CENTRO DE INVESTIGACIÓN Y DOCENCIA ECONÓMICAS, A.C.

EFFECT OF CRIME ON FUTURE CRIME: EVIDENCE FROM MEXICO CITY

TESINA

QUE PARA OBTENER EL TÍTULO DE

LICENCIADO EN ECONOMÍA

PRESENTA

SERGIO YAEL CERVANTES GIL

DIRECTOR DE LA TESINA: DR. ALEXANDER DENTLER

CIUDAD DE MÉXICO 2020

***To my mother,
father, sisters,
and people
who always
supported me.***

Content

Figures and Tables

Introduction

Crime is a phenomenon that affects society on a recurring basis and through various mechanisms. Anderson (1999) estimates the direct and indirect cost of crime for the United States to exceed a trillion dollars a year. Causes of crime vary from a lack of economic opportunities to the absence of social norms and their enforcements due to cultural diversity, residential mobility, and family disruptions (Sampson and Groves 1989, among others). Besides, crime can also change the behavior of the victims, motivating urban flight due to local crime (Levitt and Cullen 1999) or the change in daily behaviors such as the choice of the public transport to use (Lefranc 2018). Research about the behavior of crime process can give insights about how criminals plan their crimes.

The research question that this paper wants to answer is: What is the effect of crime on future crime? There are three competing answers to this question. First, it could be that there is no effect from a crime in future crime. Second, it can be a self-depressing effect. This effect has several causal mechanisms such as effective policing or when the criminal is apprehended and jailed. Finally, crime could increase after a crime. Several mechanisms can sustain a self-exciting effect. For example, the social interaction between criminals can instigate other criminals due to economies of scale when scouting an area for good rewards. In the next section, I will explain in more detail other hypotheses that explain this effect.

In the Latin American region, there is little empirical research on crime compared to advanced countries (Di Tella et al. 2010). This little research is paradoxical considering that Latin America has high crime rates compared to the world average.[1] Some estimates evaluate the cost of crime in Latin America as amounted to 12.2 percent of the region's gross domestic product (Londoño and Guerrero 1999). Furthermore, Burki and Perry (1998) point out that, if Latin America had similar crime rates as the rest of the world, per capita income would be 25% higher.

I find that there is a self-depressing effect of crime on future crime for business theft and residential burglaries. This is: previous crimes decrease the probability of crimes in the

[1] According to the survey from 2018 carried out by Latinobarómetro, 36% of people are all-time afraid of being victims of a crime

immediate future. According to the estimates, there is around a 1% to 5% decrease in the probability of a future business theft after another business theft. For residential burglaries, this decrease is around 5% to 7 %. Noteworthy that this effect does not imply that crime will be steadily decreasing but this is only the behavior of short windows of time and distance. To study this effect, I used the Regression Discontinuity Design (RDD) framework along with time and location data to give information about the behavior of crime. This effect is studied at the point of time and space when the treatment crime occurs. The extrapolation from these results can be difficult to support for other cities due to idiosyncratic characteristics of crime. Then, even though this effect could apply only to Mexico City, the understanding of the behavior of criminals as the one found here could motivate further analysis in economics, sociology, and public policy.

This research contributes to the existing literature on crime dynamics. The results differ from other researches as in Mohloer and Short (2011), Johnson (2008), and Short et al. (2008) due to the short windows of time and distance that I use. They are interested in the long-term dynamics. The difference with this work is that I try to understand the immediate effect of a crime on the occurrence of another crime. This work is presented as follow; literature review provides information about how economist have studied crime, the patters of crime, and other research that uses the same approach to study crime. Then, the section 3 describes the data and how crime is distributed in time and space. The estimation section provides information about why I use this methodology and how I interpret the data. In the results, I show the estimations for the discontinuity. Later on, I present the robustness of this work by presenting a placebo test and after then, the main conclusion is presented.

Literature Review

The seminal paper on crime is Gary Becker's (1968) publication *Crime and Punishment: An Economic Approach*[2], where he states that crime is the result of individuals making rational decisions, weighing consequences, and maximizing profits. Since then, the crime literature has tried to understand the effect of crime on variables such as the real estate market, welfare, urban planning, and so on; or the different characteristics that trigger the behavior of crime. Di Tella and Schargrodsky (2004), Levitt (2004), and Galiani et al (2010) provide evidence of key determinants to understand the response of crime to different variables to choose which of them are important to reduce it. Among those determinants are better policing, birth control, public spending, and improving the socio-economical context of criminals. These empirical studies find causal effects and give public policy recommendations based on their results. This work focuses on understanding the effect of previous crimes on future crime, using time and spatial variables.

Concerning the geographical analysis of crime, Galiani et al. (2018) developed a theoretical model of general equilibrium model and presented an empirical validation to understand the urban configuration[3] from the endogenous decision of becoming criminal or not. One implication of this work is that crime can be concentrated in "hot-spots", creating a segregated configuration of the city and lowering social welfare. Johnson (2008) shows that crime can spread locally through a contagion-like process. Mohler, et al (2011) use a space-time clustering model from seismology and show that those methods fit well for criminological applications. They conclude that crime follows a pattern of self-exciting point process similar to earthquakes, given that there is an increase in the risk of subsequent crime near the location of an initial event. Also, Reinhart and Greenhouse (2017) demonstrate through inference methods and diagnostic tools that time and spatial variables are of importance when analyzing a self-exciting point process and subsequently, the patterns of crime.

Concerning how crime is spread, Sah (1991) argues that individuals that live in areas with high crime rates can perceive a lower probability of apprehension than those living in areas

[2]Becker G.S. (1968) Crime and Punishment: An Economic Approach. In: Fielding N.G., Clarke A., Witt R. (eds) The Economic Dimensions of Crime. Palgrave Macmillan, London.

[3] Defined by the places where criminals and not criminals decide to live, either in a cheap, low protected area or an expensive and more secure area of town

with low crime rates. This is because of the resources spent in apprehending each criminal is more—which can be interpreted as a major cost for apprehension. This analysis implies that "The past crime breeds future crime". Glaeser et al (1996) point out that systematic social interactions cause criminal inertia over time, which provides evidence in favor of the existence of crime waves. One way this works is through the decreasing cost of committing a crime since criminals, like in other jobs, learn by doing and once they commit a crime, is hard to turn back. Another way to analyze criminal inertia is when the police fail to jail criminals. With this, probabilities of apprehension lower and criminals update their beliefs about being apprehended, which encourages further crime (see Posada 1994). For Latin America, Fajnzyber, Lenderman, and Loyoza (2000) provide evidence of this inertial effect in an economic model of criminal behavior. They tested for inertia by including the lagged crime rate as an explanatory variable.

