Estimation of Social Funds during the Covid - 19 Pandemic based on the Fourier Series Estimator in Semiparametric Regression for Longitudinal Data

by Kuzairi Kuzairi

Estimation of Social Funds during the Covid - 19 Pandemic based on the Fourier Series Estimator in Semiparametric Regression for Longitudinal Data

Kuzairi[a, b], Miswanto[b,*], Toha Saifudin[b], M Fariz Fadillah Mardianto[b], Aeri Rachmad[c]

[a]Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Islam Madura, Indonesia
[b]Department of Mathematics, Faculty of Science and Technology, Airlangga University, Indonesia
[c]Department of Informatics, Faculty of Engineering, University of Trunojoyo Madura, Indonesia

Abstract: Regression analysis is one of the techniques commonly used in statistics. There are three approaches in regression analysis, namely parametric, nonparametric and semiparametric. Semiparametric regression is a combination of parametric regression and nonparametric regression. One that is developed in semiparametric regression is the Fourier series approach to longitudinal data. Semiparametric regression using the Fourier series is able to overcome data that have trigonometric distribution. Longitudinal data is data with observations made as an independent subject for each subject that is observed repeatedly in different periods depending on the dependence. In the estimation of the Fourier series, the determination of optimal K uses the minimum Generalized Cross Validation (GCV), the smallest Mean Square Error (MSE) value and high determination coefficient. The application of this research was carried out on the income of social funds that came in and went out during the Covid-19 pandemic which was thought to be influential. In this study, the Fourier series semiparametric regression model for longitudinal data was formed with an optimal K value of 23, a GCV of 68.89 and an MSE value of 0.0006229203. The estimation results have met the given goodness of fit criteria.

Keywords: Fourier series Estimator, Semiparametric Regression, Social Funds, Pandemic Covid-19, Longitudinal Data

1. Introduction

The regression method is a method in Statistics used to determine the relationship between predictor variables and response variables (**Chamidah *et. al.*, 2021**). The relationship between the predictor variable and the response variable that has been stated in the regression model can be used for prediction and forecasting. For the accuracy of prediction and forecasting results, in using regression analysis it is necessary to pay attention to the approach used when estimating the regression curve. There are three approaches in regression analysis to estimate the regression curve, namely the parametric, nonparametric, and semiparametric approaches.

In parametric regression it is assumed that the shape of the regression curve is known based on previous information or past experience. Meanwhile, nonparametric regression does not provide a certain curve shape assumption or there is no information about the shape of the regression curve. The regression curve is assumed to be smooth so that nonparametric regression has high flexibility because the data is expected to find its own form of regression curve estimation without being influenced by the subjectivity of the researcher (**Eubank, 1999**). If the response variable is known to have the relationship pattern with the predictor variable, but with other predictor variables the relationship pattern is not known. So in cases like this suggest using semiparametric regression. Semiparametric regression is a combination of parametric regression and nonparametric regression (**Hardle, 1990**) and (**Mardianto, 2017**). Several studies on semiparametric regression include spline (**Chamidah *et. al.*, 2018**), local linear (**Chamidah and Rifada, 2006**) and the Fourier series (**Mardianto, 2019**) and (**Kuzairi, *et. al.*, 2020**).

Along with the development of data analysis, regression modeling has also developed for longitudinal data types. Longitudinal data are observations made as many as n subjects (cross section) independently with each subject being observed repeatedly in different time periods (time series). Longitudinal data has advantages including in the same number of subjects, the results of measurement error result in more efficient forecasting of treatment effects than cross section data because longitudinal data estimates are made for each observation and are more powerful even though they only use fewer subjects (**Wu and Zhang, 2006**). Longitudinal data is a special form of repeating measurement data. In repeated measurement data, a single measurement is collected repeatedly for each subject or experiment. Observations can be collected over time. Several studies using longitudinal data with spline truncated (**Islamiyati, *et. al.*, 2018**), Fourier series estimator (**Mardianto, 2019**) and (**Kuzairi, *et. al.*, 2020**). The application of longitudinal data develops in various fields, including economic and social.

