OPEN ACCESS
Jurnal Ilmiah Ilmu Terapan Universitas Jambi
Vol. 8 No. 2. December 2024, pp. 622-633 DOI. 10.22437/jiituj.v8i2.36246
Research Article

ESTIMATOR AND APPLIED MIXED KERNEL AND FOURIER SERIES MODELLING IN NONPARAMETRIC REGRESSION

I Wayan Sudiarsa[1,*], Ni Putu Ayu Mirah Mariati[2], Ni Made Sukma Sanjiwani[2], Dwi Prinicila Pramesuari[1]

[1] Universitas PGRI Mahadewa Indonesia, Bali, Indonesia
[2] Universitas Mahasaraswati Denpasar, Bali, Indonesia
Corresponding author email: sudiarsa@mahadewa.ac.id

Article Info
Recieved: Jul 02, 2024
Revised: Sep 28, 2024
Accepted: Oct 30, 2024
OnlineVersion: Nov 09, 2024

Abstract
There are three nonparametric regression approaches, namely parametric, nonparametric, and semiparametric regression. Nonparametric regression allows the response variable to follow a different curve from one predictor variable to another. In paired data, the components of the predictor variables and response variables are assumed to follow unknown data patterns, so they can be approached with kernel-based regression models and Fourier series. The basic components are approached with kernel functions and Fourier series functions. Errors are assumed to be normally distributed with zero mean and constant variance. The originality of this research is to obtain a mixed kernel and Fourier series model estimator and then apply it to poverty data in Bali Province. The research stage method begins with a nonparametric regression model estimator based on kernel and Fourier series. The next step is to research the regression curve estimation and obtain lemmas and theorems. The results of the function estimation are highly dependent on the bandwidth, smoothing, and oscillation parameters. In the application to the case of real data, the resulting model gives an R2 value of 0.6278, meaning that the variables used can explain the model by 62.78 percent. From the modeling results obtained, the Open Unemployment Rate has a positive effect on the percentage of poor people in Bali.

Keywords: Fourier Series, Kernel, Nonparametric Regression Modelling

INTRODUCTION

Regression analysis is one analyzes in Statistics that is used to investigate functional relationship pattern between two or more variables (Choerunnisa, Dewi, & Bariklana, 2021). There are three approaches in estimating the regression curve, namely the parametric regression approach, nonparametric regression and semi-parametric regression (Hidayat et al., 2020; Kusuma, 2020; Suwarni, 2021; Dessi & Shah, 2023; Helida, Ching, & Oyewo, 2023). In practice, the functional relationship pattern between the response variable and the predictor variable is often not known . In this condition, of course, the parametric regression model is not suitable to use (Ahmad & Asaad, 2024).

Nonparametric regression model is good to be used for data patterns of unknown form. Nonparametric regression has high flexibility, where the data is expected to find its own estimated form of the regression curve without being influenced by the researcher's subjectivity (Eubank, 1999).

There are many types of estimators in nonparametric regression model, such as kernel, splines, local polynomial, wavelets, neural network, machine learning, Fourier series (Wang & Ke, 2009; Respati, Isram, & Kusrini, 2022; Yohanie et al., 2023; Almufti, Hani, & Zeebaree, 2024; Cadiz, Manuel, & Reyes, 2024; Kalluçi, 2024; Zakiyah, Boonma, & Collado, 2024). Several studies regarding esti-mators have been carried out by several researchers (Zeng, 2019). Research on spline estimators has been carried out by other researchers such as (Mariati, 2021) Research on local polynomial estimators was conducted by (Berlinet, 2004). Research on the wavelet estimator was carried out by (Osmani, 2019). Research for the Fourier series estimator was carried out by several researchers, such as (Yilmaz, 2020). The latest method regarding kernels, namely Quick and Simple Kernel Differential Equation Regression Estimators for Data with Sparse Design has been developed by Nonparametric regression research continues to be developed by researchers (Ge & Braun, 2024). These studies generally use the same method for all predictor variables. Even though, these variables do not necessarily have the same data pattern. For this reason, it is necessary to develop mixed nonparametric regression to answer these problems. Previous researchers who have developed two types of mixed estimators in nonparametric regression are (Nurcahayani, 2021) involving Fourier series and multivariable truncated splines, (Lu, 2020) involving truncated splines and kernel, (Cheng, 2018) involving kernel and multivariable truncated linear splines and (Cheruiyot, 2020) involving truncated splines and multivariable kernel. The purpose of this research is to perform kernel-based nonparametric regression modeling and Fourier series. Further-more, it will be applied to the poverty percentage data in the Bali province. The originality of this research is the application of household data using a nonparametric regression mixed of kernels and fourier series in cross section data analysis.

The mixed kernel and Fourier series estimator is a method for estimating or approximating functions by combining two main techniques, namely kernel (from the kernel smoothing or kernel regression method) and Fourier series (to represent periodic functions). This method is commonly used in machine learning or data analysis to handle non-linear and complex data. Here is an overview of the two main components:

1. Kernel Estimation: This technique is used to estimate the probability distribution function or non-parametric regression (Linke et al., 2024). Kernel function is a weighting function used to calculate the average value of data around an estimated point (Funke, 2024). Some popular kernel functions include Gaussian, Epanechnikov, and Uniform kernels. Kernel smoothing is one of the popular techniques for dealing with non-linear regression problems, where estimation is done by giving greater weight to data that is closer to the estimated point.
2. Fourier Series: Fourier Series is a method for representing periodic functions in the form of a combination of sine and cosine functions. In function estimation, Fourier series can be used to handle data that has cyclical or periodic properties. Fourier series effectively breaks down a signal into its component frequencies, which can provide important insights in the analysis of signals or repetitive data.

