Homicide Occurrence and its Spatial Diffusion
Mexico as case of study

A dissertation submitted in partial fulfillment for the award of degree of

Master in Science

in

Spatial Transport and Environmental Economics

by

José Daniel García Hernández (2643574)

Under the supervision of

Ph.D. Wim Bernasco

SCHOOL OF BUSINESS AND ECONOMICS
VRIJE UNIVERSITEIT AMSTERDAM

August, 2021

Acknowledgement

The last months have been tough for everyone, and writing a thesis on homicide far away from the homeland makes it to realize how difficult was to reach this point. I am glad to be part of la familia García Hernandez, without them I just wouldn't be here. There were, are, and will be the most important aspect in my life, and knowing that they exist makes me know that there is something to fight for in our beautiful but particular Mexico/world.

I have no words to thank my second and international family: flat 46 and my fellow students Aldo and Faycal. Thanks to the VU Amsterdam for letting me be part of this amazing community. Thanks to my supervisor who had let me be part of this amazing experience and feel that I can contribute to change things in my country.

Finally, to my country and all what it represents: my people, food, music, kindness, friendship and all the skill it gave to face this amazing experience far from homeland, but never far from home.

Abstract

This thesis explores the impact of demographic, criminal, economic, political, time, and spatial aspects that affect the occurrence of homicide in Mexico. We use yearly data for the period 2015-2020 coming mainly from three different official sources and make a balanced panel for the entire period. Initially, we tested an OLS model per year, secondly, we compare the results of the pooled model against the fixed-effect and random-effects models. Thirdly, we looked for local and global spatial autocorrelation and tested the hypothesis of spatial dispersion of homicide through the spatial lag and spatial error models for the entire country and for organized crime regions. Finally, we tested the hypothesis of different realities in Mexico and how the occurrence of homicide is explained. We found strong evidence for non-random dispersion of homicide, as well as different spatial models for different organized crime regions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

With more than 130 million inhabitants distributed in 1,964,375 km^2[1](almost 48 times the territory of the Netherlands[2]), Mexico presents a complex situation in the evolution of crime. This can be reflected in more than 73 thousand disappeared people, of which 97% occurred during the 2006-2019 period[3], due to the war against drug cartels that started in 2006. Those disappeared people are differently distributed across the country and that distribution has changed over the years. Moreover, an increasing crime occurrence shown in the 2,585 murders in March 2020 derived from the war against drug cartels (the highest number since this indicator is measured) has brought a wave of increasing occurrence on crimes initially not related to drug cartels or homicides/disappearing. It presents different faces such as extortion against businesses, drug possession, and homicide (Estévez-Soto, 2020; C. J. Vilalta, 2010; C. Vilalta & Muggah, 2014). In general, in 2018 it was estimated that 37,807 crimes were committed per 100,000 inhabitants, on the national level (INEGI, 2019). It is essential to think that all these crimes are related in some way to each other; specifically, this thesis focuses on homicide and its relationship with other covariates which describes the Mexican context, such as regional, socioeconomic and other crimes.

1.2 Problem statement

The national problem presents different patterns depending on the region and different types of crimes occurring over time. Being a developing country implies a lack of money and scientific research on these phenomena and it could be reflected in inefficient strategies to fight against crime. For instance, Hobbs and Antonopoulo (2014) suggests that the

[1]INEGI. Anuario estadístico y geográfico de los Estados Unidos Mexicanos 2017. Retrieved on 28/12/2020. https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825097912

[2]Unite Nations Statistics. Retrieved on 28/12/2020. http://data.un.org/en/iso/nl.html

[3]According to the media release issued by the Mexican government in July 2020. https://www.gob.mx/segob/prensa/gobernacion-y-la-cnb-presentan-el-informe-relativo-a-la-busqueda-identificacion-y-registro-de-personas-desaparecidas-y-no-localizadas

biggest and most important city of Mexico in terms of population and economy (Mexico's City) did not have a strategy on how to distribute police officers within the city (in terms of arrests for drugs possession). That (nonexistent) strategy should be taken into account alongside urban planning and social development in order to find solutions to crime. This need for better policing within the city could be similar to the need for better national policing, in the form of a lack of strategy in the distribution of army or federal police within the whole country for fighting against organized crime. Even when a research of the highest quality does not necessarily imply a successful strategy and vice versa, research around this topic is needed to better understand the phenomenon of crime in general but also homicide in particular.

Differences between regions have even determined where scientific research takes place; this is mainly focusing on cities where data is available, i.e., huge and important cities (C. J. Vilalta, 2010; C. J. Vilalta, Castillo, & Torres, 2016). Following this line, no scientific research has been developed on León, Gto., although it is the fifth most important city in the country, in terms of GDP; even now with the existence of a law that forces the government to share non-sensitive data with anyone, there is still no research in this area. Nevertheless, a non-governmental organization has presented crime incidence reports of the city and has mentioned how this affects citizen behavior[4]. For example, spending money on better security systems, changing travel routes, or the time of day when people wanted to move from one place to another. From a national perspective, data from the National Statistical and Geography Institute stated that "in 2018, the total cost as a result of insecurity and crime in households represents an amount of 286.3 billion Mexican pesos, that is, 1.54% of GDP" (INEGI, 2019), which makes it financially important to invest in better data that could lead to a better strategy around the whole country. This makes it extremely important to analyze, insofar as possible and being aware of the limitations (availability of data and academic articles), different crime/homicide theories in the Mexican context in order to, on one hand, collaborate to the scientific knowledge with new scenarios for well-established theories and models; and, on the other hand, cooperate with the society to improve the Mexican reality.

The natural evolution of homicide research has led to most of the recent literature being developed around its occurrence within cities. However, the lack of data in previous years resulted in Mexico not being tested by well-defined homicide/crime theories; even when the current Mexican context demands a better understanding of the national war situation. My proposal tries to contribute to scientific knowledge and society by testing previous homicide dispersion literature on a national level with cities as units of analysis (Messner et al., 1999; Baller, Anselin, Messner, Deane, & Hawkins, 2001) as well as testing different econometric and machine learning approaches to give a base line for future research in Mexican crime analyses. To achieve this, my research question is:

[4] Observatorio Ciudadano de León (OCL). (2016). Victimización: una radiografía ciudadana de León. Retrieved on 28/12/2020. http://www.ocl.org.mx/wp-content/uploads/2016/11/Presentaci%C3%B3n-OCL-24-Noviembre-de-2016.pdf

Which are the demographic, criminal, economic, political, time, and spatial factors that explain variations in homicide rates between Mexican cities?

1.2.1 Sub-questions

Previous literature suggests three different types of variables to take into account for the analysis: Spatial concentration of homicides, different type of crimes that could lead to homicide (kidnapping, thefts, black markets, etc.), and social-cultural determinants for occurrence of homicide (demographics, space, politics, development level, etc.). Finally, to answer the research question using the mentioned variables, it was divided into 5 sub-questions as follows:

1. How can the concentration of homicide incidences between Mexican cities be measured and compared?
2. What are the relevant cultural, political, or social aspects to determine the occurrence of homicide between Mexican cities?
3. How does the correlation/effect/cause identified in the previous question depend on the type of model addressed to explain the homicide occurrence?
4. Spatial effects: How are the homicide rates in city X affected by homicide rates or by other attributes of nearby cities?
5. In a country as diverse as Mexico, How do cities can be grouped based on their social, economic, and criminal characteristics to test the feasibility of the base model within groups?

Plan:

1. Use a combination of the three social theories described in section 2 and empirical research, looking for the best model (initially, we will focus on social disorganization theory)
2. Identify and match the different types data sources that contain information on territory division, crime and socioeconomic variables.
3. Use the occurrence of crime per 100 thousand inhabitants to first measure the homicide concentration in cities and modify data in order to avoid outliers and non-normal distributions if needed.
4. In order to test the presence of city-effects on the homicide rates over cities, compare the baseline model (complete panel) with the random and fixed effects.
5. As a first approach to the spatial dispersion hypothesis, test for spatial concentration of homicide with the global and local Moran's I

6. If spatial randomness is rejected,
   - Define the best way to control for spatial heterogeneity
   - Compare the baseline model with the spatial lag and the spatial error models to test the spatial effects of crime and other variables (the data for the spatial lag model will be provided at the city level)
7. In order to look for different realities in Mexico and measure how it characterize the country, we will use the k-means model for clustering Mexican cities and test the base model within each cluster.

1.3 Areas not reviewed

As mentioned above, analyzing homicide at the city level would have the Modifiable Areal Unit Problem (MAUP) problem, as any spatial analysis. In this sense, C. J. Vilalta (2010) showed that the police sectors (zones[5]) do not show a concentration of drug possession crime in Mexico City, even when social theories state it. Moreover, only one area under its analysis does not present a mixture between commercial and residential land uses. This confirms the need for utilizing smaller zones. Howbeit, that better understanding of homicide is not feasible due to data at smaller areas than cities is not available on the national level.

Another aspect that this report does not cover is the penalty, form, and more aspects that influence homicide punishment. This lack of analysis is due to Mexico's high levels of corruption which are hard to measure. Almost by definition, a strong characteristic of corruption is to happen outside of public view.

[5] Mexico City became a State in 2018 and now those zones under study are cities

Chapter 2

Literature Review

2.1 Context: Organized Crime and Mexico

Whenever a Mexican is asked about homicides in her country, she always mentions organized crime at some extend. Academia and government also know that those topics are extremely related in this country. In order to a better understanding of crime incidence, the first step that this thesis takes is to analyse organized crime and its particularities in Mexico.

2.1.1 Background

Organized crime is not a problem that that concerns Mexico only, it has been part of all sort of countries' realities for decades. Nevertheless, this concept has been oscillating between tow main notions: 1. One that involves the activity of criminal groups with a certain degree of organization and 2. Another involving some serious illegal activities (Paoli, 2014). Nowadays, those two possible and current definitions make it hard to create effective policies or even theories against organized crime. Paoli and Beken (2014) describe how those two ways to approach organized crime basically respond to the questions "Who" and "What", respectively. Since the idea of organized crime was minted in the United States (U.S.) in the last decade of the 19th century, these two concepts are extremely linked with the U.S. history, going from a focus on markets (what) to ethnicity stereotypes (who). Going from one concept to another, finally resulted in a merging broad definition that makes it hard to analyze organized crime, but catches more actors and activities. On the other side of the western world, Europe has mainly categorized organized crime a problem that should be assessed through the "What" (markets).

The territory scope of organized crime has changed over years and it is a characteristic that both definitions must account for. According to Paoli and Beken (2014), internationalization of organized crime is something that has to be considered when trying to analyze this phenomenon. They mention that usually production, retail distribution, and other aspects of the illegal markets take place locally, and only some parts of the process, such as transport or communication between exporters and importers, are international. That internationalization of organized crime has led to an international response where countries over the planet have agreed on taking similar measures for fighting this problem together.

The constant change of the concept of organized crime has brought an underdevelopment of theories and empirical research around the phenomenon. Kleemans (2014) mentions that "*the history of organized crime research is not only the history of shifting theoretical perspectives, but it is also the history of oscillating empirical phenomena that are at the forefront of public and scientific discussion*". It describes the following theories/approaches to organized crime: 1. Alien Conspiracy Model, 2. Bureaucracy Model, 3. Illegal Enterprise, 4. Protection Theory, 5. Social Embeddedness, Social Capital, and Criminal Networks, and 6. Logistic or Situational Approach toward Organized Crime. The last is the one used in this thesis due to it responding better to the spatial variable of interest and because it has been already used to describe the phenomenon in Mexico (Estévez-Soto, 2020).

According to Fijnaut (2014), what we know about the history of organized crime presents 3 main problems: 1. There is no consensus between researchers on what should be considered or not as organized crime, 2. There is a lack of original empirical research on this topic, since it has not been studied enough in regions where there are not enough researchers who study it, and 3. The broad definition makes it hard to study it alone since other types of crime are always present. These three problems make it crucial to continue with academic literature on these topics in Mexico. Therefore, even when organized crime is not the object under analysis in this thesis, I strongly consider that is has to be included or taken into account in any analysis of crime in this country.

Situation Approach towards Organized Crime

This approach focuses on the idea that offenders act not because of motivation but because of the opportunity for committing a crime. Therefore, this approach focuses on the environment where the offence occurs more than on the offender. It follows the same idea as the the Windows broken theory, Social disorganization theory and crime pattern theory exposed in section 2.2. Even though this thesis does not try to analyze the organized crime in Mexico, considering it from the Situation Approach makes more sense due to the final aim: describing the factors that make a city more susceptible for committing homicide than another city.

2.1.2 Organized Crime in Mexico

When the prohibition era started in the U.S., the organized crime era started in Mexico as a response for the demand of alcohol and drugs in the north neighbour. Since then, Mexico started its history as producer of mainly marijuana and opium in the northern states, next to the border to the U.S., where demand was waiting for them. By 1970, Mexico became the main supplier of marijuana and heroin to the U.S. and a transit route for cocaine (Medel & Thoumi, 2014). Before 2006, when the official declaration of war against drug cartels occurred, organized crime had already high-level ex-military members and close ties with high authorities in all Mexican government levels. Therefore, the proximity between Mexico and the U.S. explains the incurrence of Mexico in drug production but its own corruption and links to organized crime explains the evolution and heyday of it.

