

# Spline and Kernel Mixed Nonparametric Regression for Malnourished Children Model in West Nusa Tenggara

Muhammad Sopian Sauri<sup>1\*</sup>, Mustika Hadijati<sup>2</sup>, Nurul Fitriyani<sup>3</sup>

<sup>1</sup>Mathematics Departement, Universitas Mataram, Indonesia, e-mail : sopiangamatika@gmail.com,

<sup>1</sup>Mathematics Departement, Universitas Mataram, Indonesia, e-mail : mustika.hadijati@unram.ac.id,

<sup>1</sup>Mathematics Departement, Universitas Mataram, Indonesia, e-mail : nurul.fitriyani@unram.ac.id

## ABSTRACT

### Article history:

Received : 23-12-2020

Revised : 16-01-2021

Accepted : 21-01-2021

### Keyword:

Bandwidth;

Kernel;

Knot;

Malnutrition;

Truncated Spline

Health sector development is essential to improve human life quality, especially in West Nusa Tenggara (NTB) Province. Based on data from the NTB Provincial Health Office from 2011 to 2016, children under five suffering from malnutrition continued to increase, caused by several factors that affected the incident. Therefore, appropriate analysis is needed to model children who suffer from malnutrition in NTB Province in 2016, consisting of 10 districts based on the variables that influence it. The analysis in this study was carried out using a nonparametric regression mixed-model spline truncated and kernel. The estimation of the nonparametric regression curve depends on the optimal knot points and bandwidths parameter. Therefore, in determining the optimal knot points and bandwidths obtained from Generalized Cross-Validation (GCV). Based on the analysis that has been done, we obtained a nonparametric regression mixed-model spline truncated and kernel optimal knot points, such as  $k_{11} = 10.13846154$  ;  $k_{12} = 72.19692308$  ;  $k_{13} = 44.25846154$  ;  $k_{14} = 91.01461538$  for each variable and optimum bandwidths, such as  $h_1 = 0.5561442$  ;  $h_2 = 1.0220133$  ;  $h_3 = 0.7163110$  ;  $h_4 = 1.2240464$  ; and  $h_5 = 1.2900146$  , with 0.003038719 the value of GCV. The mixed model acquired has a good model by considering the values of  $R^2$  and MSE. Besides, the MAPE value indicated a high degree of accuracy, so that the model obtained has an excellent forecast.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



DOI: <https://doi.org/10.30812/varian.v4i2.1003>



## A. INTRODUCTION

Building healthiness was organized to improve awareness, desire, and capability to live healthy for everyone to create high public health. To build it as soon as possible, a sound health information system is needed, especially in West Nusa Tenggara province (NTB). Improving the quality of human life in society is one of the most critical parts to make a better future. One of the indicators to monitor the health society level is to see the status of baby nutrition. The nutrition status is a description of balance condition in a particular variable form. If the condition is disturbed, so it tends to be an inference of body growth. Many found that mothers give birth with low weight, even the dead baby is caused by a mother who has malnutrition before giving birth. A baby with low weight can affect deficient nutrients, even malnourished (BPS, 2016).

According to the healthcare center of NTB province year 2016, there are many malnourished cases found in West Nusa Tenggara province. Data in 2011-2015 show that malnourished cases were found to be running down, but they increased to 403 cases in 2016. According to (Ramadani et al., 2013), some factors are causing malnourished in baby, such as the percentage of unexclusive baby care; low weight baby (<2500g); unhealth house categorized; households with clean water access; the active integrated center of

the ministry; and utilizing health facility. Another factor are the percentage of incomplete immunization given; the percentage of getting A vitamin; the percentage of household utilizing health facility; the percentage of baby with health service; the percentage of poor villagers; and the percentage of first age marriage < 15 years old (Maulani et al., 2016).

The malnourished percentage is one of the simple parameters to know the baby's nutrition status (WHO, 2010). One way to monitor the malnourished percentage is modeling to check out the relationship between malnourished and affection factors. Modeling the relationship between a dependent variable with one or more independent variables can be presented through statistics into the regression model. A regression analysis model is a statistics method related to systematic relationship pattern between the variables (Daoud, 2017).