For the econometric strategy, I used a similar approach to Dentler and Rossi (2020). They used geo-referencing data to match crimes and house sales to measure the effect of crimes on the real estate market. With the Regression Discontinuity Design framework, they give empirical evidence that crime affects the liquidity of the real estate market rather than price estimates. I use the approach they use by estimating the discontinuity in the density of crimes around a treated crime. Also, my results are consistent with the self-depressing finding that they found for crimes.

Crime literature on the causes of crime shows two different aspects of emphasis. On one hand, literature is focused on economic conditions that trigger crime. On the other, recent considerations of social factors help explaining how crime is propagated over time and location. There is a desirable interaction between theoretical and empirical studies, and it relies on the availability of methodology and data. Appropriate data sources are needed to analyse the propagation of crime, to which now I turn.

Data

The information on crimes for this work is taken from the Attorney General of Mexico City (*Procuraduría General de Justicia de la Ciudad de México*, PGJCDMX in Spanish) which is the institution where authorities register crimes. Since 2018, the government of Mexico City has adopted a policy on transparency thanks to a reform in the Law of Transparency.[4] This initiative of open data provides the geo-referencing information on crime and the date when the crime was committed and registered to authorities. The data is available from 2016 and until June 2019. In total, the database has more than 808'000 observations.

These collected data are reported by the victims to the local authorities. There are a wide variety of crimes. To list some: robberies, shoplifting, assault, burglary, domestic violence, fraud, homicide, among others. In the studies where the objective is to find the causal effect from any variable (police, public lightning, economic conditions, and so on) on crime, endogeneity is the main concern. That is because of the simultaneous determination of crime and the other variable. For example, policymakers may respond to an increase in crimes in a certain location by augmenting the spending in that area. Then, areas with more crime will have higher spending in comparison with areas with a lower crime rate, which introduces a positive bias in the regression to identify the causal effect of government spending on crime.

In related literature, the observations that are studied are property crimes, which can be defined as residential burglaries, business burglaries, residential thefts, business thefts, thefts from vehicles, and vehicle thefts.[5] I will focus on residential crimes and business crimes,[6]. The reasons for choosing those crimes are that there are incentives for victims to report the crime, which helps to avoid the tendency of underreporting the crimes (measurement error bias). There are two main incentives to do that: 1) victims can be scared that the criminals may be back to commit a crime and ask authorities for better surveillance, and 2) to claim insurance, the report must be registered by authorities. Furthermore, both crimes have the characteristic that the

[4] Transparency Law and Access to Public Information and Accountability of Mexico City, 2017, Mexico City Government, Mexico, 127 pp

[5] Burglary means entering a structure without consent but not necessarily depriving an owner of their property, which define a theft.

[6] I cannot use car thefts because, in the data set, I have no information about the characteristics of the theft. That means that I do not kwon if the theft were when the car was stopped or when it was on the road.

location is important from the criminal perspective. Surveillance in the area, streetlights, average income, and other characteristics, may incite the crime by lowering the cost (measured as the probability of getting arrested) and increasing the economic benefits (measured as the value of the stolen goods). These characteristics make it possible to link the crime to an area.

Mexico City is divided into 16 municipalities. Table 1 shows the total crimes, number of business crimes, and the number of residential crimes according to each by a municipality. It is well documented that crime is spatially and time concentrated. Then is possible to observe that crime is concentrated in a few municipalities. The difference between municipalities' crime can come from the operation of policing, geographic characteristics, socioeconomic characteristics, corruption, etc. It highlights that crime varies across geographic units. For all crime observations, most of the crime is in Cuahutémoc and for the crimes relevant to this work, most of the crimes are in Iztapalapa. The municipality with the lowest number of crimes is Milpa Alta. When I consider the population of each municipality[7], for residential crime, Benito Juarez

Table 1 Crimes per municipality

	Total				Business crime			Residential Crime		
Municipality	Count	Percent	Crime/Pp	Population	Count	Percent	Crime/Pp	Count	Percent	Crime/Pp
ALVARO OBREGON	40805	6.53	0.054	749982	3198	6.44	0.004	1101	5.63	0.001
AZCAPOTZALCO	32028	5.13	0.080	400161	3119	6.28	0.008	990	5.07	0.002
BENITO JUAREZ	55794	8.94	0.134	417416	5542	11.15	0.013	2126	10.88	0.005
COYOACAN	43139	6.91	0.071	608479	3726	7.50	0.006	1583	8.10	0.003
CUAJIMALPA DE MORELOS	8867	1.42	0.045	199224	758	1.53	0.004	322	1.65	0.002
CUAUHTEMOC	100386	16.08	0.188	532553	7358	14.81	0.014	1841	9.42	0.003
GUSTAVO A MADERO	61938	9.92	0.053	1164477	5501	11.07	0.005	2045	10.46	0.002
IZTACALCO	27314	4.37	0.070	390348	1257	2.53	0.003	867	4.44	0.002
IZTAPALAPA	92827	14.87	0.051	1827868	8015	16.13	0.004	2596	13.28	0.001
LA MAGDALENA CONTRERAS	9399	1.51	0.039	243886	329	0.66	0.001	554	2.83	0.002
MIGUEL HIDALGO	42324	6.78	0.116	364439	3636	7.32	0.010	1080	5.53	0.003
MILPA ALTA	4040	0.65	0.029	137927	116	0.23	0.001	224	1.15	0.002
TLAHUAC	13936	2.23	0.039	361593	720	1.45	0.002	546	2.79	0.002
TLALPAN	36293	5.81	0.054	677104	2708	5.45	0.004	1607	8.22	0.002
VENUSTIANO CARRANZA	35827	5.74	0.084	427263	2651	5.33	0.006	1032	5.28	0.002
XOCHIMILCO	19521	3.13	0.047	415933	1061	2.14	0.003	1030	5.27	0.002

