline estimates. We repeat the baseline estimation using subsets of labor market areas, excluding Rio de Janeiro e São Paulo, consider only non-Metropolitan, smaller and larger areas. Results ranged from 0.88% in non-metropolitan areas to 0.83% for the smaller population areas.

The existence of external returns, as found in the present study, corroborates the importance of human capital investiment as a police for reducing regional inequality as a development strategy. By presenting evidence of human capital concentration on local wages, identifing another powerful explanation for wages differences across brazilian labor markets, while at the same time contributing to evidences of human capital externalities in developing countries

2. DOES A LONGER COMMUTING TIME INCREASE THE PROBABILITY OF URBAN VIOLENCE VICTIMIZATION? EVIDENCE FROM BRAZILIAN METROPOLITAN REGIONS

2.1 Introduction

With approximately 85% of its population living in urban areas in 2010, according to the Brazilian Demographic Census, the process of urbanization in Brazil is highly advanced. This high concentration of people in urban areas results in a wide range of implications and challenges for society, from the possibilities of economic gain from agglomeration economies to the necessity of urban planning and finding solutions to the questions of mobility and pollution. Actually, It would be difficult to exaggerate the importance of Brazilian urban violence and the long commute times experienced in the Brazilian metropolitan regions.

The United Nations Office on Drugs and Crimes (UNODC, 2012) ranks Brazil as one of the most violent countries in the world, with homicide rates at approximately 27.1 (homicides per hundred thousand inhabitants) in 2011, the third highest rate among the Latin American countries (behind Colombia and Venezuela). This state of affairs reflects a general situation of high rates of violence related to other types of crime in the country; for example, the UNODC (2012) numbers for 2010 put Brazil's rates (occurrences per hundred thousand inhabitants) for robbery and theft at 554.5 and 709.3, respectively, among the three most violent Latin American countries. The situation is even worse in the biggest cities, where the homicide rates can easily reach approximately 100 homicides per one hundred thousand inhabitants, according to information from the Ministry of Health (DATASUS, 2013), and the chances of victimization by robbery or theft are substantially higher in these places as well.

The problem of urban violence is neither the only substantive problem of Brazil's large urban centers nor is it dissociated from the other urban problems in these centers. In addition to the risk of being a victim of urban violence, visitors and inhabitants of Brazil's metropolitan regions must face the problem of low mobility. The poor quality of public transportation combined with indirect public subsidies for using individual modes of transport makes short-distance travel very time-consuming (IPEA, 2013). According to the information from Pesquisa Nacional por Amostra de Domicílio (PNAD, 2009), the average commute time for the inhabitants of Brazil's metropolitan regions was approximately 40.0 minutes in 2012, a high number when compared with metropolitan regions worldwide(Pereira & Schwanen, 2013; Silveira Neto *et al.* , 2015). As shown by Silveira Neto *et al.* (2015), the commute time of the metropolitan region of São Paulo is higher than observed for the metropolitan regions of New York and Seoul, for example.

In addition to the implied waste of potentially productive time and a lower quality of life for the inhabitants of Brazilian metropolitan regions (Oliveira *et al.* , 2015), a longer commute inherently causes greater exposure to potential aggressors because individuals must spend a greater amount of time on public streets and using public transportation. According to the theory of routines activities (Cohen & Felson, 1979; Cohen & Cantor, 1981), these circumstances increase the probability of being a victim of urban violence.

Although the empirical evidence confirm the influence of public exposure associated with a longer and more frequent commute on the chance of being a victim of urban violence (Lemieux, 2011; Lemieux & Felson, 2012), the question is much less investigated for developing countries and still unexplored for Brazilian metropolitan regions. The main objective of this paper is to investigate the existence of a causal relationship between commute time and the probability of being a victim of urban violence for individuals living in Brazilian metropolitan regions, that is, to determine if a longer commute time implies a higher chance of becoming a victim of urban violence for the inhabitants in these urban centers.

To obtain this evidence, we use a non-experimental method of matching individuals based on their propensity score (associated with commute time) and a unique Brazilian household survey from 2009 that simultaneously provides information about the commute times and victimizations. Our results suggest a causal relationship between commute time and the probability of being a victim of urban robbery in Brazilian metropolitan regions, but no significant effect on theft victimization. These results are robust for different checks and a sensitive analysis.

In the next section, we present and discuss the evidence on the relationship between public exposure and its association with commute time and victimization. In section three, we present our empirical strategy. In section four, we discuss the database and initial evidence. Section five presents the main results, including the robustness checks and simulation-based sensitivity analysis. Section six comprises the concluding remarks.

2.2 Theoretical Background and Available Evidence

According to the sociological theory of routines activities (Mayhew *et al.* , 1976; Cohen & Felson, 1979; Cohen & Cantor, 1981), a greater amount of exposure to public spaces with low or ineffective guardianship creates more favorable conditions for victimization. As discussed in Cohen & Felson (1979), most criminal acts require the convergence of space and time of likely offenders and suitable targets and the absence of capable guardians. This theory focuses on the circumstances in which the criminal acts are perpetrated and how activities outside the home affect these circumstances. The opportunity theory of Cohen & Cantor (1981) claims that the key to understanding why income, race, and age appear to affect the likelihood of victimization is focusing on the mediating roles played by five factors: exposure, guardianship, proximity to potential offenders, attractiveness of the potential targets, and certain properties of the types of the crimes.

From these perspectives, a longer commute, including the time waiting for public transportation, generates a greater chance of the convergence in space and time of potential victims and likely offenders. Because this set of circumstances reduces the costs of looking for potential victims and, in an urban environment with low guardianship, criminal action, this convergence also changes the economic incentives for committing crimes, favoring the occurrence of crimes such as robbery and theft (Becker, 1968; Heineke, 1987). In other words, in an urban environment with low guardianship, a longer exposure to public spaces generated by a longer commute implies more frequent encounters between potential victims and criminals, reducing the costs associated with committing a crime.

This line of argument was originally created to explain the paradox of the improvements in social indicators accompanied by an increase in crime in the United States in the 1970s (Cohen & Felson, 1979). A similar paradox was observed in Brazil: In the 2000s, Brazil's undeniable social development was not followed by a clear reduction in crime rates. Using official data, Waiselfisz (2014) demonstrated, for example, that homicide rates increased from 25.8 in 2005 to 29.0 in 2012, and according to Corbacho *et al.*

(2015), public opinion polls in Latin America, including Brazil, in 2010, demonstrated that nearly 30% of respondents pointed to crime as the most critical problem in their country. Thus, as commute time increased during the 2000s, the increase in the exposure to public spaces, in streets and inside vehicles, generated by a longer daily commute, can be part of the explanation for the resilience of crime.

Empirical evidence on the influence of exposure to public spaces on victimization strongly supports the routine activities theory (Cohen & Cantor, 1981; Messner & Blau, 1987; Miethe *et al.* , 1987), but specific evidence about the influence of commute time on the probability of victimization is not abundant, although some explicitly worry about insurance for public transportation systems (Clarke, 1996; Smith & Clarke, 2000). Evidence confirms the aforementioned expectation. F & Minor (2002), using US census information, found an inverse relation between accessibility to jobs and violent crime in the city of Cleveland, Ohio. Lemieux (2011) and Lemieux & Felson (2012), using data from the National Crime Victimization Survey and American Time Survey, built time-adjusted measures of exposure to violent attacks that demonstrated that greater risk occurs during travel between activities, specifically, while commuting to work and school.

The evidence is still scarcer for developing countries, but the available results again appears to confirm the importance of a greater public exposure associated with a longer commute for increasing the chance of victimization. Messner *et al.* (2007), using a set of unique data on victimization for the city of Tianjin, China, and after controlling for the influence of demographic and other lifestyle variables, showed that more frequent traveling to work outside the city increased the risk of being a victim of urban theft. Focusing on Mexico City Metropolitan Area, Vilalta (2011) showed that fear of crime was higher for those traveling more than 30 minutes. On Nigerian highways, Badiora *et al.* (2015) identified robbery and theft as prevalent transit crimes and asserted that individuals moving in isolated area were most likely to be subject to a high crime risk.

As for Brazil, a few studies have observed that specific measures of exposure influence the probability of being a victim of urban violence. Beato *et al.* (2004) found that in the city of Belo Horizonte, the use of public transportation has a positive influence on the probability of being a victim of theft or robbery. Peixoto *et al.* (2007), using victimization data from the Brazilian cities of Rio de Janeiro, Recife, São Paulo, and Vitória, demonstrated that individuals who go outside of their home daily or weekly have a higher chance of being the victims of urban theft. Justus & Kassouf (2013), using victimization surveys from the city of São Paulo, found that working outside the home increases the probability of being a victim of robbery or theft. However, to the best of our knowledge, studies have neither provided evidence about the causal influence of a longer commute time on the chance of being a victim of urban violence nor considered a set of all Brazilian metropolitan regions.

2.3 Empirical Strategy and Data

2.3.1 Econometric Approach

In this paper, we estimated the causal effect of commute time on the probability of being a victim of urban violence in Brazilian metropolitan regions, for individuals who spend more than a certain amount of time commuting. A more traditional linear econometric specification for obtaining the effect of commute time on the chance of victimization would be as follows:

$$Y = \alpha + \beta C + X\gamma + \epsilon, \tag{2.1}$$

Where Y is an outcome related to victimization; C is an indicator for a longer commute time, a dummy value 1 for an individual with a long commute time and 0 otherwise; X is a set of control variables that affect the chance of victimization; and ϵ is an error term. From this perspective, the estimative of β would correspond to the effect of a longer commute time on the chance of victimization. The problem with this type of approach to non-experimental data is that it is not possible to guarantee that the error term is uncorrelated to the variable measuring the effect of commute time on victimization (C), which can make the Ordinary Least Square (OLS) estimative inconsistent and biased. Using non-linear models for categorical variables (e.g., logit and probit models) and a maximum-likelihood approach does not solve this problem and imposes particular distributions to the residual term, rendering the parametric modeling

Thus, as we have a rich set of variables that influence the commute times of individuals and the precise determinants of victimization are a much broader set, we match individuals based on their propensity score associated with their category of commute time (Angrist & Pischke, 2008) and use a semiparametric approach. Our identification strategy is based on the conditionally independence or unconfoundedness assumption (Rubin, 1974; Heckman & Robb, 1985) and the propensity score theorem of Rosenbaum & e Rubin (1983). If we denote by Y_i the observed result of an individual i for our outcome variable, the probability of being victim of urban violence, Y_i^0 Y_i^1 the potentials results of, respectively, taking the treatment (spending more than certain time in commuting to work) or not respectively, we have

$$Y_i = C_i Y_i^1 + (1 - C_i) Y_i^0. \tag{2.2}$$

The uncounfoudedness assumption implies that, conditioned on a set of individual's variables X_i , the potential results are independent of being assigned to treatment, i.e., $Y_i^1, Y_i^0 \perp C_i | X_i$, where $\perp$ means independence. As shown, for example, in Angrist & Pischke (2008), this allows to obtain the effect of the treatment (in our case, a longer commuting time) as the difference in means of the outcome variable (in our case, the probability of being victim of urban violence) by status at each at each value of X_i.

We use the propensity score theorem of Rosenbaum & e Rubin (1985) to obtain the effect of the treatment on the treated individuals. Specifically, we use the fact that $Y_i^1, Y_i^0 \perp C_i | X_i$ implies $Y_i^1, Y_i^0 \perp C_i | p(X_i)$, where $p(X_i)$ correspond to the probability of being treated or the propensity score (in our case, taking longer than a certain time for a home-to-work commute). Using this theorem and the unconfoundedness assumption, it is possible to obtain the effect of a longer commute time on the probability of being victim of urban violence in Brazilian metropolitan regions of long commuters (β_{ps}) as

$$\beta_{ps} = E\{E[Y_i | p(X_i), C_i = 1] - E[Y_i | p(X_i), C_i = 0]\}. \tag{2.3}$$

To obtain the sample correspondence of equation (3.3), we estimate $p(X_i$ using a logit model and two ways of matching the treated with the controls, the nearest-neighbor (one treated with one control), based on the estimative of the propensity score, and kernel estimation for weighting controls according to the propensity scores (one treated with weighted controls). For both types of matching, we have the general empirical estimative given by

In order to obtain the sample correspondence of equation (3.3), we estimate $p(X_i)$ using a logit model and use two ways of matching the treated with the controls, the nearest neighbor (one treated with one control) based on the estimative of propensity score and the kernel estimation for weighting controls according to propensity scores (one treated with weighted controls). For both types of matching, we have the general empirical estimative given by

$$\hat{\beta}_{ps} = \frac{1}{N_1} \sum_{i \in N_1} (Y_i^1 - \sum_{j \in N(i)} w_{ij} Y_{ij}^0), \tag{2.4}$$

where N_1 is the number of long commuters, $N(i)$ is the set of controls unities matched to long commuter i (equal to 1 in the case of the nearest-neighbor), w_{ij} is a weight given to the control unit j of the matched long commuter i (equal to 1 in the case of the nearest neighbor, and given by a kernel function in the case of a kernel estimation for weighting controls).

To verify the robustness and confidence of the results, we provide a set of additional results: by considering others' estimators, specific samples (for example, specific locations and the period of work-day), and a sensitive analysis to influence the types of non-observable factors that could affect the results.

2.3.2 Database

Our data is obtained from the PNAD, a traditional annual household survey implemented by the Instituto Brasileiro de Geografia e Estatística, the Brazilian Institute of Geography and Statistics. With a sample design to represent all individuals and households of states and metropolitan regions, this survey collects directly (face-to-face) a very rich set of information about personal, family, and household characteristics as well as labor market conditions (PNAD, 2009)

Particularly, the 2009 PNAD also presents information about individual victimization and length of commute time. The victimization information indicates if an individual was the victim of a robbery, theft, or physical aggression between September 27, 2008 and September 26, 2009. Robbery is defined as an action accompanied by physical violence or threat of violence (with or without a weapon) and motivated by the desire to take any property of value from the victims. Theft presents a similar motivation but represents an action without the offender confronting the victim. Physical aggression includes aggression toward victims with any type of motivation.

Thus, working with an outcome variable associated with urban violence that represents the mean of the probability of being the victim of each of these types of urban violence is possible. Because some thefts are only perceived after some time and might even be non-identified as this crime and physical aggression can occur at home and may be non-reported as such, we believe that robbery is less susceptible to information error and, thus, emphasized its results; however, we also present some results for theft occurrence. As non-economic factors are also a motivator for physical aggression, we do not consider this type of urban violence in this paper.

As for the variable that represents a longer commute time, we observe that the PNAD data separates commute time into four categories: up to 30 minutes, more than 30 minutes and up to 1 hour, more than 1 hour and up to 2 hours, and more than 2 hours. Notably, this information includes the time when effectively moving (inside vehicles) and when waiting for public transportation or making a transfer from one vehicle to another.

From this information and because the average commute in the Brazilian metropolitan regions was 34.6 minutes in 2009 (just over the higher limit of the first category), we created an indicator that assumes the value of 1 if the individual takes more than 1 hour to commute from their residence to their work location and 0 if this commute is less than or equal to 1 hour. In addition to allowing a meaningful distinction between individuals' commute time, theoretical reasons also justify this choice. Specifically, in an urban environment with low public guardianship, that is, a community or government that does not resolve the problems their residents incur in the public environment, this longer public exposure increases the probability of these individuals being in situations (inside and outside vehicles) where the additional spending required to avoid being a victim of crime is very high[1].

Our sample includes individuals from all ten official Brazilian metropolitan regions[2]. Considering only those aged 10 years or older who must commute to work, we have 53,303 observations for the year 2009. For each of these, we have an extensive set of variables that potentially includes determinants of victimization and of commute time.

We match individuals based on their probability of being a long commuter (propensity score matching). To determine this probability, we consider a large set of conditionings of commute time, including personal (gender, race, age, marital status, income, and education), family and household characteristics (family size and car ownership, and access to sanitation, garbage, and piped water), residential location (metropolitan regions), and employment characteristics (sectors of occupation and the occupation position). This set of variables is based on the theoretical and empirical literature of conditionings of commute time.

As demonstrated by Fujita (1989) and Duranton & Puga (2015), for example, because they affected the trade-off between access to jobs and residential space, individual, family, and household characteristics are principal determinants of residential location and, thus, of commute time. By contrast, different spatial requirements and competition across economic sectors also affect the location of sectors' jobs and, thus, an individual's commute time.

