ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

FOURIER SERIES SEMIPARAMETRIC REGRESSION MODELS (CASE STUDY: THE PRODUCTION OF LOWLAND RICE IRRIGATION IN CENTRAL JAVA)

Luh Juni Asrini[1] and I Nyoman Budiantara[2]
[1]Departement of Industrial Engineering, Widya Mandala Catholic University, Surabaya, Indonesia
[2]Department of Statistics, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia
E-Mail: juniasrini@ymail.com

ABSTRACT

Semiparametric regression model is a regression model where the shape of regression curve consists of a known pattern of parametric components and a smooth (smooth, flawless, slippery) nonparametric component which the pattern is unknown. The approach that used in estimating the nonparametric regression curves, one of which is, the Fourier series estimator. Fourier series estimator is commonly used when a data investigated patterns are not known and there is a tendency of repeating patterns. In the Fourier series estimator, the shape of nonparametric regression curve is assumed unknown and is contained in the space of continuous functions $C\ (0, \pi)$. This study aimed to analyze the shape of the estimator of the Fourier series semiparametric regression curve and applying it's to the data production of lowland rice irrigation in Central Java. Case studies are used to model the production of lowland rice irrigation in Central Java with predictor variables harvest area, the use of fertilizers, pesticides, seed, and the use of labor. Modeling aimed to determine the magnitude influence of the predictor variables on the response variable that is the number of production of lowland rice irrigation in Central Java. Modeling the production of lowland rice irrigation in Central Java with Fourier series semiparametric regression produced the coefficient value of determination $R^2 = 0.92$. It means that the magnitude influence of the predictor variables on the response variable is 92%. The performance of Fourier series semiparametric regression model was quite good in modeling the production of lowland rice irrigation in Central Java.

Keywords: fourier series, semiparametric regression, rice production.

1. INTRODUCTION

Regression analysis is one of the statistical analysis that used to investigate patterns of functional relationships between one or more variable. The main purpose of regression analysis is to find a form of regression curve estimation. There are three approaches of regression analysis to estimate the regression curve, namely parametric regression approach, nonparametric regression, and semiparametric regression.

In parametric regression approach, there is a very strong and rigid assumption that is the shape of the regression curve is known, for example linear, quadratic, cubic, polynomial of degree p, exponent, and others. In addition, knowledge of the past is needed on the characteristics of the data in order to obtain good modeling. In parametric regression models that estimated regression curve is equivalent to estimating the parameters in the model (Budiantara 2009).

Unlike the parametric regression approach, in the nonparametric regression, the form of the curve regression is assumed unknown and contained in a certain function space. Making this assumption depends on the smooth property which is owned by the function of the unknown pattern (Budiantara, 2000). Data are expected to seek their own form of the estimate, without being influenced by the subjective factor of the designer of the study. Thus, nonparametric regression approach has high flexibility (Eubank 1988).

One of the functions that can be used in estimating the nonparametric regression curve is a Fourier series. Fourier series estimator is generally used when a data investigated patterns are not known and there is a tendency of repeating patterns (Bilodeau 1992). The Research that has been done about Fourier series estimators include Fourier series estimators for nonparametric regression curve by Tripena and Budiantara (2006), Fourier series estimator in nonparametric regression weighted by Tjahjono (2009), and Fourier series estimators for nonparametric regression curve birespon by Semiati (2010).

In many cases, the relationship between variables is not always patterned parametric or nonparametric alone. According to Eubank (1988), in some other cases, often there is a relationship between variables that patterned semiparametric. In circumstances in which the shape of regression curve is composed of a known pattern of parametric components and smooth nonparametric components which the pattern is unknown, then used semiparametric regression approach. Semiparametric regression approach is a combination of parametric and nonparametric regression.

Based on the above description, this research will discuss about semiparametric regression using a Fourier series. Case studies are used to model the production of lowland rice irrigation in Central Java with predictor variables harvest area, the use of fertilizers, pesticides, seed, and the use of labor. Modeling aimed to determine the magnitude influence of the predictor variables on the response variable that is the number of production of lowland rice irrigation in Central Java. In addition, it can also be used for forecasting the amount of production of lowland rice irrigation in Central Java.

