Nereus
Núcleo de Economia Regional e Urbana da Universidade de São Paulo
The University of São Paulo Regional and Urban Economics Lab

THE SPATIAL PATTERN OF CRIME IN MINAS GERAIS: AN EXPLORATORY ANALYSIS

Eduardo S. Almeida
Eduardo A. Haddad
Geoffrey J. D. Hewings

TD Nereus 22-2003

São Paulo
2003

THE SPATIAL PATTERN OF CRIME IN MINAS GERAIS:

AN EXPLORATORY ANALYSIS

Eduardo Simões de Almeida

Dept. of Economics, University of São Paulo

Eduardo Amaral Haddad

Dept. of Economics, University of São Paulo and Regional Economics Applications Laboratory (REAL), University of Illinois

Geoffrey John Dennis Hewings

Regional Economics Applications Laboratory (REAL), University of Illinois

ABSTRACT

This paper is aimed at examining spatial pattern of crime in the state of Minas Gerais, Brazil. In methodological terms, this paper uses exploratory spatial data analysis (ESDA) to study the distribution of crime rates in more than 750 municipalities of this state for 1995. The findings reveal that crime rates are distributed non-randomly, suggesting positive spatial autocorrelation. Moreover, both global and local outliers across space are detected. Further, spatial heterogeneity represented by "hot spots" (high-high spatial association) and "cool spots" (low-low spatial association), as well as some clusters with negative spatial association (high-low and low-high) are identified. Further, crime data also have spatial trends. The pattern of spatial distribution revealed through ESDA provides an empirical foundation for the further econometric specification of multivariate models.

KEY WORDS: exploratory spatial data analysis (ESDA); crime rates; spatial regimes; spatial autocorrelation.

1 INTRODUCTION

Spatial econometrics is an emerging field in quantitative methods applied for regional science. The difference of spatial econometrics from the standard econometrics refers to characteristics of the socio-economic interaction among agents in a system and the structure of this system across space. These interactions and structures generate spatial effects in various socio-economic processes. Spatial effects are made up of **spatial heterogeneity** and the **spatial dependence**. The spatial autocorrelation refers to socio-economic interaction among agents, whereas spatial heterogeneity regards to aspects of the socio-economic structure over space (Anselin, 1988; Anselin e Bera, 1998).

In space the interaction has a multidirectional nature, generating spatial effects that violate a vital assumption of the classic linear regression model, to wit, the spherical errors assumption. Furthermore, since the heteroskedasticity is resilient to several standard procedures to correct it, it is very likely that its source comes from intricate relationship to the spatial autocorrelation. In spatial processes, it is common that heteroskedasticity generates spatial dependence and, in turn, spatial dependence also induces heteroskedasticity (Anselin e Bera, 1998).

These characteristics provoke serious difficulties for identifying proper spatial models. Consequently, the identification task may become very time demanding and cumbersome; or worst, lead to estimate wrong spatial models.

An appropriate exploratory spatial data analysis (ESDA) can help to overcome such an identification problem, furnishing clear guesses and indications about the existence of spatial regimes, preliminary spatial autocorrelation, potential regressors, spatial trends, the influence of local outliers, spatial clusters ("hot spots" and "cool spots"), etc. Hence, some ESDA work precedes a good spatial econometric modeling.

ESDA is a collection of techniques for the statistical analysis of geographic information, intended to discover spatial patterns in data and to suggest hypotheses, by imposing as little prior structure as possible. The reason for this approach stems from the drawbacks of the conventional methods such as visual inspection and standard multivariate regression analysis that *"are potentially flawed and may therefore suggest spurious relationships*". By the same token, "*human perception is not sufficiently rigorous to assess 'significant' clusters and indeed tends to be biased toward finding patterns, even in spatially random data*" (Messner *et al*, 1999, pp. 426-427). Accordingly, it is necessary to use formal tests and quantitative tools to analyze these spatial patterns to avoid misinterpretations.

ESDA, like its forerunner Exploratory Data Analysis introduced by Tukey (1977), is not aimed at testing theories or hypothesis, hence it "*may be considered as data-driven analysis*" (Anselin, 1996, p. 113). One of its roles is to shed light on future possibilities in modeling and theorizing; the primary objective is let the data speak for themselves.

This paper illustrates the ESDA approach, applying for Brazilian crime data. Nowadays crime is one of most relevant socio-economic phenomena over the world. Crime imposes immense social costs, representing pernicious effects on economic activity and quality of life. In the United States, crime costs represent more than 5 percent of the US gross domestic product (GDP). Similar estimations point out that crime-related cost in Latin America is also around 5 percent of GDP. In Mexico, the social losses related to crimes amount about 5 percent, whereas in El Salvador and Colombia these costs are about 9 percent and a little more than 11 percent, respectively. In Brazil, crime costs around 3 percent of GDP. (Fajnzylber *et al*., 2000, pp. 223-224).

However, crime is not a random activity: prior research has suggested significant spatial and temporal concentrations. A national or state average may be misleading since it reveals nothing about variability within the country or state. If crime is indeed spatially concentrated, then analysis of patterns and causes will require a different approach than in the case where it is most evenly distributed. This paper will explore where crime happens; from a theoretical standpoint, "*place might be a factor in crime, either by influencing or shaping the types and levels of criminal behavior by the people who frequent an area, or by attracting to an area people who already share similar criminal inclinations*" (Anselin *et al.*, 2000, p. 215). To this end, the exploration of the patterns of crime in Minas Gerais will consider spatial interaction among the locations to understand their heterogeneity and dependence.

