KnE Social Sciences

The 4th International Conference on Science, Mathematics, Environment, and Education

KnE Publishing

Research Article

Selecting Optimal Knot Points and Oscillation Parameters Using Generalized Cross-validation and Unbiased Risk Method in Nonparametric Regression of Combined Spline Truncated and Fourier Series

Putri Kusuma Wardani*, I Nyoman Budiantara, and Setiawan Setiawan

Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Sukolilo, Indonesia

Abstract.
A nonparametric regression approach is suitable for the cases in which the shape of the pattern between the response variables and the predictor variables is not known. There are several methods in nonparametric regression, such as spline truncated and Fourier series. In both methods, determining the optimal knot point is crucial. Optimal knot points and oscillation parameters can be selected using the generalized cross-validation (GCV) and unbiased risk (UBR) methods. This study aimed to examine the GCV and UBR methods to select optimal knot point and oscillation parameters on the data on Indonesia's economic growth rate in 2022. The estimation method used is ordinary least square (OLS). The results obtained used the GCV method because it has MSE value 1.42, which is smaller than MSE of UBR method of 10.614. The coefficient of determination for the GCV method is 89.34%. The optimal number of knot points and oscillation parameters are three and three for nonparametric regression estimator combined of spline truncated and Fourier series.

Keywords: fourier series, generalized cross-validation, nonparametric regression, spline truncated, unbiased risk

Corresponding Author: Putri Kusuma Wardani; email: putrikusumaw.19@gmail.com

Published 27 May 2025

Publishing services provided by Knowledge E

Selection and Peer-review under the responsibility of the ICoSMEE 2023 Conference Committee.

1. INTRODUCTION

The statistical method that is often used to analyze two or more variables to investigate the nature of the relationship is regression analysis. Regression analysis can be applied to data that has correlation between variables. Therefore, before carrying out a regression analysis, it is better to investigare whether the variables used have a correlation or not. The main objective in carrying out a regression analysis is to find the shape of the estimated regression curve (1).

How to cite this article: Putri Kusuma Wardani, I Nyoman Budiantara, and Setiawan Setiawan, (2025), "Selecting Optimal Knot Points and Oscillation Parameters Using Generalized Cross-validation and Unbiased Risk Method in Nonparametric Regression of Combined Spline Truncated and Fourier Series" in *The 4th International Conference on Science, Mathematics, Environment, and Education*, KnE Social Sciences, pages 220–234. DOI 10.18502/kss.v10i11.18744

There are several approaches developed by research based on the shape of the regression curve, namely parametric regression, nonparametric regression, and semi-parametric regression. The advantage of using a nonparametric approach it has high flexibility, meaning that the data is expected to be able to find its own form of estimation (2). There are several estimators that have been widely used to model regression with a nonparametric approach, namely Kernel, Wavelet, Fourier Series, Spline, and Multivariate Adaptive Regression Spline (MARS). There are two approaches to obtaining estimators, namely Penalized Least Square (PLS) and Ordinary Least Square (OLS). If the estimator obtained is based on PLS, then the main problem of the estimator is a selection of optimal smoothing parameters. Meanwhile, if the estimator is obtained based on OLS, then the main problem the estimator is selection of optimal knot points or oscillation parameters.

The nonparametric regression models currently being developed are limited to only one form of estimator for all predictor variables used. This happens because of the researcher's assumption that each of these predictors has the same data pattern. If forced to use only one estimator, it will produce an estimate that does not match the data pattern and will result in poor accuracy. The use of combined estimators can be applied in various fields, including in the social and population fields. Previous research using combined estimators included conducting spline truncated and Fourier series regression modeling in multivariable nonparametric regression on poverty data in Papua Province which concluded that the spline truncated method was better than the Fourier series method (3), developed a combined kernel and Fourier series estimator on data on the percentage of poor people in Indonesia (4), developed a combined kernel, truncated spline and Fourier series estimator using respond data (5).

Determination of the knot points and oscillation parameters in nonparametric regression spline truncated and Fourier series will greatly affect the regression curve that will be formed. Several methods can be used to determine optimal knot points and oscillation parameters. These methods are Cross-Validation (CV), Genearalized Cross-Validation (GCV), Unbiased Risk (UBR), and Generalized Maximum Likelihood (GML). This study only discussed the GCV and UBR method.