2. MATERIAL AND METHODS

2.1. Data Source

The data that used in this study are the production of lowland rice irrigation in Central Java data from Ongkos Usaha Tani Padi survey in 2008. The main source of income from as many as 691 households is derived from rice plants. The variables that examined in this study consists of the response variable and the predictor variables. Response variable in this study is the amount of rice production (kg) (Y). While selecting predictor variables based on the results of research which conducted by Zulkarnain (2004) about the factors that influence the amount of rice production in Central Java using Cobb-Douglas production function. Based on these results, becomes the predictor variables in this study were harvested area (X_1), the amount of fertilizer use (X_2), the amount of pesticide use (X_3), the number of seed use (X_4), and the amount of labor utilization (X_5).

2.2. Analysis Stage

This research was basically done in two steps. The first step is estimator curve of semiparametric regression Fourier series analysis. Then, the second step is modeling data production of lowland rice irrigation in Central Java using variables which has been determined. Framework of this research is summarized in Figure-1.

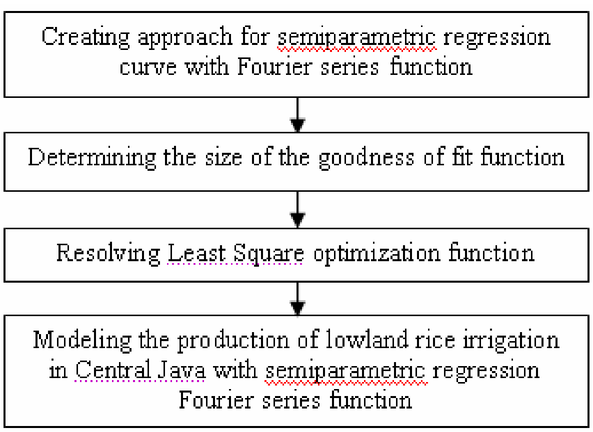

Figure-1. Framework of Research.

2.3. Semiparametric Regression

Semiparametric regression is a combination between parametric and nonparametric regression. If in a case, in which the shape of the regression curve is composed of a pattern of known parametric components and smooth nonparametric components are unknown pattern, the more appropriate approach is to use semiparametric regression (Eubank 1988). Suppose given the data pairs (x_i, z_i, y_i) and the relationship between x_i, z_i, and y_i is assumed to follow the semiparametric regression model. In this study semiparametric regression model is expressed as in equation (1).

$$\tilde{\mathbf{y}} = \boldsymbol{X}\tilde{\boldsymbol{\beta}} + g(\boldsymbol{Z}) + \tilde{\boldsymbol{\varepsilon}} \qquad (1)$$

with:

$\tilde{\mathbf{y}} = (y_1, y_2, ..., y_n)^T$, $\tilde{\boldsymbol{\beta}} = (\beta_0, \beta_1, \beta_2, \cdots, \beta_p)^T$,
$\boldsymbol{X} = (\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \cdots, \tilde{\mathbf{x}}_n)^T$, $\tilde{\mathbf{x}}_i = (x_{i1}, x_{i2}, \cdots, x_{ip})$,
$g(\boldsymbol{Z}) = (g(\tilde{\mathbf{z}}_1), g(\tilde{\mathbf{z}}_2), \cdots, g(\tilde{\mathbf{z}}_n))^T$, $\tilde{\mathbf{z}}_i = (z_{i1}, z_{i2}, \cdots, z_{iq})$,
and $\tilde{\boldsymbol{\varepsilon}} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)^T$
where:

y : response variable
β : unknown parameters on parametric regression
X : parametric linier predictor variables
Z : nonparametric predictor variables
g : unknown nonparametric regression curve function and contained in a certain function space

Respon variable y_i parametric associated with the predictor variables $x = (x_1, x_2, \cdots, x_p)$ and nonparametric associated with the predictor variables $z = (z_1, z_2, \cdots, z_q)$. Parameters $\tilde{\boldsymbol{\beta}} = (\beta_0, \beta_1, \beta_2, \cdots, \beta_p)^T \in R^{p+1}$ are not known. The main objective in semiparametric regression analysis is to obtain estimates of the regression curve $\hat{g}$ and $\hat{\beta}$. This estimate is obtained from minimizing the *Least Square* (LS):

$$\left(\tilde{\mathbf{y}} - \boldsymbol{X}\tilde{\boldsymbol{\beta}} - g(\boldsymbol{Z})\right)^T \left(\tilde{\mathbf{y}} - \boldsymbol{X}\tilde{\boldsymbol{\beta}} - g(\boldsymbol{Z})\right) \qquad (2)$$

to g functions contained in a certain function space (Budiantara 2005).

