International Journal of Scientific Research and Engineering Development-– Volume 7 Issue 6, Nov- Dec 2024
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Modeling Life Expectancy in Indonesia Using Spatial Error Regression with Queen Contiguity Weighting

Talitha Zuleika[1], Suliyanto[2], Dita Amelia[3], Sediono[4], M. Fariz Fadillah Mardianto[5]
[1,2,3,4,5]Statistics Study program, Departement of Mathematics, Airlangga University, Surabaya-Indonesia
[1]talitha.zuleika-2021@fst.unair.ac.id
[2]Corresponding author:suliyanto@fst.unair.ac.id
[3]dita.amelia@fst.unair.ac.id
[4]ahmad.s@fst.unair.ac.id
[5]m.farizfadillah.m@fst.unair.ac.id

----------**********----------

Abstract:

Health can be interpreted as an asset for each individual with a high intrinsic value for public welfare. Life expectancy is an important indicator that can be used to measure the welfare of Indonesian people in general and increase the degree of public health in particular. If an area has a low life expectancy, then there is a need for health development programs and other social programs. This is in line with one of the goals of Sustainable Development Goals (SDGs), which is on point 3 concerning healthy life and welfare for all ages. This study will model life expectancy in Indonesia using the spatial error regression method with the weight of queen contiguity. Based on the analysis carried out, the AIC value and determination coefficient, respectively, of 135.526 and 67.944%, are obtained in modeling life expectancy in Indonesia. In addition, the variables that significantly affect life expectancy in Indonesia include literacy rate variable, percentage of population under two years old who have been breastfed, and percentage of toddlers who have received complete immunization. Through spatial analysis, efforts to improve policies and public health must continue to be a priority to increase life expectancy in Indonesia by paying attention to relationships and variations between regions. Maintaining and increasing life expectancy not only improves individual quality of life but also has a positive impact on a country's progress.

Keywords — **Life Expectancy, Classical Regression, Spatial Lag Regression, Queen Contiguity.**

----------**********----------

I. INTRODUCTION

One of the tools that the government can use to measure the general welfare of the Indonesian people and specifically improve public health is life expectancy. Life expectancy at birth is an average estimate of the number of years a baby will pass from birth in a given year [1]. The countries with the highest to lowest life expectancy are developed, developing, and underdeveloped [2]. This illustrates that a country's high life expectancy is affected by the country's socioeconomic conditions. The higher the value of a region's life expectancy, the more successful the region is in developing in the health sector. The government has made various efforts to increase life expectancy in Indonesia by improving the health service infrastructure. This is done to achieve the Sustainable Development Goals (SDGs), which are on the third point regarding healthy life and promoting welfare for all ages.

The high and low life expectancy cannot be separated from many influencing factors, such as sociodemographic factors, macroeconomics, and health resources [3]. One method for analyzing the factors affecting life expectancy is regression analysis. Regression analysis is a statistical method for determining the relationship between a response variable and a predictor variable [4]. In general, regression analysis only systematically models the relationship of variables without considering spatial

dependencies (locations) throughout the region so that the influence of variables is considered the same in all locations. A statistical method applied to model relationships between variables by considering spatial aspects or places is called spatial regression [5]. There are several models in spatial regression, one of which is the spatial error regression model. Model spatial error regression occurs when there is a dependency between the error value at one location and the error value at another adjacent location [5].

Several previous studies related to modeling life expectancy in Indonesia using a spatial approach by comparing the Quantile Regression (QR) and Geographically Weighted Regression (GWR) methods [6]. The conclusion found in the study was that the GWR method had a smaller RMSE value than the QR method in each quantile [6]. Furthermore, there is another study that models life expectancy in Indonesia using the GWR method [7]. Based on the study, it can be found that the GWR method provides better results in modeling life expectancy when compared to the multiple linear regression method [7].

Life expectancy in Indonesia has a heterogeneous distribution and high disparity and contains several outliers according to existing conditions [6]. Another study said that the percentage of people under two years of age who had been breastfed and the percentage of people under five who had been fully immunized had a significant influence on life expectancy [1]. In addition, research using a panel data regression model states that high life expectancy can be influenced by the high Gross Regional Domestic Product (GDP) [8]. In addition, some studies say that the factors that significantly affect life expectancy are literacy rates and average length of schooling [9].

From the description, the researchers are interested in modeling life expectancy in Indonesia using the spatial error regression method using queen contiguity weighting. This study is expected to provide information and knowledge about life expectancy in Indonesia based on factors suspected to be influential. In addition, the research is expected to serve as an overview to the government in conducting a policy evaluation on life expectancy in Indonesia.