McCann (2013) argued, for example, that because the price of urban space varies across an urban area and a sector's activities demand different space requirements, job location tends to vary according to a sector's activities. Mulligan & Fik (1994) and Anderson and Engers (1994) have argued that sectors with a higher demand-price elasticity do not tend to agglomerate in an urban space or in the main CBD (Central Business District). More specifically, Konishi (2005) also showed that the spatial concentration in urban centers of some activities such as retail is the result of two forces: more concentrated firms tend to attract consumers who are uncertain about their tastes and expect low

[1]For a robustness check we also consider other long commute indicator that assumes value equal to 1 if the individual takes more than 30 minutes from his residence to his work location and equal to 0 if this commuting time is less then 30 minutes; results are available in the appendix.

[2]Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba, Porto Alegre and Brasília

prices; however, this situation lowers the price for the firms.

Empirical works in different contexts, including Brazil, have confirmed that these two sets of characteristics (individual, family, and household; a sector's location) are crucial to understanding the an individual's commute time (Crane, 2007; Figueroa *et al.* , 2014; Silveira Neto *et al.* , 2015; Edlund *et al.* , 2015; Fan, 2017).

Available evidence has also indicated that different urban structures are an important conditioning of commute time (Bento *et al.* , 2005; Gordon & Lee, 2013); thus, we consider the fixed effects of metropolitan areas. In addition, Silveira Neto *et al.* (2015) demonstrated that an individual's job characteristics (occupation position) are important conditionings of the commute time in the metropolitan region of São Paulo. These authors have also provided evidence that the availability of household infrastructure services provides information about the residential location in Brazilian urban centers and, thus, is associated with household heads' commute time.

2.4 Descriptive statistics and propensity score matching

In Table (2.1) we present the sample total of the victims of robbery and theft in the Brazilian metropolitan regions by place of occurrence, information potentially associated with exposure, and regular commute time. Limiting our consideration to individuals 10 years or older and eliminating missing observations, from our initial sample of 53,303 residents from the Brazilian metropolitan regions, 4,794 were victims of robbery (9.0%) and 2,596 were victims of theft (4.9%). Considering that the proportion of people 10 years or older who had been victims of robbery or theft was 2.3% in the rural areas, these numbers demonstrate a much more violent environment in the urban areas in Brazil.

Table 2.1: Distribution of Victimization by place of occurence – Brazilian Metropolitan Regions

Where	Robbery	%	Theft	%
Own house or a third person house	304	6.34	834	32.13
Commercial Estabishment	376	7.84	379	14.60
Public Way	3,600	75.09	976	37.60
Teaching Establishment	15	0.31	46	1.77
Public Transportation	442	9.22	217	8.36
Gynnasion or Sports Stadium	6	0.13	10	0.39
Other	51	1.06	134	5.16
Total	4,794	100.0	2,596	100.00

[1] Source: Author's calculation based on PNAD 2009 microdata.

The percent of robberies that occurred inside residences, in public transportation, and public thoroughfares, was 6.3%, 9.2%, and 75.1%, respectively (Table (2.1)). Given that the majority of the working population in the Brazilian metropolitan regions uses thoroughfares or public transportation to go to work, these data appear consistent with a positive relationship between commute time and the chance of being a victim of robbery. Notably, the information for the crime of theft is more evenly split among the categories of location than robbery. In particular, we observe that more than 30% of thefts occurred in residences. This is consistent with the, apparently, less crucial role of commute time in explaining this type of victimization in Brazilian metropolitan regions.

In Table (2.2), we present the distribution of victims and non-victims of robbery and theft across the commute time categories for our sample of individuals in the Brazilian metropolitan regions. According to the numbers, the shares of victims of robbery and theft in the population were, respectively, 0.09 and 0.049. Although there is not any difference between the shares of individuals with more than two hours of commute time for robbery victims and non-victims, the numbers of Table (2.2) also indicates that the percentage of robbery victims with more than one hour of commute time (15.9%) is higher than for non-victims (13.9%), a difference that is statically significant at 5% level of significance. Notably, the numbers presented in Table (2.2) do not suggest the same type of relationship for theft as it suggests for robbery.

Table 2.2: Number of Victims by Commuting Time

Commuting time (Ct)	Robbery				Theft			
	Non-Victms		Victms		Non-Victms		Victms	
	N.	Share	N.	Share	N.	Share	N.	Share
$Ct \leq 30$ min	26,354	0.543	2,533	0.528	27,445	0.541	1,442	0.555
30 min $< Ct \leq 1$ hour	15,410	0.318	1,497	0.312	16,114	0.318	793	0.305
1 hour $< Ct \leq 2$ hours	5,874	0.121	678	0.141	6,244	0.123	308	0.119
$Ct > 2$ hours	871	0.018	86	0.018	904	0.018	53	0.020
Total	48,589	0.910	4,794	0.090	50,707	0.951	2,596	0.049

[1] Source: Author's calculation based on PNAD 2009 microdata.

The association between a longer commute time and victimization can, obviously, be potentially explained by variables that affect both conditions such as gender: if men spend more time commuting and are less risk averse than women. Thus, our estimative of the influence of a longer commute time on victimization is based on a propensity score matching of individuals.

As the individuals were not distributed randomly between the two groups with different commute times and the set of control variables balanced between these groups, the first group of individuals with a short commute in Table (2.3) (unmatched sample control) are not an acceptable set of counterfactuals to the individuals with the long commute. To minimize this problem, we implement matching based on the propensity score estimates and generate a new estimative for the effect of a long commute time on the probability of being the victim of robbery or theft in Brazilian metropolitan regions. Our expectation is that by balancing the sets of variables that determine the commute time of the individuals between the two groups, we can eliminate or at least minimize the influence of the potentially omitted determinants of victimization correlated with the individuals' commute times.

For the propensity score estimative, as discussed in section three, we use the set of variables presented in Table (2.3): personal characteristics, residential characteristics, family and household characteristics, variables for sector of activities and type of job position, and identification of the metropolitan region[3]. As this paper is concerned with the effect of a long commute time on the risk of victimization, we also consider victimization only in public transportation and public thoroughfares[4], which implies a final sample of 51,556 individuals. In Table (2.3), we present the variables we used to obtain the propensity score estimative. The sets of variables are presented for the long commuters (LC), control groups (Control), and the samples of both unmatched and matched individuals, whereas the matching is performed using the nearest-neighbor criteria.

The t-statistic presented in Table (2.3) was used to test the difference in values between the treated and controls. Although the differences are statistically significant for the unmatched sample, after comparing the long commuters (LC) with their respective nearest-neighbor based on the propensity score estimative, no difference appears statistically significant at 1%. This result means that the set of characteristics is well balanced between the LC and controls, a necessary condition for measuring the effect of commuting on the probability of being a victim of violence. Note that, except for the case of the variable "dependence" (non-working to total members of the family ratio), where the difference was very small before the matching, we obtained bias reduction for all other variables above 19%.

2.5 Results

[3]For brevity reasons, we do not present the estimative for the logit model of the determinants of commute time; they are availed upon request.

[4]We thank an anonymous referee for this suggestion. The results considering victimization in a broader set of circumstances are similar and are availed upon request.

Table 2.3: Comparisons between Long Commuters (Treated) and Short Commuters (Control) in the Original (Unmatched) and the (Nearest-neighbor) Matched sample.

Variable	Unmatched Sample				Matched Sample				
	Treated	Control	*t*-test	% bias	Treated	Control	*t*-test	% bias	% bias Red.
gender	0.56	0.50	0.24	10.8	0.56	0.53	1.76	5.0	53.9
race	0.43	0.48	-7.02	-8.90	0.43	0.43	0.55	0.9	89.8
age	36.47	36.73	-1.67	-2.10	36.47	36.26	-1.08	-1.70	19.10
age squared	1468.3	1500.9	-2.61	-3.40	1468	1448	1.35	2.10	36.5
single	0.45	0.44	0.29	0.40	0.45	0.44	0.45	0.70	-102.0
highschool	0.45	0.41	6.20	7.8	0.45	0.45	-0.42	-0.70	91.1
college	0.11	0.15	-9.56	-12.7	0.11	0.11	-0.69	-1.13	91.6
Household Income	772.4	1016.0	-12.5	-18.4	772.4	773.4	-0.06	-0.10	99.6
Dependency	0.71	0.65	6.59	8.10	0.71	0.70	0.73	1.20	84.9
car	0.38	0.45	-10.6	-13.5	0.38	0.39	-0.94	0.10	88.6
family size	3.48	3.39	4.90	6.20	3.48	3.48	-0.30	-1.0	91.9
					Working Sector				
Industry	0.13	0.14	-1.12	-1.40	0.13	0.14	-0.46	-0.80	46.60
Construction	0.11	0.08	6.57	8.00	0.11	0.10	1.37	2.40	70.50
Commerce	0.16	0.21	-9.93	-13.0	0.16	0.16	-0.18	-0.30	97.80
Public Administration	0.06	0.07	-2.94	-3.80	0.06	0.06	0.54	0.90	77.60
Informal	0.19	0.22	-5.86	-7.60	0.19	0.18	0.79	1.30	83.30
Self Imployed	0.08	0.14	-14.9	-20.5	0.08	0.08	-0.30	0.00	99.80
					Infraestructure				
Sanitation	0.68	0.62	9.46	12.1	0.68	0.68	0.9	-0.59	92.10
Garbage	0.88	0.88	-0.86	-1.10	0.88	0.88	0.00	0.18	72.30
Piped Water	0.93	0.92	1.55	2.00	0.93	0.93	-0.22	0.19	84.30
					Metropolitan Regions				
Belem	0.04	0.07	-7.44	-10.1	0.04	0.04	0.33	0.50	95.20
Fortaleza	0.09	0.11	-5.34	-7.10	0.09	0.08	0.90	1.40	80.20
Recife	0.08	0.09	-2.04	-2.60	0.08	0.07	1.00	0.60	38.80
Salvador	0.10	0.12	-4.15	-5.40	0.10	0.10	0.14	0.20	95.90
Belho Horizonte	0.09	0.10	-1.06	-1.30	0.09	0.09	0.11	0.20	86.10
Rio de Janeiro	0.20	0.11	21.3	24.6	0.20	0.20	0.50	0.90	96.30
Curitiba	0.05	0.06	-4.94	-6.50	0.05	0.05	0.43	0.70	89.90
Porto Alegre	0.05	0.14	-20.8	-29.9	0.05	0.06	-0.73	-0.90	96.90
Distrito Federal	0.07	0.08	-4.34	-5.7	0.07	0.07	1.39	-2.20	60.40
Observations	7,262	44,294			7,509	44,294			

Note: The standardised bias is the difference of the sample means in the treated and non-treated (full or matched) sub-samples as a percentage of the square root of the average of the sample variances in the treated and non-treated groups, see Rosenbaum and Rubin (1985).

2.5.1 Baseline Estimates

Table (2.4) presents the results of the propensity score matching estimations. Panel A comprises the results for robbery using both the nearestneighbor and kernel-weighting approaches to matching. Panel B comprises the results for theft using the same two criteria for matching. Independent of the propensity score based matching criteria, our results indicate that a long commute is associated with a higher probability of being victims of robbery.

Specifically, a long commute generates a 2.5-perceptual point increase in the probability of being a victim of robbery according to the nearest-neighbor estimative, and a 2.0-perceptual point increase according to the kernel estimative. These results mean a significant influence of a longer commute time; using the nearest-neighbor estimative (of 0.025), the evidence indicates that approximately 27% of an individual's chance of being a victim of robbery (0.092) is attributed to their longer commute time.

Table 2.4: Propensity Score Matching Results for Robbery and Theft

	Sample	Treated	Control	Diff	St Err.	Bootstrap St Err
Panel A:	**Robbery**					
N. N. Matching	Unmatched	0.092	0.074	0.017***	0.003	
	Matched	0.092	0.066	0.025***	0.005	0.006
Kernel Matching	Unmatched	0.092	0.074	0.017***	0.003	
	Matched	0.092	0.042	0.020***	0.004	0.003
Panel B:	**Theft**					
N. N. Matching	Unmatched	0.026	0.022	0.004	0.002	
	Matched	0.026	0.022	0.004	0.003	0.003
Kernel Matching	Unmatched	0.026	0.022	0.004	0.002	
	Matched	0.026	0.022	0.004	0.002	0.002

[1] bootstrap standard erros where calculated using 200 replications for nearest neighbor matching and 50 replications for kernel matching,
[2] Signifance Levels: * 10%, ** 5% and *** 1%.

As for theft, our estimative does not indicate any positive effect for a longer commute time on the probability of being a victim, independently of the propensity score based matching criteria of proximity we use. The explanation of this result, at least in part, is probably that this type of crime tends to be much more dependent on specific circumstances and is more susceptible to information error.

The crime literature focuses on gender differences, especially in the case of public exposure (Cohen & Felson, 1979). Given that females could be considered more vulnerable in the eyes of potential offenders, we also performed estimations separated by gender Table (2.5).

Independent of the method of estimation and gender, we obtained a positive and statistically significant estimative for the effect of a long commute time on the probability of being a victim of robbery. In addition, considering the propensity score estimative

Table 2.5: Effect of Long Commuting: Gender Differences

	Panel A: Men		Panel B: Women	
Estimation	Robbery (1)	Theft (2)	Robbery (1)	Theft (2)
PS Nearest Neig. Matching	0.025***	0.003	0.022***	0.009
	(0.006)	(0.004)	(0.008)	(0.004)
	[0.007]	[0.004]	[0.009]	[0.005]
PS Kernel Matching	0.017***	0.002	0.024***	0.007***
	(0.005)	(0.003)	(0.006)	(0.003)
	[0.005]	[0.002]	[0.005]	[0.003]

[1] Note: On the linear probability model and logit estimates we use all controls use in the estimates of Table (2.3)
[2] Robust standard errors are presented in parentheses; bootstrap standard erros where calculated using 200 replications for nearest neighbor matching and 50 replications for kernel matching, they are presented in brackets.
[3] Signifance Levels: * 10%, ** 5% and *** 1%.

and using bootstrap standard deviations, a t-type test of equality of the effects (at significance level of 5%) indicates that we did not observe a significant difference in this effect between men and women.

Regarding the effect of a long commute time on the probability of being a victim of theft, we found a positive but negligible effect for women. However, using a t-type test for equality of the effects (based on bootstrap standard deviations and significance level of 5%), we note that this effect is not statistically different than the one obtained for men. Thus, this set of results indicates no gender difference related to the effect of a long commute time on the probability of being a victim of robbery or theft in Brazilian urban centers. Although a definitive explanation is not possible here, this evidence is consistent with potential less careless behavior of women compared with men[5].

2.5.2 Robustness Checks

In this subsection, we test for the robustness of our result by generating additional estimative. First, we use traditional Linear Probability Model (LPM) and logit specifications for the sample of matched individuals. Second, following Abadie & Imbens. (2002), instead of using only the propensity scores, we generate an estimative by matching individuals based on the set of variables that are potentially associated with commute time. Third, we generate an estimative using a restricted sample that includes only individuals living

[5] Consistent with this perspective, Perreault (2015) found higher rates of victimization of robbery and theft for men and women.

in residences with a complete infrastructure of household services. Fourth, we consider an estimative for a sample of individuals with specific work shifts. Finally, we control for the sense of security of individuals in their homes, neighborhoods, and cities.

In the first robustness check, we followed the suggestion of Crump *et al.* (2009) and used the sample obtained by matching each long commuter to their nearest-neighbor based on a propensity score estimative in traditional OLS and logit specifications. By using only observations of controls with a common support, we can eliminate the influence of observations without overlap in the covariates' distributions between these two groups. The first two rows of Table (2.6) present the new estimates of the effect of a long commute on the probabilities of being a victim of robbery and theft. Apart from a small reduction in the effect of a long commute time on the probability of being a victim, the results for victimization by robbery are basically the same as presented in Table (2.4), namely, a positive and statistically significant effect. As for the impact of a longer commute on the probability of being a victim of theft, the set of evidence indicates a positive but negligible effect.

