

Fourier Series Estimator For Estimation of Birespon Nonparametric Regression Curves

Agustini Tripena¹, Niken Larasati², Agus Sugandha³, Putri Raffikah⁴

^{1,2,3,4}Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Jenderal Soedirman, Purwokerto, Indonesia.

Corresponding Author: agustini.brsurbakti@unsoed.ac.id

Abstract

In the last decade the Fourier Series estimator in nonparametric regression (one response) has received a lot of attention from researchers because of its flexibility. In this case, the Derer Fourier estimator will be developed in two response nonparametric regression (Birespon). Given pairwise data (t_{1j}, y_{1j}) , $j = 1, 2, \dots, n_1$ and (t_{2j}, y_{2j}) , $j = 1, 2, \dots, n_2$. The relationship between t_{1j} , t_{2j} , y_{1j} , and y_{2j} follows the Birespon nonparametric regression model:

$$y_{1j} = f_1(t_{1j}) + \varepsilon_{1j} \text{ And } y_{2j} = f_2(t_{2j}) + \varepsilon_{2j}.$$

The shape of the regression curve f_1 is f_2 unknown and is assumed to be contained in a continuous function space $\mathbb{C}(0, \pi)$. Random ε_{1j} errors are mutually independent with zero mean and variance σ_1^2 , and ε_{1j} are also mutually independent with zero mean and variance σ_2^2 . Errors are random ε_{1j} and ε_{2j} correlated with each other $\text{Cor}(\varepsilon_{1j}, \varepsilon_{2j}) = \rho$. The regression curve $f_1(t)$ is $f_2(t)$ successively approximated by a continuous and differentiable function:

$$d_1(t_{1j}) = \gamma_1 t_{1j} + \frac{1}{2} \alpha_{01} + \sum_{k=1}^K \alpha_{1k} \cos kt_{1j}, \text{ And}$$

$$d_2(t_{2j}) = \gamma_2 t_{2j} + \frac{1}{2} \alpha_{02} + \sum_{k=1}^K \alpha_{2k} \cos kt_{2j},$$

$\gamma_1 \in \mathbb{R}$, $\gamma_2 \in \mathbb{R}$, $\alpha_{01} \in \mathbb{R}$, $\alpha_{02} \in \mathbb{R}$, $\alpha_{1k} \in \mathbb{R}$, $k = 1, 2, \dots, n_1$, $\alpha_{2k} \in \mathbb{R}$, $k = 1, 2, \dots, n_2$ are unknown parameters in the Fourier Series model. Estimated Birespon nonparametric regression curve $f_1(t)$ and $f_2(t)$ obtained from completing Weighted Penalized Least Square (PLST) optimization:

$$\underset{f_k \in \mathbb{C}(0, \pi), k=1,2}{\text{Min}} \left\{ \frac{1}{(n_1 + n_2)} (\tilde{y} - \tilde{f})' W(\rho, \sigma_1^2, \sigma_2^2) (\tilde{y} - \tilde{f}) + \lambda_1 \int_0^\pi [f_1^n(t_1)]^2 dt_1 + \lambda_2 \int_0^\pi [f_2^n(t_2)]^2 dt_2 \right\}.$$

The solution to PLST optimization is in the form of a Fourier Series estimator which can be presented in the form:

$$\hat{\tilde{f}}_\lambda(t) = \begin{pmatrix} \hat{f}_{\lambda_1}(t_1) \\ \hat{f}_{\lambda_2}(t_2) \end{pmatrix} = B(\lambda_1, \lambda_2) \begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{pmatrix}.$$

The Fourier Series Estimator in Birespon nonparametric regression is a bias for nonparametric regression curves $\tilde{f}(t) = \begin{pmatrix} f_1(t_1) \\ f_2(t_2) \end{pmatrix}$. Even though it is biased, this estimator is a linear estimator, which is very helpful in building statistical inference for nonparametric Birespon regression curves.

Keywords: Fourier Series Estimator, Birespon Nonparametric Regression, Weighted Least Square Penalization.

INTRODUCTION

Given a nonparametric regression model (one response) $y_j = f(t_j) + \varepsilon_j, t_j \in [a, b], j = 1, 2, \dots, n$. the shape of the regression curve f is assumed to be unknown and contained in a certain function space, such as space *Sobolev* $W_2^m[a, b]$, space of continuous functions $\mathbb{C}(0, \pi)$, *Hilbert space*, *Entropy space*, and others. Making this assumption depends on the *smooth nature* of the function whose pattern is unknown (Budiantara, 2000; 1999). Random ε_j error is assumed to have an independent distribution with zero mean and variance σ^2 . The problem in nonparametric regression is how to get the estimator form for the regression curve f . In nonparametric regression, generally the nonparametric estimator form is obtained using a family function approach (eg orthogonal series (Eubank, 1988), Wavelets (Antoniadis, *et al* . (2001), Kernel (Eubank, 1988) and Spline (Budiantara, 2006). Based on this view estimation theory, the family function approach is generally difficult to obtain the basic statistical basis used in the family function approach (Budiantara, 2001; 2002; 2005).

Wahba (1990), Wang (1998) and Budiantara (2002), no longer use the family function approach, but use *Penalized optimization* which combines *Goodness of fit* and *penalty*. The nonparametric regression estimator is obtained from completing the optimization:

$$\underset{f_k \in H}{\text{Min}} \{R(f) + \omega J(f)\}, \text{ for a space } H.$$

Quantity $R(f)$ and $J(f)$ respectively express *goodness of fit* and a measure of smoothness of function (*penalty*). The smoothing parameter (*bandwidth*) ω controls $R(f)$ and $J(f)$. The nonparametric regression group that often uses this *penalized approach* is Spline. Because of these advantages, several other regression groups such as the Fourier Series have also begun to use the *penalized approach* , but still combine it with the family function approach (Bilodeau, 1992; Tripenda and Budiantara, 2006).

The Fourier Series Estimator in single response nonparametric regression has received a lot of attention in recent years from several nonparametric regression researchers. This Fourier Series Estimator is generally used when the pattern of the data being investigated is unknown and there is no seasonal pattern trend (Tripena and Budiantara, 2006 and Bilodeau, 1992). In the Fourier Series estimator, a single-response nonparametric regression curve f is assumed to be unknown and contained in a continuous function space $\mathbb{C}(0, \pi)$. Random ε_j error is assumed to have an independent normal distribution with zero mean and variance σ^2 . Because $f(t)$ it is continuous on an interval $(0, \pi)$, it can be approximated by a function $F(t) = \gamma t + \frac{1}{2}\alpha_0 + \sum_{i=1}^K \alpha_i \cos it$, which $\gamma, \alpha_0, \alpha_i, i = 1, 2, \dots, K$ is the model parameters (Bilodeau, 1992). Based on this Fourier Series model, Bilodeau (1992) provides a Fourier Series estimator for a single response nonparametric regression curve, in the form:

$$\hat{F}_\omega(t) = \hat{\gamma}(\omega)t + \frac{1}{2}\hat{\alpha}_0(\omega) + \sum_{i=1}^K \hat{\alpha}_i(\omega) \cos it,$$

obtained $\hat{\alpha}(\omega) = (\hat{\gamma}(\omega), \hat{\alpha}_0(\omega), \hat{\alpha}_1(\omega), \dots, \hat{\alpha}_{iK}(\omega))'$ from the equation:

$$\hat{\alpha}(\omega) = (n^{-1}X'X + \omega G)^{-1}n^{-1}X'y,$$

where is G the diagonal matrix and X a coefficient matrix. Furthermore, Tripena and Budiantara (2016), stated that if the random *error* from a single-response nonparametric regression model is normally distributed, then the Fourier Series estimator $\hat{\alpha}(\omega)$ and $\hat{F}_\omega(t)$ that given by Bilodeau (1992), have property (i). $\hat{F}_\omega(t)$ is a bias estimator for nonparametric regression curves $f(t)$. (ii). $\hat{F}_\omega(t)$ is a class of linear estimators in observation, and (iii). $\hat{\alpha}(\omega)$ and $\hat{F}_\omega(t)$ each has a normal distribution.

In several practical problems, we often find regression models that have more than one response (Response), the pattern of the regression curve is unclear and unknown, and there are indications of seasonality. Theoretically, the Fourier Series estimator in Birespon nonparametric regression is a generalization of the Fourier Series estimator in single response nonparametric regression. Research regarding the Fourier Series estimator in nonparametric Birespon regression

has not been widely developed by researchers. In this article, we will give the form of the Fourier Series estimator in Birespon nonparametric regression and the properties of this estimator.

