Università degli Studi di Genova
DISEFIN – Series of Economic Working Papers
16126 Genova – via vivaldi 5 – Fax +39 010 209 5223

The Statistical Analysis of Crime Data at Street Level: Models Comparison

Enrico di Bella, , Lucia Leporatti, Luca Persico

wp n. 4
May 2012

“DISEFIN Working Papers on line”
series of economic working papers
published online by

Research Doctorate in
Public Economics
Ph.D School in New technologies and Social Sciences
University of Genoa

Web site:
http://www.disefin.unige.it/

The Statistical Analysis of Crime Data at Street Level: Models Comparison

Enrico di Bella
University of Genoa, Italy

Lucia Leporatti
University of Genoa, Italy

Luca Persico
University of Genoa, Italy

Abstract

The main techniques used for the spatial analysis of Urban Crime can generally be traced to crime mapping techniques, which are mere representations of crime dispersion over a specific urban area without any statistical modeling of its correlation with the urban structure of the city or any group of socio-demographic and economic variables. In this work, as a proposal to overcome the aforesaid limitation, we analyze the crime occurrences, recorded at street level, in a highly populated district of the City of Genoa, and we use different statistical models to study crime events in relationship with the context in which they happened, interpreting the urban layout of the roads network as a lattice

Keywords: *Urban Crime Analysis, Lattice Models, Space Syntax, Spatial Models*

Jel Classification: C21, C31, K42, O18, P25

1. Introduction

Crime analysis can be defined as the systematic and timely study of crime events and disorder problems, in order to assist the police and the local governments in the reduction, prevention and evaluation of criminality (Gottlieb *et al.*, 1994; Emig *et al.*, 1980, Vellani & Nahoun, 2001, Boba, 2012).

Crime is certainly a multidisciplinary topic and its analysis needs a deep comprehension of all the factors responsible for crime occurrences, particularly the demographic and economic features, as well as the spatial and temporal characteristics of crimes. Sociologists and criminologists identify as risk factors mainly socio-demographic elements, such as the residential density within urban areas, the inequality in income distribution, the level of poverty and the lack of education; on the other hand, planners and architects focus the analysis of crime on the influence on crime events, of the spatial configuration of a specific area. Both these approaches have been highly explored during the last past years; as a consequence a dense literature exists on the factors able to affect the criminality of a specific area and on the techniques introduced to evaluate and understand the relative risk of different areas.

The interest in the analysis of crime in its place (Weisburd *et al.*, 2009). has received a great boost thanks to development in computer technologies and to the availability of electronic and geographical data; as a result new spatial statistical methods have been introduced to facilitate the work of local governments and law enforcement in the fight against crimes and social disorders, pushed by the idea that "space matters".

One of the most common approach used by the police and by the local governments to understand the distribution of crime over a defined area (e.g. a city) is the crime mapping. The modern crime mapping techniques permit to visualize crime events on a map in order to analyze the distribution of crimes in the space and to identify crime incidents patterns (Boba, 2012; Weisburd & McEwen, 2010; Paulsen & Robinson, 2009). The aim of crime mapping is to geographically locate criminal occurrences in order to detect hot spots and to organize efficiently police interventions and patrolling duties in these crime concentrated areas through the use of a Geographic Information System (GIS) (Chainey & Ratcliffe, 2005, Murray *et al.*, 2001). Using different types of graphical representations, such as point maps, areal maps, line maps and buffer maps, (Boba, 2012) it is easy to detect the exact point or area in which a crime occurred with a high level of precision, moreover it is possible to obtain further information on the time of the occurrences and on the features of the victims and of the offenders. Finally by comparison the maps recorded in different periods, the authorities can study the evolution of crime over time to identify trend of long and short period.

In addition to a visual analysis of crime events, an important step towards the interpretation of spatial patterns of urban crime, is the study of the specific socio-demographic, economic and configurational characteristics for the analyzed area and, more specifically, the places where criminal events occur.

The interest in geographic criminology begins during the 19th century in France and in Belgium after the publication of the first geographical map of crime: in 1829 Michel Andrè Guerry and Adriano Balbi published a paper map representing the distribution of crime over the departments of France between 1825 and 1827. The interest of Guerry was not limited to the representation in map of statistical data; indeed, in 1833, he analyzed the distribution of crimes according to the poverty, the lack in education and the density of population of the departments but he concluded that these variables are not directly causes of crime occurrence. Some years later, the Belgian statistician and astronomer Quetelet (1984) examined crimes in relation to poverty claiming that most crimes were committed in departments where people from inferior classes lived:

the only exception was found for crimes against properties which, obviously, were more frequent in wealthy provinces. The attention given to space and context in criminology significantly developed during the 19th century thanks to the members of the School of Chicago; the most active members of the School, among the others, Robert Park, Wiliam Thomas, Ernest Burgess, Louis Wirth, Clifford Shaw and Henry McKay, gave impulse to the development of theories on the criminology of place. In particular they focused on the understanding of the causes of crime in the American cities: they identified the social disorganization (Thomas, 1966), the poverty, the racial heterogeneity and the residential mobility as reasons for the occurrence of high levels of crime. Furthermore they tried to identify similar patterns of crime among different cities, to understand if they were associated with similar socio-economic environments (Shaw & McKay, 1942).

Starting from 1960s Jacob (1961), Newman (1972) and Jeffery (1977) developed the idea of a connection between crime and the environmental features; they developed the so-called Ecological Theory of crime which supports the idea of a need for natural surveillance - the so-called *eye of the street* - (Jacobs, 1961) to reduce the number of crime committed; in detail they pointed out the importance of creating spaces where residents could gather, in order to increase the liveliness of suburbs, reducing the urban decay and the fear of crime; these ideas paved the way, during the 1980s, to the development of "situational crime prevention" (Clarke, 1983,1992,1995). According to the followers of the situational crime prevention, in order to reduce the number of crimes, it is necessary to reduce the opportunities of committing a crime, because "opportunity makes the thief" (Felson and Clarke, 1998). These ideas led to the increase in the attention, for crime analysts, for urban design details (such as turnings, street lighting, access streets, housing design) and to a deep study of the spatial configuration of the streets conducted through the Space Syntax methodology (Hillier, 1988). The Space Syntax analysis studies, through quantitative measures, the configurational properties of urban space (Hillier & Henson, 1984) and paved the way to the improvement of the Crime Prevention Through Environmental Design (CPTED) (Jeffery, 1977). Thanks to the increasing in the number of measures used in the Space Syntax Analysis, it became possible to compute the relative degree of accessibility, connection and integration of each street in its urban network. Among the others, Beavon, Brantingham and Brantingham (1994) analyzed the street structure and its dependence with crime volumes: they found out that streets with many twist and turns are fuller of crimes.

During the last thirty years a new theory on the widespread of crime emerged: according to the Routine Activity Theory the number of crimes increase if the number of opportunities for criminals rise and if the society lacks an adequate surveillance against crime (Cohen & Felson, 1979); indeed crimes are often committed in places where the victims and the offenders hold their routine activities, for example work, leisure or social interaction and where they satisfy their basic needs (Eck & Weisburd, 1995). This theory focuses on space because place is considered an explicit cause of the human actions, including committing offences. Some empirical studies are in favor of this theory: Cohen and Felson (1979) used the Routine Activity Theory to explain the increase in the number of crimes in American cities; as instance they pointed out that, with more women working, a larger number of houses were empty during the daytime and this fact led to the rise in the number of robberies increasing the vulnerability of suburbs. Roncek (1981) found out that in Cleveland streets with schools and bars are highly crime dense, while Rice and Smith (2002) and Smith and Clarke (2000) identified the places near commercial stores as particularly risky. In this context, some studies on the relationship between crimes and transports have been developed by Smith and Clarke (2000) and by Block and Davis (1996): they conclude that the structure of the public transport system can influence the number of crimes committed: in fact higher levels of crimes are recorded near stations and bus stops.

So, during recent years a new interest for a combined study of socio-demographic and spatial factors in the analysis of crime has emerged. In fact, although crime mapping is certainly the more immediate way to obtain quick information on an area, it could be interested to study the relationship between urban crimes and the socio-demographic and spatial features of an area. Indeed the study of crime in the context in which it happens could bring to the identification of local risk factors helping the local governments in drawing up policies for Urban Safety.

In the last thirty years, some authors have already explored the relationship between crimes, environment and social and demographic features. Weisburd *et al.* (2009) collected some papers presented in September 2006 at the Netherlands Institute for the Study of Crime and Law Enforcement; all the articles focus on the simultaneous analysis of crime and place and they highlight different factors responsible for crime increase.

With these considerations in our mind, in this work we propose a way to integrate a spatial analysis of crime, studying it in relationship with various context variables suggested by the scientific literature (Wilsem, 2009, Nubani & Wineman, 2005). The innovative approach herein proposed is based on the possibility of linking all these information at street level so that each street becomes a statistical unit to which we associate its crime occurrences and its particular values of the context variables. In particular we compare the effectiveness of a statistical model using three sets of context variables: socio-demographic variables, economic variables and configurational variables (representing the topographic structure of the urban layout).

