Neighborhood characteristics and the spatial concentration of crime: Evidence for Bogotá.[1]

Daniel Felipe Martínez Enríquez

Colegio Mayor Nuestra Señora del Rosario

Master Thesis

Thesis Advisors: Darwin Cortés, Juan Miguel Gallego

Abstract

Using a unique neighborhood crime dataset for Bogotá in 2011, this study uses a spatial econometric approach and examines the role of socioeconomic and agglomeration variables in explaining the variance of crime. It uses two different types of crime, violent crime represented in homicides and property crime represented in residential burglaries. These two types of crime are then measured in non-standard crime statistics that are created as the area incidence for each crime in the neighborhood. The existence of crime hotspots in Bogotá has been shown in most of the literature, and using these non-standard crime statistics at this neighborhood level some hotspots arise again, thus validating the use of a spatial approach for these new crime statistics. The final specification includes socioeconomic, agglomeration, land-use and visual aspect variables that are then included in a SARAR model an estimated by the procedure devised by Kelejian and Prucha (2009). The resulting coefficients and marginal effects show the relevance of these crime hotspots which is similar with most previous studies. However, socioeconomic variables are significant and show the importance of age, and education. Agglomeration variables are significant and thus more densely populated areas are correlated with more crime. Interestingly, both types of crimes do not have the same significant covariates. Education and young male population have a different sign for homicide and residential burglaries. Inequality matters for homicides while higher real estate valuation matters for residential burglaries. Finally, density impacts positively both crimes.

Keywords: Homicides, Residential burglaries, Neighborhoods, Spatial econometrics

JEL classification: D63, K420, C21, R1

[1] I am especially grateful to Darwin Cortés and Juan Miguel Gallego for their excellent supervision. I thank Javier Perez, Carlos Sepúlveda, Ana Isabel Moreno, Juan Daniel Oviedo and Juan Fernando Vargas for their valuable comments and suggestions. I gratefully acknowledge financial support from Ceiba-Complexity, A center of Excellence in complex-systems research for Colombia. Finally I thank the MEBOG-OBSCRIM for granting me access to the crime maps for Bogotá, 2011.

Table of Contents

1. Introduction

Crime and violence in urban environments have become one of the biggest challenges for dwellers around the world. Latin America is the region that stands outs when it comes to crime rates. It has been the only region where murder has grown during the past decade.[2] Therefore, insecurity has become an urgent topic for citizens in Latin America given the hazardous effects on economic activity, investment and more importantly on the quality of life is strikingly high. For example, the world investment and political risk report by World Bank (2010) claims that a high crime rate of a nation strongly deters investment, condemning it to stagnant development.

Still, despite the fact that many different explanations for crime have been studied, there is no consensus regarding the reasons behind the crime problem. Moreover, most studies in the literature focuses on the experience of developed countries, where institutions and socio-economical characteristics are in stark difference from those in Latin America, so there are only some lessons that can be drawn beyond certain generalities.

The amount of economic studies focusing on Latin America is small compared to developed countries and these studies for the region have stated the unique and complex dynamics of crime. They show that the deterioration of security in the region has not been uniform across countries. In some countries homicides is what affects population, whereas in other countries crimes against properties are the most worrisome. The local dimension of crime and violence is fundamental in understanding the crime phenomena in Latin America. Not only does crime concentration differ

[2] Between 2000 and 2010, the murder rate in the region grew by 11%. In other regions the rate either fell or stabilized. (UNODC report 2011)

from country to country or from city to city, but also across smaller units such as districts and neighborhoods (UNDP, 2014).

The case of Bogotá is very interesting. To start with, it is the capital of a country with a civil conflict that has lasted for at least 50 years. In spite of that, the city has managed to reduce the homicide rate in the past twenty years from 80 to 17 murders per 100.000 habitants. However, not only is the homicide rate still in an epidemic level by WHO standards, but also people feel unsafe about their property even in their neighborhoods.[3] Despite the fact that urban crime is treated as a priority in policy for the city's government, we know little regarding the economic, social, and institutional factors that make some neighborhoods safer or ridden with crime. Previous studies of crime in Bogota identified the presence and importance of crime clusters in certain parts of the city, but have mixed evidence on the incidence of neighborhood characteristics other than location. The objective of this paper is to increase the literature of urban crime in Bogotá by studying smaller units of neighborhoods, and include neighborhood characteristics (socioeconomic, land-use, visual) that are potentially relevant in explaining violent and property crimes in the city.

Using a unique neighborhood homicide and residential burglary dataset for Bogotá 2011, this paper makes five contributions. First, I overcome the challenge of data availability by introducing a crime statistic that is not standard in the literature but is spatially similar to crime rates. Second, I use a cross section dataset for crime in 2011 at a smaller unit of analysis than most studies for Bogotá with over 900 identified neighborhoods. Third, I include socioeconomic and agglomeration characteristics that remain significant even when using a spatial approach. Fourth, I identify the direct, indirect and total effect of those neighborhood characteristics on homicides and residential burglaries. Fifth, I find that the socioeconomic drivers are different between property and violent crimes in both sign and magnitude. However, agglomeration drivers have a positive sign leading to the hypothesis that more agglomeration means more crime irrespective to the type of crime.

As mentioned before, in order to achieve the objective of this paper, I use a spatial econometric approach. The use of this econometric approach is fundamental in studying events that happen in a geographical setting, where it incorporates as a covariate the notion of contiguity and neighbor dependence that are ignored in other methodologies. By including this spatial concept it is possible to solve some of the omitted variable problems and the existence of spatial autocorrelation. Additionally, the spatial econometric approach used here is statistically stronger than those in previous studies for Bogotá that do not control for spatial autocorrelation on the errors or additional heterogeneity. Finally, I will calculate the direct effect of the independent variables and the indirect effect of those variables via the spillover effect.

The rest of the paper is organized as follows. Section 2 presents a literature review making emphasis on empirical literature for Bogotá. Section 3 presents data and introduces the

[3] The victimization survey for Bogotá in 2012 established that only 32% of the respondents regard the neighborhood where they live as safe, and those same respondents blame property crimes, muggings and drug dealers as the main reason to feel unsafe.

exploratory spatial data analysis. Section 4 introduces the econometric method. Section 5 summarizes the main results. Section 6 presents final remarks.

2. Literature Review

2.1. Literature on socioeconomic characteristics and crime

The economic analysis of crime has its seminal contribution in Becker's (1968) standard model of crime as occupational choice. It describes a situation where an individual allocates his working time between the legal and the illegal sector in order to maximize his welfare. Under uncertainty, an individual chooses to commit crimes if the expected gain from a successful crime is bigger than the expected cost from being caught, punished, and diverting time from legal activities. Thus, economic research on crime has focused on either deterrence challenges or economic factors that affect the costs and benefits related to criminal activities.

The literature focusing on the benefits and costs of crime has been rich, particularly in developed countries. Ehrlich (1973) explored the effect of unemployment rates, income levels and income disparities on the incidence of crime. He finds a significant crime-inducing impact of income levels and income inequality for cities in the U.S. Interestingly, unemployment as a complementary indicator of legal income opportunities was not an important determinant of crime rates (Grogger, Freeman, 1994; Masciandaro, 1999; Imrohoroglu et. al, 2000).

Another important factor related to the effect of economic conditions on crime is the level of education of the population, which can determine the expected benefits from both the legal and crime activities. Moreover, Usher (1997) considers education to have a "civilization effect", tending to reduce globally the incidence of crime activity. A priori, one would expect that education raises the expected gains in the legal sector and therefore deterring people off crime. Lochner and Moratti (2004), and Gallipoli and Fella (2006) developed theoretical crime-education models that derives the crime-education relationship, and both find it negative. The empirical parts of these studies are usually supportive of the idea of education as a deterrent of crime (Freeman, 1994; Lochner, 1999; Lochner and Moratti, 2001; Buonanno, 2006). However, the effect of education on crime reduction is controversial in some studies. Ehrlich (1975) finds a positive relationship between the number of school years completed and property crimes committed across the U.S. in 1960.

Crime is usually a male's business and more specifically a youth male one. Criminal records show that incarcerated people tend to be males who are less educated and from poorer economic background. Criminal activities typically increase with age until the late teens and then decline (Freeman, 1991, 1996; Grogger 1991, 1995 and 1998; Lochner 1999). Most studies argue that the drop in both real earnings and employment opportunities for less educated young men are the main reason for crime involment. Moreover, the long term decline in the probability of employment for less educated population induces youth into crime.

Finally, crime and inequality is a link that has been studied several times in the literature. The usual implication is that more inequality certainly leads to more crime. The usual channel for this is that, when poor individuals who have low returns in the legal market live next to high-income individuals, they tend to commit more time to crime. There is one pure economical reason for this and another more psychological. The economical reason is that the presence of high income individuals increases the returns to time allocated to criminal activity. The psychological explanation is an "envy effect" such that it reduces the individual's moral threshold. Kelly (2000), Fajnzylber et al. (2002), Brush (2007), Choe (2008) and Menezes (2013) have found significant and positive effects from inequality on violent crime rates. However, the effect of inequality on property crimes is still debatable with Fajnzylber et al. (2002) establishing a positive robust correlation, meanwhile Neumayer (2005) finds the same correlation to be non-significant.

2.2. Literature on agglomeration variables and spatial dynamics on crime

It is well documented the existence of an agglomeration in criminal activities with varying units of analysis. Higher crime rates can be found in larger cities compared to small cities and rural areas (Glaeser and Sacerdote, 1999). In Latin America, property theft victims are located in bigger cities. Moreover, households living in cities with higher population growth are more likely to feel victimized (Gaviria et al., 2002). Glaeser et al. (1996) describe this situation as ***"****one of the oldest puzzles in the social sciences; this variance appears too high to be explained by changes in the exogenous costs and benefits of crime* ***".*** There exists a link between urbanization and crime, where the channel is the population density and the lower probability of being caught. Also, population density affects crime through higher pecuniary return to crime, social interactions and development of tastes. (Kelly, 2000)

Sherman, Gartin, and Buerger (1989) pioneered a study where crime within a city is shown to be highly concentrated in relatively few small areas for Minneapolis. These agglomerations show that there exist "hot spots" of crime in urbanized areas. Within a city, several adjacent neighborhoods are crime ridden and this could be the result of a contagious diffusion process in the past. The contagious diffusion approach is an epidemiologic explanation for crime-ridden areas clustering together. Cohen and Tita (1999) exemplify it by gang rivalries and advancing crack markets that act as drivers for the expansion of violence from one adjacent neighborhood to another.

