Borradores de ECONOMÍA

The Influence of Neighborhood Characteristics on Wages and Labor Supply in an Urban Context: The Case of a Latin-American City

Por: Leonardo Fabio Morales,
Lina Marcela Cardona

Núm. 844
2014

Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia

The Influence of Neighborhood Characteristics on Wages and Labor Supply in an Urban Context: The Case of a Latin-American City.*

Leonardo Fabio Morales†
Banco de la República
lmoralzu@banrep.gov

Lina Cardona-Sosa‡
Banco de la República
lcardoso@banrep.gov

Abstract

Using data from Medellín, the second-largest city in Colombia, in this paper, we assess how a set of neighborhood characteristics determines wages, and labor supply for workers in the city. We use GIS data to measure the quality of the environments in which workers live. The paper focuses on the impact of the following set of characteristics on labor supply and wages: availability of public transportation, crime levels, and density of economic activity. The empirical methodology consists of the estimation of linear equations for wages and worked hours, controlling the selection of individuals within the neighborhoods observed. In order to do this, in a first stage we estimate a probabilistic model of neighborhood selection from which selection correction terms are obtained; in a second stage, these correction terms are included in the linear equations for wage and worked hours. Additionally, we control the sample selection as well. We found that the endogeneity of the location decision tends to overestimate the magnitude of the effect of neighborhood characteristics on labor market outcomes. Nevertheless, the effect of some characteristics was still significant and important after we controlled the possibility of selection into neighborhoods.

Keywords: Labor Economics, Labor Supply, Urban Analysis, Housing Demand.

JEL Codes: J01, J22, O18, R21

*This work is preliminary and incomplete. The opinions expressed here are those of the authors and not of the Banco de la República de Colombia nor of its Board. All errors are authors´ responsibility. We thank Carlos Medina and other members of Banco de la República´s SGREE for their useful comments and suggestions. We also thank the participants to the 9th IZA/World Bank Conference on Employment and Development 2014 for their helpful comments and suggestions.

†Investigador Junior, Subgerencia Regional Estudios Económicos Banco de la República.

‡Investigadora, Subgerencia Regional de Estudios Económicos Banco de la República.

1 Introduction

One of the most distinctive elements that characterize Colombian and Latin American cities in general is the presence of considerable levels of spatial segregation. In this paper by "spatial segregation" we understand the existence of a clear division of the cities into big clusters of good and bad quality neighborhoods. The consequences of this type of city configuration have been studied by labor and urban economists. One of the main branches of the literature in this type of issues studies the "Spatial Mismatch Hypothesis." Broadly speaking, this hypothesis states that deficient labor outcomes are partly the result of excessive separation between individuals and their workplaces (Brueckner and Zenou, 2003).

Generally speaking, it may be possible that spatial segregation of individuals in a city causes deficient wages and labor supply. On the one hand, in segregated environments a portion of the population may be excluded from labor opportunities or networks in which information on job availability is exchanged. This type of isolation may cause an increase in the economic cost of participating in the labor market (Weinberg et al, 2004). On the other hand, there is a set of reasons discussed in the literature linking segregation and wages. These reasons range from deficient accumulation of human/social capital in bad communities (Altonji and Mansfield, 20ll), to possible discrimination against workers coming from bad neighborhoods (Rathelot, 2009; Dickerson, 2008).

Spatial segregation implies heterogeneity in neighborhood quality. Usually, isolated individuals live in low-quality neighborhoods. In this paper, we define "quality" in terms of the characteristics of the neighborhood. These characteristics are factors that may increase the cost of being employed or affect the accumulation of social and human capital, thus affecting wages. The aim of this paper is to estimate the impact of neighborhood quality on labor supply and wages, through the study of a representative sample of individuals in Medellín (the second largest city in Colombia, with a population of 3.5 million in its metropolitan area). Neighborhood quality is defined in terms of three main characteristics: (l) homicides in the neighborhood, (2) the density of economic units (business) in the neighborhood, (3) the distance to the nearest station of the city´s massive transportation system. The definition of neighborhood we use is that of "census tract polygons." These units are the building block of the census in Colombia, and they are relatively small areas for which census information is representative.

2 Literature Review

One of the first studies that explored the Spatial Mismatch Hypothesis was Kain, J. (l968). In this paper, the author proposed the existence of a relationship between segregation in the housing market and the poor results in labor outcomes of African Americans. The paper showed evidence of the negative effects of segregation on the unemployment rates for African-Americans in Detroit. In economics and other social sciences there is a research tradition regarding topics related with the Spatial Mismatch Hypothesis. The reader may refer to Holzer (1991) and Ihlanfeldt (1998) for a comprehensive review of the literature in this field.

Some empirical studies have sought to identify the relationship between a neighborhood's quality and its labor supply. A good example is Weinberg, Reagan, and Yankow, (2004). In their paper, these authors estimate a labor supply function specified in terms of some neighborhood characteristics. Weinberg, Reagan, and Yankow, (2004) are able to assess the hypothesis that the density of jobs in the neighborhoods is a factor that increases the individual labor supply. Some other papers look at the effect of community-neighborhood characteristics on wages (Altonji and Mansfiel, 2011; Cheng, 2012; Rathelot, 2009; Dickerson, 2008). The idea behind this set of papers is that wages can be explained directly or indirectly by the environments in which individuals live or have lived during their lives.

The relationship between residential environments and wages may take place through several channels. One possible channel is by altering the process of an individual's accumulation of social and human capital (Cheng, 2013). Another one is via employers' discrimination against workers living in particular areas that carry the burden of bad stereotypes (Rathelot, 2009). An example of the latter is when people may think that residents of some neighborhoods can be dangerous or cannot be trusted. In recent literature, this type of discrimination has been named redlining. The reader may find deeper explanations of redlining models in Zenou and Boccard (2000) or Zenou (2002). In addition to these explanations, the "Spatial Mismatch Hypothesis" offers a reasonable link between low-quality neighborhoods and low wages. Large distances between individuals and their jobs may affect their labor performance negatively, an argument that can be extended to other neighborhood characteristics with a negative connotation.

3 Theoretical Framework

The main idea of this study can be represented in a simple static labor supply framework, where individuals benefit from the quality of the neighborhood in which they live. A common practice in the literature on implicit prices (Rosen, 1974) is to represent an asset as a configuration of its characteristics. We represent a neighborhood as a vector $\mathbf{z} = \{z_1, z_2, ..., z_n\}$, where each z_i, $i = 1, 2, ..n$ represents a characteristic of a neighborhood. To simplify the notation, let us assume that all the variation in these characteristics can be summarized in an index $z \in [0,1]$, where $z = 0$ represents the lowest quality level and $z = 1$ represents the highest neighborhood quality.

The individuals in this framework benefit from leisure l, the quality of their neighborhood z, and a generic consumption good c. Neighborhood quality is included in the utility function because individuals obtain satisfaction from living in better neighborhoods, but also because neighborhood quality may alter marginal utility from leisure. Therefore, the representative individual´s utility function can be represented as:

$$u(c, l, z) \tag{1}$$

The budget constraint is standard, and it includes a labor - cost parameter for those individuals who work. Labor cost is a function of neighborhood quality. This represents the fact that most efficient transportation systems or the proximity to business clusters, among other characteristics, may reduce worker´s transportation expenditures. Other characteristics of the neighborhood may also alter the costs associated to the decision of working (living in a good neighborhood reduces the expenditures on an individual's security, for example). The budget constrain can be represented as:

$$1_{\{h>0\}}\left[w(x,z)\cdot h - a(x,z)\cdot h\right] + v = c + p_z \cdot z \tag{2}$$

Where $w(z, x)$ represents an individual's wage, which in this framework is a function of the individual characteristics of x (such as education and experience). Wage is a function of neighborhood quality z as well. This way of specifying wages is supported by all the literature suggesting there is an effect of segregation and residential environments on individual earnings (Zenou and Boccard (2000), Zenou (2002), Altonji and Mansfiel, 20ll; Cheng, 20l2; Rathelot, 2009; Dickerson, 2008). Additionally, $a(z, x)$ represent labor cost for individuals who work. They are also function of neighborhood quality z and individual characteristics x. A better neighborhood quality also implies better transportation systems, which reduces the individual's cost of going to work. Therefore, it is assumed that at $\frac{\partial a}{\partial z} < 0$ y $\frac{\partial a}{\partial z^2}$ > 0. p_z is the average price of an additional unit of neighborhood quality, and v represents non-labor income. All prices are relative to the price of the generic consumption good c. Individuals distribute their time, T, between work (h) and leisure (l), therefore $T = h + l$. The problem that individuals solve in this framework is maximizing (1) subject to the restriction represented by equation (2) and the time constraint. From this process, individuals obtain optimal consumption for leisure (l), consumption good (c) and neighborhood quality (z).

