Vol. 10, No.1: 20 — 25
May 2021
DOI: 10.22487/25411969.2021.v10.i1.15459

Natural Science
Journal of Science and Technology

Original article

Modelling life expectancy in Sulawesi Island by using spatial autoregressive (SAR)

Rika Yunita Pratiwi*, Nur'eni and Iman Setiawan

Department of Statistic, Faculty of Mathematics and Natural Sciences, Tadulako University, Jalan Raya Soekarno–Hatta, Tondo, Palu, 94117, Central Sulawesi, Indonesia.

Keywords: Life Expectancy, SAR, AIC

Article history:
Received 09 March 2021
Accepted 29 May 2021
Published 31 May 2021

* Coresponding Author :
rikayunita130698@gmail.com

Abstract

Life expectancy, LE, is the average estimate of the many years a person takes during life. LE in one region is different from other regions depending on the quality of life that can be achieved by its inhabitants. This research uses the Spatial Autoregressive (SAR) method. SAR model is a linear regression model where the response variable has a spatial correlation. The information of the SAR model due to the dependence of the value of observation in an area with other regions that are interconnected. The results were obtained from the value of the goodness of the model based on the smallest Akaike Information Criterion (AIC) SAR value is 328.24. The influencing factors that life expectancy are mean years of schooling (X_1), poor people (X_2), and labour force participation rate (X_5) which is obtained from the < 0.05 probability value.

INTRODUCTION

Population development both in quality and quantity is influenced by many factors. The quality of life of a country and region reflects the welfare of the people and the success of programs made by the government to improve the degree of human life. Related to quality of life there is an element of life expectancy in it. Life expectancy is also called the length of human life in the world. Life expectancy becomes an important benchmark in population as one indicator in measuring population quality. High and low life expectancy can be considered in describing the socio-economic progress of the community (Supriatna et al, 2006).

Life expectancy is the average number of years a person will live after the person reaches his xth birthday (BPS, 2013). The higher life expectancy in a region shows the higher quality of health in an area. Sulawesi Island is an island in the territory of Indonesia which is located between Kalimantan Island in the west separated by the Makassar Strait and the Maluku Islands in the east separated by the Maluku Sea, with an area of 174,600 km². According to BPS, in 2017 life expectancy in Indonesia as a whole was 71.06 years, an increase compared to 2016, which was 70.90 years, only an increase of 0.16 but showed an increase every year of life expectancy in Indonesia. In 2016, life expectancy on Sulawesi Island was 68.34 years and life expectancy in 2017 was 68.36 years, which means increasing life expectancy from year to year shows that the achievement of human development has been successfully carried out.

A regression modeling analysis to find out the factors that influence life expectancy that is influenced by regional characteristics is very important. In some cases, the dependent variable observed has a relationship with the results of observations in different regions, especially adjacent areas. The existence of spatial relationships in the dependent variable will cause the estimator to be incorrect because the error randomness assumption is violated. In overcoming these problems a regression model is needed that includes spatial relationships between regions into the model, so the model used is a spatial regression model. The fundamental component of the spatial model is the spatial weighting matrix, this matrix reflects the

relationship between one region and another (Arbia, 2005). One of the Spatial modeling is Spatial Autoregressive Model (SAR). The Spatial Autoregressive Model (SAR) is a linear regression model whose response variables are spatial correlations. This method is very important to get information about the pattern of distribution of the characteristics of an area and the relationship between locations within it. In addition, this method is also used to identify spatial modeling.

Previous research on Spatial Autoregressive (SAR) was conducted by Komang Kokom Sucahyati et al (2017) who used Spatial Autoregressive (SAR) in modeling community MCK behavior in rivers in the Blahbatuh District in 2016. The results of the study were that the SAR model obtained was well used To explain the community who used MCK behavior in the river, in Blahbatuh District in 2016, another study was conducted by Ni Wayan Dewinta Ayuni (2015) about modeling life expectancy in East Java Province using the Feed Forward Neural Network (FNN) method. The results obtained are the FFNN method produces the smallest MSEP value so it can be concluded that the modeling of life expectancy (AHH) in East Java province is best using the FFNN method.

