

## Modelling spline truncated and local polynomial for inflation sectors in Indonesia

Suparti\*, Alan Prahutama, Rukun Santoso

Statistics Department, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia

Corresponding Author Email: [suparti702@gmail.com](mailto:suparti702@gmail.com)

[https://doi.org/10.18280/mmc\\_d.390105](https://doi.org/10.18280/mmc_d.390105)

**Received:** 23 May 2018

**Accepted:** 10 July 2018

### Keywords:

*spline truncated, local polynomial, inflation sectors in Indonesia*

### ABSTRACT

The regression model can be approximated by parametric and nonparametric methods. The parametric regression method generates an excellent regression model when the shape of the curve is known, whereas if the curve shape is random, it can be approximated by using a nonparametric regression method. There is a nonparametric regression method that has been developed such as spline truncated and local polynomials. Spline truncated is segmented pieces regression, while the local polynomial is polynomial models with kernel function as weighted. The most important thing in nonparametric regression modeling is the selection of smoothing parameters. One of the selected parameters of the method is the Generalized Cross Validation (GCV) method. The purpose of this study is to generate the model of Inflation's sectors in Indonesia using Spline Truncated and local polynomial. These sectors include foodstuffs sector; food, beverages, cigarettes, and tobacco sector; housing, water, electricity, gas, and fuel sector; clothing sector; health sector; education and sports sector; as well as transportation, communication, and financial services group. The results indicated that by modeling the value of the inflation sectors in Indonesia using Spline truncated resulted in average R-square is 68.86% while for local polynomial modeling, the average R-square is 73.73%.

## 1. INTRODUCTION

Regression modeling is modeling between predictor variables and response variables. The modeling that is often done is simple linear regression modeling or multiple linear regression modeling. It includes parametric modeling which requires assumptions such as normally distributed residuals, homoscedasticity, and non-autocorrelation. Parametric modeling is preferred because it is easy to model. In contrast to the parametric approach, a nonparametric approach is an approach in the regression model that is not bound by assumptions. If the data curve spreads randomly and does not form linear, multiple or cubic patterns, then parametric modeling will result in poor model estimation. In regression modeling with nonparametric approach, the model estimation will find its form of function curve. Nonparametric regression modeling can be done with kernel approach; spline truncated, local polynomial, Fourier and Wavelet. The flexible and well-developed approach in modeling is spline truncated. The spline is one type of polynomial cut, which is a polynomial that has a segmented property. This segmented nature provides more flexibility than ordinary polynomial, making it possible to adapt more effectively to local characteristics of a function or data. Spline also has a very good ability to handle data type whose behavior fluctuates on certain sub-intervals. The concept of the spline is to define the spline order and determine the knot point. However, the most dominant aspect that determines the curve performance is knot point. One of the ways to determine optimum knot point is by using minimum GCV. There are some studies on Spline regression such as examined the nature of asymptotic on spline regression [5]. Spline regression is applied to analyze the relationship

between information technology and firms [6]. Another development of spline is to examine the biased nature of spline regression [4].

One of the approaches of nonparametric regression model that has been developed in addition to Spline and Kernel is Local Polynomial. Local Polynomial has several advantages, over another method, such as to reduce asymptotic bias and produce good estimates [2]. Local Polynomial Estimates can use WLS (Weighted Least Square) by minimizing the quadratic of error [7]. In the local polynomial regression, the smooth degree of its function is determined by its bandwidth. Optimal bandwidth determination can use GCV (Generalized Cross Validation) method [2, 7]. Research on Local Polynomial has been widely developed, on local polynomial regression in Complex Surveys [3], and nonparametric polynomial regression with error autocorrelation [9]. It also has been developed to determine the bandwidth with plug-in for local polynomial regression with correlated error [8], and with the adaptive design [1].

Inflation is a trend or movement of general price level rises continuously from one period to the next. Controlled and low inflation can support the maintenance of people's purchasing power. While unstable inflation will complicate the business world in business planning, both in production and investment activities and in determining the price of goods and services it produces. For the government, estimated inflation is needed for preparation of the Draft State Budget and the community or investors as a consideration in investment planning.

The inflation rate in Indonesia as measured by the Consumer Price Index (CPI) can be divided into seven expenditure groups, i.e. foodstuffs group; food, beverages, cigarettes and tobacco group; housing, water, electricity, gas

and fuel group; clothing group; health group; education and sports group; as well as transportation, communication and financial services group. The seven expenditure groups greatly influenced the inflation in Indonesia.

Spline truncated and local polynomial modeling in this study focuses on determining the smoothing parameters (knots and bandwidth) in trial and error. The modeling generated from this study is used to examine the modeling comparison generated on the spline and local polynomial for inflation cases in Indonesia.

## 2. LITERATUR REVIEW

### 2.1 Spline truncated

Given the regression model equation as follows:

$$y_i = f(t_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

with  $y_i$  is response variable whereas  $f(t_i)$  a function is the regression curve with  $t_i$  as the predictor variable and  $\varepsilon_i$  is a random error which is assumed to be independent normal distributed with mean zero and variance  $\sigma^2$  [5].

The spline is a segmented piecewise polynomial that has flexibility nature. This flexibility nature is what distinguishes the spline with the polynomial. The joint fusion point of the pieces or points indicating the occurrence of changes in the curves behavior at different intervals is called knots.

