Received: 02 Apr 2025 | Revised: 29 May 2025 | Accepted: 17 Jun 2025 | Available online: 15 Aug 2025
https://doi.org/10.13057/ijas.v8i1.100915

E-ISSN: 2621-086X
https://jurnal.uns.ac.id/ijas/

Indonesian Journal of Applied Statistics Vol. 8, No.1, pp. 1-12, 2025

Modeling East Java Province Poverty Cases using Biresponse Truncted Spline Regression

Rizka Amalia Putri [1*], Nindya Wulandari [1], Erlyne Nadhilah Widyaningrum [2], Morina A. Fathan [3], Nur Rezky Safitriani [3]

[1] Statistics Study Program, Department of Mathematics, Riau University, Pekanbaru, Indonesia
[2] Department of Statistics, Mulawarman University, Samarinda, Indonesia
[3] Department of Statistics, Tadulako University, Palu, Indonesia
* Corresponding Author. E-mail: rizkaamaliaputri@lecture.unri.ac.id

Abstract

An analytical method for determining the relationship between predictor and response variables is regression. For data that shows unidentified patterns, nonparametric regression is a suitable data analysis technique. A nonparametric regression technique is the truncated spline. Due to the widespread use of truncated spline with a single response variable, this study employs biresponse truncated spline, which uses two response variables to produce a better model than single-response modeling. The purpose of this study is to obtain the best model and to identify which variables influence the poverty case in East Java Province using biresponse truncated spline regression. The best knot points were chosen for this investigation using generalized cross validation (GCV). With three knot points and a model goodness of fit (R^2) of 95.83%, GCV gives the best modeling results. Applying this model to the East Java Province case of poverty using data on the poverty depth index and the percentage of the population living in poverty in 2023 reveals that the labor force participation rate (TPAK), average years of schooling (RLS), and open unemployment rate (TPT) all have a significant effect.

Keywords: biresponse truncated spline; nonparametric regression; poverty.

This is an open access article under the Creative Commons Attrribution-ShareAlike 4.0 International License

How to Cite:
R. A. Putri, N. Wulandari, E. N. Widyaningrum, M. A. Fathan, and N. R. Safitriani, "Modeling East Java Province poverty cases using biresponse truncted spline regression," *Indonesian Journal of Applied Statistics,* vol. 8, no. 1, pp. 1-12, 2025, doi: 10.13057/ijas.v8i1.100915.

1. INTRODUCTION

Regression is one of the modeling methods commonly used in statistics. In order to estimate or predict the average value of the response variable, this analysis focuses on examining the relationship between the response variable and one or more predictor variables [1]. Parametric, semiparametric, and nonparametric regression are the three categories of regression [2]. A regression curve can be used to characterize the relationship between response variables and predictors. It is assumed that the curve's shape is known while using parametric regression. Semiparametric regression is a suitable technique if some of the curve's form is known while other parts is not [3]. The use of nonparametric regression is possible if the regression curve's form is uncertain [4,5]. Spline are one type of nonparametric function estimation technique.

Spline is a model that has a very specific and effective visual and statistical interpretation [6]. Spline models are excellent at managing data that have patterns that vary at certain intervals [7]. With the knot points, the spline curve is able to overcome the problem of drastically changing data, resulting in a smoother curve [8]. Knot point position and number have a significant impact on the spline regression curve form [9].

Spline regression is referred to as univariable regression if it just includes one response variable and one predictor variable. whereas a model is referred to as multivariable if it contains one response variable and multiple predictor variables. Meanwhile, if there is more than one response variable that is correlated, then this model is called biresponse regression with the aim of producing a better model than single response regression, by considering not only the effect of predictors on the response, but also the relationship between responses. Some studies that use the biresponse or multirespon approach include spline biresponse regression on longitudinal data [10], the development of multirespon semiparametric regression [11,12].

The three main factors in building a spline regression model are determining the order of the model, the number of knots, and the position of knot placement [13]. The order of the model is determined based on the pattern that emerges from the data, while the location and number of knots are determined by changes in the data pattern at certain intervals. Some methods for selecting optimal knot points in nonparametric spline regression include cross validation (CV), unbiased risk (UBR) and generalized cross validation (GCV). GCV is a very effective method for determining optimal knot points [14] because it has asymptotic optimal properties, is not affected by transformation, and can simulate repeated observations [15].

Nonparametric regression models can be applied in various disciplines, including economics. In this field, nonparametric regression models are used to analyze issues related to poverty. Poverty is a serious problem faced by many countries, especially developing countries like Indonesia. This problem is very complex, so it requires accurate analysis for its handling [16]. In the context of the sustainable development goals (SDGs), poverty alleviation is a major goal, with the vision of creating a world without poverty, which makes it a global priority, including in Indonesia.

Biresponse truncated spline regression will be used in this study to identify the best model and identify the variables that influence the poverty status in East Java Province. The study's independent variables are the open unemployment rate (TPT), average years of schooling (RLS), and labor force participation rate (TPAK). Nonparametric regression models, with relevant variables, can be an effective alternative for analyzing poverty issues.