A research review of the CV, GCV, and UBR methods was carried out comparing the GCV and UBR methods in selecting optimal knot points in spline nonparametric regression with the best results using the GCV method (6). Compared the CV and UBR methods in selecting optimal knot points in multivariable spline nonparametric regression with the best results using the CV method (7). Selected optimal knot points and bandwidth using the CV, GCV, and UBR methods in nonparametric regression modeling of combined spline truncated estimators and Gaussian kernels with the best results using the GCV method (8). Compared the GCV and CV methods in selecting optimal oscillation parameters in nonparametric Fourier series regression (9).

Nonparametric regression models can be applied in various scientific fields, one of which is economics. In this research, a nonparametric regression model is applied to economic growth in Indonesia. Economic growth is a positive change in the level of production of goods and services which is influenced by production factors such as labor, land and capital which will change over time (10). Economic growth is also the main indicator of societal progress where economic growth is believed to be a reflection of how a nation's economy is running, where every country tries to achieve higher economic growth to provide a higher standard of living for its citizens.

From the description above, there has been no research that examines the selection of knot points and optimal oscillation parameters in a combined estimator model between spline truncated nonparametric regression and Fourier series. Therefore, researchers will conduct research using a case study of the rate of economic growth in Indonesia in 2022 in a nonparametric regression model of a combined spline truncated and Fourier series estimator with the comparison methods used, namely Generalized Cross-Validation (GCV) and Unbiased Risk (UBR) because It is not yet known whether the data that will be used is Gaussian or non-Gaussian data.

1.1. Nonparametric Regression

If given data $\left(x_{1i}, x_{2i}, \ldots, x_{pi}, y_i\right)$ and assumed relationship between the response variable and the predictor variable follow a nonparametric regression model.

$y_i = f\left(x_{1i}, x_{2i}, \ldots, x_{pi}\right) + \epsilon_i, i = 1, 2, \ldots, n$ (1)

where,

y_i = response variable at the i observation

x = predictor variable

p = many predictor variables

n = many observation

ϵ = random error

1.2. Spline Truncated Nonparametric Regression

Nonparametric regression has a high degree of flexibility. Spline Truncated nonparametric regression has knot points which are fusion points that show changes in curve behavior at different intervals (11). The advantage of Spline Truncated is that this model tends to find its own data estimates wherever the data pattern moves. This advantage occurs because in the Spline Truncated function there are knot points (1). The spline function f$\left(h_i\right)$ has order m with knots $\left(K_1, K_2, \ldots, K_m\right)$ can be expressed as follow equation (2).

$f\left(h_i\right) = \sum_{j=0}^{q} \theta_j h_i^j + \sum_{k=1}^{m} \theta_{(q+k)} \left(h_i - K_k\right)_+^q + \epsilon_i$ (2)

with truncated function

$$\left(h_i - K_k\right)_+^q = \begin{cases} \left(h_i - K_k\right)^q & , h_i \geq K_k \\ 0 & , h_i \leq K_k \end{cases}$$

Spline truncated nonparametric regression model can be written in the following equation (3).

$y_i = \sum_{l=1}^{p} \left[\sum_{j=0}^{q} \theta_j h_i^j + \sum_{k=1}^{m} \theta_{(q+k)} \left(h_i - K_k\right)_+^q + \epsilon_i\right]$ (3)

$\left(K_1, K_2, \ldots, K_m\right)$ are knot points that show patterns of behavior change of the function at different subintervals, q is the degree of the polynomial, and θ is a parameter that contains the spline function.

1.3. Fourier Series Nonparametric Regression

This Fourier series is used to estimate the regression curve which shows sin and cos waves. Fourier series nonparametric regression model can be written in the following equation (4).

$$y_i = \sum_{w=1}^{t} \left(b_w z_{wi} + \frac{1}{2} a_{0w} + \sum_{s=1}^{S} a_{ws} \cos s z_{wi}\right) + \epsilon_i$$
$$y_i = b_0 + \sum_{w=1}^{t} \left(b_w z_{wi} + \sum_{s=1}^{S} a_{ws} \cos s z_{wi}\right) + \epsilon_i \quad (4)$$

S is the number of oscillation parameters and b_w, a_{0w}, a_{1w}, ..., a_{sw} is an unknown parameter.