2.3. Empirical Review for Bogotá

Empirical literature whose main purpose is to study the urban crime in Bogotá is rather scarce and outdated. This could be because researchers focus on analyzing the existence of guerrillas and civil conflict in Colombia. Bogotá had its peak of violence almost 20 years ago which could also lead researchers to focus their attention away from Bogotá and into rural areas and other cities. Most of the literature on urban crime for Bogotá was published during a short period of time (1997-2003). These early studies are included in the set of documents *Paz Publica* and were edited, and revised by the center of studies on economic development (CEDE) from Universidad de los Andes.

Rubio (1997), Echandía et al. (2000), Rubio et al.(2000), and Llorente et al.(2001) all focus their attentions on the statistics of homicide reported by the National Police and the National Institute of Forensic Medicine. From these studies two important conclusions arise: First, the existence of a non-random pattern of homicides in Bogotá, and second: the prevalence of instrumental homicide over expressive homicide.[4] These studies also suggest the existence of criminal structures that are present and systematically are associated with “hot spots” of homicides. They concluded the existence of “hot spots” by overlapping a homicides map and a map which contained information on the presence of criminal structures and gun trafficking. This explanation is based on the contagious diffusion theory (Cohen and Tita, 1999). Interestingly, all of these studies discussed the “objective causes” of homicides yet all of them dismiss socioeconomic factors as possible determinants of crime.

The existence of some other explanations in determining homicide rates is re-evaluated by Formisano (2002) and Sanchez et al. (2003). Both studies take into account the spatial structure of the homicide data in Bogotá and analyze the homicide data with an econometrical approach. Formisano (2002) found evidence supporting the contagion and diffusion effect for homicides. He also found supporting indication for the presence of gangs, illegal drugs and lack of public lighting. However, measures such as prostitution or bars are not significant. Sanchez et al. (2003) calculates the effect of punishment measures in the reduction of crime, they show that increases in the amount of police officers and incarceration rates are the main drivers in reducing the homicide rate at the locality level. Bars and prostitution are significant but do not amount to much. Once again, no socioeconomic variable is significant in explaining the prevalence of crime in an area.

Bourguignon et al. (2003) use property crime rates for the 7 largest Colombian cities and find that a specific part of the income distribution rather than usual measures of inequality is related to property crime. In Colombia, what matters in terms of inequality causing crime is the population living under 80% of the mean income. A change in inequality for the people living above this threshold does not induce variations on property crime rates.

Gaviria et al. (2010) asses the causal relation between adolescent fertility and homicide rates when those children of adolescent mothers reach their peak crime ages. They find a positive effect on homicide only when the neighborhood has: High adolescent fertility rates, low school enrollment, and high crime rates at the time of the mother’s pregnancy. Finally, increases in school enrollment always reduce the homicide rate showing that education is an important driver in reducing homicides.

Finally, Escobar (2012) using a spatial regression and explanatory spatial descriptive analysis finds that the theories of social “disorder” partially holds in explaining the homicide rates in Bogotá. Meanwhile the concentration of social disadvantages and insulation are important in explaining higher homicides rates. There are some unexpected effects according to the author, in both the proportion of young male population and population density, Escobar finds it puzzling that more

[4] The classification of homicides arises from the motives behind it; instrumental homicides are motivated by gain, meanwhile expressive homicide is emotional, rage based and maybe purely impulsive.

densely populated places and higher proportion of young males attract less crime rates in her data.

3. Data and the Exploratory Spatial Data Analysis (ESDA)

I want to use variables that are significant and relevant at the minimum level of social interaction that is possible. Theoretically one would use "the neighborhood" since it marks the level of social interaction that does not impose a monetary cost on an individual for engaging in social interactions with others. However, this measure is purely theoretical since any individual could have different imaginary boundaries to what his neighborhood looks like. Therefore, I rely on the cadastral information that divides the city in cadastral sectors. A cadastral sector is a subdivision of the city territory which is composed of several blocks; it is limited primarily by the traffic net and geographical accidents; it also takes in account the original extension of the large properties that were divided and sold to build on them. Finally, the city is divided in 1153 cadastral sectors which are a large amount compared to other city management divisions of the territory such as the urban planning zones and localities (120 and 19 respectively). It is also important to compare the cadastral sector to the censal sector which divides the city in 692 areas. The censal sector has one main disadvantage in that it does not take in account the traffic net and sometimes group several neighborhoods or urbanizations together for being too small.

3.1. Creating the Dependent Variables

By setting the neighborhood boundaries at the cadastral sector level it is therefore necessary to find crime data at this unit of analysis. Nevertheless, it was not possible to find crime rates or crime counts at the desired unit of analysis. Most crime datasets for Bogotá that are publicly available contains crime rates at the 19 urban localities level. The Criminal Observatory of the Police Department for Bogotá (MEBOG-OBSCRIM) could not provide crime statistics at this unit of analysis for considering those to be restricted information. However, they provided me with a set of scanned maps that were compiled for a city management report on localization of crime.

The spatially located crime database is a set of scanned maps provided by the MEBOG-OBSCRIM. It contains images for 8 types of crimes in 2011. In each image, there is a map of Bogotá and a point data that represents a reported occurrence of a crime. Given the quality of the images and the scale in which these maps are presented, it is neither possible to know the precise location for each crime nor is possible to know the actual amount of crime or whether more than one crime was committed in that relative location (see figure 1, Map (A)).

I will create the crime variables for this study using only the maps from homicides and residential burglaries as measures of crime. Both of them are different in nature and both are mostly reported crimes.[5] Homicide is a violent crime while residential burglaries is a property crime that

[5] I use data from reported crimes by the Metropolitan police of Bogotá, not victimization or household surveys. Official crime statistics usually have unreported crime. Bogotá have a large amount of underreporting in robberies, muggings, and even some property thefts such as motorcycle theft and commercial theft.

not necessarily entails harm to the residents. I put geographic coordinates on the scanned maps and matched them to a map containing all the neighborhoods in Bogotá. I manipulated each crime map using an image manipulation program.[6] The manipulated maps are then censored, filtered, decomposed, and posterized in order to show only the red pixels which compose one data point.

The new maps are only in red and white, where each pixel represents a square area of 40mX40m (see figure 1, Map (B)). Given the quality of the maps and the filtering process, one red pixel does not necessarily represent one crime and several nearby red pixels doesn't clearly represent one data point. Since several nearby pixels could represent any number of crimes, I decided to create polygons around the center of each pixel. The radius of each polygon is adjusted so it closely resembles the crime data point presented by the MEBOG. If two polygons intersect they will merge into a single one. The highlight of the final map (see figure 1, Map (C)) are the orange lines that represent a polygon and circle each data point presented by the MEBOG.

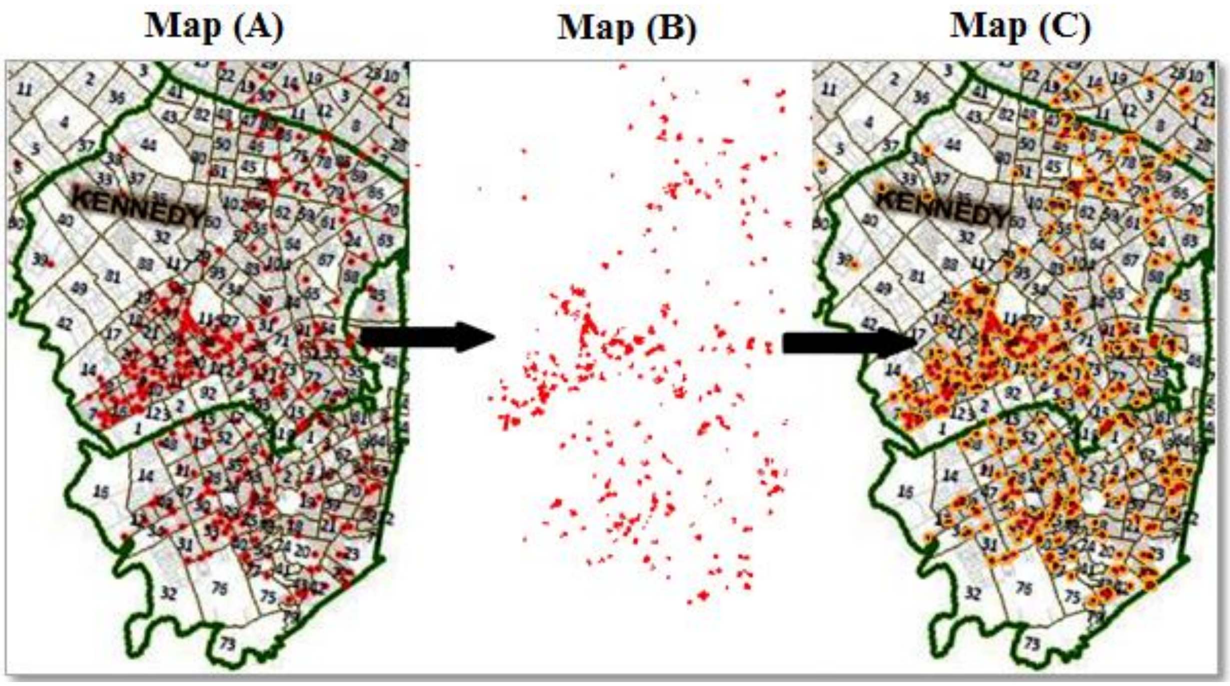
It is not possible to calculate crime rates with this dataset extracted from the maps. Because it is not possible to uniquely identify each data point that represents a crime. It is possible to count the red pixels that compose a data point but this would clearly overestimate the incidence of crime. On the other hand, it is possible to count the amount of polygons, but this option will underestimate the amount of crimes in areas where the prevalence of crime is high. In order to overcome this challenge I propose a new statistic to measure the prevalence of crime and the exposure of neighborhood residents to it. This statistic proposed here accounts for how widespread is a type of crime for the neighborhood and is defined as the area each polygon covers as percentage of the total area of the neighborhood.

