BAREKENG: Journal of Mathematics and Its Applications
March 2025 Volume 19 Issue 1 Page 0581–0594
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss1pp0581-0594

CONSUMER PRICE INDEX MODELING USING A MIXED TRUNCATED SPLINE AND KERNEL SEMIPARAMETRIC REGRESSION APPROACH

Lilik Hidayati [1*], Mustika Hadijati[2], Nur Asmita Purnamasari[3], Ristiandi[4], Ni Nyoman Dewi Kartini[5]

[1,2,3,5]*Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Mataram*
[4]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Mataram*
Jln. Majapahit No.62, Gomong, Selaparang, Mataram, 83115, Indonesia

*Corresponding author's e-mail: *lilik_hidayati@staff.unram.ac.id*

Article History:

Received: 6th August 2024
Revised: 5th November 2024
Accepted: 27th November 2024
Published:13th January 2025

Keywords:

Consumer Price Index;
Kernel;
Regression Semiparametric;
Spline Truncated.

ABSTRACT

Some semiparametric regression model approaches include spline, kernel, Fourier series, and wavelet. Semiparametric regression modelling can involve more than one independent variable (multivariable), a parametric approach is usually combined with one of the nonparametric approaches, such as combining a parametric approach with a nonparametric kernel. If a consumer price index model can be built based on the variables that influence it, predictions of consumer price percentages can be made, which it is hoped will help the government determine policies to control consumer price inflation, especially in West Nusa Tenggara Province. The data used in this research includes the consumer price index and the factors that influence it according to districts/cities in West Nusa Tenggara Province from 2022 to April 2024. The data source was obtained from secondary data at BPS West Nusa Tenggara Province. This research design uses a mixed semiparametric approach of truncated spline and kernel regression. Based on calculations, the predicted results of the consumer price index in West Nusa Tenggara Province show that the predicted data graph $(\hat{y})$ is very close to the actual data (y). Modelling the consumer price index in West Nusa Tenggara Province is a model with 2 knot points, where the model efficiency has the smallest GCV value of 0.001507. The model goodness value (R^2) is 0.99, meaning that the variables used can explain 99% of the model variability.

How to cite this article:

L. Hidayati, M. Hadijati, N. A. Purnamasari, Ristiandi and N. N. D. Kartini., "CONSUMER PRICE INDEX MODELING USING A MIXED TRUNCATED SPLINE AND KERNEL SEMIPARAMETRIC REGRESSION APPROACH," *BAREKENG: J. Math. & App.*, vol. 19, iss. 1, pp. 0581-0594, March, 2025.

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

Regression analysis has a very important role in statistical theory and its applications. A common problem that occurs is determining the functional form of correlation between two or more variables. Functional relationships between variables are also determined based on data patterns; if the data follows a certain pattern, then use a parametric regression approach. Meanwhile, if the data does not follow a certain pattern, then use a nonparametric regression model. The implementation of a combined regression approach is also said to be a semiparametric regression approach, where the correlation between each variable can be carried out using a parametric approach, while a nonparametric approach will model some of the other variables [1]. In semiparametric regression models, there are several approach models, each of which has its motivation in obtaining a form of regression estimator. Some semiparametric regression approach models include spline, kernel, Fourier series, and wavelet.

Semiparametric regression modeling can involve more than one response variable (multivariable), a parametric approach is usually combined with one of the nonparametric approaches only, such as combining a parametric approach with a nonparametric kernel [2]. The assumption used is that some of the variable data has a distribution pattern with a certain shape (for example a straight line), and some of the distribution pattern has a similar shape in a nonparametric approach, such as the pattern used in the kernel approach [3]. However, in reality the data distribution pattern is not always uniform for the nonparametric approach of each variable. Therefore, semiparametric regression modeling was developed, which can combine parametric approaches and more than one nonparametric approach in the hope of providing better models and predictions [4]. Several studies have been produced relating to mixed nonparametric and semiparametric regression models, including research on mixed truncated spline and kernel nonparametric regression models [5], [6], [7], [8] mixed kernel and Fourier series nonparametric regression models [9], [10] mixed kernel and Fourier series semiparametric regression models [11], [12], [13] mixed truncated spline and Fourier series semiparametric regression models [14].

The country of Indonesia, including West Nusa Tenggara, is starting to face very serious problems regarding the consumer price index (CPI). Controlling inflation by maintaining CPI inflation within the target range of 3.0% - 1.0% in 2023 is one of the agendas launched by the government. The consumer price index is influenced by several factors, including: 1) Price policy set by the government; 2) Level of consumer demand for goods and services; 3) Increase in community income; 4) Production costs incurred by producers; and 5) Currency exchange rates. If a consumer price index model can be built based on the variables that influence it, predictions of consumer price percentages can be made, which it is hoped will help the government determine policies to control consumer price inflation, especially in West Nusa Tenggara Province. Modeling the consumer price index can be done using a nonparametric kernel or Fourier series regression approach, as has been shown by previous studies [15], [16]. The difference between this research and previous research is the case study. So the appropriate regression model in this case is a mixed regression model of truncated splines and kernels in semiparametric regression, which is very appropriate for solving the problems in this research.

The mixed truncated spline and kernel estimator model in semiparametric regression plays a very important role in solving regression modeling problems that have a relationship between the response variable and the predictor variables. The relationships referred to are some that follow certain patterns, others that have patterns that change at certain intervals, and others that do not follow certain patterns. The parametric regression, nonparametric regression, and semi-parametric regression models that have been developed by previous studies have not been able to handle regression modeling cases such as those mentioned above. Therefore, based on the explanation that has been described, we will apply semiparametric mixed truncated spline and kernel regression modeling in modeling the consumer price index in West Nusa Tenggara Province. The aim is to obtain the best consumer price index model in West Nusa Tenggara in implementing the new policy.