RESULT

1. Biresponse Nonparametric Regression and Variance-Covariance Matrix Structure

Given paired data (t_{1j}, y_{1j}) , $j = 1, 2, \dots, n_1$ and (t_{2j}, y_{2j}) , $j = 1, 2, \dots, n_2$. The relationship between t_{1j}, t_{2j}, y_{1j} and y_{2j} is assumed to follow a nonparametric regression model with two responses (Birespon):

$$\begin{aligned} y_{1j} &= f_1(t_{1j}) + \varepsilon_{1j}, j = 1, 2, \dots, n_1 \\ y_{2j} &= f_2(t_{2j}) + \varepsilon_{2j}, j = 1, 2, \dots, n_2. \end{aligned} \quad (1)$$

The shape of the regression curve f_1 is f_2 unknown and is assumed to be contained in a continuous function space $\mathbb{C}(0, \pi)$, where:

$$\mathbb{C}(0, \pi) = \{g; g \text{ kotinu pada interval } (0, \pi)\}.$$

Random $\varepsilon_{1j}, j = 1, 2, \dots, n_1$ errors are mutually independent with zero mean and variance σ^2 , random $\varepsilon_{2j}, j = 1, 2, \dots, n_2$ errors are also mutually independent with zero mean and variance σ^2 . Errors are random $\varepsilon_{1j}, j = 1, 2, \dots, n_1$ and $\varepsilon_{2j}, j = 1, 2, \dots, n_2$ correlated with each other $Cor(\varepsilon_{1j}, \varepsilon_{2j}) = \rho$. The variance-covariance matrix structure of the error can be written as follows:

$$W^{-1}(\rho, \sigma_1^2, \sigma_2^2) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 & \sigma_{12} & \sigma_{12} & \cdots & \sigma_{12} \\ 0 & \sigma_1^2 & 0 & \vdots & \sigma_{12} & \sigma_{12} & \cdots & \sigma_{12} \\ \vdots & \vdots & \ddots & 0 & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_1^2 & \sigma_{12} & \sigma_{12} & \cdots & \sigma_{12} \\ \sigma_{21} & \sigma_{21} & \cdots & \sigma_{21} & \sigma_2^2 & 0 & \cdots & 0 \\ \sigma_{21} & \sigma_{21} & \cdots & \sigma_{21} & 0 & \sigma_2^2 & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & 0 \\ \sigma_{21} & \sigma_{21} & \cdots & \sigma_{21} & 0 & 0 & \cdots & \sigma_2^2 \end{pmatrix}.$$

For example, given quantity $\theta = \sigma_1 \sigma_2$, $r = \frac{\sigma_1}{\sigma_2}$ and correlation $\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$, then covariance σ_{12} can be written as $\sigma_{12} = \rho \sigma_1 \sigma_2 = \rho \theta$. On the other hand, variance σ_1^2 and σ_2^2 sequence can be presented as:

$$\begin{aligned} \sigma_1^2 &= \sigma_1 \sigma_2 \left(\frac{\sigma_1}{\sigma_2} \right) = \theta r, \text{ And} \\ \sigma_2^2 &= \sigma_1 \sigma_2 \left(\frac{\sigma_2}{\sigma_1} \right) = \left(\frac{1}{\left(\frac{\sigma_2}{\sigma_1} \right)} \right) = \theta \left(\frac{1}{r} \right). \end{aligned}$$

Thus the variance-covariance matrix $W^{-1}(\rho, \sigma_1^2, \sigma_2^2)$ can be presented as:

$$W^{-1}(\rho, \sigma_1^2, \sigma_2^2) = \begin{pmatrix} \theta r & 0 & \dots & 0 & \theta \rho & \theta \rho & \dots & \theta \rho \\ 0 & \theta r & 0 & \vdots & \theta \rho & \theta \rho & \dots & \theta \rho \\ \vdots & \vdots & \ddots & 0 & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \theta r & \theta \rho & \theta \rho & \dots & \theta \rho \\ \theta \rho & \theta \rho & \dots & \theta \rho & \frac{\theta}{r} & 0 & \dots & 0 \\ \theta \rho & \theta \rho & \dots & \theta \rho & 0 & \frac{\theta}{r} & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \ddots & 0 \\ \theta \rho & \theta \rho & \dots & \theta \rho & 0 & 0 & \dots & \frac{\theta}{r} \end{pmatrix}.$$

The matrix $W^{-1}(\rho, \sigma_1^2, \sigma_2^2)$ can also be presented as:

$$W^{-1}(\rho, \sigma_1^2, \sigma_2^2) = \theta \begin{pmatrix} rI_{n_1} & \vdots & \rho J_{n_1 \times n_2} \\ \dots & \vdots & \dots \\ \rho J_{n_2 \times n_1} & \vdots & \frac{1}{r} I_{n_2} \end{pmatrix}, \quad (2)$$

with the matrix I_{n_1} and I_{n_2} respectively representing the Identity matrix of size $n_1 \times n_1$ and $n_2 \times n_2$,

as well as the matrix $J_{n_1 \times n_2} = \begin{pmatrix} 1 & \vdots & 1 \\ \dots & \vdots & \dots \\ 1 & \vdots & 1 \end{pmatrix}$.

Based on this variance-covariance matrix $W^{-1}(\rho, \sigma_1^2, \sigma_2^2)$, the form of the Bivariate Fourier Series estimator in nonparametric Birespon regression will be determined.

2. Bivariate Fourier Series Estimator in Birespon Nonparametric Regression

The nonparametric regression equation Birespon (1) can be expressed in matrix form as follows:

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ \dots \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \end{pmatrix} = \begin{pmatrix} f_1(t_{11}) \\ f_1(t_{12}) \\ \vdots \\ f_1(t_{1n_1}) \\ \dots \\ f_2(t_{21}) \\ f_2(t_{22}) \\ \vdots \\ f_2(t_{2n_2}) \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1n_1} \\ \dots \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2n_2} \end{pmatrix}. \quad (3)$$

Since the regression curves $f_1 \in \mathbb{C}(0, \pi)$ and $f_2 \in \mathbb{C}(0, \pi)$, then $f_1(t)$ and $f_2(t)$ can be approximated respectively by continuous and differentiable functions $d_1(t)$ and $d_2(t)$, namely:

$$f_1(t_{1j}) = d_1(t_{1j}), j = 1, 2, \dots, n_1 \text{ And}$$

$$f_2(t_{2j}) = d_2(t_{2j}), j = 1, 2, \dots, n_2,$$

with :

$$d_1(t_{1j}) = \gamma_1 t_{1j} + \frac{1}{2} \alpha_{01} + \sum_{k=1}^K \alpha_{1k} \cos kt_{1j}, \text{ And} \quad (4)$$

$$d_2(t_{2j}) = \gamma_2 t_{2j} + \frac{1}{2} \alpha_{02} + \sum_{k=1}^K \alpha_{2k} \cos kt_{2j}. \quad (5)$$

The nonparametric regression equation Birespon (3) can be written as:

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ \dots \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \end{pmatrix} = \begin{pmatrix} \gamma_1 t_{11} + \frac{1}{2} \alpha_{01} + \sum_{k=1}^K \alpha_{1k} \cos kt_{11} \\ \gamma_1 t_{12} + \frac{1}{2} \alpha_{01} + \sum_{k=1}^K \alpha_{1k} \cos kt_{12} \\ \vdots \\ \gamma_1 t_{1n_1} + \frac{1}{2} \alpha_{01} + \sum_{k=1}^K \alpha_{1k} \cos kt_{1n_1} \\ \dots \\ \gamma_2 t_{21} + \frac{1}{2} \alpha_{02} + \sum_{k=1}^K \alpha_{2k} \cos kt_{21} \\ \gamma_2 t_{22} + \frac{1}{2} \alpha_{02} + \sum_{k=1}^K \alpha_{2k} \cos kt_{22} \\ \vdots \\ \gamma_2 t_{2n_2} + \frac{1}{2} \alpha_{02} + \sum_{k=1}^K \alpha_{2k} \cos kt_{2n_2} \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1n_1} \\ \dots \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2n_2} \end{pmatrix}. \quad (6)$$

The quantities $\gamma_1 \in \mathbb{R}$, $\gamma_2 \in \mathbb{R}$, $\alpha_{01} \in \mathbb{R}$, $\alpha_{02} \in \mathbb{R}$, $\alpha_{1k} \in \mathbb{R} \ k = 1, 2, \dots, n_1$, $\alpha_{2k} \in \mathbb{R} \ k = 1, 2, \dots, n_2$ are unknown parameters in the Fourier Series models (4) and (5). Estimated Birespon nonparametric regression curve $f_1(t)$ and $f_2(t)$ obtained from completing *Weighted Penalized Least Square* (PLST) optimization:

$$\begin{aligned} \underset{f_k \in \mathbb{C}(0, \pi), k=1,2}{\text{Min}} \left\{ \frac{1}{(n_1 + n_2)} (y - f)' W(\rho, \sigma_1^2, \sigma_2^2) (y - f) \right. \\ \left. + \lambda_1 \int_0^\pi [f_1^n(t_1)]^2 dt_1 + \lambda_2 \int_0^\pi [f_2^n(t_2)]^2 dt_2 \right\}. \quad (7) \end{aligned}$$

Vectors $y = (y'_1, y'_2)'$ are vectors of size $(n_1, n_2)x_1$ and $y_k = (y_{k1}, y_{k2}, \dots, y_{kn_k})'$ vectors of size $n_k x_1$, $k = 1, 2$. Vectors $f = (f'_1, f'_2)'$ are vectors of size $(n_1, n_2)x_1$ and $f_k = (f_{k1}, f_{k2}, \dots, f_{kn_k})'$ vectors of size $n_k x_1$, $k = 1, 2$. $W(\rho, \sigma_1^2, \sigma_2^2)$ weight matrix of size $(n_1, n_2)x(n_1, n_2)$ that depends on $\sigma^2 = (\sigma_1^2, \sigma_2^2)'$ and ρ is related to equation (2). Parameters λ_1 and λ_2 are smoothing parameters (*bandwidth*).

