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 Marcela Cardona‡
Banco de la República
lcardoso@banrep.gov

June 9, 2014

Abstract

Using data from Medellín, second largest city in Colombia, we asses in this paper how a set of neighborhood characteristics determines wages and labor supply for workers in the city. We use GIS data to construct measures of the quality of environments where workers live. This paper focuses in the impact in labor supply and wages of the following set of characteristics: availability of public transportation, crime levels and density of economic activity. The empirical methodology consist of the estimation of linear equations for wages and worked hours, and we control for the selection of individuals into the neighborhoods they are observed. In order to do this we estimate in a first stage a probabilistic model of neighborhood selection from which selection correction terms are obtained; these correction terms and included in the linear equations for wage and worked hours in a second stage. In addition, we control for sample selection as well. We find 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 is still significant and important after we control for the possibility of selection into neighborhoods.

*This work is preliminary and incomplete. All errors are authors responsability.

†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 cities, and in general cities in Latin-America, is the existence of considerable levels of spatial segregation. By spatial segregation in this paper we mean the existence of a clear division of the cities into big clusters of good quality and bad quality neighborhoods. The consequences of this type of city configuration is a topic that have been studied by the literature on labor economics and urban economics. One of the main branches of the literature in this type of issues is the one that study the "spatial mismatch hypothesis." Broadly speaking this hypothesis states that deficient labor outcomes are partly the result of excessive separation between individuals and workplaces (Brueckner y Zenou, 2003).

More generally, it may be possible that spatial segregation of individuals in a city causes deficient wages and labor supply. On one hand, in segregated environments a portion of the population may be excluded from labor opportunities or networks where information on available jobs is exchanged. This type of isolation may cause an increase in economic costs associated to participation in the labor market (Weinberg et al, 2004). On the other hand, there is a set of reasons discussed in the literature to link segregation and wages. These reasons go from scarce accumulation of human/social capital in bad communities (Altonji and Mansfield, 2011) 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 neighborhood characteristics, these characteristics are factors that may increase the cost of being employed or affect the accumulation of social and human capital and by these ways affect wages. This paper seeks to estimate the impact of neighborhood quality on labor supply and wages, using a representative sample of individuals in Medellín (second largest city in Colombia with a population of 3.5 million in the metropolitan area). The quality of the neighborhood is defined in terms of three main characteristics: (1) homicides in the neighborhood, (2) the density of economic units (business) in the neighborhood, (3) the distance to the nearest station of the massive transportation system of the city. The definition of neighborhood we use is the census tract polygons; these units are the building block of the census in Colombia and they are relative small areas for which census information is representative.

2 Literature Review

One of the first studies that explores the Spatial Mismatch hypothesis was Kain, J. (1968), in this paper the author proposes 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 line in topics related with the spatial Mismatch hypothesis, reader may refer to Holzer (1991) y Ihlanfeldt (1998) for a comprehensive review of the literature in this field.

There is some empirical literatures that seek to identify relationships between quality of the neighborhood and labor supply, a good example is Weinberg, Reagan, Yankow, (2004). In their paper these authors estimate a labor supply function specified in terms of some neighborhood characteristics. Weinberg, Reagan, Yankow, (2004) is 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). This former set of papers shares the idea that wages can be explained directly or indirectly by the environments where individuals live or have lived during their lives.

The channels through which the relationship between residential environments and wages may take place are diverse. One possible channel is that environments may alter the process of individual accumulation of social and human capital (Cheng, 2013). Another channel is that employers may discriminate against workers living in particularly areas that can carry the burden of bad stereotypes (Rathelot, 2009). One example of this type of stereotypes is that 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 as redlining. Reader may find deeper exploration 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. Excessive distance between workers and jobs may affect negatively their labor performance, and this is an argument that can be extended to other bad neighborhood characteristics.

3 Theoretical Framework

The main ideas of this research can be represented in a simple static labor supply frame, where individuals get utility from the quality of the neighborhood where they live. A common practice in the literature of implicit prices (Rosen, 1974) is representing a good as a configuration of its characteristics. We represent a neighborhood in the city as a vector $\mathbf{z} = \{z_1, z_2, ..., z_n\}$, where each z_i, $i = 1, 2, ..n$ represents a neighborhood characteristic. To simplify the notation lets assume that all the variation in these characteristics can be summarized in a index $z \in [0, 1]$, where $z = 0$ represents the lowest quality level and $z = 1$ represents the highest neighborhood quality.