Different successful strategies in countries of the world have been applied to combat organized crime in Mexico (e.g. "Operation Condor" in 1977 and "Merida Plan" in 2008). Nevertheless, its particular context with a corrupt and weak government made them all fail, Medel and Thoumi (2014) stated that "*Operation Condor pushed less daring and smaller traffickers out of business, ensuring that only the largest and best-organized groups survived*". As Paoli (2014) mentioned, even successful policies (like those implemented in the United States) could have a negative effect in other regions (Central America and, recently, West Africa), which is known as the "balloon-effect". On the other hand, policies implemented in other countries or regions also affected the evolution of Mexican drug cartels. Trust in connections between two region becomes crucial for the success in trafficking from a country to another, as mentioned in Kleemans (2014) about cases of trafficking from the Caribbean to the Netherlands. The historical population relationship between Mexico and the U.S. made it easier to find connections from one side of the border to another and then Mexican cartels negotiated with Colombian cartels to transport cocaine through Mexico instead of the Caribbean route that was under the attention of Colombian-U.S. policies to defeat Colombian Cartels in early 1990 (Medel & Thoumi, 2014).

However, those failed policies in Mexico described above, success in the US and other regions in the fight against organized crime, and corruption and a non-strategic declaration of war against organized crime in Mexico, resulted in a country immersed in a war for more than 14 years. Before the war against drug organizations started in 2006, it made sense to call them "cartels" [1], but now these groups do not have this structure but the name is just a recall for those days when the production of drugs was controlled for a few organizations. Nowadays, the war against organized crime and internationalization of drug cartels made them change from a few big family cartels to a big number of small and sophisticated businesses. Moreover, those businesses have presence and control through high level of violence in some sectors of markets in countries in Latin America (e.g. Colombia and Peru) and even Europe (Medel & Thoumi, 2014). But even only in Mexico, diversity of organized crime has reached a really complex face since it has presence in many social and economic sectors, as stated by C. Vilalta (2014): "*They can be found in the drug trade (i.e. production and trafficking), kidnapping, bribery, car theft, pirated goods, sex*

[1] From an economic perspective, a cartel is a group of organizations in the same market who work together in controlling prices or production with the aim of control the market. This could be seen as a monopoly controlled by a group of organization (the OPEC is a clear example of a cartel)

trade (adult and children), illicit organ transplanting, human smuggling, and credit card cloning.", but the list is longer .

All those factors combined made Mexican "Cartels" to fight for territory, young people, markets, resources and links with the government. Those groups are divided in among regions in the country and have links with local governments with a last consequence of a high number of homicides in the country (more than 35,000 in 2020[2]). In Mexico, media has taken the amount of homicides as an approach to the government efficiency and the last indicator for irreparable violence. Therefore, this thesis attempts to contribute to understanding this phenomenon and is expected to contribute to producing a better strategy to combat this problem. Just as a remark of the importance of this type of analysis, C. J. Vilalta (2010) mentions that "*More data and comparative research are needed of course. Nevertheless, results show that the war on drugs might be won if local authorities connect urban planning and social development policies with police strategies*".

2.2 Economic Approach and Social-crime Theories

2.2.1 Economic Approach

After the publication in 1968 of *Crime and Punishment: An Economic Approach*, by the now Nobel Prize winner Gary Becker, the study of crime has developed a new face and it could be seen as an economic phenomenon that can be represented as a market. Becker (1968) came up with a seminal contribution to economic analysis. He used two variables in the model, the probability to be discovered, apprehended, and convicted (p) and the differences of nature and extent of punishments, i.e., the size and form of punishment (f). These two variables determine three behavioral relations that would state the efficient equilibrium: Number of offenses (O), Cost of achieving p (C), and effect of changes in p and f on O. This model takes into account that the appreciation of those variables would be different from one person to another. Therefore, the optimal location of offenses and cost and revenue would be stated for the government through p and f. This location must consider the socio-cultural and economic aspects of a region. Finally, setting those variables will differ from one crime to another, depending on the severity of it (homicide should have higher p and more strict f than, for example, street robbery).

Therefore, research on crime in general and homicide in particular from an economic point of view is possible according to Becker (1968), who suggested that different type of crimes have different equilibria and, therefore, they should be analysed individually for a better understanding. He presents how p and f must be different within the model depending on the crime type. This came up when he was explaining how the appreciation from the offender of committing a crime is a function of those variables. Hence, offenses like murder must have a high enough probability and a strong enough punishment that

[2]https://www.animalpolitico.com/2021/01/mexico-homicidios-35-mil-2020/

prevent their commission (even when offenders are risk-takers they might not prefer to commit the crime). On the other hand, street robbery could have a lower probability and punishment compared to homicide. The selection of those levels of p and f will differ from each offense to another and from one society to another, because it depends on legal and penal systems, and the possible profit of committing an illegal activity versus committing a legal one.

Following that, C. J. Vilalta (2010) found that there was no spatial concentration of arrests for possession of marijuana and cocaine in Mexico City as a whole (global Moran's Index). However, six hotspots were found in form of a spatial outlier (local Moran Index); C. J. Vilalta (2010) defined spatial outlier as "a spatial unit that is significantly different from its neighbor units". Those seven hotspots were distributed in four for marijuana and three for cocaine (one hotspot for both marijuana and cocaine). This difference of only one common hotspot, in addition to different correlated explanatory variables for the group of marijuana arrests, with respect to cocaine arrests (for the last no socioeconomic correlates were identified) show clearly the difference in behavior of criminals within different crimes, seeing them as dissimilar markets.

This thesis tries to analyze the socio-cultural, economic, and regional aspects that affect the spread and commission of homicide. It aims to help in setting the optimal location of the number of homicides and the cost of achieving p. If needed, setting them per regions, based on a new strategy or the legal division of the country (Mexico is a federal country making it possible to have different constitutions per state).

2.2.2 Social-crime Theories

Different social theories have been employed over the whole homicide committing literature to explain which variables determine criminal behavior. I will focus on three frequently used in crime researches: Windows Broken, Disorganization, and Crime Pattern theories.

Windows Broken Theory

This theory gives an explanation about how and why some neighborhoods tend to be crime attractive and others are not. It was postulated by Wilson and Kelling (1982), where they explained that physical aspects of a neighborhood could "tell" criminals that there is a place where the law is not applied. They gave an example to illustrate the idea, which is an experiment made in 1969 by Philip Zimbardo where two cars were "abandoned", one in the Bronx, New York and the other one in Palo Alto, California. The car in the Bronx was attacked after 10 minutes and was completely vandalized after 24 hours, meanwhile, the one in Palo Alto passed one week without any attack. After that time, Zimbardo damaged the automobile himself and, after a few hours, the car was completely vandalized as well. In this sense, Welsh, Braga, and Bruinsma (2015) present a deep research in studies developed around this theory and policies produced by it. They refer to the main theory's idea that social and physical characteristics of a neighborhood could produce fear in citizens, what would force them to isolate themselves or leave the neighborhood, which implies a lack of control and increasing of those characteristics that "at the end" will attract more criminals.

Social Disorganization Theory

This theory was developed by the Chicago School, specifically, by Shaw and McKay. It states that three main aspects of neighborhoods would determine the occurrence of street crime. Those aspects are low socioeconomic status, racial heterogeneity, and residential instability. The lack or decrease of any of them would affect the capability of society to organize themselves and achieve common goals to fight crime. Sampson and Grove (1989) tested the theory with data for Great Britain and incorporated family disruption and urbanization as variables that determinate social disorganization. More recently, C. Vilalta and Muggah (2014) applied this theory in the city of Ciudad Juárez, Mexico as the framework that guided his model for explaining homicidal violence, nevertheless, some relationships did not result as suggested by the theory.

Crime Pattern Theory

Brantingham, Brantingham, and Andresen (2017) started saying that there are patterns in crime; its occurrence is not random or uniform over time, space, neighborhood, and social groups. In this sense, they proposed the geometry of crime and the crime patterns theory as an explanation of that distribution of crime. In this theory, frequent locations either in legal or illegal activities are seen as patterns of human behavior and they would determinate the occurrence of crime. Those patterns are determined for a whole set of conditions (social, economic, and so on). About this, Bernasco, Ruiter, and Block (2017) said that "*It states that for a crime to happen at a certain location, two conditions are necessary: the place must provide an opportunity for crime and the prospective offender must be aware of the place and the opportunity*"

2.3 Previous researches into homicide and crime

Homicide analysis in Land, McCall, and Cohen (1990) evaluated inconsistencies and results in previous researches that tried to explain the occurrence of homicide across U.S. cities, metropolitan areas, and states. Based on those studies, they constructed a baseline model that is considered to be the best approach for analyzing homicide in the U.S., according to the available data and regressors used by previous works. The 11 variables selected are: 1. Population Size, 2. Population Density, 3. Percentage Black, 4. Percentage ages 15-19, 5. Percentage Divorced, 6. Percentage kids not with both parents, 7. Median

Family Income, 8. Percentage families in Poverty, 9. Gini Index of income inequality, 10. Percentage Unemployed, and 11. South. It was found that for some empirical studies those structural society variables could have a positive, negative or null sign depending on the study. Likewise, they could have a significant impact in the theoretically expected direction or not. Land et al. (1990) found that the variance in results was determined for 5 main factors:

1. Time periods. Not all studies were carried out with many year of observation since structural society variables come mainly from census data.

2. Units of analysis. Researches used different units of analysis like cities, metropolitan areas and states, depending on what they considered as the best way to approach the occurrence of homicide (or crime).

3. Samples. Papers with states as unit of analysis use all states in their research, but those with cities and metropolitan areas tend to use only a subset of all available data.

4. Model specifications. Studies use different model specification mainly by transforming linear data to nonlinear, and by using different predefined indexes like the structural poverty index.

5. Problems of statistical inference. At that date, just like now, new methodological innovations in regression analysis were developed. Some studies included them but others did not. That made it hard to compare exact results and thus only face values were compared.

The dependent variable, homicide rate, was not transformed in their analysis, but they found a nonlinear relationship between homicide rate and population size, population density, and median family income. Therefore, a log transformation was applied to the latter. Moreover, Land et al. (1990) ran a Principal Component Analysis which found that variables tend to group in two main components: 1. *population structure component* with the population size and the population density, and 2. *resource-deprivation/affluence component* with median family income, percentage of families in poverty line, Gini index, percetange of population that is black, and percentage ages 15-19 or percentage kids not with both parents. This component analysis resulted in a six variable regression model with a R^2 between 0.5 and 0.6 at the city level, 0.55 and 0.6 at the metropolitan area level, and 0.65 and 0.85 at the state level. Moreover, they found that the two components and the percentage of population divorced present a positive significant coefficient according to the theoretical expectations. Nevertheless, the other three variables (percentage of young, unemployment rate and south) presented a more diverse impact among the 9 regressions and different signs for most of them, respect to the theoretical expectations. Finally, they concluded that the theoretical assumption of invariant relationship between regressors and regressed variables seems to be invariant over time and social space (when accounting for large samples, standard definitions and reduce collianearity).

Different variables of Land et al. (1990)'s baseline model have been explored to formalize their relation with crime. For instance, in a general model for crime, Gaigné and Zenou (2015) used the monocentric model with two types of individuals (criminals and workers) to test the relationship between agglomeration and occurrence of crime. They found that an increase in city size implies an increase in the per-capita crime rate. They concluded that, in the short run, better accessibility to jobs implies a decrease in crime. However, in the long run, job accessibility increased crime rate. This suggest to test and include the impact of density and population in the homicide model of this thesis.

2.4 Spatial Correlation

Similar to other social phenomena, different studies have suggested that crime presents a spatial concentration depending on many factors. Daily spatial behavior of people strongly determines where they will be located in the future (either for new places or for going back to already known locations). Bernasco (2019) stated that he "*demonstrated that offenders not only tend to commit crimes in and around the places they have visited before when pursuing either their legal daily activities but also have a tendency to commit crimes in and around the locations of their previous crimes*". In this sense, it is also mentioned that those spatial patterns are crucial in any strategy against crime. Moreover, if a government wants to apply a plan to prevent crime, it has to take into account the forecast of the possible future location of offenses. On the other hand, even when there is a huge amount of literature that relates crime and space, this relationship needs to be explored more deeply in order to better attend for local characteristics of countries, regions or within cities. Likewise, previous literature presents a need for a standardized way to compare different studies. In this sense, Bernasco and Steenbeek (2017) pointed out a lack of a standardized method to measure and report concentration on crime. Therefore, they present and suggest the Lorenz curve and Gini coefficient as possible methods, with The Hague as a case study. Using those indicators, Bernasco and Steenbeek (2017) propose a standardized version of them for the case when the number of places is bigger than crimes.

In addition to previous discussed crime related studies, another important variable around crime location is time and how it relates to explanatory variables. Changes in those variables over time could determine the historical amount of crime in a particular location. This importance on time could be present not only through years, it could also be present as a "cycle" during the hours of the day and the days of the week. Bernasco et al. (2017) tested the hypothesis that there is a strong correlation between time and place in street robbery, that is to say between the when and where. Over a lack of research at this spatial-time level, they identified one with Philadelphia as a case study that assessed a similar hypothesis. Nevertheless, unlike Philadelphia's study, they concluded that street robbery location is not determined by hours and days, but by potential social/economic determinants like places with cash-intensive business and/or where the probability to be apprehended is not high but the benefit of committing a crime is. However, the presence of high schools was the only factor that actually affects location choice of robbery, showing more robs around school hours than other moments of the day.