To obtain an optimization solution for Equation (7), first describe the components of *Goodness of fit*:

$$G(\beta) = G(f) = \frac{1}{n_1 + n_2} (y - f)' W(\rho, \sigma_1^2, \sigma_2^2) (y - f) \quad (8)$$

and penalties:

$$P(\beta) = P(f) = P(f_1, f_2) = \lambda_1 \int_0^\pi \left[\frac{2}{\pi} f_1^n(t_1) \right]^2 dt_1 + \lambda_2 \int_0^\pi \left[\frac{2}{\pi} f_2^n(t_2) \right]^2 dt_2 \quad (9)$$

with :

$$\beta = (\gamma_1, \gamma_2, \alpha_{01}, \alpha_{02}, \alpha_{11}, \alpha_{12}, \dots, \alpha_{1K}, \alpha_{21}, \alpha_{22}, \dots, \alpha_{2K})' \in \mathbb{R}^{2K+4}$$

First, we will describe the components of *Goodness of fit*. The Birespon nonparametric regression model in Equation (6) can be presented in the form:

$$y = B(t_1, t_2)\beta + \varepsilon$$

with :

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ \dots \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n_1} \\ \dots \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n_2} \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \frac{1}{2}\alpha_{01} \\ \alpha_{11} \\ \alpha_{12} \\ \vdots \\ \alpha_{1K} \\ \dots \\ \gamma_2 \\ \frac{1}{2}\alpha_{02} \\ \alpha_{21} \\ \alpha_{22} \\ \vdots \\ \alpha_{2K} \end{pmatrix}, \text{ And}$$

$$B(t_1, t_2) = \begin{pmatrix} t_{11} & 1 & \cos t_{11} & \cos 2t_{11} & \dots & \cos Kt_{11} & 0 & 0 & 0 & 0 & \dots & 0 \\ t_{12} & 1 & \cos t_{12} & \cos 2t_{12} & \dots & \cos Kt_{12} & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{1n_1} & 1 & \cos t_{1n_1} & \cos 2t_{1n_1} & \dots & \cos Kt_{1n_1} & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & t_{21} & 1 & \cos t_{21} & \cos 2t_{21} & \dots & \cos Kt_{21} \\ 0 & 0 & 0 & 0 & \dots & 0 & t_{22} & 1 & \cos t_{22} & \cos 2t_{22} & \dots & \cos Kt_{22} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & t_{2n_1} & 1 & \cos t_{2n_1} & \cos 2t_{2n_1} & \dots & \cos Kt_{2n_1} \end{pmatrix}$$

Goodness of fit in Equation (8) can be expressed as:

$$G(\beta) = \frac{1}{n_1 + n_2} (y - B(t_1, t_2)\beta)' W(\rho, \sigma_1^2, \sigma_2^2) (y - B(t_1, t_2)\beta) \quad (10)$$

Next, the Penalty will be derived from Equation (9). First, we derive the first penalty term from Equation (9), as follows:

Because $f_1(t_1) = \gamma_1 t_1 + \frac{1}{2} \alpha_{01} + \sum_{k=1}^K \alpha_{1k} \cos kt_1$, then:

$$\int_0^{\pi} \frac{2}{\pi} [f_1^n(t_1)]^2 dt_1 = \int_0^{\pi} \frac{2}{\pi} \left(\frac{d^2}{dt_1^2} \left[\gamma_1 t_1 + \frac{1}{2} \alpha_{01} + \sum_{k=1}^K \alpha_{1k} \cos kt_1 \right] \right)^2 dt_1$$

To solve this integral, first determine the quantity of:

$$\begin{aligned} & \left(\frac{d^2}{dt_1^2} \left[\gamma_1 t_1 + \frac{1}{2} \alpha_{01} + \sum_{k=1}^K \alpha_{1k} \cos kt_1 \right] \right)^2 \\ &= \left(\frac{d}{dt_1} \left\{ \frac{d}{dt_1} \left[\gamma_1 t_1 + \frac{1}{2} \alpha_{01} + \sum_{k=1}^K \alpha_{1k} \cos kt_1 \right] \right\} \right)^2 \\ &= \left(\frac{d}{dt_1} \left[\gamma_1 - \sum_{k=1}^K \alpha_{1k} \sin kt_1 \right] \right)^2 \\ &= \left(\sum_{k=1}^K k^2 \alpha_{1k} \cos kt_1 \right)^2 \end{aligned}$$

The final equation gives:

$$\begin{aligned} \int_0^{\pi} \frac{2}{\pi} [f_1^n(t_1)]^2 dt_1 &= \frac{2}{\pi} \int_0^{\pi} \left(\sum_{k=1}^K k^2 \alpha_{1k} \cos kt_1 \right)^2 dt_1 \\ &= \frac{2}{\pi} \int_0^{\pi} \left(\sum_{k=1}^K (k^2 \alpha_{1k} \cos kt_1)^2 + 2 \sum_{k < j}^K \sum_{k < j}^K (k^2 \alpha_{1k} \cos kt_1)(j^2 \alpha_{1j} \cos jt_1) \right) dt_1 \\ &= \frac{2}{\pi} \int_0^{\pi} \sum_{k=1}^K (k^2 \alpha_{1k} \cos kt_1)^2 dt_1 + \frac{4}{\pi} \int_0^{\pi} \sum_{k < j}^K (k^2 \alpha_{1k} \cos kt_1)(j^2 \alpha_{1j} \cos jt_1) dt_1 \quad (11) \end{aligned}$$

To solve this equation, first calculate the Integral value of the first term of Equation (11) as follows:

$$\begin{aligned} \frac{2}{\pi} \sum_{k=1}^K \int_0^{\pi} (k^2 \alpha_{1k} \cos kt_1)^2 dt_1 &= \frac{2}{\pi} \sum_{k=1}^K \int_0^{\pi} k^4 \alpha_{1k}^2 \cos^2 kt_1 dt_1 \\ &= \frac{2}{\pi} \sum_{k=1}^K k^4 \alpha_{1k}^2 \int_0^{\pi} \cos^2 kt_1 dt_1 \\ &= \frac{2}{\pi} \sum_{k=1}^K k^4 \alpha_{1k}^2 \int_0^{\pi} \left(\frac{1 + 2 \cos kt_1}{2} \right) dt_1 \end{aligned}$$

$$= \frac{1}{\pi} \sum_{k=1}^{\pi} k^4 \alpha_{1k}^2 \left[t_1 + \frac{2}{k} \sin kt_1 \right] dt_1$$

$$= \sum_{k=1}^{\pi} k^4 \alpha_{1k}^2$$

Next, the Integral value of the second term from Equation (11) is calculated as follows:

$$\frac{4}{\pi} \sum_{k < j}^K \int_0^{\pi} (k^2 \alpha_{1k} \cos kt_1) (j^2 \alpha_{1j} \cos jt_1) dt_1 = \frac{4}{\pi} \sum_{k < j}^K \int_0^{\pi} k^2 j^2 \alpha_{1k} \alpha_{1j} \cos kt_1 \cos jt_1 dt_1$$

$$= \frac{4}{\pi} \sum_{k < j}^K (kj)^2 \alpha_{1k} \alpha_{1j} \int_0^{\pi} \cos kt_1 \cos jt_1 dt_1$$

Next obtained:

$$\int_0^{\pi} \cos kt_1 \cos jt_1 dt_1 = \int_0^{\pi} \left(\frac{\cos(k+j)t_1 + \cos(k-j)t_1}{2} \right) dt_1$$

$$= \int_0^{\pi} \left(\frac{1}{2} \cos(k+j)t_1 \right) dt_1 + \int_0^{\pi} \left(\frac{1}{2} \cos(k-j)t_1 \right) dt_1$$

$$= \frac{1}{2} \cdot \frac{1}{(k+j)} (\sin(k+j)\pi - \sin 0) + \frac{1}{2} \cdot \frac{1}{k-j} (\sin(k-j)\pi - \sin 0)$$

$$= 0$$

As a result:

$$\frac{4}{\pi} \sum_{k < j}^K \int_0^{\pi} (k^2 \alpha_{1k} \cos kt_1) (j^2 \alpha_{1j} \cos jt_1) dt_1 = \frac{4}{\pi} \sum_{k < j}^K \int_0^{\pi} k^2 j^2 \alpha_{1k} \alpha_{1j} \cos kt_1 \cos jt_1 dt_1$$

$$= \frac{4}{\pi} \sum_{k < j}^K (kj)^2 \alpha_{1k} \alpha_{1j} \cdot 0 = 0$$

This result gives the *Penalty value* of the first term of Equation (9):

$$\int_0^{\pi} \frac{2}{\pi} [f_1^n(t_1)]^2 dt_1 = \sum_{k=1}^K k^4 \alpha_{1k}^2 \quad (12)$$

The next step is to derive the second *Penalty* term from Equation (9).

In a similar way to calculating the *penalty* in the first term, the *penalty value* is obtained from second term in Equation (9).

$$\int_0^{\pi} \frac{2}{\pi} [f_2^n(t_2)]^2 dt_2 = \sum_{k=1}^K k^4 \alpha_{2k}^2 \quad (13)$$

Based on Equations (12) and (13), the *Penalty* in Equation (9) can be written as:

$$\begin{aligned}
 P(\beta) &= (\lambda_1 \quad \lambda_2) \begin{pmatrix} \sum_{k=1}^K K^4 \alpha_{1k}^2 \\ \sum_{k=1}^K K^4 \alpha_{2k}^2 \end{pmatrix} \\
 &= \lambda_1 \beta_1' \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1^4 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & K^4 \end{pmatrix} \beta_1 + \lambda_2 \beta_2' \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1^4 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & K^4 \end{pmatrix} \beta_2 \\
 &= \lambda_1 \beta_1' D \beta_1 + \lambda_2 \beta_2' D \beta_2 \\
 &= (\beta_1' \quad \beta_2') \begin{pmatrix} \lambda_1 I_{K+2} & 0 \\ 0 & \lambda_2 I_{K+2} \end{pmatrix} \begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \\
 &= \beta' \lambda^* D^* \beta
 \end{aligned} \tag{14}$$

Where: $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$, $\lambda = \begin{pmatrix} \lambda_1 I_{K+2} & 0 \\ 0 & \lambda_2 I_{K+2} \end{pmatrix}$, $D^* = \begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix} = \text{Diag}(D, D)$, And

$$D = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1^4 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & K^4 \end{pmatrix}$$

By paying attention to Equation (10) and Equation (14), the PLST optimization in Equation (7) can be expressed as:

$$\begin{aligned}
 \min_{\beta \in \mathbb{R}^{2K+4}} \{G(\beta) + P(\beta)\} &= \\
 \min_{\beta \in \mathbb{R}^{2K+4}} \left\{ \frac{1}{n_1 + n_2} (y - B(t_1, t_2)\beta)' W(\rho, \sigma_1^2, \sigma_2^2) (y - B(t_1, t_2)\beta) + \beta' \lambda^* D^* \beta \right\} &= \\
 \min_{\beta \in \mathbb{R}^{2K+4}} \left\{ \frac{1}{n_1 + n_2} (y' W(\rho, \sigma_1^2, \sigma_2^2) y - 2B'(t_1, t_2) W(\rho, \sigma_1^2, \sigma_2^2) \beta \right. & \\
 \left. + \beta' B'(t_1, t_2) W(\rho, \sigma_1^2, \sigma_2^2) B(t_1, t_2) \beta) + \beta' \lambda^* D^* \beta \right\} &=
 \end{aligned}$$

$$\min_{\beta \in \mathbb{R}^{2K+4}} \left\{ \frac{1}{n_1 + n_2} y' W(\rho, \sigma_1^2, \sigma_2^2) y - \frac{2}{n_1 + n_2} B'(t_1, t_2) W(\rho, \sigma_1^2, \sigma_2^2) \beta \right. \\ \left. + \beta' \left(\frac{1}{n_1 + n_2} B'(t_1, t_2) W(\rho, \sigma_1^2, \sigma_2^2) B(t_1, t_2) + \lambda^* D^* \right) \beta \right\}$$

