Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2022, 2022:76
https://doi.org/10.28919/cmbn/7545
ISSN: 2052-2541

COMPARISON OF SIMULTANEOUSLY NONPARAMETRIC REGRESSION BASED ON SPLINE AND FOURIER SERIES ESTIMATOR RELATED SOCIAL AID DISTRIBUTION IN INDONESIA

M. FARIZ FADILLAH MARDIANTO*, SULIYANTO, ELLY PUSPORANI, ANTONIO NIKOLAS MANUEL BONAR SIMAMORA, NETHA ALIFFIA, AYUNING DWIS CAHYASARI, CHAEROBBY FAKHRI FAUZAAN PURWOKO

Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya 60115, Indonesia

Abstract: Several estimators in the nonparametric regression approach that are often used are the Spline estimator and the Fourier series. Both estimators have high flexibility and can adapt to the local nature of the data effectively. This study applies these two methods to compare the best performance of the formed model. The data used in this study are multi-response data in the form of types of government assistance and multi-predictor data in the form of factors that are thought to affect the distribution of aid receipts in the community. This research is important to optimize the distribution of welfare assistance in Indonesia, especially during the Coronavirus Disease-19 (Covid-19) pandemic. By using the Generalized Cross-Validation (GCV) criteria for the parsimony model, the estimator chosen to predict the distribution of welfare assistance in Indonesia is the Fourier series estimator with cosine and sine bases. The goodness of the nonparametric regression model with the Fourier series estimator can be seen from the minimum GCV value generated with one oscillation parameter, which is 1.483961, with a Mean Square Error (MSE) of

*Corresponding author
E-mail address: m.fariz.fadillah.m@fst.unair.ac.id
Received June 9, 2022

2.1257374, and R^2 of 0.923143. The prediction results are expected to be the government's recommendation for the restoration of community welfare through equitable distribution of aid to support the achievement of the Sustainable Development Goals (SDGs).

Keywords: poverty alleviation; well-being; nonparametric regression; Spline estimator; Fourier series estimator

2010 AMS Subject Classification: 62G08, 43A32, 41A15.

1. INTRODUCTION

Regression analysis is one of the statistical methods used to analyze the pattern of the relationship between the dependent and independent variables. However, if the pattern of the relationship between the dependent and independent variables is unknown, nonparametric regression can be used. Nonparametric regression is often chosen because it has flexibility in estimating the relationship between response variables and predictors whose function is not known and violates several assumptions in linear regression [1]. In addition, the flexibility in question is the nature of the estimator which can adapt to the local nature of the data effectively [2]. Some estimators that are often used in nonparametric regression include Spline and Fourier series. Spline is a segmented polynomial that has flexibility in modeling random data patterns. In addition, the Spline estimator can be used to overcome data patterns that show a sharp rise or fall with the help of knot points which produce a relatively smooth curve [3]. Besides of Spline estimator, the Fourier series is also a nonparametric regression estimator that is widely used because it has flexibility in modeling data patterns whose oscillation form is unknown [4].

Nonparametric regression models can be distinguished based on the number of dependent and independent variables that make up the model. If there are two or more dependent variables observed from several independent variables, the model includes multi-response regression [5]. Meanwhile, if there is one response variable with more than one predictor variable, it is called multiple predictor regression [6]. Thus, if a nonparametric regression model has more than one response variable and a predictor variable, then the model is called multi-response and multi-predictor nonparametric regression. This is interesting to study, especially in the era of developing data structures such as big data today.

This research utilizes a nonparametric regression analysis method with a Spline estimator and Fourier series for multi-response and multi-predictor data. The aim is to model the percentage of government aid recipients during the Coronavirus Disease-19 (Covid-19) pandemic. Covid-19, which has infected people around the world, has caused many problems, one of which is in the socio-economic field [7]. In addition, the pandemic poses a new challenge for Indonesia, which seeks to realize the Sustainable Development Goals (SDGs), especially in the context of overcoming poverty, and hunger, and improving the quality of health and public welfare. In Indonesia itself, this pandemic has caused an increase in poverty and hunger, as well as serious health and welfare-related problems. This situation could also affect the food security in Indonesia as a country that is currently developing a program to improve food security [8]. The number of poor people in September 2021 was recorded at 26.5 million people [9]. In addition, the pandemic has disrupted Indonesia's targets to minimize poverty and national hunger levels, as well as improve the quality of public health and welfare.

Nonparametric regression models can be distinguished based on the number of dependent and independent variables that make up the model. If there are two or more dependent variables observed from several independent variables, the model is called multi-response nonparametric regression [5]. Meanwhile, if there is one response variable with more than one predictor variable, it is called multi-predictor nonparametric regression [6]. Thus, if a nonparametric regression model has more than one response and predictor variables, then the model is called multi-response and multi-predictor nonparametric regression. This is interesting to study, especially in the era of developing data structures such as big data.

This research utilizes a nonparametric regression analysis method with a Spline estimator and Fourier series for multi-response and multi-predictor data. The aim is to model the percentage of government aid recipients during the Coronavirus Disease-19 (Covid-19) pandemic. Covid-19, which has infected people around the world, has caused many problems, one of which is in the socio-economic field [7]. In addition, the pandemic poses a new challenge for Indonesia, which

seeks to realize the Sustainable Development Goals (SDGs), especially in the context of overcoming poverty, and hunger, and improving the quality of health and public welfare. In Indonesia itself, this pandemic has caused an increase in poverty and hunger, as well as serious health and welfare-related problems. The number of poor people in September 2021 was recorded at 26.5 million people [9]. In addition, the pandemic has disrupted Indonesia's targets to minimize poverty and national hunger levels, as well as improve the quality of public health and welfare.

Therefore, it is necessary to find a way to overcome these problems, then Indonesia's SDGs targets can be achieved. The Indonesian government has established several aid schemes to reduce poverty and hunger levels, including the Social Security Administrator for Health (BPJS) for Contribution Assistance Recipients (PBI), Rice for Prosperous Families (rastra), and Beneficiary Families (KPM). BPJS PBI is part of the BPJS program which is specifically intended for Indonesian people in need where this program is made, thus all people have the same right to receive good medical treatment and also proper treatment [10]. Meanwhile, rastra is social assistance distributed by the government in the form of rice and distributed every month. With the reduction of rastra to the poor, the government hopes that social assistance in the form of rice of at least 10 kilograms per KPM each can reduce the burden of spending on KPM by fulfilling some special needs in the food sector within a certain period without being charged a ransom [11].

Government assistance during the Covid-19 pandemic has an important role in efforts to restore the socio-economic life of the community. However, in general, the distribution of welfare assistance in Indonesia tends to be uneven. Thus, it is necessary to model the percentage of poor families receiving welfare assistance in Indonesia based on the factors that influence poverty as an alternative solution to dealing with the uneven distribution of aid. The distribution of Indonesia's welfare assistance to the community is determined based on poor or vulnerable families [12].

The purpose of this study is to model the percentage of government aid recipients based on a nonparametric regression approach by comparing the performance of the Fourier and Spline series estimators. Previous research related to this method was conducted by Mariati et. al. [13] who compared the nonparametric regression analysis method with the Spline estimator and Fourier series for data with one response variable. Therefore, the novelty of this study is to compare the nonparametric regression analysis method with the Spline estimator and Fourier series for multi-response data in the form of types of government assistance and multi-predictor data in the form of factors that are thought to affect the distribution of aid receipts in the community. This model can be a recommendation and evaluation for the government to restore the socio-economic life of the community evenly for all Indonesian people to achieve targets to overcome poverty, and hunger, and improve the quality of health and welfare of the community.

