Nonparametric Regression Modeling with Multivariable Fourier Series Estimator on Average Length of Schooling in Central Java in 2023

Ludia Ni'matuzzahroh[1*], Andrea Tri Rian Dani [2]
[1]Statistics Study Program, YPPI Rembang University, Rembang, Indonesia
[2]Statistics Study Program, Department of Mathematics, Mulawarman University, Samarinda, Indonesia
*Corresponding author: ludianimatuzzahroh@gmail.com

Received: 24 March 2024
Revised: 23 April 2024
Accepted: 19 June 2024

ABSTRACT – **One of the benchmarks to see the quality of education and human resources in Indonesia is the average length of schooling. If the school average is higher, it can positively impact Indonesian society, enabling it to compete globally. There are several factors, both economic and educational factors, that influence the low average length of schooling in Central Java Province. Therefore, this research aims to model and determine what variables can influence the average length of schooling in Central Java in 2023 using a nonparametric regression approach with a multivariable Fourier series estimator. This approach is used when the form of the relationship pattern is unknown and tends to have recurring patterns. The Fourier series estimator depends on the number of oscillations, so in this study, 1 to 4 oscillations were tried, where the minimum GCV value determined the optimal oscillation. The best model was obtained on the analysis results, producing the smallest GCV value, namely the model with 3 oscillations with a GCV value of 1.027. The results of simultaneous and partial hypothesis testing showed that all predictor variables used in this research were proven to influence the Average Length of Schooling. This is also supported by the coefficient of determination value of 85.464%.**

Keywords– **Fourier series, nonparametric regression, GCV, average length of schooling**

I. INTRODUCTION

One measure of the quality of education and human resources in Indonesia is the average length of schooling. The higher the average years of schooling, the higher the Indonesian people's education level. If the Indonesian people's education level increases, it will positively affect individuals, communities, and the country. The positive impact of a high level of education along with sufficient knowledge is that it can make it easier for individuals to find work and increase their ability to work, which will also increase the potential income of individuals. This, of course, will encourage the economic growth of the Indonesian people so that life is more advanced and prosperous, and they will be able to compete in the global arena. Based on publication data at the Badan Pusat Statistik (BPS) in 2023, the Indonesian population's average schooling length reached 8.77 years [1]. This means that the Indonesian population receives an average of 8.77 years of formal education, or it can be said that most of the Indonesian population receives education up to the Junior High School (SMP) level. When compared to 2022, the average length of schooling increased by 0.08. This indicates that the enthusiasm of the Indonesian population in obtaining formal education is slowly increasing every year. Even this trend of increasing the average length of schooling in Indonesia has been going on for the past decade.

Although, in general, the national average length of schooling in 2023 is 8.77 years, several provinces have an average length of schooling below this value. Of the six provinces in Java Island, Central Java Province is the province that has the lowest average length of schooling for the last three years compared to the other five provinces, namely 8.01 in 2023, 7.93 in 2022, and 7.75 in 2021 [2]. This, of course, must be a concern for the Central Java Provincial Government in determining policies to increase the average length of schooling so that the quality of human resources becomes more advanced.

The low average years of schooling in Central Java Province can be triggered by the high poverty rate in some areas of Central Java, which makes children unable to attend school because many families still cannot afford to pay for their children's education. This is also supported by the economic disparity between regions in Central Java, where access to education and the quality of education in disadvantaged areas is still low. In addition, families with low minimum wages may find it difficult to pay for their children's education because the wages they earn are only enough to fulfil their basic daily needs, causing children to drop out of school. This, of course, can impact the school participation rate, affecting a region's average length of schooling in a region [3]. In addition to economic factors, the community is also required to have high literacy and insight to become more aware of the importance of education so that this can improve the quality of learning and the average length of schooling.

Some of the factors above are thought to affect the average length of schooling in Central Java Province; therefore, an analysis called regression analysis is needed to ascertain whether these factors have an effect. Regression analysis determines the shape of the relationship pattern between predictor variables and response variables [4]. The response variable used in this study is the Average Length of Schooling in Regencies/Cities in Central Java Province in 2023. The main purpose of regression analysis is to find the shape of the estimated regression curve [5]. In reality, the shape of the

relationship pattern between predictor variables and response variables cannot be known exactly and clearly, such as linear, quadratic, cubic, and other parametric patterns. Therefore, one approach that can be used is the nonparametric regression approach. This is because the nonparametric regression approach does not depend on the assumption of a particular curve shape and has high flexibility where the data is expected to adjust the shape of the regression curve estimate without being influenced by the researcher's subjectivity [6].