2. METHODS

2.1. Biresponse Truncated Spline Nonparametric Regression

Suppose given paired biresponse data $(y_{1i}, y_{2i}, z_{li});\ l = 1,2$ with two response variables and each response has one predictor variable. The biresponse nonparametric regression model is obtained in equation (1).

$$\begin{aligned} y_{1i} &= f(z_{1i}) + \varepsilon_{1i} \\ y_{2i} &= f(z_{2i}) + \varepsilon_{2i};\ i = 1,2,\dots,n \end{aligned} \tag{1}$$

with, $f(z_{li})$ is assumed to have an unknown shape, hence the appropriate approach with the truncated spline function in equation (2).

$$\mathrm{f}(z_{li}) = \sum_{j=0}^{m} \beta_{lj} z_{ij} + \sum_{k=1}^{r} \gamma_{lk} (z_{li} - K_{lk})_+^m + \varepsilon_i;\ i = 1,2,\dots,n;\ l = 1,2 \tag{2}$$

with a truncated function that can be described as follows.

$$(z_{li} - K_{lk})_+^m = \begin{cases} (z_{li} - K_{lk})^m, & \mathrm{z} \geq K_{lk} \\ 0 & , \mathrm{z} < K_{lk} \end{cases} \tag{3}$$

Equation (1) in matrix form is built from the vectors and matrices in equation (4).

$$\mathbf{y} = \boldsymbol{Z\theta} + \boldsymbol{\varepsilon} \tag{4}$$

Copyright © 2025, Author(s)
https://doi.org/10.13057/ijas.v8i1.100915
E-ISSN: 2621-086X

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{Z}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}$$

with, $\mathbf{y}$ is the matrix that contains the response component, $\mathbf{Z}$ is a matrix that contains nonparametric and truncated components, $\boldsymbol{\theta}$ is a matrix containing the nonparametric and truncated parameter coefficients, and $\boldsymbol{\varepsilon}$ is an error vector.

$$y_1 = (y_{11} \quad y_{12} \quad \cdots \quad y_{1n})^T$$
$$y_2 = (y_{21} \quad y_{22} \quad \cdots \quad y_{2n})^T$$

The truncated spline basis matrix is as follows.

$$\boldsymbol{Z}_1 = \begin{bmatrix} 1 & z_{11}^1 & z_{11}^2 & \cdots z_{11}^m & (z_{11}-K_{11})_+^m & \cdots & (z_{11}-K_{1r})_+^m \\ 1 & z_{12}^1 & z_{12}^2 & \cdots z_{12}^m & (z_{12}-K_{11})_+^m & \cdots & (z_{12}-K_{1r})_+^m \\ \vdots & \vdots & \vdots & \ddots \vdots & \vdots & \ddots & \vdots \\ 1 & z_{1n}^1 & z_{1n}^2 & \cdots z_{1n}^m & (z_{1n}-K_{11})_+^m & \cdots & (z_{1n}-K_{1r})_+^m \end{bmatrix}$$

$$\boldsymbol{Z}_2 = \begin{bmatrix} 1 & z_{21}^1 & z_{21}^2 & \cdots z_{21}^m & (z_{21}-K_{21})_+^m & \cdots & (z_{21}-K_{2r})_+^m \\ 1 & z_{22}^1 & z_{22}^2 & \cdots z_{22}^m & (z_{22}-K_{21})_+^m & \cdots & (z_{22}-K_{2r})_+^m \\ \vdots & \vdots & \vdots & \ddots \vdots & \vdots & \ddots & \vdots \\ 1 & z_{2n}^1 & z_{2n}^2 & \cdots z_{2n}^m & (z_{2n}-K_{21})_+^m & \cdots & (z_{2n}-K_{2r})_+^m \end{bmatrix}$$

with the following parameter vector.

$$\boldsymbol{\theta}_1 = (\beta_{10} \quad \beta_{11} \quad \cdots \quad \beta_{1m} \quad \gamma_{11} \quad \gamma_{12} \quad \cdots \quad \gamma_{1r})^T$$
$$\boldsymbol{\theta}_2 = (\beta_{20} \quad \beta_{21} \quad \cdots \quad \beta_{2m} \quad \gamma_{21} \quad \gamma_{22} \quad \cdots \quad \gamma_{2r})^T$$

and the error vector are as follows.

$$\boldsymbol{\varepsilon_1} = (\varepsilon_{11} \quad \varepsilon_{12} \quad \cdots \quad \varepsilon_{1n})^T$$
$$\boldsymbol{\varepsilon_2} = (\varepsilon_{21} \quad \varepsilon_{22} \quad \cdots \quad \varepsilon_{2n})^T$$

Then with weighted least square optimization (WLS) we can obtain the parameter estimates. Determining the weight matrix $\mathbf{V}$ in this case by calculating both the first and second responses covariance variance values and then defining the covariance variance matrix as the weight $\mathbf{V}$. The weight matrix $\mathbf{V}$ can be written.

$$\boldsymbol{V} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 & \vdots & \sigma_{12} & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \cdots & 0 & \vdots & 0 & \sigma_{12} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_1^2 & \vdots & 0 & 0 & \cdots & \sigma_{12} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{12} & 0 & \cdots & 0 & \vdots & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & \sigma_{12} & \cdots & 0 & \vdots & 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_{12} & \vdots & 0 & 0 & \cdots & \sigma_2^2 \end{bmatrix}$$

The estimator in equation (4) is obtained by solving parameter optimization with weighted least square (WLS).

$$\min_{\theta}\{\boldsymbol{\varepsilon}'\mathbf{V}\boldsymbol{\varepsilon}\} = \min_{\theta}\{(\mathbf{y}-\mathbf{Z}\boldsymbol{\theta})'\mathbf{V}(\mathbf{y}-\mathbf{Z}\boldsymbol{\theta})\}$$

To solve the above equation, partial derivation is done by initializing the function,

$$Q(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}'\mathbf{V}\boldsymbol{\varepsilon} = (\mathbf{y}-\mathbf{Z}\boldsymbol{\theta})'\mathbf{V}(\mathbf{y}-\mathbf{Z}\boldsymbol{\theta})$$

Thus, it is obtained:

$$\begin{aligned} Q(\boldsymbol{\theta}) &= (\mathbf{y}-\mathbf{Z}\boldsymbol{\theta})'\mathbf{V}(\mathbf{y}-\mathbf{Z}\boldsymbol{\theta}) \\ &= (\mathbf{y}'-\mathbf{Z}'\boldsymbol{\theta})(\mathbf{V}\mathbf{y}-\mathbf{V}\mathbf{Z}\boldsymbol{\theta}) \\ &= \mathbf{y}'\mathbf{V}\mathbf{y}-\mathbf{y}'\mathbf{V}\mathbf{Z}\boldsymbol{\theta}-\mathbf{Z}'\boldsymbol{\theta}'\mathbf{V}\mathbf{y}+\mathbf{A}\mathbf{Z}'\boldsymbol{\theta}'\boldsymbol{V}\mathbf{Z}\boldsymbol{\theta} \end{aligned}$$