2.4. Estimator Fourier series

Fourier series is a trigonometric polynomial that has flexibility, so that it can adapt effectively to the local nature of the data. Fourier series is used to describe both the curve shows a sine and cosine wave. Given the data: (t_i, y_i), $i = 1, 2, \cdots, n$ and the relationship between t_i and y_i are assumed to follow the regression model:

$$y_i = g(t_i) + \varepsilon_i \qquad (3)$$

with g is the regression curve that assumed to be contained in the space $C(0, \pi)$, ε_i is random *error* that is assumed to be independently distributed with mean zero and variance σ^2. Since g is a continuous function then g can be approached with the T function (Bilodeau 1992), where:

$$T(t) = \frac{1}{2}a_0 + bt + \sum_{k=1}^{K} a_k \cos kt \qquad (4)$$

ARPN Journal of Engineering and Applied Sciences

In a regression analysis, to estimate the curve g can be used the least squares method (LS) which minimize the sum of squared errors. In other words, estimators for g can be obtained from (Tripena, 2006):

$$\underset{g\in C(0,\pi)}{\text{Min}}\left\{\sum_{i=1}^{n}\varepsilon_i\right\}=\underset{g\in C(0,\pi)}{\text{Min}}\left\{\sum_{i=1}^{n}\left(y_i-g(t_i)\right)^2\right\} \tag{5}$$

2.5. Selection of Smoothing Parameters

In the Fourier series estimator, the parameter K is a balance controller between smoothness function. If K is large, then estimators obtained from the function will be smooth, but the ability to map the data less well. Conversely, if K is small then obtained estimators of the function will be rough. Therefore, we need a value that is not too large and not too small, in order to obtain the best estimator functions. Thus the value of K is the optimal K value. The method which often used in selecting the smoothing parameter K is the Generalized Cross Validation (GCV). GCV values expressed in the following equation:

$$\text{GCV}(\lambda)=\frac{n^{-1}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{\left\{n^{-1}\left[1-\text{trace}\left(A(\lambda)\right)\right]\right\}^2} \tag{6}$$

3. RESULT AND DISCUSSIONS

3.1 Fourier Series Estimator on Semiparametric Regression

Suppose the data given in pairs $(x_1,x_2,\cdots x_p,z_1,z_2,\cdots,z_q,y)$, with x and z is the predictor variable and y is the response variable. The relationship between x and y known to form a pattern, whereas the relationship between z and y, the form of pattern are not known. Therefore, the relationship between x_i, z_i, and y_i is assumed to follow the semiparametric regression model. In this study, semiparametric regression model assumed there are p predictor variable $x_1,x_2,\cdots x_p$ which is a parametric component and q variable $z_1,z_2,\cdots,z_q$ which is a nonparametric component. The relationship between the response variable y follows the semiparametric regression model:

$$\tilde{\mathbf{y}}=\boldsymbol{X}\tilde{\boldsymbol{\beta}}+g(\boldsymbol{Z})+\tilde{\boldsymbol{\varepsilon}} \tag{7}$$

Model (7) consists of two components, namely parametric $\boldsymbol{X}\tilde{\boldsymbol{\beta}}$ and nonparametric components $g(\boldsymbol{Z})$, and can be presented:

$$\tilde{\mathbf{y}}^*=g(\boldsymbol{Z})+\tilde{\boldsymbol{\varepsilon}} \tag{8}$$

with $\tilde{\mathbf{y}}^*=\tilde{\mathbf{y}}-\boldsymbol{X}\tilde{\boldsymbol{\beta}}$.