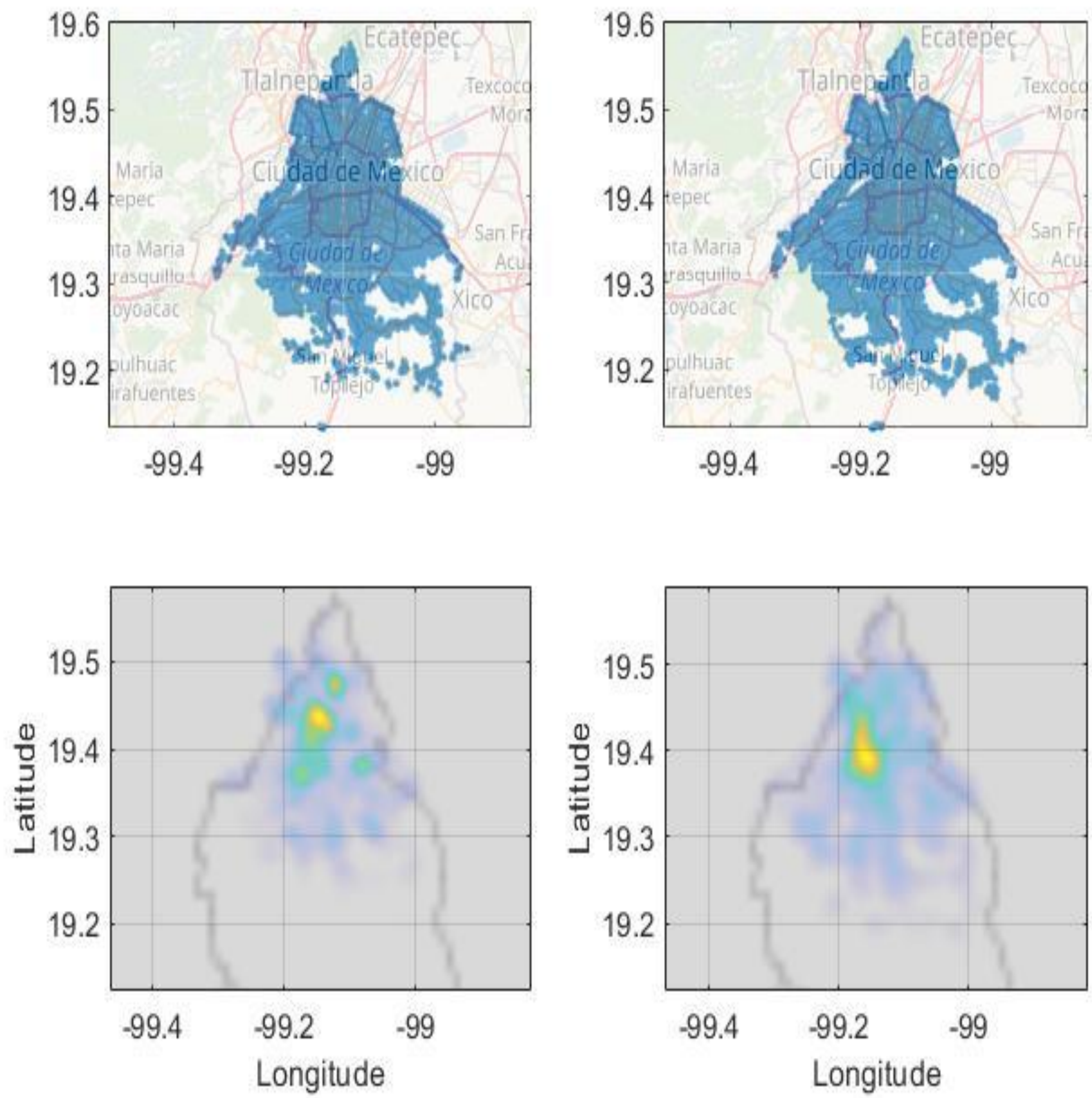
Own elaboration with data from the intercensal survey INEGI 2015

is the one with the highest number of crimes given its population. On the contrary, once I have

[7] Intercensal survey INEGI 2015

taken the population into account, Iztapalapa is the municipality with the lowest number of residential crimes in Mexico City. Figure 1 (upper) depict crimes by point where the x axis represent the shows the location and latitude and y-axis represent longitude, (down) shows the density of crimes within Mexico City, where it is possible to see that the crimes are distributed in the northern part of the city.

Figure 1 Distribution and density of crimes

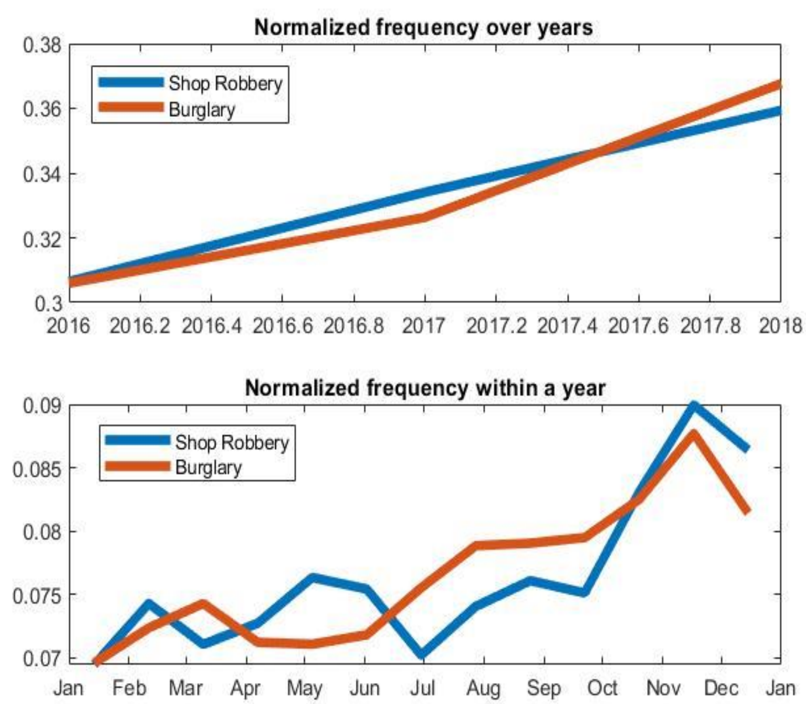


Own elaboration with data from PGJCDMX

Regarding the clustering of time, Figure 2 shows a more detailed trend across years and within years. There is a remarkable uprising trend in the frequency of both crimes over the years. Also, within years, crime increases in the last trimester of every year. Given that this work studies the impact of the crime in the immediate probability of other crime, I use short windows of time. Then, the trend of crime within the week could generate a bias even with a short bandwidth of the time.

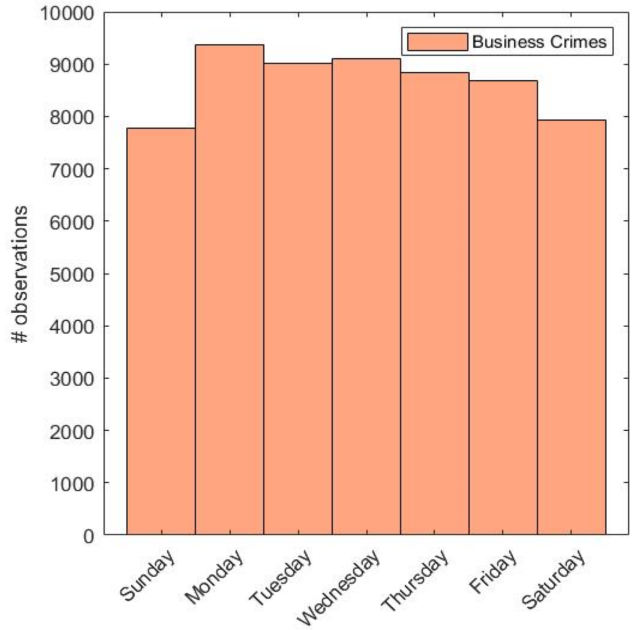
Figure 3 shows the histogram of crimes within the week. Business crimes are concentrated on labor days and residential crimes on weekends. Furthermore, in Figure 4, I plotted a histogram of the times between nearby crime separated by 200 meters or less, for all recorded residential and business crimes in Mexico City. Is possible to observe a spike almost every three days —this pattern is evident in Figure 3 for residential crimes. This may be suggesting that criminals use windows of three days or one week away to commit another crime. This graph can also suggest a self-depressing effect that comes from the decreasing trend after the spike at short times, this suggests that the likelihood of victimization within 200 meters of each crime decreases when more days are added.