Copyright © 2025, Author(s)
https://doi.org/10.13057/ijas.v8i1.100915
E-ISSN: 2621-086X

$$= \mathbf{y}'\mathbf{V}\mathbf{y} - 2\boldsymbol{\theta}'\mathbf{Z}'\mathbf{V}\mathbf{y} + \boldsymbol{\theta}'\mathbf{Z}'\mathbf{V}\mathbf{Z}\boldsymbol{\theta}$$

The equation obtained is then differentiated against $\boldsymbol{\theta}$ so that the following results are obtained.

$$\frac{\partial\big(Q(\boldsymbol{\theta})\big)}{\partial\boldsymbol{\theta}} = \frac{\partial(\mathbf{y}'\mathbf{V}\mathbf{y} - 2\boldsymbol{\theta}'\mathbf{Z}'\mathbf{V}\mathbf{y} + \boldsymbol{\theta}'\mathbf{Z}'\mathbf{V}\mathbf{Z}\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$$
$$= -2\mathbf{Z}'\mathbf{V}\mathbf{y} + 2\mathbf{Z}'\mathbf{V}\mathbf{Z}\boldsymbol{\theta}$$

After differentiating $\boldsymbol{\theta}$, the differential result is equated to zero so that the parameter estimation result is obtained.

$$-2\mathbf{Z}'\mathbf{V}\mathbf{y} + 2\mathbf{Z}'\mathbf{V}\mathbf{Z}\hat{\boldsymbol{\theta}} = 0$$
$$\mathbf{Z}'\mathbf{V}\mathbf{Z}\hat{\boldsymbol{\theta}} = \mathbf{A}\mathbf{Z}'\mathbf{V}\mathbf{y}$$
$$\hat{\boldsymbol{\theta}} = (\mathbf{Z}'\mathbf{V}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}\mathbf{y}$$

The estimate form of the spline model in biresponse nonparametric regression hence becomes.

$$\hat{\mathbf{y}} = \mathbf{Z}\hat{\boldsymbol{\theta}}$$
$$= \mathbf{Z}\big((\mathbf{Z}'\mathbf{V}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}\mathbf{y}\big)$$

2.2. Correlation Testing

The Pearson correlation coefficient is used to see the relationship between response variables, which is calculated by the formula in equation (5).

$$r_{1,2} = \frac{\sum_{i=1}^{n}(y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2)}{\left(\sum_{i=1}^{n}(y_{1i} - \bar{y}_1)^2\right)^{1/2}\left(\sum_{i=1}^{n}(y_{2i} - \bar{y}_2)^2\right)^{1/2}} \tag{5}$$

The value of the correlation coefficient in equation (5) is between -1 and 1. If it is closer to -1 or 1, the relationship between responses is stronger, while if it is close to zero, the relationship between responses is weaker [15]. The hypothesis for this test is as follows:

$H_0 : \rho_{1,2} = 0$ (there is no correlation)

$H_1 : \rho_{1,2} \neq 0$ (there is a correlation)

The hypothesis that there is significant correlations between the first and second response variables is tested using the t correlation significance test and the following formula:

$$t = \frac{r_{1,2}\sqrt{n-2}}{\sqrt{1-{r_{1,2}}^2}}$$

The test criterion is H_0 rejected is if $t \geq t_{\frac{\alpha}{2};df}$ and df stands for degrees of freedom.

2.3. Optimal Point Selection

The knot point is the joint point that displays the shift in the data pattern or function behavior. So the location and number of knot points are the most important things in truncated spline regression modeling to obtain the best spline regression model, so the optimal knot points must fit the data. In this study the authors used the GCV method in equation (6).

$$\text{GCV}(\phi_{opt}) = \left[\frac{MSE(\phi)}{\big((n)^{-1}trace[\boldsymbol{I} - \boldsymbol{A}(\phi)]\big)^2}\right] \tag{6}$$

where, I is s the identity matrix and $MSE(\phi)$ is obtained by equation (7).

$$\text{MSE}(\phi) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{7}$$

The selection of the optimal knot point value is taken from the smallest GCV value.

Copyright © 2025, Author(s)
https://doi.org/10.13057/ijas.v8i1.100915
E-ISSN: 2621-086X

2.4. The Concept of Poverty

Poverty is characterized by the inability to pay for necessities like housing, food, clothes, healthcare, and education [16]. In order to determine the number of persons living in poverty, BPS has consistently employed the basic needs approach, which defines poverty as the inability to meet one's fundamental needs, including those related to food and non-food. This method uses the head count index (HCI), which calculates the proportion of the population living below the poverty line. Other measures of poverty levels, such as the Foster-Greer-Thorbecke-developed poverty severity index (P2) and the poverty depth index (P1), are employed in addition to the HCI or P0. The poverty depth index calculates the average gap between the poor's spending and the poverty line, whereas the percentage of the poor indicates the percentage of the population that lives below the poverty line. The difference between the average expenditure of the poor and the poverty line increases with the index value.

2.5. Data Sources, Variables, and Model Implementation Steps

Secondary data was taken from the East Java Province's Badan Pusat Statistik (BPS). Cross-sectional data from 38 districts and cities in East Java Province in 2023 is used in this study. The variables include the percentage of poor population (y_1), poverty depth index ((y_2), open unemployment rate (z_1), average years of schooling (z_2), and labor force participation rate (z_3),

1. Make a scatterplot of each response variable with the predictor variables to see the relationship pattern.
2. Obtain the best model by selecting the optimal knot points from the minimum GCV.
3. Testing the goodness of the model and fulfillment of residual assumptions
4. Interpreting the model and Conclusion.

3. RESULTS AND DISCUSSION

3.1. Descriptive Statistics

The descriptive results of statistical analysis for each variable used are listed in Table 1.