II. RESEARCH METHODOLOGY

This study uses secondary data on life expectancy in Indonesia and factors suspected to affect it in 2022. The data used was obtained through the official website of the Indonesian Central Agency of Statistics (BPS). The observations used in this study consisted of 34 provinces in Indonesia, each with a code given for each province.

This study used six variables divided into one response variable (Y) and five predictor variables (X). All research variables are presented in Table 1.

TABLE 1
RESEARCH VARIABLES

Variables	Information	Scale	Variable Type
Y	Life Expectancy (Year)	Ratio	Continuous
X_1	Literacy rate (%)	Ratio	Continuous
X_2	Percentage of Population Under Two Years Old Who Have Been Breastfed (%)	Ratio	Continuous
X_3	GRDP on The Basis of Applicable Prices (IDR)	Ratio	Continuous
X_4	Percentage of Toddlers Who Have Received Complete Immunization (%)	Ratio	Continuous
X_5	Average Length of School for Residents 15 Years and Above (Year)	Ratio	Continuous

The data analysis method used in this study is spatial error regression, using queen contiguity as a weighting. The classical regression model can be written in the following form [4].

$$\begin{cases} \boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u} \\ \boldsymbol{u} = \lambda \boldsymbol{W u} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}) \end{cases} \#(1)$$

The Lagrange Multiplier (LM) test, designed to identify the impact of spatial dependence, will be employed to evaluate the appropriateness of the spatial error regression model [10]. Here are the test statistics used.

$$LM_{error} = \left[\frac{\hat{\boldsymbol{\varepsilon}}' \boldsymbol{W} \hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}/n}\right]^2 / \left[tr\left(\boldsymbol{W}^2 + \boldsymbol{W}'\boldsymbol{W}\right)\right] \sim \chi^2_{(1)} \#(2)$$

Further estimation of the error spatial regression model can be written into the following equation.

$$\hat{\boldsymbol{u}} = \hat{\lambda} \boldsymbol{W} \hat{\boldsymbol{u}} + \boldsymbol{\varepsilon}$$

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} \#(3)$$

with $\hat{\boldsymbol{\beta}} = (\boldsymbol{X'B'BX})^{-1}\boldsymbol{X'B'By}$ and $\hat{\boldsymbol{B}} = I - \hat{\lambda}\boldsymbol{W}$.

Data analysis was conducted using the GeoDa software assistance with the following steps.

1. Describes the variables of life expectancy in Indonesia by using thematic maps and conducting descriptive analysis of all research variables.
2. Conduct multicollinearity test.
3. Performing a spatial effect test consisting of Moran's I test and Breusch Pagan test.
4. Modeling the life expectancy in Indonesia using the spatial error regression method.
   - Test the fit of the spatial error regression model.
   - Estimating the parameters of the spatial model.
   - Perform individual tests of model parameters and autoregressive parameter significance tests of a spatial error regression model.
5. Interpret the model and draw conclusions.

III. RESULTS AND DISCUSSION

A. Descriptive Statistics

Descriptions of the characteristics of the life expectancy variable (Y) and the predictors that may affect it include literacy rate (X_1), percentage of population under two years old who have been breastfed (X_2), GDP (X_3), percentage of toddlers who have received complete immunization (X_4), and average length of school for residents 15 years and above (X_5) is presented in Table 2 using descriptive statistics.

TABLE 2
DESCRIPTIVE STATISTICS OF RESEARCH VARIABLES

Variabel	Mean	Variance	Minimum	Maximum
Y	70.449	6.015	65.670	75.105
X_1	96.687	13.508	81.190	99.810
X_2	10.382	0.255	9.240	11.420
X_3	76521	3653617340	21718	298360
X_4	57.23	105.94	21.21	72.93
X_5	9.247	0.668	7.310	11.300

Furthermore, the distribution of life expectancy in Indonesia can be known using GeoDa *software* through the following thematic maps.

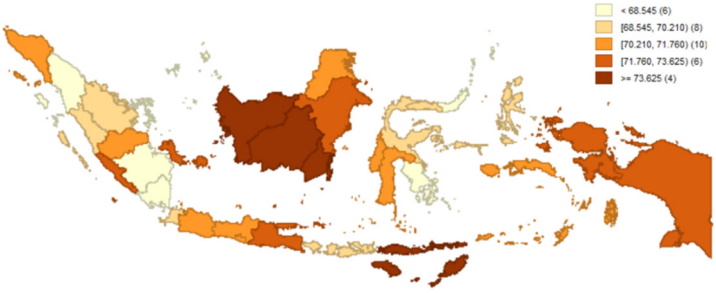

Fig. 1 Thematic Map of Life Expectancy in Indonesia in 2022

B. Multicollinearity Test

Multicollinearity tests were conducted using the calculation of Variance Inflation Factor (VIF) dan $Tolerance$ values. If the value of $VIF < 10$ and the value of $Tolerance > 0.1$, then the regression model is free from multicollinearity [11]. The results of the multicollinearity test for all predictor variables are shown below.

