Universitat Autònoma de Barcelona

THREE EMPIRICAL ESSAYS ON URBAN ECONOMICS

JORGE ANDRES DOMINGUEZ MORENO

Universitat Autònoma de Barcelona

Doctoral Thesis in Economics

THREE EMPIRICAL ESSAYS ON URBAN ECONOMICS

By

JORGE ANDRES DOMINGUEZ MORENO

Supervisor

Dr. JOSEP LLUIS RAYMOND BARA

DOCTORADO EN ECONOMÍA APLICADA

DEPARTAMENTO DE ECONOMÍA APLICADA

FACULTAD DE ECONOMÍA Y EMPRESA

UNIVERSIDAD AUTÓNOMA DE BARCELONA

2016

CONTENTS

List of Tables

CHAPTER I

CHAPTER II

CHAPTER III

List of Figures

DEPARTURE

In 1886, Robert Louis Stevenson published *The Strange Case of Dr. Jekyll and Mr. Hyde*, the story of a man who has two extremely different personalities. It is the interplay of good and evil. Some experts, such as the architectural historian Charles McKean, argue that Stevenson was inspired by his hometown, Edinburgh, to write his novel; essentially because of the explicit division between the old and the new city, the south and the north respectively.[1] Edinburgh was a city with two personalities. This is the distinction that can be found on a map of Edinburgh from the year 1820. Characteristics such as the geometry, logic and luxury of the new city contrasted with the chaos, corruption and poverty of the old city. All human beings are a mixture of good and evil, wrote Stevenson, like in cities, where the same density that spreads ideas can spread disease and generate congestion problems (Glaeser, 2011).

Today, more than half of the global population lives in an urban area. According to Edward Glaeser (2011), a city represents the reduction of physical space between people and jobs. A city is the illustration of proximity, density and closeness. Cities enable human beings to work and play together and, accordingly, their success depends on the efficiency of that physical connection.

It is interesting to observe the connection made by Stevenson between cities and human beings, as well as his way of contrasting good and evil. He seems to have in mind the original settlements of humanity, because settlement is an extension and a complement of human beings. We can observe that every city has a relatively similar contrast to the Edinburgh of 1820. Nowadays, we find areas within the cities that are less attractive to residents, even in developed countries. The contrast between good and evil may be blurred, but it does exist.

[1] In this regard, Chesterton (1928) wrote: "it seems to me that the story of Jekyll and Hyde, which is presumably presented as happening in London, is all the time very unmistakably happening in Edinburgh… and especially something that calls up that quality in Edinburgh that led an unkind observer (probably from Glasgow) to describe it as "an east-windy, west-endy place.""

In India, China, Africa and Latin America, cities reflect that contrast more clearly. Bogotá and Cali, the urban areas that we shall study in the following chapters, are used to raise the crucial concerns of cities in developing countries. In the three empirical studies that make up this thesis, the central character is the city, but the main subjects are unemployment, informality and crime.

A city is the confluence between firms and workers. Not all geographical areas within a country provide the same opportunities for production and development. Therefore, firms need to select their locations just as they select their production factors and technology (Capello, 2007). Location determines the productive capacities of firms and the productivity of the areas in which they are located. Moreover, the residence location of workers represents, implicitly, advantageous or disadvantageous opportunities in the labour market because they have to assume commuting costs. Such considerations led Lucas and Rossi-Hansberg (2002) to develop a theoretical model in which firms and workers were competing for land in different locations.

By looking at the map of a city the productive areas can be identified; however, natural concern arises about the relationship between the urban structure and labour market outcomes, principally in big cities. Bogotá, like the majority of large Latin American cities, has experienced urban problems due to the uncontrolled growth of peripheral neighbourhoods and a socio-spatial segregation process that began in the 1950s. The rapid uncontrolled urbanization of the city has resulted in severe urban sprawl and this phenomenon has increased the distance between workers and job opportunities. In Chapter 1, *Job Accessibility and Probability of Employment in Bogotá-Colombia*, we estimate the effect of job accessibility on the probability of being employed. Data used at individual level come from household surveys (2008 and 2009), while information about job location at census tract level comes from the Urban Planning Office (2008). We estimate employment probability equations to analyse the disconnection between workers and job opportunities including controls at individual level. Moreover, the paper focuses on the treatment of the location endogeneity problem using instrumental variables. The main result is that job accessibility has a significant positive effect on the probability of being employed.

Agglomeration economies and transportation costs push the location process in opposite directions, since they simultaneously induce both the concentration and the dispersion of production (Capello, 2007). Indeed, there is a strong positive relationship between productivity and economic density for different industries and levels of aggregation (Ciccone and Hall, 1996; Ciccone, 2002; Brulhart and Mathys, 2008; Melo et al., 2009; Combes et al., 2010; Morikawa, 2011). Nevertheless, most of the empirical findings concern firms in the formal sector, and the literature says little about the effect of agglomeration on the localization of informal firms (Duranton, 2009). In Chapter 2, *Agglomeration and Informal Firms in Cali-Colombia*, we estimate the effect of agglomeration on the local share of informal firms that produce legal goods but do not comply with official regulations. This issue is relevant because, like other developing countries, the informal sector in Colombia employs more than 50% of the workforce. Our results demonstrate that one standard deviation increase in agglomeration reduces the local share of informal firms by 16%. Our results are consistent with the idea that informal firms benefit less from agglomeration because of legal restrictions that block the relationship with formal firms.

There is an association between city size and crime (Glaeser and Sacerdote, 1999). High crime rates represent a significant welfare loss, reducing expected lifespan and increasing uncertainty about the future (Soares, 2010). However, crime rates are not homogeneously distributed within an urban area, and this characteristic has a strong association with neighbourhood quality. In response to crime risk, residents generally have two options: they can vote for anti-crime policies or vote with their feet. When individuals exercise the latter, local response to crime will be observed in the housing market (Gibbons, 2004; Buonanno, et al., 2012). In Chapter 3, *Homicide Rate and Housing Prices in Cali-Colombia*, we analyse this subject. Indeed, Latin America dominates the list of the world's most violent cities. In 2015, Cali (Colombia) registered 65 homicides per 100,000 people in a ranking headed by Caracas (Venezuela) with 120. The literature points out that the local response to crime will be observed in the housing market. The objective of the analysis is to estimate the relationship between housing prices and homicide rates in Cali. We found that a 10% increase in the homicide rate is related with a decrease of between 2% and 2.5% in housing prices.

A transversal challenge in urban economics - and in economics in general - is to isolate effects. Fundamental factors such as agglomeration, housing prices or transport systems seem to reinforce each other to make an urban area. From certain points of view, everything is endogenous. The problem of endogeneity arises through the correlation between the regressors and the random disturbances. So, it is our work to deal with simultaneous effects, unobserved factors, and self-selection issues, among others. Certainly, in the majority of cases, it is not easy to find valid instruments. Indeed, if we use no valid instruments, the estimation will be inconsistent. As a result, the cure may be worse than the disease.

In Chapter 1, we use the historical information of Bogotá to instrument agglomeration, particularly, the neighbourhoods of the city in 1950 and the original settlements of the city. A similar strategy was applied in literature such as in Bawn-Snow (2007), who uses the number of highways in a 1947 national interstate highway plan. Or Duranton and Turner (2012) who use the 1947 plan of the interstate highway system and a 1898 map of railroads. In this vein, García-López (2012) and García-López et al. (2015) use instrumental variables derived from maps of Roman roads and the railroad network at the end of the 19th century.

In Chapter 2 we use a geological variable to instrument agglomeration. Geological variables such as soil composition, rock depth, water capacity, soil erodibility and seismic and landslide hazard have been used for coping with endogeneity (Rosenthal and Strange, 2008; Combes et al., 2010; Combes and Gobillon, 2015). Soil characteristics were important to localize original settlements, and agglomeration processes were then developed in those areas. In this case it can be argued that the instrument is valid.

In the third chapter, in which we analyse the relationship between housing prices and homicide rates at neighbourhood level, it is difficult to find a valid instrument. Faced with the difficulty of finding valid instruments, we tried to cope with the endogeneity problem by including on the right hand side of the equation all those variables that may provoke the correlation between regressors and random disturbances.

As a final comment, I expect that both the literature mentioned and the empirical strategies implemented in the following pages, shall serve as a contribution to improve our life in cities, principally in developing countries.

References

Bawn-Snow, Nathaniel (2007). “Did Highways Cause Suburanization?”. *The Quarterly Journal of economics*, 122, 775-805.

Brulhart, Marius and Mathys, Nicole (2008). “Sectoral Agglomeration Economies in a Panel of European Regions”. *Regional Science and Urban Economics*. 38. Pp. 348-362.

Buonanno; Montolio and Raya-Vílchez (2013). “Housing Prices and Crime Perception”. *Empirical Economics*, Vol. 45, Pp. 305-321.

Capello, Roberta (2007). *Regional Economics*. Routledge London and New York.

Chesterton, G. K. (1928). *Robert Louis Stevenson*. NEW YORK DODD, MEAD & COMPANY 1928.

Ciccone, Antonio (2002). “Agglomeration Effects in Europe”. *European Economic Review*. Vol. 46. Pp. 213-227.

Ciccone, Antonio and Hall, Robert (1996). “Productivity and the Density of Economic Activity”. *American Economic Review*. 86 (1). Pp. 54-70.

Combes, Pierre-Philippe and Gobillon, Lauren (2015). “The Empirics of Agglomeration Economies”. In Duranton, Henderson and Strange (editors). *Handbook of Urban and Regional Economics*. Vol. 5. Elsevier (2015).

Combes; Duranton; Gobillon and Roux (2010). “Estimating Agglomeration Effects with History, Geology and Worker Fixed Effects”. In Glaeser, Edward (ed.) *Agglomeration Economies*. Chicago. IL: University of Chicago Press.

Duranton, Gilles (2009). “Are Cities Engines of Growth and Prosperity for Developing Countries?”. In Spence, Annez and Buckely (editors). *Urbanization and Growth*, The Commission on Growth and Development, Washington, D.C.

Duranton, Gilles and Turner, Matthew (2012). “Urban Growth and Transportation”. *Review of Economic Studies*, 01, 1-36.

García-López, Miquel-Ángel (2012).“Urban Spatial Structure, Suburbanization and Transportation in Barcelona”. *Journal of Urban Economics*, 72, Issues 2-3, 176-190.

García-López; Holl; and Viladecans-Marsal (2015).“Suburbanization and Highways in Spain when the Romans and the Bourbons still Shape its Cities”. *Journal of Urban Economics*, 85, Pp. 176-190.

Glaeser, Edward and Sacerdote, Bruce (1999). “Why is there more Crime in Cities”. *Journal of Political Economy*, Vol. 107, No. S6, Pp. S225-S258.

Glaeser, Edward (2011). *Triumph of the City*. Ed. Penguin Press, New York, 2011.

Gibbons, Steve (2004). “The Cost of Urban Property Crime”. *The Economic Journal*, Vol.114, No. 499, Pp. 441-463.

Lucas, Robert and Rossi-Hansberg, Esteban (2002). “On the Internal Structure of Cities”. *Econometrica*. Vol. 70. No. 4. Pp. 1445-1476.

Melo; Graham and Noland (2009). “A Meta-Analysis of Estimates of Urban Agglomeration Economies”. *Regional Science and Urban Economics*. 39. Pp. 332-342.

Morikawa, Masayuki (2011). “Economies of Density and Productivity in Service Industries: An Analysis of Personal Service Industries Based on Establishment-Level Data”. *Review of Economics and Statistics*. Vol. 93. Pp. 179-192.

Rosenthal, Stuart and Strange, William (2003). “Geography. Industrial Organization and Agglomeration”. *The Review of Economics and Statistics*. May 2003, 85 (2), Pp. 377-393.

Soares, Rodrigo (2010). “Welfare Costs of Crime and Common Violence: A Critical Review”. *Pontifical Catholic University of Rio de Janeiro and IZA*.

Stevenson, Robert Louis (1886). *The Strange Case of Dr. Jekyll and Mr. Hyde*. Ed. Longmans, Green & Co. London.

THE TRIPLE BRIDGE

CHAPTER I

Job Accessibility and Employment Probability in Bogotá-Colombia

Summary - *The aim of this paper is to estimate the effect of job accessibility on the probability of being employed in Bogotá-Colombia. Data used at individual level come from household surveys (2008 and 2009), while information about job location at census tract level comes from the Urban Planning Office (2008). We estimate employment probability equations to analyse the disconnection between workers and job opportunities including controls at individual level. Moreover, the paper focuses on the treatment of the location endogeneity problem using instrumental variables. The main result is that job accessibility has a significant positive effect on the probability of being employed.*

JEL-code – R10

Keywords - Job accessibility, labour markets, employment.

1. Introduction

According to Edward Glaeser (2011), cities are the absence of physical space between people and jobs. Cities represent proximity, density, closeness and their success depends on the efficient physical connection. When people live close to their work, there is potential for positive effects, including improved well-being, convenience, accessibility, air quality, and sustainability (Stoker and Ewing, 2014; Fan, 2012). Nevertheless, within cities, people or groups of people face spatial frictions in accessing jobs. This phenomenon is known as spatial mismatch and it is embedded in social structures and labour market processes. John F. Kain (1968) presented one of the first statements of the role that spatial separation might play in creating disparity in employment between different groups of people within urban areas. Since his work, several papers support this hypothesis both theoretically and empirically (e.g. Kasarda, 1990; Ihlanfeldt and Sjoquist, 1998; Gobillon and Selod, 2014).

In North America, the poor tend to be clustered in the main cities, whereas in cities in developing countries, the poor are mostly concentrated in informal settlements in the urban periphery (Kylroy, 2007). This particular location pattern of the poor probably enhances the physical disconnection between workers and job opportunities. Due to a lack of adequate data, few studies analyse the effect of physical disconnection between workers and job opportunities on the labour markets of developing countries (Gobillon and Selod, 2014).

Between 1938 and 2005, the Colombian population increased five-fold, especially the urban population which grew from 2.5 to 30 million inhabitants. Bogotá, like the majority of large Latin American cities, has experienced urban problems due to the uncontrolled growth of peripheral neighbourhoods and a socio-spatial segregation process that began in the 1950s. This rapid uncontrolled urbanization process has resulted in severe urban sprawl and an increase in the distance between workers and job opportunities.

There are some areas within Bogotá where unemployment is high, and this phenomenon persists over time. Figure 1 shows the comparison between the unemployment rates of 2003 and 2011 for 20 *localidades* (administrative districts) of Bogotá using information from quality of life surveys. The horizontal axis measures the unemployment rates of 2011 (when the average unemployment rate for Bogotá was 8.5%), while the vertical axis measures the unemployment rates of 2003 (when the average unemployment rate was 9.5%). The reference lines represent the unemployment rate of the city each year. The *localidades* with a high unemployment rate in 2003 (e.g. Ciudad Bolivar or Bosa) still have high unemployment rates in 2011. Furthermore, *localidades* with a low unemployment rate in 2003 (e.g. Teusaquillo or Usaquen) still have low unemployment rates in 2011.

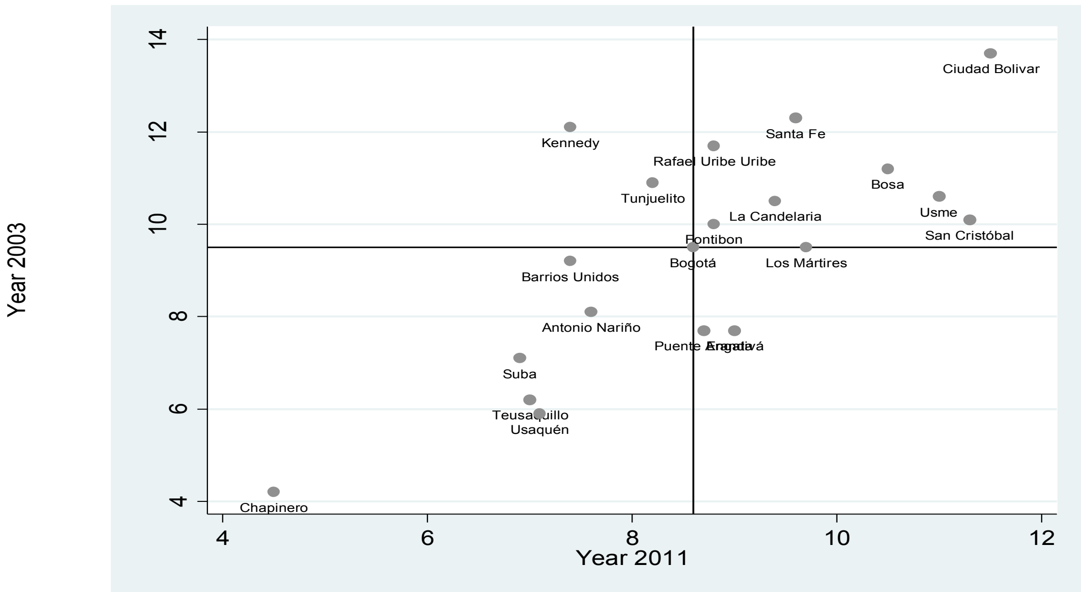


Figure 1–Unemployment rates within Bogotá by *Localidades*, years 2003 and 2011. Source: *Encuesta de Calidad de Vida Bogotá 2003* and *Encuesta Multipropósito Bogotá 2011*.

The aim of this paper is to estimate the effect of job accessibility on the probability of being employed in Bogotá. Data at individual level come from household surveys (2008 and 2009) and the information about number of jobs at census tract level comes from the Urban Planning Office (2008). Furthermore, using instrumental variables we can deal with potential endogeneity problems. We employ the distance between the census tracts and the nearest neighbourhood of Bogotá in 1950 as the source of exogenous variation.

This work is novel since – as far as we know - no previous research has been conducted on the disconnection between job opportunities and workers applied to a Latin American city. Furthermore, we provide a treatment of the location-endogeneity problem using instrumental variables. We present evidence about how geographical distances have a significant economic impact on the labour market, which means that our findings verify theoretical predictions. The main conclusion is that an improvement in job accessibility has a positive impact on the employment probability.

The paper is organized as follows. Section 2 contains a literature review. Section 3 introduces Bogotá as a case study. Section 4 presents data sources and methodological aspects. Section 5 summarizes the main results. Finally, section 6 presents the main conclusions.

2. Literature Review

An interesting work conducted by Kain (1968) analyses how housing segregation affects the distribution and level of Afro-American employment in two metropolitan areas with high levels of segregation: Chicago and Detroit. The author found that housing market segregation may contribute to the high unemployment rates of Afro-Americans. After Kain's (1968) contribution, several studies support his spatial mismatch hypothesis. These studies argue that separation between the location of the residence and job opportunities causes long commuting journeys and high commuting costs, especially in places where private transport is expensive and public transport is not efficient, in other words, conditions that increase the costs of job search (Rospabe and Selod, 2006). Fieldhouse (1999) found that employment is correlated with job density for a few ethnic groups in London, mainly Pakistani and Bangladeshi. For Portland and Atlanta, Sanchez (1999) found that transit accessibility increases labour participation. Weinberg (2000) found that black residential centralization may account for 48–62% of the black–white employment differential among 18–30 year olds living in large metropolitan areas of the United States. Raphael and Stoll (2001) present evidence about metropolitan areas where spatial mismatch contributes to the unemployment of Hispanics and Asians, but to a lesser extent than Afro-Americans. Kawabata (2003) analysed the metropolitan areas of Boston, Los Angeles and San Francisco. The author argues that job accessibility improvement drives up both the probability of being employed and the number of hours worked (the effect is higher for workers without cars).[2] In a recent paper, Gobillon and Selod (2014) note that in European cities, residential socioeconomic patterns are somehow inverted when compared with those observed in the United States. In Europe, minorities are residentially concentrated in peripheral areas. However, in both cases most of the evidence seems to support the spatial mismatch hypothesis.