2. Preliminaries

A. Regression Analysis

Predictions can be analyzed using regression. Regression analysis can also be used to analyze the relationship between predictor variables and response variables [14]. The regression model can be said to be good if it fulfills several assumptions including multicollinearity, homogeneity, linearity, and autocorrelation [15]. There are three approaches to estimating the regression curve, including the parametric regression approach, nonparametric regression, and semiparametric regression.

B. Nonparametric Regression Analysis

Nonparametric regression is used to determine the pattern of the relationship between the predictor variable and the response variable whose function is not known without having to fulfill the assumptions in modeling as is the case with parametric regression [16]. The nonparametric regression model can be expressed as follows:

$$y_i = \eta(x_i) + \varepsilon_i \; ; \; \varepsilon_i \sim N(0, \sigma^2) \tag{1}$$

$\eta(x_i)$ is a function of unknown form with the predictor of x_i and ε_i are the residual of the i-th observation which is assumed to be identical, independent, and normally distributed with mean 0 and constant variance. To estimate nonparametric regression, there are several techniques, such as Fourier and Spline series estimators [17].

C. Multi-response Multi-predictor Nonparametric Regression Analysis

The multi-response multi-predictor nonparametric regression model is expressed as follows:

$$\begin{aligned} y_{i1} &= \sum_{l=1}^{L} f_l(x_{i1l}) + \varepsilon_{i1} \\ y_{i2} &= \sum_{l=1}^{L} f_l(x_{i2l}) + \varepsilon_{i2} \\ &\vdots \\ y_{ij} &= \sum_{l=1}^{L} f_l(x_{ijl}) + \varepsilon_{ij} \end{aligned} \tag{2}$$

Where y_{ij} is the response of j-th variable on i-th data, $f_l(x_{ijl})$ is a regression function of unknown form, while x_{ijl} is l-th predictor variable, for j-th response and i-th observation, and ε_{ij} is residual. Before conducting the multi-response analysis stage, the thing that needs to be considered is the correlation between the response variables [18].

D. Fourier Series Estimator in Multi-response Nonparametric Regression

The process of getting the Fourier series estimator using the sine and cosine bases is carried out according to the number of predictors and responses in the data. The optimization method in this study is the Weighted Least Square (WLS) method. This method is not tied to the shape of the distribution function of the error, and there is no penalty. Besides, this method is easy to be used. This method only involves weighting, where the weighting that often used is based on variance [19]. Consider paired data$(x_{i1}, x_{i2}, \ldots, x_{il}, y_{ij})$. The l-th predictor variable is denoted by x_l. The first, second, and j-th response variables are represented by y_1, y_2, and y_j, respectively, with i indicating the number of observations. The multi-response multi-predictor nonparametric regression model in Equation (2) can be expressed as follows:

$$\begin{aligned} y_{i1} &= m_1 + \varepsilon_{i1}; \varepsilon_{i1} \sim IIDN(0, \sigma_1^2) \\ y_{i2} &= m_2 + \varepsilon_{i2}; \varepsilon_{i2} \sim IIDN(0, \sigma_2^2) \\ &\vdots \\ y_{ij} &= m_j + \varepsilon_{ij}; \varepsilon_{ij} \sim IIDN(0, \sigma_j^2) \end{aligned} \tag{3}$$

with $m_j = \sum_{l=1}^{L} f_l(x_{ijl})$; $l = 1,2,\dots,L$ shows many predictor variables, $i = 1,2,\dots,n$ show a lot of observations, $j = 1,2,\dots,q$ is the number of response variables. Equation (3) can be expressed in the form of a vector equation as follows:

$$\boldsymbol{y} = \boldsymbol{m} + \boldsymbol{\varepsilon}; \boldsymbol{\varepsilon} \sim IIDN(\boldsymbol{0}, \boldsymbol{\Sigma}) \quad (4)$$

with

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \boldsymbol{y}_3 \\ \vdots \\ \boldsymbol{y}_j \end{bmatrix}, \boldsymbol{m} = \begin{bmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \\ \boldsymbol{m}_3 \\ \vdots \\ \boldsymbol{m}_j \end{bmatrix}, \text{ dan } \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \boldsymbol{\varepsilon}_3 \\ \vdots \\ \boldsymbol{\varepsilon}_j \end{bmatrix}$$

$\boldsymbol{\varepsilon}$ is a vector containing identical random errors, independent and normally distributed multivariate with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma}$. In this case

$$\boldsymbol{y_j} = \begin{bmatrix} y_{1j} \\ y_{2j} \\ y_{3j} \\ \vdots \\ y_{nj} \end{bmatrix}, \boldsymbol{m_j} = \begin{bmatrix} m_{1j} \\ m_{2j} \\ m_{3j} \\ \vdots \\ m_{nj} \end{bmatrix} \text{dan } \boldsymbol{\varepsilon_j} = \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \varepsilon_{3j} \\ \vdots \\ \varepsilon_{nj} \end{bmatrix}$$

$j = 1,2,\dots,q$. In this study, the regression curve is approximated by the Fourier series function as follows:

$$m_{nj} = \sum_{l=1}^{L} \left(\frac{\alpha_{0jl}}{2} + \gamma_{jl} x_{ijl} + \sum_{k=1}^{K} \left(\alpha_{jkl} \cos k\, x_{ijl} + \beta_{jkl} \sin k\, x_{ijl} \right) \right) \quad (5)$$

with K is the oscillation parameter. If in Equation (5) eliminates the cosine component, the Fourier sine series is used. On the other hand, if Equation (5) eliminates the sine component, the Fourier cosine series is used.

For the optimization process, Equation (5) can be constructed as

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \quad (6)$$

where

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_2 & \dots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{X}_j \end{bmatrix}$$

with

$$\boldsymbol{X_j} = \begin{bmatrix} 1 & x_{1j1} & cos\, x_{1j1} & \dots & cos\, K\, x_{1j1} & sin\, x_{1j1} & \dots & sin\, K\, x_{1j1} & \cdots & \sin K\, x_{1jl} \\ 1 & x_{2j1} & cos\, x_{2j1} & \dots & cos\, K\, x_{2j1} & sin\, x_{2j1} & \dots & sin\, K\, x_{2j1} & \cdots & \sin K\, x_{2jl} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{nj1} & cos\, x_{nj1} & \dots & cos\, K\, x_{nj1} & sin\, x_{nj1} & \dots & sin\, K\, x_{nj1} & \cdots & \sin K\, x_{njl} \end{bmatrix}$$

is a predictor-related matrix that depends on the number of oscillation parameters. For Fourier sine series and Fourier cosine series adjust.

