

PAPER • OPEN ACCESS

Simulations Study Combined Estimator Fourier Series and Spline Truncated in Multivariable Nonparametric Regression

To cite this article: I Wayan Sudiarsa 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **546** 052074

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the **collection** - download the first chapter of every title for free.

Simulations Study Combined Estimator Fourier Series and Spline Truncated in Multivariable Nonparametric Regression

I Wayan Sudiarsa¹

¹Department of Mathematics, Faculty of Mathematics and Natural Sciences, Institut Keguruan dan Ilmu Pendidikan (IKIP) PGRI Bali, Indonesia

*Corresponding author: wsudiarsa72@yahoo.com

Abstract. The problems one of that appeared in everyday life is how to explain the shape of the relation pattern between a response with some predictor. Regression analysis is a statistical method used to estimate the relation pattern between response (y) with predictor (x). Development of nonparametric regression of Fourier series involving multiple predictor, has been more developed for predictors of the same pattern. We need to develop different estimator for different predictors of multivariable nonparametric regression. Theoretical research will be focused on estimator form, and applied to simulation data. The estimation of the combined regression function of the Fourier and spline truncated sequence is obtained through the optimization of penalized least square (PLS). Based on the simulation result obtained that the larger the sample size and the smaller the size of the variance, it will result in better estimation value of parameter and knot.

Keywords: simulations, Fourier series, spline truncated, Penalized Least Square, multivariable nonparametric regression

1. Introduction

In general, regression analysis is a statistical method used to explain the relation between a response variable with one or more predictor variable. Parametric regression is the most commonly used method for solving the modeling of the relation between variable if the functional relation is known or assumed to follow a particular function. In its development, parametric regression generally consists of two groups of model, namely linear regression model and nonlinear regression model. In the last decade, many studies of nonparametric regression have been used to explain the relation between response and predictor variable when the functional relation information is unknown. As computation progresses and some limitations to the parametric regression model, nonparametric regression model that do not require many assumptions become more applied to solve a problem in the various applied field [7].

Spline as a data pattern approach was introduced by Whittaker in 1923. While the spline based on an optimization problem was developed by Reinsch in 1967 [9]. The spline approach has a function basis. Commonly used functional bases include spline truncated and B-spline [6]. The truncated spline is a function where there are changes in different curve behavior pattern at different interval. Spline is one form of estimator that is also often used in nonparametric regression because it has good visual interpretation, flexible, and able to handle the character of the function is seamless [4] and [2].

Recent development relating to research on nonparametric regression models show that the Fourier series is one of the alternative estimators that have recently been reviewed and developed by



nonparametric regression researchers. [3] is a group of early researchers who examined the Fourier transform sequences for the refinement of density functions, particularly in the refinement of spectral estimators. In 1992, Bilodeau studied a Fourier-series estimator on nonparametric regression with an additive predictor component and a penalized least square to obtain the coefficients. The Fourier sequence estimate used is the sum of a linear function and a trigonometric polynomial function.

Development of nonparametric regression of Fourier series on several predictor variables collectively that have been done by [1] using Fourier series estimators for all predictors in the form of an additive model. The form of nonparametric regression model with many additives (multivariable) predictors has been developed in several predictor variables known as Generalized Additive Model (GAM) [5]. One of the open questions that emerged from [1] research is the opportunity to develop different estimators for each predictor of nonparametric regression involving many predictors. Sudiarsa [8], discovered the combined estimator form of the Fourier and spline truncated series in multivariable nonparametric regression using the penalized least squares method. The theoretical results will be applied to empirical studies that include simulated studies to investigate the goodness of the resulting regression curves.

2. Combination Shape Estimator between Fourier Series and Spline Truncated in Multivariable Additive Nonparametric Regression Model

Theorem 1

By Sudiarsa [8] page 5003 Let y_i is a multivariable additive nonparametric regression model as in Eq. (1). If the estimator of combination between Fourier series and spline truncated obtained from PLS optimization as in Eq. (5), then the estimator of $g_j(x_j)$, $j = 1, 2, 3, \dots, p$ is as follows:

$$\hat{g}_1(x_1) = \hat{b}x_1 + \frac{1}{2}\hat{a}_0 + \sum_{k=1}^K \hat{a}_k \cos kx_1,$$

and

$$\hat{g}_j(x_j) = \sum_{v=1}^m \hat{\beta}_{vj} x_j^v + \sum_{u=1}^r \hat{\alpha}_{uj} (x_j - t_{uj})_+^m, j = 2, 3, \dots, p,$$

where \hat{b} , \hat{a}_0 , \hat{a}_k , $k = 1, 2, \dots, K$, $\hat{\beta}_{vj}$, $v = 1, 2, \dots, m$ and $\hat{\alpha}_{uj}$, $u = 1, 2, \dots, r$, $j = 2, 3, \dots, p$ are obtained from equation