As regards the effect of job accessibility on women's labour results, Blumenberg (2004) studies the residential and work place location choices of women in multi-person households in Los Angeles. The author argues that women have to make more trips to various places such as schools and shops in addition to commuting to the workplace. Aslund et al. (2010) study the role of job proximity as a determinant of individual labour market outcomes for the case of Sweden. They estimate employment equations and find that the job proximity measure is significant for women's employment status, but small and insignificant for men. Furthermore, they estimate earning models and find that the estimated coefficient related with the job proximity measure is larger for men. Authors conclude that local job proximity is positively correlated with individual employment outcomes. In turn, Matas et al. (2010) analyses the effect of the urban spatial structure on female labour market outcomes in Barcelona and Madrid. The authors show that low job accessibility to public transport negatively affects the probability of being employed.

Gobillon and Selod (2007) analyse the effects of residential segregation and disconnection from jobs in Paris. The authors found that locations where the unemployment rate is highest are also characterized by segregation, but apparently not by low job accessibility. The main conclusion is that neighbourhood

[2] It is important to clarify that some authors study skills mismatch. For instance, Houston (2005) develops a framework, which conceptualizes and reconciles skills mismatch and spatial mismatch within metropolitan areas, incorporating the operation of local housing and labour markets, as well as the role of commuting. The author asserts that skills mismatch and spatial mismatch are likely to reinforce each other. In the case of the United States, Lens (2014) analyses public housing and voucher household proximity to jobs. The author finds strong evidence that public housing residents are typically much closer to low-skilled job opportunities; consequently, they are in circumstances of low-skilled competition. However, we will not discuss this issue in this paper.

segregation is the key factor that prevents unemployed workers from finding a job. Similarly, Dujardin et al. (2008) found that distance to jobs is not significant in explaining the probability of being employed in Brussels; which is coherent with the spatial structure of the city, where a large proportion of unemployed people reside close to jobs.

Due to a lack of adequate data, very few studies exist on the effect of physical disconnection to jobs on the labour market outcomes in developing countries (Gobillon and Selod, 2014). One of these is about South African cities. In this case, Rospabé and Selod (2006) found that distance to jobs has a positive impact on the probability of being unemployed.

Urban development in Latin American cities is quite different in comparison to North American or European cities. In Latin American cities informality is the rule, since housing plans have not been able to give an appropriate response to the fast population growth of capital cities since the 60s (Alinaga-Linares and Álvares-Rivadulla, 2010). Particularly in Bogotá, most of the immigrants are attracted by the opportunities that the capital city offers, but a percentage of them come having no other option since they are displaced by the violence in the countryside. Moreover, there is little information about racial or ethnic segregation in Latin America, partly because the censuses have not systematically asked questions about race and ethnicity.[3]

Olarte (2012) reveals some interesting features about Bogota's residential location pattern. In this case, the author tackles the relationship between job accessibility, employment levels and income in Bogotá. The paper seeks to explain whether accessibility, income and the number of jobs have any influence on public transport improvements or, whether on the contrary, public transport improvements determine levels of accessibility and the income levels of residents. Using Structural Equation Models (SEM), the author tries to capture both the endogeneity and the causality among variables. The author concludes that the city centre of Bogotá has declined from an urban and planning perspective, and this is

[3] In cases where information exists - such as in some Brazilian cities - racial categories appeared to be less of a determinant of residential segregation than income categories compared to US cities (Alinaga-Linares and Álvares-Rivadulla, 2010).

one of the main reasons why the spatial disconnection between workers and job opportunities has begun to become established in the city.

2.1 The location-endogeneity problem

The specialized literature points out that the location-endogeneity problem is inherent to the job accessibility analysis. Ewing, et al. (2007) present the issue in two alternative ways: environmental determinism, when the residential choice is made first and some other outcome follows; and self-selection, when people choose where to live taking into account the way that different environments affect job opportunities.

In order to tackle the location-endogeneity problem, literature recommends estimating regressions with a subsample of individuals who do not have, theoretically, the possibility to decide their residence location. Some works, such as Blumenberg (2004) or Matas et al. (2010), focus their interest on the case of married women. Alternatively, O'Reagan and Quigley (1996a, 1996b, 1998) and Dujardin et al. (2008) focus on youths still living in their parents' home, and assume that the residential location is exogenously determined by parents. However, Ihlanfeldt and Sjquist (1998) and Ihlanfeldt (2006) point out that there are two limitations to this empirical strategy: a) the results are not generalized to adults; and b) the omitted productivity characteristics in equations estimated for youths may still be correlated with measures of neighbourhood job opportunities. In this case, children share behavioural characteristics with their parents (due to either nature or nurture). To handle the bias issue more successfully, Ihlanfeldt (2006) proposes the use of simultaneous models for employment and residential location. This approximation requires the use of some variables that affect residential location but not employment location. Following this idea, some studies have included variables that measure lifestyle preferences in the residential location equation, but exclude these variables from the employment equation.

An alternative methodology is presented in papers such as those by Oreopoulus (2003), Gobillon and Selod (2007) and, more recently, Aslund et al. (2010). They use information on re-localization programs (government subsidized relocation programs) in which residence location choice is exogenous and depends on administrative decisions.

Since information on location preferences is not available, the approach in our paper is derived from literature that deals with the effect of transportation improvements on local population growth in an intra-metropolitan context. Bawn-Snow (2007), Duranton and Turner (2012) and García-López (2012) recommend the use of instrumental variables which represent characteristics of the cities at some time long ago to address the problem of endogeneity.[4] As is well known, the instrumental variables must meet two fundamental characteristics: they must predict job accessibility and they must be uncorrelated with the dependent variable.

3. Bogotá-Colombia

Bogotá spans 33 km from north to south and 16 km from east to west. The city is the main political and economic centre of Colombia. It produces 26% of Colombia's GDP. Founded in 1540, its population exceeds 7.3 million people. The central part of the city contains more than 50 percent of jobs, and most of the people commuting to this zone come from the north, the north-west, the west, and the south-west (Olarte, 2012). The urban area is 380 km^2, but the total area, which also includes 17 adjacent municipalities, covers an area of 1580 km^2. The functional integration of this space is still quite limited (Ruiz et al., 2012) so it seems more appropriate to limit the study to the urban area. The urban structure of Bogotá follows a pattern that is common in many other Latin American cities. In comparison to US cities, they are more compact and more monocentric (Ingram

[4] Bawn-Snow (2007) analyses the extent to which the construction of new limited access highways has contributed to the population decline in city centres of metropolitan areas in the United States. As an instrumental variable, the author uses the number of highways in a 1947 national interstate highway plan. Duranton and Turner (2012) investigate the role of interstate highways in the growth of US cities and use the 1947 plan of the interstate highway system and an 1898 map of railroads. García-López (2012) and García-López et al. (2015) investigate the relationship between the population growth per census tract and transportation improvements for the Barcelona Metropolitan Region, and use instrumental variables derived from maps of Roman roads, as well as from the main roads and the railroad network at the end of the 19th century.

and Carroll, 1981; Almandoz, 2008). During the last twenty years, Bogotá has re-centralized and compacted, urbanizing empty gaps (Cuervo, 2010). Recent studies on the spatial structure of employment emphasize the increasing decentralization of employment and a higher structuring role of road infrastructure axes (Romero, 2010).

The population of Bogotá was approximately 100,000 inhabitants at the beginning of the twentieth century. The urbanization process started to take on importance in the 1930s with the migration of rural families. This migratory phenomenon was essentially the consequence of the crisis in the agro-mineral exportation model and the politics of industrialization by import substitution (Dureau, et al. 2007). Furthermore, one of the most important aspects of Colombian history is the phenomenon of rural violence, which dates back to 1950. In April 1948, the liberal leader Jorge Eliecer Gaitán was assassinated in the centre of Bogotá. After this tragic political episode, which triggered massive riots in the city, violence settled in rural areas where government forces had no presence.

Dureau, et al. (2007) point out that between 1940 and 1970, the population of Bogotá grew at 6 percent annually. The main reason for this growth was the arrival of rural families fleeing war. The dramatic situation was that essentially, poor families had two options when they arrived to the city: either rent accommodation in the historical city centre (in overcrowded tenements), or self-build in peripheral territories (generally occupied illegally). Since 1970 the population growth rate per year reduced by 3 percent (Dureau, et al., 2007). However, competition for land access remains harder for new immigrants in some areas within the city. For instance, in the *localidad* Soacha*,* the population growth rate was 11 percent per year between 1985 and 1993. The case of Soacha is particular because land control was mainly deficient and illegal self-building has been the foremost way of construction for residential purposes.

4. Data and methodology

The geo-statistics zones used in our empirical exercise are census tracts, which are defined by the National Institute of Statistics (*Departamento Administrativo*

Nacional de Estadística, DANE). The census track is the smallest spatial unit for which the number of jobs is available. Specifically, the city is divided into 621 census tracts, in which 607 have employment and 14 have not (the number of jobs is not available in these census tracts because the urbanization process there has been highly informal).

Following Ihlanfeldt (2006), the effect of job accessibility on labour market outcomes can be estimated using ordinary least squares (OLS):

$$E_{is} = \alpha\, z_s + \beta\, X_i + \varepsilon_i, \qquad (1)$$

where E_{is} represents a labour market outcome for the worker i in the neighbourhood s (it is equal to 1 when the individual is employed and 0 otherwise); z_s represents the job accessibility level in the census tract s. It is expected that better access will be associated with a higher probability of being employed (Rogers, 1997). Then, if α is positive and statistically significant, it is possible to say that part of the variability of E_{is} can be attributed to the job accessibility level. X_i represents a standard set of human capital variables for the individual i in a neighbourhood s (gender, marital status, a dummy variable to identify if the worker is the head of the household, years of education, potential experience, and the squared of the potential experience. In addition, we include the number of children at home both under nine years of age and between nine and eighteen years of age. Also, a dummy variable is included which takes the value of 1 for the information of the year 2009 and 0 in the case of 2008). Consequently, we estimate the probability of being unemployed using Logit or Probit estimations (Ihlanfeldt, 2006; Matas, et al. 2010).

Firstly, we need a measure of job accessibility. This variable takes into account the spatial distribution of jobs and the distance or the cost of access to them (Hansen, 1959; Rogers, 1997; Shen, 1998). The variable is calculated as follows:

$$z_s = \int \frac{J_h}{d_{sh}}\, dh, \qquad (2)$$

where z_s represents the accessibility for location s; J_h is the number of relevant opportunities in location h; d_{sh} is the travel time, distance, or cost for a trip from s to h.

As in Matas et al. (2010), the job accessibility variable that we calculate represents the potential employment opportunities of each residential zone in the metropolitan area. The relevant variable should be the number of vacancies but, as data do not supply such a variable at this spatial scale, the total number of jobs located in each zone serves as a proxy for vacancies (Matas et al., 2010). It is expected that zones with a higher number of jobs also generate a larger number of vacancies (Rogers, 1997).

Figure 2 shows the distribution of the log job accessibility in Bogotá as is defined in equation (2).[5] The highest job accessibility levels are located on the eastern edge of the city. The city cannot extend in this direction because of the presence of a chain of mountains (Eastern Cordillera). This is why Bogotá expands to the south and north and, recently, to the west (Olarte, 2012).

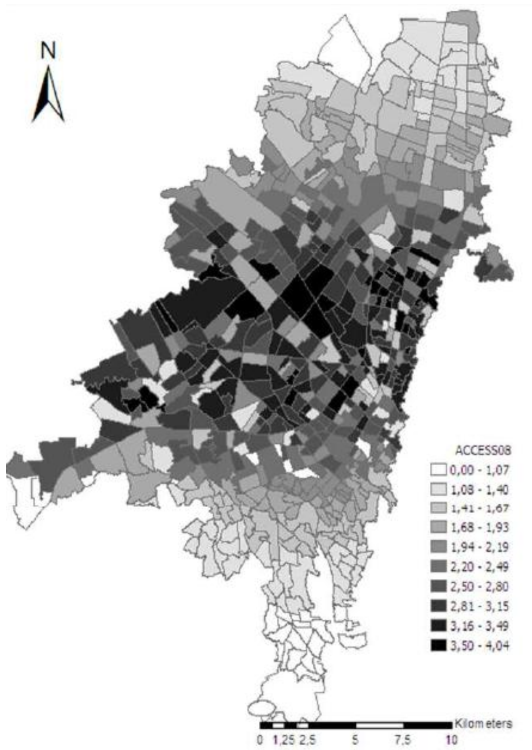


Figure 2 - Job Accessibility for Bogotá in 2008
Source: *Secretaría Distrital de Planeación* (SDP)

[5] Figure A1 in the Appendix shows the histogram of the log job accessibility.

Table 1- Descriptive Statistics

	Mean	Std. Dev.
Years of education	11	4.8
Potential experience	21	14.6
Children under 9	0.608	0.84
Children between 9 and 18	0.635	0.87
Job accessibility	2.01	0.636
Log of job accessibility	2.01	0.636
Distance to neighbourhood in 1950	1.5km	1.1km
Distance to original settlement	3.3km	1.8km
Observations	36,097	

Source: *Gran Encuesta Integrada de Hogares*, GEIH (2008-2009). *Departamento Administrativo Nacional de Estadística*, DANE.

The information on workers' characteristics and their residential location comes from household surveys of 2008 and 2009 (49% of the information is from 2008 and 51% from 2009). Overall statistics on workers are: 52% men and 48% women; 52% married or living with another person as a couple, the remaining 48% are single, divorced, widowed or separated; and 46% of the workers are head of household. Table 1 presents basic information from our database. The mean number of years of education is 11, which is to the end of high school in Colombia. The potential experience is calculated as the age of the worker minus years of education minus 7; this variable has a mean of around 21 years. Variables about the number of children are generated by counting the number of children of a certain age in the home: children aged less than nine years and aged between nine and eighteen years old. The mean and standard deviation of the job accessibility variable and the distances used as instrumental variables are presented at the bottom of the table. The mean distance between actual residence location and the nearest neighbourhood in 1950 is 1.5 kilometres, while the mean distance between actual residence location and the nearest original settlement is 3.3 kilometres.

5. Results

Table 2 presents the estimated marginal effects of the Probit estimation for the full sample and for male and female separately. In general, the marginal effects have the expected sign. Workers' characteristics such as being married, being head of household and the years of education are positively related with the probability of being employed. The variable of experience captures a quadratic relation, which means a positive but decreasing effect on the probability of being employed. The number of children aged less than 9 years old has a positive correlation in the case of men, and for women it is not significant. The number of children between 9 and 18 years old has a negative correlation with the probability of employment for both men and women. The estimated marginal effect of the log job accessibility is positive and statistically significant for the full sample as well as for men and women separately. We found that the effect is higher for women (0.0175) than for men (0.008). These results are in line with Blumenberg (2004), Aslund et al. (2010) and Matas et al. (2010) because women face more negative consequences, derived from the job opportunities disconnection. Nevertheless, the literature mentions that marginal effects estimated by this route are very likely to be biased.

Table 2- Probability of being employed

	Full sample		**Male**		**Female**	
	M. effect	z-Statistic	M. effect	z-Statistic	M. effect	z-Statistic
Job Accessibility	0.0126	4.82***	0.0080	2.45**	0.0175	4.23***
Gender (male=1)	0.0050	1.54				
Married	0.0230	6.68***	0.0250	4.57***	0.0129	2.30**
Head of household	0.0517	14.43***	0.0651	11.02***	0.0324	5.27***
Years of education	0.0038	9.15***	0.0021	4.09***	0.0061	9.46***
Potential experience	0.0042	12.26***	0.0032	7.40***	0.0046	8.18***
Potential experience 2	-0.00005	-7.55***	-0.00005	-6.90***	-0.00002	-1.63
Children under 9	0.0026	1.38	0.0055	2.13**	-0.0004	-0.14
Children between 9 and 18	-0.0051	-2.88***	-0.0044	-2.01**	-0.0063	-2.25**
Year (2009=1; 2008=0)	-0.0260	-8.40***	-0.0265	-6.75***	-0.0247	-5.09***
Observations	35,490		18,593		16,897	

The marginal effect is calculated at the mean of the independent variable.
Statistically significant: 1% ***; 5% **; 10% *.

5.1 Location-endogeneity problem: IV results

The estimated effect of the job accessibility variable could be inconsistent because of the presence of the location-endogeneity problem. In short, if workers with higher probability of being employed choose neighbourhoods with higher job accessibility levels, the estimated marginal effect may be biased. The bias would be avoided only in the unlikely event that control variables fully capture differences in productivity among workers (Ihlanfeldt, 2006). In order to deal with endogeneity, we use historical variables to instrument job accessibility.

Since 1950, rapid urbanization has caused important changes throughout Latin America; indeed, the high population growth rates and the concentration of the best agricultural land in each country have driven the rural poor to the towns and cities in search of work and basic services (Marsh, 1983). By 1940, 71% of the population lived in the countryside or villages with less than fifteen hundred people. However, by 1993, around 60% of the population lived in cities.

The instrumental variable that we use is the distance between the census tract of residence and the nearest neighbourhood in 1950. Figure 3 shows the formal neighbourhoods of Bogotá in 1950. This variable is correlated with job accessibility but not with the probability of being employed - aside from the indirect route via the job accessibility variable - (Cameron and Trivedi, 2005). According to Duranton and Turner (2011), the exclusion restriction requires the orthogonality of the dependent variable and the instrument conditional on control variables. Table A1 in the Appendix shows that the instrument can be excluded from the model of interest since the estimated marginal effect is not significant.

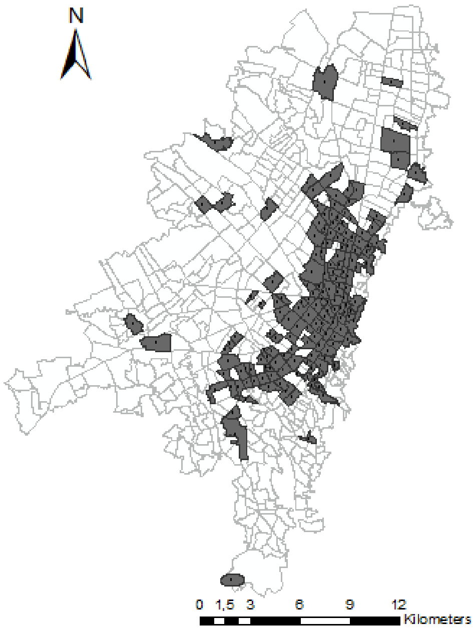


Figure 3 – Urban area of Bogotá in 1950
Source: *Gestión Urbana, Universidad Piloto de Colombia*

Table 3 - Estimated marginal effects from IV: Distance between the census tract of residence and the nearest neighbourhood in 1950

	Full sample		Male		Female	
	M. effect	z-Statistic	M. effect	z-Statistic	M. effect	z-Statistic
Job Accessibility	0.0098	2.25**	0.0056	1.00	0.0152	2.26**
Gender (male=1)	0.0050	1.52				
Married	0.0228	6.63***	0.0248	4.54***	0.0128	2.28**
Head of household	0.0517	14.42***	0.0651	11.02***	0.0324	5.27***
Years of education	0.0039	8.81***	0.0022	4.00***	0.0062	8.99***
Potential experience	0.0042	12.30***	0.0032	7.42***	0.0046	8.20***
(Potential experience)*2	-0.00005	-7.54***	-0.00005	-6.88***	-0.00002	-1.63
Children less than 9	0.0024	1.30	0.0053	2.07**	0.0003	-0.10
Children between 9 and 18	-0.0052	-2.95**	-0.0046	-2.06**	-0.0064	-2.29**
Year (2009=1; 2008=0)	-0.0264	-8.41***	-0.0268	-6.73***	-0.0251	-5.10***
Observations	35,490		18,593		16,897	
F-Stat. (weak instrument)	$(-129.15)^2$		$(-93.96)^2$		$(-88.68)^2$	
Wald test of exogeneity	0.4105		0.5773		0.6616	

The marginal effect is calculated at the mean of the independent variable.
Statistically significant: 1% ***; 5% **; 10% *.

