Modeling of Children Ever Born in Indonesia Using Fourier Series Nonparametric Regression

M A D Octavanny¹, I N Budiantara¹,², H Kuswanto¹ and D P Rahmawati¹

¹ Department of Statistics, Faculty of Mathematics, Computation and Data Science, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, 60111

* e-mail: nyomanbudiantara65@gmail.com

Abstract. Nonparametric regression aims to determine the relationship between response and predictor when the data does not follow a specific pattern. Fourier series is a nonparametric regression approach which has the flexibility to follow the characteristics of data. The purpose of this study is to obtain the estimator of the nonparametric regression using Fourier series and apply the model to the fertility data. The fertility rate represented by children ever born is one of the demographic factors that determine the decline in population growth rate. The data were obtained from the Indonesia Demographic and Health Survey 2017 with children ever born as a response. The predictors are the proportion of women graduating from junior high school, the proportion of women having sex before the age of 18, the proportion of women using a modern contraceptive method, and the infant mortality rate. The relationship between response and predictors tends to have a repetitive pattern with a certain trend. The best nonparametric regression model of children ever born in Indonesia is obtained by using 3 oscillation parameters for each predictor variable with GCV = 0.0534 and R-square = 80.04%.

Keywords: Nonparametric Regression, Fourier Series, Children Ever Born in Indonesia.

1. Introduction
Fertility is one of the most important demographic factors in reducing population growth rates [1]. The fertility rate represented by the children ever born (CEB) is one of the demographic factors that determine the decline in population growth rates. A decrease in fertility is inseparable from the success of the Indonesia family planning program set up in the 1970s. In this study, the Fourier series estimator will be applied to modeling the children ever born in Indonesia and the factors that influence it.

According to the Indonesia Demographic and Health Survey (IDHS) 2007, the national fertility rate is 2.6 and has decreased to 2.4 in 2017 [2]. This number has not been able to achieve the National Population and Family Planning Agency objectives set out in the Medium-Term National Development Plan 2015-2019, which is to reduce fertility from 2.6 to 2.28 children per woman. Based on the projected population of Indonesia between 2010 and 2035, the population will reach 273.5 million by 2020. The increase in population is due to the growing number of children born with high fertility. National fertility rates have declined, but fertility rates are not evenly distributed across all provinces in Indonesia. There are still imbalances in the development of family planning and low-quality economic growth in the region [3]. Based on these conditions, further analysis is needed to determine the factors affecting fertility rates in Indonesia which can be an useful information for policy-makers.
Several researches have been conducted to determine the cause of increasing or decreasing the fertility rate. Davis and Blake [4] declared that the use of effective contraception is one of the factors that affect low fertility. Hill, Stycos, and Back [5] stated that education and age at marriage are basic demographic factors that affect fertility. Leibenstein [6] claimed that fertility control methods delay marriage and the use of contraception. Freedman [7] argued that norms about family size and intermediate variables are influenced by mortality rates and socioeconomic structures that exist in society. In addition, environmental factors influence the infant mortality rate and fertility.

Fourier series in nonparametric regression is used for data that tend to have seasonal patterns [8] [9]. Winter [10] and Tarter et al [11] conducted research using the Fourier series nonparametric regression method. Bilodeau [8] and Pane et al [12] estimated the additive component with a function of a truncated Fourier cosine series and used Penalized Least Square to obtain the coefficients. Kim and Hart [13] reviewed point estimators based on local Fourier estimates and perform simulation studies to compare their performance with linear methods. Asrini and Budiantara [14] applied Fourier series in the semiparametric regression model to the production of lowland rice irrigation in Central Java.

Based on the formulated problem, this study aims to obtain the estimator of the nonparametric regression based on Fourier series and to model the fertility data with children ever born in Indonesia as the response using Fourier series nonparametric regression.

2. Literature Review

Regression analysis is an approach used to determine the relationship between response and predictor variables [15]. Nonparametric regression is a statistical method which used to define the relationship between response and predictor variables where the data pattern has unknown regression curve e.g. its regression curve is assumed to be smooth where the function belongs to a particular function space [16].