2. RESEARCH METHODS

2.1 Truncated Spline Estimator

The nonparametric regression estimation technique first introduced by Whittaker in 1923 was the truncated spline [17]. This technique is very effective in organizing data with changes in behavior at certain intervals. A truncated spline is also called a polynomial model which is divided into several segments [18]. Estimators of the parameters of this model can be obtained using a well-known statistical method, namely Maximum Likelihood Estimation (MLE). If the MLE method is used, the parameter estimator is obtained as follows:

$$\hat{\tilde{\boldsymbol{\theta}}} = \left[\underline{\boldsymbol{G}}(\boldsymbol{k})^{\boldsymbol{T}}\underline{\boldsymbol{G}}(\boldsymbol{k})\right]^{-\mathbf{1}}\underline{\boldsymbol{G}}(\boldsymbol{k})^{\boldsymbol{T}}\tilde{\boldsymbol{y}} \tag{1}$$

with,

$$\hat{\boldsymbol{g}}(\boldsymbol{t_i}) = \underline{\boldsymbol{G}}(\boldsymbol{k})\tilde{\boldsymbol{\theta}}.$$

2.2 Kernel Estimator

The Rosenblatt-Parzen kernel density is often referred to as a kernel estimator, which was first introduced by Rosenblatt in 1956 and Parzen in 1962 [17]. The kernel estimator is a linear estimator that is similar to general estimators and is more specific in use [19]. Thus, the form can be written in the following matrix notation:

$$\hat{\tilde{\boldsymbol{h}}}_\alpha(z) = \underline{D}(\alpha)\tilde{\boldsymbol{y}} \tag{2}$$

2.3 Generalized Cross Validation (GCV)

The optimal spline and the smoothness of the curve depend on the selection of knot points. Different intervals become places where the function changes in knot point k. Meanwhile, the bandwidth α in the kernel method functions as a smoothing parameter for smooth curve estimation [20]. Determining the optimum value can use the GCV method. The GCV value from combining the truncated spline and kernel models can be obtained in the following way:

$$GCV(k,\alpha) = \frac{MSE\ (k,\alpha)}{\left(n^{-1}tr\left(\underline{I}-\underline{M}(k,\alpha)\right)\right)^2} \tag{3}$$

with,

$$MSE(k,\alpha) = n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

2.4 Coefficient of Determination R^2

Obtaining the best model for certain criteria is the goal of regression analysis so that it can explain the correlation of each variable. Selecting the best model using the $\boldsymbol{R^2}$ value is one of the frequently used criteria. The higher the $\boldsymbol{R^2}$ value, the better the model obtained. According to [19], the coefficient of determination $\boldsymbol{R^2}$ can be defined as follows:

$$\boldsymbol{R^2} = \frac{(\hat{\boldsymbol{y}}-\bar{\boldsymbol{y}})^{\boldsymbol{T}}(\hat{\boldsymbol{y}}-\bar{\boldsymbol{y}})}{(\boldsymbol{y}-\bar{\boldsymbol{y}})^{\boldsymbol{T}}(\boldsymbol{y}-\bar{\boldsymbol{y}})} \tag{4}$$

2.5 Research Data

The data used in this research includes the consumer price index and the factors that influence it according to districts/cities in West Nusa Tenggara Province from 2022 to April 2024. The data source was obtained from secondary data at BPS West Nusa Tenggara Province. The research data is divided into dependent variables and several independent variables. Following are some of these variables listed in Table 1.

Table 1. Research Variable

Symbol	Variable	Unit
y	Consumer Price Index	%
x_1	Education	%
x_2	Food, Drinks, Tobacco	%
x_3	Clothing and Footwear	%
x_4	Housing, Water, Electricity, and Household Fuel	%
x_5	Supplies, Equipment, and Routine Household Maintenance	%
x_6	Health	%
x_7	Information, Communication, and Financial Services	%

Symbol	Variable	Unit
x_8	Recreation, Sports, and Culture	%
x_9	Provision of Food and Drinks/Restaurant	%
x_{10}	Personal Care and Other Services	%
x_{11}	Transportation	%

2.6 Research Stages

This research applies a mixed truncated spline and kernel semiparametric regression approach. The analysis steps are as follows:

a. Make a plot of each variable that influences the consumer price index to determine variables whose functional form patterns are known, as well as identify variable data patterns based on a semiparametric component approach.

b. Carrying out semiparametric regression modeling by combining parametric and nonparametric approaches.

c. Determine parameter estimates for parametric approaches and nonparametric components.

d. Determine the optimum bandwidth, optimum smoothing parameters, and optimum oscillation parameters using the GCV criteria.

e. Establish a semiparametric regression model based on the optimum bandwidth, optimum smoothing parameters, and optimum oscillation parameters obtained.

f. Testing the accuracy and goodness of the model obtained, including normality, independence, and error heteroscedasticity tests.

g. Predict the consumer price index based on the obtained model.

3. RESULTS AND DISCUSSION

3.1 Data Exploration

A plot illustration between each variable that influences the consumer price index shown in **Figure 1**.

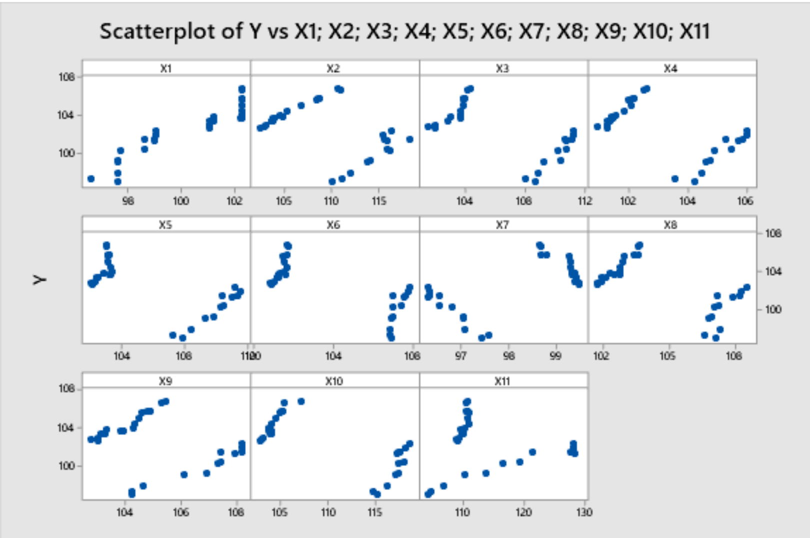

Figure 1. Scatterplot of Dependent and Independent Variables

Based **Figure 1** describes that the variable x_1 follows a linear data pattern. However, the shape of the data pattern other than variable x_1 is not known with certainty, so theoretically, these ten variables can be approached using nonparametric methods. The data pattern of variable y with variables x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10} does not have a particular pattern so it uses a nonparametric spline. Meanwhile, if the data pattern of variable y and variable x_11 changes at certain intervals, using a nonparametric kernel. The parametric component will be symbolized by x, the nonparametric spline by t, and the nonparametric kernel by z. The following complete information regarding each independent variable is written in Table 2.