In general, the spline function of  $m$  order is any function that can be written in the form of [7]:

$$f(t_i) = \sum_{j=1}^m a_j t_i^j + \sum_{k=1}^M b_k (t_i - K_k)_+^m \quad (2)$$

$$\text{with } (t_i - K_k)_+^m = \begin{cases} (t_i - K_k)^m & ; t_i \geq K \\ 0 & ; t_i < K \end{cases}$$

$a_j$  moreover,  $b_k$  are parameters and  $K_1, K_2, \dots, K_M$  are knot points,  $f(t_i)$  is a regression function, and  $t$  is a predictor variable.

If  $m = 1$  and the number of knots is one, then the above equation can be written as linear spline function:

$$f(t_i) = a_1 t_i + b_2 (t_i - K)_+^1 \quad (3)$$

this function is called the linear spline of one knot at  $t_i = K$

Eq. (3) can also be written as

$$f(t_i) = \begin{cases} a_1 t_i & ; t_i < K \\ a_1 t_i + b_2 (t_i - K) & ; t_i \geq K \end{cases} \quad (4)$$

Spline regression model can be presented in the form of:

$$y_i = \sum_{j=1}^m a_j t_i^j + \sum_{k=1}^M b_k (t_i - K_k)_+^m + \varepsilon_i \quad (5)$$

Furthermore, model (2.5) can be written as:

$$y_i = a_1 t_i + \dots + a_m t_i^m + b_1 (t_i - K_1)_+^m + \dots + b_M (t_i - K_M)_+^m + \varepsilon_i$$

If the regression model is presented in matrix form then we obtain:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{bmatrix} t_1 & \dots & t_1^m & (t_1 - K_1)_+^m & \dots & (t_1 - K_M)_+^m \\ t_2 & \dots & t_2^m & (t_2 - K_1)_+^m & \dots & (t_2 - K_M)_+^m \\ \vdots & & \vdots & \vdots & & \vdots \\ t_n & \dots & t_n^m & (t_n - K_1)_+^m & \dots & (t_n - K_M)_+^m \end{bmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_m \\ b_1 \\ \vdots \\ b_M \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Furthermore, the above equation can be written as:

$$\mathbf{y} = \mathbf{X}(K_1, \dots, K_M) \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (6)$$

with,

$$\mathbf{y} = (y_1, y_2, \dots, y_n)'$$

$$\mathbf{X}(K_1, \dots, K_M) = \begin{bmatrix} t_1 & \dots & t_1^m & (t_1 - K_1)_+^m & \dots & (t_1 - K_M)_+^m \\ t_2 & \dots & t_2^m & (t_2 - K_1)_+^m & \dots & (t_2 - K_M)_+^m \\ \vdots & & \vdots & \vdots & & \vdots \\ t_n & \dots & t_n^m & (t_n - K_1)_+^m & \dots & (t_n - K_M)_+^m \end{bmatrix}$$

$$\boldsymbol{\beta} = (a_1, \dots, a_m, b_1, \dots, b_M)' \quad \text{and} \quad \boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$$

### 2.2 Local polynomial

Based on Eq. 2 where  $f(t_i)$  is a curve with unknown form, so that nonparametric estimation is made. One approach to estimate  $f(t_i)$  is Local Polynomial method. The Local Polynomial Estimator is obtained by Taylor series containing polynomials of  $p$  degree. If  $f(t_i)$  is dragged by Taylor series with the polynomial of  $p$  degree, then we obtain

$$f(t_i) \approx f(t) + (t_i - t) f^{(1)}(t) + \dots + (t_i - t)^p f^{(p)}(t) / p! \quad (7)$$

where  $t_i \in [t - h, t + h]$ .

If given  $\beta_r(t) = \frac{f^{(r)}(t)}{r!}$  with  $r = 0, 1, 2, \dots, p$  then Eq.7 can be written as

$$f(t_i) \approx \beta_0(t) + (t_i - t) \beta_1(t) + \dots + (t_i - t)^p \beta_p(t) \quad (8)$$

Eq. 8 can be written as follows:

$\mathbf{f}(t_i) \approx \mathbf{x}\boldsymbol{\beta}$  with

$$\mathbf{x} = \begin{bmatrix} 1 & (t_i - t) & (t_i - t)^2 & \dots & (t_i - t)^p \end{bmatrix}^T \quad \text{and}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0(t) & \beta_1(t) & \beta_2(t) & \dots & \beta_p(t) \end{bmatrix}^T$$

Obtaining  $\hat{\boldsymbol{\beta}}$  estimator is done by minimizing the Weighted Least Square (WLS) criteria as follows:

$$\sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 K_h(t_i - t) \quad (9)$$

$K_h(\cdot) = \frac{K(\cdot/h)}{h}$  where  $K$  is a kernel function, and  $h$  is bandwidth, so WLS criteria can be written as follows:

$$(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T \mathbf{K}_h (\mathbf{y} - \mathbf{x}\boldsymbol{\beta}) \quad (10)$$

$\mathbf{K}_h = \text{diag}(K_h(t_1 - t), \dots, K_h(t_n - t))$ ;  $\hat{\boldsymbol{\beta}}$  estimation is given by:  $\hat{\boldsymbol{\beta}} = (\mathbf{x}^T \mathbf{K}_h \mathbf{x})^{-1} \mathbf{x}^T \mathbf{K}_h \mathbf{y}$  [2].