These proposed crime statistics are not standard measurements of crime. Thus they are not exempt from critique and thorough examination. For example, this measurement is dependent on the size of the neighborhood since smaller neighborhoods will have a higher probability to score high in the crime statistic presented. In the case for this study, most neighborhoods are relatively the same size. However other neighborhoods such as the airport and semi-rural neighborhoods are large in comparison. A boxplot graph in the appendix shows the existence of these outliers and a small interquartile range points toward a minimization of this size problem. On the other hand these crime statistics measures equally a street where there has been 5 homicides than one with a single homicide during 2011, in other words, these statistics don't account for the intensity of crime but rather how widespread it is. Therefore, it is possible to underestimate the severity of the crime phenomenon in certain neighborhoods.

The two dependent variables created here are: Homicide exposure as percentage of the neighborhood area, and residential burglaries exposure as percentage of the neighborhood area. From here on, when I refer to homicides or residential burglaries I will refer to these variables created here and not the usual crime rates.

[6] I used Gimp, which is a freely distributed piece of software for image retouching, image composition and image autoring.

Figure 1: Creating the crime variables



Map A is a selection of the original scanned image by MEBOG. Map B is the manipulated image that extracts red pixels from the original map in A. map C overlays both maps A and B, it also includes the polygon created around each occurrence.

3.2. The independent variables

The independent variables will be divided in three different categories: Socioeconomic, agglomeration and land use.

The socioeconomic variables are: average years of education in the population, unemployment rate in the neighborhood, percentage of young male population. These variables were extracted from the 2005 national census at the city block level. The real estate valuation per m2 and the inequality index were constructed using the data at the building level. Building data comes from Bogotá's real estate census in 2011.[7]

The agglomeration variables are: population density, percentage of constructed area, and the average amount of floors in each neighborhood. These variables were extracted at the neighborhood level from both the District Repository of Infrastructural Spatial Data (IDECA) and the district planning department (SDP).

Land-use variables are: percentage of nonresidential buildings in the neighborhood, a dummy for current residential construction and a dummy for current non-residential construction. Both

[7] The census contains information at the building level on the construction characteristics and the variables used by the National Department of Statistics (DANE) to obtain the real estate valuation. Those variables include the construction score as a summary of the overall current state of the building; it also includes the observational variables that allow creating the construction score represented in four main categories: structure, architectural finishing, restrooms, and kitchen.

Agglomeration and Land-use variables were taken from the IDECA at the building level and the SDP at the Neighborhood level.

Finally, I use a proxy for the visual aspect of the buildings that will characterize the current state of the building. I created the percentage of buildings rated with bad state of conservation in the building's finishing as a measure for the physical deterioration of the neighborhood. This variable comes from Bogotá's real estate census in 2011. The census classifies the state of conservation of a building based on the materials, age, and maintenance of the facades, walls and floors. They rate the building as bad conserved when it has a combination of precarious state, bad materials and old construction.

In order to combine all of the above data, it was necessary to georeference it and merge all the information to the closest measure of neighborhood, cadastral sector, which as noted above is a subdivision of the district's territory created for the main purpose of cadastral mapping. For simplicity, I will use neighborhood and cadastral sector interchangeably from here on. The collected dataset contains information on 952 cadastral sectors of the 1153 cadastral sectors in Bogotá. The main difference between the official number of sectors and the number in the collected dataset comes from the sectors located in the rural locality of Sumapaz, and certain large areas that are considered cadastral sectors yet they are in fact recreational parks, the national university, and the national administrative center.

Table 1: Descriptive Statistics

Variables	N	Mean	Sd	Min	Max	Source
Homicide area incidence	952	0.128	0.168	0	1	MEBOG-OBSCRIM
Residential burglaries area incidence	952	0.309	0.22	0	0.942	MEBOG-OBSCRIM
Unemployment rate in 2005	952	0.077	0.034	0	0.444	DANE-CENSUS 2005
Percentage of young male population aged 10-25 in 2005	952	0.287	0.057	0	0.486	DANE-CENSUS 2005
Average years of education in 2005	952	7.397	2.559	0	13.642	DANE-CENSUS 2005
Gini coefficient for the real estate valuations	952	0.251	0.115	0	0.875	DANE-REAL ESTATE CENSUS 2011
Real estate valuation per constructed m2 (millions)	952	0.858	3.794	0.029	97.621	DANE-REAL ESTATE CENSUS 2011
Percentage of buildings área	952	0.354	0.153	0	0.677	SDP/IDECA
Population density	952	0.023	0.015	0	0.092	SDP/DANE
Amount of floors	952	2.049	0.686	1	6.34	SDP/IDECA
Percentage of nonresidential área	952	0.227	0.218	0	0.999	SDP/IDECA
Dummy for current nonresidential construction	952	0.066	0.249	0	1	SDP/DANE
Dummy for current residential construction	952	0.69	0.463	0	1	SDP/DANE
Neighborhood área	952	0.386	0.437	0.008	7.234	SDP/IDECA
Percentage of bad conserved finishings	952	0.505	0.335	0	1	DANE-REAL ESTATE CENSUS 2011

Source: own calculations

3.3. ESDA: Exploratory spatial data analysis

Past literature has established the existence of a spatial structure for crime at almost any level of data aggregation. Empirical studies for Bogotá have found that homicides cluster around certain parts of the city mainly in the south and near the historic center. Those clusters of crime were found using both observational and spatial descriptive statistics. Formisano (2002), Gaviria (2008) and Escobar (2012) used the homicide rates at the censal sector level[8] and all of them found statistically significant clusters. Sanchez et al (2003) use homicide rates at the locality level and find it relevant to use a spatially weighted regression given the crime cluster in the south localities.

Since the dependents variables I propose here have not been used before in the literature for Bogotá, it is important to see if these new crime variables are in line with the studies before and check whether exists a crime clustering at this level of aggregation or not. Past empirical studies have used different measures of crime and neighborhood definition. Therefore, in order to compare "hot spot" areas found in other studies to the variables presented here, a visual check rather than a statistical test must be used.

Figure 2: Area affected by homicides in 2011 (A) Area affected by residential burglaries (B)

(A) (B)

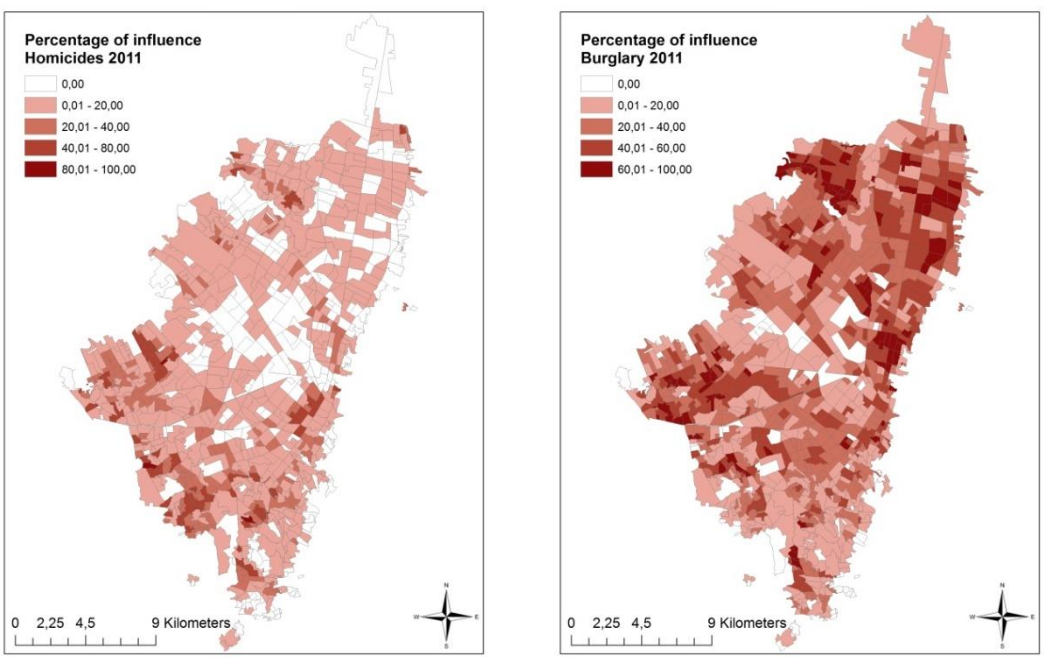


Figure 2, comparatively shows how widespread is residential burglaries compared to homicides. It is also interesting to notice how differently distributed are the more affected areas. Only a few neighborhoods in the South West part of city (Bosa Locality) and North West (Suba Locality) seem

[8] there are nearly 566 censal sectors in Bogotá.(Escobar, 2012; Formisano ,2002)

to share both types of crime. This is more evident in Figure 3 where only the top 10% of neighborhoods in terms of prevalence for both crimes are shown in the map. It can be seen by looking at homicides in figure 2 that a significant part of the top 10% are clustered in certain areas of the city, that include the localities of Bosa, Ciudad Bolivar, and Kennedy. Burglaries on the other hand, also seem to cluster around the northern part of the city in the localities of Suba and Usaquen, although those neighborhoods affected by residential burglaries look more randomly distributed than for homicides.

Figure 3 also helps to infer the existence of a north-south division of the city in terms of homicides and partly on residential burglaries. In order to show this, it is possible to draw a line east to west around the middle part of Bogotá and observe that homicides are more prevalent in the southern part of the city than in the north part. On the other hand, residential burglaries can be found mostly in the north part of the city. However, the amount of burglaries in the south cannot be ignored.

The North/South division of the city seems to explain somewhat the prevalence of homicides, yet this geographic difference by crossing the invisible middle part of the city cannot possibly explain the spatial location of crimes. There are no natural or structural barriers between the south and the north. Whatever is explaining the prevalence of homicides must be stronger in the southern part of the city and it is more relevant around certain neighborhoods in the southwest part of the city. The independent variables will play their part in estimating the influence of crime in each neighborhood. However, there is one interesting additional explanation for the clustered data on homicides or burglaries that does not come from an independent source.