Mixed Kernel and Fourier Series Estimators: Combining these two approaches means harnessing the power of kernel smoothing in handling non-linear data while using Fourier series to capture the periodic components of the data. General Steps: Data Representation: The data is approximated using a Fourier series to capture periodic patterns. Kernel Smoothing: the function generated from the Fourier series is then smoothed using a kernel to handle non-linear variations in the data. Benefits: Can handle complex patterns in data that involve both periodic and non-linear characteristics. This estimator is best suited for data that has both high and low frequency components simultaneously. This approach provides great flexibility in function estimation for data with complex structures.

RESEARCH METHOD

The data used in this research is data from the Badan Pusat Statistik (BPS) Indonesia. The data is the used research variable in this research is the response variable (Y), namely the percentage of poor people in 2020. Based on the obtained data and information, the predictor variable (X) in this research is the average length of schooling (X1), Open Unemployment Rate (X2) and Literacy Rate (X3). Badan Pusat Statistik (BPS) Indonesia uses the concept of basic needs approach in measuring poverty. By using this approach, poverty is seen as an economic inability to meet basic food and non-food needs as measured by expenditure. The Poverty Line is the sum of the Food Poverty Line and the Non-Food Poverty Line. People who have an average expenditure per capita per month below the Poverty Line are categorized as poor, while the percentage of poor people is the percentage of people who are below the Poverty Line (BPS, 2012).

The research steps are as follows:

1. Determine the next kernel function, estimate the parameters contained in the function.
2. Determine the next fourier series function, estimate the parameters contained in the function.
3. Combine the kernel function and fourier series, estimate the combined parameters.

There is given a pair of independent observation (v_i, y_i), i = 1,2,...,n, where y_i is the response variable, while v_i is the predictor variable. The y_i and v_i relationship can be modeled functionally in the form:

$$y_i = g(v_i) + \varepsilon_i , \quad (1)$$

where the regression curve $g(v_i)$ is a curve of unknown form. The $g(v_i)$curve in model (1) can be estimated using the Nadaraya-Watson kernel estimator. As a result:

$$\hat{g}_{\varphi}(v) = n^{-1}\sum_{i=1}^{n}\frac{K_{\varphi}(v - v_i)}{n^{-1}\sum_{i=1}^{n}K_{\varphi}(v - vi)}y_i ,$$

$$= n^{-1}\sum_{i=1}^{n}W_{\phi i}(v)y_i \qquad (2)$$

where $\hat{g}_{\phi}(v)$ is the kernel regression estimation function and it is the bandwidth width. Function $W_{\phi i}(v)$ is a weighting function

$$W_{\phi i}(v) = \sum_{i=1}^{n}\frac{K_{\phi}(v - v_i)}{n^{-1}\sum_{i=1}^{n}K_{\phi}(v - vi)}$$

where $K_{\phi}(v - v_i)$ is kernel function

$$K_{\phi}(v - v_i) = \frac{1}{\phi}K\left(\frac{v - v_i}{\phi}\right)$$

In this research , the used K is a Gaussian kernel function, with the equation

$$K_{\phi}(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right), -\infty < z < \infty$$

The Fourier series estimator in nonparametric regression is generally used when the pattern of the investigated data is unknown and there is a tendency to be repeated, because the Fourier series function is a continuous function, so it can be approximated by the T function, where:

$$T(t) = bt + \frac{1}{2}a_0 + \sum_{k-1}^{M}a_k \cos kt \qquad (3)$$

with the k is the oscillation parameter.

4. Model of factors influencing poverty using the mixed kernel and fourier series parameter estimation method.

RESULTS AND DISCUSSION

This research uses secondary data from the Central Statistics Agency (BPS) of year 2020; with the observation units are all nine regencies/cities in Bali Province. The used research variable in this research is the response variable (Y), namely the percentage of poor people in 2020. Based on the obtained data and information, the predictor variable (X) in this research is the average length of schooling (X1), Open Unemployment Rate (X2) and Literacy Rate (X3). The research stage begins with the introduction of kernel-based nonparametric regression and the Fourier series models. The next step is to research the estimation of regression curve, the regression curve estimator, bandwidth selection, smoothing parameter and optimum oscillation parameter. The final step is to apply the model to data on the percentage of poor people in Bali Province.