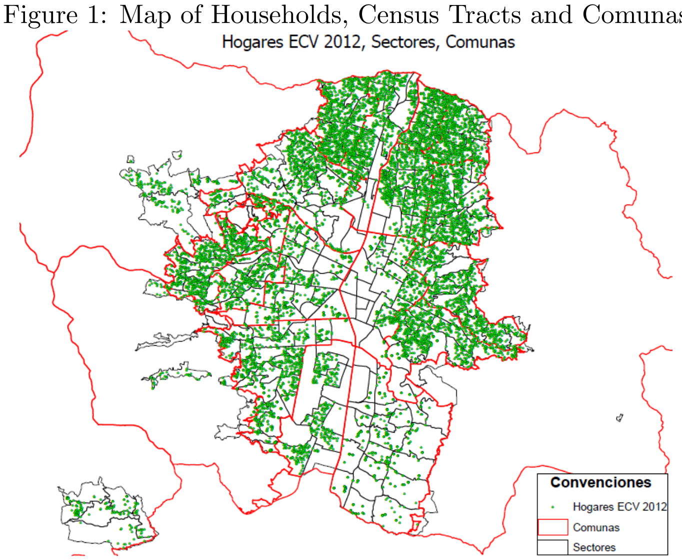
4 Data

The data used in this analysis proceeds from different sources of information: (1) Medellin's survey on household living conditions or HLCS (Encuesta de Calidad de Vida -ECVM-); (2) the city´s cartographic information updated by the Planning Department of the city, (3) homicide records from the National Police; and (4) the administrative information supplied by the local government on childcare institutions. The HLCS is an annual survey that interviews about 20.000 households in 20 "comunas" of the city (16 in the urban area and 5 in the rural area) and is representative to this level of disaggregation.

One of the advantages of this study is that all the households available in the HLCS are georeferenced, allowing us to identify the exact geographical area where most of the households are located, which is important to characterize the quality of the individual's neighborhood. By "neighborhood" we understand a small area within the city with a relatively small population. In that sense, we define a neighborhood as each one of the 243 census-tracts of the city. Each one of them accounts for 9090 inhabitants on average. The Census Tracts are small areas with enough demographic information to characterize them. Another geographical division which is important to define is that of "comunas," which are much larger than the Census Tracts and comprise several of them.

Figure (1) shows the cartographic information for the 2012's HLCS and the census tracks. "Comunas" are limited by the red line, whereas the census tracts (the smaller polygons) are defined by the black lines. Households are represented by the green dots on the map.

Figure 1: Map of Households, Census Tracts and Comunas



Several of the neighborhood characteristics are geographically defined. In order to generate those variables, we use geo-referenced information on metro stations, location of the economic units in the city (formal businesses dedicated to private or public economic activities), and location of murders.[1]. In addition, in order to construct some exclusion restrictions for the sample selection equation, we use geo-referenced information on public child-care providers in the city[2].

4.1 Policy Variables

The impact variables have been built using a methodology that considers the geographical location of the individual in relation to the neighborhood's characteristic of interest. Access to the transport system is measured through the distance in meters from the individual's residence to the nearest metro station, or to any station of the massive public transportation system in the city. As for the density of business and crime index, gravitational indexes are built using the inverse of the distance between the individual's residential location and the location of the characteristic of interest. The index that accounts for the density of business in the area is expressed in the following way:

$$A_i = \sum_{j=1}^{J} 1_{\{d(i,j)\ \leq\ D\}} \cdot \frac{1}{d\,(i,j)} \tag{A}$$

In the last expression, A_i is the density index of economic units for individual i. The expression $d(i,j)$ represents the distance between individual i and the jth amenity, assuming that there are J economic units (businesses) in the city. Parameter D represents the minimum distance at which an economic unit in the city receives a positive weight in the construction of the index for the ith individual. The indicator of function $1_{\{d(i,j)\ \leq\ D\}}$is equal to one for all business located in radius of D meters from individual's location. We estimated models with different values of D, and the specification presented in this paper uses the values of D that maximize the fit of the model. Note that A_i is a weighted summation of business in the surroundings of individual i. One intuitive interpretation of A_i is an expectation of the number of business in the neighborhood.

Regarding the variable for homicides, in addition to weighting by the inverse of the distance between the incident and the individual, we also weight by the inverse of the time between the occurrence of the homicide and the year 2012. Therefore, an expression for the density of murders for individual i, H_i, is computed in this paper using the following expression:

[1] We use data collected by the Intelligence Department of the National Police, in which all murders in the city are recorded with the address where they took place. Then, by using the address, we geo-referenced every murder.

[2] We use information provided the Program "Buen Comienzo", the largest public childcare program in the city with information on most of the public childcare providers.

$$H_i = \sum_{t=2002}^{T=2011} \left[\sum_{j=1}^{J} 1_{\{d(i,j_t)\ \leq\ D\}} \cdot \left(\frac{1}{d\left(i, j_t\right)} \right) \left(\frac{1}{T+1-t} \right) \right] \tag{H}$$

In this case the indicator of function $1_{\{d(i,j_t)\ \leq\ D\}}$is equal to one for all homicides that were reported in radius of D meters from individual's location at year t. Note that H_i is a weighted summation of homicidies in the surroundings of individual i. One intuitive interpretation of H_i is an expectation of the number of homicides in the neighborhood.

4.2 Spatial Distribution of Some Neighborhood Characteristics

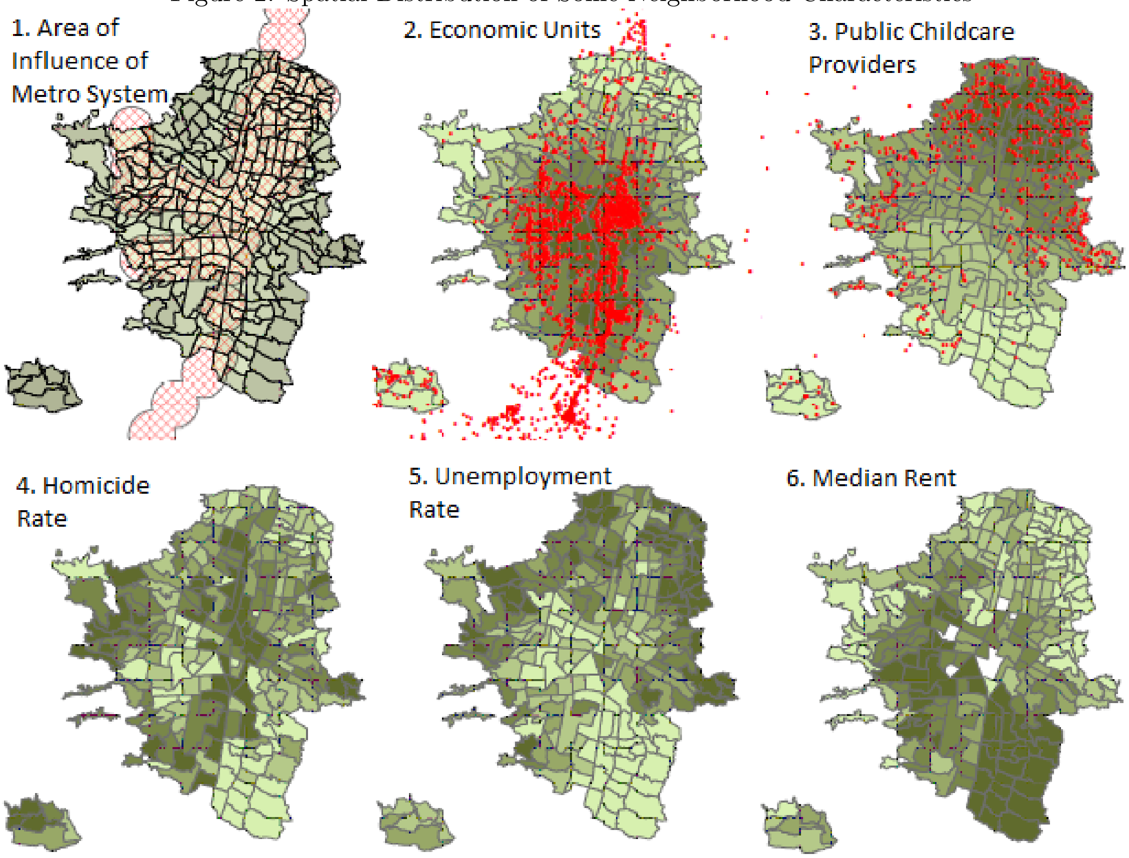
Figure (2) shows the spatial distribution of some relevant neighborhood characteristics. This is useful in order to have an idea of the composition of the city in terms of good and bad neighborhoods. The legend for each map is presented in Appendix C. The first map shows the area of influence of the city's massive transportation system. This is the set of "Metro-Cable" stations (cable air corridor), metro stations, and bus stops that feed the metro system. The area of influence of the system is defined to be around one (1) kilometer of radius from the center of the station. In this map, one can see how the massive transportation system covers most of the city. The reader can have an idea of the characteristics of this area of influence of the transportation system by comparing this map with other maps in figure 2: in fact, the areas where the average rent is higher (E.g. in the southeast on map 6) the transportation system is poor.

