

# Spline-Fourier's Method for Modelling Inflation in Indonesia

Suparti Suparti<sup>1,\*</sup>, Alan Prahutama<sup>1</sup>, Rukun Santoso<sup>1</sup> and Alvita Rachma Devi<sup>1</sup>

<sup>1</sup>Department of Statistics, Faculty of Science and Mathematics, Diponegoro University, Semarang - Indonesia

**Abstract.** Regression method is a statistical method for modelling dependent variable with independent variable. Nonparametric regression is an approach to regression analysis that is suitable for data that have an unknown curve shape. Modelling by using nonparametric regression method does not require any assumptions. Spline and Fourier methods are some of the estimators in nonparametric regression. The spline method requires optimal knots to obtain the best model. The most commonly used method to determine the optimal knots is Generalized Cross Validation (GCV). The Fourier method is a method based on the cosine and sinus series. The Fourier method is particularly suitable for data that experience repetitive patterns. This study modeled the Inflation rate in Indonesia from January 2007 to August 2017. The dependent variable is inflation rate, while the independent variable is time. From the result, linear spline regression estimation with three knots that generates R square of 60%. The best Fourier model is Fourier with  $K = 100$  that generates R square of 80.12%. The best Spline model is with 9 knots generates R square of 87.65%, so, for inflation modelling in Indonesia, the spline regression model generates a simpler model with better R-square than Fourier regression.

Keywords: **Spline; Fourier; GCV; Inflation.**

## 1 Introduction

Regression is a statistical method that is used to model the relationship between the dependent variable with the independent variable. Regression modelling can be done with three approaches, that is parametric, nonparametric and semiparametric. The parametric regression is a regression model with a known curve shape. The parametric regression can be done as simple regression or multiple regression. This regression is very strict with the assumption, such as normally distributed residuals, homocedasticity and non autocorrelation. Therefore, this regression is very easy to use, but strict with assumption. Nonparametric regression is a regression model with unknown curve shape. Meanwhile, semiparametric regression is a regression model with one or more known curves pattern, while the other curves pattern is unknown. The pattern intended here is the pattern between the dependent variable with the independent variable. So, it can be said that semiparametric regression model is a combination of parametric and nonparametric regression [1].

Some methods that are associated with nonparametric regression models are Spline regression and Fourier regression. The spline regression is a regression model based on the Spline function [2]. The Spline function is a segmented polynomial function. In spline regression modelling, the best model is determined by splines orde and knots. There are some commonly used orde in spline, such as linear, quadratic,

cubic or more. As for the determination of knot is by using Generalized Cross Validation (GCV) method.

Fourier regression is a regression model with Fourier series approach, which is a series of sine and cosine functions. The Fourier series fits very well with the pattern of data that undergoes repeated cycles. Unlike spline regression, the determination the best Fourier model is based only on the number of  $K$  [3].  $K$  is the number of sine or cosine functions that are formed.

Inflation data is time series data that is observed for several time periods. Inflation data form a fluctuating pattern over time. Modelling of inflation data by using regression model is very useful to know the movement pattern of inflation rate. Therefore, this research will model inflation data with spline regression and Fourier regression approach

## 2 Literature review

### 2.1 Nonparametric regression

Regression model is a model that explains the relationship between independent variables with dependent variable. Simple linear regression and multiple linear regression are two kind of parametric regression. Simple linear regression is a form of parametric regression with one independent variable, while multiple linear regression is a form of parametric

\* Corresponding author: [suparti702@gmail.com](mailto:suparti702@gmail.com)

regression with more than one independent variables. Regression model in general is as follows:

$$y_i = \beta_0 + \beta_1 t_{1i} + \beta_2 t_{2i} + \dots + \beta_p t_{pi} + \varepsilon_i \quad (1)$$

with  $y_i$  is a dependent variable while  $t$  is independent variable. Meanwhile, nonparametric regression is a regression model with unknown curve shape. In general, equation 1 can be changed into:

$$y_i = \eta(t_i) + \varepsilon_i \quad (2)$$

$\eta(t_i)$  is an unknown curve of regression function while  $t_i$  is an independent variable. The function curve  $\eta(t_i)$  is assumed to be smooth in a given function space [1].

