BIS Economics and Business

Volume 1 2024 | DOI: 10.31603/biseb.111
5th Borobudur International Symposium on Humanities and Social Science (BIS-HSS) 2023

UNIMMA PRESS
UNIVERSITAS MUHAMMADIYAH MAGELANG

Spatial regression of poverty in East Java Province with distance weighting matrix

S Yulianto[1*], M M Pangestu[1] and A N Ambarwati[1]

[1] Department of Statistics, Institut Teknologi Statistika dan Bisnis Muhammadiyah Semarang, Semarang, Indonesia

[*]Corresponding author email: safaat.yulianto@itesa.ac.id

Abstract

Poverty has emerged as a pressing issue in developing nations, such as Indonesia. It refers to a state where individuals are unable to meet their fundamental necessities. Between September 2020 and March 2021, there was a notable rise of 20.09 thousand individuals living in poverty in urban regions. This increase was the most significant in East Java Province when compared to other provinces in Indonesia. Analyze poverty cases using spatial analysis with distance weighting matrix. Poverty in a region is influenced by poverty in the surrounding areas, resulting in poverty data that includes spatial effects or regional aspects. To analyze data with these spatial effects or regional aspects spatial regression is employed. In this research, the variable used is the poverty level in the Eastern region of Java Province (Y), with four X variables, average years of schooling (X_1), minimum wage per regency city (X_2), open unemployment rate (X_3), and total population (X_4). The weighting matrix utilized is the Euclidean Distance Weighting Matrix, which calculated the distance weighting matrix between different areas. The method used in this study uses spatial regression model. The best Spatial Regression Model used is the Spatial Error Model (SEM). Modeling results using the Spatial Error Model (SEM) and the factors that influence poverty are average years of schooling (X_1) and open unemployment rate (X_3).

Keywords

Poverty, Spatial regression, East Java

Published:
October 20, 2024

Selection and Peer-review under the responsibility of the 5th BIS-HSS 2023 Committee

Introduction

Poverty is a state of not fulfilling one's basic rights and has become one of the issues in global [1] that occur in all developing countries, including Indonesia [2]. According to BPS East Java data, the average poverty rate in East Java ranks 14th among 34 provinces in Indonesia and has the third highest poverty rate in Java [3]. This shows that the inequality of economic growth in East Java is still relatively high and needs to be addressed immediately. East Java has a high rate of poverty because of its enormous population, which is not balanced by more employment opportunities or more even distribution of the people. Java continues to have the largest population concentration.

Data from the BPS 2020 census show that Java is home to 56.10% of Indonesia's total population, with East Java accounting for 26.11% of that number. Between September 2020 and March 2021, there was a 20.09 thousand increase in the number of impoverished individuals living in metropolitan areas. This is the highest increase of poverty percentage among other provinces. As a result, there is an imbalance in the distribution of development and income gaps, as well as uneven economic growth.

The phenomenon of increasing poverty in East Java is still quite high compared to 34 provinces in Indonesia and several provinces in Java [4]. This has become the focus of the government on how to overcome poverty in East Java. The impact of poverty on the economy is very detrimental, and to overcome the problem of poverty in Indonesia, it is necessary to know the factors that cause poverty. Previous research on the analysis of factors affecting poverty using a multiple linear regression analysis approach shows that the provincial minimum wage has a negative and significant effect on poverty, the human development index has a positive and insignificant effect on poverty, economic growth has a negative and insignificant effect on poverty, unemployment has a positive and significant effect on poverty, and at the same time, all independent variables affect poverty in East Java [5].