According to (Hardle et al., 2004), there are two approaches that can be used to determine the regression curve, namely the parametric regression approach and nonparametric regression approach. Parametric regression has an assumption to be filled up like normal distribution normal and constant variance. In applying parametric regression, a deviation to the assumption often happens like the normal distribution. Therefore, to avoid tight and robust assumptions, the statistics technique is not linked to the tight assumption and certain regression. One of the alternative ways to solve it is the nonparametric regression approach. This approach is used when the first information relates to a curve regression limit or not (Eubank, 1999).

Nonparametric approach methods that are often used are truncated spline and kernel estimator. Generally, growth patterns for babies tend to have changed at certain ages. The pattern has a form that cannot be determined so that when estimated using parametric regression, the results are inaccurate. Therefore, data case related to the percentage of children's babies suffering from malnutrition uses nonparametric regression (Pratiwi, 2017). Initial analysis shows that the pattern formed between the data on the percentage of malnourished children under five with several variables that influence it fluctuates at certain intervals. This characteristic is following the spline approach, which has high flexibility without the subjectivity of the researcher (Eubank, 1999). Besides, other variables that also influence it do not show any particular pattern in the data, so this pattern is compatible with the kernel approach, which can model data without certain patterns. Besides, this approach has also received special attention from researchers because it has a relatively fast convergence speed compared to other approaches (Hardle et al., 2004).

Two basic assumptions need to be considered when the nonparametric regression model is explored. The first assumption is the patterns in each independent multivariable are considered to have the same pattern. The second assumption is that researchers only use one form of the model estimator for each independent variable. In applying it in various cases, data patterns often differ from each of the independent variables. Therefore, if only one estimator is used to estimate the nonparametric regression curve, the estimator generated does not match the data pattern. As a result, the result regression model's estimation is less precise and tends to produce large errors (Budiantara et al., 2015). Based on the description explained, this study was conducted to model the percentage of malnourished children baby in the NTB Province using a nonparametric mixed truncated spline and kernel regression model.

## **B. LITERATURE REVIEW**

### **1. Nonparametric Regression**

Nonparametric regression is one of the approaches used to determine the relationship pattern between dependent and independent variables whose regression curve is unknown, or there is no complete past information about the shape of the data pattern. This approach has high flexibility because it is expected to find its own form of regression curve estimation without being influenced by the researcher's subjectivity factor. The general nonparametric regression model is as equation 1.

$$y_i = f(x_i) + \varepsilon_i \quad , \quad i = 1, 2, \dots, n \tag{1}$$

$y_i$  is the dependent variable,  $x_i$  is the independent variable,  $f(x_i)$  is a regression function of unknown shape, and  $\varepsilon_i$  is an error that is assumed to be random with zero mean and constant variance (Eubank, 1999).

**2. Mixed Estimator Nonparametric Regression Truncated Spline and Kernel**

Consider data  $(t_i, z_i, y_i)$  and the relations between independent variables  $(t_i, z_i)$  and a dependent variable  $(y_i)$  are assumed to follow the nonparametric regression model. In general, the nonparametric regression model is defined as equation 2.

$$y_i = \mu(t_i, z_i) + \varepsilon_i \quad , \quad i = 1, 2, \dots, n \tag{2}$$

The shape of the regression curve  $\mu(t_i, z_i)$  is assumed to be unknown and smooth, meaning continuous and differentiable. A random error  $\varepsilon_i$  has a normal distribution with zero mean and constant variance. The regression curve  $\mu(t_i, z_i)$  is assumed to be additive, meaning it can be written as equation 3.

$$\mu(t_i, z_i) = f(t_i) + g(z_i) \tag{3}$$

The main problem is sounding nonparametric mixed curve regression. It is how to get estimation curve regression form as defined below, with a vector of bandwidth parameters  $\mathbf{h}$  and a vector of knot points  $\mathbf{k}$ .

$$\hat{\mu}_{\mathbf{h}, \mathbf{k}}(t_i, z_i) = \hat{f}_{\mathbf{k}}(t_i) + \hat{g}_{\mathbf{h}}(z_i) \tag{4}$$

Regarding obtain an estimator of mixed spline truncated and kernel regression, regression  $f_{\mathbf{k}}(t_i)$  is approached using the function of the truncated spline with knot points  $\mathbf{k} = (k_1, k_2, \dots, k_k)^T$ , and then, regression curve  $g_{\mathbf{h}}(z_i)$  is approached using kernel's function. For example, given a basis for truncated spline space as follow, with  $I$  being an indicator function.