Table 3 presents the results of the IV model. We found that the estimated marginal effect of job accessibility is negative and significant. At the bottom of Table 3 we show the result of the Wald test for exogeneity of the instrumented variable (the result indicates that there is evidence to not reject the null hypothesis of exogeneity). Additionally, we report the F-statistic to test if the instrument is weak; the result is above Stock and Yogo (2005)'s rule of thumb (F-stat. > 10), consequently the null hypothesis of weak instrument is rejected. Table 2A in the Appendix presents the first-stage regression of the model, where we check if the correlation between job accessibility and the instrument is negative and significant.

Table 4 summarizes the estimated coefficients of log job accessibility obtained from both estimation strategies. Results from the instrumental variables method show that the marginal effect is positive and significant. We found that if the log of job accessibility variable increases by 1%, the probability of being employed increases by 0.001% for the full sample, 0.006% for men and 0.015% for women. We conclude that the effect of job accessibility on the probability of being employed is higher for women than for men. As has been pointed out in other literature (e.g. Blumenberg, 2004; Matas et al., 2010), we provide evidence that women are more sensitive to local labour market conditions in Bogotá than men.

Table 4- Marginal effects of log job accessibility

	Full sample		**Male**		**Female**	
	M. effect	z-Statistic	M. effect	z-Statistic	M. effect	z-Statistic
Probit	0.0126	4.82***	0.0080	2.45**	0.0175	4.23***
IV Probit	0.0098	2.25**	0.0056	1.00	0.0152	2.26**

The marginal effect is calculated at the mean of the independent variable.
Statistically significant: 1% ***; 5% **; 10% *.

An alternative instrument that we can use is the distance between census tracts and the original settlements of the city. Figure A2 in the Appendix shows the original settlements of Bogotá, and Table A3 shows the approximate foundation year of the settlements. Then we estimate the IV model using this variable as an instrument for job accessibility. Table A4 in the Appendix shows that the instrument can be excluded from the model of interest since the estimated marginal effect is not significant. We found that the estimated marginal effect of job accessibility on the probability of being employed is positive and significant.

5.2 Educational levels

Job accessibility is a proxy of the spatial distribution of job opportunities within the city, and we found a positive relationship between this variable and the employment probability. However, it is important to take into account that some workers have more of a capacity to move or commute within the city. Particularly, it is likely that highly skilled workers are less sensitive to local labour markets conditions. Table 5 shows the estimated marginal effects for workers who have 11 years of education or more (skilled) and workers with less than 11 years of education (unskilled). In Colombia, 11 years of education means that the individual has finished high school. Results for the full sample indicate that the marginal effect of workers who have less than 11 years of education is higher than the marginal effect of workers who have 11 or more years of education, 0.0146 and 0.0072 respectively. When we separate regressions for men and women we find that the marginal effect is positive and significant for unskilled men (0.0161), and for skilled women (0.0161). At the bottom of the table we report the F-statistic. In all cases there is evidence to reject the null hypothesis of weak instruments (Stock and Yogo (2005)'s rule of thumb (F-stat. > 10)).

Table 5- Estimated marginal effect for skilled and low-skilled workers

	Full sample		**Male**		**Female**	
	M. effect	z-Statistic	M. effect	z-Statistic	M. effect	z-Statistic
Less than 11 years of education	0.0146	2.00**	0.0161	1.69*	0.0131	1.18
F-Stat. (weak instrument)		$(-79.08)^2$		$(-58.67)^2$		$(-52.87)^2$
More than 11 years of education	0.0072	1.35	-0.0006	-0.09	0.0161	1.93*
F-Stat. (weak instrument)		$(-101.79)^2$		$(-73.17)^2$		$(-70.74)^2$

The marginal effect is calculated at the mean of the independent variable.
Statistically significant: 1% ***; 5% **; 10% *.

5.3 Impact of accessibility increase on the probability of employment

Using the estimated coefficients it is possible to calculate the change in the probability of being employed when the job accessibility for all the individuals in the sample is set at least equal to the average value of this variable for the census tracts in the highest quintile of its distribution (Matas, et al. 2010). In our case, this modification implies an increase of 50% in job accessibility. Subsequently, we compare the predictions of the model in Table 4 with the simulated predictions.

Table 6 reports the difference between both predictions. The first column is the predicted value of the employment probability using the observed values of job accessibility (baseline prediction). The second column shows the predicted growth in the employment probability with the simulated job accessibility increase. The third column reports the difference. The results show that an increase in job accessibility has a positive effect on the probability of employment, and this effect is higher for women than for men.

Table 6- Impact of accessibility increase on the probability of employment

	Baseline prediction	Simulated value	Difference
Full sample	90.04	91.25	1.21
Male	91.13	91.94	0.81
Female	88.86	90.49	1.63

6. Conclusions

The urban structure is a potential determinant of labour market results because the social or physical distance affects job search costs. Indeed, the rapid urbanization of cities in Latin America may increase the distance between job opportunities and workers. In this paper, we test whether job accessibility matters for individual employment in Bogotá - Colombia. Since 1950, this city has experienced an uncontrolled growth phenomenon of peripheral neighbourhoods and a socio-spatial segregation process, probably affecting labour market outcomes.

For the empirical exercise, we use information about job location for the year 2008 while the information about characteristics of the workers comes from household surveys from the years 2008 and 2009. After controlling individual characteristics of workers, the main result is that job accessibility is a significant determinant of the probability of being employed, especially for women. Therefore, investments in public infrastructure that increase accessibility to employment would lower unemployment.

To control potential simultaneity problems the regressions were estimated using instrumental variables related with historical characteristics of the city.

References

Alinaga-Linares and Álvarez-Rivadulla (2010). "Residential Segregation in Bogotá across Time and Scales". *Lincoln Institute of Land Policy, Working Paper*, Product Code: WP10LA1.

Almandoz, Arturo (2008) "Despegues sin madurez. Urbanización, industrialización y desarrollo en la Latinoamérica del siglo XX" *EURE*, Vol XXXIV, nº 102, pp. 61-76

Aslund, Olof; Osth, John, and Zenou, Yves (2010). "How Important is Access to Jobs? Old Question - Improved Answer". *Journal of Economic Geography*, 10 (3): 389-422.

Bawn-Snow, Nathaniel (2007). "Did Highways Cause Suburanization?". *The Quarterly Journal of economics*, 122, 775-805.

Blumenberg, Evelyn (2004). "En-gendering Effective Planning: Spatial Mismatch, Low-Income Women, and Transportation Policy". *Journal of the American Planning Association*, 70(3): 269-281.

Cameron, Colin and Trivedi, Pravin (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Cuervo, Luís M. (2010) "Expansión metropolitana y globalización en Bogotá". *V Encuentro Red Iberoamericana de Investigadores sobre globalización y territorio*. Toluca, México

Dujardin, Claire; Selod, Harris; and Thomas, Isabelle (2008). "Residential Segregation and Unemployment: The Case of Brussels". *Urban Studies*, 45(1), 89-113, January 2008.

Duranton, Gilles and Turner, Matthew (2011). "The Fundamental Law of Road Congestion: Evidence from US Cities". *American Economic Review*, 101, Pp. 2616-2652.

Duranton, Gilles and Turner, Matthew (2012). "Urban Growth and Transportation". *Review of Economic Studies*, 01, 1-36.

Dureau, Françoise; Barbary, Olivier;Gouëset, Vincent; Pissoat, Olivier and Lulle, Thierry (2007). *Ciudades y Sociedades en Mutación: LecturasCruzadas sobre Colombia*. Institut de Recherche pour le Développement (IRD), Institut Francais D'Études Andines (IFEA) and Universidad Externado de Colombia.

Ewing, Reid; Bartholomew, Keith; Winkelman; Walters, Jerry; and Chen, Don (2007). *Growing Cooler: The Evidence on Urban Development and Climate Change*. Publisher: Urban Land Institute.

Fan, Yingling (2012). "The Planners War against Spatial Mismatch: Lessons Learned and Ways Forward". *Journal of Planning Literature*, 27, Pp. 153-169.

Fieldhouse, Edward (1999). "Ethnic Minority, Unemployment and Spatial Mismatch: The Case of London". *Urban Studies*, 36, 1569-1596.

García-López, Miquel-Ángel (2012)."Urban Spatial Structure, Suburbanization and Transportation in Barcelona". *Journal of Urban Economics*, 72, Issues 2-3, 176-190.

García-López; Holl; and Viladecans-Marsal (2015)."Suburbanization and Highways in Spain when the Romans and the Bourbons still Shape its Cities". *Journal of Urban Economics*, 85, Pp. 176-190.

Glaeser, Edward (2011). *Triumph of the City*. Ed. Penguin Press, New York, 2011.

Gobillon, Lauren and Selod Harris (2014). "Spatial Mismatch, Poverty and Vulnerable Populations". *Handbook of Regional Science*, pp. 93-107. Springer-Verlag Berlin Heidelberg, 2014.

Gobillon, Lauren and Selod, Harris (2007). "The Effect of Segregation and Spatial Mismatch on Unemployment: Evidence from France". *CEPR*, Working Paper 6198.

Hansen, Walter (1959). "How Accessibility Shapes Land Use". *Journal of the American Institute of Planners*, 25, pp. 73-76.

Houston, Donald (2005). "Employability, Skills Mismatch and Spatial Mismatch in Metropolitan Labour Markets". *Urban Studies*, 42, pp. 221-243.

Ihlanfeldt, Keith (2006). "A Primer on Spatial Mismatch within Urban Labor Markets". *A Companion to Urban Economics*. Edited by Richard J. Arnott and Daniel P. McMillen, Blackwell Publishing Ltd.

Ihlanfeldt, Keith and Sjquist, David (1998). "The Spatial Mismatch Hypothesis: a Review of Recent Studies and their Implications for Welfare Reform". *Housing policy Debate, 9(4), pp. 849-892.*

Ingram, Gregory and Carroll, Alan (1981) "The Spatial Structure of Latin American Cities". *Journal of Urban Economics*, 9, pp. 257-273.

Kain, John (1968). "Housing Segregation, Negro Employment, and Metropolitan Descentralization". *The Quarterly Journal of Economics*, Vol. 82, No. 2 (May, 1968), pp. 175-197.

Kain, John (1992). "The Spatial Mismatch Hypothesis: Three Decades Later". *Housing Policy Debate*, 3 (2), pp. 371-460.

Kasarda, John D. (1990). "City Jobs and Residents on a Collision Course: The Urban Underclass Dilemma". *Economic Development Quarterly*, 4(4), pp. 313-319.

Kawabata, Mizuki (2003). "Job Access and Employment among Low-skilled Autoless Workers in US Metropolitan Areas". *Environment and Planning A 2003*, volume 35, pp. 1651-1668.

Kilroy, Austin (2007). "Intra-Urban Spatial Inequality: Cities as "Urban Regions"". *World Development Report, Reshaping Economic*. Geography, Background Paper, Washington, DC: World Bank.

Lens, Michael (2014). "Employment Accessibility Among Housing Subsidy Recipients". Housing Policy Debate, Vol. 24, No. 4, Pp. 671-691.

Manski, Charles (2003). "Identification problems in the Social Sciences and Everyday Life". *Southern Economic Journal,* 2003, 70(1), 11-21.

Martinez, Sergio (2010). "Síntesis de la Problemática de las Áreas Desarrolladas Informalmente". *Secretaria Distrital de Planeación*, Alcaldía Mayor de Bogotá: http://www.slideshare.net/smmtocan/evolucion-urbana-informal-en-bogota

Marsh, Robin Ruth (1983). *Development Strategies in Rural Colombia: The Case of Caquetá*. UCLA Latin American Center Publications, University of California, Los Angeles.

Matas, Anna; Raymond, Josep-Lluis; and Roig, Josep-Lluis (2010). "Job Accessibility and Female Employment Probability: The Cases of Barcelona and Madrid". *Urban Studies*, 47 (4) 769-787, April 2010.

Olarte, Carlos Augusto (2012). "The Relationship among Improvements of Urban Transport System, Income and Job Accessibility: an Endogenous Question?". *Centre d'Economie de la Sorbonne* (CES); University of Paris 1 Pantheon-Sorbonne.

O'Reagan, Katherine and Quigley, John (1996a). "Teenage employment and the spatial isolation of minority and poverty households". *Journal of Human Resources*, 31, pp. 692-702.

O'Reagan, Katherine and Quigley, John (1996b). ""Spatial effects upon teenage outcomes: the Case of New jersey Teenagers". *New England Economic Review*, May/June, pp. 41-58.

O'Reagan, Katherine and Quigley, John (1998). "Where Youth Live: Economic Effects of Urban Space on Employment Prospects". *Urban Studies*, 35, pp. 1187-1205.

Oreopoulos, Philip (2003). "The Long-Tun Consequences of Living in a Poor Neighborhood". *The Quarterly Journal of Economics*, November 2003.

Raphael, Steven (1998). "The Spatial Mismatch Hypothesis and Black Youth Joblessness: Evidence from the San Francisco Bay Area". *Journal of Urban Economics*, 43, pp. 79-111.

Raphael, Steven and Stoll, Michael (2001). "Can boosting minority car-ownership rates narrow inter-racial employment gaps?" In: Rothenberg Pack J, Gale W (eds) *Brookings-Wharton Papers on Urban Economic Affairs 2001*, Brooking Paper Institution Press, 99-145.

Rogers, Cynthia (1997). "Job Search and Unemployment Duration: Implications for the Spatial Mismatch Hypothesis". *Journal of Urban Economics*, 42, pp. 109-132.

Romero, J.A. (2010) "Transformación Urbana de la Ciudad de Bogotá, 1990-2010: efecto espacial de la liberalización del comercio". *Perspectiva Geográfica* Vol 15, pp. 85-112.

Rospabé, Sandrine and Selod, Harris (2006). "Urban Unemployment in the Cape Metropolitan Area, in Post Apartheid South Africa". In: Bhorat, H. and Kanbur, R. (ed), *Human Science Research Council*, Pretoria, South Africa, chapter 7, pp. 262-287.

Ruiz, Nancy; Roca, Josep and Moix, Montserrat (2012) "Análisis de la estructura metropolitana mediante la detección de subcentros de empleo. El caso de Bogotá" *ACE* nº 18, pp. 297-324.

Sanchez, Thomas (1999). "The Connection between Public Transport and Employment: The Cases of Portland and Atlanta". *Journal of the American Planning Association*, 65, pp. 284-296.

Shen, Qinq (1998). “Location Characteristics of Inner-city neighborhoods and employment accessibility of Low-wage Workers”. *Environment and Planning B*: Planning and Design 1998, volume 25, pages 345-365.

Stock, James and Yogo, Motohiro (2005). “Testing for Weak Instruments in Linear IV Regressions”. In Andrews and Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge University Press, Cambridge, Pp. 80-108.

Stoker, Philip and Ewing, Reid (2014). “Job-Worker Balance and Income Match in the United States”. *Housing Policy Debate*, Vol. 24, No. 2, Pp. 485-497.

Townshend, Ivan and Walker, Ryan (2002). “The Structure of Income Residential Segregation in Canadian Metropolitan Areas”. *Canadian Journal of Regional Science*, XXV: 1, pp. 25-52.

Weinberg, Bruce (2000). “Black Residential Centralization and the Spatial Mismatch Hypothesis”. *Journal of Urban Economics”*. Vo. 48 (1), pp. 110-134.

Appendix

Figure A1- Histogram of log job accessibility

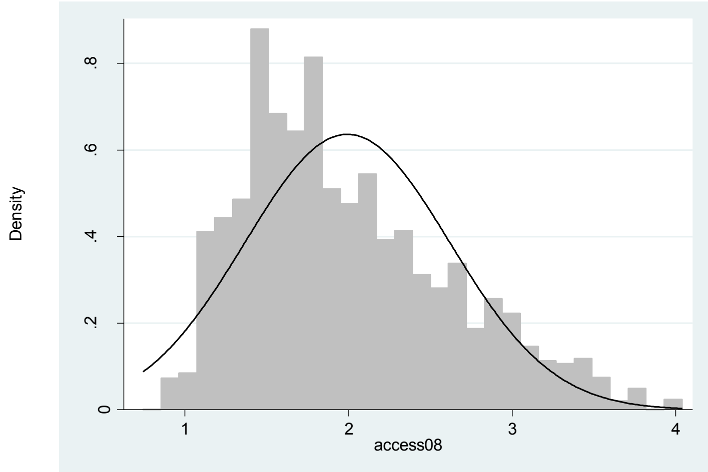


Table A1 – Exclusion restriction: Distance between the census tract of residence and the nearest neighbourhood in 1950

	Total		**Male**		**Female**	
	M. effect	z-Statistic	M. effect	z-Statistic	M. effect	z-Statistic
Job Accessibility	0.0141	4.44***	0.0094	2.33**	0.0188	3.75***
Dist. Neighbourhood 1950	**0.0013**	**0.82**	**0.0012**	**0.56**	**0.0011**	**0.44**
Gender (male=1)	0.0050	1.54				
Married	0.0230	6.68***	0.0250	4.56***	0.0129	2.30**
Head of household	0.0517	14.43***	0.0651	11.02***	0.0324	5.27***
Years of education	0.0038	9.13***	0.0021	4.08***	0.0061	9.46***
Potential experience	0.0042	12.26***	0.0032	7.40***	0.0046	8.17***
Potential experience 2	-0.00005	-7.54***	-0.00005	-6.90***	-0.00002	-1.62
Children under 9	0.0026	1.38	0.0055	2.13**	-0.0004	-0.15
Children between 9 and 18	-0.0050	-2.88***	-0.0044	-2.01**	-0.0063	-2.25**
Year (2009=1; 2008=0)	-0.0260	-8.41***	-0.0265	-6.76***	-0.0248	-5.10***
Observations	35,490		18,593		16,897	

The marginal effect is calculated at the mean of the independent variable.
Statistically significant: 1% ***; 5% **; 10% *.