2.4.1 Homicide diffusion

Homicide has been studied from a very diverse perspective. Recently, its spatial diffusion has become an important research branch of analysis. According to Messner et al. (1999), in this process "*a crucial parameter pertains to the location of the initial "shock" to the system (the "innovation") relative to the location of the "adopters" across space and over time. Adopters are either immediate neighbors, members of the so-called mean information field (contagious diffusion), or located in nodes connected to the origin of the innovation in a hierarchical network fashion (hierarchical diffusion)*". Since the Exploratory Spatial Data Analysis (ESDA) is useful for the study of diffusion process, one first approach was to look for local patterns through the local and global Moran's I carried out by Messner et al. (1999). It was found that homicide seemed to not be randomly distributed over time and space in the counties of the metropolitan area of St. Louis and its surrounding areas. They also found evidence for the existence of areas that serve as barriers for homicide diffusion. Those were the most affluent areas, or those more rural or agricultural areas. As Messner et al. (1999) mentioned, "*The patterns of spatial distribution revealed through ESDA provide an empirical foundation for the specification of multivariate models which can provide formal tests for diffusion processes*". Theories and methods had to be developed to be able to state this spatial dispersion of homicide.

Following those results and techniques, Baller et al. (2001) mentions that if a spatial process is present and models do no account for it, estimators may be biased. Therefore, they used the spatial lag and error models to account for spatial dispersion of homicide in the U.S. between its counties (controlling for spatial heterogeneity and social and cultural covariates described in Land et al. (1990)). Previous results about the not random distribution of crime were confirmed. In addition, they came up with the hypothesis of different possible causal process within the country, and how spatial analysis can reveal regions with different effects of predictors between them. Following that idea, models were run separately for the southern and northern areas and found that, on one hand, the spatial lag model fits better in the south and, on the other hand, the spatial error model fits better in the north. Nevertheless, theoretically, it is challenging to find an explanation for those not included variables that affect the spatial diffusion of homicide in the northern counties.

2.5 Previous Crime/Homicide related studies in Mexico

The three social-crime theories described above were designed for inner city crimes, such as theft. The main differences are the social variables and statistic method that each paper took into account; even when the idea is to link them in a single model, there are studies that use spatial correlation of the variable under study to find hostspots and then look for a linear correlation or significance with the regressors. This is the case of C. J. Vilalta (2010), who found a correlation between marijuana arrests hotspots and socioeconomic variables (college education, poor housing conditions, and female-headed households); meanwhile no correlation was found between cocaine arrest hostspots and socioeconomic characteristic, by using a mix between Window broken and Disorganization theories with the Moran's I and Pearson's correlation coefficients. Nevertheless, multiple studies have been carried out on a higher level in this type of topics and a few of them took place in Mexico. For instance, C. J. Vilalta (2012) debates about the possibility of having a national behavior against spatial clusters with different patterns compared with the national trends, due to differences between local governments or judicial trusts. Even when this is not a study about crime but judicial data and policy analysis, it gives a possible idea to analyze different geographical patterns in the entire country at a lower level, like metropolitan areas or cities. He used the spatial lag model just like Baller et al. (2001) for homicides in the U.S.

Thinking on those three mentioned theories as an explanation for possible spatial concentration of homicide in Mexico between cities could give a new perspective of Latino America crime issues. C. Vilalta (2014) suggests that the initial conditions of the Mexican states when the war against organized crime started determined in many cases the output in terms of crime. This suggests that it could be possible to compare a city average of the variables used in the social-crime theories. Thereby, spatial concentration of crime and its relation with social variables must be tested at a municipal level over time (to try to find time spatial concentration or trends). This hypothesis has not been tested formally in Mexico at the city-level, nevertheless, C. Vilalta and Muggah (2014) already tested it in the city of Ciudad Juaréz, Chihuahua in the north of Mexico. They followed the social disorganization theory at the police zones level and the geographical electoral level. The first level was used for the spatial analysis trough the global Moran's I, and the second level for the binary logistic regression model, where was tried to identify the determinant covariates for making a police sector to be above the city average of homicide. They found evidence for possible spatial autocorrelation of homicide within Ciudad Juárez; nevertheless, some of the logistic coefficients resulted in an effect sign opposite to what the Social Disorganization Theory would predict.

In comparison with high-income level countries, just a few pieces of research have been developed in Latin America's crime spatial patterns, Vilalta, Castillo and Torres (2016) is a no-scientific work (belongs to Inter-American Developing Bank) which presents a comparison between different patterns in and along cities in this zone of the continent. This study shows the need for more scientific papers.

Chapter 3

Data and Scope

Gathering and joining the used data for realizing this work was challenging since it is available for different time periods and it comes from different sources according to its nature: Crime, Organized Crime, Social-demographic, economic activity and territory division, each one of them at the city-level. In the following sections can be found a description of these data sets and how they were processed and merged to create a yearly panel data at the city level for the period 2015-2020.

3.1 Territory Division and Unit of Analysis

As a country in constant growth, Mexico has also suffered several changes in its territorial internal division. Its continuous growth has brought a division of different cities within the country according to local characteristics and social representative preferences. This makes it impossible to simply take a census taken in year x and compare it with another from a different year because new cities were added to the division of Mexico in cities. Carrying out a research using cities as object under study must take into account this division change of cities over time. In this sense, this thesis will use the territory division from the National Geoestatistical Framework from 2010[1]. This decision was made based on the period under study (2015-2020) and the frequency of collection of needed datasets. In order to match datasets over time, if a city A is divided into City B and City C, all variables for cities B and C will be computed and aggregated for City A.

By 2010 the national territory of the United States of Mexico was formed by 32 states[2] with 2,456 cities across the country. Within the cities, another 2 subdivisions exist: 1. *Localidad*, which is the lower official territorial division in Mexico and represents territory occupied for one or more households and this place is recognized by a name given by law or custom, and 2. *AGEB*: which constitutes of the basic unit of the National Geostatistical Framework. Those three different levels of division represent possible units of analysis; however, the decision of choosing the ideal unit of analysis rests on the liter-

[1]https://www.inegi.org.mx/temas/mg/#Descargas

[2]Mexico Federal District (D.F.) is considered a State due to in 2016 it became one

ature suggestions for the analysis of organized crime and homicide spatial diffusion and the availability of data. In this sense, resent research has focused on the lower possible geographic unit, like neighborhood or blocks (Bernasco, 2019; Bernasco et al., 2017; C. Vilalta & Muggah, 2014), suggesting that the ideal unit would be the AGEB-level. However, the availability of data limits the possible units of study because Mexico does no have crime data at the AGEB-level (the lowest level with data available is the city-level). On the other hand, when this topic started to being studied in the U.S., research was carried out at the County (Baller et al., 2001; Messner et al., 1999) due to the lack of previous research and data availability. Compared to this first research, using data at the city-level as a beginning of the area of analysis of spatial diffusion of homicide in Mexico gives a promising future (counties are bigger than cities). Therefore, using data at the city-level brings the advantages that 1. it is possible to find other data at this level (social, economics, demographic and crime data), and 2. the political, economic and cultural division of regions in Mexico is well represented by the characteristics related to the cities.

3.2 Crime Data

Analysing crime data involves different challenges, from its availability to the correct classification by the authorities in turn. Therefore, when we try to assess how crime behaves in a particular context, we have to take into account the possible noise that accompanies the data. In this sense, for the specific case of Mexico, homicides is a good indicator since it does not need to be reported to the public prosecutor, instead, the government has to investigate it, which suggests a level of non-reporting lower than other crimes. Furthermore, México-Evalua (2020) mentions that homicide data are usually compiled on the basis of national criminal codes, which brings a better understanding and easier classification of the crime (minimizes the probability of error when people are in charge of classification). Moreover, crime data comes from the Mexican National Secretariat of the Public Security System[3] whom in 2018 updated and improved the methodology for reporting crimes to make it more trustworthy when analyzing Mexican reality (this improvement was made retroactive in 2015).

As mentioned earlier, homicide data from the SNSSP uses a legal classification and it distinguishes between culpable (unintentional or accidental) homicides from malicious (intentional) homicides. México-Evalua (2020) evaluated the reliability of malicious homicides through its Criminal Statistics Reliability Index and concluded that only five[4] states out of 32 present a index below 7 and above 5 (where 10 is the highest and 0 the lowest). This enforce the idea that SNSSP's homicide data can be used for a first analysis at the

[3]https://www.gob.mx/sesnsp/acciones-y-programas/datos-abiertos-de-incidencia-delictiva?state=published

[4]Oaxaca 6.9, Puebla 6.9, Guanajuato 6.84, Tabasco 5.75 and Tamaulipas 5.08

Table 3.1: Missing Data for Occurrence of Crime.

Year	Missing States	Missing Cities
2017	Oaxaca	5
2016	Oaxaca and Chiapas	540 (5 and 535)
2015	Oaxaca and Chiapas	540 (5 and 535)

city level. This data is updated monthly and we used the available for the period January 2015 – December 2020. Due to availability at the city level, all crime data represent crime-incidence and not victim-incidence, *i.e.*, representing the number of crimes reported to the Mexican government across the country per municipality and not the number of victims. Categories are stated based on the legal definition of each crime-type; that definition also sets whether the crimes must be processed by the municipal, state or federal administration. For the analysis of homicide occurrence in this thesis I grouped the categories "feminicides" and "malicious homicides" in one single category: "homicides"[5].

Classification of crimes is divided between 40 different categories with 55 subcategories in total. Likewise, it has the modality or the form in which the crime was committed (with a firearm or a knife, etc.). I am aware that recent studies have pointed out that legal crime categorization presents a challenging scenario when trying to represent contextual characteristics of committing a crime. Kuang, Brantingham, and Bertozzi (2017) found that the legal categorization of crime into discrete groups brings a massive loss of information. That loss is due to crime occurrence depending on different behavior and situations that might not be represented by the legal definition and a social or contextual categorization could be the best approach. Nevertheless, discrete categories have been used as the best existing approach for analysing crime patterns (Baller et al., 2001; Messner et al., 1999; C. Vilalta & Muggah, 2014). Based on this, the analysis of homicide in this thesis will be addressed by using legal crime categorization for the final good affected by the crime commission and, as specific crime analysis, some particular subcategories that the current situation in Mexico suggests as possible explanation for the occurrence of homicide (for a detailed description see Appendix A).

As stated before, by 2010 the national territory of the United States of Mexico is formed by 32 states with 2,456 cities across the country and the period under study due to availability of crime data is 2015-2020. Nevertheless, for this dataset the panel data is not balanced, it presents missing data (See Table 3.1). In order to solve the missing data issue, I used the mean-imputation method per city for completing the data set and have a balanced panel, where the mean of the available years is computed and used in the missing year for those years.

[5]Unintentional or accidental homicide was not included due its nature does not necessarily represent/reflects the phenomenon of violence that is the aim of this work

3.2.1 Organized Crime Data

As mentioned in the section 2.1, research on organized crime presents a number of challenges due to the ambiguity of its definition and the lack of previous works. That makes it harder to collect data for studying a non completely clear phenomenon. Moreover, Hobbs and Antonopoulo (2014) mention how difficult is to get data on organized crime since some times even ONGs dedicated to something related to this topic, do not share their data. This lack of data on organized crime makes it really hard to control for this variable, compare research or even carry out one about the same topic in different parts of the world. The most complete and available data on organized crime for the purpose of this thesis is the data coming from the platform Lantia Intelligence[6]. Nevertheless, the non-clear specification of that date due to author right makes it not matchable with other official data used in this thesis. Therefore, I will use the geographical division that the government announced back in 2013 as a representation of the organized crime type or criminal groups located in a region and that will help in the war strategy against Organized Crime[7]. This division responds to the presence of different cartels and their activities along states and could also be an approach to the presence of organized crime in different regions. Regions were set as follows:

- Zone 1: Baja California, Baja California Sur, Chihuahua, Sinaloa y Sonora;
- Zone 2: Coahuila, Durango, Nuevo León, San Luis Potosí y Tamaulipas;
- Zone 3: Aguascalientes, Colima, Guanajuato, Jalisco, Michoacán, Nayarit, Querétaro y Zacatecas;
- Zone 4: Distrito Federal, Estado de México, Guerrero, Hidalgo, Morelos, Puebla y Tlaxcala, and
- Zone 5: Campeche, Chiapas, Oaxaca, Quintana Roo, Tabasco, Veracruz y Yucatán.

3.3 Social and Demographic Data

Data for social and demographic variables has been collected from the National Censuses of 2010 and 2020[8]. Those datasets present different variables that could be derived at the city level, such as households with a single mother or father, number of men or women, inhabitants age-range, characteristics of immigrants, indigenous languages spoken and people who know them, illiterate population, school grade/level, population part of an

[6]Company that has studied the main variables on organized crime in Mexico since 2008. https://lantiaintelligence.com/

[7]http://static.adnpolitico.com/gobierno/2013/01/10/la-segob-establece-5-regiones-en-la-estrategia-de-seguridad

[8]https://www.inegi.org.mx/programas/ccpv/2020/

economic activity, population with national health insurance (this could be an approximation of formal workers), religion, number of inhabited houses, number of inhabitants per house, amenities or services in the house (concrete floor, potable water, electricity, and so on), IT devices or household appliances, and so on. The frequency of censuses brings missing data issue in the between collection years. In order to be able to use all available crime data and to solve the missing data of years between census I used a linear interpolation with a constant rate of growth between the two data points. Likewise, in order to account for the creation of new cities and differences in number of cities between censuses, the process of computation and aggregation described in section 3.1 was applied.