Estimation of Kernel-Based Nonparametric Regression and Fourier Series Models

There is given paired data $\left(v_{1i}, v_{2i}, \ldots, v_{pi}, t_{1i}, t_{2i}, \ldots, t_{qi}, y_1\right)$ i = 1,2,.. ,*n* which has an assumed relationship to follow a nonparametric regression model. The variables are the predictor variables and y, which are the response variables. The form of nonparametric regression model is

$$y_i = \mu\left(v_{1i}, v_{2i}, \ldots, v_{pi}, t_{1i}, t_{2i}, \ldots, t_{qi}\right) + \varepsilon_i$$

$$= \mu\left(v_1, t_i\right) + \varepsilon_i \quad (4)$$

where $v_i = \left(v_{1i}, v_{2i}, \ldots, v_{pi}\right)$ and $t_i = \left(t_{1i}, t_{2i}, \ldots, t_{qi}\right)$. The shape of the regression curve in model is assumed to be unknown and the curve is smooth in the sense of being continuous and differentiable. Random errors ε_i are normally distributed with E$\left(\varepsilon_i\right)$ = 0 and Var $\left(\varepsilon_i\right)$ = σ^2.In addition, the regression curve $\mu\left(v_1, t_i\right)$ is assumed to be additive, so it can be written in the form

$$\mu\left(v_1, t_i\right) = \sum_{j=1}^{p} g_j\left(v_{ji}\right) + \sum_{s=1}^{q} h_s\left(t_{si}\right) \quad (5)$$

The main problem with the nonparametric regression curve estimator is obtaining the shape of the regression curve estimate $\mu\left(v_1, t_i\right)$

$$y = \sum_{j=1}^{p} \hat{g}_j\left(v_{ji}\right) + \sum_{s=1}^{q} \hat{h}_s\left(t_{si}\right) + \varepsilon \quad (6)$$

curve $g_j\left(v_{ji}\right)$ is approximated by a kernel function and the regression curve $h_s\left(t_{si}\right)$ is approximated by a Fourier series function.

Lemma 1. If the kernel function component in equation (5) is estimated with the estimator component in equation (2), then

$$\sum_{j=1}^{p} \hat{g}_{j\phi_j}\left(v_j\right) + \mathrm{V}(\Phi) y \quad (7)$$

where

$$\hat{g}_{j\phi_j}\left(v_j\right) = \begin{bmatrix} \hat{g}_{j\phi_j}\left(v_{j1}\right) \\ \hat{g}_{j\phi_j}\left(v_{j2}\right) \\ \vdots \\ \hat{g}_{j\phi_j}\left(v_{jn}\right) \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \text{ and}$$

$$\mathrm{V}(\Phi) = \begin{bmatrix} n^{-1}\sum_{j=1}^{p} W_{\phi_j^1}\left(v_{j1}\right) & n^{-1}\sum_{j=1}^{p} W_{\phi_j^2}\left(v_{j1}\right) & \cdots & n^{-1}\sum_{j=1}^{p} W_{\phi_j^n}\left(v_{j1}\right) \\ n^{-1}\sum_{j=1}^{p} W_{\phi_j^1}\left(v_{j2}\right) & n^{-1}\sum_{j=1}^{p} W_{\phi_j^2}\left(v_{j2}\right) & \cdots & n^{-1}\sum_{j=1}^{p} W_{\phi_j^n}\left(v_{j2}\right) \\ \vdots & \vdots & \ddots & \\ n^{-1}\sum_{j=1}^{p} W_{\phi_j^1}\left(v_{jn}\right) & n^{-1}\sum_{j=1}^{p} W_{\phi_j^2}\left(v_{jn}\right) & \cdots & n^{-1}\sum_{j=1}^{p} W_{\phi_j^n}\left(v_{jn}\right) \end{bmatrix} \quad (8)$$

With a vector of $\hat{g}_{j\phi_j}\left(v_j\right)$ in size n x 1, y vector in size n x l, and matrix $\mathrm{V}(\Phi)$ in size *n x n*

Lemma 2. If the components of the Fourier series function in equation (5) are approximated by equation (3), then

$$\sum_{i=1}^{q} \mathrm{T}_s\left(t_s\right) = \mathrm{X}a \quad (9)$$

where X = [X_1, X_2, … X_p].

$$\mathrm{a} = [\mathrm{a'}_1 \ \mathrm{a'}_2 \ \ldots \ \mathrm{a'}_p]' \text{, dan } \mathrm{T_s}(t_s) = \begin{bmatrix} \mathrm{T_s}(t_{s1}) \\ \mathrm{T_s}(t_{s2}) \\ \vdots \\ \mathrm{T_s}(t_{sn}) \end{bmatrix} \quad (10)$$

With a vector $T_s(t_s)$ in size n x 1, matrix a is in size *p(M*+2) x l and matrix X is in size *nx(p(M* + 2)). Then

$$X_s = \begin{bmatrix} t_{s1} & 1 & \cos t_{s1} & \cos 2t_{s1} & \cdots & \cos Mt_{s1} \\ t_{s2} & 1 & \cos t_{s2} & \cos 2t_{s2} & \cdots & \cos Mt_{s2} \\ t_{s3} & 1 & \cos t_{s2} & \cos 2t_{s3} & \cdots & \cos Mt_{s3} \\ \vdots & \vdots & \vdots & & \cdots & \vdots \\ t_{sn} & 1 & \cos t_{sn} & \cos 2t_{sn} & \cdots & \cos Mt_{sn} \end{bmatrix}$$

and

$$\mathrm{a}_s = \begin{bmatrix} b_s \\ \frac{1}{2}a_{0s} \\ a_{1s} \\ \vdots \\ a_{Ms} \end{bmatrix} \quad (11)$$

With an X matrix in size $n\times(M+2)$, and a_s vector in size $(M+2)\times 1$.