Shape function g is unknown, and is estimated using Fourier series estimator. In general, nonparametric regression model Fourier series which given by Bilodeau (1996) with q predictor variables on the response to-i are:

$$g(z_{il})=\frac{1}{2}a_0+b_lz_{il}+\sum_{k=1}^{K}a_{lk}\cos kz_{il} \tag{9}$$

with $i=1,2,\cdots,n$ and $l=1,2,\cdots q$

Fourier series model (9) when substituted into the equation (8), assuming $g(\boldsymbol{Z})$ an additive function, then it can be written in matrix form as follows:

$$\tilde{\mathbf{y}}^*=\boldsymbol{C}\tilde{\boldsymbol{\alpha}}+\tilde{\boldsymbol{\varepsilon}} \tag{10}$$

where $\tilde{\mathbf{y}}^*=\left(y_1^*,y_2^*,\cdots,y_n^*\right)^T$,

$$C=\begin{pmatrix}1 & x_{11} & \cos x_{11} & \cos 2x_{11} & \cdots\cos Kx_{11} & x_{12} & \cos x_{12} & \cos 2x_{12} & \cdots\cos Kx_{12} & \cdots x_{1q} & \cos x_{1q} & \cos 2x_{1q} & \cdots\cos Kx_{1q}\\ 1 & x_{12} & \cos x_{12} & \cos 2x_{12} & \cdots\cos Kx_{12} & x_{22} & \cos x_{22} & \cos 2x_{22} & \cdots\cos Kx_{22} & \cdots x_{2q} & \cos x_{2q} & \cos 2x_{2q} & \cdots\cos Kx_{2q}\\ \vdots & \vdots & \vdots & \vdots & \ddots\ \vdots & \vdots & \vdots & \vdots & \ddots\ \vdots & \ddots\ \vdots & \vdots & \vdots & \ddots\ \vdots\\ 1 & x_{n1} & \cos x_{n1} & \cos 2x_{n1} & \cdots\cos Kx_{n1} & x_{n2} & \cos x_{n2} & \cos 2x_{n2} & \cdots\cos Kx_{n2} & \cdots x_{nq} & \cos x_{nq} & \cos 2x_{nq} & \cdots\cos Kx_{nq}\end{pmatrix},$$

and

$$\tilde{\boldsymbol{\alpha}}=\left(a_0^* \;\; b_1 \;\; a_{11} \;\; a_{12} \;\; \cdots \;\; a_{1K} \;\; b_2 \;\; a_{21} \;\; a_{22} \;\; \cdots \;\; a_{2K} \;\; \cdots \;\; b_q \;\; a_{q1} \;\; a_{q2} \;\; \cdots \;\; a_{qK}\right)^T$$

with $a_0^*=\dfrac{q}{2}a_0$.

Estimates for the parameters $\tilde{\boldsymbol{\alpha}}$ obtained by the method of least squares (LS). Defined goodness of fit, namely Least Squares criteria which is the square of the error in equation (10), and is denoted by $\psi(\tilde{\boldsymbol{\alpha}})$:

$$\psi(\tilde{\boldsymbol{\alpha}})=(\tilde{\mathbf{y}}^*-\boldsymbol{C}\tilde{\boldsymbol{\alpha}})^T(\tilde{\mathbf{y}}^*-\boldsymbol{C}\tilde{\boldsymbol{\alpha}}) \tag{11}$$

Estimator $\hat{\tilde{\boldsymbol{\alpha}}}$ obtained through the differentiation towards $\tilde{\boldsymbol{\alpha}}$. The minimum value $\psi(\tilde{\boldsymbol{\alpha}})$ is achieved when the result of the differential equal to zero, as follows:

$$\hat{\tilde{\boldsymbol{\alpha}}}=(\boldsymbol{C}^T\boldsymbol{C})^{-1}\boldsymbol{C}^T\tilde{\mathbf{y}}^* \tag{12}$$

Based on equation (8) and (10), it can be stated that:

$$\begin{aligned}\hat{g}(\boldsymbol{Z})&=\boldsymbol{C}\hat{\tilde{\boldsymbol{\alpha}}}\\ &=\boldsymbol{C}(\boldsymbol{C}^T\boldsymbol{C})^{-1}\boldsymbol{C}^T\tilde{\mathbf{y}}^*\\ &=\boldsymbol{A}\tilde{\mathbf{y}}^*\end{aligned} \tag{13}$$

If $g(\boldsymbol{Z})$ suspected with $\hat{g}(\boldsymbol{Z})$, then from equation (7) is obtained:

$$(\boldsymbol{I}-\boldsymbol{A})\tilde{\mathbf{y}}=(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}\tilde{\boldsymbol{\beta}}+\tilde{\boldsymbol{\varepsilon}} \tag{14}$$

with $\boldsymbol{I}$ is identity matrix .