### 2.3 Selection of smoothing parameter using Generalized Cross Validation (GCV)

The selection of the smoothing parameter on Spline Truncated modeling is by using GCV [7]. The smoothing parameter in Spline truncated model is knot point, whereas the smoothing parameter in a local polynomial is bandwidth. The selection of the optimal smoothing parameter is very important so that the estimation result is not over-smoothing or under-smoothing. Suppose an estimation curve with a local polynomial produces a function, in which any bandwidth  $h$  has an  $n \times n$  size matrix  $A$  such that  $\hat{\mathbf{y}} = \mathbf{A}\mathbf{y}$  then the value of  $GCV(h) = \frac{MSE(h)}{(n^{-1} \text{tr}[I-A])^2}$  with  $MSE(h) = (n)^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)$ . The smallest GCV value will provide an optimal bandwidth  $h$ . Likewise, if at any point of the knot point  $K$  on the estimation curve by spline there is an  $n \times n$  size matrix  $A$  such that  $\hat{\mathbf{y}} = \mathbf{A}\mathbf{y}$  then it is obtained  $GCV(K) = \frac{MSE(K)}{(n^{-1} \text{tr}[I-A])^2}$  with  $MSE(K) = (n)^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)$ .

## 3. RESULTS AND DISCUSSION

The Scatter plots of the seven categories of inflation expenditures include foodstuffs sector; food, beverages, cigarettes and tobacco sector; Housing, Water, Electricity, Gas And Fuel Sector; Clothing Sector; Health Sector, Education, Recreation and Sports Sector; Transportation, Communications and Financial Services Sector from January 2007 to August 2017. Based on the figure in Appendix 1, it appears that inflation in the foodstuff sector reached the highest compared to other sectors. Inflation rate movements from various sectors show the value fluctuate from month to month.

While in the food, beverages, cigarettes and tobacco sector ( $y_2$ ); Housing, Water, Electricity, Gas, and Fuel ( $y_3$ ) and clothing ( $y_4$ ) tend to be stationary. However, the foodstuff sector ( $y_1$ ) tends to have a fluctuating value. The period from 2010 to 2011 tend to be higher, and also 2013 and 2014 tend to be higher than in other periods. It is also seen that the inflation in Transport, Communication, and Financial Services sector ( $y_7$ ) tends to be higher since 2013, while from 2009 to 2010 it is lower than other inflation sectors. Inflation in the Transportation, Communication and Financial Services sector ( $y_7$ ) tend to be more volatile than the Health sector ( $y_5$ ); Education, recreation and sports ( $y_6$ ). The Health Sector ( $y_5$ ) and Education, Recreation and Sports ( $y_6$ ) in 2009 to 2016 tend to be stationary in value compared to Transport, Communication, and Financial Services sector ( $y_7$ ). Foodstuff ( $y_1$ ) is more volatile and higher than other sectors. The following descriptive statistics are presented for all variables.

Based on Table 1. it is seen that the average inflation rate for the foodstuff sector is also higher than in other sectors. While for the transportation, communication and financial services sector have the lowest average value of inflation but has the highest standard deviation among other sectors. It shows that the variance or diversity in transportation, communication, and financial services sector is the highest.

**Table 1.** Descriptive Statistics for seven groups of inflation expenditures from January 2007-August 2017

Sectors	SE				
	Mean	Mean	Stdev	Min	Max
Foodstuff	8.81	0.382	4.317	1.452	20.019
Food, Beverages, Cigarettes, and Tobacco	7.071	0.165	1.862	4.25	12.931
Housing, Water, Electricity, and Gas	4.933	0.219	2.474	1.183	12.398
Clothing	5.18	0.253	2.858	-0.413	12.267
Health	4.59	0.151	1.707	2.187	9.691
Education, Recreation and Sports	4.87	0.159	1.799	2.747	10.408
Trasportation, Communication and Financial Services	3.715	0.461	5.215	-6.852	16.333

To get the estimation of spline truncated regression model, the following steps are needed:

1. Determining the spline order, i.e. linear, quadratic or cubic
2. Choosing the optimum knots with GCV method
3. Conduct an estimation of model parameters

The optimal estimation of spline truncated model depends on the knot point optimal selection, where the optimal point determination is based on the smallest GCV value. In this estimation process, the GCV values calculated on the basis of one knot ( $K = 1$ ), two knots ( $K = 2$ ) and three knots ( $K = 3$ ) are performed on linear spline ( $m = 1$ ), quadratic ( $m = 2$ ) and cubic ( $m = 3$ ). The GCV value is obtained by trial and error for each possible value. Below is presented the optimal GCV value for each sector with a certain order.

Based on the Appendix 2, the smallest GCV value for foodstuff sector is linear with three-knot points, i.e. at  $k_1=18$ ,  $k_2=36$ , and  $k_3=43$  with GCV 8.701. In the food, cigarettes and tobacco sector, the smallest GCV value is 0.434 with  $k_1=23$ ,  $k_2=56$ , and  $k_3=102$ . In the Housing, Electricity and Water sector, the smallest GCV value is 1,967 with  $k_1=23$ ,  $k_2=34$ , and  $k_3=96$  for linear order. In Clothing sector inflation, the smallest GCV value is 3,835 for  $k_1=17$ ,  $k_2=38$ , and  $k_3=56$  for linear order. In health sector inflation, the smallest value of GCV is 0.582 for  $k_1=23$ ,  $k_2=39$ , and  $k_3=101$  for linear order. For Education, Recreation and Sports sector the smallest GCV value is 0.252 for  $k_1=32$ ,  $k_2=51$  and  $k_3=124$  for cubic order. For the inflation of transport, communications and financial services sector, the smallest GCV value is 15.999 with  $k_1=18$ ,  $k_2=34$ , and  $k_3=86$  for linear order. The estimation graph of spline truncated can be seen in Appendix 3. The result of model estimation for each sector is as follows:

Spline model in Foods Sector as follows as:

$$\hat{y}_i = 7.713 + 0.599t_i - 1.485(t_i - 18)_+ + 2.023(t_i - 36)_+ - 1.202(t_i - 43)_+$$