Table 1. Descriptive of response variables and predictor variables

Variable	Mean	StatDev	Min	Max	Range
y_1	10.293	4.321	3.310	21.760	18.450
y_2	1.493	0.808	0.350	4.500	4.150
z_1	4.663	1.429	1.710	8.050	6.340
z_2	8.376	1.658	5.070	11.820	6.750
z_3	73.159	3.767	66.890	81.640	14.750

3.2. Correlation Test

An assumption that needs to be fulfilled in birresponse nonparametric regression is that there is a correlation between the two response variables. Therefore, a correlation test will be conducted between y_1 (percentage of poor people) and y_2 (poverty depth index). Based on the results of the analysis, a Pearson correlation of 0.89835 was obtained, indicating that there is a strong relationship between the two response variables. Then, the t-value is 12.271 and t_{table} is 1.687. As well as a *p-value* of 2.009×10^{-14}, with $\alpha = 0.05$. Based in the t test dan *p-value* obtained, there is a significant correlation between variables y_1 and y_2. So the assumption that the two response variables must be correlated has been fulfilled.

3.3. Identify Relationship Patterns

First, to create a birresponse model, we must pay attention to the scatterplot pattern between y_1 and y_2 with each variables. Figure 1 is a *scatterplot* y_1 and y_2 with each predictor variables. Based on the figure, it can be seen that the scatterplots of y_1 and y_2 with several predictor variables z_1, z_2, and z_3 end not to form a certain pattern. Thus, the y_1 and y_2 data with the factors that are thought to affect it will be modeled using biresponse truncated spline nonparametric regression.

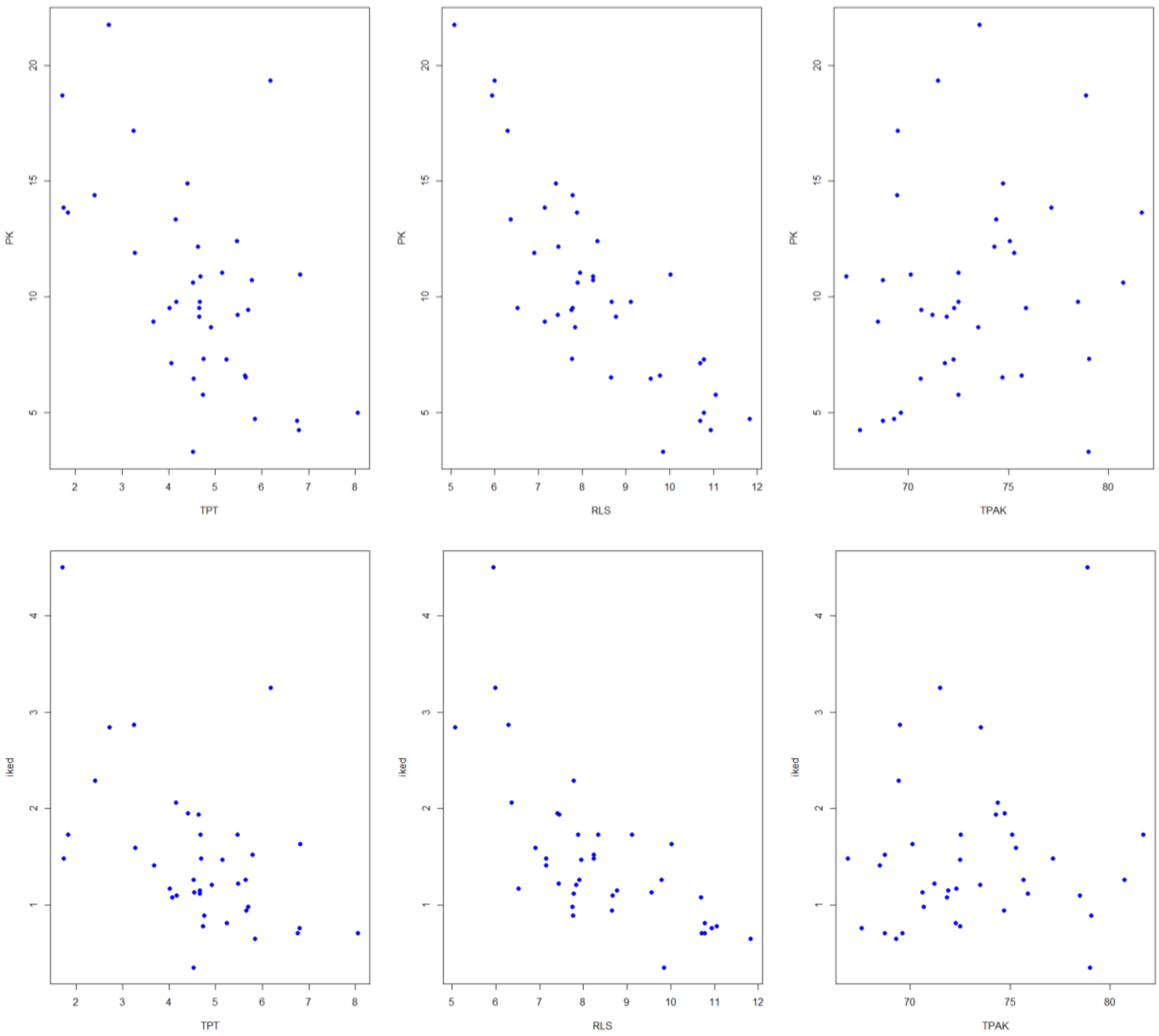

Figure 1. Scatterplot y_1 and y_2 against z_1, z_2, and z_3

3.4. Optimal Knot Point Selection

The best model is selected by comparing the smallest GCV value of many knot points in Table 2. In this regression analysis, the upper limit of three knots was chosen to avoid overfitting, even though the GCV value tends to drop as the number of knots grows, and to preserve a balance between model complexity and the interpretability of the results.