Some tests in the Spatial Autoregressive Model one of which is the Moran index. This method is very important to get information about the pattern of distribution of the characteristics of an area and the relationship between locations within it. In addition, this method is also used to identify spatial modeling. In this study the authors will model life expectancy on Sulawesi Island using Spatial Autoregressive (SAR) to see spatial factors on life expectancy (AHH) on Sulawesi Island in 2017, so that the model obtained can be an evaluation and consideration material in compiling policy which is more effective.

MATERIALS AND METHODS

The data used in this study are secondary data obtained from Badan Pusat Statistik (BPS) in each province on Sulawesi Island in 2017 with observation units of 81 regencies/cities in Sulawesi with 6 variables as follows :

X1 : Mean Year Schooling
X2 : Poor People
X3 : Expected Year Schooling
X4 : Unemployment Rate
X5 : Labour Force Participation Rate
X6 : Population Growth Rate

The method that used in this study was Spatial Autoregressive (SAR) analysis, with the following stages of analysis carried out are as follows: 1). Retrieval of life expectancy data and 6 variables studied, 2). Conduct descriptive analysis on life expectancy data and 6 variables studied, 3). Perform multiple linear regression modeling on life expectancy data, 4). Testing classic assumptions (normality using the Shapiro-Wilk test, multicollinearity using the VIF test, and heteroscedasticity using the Breusch-Pagan test), if the violation will be transformed, 5). Test the assumption of autocorrelation (autocorrelation using the Durbin Watson test), 6). Creating a distance inverse weighting matrix, with the distance used is the value of longitude and latitude of each region, 7). Test the spatial effect of the response variable (Y) with the Moran's index test, 8). Test the effects of spatial dependency by using the Lagrange Multiplier (LM) test, 9). Perform spatial regression modeling with the Spatial Autoregressive (SAR) method on life expectancy data, 10). Test the assumption of the SAR model (normality using the Shapiro-Wilk test, heteroscedasticity using the Breusch-Pagan test, and autocorrelation using the Durbin Watson test), 11). Determine the AIC value, and 12). Interpretation.

RESULT AND DISCUSSION

The variable of life expectancy (Y) has the lowest value is 62.86 years which is precisely in the region of Pokuwato Regency and the highest value which is precisely in the region of 73.02 years which is precisely in the area of Kendari City with a mean of 68.42, standard deviation of 2.33 , and variance of 5.46. Mean Year Schooling (X_1) has the lowest value of 5.98 years, which is precisely in the region of Jeneponto Regency, the highest value of 11.26 years precisely in the Palu City region shows that the average length of school is Palu City with a mean of 7.97, standard deviation of 1.03, and variance of 1.07. Poor people (X_2) has the lowest value of 4.59% which is precisely in the Makassar City area, the highest value of 21.85% which is precisely in the area of Boalemo Regency, with a mean of 2.09, a standard deviation of 4.47, and a variance of 19.98. The following are descriptive statistics of all variables used:

Table 1. Descriptive Statistics of The Variables Used

Var	Min Value	Mak Value	Mean	Std. Deviasi	Varians
Y	62.86	73.02	68.42	2.33	5.46
X_1	5.98	11.26	7.97	1.03	1.07
X_2	4.59	21.85	2.09	4.47	19.98
X_3	11.16	14.79	12.6	0.77	0.59
X_4	0.47	9.85	4.29	2.12	4.50
X_5	49.17	81.06	64.89	6.01	36.20
X_6	0.02	3.22	1.37	0.74	0.55

Expectancy year schooling (X_3) has the lowest value of 11.16 years, precisely in the region of Bolaang Mongondow Regency, the highest value of 14.79 years, which is exactly what in the Bau-Bau City area, with a mean of 12.6, a standard deviation of 0.77, and a variance of 0.59. Unemployment Rate (X_4) has the lowest value of 0.47% which is precisely in the area of Bombana Regency, the highest value of 9.85 which is precisely in the Bitung City region, with a mean of 4.29, a standard deviation of 2.12, and a variance of 4.50. Labour force participation rate (X_5) has the lowest value of 49.17% which is precisely in the Sindreng Rappang Regency area, the highest value of 81.06% which is precisely in the East Kolaka Regency area, with a mean of 64.89, a standard deviation of 6.01, and a variance of 36.20. Population Growth Rate (X_6) has the lowest value of 0.02% which is precisely in the area of North Luwu Regency, the highest value of 3.22% which is precisely in the area of Bombana Regency, with a mean of 1.37, a standard deviation of 0.74, and a variance of 0.55.