One way to identify poverty factors is through regression modeling analysis. However, characteristics of poverty are most likely influenced by the variety of locational factors in addition to explanatory variables. The foundation of the study of spatial data analysis is Tobler's first law of geography [6], which asserts that everything is related to everything else, but that something closer will have greater influence than something farther away. In spatial data, often observations in one location (space) often observations in other neighboring locations. Furthermore, research on the poverty rate model of East Java Province with spatial regression analysis, obtained the result that the variable number of disabled people and unprotected drinking water sources affect the poverty rate of districts/cities in East Java Province [7]. Another study using Spatial Panel Model shows that by using Spatial Autoregressive (SAR) and Spatial Error Model (SEM), variables that affect poverty include gross domestic product, life expectancy and average years of schooling [8]. The percentage of the population aged 15 and over who has completed primary school, the percentage of the poor population aged 15 and over who graduated from elementary or junior high school, the percentage of branded packaged drinking water sources, the percentage of refilled drinking water sources, the percentage of the population in East Java with a morbidity rate, and the percentage of the population in East Java with health insurance are the variables that affect poverty, according to research using Spatial Autoregressive Moving Average (SARMA) to model the percentage of the poor population in Districts/Cities in East Java Province in 2021 [9].

This description piques the curiosity of academics who wish to use a distance weighting matrix in conjunction with geographical analysis to examine these poverty cases. The open unemployment rate (TPT), average years of schooling (RLS), district/city minimum salary, and total population are among the criteria used by researchers to determine the percentage of impoverished individuals in East Java Province, which has 29 districts and 9 cities in 2021.

Method

Spatial regresion

By taking location or spatial correlations into account, spatial regression is a statistical technique used to ascertain the relationship between dependent variables and independent factors [10]. The basis for the development of spatial regression methods is the classical linear regression method. The development is based on the influence of location or space on the data analyzed. Spatial regression model that consists of:

1. Spatial autoregressive models (SAR)

Using cross-sectional data, the SAR model combines a basic regression model with a geographical lag on the response variable [11]. This model is formed if $\rho \neq 0$ and $\lambda = 0$, the SAR model equation is as follows:

$$y = \rho W y + X\beta + \epsilon$$
$$y = (1 - \rho W)^{-1} X\beta + (1 - \rho W)^{-1}\varepsilon \tag{1}$$

2. Spatial error models (SEM)

The association between the error value at one site and the error value in the vicinity produces the spatial error model [12]. The SEM model is formed if the value of $\rho = 0$ dan $\lambda \neq 0$. The spatial error model's (SEM) formula is the following:

$$Y = X\beta + (I - \lambda W)^{-1}\,\varepsilon \tag{2}$$

3. Spatial autoregressive moving average (SARMA)

Regression based on spatial areas (Spatial Autoregressive Moving Average, or SARMA) assumes that spatial influences also affect the residuals and the dependent variable [9]. In addition, the SARMA model is commonly used in the analysis of cross-sectional data with a spatial weighting matrix as a form of relationship between regions [13]. This model is formed when the value of $\rho \neq 0$ and $\lambda \neq 0$, and the SARMA equation is written as follows:

$$y = \rho W y + X\beta + (1 - \lambda W)^{-1}\varepsilon \tag{3}$$

Where :

ρ = spatial autoregressive lag coefficient
λ = error coefficient
W = spatial weighting matrix (n x n)
y = response variable (n x 1)
X = predictor variable (n x k)
B = estimated parameter (k x 1)
E = error (n x 1) $\varepsilon \sim N(0, I\sigma)^2$

Spatial weighting matrix

A geographic weight matrix, which depicts the relationship between a region and other areas based on data about the position of a (n x n) area, and a matrix, which describes the relationship of observation locations, are the essential components of the spatial analysis model. There are several ways to determine the spatial weighting matrix [14], including contiguity weight (intersection between regions) and distance weight (distance between regions).

The Euclidian distance weight matrix is employed in this investigation. The definition of Euclidean distance, which is the search for a distance weight matrix between two regions, is as follows: [15]

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \tag{4}$$

Spatial autocorrelation test

Spatial dependence occurs because of the dependence of regional data. The existence of spatial dependence suggests that the value of characteristics in one area is correlated with the values of characteristics in adjacent or surrounding places. To determine whether there is autocorrelation or geographic dependence between observations or places, utilize Moran's I coefficient [16]. Positive or negative spatial autocorrelation is possible. When data from nearby sites are comparable, they tend to cluster together, as indicated by positive spatial autocorrelation. Conversely, negative spatial autocorrelation suggests that nearby sites tend to disperse and have disparate value. The hypothesis used in this test is:

$H_0 : I = 0$ vs $H_1 : I \neq 0$

The Moran's index equation is as follows:

$$I = \frac{\sum_{i=1}^{n} W_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^{n} W_{ij}(X_j - \bar{X})^2} \tag{5}$$

To ascertain whether or not there is spatial autocorrelation, a significance test of Moran's index is conducted using the hypothesis:

H_0 : no spatial autocorrelation vs H_1 : spatial autocorrelation exists

The test statistic used is:

$$Z(I) = \frac{I - E(I)}{\sqrt{Var\,(I)}} \tag{6}$$

Where $E(I) = I_0 = -\frac{1}{n-1}$; $Var(I) = \frac{n[(n^2-3n+3)S_1 - nS_2 + 2S_0^2]}{(n-1(n-2)(n-3)S_0^2} - [E(I)]^2$; $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}$

H_0 is rejected if $Z(I) > Z\alpha/2$

For a normalized spatial weight matrix, the value of Moran's index falls between -1 and 1. When I < Io, there is negative spatial autocorrelation; when I > Io, there is positive autocorrelation. When the Moran index is zero, there is no clustering. With Moran's Scatterplot, patterns of dispersion and clustering between places can also be seen. The association between standardized observations can be seen using Moran's Scatterplot. The types of spatial relationships, according to [17], the following are the quadrants of the Moran scatter:

1. sites with high observed values are encircled by sites with high observed values in Quadrant I (High-High).
2. sites with low observed values are encircled by sites with high observed values in Quadrant II (Low-High).
3. Quadrant III (Low-Low) displays low observed value locations encircled by low observed value locations.
4. Quadrant IV, or "High-Low," displays areas with high observed values encircled by areas with low observed values.

Spatial dependence test

The Lagrange Multiplier (LM) test is used to determine which spatial regression model is best by testing for spatial effects or dependencies. Lagrange Multiplier lag and Lagrange Multiplier error make up the Lagrange Multiplier test. The proper model is the spatial autoregressive (SAR) model if the Lagrange Multiplier lag is large, and the spatial error model (SEM) model if the Lagrange Multiplier error is considerable. The statistical Lagrange Multiplier lag test is used to visualize the lag's spatial dependence:

$$LM_{lag} = \frac{(\varepsilon^T W_y)^2}{S^2((WX\beta)^T M(WX\beta) + TS^2)} \tag{7}$$

The test statistic of the lagrange multiplier of the error test statistic is employed in order to observe the spatial dependence of the errors:

$$LM_{error} = \frac{(\varepsilon^T W_\varepsilon / S^2)^2}{T} \tag{8}$$

The following phases were involved in finishing the investigation and examination of the poverty statistics in the Province of East Java:

1. An explanation of the East Java Province's poverty rate in 2021
2. Carrying out an examination of spatial regression:
   a. Selecting a spatial weight matrix
   b. Use Moran's I test [18], to verify the presence of geographical effects between regions, testing the null hypothesis that there is no spatial autocorrelation.
   c. To ascertain whether spatial interaction is present on lags or residuals, run a Lagrange Multiplier test [10].
3. Estimated the parameter of the spatial regression model
4. Perform a significance test on the parameters.

5. Interpret the model

Results and Discussion

Description of data

This thematic map is created using data regarding the percentage of poverty in the Province of East Java. The distribution of poverty in the districts and municipalities of East Java Province is shown in Figure 1, with the areas having the highest percentage of poverty represented by the darkest hues. Bangkalan Regency, Sampang Regency, Sumenep Regency, and Probolinggo Regency are these regions. However, the areas with the lightest hue are those with the lowest rates of poverty, such as Batu City, Tulungagung Regency, Banyuwangi Regency, Kediri City, Blitar City, Malang City, Probilinggo City, Pasuruan City, Mojokerto City, Madiun City, Surabaya City, and Sidoarjo Regency.