$$\{1, t, t^2, \dots, t^m, (t - k_1)I(t \geq k_1), (t - k_2)I(t \geq k_2), \dots, (t - k_k)I(t \geq k_k)\} \tag{5}$$

Regression curve  $f_{\mathbf{k}}(t_i)$  can be written as follows, with  $\theta_0, \theta_1, \dots, \theta_m, \phi_1, \phi_2, \dots, \phi_k$  being unknown parameters.

$$f_{\mathbf{k}}(t_i) = \theta_0 + \theta_1 t_i + \dots + \theta_i^m + \phi_1 (t_i - k_1)^m I(t_i \geq k_1) + \dots + \phi_k (t_i - k_k)^m I(t_i \geq k_k) \tag{6}$$

Moreover, the estimation of the kernel's regression curve  $g(z_i)$  can be presented as the following formula.

$$\hat{g}_{\mathbf{h}}(z_i) = n^{-1} \sum_{i=1}^n \left\{ \frac{K_{\mathbf{h}}(z_i - z)}{n^{-1} \sum_{i=1}^n K_{\mathbf{h}}(z_i - z)} \right\} y_i = n^{-1} \sum_{i=1}^n W(z) y_i \tag{7}$$

$W(z)$  and  $K_{\mathbf{h}}(z - z_i)$  can be written as the following equations, respectively, with a kernel function  $K(u)$ .

$$W(z) = \frac{K_{\mathbf{h}}(z_i - z)}{n^{-1} \sum_{i=1}^n K_{\mathbf{h}}(z_i - z)} \quad \text{and} \quad K_{\mathbf{h}}(z_i - z) = \frac{1}{h} K\left(\frac{z_i - z}{h}\right)$$

There are several types of kernel functions, such as the uniform kernel, triangle kernel, quartic kernel, Gaussian kernel, and others (Hardle et al., 2004).

Kernel estimators are divided into several types, including Nadaraya Watson estimator, Priestley-Chao estimator, and Gasser-Muller estimator. The kernel estimator that is often used for random cases is the Nadaraya-Watson kernel estimator. The following Equation shows the multivariate form of the Nadaraya-Watson estimator (Nadaraya, 1964).

$$\hat{g}_h(z_i) = \frac{\frac{1}{n} \sum_{i=1}^n \left\{ \prod_l K_h \left( \frac{z_l - z_{li}}{h_l} \right) \right\} y_i}{\frac{1}{n} \sum_{i=1}^n \left\{ \prod_l K_h \left( \frac{z_l - z_{li}}{h_l} \right) \right\}}, \quad l = 1, 2, \dots, q \tag{8}$$

### 3. Estimation of Mixed Nonparametric Regression Model Truncated Spline and Kernel

According to Ratnasari et al. (2016), to acquire the estimation curve of nonparametric regression using mixed truncated spline and kernel, given some lemmas and a theorem. Given the nonparametric multivariable regression model as in Equation (3). If the function  $f_j(t_{ji})$ ,  $j = 1, 2, \dots, p$  approached using linear truncated spline function with knot point  $\mathbf{k}_j = (k_{j1}, k_{j2}, \dots, k_{jk})^T$ , it could be noted:

$$\sum_{j=1}^p \mathbf{f}_j(t_{ji}) = \mathbf{G}(\mathbf{k})\boldsymbol{\theta} \tag{9}$$

with,

$$\mathbf{f}_j(t_{ji}) = (f_1(t_{1i}), f_2(t_{2i}), \dots, f_p(t_{pi}))^T$$

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \dots, \boldsymbol{\theta}_p^T)^T$$

$$\boldsymbol{\theta}_j = (\theta_{j0}, \theta_{j1}, \phi_{j1}, \phi_{j2}, \dots, \phi_{jk})$$

$$\mathbf{G}(\mathbf{k}) = (\mathbf{G}(k_1), \mathbf{G}(k_2), \dots, \mathbf{G}(k_p))$$

$$\mathbf{G}(k_j) = \begin{pmatrix} t_{j1} & (t_{j1} - k_{j1})_+ & \dots & (t_{j1} - k_{jk})_+ \\ t_{j2} & (t_{j2} - k_{j1})_+ & \dots & (t_{j2} - k_{jk})_+ \\ \vdots & \vdots & \ddots & \vdots \\ t_{jn} & (t_{jn} - k_{j1})_+ & \dots & (t_{jn} - k_{jk})_+ \end{pmatrix}$$