Modeling of Multiple Linear Regression

$$Y = 49.056 + 0.658\,X_1 - 0.185X_2 + 0.453X_3 + 0.160\,X_4 + 0.156\,X_5 - 0.145\,X_6 + \varepsilon_i$$

Classical Assumption Test

Normality Test

The Shapiro-Wilk value of 0.1083 is obtained and the significance value of $0.1083 > \alpha = 0.05$ so that H_0 is accepted. This shows that the data used in the study were normally distributed.

Table 2. Shapiro-Wilk Normality Test Results

Test	Value Shapiro-Wilk	p-value
Shapiro-Wilk	0.97467	0.1083

Multikolinearity Test

All independent variables (X) in the study have a Variance Inflation Factor $(VIF) < 10$, so it can be concluded that there is no multicollinearity in the independent variables used. Therefore, all of these independent variables can be used in modeling using Spatial Autoregressive (SAR) regression

Table 3. Variance Inflation Factor (VIF) Test Results

Var	X_1	X_2	X_3	X_4	X_5	X_6
VIF	1.650	1.445	1.139	1.523	1.352	1.239

Heteroskedasticity Test

Based on table 4, obtained a free degree of 6, the p-value of $0.4683 > \alpha = 0.05$ so that H_0 is accepted and the BP value is 5.609. This shows that there is no heteroscedasticity in the model, meaning that the data in the study are homogeneous.

Table 4. Breusch-PaganTest Results

Test	DB	Value BP	p-value
Breusch-Pagan	6	5.609	0.4683

Fulfillment of Autocorrelation Test

Based on table 5, a DW value of 1,466 is obtained, while from a DW table with a significant level of 0.05 a dL value of 1.4842 is obtained, and a dU of 1.8008. Decision making for autocorrelation can be seen in table 1. Based on the table the DW value is in the area between $0 < DW < dL$, so reject H_0 which means accept H_1, then there is a positive autocorrelation.

Table 5. Durbin WatsonTest Results

Test	Value DW
Durbin Watson	1.466

Spatial Weighting Matrix

In this study, the approach used to determine the spatial weighting matrix is the distance function. This distance is flexible enough to be modified in overcoming data weaknesses. Spatial weighting matrix symbolized by W denotes the relationship of the observation area n x n.. W_{ij} is the element of the matrix **W** in the -*i* row of the -*j* column *i,j* = 1,2, ...,.The weighting matrix **W** with order 81 x 81 calculated by using R software are in the following:

$$\mathbf{W} = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & \cdots & 81 \end{matrix} \\ \begin{bmatrix} 0.0000000 & 0.5810911 & 2.5376279 & \cdots & 0.1617132 \\ 0.5810911 & 0.0000000 & 0.6729718 & \cdots & 2.3328748 \\ 2.5376279 & 0.6729718 & 0.0000000 & \cdots & 0.8261699 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0.1617132 & 2.3328748 & 0.8261699 & \cdots & 0.0000000 \end{bmatrix} \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ 81 \end{matrix} \end{array}$$

Manhattan distance is one of the formulas for calculating the shortest distance between two points (Budianto, 2011). Manhattan distance is used to calculate the absolute difference between the coordinates of a pair of objects. Called Manhattan, it is based on the city of Manhattan which is arranged into blocks, so it is often called a city block distance or absolute value distance. The **W** matrix above is the **W** matrix in the regions of South Bolaang Mongondow, North Bolaang Mongondow and Kotamobago districts. For example, a sample of Manhattan distances between two points, namely South Bolaang Mongondow Regency and North Bolaang Mongondow District can be seen as follows:

$$W_{12} = \frac{1}{(d_{12})^2}$$
$$= \frac{1}{(|x_2-x_1|+|y_2-y_1|)^2}$$
$$= \frac{1}{\left(\begin{matrix}|0.383658-0.906708|+\\ |124.0479-123.2591|\end{matrix}\right)^2}$$
$$= \frac{1}{(|-0.52305|+|0.788781|)^2}$$
$$= \frac{1}{(|0.52305+0.788781|)^2}$$
$$= \frac{1}{(1.31183)^2}$$
$$= \frac{1}{1.720900}$$
$$= 0.58109$$

Moran's I Test

Based on table 6, the $\text{p} - \text{value}\, 0.0001 < \alpha = 0.05$ so that H_0 is rejected and the statistical value of the moran's I is 0.210324. This shows that there is spatial dependence, so that spatial regression will be used to model life expectancy on Sulawesi Island.

