Theoretical Study of Fourier Series Estimator in Semiparametric Regression for Longitudinal Data Based on Weighted Least Square Optimization

by Kuzairi Kuzairi

Advances in Social Science, Education and Humanities Research, volume 410
1st International Multidisciplinary Conference on Education, Technology, and Engineering (IMCETE 2019)

Theoretical Study of Fourier Series Estimator in Semiparametric Regression for Longitudinal Data Based on Weighted Least Square Optimization

1st Kuzairi
Department of Mathematics, Faculty of Mathematics and Natural Sciences
The Islamic University of Madura
Pamekasan 69317, Indonesia.
Ph.D. Student in Department of Mathematics, Faculty of Sciences and Technology
Airlangga University
Surabaya 60115, Indonesia
kuzairi81@gmail.com

2nd N Chamidah[2]
Department of Mathematics, Faculty of Sciences and Technology
Airlangga University
Surabaya 60115, Indonesia
nur_c@fst.unair.ac.id

3rd I N Budiantara
Department of Statistics, Sepuluh Nopember Institute of Technology
Surabaya 60111, Indonesia.

***Abstract*—Semiparametric regression approach is a combination of two components, namely the parametric regression component and the nonparametric regression component. The data used in this study is longitudinal data. Longitudinal data is data obtained from repeated observations of each subject at different time intervals. This data correlates to the same subject and is independent between different subjects. In this study the parametric component is assumed to be linear and the nonparametric component is approximated by the Fourier Series function. In this study, we determine the estimator for semiparametric regression parameters longitudinal data using Weighted Least Square (WLS). In the semiparametric regression based on Fourier series estimator for longitudinal data, the optimal oscillation parameter *k* will be selected. To get the estimation of model parameters, the WLS optimization is performed and GCV method is used to determine the optimal *k*. After obtaining the optimal oscillation parameters from the minimum GCV, the oscillation parameters are used again in the Fourier series semiparametric regression modeling. The criteria for goodness of the model use R^2 and the value of MSE. The best model is a model that has a high R^2 value and a small MSE value.**

Keywords: Fourier, semiparametric regression, longitudinal data based

I. INTRODUCTION

Regression analysis is one method that often used in Statistics. The purpose of the regression analysis is to determine the pattern of the relationship between the response variable and the predictor variable, in order to estimate the shape of the regression function. [1]. In regression analysis the relationship pattern between the response variable and the predictor variable, is not always parametric patterned such as linear, quadratic, cubic and others. There are several cases where the pattern of relationships between response variables and predictor variables have uncertain pattern, it can be solving with nonparametric regression such as spline [2], local polynomial [1], local linear [3], kernel [4], and Fourier series [5]. Evenly, in some other cases, there are some pattern that have uncertain pattern, and the others have certain pattern like linear. So, the pattern can be solved based on semiparametric regression analysis [6]. Semiparametric regression is a combination of parametric regression and nonparametric regression. Recently, semiparametric regression is not only developing in cross section data, but also longitudinal data.

Longitudinal data is the data that be obtained from repeated observations (repeated measures) for more than one subjects on several individuals (cross-sectional) in a row (time series), with the assumption that observations in the same object are interdependent but observations between one object and the other are mutually independent [7]. Research using longitudinal data is usually more complex and requires greater costs than cross-sectional research, but is more reliable in finding answers about the dynamics of changes that occur in certain objects. Longitudinal data analysis can be done with parametric, nonparametric or semiparametric approaches. However, because the object is observed repeatedly in different periods of time, it causes the curve model of the relationship between variables is not clear, so the approach that can be used to see the effect of time on the response is nonparametric regression.

In this paper, we discuss about semiparametric regression for longitudinal data with linear function as parametric component and Fourier series estimator as nonparametric component. Fourier series is a Mathematical function that often be used in nonparametric regression specially to make prediction for trend – seasonal data pattern [8]. Moreover, the advantage of the Fourier series is that it is able to overcome the data patterns that have oscillation pattern [9]. Research about Fourier series estimators [5], [10], [11], [12]. All of recent study is about cross section data. In this case, we develop Fourier series estimator for longitudinal data, based on previous study that Fourier series estimator has been developed in nonparametric regression for longitudinal data [13]. In this study, the parametric component and the nonparametric component parameters in semiparametric regression can be estimated based on Weighted Least Square (WLS) optimization. This theoretical study in this paper is a fundamental result for the development of Fourier series estimator in semiparametric regression for longitudinal data.