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \\ \vdots \\ \boldsymbol{\beta_j} \end{bmatrix}$$

with

$$\boldsymbol{\beta_j} = \begin{bmatrix} \alpha^*_{0j1} \\ \gamma_{j1} \\ \alpha_{1j1} \\ \vdots \\ \alpha_{Kj1} \\ \beta_{1j1} \\ \vdots \\ \beta_{Kj1} \\ \vdots \\ \beta_{Kjl} \end{bmatrix}$$

in this case $\alpha^*_{0jl} = \frac{1}{2}\alpha_{0jl}$. If Equation (5) is substituted into Equation (3), we get

$$\begin{aligned} y_{i1} &= \sum_{l=1}^{L} \left(\frac{\alpha_{01l}}{2} + \gamma_{1l} x_{i1l} + \sum\nolimits_{k=1}^{K} (\alpha_{1kl}\, cos\, k\, x_{i1l} + \beta_{1kl}\, sin\, k\, x_{i1l}) \right) + \varepsilon_{i1} \\ y_{i2} &= \sum_{l=1}^{L} \left(\frac{\alpha_{02l}}{2} + \gamma_{2l} x_{i2l} + \sum\nolimits_{k=1}^{K} (\alpha_{2kl}\, cos\, k\, x_{i2l} + \beta_{2kl}\, sin\, k\, x_{i2l}) \right) + \varepsilon_{i2} \\ &\vdots \\ y_{ij} &= \sum_{l=1}^{L} \left(\frac{\alpha_{0jl}}{2} + \gamma_{jl} x_{ijl} + \sum\nolimits_{k=1}^{K} (\alpha_{jkl}\, cos\, k\, x_{ijl} + \beta_{jkl}\, sin\, k\, x_{ijl}) \right) + \varepsilon_{ij} \end{aligned} \tag{7}$$

with $i = 1,2, \dots, n,$ and $k = 1,2, \dots, K.$ Random errors are correlated with each other. The weighting form is based on the covariance variance matrix of the error which can be written $\boldsymbol{W} = \boldsymbol{V^{-1}}$ with

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{\Sigma_1} & \boldsymbol{\Sigma_{12}} & \cdots & \boldsymbol{\Sigma_{1j}} \\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_2} & \cdots & \boldsymbol{\Sigma_{2j}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma_{j1}} & \boldsymbol{\Sigma_{j2}} & \cdots & \boldsymbol{\Sigma_j} \end{bmatrix}$$

with

$$\boldsymbol{\Sigma_j} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_j^2 \end{bmatrix}$$

and

$$\boldsymbol{\Sigma_{ij}} = \begin{bmatrix} \sigma_{ij} & \sigma_{ij} & \cdots & \sigma_{ij} \\ \sigma_{ij} & \sigma_{ij} & \cdots & \sigma_{ij} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{ij} & \sigma_{ij} & \cdots & \sigma_{ij} \end{bmatrix}; i \neq j$$

σ_j^2 is the j-th response error variance and σ_{ij} is the j-th response error covariance with other responses. The $\boldsymbol{W}$ matrix is used in the optimization of Weighted Least Square (WLS) to get the $\boldsymbol{\beta}$ parameter estimator. WLS optimization is done by minimizing the goodness of fit of the nonparametric multi-response multi predictor regression model with the Fourier series approach.

$$\min_{\boldsymbol{m}\in C(0,\pi)}[R(\boldsymbol{m})] = \min_{\boldsymbol{m}\in C(0,\pi)}\{(\boldsymbol{y}-\boldsymbol{m})^T\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{m})\}$$

$\boldsymbol{m}$ loading $\boldsymbol{\beta}$ parameters, thus

$$\min_{\boldsymbol{\beta}\in \boldsymbol{R}^{j(l+1)}}[R(\boldsymbol{\beta})] = \min_{\boldsymbol{\beta}\in \boldsymbol{R}^{j(l+1)}}\{(\boldsymbol{y}-\boldsymbol{X\beta})^T\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{X\beta})\} \tag{8}$$

By elaborating on the optimization of Equation (8), we get

$$\begin{aligned} R(\boldsymbol{\beta}) &= (\boldsymbol{y}-\boldsymbol{X\beta})^T\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{X\beta}) \\ &= (\boldsymbol{y}^T-\boldsymbol{\beta}^T\boldsymbol{X}^T)\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{X\beta}) \\ &= (\boldsymbol{y}^T\boldsymbol{W}-\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{W})(\boldsymbol{y}-\boldsymbol{X\beta}) \\ &= \boldsymbol{y}^T\boldsymbol{W}\boldsymbol{y}-\boldsymbol{y}^T\boldsymbol{W}\boldsymbol{X\beta}-\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}+\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X\beta} \end{aligned}$$

Getting an estimator of the $\boldsymbol{\beta}$ parameter, is done by performing the derivative $R(\boldsymbol{\beta})$ on $\boldsymbol{\beta}$.

$$\frac{\partial R(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \boldsymbol{0}$$

$$\frac{\partial\{(\boldsymbol{y}-\boldsymbol{X\beta})^T\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{X\beta})\}}{\partial \boldsymbol{\beta}} = \boldsymbol{0}$$

$$-2\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}+2\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{0}$$

$$-\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{0}$$

$$\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}$$

An estimator for the parameters of the multi-response multi-predictor nonparametric regression model was obtained using the Fourier series approach.

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} \tag{9}$$

In this case the estimator for the regression curve is

$$\begin{aligned}
\hat{y}_{i1} &= \sum_{l=1}^{L}\left(\frac{\hat{\alpha}_{01l}}{2} + \hat{\gamma}_{1l}x_{i1l} + \sum\nolimits_{k=1}^{K}\left(\hat{\alpha}_{1kl}\cos k\,x_{i1l} + \hat{\beta}_{1kl}\sin k\,x_{i1l}\right)\right)\\
\hat{y}_{i2} &= \sum_{l=1}^{L}\left(\frac{\hat{\alpha}_{02l}}{2} + \hat{\gamma}_{2l}x_{i2l} + \sum\nolimits_{k=1}^{K}\left(\hat{\alpha}_{2kl}\cos k\,x_{i2l} + \hat{\beta}_{2kl}\sin k\,x_{i2l}\right)\right)\\
&\ \vdots\\
\hat{y}_{ij} &= \sum_{l=1}^{L}\left(\frac{\hat{\alpha}_{0jl}}{2} + \hat{\gamma}_{jl}x_{ijl} + \sum\nolimits_{k=1}^{K}\left(\hat{\alpha}_{jkl}\cos k\,x_{ijl} + \hat{\beta}_{jkl}\sin k\,x_{ijl}\right)\right)
\end{aligned} \tag{10}$$

If in the form of vector, then

$$\begin{aligned}
\widehat{\boldsymbol{y}} &= \widehat{\boldsymbol{m}}\\
&= \boldsymbol{X}\widehat{\boldsymbol{\beta}}\\
&= \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}\\
&= \boldsymbol{A}(\boldsymbol{K})\boldsymbol{y}
\end{aligned}$$

With the hat matrix defined by

$$\boldsymbol{A}(\boldsymbol{K}) = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W} \tag{11}$$

Based on the results obtained, it can be seen that this estimator depends on the oscillation parameters. Selection of optimal oscillation parameters using the Generalized Cross Validation (GCV) method:

$$\boldsymbol{GCV}(\boldsymbol{K}) = \frac{\boldsymbol{MSE}(\boldsymbol{K})}{\left(n^{-1}trace(\boldsymbol{I}-\boldsymbol{A}[\boldsymbol{K}])\right)^2} \tag{12}$$

$$\boldsymbol{MSE}[\boldsymbol{K}] = n^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})$$

E. Nonparametric Regression with Spline Estimator Approach

Regression curve $f_l(x_{ijl})$ approximated by linear truncated Spline function and knot points of $h_1, h_2, \dots, h_U$:

$$f_l(x_{ijl}) = \alpha_{ijl}x_{ijl} + \sum_{u=1}^{U} \beta_{iju}(x_{ijl} - H_{iU})_+^1 \quad (13)$$

The truncated function of $(x_{ijl} - H_{iU})_+^1$ on Equation (13) is defined by

$$(x_{ijl} - H_{iU})_+^1 = \begin{cases} (x_{ijl} - H_{iU})_+^1, x_{ijl} \geq H_{iU} \\ 0 \qquad , x_{ijl} < H_{iU} \end{cases} \quad (14)$$

As a result, the nonparametric regression truncated Spline multi-response multi predictor is obtained which can be presented in Equation (15) as follows:

$$y_{ij} = \sum_{l=1}^{L} \left\{ \alpha_{ijl}x_{ijl} + \sum_{u=1}^{U} \beta_{iju}(x_{ijl} - H_{iU})_+^1 \right\} + \varepsilon_{ij} \quad (15)$$