Besides, if the kernel's function  $g_j(z_{ji})$ ,  $j = 1, 2, \dots, q$  approached using kernel Nadaraya-Watson, then based on Equation (8), kernel estimator for multivariable regression curve is given by the following matrix form.

$$\sum_{l=1}^q g_l(z_{li}) = \mathbf{D}(\mathbf{h})\mathbf{y} \tag{10}$$

with vector  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ . Moreover, Equation (9) can be written in the following matrix form.

$$\sum_{j=1}^p \mathbf{f}_j(t_j; \mathbf{k}; \mathbf{h}) = \mathbf{K}(\mathbf{k}, \mathbf{h})\mathbf{y} \tag{11}$$

with  $\mathbf{K}(\mathbf{k}, \mathbf{h}) = \mathbf{G}(\mathbf{k})(\mathbf{G}(\mathbf{k})^T \mathbf{G}(\mathbf{k}))^{-1} \mathbf{G}(\mathbf{k})^T (\mathbf{I} - \mathbf{D}(\mathbf{h}))$ . Therefore, the mixed truncated spline and multivariable kernel estimator as follows.

$$\begin{aligned}\hat{\mu}(t_{1,\dots,t_p}, z_1, \dots, z_q; \mathbf{k}; \mathbf{h}) &= \mathbf{K}(\mathbf{k}, \mathbf{h}) + \mathbf{D}(\mathbf{h}) \\ &= \mathbf{M}(\mathbf{k}, \mathbf{h})\mathbf{y}\end{aligned}\quad (12)$$

#### 4. Choosing Knot Points and Optimal Bandwidths on Mixed Truncated Spline and Kernel

The estimation of nonparametric curve regression mixed truncated spline and kernel depends on knot point and optimal bandwidth parameters based on the previous explanation. According to Budiantara et al. (2015), a method that can be used to choose knot point and optimal parameter bandwidth is Generalized Cross-Validation (GCV). This method is a modified form of Cross-Validation (CV). A case data by Fitriyani & Budiantara (2014) showed that GCV has a better result than CV, taking into account the model's goodness criteria. The following formula gives the GCV function as in the equation 13.

$$GCV(\mathbf{k}, \mathbf{h}) = \frac{n^{-1} \|y_i - \hat{\mu}(t_{1,\dots,t_p}, z_1, \dots, z_q; \mathbf{k}; \mathbf{h})\|^2}{(n^{-1} \text{trace}[\mathbf{I} - \mathbf{K}(\mathbf{k}, \mathbf{h}) - \mathbf{D}(\mathbf{h})])^2}\quad (13)$$

with  $\mathbf{I}$  indicates an identity matrix. Optimal parameter knot point  $\mathbf{k}_{\text{opt}} = (k_{j1(\text{opt})}, k_{j2(\text{opt})}, \dots, k_{jp(\text{opt})})^T$  and optimal parameter bandwidth  $\mathbf{h}_{\text{opt}}$  gained from optimizing as defined equation 14:

$$\mathbf{G}(\mathbf{k}_{\text{opt}}, \mathbf{h}_{\text{opt}}) = \underset{\mathbf{k}, \mathbf{h}}{\text{Min}}\{\mathbf{G}(\mathbf{k}, \mathbf{h})\}\quad (14)$$

#### 5. Goodness Criteria

One of the objectives expected by conducting regression analysis is to obtain the best model that can explain the relationship between the dependent variable and the independent variable based on specific criteria. Criteria used in measuring the goodness of fit of a regression model are to use the coefficient of determination  $R^2$  and Mean Square Error (MSE). The smaller the MSE value obtained, the better the model obtained. Conversely, the model obtained will be better if the value of  $R^2$  obtained is considerable (Cheng et al., 2014).