The second map (from left to right) shows the economic units of the city represented by red dots. The economic units are the best proxy variables we have to represent the spatial distribution of labor demand in the city. The map's background shows the distribution of a density index computed as indicated in the previous section. From the map, one can see that most of the economic units are concentrated in the center and south of the city. The third map represents the location of public childcare providers in the city. This variable is used as an exclusion restriction in our sample selection equation, a procedure that is explained in the following section. Map (4) represents the distribution of the homicide rate per census tract in the city. Map (5) represents the distribution of unemployment rate by census tract in the city. Finally, map (6) represents the distribution of the median rent by census tract in the city.

5 Methodology

There are three estimable equations that are derived from the economic optimization process sketched in section number 3: an equation for the optimal labor supply, a wage equation, and a residential

Figure 2: Spatial Distribution of Some Neighborhood Characteristics



location demand equation[3]. The main purpose of this paper is to estimate the unbiased effect of the neighborhood characteristics on the first two equations, which can be represented by the following expressions:

$$\ln(h_{is}) = \alpha_h + \pi \ln(w_i) + X_i \beta_h + Z_s \gamma_h + \varepsilon_i^h \tag{3}$$

$$\ln(w_i) = \alpha_w + X_i \beta_w + Z_s \gamma_w + \varepsilon_i^w \tag{4}$$

In which h_{is} represents the hours worked by individual i, who lives in the neighborhood s. Additionally, w_i represents the hourly wage of individual i. The matrix X_i contains the characteristics of an individual, while matrix Z_s contains the characteristics of the neighborhoods s. The interest of this study is the estimation of parameters in vectors γ, which describe the impact of neighborhood characteristics on labor supply and wage.

5.1 Self-Selection into Neighborhood Bias Correction

A possible source of bias in the estimation of equations (3) and (4) is that individuals choose the neighborhoods where they live. This can be seen as a self-selection process that can bias the coefficients in (3) and (4), specifically the ones in vectors γ. The bias would take place if this process of self-selection into the neighborhoods is driven by unobserved factors correlated with perturbation terms ε_i. It could be the case that an individual with a high interest in finding a job moves to areas with more economic units around, a motivation that will be affecting both the choice of a neighborhood and the labor supply. In order to control for this selection process, we estimate generalized selection models. This methodology allows us to specify a selection equation for any possible neighborhood in the city (census tract) using discrete choice selection models. There are several alternatives in the literature for the estimation of generalized selection models. The reader may find a survey of the alternatives available in Bourguignon et al (2007).

The idea of a generalized selection model is to specify a main (lineal) equation together with a multinomial selection equation. The models we estimate in this paper consist of two stages. The first stage is a discrete choice model of neighborhood choice (census tract). In the second stage, we estimate the labor supply and wage equations augmented with correction selection factors. The specifications for the labor supply are determined by functions of alternative specific probabilities estimated from the first stage. A more detailed description of the methodology is offered in the following paragraphs.

[3] This is an alternative way of presenting the neighborhood quality demand z because each neighborhood in the city has a particular configuration of characteristics that correspond to a unique value of z.

A reasonable hypothesis is that, at least partially, errors in equation (3) and (4) are correlated to unobservable factors driving the choice for residential location by workers in the city. Therefore, it is important to control for the possibility of selection into a neighborhood, and by doing so, correct the estimation bias. This practice has started to gain strength in the literature given that more and more researchers are considering the location of the individual as an endogenous factor. The reader may refer to Lall and Mengistae (2005) for a generalized selection correction model applied to firm-location decisions in India.

The part of our model that explains the process of neighborhood selection is in its own right a model of residential location demand, in which rational individuals choose the neighborhood that maximize their utility level. The level of utility associated to each alternative is a latent variable in a discrete choice model, in our case a conditional logit.

This study assumes that the individual i chooses a place to live from a set $S = \{s_1, s_2, ...s_k\}$, where each of the elements of the set S represents one of the neighborhoods in the city. In particular, each of the neighborhoods is defined as a census tract of the city. Assuming that each individual i derives a utility level y_{is}^* from choosing the neighborhood s, this level of utility is modeled as a linear function in the parameters as follows:

$$y_s^* = z_s\vartheta + \sum_l [x_{i,l} \times z_s{'}]\, \theta^I + u_{is}, \quad s = 1,..,K \tag{5}$$

where $x_{i,l}$ represents the l-th characteristic of the individual interacting with each of the elements in the z vector. The whole expression $\sum_l [x_{i,l} \times z_s{'}]$ contains the interactions among the characteristics of the s choice and the individual variables x_l of individual i. Vector θ^I includes the coefficients of these interactions. This is important because it is a way to increase the heterogeneity of the utility associated to each alternative. In this way the marginal utility of a particular neighborhood characteristic depends on the characteristics of the individual l. For instance, notice that the availability of public transportation or another amenity would provide different levels of utility to different households according to their demographic characteristics (i.e., income, household composition, etc.). Identification of systems of equations were one of them is a selection equation rely on the nonlinearity of this equation. It is a common practice as well using exclussion restrictions; these are variables included in the selection equation and excluded from the outcome equation. The exclussion restrictions that we use in this paper are a series of variables that explain the residential location decision, but presumably not, any labor outcome. The set of variables we use for this purpose are: the availability of desirable amenities as shopping malls, recreation and sport centers, and cultural facilities and libraries.

By assuming that u_{is} follows a Gumbel distribution, a model of "residential location demand" is derived as a Conditional Logit. This model follows a multinomial specification which is very convenient because there is only one parameter per alternative. This is a special characteristic of conditional models in which characteristics vary by alternative and not by individual. To simplify the notation, let us call ω_{is} any of the dependent variables of the models that will be estimated (logarithm for monthly wage, logarithm for hourly wage, logarithm for hours worked). For each individual i we are able to observe ω_{is}, only when the alternative s is chosen. The value of ω_{is} conditional on another alternative being chosen is a counterfactual. The neighborhood s is chosen only when:

$$y_s > \max_{s \neq s'} \{y_{s'}\} \tag{6}$$

Following MacFadden (1974) and under the assumption that the errors u_{is} are Gumbel independent and identically distributed, the probability associated to each alternative follows a logistic distribution that is closed and can be easily computed. Therefore, the probability that an individual i chooses alternative s can be written as:

$$P(s) = \frac{\exp\left(z_s'\theta + \sum_l \left[x_{i,l} \times z_s'\right] \theta^I\right)}{\sum_{j \neq s} \exp\left(z_j'\theta + \sum_l \left[x_{i,l} \times z_j'\right] \theta^I\right)} \tag{7}$$

The literature proposes different approaches in order to obtain unbiased estimators of equations (3) and (4). (For more details, the reader may refer to Bourguignon et. al., (2OO7)). The approach followed in this study is the one implemented by Dubin and McFadden (l984). This methodology considers the inclusion of the conditional expectations of the error term in eq (3) and eq (4) given the unobservable factors associated to each residential location alternative. Dubin and McFadded (l984) found that, under standard assumptions, the conditional expectation of the error term ε_i, is given by the following expression:

$$E\left[\varepsilon_{is} | u_{i1}, u_{i2}, ..., u_{iK}\right] = \sum_{s \neq j} \gamma_j \left[\frac{P_{ij} \ln(P_{ij})}{1 - P_{ij}} + \ln(P_{is})\right] \tag{8}$$

where P_{ij} is the probability of observing an individual i in the neighborhood j.

