Spatial analysis of factors influencing Gross Regional Domestic Product (GRDP) in East Java: a spatial durbin error model analysis

N Kholifia, S Rahardjo*, M Muksar, N Atikah, and D L Afifah

Department of Mathematics, Faculty of Mathematics and Science, Universitas Negeri Malang

*Corresponding author: swasono.rahardjo.fmipa@um.ac.id

Abstract. Regression analysis is not always a suitable solution if the analyzed data contains spatial effects. In overcoming the spatial effect on the data, a statistical method that can overcome it is needed. Spatial regression is a method used for data that has a location effect. One of the spatial regression model that can be used is the spatial Durbin error model. Spatial Durbin error model can overcome the spatial autocorrelation relationship in the independent variables and overcome the spatial error between regions. Spatial effects influence the rate of economic growth in an area, so it is different in each region. The quality of economic growth is an essential indicator in measuring the welfare of an area’s people. The Gross Regional Domestic Product (GRDP) can measure the rate of economic growth. Many factors affect the size of the GRDP, including the total workforce, the number of industries, the number of labour, general allocation funds, regional revenue, and regional expenditure. This study uses the Spatial Durbin Error Model to model GRDP and map the level of economic growth in 38 districts/cities in East Java.

1. Introduction

Spatial regression is the development of simple linear regression, where there is a spatial effect on the data to be analyzed. The spatial effects that arise are spatial heterogeneity and spatial dependence [1]. The spatial heterogeneity shows the differences in characteristics from one region to another, while the spatial dependence shows the dependence between adjacent areas. According to [2], spatial dependence represents a situation in which the observation values of the nearest neighbour affect the values observed in a location or region. Data containing spatial effects will not produce biased and inconsistent estimators when analyzed using OLS [3].

There are several types of spatial regression, including Spatial Autoregressive (SAR), Spatial Error Model (SEM), and Spatial Durbin Model (SDM). The particular case of spatial autoregressive, which is shown by the spatial effect on the dependent variable and the independent variable can be modelled with the Spatial Durbin Model (SDM). Meanwhile, an alternative SEM model that does not allow spatial lag influence on the dependent variable is the Spatial Durbin Error Model (SDEM). In its application, a local multiplier is used in the SDM model, which is then developed for SDEM, which uses a local multiplier to simplify modelling [3]-[4]. The SDEM model has been applied several times as in research that discusses the factors that influence the human development index in Central Java Province [5], research on Spatial Durbin Error Model with finite distribution lags [6], research on Bayesian estimation on Spatial Durbin Error Model [7], modelling energy efficiency in China using fixed-effect stochastics SDEM [8], [9] conducting research that developed the Bayesian Lasso prior.
type of SDEM and used it to model economic growth, and research on the effect of media promotion on the development of small business innovation in various regions. [10].

Economic growth is a process in which income increases without linking it to growth rates [11]. The success of a country’s economic development can be measured from its economic growth. With a high level of economic growth, the level of welfare of the people is also high and vice versa.

Indonesia is one of the countries included in the category of developing countries in the world. With an area divided into 34 provinces, the Indonesian government strives to improve the welfare of its population. Gross Regional Domestic Gross (GRDP) is an indicator used to determine the level of economic growth in a region. East Java is the province with the second-highest GRDP value after DKI Jakarta. The order under East Java is West Java, Central Java, and Riau [12]. Of the top five, the four provinces with the highest GRDP value are on the island of Java and are located close to each other. It indicates a neighbourhood effect or spatial effect.

In this study, modelling will be carried out to determine the spatial effect of the Gross Regional Domestic Product (GRDP) in East Java and the factors that influence it significantly by using the Spatial Durbin Error Model (SDEM).

2. Literature Review

2.1. Spatial Autocorrelation

Spatial autocorrelation occurs because of the similarity of characteristics that occur in adjacent locations [13]. The value of spatial autocorrelation will measure the proximity between locations and the similarity of characteristics to the location [14].

The spatial autocorrelation test using the Moran’s I test statistic, as follows [15]:

\[
I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sigma^2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}}
\]

\(w_{ij}\) : elements of the spatial weighted matrix locations i and j
\(x_i\) : attribute value at the location of concern (location i)
\(x_j\) : attribute value at the location of concern (location j)
\(\bar{x}\) : the average value of the attribute at the location i of the n observation location
\(n\) : the number of observation sites

2.2. Spatial Durbin Error Model

Spatial Durbin Error Model has a spatial lag effect on the independent variables and error. Spatial Durbin Error Model is expressed in the following equation [2],

\[
y = X\beta + WX\theta + u + \epsilon
\]

with \(u = \lambda W u + \epsilon\) and \(\epsilon \sim N(0, \sigma^2)

\(y\) : dependent variable vector
\(X\) : independent variable matrix
\(\beta\) : regression parameter vector
\(W\) : spatial weighting matrix
\(\theta\) : the vector of the independent variable spatial lag parameter
\(u\) : spatial error
\(\lambda\) : spatial coefficient of error
\(\epsilon\) : error vector of the model

3. Method
3.1. Types, Data Sources and Research Variables

The Spatial Durbin Model (SDM) data used in this study are secondary data obtained from the Central Bureau of Statistics of East Java in 2019. The data used is Gross Regional Domestic Product (GRDP) as a "variable, number of industries \((X_1)\), general allocation funds \((X_2)\), local revenue \((X_3)\), profit sharing fund \((X_4)\), open unemployment rate \((X_5)\), and labour force participation rate \((X_6)\). The software used in this study is R 4.0.2 software.