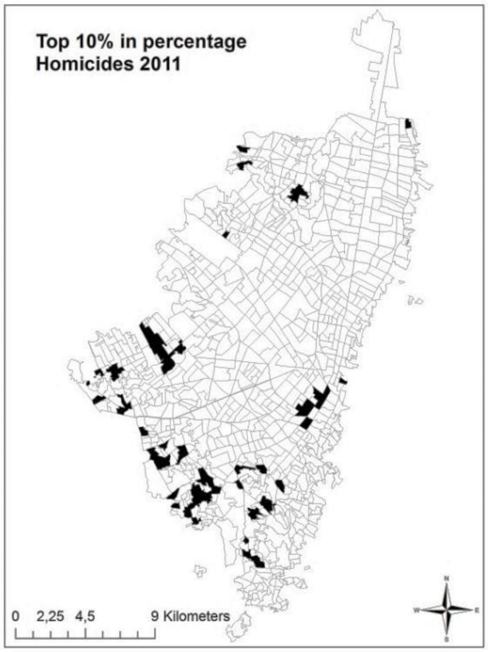
Homicide in the neighboring areas might have some effects on homicides in a specific area. It could also be that there exist an underlying variable that is not observable but because of this variable the neighborhoods seem to cluster together, as an example there is a neighborhood watch whose duty not only deters crime in its neighborhood but also in neighboring areas.

In order to test the importance of neighboring areas in the study of crime, it is necessary to begin with Tobler's (1970) first rule of geography ***"Everything is related to everything else, but near things are more related than distant things"*** it is possible to use the inverse Euclidean distance to measure the relation between neighborhoods. Inverse-distance matrices allow for all places to affect each other. These effects between neighborhoods can also be restricted so areas outside a given radius are specified to have a zero effect. This limits the dependence on more neighbors but increases the probable size of the effect. The restriction can be ultimately reduced to contiguity matrices that only allow contiguous neighbors to affect each other directly.

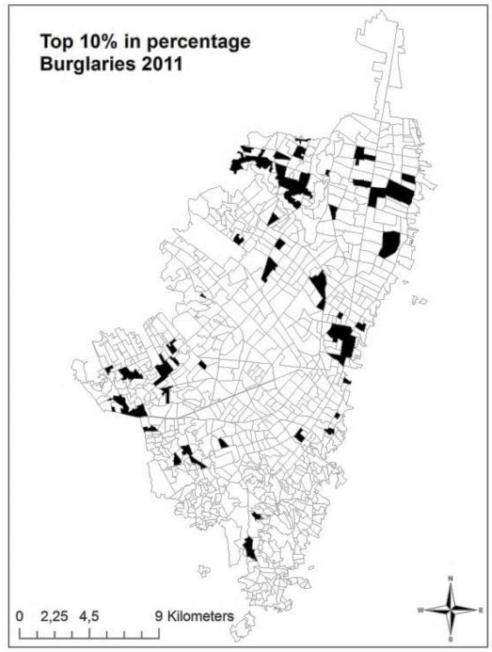
There exists a test for this between neighborhoods correlation. However, it is necessary to establish what type of connection between the neighborhoods is going to be used. This decision needs to be made a priori and there is no straightforward method to decide on it. I will test two different types of relationship between the neighborhoods. First, an inverse-distance relationship that values directly the distance to all neighborhoods. Second, a contiguity relationship that values only the neighbors which share a common border.

Figure 3: Top 10% areas affected by homicides (A) Top 10% areas affected by burglaries (B)

(A)



(B)



I use the Global Moran's I test which measures global spatial autocorrelation. The Moran's Index ranges from -1 to 1, where negative values indicate dispersion and positive values indicate agglomeration. A zero value would indicate a random spatial pattern. The Moran's I can be transformed to Z-scores in order to test the null: random spatial pattern in the data.

Table 2 shows a measure for the persistence of the correlation between neighborhoods and neighboring areas, in both types of crimes (homicides and burglaries) and for two different types of weight matrices (Contiguity and Inverse Distance). Given the Z-scores higher than 20, there is a less than 1% likelihood that these clustered patterns could be the result of random chance. Since the queen contiguity has larger Moran's I than the Inverse distance, for the remainder of the study I will use queen contiguity as the default distance.

The local Indicators of spatial association (LISA) such as the Local Moran I's allow identifying the spatial clustering of neighborhoods where the variable has a similar magnitude. LISA Moran I's are specific to each area, it calculates Z-scores and p-values which are also specific to each observation where the null hypothesis is no spatial clustering at observation i.

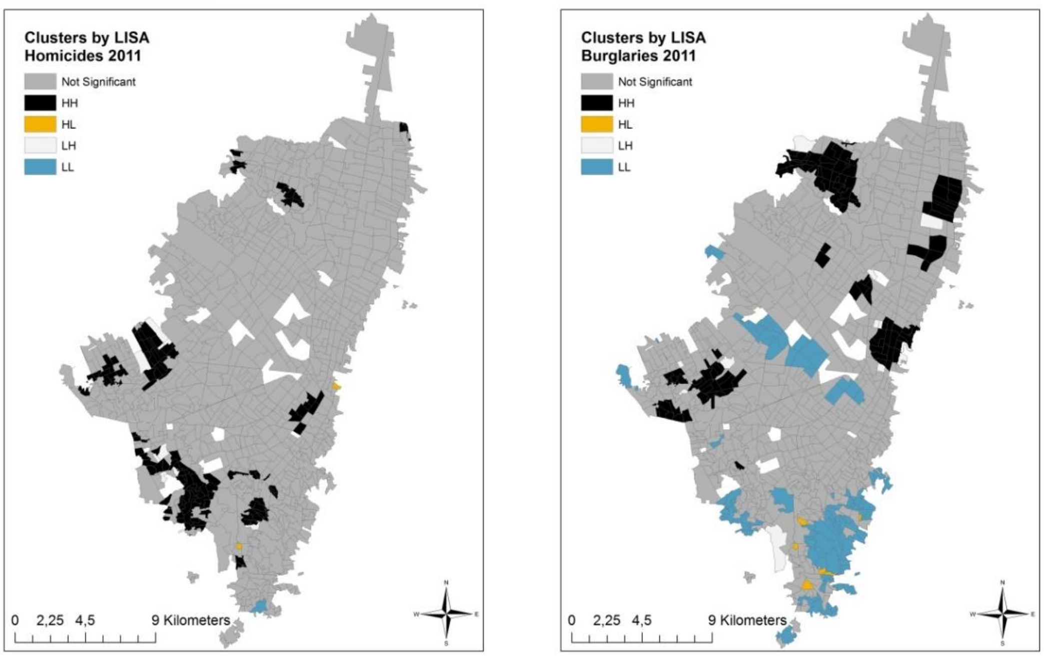
The LISA is useful for showing group of areas where significant spatial autocorrelation exists; nonetheless this is purely descriptive, but can still be interpreted as local pockets of non-stationarity or hot spots. On the homicide data there are evident and significant high-high clusters in the south part of the city, while on the residential burglaries there are mostly high-high clusters in the north part of the city and low-low clusters in the southeast part of the city. The low-low cluster around the middle of the city is where the industrial zone is located; therefore the residential burglaries will be at a minimum by default.

Table 2: Global Moran's I Statistics

weight matrix	queen contiguity[9]	Inverse distance	queen contiguity	Inverse distance
	Homicides		Burglaries	
Moran's Index:	0.429387	0.374499	0.431135	0.372688
Expected Index:	-0.001052	-0.001052	-0.001052	-0.001052
Variance:	0.000432	0.000244	0.00043	0.000243
z-score:	20.717	24.037	20.849	23.976
p-value:	0	0	0	0

Figure 4: LISA clusters, affected by homicides (A) LISA clusters, affected by burglaries (B)

(A) (B)



4. Econometric Approach

My basic regression considers socioeconomic variables that may affect the widespread of homicides and residential burglaries. Then, I extend the basic model with three different specifications. The core model includes as explanatory variables: unemployment rate, percentage

[9] Queen Contiguity is the weight matrix that assigns 1 to contiguous neighborhoods regardless of the size of the common boundary, and 0 to any non-contiguous neighborhood.

of young male population, average years of education, level of real state value inequality, and the average real estate valuation. The other three different specifications are the following. First, I consider agglomeration variables which are related to the amount of residents and the land exploitation, namely: population density, amount of floors, and percentage of constructed area. The second extension adds variables related to both building and land current use in the neighborhood, these variables are: percentage of non-residential area, current non-residential construction, current residential construction and size of the neighborhood. Finally, the third specification adds a component related to the current visual aspect of the buildings that could signal criminals into targeting well conserved properties.

$$Crimevar2011_i = \beta_{k_1} Sociecon_i + \gamma_{k_2} Agglom + \phi_{k_3} Land\ use + \theta_{k_4} visual\ aspect + \epsilon_i \quad (1)$$

The regression presented in equation (1) is in line with most of the empirical literature in their covariates. However, the spatial structure of the crime phenomenon has not been included. In order to check if it is necessary to include any spatial structure I ran an OLS regression and made some diagnostics on both the predicted values and the errors. I found that the inclusion of a spatial structure is necessary for the case of homicides and residential burglaries.[10] In order to explain how to solve the spatial autocorrelation problem, it is necessary to exemplify this by starting from a canonical OLS regression

$$Y_i = X_i \beta + \varepsilon_i \quad (2)$$

If errors are $iid(0, \sigma^2 I_N)$ then β is BLUE. However, in the presence of spatial autocorrelation the variance-covariance matrix $\Sigma = E[\varepsilon' \varepsilon]$ contains N variances and $\frac{N(N-1)}{2}$ off-diagonal parameters that follow a spatial ordering. In order to account for this autocorrelation, it is possible to impose restrictions on Σ and estimate parameters that allow us to characterize the spatial structure of the data. This is done by including spatial lags in the regression model. These spatial lags are obtained as the interaction of a spatial weight matrix W with the vector of observations on a random variable (Y, X or ε).