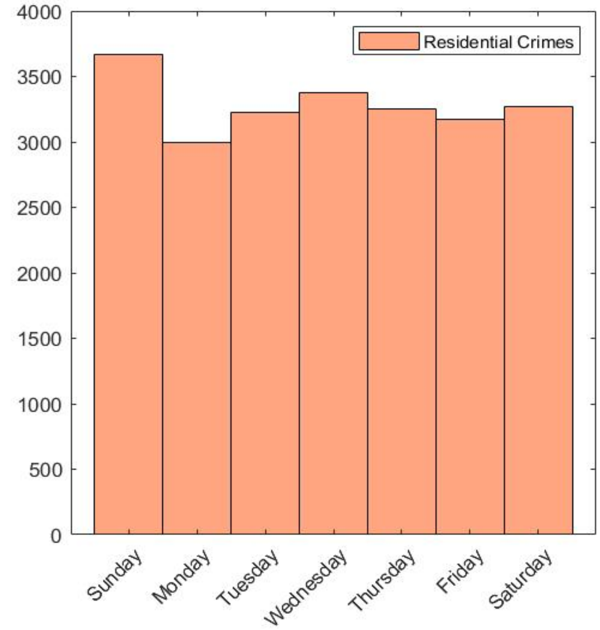
Figure 2 Frequency over and within years.



Own elaboration with data from PGJCDMX

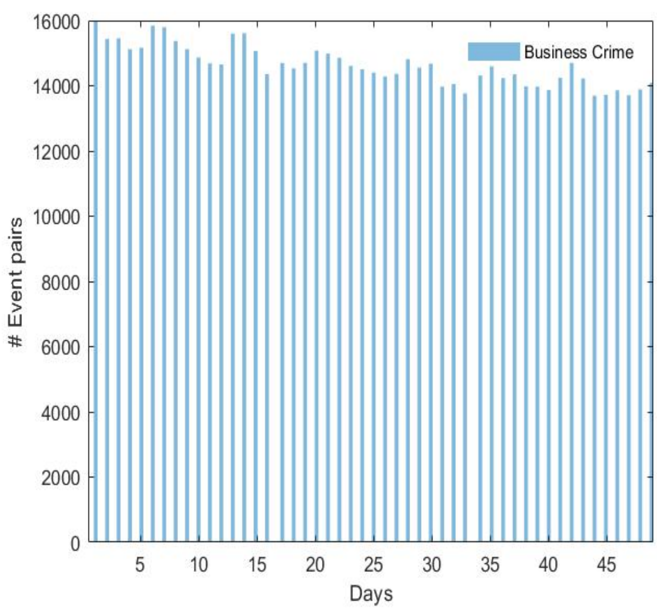
Figure 3 Histogram within weeks

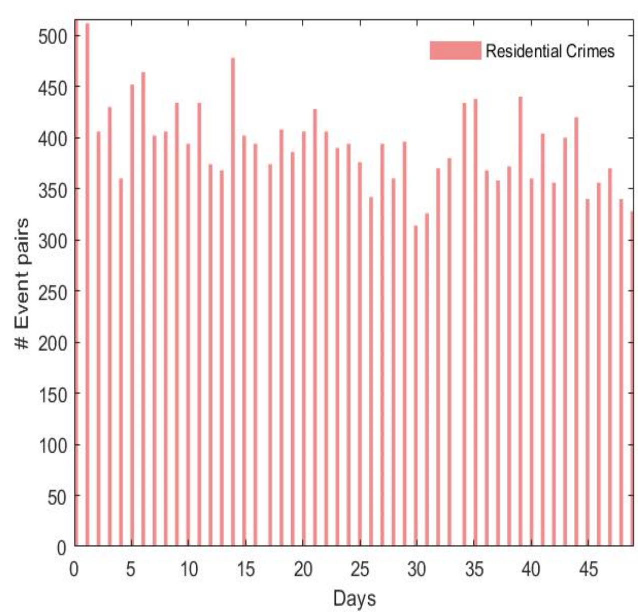




Own elaboration with data from PGJCDMX

Figure 4 Histogram of time between crimes





Own elaboration with data from PGJCDMX

The data taken from PGJCDMX has the time and the coordinates when the crime was committed (according to the victims). The representation of the latitude and longitude is given by the vector $L = [L_1\ L_2]$ where L_1 (L_2) represent latitude (longitude). I use the georeferencing data to transform this information into distance information. I linked the vector of crimes C to get the information of the distance between all of them. The number of the pairwise link between crimes has size $C \times C = I$ indexed by i to refer to a specific pair of crimes. Similarly, I will link the time distance between crimes according to $T_i = T_c - T_{-c}$. This means that the pair of a crime linked with itself will lead to a distance of time and location of zero. In this way, T_i is positive when an observation is compared with a crime that occurred after the particular observation.

Then, I use the data to get the Euclidian distance. This is

$$D_i = \sqrt{(L_{c1} - L_{-c1})^2 + (L_{c2} - L_{-c2})^2}$$

Where $-c$ is a crime observation different from c. I take logarithm to get $d_i = \log(D_i + \epsilon)$, where ϵ is a positive value that allows me to take the logarithm. The realization $X_i = T_i, D_i$ have a distribution $F(X)$ and the probability density function $f(X)$.