Although it is not possible to express a precise causal effect of crime occurrences from many of the variables whose use we propose in the following, it is anyhow reasonable to sustain their correlation. The models presented in this work should be thought as instruments to identify concomitant variables which should help the public decision makers (i.e. Prefect, Municipal Councillor and Directors) to understand why in some streets there are more occurrences than in others and to suggest on which variables they should act to give a proper response to crime hot spots.

In this work we compare the results of some of the most commonly used statistical models whose characteristics are adequate for a first explorative study on the data available. We started the analysis with the construction of a traditional stepwise linear model, passing through spatial models (simultaneous autoregressive and conditional autoregressive) and concluding the discussion using count data models (negative binomial and zero inflated negative binomial). Proper and more advanced models (e.g. Besag *et al.* 1991) are under implementation and their results will be published in some future works.

2. Data Available

The data used in this work derive from three databases owned by different Institutions and their convergence to a unique dataset has been possible thanks to an official collaboration between the Municipality of Genoa and the University of Genoa; all data are available at street level.

The first dataset, provided by the Territorial Office of the Municipality of Genoa, contains the urban graph of a highly populated district of the City of Genoa: each street, called axis, can be geocoded using a coordinate system and it is defined by an initial and final point. The streets are therefore part of a connected urban network, or graph, composed, in total, by 83 streets; the network configuration has been analyzed according to a technique named Space Syntax Analysis (SSA). SSA provides a method for partitioning a spatial system into relatively independent but connected subspaces so that the importance of these subspaces can be measured in terms of their relative nearness or accessibility (Hillier & Hanson, 1984). A vast number of variables can be measured on each single axis of an Urban Graph through the means of the SSA; among them, we chose a set of variables which proved itself to be particularly effective in this context (di Bella *et al.*, 2011): Integration (a measure of the centrality of an axis in the urban graph), Choice (the ratio of geodetic paths in the urban graph including a specific axis) and the Line-Length (a measure of the length of a street). We decided to include these variables in the analysis because some literature inspired by the Ecological Theory (Newman, 1973) have found strong correlations between certain types of crime and spatial and configurational features of the streets (Rengert, 1980; Shu, 2000; Hillier & Shu, 2000).

The second dataset contains the complaints collected by the local Carabinieri Station from 01/01/2009 up to 27/07/2010; all the crime events are geo-coded at street level, on the basis of the place in which the crime occurred. These data are classified according to a standardized sorting of types of crime in three main different categories: violent crimes (e.g. attacks and murders), predatory crimes (crimes agaist property: thefts, robberies) and damages and fires (actions of vandalism).

Table 1: Distribution of crimes on the 83 streets

	CRIMES_ TOT		CRIMES_DAMAGES		CRIMES_PREDATORY		CRIMES_VIOLENT	
	%	NUM OF STREETS	%	NUM OF STREETS	%	NUM OF STREETS	%	NUM OF STREETS
0	13%	11	29%	24	18%	15	70%	58
1-10	46%	38	48%	40	57%	47	30%	25
11-20	12%	10	6%	5	12%	10	0%	0
21-30	12%	10	10%	8	2%	2	0%	0
>31	17%	14	7%	6	11%	9	0%	0
TOTAL	100%	83	100%	83	100%	83	100%	83

This division permits to identify different patterns of crime for different typologies of offences; in fact, different types of crime are often associated with different social and spatial features (Dunn, 1980). As Hillier and Sahbaz (2008) said different crimes are facilitated by different kinds of space; for example muggings are more frequent in crowded streets, while burglary is easier in isolated and uncrowded streets. The largest category of crimes is composed by crimes against properties accounting for 916 crimes in the period considered; the second category of crimes is composed by damages and fires (700), while the smallest typology of crimes is formed by violent crimes (46). As is evident from Table 1 a lot of streets count for zero crimes, in particular, for what concerns the violent crimes for which the 70% of the streets don't record any crimes.

The third and last dataset, provided by the Demographic Office of the Municipality of Genoa, refers to demographic and economic characteristics of the district at street level (e.g. residents per axis, number and types of commercial activities, number of recreation activities).

As a first step of the analysis we have selected a set of variables from the dataset available according to their explicative capacity. For what concerns the graph of Genoa and the SSA measures, the variables considered in the analysis are Integration, Line-Length and Choice (di Bella *et al.*, 2011). According to the demographic variables we have selected as particularly influencing the number of residents for each street, the number of residents for nationality and the percentage of foreign people; in particular we introduced in the analysis the foreign people coming from the main ethnic communities in Genoa, which are Ecuador, Albania, Romania, Morocco and Peru. Finally we have considered the socio-economic layout of the street, in particular the total number of shops, the number of recreation activities and the number of shops authorized to sell alcohol as a measure of the alcoholic risk of the street (Gorman *et al.*, 2001).

These variables have gone through additional re-elaboration, in order to create new variables, from the aforesaid ones, particularly important for the description of some demographic and economic phenomena. In detail we compute the variables Residential Vocation, obtained as the ratio between the number of shops with a mainly residential vocation (i.e. whose main business is for local residents) on the total number of shops, and the Commercial Vocation as the ratio between the shops with a mainly commercial vocation (i.e. whose main business attracts clients from other areas) on the total number of shops. Moreover, in order to analyze properly the presence of foreign people in the district, we use the division of the nationalities of origin of the residents according to the level of the Human Development Index (HDI) proposed by the United Nations Development Programme (UNDP, 2011); we computed the number of residents in each street for each level of HDI in order to detect a possible dependence of crime on the medium and low level of this index (variable HDI_LOW).

Although data are available at single-street level, in order to protect the privacy of the information, data are herein presented in an aggregate form, without any particular mention of the names of streets.

Table 2 and Table 3 contain a brief description of the variables included in the stepwise procedure and some descriptive statistics. As is clear from Table 3 the lengths of street and population size vary significantly across the streets.

Table 2 Variables considered

Variables code	Variable description
CRIMES_ TOT	Total number of crimes
CRIMES_DAMAGES	Number of damages and fires
CRIMES_PREDATORY	Number of crimes against properties (theft, robberies)
CRIMES_VIOLENT	Violent crimes against person (murders, attacks)
LINE LENGTH	Standardized measures of the length of a street
CHOICE	The ratio of geodetic paths in the urban graph including a specific axis; it is a standardized variable
INTEGRATION	Centrality of an axis in the urban graph. It is a standardized variable
NUMBER OF SHOPS	Total number of shops in the street
COMM_VOCATION	Ability of the shops to attracts clients from other areas
RES_VOCATION	Shops with a main business addressed to local residents
RES_0.18	Number of residents (age: 0-18) in percentage of the total number of residents
RES_65	Number of residents (age: > 65) in percentage of the total number of residents
RES_TOT	Total number of residents
ALCOHOL_SHOPS	Number of shops authorized to sell alcohol divided by the line length
HDI_LOW	Number of born in countries with low human development index level
PERC_FOREIGN	Percentage of foreign people
PERC_ECUADOR	Percentage of residents born in Ecuador of the total number of foreign people.
PERC_ALBANIA	Percentage of residents born in Albania of the total number of foreign people
PERC_ROMANIA	Percentage of residents born in Romania of the total number of foreign people
PERC_MOROCCO	Percentage of residents born in Morocco of the total number of foreign people
PERC_PERU	Percentage of residents born in Peru of the total number of foreign people

Table 3 Variables considered – descriptive statistics –

	Mean	Standard Deviation	Min	Max
CRIMES_ TOT	20.54	31.23	0.00	160.00
CRIMES_DAMAGES	8.43	12.34	0.00	74.00
CRIMES_PREDATORY	11.54	19.29	0.00	81.00
CRIMES_VIOLENT	0.55	1.17	0.00	6.00
LINE LENGTH	425.37	488.74	16.08	3,031.32
CHOICE	0.00	1.00	-0.38	5.35
INTEGRATION	0.00	1.00	-5.30	1.37
NUMBER OF SHOPS	10.19	23.44	0.00	133.00
COMM_VOCATION	1.95	3.92	0.00	17.13
RES_VOCATION	2.53	4.35	0.00	18.73
RES_0.18	0.14	0.05	0.00	0.37
RES_65	0.26	0.08	0.00	0.56
RES_TOT	519.08	608.82	3.00	3,192.00
ALCOHOL_SHOPS	0.00	0.00	0.00	0.02
HDI_LOW	38.45	56.68	0.00	302.00
PERC_FOREIGN	0.10	0.07	0.00	0.40
PERC_ECUADOR	0.21	0.18	0.00	0.61
PERC_ALBANIA	0.08	0.11	0.00	0.67
PERC_ROMANIA	0.09	0.13	0.00	1.00
PERC_MAROCCO	0.03	0.07	0.00	0.33
PERC_PERU’	0.07	0.10	0.00	0.60

3. Statistical Models

In this work we estimate five different models on the three typologies of crimes (crimes against properties, damages and fires, violent crimes) and on the total number of crimes, for a total of 20 models. To keep the results obtained comparable among the models estimated, we defined a set of explicatory variables, from the ones described in section 2, using a mixed stepwise procedure to select them. The variables selected by the stepwise procedure are therefore used in all the subsequent models for each typology of crime.