In the literature, there are some studies relating, explicit or implicitly, space and crime. We begin here with a very brief overview of the literature on the role of the geographic space in the study of crime.[1] Place-based theories of crime seek to explain the relationship between place and crime. According to Anselin *et al.* (2000, p. 216), "*routine activities that bring together potential offenders and criminal opportunities are specially effective in explaining the role of place in encouraging or inhibiting crime. The resulting crime locales often take the form of facilities – places that people frequent for a specific purpose – that are attractive to offenders or conducive to offending. Facilities might*

[1] For a more detailed review of literature on space and crime, see Messner *et al*. (2000), Messner and Anselin (2001) and Anselin *et al.* (2000).

provide an abundance of criminal opportunities (...). Or they might be the sites of licit behaviors that are associated with increased risk of crime (...)".

Ecological theories seek, in turn, to explain variations in crime rates through the differing incentives, pressures and deterrents that individuals face in different environments (different locations). The most famous ecological theory is the economic theory of crime developed by Becker (1968) and Ehrlich (1973). This theory points out that "*individuals allocate time between market and criminal activity by comparing the expected return from each, and taking account of the likelihood and severity of punishment*" (Kelly, 2000, p. 530). Ecological theory highlights the economic factors within an individual cost-benefit analysis of the criminal activity.

Glaezer and Sacerdote (1996) furnish another theory of crime, linked implicitly to space. They investigate why crime rates are much larger in large cities than in small cities and rural areas. Their findings indicate that city size and urbanization rates are important variables to consider in crime studies.

Using an ESDA approach, similar to the one developed in this paper, we found studies analyzing homicide rates in Saint Louis metropolitan area (Messner *et al.*, 1999; Messner and Anselin, 2001). The authors found that there is the presence of potential diffusion processes in criminal activity.

In the Brazilian literature, there is no study investigating the spatial patterns of crime, adopting this set of spatial statistical tools. Consequently, this paper is pioneer in doing this kind of investigation, using a vast collection of exploratory spatial statistics methods in order to extract information from Brazilian crime data.

The rest of the paper is divided as follows. Section 2 discusses the data used in the spatial analysis. Section 3 presents the results of the application of the ESDA approach. The conclusions and final remarks are shown in section 4.

2 THE DATA

The crime data for this paper come from the Secretaria de Segurança Pública do Estado de Minas Gerais (Public Security Secretariat of the State of Minas Gerais). The data consist of the distribution of crime rates in 754 municipalities of the state of Minas Gerais for 1995.[2] The crime rate used here is aggregated by municipality of residence (areal unit) and expressed as a rate of homicides and homicide attempt per 100,000 resident population.

In the light of this, the selection of the areal unit is a sensitive problem in spatial analysis. In other words, the choice of the spatial aggregation of the data influences the stability of parameter estimates obtained from a multivariate regression. In the literature, this problem is referred as modifiable areal unit problem (MAUP). According to Fotheringham *et al*. (2000, p. 237), "*one problem that has long been identified in the analysis of spatially aggregated data is that for some analyses the results depend on the definition of the areal units for which data are reported*".

As the crime data come from an official source, there is potentially a problem of underreporting. That is to say, only a fraction of all crimes makes its way to official statistics. However, the use of homicides and homicide attempts as a crime rate shows the property of precision: underreporting is low for this kind of data, unlike crime as theft or

[2] Minas Gerais is an important Brazilian state: it has Brazil's third largest GDP, besides being the country's second largest population. Moreover, in geographic dimension, Minas Gerais is Brazil's fourth largest state.

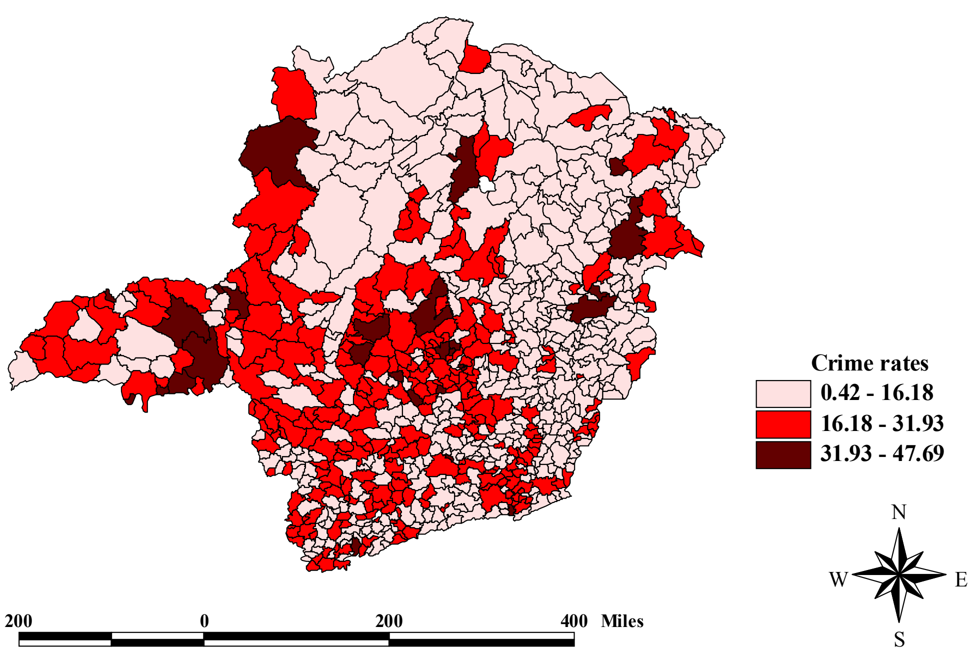
rape (Fajnzylber *et al.*, p 235). Besides, the incidence of homicide is considered as a proxy for crime rate in most studies.