Table A2 - First stage: Dependent variable log job accessibility instrumented by distance between census tract of residence and nearest neighbourhood of 1950

	Total		Male		Female	
	Coeff.	z-Statistic	Coeff.	z-Statistic	Coeff.	z-Statistic
Dist. Neighbourhood 1950	-0.3097	-129.15***	-0.3068	-93.96***	-0.3130	-88.68***
Gender (male=1)	-0.0094	-1.62				
Married	-0.0313	-5.31**	-0.0304	-2.92***	-0.0308	-3.35
Head of household	-0.0059	-0.94	0.0005	0.04	-0.0068	-0.65
Years of education	0.0302	44.14***	0.0306	33.68***	0.0300	28.37***
Potential experience	0.0040	6.34***	0.0021	2.38**	0.0057	6.10***
Potential experience2	0.00001	0.23	-0.00003	1.87*	-0.00002	-0.50
Children under 9	-0.0353	-11.31***	-0.0326	-7.42***	-0.0380	-8.37***
Children between 9 and 18	-0.0312	-10.31***	-0.0290	-7.04***	-0.0344	-7.56***
Year (2009=1; 2008=0)	-0.0808	-15.19***	-0.0794	-10.93***	-0.0820	-10.49***
Observations	35,490		18,593		16,897	
R-squared	0.42		0.42		0.42	

The marginal effect is calculated at the mean of the independent variable.
Statistically significant: 1% ***; 5% **; 10% *.

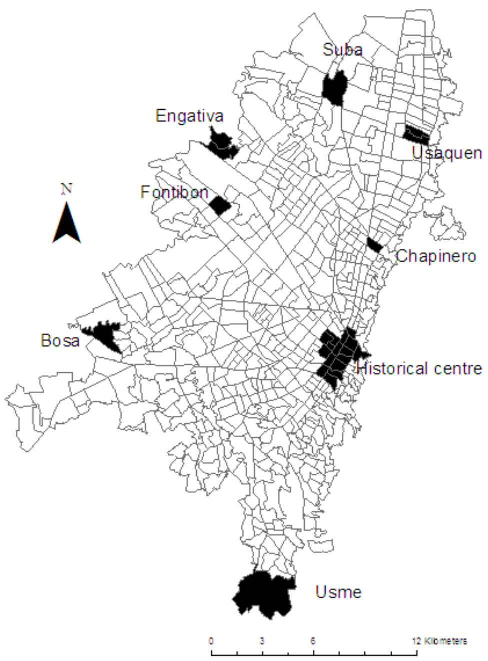


Figure A2- Original settlements of Bogotá

Table A3 - Year of the original settlements

Original settlement	year	Description
Historical city centre	1538	Gonzalo Jiménez de Quesada founded what would become the city of Bogotá.
Bosa	1538	Meeting place of three conquerors: Jiménez de Quesada, Nicolás de Federmán and Sebastián de Belalcázar.
Chapinero	1812	The place began to be populated by artisans and potters after independence.
Engativa	Before 1492	It dates back to pre-Columbian times.
Fontibon	Aprox. 1525	During the colonial period it was somewhere between Bogotá and the Magdalena River.
Suba	1550	The Spanish settlement was founded by Antonio Cardoso.
Usaquen	1539	This was a "*Muisca*" town in pre-Hispanic times and was founded in 1539 as an Indian village.
Usme	1650	The centre of a rural area dedicated to agriculture, provides an important part of the food for the capital.

Table A4 – Exclusion restriction: Distance between the census tract of residence and the nearest original settlement

	Total		**Male**		**Female**	
	M. effect	z-Statistic	M. effect	z-Statistic	M. effect	z-Statistic
Job Accessibility	0.0114	4.22***	0.0068	2.02**	0.0163	3.81***
Dist. original settlement	**-0.0017**	**-1.90***	**-0.0016**	**-1.46**	**-0.0017**	**-1.27**
Gender (male=1)	0.0051	1.55				
Married	0.0229	6.68***	0.0250	4.58***	0.0129	2.30**
Head of household	0.0516	14.41***	0.0649	10.99***	0.0324	5.27***
Years of education	0.0037	8.93***	0.0020	3.93***	0.0061	9.32***
Potential experience	0.0042	12.27***	0.0032	7.42***	0.0046	8.18***
(Potential experience)*2	-0.00005	-7.58***	-0.00005	-6.94***	-0.00002	-1.64
Children under 9	0.0027	1.43	0.0055	2.16**	-0.0005	-0.17
Children between 9 and 18	-0.0049	-2.80***	-0.0043	-1.94*	-0.0062	-2.20**
Year (2009=1; 2008=0)	-0.0264	-8.52***	-0.0269	-6.86***	-0.0251	-5.16***
Observations	35,490		18,593		16,897	

The marginal effect is calculated at the mean of the independent variable.
Statistically significant: 1% ***; 5% **; 10% *.

Table A5 - Estimated marginal effects from IV: Distance between the census tract of residence and the nearest original settlement

	Total		Male		Female	
	M. effect	z-Statistic	M. effect	z-Statistic	M. effect	z-Statistic
Job Accessibility	0.0313	3.05***	0.0269	1.99**	0.0365	2.34**
Gender (male=1)	0.0053	1.61				
Married	0.0239	6.86***	0.0263	4.71***	0.0138	2.44**
Head of household	0.0519	14.43***	0.0648	10.94***	0.0328	5.33***
Years of education	0.0029	4.98***	0.0013	1.68*	0.0053	5.86***
Potential experience	0.0041	11.85***	0.0031	7.22***	0.0045	7.80***
(Potential experience)*2	-0.00005	-7.60***	-0.00005	-6.98***	-0.00002	-1.62
Children less than 9	0.0035	1.79*	0.0063	2.36**	0.0013	-0.46
Children between 9 and 18	-0.0043	-2.36**	-0.0037	-1.61	-0.0055	-1.90*
Year (2009=1; 2008=0)	-0.0233	-6.87***	-0.0239	-5.57***	-0.0220	-4.10***
Observations	35,490		18,593		16,897	
F-Stat. (weak instrument)	$(-48.51)^2$		$(-34.78)^2$		$(-33.77)^2$	
Wald test of exogeneity	0.0572		0.1446		0.2034	

The marginal effect is calculated at the mean of the independent variable.
Statistically significant: 1% ***; 5% **; 10% *.

Table A6 - First stage: Dependent variable log job accessibility instrumented by distance between the census tract of residence and the nearest original settlement

	Total		Male		Female	
	Coeff.	z-Statistic	Coeff.	z-Statistic	Coeff.	z-Statistic
Dist. original settlement	-0.0842	-48.51***	-0.0823	-34.78***	-0.0862	-33.77***
Gender (male=1)	-0.0101	-1.46				
Married	-0.0452	-6.39***	-0.0556	-4.45***	-0.0457	-4.18***
Head of household	-0.0083	-1.11	0.0118	0.95	-0.0196	-1.57
Years of education	0.0373	45.36***	0.0375	34.30***	0.0370	29.40***
Potential experience	0.0054	7.13***	0.0036	3.45***	0.0069	6.22***
(Potential experience)*2	0.00001	0.50	-0.00001	1.71*	0.00001	0.63
Children under 9	-0.0407	-10.81***	-0.0386	-7.35***	-0.0417	-7.60***
Children between 9 and 18	-0.0333	-9.15***	-0.0315	-6.40***	-0.0350	-6.36***
Year (2009=1; 2008=0)	-0.1540	-24.22***	-0.1534	-17.67***	-0.1541	-16.50***
Observations	35,490		18,593		16,897	
R-squared	0.18		0.18		0.17	

The marginal effect is calculated at the mean of the independent variable.
Statistically significant: 1% ***; 5% **; 10% *.

CHAPTER II

Agglomeration and Informal Firms in Cali-Colombia

Summary - *We estimate the effect of agglomeration on the local share of informal firms which produce legal goods but do not comply with official regulations in Cali, Colombia. This issue is relevant because, similar to other developing countries, the informal sector in Colombia employs more than 50% of the workforce. Our results demonstrate that one standard deviation increase in agglomeration reduces the local share of informal firms by 16%. Results are consistent with the idea that informal firms benefit less from agglomeration because of legal restrictions that block the relationship with formal firms.*

JEL-code – R10, R14

Keywords – agglomeration effects, informality, firm location

1. Introduction

Cities can be studied as market responses to production and income opportunities (Mills, 1967). Accordingly, the size and growth of urban areas can be interpreted as responses to these opportunities.[6] Indeed, there is a strong positive relationship between productivity and economic density for different industries and levels of aggregation (Ciccone and Hall, 1996; Ciccone, 2002; Brulhart and Mathys, 2008; Melo et al., 2009; Combes et al., 2010; Morikawa, 2011).

Theory argues that spatial production externalities and commuting costs are among the main forces that shape the city's internal structure. Externalities include effects that increase firms' production, and therefore workers' income, when the size of the local economy grows.[7] We now have strong evidence that firms' productivity is positively related with the volume of nearby employment (Rosenthal and Strange, 2003; Combes and Gobillon, 2014). Most of the empirical findings concern firms in the formal sector and accordingly, the literature says little about the effect of agglomeration on informal firms' location (Duranton, 2009). However, this effect is crucial for developing countries where informal work is the main option for low-educated workers facing unemployment. Indeed, in several countries, more than 50% of employment is in the informal sector (Maloney, 2004; Perry et al., 2007).

One of the main factors that explain informality is a significant cost related to the formalization. An entrepreneur evaluates costs and benefits and could find that it is efficient to choose informality. Indeed, the literature points out that high taxes and social security contributions are costly regulations that lead entrepreneurs not to set up formal business and hence not to register their firms (de Soto, 2000;

[6] In developing countries however, migration from rural to urban areas has been associated with push rather than pull factors, because the population is expelled from rural areas rather than attracted to urban areas by the prospects of better living standards (Bairoch, 1988; Barrios et al., 2006).

[7] As is well known, the literature focuses on technological spillovers, labour pooling and intermediate input linkages (Marshall, 1890; Ellison et al., 2010), and sharing, matching and learning effects, as in Duranton and Puga (2004).

Maloney, 2004).[8] Nevertheless, there is a threshold in the level of the economic activity when the incentives of formalization become more important.

The informal sector comprises a heterogeneous mixture of self-employed entrepreneurs, small and short-life firms that, although they produce legal goods, do not comply with legal regulations. When firms choose to be informal they do not have access to all markets where property rights are secure and well-defined. When a firm is informal, the owner faces restrictions in the financial system: they cannot make long-time capital investments and cannot use their property as collateral to secure loans (Feige, 1990; de Soto, 2000; Sindzingre, 2006). As a result, informal firms have lower productivity levels, lower fixed assets per worker and less access to government services than formal firms (Cárdenas and Mejia, 2007; Santa Maria and Rozo, 2009; Cardenas, 2009). In addition, they do not comply with labour regulations, may practice smuggling and, frequently, do not carry accounting.[9]

We have evidence that formal and informal firms (of similar size and belonging to the same economic sector) display different locational patterns within an urban area (Moreno-Monroy and García, 2015). Furthermore, there is evidence (for São Paulo) that informality rates decrease on average 15% faster in areas with new transport infrastructure (Moreno-Monroy and Roman, 2015) in line with the idea that informality is a choice based on a cost-benefit calculation of the entrepreneur.

The objective of the present paper is to estimate the causal effect of agglomeration on the probability that a firm – given a location - chooses to be informal. We deal with endogeneity issues by using soil information related to earthquake risk, which reduces the height of buildings and therefore increases the cost of agglomeration. The analysis focuses on Cali, in the west of Colombia, where the informal sector employs 60% of the workforce. Using the registration of economic activities as a criterion to identify formal and informal firms (Schneider,

[8] There are different perspectives to analyse informality. Dualism assumes that informal firms do not have linkages with formal firms. Structuralism assumes that formal and informal firms are intrinsically linked (formal firms aim to reduce their input costs by promoting informal activities). Legalism focuses on the regulatory environment of the relationship between formal and informal firms (Chen, 2006; Perry et al., 2007).

[9] For example, around 57% of informal firms in Cali do not carry accounting (Santa Maria and Rozo, 2009).

2005), we identify informal firms when they are not registered in the Chamber of Commerce, in such a way that the percentage of informal firms is 42%.

We find that the effect of agglomeration is strongly negative. The share of informal firms diminishes by 16% when agglomeration increases by one standard deviation. Results in this paper shed light on how formal firms tend to be localized in high density commercial and industrial areas, while informal firms are localized in low density and peripheral areas where the land for production is cheaper and where they can avoid the control of authorities.

This paper proceeds as follows: Section 2 contains the theoretical framework. Section 3 presents the data and a discussion about the results. Section 4 concludes.

2. Theoretical framework

A linear city model where employment clustering is determined by an agglomeration externality was introduced by Fujita and Ogawa (1982).[10] An improvement was made by Lucas and Rossi-Hansberg (2002) where, in a circular city model, firms and workers compete for land at different locations, and the external agglomeration effects lead firms to outbid residential use for land near production centres. The interaction between agglomeration effects and commuting costs is then the main determinant of urban structure: firms have an incentive to be close to each other to obtain benefits from agglomeration, whereas workers prefer proximity to the workplace to minimize commuting costs. Then, market prices (land rents and wages) give firms and households incentives for making land use decisions.

Workers consume residential land and a good which is produced using labour and land. If productivity increases with employment levels in neighbouring locations, firm production per unit of land at location s, $x(s)$, is expressed as $x(s) = A\, {z_s}^{\gamma} n^{\alpha}$, where z_s represents the agglomeration effect on production at

[10] The Alonso-Mills-Muth model has served as the most important base to analyse urban spatial structures, but assumes employment clustering in a city centre (Alonso, 1964; Mills, 1967; Muth, 1969).

location s, A is a productivity constant, n is the number of employees and $\alpha < 1$.[11] The profit per unit of land at location s is represented as $q(s) = A\, {z_s}^{\gamma} n^{\alpha} - w(s)n$, where $w(s)$ is the wage rate. Firms choose employment n to maximize profits. From the first-order condition we obtain $n = \hat{n}(w, z)$ and $q = \hat{q}(w, z)$. Therefore, given w and z, the business bid-rent is determined.[12] The model implies that land use depends on the difference between bids made by households and firms.

For simplicity, let us assume that informal and formal firms are identical except that $\gamma = 0$ for informal firms, because they are not allowed to have formal business contracts. The model predicts then that these firms will be less productive and smaller. Now let us suppose that firms can choose to be formal or informal at a given cost $c > 0$. Given this assumption, firms will be less likely to choose being informal given higher levels of agglomeration. This happens because informal firms benefit less from positive technological and pecuniary externalities that arise in close spatial proximity due to legal restrictions that block the relationship with formal firms. Formal property allows assets to be identified and linked to other assets in the economy. Then, in order to benefit from external effects, the owner has incentives to formalize the business. As a result, we will find formal firms located in high density areas meanwhile informal firms will be located in low density areas. Accordingly, we aim to estimate the causal effect of agglomeration on the probability of being informal (controlling for the firm size and the economic sector).

3. Empirical results

3.1. Data

We focus on Cali, which is the third city in terms of population in Colombia. It was founded on 25 July 1536 and is located in the west of Colombia, in the Cauca Valley. The metropolitan area of Cali has a population of about 2,200,000 with a density of 21,295 persons per square kilometre (in 2005). The city has two natural limits: the

[11] The agglomeration effect, z_s, is calculated using the employment at neighbouring locations. The function is assumed linear and decays exponentially at a rate δ with the distance from s.

[12] The bid-rent is defined as the rent per unit of land that a firm will be willing to pay.

limit to the east is the Cauca River and the limit to the west is the Western Mountain Range (*cordillera occidental*).[13] The urban area is divided into 338 administrative neighbourhoods with an average of 0.36 km^2.

Our main source of information is the Economic Census carried out by the National Institute of Statistics (*Departamento Administrativo Nacional de Estadística,* DANE). The database contains population data as well as establishment-level information, including employment, economic sector (2 digit-level), geographical location and compliance of legal requirements. From now on, we will refer to establishments as firms.

Information about the compliance of legal requirements enables us to identify informal firms which, although they produce legal goods and services (e.g. bread, shoe repair, groceries shop), do not fully comply with legal regulations.[14] This means that informal firms are known, but not registered in the Chamber of Commerce (*Camara de Comercio de Cali*), which certifies firm ownership. These firms evade taxes, have less rigorous bookkeeping, do not contribute to social security and face restrictions to formal financial credit (Cárdenas and Rozo, 2009).[15] Furthermore, informal firms are not allowed to have commercial or financial relations with formal or public firms. In Cali, there are 22,208 informal firms, about 43% of 51,457 firms.

We measure agglomeration, z_s, as a weighted average of the number of jobs at locations h, with weights that are a decreasing function of distance between s and h (see also Koster and Rouwendal, 2013).[16] To be specific, z_s is defined as follows:

$$z_s = \int J_h e^{-\delta d_{sh}} \, dh, \qquad (1)$$

where J_h denotes the number of jobs at location h, d_{sh} denotes the distance between locations s and h, and $\delta > 0$ is a given decay parameter. The external effect is more localized for higher values of δ, which implies that the value to a firm of locating near other producers is also higher. Panel (a) in Figure 1 shows the agglomeration variable for $\delta = 2$. We can observe that employment density is high

[13] Figure A1 in the Appendix shows the geographical location of the city.

[14] At the moment of collecting information, business owners are informed that if they declare that the business fails to comply with legal regulations, they would not experience any negative legal consequences (the information provided is confidential). Informality is not persecuted, which leads business owners to provide truthful information.

[15] Business informality is closely related to labour informality. Informal workers are characterized by lower levels of education and wages.

[16] Fujita and Ogawa (1982) define the externality effect as a potential of employment.

near the city centre; as we move away from the centre we find areas of residential use. Panel (b) shows that the proportion of informal firms is higher in areas away from the city centre; and panel (c) shows that the east of the city is affected by a liquefaction risk, which is present in areas where the soil is saturated with water and then acts like a liquid when shaken by an earthquake.[17]

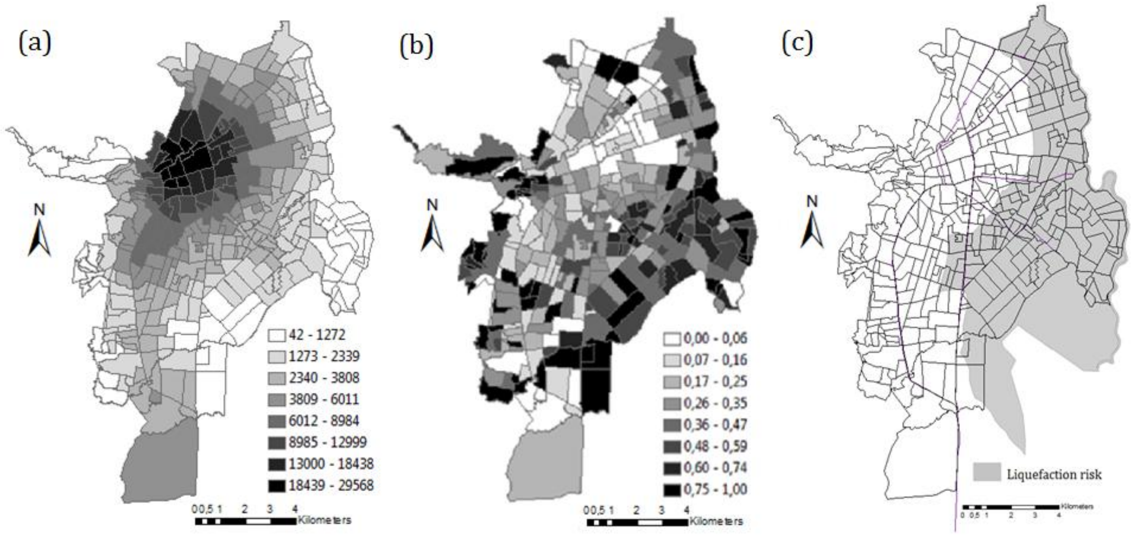


Figure 1- (a) Agglomeration, (b) informality share and (c) liquefaction risk

Table 1- Descriptive statistics

	Mean	Std. Dev.	Minimum	Maximum
Local share of informal firms	0.42	0.31	0	1
ln(Agglomeration, δ =2)	7.81	0.87	4.35	10.30
Firm's size	5	35	1	2415
Jobs	727	1417	0	16087
Formal jobs	551	1191	0	13172
Informal jobs	176	324	0	3319
Job density (per km^2)	3109	5794	0	43435
Population	5972	5763	0	45702
Population Density (per km^2)	21296	13022	0	56716
Neighbourhood Area (km^2)	0.36	0.50	0.02	7.84
Distance from CBD (km)	3.79	2.47	0	16.43
Distance from main corridors (km)	0.56	0.62	0	6.55

Note: 338 Neighbourhoods.