This multi-response multi-predictor regression model contains j responses with as many as n observations and can be described as follows

$$\begin{aligned} y_{i1} &= \sum_{l=1}^{L} \{\alpha_{i1l}x_{i1l} + \sum_{u=1}^{U} \beta_{i1u}(x_{i1l} - H_{iU})_+^1\} + \varepsilon_{i1} \\ y_{i2} &= \sum_{l=1}^{L} \{\alpha_{i2l}x_{i2l} + \sum_{u=1}^{U} \beta_{i2u}(x_{i2l} - H_{iU})_+^1\} + \varepsilon_{i2} \\ &\vdots \\ y_{ij} &= \sum_{l=1}^{L} \left\{ \alpha_{ijl}x_{ijl} + \sum_{u=1}^{U} \beta_{iju}(x_{ijl} - H_{iU})_+^1 \right\} + \varepsilon_{ij} \end{aligned} \quad (16)$$

The multi-predictor multi-response regression model in Equation (16) can be presented in the form of a matrix such as Equation (6), the difference is the matrix structure of $\boldsymbol{X}$ and $\boldsymbol{\beta}$ vector, where

$$\boldsymbol{X_j} = \begin{bmatrix} x_{1j1} & (x_{1j1} - H_{11})_+^1 & \dots & (x_{1j1} - H_{1U})_+^1 & \dots & x_{1jl} & \cdots & (x_{1jl} - H_{1U})_+^1 \\ x_{2j1} & (x_{2j1} - H_{21})_+^1 & \dots & (x_{2j1} - H_{2U})_+^1 & \dots & x_{2jl} & \cdots & (x_{2jl} - H_{2U})_+^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{nj1} & (x_{nj1} - H_{n1})_+^1 & \dots & (x_{nj1} - H_{nU})_+^1 & \dots & x_{njl} & \cdots & (x_{njl} - H_{nU})_+^1 \end{bmatrix}$$

and

$$\boldsymbol{\beta_j} = \begin{bmatrix} \alpha_{ij1} \\ \beta_{ij1} \\ \vdots \\ \beta_{ijU} \\ \vdots \\ \alpha_{ijl} \\ \vdots \\ \beta_{ijU} \end{bmatrix}$$

The results of the multi-response multi-predictor Spline estimator in the form of a matrix equation are the same as the results of the multi-response multi-predictor Fourier series estimator with a linear truncated Spline estimation curve as follows:

$$\hat{y}_{i1} = \sum_{l=1}^{L}\left\{\hat{\alpha}_{i1l}x_{i1l} + \sum_{u=1}^{U}\hat{\beta}_{i1u}(x_{i1l} - H_{iU})_{+}^{1}\right\} + \varepsilon_{i1}$$

$$\hat{y}_{i2} = \sum_{l=1}^{L}\left\{\hat{\alpha}_{i2l}x_{i2l} + \sum_{u=1}^{U}\hat{\beta}_{i2u}(x_{i2l} - H_{iU})_{+}^{1}\right\} + \varepsilon_{i2} \quad (17)$$

$$\vdots$$

$$\hat{y}_{ij} = \sum_{l=1}^{L}\left\{\hat{\alpha}_{ijl}x_{ijl} + \sum_{u=1}^{U}\hat{\beta}_{iju}\left(x_{ijl} - H_{iU}\right)_{+}^{1}\right\} + \varepsilon_{ij}$$

Based on the results obtained, it can be seen that this estimator depends on the knot point. Selection of optimal knot points using the Generalized Cross Validation (GCV) method:

$$\boldsymbol{GCV}(\boldsymbol{H}) = \frac{\boldsymbol{MSE}(\boldsymbol{H})}{\left(n^{-1}trace(\boldsymbol{I}-\boldsymbol{A}[\boldsymbol{H}])\right)^2} \quad (18)$$

$$\boldsymbol{MSE}[\boldsymbol{H}] = n^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})^T\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{X\beta})$$

with

$$\boldsymbol{A}[\boldsymbol{H}] = \boldsymbol{X}(\boldsymbol{X^T}\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}$$

F. Size of Model's Goodness

In determining the size of goodness can be known by the value of the coefficient of determination or R^2. A small value of R^2 means that the ability of the independent variables (free) in explaining the variation of the dependent variable is very limited. The coefficient of determination (R^2) describes the accuracy of the regression curve to determine the variation of the response variable which can be explained by several predictor variables [17]. The coefficient of determination can be calculated by the following formula:

$$R^2 = \frac{(\hat{\boldsymbol{y}}-\overline{\boldsymbol{y}})^T(\hat{\boldsymbol{y}}-\overline{\boldsymbol{y}})}{(\boldsymbol{y}-\overline{\boldsymbol{y}})^T(\boldsymbol{y}-\overline{\boldsymbol{y}})} ; 0 \leq R^2 \leq 1 \quad (19)$$

where $\hat{\boldsymbol{y}}$ is a vector containing the estimation results for the response, $\boldsymbol{y}$ is a vector containing the response data, and $\overline{\boldsymbol{y}}$ is a vector containing the average for each response. In addition to R^2, the goodness of the model is also measured by the Mean Square Error (MSE) value. MSE is the estimated value of the error variance, so a good predictive model is indicated by a large R^2 value and a small MSE.

G. Social Assistance in Indonesia

The Indonesian government has established several aid schemes to reduce poverty and hunger levels, including the BPJS for Contribution Assistance Recipients (PBI), Rice for Prosperous Families (Rastra), and Beneficiary Families (KPM). BPJS PBI is part of the BPJS program which is specifically intended for Indonesian people in need where this program is made, thus all people have the same right to receive good medical treatment and also proper treatment [20]. Meanwhile, rastra is social assistance distributed by the government in the form of rice and distributed every month. With the reduction of rastra to the poor, the government hopes that social assistance in the form of rice of at least 10 kilograms per KPM each can reduce the burden of spending on KPM by fulfilling some special needs in the food sector within a certain period without being charged a ransom fee.

3. METHODOLOGY

A. Data and Research Variable

The data used in this study is secondary data consisting of the dependent variable multi-response and independent multi predictor. The dependent variable used is the types of Indonesian government assistance, while the independent variable is the factors that are thought to influence the government's decision to distribute aid. The research variables include data collected from 34 provinces in Indonesia. The data are sourced from Indonesia in Figures 2019 [10] and Indonesia in Figures 2020 [21] which are in sample data and out sample data, respectively. The research variables are shown in Table 1 below.

Variable Type	Variable Name	Scale
Independent	Families with the floor of the house in the form of dirt	Percent
	Families that do not have electricity	Percent
	Number of people with health complaints	Percent
	No medical expenses	Percent
	No transportation costs	Percent
	Theft	Percent
	Unemployment rate	Percent
Dependent	BPJS PBI	Percent
	Rastra	Percent
	KPM	Number of families

TABLE 1. Research Variable

B. Analysis Procedure

The analytical steps of this research include:

1. Making a data plot between predictor variables and response variables to determine whether the relationship between the two variables is linear or not.
2. Analyzing with descriptive statistics on each response variable and predictor variable.
3. Testing the correlation between the response variables using the Bartlett Sphericity test.
4. Analyzing with nonparametric regression with the Fourier estimator approach.
    a. Determining the optimal trigonometric function for the Fourier series based on the minimum GCV value.
    b. Determining the optimal oscillation parameter (k) that results in the minimum GCV of the selected Fourier series trigonometric function. The parsimony principle can be applied in the selection of these oscillation parameters.
    c. Determining the estimator based on the optimal k value.
    d. Calculating R^2 and MSE based on the optimal k to measure the criteria for the goodness of the model.
5. Analyzing with nonparametric regression with the Spline estimator approach.
    a. Selecting the optimal knot point (h) using the GCV method. The parsimony principle can be applied in the selection of this knot point.

b. Forming a nonparametric Spline truncated regression model based on the selected h.

c. Looking for regression model estimation.

d. Calculating R^2 and MSE based on the optimal h to measure the criteria for the goodness of the model.

