

BAREKENG: Journal of Mathematics and Its ApplicationsJune 2023Volume 17 Issue 2Page 0727–0734P-ISSN: 1978-7227E-ISSN: 2615-3017

doi https://doi.org/10.30598/barekengvol17iss2pp0727-0734

# MODELLING OF POVERTY PERCENTAGE IN EAST JAVA PROVINCE WITH SEMIPARAMETRIC REGRESSION APPROACH

## Idrus Syahzaqi<sup>1\*</sup>, Salman Alfarizi P. A.<sup>2</sup>, Kartika Fithriasari<sup>3</sup>

<sup>1,2,3</sup> Department of Statistics, Institut Teknologi Sepuluh Nopember Raya ITS Street, Sukolilo, Surabaya City, 60111, Indonesia

Corresponding author's e-mail: \* idrus.syahzaqi@gmail.com

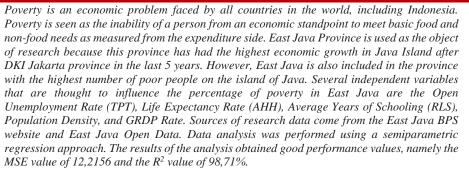
#### ABSTRACT

#### Article History:

Received: 5<sup>th</sup> December 2022 Revised: 9<sup>th</sup> April 2023 Accepted: 12<sup>th</sup> April 2023

#### Keywords:

East Java; MSE value; Percentage of Poverty; R<sup>2</sup> value; Semiparametric Regression.





This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.

*How to cite this article:* 

I. Syahzaqi, S. Alfarizi P. A and K. Fithriasari., "MODELLING OF POVERTY PERCENTAGE IN EAST JAVA PROVINCE WITH SEMIPARAMETRIC REGRESSION APPROACH," *BAREKENG: J. Math. & App.*, vol. 17, iss. 2, pp. 0727-0734, June, 2023.

Copyright © 2023 Author(s) Journal homepage: https://ojs3.unpatti.ac.id/index.php/barekeng/ Journal e-mail: barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id

**Research Article** • **Open Access** 

## 1. INTRODUCTION

Poverty is a macroeconomic conflict faced by all countries in the world, including Indonesia [1]. The source of the problem of poverty in Indonesia is the high disparity between regions caused by the uneven level of income distribution [2]. This creates a widening gap between the rich and the poor, so poverty reduction is a major concern for human development [3]. According to [4], poverty is seen as an economic inability to meet basic food and non-food needs as measured from the expenditure side. Poor people are residents who have an average expenditure per capita per month below the poverty line.

Indonesia is a vast archipelagic country consisting of several provinces with the characteristics of people with different economic conditions. So, to solve the problem of poverty in Indonesia, it must start from the lower level, starting at the district, provincial and national levels. East Java was chosen as the focus of this research based on several reasons. First, in the government's official publication through the Central Bureau of Statistics [5], in the last 5 years, East Java has been the province with the highest economic growth on the island of Java after the province of DKI Jakarta. But on the other hand, the number of poor people in East Java in 2019 was 4.112 million people, which means that East Java is the province with the largest number of poor people on the island of Java. Even though, according to data from BPS, economic growth in East Java has continued to increase from year to year, this increase in economic growth has not been matched by a significant reduction in the poverty rate. Second, the population always increases every year without being matched by an even distribution of population distribution. So far, most of the population is still concentrated in Java. Data from the 2020 BPS census shows that 151,591,262 people out of Indonesia's total population of 270,203,917 live on the island of Java. Of this number, 40,665,696 residents live in East Java Province, the second province with the largest population after West Java.

Unemployment has several types, one of which is open unemployment. Open unemployment defines as part of the workforce that is currently not working and is actively looking for work. Meanwhile, Adzim [6] explains that the open unemployed are those who do not have a job because they are looking for work, preparing for a business, or because they feel it is impossible to get a job, as well as those who already have a job but have not started working yet. To measure the percentage level of unemployment in a region, the open unemployment rate is generally used. The open unemployment rate is the percentage of the number of unemployed to the total labour force.

Life expectancy at birth is the average number of years a new born will live in a given year. Life expectancy is a tool to evaluate the government's performance in improving the welfare of the population in general and improving health status in particular. Low life expectancy in an area must be followed by health development programs and other social programs.