Individuals in this framework get utility 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 of 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 constrain is standard, it includes a labor cost parameter for those individuals who work. Labor cost are 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 the 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 in 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 wage, which in this framework is a function of individual characteristics x as education and experience, and it is a function of neighborhood quality z. This way of specifying wage is supported by all the literature that establish an effect of segregation and residential environments on individual earnings (Zenou and Boccard (2000), Zenou (2002), Altonji and Mansfiel, 2011; Cheng, 2012; Rathelot, 2009; Dickerson, 2008). In addition, $a(z, x)$ represent labor cost for individuals who work. They are also function of neighborhood quality z and individual characteristics x. It is assumed that better neighborhood quality reduces the labor cost because it implies better transportation systems, no additional expenditures in individual's security, etc. Therefore, it is assumed that $a\acute{} < 0$ y $a\acute{} > 0$. In addition, p_z is the average price of an additional unit of neighborhood quality y v represent non labor income. Al 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 de restriction represented by equation (2) and the restriction time. From this process the individual obtain optimal consumption for leisure (l), consumption good (c) and neighborhood quality (z).

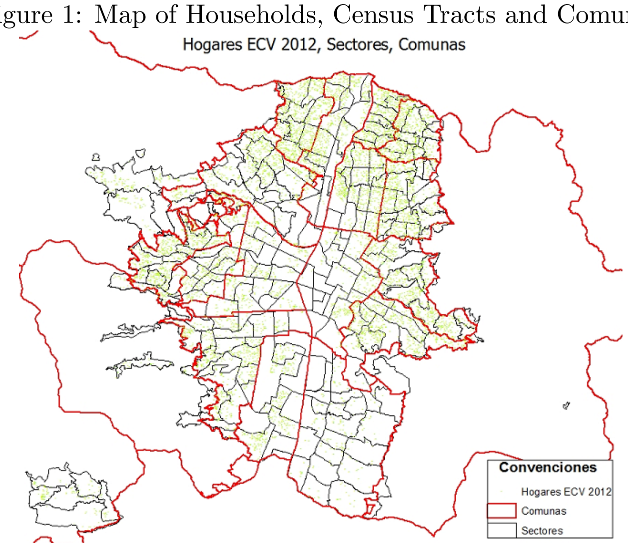
4 Data

The data used in this analysis comes from different sources of information: (1) the Medellin's household living conditions survey or HLCS (Encuesta de Calidad de Vida -ECVM-), (2) the city´s map information updated by the city planning department, (3) homicides records from the National Police and (4) the administrative information of childcare institutions supplied by the local government. 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 the survey is representative to this level of disaggregation.

One of the advantages of this paper is that all the households available in the HLCS are georeferenced and it is possible to identify the exact geographical area where most of the households are located, which is crucial to know the characteristics of individual's neighborhood. In this study we define a neighborhood as a small area within the city with a relative small population. In that sense, we use the 243 census tracts of the city, which account for 9090 inhabitants each one (on average). The Census Tracts are small areas with enough demographic information to characterize each one of them. Another geographical division that is important to define is the "comuna" which are much larger than the Census Tracts and grouped several of them.

Figure (1) shows household maps of 2012's HLCS and the census tracks. The "comunas" are delimited by the red line, the census tracts are the smaller polygons defined by the black lines. The households are represented by green dots in the map.

Figure 1: Map of Households, Census Tracts and Comunas



In this study we use several variables geographically defined. In order to generate these variables, we use georeferenced information on metro stations, information on the location of economic units in the city (formal business dedicated to private or public economic activities), and georeferenced information on murders in the city[1]. In addition, in order to construct some exclussion restrictions for the sample selection equation, we use georeferenced information on public childcare providers in the city[2].