II. THE FOURIER ESTIMATOR IN SEMIPARAMETRIC REGRESSION FOR LONGITUDINAL DATA BASED ON WEIGHTED LAST SQUARE (WLS)

Semiparametric regression for longitudinal data as follows [12]:

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

In equation (1), the semiparametric regression approach used consists of parametric regression components which are approximated by linear functions with predictor variables x as much as p , and nonparametric regression functions with predictor variables t as much as q . Response variable denoted by y . Regression coefficient for parametric component denoted by β . Regression curve for nonparametric component represented with $g_q(t_{qij})$, here j depends on the number of observations. An error random which independent and identically distributed with mean 0, and variance σ^2 denoted by ε [11]. Regression curve in (1) approached by Fourier series estimator. Fourier series is a function of trigonometric polynomials which has a high degree of flexibility. Fourier series is a curve that shows the sines and cosines functions. If given $g(t)$ is a function that can be integrated and differentiable in intervals $[a, a+2L]$, then the Fourier series representation of the interval associated with $f(t)$ containing the cosines Fourier based on Bilodeau, 1992 [5], with definition as follows :

Definition 1

if $g(t)$ the function is even, or if $g(-t) = g(t)$ then the Fourier coefficient. $b_n = 0$ Thus the Fourier series is called the Fourier cosine series. If $g(t)$ it can be integrated at intervals [0, L], then the Fourier cosine series is as follows:

$$g(t) = \frac{a_0}{2} + \sum_{k=1}^{K} a_k \cos k^* t \quad (2)$$

with $k^* \approx \frac{n\pi}{L} : n = 1, 2, 3, \ldots$ The Fourier coefficient is determined by the formula as follows:

$$a_0 = \frac{2}{L}\int_0^L g(t)\,dt; a_n = \frac{2}{L}\int_0^L g(t)\cos k^* t\,dt$$

The Fourier series in equation (2) only accommodate seasonal pattern. For accommodating trend and seasonal pattern, [5], appended and modified equation (2) with linear function as follows:

$$g(t_i) = bt_i \frac{a_0}{2} + \sum_{k=1}^{K} a_k \cos k^* t_i \quad (3)$$

One of estimator used in semparametric regression is Fourier series that be proposed by Bilodeau on 1992 at the first time [5]. In this case, the Fourier series estimator is applied for longitudinal data. Longitudinal data is the data structure that contains elements of cross section and time series data. The advantage of using longitudinal data, can to find out the changes that occur in subjects, because the observations are repeated for each subject [7]. Given data $\left(x_1, x_2, \ldots, x_p, t_1, t_2, \ldots, t_q, \text{and } y_i\right)$, relationship between x_p and y_i assumed to follow the parametric regression model, and the relationship between t_q and y_i assumed to follow a nonparametric regression model, so the semiparametric regression estimator Forier series for longitudinal data as follows :

$$y_{ij} = \beta_{0i} + \sum_{p=1}^{P} \beta_{pi} x_{pij} + \sum_{q=1}^{Q} (b_{qi} t_{qij} + \frac{a_{0qi}}{2} + \sum_{k=1}^{K} (a_{kqi} \cos kt_{qij})) + \varepsilon_{ij} \quad (4)$$

with $\beta_{0i}, \beta_{1i}, \ldots, \beta_{pi}$ are parameters in parametric regression that will be estimated, $b_{1i}, b_{2i}, \ldots, b_{qi}, a_{01i}, a_{02i}, \ldots, a_{0qi}, a_{k1i}, a_{k2i}, \ldots, a_{kqi}$ are parameter in nonparametric regression that the values will be estimated, presents the number of oscillation parameter that be inputted. A random error that independent and identically distributed is denoted by ε_{ij} .