Estimation

The main objective of this work is to understand which are the effects of crime on future crime. Then, I want to compare a location (over time) that has recently experienced a crime event with a location that has not. The first is the treatment observation and the other is the control. The framework of the RDD allows me to estimate the average treatment effect which compares the mean outcome of the treated group with the mean outcome of the control group. This methodology contrasts the mean outcome of the treatment group with the mean outcome of the control group. I use index T_i to assign the link of crimes i according to the next rule: treatment group if T_i at or above the threshold τ_0 and control group if is below the threshold τ_0. Then, τ_0 will force the observations to be either control or treatment group and is predetermined and public knowledge. Identification of the average treatment effect comes from the comparison of the mean from marginally below and above τ_0. Therefore, the comparison through extrapolation is what gives RDD the possibility of inference. The identification requires that the conditional outcomes are continuous around the threshold so that extrapolation is valid. Suppose that I let $f_i(T_i)$ denote the density function around the running time variable, then I could test for a jump in the density by calculating

$$\theta_i = \lim_{T\uparrow 0} \log\big(f_i(T)\big) - \lim_{T\downarrow 0} \log\big(f_i(T)\big)$$

And test if $\theta_i = 0$. The problem with this assumption is that I can only see one realization per linked crimes. Then, I pool observations around crimes to generate the distribution function F with the density function f. The method I use to generate this distribution function is similar to McCrary (2008) with the only difference that I made it with two variables: time and distance. However, the "cookbook" receipt provided by him is used: construct a histogram and make it softer by using linear regression. Later on, I explain this in more detail. When the distribution function is made, I follow the arguments from Dentler and Rossi (2020) where I can test for a jump in the density around the threshold with

$$\theta = \lim_{T\uparrow 0} \log\big(f(T)\big) - \lim_{T\downarrow 0} \log\big(f(T)\big)$$

where I want to test $\theta = 0$. Here, a rejection of the hypothesis suggests that crime affects the likelihood of another crime. I pooled observations around crimes to generate the distribution F

with density f. I used observations that occurred after the treatment crime and are within the bandwidth of time and location.

There can be a possible impact of time on the outcome. Serial dependence can create a bias in the estimator due to the general trend of crime if the window of the running variable is too large. I sort those issues by using the triangular kernel function to downweigh observations that are far from the threshold and use a reasonably small varying bandwidth. I set up windows bandwidth for time, denoted as v_T; and distance denoted as v_D. I use data that is not more than 400 meters away and 75 days from the crime to avoid this bias. Besides, I include an extra time variable that will reduce the bias that comes from the trend of crime. This is, I added variables that control for the day of the week given the presence of days with more observations.

Down-weighting observation far from the threshold ensures that the estimation is effectively local. This reduces irrelevant influences on the point of interest. I use triangular kernel so it integrates into one. Particularly, I use $K_T(u) = \max(0,1-|u|)$ and $K_D(u) = 2\max(0,1-|u|)$. Then, the matrix of weights is defined by

$$w(X) = \frac{1}{v_T v_D} K_T\left(\frac{T}{v_T}\right) K_D\left(\frac{D}{v_D}\right)$$

To estimate the discontinuity in the density of crimes RDD is similar to a non-equivalent group design, where the major issue is model misspecification when the regression equation does not reflect data distribution. Then, a local polynomial approximation tends to accommodate the jumps around the threshold to avoid misspecification. The model I use is

$$m^T(X,\delta) = \delta + \delta_T T + (\Delta + \Delta_T T)\mathbf{1}(T > 0)$$

The first part of the estimation captures the level (δ) and trend ($\delta_T T$) of crimes that are below the treatment ($T < 0$). Then, the variation that I am interested in is the change in the level after the treatment of crime ($T > 0$) which is explicitly coefficient Δ. This change in the level around the treatment is θ. The trend after the treatment is captured by $\Delta_T T$. Also, we have seen that criminals may prefer some days to commit a crime. Let say, that in the data, a lot of crime happens on Monday; then is probable that the recurrently decrease from Monday to Tuesday, may bias the estimator. To solve, I ran separate regressions for each day to check if there was a significant effect that may bias the estimator.

Even when is possible to incorporate the distance between crimes into the model, the interpretation of that relationship does not come straightforward. That is because the relation of the pair of crimes independent of when occurred the crime can suggest if is there is a lower density, that criminals anticipated that crime and avoided the area. Furthermore, the selection of crimes could be bias by the confounding factors that come from the spaces where there are not many people, say mountains, dams, etc., in the south of the city or any other place.

As I cannot observe the density of crimes, it is not possible to run a usual regression. I bin observations with the methodology proposed by McCrary (2018). This means that I count the frequency of observations in each bin to create regressands. I use a linear model with a break in coefficients and use the regressands from the binning technique. This test estimates the discontinuity at the cutoff in the density function of the running variable. Wald test support the null hypothesis that discontinuity is zero. This estimator is constructed in two steps. First, has to obtain a finely gridded histogram (through a grid previously defined). And second: smooth that histogram using local linear regression. This technique allows smoothing the density separately on either side of the cutoff. This methodology is different from a simple histogram or kernel density estimator because it allows for point estimation. (i.e. a kernel estimation from point to the left and the right of discontinuity will make that, at boundaries, the estimator is biased as stated by Marron and Ruppert (1994). In the appendix, I show the construction of the binning estimation.

Results

I take out incomplete data and restrict the period of analysis (from January 2016 to December 2018). This restriction is because of the change in the administration of Mexico City on the 1st of December of 2018, which can change some of the criteria when the crimes are registered. I analyzed the data by using equi-spaced windows of time and location. The shorter window is 10 days (after the crime) and the maximum is 30 days, with the intermediate windows of 15, 20, and 25 days. These variations on the days allows to show the tradeoff between efficiency vs bias due to the long-term effect because wider bandwidth have more observations to estimate the discontinuity effect. With the shorter windows, I try to reduce the variance of the short-run estimate by accepting a potential bias due to long-run effects. Similar for distance, the shorter window is 200 meters (around the crime, which means a diameter of 400 meters) and 450 meters away, with intervals of 50 meters. The window of 200 meters is comparable with two street blocs. The windows chosen here are similar to bins used by Mohler et al. (2011) to model the self-excitement point process.

Table 2 presents the summary statistics from both crimes. Column 1 shows the summary statistics for all crimes taking account of all the data. The mean distance supports the use of a window of 200 meters. Columns 2 and 3 show the means of the larger bandwidth of 30 days before and after the crime event. The same analysis is for columns 5 and 6, but only with the 10 days away from each crime. The decrease in the mean may suggest that there is a self-depressing effect when we make the windows adequately short. Nevertheless, is still not possible to be sure of any effect, the use of the triangular kernel to downweigh the observations far from the treatment event, and the use of more windows will explain better the jumps in the density estimation.

Now, I discuss the results from the linear model presented in the previous section. Tables 3 and 4 show the results from observations for business residential crimes, respectively. The information shown here only corresponds to the parameter of interest θ, since the other coefficients are nuisance parameters. This parameter shows the discontinuity of the density estimation when the observation of that particular number of days and meters are taken into account for the estimation. The days where the pair of crimes happened are taken into account in the regression but, given that there is not a clear pattern of coefficients statistically significant, they are not shown.