4.1 Policy Variables Measurement

The impact variables have been built using a methodology that considers the geographical location of the individual relative to the neighborhood's characteristic of interest. The access to the transport system is measured through the distance in meters from the individual´s residence to his/her nearest metro station or any station of the massive public transportation system in the city. In the case of the density of business and the crime index, gravitational indexes are built using the inverse of the distance between the individual´s residential location and the characteristic of interest. In the case of business, the index can be 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 activity 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 (business) in the city. The 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. We estimate models with different values of D, the especification presented in this paper use the values of D that maximize the fit of the model.

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

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

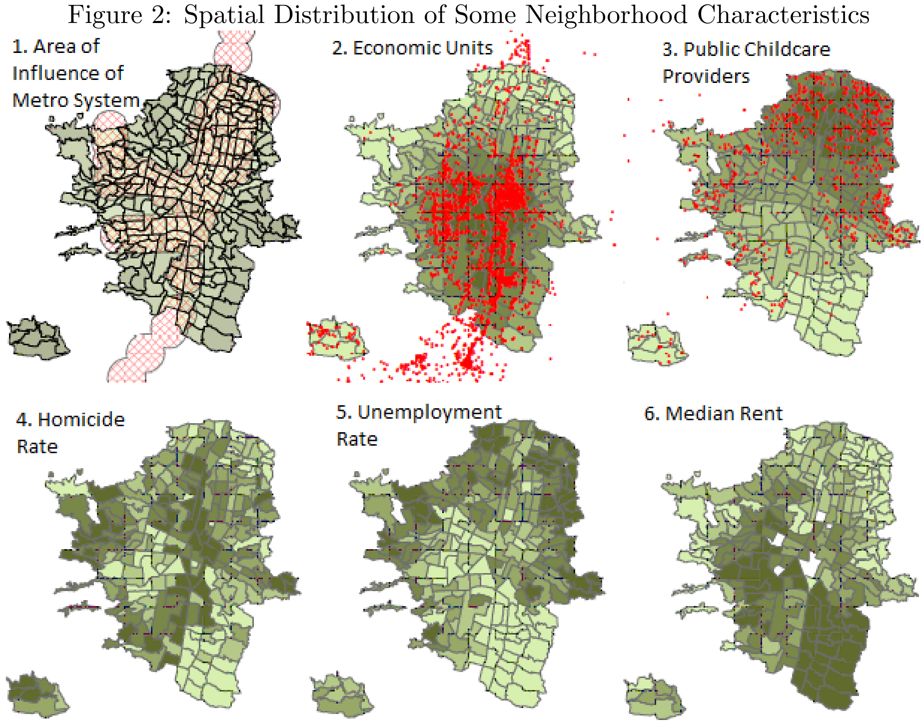
[1] We use data collected by Intelligence department of National Police, where all murders in the city are recorded with the address where it happened. Then, using the address we georefereced each murder.

[2] We use information provided by Good Start Program. This is a big public childcare program in the city that collects information on most of the public childcare providers.

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 of each map is presented Appendix C. The first map shows the area of influence of the massive transportation system in the city. This is the set of metro stations, bus stations that feed the metro system, and metro-cable stations (cable air corridor). The area of influence is defined 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 transportion system by comparing this map with other maps in figure 2. For example, the areas where the average rent is higher (south-east in map 6) the transportion system is lacked.

Figure 2: Spatial Distribution of Some Neighborhood Characteristics



The second map shows the economic units of the city represented by red dots. The economic units are the best proxy variable 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 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 exclusion restriction in our sample selection equation; this procedure is explained in the next section. Map number four represents the distribution of the homicide rate per census tract in the city. Map number five represents the distribution of unemployment rate by census tract in the city. Finally, map number six represents the distribution of the median rent by census tract in the city.

5 Metodology

There are three estimable equations that would be derived from the economic optimization process sketched in section number 3: an equation for the optimal labor supply, a wage equation, and a residential location demand equation[3]. The main purpose of this paper is the estimation of the unbiased effect of some neighborhood characteristics in the first two equations, which can be represented by the following expressions:

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

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

Where h_{is} represents the hours worked by individual i, which is observed living in neighborhood s. In addition, w_i represents the hourly wage of individual i. The matrix X_i contains individual's characteristics, the matrix Z_s contains characteristics of the neighborhood s. The interest of this study is the estimation of parameters in vectors γ, they describe the impact of neighborhood characteristics on labor supply and wage.

