Semiparametric regression based on three forms of trigonometric function in fourier series estimator

M F F Mardianto*, E Tjahjono and M Rifada
Study Program of Statistics, Department of Mathematics, Airlangga University, Surabaya, Indonesia.

*Corresponding author: m.fariz.fadillah.m@fst.unair.ac.id

Abstract. The semiparametric regression is one of the three forms of regression analysis which is made up of parametric and nonparametric. While the parametric is based on linear estimator, this nonparametric component is an innovation. This research proposes all the possible trigonometric basis usually used in Fourier series as nonparametric component estimator, its advantage, which includes its ability to overcome data with oscillation patterns. This study discusses nonparametric regression based on complete and sine Fourier series. Both estimators are developed using the cosine Fourier series concept. The outputs are two estimators which are used for parametric and nonparametric components with the corresponding form in semiparametric regression. In addition, all of these can be applied in real problems, and the best estimator is determined based on the smallest GCV and MSE for an oscillation parameter which gives the highest coefficient of determination for the selected one.

1. Introduction
Regression is a statistical method used to model the relationship between response and predictor variables. Its purpose is to make a prediction based on a function that estimates the data pattern. There are three approaches in regression analysis based on data pattern through mathematical function - parametric, nonparametric, and semiparametric regression [1]. Parametric regression is an analysis in which the data pattern is known and the most popular is the linear regression. However, in a situation where the data pattern is unable to satisfy the assumption test, or in situation where there are many insignificant parameters, one of the ways of solving this problem is through nonparametric regression. It has a high flexibility when it comes to modelling data pattern. There are estimators that are used in nonparametric regression and these include spline [2], kernel [3], local polynomial [4], and Fourier series [5–8].

The semiparametric regression approach combines both the parametric and nonparametric regression. This approach, in general, serves as a solution when some data pattern is linear based on linearity test, while the data pattern of others ones are unknown [9]. Recently, the semiparametric regression has developed theoretically and in application wise. According to Hardle et al [9],
semiparametric regression is a partial model in regression analysis due to the fact that it has both parametric and nonparametric components. In this study, the semiparametric regression approach used was based on linear estimator for parametric regression component and Fourier series for nonparametric regression. The process of combining these two components gives a high flexibility in modelling trend-seasonal data pattern in a cross section data, as well as in time series regression [1]. Also, the trend pattern can be represented using linear estimator and the seasonal pattern can be represented by trigonometric basis using Fourier series with periodic properties. In other words, Fourier series can be used to solve problems with data that has oscillation pattern.

There are numerous studies on semiparametric regression with Fourier series estimator, such as the ones conducted by Bilodeau [5], Asrini [10], Pane et al, [11] and other researchers. However, the Fourier series used in those studies only included cosine series. But in mathematical terms, there are sines basis aside the cosines and the analytical concept to get estimator is the same for both components. This study critically considers the analytical concept to obtain estimator in Fourier series. In addition, as a novelty of this study, the Fourier series used is not only the cosine series, but also the sine. In other words, the Fourier series used include the cosine and sine basis, or complete. Biedermann, et al [7] and Dette, et al [8] reported that the complete Fourier series estimator is rarely used in nonparametric and semiparametric regression studies, although the results model trend and seasonal data pattern with parsimony and always smooth.

2. Fourier series estimator in semiparametric regression based on Bilodeau

Fourier series is a trigonometric polynomial function that often be used in mathematical modeling. Fourier series was first introduced by Joseph Fourier in 1822 for modeling heat equations in metal plates. The Fourier idea is modeling the heat source of a metal plate as a simple sine and cosine superposition [12]. Along with the development of research and flexibility of the Fourier series, currently Fourier ideas are not only used in heat equation modeling, but also modeling in various fields. Recently, Fourier series is used not only in mathematical modeling, but also in statistical modeling such as time series and regression, specially nonparametric and semiparametric regression.

Bilodeau [5] in 1992 proposed cosine Fourier series in regression analysis for smoothing method. Bilodeau’s research give motivation the development of Fourier series in nonparametric regression. The development of nonparametric regression is semiparametric regression that be discussed in this research.