1.4. Combined Estimators Nonparametric Regression

Combined nonparametric regression is a multipredictor nonparametric regression whose regression curve is additive where the regression curve is approach bt two or more types of nonparametric regression curves. If given data $(y_i, h_{1i}, h_{2i}, \ldots, h_{pi}, z_{1i}, z_{2i}, \ldots, z_{ti})$, where i = 1, 2,..., n with $h_{1i}, h_{2i}, \ldots, h_{pi}, z_{1i}, z_{2i}, \ldots, z_{ti}$ as the predictor variable and y_i as the response variable which is assumed to follow the nonparametric regression of combined estimators spline truncated and Fourier series model as in equation (5).

$$y_i = \mu\left(h_{1i}, h_{2i}, \ldots, h_{pi}, z_{1i}, z_{2i}, \ldots, z_{ti}\right) + \epsilon_i \quad (5)$$

Nonparametric regression of combined spline truncated and Fourier series model is shown in equation (6).

$$y_i = c + \sum_{l=1}^{p}\left[\sum_{j=0}^{q}\theta_j h_i^j + \sum_{k=1}^{m}\theta_{(q+k)}\left(h_i - K_k\right)_+^q\right] + \left[\sum_{w=1}^{t} b_w z_{wi} + \frac{1}{2}a_0 + \sum_{s=1}^{S} a_{ws}\cos s z_{wi}\right] + \epsilon_i \quad (6)$$

where $c = \theta_0 + b_0$ is a model constant and random error (ϵi) assumed to be identical, independent, and normally distributed with zero mean and variance σ2.

The combined estimator nonparametric regression of spline truncated and Fourier series is obtained using the Ordinary Least Square (OLS) method as shown in equation (7).

$$\vec{\beta} = \left(X^T X\right)^{-1} X^T \vec{y} \quad (7)$$

The estimation curve of the combined estimators nonparametric regression of spline truncated and Fourier series estimators is as follows.

$$\vec{\mu}(h, z) = X\vec{\beta}$$

$$\vec{\mu}(h, z) = \left(X^T X\right)^{-1} X^T \vec{y}$$

so, obtained

$$\vec{\mu}(h, z) = A\vec{y}$$

where $\vec{\mu}(h, z)$ is a combined curve of spline truncated and Fourier series and A is a combined function matrix of knot points and oscillation parameters.

1.5. Selecting Optimal Knot Points and Oscillation Parameters

Selecting optimal knot points and oscillation parameters in nonparametric regression combined estimator spline truncated and Fourier series can use Generalized Cross-Validation (GCV) and Unbiased Risk (UBR).

Generalized Cross-Validation (GCV). Generalized Cross Validation (GCV) is a method that is very often used to select optimal knot points and oscillation parameters. The selection of the best and optimal model is obtained from the minimum GCV value. The GCV method is a modified form of the Cross-Validation (CV) method. GCV is a generalization of the CV method which has been weighted with the CV equation shown in equation (8).

$CV = n^{-1} \sum_{i=1}^{n} \left(\frac{y_i - \hat{\mu}(h,z)}{1-A} \right)$ (8)

Given the GCV function in equation (9).

$GCV = \frac{MSE}{\left[n^{-1}\ trace\ I-A\right]}$ (9)

where I is the identity matrix, n is the number of observations, and A is the matrix for estimating the estimator. MSE is described in the following equation (10).

$MSE = n^{-1} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2$ (10)

Equation (8) after being weighted is written as in equation (11).

$GCV = n^{-1} \sum_{i=1}^{n} \left(y_i - \hat{\mu}(h,z) \right)^2 \frac{1-A}{\left[n^{-1}\ trace\ I-A\right]}$ (11)

So in general, Wu and Zhang (12) make the GCV equation as follows.