One of the optimization methods in statistics is the *Penalized Least Squares* (PLS) method. This method combines the goodness of fit with the penalty function which both of them is controlled by the smoothing parameter. The PLS method can be written as follow:

$$\min_a \left\{ R(\mathrm{a}) + \sum_{s=1}^{q} \lambda_s J(T_s(t_{si})) \right\} \quad (12)$$

where R(a) is the goodness of fit, the penalty function of $J(T_s(t_{si}))$ and the smoothing parameter λ_s.

Lemma 3. If the regression model is given by equation (4), the kernel function is given by Lemma 2.1 and the Fourier series function is given by Lemma 2.2, then the goodness of fit function is

$$R(\mathrm{a}) = n^{-1}\left\| (\mathrm{I} - \mathrm{V}(\Phi))\mathrm{y} - \mathrm{Xa} \right\|^2 \quad (13)$$

Where, I is an identity matrix of size *n x n*.

Lemma 4. If the Fourier series function is given by Lemma 2, then the penalty function and smoothing parameter for the PLS method are

$$\sum_{s=1}^{q} \lambda_s J(T_s(t_{si})) = \mathrm{a}'\mathrm{D}(\lambda)\mathrm{a} \quad (14)$$

where $\mathrm{D}(\lambda) = diag\,(\mathrm{D}_1(\lambda)\ \mathrm{D}_2(\lambda)\ \ldots\ \mathrm{D}_p(\lambda))$ and $\mathrm{D}_s(\lambda) = diag\,(0\ \ 0\ \ \lambda_s 1^4\ \lambda_s 2^4\ \ldots\ \lambda_s M^4)$. The D(λ) matrix is (*p x* (*M* + 2)) x (*p x* (*M* + 2)) and the D(λ) matrix is (M + 2) x (M + 2).

Theorem 1. If the goodness of function is given by Lemma 3 while the penalty function and smoothing parameter are given by Lemma 4, then the estimation for parameter a is obtained from optimization:

$$\min_a \left\{ R(\mathrm{a}) + \sum_{s=1}^{q} \lambda_s J(T_s(t_{si})) \right\} = \min_a \{Q(a)\} \quad (15)$$

which

Q(a) = $n^{-1}\left\| \left(\mathrm{I} - \mathrm{V}(\Phi)\right)\mathrm{y} - \mathrm{Xa} \right\|^2$ + a'D(λ)a

Theorem 2. If the estimate parameter a is given by Theorem 1 then the new nonparametric regression estimator $\mu(v_i, t_i)$ is given by

$$\hat{\mu}_{\Phi 2M}(v_i, t_i) = \sum_{j=1}^{p} \hat{g}_{j\phi_j}(v_j) + \sum_{j=1}^{q} \hat{h},_{\lambda,M}(t_s) \qquad (16)$$

which $\sum_{j=1}^{q} \hat{h},_{\lambda,M}(t_s) = \mathrm{X}\hat{a}(\lambda)$, and

$$\hat{a}(\lambda) = \left(\mathrm{X'X} + n\mathrm{D}(\lambda)\right)^{-1} \mathrm{X'}\left(\mathrm{I} - \mathrm{V}(\Phi)\right)\mathrm{y}$$

The matrix V(Φ) is given by Lemma 2.1, the matrix X is given by Lemma 2 and I is the identity matrix.

Lemma 4. If the estimator $\hat{\mu}_{\Phi 2M}(v_i, t_i)$, $\sum_{j=1}^{p} \hat{g}_{j\phi_j}(v_j)$, $\sum_{j=1}^{q} \hat{h},_{\lambda,M}(t_s)$, and $\hat{a}(\lambda)$ is given by Theorem 2.2, then

$$\sum_{j=1}^{q} \hat{h},_{\lambda,M}(t_s) = \mathrm{S}(\Phi, \lambda, \mathrm{M})\mathrm{y} \text{ and}$$

$$\hat{\mu}_{\Phi,\lambda,M}(v_i, t_i) = \mathrm{Z}(\Phi, \lambda, \mathrm{M})\mathrm{y}$$

which

$$\mathrm{Z}(\Phi, \lambda, \mathrm{M})\mathrm{y} = \mathrm{X}\left(\mathrm{X'X} + n\mathrm{D}(\lambda)\right)^{-1} \mathrm{X'}\left(\mathrm{I} - \mathrm{V}(\Phi)\right)\mathrm{y} \text{ and}$$

$$\mathrm{Z}(\Phi, \lambda, \mathrm{M})\mathrm{y} = \left(\mathrm{V}(\Phi) + \mathrm{S}(\Phi, \lambda, \mathrm{M})\right)\mathrm{y}$$

Nonparametric regression curve estimators of Kernel basis and Fourier series $\hat{\mu}_{\Phi,\lambda,M}(v_i, t_i)$ are highly dependent on bandwidth, M oscillation parameter, and smoothing parameter. This research uses the Generalized Cross Validation (GCV) method to obtain the optimum bandwidth, oscillation parameter M, and smoothing parameter, as follows:

$$GCV(\Phi,\lambda,M) = \frac{MSE(\Phi, \lambda, M)}{\left(n^{-1} trace\left(\mathrm{I} - \mathrm{Z}(\Phi, \lambda, M)\right)\right)^2}$$

which

$MSE(\Phi, \lambda, M) = n^{-1} y'\left(\mathrm{I} - \mathrm{Z}(\Phi, \lambda, M)\right)' \left(\mathrm{I} - \mathrm{Z}(\Phi, \lambda, M)\right)\mathrm{y}$. Optimum *Bandwidth* $\Phi_{opt} = \left(\phi_{1(opt)}, \phi_{2(opt)}, \ldots, \phi_{p(opt)}\right)$ optimum smoothing parameter $\lambda_{(opt)} = \left(\lambda_{1(opt)}, \lambda_{2(opt)}, \ldots, \lambda_{q(opt)}\right)$ and optimum oscillation parameter *M* are obtained from GCV $\left(\Phi_{(opt)}, \lambda_{(opt)}, M_{(opt)}\right) = \underset{\Phi,\lambda,M}{Min}\left(GCV(\Phi, \lambda, M)\right)$.