3. EMPIRICAL RESULTS OF ESDA [3]

3.1. Distribution of Crime Rates

We begin the analysis with the choropleth map of the data crime. Map 1 shows the data for 1995. The spatial pattern of the crime rates is illustrated in this map, with the darkest shade corresponding to the highest rate range. The suggestion of spatial clustering of similar values that follows from the visual inspection of this map needs to be confirmed by formal tests, which are presented in Section 3.3.

Map 1. Crime Rates in Minas Gerais in 1995



[3] Most results of this section were obtained through SpaceStat™ extension for ArcView™ (see Anselin, 1999b). Other results were generated in the ArcGIS™ and in the CrimeStat (see Levine, 1999).

In order to describe the distribution of crime rates, we present summary statistics in Table 1.

Table 1. Summary statistics for crime rates in Minas Gerais

Statistics	*Crime rate in 1995*
Minimum	0.42
Maximum	47.69
Mean	15.45
Std Deviation	7.91
Skewness	0.89
Kurtosis	3.79
1-st quartile	9.63
Median	13.98
3-rd quartile	19.83

3.2. Tests for global spatial autocorrelation

The first step in a study of ESDA is to test this hypothesis: are the spatial data randomly distributed? To do that, it is necessary to use global autocorrelation statistics. There are many possible spatial weights matrices, depending on the choice of the nonzero elements for pairs of correlated observations. The value of a global autocorrelation statistics depends on the elements of the spatial weights matrix. For the analysis of the crime rates in Minas Gerais, three different types of spatial weights matrices will be constructed: rook, queen and inverse distance.

The spatial correlation coefficient Moran's I was proposed in 1948.[4] The underlying hypothesis is spatial randomness, that is, there is the absence of spatial

[4] Formally, this statistics is given by:

dependence in the data. Intuitively, spatial randomness can be expressed as follows: values of an attribute at a location do not depend on values of an attribute at neighboring locations.

Moran's I has an expected value of $-[1/(n-1)]$, that is, the value that would be obtained if there was no spatial pattern to the data.[5] The calculated value of I should be equal to this expectation, within the limits of statistical significance, if the y_i is independent of the values of y_j, $j \in J$ (and J is the set of neighboring locations). Values of I that exceed $-[1/(n-1)]$ indicate positive spatial autocorrelation. Values of I below the expectation indicate negative spatial autocorrelation. As the number of locations increases, this expectation approaches zero, which is the expectation for an ordinary correlation coefficient.

Table 2 reports the global Moran's I statistics for all municipalities of Minas Gerais in 1995. We used the criterion of binary neighborhood, namely, if two locations are neighbors (that is, they have a boundary in common of non-zero length), a value of 1 is assigned; otherwise, a value of zero is assigned. There are two conventions used in the construction of a binary spatial weights matrix: rook and queen conventions. In the rook convention, only common boundaries are considered in the computation of spatial weights matrix, while, in the queen convention, both common boundaries and common nodes are considered.[6]

$$I = \frac{n}{\sum\sum w_{ij}} \frac{\sum\sum w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum (y_i - \bar{y})^2}$$

where n is the number of locations, y_i is the data value of attribute in analysis (in our case, crime rate), w_{ij} is a spatial weight for the pair of locations i and j.

[5] It is noteworthy perceiving that most correlation measures have an expected value of zero.

[6] For more details on the concept of binary neighborhood, see Anselin (1988).

The statistical evidence in Table 2 casts doubt on the assumption of spatial randomness of the crime data for Minas Gerais. In fact, we can reject the hypothesis of no spatial autocorrelation at 0.1% significance level for 1995. These results are invariant with regards to convention of binary neighborhood used for the construction of the spatial weights (queen or rook). In addition, Moran's *I* provides clear indication that the spatial autocorrelation for crime rate in Minas Gerais is positive.

Table 2 – Global Moran's *I* Statistics for Crime Rates in Minas Gerais, using Binary Spatial Weight Matrix

Year	*Convention*	*I statistic*	*Probability level**
1995	Queen	0.3023	0.001
1995	Rook	0.3034	0.001

Note: *Empirical pseudo-significance based on 999 random permutations.

Table 3 reports the global Moran's *I* measure, using inverse distance convention through the formulas.[7]

[7] The inverse distance weights between two neighboring locations w_{ij} is given by the reciprocal of the distance of the two locations:

$$w_{ij} = \frac{1}{d_{ij}}$$

where d_{ij} is the distance between locations i and j.