Consider the pairs of data \((x_1, x_2, ..., x_p, t_1, t_2, ..., t_r, y)\), with \(x\) and \(t\) are predictor variable and \(y\) is response variable. The relationship between \(x\) and \(y\) is known to form a pattern, whereas the relationship between \(t\) and \(y\) pattern forms is unknown. Therefore, the relationship between \(x, t\) and \(y\) is assumed to follow a semiparametric regression model. In this study, the semiparametric regression model is assumed \(p\) to be a predictor of parametric components \(x_1, x_2, ..., x_p\) that be estimated by linear function and \(r\) predictors of nonparametric components \(t_1, t_2, ..., t_r\). The relationship between response variables \(y\) follows the semiparametric regression model as follows:

\[
y_i = \alpha_0 + \alpha_1 x_{i1} + ... + \alpha_p x_{ip} + \sum_{j=1}^{r} m(t_{ij}) + \epsilon_i \sim N(0, \sigma^2)
\]

(1)

with \(i = 1, 2, ..., n\) present the number of observation. In equation (1) \(\alpha_0 + \alpha_1 x_{i1} + ... + \alpha_p x_{ip}\) are parametric component and \(m(t_{ij})\) is nonparametric component. In this case \(m(t_{ij})\) is estimated by cosine Fourier series with definition as follows:

**Definition 1.**
If \( m(t) \) the even function, or if \( m(-t) = m(t) \), then the Fourier coefficient for sine basis, \( b_n = 0 \). Thus the Fourier series is called the cosine Fourier series. If \( m(t) \) it can be integrated at intervals \([0, L]\), then the Fourier series of cosine is as follows:

\[
m(t) = \frac{\beta_0}{2} + \sum_{s=1}^{\infty} \beta_s \cos k^* t
\]

with \( k^* \approx \frac{\pi}{L} : s = 1, 2, 3, \ldots \) the Fourier coefficient is determined by the following formula:

\[
\beta_0 = \frac{2}{L} \int_0^L m(t) \, dt; \beta_s = \frac{2}{L} \int_0^L m(t) \cos k^* t \, dt
\]

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

\[
m(t_i) = \frac{\beta_0}{2} + y t_i + \sum_{k=1}^{K} \beta_k \cos k t_i
\]

For \( j = 1, 2, \ldots, r \) predictors equation (3) becomes

\[
m(t_{ij}) = \frac{\beta_{0j}}{2} + y_j t_{ij} + \sum_{k=1}^{K} \beta_{kj} \cos k t_{ij}
\]

So, using Fourier series that be proposed by Bilodeau [5], the Fourier series in semiparametric regression equation is given with substitute equation (4) to equation (1) as follows:

\[
y_i = \alpha_0 + \alpha_1 x_{i1} + \ldots + \alpha_p x_{ip} + \sum_{j=1}^{r} \left( \frac{\beta_{0j}}{2} + y_j t_{ij} + \sum_{k=1}^{K} \beta_{kj} \cos k t_{ij} \right) + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)
\]

with \( \alpha_0, \alpha_1, \ldots, \alpha_p \) are parameters in parametric regression that the values will be estimated, \( \beta_{0j}, \beta_{kj}, y_j \) are parameters in nonparametric regression that the values will be estimated, \( k = 1, 2, \ldots, K \) presents the number of oscillation parameter that be inputted. A random error that independent and identically distributed is denoted by \( \epsilon_i \).