In Bourguignon et al (2007), the authors evaluate different alternatives proposed in the literature for estimating selection correction models when the selection equation is specified as a multinomial logit. In order to do this, the authors assess the precision and unbiasedness of the models through Monte Carlo experiments. The main result shows that in most of the cases the methodology proposed by Dubin and McFaddedn (l984) presents a better performance than other methodologies such as the one proposed by Lee (l983). Bourguignon et al (2007) concludes that the methodologies of the Dubin-McFadden type have a good performance, in general. In fact, the Monte Carlo experiments indicated that correction models of selection bias based on a multinomial logit provide a satisfactory correction bias.

The idea of a multinomial selection model is not restricted to the use of a multinomial logit. Other models share the same assumptions about the error distribution of the selection equation. For instance, when the choice is about geographical location, the conditional logit is very convenient, since it allows modeling the utility of each alternative in a tractable and realistic way. There are relatively few studies using selection correction models where the categories of selection are spatial locations. Up to the knowledge of these authors, there is one previous application of generalized selection models using a conditional logit in the selection equation: the one by Lall and Megistae (2005). In this paper, the authors model a firm's choice of location by using a conditional logit to estimate the selection equation.

The specification proposed in this study for each of the labor outcomes ω_{is} is as follows:

$$\omega_{is} = \alpha + X_i\beta + Z_s\gamma + \sum_{s\neq j}\gamma_j \left[\frac{\hat{P}_{ij}\ln\left(\hat{P}_{ij}\right)}{1-\hat{P}_{ij}} + \ln\left(\hat{P}_{is}\right)\right] + \eta_{is} \tag{9}$$

where probabilities $\widehat{P}_{ij}$ for an individual i are the predicted probabilities for each alternative of the conditional logit after the parameters of each equation (7) are estimated.

5.1.1 Sampling of the choice set

Although it is possible to estimate a conditional logit for all possible neighborhoods in the city (243 census tracts in total), a model with that many alternatives may be difficult to manage. In this paper, we follow a common result found in previous literature (McFadden, l975) which shows that under certain conditions, the maximum likelihood function of a model with all the alternatives is equivalent to the one of a model in which the set of alternatives is built through a random sampling process.

In the literature, there are several methodologies for random sampling of a choice set. One of the most used is dividing the entire set of alternatives into smaller sets or partitions, and after that selecting randomly one alternative from each partition. The random subset will be formed by a random category from each partition, jointly with the individual's observed choice in the sample. The literature offers different ways to partition the choice set. In this study, we use the "comunas" as partitions, a geographical unit grouping several census tracts. In that way the number of alternatives for the estimation of the conditional logit is 20, while the subset of choices is formed by the neighborhood that the individual chose and other l9 alternatives (one for each "comuna" in the city) randomly chosen among the different census tracts within each "comuna."

5.1.2 Selection into Labor Force and Correction of the Selection Bias

The estimation of wages and labor supply can suffer from sample selection bias since wages and hours worked are observed only for the share of the population that is employed. That is, those individuals who manage to find a job can be the more educated, the best qualified, or the ones with better abilities, therefore the estimation of wages is biased. In order to control for this potential source of bias, we use standard assumptions of the literature in labor economics and we estimate our second stage equations as a regular Heckman selection correction model, augmented in this second stage with the correction parameters of the selection into the neighborhoods' process [4]. The exclusion restrictions we use in the first-stage equation of the process for sample selection are household variables that we claim to be important determinants of the labor participation, but they are relatively orthogonal to the wage and hours worked. The first variable is the density of public childcare providers in the neighborhood. This variable is generated in the same way as other neighborhood characteristics (see section 4.l). An interaction of this variable with the gender dummy is also included. 0ther variables that describe household composition in terms of children and recent childbirths are also included in the sample selection equation as exclusion restrictions.

6 Results

In this section, we present the results of the estimation of equations 3 and 4, and a table of summary statistics. All the equations are estimated for a sample of individuals who were at least 25 years of age at the time of the interview. In this section, we only present the estimation of the second stage equations. As the reader may recall, selection correction factors were generated from two first-stage equations. The first one is an equation of neighborhood selection, and the other one is a sample selection equation, where the estimation sample is the sample of individuals who have a job. The result of these two former estimations is presented in Appendixes B and C, respectively.

Table 1 shows the descriptive statistics for the estimation sample. The results suggest that, on average, the individuals in our sample have 10.2 years of education, and 43% of the sample attends an educational institution. Twenty percent of the sample has a college degree, while 11% have a junior/community college degree. In addition, most of the individuals in the sample considered themselves as having a mestizo ethnic background.

Table 2 presents the estimation of the three outcomes studied in this paper. Panel [1] presents the results for the estimates of wage, panel [2] presents the estimates for hourly wages, and panel [3]

[4] Each equation includes 21 correction parameters, the traditional sample selection correction parameter, and other 20, one per each alternative in the neighborhood choice set of the residential demand model.

Table 1: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Hours Worked (principal)	14317	49.01837	13.93248	1	72
Montly wage	14317	891586.7	1081087	0	2.0E+07
Nonlabor Income[5]	14317	12.42569	17.57532	0	220
Years of education	14317	10.26346	4.6879	0	21
Attends Educational Establishment	14317	0.043305	0.20355	0	1
Potential Experience	14317	26.24796	13.53844	0	92
Square potential Experience	14317	872.2318	847.2213	0	8464
Complete High School	14317	0.303276	0.459689	0	1
Junior/Community College	14317	0.112733	0.316277	0	1
College	14317	0.204652	0.403461	0	1
Race: Mestizo	14317	0.740239	0.438519	0	1
Race: White	14317	0.214989	0.410829	0	1
Race: Missing	14317	0.016135	0.125998	0	1
Sector: Primary	14317	0.013201	0.114139	0	1
Sector: Industry and Utilities	14317	0.172662	0.377968	0	1
Sector: Construction	14317	0.06845	0.252526	0	1
Sector: Services and Commerce	14317	0.246839	0.431188	0	1
Sector: Transport and Communicatio	14317	0.063631	0.244102	0	1
Laborer-Company Worker	14228	0.039992	0.195946	0	1
Laborer-Government Employee	14228	0.029379	0.168872	0	1
Domestic Worker	14228	0.325766	0.468677	0	1
Self-Employed	14228	0.02713	0.162467	0	1
Density of Homicide	14317	282.2217	144.1982	1	597
Density of Economic Activity	14317	145.5417	151.492	0	655
Minimum Distance to Metro	14317	1279.406	992.4218	9	6621
Density of Child Care Centers	14317	7.083188	5.76602	0	30
Any child born alive in the last 2 year	14317	0.033736	0.180555	0	1
Any child born alive in the last 5 year	14317	0.102256	0.302995	0	1
Children under 6 years at home	14317	0.277013	0.558243	0	5
Children between 6 and 17 years at h	14317	0.693581	0.902433	0	7

Notes

1. The parameter D in the construction of the index (formula A) was set to 2 km, weights within a radius of 1 km centered in the household are 1.
2. The parameter D in the construction of the index (formula H) was set to 2 km, weights within a radius of 1 km centered in the household are 1.
3. Euclidean distance in meters
4. The parameter D in the construction of the index (formula H) was set to 500 mt, weights within a radius of 300 mt centered in the household are 1.
5. Nonlabor Income is in $100.000 Colombian pesos

presents the results of hours worked. Each panel of Table 2 presents two different sets of coefficients. The first set ignores the process of selection into neighborhoods while in the second specification we control for it. As the reader may recall, in this second specification we included 20 parameters for selection correction, one for each alternative of the choice set that resulted from the sampling process explained in section 5.1.1. In addition, the table also presents the estimated coefficient of the lambda parameter; this is the traditional Heckman sample selection coefficient. In order to enhance the interpretation of the coefficients estimated in this model, Table 3 offers a summary of the effects on labor outcomes generated as a result of a standard deviation increase of one in the policy variables.