6. Comparing the performance of the Fourier series and the selected Spline to get the best model.
7. Making conclusions and recommendations based on the best model.

4. MAIN RESULTS

A. Descriptive Statistics

In determining the distribution of welfare assistance in Indonesia, a descriptive statistical analysis was carried out. Based on the results of descriptive statistical analysis, Table 2 presents a summary of the research variables in the form of the average value, maximum value, and minimum value, each of which is taken by one province from each research variable.

Variables	**Averages**	**Maximum Score**		**Minimum Score**	
		Score	**Province**	**Score**	**Province**
BPJS BPI (Y_1)	34.69	77.77	Aceh	18.55	Riau Islands
Rastra (Y_2)	51.89	77.81	Aceh	0.44	Jakarta
KPM (Y_3)	441.774	2.846.829	East Java	21.013	North Kalimantan
Family with a ground floor (X_1)	3.447	22.13	Papua	0.11	Bangka Belitung
Families that do not have electricity (X_2)	2.511	27.63	Papua	0	Jakarta
Number of people with health complaints (X_3)	29.63	44.39	West Nusa Tenggara	15.2	Papua
No medical expenses (X_4)	1.519	3.69	Jambi	0.02	North Kalimantan
No transportation costs (X_5)	0.4444	1.78	Papua	0.03	Bangka Belitung
No facilities of transportation (X_6)	0.1832	1.27	Papua	0	Riau Islands
Theft (X_7)	2.645	10.86	Jakarta	0.4	Aceh
Percentage of the unemployment rate (X_8)	4.712	8.11	Banten	1.57	Bali

TABLE 2. Descriptive Statistics of the Distribution of Welfare Assistance in Indonesia

Based on Table 2, Papua Province is in the highest order of the four research variables with the percentage of families having the largest building floor in the form of land, not having electricity, no transportation costs, and no transportation facilities in 2019 compared to other provinces in Indonesia. This is in contrast to Aceh Province which has the highest percentage of BPJS PBI and Rastra recipients, and Jakarta Province which has the lowest percentage of not having electricity compared to other provinces in Indonesia. This means that the priority in the distribution of rastra assistance to the province of Papua will be higher than that of Aceh and Jakarta. Therefore, predictions are needed to determine the distribution of welfare assistance in Indonesia in the coming period, then it can be used in policy evaluation by the government.

Furthermore, because the percentages of research variables related to response and predictor are modeled simultaneously, it must be statistically confirmed that per category of research variables are correlated together or there is a dependent relationship between the percentages of research variables with one another. The test is carried out using the Bartlett Sphericity test involving the correlation matrix ($\boldsymbol{R}$) as follows.

	y_1	y_2	y_3
y_1	1	0.359	-0.049
y_2	0.359	1	0.329
y_3	-0.049	0.329	1

TABLE 3. Y Dependent Variable Correlation Matrix Structure

Table 3 shows that not all relationships between variables have a high correlation, so it is necessary to confirm the relationship with the Bartlett Sphericity test with the following hypothesis.

H_0 : Between variables are correlated with each other

H_1 : Between variables are not correlated with each other

Test statistics:

$$X^2_{count} = -\left\{n-1-\frac{2j+5}{6}\right\} ln|\boldsymbol{R}|$$

$$= -\left(34-1-\left(\frac{11}{6}\right)\right) ln|0.748902122| = 9.011747612$$

Compared with $X^2_{table} = X^2_{a.\frac{1}{2}j(j-1)} = X^2_{0.05.\frac{1}{2}3(3-1)} = X^2_{0.05.3} = 7.8147$

Because $X_{count}^2 > X_{table}^2$, then H_0 failed to be rejected so the response variables were dependent. Thus, simultaneous modeling can be performed for all response variables.

B. Simultaneous Modeling Based on Nonparametric Regression Approach with Fourier Series Estimator

The nonparametric regression approach with the Fourier series estimator has an oscillation parameter (k). The selection of k in each basis of the Fourier series function takes place iteratively until the minimum GCV value is obtained. Furthermore, the basis of the best Fourier series function is chosen by considering the parsimony principle, where a good model contains an optimal number of parameters. By using the R software, the minimum GCV calculation results for the in-sample data are presented in Table 4 as follows.

Base	k	GCV	MSE	R^2
Cos Sin	1	1.483961	2.1257374	0.923143
Cos	1	1.724324	2.5864239	0.220647
Sin	1	1.518111	2.6459449	0.356552

TABLE 4. Performance Comparison of the Three Base Functions of the Fourier Series

Table 4 shows that the minimum GCV value on the three bases of the Fourier series function is obtained when the value of $k = 1$. If further reviewed, the GCV generated in the Fourier series with the cosine and sine bases has the smallest value among other bases, which is 1.483961. By considering the parsimony principle, it can be seen in Table 4 that with $k = 1$, a Fourier series function with cosine and sine bases can produce a model estimate with an R^2 of 0.923. Thus, it was decided to choose the cosine and sine bases to perform simultaneous modeling based on nonparametric regression with a Fourier series estimator to predict the distribution of welfare assistance in Indonesia.

C. Simultaneous Modeling Based on Nonparametric Regression Approach with Linear Truncated Spline Estimator

The nonparametric regression approach using the best linear truncated Spline estimator is obtained from the selection of the optimum knot points. The optimum knot point is obtained at the minimum GCV value. Furthermore, the goodness of the results of simultaneous modeling using the Spline estimator is reviewed by looking at the MSE and R^2 values.

Number of Knots	GCV	MSE	R^2
1 Knot	34.92438	1.23522	0.9095852
2 Knot	37.23868	2.64966	0.4430581
3 Knot	40.17619	4.20778	0.1954934

TABLE 5. Comparison of the Performance of the Model Formed by Many Knot Points

Based on Table 5, the best model is the nonparametric regression model of multi-response linear truncated Spline with 1 knot point. This is because the model produces the smallest GCV, which is 34.92438. In addition, the MSE and R^2 values generated by this model are also the smallest among other models.

D. Determination of the Best Model

Simultaneous modeling results based on a nonparametric regression approach using Fourier series and Spline estimators are compared to determine the best estimator to predict the distribution of welfare assistance in Indonesia.

Estimator	GCV	MSE	R^2	Information
Fourier Series	1.483961	2.1257374	0.923143	$k = 1$ (Optimal oscillation parameters)
Spline	34.92438	1.23522	0.909585	The optimum number of knots = 1

TABLE 6. Fourier Series and Spline of Performance Comparison

Based on Table 6, it appears that the estimator of the Fourier series basis of cosine and sine has a GCV of 1.483961. This value is smaller than the GCV of the linear truncated Spline estimator of 34.92438. The R^2 value of the cosine and sine basis Fourier series estimator is also greater than the linear truncated Spline estimator R^2. By using the goodness-of-model principle, the estimator chosen to predict the distribution of welfare assistance in Indonesia is the Fourier series estimator with cosine and sine bases.