3.4 Economic Activity Data

In order to account for city Economic Activity, data from the National Directory of Economic Units (DENUE, for its acronyms in Spanish) was collected. These data sets contain information of all economic units in Mexico (formal and informal businesses), from a small local grocery shop to an international company for the period 2015-2020. The DENUE is been updated different times a year but for the realization of this thesis I use the last version for each year as follows: October for 2015 and 2016, and November for 2017, 2018, 2019 and 2020. This data contains information at the economic unit level about the number of employees, type of unite (fixed or movable), type of economic activity, type of settlement, type of roads where the unit is located, date of incorporation to DENUE, and so on. Nevertheless, based on the available data for the territory division at the city level and years of collection, some disparities of variables classifications between yearly datasets made difficult to use a couple of variables. That disparities were present in the sense of new categories in variables like type of settlement or type of roads where the unit is located. Therefore, the only matchable variables were number of employees, type of unite (fixed o movable), and type of economic activity. Likewise, in order to account for the creation of new cities and differences in number of cities between DENUE datasets, the process of computation and aggregation described in section 3.1 was applied.

3.5 Data Merging

When this type of research is carried out in developing countries the merging process presents a challenging scenario due to different problems. The first and usually present is the missing data problem, that was solved using the mean-imputation method per city, as stated before. Another challenge faced was the addition of 13 cities when comparing the political division of Mexico in 2010 with 2020[9] (stated in section 3.1). As we know, when trying to analyze a phenomenon over time, the permanence of the agents becomes crucial. In this thesis the object of analysis are the cities and, as Mexico grows over time,

[9] 1 in Baja California, 1 in Campeche, 6 in Chiapas, 3 in Morelos and 2 in Quintana Roo

they have also changed over the years. In order to minimize the loss of information and to have a fixed panel data, the political division of 2010 was used as the definition of the agents that will drive this thesis and all data sets follow this criterion. It is important to mention that originally all the data sets follow the statistical format provided by the National Institute of Statistics and Geography, which gives a plus in the coincidence and comparability of the data sets.

Chapter 4

Methodology

4.1 Baseline Model

Different model specifications have been analysed for explaining the occurrence of homicide over time, neighborhoods, cities, and space diffusion. The exploratory phase and first approach to this phenomenon driven by this thesis will be through a Ordinary Last Square (OLS) specified in equation 4.1.

$$y_i = \sum_k X_{k,i}\beta_k + \epsilon_i \tag{4.1}$$

Where y_i is the homicide events per 100,000 inhabitants, $X_{k,i}$ is an element in the $X_1, ..., X_k$ vector of covariates, β_k is an element in the $\beta_1, ..., \beta_k$ vector of regression coefficients, and ϵ_i is the stochastic error term.

The base model in this thesis follows mainly the covariates specified in Land et al. (1990) and C. Vilalta and Muggah (2014), as well as considering findings by other authors (see Table 4.1), and the availability and relevance of data at the municipality level for Mexican context. The specific model is presented in equation 4.2. Initially, a general OLS model for each year and an aggregated one for the period 2015-2020 were analyzed.

$$\begin{aligned} lhomicide_i = \beta_0 + \beta_1 LPOBTOT + \beta_2 LPopDensity + \beta_3 SIndi + \beta_4 SLand + \beta_5 SFemaleAh \\ + \beta_6 SISSSTE + \beta_7 SMale + \beta_8 S65More + \beta_9 SUnemploy + \beta_{10} SDivorced \\ + \beta_{11} SFem6to11NAS + \beta_{12} SYearsSchool + \beta_{13} DummyR1 + \beta_{14} DummyR2 \\ + \beta_{15} DummyR3 + \beta_{16} DummyR4 + \epsilon_i \end{aligned} \tag{4.2}$$

Where ϵ_i and β_k are as defined in 4.1, $homicide_i$ is the homicide rate and,

- $LPOBTOT$: Log of population size

- $LPopDensity$: Log of population density
- $SIndi$: Percentage of indigenous population
- $SLand$: Percentage of houses with land floor
- $SFemaleAh$: Percentage of houses with female ahead
- $SISSSTE$: Percentage of population with government social security
- $SMale$: Percentage of males
- $S65More$: Percentage of population above 64 years old
- $SUnemploy$: Unemployment rate for population aged 12-130
- $SDivorced$: Share of divorced population
- $SFem6to11NAS$: Share of females aged 6-11 who do not attend school
- $YearsSchool$: Average years of schooling per city
- $DummyR1$: Dummy for organized crime region 1
- $DummyR2$: Dummy for organized crime region 2
- $DummyR3$: Dummy for organized crime region 3
- $DummyR4$: Dummy for organized crime region 4

From the 11 variables selected by Land et al. (1990) 3 were not available at the city-level in Mexico (Median Family Income, Percentage families in Poverty, and Gini Index of income inequality). This thesis accounted for them by using social variables described in C. Vilalta and Muggah (2014), nevertheless, some of them resulted into a non statistically nor economically significant at the city level when were included in the Land et al. (1990)'s specification. On the other hand, $S65More$ and $SMale$ were selected together to account for the median age variable, in addition to hoping that it captures the strong Mexican culture of relying on the elderly for the education of children and as a pillar to keeping family unite, and so, stronger ties in their networks. Along the same lines of accounting for the Mexican reality, percentage of black people will be addressed by the percentage of indigenous people per city. Since this groups of people present a similar segregation and discrimination that make them more vulnerable according to the social disorganization theory. Southern dummy variable will be accounted by the region variables of organized crime zones described in section3.2.1, which captures criminal behavior of organizations located across the country. Finally, it is important to point out that this thesis will use the unemployment rate following the specification of Land et al. (1990) and the C. Vilalta and Muggah (2014) specification that used the employment rate.

Finally, this first approach will test the Principal Component Analysis driven by Land et al. (1990) and Baller et al. (2001). The two components in this thesis will test and try to capture the variance of certain variables in two components as follows 1. Resource deprivation component will consist in percentage of indigenous people, houses female ahead, percentage of population with ISSSTE, percentage of population with land floor, average years of schooling, and share of females aged 6-11 who do not attend school; on the other hand, the population structure component consists of population size (log), and population density (log).

4.2 Panel Data Model

Once we have a treated our panel data with a pooled model, it is time for testing error components models, such as fixed effects and random effects models. The general specification for those models is shown in equation 4.3.

$$y_{i,t} = \sum_k X_{k,i,t}\beta_k + \epsilon_{i,t} \qquad with\ \epsilon_{i,t} = \eta_i + \tau_t + u_{i,t} \tag{4.3}$$

Where y_i is the homicide events per 100,000 inhabitants, $X_{k,i}$ is an element in the $X_1, ..., X_k$ vector of covariates defined in equation 4.2, β_k is an element in the $\beta_1, ..., \beta_k$ vector of regression coefficients, ϵ_i is the stochastic error term, η_i is the city effect unobserved heterogeneity, τ_t is the time effect, and $u_{i,t}$ is the idiosyncratic error term.

Considering the previous equation 4.3, it can be inferred the general Fixed Effects Model described in equation 4.4. This model accounts for individual heterogeneity. In order to avoid potential heteroskedastic and potential correlation over time within cities, this thesis will use the clustered standard error model.

$$y_{i,t} = \sum_k X_{k,i,t}\beta_k + \epsilon_{i,t} \qquad with\ \epsilon_{i,t} = \eta_i + u_{i,t} \tag{4.4}$$

Where it fits the model description for equation 4.3, with the particularity that β_0 should be normalized or impose a restriction to η_i (usually the coefficient is imposed $\beta_0 = 0$).

On the other hand, due to in fixed effects models time-invariant covariates can not be included and most variation in regressors is usually between individual rather than over time, estimators might no be very precise. Therefore, this thesis also explores the possibility of random effect. Equation 4.3 shows the general definition for the Random Effects model addressed in this thesis (two-way error component).

In order to decide whether or not to use the random and fixed effects models, this thesis tested the existence of individual fixed effects through a simple F-test where the polled model was the null hypothesis and the individual fixed effects the alternative hypothesis. For the existence of random effects, this thesis uses the simple Breusch-Pagan Lagrange Multiplier test, where the null hypothesis it the structure of pooled model, and the alternative hypothesis is the existence of random effects. Finally, for deciding between fixed effects or random effects models, this thesis uses the Hausman-Wu test for non-clustered models, where the null hypothesis is that both estimators are consistent, but the Random Effects estimator is more efficient, on the other hand, the alternative hypothesis is that only fixed effects estimator is consistent.

4.3 Exploratory Spatial Data Analysis of Homicides Rates

The understanding of spatial dispersion of homicide across cities in Mexico follows the Exploratory Spatial Data Analysis stated by Messner et al. (1999). It uses descriptive and traditional graphs as well as visualizations through maps and hypothesis tests for spatial patterns. The first step in any spatial analysis is to look for spatial autocorrelation, which will be done trough the global and local Moran's I statistics. Global autocorrelation happens when a local shock affects the whole system meanwhile local autocorrealtion occurs when a local sock affects only neighbours. On one hand, the global Moran's I can tell if global autocorrelation is present by a positive and significant statistic. On the other hand, rejecting the null hypothesis of spatial randomness with the local Moran's I indicates whether a city is surrounding by similar high/low values of homicide rate or by contrary values (either high-low or low-high relation). In order to visualize and identify spatial autocorrelation and spatial outliers, this thesis will use the Moran scatterplots and Moran scatterplots maps. An additional test for local spatial autocorrelation was implemented in this thesis, the Getis–Ord Gi statistic, which tells how large is the homicide of the neighborhood of a city, compared to the average neighborhood in terms of homicide rate. This statistic can give a better visualization of homicide dispersion.

Following previous researches in homicide, the neighbour definition in this thesis will respond to neighbour-base contiguity, *i.e.*, cities whose boundaries are shared with more than a simple point. This thesis does not use the distance-based contiguity due to in general in Mexico, the bigger a city is, the easier it to go to another boundary city, and the other way around. Therefore, distance-based contiguity could be not very precise. Using a row-standardized spatial weight matrix W, the standardized Moran's I general specification is shown in equation 4.5. Where $\tilde{x} = x - \bar{x}$ is a vector with the sum of homicide rates per city for all available years.

$$I_s = \frac{\tilde{x}'W\tilde{x}}{\tilde{x}'\tilde{x}} \tag{4.5}$$

Following results of previous works of possible spatial clusters of homicide rates (Messner et al., 1999; Baller et al., 2001; C. Vilalta & Muggah, 2014), the initial hypothesis in this thesis is that homicide rate will present a positive and statistically significant spatial autocorrelation. Nevertheless, even when that happens, a spatial effect is no the only explanation for it. The presence of spatial heterogeneity trough particularities happening in different areas of Mexico could be the reason. Therefore, when testing for spatial dependence we have to control for possible spatial heterogeneity and see if it persists. In this sense, this thesis uses a geographical division that the government announced back in 2013 and has been described in subsection 3.2.1. That geographical division responds to the presence of different organized crime groups and their activities. In other words, this could also be an approach to the presence of particular groups of organized crime in different regions and how their particular activities in a region form a cluster for some crimes, particularly homicide.

4.4 Spatial Model

When looking for spatial effects of homicide rates in spatial models, the spatial dependence is introduced by another term conformed by either a weighted average matrix of values of the homicide rates in neighbouring locations or through a spatial dependence in the regression error term. A general model specification could be seen in equation 4.6.

$$\begin{aligned} y &= \rho W y + X\beta + W X \gamma + \epsilon \\ \epsilon &= \lambda W \epsilon + \mu \end{aligned} \tag{4.6}$$

Where y and X follows the specifications in equation 4.2, W is a row-standardised weight matrix, ϵ is the spatial error term, μ is the stochastic error term and, ρ , γ, β and, λ are parameters to be estimated.

Based on the definition in 4.6, the spatial lag model can be defined as in equation 4.7. The spatial lag model incorporates a spatial dependence in the homicide rate variable as well as a global autocorrelation of homicide. We use the Lagrange multiplier test when testing for the correct specification of the spatial lag model against the OLS model. The null hypothesis is that the spatial lag model is the correct specification for capturing the spatial effect. Therefore, rejecting the null hypothesis implies that the spatial lag model is not the appropriate model for capturing the spatial effect, in case the spatial lag effect (ρ) is statistically significant.

$$\begin{aligned} y &= \rho W y + X\beta + \mu \\ \rho &\neq 0, \gamma = 0, \lambda = 0 \end{aligned} \tag{4.7}$$

Likewise, the spatial error model can be defined as in equation 4.8. The spatial error model incorporates the spatial effect as omitted spatially correlated variables uncorrelated to X. We use the spatial Hausman test when testing for correct specification of the spatial error model. The null hypothesis is that spatial effect can be capture by the spatial error. Therefore, rejecting the null hypotheses suggests that the spatial error model is not the best spatial model for capturing the spatial effect, in case the spatial effect (λ) is statistically significant.

$$\begin{aligned} y = X\beta + \epsilon, \quad & with\ \epsilon = \lambda W \epsilon + \mu \\ & \rho = 0, \gamma = 0, \lambda \neq 0 \end{aligned} \tag{4.8}$$

According to the idea of previous section, even if spatial randomness is rejected with the ESDA, the spatial dispersion of homicide could not be the only explanation for the nonexistence of spatial randomness and we have to control for spatial heterogeneity. Therefore, a formal spatial model must take into account the possible existence of spatial heterogeneity. As mentioned before, controlling for spatial heterogeneity is important due to a possible spatial correlation that is not related to a spatial dispersion of homicide but a geographical distribution of it trough other determinants. Therefore, to be theoretically able to contrast the spatial lag and the error models, they will include the control variable for heterogeneity described in 3.2.1 in the covariates matrix X.