In nonparametric regression, researchers often use one estimator, namely the Fourier series estimator. The Fourier series estimator is generally used when the data under investigation has an unknown relationship pattern and tends to have a repeating pattern [7]. The Fourier series is very dependent on the determination of the number of oscillations, where the longer the oscillation, the more cosine waves are produced, causing the model to be more complex and the oscillations are closer to the actual data pattern. One method that can be used in determining the optimal oscillation is the Generalized Cross-Validation (GCV) method. Theoretically, this GCV method is asymptotically optimal and has a formula that does not contain σ^2 variance for unknown populations, and invariance to transformation [8]. Several studies have been conducted previously using the Fourier Series estimator in nonparametric regression, including [9]-[14].

Based on the description above, nonparametric regression modeling of multivariable Fourier Series estimators will be applied to the case study of Average Length of Schooling in Regencies/Cities in Central Java Province in 2023. Then, proceed with simultaneous and partial hypothesis testing to achieve the objectives of this study, namely to determine whether the Percentage of Poor Population, Community Literacy Development Index, Gini Ratio, Minimum Wage, and School Participation Rate are factors that can affect the Average Length of Schooling variable.

II. LITERATURE REVIEW

A. Nonparametric Regression

Nonparametric regression is one of the approaches used to estimate the shape of the regression curve when the data pattern is unknown. This is because nonparametric regression can find its form of regression curve estimation without being influenced by the subjectivity of the researcher and without knowing past information about the data pattern used so nonparametric regression can be said to have high flexibility [6]. If there are paired data (x_{ij}, y_j) with $i = 1, 2, \dots, m$ predictor variables and $j = 1, 2, \dots, n$ observations in each variable and the shape of the relationship pattern of the two variables is assumed to follow multivariable nonparametric regression, it can be generally written:

$$y_j = f(x_{ij}) + \varepsilon_j \quad ; i = 1, 2, \dots, m \;\; ; j = 1, 2, \dots, n \tag{1}$$

where y_j is response variable, x_{ij} is the predictor variable, $f(x_{ij})$ is a function whose regression curve shape is unknown, and ε_j is a random error assumed to be identical, independent, and normally distributed with zero mean and variance σ^2.

In nonparametric regression, several estimators can be used to estimate the regression curve when the shape of the regression curve is unknown. Often used estimators include Fourier Series, Truncated Spline, Kernel, K-Nearest Neighbor, etc.

B. Multivariable Fourier Series Estimator

The Fourier series estimator is often used in nonparametric regression to estimate the regression curve when the shape of the curve is unknown, and there is a tendency to have a repeating pattern. The Fourier series is a trigonometric polynomial that can adapt effectively or has high flexibility to the local nature of the data [15]. This is because there are oscillations (c) in the Fourier series, where the longer the oscillation, the more waves are produced.

If $f(x_{ij})$ in Equation (1) is a function whose exact shape is unknown and is assumed to form a repeating relationship pattern, then the function can be approximated by a multivariable Fourier Series. In general, the form of a nonparametric regression model with a multivariable Fourier series estimator is as in Equation (2) below [14]:

$$y_j = \frac{1}{2}a_0 + \sum_{i=1}^{m}\left(b_i x_{ij} + \sum_{c=1}^{C} a_{ci} cos c x_{ij}\right) + \varepsilon_j \quad ; i = 1, 2, \dots, m \;\; ; j = 1, 2, \dots, n \;\; ; c = 1, 2, \dots, C \tag{2}$$

with a_0, b_i, and a_{ci} are parameters of the model.

If Equation (2) above is expressed in matrix form, the result can be seen as in Equation (3).

$$\vec{\boldsymbol{y}} = \boldsymbol{X}\vec{\boldsymbol{\beta}} + \vec{\boldsymbol{\varepsilon}} \tag{3}$$

Based on the matrix form in Equation (3) above, it can be seen that $\vec{\boldsymbol{y}}$ s a response vector of size $n \times 1$, $\boldsymbol{X}$ is a matrix containing nonparametric components of multivariable Fourier series of size $n \times m(C+2)$, $\vec{\boldsymbol{\beta}}$ is a parameter vector of size $m(C+2) \times 1$, and $\vec{\boldsymbol{\varepsilon}}$ is an error vector of size $n \times 1$ which is assumed to be identical, independent, and normally distributed with mean zero and variance σ^2.