Spline model in fast foods, cigarette, and tobacco Sector as follow as:

$$\hat{y}_i = 4.584 + 0.395t_i - 0.498(t_i - 23)_+ + 0.28(t_i - 56)_+ - 0.24(t_i - 102)_+$$

Spline model in Housing, Electricity, and Water Sector as follow as:

$$\hat{y}_i = 4.584 + 0.395t_i - 0.498(t_i - 23)_+ + 0.28(t_i - 56)_+ - 0.24(t_i - 102)_+$$

The step in modeling using local polynomial is to determine the optimum bandwidth by the GCV method. The kernel function used in this study is the Gaussian kernel. Similar to determine the optimum knot point in Spline, determining the optimum bandwidth in a local polynomial is also through trial and error. For modeling using local polynomial, the number of order used is not restricted, as long as it is a positive integer. The following GCV values are presented based on the specified bandwidth-bandwidth.

**Table 3.** The value of GCV to determine the bandwidth of local polynomial

Foodstuff				Food, Cigarettes and Tobacco			
X <sub>m</sub>	h	order	GCV	X <sub>m</sub>	h	order	GCV
75	80	8	13.03064	65	100	8	0.8723
65	100	12	13.7409	85	90	12	1.0513
90	100*	18	7.6064	86	305	10*	0.7711
90	200	14	9.2377	80	250	12	0.9833
120	100	10	13.78	84	120	16	1.09061
Housing, Water, Electricity, and Gas				Clothing			
X <sub>m</sub>	h	order	GCV	X <sub>m</sub>	h	order	GCV
80	100	10	3.8893	95	110	10*	2.3077
75	120	6	3.25993	75	90	8	6.842
55	100*	10	2.3468	105	200	12	3.4297
95	120	10	9.09	90	200	10	4.536
95	110	14	11.189	94	200	14	179.93
Health				Education, Recreation, and Sports			
X <sub>m</sub>	h	order	GCV	X <sub>m</sub>	h	order	GCV
94	90	8	1.078	75	100	10	0.2626
95	90	12	1.2598	95	165	8	0.3576
65	120*	12	0.706	95	200	8	0.2741
65	125	8	1.0119	95	200	10	0.2737
75	200	10	0.76015	95	175	12*	0.2336
Transportation, Communication and Financial Services							
X <sub>m</sub>	h	order	GCV	X <sub>m</sub>	h	order	GCV
75	20	5	54.0034	50	200	10*	14.571
75	100	8	20.718	35	250	10	16.012

\* Optimal bandwidth

Estimation Model of local polynomial regression for inflation sector in Indonesia as follows as:

Local Polynomial model in Foods Sector as follows as:

$$\begin{aligned} \hat{y}_{1i} = & 7.95 - (3.77 \times 10^{-25})(t_i - 90) + (1.7 \times 10^{-21})(t_i - 90)^2 + (2.15 \times 10^{-20}) \\ & (t_i - 90)^3 - (1.33 \times 10^{-18})(t_i - 90)^4 - (2.93 \times 10^{-20})(t_i - 90)^5 - \\ & (2.47 \times 10^{-17})(t_i - 90)^6 + (3.42 \times 10^{-15})(t_i - 90)^7 - (3.55 \times 10^{-13})(t_i - 90)^8 \\ & - (5.38 \times 10^{-13})(t_i - 90)^9 - (1.31 \times 10^{-14})(t_i - 90)^{10} + (5.47 \times 10^{-16})(t_i - 90)^{11} \\ & + (1.96 \times 10^{-17})(t_i - 90)^{12} - (1.11 \times 10^{-20})(t_i - 90)^{13} - (7.01 \times 10^{-21})(t_i - 90)^{14} \\ & - (9.08 \times 10^{-23})(t_i - 90)^{15} - (2.26 \times 10^{-25})(t_i - 90)^{16} + (2.62 \times 10^{-27})(t_i - 90)^{17} \\ & + (1.36 \times 10^{-29})(t_i - 90)^{18} \end{aligned}$$

Local Polynomial model in fast foods, cigarette, and tobacco Sector as follow as:

$$\begin{aligned} \hat{y}_{2i} = & 7.22 + (1.95 \times 10^{-11})(t_i - 86) - (1.29 \times 10^{-8})(t_i - 86)^2 + (1.61 \times 10^{-4})(t_i - 86)^3 \\ & - (2.64 \times 10^{-6})(t_i - 86)^4 - (2.37 \times 10^{-7})(t_i - 86)^5 + (4.92 \times 10^{-10})(t_i - 86)^6 + \\ & (9.08 \times 10^{-11})(t_i - 86)^7 + (4.9 \times 10^{-13})(t_i - 86)^8 - (6.61 \times 10^{-15})(t_i - 86)^9 - \\ & (4.87 \times 10^{-17})(t_i - 86)^{10} \end{aligned}$$

Local Polynomial model in Housing, Electricity, and Water Sector as follow as:

$$\begin{aligned} \hat{y}_{3i} = & 2.27 - (6.25 \times 10^{-11})(t_i - 55) - (7.66 \times 10^{-10})(t_i - 55)^2 - (8.04 \times 10^{-5})(t_i - 55)^3 \\ & + (1.35 \times 10^{-5})(t_i - 55)^4 - (2.38 \times 10^{-9})(t_i - 55)^5 - (1.02 \times 10^{-8})(t_i - 55)^6 + \\ & (5.24 \times 10^{-11})(t_i - 55)^7 + (2.06 \times 10^{-12})(t_i - 55)^8 - (1.5 \times 10^{-14})(t_i - 55)^9 - \\ & (3.66 \times 10^{-17})(t_i - 55)^{10} \end{aligned}$$