Table 2. Smallest GCV comparison

Knot	Smallest GCV
1	0.5897
2	0.5099
3	0.4338

Copyright © 2025, Author(s)
https://doi.org/10.13057/ijas.v8i1.100915
E-ISSN: 2621-086X

Table 2 shows that the smallest GCV value is 0.4338 resulting from three knot points.

3.5. Formed Biresponse Truncated Spline Regression Model

Based on the above results, the best biresponse truncated spline nonparametric regression equation is obtained with three knot points.

$$\hat{y}_{1i} = \hat{\beta}_{10} + \hat{\beta}_{111}z_{11i} + \hat{\beta}_{121}z_{21i} + \hat{\beta}_{131}z_{31i} + \hat{\gamma}_{111}(z_{11i} - K_{111})_+ + \hat{\gamma}_{112}(z_{11i} - K_{112})_+ + \hat{\gamma}_{113}(z_{11i} - K_{113})_+ + \hat{\gamma}_{121}(z_{21i} - K_{121})_+ + \hat{\gamma}_{122}(z_{21i} - K_{122})_+ + \hat{\gamma}_{123}(z_{21i} - K_{123})_+ + \hat{\gamma}_{131}(z_{31i} - K_{131})_+ + \hat{\gamma}_{132}(z_{31i} - K_{132})_+ + \hat{\gamma}_{133}(z_{31i} - K_{133})_+$$

$$\hat{y}_{2i} = \hat{\beta}_{20} + \hat{\beta}_{211}z_{12i} + \hat{\beta}_{221}z_{22i} + \hat{\beta}_{231}z_{32i} + \hat{\gamma}_{211}(z_{12i} - K_{211})_+ + \hat{\gamma}_{212}(z_{12i} - K_{212})_+ + \hat{\gamma}_{213}(z_{12i} - K_{213})_+ + \hat{\gamma}_{221}(z_{22i} - K_{221})_+ + \hat{\gamma}_{222}(z_{22i} - K_{222})_+ \hat{\gamma}_{223}(z_{22i} - K_{223})_+ + \hat{\gamma}_{231}(z_{32i} - K_{231})_+ + \hat{\gamma}_{232}(z_{32i} - K_{232})_+ + \hat{\gamma}_{233}(z_{32i} - K_{233})_+$$

Furthermore, the biresponse truncated spline nonparametric regression model on the data of poverty cases in East Java Province in 2023 is obtained as follows:

$$\hat{y}_{1i} = 32.315 + 2.956z_{11i} - 25.869(z_{11i} - 2.847)_+ - 1.036(z_{11i} - 3.004)_+ + 23.798(z_{11i} - 3.133)_+ + 1.868z_{21i} - 95.929(z_{21i} - 6.309)_+ + 144.905(z_{21i} - 6.448)_+ - 52.3749(z_{21i} - 6.658)_+ + - 0.418z_{31i} + 49.306(z_{31i} - 69.599)_+ - 92.593(z_{31i} - 69.900)_+ +43.555(z_{31i} - 70.201)_+$$

$$\hat{y}_{2i} = -0.050 - 0.884z_{12i} + 0.915(z_{12i} - 2.847)_+ + 0.296(z_{12i} - 3.004)_+ - 0.322(z_{12i} - 3.133)_+ + - 0.125z_{22i} - 4.580(z_{22i} - 6.309)_+ - 0.329(z_{22i} - 6.4480)_+ + 4.875(z_{22i} - 6.658)_+ + 0.091z_{32i} - 0.119(z_{32i} - 69.599)_+ - 0.055(z_{32i} - 69.900)_+ + 0.048(z_{32i} - 70.201)_+$$

3.6. Model Fit Test and Residual Assumptions

The best model obtained has an R^2 value of 95.83%. Furthermore, there are several residual assumptions that must be met from the residuals in the model, namely the residual normality test on each response variable, the independence test on the residuals of each response variable, the homoscedasticity test on each response variable residual, and the normal multivariate test for both response variables. All residual assumption tests are met in this model.

3.7. Model Interpretation

Model for Variable z_1

Assuming the data other than variable z_1 and z_2 are denoted as c1 and c2, the general model is obtained as follows.

$$\hat{y}_{1i} = 2.956z_{11i} - 25.869(z_{11i} - 2.847)_+ - 1.036(z_{11i} - 3.004)_+ + 23.798(z_{11i} - 3.133)_+ + c_1$$
$$\hat{y}_{2i} = -0.884z_{12i} + 0.915(z_{12i} - 2.847)_+ + 0.296(z_{12i} - 3.004)_+ - 0.322(z_{12i} - 3.133)_+ + c_2$$

where, the truncated function for variable z_1 and z_2 are as follows:

$$\hat{y}_{1i} = \begin{cases} 2.956z_{11i} + c_1 & , z_{11i} \leq 2.869 \\ 26.800 + 28.825z_{11i} + c_1 & , 2.869 < z_{11i} \leq 3.004 \\ 29.912 + 27.789z_{11i} + c_1 & , 3.004 < z_{11i} \leq 3.133 \\ -44.647 + 51.587z_{11i} + c_1 & , z_{11i} > 3.133 \end{cases}$$

$$\hat{y}_{2i} = \begin{cases} -0.884z_{12i}+c_2 & , z_{12i} \leq 2.869 \\ -2.605+0.031z_{12i}+c_2 & , 2.869 < z_{12i} \leq 3.004 \\ -3.494 + 0.327z_{12i}+c_2 & , 3.004 < z_{12i} \leq 3.133 \\ -2.485+0.005z_{12i}+c_2 & , z_{12i} > 3.133 \end{cases}$$

The following is a list of regions in East Java Province based on both poverty variables by TPT, according to the selected knot points.