C. Generalized Cross-Validation (GCV)

The Fourier series estimator depends on determining the number of oscillations (c). Therefore, one method that has been developed by researchers and can be used in determining the optimal oscillation in multivariable Fourier series

nonparametric regression is the Generalized Cross-Validation (GCV) method. Theoretically, the GCV method is asymptotically optimal, has a formula that does not contain variance σ^2 for unknown populations, and invariance to transformation [8][16]. The optimal oscillation parameter is obtained when the GCV value is the minimum. Equation (4) shows the form of the GCV formula for the multivariable Fourier series nonparametric regression model.

$$GCV(c) = \frac{MSE(c)}{\left(n^{-1}tr\left(\boldsymbol{I} - \boldsymbol{A}(c)\right)\right)^2} \tag{4}$$

where,

$$MSE(c) = n^{-1}\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2 \tag{5}$$

and

$$\boldsymbol{A}(c) = \boldsymbol{X}[\boldsymbol{X}^T\boldsymbol{X}]^{-1}\boldsymbol{X}^T \tag{6}$$

where c is oscillation parameter, $\boldsymbol{I}$ is the identity matrix with size $n \times n$, y_j is the response variable for the j^{th} observation, and $\hat{y}_j$ is the predicted value of the response variable for the j^{th} observation.

D. Coefficient of Determination (R^2)

The coefficient of determination (R^2) is one of the criteria that can be used to see how much the contribution of predictor variables can explain the response variable. Besides that, it can also be used to select the best model. Therefore, the coefficient of determination in nonparametric regression modelling can be used to achieve the objectives of regression analysis, where the goal is to get the best model. The greater the coefficient of determination obtained, the better the model obtained [17], which means that the predictor variables in the model have been able to explain the response variable. The coefficient of determination (R^2) formula is as in Equation (7).

$$R^2 = 1 - \left(\frac{\sum_{j=1}^{n}\left(\hat{y}_j - \bar{y}_j\right)^2}{\sum_{j=1}^{n}\left(y_j - \bar{y}\right)^2}\right) \tag{7}$$

where y_j is the j^{th} response variable , $\hat{y}_j$ is the predicted value of the j^{th} response variable, and $\bar{y}$ is the average of the response variable.

E. Simultaneous and Partial Hypothesis Testing

Hypothesis testing is part of a series of regression analyses whose purpose is to determine whether the predictor variables in the model have a significant effect on the response variable, where what is tested in this hypothesis testing is the model parameters. In this parameter, hypothesis testing can be done in two ways: simultaneously and partially or individually.

Simultaneous hypothesis testing is a test for regression model parameters carried out simultaneously using the F test statistic. Based on the nonparametric regression model with a multivariable Fourier series estimator in Equation (2), the hypothesis used to test the model parameters simultaneously is:

$\boldsymbol{H_0}$: $\boldsymbol{b_1} = \boldsymbol{b_2} = \cdots = \boldsymbol{b_m} = \boldsymbol{a_{11}} = \boldsymbol{a_{21}} = \cdots = \boldsymbol{a_{Cm}} = \boldsymbol{0}$

$\boldsymbol{H_1}$: there is at least one $\boldsymbol{b_i} = \boldsymbol{a_{ci}} \neq \boldsymbol{0}$, where $\boldsymbol{i} = \boldsymbol{1}, \boldsymbol{2}, \ldots, \boldsymbol{m}$; $\boldsymbol{c} = \boldsymbol{1}, \boldsymbol{2}, \ldots, \boldsymbol{C}$

with the F test statistic as in Equation (8) [14].

$$F = \frac{MSR}{MSE} = \frac{\left(\frac{\sum_{j=1}^{n}\left(\hat{y}_j - \bar{y}\right)^2}{m + Cm}\right)}{\left(\frac{\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2}{n - (m + Cm) - 1}\right)} \tag{8}$$

where $i = 1, 2, \ldots, m$ is the number of predictor variables, $c = 1, 2, \ldots, C$ are the number of oscillations, and $j = 1, 2, \ldots, n$ are the number of observations. Rejection region of H_0 that is Reject H_0 if $F > F_{(\alpha;m+Cm;n-(m+Cm)-1)}$ or p-value < α which indicates that at least one parameter has a significant effect on the response variable.

Partial hypothesis testing is a test for regression model parameters carried out individually using the t-test statistic. Based on the nonparametric regression model with a multivariable Fourier series estimator in Equation (2), the hypothesis used to test the model parameters partially is:

$\boldsymbol{H_0}$: $\boldsymbol{b_i}, \boldsymbol{a_{ci}} = \boldsymbol{0}$

H_1 : $b_i \neq 0$, $a_{ci} \neq 0$ where $i = 1, 2, \ldots, m$; $c = 1, 2, \ldots, C$

with the t-test statistic as in Equation (9) [18].

$$t = \frac{\hat{\beta}_i}{SE\left(\hat{\beta}_i\right)} \tag{9}$$

where $SE(\hat{\beta}_i)$ is the standard error of $\hat{\beta}_i$ which is obtained from $\sqrt{var(\hat{\beta}_i)}$, then $i = 1, 2, \ldots, m$ are the number of predictor variables, and $c = 1, 2, \ldots, C$ are the number of oscillations. The H_0 rejection area is to Reject H_0 if $|t| > t_{(\frac{\alpha}{2};n-(m+Cm)-1)}$ or p-value < α.