## 2.2 Fourier regression model

Fourier Regression one of the approaches in nonparametric regression based on Fourier series, that a series of sine or cosine functions. Fourier regression is particularly suitable for data that has repetitive patterns in a given period. The Fourier series functions was expressed in equation 3 [3].

$$f(t) = \frac{1}{2} \alpha_0 + \gamma t + \sum_{k=1}^K \alpha_k \cos kt \quad (3)$$

The form of equation (1) can be expressed in matrix:

$$\mathbf{y} = \mathbf{t}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (4)$$

if

$$\mathbf{y} = \mathbf{f}(\mathbf{t}) + \boldsymbol{\varepsilon} \quad (5)$$

then

$$\mathbf{f}(\mathbf{t}) = [f(t_1) f(t_2) \dots f(t_n)]^T \quad (6)$$

with  $f(t)$  is an unknown curve shape.  $f(t)$  is approached by using Fourier series as follows [4]:

$$f(t_i) = \frac{1}{2} \alpha_0 + \gamma t_i + (\alpha_{i1} \cos t_i + \alpha_{i2} \cos 2t_i + \dots + \alpha_{iK} \cos Kt_i) \quad (7)$$

If

$$\mathbf{f}(\mathbf{t}) = \mathbf{A}\boldsymbol{\theta} \quad (8)$$

then

$$\mathbf{A} = \begin{bmatrix} 1 & t_1 & \cos t_1 & \cos 2t_1 & \dots & \cos Kt_1 \\ 1 & t_2 & \cos t_2 & \cos 2t_2 & \dots & \cos Kt_2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & \cos t_n & \cos 2t_n & \dots & \cos Kt_n \end{bmatrix} \quad (9)$$

$$\boldsymbol{\theta} = [\phi \quad \gamma \quad \alpha_{11} \quad \alpha_{12} \quad \dots \quad \alpha_{1K}]^T \quad (10)$$

with

$$\phi = \frac{n}{2} \alpha_0 \quad (11)$$

The estimation of nonparametric regression model is by using OLS (Ordinary Least Square) as follows:

$$\begin{aligned} \boldsymbol{\psi}(\boldsymbol{\theta}) &= \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \\ \hat{\boldsymbol{\theta}} &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \end{aligned} \quad (12)$$

Equation 12 is a nonparametric regression estimator by using the Fourier series approach [5].

## 2.3 Spline regression

Spline regression is a nonparametric regression approach based on spline function. The spline function is a segmented polynomial, so it has a high flexibility. In general, the spline regression model can be expressed in equation 13:

$$y_i = S(t) + \varepsilon_i$$

$$S(t) = \sum_{p=1}^m \beta_p t^p + \sum_{j=1}^r \beta_{j+m} (t - k_j)_+^m \quad (13)$$

with

$$(t - k_j)_+^m = \begin{cases} (t - k_j)^m & , t \geq k_j \\ 0 & , t < k_j \end{cases} \quad (14)$$

where  $\beta$  is constants and  $k_1, k_2, \dots, k_r$  are knots,  $m$  is the orde of the spline regression model and  $r$  is the number of knos. Equation (13) can be expressed in matrix form as follows [6]:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & \dots & t_1^m & (t_1 - k_1)^m & \dots & (t_1 - k_r)^m \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \dots & t_n^m & (t_n - k_1)^m & \dots & (t_n - k_r)^m \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{m-1} \\ \beta_m \\ \vdots \\ \beta_r \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

or

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (15)$$

Estimation of spline model use *Ordinary Least Square* (WLS) generate this following estimation:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \quad (16)$$

So, the estimation model is:

$$\hat{\mathbf{y}} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

or

$$\hat{\mathbf{y}} = \mathbf{A}(k) \mathbf{y} \quad (17)$$

There are several methods to select optimal knot and the most commonly used method is Generalized Cross Validation (GCV). GCV function is defined as:

$$GCV(k) = MSE(k) / (n^{-1} \text{tr}[\mathbf{I} - \mathbf{A}(k)])^2, \quad (18)$$