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Estimates for the parameters $\tilde{\beta}$ can be obtained by Least Squares (LS) optimization method. Defined by goodness of fit, namely Least Squares criteria which is the square of the error equation (14), and is denoted by $\xi(\tilde{\beta})$ the following:

$$\xi(\tilde{\beta}) = \left((\boldsymbol{I}-\boldsymbol{A})\tilde{\mathbf{y}} - (\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}\tilde{\beta}\right)^T \left((\boldsymbol{I}-\boldsymbol{A})\tilde{\mathbf{y}} - (\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}\tilde{\beta}\right) \quad (15)$$

Estimator $\hat{\hat{\beta}}$ is obtained through differentiation of the equation $\xi(\tilde{\beta})$ towards $\tilde{\beta}$. The minimum valu $\xi(\tilde{\beta})$ is achieved when the result of the differential equal to zero.

$$\hat{\hat{\beta}} = \left(\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{A})^T(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}\right)^{-1} \boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{A})^T(\boldsymbol{I}-\boldsymbol{A})\tilde{\mathbf{y}} \quad (16)$$

Equation (13) when substituted $\hat{\hat{\beta}}$ value will be obtained for the curve $\hat{g}$ estimator as follows:

$$\hat{g}(\boldsymbol{Z}) = \boldsymbol{A}\left(\tilde{\mathbf{y}} - \boldsymbol{X}^T\left(\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{A})^T(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}\right)^{-1} \boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{A})^T(\boldsymbol{I}-\boldsymbol{A})\tilde{\mathbf{y}}\right) \quad (17)$$

Thus, the estimates is obtained for semiparametric regression curve Fourier series which given in equation (18).

$$\hat{\hat{\phi}} = \boldsymbol{X}\hat{\hat{\beta}} + \hat{g}(\boldsymbol{Z}) \quad (18)$$

3.2 Description of Irrigation Rice Production Data in Central Java

The relationship between the area harvested and the amount of rice production can be seen from the scatterplot in Figure-2. Based on Figure-2 it can be seen that the relationship between the number of harvested area of rice production tends to be linear. If the harvested area increased, the amount of rice production is also likely to increase. Therefore, the harvested area of rice production modeled with parametric linear.

Based on the scatterplot in Figure-3 shows the relationship between the amount of fertilizer to rice production number. Based on Figure-3 it can be seen that the use of fertilizer is less likely to result in the production of rice which tends to be slightly too. However, the use a lot of fertilizers tends to decrease production. The same data pattern also occurs in the relationship between the amount of rice production with the amount of pesticide use (Figure-4), the relationship between the amount of rice production with the use of the number of seeds (Figure-5), and the relationship between the amount of rice production with the amount of labor utilization (Figure-6). Established relationships tend not to follow the certain patterns or unknown pattern. Therefore, the relationship between the amount of rice production with the amount of fertilizer use, the amount of pesticide use, the amount of seed use, and the amount of labor utilization is modeled as a nonparametric.

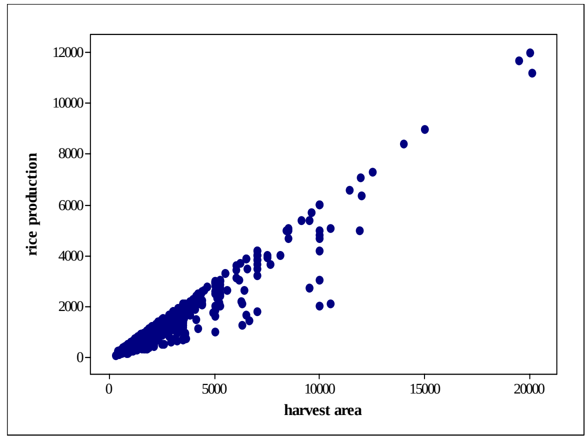

Figure-2. Scatterplot harvest area (X_1) vs rice production (Y).