III. METHODOLOGY

This section will describe the data source, research variable, and analysis step.

A. Data Source and Research Variable

This study uses secondary data from data published by the Badan Pusat Statistik (BPS) Central Java, with an observation unit of 35 regencies/cities in Central Java Province in 2023. The variables in this study consist of five predictor variables (X) and one response variable (Y), and an explanation of these variables is presented in Table 1.

Table 1 Research Variables

Variable	Variable Symbol	Description
Response	Y	Average Length of Schooling
Predictor	X_1	Percentage of Poor Population
	X_2	Community Literacy Development Index
	X_3	Gini Ratio
	X_4	Minimum Wage
	X_5	School Participation Rate

Below are operational definitions of several research variables used:

1. Average Length of Schooling
   The average length of schooling is the average number of years spent by residents aged 25 years and over to complete all levels of formal education ever undertaken [19]. Based on this, the average length of schooling can be used to determine the quality of education in an area.
2. Percentage of Poor Population
   The percentage of poor people is the percentage of people who are below the poverty line. The percentage of poor people is those whose income or consumption is below the poverty line, namely the population group who cannot afford to buy a package of necessities. The percentage of poor people variable is related to the average length of schooling variable [20] – [22].
3. Community Literacy Development Index
   The community literacy development index measures the efforts of provincial and district/city regional governments in fostering and developing libraries as a vehicle for lifelong learning to achieve a community literacy culture. If people have high literacy and insight, it is hoped that they will become more aware of the importance of education so that this can improve the quality of learning and the average length of schooling.
4. Gini Ratio
   The Gini ratio is an index used to see inequality or to measure the degree of income inequality or other income distribution ranging from 0 to 1.
5. Minimum Wage
   The minimum wage is the lowest monthly wage, consisting of the basic wage, including fixed allowances set by the governor so that workers/laborers can get a decent living physically.
6. School Participation Rate
   The school participation rate is the proportion of the population in a certain age group at a certain level of education who are still attending school compared to those in a certain age group [19].

B. Step of Analysis

The following are the steps taken to complete the research objectives.

1.Conduct descriptive statistical analysis for all variables

2.Make a scatter plot for each predictor variable against the response variable

3.Modelling using a nonparametric regression approach with a multivariable Fourier series estimator

4.Conduct simultaneous and partial hypothesis testing.

IV. RESULTS AND DISCUSSIONS

In this section, the analysis will be conducted for nonparametric regression modeling with multivariable Fourier series estimators.

A. Descriptive Statistics

Before modeling, the first step taken in this study is to conduct descriptive statistical analysis, which aims to obtain general information from each variable used. The results of the descriptive statistical analysis have been presented in Table 2.

Table 2 Descriptive Statistics

Variable Symbol	Variable	Mean	Minimum Value	Maximum Value	Standard Deviation
Y	Average Length of Schooling	8.255	6.400	11.240	1.310
X_1	Percentage of Poor Population	10.397	4.230	16.340	3.265
X_2	Community Literacy Development Index	59.300	35.880	94.480	13.240
X_3	Gini Ratio	0.359	0.309	0.419	0.032
X_4	Minimum Wage	2198563	1958170	3060349	225260
X_5	School Participation Rate	72.680	56.220	88.390	7.340

The second step before modeling is to make a scatter plot which aims to see the pattern of the relationship between the response variable (Y) and each predictor variable (X). The scatter plot between the Average Length of Schooling variable and each predictor variable that is thought to affect it is shown in Figure 1.