6.1 Results of Estimated Wage equations

The first result of panel [1] uses the log of monthly wages as dependent variable. The estimates suggest that without correcting for selection into neighborhoods, all the three policy variables are significant at the 10% level with the expected signs. Following the intuition of the theoretical framework presented in section 3, one would expect that individuals who live in low quality neighborhoods would have lower earnings. 0nce we control for the endogeneity of the residential location decision, all the coefficients for the policy variables are smaller in magnitude and the coefficient of the variable distance to a metro station is no longer significant. We observed something very similar in the case of the hourly wage log in panel [2]. Without correcting for selection into neighborhoods, all the three policy variables are significant, but once we control for this sort of selection, the magnitude of the coefficients is reduced, and only the effect of the density of economic units in the neighborhood remains significant. These findings are interesting because they tell us that some of the effects that can be interpreted as redlining (discrimination) or neighborhood effects (low human/social capital in some neighborhoods) may result from a self-selection effect of individuals into their neighborhoods.

The results of our preferred specification, the one where the selection into neighborhood is modeled (specification 2), evidences that the density of business in an individual's neighborhood has a positive and significant effect on the monthly labor earnings. As shown in the second panel of Table 3, a standard deviation (SD) increase of one in the number of businesses in the neighborhood (as defined in section 4.1) increases the monthly wage in 5.4% [5]. This variable is also highly significant when the dependent variable used is the log of hourly wage (panel [2]). In this case, a 1 SD increase in the number of businesses in the neighborhood (as defined in section 4.1) raises the hourly wage by 3.7%. There can be plenty of reasons to explain this positive effect, such as the ones proposed by the mismatch hypothesis, which suggests that individuals in better neighborhoods are expected to have better wages because they may enjoy the possibilities of enhanced levels of social/human capital in

[5] This effect is obtained by multiplying the coefficient by one standard deviation of the variable

those neighborhoods. This could be an effect of demand as well, in the sense that a higher amount of labor demanders in a given location with a fixed area may create incentives for companies to offer higher salaries.

The density of murders is another variable which is statistically significant at the 10% level in our preferred specification of the log of monthly wages. In this case, a 1 SD increase in this violence index reduces the individuals' monthly wages by 2%. This phenomenon is rather a quantity effect than a price effect because this variable is not significant in the regression with log of wage rate per hour as a dependent variable, but it is significant in the labor supply equation, as we will see below. No other policy variables were found to be significant in our preferred specification for wages.

The control variables in our preferred specification for wages showed to have the expected signs and significance. In the case of hourly wages, we find an important positive and significant return to an additional year of education. In addition, the dummy variables for completed junior college and completed college were found to be significant and to have important effects on wages: university/college degrees increase the wage per hour by 56%, as compared to individuals who have an educational level under high school. The dummy variable for those who considered themselves as whites is significant and positive, which implies a positive wage gap of 12% as compared to individuals belonging to a minority ethnic background (afro-descendant, indigenous population). Many fixed effects of occupational characteristics and economic sectors are significant. Several selection correction parameters are significant as well.

6.2 Results of Estimated Labor Supply Equations

The first estimation of panel [3], the one using the log of hours worked as dependent variable, shows that, without correcting for selection into neighborhoods, the density variables of economic units and homicides were found to be significant and with the expected effects. As in the previous estimation, one would expect that individuals living in high quality neighborhoods work more hours, as better labor market outcomes are usually associated in the literature to higher neighborhood quality. 0nce we control for the endogenous residential location decision, the coefficients of these variables are smaller in magnitude, but still significant at the 10% level. These results are interesting because they tell us that the effects that are usually interpreted as redlining or contextual effects on the individuals' labor supply can be overestimated if the selection into neighborhood process is not considered. The preferred specification for the estimation of labor supply evidences that labor supply is sensitive to neighborhood quality. As shown in Table 3, a 1 SD increase in the density of homicides reduces the number of hours worked by almost 1%, an effect which is significant at the 5% level. In addition, the density of economic units in the neighborhood has a positive and significant effect on labor supply. A 1 SD increase in the density index of business in the neighborhood increases the hours worked by 1.3%.

The estimated coefficients of the covariates in the preferred specification for labor supply have the expected signs and significance. We find negative and significant effects of college degree on hours worked. In fact, having a university/college degree reduces the hours worked in more than 10% as compared to individuals with less than high school. This effect can be explained because more educated individuals have good quality jobs with fixed schedules, while unskilled workers with low education have informal jobs or need to work more hours given the low wage rates per hour they may earn. We also find that potential experience increases the labor supply in a non-linear way. Many fixed effects of occupational characteristics and economic sector are significant. Several selection correction parameters are significant as well.

Table 2: Estimation Results

Variables	[1]:Log(Monthly Wage)				[2]:Log(Wage/Hours)				[2]:Log(Hours)			
	coef	se	coef	se[5]	coef	se	coef	se[5]	coef	se	coef	se[5]
Log(Wage/Hour)									0.04294	0.00340***	0.04238	0.00603***
Non Labor Income	0.00742	0.00064***	0.00488	0.00100***	0.00753	0.00066***	0.00491	0.00100***	-0.00028	0.00029	-0.00006	0.00036
Density of Homicide[1]	-0.00032	0.00008***	-0.00014	0.00008*	-0.00021	0.00008**	-0.00005	0.00008	-0.00008	0.00003***	-0.00006	0.00003*
Density of Economic Activity[2]	0.00039	0.00007***	0.00035	0.00007***	0.00027	0.00007***	0.00024	0.00007***	0.00009	0.00003***	0.00009	0.00002***
Minimum Distance to Metro[3]	-0.00002	0.00001*	-0.00001	0.00001	-0.00002	0.00001*	-0.00001	0.00001	0.00000	0.00000	0.00000	0.00000
Educational Attainment	0.02387	0.00650***	0.02199	0.00754***	0.02335	0.00670***	0.02158	0.00744***	-0.00195	0.00246	-0.00192	0.00285
Attends Educational Establishment	-0.21303	0.05523***	-0.20921	0.05839***	-0.17287	0.05710***	-0.17066	0.06312***	-0.04482	0.02113**	-0.04292	0.02702
Potential Experience	0.02365	0.00338***	0.02134	0.00345***	0.02062	0.00348***	0.01872	0.00336***	0.00306	0.00126**	0.00272	0.00151*
Potential Experience2	-0.00044	0.00010***	-0.00041	0.00010***	-0.00043	0.00010***	-0.00040	0.00010***	-0.00003	0.00003	-0.00003	0.00004
Complete Secondary	0.03426	0.03963	0.03323	0.04145	0.05656	0.04095	0.05538	0.04276	-0.01942	0.01547	-0.01929	0.01663
Technician Technological	0.24443	0.06280***	0.22677	0.06302***	0.27306	0.06482***	0.26001	0.06393***	-0.02873	0.02420	-0.03305	0.02413
Higher Education	0.58839	0.08026***	0.47786	0.09374***	0.64948	0.08274***	0.56518	0.09477***	-0.08070	0.03235**	-0.10332	0.03436***
Race: Mestizo	0.00379	0.05529	-0.00904	0.05504	0.03359	0.05729	0.01920	0.05553	-0.03320	0.02182	-0.03090	0.02009
Race: White	0.11837	0.05792**	0.09433	0.05715*	0.14899	0.06002**	0.12521	0.05818**	-0.03928	0.02286*	-0.03827	0.02132*
Race: Missing	-0.03005	0.09239	-0.03816	0.06827	-0.03881	0.09554	-0.05208	0.07101	-0.00012	0.03610	0.00574	0.03452
Sector: Primary	0.03970	0.08221	0.00884	0.12338	-0.01340	0.08372	-0.04355	0.12603	0.05312	0.03202*	0.05369	0.03670
Sector: Utilities+Industry	0.12667	0.02669***	0.12137	0.02120***	0.08223	0.02718***	0.07603	0.02089***	0.04039	0.01040***	0.04159	0.00952***
Sector: Construction	0.08740	0.03999**	0.08034	0.03657**	0.07640	0.04084*	0.07071	0.03621*	0.00913	0.01562	0.00802	0.01529
Sector: Service/Commercial	-0.04292	0.02405*	-0.04713	0.02493*	-0.07953	0.02446***	-0.08411	0.02801***	0.04008	0.00935***	0.04058	0.01015***
Sector: Transport/Communications	0.10812	0.04023***	0.09625	0.04120**	-0.01665	0.04104	-0.02696	0.04235	0.12480	0.01570***	0.12366	0.01378***
Laborer/Government Employee	0.20419	0.04917***	0.21629	0.05691***	0.21667	0.05010***	0.22929	0.05645***	-0.02183	0.01919	-0.02277	0.01905
Domestic Worker	0.18414	0.05617***	0.14010	0.03346***	0.24359	0.05674***	0.20911	0.03768***	-0.06761	0.02169***	-0.07511	0.03445**
Self Employed	-0.28102	0.02148***	-0.28893	0.01820***	-0.17889	0.02186***	-0.18476	0.01782***	-0.09267	0.00839***	-0.09451	0.00919***
Employer/Family	-0.07404	0.06151	-0.08744	0.07481	0.10501	0.06261*	0.09519	0.07248	-0.18252	0.02393***	-0.18551	0.03254***
Gender (Man=1)	0.29430	0.07622***	0.27225	0.08608***	0.32348	0.07817***	0.30676	0.08341***	-0.01213	0.02621	-0.01776	0.03189
Correction parameter 2			-0.32737	0.25158			-0.19601	0.24896			-0.11732	0.11200
Correction parameter 3			0.24513	0.24609			0.26346	0.24193			-0.02822	0.09815