Modeling the Percentage of Poor Population in Indonesia Using Kernel Nonparametric Regression and Fourier Series

The percentage of poor people in Bali in 2020 reached 11.47 percent. There were 8 regencies/cities that have a high percentage of poor people compared to other regency, namely Karangasem (19.27 percent), Klungkung (18.01 percent), Jembrana (17.75 percent). The average percentage of poor people for the three regencies was 18.3 percent. The predictor variable of the average length of schooling is obtained 7.80, open unemployment rate obtained an average of 5.65 and Literacy Rate obtained an average of 88.74.

Furthermore, a GCV comparison was also carried out to determine the predictor variables following the kernel regression and the Fourier series. Of all possibilities, the minimum GCV is obtained if the kernel components are X1, X3 and the Fourier series component is X2. Thus, the data of poor people percentage is approximated by a kernel-based nonparametric regression and Fourier series. There are predictor variables for average length of schooling and Literacy Rate approximated by the kernel function, while the predictor variable for open unemployment rate is approximated by the Fourier series function. Furthermore, variables which are approximated by the kernel function are symbolized by *v* and variables which are approximated by the Fourier series function are symbolized by *t*. A complete list of each predictor variable can be seen in Table 4.1

Table 1. Used Predictor and Approaches Variables

No	Predictor Variables	Approaches	Symbol
1	Average Length of Schooling	Kernel Function	v_1
2	Open Unemployment Rate	Fourier Series Function	t_1
3	Literacy Rate	Kernel Function	v_2

Determination of bandwidth values, smoothing and oscillation parameters also uses GCV value comparison. The following is a comparison table of GCV values for each model with oscillation parameters 1, 2, 3, 4 and 5.

Table 2. Comparison of GCV values for the model with oscillation parameters

No.	ϕ_1	ϕ_2	λ	M	GCV
1	0,047	4,146	10^{-3}	1	42,888
2	0,047	4,146	10^{-3}	2	43439
3	0,047	4,146	1	3	43,617
4	0,047	4,146	10^{-3}	4	43,063
5	0,047	4,146	10^{-3}	5	43,419

Based on Table 2, the minimum GCV is 42.888, so that ϕ_1 = 0.047; ϕ_2 = 4.146; M = I and λ = 10-3. Estimation of the Fourier series components is presented in Table 3 below.

Table 3. Estimation of the Fourier Series Components

Parameter	E(λ)	$a_0(\lambda)$	$a_1(\lambda)$	λ
Estimation	-0,148	2,140	1,799	0,001

The kernel-based nonparametric regression model and the Fourier series obtained result as follows:

$$\hat{y}_i = \sum_{i=1}^{100} \frac{\frac{1}{0,047} K\left(\frac{v_1 - v_{1i}}{0,047}\right)}{\sum_{i=1}^{100} \frac{1}{0,047} K\left(\frac{v_1 - v_{1i}}{0,047}\right)} y_i + \sum_{i=1}^{100} \frac{\frac{1}{4,146} K\left(\frac{v_2 - v_{2i}}{4,146}\right)}{\sum_{i=1}^{100} \frac{1}{4,146} K\left(\frac{v_j - v_{2i}}{4,146}\right)} y_i - 0,148 t_{1i} + 1,070 + 1799 \cos t_{1i}$$

Model for data pattern that follows the Fourier series curve can be interpreted. Meanwhile, the data pattern that follows the kernel function cannot be interpreted. The following is a model for the Open Unemployment Rate variable:

$$\hat{y}_i = -0,148_{1i} + 1,070 + 1,799 \cos t_{1i}$$

which

$$c = \sum_{i=1}^{100} \frac{\frac{1}{0,047} K\left(\frac{v_1 - v_{1i}}{0,047}\right)}{\sum_{i=1}^{100} \frac{1}{0,047} K\left(\frac{v_1 - v_{1i}}{0,047}\right)} y_i + \sum_{i=1}^{100} \frac{\frac{1}{4,146} K\left(\frac{v_2 - v_{2i}}{4,146}\right)}{\sum_{i=1}^{100} \frac{1}{4,146} K\left(\frac{v_j - v_{2i}}{4,146}\right)} y_i$$

Therefore, by assuming that other than constant Open Unemployment Rate data, when the Open Unemployment Rate increases by 1 point, the percentage of poor people will increase by 1.90 percent.

Modeling data on the percentage of poor people using the kernel-based estimator and the Fourier series produces an R^2 value of 62.78 percent. From the modeling results obtained the Open Unemployment Rate has a positive effect on the percentage of poor people in Bali. Thus, to prevent an increase in the percentage of poor people, the government must make efforts to reduce the increase of Open Unemployment Rate. Efforts that can be taken are for example opening new jobs, providing capital loans to open businesses independently and so on.