(a) Pacitan District, Sumenep District, Pamekasan District, Sampang District, Ngawi District are districts/cities with TPT less than or equal to 2.869.
(b) There were no districts/cities with a TPT between more than 2.869 and less than 3.004.
(c) There were no districts/cities where the TPT was between more than 3.004 and less than 3.133.
(d) Probolinggo District, Lumajang District, Kediri City, Bondowoso District, Magetan District, Tuban District, Trenggalek District, Jombang District, Probolinggo City, Bojonegoro District, Ponorogo District, Batu City, Situbondo District, Mojokerto District, Nganjuk District, Mojokerto City, Banyuwangi District, Madiun District, Blitar City, Lamongan District, Jember District, Pasuruan District, Pasuruan City, Tulungagung District, Blitar District, Malang District, Kediri District, Madiun City, Bangkalan District, Surabaya City, Gresik District, Malang City, and Sidoarjo District are districts/municipalities with a TPT of more than 3.133.

Based on the model, it is known that :

(i) When the TPT is less than 2.869%, every 1% increase in TPT in region (a) tends to increase the percentage of poor people in region (a) by 2.956% and when a 1% increase in TPT in region (a) also decreases the poverty depth index in region (a) by 0.884%.
(ii) When the TPT lies between 2.869% and 3.004%, every 1% increase in the TPT in region (b) tends to increase the percentage of poor people in region (b) by 28.825% and when the TPT increases by 1%, it increases the poverty depth index in region (b) by 0.031%.
(iii) When the TPT lies between 3.004% and 3.133%, every 1% increase in the TPT in region (c) tends to increase the percentage of poor people by 27.789% and when the TPT increases by 1%, it increases the poverty depth index in region (c) by 0.327%.
(iv) When the TPT is more than 3.133%, every 1% increase in the TPT in region (d) tends to increase the percentage of poor people by 51.587%. And when the increase in the TPT by 1 increases the poverty depth index by 0.005%.

Based on the interpretation of this model, it is found that the TPT has a positive effect on both response variables. The higher the percentage of poor people and also the poverty depth index, the higher the TPT. Therefore, the government can take steps to reduce the percentage and depth index of poverty so that the TPT can also decrease.

Model for Variable z_2

Assuming the data other than the variable z_2 is denoted as c1 and c2, the general model is obtained as follows.

$$\hat{y}_{1i} = 1.868z_{21i} - 95.929(z_{21i} - 6.309)_+ + 144.905(z_{21i} - 6.448)_+ - 52.374(z_{21i} - 6.658)_+ + c_1$$

$$\hat{y}_{2i} = -0.125z_{22i} - 4.580(z_{22i} - 6.309)_+ - 0.329(z_{22i} - 6.448)_+ + 4.875(z_{22i} - 6.658)_+ + c_2$$

where, the truncated function for the RLS variable is as follows:

$$\hat{y}_{1i} = \begin{cases} 1.868z_{21i}+c_1 & , z_{21i} \leq 6.309 \\ 605.197 - 4.705z_{21i}+c_1 & , 6.309 < z_{21i} \leq 6.448 \\ -329.150 + 140.2z_{21i}+c_1 & , 6.488 < z_{21i} \leq 6.658 \\ 19.562+87.825z_{21i}+c_1 & , z_{21i} > 6.658 \end{cases}$$

$$\hat{y}_{2i} = \begin{cases} -0.1025z_{22i}+c_2 & , z_{22i} \leq 6.309 \\ 28.895 - 4.705z_{22i}+c_2 & , 6.309< z_{22i} \leq 6.448 \\ 31.016 - 5.034z_{22i}+c_2 & , 6.448< z_{22i} \leq 6.658 \\ -1.442 - 0.159z_{22i}+c_2 & , z_{22i}> 6.658 \end{cases}$$

The following is a list of regions in East Java Province based on both poverty variables by RLS, according to the selected knot points.

(a) Sampang District, Sumenep District, Probolinggo District And Bangkalan District are district/cities with RLS less than equal to 6.309 years.
(b) Bondowoso district is a district with RLS between more than 6.309 years and less than 6.448 years.
(c) Jember district is a district/cities with an RLS of more than 6.448 years to less than equal to 6.658 years.
(d) Situbondo District, Lumajang District, Pamekasan District, Tuban District, Pasuruan District, Bojonegoro District, Malang District, Banyuwangi District, Ngawi District, Ponorogo District, Blitar District, Pacitan District, Trenggalek District, Madiun District, Nganjuk District, Kediri District, Lamongan District, Tulungagung District, Tulungagung District, Magetan District, Jombang District, Mojokerto District, Probolinggo City, Pasuruan Citu, Batu City, Gresik District, Kediri City, Surabaya City, Blitar City, Sidoarjo District, Malang City, Mojokerto District, and Madiun City are district/cities with RLS of more than 6.658 years.

Based on the model, it is known that:

(i) When the RLS is less than 6.309 years, every 1 year increase in RLS in region (a) tends to increase the percentage of poor people in region (a) by 1.868% and every 1 year increase in RLS in region (a) but tends to decrease the poverty depth index in region (a) by 0.1025.
(ii) When the RLS in region (b) lies between 6.309 to 6.488 years, every increase in RLS by 1 year in region (b) tends to reduce the percentage of poor people in region (b) by 4.705% and reduce the poverty depth index in region (b) by 4.705.
(iii) When the RLS in region (c) lies between 6.488 and 6.658 years, then each increase in the RLS in region (c) by 1 year tends to increase the percentage of poor people in region (c) by 140.2% but decrease the poverty depth index in region (c) by 5.034.
(iv) Meanwhile, when the RLS in region (d) is more than 6.658 years, every increase in RLS in region (d) by 1 year tends to increase the percentage of poor people in region (d) by 87.82% but decrease the poverty depth index in region (d) by 0.159.

It is known that the RLS variable has a negative effect on the poverty depth index, in accordance with economic theory. This means that the higher the RLS, the lower the poverty depth index. Therefore, the government can reduce poverty by increasing RLS. However, on the poverty percentage variable, RLS has a positive effect. This does not mean that to reduce poverty, the government must reduce RLS. This is because poverty can be reduced by the influence of other predictor variables, even though RLS is decreasing.