5.1 Self-Selection into Neighborhoods Bias Correction

A possible source of bias in the estimation of equation (3) and (4) is that individuals choose the neighborhoods where they live. This can be seen as a selection process which can bias the coefficients in (3) and (4), especially those ones in vectors γ. The bias would take place if this self-selection into neighborhoods process is driven by unobserved factors that are correlated with perturbation terms ε_i. In order to control for this selection process, we estimate generalized selection models. This methodology allows 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, reader may find a survey of the alternatives available in the literature in Bourguignon et. al (2007).

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

The idea of a generalized selection model is specifying 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, which are functions of choice probabilities and are generated from the first stage. A more detailed description of the methodology is offered in the following paragraphs.

A reasonable hypothesis is that, at least partially, errors in equation (3) and (4) are correlated with unobservables driving the residential location decision of workers in the city. Therefore, it is important control for the possibility of selection into neighborhood bias. This practice has started to gain strength in the literature since more and more researchers are considering the possibility that location is an endogenous factor. Reader may see for example Lall and Mengistae (2005) in a generalized selection correction model applied to problem of firms location decisions in India.

The part of our model that explain the process of selection into neighborhoods is in its own right a model of residential location demand, where 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 a neighborhood of the city. In particular, each of the neighborhoods is defined as a census tract of the city. Assuming that each individual i derives an utility level y^*_{is} from choosing the neighborhood s, this level of utility is modelled as a linear function in the parameters as follows:

$$y^*_s = z_s\acute{}\theta + \sum_l \left[x_{i,l} \times z_s\acute{}\right] \theta^I + u_{is}, \quad s = 1, .., K \tag{5}$$

where $x_{i,l}$ represents the l-th characteristic of the individual that is interacted with each of the elements in the z vector. The whole term $\sum_l \left[x_{i,l} \times z_s\acute{}\right]$ contains the interactions between the characteristics of the s choice and the individual variables x_l of individual i. The 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 individual´s characteristics i. Notice for instance that the availability of public transportation or another amenity would provide different utility to different households according to their demographic characteristics (i.e., income, household composition, etc).

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 especial characteristic of conditional models where characteristics vary by alternative and not by individual. To simplify notation let us call ω_{is} any of the dependent variables of the models that will be estimated (logarithm of monthly wage, logarithm of hourly wage, logarithm of worked hours). 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 other 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 than individual i choose 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 to produced unbiased estimators of equations (3) and (4) (for more details the reader may refer to Bourguignon et. al., (2007)). The approach followed in this study is the one implemented by Dubin and McFadden (1984). This methodology consists in the inclusion of the conditional expectations of the error term in equations (3) and (4) given unobservable associated to each residential location alternative. Dubin and McFadded (1984) 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) 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 evaluate the precision and unbiasness of the models through Monte Carlo experiments. The main result shows that in most of the cases the methodology proposed by Dubin and McFaddedn (1984) presents a better performance than other methodologies like the method proposed by Lee(1983). Bourguignon et. al. (2007) concludes that type Dubin-McFadden methodologies have in general a good performance. In fact, the Monte Carlo experiments indicate that correction models of selection bias based on a multinomial logit provides a satisfactory correction bias.

The idea of a multinomial selection model is not restricted to the use of a multinomial logit, there are other models that share the same assumptions about the error distribution of the selection equation. For instance, when the choice is a geographical location, the conditional logit is very convenient since it allows to model the utility of each alternatives 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 authors model the location choice of firms and they also use 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

Even though it is possible to estimate a conditional logit for all the possible neighborhoods in the city (243 census tracts in total), a model with that many alternatives can be difficult to manage. In this paper we follow a common result found in previous literature (McFadden, 1975) 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 where the set of alternatives is built through a random sampling process.

