Three form fourier series estimator semiparametric regression for longitudinal data

by Kuzairi Kuzairi

Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Three form fourier series estimator semiparametric regression for longitudinal data

To cite this article: Kuzairi *et al* 2020 *J. Phys.: Conf. Ser.* **1538** 012058

Three form fourier series estimator semiparametric regression for longitudinal data

Kuzairi[1,2], Miswanto[2], I Nyoman Budiantara[3]

[1]Departement of Mathematics, Faculty of Mathematics and Natural Science, The Islamic Uniersity of Madura, Pamekasan 69317, Indonesia,
[2]Department of Mathematics, Faculty of Sciences and Technology, Airlangga University Surabaya 60115, Indonesia.
[3]Department of Statistics, Sepuluh Nopember Institute of Technology Surabaya 60111, Indonesia.

E-mail: miswanto@fst.unair.ac.id

Abstract. Analysis of regressionis one technique that is often used in statistical analysis. There are three regression analysis approaches, such as parametric regression, nonparametric regression and semiparametric regression. Semiparametric regression consists of parametric components and nonparametric components. Parametric component that used such as linear estimator and nonparametric component by using a Fourier series estimator. Semiparametric regression approach that use Fourier series, have an advantages which is can resolve oscillation data pattern. This study compares the three Fourier series estimators such as sine, cosine, and combination between cosine and sine or complete estimator for longitudinal data. Longitudinal data can explain more complete information than cross section data or time series data. The purpose of this study is to introduce another Fourier series for the application of electricity consumption in Madura island. The results of this study indicated the optimal model in predicting electricity consumption in Madura island. The best estimator is the Fourier series estimator with the smallest Generalized Cross Validation (GCV) and Mean Square Error (MSE), and the biggest determination coefficient values by considering the parsimony of the model.

1. Introduction

Regression analysis is one of the tools often used in statistical science to determine the pattern of relationship between two or more variables. In regression analysis the relationship pattern between the response variable and the predictor variable is not always the parametric pattern used when the form of the function is known based on the theory and past experience, such as linear, quadratic, cubic, exponential and others [1]. There are many cases where the pattern of relationship between the response variable and the predictor variable is nonparametric pattern, which is used if the shape of the regression function is assumed to be unknown [2]. Several studies on nonparametric regression with spline estimators have been discussed in [3,4], local linear estimators by [5], kernels by [6,7] and the Fourier series by [8,9,10].

Semiparametric regression is a combination of parametric regression and nonparametric regression [6]. The combinationin this case is intended that the semiparametric regression includes both parametric regression and nonparametric regressionmodel. Several studies on semiparametric regression include spline in [11], local linear estimators [12] and Fourier series estimators in [4,13,14]. One of the advantages of the semiparametric regression approach using the Furier series is that it is able to overcome data that have trigonometric distribution, in this case sine and cosine. Data patterns that are consist of the fourier series approach are repetitive data patterns that are repetitions of different predictor variables [15]

Longitudinal data in its development are widely used for regression analysis. The study showed that longitudinal data is more difficult than cross section data which known as very commonly use in regression analysis. Longitudinal have several advantage such as it is reliable in evaluating the dynamics change and it is able to provide more complete information. Beside that it can determine changes occur in individuals, did not need a lot of subject and the last advantages is more efficient assessment in each observation [16]. Some researchers use longitudinal data with Spline Trancated estimators have been discussed in [17], Fourier series estimators [9].

Electricity is inseparable part of human life. Electricity has become a primary need for human life. electricity usage has been in various sectors of human life including educational activities, offices, the economy, households, and etcetera. Deficiancy outage within a few hours will influence in state losses and will also have an impact on the economy, education, social, and even psychological. Like other energy sources, electricity also has a capacity and limited capacity. Electricity usage is the maximum electrical energy usage load that is recorded based on daily, weekly and even annual time

2. Three forms of fourier series estimator in semiparametric regression for longitudinal data

The semiparametric regression model is a modeling approach that was initiated by Engle [18] and over the last few decades many have been developed both in theory and application. The semiparametric regression models for longitudinal data are as follows:

$$y_{ij} = \beta_{0i} + \sum_{p=1}^{P} \beta_{pi}\left(x_{pij}\right) + \sum_{l=1}^{r} g_l\left(t_{lij}\right) + \varepsilon_{ij} \text{ with } i = 1,2,\ldots,n\,,\ j = 1,2,\ldots,m \tag{1}$$

in equation (1) where y_{ij} is the response variable i, observation j, the parametric component is approximated by linear functions with predictor variables x to p and nonparametric components are approximated to fourier series with predictor variables t to l. is rendom error, where. Fourier series estimators are trigonometric polynomials that have high flexibility, so they can adapt effectively to the nature of local data. Fourier series estimators are used to describe both curves, namely sine and cosine waves