One of the popular methods for estimating regression models with nonparametric and semiparametric approaches is the Fourier series. The Fourier series is a trigonometric polynomial that has flexibility. The Fourier series is best used to describe curves representing sine and cosine waves. The appropriate data to be analyzed using the Fourier series is if the data being investigated has

no known pattern and there is a tendency for the data pattern to repeat itself (**Bilodeau, M, 1992**). However, these studies are only on cross section data or data observed at a certain time. For special cases, semiparametric regression can be used on longitudinal data.

In the semiparametric regression research for longitudinal data based on the Fourier series estimator, it is applied to the income of incoming social funds and outgoing funds during the Covid-19 pandemic. As we know and it has occurred in early 2020, Covid-19 has become a troubling problem world health. This case began with information from the World Health Organization (WHO) on December 31, 2019, which states that there have been cases of pneumonia cluster cases with unclear etiology in Wuhan City, Hubei Province, China. This problem continues to develop until it is known that the cause of this pneumonia cluster is the novel coronavirus. This case continues to grow until there are reports of deaths and importations outside China. On January 30, 2020, WHO declared Covid-19 a Public Health Emergency of International Concern (PHEIC). On February 12, 2020, WHO officially designated the novel coronavirus disease in humans as Coronavirus Disease (COVID-19). The development of the Covid -19 pandemic has greatly affected the economy, even in all lines of community activity, it has decreased significantly, especially in obtaining social funds (alms) for the construction of mosques on Madura Island which consists of four districts, namely Sumenep, Pamekasan, Sampang and Bangkalan. Alms giving is a gift given by one person to another spontaneously and voluntarily without being limited by a certain time and amount. The benefits of estimating social fund revenue (alms) to determine the available fund reserves. Research on semiparametric regression based on the Fourier series estimator has been studied a lot but is limited to cross section data. Therefore, the purpose of this study is how to estimate the semiparametric regression curve on longitudinal data using the Fourier series estimator. Next, apply the regression model to the data on the acquisition and expenditure of social funds (alms) for the construction of mosques on Madura Island during the Covid-19 pandemic.

2. Semiparametric Regression Model of Fourier Series Estimator For Longitudinal Data

Longitudinal data is data obtained from measurements or observations made as many as n independent subjects with each subject being observed repeatedly in different interdependent periods (**Wu and Zhang, 2006**). If y_{ij} expressing observations for subject to -i and time to -j, x_{pij} is the predictor variable to -p, the parametric component represents observations for subject to -i and time to -j. and t_{qij} is the nonparametric component of the predictor variable which states the observations for the subject to -i and time to -j, n states the number of subjects and m represents the number of events in the subject to -i, the longitudinal data structure in this study is presented in Table 1.

Table 1. The structure of longitudinal data

Subject	Response	Predictors							
		Parametric				Nonparametric			
	y_{ij}	x_{1ij}	x_{2ij}	...	x_{pij}	t_{1ij}	t_{2ij}	...	t_{qij}
1st Subject	y_{11}	x_{111}	x_{211}	...	x_{p11}	t_{111}	t_{211}	...	t_{q11}
	y_{12}	x_{121}	x_{212}	...	x_{p12}	t_{112}	t_{212}	...	t_{q12}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	y_{1m}	x_{11m}	x_{21m}	...	x_{p1m}	t_{11m}	t_{21m}	...	t_{q1m}
2nd Subject	y_{21}	x_{121}	x_{221}	...	x_{p21}	t_{121}	t_{221}	...	t_{q21}
	y_{22}	x_{122}	x_{222}	...	x_{p22}	t_{122}	t_{222}	...	t_{q22}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	y_{2m}	x_{12m}	x_{22m}	...	x_{p2m}	t_{12m}	t_{22m}	...	t_{q2m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
nnd Subject	y_{n1}	x_{1n1}	x_{2n1}	...	x_{pn1}	t_{1n1}	t_{2n1}	...	t_{qn1}
	y_{n2}	x_{1n2}	x_{2n2}	...	x_{pn2}	t_{1n2}	t_{2n2}	...	t_{qn2}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	y_{nm}	x_{1nm}	x_{2nm}	...	x_{pnm}	t_{1nm}	t_{2nm}	...	t_{qnm}

