Crime Modeling using Spatial Regression Approach

Ansari Saleh Ahmar1,2*, Adiatma1, M. Kasim Aidid1
1Department of Statistics, Universitas Negeri Makassar, Indonesia
2AHMAR Institute, Makassar, Indonesia

*ansarisaleh@unm.ac.id

Abstract. Act of criminality in Indonesia increased both variety and quantity every year. As murder, rape, assault, vandalism, theft, fraud, fencing, and other cases that make people feel unsafe. Risk of society exposed to crime is the number of reported cases in the police institution. The higher of the number of reporter to the police institution then the number of crime in the region is increasing. In this research, modeling criminality in South Sulawesi, Indonesia with the dependent variable used is the society exposed to the risk of crime. Modelling done by area approach is the using Spatial Autoregressive (SAR) and Spatial Error Model (SEM) methods. The independent variable used is the population density, the number of poor population, GDP per capita, unemployment and the human development index (HDI). Based on the analysis using spatial regression can be shown that there are no dependencies spatial both lag or errors in South Sulawesi.

1. Introduction
Criminology is a knowledge of crime or criminality that can be viewed from the perspective of biological, sociological, and others. This knowledge also provides two meanings for the term of crime or criminality, namely juridical and sociological. According to Bonger legally, crimes means the act of anti-social consciously gets a reaction from the state for the provision of the pain and then as a reaction to the legal definitions of crimes [1]. Sociologically, the crime is a human behavior that is created by the community.

Criminal acts in Indonesia increased both variety and quantity every year. As murder, rape, assault, vandalism, theft, fraud, fencing, and other cases that make people feel unsafe. According Abdulsyani that crime cases that happened was caused by various factors, both internal and external [2]. For example the level of education obtained, the salary or wages that are not sufficient, and family relations. These factors are found in juvenile detention centers in major countries such as Australia [3] and United States [4]. Additionally, Kakamu, Polasek, & Wago in their study in Japan also noted that the level of crime cases that occur are affected by unemployment, income, and level of arrest [5]. According to the research done before, it can be said that the dominant factor that influence the criminal case is an economic phenomenon.

One way that is used in defining the relationship between economic phenomena regression analysis. To determine the economic impact, especially on the level of crime cases using regression analysis, spatial linkages between regions Noteworthy because it is associated with a location / region. Spatial linkages between regions it tends often occurs due to the location adjacent to each other. To determine the effect of the relationship between the location of the existing response, then each region...
can not be regarded as an observation unit that stands alone and is not associated with the location area more so that the necessary spatial analysis.  

Research using regression analysis with spatial approach had been done already. Kusrini, Suhartono, & Sari using regression approach spatially to model criminal cases in the city of Surabaya, the results of the study can be concluded that the measurement of the risk of a criminal act did not indicate any dependencies spatial variables significant effect is the population of junior high school education, which has negative and the level of per capita income of the population negatively [6]. Septiana & Wulandari using regression method spatially to model the dropout age from high school in East Java province, on the results of these studies it is concluded that there are dependencies in lag and error, where the model of SAR variables that have a significant effect is poor, the negative effect and the location of the house in the countryside a positive effect [7]. While SEM models, the variables are significant and positive impact is the location of the house in the countryside.

Research modeling criminality in South Sulawesi with spatial regression approach has not been done. Therefore, researchers wanted to do modeling criminality in the province of South Sulawesi with spatial regression approach that can be known or not existing spatial dependencies criminality in the province of South Sulawesi.

2. Literature

2.1. Spatial Regression Model

According to the Tobler’s first law [8] that “Everything is related to everything else, but near thing are more related than distant things”. The law that became the reference of the regional science studies. Spatial effect is usually found on a region to another, meaning that the observations in a region depends on the observation in an area that became neighbors or adjacent areas. Spatial regression is a statistical method used to determine the relationship between the response variable and the predictor variables taking into account the linkages between regions.

Anselin developed a general model of spatial regression using cross section data [9]. The general model of spatial regression equation written in the form (1).

\[ y = \rho W_1 y + X\beta + u, \quad \text{with} \quad u = \lambda W_2 + \epsilon \]  

\[ y = (I_n - \rho W_1)^{-1} X\beta + (I_n - \rho W_1)^{-1} (I_n - \lambda W_2)^{-1} \epsilon 
\]

\[ \epsilon \sim N(0, \sigma^2 I_n) \]

From the general form of spatial regression, there are some models that can be formed in cross-section data, namely:

1. If \( \rho = 0 \) and \( \lambda = 0 \) then the equation becomes:

\[ y = X\beta + \epsilon \]  

This equation is called the classical regression model by ignoring the spatial effects.

2. If \( \rho \neq 0 \) and \( \lambda = 0 \) then the equation becomes:

\[ y = \rho W_1 y + X\beta + \epsilon 
\]

\[ y = (I_n - \rho W_1)^{-1} X\beta + (I_n - \rho W_1)^{-1} \epsilon 
\]

This equation is called Spatial Autoregressive Model (SAR).