Table 3 - Global Moran's *I* Statistics for Crime Rates in Minas Gerais, using Inverse Distance as Spatial Weights Matrix

Year	*Convention*	*I statistics*	*Randomization Significance (Z)*	*Normality Significance (Z)*
1995	Inverse distance	0.1260	14.10	14.09

The *I* value for crime rate in 1995, in turn, was about 0.13 with a normal *Z* value of 14.10 and a randomization-based *Z* value of 14.09, both of which are highly significant. Since the computed value of *I* exceeds its theoretical value, there is again evidence of positive spatial autocorrelation. That is, municipalities with a high crime rate are also adjacent to municipalities with a high crime rate. In an analogous manner, municipalities with a low crime rate are adjacent to municipalities with a low crime rate as well. That is the intuitive meaning of positive spatial autocorrelation.

Another global autocorrelation statistics was proposed by Geary in 1954, representing an alternative statistic for spatial autocorrelation.[8] A different measure of covariation is chosen, namely, the sum of squared differences between pairs of data values of the attribute in study. Again the underlying assumption is spatial randomness. Note that both *I* and *C* take on the classic form of any autocorrelation coefficient: the

[8] The formula of this statistics is given by:

$$C = \frac{n-1}{2\sum\sum w_{ij}} \frac{\sum\sum w_{ij}(y_i - y_j)^2}{\sum(y_i - \bar{y})^2}$$

numerator term of each is a measure of covariance among the y_is, whereas the denominator term is a measure of variance (Cliff and Ord, 1981).

C is asymptotically normally distributed as *n* increases; its value ranges between 0 and 2. The expected value of *C* (theoretical value) is 1. Values less than 1 (i.e., between 0 and 1) indicate positive spatial autocorrelation, while values greater than 1 (i.e., between 1 and 2) indicate negative spatial autocorrelation. Table 4 below presents the results for *C* statistics:

Table 4 – Global Geary "*C*" Statistics for crime rates in Minas Gerais

Year	*Criterion*	*C statistics*	*Normality Significance (Z)*
1995	Inverse distance	0.78	8.24

For crime rates in 1995, the *C* value for crime rate was 0.78 with a *Z* value of 8.24, thereby indicating positive spatial autocorrelation as well.

We corroborate the spatial non-randomness hypothesis for crime rates in Minas Gerais, using both Moran's *I* and Geary *C*. As a matter of fact, both indicate that crime data are not randomly distributed across space; indeed, there is highly significant positive spatial autocorrelation.

An alternative approach to visualize spatial association is based on the concept of a Moran Scatterplot, which shows the spatial lag (i.e. the average of the attribute for the neighbors) on the vertical axis and the value at each location on the horizontal axis. According to Anselin (1999a, p. 261), "*when the variables are expressed in standardized form (i.e. with mean zero and standard deviation equal to one), this allows for an* *assessment of both global spatial association (the slope of the line) as well as local spatial association (local trends in the scatterplot)*".

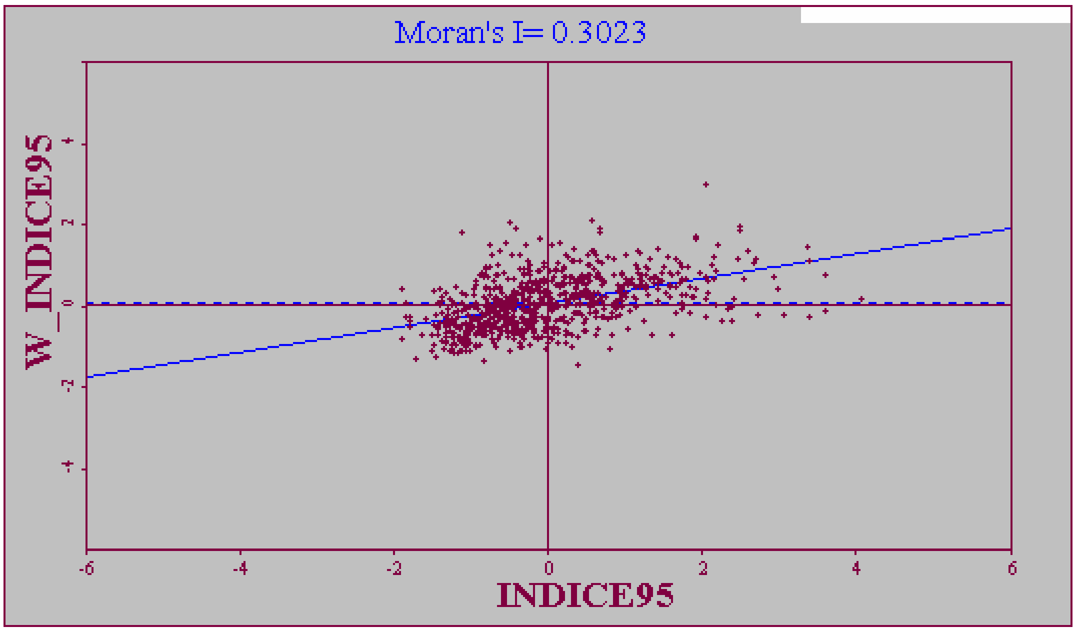
Thus, Moran's I provides a formal indication of the degree of linear association between a vector of observed values $\boldsymbol{y}$ and a weighted average of the neighboring values, or spatial lag, $\boldsymbol{Wy}$. When the spatial weights matrix is row-standardized such that the elements in each row sum to 1, the Moran's I is interpreted as a coefficient in a regression of $\boldsymbol{Wy}$ on $\boldsymbol{y}$ (but not of $\boldsymbol{y}$ on $\boldsymbol{Wy}$). As the slope is positive in the Moran Scatterplot (see Figure 1), once again we corroborate, diagrammatically, the existence of positive global spatial association.