Equation (4) can be formed as matrices equation as follows

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{T\eta} + \mathbf{\varepsilon} \quad (5)$$

with

$$y = \left(y_{11}, y_{12}, \ldots, y_{1m}, \ldots, y_{n1}, y_{n2}, \ldots, y_{nm}\right)^T$$

$$\beta = \left(\beta_{01}, \beta_{11}, \ldots, \beta_{p1}, \ldots, \beta_{0n}, \beta_{1n}, \ldots, \beta_{pn}\right)^T,$$

$$\varepsilon = \left(\varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{1m}, \ldots, \varepsilon_{n1}, \varepsilon_{n2}, \ldots \varepsilon_{nm}\right)^T$$

$$\eta = \left(b_{11}, {a_{011}}/{2}, a_{111}, \ldots, a_{k11}, \ldots, b_{q1}, {a_{0q1}}/{2}, a_{1q1}, \ldots, a_{kq1}, \ldots, b_{1n}, {a_{01n}}/{2}, a_{11n}, \ldots, a_{k1n}, \ldots, b_{qn}, {a_{0qn}}/{2}, a_{1qn}, \ldots, a_{kqn}\right)^T$$

$$X = \begin{pmatrix} 1 & x_{111} & x_{211} & \cdots & x_{p11} & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 1 & x_{112} & x_{212} & \cdots & x_{p12} & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{11m} & x_{21m} & \cdots & x_{p1m} & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & x_{1n1} & x_{2n1} & \cdots & x_{pn1} \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & x_{1n2} & x_{2n2} & \cdots & x_{pn2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & x_{1nm} & x_{2nm} & \cdots & x_{pnm} \end{pmatrix}$$

$$T=\begin{bmatrix}
t_{111} & 1 & \cos t_{111} & \cdots & \cos kt_{111} & \cdots & t_{q11} & 1 & \cos t_{q11} & \cdots & \cos kt_{q11} & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
t_{112} & 1 & \cos t_{112} & \cdots & \cos kt_{112} & \cdots & t_{q12} & 1 & \cos t_{q12} & \cdots & \cos kt_{q12} & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\
t_{11m} & 1 & \cos t_{11m} & \cdots & \cos kt_{11m} & \cdots & t_{q1m} & 1 & \cos t_{q1m} & \cdots & \cos kt_{q1m} & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & t_{1n1} & 1 & \cos t_{1n1} & \cdots & \cos kt_{1n1} & \cdots & t_{qn1} & 1 & \cos t_{qn1} & \cdots & \cos kt_{qn1} \\
0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & t_{1n2} & 1 & \cos t_{1n2} & \cdots & \cos kt_{1n2} & \cdots & t_{qn1} & 1 & \cos t_{qn2} & \cdots & \cos kt_{qn2} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & t_{1nm} & 1 & \cos t_{1nm} & \cdots & \cos kt_{1nm} & \cdots & t_{qn1} & 1 & \cos t_{qnm} & \cdots & \cos kt_{qnm}
\end{bmatrix}$$

The estimation of Fourier series the semiparametric regression for longitudinal data model (4) using WLS (weighted Least Square). Theorem 1 presents Fourier series estimator for semiparametric regression for longitudian data as follows:

Theorem 1

If a semiparametric regression equation is given for longitudinal data with a regression curve approximated by the Fourier series as in equation (4), then the Fourier series estimator is as follows:

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

Proof :

Semiparametric regression equations for longitudinal data in equation (4) are given which are ε normally distributed with mean 0 and variance σ^2

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

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

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

So obtained

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

Theorem 2

If the semiparametric regression equation for longitudinal data approximated by the Fourier series estimator is presented in the form of a vector equation as in equation (5) then:

i. Estimators for parametric component parameters that do not contain other parameters are as follows:

$$\begin{aligned}\hat{\boldsymbol{\beta}}&=\mathbf{M}(\mathbf{X}^{T}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{T}\mathbf{W}\mathbf{y}-\mathbf{M}(\mathbf{X}^{T}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{T}\mathbf{W}\mathbf{T}(\mathbf{T}^{T}\mathbf{W}\mathbf{T})^{-1}\mathbf{T}^{T}\mathbf{W}\mathbf{y} \qquad (7)\\ &=\mathbf{M}(\mathbf{X}^{T}\mathbf{W}\mathbf{X})^{-1}\{\mathbf{X}^{T}-\mathbf{X}^{T}\mathbf{W}\mathbf{T}(\mathbf{T}^{T}\mathbf{W}\mathbf{T})^{-1}\mathbf{T}^{T}\}\mathbf{W}\mathbf{y}\end{aligned}$$