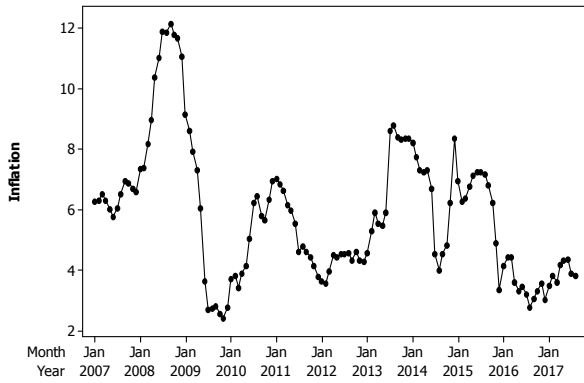
where

$$MSE(k) = n^{-1} \sum_{i=1}^n \{y_i - \hat{f}(t_i)\}^2; \quad (19)$$

$\mathbf{A}(k)$  is the projection matrix at the  $k$ -th knot;  $y_i$  is the  $i$ -th response variable.  $\mathbf{I}$  is the matrix identity. The selection of optimal knot is by the minimum value of GCV.

## 3 Results and discussion

The data used in this study is year on year inflation data of Indonesia. This study used inflation data from January 2007 to August 2017. Figure 1 shows the scatterplot of inflation data in Indonesia.