While the overall tendency depicted in the Moran Scatterplot is one of positive spatial association, there are many municipalities that show the opposite, that is, low values surrounded by high values (Low-High negative association), portrayed in the upper left quadrant. In addition, there are many municipalities that represent high values surrounded by low values (High-Low negative association), portrayed in the lower right quadrant.

The indication of global patterns of spatial association may correspond to the local analysis, although this is not necessarily the case. In fact, there are two cases. The first case occurs when no global autocorrelation hides several significant local clusters. The opposite case is when "*a strong and significant indication of global spatial association may hide totally random subsets, particularly in large dataset*" (Anselin, 1995, p. 97).

The global indicators of spatial association are not capable of identifying local patterns of spatial association, such as local spatial clusters or local outliers in data that are statistically significant. To overcome this obstacle, it is necessary to implement a spatial clustering analysis.

Figure 1. Moran scatterplot for crime in 1995



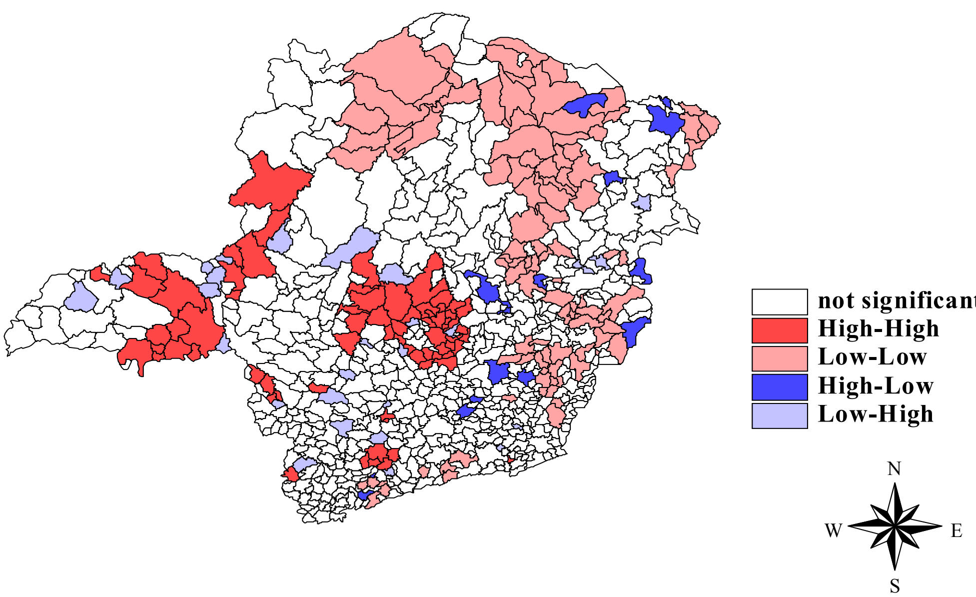
3.3. Spatial Clustering Analysis

Anselin (1995) suggested a new kind of indicator for capturing spatial clusters, known as a local indicator of spatial association (LISA). The intuitive interpretation is that LISA provides for each observation an indication of the extent of significant spatial clustering of similar values around that observation.

LISA (like local Moran) can be used as the basis for testing the null hypothesis of local randomness, that is, no local spatial association (Anselin, 1995, p. 95). LISA statistics have two basic functions. First, it is relevant for the identification of significant local spatial clusters; second, it is important as a diagnostic of local instability (spatial outliers) in measures of global spatial association (Anselin, 1995, p. 102). Map 2 shows the significance of the local Moran statistics. [9]

There are various LISA statistics in the spatial analysis literature.[10] We adopted here the local version of Moran's *I*, because it allows for the decomposition of the pattern of spatial association into four categories, corresponding to the four quadrants in the Moran Scatterplot (see Figure 1). Map 2 combines the information of the Moran scatterplot and the LISA statistics. It illustrates the classification into four categories of spatial association that are statistically significant in terms of the LISA concept.

Map 2. Moran Significance Map for Crime Rates in Minas Gerais



[9] Following Anselin (1995), local Moran statistic for an observation *i* can be stated as

$$I_i = z_i \sum_j w_{ij} z_j$$

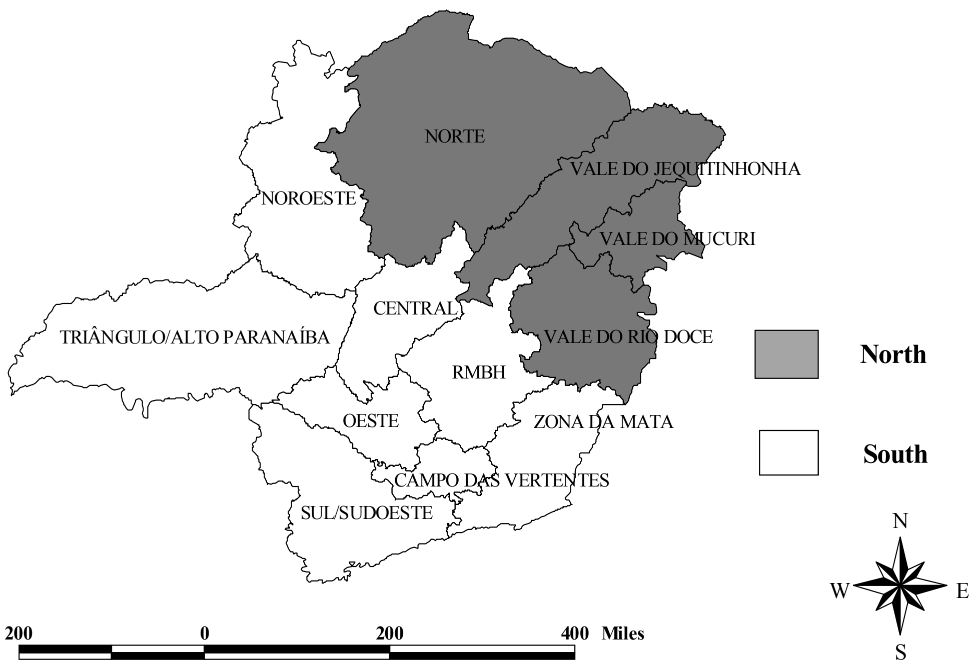
where the observation z_i, z_j are in deviation from the mean, and the summation over *j* is such that only neighboring values $i \in J_i$ are included, where J_i is set of neighbors of *i*.