ii. Estimators for nonparametric component parameters that do not contain other parameters are as follows:

$$\begin{aligned}\hat{\boldsymbol{\eta}}&=\mathbf{N}(\mathbf{T}^{T}\mathbf{W}\mathbf{T})^{-1}\mathbf{T}^{T}\mathbf{W}\mathbf{y}-\mathbf{N}(\mathbf{T}^{T}\mathbf{W}\mathbf{T})^{-1}\mathbf{T}^{T}\mathbf{W}\mathbf{X}(\mathbf{X}^{T}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{T}\mathbf{W}\mathbf{y} \qquad (10)\\ &=\mathbf{N}(\mathbf{T}^{T}\mathbf{W}\mathbf{T})^{-1}\{\mathbf{T}^{T}-\mathbf{T}^{T}\mathbf{W}\mathbf{X}(\mathbf{X}^{T}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{T}\}\mathbf{W}\mathbf{y}\end{aligned}$$

iii. Estimators for semiparametric regression curves in vector equations are as follows

$$\hat{\mathbf{y}}=\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{T}\hat{\boldsymbol{\eta}} \qquad (11)$$

Proof

Parameter estimation in the Fourier series semiparametric regression model for longitudinal data obtained by optimization of Weighted Least Square (WLS) is done to minimize the goodness of fit of the semiparametric regression model with the Fourier series approach for longitudinal data

$$\min_{f\in C(0,\pi)}[R(f)]=\min_{f\in C(0,\pi)}\boldsymbol{\varepsilon}^{T}\mathbf{W}\boldsymbol{\varepsilon}=\min_{f\in C(0,\pi)}\{(\mathbf{y}-\mathbf{f}(\mathbf{x},\mathbf{t}))^{T}W(\mathbf{y}-\mathbf{f}(\mathbf{x},\mathbf{t}))\} \qquad (12)$$

by elaborating on equation (12), WLS optimization from is given as fallows :

$$R(\boldsymbol{\beta},\boldsymbol{\eta})=\mathbf{y}^{T}\mathbf{W}\mathbf{y}-2\mathbf{y}^{T}\mathbf{W}\mathbf{X}\boldsymbol{\beta}-2\boldsymbol{\eta}^{T}\mathbf{T}^{T}\mathbf{W}\mathbf{y}+2\boldsymbol{\beta}^{T}\mathbf{X}^{T}\mathbf{W}\mathbf{T}\boldsymbol{\eta}+\boldsymbol{\beta}^{T}\mathbf{X}^{T}\mathbf{W}\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\eta}^{T}\mathbf{T}^{T}\mathbf{W}\mathbf{T}\boldsymbol{\eta}$$
(13)

for obtaining estimator from $\boldsymbol{\beta}$, can be determined by doing partial derivatives, $R(\boldsymbol{\beta},\boldsymbol{\eta})$ to $\boldsymbol{\beta}$ with condition $\partial R(\boldsymbol{\beta},\boldsymbol{\eta})/\partial\boldsymbol{\beta}$ equals to $\mathbf{0}$, so it can be resulted as follows :

$$\hat{\boldsymbol{\beta}}=\left(\mathbf{X}^{T}\mathbf{W}\mathbf{X}\right)^{-1}\left\{\mathbf{X}^{T}\mathbf{W}\mathbf{y}-\mathbf{X}^{T}\mathbf{W}\mathbf{T}\hat{\boldsymbol{\eta}}\right\} \qquad (14)$$

With similar step, for obtaining estimator from $\boldsymbol{\eta}$ can be determined by doing partial derivatives $R(\boldsymbol{\beta},\boldsymbol{\eta})$ to $\boldsymbol{\eta}$ with condition $R(\boldsymbol{\beta},\boldsymbol{\eta})$ equal to $\mathbf{0}$. So, it can be resulted as follows :

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

Proof

i. Estimator in equation (14) and (15) are still include parameter. According to Statistical inference theory, this condition will imply to unsatisfied sufficiency criteria. So, in this study, parameter vector estimator for parametric and nonparametric component that be determined so that it is free from parameters. The procedure is substitution method. To get $\hat{\boldsymbol{\beta}}$ that free from parameter substitute equation (15) into equation (14)