Table 2: Estimation Results. Continued from previous page.

Variables	[1]:Log(Monthly Wage)				[2]:Log(Wage/Hours)				[2]:Log(Hours)			
	coef	se	coef	se[5]	coef	se	coef	se[5]	coef	se	coef	se[5]
Correction parameter 4			0.20630	0.23191			0.02859	0.23250			0.17301	0.10909
Correction parameter 5			0.43365	0.24142 *			0.27264	0.24820			0.14670	0.07477 **
Correction parameter 6			-0.13826	0.28664			-0.25038	0.28141			0.11972	0.10988
Correction parameter 7			0.20774	0.14426			0.26989	0.14857 *			-0.07038	0.05451
Correction parameter 8			0.16161	0.21014			0.18995	0.21299			-0.03415	0.08144
Correction parameter 9			0.00757	0.15721			0.00217	0.16136			0.00287	0.05579
Correction parameter 10			0.11991	0.19523			0.11736	0.18956			-0.00153	0.06513
Correction parameter 11			0.71132	0.20355 ***			0.73684	0.21234 ***			-0.05567	0.06685
Correction parameter 12			0.21835	0.19084			0.28689	0.18393			-0.07909	0.06050
Correction parameter 13			0.36338	0.21661 *			0.40946	0.22383 *			-0.06153	0.07003
Correction parameter 14			-0.63668	0.28743 **			-0.47709	0.28436 *			-0.13462	0.07785 *
Correction parameter 15			0.12189	0.14945			0.15015	0.14863			-0.03456	0.04898
Correction parameter 16			-0.21462	0.16220			-0.17923	0.16644			-0.02768	0.06270
Correction parameter 17			-0.54172	0.49869			-0.33247	0.50852			-0.18683	0.20229
Correction parameter 18			2.52874	1.45711 *			3.26084	1.32836 **			-0.86956	0.46571 *
Correction parameter 19			-0.46977	0.56488			-0.56663	0.58034			0.11561	0.17465
Correction parameter 20			-0.09205	0.19439			-0.01620	0.19867			-0.07359	0.05764
Correction parameter 1			0.00922	0.01667			0.01192	0.01636			-0.00319	0.00523
Constant	12.44535	0.13771 ***	12.82882	0.34518	12.44535	0.13771 ***	8.88213	0.35775 ***	3.58042	0.05354 ***	3.50146	0.13736 ***
Lambda	0.19387	0.17246	0.168	0.194	0.19387	0.17246	0.406	0.187 **	-0.16605	0.05456 ***	-0.1731912	0.073553 **

Notes:

1. The parameter D in the construction of the index (formula A) was set to 2 km, weights within a radius of 1 km centered in the household are 1.
2. The parameter D in the construction of the index (formula H) was set to 2 km, weights within a radius of 1 km centered in the household are 1.
3. Euclidean distance in meters
4. Coeficient of the Inverse Mill's Ratio in the standard sample selection model
5. Standard errors computed using bootstrap with 100 repetitions

It is important to notice the differences between the effects of policy variables identified by the generalized selection model and those obtained from more naïve specifications. The reader can have a better idea of this by looking at Table 3. Let us consider the effects of the density of homicides on hourly wages: under the naïve specification we find important negative effects of this violence index on wages (3 percentage points per standard deviation), which could be interpreted as some sort of discrimination against workers living in bad neighborhoods. This is the basic intuition of the redlining hypothesis. Using our generalized selection model, we find that this effect is completely gone once we control for the endogeneity of the residential location decision. Regarding the individuals' labor supply, we still find a significant effect of homicides, but the magnitude of the effect decreases in almost 25%. We observe a very similar pattern when it comes to the density of economic activity: the effects are still significant, but the magnitude is substantially smaller in the preferred specification. For example, the reduction of the magnitude of this effect is more than 10% in the wage equations, and around 5% in the labor supply equation. Finally, it is important to mention that when using the naïve specification we find a negative effect of the distance of metro station on wages, which is significant at least at the 10% level, but once we correct for selection into neighborhoods, the effect disappears completely.

Table 3: Effects of policy variables.

Dependent Variable	Without Selection into Neighborhood Correction			With Selection into Neighborhood Correction		
	Density of Homicide					
	Δy/Δx	(Δy/Δx)*sd(x)	t	Δy/Δx	(Δy/Δx)*sd(x)	t
[1]:Log(Monthly Wage)	-0.00032	-0.04663	-4.122	-0.00014	-0.02072	-1.771
Standard Error		0.00008			0.00008	
[2]: Log(Wage/Hours)	-0.00021	-0.02949	-2.522	-0.00005	-0.00720	-0.590
Standard Error		0.00008			0.00008	
[3]: Log(Hours)	-0.00008	-0.01135	-2.599	-0.00006	-0.00875	-1.939
Standard Error		0.00003			0.00003	
	Density of Economic Activity					
	Δy/Δx	(Δy/Δx)*sd(x)	t	Δy/Δx	(Δy/Δx)*sd(x)	t
[1]:Log(Monthly Wage)	0.00039	0.05920	5.435	0.00035	0.05356	4.963
Standard Error		0.00007			0.00007	
[2]:Log(Wage/Hours)	0.00027	0.04138	3.669	0.00024	0.03662	3.365
Standard Error		0.00007			0.00007	
[3]:Log(Hours)	0.00009	0.01357	3.163	0.00009	0.01291	3.422
Standard Error		0.00003			0.00002	
	Minimum Distance to Metro					
	Δy/Δx	(Δy/Δx)*sd(x)	t	Δy/Δx	(Δy/Δx)*sd(x)	t
[1]:Log(Monthly Wage)	-0.000018	-0.00277	-1.655	-0.000012	-0.00183	-1.212
Standard Error		0.00001			0.00001	
[2]:Log(Wage/Hours)	-0.000020	-0.00297	-1.704	-0.000015	-0.00220	-1.355
Standard Error		0.00001			0.00001	
[3]:Log(Hours)	0.000004	0.00058	0.882	0.000005	0.00072	1.131
Standard Error		0.000004			0.00000	

Notes:
Left hand side effects come from a model that ignores the endogeneity of the neighborhood choice.
Right hand side effects are based on our preferred specification, where the endogeneity of neighborhood choice is controlled for by using a model of multinomial selection.