[17] Figure A2 in the Appendix shows the map for job and population density (per km^2) and Figure A3 shows cumulative population and employment (formal and informal) as a function of distance from the city centre in kilometres.

Table 1 presents the descriptive statistics of the main variables by neighbourhood. On average, the share of informal firms in a neighbourhood is 0.42, with a standard deviation of 0.31. On average, a firm has 5 workers and the city has around 727 jobs per neighbourhood, of which 551 are formal jobs and 176 are informal jobs. The area of a neighbourhood is, on average, 360 square meters, the distance between a centroid and the Central Business District is approximately 3.8 *km*, and the distance between a centroid and the main corridors is 560 meters.

3.2. Methodology

We estimate a Probit model where the dependent variable, y_{is}, is equal to 1 for an informal firm i in a neighbourhood s, and equal to 0 otherwise. We are interested in the effect of agglomeration z_s, hence we assume:

$$y_{is} = z_s\rho \; + X_{is}\beta + \varepsilon_{is}, \qquad (2)$$

where X represents firms' characteristics such as industrial sector (2 digit-level ISIC) and the number of workers. Moreover, it includes spatial variables such as distance from the city centre (in discrete categories to capture non-monocentric effects), distance from the main corridors (also in discrete categories), and X and Y coordinates, which are included to control for unobserved factors that smoothly vary over space. Unobserved variables are represented by ε_i.

The agglomeration variable, z_s, depends on the value of δ, which is unknown.[18] We have estimated (2) for different values of δ (from 0.1 to 5). The maximum fit occurs when δ is 2. We report results using this value, which implies that the agglomeration effect disappears within two kilometres (panel (a) in Figure 1 shows the geographical distribution of agglomeration) and Table 1 shows that the logarithm of this variable has a mean of 7.81 with a standard deviation of 0.87 (we use this value to interpret the effect of the variable on the local share of informal firms).[19]

[18] The parameter δ can be estimated with non-linear regression estimation procedures, but this is cumbersome.

[19] Figure A4 in the Appendix shows that the logarithm of agglomeration is approximately normally distributed.

3.3 Main results

The estimated marginal effects using information from 51,454 firms are presented in Table 2 (standard errors are clustered by neighbourhood). We control for distance from CBD, distance from main corridors, distance from the liquefaction limit and geographical coordinates X and Y. In the estimation of column (1) we find that the marginal effect of agglomeration is -0.1406, which means that one standard deviation increase in agglomeration reduces the probability of being informal by 10%.[20] In column (3) we control for size and industrial sector of the firm (in order to control the productivity); in this case the marginal effect of agglomeration is less negative (-0.1026), which means that one standard deviation increase in agglomeration reduces the probability of being informal by 7.6%.[21]

This estimated effect is likely to be biased because of the presence of omitted variables which are correlated with agglomeration and the probability informality (Ellison and Glaeser, 1999; Bayer and Timmins, 2007). Omitted variables may be related to the educational level of employees and fixed assets of the firm. Moreover, there is a reverse causation problem because informal firms have incentives to choose locations where the agglomeration level is lower because the land is cheaper and they try to avoid the control of authorities.

Using an instrumental variable that is correlated with agglomeration but uncorrelated with any unobserved locational advantage may correct the bias. Geological variables such as soil composition, rock depth, water capacity, soil erodibility and seismic and landslide hazard have been used for coping with endogeneity (Rosenthal and Strange, 2008; Combes et al., 2010; Combes and Gobillon, 2015). Characteristics of soil were important to localize original settlements, and agglomeration processes have then developed in those areas. In that case, the instrument is relevant. Our instrument is based on liquefaction risk, which refers to the strength and stiffness of the soil when it is affected in the case of earthquakes. Liquefaction risk is present in areas where the soil is saturated with water and then acts like a liquid when shaken by an earthquake. Earthquake waves cause water pressure to increase in the sediment, so sand grains lose

[20] -0.1406[ln(3648+4020)-ln(3648)]=-0.10.

[21] -0.1026[ln(3648+4020)-ln(3648)]=-0.076.

contact with each other and the soil loses its ability to support high buildings. It can be argued that the instrument is exogenous because there is no reason to say why liquefaction risk may have a direct effect on the registration of firms in the Chamber of Commerce.

Table 2- Marginal effects on probability of being informal

	(1)	(2)	(3)	(4)
ln(Agglomeration)	-0.1406	-0.3111	-0.1026	-0.2207
	(0.0247)***	(0.0691)***	(0.0200)***	(0.0724)***
Size of the firm	no	no	yes	yes
Industrial sectors (47)	no	no	yes	yes
Distance from CBD	yes	yes	yes	yes
Distance from main corridors	yes	yes	yes	yes
Distance from liquefaction limit	yes	yes	yes	yes
Geographical coordinates	yes	yes	yes	yes
Number of observations	51,454	51,454	51,454	51,454
F-Statistic (weak instrument)		33.18		34.22

Note: Marginal effects at the means of the independent variables. Clustered standard errors by neighbourhood in parenthesis.*** p<0.01, ** p<0.05.

Panel (c) in Figure 1 shows areas with liquefaction risk in Cali. This risk is present in the east of the city because of the Cauca River.[22] For most neighbourhoods the risk is either zero or one. When a proportion of the neighbourhood is affected we use the share of the area where the risk is present. Clearly, the instrument does not vary randomly over space. We therefore emphasize that we control for X and Y coordinates.[23] As shown in the estimation of column (2), without controlling for firm size and industrial sector, we find that one standard deviation increase in agglomeration reduces the probability of being informal by 23%.[24] In column (4) we control for firm size and industrial sector and the reduction of the probability of being informal is 16%.[25] We confirm that firm size is negatively related to the probability of being informal, as mentioned in the literature. In our results we obtain that the IV estimated coefficient is more negative than the OLS estimated coefficient, and the reason for this is that there is

[22] It is one of the main rivers of Colombia measuring some 1350km in length.
[23] Table A1 in the Appendix shows the results of the first-stage estimation. The F-statistic for weak instruments is higher than 10, which implies that the instrument has a strong negative effect on agglomeration.
[24] -0.3111[ln(3648+4020)-ln(3648)]= -0.23.
[25] -0.2207[ln(3648+4020)-ln(3648)]= -0.16.

a negative correlation between the instrumented variable and the instrument that we are using. At the end of the Appendix we formally explain this subject.

Table A2 in the Appendix shows the estimated marginal effects of the regressions using only agglomeration of formal jobs, and we find that the results are similar (14.5%). These results allow us to conclude that the urban structure is determined by formal agglomeration and informal firms will occupy spaces that formal firms do not occupy. This means that informal firms have to make decisions in a different set of constraints, including those that link them to the formal sector, and supports the hypothesis that informal firms face restrictions that do not allow them to benefit from agglomeration externalities. These firms are marginalized from accessing the same set of external effects or participating in the same economic transactions as their formal counterparts (Moreno-Monroy, 2012). In short, formalization is fundamental in order to reap all the benefits associated with property rights.

3.4. Sensitivity analysis

We re-estimate the effect of agglomeration on the probability of being informal using different specifications. It is important to bear in mind that we have 338 neighbourhoods. There is extreme variation in the level of agglomeration (in log from 4.3 to 10.3). The result may be sensitive to extreme 'outliers'. We have estimations of the model excluding 20 neighbourhoods with the lower and higher agglomeration levels (the variation in log is now from 6.4 to 9.2). The results are similar: when the agglomeration variable increases by one standard deviation the probability of being informal diminishes by 10% in the Probit model and 18% in the IV specification.

The validity of our instrument can be questioned as it is non-random over space. We therefore re-estimate the model for observations within 1*km* and 500 meters of the liquefaction limit. The results are shown in Table 3 respectively. It should be noted that we also control for X and Y coordinates because the liquefaction limit is north-south. The marginal effects from IV are -0.30 and -0.28, which means that the probability of being informal diminishes by 19% when the agglomeration variable increases by one standard deviation. It is important to note that using the information of neighbourhoods within 500 meters of the limit the marginal effect is not significant. In this case, the result of the Probit model (-0.1841) is more efficient. This means that the probability of being informal diminishes by 13% when agglomeration increases by one standard deviation. Our results point out that the marginal effect of agglomeration is robust.

Table 3-Marginal effects controlling the distance from the liquefaction limit

	1 km from liquefaction limit		0.5 km from liquefaction limit	
	Probit	IV	Probit	IV
ln(Agglomeration)	-0.1358	-0.3089	-0.1841	-0.2777
	(0.0344)***	(0.0975)***	(0.0551)***	(0.1490)*
Number of observations	12,850	12,850	6,898	6,898
F-Statistic (weak instrument)		38.94		19.09

Note: Marginal effects at the means of the independent variables. Clustered standard errors by neighbourhood in parenthesis. *** p<0.01; ** p<0.05; * p<0.10. We use the same control variables as in previous estimations.

Table 4 presents the estimations for three economic sectors: industry, commerce and services. We find that the effect of formal agglomeration on the share of local informal firms is negative in all cases and is less negative in the service sector. This is because relations between formal and informal firms could be more likely in this sector. Informal firms offer distribution services and sell inputs through nonregulated transactions or sub-contracting arrangements (Chen, 2006), and the control of authorities is more difficult than in industry or commerce.

Table 4-Marginal effects by economic sector: Industry, Commerce and Services

		All neighbourhoods	Within 1*Km*
Industry	ln(formal agglomeration)	-0.1907	-0.1690
		(0.1165)***	(0.1737)
	Number of observations	4,753	1,353
Commerce	ln(formal agglomeration)	-0.2540	-0.4197
		(0.1010)**	(0.1855)**
	Number of observations	26,163	6,602
Services	ln(formal agglomeration)	-0.1454	-0.1506
		(0.0802)*	(0.0665)**
	Number of observations	13,047	11,708

Note: Marginal effects at the means of the independent variables. Clustered standard errors by neighbourhood in parenthesis. *** p<0.01;** p<0.05; * p<0.10. We use the same set of control variables as in previous estimations.

4. Conclusion

Literature provides evidence about the relationship between spatial density and aggregate increasing returns. Moreover, the structure of a city is determined by a production externality under which employment at any site is more productive the higher the employment at neighbouring sites. As profit increases with productivity, firms should locate where their expected profit is the highest. Nevertheless, when the property-rights system fails, the market cannot work efficiently, because the owner has de facto rights to their property but does not have a legal enforceable title.

The literature does not say too much about the structure of cities where the percentage of the informal sector is significant. In particular, we analyse the case of a Colombian city where 42% of firms are informal. We shed light on how formal firms tend to be localized in high density commercial areas, while informal firms are localized in low density and peripheral areas where the land is cheaper and where they can avoid the control of authorities.

We aim to estimate the effect of formal agglomeration on the probability of being informal, which could be interpreted as a local share of informal firms. The main result is that when agglomeration increases by one standard deviation the local share of informal firms diminishes by 16%. This may happen because informal firms have less opportunity to benefit from agglomeration effects because of legal restrictions that block their relationship with formal firms. The result explains why formal and informal firms display different locational patterns in the urban structure. We conduct an IV analysis in order to tackle the potential endogeneity problem.

References

Alonso, William (1964). *Location and Land Use: Towards a General Theory of Land Rent*. Harvard University Press. Cambridge. MA.

Bairoch, Paul (1988). *Cities and Economic Development: From the Dawn of History to the Present*. Chicago: The University of Chicago Press.

Bayer, Patrick and Timmins, Christopher (2007). "Estimating Equilibrium Models of Sorting Across Locations". *Economic Journal*. Vol. 117. Pp. 353-374.

Brulhart, Marius and Mathys, Nicole (2008). "Sectoral Agglomeration Economies in a Panel of European Regions". *Regional Science and Urban Economics*. 38. Pp. 348-362.

Barrios, Bertinelli and Strobl (2006). "Climate Change and Rural-Urban Migration: The Case of Sub-Saharan Africa". *Journal of Urban Economics*, 60, Pp. 357-371.

Cárdenas, Mauricio and Mejía, Carolina (2007). "Informalidad en Colombia: Nueva Evidencia". *Fedesarrollo. Working Papers Series*. No. 35.

Chen, Martha (2006). "Rethinking the Informal Economy: Linkages with the formal economy and the formal Regulatory Environment". In Basudeb, Ravi and Ostrom (ed.) *Linking the Formal and Informal Economy*. Oxford University Press.

Ciccone, Antonio (2002). "Agglomeration Effects in Europe". *European Economic Review*. Vol. 46. Pp. 213-227.

Ciccone, Antonio and Hall, Robert (1996). "Productivity and the Density of Economic Activity". *American Economic Review*. 86 (1). Pp. 54-70.

Combes; Duranton; Gobillon and Roux (2010). "Estimating Agglomeration Effects with History, Geology and Worker Fixed Effects". In Glaeser, Edward (ed.) *Agglomeration Economies*. Chicago. IL: University of Chicago Press.

Combes, Pierre-Philippe and Gobillon, Lauren (2015). "The Empirics of Agglomeration Economies". In Duranton, Henderson and Strange (editors). *Handbook of Urban and Regional Economics*. Vol. 5. Elsevier (2015).

Duranton, Gilles (2009). "Are Cities Engines of Growth and Prosperity for Developing Countries?". In Spence, Annez and Buckely (editors). *Urbanization and Growth*, The Commission on Growth and Development, Washington, D.C.

Duranton, Gilles and Puga, Diego (2004). “Micro-foundations of Urban Agglomeration Economies”. In Henderson and Thisse (editors). *Handbook of Urban and Regional Economics*, Vol. 4.

Ellison, Glenn and Glaeser, Edward (1999). “The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?”. *American Economic Review*, 89, Pp. 311-316.

Ellison, Glaeser and Kerr (2010). “What Causes Industry Agglomeration? Evidence from CoagglomerationPatterns”. *American Economic Review*, 100, Pp. 1195-1213.

Feige, Edgar (1990). “Defining and Estimating Underground and Informal Economies: The New Institutional Economics Approach”. *World Development*. 18. Pp. 989-1002.

Fujita, Masahisa and Ogawa, Hideaki (1982).“Multiple Equilibria and Structural Transition of Non-Monocentric Urban Configurations”. *Regional Science and Urban Economics*. Vol.12: Issue 2. Pp. 161-196.

Koster, Hans and Rouwendal, Jan (2013). “Agglomeration, Commuting Costs, and the Internal Structure of Cities”. *Regional Science and Urban Economics*. 43. Pp. 352-366.

Lucas, Robert and Rossi-Hansberg, Esteban (2002). “On the Internal Structure of Cities”. *Econometrica*. Vol. 70. No. 4. Pp. 1445-1476.

Maloney, William (2004). “Informality Revisited”. *World Development*, 32 (7). Pp. 1159-1178.

Marshall, Alfred (1890). *Principles of Economics*. Macmillan. London. 1890.

Melo; Graham and Noland (2009). “A Meta-Analysis of Estimates of Urban Agglomeration Economies”. *Regional Science and Urban Economics*. 39. Pp. 332-342.

Mills, Edwin (1967). “An Aggregative Model of Resource Allocation in a Metropolitan Area”. *The American Economic Review*. Vol. 57. No. 2. Pp. 197-210.

Moreno-Monroy, Ana (2012). “Critical Commentary. Informality in Space: Understanding Agglomeration Economies during Economic Development”. *Urban Studies*. 49(10). Pp. 2019-2030.

Moreno-Monroy, Ana and García, Gustavo (2015). "Intra-metropolitan Agglomeration of Formal and Informal Manufacturing Activity: Evidence from Cali, Colombia". *Working Papers*, Universitat Rovira i Virgili, Department of Economics.

Moreno-Monroy Ana and Roman, Frederico (2015). "The Impact of Public Transport Expansions on Informality: the Case of the São Paulo Metropolitan Region". *ERSA conference papers*, European Regional Science Association.

Morikawa, Masayuki (2011). "Economies of Density and Productivity in Service Industries: An Analysis of Personal Service Industries Based on Establishment-Level Data". *Review of Economics and Statistics*. Vol. 93. Pp. 179-192.

Muth, Richard (1969). *Cities and Housing*. University of Chicago Press. Chicago.

Perry, Maloney, Arias, Fajnzylber, Mason and Saavedra-Chaduvi (2007). *Informality: Exit and Exclusion*. World Bank. Latin American and Caribbean Studies. Washington DC.

Posada, Estevan and Mejia, Daniel (2007). "Informalidad: Teoría e Implicaciones de Política". *Borradores de Economía*. No. 455. Banco de la República, Bogotá – Colombia.

Rosenthal, Stuart and Strange, William (2003). "Geography. Industrial Organization and Agglomeration". *The Review of Economics and Statistics*. May 2003, 85 (2), Pp. 377-393.

Rosenthal, Stuart and Strange, William (2008). "The Attenuation of Human Capital Spillovers". *Journal of Urban Economics*, 64 (2), Pp. 373-389.

Santa María, Mauricio and Rozo, Sandra (2009). "Qualitative and Quantitative Analysis of Firm Informality in Colombia". *Desarrollo y Sociedad*. Pp. 269-296.

Schneider, (2005). "Shadow Economies Around the World: What do we Really Know?". *European Journal of Political Economy*, Vol. 21 (3), Pp. 598-642.

Sindzingre, Alice (2006). "The Relevance of the Concepts of Formality and Informality: a Theoretical Appraisal". In Basudeb, Ravi and Ostrom (editors). *Linking the Formal and Informal Economy*. Oxford University Press.

Soto, Hernando de (2000). *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*. New York: Basic Books.

Appendix

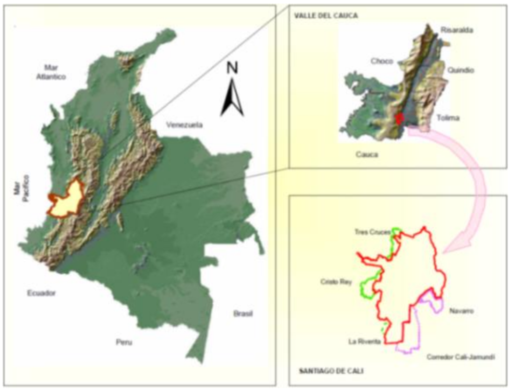


Figure A1- Study area: Cali, Colombia
Source: Colombian geological survey.