The average length of schooling (RLS) is a number that describes the length (years) of school experienced by people aged 25 years and over. The Central Bureau of Statistics suggests that RLS is defined as the number of years used by the population to undergo formal education. RLS can be used to determine the level and quality of community education in an area, while HLS is defined as the length of schooling (years) that is expected to be experienced by children at a certain age in the future. In addition, there are several factors that cause children to drop out of school, such as environmental factors, understanding of the importance of education, culture, availability of educational facilities/infrastructure, and others. This is in line with states that children who drop out of school are generally unable to pay tuition fees and do not receive information about scholarships, both regarding scholarship sources and how to access them, so the response is to build a Scholarship Village [7].

A large population can move the market from demand through a multiplier effect due to aggregate demand. Population in the economic development of a region is a fundamental problem because uncontrolled population growth can result in failure to achieve economic development goals, namely community welfare and poverty alleviation [8]. Population density is a condition that is said to be denser if the number of people in a certain space is more than the area of the room [9]. Population density is the ratio between the number of residents and the area inhabited [10]. The increase in population density was influenced by the number of residents, which continued to increase yearly. Population density accompanied by a high accumulation of human capital will encourage increased economic activity. The skills and knowledge possessed by individuals will affect their performance, people with high human capital will be able to produce new technological ideas that can encourage increased output.

728

GRDP is income generated through goods and services by all economic activities in an area in a certain period. The higher the GRDP in a region, the greater the level of revenue for the region. However, GRDP does not guarantee that all residents enjoy prosperity. GRDP is only an overview of people's welfare. The increase in GRDP cannot be concluded whether the condition of the low-income population has improved or not [11].

Based on the description above, this study will apply semiparametric regression estimation using the Fourier series estimator on the poverty percentage data for East Java Province on several factors, namely the Open Unemployment Rate (TPT), Life Expectancy Rate (AHH), Average Years of Schooling (RLS), Population Density, and GRDP Rate.

## 2. RESEARCH METHODS

The research method contains explanations in the form of paragraphs about the research design or descriptions of the variables, data sources, and data analysis.

#### 2.1 Variables and Data Sources

The data used in this study is secondary data consisting of the variables Percentage of Poor Population, Open Unemployment Rate (TPT), Life Expectancy Rate (AHH), Average Years of Schooling (RLS), Population Density, and GRDP Rate in East Java in 2021. Types of data for each variable and data source can be seen in the following table:

Variable	Description	Data Type	Data Source
Y	Percentage of Poor Population	Continuous	East Java BPS
$X_1$	TPT	Continuous	East Java Open Data
$X_2$	AHH	Continuous	East Java Open Data
$X_3$	RLS	Continuous	East Java Open Data
$X_4$	Population Density	Continuous	East Java Open Data
$X_5$	GRDP Rate	Continuous	East Java BPS

Table 1. Variables and data sources

#### Variables Definitions:

- 1. The percentage of poverty is the percentage of the population below the poverty line (GK). GK is the sum of the Food Poverty Line (GKM) and the Non-Food Poverty Line (GKNM). GKM is the value of spending on minimum food needs which is equivalent to 2100 kilo calories per capita per day. GKNM is the minimum requirement for housing, clothing, education, and health.
- 2. The Open Unemployment Rate (TPT) is the percentage of the number of unemployed to the total labour force. It is said to be unemployed if someone does not have a job and is looking for work, someone does not have a job and is preparing for a business, someone does not have a job and is not looking for work because they feel it is impossible to get a job, someone already has a job but has not started working.
- 3. Life Expectancy at Birth (AHH) is defined as the average number of years that a person can live from birth. AHH reflects the degree of health of a community.
- 4. The average length of schooling (RLS) is defined as the number of years spent by the population in formal education. It is assumed that under normal conditions, the average length of schooling in a region will not decrease. The population coverage calculated in calculating the average length of schooling is the population aged 25 years and over.
- 5. Population density shows the number of inhabitants per square kilometre of area. The intended area is the total area of land in an administrative area.
- 6. The GRDP rate shows the growth in the production of goods and services in an economic area within a certain time interval.