Table 2.6: Robustness Check

Outcome:	Robbery	Theft
Estimation	(1)	(2)
LPM	0.020***	0.006**
	(0.004)	(0.002)
Logit	0.019***	0.006***
	(0.003)	(0.002)
Bias Ajusted Variable Matching	0.026***	0.004
	(0.005)	(0.002)
	[0.006]	[0.003]

[1] Note: On the linear probability model and logit estimates we use a restricted subsample of individuals with a complete set of household infraestructure (sanitation, garbage pickup and piped water), and all controls use in the estimates of Table (2.3)
[2] For the bias ajusted variable matching the coeficient corresponde to the sample average treatment effect.
[3] Robust standard errors are presented in parentheses; bootstrap standard erros where calculated using 200 replications for nearest neighbor matching and 50 replications for kernel matching, they are presented in brackets.
[4] Signifance Levels: * 10%, ** 5% and *** 1%.

Our second robustness check applies the bias-corrected matching estimator proposed by Abadie & Imbens. (2002). Instead of matching by only propensity scores, this estimator matches the observations based on all the sets of variables presented in Table (2.3). The estimative of the effect of a long commute time on the probability of victimization by robbery and by theft obtained using this estimator are presented in the third row of Table (2.6). The estimative is quite similar to those obtained using the propensity score matching and nearestneighbor-proximity criteria (see Table (2.5)); specifically, a long commute time implies a 2.6-perceptual point increase in the probability of being a victim of robbery but has no effect on the probability of being a victim of theft.

The next robustness check applies to the propensity score matching using a sample of individuals whose residences present the complete set of household infrastructure services: sewer, piped water. and garbage collection. The purpose of this check is to verify if our results only reflect the more violent and poorer neighborhoods located on the fringes of Brazil's metropolitan regions, which commonly have less access to household infrastructure services and where residents spend more time commuting. The results are presented in Table (2.7), Panel A.

Notably, the estimative of the influence of a long commute time on the probability of being a victim of robbery present a small reduction but continue to be positive and statistically significant. The result indicates that this potential source of influence cannot explain our main previous results. More specifically, even after eliminating the most critical differences associated with the household infrastructure services, which means discarding the poorest neighborhoods located on the fringes of the urban centers, using propensity score matching based on the nearest-neighbor, we estimate that a longer commute time implies an 2.2-perceptual point increase in the probability of being a victim of robbery in the Brazilian metropolitan regions.

According to Felson & Poulsen (2003), crime occurrence varies significantly by time of day. Thus, in the next robustness checks, we restricted the sample to individuals with work shifts only from 5 AM to 10 PM. We were concerned about the possibility that individuals with long commute times, compared with short commute times, were also more exposed to night periods. Unfortunately, due to dataset restrictions, a direct assessment of which individuals worked day shifts was not possible. However, combining the information about work shift period with the most common schedule of working 40 hours per week, it is possible to attenuate this limitation.

Finally, to eliminate, or at least attenuate, the potential problems associated with the failure to observe individual behaviors that could directly affect the probability of becoming a victim, we included additional information about an individual's sense of security in their own home, neighborhood, and city. This information was available in the PNAD database for 2009; specifically, in this survey, individuals were required to respond to questions about their sense of security at home, in their neighborhood, and in their city.

Perhaps, for example, individuals who feel secure might exhibit more careless behavior when commuting, something we cannot observe, which might render them more susceptible to victimization. By contrast, because identifying the exact address of individ-

Table 2.7: Robustness Check 2

Outcome:	Robbery	Theft
Panel A: Complete Infraestructure	(1)	(2)
PS Nearest Neigbor Matching	0.022**	0.008*
	(0.005)	(0.003)
	[0.007]	[0.004]
PS Kernel Matching	0.018**	0.006**
	(0.004)	(0.003)
	[0.004]	[0.002]
Panel B: up to 40 weekly hours day shift		
PS Nearest Neigbor Matching	0.022*	0.001*
	(0.008)	(0.04)
	[0.009]	[0.006]
PS Kernel Matching	0.020**	0.005
	(0.006)	(0.003)
	[0.004]	[0.003]
Panel C: Including Perception of Safety		
PS Nearest Neigbor Matching	0.014**	0.004
	(0.005)	(0.003)
	[0.006]	[0.003]
PS Kernel Matching	0.019***	0.004
	(0.004)	(0.002)
	[0.005]	[0.004]

[1] All estimations consider the greater then one hour commuting as treatment.
[2] Robust standard errors are presented in parentheses; bootstrap standard erros where calculated using 200 replications for nearest neighbor matching and 200 replications for kernel matching, they are presented in brackets.
[3] Signifance Levels: * 10%, ** 5% and *** 1%.

uals' residences is not possible, the perception of safety could also provide an indicator of real existing security in their neighborhood and city; that is, the individuals feel safer because, in fact, they live in a location with greater guardianship levels.

We included these new data about individuals' sense of security in their home, neighborhood, and city as additional variables affecting the propensity score estimative of a long commute time. Next, we obtained new matching. The results are presented in Table (2.7), Panel C. Notably, the estimative are similar to those presented in Table (2.4), suggesting that the declared sense of security or perception of risk is not sufficient to explain the relationship between the long commute time and victimization we obtained.

2.5.3 Simulation-based sensitivity analysis

Even though the aforementioned matching quality and robustness check results support the validity of our propensity score matching estimates, these results rely on the conditional independence assumption (CIA). As this identifying assumption is non-testable by nature, one may still question its plausibility in our case and argue that our results are being affected by an omitted variable strongly correlated with commuting duration.

To remove this suspicion, we apply the simulation-based sensitivity analysis proposed by Ichino & Nannicini (2008). This analysis assesses the bias of our estimates when the CIA is assumed to fail in some specific ways. This analysis assesses the bias of our estimates when the CIA is assumed to fail in specific ways. A failure in the CIA implies $Pr(C = 1|Y_0, Y_1, X) \neq Pr(C = 1|X)$, that is, the assignment to treatment is not unconfounded for a given X, even when adding the assumption that the CIA holds for a given X and an unobserved binary covariate U, implying $Pr(C = 1|Y_0, Y_0, X, U) = Pr(C = 1|X, U)$. Even though U is a unobservable confounding factor, Ichino & Nannicini (2008) proposes a characterization of it's distribution, obtained by specifying the following parameters $p_{ij} \equiv Pr(U = 1|C = i, Y = j, X) = Pr(C = 1|C = i, Y = j); i, j \in \{0, 1\}$, which define the probability that $U = 1$ in each of the four groups defined by the treament status (C) and the outcome value (Y).

Ichino & Nannicini (2008) implement the sensitivity analysis by estimating a logit model of $Pr(Y = 1|C = 0; U; X)$ in every iteration, with the effect of U on the relative probability to have a positive outcome in the case of no treatment (the observed "outcome effect" of the simulated U). This is the average estimated odds ratio of the variable U, denoted by Γ. Similarly, by estimating the logit model of $Pr(C = 1|U; X)$, the average odds ratio of U would measure the effect of U on the relative probability of being assigned to the treatment $T = 1$ (the observed "selection effect" of U) and is denoted by Λ . Following this reasoning, we proceeded with the sensibility analysis, calibrating 0 to mimic other observed covariates and considering parameter specifications that would ultimately drive the effect of a long commute time on the chance of being a victim to zero, and assess its plausibility (Table (2.8)).

As presented in Table (2.8), an unobserved confounder U , similar to the observable covariates, would not suffice to reduce the effect of a longer commute time (ATT) to zero. By contrast, the effect remains virtually unaltered, which is plausible given the small outcome and treatment effects of these unconfounding factors. The confounding factor U required to reduce the effect of commuting on the chance of being a victim to zero would have $\Gamma = 2.3$ and a $\Lambda = 7.0$. More precisely, U must increase the relative probability of having Y (probability of being a victim of robbery) above the mean by a factor greater

Table 2.8: Sensitivity Analysis: Effect of Calibrated Confounders

	$Pr(U = 1 \| C = i, Y = j)$				Γ	Λ	ATT	SE
	p11	p10	p01	p00				
No confouder	.00	.00	.00	.00	-	-	0.025	0.005
Neutral Confouder	.50	.50	.50	.50	1.003	0.997	0.023	0.006
Confouder like								
Gender $(male = 1)$	0.55	0.57	0.58	0.57	1.052	1.009	0.022	0.006
Race $(white = 1)$	0.42	0.44	0.43	0.41	0.745	0.839	0.022	0.006
Single	0.39	0.45	0.34	0.45	0.638	1.000	0.023	0.006
Highschool	0.46	0.44	0.45	0.40	1.220	1.171	0.023	0.006
College	0.11	0.11	0.14	0.15	0.898	0.684	0.022	0.006
Own a Car	0.33	0.38	0.37	0.45	0.704	0.762	0.022	0.006
Sanitation	0.64	0.68	0.59	0.63	0.853	1.286	0.023	0.006
Garbage	0.88	0.88	0.86	0.89	0.789	0.964	0.024	0.006
Pipewater	0.92	0.93	0.90	0.92	0.758	1.082	0.023	0.006
“killer confounder”	0.80	0.80	0.55	0.35	2.277	7.021	-0.001	0.006

[1] Neutral Confouder in the sense that set of the effect on the untreated outcome is zero $(p01 - p00 = 0)$ and the effect on the selection into treatment is also zero $(p1. - p0. = 0)$.

[2] Γ is the average estimated odds ratio of U in the logit model of $Pr(y = 1|C = 0; U; X)$; Λ is the average estimated odds ratio of U in the logit model of $Pr(C = 1|U; X)$;“ATT” is the average of the simulated ATTs; “SE” is the standard error calculated as shown in Ichino et al., 2008.

than 2.3, and the relative probability of being treated (being a long commuter) by 7.0 times. The presence of a confounder with similar characteristics among the unobservable factors is considered implausible, given that important covariates for crime victimization, such as gender and race, had little effect on the simulated U. These simulation exercises support the robustness of the matching estimates.

2.6 Concluding Remarks

Because their associated impact on urban life quality, urban violence and long commuting time to work location are certainly among the biggest urban problems of Brazilian Metropolitan Regions. The set of evidence obtained in this research indicates that these problems are not dissociated; specifically, using a unique household survey that have collected both information about commuting time and victimization (the supplement of PNAD (2009)), we obtain robust evidence that a long commuting time for individuals living in the Brazilian Metropolitan Regions increases the probability of these individuals being victims of robbery. This main result is consistent with both routine activities theory (Cohen & Felson, 1979; Cohen & Cantor, 1981) and the economic incentives approach to crime (Becker, 1968), once a longer commuting time increases the exposure of individuals to less security locations and implies higher expected gains for the potential criminal.

Our results were obtained by using propensity score matching techniques to create counterfactuals for the long commuters, and indicate that individuals with more than one hour of commute time have an overall 2.5-perceptual point increase in the probability of being a victim of robbery, with no robust effect on theft. This result means that approximately 27% of a long commuter's probability of being a victim of robbery can be attributed to their longer commute time. The estimative are robust to different robustness checks, including the estimative that excludes the poorest neighborhoods, located on the fringes of the Brazilian metropolitan regions, where a longer commute can co-occur with urban violence, and an estimative that excluded most of the commuters that travel during the late-night periods. The performed sensitivity analysis also strongly suggests that the presence of unobservable factors would not be a sufficient impetus for our results.

Our results can help with understanding the patterns of increasing urban violence in Brazilian centers during the period of significant improvement in social conditions in the years since 2000. Whereas the improved labor market conditions of the 2000s could have potentially reduced the number of criminals, the higher regular exposure associated with increasing commute times in the same period reduced the cost of committing a crime in these urban centers.

There are also clear policy implications from the results. In an urban environment characterized by the public thoroughfare congestion associated with the increasing use of cars (Pereira & Schwanen, 2013), our set of evidence offers support for the idea that a safer public transportation system must be principal part of the public policy to incentivize its use. Finally, our evidence indicates that without considering the effect of a long commute time on the probability of victimization, the urban-policy makers of Brazil's metropolitan regions are underestimating the welfare gains associated with a more effective and safer transport system.

3. SOCIAL HOUSING PROGRAMS IMPACTS ON COMMUTING TIME: EVIDENCE FROM BRAZIL

3.1 Introduction

Housing is one of the main factors for family well-being and for the development of a full life. In Brazil, its importance was recognized to the point of being included as a social right in the Brazilian Federal Constitution. Despite this recognition, the housing deficit in Brazil is still large. According to João Pinheiro Foundation's 2016 report, in 2013 the housing deficit was estimated at 5.846 million households, of which 85.7% (5.010 million) were in urban areas (Fundação João Pinheiro, 2016). Also, there was an increase in the number of households in housing deficit condition in 2014, totaling 6.068 million units. This growth has raised the share of urban housing deficit to about 87.6%, mainly due to a reduction of the rural housing deficit in the same period.

The unsatisfied match between importance and provision of housing is explained by the specific conditions in which it's acquired. As stated by Duarte *et al.* (2017), it's purchase involves resources incompatible with families' annual income and depends crucially on the economic environment and stability, often absent features in emerging countries. Despite the great challenge, governments from both emerging and developed countries regularly implement housing policies to foster adequate housing provision. In Brazil, over decades, several programs have emerged with the explicit goal of reducing the housing deficit. However, such programs were of limited scope, or they were prone to economic cycles instabilities given the labor income sources of resources for the program.

From 2009 onwards, the main housing program in Brazil was the Programa Minha Casa Minha vida (PMCMV). Unprecedented on its scale, with about 3.5 million units contracted and 1.7 million delivered until 2014 (Ministry of Cities, 2016), the program was particularly focused on low-income families, offering various levels of below market interest rates and subsidies. The PMCMV was also designed as an anticyclical measure in response to the 2008 international crisis (Amore, 2015). This dual purpose has generated discrepancies between the program and previous housing plan guidelines, such as the National Housing Plan that created the National Social Interest Housing Fund. As discussed by Krause *et al.* (2013), in its initial structure the PMCMV did not have fundamental concerns with planning participation and social control, and does not reinforce the society performance review of the program, focusing mostly in the delivery of standardized products.

Although the program undoubtedly improve household well-being through an improvement in residence features as shown recently by Duarte *et al.* (2017), this lack of planning could be leading to an increase in social exclusion. For instance, Pequeno & Rosa (2015), in a study funded by the Ministry of Cities, shown that, for the Fortaleza Metropolitan Area, most of the housing units for lower income brackets were located predominantly in peripheral regions or in municipalities adjacent to the capital. Most of the beneficiaries reported, through questionnaires applied in these housing complexes, a deterioration in relation to their former residences in access to public transport services, access to work, health and school services, commerce and leisure. For those located the metropolitan area fringes, up to 86.7% declared worsening access to public transportation; 79.6% in access to health posts; 77% worsening access to trade; 63.7% in access to work; 69.9% in access to school and 55.8% in access to leisure.

In fact, one of the main dimensions that increase exclusion perception is a worsening in urban mobility with reflexes in commuting time. As discussed in Lucas (2012), transportation poverty is an intersection between transport disadvantages such as, absence of own vehicle, high transport costs and poor transportation services, with social disadvantages as, low income, absence of employment opportunities and poor housing. This combination leads to the inaccessibility of goods and services, social capital, social networks and the possibility of decisions, leading finally to social exclusion. In this context, modern housing policies must associate the source of housing with urban policies to prevent this exclusion process from occurring, or in Brazilian case, to continue.

In a country like Brazil, that already has a high commuting time with upward trends (Silveira Neto *et al.* , 2015), the improvement in housing conditions from housing programs could be overshadow by the increase in exclusion. In fact, commuting time is already one of the main urban problems of the country. According to the more up-to-date information of PNAD (IBGE, 2012), the average commuting time for the inhabitant of Brazilian metropolitan regions was around 40.0 minutes in 2012, a very high number if compared to metropolitan regions around the world (Pereira & Schwanen, 2013; Silveira Neto *et al.* , 2015; Pero & Stefanelli, 2015). A combination of low quality public transport together with public indirect subsidies for using individual transport make short distance locomotion a very high time demand action (IPEA, 2013).

A specific problem arising from exclusion through the worsening of accessibility is the possibility of increasing urban spatial mismatch. In fact, the social geographies of developing cities tend to be present spatial mismatches specially for the poor (Cervero, 2013). According to Gobillon & Selod (2014), there are at least five theoretical mechanisms that can make the distance to job opportunities harmful: increase in commuting costs for the individual, leading to less job offers to be economically viable to be accepted; decrease in job search efficiency because individuals may have very little information on which places have suitable job offers;job search costs can be large and may deter individuals from looking for a job in places that are distant from their residence; individuals residing in areas that are far from job centers and where housing is more affordable may have less incentive to actively look for a job, reducing job search effort; and finally, employers may consider that long commutes deteriorate the productivity because distant workers are more likely to be late or tired and longer commuting time convey a series of problems such as worsening in reported health Oliveira *et al.* (2015).