Table 2. Summary Statistics

Business Crime					
	All	**Before**	**After**	**10 Days Before**	**10 Days After**
Mean of distance	0.2284	0.2324	0.2334	0.2306	0.2306
Std of distance	0.1535	0.1515	0.1516	0.1524	0.1525
Mean of observations	0.9752	0.4844	0.4908	0.1576	0.1582
Std of observations	0.1554	0.4998	0.4999	0.3644	0.3650
Residential Crime					
	All	**Before**	**After**	**10 Days Before**	**10 Days After**
Mean of distance	0.2862	0.3215	0.3215	0.3210	0.3208
Std of distance	0.1550	0.1252	0.1252	0.1262	0.1262
Mean of observations	0.8840	0.4360	0.4479	0.1436	0.1447
Std of observations	0.3203	0.4959	0.4973	0.3507	0.3518

Own elaboration with data from PGJCDMX

The coefficients for all the windows are positive and statistically significant. It is in a range of $\theta \in (-0.0028, -0.055)$ for business crimes.[8] This means that there is a reduction between 1% and 5% of the probability of another business crime after the first crime. When we look to the coefficients of the same distance, is possible to observe that the larger the bandwidth of days is, the lower the probability of a jump in the density function. This demonstrates the trade-off between efficiency and accuracy have given that, when more pair of observations is added to estimate the density jump, the estimator may be biased by the long-term effects. When more observations are added due to a wider bandwidth of distance, there is not a clear trend in the probability of a discontinuity. The only observation here is that the probability of a new crime is lower.

The results for residential crimes are shown in Table 4. The parameter of interest for some certain windows of analysis is $\theta \in (-0.0771, -0.0418)$ which is equivalent to a decrease

[8] As the coefficient is log-transformed, can be interpreted with the transformation $(e^{\theta} - 1) \times 100\%$

of 4% to 7% in the probability of a new crime. For this particular crime, there is research that states that burglars will repeatedly attack clusters of nearby targets because they are aware of the local vulnerabilities (Besnarsco and Nieuwbeerta 2005), This behavior of crime leads to the formation of crime clusters in space and time. Nevertheless, in the resulting estimates of Table 4, the trend suggests that there will be less crime right after another crime. This contradicts the findings by Mohler et al. (2011), Farrell and Pease (2001), and Johnson et al (2007) where "evidence indicates that an elevated risk exists for both a house that has been recently burgled and its neighboring houses". The reasons that can explain this difference in the results are mainly that they are looking for repeated victimization, which must include a large bandwidth of time, while I try to find the immediate impact of crime on crime using the smallest window of time possible.

Table 3					
Discontinuity for business crimes					
Days	**10**	**15**	**20**	**25**	**30**
Distance	**0.2**	**0.2**	**0.2**	**0.2**	**0.2**
Pair of crimes	**94914**	**110267**	**126522**	**142992**	**161181**
θ	-0.0550** (0.0222)	-0.0109 (0.0086)	-0.0227* (0.0132)	-0.0138** (0.0069)	-0.0047 (0.0068)
R2	0.9992	0.9999	0.9996	0.9999	0.9999
Distance	**0.25**	**0.25**	**0.25**	**0.25**	**0.25**
Pair of crimes	**189535**	**220101**	**252182**	**285195**	**321622**
θ	-0.0370* (0.0207)	-0.0108 (0.0076)	-0.0079 (0.0115)	-0.0085 (0.0062)	-0.0028 (0.0063)
R2	0.9993	0.9999	0.9997	0.9999	0.9999
Distance	**0.3**	**0.3**	**0.3**	**0.3**	**0.3**
Pair of crimes	**282674**	**328191**	**376161**	**425097**	**479092**
θ	-0.0427** (0.0168)	-0.0097 (0.0066)	-0.0149 (0.0100)	-0.0097* (0.0053)	-0.0046 (0.0049)
R2	0.9994	0.9999	0.9997	0.9999	0.9999
Distance	**0.35**	**0.35**	**0.35**	**0.35**	**0.35**
Pair of crimes	**373407**	**434205**	**498211**	**563091**	**634921**
θ	-0.0345** (0.0155)	-0.0078 (0.0059)	-0.0074 (0.0088)	-0.0070 (0.0050)	-0.0048 (0.0047)
R2	0.9995	0.9999	0.9998	0.9999	0.9999
Distance	**0.4**	**0.4**	**0.4**	**0.4**	**0.4**
Pair of crimes	**464541**	**540541**	**620386**	**701473**	**790863**
θ	-0.0387*** (0.0137)	-0.0100* (0.0052)	-0.0104 (0.0076)	-0.0080** (0.0040)	-0.0024 (0.0040)
R2	0.9879979	0.98	0.98	0.98	0.96
Distance	**0.45**	**0.45**	**0.45**	**0.45**	**0.45**
Pair of crimes	**554454**	**645193**	**741250**	**838404**	**945223**
θ	-0.0430*** (0.0126)	-0.0106** (0.0045)	-0.0127* (0.0068)	-0.0097*** (0.0031)	-0.0048 (0.0031)
R2	0.9997	0.9999	0.9998	0.9999	0.9999

Standard errors in brackets. The pair of crime are the number of linked crimes within each window of time and distance.
*, ** and *** represent 10%, 5% and 1% significance.
Own elaboration with data from PGJCDMX

Table 4

Discontinuity for residential crimes

Days	**10**	**15**	**20**	**25**	**30**
Distance	**0.2**	**0.2**	**0.2**	**0.2**	**0.2**
I	**22483**	**24147**	**25742**	**27399**	**28886**
θ	0.0218 (0.0531)	-0.0699 (0.0453)	-0.0110 (0.0328)	-0.0332 (0.0339)	-0.0284 (0.0307)
R2	0.9974	0.9980	0.9978	0.9981	0.9980
Distance	**0.25**	**0.25**	**0.25**	**0.25**	**0.25**
I	**24300**	**26694**	**29028**	**31466**	**33754**
θ	0.0106 (0.0443)	-0.0743* (0.0385)	-0.0190 (0.0292)	-0.0327 (0.0293)	-0.0274 (0.0269)
R2	0.9969	0.9977	0.9974	0.9977	0.9974
Distance	**0.3**	**0.3**	**0.3**	**0.3**	**0.3**
I	**26317**	**29581**	**32803**	**36104**	**39240**
θ	0.0182 (0.0412)	-0.0679* (0.0348)	-0.0097 (0.0264)	-0.0245 (0.0262)	-0.0222 (0.0243)
R2	0.9964	0.9976	0.9972	0.9975	0.9969
Distance	**0.35**	**0.35**	**0.35**	**0.35**	**0.35**
I	**28482**	**32770**	**37050**	**41382**	**45578**
θ	-0.0006 (0.0343)	-0.0689** (0.0330)	-0.0179 (0.0252)	-0.0309 (0.0250)	-0.0352 (0.0228)
R2	0.9957	0.9971	0.9964	0.9970	0.9965
Distance	**0.4**	**0.4**	**0.4**	**0.4**	**0.4**
I	**31057**	**36529**	**42046**	**47558**	**52982**
θ	-0.0110 (0.0332)	-0.0771*** (0.0298)	-0.0263 (0.0225)	-0.0378 (0.0238)	-0.0449** (0.0223)
R2	0.99492229	0.9967	0.9956	0.9964	0.9957
Distance	**0.45**	**0.45**	**0.45**	**0.45**	**0.45**
I	**33979**	**40802**	**47674**	**54530**	**61337**
θ	-0.0045 (0.0322)	-0.0595** (0.0291)	-0.0227 (0.0232)	-0.0301 (0.0228)	-0.0418** (0.0205)
R2	0.9936	0.9952	0.9942	0.9947	0.9938