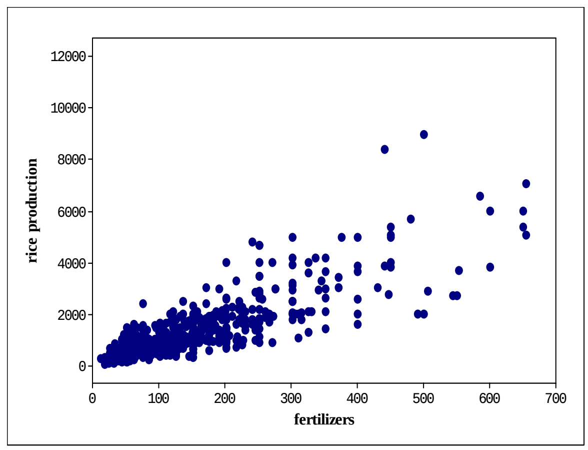

Figure-3. Scatterplot fertilizers (X_2) vs rice production (Y).

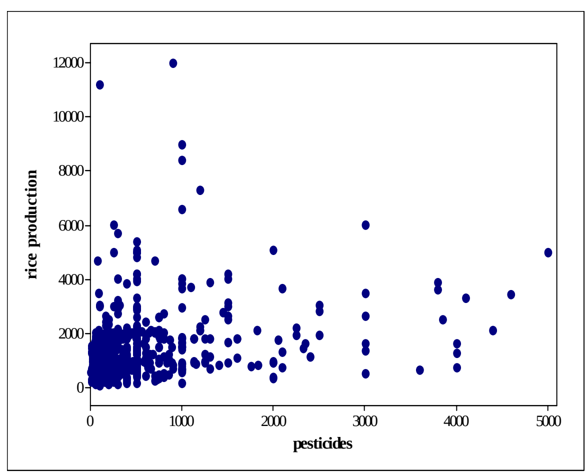

Figure-4. Scatterplot pesticides (X_3) vs rice production (Y).

VOL. 9, NO. 9, SEPTEMBER 2014 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

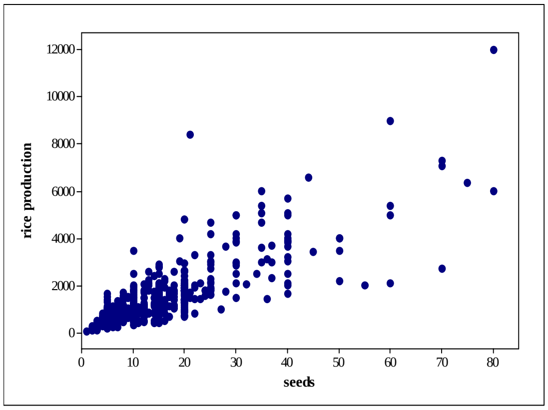

Figure-5. Scatterplot seeds (X_4) vs rice production (Y).

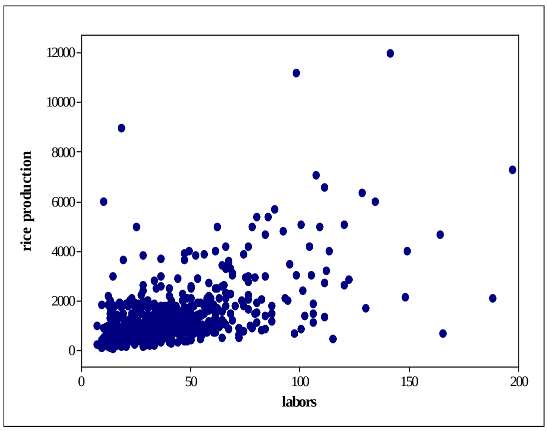

Figure-6. Scatterplot labors (X_5) vs rice production (Y).