Local Polynomial model in Clothes Sector as follows as:

$$\begin{aligned} \hat{y}_{4i} = & 2.84 + (1.29 \times 10^{-11})(t_i - 95) - (2.44 \times 10^{-9})(t_i - 95)^2 + (9.38 \times 10^{-5})(t_i - 95)^3 \\ & + (1.16 \times 10^{-5})(t_i - 95)^4 - (2.25 \times 10^{-7})(t_i - 95)^5 - (1.73 \times 10^{-8})(t_i - 95)^6 - \\ & (3.46 \times 10^{-11})(t_i - 95)^7 + (6.83 \times 10^{-12})(t_i - 95)^8 + (9.75 \times 10^{-14})(t_i - 95)^9 + \\ & (3.97 \times 10^{-16})(t_i - 95)^{10} \end{aligned}$$

Local Polynomial model in Health Sector as follows as:

$$\begin{aligned} \hat{y}_{5i} = & 2.66 - (4.11 \times 10^{-13})(t_i - 65) + (2.09 \times 10^{-11})(t_i - 65)^2 + (4.68 \times 10^{-9})(t_i - 65)^3 \\ & + (2.94 \times 10^{-8})(t_i - 65)^4 + (4.78 \times 10^{-8})(t_i - 65)^5 + (6.67 \times 10^{-9})(t_i - 65)^6 - \\ & (7.25 \times 10^{-11})(t_i - 65)^7 - (5.76 \times 10^{-12})(t_i - 65)^8 + (2.96 \times 10^{-14})(t_i - 65)^9 + \\ & (1.67 \times 10^{-15})(t_i - 65)^{10} - (3.64 \times 10^{-18})(t_i - 65)^{11} - (1.61 \times 10^{-19})(t_i - 65)^{12} \end{aligned}$$

Local Polynomial model in Education, Vacation, Sports Sector as follow as:

$$\hat{y}_{6t} = 4.26 - (1.46 \times 10^{-13})(t_i - 95) + (8.02 \times 10^{-11})(t_i - 95)^2 - (7.68 \times 10^{-6})(t_i - 95)^3 - (4.86 \times 10^{-7})(t_i - 95)^4 + (5.25 \times 10^{-8})(t_i - 95)^5 + (4.95 \times 10^{-10})(t_i - 95)^6 - (5.25 \times 10^{-8})(t_i - 95)^7 + (4.95 \times 10^{-10})(t_i - 95)^8 - (5.02 \times 10^{-11})(t_i - 95)^9 - (4 \times 10^{-13})(t_i - 95)^{10} + (1.15 \times 10^{-14})(t_i - 95)^{11} + (1.16 \times 10^{-16})(t_i - 95)^{12}$$

Local Polynomial model in Transportation, Communication and Financial Services Sectors as follow as:

$$\hat{y}_{7t} = -3.75 + (1.24 \times 10^{-10})(t_i - 50) + (8.89 \times 10^{-9})(t_i - 50)^2 + (2.95 \times 10^{-5})(t_i - 50)^3 + (2.34 \times 10^{-5})(t_i - 50)^4 - (5.25 \times 10^{-7})(t_i - 50)^5 - (1.8 \times 10^{-8})(t_i - 50)^6 + (3.83 \times 10^{-10})(t_i - 50)^7 + (2.52 \times 10^{-12})(t_i - 50)^8 - (8.87 \times 10^{-14})(t_i - 50)^9 + (4.65 \times 10^{-16})(t_i - 50)^{10}$$

Based on the model obtained, the estimated inflation rate using local polynomials results in different local point values in each sector of inflation. The degree of polynomial produced reaches 10 to 12 degree. In local polynomial modeling, the determination of local points and the polynomial degree greatly determines the accuracy of the model. The resulting model of the local polynomial is more complicated than the spline. However, the accuracy of modeling inflation sectors in Indonesia using local polynomials is better than spline truncated. The estimation graph of the local polynomial can be seen in Appendix 4. The table shows the R-square value for each sector as follows:

**Table 4.** R-square model of spline truncated and local polynomial

Sectors	R-Square	
	Spline Truncated	Local Polynomial
Foodstuff	56.6%	65.17%
Food, Beverages, Cigarettes, and Tobacco	88.4%	80.83%
Housing, Water, Electricity, and Gas	70.1%	69.14%
Clothing	56.3%	75.53%
Health	81.4%	81.49%
Education, Recreation, and Sports	84%	93.78%
Transportation, Communication and Financial Services	45.2%	50.19%
<b>Mean</b>	<b>68.86%</b>	<b>73.73%</b>

Based on the Table, it can be seen that the average R-square for the local polynomial model is higher than the spline truncated model. The R-square value in each sector has similar value characteristics, for example for the health sector reaching more than 80%, while for transportation, communication, and financial services sector does not reach 52%. The greatest R-square values in spline truncated models are in the Food, Beverages, Cigarettes and Tobacco sectors. While for a local polynomial is Education, Recreation and Sports. The lowest R-square values for spline truncated and local polynomial models are in the transportation, communication, and Financial Services sector. Therefore,

other nonparametric modeling methods are needed for modeling transportation, Communication, and Financial Services sector.

#### 4. CONCLUSION

Based on the results of the previous analysis, it can be concluded that spline truncated and local polynomial modeling requires a smoothing parameter. In addition to the smoothing parameters, the order of the model also determines the model match. In inflation modeling, in Indonesia, the local polynomial approach results in a better average R-square accuracy than the truncated spline. Meanwhile, modeling in the Transportation, Communication and Financial Service produces the lowest R-square value, so another nonparametric modeling is required for it.