Semiparametric regression is a combination of parametric regression and nonparametric regression. Given the data paired $\left(y_{ij}, x_{pij}, t_{qij} \right)$ with is the response variable for subject to -i and observation to -j, x_{pij} is the predictor variable to -p the parametric component for subject to -i and observations to -j, t_{qij} is the predictor variable to -q the nonparametric component with no known

effect on the response for subject to $-i$ and observations to $-j$, where, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, $p = 1, 2, \ldots, P$, $q = 1, 2, \ldots, Q$. In general, the semiparametric longitudinal data regression model can be presented in the following equation:

(1)
$$y_{ij} = \beta_{0i} + \sum_{p=1}^{P} \beta_{pi} x_{pij} + \sum_{q=1}^{Q} g_{qi}\left(t_{qij}\right) + \varepsilon_{ij}$$

where, β_{0i}, β_{pi} is a parameter of parametric regression, $g_{qi}\left(t_{qij}\right)$ is a nonparametric regression function for subject to $-i$ and observations to $-j$, predictors of $-q$ and is the residual for subject to $-i$ and observations to $-j$, where ε_{ij} are identical, independent random errors with normal distribution of mean of 0 and variance σ^2.

In semiparametric regression modeling for longitudinal data on nonparametric functions such as equation (1) it can be approximated by the Fourier series function which approaches the regression curve $g_{qi}\left(t_{qij}\right)$ with the following equation:

(2)
$$g_{qi}\left(t_{qij}\right) = \sum_{q=1}^{Q}(b_{qi} t_{qij} + \frac{a_{0qi}}{2} + \sum_{k=1}^{K}(a_{kqi} \cos k t_{qij}))$$

where, b_{qi}, $\frac{a_{0qi}}{2}$, a_{kqi}, the parameter of the nonparametric regression component, t_{qij} is the predictor variable to $-q$ the nonparametric component for the subject to $-i$ and observations to $-j$ and $k = 1, 2, \ldots, K$ is the oscillation parameter which is a measure of smoothing. If equation (1) is substituted with equation (2), the semiparametric regression equation based on the Fourier series estimator for longitudinal data will be obtained as follows:

(3)
$$y_{ij} = \beta_{0i} + \sum_{p=1}^{P} \beta_{pi} x_{pij} + \sum_{q=1}^{Q}(b_{qi} t_{qij} + \frac{a_{0qi}}{2} + \sum_{k=1}^{K}(a_{kqi} \cos k t_{qij})) + \varepsilon_{ij}$$

K is the oscillation parameter. Parameters whose values can be determined based on the Weighted Least Square (WLS) which has been researched by (**Kuzairi *et. al.*, 2021**). WLS optimization form is given as follows:

(4)
$$\min_{g \in C(0,\pi)}[R(g)] = \min_{g \in C(0,\pi)} \boldsymbol{\varepsilon}^{\mathsf{T}} \mathbf{W} \boldsymbol{\varepsilon} = \min_{g \in C(0,\pi)} \{(\mathbf{y} - g(\mathbf{x}, \mathbf{t})^{T} \mathbf{W}(\mathbf{y} - g(\mathbf{x}, \mathbf{t}))\}$$

g contains the $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ parameters so that the equation is given as follow:

(5)
$$\min_{g \in C(0,\pi)}[R(g)] = \min_{g \in C(0,\pi)} \boldsymbol{\varepsilon}^{\mathsf{T}} \mathbf{W} \boldsymbol{\varepsilon} = \min_{g \in C(0,\pi)} \{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{T}\boldsymbol{\eta})^{\mathsf{T}} \mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{T}\boldsymbol{\eta})\}$$

where $\mathbf{W}$ is the weighted matrix

Then the estimation results from equation (3) are as follows:

(6)
$$\hat{y}_{ij} = \hat{\beta}_{0i} + \sum_{p=1}^{P} \hat{\beta}_{pi} x_{pij} + \sum_{q=1}^{Q}(\hat{b}_{qi} t_{qij} + \frac{\hat{a}_{0qi}}{2} + \sum_{k=1}^{K}(\hat{a}_{kqi} \cos k t_{qij}))$$

3. Studies Parameter Estimator on the Semiparametric Regression Model of Fourier Series Estimator for Longitudinal Data

Regression curve estimates for the parametric and nonparametric components can be solved based on the appropriate WLS optimization method to obtain the parameter estimators $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$, whose weights are determined by (**Wu and Zhang, 2006**). This optimization is done to minimize the goodness of fit. then by doing the optimization of equation (5) in order to obtain the following equation:

(7)
$$R(\boldsymbol{\beta}, \boldsymbol{\eta}) = \mathbf{y}^{\mathsf{T}} \mathbf{W} \mathbf{y} - 2\mathbf{y}^{\mathsf{T}} \mathbf{W} \mathbf{X} \boldsymbol{\beta} - 2\boldsymbol{\eta}^{\mathsf{T}} \mathbf{T}^{\mathsf{T}} \mathbf{W} \mathbf{y} + 2\boldsymbol{\beta}^{\mathsf{T}} \mathbf{X}^{\mathsf{T}} \mathbf{W} \mathbf{T} \boldsymbol{\eta} + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{X}^{\mathsf{T}} \mathbf{W} \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\eta}^{\mathsf{T}} \mathbf{T}^{\mathsf{T}} \mathbf{W} \mathbf{T} \boldsymbol{\eta}$$

To obtain an estimator of the parameter $\boldsymbol{\beta}$, it is done by performing individual derivatives $R(\boldsymbol{\beta}, \boldsymbol{\eta})$ against $\boldsymbol{\beta}$. so that the following equation is obtained:

(8)
$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathsf{T}} \mathbf{W} \mathbf{X}\right)^{-1}\left\{\mathbf{X}^{\mathsf{T}} \mathbf{W} \mathbf{y} - \mathbf{X}^{\mathsf{T}} \mathbf{W} \mathbf{T} \hat{\boldsymbol{\eta}}\right\}$$

_____

As well, to obtain an estimator of the parameter η is done by performing individual derivatives $R(\boldsymbol{\beta}, \boldsymbol{\eta})$ against η, so that the following equation is obtained:

(9)
$$\hat{\boldsymbol{\eta}} = \left(\mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{T}\right)^{-1}\left\{\mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{y} - \mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}}\right\}$$

Estimators in equations (8) and (9) are not yet free of parameter, so an estimator that is free from parameters with mutual substitution must be sought. To get $\hat{\boldsymbol{\beta}}$ that is free of parameter, substitute equation (9) to equation (8) as follow:

(10)
$$\hat{\boldsymbol{\beta}} = \mathbf{M}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\left\{\mathbf{X}^{\mathrm{T}} - \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\right\}\mathbf{W}\mathbf{y} = \mathbf{A}(K)\mathbf{y}$$

with
$$\mathbf{M} = (\mathbf{I} - \left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}$$

To get $\hat{\boldsymbol{\eta}}$ that is free of parameter, substitute equation (8) to equation (9) as follow:

(11)
$$\hat{\boldsymbol{\eta}} = \mathbf{N}\left(\mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{T}\right)^{-1}\left\{\mathbf{T}^{\mathrm{T}} - \mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right\}\mathbf{W}\mathbf{y} = \mathbf{B}(K)\mathbf{y}$$

with
$$\mathbf{N} = (\mathbf{I} - \left(\mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{T})^{-1}$$

