MODELLING JAKARTA COMPOSITE INDEKS USING SPLINE TRUNCATED

Alan Prahutama[1], Suparti[2], Sugito[3] dan Tiani Wahyu Utami[4]

[1,2,3] Statistics Department, Diponegoro University, Semarang, Indonesia
[4]Statistics Department, University of Muhammadiyah Semarang, Semarang, Indonesia
[1]alan.prahutama@gmail.com, [2]suparti702@gmail.com, [3]sugitostat@gmail.com, [4]tiani.utami@gmail.com,

Abstract

Regression analysis can be done by parametric and nonparametric approach. The nonparametric approach does not assume an assumption compared to parametric. One nonparametric approach is the spline truncated. Spline is a polynomial piece that provides high flexibility. Spline modeling requires spline and knots. To determine the knots using General Cross Validation (GCV). In this study modeled the value of Jakarta Composite Index (JCI). JCI provides benefits to know the overall stock price in the stock exchange Indonesia. In this study the best spline model is linear with three knots with R square is 94.34%.

Keywords: Jakarta Composite's Index, Spline truncated, GCV.

Introduction

Regression is one of the statistical methods to model the relationship between response variables and predictor variables. The parametric regression approach is easy, but very strict with assumptions. In contrast to parametric approach, nonparametric approach is complex to do, but do not require assumptions. Parametric modeling is done when the data pattern is known, while nonparametric approach can be done if the data pattern is unknown. Some nonparametric regression modeling methods have been widely used, among others, using spline truncated, local polynomial, Fourier series, Wavelet, Kernel and others. Spline regression truncated is a segmented regression model-segment in the form of piecewise polynomial. This segmented nature provides spline benefits compared to other methods. Spline truncated modeling procedures include determining the spline order and selecting the optimum knot point. Determination of optimum knot point using CV (Cross Validation) and GCV (General Cross Validation).
Composite Stock Price Index (CSPI) is an indicator that presents the market price in Indonesia Stock Exchange. IHSG values tend to fluctuate and have high volatility over time. In this research will be modeled the value of IHSG using Spline truncated.

Method

2.1. Nonparametric regression

Nonparametric regression is one of the approaches used to find out the relationship pattern between explanatory variable and unrecognized response of regression curve. In general, nonparametric regression has the following function form:

$$y_i = f(t_i) + \varepsilon_i \ , \quad \mathrm{i} = 1, 2, \cdots, \mathrm{n}$$

with y_i is response variable and $f(t_i)$ is curve of regression with t_i is prediktor variable and ε_i is residual of model (Wahba, 1990).

2.1. Spline Regression

Spline is a segment of a segmented polynomial (piecewise polynomial) that has flexibility properties. The nature of flexibility is what distinguishes the spline with the polynomial. The joint fusion point of the pieces or points indicating changes in the behavior of the curve at different intervals.
In general the spline function of the order is any function that can be written in form (Eubank, 1988):

$$f(t) = \sum_{j=1}^{m} \alpha_j t^j + \sum_{k=1}^{M} \beta_k \left(t - K_k\right)_+^m$$

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

α_j and β_k are parameter of spline regression then $K_1, K_2, \cdots, K_M$ are knot.

If $m=1$ and the number of knot is one then the form of spline model is spline linear as follow as

$$f(t)=\alpha_1 t+\beta_2 (t-K)_+^1$$

It is called spline linear with one knot for t = K

It can be written as:

$$f(t)=\begin{cases}\alpha_1 t & ;\ t<K \\ \alpha_1 t+\beta_2 (t-K) & ;\ t\geq K\end{cases}$$

Spline regression model can be written as:

$$y_i=\sum_{j=1}^{m}\alpha_j t_i^{\,j}+\sum_{k=1}^{M}\beta_k (t_i-K_k)_+^m+\varepsilon_i$$

Then it can be written as :

$$y_i=\alpha_1 t_i+\cdots+\alpha_m t_i^m+\beta_1 (t_i-K_1)_+^m+\cdots+\beta_M (t_i-K_M)_+^m+\varepsilon_i$$