3.3 Semiparametric Regression Model Production of Lowland Rice Irrigation In Central Java

Based on the description of the data using a scatterplot, it can be seen that the relationship between the amount of rice production (Y) with the harvested area (X_1) has a parametric linear patterns and the relationships amount of rice production with the use of fertilizers (X_2), the use of pesticides (X_3), the use of seeds (X_4), and the use of labor (X_5) has a nonparametric pattern, so that the model that used to model the production of lowland rice irrigation in Central Java is semiparametric model. The semiparametric model used is additive models that expressed in the following equation:

$$y_i = \beta_0 + \beta x_{i1} + g_1(x_{i2}) + g_2(x_{i3}) + g_3(x_{i4}) + g_4(x_{i5}) + \varepsilon_i \ .$$

In the scatterplot that illustrates the relationship between the amount of rice production (Y) with the use of fertilizers (X_2), the use of pesticides (X_3), the use of seeds (X_4), and the use of labor (X_5) can be seen there is a repeating pattern, namely there is a repetition of response values on different predictor variable values. Therefore, the Fourier series approach is used which is an approach in nonparametric regression that used when the data were investigated pattern is unknown and there is a tendency of repeating patterns. Based on the description, the model that used to model the production of lowland rice irrigation in Central Java is semiparametric model Fourier series. Estimation of the model with a Fourier series approximation is obtained as follows:

$$y_i = \omega + \beta x_{1i} + b_1 x_{2i} + \sum_{k=1}^{K} a_{1k} \cos kx_{2i} + b_2 x_{3i} + \sum_{k=1}^{K} a_{2k} \cos kx_{3i} + b_3 x_{4i} + \sum_{k=1}^{K} a_{3k} \cos kx_{4i} +$$

$$b_4 x_{5i} + \sum_{k=1}^{K} a_{4k} \cos kx_{5i} + \varepsilon_i$$

where $\omega = \beta_0 + a_0^*$.

In semiparametric regression Fourier series, is highly dependent on the parameter K. To obtain the best estimator of Fourier series in semiparametric regression, it is necessary to select the optimal K parameter. Selection of the optimal K is performed using GCV method that the formula shown in equation (2.10). Selected the optimal K value is the K value that produces the minimum GCV value. The results of the analysis for the value of K is given in the following Table:

Table-1. K Optimal Selection Table with GCV and R^2 Value.

K	GCV	R^2	K	GCV	R^2
1	803.651.122	0,9207436	11	24.954.525	0,9270503
2	376.734.213	0,9224818	12	21.252.790	0,9273142
3	218.226.828	0,9232129	13	18.124.960	0,9283017
4	142.452.936	0,9235122	14	15.610.737	0,9292763
5	99.715.206	0,9241207	15	13.693.299	0,9295604
6	73.424.580	0,9248170	16	12.051.575	0,9301407
7	56.587.668	0,9249706	17	10.674.274	0,9307426
8	44.889.662	0,9251776	18	9.424.114	0,9319694
9	36.441.411	0,9254161	19	8.412.009	0,9328027
10	29.827.097	0,9264609	20	7.577.149	0,9333461

VOL. 9, NO. 9, SEPTEMBER 2014 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

K value is the sum of the cosine wave oscillations in the model. The greater value of *K* will result in more complex models and oscillation of the estimate curve will be closer and follow the pattern of the actual data, so that the bias gets smaller and the variants gets larger. In this study the value of *K* is limited to *K* = 20 since based on Table-1, GCV value gets smaller and the value of R^2 gets larger in line with the increase of K value, however the increase in the value of R^2 is not too significant, so that the iteration process is stopped at a value of *K* = 20. Based on Table-1, it can be seen that the optimal value of *K* is 20, because of the value of *K* = 20, it is resulting the minimum value GCV. When viewed from the pattern of GCV values that obtained for the larger of *K* value so that it is resulting the value of the GCV which is getting smaller. However, at the same time the resulting model will be more complex. This happens because of the possibility of this research only focuses on semiparametric regression Fourier series without smoothing. Thus, the selection of the optimal value of *K* based on the minimum GCV value can not be done. Therefore, for each value of *K*, can be seen from the R^2 value of the model. For simplicity of the model, then chosen the value of *K* = 1, since the value of *K* = 1, the model has been able to produce a value of R^2 = 92.07%. That is, the diversity of response variable has been able to explain by the predictor variables was 92.07%. Thus, the estimator of semiparametric regression curves Fourier series with a value of *K* = 1 is good enough to model the production of lowland rice irrigation in Central Java. The model can be written as follows:

$$\begin{aligned}\hat{y} &= \hat{\omega} + \hat{\beta}_1 x_1 + \hat{b}_1 x_2 + \hat{a}_1 \cos x_2 + \hat{b}_2 x_3 + \hat{a}_2 \cos x_3 + \hat{b}_3 x_4 \\ &\quad + \hat{a}_3 \cos x_4 + \hat{b}_4 x_5 + \hat{a}_4 \cos x_5\end{aligned}$$

$$\begin{aligned}\hat{y} &= -69{,}9 + 0{,}495 x_1 + 1{,}68 x_2 - 2{,}94 \cos x_2 - 0{,}028 x_3 + 12{,}47 \cos x_3 + \\ &\quad -4{,}76 x_4 - 13{,}26 \cos x_4 - 2{,}44 x_5 - 8{,}19 \cos x_5\end{aligned}$$

4. CONCLUSIONS

Semiparametric regression model $\tilde{\mathbf{y}} = \boldsymbol{X}\tilde{\boldsymbol{\beta}} + g(\boldsymbol{Z}) + \tilde{\boldsymbol{\varepsilon}}$, when the nonparametric component regression curve g was approached by a Fourier series, so that the estimators for the parameters of the parametric component $\boldsymbol{\beta}$ is given by $\hat{\tilde{\boldsymbol{\beta}}} = \left(\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{A})^T(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{A})^T(\boldsymbol{I}-\boldsymbol{A})\tilde{\mathbf{y}}$ for a matrix **A** and **X.** While the nonparametric component estimator is given by $\hat{g} = \left(\boldsymbol{A} - \boldsymbol{A}\boldsymbol{X}^T\left(\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{A})^T(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}^T\right)^{-1}\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{A})^T(\boldsymbol{I}-\boldsymbol{A})\right)\tilde{\mathbf{y}}$

Estimation of semiparametric regression curve is given by the Fourier series:

$$\hat{\tilde{\boldsymbol{\phi}}} = \left(\boldsymbol{X}\left(\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{A})^T(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{A})^T(\boldsymbol{I}-\boldsymbol{A}) + \boldsymbol{A} - \boldsymbol{A}\boldsymbol{X}^T\left(\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{A})^T(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{A})^T(\boldsymbol{I}-\boldsymbol{A})\right)\tilde{\mathbf{y}}$$

Estimator of Fourier series semiparametric regression curve for data production of lowland rice irrigation in Central Java, produces the optimal value of *K* is equal to 1 and produces the value of R^2 = 92.07% with the model equations:

$$\begin{aligned}\hat{y} &= -69{,}9 + 0{,}495 x_1 + 1{,}68 x_2 - 2{,}94 \cos x_2 - 0{,}028 x_3 + 12{,}47 \cos x_3 + \\ &\quad -4{,}76 x_4 - 13{,}26 \cos x_4 - 2{,}44 x_5 - 8{,}19 \cos x_5\end{aligned}$$

It means that the magnitude influence of the predictor variables harvest area, the use of fertilizers, pesticides, seed, and the use of labor on the response variable, the number of production of lowland rice irrigation in Central Java is 92%.

REFERENCES

[1] Bilodeau M. 1992. Fourier Smoother and Additive Models. *The Canadian Journal of Statistics.* 3, 257-259.

[2] Budiantara I.N. 2005. Model Keluarga Spline Polinomial Truncated Dalam Regresi Semiparametrik. Papers of the National Seminar on Mathematics, Department of Mathematics, Diponegoro University, Semarang.

[3] Budiantara I N. 2009. Spline Dalam Regresi Nonparametrik Dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang. Inauguration Speech for Professor Position in the Field of Mathematics Statistics and Probability, Statistics Department, Faculty of Science, ITS, ITS Press, Surabaya.

[4] Eubank R.L. 1988. Spline Smoothing and Nonparametric Regression. Marcel Dekker, New York.

[5] Semiati Rini. 2010. Regresi Nonparametrik Deret Fourier Birespon. Thesis, ITS, Surabaya.

[6] Susianto Yuni. 2008. Model Regresi Semiparametrik Kernel Untuk Menduga Produksi Padi Sawah Di Jawa Tengah. Thesis, ITS, Surabaya.

[7] Tjahjono Eko. 2009. Estimator Deret Fourier Terbobot Dalam Regresi Nonparametrik. Thesis, ITS, Surabaya.

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