3.2. Data Analysis Steps

The steps taken are as follows:

a. Multicollinearity testing using the Variance Inflation Factor (VIF) value.
b. Spatial autocorrelation test using Moran's I test.
c. Modelling using SDEM.
d. Testing the model parameters and test the significance of model parameters.
e. Testing the model residuals obtained includes:
   1) Testing for normality uses the Jarque Bera test.
   2) The test for spatial heterogeneity used the Breusch Pagan test.
   3) Testing of spatial autocorrelation using Moran's I test.
f. Mapping and data interpretation.

4. Results and Discussion

4.1. Multicollinearity Detection

Multicollinearity detection can be seen from the Variance Inflation Factor (VIF) value. Based on the results of linear regression analysis, the VIF value was obtained, which is presented in Table 1.

| Variable | VIF value |
|----------|-----------|
| \(X_1\)  | 1.73      |
| \(X_2\)  | 1.26      |
| \(X_3\)  | 1.28      |
| \(X_4\)  | 1.12      |
| \(X_5\)  | 1.81      |
| \(X_6\)  | 1.25      |

Based on the results in Table 1, it is known that the VIF value of all independent variables is less than 10, so it can be concluded that there is no multicollinearity.

4.2. Spatial Autocorrelation Test

The autocorrelation test using Moran's I test statistics with the following hypotheses:

\[ H_0: I = 0 \quad \text{(no autocorrelation between locations)} \]
\[ H_A: I \neq 0 \quad \text{(there is autocorrelation between locations)} \]

where I is Moran's I index.

The results of Moran's I testing are presented in Table 2.

| Variable | I    | E(I)  | p-value   |
|----------|------|-------|-----------|
| error    | 0.1674 | -0.0270 | 0.0023    |
| \(Y\)    | 0.4412 | -0.0270 | 6.2361×10^{-5} |
| \(X_1\)  | 0.5852 | -0.0270 | 5.5087×10^{-7} |
| \(X_2\)  | -0.0462 | -0.0270 | 0.8819    |
| \(X_3\)  | 0.0764 | -0.0270 | 0.3623    |
| \(X_4\)  | -0.0391 | -0.0270 | 0.7162    |

The results of Moran's I testing are presented in Table 2.
Based on Table 2, the p-value of Moran's I on the error is less than the real level of 5%, so $H_0$ is rejected. These results indicate that there is spatial autocorrelation on the error. Besides, all variables are known to be positive and greater than the expected value so that the spatial pattern that is formed is a clustered pattern with positive spatial autocorrelation. As a result of the spatial autocorrelation, the use of multiple regression analysis is not appropriate, so the analysis using spatial regression analysis.

4.3. Spatial Durbin Error Model (SDEM)

Estimated parameters of the SDEM model are presented in Table 3.

| Variable | Coefficient | p-value |
|----------|-------------|---------|
| Intercept | 1.8997      | 0.8641  |
| $X_1$    | 0.4031      | 0.0295  |
| $X_2$    | 1.1332      | 0.0336  |
| $X_3$    | 0.3433      | 0.0184  |
| $X_4$    | 0.1217      | 0.0482  |
| $X_5$    | 0.0629      | 0.0187  |
| $X_6$    | 0.8442      | 0.0474  |
| $WX_1$   | 0.6546      | 0.0104  |
| $WX_2$   | 2.4845      | 0.0242  |
| $WX_3$   | -2.9501     | 7.622×10^{-5} |
| $WX_4$   | 0.1735      | 0.01072 |
| $WX_5$   | -0.0238     | 0.0099  |
| $WX_6$   | -0.7605     | 0.0089  |
| $\lambda$| -0.9021     | 0.0086  |

Based on the results in Table 3, it is known that the p-value on the spatial parameter error ($\lambda$) and several lags of the independent variables is significant, meaning that there is an effect of error and independent variables at neighboring locations ($j$) on the independent variables at the observation location ($i$). The estimation equation for the Spatial Durbin Error Model (SDEM) model that is formed is as follows:

$$
\hat{y}_i = 1.8997 - 0.9021 \sum_{j=1}^{38} w_{ij}u_j + 0.4031x_{1i} + 1.1332x_{2i} + 0.3433x_{3i} + 0.1217x_{4i} + 0.0629x_{5i} + 0.8442x_{6i} + 0.6546 \sum_{j=1}^{38} w_{ij}x_{1j} + 2.4845 \sum_{j=1}^{38} w_{ij}x_{2j} - 2.9501 \sum_{j=1}^{38} w_{ij}x_{3j} + 80.1735 \sum_{j=1}^{38} w_{ij}x_{4j} - 0.0238 \sum_{j=1}^{38} w_{ij}x_{5j} - 0.7605 \sum_{j=1}^{38} w_{ij}x_{6j} + \varepsilon_i.
$$

Based on the SDEM estimation results, it is known that direct and indirect variables influence the GDRP. Variables that have a direct and indirect effect are the number of industries, general allocation funds, local revenue, profit-sharing fund, open unemployment rate, and labour force participation rate.