$GCV = \frac{n^{-1} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2}{\left[n^{-1}\ trace\ I-A\right]}$

$GCV = \frac{n^{-1}\left[\vec{Y}(I-A)^T(I-A)\right]}{\left[n^{-1}\ trace\ I-A\right]}$ (12)

Unbiased Risk. The Unbiased Risk method is a method that can also be used to measure the selection of knot points and optimal oscillation parameters in a mixed truncated spline and Fourier series estimator by adding information about σ^2. Unbiased Risk refers to a method for evaluating and comparing models using an unbiased or unbiased risk assessment. Given the loss function equation (13).

$L = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{\mu}_i(h,z) - \mu_i \right)$ (13)

The definition of the risk function which is the expected loss function is shown in equation (14).

$R = E(L) = E\left(\frac{1}{n} \sum_{i=1}^{n} \left(\hat{\mu}_i(h,z) - \mu_i \right)^2 \right)$ (14)

E is the expected value.

$$R = \frac{1}{n} E \left\| A\left(\vec{\mu} + \vec{\epsilon}\right) - \vec{\mu} \right\|^2$$
$$R = \frac{1}{n} E \left\| \left(A\vec{\mu} + A\vec{\epsilon}\right) - \vec{\mu} \right\|^2$$
$$R = \frac{1}{n} \left\| I - A\vec{\mu} \right\|^2 + 0 + \frac{1}{n} E\left(\vec{\epsilon}^T A^T A \vec{\epsilon}\right) \quad (15)$$
$$R = \frac{1}{n} \left\| I - A\vec{\mu} \right\|^2 + \frac{\sigma^2}{n} trace\left(A^T A\right)$$

Next, it will be proven whether $E\left(\hat{R}\right) = R$ where $\hat{R}$ is the Unbiased Risk (UBR) criterion where the value of $\hat{R}$ is defined in equation (16).

$$\hat{R} = \frac{1}{n}\vec{y}^T \left(I - A^T \left(I - A\right)\vec{y} - \frac{\sigma^2}{n} trace\left(I A^T \left(I A\right)\right) + \frac{\sigma^2}{n} trace\left(A^T A\right)\right) \quad (16)$$

Then we will look for the expected value of the UBR criteria which is shown in the following equation.

$$E\left(\hat{R}\right) = \frac{1}{n}\vec{\mu}^T \left(\left(I A^T\right)\left(I A\right)\vec{\mu} + \frac{\sigma^2}{n} trace\left(A^T A\right)\right) \quad (17)$$
$$E\left(\hat{R}\right) = \frac{1}{n}\left\| I A \vec{\mu} \right\|^2 \frac{\sigma^2}{n} trace\left(A^T A\right)$$

The UBR formula is given in equation (18).

$$UBR = \frac{1}{n}\left|\left(I - A\right)\vec{y}\right|^2 \frac{\hat{\sigma}^2}{n} trace\left[I - A\right] + \frac{\hat{\sigma}^2}{n} trace\left[A\right]^2 \quad (18)$$

where I is the identity matrix, n is the number of observations, and A is the matrix for estimating the estimator $\hat{\sigma}^2$ is described in equation (19).

$$\hat{\sigma}^2 = \frac{\left|\left(I - A\right)\vec{y}\right|^2}{trace\left[\left|\left(I - A\right)\vec{y}\right|\right]} \quad (19)$$

1.6. Criteria for Selection of The Best Model

One of the goals of regression analysis is to obtain the best model that is able to explain the relationship between response variables and predictor variables based on certain criteria. One of the criteria used in selecting the best model is to use the Mean Square Error (MSE) value with the formula used shown in equation (20).

$$MSE = n^{-1} \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \quad (20)$$

with n indicating the number of observation data. MSE is a criterion for the goodness of a model that has a non-negative value. The smaller the MSE value produced by a model, the better the model.

Apart from MSE, the best model selection criteria can also be used coefficient of determination R^2, where the higher the value of the coefficient of determination R^2 produced by a model, the better the predictor variables in the model are at explaining the variability of the response variable with the coefficient of determination R^2 formula shown in equation (21).

$$R^2 = 1 - \frac{\text{Sum Square Error}}{\text{Sum Square Total}} = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2} \quad (21)$$

2. METHOD

The data used in this study is secondary data obtained from the publications of the Statistics Indonesia. The observation unit covers 34 provinces Indonesia in 2022 (Table 1).