3.1 Stepwise Models

The stepwise procedure permits to select automatically the most explicative variables among all the ones available. The most widely used approaches are three: the forward procedure, the backward procedure and a mixed model. Forward selection starts with no variables in the model, trying out the variables one by one and including them if they are *statistically significant*. Conversely backward elimination starts with all candidate variables included in the model and it tests them one by one for statistical significance, deleting it if not significant. In this work we used a mixed method which is a combination of the above, testing at each stage for variables to be included or excluded. As a measure of goodness of fit we adopt Akaike's Information Criterion (AIC) which includes a penalty that is an increasing function of the number of estimated parameters (Akaike, Hirotugu, 1974).

In order to make the results comparable, we collected all the most explicative variables, derived from the stepwise procedure, from the three different categories of crimes and we created a simple linear model for each crime typology.

The more explicative variables selected by the stepwise model are related to the three different categories of variables; in fact the model selects some spatial, some demographic and some socio-economic variables as highly explicative. In particular the most representative spatial variables are Line-Length, Integration and Choice: the first and the second variables have a positive influence on crimes: this means that in more integrated and long street higher number of crimes occurs. On the contrary high level of the variable Choice acts as a deterrent on committing crimes. According to demographic variables the most important one is the number of residents (RES_TOT) while the division in range of ages does not appear significant (RES_0.18, RES_>65); also the measures of the Human Development Index of the countries from which the residents come from have a significant influence on crimes: in particular residents from country of origin with low level of HDI seem to increase the number of damages and fires occurred. The number of shops in a street is positive correlated with the number of crimes because of the capacity of shops to attract people and to increase the opportunities of committing crimes.

All the four models look explicative, with high levels of R squared especially for what concerns the crimes against properties, which form the most numerous class, where R^2 reaches a value of 84%.

Table 4 Linear Model – Damages and fires –

CRIMES_DAMAGES	Estimate	Std Error	t value	Pr(>|t|)
(Intercept)	0.4364	1.8651	0.2340	0.8156
INTEGRATION	2.4530	0.9494	2.5840	0.0117 *
CHOICE	-2.2690	1.0662	-2.1280	0.0366 *
LINE LENGTH	0.0135	0.0025	5.2930	0.0000 ***
RES_TOT	-0.0031	0.0040	-0.7850	0.4350
NUMBER OF SHOPS	0.0721	0.0730	0.9890	0.3259
HDI_LOW	0.1036	0.0482	2.1490	0.0349 *
PERC_FOREIGN	-8.3721	14.3093	-0.5850	0.5603
R^2	0.66			

Table 5 Linear Model – Crimes against properties –

CRIMES_PREDATORY	Estimate	Std Error	t value	Pr(>|t|)
(Intercept)	0.7337	1.9782	0.3710	0.7118
INTEGRATION	2.7494	1.0069	2.7300	0.0079 **
CHOICE	-1.7685	1.1308	-1.5640	0.1220
LINE LENGTH	0.0093	0.0027	3.4580	0.0009 ***
RES_TOT	-0.0005	0.0042	-0.1170	0.9071
NUMBER OF SHOPS	0.5720	0.0774	7.3920	0.0000 ***
HDI_LOW	0.0433	0.0511	0.8460	0.4001
PERC_FOREIGN	-4.0555	15.1767	-0.2670	0.7900
R^2	0.84			

Table 6 Linear Model – Violent crimes –

CRIMES_VIOLENT	Estimate	Std Error	t value	Pr(>|t|)
(Intercept)	-0.1489	0.2149	-0.6930	0.4904
INTEGRATION	0.0935	0.1094	0.8550	0.3954
CHOICE	-0.1625	0.1228	-1.3230	0.1899
LINE LENGTH	0.0007	0.0003	2.5010	0.0146 *
RES_TOT	0.0002	0.0005	0.5350	0.5943
NUMBER OF SHOPS	0.0232	0.0084	2.7550	0.0074 **
HDI_LOW	-0.0003	0.0056	-0.0520	0.9590
PERC_FOREIGN	0.3865	1.6485	0.2340	0.8153
R^2	0.51			

Table 7 Linear Model – Total crimes –

CRIMES_TOT	Estimate	Std Error	t value	Pr(>|t|)
(Intercept)	0.9979	3.6086	0.2770	0.7829
INTEGRATION	5.2948	1.8369	2.8830	0.0051 **
CHOICE	- 4.2432	2.0628	-2.0570	0.0432 *
LINE LENGTH	0.0236	0.0049	4.7850	0.0000 ***
RES_TOT	- 0.0033	0.0077	-0.4320	0.6667
NUMBER OF SHOPS	0.6735	0.1412	4.7720	0.0000 ***
HDI_LOW	0.1443	0.0933	1.5480	0.1259
PERC_FOREIGN	-11.8030	27.6851	-0.4260	0.6711
R^2	0.80			

In spite of the high level of goodness of fit of the models, the stepwise model presents some relevant limitations. The principal limit of this model is the fact that it does not take into account the autoregressive spatial component that can be significant in spatial data on crimes; in fact a vast literature has explored the spatial dependency pattern of crime events, pointing out that units tend to be influenced by neighbor ones (e.g. Morenoff *et al.*, 2001; Wilsem, 2009; Cohen & Tita, 1999). This limit can be ride out using an appropriate autoregressive spatial model; as a result, in section 3.2 and 3.3, we describe the construction of a SAR and of a CAR model. Another problem is the fact that the prevision on crimes could also assume negative values: in this context the negative predictions have been considered as zero crimes.

3.2 Simultaneous Autoregressive Models

The introduction of a autoregressive model to study criminality is largely justified by the fact that the activities in one area have an influence in the neighborhood (Morenoff *et al*., 2001; Sampson, 2004); many applications of spatial autoregressive models in criminology involve the study of the gangs wars or drug markets, which are phenomena particularly spatially dependent (e.g Decker 1996; Cohen & Tita, 1999; Griffiths & Chavez, 2004, Tita & Greenbaum, 2009).

Simultaneous autoregressive models assume that the response variable at each location i is a function of the explanatory variables at i but that it depends also on the values of the response variable at close locations *j* as well (Cressie, 1993; Lichstein *et al.* 2002). The specification of the spatial dependence in crime analysis often follows the Tobler's First Law of Geography (Tobler, 1970) which states that "everything is related to everything else, but near things are more related than distant things". In SAR model, the neighborhood relationship is formally expressed using a $n \ x \ n$ matrix of spatial weights (**W**), where elements (w_{ij}) represent a measure of the connection between locations i and j. In particular in this work we have considered as neighbors all the adjacent streets building up a symmetric matrix of neighborhood. Other kind of neighborhood matrices can be constructed for example defining them in terms of distance instead of in term of adjacency (Doreian, 1980). After the definition of the neighborhood structure we have given a weight of one to neighbor streets and a weight of zero to streets that are not considered close one to each other. At this point it is possible to build up a spatial model that takes into account the spatial pattern of the residuals, modeling the eventual spatial dependency of crime.

If the vector of response variables is multivariate normal, we can formally express the SAR model as follows (Cressie, 1993):

$$Y_i = \mu_i + \delta_i \tag{1}$$

where Y_i is the random process in i, μ_i is the mean in i, and δ_i is the i-th value of a vector of normally distributed random errors with zero mean and covariance matrix:

$$\boldsymbol{\Sigma} = [(\boldsymbol{I} - \rho\boldsymbol{N})^T \boldsymbol{D}^{-1} (\boldsymbol{I} - \rho\boldsymbol{N})]^{-1} \sigma^2 \tag{2}$$

where ρ represents the spatial autocorrelation parameter and σ^2 is the variability measure, $\boldsymbol{N}$ is a weighted neighborhood matrix and $\boldsymbol{D}$ is a diagonal matrix used to account for nonhomogeneous variance of the marginal distributions; in this context the $\boldsymbol{D}$ matrix is considered as a identity matrix $\boldsymbol{I}_{nxn}$. The small scale variation due to interactions with neighbors is modeled by fitting an auto-regressive covariance model to **Σ**. To take into account the covariance structure of data we can express the model with the following equation:

$$Y = \boldsymbol{X}\beta + \rho\boldsymbol{N}(y - \boldsymbol{X}\beta) + \boldsymbol{W}^{\frac{1}{2}}\varepsilon \tag{3}$$

where $\boldsymbol{X}\beta$ is the linear trend, $\rho\boldsymbol{N}(y - \boldsymbol{X}\beta)$ the covariance structure and $\boldsymbol{W}^{1/2}\varepsilon$ the noise of model.