4.5 Clustering Model

It is a common place in Mexico to think that it is a multi-diverse country that has cities with a economic development similar that Portugal and, at the same time, other cities with development indexes alike low-income countries in Africa. This diversity presents tough challenges when trying to analyze any social phenomenon just like homicide. While it is true that there are some clusters of cities alike, they do not gather in a single place along the country. Instead, similar cities can be found in the north, west and south, as well as surrounding by similar cities o completely different ones. The location of these conglomerates responds more to the economic activity, and to the social, political, or geographical characteristics of the cities than to a simple spatial distribution of them. For this reason, this thesis will stress that multi-diverse hypothesis by grouping cities according to social, economical, and criminal attributes.

Following the idea of different possible causal processes within the country in different geographical areas presented by Baller et al. (2001). In this section a different hypothesis of dissimilar impact of the predictors will be tested. This division of cities will not be based on the geographic division, but rather on a socioeconomic grouping by running a K-mean clustering model. Covariates used in the clustering model will be selected based in their non-correlation with dependent and independent variables in the base line model described in equation 4.2. This new grouping of cities is expected to reflect socioeconomic and criminal characteristics between cities. The 98 variables included were normalized around the zero value and will come from the economic, social and crime data sets at the city-level mentioned in section 3, such as street robbery, share of businesses dedicated to agriculture and so on (all crime variables are presented as a rate per 100 thousand inhabitants). Including the business variables is important due to it could reflect particular characteristics of a city. Just as a remark, the amount of economic units/businesses in Mexico went from 4,349,900 in 2015 to 4,821,703 in 2020 (almost 11% of growth), but this change has its particularities per city, based on the economic characteristics of them. Due to absolute variables tend to be correlated, this work uses variables as share of the total amount for each category, for instance the share of fixed businesses respect all fixed and semi-fixed units. The selection of the most appropriated number of clusters will be through the withinss (total within-cluster sum of squares). This indicator gives aggregated distance of every city to the center of the cluster that is belongs to. Therefore, a lower withnss indicates a higher concentration within the clusters.

In order to recognize the importance of each variable in the clustering model process, a random forest classification model will be used. For this variable importance selection we use the Mean Decrease Accuracy of variable permutation in the Random Forest process. Finally, a robust base line model described in equation 4.2 will be tested for each cluster and they will be compared along the groups.

Table 4.1: Summary of covariates used by previous researches on Crime and Homicide

Authors, Method and Unit of Analysis	Variables	Main Results
Land et al. (1990). OLS for 3 periods at the level of U.S. cities, metropolitan areas and states	1. Population Size, 2. Population Density, 3. Percentage Black, 4. Percentage ages 15-29, 5. Percentage Divorced, 6. Percentage kids not with both parents, 7. Median Family Income, 8. Percentage families in Poverty, 9. Gini Index of income inequality, 10. Percentage Unemployed, and 11. South	The theoretical assumption of invariant relationship between regressors and regressed variables seems to be invariant over time and social space when controlling for multicollinearity between regressors
Baller et al. (2001). Spatial lag and error models at the county-level	Resource deprivation component consists of 1. percentage black, 2. median family income (log), 3. Gini index of income inequality, 4. percent of families in poverty, and 5. the percent of families that are female headed. The population structure component comprises 6. population size (log) and 7. population density (log). The models also include 8. median age, 9. unemployment rate, 10. percent divorced, and a 11. Southern dummy variable, 12. spatial lag and error effects	1. Homicide is strongly clustered in space, 2. Formal space models are needed to explain the clustering, 3. Presence of regional differences in the effect of regressors on homicide rates
Gaigné and Zenou (2015). Monocentric model. City with individuals as unit of analysis	Firms, criminals and workers	1. Accessibility to jobs decreases crime in the short run but increases it in the long run 2. Per-capita crime rate increase with city size
C. Vilalta and Muggah (2014). Logistic binary regression model and Moran's I coefficient. At the police and electoral levels	1. Population bornt in another state, 2. Females between 6 and 11 that do not attend school, 3. Population above 15 with more than 9 years of schooling, 4. Average schooling among male population, 5. Population with employment, 6. Population ascribed to the ISSSTE social security, 7. Population ascribed to Seguro popular, 8. Population over 12 that is married, 9. Vacant housing, 10. Temporary housing, 11. Occupied home units with land floor, 12. Occupied home units with no access to water inside the premises	1. Evidence for spatial autocorrelation, and 2. Some variables are correlated in opposite sign of what is expected by the Social Disorganization Theory

Chapter 5

Results

5.1 Baseline Model

Both of Land et al. (1990) and C. Vilalta and Muggah (2014) models were tested separately as a starting point, nevertheless neither of them resulted in a significant and relevant model. The first differs in the variable selection according to Mexican society context and data available. Covariates like south and percentage of black people does not have a meaning in Mexican context, and variables like Gini coefficient and Median family income are not available in Mexican official data set for the years under study. On the other hand, the second differs in the model specification due to C. Vilalta and Muggah (2014) uses a logistic model, as well as in the economic and statistics significance due to variables like population born in another state, and population with employment resulted in non-significant coefficients. However, analyzing those two models and particularities of Mexican context resulted in the model specified in equation 4.1, which will be the base line model for all methods used in this work.

Following Land et al. (1990) and Baller et al. (2001), we tested a Principal Components Analysis for trying to get the two components found in their models. Nevertheless, our data did not fit as expected for this method and was not possible to capture the variance in only two components. Therefore, we will continue with all variables included in the model definition. First approach in this thesis uses the Ordinary Least Square model specified in equation 4.1. It was tested for all year between 2015 and 2020, as well as for the entire panel as a whole, where every row represents a city's data in the t year ($t \in [2015, 2020]$). Nevertheless, a robust regression model was implemented due to all seven models resulted in a violation of the non-heteroskedasticity assumption according to the Breusch-Pagan test (see appendix A). Results of the robust regression and a comparison of the residual standard errors (RSE) against the OLS models are shown in table 5.1. RSEs for the OLS models are almost 3 times bigger than robust models for all models, and thus the robust regression model performances better than OLS and suggests that the first is the best option for this data. Interestingly, for the whole period model (column 7 in table 5.1), the 12 covariates, intercept, and regional dummies were statistically significant at the

0.001 level, which gives a strong support to variables selection in the model specification process. However, due to variability in homicide rates over years, explanatory variables like the share of males and unemployment were not significant for four and two years respectively. According to that variability in results, we will focus our analysis in the 7th model in table 5.1.

According to the report on money laundry of the Mexican Financial Intelligence Unit, by 2020 19 organized groups where detected in the national territory[1] and it is possible to see that their geographical location have a particular aspect for every organized crime region. Region 5 in the south is the reference region for the analysis and also the one with the lowest homicide rates over years. This region is characterized for having a 6 different criminal organizations along the region, but neither of them control big regions, mainly because this region is not high-profit area due to the underdevelopment respect the rest of the country. In addition, ONGs and journalists[2] in Mexico have argued that the "*peace*" present in region 5 is because local police offices have used methods against suspects that attempt their human rights, such as torture; therefore, population is afraid of committing any act that can be interpreted as an even minor crime, because they have no guarantee that their human rights will be respected. With region 4, on the other hand, starts the geographical location of big, old and powerful organized groups such as the Cartel of Michoacan, whom controls some cities in the region known as "*Tierra Caliente*" and not even military goes into that area. This region has in average 4.15 more homicides per 100,000 inhabitants. Region 3 is a relatively new crime area, even when it is the origin of the most powerful cartel in Mexico at the moment, the Cartel de Jalisco Nueva Generación (CJNG). This cartel has a constant dispute in the north with the Cartel de Sinaloa, in the south with the Cartel de Michoacán and to the east with the new Cartel de Santa Rosa de Lima. The last organized group is a perfect example of the complexity of organized crime in Mexico, it started as an organization that controls the black market of gasoline, in Salamanca, Guanajuato, but when the government applied a strategy to defeat the whole black market, they used their organizational structure to enter a new market and thus obtain the income lost with the implementation of the government's strategy. Thus starting with the collection of a fee to all businesses in their area and if any refused to pay the fee, then they were killed. At the same time, they went to war against the CJNG to control the area. This region 3 has a homicide rate higher than region 5 by 5.84. The region 2 is mainly occupied by two organized groups the Cartel del Golfo and the Cartel del Noreste, the first is one of the oldest organized crime in Mexico, It started in the Prohibition era, with the smuggling of alcohol into the US. This region has in average a higher homicide rate than region 5 by 2.34. Finally, with an average of 11.52 more homicide rate than region 5, region 1 is the most violent zone in Mexico. This region is mainly controlled by the Cartel de Sinaloa (also known as the Cartel del Pacífico) but

[1]https://www.eluniversal.com.mx/estados/el-cjng-el-de-mayor-penetracion-en-el-pais
[2]https://www.animalpolitico.com/hojas-en-el-cenicero/torturar-en-nombre-de-la-paz/

maints a constant fight against the CJNG. Joaquin "*El Chapo*" Guzmán, the head of this cartel for a long period (now he is in prison in Colorado, US) was considered one of the richest people on the globe by Forbes lists in 2009-2012 editions.

LPOBTOT, SLand, SFemaleAh, SISSSTE, SDivorced, SFem6to11NAS, and YearsSchool covariates resulted with the sign impact expected by previous works and social crime theories. SUnemploy presents a negative sign, which differs with what is expected by theories but follows previous articles. Nevertheless, four covarite coefficients suggest a deeper analysis: 1. LPopDensity, contrary to previous works, density logged presents a negative sign, which implies that a decrease of 1% in cities density is related with an decrease of 0.91 in the homicide rate. This can be explained by the fact that organized crime homicide reports occur usually, but not exclusively, in places where the crime can be hidden from police or ONGs, that is places with not a high population density. 2. SIndi, we used the percentage of indigenous people in a city trying to capture a similar relationship than the black variable in previous works; but it ended up to have a negative sign with a coefficient of -0.05, contrary to previous works, this could be explained by the fact that indigenous people are more concentrated in the south but in general it is also the poorest zone of Mexico, and so a not very attractive area for organized crime. Another important variable is the percentage of males in a city, the model shows that an increase of one percent in the share of males implies an increase of 0.62 homicides per 100,000 inhabitants. Finally, the share of elderly population resulted to have a negative impact in homicide rates, specifically, an increase of one percentage in the share of old people is related with a decrease of 0.47 in the homicide rate per 100 thousand inhabitants. This is particularly interesting because it captures the deep cultural tradition that Mexicans depend on older people. In general, having a grandfather or grandmother makes the family ties stronger.

Finally, it is interesting how in models for years after 2018 the dummy variable for Region 2 became non-significant, and so impacts of some covariates had a drastic change. SFem6to11NAS changed from 6.27 to 10.57, SLand from 0.20 to 0.36. Those changes could be explained by the fact that Mexico had presidential elections in 2018 and the political party in charge changed. Due to previous agreements that the government had with organized groups, it is well known by Mexican media that a new administration usually brings regional reactions from organized crime groups and could change homicide behavior in the first years of the new political party or group in power.

5.2 Panel Data Model

Next step in this analysis process is to test the suitability of error component models and look for possible specific effects on cities. Table 5.3 shows the three different panel data models explored (Pooled model, Fixed Effects Model and Random Effects Model). All three models used a clustered standard errors for cities, which accounts for possible heterogeneity or potential correlation over time within cities in the error term. Interestingly, all coefficients of both, the pooled and the random effects models, are statistically significant at the 5% level or better, nevertheless, none of the fixed effects model coefficients is statistically significant. In general terms, the random effects model keeps a similar impact size than the pooled model. In the fixed effects model all invariant variables are omitted due to invariant variables cancel themselves, the Region dummy variables and the density were omitted for the model (density has a marginal change from a year to another).

Next step is using tests to determine which model is the best. Firstly, the F test for individual fixed effects rejects the null hypothesis and there fore, we can assume that fixed effects is a better choice for our specified model. Secondly, when testing for the existence of fixed effects, we failed to reject the null hypothesis of the Breusch-Pagan Lagrange multiplier test, which is that there is not panel effect and, therefore the pooled model is better, *i.e.*, we can infer that randoms effects exist. Finally, due to both of the fixed and random effects are present, we ran a Hausman test for the non-clustered standard errors and conclude that the fixed effect is the best option for our data. However, none of the coefficients was significant.

5.3 ESDA and spatial models of City Homicides Rates

After testing for error components, this thesis also examines the possibility of spatial dispersion of homicide. First, we analyzed the possibility of a global autocorrelation through the global Moran's I for homicide rates in the period 2015-2020 and a sum for the whole period (see Table 5.5). As was expected for previous literature and social crime theories, all 7 coefficients are statistically significant at the 0.001 level and reject the null hypothesis of spatial randomness. That gives strong evidence of a significant spatial pattern of homicide in cities across Mexico. Moreover, figure 5.1 shows the results of the local clusters of the accumulated homicide rate through the local Moran's I test. Those local clusters were result of a statistically significance at the 0.1 level and represents 4 possible scenarios: 1. a city with high homicide rate surrounding by cities with high homicide rates (high-high), 2. a city with high homicide rate surrounding by cities with low homicide rates (high-low), 3. low-low relationship, and low-high relationship. It can be seen in that figure that organized crime regions one, two, and three share a large area with local clusters of cities with high rate of homicide surrounding by cities with high homicide rates. Regions 3 and 4 also share a big area of high-high homicide rate cities. On the other hand, region 5 is the only region with the presence of cities with low-low homicide rates. This spatial distribution of local clusters supports the use of the regional variable for controlling for a potential spatial heterogeneity and shows the possible differences in country as big and diverse as Mexico.