After getting an estimator for the parametric component and the nonparametric component, then determine the semiparametric regression model estimator with the Fourier Series estimator approach for longitudinal data as follows:

(12)
$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{T}\hat{\boldsymbol{\eta}} = \mathbf{C}(K)\mathbf{y}$$

with $\mathbf{C}(K)\mathbf{y} = \mathbf{X}\mathbf{A}(K)\mathbf{y} + \mathbf{T}\mathbf{B}(K)\mathbf{y}$. $\mathbf{A}(K)\mathbf{y}$ is the hat matrix for parametric components. $\mathbf{B}(K)\mathbf{y}$ is the hat matrix for nonparametric components. Hat matrix for semiparametric regression models with the Fourier series estimator for longitudinal data approach is symbolized as $\mathbf{C}(K)\mathbf{y}$. In this case, k shows the oscillation parameter that is contained in the matrix $\mathbf{C}(K)\mathbf{y}$.

In semiparametric regression modeling with a Fourier series estimator for longitudinal data, the thing to attention is to determine the optimal selection of oscillation parameters. Selection of optimal oscillation parameters usually uses the Generalized Cross Validation (GCV) method. The GCV method is generally defined as follows:

(13)
$$GCV\left(k_1, k_2, ..., k_r\right) = \frac{MSE\left(k_1, k_2, ..., k_r\right)}{\left((nm)^{-1} trace\left(\mathbf{I} - \mathbf{C}(k_1, k_2, ..., k_r)\right)\right)^2}$$

With

(14)
$$MSE\left(k_1, k_2, ..., k_r\right) = (nm)^{-1} \mathbf{y}^T \left(\mathbf{I} - \mathbf{C}(k_1, k_2, ..., k_r)\right)^T \left(\mathbf{I} - \mathbf{C}(k_1, k_2, ..., k_r)\right)\mathbf{y}$$

where $\mathbf{C}(k_1, k_2, ..., k_r)$ is the hat matrix, the GCV value depends on the Mean Square Error (MSE) value because the numerator for the GCV formula is the MSE formula. The measure of the goodness of the model is also determined by the coefficient of determination (R2) which shows the percentage contribution of the predictor variable to the response variable. The best model that can be used for prediction meets the goodness criteria of the model. The criteria for the goodness of the model are the smallest GCV value for the optimal oscillation parameter, the smallest MSE value, and the large determination coefficient value. The coefficient of determination can be presented in the following equation:

$$R^2 = \frac{\left(\hat{\mathbf{y}} - \bar{\mathbf{y}}\right)^{\mathrm{T}}\left(\hat{\mathbf{y}} - \bar{\mathbf{y}}\right)}{\left(\mathbf{y} - \bar{\mathbf{y}}\right)^{\mathrm{T}}\left(\mathbf{y} - \bar{\mathbf{y}}\right)}; 0 \leq R^2 \leq 1$$

(15)

Where $\hat{\mathbf{y}}$ is a vector that contains the estimation results for all subjects and $\bar{\mathbf{y}}$ is a vector that has the average value in each subject. So that the model can be used to estimate according to the goodness of fit criteria. the goodness of fit criterion is the smallest GCV value for the optimal oscillation parameter, the smallest MSE value, and the large determination coefficient value.

4. Result

The data is obtained from the social fund registrar (alms) of mosques on Madura Island, which consists of four districts, namely Bangkalan, Sampang, Pamekasan and Sumenep. The data obtained is longitudinal data regarding the contribution of the acquisition and expenditure of social funds (alms) during the Covid-19 pandemic. The data period starts from January 2020 to December 2020. Procedures in data analysis relating to the estimation of income and expenditure of social funds (alms) for the construction of mosques when the Covid-19 pandemic based on the Fourier series estimator in the semiparametric regression for longitudinal data is given as follows:

1. Literature study related to obtaining social funds for mosque construction and its relationship with predictor variables at the time of the Covid-19 pandemic.
2. Identifying data patterns using a scatter plot for cross section data and time series plots for time series data.
3. Determine the GCV value for each input oscillation parameter.
4. Choose the optimal k value based on the minimum GCV value.
5. Choosing the optimal K value based on the minimum GCV value and with other good criteria such as MSE and R2.
6. Construct the selected model in the form of mathematical equations.
7. Make conclusions based on the estimation results.