If it is written as matrix can be follow:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}=\begin{bmatrix} t_1 & \cdots & t_1^m & (t_1-K_1)_+^m & \cdots & (t_1-K_M)_+^m \\ t_2 & \cdots & t_2^m & (t_2-K_1)_+^m & \cdots & (t_2-K_M)_+^m \\ & \vdots & & \vdots & & \vdots \\ t_n & \cdots & t_n^m & (t_n-K_1)_+^m & \cdots & (t_n-K_M)_+^m \end{bmatrix}\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \\ \beta_1 \\ \vdots \\ \beta_M \end{pmatrix}+\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Then it can be written as:

$$\tilde{y}=\mathbf{X}(K_1,\cdots,K_M)\tilde{\delta}+\tilde{\varepsilon}$$

with

$$\tilde{y}=(y_1,y_2,\cdots,y_n)',$$

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

,

$\tilde{\delta}=(\alpha_1,\cdots,\alpha_m,\beta_1,\cdots,\beta_M)'$ and

$\tilde{\varepsilon}=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)'$

2.3. Determine Optimal Knot using General Cross Validation (GCV)

Selection of the optimal knot $K_1,K_2,\cdots,K_M$ is very important in nonparametric regression. The knot is a common fusion point where there are behavioral changes at different intervals (Budiantara, 2006). Therefore, to obtain the optimal spline should be selected the optimal knot point. If the optimal knot point is obtained, it will give the best spline. One of the optimal knot selection methods is Generalized Cross Validation or GCV (Budiantara, 2000). The corresponding spline model corresponding to the optimal knot point is obtained from the smallest GCV value.

The GCV function is defined as:

$$GCV(K_1,K_2,\cdots,K_M)=\frac{MSE(K_1,K_2,\cdots,K_M)}{\left(n^{-1}tr[I-\mathbf{A}(K_1,K_2,\cdots,K_M)]\right)^2} \quad (2.3)$$

where

$$MSE(K_1,K_2,\cdots,K_M)=n^{-1}\sum_{j=1}^{p}\left(y_j-\hat{f}_{(K_1,K_2,\cdots,K_p)}(t_j)\right)^2 \quad (2.4)$$

and $\mathbf{A}(K_1,\cdots,K_M)$ is part of equation of $\hat{y}=\mathbf{A}(K_1,\cdots,K_M)y$. (2.5)

3.Results and Discussions

Statistics descriptive of the Jakarta Composite Index from July 12th 2016 to 2017 has mean is 5756.6; variance is 164803.9; minimum is 5027.7 and maximum is 6689.3. The scatter plot of the data as follow as in Figure 1. Based on Figure 1, it shows that the plot has trend increased model, but in period arround 384, has declined.

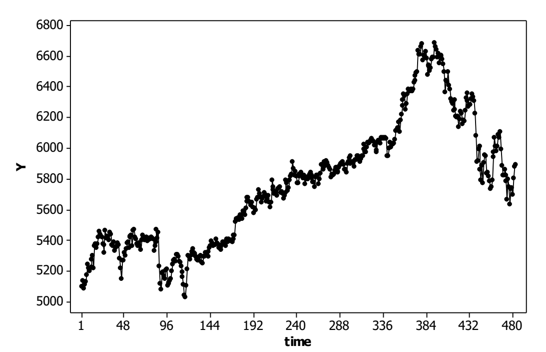

Figure 1. Scatter plot of the data

First step for modelling spline regression is determine the knots. To determine the knots, we could used GCV. We find the minimum of GCV, it shows that optimum knot. We find for one up three orde of spline, then we get one to three knots.

Table 1 shows that the GCV value of spline.

Table 1. GCV value of spline model

orde	Number of Knot	Value of knot	GCV
1	1	408	23591.6
	2	123; 400	1226.3
	3	56;114;400	11174.89
2	1	363	14819.7
	2	399;398	12197.6
	3	50; 105; 400	11260.02
3	1	329	16224.8
	2	116;306	16466.6
	3	50;99;380	16599.5

Based on Table 1, the best knot for spline model in modelling JCI is linear with three knot. In 2 orde for three knot as well as 1 orde three knot. For 3 orde, it has simililar GCV for all of knots. So the best estimation spline regression model for JCI in one orde for three knots as follow as:

$$y_i = 5229.9 + 3.39t_i - 7.042(t_i - 56) + 8.04(t_i - 114) - 13.5(t_i - 114)$$

With R^2=93.34% and MSE=10944.73, the graph of it is can be shown as in figure 2.

If we would like to compare with anothe degree, such as 2 orde or 3 orde with the same knots as well as one orde can be done.

Then we compare model with others model with same knots but different orde.