TABLE 1: Research Variable.

Variable	Variable Type	Information
y	Response	Growth Rate of Gross Regional Domestic Product
h_1	Predictor	Labour Force Participation Rate
h_2	Predictor	Mean Years Schooling
h_3	Predictor	An Improved Drinking Water
z_1	Predictor	Domestic Direct Investment

The analysis step in this study is step (1) to obtain the nonparametric regression of combined estimators spline truncated and Fourier series and obtain optimization of the GCV and UBR methods. (2) Make a scatterplot to determine variables with spline truncated and Fourier series. (3) Obtain optimum knot points and oscillation parameters using GCV and UBR methods as well MSE value and coefficient of determination (R^2) for comparison. (4) Create model and scenarios from the output results and make conclusions and suggestions.

3. RESULTS AND DISCUSSIONS

3.1. GCV and UBR Methods for Combined Estimators

The GCV method formula in equation (9) will be modified in the form of a combined estimator written in equation (22).

$$GCV_{(K,S)} = \frac{MSE_{(K,S)}}{\left[n^{-1}\ trace\ I - A_{(K,S)}\right]} \quad (22)$$

Equation (18) shows the UBR formula used to find optimal knot points and oscillation parameters.

$$UBR_{(K,S)} = \frac{1}{n}\left|\left(I - A_{(K,S)}\right)\vec{y}\right|^2 \frac{\hat{\sigma}^2}{n} trace\left[I - A_{(K,S)}\right] + \frac{\hat{\sigma}^2}{n} trace\left[A_{(K,S)}\right]^2 \quad (18)$$

where,

$$A_{(K,S)} = X\left(X^TX\right)^{-1}X^T \qquad \hat{\vec{\mu}}_{(h,z)} = X\left(X^TX\right)^{-1}X^T\vec{y}$$

$$MSE_{(K,S)} = n^{-1}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \qquad = A(K,S)\vec{y}$$

$$= n^{-1}\sum_{i=1}^{n}\left(y_i - \hat{f}(h,z)\right)^2$$

I is the identity matrix, $A_{(K,\ S)}$ is a combined function matrix of knot points (K) and oscillation parameters (S), n is the number of observations, K is the optimal knot point, S is the optimal oscillation parameter.

3.2. Application to Data On The Rate of Economic Growth in Indonesia

In this section we will discuss the selection of knot points and optimal oscillation parameters using the GCV and UBR methods using many 1, 2, 3 knot points and 1, 2, 3 oscillation parameters applied to data on provincial economic growth rates Indonesia in 2022. Before carrying the analysis, it os necessary to pay attention to the data patterns as shown in the following figure (Figure 1).

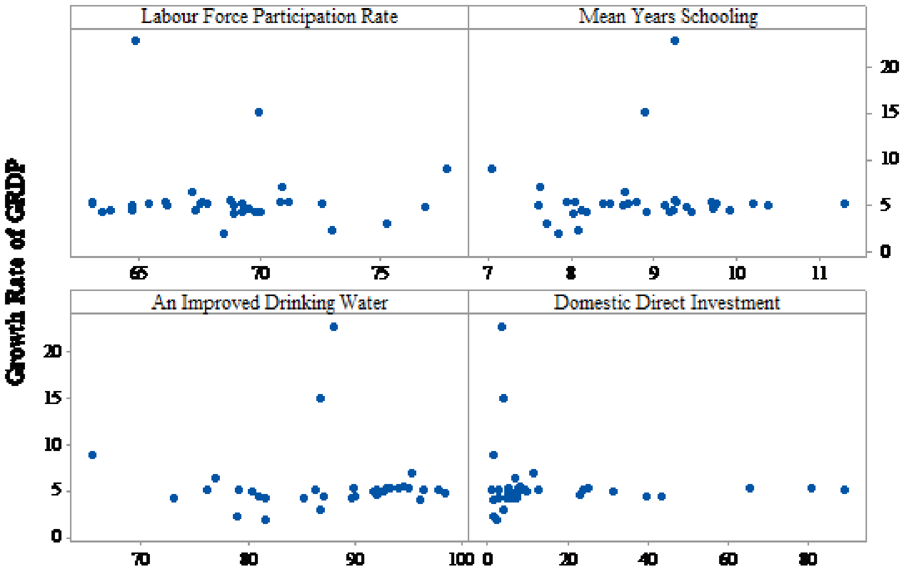

Figure 1: Scatterplot of Growth Rate GRDP with LFPR, MYS, AIDW, and DDI.