Equation (5) can be formed as matrices equation as follows

\[
y = X\alpha + T\beta + \epsilon
\]

With
\[ y = (y_1, y_2, ..., y_p)^T; \alpha = (\alpha_0, \alpha_1, ..., \alpha_p)^T; \varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_p)^T; \]

\[ X = \begin{bmatrix}
1 & x_{11} & x_{12} & \cdots & x_{1p} \\
1 & x_{21} & x_{22} & \cdots & x_{2p} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{n1} & x_{n2} & \cdots & x_{np}
\end{bmatrix} \]  \hspace{1cm} (7)

The differences matrix and vector between recent estimator and proposed estimators are presented as follows:

\[ \beta = \left( \frac{\delta \alpha_0}{2}, \beta_1 y_1, ..., \beta_l y_l, ..., \beta_r \right)^T; \]

\[ T = \begin{bmatrix}
1 & t_{11} \cos t_{11} & \cdots & \cos K t_{11} & 1 & t_{1r} \cos t_{1r} & \cdots & \cos K t_{1r} \\
1 & t_{21} \cos t_{21} & \cdots & \cos K t_{21} & 1 & t_{2r} \cos t_{2r} & \cdots & \cos K t_{2r} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & t_{n1} \cos t_{n1} & \cdots & \cos K t_{n1} & 1 & t_{nr} \cos t_{nr} & \cdots & \cos K t_{nr}
\end{bmatrix} \]  \hspace{1cm} (8)

In this case there are two estimators for vectors that include parameters based on Ordinary Least Square (OLS) optimization result. Theorem 2 presents the Fourier series estimator for semiparametric regression as follows:

**Theorem 2.**

Consider semiparametric regression equation with curve regression appoached by Fourier series that be presented in equation (5). The estimator for approaching curve regression is given as follows:

\[ \hat{y}_i = \hat{a}_0 + \hat{a}_1 x_{i1} + \cdots + \hat{a}_p x_{ip} + \sum_{j=1}^{r} \left( \frac{\hat{\beta}_0 j}{2} + \hat{\gamma}_j t_{ij} \right) + \sum_{k=1}^{K} \hat{\beta}_k \cos K t_{ij} \]  \hspace{1cm} (9)

with the values of \( \hat{a}_0, \hat{a}_1, ..., \hat{a}_p, \hat{\beta}_0 j, \hat{\gamma}_j, \) and \( \hat{\beta}_k \) can be determined based on OLS result. 

3. **An alternative fourier series estimator in semiparametric regression**

In this section is proposed an alternative Fourier series estimator in semiparametric regression that be approached by Fourier series [7,8] and sines Fourier series that be proposed by authors. So, there are three comparisons for Fourier series estimator in semiparametric regression as a consideration to determine the best estimator based on goodness of model criteria that be discussed in the section 4.

Consider the pairs of data \( (x_1, x_2, ..., x_p, t_1, t_2, ..., t_r, y') \), with \( x \) and \( t \) are predictor variable and \( y \) is response variable. The relationship between \( x \) and \( y \) is known to form a pattern, whereas the relationship between \( t \) and \( y \) pattern forms is unknown. The general form from semiparametric regression is given by equation (1). In this case \( m(t_{ij}) \) is estimated by Fourier series and sine Fourier series with definition as follows:

**Definition 3.**

If \( m(t) \) is a function which can be integrated and differentiable at the interval \( [a, a + 2L] \), then the Fourier series representation at that interval relating to \( m(t) \) which contains the components of trigonometric sine and cosine is as follows:
\[ m(t) = \frac{\beta_0}{2} + \sum_{n=1}^{\infty} (\beta_n \cos k^* t + \delta_n \sin k^* t) \]  

(10)

with \( k^* \approx \frac{ \pi t }{ L } \): \( n = 1, 2, 3, \ldots \) the Fourier coefficient is determined by the following formula:

\[
\beta_0 = \frac{1}{L} \int_{a}^{a+2L} m(t) \, dt; \quad \beta_n = \frac{1}{L} \int_{a}^{a+2L} m(t) \cos k^* t \, dt; \quad \delta_n = \frac{1}{L} \int_{a}^{a+2L} m(t) \sin k^* t \, dt
\]

The Fourier series in equation (10) only accommodate seasonal pattern. For accommodating trend and seasonal pattern, it can be appended and modified with linear function, so equation (10) becomes as follows:

\[ m(t_i) = \frac{\beta_0}{2} + \gamma t_i + \sum_{k=1}^{K} (\beta_k \cos k t_i + \delta_k \sin k t_i) \]