$$\hat{\underline{a}}(K, t, \lambda) = (\hat{b}, \hat{a}_0, \hat{a}_1, \dots, \hat{a}_K)' = A(K, t, \lambda) \underline{y},$$

$$\hat{\underline{\beta}}(K, t, \lambda) = (\hat{\beta}_{11}, \dots, \hat{\beta}_{m2}, \hat{\alpha}_{12}, \dots, \hat{\alpha}_{r2}, \dots, \hat{\beta}_{1p}, \dots, \hat{\beta}_{mp}, \hat{\alpha}_{1p}, \dots, \hat{\alpha}_{rp})' \\ = B(K, t, \lambda) \underline{y},$$

$$A(K, t, \lambda) = \left(S(K, \lambda) W I - S(K, \lambda) W X \left(I - (X X)^{-1} X' W S(K, \lambda) W X \right)^{-1} (X X)^{-1} X' (I - W S(K, \lambda) W') \right),$$

$$B(K, t, \lambda) = \left(I - (X X)^{-1} X' W S(K, \lambda) W X \right)^{-1} (X X)^{-1} X' (I - W S(K, \lambda) W'),$$

$$S(K, \lambda) = (W W + n \lambda D)^{-1}.$$

Hence, the estimator of combination between Fouries series and spline truncated is

$$\hat{\mu}(x_{1i}, x_{2i}, \dots, x_{pi}) = \hat{g}_1(x_1) + \sum_{j=2}^p \hat{g}_j(x_j).$$

3. Simulations Study Combination Estimator between Fourier Series and Spline Truncated in Multivariable Additive Nonparametric Regression Model

Simulations were performed using data generated from the Fourier series function with various K oscillation and spline truncated linear polynomial functions with knot point variation. Model of experimental function simulation $K_1 r_1$ is a combined function between $g_1(x_{1i})$: Fourier series function with $K=1$ and $g_2(x_{2i})$: linear spline truncated linear polynomial function with one knot point. The regression equations designed for this simulation study are as follows:

$$y_i = g_1(x_{1i}) + g_2(x_{2i}) + \varepsilon_i, i = 1, 2, \dots, n \quad (1)$$

$$= 0,1x_{1i} + 1 + 0,9 \cos\left(\frac{2\pi x_{1i}}{100}\right) + 2x_{2i} - 4(x_{2i} - 0,45)_+ + \varepsilon_i$$

Furthermore, from equation (1) is raised: $x_{1i} \sim U(0,1)$, $x_{2i} \sim U(0,1)$ and $\varepsilon_i \sim N(0, \sigma^2)$ with $\sigma^2 = 0.1$ and sample size $n = 100$. From this equation will be obtained data generated (x_{1i}, x_{2i}, y_i) , $i = 1, 2, \dots, 100$. Plot between response y_i with predictors x_{1i} and y_i with predictors x_{2i} is presented in Figure 1.

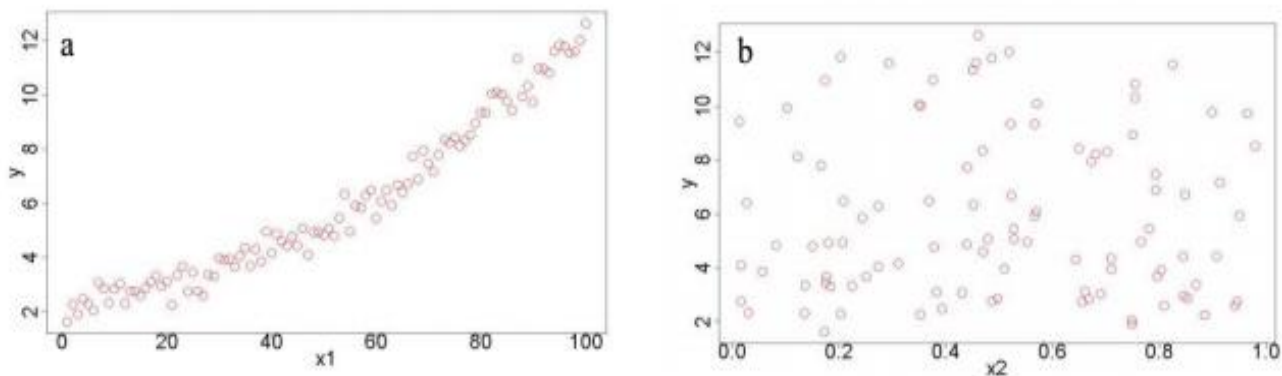


Figure1. Between Plot y with x_1 (a), and y with x_2 (b)

Plot of dimensional three (x_{1i}, x_{2i}, y_i) is presented in Figure 2.