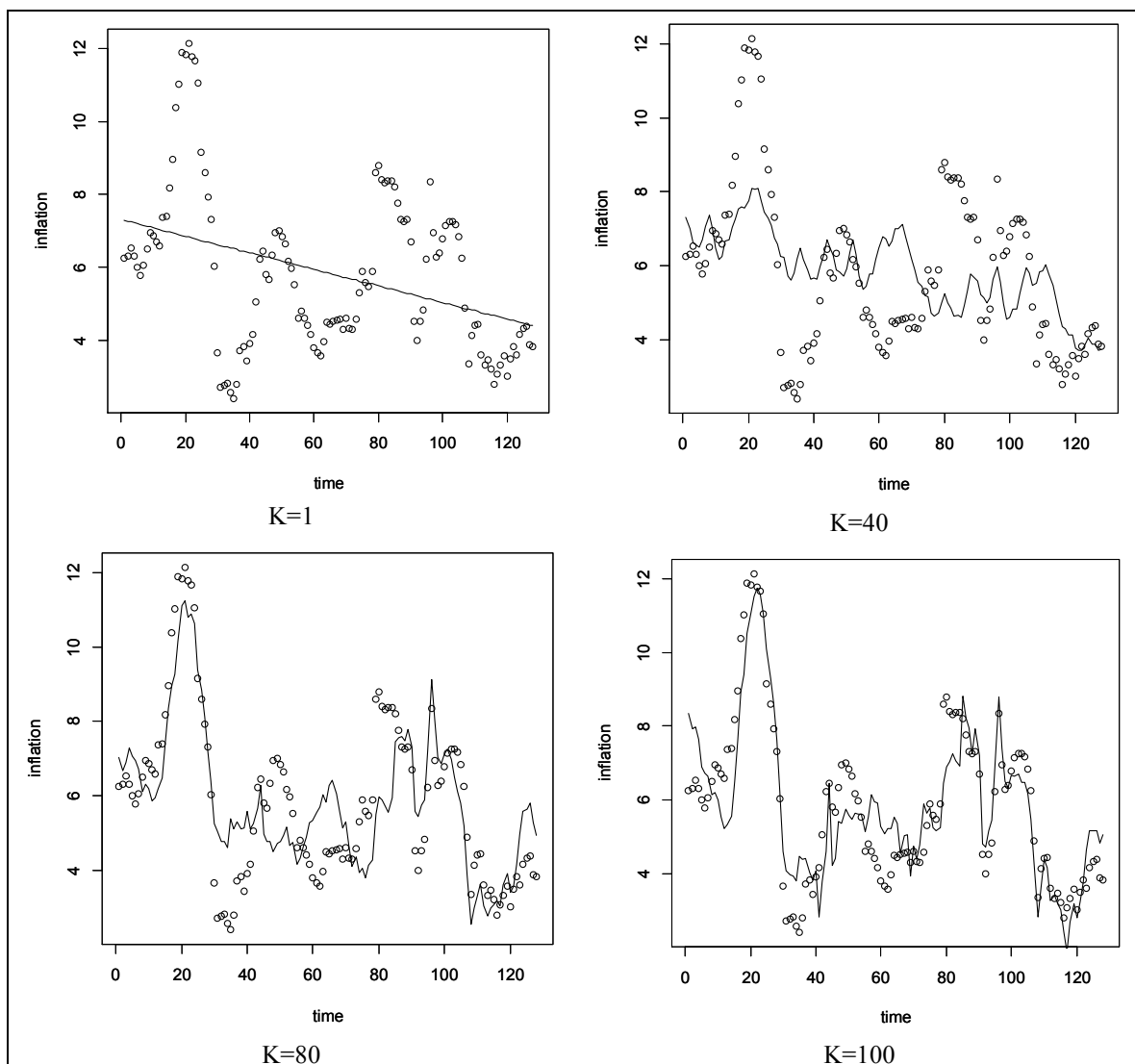
**Fig. 1.** Scatterplot data

Based on Figure 1, it shows that the pattern of data fluctuate up and down. The highest inflation rate reached 12.14, occurred in September 2008. Meanwhile the

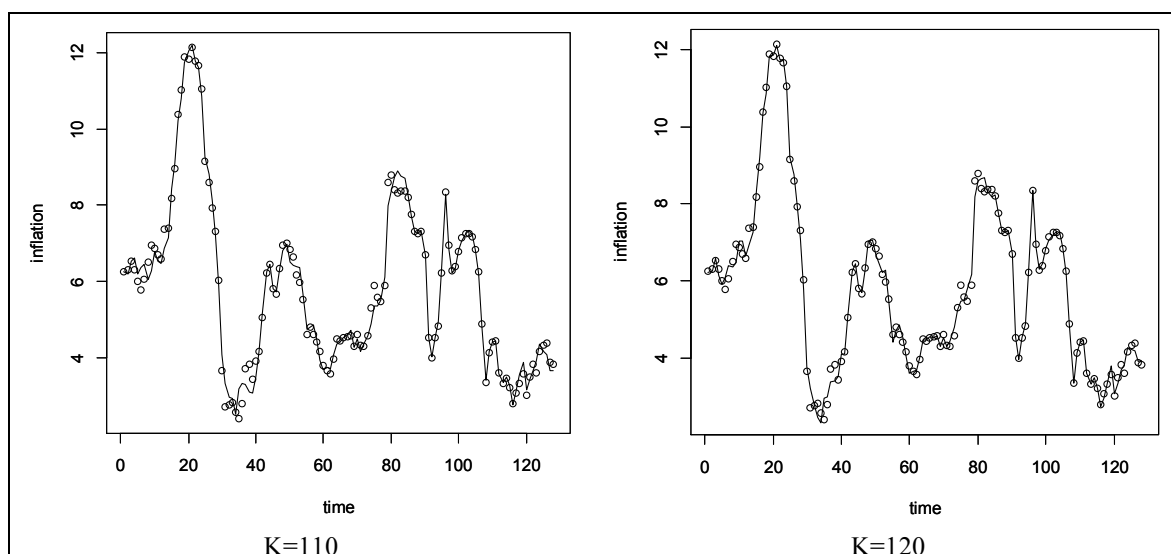
minimum inflation rate reached 2.41 occurred in November 2009. The volatile inflation rate cause economic instability. The average of inflation rate reached 5.842 with the standard deviation of 2.232.