In this study, there are 3 response variables and 8 predictor variables. Based on the value of the optimum oscillation parameter (k), which is 1, it is obtained a nonparametric regression model with a Fourier series approach using cosine and sine bases whose general form is presented in Equation (20) as follows:

$$\hat{y}_{i1} = \sum_{l=1}^{8}\left(\frac{\hat{\alpha}_{01l}}{2} + \hat{\gamma}_{1l}x_{i1l} + \hat{\alpha}_{11l}\cos x_{i1l} + \hat{\beta}_{11l}\sin x_{i1l}\right)$$

$$\hat{y}_{i2} = \sum_{l=1}^{8}\left(\frac{\hat{\alpha}_{02l}}{2} + \hat{\gamma}_{2l}x_{i2l} + \hat{\alpha}_{21l}\cos x_{i2l} + \hat{\beta}_{21l}\sin x_{i2l}\right) \quad (20)$$

$$\hat{y}_{i3} = \sum_{l=1}^{8}\left(\frac{\hat{\alpha}_{03l}}{2} + \hat{\gamma}_{3l}x_{i3l} + \hat{\alpha}_{31l}\cos x_{i3l} + \hat{\beta}_{31l}\sin x_{i3l}\right)$$

Based on the results of calculations using R software, the parameter values in the model are presented in full in Table 7 as follows:

$j = 1$		Intercept = 241.3421		
		$\hat{\gamma}_{1l}$	$\hat{\alpha}_{11l}$	$\hat{\beta}_{11l}$
$\hat{y}_{i1}$	x_{i11}	-1.1027	-6.2379	-0.8252
	x_{i12}	1.0922	-2.539	5.246
	x_{i13}	0.3353	1.1056	-3.7754
	x_{i14}	-18.2259	-14.7898	-11.3188
	x_{i15}	-230.1675	-112.0103	212.0973
	x_{i16}	51.786	-51.5935	-57.9635
	x_{i17}	-1.3812	7.5776	-3.6989
	x_{i18}	-0.9849	2.8838	1.6745
$j = 2$		Intercept = 912.2795		
		$\hat{\gamma}_{2l}$	$\hat{\alpha}_{21l}$	$\hat{\beta}_{21l}$
$\hat{y}_{i2}$	x_{i11}	-3.2137	-2.5255	-6.9029
	x_{i12}	-1.7077	0.5383	-7.5305
	x_{i13}	1.5645	7.4605	-4.0603
	x_{i14}	6.6974	10.1275	0.1689
	x_{i15}	720.7702	564.6149	-550.2312
	x_{i16}	-2.297.859	-1.452.055	2.075.36
	x_{i17}	-8.3827	9.6926	-4.1963
	x_{i18}	0.1163	-3.9152	-3.4875
$j = 3$		Intercept = -6565444		
		$\hat{\gamma}_{3l}$	$\hat{\alpha}_{31l}$	$\hat{\beta}_{31l}$
$\hat{y}_{i3}$	x_{i11}	168.055.5	5.838.495	685.052.2
	x_{i12}	-267.292.8	-153.028	-802.420.6
	x_{i13}	54.701.93	14.479.5	64.175.43
	x_{i14}	38.769.57	-30.343.48	111.369.9
	x_{i15}	1.167.304	7.447.686	-1.079.865
	x_{i16}	7.554.179	-2.463.776	-8.174.513
	x_{i17}	-95.688.12	424.853.3	-58.235.19
	x_{i18}	184.929.3	104.927.6	-371.741

TABLE 7. Parameters Value in Model

Furthermore, the estimator in Equation (10) whose parameter estimator values are presented in Table 7 is used to predict the distribution of welfare assistance in Indonesia.

E. Prediction of Welfare Distribution in Indonesia in 2020

The distribution of welfare assistance in Indonesia in the following year is predicted to use 2020 data as out-sample data based on a nonparametric regression approach using a Fourier series estimator whose equation has been formulated previously. Prediction results are compared with actual data in 2020 as presented in Table 8 as follows:

Province	y_{i1}	$\hat{y}_{i1}$	y_{i2}	$\hat{y}_{i2}$	y_{i3}	$\hat{y}_{i3}$
Aceh	82.89	84.87	67.52	68.66	451,940	452,548.31
North Sumatera	31.1	30.97	38.2	38.77	769,342	769,564.22
West Sumatera	36.7	34.17	39.89	41.22	295,752	295,866.52
Riau	23.82	26.73	35.38	36.18	290,601	290,542.41
Jambi	23.54	25.44	37.66	38.61	189,741	189,965.32
South Sumatera	32.26	31.86	38.75	39.22	569,116	569,228.96
Bengkulu	33.18	32.31	42.96	43.25	143,305	143,240.62
Lampung	37.51	37.93	53.94	55.18	777,161	777,912.76
Bangka Belitung Islands	29.98	28.22	32.62	33.76	56,937	56,986.12
Riau Islands	23.27	25.33	14.6	15.27	63,483	63,553.12
Jakarta	52.3	51.82	14.16	12.21	198,393	198,721.52
West Java	30.46	30.28	37.48	38.42	3,311,983	3,312,762.31
Central Java	40.84	42.23	45.19	46.72	3,423,462	3,423,567.42
Yogyakarta	47.27	49.55	51.39	52.31	355,098	353,421.11
East Java	33.74	32.92	40.07	42.15	3,515,152	3,515,672.44
Banten	30.78	30.62	45.5	47.21	563,134	563,222.12
Bali	35.95	33.25	25.67	26.42	169,490	169,531.37
West Nusa Tenggara	42.84	44.71	65.53	67.43	577,559	577,932.12
East Nusa Tenggara	46.67	48.71	60.9	64.21	504,194	503,213.22
West Kalimantan	27.47	27.61	39.64	40.12	303,557	303,642.15
Central Kalimantan	33.36	32.56	15.99	19.22	100,264	100,432.46
South Kalimantan	31.86	31.53	44.5	45.33	188,370	188,564.91
East Kalimantan	22.64	21.65	36.73	38.12	125,079	125,182.45
North Kalimantan	39.15	40.11	28.92	31.22	28,641	28,754.91
North Sulawesi	45.12	45.22	44.33	44.98	164,371	164,422.64
Central Sulawesi	45.63	46.91	60.55	62.11	221,982	222,654.37
South Sulawesi	50.17	50.12	46.16	47.91	647,922	647,931.44

Province	y_{i1}	$\hat{y}_{i1}$	y_{i2}	$\hat{y}_{i2}$	y_{i3}	$\hat{y}_{i3}$
Southeast Sulawesi	46.45	48.46	51.06	50.16	199,591	199,782.27
Gorontalo	62.09	60.23	49.53	49.77	113,750	113,827.43
West Sulawesi	55.29	56.12	63.11	66.54	101,617	101,699.23
Maluku	32.9	32.12	51.21	51.98	122,879	122,542.36
North Maluku	31.38	31.32	26.89	29.59	51,252	51,398.64
West Papua	47.15	48.79	57.25	58.25	75,822	75,984.58
Papua	30.43	29.87	47.88	48.87	326,408	327,654.19

TABLE 8. Comparison of Actual Data (Out Sample) and Predicted Data

In facilitating the process of interpreting the prediction results, a ranking of the distribution of welfare assistance was carried out in 2020 and the predicted results in 2020. The ranking results for the three response variables, namely the percentage of BPJS PBI recipients (y_{i1}), percentage of rastra recipients (y_{i2}), and the number of Beneficiary Families (KPM) (y_{i3}) presented respectively in Table 9, Table 10, and Table 11 as follows.