In the literature, there are several methodologies of 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 partitioning the choice set. In this study we use as partitions the "comunas", a geographical unit that groups several census tracts in its interior. In that way the number of alternatives for the estimation of the conditional logit is 20, and the subset of choices is formed by the neighborhood that the individual chose and other 19 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 Bias Correction

In the estimation of wage and labor supply equations there exist also the possibility of sample selection bias since wages and worked hours are observed only for the share of the population that is working. In order to control for this potential source of bias, we use standard assumptions of the literature in labor economics and estimate our second stage equations as a regular Heckman selection correction model, augmented with the correction parameters of the selection into neighborhood process[4]. The exclusion restrictions we use in the first stage equation of the process of 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 worked hours. The first variable is the density of public childcare providers in the neighborhood, the variable is generated in the same way as other neighborhood characteristics (see section 4.1); an interaction of this variable with the gender dummy is also included. Other variables that describe the 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 we present a table of summary statistics. All the equations are estimated for a sample of individuals who were at least 25 years old 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 are generated from two first stage equations. The first one is an equation of neighborhood selection and another 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 Appendix B and C, respectively.

Table 1 presents summary statistics of the estimation sample. The individuals in our sample have 10.2 years of education, 43% of our estimation sample attend to an educational institution, 20% of the sample have college degree, and 11% of the sample have junior/community college degree. In addition, most of the sample considered themselves having a mestizo ethnic background.

Table 2 present the estimation of three outcomes, panel [1] presents the results of an estimation with dependent variable log of monthly wage, panel [2] present the results of an estimation with dependent variable log of per hour wage, and panel [3] presents the results of an estimation with dependent variable log of worked hours. In each panel table 3 presents a set of two different results.

[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

The first estimation in each panel ignores the selection into neighborhood process, and the second present an estimation where we control for the selection into neighborhood process. As reader may recall in this second specification we include 20 selection correction parameters one per each alternative of the choice set that results from the sampling process explained in section 5.1.1. All the selection correction parameters are presented, the parameter lambda is the traditional Heckman sample selection coefficient.

6.1 Results of Estimated Wage equations

The first estimation of panel [1], the one with log of monthly wage as dependent variable, show that without controlling for the process of selection into neighborhoods, all three policy variables are significant (10% of significance at least) and with expected effects. Following the intuition of the theoretical framework presented in section 3, one would expect that individuals who lived in low quality neighborhoods will have lower labor earnings. Once we control for the endogenous residential location decision, all policy variable coefficients are smaller in magnitude and the coefficient of the variable distance to metro station is no longer significant. We observed something very similar in the case of log of hourly wage, panel [2]. Without controlling for selection into neighborhoods all three policy variables are significant, but once we control for selection into neighborhoods the magnitude of the coefficients reduces considerable, and only the effect of the density of economic units in the neighborhood remains significant. This set of results is 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) can be the result of a self-selection effect of individuals into their neighborhoods.

The results of our preferred specification, the one where the selection into neighborhood is modeled, show evidence that the density of business in the individual's neighborhood has a positive and significant effect in the monthly labor earnings, an increment of one standard deviation in the number of business in the neighborhood (as defined in section 4.1) increment the monthly wage in 5.4%[5]. This variable is also highly significant when the dependent variable is log of hourly wage (panel [2]). In this case an increment of one standard deviation in the number of bussines in the neighborhood (as defined in section 4.1) raises the hourly wage in 3.7%. There can be plenty of reasons to explain this positive effect, from the literature of mismatch hypothesis we could say 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 those neighborhoods. This could be also a demand effect, in the sense that higher amount of labor demanders in a physical space may create firms' incentives to offer higher

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

salaries.

Other variable that is significant in the preferred specification of log of monthly wages equation is the density of murders, in this case, an increment of one standard deviation in this violence index causes an reduction of 2% in monthly wage, this is probably an quantity effect and not a price effect because this variable is not significant in the regression with log of wage rate per hour as dependent variable, but it is significant in the labor supply equation, as we will see in the analysis of results of the labor supply equation. No other policy variables are significant in our preferred specification.