Fourier series are trigonometric polynomial functions that have high flexibility to deal with data that has a repeating pattern [17]. Fourier series are good for describing curves that show sine and cosine waves. This estimator is commonly used if there is a recurring tendency [8]. A nonparametric regression based on Fourier series estimator follows the model below:

\[ y_i = g(x_{i1}, x_{i2}, \ldots, x_{ip}) + \epsilon_i = \sum_{j=1}^{p} g_j(x_{ij}) + \epsilon_i, \quad i = 1, 2, \ldots, n \]  

(1)

Regression curve \( g_j(x_{ij}) \), \( j = 1, 2, \ldots, p \) is assumed in a continuous functional space \( C(0, \pi) \). Hence, \( g_j(x_{ij}) \) can be approached by Fourier series function with trend defined by:

\[ g_j(x_{ij}) = d_j x_{ij} + \frac{1}{2} c_{oj} + \sum_{h=1}^{H} c_{hj} \cos hx_{ij}, \quad j = 1, 2, \ldots, p \]  

(2)

The regression model in equation (2) can be rewritten in matrix form as:

\[ y = X\gamma + \epsilon \]  

(3)

In consequence, the nonparametric regression model based on Fourier series estimator is expressed as:

\[ \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} X_1(H) & \ldots & X_p(H) \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \]

where \( y \) is the \( n \times 1 \) dimensional vector of response, \( X \) is the \( n \times (H+2) \) dimensional matrix with \( j \)th entry is given by:
\[
X_j(H) = \begin{bmatrix}
x_{j1} & 1 & \cos x_{j1} & \ldots & \cos H x_{j1} \\
x_{j2} & 1 & \cos x_{j2} & \ldots & \cos H x_{j2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
x_{jn} & 1 & \cos x_{jn} & \ldots & \cos H x_{jn}
\end{bmatrix}, \; j = 1, 2, \ldots, p,
\]

\( \gamma \) is the \((H + 2) \times 1\) dimensional vector with \(j\)th entry is given by:
\[
\gamma_j = \left[ d_j \cdot \frac{1}{2} c_{o,j} \quad c_{ij} \quad \ldots \quad c_{ij} \right]^T, \; j = 1, 2, \ldots, p,
\]
and \( \varepsilon \) is the \(n \times 1\) dimensional vector of random error.

This research will use the least square method in order to obtain the estimation of the regression curve. The advantages of the least square method are simple and free of distribution.

3. Result and Discussion

3.1. Fourier Series Estimator
Based on the model in equation (2), the estimators of nonparametric regression model are obtained by using Least Square (LS) method. The estimation of \( \gamma \) can be obtained through LS optimization as follows:

\[
\min_{\gamma} \left\{ (y - X(H)\gamma)^T (y - X(H)\gamma) \right\}
\]  
(4)

To solve the LS optimization in equation (4), consider:

\[
Q(\gamma) = (y - X(H)\gamma)^T (y - X(H)\gamma) \\
= y^T y - 2\gamma^T X^T (H)y + \gamma^T X^T (H)X(H)\gamma \]

The estimation of \( \gamma \) is obtained by using a partial derivative of \( Q(\gamma) \) to \( \gamma \). Thus, the derivation of \( Q(\gamma) \) respect to \( \gamma \) is:

\[
\frac{\partial Q(\gamma)}{\partial \gamma} = 0 \\
-2X^T (H)y + 2X^T (H)X(H)\hat{\gamma} = 0 \\
-X^T (H)y + X^T (H)X(H)\hat{\gamma} = 0 \\
X^T (H)X(H)\hat{\gamma} = X^T (H)y \\
\hat{\gamma} = \left( X^T (H)X(H) \right)^{-1} X^T (H)y
\]  
(5)

Consequently, the estimator of \( \gamma \) is given by:

\[
\hat{\gamma} = \left( X^T (H)X(H) \right)^{-1} X^T (H)y = \Lambda(H)y
\]  
(6)

where \( \Lambda(H) = \left( X^T (H)X(H) \right)^{-1} X^T (H) \).

A different oscillation parameter results on varying parameter \( \gamma \). Generalized Cross Validation (GCV) method is one method that used for the selection of an optimal oscillation parameter. After we obtain the estimator of Fourier series in nonparametric regression in equation (6), the GCV for each oscillation obtained by:
\[ GCV(H) = \frac{MSE(H)}{\left[ n^{-1} trace\left( I - A(H) \right) \right]^2} \]  

with \( MSE(H) = n^{-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \). 

3.2. Modeling of Children Ever Born in Indonesia

Nonparametric regression model based on Fourier series estimator will be implemented to fertility data represented by children ever born (CEB). The data of this research were obtained from the Indonesia Demographic and Health Survey 2017. The dataset comprises of 34 provinces in Indonesia with CEB as a response variable \((y)\). Predictor variables are the proportion of women graduating from junior high school \((x_1)\), the proportion of women having sex before the age of 18 \((x_2)\), the proportion of women using a modern contraceptive method \((x_3)\), and the infant mortality rate \((x_4)\).