TABLE 3
VIF VALUE AND TOLERANCE OF ALL PREDICTOR VARIABLES

Variable	VIF	$Tolerance$
X_1	1.95	0.513
X_2	1.22	0.819
X_3	1.46	0.685
X_4	1.09	0.917
X_5	2.28	0.439

According to Table 3, all predictor variables have a VIF value of less than 10 and a $Tolerance$ value of more than 0.1. Thus, it can be concluded that there is no case of multicollinearity in the predictor variables.

C. Spatial Effect Test

Two spatial effect tests will be carried out: the Moran's I test and the Breusch Pagan test. Moran's I test is used to detect spatial autocorrelation globally [12] and the Breusch Pagan test is used to determine whether there is spatial heterogeneity or diversity between regions in the data [13]. Both tests were conducted using GeoDa software assistance. The table below presents the test results based on the analysis conducted.

TABLE 4
SPATIAL EFFECT TEST RESULTS

Test	$P-Value$
Moran's I Test	0.00037
Breusch Pagan Test	0.02776

According to Table 4, it can be seen that Moran's test I has a $p-value$ of 0.00037. The decision made in the test was to reject H_o with an α of 0.05. It can be concluded that life expectancy in Indonesia exhibits spatial autocorrelation. Then, in Table 4, the $p-value$ of the Pagan Breusch test was found to be 0.02776, so the test resulted in the decision to reject H_o with an α of 0.05. Therefore, it can be concluded that there is spatial heterogeneity in the data.

D. Modeling Life Expectancy in Indonesia with Spatial Error Regression Weighted Queen Contiguitiy

Before modeling using spatial error regression, a model suitability test will be carried out using the Lagrange Multiplier (LM) test, with the test criteria being rejected H_0 if the value of $p-value < \alpha = 5\%$. Based on the analysis that has been carried out, a $p-value$ of 0.00151 is obtained. Thus, the test results in a decision to reject H_0, so it can be concluded that the spatial error regression model is appropriate.

The results of estimating the parameters of the spatial error regression model are presented in Table 5 below.

TABLE 5
RESULTS OF ESTIMATION OF SPATIAL ERROR REGRESSION MODEL PARAMETERS

Parameter	Estimated Value
$\hat{\lambda}$	0.607788
$\hat{\beta}_0$	4.42898
$\hat{\beta}_1$	0.361438
$\hat{\beta}_2$	2.39855
$\hat{\beta}_3$	0.0000082
$\hat{\beta}_4$	0.084453
$\hat{\beta}_5$	0.0810301

Based on Table 5, the estimator for the spatial error regression model of life expectancy in Indonesia can be expressed by the following equation.

$$\hat{Y}_i = 4.42898 + 0.361438\,X_{1i} + 2.39855\,X_{2i} + 0.0000082\,X_{3i} + 0.084453\,X_{4i} + 0.0810301\,X_{5i};\ i = 1,2,3,\ldots,34 \ \#(4)$$

Subsequently, a partial test of the erroe spatial regression model will be carried out. The critical area used in this test is rejected H_0 when the $p-value < \alpha = 5\%$. The following is the result of a partial test of the error spatial regression model parameters in Table 6.

TABLE 6
RESULTS OF PARTIAL TEST OF SPATIAL ERROR REGRESSION MODEL PARAMETERS

Parameter	$P-Value$	Decisions
$\hat{\beta}_0$	0.68973	-
$\hat{\beta}_1$	0.00001	Reject H_0
$\hat{\beta}_2$	0.00001	Reject H_0
$\hat{\beta}_3$	0.08079	Accept H_0
$\hat{\beta}_4$	0.00227	Reject H_0
$\hat{\beta}_5$	0.83651	Accept H_0

According to Table 6, variables influencing the decision to reject H_o and produce conclusions had a partially significant effect on the response variable consisting of the literacy rate variable (X_1), percentage of population under two years old who have been breastfed (X_2), and percentage of toddlers who have received complete immunization (X_4).