Percentage of BPJS PBI Recipients in 2020 (*Out Sample*)			**Prediction of the Percentage of BPJS PBI Recipients in 2020**		
Province	y_{i1}	**Rank**	**Province**	$\hat{y}_{i1}$	**Rank**
Aceh	82.89	1	Aceh	84.87	1
Gorontalo	62.09	2	Gorontalo	60.23	2
West Sulawesi	55.29	3	West Sulawesi	56.12	3
Jakarta	52.30	4	Jakarta	51.82	4
South Sulawesi	50.17	5	South Sulawesi	50.12	5
Yogyakarta	47.27	6	Yogyakarta	49.55	6
West Papua	47.15	7	West Papua	48.79	7
East Nusa Tenggara	46.67	8	East Nusa Tenggara	48.71	8
Southeast Sulawesi	46.45	9	South-east Sulawesi	48.46	9
Central Sulawesi	45.63	10	Central Sulawesi	46.91	10
North Sulawesi	45.12	11	North Sulawesi	45.22	11
West Nusa Tenggara	42.84	12	West Nusa Tenggara	44.71	12
Central Java	40.84	13	Central Java	42.23	13
North Kalimantan	39.15	14	North Kalimantan	40.11	14
Lampung	37.51	15	Lampung	37.93	15
West Sumatera	36.70	16	West Sumatera	34.17	16
Bali	35.95	17	Bali	33.25	17
East Java	33.74	18	East Java	32.92	18
Central Kalimantan	33.36	19	Central Kalimantan	32.56	19

Percentage of BPJS PBI Recipients in 2020 (*Out Sample*)			**Prediction of the Percentage of BPJS PBI Recipients in 2020**		
Province	y_{i1}	**Rank**	**Province**	$\hat{y}_{i1}$	**Rank**
Bengkulu	33.18	20	Bengkulu	32.31	20
Maluku	32.90	21	Maluku	32.12	21
South Sumatera	32.26	22	South Sumatera	31.86	22
South Kalimantan	31.86	23	South Kalimantan	31.53	23
North Maluku	31.38	24	North Maluku	31.32	24
North Sumatera	31.10	25	North Sumatera	30.97	25
Banten	30.78	26	Banten	30.62	26
West Java	30.46	27	West Java	30.28	27
Papua	30.43	28	Papua	29.87	28
Bangka Belitung Islands	29.98	29	Bangka Belitung Islands	28.22	29
West Kalimantan	27.47	30	West Kalimantan	27.61	30
Riau	23.82	31	Riau	26.73	31
Jambi	23.54	32	Jambi	25.44	32
Riau Islands	23.27	33	Riau Islands	25.33	33
East Kalimantan	22.64	34	East Kalimantan	21.65	34

TABLE 9. Comparison of BPJS PBI Percentage Ranks in 2020 as Out Sample Data with Predicted Percentage of BPJS PBI Recipients in 2020 by Province

Percentage of Rastra Recipients in 2020 (*Out Sample*)			**Predicted Percentage of Rastra Recipients in 2020**		
Province	y_{i2}	**Rank**	**Province**	$\hat{y}_{i2}$	**Rank**
Aceh	67.52	1	Aceh	68.66	1
West Nusa Tenggara	65.53	2	West Nusa Tenggara	67.43	2
West Sulawesi	63.11	3	West Sulawesi	66.54	3
East Nusa Tenggara	60.9	4	East Nusa Tenggara	64.21	4
Central Sulawesi	60.55	5	Central Sulawesi	62.11	5
West Papua	57.25	6	West Papua	58.25	6
Lampung	53.94	7	Lampung	55.18	7
Yogyakarta	51.39	8	Yogyakarta	52.31	8
Maluku	51.21	9	Maluku	51.98	9
South-east Sulawesi	51.06	10	South-east Sulawesi	50.16	10
Gorontalo	49.53	11	Gorontalo	49.77	11
Papua	47.88	12	Papua	48.87	12
South Sulawesi	46.16	13	South Sulawesi	47.91	13
Banten	45.5	14	Banten	47.21	14

Percentage of Rastra Recipients in 2020 (*Out Sample*)			**Predicted Percentage of Rastra Recipients in 2020**		
Province	y_{i2}	**Rank**	**Province**	$\hat{y}_{i2}$	**Rank**
Central Java	45.19	15	Central Java	46.72	15
South Kalimantan	44.5	16	South Kalimantan	45.33	16
North Sulawesi	44.33	17	North Sulawesi	44.98	17
Bengkulu	42.96	18	Bengkulu	43.25	18
East Java	40.07	19	East Java	42.15	19
West Sumatera	39.89	20	West Sumatera	41.22	20
West Kalimantan	39.64	21	West Kalimantan	40.12	21
South Sumatera	38.75	22	South Sumatera	39.22	22
North Sumatera	38.2	23	North Sumatera	38.77	23
Jambi	37.66	24	Jambi	38.61	24
West Java	37.48	25	West Java	38.42	25
East Kalimantan	36.73	26	East Kalimantan	38.12	26
Riau	35.38	27	Riau	36.18	27
Bangka Belitung Islands	32.62	28	Bangka Belitung Islands	33.76	28
North Kalimantan	28.92	29	North Kalimantan	31.22	29
North Maluku	26.89	30	North Maluku	29.59	30
Bali	25.67	31	Bali	26.42	31
Central Kalimantan	15.99	32	Central Kalimantan	19.22	32
Riau Islands	14.6	33	Riau Islands	15.27	33
Jakarta	14.16	34	Jakarta	12.21	34

TABLE 10. Comparison of Ranking Percentage of Rastra Recipients in 2020 as Out-Sample Data with Predicted Percentage of Rastra Recipients in 2020 by Province

Number of KPM in 2020 (*Out Sample*)			**Predicted Number of KPM in 2020**		
Province	y_{i3}	**Rank**	**Province**	$\hat{y}_{i3}$	**Rank**
East Java	3,515,152	1	East Java	3,515,672.44	1
Central Java	3,423,462	2	Central Java	3,423,567.42	2
West Java	3,311,983	3	West Java	3,312,762.31	3
Lampung	777,161	4	Lampung	777,912.76	4
North Sumatera	769,342	5	North Sumatera	769,564.22	5
South Sulawesi	647,922	6	South Sulawesi	647,931.44	6
West Nusa Tenggara	577,559	7	West Nusa Tenggara	577.932,12	7
South Sumatera	569,116	8	South Sumatera	569,228.96	8
Banten	563,134	9	Banten	563,222.12	9
East Nusa Tenggara	504,194	10	East Nusa Tenggara	503,213.22	10
Aceh	451,940	11	Aceh	452,548.31	11

Number of KPM in 2020 (*Out Sample*)			**Predicted Number of KPM in 2020**		
Province	y_{i3}	**Rank**	**Province**	$\hat{y}_{i3}$	**Rank**
Yogyakarta	355,098	12	Yogyakarta	353,421.11	12
Papua	326,408	13	Papua	327,654.19	13
West Kalimantan	303,557	14	West Kalimantan	303,642.15	14
West Sumatera	295,752	15	West Sumatera	295,866.52	15
Riau	290,601	16	Riau	290,542.41	16
Central Sulawesi	221,982	17	Central Sulawesi	222,654.37	17
Southeast Sulawesi	199,591	18	Southeast Sulawesi	199,782.27	18
Jakarta	198,393	19	Jakarta	198,721.52	19
Jambi	189,741	20	Jambi	189,965.32	20
South Kalimantan	188,370	21	South Kalimantan	188,564.91	21
Bali	169,490	22	Bali	169,531.37	22
North Sulawesi	164,371	23	North Sulawesi	164,422.64	23
Bengkulu	143,305	24	Bengkulu	143,240.62	24
East Kalimantan	125,079	25	East Kalimantan	125,182.45	25
Maluku	122,879	26	Maluku	122,542.36	26
Gorontalo	113,750	27	Gorontalo	113,827.43	27
West Sulawesi	101,617	28	West Sulawesi	101,699.23	28
Central Kalimantan	100,264	29	Central Kalimantan	100,432.46	29
West Papua	75,822	30	West Papua	75,984.58	30
Riau Islands	63,483	31	Riau Islands	63,553.12	31
Bangka Belitung Islands	56,937	32	Bangka Belitung Islands	56,986.12	32
North Maluku	51,252	33	North Maluku	51,398.64	33
North Kalimantan	28,641	34	North Kalimantan	28,754.91	34

TABLE 11. Comparison of the Number of KPM in 2020 as Out-Sample Data with Predicted Number of KPM Recipients in 2020 by Province

Based on the results of the ranking comparison, there is no significant shift in ranking between the distribution of welfare assistance in 2020 and its prediction by province. This can be used as an indication that the model based on multi-response multi-predictor nonparametric regression with Fourier series to predict the distribution of welfare assistance in Indonesia is good.