The following is a scatterplot and time series plot between the response variable (y) and each predictor variable (x and t):
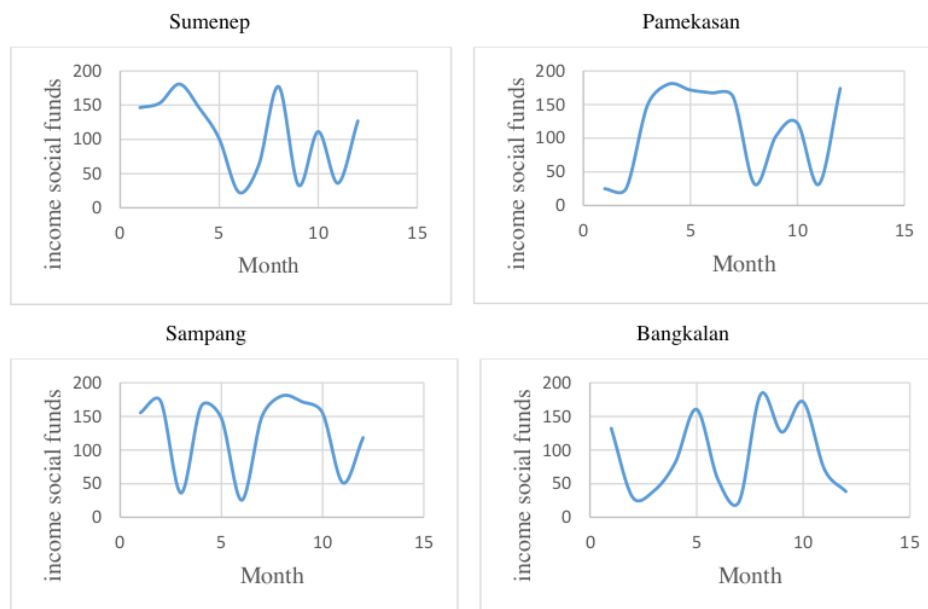
Fig 1. Time Series Plot every subjects between response and parametric component variable

In Fig 1.The data period starts in January 2020 to December 2020, it can be seen that the income of social funds (alms) in each month in Sumenep has a pattern that tends to fluctuate so that it can be seen from January to March there is an increase but the lowest peak occurs in June, during the Covid-19 pandemic. At the beginning of the month in Pamekasan it was relatively low but from March to August it was stable. At the beginning of the month and the following month the data period in Sampang district was relatively stable but the lowest peaks occurred in March, June and November. In Bangkalan district, the data period tends to fluctuate and the lowest peak is in July and the highest peak occurs in August.
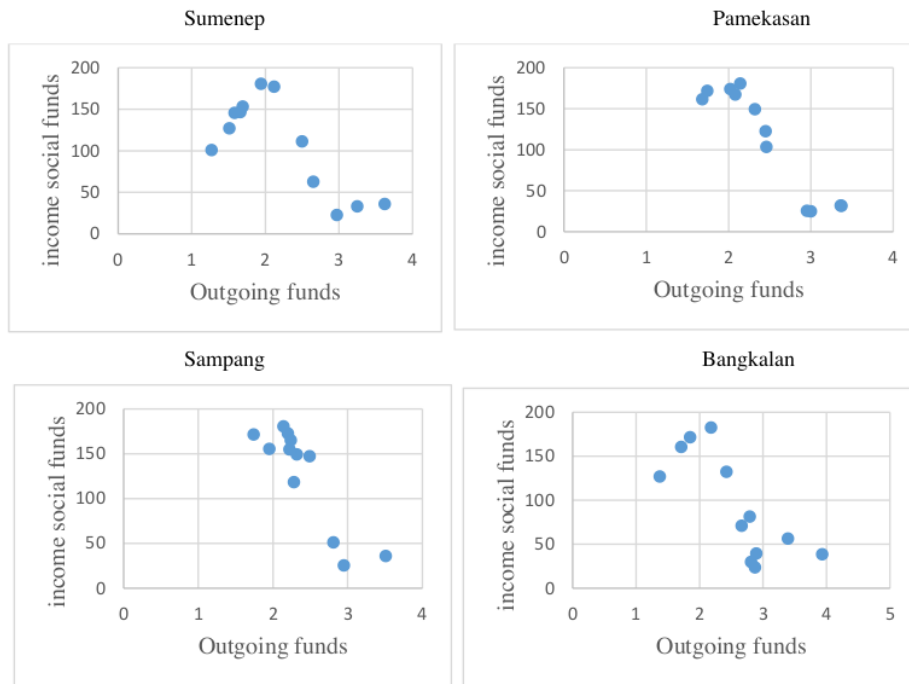
Fig 2. Scatter Plot every subjects between response and nonparametric component variable