With a little elaboration and using partial derivatives, we obtain the normal equation:

$$-\frac{2}{n_1 + n_2} B'(t_1, t_2) W(\rho, \sigma_1^2, \sigma_2^2) + 2 \left(\frac{1}{n_1 + n_2} B'(t_1, t_2) W(\rho, \sigma_1^2, \sigma_2^2) B(t_1, t_2) + \lambda^* + D^* \right) \beta = 0$$

The estimator for β is given by:

$$\hat{\beta} = \left(\frac{1}{N} B'(t_1, t_2) W(\rho, \sigma_1^2, \sigma_2^2) B(t_1, t_2) + \lambda^* + D^* \right)^{-1} \frac{1}{N} B'(t_1, t_2) W(\rho, \sigma_1^2, \sigma_2^2) y,$$

with $N = n_1 + n_2$

The Fourier Series estimator for the Birespon nonparametric regression curve, is given by:

$$\hat{f}_{\lambda}(t_1, t_2) = \begin{pmatrix} \hat{f}_{\lambda_1}(t_1) \\ \hat{f}_{\lambda_2}(t_2) \end{pmatrix} = B(t_1, t_2) \hat{\beta}_{\lambda} \\ = B(t_1, t_2) \left(\frac{1}{N} B'(t_1, t_2) W(\rho, \sigma_1^2, \sigma_2^2) B(t_1, t_2) + \lambda^* D^* \right)^{-1} \frac{1}{N} B'(t_1, t_2) W(\rho, \sigma_1^2, \sigma_2^2) y \\ = H(\lambda_1, \lambda_2) y \quad (15)$$

With $H(\lambda_1, \lambda_2) =$

$$B(t_1, t_2) \left(\frac{1}{N} B'(t_1, t_2) W(\rho, \sigma_1^2, \sigma_2^2) B(t_1, t_2) + \lambda^* D^* \right)^{-1} \frac{1}{N} B'(t_1, t_2) W(\rho, \sigma_1^2, \sigma_2^2)$$

3. Properties of the Bivariate Fourier Series Estimator

Equation (15), provides a Fourier Series estimator for the Birespon nonparametric regression curve which can basically be expressed as a linear combination of observations (responses) $y =$

$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, namely:

$$\hat{f}_{\lambda}(t_1, t_2) = \begin{pmatrix} \hat{f}_{\lambda_1}(t_1) \\ \hat{f}_{\lambda_2}(t_2) \end{pmatrix} = H(\lambda_1, \lambda_2) y = H(\lambda_1, \lambda_2) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

This equation shows that the Fourier Series estimator for the Birespon nonparametric regression curve belongs to the class of linear estimators. Linear estimators in Inference Statistics, especially

estimation theory, are very preferred, because this linear nature will make it easier to obtain inferences for non-parametric Birespon regression curves. Even though this Fourier Series estimator is classified as a linear estimator, this Fourier Series estimator is biased. This can be shown as follows.

From Equation (15), we obtain:

$$\begin{aligned} E \left[\widehat{f}_{\lambda}(t_1, t_2) \right] &= E \left[\begin{pmatrix} \widehat{f}_{\lambda_1}(t_1) \\ \widehat{f}_{\lambda_2}(t_2) \end{pmatrix} \right] \\ &= E \left[H(\lambda_1, \lambda_2) y \right] \\ &= E \left[H(\lambda_1, \lambda_2) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right] \\ &= H(\lambda_1, \lambda_2) \begin{pmatrix} E(y_1) \\ E(y_2) \end{pmatrix} \\ &= H(\lambda_1, \lambda_2) \begin{pmatrix} f_1(t_1) \\ f_2(t_2) \end{pmatrix} \\ &\neq \begin{pmatrix} f_1(t_1) \\ f_2(t_2) \end{pmatrix} = f \end{aligned}$$

4. Generalized Cross Validation (GCV) and Mean Square Error (MSE)

The shape of the Fourier Series estimator is strongly influenced by smoothing parameters λ . Therefore, λ optimal selection needs to be done to get a good Fourier Series estimator. A value will be chosen λ that minimizes the loss function $L(\lambda)$, but nonparametric regression cannot be done directly because $L(\lambda)$ it still contains an g unknown function. The methods often used in selecting smoothing parameters λ are GCV and MSE, Green and Silverman, 1994). The GCV method is one of the methods used to obtain estimates of smoothing parameters λ obtained by minimizing

$$GCV(\lambda) = n^{-1} \sum_{i=1}^n \frac{MSE(\lambda)}{\left(1 - n^{-1} \sum_{i=1}^n a_{ii}(\lambda)\right)^2} \quad (16)$$

where $MSE(\lambda) = n^{-1} \sum_{i=1}^n (y_i - \hat{g}_\lambda(t_i))^2$ and $a_{ii}(\lambda)$ are the diagonal elements of the matrix $S(\lambda)$ satisfying $\hat{g}_\lambda(t) = S(\lambda)y$.

One measure for the accuracy of estimating the regression curve function is to use the Mean Square Error (MSE) which is defined as follows: $MSE(\lambda) = n^{-1} \sum_{i=1}^n (y_i - \hat{g}_\lambda(t_i))^2$. (17)

5. Application of the Fourier Series to data on rainy days and rainfall in cities Cilacap

In this research, an application was carried out to provide an overview of the Fourier Series regression model. The application in this research was carried out to evaluate the goodness of the GCV and MSE methods. Measurement reliability is based on the smallest E value obtained using the GCV and MSE methods, using Veiten Software. Given two experimental curves, namely:

1. Trigonometric Functions : $g(t) = \sin(2\pi t^3)$
2. Exponential Function : $g(t) = 4,26(e^{-t} - 4e^{-2t} + 3e^{-3t})$.

The simulation procedures are described as follows:

Building a regression model $y_i = g(t_i) + \varepsilon_i$, $i = 1, 2, \dots, n$, and sample size $n = 50$, $n = 100$, $n = 200$ with values $K = 5$, $K = 20$. Given $t_i = \frac{i}{n}$, by $i = 1, 2, 3, \dots, n$. Generating $\varepsilon_i \sim N(0, \sigma^2)$, by $\sigma^2 = 0, 1$. After generating the data, the **X matrix**, **D matrix** and hat matrix are calculated $S(\lambda) = X(n^{-1}X'X + \lambda D)^{-1}n^{-1}X'$.

Next, the estimator is calculated $\hat{a}(\lambda) = (n^{-1}X'X + \lambda D)^{-1}n^{-1}X'y$, and determine the Fourier Series estimator with $\hat{g}_\lambda(t) = S(\lambda)y$, so that we get

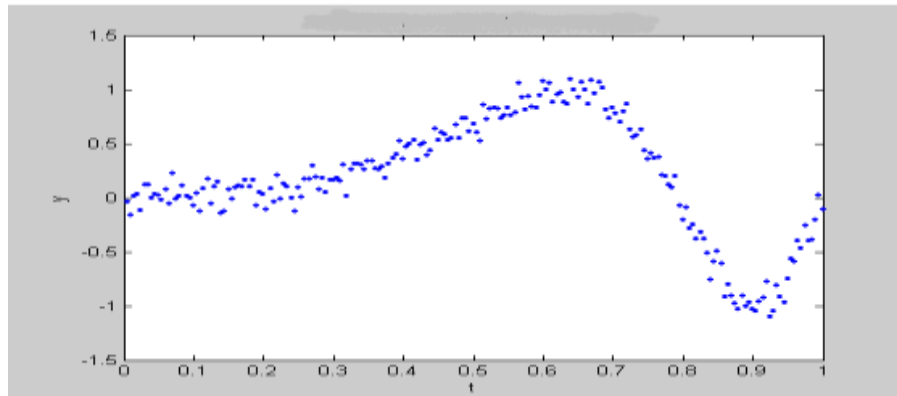
$$\hat{g}_\lambda(t) = \hat{b}(\lambda)t + \frac{1}{2}\hat{a}_0(\lambda) + \sum_{k=1}^K \hat{a}_k(\lambda)\cos kt.$$

Select smoothing parameters using the UBR, CV and GCV

methods and compare the MSE values of the estimators where the optimal smoothing parameters are selected using the UBR, CV and GCV methods.

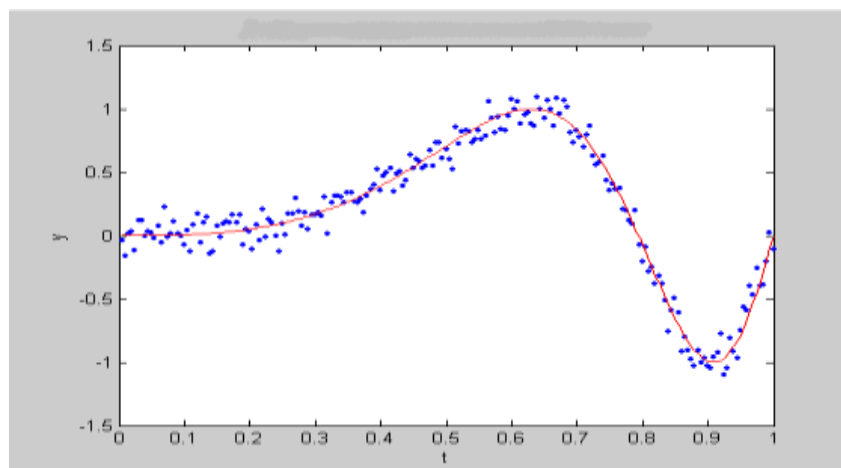
4.4.1. Trigonometric Function Simulation : $g(t) = \sin(2\pi t^3)$.