Definition 1

If given $g(t)$ is a function that can be integrated and differentiated at intervals $[a, a+L]$, then the representation of the Fourier series at that interval related to $g(t)$ containing the trigonometric components of sines and cosines is as follows:

$$g(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n \cos k^* t + b_n \sin k * t\right) \tag{2}$$

with $k^* \approx \frac{n}{L}: n = 1,2,\ldots$ As for the Fourier coefficient, it is determined by the following formula:

$$a_0 = \frac{1}{L}\int_a^{a+2L} g(t)dt\ ,\ a_n = \frac{1}{L}\int_a^{a+2L} g(t)\cos k^* tdt\ ,\ b_n = \frac{1}{L}\int_a^{a+2L} g(t)\sin k^* tdt \text{ [15]}$$

Definition 2

If $g(t)$ *is an odd function, or if* $g(-t)=-g(t)$, *then the Fourier coefficient* $a_n=0$. *Those the Fourier series is called the Fourier sine series. If* $g(t)$ *can be integrated in the interval* $[0,L]$, *then the Fourier sine series is as follows*

$$g(t)=\frac{a_0}{2}+\sum_{n=1}^{\infty}\left(b_n \sin k^* t\right) \tag{3}$$

with $k^*\approx\frac{n}{L}: n=1,2,\ldots$ *As for the Fourier coefficient, it is determined by the following formula:*

$$a_0=\frac{2}{L}\int_a^L g(t)\,dt\ ,\ b_n=\frac{2}{L}\int_a^L g(t)\sin k^* t\,dt$$ [15]

Definition 3

If the $g(t)$ *is function, or if* $g(-t)=g(t)$, *then the Fourier coefficient* $b_n=0$. *Those the Fourier series is called the Fourier cosine series. If* $g(t)$ *can be integrated in the interval* $[0,L]$, *then the Fourier cosine series is as follows:*

$$g(t)=\frac{a_0}{2}+\sum_{n=1}^{\infty}\left(a_n \cos k^* t\right) \tag{4}$$

with $k^*\approx\frac{n}{L}: n=1,2,\ldots$ *As for the Fourier coefficient, it is determined by the following formula:*

$$a_0=\frac{2}{L}\int_a^L g(t)\,dt\ ,\ a_n=\frac{2}{L}\int_a^L g(t)\cos k^* t\,dt$$ [15]

Fourier series semiparametric regression models with sine and cosine bases are given for the longitudinal data in the following equation:

$$y_{ij}=\beta_{0i}+\sum_{p=1}^{P}\beta_{pi}x_{pij}+\sum_{l=1}^{r}(\gamma_{li}t_{lij}+\frac{a_{0li}}{2}+\sum_{k=1}^{K}(a_{kli}\cos kt_{lij}+b_{kli}\sin kt_{lij}))+\varepsilon_{ij} \tag{5}$$

Fourier series estimators in semiparametric regression with sine and cosine bases for longitudinal data are as follows:

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

3. The Criteria for Goodness of Fit

The goodness indicator model of semiparametric regression by Fourier series in longitudinal data can be seen asfollow:

Mean Square Error (MSE) dan*Generalized Cross Validation* (GCV)

Mean Square Error (MSE) is the estimated value of the error variance determined by the following equation:

$$MSE[k_1,k_2,\ldots,k_n]=(nm)^{-1}y^T(I-\mathbf{D}[k_1,k_2,\ldots,k_n])^T(I-\mathbf{D}[k_1,k_2,\ldots,k_n])y \tag{7}$$

The model is considered good if the MSE value is minimum.