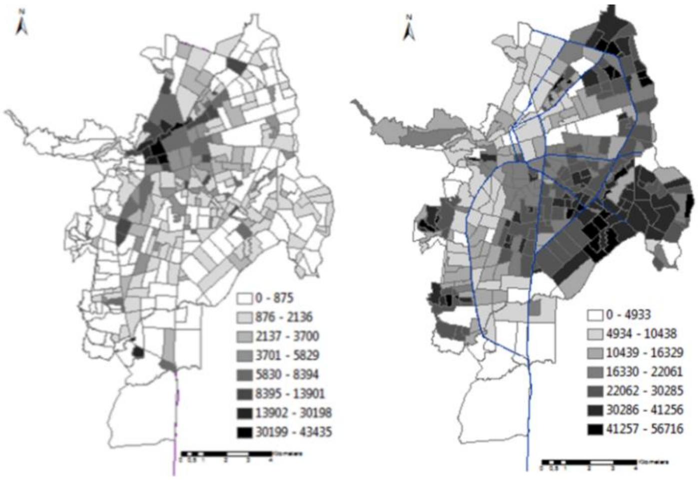


Figure A2-Jobs and population density per km^2 by neighbourhood
Source: Census 2005, DANE.

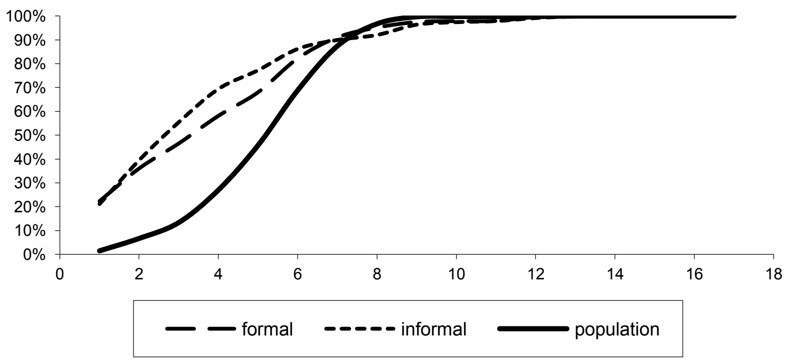


Figure A3- Cumulative employment and population

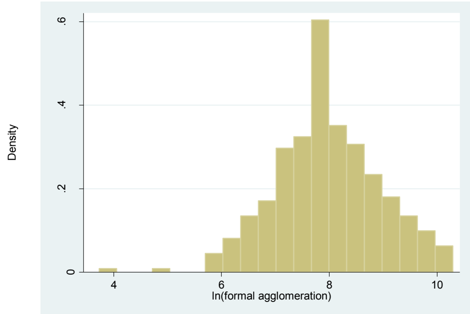


Figure A4- Distribution of log agglomeration

Table A1- First stage estimation

	Coefficient
Liquefaction	-0.4565
	(0.078)***
Size of the firm	yes
Industrial sectors (47)	yes
Distance form CBD	yes
Distance from main corridors	yes
Distance from liquefaction limit	yes
Geographical coordinates	yes
Adj. R-Squared	0.91
F-Statistic (weak instrument)	34.22
Number of observations	51,454

Note: Marginal effects at the means of the independent variables. Standard errors clustered by neighbourhood in parenthesis. *** p<0.01.

Table A2-Marginal effects on probability of being informal (formal agglomeration)

	(1)	(2)	(3)	(4)
ln(formal agglomeration)	-0.1666	-0.2805	-0.1217	-0.1997
	(0.0250)***	(0.0537)***	(0.0213)***	(0.0889)***
Size of the firm	no	no	yes	yes
Industrial sectors (47)	no	no	yes	yes
Distance from CBD	yes	yes	yes	yes
Distance from main corridors	yes	yes	yes	yes
Distance from liquefaction limit	yes	yes	yes	yes
Geographical coordinates	yes	yes	yes	yes
Number of observations	51,454	51,454	51,454	51,454
F-Statistic (weak instrument)		85.00		85.19

Note: Marginal effects at the means of the independent variables. Clustered standard errors by neighbourhood in parenthesis. *** p<0.01.

When formal agglomeration increases by one standard deviation the probability of being informal diminishes by 8.8%, according to the Probit model and 14.5% according to the IV Probit model.

Table A3- Restriction condition test

ln(Agglomeration)	-0.1026
	(0.0200)***
Liquefaction risk	-0.0586
	(0.0355)
Size of the firm	yes
Industrial sectors (47)	yes
Distance from CBD	yes
Distance from main corridors	yes
Distance from liquefaction limit	yes
Geographical coordinates	yes
Number of observations	51,454

Note: Marginal effects at the means of the independent variables. Clustered standard errors by neighbourhood in parenthesis.*** p<0.01, ** p<0.05.

A note about the IV estimation

We assume a simplified model:

$$y = \beta x + u$$

$$x = \phi z + \eta$$

Where y represents the probability of informality, x represents agglomeration and z represents the liquefaction risk. From the OLS estimation we obtain:

$$\hat{\beta} \to \beta + \frac{\sigma_{xu}}{\sigma_x^2}$$

We know that $\sigma_{xz} = \phi\sigma_z^2$ and $\phi = \frac{\sigma_{xu}}{\sigma_z^2}$. Consequently, ϕ will be negative if $\sigma_{xu} < 0$, and will be positive if $\sigma_{xu} > 0$. In this case it is verified that:

$\hat{\beta}_{OLS} \to \beta + \frac{\sigma_{xu}}{\sigma_x^2}$

$\tilde{\beta}_{IV} \to \beta$, assuming that z is a valid instrument.

Hence it follows that:

$\hat{\beta}_{OLS} - \tilde{\beta}_{IV} \to \frac{\sigma_{xu}}{\sigma_x^2} = \phi \frac{\sigma_z^2}{\sigma_x^2}$

In our case, we expect a negative correlation between agglomeration and liquefaction risk, which means that $\phi < 0$. As a result, we have that $\hat{\beta}_{OLS} - \tilde{\beta}_{IV}$ is a negative value.

CHAPTER III

Homicide Rates and Housing Prices in Cali, Colombia

Summary - *Latin America dominates the list of most violent cities in the world. In 2015, Cali (Colombia) registered 65 homicides per 100,000 people in a ranking topped by Caracas (Venezuela) with 120. However, the crime rates are not homogeneously distributed within an urban area and the literature points out that the local response to crime will affect the housing market. The objective of the paper is to estimate the relationship between housing prices and homicide rates in Cali. We found that a 10% increase in the homicide rate is related to a decrease of between 2% and 2.5% in housing prices.*

JEL-code – R30, R31

Keywords – Homicide rates, housing prices, crime.

1. Introduction

The housing market deals with characteristics such as heterogeneity, durability and spatial fixity. A house is a heterogeneous commodity (e.g. design, age, etc.) and it is difficult to define a unit measure of analysis; indeed, price in the transaction is the product of the price and the quantity of the dwelling. The second characteristic is durability, because a particular homeowner acts both as an investor and as a consumer. Therefore, expectations about the future are important, as well as wealth and income. The third characteristic means that a dwelling is located in a neighbourhood and both the characteristics of the neighbourhood as well as the accessibility to other neighbourhoods are fundamental determinants. Consequently, externalities such as level of noise, police protection, crime, etc., are very important factors in the housing market since it will take a long time to adjust (Fallis, 1985).

High crime rates represent a significant welfare loss, a reduction of the expected lifespan and an increase in uncertainty about the future (Soares, 2010). Furthermore, the quantity of money allocated to maintaining justice and prison systems is relevant. According to CCSP-JP (2016),[26] Latin America dominated the list of most violent cities in the world. Figure 1 shows the 50 cities with the highest homicide rates and populations higher than 300,000 people. The horizontal axis represents the homicide rate per 100,000 people, while the size of the circle represents the number of homicides. The ranking is led by Caracas (Venezuela), San Pedro Sula (Honduras) and San Salvador (El Salvador). Two Colombian cities appear in this shocking top ten: Palmira and Cali. The former is an intermediate city of 1.5 million inhabitants, and Cali is the third city in terms of population in Colombia (2.3 million after Bogotá and Medellín).

There is an association between city size and crime (Glaeser and Sacerdote, 1999). Furthermore, crime rates are not homogeneously distributed within an urban area and this characteristic shows a clear association with the quality of the

[26] Council for Public Security and Criminal Justice (CCSP-JP by its Spanish acronym).

neighbourhood. In response to crime risk, residents generally have two options: they can vote for anti-crime policies or vote with their feet. When individuals exercise the latter option, local response to crime will be observed in the housing market (Gibbons, 2004; Buonanno, et al., 2012). Indeed, the fear of crime through its indirect effects on housing prices may also hinder local regeneration and cause a downward spiral in the quality of the neighbourhood (Gibbons, 2004).

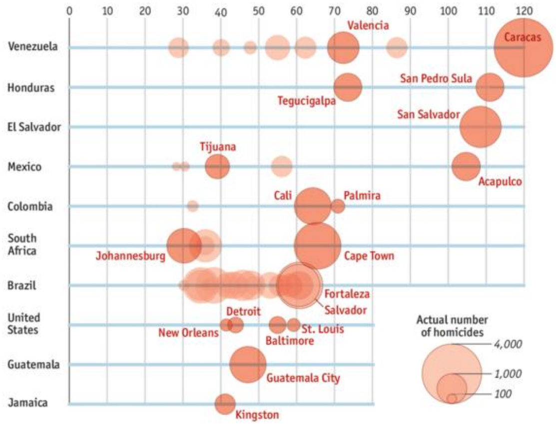


Figure 1-Homicides per 100,000 people

Source: The Economist (2016) with information from 2015.

There is evidence of a relationship between urban crime and housing prices for European and North American cities. Soares (2010) points out that death due to violence is 200 percent more common in Latin America than in North America and 450 percent more common than in Western Europe. The objective of this paper is to evaluate the relationship between homicide rates and housing prices in Cali. We use a cross section of cadastral housing prices of 2012 and the average homicide rates of the period 2000-2010 at a neighbourhood level.

The analysis is performed using two estimation strategies. In the first, we estimate a model where the dependent variable is housing prices and the main explanation variable is the homicide rate: we found that a 10% increase in the homicide rate is related to a 2.4% decrease in housing prices. In the second strategy, we estimate a two-stage model: in one stage we estimate the hedonic price at a neighbourhood level and in the second stage we estimate a regression to test if the homicide rate has a negative relationship with the estimated hedonic price. We found that, on average, a 10% increase in the homicide rate is related to a decrease of between 2% and 2.5% in housing prices.

Additionally, we discuss a methodological issue: when the statistical information is clustered (e.g. classrooms, neighbourhoods, economic sectors, etc.) the literature recommends estimating models using cluster standard errors. Nevertheless, there are some potential harmful consequences from the formation of false clusters in the estimated standard errors. In this paper we present a simulation exercise to show the magnitude and the direction of the bias in the standard errors estimation.

The remainder of the paper is organized as follows: section 2 relates the paper to the existing literature, section 3 describes data and provides the estimation results, and section 4 concludes.

2. Literature Review

Fear of crime has a powerful influence on perceptions of area deprivation and may discourage home-buyers, inhibit local regeneration and catalyse a downward spiral in neighbourhood status (Gibbons, 2004). The literature has tried to measure the effect of crime on housing prices using two main methodologies:

- The contingent valuation tries to estimate the value of goods that are not transacted in a market. The strategy is to ask how much people would be willing to pay for it. For example, in Cohen et al. (2004), respondents were asked if they would be willing to vote for a proposal requiring each household in their community to pay a certain amount to be used to prevent one in ten crimes in their community. Meanwhile, in Atkinson et al. (2005), respondents were told the characteristics of a type of crime and the current risk of victimization, and then asked to express their willingness to pay to reduce the chance of being a victim of this offence by 50 percent over the next 12 months.
- In the hedonic models (Rosen, 1974) a house may derive its value from the quality of its physical characteristics (e.g. living space, number of bedrooms, garage, amenities), and also from its location. Furthermore, the level of crime and violence in the surrounding area may be additional attributes of the property, and individuals may be willing to pay more to live in an area with lower levels of crime. Then, an estimate of how much

the attribute *lower-level-of-crime* is worth in the housing price provides an estimation of the cost of crime.

Asking individuals how they would react in a certain hypothetical situation is not the same as how they will react in a real decision-making situation. Consequently, economists have long been sceptical of information extracted from stated preferences, rather than revealed ones (Carson et al., 2001; Levitt and List, 2007). This is why hedonic price models have been most used to estimate the relationship between crime and housing prices. Hedonic models rely on the preferences revealed by market behaviour by analysing the actual amount that people pay to avoid living in high crime areas. Some literature that uses this methodology found a negative and significant relationship between crime and housing prices: For Rochester, New York, one standard deviation increase in the crime rate caused a 3% reduction in house prices (Thaler, 1976). The measures of crime used by the author are: total offences, property crimes, crimes against persons, and property crimes committed in or around homes. Hellman and Naroff (1979) obtain an elasticity of property value with respect to crime equal to -0.63 for Boston. For Jacksonville, Florida, Lynch and Rasmussen (2001) found an elasticity of -0.05 for violent crimes. Meanwhile, Atlanta, Bowes and Ihlanfeldt (2001) argue that crime may be higher in train station areas; moreover, train stations may be the source of more crime in higher than in lower-income neighbourhoods. Authors found that an additional crime per-acre per-year decreases housing prices by around 3%. Besley and Mueller (2012) present evidence that supports the assumption that housing prices depend on the level and persistence of historical crime rates. Authors argue that houses are assets whose prices reflect the present and future expected attractiveness of living in an area. They use information for 11 regions of Northern Ireland to evaluate the increased housing prices in response to a reduction in murders. Estimating a Markov switching model, the authors predict that peace in Northern Ireland leads to an increase in housing prices of between 1.3 percent and 3.5 percent (the result is heterogeneous across regions).

Table 1-Literature review

Author	Place and time	Instruments for crime	Results
Studies that do not instrument for crime			
Thaler (1978)	Rochester, New York (1971).		One standard deviation increase in the crime rate caused a 3% reduction in house prices.
Hellman and Naroff (1979)	Boston (1976).		They report a negative elasticity of property value with respect to total crime (-0.63).
Lynch and Rasmussen (2001)	Jacksonville, Florida.		They found an elasticity of -0.05 for violent crimes.
Bowes and Ihlanfeldt (2001)	Atlanta (1991-1994).		An additional crime per acre per year in a given census tract has the effect of reducing house prices by around 3%.
Shapiro and Hassett (2012)	Seattle, Milwaukee, Huston, Dallas, Boston, Philadelphia, Chicago and Jacksonville.		A 10% reduction in homicides would lead to a 0.83% increase in housing values the following year.
Besley and Mueller (2011)	11 regions of Northern Ireland (1984-2009).		Property prices depend on the level and persistence of historical crime rates.
Frischtak and Mandel (2012)	Rio de Janeiro.		Homicides dropped by between 10% and 25% and robberies by between 10% and 20%, while the selling price of the properties increased by between 5% and 10%, and was proportionally higher in low-income neighbourhoods.
Studies that do instrument for crime			
		Instruments for crime	
Rizzo (1979)	Chicago (1970).	Proportion of population between ages 15 and 24; median years of schooling; unemployment rate; population density; proportion of population receiving welfare; ratio of males to females; and the labour force participation rate.	The estimated elasticity of crime with respect to prices is -0.23.
Gibbons (2004)	London (2001).	Crimes on non-residential properties; Spatial lags of the crime density; Distance to the nearest alcohol licensed premises.	A one-tenth standard deviation increase in the recorded density of incidents of criminal damage has a capitalized cost of just under 1% of property values
Tita, Petras and Greenbaum (2006)	Columbus, Ohio (1995-1998).	Homicide rate.	Negative significant relationship between prices and violent crimes.
Ceccato and Wilhelmsson (2011)	Stockholm (2008).	Murders as an instrument for crime.	If total crime increases by 1%, apartment prices are expected to fall by 0.04%.
Buonanno et al. (2013)	Barcelona (2004-2006).	Victimization rate 20 years ago; Share of youth aged between 15 and 24.	One standard deviation increase in perceived security is associated with a 0.57% increase in the valuation of districts.

Previous literature assumes crime as an exogenous regressor. Relaxing that assumption, Rizzo (1979) found similar qualitative results for Chicago and points out that in the housing market people reveal the cost of crime as they themselves perceive it. Gibbons (2004) estimates the impact of recorded domestic property crime on property prices in the London area. This author has considered information for five types of crime: burglary in a dwelling, burglary in other buildings, criminal damage to a dwelling, criminal damage to other buildings, and theft from shops. As a result, a one-tenth standard deviation increase in the recorded density of incidents of criminal damage has a capitalized cost of just under 1% of property values. This means that incoming residents perceive criminal damage as a deterioration of the neighbourhood. The difference with previous papers is that the author pays attention to identification issues and deals with the endogeneity problem using instrumental variables. Ceccato and Wilhelmsson (2011) analyse the relationship between apartment prices and different measures of crime in Stockholm. The authors found that when total crime increases by 1%, apartment prices are expected to fall by 0.04%. Buonanno et al. (2013), for Barcelona, found that one standard deviation increase in perceived security is associated with a 0.57% increase in the valuation of districts (the authors deal with endogeneity using instrumental variables).

Although comparing results is somewhat arbitrary because of the differences in the types of crimes, there is evidence of a negative relationship between crime and housing prices. This means that high crime rates deter new residents and motivate those who can to move out to lower-crime rate neighbourhoods (Gibbons, 2004).

3. Data and Results

Cadastral information in Colombia is one of the oldest and most extensive in Latin America. Nevertheless, it is limited in order to formulate politics (DAHM, 2012). We use the cadastral housing prices of 2012 in order to estimate the relationship with homicide rates at neighbourhood level. Homicide rate is defined as the number of homicides committed in a year per 100,000 people, excluding homicides committed as a result of armed conflict.

Table 2 summarises the key variables in the housing price and homicide data. The mean log of housing prices is 17.16 with a standard deviation of 1.21 (Figure 2 shows the distribution of this variable); the homicide rate that we use is an average of ten years (2000-2010) at neighbourhood level. This variable has a mean of 104 with a standard deviation equal to 155.[27] The first map on Figure 3 shows the average housing prices per square meter in Colombian Pesos while the second map shows the homicide rates at neighbourhood level. Our hypothesis is that persistent cases of homicide will be capitalized in housing prices. The urban area is divided into 338 neighbourhoods: the average size of a neighbourhood is 360 square meters; the average distance between the Central Business District (CBD) and the centroid of a neighbourhood is 4.76 kilometres; and the average distance from the centroid of a neighbourhood to the closest main road is 0.41 kilometres.

Table 2- Summary statistics

	Mean	Std. Dev.	Minimum	Maximum
ln(Housing price)	17.16	1.21	10.99	25.20
Homicide rate	104	155	0	1628
Area (km^2)	0.36	0.50	0.02	7.84
Distance to CBD (km)	4.76	2.12	0	9.78
Distance to main roads (km)	0.41	0.59	0	5.23

Note: 338 Neighbourhoods.

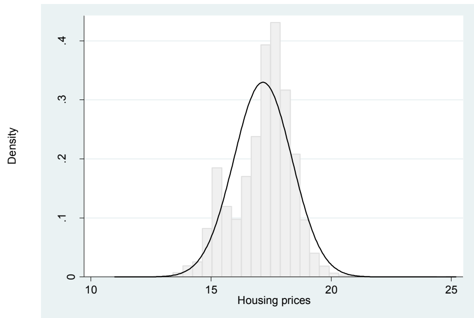


Figure 2- Distribution of log housing prices

[27] Appendix 1 shows information about neighborhoods within *comuna* 6 (the city is divided into 22 *comunas*).