Table 2. Parametric and Nonparametric Components

No.	Independent Variable	Approach Method	Variable Symbol
1.	Education	Linear parametric	x
2.	Food, Drinks, Tobacco	Truncated Spline	t_1
3.	Clothing and Footwear	Truncated Spline	t_2
4.	Housing, Water, Electricity and Household Fuel	Truncated Spline	t_3
5.	Supplies, Equipment, and Routine Household Maintenance	Truncated Spline	t_4
6.	Health	Truncated Spline	t_5
7.	Information, Communication and Financial Services	Truncated Spline	t_6
8.	Recreation, Sports and Culture	Truncated Spline	t_7
9.	Provision of Food and Drinks/Restaurant	Truncated Spline	t_8
10.	Personal Care and Other Services	Truncated Spline	t_9
11.	Transportation	Kernel	z

3.2 Mixed Semiparametric Regression Models

Mixed truncated spline and kernel semiparametric regression modeling for consumer price index data can be expressed as follows:

$$\begin{aligned} y_i &= \mu(x_i, t_{1i}, t_{3i}, t_{4i}, t_{5i}, t_{6i}, t_{7i}, t_{8i}, t_{9i}, z_i) + \varepsilon_i \\ &= f(x_i) + g_1(t_{1i}) + g_2(t_{2i}) + g_3(t_{3i}) + g_4(t_{4i}) + g_5(t_{5i}) + g_6(t_{6i}) + g_7(t_{7i}) + g_8(t_{8i}) \\ &\quad + g_9(t_{9i}) + h(z_i) + \varepsilon_i. \end{aligned}$$

Next, it is approached with the following function:

$$\begin{aligned} y_i = {} & \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \lambda_{11}(t_{1i} - K_{11}) + \lambda_{12}(t_{1i} - K_{12}) + \cdots + \lambda_{1k}(t_{1i} - K_{ik}) + \theta_{21} t_{2i} \\ & + \lambda_{21}(t_{2i} - K_{21}) + \lambda_{22}(t_{2i} - K_{22}) + \cdots + \lambda_{2k}(t_{2i} - K_{2k}) + \theta_{31} t_{3i} + \lambda_{31}(t_{3i} - K_{31}) \\ & + \lambda_{32}(t_{3i} - K_{32}) + \cdots + \lambda_{3k}(t_{3i} - K_{3k}) + \theta_{41} t_{4i} + \lambda_{41}(t_{4i} - K_{41}) + \lambda_{42}(t_{4i} - K_{42}) \\ & + \cdots + \lambda_{4k}(t_{4i} - K_{4k}) + \theta_{51} t_{5i} + \lambda_{51}(t_{5i} - K_{51}) + \lambda_{52}(t_{5i} - K_{52}) + \cdots + \lambda_{5k}(t_{5i} - K_{5k}) \\ & + \theta_{61} t_{6i} + \lambda_{61}(t_{6i} - K_{61}) + \lambda_{62}(t_{6i} - K_{62}) + \cdots + \lambda_{6k}(t_{6i} - K_{6k}) + \theta_{71} t_{7i} + \lambda_{71}(t_{7i} - K_{71}) \\ & + \lambda_{72}(t_{7i} - K_{72}) + \cdots + \lambda_{7k}(t_{7i} - K_{7k}) + \theta_{81} t_{8i} + \lambda_{81}(t_{8i} - K_{81}) + \lambda_{82}(t_{8i} - K_{82}) \\ & + \cdots + \lambda_{8k}(t_{8i} - K_{8k}) + \theta_{91} t_{9i} + \lambda_{91}(t_{9i} - K_{91}) + \lambda_{92}(t_{9i} - K_{92}) + \cdots + \lambda_{9k}(t_{8i} - K_{8k}) \\ & + \sum_{i=1}^{n} \frac{\frac{1}{\alpha} K\left(\frac{z - z_i}{\alpha}\right)}{\sum_{i=1}^{n} \frac{1}{\alpha} K\left(\frac{z - z_i}{\alpha}\right)} yi + \varepsilon_i. \end{aligned}$$

The best model results were obtained through a comparison of GCV values. The GCV criteria determine the optimum bandwidth, optimum smoothing parameters, and optimum oscillation parameters. This research will carry out various experiments on the number of knot points, namely 1, 2, up to 3 knot points.

3.2.1 Model with 1 Knot Truncated Spline Component

Modeling with a truncated spline component with 1 knot point can be shown in the following equation:

$$\begin{aligned} y_i = {} & \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \lambda_{11}(t_{1i} - K_{11}) + \theta_{21} t_{2i} + \lambda_{21}(t_{2i} - K_{21}) + \theta_{31} t_{3i} + \lambda_{31}(t_{3i} - K_{31}) \\ & + \theta_{41} t_{4i} + \lambda_{41}(t_{4i} - K_{41}) + \theta_{51} t_{5i} + \lambda_{51}(t_{5i} - K_{51}) + \theta_{61} t_{6i} + \lambda_{61}(t_{6i} - K_{61}) + \theta_{71} t_{7i} \\ & + \lambda_{71}(t_{7i} - K_{71}) + \theta_{81} t_{8i} + \lambda_{81}(t_{8i} - K_{81}) + \theta_{91} t_{9i} + \lambda_{91}(t_{9i} - K_{91}) \end{aligned}$$

$$+\sum_{i=1}^{n}\frac{\frac{1}{\alpha}K\left(\frac{z-z_i}{\alpha}\right)}{\sum_{i=1}^{n}\frac{1}{\alpha}K\left(\frac{z-z_i}{\alpha}\right)}Y_i+\varepsilon_i.$$

The GCV values obtained are listed in **Table 3**.