Aside from the minimum MSE seen, GCV indicators are also very influential for the best models. GCV values are expressed in the following equation

$$GCV(k_1,k_2,\ldots,k_n)=\frac{MSE(k_1,k_2,\ldots,k_n)}{((nm)^{(-1)}trace(I-\mathbf{D}[k_1,k_2,\ldots,k_n]))^2} \tag{8}$$

One of the criteria used in the selection of the best model is to use the coefficient of determination R^2. The coefficient of determination R^2 is a quantity that describes the percentage of variation in the response variable explained by the predictor variable, given the formula of the coefficient of determination as follows:

$$\mathbf{R}^2 = \frac{(\hat{\mathbf{y}} - \bar{\mathbf{y}})^{\mathrm{T}} (\hat{\mathbf{y}} - \bar{\mathbf{y}})}{(\mathbf{y} - \bar{\mathbf{y}})^{\mathrm{T}} (\mathbf{y} - \bar{\mathbf{y}})} \tag{9}$$

4. The Data and Procedure

The procedure in analyzing data related to electricity usage in Maduraisland by using semiparametric regression for longitudinal data based on three Fourier series estimator forms as follows:

a. Literature study about the usage of electricity usage at Madura island, and its relationship to the predictor variables.
b. The variable analysed by statistic descriptive approach based on minimum, average and maximum values.
c. Maintain the general form of three Fourier series estimators including cosines, sines or both of them.
d. Set the GCV value for oscillation parameters in datatraining.
e. Chose the smallest GCV value, after that select the MSE and determination coefficient.
f. Compare three Fourier series estimators that will be usein Madura island electricity.
g. Determine the smallest GCV value, MSE value and the largest determinationcoefficient value as the best model.

5. Result and Discussion

Theoritrical Result Based on WLS Optimation

Semiparametric regression is a combination of parametric regression and nonparametric regression. Parametric regression in this case is approximated by a linear function while nonparametric regression is approximated by the Fourier series, namely cosines, sines and combined sines and cosines. The smallest square of the regression curve error which is approached by WLS.The semiparametric regression equation for longitudinal data is as in equation (1). Equation (1) can be written in interesting form as follows:

$$\mathbf{Y} = \mathbf{X\beta} + \mathbf{T\eta} + \boldsymbol{\varepsilon} \tag{10}$$

for the vector component y, x, β, ε is same for each Fourier series, but the difference is $T_{\sin}, T_{\cos}, T_{\sin\cos}$ and $\eta_{\sin}, \eta_{\cos}, \eta_{\sin\cos}$. Then the matrix form containing nonparametric components based on sinus, cosinus and combination sinus and cosinus as in equation (5). so, $T_{\sin}, T_{\cos}, T_{\sin\cos}$ and $\eta_{\sin}, \eta_{\cos}, \eta_{\sin\cos}$ Such as follows:

Then the $T_{\sin}$ matrix and the parameter $\eta_{\sin}$ will be given :

$$T_{\sin} = \begin{bmatrix} t_{111} & 1 & \sin t_{111} & \cdots & \sin kt_{111} & \cdots & t_{l11} & 1 & \sin t_{l11} & \cdots & \sin kt_{l11} & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ t_{112} & 1 & \sin t_{112} & \cdots & \sin kt_{112} & \cdots & t_{l12} & 1 & \sin t_{l12} & \cdots & \sin kt_{l12} & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ t_{11m} & 1 & \sin t_{11m} & \cdots & \sin kt_{11m} & \cdots & t_{l1m} & 1 & \sin t_{l1m} & \cdots & \sin kt_{l1m} & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & t_{1n1} & 1 & \sin t_{1n1} & \cdots & \sin kt_{1n1} & \cdots & t_{ln1} & 1 & \sin t_{ln1} & \cdots & \sin kt_{ln1} \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & t_{1n2} & 1 & \sin t_{1n2} & \cdots & \sin kt_{1n2} & \cdots & t_{ln1} & 1 & \sin t_{ln2} & \cdots & \sin kt_{ln2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & t_{1nm} & 1 & \sin t_{1nm} & \cdots & \sin kt_{1nm} & \cdots & t_{ln1} & 1 & \sin t_{lnm} & \cdots & \sin kt_{lnm} \end{bmatrix}$$

$$\eta_{\sin} = \left(\gamma_{11,} \frac{a_{011}}{2}, b_{111}, ..., b_{k11}, ..., \gamma_{l1}, \frac{a_{0l1}}{2}, b_{1l1}, ..., b_{kl1}, ..., \gamma_{1n,} \frac{a_{01n}}{2}, b_{1n1}, ..., b_{k1n}, ..., \gamma_{\ln}, \frac{a_{0\ln}}{2}, b_{1\ln}, ..., b_{k\ln} \right)^T$$