As is clear from the results tables of the SAR model the majority of the coefficients look significant as they have been already selected through a stepwise procedure. Choice maintains negative influence on crimes, in particular for the damages and fires: this can be reasonable, in fact vandalism actions often occurred in not crowded roads. On the contrary, obviously, the longer the street is, the riskier it is in terms of absolute number of crimes. It could be interesting to divide the number of crimes occurred in a street by the Line-Length of the street in order to obtain a measure of risk independent from the length of the street. The presence of shops have a positive influence on the occurrences of crimes: this connection is particularly obvious for the crimes against properties like thefts and muggings who often occurred near shopping centers or markets.

It is important to specify that the autoregressive component Rho does not result significant for the three typology of crimes (see Tables 8 - Tables 10): it is significant only for the model of total number of crimes probably because this is the larger category of crimes including all the three previous mentioned subcategories. So it does not seem to exist a significant autoregressive pattern in crime data analyzed.

Table 8 SAR Model – Damages and fires –

CRIMES_DAMAGES	Estimate	Std Error	t value	Pr(>|t|)
(Intercept)	0.3092	1.8501	0.1671	0.8673
INTEGRATION	2.6876	0.9411	2.8559	0.0043 **
CHOICE	-2.1610	1.0499	-2.0582	0.0396 *
LINE LENGTH	0.0131	0.0024	5.5306	0.0000 ***
RES_TOT	-0.0026	0.0036	-0.7134	0.4756
NUMBER OF SHOPS	0.0251	0.0688	0.3645	0.7155
HDI_LOW	0.1093	0.0436	2.5065	0.0122 *
PERC_FOREIGN	-6.3366	13.7003	-0.4625	0.6437
Rho	0.0609			0.1507
R^2	0.68			
AIC	580.2			

Table 9 SAR Model – Crimes against properties –

CRIMES_PREDATORY	Estimate	Std Error	z value	Pr(>|z|)
(Intercept)	0.8075	1.9553	0.4130	0.6796
INTEGRATION	2.6879	0.9954	2.7003	0.0069 **
CHOICE	-1.7540	1.1109	-1.5789	0.1144

LINE LENGTH	0.0093	0.0025	3.7068	0.0002 ***
RES_TOT	-0.0013	0.0038	-0.3344	0.7381
NUMBER OF SHOPS	0.5300	0.0729	7.2734	0.0000 ***
HDI_LOW	0.0669	0.0463	1.4458	0.1482
PERC_FOREIGN	-7.6952	14.5117	-0.5303	0.5959
Rho	0.0585			0.1344
R^2	0.85			
AIC	589.79			

Table 10 SAR Model –Violent crimes –

CRIMES_VIOLENT	Estimate	Std Error	z value	Pr(>lzl)
(Intercept)	-0.1480	0.2018	-0.7334	0.4633
INTEGRATION	0.1040	0.1021	1.0189	0.3083
CHOICE	-0.1786	0.1148	-1.5551	0.1199
LINE LENGTH	0.0007	0.0003	2.5878	0.0097 **
RES_TOT	0.0003	0.0004	0.6348	0.5255
NUMBER OF SHOPS	0.0245	0.0080	3.0842	0.0020 **
HDI_LOW	-0.0008	0.0053	-0.1554	0.8765
PERC_FOREIGN	0.3744	1.5515	0.2413	0.8093
Rho	-0.0161			0.6728
R^2	0.51			
AIC	223.34			

Table 11 SAR Model – Total crimes –

CRIMES_TOT	Estimate	Std Error	z value	Pr(>lzl)
(Intercept)	0.8837	3.5863	0.2464	0.8054
INTEGRATION	5.4727	1.8055	3.0311	0.0024 **
CHOICE	- 3.9549	2.0088	-1.9688	0.0490 *

LINE LENGTH	0.0232	0.0045	5.1937	0.0000 ***
RES_TOT	- 0.0038	0.0067	-0.5681	0.5700
NUMBER OF SHOPS	0.5425	0.1303	4.1620	0.0000 ***
HDI_LOW	0.1885	0.0810	2.3260	0.0200 *
PERC_FOREIGN	-15.0836	25.9865	-0.5804	0.5616
Rho	0.0823			0.0214
R^2	0.82			
AIC	686.52			

The goodness of fit of the models remain high, ranging from 51% for violent crimes to 85% for crimes against properties. Also the SAR model produce negative predictions: for this reason all the streets with negative predictive values are considered as street with no crimes estimated.

As we said before, the spatial autoregressive component does not result significant; this fact can have various reasons. First of all it is possible that the Space Syntax Variables capture themselves the spatial correlation structure; furthermore the small dimension of the district can affect the analysis. The next step of our analysis consists in the computation of the CAR models, in order to understand if the changing of the covariance matrix could modify the results.

3.3 Conditional Autoregressive Models

What is different between a SAR and a CAR model is the covariance matrix; in particular for the CAR model it is expressed by the following equation:

$$\boldsymbol{\Sigma} = (\boldsymbol{I} - \rho \boldsymbol{N})^{-1} \boldsymbol{D} \sigma^2 \tag{4}$$

where ρ represents the spatial autocorrelation and σ^2 is the variability measure, $\boldsymbol{N}$ is a weighted neighborhood matrix and $\boldsymbol{D}$ is a diagonal matrix used to account for nonhomogeneous variance of the marginal distributions.

Results on the CAR model are much similar to the ones on the SAR model in terms of sign and significance of the coefficients and in terms of goodness of fit. Also using a CAR model the spatial autoregressive component does not result significant.

Table 12 CAR Model – Damages and fires –

CRIMES_DAMAGES	Estimate	Std Error	z value	Pr(>|z|)
(Intercept)	0.2975	1.8384	0.1618	0.8714
INTEGRATION	2.5760	0.9361	2.7519	0.0059 **
CHOICE	-2.2420	1.0467	-2.1421	0.0322 *
LINE LENGTH	0.0132	0.0024	5.5355	0.0000 ***
RES_TOT	-0.0027	0.0037	-0.7480	0.4545
NUMBER OF SHOPS	0.0402	0.0691	0.5811	0.5612
HDI_LOW	0.1074	0.0444	2.4202	0.0155 *
PERC_FOREIGN	-6.8572	13.6883	-0.5010	0.6164
Rho	0.0819			0.2117
R^2	0.68			
AIC	580.7			

Table 13 CAR Model –Crimes against properties –

CRIMES_PREDATORY	Estimate	Std Error	z value	Pr(>|z|)
(Intercept)	0.7183	1.9521	0.3680	0.7129
INTEGRATION	2.6436	0.9931	2.6619	0.0078 **
CHOICE	-1.8007	1.1102	-1.6219	0.1048
LINE LENGTH	0.0093	0.0025	3.6748	0.0002 ***
RES_TOT	-0.0010	0.0039	-0.2544	0.7992
NUMBER OF SHOPS	0.5404	0.0732	7.3820	0.0000 ***

HDI_LOW	0.0599	0.0469	1.2753	0.2022
PERC_FOREIGN	-6.5398	14.4977	-0.4511	0.6519
Rho	0.0849			0.1756
R^2	0.85			
AIC	590.2			

Table 14 CAR Model – Violent crimes –

CRIMES_VIOLENT	Estimate	Std Error	z value	Pr(>|z|)
(Intercept)	-0.1483	0.2008	-0.7384	0.4603
INTEGRATION	0.1080	0.1012	1.0672	0.2859
CHOICE	-0.1869	0.1140	-1.6397	0.1011
LINE LENGTH	0.0007	0.0003	2.5610	0.0104 *
RES_TOT	0.0003	0.0004	0.6777	0.4980
NUMBER OF SHOPS	0.0252	0.0079	3.1744	0.0015 **
HDI_LOW	-0.0011	0.0053	-0.2065	0.8364
PERC_FOREIGN	0.3624	1.5433	0.2348	0.8144
Rho	-0.0481			0.6045
R^2	0.51			
AIC	223.25			

Table 15 CAR Model – Total crimes –

CRIMES_TOT	Estimate	Std Error	z value	Pr(>|z|)
(Intercept)	0.6836	3.5769	0.1911	0.8484
INTEGRATION	5.2259	1.8084	2.8898	0.0039 **
CHOICE	- 4.2352	2.0197	-2.0969	0.0360 *
LINE LENGTH	0.0233	0.0045	5.1181	0.0000 ***
RES_TOT	- 0.0035	0.0069	-0.5015	0.6160
NUMBER OF SHOPS	0.5906	0.1320	4.4729	0.0000 ***

HDI_LOW	0.1705	0.0840	2.0297	0.0424	*
PERC_FOREIGN	-13.1499	26.1386	-0.5031	0.6149	
Rho	0.1000			0.0505	
R^2	0.82				
AIC	687.99				

We can identify some problems in the use of spatial autoregressive model on these data. First of all, data are assumed to be normally distributed but this distribution does not fit well with the distribution of crime, a count dependent variable; for this reason in the next paragraph we will consider different models, more suitable for count data, based on a Negative Binomial distribution. Moreover the spatial autoregressive component Rho seems to be not significant: this may be due to the fact that the Space Syntax variables included in the model implicitly take into account the spatial autoregressive component. Furthermore the presence of a lot of streets with no crimes occurred can distort the results: a zero inflated model will be considered to face this problem.