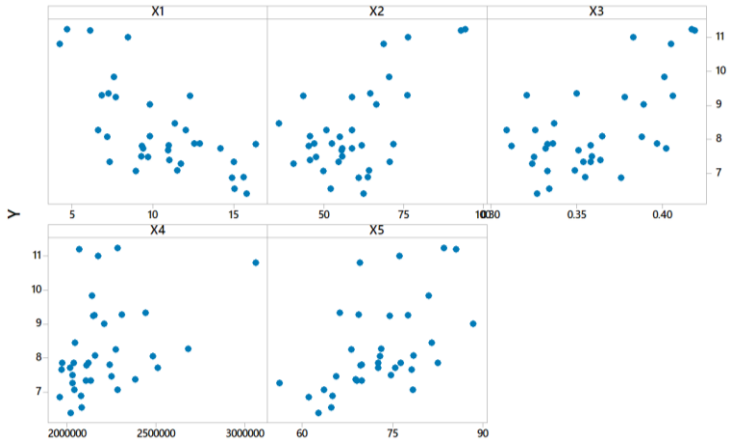

Figure 1 Scatter Plot Between Response Variable (Y) and Predictor Variables (X)

Based on Figure 1, the scatterplot between the response variable (Y) and each predictor variable (X) does not appear to follow linear, quadratic, cubic, and other parametric patterns, so nonparametric is an alternative that can be used and is assumed to follow the Fourier Series estimator. This is because the resulting scatter plot shape shows a tendency of repeating patterns, where the shape of this pattern is characteristic of the Fourier series estimator. Based on this, this study uses a nonparametric regression approach with a multivariable Fourier series estimator to model the problem of the Average Length of Schooling in Central Java Province in 2023.

The Fourier Series Estimator is part of nonparametric regression, where this estimator is one of the approaches/approximations that can be used in nonparametric regression. Based on this, as long as the form of the relationship pattern produced between each predictor variable and the response variable is not known with certainty and clarity, you can still use the Fourier Series estimator.

B. Nonparametric Regression Modeling with Multivariable Fourier Series Estimator

This section uses nonparametric regression modeling with a multivariable Fourier series estimator to model the problem of the Average Length of Schooling in Central Java Province in 2023. In this nonparametric Fourier series regression modeling, it is necessary to consider the number of oscillations used. Oscillation is a constituent component of the Fourier series, which refers to the periodic change of the cosine wave in the model. In this study, the number of oscillations used is limited to one to four oscillations, where each predictor variable will have the same number of oscillations. The number of oscillations used can be more than four, but if too many oscillations are used, the resulting

model does not follow the principle of parsimony because there will be too many parameters that need to be estimated.

The results of modeling the Average Length of Schooling using a nonparametric regression approach with a multivariable Fourier series estimator are presented in Table 3.

Table 3 GCV Value of Nonparametric Regression of Multivariable Fourier Series Estimator

Number of Oscillations	GCV	R^2
1	2.662	84.226%
2	1.637	84.287%
3	1.027	85.464%
4	12.946	70.505%

Based on Table 3, it can be seen that the minimum GCV value is obtained when the number of oscillations used is three oscillations, with a minimum GCV value of 1.027.

The parameter estimation results of the nonparametric regression modeling of the multivariable Fourier series estimator with three oscillations are shown in Table 4.

Table 4 Parameter Estimation Results of Nonparametric Regression of Multivariable Fourier Series Estimator

Parameter	Parameter Estimation	Parameter	Parameter Estimation	Parameter	Parameter Estimation
$\hat{a}_0$	15.185				
$\hat{b}_1$	-0.158	$\hat{b}_3$	1.789	$\hat{b}_5$	0.059
$\hat{a}_{11}$	14.294	$\hat{a}_{13}$	-41.867	$\hat{a}_{15}$	51.145
$\hat{a}_{21}$	-31.777	$\hat{a}_{23}$	258.627	$\hat{a}_{25}$	5.210
$\hat{a}_{31}$	-48.867	$\hat{a}_{13}$	855.552	$\hat{a}_{35}$	-4.418
$\hat{b}_2$	0.020	$\hat{b}_4$	$7.389e^{-7}$		
$\hat{a}_{12}$	-15.598	$\hat{a}_{14}$	-19.031		
$\hat{a}_{22}$	-4.451	$\hat{a}_{24}$	0.170		
$\hat{a}_{32}$	12.265	$\hat{a}_{34}$	-19.031		

Based on the estimation results in Table 4, the form of the nonparametric regression model with the multivariable Fourier series estimator can be written as follows:

$$\hat{y}_j = 15.185 - 0.158x_{1j} + 14.294cosx_{1j} - 31.777cos2x_{1j} - 48.867cos3x_{1j} + 0.020x_{2j} - 15.598cosx_{2j} - 4.451cos2x_{2j} + 12.265cos3x_{2j} + 1.789x_{3j} - 41.867cosx_{3j} + 258.627cos2x_{2j} + 855.552cos3x_{3j} + 7.389e^{-7}x_{4j} - 19.031cosx_{4j} + 0.170cos2x_{4j} - 19.031cos3x_{4j} + 0.059x_{5j} + 51.145cosx_{5j} + 5.210cos2x_{5j} - 4.418cos3x_{5j}$$

where, $j = 1, 2, \dots, n$.