Below are presented the simulation results for trigonometric functions $g(t) = \sin(2\pi t^3)$. Plots of (t_i, y_i) , experimental curves of trigonometric functions where $n = 200$, and $\sigma^2 = 0, 1$, and combined plots of generation data and function curves are presented respectively. in Figure 4.1, and Figure 4.2



..... (Dots) : Generation Data

Figure 1 Plots (t_i, y_i) for Trigonometric Functions $g(t) = \sin(2\pi t^3)$,
with $n = 200$ $\sigma^2 = 0,1$



..... (Dotted): Generation Data

— (Red Line): Function $g(t)$

Figure 2 Trigonometric $g(t) = \sin(2\pi t^3)$ Plots and Functions (t_i, y_i) , with $n = 200$
 $\sigma^2 = 0,1$

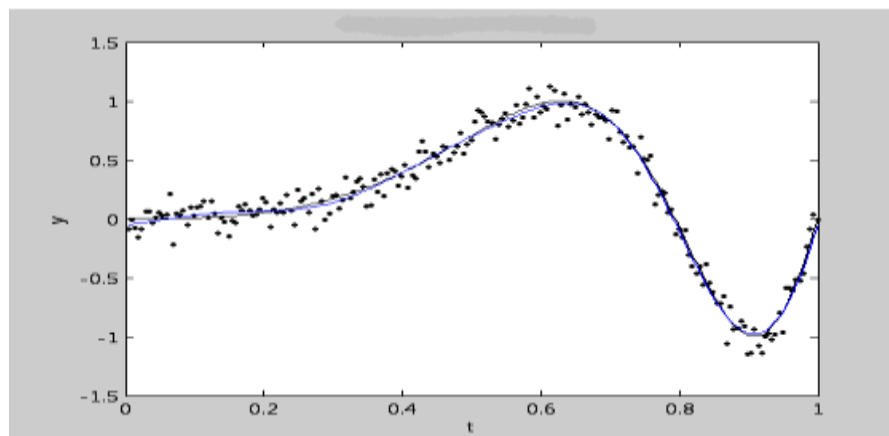
Next, a Fourier Series estimator plot is given for the generation data from one of the replications for trigonometric functions $g(t)$ with $n = 200$, $\sigma^2 = 0,1$, $K = 5$ which are respectively given in Figure 4.3, Figure 4.4, Figure 4.5, and Figure 4.6.

Next, the selection of optimal smoothing parameters λ is selected using the UBR method which was previously studied in sub-chapter 4.3. For the UBR method the equation used is as follows:

$$UBR(\lambda) = n^{-1} \tilde{y}' (\mathbf{S}(\lambda) - \mathbf{I})' (\mathbf{S}(\lambda) - \mathbf{I}) \tilde{y} + n^{-1} \sigma^2 \text{trace}(\mathbf{S}'(\lambda) \mathbf{S}(\lambda)) + \\ - n^{-1} \sigma^2 \text{trace}(\mathbf{S}(\lambda) - \mathbf{I})' (\mathbf{S}(\lambda) - \mathbf{I})$$

with $\mathbf{S}(\lambda)$ is the hat matrix obtained from the Fourier Series $\hat{g}_\lambda(t)$

Below are the intermediate plot results (t_i, y_i) for Fourier Series estimation for trigonometric functions with $n = 200$, $\sigma^2 = 0,1$, $K = 5$, presented in Figure 4.3 as follows:



- (Dots) : Generation Data
- (Black Line) : Function $g(t)$
- (Blue Line): Fourier Series with UBR Method

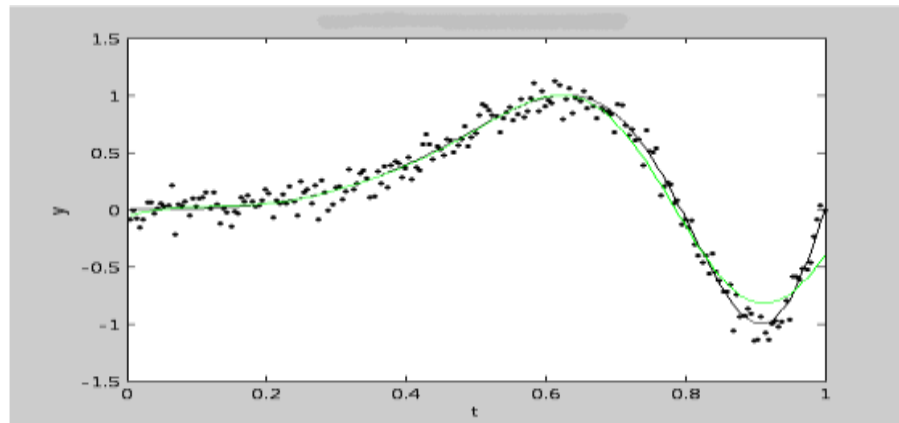
Figure 3 Plot (t_i, y_i) , for Trigonometric Functions $g(t)$ and Fourier Series Estimator with UBR Method, $n = 200$, $\sigma^2 = 0,1$, $K = 5$

Next, to select the λ optimal smoothing parameters using the CV method, the following equation is used:

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{g}_\lambda(t_i))^2$$

Given a plot (t_i, y_i) on the Fourier Series estimator for trigonometric functions $g(t)$ using the CV method with $n = 200$, $\sigma^2 = 0,1$,

$K = 5$ is presented in Figure 4.4 as follows:



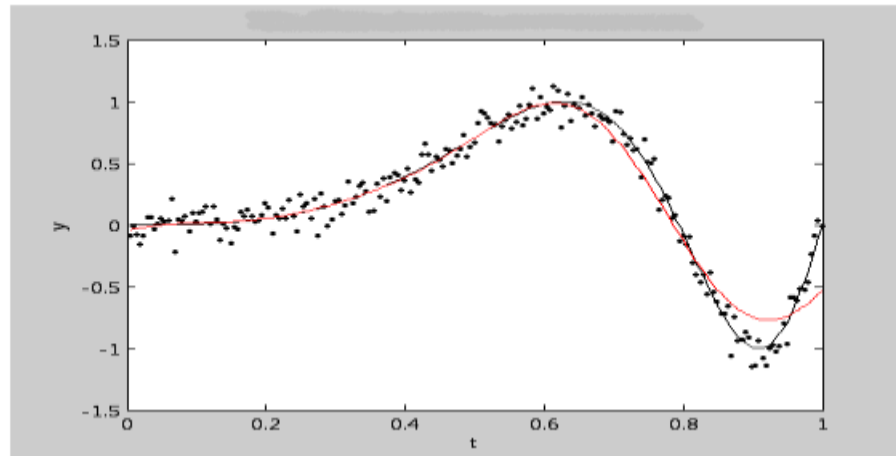
- (Dotted): Generation Data
- (Black Line) : Function $g(t)$
- (Green Line): Fourier Series with CV Method

Figure 4 Plot (t_i, y_i) , Trigonometric Functions $g(t)$ and Fourier Series Estimator with CV Method for $n = 200$, $\sigma^2 = 0,1$, $K = 5$

Next, to select λ optimal smoothing parameters using the GCV method, the following equation is used:

$$GCV(\lambda) = n^{-1} \sum_{i=1}^n \frac{(y_i - \hat{g}_\lambda(t_i))^2}{\left(1 - n^{-1} \sum_{i=1}^n a_{ii}(\lambda)\right)^2}$$

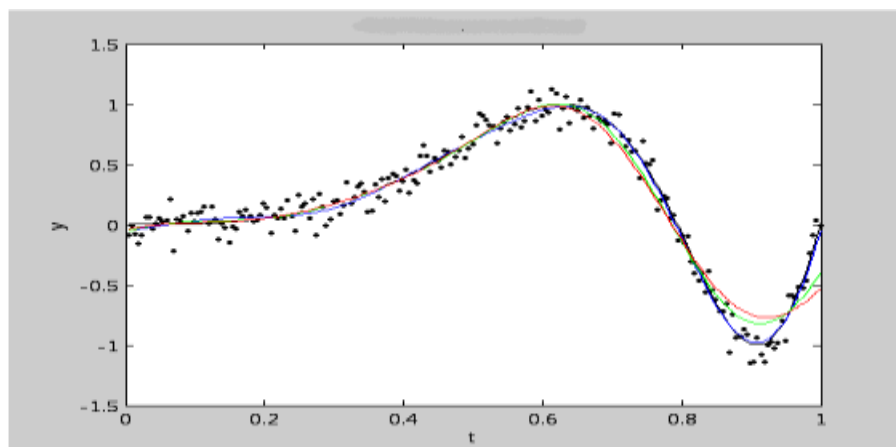
Given a plot (t_i, y_i) on the Fourier Series estimator for trigonometric functions $g(t)$ using the GCV method with $n = 200$, $\sigma^2 = 0,1$, $K = 5$ is presented in Figure 4.5 as follows:



- (Dotted): Generation Data
- (Black Line): Function $g(t)$
- (Red Line): Fourier Series with GCV Method

Figure 5 Plot (t_i, y_i) , Trigonometric Functions $g(t)$ and Fourier Series Estimator using the GCV Method for $n = 200$, $\sigma^2 = 0,1$, $K = 5$

Next, a plot is given between (t_i, y_i) the function $g(t)$ and the Fourier Series estimator where λ the optimum is selected using the UBR, CV and GCV methods for $n = 200$, $\sigma^2 = 0,1$, $K = 5$, presented in Figure 4.6 as follows:



- (Dots) : Generation Data
- (Black Line): Function $g(t)$

- (Blue Line): Fourier Series with UBR Method
— (Green Line): Fourier Series with CV Method
— (Red Line): Fourier Series with GCV Method