### 3.1 Modelling inflation in Indonesia by using fourier regression

The next step is modelling inflation rate data in Indonesia by using Fourier series. The dependent variable is the inflation rate, while the independent variable is time. In Fourier regression modelling, determining the optimal K is very important. K is an integer. The more K values indicates the more model parameters in the Fourier regression. Figure 2 shows the result of parameter estimation using Fourier series. It can be concluded that if a model has many K values, the Fourier regression models become more complex. The



**Fig. 2.** Scatterplot of Fourier series estimation for various number of K



**Fig. 2.** Scatterplot of Fourier series estimation for various number of K (continuation)

more K values show the more parameter that must be estimated. Basically, the best model is a parsimony Fourier regression models and produce a good accuracy value. Parsimony model means a simple model, not complex. This study used accuracy values of  $R^2$  and MSE.

Based on Figure 2, the estimation value that is close to the actual data is  $K = 120$ , but the required parameter is 121. As for the number of  $K = 1$  does not produce a good model estimation. It has been experimented for the number of  $K$  ranging from 2, 4, and 5 but the estimation is not as good as  $K = 1$ .

**Table 1.**  $R^2$  and MSE value of Fouries regression model for various K

Number of K	$R^2$ (%)	MSE
K=1	14.308	4.2375
K=5	14.37	4.2344
K=10	17.738	4.0678
K=40	32.518	3.3369
K=60	40.58	2.9378
K=70	66.01	1.678
K=80	66.41	1.6607
K=90	69.85	1.49
K=100	80.12	0.98307
K=110	98.78	0.06
K=120	99.4	0.0295

Based on Table 1, for  $K = 1$  to  $K = 60$ , the  $R^2$  value is not very good. For  $K = 70$ , the  $R^2$  value is above 50%, but the increase of  $R^2$  value for  $K = 70$  to  $K = 80$  is not very significant. For  $K = 90$  to  $K = 100$ , there is a significant increase in  $R^2$  value. For the following example, take  $K = 5$ . The estimation Fourier regression model is:

$$\hat{f}(t) = \frac{1}{2}(7.3) - 0.023t - 0.0117 \cos t - 0.01 \cos 2t + 0.035 \cos 3t - 0.013 \cos 4t - 0.068 \cos 5t$$

The next step is modeling inflation rate data by using spline truncated regression.

### 3.2 Modelling inflation in Indonesia by using spline truncated

In spline truncated modelling, the most important step is to determine the optimal knot. The method for selecting optimal knot in this study is GCV method. This study limits the selection of optimal knot only up to 3 knots. The orde used in this study is linear, quadratic and cubic. The GCV values for various knots and orde were presented in Table 2.

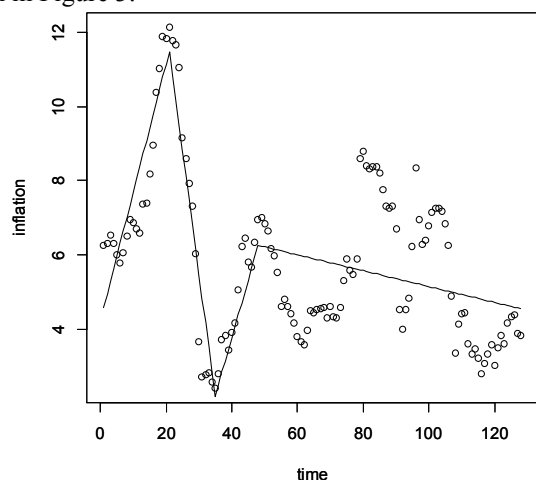
**Table 2.** GCV values for various knots and orde

Number of knot	Orde	Knot	GCV
1	Linear	102	4.24538
	Quadratic	78	3.9104
	Cubic	31	3.111178
2	Linear	30; 23	2.4837
	Quadratic	33; 34	2.35471
	Cubic	21; 22	2.71633
3	Linear	21; 35; 84	1.81878
	Quadratic	24; 69; 109	2.671
	Cubic	20; 39; 70	2.9365

Based on Table 2, the best spline model is linear with three knots. The following is the result of parameter estimation.