From the multivariable Fourier series estimator nonparametric regression model with 3 oscillations formed above, the coefficient of determination (R^2) is 85.464%.

C. *Simultaneous Hypothesis Testing*

After modeling and obtaining the best model with minimum GCV, namely the multivariable Fourier series estimator nonparametric regression model with 3 oscillations, the next step is to test the hypothesis of the best model parameters simultaneously. Simultaneous hypothesis testing is used to see whether the predictor variables used simultaneously significantly affect the Average Length of Schooling Variable or not. In this simultaneous hypothesis testing using the *F* test statistic, with the following hypothesis:

H_0 : $b_1 = b_2 = \cdots = b_m = a_{11} = a_{21} = \cdots = a_{Cm} = 0$

H_1 : there is at least one $b_i = a_{ci} \neq 0$; $i = 1, 2, \dots, 5$; $c = 1, 2, 3$

The significance level used is 0.05 or 5%, and the $\boldsymbol{H_0}$ rejection area is Reject $\boldsymbol{H_0}$ if $\boldsymbol{F > F_{(\alpha;m+Cm;n-(m+Cm)-1)}}$. The results of simultaneous hypothesis testing are presented in the ANOVA table as in Table 5 below.

Table 5 ANOVA

Source	df	Sum of Square	Mean of Square	F_{hit}
Regression	20	49.869	2.493	4.116
Error	14	8.481	0.606	
Total	34	58.351		

Based on the ANOVA results in Table 5 above, a value of $\boldsymbol{F}$ by 4.116 where the value is greater than the value of $\boldsymbol{F_{(0.05,20,14)}}$ that is 2.388, and produces a p-value of 0.005 which is smaller than $\boldsymbol{\alpha = 0.05}$. Therefore, it can be decided that Reject $\boldsymbol{H_0}$, which means there is at least one significant parameter. The predictor variables used in this modeling simultaneously significantly affect the response variable, in this case, the Average Length of Schooling variable.

Table 6 *t*-Test

Variable	Parameter	t_{hit}	p-value	Decision	Conclusion
Constant	$\hat{a}_0$	30.845	$2.844e^{-14}$	Reject H_0	Significant
X_1	$\hat{b}_1$	-0.295	0.772	Failure to Reject H_0	Not Significant
	$\hat{a}_{11}$	26.368	$2.466e^{-13}$	Reject H_0	Significant
	$\hat{a}_{21}$	-55.714	$7.722e^{-18}$	Reject H_0	Significant
	$\hat{a}_{31}$	-85.672	$1.900e^{-20}$	Reject H_0	Significant
X_2	$\hat{b}_2$	0.036	0.972	Failure to Reject H_0	Not Significant
	$\hat{a}_{12}$	-26.894	$1.881e^{-13}$	Reject H_0	Significant
	$\hat{a}_{22}$	-8.747	$4.772e^{-7}$	Reject H_0	Significant
	$\hat{a}_{32}$	24.806	$5.703e^{-13}$	Reject H_0	Significant
X_3	$\hat{b}_3$	3.262	$5.680e^{-3}$	Reject H_0	Significant
	$\hat{a}_{13}$	-68.559	$4.272e^{-19}$	Reject H_0	Significant
	$\hat{a}_{23}$	528.866	$1.648e^{-31}$	Reject H_0	Significant
	$\hat{a}_{13}$	1846.326	$4.128e^{-39}$	Reject H_0	Significant
X_4	$\hat{b}_4$	$1.219e^{-6}$	0.999	Failure to Reject H_0	Not Significant
	$\hat{a}_{14}$	-34.034	$7.295e^{-15}$	Reject H_0	Significant
	$\hat{a}_{24}$	0.405	0.691	Failure to Reject H_0	Not Significant
	$\hat{a}_{34}$	-42.509	$3.336e^{-16}$	Reject H_0	Significant
X_5	$\hat{b}_5$	0.112	0.912	Failure to Reject H_0	Not Significant
	$\hat{a}_{15}$	102.729	$1.502e^{-21}$	Reject H_0	Significant
	$\hat{a}_{25}$	11.046	$2.687e^{-8}$	Reject H_0	Significant
	$\hat{a}_{35}$	-6.404	$1.644e^{-5}$	Reject H_0	Significant

D. Partial Hypothesis Testing

The next hypothesis test is partial hypothesis testing, which aims to see the parameters of the best model with a significant effect. In this partial hypothesis testing using the *t*-test statistic, with the following hypothesis:

$\boldsymbol{H_0} : \boldsymbol{b_i, a_{ci}} = \boldsymbol{0}$

$\boldsymbol{H_1} : \boldsymbol{b_i} \neq \boldsymbol{0}\,, \boldsymbol{a_{ci}} \neq \boldsymbol{0}\ ;\ \boldsymbol{i} = \boldsymbol{1, 2, \dots, 5}\ ;\ \boldsymbol{c} = \boldsymbol{1, 2, 3}$

The significance level used is 0.05 or 5% and the $\boldsymbol{H_0}$ rejection area is Reject $\boldsymbol{H_0}$ if $|\boldsymbol{t}| > \boldsymbol{t}_{(\frac{\alpha}{2};\boldsymbol{n-(m+Cm)-1})}$. Then the results of partial hypothesis testing are presented in Table 6. Based on the *t*-test results in Table 6, out of 21 parameters, there are only 5 parameters that are not significant, or it can be said that most of the parameters contained in the best model are significant. Based on this, all predictor variables used in this study significantly affect the response variable, namely the Average Length of Schooling in Regencies/Cities in Central Java Province.

V. CONCLUSIONS AND SUGGESTIONS

This section will explain the conclusion and suggestions for future research based on the modeling results to hypothesis testing. The conclusions of this research are as follows:

1. The application of a nonparametric regression model with a multivariable Fourier series estimator has been successfully conducted. The best model selection is obtained based on the minimum GCV value and supported by the highest coefficient of determination (R^2) value. The best model obtained is a model with three oscillations, with the following model form:

$$\begin{aligned}\hat{y}_j = {} & \boldsymbol{15.185 - 0.158x_{1j} + 14.294cosx_{1j} - 31.777cos2x_{1j} - 48.867cos3x_{1j} + 0.020x_{2j} - 15.598cosx_{2j}} \\ & \boldsymbol{-4.451cos2x_{2j} + 12.265cos3x_{2j} + 1.789x_{3j} - 41.867cosx_{3j} + 258.627cos2x_{2j} + 855.552cos3x_{3j}} \\ & \boldsymbol{+7.389e^{-7}x_{4j} - 19.031cosx_{4j} + 0.170cos2x_{4j} - 19.031cos3x_{4j} + 0.059x_{5j} + 51.145cosx_{5j}} \\ & \boldsymbol{+5.210cos2x_{5j} - 4.418cos3x_{5j}}\end{aligned}$$

   where $\boldsymbol{j} = \boldsymbol{1, 2, \dots, n}$, with a GCV value of 1.027 and R^2 of 85.464%, which means that the predictor variables contained in the model can explain the response variable by 85.464%, while other variables outside the model influence the remaining 14.536%.
2. Based on simultaneous and partial hypothesis testing, it can be said that the Percentage of Poor Population, Community Literacy Development Index, Gini Ratio, Minimum Wage, and School Participation Rate simultaneously and partially have a significant effect on Average Years of Schooling in Regencies/Cities in Central Java in 2023.

Nonparametric regression is very flexible in finding the form of an estimated regression curve but has minimal interpretation. So, in this case, it is impossible to interpret the model formed and can only estimate the regression curve. This means the best estimate for y $(\hat{\boldsymbol{y}})$ is obtained from the nonparametric regression model with the Fourier Series estimator.

Based on a series of studies that have been carried out, the suggestions that can be conveyed for further research are the number of oscillations that are tried can be different for each predictor variable, this is because the shape of the relationship pattern between the response variable and each predictor variable has different characteristics.

ACKNOWLEDGEMENT

This research was funded by YPPI Rembang University, Rembang, Indonesia.

REFERENCES

[1] Badan Pusat Statistik Indonesia, "[Metode Baru] Rata-Rata Lama Sekolah," 2023, accessed on March 4th 2024, from https://www.bps.go.id/id/statistics-table/2/NDE1IzI%253D/-metode-baru--rata-rata-lama-sekolah.html

[2] BPS Provinsi Jawa Tengah, "Provinsi Jawa Tengah dalam Angka 2024," vol. 49, BPS Provinsi Jawa Tengah: Jawa Tengah, 2024, pp. 467.

[3] BPS Provinsi Jawa Tengah, Statistik Daerah Provinsi Jawa Tengah 2023. Jawa Tengah: Badan Pusat Statistik Provinsi Jawa Tengah, 2023.

[4] H. Nurcahyani, I. N. Budiantara, and I. Zain, "The Curve Estimation of Combined Truncated Spline and Fourier Series Estimators for Multiresponse Nonparametric Regression," Mathematics, vol. 9, no. 1141, pp. 1–22, 2021.