:	District / City	<b>Respond Variable</b>	Predictor Variable				
1		Y	X <sub>1</sub>	$\mathbf{X}_2$	<b>X</b> <sub>3</sub>	X4	X5
1	Pacitan	$Y_1$	$X_{1,1}$	$X_{2,1}$	$X_{3,1}$	$X_{4,1}$	$X_{5,1}$
2	Ponorogo	$Y_2$	$X_{1,2}$	$X_{2,2}$	$X_{3,2}$	$X_{4,2}$	$X_{5,2}$
:	:	:	:	:	:	:	:
n=38	Batu City	Y <sub>38</sub>	$X_{1,38}$	X <sub>2,38</sub>	X <sub>3,38</sub>	$X_{4,38}$	X <sub>5,38</sub>

Table 2. Data structure

#### 2.2 Analysis Technique

This study uses a semiparametric regression approach, which is a combination of parametric regression and nonparametric regression, to determine the factors that are thought to influence the percentage of poor people in Indonesia. The parametric approach that can be used is the linear regression method. The parametric approach is chosen when the regression curve forms a certain pattern, for example, linear. Meanwhile, the nonparametric approach can be an alternative when the regression curve is unknown or does not have a certain pattern. One of the estimators that can be used in a nonparametric approach is the Fourier Series. The steps in this research are as follows:

- 1. Conduct a descriptive analysis of the variables Percentage of Poor Population, Open Unemployment Rate (TPT), Life Expectancy Rate (AHH), Average Years of Schooling (RLS), Population Density, and GRDP Rate in East Java in 2021 to find out the characteristics of the data by steps as follows:
  - a. Describes the Percentage of Poor Population, Open Unemployment Rate (TPT), Life Expectancy Rate (AHH), Average Years of Schooling (RLS), Population Density, and GRDP Rate in East Java, including average, maximum, and minimum values.
  - b. Make a scatter plot between response variables and predictors to identify which data patterns are classified as parametric regression and nonparametric regression.
- 2. Analyse the variables that are thought to influence the Percentage of the Poor Population in East Java Province using a semiparametric regression approach with the following steps:
  - a. Create a semiparametric regression model. Given the response variable Y with parametric components  $X_1, X_2, ..., X_p$  and nonparametric components  $t_1, t_2, ..., t_r$ :

$$y_{i} = \beta_{0} + \beta_{1} x_{1i} + \beta_{2} x_{2i} + \dots + \beta_{p} x_{pi} + \sum_{r=1}^{R} f(t_{ri}) + \varepsilon_{i}$$
(1)

Where i = 1, 2, ..., n.

b. Approach nonparametric components with a Fourier Series function.

$$f(t_{ri}) = b_r t_{ri} + \frac{\alpha_0}{2} + \sum_{k=1}^{K} (\alpha_{kr} \cos k t_{ri})$$
(2)

Where i = 1, 2, ..., n.

- c. Model the data using semiparametric regression, where parametric components are approximated by linear functions and nonparametric components are approximated by Fourier Series.
- d. Choose the optimal k oscillation parameters seen from the minimum GCV value to form the best model.
- e. Calculate MSE and the coefficient of determination  $(R^2)$  as criteria for the goodness of the model.
- f. Interpret the results of semiparametric regression modelling that have been obtained and draw final conclusions.

#### 3. RESULTS AND DISCUSSION

#### 3.1. Statistics Descriptive

To find out the characteristics of the data on the response variables and predictors used, it can be seen from the results of the descriptive statistics in **Table 3** below:

No.	Variable	Mean	Maximum	Minimum
1	Y	11.322	23.760	4.090
2	$X_1$	70.384	80.570	64.240
3	$X_2$	71.718	74.180	66.890
4	$X_3$	8.061	11.370	4.860
5	$X_4$	2.004	8.504	0.303
6	$X_5$	4.60	30.76	0.26

Table 3.	Descriptive	statistics of	of each	variable
----------	-------------	---------------	---------	----------