Another way to account for local autocorrelation is the Getis-Ord Gi test. Figure 5.2 shows the results for this test and can be seen qualitatively how this differs from

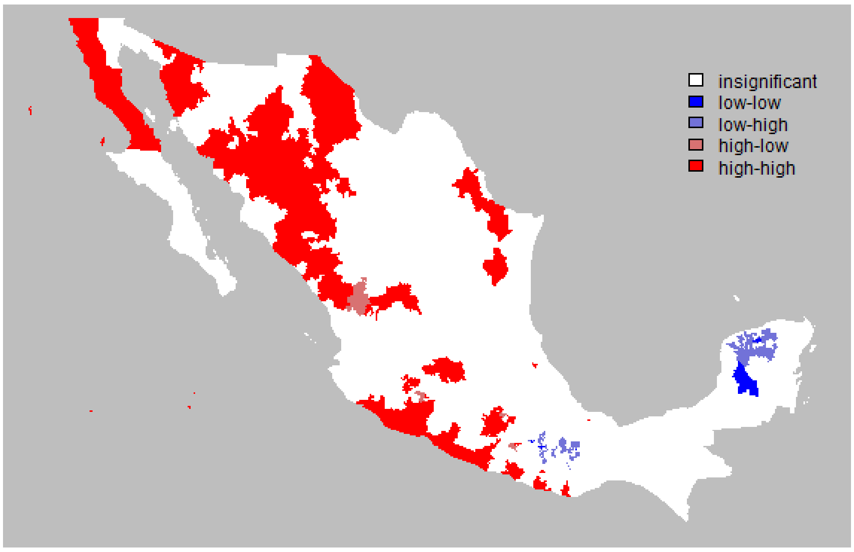

Figure 5.1: Local Moran's I for the entire period 2015-2020

the local Moran's I test in the way that Getis-Ord Gi test accounts for the level of local accumulated homicide rate clustering for a city compared with the national city average. Interestingly, at some extent it addresses one of the most fundamental concepts of the spatial diffusion process, the initial shock. The initial shock represents the origin of the homicide diffusion process (innovation's location) that will be spread and adopted by neighbor cities (adopters) over time and space. In this sense, figure 5.2 shows a sort of diffusion process, for instance in the northwest there is a probable initial shock around the cities of Chihuahua and Delicias, and from those cities the homicide seems to be spread over space. And so, almost all region one is covered by a big local cluster of high homicide rates. Moreover, region three seems to have two initial shocks, one in the center of Mexico, in the states of Guanajuato, Michoacá and Jalisco, which is not surprising due to it is the region where three cartels have been fighting for the territory, and the second coast, where another fight for territory is happening but mostly between two organized crime groups. It is also interesting how there is local cluster of low crime in the state of Oaxaca surrounding for high homicide rate cities.

5.3.1 Spatial Homicide Models

Now we have evidence for global and local spatial autocorrelation of homicide between Mexican cities, the next step is to test the spatial lag and the spatial error models in our data (equations 4.7 and 4.8). Running those models per year drives to very volatile coefficients' significance between them due to a high variance in homicide rates along years. In addition, even when all spatial lag and error effects were statistically significant at the 0.001 level for all yearly models, only those for 2018 failed at rejecting the null hypothesis of the spatial Hausman autocorrelation test for the spatial error model and the Lagrange Multiplier test for the spatial lag model. This means that only for the

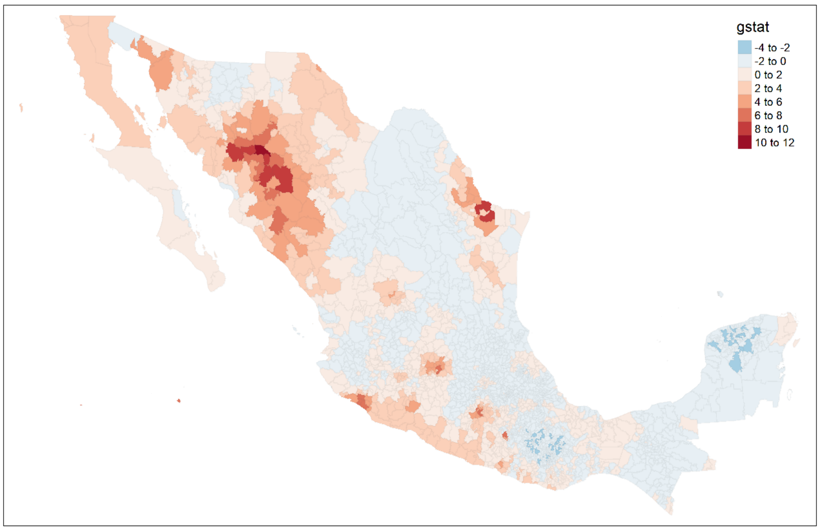

Figure 5.2: Getis-Ord Gi for the entire period 2015-2020

year 2018 our spatial models were the correct specification to capture the spatial effect of homicide. Therefore, this thesis followed the idea of Baller et al. (2001) about smoothing the variance of homicide and uses a six-year average homicide rate for the 2015-2020 period as the regressed variable with covariates from the 2020 data set[3]. All spatial effects for those two models were statistically significant at the 0.001 lever. Nevertheless, neither of them rejected the null hypothesis of misspecification for the spatial lag and error models (Lagrange multiplier and spatial Hausman tests, respectively).

These results are not surprising because Mexico is a very diverse country and finding a model that captures the entire national dynamic is complicated. Therefore, in order to account for regional differences within the country, this thesis follows the idea carried out by Baller et al. (2001) of looking for regional spatial models that explain better the spatial process of homicide. Consequently, the spatial lag and the spatial error model specifications were tested within each one of the five organized crime regions described in section 3.2.1. Table 5.6 shows the results of the spatial models that better adjusted to the regional spatial dynamics within the regions.

Spatial Homicide Dispersion Within Regions

Based on the results for spatial autocorrelation and the spatial models for the entire country, this thesis examines a disaggregated modeling strategy for each of the organized crime regions. Both of the two spatial model results are shown in table 5.6. Selection of the best model approach for the homicide rate in regions was based on the specification tests for each model described in section 4.4. Interestingly, alike previous literate's results for the US regions south and non-south (Baller et al., 2001), Mexican regions also present

[3]This thesis uses six instead of three years average due to in 2018 Mexico had a presidential secession and it usually brings regional reactions from organized crime and it could have changed the homicide behavior for the first years of the presidential period

different spatial modelling specifications for different regions. In this sense, regions one and five suggest a need for a spatial error model specification, meanwhile regions two and three suggests a spatial lag model specification. Region two suggests that neither of them is the spatial model specification needed for addressing the spatial diffusion process of homicide between Mexican cities. Nevertheless, the spatial lag model specification was chosen as the best model specification due to it was the closest to reject the null hypothesis of the specification test.

Begging with results for regions with spatial error specification, on one hand, we can see in region one an important change in significance and magnitude of the coefficients' covariates respect the robust regression for non-spatial analyses. Only three regressors are statistically significant, but their signs are consistent with results for non-spatial analyses. However, those coefficients increased dramatically their impact in the homicide rate. Moreover, the spatial error coefficient (λ) is statistically significant at the 0.01 level. On the other hand, the only region with a low-low local cluster in Mexico, region five, presents the smallest changes in coefficients' magnitude respect the non-spatial models. It also has the highest number of significant variables and the spatial error coefficient is significant at the 0.001 level. These results indicate that the residual spatial autocorrelation in regions one and five are better accounted for by the unmeasured predictor variables.

Now we will focus on regions two and three, where the best spatial fit was the spatial lag model. Firstly, region two resulted in a model where the only significant variable is the spatial lag coefficient (ρ), with a significance level of 0.001. This implies that the only variable that affects the rate of homicide in a city is the homicide rates of its neighbors. Secondly, region three resulted in a model with only three significant variables, which do no present big changes compared to the non-spatial models. The third variable, the unemployment rate is interesting due to it presents a bigger coefficient with an opposite sign than the non-spatial models. This change in sign makes it consistent with social theories.

5.4 Clustering Model

The final stage in this thesis is to look for clusters derived from social, economical and criminal characteristics of the cities. All 98 variables included in this process must be uncorrelated any covariate used in our base model (equation 4.2). The optimal selection of the number of clusters that dive better our data indicated that either five or six clusters are at the elbow of the relationship between the number of clusters against the total within-cluster sum of squares. In order to decide between them two, we took a look at the clusters size. Having five clusters of cities will derive in the first one with 378, second one with 276, third one with 1,143, fourth one with 65, and the fifth one with 594. Meanwhile if we go for the six cluster choice, it will basically divide the 594 cities one into two and won't change the biggest one. We are looking for clusters that capture the closeness of cities, but at the same time we are looking for a meaningful clustering, in terms of practicality. Therefore, the five clusters option was selected as the grouping that will drive this part of the analysis.

Two possible visualizations are shown in figures 5.3 and 5.4. The first describes the location of clusters along Mexico. Cluster number one is also distributed along the country but in cities with a medium economic development. Cluster number 3 is highly concentrated in the north in cities extremely industrialized and with a strong economic link to the US. Cluster number three is distributed along the country in areas with a low level of economic development. Cluster number four is mainly located in the south, in the zone where economic development is particularly low. Finally, cluster number five represents the mainly the metropolitan area of Ciudad de Mexico and other cities with high economic development for Mexican standards.

The second figure describes the results of the classification with random forest. It shows the particular differentiation between cities for the three variables in the top four of variables that influenced the most in the clustering process (variable importance estimated by using the Mean Decrease Accuracy of variable permutation in the Random Forest process.): 1. Private dwellings inhabited without any good, 2. Percentage of businesses in the city that belong to the manufacturing industries, and 3. Occurrence of domestic violence per 100 thousand inhabitants. The average behavior per cluster for that three specific characteristics is (as a comparison between clusters):

- cluster 1: high rate of domestic violence, low share of Private dwellings inhabited without any good, and medium share of Manufacturing industries;
- cluster 2: medium share of Private dwellings inhabited without any good, and low share of Manufacturing industries;
- cluster 3: medium rate of domestic violence, medium share of Private dwellings inhabited without any good, and medium share of Manufacturing industries;
- cluster 4: high rate of domestic violence, low share of Private dwellings inhabited without any good, and low share of Manufacturing industries, and
- cluster 5: medium rate of domestic violence, high share of Private dwellings inhabited without any good, and high share of Manufacturing industries.

The base model of this thesis has been tested within each cluster by using the OLS model and a robust regression model. Results are shown in table 5.8 where can be seen that the two biggest clusters are those with more covariates that are statistically significant. Likewise, most of the magnitudes are similar to the non-clustered models, nevertheless some particular coefficients are interesting for some clusters. First, Cluster one shows a huge significant coefficient for the share of females aged 6-11 who do not attend school, suggesting that an increase by 1% in the share of females aged 6-11 who do not attend

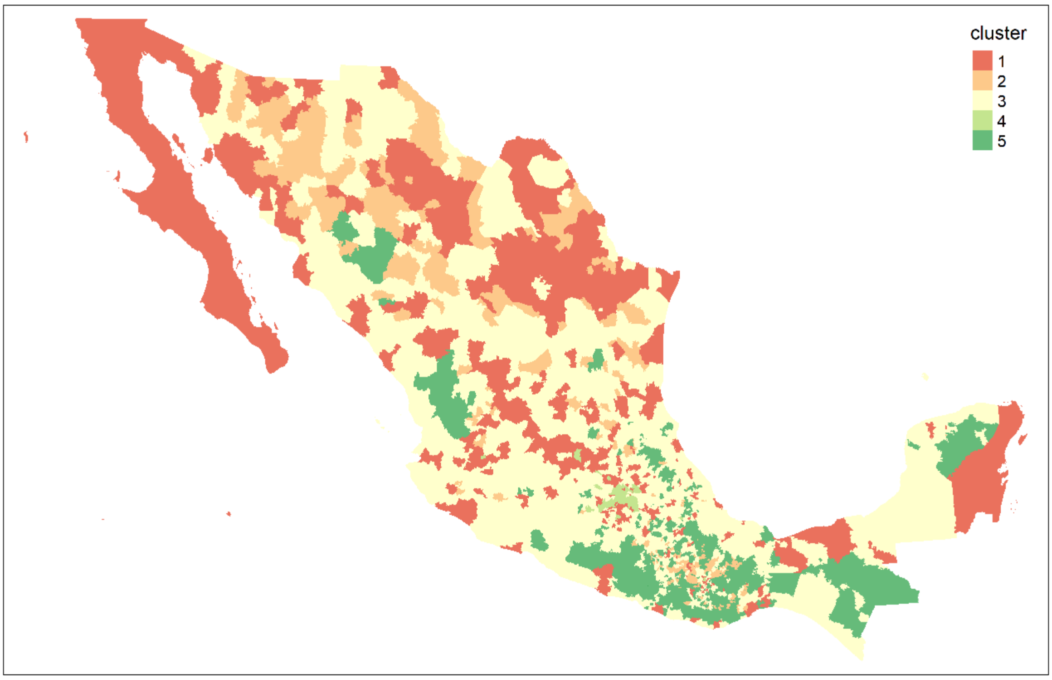

Figure 5.3: Clusters distribution along Mexico (2020)

school is related with an increase of 34.38 homicides per 100 thousand inhabitants. Cluster number three is the only one that presents a positive and significant relationship for all non-spatial models explored in this thesis, which is compatible with social crime theories. Cluster four concentrates the entire significant effects in covariates related with education. Finally, cluster five, where concentrates the higher significant effect in the dummy variable for region 1. Specifically, can be inferred that cities in cluster five that belongs to region one, present 74 homicide per 100 thousand inhabitants. All those results might be taken with precaution because they could present the selection bias, nevertheless, they do give an insight on the huge Mexican diversity and how it is correlated with the phenomenon of homicide.