**Figure 1.** Scatterplot between Children Ever Born with Each Predictor.

Figure 1 illustrated the partial relationship between CEB and each predictor variable. It can be seen that the relationship between children ever born and are the proportion of women graduating from junior high school has a repetitive pattern with a certain trend line. Similar pattern is observed on the relationship between the response and the other predictors. For this reason, we proposed a nonparametric regression approach based on Fourier series estimator. From the result of various oscillation parameters, it is obtained GCV values which are presented in table 1.

| Oscillation Parameter \((H)\) | GCV \((H)\) |
|-------------------------------|------------|
| 1                             | 0.2485     |
| 2                             | 0.0639     |
| 3                             | **0.0534** |

This modeling is very dependent on oscillation parameters. In this study limited to only using 3 oscillation parameters. The selection of the optimum parameter using minimum GCV criteria.
the result of table 1, it is obtained minimum GCV value is 0.0534 achieved at 3 oscillation parameters for each predictor. The result of parameter estimation is presented in table 2.

Table 2. The Results of Parameter Estimation with Optimal Oscillation Parameter (H).

| Parameter | Estimation | Parameter | Estimation |
|-----------|------------|-----------|------------|
| $\gamma_0$ | 57.06      | $d_3$     | 27.83      |
| $d_1$     | -199.32    | $c_{13}$  | 29.28      |
| $c_{11}$  | 32.87      | $c_{23}$  | -17.25     |
| $c_{21}$  | -106.71    | $c_{33}$  | 14.94      |
| $d_2$     | -2.25      | $d_4$     | 0.01       |
| $c_{12}$  | 43.90      | $c_{14}$  | 0.02       |
| $c_{22}$  | 12.66      | $c_{24}$  | -0.01      |
| $c_{32}$  | -12.96     | $c_{34}$  | -0.02      |

Based on the parameter estimation in table 2, so the nonparametric regression model based on Fourier series estimator can be written as follows:

$$
\hat{y}_i = 57.06 - 199.32x_{i1} + 32.87\cos x_{i1} - 106.71\cos 2x_{i1} - 4.21\cos 3x_{i1} - 2.25x_{i2} + 43.9\cos x_{i2} +
+12.66\cos 2x_{i2} - 12.96\cos 3x_{i2} + 27.83x_{i3} + 29.28\cos x_{i3} - 17.25\cos 2x_{i3} + 14.94\cos 3x_{i3} +
+0.01x_{i4} + 0.02\cos x_{i4} - 0.01\cos 2x_{i4} - 0.02\cos 3x_{i4}
$$

Nonparametric regression modeling of children ever born in Indonesia based on Fourier series estimator obtained R-square equal to 80.04% and MSE equal to 0.0078. According to the nonparametric regression model, the estimated and actual value of children ever born in Indonesia are presented in table 3. The result from table 3 can be illustrated by a line graph to show the variation of the actual and estimated value. Figure 2 is the comparison between actual and estimated children ever born in Indonesia. Based on the diagram of comparison in figure 2, there is a small difference between the estimated value and the actual data. Nonparametric regression based on Fourier series approach has been able to recognize the repetitive pattern with a certain trend line. Thus, it can be stated that the estimation results of this study are able to predict children ever born in Indonesia.

Figure 2. The Comparison between Actual and Estimated Children Ever Born in Indonesia.
### Table 3. The Value of Actual and Estimated Children Ever Born in Indonesia.