In the table above, it can be concluded that the poverty percentage variable (Y) has an average of 11,322 with a maximum value of 23,76 in Bangkalan Regency and a minimum value of 4,090 in Batu City. Meanwhile, the Open Unemployment Rate ( $X_1$ ) has an average of 70,384 with a maximum score of 80,570 in Pacitan Regency and a minimum score of 64,24 in Jombang Regency. Then the Life Expectancy Rate ( $X_2$ ) has an average of 71,718 with a maximum value of 74,180 in Madiun City and a min value of 66,890 in Banyuwangi Regency. Then, the average length of schooling ( $X_3$ ) has an average of 8,061 with a maximum score of 4,860 in Bangkalan Regency. Population Density ( $X_4$ ) has an average of 2,004 with a maximum value of 8,504 in Pasuruan City and a minimum value of 0,303 in Jember Regency. GRDP rate ( $X_5$ ) has an average value of 4,60, a maximum value of 30,76 in Pasuruan Regency, a minimum value of 0,26 in Bangkalan Regency.

The next descriptive analysis is to make a scatter plot. This is done to identify the pattern of the relationship between the response variable and each predictor variable. Furthermore, it can be used to determine variables as parametric components and nonparametric components. Figure 1 shows the shape of the relationship pattern between the Poor Population Percentage variable and the predictor variables, namely TPT ( $X_1$ ), AHH ( $X_2$ ), RLS ( $X_3$ ), Population Density ( $X_4$ ), and GRDP Rate ( $X_5$ ).

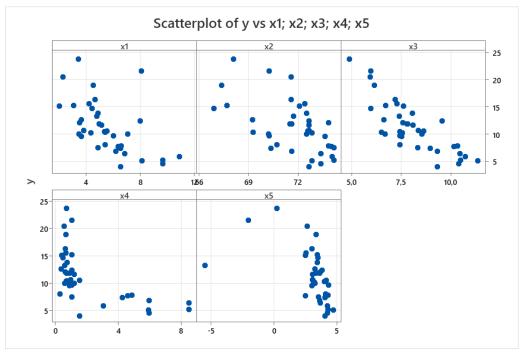


Figure 1. Scatter plot results between response variables and predictors

From **Figure 1**, it can be seen that relationship between TPT, AHH, and RLS forms a pattern that tends to be linear and inversely proportional to the percentage of poor people in East Java Province. This shows that the higher the TPT, AHH, and RLS factors, the lower the percentage of poor people. While the relationship between the variable population density and GRDP figures for the percentage of poor people does not form a linear (non-linear) pattern. The results of the scatter plot identification are presented in **Table 4**.

Table 4. Identify	scatter	plots
-------------------	---------	-------

<b>Response Variable</b>	Predictor Variable	Identification Results
Percentage of Poor Population	TPT	Linear Parametric
Percentage of Poor Population	AHH	Linear Parametric
Percentage of Poor Population	RLS	Linear Parametric
Percentage of Poor Population	Population Density	Nonparametric
Percentage of Poor Population	GRDP Rate	Nonparametric

The relationship pattern between the variable percentage of poor people and the TPT, AHH, and RLS variables appears to form a linear pattern. From the results of the scatter plot, the TPT, AHH, and RLS variables are approached with parametric regression. Meanwhile, the pattern of relationship between the variable percentage of poor people and population density and GRDP rate tends not to form a specific pattern. This unknown pattern of relationships makes solving this problem a nonparametric regression approach.

## 3.2. Semiparametric Regression Method

To get the best estimator in the semiparametric regression model with the Fourier Series approach, the optimal *K* is selected using the GCV method. The value of *K* is chosen based on the minimum GCV value. The value of *K* is the number of cosine wave oscillations in the regression model with the Fourier Series approach. The greater the value of *K*, the more complex the model and the tighter the oscillations of the regression curve. For this reason, the value of *K* is limited to K = 5. Table 5 is a table of GCV gains for values of K = 1, 2, 3, 4, and 5.