$T_{\cos}$ matrix and the parameter $\eta_{\cos}$ will be given

$$T_{cos} = \begin{bmatrix} t_{111} & 1 & \cos t_{111} & \cdots & \cos kt_{111} & \cdots & t_{l11} & 1 & \cos t_{l11} & \cdots & \cos kt_{l11} & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ t_{112} & 1 & \cos t_{112} & \cdots & \cos kt_{112} & \cdots & t_{l12} & 1 & \cos t_{l12} & \cdots & \cos kt_{l12} & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ t_{11m} & 1 & \cos t_{11m} & \cdots & \cos kt_{11m} & \cdots & t_{l1m} & 1 & \cos t_{l1m} & \cdots & \cos kt_{l1m} & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & t_{1n1} & 1 & \cos t_{1n1} & \cdots & \cos kt_{1n1} & \cdots & t_{ln1} & 1 & \cos t_{ln1} & \cdots & \cos kt_{ln1} \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & t_{1n2} & 1 & \cos t_{1n2} & \cdots & \cos kt_{1n2} & \cdots & t_{ln1} & 1 & \cos t_{ln2} & \cdots & \cos kt_{ln2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & t_{1nm} & 1 & \cos t_{1nm} & \cdots & \cos kt_{1nm} & \cdots & t_{ln1} & 1 & \cos t_{lnm} & \cdots & \cos kt_{lnm} \end{bmatrix}$$

$$\eta_{\cos} = \left(\gamma_{11,} \frac{a_{011}}{2}, a_{111}, ..., a_{k11}, ..., \gamma_{l1}, \frac{a_{0l1}}{2}, a_{1l1}, ..., a_{kl1}, ..., \gamma_{1n,} \frac{a_{01n}}{2}, a_{1n1}, ..., a_{k1n}, ..., \gamma_{\ln}, \frac{a_{0\ln}}{2}, a_{1\ln}, ..., a_{k\ln} \right)^T$$

$T_{\sin\cos}$ matrix and the parameter $\eta_{\sin\cos}$ will be given

$$T_{\sin\cos} = \begin{pmatrix} T_1 & 0 & \cdots & 0 \\ 0 & T_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & T_l \end{pmatrix}$$

$$\eta_{\sin\cos} = \left(\eta_1^T, \eta_2^T, ..., \eta_l^T \right)^T$$

With

$$T_1 = \begin{pmatrix} t_{111} & 1 & \cos t_{111} & \cdots & \cos kt_{111} & \sin t_{111} & \cdots & \sin kt_{111} \\ t_{112} & 1 & \cos t_{112} & \cdots & \cos kt_{112} & \sin t_{112} & \cdots & \sin kt_{112} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ t_{11m} & 1 & \cos t_{11m} & \cdots & \cos kt_{11m} & \sin t_{11m} & \cdots & \sin kt_{11m} \end{pmatrix}$$

$$\eta_1 = \left(\gamma_{11,} \frac{a_{011}}{2}, a_{111}, ..., a_{k11}, b_{111}, ..., b_{k11} \right)^T$$

$$T_l = \begin{pmatrix} t_{ln1} & 1 & \cos t_{ln1} & \cdots & \cos kt_{ln1} & \sin t_{ln1} & \cdots & \sin kt_{ln1} \\ t_{ln2} & 1 & \cos t_{ln2} & \cdots & \cos kt_{ln2} & \sin t_{ln2} & \cdots & \sin kt_{ln2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ t_{lnm} & 1 & \cos t_{lnm} & \cdots & \cos kt_{lnm} & \sin t_{lnm} & \cdots & \sin kt_{lnm} \end{pmatrix}$$

$$\eta_l = \left(\gamma_{\ln,} \frac{a_{0\ln}}{2}, a_{1\ln}, ..., a_{k\ln}, b_{1\ln}, ..., b_{k\ln} \right)^T$$

By using the WLS method the error is minimized through the following equation:

$$\mathbf{min}(\boldsymbol{\beta}, \boldsymbol{\eta}) = \mathbf{min}\,\boldsymbol{\varepsilon}^{\mathbf{T}}\mathbf{w}\boldsymbol{\varepsilon} = \mathbf{min}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{T}\boldsymbol{\eta})^{\mathbf{T}}\, \mathbf{w}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{T}\boldsymbol{\eta}) \tag{11}$$