7 Conclusions

In the literature for labor economics there are several studies linking an individual's labor outcomes to residential segregation or more general measurements for neighborhood quality. Many empirical studies have found evidence to support that hypothesis (Dickerson, 2008; Weinberg et al, 2004; Altonji and Mansfield, 2011) suggesting that residential location matters on the determination of wages and labor supply because individuals living in segregated or bad quality neighborhoods tend to do worse in the labor market than others living in better neighborhoods.

In this paper, in an urban context, we estimate labor supply and wages paying especial attention to the individual's residential location decision. In order to do this, we estimate a generalized selection model that allows us to control for the possibility of self-selection into neighborhoods. 0ne of the most important conclusions of the paper is that self-selection into residential locations matters, since once we control for an individual's self-selection into the neighborhoods, the effect of neighborhood amenities on an individual's labor outcomes is reduced. Hence, hypothesis that the more basic specification seems to support (significant effects of transportation means availability on wages, for instance) are no longer supported in our final –preferred– specification once we control for selection.

Once we correct for neighborhood selection, some neighborhood characteristics still have significant effects on determining wages and labor supply. The effect of the density of economic activity (density of business) is significant and robust in all our estimated models, even after controlling for the endogeneity of an individual's residential location. A 1 SD increase in the number of businesses in the neighborhood raises the hourly wage in 3.7%. Similarly, a 1 SD increase in the density index of economic activity (business) increases the hours worked by 1.3%. We also find a negative effect of the density of homicides on an individual's labor supply. A 1 SD increase in the density index of homicides reduces an individual's hours worked by almost 1%.

References

[1] Altonji Joseph and Mansfield Richard (2011) "The Contribution of Family, School and Community Characteristics to Inequality in Education and Labor Market Outcomes". Working paper. (http://aida.wss.yale.edu/~jga22/website/research.htm).

[2] Bourguignon, François y Fournier, Martin y Gurgand, Marc (2007), "Selection Bias Corrections Based on the Multinomial Mogit Model: Monte Carlo Comparisons", Journal of Economic Surveys, volumen 21, N.o 1, 174-205

[3] Brueckner, J. y Zenou, Y. (2003), "Space and Unemployment: The Labor-Market Effects of Spatial Mismatch", en Journal of Labor Economics, volumen 21, N.o 1, 242-266

[4] Claire Dujardin, Florence Goffette-Nagot, Neighborhood effects on unemployment?: A test à la Altonji, Regional Science and Urban Economics, Volume 40, Issue 6, November 2010, Pages 380-396.

[5] Zhiming Cheng, Haining Wang (2013). Do neighbourhoods have effects on wages? A study of migrant workers in urban China, Habitat International, Volume 38, April 2013, Pages 222-231.

[6] Dickerson, Niki (2008). Occupational and residential segregation: The confluence of two systems of inequality. Labor Studies Journal 33 (4): 393-411.Jeffrey A. Dubin and Daniel L. McFadden (1984). An Econometric Analysis of Residential Electric Appliance Holdings and Consumption. Econometrica Vol. 52, No. 2 (Mar., 1984), pp. 345-362

[7] Holzer, Harry J. 1991. The spatial mismatch hypothesis: What has the evidence shown? Urban Studies 28 (February): 105–22.

[8] Lee, L. F. (1983) Generalized econometric models with selectivity. Econometrica 51: 507–512.

[9] Ihlanfeldt, Keith R. 1998. The spatial mismatch hypothesis: A review of recent studies and their implications for welfare reform. Housing Policy Debate 9:849–92.

[10] KAIN, J. (1968), "Housing Segregation, Negro Employment, and Metropolitan Decentralization", en The Quarterly Journal of Economics, volumen 82, N.o 2, 175-197.

[11] McFadden, D. (1978): *Spatial Interaction Theory and Planning Models*. Modeling the Choice of Residential Location, pp. 75–96.

[12] Mouw, T. (2000), "Job Relocation and the Racial Gap in Unemployment in Detroit and Chicago, 1980 to 1990", American Sociological Review, volumen 65, N.o 5, 730-753.

[13] Medina Carlos y Tamayo Jorge (2011). An Assessment of How Urban Crime and Victimization Affects Life Satisfaction, Serie Borradores deEconomía # 640.

[14] Medina Carlos, Posso Christian, Tamayo Jorge (2013). The Effect of Adult Criminals' Spillovers on the Likelihood of Youths Becoming Criminals. Borradores de Economía 755.

[15] Morales Leonardo (2013). Peers Effects on a Fertility Decision: An Application for Medellín. Borradores de Economía 777.

[16] Rattelot, Roland (2009). "Wrong Origin or Wrong Neighborhood : Explaining Lower Labor Market Performance of French Individuals of African Origin," Working Papers 2009-04, Centre de Recherche en Economie et Statistique.

[17] Somik V. Lall and Taye Mengistae (2005). The Impact Of Business Environment And Economic Geography On Plant-Level Productivity : An Analysis Of Indian Industry. World Bank Policy Research Working Papers. July 2005

[18] Weinberg, Bruce, Reagan, Patricia Benton y Yankow, Jeffrey Jon, (2004) "Do Neighborhoods Affect Hours Worked? Evidence from Longitudinal Data", Journal of Labor Economics 22,4 (October 2004): 891-825.

[19] Zenou, Y. (2002): "How do firms redline workers?," Journal of Urban Economics, 52(3), 391-408.

[20] Zenou, Y., and N. Boccard (2000): "Racial Discrimination and Redlining in Cities," Journal of Urban Economics, 48(2), 260-285.