The spatial weight matrix W is a NxN matrix, representing the spatial relationship among regions. This matrix is such that:

- The relationship of Region i with itself must be equal to zero ($w_{ij} = 0$). This means that the W matrix has zeros in the main diagonal. The explanation behind this is that the geographical distance between region i with itself is zero.
- The relationship of Region i with its neighbor j must be nonzero($w_{ij} \neq 0$). [11] In the matrix W the cell defined by row i and column j includes the value w_{ij}, conversely, row j and column i includes $w_{ij} = w_{ji}$. Then, the matrix W is symmetric.

[10] Appendix 1 includes the OLS regression and diagnostic tests. All Maximum likelihood tests reject the null of spatial independence.
[11] Region j is neighbor of Region i when they have a common border regardless of the border size.

The spatial autocorrelation is then modeled by specifying different functional relationships using the matrix W. This matrix may enter into the Equation affecting either the dependent variable (Y), the independent variables (X), the error term (u), or any combination of the previous ones (Elhorst, 2010). Formally, a model that includes the matrix W affecting all those variables simultaneously is:

$$Y_i = \lambda W Y_i + \theta W X_i + \beta X_i + u \qquad (3)$$
$$u = \rho W u_i + e_i$$

Where Y is one of the two crime variables discussed in Section 3. The first one is the proportion of the neighborhood area affected by homicides. The second one is the proportion of the neighborhood area affected by residential burglary. W is the NxN spatial weight matrix that parameterizes the distance between the N neighborhoods. X is the NxK matrix of observations of K independent variables. Finally u are spatially correlated residuals and e are independent and identically distributed residuals.

Assuming that the matrices $(I_N - \lambda W)$, $(I_N - \rho W)$ are not singular the last two equations boil down to the following reduced form

$$Y = (I_N - \lambda W)^{-1} X\beta + (I_N - \lambda W)^{-1} X W \theta + (I_N - \lambda W)^{-1} (I_N - \rho W)^{-1} \varepsilon \qquad (4)$$

It is possible to always assume that the matrices are not singular when the W matrix is normalized. Following Kelejian and Prucha(2010), and Plumper and Neumayer (2012) I decided to avoid the normally used row normalization in favor of minmax normalization. Unless it is theoretically necessary a row-normalized weight matrix will lead to a misspecified model. Minmax normalization divides the matrix W by one single scalar rather than the row sum for each observation. Therefore, the matrix does not impose the assumption of homogeneous total exposure which in turn does not change the relative relevance for different neighboring regions.

For the sake of simplicity, I will assume $\theta = 0$. This assumption implies no direct relationship between the variables X of neighbors of Region i and the dependent variable of region i (Y_i) . This means that X_j affects Y_i only through Y_j.

This model where $\lambda \neq 0, \rho \neq 0 \; and \; \theta = 0$ is called the SARAR model. It is a model with an AR process in both the dependent variable and the error term. It was studied for the first time by Anselin and Florax (1995).

$$Y = (I_N - \lambda W)^{-1} X\beta + (I_N - \lambda W)^{-1} (I_N - \rho W)^{-1} \varepsilon \qquad (5)$$

Notice that the estimation of equation (5) has two econometric problems that affect each other, namely, endogeneity and heteroscedasticity. The endogeneity problem arises because of the expected value of the error term on Equation (5) is nonzero. This makes the spatial lag to be endogenous, irrespective of ε being $iid(0, \sigma^2 I_N)$. Then, we need to estimate equation (5) using either Maximum likelihood, Bayesian Methods or Instrumental Variables.

The heteroscedasticity problem emerges from the spatial structure of data that prevents the variance-covariance matrix to be a scalar matrix. [12] In a normal OLS regression we would use Huber White robust standard errors to partially control the problems of heteroscedasticity and normality. In the case of a spatial non-linear regression, the spatial structure of the data is controlled by the spatial lags. However, Kelejian and Prucha (2007) suggest a nonparametric heteroscedasticity- autocorrelation consistent estimator of the variance-covariance matrix (SHAC) that partially controls for autocorrelation and heteroscedasticity that are not geographically dependent. This SHAC estimator is currently only available for the IV GMM method introduced by Kelejian and Prucha (1999, 1998, 2004 and 2009) and Arraiz et al.(2009).

The procedure to estimate the spatial components in this method is based on a generalized spatial two stages procedure and it is included in the Stata command *spreg* (Drukker, Prucha, and Raciborski, 2011). This generalized-spatial-two-stage procedure (GS2SLS) generates an estimator that requires instruments. Kelejian and Prucha (1998, 1999) suggest the use of the linearly independent columns of WX, W^2X, W^3X, … as instrument H for the spatial lag on the errors.

A simple way to understand this procedure is to enumerate in a few steps how they derive the estimators.

Step 1a (First Estimation of rho). Use the instrument matrix H to estimate the spatial lag on the error term that will affect in turn the estimator of the spatial lag on the dependent variable.

Step 1b (First estimation of lambda) Takes the residuals from Step 1a to estimate the spatial lag of the dependent variable.

Step 2a (Second –and corrected- estimation of rho) Generalized spatial two-stage least squares (GS2LS) Estimator. They compute a GS2LS estimator of the spatial lag on the residuals. This estimator is defined as the two stages least squares of the normalized version of the equation (5) transformed by Cochrane-Orcutt[13], the only change imposed here is to use the estimator for the spatial lag on the dependent variable obtained from step 1b.

Step 2b (Second –and corrected. Estimation of lambda) Efficient GMM estimator of the spatial lag on the dependent variable based on the GS2LS residuals.

[12] The variance-covariance matrix of equation XX is

$$E(\varepsilon'\varepsilon) = (I_N - \rho W)^{-1}(I_N - \lambda W)^{-1}\, E(\varepsilon'\varepsilon)(I_N - \rho W')^{-1}(I_N - \lambda W')^{-1}$$

[13] The Cochrane-Orcutt transformation adjusts a lineal model for serial correlation on the error term. It is only possible to do the transformation if the errors follow an autoregressive process, assumption that is given by the min-max normalization of the weights matrix that bound the spatial parameter to (-1, 1). Since ρ is not known, then this parameter will be estimated first on the standard model, obtaining the errors and then regressing the errors on the lagged errors (spatially lagged) and finally constructing the quasi-diference.

The procedure generates consistent estimates when the disturbances are heteroscedastic according to Arraiz et al. (2009) and the estimates are still consistent under homoscedasticity therefore there is no apparent loss in using this econometric approach, aside from computational time.

The coefficients in the GS2LS model can only be interpreted for their sign and significance but not for their magnitude. More generally, the coefficients for the independent variables in this spatial regression do not represent the marginal effect of a change in the independent variable on the dependent. This happens because of the simultaneous effect on the area and also on every other neighborhood in the weight matrix that in turn affect the area through the spatial lag.

In order to find the marginal effects for spatial models, Lesage and Pace (2010) discuss and derive two marginal effects that measure the changes in the exogenous variables; one assumption must be made in this case. The predicted values for the dependent variable uses a reduced form extracted from equation (5) that does not take in account the feedback effects on the error.

$$\hat{Y} = E[y|X, W, M] = (I - \lambda W)^{-1} X\beta \tag{6}$$

Consistent with the terminology of Lesage and Pace (2010) the first marginal effect to be considered here is called the average total direct impact (ATDI), which is the change of a delta amount on the independent variable for exactly one neighborhood in the sample at the time. The ATDI can be calculated by using the difference between predicted values, after and before the delta change in the independent variable focused in the area of interest. Lesage and Pace define the corresponding summary measure for ATDI.

$$n^{-1} \sum_{i=1}^{n} \frac{\partial \widehat{Y}_i(X_k + \delta i)}{\partial \delta} = n^{-1} \sum_{i=1}^{n} \frac{\partial \widehat{Y}_i(X_k)}{\partial X_{ik}} \tag{7}$$

Simultaneously changing in one unit the independent variable for all areas yields a different result. The effect of simultaneous change in one delta is called the average total impact (ATI). The ATI can be calculated in the same way as the ATDI. The average difference in these vectors of predicted values is the ATI, where delta is the magnitude of the simultaneous effect on the independent variables.

$$n^{-1} \sum_{i=1}^{n} \frac{\partial \widehat{Y}_i(X_k + \delta e)}{\partial \delta} = n^{-1} \sum_{i=1}^{n} \sum_{r=1}^{n} \frac{\partial \widehat{Y}_i(X_k)}{\partial X_{rk}} \tag{8}$$

In most of the recent literature The ATI has been simply called the Total Effect meanwhile the ATDI has taken the name of Direct Effect. Consequently, the difference between the total effect and the direct effect is named the Indirect Effect.

Elhorst (2010) notes that for the SAR model all coefficients β^k are multiplied by the same matrix $(I_n - \lambda W)^{-1}$ this means the ratio between indirect and direct effects is the same for all covariates. The sign of the marginal effects will depend on the coefficient for each explanatory variable. An implication of the SAR model is that all the marginal effects will have the same sign. In the case at hand the direct effect will be smaller than the total effect because the direct effect does not incorporate the feedback effects of having all neighborhoods implement the change simultaneously as does the total effect.

The resulting econometric approach described here differs significantly to those in previous studies for Bogota. Not only is different in the crime statistic to use and the neighborhood characteristics included, but also is different in the spatial approach and the differential effects it finds. By going further than local indicators of spatial association and geographically weighted regression this study potentially derives a more refined relation of correlation between neighborhood crime and contiguous neighborhoods. However, the potential shortcoming of the dependent variable being size dependent cannot be ignored, and explains the inclusion of the neighborhood size which, at least, partially controls for this issue.

Main Results

The estimates from the SARAR model are shown in Table 3 for the homicides and in Table 4 for the residential burglaries. Given the nonlinearity of the models, it is not possible to interpret the regression coefficients as marginal effects but it is still possible to interpret the sign of the coefficient as in other nonlinear models such as Probit, Logit and etc.

In model 1, I include socio-economic variables. In model 2, I add the variables related to the agglomeration process. Model 3 and 4 add respectively variables for current use of the land and the proxy for physical deterioration. First, all the models consistently show that the dependent spatial lag is significant and positive which lead us to conclude that there is a geographical concentration in both types of crimes. These results are consistent with the findings from previous studies using spatial autocorrelation analysis. Crime does not occur at random and it occurs with a spatial pattern as seen in the ESDA section.