(11)

For \( j = 1, 2, \ldots, r \) predictors equation (11) becomes

\[ m(t_{ij}) = \frac{\beta_{0j}}{2} + \gamma_j t_{ij} + \sum_{k=1}^{K} (\beta_{kj} \cos k t_{ij} + \delta_{kj} \sin k t_{ij}) \]

(12)

So, the Fourier series in semiparametric regression equation is given with substitute equation (12) to equation (1) as follows:

\[ y_i = \alpha_0 + \sum_{l=1}^{p} \alpha_l x_{il} + \sum_{j=1}^{r} \left( \frac{\beta_{0j}}{2} + \gamma_j t_{ij} + \sum_{k=1}^{K} (\beta_{kj} \cos k t_{ij} + \delta_{kj} \sin k t_{ij}) \right) + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \]

(13)

with \( \alpha_0 \) and \( \alpha_l \) are parameters in parametric regression that the values will be estimated, \( \beta_{0j}, \beta_{kj}, \gamma_j \), and \( \delta_{kj} \) are parameters in nonparametric regression that the values will be estimated. By the same analog for sine Fourier series, the semiparametric regression is given as follows:

\[ y_i = \alpha_0 + \alpha_1 x_{i1} + \ldots + \alpha_p x_{ip} + \sum_{j=1}^{r} \left( \frac{\beta_{0j}}{2} + \gamma_j t_{ij} + \sum_{k=1}^{K} \delta_{kj} \sin k t_{ij} \right) \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \]

(14)

Semiparametric regression in equation (13) and (14) can be formed as matrices equation that be presented in equation (6) with the components of vector \( y, \alpha \), and \( \varepsilon \), also matrix \( X \) is presented in equation (7). However, the structure of vector and matrix for nonparametric components are given in equation (15) for complete Fourier series and equation (16) for sine Fourier series.
\[
\beta = \left( \frac{\beta_1}{2}, \gamma_1, \beta_{11}, \ldots, \beta_{k1}, \delta_{11}, \ldots, \delta_{k1}, \ldots, \frac{\beta_r}{2}, \gamma_r, \beta_{1r}, \ldots, \beta_{kr}, \delta_{1r}, \ldots, \delta_{kr} \right)^T;
\]
\[
T = [T_1 \quad T_2 \quad \ldots \quad T_r]
\]
with
\[
T_r = \begin{bmatrix}
1 & t_{1r} & \cos t_{1r} & \cdots & \cos K t_{1r} & \sin t_{1r} & \cdots & \sin K t_{1r} \\
1 & t_{2r} & \cos t_{2r} & \cdots & \cos K t_{2r} & \sin t_{2r} & \cdots & \sin K t_{2r} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\
1 & t_{nr} & \cos t_{nr} & \cdots & \cos K t_{nr} & \sin t_{nr} & \cdots & \sin K t_{nr}
\end{bmatrix}
\]

and for sine Fourier series
\[
\beta = \left( \frac{\beta_1}{2}, \gamma_1, \beta_{11}, \ldots, \beta_{k1}, \delta_{11}, \ldots, \delta_{k1}, \ldots, \frac{\beta_r}{2}, \gamma_r, \beta_{1r}, \ldots, \beta_{kr}, \delta_{1r}, \ldots, \delta_{kr} \right)^T;
\]
\[
T = \begin{bmatrix}
1 & t_{11} & \sin t_{11} & \cdots & \sin K t_{11} & 1 & t_{1r} & \sin t_{1r} & \cdots & \sin K t_{1r} \\
1 & t_{21} & \sin t_{21} & \cdots & \sin K t_{21} & 1 & t_{2r} & \sin t_{2r} & \cdots & \sin K t_{2r} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & t_{n1} & \sin t_{n1} & \cdots & \sin K t_{n1} & 1 & t_{nr} & \sin t_{nr} & \cdots & \sin K t_{nr}
\end{bmatrix}
\]

Lemma 4 is given and proofed in this study to determine the form of parameter vector estimator for parametric and nonparametric component.