Table 3. GCV Value on Models with 1 Knot Point

Linear Parametric	Spline									Kernel	GCV
β_1	Knot Point (K)									Bandwidth	
x	K_1	K_2	K_3	K_4	K_5	K_6	K_7	K_8	K_9	α	
-1.501	106.96	104.30	102.37	104.80	102.79	97.22	103.67	104.32	107.37	0.040	0.011039
-1.349	106.20	103.84	10212	104.34	102.46	97.07	103.34	104.07	106.62	0.042	0.019503
-0.978	105.44	103.37	101.89	103.89	102.12	96.92	103.02	103.81	105.87	0.043	0.021315
-0.705	104.68	102.91	101.64	103.44	101.78	96.77	102.69	103.55	105.11	0.050	0.038038
-1.586	111.51	107.10	103.82	107.51	104.83	98.12	105.62	105.87	111.88	0.044	0.054243

Table 3 shows that the smallest GCV value is 0.011039 and provides estimates of the parameters in the spline components listed in **Table 4** below:

Table 4. Parameter Estimation with 1 Knot Point

Variable	Parameter	Estimation
t_1	θ_{11}	-0.0164
	λ_{11}	1.0486
t_2	θ_{21}	-1.9252
	λ_{21}	-0.3032
t_3	θ_{31}	-2.5397
	λ_{31}	0.7894
t_4	θ_{41}	1.6667
	λ_{41}	1.4069
t_5	θ_{51}	-0.4613
	λ_{51}	0.1778
t_6	θ_{61}	-0.5580
	λ_{61}	5.7414
t_7	θ_{71}	-1.1938
	λ_{71}	2.1619
t_8	θ_{81}	-0.1695
	λ_{81}	-1.6982
t_9	θ_{91}	-1.0670
	λ_{91}	-0.1367

Based on the results of **Table 3** and **Table 4**, a model with a truncated spline component of 1 knot point can be written, namely:

$$\begin{aligned}\hat{y}_i = {} & 183.3636-1.50131x_i-0.0164t_{1i}+1.0486(t_{1i}-106.96)-1.9252t_{2i}-0.3032(t_{2i}-104.30)\\ & -2.5397t_{3i}+0.7894(t_{3i}-102.37)+1.6667t_{4i}+1.4069(t_{4i}-104.80)-0.4613t_{5i}\\ & +0.1778(t_{5i}-102.79)-0.5580t_{6i}+5.7414(t_{6i}-97.22)-1.1938t_{7i}\\ & +2.1619(t_{7i}-103.67)-0.1695t_{8i}-1.6982(t_{8i}-104.32)-1.0670t_{9i}\\ & -0.1367(t_{9i}-107.37)+\sum_{i=1}^{28}\frac{\frac{1}{0.040}K\left(\frac{z-z_i}{0.040}\right)}{\sum_{i=1}^{n}\frac{1}{0.040}K\left(\frac{z-z_i}{0.040}\right)}Y_i\end{aligned}$$

3.2.2 Model with 2 Knot Truncated Spline Components

Modeling with a truncated spline component with 2 knot points can be shown in the following equation:

$$
\begin{aligned}
y_i = {} & \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \lambda_{11}(t_{1i} - K_{11}) + \lambda_{12}(t_{1i} - K_{12}) + \theta_{21} t_{2i} + \lambda_{21}(t_{2i} - K_{21}) + \lambda_{22}(t_{2i} - K_{22}) \\
& + \theta_{31} t_{3i} + \lambda_{31}(t_{3i} - K_{31}) + \lambda_{32}(t_{3i} - K_{32}) + \theta_{41} t_{4i} + \lambda_{41}(t_{4i} - K_{41}) + \lambda_{42}(t_{4i} - K_{42}) + \theta_{51} t_{5i} \\
& + \lambda_{51}(t_{5i} - K_{51}) + \lambda_{52}(t_{5i} - K_{52}) + \theta_{61} t_{6i} + \lambda_{61}(t_{6i} - K_{61}) + \lambda_{62}(t_{6i} - K_{62}) + \theta_{71} t_{7i} \\
& + \lambda_{71}(t_{7i} - K_{71}) + \lambda_{72}(t_{7i} - K_{72}) + \theta_{81} t_{8i} + \lambda_{81}(t_{8i} - K_{81}) + \lambda_{82}(t_{8i} - K_{82}) + \theta_{91} t_{9i} \\
& + \lambda_{91}(t_{9i} - K_{91}) + \lambda_{92}(t_{9i} - K_{92}) + \sum_{i=1}^{n} \frac{\frac{1}{\alpha} K\left(\frac{z - z_i}{\alpha}\right)}{\sum_{i=1}^{n} \frac{1}{\alpha} K\left(\frac{z - z_i}{\alpha}\right)} Y_i + \varepsilon_i.
\end{aligned}
$$

The resulting GCV values are listed in **Table 5**.

Table 5. GCV Value on Model with 2 Knot Points

Linear Parametric	**Spline**									**Kernel**	**GCV**
$\boldsymbol{\beta_1}$	**Knot Point (K)**									**Bandwidth**	
$\boldsymbol{x}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$	$\boldsymbol{\alpha}$	
0.2291	103.63	102.26	101.30	102.81	101.31	96.56	102.24	103.20	104.07	0.028	0.001507
	106.08	103.77	102.09	104.27	102.40	97.05	103.29	10403	106.50		
1.133	103.63	102.26	101.30	102.81	101.31	96.56	10224	103.20	104.07	0.044	0.002642
	107.31	104.52	102.48	105.01	102.95	97.29	103.82	104.44	107.72		
-0.114	110.99	106.78	103.66	107.20	104.59	98.02	105.39	105.69	111.36	1.089	0.003242
	112.22	107.53	104.05	107.93	105.14	98.26	105.92	106.10	112.57		
-2.134	106.08	103.77	102.09	104.27	102.40	97.05	103.29	104.03	107.72	0.035	0.004192
	113.44	108.28	104.44	108.66	105.69	98.50	106.44	106.52	117.43		
-2.879	107.31	104.52	102.48	105.01	102.95	97.29	103.82	104.44	107.72	0.051	0.005112
	117.12	110.54	105.62	110.86	107.33	99.23	108.02	107.76	117.43		

Table 5 shows that the smallest GCV value is 0.001507 and provides estimates of the parameters in the spline components listed in **Table 6** below:

Table 6. Parameter Estimation with 2 Knot Points

Variable	Parameter	Estimation
t_1	θ_{11}	1.5054
	λ_{11}	1.3069
	λ_{12}	0.1295
t_2	θ_{21}	-1.0255
	λ_{21}	-0.0440
	λ_{22}	1.7958
t_3	θ_{31}	-0.3889
	λ_{31}	-0.4072
	λ_{32}	-2.8178
t_4	θ_{41}	-1.3708
	λ_{41}	-0.0073
	λ_{42}	0.8788
t_5	θ_{51}	-2.1571
	λ_{51}	0.3333
	λ_{52}	-0.5816
t_6	θ_{61}	-0.2665
	λ_{61}	0.8782
	λ_{62}	-0.1733
t_7	θ_{71}	1.2418

Variable	Parameter	Estimation
	λ_{71}	-0.4986
	λ_{72}	-1.9169
t_8	θ_{81}	-1.3297
	λ_{81}	1.9603
	λ_{82}	-0.7795
t_9	θ_{91}	1.4145
	λ_{91}	2.8520
	λ_{92}	-0.2720

Based on the results of **Table 5** and **Table 6**, a model with a truncated spline component of 2 knot points can be written, namely:

$$\begin{aligned}\hat{y}_i &= -15.6575 + 0.2291x_i + 1.5054t_{1i} + 1.3069(t_{1i} - 103.63) + 0.1295(t_{1i} - 106.08) - 1.0255t_{2i}\\ &- 0.0440(t_{2i} - 102.26) + 1.7958\,(t_{2i} - 103.77) - 0.3889t_{3i} - 0.4072(t_{3i} - 101.30)\\ &- 2.8178(t_{3i} - 102.09) - 1.3708t_{4i} - 0.0073(t_{4i} - 102.81) + 0.8788\,(t_{4i} - 104.27)\\ &- 2.1571t_{5i} + 0.3333(t_{5i} - 101.31) - 0.5816(t_{5i} - 102.40) - 0.2665t_{6i}\\ &+ 0.8782(t_{6i} - 96.56) - 0.1733\,(t_{6i} - 97.05) + 1.2418t_{7i} - 0.4986(t_{7i} - 102.24)\\ &- 1.9169(t_{7i} - 103.29) - 1.3297t_{8i} + 1.9603(t_{8i} - 103.20) - 0.7795\,(t_{8i} - 104.03)\\ &+ 1.4145t_{9i} + 2.8520(t_{9i} - 104.07) - 0.2720(t_{9i} - 106.50) + \sum_{i=1}^{28} \frac{\frac{1}{0.028}K\left(\frac{z - z_i}{0.028}\right)}{\sum_{i=1}^{n}\frac{1}{0.028}K\left(\frac{z - z_i}{0.028}\right)}Y_i\end{aligned}$$

3.2.3 Model with 3 Knot Truncated Spline Components

Modeling with a truncated spline component with 3 knot points can be shown in the following equation:

$$\begin{aligned}y_i &= \beta_0 + \beta_1 x_i + \theta_{11}t_{1i} + \lambda_{11}(t_{1i} - K_{11}) + \lambda_{12}(t_{1i} - K_{12}) + \lambda_{13}(t_{1i} - K_{13}) + \theta_{21}t_{2i} + \lambda_{21}(t_{2i} - K_{21})\\ &+ \lambda_{22}(t_{2i} - K_{22}) + \lambda_{23}(t_{2i} - K_{23}) + \theta_{31}t_{3i} + \lambda_{31}(t_{3i} - K_{31}) + \lambda_{32}(t_{3i} - K_{32}) + \lambda_{33}(t_{3i} - K_{33})\\ &+ \theta_{41}t_{4i} + \lambda_{41}(t_{4i} - K_{41}) + \lambda_{42}(t_{4i} - K_{42}) + \lambda_{43}(t_{4i} - K_{43}) + \theta_{51}t_{5i} + \lambda_{51}(t_{5i} - K_{51})\\ &+ \lambda_{52}(t_{5i} - K_{52}) + \lambda_{53}(t_{5i} - K_{53}) + \theta_{61}t_{6i} + \lambda_{61}(t_{6i} - K_{61}) + \lambda_{62}(t_{6i} - K_{62})\\ &+ \lambda_{63}(t_{6i} - K_{63}) + \theta_{71}t_{7i} + \lambda_{71}(t_{7i} - K_{71}) + \lambda_{72}(t_{7i} - K_{72}) + \lambda_{73}(t_{7i} - K_{73}) + \theta_{81}t_{8i}\\ &+ \lambda_{81}(t_{8i} - K_{81}) + \lambda_{82}(t_{8i} - K_{82}) + \lambda_{83}(t_{8i} - K_{83}) + \theta_{91}t_{9i} + \lambda_{91}(t_{9i} - K_{91}) + \lambda_{92}(t_{9i} - K_{92})\\ &+ \lambda_{93}(t_{9i} - K_{93}) + \sum_{i=1}^{n} \frac{\frac{1}{\alpha}K\left(\frac{z - z_i}{\alpha}\right)}{\sum_{i=1}^{n}\frac{1}{\alpha}K\left(\frac{z - z_i}{\alpha}\right)}Y_i + \varepsilon_i.\end{aligned}$$

The resulting GCV values are listed in **Table 7**.

Table 7. **GCV Value on Model with 3 Knot Points**

Linear Parametric	Spline									Kernel	GCV
$\boldsymbol{\beta_1}$	Knot Point ($\boldsymbol{K}$)									Bandwidth	
$\boldsymbol{x}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$ $\boldsymbol{K_3}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$ $\boldsymbol{K_3}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$ $\boldsymbol{K_3}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$ $\boldsymbol{K_3}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$ $\boldsymbol{K_3}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$ $\boldsymbol{K_3}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$ $\boldsymbol{K_3}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$ $\boldsymbol{K_3}$	$\boldsymbol{K_1}$ $\boldsymbol{K_2}$ $\boldsymbol{K_3}$	$\boldsymbol{\alpha}$	