Control variables in wage equation (preferred specification) have expected and significant coefficients. In the case of the log wage-hour equation, for instance, we find important positive and significant returns of an additional year of education. In addition, dummy variables for completed junior college and completed college are significant and have important effects in wages, university/college degree increments wage per hour in 56% in comparison with individuals who have educational attainment less than high school. The dummy variable for white self-reported ethnic background is significant and positive, which imply a positive wage gap of 12% in comparison to individuals belonging to a minority ethnic background (afrodescendant, 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 Wage equations

The first estimation of panel [3], the one with log of worked hours as dependent variable, show that without controlling for the process of selection into neighborhoods, the variables density of economics units and density of homicides are significant and with expected effects. As in the case of wage, one would expect that individuals who live in high quality neighborhoods will work more hours. Once we control for the endogenous residential location decision, the coefficients of these variables are smaller in magnitude, but still significant (10% level at least). This results is interesting because is telling us that the effects that can be interpreted as redlining or contextual effects on the individual´s labor supply can be overestimated if the selection into neighborhood process is not taken into account. The preferred specification for the labor supply equations shows evidence that labor supply is sensible to neighborhood quality. An increment of one standard deviation in the density of homicides reduces in almost one percent the number of worked hours. In addition, the density of economic units in the neighborhood has a positive and significant effect in labor supply. An increment of one standard deviation in the density index of business in the neighborhood increment the worked hours in 1.3%.