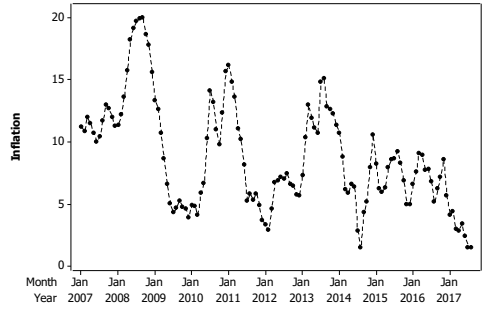
#### ACKNOWLEDGMENT

This research founded by Ministry of Reseach, Technology and Higher Education as “Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT)” in the First year. Thanks to Statistics Laboratory of Statistics Department, Diponegoro University, Semarang, Indonesia.

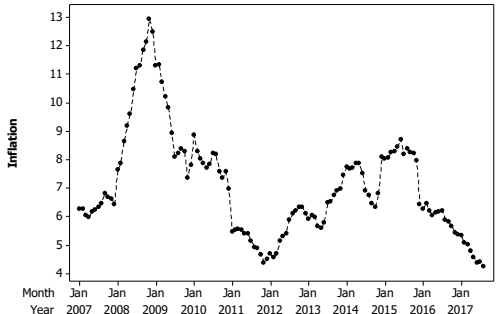
#### REFERENCES

- [1] Delaigle A, Fan J, Carroll RJ. (2009). A design-adaptive local polynomial estimator for the errors in variable problem. *Journal of the American Association* 104 (485): 348-359. <http://dx.doi.org/10.1198/jasa.2009.0114>
- [2] Welsh AH, Yee TY. (2005). Local regression for vector responses. *Journal of Statistical Planning and Inference* 136(9): 3007-3031. <http://dx.doi.org/10.1016/j.jspi.2004.01.024>
- [3] Xue L. (2010). Empirical likelihood local polynomial regression analysis of clustered data. *Scandinavian Journal of Statistics: Theory and Applications* 37(4): 644-663. <https://doi.org/10.1111/j.1467-9469.2009.00677.x>
- [4] Dette H, Melas VB. (2010). A note on all-bias designs with applications in spline regression models. *Journal of Statistical Planning and Inference* 140(7): 2037-2045. <http://dx.doi.org/10.1016/j.jspi.2010.01.047>
- [5] Huang JZ. (2003). Asymptotics for polynomial spline regression under weak conditions. *Statistics & Probability Letter* 65(3): 207-216. <http://dx.doi.org/10.1016/j.spl.2003.09.003>
- [6] Bryson KMO, Ko M. (2004). Exploring the relationship between information technology investment and firm performance using regression spline analysis. *Information and Management* 42(1): 1-13. <https://doi.org/10.1016/j.im.2003.09.002>
- [7] Takezawa K. (2006). *Introduction to Nonparametric Regression*. John Wiley & Sons, New Jersey.
- [8] Fransisco MF, Opsomer FJ, Fernandez MV. (2004). Plug in bandwidth selector for local polynomial regression estimator with correlated error. *Nonparametric Statistics* 16(1-2): 127-151. <https://doi.org/10.1080/10485250310001622848>
- [9] Xiao Z, Linton OB, Carroll RJ, Mammen E. (2003).

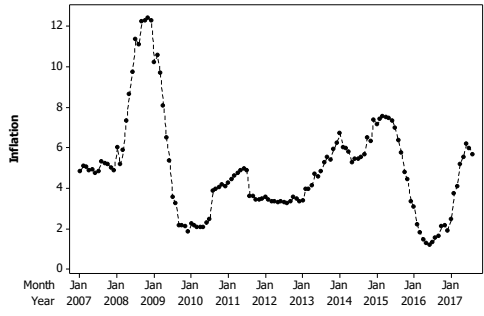
**Appendix 1. Scatter Plots of the Data**



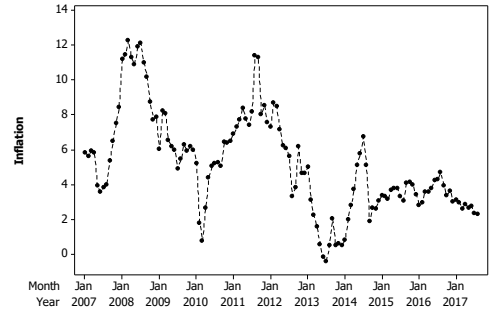
(1)



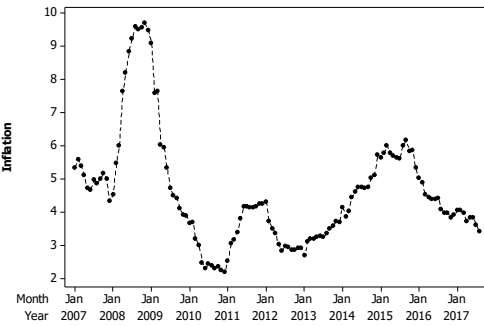
(2)



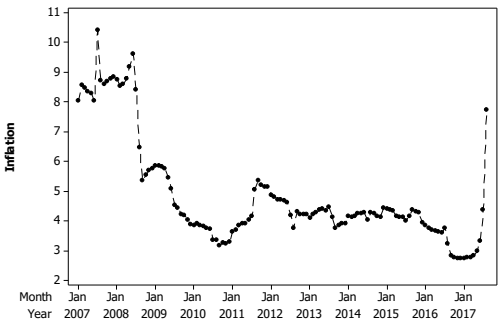
(3)