Figure 6 Trigonometric Function $g(t)$ Plot of Fourier Series Estimator using UBR, CV, GCV Methods for $n = 200$, $\sigma^2 = 0,1$, $K = 5$

Evaluation of the goodness of the UBR, CV and GCV methods in selecting smoothing parameters λ is reviewed based on the smallest MSE value produced by each method. The following data simulation results, for optimal values λ and MSE of the UBR, CV and GCV methods are presented in Tabel 1 as follows:

Table 1 Simulation Results **Optimal** Value λ and MSE of the UBR, CV and GCV Methods in the Fourie Series Estimator with Function $g(t) = \sin(2\pi t^3)$, where $n = 50, n = 100, n = 200$, $\sigma^2 = 0,1$ and $K = 5, K = 20$

N	var	k	UBR method		CV method		GCV method	
			$\lambda_{optimal}$	MSE UBR	$\lambda_{optimal}$	MSE CV	$\lambda_{optimal}$	MSE GCV
50	0.1	5	0.2067	0.0019053	0.107620	0.004077	0.015056	0.021495
		20	0.08617	0.014198	0.045473	0.019947	0.003230	0.0359330
100	0.1	5	0.20015	0.0021103	0.102020	0.0040874	0.014231	0.0179270

		20	0.094627	0.013752	0.048525	0.020311	0.003442	0.0359630
200	0.1	5	0.1808	0.0023361	0.091257	0.0044296	0.012714	0.0179280
		20	0.10285	0.013313	0.052058	0.020157	0.003691	0.0375740

Visually, the MSE value in Table 4.1 can be displayed in the bar diagram in Figure 4.7

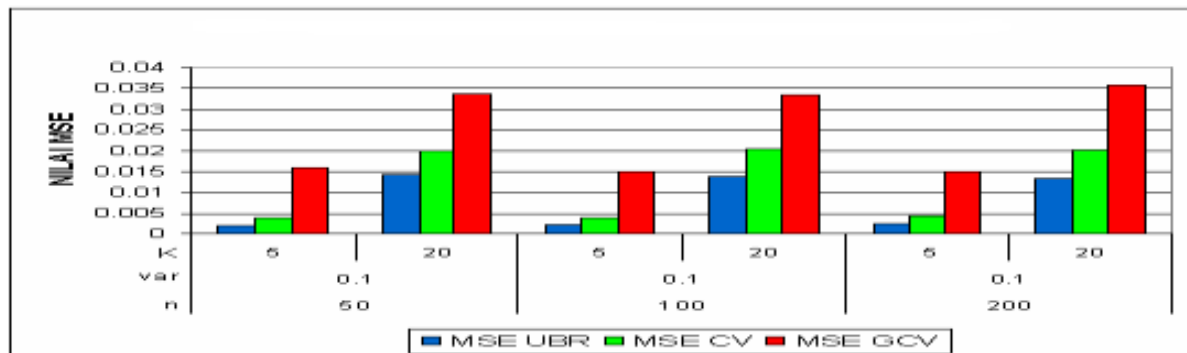


Figure 7 : Bar Diagram of MSE Values for Trigonometric Functions $g(t)$, for

$n = 50, n = 100, n = 200, \sigma^2 = 0.1$ and $K = 5, K = 20$

By paying attention to the MSE values from the UBR, CV and GCV methods in Table 4.1 and Figure 4.7, it can be concluded as follows:

1. At sample sizes $n = 50, n = 100, n = 200, \sigma^2 = 0.1$ and $K = 5, K = 20$, the MSE UBR value is smaller than the MSE CV and GCV values, so the UBR method is better than the CV and GCV methods
2. At sample sizes $n = 50, n = 100, n = 200, \sigma^2 = 0.1$ and $K = 5, K = 20$, the MSE CV value obtained

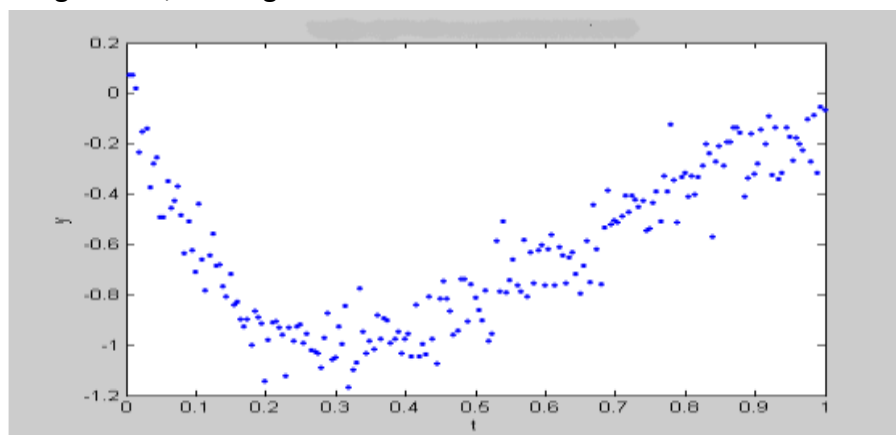
is smaller than the GCV method, so the CV method is better than the GCV method.

Based on the simulation results of the Fourir Series estimator with trigonometric functions in Table 4.1 and Figure 4.7, it can be explained that the UBR method is more suitable than the CV and GCV methods. The same thing also happens for experiment two for samples $n = 50$, $n = 100$, $n = 200$, $\sigma^2 = 0,1$ and $K = 5$, $K = 20$

The complete results for this can be seen in Appendix 5.

4.4.2. Exponential Function Simulation: $g(t) = 4,26(e^{-t} - 4e^{-2t} + 3e^{-3t})$

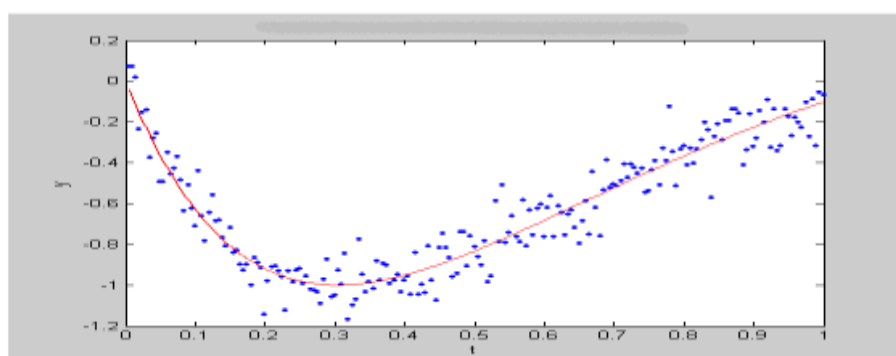
The following describes the simulation results for the exponential function $g(t) = 4,26(e^{-t} - 4e^{-2t} + 3e^{-3t})$. The plot between (t_i, y_i) , the experimental curve of the trigonometric function where $n = 200$, and $\sigma^2 = 0,1$, the combined plot of both are presented respectively in Figure 4.8, and Figure 4.9.



..... (Dotted): Generation Data

Figure 4.8 Plot (t_i, y_i) and Exponent Function $g(t) = 4,26(e^{-t} - 4e^{-2t} + 3e^{-3t})$

with $n = 200$ $\sigma^2 = 0,1$



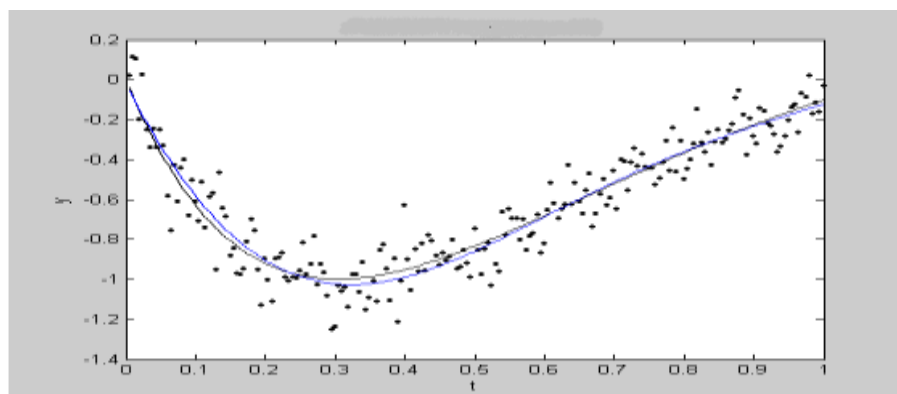
..... (Dotted): Generation Data
 — (Red Line): Function $g(t)$

Figure 9 Plot (t_i, y_i) and Exponent Function $g(t) = 4,26(e^{-t} - 4e^{-2t} + 3e^{-3t})$
 with $n = 200$ $\sigma^2 = 0,1$.