Control variables in labor supply equations (preferred specification) have expected coefficients. We find negative and significant effects of college degree in worked hours. University/college degree reduced worked hours in almost 11% in comparison with individuals with educational attainment less than high school. This effect can be explained because more educated individuals have good quality jobs with fixed schedules, on the other hand, unskilled workers with low education have informal jobs or need to work more hours given they have jobs with low wage rates per hour. Potential experience increment labor supply in a nonlinear 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	t	coef	se	t	coef	se	t	coef	se	t	coef	se	t	coef	se	t
Non Labor Income	0.00750	0.00064	11.70	0.00496	0.00075	6.63	0.00757	0.00066	11.44	0.00495	0.00077	6.43	-0.00023	0.00030	-0.77	-0.00001	0.00033	-0.04
Density of Homicide[1]	-0.00033	0.00008	-4.19	-0.00015	0.00008	-1.86	-0.00021	0.00008	-2.56	-0.00005	0.00008	-0.64	-0.00008	0.00003	-2.58	-0.00006	0.00003	-1.93
Density of Economic Activity[2]	0.00039	0.00007	5.47	0.00036	0.00007	4.89	0.00027	0.00007	3.70	0.00024	0.00008	3.23	0.00009	0.00003	3.14	0.00008	0.00003	2.93
Minimum Distance to Metro[3]	-0.00002	0.00001	-1.69	-0.00001	0.00001	-1.13	-0.00002	0.00001	-1.73	-0.00001	0.00001	-1.28	0.00000	0.00000	0.88	0.00000	0.00000	1.08
Years of Education	0.02473	0.00650	3.80	0.02283	0.00646	3.53	0.02380	0.00669	3.56	0.02202	0.00666	3.31	-0.00171	0.00246	-0.70	-0.00168	0.00247	-0.68
Attends Educational Establishment	-0.20577	0.05525	-3.72	-0.20201	0.05495	-3.68	-0.16927	0.05709	-2.97	-0.16712	0.05681	-2.94	-0.04252	0.02122	-2.00	-0.04063	0.02128	-1.91
Potential Experience	0.02311	0.00338	6.84	0.02080	0.00337	6.17	0.02035	0.00348	5.85	0.01845	0.00347	5.31	0.00290	0.00126	2.30	0.00256	0.00127	2.02
Potential Experience[2]	-0.00042	0.00010	-4.37	-0.00039	0.00010	-4.02	-0.00042	0.00010	-4.24	-0.00039	0.00010	-3.96	-0.00003	0.00003	-0.83	-0.00002	0.00003	-0.62
Complete Secondary	0.03176	0.03957	0.80	0.03075	0.03939	0.78	0.05526	0.04088	1.35	0.05408	0.04070	1.33	-0.02008	0.01552	-1.29	-0.01994	0.01556	-1.28
Junior/Community College Degree	0.23611	0.06278	3.76	0.21847	0.06275	3.48	0.26877	0.06478	4.15	0.25572	0.06476	3.95	-0.03137	0.02429	-1.29	-0.03568	0.02446	-1.46
College Degree+	0.57687	0.08025	7.19	0.46637	0.08337	5.59	0.64346	0.08271	7.78	0.55910	0.08587	6.51	-0.08503	0.03246	-2.62	-0.10757	0.03384	-3.18
Race: Mestizo	0.00598	0.05518	0.11	-0.00686	0.05490	-0.12	0.03462	0.05717	0.61	0.02023	0.05688	0.36	-0.03248	0.02190	-1.48	-0.03018	0.02197	-1.37
Race: White	0.12061	0.05782	2.09	0.09656	0.05755	1.68	0.15005	0.05989	2.51	0.12625	0.05963	2.12	-0.03858	0.02294	-1.68	-0.03755	0.02302	-1.63
Race: Missing	-0.02427	0.09225	-0.26	-0.03239	0.09179	-0.35	-0.03587	0.09537	-0.38	-0.04916	0.09492	-0.52	0.00148	0.03622	0.04	0.00735	0.03632	0.20
Sector: Primary	0.03944	0.08221	0.48	0.00856	0.08180	0.10	-0.01346	0.08372	-0.16	-0.04363	0.08340	-0.52	0.05297	0.03203	1.65	0.05353	0.03204	1.67
Sector: Utilities+Industry	0.12665	0.02669	4.75	0.12135	0.02656	4.57	0.08215	0.02718	3.02	0.07595	0.02708	2.80	0.04045	0.01040	3.89	0.04165	0.01040	4.00
Sector: Construction	0.08657	0.03997	2.17	0.07952	0.03981	2.00	0.07588	0.04082	1.86	0.07020	0.04070	1.73	0.00894	0.01563	0.57	0.00783	0.01565	0.50
Sector: Service/Commercial	-0.04297	0.02406	-1.79	-0.04719	0.02394	-1.97	-0.07953	0.02446	-3.25	-0.08412	0.02437	-3.45	0.04005	0.00934	4.29	0.04056	0.00934	4.34
Sector: Transport/Communications	0.10760	0.04022	2.68	0.09571	0.04003	2.39	-0.01695	0.04103	-0.41	-0.02728	0.04088	-0.67	0.12462	0.01571	7.93	0.12347	0.01572	7.86
Laborer/Government Employee	0.20405	0.04916	4.15	0.21617	0.04891	4.42	0.21639	0.05010	4.32	0.22903	0.04990	4.59	-0.02177	0.01920	-1.13	-0.02270	0.01920	-1.18
Domestic Worker	0.18547	0.05623	3.30	0.14140	0.05627	2.51	0.24422	0.05681	4.30	0.20971	0.05694	3.68	-0.06697	0.02165	-3.09	-0.07446	0.02174	-3.43
Self Employed	-0.28083	0.02148	-13.07	-0.28876	0.02138	-13.51	-0.17875	0.02186	-8.18	-0.18464	0.02178	-8.48	-0.09264	0.00839	-11.04	-0.09448	0.00839	-11.26
Employer/Family	-0.07402	0.06151	-1.20	-0.08745	0.06118	-1.43	0.10512	0.06261	1.68	0.09528	0.06235	1.53	-0.18257	0.02393	-7.63	-0.18557	0.02392	-7.76
Gender (Man=1)	0.27437	0.07659	3.58	0.25244	0.07629	3.31	0.31327	0.07853	3.99	0.29666	0.07827	3.79	-0.01827	0.02633	-0.69	-0.02389	0.02642	-0.90
Correction parameter 2				-0.32862	0.25781	-1.27				-0.19709	0.26267	-0.75				-0.11732	0.10071	-1.16
Correction parameter 3				0.24407	0.24159	1.01				0.26296	0.24620	1.07				-0.02862	0.09446	-0.30
Correction parameter 4				0.20477	0.24669	0.83				0.02706	0.25143	0.11				0.17308	0.09647	1.79