After a partial test of the model, the next stage is to conduct an autoregressive parameter significance test on the spatial model. This will be carried out using the Likelihood Ratio Test (LRT). The test criteria used are rejected H_0, if the value of $p-value < \alpha = 5\%$. The analysis yielded a p-value of 0.00026. Thus, the test results in a decision to reject H_0 and it can be inferred that the autoregressive parameters within the spatial error regression model are significant.

E. Goodness of Fit Test

The goodness of fit test on the spatial error regression model will use the value of R^2 dan AIC. A value of R^2 was obtained of 0.67944 which shows that all predictor variables can simultaneously explain the life expectancy of 67.944%. In addition, an AIC value of 135.526.

F. Residual Normality Assumption Test

The Kolmogorov-Smirnov test will be used to assess residual normality in the spatial error regression model. The following hypotheses are applied to determine whether the residuals of the

spatial error regression model are normally distributed.

H_o: normally distributed residual

H_1: residual is not normally distributed

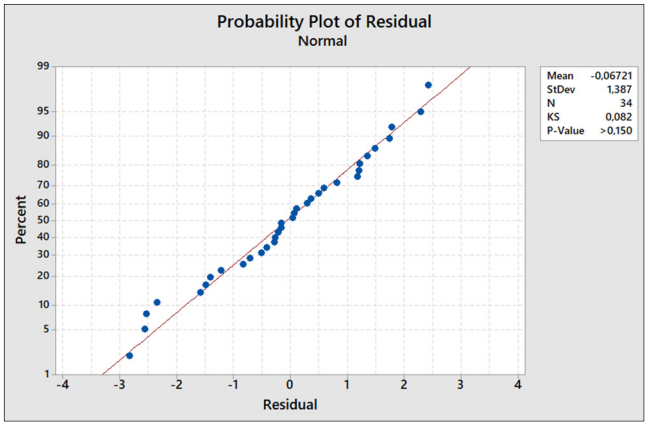

Fig. 2 Residual Normality Test Results Plot

According to Fig. 2, a $p-value$ greater than 0.150 was obtained. With an α value of 0.05, the test results in the decision to accept H_o which concludes that the residual classical regression model is normally distributed.

G. Spatial Error Regression Model Interpretation

Based on the spatial regression error model, it is possible to model life expectancy in West Sulawesi Province, the province with the lowest life expectancy value in Indonesia in 2022. Here is a calculation that can be used.

$$\hat{Y}_{30} = 4.42898 + 0.361438\,X_{1,30} + 2.39855\,X_{2,30} + 0.0000082\,X_{3,30} + 0.084453\,X_{4,30} + 0.0810301\,X_{5,30}\ \#(5)$$

The life expectancy in West Sulawesi Province was 65.67 years, with a literacy rate of 93.82%, a percentage of the population under two years old who have been breastfed 9.94%, a GDP of 37070 thousand rupiahs, a percentage of toddlers who have received complete immunization 53.67%, and an average school age of 9.94 years.

Then, there can also be a life expectancy modeling in Yogyakarta Province, the province with the highest life expectancy value in Indonesia in 2022. Here is a calculation that can be used.

$$\hat{Y}_{14} = 4.42898 + 0.361438\,X_{1,14} + 2.39855\,X_{2,14} + 0.0000082\,X_{3,14} + 0.084453\,X_{4,14} + 0.0810301\,X_{5,14}\ \#(6)$$

The life expectancy in Yogyakarta Province was 75.105 years, with a literacy rate of 95.15%, a percentage of the population under two years old who have been breastfed 10.82%, a GDP of 44045 thousand rupiahs, a percentage of toddlers who have received complete immunization 65.09%, and an average school age of 10.07 years.

Through the partial test of the spatial error regression model, three variables affect the life expectancy rate in Indonesia, including the literacy rate variable (X_1), percentage of population under two years old who have been breastfed (X_2), and percentage of toddlers who have received complete immunization (X_4) presented in the following equation.

$$\hat{Y}_i = 4.42898 + 0.361438\,X_{1i} + 2.39855\,X_{2i} + 0.084453\,X_{4i}\ \#(5)$$

Through this equation, if the literacy rate in *i*-province increases by 1%, it will result in the life expectancy rate in *i*-province will increase by 0.351438%. Then, if the percentage of population under two years old who have been breastfed in the *i*-province has increased by 1%, this will result in the life expectancy in the province having increased by 2.3985%. Furthermore, if the percentage of toddlers who have received complete immunization in the *i*-province increases by 1%, then the life expectancy of the *i*-province will increase by 0.084453%.