Annex A: Residential Location Choice Model

Variable	Residential Location Model			Variable (Continued)	Residential Location Model		
	coef	se	t		coef	se	t
Neighborhood Median Income	0.0218	0.0020	10.87	(SM)xUniversity	-0.0058	0.0075	-0.77
Neighborhood Average Rent	-0.0472	0.0078	-6.04	(SM)x2nd Quartile of Income	0.0128	0.0093	1.37
Homicide Rate (HR)	-0.0225	0.0019	-11.74	(SM)x3rd Quartile of Income	-0.0038	0.0094	-0.40
(HR)x1{Married}	-0.0022	0.0017	-1.28	(SM)x4th Quartile of Income	0.0253	0.0084	3.02
(HR)xUniversity	0.0039	0.0021	1.84	(SM)x1{Automobile}	0.0696	0.0083	8.34
(HR)x2nd Quartile of Income	0.0014	0.0024	0.58	Nightclubs and Casinos* (NC)	-0.0238	0.0084	-2.84
(HR)x3rd Quartile of Income	-0.0009	0.0025	-0.38	(NC)x1{Married}	0.0136	0.0070	1.94
(HR)x4th Quartile of Income	0.0009	0.0025	0.35	(NC)xUniversity	-0.0442	0.0086	-5.16
(HR)x1{Automobile}	0.0069	0.0025	2.78	(NC)x2nd Quartile of Income	0.0088	0.0108	0.82
Economic Activity (EA)	0.0053	0.0009	5.84	(NC)x3rd Quartile of Income	0.0072	0.0107	0.67
(EA)x1{Married}	-0.0004	0.0008	-0.49	(NC)x4th Quartile of Income	-0.0090	0.0097	-0.93
(EA)xUniversity	0.0037	0.0011	3.27	(NC)x1{Automobile}	-0.0287	0.0092	-3.13
(EA)x2nd Quartile of Income	-0.0034	0.0011	-3.09	% of Population with University (%U)	-4.9899	0.2833	-17.61
(EA)x3rd Quartile of Income	-0.0015	0.0012	-1.32	(%U)x1{Married}	-0.2266	0.2261	-1.00
(EA)x4th Quartile of Income	-0.0026	0.0012	-2.10	(%U)xUniversity	3.8913	0.2876	13.53
(EA)x1{Automobile}	-0.0029	0.0013	-2.21	(%U)x2nd Quartile of Income	-1.6420	0.3265	-5.03
Distance to Station (DS)	0.0003	0.0000	14.83	(%U)x3rd Quartile of Income	-2.2918	0.3283	-6.98
(DS)x1{Married}	0.0001	0.0000	3.34	(%U)x4th Quartile of Income	0.8079	0.3195	2.53
(DS)xUniversity	-0.0001	0.0000	-4.01	(%U)x1{Automobil}	4.9746	0.3310	15.03
(DS)x2nd Quartile of Income	0.0000	0.0000	-1.26	Unemployment Rate (UR)	6.4360	0.7485	8.60
(DS)x3rd Quartile of Income	0.0000	0.0000	-1.71	(UR)x1{Married}	-0.8669	0.7216	-1.20
(DS)x4th Quartile of Income	-0.0001	0.0000	-3.20	(UR)xUniversity	-0.8733	1.0337	-0.84
(DS)x1{Automobile}	0.0001	0.0000	3.74	(UR)x2nd Quartile of Income	0.4089	0.9474	0.43
Child Care Centers* (CC)	0.0107	0.0006	16.76	(UR)x3rd Quartile of Income	0.3845	0.9869	0.39
(CC)x1{Married}	0.0006	0.0006	0.93	(UR)x4th Quartile of Income	-0.3931	1.0897	-0.36
(CC)xUniversity	-0.0052	0.0009	-5.90	(UR)x1{Automobile}	-1.3736	1.2846	-1.07
(CC)x2nd Quartile of Income	-0.0007	0.0008	-0.86	Ethnic Minority (EM)	0.8004	0.3629	2.21
(CC)x3rd Quartile of Income	-0.0009	0.0008	-1.03	(EM)x1{Married}	-0.3051	0.3535	-0.86
(CC)x4th Quartile of Income	0.0005	0.0009	0.51	(EM)xUniversity	-0.6835	0.5437	-1.26
(CC)x1{Automobile}	0.0035	0.0011	3.08	(EM)x2nd Quartile of Income	-1.3572	0.4584	-2.96
Recreation/Sports Centers* (RS)	-0.0168	0.0107	-1.57	(EM)x3rd Quartile of Income	-0.9290	0.4810	-1.93
(RS)x1{Married}	-0.0057	0.0097	-0.58	(EM)x4th Quartile of Income	-1.4600	0.5456	-2.68
(RS)xUniversity	-0.0201	0.0129	-1.56	(EM)x1{Automobile}	0.4194	0.6611	0.63
(RS)x2nd Quartile of Income	0.0269	0.0132	2.03	Children per Woman (CW)	-0.9106	0.1116	-8.16
(RS)x3rd Quartile of Income	0.0225	0.0136	1.66	(CW)x1{Married}	-0.1273	0.1067	-1.19
(RS)x4th Quartile of Income	0.0097	0.0144	0.67	(CW)xUniversity	-1.0544	0.1522	-6.93
(RS)x1{Automobile}	0.0081	0.0150	0.54	(CW)x2nd Quartile of Income	0.0109	0.1413	0.08
Cultural Centers and Libraries* (CL)	-0.1277	0.0123	-10.35	(CW)x3rd Quartile of Income	-0.4538	0.1498	-3.03
(CL)x1{Married}	0.0000	0.0110	0.00	(CW)x4th Quartile of Income	0.0171	0.1578	0.11
(CL)xUniversity	-0.0212	0.0146	-1.45	(CW)x1{Automobile}	-0.5907	0.1834	-3.22
(CL)x2nd Quartile of Income	0.0668	0.0153	4.37	% de Involuntary Fasting (IF)	2.1097	0.6920	3.05
(CL)x3rd Quartile of Income	0.0456	0.0157	2.90	(IF)x1{Married}	2.5616	0.6803	3.77
(CL)x4th Quartile of Income	0.0725	0.0164	4.44	(IF)xUniversity	-3.9475	1.0555	-3.74
(CL)x1{Automobile}	-0.0145	0.0170	-0.85	(IF)xCuartil 2 de Ingreso	-3.5860	0.8785	-4.08
Shopping Malls* (SM)	0.0109	0.0073	1.49	(IF)xCuartil 3 de Ingreso	-4.8278	0.9366	-5.15
(SM)x1{Married}	-0.0076	0.0060	-1.27	(IF)xCuartil 4 de Ingreso	-6.9633	1.0511	-6.62
				(IF)x1{Automobile}	-1.1011	1.3231	-0.83

Notes:

The neighborhood median income and average rent are in $100000 Colombian pesos of 2012

Income interactions are built with women non-labor income.

Annex B: Sample Selection Equations

Variables	[1]:Log(Monthly Wage)			[2]:Log(Wage/Hours)			[3]:Log(Hours)		
	coef	se	t	coef	se	t	coef	se	t
Non Labor Income	-0.00331	0.00049	-6.77	-0.00331	0.00049	-6.77	-0.00666	0.00052	-12.88
Density Index of Homicides	0.00014	0.00007	1.87	0.00014	0.00007	1.87	-0.00007	0.00008	-0.96
Density Index of Economic Activity	-0.00003	0.00007	-0.45	-0.00003	0.00007	-0.45	0.00005	0.00008	0.71
Minimum Distance to Metro	0.00002	0.00001	1.52	0.00002	0.00001	1.52	0.00000	0.00001	0.26
Educational Attainment	-0.02961	0.00464	-6.38	-0.02961	0.00464	-6.38	-0.03231	0.00480	-6.73
Attends Educational Establishment	-0.32245	0.04595	-7.02	-0.32245	0.04595	-7.02	-0.38156	0.04880	-7.82
Potential Experience	0.00771	0.00233	3.31	0.00771	0.00233	3.31	0.00564	0.00248	2.27
Potential Experience^2	-0.00064	0.00003	-19.79	-0.00064	0.00003	-19.79	-0.00066	0.00003	-19.36
Complete Secondary	0.08771	0.03354	2.62	0.08771	0.03354	2.62	0.08500	0.03484	2.44
Junior/Community College Degree	0.30288	0.04797	6.31	0.30288	0.04797	6.31	0.36025	0.05008	7.19
Higher Education	0.40825	0.05755	7.09	0.40825	0.05755	7.09	0.60639	0.06038	10.04
Race: Mestizo	-0.08704	0.05159	-1.69	-0.08704	0.05159	-1.69	-0.10142	0.05408	-1.88
Race: White	-0.09056	0.05389	-1.68	-0.09056	0.05389	-1.68	-0.10209	0.05653	-1.81
Race: Missing	-0.21985	0.08002	-2.75	-0.21985	0.08002	-2.75	-0.22408	0.08380	-2.67
Gender (Man=1)	0.60751	0.02696	22.54	0.60751	0.02696	22.54	0.71813	0.02881	24.92
Any child born alive in the last 2 years	-0.06618	0.05956	-1.11	-0.06618	0.05956	-1.11	-0.05907	0.06264	-0.94
Any child born alive in the last 5 years	-0.06486	0.04000	-1.62	-0.06486	0.04000	-1.62	-0.10513	0.04194	-2.51
Children under 6 years at home	-0.00601	0.01804	-0.33	-0.00601	0.01804	-0.33	-0.01223	0.01871	-0.65
Children between 6 and 17 years at home	0.00694	0.00970	0.72	0.00694	0.00970	0.72	0.00050	0.01014	0.05
Density of child care public provaiders (CH)[4]	-0.00783	0.00230	-3.40	-0.00783	0.00230	-3.40	-0.00784	0.00236	-3.32
{Gender} x {CH}	0.02152	0.00294	7.31	0.02152	0.00294	7.31	0.02171	0.00312	6.96
Constant	0.52897	0.08136	6.50	0.52897	0.08136	6.50	0.79869	0.08578	9.31
Observations	27,950			27,950			26,678		
R^2									

Notes

1. The parameter D in the construction of the index (formula A) was set to 2 km, weights within a radius of 1 km centered in the household are 1.
2. The parameter D in the construction of the index (formula H) was set to 2 km, weights within a radius of 1 km centered in the household are 1.
3. Euclidean distance in meters
4. The parameter D in the construction of the index (formula H) was set to 500 mt, weights within a radius of 300 mt centered in the household are 1.

Annex C: Legend Map 1