-1.678	106.75	104.18	102.30	104.67	102.70	97.18	103.58	104.25	107.16	0.040	0.004373
	114	108.62	104.62	109	105.94	98.61	106.68	106.71	114.34		
	115.45	109.51	105.08	109.86	106.59	98.90	107.30	107.20	115.77		
-1.821	103.85	102.4	101.37	102.94	101.41	96.61	102.34	103.27	104.29	0.550	0.005277
	109.65	105.96	103.23	106.40	104	97.75	104.82	105.23	110.03		
	116.9	110.4	105.55	110.73	107.23	99.18	107.92	107.69	117.21		
-0.387	109.65	105.96	103.23	106.40	104	97.75	104.82	105.23	110.03	0.025	0.007024
	114	108.62	104.62	109	105.94	98.61	106.68	106.71	114.34		
	116.90	110.40	105.55	110.73	107.23	99.18	107.92	107.69	117.21		
-1.985	106.75	104.18	102.30	104.67	102.70	97.18	103.58	104.25	107.16	0.041	0.009727
	112.55	107.73	104.16	108.13	105.29	98.32	106.06	106.22	112.90		
	114	108.62	104.62	109	105.94	98.61	106.68	106.71	114.34		
-1.386	106.75	104.18	102.30	104.67	102.70	97.18	103.58	104.25	107.16	0.034	0.011891
	111.10	106.84	103.69	107.27	104.64	98.04	105.44	105.73	111.47		
	114	108.62	104.62	109	105.94	98.61	106.68	106.71	114.34		

Table 7 shows that the smallest GCV value is 0.004373 and provides estimates of the parameters in the spline components listed in **Table 8** below:

Table 8. Parameter Estimation with 3 Knot Points

Variable	Parameter	Estimation
t_1	θ_{11}	-0.1508
	λ_{11}	1.4593
	λ_{12}	-2.2736
	λ_{13}	0.2963
t_2	θ_{21}	-2.0102
	λ_{21}	-0.1914
	λ_{22}	1.7161
	λ_{23}	0.6999
t_3	θ_{31}	-0.8183
	λ_{31}	0.1360
	λ_{32}	-0.4543
	λ_{33}	0.7867
t_4	θ_{41}	5.1870
	λ_{41}	-6.5744
	λ_{42}	-0.3240
	λ_{43}	-4.8175
t_5	θ_{51}	5.2118
	λ_{51}	4.0952
	λ_{52}	-2.6196
	λ_{53}	-3.2845
t_6	θ_{61}	5.8172
	λ_{61}	-1.2840
	λ_{62}	0.7062
	λ_{63}	1.1043
t_7	θ_{71}	2.0096
	λ_{71}	2.5063
	λ_{72}	-1.2964
	λ_{73}	5.3417
t_8	θ_{81}	2.0719
	λ_{81}	-7.7212
	λ_{82}	4.8211

Variable	Parameter	Estimation
	λ_{83}	-5.2159
t_9	θ_{91}	-2.4928
	λ_{91}	-1.9460
	λ_{92}	2.4650
	λ_{93}	0.0204

Based on the results of **Table 7** and **Table 8**, a model with a truncated spline component of 3 knot points can be written, namely:

$$\begin{aligned}\hat{y}_i = {} & 289.502 - 1.678x_i - 0.1508t_{1i} + 1.4593(t_{1i} - 106.75) - 2.2736(t_{1i} - 114) \\ & + 0.2963(t_{1i} - 115.45) - 2.0102t_{2i} - 0.1914(t_{2i} - 104.18) + 1.7161(t_{2i} - 108.62) \\ & + 0.69999(t_{2i} - 109.51) - 0.8183t_{3i} + 0.1360(t_{3i} - 102.30) - 0.4543(t_{3i} - 104.62) \\ & + 0.7867(t_{3i} - 105.08) + 5.1870t_{4i} - 6.5744(t_{4i} - 104.67) - 0.3240(t_{4i} - 109) \\ & - 4.8175(t_{4i} - 109.86) + 5.2118t_{5i} + 4.0952(t_{5i} - 102.70) - 2.6196(t_{5i} - 105.94) \\ & - 3.2845(t_{5i} - 106.59) + 5.8172t_{6i} - 1.2840(t_{6i} - 97.18) + 0.7062(t_{6i} - 98.61) \\ & + 1.1043(t_{6i} - 98.90) + 2.0096t_{7i} + 2.5063(t_{7i} - 103.58) - 1.2964(t_{7i} - 106.68) \\ & + 5.3417(t_{7i} - 107.30) + 2.0719t_{8i} - 7.7212(t_{8i} - 104.25) + 4.8211(t_{8i} - 106.71) \\ & - 5.2159(t_{8i} - 107.20) - 2.4928t_{9i} - 1.9460(t_{9i} - 107.16) + 2.4650(t_{9i} - 114.34) \\ & + 0.0204(t_{9i} - 115.77) + \sum_{i=1}^{n} \frac{\frac{1}{0.040} K\left(\frac{z - z_i}{0.040}\right)}{\sum_{i=1}^{n} \frac{1}{0.040} K\left(\frac{z - z_i}{0.040}\right)} Y_i.\end{aligned}$$

Next, compare the GCV values of the 3 knot points obtained as in **Table 9** below:

Table 9. GCV Values for Models with 1, 2, 3 Knot Points

No.	Model	GCV
1.	Model with 1 Knot Point	0.011039
2.	Model with 2 Knot Points	0.001507
3.	Model with 3 Knot Points	0.004373

Based **Table 3**, **Table 5**, and **Table 7** show that the model with a truncated spline component with 2 knot points gives the smallest GCV value, namely 0.001507. Thus, the value used in modeling the consumer price index in West Nusa Tenggara Province is a model with 2 knot points.

3.3 Model Accuracy Testing

The residual assumption is used as a condition for model feasibility. The residual assumptions that must be met are that the residuals are normally distributed, independent, and have the following error heteroscedasticity.

3.3.1 Normality Test

The normality test is carried out using the hypothesis in the Kolmogorov-Smirnov test as follows:

H_0 : Errors are normally distributed ($\mu = 0$)

H_1 : Errors are not normally distributed ($\mu \neq 0$)

The following is the calculation of the Kolmogorov-Smirnov test results which are listed in **Table 10**.

Table 10. Kolmogorov-Smirnov Test

Test Statistical Value	$p-$value	Decision
0.093	0.200	Reject H_0

Table 10 shows a value of $Dtest = 0.093$ with $\alpha = 0,05$. The comparison of the values of $D < D_{table}$ is $0.093 < 0.250$, so the decision fails to reject H_0. Apart from that, the p value is 0.200, which is greater than 0.05, causing the decision to fail to reject H_0. Thus, the conclusion is that the error is normally distributed.