Journal of Physics: Conference Series **1538** (2020) 012058 doi:10.1088/1742-6596/1538/1/012058

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

To get the estimator of the parameter $\boldsymbol{\beta}$ and parameter $\boldsymbol{\eta}$ by doing individual derivatives $\mathbf{R}(\boldsymbol{\beta},\boldsymbol{\eta})$ against β and η by doing the translation in equation (11), the final result becomes

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\left\{\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{y} - \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{T}\hat{\boldsymbol{\eta}}\right\} \quad (12)$$

and

$$\hat{\boldsymbol{\eta}} = \left(\mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{T}\right)^{-1}\left\{\mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{y} - \mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}}\right\} \quad (13)$$

Estimators in equations (12) and (13) arefreeparameter so an estimator that is free from parameters with mutual substitution must be sought. To get free parameters, substitute equation (12) for equation (13). So the final result becomes

$$\hat{\boldsymbol{\beta}} = \mathbf{M}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\left\{\mathbf{X}^{\mathrm{T}} - \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\right\} \quad (14)$$

$$\hat{\boldsymbol{\beta}} = \mathbf{B}(\mathbf{K})\mathbf{y}$$

and

$$\hat{\boldsymbol{\eta}} = \mathbf{N}\left(\mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{T}\right)^{-1}\left\{\mathbf{T}^{\mathrm{T}} - \mathbf{T}^{\mathrm{T}}\mathbf{W}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right\}\mathbf{W}\mathbf{y} \quad (15)$$

$$\hat{\boldsymbol{\eta}} = \mathbf{C}(\mathbf{K})\mathbf{y}$$

After getting the estimator for parametric and nonparametric components, the next determinant is the semiparametric regression model estimator by using the Fourier Series approach

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

was obtained

$\hat{\mathbf{y}} = \mathbf{D}(\mathbf{K})\mathbf{y}$ with $\mathbf{D}(\mathbf{K}) = \mathbf{X}\mathbf{B}(\mathbf{K}) + \mathbf{T}\mathbf{C}(\mathbf{K})$

The Application result

The application of a fourier series semiparametric regression estimator for longitudinal data is electricity consumption on the island of Madura. Focus of this study is to investigate the pattern of the relationship between electricity usage with variables that are considered influential. Some variables that are considered influential include observation time, extinction time, loss power, temperature and humidity.an example of the characteristics of electricity usage in Madura island is Sumenep city.

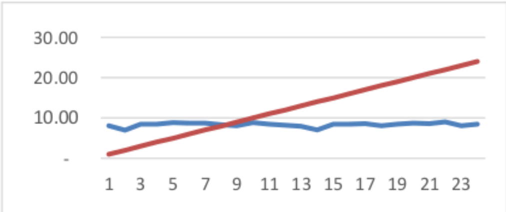

Figure 1. The use of electricity to time of observation in the year 2016 - 2017 in Sumenep

In Figure 1 shows the relationship between electricity consumption and the observation time at Sumenep forming a regular pattern or forming a linear pattern.

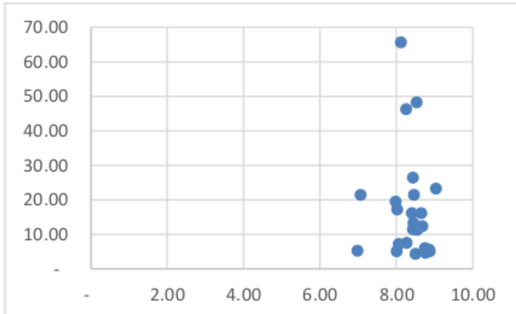

Figure 2. The use of electricity to extinction in the year 2016-2017 in Sumenep

In Figure 2 the relationship between electricity usage and the duration of blackouts tends to form irregular pattern, the longest blackouts in Sumenep is 65 hours 62 minutes, then the lowest extinguishing is 4 hours 35 minutes.