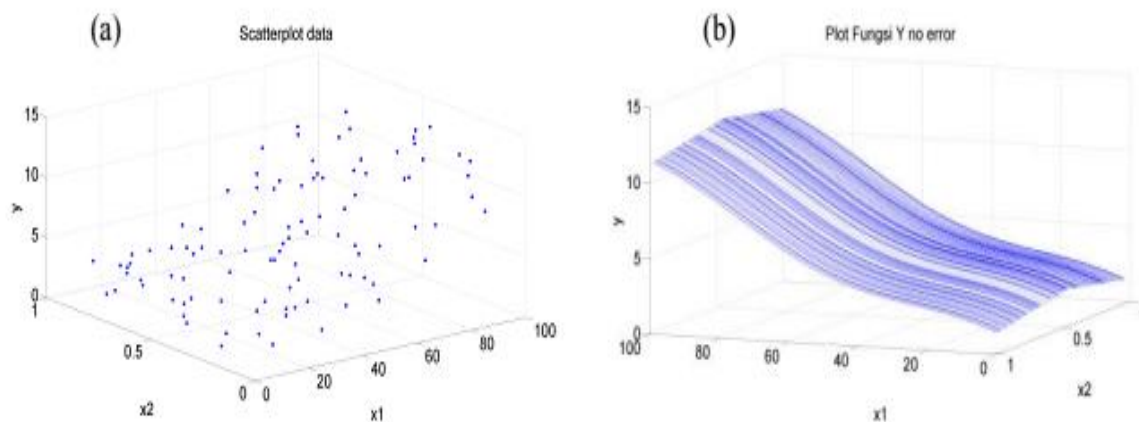


Figure 2. Scatterplot of y with x_1 and x_2 (a), and regression curve y (b)

Based on Figures 1 and Figure 2 it can be described that the relationship pattern between y_i with x_{1i} tends to change in sub interval. While the relationship between y_i and x_{2i} not following a certain pattern. So we tried a combination model of Fourier series and spline truncated. The combined regression model of the Fourier and spline truncated series is as follows:

$$y_i = bx_{1i} + \frac{1}{2}a_0 + \sum_{k=1}^3 a_k \cos k \left(\frac{2\pi x_{1i}}{n} \right) + \beta x_{2i} + \sum_{u=1}^3 \alpha_u (x_{2i} - t_u)_+ + \varepsilon_i \quad (2)$$

Furthermore, the result of simulation study on the models $K_1 r_1$ which consists of the output the R programs presented in full. Table 1 presented the minimum GCV , Smoothing parameters λ optimal, R^2 and MSE .

Table 1. Value of GCV , smoothing parameters λ optimal, R^2 and MSE Function $K_1 r_1$

Replay	Osila tion: K	The many Knot: r	GCV	$\lambda_{optimal}$	R^2	MSE	Location Knot		
							1	2	3
1	3	3	0.104	0.0605	98.96	0.095	0.23	0.52	0.81
2	1	1	0.110	0.0007	98.82	0.105	0.45	*	*
3	3	1	0.110	0.0015	98.83	0.102	0.46	*	*
4	1	3	0.108	0.0008	98.85	0.200	0.23	0.54	0.80
5	1	1	0.095	0.0006	99.04	0.091	0.44	*	*
6	2	1	0.089	0.0055	99.05	0.084	0.44	*	*
7	1	1	0.073	0.0005	99.21	0.073	0.43	*	*
8	1	3	0.099	0.0006	99.00	0.094	0.26	0.51	0.79
9	2	1	0.103	0.0030	98.91	0.097	0.44	*	*
10	3	1	0.096	0.0015	99.05	0.090	0.47	*	*

The best Fourier and spline truncated nonparametric regression models are obtained from optimum K values and optimum knots, using the smallest GCV . The GCV Plot (λ) at $K=1$ and $r=1$ repeat 7 is presented in Figure 3 as follow:

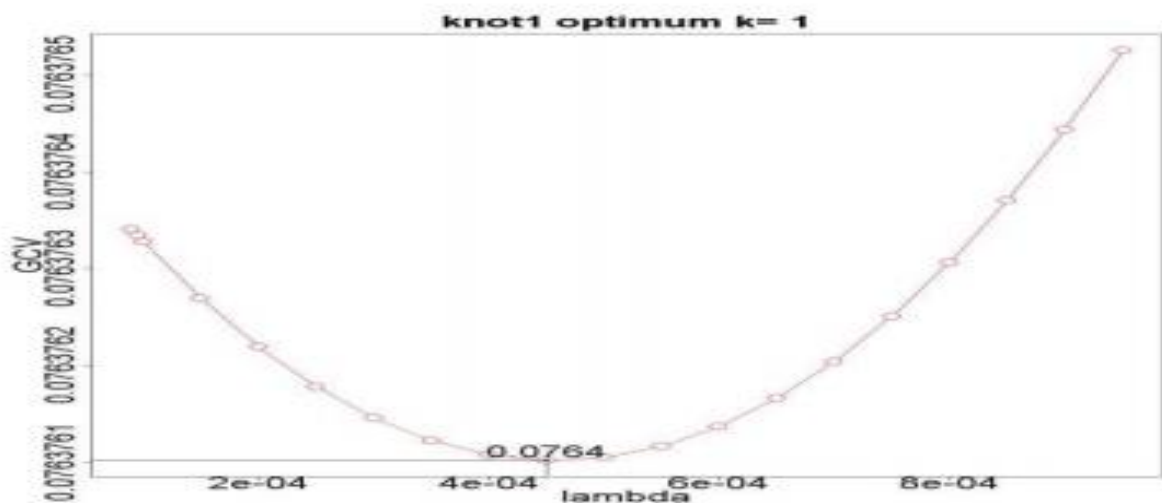


Figure 3. Plot GCV (λ) of the function models $K_1 r_1$

Referring to Figure 3 and Table 1 show that the minimum $GCV(\lambda)$ minimum is 0.076 and value λ optimal is 0.00045, with $K = 1$ and the number of knot one knot and the location of the knot point is 0.43. The combined estimators of the Fourier and spline truncated series depend on the value of the smoothing parameters λ . Figure 4(a) present the combined regression curve estimators of Fourier and spline truncated series with value $\lambda \rightarrow 0$ (small). Figure 4(b) present the combined regression curve estimators of Fourier and spline truncated series with value $\lambda \rightarrow \infty$ (large). While Figure 4(c) present the estimators of the combined regression curve of Fourier and spline truncated series with λ optimal, as follow:

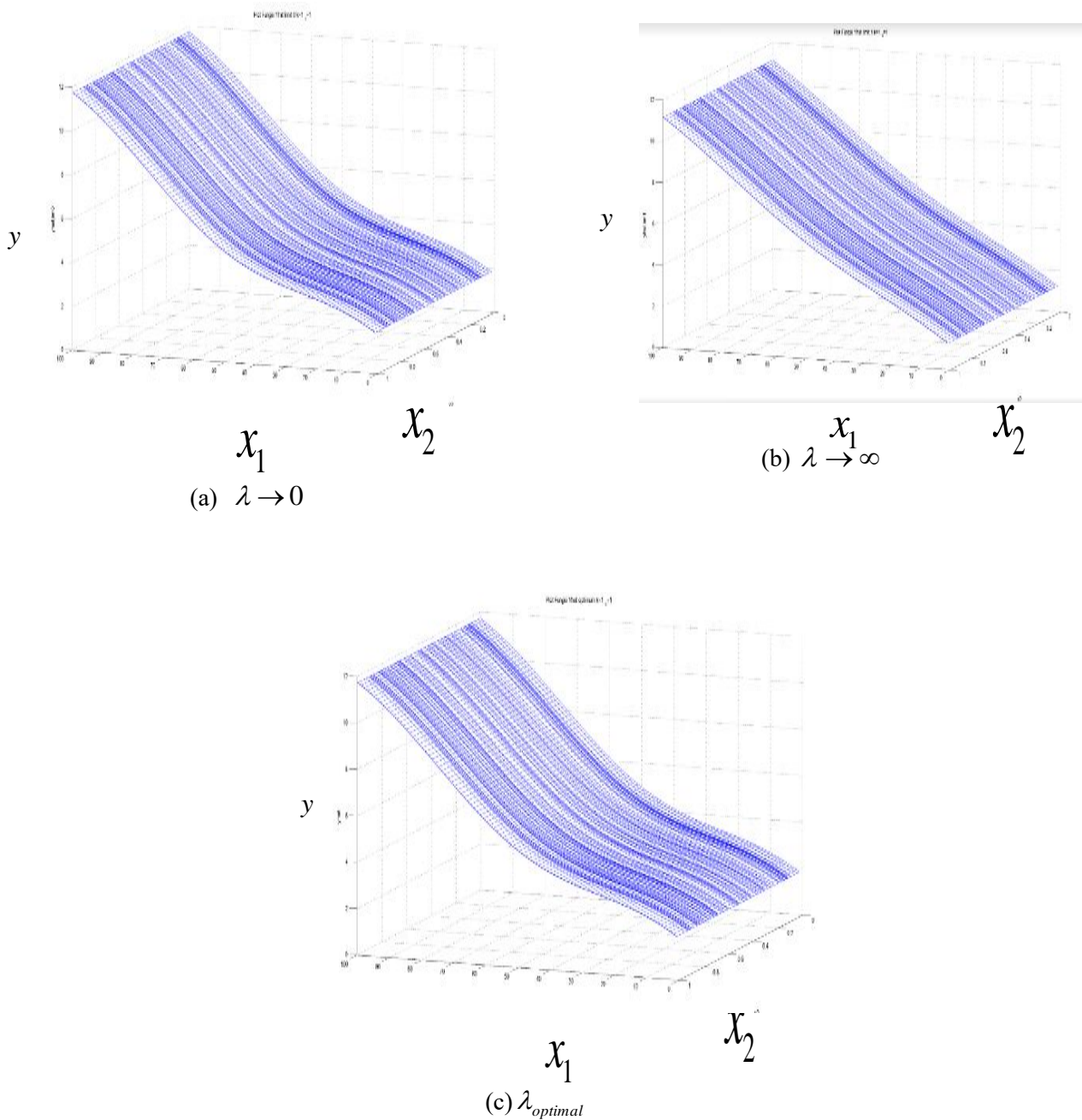


Figure 4. Plot of regression models $K_1 r_1$ based on λ optimal and GCV minimum.