[5] B. W. Silverman, "Some Aspects of The Spline Smoothing Approach to Non-parametric Regression Curve Fitting," Journal of the Royal Statistical Society, Series B (Methodological), vol. 47, no. 1, pp. 01–52, 1985.

[6] R. L. Eubank, Nonparametric Regression and Spline Smoothing. New York: Marcel Dekker, 1999.

[7] M. Bilodeau, "Fourier Smoother and Additive Models," The Canadian Journal of Statistics, vol. 20, no. 3, pp. 257–259, 1992.

[8] Y. Wang, "Smoothing Spline Models with Correlated Random Errors," Journal of the American Sttaistical Association, vol. 93, pp. 341–348, 1998.

[9] A. Prahutama, "Model Regresi Nonparametrik dengan Pendekatan Deret Fourier pada Kasus Tingkat Pengangguran Terbuka di Jawa Timur," Prosiding Seminar Nasional Statistika Undip, vol. 10, pp. 69–76, 2013.

[10] N. P. A. M. Mariati, I. N. Budiantara, and V. Ratnasari, "Modeling Poverty Percentages in the Papua Islands using Fourier Series in Nonparametric Regression Multivariable," Journal of Physics: Conference Series, vol. 1397, no. 1, pp. 1–7, 2019.

[11] N. Y. Adrianingsih, A. T. R. Dani, and A. Ainurrochmah, "Pemodelan dengan Pendekatan Deret Fourier pada Kasus Tingkat Pengangguran Terbuka di Nusa Tenggara Timur," Prosiding Seminar Edusaintech, vol. 4, pp. 400–407, 2020.
[12] M. A. D. Octavanny, I. N. Budiantara, H. Kuswanto, and D. P. Rahmawati, "Modeling of Children Ever Born in Indonesia Using Fourier Series Nonparametric Regression, " Journal of Physics: Conference Series, vol. 1752, no. 1, pp. 1–7, 2021.
[13] A. T. R. Dani, A. F. Dewi, and L. Ni'matuzzahroh, "Studi Simulasi dan Aplikasi: Estimator Deret Fourier pada Pemodelan Regresi Nonparametrik," Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya, vol. 2, pp. 279–288, 2022.
[14] Rahmania, Sifriyani, and A. T. R. Dani, "Modeling Open Unemployment Rate in Kalimantan Island using Nonparametric Regression with Fourier Series Estimator," Barekeng: jurnal ilmu matematika dan terapan, vol. 18(1), pp. 0245–0254, 2024.
[15] L. J. Asrini and I. N. Budiantara, "Fourier Series Semiparametric Regression Models (Case Study: The Production of Lowland Rice Irrigation in Central Java," ARPN Journal of Engineering and Applied Sciences, vol. 9, no. 9, pp. 1501–1506, 2014.
[16] G. Wahba and Y. Wang, "Spline Function," Encyclopedia of Statistical Sciences, pp. 1–27, 2014.
[17] A. T. Damaliana, I. N. Budiantara, and V. Ratnasari, "Comparing Between mGCV and aGCV Methods to Choose the Optimal Knot Points in Semiparametric Regression with Spline Truncated Using Longitudinal Data," IOP Conference Series: Materials Science and Engineering, vol. 546(3), pp. 1–10, 2019.
[18] J. Racine, "Consistent Significance Testing for Nonparametric Regression," Journal of Business & Economic Statistics, vol. 15, no. 3, pp. 369–378, 1997.
[19] Badan Pusat Statistik Provinsi Jawa Tengah, Statistik Pendidikan Provinsi Jawa Tengah. Jawa Tengah: Badan Pusat Statistik Provinsi Jawa Tengah, 2022.
[20] L. Ni'matuzzahroh and A. T. R. Dani, "Pemodelan Rata-Rata Lama Sekolah di Provinsi Nusa Tenggara Timur (NTT) Menggunakan Pendekatan Regresi Nonparametrik Spline Truncated," Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya, vol. 2, pp. 289–301, 2022.
[21] F. A. Suhendar, R. V. Sari, T. Pangesti, Z. M. G. Putra, and A. P. A. Santoso, "The Impact of Poverty in Indonesia on Education," Jurnal Ilmu Sosial dan Pendidikan (JISIP), vol 8, no. 2, pp. 1119–1125, 2024.
[22] D. Suryadarma and A. Suryahadi, The Contrasting Role of Ability and Poverty on Education Attainment: Evidence from Indonesia. Jakarta: SMERU Research Institute, 2009.