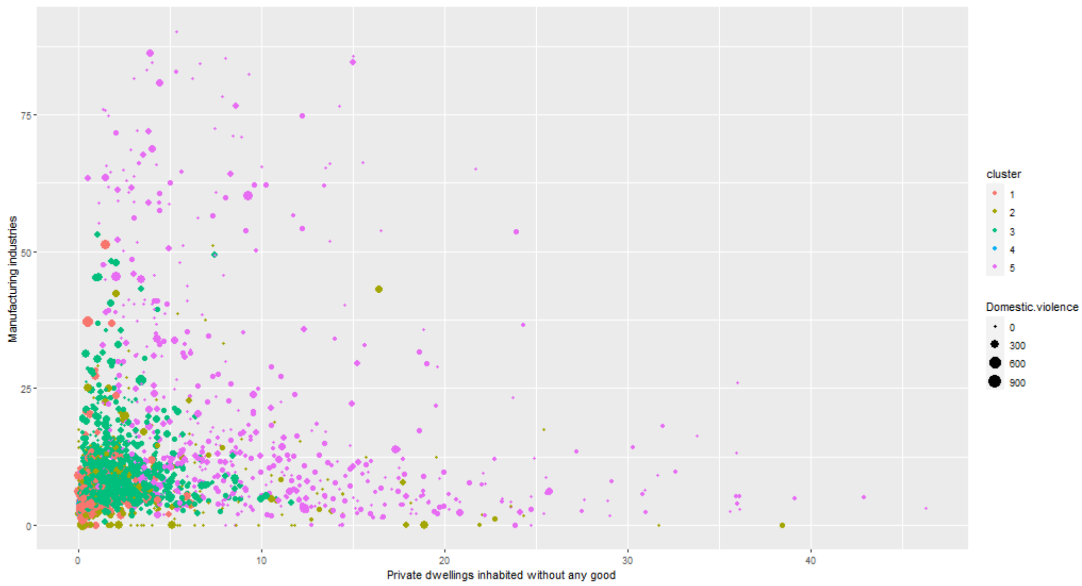

Figure 5.4: Characterization of clusters in socioeconomic and crime characteristics (2020)

Table 5.1: Social covariates effect on Homicide rates for Robust Regression Model

Covariates	(1) 2015	(2) 2016	(3) 2017	(4) 2018	(5) 2019	(6) 2020	(7) 2015-2020
Intercept	-46.20***	-69.81***	-45.78***	-30.13*	-38.68*	-32.87*	-52.09***
	(9.43)	(10.55)	(11.73)	(15.28)	(15.07)	(13.34)	(5.06)
LPOBTOT	1.81***	2.66***	2.94***	2.39***	2.44***	2.68***	2.46***
	(0.18)	(0.21)	(0.23)	(0.31)	(0.32)	(0.29)	(0.10)
LPopDensity	-1.16***	-0.88***	-0.83***	-0.90**	-0.66*	-0.69*	-0.91***
	(0.19)	(0.21)	(0.24)	(0.31)	(0.32)	(0.29)	(0.10)
SIndi	-0.03***	-0.05***	-0.05***	-0.06***	-0.06***	-0.06***	-0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
SLand	0.20***	0.28***	0.20***	0.36***	0.29***	0.29***	0.25***
	(0.03)	(0.04)	(0.05)	(0.06)	(0.06)	(0.06)	(0.02)
SFemaleAh	0.21***	0.34***	0.17*	0.30***	0.23**	0.22**	0.29***
	(0.06)	(0.06)	(0.07)	(0.09)	(0.08)	(0.07)	(0.03)
SISSSTE	-0.73***	-0.55***	-1.13***	-0.94***	-0.83***	-0.36	-0.73***
	(0.13)	(0.15)	(0.17)	(0.22)	(0.22)	(0.20)	(0.07)
SMale	0.58***	0.92***	0.42	0.23	0.47	0.32	0.62***
	(0.18)	(0.20)	(0.22)	(0.29)	(0.28)	(0.25)	(0.09)
S65More	-0.35***	-0.26**	-0.40***	-0.62***	-0.62***	-0.56***	-0.47***
	(0.09)	(0.10)	(0.11)	(0.14)	(0.14)	(0.12)	(0.05)
SUnemploy	-0.73**	-0.78*	-0.91*	-1.10*	0.61	0.21	-0.73***
	(0.28)	(0.32)	(0.37)	(0.48)	(0.46)	(0.40)	(0.15)
SDivorced	0.82***	1.01***	1.59***	1.68***	1.77***	1.51***	1.44***
	(0.19)	(0.21)	(0.22)	(0.28)	(0.27)	(0.23)	(0.09)
SFem6to11NAS	6.58***	4.80***	6.27***	10.57***	6.09**	5.93***	7.44***
	(1.18)	(1.34)	(1.49)	(1.95)	(1.93)	(1.71)	(0.64)
YearsSchool	-0.32	-0.98***	-1.16***	-1.33***	-1.76***	-1.57***	-1.11 ***
	(0.21)	(0.23)	(0.26)	(0.34)	(0.34)	(0.30)	(0.11)
DummyR1	13.09***	9.71***	17.19***	12.75***	11.20***	8.55***	11.52***
	(0.99)	(1.13)	(1.27)	(1.68)	(1.70)	(1.54)	(0.56)
DummyR2	2.94***	1.87*	3.00**	1.86	2.14	1.13	2.34***
	(0.76)	(0.86)	(0.98)	(1.30)	(1.31)	(1.19)	(0.43)
DummyR3	5.38***	4.44***	5.77***	7.63***	7.22***	5.86***	5.84***
	(0.64)	(0.72)	(0.81)	(1.07)	(1.08)	(0.98)	(0.36)
DummyR4	5.39***	3.21***	4.60***	4.09***	4.61***	2.01**	4.15***
	(0.50)	(0.56)	(0.63)	(0.84)	(0.85)	(0.78)	(0.28)
Observations	2,456	2,456	2,456	2,456	2,456	2,456	14,736
RSE-OLS	21.43	21.1	21.4	35.6	26.9	43.0	29.6
RSE-robust	7.7	8.8	9.9	13.6	13.7	2.8	10.9

*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$

Standard errors in parenthesis

Table 5.3: Covariates effect on Homicide rates for Panel Data Models (Clustered SE)

Covariates	(1) Pooled	(2) Fixed	(3) Random
Intercept	-75.88 ***	-240.80	-91.62 ***
	(12.61)	(470.18)	(20.37)
LPOBTOT	1.21 ***	10.27	1.14 **
	(0.28)	(34.19)	(0.42)
LPopDensity	-1.18 ***		-1.26 **
	(0.27)		(0.45)
SIndi	-0.08 ***	-0.10	-0.06 ***
	(0.01)	(0.41)	(0.01)
SLand	0.36 ***	-0.13	0.29 ***
	(0.04)	(0.21)	(0.07)
SFemaleAh	0.55 ***	0.69	0.67 ***
	(0.08)	(0.66)	(0.12)
SISSSTE	-0.96 ***	-3.27	-0.99 ***
	(0.25)	(2.04)	(0.29)
SMale	1.31 ***	1.32	1.55 ***
	(0.24)	(3.91)	(0.40)
S65More	-0.78 ***	0.56	-0.78 ***
	(0.13)	(2.99)	(0.18)
SUnemploy	-0.82 *	-0.08	-1.34 **
	(0.40)	(2.00)	(0.58)
SDivorced	2.54 ***	3.04	2.68 ***
	(0.24)	(2.44)	(0.36)
SFem6to11NAS	12.30 ***	11.21	14.03 ***
	(2.57)	(8.96)	(3.70)
YearsSchool	-2.35 ***	6.70	-2.20 ***
	(0.33)	(3.51)	(0.49)
DummyR1	22.61 ***		21.92 ***
	(1.80)		(3.01)
DummyR2	7.06 ***		7.03 ***
	(1.52)		(1.95)
DummyR3	10.37 ***		10.32 ***
	(0.81)		(1.34)
DummyR4	6.44 ***		6.58 ***
	(0.56)		(0.88)
Years	6	6	6
Cities	2,456	2,456	2,456
R^2	0.104	0.015	0.103
σ_u/σ_e		31.03/25.46	15.10/25.46
ρ		0.60	0.26

*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$ (Standard errors in parenthesis)

Table 5.5: Global Moran's I Statistics: Homicide Rates

Year	I statistic	Interpretation
	A. Annual homicide rate	
2015	0.307 ***	Spatial Clustering
2016	0.349 ***	Spatial Clustering
2017	0.389 ***	Spatial Clustering
2018	0.239 ***	Spatial Clustering
2019	0.279 ***	Spatial Clustering
2020	0.167 ***	Spatial Clustering
	B. Average homicide rate	
2015-2020	0.460***	Spatial Clustering

*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$ (two-tailed tests).
Empirical Significance based on 999 Random Monte-Carlo Simulations.

Table 5.6: Covariates effect on Homicide rates for spatial models per region

Covariates	(1) Region 1	(2) Region 2	(3) Region 3	(4) Region 4	(5) Region 5
Intercept	50.8	23.55	-43.52	-62.62*	-17.74
	(108.14)	(93.65)	(41.87)	(29.87)	(21.20)
LPOBTOT	1.63	-0.46	1.27	0.86	1.61***
	(2.67)	(2.01)	(1.01)	(0.48)	(0.48)
LPopDensity	-2.80	-1.47	0.22	-1.88***	-0.74
	(2.58)	(1.51)	(0.90)	(0.56)	(0.60)
SIndi	-0.23	-0.05	-0.19*	-0.00	-0.06***
	(0.38)	(0.13)	(0.10)	(0.02)	(0.01)
SLand	2.85***	0.59	0.37	-0.10	0.17*
	(0.84)	(0.55)	(0.31)	(0.11)	(0.08)
SFemaleAh	0.82	0.39	0.03	0.28*	0.40***
	(0.60)	(0.46)	(0.18)	(0.13)	(0.11)
SISSSTE	-0.66	5.56	0.42	-0.32	-0.78
	(0.71)	(3.60)	(2.69)	(0.28)	(0.94)
SMale	-0.30	-0.80	0.51	1.05	0.30
	(1.75)	(1.67)	(0.79)	(0.58)	(0.38)
S65More	-0.90	0.58	-0.15	-1.02***	-0.47**
	(0.96)	(0.63)	(0.4)	(0.25)	(0.19)
SUnemploy	-4.83*	-0.80	3.49*	-0.47	-0.11
	(2.13)	(2.30)	(1.59)	(0.78)	(0.59)
SDivorced	2.98	0.89	1.96**	2.13***	0.92*
	(1.92)	(1.19)	(0.73)	(0.43)	(0.39)
SFem6to11NAS	-7.12	19.67	7.74	9.35*	0.62
	(9.49)	(11.78)	(5.54)	(4.32)	(2.81)
YearsSchool	-6.99*	0.18	-1.38	-0.60	-1.83***
	(3.14)	(1.92)	(0.89)	(0.54)	(0.52)
Cities	2,456	2,456	2,456	2,456	2,456
Spatial error (λ)	0.34**	NI	NI	NI	0.37***
Spatial Lag (ρ)	NI	0.39***	0.75***	0.64***	NI
Spatial Hausman test	14.69				21.48
LM test		3.47	0.02	5.56*	

*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$ (Standard errors in parenthesis)

Table 5.8: Social covariates effect on Homicide rates for Robust Regression Model Within clusters

Covariates	(1) Cluster 1	(2) Cluster 2	(3) Cluster 3	(4) Cluster 4	(5) Cluster 5
Intercept	-286.22 ***	-46.68	-35.60 *	29.39	0.4
	(-47.81)	(-30.17)	(-16.99)	(-126.37)	(-14.36)
LPOBTOT	0.74	2.35	2.44 ***	-0.48	0.96 *
	(-0.68)	(-1.26)	(-0.35)	(-0.92)	(-0.41)
LPopDensity	-0.38	-1.29	-0.93 **	-0.23	-1.40 ***
	(-0.66)	(-1.05)	(-0.34)	(-1.11)	(-0.4)
SIndi	-0.06	-0.04	-0.07 ***	-0.37	-0.04 ***
	(-0.07)	(-0.03)	(-0.02)	(-0.29)	(-0.01)
SLand	1.50 ***	0.35	0.67 ***	-1.1	0.19 ***
	(-0.41)	(-0.19)	(-0.09)	(-0.87)	(-0.05)
SFemaleAh	0.94 ***	0.32	0.1	0.53	0.20 **
	(-0.26)	(-0.17)	(-0.09)	(-0.51)	(-0.07)
SISSSTE	-0.41	-0.90 **	-0.70 *	-0.85	0.21
	(-0.57)	(-0.29)	(-0.3)	(-0.55)	(-0.93)
SMale	5.03 ***	0.85	0.23	0.53	-0.03
	(-0.89)	(-0.52)	(-0.32)	(-2.39)	(-0.26)
S65More	0.14	-0.3	-0.17	0.52	-0.46 **
	(-0.42)	(-0.24)	(-0.16)	(-0.84)	(-0.16)
SUnemploy	3.05	-0.74	1.16 *	3.79	-0.21
	(-2.51)	(-0.84)	(-0.56)	(-5.02)	(-0.35)
SDivorced	2.37 **	0.39	2.28 ***	-0.65	0.64 *
	(-0.79)	(-0.47)	(-0.28)	(-1.59)	(-0.29)
SFem6to11NAS	34.38 ***	-0.01	11.10 ***	54.53 ***	4.45 **
	(-8.72)	(-3.43)	(-2.65)	(-13.21)	(-1.6)
YearsSchool	-1.68	-2.08 *	-2.02 ***	-4.56 **	-0.78
	(-0.93)	(-0.87)	(-0.36)	(-1.43)	(-0.41)
DummyR1	1.95	20.13 ***	19.32 ***		74.26 ***
	(-3.39)	(-4.56)	(-1.8)		(-4.55)
DummyR2	-1.59	4.02	3.22 *		9.19 *
	(-3.05)	(-4.16)	(-1.29)		(-3.68)
DummyR3	7.50 *	6.4	8.62 ***	-12.19	0.61
	(-3.07)	(-4.58)	(-0.97)	(-9.99)	(-2.54)
DummyR4	3.01	1.59	6.36 ***	-11.95	4.93 ***
	(-2.52)	(-3.94)	(-0.85)	(-8.32)	(-0.9)
Observations	378	276	1,143	65	594
RSE-OLS	23.65	19.61	14.71	7.19	6.99
RSE-robust	12.73	10.27	9.15	5.00	6.99

$^{*}p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$

Standard errors in parenthesis

Chapter 6

Conclusions and Future Work

The final goal of this thesis is to provide a general exploration of the phenomenon of homicide in Mexico from different approaches. Thus, various noteworthy findings for each type of analysis were identified.