Have a period from Figure 4(a) it can be visually described that the combined model approach of the Fourier and spline lines is truncated with λ close to zero gives a roughly (local). While figure 4(b) with value $\lambda \rightarrow \infty$ (Large) providing a smooth (global) approximation. As well λ optimal approach is best seen from figure 4(c).

Table 2 presents the estimated parameters and optimal knot points of the experimental model $K_1 r_1$ occurred in replications of 2, 2.5, and 7.

Table 2. Parameter Estimation and Knot Equation Function Model $K_1 r_1$.

No	Parameters Models and Knot	Value Parameters and Knot Simulation Models	Estimations Parameters and Knot Location		
			Replay 2	Replay 5	Replay 7
1	b	0.1	0.09	0.10	0.10
2	a_0	1	0.86	0.89	0.77
3	a_1	0.9	0.89	0.88	0.91
4	β_1	2	2.15	2.21	2.35
5	α_1	-4	-4.23	-4.39	-4.33
6	t_1	0.45	0.45	0.44	0.43

Based on the simulation results in Table 2, the estimation of the combined function of the Fourier and spline truncated series in the multivariable nonparametric regression in the 7th repetition is as follows:

$$\hat{y}_i = 0.1x_{1i} + 0.77 + 0.91\cos\left(\frac{2\pi x_{1i}}{100}\right) + 2.35x_{2i} - 4.33(x_{2i} - 0.43)_+ \quad (3)$$

It can be seen from Table 2 that the value of parameter estimation and knot point is close to the value of the specified model or simulated model. This model has a value of $R^2 = 99.21168$ and $MSE = 0.072528$. Models simulation $K_1 r_1$ is a combined model of Fourier series with $K=1$ and a linear spline truncated linear polynomial function with one knot point, $n=100$ and $\sigma^2 = 0.1$, similarly presented in boxplot form. Simulation data is generated with variation of sample size $n=50$, $n=100$ and $n=200$ variation of variance size, $\sigma^2 = 0.1$, $\sigma^2 = 0.5$ and $\sigma^2 = 1$, variation $K=1$, $K=2$ and $K=3$ and the variation of the knot point size one knot point, two knot points and three knots point. Figure 5(a) and (b) presents the results of a simulated model study of Fourier and spline truncated series $K_1 r_1$ for the exactness of the sample size n and parameter estimation b , as follows:

Compatibility models for sample size variation $n_1 = 50$, $n_2 = 100$ and $n_3 = 200$.

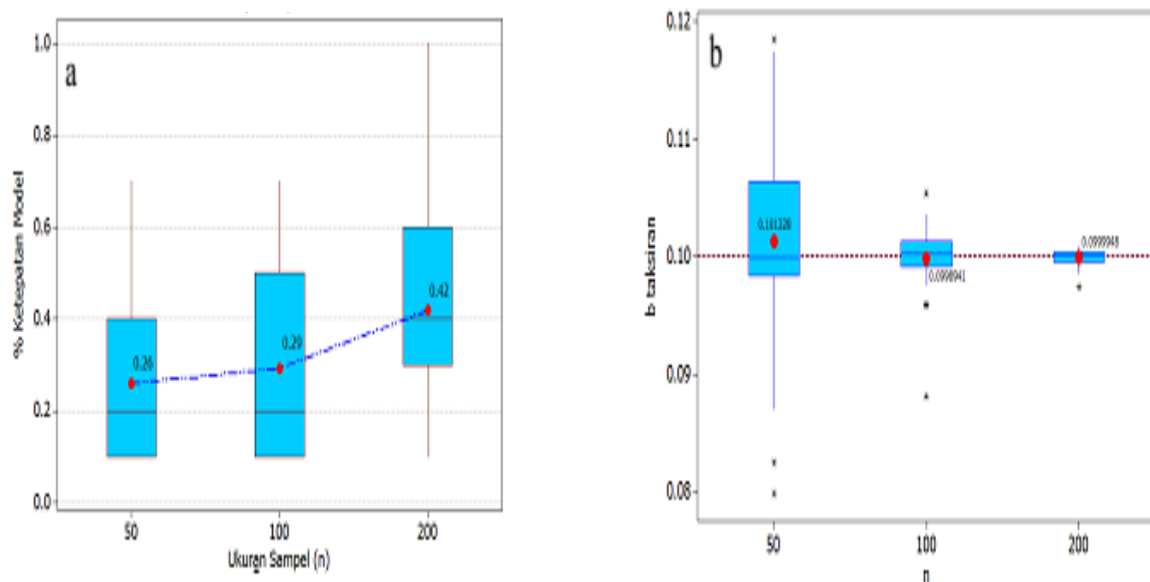


Figure 5. Compatibility models for sample size n (a), and estimate b (b).