Standard errors in brackets. The pair of crime are the number of linked crimes within each window of time and distance.
*, ** and *** represent 10%, 5% and 1% significance.
Own elaboration with data from PGJCDMX

Discussion

I showed that there is a self-depressing effect in and business crimes as defined before. The methodology used here is based on linking the pair of crimes that satisfies the conditions of the bandwidths according to the days and distance between them. Then, a grid is used to accommodate those observations and a histogram is done based on the number of observations in each bin. The normalized and log-transformed observations falling into the bins are treated as an outcome variable and the midpoints of the histogram bins are treated as a regressor. To accommodate the discontinuity, local linear smoothing is conducted for the bins to the right and left of the “zero-day” condition, which is the comparison of the crimes with the crimes that happened the same day.

The difference with other methodologies like the auto-regressive process is that it only uses the time dimension and then, they are concentrated in a specific location (Cesario et al. 2016). To explain the dynamics of crime along the two dimensions, Mohler et al. (2011) uses a self-exciting point process from seismology and find evidence that crime can be modeled under this approach. They used the whole data of burglary crime in L.A. for two years. The density estimated by Monte Carlo simulations for burglaries shows a pattern where there is a spike around 1-2 days which they explain as “the presence of crime sprees, where most likely the same burglar visited several neighboring houses within a time span of a few minutes to several days”. In line with my results, after 7-10 days, the elevated risk of a new crime drops to an intermediate level and stays relatively flat.

In this work, to capture the immediate effect of the crimes, I use smaller bandwidth of time and space along with the triangular kernel to downweigh the observations far from the treatment crime. This difference, in addition to the evidence provided by Mohler et al (2011) of a decrease in the probability of new crime for 7-10 days, gives more support to the findings of a self-depressing effect found here. I addition and to compare the results, I estimated the discontinuity for a window of 5 days and 200 meters for business crime and found a self-excitement effect. Nevertheless, for burglaries, the coefficients remained insignificant. This can be explained for the controls of the days of the weeks, which also as explained by Mohler et al. “the routine of the burglar and/or the victim is such that a particular day of the week is a preferable time to commit the burglary”.

Another problem arises with the registration of the crime. When a burglary happens, victims usually do not know the exact time when the crime was committed. Usually, the registration of the "moment of crime" is the moment when the victims realized of the burglary (Bernasco and Nieuwbeerta 2005). The way to account for this issue is to kick out the observations that occur on the same day. This is: the crime will not be counted into the bins if the difference of days between them is zero.

To provide credence to the causal inference that the effect of crimes on future crimes are captured, I ran a placebo-styled test. Instead of using crime events as treatment, I used an "event" that should not incur any response. Then I used the same locations from the crime events but 365 days before. For residential crime, if I allow for 10% false positives, indicates that it is a good placebo, and this suggests that the variations in my results ought to be the crimes. If I follow this rule for the placebo test, for business crimes, does not meet this condition. This can be explained by the process of bootstrapping, where it is not resampling the treatment crimes but only the observations around it.

In this research, I did not take into account other variables that could be explaining the increase or decrease in the probability of a future crime. The trends in crime rates can be improved when additional variables are added. For example, the residential units, proximity to the city center, ethnic heterogeneity and real estate value can adjust the probability of a future crime. Along with the quantitative analysis, qualitative analysis must be carried out to validate the findings. Sociological research has been studying criminals to understand the dynamics of peers and delinquent behavior (Sah 1991; Matsueda 1998) and this research can give insights about which variables are important to increase/reduce the probability of a criminal committing another crime.

Conclusions

What is the effect of crime on future crime? I this work, I used the testing for density discontinuities used in the RDD framework to show that there is a self-depressing effect for residential and business crimes. The extrapolation of these results may be plausible for cities that are similar to Mexico City. The population of a city, crime rates, and law enforcement are some variables that differentiate the cities where it can be extrapolated. The policy implemented to stop crime varies from town to town. From policing, neighborhood vigilance, public lighting, etc. could be the responses aimed to reduce crime. Then, the effect of those responses can be more (less) effective depending on things such as culture, the mix of those policies, the law system, etc. Then if a city shares some of these characteristics, is possible to say that criminals behave with the dynamics shown in this work.

Even when the results in this work accommodates better in the literature of a self-depressing effect of crime on future crime, the result is not an absolute one. As I have shown in the previous section, there is evidence in related literature that I could verify that the dynamic of the crime depends on the windows and variables that are of interest. In this work, I wanted to show which is the immediate effect of the crime in the dynamics of itself. To avoid the bias generated by the long-term trends, I used short bandwidths of time and distance. Even with that, when increasing the windows of time and distance, the bias could be in both directions.

Then, with the lower probability in the risk of a future crime, many explanations can accommodate this effect. A self-depressing effect suggests that criminals are having a worse time immediately after they commit a crime. This could be because of an effective (and quickly) response of police or any other kind of punishment. Also, it can be explained by any form of planning in crimes, where criminals choose the place to commit a crime in a certain kind of intervals in which case, crime could be having a self-depressing effect immediately after the crime but the contrary in some distance of time and space away.

The study of these “cold-spots” could be also objective for further research to understand how a criminal who had already stopped committing crimes, get back to the same pattern as before. This kind of information could be more feasible to find the patterns that will eventually help in the forecasting of crime. Also, further research should compare the different methodology that tries to give insights about the dynamics of crime to say how accurate are when predicting where crimes will occur in the future.

One possible implication from the results given in this work is that the policing in Mexico City and other variables —which can include neighborhood vigilance, apprehension of criminals, and so on— are indeed reducing crime, even when it is only in the short run. A better understanding of these dynamics could help translate this reduction in the likelihood of a future crime into longer-term effects. This desirable goal of predicting the crimes instead of just reacting to them and one thing is for sure, a better understanding of crime will eventually be translated into a reduction of crime and, consequently, higher welfare in society.

References

Anderson, D. A. (1999). The aggregate burden of crime. The Journal of Law and Economics, 42(2), 611-642.