Below is a plot of the Fourier Series estimator for generation data from one of the replications for the exponential function $g(t)$ with $n = 200$, $\sigma^2 = 0,1$, $K = 5$ which are respectively given in Figure 4.10, Figure 4.11, Figure 4.12, and Figure 4.13. The UBR equation method used is as follows:

$$UBR(\lambda) = n^{-1} \tilde{y}'(S(\lambda) - I)'(S(\lambda) - I) \tilde{y} + n^{-1} \sigma^2 \text{trace}(S'(\lambda) S(\lambda)) + n^{-1} \sigma^2 \text{trace}(S(\lambda) - I)'(S(\lambda) - I)$$

Figure 4.10 presents a plot between (t_i, y_i) the exponential function $g(t)$ and the Fourier Series estimator where λ the optimum is selected using the UBR method. The smoothing parameters using the UBR method at certain intervals coincide with each other so it can be said that the UBR method still follows the function $g(t)$. Simulation results for an exponential function $g(t)$ with $n = 200$, $\sigma^2 = 0,1$, $K = 5$ are presented in Figure 4.10 as follows:



..... (Dotted): Generation Data
 — (Black Line): Function $g(t)$
 — (Blue Line): Fourier Series with UBR Method

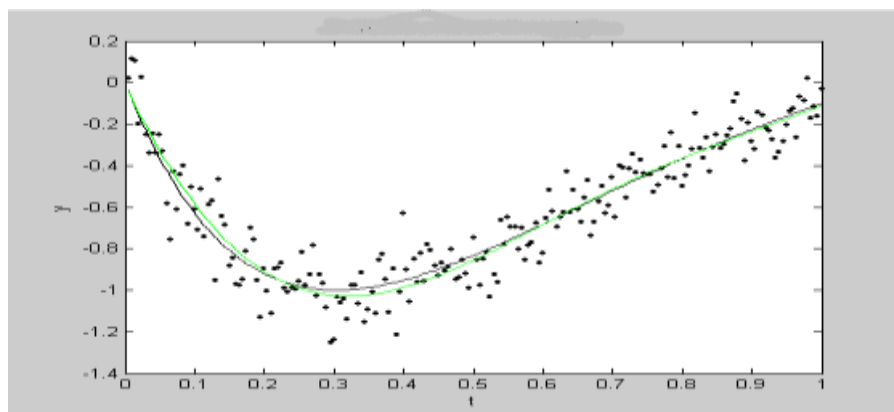
Figure 10 Exponential Function Plot (t_i, y_i) $g(t)$ and Fourier Series Estimator with the UBR Method for $n = 200$, $\sigma^2 = 0,1$, $K = 5$

Next, to select the λ optimal smoothing parameters using the CV method, the following equation is used:

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{g}_\lambda(t_i))^2$$

Figure 4.11 presents a plot between (t_i, y_i) the exponential function $g(t)$ and the Fourier Series estimator where the λ optimal smoothing parameters are selected using the CV method for $n = 200$, $\sigma^2 = 0,1$, $K = 5$. The CV smoothing parameters are very close to the function $g(t)$, even at certain intervals they coincide with each other.

Simulation results for the Fourier Series estimator plot for an exponential function $g(t)$ with $n = 200$, $\sigma^2 = 0,1$, $K = 5$ are presented in Figure 4.11 as follows:



- (Dotted): Generation Data
- (Black Line): Function $g(t)$
- (Green line): Fourier Series with CV Method

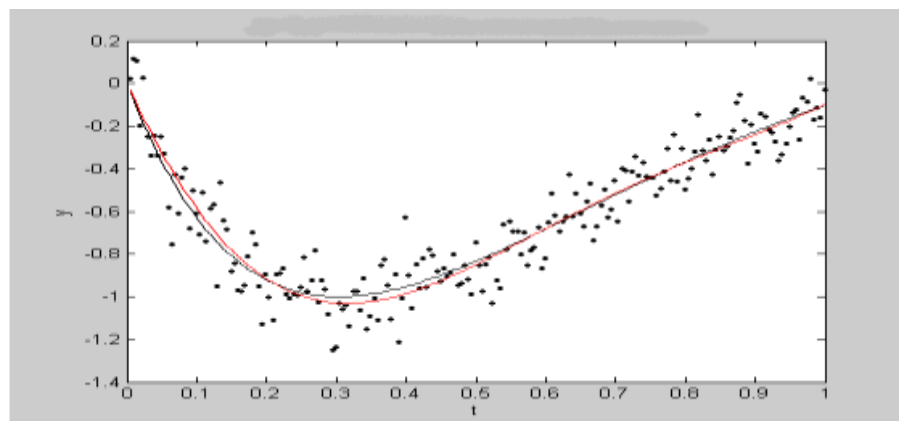
Figure 11 Plot (t_i, y_i) , Exponential Function $g(t)$ and Fourier Series Estimator with CV Method for $n = 200$, $\sigma^2 = 0,1$, $K = 5$

Next, choose the λ optimal smoothing parameters using the GCV method. For the GCV method the equation used is as follows:

$$GCV(\lambda) = n^{-1} \sum_{i=1}^n \frac{(y_i - \hat{g}_\lambda(t_i))^2}{\left(1 - n^{-1} \sum_{i=1}^n a_{ii}(\lambda)\right)^2}$$

Figure 4.12 presents an intermediate plot (t_i, y_i) of the exponential function $g(t)$ and a Fourier Series estimator where the λ optimal smoothing parameters are selected using the GCV

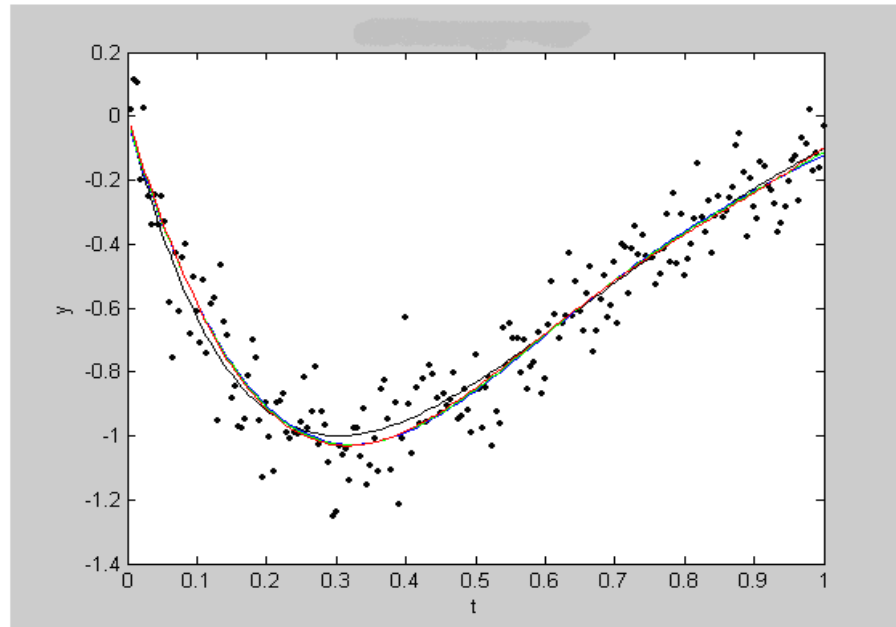
method for $n = 200$, $\sigma^2 = 0,1$, $K = 5$. The GCV smoothing parameters are very close to the function $g(t)$. The simulation results using the GCV method and the Fourier Series estimator plot for the exponential function $g(t)$ are presented in Figure 4.12 as follows:



- (Dotted): Generation Data
- (Black Line): Function $g(t)$
- (Red Line): Fourier Series with GCV Method

Figure 12 Plot (t_i, y_i) , Exponential Function $g(t)$ and Fourier Series Estimator using GCV Method for $n = 200$, $\sigma^2 = 0,1$, $K = 5$

Next is the plot between (t_i, y_i) the exponential function $g(t)$ and the Fourier Series estimator where the λ optimal smoothing parameters are selected using the UBR, CV and GCV methods for $n = 200$, $\sigma^2 = 0,1$, $K = 5$. The combination of these three methods looks very close to the Fourier Series estimator, even at certain intervals they coincide with each other. so it can be said that the UBR, CV and GCV methods still follow the function $g(t)$. The simulation results using the UBR, CV and GCV methods and the Fourier Series estimator plot for the exponential function $g(t)$ are presented in Figure 13 as follows:



- (Dots) : Generation Data
- (Black Line) : Function $g(t)$
- (Blue Line : Fourier Series using the UBR Method
- Green Line : Fourier Series with CV Method
- Red Line): Fourier Series with GCV Method

Figure 13 Plot of (t_i, y_i) the Exponent Function of $g(t)$ the Fourier Series Estimator using the UBR, CV, GCV Method, $n = 200$, $\sigma^2 = 0,1$, and $K = 5$

Below we investigate the MSE value for the Fourier Series estimator where the value λ . optimal for each is selected by UBR, CV and GCV methods. Table 4.2 and Figure 4.14 present the MSE values for each method

$n = 50, n = 100, n = 200, \sigma^2 = 0,1$ and $K = 5, K = 20$

Table 4.2 Simulation Results **Optimal** value λ and MSE of UBR, CV and GCV methods on the Fourie Series Estimator with Functions $g(t) = 4, 26(e^{-t} - 4e^{-2t} + 3e^{-3t})$, $n = 50, n = 100, n = 200$, $\sigma^2 = 0,1$ and $K = 5, K = 20$

n	var	k	UBR method		CV method		GCV method	
			$\lambda_{optimal}$	MSE UBR	$\lambda_{optimal}$	MSE CV	$\lambda_{optimal}$	MSE GCV
50	0.1	5	0.208130	0.0029387	0.108360	0.0056457	0.015160	0.02517
		20	0.023513	0.0011852	0.012408	0.0013142	0.000881	0.0015975
100	0.1	5	0.018596	0.0010388	0.009479	0.0010671	0.001322	0.0010937
		20	0.017993	0.0012463	0.009227	0.001326	0.000654	0.0014255
200	0.1	5	0.025844	0.0014638	0.013044	0.0016284	0.001807	0.0018173

		20	0.019389	0.0012701	0.009814	0.0013433	0.000696	0.0014924

It can be seen from Table 4.2 for each method and all sample sizes of $n = 50$, $n = 100$, $n = 200$, $\sigma^2 = 0,1$ and $K = 5$, $K = 20$ in the exponential function that in general it is found that the MSE value of the UBR method tends to be smaller than the MSE method CV and MSE GCV methods.