3.4 Negative Binomial Model

In this work we consider as dependent variable the number of crime complaints. So, after a preliminary analysis with standard regression models, it is natural to consider models for counting data. The count models are used with non-negative integer responses to express the number of occurrences of an event. In a Poisson regression model the probability of the area *i* having y_i occurrences is given by:

$$P(y_i) = \frac{EXP(-\lambda_i)\lambda_i^{y_i}}{y_i!} \qquad (5)$$

where λ_i is the area parameter for i . To estimate the Poisson model it is necessary to specify the Poisson parameter λ_i, that represents the expected number of events per period, as function of the explicative variables:

$$\lambda_i = EXP(\beta X_i) \tag{6}$$

where X_i is a vector of explanatory variables and β is a vector of estimated parameters. The Negative Binomial Regression Model is obtained rewriting the previous equation adding a term of error ε_i with mean 1 and variance α^2:

$$\lambda_i = EXP(\beta X_i + \varepsilon_i) \tag{7}$$

The α parameter is considered an over-dispersion parameter. The addition of this term permits the variance to be different from the mean. The negative binomial distribution has the form:

$$P(y_i) = \frac{\Gamma(1/\alpha + y_i)}{\Gamma(1/\alpha) y_i!} \left(\frac{1/\alpha}{1/\alpha + \lambda_i}\right)^{\frac{1}{\alpha}} \left(\frac{\lambda_i}{1/\alpha + \lambda_i}\right)^{y_i} \tag{8}$$

where $\Gamma(.)$ is the Gamma function. Through the likelihood function (9) it is possible to estimate the parameters.

$$L(\lambda_i) = \prod \frac{\Gamma(1/\alpha + y_i)}{\Gamma(1/\alpha) y_i!} \left(\frac{1/\alpha}{(1/\alpha) + \lambda_i}\right)^{1/\alpha} \left(\frac{\lambda_i}{(1/\alpha) + \lambda_i}\right)^{y_i} \tag{9}$$

In practice, the Negative Binomial Model is derived from the Poisson Gamma mixture model and it is used to model count data when the mean and the variance cannot be considered equal (Hilbe, 2007). In this case, when the response variance is greater than the mean we talk about over-dispersion. Over-dispersion is caused by positive correlation between responses or by an excess variation between response probabilities or counts (Hilbe, 2007) and it brings to an underestimation of the standard errors. As is evident from Table 16 for all the response variables the variance is considerably greater than the mean; for this reason the use of Negative Binomial Model instead of a Poisson Model seems to be desirable.

Table 16 Descriptive statistics of the response variables

	Mean	**Variance**	**Max**
CRIMES_TOT	20.54	975.62	160.00
CRIMES_DAMAGES	8.43	152.32	74.00
CRIMES_PREDATORY	11.54	372.25	81.00
CRIMES_VIOLENT	0.55	1.38	6.00

We can interpret the negative binomial regression coefficient as follows: for a one unit change in the predictor variable, the difference in the logs of expected counts of the response variable is expected to change by the respective regression coefficient, assuming that the other variables are kept constant.

Using a Negative Binomial Model the significance of some of the coefficients changes: in the model on damages and fires some variables are no longer significant, in particular the variables Integration and HDI_LOW.

Table 17 NB Model – Damages and fires –

CRIMES_DAMAGES	Estimate	Std Error	t value	Pr(>|t|)
(Intercept)	1.1970	0.2464	4.8600	0.0000
INTEGRATION	0.1861	0.1449	1.2800	0.1990
CHOICE	-0.1682	0.0963	-1.7500	0.0810 .
LINE LENGTH	0.0008	0.0002	4.6300	0.0000 ***
RES_TOT	0.0001	0.0003	0.2900	0.7710
NUMBER OF SHOPS	0.0027	0.0065	0.3800	0.7010
HDI_LOW	0.0057	0.0041	1.4000	0.1610

PERC_FOREIGN	-1.0521	1.8586	-0.5700	0.5710
lnalpha	1.9916	0,2379		
alpha	7.3278	1.7440		
R^2	0.63			
LR chi2 =55.17 Prob > chi2 = 0.0000				
Likelihood-ratio test of alpha = 0: chibar2(01) = 284.38 Prob>=chibar2 = 0.000				

Table 18 NB Model – Crimes against properties –

CRIMES_PREDATORY	Estimate	Std Error	t value	Pr(>|t|)
(Intercept)	1.3920	0.2054	6.7800	0.0000 ***
INTEGRATION	0.4538	0.1448	3.1300	0.0020 **
CHOICE	-0.1618	0.0742	-2.1800	0.0290 *
LINE LENGTH	0.0007	0.0002	4.0000	0.0000 ***
RES_TOT	0.0003	0.0002	1.2300	0.2200
NUMBER OF SHOPS	0.0129	0.0047	2.7200	0.0060 **
HDI_LOW	0.0005	0.0031	0.1700	0.8610
PERC_FOREIGN	-0.3233	1.4940	-0.2200	0.8290
lnalpha	1.8057	0.2267		
alpha	6.0844	1.3795		
R^2	0.72			
LR chi2 =74.76 Prob > chi2 = 0.0000				
Likelihood-ratio test of alpha=0: chibar2(01) = 240.17 Prob>=chibar2 = 0.000				

Table 19 NB Model – Violent crimes –

CRIMES_VIOLENT	Estimate	Std Error	t value	Pr(>|t|)
(Intercept)	-1.7780	0.4574	-3.8900	0.0000 ***
INTEGRATION	0.3665	0.2797	1.3100	0.1900

CHOICE	-0.2178	0.1567	-1.3900	0.1650	
LINE LENGTH	0.0009	0.0003	2.9700	0.0030	**
RES_TOT	0.0003	0.0004	0.8100	0.4190	
NUMBER OF SHOPS	0.0130	0.0079	1.6400	0.1000	
HDI_LOW	0.0000	0.0055	0.0000	0.9960	
PERC_FOREIGN	-0.0812	3.6670	-0.0200	0.9820	
lnalpha	-2.6480	3.6796			
alpha	0.0708	0.2605			
R^2	0.43				
LR chi2 =32.88 Prob > chi2 = 0.0000					
Likelihood-ratio test of alpha=0: chibar2(01) = 0.08 Prob>=chibar2 = 0.386					

In order to test the effective presence of over-dispersion, we analyzed the dispersion parameter alpha. If the dispersion parameter equals zero, the model can be reduced to a Poisson model because the over-dispersion in data is not significant and the Poisson model is less expensive in term of computation. To test the significance of alpha a Likelihood Ratio chi-square test is used; if the test statistic assumes a high value, this suggest that the response variable is over-dispersed and that the application of a Negative Binomial Model is appropriate. As is clear from the tables of results, over-dispersion is evident for damages and fires, crimes against properties and for the total number of crimes. Conversely the parameter alpha does not result significant for violent crimes; for this reason, for this typology of crimes, it could reasonable to use a Poisson Model. Table 20 reports the results of the Poisson model run on the violent crimes data: the table records the Pearson dispersion statistic, defined as the ratio between the Pearson statistic and the degree of freedom (number of observation less the predictor, 75 in our example). If there is no over-dispersion in data the statistic assumes a value of one; on the contrary, value of approximately 6.25 identify over-dispersion (Hilbe, 2007). This statistic, computed for all the response variables, highlights the presence of over-dipersion in all the model, except for the model built on violent crimes, for which is therefore more appropriate a Poisson Model rather that a Negative Binomial model.

Response Variable	**Pearson**	**(1/df)Pearson**
CRIMES_TOT	822.14	11.72
CRIMES_DAMAGES	549.68	7.33
CRIMES_PREDATORY	468.07	6.24
CRIMES_VIOLENT	78.16	1.04

Table 20 Poisson Model – Violent crimes –

CRIMES_VIOLENT	Estimate	Std Error	t value	Pr(>|t|)
(Intercept)	-1.7930	0.4452	-4.0300	0.0000 ***
INTEGRATION	0.3586	0.2725	1.3200	0.1880
CHOICE	-0.1940	0.1194	-1.6200	0.1040
LINE LENGTH	0.0009	0.0003	3.1400	0.0020 **
RES_TOT	0.0003	0.0004	0.8700	0.3860
NUMBER OF SHOPS	0.0132	0.0076	1.7300	0.0840
HDI_LOW	-0.0004	0.0052	-0.0700	0.9410
PERC_FOREIGN	0.0063	3.5940	0.0000	0.9990
R^2	0.57			
Deviance = 72.2304	(1/df) Deviance = .9631			
Pearson = 78.1629	(1/df) Pearson = 1.0421			
AIC = 1.7704	BIC = -259.18			

Table 21 BN Model – Total crimes –.