[10] For more details on examples of LISAs, see Anselin (1995).

Beginning with the Moran Significance Map for crime (Map 2), we find evidence of spatial grouping. Overall, there are some clusters of municipalities with high crime rates, as well as neighbors with high crime rates in the Triângulo/Alto Paranaíba and the Central region (mainly in the Metropolitan Area of Belo Horizonte), besides the South-Southwestern Minas Gerais and the Northwestern Minas. There are also some municipalities in these regions that are LH: municipalities with a low crime rate surrounded by municipalities with a high crime rate. In general terms, it seems that there are groupings of crime around the larger cities of Minas Gerais (or population agglomerations with a high urbanization rate). By contrast, most clusters of municipalities with low crime (LL) are located in Northern Minas and Eastern Minas. In these regions, it is possible to observe some clusters of municipalities HL: municipalities with high crime rates surrounded by municipalities with low crime rates (see Map 3).

Roughly speaking, if we divide Minas Gerais into two parts ("North" and "South"), we can observe that most "hot spots" (HH) lie in the Southern part, while most "cool spots" (LL) are in the Northern part. Let us define two spatial regimes, "North" and "South", for crime rates in Minas Gerais (see Map 3). In this division, "South" is made up of the following planning regions: Triângulo/Alto Paranaíba, Central, Região Metropolitana de Belo Horizonte (RMBH), Oeste, Sul-Sudoeste, Campo das Vertentes, Noroeste e Zona da Mata. "North" is compounded of the following planning regions: Norte, Vale do Jequitinhonha, Vale do Mucuri and Vale do Rio Doce.

Map 3. The Planning Regions in Minas Gerais



To find more evidences about spatial regimes, we implement the spatial ANOVA to test difference in means of crime rates between the regions "North" and "South". The basic regression consists of a dummy variable regression of crime rates on a constant term and the geographical treatment indicator as follows:

$$y_i = \alpha + \delta REG + \varepsilon$$

where α is the overall mean of the regression, δ is a parameter to be estimated and *REG* is a dummy (geographical treatment indicator), which takes on a value 1 for "South" and 0 for "North"; ε is an error term.

Table 6. OLS regression of Spatial ANOVA

Independent variables	
Constant	11,56
	(21,44)***
REG	5,26
	(8,38)***
Adj. R-squared	0,084

Note: t-ratios in parentheses. * p<=0.1; ** p<=0.05 ; ***p<=0.01.

Table 6 shows the regression results. The positive and highly significant value at one percent for the coefficient REG indicates that there is a considerable discrepancy between the mean of crime in the "South" and the overall mean represented by the constant, which is also highly significant. As the signal of the categorical variable is positive, this indicates that the crime in "South" is higher in about 5 points than the overall mean.[11]

[11] It is worth noting that the adjusted R^2 is low, because no explanatory variable is included into this regression.

Table 7. Diagnostics for Spatial ANOVA Regression

	Model	
TEST	*VALUE*	*PROB.*
1. MULTICOLLINEARITY		
Condition number	3.64	-
2. NORMALITY		
Jarque-Bera	182,28	0,00
3. HETEROSKEDASTICITY		
Koenker-Basset	1,73	0,19
4. SPATIAL DEPENDENCE		
Moran's I	10,49	0,00
Lagrange Multiplier (error)	106,27	0,00
Robust LM (error)	0,40	0,53
Lagrande Multiplier (lag)	109,71	0,00
Robust LM (lag)	3,84	0,05

Table 7 reports the diagnostics of the regression. The diagnostics reveal that there is no multicollinearity problem, but the errors are non-normal, as detected by the Jarque-Bera test. Once the errors are not normal, the appropriate test presented is the Koenker-Bassett test that signals no heteroskedasticity problem. This is an important result, because the assumption in ANOVA is that the variances in the sub-groups are constant. According to the Koenker-Bassett test, we cannot reject the null on homoskedasticity. Besides, we can identify evidence of spatial autocorrelation by means of several tests. The high significance of LM_{LAG} and robust LM_{LAG} tests signal that a spatial lag must be included into the model. Hence, we estimate a spatial lag ANOVA regression, defined as follows:

$$y_i = \alpha + \rho W y_i + \delta REG + \varepsilon$$

where ρ is a parameter to be estimated, W is a binary spatial weight matrix and other variables as before. Estimating this regression by Maximum Likelihood (ML), we obtain the results, as reported in Table 8.