Empirical analysis seems to confirm these mechanisms around the world, Gobillon & Selod (2014) provides a review of empirical studies focused on the spatial mismatch mechanisms. In the Brazilian case, for the São Paulo metropolitan regions, for example, Haddad & Barufi (2017) using an origin-destination survey combined with a instrumental variable approach show that proximity to the center or accessibility to jobs will be translated more easily in real opportunities in the labor market, increasing individual wages. Also for the São Paulo Metropolitan Region, Moreno-Monroy & Ramos (2015) combine a difference-in-difference method with an instrumental variable strategy, using a historical network plan for the SPMR as an instrument to identify the impact of public transport expansions on local informality rates. The authors found that the average informality rate decreased 16 percent faster than areas that should have received infrastructure but did not because of delays. These results highlight the importance of accessibility in individual economic outcomes.

Therefore, the objective of this paper is to assess whether home financing, specially through the recent housing program in Brazil, which focused initially only on the mass production of housing proposed by the civil construction sectors, leads to an increase in commuting time, so that concerns about social exclusion as a byproduct of such policies are justified. The contributions of the paper are twofold. First, given the prominence of the current housing program in the Brazilian governments' agenda, combined with an absence of more rigorous assessments in impact of housing policies on commuting time, the paper fills an important gap in the national urban economics literature. Second, given the inexistence of research about the links between housing programs and the potential worsening in accessibility to jobs and services, we contribute to provide evidence about the assertion that social housing programs could have impacts on spatial mismatch.

To achieve this goal, we used information from the National Household Sample Survey (PNAD) for 2014. Focusing on household heads commuting time and using the PMCMV income brackets, we seek a robust evaluation of the program. Our results indicate that, for Brazil, the effect of the house financing programs on commuting time is positive and significant, although it is not a widespread phenomenon, being concentrated at the lower-income levels of non-metropolitan urban areas. These results were robust to different specification, estimation strategies and sensitivity analysis.

The remainder of the paper is organized as follows. The next section presents a description the Brazilian housing programs. Section 3 presents previous empirical evidence on the subject. Section 4 describes the empirical strategy and the treatment of the underlying data and Section 5 presents and analyzes the results while performing robustness checks and sensitivity analysis. The conclusions are presented in the last section.

3.2 Brazilian Social Housing Programs

One of the major recent demographic changes experienced in Brazil was a rapid urbanization process, the share of urban population increased from 45% to 84% just in the last four decades of the twentieth century, approaching developed country levels. This process combined with a high population growth rate exerted intense pressure on public policies in different dimensions of social life, particularly in the absence of adequate housing provision, leading to an increase in the housing deficit in the period.

The first national policy with the explicit goals of reducing the national housing deficit and to provide housing and residences for low-income families in Brazil came only in 1946, with the *Fundação Casa Popular* (FCP), a federal institution designed to guarantee urban residences for a quickly growing urban population. In the same year, with the realization that it was not possible to address the housing problem without attacking the barriers represented by the lack of physical infrastructure and basic sanitation, the FCP's competences were broadened to include work in complementary areas such as: urban water supply, sewerage, electricity supply and social assistance. This expansion of competences somehow recognized problems of social exclusion, however, as argued by Azevedo & Andrade (1982), since its origin, this initial effort for housing provision for low-income families experienced serious limitations such as, lack of resources and bad administration, characterized by political uses and influence in the FCP decision making.

This had direct implications in its mediocre performance; during almost 20 years of duration, the actions of the FCP were responsible for only around 17,000 residences (Motta, 2010).

Following this failure, the Brazilian government created the *Sistema Financeiro da Habitação* (SFH) in 1964, during the military regime. Differently from FCP, the SFH represented an ambitious attempt of not only regulating the housing credit financing, but to organize, centralize and direct financial resources to construction of residences in Brazil. Although not exclusively focused on low-income families, the SFH presented a clear branch direct to them (SFH-FGTS), founded by resources from the *Fundo de Garantia por Tempo de Serviços* (FGTS), a mandatory found collected from employed workers and initially used for providing resources assistance in case of unemployment (Azevedo & Andrade, 1982). However, as stated by (Cardoso, 2003) the SFH was incapable of meet the lower-income population demands because of the intrinsic contradictions of the two main objectives of housing policy, first to leverage economic growth and to meet the housing demand of the low-income population. Business sectors related to the housing market, such as construction companies, forced a prioritizing of upper-income sector financing, which formed an effective demand, leading to a bias in SFH to beneficiate middle and high-income families (Amore, 2015).

The financing of lower income groups proved to be inadequate for the most impoverished populations and generated a bottleneck represented by the absence of subsidies combined with and the requirement of real correction of debts, given the excessive cost of housing in relation to income levels (Cardoso, 2003). These problems and difficulties associated with the housing financing low-income families motived the creation of alternative programs in the 1970's. The programs, such as *Profilurb*, *Pró-Morar*, and *João de Barro*, however, were mostly directed to help low-income families evolved in self-building and to re-urbanization of poor areas (mainly slums) and presented reduced impacts in terms of new residences to low-income families (Azevedo & Andrade, 1982). Despite these problems, Studies on the BNH indicate that during its existence approximately 4.5 million homes were financed, with 48,8% of the total destined to the average sectors, and 33,5% formally destined to the low-income groups (Cardoso, 2003; Amore, 2015).

Between 1986 and 1994, due to the political instability arising from the democratization and impeachment of the first democratically elected president after the military dictatorship, a vacuum was created in terms of housing policies. During this period there were several changes in the designation of the agency or ministry responsible for the management of housing policies. Also, the late 90's fiscal crisis leads to mild investments in the sector and to a pulverization of urban and housing policies in the state and local levels. The centralization of urban and housing policies returns with the creation of the Ministry of Cities, in 2003, responsible for building a construction of a system of cities and housing interest named National System of Social Interest Housing (SNHIS) and the National Housing Fund Housing of Social Interest (FNHIS). In the SNHIS/FNHIS system funds were articulated at different federative levels, all socially controlled with popular participation and with planned actions, obligatory to the federated entities that wanted to qualify in the system and receive federal funds. The Plans would be responsible for clearly defining the housing needs of each municipality, as well as presenting a strategy to address them.

Despite this careful constructed plan and even though the deficit was still a severe problem at the time of the launch of the PMCMV, its conception was an economic one, as a countercyclical response to the international crisis that began in 2008. Launched in 2009 the Minha Casa Minha Vida program (PMCMV) was unprecedented on its scale, with about 3.5 million units contracted and 1.7 million delivered until 2014 (Ministry of Cities, 2016). One improvement in relation to previous programs was an increase in the focus on low-income families, offering various levels of below market interest rates and direct subsidies, since the housing deficit in Brazil at the time was around 7,2 million residencies, 90% concentrated in low-income (Amore, 2015).

However, given the speed needed in a countercyclical policy, the government neglect the well thought out plans from the SNHIS/FNHIS and embraced the mass production of housing proposed by the civil construction sectors. This view change becomes explicit with concentration of decision making for the program in the Treasury (*Ministério da Fazenda*) and the Chief of Staff (*Casa Civil*). The main problem of this decision was the underutilization of all knowledge accumulated by the Ministry of Cities since 2003. Amore (2015) points out the concern the errors of the BNH were going to be repeated, with peripheral production in places poorly served by urban infrastructure, disarticulated of urban planning and regulations of land use and occupation. In this context, many were worried that the program would not improve the urban services provided for the poor, and in the worst scenario, could increase social exclusion. (Rolnik and Nakano, 2009).

The specific impact of the program design on the worsening of the beneficiaries' location was due to the change in the role of public agents in the PMCMV in relation to previous programs. In the PMCMV the public agents, and their knowledge of social demands, lost prominence in favor of the financial institution and construction companies. According to Rufino (2015), the financial logic prevailing in the program was based on the role of construction companies in proposing ventures with the financial institution based on established minimum construction standards. In this logic, the choices were based on cheaper land, increase in scale and standardization of housing units to increase financial viability. The combination of these conditions is possible only in places further away from urban centers, were cheaper land is available.

Due to this urban insertion pattern, the production of the program had a relevant role in the emergence of new peripheral frontiers, constituted from the implantation of new housing projects in discontinuous urban strips. This process was particularly evident in peripheral regions of the metropolitan areas, with lower levels of integration to the remainder of the metropolitan area such as, for example, *Itanhaém* in the *Baixada Santista* region, and in non-metropolitan areas spaces such as *São Carlos*, São Paulo, or *Parauapebas*, in the state of Pará[1].

Still according to Rufino (2015), most of the program's ventures did not include commercial and local services, leading to critical situations on the new expanded urban frontier than that of the consolidated peripheries in terms of monofunctionally[2] and segregation from the implantation of large closed housing projects with little articulation with the surroundings. However, the construction of projects of this nature also generally spawn informal service sector markets in its surroundings of even improvised in the own units, with a view to offering basic products to the same time that it has consolidated as an income-generating alternative and survival under a degree of absolute precariousness. As previously discussed, this worsening is access to job as services leads to one of the PMCMV contradictions. On the one hand, the program leads to advances in formalizing and increase in housing conditions, while fostering new forms of informality.

3.3 Social Housing Programs Impacts: Previous Evidence

For the past forty years, urban economics has used a powerful tool in explaining the urban structure. The Alonso-Muth-Mills standard model for analyzing household location choice base on compensation between the costs of commuting with the higher price of the dwelling near the Central Business Districts (CDB). One of the model's theoretical prediction is, an increase in commuting costs raises the price per square foot of housing, land rent, and structural density at central locations while lowering the values of these variables at more distant points (Brueckner, 1987; Duranton & Puga, 2015).

One of the main assumptions of the model is a uniform transport infrastructure within the radius of the CBD, making the commuting cost associated with distance and the opportunity cost of time. In the Brazilian case, a deterioration of the transport infrastructure is observed with the increase of the distance of the CBD, implying a high opportunity cost for commuting, especially great problem for the residents of the periph-

[1] See Rufino (2015) for maps, pictures and more examples on this urban insertion pattern.
[2] Areas with only one type of zoning, for example, only residential without services nearby.

eries. Pero & Stefanelli (2015) shows data for 2013, the poorest decile took approximately 20% more than the richest decile. In fact, Brazil is one of the countries with the longest average time to work reflecting the low level of collective transportation infrastructure in the main urban agglomerations (Pereira & Schwanen, 2013).

Empirical evidence on the direct effect of housing programs in commuting time is scarce. The available evaluations use different empirical strategies, depending on the specific program, and present different focus in terms of characteristics associated with housing service, depending on the set of available information. Most evaluations, however, can be divided in two groups, the ones focusing on impacts of housing programs in housing and socio-economic conditions and the ones focusing in the spatial mismatch implication of housing policies. In fact, most housing policies are justified as to tackle spatial work-residence mismatch and poverty concentration.

On the first group, for example, we have Waddington *et al.* (2009) meta evaluation of the impact of the availability of piped water and sewage in the low-income families' residences on health conditions in a set of developing countries and Ruprah (2011), which evaluated social housing programs in Latin American countries (Chile, Colombia, Equator, Costa Rica, Nicaragua, Panama, and Peru) findings positive impacts of the programs in a set of indicators such as ownership, quality of the residences, labor supply and income. For Mexico, Cattaneo *et al.* (2009), used information from a natural experiment to evaluate the impacts of the program Piso Forte, a program designed to substitute floor of land for floor of cement, on child health and on satisfaction of adults, finding positive effects on child health and satisfaction of adults. For Colombia, Rodriguez (2008) using a propensity score matching strategy, analyzed a program of social housing for families relocated because of urban violence in Colombia. The author finds positive impact on the quality of physical characteristics of the residences but non-effect on the education, health, and income indicators of beneficiary families.

As for the brazilian case, Duarte *et al.* (2017) shown that government housing policies like PMCMV had an impact on the welfare of recipient families translated in improvement in household characteristics. The authors found significant positive impacts of the recent housing programs in Brazil on the beneficiated families: these families tend to live in bigger residence and to have better accessing to important infrastructure residential services such as, piped water, sewage and garbage collection. Also, their results indicate that the locations of the new residences do not present worse accessibility to job location when considering the households' heads. However, the study only focusses on one income group and do not distinguish between metropolitan and urban regions.

On the second group, we can mention Matas *et al.* (2010) study for Madrid and Barcelona which shows that low job accessibility in public transport negatively affects employment probability. For developing countries, (Rosbape & Selod, 2006) show that in Johannesburg for instance, the average commute is around 80 minutes one way, and a national household survey shows that commuters in the poorest income bracket spend about 35% of their earnings on commuting. More recently, for Brazil, Haddad & Barufi (2017) used an origin-destination survey combined with an instrumental variable approach to evaluate the effect of location on wages. The author show that proximity to the center or accessibility to jobs will be translated more easily in real opportunities in the labor market, increasing individual wages. Also for the São Paulo Metropolitan Region, (Moreno-Monroy & Ramos, 2015) combine a difference-in-difference method with an instrumental variable strategy, using a historical network plan for the SPMR as an instrument to identify the impact of public transport expansions on local informality rates. The authors found that the average informality rate decreased 16 percent faster than areas that should have received infrastructure but did not because of delays.

However, as discussed earlier, the recent housing programs focused on mass production of housing without urban infrastructure, urban planning and regulations of land use occupation, in this context, these programs could have a dual effect. On one hand, welfare increases trough improvements in dwellings as shown by Duarte *et al.* (2017) and, on the other, welfare could be reduced by a longer commuting time, which implies larger opportunity costs and a higher chance of urban spatial mismatch.

3.4 Empirical Strategy and Data

Our initial hypothesis is that the Brazilian housing programs could be increasing the commuting time of its beneficiaries, since most of the program's ventures are on the edges of cities or in adjacent municipalities (Pequeno & Rosa, 2015; Rufino, 2015). If this hypothesis is confirmed, Brazilian social housing programs, besides improving the habitability conditions, can contribute to increase social exclusion of beneficiaries.

3.4.1 Propensity Score Matching

A traditional linear econometric specification for obtaining the effect of housing financing on the commuting time would be the following one:

$$Y = \alpha + \beta F + X\gamma + \varepsilon, \tag{3.1}$$

where Y is commuting time measure, F is an indicator for house financing, X is a set of control variables that affect commuting time chosen based on the commuting determinants literature, and ε is an error term. In this perspective, ideally, the estimative of β would correspond to the effect of house financing on commuting time. To circumvent the known problem with this kind of approach to non-experimental data and to obtain a more casual estimate of house financing on the commuting time we used the rich set of variables that influences individual available in PNAD to match individuals based on their propensity score (Angrist & Pischke, 2008), reducing potential influence of omitted variables. Our identification strategy is based on the conditionally independence or unconfoundedness (Rubin, 1974; Heckman & Robb, 1985) assumption and the propensity score theorem (Rosenbaum & e Rubin, 1985). If we denote by Y_i the observed result of individual i for our outcome variable, the probability of presenting long commuting, and Y_i^1 and Y_i^0 the potentials result of, respectively, taking the treatment (beneficiary of house financing) or not respectively, we have:

$$Y_i = F_i Y_i^1 + (1 - F_i) Y_i^0 \tag{3.2}$$

The uncounfoudedness assumption implies that, conditioned on a set of individual's variables X_i , the potential results are independent of being assigned to treatment, i.e., $Y_i^1, Y_i^0 \perp C_i | X_i$, where $\perp$ means independence. As shown, for example, in Angrist & Pischke (2008), this allows to obtain the effect of the treatment (in our case, if the household financed his house) as the difference in means of the outcome variable (in our case, the probability of presenting a larger commuting time) by status at each at each value of X_i. Also, using the Propensity Score Theorem of (Rosenbaum & e Rubin, 1985) to obtain the effect of treatment on treated individuals we use the fact that $Y_i^1, Y_i^0 \perp C_i | X_i$ implies $Y_i^1, Y_i^0 \perp C_i | p(X_i)$, where $p(X_i)$ correspond to the probability of being treated or the propensity score. In other words, conditioned on the propensity score, the potential results are independent of being assigned to the treatment. Using this theorem and the unconfoundedness assumption, it is possible to obtain the effect of a house financing on the commuting times as:

$$E(Y_i^1 - Y_i^0) = E\{E[Y_i | p(X_i), C_i = 1] - E[Y_i | p(X_i), C_i = 0]\}. \tag{3.3}$$

To obtain the sample correspondence of equation (3), we estimate $p(X_i)$ using a logit model and use two ways of matching the treated with the controls, the nearest neighbor (one treated with one control) based on the estimative of propensity score and the kernel estimation for weighting controls according to propensity scores (one treated with weighted controls). In our approach the treatment group will consist of those household with financed house, with the remainder households as control. As for the outcome, we observe that the PNAD dataset presents commuting time of the individuals in a categorized form; up to 30 minutes, more than 30 minutes and up to 1 hour, more than 1 hour and up to 2 hours, and more than 2 hours. From this information and because the average of commuting time of Brazilian metropolitan regions was less than an hour for 2012, we built the outcome dummy variable assuming value equal to 1 if the individual takes more than 30 minutes commuting to work and equal to 0 otherwise.