Table 6. Moran's I Test Results

Test	Moran's I Statistic	P-value
Moran's I	0.210324	0.0001

Lagrange Multiplier (LM) Test

The Lagrange Multiplier (LM) test results for the Spatial Autoregressive Model (SAR) obtained a p-value of $0.0006 < \alpha = 0.05$ so that H_0 is rejected and the degree of freedom is 1. This shows that there is a spatial lag dependency so needs to be continued in the formation of the Spatial Autoregressive (SAR) model.

Table 7. Lagrange Multiplier (Lag) Test Results

Dependensi Spasial Test	DB	p-value
Lagrange Multiplier (Lag)	1	0.0006

Based on table 8, shows that mean year schooling (X_1), poor people (X_2), labour force participation rate (X_5) and spatial remainder (ρ) significantly influence the percentage of life expectancy on Sulawesi Island at $\alpha = 0.05$. So it can be concluded that obtained by looking at the p-value of the variables used in the study has a value smaller than $\alpha = 0.05$, so the SAR model equation is obtained as follows:

$\hat{Y}_i = 12.004 + 0.543 \sum_{j=1,i\neq j}^{81} W_{ij}Y_i + 0.643X_{1i} - 0.181X_{2i} + 0.484X_{3i} + 0.083X_{4i} + 0.149X_{5i} - 0.053X_{6i} + U_i$

Notes:

$\hat{Y}_i$: Estimator Value of the dependent variable Life Expectancy in the Regency/City to-i

Y_i : The dependent variable size $n \times 1$ from the Regency /City to-i

X_{1i} : Mean Year Schooling in the Regency /City to-i

X_{2i} : Poor people in the Regency/City to-i

X_{3i} : Expected Year Schooling in the Regency /City to-i

X_{4i} : Unemployment Rate in the Regency/City to-i

X_{5i} : Labour Force Participation Rate in the Regency/City to-i

X_{6i} : Population Growth Rate in the Regency/City to-i

U_i : Error spasial from the Regency/City to-i

W_{ij} : The element of the matrix **W** in the *-i* row of the *-j* column $i,j = 1,2,\dots,n$

Table 8. Estimation and Testing of SAR Model Parameters

Var	Koefisien	Std. Error	z-value	p-value
Constant	12.004	8.767	1.369	0.170
X_1	0.643	0.223	2.885	0.003*
X_2	-0.181	0.048	-3.780	0.000*
X_3	0.484	0.248	1.945	0.051
X_4	0.083	0.105	0.796	0.425
X_5	0.149	0.034	4.321	0.000*
X_6	-0.053	0.270	-0.197	0.843
ρ	0.543	0.110	4.909	0.000*

*) significant α=0.05

Spatial Autoregressive (SAR)

In general, the SAR model equation above can be interpreted as follows:

1. The mean year schooling has a positive relationship with life expectancy which means, if other factors are considered constant, if the Mean Year Schooling (X_1) rises by one unit it will add a life expectancy (Y) of 0.643. In this study, the Mean Year Schooling variables has an influence on life expectancy on Sulawesi Island. Based on the $\text{p} - \text{value}(0.003) < \alpha = 0.05$.
2. Poor people has a negative relationship with life expectancy, which means, poor population (X_2) increases by one unit, so it will reduce life expectancy (Y) by 0.181. In this study, the variable of the poor people has an influence on life expectancy on Sulawesi Island. Based on the $\text{p} - \text{value}\,(0.000) < \alpha = 0.05$.
3. Expected Year Schooling have a positive relationship with life expectancy, which means, long-term school expectations (X_3) go up by one unit, so it will add a life expectancy (Y) of 0.484. In this study, the variable of old school expectations has an influence on life expectancy

on Sulawesi Island. Based on the $p-value(0.051) > \alpha = 0.05$.