Figure 1 shows that the relationship pattern between the Growth Rate of GRDP variable and the Labor Force Participation Rate tends not to form a particular pattern.

The relationship pattern appears to experience changes in data behavior at certain sub-intervals, so that the Labor Force Participation Rate variable will be approached with a truncated spline estimator. The relationship between the Growth Rate of GRDP variable and Mean Years of Schooling tends not to form a particular pattern. The relationship pattern appears to experience changes in data behavior at certain sub-intervals, so that the Mean Years Schooling variable will be approached with a truncated spline estimator. The relationship pattern between the Growth Rate of GRDP variable and An Improved Drinking Water tends not to form a particular pattern. The relationship pattern appears to experience changes in data behavior at certain sub-intervals, so that the variable An Improved Drinking Water will be approached with a truncated spline estimator. The relationship pattern between the Growth Rate of GRDP variable and Domestic Direct Investment at first glance forms a repeating pattern with an upward trend. Therefore, the Domestic Direct Investment variable will be approached using a Fourier series estimator.

3.3. Selecting Optimal Knot Points and Oscillation Parameters Using GCV and UBR

Table 2 show the results of the smallest GCV and UBR at 1, 2, 3 knot point and oscillation parameter 1, 2, 3.

TABLE 2: GCV and UBR with 1 Knot Point and Oscillation.

Knot Point	Oscillation	GCV	UBR
1	1	17.034	194.730
1	2	15.802	194.647
1	3	16.305	194.701
2	1	16.137	7.410 x e^{-14}
2	2	15.708	7.005 x e^{-14}
2	**3**	15.807	**5.607 x e^{-14}**
3	1	7.488	1.136 x e^{-12}
3	2	5.646	1.036 x e^{-12}
3	**3**	**5.068**	1.229 x e^{-12}

The minimum GCV value is 5.068 with three knot points and three oscillation parameters and The minimum UBR value is 5.607 x e^{-14} with two knot points and three oscillation parameters.

3.4. Optimal GCV and UBR Method Parameter Estimation

The results of the parameter estimation in nonparametric regression of combined estimators spline truncated, and Fourier series using GCV and UBR method are shown in Table 3.

TABLE 3: Optimal GCV and UBR Model Parameter Estimation.

Variable	Parameter	Estimation GCV	Estimation UBR
constan	c	-4.781	0.002
h_1 = Labour Force Participation Rate	θ_1	0.269	0.005
	θ_{11}	-1.870	-0.136
	θ_{12}	-28.232	-6.155x10^{-16}
	θ_{13}	35.681	
h_2 = Mean Years Schooling	θ_2	0.214	0.232
	θ_{21}	-8.516	0.217
	θ_{22}	6.727	4.567x10^{-16}
	θ_{23}	9.229	
h_3 = An Improved Drinking Water	θ_3	-0.142	0.068
	θ_{31}	7.221	-0.077
	θ_{32}	-17.918	0
	θ_{33}	10.644	
z_1 = Domestic Direct Investment	b_1	-0.045	-0.022
	a_{11}	-0.385	-1.580
	a_{12}	1.980	1.578
	a_{13}	-0.988	-1.270

Table 4 shows the optimal knot point values and oscillation parameters for the GCV and UBR methods.

TABLE 4: Optimal knots and oscillation points.