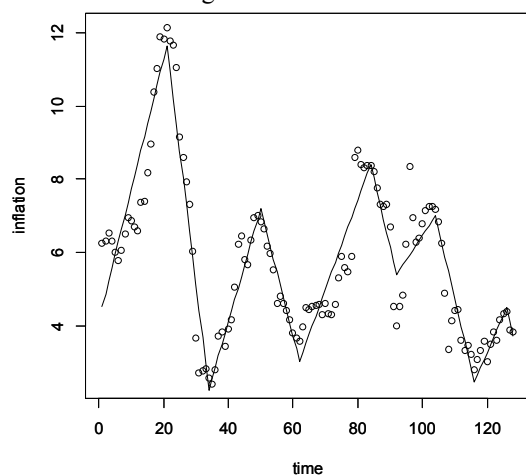
$$\hat{f}(t) = 4.242 + 0.345t_i - 1.009(t_i - 21)_+ + 0.978(t_i - 35)_+ - 0.335(t_i - 48)_+$$

The linear spline model with three knot generated  $R^2$  value of 60% and MSE of 1.9684. The scatterplot of linear spline model estimation with three knots can be seen in Figure 3.



**Fig. 3.** Linear spline regression estimation with three knot.

Basically, to determine the optimal knot on the spline regression, it will be very easy if the orde is linear. We can see the changes of the data on the scatterplot. Based on Figure 1, the value that experience change is on the 21-st, 34-th, 50-th, 62-nd, 80-th, 92-nd, 104-th, 116-th, and 126-th observation. So it can be determined that it has nine knots. The estimation results by using nine knots can be seen in Figure 4.



**Fig. 4.** Linear spline regression estimation with nine knots

The spline regression model for inflation rate in Indonesia with nine knot is:

$$\hat{f}(t) = 4.153 + 0.356t_i - 1.08(t_i - 21)_+ + 1.034(t_i - 34)_+ - 0.657(t_i - 50)_+ + 0.593(t_i - 62)_+ - 0.624(t_i - 84)_+ + 0.514(t_i - 92)_+ - 0.515(t_i - 104)_+ + 0.584(t_i - 116)_+ - 0.592(t_i - 126)_+$$

The  $R^2$  value for that model is 87.65% and the MSE value is 0.6151.

## 4 Conclusion

Inflation modelling in Indonesia by using Fourier regression resulted in a more complex model estimation compared to spline regression. Fourier regression modelling requires  $K = 100$  to obtain an  $R^2$  value of 80.12%. It means Fourier regression requires 101 parameters to model. While spline regression requires only 11 parameters to obtain the inflation model with  $R^2$  of 87.65%. The linear spline model with three knots yields  $R^2$  of 60%. Fourier regression models need to be studied in the recurring determination so that the optimal Fourier regression model will be obtained. Model Fourier can be applied to model CO, NO and another ambient air. So that it is useful to create low carbon city as a Sustainable city.

The authors would like to thank Ministry of Research, Technology and Higher Education for funding this research with PTUPT (applied research college) scheme for 2018 with contract No:101-124/UN7.P4.3/PP/2018.

## References

1. W. Hardle, *Applied Nonparametric Regression*, Cambridge University, (1990).
2. R.L., Eubank, C. Huang, Y.M. Maldonado, N. Wang, S. Wang, R.J. Buchanan, *J. R. Statist. Soc.*, **66**, Part 3, pp. 653-667, (2004).
3. M. Bilodeau, *The Canadian Journal of Statistics*, Vol. **3**, pp: 257-259, (1992).
4. A. Prahutama, Suparti, T.W. Utami, *IOP Journal of Physics Conference Series*, **974**, (2018).
5. A. Tripena, I.N. Budiantara, *International Conference on Natural Sciences and Applied Natural Sciences*, Ahmad Dahlan University, Yogyakarta, (2007).
6. B. Lestari, I.N. Budiantara, S. Sunaryo, and M. Mashuri, *Journal of Mathematics and Statistics*, Vol. **6**, No. 3, hal. 327-332, (2010)