Model for Variable z_3

Assuming the data other than the variable z_3 are denoted as c1 and c2, the general model is obtained as follows.

$$\hat{y}_{1i} = -0.418z_{31i} + 49.306(z_{31i} - 69.599)_+ - 92.593(z_{31i} - 69.900)_+ +43.555(z_{31i} - 70.201)_+ + c_1$$

$$\hat{y}_{2i} = 0.091z_{32i} - 0.119(z_{32i} - 69.599)_+ - 0.055(z_{32i} - 69.900)_+ + 0.048(z_{32i} - 70.201)_+ + c_2$$

where, the truncated function for the TPAK variable is as follows:

Copyright © 2025, Author(s)
https://doi.org/10.13057/ijas.v8i1.100915
E-ISSN: 2621-086X

$$\hat{y}_{1i} = \begin{cases} -0.418z_{31i}+c_1 & , z_{31i} \leq 69.599 \\ -3431.648+48.888z_{31i}+c_1 & , 69.599< z_{31i} \leq 69.900 \\ 2957.062 - 43.705z_{31i}+c_1 & , 69.900< z_{31i} \leq 70.201 \\ -100.542 - 0.15z_{31i}+c_1 & , z_{31i}> 70.201 \end{cases}$$

$$\hat{y}_{2i} = \begin{cases} 0.091z_{32i}+c_2 & , z_{32i} \leq 69.599 \\ 8.282 - 0.028z_{32i}+c_2 & , 69.599< z_{32i} \leq 69.900 \\ 12.126 - 0.083z_{32i}+c_2 & , 69.900< z_{32i} \leq 70.201 \\ 8.756 - 0.035z_{32i}+c_2 & , z_{32i}> 70.201 \end{cases}$$

The following is a list of regions in East Java Province based on the two poverty variables according to TPAK, according to the selected knot points.

(a) Surabaya City, Nganjuk City, Malang City, Lumajang District, Kediri District, Madiun City, Ngawi District, and Probolinggo District are districts/cities with a TPAK of less than or equal to 69.599%.
(b) Sidoarjo district is a district/cities with a TPAK of more than 69.599% to less than or equal to 69.900%.
(c) Gresik district has a TPAK of more than 69.900% to less than or equal to 70.201%.
(d) Probolinggo City, Malang District, Pasuruan District, Bangkalan District, Jombang District, Blitar City, Madiun District, Mojokerto City, Mojokerto District, Jember District, Kediri City, Blitar District, Sampang District, Bojonegoro District, Bondowoso District, Tulungagung District, Tuban District, Bondowoso District, Tulungagung District, Tuban District, Lamongan District, Situbondo District, Pasuruan District, Ponorogo District, Pamekasan District, Magetan District, Sumenep District, Batu City, Banyuwangi Distrik, Trenggalek Distrik, and Pacitan Distrik are districts/cities with a TPAK of more than 70.201%.

Based on the model, it is known that:

(i) When the TPAK in region (a) is less than 69.599%, every 1% increase in the TPAK in region (a) tends to decrease the percentage of poor people in region (a) by 0.418% and every 1% increase in the TPAK in region (a) tends to increase the poverty depth index in region (a) by 0.091%.
(ii) When the TPAK in region (b) lies between 69.599% and 69.900%, every 1% increase in the TPAK in region (b) tends to increase the percentage of poor people in region (b) by 48.888% but decrease the poverty depth index in region (b) by 0.028. However, this does not mean that to reduce poverty, the government must reduce TPAK. This is because there are other variables that can reduce poverty when TPAK is falling, and only Sidoarjo district is in this range.
(iii) Meanwhile, when the TPAK in region (c) lies between 69.900% and 70.201%, each 1% increase in TPAK in region (c) tends to reduce the percentage of poor people in region (c) by 43.705% and also reduce the poverty depth index in region (c) by 0.083.
(iv) When the TPAK in region (d) is more than 70.201%, every 1% increase in the TPAK in region (d) tends to reduce the percentage of poor people in region (d) by 0.15% and reduce the poverty depth index in region (d) by 0.035%.

According to economic theory, it is well known that the TPAK variable has a negative impact on both the poverty depth index and the percentage of poor people. This indicates that the poverty depth index and the percentage of poor people decrease with increasing TPAK. The government can therefore reduce poverty rates by attempting to increase TPAK. This conclusion is consistent with studies [17] that showed the significant effect of TPAK on poverty rates.

3.8. Discussion

According to this study, the most effective model for estimating the poverty depth index and the percentage of poor people in East Java Province in 2023 is the biresponse truncated spline regression model with three knots. Poverty rates were found to be significantly impacted by the three independent variables under analysis there are TPT, RLS, and TPAK. This results is consistent with earlier study [18],

Copyright © 2025, Author(s)
https://doi.org/10.13057/ijas.v8i1.100915
E-ISSN: 2621-086X

which similarly employed a truncated spline model and found that TPT significantly reduced poverty, with a coefficient of determination (R^2) of 88.39%. However, a different study [19] that employed the B-spline approach with a comparable example only managed to generate an R^2 of 67.79%, although both studies only used one response variable.

The constructed model has stronger and more effective predictive capacities in explaining the variation in poverty data in the region, as evidenced by the large rise in the R^2 value to 95.83% that was obtained by using two response variables in this study.