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

So, it can be resulted that

$$\begin{aligned}\hat{\boldsymbol{\beta}}&=\mathbf{M}\left(\mathbf{X}^{T}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{T}\mathbf{W}\mathbf{y}-\mathbf{M}\left(\mathbf{X}^{T}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{T}\mathbf{W}\mathbf{T}\left(\mathbf{T}^{T}\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}^{T}\mathbf{W}\mathbf{y}\\ &=\mathbf{M}\left(\mathbf{X}^{T}\mathbf{W}\mathbf{X}\right)^{-1}\left\{\mathbf{X}^{T}-\mathbf{X}^{T}\mathbf{W}\mathbf{T}\left(\mathbf{T}^{T}\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}^{T}\right\}\mathbf{W}\mathbf{y}\\ &=\mathbf{A}(\mathbf{K})\mathbf{y}\end{aligned} \qquad (16)$$

with

$$\mathbf{A}(\mathbf{K})=\mathbf{M}(\mathbf{X}^{T}\mathbf{W}\mathbf{X})^{-1}\{\mathbf{X}^{T}-\mathbf{X}^{T}\mathbf{W}\mathbf{T}(\mathbf{T}^{T}\mathbf{W}\mathbf{T})^{-1}\mathbf{T}^{T}\}\mathbf{W}$$

ii. To get $\hat{\boldsymbol{\eta}}$ that free from parameter substitute equation (14) into equation (15).

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

So, it can be resulted that

$$\begin{aligned}\hat{\boldsymbol{\eta}} &= \mathbf{N}\left(\mathbf{T}^T\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}^T\mathbf{W}\mathbf{y} - \mathbf{N}\left(\mathbf{T}^T\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}^T\mathbf{W}\mathbf{X}\left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y} \\ &= \mathbf{N}\left(\mathbf{T}^T\mathbf{W}\mathbf{T}\right)^{-1}\left\{\mathbf{T}^T - \mathbf{T}^T\mathbf{W}\mathbf{X}\left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^T\right\}\mathbf{W}\mathbf{y} \\ &= \mathbf{B}(\mathbf{K})\mathbf{y}\end{aligned} \qquad (17)$$

with

$$\mathbf{B}(\mathbf{K}) = \mathbf{N}\left(\mathbf{T}^T\mathbf{W}\mathbf{T}\right)^{-1}\left\{\mathbf{T}^T - \mathbf{T}^T\mathbf{W}\mathbf{X}\left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^T\right\}\mathbf{W}$$

iii. After obtaining an estimator for parametric and nonparametric components, then determine the estimator of the semiparametric regression model with the Fourier series for longitudinal data as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{T}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad \varepsilon_i \sim IIN\left(0, \sigma^2\right)$$

$$E(\mathbf{y}) = E(\mathbf{X}\boldsymbol{\beta} + \mathbf{T}\boldsymbol{\eta} + \boldsymbol{\varepsilon})$$

$$E(\mathbf{y}) = E(\mathbf{X}\boldsymbol{\beta} + \mathbf{T}\boldsymbol{\eta}) + E(\boldsymbol{\varepsilon}) \quad \text{so that}$$

$$E(\mathbf{y}) = E(\mathbf{X}\boldsymbol{\beta} + \mathbf{T}\boldsymbol{\eta})$$

$$\begin{aligned}\hat{\mathbf{y}} &= \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{T}\hat{\boldsymbol{\eta}} \\ &= \mathbf{X}\mathbf{A}(\mathbf{K})\mathbf{y} + \mathbf{T}\mathbf{B}(\mathbf{K})\mathbf{y} \\ &= (\mathbf{X}\mathbf{A}(\mathbf{K}) + \mathbf{T}\mathbf{B}(\mathbf{K}))\mathbf{y} \\ &= \mathbf{C}(\mathbf{K})\mathbf{y}\end{aligned} \qquad (18)$$

with $\mathbf{C}(\mathbf{K}) = \mathbf{X}\mathbf{A}(\mathbf{K}) + \mathbf{T}\mathbf{B}(\mathbf{K})$ is a hat matrix