#### 6. Residual Assumptions and Accuracy of Prediction

The mixed truncated spline and kernel model residuals must fulfill several assumptions, including being identical and independent. The purpose of testing this residual assumption is to provide certainty that the regression equation obtained has estimation accuracy, is unbiased and consistent. Moreover, one of the regression analysis objectives is to predict the average value of the dependent variable based on data from known independent variables. Therefore, the accuracy of predictions is essential to note. Several calculations can be used to calculate prediction errors based on the model, one of which is the Mean Absolute Percentage Error (MAPE) value (Wei, 2006). This measure is calculated by dividing the absolute error in each period by the actual data values. Furthermore, the percentage is calculated on average. This approach is useful when an evaluation measure of the accuracy of the prediction variable is desired and shows how much prediction error is generated (Khair et al., 2017).

### C. RESEARCH METHODS

This research used secondary data, taken from the Public Health Office and Central Bureau of Statistics of West Nusa Tenggara Province year 2016. The variables used in this study consisted of malnutrition status in children under five ( $y$ ) as the dependent variable and nine variables as independent variables, namely the percentage of infants who never received exclusive breastfeeding ( $x_1$ ); the number of houses in healthy categories ( $x_2$ ); the number of active integrated service post (*posyandu*) ( $x_3$ ); the percentage of households using health facilities ( $x_4$ ); the number of babies with low body weight <2500 grams ( $x_5$ ); the number of

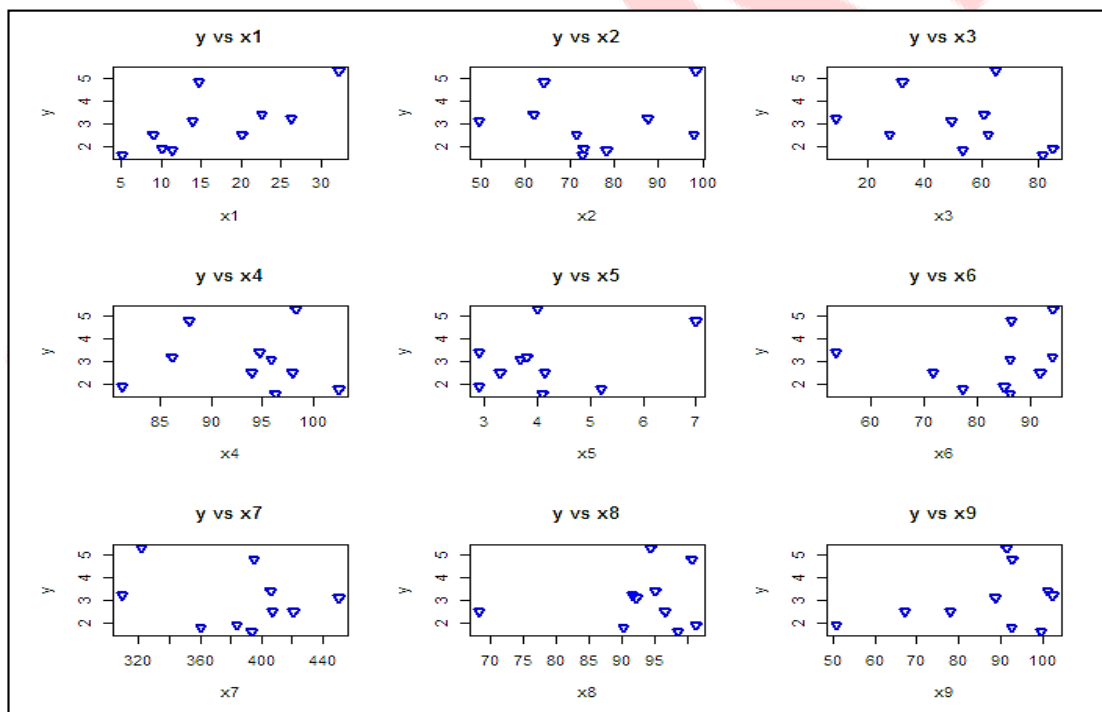


households with access to clean water ( $x_6$ ); the percentage of children under five who received complete immunization ( $x_7$ ); the percentage of infants who received vitamin A ( $x_8$ ); and the percentage of infants who received health services ( $x_9$ ).

The research steps were divided into several stages, including (1) determining variables and collecting data; (2) forming scatter plots of data and determining the appropriate approach based on the scatter plots; (3) modeling the data based on the kernel characteristic and determining the optimal bandwidth based on the minimum GCV; (4) modeling the data using a nonparametric mixed regression model truncated spline and kernel, with various knot points and using bandwidths obtained in step (3), and calculating the GCV value for each of the determined model; (5) determining the optimal knots and bandwidths and modeling the data; (6) determining the criteria for the goodness of the model and testing the residual assumptions; (7) determining predictions and testing the accuracy of predictions based on MAPE value; and (8) concluding.