CRIMES_TOT	Estimate	Std Error	t value	Pr(>|t|)	
(Intercept)	2.0080	0.2044	9.8200	0.0000	***
INTEGRATION	0.2965	0.1361	2.1800	0.0290	*
CHOICE	-0.1537	0.0792	-1.9400	0.0520	*
LINE LENGTH	0.0008	0.0002	4.5900	0.0000	***
RES_TOT	0.0002	0.0002	0.9000	0.3700	
NUMBER OF SHOPS	0.0093	0.0055	1.6800	0.0920	
HDI_LOW	0.0023	0.0034	0.6800	0.4940	
PERC_FOREIGN	-0.2915	1.4492	-0.2000	0.8410	
lnalpha	2.5810	0.2073			
alpha	13.2110	2.7390			
R^2	0.71				
LR chi2 = 67.93 Prob > chi2 = 0.0000					
Likelihood-ratio test of alpha=0: chibar2(01) = 603.28 Prob>=chibar2 = 0.000					

Even though R^2 measure is not appropriate in the context of count data model, we compute it in order to have a common measure of the goodness of fit of the models. The coefficient of determination decreases a little bit particularly for the crimes against properties; although this drop, all the model still preserve high level of goodness of fit.

The explicative capacity of the models is confirmed by the LR test statistic that tests the null hypothesis: “all regression coefficients in the model are simultaneous equal to zero”. It is calculated as negative two times the difference of the likelihood for the null model and the fitted model where the null model corresponds to the last iteration from fitting constant-only model. In order to test the null hypothesis, we observe the “Prob>Chi2” (In Table 17-20) which represents the probability of getting a LR test statistic as extreme or more extreme than the observed under the null hypothesis and we compare the *p*-value with the usual level of 0.05. As is evident from Table 17-20 all the models seem significant, in fact the *p*-value from the LR test are lower than 0.00001; as a result at least one of the regression coefficients in all the models is not equal to zero.

Even though this typology of model seems suitable for this data, it presents some limits: firstly it does not take into account a possible spatial autocorrelation in data, moreover it does not consider the fact that a lot of streets don't register any crimes. We will talk about the first limit in the conclusion section. For what concerns the second limit, a solution can be found, in the context of count data models, using a zero inflated model, which permits to take into account the massive presence of streets without any crime recorded.

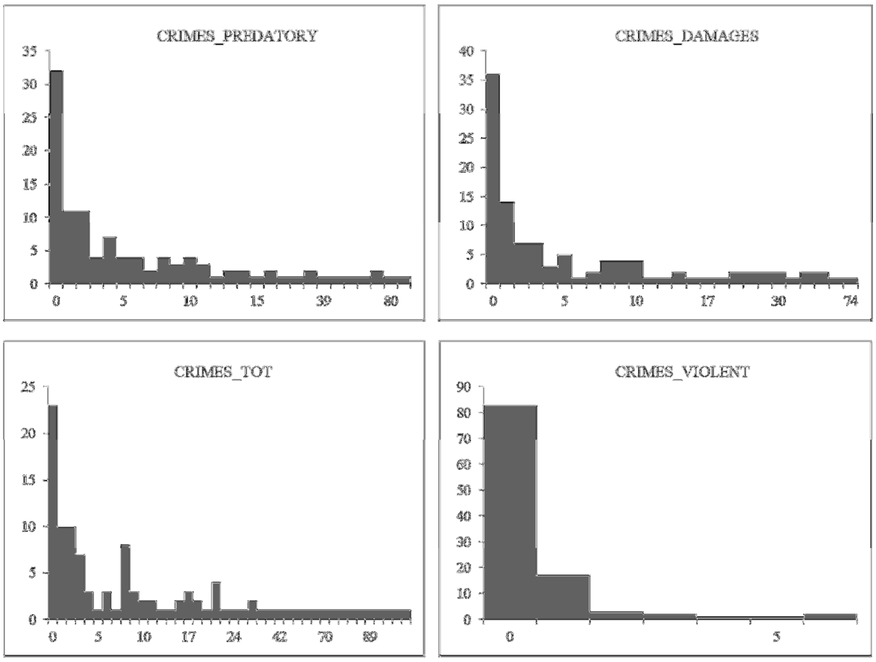
3.5 Zero Inflated Negative Binomial Model

As just mentioned this model has been implemented to take into account the fact that a lot of streets don't register any crimes. This models were introduced by Lambert (1992) to face the problem of data containing an excessive number of zero observations. Zero inflated models are used when a "zero" observation can arise from two different situations: the failing to observe an event or the inability to observe an event (Washington, Karlafftis, Mannering, 2003). The Zero Inflated Negative Binomial (ZINB) regression model follows the formula:

$$
\begin{cases}
y_i = 0 & p_i + (1-p_i)\left[\dfrac{1/\alpha}{\frac{1}{\alpha}+\lambda_i}\right]^{1/\alpha} \\
y_i = y & (1-p_i)\left[\dfrac{\Gamma\left(\frac{1}{\alpha}+y\right)u_i^{1/\alpha}(1-u_i)^{y}}{\Gamma\left(\frac{1}{\alpha}\right)y!}\right], y = 1,2,3\ldots
\end{cases}
\qquad (10)
$$

where $u_i = (1/\alpha)/[\,(1/\alpha) + \lambda_i]$. In order to estimate the parameters, the maximum likelihood methods is used.

Figure 1: Distribution of the number of crime complaints



As is evident from Figure 1, all the typologies of crime account for an excessive number of zero; this phenomenon is particularly clear for violent crimes. For this typology of crime, because of the conclusion on the over-dispersion parameter of the previous paragraph, we use a Zero Inflated Poisson Model rather than a Zero Inflated Negative Binomial Model.

In this context the zero inflation has been assumed to follow a logit process. As we can see from the tables of results (Table 22-25) all the regression are significant, as the *p*-value of the LR Chi2 test shows. Also in these models the R^2 statistics must be read with caution because they are not a proper measures of the goodness of fit in count models; however the LR test on the significance of the regression confirm that at least one of the coefficients is significantly different from zero, for all the models. In order to compare the ZINB to the standard NB model we can use the Vuong Test (Vuong, 1989). This test compares the zero-inflated negative binomial model to a standard negative binomial model. In our models the z-values are not significant, therefore the Vuong test shows that the zero-inflated negative binomial is not better than the standard negative binomial.

Table 22 ZIBN Model – Damages and fires–

CRIMES_DAMAGES	Estimate	Std Error	t value	Pr(>|t|)	
(Intercept)	1.4960	0.4829	3.1000	0.0020	**
INTEGRATION	0.4741	0.1542	3.0700	0.0020	**
CHOICE	-0.1757	0.1209	-1.4500	0.1460	

LINE LENGTH	0.0009	0.0003	3.0700	0.0020	**
RES_TOT	0.0002	0.0006	0.3900	0.7000	
NUMBER OF SHOPS	-0.0013	0.0081	-0.1600	0.8690	
HDI_LOW	0.0045	0.0067	0.6700	0.5040	
PERC_FOREIGN	-2.3940	4.1587	-0.5800	0.5650	
Inflate_const	-1.3375	0.4251	-3.1500	0.0020	
lnalpha	-0.5518	0.3589	-1.54	0.124	
alpha	0.5759	0.2067			
R^2	0.60				
LR chi2 =42.74 Prob > chi2 = 0.0000					
Vuong test of ZIBN vs NB: z = 0.73 Prob>z = 0.2338					

The values and the significance of the coefficients change, particularly for the model on the total number of crimes in which the variables Choice and Number of Shops lose their significance. The coefficient of determination significantly decreases in all the models.

Table 23 ZIBN Model – Crimes against properties –.

CRIMES_PREDATORY	Estimate	Std Error	t value	Pr(>ltl)	
(Intercept)	1.1480	0.3053	3.7600	0.0000	***
INTEGRATION	0.6233	0.1387	4.4900	0.0000	***
CHOICE	-0.1500	0.1021	-1.4700	0.1420	
LINE LENGTH	0.0007	0.0003	2.6500	0.0080	**
RES_TOT	0.0006	0.0005	1.2000	0.2310	
NUMBER OF SHOPS	0.0142	0.0072	1.9700	0.0490	*
HDI_LOW	-0.0004	0.0054	-0.0700	0.9410	
PERC_FOREIGN	-0.2266	2.4170	-0.0900	0.9250	
Inflate_const	-2.4332	0.8284	-2.9400	0.0030	

lnalpha	-0.7729	0.3305	-2.3400	0.0190
alpha	0.4617	0.1526		
R^2	0.58			
LR chi2 =78.88 Prob > chi2 = 0.0000				
Vuong test of ZIBN vs NB: z = 0.50 Pr>z = 0.3097				

Table 24 ZIP Model – Violent crimes –

CRIMES_VIOLENT	Estimate	Std Error	t value	Pr(>|t|)
(Intercept)	-1.6106	0.5029	-3.2000	0.0010 ***
INTEGRATION	0.3578	0.2818	1.2700	0.2040 ***
CHOICE	-0.0179	0.0933	-0.1900	0.8470
LINE LENGTH	0.0007	0.0003	2.1000	0.0360 **
RES_TOT	0.0004	0.0004	1.0200	0.3070
NUMBER OF SHOPS	0.0175	0.0074	2.3600	0.0180 *
HDI_LOW	0.0010	0.0050	0.2100	0.8370
PERC_FOREIGN	-1.9789	4.3849	-0.4500	0.6520
Inflate_const	-2.2713	1.0270	-2.2100	0.0270
R^2	0.27			
LR chi2 = 54.33 Prob > chi2 = 0.0000				
Vuong test of ZIP vs standard Poisson: z = 1.37 Pr>z = 0.0853				

Table 25 ZIBN Model – Total crimes –.