3.3.2 Independence Test

The independence test is carried out using the hypothesis in the Durbin-Watson test as follows:

$H_0 :\ \rho_1 = \rho_2 = \cdots = \rho_n = 0$ (There is no autocorrelation)

$H_1 : \rho_i \neq 0, i = 1, 2, 3, \dots, n$ (There is autocorrelation)

The following is the calculation of the Durbin-Watson test results which are listed in **Table 11**.

Table 11. Durbin-Watson Test

d value	Durbin Watson table values	
	dL	dU
1.981	1.06	1.91

Based **Table 11** explained that the Durbin-Watson test from the principal component regression method obtained a value of $dU < d < 4 - dU$, namely $1.91 < 1.981 < 2.09$ so that the decision failed to reject H_0, which gave the conclusion that there was no autocorrelation and the independent assumption was met.

3.3.3 Heteroscedasticity Test

The heteroscedasticity test is carried out using the hypothesis in the Glejser test as follows:

H_0 : Heteroscedasticity does not occur.

H_1 : Heteroscedasticity occurs.

The following is the calculation of the Glejser test results, which are listed in **Table 12**.

Table 12. Glejser Test

Variable	t_{count}	$p\ value$	Decision
Error	1.146	0.262	Failed to reject H_0

Based **Table 12** the value obtained is $t_{0.05;27} = 2.052$ with $\alpha = 0.05$. Comparison of the calculated $t_{count} < t_{0.05;27}$, namely $0.146 < 2.052$, so the decision fails to reject H_0. Apart from that, the p value is 0.262 which is greater than 0.05, causing the decision to fail to reject H_0. Thus, the conclusion is that heteroscedasticity does not occur or the error variance of the method used is consistent.

3.4 Consumer Price Index Prediction

The model is used to predict consumer price index percentage data. The plot between y and $\hat{y}$ for 28 data is shown in **Figure 2** below.

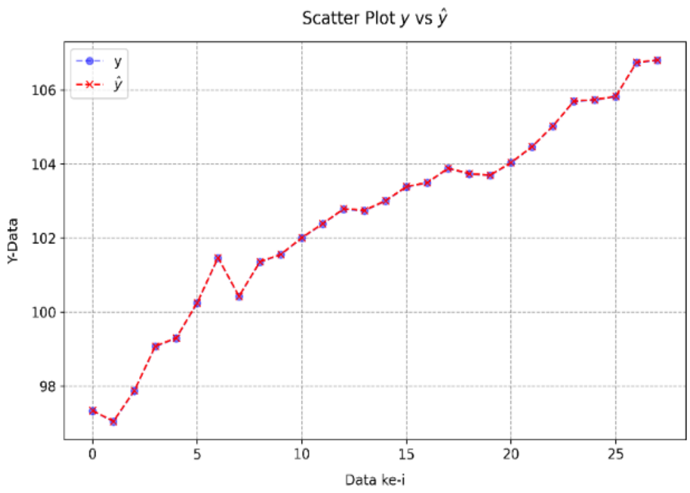

Figure 2. Plot of Consumer Price Index Data

` **Figure 2** shows that the predicted data graph $(\hat{y})$ is very close to the actual data (y). The model goodness value (R^2) in the model with 2 knot points is 0.99 or 99%. Thus, the variable groups are education (X_1), food, drink, and tobacco (X_2), clothing and footwear (X_3), housing, water, electricity, and household fuel (X_4), supplies, equipment, and routine maintenance household (X_5), health (X_6), information, communication, and financial services (X_7), recreation, sports, and culture (X_8), provision of food and drinks/restaurants (X_9), personal care and other services (X_{10}), and transportation (X_{11}) are able to explain 99% of the variability of the consumer price index dependent variable. Variability in a regression model is critical to evaluating the model's goodness-of-fit, prediction consistency, and reliability. The R^2 value of 99% shows that the semiparametric mixed truncated spline and kernel regression model is very good at fitting consumer price index data. If compared, this value is better than previous research regarding nonparametric mixed regression of trauncated linear splines and kernel functions which obtained a model goodness-of-fit value of 92.02%.

4. CONCLUSIONS

Several conclusions were obtained based on data analysis and discussions that have been carried out, namely as follows:

a. Model with 2 knot points on the consumer price index data in West Nusa Tenggara Province:

$$
\begin{aligned}
\hat{y}_i = & -15.6575 + 0.2291x_i + 1.5054t_{1i} + 1.3069(t_{1i} - 103.63) + 0.1295(t_{1i} - 106.08) \\
& - 1.0255t_{2i} - 0.0440(t_{2i} - 102.26) + 1.7958\,(t_{2i} - 103.77) - 0.3889t_{3i} \\
& - 0.4072(t_{3i} - 101.30) - 2.8178(t_{3i} - 102.09) - 1.3708t_{4i} - 0.0073(t_{4i} - 102.81) \\
& + 0.8788\,(t_{4i} - 104.27) - 2.1571t_{5i} + 0.3333(t_{5i} - 101.31) - 0.5816(t_{5i} - 102.40) \\
& - 0.2665t_{6i} + 0.8782(t_{6i} - 96.56) - 0.1733\,(t_{6i} - 97.05) + 1.2418t_{7i} \\
& - 0.4986(t_{7i} - 102.24) - 1.9169(t_{7i} - 103.29) - 1.3297t_{8i} + 1.9603(t_{8i} - 103.20) \\
& - 0.7795\,(t_{8i} - 104.03) + 1.4145t_{9i} + 2.8520(t_{9i} - 104.07) - 0.2720(t_{9i} - 106.50) \\
& + \sum_{i=1}^{28} \frac{\frac{1}{0.028} K\left(\frac{z - z_i}{0.028}\right)}{\sum_{i=1}^{n} \frac{1}{0.028} K\left(\frac{z - z_i}{0.028}\right)} Y_i
\end{aligned}
$$

b. The prediction results of the consumer price index in West Nusa Tenggara Province based on the model obtained show that the predicted data graph $(\hat{y})$ is very close to the actual data (y). In addition, the model has met the feasibility requirements by testing the assumptions of normality, independence, and error heteroscedasticity.

c. Modeling the consumer price index in West Nusa Tenggara Province is a model with 2 knot points, where the model efficiency has the smallest GCV value of 0.001507. The goodness of model value (R^2) is 0.99, meaning that the variables used can explain 99% of the model variability.