### D. RESULTS AND DISCUSSION

The choice of truncated spline and kernel nonparametric regression approach for each independent variable can be seen from how the shape of the relationship patterns produced by factors that affect malnourished under-five years children. The scatter plots between the dependent variable ( $y$ ) and each independent variable ( $x_j$ ) are shown in Figure 1.



**Figure 1.** Scatter plots between the dependent variable and each independent variable

It can be examined from Figure 1 that the relationships generated by  $y$  with  $x_1, x_2, x_3,$  and  $x_4$  forming specific patterns so that the approach used for these relationship patterns was the truncated spline nonparametric regression approach. The pattern of relationships generated  $y$  with  $x_5, x_6, x_7, x_8,$  and  $x_9$  the distribution of data produced does not form a particular pattern. The right approach to solve the relationship pattern is the nonparametric kernel regression approach.

According to (Budiantara et al., 2015), the main thing to do in processing the nonparametric kernel regression approach is by determining the optimum bandwidths. If the chosen bandwidth is too small, it will produce an under-smoothing curve regression. Otherwise, if the chosen bandwidth is too big, it will create

an over-smoothing curve regression. Therefore, it is necessary to determine the optimum bandwidth that corresponds to the kernel estimation that is carried out. Based on the data's  $y$  predetermined, namely the Nadaraya-Watson kernel nonparametric statistical approach, the optimum bandwidths were selected to produce the right regression curve. Here are given the optimal bandwidths with the smallest GCV for each independent variable  $x_5, x_6, x_7, x_8$ , and  $x_9$ , then, named  $z_5, z_6, z_7, z_8$ , and  $z_9$ .

**Table 1. Optimal bandwidths**

Independent variables	Optimal bandwidths	
$z_5$	$h_5$	0.5561442
$z_6$	$h_6$	1.0220133
$z_7$	$h_7$	0.716311
$z_8$	$h_8$	1.2240464
$z_9$	$h_9$	1.2900146
<b>Generalized Cross-Validation (GCV)</b>		4.0140741

The following formula represents the kernel estimator related to the mentioned independent variables regarding the optimal bandwidths acquired.

$$\hat{g}_h(z_5, z_6, z_7, z_8, z_9; \mathbf{h}) = \frac{\sum_{i=1}^{10} \left[ K\left(\frac{z_5 - z_{i,5}}{0.5561}\right) \cdot K\left(\frac{z_6 - z_{i,6}}{1.0220}\right) \cdot K\left(\frac{z_7 - z_{i,7}}{0.7163}\right) \cdot K\left(\frac{z_8 - z_{i,8}}{1.2240}\right) \cdot K\left(\frac{z_9 - z_{i,9}}{1.2900}\right) \cdot y_i \right]}{\sum_{i=1}^{10} \left[ K\left(\frac{z_5 - z_{i,5}}{0.5561}\right) \cdot K\left(\frac{z_6 - z_{i,6}}{1.0220}\right) \cdot K\left(\frac{z_7 - z_{i,7}}{0.7163}\right) \cdot K\left(\frac{z_8 - z_{i,8}}{1.2240}\right) \cdot K\left(\frac{z_9 - z_{i,9}}{1.2900}\right) \right]}$$

with  $\mathbf{h} = (h_5, h_6, h_7, h_8, h_9)^T$  given in Table 1.

Moreover, after obtaining a nonparametric kernel regression model, the next step is to make a mixed model associated with the related independent variables  $x_1, x_2, x_3$ , and  $x_4$  with a truncated spline approach with the kernel's effect model produced. The best model selection is made by taking into account the knot points and the obtained bandwidths that produce minimal GCV. In Table 2, given the optimal knot points, related to the optimum bandwidths acquired, for each independent variable  $x_1, x_2, x_3$ , and  $x_4$ , which are then named  $t_1, t_2, t_3$ , and  $t_4$ .