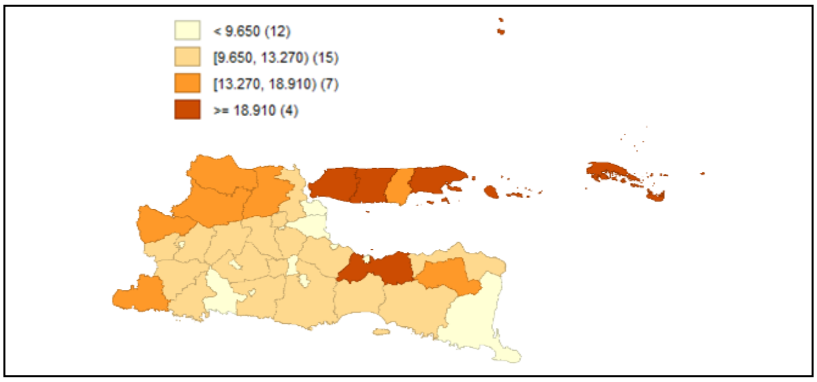

Figure 1. Poverty distribution map of East Java Province in 2021

Additionally, Moran's scatterplot can be used to visually represent the spatial dependence across regions. This is a picture of a Moran's scatterplot:

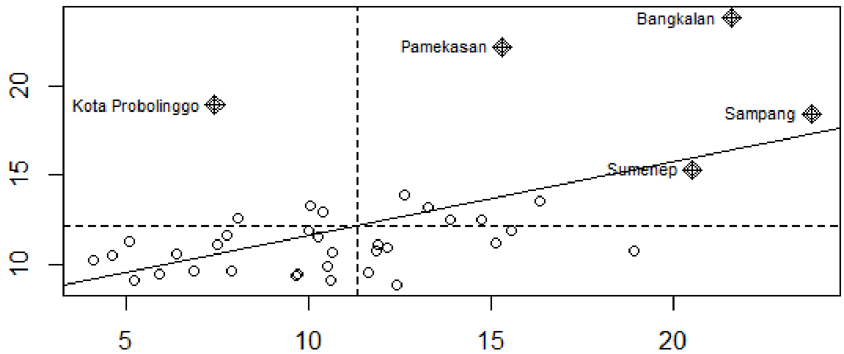

Figure 2. Moran's scatterplot

Figure 2 illustrates how each of these can be explained as follows:

1. Quadrant I: High-High (HH) shows that locations with a high proportion of the impoverished are encircled by other areas with a high proportion of the impoverished. The regions in quadrant I are the following: Bojonegoro Regency,

Bondowoso Regency, Situbondo Regency, Lamongan Regency, Sumenep Regency, Tuban Regency, Sampang Regency, Pamekasan Regency, and Bangkalan Regency.

2. Quadrant II: Low-High (LH) shows that places with a high percentage of the impoverished are encircled by those with a low percentage. Probolinggo City, Lumajang Regency, Jember Regency, and Banyuwangi Regency are the regions in quadrant II.
3. Quadrant III: Low-Low (LL) demonstrates that low-poverty communities are encircled by low-poverty surrounding areas. The regions in quadrant III are Kediri City, Blitar City, Malang City, Pasuruan City, Mojokerto City, Madiun City, Surabaya City, and Batu City. Ponorogo Regency, Tulungagung Regency, Blitar Regency, Malang Regency, Pasuruan Regency, Sidoarjo Regency, Mojokerto Regency, Jombang Regency, and Magetan Regency are also included.
4. Quadrant IV: High-Low (HL) shows that places with a high proportion of the impoverished are encircled by areas with a low proportion of the impoverished. Quadrant IV comprises the following regencies: Pacitan, Trenggalek, Kediri, Probolinggo, Nganjuk, Madiun, Ngawi, and Gresik. Regencies.

Test of spatial autocorrelation

The Moran index test is performed to detect the presence of spatial influence (spatial autocorrelation) so that spatial autocorrelation modeling can be performed. The results of the Moran index test are show in Table 1.