III. The Criteria for Goodness Of Fit

Generalized cross validation (GCV), the selection of optimal bandwidth is very important in regression analysis. choosing a bandwidth that is too large results in the plot estimation results of the model will move away from the initial data plot so that it becomes very smooth (oversmoothing).
by formula

$$GCV(\mathbf{K}) = \frac{MSE(\mathbf{K})}{\left((nm)^{-1} trace\left(I - C(\mathbf{K})\right)\right)^2} \qquad (19)$$

Coefficient of determination (R^2), a value or measure used to measure how well the regression line matches the actual data is called the coefficient of determination (R^2). The value of the coefficient of determination can be obtained by the following formula:

$$R^2 = \frac{\left(\hat{\mathbf{y}} - \bar{\mathbf{y}}\right)^T\left(\hat{\mathbf{y}} - \bar{\mathbf{y}}\right)}{\left(\mathbf{y} - \bar{\mathbf{y}}\right)^T\left(\mathbf{y} - \bar{\mathbf{y}}\right)} \qquad (20)$$

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

$$MSE(\mathbf{K}) = (nm)^{-1} y^T\left(I - C(\mathbf{K})\right)^T W\left(I - C(\mathbf{K})\right) y \qquad (21)$$

The best model that can be used for prediction met the goodness of criteria. The goodness of criteria is the smallest GCV value for an optimal oscillation parameter, the smallest Mean Square Error (MSE) value, and the big of determination coefficient value.

IV. Conclusion

Given the Fourier series semiparametric regression models for longitudinal data in equation (4) are random errors are normally distributed with mean 0 and variance σ^2. Based on the analysis conducted it can be concluded in the estimation of the semiparametric regression model for longitudinal data based on the fourier series estimator the parameter estimation for the parametric component is in equation (7). while the estimated nonparametric component is in equation (10). The k parameter can be selected based on the minimum GCV at an optimal k. The choice of minimum GCV affects the small MSE value and (R^2) is high so that the model can be used according to application

References

[1] Fan, J. and Gijbels, I. (1997). Local Polynomial Modelling and Its Applications. London: Chapman and Hall.

[2] Chamidah, N. and 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.

[3] Chamidah, N., Tjahjono, E., Fadilah, A.R., and Lestari, B., (2018). Standard Growth Charts for Weight of Children in East Java Using Local Linear Estimator, Journal of Physics: Conference Series.

[4] Chamidah, N.and 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.

[5] Bilodeau, M. (1992). Fourier Smoother and Additive Models, The Canadian Journal of Statistics, Vol. 3, hal. 257-259.

[6] Eubank, R.L. (1999). Nonparametric Regression and Spline Smoothing Second Edition. Texas: Departmen of Statistics Southern Methodist Dallas University.

[7] Wu, H. and Zhang, J. T. 2006. Nonparametric Regression Methods for Longitudinal Data Analysis. New York: John Wiley and Sons, Inc.Another reference

[8] Mardianto, M. F. F., Kartiko, S. H., and Utami, H. (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)

[9] Tripena A. and Budiantara I.N. 2006. Fourier Estimator in Nonparametric Regression. International Conference on Natural Sciences and Applied Natural Scienes, Ahmad Dahlan University, Yogjakarta.

[10] Asrini L. J. and Budiantara I. N. 2014. Fourier series semiparametric regression models (case study: the production of lowland rice irrigation in Central Java). ARPN Journal of Engineering and Applied Sciences. 9(9) : 1501-1506.

[11] Pane R., Budiantara I. N., Zain I. and Otok B. W. 2013. Parametric and Nonparametric Estimators in Semiparametric Regression and Their Characteristics Journal of Applied Mathematical Science. 8(102): 5053- 5064.

[12] Mardianto, M. F. F., Tjahjono, E and Rifada, M. (2019). Semiparametric regression based on three forms of trigonometric function in fourier series estimator. Journal of Physics: Conference Series

[13] Mardianto, M. F. F., Kartiko, S. H., and 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.

Theoretical Study of Fourier Series Estimator in Semiparametric Regression for Longitudinal Data Based on Weighted Least Square Optimization

ORIGINALITY REPORT

19%	17%	17%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

8%

★ core.ac.uk

Internet Source

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		