By using prediction results based on nonparametric regression using a Fourier series estimator, several recommendations are formulated regarding the distribution of the Indonesian government aid program, including as follows:

1. Monitoring and collecting data on the poor or vulnerable to poverty is maximized from agencies with the closest scope such as at the RT level, then government assistance programs, especially BPJS PBI, Rastra, and KPM can be distributed on target.
2. Trying to create a society that supports the achievement of the SDGs in minimizing the level of poverty and poverty.
3. Making enforcement of regulations strictly and rules to prevent misuse of government assistance, especially BPJS PBI, Rastra, and KPM.
4. Provinces in Indonesia that are ranked at the top in the category of providing government assistance programs need to be maintained in stabilizing the distribution of government assistance which includes BPJS PBI, Rastra, and KPM. In addition, it is also necessary to pay attention to their welfare and ensure that the distribution of government assistance must be properly targeted.
5. The prediction results can be used as a reference in determining the priority of racial division by the province in the following year. Besides, this research results can be developed using another statistical method that considers the geographic or spatial aspect, which is geographically weighted regression that has been done by Sediono et al. [22].

5. CONCLUSION

From the results of the comparative analysis of predictions of the percentage of poor families receiving government assistance which includes BPJS PBI, Rastra, and KPM in 2019 and data on the percentage of poor families receiving government assistance which includes BPJS PBI, Rastra, and KPM in 2020, there is no significant difference. Thus the model obtained is good. A related recommendation is to pay attention to the factors that have a positive influence on the distribution of the percentage of poor families receiving government assistance which includes BPJS PBI, Rastra, and KPM as an effort to improve food security and provide social protection evenly, thus the SDGs target can be achieved to move Indonesia up in the world. Society 5.0 era.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] M.F.F. Mardianto, S.H. Kartiko, H. Utami, The performance of nonparametric regression for trend and seasonal pattern in longitudinal data, ARPN J. Eng. Appl. Sci. 15 (2020), 1111-1115.

[2] N. Murbarani, Y. Swastika, A. Dwi, et al. Modeling of the percentage of aids sufferers in east java province with nonparametric regression approach based on spline truncated estimator, Indones. J. Stat. Appl. 3 (2019), 139–147. https://doi.org/10.29244/ijsa.v3i2.209.

[3] A.R. dos S. Sousa, M.T.F. Severino, F.G. Leonardi, Model selection criteria for regression models with splines and the automatic localization of knots, arXiv:2006.02649. (2020). http://arxiv.org/abs/2006.02649.

[4] M.F.F. Mardianto, E. Tjahjono, M. Rifada, Statistical modelling for prediction of rice production in Indonesia using semiparametric regression based on three forms of Fourier series estimator. ARPN J. Eng. Appl. Sci. 14 (2019), 2763-2770.

[5] B. Lestari, Fatmawati, I.N. Budiantara, et al. Smoothing parameter selection method for multiresponse nonparametric regression model using smoothing Spline and Kernel estimators approaches, J. Phys.: Conf. Ser. 1397 (2019), 012064. https://doi.org/10.1088/1742-6596/1397/1/012064.

[6] M.H. Balali, N. Nouri, M. Rashidi, et al. A multi-predictor model to estimate solar and wind energy generations, Int. J. Energy Res. 42 (2017), 696–706. https://doi.org/10.1002/er.3853.

[7] M. Khan, S.F. Adil, H.Z. Alkhathlan, et al. COVID-19: A global challenge with old history, epidemiology and progress so far, Molecules. 26 (2020), 39. https://doi.org/10.3390/molecules26010039.

[8] M.F.F. Mardianto, Suliyanto, F. Effendy, et al. Classification of food menu and grouping of food potential to support the food security and nutrition quality, Commun. Math. Biol. Neurosci., 2022 (2022), 2. https://doi.org/10.28919/cmbn/6801.

[9] Central Bureau of Statistics, Badan Pusat Statistik. The Percentage of Poor Population in September 2021 Decreased into 9.71 Percent, https://www.bps.go.id/pressrelease/2022/01/17/1929/persentase-penduduk-miskin-september-2021-turun-menjadi-9-71-persen.html. (2022).

[10] Central Bureau of Statistics. Statistics Indonesia 2019. Central Bureau of Statistics, Jakarta, (2020).

[11] The National Team for The Acceleration of Poverty Reduction, The Impact of the Transition of the Rastra Program to BNPT, http://www.tnp2k.go.id/articles/the-impact-of-the-transition-of-the-rastra-program-to-bpnt. (2021).

[12] Coordinating Ministry for Human Development and Cultural Affairs, General Guidelines for Welfare Rice Social Assistance, Central Rastra Coordination Team, Jakarta, (2018).

[13] N.P.A.M. Mariati, I.N. Budiantara, V. Ratnasari, The application of mixed smoothing spline and Fourier series model in nonparametric regression, Symmetry. 13 (2021), 2094. https://doi.org/10.3390/sym13112094.

[14] R. A. Gordon, Regression analysis, SAGE Publications, Newbury Park, 2020.

[15] A.F. Schmidt, C. Finan, Linear regression and the normality assumption, J. Clinic. Epidemiol. 98 (2018), 146–151. https://doi.org/10.1016/j.jclinepi.2017.12.006.

[16] A.E. Marques, P.A. Prates, A.F.G. Pereira, et al. Performance comparison of parametric and non-parametric regression models for uncertainty analysis of sheet metal forming processes, Metals. 10 (2020), 457. https://doi.org/10.3390/met10040457.

[17] M.F.F. Mardianto, E. Tjahjono, M. Rifada, Semiparametric regression based on three forms of trigonometric function in Fourier series estimator, J. Phys.: Conf. Ser. 1277 (2019), 012052. https://doi.org/10.1088/1742-6596/1277/1/012052.

[18] P. Tavallali, P. Tavallali, M. Singhal, Optimization of Hierarchical Regression Model with Application to Optimizing Multi-Response Regression K-ary Trees, Proc. AAAI Conf. Artif. Intell. 33 (2019), 5133–5142. https://doi.org/10.1609/aaai.v33i01.33015133.

[19] K. Takezawa, Introduction to nonparametric regression, John Wiley & Sons, New Jersey, (2006).

[20] Social Security Administrator for Health, Participant, (2020). https://bpjs-kesehatan.go.id/bpjs/index.php/pages/detail/2014/11.

[21] Central Bureau of Statistics. Statistics Indonesia 2020. Central Bureau of Statistics, Jakarta, (2021).

[22] Sediono, M.F.F. Mardianto, S.M. Ulyah, et al. The modelling of earthquake magnitude in the southern part of Java Island using geographically weighted regression, Commun. Math. Biol. Neurosci. 2022 (2022), 13 https://doi.org/10.28919/cmbn/6753