**Lemma 4.**
Consider matrix equation for semiparametric regression that be presented in equation (6). Based on OLS optimization result can be gotten estimator parameter vector for parametric and nonparametric component that free from parameter $\beta$.

To proof Lemma 4, OLS optimization form is given as follows:
\[
\min (\alpha, \beta) = \min \varepsilon^T \varepsilon = \min Q(\alpha, \beta) = \min (y - X\alpha + T\beta)^T(y - X\alpha + T\beta) \tag{17}
\]
by elaborating on equation (17), it can be obtained:
\[
Q(\alpha, \beta) = y^T y - 2y^T X \beta - 2\beta^T T^T y + 2\alpha^T X^T T \beta + \alpha^T X^T X \alpha + \alpha^T T^T T \alpha \tag{18}
\]
For obtaining estimator from $\alpha$, can be determined by doing partial derivatives $Q(\alpha, \beta)$ to $\alpha$ with condition $\partial Q(\alpha, \beta)/\partial \alpha$ equals to 0. So, it can be resulted as follows:
\[
\hat{\alpha} = (X^T X)^{-1} \{X^T y - X^T T \hat{\beta}\} \tag{19}
\]
With similar step, for obtaining estimator from $\beta$, can be determined by doing partial derivatives $Q(\alpha, \beta)$ to $\beta$ with condition $\partial Q(\alpha, \beta)/\partial \beta$ equals to 0. So, it can be resulted as follows:
\[
\hat{\beta} = (X^T X)^{-1} \{X^T y - X^T T \hat{\alpha}\} \tag{20}
\]
Estimator in equation (19) and (20) 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 $\hat{\alpha}$ that free from parameter substitute equation (20) into equation (19).

$$\hat{\alpha} = (X^TX)^{-1} [X^Ty - X^T((T^T)^{-1}(T^Ty - T^TX\hat{\alpha}))]$$

So, it can be resulted that

$$\hat{\alpha} = (X^TX)^{-1}X^Ty - (X^TX)^{-1}X^TT(T^T)^{-1}T^Ty + (X^TX)^{-1}X^TT(T^T)^{-1}T^TX\hat{\alpha}$$

Furthermore, the part that contains the parameters grouped in one segment as follows:

$$\hat{\alpha} - (X^TX)^{-1}X^TT(T^T)^{-1}T^TX\hat{\alpha} = (X^TX)^{-1}X^Ty - (X^TX)^{-1}X^TT(T^T)^{-1}T^Ty$$

which can be expanded as follows:

$$\hat{\alpha}(I - (X^TX)^{-1}X^TT(T^T)^{-1}T^TX) = (X^TX)^{-1}X^Ty - (X^TX)^{-1}X^TT(T^T)^{-1}T^Ty$$

The parts that do not contain parameter estimators are moved to the right-hand side as follows:

$$\hat{\alpha} = (I - (X^TX)^{-1}X^TT(T^T)^{-1}T^TX)^{-1}((X^TX)^{-1}X^Ty - (X^TX)^{-1}X^TT(T^T)^{-1}T^Ty)$$

If defined $M = (I - (X^TX)^{-1}X^TT(T^T)^{-1}T^TX)^{-1}$, so can be determined estimator for parametric component as follows:

$$\hat{\alpha} = M(X^TX)^{-1}X^Ty - M(X^TX)^{-1}X^TT(T^T)^{-1}T^Ty$$

$$= M(X^TX)^{-1} \{X^T - X^TT(T^T)^{-1}T\}y$$

$$= A(K)y$$

(21)

with $A(K) = M(X^TX)^{-1} \{X^T - X^TT(T^T)^{-1}T\}$

To get $\hat{\beta}$ that free from parameter substitute equation (19) into equation (20).