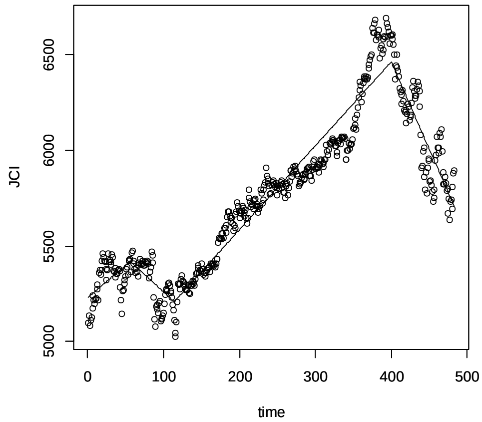

Figure 2. Scatterplot of model spline linaer with three knots

The model of JCI spline quadratic with three knots is follow as:

$$y_i = 5086.23 + 17.57t_i - 0.224t_i^2 + 0.32(t_i - 56) - 0.0998(t_i - 114) - 0.167(t_i - 114)$$

R-square of this model is 90.64 with MSE model is 15393.6. The scatter plot of this model follow as:

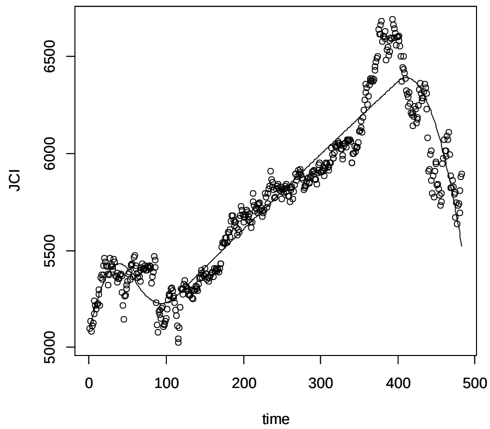

Figure 3. Scatterplot of model spline quadratic with three knots

Then the model spline for cubic with three knots as follow as:

$$y_i = 5043.82 + 25.863t_i - 0.572t_i^2 + 0.0037t_i^3 + 0.0038(t_i - 56) + 5.8\times10^{-5}(t_i - 114) - 0.00087(t_i - 114)$$

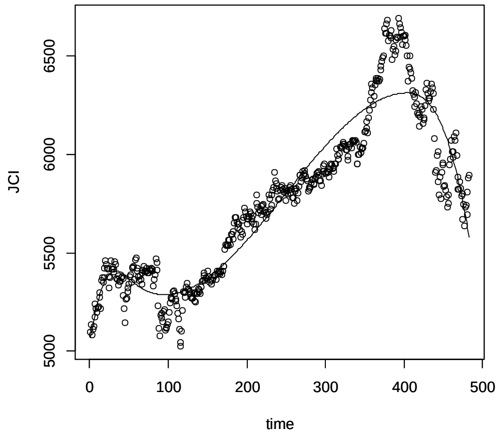

Figure 4. Scatterplot of model spline quadratic with three knots

R-square of spline model cubic with three knots is 89.58 with MSE model 17133.

Based on spline model, the best model for modelling JCI is spline linearwith three knots. Somde of model such as quadratic and cubic with three knots, for same knots with linear, it shows that the models have similiarity results as R square.

References:

1. Eubank, R.L., Huang, C., Maldonado, Y.M., Wang, N., Wang, S., and Buchanan, R.J. (2004). Smoothing Spline Estimation in Varying-Coefficient Models, *J. R. Statist. Soc*., 66, Part 3, pp. 653-667.
2. Hu, Z., Wang, N., dan Carroll, R.J. (2004). Profile Kernel Versus Backfitting In The Partially Linier Models For Longitudinal Or Clustered Data, *Biometrika*, Vol. 91, No. 2, hal. 251-262.
3. Huang, J.Z. (2003). Asymptotics for Polynomial Spline Regression under Weak Conditions, *Statistics and Probability Letters*, 65: 207-216.
4. Ibrahim, N.A. dan Suliadi. (2008). Analyzing Longitudinal Data Using Gee –Smoothing Spline, *Proceedings of the 8th WSEAS International Conference on Applied Computer and Applied Computational science*, hal. 26-33.
5. Lin, X., Wang, N., Wels, A.H., Carroll, R.J. (2004). Equivalent Kernels Of Smoothing Splines In Nonparametrics Regression For

Clustered/Longitudinal Data. *Biometrika*, Vol. 91, No. 1, hal. 177-193.
6. Takezawa, K. (2006). *Introduction to Nonparametric Regression.* New Jersey: John Wiley & Sons,Inc.
7. Wu, H. dan Zhang, J.T. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis*. A John-Wiley and Sons Inc. Publication, New Jersey.