CRIMES_TOT	Estimate	Std Error	t value	Pr(>|t|)
(Intercept)	1.8046	0.3917	4.6100	0.0000 ***
INTEGRATION	0.4898	0.1295	3.7800	0.0000 ***
CHOICE	-0.1467	0.1290	-1.1400	0.2550
LINE LENGTH	0.0008	0.0003	2.6000	0.0090 **
RES_TOT	0.0006	0.0006	1.0000	0.3160

NUMBER OF SHOPS	0.0088	0.0085	1.0400	0.3000
HDI_LOW	0.0008	0.0067	0.1100	0.9090
PERC_FOREIGN	-0.8664	2.9720	-0.2900	0.7710
Inflate_const	-2.7850	1.0332	-2.7000	0.0070
lnalpha	-0.2851	0.2840	-1.0000	0.3160
alpha	0.7520	0.2136		
R^2	0.56			
LR chi2 =61.92 Prob > chi2 = 0.0000				
Vuong test of ZIBN vs NB: z = 0.37 Pr>z = 0.3539				

4. Models comparison

As is evident from Table 26, the linear regression and the spatial autoregressive models bring to similar results in terms of sign and significance of the coefficients and goodness of fit. The most significant variable is the length of a street: in fact it is obvious that longer streets register a higher number of crimes. The integration of a street is significantly and positively related to damages and fires and to crimes against properties, while the Choice is a deterrent for vandalism actions. The presence of shops tends to be positive related to crimes, especially for crimes against properties. The percentage of foreign people resident in a street does not seem to be related to crimes in this district: anyway we have decided to include this variable in the model because it could be highly significant in some other districts of Genoa. The goodness of fit is high reaching a value of 84% for crimes against properties in the linear and spatial autoregressive models.

The Negative Binomial and the Negative Binomial Zero Inflated bring to different results in terms of significance and R^2; some coefficients, especially the number of residents from countries of origin with low level of HDI is no longer significant; however all the models seem to confirm the general influence of the explicative variables on crime.

Table 26 summarizes the comparison between models: as is evident the sign of the coefficient doesn't change between the models, showing robustness. In addition the significance of the co-efficient is almost the same; this is a confirm of the effectiveness of the variables used.

Table 26: comparison between models

[redacted]

5. Conclusions and further work

These preliminary results should be read carefully for various reasons. Firstly this analysis focuses on a small district of Genoa: it will be interesting to extent these results to other city districts and to other cities with different spatial, economic and demographic features in order to validate the effectiveness of the variables in different contexts of application. Moreover some important variables have not been included in the models. Particularly, the income of the residents is, for the time being, not available; many researchers have pointed out a strict relationship between the number of crimes and the income or relative deprivation of people (Land, et al, 1990; Messner and Rosenfeld, 1999; Sampson, Morenoff, and Gannon-Rowley, 2002, Mears, Bhati, 2006). For this reason it could be relevant to include in the models a variable, such as the value of the houses, as a proxy of the income of the residents in an area.

Another problem to be faced is that crimes complaints recorded by police force are only a proxy of the actual number of crimes. The problem of crimes undetected can be solved collecting additional information throught appropriate surveys.

Moreover, the spatial dependence structure is still under focus: we expect that the weakness of spatial autocorrelation is due to the fact that a relevant part of it is implicitly specified inside the Space Syntax variables used in model and further work will be done in this direction. In order to verify this conjecture, future analyses will be focused on the construction of a model which takes into account all the three aspect of our study: counting dependent variable, zero inflation and spatially structured random effects. Particular attention is given to the Besag (1991) model; this Bayesian model is mainly used in epidemiological studies but it can be suitable also in crime analysis. The development of a Bayesian approach to model spatial data has significantly increased since 1990s with many applications in disease mapping (Mollie, 1996 ; Congdon 2002). Some authors have already used this approach in the analysis of crime (Law and Haining, 2004; Cohen 1998; Berry, Evett and Pinchin 1992). For this reason the next step of the analysis will focus on the construction of a Bayesian model to express the relationship between crimes and spatial and socio demographic variables using this innovative approach.

Acknowledgements

This work is one of the results of the Project “Territori Urbani e Politiche Pubbliche di Sicurezza” financed by the Municipality of Genoa – Direzione Città Sicura (Safe city Directorate)

References

Akaike, Hirotugu (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6): 716–723.

Baumer, E., Horney, J., Felson, R. & Lauritsen, J.L. (2003). Neighbourhood disadvantage and the nature of violence. Criminology, 41, 39-71.

Beavon, D. J.K., Brantingham P.L. and Brantingham P.J. (1994). The influence of street networks on the patterning of property offences. In R.V. Clarke (Ed.), *Crime Prevention Studies*, Vol. 2, New York: Willow Tree Press.

Berry, D.A., Evett, I.W. and Pinchin, R. (1992) Statistical inference in crime investigations using deoxyribonucleic acid profiling (with discussion). Applied Statistics, 41, 499–531.

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice System, Journal of the Royal Statistical Society, Series B, Vol 36 (2): 192-236.

Besag, J., York, J. and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics, Ann. Inst. Statist. Math., 43, 1-59.

Block, R. and Davis. S. (1996). The environs of rapid transit stations: A focus for street crime or just another risky place? In: R. Clarke (Ed.) *Preventing Mass Transit Crime*, Monsey, NY: Criminal Justice Press.

Boba, R. (2012). *Crime analysis with crime mapping*, third edition, SAGE Publications Inc.

Brantingham, P.L., and Brantingham, P.J.(1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. Journal of Environmental Psycology, 13, 3-28.

Chainey, S. & Ratcliffe, J. (2008), GIS and Crime Mapping, England, WILEY.

Clarke, R. V. (1983), Situational Crime Prevention: its theoretical basis and practical scope. In: M. Tonry & N. Morris (Eds), Crime and justice: an annual review of research (Vol 14, pp. 225-256).The Chicago University Press.

Clarke, R. V. (1992), Situational Crime Prevention: Successful case studies. Albany, NY: Harrow and Heston.

Clarke, R. V. (1995), Situational Crime Prevention. In: M. Tonry & D. Farrington (Eds), Building a safer society: Strategic approaches to crime prevention. Crime and Justice: A review of research (Vol.19, pp.91-150). The Chicago University Press.

Cohen, J, Nagin, D, Wallstrom, G, Wasserman, L. (1998). Hierarchical Bayesian analysis of arrest rates. *Journal of the American Statistical Association*, 93:1260-1270.

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

Cohen, L. & Felson, . (1979). "Social Change and Crime Rates" American Sociological Review 44: 588-08.

Congdon P. (2002) A model for mental health needs and resourcing in small geographic areas: a multivariate spatial perspective, *Geographical Analysis*, 34, 2, 168-186

Cressie, N.A.C.(1993), Statistics for Spatial Data, Rev Ed., Wiley.

Decker, S. (1996), Collective and normative features of gang violence. Justice Quarterly, 13, 243-264

di Bella, E., Corsi, M., Persico, L.(2011). A Multivariate Analysis Of The Space Syntax Output For The Definition Of Strata In Street Security Surveys, DISEFIN Series of economic working papers, 5.

Doreian, P. (1980), Linear models with spatially distributed data: Spatial disturbances or spatial effects. Sociological Methods and Research, 9, 29-60.

Dunn, C., (1980) Crime Area Research, in: Georges, D., and Harris, K., (eds.), Crime: A Spatial Perspective, New York, Columbia Press.

Eck, J. E., and Weisburd, D. (1995). Crime places in Crime Theory. In: J. E. Eck & D. Weisburd (Eds.), Crime and Place. Monsey, NY: Criminal Justice Press.

Eisner, M., Wikstrom, P.O., (1999), Violent crime in the urban community: a comparison of Stockholm and Basel. European Journal on Criminal Policy and Research. Vol. 7, No 4,p. 427-442.

Emig, M., Heck, R., & Kravitz, M. (1980).Cryme analysis: A selected bibliography.Washington, DC.National Criminal Justice Reference Service.

Felson, M. and Clarke, R.V. (1998), Opportunity makes the thief: Practical theory for crime prevention. Police Research Series Paper 98. London: Police and Reducing Crime Unit; Research, Development and Statistics Directorate. Journal of Studies on Alcohol.

Gorman, D.M., Speer, P.W., Gruenewald, P.J., Labouvie, E.W. (2001), Spatial dynamics of alcohol availability, neighborhood structure and violent crime, Volume 62, Issue 5, Pages 628-636

Gottlieb, S., Arenberg,s., & Singh, R.(1994).Crime analysis. From first report to final arrest. Montclair, CA: Alpha Publishing

Griffiths, E. & Chavez, J. M. (2004), Communities, street guns and homicide trajectories in Chicago, 1980-1995: Merging methods for examinig homicide trends across space and time. Criminology, 42, 941-975.