In Fig 2. Scatterplots for each response variable with the predictor variables do not show any trend patterns or the shape of the regression curve is not yet known. Regression curves identified through scattered plots form a random pattern or it can be said that they do not follow a certain pattern and tend to repeat themselves. Therefore, this study will use a semiparametric regression approach based on the Fourier series estimator for longitudinal data.

In this study the predictors consisted of one parametric component and one nonparametric component, and consisted of four subjects, the first subject $\hat{y}_{1j}$ was Sumenep district, the second subject $\hat{y}_{2j}$ was Pamekasan district, the third subject $\hat{y}_{3j}$ was Sampang district and the fourth subject $\hat{y}_{4j}$ was Bangkalan district. The results of the optimal GCV value that be calculated from R software used training data are presented in Table 2.

Table 2. GCV Value

k	GCV Value
21	114,65
22	71,37
23	68.89
24	72,07
25	92,13

Based on Table 2. The minimum GCV value for semiparametric regression with the Fourier series estimator for longitudinal data is 68.89 with optimal K equal to 23. Based on the results of calculations using R software, the parameter values can be written in the model as follows:

$$\hat{y}_{1j} = 109.032 + 6.511x_{11j} + 13.289t_{11j} - 0.807\cos t_{11j} + ... + 3.810\cos 23t_{11j}$$

$$\hat{y}_{2j} = 107.744 - 5.402x_{12j} + 7.854t_{12j} - 0.097\cos t_{12j} + ... + 3.872\cos 23t_{12j}$$

$$\hat{y}_{3j} = 110.049 + 7.538x_{13j} + 3.853t_{13j} - 0.744\cos t_{13j} + ... - 2.898\cos 23t_{13j}$$

$$\hat{y}_{4j} = 112.361 + 2.003x_{14j} + 4.358t_{14j} - 0.780\cos t_{14j} + ... - 1.313\cos 23t_{14j}$$

The estimation results obtained by the value of R2 and MSE from semiparametric regression with the Fourier series estimator for longitudinal data, respectively, are 0.9999788 for R2 and 0.0006229203 for MSE. The GCV value is 68.89, so it shows that the model obtained has a very accurate performance for estimation.

5. Conclusion

The Covid-19 pandemic has an impact on obtaining income from social funds on Madura Island. Based on the data pattern, the related data can be modeled by semiparametric regression for longitudinal data based on the Fourier series estimator. The K oscillation parameter can be selected based on the minimum GCV value. The selection of the minimum GCV affects the MSE value which is small and the coefficient of determination is large, so that the model can be used according to the application. Based on the results of data analysis, it was obtained that the GCV value was 68.89, the MSE value was 0.0006229203 and the R2 value was 0.9999788 so that it can be said that the model is good for estimating according to the goodness of fit criteria.