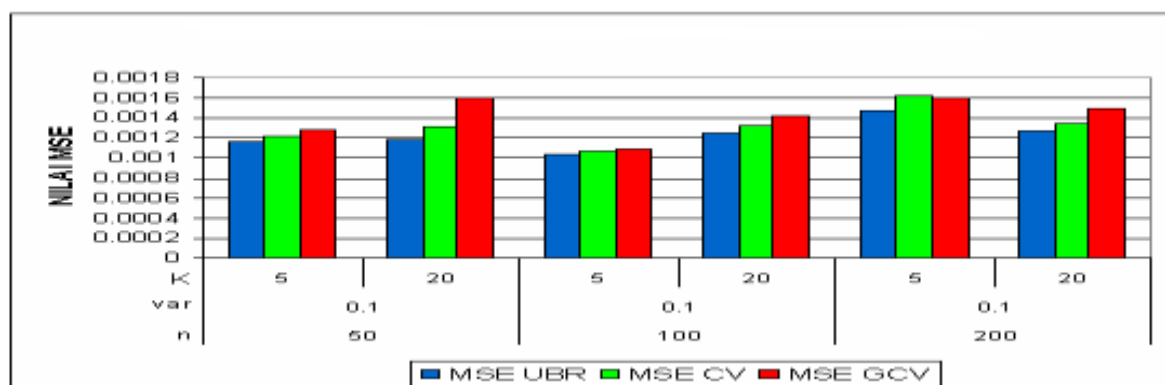


Figure 14 : Bar Diagram of MSE Values for Exponential Functions

$$g(t) = 4,26(e^{-t} - 4e^{-2t} + 3e^{-3t})$$

$$n = 50, n = 100, n = 200, \sigma^2 = 0,1 \text{ and } K = 5, K = 20$$

Visually, the MSE values for each method are shown in Table 4.2 and Figure 4.14 for all sample sizes $n = 50$, $n = 100$, $n = 200$, $\sigma^2 = 0,1$ and $K = 5$, $K = 20$ for exponential functions $g(t) = 4,26(e^{-t} - 4e^{-2t} + 3e^{-3t})$.

Based on the simulation results of the Fourir Series estimator with the exponential function Table 4.2 and Figure 4.14 it can be explained as follows:

1. The MSE value of the UBR method is smaller than the MSE of the CV method and MSE of the GCV method, with sample sizes of $n = 50$, $n = 100$, $n = 200$, $\sigma^2 = 0,1$ and $K = 5$, $K = 20$.
2. Meanwhile, the MSE value of the CV method is smaller than the GCV method with sample sizes of $n = 50$, $n = 100$, $n = 200$, $\sigma^2 = 0,1$ and $K = 5$, $K = 20$.



3. If the K value becomes greater, the MSE value becomes greater for both the UBR, CV and GCV methods, but this has no effect because the MSE for the UBR method is smaller.

CONCLUSION

Overall it is good to use trigonometric functions and exponential functions, if the K value is greater then the MSE value becomes greater for both UBR, CV and GCV methods for sample sizes $n = 50$, $n = 100$, $n = 200$, $\sigma^2 = 0,1$ and $K = 5$, $K = 20$. So it can be concluded that even though K is getting bigger it has no effect on the UBR method, because the UBR method tends to be smaller. So of these three methods the UBR method is more suitable and better than the CV method and GCV method for the Fourir Series estimator in nonparametric regression. The complete results for this can be seen in Appendix 5.

REFERENCE

- Budiantara, I N., dan Subanar, 1997, Pemilihan Parameter Penghalus Dalam Regresi Spline, *Majalah Ilmiah Matematika dan Ilmu Pengetahuan Alam*, 7,37 – 49.
- Budiantara, I N., 2000 *Estimator Spline Dalam Regresi Nonparametrik dan Semiparametrik*, Disertasi Doktor pada Universitas Gadjah Mada, Yogyakarta
- Budiantara, I. N., Subanar, and Soejoeti, Z., 1997, *Weighted Spline Estimator*, *Bulletin of the International Statistical Insitute*, 51, 333-334.
- Budiantara, I. N., dan Subanar, 1997, *Sifat Asimtotik Estimator Spline Terbobot*, *Majalah Berkala Matematika dan Ilmu Pengetahuan Alam (BMIPA)*, Universitas Gadjah Mada, 2, 23-36.
- Budiantara, I. N., 2001b, *Estimasi Parametrik dan Nonparametrik untuk Pendekatan Kurva Regresi*, *Makalah Pembicara Utama pada Seminar Nasional Statistika V*, Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS), Surabaya.
- Budiantara, I. N., 2002a, *Estimator Tipe Penalized Likelihood*, *Jurnal Natural FMIPA Unibraw*, Edisi Khusus, 231-235.
- Budiantara, I. N, 2004, *Spline Historis, Motivasi, dan Perannya Dalam Regresi Nonparametrik*, *Makalah Pembicara Utama pada Konferensi Nasional Matematika XII*, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana (UNUD), Denpasar.
- Dani, T. A., & Adrianingsih, Y. N. (2021). *Pemodelan Regresi Nonparametrik dengan Estimator Spline Truncated dan Deret Fourier*. *Jambura Journal Of Mathematics*, 26-36.
- Dani, A. T. R., & Adrianingsih, N. Y. (2021). *Pemodelan Regresi Nonparametrik dengan Estimator*



- Spline Truncated vs Deret Fourier. Jambura Journal of Mathematics, 3(1), 26–36.*
<https://doi.org/10.34312/jjom.v3i1.7713>
- Dewanti, D., Suparti., dan Prahutama, A. *Pemodelan Indeks Harga Saham Gabungan (IHSG) dan Jakarta Islamic Index (JII) Menggunakan Regresi Birespon Spline Truncated Berbasis GUI R. Jurnal Statistika. 2020;8(2):134-143.*
- Eubank, R.L., 1988, *Spline Smoothing and Nonparametric Regression*, Mercel Dekker, New York.
- Faisol, , Ukhrowi, P., Mardianto, M. F., Yudistira, I., & Kuzairi, . (2022). *Comparison Of Salinity And Seawater Temperature Predictions Using Var And Biresponse Fourier Series Estimator . Barakeng: Jurnal Ilmu Matematika dan Terapan, 1465- 1476*
- Fatima, S. (2019). *Estimator Deret Fourier Dalam Regresi Nonparametrik Birespon Untuk Prediksi Beban Puncak Listrik Di Madura Ketika Siang dan Malam Hari. Pamekasan: Universitas Islam Madura*
- Fernandes, A.A.R., Budiantara, I.N., Otok, B.W., dan Suhartono. *Spline Estimasi for Bi-responses Nonparametric Regression Model for Longitudinal Data. Jurnal Applied Mathematical Sciences. 2014;8(114):5653-5665*
- Hardle, W.,1990, *Applied Nonparametric Regression*, Cambridge University Press, New York.
- Kawata, T., 1972, *Fourier Analysis in Probability Theory*, Academic Press, New York.
- Kuzairi, , Azizah, N., Mardianto, M. F., Yudistira, I., Faisol, , Yulianto, T., & Amalia, R. (2022). *The Parametric And Nonparametric Estimator In Semiparametric Regression For Longitudinal Data With Spline Approach. Kursor, 187-194*
- Mardianto, M. F., & Budiantara, I. (2014). *Estimasi Model Regresi Semiparametrik Birespon Dengan Pendekatan Deret Fourier. Proseding Seminar Matematika Unud Denpasar*
- Mardianto, M., Tjahjono, E., & Rifada, M. (2019). *Semiparametric Regression Based on Three Forms of Trigonometric Function in Fourier Series Estimator. IOP Publishing.*
- Nurhuda, G., Wasono., dan Nohe, D.A. *Pemodelan Regresi Nonparametrik Estimator Spline Truncated pada Data Simulasi. Jurnal Matematika, Statistika. dan Komputasi. 2022;19(1):172-182.*
- Pratiwi, L. P. S. (2020). *Pemilihan Titik Knot Optimal Model Spline Truncated dalam Regresi Nonparametrik Multivariabel dengan GCV. Jurnal Matematika, 10(2), 78.*
<https://doi.org/10.24843/jmat.2020.v10.i02.p125>
- Padatuan, A. B., Sifriyani, S., & Prangga, S. (2021). *Pemodelan Angka Harapan Hidup dan Angka Kematian di Kalimantan dengan Regresi Nonparametrik Spline Birespon. BAREKENG: Jurnal Ilmu Matematika Dan Terapan, 15(2), 283– 296.*
<https://doi.org/10.30598/barekengvol15iss2pp283-296>
- Rahmawati, R.D., Suparti., dan Prahutama, A. *Pemodelan Inflasi dan Indeks Harga Saham*



- Gabungan Menggunakan Regresi Nonparametrik Birespon Spline Truncated dengan Pembobot Invers Matriks Variansi-Kovariansi Error Respon. Jurnal Statistika. 2019;7(1):20-38.*
- Sihotang, F.R., Sifriyani., dan Prangga, S. Aplikasi Regresi Nonparametrik Spline Birespon pada Data Kualitas Air di DAS Mahakam. Jurnal Media Bina Ilmiah. 2021;16(3):6515-6525*
- Salam, O. N., Sukmawaty, Y., & Halida, A. (2022). Estimasi Model Regresi Nonparametrik dengan Metode B-Spline. Open Journal Systems, 16(10), 7631–7638. <http://ejurnal.binawakya.or.id/index.php/MBI>*
- Tripina, A, Agung Prabowo, Yosita Lianawati, Abdul Talib Bon (2021). Estimated Spline in Nonparametric Regression with a Generalized Cross Validation and Unbiased Risk Approach, Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management Singapore, 3788-3798*
- Tripina, A. and Budiantara, I N., 2006, Fourier Estimator in Nonparametric Regression, International Conference On Natural Sciences and Applied Natural Sciences, Ahmad Dahlan University, Yogyakarta.*
- Tripina, A. 2005. Pendekatan Model Regresi Spline Linier . Jurusan MIPA, Fakultas Sains dan Teknik, UNSOED.*
- Tripina, A., 2011. Analisis Regresi Spline Kuadratik. Prosiding Seminar Nasional Matematika dan Pendidikan Matematika Jurusan Pendidikan Matematika FMIPA UNY. Yogyakarta. ISBN : 978-979 -16353-6-3*
- Wahba, G., 1990. Spline Models For Observasion Data, SIAM Pensylvania.*
- Zahrah, M. (2020). Prediksi Tingkat Hunian Hotel Di Kabupaten Pamekasan Berdasarkan Estimator Deret Fourier dengan dan Tanpa Tren. Pamekasan: Universitas Islam Madura.*