$$\hat{\beta} = (T^T)^{-1} [T^Ty - T^TX((X^TX)^{-1}X^Ty - X^T\hat{\beta})]$$

So, it can be resulted that
\[ \hat{\beta} = (T^T T)^{-1} T^T y - (T^T T)^{-1} T^T X (X^T X)^{-1} X^T y + (T^T T)^{-1} T^T X (X^T X)^{-1} X^T \hat{\beta} \]

Furthermore, the part that contains the parameters grouped in one segment as follows:
\[ \hat{\beta} - (T^T T)^{-1} T^T X (X^T X)^{-1} X^T \hat{\beta} = (T^T T)^{-1} T^T y - (T^T T)^{-1} T^T X (X^T X)^{-1} X^T y \]

which can be expanded as follows:
\[ \hat{\beta} (I - (T^T T)^{-1} T^T X (X^T X)^{-1} X^T T) = (T^T T)^{-1} T^T y - (T^T T)^{-1} T^T X (X^T X)^{-1} X^T y \]

The parts that do not contain parameter estimators are moved to the right-hand side as follows:
\[ \hat{\beta} = (I - (T^T T)^{-1} T^T X (X^T X)^{-1} X^T T)^{-1} \left( (T^T T)^{-1} T^T y - (T^T T)^{-1} T^T X (X^T X)^{-1} X^T y \right) \]

If defined \( N = (I - (T^T T)^{-1} T^T X (X^T X)^{-1} X^T T)^{-1} \), so can be determined estimator for nonparametric component as follows:
\[
\hat{\beta} = N (T^T T)^{-1} T^T y - N (T^T T)^{-1} T^T X (X^T X)^{-1} X^T y \\
= N (T^T T)^{-1} \left( T^T - T^T X (X^T X)^{-1} X^T \right) y \\
= B(K) y \tag{22}
\]

with \( B(K) = N (T^T T)^{-1} \left[ T^T - T^T X (X^T X)^{-1} X^T \right] \)

After obtaining an estimator for parametric and nonparametric components, then determine the estimator of the semiparametric regression model with the Fourier series as follows:
\[
\hat{y} = X \hat{\alpha} + T \hat{\beta} \\
= X A(K) y + T B(K) y \\
= (X A(K) + T B(K)) y \tag{23}
\]

with \( H(K) = X A(K) + T B(K) \).

Equation (23) is a matrix equation for the Fourier series estimator in semiparametric regression which applies to three Fourier series forms with adjustment of matrices and vectors in nonparametric components. It appears that the parameter estimators in vector form for the parametric components presented in equation (21) and the nonparametric components presented in equation (22) are free from parameters. At the end of this section, consider the theorem that be related to the estimator of semiparametric regression curves based on the complete Fourier series in Theorem 5 and Fourier sine series in Theorem 6.
Theorem 5.
Consider semiparametric regression equation with curve regression approached by Fourier series that be presented in equation (13). The estimator for approaching curve regression is given as follows:

\[
y_i = \tilde{a}_0 + \sum_{l=1}^{p} \tilde{a}_l x_{il} + \sum_{j=1}^{r} \frac{\tilde{\beta}_0 j}{2} + \tilde{\gamma}_j t_{ij} + \sum_{k=1}^{K} (\tilde{\beta}_{kj} \cos kt_{ij} + \tilde{\delta}_{kj} \sin kt_{ij})
\]  

with the values of \(\tilde{a}_0, \tilde{a}_1, \ldots, \tilde{a}_p, \tilde{\beta}_0 j, \tilde{\gamma}_j, \tilde{\beta}_{kj}, \) and \(\tilde{\delta}_{kj}\) can be determined based on OLS result. ■

Theorem 6.
Consider semiparametric regression equation with curve regression approached by Fourier series that be presented in equation (14). The estimator for approaching curve regression is given as follows:

\[
\hat{y}_i = \tilde{a}_0 + \tilde{a}_1 x_{i1} + \ldots + \tilde{a}_p x_{ip} + \sum_{j=1}^{r} \frac{\tilde{\beta}_0 j}{2} + \tilde{\gamma}_j t_{ij} + \sum_{k=1}^{K} \tilde{\delta}_{kj} \sin kt_{ij}
\]

with the values of \(\tilde{a}_0, \tilde{a}_1, \ldots, \tilde{a}_p, \tilde{\beta}_0 j, \tilde{\gamma}_j, \) and \(\tilde{\delta}_{kj}\) can be determined based on OLS result. ■