References (APA)

[1]. Bilodeau, M., Fourier Smoother and Additive Models. *Canadian Journal of Statistics* 20 (3), (1992) 257-269.

[2]. Chamidah N, Ari Widyanti, Fitri Trapsilawati, Utami Dyah Syafitri, Local Linear Negative Binomial Nonparametric Regression for Predicting The Number of Speed Violations on Toll Road: A Theoretical Discussion. Commun. Math. Biol. Neurosci. 2021, 2021:10

[3]. Chamidah N, Rifada M., Local linear estimator in bi-response semiparametric regression model for estimating median growth charts of children, Far East Journal of Mathematical Sciences, 99 (8), (2016) pp. 1233-1244.

[4]. Chamidah N, Kurniawan A, Zaman B, Muniroh L., Least Square – Spline Estimator in multi response semiparametric regression model or estimating median growth charts of children in East Java, Far East Journal of Mathematical Sciences, vol. 107 No. 2. (2018) pp 295 – 307.

[5]. Hardle. W, Applied Nonparametric Regression, Cambridge University Press, New York. 1990

[6]. Islamiyati A, Fatmawati, Chamidah, N., Estimation of Covariance Matrix on Bi-Response Longitudinal Data Analysis with Penalized Spline Regression, Journal of Physics: Conference Series, 2018

[7]. Kuzairi, Miswanto, Mardianto M F F,. Semiparametric regression based on fourier series for longitudinal data with Weighted Lest Square (WLS) optimization. Journal of Physics: Conference Series. 2021

[8]. Kuzairi, Miswanto, Budiantara, I., N., Three form fourier series estimator semiparametric regression for longitudinal data. Journal of Physics: Conference Series, 2020

[9]. Mardianto, M.F.F., Modeling Factors that Influence Health Index in Indonesia using Multipredictor Semiparametric Regression with Fourier Series Approach, Proceeding of The 7th Annual Basic Science International Conference 2017, Malang.

[10]. Mardianto, M.F.F., Tjahjono, E., and Rifada, M., Statistical modelling for prediction of rice production in Indonesia using semiparametric regression based on three forms of Fourier series estimator. ARPN Journal of Engineering and Applied Sciences. 14(15) (2019): 2763-2770.

[11]. R. L. Eubank, Nonparametric Regression and Spline Smoothing 2nd Edition. Marcel Deker. New York, United State of America, 1999.

[12]. Wu, H., and Zhang, J. T., Nonparametric Regression Methods for Longitudinal Data Analysis. Willey-Interscience, New Jersey. 2006

Estimation of Social Funds during the Covid - 19 Pandemic based on the Fourier Series Estimator in Semiparametric Regression for Longitudinal Data

1 Kuzairi, Miswanto, I Nyoman Budiantara. "Three form fourier series estimator semiparametric regression for longitudinal data", Journal of Physics: Conference Series, 2020
Publication
3%

2 jurnalsaintek.uinsby.ac.id
Internet Source
2%

3 abd88079-bdc5-4274-9638-f3715aab13b0.filesusr.com
Internet Source
2%

4 9lib.net
Internet Source
2%

5 Muhammad Danil Pasarella, Sifriyani Sifriyani, Fidia Deny Tisna Amijaya. "NONPARAMETRIK REGRESSION MODEL ESTIMATION WITH THE FOURIER SERIES THE FOURIER SERIES APPROACH AND ITS APPLICATION TO THE ACCUMULATIVE COVID-19 DATA IN
2%

INDONESIA", BAREKENG: Jurnal Ilmu
Matematika dan Terapan, 2022
Publication

| 6 | www.turcomat.org
Internet Source | 2% |

| 7 | Submitted to Universitas Negeri Surabaya The State University of Surabaya
Student Paper | 2% |

| 8 | iopscience.iop.org
Internet Source | 2% |