Based on Figure 5(a), it is known that if the larger sample size of n , then the results obtained better model accuracy or more in accordance with the simulation model. While Figure 5(b) shows that if the larger sample size n , then obtained the average estimation of the parameter b tend to be better or closer to the real value. Furthermore, in Figure 6(a) and (b), the results of simulation study of combined models of Fourier series and spline truncated $K_1 r_1$ for parameter estimation a_0 and a_1 , as follow:

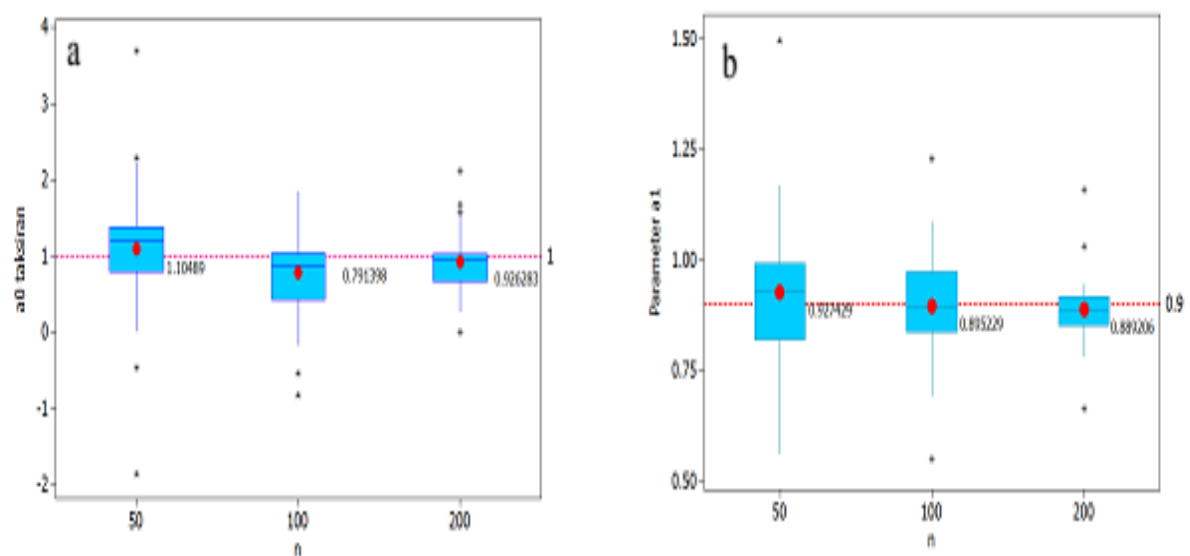


Figure 6. Estimation a_0 (a), and Estimation a_1 (b)

Referring to Figure 6(a) and (b), obtained data that if the larger sample size n , then tend to average parameter estimation a_0 and a_1 the better or closer to the real value. Next figure 7(a), (b) and (c) presented the result of simulation study of combined model estimator of Fourier and spline truncated series $K_1 r_1$, for parameter estimation β , α_1 and knot point value, as follow:

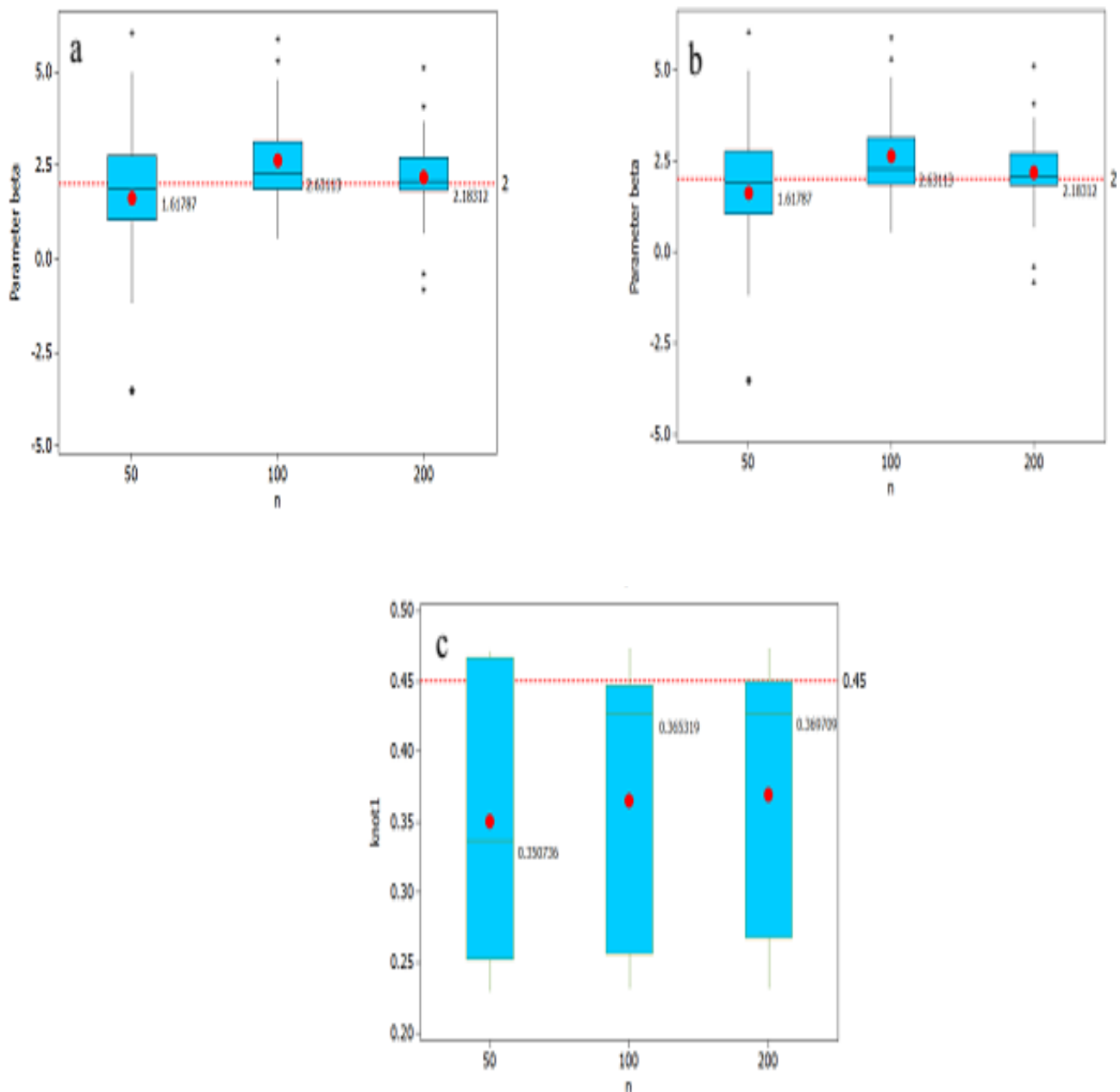


Figure 7. Estimation β (a), Estimation of the knot point t (b), and Estimation α_1 (c)

Figure 7a and b show simulation data information that if the large the sample size n , then to the average parameter estimation β and α_1 the better or closer to the real value. From Figure 7c, it can be seen that the bigger of sample sizes, n , the bigger the estimation of knot point t the better or closer to the true knot point value. Next plot the result of the combined estimator of the Fourier and spline truncated series in the simulation experiment function $K_1 r_1$ Figure 8(a) and (b) show plot regression models f , and regression models \hat{f} , while Figure 8(c) presented the regression models f with \hat{f} .