4. CONCLUSIONS

The best birespons truncated spline nonparametric regression model was created for the East Java Province poverty case in 2023 with three knot points, according to the data analysis that was completed. Where, the above model produces an R^2 of 95.83%.

$$\hat{y}_{1i} = 32.315 + 2.956z_{11i} - 25.869(z_{11i} - 2.847)_+ - 1.036(z_{11i} - 3.004)_+ + 23.798(z_{11i} - 3.133)_+ + 1.868z_{21i} - 95.929(z_{21i} - 6.309)_+ + 144.905(z_{21i} - 6.448)_+ - 52.3749(z_{21i} - 6.658)_+ + - 0.418z_{31i} + 49.306(z_{31i} - 69.599)_+ - 92.593(z_{31i} - 69.900)_+ + 43.555(z_{31i} - 70.201)_+$$

$$\hat{y}_{2i} = -0.050 - 0.884z_{12i} + 0.915(z_{12i} - 2.847)_+ + 0.296(z_{12i} - 3.004)_+ - 0.322(z_{12i} - 3.133)_+ + - 0.125z_{22i} - 4.580(z_{22i} - 6.309)_+ - 0.329(z_{22i} - 6.448)_+ + 4.875(z_{22i} - 6.658)_+ + 0.091z_{32i} - 0.119(z_{32i} - 69.599)_+ - 0.055(z_{32i} - 69.900)_+ + 0.048(z_{32i} - 70.201)_+$$

In East Java Province, poverty has been successfully modeled using the implementation of nonparametric regression, particularly the birespons truncated spline model.

REFERENCES

[1] D. N. Gujarati, *Basic Econometric*, 4th edition, New York: McGraw-Hill Companies, Inc., 2003.

[2] Sriliana, I. N. Budiantara, and V. Ratnasari, "A truncated spline and local linear mixed estimator in nonparametric regression for longitudinal data and its application," *Symmetry*, vol. 14, 2022, doi: 10.3390/sym14122687.

[3] D. P. Rahmawati, I. N. Budiantara, D. D. Prastyo, and M. A. D. Octavanny, "Mixture spline smoothing and kernel estimator in multiresponse nonparametric regression," *IAENG Int. J. Appl. Math*, vol. 51, no. 28, 2021, https://www.iaeng.org/IJAM/issues_v51/issue_3/IJAM_51_3_28.pdf.

[4] V. Ratnasari, I. N. Budiantara, and A. T. R. Dani, "Nonparametric regression mixed estimators of truncated spline and gaussian kernel based on cross-validation (CV), generalized cross-validation (GCV), and unbiased risk (UBR) methods," *Int. J. Adv. Sci. Eng. Inform. Technol*, vol. 11, pp. 2400-2406, 2021, https://ijaseit.insightsociety.org/index.php/ijaseit/article/view/14464.

[5] S. Sifriyani, S. H. Kartiko, I. N. Budiantara, and G. Gunardi, "Development of nonparametric geographically weighted regression using truncated spline approach," *Songklanakarin J. Sci. Techno.*, vol. 40, pp. 909-920, 2018, doi: 10.14456/sjst-psu.2018.98.

[6] A. T. R. Dani, L. Ni'matuzzahroh, V. Ratnasari, & I. N Budiantara, "Pemodelan regresi nonparametrik spline truncated pada data longitudinal," *Inferensi*, vol.4, no.1, pp. 47, 2021, doi:10.12962/j27213862.v4i1.8737.

Copyright © 2025, Author(s)

[7] R. L Eubank, *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, Inc., New York., 1999.

[8] E. Montoya, N. Ulloa, and V. Miller, "A simulation study comparing knot selection methods with equally spaced knots in a penalized regression spline", *International Journal of Statistics and Probability*, vol. 3, no. 3, 2014.

[9] T. Similia, and J. Tikka, *Input Selection and Shrinkage in Multiresponse Linear Regression*. Preprint Submitted to Elsevier, 2007.

[10] A. A. R. Fernandes, *Estimator Spline dalam Regresi Nonparametrik Birespon untuk Data Longitudinal (Studi Kasus pada Pasien Penderita TB Paru di Malang)* (Doctoral dissertation, Institut Teknologi Sepuluh Nopember), 2016.

[11] L. Hidayah, N. Cahmidah, and I. N. Budiantara, "Spline truncated estimator in multiresponse semiparametric regression model for computer based national exam in West Nusa Tenggara," *IOP Conference Series: Materials Science and Engineering*, vol. 546, no. 5, pp. 20-29, 2019, doi: 10.1088/1757-899X/546/5/052029.

[12] D. Amelia, *Model Regresi Nonparametrik Multirespon Spline Truncated Untuk Data Longitudinal*. Tesis Jurusan Statistika FMIPA Institut Teknologi Sepuluh November. Surabaya, 2016.

[13] G. Wahba, *Spline Models for Observation Data*. Penylvania: SIAM., 1990.

[14] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, Inc., New York., 1999.

[15] N. R. Draper, and H. Smith, *Applied Regression Analysis*. New York: John Wiley & Sons, 1998.

[16] Badan Pusat Statistik, *Jumlah penduduk miskin, persentase penduduk miskin dan garis kemiskinan* 1970-2013, 2015.

[17] D. Desmawan, A. K. Salsabila, L. Amalia, R. A. Anargya, R. S. Kirana, V. Valentina, "Analisis pengaruh tingkat partisipasi angkatan kerja dan upah minimum provinsi terhadap kemiskinan di Provinsi Banten," *Jurnal Manajemen Akuntansi (JUMSI)*, vol. 3, no. 2, pp. 649-657, 2023, doi: 10.36987/jumsi.v3i2.4124.

[18] N. P. A. Mariati, I. W. Sudiarsa, dan N. M. S. Sanjiwani, "Perbandingan Regresi Linier Berganda dengan Spline Truncated (Studi Kasus: Kemiskinan Di Provinsi Papua)," *Widyadari*, vol. 23, no. 2, pp. 240-246, 2022, doi: 10.5373/jardcs/v12i8/20202612.

[19] A. S. Rahmawati, D. Ispriyanti, and B. Warsito, "Pemodelan Kasus Kemiskinan di Jawa Tengah Menggunakan Regresi Nonparametrik Metode B-Spline," *Jurnal Gaussian*, vol. 6, no. 1, pp. 11-20, 2017, doi: 10.14710/j.gauss.6.1.11-20.

Copyright © 2025, Author(s)
https://doi.org/10.13057/ijas.v8i1.100915
E-ISSN: 2621-086X