3.4.2 Sensibility Analysis

Even with good matching quality, propensity score matching results relies on the conditional independence assumption (CIA). As this identifying assumption is non-testable by its nature, one may still question its plausibility and argue that our results are being affected by an omitted variable strongly correlated with house financing. With the purpose of remove this suspicion, we apply the sensitivity analysis proposed by Becker & Caliendo (2007) for binary outcome based on the bounds approach of Rosenbaum (2002). This analysis aims at assessing the sensibility of our estimates when it presents some level of bias.

Suppose that the probability of participation in a program is given by $P_i = P(x_i, u_i) = P(D = 1|x_i, u_i) = F(\beta x_i + \gamma u_i)$, where x_i are the observable characteristics, u_i are the unobserved ones, F is a distribution function and γ is the effect of the unobserved characteristics in the participation decision (in our case the home financed decision). With no bias we have $\gamma = 0$, and the probability of participation depends solely on x_i, however if we have some confounding factors leading to bias, two individuals with the same x, will have different treatment probabilities. Suppose a matched pair of individuals i and j, and assume F as a logistic distribution. In this case, the odds that individuals receive treatment are then given by $P_i/(1 - P_i)$ and $P_j/(1 - P_j)$ and the odds ratio between then will be:

$$\frac{\frac{P_i}{(1-P_i)}}{\frac{P_j}{(1-P_j)}} = \frac{P_i(1-P_j)}{P_j(1-P_i)} = \frac{\exp(\beta x_i + \gamma u_i)}{\exp(\beta x_j + \gamma u_j)} = \exp\{\gamma(u_i - u_j)\}, \tag{3.4}$$

where the last equality comes from the matching procedure of the individuals with same observable characteristics. The sensibility analysis focus on how changes in in γ and $u_i - u_j$ affect the treatment effect found by the propensity score matching. Rosenbaum

(2002) finds upper and lower bounds of the odds ratio based on the values of γ as:

$$\frac{1}{e^{\gamma}} \leq \frac{P_i(1-P_j)}{P_j(1-P_i)} \leq e^{\gamma} \tag{3.5}$$

When both matched individuals have the same probability of participating we have $e^{\gamma} = 1$. otherwise, if for example $e^{\gamma} = n$, individuals who appear to be similar (in terms of x) could differ in their odds of receiving the treatment by a factor greater than one. In this sense, e^{γ} is a measure of the degree of departure from a study that is free of hidden bias (Becker & Caliendo, 2007).

However, the focus of Rosenbaum (2002) is on continuous outcome variables. For binary outcomes, Aakvik (2001) suggest the usage of the Mantel and Haenszel (MH) test statistics. According to Becker & Caliendo (2007), by observing the outcome y for both participants and nonparticipants we can make inferences on the effects of a treatment. If y is unaffected by different treatment assignments, the treatment has no effect. However, if y is different for different assignments, then the treatment has some positive (or negative) effect, and to be significant, the treatment effect must cross some test statistic $t(d, y)$. The authors use the MH test statistics, given by:

$$Q_{MH} = \frac{|Y_1 - \sum_{s=1}^{S} E(Y_{1s})| - 0.5}{\sqrt{\sum_{s=1}^{S} var(Y_{1s})}} \tag{3.6}$$

The MH nonparametric test compares the successful (in terms of outcome variable) number of individuals in the treatment group with the same expected number, given that the treatment effect is zero. Under the null hypothesis of no treatment effect, the distribution of y is hypergeometric, allowing to compute the expected value and variance of y.

$$Q_{MH} = \frac{|Y_1 - \sum_{s=1}^{S} \left(\frac{N_{1s}Y_s}{N_s}\right)| - 0.5}{\sqrt{\sum_{s=1}^{S} \frac{N_{1s}N_{0s}Y_s)(N_s - Y_s}{N_s^2(N_s-1)}}} \sim N(0,1), \tag{3.7}$$

where N_{1s} and N_{0s} as the numbers of treated and nontreated individuals in stratum s, and $N_s = N_{0s} + N_{1s}$. Y_{1s} is the number of successful participants, Y_{0s} is the number of successful nonparticipants and Y_s is the number of total successes in stratum s.

To use such a test statistic, we must first make the individuals in the treatment and control groups as similar as possible, because this test is based on random sampling. Since our matching procedure accomplishes this task, we can discuss the possible influences of $e^{\gamma} > 1$. For fixed $e^{\gamma} > 1$ and $u \in (0, 1)$, Rosenbaum (2002) shows that the test statistic QMH can be bounded by two known distributions. If $e^{\gamma} = 1$ the bounds are equal to the base scenario of no hidden bias. With increasing e^{γ}, the bounds move apart, reflecting uncertainty about the test statistics in the presence of unobserved selection bias, the value of the QM statistic is then corrected considering the position of the bias and the change in the expected values of success cases. Two scenarios are especially useful, one with a supposition of overestimation of the treatment effect, and another were we have underestimated the treatment effect (Becker & Caliendo, 2007). By evaluating the significance values of the test statistics for increasing values of γ, me may assess the sensibility of the treatment effect.

3.4.3 Data

To obtain the impact of the social housing programs on commuting time in Brazil, we use the official household survey of the National Household Sample Survey (PNAD) for 2014 provided by The Brazilian Institute of Geography and Statistics (IBGE). In addition to personal characteristics such as age, gender, race, education, family income, family structure, civil status, car ownership, working sector and commuting time, this nationwide survey provides information about numerous variables that characterize the households, including information about home ownership status, which include if the family lives in a financed house. As we match individuals based on their probability of being beneficiary of a house financing (the propensity score matching), for determining this probability we consider a large set of conditionings of house financing.

However, there are data limitations. The data does not inform the household purchase year, making it impossible to know when financing payments began, nor it informs if the housing financing is effectively associated with public program financing or with private financing. The MCMV program initiated in 2009 and it possible that there were families in 2014 that are still paying parcels associated with previous federal government programs. So, to circumvent those issues, we follow Duarte *et al.* (2017) strategies. First, they focus on families with total income up to R$ 1,600.00 (one thousand and six hundred reais, Brazilian currency) per month. This value corresponded to the upper limit of the lowest income bracket group eligible for the program (branch I of MCMV) in 2014. The argument is that for those families within this income bracket have high subsidies levels with government programs making market financing virtually inexistent for them. In addition, we introduced a second strategy, based on Ministry of Cities data on the number of residences financed and already delivered by the MCMV program, we take advantage of the unequal geographic distribution of the program to focus on household heads living in areas where the concentration of financed units is greater.

In this setup, our main treatment group consists ofeligible families in the lowest income branch, that is, families receiving no more than R$ 1,600.00, which are owners of a residence but are still paying the financing it received. The control group is constituted by eligible families that own a residence without paying any financing. Although our focus is on the impact of the poorer group given their social vulnerability, we also estimate the impacts of the housing policies in other income branches eligible for the PMCMV program, namely families with income between R$1,600.00 and R$ 5,000.00 reais. Our sample includes household heads of all urban areas in the country and of the ten official Brazilian Metropolitan areas. For each of these observations, we have an extensive set of variables that includes the potential determinants of house financing and commuting time, including individual characteristics such as gender, age, marital status, educational level, income level and working sector (Fujita, 1989; Lee & Mcdonald, 2003; Silveira Neto *et al.* , 2015). Variables associated to location of the household in the metropolitan region an urban area are also considered, which includes both family characteristics and variables associated to regular activities (Fujita, 1989; Silveira Neto *et al.* , 2015).

Before presenting the empirical results, we will analyze the differences between the treated and untreated groups, referred from now on as beneficiaries and non beneficiaries, respectively. Table (3.1) presents some descriptive statistics for beneficiaries and non-beneficiaries of public housing financing considering all incomes, where we have a series of age dummies defined in the following manner. *Age2* represents individuals with wage between 20 and 35 years; *Age3* Between 35 and 50 years and *Age4* for those with 50 years old or more. Also, we have a group of educational dummies defined as: *Edu1* for individuals with less than 4 years of education; *Edu2* for those between 4 and 7 years; *Edu3* for those between 8 and 10 years; *Edu4* for 11 to 14 years of education and finally *Edu5* for individuals with more than 14 years of education. *Ln Income*, is the $\log$ of household per capita income, Family size is the number of individuals in the household.

Table 3.1: Descriptive Statistics

	Unmatched			Matched		
	Treated	Control	t-test	Treated	Control	t-test
Man	0.44	0.36	3.59***	0.44	0.44	-0.30
White	0.38	0.30	3.98***	0.38	0.37	0.18
Married	0.06	0.07	-0.53	0.06	0.05	1.29
Age1 (<20)	0.02	0.03	-3.85***	0.02	0.02	-0.40
Age2 [20-35)	0.50	0.57	-3.35*	0.50	0.54	-1.30
Age3 [35-50)	0.41	0.39	1.08	0.41	0.47	-2.20
Age4 [50-)	0.44	0.39	2.30**	0.44	0.38	1.98
Edu1 [0-4] yrs	0.14	0.20	-3.32***	0.14	0.14	0.00
Edu2 [4-8) yrs	0.24	0.27	-1.99**	0.24	0.24	0.00
Edu3 [8-11) yrs	0.18	0.20	-1.35	0.179	0.17	0.23
Edu4 [11-14] yrs	0.42	0.31	5.80***	0.423	0.41	0.30
Edu5 (14+] yrs	0.06	0.04	3.33	0.06	0.07	-0.70
Ln Income	6.99	6.89	5.77***	6.99	7.01	-0.84
Family Size	2.70	2.68	0.41	2.70	2.69	0.07
Informal	0.22	0.28	-3.12***	0.22	0.24	-0.56
Self Employed	0.15	0.20	-3.33***	0.15	0.13	0.86
Enterpreneur	0.01	0.01	0.43	0.01	0.01	-0.30
Industry	0.14	0.11	3.02***	0.14	0.15	-0.17
Construction	0.11	0.17	-3.89***	0.11	0.12	-0.65
Commerce	0.18	0.19	-0.41	0.18	0.16	0.94
Metropolitan	0.36	0.38	-0.89	0.36	0.36	0.06
Urban	0.64	0.62	0.89	0.64	0.64	-0.06

Note: Based on Matching Logit and Nearest Neighbor Matching criterion; All incomes sample
Significance levels: *** 1% , ** 5% and * 10% .
Source: author's own calculation based on the PNAD dataset.

The information provided in Table (3.1) indicate some differences between the two groups related to individual characteristics when considering unmatched samples, confirming that only the income criteria (focus on branch I) is not enough for balancing the treated and control samples. Note, for example, that the household heads of beneficiary families present greater proportion of white and higher schooling individuals. However, after matching, those differences area eliminated.

Table 3.2: Commuting time distribution between Beneficiaries and Non-beneficiaries

	Full Sample			Metro Area			Urban Area		
Commuting time	Treated Mean	Control Mean	Diff.	Treated Mean	Control Mean	Diff.	Treated Mean	Control Mean	Diff.
Up to 30 minutes	0.664	0.652	0.012	0.529	0.496	-0.032***	0.768	0.766	-0.002
30 minutes to one hour	0.226	0.242	-0.016	0.309	0.328	0.018*	0.170	0.173	0.003
1 hour to 2 hours	0.090	0.086	0.004	0.133	0.144	0.011	0.046	0.045	-0.001
More than 2 hours	0.019	0.018	0.001	0.027	0.031	0.002	0.014	0.015	0.001

Significance levels: *** 1% , ** 5%, and * 10% .
Source: author's own calculation based on the PNAD dataset.

In Table (3.2) we present the distribution of both treated and control groups across commuting time indicators. As we can see, most household heads have at most 30 minutes commuting time (66.4% for the beneficiaries and 65.2% for non-beneficiaries respectively). Also, there are no statistical significant differences between the treated and control groups. These comparisons seem to suggest that critics of the programs are wrong in pointing to an increase in the commuting of beneficiaries. However, these results do not present directly "comparable" individuals, that is, they do not consider the matching in the personal characteristics and, particularly, on states and metropolitan regions of the country that present different commuting patterns. Thus, a more rigorous investigation is necessary.

3.5 Results

In this section we present our estimates of the impact of Brazilian low-income housing financing programs on commuting time of beneficiary families obtained by using the propensity score matching approach. Then, to assess the robustness of our results, we repeat propensity score estimates for a series of alternative samples. Finally, we perform Becker & Caliendo (2007) sensitivity analysis to assess the robustness of our results to the existence of unobserved confounding factors.

3.5.1 Propensity Score Matching Results

The logit model estimated coefficients, which use the set of variables in the Table (3.1), are presented in Table (3.3), where the dependent variable is a dummy indicating if a family is a beneficiary of the programs. The first two column presents estimates considering the full sample, the following two presents estimates considering families with income up to R$ 1,600 per month, and the last two columns present results for families with income between R$1,600 and R$ 5,000 per month. As discussed before, in this lower income group we have a potentially better comparison, because it is much improbable that an owner of this group had ever obtained a public house financing.

We can directly notice the importance of several variables in the probability of being a beneficiary of house financing. such as gender, with male household heads less likely to live in a financed house, also, older and better educated household are more likely to have a financed house.

Table 3.3: Conditionings of being beneficiary of public housing financing - logit model

	All households		Branch I[a]		Branch II[b]	
	Coef.	SE.	Coef.	SE.	Coef.	SE.
Men	-0.096*	0.051	-0.336***	0.108	-0.058	0.07
White	-0.024	0.052	-0.063	0.103	0.033	0.069
Married	-0.143	0.098	-0.101	0.192	-0.286**	0.14
Single	-0.089	0.060	-0.101	0.125	-0.127	0.082
Age2	0.560***	0.195	0.354	0.284	0.528*	0.273
Age3	0.538***	0.196	0.497*	0.288	0.490*	0.275
Age4	0.036	0.204	0.204	0.312	0.011	0.286
Edu2	0.239**	0.110	0.267	0.170	0.247	0.153
Edu3	0.318***	0.112	0.417**	0.178	0.287*	0.155
Edu4	0.530***	0.105	0.685***	0.169	0.470***	0.146
Edu5	0.655***	0.117	0.694***	0.245	0.580***	0.161
Income	0.503***	0.036	0.183	0.131	0.554***	0.105
Family Size	-0.033*	0.019	0.070*	0.037	-0.090***	0.027
Work Type	Yes		Yes		Yes	
Work Sector	Yes		Yes		Yes	
State FE	Yes		Yes		Yes	
Metro FE	Yes		Yes		Yes	
N	38,278		17,605		16,795	

[a] Branch I consists of households up to R$ 1600 monthly family income
[b] Branch II consists of households with monthly family income between R$ 1600 and R$ 5,000
Significance levels: *** 1% , ** 5% and * 10% .
Source: author's own calculation based on the PNAD dataset.

Table (3.4) contains statistics that summarize the quality of the Nearest Neighbor PSM implementation considering the sets of all, metropolitan areas and urban areas households. Column (1) of Table (3.4), displays the pseudo R^2 from the estimation of the conditional treatment probability (propensity score) on both raw and matched samples (i.e., before and after matching), we can see that the model has significantly less power to explain treatment status after matching. This is complemented by the Likelihood-Ratio (LR) test of the joint insignificance of all the regressor in column (2), which also suggests that the matched sample is well balanced in the observed variables. Additionally, substantial reduction in mean and median absolute bias (treated versus control differences in covariate means and medians before and after matching), as computed as in Rosenbaum & e Rubin (1985). Also in Table (3.4), the same pattern can be observed for the matching quality for the metropolitan and urban sets of households.