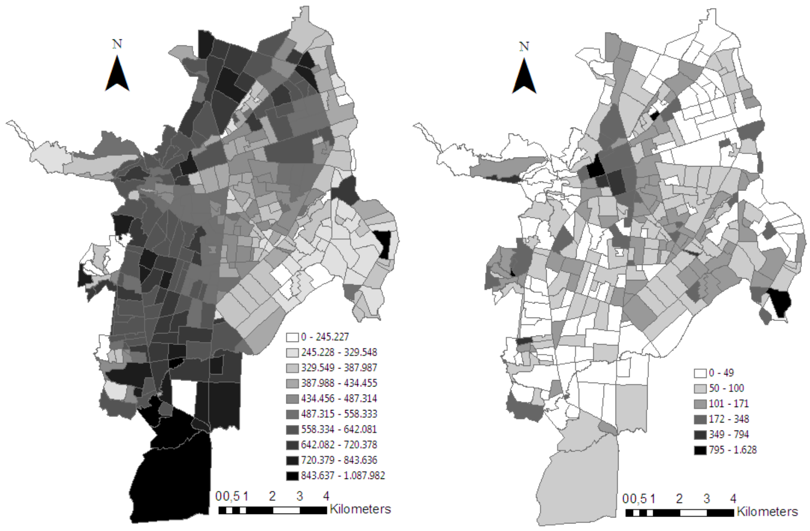


Figure 3- Housing prices per square meter and homicide rates in Cali

Hellman and Naroff (1979) present a model where z represents a composite good (numeraire), Q the quantity of housing attributes per unit of land, $Hr(j)$ the homicide rate in a neighbourhood j, Y represents income, $P(j)$ is the price per unit of housing and $T(j)$ the transportation cost. The problem for the typical household is to maximise $U = U[z, Q, Hr(j)]$ subject to $Y = z + P(j)Q + T(j)$. We assume that the homicide rate enters the utility function only for the household residing at that location and $\partial U/\partial Hr < 0$. Using a Cobb-Douglas utility function $U = z^{\alpha}\, Q^{\beta}\, Hr^{-\gamma}$ the indirect utility function can be obtained, which can be solved for P to get the bid-rent function:

$$P = \left[\left(\frac{\alpha}{\alpha+\beta}\right)^{\alpha} \left(\frac{\beta}{\alpha+\beta}\right)^{\beta} \frac{1}{\bar{U}}\right]^{1/\beta} (Y-T)^{(\alpha+\beta)/\beta} Hr^{-\gamma/\beta} \;=\; C_{*}(Y-T)^{(\alpha+\beta)/\beta} Hr^{-\gamma/\beta} \quad (1),$$

where $\bar{U}$ equals the constant and equal utility level for each household in the city. The bid-rent is positively related to income and negatively related to the cost of transportation. When the homicide rate increases beyond a threshold level $(Hr > 1)$, the bid rent shifts down:[28]

$$\frac{\partial P}{\partial Hr} = \frac{-\gamma}{\beta} C_{*}(Y-T)^{(\alpha+\beta)/\beta} Hr^{(-\gamma/\beta)-1} < 0 \qquad (2),$$

The impact on housing prices in any neighbourhood j is the combined effect of a decrease in bid-rents and a resulting decline in density. The total housing price in j is given by:

[28] A value of one is given when the crime level is at a minimum acceptable level, or when it has no perceptible impact on utility.

$$V(j) = P(j)\, D(j) = B_*(Y - T)^{(2\alpha+\beta-\alpha\delta)/\beta(1-\delta)} Hr^{(-2\gamma-\gamma\delta)/\beta(1-\delta)} \quad (3)$$

Taking the log of both sides of the equation results in an equation where housing price is a linear function of income, distance from the Central Business District (CBD), and the homicide rate. The value of the coefficient of the homicide rate variable represents the elasticity of housing prices with respect to the homicide rate. The model can be estimated as follows:

$$V_{ij} = \mu + \delta Hr_j + \sum_k \beta_k Q_{ik} + u_{ij} \quad (4)$$

where V_{ij} is the logarithm of housing price i in a neighbourhood j; α is the constant term; Hr_j is the homicide rate in the neighbourhood j; Q_{ik}'s is the attribute k of a house i; β_k's are the estimated price associated with each attribute k; and u_{ij} is an error term.

We can estimate the standard errors of the coefficients of this equation by OLS assuming that all observations in the database are unrelated. However, the correct standard error estimation procedure is given by the underlying structure of the data. Indeed, some economic phenomena do not affect observations individually, but they affect groups of observations within each cluster. In our case, it is recommended to use clusters at a neighbourhood level, which means a data structure where unobservable elements of housing prices within a cluster are correlated, while they are uncorrelated across clusters. One cause of this kind of correlation is because some regressors take the same value for all observations within clusters. Indeed, Hr_j is perfectly correlated within neighbourhood (Cameron and Miller, 2015). Table 3 shows the results for different strategies of estimation: OLS model; the heteroskedasticity correction suggested by White (1980) which is known as robust standard errors; cluster-robust standard errors;[29] and a Random Effects (RE) model where j represents clusters and i represents individual housing prices as follows:

$$V_{ij} = \mu + \delta Hr_j + \textstyle\sum_k \beta_k Q_{ik} + (\alpha_j + u_{ij}),$$

In the RE model the estimation of standard errors is valid because it considers correlations of residuals within each cluster. Furthermore, if residuals α_j are

[29] For a regressor k, the variance inflation factor, $\tau_k \approx 1 + \rho_{xk}\rho_u(\bar{N}_j - 1)$, is increasing in: the within-cluster correlation of the regressor, ρ_{xk}; the within-cluster correlation of the error, ρ_u; and the number of observations in each cluster, $\bar{N}_j$ (Cameron and Miller, 2015).

correlated with explanatory variables, asymptotically, coefficients of the RE model tend to coefficients of a fixed effects model. So, the correlation between random effects and explanatory variables is corrected. Therefore, a consistent estimate with N is obtained.[30]

We found a significant negative relationship between housing prices and homicide rate. The elasticity is around -0.24, which means that a 10% increase in the homicide rate is related to a 2.4% decrease in housing prices.[31] The coefficient of area is significantly positive, while the coefficient of squared area is significantly negative (this variable is included in order to control for nonlinear effects). We include a geological variable called liquefaction risk. Geological variables such as soil composition, rock depth, water capacity, soil erodibility and seismic and landslide hazard have been used in the literature about urban structures (Rosenthal and Strange, 2008; Combes et al., 2010; Combes and Gobillon, 2015). Characteristics of soil were important to localize original settlements and agglomeration processes have then developed in those areas. Liquefaction risk refers to strength and hardness of the soil when it is affected in the case of earthquakes. The risk is present in areas where the soil is saturated with water and then acts like a liquid when shaken by an earthquake. Earthquake waves cause water pressure to increase in the sediment, so sand grains lose contact with each other and the soil loses its ability to support high buildings. Nobody would be willing to pay the same price for two equivalent houses when one of them is located in an area where risk is present and the other located in an area without the risk. The coefficient of this variable is significantly negative. We include the distance to main roads and we also found a significantly negative coefficient.

[30] A strong assumption of the RE model is the dependence pattern within each cluster: if $u_{ij} = \alpha_j + \varepsilon_{ij}$ and $u_{hj} = \alpha_j + \varepsilon_{hj}$ then $E(u_{ij}, u_{hj}) = \sigma_\alpha^2$ for all j.

[31] $Elasticity = \frac{dPrice/Price}{dHr/Hr} = \hat{\beta}_{Hr}(Hr)$.

Table 3- Dependent variable: log housing prices

	OLS	Robust	Cluster	RE
Homicide rates	-0.0037***	-0.0037***	-0.0037***	-0.0031***
	(-229.35)	(-188.17)	(-6.77)	(-9.42)
ln(Area in square meters)	1.27***	1.27***	1.27***	1.39***
	(370.37)	(257.79)	(18.95)	(465.55)
ln(Area in square meters)2	-0.02***	-0.02***	-0.02**	-0.03***
	(-50.03)	(-35.78)	(-2.34)	(-79.29)
Liquefaction risk	-0.33***	-0.33***	-0.33***	-0.29***
	(-222.15)	(-211.40)	(-8.36)	(-8.98)
Distance from main roads	-0.0003***	-0.0003***	-0.0003***	-0.0002***
	(-245.23)	(-210.46)	(-8.98)	(-9.52)
Constant	12.82***	12.82***	12.82***	12.32***
	(2001.17)	(1388.66)	(119.70)	(356.56)
R-squared	84%	84%	84%	84%
N	504,617	504,617	504,617	504,617

Note: t-statistics in parenthesis. Significant level: *** 1%; ** 5%; * 10%.

Results reveal that in the presence of correlation within each cluster (neighbourhood) the OLS standard errors can overvalue estimator precision. The literature points out that when the number of clusters is large, statistical inference after OLS should be based on cluster-robust standard errors (Cameron and Miller, 2015). An alternative is to estimate a RE model because it takes into account the within dependency and in this way provides efficient estimates. Furthermore, the RE model is consistent in N even in the case of correlation between the random effects and the explanatory variables.

With the estimated coefficients, the increase of housing prices when the homicide rate diminishes can be simulated. In Figure 4, the horizontal axis measures the homicide rate, while on the vertical axis we have the average increase of a housing prices index in percentage. In this graph, 100 is the observed value of the housing prices for the actual homicide rate of 64.45. As the graph shows, when the homicide rate is “zero” the estimated real estate value reaches 122.64 (22.64% of increase), which can be interpreted as an increase in the welfare of the people.

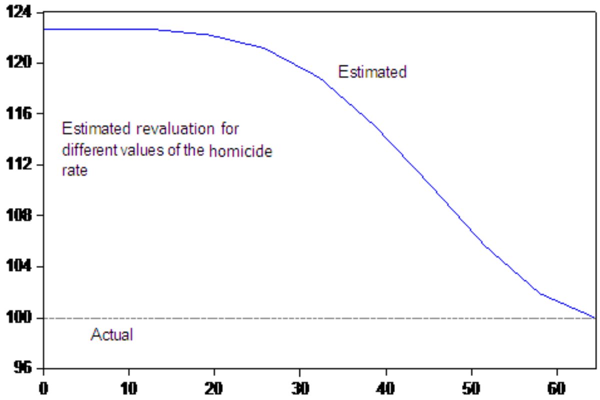


Figure 4- Estimated housing prices for different values of the homicide rate

3.1 An alternative estimation: Hedonic prices

A standard method to estimate the effect of crime on housing prices is a two-stage model. In the first stage, we estimate the shadow price of the location of houses and in the second, we test if crime can explain some of the variability of the estimated locational valuation (Thaler, 1978; Gibbons, 2004; Buonanno, et al., 2012; among others). Table 4 shows the estimation of the hedonic housing price taking into account the size of the property (area in square meters and the square of the area to take into account nonlinear effects) in order to obtain the hedonic housing price at the neighbourhood level (see Figure 5).[32]

Table 4- Hedonic first stage: Dependent variable log housing prices

ln(Area in square meters)	1.40***
	(258.28)
ln(Area in square meters)2	-0.03***
	(-47.03)
Fixed effects by neighbourhood	Yes
Constant	-28.44***
	(-10.69)
R-squared	89%
N	524,400

Note: Robust t-statistics in parenthesis.
Significant level: *** 1%; ** 5%; * 10%

[32] In Appendix 2 we present a simulation procedure to analyse the potential consequences on the estimated standard errors derived from forming false clusters. As is shown in Table 3, it has important implications on the significance level of estimated coefficients.

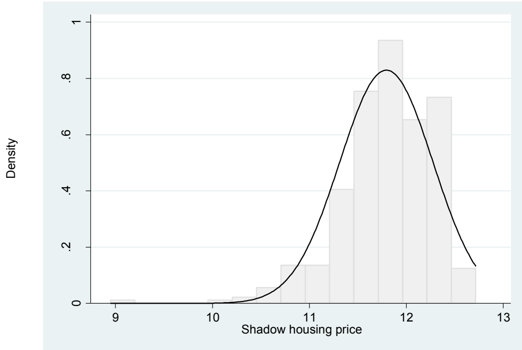


Figure 5-Hedonic housing prices at neighbourhood level

In the second stage we test if the homicide rate explains some of the variability of the hedonic price at neighbourhood level. It is difficult to find exogenous variables in order to explain housing prices. We use liquefaction risk as a control variable. As we argued in the previous section, nobody would be willing to pay the same price for two equivalent residences when one of them is located in an area where risk is present and the other located in an area without the risk, so we expect a negative coefficient.

Table 4 shows the second stage estimation. The estimated regression in Table 5 excludes neighbourhoods with more than 400, 300, 200 and 100 homicides per 100,000 people in order to check if the coefficient is stable across the distribution of the variable. We control for liquefaction risk and the coefficient associated with this variable is negative as expected. Figure 6 shows the distribution of homicide rates at neighbourhood level up to a level of 400; indeed, we have few neighbourhoods with homicide rates of 500, 1000 and 1628.

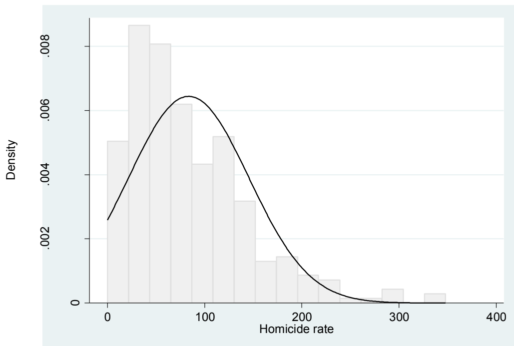


Figure 6-Homicide rates at neighbourhood level

Table 5- Hedonic second stage

	<400	<300	<200	<100
Homicide rate	-0.0024***	-0.0026***	-0.0033***	-0.0044
	(-6.33)	(-6.38)	(-6.78)	(-4.80)
Liquefaction risk	-0.2737***	-0.2661***	-0.2638***	-0.2365
	(-6.69)	(-6.46)	(-6.38)	(-4.79)
Distance from main roads	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(-7.16)	(-7.12)	(-7.35)	(-6.79)
Constant	12.24***	12.25***	12.30***	12.34
	(282.15)	(277.15)	(258.25)	(227.56)
R-squared	28%	29%	29%	23%
N	319	317	302	215

Note: Robust t-statistics in parenthesis. Significant level: *** 1%; ** 5%; * 10%

In Table 6 we report the "price-homicide" elasticity derived from the hedonic second stage estimation:

$$E_h = \frac{dPrice/Price}{dHr/Hr} = \hat{\beta}_{Hr}(Hr),$$

where E_h represents the elasticity for the housing price for a neighbourhood in strata h; $\hat{\beta}_{Hr}$ is the estimated coefficient; and Hr is the homicide rate.

We found that a 10% increase in the homicide rate is related to a 2% decrease in housing prices (excluding neighbourhoods with homicide rates higher than 400). The decrease in housing prices is 2.1%, 2.5% and 2.1% excluding neighbourhoods with homicide rates of more than 300, 200 and 100 respectively. However, as some literature points out, the homicide rate has higher consequences in neighbourhoods of lower socioeconomic strata. For that reason, we differentiate results by socioeconomic strata.

Table 6- Housing prices, homicide rate and elasticity by strata

				E_h			
Strata	Observations	Average price	Hr	<400	<300	<200	<100
1	12.74%	327,914	119	-0.26	-0.28	-0.30	-0.21
2	23.52%	384,455	128	-0.23	-0.25	-0.30	-0.25
3	28.52%	533,101	111	-0.20	-0.21	-0.24	-0.22
4	10.19%	639,195	56	-0.13	-0.14	-0.19	-0.19
5	19.74%	684,311	63	-0.15	-0.16	-0.19	-0.18
6	5.29%	784,783	48	-0.11	-0.12	-0.16	-0.15
All	100%	533,051	104	-0.20	-0.21	-0.25	-0.21

In Colombia, residential areas have been divided into six socio-economic status (strata) since 1988. The stratification system, where 1 is the lowest strata and 6 is the highest, divides the city into areas of wealth and poverty. With this socio-economic stratification system, the public administration guarantees that the upper strata pay a higher rate for services (electricity, water and sewage) to subsidize the cost of services for the lower strata. Then, the stratification makes poor families settle in areas where they can afford to pay housing and basic services. Income-based class division explicitly categorized via public policy is strange, but, in our case, we can use it to differentiate the negative effect of homicide rate on housing prices according to different wealth levels because the stratification classifies neighbourhoods with similar characteristics. The second column in Table 6 shows that around 65% of dwellings in Cali are in strata 1, 2 and 3.

The third column in Table 6 shows the average price of the square meter by strata; column 4 shows the homicide rate per 100,000 people (it is clear that the rate is higher in lower strata); and the last four columns show that the price-homicide rate elasticity is more negative in lower strata. In any case, the fact that the elasticity decreases with strata may result from the definition of elasticity: (percentage change price) / (percentage change homicide rate). This means that if the homicide rate goes from 1 to 2 in a wealthy neighbourhood, the percentage change is 100%. If the homicide rate goes from 100 to 150 in a very poor neighbourhood, the percentage of variation is 50%. Lastly, if, in a very wealthy neighbourhood, the homicide rate goes from 0 to 1, the percentage of variation is infinite, so that elasticity must be zero.

In Colombia, crime is constant across the income distribution. Nevertheless, the rich are most often victims of kidnappings and the poor are most often victims of homicides. For that reason, the rich are more likely to adopt costly protective behaviour, neighbourhood watch programs, install anti-theft devices at home, hire private security personnel or migrate (Gaviria and Vélez, 2002). In Brazil, homicide victimization is also more common in lower socioeconomic strata (Soares, 2006). Finally, in Argentina, most of the increase in burglary rates was shouldered by the poor, since the rich were able to adopt effective protective strategies (Di Tella et al., 2010).

3.2 A comment about endogeneity

The problem of endogeneity arises through the correlation between the regressors and the random disturbances. Consequently, treating the crime rate as an exogenous variable may result in a biased estimated elasticity because crime occurs disproportionately in poorer neighbourhoods with low housing prices or, conversely, if criminals target areas where housing prices are higher. In both cases, the behaviour of neighbours will depend on their individual characteristics and these may well be systematically related to unobserved determinants of housing prices (Gibbons, 2004; Buonanno, et al., 2013). We may infer a causal relationship between local characteristics and housing prices, when in fact it is the unobserved component that drives neighbourhood characteristics.

To estimate the impact of crime on housing prices is empirically challenging because of omitted variables. Furthermore, it is difficult to find valid instruments. To mention a couple of examples, Gibbons (2004) used instruments such as: crimes on non-residential properties, spatial lags of crime, or the distance to the nearest public house or wine bar. Buonnano et al. (2012) uses the victimization rate 20 years ago and the share of youth aged between 15 and 24 as instruments of crime rates. Nevertheless, it is hard to instrument murders.

In the present paper, the explanatory variable refers to homicides, but not theft or crime in general. Homicide and theft could be correlated, but motives of homicide may differ from the motivations of theft. Indeed, the propensity to report a theft varies with the severity of the incidence. Faced with the difficulty of finding valid instruments, we have tried to cope with the endogeneity problem by including in the right hand side of the equation all the variables that may provoke the correlation between regressors and random disturbances.

4. Conclusions

High homicide rates discourage new residents in a neighbourhood and encourage those people who can to move out to neighbourhoods with a lower homicide rate. This phenomenon has consequences in the housing market. The objective of this paper is to estimate the relationship between homicide rates and housing prices in Cali, Colombia. We use cadastral information of housing prices from 2012 and the average homicide rates of the period 2000-2010 at neighbourhood level.