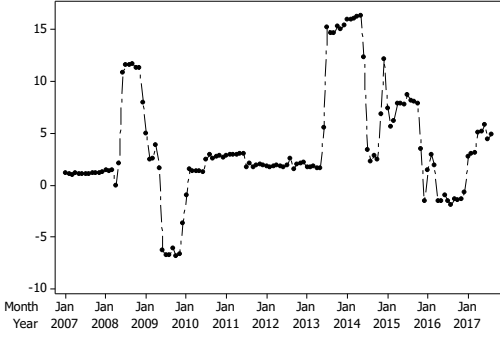
(4)



(5)



(6)



(7)

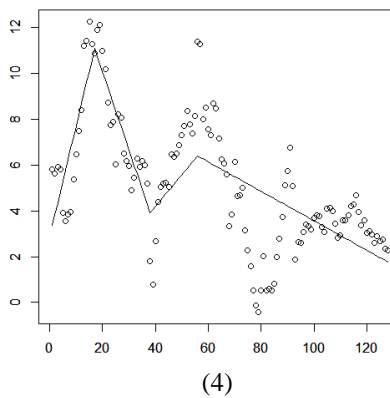
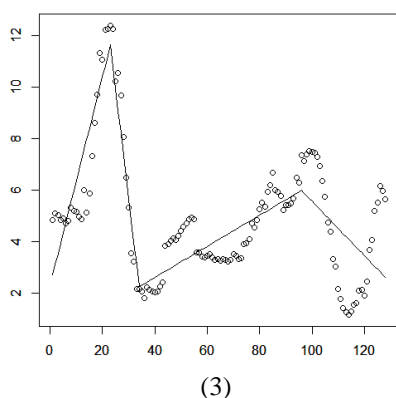
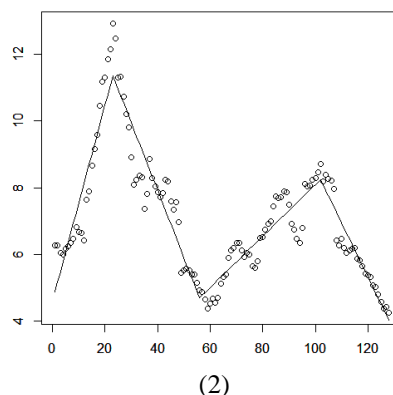
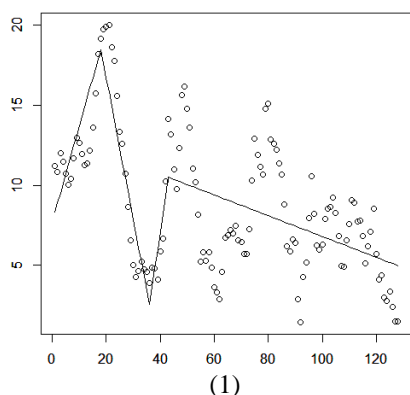
Note:  
 (1): Foodstuff Sector Inflation  
 (2): Food, Cigarettes and Tobacco Sector Inflation  
 (3): Housing, Water, Electricity and Gas Sector Inflation  
 (4): Clothing Sector Inflation  
 (5): Health Sector Inflation  
 (6): Education Sector Inflation  
 (7): Transportation, Communications and Financial Services Sector Inflation

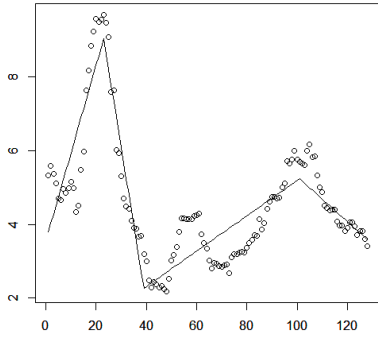
Appendix 2. Value of GCV to determine Knot of Spline Truncated

Sector	Order	1 Knot		2-knot points			3-knot points			
		Knot	GCV	Knot 1	Knot 2	GCV	Knot 1	Knot 2	Knot 3	GCV
Foodstuff	Linear	35	13.754	18	30	10.591	18	36	43	<b>8.701*</b>
	Quadratic	103	13.587	30	39	10.709	13	26	32	8.891
	Cubic	30	12.183	40	59	12.545	22	29	40	9.310
Food, Cigarettes and Tobacco	Linear	20	2.486	22	59	1.367	23	56	102	<b>0.434*</b>
	Quadratic	26	2.173	30	99	1.077	15	39	66	0.725
	Cubic	43	0.998	20	39	0.557	8	47	67	0.723
Housing, Water, Electricity and Gas	Linear	43	5.507	23	34	2.711	23	34	96	<b>1.967*</b>
	Quadratic	23	5.399	23	56	3.924	35	49	80	3.658
	Cubic	34	4.226	38	86	3.589	16	39	105	2.479
Clothing	Linear	15	4.854	16	34	4.202	17	38	56	<b>3.835*</b>
	Quadratic	30	4.980	25	66	4.860	23	39	80	4.148
	Cubic	30	4.894	30	59	4.663	33	59	100	3.928
Health	Linear	49	2.096	22	43	0.912	23	39	101	<b>0.582*</b>
	Quadratic	23	2.065	40	89	1.435	32	49	100	0.915
	Cubic	40	1.162	11	49	1.189	13	32	106	0.789
Education, Recreation and Sports	Linear	36	0.625	33	62	0.641	13	32	126	0.383
	Quadratic	55	0.718	43	127	0.630	25	36	60	0.488
	Cubic	18	0.675	20	122	0.372	32	51	124	<b>0.252*</b>
Trasportation, Communication and Financial Services	Linear	89	23.119	43	87	19.580	18	34	86	<b>15.999*</b>
	Quadratic	71	22.081	23	66	20.461	13	30	103	16.141
	Cubic	30	22.127	23	106	18.528	13	30	103	16.141