Table 3.4: Balanced Quality Before and After Matching

Sample		Ps R^2 (1)	LR χ^2 (2)	$p > \chi^2$ (3)	Mean Bias (4)	Med. Bias (5)	B[a] (6)	R[b] (7)
All	Unmatched	0.059	981.6	0.00	16.1	10.4	71.2*	0.90
	Matched	0.001	5.21	1.00	1.30	0.90	7.00	1.01
Metro Regions	Unmatched	0.058	354.9	0.00	15.5	10.6	68.8*	1.06
	Matched	0.003	5.85	0.99	2.00	1.70	11.9	1.03
Urban Regions	Unmatched	0.054	378.5	0.00	15.3	11.3	67.0*	0.96
	Matched	0.005	12.54	0.82	2.90	2.50	16.5	1.02

[a] absolute standardized difference of the means of the linear index of the propensity score in the treated and (matched) non-treated group)
[b] the ratio of treated to (matched) non-treated variances of the propensity score index).
Note: Rubin (2002) recommends that B be less than 25 and that R be between 0.5 and 2 for the samples to be considered sufficiently balanced.
Significance levels: *** 1% , ** 5% and * 10% .
Source: author's own calculation based on the PNAD dataset.

Given the above good balance of the variables between the two groups, we present the results of the impact of government housing financing programs in Brazil on the commuting time of beneficiary families in the following Table (3.5). In addition to considering the two income groups, the results are also presented separately for metropolitan regions and non-metropolitan urban regions. This is justified by the intrinsic differences between these regions in terms of available land for large housing project, and commuting time.

First, notice in Panel A of Table (3.5) that, in the full sample, the results point to positive and statically significant impact of the public housing financing program in Brazil on commuting time, especially in the second income bracket, whith an increase in the probability of having larger than 30 minutes commuting time by approximately 5 percentage points. Given that almost 35% off individuals spend more than 30 minutes commuting, this result is particularly important. This corroborate the concerns raised by Amore (2015), that these programs are moving away its beneficiaries, with could lead to an increase in social exclusion if urban infrastructure is lacking in the destination areas.

However, also from Panel A we see that the positive effect is concentrated in household from the second income brackets. A possible explanation for these results are that the poor already have higher commuting than those with higher incomes (Pereira & Schwanen, 2013), since they live predominantly in peripheries or places with low trans-

Table 3.5: Impact of house financing on commute time in Brazil

	All Households	Branch I[a]	Branch II[b]
	Panel A: Full Sample		
House financing	0.038*	0.027	0.051**
	(0.015)	(0.038)	(0.023)
Controls	Yes	Yes	Yes
N - obs	38,278	17,605	16,795
	Panel B: Metropolitan Regions		
House financing	0.024	-0.016	0.017
	(0.027)	(0.061)	(0.04)
Controls	Yes	Yes	Yes
N - obs	15,552	6,113	7,260
	Panel C: Urban Non-Metropolitan Regions		
House financing	0.033*	0.101**	0.007
	(0.019)	(0.046)	(0.025)
Controls	Yes	Yes	Yes
N - obs	19,621	8,771	8,761

[a] Branch I consists of households up to R$ 1600 monthly family income
[b] Branch II consists of households with monthly family income between R$ 1600 and R$ 5,000
Significance levels: *** 1% , ** 5% and * 10%, robust standard erros in parenthesis.
Source: author's own calculation based on the PNAD dataset.

port density (Lucas, 2012). In this case, the housing program would not be further worsening the access of this group. Another possible explanation would be that by accepting this new location, the poorer group can enter the context of informal work close to home, as it seems to be the case. In Figure (3.1) we see that, for our sample, lower income deciles have smaller average commuting times. Unfortunately, to test this hypothesis it would be necessary to know the impact of the housing program in the choice of the workplace, information not available in the database used.

In Panel B we focus on metropolitan household. We expect no significant worsening in commuting time because of house financing given the already high commuting time and high land prices for housing development near the central areas. Rufino (2015) makes clear that the only exceptions occurred when public entities donated unused areas marked for housing projects under previous programs. As we can see from the panel, there is no significant effect of house financing in commuting time in any of the income brackets. As a possible explanation, we have the tendency in Brazilian Metropolitan Regions of the increase of commuting time even prior to the 2009 expansion of housing program. In fact,

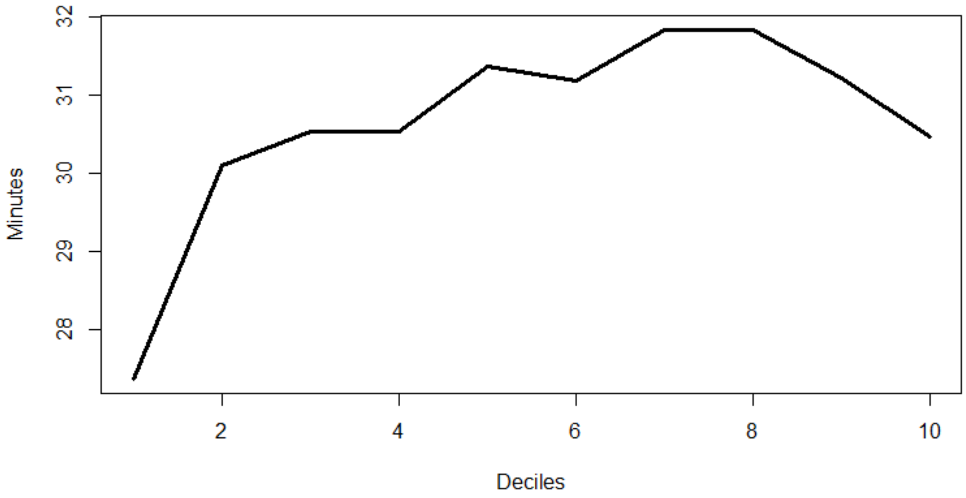

Figure 3.1: Average Commuting Time by Income Deciles (2014)
Source: PNAD microdata.

Delgado *et al.* (2016) using data for 2000 and 2010 population Census finds that most of commuting increase in metropolitan regions comes from non-migrants, that is, individuals who lived in the RM for more than 10 years, indicating that the suburbanization trend precedes the major housing program.

Finally, in Panel C, where we focus on families located in non-metropolitan urban areas, we find positive and significant effects for the low-income group. Households with home financing have a 10 percentage points higher probability of presenting a larger than 30 minutes commuting in comparison with no home financing households. Here we have an inverse situation of the metropolitan regions. Because cities outside metropolitan areas are usually smaller, most individuals do not exhibit excessively long commuting (in our sample 51% of individuals had commuting greater than 30 minutes in the metropolitan regions versus only 23% outside the metropolitan regions). Given the predominance of housing projects located in urban fringes or even disconnected from existing urban network not accompanied by connections and urban transportation planning, the program beneficiaries would be disadvantaged in terms of accessibility.

As stated by Carvalho & Rigotti (2015), medium-sized cities are becoming increasingly important in the urban network due to the historical processes of decentralization and productive restructuring. Non-metropolitan urban areas are the new destination of the poorest, due to the difficulty of insertion into the labor market in the large centers, concomitantly with the difficulties of staying in rural areas. In this case one of the viable options is the move to the middle-center peripheries (Matos & Baeninger, 2004).

Examples of this trend can be found in Sathler & Miranda (2011) for the case of São Paulo, Carvalho & Rigotti (2015) for mid-sized cities of Minas Gerais and in Macdonald & Winklerprins (2014) in the State of Pará.

So far, our results indicate that our hypothesis that the housing program could be increasing the commuting time of its beneficiaries can be confirm, but it's not a widespread phenomenon, being mostly concentrated in smaller income beneficiaries located in urban non-metropolitan region. To assess the validity of these initial results, in next section we perform robustness check and a sensibility analysis.

3.5.2 Robustness and Sensibility Analysis

Our method assumes specific weights for the observations in the control group and the obtained estimative considers a sample only delimited by family per capita income, ignoring the possibility of some families of control group have been beneficiated by past programs. Maybe more important, the set of above evidence was obtained assuming the conditional independence assumption (CIA), i.e., that conditional on observables characteristics, the potential results of receiving or not the housing financing does not depend on the fact that the family have received or do not the government financial help, a hypothesis not directly testable. In this section, we, first, provided some robustness checks for the method and the sample we used. Then, we evaluate the potential influence of non-observables factors on the results using Becker & Caliendo (2007) approach base on Rosenbaum (2002).

Initially we provide some robustness checks for the method and the sample we used. First, we consider as control group only renters, since those who rent probably never resorted to house financing before. Also, the usage of renters can overestimate the effects, since this group tends to live closer to the work places. In Panel A, B and C, we present results for this robustness check. In comparison with the main results in Table (3.5), we can see that results remain unchanged.

In the same line as the previous test, we use as controls only households with more informal occupancy conditions, such as ceded by the employer - defined as for the house assigned free of charge by the resident's employer, institution or non-resident person (relative or not), even if for a fee occupancy or conservation – and other unspecified occupancy conditions - defined by housing in abnormal conditions such as invasions, for example. It is possible to affirm more vehemently that families in these conditions of occupation have not been exposed to housing financing, improving the clear definition between treatment and control groups. Results using this subsample are presented in Table (3.7).

As we can see, the results for the full sample remain positive and significant for

Table 3.6: Impact of house financing on commute time in Brazil: Renters as Control

	All Households	Branch I[a]	Branch II[b]
	Panel A: Full Sample		
House financing	0.031*	0.045	0.058**
	(0.019)	(0.038)	(0.027)
Controls	Yes	Yes	Yes
N - obs	12,887	5,237	6,153
	Panel B: Only Metropolitan Regions		
House financing	0.013	-0.049	0.041
	(0.029)	(0.063)	(0.04)
Controls	Yes	Yes	Yes
N - obs	5,307	1,871	2,569
	Panel C: Only Urban Non-Metropolitan Regions		
House financing	0.070***	0.099**	0.004
Controls	Yes	Yes	Yes
	(0.024)	(0.044)	(0.031)
N - obs	7,359	3,117	3,513

[a] Branch I consists of households up to R$ 1600 monthly family income
[b] Branch II consists of households with monthly family income between R$ 1600 and R$ 5,000
Significance levels: *** 1% , ** 5% and * 10%, robust standard erros in parenthesis.
Source: author's own calculation based on the PNAD dataset.

higher income households, while results for metropolitan areas are still not significant for the lower income groups (Panel B). More importantly, the effect for urban household are even larger than those of Table (3.5). This is additional indication that lower income households have a worse off accessibility to work as an effect of the financed house location.

We now turn to the potential influence of non-observable factors. As we have discussed, the idea behind the Becker & Caliendo (2007) strategy is to observe how much the initial results change when it is considered existence of bias given by non-observance of the CIA hypothesis in some specific ways. In the following Table 8, we present results for the Mantel Haenszel generated bounds for the baseline result of Table (3.5) (first Column of Panel A).

As can be noted in Table (C3), in a study free of hidden bias, i.e., where $e^{\gamma} = 1$, the Q_{mh} test statistic is 2.36 and would constitute convincing evidence that house financing positively influences commuting time. Following Becker & Caliendo (2007) interpretation, if we have a positive (unobserved) selection, in that those most likely to financed

Table 3.7: Impact of house financing on commute time in Brazil: Robustness Checks II – Others Housing conditions

	All Households	Branch I[a]	Branch II[b]
	Panel A: All Formal Workers		
House financing	0.044*	0.014	0.099***
	(0.024)	(0.040)	(0.037)
N - obs	4,111	1,832	1,813
	Panel B: Metropolitan Regions		
House financing	0.057	0.038	0.112**
	(0.049)	(0.078)	(0.055)
N - obs	1,626	581	773
	Panel C: Urban Non-MetroRegions		
House financing	0.064**	0.119**	0.069*
	(0.031)	(0.053)	(0.041)
N - obs	2,054	908	925

[a] Branch I consists of households up to R$ 1600 monthly family income
[b] Branch II consists of households with monthly family income between R$ 1600 and R$ 5,000
Significance levels: *** 1% , ** 5% and * 10%, robust standard erros in parenthesis.
Source: author's own calculation based on the PNAD dataset.

also have a higher probability of presenting long commuting, then the estimated treatment effects overestimate the true treatment effect. So next, we assume the existence of positive unobserved heterogeneity and assess the sensibility of the treatment effect in the Q^{+}_{mh} column.

We can see that the test statistics is significant up to a $\Gamma = 1.05$, indicating that the effect in unsensitive for a bias that would increase the outcome odds in 5%, but sensitive to a 10% bias, leading to a critical Γ value of 1.10. This result does not imply that the CIA does not hold, instead, states only that the confidence interval for the effect would include zero if an unobserved variable caused the odds ratio of treatment assignment to differ between the treatment and comparison groups by 1.10. However, results are sensitive to possible deviations from the identifying unconfoundedness assumption, and hence some caution when interpreting the results in necessary. We repeat this sensibility analysis for the remainder of the statistically significant results in Table 5, namely the positive impacts for the higher income bracket beneficiaries, in the full sample, and for the non-metropolitan urban area poorer beneficiaries. The former effect (0.051), is robust to a

Table 3.8: Mantel Haenszel bounds for sensitivity of average treatment effects on the treated for the baseline results

$\Gamma = e^{\gamma}$	Q^{+}_{mh} [a]	p^{+}_{mh} [b]	Q^{-}_{mh} [c]	p^{-}_{mh} [d]
1.00	2.360	0.009	2.36	0.009
1.05	1.627	0.050	3.093	0.000
1.10	0.929	0.176	3.792	0.000
1.15	0.262	0.396	4.461	0.000
1.20	0.309	0.378	5.103	0.000
1.25	0.921	0.178	5.718	0.000
1.30	1.510	0.065	6.311	0.000
1.35	2.076	0.018	6.881	0.000
1.40	2.623	0.004	7.432	0.000

[a] Q^{+}_{mh} is the Mantel-Haenszel statistic under the assumption of overestimation of treatment effect
[b] p^{+}_{mh} is the significance level under the assumption of overestimation of treatment effect
[c] Q^{-}_{mh} is the Mantel-Haenszel statistic under the assumption of underestimation of treatment effect
[d] p^{-}_{mh} is the significance level under the assumption of underestimation of treatment effect
Source: author's own calculation based on the PNAD dataset.

10% bias . Also, the results found for the sub-sample of the all urban non-metropolitan beneficiaries (0.033) and the lower income bracket beneficiaries for the same subsample (0.101) are robust to a 5% and 20% bias, respectively .

Likewise, we can assume the existence of negative unobserved heterogeneity and assess the sensibility of the treatment effect in the Q^{-}_{mh} column. However, as stated by Becker & Caliendo (2007) for positive estimated treatment effect, as in our case, the bounds under this assumption are somewhat less interesting because the effect is significant under $\Gamma = 1$ and becomes even more significant for increasing values of Γ if we have underestimated the true treatment effect.

3.6 Concluding Remarks

Due to both its very quick urbanization process and deficient performance of internal housing credit market, different Brazilian governments have adopted social housing programs with the objective of providing minimum housing quality for low-income families. Although these programs undoubtedly improve household well-being through an improvement in residence features as shown by Duarte *et al.* (2017), this lack of planning could be leading to an increase in social exclusion working through increases in commuting time.

By using a propensity score matching strategy, the present research estimated of the impact of housing credit in commuting time, one of the main measures of social exclusion. Our results indicate that the Brazilian social housing credit program increase commuting time, specifically, having a financed house increases the odds of having larger than 30 minutes commuting time in 3.8 percentage points, this result was robust to some degree of bias. However, this effect it's not widespread, our results do indicate that house financing increases the odds of larger commuting time mostly for lower income people living not in metropolitan regions, but in urban non-metropolitan areas. Unfortunately, our dataset limitations do not permit the clear identification of public subsidized housing programs such as MCMV, which limits the scope of the conclusion. However, the findings of this research suggest that there is a non-negligible effect of housing financing on accessibility conditions, especially for low income beneficiaries and special attention for this group is needed in designing housing policies.

Bibliography

AAKVIK, ARILD. 2001. Bounding a Matching Estimator: The Case of a Norwegian Training Program. *Oxford Bulletin of Economics and Statistics*, **63**(1), 115–43.

ABADIE, A., & IMBENS., G. 2002. Simple and bias-corrected matching estimators. *Technical report, Department of Economics, University of California, Berkeley.*

ACEMOGLU, DARON, & ANGRIST, JOSHUA. 2001. How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws. *Pages 9–74 of: NBER Macroeconomics Annual 2000, Volume 15*. National Bureau of Economic Research, Inc.