ACKNOWLEDGMENT

Thank you to LPPM Mataram University for the 2023 PNBP funding support for capacity-building research with the research title "Consumer Price Index Modeling Using Mixed Truncated Spline and Kernel Semiparametric Regression Approach".

REFERENCES

[1] L. Hidayati, N. Chamidah, and I. N. Budiantara, "Estimasi Selang Kepercayaan Nilai Ujian Nasional Berbasis Kompetensi Berdasarkan Model Regresi Semiparametrik Multirespon Truncated Spline," *MEDIA STATISTIKA*, vol. 13, no. 1, pp. 92–103, Jun. 2020, doi: 10.14710/MEDSTAT.13.1.92-103.

[2] L. Hidayati, N. Chamidah, and I. N. Budiantara, "Bi-Response Semiparametric Regression Model Based on Spline Truncated for Estimating Computer Based National Exam in West Nusa Tenggara," In *Proceeding 1st International Conference on Mathematics and Islam (ICMIs 2018), SCITEPRESS–Science and Technology Publications*, p. 357, 2020.

[3] L. Hidayati, N. Chamidah, and I. N. Budiantara," Spline Truncated Estimator in Multiresponse Semiparametric Regression Model for Computer Based National Exam in West Nusa Tenggara. *IOP Conf Ser Earth Environ Sci*, vol. 546, no. 5, p. 052029, Jun. 2019, doi: 10.1088/1757-899X/546/5/052029.

[4] L. Hidayati, N. Chamidah, and I. N. Budiantara," Confidence Interval of Multiresponse Semiparametric Regression Model Parameters Using Truncated Spline. *Int. J. Acad. Appl. Res. (IJAAR)*, vol. 4, no. 1, 2020.

[5] I. N. Budiantara, V. Ratnasari, M. Ratna, and I. Zain, "The Combination of Spline and Kernel Estimator for Nonparametric Regression and its Properties," *Applied Mathematical Sciences*, vol. 9, no. 122, pp. 6083–6094, 2015, doi: 10.12988/AMS.2015.58517.

[6] V. Ratnasari, I. N. Budiantara, M. Ratna, and I. Zain, "Estimation of Nonparametric Regression Curve Using Mixed Estimator of Multivariable Truncated Spline and Multivariable Deret," *Global Journal of Pure and Applied Mathematics*, vol. 12, no. 6, pp. 5047–5057, 2016.

[7] M. Hadijati, Irwansyah, and N. Fitriyani, "*Particle Swarm Optimization pada Model Nonparametrik Campuran Spline Truncated dan Kernel (Studi Kasus: Pemodelan Indeks Pembangunan Manusia Provinsi Nusa Tenggara Barat)*," 2019.

[8] N. Fitriyani, M. Hadijati, L. Harsyiah, and M. S. Sauri, "Mixed Truncated Spline and Kernel Nonparametric Regression Model on Population Growth Rate in West Nusa Tenggara Province," *IOP Conf Ser Mater Sci Eng*, vol. 1115, no. 1, p. 012054, Mar. 2021, doi: 10.1088/1757-899X/1115/1/012054.

[9] I. N. Budiantara et al., "Modeling Percentage of Poor People in Indonesia Using Kernel and Fourier Series Mixed Estimator in Nonparametric Regression," *Investigacion Operacional*, vol. 40, no. 4, pp. 538–550, 2019.

[10] N. Afifah, "*Estimator Campuran Kernel dan Deret Fourier dalam Regresi Nonparametrik (Studi Kasus: Pemodelan Persentase Penduduk Miskin di Indonesia)*," Institut Teknologi Sepuluh Nopember, Surabaya, 2017.

[11] K. Nisa, I. N. Budiantara, and A. T. Rumiati, "Multivariable Semiparametric Regression Model with Combined Estimator of Fourier Series and Kernel," *IOP Conf Ser Earth Environ Sci*, vol. 58, no. 1, p. 012028, Mar. 2017, doi: 10.1088/1755-1315/58/1/012028.

[12] N. Afifah, I. N. Budiantara, and I. N. Latra, "Mixed Estimator of Kernel and Fourier Series in Semiparametric Regression," *J Phys Conf Ser*, vol. 855, no. 1, p. 012002, Jun. 2017, doi: 10.1088/1742-6596/855/1/012002.

[13] M. Hadijati, Z. W. Baskara, L. Harsyiah, and N. Fitriyani, "*Model Regresi Nonparametrik Campuran Kernel dan Deret Fourier pada Persentase Kemiskinan di Provinsi Nusa Tenggara Barat*," 2021.

[14] K. Nisa, "*Model Regresi Semiparametrik Campuran Spline Truncated dan Deret Fourier (Studi Kasus : Angka Harapan Hidup Provinsi Jawa Timur)*," Institut Teknologi Sepuluh Nopember, Surabaya, 2017.

[15] D. R. S. Saputro, A. Sukmayanti, and P. Widyaningsih, "The Nonparametric Regression Model Using Fourier Series Approximation and Penalized Least Squares (PLS) (Case on Data Proverty in East Java)," *J Phys Conf Ser*, vol. 1188, no. 1, p. 012019, Mar. 2019, doi: 10.1088/1742-6596/1188/1/012019.

[16] Rory, "*Regresi Campuran Nonparametrik Spline Linier Truncated dan Fungsi Kernel untuk Pemodelan Data Kemiskinan di Provinsi Papua*," Institut Teknologi Sepuluh Nopember, Surabaya, 2016.

[17] W. Härdle, *Applied Nonparametric Regression*. Cambridge University Press, Oct. 1990, doi: 10.1017/CCOL0521382483.

[18] L. Hidayati, D. Agustini, Ripai, and Awaludin, "Pemodelan Produktivitas Padi Menggunakan Regresi Semiparametrik Spline Truncated," *Journal of Innovation Research and Knowledge*, vol. 2, no. 3, pp. 913–916, Aug. 2022, doi: 10.53625/JIRK.V2I3.3404.

[19] R. L. Eubank, *Nonparametric Regression and Spline Smoothing.* CRC Press, 1999. doi: 10.1201/9781482273144.

[20] W. Härdle and M. Müller, "*Multivariate and Semiparametric Kernel Regression*," SFB 373 Discussion Papers, 1997.