**Table 2. Optimal knot points**

Independent variables	Optimal knot points	
$t_1$	$k_{11}$	10.1385
$t_2$	$k_{21}$	72.1969
$t_3$	$k_{31}$	44.2585
$t_4$	$k_{41}$	91.0146
<b>Generalized Cross-Validation (GCV)</b>		0.003038719

Based on the optimal knot points and bandwidths, the obtaining parameter estimation for mixed model truncated spline and kernel nonparametric regression with one knot point for each related independent variables and optimum bandwidths are listed in Table 3.

**Table 3.** Estimator for model parameters

Independent variables	Parameter	Estimator
Constant	$\theta_0$	-10.89
$t_1$	$\theta_{11}$	0.32
	$\phi_{11}$	-14.64
$t_2$	$\theta_{21}$	0.03
	$\phi_{21}$	0.16
$t_3$	$\theta_{31}$	-0.0003
	$\phi_{31}$	0.03
$t_4$	$\theta_{41}$	0.09
	$\phi_{41}$	-3.76

The estimation results produced, mixed nonparametric regression truncated spline and kernel model given as follows, with kernel function  $K(u) = \frac{15}{16}(1-u^2)^2$ ,  $-1 \leq u \leq 1$  that influence kernel shape used was quartic function (Eubank, 1999).

$$\hat{y} = -10.89 + 0.32t_{1i} + 0.03t_{2i} - 0.0003t_{3i} + 0.09t_{4i} - 14.64(t_{1i} - 10.1385)_+^1 + 0.16(t_{2i} - 72.1969)_+^1 + 0.03(t_{3i} - 44.2585)_+^1 - 3.76(t_{4i} - 91.1046)_+^1 + \frac{\sum_{i=1}^{10} \left[ K\left(\frac{z_5 - z_{i,5}}{0.5561}\right) \cdot K\left(\frac{z_6 - z_{i,6}}{1.0220}\right) \cdot K\left(\frac{z_7 - z_{i,7}}{0.7163}\right) \cdot K\left(\frac{z_8 - z_{i,8}}{1.2240}\right) \cdot K\left(\frac{z_9 - z_{i,9}}{1.2900}\right) \cdot y_i \right]}{\sum_{i=1}^{10} \left[ K\left(\frac{z_5 - z_{i,5}}{0.5561}\right) \cdot K\left(\frac{z_6 - z_{i,6}}{1.0220}\right) \cdot K\left(\frac{z_7 - z_{i,7}}{0.7163}\right) \cdot K\left(\frac{z_8 - z_{i,8}}{1.2240}\right) \cdot K\left(\frac{z_9 - z_{i,9}}{1.2900}\right) \right]}$$

The mixed truncated spline and kernel nonparametric regression model obtained has a 99.998838% value of  $R^2$ . This value means that the model can explain the variability of babies suffering from malnutrition in West Nusa Tenggara Province by 99.998838%, with a Mean Square Error (MSE) value of 0.008353. The  $R^2$  value indicates that the independent variable used has a significant influence on babies suffering from malnutrition in the province of West Nusa Tenggara, and the rest is the influence of other variables. Therefore, it can be concluded that the mixed model attained is excellent and proper to use.

Furthermore, the residual assumption tests are needed to test the feasibility of the model derived. The analysis results show that the residual model has fulfilled the identical and independent assumptions for the residual model to be used for prediction. As one of the regression analysis objectives, Figure 2 below reveals the comparative line charts between the actual data and the prediction value attained, based on data from known independent variables.