Table 2: Estimation Results, contunied form previous page

Variables	[1]:Log(Monthly Wage)						[2]:Log(Wage/Hours)						[2]:Log(Wage/Hours)					
	coef	se	t	coef	se	t	coef	se	t	coef	se	t	coef	se	t	coef	se	t
Correction parameter 5				0.43287	0.22179	1.95				0.27251	0.22607	1.21				0.14631	0.08677	1.69
Correction parameter 6				-0.13968	0.25407	-0.55				-0.25184	0.25893	-0.97				0.11974	0.09931	1.21
Correction parameter 7				0.20648	0.13698	1.51				0.26911	0.13962	1.93				-0.07078	0.05360	-1.32
Correction parameter 8				0.16098	0.19452	0.83				0.18986	0.19828	0.96				-0.03458	0.07613	-0.45
Correction parameter 9				0.00591	0.14878	0.04				0.00126	0.15164	0.01				0.00220	0.05823	0.04
Correction parameter 10				0.11918	0.15098	0.79				0.11691	0.15384	0.76				-0.00170	0.05905	-0.03
Correction parameter 11				0.71000	0.17553	4.04				0.73566	0.17888	4.11				-0.05571	0.06874	-0.81
Correction parameter 12				0.21805	0.16144	1.35				0.28653	0.16453	1.74				-0.07910	0.06320	-1.25
Correction parameter 13				0.36157	0.18766	1.93				0.40771	0.19129	2.13				-0.06165	0.07346	-0.84
Correction parameter 14				-0.63994	0.20937	-3.06				-0.47909	0.21340	-2.24				-0.13567	0.08203	-1.65
Correction parameter 15				0.12151	0.13650	0.89				0.14988	0.13912	1.08				-0.03469	0.05341	-0.65
Correction parameter 16				-0.21604	0.14730	-1.47				-0.18021	0.15012	-1.20				-0.02812	0.05767	-0.49
Correction parameter 17				-0.54403	0.51558	-1.06				-0.33271	0.52555	-0.63				-0.18818	0.20168	-0.93
Correction parameter 18				2.52618	1.08039	2.34				3.25773	1.10120	2.96				-0.87077	0.42294	-2.06
Correction parameter 19				-0.46952	0.45641	-1.03				-0.56551	0.46516	-1.22				0.11483	0.17862	0.64
Correction parameter 20				-0.09289	0.14469	-0.64				-0.01673	0.14745	-0.11				-0.07389	0.05665	-1.30
Correction parameter 1				0.00918	0.01321	0.69				0.01193	0.01346	0.89				-0.00322	0.00517	-0.62
Lambda[4]	0.15	0.17	0.85	0.12	0.17	0.70	0.40	0.18	2.27	0.38	0.18	2.16	-0.18	0.05	-3.28	-0.19	0.05	-3.40
Log(Wage/Hours)													0.04	0.00	12.62	0.04	0.00	12.39
Constant	12.47	0.14	90.40	12.85	0.29	44.13	8.45	0.14	59.55	8.89	0.30	29.90	3.59	0.05	66.81	3.51	0.11	30.59
Observations	27,950			27,950			27,950			27,950			26,678			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. Coeficient of the Inverse Mill's Ratio in the standard sample selection model

7 Conclusions

There are several branches of the labor economics literature that links labor outcomes to residential segregation or more general measures of neighborhood quality. Many empirical studies have found evidence to support hypothesis based on this relationship (Dickerson, 2008; Weinberg et al, 2004; Altonji and Mansfield, 2011). From a more general perspective, this empirical evidence tell us that residential location matters in the determination of labor outcomes as wages and labor supply because individuals 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 model wage and labor supply determination paying especial attention to the individual´s residential location decision. In order to do this we estimate generalized selection models and in this way we are able to control for the possible self-selection into neighborhoods bias. One of the most important conclusions of the paper is that self-selection into residential locations matters, coefficients of variables of interest have smaller impacts once we control for the process of self-selection. Because of these reason hypothesis that naive specifications seems to support (productivity effects of transportation means availability, for instance) are no longer supported in our final specification.

After controlling for self-selection into residential location, some neighborhood characteristics have still a significant effects in the determination of 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 individual´s location. An increment of one standard deviation in the number of business in the neighborhood raises the hourly wage in 3.7%. Similarly, an increment of one standard deviation in the density index of economic activity (business) increments the worked hours in 1.3%. We also find a significant and negative effect of the homicides density on labor supply. An increment of one standard deviation in the homicides density index reduces individual´s worked hours in 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