Bernasco, W., & Nieuwbeerta, P. (2005). How do residential burglars select target areas? A new approach to the analysis of criminal location choice. British Journal of Criminology, 45(3), 296-315.

Braga, A., Papachristos, A., & Hureau, D. (2012). Hot spots policing effects on crime. Campbell Systematic Reviews, 8(1), 1-96.

Burki, S. J., & Perry, G. E. (1998). Beyond the Washington consensus: institutions matter. The World Bank.

Cesario, E., Catlett, C., & Talia, D. (2016, August). Forecasting crimes using autoregressive models. In 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech) (pp. 795-802). IEEE.

Cracolici, M. F., & Uberti, T. E. (2009). Geographical distribution of crime in Italian provinces: a spatial econometric analysis. Jahrbuch für Regionalwissenschaft, 29(1), 1-28.

Cullen, J. B., & Levitt, S. D. (1999). Crime, urban flight, and the consequences for cities.

Dentler, Alexander, and Enzo Rossi. (2020) Shooting up liquidity: The effect of crime on the Real Estate Market. Working Paper. CIDE-Swiss National Bank.

Di Tella, R., & Schargrodsky, E. (2004). Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. American Economic Review, 94(1), 115-133.

Di Tella, R., & Schargrodsky, E. (Eds.). (2010). The economics of crime: lessons for and from Latin America. University of Chicago Press.

Dills, A. K., Miron, J. A., & Summers, G. (2008). What do economists know about crime? (No. w13759). National Bureau of Economic Research.

Galiani, S., Cruz, I. L., & Torrens, G. (2018). Stirring up a hornets' nest: Geographic distribution of crime. Journal of Economic Behavior & Organization, 152, 17-35.

Glaeser, E. L., Sacerdote, B., & Scheinkman, J. A. (1996). Crime and social interactions. The Quarterly journal of economics, 111(2), 507-548.

Hausman, C., & Rapson, D. S. (2018). Regression discontinuity in time: Considerations for empirical applications. Annual Review of Resource Economics, 10, 533-552.

Levitt, S. D. (2004). Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not. Journal of Economic perspectives, 18(1), 163-190.

Loayza, N., Fajnzylber, P., & Lederman, D. (2000). Crime and victimization: An economic perspective. Economía Journal, 1(Fall 2000), 219-302.

Londoño, J. L., & Guerrero, R. (2000). Violencia en América Latina: epidemiología y costos. Asalto al desarrollo. Violencia en América Latina, 11-57.

Matsueda, R. L. (1989). The dynamics of moral beliefs and minor deviance. Social Forces, 68(2), 428-457.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. Journal of econometrics, 142(2), 698-714.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. Journal of the American Statistical Association, 106 (493), 100-108.

Otsu, T., Xu, K. L., & Matsushita, Y. (2013). Estimation and inference of discontinuity in density. Journal of Business & Economic Statistics, 31(4), 507-524.

Reinhart, A., & Greenhouse, J. (2017). Self-exciting point processes with spatial covariates: modeling the dynamics of crime.

Sah, R. K. (1991). Social osmosis and patterns of crime. Journal of Political Economy, 99(6), 1272-1295.

Sampson, R. J., & Groves, W. B. (1989). Community structure and crime: Testing social-disorganization theory. American journal of sociology, 94(4), 774-802.

Sandler, D. H., & Sandler, R. (2014). Multiple event studies in public finance and labor economics: A simulation study with applications. Journal of Economic and Social Measurement, 39(1-2), 31-57.

Short, M. B., D'orsogna, M. R., Pasour, V. B., Tita, G. E., Brantingham, P. J., Bertozzi, A. L., & Chayes, L. B. (2008). A statistical model of criminal behavior. Mathematical Models and Methods in Applied Sciences, 18, 1249-1267.

Databases

Crime data:
https://datos.cdmx.gob.mx/explore/dataset/carpetas-de-investigacion-pgj-cdmx/export/

The population of Mexico City:
https://www.inegi.org.mx/programas/intercensal/2015/default.html#Tabulados

Appendix

Describing the work from McCrary, the binning estimator generates regressands by counting the number of observations in each bin. The implementation comes in two steps. The first one is to construct the histogram with the bins carefully defined so no one extreme includes points on both parts of the discontinuity. The second step is to use the midpoints of the histogram bins as a regressor for local linear regression. The normalization of observations within the bin is used as a dependent variable. To get the possible discontinuity, local linear smoothing is conducted separately from the bins around the treatment τ_0. Nevertheless, I estimate both sides of the logarithmic transformation jointly to avoid negative point estimates for the density.

First, I create the histogram based on the frequency table of a discretized version of time and location. For that, I create a grid of time and distance equispaced by ϕ_D (ϕ_T) for distance (time) covering the support of the histogram. This is defined by $\left\{T_j^{bin}\right\}_{j=1}^{J} = [(2j - J - 1)\phi_T]$ for time and $\left\{D_k^{bin}\right\}_{k=1}^{K} = [(2k - 1)\phi_D]$ for distance, where J and K are even. Then, Y_{kj} define the normalized cellsize for the jkth bin as:

$Y_{jk} = \frac{1}{I(\phi_D \phi_T)}\{(T, D) \in R^2 | \sum_i (T_j^{bin} - \phi_T \leq T_i < T_j^{bin} + \phi_T \ \wedge \ D_k^{bin} - \phi_D \leq D_i < D_k^{bin} +$
$\phi_T \ )\}$

In words, I have an equispaced grid for both variables $X_{jk} = \{T_j^{bin}, D_k^{bin}\}$ and the counted observations for each bin is Y_{jk}[9]. Then, the histogram is the scatterplot (X_{jk}, Y_{jk}).

For the second step, I smooth the histogram by using linear regression. The problem becomes

$$\min_{\delta}\left\{\sum_{jk}^{JK} w(X_{jk})\left(y_{jk} - v(X_{jk}, \delta)\right)^2\right\}$$

[9] I denote $y_{jk} = \log{(Y_{jk} + \epsilon)}$ as explained before, where ϵ prevent undefinition.

This step smooths the histogram by estimating a weighted regression using the midpoint X_{jk} to explain the height of each bin. Triangular Kernel Function gives more weight to the bins nearest where I try to estimate the density.

According to Otsu et al. (2013), the are four issues with the t-statistic test for discontinuity $(\theta = 0)$ proposed by McCrary. First, is necessary to find the formula and estimator of the variance for each form of the kernel function. Second, the local linear estimator based on a non-negative sample may produce negative estimates. Third, it does not generate a confidence test for the null hypothesis $\theta = f_{left} - f_{right}$ since the test is constructed to test the log difference. And, finally, the Wald test will take different values in finite samples. The way to sort these issues is to test jointly the left and right log-density and apply the bootstrapping technique in a direct least-squares estimation.