IV. CONCLUSIONS

Based on the research that has been conducted, it can be concluded that modeling life expectancy in Indonesia using spatial regression error results in an AIC value of 135.526 and the resulting coefficient of determination of 67.944%. Moreover, it can be observed that three predictor variables have a significant impact on life expectancy in Indonesia, namely the literacy rate variable (X_1), percentage of population under two years old who have been

breastfed (X_2), and percentage of toddlers who have received complete immunization (X_4).

According to the findings from the research conducted, researchers recommend that the government pay attention to spatial dependence between regions in Indonesia in developing targeted strategies to increase life expectancy in Indonesia based on the factors that affect it. This strategy can include infrastructure development in the education and health sector and improving the socialization of healthy living among Indonesian people. In addition, further research can be conducted on life expectancy in Indonesia by adding other predictor variables suspected to affect life expectancy.

ACKNOWLEDGMENT

With humility, the author would like to thank Mr. Suliyanto, Mrs. Dita Amelia, Mr. Sediono, and Mr. M. Fariz Fadillah Mardianto for guiding and directing the author in completing this research. The author also thanked all parties who had provided support both directly and indirectly.

REFERENCES

[1] A. P. Sugiantari and I. N. Budiantara, "Analisis Faktor-faktor yang mempengaruhi angka harapan hidup di Jawa Timur menggunakan Regresi Semiparametrik Spline," *Jurnal sains dan Seni ITS,* pp. D37-D41, 2013.

[2] E. Anggraini and U. Lisyaningsih, "Disparitas Spasial Angka Harapan Hidup di Indonesia Tahun 2010," *Jurnal Bumi Indonesia,* 2013.

[3] A. A. B. Wirayuda and M. F. Chan, "A Systematic Review of Sociodemographic, Macroeconomic, and Health Resources Factors on Life Expectancy," *Asia Pacific Journal of Public Health,* p. 335–35, 2021.

[4] A. C. Rancher and G. B. Schaalje, Linear Model in Statistics, United States of America: John Wiley and Sons, 2008.

[5] L. Anselin, Spatial Econometrics: Methods and Models, Dordrecht: Kluwer Academic Publisher, 1988.

[6] R. S. Irevanie, "Perbandingan Metode Quantile Regression (QR) dan Geographically Weighted Regression (GWR) Pada Data Angka Harapan Hidup di Indonesia," Institut Teknologi Sepuluh, Fakultas Matematika dan Ilmu Pengetahuan Alam, Surabaya, 2017.

[7] N. T. Suhesti, "Analisis Spasial Angka Harapan Hidup di Indonesia Tahun 2016 (Doctoral dissertation, Program Studi Statistika Program Diploma IV)," 2018.

[8] T. M. Crisdyastuti, "Pengaruh Pendapatan Per Kapita, Akses Air Bersih, dan Tingkat Pendidikan Terhadap Angka Harapan Hidup di Jawa Timur Tahun 2002-2007," Universitas Airlangga, Surabaya, 2010.

[9] N. W. D. Ayuni, "Pemodelan Angka Harapan Hidup Di Provinsi Jawa Timur Tahun 2007 Dan 2011 Berdasarkan Angka Melek Huruf, Rata Rata Lama Sekolah, Dan Pengeluaran Perkapita," *Jurnal Matematika,* pp. 12-23, 2013.

[10] I. F. Fatati, H. Wijayanto and A. M. Sholeh, "Analisis Regresi Spasial dan Pola Penyebaran pada Kasus Demam Berdarah Dengue (DBD) di Provinsi Jawa Tengah," *Media Statistika,* pp. 95-105, 2017.

[11] E. Yaldi, J. K. P. Pasaribu, E. Suratno, M. Kadar, Gunardi, R. Naibaho, S. K. Hati and V. Aryati, "Penerapan Uji Multikolinieritas dalam Penelitian Manajemen Sumber Daya Manusia," *Jurnal Ilmiah Manajemen Dan Kewirausahaan (JUMANAGE),* pp. 94-102, 2022.

[12] T. Wuryandari, A. Hoyyi, D. S. Kusumawardani and D. Rahmawati, "Identifikasi autokorelasi spasial pada jumlahpengangguran di jawa tengah menggunakan indeks moran," *Media Statistika,* pp. 1-10, 2014.

[13] A. Arif, M. A. Tiro and M. Nusrang, "Perbandingan Matriks Pembobot Spasial Optimum dalam Spatial Error Model (SEM)(Kasus: Indeks Pembangunan Manusia Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2015)," *VARIANSI: Journal of Statistics and Its application on Teaching and Research,* pp. 66-76, 2019.