4. The goodness indicator of estimator result
The goodness indicators that often be used in semiparametric regression based on Fourier series is Mean Square Error (MSE), Generalized Cross Validation (GCV), and determination coefficient (R²). All of goodness indicators can be applied to Fourier series estimator in semiparametric regression. In Fourier series estimator in semiparametric regression, an optimal oscillation parameter (K) is determined. In determining optimal oscillation parameter can be used Generalized Cross Validation (GCV) formula. GCV often be used because have asymptotically optimal properties [2]. For determining an optimal oscillation parameter can be seen based on the smallest GCV value. The formula of GCV given as follows:

\[
\text{GCV (K)} = \frac{\text{MSE (K)}}{(n^{-1} \text{trace}(I - H(K)))^2}
\]

with \(\text{MSE (K)} = n^{-1} y^T (I - H(K))^T (I - H(K)) y\).

By choosing of an optimal oscillation parameter will give impact to produce a determination coefficient with high value, or approximate to 100%. The determination coefficient formula given as follows:

\[
R^2 = \frac{(\hat{y} - \bar{y})^T (\hat{y} - \bar{y})}{(y - \bar{y})^T (y - \bar{y})} ; \quad 0 \leq R^2 \leq 1
\]

with \(\hat{y}\) is a vector that include of estimation result for all of subjects, and \(\bar{y}\) is a vector that include mean value for each subject. 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.
5. Conclusion
This study gives new knowledge about semiparametric regression based on Fourier series estimator, because aside from cosine Fourier series but there are also sine and complete Fourier series estimator, which are the other alternatives. The estimator form and how to obtain it is not much different, because it is enough to use an analog estimator of the existing cosine Fourier series in semiparametric regression. All of estimators could be applied in real problems involving data with trend and seasonal pattern, hence, it is suitable for Fourier series estimator in semiparametric regression concept. The best estimator is determined based on the smallest GCV and MSE for an oscillation parameter, also the highest coefficient of determination. Therefore, the comparison between the three estimators can be observed based on the smallest GCV and MSE, as well as the highest coefficient of determination for the selected oscillation parameter in every estimator.

Acknowledgements
The authors specially appreciate Airlangga University for funding this publication.

References
[1] Tjahjono E, Mardianto M F F and Chamidah N 2018 Far East Journal of Mathematical Sciences (FJMS) 103 1251
[2] Eubank R L 1999 Spline Smoothing and Nonparametric Regression 2nd Edition (New York: Marcel Dekker)
[3] Hardle W 1990 Applied Nonparametric Regression (New York: Cambridge University Press)
[4] Lin Z, Li D and Chen J 2008 Journal Multivariate Analysis 99 2339
[5] Bilodeau M 1992 The Canada Journal of Statistics 3 257
[6] Budiantara I N, Ratnasari V, Zain I, Ratna M and Mardianto M F F 2015 ATABS Journal 5 21
[7] Biedermann S, Dette H and Hoffmann P 2009 Ann Inst Stat Math Journal 61 143
[8] Dette H, Melas V B and Shpilev P 2016 Journal of Computational Statistics and Data Analysis 26 1
[9] Hardle W, Muller M, Werwatz A and Sperlich S 2004 Nonparametric and Semiparametric Models (New York: Springer Verlag)
[10] Asrini L J 2014 ARPN Journal of Engineering and Applied Sciences 9 1501
[11] Pane R, Budiantara I N, Zain I and Otok B W 2013 Journal of Applied Mathematical Science 8 5053
[12] Suslov S K 2003 An introduction to Basic Fourier Series (Arizona: Springer-Science)
[13] Takezawa K 2006 Introduction to Nonparametric Regression (New Jersey: John Wiley and Sons Inc)