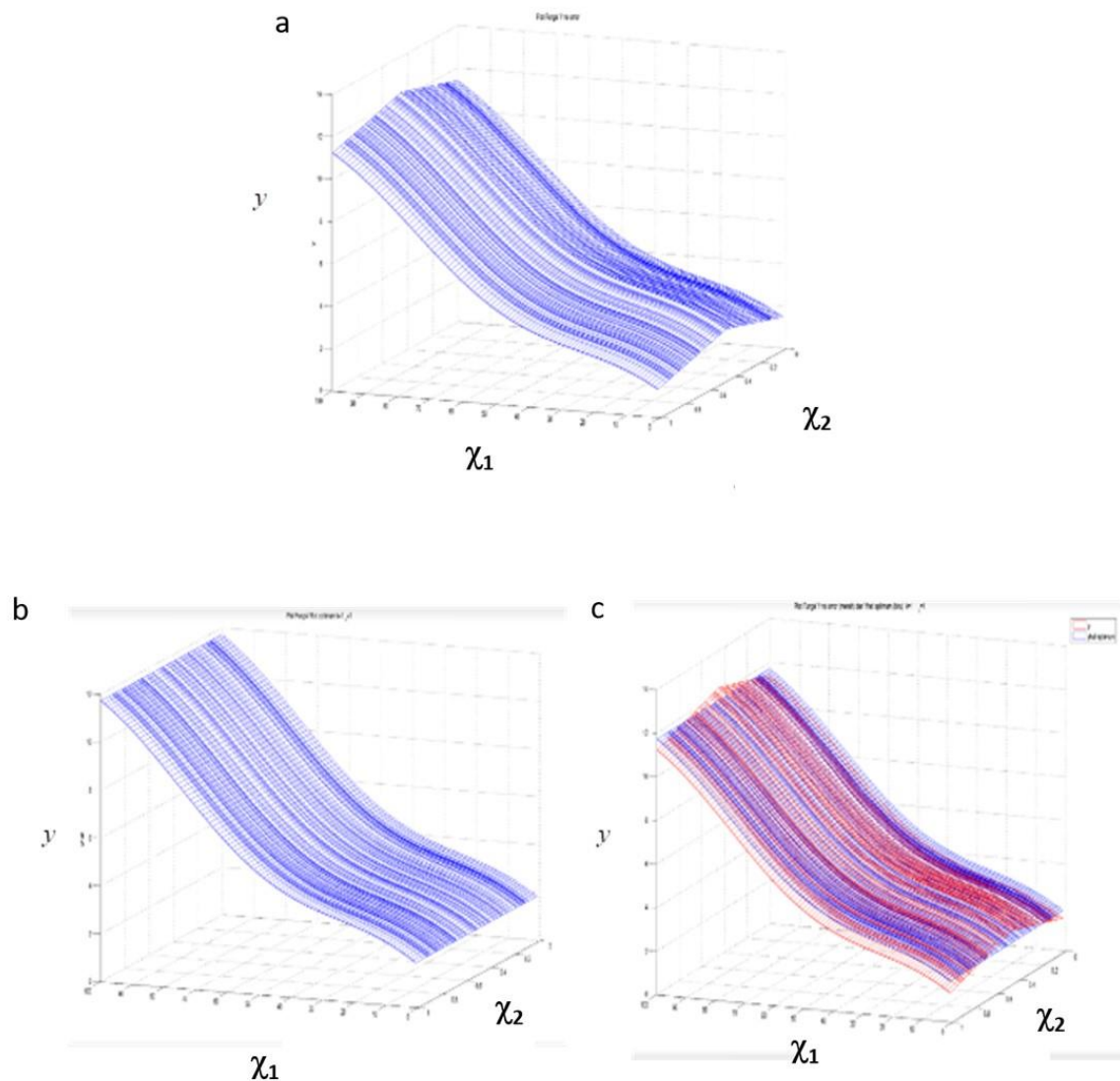


Figure 8. Plot model Regression f (a); model regression \hat{f} (b), and model regression f with \hat{f} (c)

Based on Figure 8(b) it can be shown that between f and \hat{f} tend to be almost the same size (squeeze). This shows that the combined estimator of the Fourier and spline truncated series is very good for use in modeling this data.

4. Conclusion

Based on simulation result that have been describe before, the obtained the following conclusions: If sample size n the greater it is, then the combined model estimator of the Fourier and spline truncated series the, for all variations of variance, σ^2 , number of knot point and oscillation parameter K . If the variance σ^2 the smaller, the combined model estimator of the Fourier and spline truncated series is getting better or closer to the actual models, for all combination of sample n , the number of knot points and the oscillation parameter K . If the knot point is used more and more, then there is a tendency of combined estimators of Fourier and spline truncated series the better, for all combinations of sample n , variance σ^2 and the oscillation parameter K . However, too many knots will also produce very complex models, thus, it is needed to select the optimal knot point. If the size of the oscillation parameter K is more numerous, then the combined model estimator of the Fourier and spline truncated series is not getting better. Therefore, it will be necessary to select the optimal parameter of K oscillation. If the fining parameters are very large ($\lambda \rightarrow \infty$), the combined model estimate of the Fourier and spline truncated series is very smooth (global). If the fining parameter are very small ($\lambda \rightarrow 0$), the combined model estimator of the Fourier and spline truncated series is very rough (local).

References

- [1] Bilodeau, M., Fourier Smoother and Additive Models, *The Canadian Journal of Statistics*, 3 (1992) 257-269
- [2] Budiantara, I. N., 2007, Inferensi Statistik untuk Model Spline, *Jurnal Mat-Stat*, 7, 1-14
- [3] De Jong, P., The Fast Fourier Transform Spectral Estimator, *Journal of Statistical Soc. B* 39 (1977) 3,327-330
- [4] Eubank, R.L., *Nonparametric Regression and Spline Smoothing*, Second Edition, New York, Marcal Dekker, Inc 1999
- [5] Hastie, T.J. and Tibshirani, R.J., *Generalized Additive Models*, Chapman and Hall, New York 1990
- [6] Lee, T. C. M., 2004 Improved Smoothing Spline Regression by Combining Estimatics of Different Smoothnes, *Statistics and Probability Letters*, 67. 133-140
- [7] Machler, M.B., Variational Solution of Penalized Likelihood Problem and Smooth Curve Estimation, *Annals of Statistics*, 23 (1995) 1496-1517
- [8] Sudiarsa, I.W., Budiantara, I.N., Suhartono and Santi Wulan, P. Combined Estimator Fourier Series and Sline Truncated in Multivariable Nonparametric Regression. *Journal of Applied Mathematical Sciences*. 9 (2015) 4997-5010
- [9] Wahba G.,1990, *Spline Models For Observasion Data*, SIAM Pensylvania