| No | Province                  | Estimated CEB | Actual CEB |
|----|---------------------------|---------------|------------|
| 1  | Aceh                      | 1.865         | 1.806      |
| 2  | Sumatera Utara            | 1.821         | 1.937      |
| 3  | Sumatera Barat            | 1.551         | 1.599      |
| 4  | Riau                      | 1.681         | 1.902      |
| 5  | Jambi                     | 1.717         | 1.722      |
| 6  | Sumatera Selatan          | 1.763         | 1.888      |
| 7  | Bengkulu                  | 1.701         | 1.820      |
| 8  | Lampung                   | 1.697         | 1.716      |
| 9  | Kepulauan Bangka Belitung | 1.676         | 1.619      |
| 10 | Kepulauan Riau            | 1.734         | 1.638      |
| 11 | DKI Jakarta               | 1.586         | 1.408      |
| 12 | Jawa Barat                | 1.688         | 1.626      |
| 13 | Jawa Tengah               | 1.588         | 1.536      |
| 14 | DI Yogyakarta             | 1.236         | 1.291      |
| 15 | Jawa Timur                | 1.655         | 1.507      |
| 16 | Banten                    | 1.742         | 1.747      |
| 17 | Bali                      | 1.642         | 1.573      |
| 18 | Nusa Tenggara Barat      | 1.740         | 1.678      |
| 19 | Nusa Tenggara Timur       | 2.034         | 2.122      |
| 20 | Kalimantan Barat         | 1.948         | 1.888      |
| 21 | Kalimantan Tengah         | 1.887         | 1.879      |
| 22 | Kalimantan Selatan        | 1.779         | 1.721      |
| 23 | Kalimantan Timur          | 1.706         | 1.897      |
| 24 | Kalimantan Utara          | 1.905         | 1.931      |
| 25 | Sulawesi Utara            | 1.662         | 1.692      |
| 26 | Sulawesi Tengah           | 1.797         | 1.857      |
| 27 | Sulawesi Selatan          | 1.828         | 1.718      |
| 28 | Sulawesi Tenggara         | 2.004         | 2.001      |
| 29 | Gorontalo                 | 1.868         | 1.787      |
| 30 | Sulawesi Barat            | 1.905         | 1.953      |
| 31 | Maluku                    | 1.895         | 1.847      |
| 32 | Maluku Utara              | 2.028         | 2.041      |
| 33 | Papua Barat               | 2.101         | 2.054      |
| 34 | Papua                     | 2.171         | 2.201      |

### 4. Conclusion and Suggestion
In summary, it can be concluded that the children ever born in Indonesia can be modeled using nonparametric regression based on the Fourier series approach. The best model uses 3 oscillation parameters for each predictor with GCV equal to 0.0534. R-square of this model reaches 80.04% and MSE equal to 0.0078. Based on the conclusion, we suggest developing a nonparametric regression based on the Fourier series estimator for a higher oscillation parameter to improve the model performance.
References
[1] Iswarati 2009 Analisa Lanjut SDKI 2007: Proximate Determinant Fertilitas di Indonesia (Jakarta: Puslitbang KB dan Kesehatan Reproduksi BKKBN)
[2] Jatmiko Y A and Wahyuni S 2019 Determinan Fertilitas di Indonesia: Hasil SDKI 2017 Jurnal Euclid 6 95-106
[3] Andini N and Ratnasari V 2018 Pemetaan Total Fertility Rate (TFR) di Jawa Timur Menggunakan Pendekatan Regresi Logistik Biner dengan Efek Interaksi Jurnal Sains dan Seni ITS 7 D168-D175
[4] David K and Blake J 1956 Social Structure and Fertility: An Analytic Framework Economic Development and Cultural Change 4 211-235
[5] Hill R J, Stycos M and Back K W 1959 The Family and Population Control: A Puerto Rican Experiment in Social Change (Chapel Hill: University of North Carolina Press)
[6] Leibenstein H 1974 An Interpretation of the Economic Theory of Fertility: Promising Path or Blind Alley? Journal of Economic Literature 12 457-479
[7] Freedman R 1961 The Sociology of Human Fertility: a Trend Report and Bibliography Current Sociology 10 35-68
[8] Bilodeau M 1992 Fourier Smoother and Additive Models The Canadian Journal of Statistics 3 257-269
[9] Tripena A 2013 Estimator Deret Fourier untuk Estimasi Kurva Regresi Nonparametrik Birespon Magistra 84 6-15
[10] Winter B B 1981 Fourier Series Methods in Nonparametric Estimation Series Statistics 12 339-355
[11] Tarter M E, Freeman W and Hopkins A 1986 A Fortran Implementation of Univariate Fourier Series Density Estimation Communications in Statistics - Simulation and Computation 15 855-870
[12] Pane R, Budiantara I N, Zain I and Otok B W 2014 Parametric and Nonparametric Estimators in Fourier Series Semiparametric Regression and Their Characteristics Applied Mathematical Sciences 8 5053-5064
[13] Kim J and Hart J D 2011 A Change-Point Estimator Using Local Fourier Series Journal of Nonparametric Statistics 23 83-98
[14] 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 1501-1506
[15] Draper N R and Smith H 1998 Applied Regression Analysis (New York: John Wiley & Sons)
[16] Eubank R L 1999 Nonparametric Regression and Spline Smoothing (New York: Marcel Dekker, Inc)
[17] Budiantara I N, Ratnasari V, Ratna M, Wibowo W, Afifah N, Rahmawati D P and Octavanny M A D 2019 Modeling Percentage of Poor People in Indonesia Using Kernel and Fourier Series Mixed Estimator in Nonparametric Regression Investigacion Operacional 40 538-550