Table 1. The Moran's I index test results

Variable	Morans'I	Expectation Index	p-value
Percentage of Poverty	0.060	- 0.027	0.002

Table 1 indicates that there is spatial autocorrelation in the percentage of impoverished individuals in East Java Province in 2021, with a p-value less than α (0.1). Given that the expectation value (I0) acquired is -0.027 and the obtained Morans' I value is 0.060, it may be concluded that I > I0. This demonstrates that there is a clustering pattern and positive autocorrelation.

Test of spatial dependence

The Lagrange Multiplier test (LM test) is used as a basis for selecting an appropriate spatial regression model.

Table 2. Lagrange multiplier test results

Test	p-value
Lagrange Multiplier (lag)	0.628
Lagrange Multiplier (error)	0.095*

From Table 2, it is obtained information that the p-value of the Lagrange Multiplier error result is smaller than the α value (0.1), while the p-value of the Lagrange Multipliear lag result is larger than the α value (0.1), so it can be concluded that there is a spatial dependence on the error, so that the spatial model can be analyzed using the Spatial Error Model (SEM) model.

Spatial error model (SEM)

Based on the result of the Lagrange Multiplier test results, a significant spatial regression model is the Spatial Error Model (SEM). The next step is to test the model. Table 3 shows that the open unemployment rate (X3), average years of education (X1), and percentage of the population living in poverty are the three variables that significantly affect the dependent variable (% of the poor population). While the variable minimum wage per city district (X2), and population (X4) does not significantly affect the dependent variable. Thus, the next step is to remove the insignificant variables from the model, then regress again and obtain the following results in Table 4.

Table 3. Lagrange multiplier test for SEM model

Variable	Coeff	Std. Error	Z - value	Prob.
Constant	35.873	2.804	12.792	0.000
Lambda	0.656	0.191	3.442	0.001
X1	-3.296	0.471	-7.002	0.000*
X2	2.766 x 10-7	7.599 x 10-7	0.364	0.716
X3	0.479	0.341	1.406	0.016*
X4	-9.881 x 10-7	7.537 x 10-7	-1.311	0.190

Table 4. Results of the best SEM model

Variable	Coefficient	Std. Error	Z - Value	Prob.
Constant	34.116	2.537	13.447	0.000
Lambda	0.688	0.177	3.896	0.000
X1	-3.008	0.411	-7.323	0.000*
X3	0.311	0.315	0.988	0.032*

Conclusion

Based on the results of the analysis and discussion in this study, it can be concluded that the presentation of the poor population in East Java Province in 2021 there are 4 areas that are in the category of very high presentation of the poor. These areas are Bangkalan Regency, Sampang Regency, Sumenep Regency, and Probolinggo Regency.

The spatial regression model equation that is produced is because the regression model that is utilized is the spatial error model (SEM):

$$\widehat{Y_i} = 34.116 - 3.008Xi_1 + 0.311Xi_3 + 0.688 \sum_{j=1,i\neq j}^{38} W_{ij}y_j$$

In the formula, the proportion of the impoverished population in the districts and cities of East Java Province in 2021 is influenced by the open unemployment rate and the average number of years of education. The lambda coefficient (λ=0.688) indicates the degree to which an area is impacted by its surroundings.

References

[1] S. Soekanto and B. Sulistyowati, Sosiologi : Suatu pengantar, Ed. Revisi. Jakarta: Rajagrafindo Persada, 2017.