Table 8. ML Estimation of the Spatial Lag ANOVA

Independent variables	
Constant	6,46
	(8,76)***
Spatial Lag (ρ)	0,43
	(9,25)***
REG	3,00
	(4,82)***
Adj. R-squared	0,15

Note: t-ratios in parentheses. * p<=0.1; ** p<=0.05 ; ***p<=0.01.

All estimates are highly significant at the one percent level. The constant term and the regional dummy (*REG*) are lower than in the previous case (see Table 6), whereas the coefficient for the spatial lag is 0,43, indicating positive spatial autocorrelation. The intuitive meaning of ρ is that it is important to take into account the neighbor's crime rates to analyze crime over space. The results are relatively different from the previous ones, but this is in line with the fact that the omission of the spatial lag provokes bias in the estimates. Consequently, these estimates can be considered unbiased.

3.4. Outlier Analysis

In an interactive ESDA setting, it is important to identify outliers or high leverage points that spuriously influence the global spatial association measure. Outliers are observations that do not follow the same process of spatial dependence as the majority of the data. Besides, following Anselin (1996, p. 117), "*outliers could thus be considered pockets of local non-stationarity, especially if they correspond to spatially contiguous locations or boundary points*".

It is worth differentiating between global outliers and local outliers. A global outlier is a sample point that has a very high or a very low value relative to all of the values in a dataset. A local (or spatial) outlier is a sample point that has a value that is within the normal range, but it is very high in comparison with its surrounding neighbors.

To detect local outliers, it is necessary to use again the Moran Significance Map (see Map 2). There are two types of local outliers: that municipality that has a high crime rate surrounded by municipalities with a low crime rate (local outlier of type HL); and a municipality with a low crime rate surrounded by municipalities with a high crime rate (local outlier of type LH). We can identify various local outliers across the regions of Minas Gerais. Without naming all of them, we can cite, for example, in the North, the following municipalities as outliers HL: Taiobeiras, Padre Paraíso e Mantena; we find, in turn, the following municipalities as outlier LH: in Triângulo Mineiro, Gurinhatã, Centralina and Conquista; in Center, Felixlândia and Pedro Leopoldo; in South and Southwestern, Carmo da Cachoeira, Campos Gerais and Guape.

3.5. Spatial Trend Analysis

A surface may be decomposed into two main components: a deterministic global trend, sometimes referred as a fixed mean structure, and a random short-range variation.

The spatial trend analysis is aimed at finding out and identifying global spatial trends in the database.

The Trend Surface model tries to capture a spatial drift in the mean. The general model is simple and it is based on a polynomial regression in coordinates of the observations, z_{1i} and z_{2i}. For a quadratic specification, we have:

$$y_i = \alpha + \beta_1 z_{1i} + \beta_2 z_{2i} + \beta_3 z_{1i}^2 + \beta z_{2i}^2 + \beta_5 z_{1i} z_{2i} + \varepsilon$$

This regression explores changes in crime rates as a function of the coordinates, z_{1i}, z_{2i}, squares of z_{1i} and z_{2i}, and an interaction component, z_{1i} z_{2i}. The results of this regression are presented in Table 9.

All coefficients for z_{2i} are significant, indicating a quadratic West-East trend in form of an inverted “U” shape. On the other side, there appear to be no quadratic spatial trend in the South-North direction.

Table 9. OLS Regression of Trend Surface Model

Independent variables	
Constant	-255,46
	(-2,10)**
z_{1i}	-6,37
	(-1,30)
z_{2i}	-10,17
	(-2,54)**
z_{1i}^2	-0,07
	(-1,26)
z_{2i}^2	-0,34
	(-4,45)***
$z_{1i}.z_{2i}$	0,07
	(0,75)
Adj. R-squared	0,15

Note: t-ratios in parentheses. * p<=0.1; ** p<=0.05 ; ***p<=0.01.

The diagnostics are given in Table 10. The very high score for the multicollinearity is due to the strong functional relation among the coordinate terms in the regression, so the t-values and the goodness of the fit indicator may be suspect. The residuals are not normally distributed, but there is no evidence for heteroskedasticity. However, there is continuing evidence for spatial autocorrelation.

Table 10. Diagnostics for Trend Surface Model

	Model	
TEST	*VALUE*	*PROB.*
1. MULTICOLLINEARITY		
Condition number	2456,82	-
2. NORMALITY		
Jarque-Bera	165,61	0,00
3. HETEROSKEDASTICITY		
Koenker-Basset	7,74	0,02
4. SPATIAL DEPENDENCE		
Moran's I	8,43	0,00
Lagrange Multiplier (error)	63,08	0,00
Robust LM (error)	0,50	0,48
Lagrande Multiplier (lag)	63,48	0,00
Robust LM (lag)	0,89	0,35

4. CONCLUSIONS AND FINAL REMARKS

Our application of ESDA to crime rates in Minas Gerais leads to various substantively important conclusions. First, the hypothesis of spatial randomness is clearly rejected. Statistically significant spatial clusters are observed for crime data in 1995. The conclusion is that there is positive spatial autocorrelation. In other words, crime is not distributed evenly and randomly over space in Minas Gerais.