3. If \( \rho = 0 \) and \( \lambda \neq 0 \) then the equation becomes:

\[ y = X\beta + u, \quad \text{with} \quad u = \lambda W_2 + \epsilon 
\]

\[ y = X\beta + (I_n - \lambda W_2)^{-1} \epsilon 
\]

This equation is called Spatial Error Model (SEM).
4. If $\rho \neq 0$ and $\lambda \neq 0$ then the equation becomes

$$y = \rho \mathbf{W}_1 y + \mathbf{X} \beta + u,$$

with $u = \lambda \mathbf{W}_2 + \varepsilon$

$$y = (I_N - \rho \mathbf{W}_1)^{-1} \mathbf{X} \beta + (I_N - \rho \mathbf{W}_1)^{-1} u$$

$$y = (I_N - \rho \mathbf{W}_1)^{-1} \mathbf{X} \beta + (I_N - \rho \mathbf{W}_1)^{-1} (I_N - \lambda \mathbf{W}_2)^{-1} \varepsilon$$

(5)

This equation is called a General Spatial Model or Spatial Autocorrelation (SAC).

2.2. Spatial Dependencies

Anselin defines the spatial dependencies as their functional relationship between what happens at one point in space and what happens elsewhere [9]. To see the magnitude of the spatial dependencies can use the index Morans's I defined in equation (6).

$$I = \frac{\mathbf{e}^t \mathbf{W} \mathbf{e}}{e^t \mathbf{e}}$$

(6)

The weighting matrix that can use the index has not been standardized Moran's I defined in equation (7).

$$I = \frac{\mathbf{N} \mathbf{e}^t \mathbf{W} \mathbf{e}}{\mathbf{e}^t \mathbf{e}}$$

(7)

To see if the magnitude of the spatial dependencies ($I_j$) is significant in the data can be done by testing on Moran's I index the following hypotheses.

$H_0: I_j = 0$ (There are no spatial dependencies)

$H_1: I_j \neq 0$ (There is a spatial dependencies)

Then the test statistic used in tests of significance on Moran's I index by Cliff and Ord [10] defined in equation (8).

$$Z = \frac{|I_j - \mathbb{E}(I_j)|}{\text{var}(I_j)^{1/2}}$$

(8)

2.3. Spatial Weighting

The weighting matrix/spatial weights ($W$) can be obtained by distance information of the adjacency (neighborhood) or the distance from one location to another. There are several methods for defining the intersection relations (contiguity) between these locations. According Lesage that the method include [11]:

1. Linear contiguity (the intersection edge);
2. Rook contiguity (intersection side);
3. Bisop contiguity (intersection angle);
4. Double linear contiguity (the intersection of two edges);
5. Double rook contiguity (the intersection of the two sides);
6. Queen contiguity (side-angle intersection);
3. Methods

The method used in this study is cross-sectional spatial using data sourced from BPS-South Sulawesi province, Indonesia. The steps are as follows.

1. Exploration data on thematic maps to determine the spatial distribution patterns and dependencies on each variable and scatterplot to determine the pattern of variables and relationships.
2. Spatial regression modeling with the method of ordinary least squares (OLS), which includes estimation parameters, significant estimation models, test assumptions residual (identical, independent, and normal distribution).
3. Determine the spatial weighting (W) is using a chess queen (queen contiguity).
4. Test spatial dependency or spatial correlation between observations that are close together with the index moran or Morans' I.
   Hypotheses used to test the spatial dependencies are as follows.
   \[ H_0 : I = 0 \] (There is no spatial dependencies)
   \[ H_1 : I \neq 0 \] (There is a spatial dependencies)
5. The identification of the existence of spatial effect by using Lagrange Multiplier test (LM). LM Tests conducted to find out which models are in accordance with the procedure.
6. Process modeling, ie the data is modeled with Autoregressive Spatial Model (SAR), Spatial Error Model (SEM), or Autoregressive Spatial Moving Average (Sarma).
7. To examine the assumption of spatial models.
8. Summing up the results obtained.

4. Result

4.1. Administrative Map of South Sulawesi Province

Administrative map of South Sulawesi province can be seen in Figure 1, while the variable data dependent and independent variables used can be seen as follows.

![Figure 1. Administrative map of South Sulawesi Province, Indonesia](image)

4.2. Exploration Data

Figure 2 below is a map of the affected population thematic risk of crime in the province of South Sulawesi, Indonesia, where seen no degradation of color which indicates the values of the communities affected by the risk of crime. The more intense the color of the region, the higher the risk
of the communities affected by crime in the region. To determine the risk level of the communities affected by crime then categorized into five categories: very low, low, medium, high and very high.