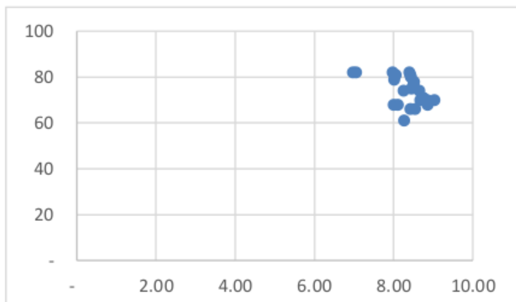

Figure 3. The use of electricity to humidity in the year 2016 - 2017 in Sumenep

Figure 3 shows the relationship between electricity usage and humidity in Sumenep forming irregular patterns, so that the greatest humidity in Sumenep is 82 RH while the lost power is 61 RH

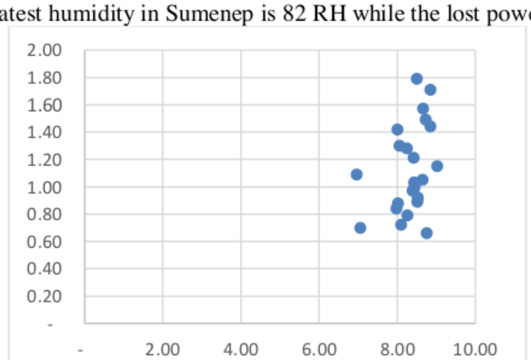

Figure 4. Percentage of electricity usage to loss power in the year 2016 - 2017 in Sumenep

Figure 4 shows the relationship between the percentage of electricity consumption and the percentage of power loss in Sumenep forming an irregular pattern.The greatest power loss in Sumenep is 1.79 percent and lowest power loss is 0.66 percent.

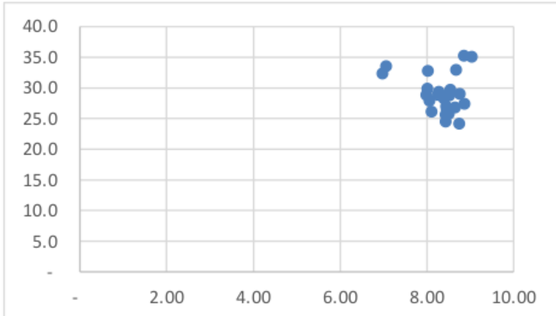

Figure 5. The use of electricity to temperature in the year 2016 - 2017 in Sumenep

According to figure the relationship between electricity consumption and temperature in Sumenep forming irregular patterns so that the highest temperature in Sumenep is 35.3°C, while the lost power is 24.2°C.

The results of the optimal GCV value that calculated to R software using training data are presented in the following table:

Table 1. GCV value based on Fourier series that include cosines and sines estimator.

k	GCV Value
6	3.259924
7	0.091499
8	2.241445

Table 2. GCV value based on Fourier sines series estimator.

k	GCV Value
13	2.955666
14	0.113803
15	1.092435

Table 3. GCV value based on Fourier cosines series estimator.

k	GCV Value
32	1.166078
33	0.828635
34	1.493572

Based on Table1. the minimum GCV value is 0.091499 with k equal to 7 is chosen. Based on optimal osilation parameter value (k) as equal as 7, Based on the results of calculations using R software, the parameter values in the model can be written as follows:

$$\hat{y}_{1j} = 6.069 - 4.64x_{1j1} + 9.299 + 1.369t_{1j1} - 8.3157\cos t_{1j1}$$
$$-2.195\cos 2t_{1j1} + \ldots + 2.1844\sin 6t_{1j4} + 1.383\sin 7t_{1j4}$$
$$\hat{y}_{2j} = 12.409 - 10.069x_{2j1} + 6.5049 + 6.246t_{2j1} - 5.7119\cos t_{2j1}$$
$$-4.8311\cos 2t_{2j1} + \ldots - 1.7806\sin 6t_{2j4} + 6.410\sin 7t_{2j4}$$
$$\hat{y}_{3j} = 4.922 - 363x_{3j1} + 4.992 + 2.0706t_{3j1} - 1.525\cos t_{3j1}$$
$$-3.846\cos 2t_{3j1} + \ldots + 2.174\sin 6t_{3j4} + 4.971\sin 7t_{3j4}$$

$$\hat{y}_{4j} = 4.768 - 3.71x_{4j1} - 4.941 + 2.4351t_{4j1} + 4.0792\cos t_{4j1}$$
$$+4.163\cos 2t_{4j1} + ... - 2.220\sin 6t_{4j4} - 2.4845\sin 7t_{4j4}$$