Regarding the socioeconomic variables that are first introduced in model 1, the results for homicides are consistent in the percentage of young male population, the years of education and the inequality measure. The percentage of young male population is positively and significantly correlated with homicides, the average years of education is negative and significant, and the Gini coefficient is positive and significant. The results found are in the same line as the ones in Lochner (1999) who asserts that:

> *"Crime is primarily a problem among young uneducated men. Individuals with low skill level are more likely to participate in criminal activities because the returns they can earn from work or school are low. Both high school graduation and ability directly lower criminal propensities"* (p. 34)

After including the socio-economic variables and inequality, I add three agglomeration variables that are significant in all the models when included, I find that an increase in the population density, amount of floors, and percentage of constructed area show a significant positive signal.

Increases in population density are related to higher crime rates and are consistent with Glaeser and Sacerdote (1996).

Model 3 introduces some land-use variables and includes the size of the neighborhood as a control variable. The percentage of area affected by homicides does not seem to be correlated with current constructions or the size of their non-residential areas. The coefficient of the size of the neighborhood is negative and significant which would suggest that smaller neighborhoods are more prone to crime. However, this could be the result of how the dependent variable is constructed, given that one isolated homicide represents a higher percentage of crime-ridden area for a small neighborhood than a larger one.

Finally, Model 4 includes one last final variable, the percentage of buildings rated as having bad conserved finishings. This variable is intended to play a role for the visual aspect and the current state of investment in those neighborhoods. In the case of Homicides this aspect seems to be of non-importance.

Table 3: Spatial regressions using the min-max neighbors matrix on Homicides

Variables	(1)	(2)	(3)	(4)
Dependent spatial lag	0.794***	0.862***	0.975***	0.962***
	(0.275)	(0.195)	(0.207)	(0.206)
Unemployment rate in 2005	-0.157	-0.133	-0.183	-0.191
	(0.187)	(0.156)	(0.150)	(0.150)
Percentage of young male population aged 10-25 in 2005	0.530***	0.264***	0.258***	0.241**
	(0.135)	(0.094)	(0.094)	(0.097)
Average years of education in 2005	-0.016***	-0.018***	-0.017***	-0.015***
	(0.003)	(0.002)	(0.002)	(0.003)
Gini coefficient for the real estate valuations	0.005	0.103***	0.104***	0.100***
	(0.036)	(0.031)	(0.033)	(0.033)
Real estate valuation per constructed m2	-0.002***	-0.001**	-0.000	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)
Percentage of buildings area		0.177***	0.171***	0.166***
		(0.045)	(0.055)	(0.056)
Population density		2.843***	2.632***	2.600***
		(0.672)	(0.849)	(0.854)
Amount of floors		0.012**	0.011*	0.013**
		(0.006)	(0.006)	(0.006)
Percentage of non-residential area			-0.018	-0.016
			(0.027)	(0.027)
Dummy for current non-residential construction			0.028	0.028
			(0.020)	(0.020)
Dummy for current residential construction			0.001	0.002
			(0.012)	(0.012)

Neighborhood area			-0.021***	-0.020***
			(0.007)	(0.007)
Percentage of bad conserved finishings				0.020
				(0.020)
Constant	0.076***	-0.022	-0.009	-0.026
	(0.025)	(0.021)	(0.023)	(0.027)
Error spatial lag	1.045***	0.734***	0.615**	0.618**
	(0.306)	(0.249)	(0.264)	(0.263)
Observations	952	952	952	952

Standard errors in parentheses

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

Source: own calculations

Residential Burglaries, as seen in the ESDA section, have a completely different spatial pattern than Homicides. Residential Burglaries are also different in the correlation to the variables presented over the 4 models introduced above.

The significant socio-economic variables correlated to the residential burglaries include the percentage of young males, the average years of education and the real estate value per constructed meter. But, both years of education and young males have the opposite sign to what is found for the homicide data. It is interesting that higher amount of years of education are positively correlated to residential burglaries and higher percentage of young male population actually decreases burglaries. The real estate valuation per m2 is also positive and significant while the inequality measure is non-significant.

The significance and the signs in these socio-economic variables suggest that the economic gains to be obtained in those neighborhoods are larger. Therefore, those areas are more prone to burglaries given they are more educated, have more experience and accumulated goods (smaller percentage of young males), and their constructions are more valuable, in summary those areas are more likely wealthier than the rest and are subject to more burglaries.

The Agglomeration variables are also significant and positive for the residential burglaries as in the homicide data. The more population a place holds the larger the probability of a crime occurring in the neighborhood. It could be a problem of “crime premium” as noted in Glaeser and Sacerdote (1996). “*Theft becomes easier as the potential criminal’s environment becomes more densely populated with victims”.*

Land-use variables introduced in model 3 for the case of residential burglaries are treated more like controls than potential explanatory covariates. The percentage of non-residential buildings in the area and the neighborhood size are logically negative and potentially significant. The last variable introduced in model 4, the percentage of bad conserved finishings, shows an important visual effect that could signals criminals to avoid areas that combines (old buildings, low investment and maintenance, and poor materials).

Table 4: Spatial regressions using the min-max neighbors matrix on residential burglaries

Variables	(1)	(2)	(3)	(4)
Dependent spatial lag	0.528***	0.568***	0.744***	0.776***
	(0.158)	(0.110)	(0.115)	(0.115)
Unemployment rate in 2005	0.200	0.244	0.083	0.108
	(0.285)	(0.203)	(0.196)	(0.198)
Percentage of young male population aged 10-25 in 2005	0.089	-0.359***	-0.314***	-0.249**
	(0.122)	(0.109)	(0.107)	(0.110)
Average years of education in 2005	0.021***	0.018***	0.017***	0.012***
	(0.003)	(0.003)	(0.003)	(0.004)
Gini coefficient for the real estate valuations	-0.078	0.071	0.047	0.059
	(0.060)	(0.049)	(0.051)	(0.051)
Real estate valuation per constructed m2	-0.001*	0.001**	0.002***	0.001**
	(0.001)	(0.001)	(0.001)	(0.001)
Percentage of buildings area		0.243***	0.311***	0.326***
		(0.057)	(0.063)	(0.063)
Population density		5.364***	3.962***	4.097***
		(0.676)	(0.762)	(0.780)
Amount of floors		0.025**	0.027**	0.021*
		(0.011)	(0.011)	(0.011)
Percentage of nonresidential area			-0.117***	-0.123***
			(0.031)	(0.031)
Dummy for current non-residential construction			-0.036	-0.037
			(0.023)	(0.023)
Dummy for current residential construction			-0.003	-0.008
			(0.014)	(0.014)
Neighborhood area			-0.020	-0.025*
			(0.014)	(0.014)
Percentage of bad conserved finishings				-0.072***
				(0.028)
Constant	0.084***	-0.079***	-0.041	0.018
	(0.026)	(0.028)	(0.030)	(0.039)
Error spatial lag	1.384***	1.066***	0.869***	0.809***
	(0.162)	(0.141)	(0.154)	(0.160)
Observations	952	952	952	952

Standard errors in parentheses

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

Source: own calculations

The regressions above show that different type of crimes does not necessarily happen in the same neighborhoods, moreover socioeconomic factors are in stark contrast in explaining homicides and residential burglaries. In common, both type of crimes occur in more densely populated areas and suffer from a spatial diffusion process. It is interesting that several variables remain significant when including the spatial structure of crime, both dependent lag and errors lag. This could potentially mean that is not only undetermined hot spots what matters for crime, but rather is hot spots and a set of neighborhood characteristics such as education, age, density and income distribution which determines the reach of crime for a neighborhood.

Only looking at the coefficient does not give us a clear picture of the magnitude of every covariate on each crime. Therefore we turn to the marginal effects devised in the spatial econometric literature.

4.1. Marginal Effects

Using the Marginal effects described in the Econometric Approach, which are based on Lesage and Page (2010). I calculate all three different effects for both types of crimes. I use increases depending on the sample standard deviation. The increases showed here are of 10% and 50% in the standard deviation for each neighborhood.[14]

Table 5 contains the marginal effects for an increase in 10% of the standard deviation (S.D) in each significant variable on model 4 for homicides. Table 6 contains the marginal effects for an increase in half a standard deviation.

Simultaneously increasing by 10%S.D the percentage of young male population in all neighborhoods increases in 0.21 pp (percentage points) the average predicted area of influence in homicide, this effect is the total marginal effect and it includes the feedback effects from other neighboring areas. When only taking in account the average direct effect the increase in young male population accounts for a 0.14 pp increase in homicide coverage. The indirect effect is then calculated as the difference between total and direct effects, it accounts for 0.07pp which can be taken as the average spillover effect from the neighbors having an increase of 10%S.D in their young male population.

On the average years of education, an increase of 10%S.D for all the population in the neighborhood yields a total average effect of -0.6pp. The average direct effect amounts to -0.4pp and the indirect effect is -0.2pp.

An increase in 10% S.D of the inequality Gini index on the real estate valuation is correlated with a 0.18pp increase in homicide coverage, with 0.12pp being an average direct impact.

For the agglomeration variables that are all significant in the case of homicide coverage. I found that an increase of 10% S.D of constructed area is correlated with an increase of 0.39pp as the total effect and 0.26pp in direct effect. An additional level in the average amount of floors lead to a total average increase of 0.14pp and an average direct impact of 0.09pp and the increase of 10%S.D more people per square meter leads to an increase of over 0.61pp in total effect and

[14] Increases of 3% and 1 standard deviation are also available in the appendix

nearly 0.41pp in average direct effect. Table 6 contains the increase of half standard deviation and the effects found are in the same line as the ones in table 5.