4.4. Spatial Autocorrelation Test

The spatial autocorrelation test was carried out to determine whether the error in the SDEM model contained spatial interactions with the error model. The results of the Moran's I test on the error of the SDEM model can be seen in Table 4.

| I | E (I) | p-value |
|---|------|---------|

$x_5$ 0.2199 -0.0270 0.0572
$x_6$ 0.2319 -0.0270 0.0452
Based on Table 4, the p-value of Moran's I is more than the real level of 5%, so \( H_0 \) is accepted. These results indicate that there is no autocorrelation in error.

4.5. Spatial Heterogeneity Test

The spatial heterogeneity test is used to determine variance instability spatially. This test is performed using the Breuch-Pagan test in Table 5.

| \( BP \) | \( p\)-value |
|--------|-------------|
| 4.8637 | 0.3016      |

The p-value in the Breusch-Pagan test is known to be greater than the real level of 5%, so \( H_0 \) is accepted. These results indicate that there is no spatial heterogeneity in error.

4.6. Normality Test

The normality test was carried out with the Jarque Bera test statistic. The test results can be seen in Table 6.

| \( JB \) | \( p\)-value |
|--------|-------------|
| 5.6365 | 0.05971     |

The p-value in the Jarque Bera test is known to be greater than the real level of 5%, so \( H_0 \) is accepted. These results indicate that the error in the SDEM model is normally distributed.

4.7. Mapping

Based on the results of the analysis with the Robust SEM approach, the predictive value of MSEs is divided into five classes where the classification can be seen in Table 7.

| Classification | Category | Colour |
|----------------|----------|--------|
| Class 1        | Very high| Red    |
| Class 2        | High     | Orange |
| Class 3        | Low      | Green  |
| Class 4        | Very low | Blue   |

The results of the prediction mapping with SDEM model of 38 cities/regencies in East Java are presented in Figure 1 below.
5. Conclusion

Based on the analysis, it was found that the variable number of industries, general allocation funds, local revenue, profit-sharing fund, open unemployment rate, and labour force participation rate had a significant positive impact on GRDP.

Based on the prediction level of GRDP obtained by the SDEM model approach, there are three cities/regencies that belong to the "very high" category because the three cities/regencies have a large number of industries, general allocation funds, and profit-sharing funds compared to other cities/regencies. There are four cities/regencies that belong to the "high" category because Pasuruan Regency has a large number of industries and these four cities/regencies have quite large local revenue and profit-sharing funds. Ten regencies belong to the "low" category because the ten cities have sufficient or relatively low scores for some of their independent variables. There are 21 cities/regencies in the "very low" category. These cities/regencies have relatively low value for each of the independent variables.

References

[1] Anselin L 1988 Spatial Econometrics: Methods and Models (London: Kluwer Academics Publishers)
[2] LeSage J P and Pace R K 2009 Introduction to Spatial Econometrics (London: Chapman & Hall)
[3] LeSage J and Pace R K 2008 Introduction to Spatial Econometrics (Boca Raton, London and New York: CRC Press)
[4] Karim A and Wasono R 2017 Adv Sci Lett 23(7) 6530-6532
[5] Septiawan A R, Handajani S S, and Martini T S 2018 J Phys Conf Ser 1025 index in Province of Central Java
[6] Han X and Lee L 2013 Reg Sci Urban Econ 43 816-837
[7] Lacombe D J, Halloway G J, and Shaughnessy T M 2014 Int Reg Sci Rev 37(3) 247-250
[8] Jiang L, Folmer H, Ji M and Tang J 2017 The Ann Reg Sci 58 301-319
[9] Li J and Han X 2019 Reg Sci Urban Econ 77 38-74
[10] Yinqi L and Baizhou L 2017 Sci Technol Prog Policy 3
[11] Rapanna P and Sukarno Z 2017 Ekonomi Pembangunan (Makassar: CV Sah Media)
[12] Badan Pusat Statistika 2018 Data Produk Domestik Regional Bruto.
[13] Lee J and Wong D W S 2001 Statistical Analysis ArcView GIS (New York: John Wiley & Sons, Inc)
[14] Kosfeld R 2006 Spatial Econometrics. Institute of Economics (Germany: University of Kassel)
[15] Wang B, Zhang Y, Zhan C, and Yang X 2020 *J Clean Prod* 270 122598
[16] Badan Informasi Geospasial 2017 Indonesia Geospasial Portal: File SHP Wilayah Jawa Timur