The best estimator based on Fourier series is longitudinal data which have sine and cosine in semiparametric regression. It will be the simplest estimators form when Fourier sine compared with Fourier cosine series. For estimators based on Fourier sines and cosines, optimal oscillation parameters can be reached when $k$equal to 7. For Fourier sine series estimator, based on Table 2, optimal oscillation parameters can be reached when$k$ equals 14. For Fourier series estimators that include cosines, based on Table 1 , the optimal oscillation parameter can be reached when $k$equals to 33

6. Conclusion

The usage of electricity in Madura island depends on the capacity of the existing power. To make plans and targets, the addition of power on the island of Madura is modeled based on the best semiparametric regession for longitudinal data based on the Fourier series estimator. The k parameter can be selected based on the minimum GCV value. the selection of minimum GCV has an effect on the small MSE value, so the model can be used according to application. The government must manage the electricity that happened at Madura island. One way that can be doneby government is by maintaining the stability of electricity supply in each regency regulating.

Acknowledgements

The author would like to thank all those who have helped with the completion of this paper and specially appreciate Airlangga University for funding this publication.

References

[1] Draper, Norman, Smith, Harry 1992 Analisis Regresi Terapan (Edisi Kedua) Jakarta: PT. Gramedia Pustaka
[2] Eubank R L 1999 Spline Smoothing and Nonparametric Regression 2nd Edition, Marcel Dekker, New York.
[3] Wahba G 1990 Spline Model for Observational Data. SIAM, XII, Philadelphia.
[4] Chamidah N, Lestari B 2016 Spline estimator in homoscedastic multiresponse nonparametric regression model in case of unbalanced number of observations, *Far East Journal of Mathematical Sciences*, 100 (9), pp. 1433-145.
[5] Chamidah N, Tjahjono E, Fadilah A R, Lestari B 2018 Standard Growth Charts for Weight of Children in East Java Using Local Linear Estimator, *Journal of Physics: Conference Series*, **1097** (1), 012092.
[6] Hardle W 1990 Applied Nonparametric Regression, Cambridge University Press, New York.
[7] Chamidah N, Saifudin T 2013 Estimation of children growth curve based on kernel smoothing in multi-response nonparametric regression, *Applied Mathematical Sciences*, **7** (37-40), pp. 1839-1847.
[8] Bilodeau M 1992 Fourier Smoother and Additive Models, *The Canadian Journal of Statistics*, Vol. **3**, hal. 257-259.
[9] Mardianto M F F, Kartiko S H, Utami H 2019 Prediction the Number of Students in Indonesia who Study in Tutoring Agency and Their Motivations based on Fourier Series Estimator and Structural Equation Modelling. *International Journal of Innovation, Creativity and change*. Volume 3, issue 3. Science, applied, applied science, teaching and education.
[10] M Fariz Fadillah Mardianto, Sri Haryatmi Kartiko, Herni Utami. (2019). Forecasting Trend-Seasonal Data Using Nonparametric Regression with Kernel and Fourier Series Approach. *Proceedings of the Third International Conference on Computing, Mathematics and Statistics (iCMS2017)* pp 343-349

[11] Chamidah N, Kurniawan A, Zaman B, Muniroh L 2018 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. pp 295 – 307.

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

[13] Pane R, Budiantara I N, Zain I, Otok B W 2013 Parametric and Non- parametric Estimators in Fourier Series Semiparametric Regression and Their Characteristics. *Applied Mathematical Sciences*, **8**(102), 5053–5064.

[14] Mardianto M F F, Eko Tjahjono, Marisa Rifada 2019 Semiparametric regression based on three forms of trigonometric function in fourier series estimator. *Journal of Physics: Conference Series*.

[15] Suslov S K 2003 An Introduction to Basic Fourier Series (Arizona: *Springer-Science*)

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

[17] Islamiyati A, Fatmawati, Chamidah N 2018 Estimation of Covariance Matrix on Bi-Response Longitudinal Data Analysis with Penalized Spline Regression, *Journal of Physics: Conference Series*, *979* (1), art. no. 012093, . Cited 1 time.

[18] Engle R F, Grangger C W J, Rice J, Weiss A 1986 Semiparametric Estimates of Relation Between Weather and Electric Sale. *Journal of the American Statistical Association* **81**, 310–320.

Three form fourier series estimator semiparametric regression for longitudinal data

ORIGINALITY REPORT

20%	16%	11%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

10%

★ download.atlantis-press.com

Internet Source

Exclude quotes	On	Exclude matches	< 3%
Exclude bibliography	On		