Table 5: Marginal effects of a 0.10*Standard Deviation, using the min-max neighbors weight matrix on Homicides

Variables	Marginal effects		
	Direct	Indirect	Total
Percentage of young male population aged 10-25 in 2005	0.0014	0.0007	0.0021
Average years of education in 2005	-0.004	-0.002	-0.006
Gini coefficient for the real state valuations	0.0012	0.0006	0.0018
Percentage of buildings área	0.0026	0.0013	0.0039
Population density	0.0041	0.002	0.0061
Amount of floors	0.0009	0.0005	0.0014
Neighborhood área	-0.0009	-0.0004	-0.0013

Source: own calculations

Table 6: Marginal effects of a 0.50*Standard Deviation, using the min-max neighbors weight matrix on Homicides

Variables	Marginal effects		
	Direct	Indirect	Total
Percentage of young male population aged 10-25 in 2005	0.0070	0.0034	0.0104
Average years of education in 2005	-0.0202	-0.0098	-0.0300
Gini coefficient for the real state valuations	0.0059	0.0029	0.0088
Percentage of buildings área	0.0130	0.0063	0.0193
Population density	0.0205	0.01	0.0305
Amount of floors	0.0046	0.0022	0.0068
Neighborhood área	-0.0045	-0.002	-0.0065

Source: own calculations

Table 7 contains the marginal effects for an increase in 10% of the standard deviation (S.D) in each significant variable on model 4 for residential burglaries. Table 8 contains the marginal effects for an increase in half a standard deviation.

For the case of residential burglaries, the marginal effect of an increase of the young male population in 10% S.D is a reduction of 0.19pp in the total effect and 0.14pp in the direct effect. Meanwhile, an increase of a 10%S.D in the average education years of the population leads to an increase of 0.41pp in the total effect and 0.30pp in the average direct effect. In the case of residential burglaries the value in millions per square meter is also positively correlated, an increase of 10% S.D in the real estate valuation increases in 0.07pp the total effect and 0.05pp the average direct effect.

The agglomeration variables are all positive for the case of the influence of residential burglaries. An increase of 10% S.D in the amount of buildings in all the neighborhoods is correlated with a total effect of 0.69pp and 0.51pp in direct effect. An increase in 10%S.D in the average amount of floors lead to a total impact of 0.20pp and an average direct impact of 0.15pp and the increase of 10% S.D more people per square meter leads to an increase of over 0.87pp in total effect and nearly 0.64pp in average direct effect.

Finally, an increase in the percentage of bad conserved finishings by 10% S.D leads to a deterrence for residential burglaries with a total effect of -0.33pp and an average direct effect of -0.25pp. Table 8 includes the marginal effect of an increase of half a standard deviation. The results are in the same line as the increase of a 10%S.D.

Table 7: Marginal effects of an increase of 0.10*Standard Deviation using the min-max neighbors weight matrix on Residential burglaries

Variables	Marginal effects		
	Direct	Indirect	Total
Percentage of young male population aged 10-25 in 2005	-0.0014	-0.0005	-0.0019
Average years of education in 2005	0.0030	0.0011	0.0041
Real state valuation per constructed km2	0.0005	0.0002	0.0007
Percentage of buildings área	0.0051	0.0018	0.0069
Population density	0.0064	0.0023	0.0087
Amount of floors	0.0015	0.0005	0.0020
Percentage of nonresidential área	-0.0027	-0.0010	-0.0037
Neighborhood área	-0.0011	-0.0004	-0.0015
Percentage of bad conserved finishings	-0.0025	-0.0008	-0.0033

Source: own calculations

Table 8: Marginal effects of an increase of a 0.50*Standard Deviation using the min-max neighbors weight matrix on Residential burglaries

Variables	Marginal effects		
	Direct	Indirect	Total
Percentage of young male population aged 10-25 in 2005	-0.007	-0.003	-0.010
Average years of education in 2005	0.015	0.006	0.021
Real state valuation per constructed km2	0.003	0.001	0.004
Percentage of buildings área	0.026	0.009	0.035
Population density	0.032	0.012	0.044
Amount of floors	0.008	0.003	0.010
Percentage of nonresidential área	-0.014	-0.005	-0.019
Neighborhood área	-0.006	-0.002	-0.008
Percentage of bad conserved finishings	-0.013	-0.004	-0.017

Source: own calculations

Table 6 and 8 both show an increase of half a standard deviation for each neighborhood characteristic. In order to compare which covariate has a higher impact I divide the marginal effect by the average of the dependent variable. This comparison allows us to infer which significant characteristic has a larger effect between types of crimes.

Table 9 includes the percentage that represents an increase of half a standard deviation on the dependent variables. It is noticeable that the effects represent a larger percentage for homicides than residential burglaries. Education has a large effect in reducing homicide influence. By comparison, the inducing effect of education on residential burglaries is almost four times smaller in magnitude. Density measures such as the percentage of buildings area and population density also have a large crime inducing effect.

Table 9: Representing the increase of half a standard deviation as percentage of the dependent variable

Variables	Effect / Homicide		Effect / residential burglaries	
	Direct	Total	Direct	Total
Percentage of young male population aged 10-25 in 2005	5.47%	8.12%	-2.31%	-3.13%
Average years of education in 2005	-15.82%	-23.49%	4.85%	6.58%
Gini coefficient for the real state valuations	4.63%	6.87%		
Percentage of buildings área	10.19%	15.13%	8.19%	11.11%
Population density	16.08%	23.88%	10.38%	14.08%
Amount of floors	3.57%	5.30%	2.40%	3.26%
Neighborhood área	-3.46%	-5.13%	-1.81%	-2.46%
Real estate valuation per m2			0.82%	1.12%
Percentage of nonresidential buildings			-4.40%	-5.97%
Percentage of bad conserved finishings			-3.98%	-5.40%

Source: own calculations

5. Final Remarks

This paper use a spatial econometric method that controls for the spatial structure identified for the crime phenomenon in Bogotá, Colombia. It uses cadastral sectors as a neighborhood measure and includes socioeconomic, agglomeration, land use and visual aspect variables as covariates. It uses two different measures for crime, one is violent crime represented in homicides and the other one is property crime represented in residential burglaries. The final specification used here shows that both the spatial structure and several neighborhood characteristics are significant in explaining the spatial incidence of crimes. In the case of neighborhood characteristics, there is a significant correlation between both socio-economic and agglomeration variables with crime. However, homicides and residential burglaries have a different relation with education, age and income distribution variables in both sign and magnitude. Agglomeration variables on the other hand have the same positive sign leading to the hypothesis that more density means more crime irrespective of the type.

Interestingly enough when dividing the effect of significant covariates between direct, indirect and total effect and comparing similar increases on the covariates. It is possible to see that similar increases on the covariates have larger impact on homicides than residential burglaries, which reinforces the positive driver behind education attainment. Education reduces homicide more than induces property crime through income gains. It is also possible to see the relevance of population density and other agglomeration variables and the hypothesis that more densely populated places attract criminals. Criminals seem to lower their probability of being caught and enhance the probability of finding a vulnerable victim. Both of these findings could be important tools in improving the current tactics of fighting crime in Bogotá.

Finally, it is important to stress that this study have two important shortcomings and therefore should be treated as a characterization rather than a full impact evaluation. First, the use of a non-standard measure of crime makes it non comparable to most studies in the literature, so it will be interesting to either validate the approach presented here or use usual crime statistics at the cadastral sector as unit of analysis. Second, this paper is a still picture of crime in Bogotá for 2011 and it would be interesting to include a time dynamic since crime trends shows interesting patterns of crime for certain periods of time.

Bibliography

Anselin, L., & Florax, R. J. (Eds.). (1995). New Directions in Spatial Econometrics. Springer Verlag.

Arraiz, I., Drukker, D. M., Kelejian, H. H., & Prucha, I. R. (2009). A SPATIAL CLIFF-ORD-TYPE MODEL WITH HETEROSKEDASTIC INNOVATIONS: SMALL AND LARGE SAMPLE RESULTS*. Journal of Regional Science, 50(2), 592-614.

Becker, C. (1968). Punishment: An Economic Approach, 76J. Pol. Econ, 169(10.2307), 1830482169.

Brush, J. (2007). Does income inequality lead to more crime? A comparison of cross-sectional and time-series analyses of United States counties. Economics letters, 96(2), 264-268.

Bourguignon. F., Nuñez, J. & Sanchez, F. (2003) A STRUCTURAL MODEL OF CRIME AND INEQUALITY IN COLOMBIA. Journal of the European Economic Association, *1*(2-3), 440-449.

Buonanno, P., & Leonida, L. (2006). Education and crime: evidence from Italian regions. Applied Economics Letters, 13(11), 709-713.

Choe, J. (2008). Income inequality and crime in the United States. Economics Letters, 101(1), 31-33.

Cohen, J., & Tita, G. (1999). Diffusion in homicide: Exploring a general method for detecting spatial diffusion processes. Journal of Quantitative Criminology, 15(4), 451-493.

Drukker, D. M., Prucha, I. R., & Raciborski, R. (2011). A command for estimating spatial-autoregressive models with spatial-autoregressive disturbances and additional endogenous variables. The Stata Journal, 1(1), 1-13.

Echandía, C.,Escobedo, R., & Rubio, M. (2001). Violencia homicida en Bogotá: más que intolerancia. Centro de Estudios sobre Desarrollo Economico (CEDE), Universidad de Los Andes.

Ehrlich, I. (1973). Participation in illegitimate activities: A theoretical and empirical investigation. The Journal of Political Economy, 521-565.

Ehrlich, I. (1975). On the relation between education and crime. In Education, income, and human behavior (pp. 313-338). National Bureau of Economic Research.

Elhorst, J. P. (2010). Applied spatial econometrics: raising the bar. Spatial Economic Analysis, 5(1), 9-28.

Escobar, G. (2012). El uso de la teoría de la desorganización social para comprender la distribución de homicidios en Bogotá, Colombia. Revista INVI, 27(74), 21-85.

Fajnzylber, P., Lederman, D., & Loayza, N. (2002). What causes violent crime?. European Economic Review, 46(7), 1323-1357.

Formisano, M. (2002). Econometría espacial: características de la violencia homicida en Bogotá. CEDE.

Freeman, R. B. (1991). Crime and the employment of disadvantaged youths (No. w3875). National Bureau of Economic Research.

Freeman, R. B. (1994). Crime and the job market (No. w4910). National Bureau of Economic Research.

Freeman, R. B. (1996). Why do so many young American men commit crimes and what might we do about it? (No. w5451). National Bureau of Economic Research.

Gallipoli, G., & Fella, G. (2006). Education and Crime over the Lifecycle (No. 192). Society for Computational Economics.