![Map Thematic Risk Residents Affected by Crime in South Sulawesi](image)

**Figure 2.** Map Thematic Risk Residents Affected by Crime in South Sulawesi

Risk population exposed to very low crime category (113-420 people) are in Bantaeng, Jeneponto, Luwu, Takalar and Wajo. Risk population exposed to criminality with the low category (239-295 people) are in Enrekang, Sidrap, Pangkep and Sinjai. Risk population being affected by crime category (317-449 people) are in Bone regency, Tana Toraja, Barru, Selayar and Palopo. Risk population exposed to high criminality by category (486-584 people) are in Pinrang, Soppeng, Bulukumba and East Luwu Regency. And the risk of the communities affected by crime category is very high (686-4319 inhabitants) is located in the City of Pare-pare, North Luwu, Maros, Gowa and Makassar.

4.3. Moran’s I

By value $\alpha = 0.1$, then the value $Z(I) = 0.41851 < Z_{\alpha/2} = 1.645$, it means failed to reject $H_0$. This indicates that there is no spatial dependencies.

| Variabel | Moran’s I | Z(I) |
|----------|-----------|------|
| $Y$      | 0.11889   | 0.41851|
| $X_1$    | -0.01774  | 0.28188|
| $X_2$    | 0.08274   | 0.38236|
| $X_3$    | -0.06733  | 0.23229|
| $X_4$    | -0.03210  | 0.26752|
| $X_5$    | 0.11609   | 0.41571|

$Z_{0.05} = 1.645$, $I_0 = -0.0454$

Table 1 above shows that all of the dependent and independent variables generate value $Z(I) < Z_{\alpha/2} = 1.645$, it means failed to reject $H_0$. This indicates that there is no inter-regional spatial dependencies. Also seen, $I_0$ that all grades Moran’s I value is greater than $I_0$ that means that all the
dependent and independent variables indicating positive spatial dependency which means adjacent sites have similar values and tend to cluster. Unless the GDP per capita variable (X3) where the value of Moran's I is worth less than $I_0$ and negative value, the indicated that spatial dependencies have significant negative adjacent locations have different values.

4.4. Lagrange Multiplier (LM)

Selection of spatial models made by Lagrange Multiplier (LM) as the initial identification. Lagrange Multiplier (LM) is used to detect the spatial dependencies with more specific, ie with dependencies in the lag, error or both (lag and error). When LM LM lag and no significant error does occur it can be concluded either dependency on lag and error. Test conducted on a weighted spatial dependencies queen contiguity. The test results Lagrange Multiplier (LM) in Table 2. Then berdasarkan Table 2 shows that by using a weighted queen contiguity p-value Moran's I = not significant at $\alpha = 10\%$. This means that there is no spatial dependencies.

| Spatial Dependency Test | Value     | P-value   |
|-------------------------|-----------|-----------|
| Moran's I               | -0.3236764| 0.7461832 |
| Lagrange Multiplier (lag)| 0.9296950 | 0.3349422 |
| Lagrange Multiplier (error)| 0.5521177 | 0.4574537 |
| Lagrange Multiplier (SARMA)| 3.1437015 | 0.2076605 |

5. Conclusion

Based on the analysis and discussion drawn some conclusions about the risk of the communities affected by crime in the province of South Sulawesi showed no spatial dependencies in both the lag and error. Based on the results of this study, it presented some suggestions as follows: (1) Subsequent research include variables that are considered essential and are not included in this study. (2) Researcher further use customized methods as weighting. (3) Provincial and Local Government should be more serious in taking into pemerata an income distribution.

References

[1] Bonger, W. A. 1981. Pengantar Tentang Kriminologi. Jakarta: Ghalia Indonesia.
[2] Abdulsyani. 1987. Sosiologi Kriminalitas. Bandung: Remaja Karya.
[3] Cunneen, C. and White, R. 2011. Juvenile justice: Youth and crime in Australia. Oxford University Press.
[4] Krisberg, B. 2005. Juvenile justice: Redeeming our children. Sage.
[5] Kakamu, K., Polasek, W. and Wago, H. 2008. Spatial interaction of crime incidents in Japan. Mathematics and Computers in Simulation. 78(2), pp. 276–282.
[6] Kusrini, D. E., Suhartono, S. and Sari, D. M. 2013. Pemodelan Kasus Tindak Pidana di Kota Surabaya dengan Pendekatan Regresi Spasial. J. Sains dan Seni ITS, vol. 2, no. 2, pp. D135–D140.
[7] Septiana, L. and Wulandari, S. P. 2009. Pemodelan Remaja Putus Sekolah Usia SMA di Provinsi Jawa Timur dengan Menggunakan Metode Regresi Spasial. Institut Teknologi Sepuluh Nopember.
[8] Miller, H. J. 2004. Tobler’s first law and spatial analysis. Annals of the Association of American Geographers. 94(2), pp. 284–289.
[9] Anselin, L. 1988. Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic.
[10] Cliff, A. D. and Ord, J. K. 1981. Spatial processes: methods and applications. London: Pion.
[11] LeSage, J. P. 1999. The theory and practice of spatial econometrics. University of Toledo. Toledo, Ohio, 28, p. 33.