4. Open unemployment rate has a positive relationship with life expectancy, which means, open unemployment rate (X_4) increases by one unit, so it will add life expectancy (Y) of 0.083. In this study, the variable open unemployment rate has an influence on life expectancy on Sulawesi Island. Based on the $p-value(0.425) > \alpha = 0.05$.
5. The labour force participation rate has a positive relationship with life expectancy which means, the labor force participation rate (X_5) rises by one unit then it will add life expectancy (Y) of 0.149. In this study, the variable labor force participation rate has an influence on life expectancy on Sulawesi Island. Based on the $p-value(0.000) < \alpha = 0.05$.
6. The population growth rate has a negative relationship with life expectancy, which means that the population growth rate (X_6) increases by one unit so it will reduce life expectancy (Y) by 0.053. In this study, the variable population growth rate has an influence on life expectancy on Sulawesi Island. Based on the $p-value(0.843) > \alpha = 0.05$.

The rho coefficient (ρ) shows that an area surrounded by other regionsas many as n, so the influence of each surrounding area can be measured by 0.543, which means that life expectancy on the island of Sulawesi, it is influenced by neighborliness between regions.

Assumption Testing of SAR Model

Normalitas Test

The Shapiro-Wilk value of 0.40947 was obtained and the significance value of $0.2031 > \alpha = 0.05$, so H_0 was accepted. This shows that the data used in the study were normally distributed.

Table 9. Shapiro-Wilk Test Results

Test	Value Shapiro-Wilk	*p-value*
Shapiro-Wilk	0.97894	0.2031

Heteroskedasticity Test

Table 10. Breusch-Pagan Test Results

Test	DB	Value BP	p-value
Breusch-Pagan	6	10.485	0.105

Autocorrelation Test

Based on table 11, a DW value of 1.8157 is obtained, while a DW table with a significant level of 0.05 obtained a dL value of 1.4842, and a dU of 1.8008. Decision making on the presence or absence of autocorrelation can be seen in table 1. Based on the table the DW value is in the area between $dU < DW < 4 - du$, so accept H_0, it can be concluded that there is no positive or negative autocorrelation.

Table 11. Durbin Watson Test Results

Test	Value DW
Durbin Watson	1.8429

Good Measure Of Models

The best spatial regression model is chosen by looking at the value of the goodness of the model. The goodness of a model can be seen from the resulting AIC value. The smaller AIC value compared to other models indicates that this model is better than other models. The measure of the goodness of the model is presented in the table as follows:

Table 12. Good measure of models

Model	AIC
Classic	340.28
SAR	328.24

The AIC value produced by the SAR model is smaller than the classical regression model. So it can be concluded that overall the SAR model is better used to model the life expectancy of the Regency / City on Sulawesi Island.

ACKNOWLEDGEMENT

The author would like to thank to the Central Statistics Agency for allowing the authors to retrieve data to complete this study.

REFERENCES

Arbia, G. (2005). Spatial Econometrics: Statistical Foundation and Application to Regional Convergence. Berlin: Springer.

[BPS] Badan Pusat Statistik (2017). Data dan informasi angka harapan hidup: Badan Pusat Statistik.

[BPS] Badan Pusat Statistika (2013). Angka Harapan Hidup. Jakarta: Badan Pusat Statistik.

Budianto. (2012). Implementasi Algoritma Breadth First Search Dan Obstancle Detection Dalam Penelusuran Eabrin Dinamis Menggunakan Robot Lego. Semarang: Ilmu Komputer Universitas Diponogoro.

Komang, KSDP, Made S, I Wayan S. (2017). Metode Spatial Autoregressive Dalam Memodelkan Masyarakat Yang Berperilaku MCK di Sungai.

Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana.

Ni Wayan, DA. (2015). Pemodelan Angka Harapan Hidup di Provinsi Jawa Timur Dengan Menggunakan Metode Feed Forward Neural Network (FFNN), Jurusan Akuntansi, Politeknik Negeri Bali.

Supriatna, N., Ruhima, M. & Kosim. (2006). Ilmu Pengetahuan Sosial (Geografi, Sejarah, Sosiologi, Ekonomi). Bandung: Grafindo Media Pratama.