The analysis was performed using two estimation strategies. In the first, we estimated a model where the dependent variable was housing prices and one of the explanation variables was the homicide rate: we found that a 10% increase in the homicide rate is related to a 2.4% decrease in housing prices. In the second strategy, we estimated a two-stage model: in one stage we estimated the hedonic price at neighbourhood level and then we estimated a regression to test if the homicide rate has a negative relationship with the estimated hedonic price. We found that, on average, a 10% increase in the homicide rate is related to a decrease of between 2% and 2.5% in housing prices.

In order to diminish the homicide rates judiciary and police procedures need to be strengthened, sites where crime is concentrated need to be identified, criminal gangs must be deactivated, and the carrying of weapons must be restricted (most crimes occur with a firearm). In order to improve estimations, future work will be to include house characteristics and evaluate strategies to deal with endogeneity problems.

References

Atkinson; Healey and Mourato (2005). "Valuing the Cost of Violent Crime: A Stated Preference Approach". *Oxford Economic Papers*, 57, Pp. 559-585.

Besley, Timothy and Mueller, Hannes (2012). "Estimating the Peace Dividend: The Impact of Violence on Housing Prices in Northern Ireland". *American Economic Review*, 102 (2), Pp. 810-813.

Bowes, Davis and Ihlanfeldt, Keith (2001). "Identifying the Impacts of Rail Transit Stations on Residential property Values". *Journal of Urban Economics*, 50, Pp. 1-25.

Buonanno, Montolio and Raya-Vílchez (2013). "Housing Prices and Crime Perception". *Empirical Economics*, Vol. 45, Pp. 305-321.

Cameron, Colin and Miller, Douglas (2015). "A Practitioner's Guide to Cluster-Robust Inference". *The Journal of Human Resources*, 50, Pp. 317-372.

Carson; Flores and Meade (2001). "Contingent Valuation: Controversies and Evidence". *Environmental and Resource Economics*, 19(2), Pp. 173-210.

Ceccato, Vania and Wilhelmsson, Mats (2011). "The Impact of Crime on Apartment Prices: Evidence from Stockholm, Sweden". *Geografiska Annaler: Series B*, Swedish Society for Antropology and Geography, Pp. 81-103.

CCSP-JP (2016). "Caracas, Venezuela, the Most Violent City in the World".

Cohen; Rust; Steen and Tidd (2004). "Willingness to Pay for Crime Control Programs". *Criminology*, 42(1), Pp. 89-109.

DAHM (2012). "El Catastro en Cali: Evolución y Recomendaciones de Política". *Departamento Administrativo de Hacienda Municipal*, Alcaldía de Santiago de Cali.

Di Tella, Edwards and Schargrodsky (2010). "The Economics of Crime: Lessons for and from Latin America". In *The Economics of Crime: Lessons for and from Latin America*, National Bureau of Economic Research. University of Chicago Press.

Fallis, George (1985). *Housing Economics*. Butterworths, Toronto-Canada, 1985.

Frischtak, Claudio and Mandel, Benjamin (2012). "Crime, House Price, and Inequality: The Effect of UPPs in Rio". *Federal Reserve Bank of New York Staff Reports*, No. 542.

Gaviria, Alejandro and Vélez, Carlos (2002). “Who Bears the Burden of Crime and Violence in Colombia?”. In *Colombia Poverty Report*, Vol. 2, 146-161. Washington, D.C. World Bank.

Gibbons, Steve (2004). “The Cost of Urban Property Crime”. *The Economic Journal*, Vol.114, No. 499, Pp. 441-463.

Glaeser, Edward and Sacerdote, Bruce (1999). “Why is there more Crime in Cities”. *Journal of Political Economy*, Vol. 107, No. S6, Pp. S225-S258.

Hellman, Daryl and Naroff, Joel (1979). “The Impact of Crime on Urban Residential Property Values”. *Urban Studies*, 16, Pp. 105-112.

Levitt, Steven and List, John (2007). “What do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?”. *The Journal of Economic Perspectives*, Vol. 21, No. 2, Pp. 153-174.

Rizzo, Mario (1979). “The Effect of Crime on Residential Rents and Property Values”. *The American Economist*, Vol. 23, No. 1, Pp. 16-21.

Rosen, Sherwin (1974). “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition”. *Journal of Political Economy*, 82(1), Pp. 34-55.

Soares, Rodrigo (2006). “The Welfare Cost of Violence across Countries”. *Journal of Health Economics*, Vol. 25 (5), Pp. 821-846.

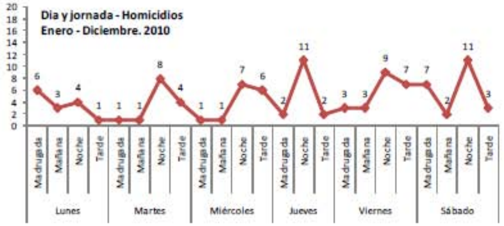
Soares, Rodrigo (2010). “Welfare Costs of Crime and Common Violence: A Critical Review”. *Pontifical Catholic University of Rio de Janeiro and IZA*.

Thaler, Richard (1976). “A Note on the Value of Crime Control: Evidence from the Property Market”. *Journal of Urban Economics*, 5, Pp. 137-145.

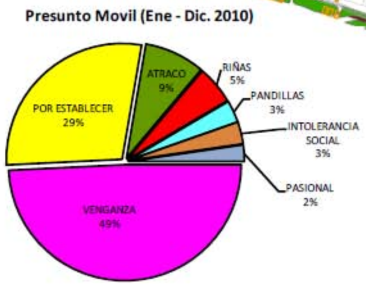
The Economist (2016). “The World’s most Violent Cities”. February 3rd 2016, http://www.economist.com/blogs/graphicdetail/2016/02/daily-chart-3

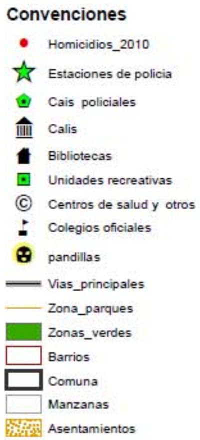
White, Halbert (1980). “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity”. *Econometrica*, 48, Pp. 817-838.

Proyecciones de la población de Cali según total, cabecera y resto															
	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Total Cali	2.108.257	2.133.143	2.158.078	2.183.065	2.207.994	2.232.996	2.388.927	2.431.965	2.475.249	2.518.808	2.562.632	2.606.721	2.651.076	2.695.696	2.740.581
Comuna 6	175.202	180.216	185.229	190.239	195.242	200.045	216.640	223.077	229.497	235.905	241.799	347.718	253.662	259.631	265.627



Mecanismo	Casos
ARMA DE FUEGO	118
ARMA BLANCA	9
OTRAS ARMAS	3
Total general	130





HISTORICO - HOMICIDIOS POR BARRIOS, POR AÑOS ENERO - DICIEMBRE 2010														
Comuna	Codbarrio	Barrio	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Suma
6	1	SAN LUIS	19	10	19	13	25	3	6	5	6	16	5	**127**
6	2	JORGE ELIECER GAITAN	9	9	4	11	4	11	2	6	5	6	4	**71**
6	3	PASO DEL COMERCIO	2	2	5	3	2	4		3			5	**26**
6	4	LOS ALCAZARES	7	5	3	8	9	6	6	7	7	9	10	**77**
6	5	PETECUY PRIMERA ETAPA	4	16	24	16	17	10	11	4	9	12	25	**148**
6	6	PETECUY SEGUNDA ETAPA	10	7	3	14	3	4	8	8	8	12	13	**90**
6	7	LA RIVERA 1	3	2	11	5	3	3		1	2	2		**32**
6	8	LOS GUADUALES	6	2	4	2	8	1	4	1	1	5	2	**36**
6	9	PETECUY TERCERA ETAPA	6	4	5	3	8	6	3	7	6	12	8	**68**
6	10	CIUDADELA FLORALIA	14	22	28	19	24	16	11	14	9	34	26	**217**
6	11	OASIS DE COMFANDI						1	1				1	**3**
6	12	URB.TEJARES DE SALOMIA						1	1	1			1	**4**
6	13	LARES DE COMFENALCO								1		2	8	**11**
6	94	QUINTAS DE SALOMIA					1	2		1			1	**5**
6	95	FONAVIEMCALI		2		3				2		1	1	**9**
6	96	SAN LUIS II	3	12	2	2	10	5	6	10	7	7	10	**74**
6	97	URBANIZACION CALIMIO	2		2	5	3	2	3	3	7	8	10	**45**
6	98	SECTOR PUENTE DEL COMERCIO		1	4	2	1							**8**
6	99	PORTAL DE COMFANDI						1	1		1	1		**4**
TOTAL COMUNA - POR BARRIOS - POR AÑOS			**85**	**94**	**114**	**106**	**118**	**76**	**63**	**74**	**68**	**127**	**130**	**1055**

Appendix 1 – Information about *Comuna 6*

Source: *Observatorio Social de Cali*

Appendix 2 - Potential consequences on the estimated standard errors derived from forming false clusters

When the statistical information is clustered (e.g. classrooms, neighbourhoods, economic sectors) the specialized literature recommends estimating models using cluster standard errors. In this section, we have carried out a simulation exercise to show potential consequences on the estimated standard errors derived from forming false clusters. The starting point is the OLS estimation of the regression model in Table 3:

$$y = X\beta + u,$$

where y represents the log of housing price and the explanatory variables are the area, the square of the area, and the fixed effects of neighbourhoods. Using the OLS estimation of $\beta = \hat{\beta}$ and the standard error of disturbances $\hat{\sigma}_u = 0.39$, we generate a new dependent variable y^s as follows:

$$y^s = X\hat{\beta} + u^s,$$

where u^s is a normal independent random number distribution with a zero mean and a standard deviation of $\hat{\sigma}_u = 0.39$. These random disturbances verify the standard hypothesis of the regression model; then, the population matrix of variances and covariances of the estimated coefficients is obtained as follows:

$$cov(\hat{\beta}) = (0.39)^2(X'X)^{-1}$$

Once the simulated dependent variable, y^s, has been generated, we estimate the equation with errors clustered at neighbourhood level, which means that we are dealing with false clusters. The new estimated covariance matrix is $cov(\tilde{\beta})$. This strategy facilitates the comparison between the population covariance matrix $cov(\hat{\beta})$ and the estimated covariance matrix $cov(\tilde{\beta})$. We estimate the equation with clustered errors 1,000 times. For each standard error of the coefficient, we calculate the percentage of error, pe_j, as follows:

$$pe_j = \left[\frac{\sqrt{var(\tilde{\beta}_j)} - \sqrt{var(\hat{\beta}_j)}}{\sqrt{var(\hat{\beta}_j)}}\right].100$$

Figure 4 presents the percentage of error. The area error seems to be free from bias, however, the absolute value of the error is 6.3%. The worrying results for the cluster option are those related to neighbourhood fixed effects, which imply an important negative bias. Figure 1A shows the distributions of the percentage of error, pe_j, for two groups of variables: area of the house and neighbourhood fixed effects. The average of the distribution of errors for neighbourhood fixed effects is -96%. The main conclusion is that false cluster can have an important cost. Nevertheless, in this paper the statistical significance of the coefficients is invariant to the option selected as is shown in Table 3 of the main text.

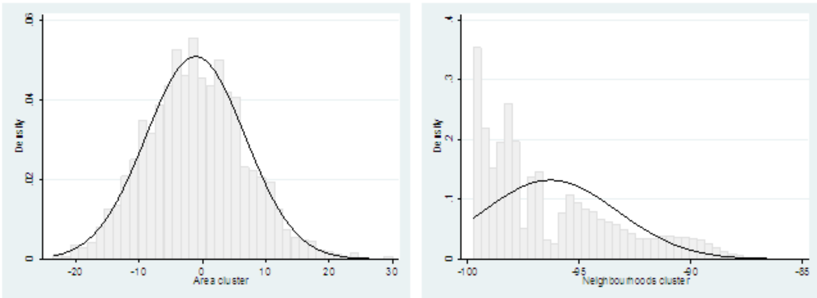


Figure 1A-Consequences of forming false clusters

Figure 2A shows results from a similar simulation process using the heteroskedasticity correction suggested by White (1980). The estimations of the covariance matrix are denominated *robust*. As in the previous exercise, the robust correction is unnecessary but enables us to analyse the potential effects of making such corrections.

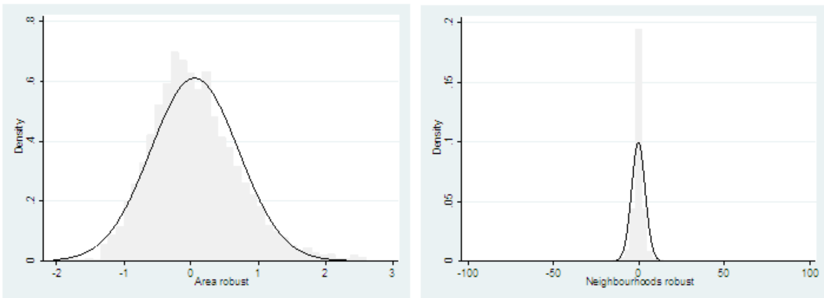


Figure 2A-Consequences of unnecessary robust correction

ARRIVAL

The three empirical studies that make up this thesis analysed different and significant urban phenomena: unemployment, informality and crime.

Workers who reside in some areas within a city do not have enough job opportunities. Given these circumstances, specialized literature has taken geographic distance into account as a potential determinant of the labour market results. The first chapter was aimed at testing whether job accessibility matters for individual employment in Bogotá (Colombia). Since 1950, this city has experienced an uncontrolled growth phenomenon in peripheral neighbourhoods and a socio-spatial segregation process that probably have affected labour market outcomes.

In the empirical exercise, we use information about job location for the year 2008. Information about workers' characteristics comes from household surveys from the years 2008 and 2009. After controlling for individual characteristics, the main result of this research is that job accessibility is a significant determinant of the probability of being employed, especially for women. Therefore, investments in public infrastructure that increase accessibility to employment would significantly lower frictional unemployment. To control for potential simultaneity problems the regressions were estimated using instrumental variables (distance between the census tract of residence and the nearest neighbourhood in 1950).

The literature provides evidence about the relationship between spatial density and aggregate increasing returns. Moreover, the structure of a city is determined by a production externality under which employment at any site is more productive the higher the employment at neighbouring sites. As profit increases with productivity, firms should locate where their expected profit is the highest. Nevertheless, when the property-rights system fails, the market cannot work efficiently, because the owner has de facto rights to their property but does not have a legal enforceable title.

The literature does not say too much about the structure of cities where the percentage of the informal sector is significant. In the second chapter we analysed the case of Cali (Colombia) where 42% of firms are informal. We shed light on how formal firms tend to be located in high density commercial areas, meanwhile informal firms are located in low density and peripheral areas where the land is cheaper and where they can avoid the control of authorities.

We set out to estimate the effect of formal agglomeration on the probability of being informal, which could be interpreted as a local share of informal firms. The main result is that when agglomeration increases by one standard deviation the local share of informal firms diminishes by 16%. This may happen because informal firms have less opportunity to benefit from agglomeration effects because of legal restrictions that block their relationship with formal firms. The result explains why formal and informal firms display different locational patterns in the urban structure. We conducted an IV analysis in order to tackle the potential endogeneity problem.

High homicide rates discourage new residents inform moving to a neighbourhood and encourage those people who can to move out to neighbourhoods with lower homicide rates. This phenomenon has consequences in the housing market. The objective of the third chapter was to evaluate the relationship between homicide rates and housing prices in Cali, Colombia. We use cadastral information of housing prices from 2012 and the average homicide rates of the period 2000-2010 at neighbourhood level.

The analysis was performed using two estimation strategies. In the first, we estimated a model where the dependent variable was housing prices and one of the explanation variables was the homicide rate: we found that a 10% increase in the homicide rate is related with a 2.4% decrease in housing prices. In the second strategy, we estimated a two-stage model: in one stage we estimated the hedonic price at neighbourhood level and next we estimated a regression to test if the homicide rate has a negative relationship with the estimated hedonic price. We found that, on average, a 10% increase in the homicide rate is related with a decrease of between 2% and 2.5% in housing prices.

Policy implications

Today, more than half of the global population live in urban areas. By the middle of the present century the percentage will be close to 70%. From this perspective, we are forced to think about improving the efficiency of cities. Among the policy implications derived from the results of this thesis, we will mention three:

1. Access to jobs has significant effects on employment. Thus, by improving commuting within the city we can bring about better connections between residence locations and job opportunities. We know that urban areas develop in polycentric structures. Therefore, public policy has to take into account the improvement of commuting routes to diminish the negative effects of decentralized urban structures.
2. Informality encompasses a heterogeneous range of activities, and finding strategies to encourage formalization is a complex task. In Latin America, it is a lifestyle. The Colombian government has approved laws to facilitate formalization, essentially reducing the costs of registration and introducing programs to advise entrepreneurs about the advantages of being formal (e. g. subsidies, tax reductions and financial access). Taking into account the results obtained in Chapter 2, we could design a system to control the relationship between formal and informal firms. Several formal firms buy goods and services from informal firms because it is cheaper, so informal firms benefit somehow from the agglomeration processes. It is essential that informal firms formalize in order to benefit efficiently from these processes.
3. Amenities have positive effects on housing prices and quality of life; moreover, crime generates negative effects. The best strategy that a government has to reduce crime is through education and job opportunities. Moreover, judiciary and police procedures need to be strengthened; sites where crime is concentrated must be identified, criminal gangs must be deactivated, and the carrying of weapons must be restricted (most crimes occur with a firearm).

Future work

There are limitations in this research. Consequently, these limitations help to identify future research lines.

In order to improve the research about job accessibility and its effect on labour market results, we could consider information about commuting times using different transport systems (e.g. bus systems or personal vehicles). With adequate statistical information, it would be possible to estimate the approximate change in welfare resulting from improvements in transport infrastructure. In addition, we could complement the results of job accessibility analysis with estimations of its effect on wages. Indeed, residential segregation or social mobility processes are an important dimension of socioeconomic disconnection and labour force participation.

The debate about informality in developing countries is huge. In the second chapter of this thesis, we studied a particular urban consequence of the higher levels of informality. But, an analysis of co-agglomeration or co-clustering processes could also be considered, in order to identify patterns where informal companies have incentives to follow formal firms in certain industries. Furthermore, we know that low education levels, poverty and inequality are factors intrinsically related with informality. Therefore, in order to improve the analysis, we would like to take into account the human capital of firms in order to control for more characteristics. Furthermore, knowing that the urban planning office has to include the location of informal firms as an important dimension in the politics of land use, it would be interesting to complement the research of agglomeration and informal firms with an analysis from an institutional economic perspective.

Finally, regarding the subject of the third chapter, in order to improve estimations, we know that additional house characteristics would need to be included (unfortunately these variables are not available) and strategies to deal with potential endogeneity problems need to be evaluated.

Acknowledgments

Above all I thank Josep Lluis Raymond, who has supervised this research project. The author gives thanks to Jos van Ommeren from the Department of Spatial Economics (Vrije Universiteit Amsterdam) who has supervised the second chapter of this thesis and to Ivan Muñiz for his helpful constructive criticisms. Finally, thanks to the Departamento Administrativo de Ciencia, Tecnología e Innovación in Colombia and to the Agència de Gestió d'Ajuts Universitaris i de Recerca for their financial backing in different stages of the doctoral academic program.