Knot Point			Oscillation	GCV	UBR
h_1	h_2	h_3	S		
63.08	7.02	65.39	3		5.607 x e^{-14}
77.75	11.31	98.42			
72.062	9.647	85.612	3	5.068	
73.259	9.997	88.309			
73.858	10.172	89.657			

Nonparametric regression of Combined Spline Truncated and Fourier Series model using optimal values in the GCV method.

$$y_i = -4,781 + 0.269h_{1i} - 1,87\left(h_{1i} - 72,062\right)_+ - 28,232\left(h_{1i} - 73,259\right)_+ + 35,681\left(h_{1i} - 73,858\right)_+ + 0,214h_{2i} - 8,516\left(h_{2i} - 9,647\right)_+ + 6,727\left(h_{2i} - 9,997\right)_+ + 9,229\left(h_{2i} - 10,172\right)_+ - 0,142h_{3i} + 7,221\left(h_{3i} - 85,612\right)_+ - 17,919\left(h_{3i} - 88,309\right)_+ + 10,644\left(h_{3i} - 89,657\right)_+ - 0,045z_{1i} - 0,385\cos z_{1i} + 1,98\cos 2z_{1i} - 0,988\cos 3z_{1i} + \epsilon_i$$

Nonparametric regression of Combined Spline Truncated and Fourier Series model using optimal values in the UBR method.

$$\hat{y}_i = 0,002 + 0,005h_{1i} - 0,136\left(h_{1i} - 63.08\right)_+ - 6,155 \times 10^{-16}\left(h_{1i} - 77.75\right)_+ + 0,232h_{2i} + 0,217\left(h_{2i} - 7.02\right)_+ + 4,567 \times 10^{-16}\left(h_{2i} - 11.31\right)_+ + 0,068h_{3i} - 0,077\left(h_{3i} - 65.39\right)_+ + 0\left(h_{3i} - 98.42\right) - 0,022z_{1i} - 1,580\cos z_{1i} + 1,578\cos 2z_{1i} - 1,270\cos 3z_{1i}$$

3.5. Comparison of GCV and UBR Methods

Selecting of optimal knot points and oscillation parameters in done by comparing the GCV and UBR methods. In the results of the analysis of the previous subchapter, the following results were obtained.

Table 5: Comparison Table.

	GCV	UBR
Mean Square Error	1,420	10,614
Coefficient of Determination (R^2)	89,34%	20,27%
Knot Points	3	2
Oscillation Parameter	3	3

It is known that the GCV method has a smaller MSE than the UBR method. That is, the best number of optimal knot points and oscillation parameters use the GCV method, namely as many as three knot points and three oscillation parameters.

3.6. Best Model Scenario

Parameter estimation will form the nonparametric regression model equation of combined estimators spline truncated and Fourier series as follow

$$
\begin{aligned}
y_i = & -4,781 + 0.269h_{1i} - 1,87\left(h_{1i} - 72,062\right)_+ - 28,232\left(h_{1i} - 73,259\right)_+ + \\
& 35,681\left(h_{1i} - 73,858\right)_+ + 0,214h_{2i} - 8,516\left(h_{2i} - 9,647\right)_+ + \\
& 6,727\left(h_{2i} - 9,997\right)_+ + 9,229\left(h_{2i} - 10,172\right)_+ - 0,142h_{3i} + \\
& 7,221\left(h_{3i} - 85,612\right)_+ - 17,919\left(h_{3i} - 88,309\right)_+ + 10,644\left(h_{3i} - 89,657\right)_+ - \\
& 0,045z_{1i} - 0,385\cos z_{1i} + 1,98\cos 2z_{1i} - 0,988\cos 3z_{1i} + \epsilon_i
\end{aligned}
$$

A nonparametric regression model scenario using a mixture of truncated spline and Fourier series estimators will be carried out. The model scenario is divided into three parts, namely by looking at the growth rate of GRDP with the provinces that have rate figures GRDP growth below 3%, provinces that have a growth rate of GRDP between 3% and 4%, and provinces that have a growth rate of GRDP above 5 percent.

Provinces that have growth rate of GRDP below 3%. There are two provinces in Indonesia that have growth rate of GRDP below 3%, namely West Sulawesi and Papua. In order for the growth rate of GRDP to increase to around 5%, the labor force participation rate must be around 77.34%, mean years schooling must be around 9.64 years, the percentage of households that have an improved drinking water drinking water must be worth about 84.65%, and domestic direct investment should be worth about 3.67 trillion.

Provinces that have a growth rate of GRDP between 3% and 4%. There are eleven provinces in Indonesia that have a growth rate of GRDP between 3% and 4%. In order for the growth rate of GRDP to increase to around 5%, the labor force participation rate must be around 76.71%, mean years schooling should be worth around 10.18 years, the percentage of households that have an improved drinking water drinking water should be worth around 85.23%, and domestic direct investment should be worth around 28.56 trillion.