AHREND, RUDIGER, FARCHY, EMILY, KAPLANIS, IOANNIS, & LEMBCKE, ALEXANDER C. 2014. What Makes Cities More Productive? Evidence on the Role of Urban Governance from Frive OECD Countries. *OECD Regional Development Working Papers, 2014/05, OECD Publishing, Paris.*

AMORE, C S. 2015. *"Minha Casa Minha Vida" para iniciantes*. Rio de Janeiro : Letra Capital. Chap. 1, pages 11–28.

ANGRIST, J., & PISCHKE, J. S. 2008. *Mostly Harmless Econometrics: An Empiricists Companion*. Princeton University Press.

ARAUJO JUNIOR, IGNACIO, & SILVEIRA NETO, RAUL. 2004. Concentração Geográfica de Capital Humano, Ganhos de Produtividade e Disparidades Regionais de Renda: Evidências para o Brasil Metropolitano. 297–314.

AZEVEDO, SÉRGIO, & ANDRADE, LUÍS AURELIANO GAMA. 1982. *Habitação e Poder: da Fundação Casa Popular ao Banco Nacional de Habitação*. Rio de Janeiro: Zahar Editores.

BADIORA, ADEWUMI ISRAEL, OJEWALE, OLUWOLE SAMUEL, & OKUNOLA, OLASUNKANMI HABEE. 2015. Perceived Risk and Fear of Crime in Public Transport

Nodes: The Experience from Nigerian Transit Environment. *International Journal of Criminal Justice Sciences*, **10**(2), 139–151.

BARBOSA FILHO, FERNANDO DE HOLANDA, & PESSOA, SAMUEL. 2008. Retorno da Educação no Brasil. *pesquisa e planejamento econômico*, **38**(1), 97–125.

BARROS, RICARDO PAES, FRANCO, SAMUEL, & MENDONÇA, ROSABE. 2007. A Recente Queda da Desigualdade de Renda e o Acelerado Progresso Educacional Brasileiro da Última Década. *Texo para Discussão - IPEA*, **1304**, 40.

BARUFI, ANA MARIA BONOMI, HADDAD, EDUARDO AMARAL, & NIJKAMP, PETER. 2016. Industrial scope of agglomeration economies in Brazil. *Annals of Regional Science*, **56**(3), 707–755.

BEATO, F, C., PEIXOTO, B. T., & ANDRADE, M. V. 2004. Crime, oportunidade e vitimizao. *Revista Brasileira de Cincias Sociais*, **19**, 73–89.

BECKER, G. S. 1968. Crime and punishment: An economic approach. *Journal of Political Economy*, **76**(2), 169–217.

BECKER, SASCHA O., & CALIENDO, MARCO. 2007. Sensitivity analysis for average treatment effects. *Stata Journal*, **7**(1), 71–83.

BENTO, ANTONIO M., CROPPER, MAUREEN L., MOBARAK, AHMED MUSHFIQ, & VINHA, KATJA. 2005. The Effects of Urban Spatial Structure on Travel Demand in the United States. *The Review of Economics and Statistics*, **87**(3), 466–478.

BROERSMA, LOURENS, EDZES, ARJEN J. E., & VAN DIJK, JOUKE. 2015. Human Capital Externalities: Effects for Low-Educated Workers and Low-Skilled Jobs. *Regional Studies*, **3404**(October), 1–13.

BRUECKNER, JAN K. 1987. Chapter 20 The structure of urban equilibria: A unified treatment of the muth-mills model. *Handbook of Regional and Urban Economics*, **2**, 821 – 845. Urban Economics.

CARDOSO, ADAUTO LÚCIO. 2003. Política Habitacional no Brasil. *Revista Proposta*, fev.

CARVALHO, RODRIGO COELHO DE, & RIGOTTI, JOSÉ IRINEU RANGEL. 2015. As migrações nas cidades médias de Minas Gerais e seus impactos no crescimento e na composição por sexo e idade da população no período 1980-2010. *Revista Brasileira de Estudos de População*, **32**(08), 235 – 256.

CATTANEO, MATIAS D., GALIANI, SEBASTIAN, GERTLER, PAUL J., MARTINEZ, SEBASTIAN, & TITIUNIK, ROCIO. 2009. Housing, Health, and Happiness. *American Economic Journal: Economic Policy*, **1**(1), 75–105.

CERVERO, ROBERT. 2013. Linking urban transport and land use in developing countries. *Journal of Transport and Land use (JTLU)*, **Vol. 6**, 7–24.

CICCONE, ANTONIO, & PERI, GIOVANNI. 2002. *Identifying Human Capital Externalities: Theory with an Application to US Cities*. IZA Discussion Papers 488. Institute for the Study of Labor (IZA).

CLARKE, RONALD V. 1996. *Preventing Mass Transit Crime*. Willow Tree Press, Inc.

COHEN, L. E., & CANTOR, D. 1981. Residential Burglary in the United States: Life-Style and Demographic Factors Associated With the Probability of Victimization. *Journal of Research in Crime and Deliquency*, **18**(January), 113–127.

COHEN, L. E., & FELSON, M. 1979. Social Change and Crime Rate Trends: a Routine Activity Approach. *American Sociological Review*, **44**, 588–608.

COMBES, PIERRE PHILIPPE, DURANTON, GILLES, & GOBILLON, LAURENT. 2008. Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, **63**(2), 723–742.

COMBES, PIERRE-PHILIPPE, DURANTON, GILLES, GOBILLON, LAURENT, & ROUX, SÉBASTIEN. 2010. *Estimating agglomeration economies with history, geology, and worker effects*. Vol. I.

COMBES, PIERRE PHILIPPE, DURANTON, GILLES, & GOBILLON, LAURENT. 2011. The identification of agglomeration economies. *Journal of Economic Geography*, **11**(2), 253–266.

CORBACHO, ANA, PHILIPP, JULIA, & RUIZ-VEGA, MAURICIO. 2015. Crime and Erosion of Trust: Evidence for Latin America. *World Development*, **70**, 400 – 415.

CRANE, RANDALL. 2007. Is There a Quiet Revolution in Women's Travel? Revisiting the Gender Gap in Commuting. *Journal of the American Planning Association*, **73**(3), 298–316.

CRUMP, R. K., HOTZ, V. J., IMBENS, G. W., & MITNIK, O. A. 2009. Dealing with limited overlap in estimation of average treatment effect. *Biometrika*, **96**(1), 187–199.

DATASUS. 2013. *Departamento de Informática do SUS*. Tech. rept. Ministério da Saúde.

DELGADO, P., DESCHAMPS, M., MOURA, R., & CINTRA, A. 2016. *Mobility in Brazilian Metropolitan Regions: Migration and Commuting Processes*. Brasília :IPEA. Chap. 10, pages 215–238.

DUARTE, GISLEIA BENINI, SILVEIRA-NETO, RAUL, & ALVES, JANIELLE. 2017. *Welfare Impacts of Social Housing Programs: Evidence from the Brazilian Experience*.

DURANTON, GILLES, & PUGA, DIEGO. 2015. Urban Land Use. vol. 5. Elsevier.

EDLUND, LENA, MACHADO, CECILIA, & SVIATSCHI, MARÍA MICAELA. 2015. Bright Minds, Big Rent: Gentrification and the Rising Returns to Skill. *NBER Working Paper No. w21729*.

F, WANG., & MINOR, W. W. 2002. Where the Jobs Are: Employment Access and Crime Patterns in Cleveland. *Annals of The Association of American Geographers*, **92**(3), 435–450.

FALCÃO, NATASHA, & SILVEIRA NETO, RAUL. 2007. *Concentração Espacil de Capital Humano e Externalidades: o caso das cidades brasilerias*. Anais do XXXV Encontro Nacional de Economia [Proceedings of the 35th Brazilian Economics Meeting]. ANPEC - Associação Nacional dos Centros de Pósgraduação em Economia [Brazilian Association of Graduate Programs in Economics].

FAN, YINGLING. 2017. Household structure and gender differences in travel time: spouse/partner presence, parenthood, and breadwinner status. *Transportation*, **44**(2), 271–291.

FELSON, M., & POULSEN, E. 2003. Simple Indicators of crime by time of Day. *International Journal of Forecasting*, **19**, 595–601.

FIGUEROA, MARIA J., NIELSEN, THOMAS A. SICK, & SIREN, ANU. 2014. Comparing urban form correlations of the travel patterns of older and younger adults. *Transport Policy*, **35**, 10 – 20.

FLORIDA, RICHARD, & MELLANDER, CHARLOTTA. 2014. The Geography of Inequality: Difference and Determinants of Wage and Income Inequality across US Metros. *Regional Studies*, **0**(February 2015), 1–14.

FUJITA, MASAHISA. 1989. *Urban Economic Theory*. Cambridge University Press.

FUNDAÇÃO JOÃO PINHEIRO. 2016. Déficit habitacional no Brasil 2013-2014. *Fundação João Pinheiro. Centro de Estatística e Informações*, 192.

GALEANO, EDILEUZA, & FEIJÓ, CARMEN. 2013. A estagnação da produtividade do trabalho na indústria brasileira nos anos 1996-2007 :. *Nova Economia - Belo Horizonte*, **23**(1), 9–50.

GLAESER, EDWARD L. 1999. Learning in Cities. *Journal of Urban Economics*, **46**(2), 254–277.

GLAESER, EDWARD L., KALLAL, HEDI D., SCHEINKMAN, JOSÉ A., & SHLEIFER, ANDREI. 1992. Growth in Cities. *Journal of Political Economy*, **100**(6), 1126–1152.

GOBILLON, LAURENT, & SELOD, HARRIS. 2014. Spatial mismatch, poverty, and vulnerable populations. *Handbook of Regional Science*, 93–107.

GORDON, P., & LEE, B. 2013. Spatial structure and travel trends in commuting and non-commuting travel in US Metropolitan Areas. *In:* HICKMAN, R., BONILLA, D., GIVONI, M., & BANISTER, D. (eds), *International Handbook on Transport and Development*. Cheltenham, UK: Edward Elgar.

GROOT, STEFAN P T, DE GROOT, HENRI L F, & SMIT, MARTIJN J. 2014. Regional wage differences in the netherlands: Micro evidence on agglomeration externalities. *Journal of Regional Science*, **54**(3), 503–523.

HADDAD, EDUARDO AMARAL, & BARUFI, ANA MARIA BONOMI. 2017. From rivers to roads: Spatial mismatch and inequality of opportunity in urban labor markets of a megacity. *Habitat International*, **68**, 3–14.

HECKMAN, J. J., & ROBB, R. 1985. Alternative Methods for Evaluating the Impact of Interventions. *Journal of Econometrics*, **30**, 239–267.

HEINEKE, J. M. 1987. *Economic Models of Criminal Behavior*. Amsterdam: North Holland Publishing Company.

HENDERSON, J. VERNON. 2010. Cities and development. *Journal of Regional Science*, **50**(1), 515–540.

HEUERMANN, DANIEL. 2011. Human Capital Externalities in Western Germany. *Spatial Economic Analysis*, **6**(2), 139–165.

IBGE. 2012. *Pesquisa Nacional por Amostra de Domiclios - 2012*. Tech. rept. Instituto Brasileiro de Geografia e Estatstica.

IBGE. 2013. *Regiões de Influência das CIDADES*.

ICHINO, A. MEALI, F., & NANNICINI, T. 2008. From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics*, **23**, 305–327.

IPEA. 2013. *Indicadores de Mobilidade Urbana: PNAD 2012*. Tech. rept. Instituo de Pesquisa Econmica Aplicada.

JACOBS, JANE. 1970. *The Economy of Cities*. Randon House Inc.

JOVANOVIC, BOYAN, & ROB, RAFAEL. 1989. The Growth and Difussion of Knowledge. *Review of Economic Studies*.

JUSTUS, MARCELO, & KASSOUF, ANA LÚCIA. 2013. Evidence of the effect of wealth observed by criminals on the risk of becoming a victim of property crimes. *EconomiA*, **14**(2), 88 – 101.

KONISHI, HIDEO. 2005. Concentration of competing retail stores. *Journal of Urban Economics*, **58**(3), 488 – 512.

KRAUSE, C, BALBIM, R, & LIMA NETO, V C. 2013. Minha Casa, Minha Vida, Nosso Crescimento: onde fica a política habitacional? *Texto para discussão 1853 Instituto de Pesquisa Econômica Aplicada.- Brasília : Rio de Janeiro*.

LEE, BUN SONG, & MCDONALD, JOHN F. 2003. Determinants of Commuting Time and Distance for Seoul Residents: The Impact of Family Status on the Commuting of Women. *Urban Studies*, **40**(7), 1283–1302.

LEMIEUX, A. M., & FELSON, M. 2012. Risk of Violent Crime Victimization During Major Daily Activities. *Violence and Victims*, **27**(5), 635–655.

LEMIEUX, ANDREW MICHAEL. 2011. *Risks of Violence in Major Daily Activities: United States, 2003-2005*. Ph.D. thesis, Graduate School-Newark Rutgers, the State University of New Jersey.

LIU, ZHIQIANG. 2007. The external returns to education: Evidence from Chinese cities. *Journal of Urban Economics*, **61**(3), 542–564.

LUCAS, KAREN. 2012. Transport and social exclusion: Where are we now? *Transport Policy*, **20**, 105–113.

LUCAS, ROBERT E. 1988. On the mechanics of economic development. *Journal of Monetary Economics*, **22**(1), 3–42.

MACDONALD, TRILBY, & WINKLERPRINS, ANTOINETTE M.G.A. 2014. Searching for a better life: Peri-urban migration in western Para State, Brazil. *Geographical Review*, **104**(3), 294–309.

MANDA, D, GERMANO, M, & MWANGI, M. 2002. Human Capital Externalities and Returns to Education in Kenya. *KIPPRA Discussion Paper No. 13*, 26.

MATAS, ANNA, RAYMOND, JOSEP-LLUIS, & ROIG, JOSEP-LLUIS. 2010. Job Accessibility and Female Employment Probability: The Cases of Barcelona and Madrid. *Urban Studies*, **47**(4), 769–787.

MATOS, R., & BAENINGER, R. 2004. Migração e urbanização no Brasil: processos de concentração e desconcentração espacial e o debate recente. *O Laboratório*.

MAYHEW, P., CHARKE, R.V. G. AND, STURMAN A., & HOUGHH, J. M. 1976. *Crime as opportunity.* Tech. rept. Home Officice Research Study 34.

MCCANN, P. 2013. *Urban and Regional Economics*. New York: Oxford University Press.

MESSNER, S. F., & BLAU, J. R. 1987. Routine Leisure Activities and Rates of Crime: A Macro-Level Analysis. *Social Forces*, **65**, 1032–1052.

MESSNER, S. F., LU, Z., ZHANG L., & LIU, J. 2007. Risk of Criminal Victimization in Contemporary Urban China: An Application of Lifestyle/Routine Activities Theory. *Justice Quarterly*, **24**(3), 496–522.

MIETHE, T. D., STAFFORD, M. C., & LONG, J. S. 1987. Social Differentiation in Crime Victimization: A Test of Routine Activities/Lifestyle Theories. *American Sociological Review*, **52**, 184–194.

MORENO-MONROY, ANA ISABEL, & RAMOS, FREDERICO. 2015 (Oct.). *The impact of public transport expansions on informality: the case of the São Paulo Metropolitan Region*. ERSA conference papers ersa15p1551. European Regional Science Association.

MORETTI, ENRICO. 2004a. Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, **121**(1-2), 175–212.

MORETTI, ENRICO. 2004b. Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions. *American Economic Review*, **94**(3), 656–690.

MOTTA, L. D. 2010. A Questão da Habitação no Brasil: Políticas Públicas, Conflitos Urbanos e o Direito à Cidade.

MULLIGAN, GORDON F., & FIK, TIMOTHY J. 1994. PRICE AND LOCATION CONJECTURES IN MEDIUM- AND LONG-RUN SPATIAL COMPETITION MODELS*. *Journal of Regional Science*, **34**(2), 179–198.

NARITOMI, JOANA, SOARES, RODRIGO R, & ASSUNÇÃO, JULIANO J. 2012. Institutional Development and Colonial Heritage within Brazil. *The Journal of Economic History*, **72**(2), 393–422.

OLIVEIRA, RODRIGO, MOURA, KLEBSON, VIANA, JORGE, TIGRE, ROBSON, & SAMPAIO, BRENO. 2015. Commute duration and health: Empirical evidence from Brazil. *Transportation Research Part A: Policy and Practice*, **80**(C), 62–75.