Gaviria, A., & Pagés, C. (2002). Patterns of crime victimization in Latin American cities. Journal of Development Economics, 67(1), 181-203.

Gaviria, A., Medina, C., & Tamayo, J. A. (2010). Assessing the link between adolescent fertility and urban crime. Borradores de Economía, 594.

Glaeser, E. L., & Sacerdote, B. (1999). Why Is There More Crime in Cities?. Journal of Political Economy, 107(S6).

Glaeser, E. L., Sacerdote, B., & Scheinkman, J. A. (1996). Crime and Social Interactions. The Quarterly Journal of Economics, 111(2), 507-48.

Grogger, J. (1995). The effect of arrests on the employment and earnings of young men. The Quarterly Journal of Economics, 51-71.

Grogger, J. (1998). Market Wages and Youth Crime. Journal of Labor Economics, 16(4), 756-791.

Grogger, J. T. (1994). Criminal opportunities, youth crime and young men's labor supply. Department of Economics, University of California.

İmrohoroğlu, A., Merlo, A., & Rupert, P. (2000). On the political economy of income redistribution and crime. International Economic Review, 41(1), 1-26.

Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. The Journal of Real Estate Finance and Economics, 17(1), 99-121.

Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. International economic review, 40(2), 509-533.

Kelejian, H. H., & Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. Journal of Econometrics, 118(1), 27-50.

Kelejian, H. H., & Prucha, I. R. (2009). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics, 157(1), 53-67.

Kelly, M. (2000). Inequality and crime. Review of Economics and Statistics, 82(4), 530-539.

LeSage, J. P., & Pace, R. K. (2010). Spatial econometric models. In Handbook of Applied Spatial Analysis (pp. 355-376). Springer Berlin Heidelberg.

Llorente, M. V., Escobedo, R., Echandía, C., & Rubio, M. (2001). Violencia homicida y estructuras criminales en Bogotá. Análisis Político, 44, 17-38.

Lochner, L. (1999). Education, work, and crime: theory and evidence. Rochester center for economic research working paper, (465).

Lochner, L., & Moretti, E. (2004). The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. American Economic Review, 155-189.

Masciandaro, D., & Ruozi, R. (1999). Mercati e illegalità: economia e rischio criminalità in Italia (Vol. 11). EGEA.

Menezes, T., Silveira-Neto, R., Monteiro, C., & Ratton, J. L. (2013). Spatial correlation between homicide rates and inequality: Evidence from urban neighborhoods. Economics Letters, 120(1), 97-99.

Neumayer, E. (2005). Inequality and violent crime: Evidence from data on robbery and violent theft. Journal of Peace Research, 42(1), 101-112.

Plümper, T., Neumayer, Eric., **W** (2012). EPSA 2013 Annual General Conference Paper 165.

Rubio, M. (1997). Criminalidad Urbana en Colombia. Mimeo online. Bogotá: Centro de Estudios sobre Desarrollo Economico (CEDE) , Universidad de Los Andes.

Rubio, M. (2000). Caracterización de la violencia homicida en Bogotá. Mimeo online. Bogotá: Alcaldía de Bogotá y Paz Pública-CEDE-Universidad de Los Andes.

Sánchez, F., Espinosa, S., & Rivas, Á. (2003). Garrote o zanahoria?: factores asociados a la disminución de la violencia homicida y el crimen en Bogotá, 1993-2002. CEDE.

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place*. Criminology, 27(1), 27-56.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic geography, 234-240.

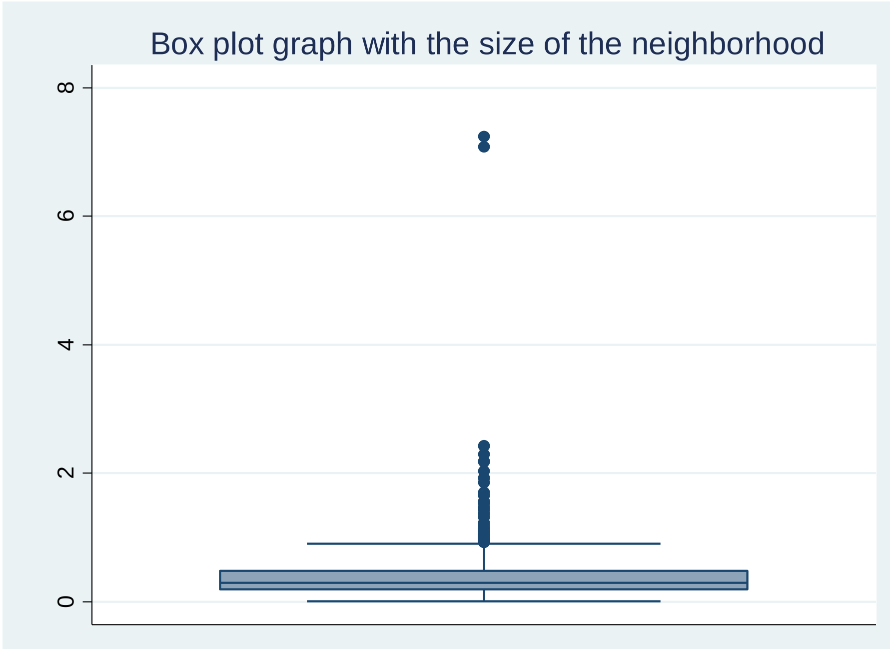
United Nations Development Programee. (2014). Regional human development report 2013-2014: Citizen Security with a Human Face: Evidence and Proposals for Latin America.

Usher, D. (1997). Education as a Deterrent to Crime. Canadian Journal of Economics, 367-384.

World Bank (2010). World Investment and Political Risk Report. Washington, DC

Appendix

1. Box plot of neighborhood sizes



2. OLS regression on both crime variables and diagnostic tests on spatial error and spatial lag

Table 10: Linear Regression for all independent variables

Variables	homicide	residential burglaries
Unemployment rate in 2005	-0.163	-0.098
	(0.155)	(0.204)
Percentage of young male population aged 10-25 in 2005	0.283***	-0.245**
	(0.082)	(0.111)
Average years of education in 2005	-0.020***	0.015***
	(0.003)	(0.004)
Gini coefficient for the real estate valuations	0.127***	0.121**
	(0.035)	(0.050)
Real estate valuation per constructed m2	-0.001	0.001
	(0.000)	(0.001)
Percentage of buildings área	0.193***	0.306***
	(0.055)	(0.063)
Population density	3.041***	5.839***
	(0.735)	(0.740)
Amount of floors	0.028***	0.023**
	(0.007)	(0.010)
Percentage of nonresidential área	0.007	-0.077**
	(0.027)	(0.032)
Dummy for current non-residential construction	0.032	-0.052**
	(0.022)	(0.025)
Dummy for current residential construction	0.006	0.007
	(0.013)	(0.014)
Neighborhood área	-0.003	-0.013
	(0.006)	(0.011)

Percentage of bad conserved finishings	0.055***	-0.114***
	(0.020)	(0.029)
Constant	-0.055*	0.033
	(0.028)	(0.041)
Observations	952	952
R-squared	0.333	0.409

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Source: own calculations

Diagnostics on the queen matrix after OLS regression

Test	Statistic	df	p-value
Spatial error:			
Moran's I	38.01	1	0
Lagrange multiplier	162.122	1	0
Robust Lagrange multiplier	47.867	1	0
Spatial lag:			
Lagrange multiplier	134.44	1	0
Robust Lagrange multiplier	20.185	1	0

Marginal effects of a 0.30*Standard Errors, using the min-max neighbors weight matrix on Homicides

Variables	Marginal effects		
	Direct	Indirect	Total
percentage of young male population aged 10-25 in 2005	0.0042	0,002	0.0062
average years of education in 2005	-0.0121	-0,0059	-0.0180
gini coefficient for the real state valuations	0.0035	0,0018	0.0053
percentage of buildings area	0.0078	0,0038	0.0116
population density	0.0123	0,006	0.0183

Marginal effects of a 1 Standard Errors, using the min-max neighbors weight matrix on Homicides

Variables	Marginal effects		
	Direct	Indirect	Total
percentage of young male population aged 10-25 in 2005	0.0138	0,0067	0.0205
average years of education in 2005	-0.0400	-0,0194	-0.0594
gini coefficient for the real state valuations	0.0117	0,0057	0.0174
percentage of buildings area	0.0258	0,0124	0.0382
population density	0.0407	0,0196	0.0603

amount of floors	0.0027	0,0014	0.0041	amount of floors	0.0090	0,0044	0.0134
neighborhood area	-0,0027	-0,0012	-0,0039	neighborhood area	-0,009	-0,004	-0,013
Observations	952	952	952	Observations	952	952	952

Source: own calculations

Source: own calculations

Marginal effects of a 0.30*Standard Errors, using the min-max neighbors weight matrix on residential burglaries

Variables	Marginal effects		
	Direct	Indirect	Total
percentage of young male population aged 10-25 in 2005	-0,0042	-0,0015	-0,0057
average years of education in 2005	0,0090	0,0033	0,0123
real state valuation per constructed km2	0,0015	0,0006	0,0021
percentage of buildings area	0,0153	0,0054	0,0207
population density	0,0192	0,0069	0,0261
amount of floors	0,0045	0,0015	0,0060
percentage of nonresidential area	-0,0081	-0,0030	-0,0111
neighborhood area	-0,0033	-0,0012	-0,0045
percentage of bad conservated finishings	-0,0075	-0,0024	-0,0099

Source: own calculations

Marginal effects of a 1 Standard Error, using the min-max neighbors weight matrix on residential burglaries

Variables	Marginal effects		
	Direct	Indirect	Total
percentage of young male population aged 10-25 in 2005	-0,014	-0,005	-0,019
average years of education in 2005	0,030	0,011	0,041
real state valuation per constructed km2	0,005	0,002	0,007
percentage of buildings area	0,051	0,018	0,069
population density	0,064	0,023	0,087
amount of floors	0,015	0,005	0,020
percentage of nonresidential area	-0,027	-0,010	-0,037
neighborhood area	-0,011	-0,004	-0,015
percentage of bad conservated finishings	-0,025	-0,008	-0,033

Source: own calculations