The open unemployment rate significantly affects the poverty rate. Open unemployment refers to the part of the workforce that is actively looking for work but has not yet found a job. When the unemployment rate is high, more people lose stable income, which then worsens the economic conditions of individuals and their families. This ultimately increases the poverty rate. The open unemployment rate and the poverty rate are closely related, because high unemployment tends to increase the poverty rate. When many people are unemployed, they lose a stable source of income, which can ultimately lead to poverty. Reducing unemployment is one effective way to reduce poverty. The government has an important role in designing policies that can reduce unemployment and its impact on poverty. Here are some policies that are often implemented by the government:

1. Job Creation Program
    a. Labor-Intensive Program: The government can launch a labor-intensive program that aims to create jobs for the community, especially in the infrastructure sector and public projects. Examples include the construction of roads, bridges, or environmental improvement projects.
    b. Incentives for the MSME Sector: Encourage the growth of Micro, Small, and Medium Enterprises (MSMEs) through tax incentive policies, subsidies, and entrepreneurship training. MSMEs are often the largest absorbers of labor in a country's economy.
2. Skills Training and Vocational Education
    a. Upskilling and Reskilling: The government can focus on retraining unemployed workers by providing new skills needed by the market. This training can be done in the form of vocational education programs and technology-based training.
    b. Industry-Based Education: Ensuring that the education and training system is directly linked to the needs of the labor market, so that graduates have relevant skills and are more easily absorbed by industry.
3. Sectoral Economic Development
    a. Development of the Agriculture, Fisheries, and Tourism Sectors: In rural areas, unemployment is often caused by a lack of jobs in the formal sector. The government can focus on developing potential sectors such as agriculture, fisheries, and tourism that can create many jobs.
    b. Support for the Creative and Digital Economy: Encouraging the creative and digital economy can also be an effective policy. The government can provide assistance in the form of training, funding, and digital infrastructure to encourage more job opportunities in this sector.
4. Monetary and Fiscal Policy to Encourage Investment
    a. Foreign and Domestic Investment: Encouraging investment through conducive fiscal and monetary policies can open up more jobs. For example, the government can provide tax incentives to companies that open new jobs or invest in underdeveloped areas.
    b. Village Fund Program: One of the policies in Indonesia is the Village Fund which is aimed at infrastructure development in rural areas, with the hope of reducing local unemployment through job creation in local development projects.
5. Social Protection
    a. Social Assistance (Bansos): Social assistance programs such as the Family Hope Program (PKH), Non-Cash Food Assistance (BPNT), and Pre-Employment Cards are government efforts to maintain the purchasing power of poor and vulnerable people so that they still have access to basic needs such as education and health, even though they are unemployed.
    b. Subsidies and Welfare: Programs such as subsidies for basic necessities, electricity, or education can help the poor reduce their spending burden, even though they are unemployed.

6. Employment Policy and Labor Protection
    a. Minimum Wage Policy: Setting a decent minimum wage can help prevent poor workers or those who work from remaining in poverty. However, this policy must be balanced with business capabilities so as not to cause a reduction in the workforce.
    b. Employment Flexibility: Encourage more flexible employment policies, such as supporting freelancers and informal workers with social security and access to legal protection.
7. Infrastructure and Urbanization Policy
    a. Infrastructure Development: Massive infrastructure development in various regions can open up jobs and increase economic mobility, especially in disadvantaged areas.
    b. Industrial Area Development: Building industrial areas and accelerating development in small towns and rural areas can be a long-term solution to address unemployment outside big cities.

Government policies in dealing with unemployment must be holistic and involve various sectors. Not only creating jobs through physical development, but also by improving the quality of human resources through education, training, and support for industries that can absorb a lot of workers. By reducing unemployment, the government can directly reduce poverty rates and improve people's welfare.

CONCLUSION

It is given paired data $(v_{1i}, v_{2i}, \ldots v_{pi}, t_{1i}, t_{2i}, \ldots t_{qi}, y_i), i = 1, 2, \ldots, n,$ following a mixed regression model, that is $y = \mu(v_i, t_i) + \varepsilon = \sum_{j=1}^{p} g_j(v_j) + \sum^{q} h_s(t_s) + \varepsilon$ Component $g_j(v_j)$ is approximated by kernel function and component $h_s(t_s)$ is approximated by Fourier series function. Error ε is assumed to be normally distributed with zero mean and constant variance. The function estimator results obtained: $\hat{y} = \hat{\mu}_{\Phi 2M}(v_i, t_i) = Z(\Phi, \lambda, M)y$

Which

$Z(\Phi, \lambda, M) = (V(\Phi) + S(\Phi, \lambda, M))$, and

$S(\Phi, \lambda, M) = X(X'X + nD(\lambda))^{-1} X'(I - V(\Phi))$

The non-parametric regression model based on the kernel and Fourier series is applied to the percentage of poor people in Bali Province in 2020 with the response variable is the percentage of poor people (y), while the predictor variables are average length of schooling (v_1) and literacy rate (v_2) and the Open Unemployment Rate (t_1). From the results of the model, it is interpreted that the percentage of poor people will increase by 1.90 percent for every 1 increase in the Open Unemployment Rate. This modeling produces an R^2 of 0.6278. This means that the used variables can explain the model by 62.78 percent.