PEIXOTO, B. T., ANDRADE, M. V., & MORO, S. 2007. Violncia Urbana: Uma anlise Comparativa da vitimizao em So Paulo, Rio de Janeiro, Recife e Vitria. *Centro de Desenvolvimento e Planejamento Regional (CEDEPLAR). Universidade Federal de Minas Gerais. Discussion Paper: 323.*

PEQUENO, R, & ROSA, S V. 2015. *Inserção urbana e segregação espacial: análise do Programa Minha Casa Minha Vida em Fortaleza*. Rio de Janeiro : Letra Capital. Chap. 6, pages 131–164.

PEREIRA, R. H. N., & SCHWANEN, T. 2013. Tempo de Deslocamento Casa-Trabalho no Brasil (1991-2009): diferenas entre regies metropolitanas, nveis de renda e sexo. *IPEA, discussion paper 1813*.

PERO, VALÉRIA, & STEFANELLI, VICTOR. 2015. A Questão da Mobilidade Urbana nas Metrópoles Brasileiras. *Revista de Economia Contemporânea*, **19**(3), 366–402.

PERREAULT, SAMUEL. 2015. Criminal victimization in Canada, 2014. *Juristat*, **35**(1).

PNAD. 2009. *Pesquisa Nacional por amostra de domicílios*. Tech. rept. Instituto Brasileiro de Geografia e Estatística.

RAUCH, JAMES. 1993. Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities. *Journal of Urban Economics*, **34**(3), 380–400.

ROBACK, J.A. 1980. *The Value of Local Urban Amenities: Theory and Measurement.*

ROBACK, JENNIFER. 1982. Wages, Rents, and the Quality of Life. *Journal of Political Economy*, **90**(6), 1257–1278.

ROCHA, ROBERTA, SILVEIRA NETO, RAUL, & GOMES, SÓNIA MARIA. 2011. Maiores Cidades, Maiores Habilidades Produtivas: Ganhos de Aglomeração ou Atração de Habilidosos? Uma Análise para as Cidades Brasileiras. *Revista Economica do Nordeste*, 675–695.

RODRIGUEZ, O. 2008 (Aug.). *Evaluacion del Impacto de los programas de Vivienda Rural y Urbana para la poblacion en situacion de desplazamiento.* Tech. rept. Econometría S.A.

ROSBAPE, S., & SELOD, H. 2006. *Urban Unemployment in the Cape Metropolitan Area, in Post- Apartheid South Africa.* Petroria : Human Science Research Council. Chap. 7, pages 262–287.

ROSENBAUM, P. R., & E RUBIN, D. B. 1983. The Central Role of The Propensity Score in Observational Studies for Causal Effects. *Biometrika.*, **70**(1), 41–55.

ROSENBAUM, P. R., & E RUBIN, D. B. 1985. Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician*, **39**(1), 33–38.

ROSENBAUM, PAUL R. 2002. *Observational Studies.* New York: Springer.

RUBIN, D. B. 1974. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, **66**(5), 688–701.

RUBIN, D B. 2002. Using prpensity scores to help design ovservational studies: Application to the tobacco litigation. *Health services & outcomes research methodology*, **2**, 169–188.

RUFINO, MARIA BEATRIZ CRUZ. 2015. *Um olhar sobre a produção do PMCMV a partir de eixos analíticos.* Rio de Janeiro : Letra Capital. Chap. 3, pages 51–70.

RUPRAH, INDER. 2011 (May). *The Welfare Impacts of Social Housing Programs in Latin America: A Meta-impact Analysis.* OVE Working Papers 0511. Inter-American Development Bank, Office of Evaluation and Oversight (OVE).

SACHSIDA, ADOLFO, LOUREIRO, PAULO ROBERTO AMORIM, & MENDONÇA, MÁRIO JORGE CARDOSO. 2004. Um estudo sobre retorno em escolaridade no Brasil. *Revista Brasileira de Economia*, **58**(2), 249–265.

SATHLER, DOUGLAS, & MIRANDA, VITOR. 2011. A desconcentração demográfica paulista em perspectiva. *Cadernos Metrópole.*, **12**(24).

SILVA, VITOR HUGO MIRO COUTO, DE FRANÇA, JOÃO MÁRIO SANTOS, & NETO, VALDEMAR RODRIGUES DE PINHO. 2016. Capital Humano E Desigualdade Salarial No Brasil : Uma Análise De Decomposição Para O Período 2001-2012 . *Estudos Econômicos*, **46**(3), 579–608.

SILVEIRA NETO, R., DUARTE, G., & PAZ, A. 2015. Gender and Commuting Time in So Paulo Metropolitan Region. *Urban Studies*, **52**(2), 298–213.

SMITH, MARTHA J., & CLARKE, RONALD V. 2000. Crime and Public Transport. *Crime and Justice*, **27**, 169–233.

STOCK, JAMES H., & YOGO, MOTOHIRO. 2002 (November). *Testing for Weak Instruments in Linear IV Regression*. Working Paper 284. National Bureau of Economic Research.

TAVARES, PRISCILLA ALBUQUERQUE, & MENEZES FILHO, NAÉRCIO AQUINO. 2011. Human Capital and the Recent Decline of Earnings Inequality in Brazil. *Brazilian Review of Econometrics*, **31**(2), 231–257.

UNODC. 2012. *Global Study on Homicide, 2012*. Tech. rept. United Nation Office on Drugs and Crime.

VILALTA, C. J. 2011. Fear of crime in public transport: reaserach in Mexico City. *Crime Prevention and Community Safety*, **13**(3), 171–186.

WADDINGTON, H, SNILSTVEIT, B, & FEWTRELL, L. 2009. Water, sanitation and hygiene interventions to combat childhood diarrhea in developing countries. *International Initiative for Impact Evaluation*.

WAISELFISZ, JULIO JACOBO. 2014. *Mapa da Violência 2014: Os Jovens do Brasil*. Flasco Brasil.

Apêndice A

Table A1: Human Capital Concentration and Local Wages (Mincerian Equation for Educational Groups)

	Depent Variable.: Individual Hourly Wage			
	OLS		Panel	
Explanatory Variables	Coef.	Stand. Dev.	Coef.	Stand. Dev.
age (years)	0.044***	(0.000)	0.055***	(0.000)
age^2	-0.001***	(0.000)	-0.001***	(0.000)
tenure (months)	0.002***	(0.000)	0.001***	(0.000)
tenure^2	0.000***	(0.000)	0.000***	(0.000)
< Middle School	0.177***	(0.002)	-0.252***	(0.006)
Middle School	0.296***	(0.002)	-0.256***	(0.006)
High School	0.581***	(0.002)	-0.240***	(0.006)
College	1.397***	(0.002)	-0.187***	(0.006)
Firm Size (Employee Number)				
(0;4)	-0.192***	(0.012)	-0.125***	(0.012)
(5;9)	-0.123***	(0.012)	-0.090***	(0.012)
(10;19)	-0.041***	(0.012)	-0.044***	(0.012)
(20;49)	0.051***	(0.012)	-0.002***	(0.012)
(50;99)	0.136***	(0.012)	0.047***	(0.012)
(100;249)	0.229***	(0.012)	0.095***	(0.012)
(250;499)	0.275***	(0.012)	0.131***	(0.012)
(500;999)	0.298***	(0.012)	0.153***	(0.012)
1000+	0.326***	(0.012)	0.168***	(0.012)
Ignored	0.305***	(0.012)	0.178***	(0.012)
Intercept	0.250***	(0.072)	0.543***	(0.047)
Sector Dummies	Yes		Yes	
$\lambda_{j,t}$ *Dummies*	Yes		Yes	
$\lambda_{j,t} \times$ *College Dummies*	Yes		Yes	
R^2	0.452		0.455	
N	8,928,383		9,017,782	

Significance levels: *** 1% , ** 5% and * 10% .

Table A2: Correlation Table

	w indec	hc	densi.	ln area	share ind.	diver.	spec.	compet.	ski. shock	unski. shock
wage index	1									
humancap	0.06*	1								
emp. dens	0.13*	0.37*	1							
ln area	0.26*	-0.10*	-0.21*	1						
share ind	-0.32*	0.16*	0.28*	-0.19*	1					
diversity	-0.03	0.27*	0.13*	-0.11*	0.19*	1				
speciali.	0.13*	0.04	0.35*	0.03	0.13*	0.03	1			
competition	0.11*	0.15*	0.30*	-0.00	0.10*	0.01	0.11*	1		
skilled shock	-0.30*	0.13*	0.09*	-0.00	0.38*	-0.03	0.00	0.09*	1	
Unskilled shock	-0.25*	0.12*	0.09*	-0.07*	0.21*	-0.06*	-0.27*	0.08*	0.70*	1

Significance level: * 1%.

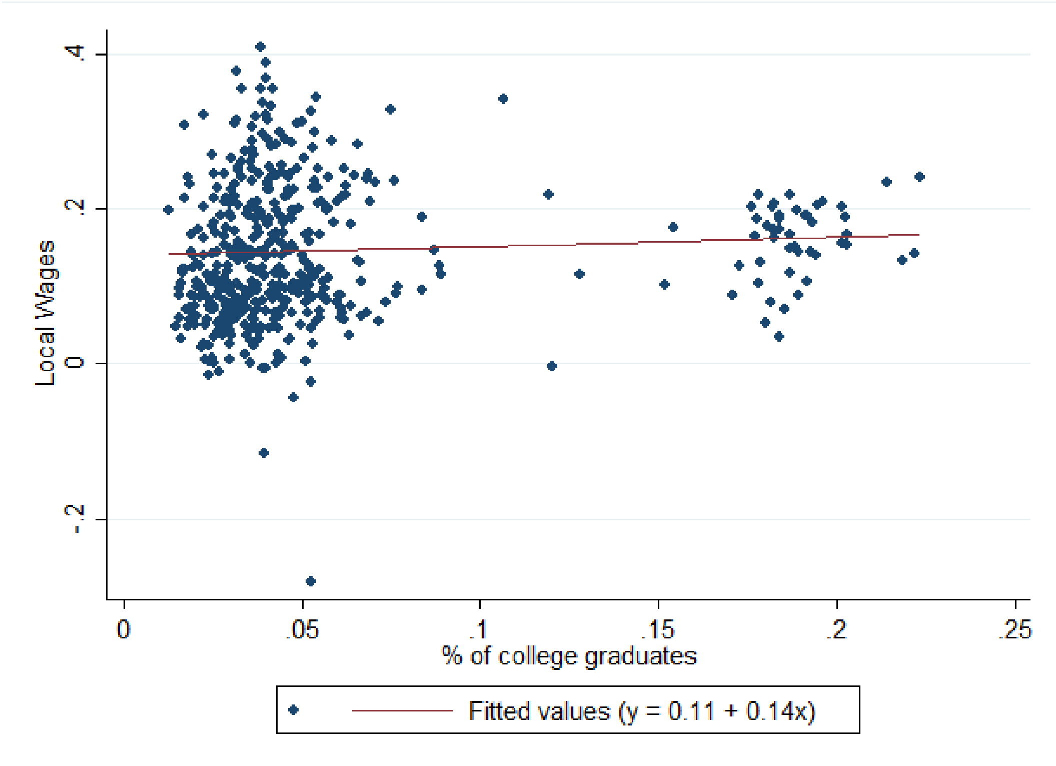

Figure 3.2: Local wages ($\hat{\lambda}$) relation with share of college graduates by Regic.

Apêndice B

Another way to evaluate the robustness of results is found using a different definition of treatment, in this case, considering that treated individuals have over 30 minutes of travel time. The results of the propensity score estimates can be seen in Table (B1) below. As can be seen, even with this more lenient definition, i. e, considering as treated a much larger number of individuals, we still observe significance of the postulated relationship for robbery.

Obviously, the impact is smaller since the average length of exposure to public spaces is reduced, in line with the exposure theory. In fact, this already expected reduction intuitively serves as an additional evidence of the relationship between probability of being robbed and the exposure time measured via commuting time.

Table B1: Propensity Score Matching Results for Robbery and Theft (30 minutes or more)

	Sample	Treated	Control	Diff	St Err.	Bootstrap St Err
Panel A:	**Robbery**					
N. N. Matching	Unmatched	0.083	0.072	0.011***	0.002	
	Matched	0.083	0.068	0.014***	0.003	0.003
Kernel Matching	Unmatched	0.083	0.072	0.011***	0.002	
	Matched	0.083	0.071	0.012***	0.002	0.002
Panel B:	**Theft**					
N. N. Matching	Unmatched	0.025	0.021	0.005**	0.001	
	Matched	0.026	0.020	0.006**	0.002	0.002
Kernel Matching	Unmatched	0.025	0.021	0.005***	0.001	
	Matched	0.025	0.020	0.005***	0.001	0.001

[1] bootstrap standard erros where calculated using 200 replications for nearest neighbor matching and 50 replications for kernel matching,

[2] Signifance Levels: * 10%, ** 5% and *** 1%.

Apêndice C

Table C1: Mantel Haenszel bounds for sensitivity of average treatment effects on the treated for the baseline results

$\Gamma = e^{\gamma}$	${Q_{mh}^{+}}$ [a]	${p_{mh}^{+}}$ [b]	${Q_{mh}^{-}}$ [c]	${p_{mh}^{-}}$ [d]
1.00	2.556	0.005	2.556	0.005
1.05	2.057	0.020	3.058	0.001
1.10	1.581	0.057	3.535	0.000
1.15	1.126	0.130	3.992	0.000
1.20	0.691	0.245	4.430	0.000
1.25	0.273	0.392	4.851	0.000
1.30	0.030	0.488	5.256	0.000
1.35	0.416	0.339	5.646	0.000
1.40	0.788	0.215	6.023	0.000

[a] Q_{mh}^{+} is the Mantel-Haenszel statistic under the assumption of overestimation of treatment effect

[b] p_{mh}^{+} is the significance level under the assumption of overestimation of treatment effect

[c] Q_{mh}^{-} is the Mantel-Haenszel statistic under the assumption of underestimation of treatment effect

[d] p_{mh}^{-} is the significance level under the assumption of underestimation of treatment effect

Source: author's own calculation based on the PNAD dataset.

Table C2: Mantel Haenszel bounds for sensitivity of average treatment effects on the treated for the baseline results

$\Gamma = e^{\gamma}$	Q^{+}_{mh} [a]	p^{+}_{mh} [b]	Q^{-}_{mh} [c]	p^{-}_{mh} [d]
1.00	2.001	0.023	2.001	0.023
1.05	1.557	0.060	2.446	0.007
1.10	1.134	0.128	2.871	0.002
1.15	0.729	0.233	3.277	0.001
1.20	0.343	0.366	3.667	0.000
1.25	-0.028	0.511	4.042	0.000
1.30	0.275	0.392	4.403	0.000
1.35	0.617	0.268	4.750	0.000
1.40	0.948	0.172	5.086	0.000

[a] Q^{+}_{mh} is the Mantel-Haenszel statistic under the assumption of overestimation of treatment effect

[b] p^{+}_{mh} is the significance level under the assumption of overestimation of treatment effect

[c] Q^{-}_{mh} is the Mantel-Haenszel statistic under the assumption of underestimation of treatment effect

[d] p^{-}_{mh} is the significance level under the assumption of underestimation of treatment effect

Source: author's own calculation based on the PNAD dataset.

Table C3: Mantel Haenszel bounds for sensitivity of average treatment effects on the treated for the baseline results

$\Gamma = e^{\gamma}$	Q^{+}_{mh} [a]	p^{+}_{mh} [b]	Q^{-}_{mh} [c]	p^{-}_{mh} [d]
1.00	2.712	0.003	2.712	0.003
1.05	2.461	0.007	2.969	0.001
1.10	2.220	0.013	3.212	0.001
1.15	1.990	0.023	3.445	0.000
1.20	1.771	0.038	3.669	0.000
1.25	1.561	0.059	3.884	0.000
1.30	1.359	0.087	4.092	0.000
1.35	1.165	0.122	4.292	0.000
1.40	0.978	0.164	4.486	0.000

[a] Q^{+}_{mh} is the Mantel-Haenszel statistic under the assumption of overestimation of treatment effect

[b] p^{+}_{mh} is the significance level under the assumption of overestimation of treatment effect

[c] Q^{-}_{mh} is the Mantel-Haenszel statistic under the assumption of underestimation of treatment effect

[d] p^{-}_{mh} is the significance level under the assumption of underestimation of treatment effect

Source: author's own calculation based on the PNAD dataset.