Guerry, A. M. (1833), Essai sur la statistique morale de la France, Paris, Cochard.

Hilbe J. M. (2007), Negative Binomial Regression, Cambridge University Press.

Hillier, B. and Hanson, J. (1984). The Social Logic of Space. Cambridge University Press.

Hillier, B., (1988), Against Enclosure , N. Teymur, T.A. Markus, T. Wooley (Eds.), Rehumanizing Housing, Butterworths, London, 63-88.

Hillier, B. and Shu, S., (2000) Crime and Urban Layout: The Need for Evidence, in: Ballintyne, S., Pease, K. and McLaren, V., Secure Foundations: Key Issues in Crime Prevention, Crime Reduction and Community Safety, London, IPPR.

Hillier, B., and O. Sahbaz. (2008). An evidence based approach to crime and urban design. In Urban Sustainability for the 24 Hour City, Edited by R Cooper, C Boyko, G Evans and M Adams. London: Routledge.

Jacobs J. (1961), The Death and Life of Great American Cities, Vintage, New York

Jeffery, C. (1977), Crime Prevention Through Environmental Design. Beverly Hills, CA, Sage Publishing.

Lambert, D. (1992), Zero inflated Poisson Regression with an application to defects in manifacturing, Technometrics 34, 1-14.

Land, Kenneth C., Patricia L. McCall, and Larry E. Cohen. (1990).Structural covariates of homicide rates: Are there invariances across time and social space. American Journal of Sociology 95(4): 922–63.

Law, J., and R. P. Haining (2004). A Bayesian Approach to Modelling Binary Data: The case of High-Intensity Crime Areas. Geographical Analysis 36:3 197-216.

Lichstein, J.W., Simons, T.R., Shriner, S.A. & Franzreb, K.E. (2002) Spatial autocorrelation and autoregressive models in ecology. Ecological Monographs, 72, 445–463.

Mears, D.P. , Bhati, A.S. (2006) , No community is an Island: The effects of resource deprivation on urban violence in spatially and socially proximate communities, Criminology, Volume 44, Issue 3, August 2006, Pages 509-548

Meneroff, J.D., Sampson, R.J., & Raudenbush, S. (2001). Neighnorhood inequality, collective efficacy, and the spatial dynamics of urban violence. Criminology, 39, 517-560.

Messner, S. F., Rosenfeld R. (1999). Social structure and homicide: Theory and research. In Homicide: A Sourcebook of Social Research, eds. M. Dwayne Smith and Margaret A. Zahn. Thousand Oaks, CA: Sage Publications.

Mollie, A. (1996). Bayesian mapping of disease. In: Gilks W.R., Richardson S., and Spiegelhalter D.J. (Eds.), *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London, pp. 359-379.

Moran, P.A.P. (1950): Notes on Continuous Stochastic Phenomena, Biometrika 37: 17-33.

McClintock, F.H., and P.O. Wikstrom (1992), The comparative study of urban violence. Criminal Violence in Edinburgh and Stockholm. British Journal of Criminology, 32, 505-520.

Murray, A.T., McGuffog, I., Western, J.S., Mullins, P. (2001), Exploratory spatial data analysis techniques for examining urban crime, British Journal of Criminology, Volume 41, Issue 2, March 2001, Pages 309-329.

Nubani, L., Wineman, J.(2005), The Role of Space Syntax in Identifying the Relationship Between Space and Crime, A. van Nes (Ed.), Proceedings, 5th International Space Syntax Symposium, Delft, 413-422.

Newman O. (1972), Defensible Space: Crime Prevention Trough Urban Design, New York, MACMILLAN.

Paulsen, D. J. and Robinson, M. B. (2009), Crime mapping and Spatial Aspects of Crime, PEARSON.

Quetelet, A. (1984), Research on the propensity for crime at different ages, translated and with and indtrodution by S.F. Sylvester. Cincinnati: Anderson Publishing.

Rengert, G.,(1980) Spatial Aspects of Criminal Behavior, in: Georges, D., and Harris, K., (eds.), Crime: A Spatial Perspective, New York, Columbia Press.

Rice, K.J. and Smith, W.R. (2002), Socioeconomic Models of Automotive Theft: Integrating Routine Activity and Social Disorganization Approaches, *Journal of Research in Crime and Delinquency*, 39: 304--336.

Roncek, D.W., (1981). Dangerous places: Crime and residential environment. *Social Forces,* 60, 74-96.

Sampson, Robert J., Jeffrey D. Morenoff, and Thomas Gannon-Rowley. (2002). Assessing “neighborhood effects”: Social processes and new directions in research. Annual Review of Sociology 28:443–78.

Sampson, R. J. (2004), Networks and neighbourhoods: The implications of connectivity for thinking about crime in the modern city. In: H. McCarthy, P. Miller & P. Skidmore (Eds), Network Logic: Who Governs in an Interconnected World?(pp. 157-166). London: Demos.

Shaw, C. & H. McKay. (1942). Juvenile Delinquency and Urban Areas. Chicago: Univ. Press.

Smith, M. and Clarke, R.V. (2000) Crime and Public Transport. Crime and Justice, Vol. 27, 169-233.

Shu, S., (2000) Housing Layout and Crime Vulnerability, Urban Design International, (5.3-4), p.177-188.

Tita G. E. and Greenbaum, R.T. (2009), Crime, Neighborhoods, and Units of Analysis: Putting Space in Its Place, D. Weisburd *et al.* (eds.), Putting Crime in its Place, Springer.

Thomas, W. I. (1966), Social disorganization and social reorganization. In: M. Janovitz (Ed.), On social organization and social personality: Selected papers (pp.3-11). The University of Chicago Press.

Tobler, W. R., (1970), A computer movie simulating urban growth in the Detroit region, Economic Geography, 46, 234-240.

Vellani, K. and Nahoun, J. (2001) Applied Crime Analysis, Paperback Saint Louis, Missouri, U.s.a.: Butterworth-heinemann.

Vuong, Q. (1989), Likelihood ratio tests for model selection and non-nested hypotheses, Econometrica, 57, 307-334.

Washington, S., Karlaftis, M., Mannering, F. (2003), Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC, Boca Raton.

Weisburd, D., Bernasco, W., Bruinsma, G.J.N. (2009). Putting Crime in its Place, Units of Analysis in Geographical Criminology, Springer.

Weisburd, D. and McEwen, T. (2010), Crime Mapping and Crime Prevention, U.S.A., LYNNE RIENNER.

Wilsem, van J. (2009), Urban Streets as Micro Context to Commit Violence, D. Weisburd *et al.* (eds.), Putting Crime in its Place, Springer.

Working Papers recently published

(The complete list of working papers can be found at http.//www.disefin.unige.it)

n.3/2012 Rosella Levaggi, Francesco Menoncin, "Paternalistic Goods to Improve Income Distribution: a Political Economy Approach", April 2012

n.2/2012 Gabriele Cardullo, "Public Sector Wage Bargaining, Un-employment, and Inequality ", February 2012

n.1/2012 Amedeo Fossati, "The Italian Tradition in Public Finance: an Annotated Bibliografy of Mauro Fasiani", January 2012

n.7/2011 Amedeo Fossati, Marcello Montefiori, "Adverse Selection in Elderly Care", November 2011

n.6/2011 Paolo Cremonesi, Enrico di Bella, Marcello Montefiori, Luca Persico, "A self-reported work sampling to assess the Emergency Department's costs", October 2011

n.5/2011 Enrico di Bella, Luca Persico, Matteo Corsi, "A Multi-vari-ate Analysis Of The Space Syntax Output For The Defi-nition Of Strata In Street Security Surveys", September 2011n.4/2011 E. Briata, "Marginal tax rates, tax revenues and inequality. Reagan's fiscal policy", July 2011

n.4/2011 Elena Briata, "Marginal tax rates, tax revenues and ine-qual-ity. Reagan's fiscal policy", July 2011n.3/2011 France-sco Copello, Cristiana Pellicanò, "Esemplificazione della Da-ta Envelopment Analysis per la valutazione di efficienza in una grande azienda ospedaliera universitaria"

n.2/2011 Stefano Capri, Rosella Levaggi, "Shifting the risk in pric-ing and reimbursement schemes? A model of risk-sharing agreements for innovative drugs"

n.1/2011 Cinzia Di Novi, "The Indirect Effect of Fine Particulate Matter on Health through Individuals' Life-style" n.4/2010 Angelo Baglioni, Andrea Monticini, "Why does the Interest Rate Decline Over the Day? Evidence from the Liquidity Crisis"

n.3/2010 Amedeo Fossati: "The double taxation of savings: the Italian debate revisited"

n.2/2010 Andrea Monticini, David Peel, Giacomo Vaciago: "The impact of ECB and FED announcements on the Euro Interest Rates"

n.1/2010 Amedeo Fossati: "Vilfredo Pareto and the methodology of the Italian tradition in public finance"