First, there is a lack of prior literature on this topic in Mexico, and model definitions for other countries or for analyzing homicide within a city in Mexico were insignificant from a city-level perspective. That makes it really difficult to find the definition of the model that captures the Mexican context. This thesis provides a possible base model built from previous literature. Howbeit, it responds differently depending on the type of analysis driven. Future work should explore another specification of models to more efficiently address homicide occurrence in different societies in Mexico.

Second, social disorganization theory and crime patterns theory seem to be useful for the characterization of homicide in Mexico. However, the high variance of the significance of the coefficients between different model specifications suggests, for future works, a complete analysis in the definition of a model with less variance.

Third, unsurprisingly, this thesis found strong evidence of the spatial dispersion of homicide in Mexico at the city level and, therefore, homicide is not randomly distributed in space. Furthermore, the ESDA reveals that homicides are concentrated locally in different areas of Mexico. The highest and oldest concentration is in the northwest, where the Cartel de Sinaloa controls the zone. However, cold areas were also found, particularly in the state of Oaxaca, where small and well-organized communities have stopped the spread of organized crime and, therefore, homicide.

Fourth, no spatial model was well specified when analyzing the entire country, however, the residual spatial autocorrelation is highly significant for all years. Within organized crime regions, significantly different model specifications were found. For regions one and five the most appropriate model was the spatial error model, for regions two and three the spatial lag model was the best option and for region four neither of the two spatial models was the specification that explains the homicide dispersion process. These results suggest that homicide is not only explained by socioeconomic covariates. In order to control for the real impact of organized crime, future works could explore the specific location per city of different organized crime groups at the city level and its relation with homicide. Moreover, since yearly periods bring a high level of variability, future works should consider having a bigger space between time periods and smooth the homicide rate.

Fifth, as expected, homicide presents different patterns in different types of societies. It was found that the significance of coefficients changes a lot between clusters or organized crime regions, suggesting that different model specifications might be needed for different clusters and/or organized crime regions. The cluster with less homicide concentration is where the previous literature's model for the US specification fitted the best. Likewise, industrialized low-income cities are extremely vulnerable in the north. Future works should explore theoretically and empirically this possible difference between different regions contexts.

Finally, future works in Mexico that analyze this phenomenon over time must include a control for public policies for fighting against organized crime within a region and over years.

References

Baller, R. D., Anselin, L., Messner, S. F., Deane, G., & Hawkins, D. F. (2001). Structural Covariates of U.S. County Homicide Rates: Incorporating Spatial Effects. *Criminology*, *39*(3), 561–590. doi: 10.1111/j.1745-9125.2001.tb00933.x

Becker, G. S. (1968). Crime and Punishment: an Economic Approach. *Journal of Political Economy*, *76*(2), 169–217. doi: 10.1086/259394

Bernasco, W. (2019). Adolescent offenders' current whereabouts predict locations of their future crimes. *PLoS ONE*, *14*(1), e0210733. doi: 10.1371/journal.pone.0210733

Bernasco, W., Ruiter, S., & Block, R. (2017). Do Street Robbery Location Choices Vary Over Time of Day or Day of Week? A Test in Chicago. *Journal of Research in Crime and Delinquency*, *54*(2), 244–275. doi: 10.1177/0022427816680681

Bernasco, W., & Steenbeek, W. (2017). More Places than Crimes: Implications for Evaluating the Law of Crime Concentration at Place. *Journal of Quantitative Criminology*, *33*, 451-467. doi: 10.1007/s10940-016-9324-7

Brantingham, P. J., Brantingham, P. L., & Andresen, M. A. (2017). The geometry of crime and crime pattern theory. In R. Wortley & M. Townsley (Eds.), *Environmental criminology and crime analysis* (p. 98-116). New York, NY: Routledge.

Estévez-Soto, P. R. (2020). Organised Crimes and Repeat Victimisation: Modelling victimisation pattern of extortion against Mexican business. *Doctoral thesis (Ph.D), UCL (University College London).*

Fijnaut, C. (2014). SEARCHING FOR ORGANIZED CRIME IN HISTORY. In L. Paoli (Ed.), *The oxford handbook of organized crime* (p. 53-95). New York, NY: Oxford University Press.

Gaigné, C., & Zenou, Y. (2015). Agglomeration, city size and crime. *European Economic Review*, *80*, 62–82. doi: 10.1016/j.euroecorev.2015.08.014

Hobbs, D., & Antonopoulo, G. A. (2014). HOW TO RESEARCH ORGANIZED CRIME. In L. Paoli (Ed.), *The oxford handbook of organized crime* (p. 96-117). New York, NY: Oxford University Press.

INEGI. (2019, October). Encuesta nacional de victimización y percepción sobre seguridad pública (envipe) 2019: Principales resultados. *Instituto Nacional de Estadística y Geografía*.

Kleemans, E. R. (2014). THEORETICAL PERSPECTIVES ON ORGANIZED CRIME. In L. Paoli (Ed.), *The oxford handbook of organized crime* (p. 32-52). New York, NY: Oxford University Press.

Kuang, D., Brantingham, P., & Bertozzi, A. L. (2017). Crime topic modeling. *Crime Science*, *6*(12). doi: 10.1186/s40163-017-0074-0

Land, K. C., McCall, P. L., & Cohen, L. E. (1990). Structural Covariates of Homicide Rates: Are There Any Invariances across Time and Social Space? *American Journal of Sociology*, *95*(4), 922–963. doi: 10.1086/229381

Medel, M., & Thoumi, F. E. (2014). MEXICAN DRUG "CARTELS". In L. Paoli (Ed.), *The oxford handbook of organized crime* (p. 196-218). New York, NY: Oxford University Press.

Messner, S. F., Anselin, L., Baller, R. D., Hawkins, D. F., Deane, G., & Tolnay, S. E. (1999). The Spatial Patterning of County Homicide Rates: An Application of Exploratory Spatial Data Analysis. *Journal of Quantitative Criminology*, *15*(4), 423–450. doi: 10.1023/A:1007544208712

México-Evalua. (2020). Fallas de Origen 2020. In (p. 1-93). Polanco, Ciudad de México: México Evalúa.

Paoli, L. (2014). Introduction. In L. Paoli (Ed.), *The oxford handbook of organized crime* (p. 13-31). New York, NY: Oxford University Press.

Paoli, L., & Beken, T. V. (2014). Organized Crime: A Contested Concept. In L. Paoli (Ed.), *The oxford handbook of organized crime* (p. 13-31). New York, NY: Oxford University Press.

Sampson, R. J., & Grove, W. B. (1989). Community Structure and Crime: Testing Social-Disorganization Theory. *American Journal of Sociology*, *94*(4), 774–802. doi: 10.1086/229068

Vilalta, C. (2014). How Did Things Get So Bad So Quickly? An Assessment of the Initial Conditions of the War Against Organized Crime in Mexico. *European Journal on Criminal Policy and Research*, *20*(1), 137–161. doi: 10.1007/s10610-013-9218-2

Vilalta, C., & Muggah, R. (2014). Violent disorder in Ciudad Juarez: a spatial analysis of homicide. *Trends in Organized Crime*, *17*, 161–180. doi: 10.1007/s12117-014-9213-0

Vilalta, C. J. (2010). The spatial dynamics and socioeconomic correlates of drug arrests in Mexico City. *Applied Geography*, *30*(2), 263–270. doi: 10.1016/j.apgeog.2009.06.001

Vilalta, C. J. (2012). The Spatial Dependence of Judicial Data. *Applied Spatial Analysis and Policy*, *5*(4), 273–289. doi: 10.1007/s12061-011-9070-z

Vilalta, C. J., Castillo, J. G., & Torres, J. A. (2016). Violent Crime in Latin American Cities. *Inter-American Development Bank*.

Welsh, B. C., Braga, A. A., & Bruinsma, G. J. N. (2015). Reimagining Broken Windows: From Theory to Policy. *Journal of Research in Crime and Delinquency*, *52*(4), 447–463. doi: 10.1177/0022427815581399

Wilson, J. Q., & Kelling, G. L. (1982). Broken windows: the police and neighbourhood safety. *Atlantic Monthly*, 1–8.

Appendix A

Crime Data Description

Table A.1: Crime categories.

Categories
Abortion
Against the environment
Breach of family assistance obligations
Child trafficking
Corruption of minors
Crimes committed by public servants
Damage to property
Dispossession
Domestic violence
Electoral
Equal rape
Evasion of prisoners
Extortion
Falsehood
Falsification
Feminicide
Fraud
Gender violence in all its forms other than family violence
Homicide
Incest
Injuries
Kidnapping
Narcomenudeo
Other crimes against heritage
Other crimes against society
Other crimes against the family
Other crimes of the Common Law
Other crimes that threaten life and bodily integrity
Other crimes that threaten sexual freedom and security
Other crimes that violate personal freedom
Rapture
Sexual abuse
Sexual bullying
Sexual harassment
Simple rape
Stealing
Threats
Trafficking in persons
Trespassing
Trust abuse

Appendix B

Appendix Title1

Figure B.1: Strategical crime regions

Appendix C

OLS models

Table C.1: Social covariates effect on Homicide rates

Covariates	(1) 2015	(2) 2016	(3) 2017	(4) 2018	(5) 2019	(6) 2020	(7) 2015-2020
(Intercept)	-5.69***	-8.22***	-5.23***	-1.60	-3.60**	-2.91*	-5.11***
	(1.32)	(1.32)	(1.29)	(1.40)	(1.35)	(1.31)	(0.55)
LPOBTOT	0.34***	0.44***	0.46***	0.36***	0.35***	0.41***	0.39***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)
LPopDensity	-0.17***	-0.11***	-0.09***	-0.07*	-0.04	-0.06*	-0.09***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)
SIndi	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
SLand	0.03***	0.03***	0.02***	0.03***	0.03***	0.03***	0.03***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
SFemaleAh	0.03***	0.04***	0.01	0.02**	0.02*	0.01*	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
SISTEE	-0.05**	-0.01	-0.08***	-0.05**	-0.05*	-0.02	-0.04***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
SMale	0.06*	0.10***	0.04	-0.00	0.04	0.02	0.05***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.01)
S65More	-0.06***	-0.03**	-0.05***	-0.07***	-0.06***	-0.05***	-0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
SUnemploy	-0.05	-0.08	-0.07	-0.09*	0.04	-0.00	-0.06***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)
SDivorced	0.11***	0.11***	0.18***	0.15***	0.17***	0.14***	0.15***
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.01)
SFem6to11NAS	0.73***	0.61***	0.63***	0.55**	0.44*	0.51**	0.63***
	(0.16)	(0.17)	(0.16)	(0.18)	(0.17)	(0.17)	(0.07)
YearsSchool	-0.07*	-0.14***	-0.14***	-0.16***	-0.19***	-0.17***	-0.14***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)
DummyR1	1.14***	0.83***	1.12***	0.91***	0.76***	0.60***	0.86***
	(0.14)	(0.14)	(0.14)	(0.15)	(0.15)	(0.15)	(0.06)
DummyR2	0.55***	0.43***	0.53***	0.31**	0.38**	0.18	0.40***
	(0.11)	(0.11)	(0.11)	(0.12)	(0.12)	(0.12)	(0.05)
DummyR3	0.79***	0.57***	0.80***	0.71***	0.66***	0.50***	0.67***
	(0.09)	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)	(0.04)
DummyR4	0.77***	0.44***	0.60***	0.36***	0.48***	0.18*	0.47***
	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.03)
Observations	2456	2456	2456	2456	2456	2456	14736
R^2	0.333	0.353	0.398	0.284	0.283	0.283	0.312
R^2-adjusted	0.329	0.349	0.394	0.279	0.278	0.278	0.311

*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$ (two-tailed tests).
Standard errors in parenthesis.