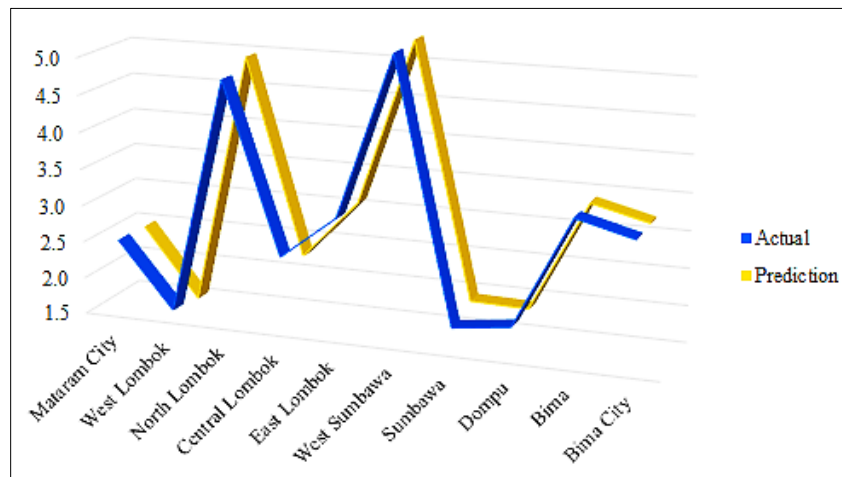


Figure 2. Line charts actual and prediction data

Based on the prediction acquired, MAPE value was then used to measure the accuracy of predictive values mixed spline truncated and kernel model. The results represent that the resulting model is the best and has a high degree of accuracy by considering the MAPE value that did not exceed the 10% margin of error (Khair et al., 2017). Therefore, the model obtained has the good prediction with a MAPE value of 0.306711%.

## E. CONCLUSIONS AND SUGGESTION

Based on the implementation of a mixed truncated spline and kernel nonparametric regression model on the percentage of malnourished children in West Nusa Tenggara Province, a good model was found with an MSE value of 0.008353 and  $R^2$  a value of 99.998838%. In addition, this model produced a value of predictive accuracy, measured by MAPE value of 0.3067109%. The MAPE value indicated a high degree of accuracy so that the model obtained has an excellent forecast. Furthermore, future research is suggested to conduct mixed model spline truncated and kernel with more than one knot point.

## ACKNOWLEDGMENT

This work was supported by the Statistics Areas of Expertise, Universitas Mataram, using PNPB Universitas Mataram year 2019 funding.

## REFERENCES

- BPS, C. B. of S. (2016). *Official Statistics News*. Central Bureau of Statistics, West Nusa Tenggara Province. <https://ntb.bps.go.id/>
- Budiantara, I. N., Ratnasari, V., Ratna, M., & Zain, I. (2015). The Combination of Spline and Kernel Estimator for Nonparametric Regression and its Properties. *Applied Mathematical Sciences*, 9(122), 6083–6094.
- Cheng, C. L., Shalabh, & Garg, G. (2014). Coefficient of Determination for Multiple Measurement Error Models. *Journal of Multivariate Analysis*, 126, 137–152.
- Daoud, J. I. (2017). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*, 949(1), 1–6.
- Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing* (Second). Marcel Dekker, Inc.
- Fitriyani, N., & Budiantara, I. N. (2014). Cross Validation and Generalized Cross Validation Methods in Spline Nonparametric Regression: Case Study of Fertility Data in East Java (Translated). *Seminar Nasional Pendidikan Sains*, 1089–1095.
- Hardle, W., Muller, M., Sperlich, S., & Werwatz, A. (2004). *Nonparametric and Semiparametric Models: An Introduction*. Humboldt-Universität zu Press.
- Khair, U., Fahmi, H., Hakim, S. Al, & Rahim, R. (2017). Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error. *Journal of Physics: Conference Series*, 930(1).
- Maulani, A., Herrhyanto, N., & Suherman, M. (2016). Application of Geographically Weighted Regression (GWR) Model to Determine Factors Affecting Malnutrition Cases in Children under Five in West Java (Translated). *Jurnal EurekaMatika*, 4(1), 46–63.
- Nadaraya, E. A. (1964). On Estimating Regression (Translated by Bernard Seckler). *Theory of Probability and Application*, 9, 141–142.
- Pratiwi, L. P. S. (2017). Spline Truncated Modeling in Birespon Nonparametric Regression (Translated). *Konferens*

*i Nasional Sistem & Informatika*, 441–445.

Ramadani, I. R., Rahmawati, R., & Hoyyi, A. (2013). Analysis of Factors Affecting Malnutrition of Toddlers in Central Java using the Spatial Durbin Model Method (Translated). *Jurnal Gaussian*, 2(4), 333–342.

Ratnasari, V., Budiantara, I. N., Ratna, M., & Zain, I. (2016). Estimation of Nonparametric Regression Curve using Mixed Estimator of Multivariable Truncated Spline and Multivariable Kernel. *Global Journal of Pure and Applied Mathematics*, 12(6), 5047–5057.

Wei, W. W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Education, Inc.

WHO, W. H. O. (2010). *Nutrition Landscape Information System (NLIS) Country Profile Indicators: Interpretation Guide*. <https://www.who.int/nutrition>

VARIAN