[2] M. I. Rizki and T. A. Taqiyyuddin, "Pemodelan Regresi Spatial Autoregressive Fixed Effect Model Data Panel Pada Tingkat Kemiskinan Di Provinsi Jawa Barat," J Stat. J. Ilm. Teor. dan Apl. Stat., vol. 14, no. 1, pp. 44–51, 2021, doi: 10.36456/jstat.vol14.no1.a3816.
[3] B. P. S. (BPS), "Provinsi Jawa Timur Dalam Angka 2021," BPS Provinsi Jawa Timur, 2021.
[4] B. P. S. (BPS), Perkembangan Pembangunan Provinsi Jawa Timur. BPS Provinsi Jawa Timur, 2022.
[5] L. Priseptiawan and W. P. Primandhana, "Analisis faktor-faktor yang mempengaruhi kemiskinan," Forum Ekon., vol. 24, no. 1, pp. 45–53, 2022.
[6] S. Yulianto, A. Djuraidah, and A. H. Wigena, "Model Otoregresif Simultan Bayes untuk Analisis Data Kemiskinan," Bandung: Statistika, Universitas Padjadjaran, 2011, pp. 406–413.
[7] S. Yulianto and C. A. Ayuwida, "Model tingkat kemiskinan Provinsi Jawa Timur dengan analisis regresi spasial," Semin. Nas. Mat. Dan Pendidik. Mat. (6 th Senat., vol. 6, pp. 121–127, 2021.
[8] S. Yulianto and G. E. Romandilla, "Pemodelan Regresi Data Panel pada Data Kemiskinan di Provinsi Jawa Timur," Semin. Nas. Mat. Dan Pendidik. Mat. (7th Senat., vol. 7, no. November 2022, pp. 29–36, 2022.
[9] W. Pramesti, F. Fitriani, and K. L. Nirmala, "Spatial Autoregresive Moving Average Pada Pemodelan Persentase Penduduk Miskin Kabupaten/Kota Di Provinsi Jawa Timur Tahun 2020," J Stat. J. Ilm. Teor. dan Apl. Stat., vol. 15, no. 1, pp. 158–166, 2022, doi: 10.36456/jstat.vol15.no1.a5838.
[10] L. Anselin, Spatial Econometrics : Methods and Models. Dordrecht: Kluwer Academic Publishers, 1988.
[11] K. K. S. Dewi P, M. Susilowati, and I. W. Sumarjaya, "Metode Spatial Autoregressive Dalam Memodelkan Masyarakat Yang Berperilaku Mck Di Sungai," E-Jurnal Mat., vol. 6, no. 4, pp. 233–240, 2017, doi: 10.24843/mtk.2017.v06.i04.p171.
[12] R. Rahmawati, D. Safitri, and O. U. Fairuzdhiya, "ANALISIS SPASIAL PENGARUH TINGKAT PENGANGGURAN TERHADAP KEMISKINAN DI INDONESIA (Studi Kasus Provinsi Jawa Tengah)," Media Stat., vol. 8, no. 1, pp. 23–30, 2015.
[13] N. M. L. Lispani, I. W. Sumarjaya, and I. K. G. Sukarsa, "Pemodelan Jumlah Tindak Kriminalitas Di Provinsi Jawa Timur Dengan Analisis Regresi Spatial Autoregressive and Moving Average," E-Jurnal Mat., vol. 7, no. 4, p. 346, 2018, doi: 10.24843/mtk.2018.v07.i04.p224.
[14] K. S. Suryowati, M. Nahak, and R. D. Bekti, "Penerapan Model Spasial Menggunakan Matriks Pembobot Queen Contiguity dan Euclidean Distance Terhadap Kasus Gizi Buruk Balita di Provinsi Nusa Tenggara Timur," J Stat. J. Ilm. Teor. dan Apl. Stat., vol. 16, no. 1, pp. 298–308, 2023, doi: 10.36456/jstat.vol16.no1.a7871.
[15] F. B. Lega, R. Dwi Bekti, J. Statistka, and A. Yogyakarta, "Kajian Pengaruh Matriks Pembobot Spasial Dalam Model Spasial Data Panel Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Tingkat Kemiskinan Di Provinsi Nusa Tenggara Timur," J. Stat. Ind. dan Komputasi, vol. 8, no. 1, pp. 1–14, 2023.
[16] J. LeSage and R. K. Pace, Introduction to Spatial Econometrics. CRC Press, Taylor and Francis Group, 2009.
[17] S. Sukarna, W. Sanusi, and H. Hardiono, "Analisis Moran's I, Geary's C, dan Getis-Ord G pada Penerapan Jumlah Penderita Kusta di Kabupaten Gowa," J. Math. Comput. Stat., vol. 2, no. 2, p. 151, 2020, doi: 10.35580/jmathcos.v2i2.12577.
[18] J. Lee and D. W. S. Wong, Statistical Analysis with Arcview Gis. New York: John Wiley & Sons, 2005.