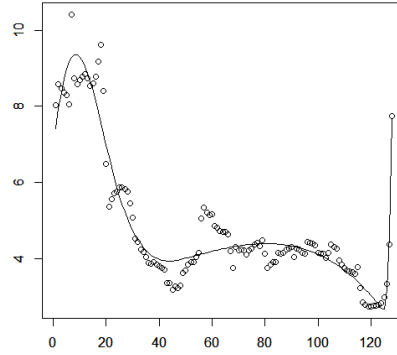
\*) The Optimal Knot

Appendix 3. Scatter Plots of Estimation of Spline Truncated

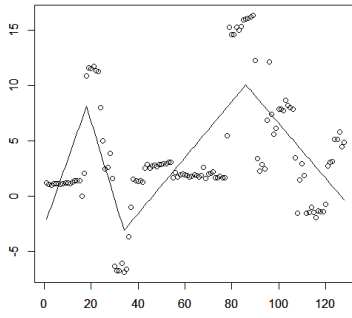




(5)



(6)

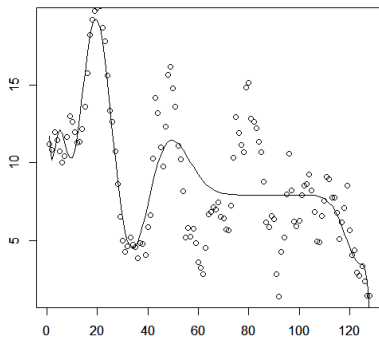


(7)

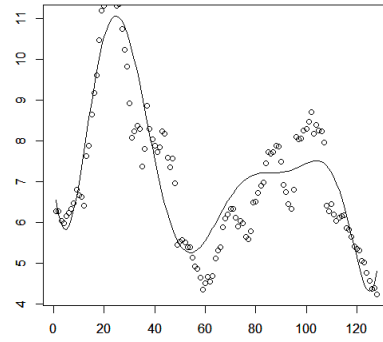
Note:

- (1): Foodstuff Sector Inflation
- (2): Food, Cigarettes and Tobacco Sector Inflation
- (3): Housing, Water, Electricity and Gas Sector Inflation
- (4): Clothing Sector Inflation
- (5): Health Sector Inflation
- (6): Education Sector Inflation
- (7): Transportation, Communications and Financial Services Sector Inflation

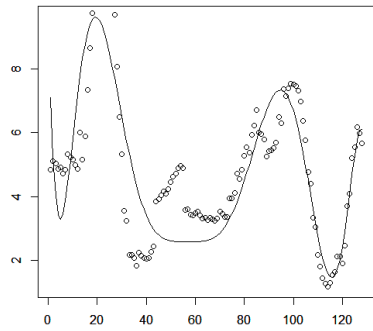
**Appendix 4. Scatter Plots of Estimation of Local Polynomial**



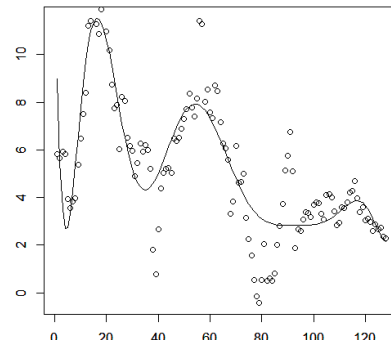
(1)



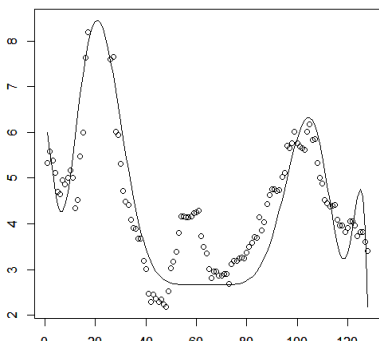
(2)



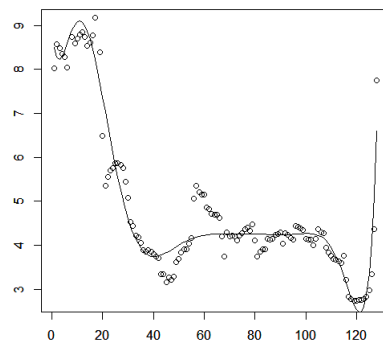
(3)



(4)

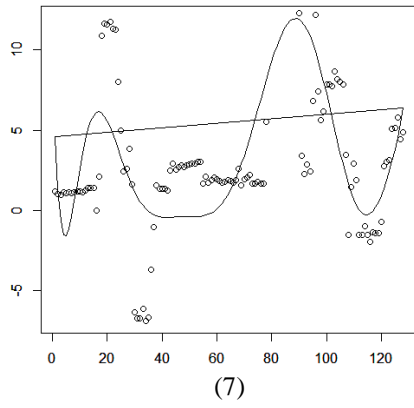


(5)



(6)





Note:

- (1): Foodstuff Sector Inflation
- (2): Food, Cigarettes and Tobacco Sector Inflation
- (3): Housing, Water, Electricity and Gas Sector Inflation
- (4): Clothing Sector Inflation
- (5): Health Sector Inflation
- (6): Education Sector Inflation
- (7): Transportation, Communications and Financial Services Sector Inflation