This research uses the basis of two nonparametric regression curves. Subsequent studies may use a basis of more than two nonparametric components that match the data pattern. In future research, it is expected to add predictor variables so that they can better represent the percentage of poor people. In addition, this research was developed using other kernel components such as the uniform kernel, triangle and so on. It is recommended that further research add predictor variables that influence the percentage of poverty in Bali, such as the rate of economic growth and the rate of inflation.

ACKNOWLEDGMENTS

The authors thank the editor and the reviewers for their constructive and helpful comments.

AUTHOR CONTRIBUTIONS

Conceptualization, I.W.S., and N.P.A.M.M.; Methodology, I.W.S., and N.P.A.M.M; Software, I.W.S., and N.P.A.M.M.; Validation, I.W.S., and N.P.A.M.M.; Formal Analysis, I.W.S., and N.P.A.M.M.; Investigation, I.W.S., and N.P.A.M.M.; Resources, I.W.S., and N.P.A.M.M.; Data

Curation, N. M.S.S.; Writing – Original Draft Preparation, D.P.W; Writing – Review & Editing, N. M.S.S.; Visualization, N.M.S.S.; Supervision, I.W.S.,; Project Administration, I.W.S.,; Funding Acquisition, I.W.S.

CONFLICTS OF INTEREST

The author(s) declare no conflict of interest.

REFERENCES

Ahmad, H. B., Asaad, R. R., Almufti, S. M., Hani, A. A., Sallow, A. B., Zeebaree, S. R. M. (2024). Smart Home Energy Saving With Big Data And Machine Learning. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, 8(1), 11-20. https://doi.org/10.22437/jiituj.v8i1.32598.

Abdulrahman, S. M., Hani, A. A., Zeebaree, S. R., Asaad, R. R., Majeed, D. A., Sallow, A. B., & Ahmad, H. B. (2024). Intelligent home iot devices: an exploration of machine learning-based networked traffic investigation. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, *8*(1), 1-10. https://doi.org/10.22437/jiituj.v8i1.32767.

Afifah, N., Budiantara, I.N., and Latra, I.N. (2017), "Mixed Estimator of Kernel and Fourier Series in Semiparametric Regression", *Journal of Physics: Conference Series*, 855(1). https://doi.org/10.1088/1742-6596/855/1/012002.

Berlinet, A.; Thomas-Agnan, C. (2004). Reproducing Kernel Hilbert Spaces in Probability and Statistics; Kluwer Academic: Norwell, MA, USA. https://scholar.archive.org/work/lcildyh235gevmzhvq4gyx7ubq/access/wayback/http://tocs.ulb.tu-darmstadt.de/118341111.pdf .

BPS. Penghitungan dan Analisis Kemiskinan Makro Indonesia Tahun 2012. (2012). Jakarta. https://media.neliti.com/media/publications/49346-ID-penghitungan-dan-analisis-kemiskinan-makro-indonesia-tahun-2012.pdf .

Cadiz, M. C. D., Manuel, L. A. F., Reyes, M. M., Natividad, L. R., & Ibarra, F. P. (2024). Technology integration in Philippine higher education: A content-based bibliometric analysis. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, *8*(1), 35-47. https://doi.org/10.22437/jiituj.v8i1.31807.

Cheng, M.-Y.; Huang, T.; Liu, P.; Peng, H. (2018). Bias reduction for nonparametric and semiparametric regression models. Stat. Sin. 28, 2749–2770. https://doi.org/10.5705/ss.202017.0058.

Cheruiyot, L.R. (2020). Local linear regression estimator on the boundary correction in nonparametric regression estimation. J. Stat. Theory Appl. 19, 460–471. https://doi.org/10.2991/jsta.d.201016.001.

Choerunnisa, R. A., Dewi, R. R., Bariklana, M., & Widodo, E. (2021). Analisis faktor yang mempengaruhi tingkat produksi jahe di indonesia menggunakan metode regresi linier berganda. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, *5*(2), 231-242. https://doi.org/10.22437/jiituj.v5i2.15958.

Dessi, L. C., & Shah, M. (2023). Application of the numbered head together type cooperative learning model to improve student learning outcomes in mathematics subjects. *Interval: Indonesian Journal of Mathematical Education*, *1*(2), 67-72. https://doi.org/10.37251/ijome.v1i2.773.

Eubank, R.L. (1999). *Nonparametric Regression and Spline Smoothing*, 2nd ed.; Marcel Dekker, Inc.: New York, NY, USA.

Funke, B. (2024). *On Uniform Consistency of Nonparametric Estimators Smoothed by the Gamma Kernel*.

Ge, C., & Braun, W. J. (2024). *Quick and Simple Kernel Differential Equation Regression Estimators for Data with Sparse Design*.

Helida, Y., Ching, C. P., & Oyewo, A. (2023). Development of a simple stirling engine demonstration tool on the subject of thermodynamics. *Journal of Educational Technology and Learning Creativity*, *1*(2), 59-67. https://doi.org/10.37251/jetlc.v1i2.790.

Hidayat, R., Budiantara, I. N., Otok, B. W., & Ratnasari, V. (2020). The regression curve estimation by using mixed smoothing spline and kernel (MsS-K) model. *Communications in Statistics - Theory and Methods*, 1–12. https://doi.org/10.1080/03610926.2019.1710201.