Provinces that have a growth rate of GRDP above 5%. There are 21 provinces in Indonesia that have a growth rate of GRDP above 5%. In order for the growth rate of GRDP to increase to around 7%, the labor force participation rate must be around 78%, mean years schooling must be around 9.75 years, the percentage of households that have an improved drinking water should be worth around 96.5%, and domestic direct investment should be worth around 30.28 trillion.

4. CONCLUSION

The results of the comparison of the GCV and UBR methods on the selection of optimal knot points and oscillation parameters applied to economic growth Indonesia in 2022 for the GCV method obtained an MSE value of 1.42 with a coefficient of determination (R^2) value of 89.34% with many knot points as three and three oscillation parameters. For the UBR method, MSE value is 10.614 with a coefficient of determination (R^2) of 20.27% with two knot points and three oscillation parameters. That means GCV method is the best method for selecting knot points and oscillation parameters in nonparametric regression of combined spline truncated and Fourier series with used three knot points and three oscillation parameters.

References

[1] Budiantara IN. Splines in Nonparametric and Semiparametric Regression: A Statistical Modeling of the Present and Future. Inauguration Speech for the Position of Professor in the Department of Statistics, Institut Teknologi Sepuluh Nopember Surabaya; 2009.

[2] Budiantara N, Ratna M, Zain I, Wibowo W. Modeling the Percentage of Poor People in Indonesia Using Spline Nonparametric Regression Approach. International Journal Basis & Applied Sciences; 2012. pp. 118–24.

[3] Mariati NPAM. Fourier Series and Truncated Spline Regression Modeling in Multivariable Nonparametric Regression (Application: Poverty Data in Papua Province). Surabaya: Institut Teknologi Sepuluh November; 2015.

[4] Afifah N. Mixed Kernel and Fourier Series Estimators in Nonparametric Regression (Case Study: Modeling the Percentage of Poor Population in Indonesia). Surabaya: Institut Teknologi Sepuluh November; 2017.

[5] Gabrela PP. Birespon Nonparametric Regression Modeling Using Mixed Truncated Spline, Kernel, and Fourier Series Estimators (Case Study: Modeling the Percentage of Poor Population and Poverty Depth Index in Papua Province). Surabaya: Institut Teknologi Sepuluh November; 2021.

[6] Sari SUR. Comparison of Multivariable Spline Nonparametric Regression Models Using Generalized Cross Validation (GCV) and Unbiased Risk (UBR) Methods in Selection of Optimal Knot Points (Case Study of Maternal Mortality Rate Data in East Java). Surabaya: Institut Teknologi Sepuluh November; 2016.

[7] Devi AR, Budiantara IN, Ratnasari V. Unbiased risk and cross-validation method for selecting optimal knots in multivariable nonparametric regression spline truncated (case study: unemployment rate in Central Java, Indonesia, 2015). AIP Conf Proc. 2018;:1–9.

[8] Dani ATR. Selection of Optimal Knot Points and Optimal Bandwidth Using Cross-Validation (CV), Generalized Cross-Validation (GCV), and Unbiassed Risk (UBR) in Nonparametric Regression Modeling Mixed Truncated Spline Estimator and Gaussian Kernel. Surabaya: Institut Teknologi Sepuluh November; 2021.

[9] Amrullah MN, Amalia SH. Comparison of Generalized Cross Validation (GCV) Methods with Cross Validation (CV) to Determine Optimal Knots in Fourier Series Nonparametric Regression (Case Study: poverty Rate in North Sumatra Province). Jurnal Litbang Edusaintech. 2022;3(1):1–10.

[10] Abdullah, Shah T, Sargani GR, Ali A, Siraj W. The Effect of Increase in Population on the Economic Growth of. Developing Country Studies. 2015;5(17):87–94.

[11] Eubank RL. Nonparametric Regression and Spline Smoothing. New York: Marcel Dekker; 1999. https://doi.org/10.1201/9781482273144.

[12] Wu H, Zhang JT. Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches. New York: John Wiley & Sons;