Table 5. GCV value			
K Value	GCV		
1	1905456.3		
2	369167.4		
3	1648242		
4	572534.6		
5	1282733.1		

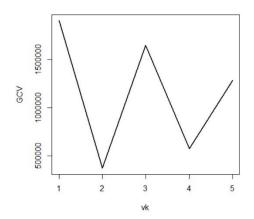


Figure 2. Changes in GCV values in the Fourier Series for K = 1 to K = 5

By paying attention to K=2, the minimum GCV is obtained when compared to other GCV values. The semiparametric regression model with the Fourier Series approach with K=2 was chosen to model the Percentage of Poor Population in East Java. The estimation of the semiparametric regression model with the Fourier Series approach for K=2 is as follows:

$$\hat{y}_{i} = \hat{\beta}_{0} + \hat{\beta}_{1}x_{1i} + \hat{\beta}_{2}x_{2i} + \hat{\beta}_{3}x_{3i} + \frac{\hat{\alpha}_{0}}{2} + \hat{b}_{1}t_{1i} + \hat{\alpha}_{11}\cos t_{1i} + \hat{\alpha}_{21}\cos 2t_{1i}$$

$$+ \hat{b}_{2}t_{2i} + \hat{\alpha}_{12}\cos t_{2i} + \hat{\alpha}_{22}\cos 2t_{2i}$$
(3)

Based on the estimator **Table 6**, a semiparametric regression model with the Fourier Series approach for the Percentage of Poor Population in East Java Province is obtained as follows:

$$\hat{y}_{i} = 20,08 + 0,08 x_{1i} - 0,01 x_{2i} - 2,37 x_{3i} + 9,7 + 0,14 t_{1i} - 0,1 \cos t_{1i} -0,23 \cos 2t_{1i} - 0,47 t_{2i} + 0,55 \cos t_{2i} + 0,48 \cos 2t_{2i}$$
(4)

<b>Table 6. Regression</b>	coefficient	parameter	value
----------------------------	-------------	-----------	-------

Parametric Parameters	Value	Nonparametric Parameters	Value
$\beta_0$	20.08	$\frac{\alpha_0}{\alpha_0}$	9.7
		2	
$\beta_1$	0.08	$b_1$	0.14
$\beta_2$	-0.01	$\alpha_{11}$	-0.1
$\beta_3$	-2.37	$\alpha_{21}$	-0.23
		$b_2$	-0.47
		$\alpha_{12}$	0.55
		$\alpha_{22}$	0.48

This model has a value of MSE = 12,2156 and minimum GCV = 369167,4. These values are smaller when compared to other *K* values. The coefficient of determination or  $R^2$  that this model has is high, namely 0,9871. This means that the diversity of response values has been able to be explained by the predictor variable of 98,71%.

#### **3.3. Interpretation**

The resulting model will present several model interpretations of the variables in the parametric regression. The following describes the interpretation of the model and description of the behavior of several variables.

- 1. The interpretation of the TPT variable  $(X_1)$  has a value of 0.8 which means that if there is an increase in TPT by one percent, the poverty percentage value will increase by 0.8 percent. This is in accordance with research conducted by [12], which states that the results of the interpretation regarding the relationship between TPT and the percentage of poor people show a positive value, meaning that the percentage of poor people increases when the TPT variable increases. The fact that the high percentage of poor people is accompanied by a high unemployment rate can be caused by a lack of employment opportunities and the low quality of human resources. Increasing the TPT variable can have a direct impact on poverty, so it is necessary to increase skills and expand employment opportunities to reduce this variable so that poverty decreases.
- 2. The coefficient of the AHH variable  $(X_2)$  has a value of -0.01, which means that if there is an increase in AHH by one percent, the poverty percentage value will decrease by 0.01 percent. This study resulted in the variable life expectancy showing a negative sign on the percentage of poverty. The higher the life expectancy, the higher the quality of public health, the higher the productivity level. An increased level of community productivity can encourage economic growth, which can ultimately reduce poverty, meaning that the higher the life expectancy, the lower poverty will be [13].
- 3. The interpretation of the RLS variable (X<sub>3</sub>) has a value of -2.37 which means that if there is an increase in TPT by one percent, the poverty percentage value will increase by 2.37 percent. This has been explained by [14] and [15] that education plays an important role in tackling and reducing poverty in the long run, either indirectly through productivity improvements, or directly through training the poor with the skills needed to increase their productivity and in time will increase their income.