Secondly, we could identify considerable spatial regime heterogeneity. Some of the observed local patterns of spatial association reveal clustering both positive autocorrelation ("hot spots" and "cool spots") and negative autocorrelation. Taking into account these different regimes, it seems to be reasonable to propose a "North-South" division of Minas Gerais. In general terms, in the "North", there would be most "cool spots" (with some low-high spots as well) and, in the "South", there would be most "hot spots" (with some high-low spots). Ultimately, the Triângulo Mineiro and the Belo Horizonte Metropolitan Area host more intense "hot spots". Overall, we perceived that spatial pattern of crime in Minas Gerais has a tendency to concentrate around larger population agglomerations. Therefore, there seems to be a possible association between the crime rate and the urbanization rate. The Spatial ANOVA analysis reinforces the evidence of the existence of these two spatial regimes.

Thirdly, we identified various global outliers and local outliers for 1995. Most local outliers HL are located in the "North", while most local outliers of type LH are located in the "South". We found out that the global outliers detected exert little influence in the computation of Moran's *I*, thereby not altering the signal of the spatial association.

Fourthly, with the help of the trend surface model, we concluded that there is an accentuated West-East global trend in the form of an inverted "U" shape in the crime data for Minas Gerais. Nevertheless, there is no evidence for a South-North trend.

Fifthly, as we showed evidences that crime in Minas Gerais is displayed spatially in the form of spatial clusters and spatial regimes, the policy intervention in terms of crime fight must be the responsibility of the State government, which has, on the one hand, power of establishing state police and, the other, coordinating the municipal efforts to fight criminal activities that spillover the municipal borders.

Finally, the procedure for identifying clusters, outliers and trends discussed in this paper are only initial steps in the understanding of the patterns of crime. Richer econometric models need to be considered to find out the determinants of the crime within a spatial setting. Consequently, the next step would be to insert covariates to explain the crime in Minas Gerais, using spatial econometric models.

In sum, the main conclusion of this paper can be stated in a unique phrase: space matters in the study of crime in Minas Gerais. So, it would be good that we begin to acknowledge this fact in our models.

REFERENCES

Anselin, L., Cohen, J., Cook, D., Gorr, W., Tita, G. Spatial analyzes of crime. In: Criminal Justice 2000, Volume 4, *Measurement and Analysis of Crime and Justice*, edited by David Duffee, Washington, DC: National Institute of Justice, pp. 213-262, 2000.

Anselin, L. Interactive Techniques and Exploratory Spatial Data Analysis. In *Geographic Information System: Principles, Techniques, Manegement and Applications*, edited by P. A. Longley, M.F. Goodchild, D. J. Maguire and D. W. Rhind. New York: John Wiley, pp. 251-264, 1999a.

Anselin, L. Spatial Data Analysis with SpaceStat and ArcView. Mimeo, University of Illinois, 3rd. edition, 1999b.

Anselin, L. The Moran Scatterplot as an ESDA tool to assess Local Instability in Spatial Association. In *Spatial Analytical Perspectives on GIS in Environmental and Socio-Economic Sciences*. London: Taylor and Francis, pp. 111-125, 1996.

Anselin, L. Local Indicators of Spatial Association – LISA. *Geographical Analysis*, 27(2): 93-115, 1995.

Anselin, L. *Spatial Econometrics*. Boston: Kluwer Academic, 1988.

Anselin, L., IBNU, S., SMIRNOV, O., REN, Y. Visualizing Spatial Autocorrelation with Dynamically Linked Windows. *Computing Science and Systems* 33 (forthcoming), 2002.

Anselin, L. and Bera, A. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah A. and Giles D. E. (eds.) *Handbook of Applied Economic Statistics*, Marcel Dekker, New York, 237-289.

Becker, G. Crime and punishment: an economic approach. *Journal of Political Economy*, 76(2): 169-217, 1968.

Cliff, A. and Ord, J. K. *Spatial Processes: Models and Applications*. London: Pion, 1981.

Ehrlich, I. Participation in illegitimate activities: a theoretical and empirical investigation. *Journal of Political Economy*, 81(3): 521-565, 1973

Fajnzylber, P. , Lederman, D. and Loayza, N. Crime and victimization: an economic perspective. *Economia*, 2000.

Fotheringham, A. S., Brunsdom, C. and Charlton, M. *Quantitative* Geography*: perspectives on spatial data analysis*. London: Sage Publications, 2000.

Glaeser, E. L. and Sacerdote, B. *Why is There More Crime in Cities?* National Bureau of Economic Research working paper no. 5430, 1996

Kelly, M. Inequality and Crime. *The Review of Economics and Statistics*, 82(4):530-539, 2000.

Levine, N. *CrimeStat: a spatial statistical program for the analysis of crime incident locations*. Ned Levine & associates Annandale, VA and the National Institute of Justice, Washington, DC. August, 1999.

Messner, S. F. , Anselin, L, Baller, R. D., Hawkins, D. F., Deane, G., and Tolnay, S. E. The Spatial Patterning of County Homicide Rates: an application of exploratory spatial data analysis. *Journal of Quantitative Criminology*, 15(4): 423-450, 1999.

Messner, S. F. and Anselin, L. *Spatial Analyses homicide with areal data*. Mimeo, University of Illinois, 2001.

Odland, J. *Spatial Autocorrelation*. London: Sage Publications, 1988.

Tukey, J. W. *Exploratory Data Analysis*. Reading: Addison-Wesley, 1977.