Ismail, Y., Mir, S.A., and Nazir, N. (2018), Utilization of Parametric and Nonparametric Regression Models for Production, Productivity and Area Trends of Apple (Malus domestica) in Jammu and Kashmir, India, *International Journal of Current Microbiology and Applied Sciences.* 7(04), 267–276.

Kalluci, E. (2024). *Modeling inflation dynamics using the logistic model: insights*. *8*(1), 364–378.

Kohn, R., Ansley, C.F., and Tharm, D., (2014), The Performance of Cross Validation and Maximum Likelihood Estimator of Spline Smooting Parameter. *Journal of The American Statistics Assosiations*, *86*(416), 1042–1050.

Kusuma, R. S. (2020). Improving students' basic asking skills by using the discovery learning model. *Tekno - Pedagogi : Jurnal Teknologi Pendidikan*, *10*(2), 8-13. https://doi.org/10.22437/teknopedagogi.v10i2.32743.

Linke, Y., Borisov, I., Ruzankin, P., Kutsenko, V., Yarovaya, E., & Shalnova, S. (2024). *Multivariate Universal Local Linear Kernel Estimators in Nonparametric Regression: Uniform Consistency*. 1–23.

Liu, X., and Preve, D. (2016). Measure of location-based estimators in simple linear regression. *Journal of Statistical Computation and Simulation*, *86*(9), 1771–1784. https://doi.org/10.1080/00949655.2015.1082131.

Lu, M.., Liu, Y., Li, C.-S. (2020). Efficient estimation of a linear transformation model for current status data via penalized splines. *Stat. Meth. Medic. Res. 29*, 3–14. https://doi.org/10.1177/0962280218820406.

Mariati, M.P.A.M., Budiantara, I.N., Ratnasari, V. (2021). The application of mixed smoothing spline and Fourier series model in nonparametric regression. *Symmetry. 13*, 2094. https://doi.org/10.3390/sym13112094.

Mozumder, S. U., Rutherford, M., & Lambert, P. (2017). Direct likelihood inference on the cause-specific cumulative incidence function: A flexible parametric regression modelling approach. *Statistics in Medicine. 37*, 1–16. https://doi.org/10.1002/sim.7498.

Nurcahayani, H., Budiantara, I. N., Zain, I. (2021). The Curve Estimation of Combined Truncated Spline and Fourier Series Estimators for Multiresponse Nonparametric Regression. *Mathematics, 9*(10), 1141; https://doi.org/10.3390/math9101141 .

Osmani, F., Hajizadeh, E., & Mansouri, P. (2019). Kernel and regression spline smoothing techniques to estimate coefficient in rates model and its application in psoriasis. *Medical Journal of the Islamic Republic of Iran*, *33*, 90. https://doi.org/10.47176/mjiri.33.90 .

Respati, S., Isram, M., & Kusrini, S. (2022). *Estimation of Queue Length at Signalized Intersections Using Artificial Neural Network*. *6*, 201–213.

Storlie, C. B., Lane, W. A., Ryan, E. M., Gattiker, J. R., and Higdon, D. M. (2015). Calibration of Computational Models with Categorical Parameters and Correlated Outputs via Bayesian Smoothing Spline ANOVA. *Journal of the American Statistical Association*, 110 (509), 68–82. http://doi.org/10.1080/01621459.2014.979993.

Sudiarsa, I W., Budiantara, I. N., Suhartono; Pumami, S. W. (2015). Combined Estimator Fourier Series and Spline Truncated in Multivariabel Nonparametrik Regression. *Applied Mathematical Science. 9*(100), 4997-5010. http://dx.doi.org/10.12988/ams.2015.55394 .

Suwarni, R. (2021). Analysis the process of observing class iv students in thematic learning in primary schools. *Tekno - Pedagogi : Jurnal Teknologi Pendidikan*, *11*(1), 26-32. https://doi.org/10.22437/teknopedagogi.v11i1.32717.

Yilmaz, E., Ahmed, S.E., Aydin, D. (2020) A-Spline regression for fitting a nonparametric regression function with censored data. *Stats. 3*, 11. https://doi.org/10.3390/stats3020011 .

Wang, Y., & Ke, C. (2009). *Smoothing Spline Semiparametric Nonlinear Regression Models*. *18*(1), 165–183. https://doi.org/10.1198/jcgs.2009.0010.

Yu, X., Lin, J., Jiang, F., Du, J., & Han, J. (2018). A Cross-Domain Collaborative Filtering Algorithm Based on Feature Construction and Locally Weighted Linear Regression. *Computational Intelligence and Neuroscience*, 2018, hal.1–12.

Yohanie, D. D., Botchway, G. A., Nkhwalume, A. A., & Arrazaki, M. (2023). Thinking process of mathematics education students in problem solving proof. *Interval: Indonesian Journal of Mathematical Education*, *1*(1), 24-29. https://doi.org/10.37251/ijome.v1i1.611.

Zakiyah, Z., Boonma, K., & Collado, R. (2024). Physics learning innovation: Song and animation-based media as a learning solution for mirrors and lenses for junior high school students. *Journal of Educational Technology and Learning Creativity*, *2*(2), 54-62. https://doi.org/10.37251/jetlc.v2i2.1062.

Zeng, X., Xia, Y. (2019). Asymptotic distribution for regression in a symmetric periodic Gaussian kernel Hilbert space. *Stat. Sin*. 29, 1007–1024. https://doi.org/10.5705/ss.202016.0440.