## 4. CONCLUSIONS

Based on the results of research and data analysis using simple linear regression techniques that have been carried out in this study, several conclusions can be drawn, namely:

- 1. In the descriptive analysis, it was found that the poverty percentage factor (Y) had an average of 11.322 with a maximum value of 23.76 in Bangkalan Regency and a minimum value of 4.090 in Batu City. The Open Unemployment Rate (X<sub>1</sub>) has an average of 70.384, with a maximum value of 80.570 in Pacitan Regency and a minimum value of 64.24 in Jombang Regency. Life expectancy (X<sub>2</sub>) has an average of 71.718, with a maximum value of 74.180 in Madiun City and a minimum value of 66.890 in Banyuwangi Regency. The average length of schooling (X<sub>3</sub>) has an average of 8.061, with a maximum score of 11.370 in Mojokerto City and a minimum value of 8.504 in Pasuruan City and a minimum value of 0.303 in Jember Regency. GRDP rate (X<sub>5</sub>) has an average of 4.60, with a maximum value of 30.76 in Pasuruan Regency and a minimum value of 0.26 in Bangkalan Regency.
- 2. A semiparametric regression model with the Fourier Series approach for the Percentage of Poor Population in East Java Province is obtained as follows:

$$\hat{y}_i = 20,08 + 0,08 x_{1i} - 0,01 x_{2i} - 2,37 x_{3i} + 9,7 + 0,14 t_{1i} - 0,1 \cos t_{1i} - 0,23 \cos 2t_{1i} - 0,47 t_{2i} + 0,55 \cos t_{2i} + 0,48 \cos 2t_{2i}$$

This model yields an MSE value of 12.2156 with a minimum GCV value of 369,167.4 and an  $R^2$  value of 98.71%.

## ACKNOWLEDGMENT

The authors give high appreciation for East Java Open Data and BPS RI which has provided data and Institut Teknologi Sepuluh Nopember that have supported this research.

#### REFERENCES

- [1] N. Sakinah, H. Pudjianto, and U. J. Soedirman, "Determinants of Poverty in East Java Metropolitan Area in 2010-2016," no. 26, pp. 32–40, 2016.
- [2] I. Introduction, "What's Happened to Poverty and Inequality in Indonesia over Half a Century?," vol. 38, no. 1, pp. 68–97, 2021.
- [3] W. S. Report, INEQUALITY IN A RAPIDLY CHANGING. 2020.
- [4] P. N. Sambas, "Analysis of Poverty in Indonesia," pp. 7368–7373, 2021.
- [5] I. Yulivan *et al.*, "PT. PINDAD in Supporting The Defense Economy in East Java Indonesia," no. 3, pp. 10–17, 2022, doi: 10.55683/jrbee.v4i3.375.
- [6] F. Adzim, S. Dian, W. Prajanti, M. I. Ekonomi, F. Ekonomika, and U. Diponegoro, "Assessing the Open Unemployment Rate in Bali : Regional Panel Data in Bali Evidence from," vol. 10, no. 5, pp. 252–265, 2021.
- [7] M. Fisher, A. Maryudi, and M. A. K. Sahide, "Forest and Society : Initiating a Southeast Asia Journal for Theoretical, Empirical , and Regional Scholarship," vol. 1, no. April, pp. 1–6, 2017.
- [8] H. R. Espacios, L. O. S. Autores, A. Asrol, and H. Ahmad, "Analysis of factors that affect poverty in Indonesia," 2018.
- [9] P. O. Hakim I, Sunoko K, "Spatial territoriality in the Semanggi embankment area Spatial territoriality in the Semanggi embankment area," 2020, doi: 10.1088/1755-1315/447/1/012029.
- [10] Voit E, Population Systems. New York, 2020. doi: 10.1016/B978-0-12-803033-2/00005-4.
- [11] F. Effect and R. E. Model, "THE EFFECT OF ECONOMIC GROWTH ON INCOME INEQUALITY :," 2020.
- [12] U. F. Muhammad and J. David, "Relationship between Poverty and Unemployment in Niger State," vol. 8, no. 1, pp. 71–78, 2019.
- [13] E. Science, "Impact of life expectancy, literacy rate, opened unemployment rate and gross domestic regional income per capita on poverty in the districts / city in Central Sulawesi Province Impact of life expectancy, literacy rate, opened unemployment rate and gro," 2018.
- [14] A. M. Arsani, B. Ario, and A. F. Ramadhan, "Impact of Education on Poverty and Health : Evidence from Indonesia," vol. 9, no. 6, 2020.
- [15] S. Van Der Berg, Poverty and education, 10th ed. Paris: UNESCO IIEP, 2008.

734